

Další: [Předmluva](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [Základy statistiky pro biomedicínské obory](#)

Obsah

- [Obsah](#)
- [Předmluva](#)
- [1. Úvod do statistické metodologie](#)
 - [1.1 Obsah a význam statistiky](#)
 - [1.2 Populace a výběr](#)
 - [1.3 Výběrová šetření](#)
- [2. Statistika v biomedicínském výzkumu](#)
 - [2.1 Obecné schéma výzkumného projektu](#)
 - [2.2 Plánování a návrh výzkumného projektu, statistické hledisko](#)
 - [2.3 Sběr a zpracování dat](#)
 - [2.4 Analýza, interpretace a prezentace výsledků](#)
 - [2.5 Publikace výsledků výzkumu](#)
- [3. Základy teorie pravděpodobnosti](#)
 - [3.1 Pravděpodobnost a relativní četnost](#)
 - [3.2 Pravidla pro počítání s pravděpodobnostmi](#)
 - [3.3 Bayesův vzorec](#)
 - [3.4 ROC křivka](#)
 - [3.5 Šance, pravděpodobnost a věrohodnost](#)
- [4. Náhodná veličina a rozdělení pravděpodobnosti](#)
 - [4.1 Diskrétní a spojitá náhodná veličina](#)
 - [4.2 Normální rozdělení](#)
 - [4.3 Binomické rozdělení](#)
 - [4.4 Poissonovo rozdělení](#)
- [5. Výběr a popisné statistiky](#)
 - [5.1 Výběrové rozdělení](#)
 - [5.2 Míry polohy](#)
 - [5.2.1 Průměr](#)
 - [5.2.2 Modus](#)
 - [5.2.3 Medián](#)
 - [5.2.4 Geometrický průměr](#)
 - [5.3 Míry variability](#)
 - [5.3.1 Rozpětí](#)
 - [5.3.2 Rozptyl, směrodatná odchylka a variační koeficient](#)
 - [5.3.3 Kvantily, kvartily, decily a percentily](#)
- [6. Odhady populačních průměrů a pravděpodobností](#)
 - [6.1 Rozdělení výběrového průměru](#)
 - [6.2 Intervaly spolehlivosti](#)

- [6.2.1 Intervalový odhad průměru](#)
- [6.2.2 Intervalový odhad populační pravděpodobnosti](#)
- [6.3 Rozsah výběru](#)
- [7. Testování hypotéz](#)
 - [7.1 Nulová a alternativní hypotéza](#)
 - [7.2 Statistická vs. klinická významnost](#)
 - [7.3 Chyba I. a II. druhu a síla testu](#)
 - [7.4 Testování hypotéz o průměru pro jeden výběr](#)
- [8. Porovnání dvou skupin](#)
 - [8.1 Nezávislé výběry](#)
 - [8.1.1 Porovnání průměrů](#)
 - [8.1.2 Interval spolehlivosti pro rozdíl mezi dvěma průměry](#)
 - [8.1.3 Porovnání populačních pravděpodobností](#)
 - [8.2 Párová pozorování](#)
 - [8.2.1 Rozdíl průměrů](#)
- [9. Neparametrické metody](#)
 - [9.1 Základní pojmy](#)
 - [9.2 Kvantilový, mediánový a znaménkový test](#)
 - [9.3 Wilcoxonův párový test](#)
 - [9.4 Mannův-Whitneyův test](#)
- [10. Analýza kategoriálních dat](#)
 - [10.1 \$\chi^2\$ test dobré shody](#)
 - [10.2 Kontingenční tabulky](#)
 - [10.2.1 Test hypotézy o shodnosti struktury](#)
 - [10.2.2 Test hypotézy o nezávislosti](#)
 - [10.2.3 Test hypotézy o symetrii](#)
- [11. Měření závislosti](#)
 - [11.1 Obecné principy](#)
 - [11.2 Lineární regrese a korelace](#)
 - [11.3 Závislost kvalitativních znaků, Spearmanův korelační koeficient](#)
- [12. Analýza rozptylu](#)
 - [12.1 Porovnání průměrů pro tři a více skupin](#)
 - [12.2 Ověřování předpokladů modelu](#)
 - [12.3 Mnohonásobná porovnávání](#)
 - [12.4 Obecné poznámky k analýze rozptylu](#)
- [13. Statistické tabulky](#)
- [Literatura](#)
- [Rejstřík](#)
- [O tomto dokumentu ...](#)

Vytvořeno překladačem L^AT_EX₂HTML.

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [1. Úvod do statistické metodologie](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [Obsah](#)

Předmluva

Publikace "Základy statistiky pro biomedicínské obory" je první částí připravované edice "Biomedicínská statistika". Její vznik byl inspirován mezinárodními kurzy, které probíhaly a probíhají v Evropském centru pro medicínskou informatiku, statistiku a epidemiologii UK a AV ČR. EuroMISE centrum v rámci Univerzity Karlovy spravuje Matematicko-fyzikální fakulta a v rámci Akademie věd ČR Ústav informatiky, v jehož prostorách EuroMISE centrum sídlí. Ve spolupráci s Katedrou pravděpodobnosti a matematické statistiky MFF UK jsou v EuroMISE centru vytvářeny nové přístupy pro rozvoj moderní interdisciplinární výuky biomedicínské statistiky v pregraduální i postgraduální výuce.

V letech 1993-1995 byly mezinárodní kurzy EuroMISE centra podporovány evropským projektem EuroMISE (European Education in Medical Informatics, Statistics and Epidemiology) programu Tempus-Phare. Později pokračovaly za podpory spolupracujících zahraničních institucí, v roce 1997 navíc s podporou Centra pro studium vysokého školství.

První díl edice "Biomedicínská statistika" uvádí se souhlasem spolupracujících zahraničních učitelů také některé příklady, které prezentovali v mezinárodních kurzech EuroMISE centra. Tímto jmenovitě děkuji Prof. Dr. N. Victorovi a Dr. R. Hollemu z univerzity v Heidelbergu a Dr. M. O'Regan a Dr. A. Kellymu z Trinity College v Dublinu.

Cílem publikace je rovněž usnadnit správné využívání statistických metod lékařům, biologům, genetikům, farmaceutům, psychologům a sociologům, kteří statistiku potřebují ve své vědecko-výzkumné práci. Výklad statistických pojmů a metod je prováděn tak, aby mu snadno porozuměli i studenti lékařských, přírodovědeckých a farmaceutických fakult (nepoužívají se matematické prostředky nad rámec středoškolské matematiky). Interpretace výsledků řešených příkladů však pomůže i studentům statistiky snáze proniknout do aplikací statistiky v biomedicínských oborech.

S podporou dalšího evropského projektu 4. rámcového programu IT EDUCTRA (Information Technologies Education and Training) dostala tato publikace kromě knižní podoby i podobu elektronickou. Za velkou práci spojenou s realizací elektronické podoby publikace děkuji zejména Mgr. J. Neustadtovi, Mgr. P. Rošickému a všem ostatním spolupracovníkům z EuroMISE centra UK a AV ČR, kteří přispěli k jejímu vzniku.

Závěrem bych chtěla poděkovat za cenné připomínky k rukopisu recenzentům Prof. MUDr. V. Benckovi, DrSc., Mgr. K. Hrachovi a Ing. J. Machkovi, CSc. Kromě recenzentů za rady pro dílčí úpravy rukopisu děkuji i Doc. RNDr. K. Zvárovi, CSc. a Mgr. J. Svatošovi.

Doc. RNDr. Jana Zvárová, CSc.

Další: [1. Úvod do statistické metodologie](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [Obsah](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Základy statistiky pro biomedicínské obory

Jana Zvárová

- [Obsah](#)
- [Předmluva](#)
- [1. Úvod do statistické metodologie](#)
- [2. Statistika v biomedicínském výzkumu](#)
- [3. Základy teorie pravděpodobnosti](#)
- [4. Náhodná veličina a rozdělení pravděpodobnosti](#)
- [5. Výběr a popisné statistiky](#)
- [6. Odhady populačních průměrů a pravděpodobností](#)
- [7. Testování hypotéz](#)
- [8. Porovnání dvou skupin](#)
- [9. Neparametrické metody](#)
- [10. Analýza kategoriálních dat](#)
- [11. Měření závislosti](#)
- [12. Analýza rozptylu](#)
- [13. Statistické tabulky](#)
- [Literatura](#)
- [Rejstřík](#)
- [O tomto dokumentu...](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [O tomto dokumentu ...](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [Literatura](#)

Rejstřík

A analýza rozptylu

[12.1 Porovnání průměrů pro tři a více skupin](#)

C census

viz [sčítání lidu](#)

Č četnost

absolutní

[2.4 Analýza, interpretace a prezentace výsledků](#) | [3.1 Pravděpodobnost a relativní četnost](#)

kumulativní

absolutní

[2.4 Analýza, interpretace a prezentace výsledků](#)

relativní

[2.4 Analýza, interpretace a prezentace výsledků](#) | [5.3.3 Kvantily, kvartily, decily a percentily](#)

kumulativní

[2.4 Analýza, interpretace a prezentace výsledků](#)

očekávaná

[10.2.1 Test hypotézy o shodnosti struktury](#)

relativní

[2.4 Analýza, interpretace a prezentace výsledků](#) | [3.1 Pravděpodobnost a relativní četnost](#) | [6.2.2 Intervalový odhad populační pravděpodobnosti](#)

člen absolutní

[11.2 Lineární regrese a korelace](#)

D decil

[4.2 Normální rozdělení](#) | [5.3.3 Kvantily, kvartily, decily a percentily](#)

dedukce

[1.1 Obsah a význam statistiky](#)

dokumentace

[2.2 Plánování a návrh výzkumného projektu, statistické hledisko](#)

dotazník

[2.2 Plánování a návrh výzkumného projektu, statistické hledisko](#)

E**efekt**

[7.1 Nulová a alternativní hypotéza](#)

F**funkce****distribuční****empirická**

viz [četnost relativní kumulativní](#)

distribuční

[4.1 Diskrétní a spojitá náhodná veličina](#)

frekvenční

viz [hustota](#)

G**graf****bodový**

[2.4 Analýza, interpretace a prezentace výsledků](#)

Box and Whisker

viz [krabicový](#)

koláčový

viz [kruhový](#)

krabicový

[2.4 Analýza, interpretace a prezentace výsledků](#)

kruhový

[2.4 Analýza, interpretace a prezentace výsledků](#)

normální

[5.3.3 Kvantily, kvartily, decily a percentily](#)

sloupcový

[2.4 Analýza, interpretace a prezentace výsledků](#)

spojnicový

[2.4 Analýza, interpretace a prezentace výsledků](#)

výšečový

viz [kruhový](#)

graf

[2.4 Analýza, interpretace a prezentace výsledků](#)

H

histogram

[2.4 Analýza, interpretace a prezentace výsledků](#)

hladina významnosti

dosažená

[7.1 Nulová a alternativní hypotéza](#)

hladina významnosti

[7.1 Nulová a alternativní hypotéza](#)

hodnota

prediktivní

negativního testu

[3.3 Bayesův vzorec](#)

pozitivního testu

[3.3 Bayesův vzorec](#)

předpovídaná

[12.2 Ověřování předpokladů modelu](#)

střední

viz [průměr](#)

hustota

[4.1 Diskrétní a spojitá náhodná veličina](#)

hypotéza

alternativní

[7.1 Nulová a alternativní hypotéza](#)

nulová

[7.1 Nulová a alternativní hypotéza](#)

**Ch
chyba**

I. druhu

[7.3 Chyba I. a II. druhu a síla testu](#)

II. druhu

[7.3 Chyba I. a II. druhu a síla testu](#)

směrodatná

průměru

[6.1 Rozdělení výběrového průměru](#)

rozdílů průměrů

[8.1.1 Porovnání průměrů](#)

směrnice

[11.2 Lineární regrese a korelace](#)

středníviz [směrodatná](#)**I
indukce**[1.1 Obsah a význam statistiky](#)**interval****konfidenční**viz [interval spolehlivosti](#)**spolehlivosti****pro populační pravděpodobnost**[6.2.2 Intervalový odhad populační pravděpodobnosti](#)**pro průměr**[6.2.1 Intervalový odhad průměru](#)**pro směrnici**[11.2 Lineární regrese a korelace](#)**třídní**[2.4 Analýza, interpretace a prezentace výsledků](#)**J
jev****doplňkový**viz [opačný](#)**jistý**[3.1 Pravděpodobnost a relativní četnost](#)**náhodný**[3.1 Pravděpodobnost a relativní četnost](#)**nemožný**[3.1 Pravděpodobnost a relativní četnost](#)**opačný**[3.2 Pravidla pro počítání s pravděpodobnostmi](#)**jevy****neslučitelné**[3.2 Pravidla pro počítání s pravděpodobnostmi](#)**nezávislé**[3.2 Pravidla pro počítání s pravděpodobnostmi](#)**slučitelné**[3.2 Pravidla pro počítání s pravděpodobnostmi](#)

K**koeficient****determinace**

[11.2 Lineární regrese a korelace](#)

korelační**Pearsonův**

[11.2 Lineární regrese a korelace](#)

Spearmanův

[11.3 Závislost kvalitativních znaků, Spearmanův korelační koeficient](#)

spolehlivosti

[6.2.1 Intervalový odhad průměru](#)

variační

[4.2 Normální rozdělení](#) | [5.3.2 Rozptyl, směrodatná odchylka a variační koeficient](#)

korekce**Sheppardova**

[5.3.2 Rozptyl, směrodatná odchylka a variační koeficient](#)

Yatesova

[10.2.2 Test hypotézy o nezávislosti](#)

kovariance

[11.2 Lineární regrese a korelace](#)

křivka ROC

[3.4 ROC křivka](#)

kvantil

[4.2 Normální rozdělení](#) | [5.3.3 Kvantily, kvartily, decily a percentily](#)

kvartil

[4.2 Normální rozdělení](#) | [5.3.3 Kvantily, kvartily, decily a percentily](#)

M**medián**

[4.2 Normální rozdělení](#) | [5.2.3 Medián](#)

metoda**Bonferroniho**

[12.3 Mnohonásobná porovnávání](#)

nejmenších čtverců

[11.2 Lineární regrese a korelace](#)

metody**observační**

[2.2 Plánování a návrh výzkumného projektu, statistické hledisko](#)

míry**polohy**

[4.2 Normální rozdělení](#) | [5.2 Míry polohy](#)

variability

[4.2 Normální rozdělení](#) | [5.3 Míry variability](#)

modus

[4.2 Normální rozdělení](#) | [5.2.2 Modus](#)

N**negativita nesprávná**

[3.3 Bayesův vzorec](#)

O**odchylka směrodatná**

[4.1 Diskrétní a spojitá náhodná veličina](#) | [5.3.2 Rozptyl, směrodatná odchylka a variační koeficient](#)

odds ratio

viz [podíl šancí](#)

odhad**intervalový**

viz [interval spolehlivosti](#)

sdužený**rozptylu**

[8.1.1 Porovnání průměrů](#)

směrodatné odchylky

[8.1.1 Porovnání průměrů](#)

skupinových efektů

[12.2 Ověřování předpokladů modelu](#)

opora výběru

[1.3 Výběrová šetření](#)

P**papír pravděpodobnostní**

[5.3.3 Kvantily, kvartily, decily a percentily](#)

parametr

[1.2 Populace a výběr](#)

percentil

[4.2 Normální rozdělení](#) | [5.3.3 Kvantily, kvartily, decily a percentily](#)

podíl šancí

[3.5 Šance, pravděpodobnost a věrohodnost](#)

pokus**dvojitě slepý**

[1.2 Populace a výběr](#)

kontrolovaný

[1.2 Populace a výběr](#)

náhodný

[3.1 Pravděpodobnost a relativní četnost](#)

slepý

[1.2 Populace a výběr](#)

pokus

[1.2 Populace a výběr](#)

polygon četností

[2.4 Analýza, interpretace a prezentace výsledků](#)

poměr věrohodností

[3.5 Šance, pravděpodobnost a věrohodnost](#)

populace

[1.2 Populace a výběr](#)

pořadí

[9.1 Základní pojmy](#)

pozitivita nesprávná

[3.3 Bayesův vzorec](#)

pravděpodobnost

aposteriorní

[3.3 Bayesův vzorec](#)

apriorní

[3.3 Bayesův vzorec](#)

podmíněná

[3.2 Pravidla pro počítání s pravděpodobnostmi](#)

potestová

viz [aposteriorní](#)

pretestová

viz [apriorní](#)

pravděpodobnost

[3.1 Pravděpodobnost a relativní četnost](#)

pravidlo

o násobení pravděpodobností

[3.2 Pravidla pro počítání s pravděpodobnostmi](#)

o sčítání pravděpodobností

[3.2 Pravidla pro počítání s pravděpodobnostmi](#)

o úplné pravděpodobnosti

[3.3 Bayesův vzorec](#)

prevalence

[3.3 Bayesův vzorec](#)

procento

viz [četnost relativní](#)

proměnná

nezávisle

[11.2 Lineární regrese a korelace](#)

závisle

[11.2 Lineární regrese a korelace](#)

průměr

aritmetický

viz [průměr \(výběrový\)](#)

geometrický

[5.2.4 Geometrický průměr](#)

vážený

[5.2.1 Průměr](#)

průměr

[4.1 Diskrétní a spojitá náhodná veličina](#)

průměr (výběrový)

[5.2.1 Průměr](#)

přesnost skriningového testu

[3.3 Bayesův vzorec](#)

pyramida věková

[2.4 Analýza, interpretace a prezentace výsledků](#)

R

reziduum

[11.2 Lineární regrese a korelace](#)

rozdělení

alternativní

[4.3 Binomické rozdělení](#)

logaritmicko-normální

[4.2 Normální rozdělení](#)

normální

normované

viz [standardizované](#)

standardizované

[4.2 Normální rozdělení](#)

normální

[4.2 Normální rozdělení](#)

Studentovo t

[6.2.1 Intervalový odhad průměru](#) | [7.4 Testování hypotéz o průměru pro jeden výběr](#)

výběrové četnosti

[2.4 Analýza, interpretace a prezentace výsledků](#)

výběrového průměru[6.1 Rozdělení výběrového průměru](#)**rozhovor**[2.2 Plánování a návrh výzkumného projektu, statistické hledisko](#)**rozpětí****kvartilové**[5.3.3 Kvantily, kvartily, decily a percentily](#)**rozpětí**[5.3.1 Rozpětí](#)**rozptyl****reziduální**[11.2 Lineární regrese a korelace](#)**rozptyl**[4.1 Diskrétní a spojitá náhodná veličina](#) | [5.3.2 Rozptyl, směrodatná odchylka a variační koeficient](#)**S****sčítání lidu**[1.2 Populace a výběr](#)**senzitivita**[3.3 Bayesův vzorec](#)**shody**[9.1 Základní pojmy](#)**síla testu**[7.3 Chyba I. a II. druhu a síla testu](#)**složka náhodná**[11.2 Lineární regrese a korelace](#)**směrnice**[11.2 Lineární regrese a korelace](#)**součet čtverců****celkový**[11.2 Lineární regrese a korelace](#)**regresní**[11.2 Lineární regrese a korelace](#)**reziduální**[11.2 Lineární regrese a korelace](#)**specifická**[3.3 Bayesův vzorec](#)**statistika****induktivní**[1.1 Obsah a význam statistiky](#)

popisná[1.1 Obsah a význam statistiky](#)**testová**[7.1 Nulová a alternativní hypotéza](#)**statistika**[1.1 Obsah a význam statistiky](#) | [1.3 Výběrová šetření](#)**strom života**viz [pyramida věková](#)**stupnice****dichotomická**[1.2 Populace a výběr](#)**intervalová**[1.2 Populace a výběr](#)**logaritmická**[2.4 Analýza, interpretace a prezentace výsledků](#)**nominální**[1.2 Populace a výběr](#)**ordinální**[1.2 Populace a výběr](#)**poměrová**[1.2 Populace a výběr](#)**Š****šance**[3.5 Šance, pravděpodobnost a věrohodnost](#)**šetření****úplné**[1.2 Populace a výběr](#)**výběrové**[1.2 Populace a výběr](#)**šetření**[1.2 Populace a výběr](#)**šíře variací**viz [rozpětí](#)**T****tabulka****analýzy rozptylu**[11.2 Lineární regrese a korelace](#) | [12.1 Porovnání průměrů pro tři a více skupin](#)**čtyřpolní**[10.2 Kontingenční tabulky](#)

kombinační[2.4 Analýza, interpretace a prezentace výsledků](#)**kontingenční**[10.2 Kontingenční tabulky](#)**korelační**[2.4 Analýza, interpretace a prezentace výsledků](#)**prostá**[2.4 Analýza, interpretace a prezentace výsledků](#)**skupinová**[2.4 Analýza, interpretace a prezentace výsledků](#)**test****F**[12.1 Porovnání průměrů pro tři a více skupin](#)**homogenity**[viz test hypotézy o shodnostistruktury](#)**hypotézy****o nezávislosti**[10.2.2 Test hypotézy o nezávislosti](#)**o shodnosti struktury**[10.2.1 Test hypotézy o shodnosti struktury](#)**o symetrii**[10.2.3 Test hypotézy o symetrii](#) **χ^2 dobré shody**[10.1 \$\chi^2\$ test dobré shody](#)**kvantilový**[9.2 Kvantilový, mediánový a znaménkový test](#)**Mannův-Whitneyův**[9.4 Mannův-Whitneyův test](#)**McNemarův**[viz test hypotézy o symetrii](#)**mediánový**[9.2 Kvantilový, mediánový a znaménkový test](#)**t****dvouvýběrový**[8.1.1 Porovnání průměrů](#)**jednovýběrový**[7.4 Testování hypotéz o průměru pro jeden výběr](#)**párový**[8.2.1 Rozdíl průměrů](#)**Wilcoxonův**

dvouvýběrový

viz [Mannův-Whitneyův](#)

párový

[9.3 Wilcoxonův párový test](#)

znaménkový

[9.2 Kvantilový, mediánový a znaménkový test](#)

třídění statistické

[2.4 Analýza, interpretace a prezentace výsledků](#)

U

ukazatel (statistický)

[2.4 Analýza, interpretace a prezentace výsledků](#)

V

variabilita

mezi skupinami

[12.1 Porovnání průměrů pro tři a více skupin](#)

uvnitř skupin

[12.1 Porovnání průměrů pro tři a více skupin](#)

veličina

viz [znak](#)

veličina náhodná

diskrétní

[4.1 Diskrétní a spojitá náhodná veličina](#)

spojitá

[4.1 Diskrétní a spojitá náhodná veličina](#)

výběr

mechanický

[1.3 Výběrová šetření](#)

náhodný

prostý

[1.3 Výběrová šetření](#)

náhodný

[1.3 Výběrová šetření](#)

oblastní

[1.3 Výběrová šetření](#)

reprezentativní

[1.2 Populace a výběr](#)

selektivní

[1.3 Výběrová šetření](#)

skupinový

[1.3 Výběrová šetření](#)

stratifikovaný

viz [oblastní](#)

systematický

viz [mechanický](#)

vícetupňový

[1.3 Výběrová šetření](#)

záměrný

[1.3 Výběrová šetření](#)

výběr

[1.2 Populace a výběr](#)

významnost statistická

[7.1 Nulová a alternativní hypotéza](#)

vzorec Bayesův

[3.3 Bayesův vzorec](#)

Z

znak

kvalitativní

nominální

[1.2 Populace a výběr](#)

ordinální

[1.2 Populace a výběr](#)

kvalitativní

[1.2 Populace a výběr](#)

kvantitativní

diskrétní

[1.2 Populace a výběr](#)

spojitý

[1.2 Populace a výběr](#)

kvantitativní

[1.2 Populace a výběr](#)

znak

[1.2 Populace a výběr](#)

Další: [2. Statistika v biomedicínském výzkumu](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [Předmluva](#)

Kapitola

- [1.1 Obsah a význam statistiky](#)
- [1.2 Populace a výběr](#)
- [1.3 Výběrová šetření](#)

1. Úvod do statistické metodologie

Statistika se těší pochybnému vyznamenání tím, že je nejvíce nepochopeným vědním oborem. Neznamená to však, že je nejméně známá. Nepochopení nějaké věci totiž předpokládá, že se o ní něco ví, nebo přinejmenším se myslí, že se ví. O statistice však panuje všeobecné mínění, že z každého, kdo se naučil ve škole trochu počítat, lze bez obtíží udělat statistika prostě tím, že se mu tak říká.

H. Levinson

1.1 Obsah a význam statistiky

Lékaři i výzkumní pracovníci v biologii a medicíně se často domnívají, že hlubší znalosti statistické metodologie nejsou nezbytné. Vždyť spočítat základní charakteristiky dat a *t* test je jednoduché a vše složitější obstará statistik a počítač. Důvodů, proč si myslíme, že tomu tak není, je několik.

1.

Statistika je v určitém smyslu jazykem pro shromažďování dat, manipulaci s nimi a jejich kvantitativní interpretaci. Lékař v podstatě dělá totéž. Proto by měl tento jazyk také ovládat.

2.

Otázky, které lékař klade, jsou často *statistického charakteru*. Například otázky typu: "Je nový lék lepší než dosud užívaný? U kolika nemocných jej musíme aplikovat, abychom prokázali jeho příznivý efekt?" jsou statistického charakteru. Je proto dobré vědět, jakými metodami se dobrat odpovědi a co přesně výsledek analýzy znamená.

3.

Exploze výpočetní techniky, která zasáhla do zdravotnictví už i u nás, umožňuje také laikům zpracování dat pomocí náročných a donedávna prakticky neproveditelných statistických postupů. Tyto skvělé možnosti mají svou negativní stránku ve zvýšeném nebezpečí výběru nesprávné metody zpracování dat a z toho plynoucích nesprávných závěrů.

4.

V publikovaných článcích v *biomedicínských časopisech* je používání statistických metod pravidlem. Například v roce 1983 Emerson a Colditz při posuzování původních článků publikovaných v *New England Journal of Medicine* ukázali, že 70 % článků obsahovalo nějakou statistickou analýzu. Proto jsou základní znalosti ze

statistiky nezbytné, a to nejen pro vlastní produkci článků, ale i pro jejich studium a kritické hodnocení.

Zamysleme se nad častým znevažováním významu statistiky, například výroky typu: *Jsou tři druhy lži: lež prostá, lež odsouzeníhodná a statistika*, které naznačují, že statistika je zvlášť rafinovanou formou lži. Tento obecně rozšířený názor je však podivuhodným způsobem spojen se zcela opačným postojem, totiž že statistické výsledky se zároveň považují za vrchol nevyvratitelného, neboť mají magické kouzlo matematické přesnosti. Vždyť co může být přesvědčivějšího než číselný údaj, často získaný složitými výpočty na dnešních počítačích?

Kde hledat kořeny tohoto zvláštního postoje ke statistice? Mohou za to především dvě skutečnosti: za prvé nedostatečná znalost cílů, metod a možností statistiky, a za druhé, že za statistiku se pokládá i to, co je ve skutečnosti pseudostatistikou. Protože vládne mylný názor, že pro statistické zpracování dat postačují jen obstojně znalosti z matematiky, stává se tak statistika polem působení mnoha amatérů. Výsledek jejich práce pak vnáší do biologie a medicíny nesprávné statistické závěry. Proto je v posledních letech čím dal větší pozornost v renomovaných lékařských a biologických časopisech věnována i kvalitě statistického zpracování.

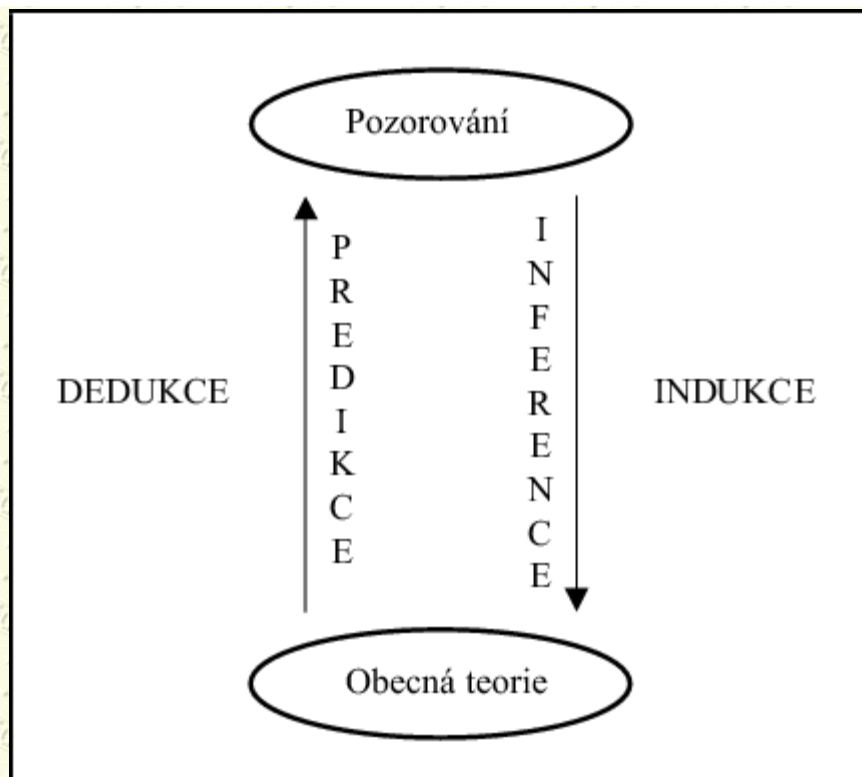
Náš výklad poskytne zájemcům úvod do myšlenkových základů a rutinních postupů statistiky, neboť dnešní statistika se již zdaleka neomezuje na prostý popis nashromážděných dat, ale umožňuje zobecňovat poznatky z dat a podporovat rozhodování.

Původní význam slova **statistika** souvisí se státem, s jeho administrativním spravováním, kdy docházelo k zaznamenávání údajů k vojenským a daňovým účelům. I dnes slovo statistika vyjadřuje popis souboru či jeho sumarizaci, například statistika nemocnosti nebo statistika zdravotnických služeb. V tomto smyslu funguje například Český statistický úřad. Sbírá některé údaje o všech nemocnicích, o všech obcích, dělá úplný popis zvoleného souboru statistických jednotek. **Popisná (deskriptivní) statistika** se zabývá uspořádáním souborů, jejich popisem a účelnou sumarizací.

Problém vědeckého myšlení je spjat s hledáním pravdy. Na čem je vlastně založen náš závěr, že něco je pravdivé? Máme dva hlavní prostředky, které v procesu vytváření vědeckých závěrů používáme. Jsou to hlavně pozorování vnějšího světa a naše schopnosti vytvářet z těchto pozorování zobecňující poznatky. Například při sledování účinku léku na hodnotu krevního tlaku u nemocných s hypertenzní chorobou nelze přeměřit všechny nemocné v naší populaci. Místo celého souboru nemocných se vyšetří jen určitý vzorek nemocných s hypertenzní chorobou (výběr) a z těchto pozorování se snažíme odvodit závěry pro všechny pacienty trpící hypertenzní chorobou (populaci), kteří jsou použitým vzorkem reprezentováni.

Proces zobecňování poznatků, například přenášením závěrů z výběru na celou populaci, nazýváme **induktivním způsobem usuzování (indukcí)**. Induktivní způsob myšlení má prvořadý význam pro postavení člověka v reálném světě. Schopnost učit se ze zkušenosti umožňuje lidem včas se přizpůsobit měnícím se životním podmínkám. Je jedním ze základních typů lidského myšlení, přičemž druhým základním typem je **deduktivní myšlení (dedukce)**. Při deduktivním myšlení z obecných zákonitostí (teorie) činíme závěry (predikce) pro jednotlivé případy (pozorování). Toto myšlení nalézá své uplatnění zejména v matematice. Imponující nezvratnost deduktivních důkazů je však dosahována za cenu toho, že nic nevypovídají o reálném světě. Praktický význam nabyvá deduktivní myšlení jen jako článek myšlenkového řetězce, ve kterém se uplatňují i jiné typy myšlení. Na obrázku [1.1](#) je znázorněn základní rozdíl mezi deduktivním a induktivním myšlením.

Obrázek 1.1: Indukce a dedukce



Závěry induktivních myšlenkových pochodů jsou však ovlivněny subjektivními postoji a mají pouze omezenou platnost. **Induktivní statistika** se zabývá metodami jak takové poznatky přenášet a umožňuje z pozorovaných dat vytvářet obecné závěry s udáním *stupně jejich spolehlivosti*. Výpočet stupně spolehlivosti závěrů je však objektivní, neboť je založen na poznacích teorie pravděpodobnosti a nezávisí na subjektivním názoru hodnotitele.

1.2 Populace a výběr

Hlavní roli v metodách induktivní statistiky hrají dva důležité pojmy - *populace* a *výběr*.

Populace (základní soubor) je zadána přesným stanovením jejích prvků. Prvky jsou buď dány výčtem, nebo vymezením některých společných vlastností. U demografických souborů bývá vymezení dáno příslušností k jisté územní oblasti, věkem, pohlavím, zaměstnáním nebo časovým intervalem sledování (například všechny děti narozené v roce 1990 a k 1. 1. 1998 žijící na území ČR). Vymezená populace může mít *konečný rozsah* (obecně značíme symbolem N), jak je tomu u všech demografických souborů, nebo *nekonečný rozsah*. Nekonečný rozsah mají hypotetické populace, vymezené vlastnostmi svých prvků (například přejeme-li si změřit hladinu cholesterolu ve vzorku krve, pak bychom teoreticky mohli měřit vzorek nekonečněkrát). **Parametr (populační charakteristika)** je číselná charakteristika populace, např. průměrná výška osmiletých dívek v ČR. Je obvykle pokládán za pevné číslo, ale jeho hodnota je obecně neznámá.

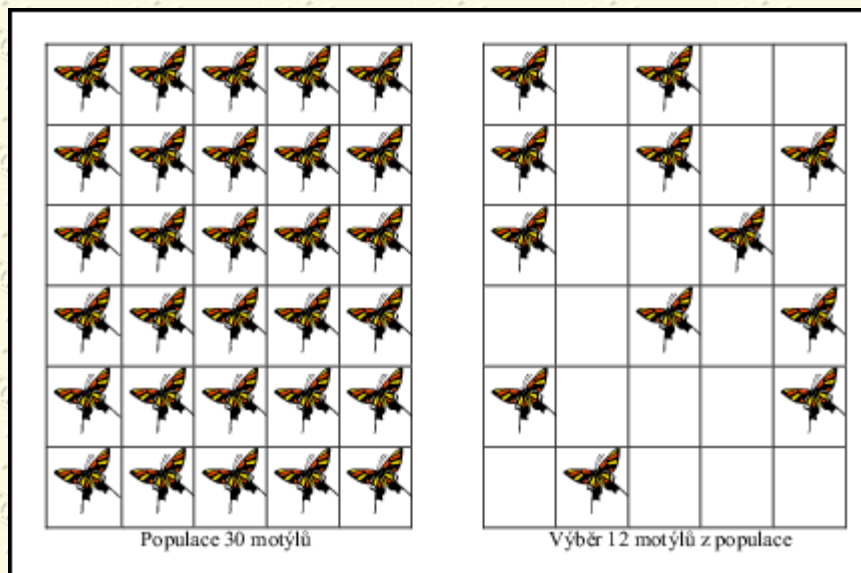
Informace o populaci získáváme prostřednictvím statistického výzkumu. Je užitečné rozlišovat mezi dvěma typy *statistických studií* - pokusem a šetřením.

V **pokus**u plánovitě měníme faktory (či jejich hladiny, úrovně) a sledujeme jejich vliv. Jasným příkladem *klinického pokusu* je srovnávací studie účinků chirurgické a konzervativní léčby určitého onemocnění. Pacienty náhodně rozdělíme do dvou skupin, z nichž jednu léčíme chirurgicky a druhou konzervativně. Výsledky léčby vzájemně

porovnáme. Podobně, pro zjištění účinku vitamínu C na určité virové onemocnění, nemocné náhodně rozdělíme do dvou skupin. Pouze jedné skupině podáváme vitamin C. Tento typ klinického pokusu nazýváme **kontrovaný pokus**. Jestliže lékař při vyhodnocování účinku léčby neví, kteří pacienti vitamin C dostávali, jde o **slepý pokus**. V případě, že ani lékař ani pacienti nevědí, kdo vitamin C dostával (u kontrolní skupiny je podáváno placebo), jde o **dvojitě slepý pokus**.

Při statistickém **šetření** je výzkumný pracovník pouze pasivním pozorovatelem, který zasahuje co nejméně do průběhu šetření. Sledujeme-li znaky u všech prvků (jednotek statistického šetření) populace, provádíme **úplné šetření**. Příkladem úplného šetření na demografických souborech je provádění soupisů, např. soupis pacientů v lůžkových zařízeních. Soupis se provádí vždy k určitému **rozhodnému okamžiku** a je třeba jej pečlivě organizačně připravit. Pro soupis obyvatelstva se vžil název **sčítání lidu (census)** a v našich zemích se provádí v přibližně desetiletých intervalech. Úplné šetření také provádíme při evidenci povinně hlášených nemocí nebo při sledování důležitých demografických jevů, jako je narození nebo úmrtí. Úplné šetření provádíme v praxi poměrně zřídka. Vzhledem ke stanoveným cílům bývá úplné šetření na rozsáhlejších populacích organizačně, ekonomicky a časově tak náročné, že je nelze uskutečnit. Proto zpravidla přistupujeme k **výběrovému šetření**. Při výběrovém šetření zjišťujeme požadované vlastnosti pouze u některých prvků populace, které vytvářejí **výběr** (viz obrázek 1.2). Počet prvků ve výběru neboli rozsah výběru obecně označujeme symbolem n .

Obrázek 1.2: Populace a výběr



Pokud zkoumaný výběr dobře odráží strukturu celého zkoumaného souboru, nazýváme jej **reprezentativním výběrem**. Za určitých předpokladů můžeme závěry z výběrů pomocí **statistické indukce** zobecňovat na celou populaci. Jak již bylo řečeno, induktivní závěry jsou provázeny určitou nejistotou. Metody statistické indukce umožňují však objektivně vyjádřit míru nejistoty zobecněných závěrů pomocí počtu pravděpodobnosti.

Vlastnosti, sledované na prvcích (jednotkách statistického šetření) výběru či populaci, nazýváme **znaky (veličiny)**. Znaky, které sledujeme, dělíme na kvalitativní a kvantitativní. Jsou-li varianty zkoumané vlastnosti dány slovním vyjádřením, jde o **znak kvalitativní**. Jsou-li varianty vyjádřeny číslem, jde o **znak kvantitativní**. Kvantitativní znak je buď **spojitý**, jestliže jednotlivé varianty znaku mohou nabývat jakékoliv hodnoty z určitého intervalu nebo rozmezí (výška, hmotnost apod.), či **diskrétní** (nespojité), jsou-li varianty znaku vyjádřeny oddělenými čísly (počet

onemocnění, počet zemřelých apod.). Kvantitativní znaky měříme na intervalové nebo poměrové stupnici. Tyto stupnice umožňují nejen uspořádat pozorování vzhledem ke stupni vyšetřované vlastnosti, ale navíc určují jejich přesnou vzdálenost.

Na **intervalové stupnici** je nulová pozice věcí volby. Příkladem měření na této stupnici je teplota ve stupních Celsia nebo libovolná kalendářní stupnice.

Nulová pozice na **poměrové stupnici** je pevně dána a vyjadřuje naprostou nepřítomnost měřené vlastnosti. Měřením na této stupnici je hmotnost v kg, příjem v Kč nebo ukazatel úmrtnosti (počet zemřelých na 1 000 exponovaných). Na rozdíl od intervalové stupnice při porovnání dvou hodnot na poměrové stupnici má smysl se ptát, kolikrát je první hodnota větší než druhá.

Pojem měření bývá často rozšiřován i na kvalitativní znaky. K měření kvalitativních znaků používáme nominální nebo ordinální (pořadovou) stupnici. Vlastnosti uvedených dvou stupnic popíšeme nyní podrobněji.

Nominální stupnice sestává ze dvou či více vzájemně se vylučujících kategorií (tříd). V případě, že kategorie jsou právě dvě, stupnice se nazývá **dichotomická**. Příklady kvalitativních znaků měřených na nominální stupnici jsou: *rodinný stav muže* s kategoriemi {svobodný, ženatý, rozvedený, vdovec}; *pohlaví* s kategoriemi {muž, žena}; *operace žaludku* s kategoriemi {resekce pro vřed, resekce pro karcinom, jiná radikální, jiná paliativní}. Pro identifikaci tříd **nominálních znaků** často používáme čísla. Přiřazená čísla jsou ale jen symboly (kódy), které nelze použít pro obvyklé aritmetické operace. Příkladem měření na nominální stupnici je tak i známá mezinárodní klasifikace nemocí, úrazů a příčin smrti. Přiřazená čísla zde jen usnadňují administrativní činnost lékaře a zpracování údajů.

Ordinální stupnice je bohatší než nominální stupnice. Hodnoty znaků jsou seskupeny do neslučitelných kategorií (tříd), které jsou vzájemně uspořádány. Čísla přiřazená kategoriím pro identifikaci odrážejí toto uspořádání, ale neposkytují žádnou informaci o vzdálenosti kategorií. Příklady **ordinálních znaků** jsou: *invalidita* s kategoriemi {úplná, částečná, žádná}; *bolest hlavy* s kategoriemi {často, někdy, zřídka, nikdy}; *spokojenost se zdravotnickými službami* s kategoriemi {velmi spokojen, spokojen, mírně spokojen, mírně nespokojen, nespokojen, velmi nespokojen}.

1.3 Výběrová šetření

Výběry jsou v praxi používány proto, že je jen zřídka možné z důvodů omezeného času či prostředků vyšetřit všechny jednotky uvažované populace. Sčítání lidu, prováděné u nás Českým statistickým úřadem (ČSÚ), je výjimečnou možností vyšetřit celou populaci, i když vyplnění krátkého dotazníku představuje jen velmi nízkou úroveň měření. Pokud by chtěl ČSÚ získat podrobnější informaci o jednotlivcích, dal by přednost práci s výběrem před celou populací, protože výsledná informace by byla dokonce lepší než při zkoumání populace jako celku! Na první pohled to může vypadat poněkud paradoxně, ale po chvíli uvažování se to vyjasní. Zjištění komplexní informace by vyžadovalo vysoce vycvičený sbor tazatelů, který ale pro úplné sčítání lidu není k dispozici. Proto se můžeme o populaci dozvědět více z rozhovorů vedených profesionálními tazateli s reprezentativním výběrem než z rozhovorů s celou populací, které by ale vedli nekompetentní tazatelé.

V některých případech je proces měření destruktivní, takže pokud bychom změřili celou populaci, měli bychom přesnou znalost o ničem, protože bychom ji přitom zničili! Hypotetické populace tvořené měřeními nelze už z podstaty věci vyšetřit úplně - bez ohledu na to, kolik měření jsme provedli na vzorku krve, je to pouze výběr, protože jsme jich mohli, alespoň teoreticky, provést nekonečněkrát více.

Statistika (výběrová charakteristika) je rovněž pojem užívaný pro číselnou charakteristiku výběru, např. průměrná výška 200 náhodně vybraných osmiletých českých dívek. Hodnota téže statistiky bude obvykle v různých výběrech různá.

Výběry pořizujeme metodami náhodného (pravděpodobnostního) výběru, nebo metodami záměrného výběru. Chceme-li z výběru usuzovat na vlastnosti populace, měl by být reprezentativní tj. odrážet svým složením vzhledem ke sledovaným znakům vlastnosti celé populace. Výběr, který není reprezentativní, se nazývá selektivní výběr.

Selektivní výběr dává zkreslený obrázek o studované populaci. Příkladem selektivního výběru je vzorek závodně sportujících volejbalistů ve věku 15-16 let, ze kterého chceme činit závěr o výšce v české populaci chlapců ve věku 15-16 let. Příčina selektivity tkví ve skutečnosti, že ve volejbalu se lépe uplatní větší chlapci, a proto je vybraný vzorek volejbalistů odlišný ve struktuře výšky od chlapců v tomto věku v celé české populaci.

Vybírání prvků z populace pomocí **záměrného výběru** se opírá o expertní stanoviska a různé odhady jak získat reprezentativní výběry. Takto získané výběrové soubory jsou často ovlivněny subjektivními hledisky experta, či dalšími faktory ovlivňujícími pořizované výběry. Přesnost zobecňujících závěrů se spíše opírá o expertní hledisko než o statistickou metodologii.

Metody **náhodného výběru** umožňují vybírat prvky z populace zcela náhodně a nezávisle na našem úsudku. **Oporou (rámcem) výběru** nazýváme technickou dokumentaci, která umožňuje proces vybírání, např. různé seznamy, mapy nebo kartotéky pacientů. Podle způsobu provedení rozlišujeme několik druhů náhodného výběru.

Druhy náhodného výběru

● Prostý náhodný výběr

Prostý náhodný výběr. Provádí se různými technikami losování, které musí zaručit, aby každý prvek populace měl stejnou možnost být zařazen do výběru. Tento postup se usnadní, když prvky populace jsou očíslovány. Pak je možno provést výběr pomocí tabulek náhodných čísel.

● Mechanický výběr

Mechanický (systematický) výběr. Je založen na určitém, předem daném uspořádání prvků populace. Do výběrového souboru zařadíme všechny prvky, které jsou od sebe vzdáleny o zvolený výběrový krok, přičemž první prvek vybereme prostým náhodným výběrem. Například z abecedně uspořádané kartotéky pacientů u praktického lékaře vybíráme s krokem dvacet. První kartu vylosujeme mezi prvními dvaceti kartami, třeba devátou. Potom je výběr tvořen pacienty s pořadím karet 9,29,49,69 atd. Při mechanickém výběru musíme dát pozor, aby uspořádání prvků nesouviselo se sledovaným znakem. Zjišťujeme například znak "zaměstnání" u pacientů stomatologického pracoviště. Pacienti přicházejí podle objednávaných a jejich záznamy jsou časově uspořádány. Provedení výběru s krokem rovným dennímu počtu objednaných pacientů (předpokládejme, že je každý den stejný), může vést k silně selektivnímu výběru. Stane se, že vybereme pouze pacienty, kteří přicházejí k ošetření v určitou denní dobu, což může úzce souviset s typem jejich zaměstnání.

● Oblastní výběr

Oblastní (stratifikovaný) výběr. Studovaná populace je rozdělena do dílčích oblastí. Oblasti jsou vytvořeny tak, aby byly uvnitř homogenní (ve sledovaných znacích se příliš neliší) a mezi sebou heterogenní (sledované znaky se značně liší). Při šetření na obyvatelstvu jsou oblasti vytvořeny např. územními celky, věkovými skupinami nebo socioekonomickým statutem. Z každé oblasti vybereme vzorek metodou prostého náhodného výběru nebo mechanického výběru. Procento vybraných prvků z oblastí může být buď pro

všechny oblasti stejné, nebo se mezi oblastmi liší. Někdy máme pádné důvody vybírat z některé oblasti relativně méně prvků, např. při zvýšených finančních nákladech na šetření nebo obtížné dostupnosti údajů. Konečný výběrový soubor vytvoříme spojením vzorků ze všech oblastí.

• Skupinový výběr

Skupinový výběr. Je-li základní soubor velmi rozsáhlý, čítající statisíce nebo miliony osob, lze prostý náhodný výběr uskutečnit jen velmi obtížně. V tomto případě používáme často skupinového výběru. Nevybíráme jednotlivé osoby, nýbrž celé skupiny osob, které tvoří buď přirozené nebo umělé agregáty. Tyto skupiny mohou být malé i větší (rodina, škola, závod, zdravotní obvod) nebo i značně rozsáhlé (obce, okresy). Je žádoucí, aby skupiny byly pokud možno stejně velké a osoby uvnitř každé skupiny různorodé. Dále se požaduje, aby variabilita mezi skupinami byla co nejmenší, což je tedy obráceně, než je tomu u oblastního výběru. Když byl proveden výběr skupin, lze dále pokračovat dvojím způsobem: buď vyšetříme vyčerpávajícím způsobem všechny osoby vybraných skupin, nebo zvolíme metodu vícestupňového výběru.

• Vícestupňový výběr

Vícestupňový výběr. Je založen na existenci určitého hierarchického popisu prvků základního souboru. K těmto prvkům se postupně dostáváme přes vyšší výběrové jednotky, například: města - bloky - domy - domácnosti; okresy - závody - dílny - zaměstnanci. Každá výběrová jednotka je skupinou výběrových jednotek nižšího řádu. Postupně vybíráme jednotky prvního stupně (primární jednotky), z nich potom jednotky druhého stupně (sekundární jednotky), z nich jednotky třetího stupně atd., až dojdeme k základním jednotkám statistického šetření. Postupné výběry provádíme často metodou prostého náhodného výběru, lze však uplatnit i výběr mechanický nebo oblastní. Vícestupňový výběr je vhodný v situacích, kdy úplná opora výběru není dostupná před začátkem výběrového postupu. Jeho výhody jsou především ekonomického charakteru.

[Další](#) [Výš](#) [Předchozí](#) [Obsah](#) [Rejstřík](#)

Další: [2. Statistika v biomedicínském výzkumu](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [Předmluva](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [3. Základy teorie pravděpodobnosti](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [1. Úvod do statistické metodologie](#)

Kapitola

- [2.1 Obecné schéma výzkumného projektu](#)
- [2.2 Plánování a návrh výzkumného projektu, statistické hledisko](#)
- [2.3 Sběr a zpracování dat](#)
- [2.4 Analýza, interpretace a prezentace výsledků](#)
- [2.5 Publikace výsledků výzkumu](#)

2. Statistika v biomedicínském výzkumu

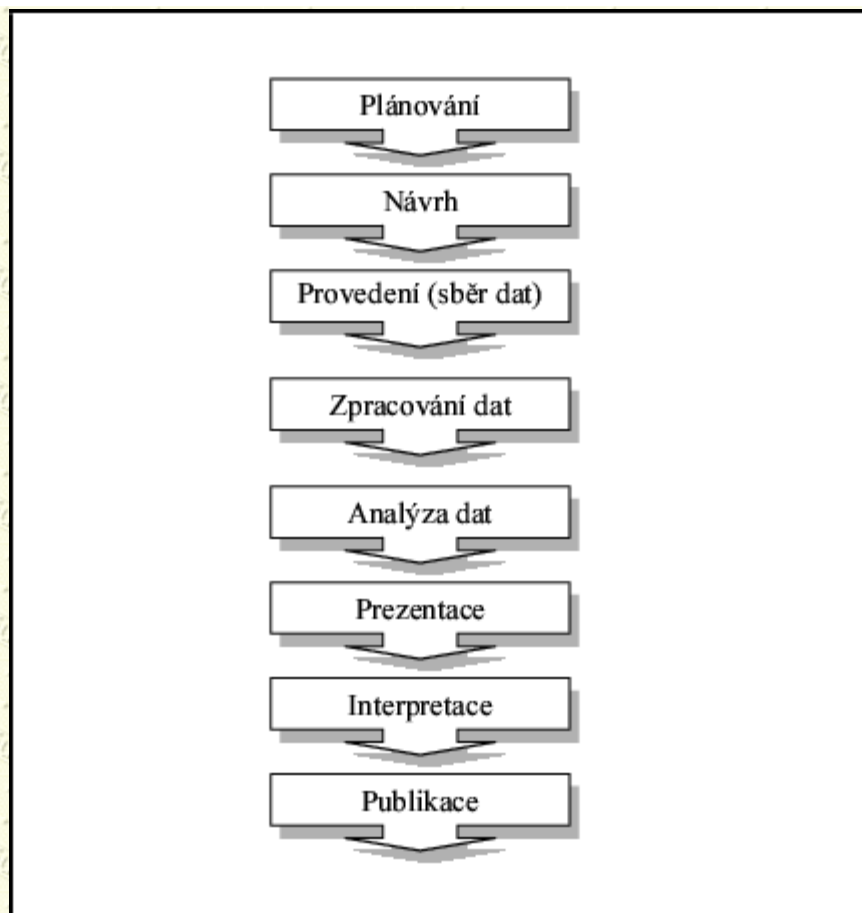
Věda se staví z faktů, podobně jako dům z kamenů. Ale pouhá sbírka faktů je vědou asi tak, jako je hromada kamenů domem.

J. H. Poincare

2.1 Obecné schéma výzkumného projektu

Každý výzkumný projekt lze schematicky popsat pomocí dílčích stadií uvedených na obrázku [1.1](#).

Obrázek 1.1: Obecné schéma dílčích stadií výzkumného projektu



Statistický způsob myšlení může přispět k řešení výzkumného projektu prakticky v každém stadiu. Hlavní roli však statistika sehrává v oblasti návrhu výzkumného projektu, při analýze a interpretaci dat. Všimněme si z hlediska statistiky postupně jednotlivých stadií výzkumného projektu.

2.2 Plánování a návrh výzkumného projektu, statistické hledisko

Je pochopitelné, že v mnoha výzkumných projektech musíme počítat s tím, že nemůžeme studovat celou populaci, která nás zajímá. Například nemůžeme zařadit do výzkumu všechny těhotné ženy či všechny osoby, které žijí v určité geografické oblasti. Pokud by například cílem našeho výzkumu bylo zjišťovat, jaká je souvislost mezi přírůstkem hmotnosti ženy během těhotenství a hmotností novorozence, můžeme údaje zjistit jen na určitém vzorku (výběru) těhotných žen. Cílem výzkumu je potom zobecňovat poznatky z tohoto výběru na všechny těhotné ženy.

Vidíme tedy, že hlavní úkoly při plánování a návrhu výzkumného projektu jsou:

1. Přesná formulace cíle a účelu výzkumu
2. Vymezení pojmů a metod pro
 - studovanou populaci;

- sledované znaky;
- sběr dat;
- statistickou analýzu a technické zpracování dat.

Například lékař zformuluje problém, který podle jeho názoru vyžaduje další výzkum. Je samozřejmé, že se má dobře seznámit s výsledky dřívějšího výzkumu v této oblasti a že má v hrubých rysech formulovat cíl a účel výzkumu.

Uvedme tři příklady takových formulací cíle a účelu výzkumu.

1. Cíl: Vytypovat osoby, které se snadno mohou nakazit chřipkou.
Účel: Provést preventivní opatření.
2. Cíl: Získat poznatky o účincích dvou léků užívaných při léčbě Bechtěrevovy choroby.
Účel: Použít lepší z obou léků.
3. Cíl: Zjistit příčiny vysoké dětské úmrtnosti v dané oblasti.
Účel: Provést potřebná zdravotnická opatření.

Při vytváření přesné formulace cílů výzkumu, kdy se předpokládá užití statistických metod, je vhodné zahájit spolupráci lékaře a erudovaného statistika na samém počátku plánování a návrhu výzkumného projektu. Statistik má být obeznámen s nejdůležitějšími aspekty problému z lékařské stránky a lékař musí získat základní statistické znalosti. Společně pak formulují cíle ve tvaru hypotéz, které mají být kriticky prověřeny pomocí dat. Zásadní chybou je vytvářet hypotézy až podle druhu nashromážděných dat a na stejných datech je prověřovat.

Většinou již při formulaci cíle a účelu výzkumu přesně vymezujeme (místně, časově, věcně) populaci, které se výzkum týká. Někdy však uvažujeme populace čistě hypotetické, kdy vymezujeme populaci pouze vlastnostmi osob, které do populace zařazujeme (např. žena v reprodukčním věku, běloška, krevní skupina A, Rh faktor negativní). V této fázi rovněž rozhodujeme, zda prováděné statistické šetření bude úplné, nebo výběrové. V případě výběrového šetření stanovíme metodu výběru a odhadneme potřebný rozsah výběru. Zvolená metoda výběru předurčuje kvalitu prováděných závěrů. Nežádoucí jsou takové metody, které vedou k vytváření selektivních výběrů, např. výběr pacientů podle úplnosti lékařské dokumentace.

Znaky, které chceme zjišťovat, jsou vybrány s ohledem na stanovené cíle šetření a dané možnosti. Často je obtížné zjistit některé znaky, a musíme se proto spokojit s jinými, které jsou dostupnější a mohou je zcela nebo alespoň částečně nahradit. Každý sledovaný znak musí být přesně a jednoznačně určen. Například pro znak "měsíční příjem" musíme udat, zda sledovaná osoba má uvést hrubý, nebo čistý měsíční příjem, a to z vlastního výdělku, nebo jako průměrný výdělek připadající na jednoho člena rodiny. Dále je třeba určit povahu sledovaných znaků (zda jsou kvalitativní nebo kvantitativní) a zvolit stupnici měření. Volba měrných stupnic ovlivňuje množství informace obsažené v údajích a má rozhodující význam i pro použité metody statistického zpracování.

Metody sběru dat lze zhruba rozdělit do tří kategorií. K **observačním metodám** patří jak přímé vizuální pozorování, tak i složitější metody, které vyžadují speciální znalosti a techniku, např. klinické vyšetření, biochemické vyšetření nebo mikrobiologické vyšetření.

Rozhovor a dotazník shromažďují údaje prostřednictvím záměrně cílených otázek. Získané informace mohou však být zkresleny nepochopením otázek, špatným záznamem odpovědi a při rozhovoru rovněž vlivem sociální interakce.

Použití **dokumentace** je poměrně jednoduchý způsob sběru dat a často jediný, který umožňuje získat informace z minulosti. Zdravotnickému výzkumu běžně slouží dva typy dokumentace: původní zdravotnická dokumentace, jako je záznam o zdraví a nemoci, hlášení o narození dítěte nebo list o prohlídce mrtvého, a údaje rutinní zdravotnické statistiky a rutinních statistik jiných odvětví. Rutinní zdravotnická statistika poskytuje údaje o zdravotním stavu obyvatelstva (o celkové nemocnosti, hospitalizované nemocnosti, nemocnosti spojené s pracovní neschopností, příčinách smrti a příčinách invalidity) a údaje o zdravotnických službách (o síti zdravotnických zařízení, pracovnících ve zdravotnictví, zdravotnických školách a o lékařských a farmaceutických fakultách). Z jiných rutinních statistik používá zdravotnický výzkum zejména demografických, ekonomických a meteorologických údajů.

Při plánování výzkumu je třeba vyjasnit, jakými statistickými metodami se budou nashromážděná data zpracovávat. Úvahy o statistické analýze dat často odhalí nedostatky ve volbě znaků nebo měrných stupnic, navržené metodě výběru apod. Navíc je nezbytné rozhodnout, jakou technikou budou data zpracována. V dnešní době již jde téměř vždy o zpracování na počítači, a musíme tedy vážit i to, jaký software je pro zpracování nejvhodnější. Toto rozhodnutí je důležité pro způsob záznamu údajů.

2.3 Sběr a zpracování dat

Pokud připravujeme sběr velkého množství dat, obvykle nejprve provedeme předběžné šetření (pilotní studii), tj. na malém vzorku ze studované populace ověříme navržené metody pozorování, měření a zpracování údajů. Na základě výsledků předběžného šetření přistupujeme ke sběru dat v plánovaném rozsahu. V tomto stadiu musíme podrobně zaznamenat, proč některé prvky výběru byly vyřazeny ze studie (např. osoby, které odmítly vyplnit dotazník), a ověřit kvalitu získaných údajů. Dále hledáme formální chyby (např. muž uvádí komplikovaný porod). Věcné chyby odstraňujeme zpravidla logickou kontrolou. Výsledky měření mohou být zatíženy náhodnými a systematickými chybami. Náhodné chyby vznikají zejména nepozorností a vedou jak k nadhodnocování, tak i podhodnocování správné hodnoty sledovaného znaku. Systematické chyby jsou způsobeny nejasně položenými otázkami, nesprávným seřízením přístroje apod. Působí pouze jedním směrem.

Při zpracování dat stále větší roli sehrávají počítače. Je mnoho výhod, které vyplývají z možností zpracování dat na počítačích, ale jsou zde i nevýhody. K největší výhodě patří skutečnost, že dnes již můžeme s pomocí počítače aplikovat řadu statistických metod, které dříve pro jejich výpočetní složitost nebylo možné použít. Ukažme některé další výhody a nevýhody počítačového zpracování.

Výhody počítačového zpracování

● Přesnost a rychlost.

Dobré počítačové programy (software) nám dají velmi rychle správné výsledky. Dřívější ruční zpracování dat bylo často zatíženo aritmetickými chybami a bylo časově velmi náročné.

● Univerzálnost.

Počítače zpřístupňují širokou škálu statistických metod a umožňují provést velmi rychle i rozsáhlé komplexní statistické analýzy.

● Grafika.

Počítače umožňují snadné grafické zobrazení pozorovaných dat a výsledků statistického zpracování.

● Flexibilita.

Velkou výhodou počítačů je, že umožňují rychle provést nové zpracování při změnách v datech či transformaci některých veličin.

- **Nové veličiny.**

Snadno lze vytvářet nové veličiny pomocí požadovaných transformací.

- **Velikost datových souborů.**

Počítače umožňují zpracování velmi rozsáhlých souborů dat pomocí vhodného softwaru, což bylo ještě před deseti lety velmi obtížné.

- **Snadný přenos dat.**

Jakmile se jednou data dostala do počítače, lze je snadno přenést elektronicky (například pomocí Internetu) na jiné místo, nebo mohou být uložena na disketě a zaslána poštou.

Kromě uvedených výhod má však počítačové zpracování i svá úskalí, z nichž některá uvádíme.

Nevýhody počítačového zpracování

- **Chyby v softwaru.**

Ne všechny statistické programy jsou spolehlivé. Některé mohou poskytovat chybné výsledky zpracování, protože programátor udělal chyby při tvorbě programu či neporozuměl statistické metodě. Je dobré používat ty statistické programy, které mají dobrou pověst a jsou používány již dostatečně dlouho, takže byla postupně odstraněna většina jejich chyb. K takovým programům patří například BMDP, SAS, SPSS, STATISTICA, S PLUS, STATGRAPHICS a další.

- **Univerzálnost.**

Univerzálnost byla již uváděna jako výhoda, ale může být i nevýhodou. Protože je nabízena řada statistických metod pro zpracování dat, snadno se stane, že bude k vlastnímu zpracování vybrána nevhodná metoda. Je velmi důležité, aby každý, kdo používá statistický software, si byl vědom úrovně svých statistických znalostí a užíval pouze ty metody, kterým rozumí. Pokud chce aplikovat statistickou metodu, kterou nezná, měl by požádat před zpracováním o radu zkušeného statistika.

- **Černá skříňka.**

Použití počítače nás může vzdalovat od vlastních dat. Statistická analýza se provádí automaticky, nová data se zpracovávají a výsledky se ukládají, aniž by byly posouzeny člověkem. Protože většinou výsledky zachycují jen průměrné efekty, může se zcela ztratit citlivost k individuálním pozorováním.

- **Špatná data plodí špatné závěry.**

Jestliže data jsou nasbírána špatně (například jsou špatně kladené otázky v dotazníku), nelze očekávat, že závěry z takových dat budou správné. Dále mohou být data poškozena tím, že se špatně zpracovávají datové soubory, kde některé údaje chybí, když data jsou chybně vložena do počítače nebo se vyskytly chyby již při samotném sběru dat.

2.4 Analýza, interpretace a prezentace výsledků

Při statistické analýze výsledků zpracování používáme metod popisné (deskriptivní) a induktivní statistiky. Metody popisné statistiky umožňují přehledné uspořádání dat (statistické třídění) a výpočet potřebných ukazatelů.

Statistické třídění slouží k rozdělení prvků ze souboru do skupin (tříd) podle předem určených třídících znaků. Podle počtu třídících znaků rozlišujeme třídění jednostupňové (jeden třídící znak) a třídění víceúrovňové neboli

kombinační (dva a více třídicích znaků). Například třídění novorozenců podle pohlaví je jednostupňové, třídění zemřelých osob podle věku, pohlaví a zaměstnání je třístupňové. *Třídy* při třídění podle kvantitativního znaku jsou určeny pomocí třídních intervalů. *Třídní intervaly* musí pokrýt všechny hodnoty kvantitativního znaku, ale vzájemně se nepřekrývají. *Horní a dolní hranice* třídního intervalu jsou dány číselně a jejich rozdíl je délka třídního intervalu. *Střed* třídního intervalu je většinou určen jako aritmetický průměr hranic intervalu. Jsou-li intervaly stejně dlouhé (ekvidistantní), lze třídění zadat přímo středy třídních intervalů. Počet třídních intervalů se zpravidla pohybuje v rozmezí od pěti do dvaceti. Volba počtu intervalů je provedena s ohledem na rozsah souboru a rozpětí (variační šíři), což je rozdíl mezi největší a nejmenší naměřenou hodnotou kvantitativního znaku. Výsledky třídění zaznamenáváme do statistických tabulek.

Počet prvků zařazených do určité třídy nazýváme *absolutní četnost*. *Relativní četnost* je dána podílem absolutní četnosti k rozsahu výběru. Zpravidla ji vyjadřujeme v procentech. Uspořádané četnosti absolutní nebo relativní podle třídních intervalů vytvářejí *rozdělení výběrové četnosti*. U kvantitativních znaků používáme ještě *kumulativní četnost*, která vznikne postupným načítáním četností. Může být opět *absolutní*, nebo *relativní*. Absolutní kumulativní četnost posledního intervalu je tedy rovna rozsahu souboru a jeho relativní kumulativní četnost je 100 %.

Souhrnnou informaci o sledovaných údajích poskytují statistické *ukazatele*. Charakterizují frekvenci sledovaných jevů, nakupení hodnot měřených znaků v určitých místech, jejich variabilitu apod. Metody induktivní statistiky využívají popisu statistického souboru a umožňují vytvářet zobecňující závěry. Tyto závěry se týkají ověřování formulovaných hypotéz či odhadů neznámých parametrů základního souboru. Jsou doplněny údajem o možné velikosti chyby vyvolané tím, že jsme použili jen části souboru, vybrané více méně náhodně.

Interpretace výsledků výzkumu závisí na typu statistického šetření a je velmi úzce spjata s konkrétně dosaženými výsledky. Spolupráce lékaře či biologa a statistika je v této fázi velmi potřebná, zejména při interpretaci výsledků mnohorozměrných statistických metod.

Prezentace výsledků statistického šetření je dána slovním výkladem, statistickými tabulkami a grafickým znázorněním. Statistické tabulky zaznamenávají výsledky šetření v přehledném a snadno srozumitelném tvaru. Číselné údaje v tabulce jsou uspořádány do vodorovných řádků a svislých sloupců. Obsah sloupců vyjadřuje hlavička tabulky umístěná v prostoru nad sloupci. Obsah řádků vyjadřuje legenda tabulky, uváděná obvykle po levé straně tabulky. Tabulka má mít název, který stručně a výstižně vyjadřuje její obsah. Podle stupně třídění rozeznáváme *prostou tabulku* (viz tab. 1.1), která uvádí netříděná data, dále *skupinovou tabulku* (viz tab. 1.2), ve které jsou data roztríděná podle jednoho znaku a *kombinační tabulku* (viz tab. 1.3), která třídí soubor podle dvou a více znaků. Zvláštním druhem kombinační tabulky je *korelační tabulka* (viz tab. 1.4), kterou užíváme ke studiu závislosti dvou kvantitativních znaků.

Tabulka 1.1: Tělesná výška 27 dětí ve věku 9,5-10 let

Tělesná výška v cm								
130	140	136	141	139	133	149	151	139
136	138	142	127	147	139	135	141	143
132	146	151	146	141	141	131	142	141

Tabulka 1.2: Rozdělení dětí ve věku 9,5 - 10 let podle tělesné výšky (délka třídního intervalu 5 cm)

Střed třídního intervalu	Počet dětí
125	1
130	3
135	4
140	12
145	4
150	3
Celkem	27

Tabulka 1.3: Rozdělení dětí ve věku 9,5 - 10 let podle pohlaví a tělesné výšky (délka třídního intervalu 5 cm)

Střed třídního intervalu	Počet chlapců	Počet dívek	Počet dětí
125	1	-	1
130	1	2	3
135	3	1	4
140	7	5	12
145	1	3	4
150	2	1	3
Celkem	15	12	27

Každé políčko tabulky má být pokud možno vyplněno číslem nebo jiným symbolem. Tato zásada nemusí být vždy dodržena, např. u korelačních tabulek. Vodorovná čárka (-) značí, že se nevyskytl žádný případ, nula s uvedeným počtem desetinných míst znamená, že hodnota je tak malá, že v požadované přesnosti ji lze považovat za nulovou. Tečka (.) se používá v případě, že hodnotu neznáme nebo ji nelze zjistit. Není-li zápis možný z logických důvodů, používáme ležatý křížek (x).

Tabulka 1.4: Vztah mezi věkem manželky a věkem manžela

Věk	Věk manželky
-----	--------------

manžela	15-24	25-34	35-44	45-54	55-64	65-74	75 a více	Celkem
15-24	193	50	1	-	-	-	-	244
25-34	231	1162	108	4	-	-	-	1505
35-44	12	408	977	92	4	-	-	1493
45-54	1	36	320	652	66	3	-	1078
55-64	-	5	37	211	358	34	1	646
65-74	-	1	6	24	105	133	10	279
75 a více	-	-	1	4	10	32	25	72
Celkem	437	1664	1450	987	543	202	36	5317

Grafické znázorňování je velmi účinný způsob, jak prezentovat statistické údaje. Grafy nejsou tak přesné jako tabulky, ale rychle a lépe poskytnou názornou představu o důležitých tendencích a souvislostech. Pomocí grafů můžeme například odhadovat trendy a kolísání časových řad nebo několik časových řad vzájemně srovnávat. Graf však slouží i k znázornění postupů induktivní statistiky, např. sekvenčních rozhodovacích pravidel.

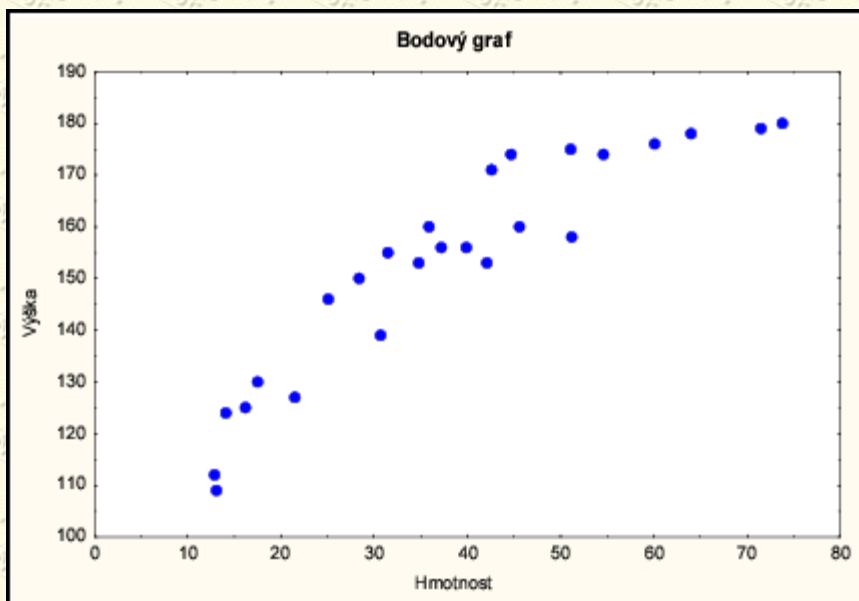
Graf představuje přepsání číselných údajů do soustavy geometrických obrazců. Jejich smysl vykládáme pomocí souřadnic, stupnic a grafické sítě. Základem grafu může být i statistická mapa nebo jiný obrázek. Nositelkou stupnice je čára - přímka nebo křivka. Dostáváme tak stupnice *přímočaré* a *křivočaré*. Z křivočarých stupnic se často setkáváme s kruhovou stupnicí. Na každé stupnici jsou vyznačeny kóty, kterým jsou přiřazena čísla. Vzdálenost mezi dvěma kótami je **grafický interval**, rozdíl mezi jejich číselným označením je **číselný interval**. Poměr mezi grafickým a číselným intervalem se nazývá **modul stupnice**. Podkladem grafického znázornění je většinou soustava souřadnic, v níž horizontální osa (x) se nazývá **osa úseček (abscisa)** a vertikální osa (y) je **osa souřadnic (ordináta)**. Poloha libovolného bodu je určena délkou kolmice k ose x a k ose y. Číselně polohu bodu vyjádříme vzhledem k zvoleným stupnicím na obou osách. Jestliže stejnému číselnému intervalu v libovolném místě stupnice odpovídá stejný grafický interval, jde o **rovnoměrnou stupnici**. U **nerovnoměrné stupnice** stejným číselným intervalům odpovídají nestejně grafické intervaly. Například nerovnoměrná stupnice, pro kterou grafický interval je dán rozdílem logaritmu čísel, se nazývá **logaritmická stupnice**.

Kromě pravoúhlé soustavy souřadnic se používá i **polárních souřadnic**. Tato soustava určuje polohu libovolného bodu pomocí jeho vzdálenosti od počátku a velikosti úhlu, který je měřen od zadaného směru. Uvedeme nyní některé často používané grafy.

Druhy grafů

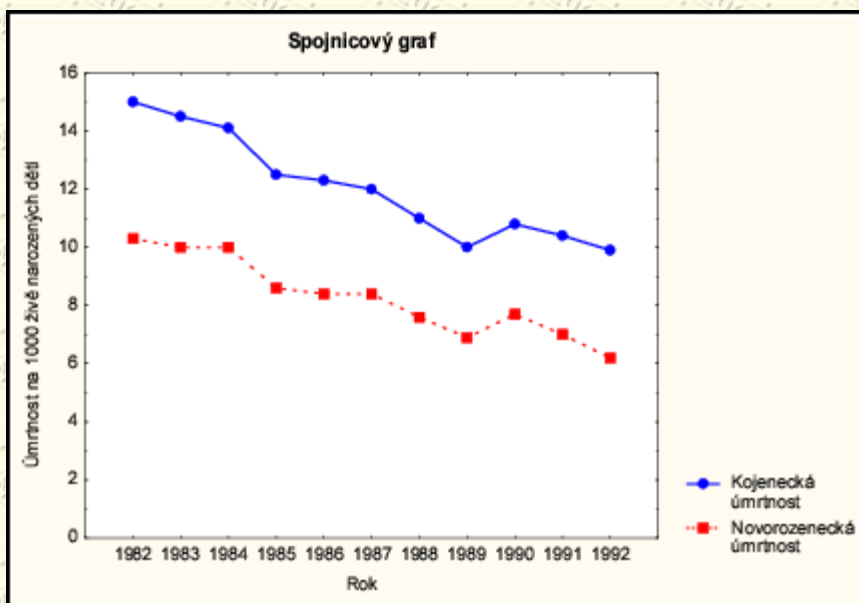
● Bodový graf

Bodový graf (viz obr. 1.2) znázorňuje naměřené hodnoty pomocí bodů v soustavě pravoúhlých souřadnic. Chceme-li v jednom bodovém grafu odlišit hodnoty různých kategorií, použijeme rozdílných symbolů (trojúhelníčky, kroužky, křížky) nebo různých barev. Bodový graf se užívá zejména ke znázornění závislosti dvou znaků.

Obrázek 1.2: Vztah mezi hmotnostmi a výškou u 25 pacientů

• Spojnicový graf

Spojnicový graf (viz obr. 1.3) vyjadřuje velmi často průběh časové řady. Slouží však také ke znázornění rozdělení absolutních nebo relativních četností spojitého znaku a v tomto případě se nazývá *polygon četností*.

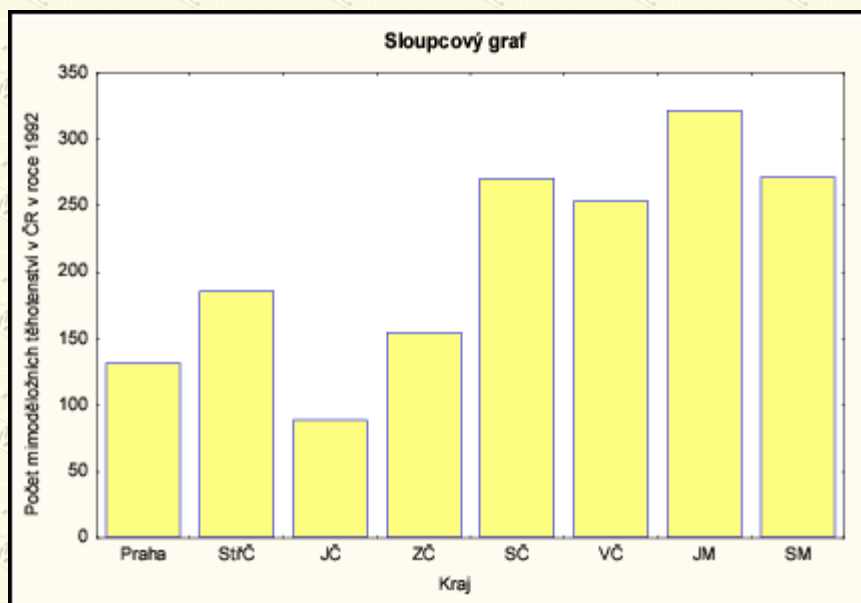
Obrázek 1.3: Kojenecká a novorozenecká úmrtnost na 1 000 živě narozených dětí

• Sloupcový graf

Sloupcový graf (viz obr. 1.4) je velmi rozšířený pro svou názornost a jednoduchost. Číselné hodnoty jsou vyjádřeny pomocí obdélníkových sloupců. Sloupce v grafu obvykle zakreslujeme ve svislé poloze. Ve

vodorovné poloze je umísťujeme v případě, že text ke sloupcům je příliš dlouhý. Chceme-li v grafu současně srovnávat v daném znaku více souborů, můžeme do téže třídy umístit i více sloupců. Sloupce pak odlišujeme barevně nebo různým šrafováním. Při stejné velikosti tříd je šířka sloupců konstantní a výška odpovídá velikosti nebo četnosti znázorňovaného jevu. Při nesteréjné velikosti tříd musí být šířka sloupce úměrná velikosti třídy a plocha odpovídat četnosti.

Obrázek 1.4: Počet mimoděložních těhotenství v ČR v roce 1992 podle kraje

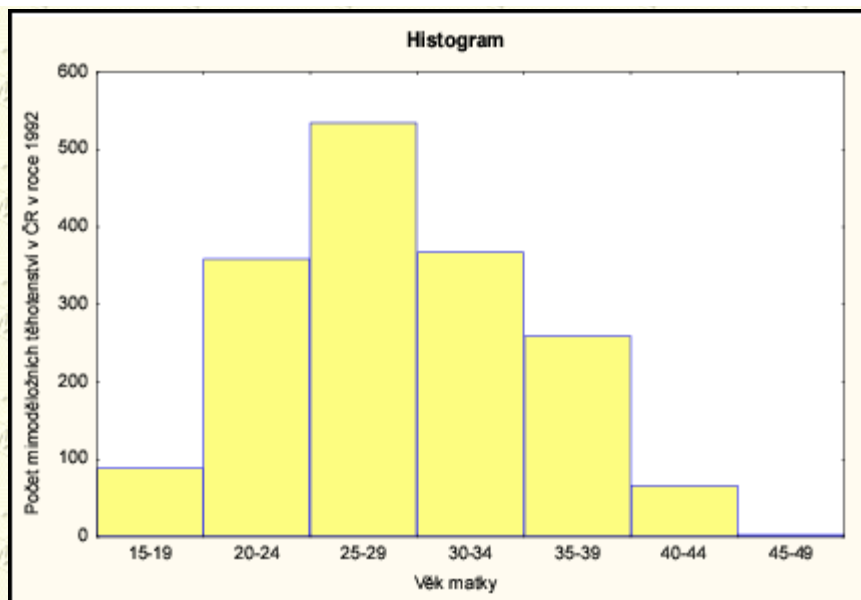


• Histogram

Histogram (viz obr. 1.5) se používá ke znázornění rozdělení absolutních nebo relativních četností spojitého znaku. Je to sloupcový graf, který lze charakterizovat následovně:

1. Sloupce v histogramu jsou vždy vertikální. Jejich výška odpovídá četnosti (absolutní nebo relativní).
2. Stupnice na vodorovné ose grafu je vždy ve stejných jednotkách (obecné sloupcové grafy, které obvykle obsahují kvalitativní veličiny, nemusí mít měřítko základny).
3. Šířka sloupců v histogramu má význam - základna každého sloupce zahrnuje třídu hodnot veličiny. Četnost tedy odpovídá ploše sloupce (tj. šířce sloupce \times výšce).

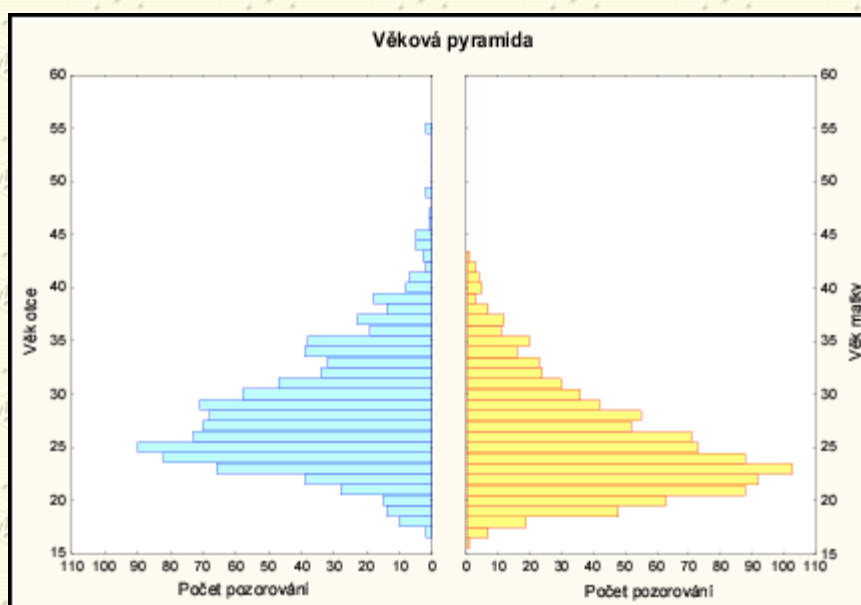
Obrázek 1.5: Počet mimoděložních těhotenství v ČR v roce 1992 podle věku matky



• Věková pyramida

Věková pyramida (strom života) (viz obr. 1.6) znázorňuje věkové složení obyvatelstva. Jedná se opět o zvláštní typ sloupcového grafu.

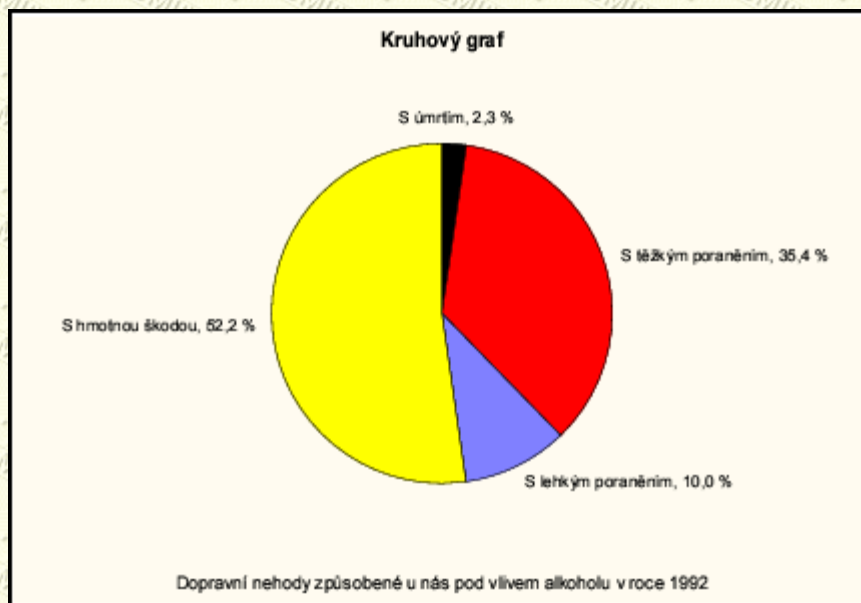
Obrázek 1.6: Věkové složení rodičů v souboru novorozenců



• Kruhový graf

Kruhový (výšečový, koláčový) graf (viz obr. 1.7) zachycuje strukturu souboru. Plocha kruhu představuje celý soubor a jednotlivé části jsou znázorněny kruhovými výsečemi. Protože 360° odpovídá 100 % plochy kruhu, představuje výseč o středovém úhlu $3,6^\circ$ jedno procento. Výseče, které představují jednotlivé složky souboru, odlišujeme různým šrafováním nebo barevně.

Obrázek 1.7: Struktura závažnosti dopravních nehod způsobených u nás v roce 1992 pod vlivem alkoholu



• Krabicový graf

Krabicový graf (Box and Whisker Plot) (viz obr. 5.2) znázorňuje nejčastěji medián (viz kapitola 4), kvartilové rozpětí, nejmenší a největší hodnoty, případně odlehle hodnoty. Obdélník vymezuje kvartily, jeho příčka (nebo čtvereček uvnitř) medián. Tykadla sahají k nejmenšímu a největšímu pozorování. Pokud by však některé tykadlo mělo být delší než 1,5násobek kvartilového rozpětí, sahá jen k tomuto násobku a vzdálenější pozorování jsou vyznačena zvlášť jako odlehle. Někdy však krabicové grafy znázorňují jiné charakteristiky dat. Například místo mediánu je zakreslen průměr, místo kvartilů jsou násobky směrodatné chyby a konce tykadel znázorňují násobek směrodatné odchylky.

2.5 Publikace výsledků výzkumu

Během tohoto století se biomedicínský výzkum velmi rozvinul a jeho závěry stále výrazněji ovlivňují život lidského společenství. Publikace výsledků výzkumu, zejména v předních odborných časopisech, rychle šíří nové poznatky po celém světě. Články publikované v recenzovaných časopisech by měly zaručit jak vědeckou kvalitu, tak i praktický význam publikovaných závěrů. Bohužel, ne vždy to však platí. Příklady nekvalitně navržených výzkumných projektů a nesprávné statistické analýzy nashromážděných poznatků lze nalézt téměř v každém časopise publikujícím výsledky biomedicínského výzkumu. Například Altman [1] poukazuje na to, že chybné užívání či dokonce zneužití statistiky je neetické. Jako obecně etické důsledky nekvalitního biomedicínského výzkumu (nikoliv pouze z hlediska statistiky) uvádí například

- zneužití pacientů tím, že jsou vystaveni nepřiměřenému riziku a nepohodlí,
- zneužití zdrojů, včetně času výzkumných pracovníků, kteří by se mohli věnovat smysluplnějším činnostem,
- důsledky plynoucí z publikování zavádějících závěrů, které mohou vést i k provádění dalších zbytečných studií v budoucnosti.

V extrémních případech se může jednat i o přímý zásah do péče o nemocného. Například se může jednat o zavedení

léčby na základě závěrů o úspěšnosti zkoumané terapie, učiněných z nekontrolovaných studií, i když později randomizované studie ukáží, že terapie je neúčinná. Tyto poznámky vedou ke třem nevyhnutelným závěrům. Za prvé, výzkumný pracovník musí věnovat velkou pozornost plánování, provádění, analýze a interpretaci při výzkumu. Za druhé, musí pečlivě číst a interpretovat publikované výsledky od jiných výzkumných pracovníků, zda pocházejí z pečlivě prováděných a správně analyzovaných výzkumných studií. Za třetí, standard statistických metod v publikovaných pracích může být ovlivněn i vydavatelskou a recenzní politikou různých biomedicínských časopisů. V následující části shrneme některá doporučení, podle čeho hodnotit kvalitu biomedicínských výzkumných prací.

Souhrn

Články publikované v odborných časopisech téměř vždy obsahují souhrn před vlastním textem článku. Souhrn má ukázat cíle výzkumu a informovat čtenáře tak, aby se sám mohl rozhodnout, zda číst celý článek, či nikoliv. Některé časopisy, například *The New England Journal of Medicine*, nyní požadují strukturované souhrny, které obsahují informace o cílech výzkumu, základních postupech, nejdůležitějších poznatcích a hlavních závěrech výzkumu.

Dilema, zda číst celý článek, či nikoliv, mohou pomoci rozřešit odpovědi na dvě otázky:

1.

Pokud byla výzkumná studie správně navržena a analyzována, jsou pro mne její výsledky tak důležité a cenné, abych se s nimi podrobně seznámil?

2.

Pokud výsledky jsou statisticky významné na uvedené hladině, má prokazovaná změna rovněž biologický či klinický význam? Pokud výsledky nejsou statisticky významné, byl rozsah výběru dostatečně velký, aby odhalil požadovaný rozdíl?

V případě, že jsou odpovědi na obě otázky "ano", má smysl se s celým článkem podrobně seznámit.

Úvod

Úvod publikované práce je většinou krátký. Autoři obvykle stručně popíší předchozí biomedicínský výzkum související s danou tematikou a uvedou, proč byla výzkumná studie potřebná. V některých časopisech se však hlavní ospravedlnění, proč byla výzkumná studie provedena, objevuje až v závěru publikací v části "Diskuse". V úvodu by měl být čtenáři sdělen srozumitelně hlavní cíl studie. Čtenář by měl také vědět, zda cíl byl stanoven dříve, než se nashromáždila data, či zda byl formulován, až když byla data nashromážděna. Druhý případ obvykle nastává, když je snaha dále využít náhodně zjištěné nálezy. Nejasně stanovené cíle výzkumu jsou velmi častým důvodem pro zamítnutí článku k publikaci vydavatelem časopisu (Marks et al., 1988). Kromě toho, že je v úvodu stanoven cíl studie, někdy zde nalezneme i další informace o studii, například kde a jak dlouho studie probíhala a jaké subjekty byly do studie zařazeny. Tato informace, která je někdy zařazena až v sekci "Metody", pomáhá čtenáři rozhodnout, zda místo studie a charakter subjektů mohou poskytnout závěry i pro řešení problémů v jeho vlastní práci. Časové období, které studie pokrývá, je také velmi důležité. Například pokud studie určité terapie probíhá příliš dlouho, pacienti vstupující do studie na jejím počátku se mohou značně lišit od těch, kteří vstupují na konci studie.

Metody

Popis použitých metod poskytuje podrobnou informaci o tom, jak byla studie provedena. Již znalost návrhu poskytuje velkou část potřebné informace, která se týká osob, zvířat a materiálu, se kterými se ve studii pracovalo. Použité postupy měření musí být popsány dosti podrobně, aby čtenář věděl, jak byla prováděna. Pokud některé metody měření jsou nové, či vyžadují interpretaci, měla by být sdělena informace o spolehlivosti výsledků.

V metodické části by rovněž měl být uveden rozsah výběru a statistické metody použité při analýze dat. Nejdůležitější na celé výzkumné studii však je, jak byla navržena. Výzkumník totiž může nasbíraná data znovu analyzovat jinou metodou, ale nemůže změnit návrh studie, když jsou již data sesbírána. Všimněme si nyní podrobněji některých problémů souvisejících s metodickou částí výzkumné práce.

Nejprve se zaměříme na to, jak získáváme objekty (například pacienty, zvířata apod.) do studie a jaké zásahy na nich byly prováděny. Například randomizace ve výběru pacientů při zařazování do skupin s různou terapií značně zvyšuje možnost pro zobecňování poznatků a pro nestranné závěry. Připomeňme, že randomizace je přesnou metodou, která zajišťuje, že každý subjekt (nebo léčebný postup) má známou pravděpodobnost, že bude vybrán. Dále potřebujeme znát přesná pravidla pro zařazení subjektu do studie či jeho vyloučení ze studie. V některých studiích, například longitudinálních studiích, se setkáváme s tím, že některé objekty zařazené do studie (například pacienti) v průběhu studie vypadnou. Popis způsobu sledování a důvodu vypadnutí ze studie musí být dostatečně podrobně popsán, aby si čtenář dokázal udělat představu o možných následcích ze ztráty těchto pozorování.

Důležitou informací v metodické části tvoří terminologie, způsob značení a popis použitých metod. Čtenáři mohou, ale nemusí souhlasit s vymezením některých pojmů. Musí rovněž znát podrobný popis metod zpracování dat a to, jak se zacházelo s chybějícími údaji.

V neposlední řadě je třeba stanovit rozsah výběru, který je potřebný k určení předem specifikovaného rozdílu či efektu. Určení rozsahu výběru se vztahuje ke zjištění síly studie. Stanovení síly je důležité zejména v negativních studiích, tedy takových studiích, které nezjistily očekávaný rozdíl nebo souvislost. Doporučujeme proto čtenářům, aby se dále nezabývali výsledky negativních studií, které neposkytují žádnou informaci o síle studie.

Při hodnocení navržených statistických metod přihlížíme k tomu, zda data neodporují předpokladům pro jejich aplikaci. Dále se podíváme, zda studie se nepodobá "rybářské expedici", kdy nejsou jasně stanoveny cíle studie a hledají se jenom "signifikantní" výsledky. Problémem takového přístupu je, že může zachycovat odchylky náhodně vzniklé právě jenom v těchto datech a vede k vytváření závěrů, které se při opakování studie nemusí znovu vyskytnout. Dále musíme při hodnocení výsledků statistických testů přihlížet k problému mnohonásobného porovnávání. Často jsou výsledky nějakého testu neoprávněně interpretovány nezávisle na výsledcích jiných testů, prováděných na stejných datech.

Výsledky

Část publikace shrnující výsledky výzkumu má obsahovat závěry studie odpovídající stanoveným cílům studie. Autoři většinou uvádějí tabulky a grafy sumarizující data a popisující výsledky studie. Jsou zde uváděny výsledky zpracování jak metodami deskriptivní, tak i induktivní statistiky. Další poznatky výzkumu popisují volným textem.

Diskuse a závěry

Diskuse a závěry v publikaci by měly být pro čtenáře snadno srozumitelné. Důležité je, všítat si, zda je konzistence mezi cíli studie popsanými v úvodu, uváděnými výsledky a poznámkami v diskusi. Dále je třeba si všítat, zda autoři extrapolovali své závěry nad rámec analyzovaných dat. Například doporučují lékové dávky, které se v datech studie nevyskytovaly. Všimněme si rovněž, zda autoři upozornili i na slabá místa ve studii a diskutovali i otázky, které ve studii zůstaly nezodpovězeny.

Upozornili jsme na hlavní problémy publikace výsledků biomedicínského výzkumu. Podrobněji se některým z těchto otázek budeme věnovat později pro konkrétní návrhy studií.

[Další](#) [Výš](#) [Předchozí](#) [Obsah](#) [Rejstřík](#)

Další: [3. Základy teorie pravděpodobnosti](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [1. Úvod do statistické metodologie](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [4. Náhodná veličina a rozdělení pravděpodobnosti](#) **Výš:** [Základy statistiky pro biomedicínské obory](#)

Předchozí: [2. Statistika v biomedicinském výzkumu](#)

Kapitola

- [3.1 Pravděpodobnost a relativní četnost](#)
- [3.2 Pravidla pro počítání s pravděpodobnostmi](#)
- [3.3 Bayesův vzorec](#)
- [3.4 ROC křivka](#)
- [3.5 Šance, pravděpodobnost a věrohodnost](#)

3. Základy teorie pravděpodobnosti

Teorie pravděpodobnosti není v podstatě nic jiného než vyjádření obecného povědomí počítáním.

P. S. de Laplace

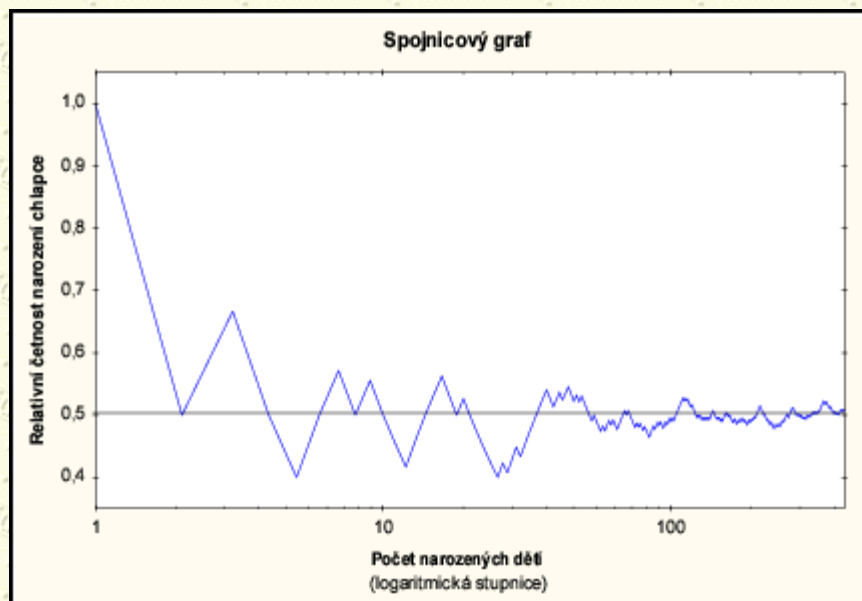
3.1 Pravděpodobnost a relativní četnost

V praxi se často setkáváme s pokusy, jejichž výsledky nejsou jednoznačně předurčeny podmínkami, za kterých probíhají. Takové pokusy, které jsou (alespoň teoreticky) neomezeně mnohokrát opakovatelné, nazýváme **náhodné pokusy**. Jednoduchými příklady náhodných pokusů jsou: házení hracími kostkami nebo mincemi, tahání losů z osudí, míchání karet. I tyto jednoduché náhodné pokusy hrají důležitou roli v mnoha reálných situacích. Setkáváme se s nimi ve výběrových šetřeních, kdy třeba prostý náhodný výběr provádíme vhodnou technikou losování. Můžeme pak určit pravděpodobnosti, s jakými dostáváme výběry strukturou odlišné od struktury základního souboru (populace). Avšak teprve náhodné pokusy jiného druhu činí z počtu pravděpodobnosti důležitou vědu. Téměř všechny experimenty prováděné v biologii a medicíně mají svou náhodnou stránku a počtem pravděpodobnosti se zabýváme proto, abychom ji vhodně vyjádřili či zachytili. Přístupným způsobem pro pracovníky v biomedicínských oborech jsou širší základy teorie pravděpodobnosti sepsány v publikaci K. Zváry a J. Štěpána [22].

Náhodným jevem rozumíme jakékoli tvrzení o výsledku, o kterém lze po uskutečnění pokusu či pozorování rozhodnout, zda je či není pravdivé. Náhodné jevy označujeme velkými písmeny latinské abecedy, ke kterým můžeme připisovat indexy. Zabýváme-li se třeba náhodným jevem A "narození chlapce", můžeme říci o tomto jevu v okamžiku početí poměrně málo, ba ani těsně před porodem není předpověď pohlaví budoucího novorozence příliš spolehlivá. Předpokládejme, že postupně zaznamenáváme pohlaví narozených dětí a dostáváme následující posloupnost: $A, \neg A, A, \neg A, \neg A, A, A, \neg A, \dots$, kde $\neg A$ je náhodný jev, že se chlapec nenarodí, tj. "narození dívky". Četnost, s jakou nastává náhodný jev A pro libovolně dlouhou posloupnost pozorování, můžeme charakterizovat podílem r/n , kde n je délka posloupnosti (rozsah výběru) a r je počet narozených chlapců. Číslo r nazýváme **absolutní četnost** a podíl r/n **relativní četnost** výskytu náhodného jevu A ve výběru o rozsahu n . Grafické znázornění relativních

četností výskytu náhodného jevu A (narození chlapce) v závislosti na rozsahu výběru n je uvedeno na obrázku 3.1. Vidíme, že se vzrůstajícím rozsahem výběru se relativní četnosti ustalují v blízkosti hodnoty 0,5.

Obrázek 3.1: Relativní četnost jevu "narození chlapce" v závislosti na celkovém počtu novorozenců (logaritmická stupnice)



Na základě tohoto příkladu si lze představit, že každý náhodný jev A je charakterizován číslem $P(A)$, které je mírou četnosti výskytu tohoto jevu a nazývá se *pravděpodobnost* náhodného jevu A . V následujících bodech shrneme základní vlastnosti pravděpodobnosti.

- Pravděpodobnost $P(A)$ náhodného jevu A nabývá hodnot mezi nulou a jedničkou, tj. $0 \leq P(A) \leq 1$.

V případě, že A je *jistý jev* (jev A nastává vždycky), je pravděpodobnost $P(A) = 1$. V případě, že A je *nemožný jev* (jev A nikdy nenastane), je pravděpodobnost $P(A) = 0$.

- Při mnohonásobném nezávislém opakování náhodného pokusu je prakticky jisté, že se relativní četnost výskytu náhodného jevu A jen nepatrně liší od pravděpodobnosti $P(A)$.
- Když se pravděpodobnost $P(A)$ jen nepatrně liší od nuly, je prakticky jisté, že při jediném pokusu jev A nenastane.
- Když se pravděpodobnost $P(A)$ jen nepatrně liší od jedné, je prakticky jisté, že při jediném pokusu jev A nastane.

V praxi pravděpodobnosti náhodných jevů odhadujeme pomocí relativních četností. Kvalita těchto odhadů vzrůstá s rostoucím počtem provedených pokusů. Podrobnější informace z teorie pravděpodobnosti přístupnou formou lze nalézt v [22].

Všimněme si statistických údajů, ze kterých odhadujeme pravděpodobnost náhodného jevu "narození chlapce" v naší populaci. V tabulce 3.1 jsou uvedeny relativní četnosti "narození chlapce" určené z celkového počtu živě

narozených dětí v Československu v letech 1966-1975. Je patrné, že relativní četnosti jevu "narození chlapce" se v jednotlivých letech téměř neliší. Lze tedy usuzovat, že ani pravděpodobnost, že narozené dítě bude chlapec, se průběhem let nemění a je o něco vyšší než 0,51.

Tabulka 3.1: Relativní četnosti "narození chlapce" z celkového počtu živě narozených dětí v Československu v letech 1966-1975

Rok	Počet živě narozených dětí	Relativní četnost jevu "narození chlapce"
1966	222 615	0,5131
1967	215 985	0,5146
1968	213 807	0,5138
1969	222 934	0,5145
1970	228 531	0,5125
1971	237 242	0,5134
1972	251 455	0,5133
1973	274 703	0,5144
1974	291 367	0,5135
1975	289 425	0,5108
Celkem	2 448 064	0,5133

3.2 Pravidla pro počítání s pravděpodobnostmi

V praxi většinou nesledujeme výskyt jednoho náhodného jevu, ale zajímáme se o více jevů současně a o jejich vzájemné interakce. Ze souvislosti pravděpodobnosti a relativní četnosti můžeme odvodit základní pravidla pro počítání s pravděpodobnostmi. Vzájemné vztahy mezi náhodnými jevy jsou vyjádřeny v následujících symbolech. Náhodný jev $C = (A, B)$ nastává v případě, že nastanou oba jevy A a B současně. Náhodný jev $D = (A \text{ nebo } B)$ nastává v případě, že nastane alespoň jeden z jevů A a B , tj. buď jev A , nebo jev B , či oba jevy A a B současně (tj. jev C). Pravděpodobnost jevu $D = (A \text{ nebo } B)$ dovedeme vyjádřit pomocí pravděpodobností jevů A , B a $C = (A, B)$ jako

$$P(A \text{ nebo } B) = P(A) + P(B) - P(A, B). \quad (3.1)$$

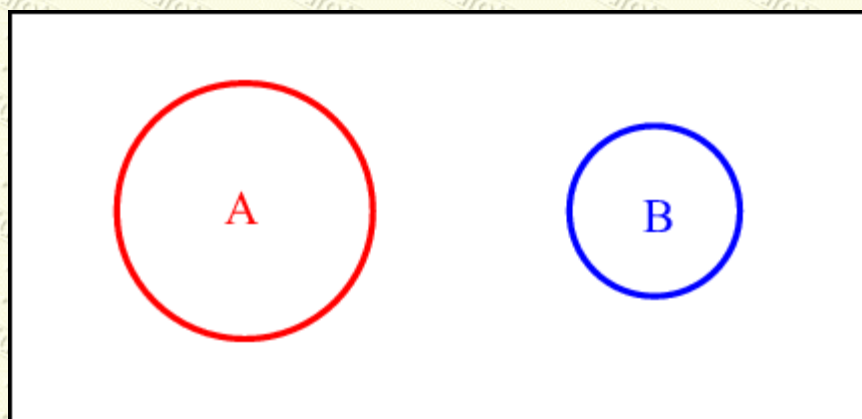
Příklad 3.1 V náhodně vybrané skupině 140 mužů ve věku 40-50 let ohrožených ateriální hypertenzí se vyskytl rizikový faktor "zvýšený cholesterol" (jev A) ve 37 případech a rizikový faktor "kouření" (jev B) v 96 případech. Ve 31 případech jsme zjistili současný výskyt obou rizikových faktorů. Odhadněte pomocí relativních četností pravděpodobnosti výskytu jevů A , B , $C = (A, B)$ a

$$D = (A \text{ nebo } B)$$

Řešení: Pravděpodobnost výskytu faktoru "zvýšený cholesterol" je odhadnuta jako $P(A) = 37/140 = 0,2643$ a faktoru "kouření" jako $P(B) = 98/140 = 0,7000$. Odhad pravděpodobnosti současného výskytu obou faktorů je $P(A, B) = 31/140 = 0,2214$. Odhad pravděpodobnosti výskytu "zvýšeného cholesterolu" nebo "kouření" je $P(A \text{ nebo } B) = 0,2643 + 0,7000 - 0,2214 = 0,7429$. \diamond

Často nám pro objasnění vzájemných souvislostí mezi pravděpodobnostmi náhodných jevů pomáhá jejich grafické znázornění pomocí Vennových diagramů. Na obrázcích 3.2 až 3.4 je znázorněno, že uvnitř obdélníku leží všechny možné výsledky náhodných pokusů či pozorování. Kruh s označením A reprezentuje jen takové výsledky, které vytvářejí jev A , podobně kruh s označením B reprezentuje výsledky, které vytvářejí jev B .

Obrázek 3.2: Neslučitelné jevy



Dva jevy A a B jsou **neslučitelné**, jestliže nemohou nastat oba současně (viz obrázek 3.2). Současný výskyt jevů A a B je vyjádřen jevem $C = (A, B)$. Pro neslučitelné jevy A a B je jev $C = (A, B)$ nemožný a jeho pravděpodobnost je rovna nule, tj. $P(A, B) = 0$. Proto pravděpodobnost výskytu alespoň jednoho ze dvou neslučitelných jevů A a B se rovná součtu jejich pravděpodobností

$$P(A \text{ nebo } B) = P(A) + P(B). \quad (3.2)$$

Tento vztah se nazývá **pravidlem o sčítání pravděpodobností**.

Speciálním případem dvou neslučitelných jevů jsou jevy opačné, např. "narození chlapce" a "narození dívky". Při určování pohlaví novorozence jev "narození chlapce" nastává vždy, když nenastane jev "narození dívky". Obecně rozumíme **opačným (doplňkovým) jevem** k jevu A takový jev (značíme ho $\neg A$), který nastává právě tehdy, když jev A nenastává. Tedy $P(A \text{ nebo } \neg A) = 1$ a z pravidla o sčítání pravděpodobností dostaneme

$$P(\neg A) = 1 - P(A). \quad (3.3)$$

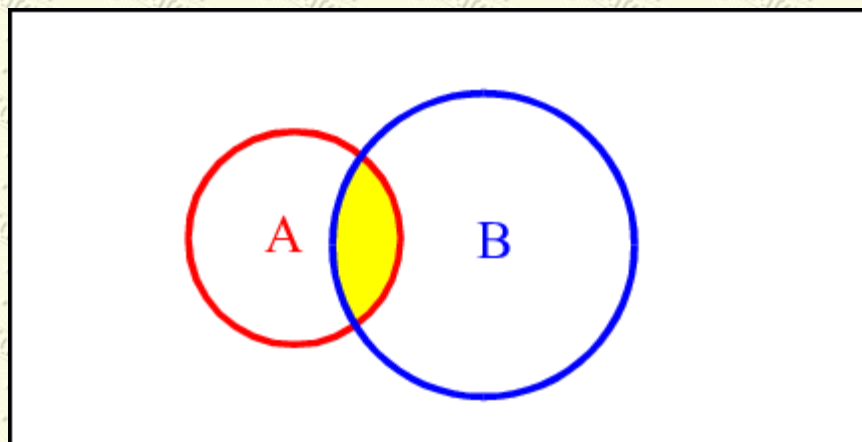
Příklad 3.2 Jestliže pravděpodobnost jevu A "narození chlapce" je rovna $P(A) = 0,51$, spočítejte pravděpodobnost jevu $\neg A$ "narození dívky".

Řešení: Jev $\neg A$ "narození dívky" je opačným jevem k jevu A "narození chlapce". Proto $P(\neg A) = 1 - P(A) = 1 - 0,51 = 0,49$. \diamond

Pravidlo o sčítání pravděpodobností lze snadno rozšířit na libovolný počet vzájemně neslučitelných jevů A_1, A_2, \dots, A_k . Označíme-li D výskyt aspoň jednoho z těchto jevů, tj. $D = (A_1 \text{ nebo } A_2 \text{ nebo } \dots \text{ nebo } A_k)$, potom pravidlo o sčítání pravděpodobností má tvar

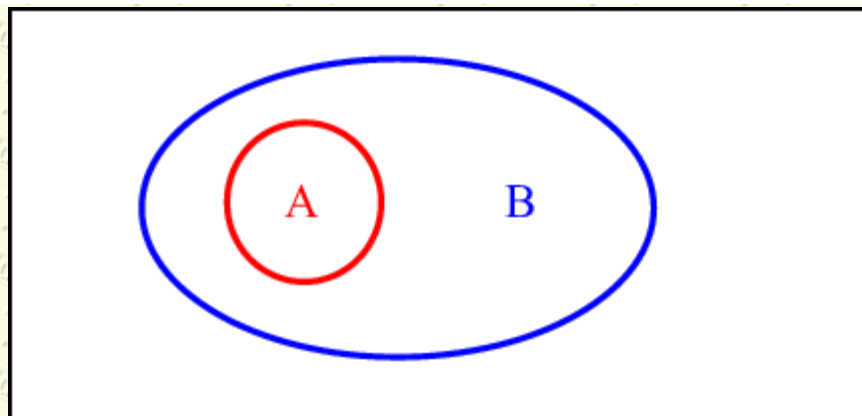
$$P(D) = P(A_1) + P(A_2) + \dots + P(A_k) = \sum_{i=1}^k P(A_i). \quad (3.4)$$

Obrázek 3.3: Slučitelné jevy



Na obrázku 3.3 jsou znázorněny dva jevy A a B , které se mohou vyskytovat současně. Tyto jevy tedy nejsou neslučitelné (tj. jsou *slučitelné*), a proto pro výpočet $P(A \text{ nebo } B)$ nelze použít pravidlo o sčítání pravděpodobností, ale obecnější vzorec (3.1).

Obrázek: Jevy, pro které je podmíněná pravděpodobnost $P(A|B) = 1$



V některých situacích se zajímáme o výskyt jevu A jen v případě, že nastal určitý jev B , který má kladnou pravděpodobnost (tj. může opravdu nastat). Víme-li že nastal jev B , může se tím změnit i pravděpodobnost výskytu jevu A . Všechny jevy neslučitelné s B se stanou nemožnými a jevy deterministicky určené B se stanou jistými (viz obr. 3.4, kde jev A nastává vždy, když nastane jev B). Ostatní jevy se mohou vyskytnout s pravděpodobnostmi, které mohou být odlišné od původních. Pravděpodobnosti jevů, zjištěné za podmínky výskytu jevu B , se nazývají podmíněné pravděpodobnosti vzhledem k jevu B . **Podmíněná pravděpodobnost** jevu A vzhledem k jevu B je definována jako

$$P(A|B) = \frac{P(A, B)}{P(B)}. \quad (3.5)$$

Tudíž pravděpodobnost současného výskytu dvou jevů A a B lze vyjádřit jako $P(A, B) = P(A|B)P(B)$.

Příklad 3.3 Odhadněte podmíněnou pravděpodobnost výskytu faktoru "zvýšený cholesterol" (jev A) za podmínky výskytu faktoru "kouření" (jev B) z údajů uvedených v příkladu 3.1.

Řešení: Odhad podmíněné pravděpodobnosti $P(A|B)$ spočteme, jestliže za pravděpodobnosti jevů A a B dosadíme jejich odhady pomocí relativních četností. Dostaneme

$$P(A|B) = \frac{0,2214}{0,7000} = 0,3163.$$

Stejný výsledek musíme dostat, jestliže z celkového počtu 98 případů, ve kterých nastal jev B , stanovíme počet případů, ve kterých zároveň nastal jev A . Těchto případů je 31. Odhad podmíněné pravděpodobnosti je tedy

$$P(A|B) = \frac{31}{98} = 0,3163$$

a vyjadřuje relativní četnost jevu A mezi případy, kdy nastal jev B . \diamond

Dva jevy A a B jsou **nezávislé**, jestliže výskyt jednoho jevu neovlivňuje výskyt druhého jevu. Matematické vyjádření tohoto faktu zapíšeme pomocí podmíněné pravděpodobnosti jako $P(A|B) = P(A)$ nebo obdobně

$$P(A|B) = P(A)$$

$P(B|A) = P(B)$. Vidíme tedy, že pro nezávislé jevy A, B platí

$$P(A, B) = P(A)P(B). \quad (3.6)$$

Tento vztah se nazývá *pravidlem o násobení pravděpodobnosti*.

Příklad 3.4 Zjistěte, zda faktory A a B uvedené v příkladu 3.1 se vyskytují nezávisle, jestliže vypočtené relativní četnosti považujeme za skutečné pravděpodobnosti.

Řešení: V případě nezávislosti faktorů A a B platí $P(A|B) = P(A)$. Z dat příkladu 3.1 dostáváme $P(A|B) = 0,3163$, což se liší od pravděpodobnosti $P(A) = 0,2643$. Jevy A a B tedy nejsou nezávislé. \diamond

Příklad 3.5 Označme A jev, že "první novorozenec narozený v příštím kalendářním roce v ČR je chlapec" a B jev, že "druhý novorozenec narozený v příštím kalendářním roce v ČR je chlapec". Vyloučíme-li vícečetné porody, spočítejte pravděpodobnost jevu C , že "oba novorozenci jsou chlapci" za předpokladu, že pravděpodobnost narození chlapce je 0,51.

Řešení: Výskyt jevu A neovlivňuje výskyt jevu B , tudíž jevy A a B jsou nezávislé. Pravděpodobnost jevu $C = (A, B)$, že oba novorozenci jsou chlapci, je tedy podle (3.6) rovna

$$P(C) = P(A)P(B) = 0,51 \cdot 0,51 = 0,2601. \quad \diamond$$

Příklad 3.6 Pravděpodobnost jevu A "osoba má pravé oko modré" je rovna 0,3 a pravděpodobnost jevu B "osoba má levé oko modré" je také rovna 0,3. Jestliže pravděpodobnost, že "osoba má pravé oko modré" za podmínky, že nastal jev "osoba má levé oko modré" je rovna 1, spočítejte pravděpodobnost jevu C "osoba má obě oči modré".

Řešení: Jevy A a B nejsou nezávislé, neboť $P(A) = 0,3$ a $P(A|B) = 1$. Proto pravděpodobnost jevu $C = (A, B)$ spočteme jako

$$P(C) = P(A|B)P(B) = 1 \cdot 0,3 = 0,3.$$

Pravděpodobnost, že "osoba má obě oči modré" je tedy rovněž 0,3. \diamond

Pravidlo o násobení pravděpodobností lze rozšířit na libovolný počet nezávislých jevů A_1, A_2, \dots, A_k . Označíme-li C jev, který spočívá v současném výskytu těchto jevů, tj. $C = (A_1, A_2, \dots, A_k)$, potom pravidlo o násobení pravděpodobností má tvar

$$P(C) = P(A_1, A_2, \dots, A_k) = P(A_1)P(A_2) \cdots P(A_k). \quad (3.7)$$

Příklad 3.7 Za předpokladu, že pravděpodobnost narození chlapce je 0,51, spočtěte, jaká je pravděpodobnost, že v sérii čtyř po sobě narozených dětí (vícečetné porody vyloučíme), bude právě jeden chlapec.

Řešení: Označte C jev, že mezi čtyřmi novorozenci je právě jeden chlapec. Konkrétní možnosti, které vytvářejí jev C , jsou dány jevy C_1, C_2, C_3 a C_4 , kde $C_1 = (A, \neg A, \neg A, \neg A)$ je jev, kdy chlapec se narodí jako první, a podobně zbývající jevy $C_2 = (\neg A, A, \neg A, \neg A)$, $C_3 = (\neg A, \neg A, A, \neg A)$ a $C_4 = (\neg A, \neg A, \neg A, A)$ vyjadřují, v jakém pořadí se chlapec narodí. Jevy C_1, C_2, C_3 a C_4 jsou vzájemně neslučitelné. Z pravidla o sčítání pravděpodobností dostaneme

$$\begin{aligned} P(C) &= P(C_1 \text{ nebo } C_2 \text{ nebo } C_3 \text{ nebo } C_4) = \\ &= P(C_1) + P(C_2) + P(C_3) + P(C_4). \end{aligned}$$

Pravděpodobnosti jevů C_1, C_2, C_3 a C_4 jsou všechny stejné a jsou vypočteny pomocí pravidla o násobení pravděpodobností. Například

$$P(C_1) = P(A)P(\neg A)P(\neg A)P(\neg A) = 0,51 \cdot 0,49 \cdot 0,49 \cdot 0,49 = 0,06.$$

Tedy $P(C) = 4 \cdot 0,06 = 0,24$ je pravděpodobnost jevu, že mezi čtyřmi novorozenci bude právě jeden chlapec. \diamond

3.3 Bayesův vzorec

Předpokládejme, že náhodné jevy B_i , kde $i = 1, 2, 3, \dots, k$, jsou vzájemně neslučitelné a v každém pokusu nastává právě jeden z nich, takže musí platit

$$P(B_1 \text{ nebo } B_2 \text{ nebo } \dots \text{ nebo } B_k) = \sum_{i=1}^k P(B_i) = 1.$$

Známe-li podmíněné pravděpodobnosti $P(A|B_i)$ jevu A za podmínky výskytu jevu B_i pro $i = 1, 2, \dots, k$, potom pravděpodobnost jevu A lze vyjádřit vztahem

$$P(A) = \sum_{i=1}^k P(A, B_i) = \sum_{i=1}^k P(A|B_i)P(B_i), \quad (3.8)$$

nazývaným **pravidlo o úplné pravděpodobnosti**.

Příklad 3.8 Předpokládejme, že pravděpodobnost "úrazu" (jev A) u "dítěte" (jev B_1) je $P(A|B_1) = 0,2$, pravděpodobnost úrazu u "osoby v reprodukčním věku" (jev B_2) je _____ a pravděpodobnost úrazu _____

$$P(A|B_2) = 0,1$$

u "osoby v postreprodukčním věku" (jev B_3) je $P(A|B_3) = 0,4$. Pravděpodobnosti, že osoba bude patřit do některé z těchto skupin, jsou $P(B_1) = 0,25$, $P(B_2) = 0,60$ a $P(B_3) = 0,15$. Spočítejte pravděpodobnost úrazu v dané populaci.

Řešení: Jevy B_1 , B_2 a B_3 jsou vzájemně neslučitelné a v každém případě nastává právě jeden z nich. Ze znalosti podmíněných pravděpodobností výskytu úrazu v jednotlivých věkových kategoriích obyvatelstva a ze znalostí pravděpodobností těchto kategorií spočteme pravděpodobnost úrazu v populaci jako

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3) = \\ &= 0,20 \cdot 0,25 + 0,10 \cdot 0,60 + 0,40 \cdot 0,15 = 0,17. \end{aligned}$$

Pravděpodobnost úrazu v populaci je tedy 17%. \diamond

Bayesův vzorec udává, jakým způsobem vypočítáme pravděpodobnosti $P(B_j|A)$ jevu B_j za podmínky, že nastal jev A , jestliže známe *apriorní* pravděpodobnosti $P(B_i)$ a *podmíněné* pravděpodobnosti $P(A|B_i)$ pro všechny jevy B_i , $i = 1, 2, \dots, k$. Bayesův vzorec má tvar

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^k P(A|B_i)P(B_i)}. \quad (3.9)$$

Odvození Bayesova vzorce provedeme snadno pomocí vztahů

$$\begin{aligned} P(B_j, A) &= P(A, B_j) = P(A|B_j)P(B_j), \\ P(B_j|A) &= \frac{P(B_j, A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{P(A)}, \end{aligned}$$

$$\text{kde } P(A) = \sum_{i=1}^k P(A, B_i) = \sum_{i=1}^k P(A|B_i)P(B_i).$$

Příklad 3.9 Pravděpodobnost, že "osoba je kuřák" (jev A) ve skupině "osob s chronickou bronchitidou" (jev B_1) je $P(A|B_1) = 0,75$ a pravděpodobnost, že "osoba je kuřák" ve skupině "osob bez chronické bronchitidy" (jev B_2) je $P(A|B_2) = 0,50$. Pravděpodobnost "výskytu osoby s chronickou bronchitidou" v populaci budiž $P(B_1) = 0,40$ a pravděpodobnost "výskytu osoby bez chronické bronchitidy" v populaci $P(B_2) = 0,60$. Spočítejte pravděpodobnost výskytu chronické bronchitidy u kuřáka.

Řešení: Pomocí Bayesova vzorce dostaneme, že pravděpodobnost výskytu chronické bronchitidy u kuřáka je

$$\begin{aligned}
 P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = \\
 &= \frac{0,75 \cdot 0,40}{0,75 \cdot 0,40 + 0,50 \cdot 0,60} = 0,50. \diamond
 \end{aligned}$$

Bayesův vzorec se často používá v populačních etiologických studiích a v některých matematických modelech diagnostického, terapeutického či prognostického lékařského rozhodování. Bayesův vzorec tak nachází uplatnění v případech, kdy chceme vyhodnotit kvalitu diagnostického testu či skriningového testu (screening test). Implicitní předpoklad pro provádění skriningu je, že včasná detekce nemoci povede k příznivější prognóze nemoci, neboť bude včas zahájena léčba. Některé nemoci nejsou vhodnými kandidáty pro provádění skriningu. Aby byl skrining vhodný, musí být onemocnění závažné a léčba zahájená před rozvinutím příznaků musí být příznivější vzhledem ke snížení mortality či morbidit než v případě, že onemocnění zachytíme již v pokročilém stadiu. Navíc prevalence nemoci v preklinickém stadiu musí být dostatečně vysoká v populaci, na které je skrining prováděn. Dalším problémem skriningu je, jak dobře vyvážit užitek včasné detekce nemoci u osob, které ji skutečně mají, a nepříznivé důsledky, které vniknou tím, že skriningový test určí jako nemocné i ty osoby, které nemoc nemají. Ukážeme dále postupy jak hodnotit kvalitu skriningového testu pro detekci nemoci D .

Označme D^+ jev, že osoba nemoc má, a D^- jev, že osoba sledovanou nemoc nemá. Pozitivní výsledek skriningového testu označme T^+ a negativní výsledek T^- . Pravděpodobnost výskytu nemoci v populaci $P(D^+)$ se nazývá *prevalence (apriorní pravděpodobnost, pretestová pravděpodobnost)*. Naměřené kombinace výsledků skriningového testu pro nemocné a osoby bez sledované nemoci můžeme zapsat do následující tabulky (viz tab. 3.2). Přitom a je počet nemocných osob, u nichž test reagoval pozitivně, a c je počet nemocných osob, u nichž test reagoval negativně. Podobně b je počet osob bez nemoci s pozitivní reakcí na test a d je počet osob bez nemoci s negativní reakcí na test.

Tabulka 3.2: Výsledky skriningového testu

Výsledek skriningového testu	Nemoc		Celkem
	přítomna (D^+)	nepřítomna (D^-)	
T^+	a	b	$a + b$
T^-	c	d	$c + d$
Celkem	$a + c$	$b + d$	n

Senzitivita (sensitivity) a specifita (specificity) jsou dvě míry pro hodnocení skriningového testu.

Senzitivita SE je definována jako pravděpodobnost $P(T^+|D^+)$, že test bude pozitivní u nemocných. Odhaduje se jako

$$SE = \frac{a}{a + c}. \quad (3.10)$$

Specifická SP je definována jako pravděpodobnost $P(T^-|D^-)$, že test je negativní u osob bez nemoci. Odhaduje se jako

$$SP = \frac{d}{b+d}. \quad (3.11)$$

Kromě senzitivity a specifity nás rovněž zajímají pravděpodobnosti, s jakými skrínigový test reaguje negativně u nemocných, resp. pozitivně u osob bez nemoci, nazývané nesprávná negativita (false negativity), resp. nesprávná pozitivita (false positivity).

Nesprávná pozitivita FP diagnostického testu se odhaduje jako

$$FP = \frac{b}{b+d}. \quad (3.12)$$

a **nesprávná negativita FN** diagnostického testu jako

$$FN = \frac{c}{a+c}. \quad (3.13)$$

Vidíme, že součet senzitivity a nesprávné negativity je jedna, tj. $SE + FN = 1$. Podobně i součet specifity a nesprávné pozitivity je jedna, tj. $SP + FP = 1$.

Zřejmě je žádoucí, aby skrínigový test byl vysoce senzitivní a vysoce specifický. Většinou to však není možné, a proto jde o to, jak vyvážit senzitivitu a specifitu skrínigového testu. Nesmíme také zapomínat, že náklady na skrínigový test nejsou jen náklady vztažené přímo k prováděnému skrínigu, ale také náklady, které vznikají vzhledem k dalším procedurám prováděných u těch osob, které ve skrínigovém testu reagovaly pozitivně.

Prediktivní hodnoty (predictive values) skrínigového testu měří, zda osoba podrobená skrínigovému testu je skutečně nemocná.

Prediktivní hodnota pozitivního testu PV^+ je pravděpodobnost $P(D^+|T^+)$, že osoba je opravdu nemocná, když test reagoval pozitivně. Odhadujeme ji jako

$$PV^+ = \frac{a}{a+b}. \quad (3.14)$$

Podobně, **prediktivní hodnota negativního testu PV^-** je pravděpodobnost $P(D^-|T^-)$, že osoba nemá sledovanou nemoc při negativním výsledku testu. Odhaduje se jako

$$PV^- = \frac{d}{c+d} \quad (3.15)$$

Senzitivita a specificita jsou charakteristiky samotného testu. Prediktivní hodnoty jsou ale velmi silně závislé na prevalenci nemoci. Podle Bayesova vzorce můžeme prediktivní hodnoty vyjádřit následovně

$$PV^+ = \frac{SE P(D^+)}{SE P(D^+) + (1 - SP)(1 - P(D^+))} \quad (3.16)$$

a

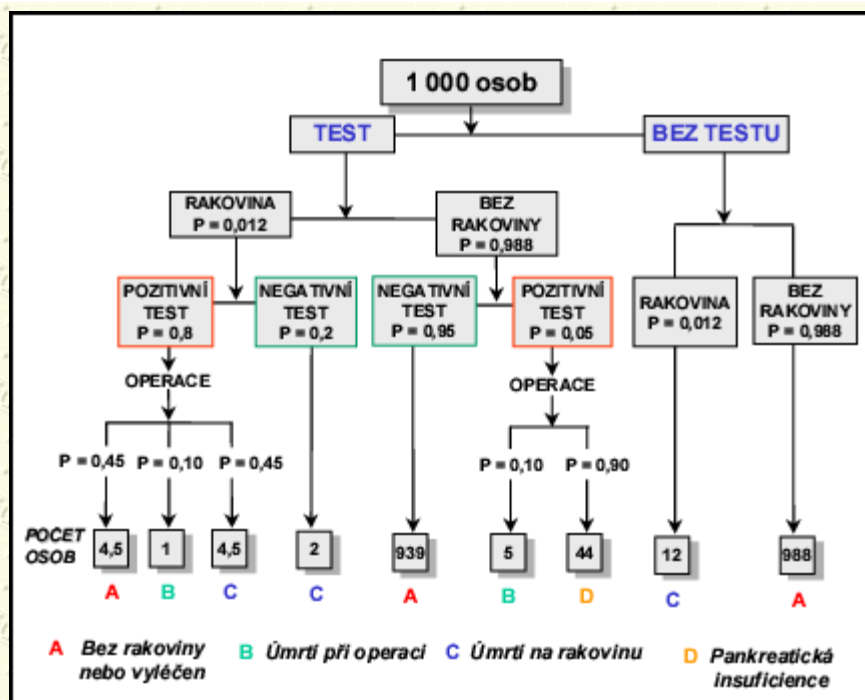
$$PV^- = \frac{SP(1 - P(D^+))}{SP(1 - P(D^+)) + (1 - SE)P(D^+)} \quad (3.17)$$

Například použijeme-li test s 95% senzitivitou a 95% specificitou k předpovědi nemoci, která má prevalenci 1 %, potom pozitivní prediktivní hodnota je $PV^+ = 0,16$. To tedy znamená, že jenom 16 % lidí, u kterých byl test pozitivní, jsou skutečně nemocní, a 84 % z nich nemoc nemá, i když test vyšel pozitivně. Praktický význam diagnostického testu tedy záleží na senzitivitě, specificitě a prevalenci $P(D^+)$, které plně určují prediktivní hodnoty.

Když známe výsledek diagnostického testu, můžeme počítat **aposteriorní (potestovou) pravděpodobnost** výskytu nemoci D^+ . Tyto aposteriorní pravděpodobnosti jsou dány odpovídajícími prediktivními hodnotami, tj. pro pozitivní test PV^+ a pro negativní test PV^- , a jsou předmětem prvořadého zájmu pro kliniky. Dobrý skrínigový či diagnostický test je takový, jehož výsledek zvyšuje kvalitu předpovědi o výskytu nemoci D^+ oproti předpovědi založené pouze na prevalenci nemoci. **Přesnost skrínigového testu** udává pravděpodobnost, s jakou test poskytuje správné závěry v populaci podrobené skrínigu. Odhadujeme ji jako podíl $(a + d)/n$.

Nevhodné užití skrínigového testu na příkladu detekce rakoviny pankreatu uvedli Sisson, Schoomaker a Ross (1976). Skrínig měl na základě výsledku testu poskytnout podklady pro rozhodnutí, zda osoba má či nemá rakovinu pankreatu. Prevalence nemoci v populaci byla $P(D^+) = 0,012$, senzitivita $SE = 0,8$ a specificita $SP = 0,95$. Pokud byla diagnostikována rakovina, byla u pacienta prováděna operace. Operace přinášela pacientovi riziko, že zemře, s pravděpodobností 0,10. Pravděpodobnost, že pacient po operaci stejně zemře na rakovinu, byla 0,45 a zlepšení jeho stavu po operaci nastalo také s pravděpodobností 0,45. V případě, že by byla operována osoba bez rakoviny, pravděpodobnost úmrtí kvůli operaci byla 0,10 a pravděpodobnost přežití po operaci, ale se zhoršeným zdravotním stavem kvůli pankreatické insuficienci, byla 0,90. Výsledky skrínigu jsou ukázány na obrázku 3.5. Vidíme, že pokud by se skrínig neprováděl, potom očekáváme, že z celkového počtu 1000 osob 12 osob zemře na rakovinu. V případě provedení skrínigu u 1000 osob očekávaný počet úmrtí vzroste na 12,5 a navíc 44 osob, které rakovinu pankreatu neměly, bude operováno a projeví se všechny negativní důsledky tohoto zásahu na jejich zdraví. Očekávané počty osob již nemusí být nutně celočíselné hodnoty.

Obrázek 3.5: Výsledky skrínigu



3.4 ROC křivka

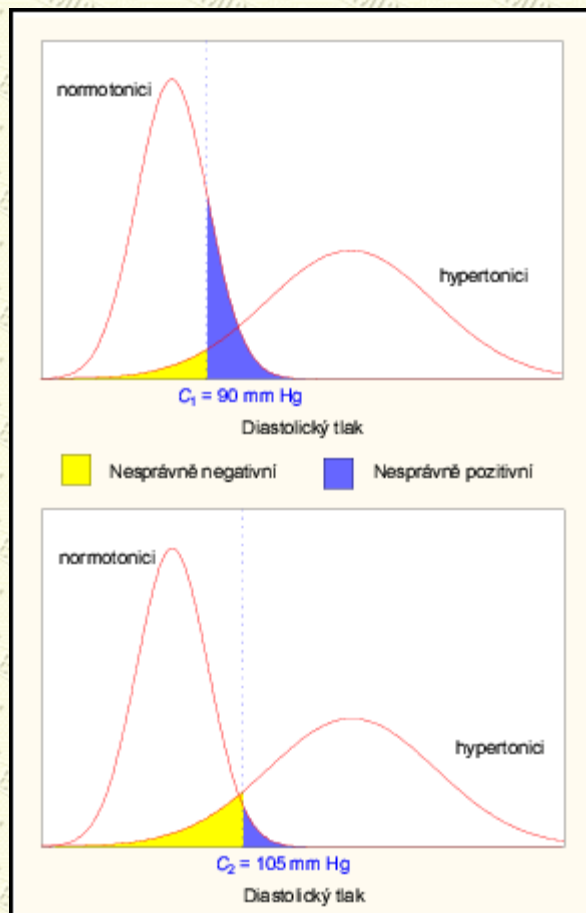
Jak jsme viděli v předcházející části, senzitivita a specifita jsou charakteristiky samotného testu, ale prediktivní hodnoty jsou velmi ovlivněny tím, jak často se nemoc vyskytuje v populaci. Například test s 95% senzitivitou a 95% specificitou, který použijeme pro nemoc s prevalencí 1 %, vykáže 16% prediktivní hodnotu pozitivního testu (tedy 16 % lidí s pozitivním testem má skutečně uvažovanou nemoc), zatímco prediktivní hodnota negativního testu je 99,9 %. To znamená, že u osoby s negativním testem si můžeme být prakticky jisti, že uvažovanou nemoc nemá.

Když uvažujeme o senzitivě a specificitě jako o charakteristikách diagnostického testu, všimněme si podrobněji následujícího příkladu. Chceme zjistit, zda pacient je hypertonik, či normotonik, podle hodnoty jeho diastolického tlaku. Zvolme jako dělicí bod hodnotu $C_1 = 90$ mm Hg. Pacienty s tímto a vyšším tlakem budeme klasifikovat jako hypertoniky, ostatní pacienty jako normotoniky. Diastolický tlak lze považovat za spojitou veličinu s normálním rozdělením (viz 4.2) a jeho rozdělení ve skupině normotoniků a hypertoniků je znázorněno na obrázku 3.6. Na tomto obrázku vidíme vyznačené plochy, které ukazují podíl nesprávně pozitivních závěrů (normotonika zařadíme mezi hypertoniky) a podíl nesprávně negativních závěrů (hypertonika zařadíme mezi normotoniky). Tato situace je znázorněna

- pro dělicí bod $C_1 = 90$ mm Hg,
- pro dělicí bod $C_2 = 105$ mm Hg.

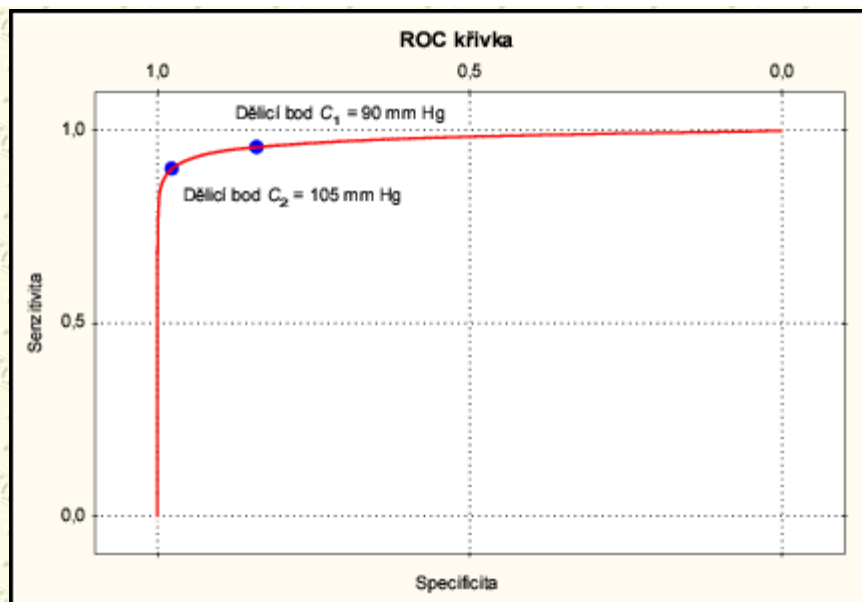
Obrázek 3.6: Grafické zobrazení vlivu dělicí hranice na nesprávně pozitivní a nesprávně negativní závěry: a) při dělicím bodu $C = 90$ mm Hg je vyšší nesprávná pozitivita

testu b) při dělicím bodu $C_2 = 105$ mm Hg je vyšší
nesprávná negativita testu.



Naším cílem je najít takový dělicí bod, abychom docílili vhodné rovnováhy mezi nesprávně pozitivními a nesprávně negativními závěry. Z obrázku 3.6 můžeme snadno porovnat důsledky volby dělicího bodu C_1 oproti bodu C_2 . Volba dělicího bodu je tedy závislá na tom, jaké důsledky přinášejí nesprávná rozhodnutí a jak často se objevují. Proto při rozhodování o volbě dělicího bodu hrají roli i váhy, které přisuzujeme nesprávným rozhodnutím. Na obrázku 3.7 je nakreslena křivka, která pro zvolený dělicí bod dopočítává senzitivitu a specifitu testu. Nazývá se **ROC (Receiver Operating Characteristic) křivka**. (Nepřehlédněte poněkud nestandardní volbu měřítka na vodorovné ose!) Na ROC křivce uvádíme dva body, které přísluší zvoleným dělicím bodům $C_1 = 90$ mm Hg a $C_2 = 105$ mm Hg.

Obrázek 3.7: ROC křivka a dělicí body C_1 a C_2



3.5 Šance, pravděpodobnost a věrohodnost

Řekneme-li že šance $O(A)$ (odds) závodního koně na první místo v dostihovém závodě (jev A) je 1 ku 4, znamená to, že kůň závod vyhraje s pravděpodobností $P(A) = 1/5 = 0,20$. Abychom vyjádření pomocí šance převedli na vyjádření pomocí pravděpodobnosti, sečteme vlastně čísla $1 + 4 = 5$ a dostaneme tak jmenovatel zlomku pro vyjádření pravděpodobnosti výhry, tj. $1/5$.

Pro libovolný náhodný jev A tedy platí: **šance** $O(A)$ výskytu jevu A je

$$O(A) = \frac{P(A)}{P(\neg A)} = \frac{P(A)}{1 - P(A)} \quad (3.18)$$

a

$$P(A) = \frac{O(A)}{1 + O(A)} \quad (3.19)$$

V medicíně často používáme šance pro výpočet **podílu šancí (odds ratio) OR**, který udává podíl šance, že se vyskytne nějaký jev A za určité podmínky (jev B), k šanci, že se jev A vyskytne, když podmínka neplatí (jev $\neg B$). Podíl šancí se tedy vypočte jako

$$OR = \frac{O(A|B)}{O(A|\neg B)} \quad (3.20)$$

kde samozřejmě

$$O(A|B) = \frac{P(A|B)}{P(\neg A|B)} \text{ a } O(A|\neg B) = \frac{P(A|\neg B)}{P(\neg A|\neg B)}.$$

Řekneme-li například, že ve statistické studii o rakovině plic bylo zjištěno, že šance na výskyt rakoviny plic (jev A) u kuřáků (jev B) je 5 ku 4 ($5/4$) a šance na výskyt rakoviny u nekuřáků (jev $\neg B$) je 1 ku 8 ($1/8$), potom podíl šancí je

$$\frac{5/4}{1/8} = \frac{40}{4} = 10,$$

což znamená, že šance dostat rakovinu plic je $10\times$ větší u kuřáků než u nekuřáků.

Příbuzným pojmem k podílu šancí je **věrohodnostní poměr (likelihood ratio) LR** , který udává podíl pravděpodobnosti, že se vyskytne nějaký jev A za určité podmínky (jev B), k pravděpodobnosti, že se jev A vyskytne, když podmínka neplatí (jev $\neg B$). Má-li například pacient náhlou ztrátu paměti (jev A), chceme znát věrohodnostní poměr výskytu jevu A v případě, že má mozkový nádor (jev B), tj. podíl pravděpodobnosti, s jakou ztráta paměti vzniká při nádoru mozku, k pravděpodobnosti, s jakou vzniká v ostatních případech ($\neg B$). Věrohodnostní poměr je tedy podíl podmíněných pravděpodobností

$$LR = \frac{P(A|B)}{P(A|\neg B)}. \quad (3.21)$$

Použití věrohodnostního poměru je možná praktičtější než počítat pravděpodobnost mozkového nádoru. Věrohodnostní poměr užíváme i při hodnocení skriningových a diagnostických testů. Například věrohodnostní poměr pozitivního skriningového testu je dán jako $P(T^+|D^+)/P(T^+|D^-)$. Podobně věrohodnostní poměr negativního testu spočteme jako $P(T^-|D^+)/P(T^-|D^-)$.

[Další](#) [Výš](#) [Předchozí](#) [Obsah](#) [Rejstřík](#)

Další: [4. Náhodná veličina a rozdělení pravděpodobnosti](#) **Výš:** [Základy statistiky pro biomedicínské obory](#)

Předchozí: [2. Statistika v biomedicínském výzkumu](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [5. Výběr a popisné statistiky](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [3. Základy teorie pravděpodobnosti](#)

Kapitola

- [4.1 Diskrétní a spojitá náhodná veličina](#)
- [4.2 Normální rozdělení](#)
- [4.3 Binomické rozdělení](#)
- [4.4 Poissonovo rozdělení](#)

4. Náhodná veličina a rozdělení pravděpodobnosti

*Statistik je ten, kdo s hlavou v rozpálené troubě a s nohama v nádobě s ledem na dotaz, jak se cítí, odpoví:
"V průměru se cítím dobře."*

anonym

4.1 Diskrétní a spojitá náhodná veličina

Většina náhodných pokusů a pozorování prováděných v biologii a medicíně má výsledek vyjádřený reálným číslem. Tato čísla vytvářejí hodnoty reálné náhodné veličiny. Náhodné veličiny označujeme zpravidla velkými písmeny z konce latinské abecedy, například X, Y, Z , a jejich hodnoty odpovídajícími malými písmeny, například x, y, z . K základním charakteristikám náhodné veličiny řadíme průměr (μ) a rozptyl (σ^2) - viz dále. Jak často určité

hodnoty náhodné veličiny nastávají, je exaktně matematicky popsáno pomocí rozdělení pravděpodobnosti. V praxi se zpravidla setkáváme s náhodnými veličinami dvojího typu - diskrétními a spojitými náhodnými veličinami.

Diskrétní náhodná veličina X může nabýt jen konečného x_1, x_2, \dots, x_k nebo spočetného x_1, x_2, x_3, \dots počtu hodnot. Každé hodnotě x_i je přiřazena pravděpodobnost $P(X = x_i) > 0$ a součet těchto pravděpodobností pro všechny hodnoty x_i je roven jedné. Pravděpodobnosti $P(X = x_i)$ charakterizují diskrétní pravděpodobnostní rozdělení. Pro diskrétní náhodnou veličinu X s konečným počtem hodnot spočteme **průměr (střední hodnotu)**

$$\mu = \sum_{i=1}^k x_i P(X = x_i) \quad (4.1)$$

a **rozptyl**

$$\sigma^2 = \sum_{i=1}^k (x_i - \mu)^2 P(X = x_i). \quad (4.2)$$

Druhá odmocnina z rozptylu se nazývá **směrodatná (standardní) odchylka** $\sigma \geq 0$.

V případě, že diskrétní náhodná veličina nabývá početného počtu hodnot, dostáváme obdobně

$$\mu = \sum_{i=1}^{\infty} x_i P(X = x_i) \quad (4.3)$$

a

$$\sigma^2 = \sum_{i=1}^{\infty} (x_i - \mu)^2 P(X = x_i). \quad (4.4)$$

Příklad 4.1 Vylučovatelství skupinově specifických substancí ABH je podmíněno dominantní alelou *Se*, nevylučovatelství je podmíněno recesivní alelou *se*. Jestliže rodiče jsou heterozygotní vylučovatelé (*Se, se*), jejich potomek může být nevylučovatel (*se, se*), homozygotní vylučovatel (*Se, Se*) nebo heterozygotní vylučovatel (*Se, se*) s pravděpodobnostmi uvedenými v tabulce 4.1.

Tabulka 4.1: Vylučovatelství skupinově specifických substancí ABH

Genotyp potomka	Počet alel <i>Se</i>	Pravděpodobnost
Nevylučovatel (<i>se, se</i>)	0	0,25
Heterozygotní vylučovatel (<i>Se, se</i>)	1	0,50
Homozygotní vylučovatel (<i>Se, Se</i>)	2	0,25

Vidíme, že výsledky pozorování lze popsat buď pomocí nominálního znaku "genotyp potomka", rozděleného do tří tříd: nevylučovatel (*se, se*), heterozygotní vylučovatel (*Se, se*), homozygotní vylučovatel (*Se, Se*), či méně podrobně pomocí dichotomického znaku - vylučovatel, nevylučovatel. Můžeme však také sledovat kvantitativní znak "počet alel *Se*", jehož hodnoty lze interpretovat jako výsledky náhodných pozorování. "Počet alel *Se*" je diskrétní náhodná veličina *X* nabývající hodnot $x = 0, 1, 2$ s pravděpodobnostmi $P(X = 0) = 0,25$, $P(X = 1) = 0,5$, $P(X = 2) = 0,25$, tudíž s průměrem (střední hodnotou)

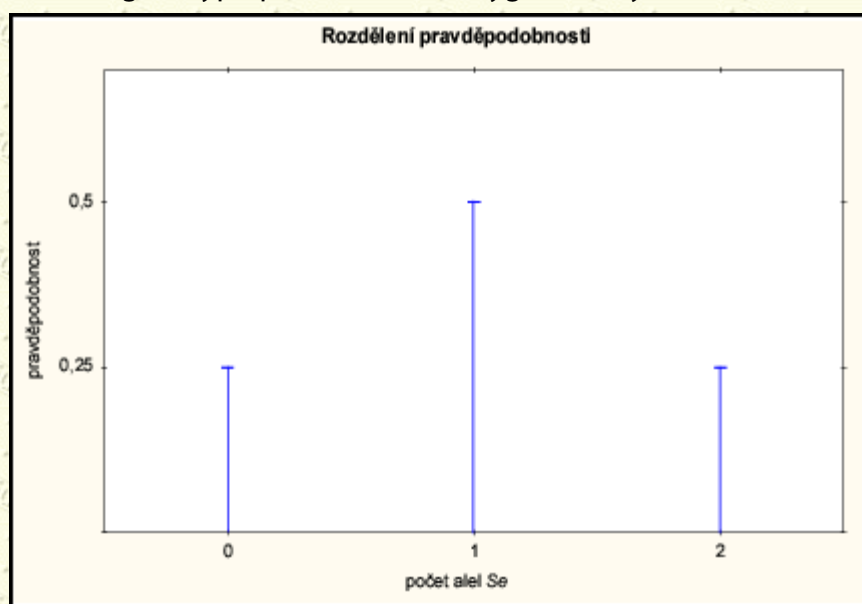
$$\mu = 0 \cdot 0,25 + 1 \cdot 0,50 + 2 \cdot 0,25 = 1$$

a rozptylem

$$\sigma^2 = (0 - 1)^2 \cdot 0,25 + (1 - 1)^2 \cdot 0,5 + (2 - 1)^2 \cdot 0,25 = 0,50.$$

Její pravděpodobnostní rozdělení je znázorněno tyčkovým grafem na obrázku 4.1.

Obrázek 4.1: Rozdělení pravděpodobnosti "počtu alel Se" v genotypu potomka heterozygotních vylučovatelů

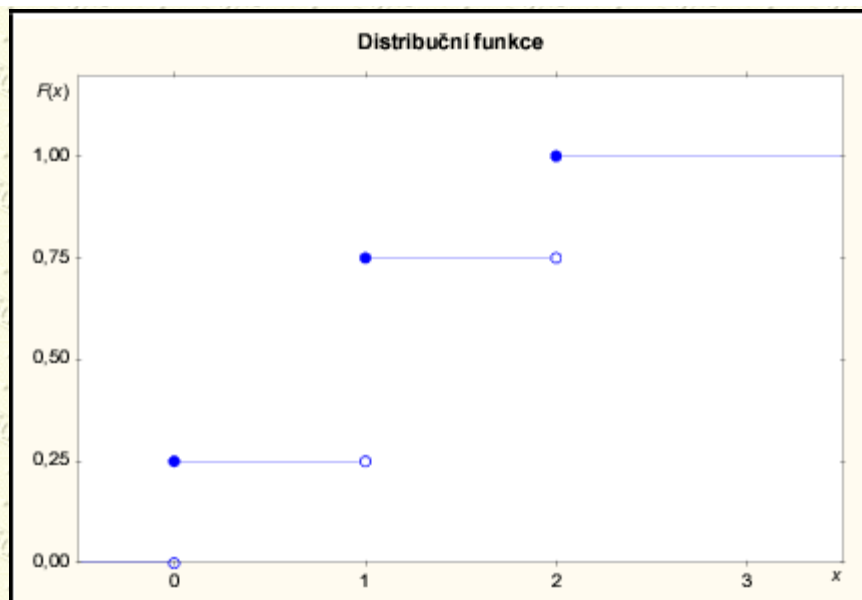


Velmi často popisujeme rozdělení pravděpodobnosti náhodné veličiny X pomocí distribuční funkce $F(x)$. **Distribuční funkce $F(x)$** vyjadřuje pro každé reálné x pravděpodobnost, že náhodná veličina X nabude hodnoty menší nebo rovné x , tedy

$$F(x) = P(X \leq x), \quad \text{pro } -\infty < x < +\infty. \quad (4.5)$$

Na obrázku 4.2 je graficky znázorněna distribuční funkce diskrétní náhodné veličiny X "počet alel Se". \diamond

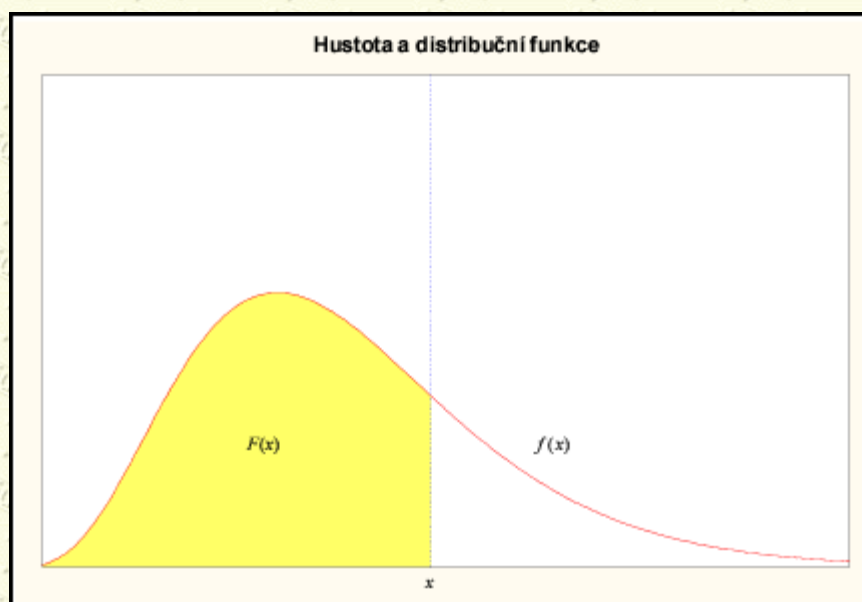
Obrázek 4.2: Distribuční funkce $F(x)$ náhodné veličiny X "počet alel Se" v genotypu potomka heterozygotních vylučovatelů



Jestliže je náhodná veličina spojitá, nemá smysl uvažovat pravděpodobnosti jednotlivých hodnot x , protože tyto pravděpodobnosti jsou obecně rovny nule. Například pravděpodobnost, že výška dospělého muže je *přesně* 180 cm, je rovna nule, protože v obecně nekonečné populaci muži s výškou 180 cm tvoří zcela zanedbatelnou část. Avšak uvažujeme-li malý interval kolem středu 180 cm, bude již nenulová pravděpodobnost, že naměříme výšku z tohoto intervalu. V praxi měření výšky provádíme většinou s přesností na centimetry, a proto naměřená výška 180 cm znamená, že skutečná výška muže leží přibližně v rozmezí od 179,5 cm do 180,5 cm.

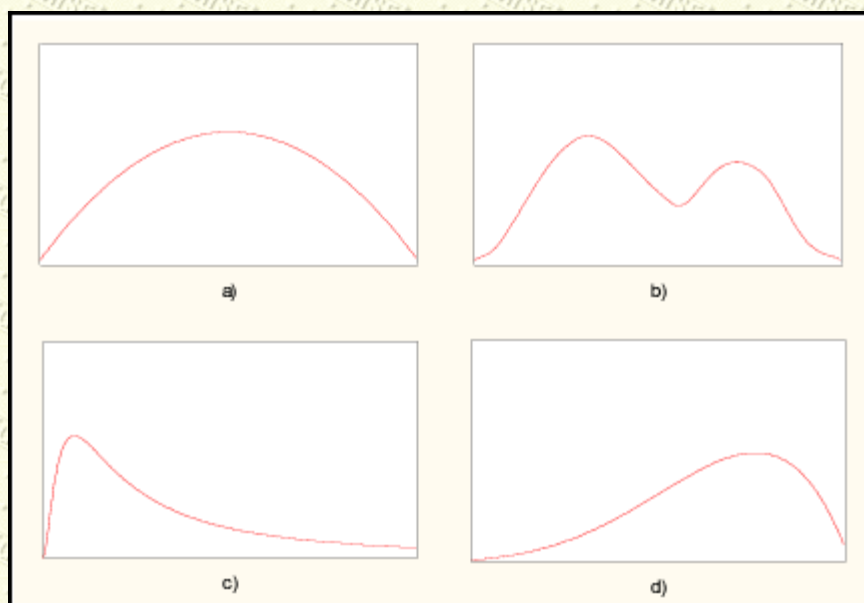
Spojitá náhodná veličina X nabývá všech hodnot x z určitého intervalu. Její pravděpodobnostní rozdělení je popsáno pomocí reálné nezáporné funkce $f(x)$, která se nazývá **hustota (frekvenční funkce)** rozdělení. Pomocí hustoty počítáme hodnoty průměru a rozptylu spojitě náhodné veličiny. Distribuční funkci spojitě náhodné veličiny $F(x)$ lze graficky vyjádřit plochou pod hustotou $f(x)$, viz obrázek 4.3.

Obrázek 4.3: Velikost vybarvené plochy odpovídá hodnotě distribuční funkce F v bodě x



Podle tvaru rozeznáváme symetrické a asymetrické pravděpodobnostní rozdělení. Symetrické rozdělení často mívá "tělesná výška" ve sledované populaci osob. Asymetrické rozdělení je buď levostranně nebo pravostranně asymetrické. Pravostranně asymetrické rozdělení (má delší pravý chvost) často pozorujeme pro "tělesnou hmotnost" osob sledované populace. Pravděpodobnostní rozdělení jsou většinou jednovrcholová, ale mohou být i dvouvrcholová nebo obecně vícevrcholová. Zpravidla je vícevrcholovost způsobena nehomogenitou dat, například dvouvrcholovost pozorujeme u pravděpodobnostního rozdělení "tělesné výšky" pro populaci složenou z mužů a žen. Na obrázku 4.4 jsou uvedeny příklady některých spojitých rozdělení pravděpodobnosti.

Obrázek 4.4: Příklady spojitých rozdělení a) symetrické jednovrcholové rozdělení, b) dvouvrcholové rozdělení, c) pravostranně asymetrické rozdělení, d) levostranně asymetrické rozdělení



4.2 Normální rozdělení

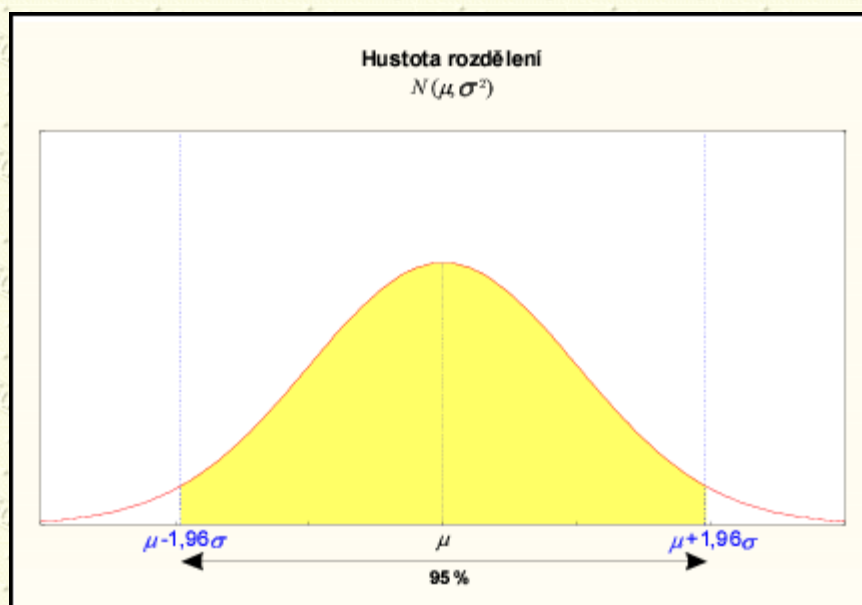
V praxi se často setkáváme s normálním rozdělením u řady veličin popisujících výsledky biochemického vyšetření, např. u tělesné výšky, diastolického tlaku, vitální kapacity plic nebo sledování velikosti chyby měření. Normální rozdělení je také známo jako Gaussovo rozdělení (podle svého objevitele Gausse). Slovo "normální" zde ovšem není použito ve svém obvyklém smyslu "obyčejné", "běžné", či v lékařském významu "bez nemoci". Jeho použití se vztahuje k staršímu významu "řídící se zákonem nebo modelem".

Hustota normálního rozdělení veličiny X má tvar

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (4.6)$$

kde $\pi = 3,141 \dots$ a $e = 2,718 \dots$ jsou matematické konstanty a μ a $\sigma > 0$ jsou konstanty určující polohu křivky (4.6) na ose x (μ) a její "roztážení" podél osy x (σ), tj. průměrnou hodnotu a míru variability. Takovým konstantám se říká *parametry*. Známe-li parametry μ a σ , je normální rozdělení plně určeno. To, že veličina X má normální rozdělení s průměrem μ a rozptylem σ^2 , se proto symbolicky zapisuje jako $X \sim N(\mu, \sigma^2)$. Pro veličinu X s normálním rozdělením lze histogram výsledků velkého počtu n nezávislých pozorování vyrovnat křivkou (4.6) - viz obrázek 5.1.

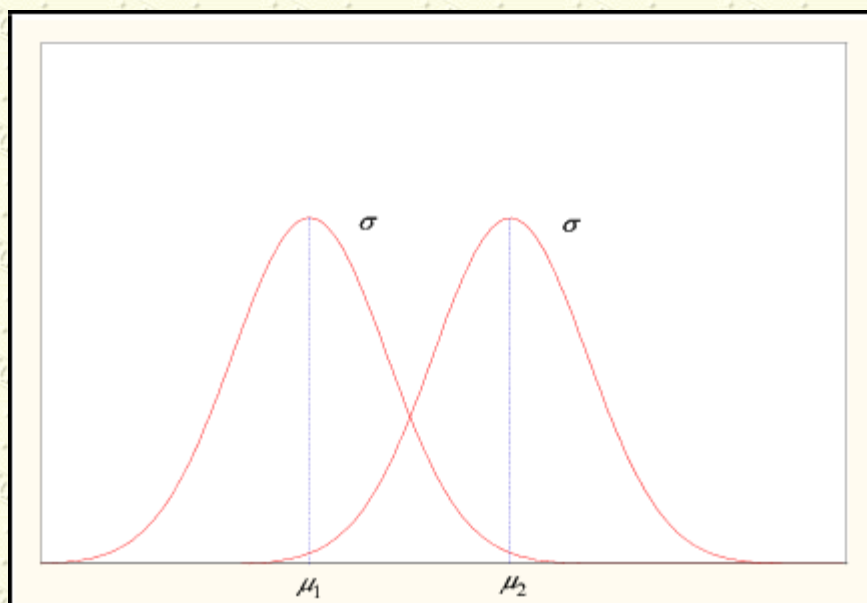
Obrázek 4.5: Hustota normálního rozdělení



Grafické znázornění normálního rozdělení je dáno touto symetrickou jednovrcholovou hustotou, která je zvonovitého tvaru a nikde neprotíná vodorovnou osu (viz obrázek 4.5). Parametr ležící pod vrcholem hustoty je průměr μ , parametr σ je směrodatná odchylka a jeho druhá mocnina rozptyl σ^2 (variance) veličiny X . Plocha pod křivkou hustoty normálního rozdělení je rovna jedné. Pravděpodobnost, že náhodná veličina nabude hodnot z určitého intervalu, je rovna ploše pod hustotou nad tímto intervalem. Například pro interval s hranicemi $\mu - 1,96\sigma$ a $\mu + 1,96\sigma$ má tato plocha velikost 0,95. Náhodná veličina X nabývá tedy hodnot z tohoto intervalu s 95% pravděpodobností a pouze s 5% pravděpodobností leží její hodnoty mimo uvedený interval (viz obrázek 4.5).

Parametr μ , průměr náhodné veličiny, určuje polohu rozdělení na číselné ose. Na obrázku 4.6 jsou zakresleny dvě hustoty normálního rozdělení s různými průměry a stejnými směrodatnými odchylkami. Vzhledem k symetrii normálního rozdělení je parametr současně modem a mediánem. *Modus* je hodnota náhodné veličiny, která se vyskytuje nejčastěji. U spojitého rozdělení ji hledáme pod vrcholem hustoty. *Medián* je prostřední hodnota v tom smyslu, že rozděluje plochu pod hustotou na dvě stejně velké části. Průměr, modus a medián se nazývají *míry polohy*. Modus a medián se uplatňují zejména při popisu asymetrických a vícevrcholových rozdělení.

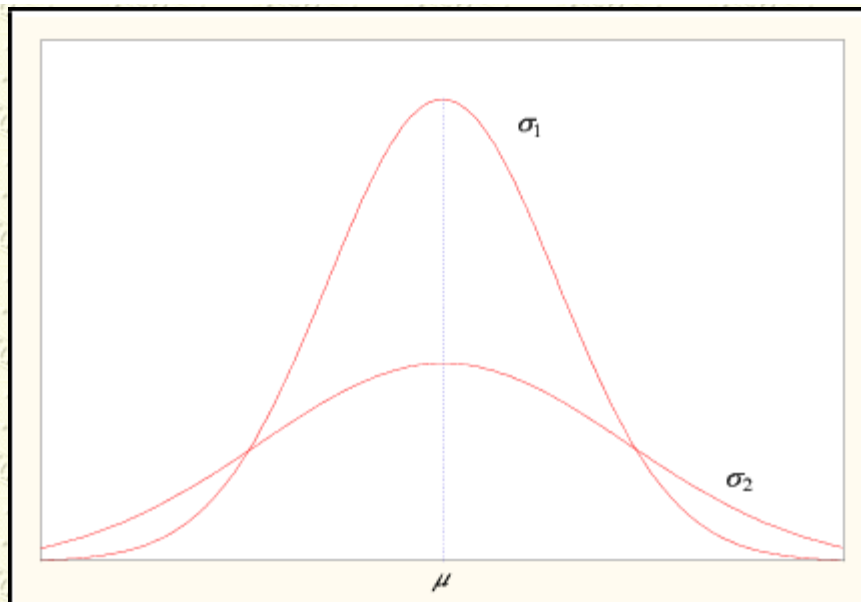
Obrázek: Hustoty normálních rozdělení s různými průměry $\mu_1 < \mu_2$ a stejnými směrodatnými odchylkami



Poznámka: Je-li veličina normálně rozdělena, pak průměr, medián a modus jsou shodné, tj. mají stejnou hodnotu. Je-li veličina zešikmena napravo (tj. má pravostranně asymetrické rozdělení neboli delší pravý chvost), pak modus $<$ medián $<$ průměr. Analogicky pro veličinu zkosenou doleva. (Viz obrázek [4.4](#).)

Parametr σ , směrodatná odchylka náhodné veličiny, určuje tvar hustoty normálního rozdělení. Čím je směrodatná odchylka větší, tím je hustota plošší, čím je menší, tím je křivka strmější. Na obrázku [4.7](#) jsou znázorněny dvě hustoty normálního rozdělení se stejnými průměry a různými směrodatnými odchylkami ($\sigma_1 < \sigma_2$). Vidíme, že pro větší směrodatnou odchylku se mohou snáze vyskytovat hodnoty náhodné veličiny vzdálenější od průměru. Proto směrodatnou odchylku σ (stejně jako rozptyl σ^2) řadíme mezi *míry variability*.

Obrázek: Hustoty normálních rozdělení se stejnými průměry a různými směrodatnými odchylkami $\sigma_1 < \sigma_2$



V normálním rozdělení:

1.

téměř 70 % hodnot leží ve vzdálenosti menší než 1 směrodatná odchylka od průměru, přesněji

68,27 % leží mezi $\mu \pm 1\sigma$;

2.

95 % hodnot leží ve vzdálenosti menší než 2 směrodatné odchylky od průměru, přesněji

95 % leží mezi $\mu \pm 1,96\sigma$;

3.

99 % hodnot leží ve vzdálenosti menší než 3 směrodatné odchylky od průměru, přesněji

99 % leží mezi $\mu \pm 2,576\sigma$.

Pro srovnání variability některých pravděpodobnostních rozdělení je vhodným ukazatelem **variační koeficient**, který je dán podílem směrodatné odchylky a průměru. Je bezrozměrný a obvykle se vyjadřuje v procentech, tj.

$(\sigma/\mu) \cdot 100\%$. Variační koeficient používáme pro srovnání několika náhodných veličin s velmi odlišnými průměry nebo pro srovnání variability veličin měřených v různých jednotkách.

Důležitými ukazateli variability jsou tzv. **kvantily**. $100P\%$ kvantil x_p je číslo, které odděluje $100P\%$ nejmenších hodnot náhodné veličiny X , tj. distribuční funkce v bodě x_p je $F(x_p) = P$. Tedy 50% kvantil $x_{0,50}$ je totéž co medián.

Dobry popis rozdělení pravděpodobnosti dostaneme stanovením dostatečného počtu kvantilů. Kvantily zaznamenané po dvaceti pěti procentech nazýváme **kvartily**, po deseti procentech **decily** a po jednom procentu

percentily. Tedy 25% kvantil je 1. kvartil (dolní kvartil), 10% kvantil je 1. decil a podobně 1% kvantil je 1. percentil. Medián je totéž co 50% kvantil, 2. kvartil, 5. decil nebo 50. percentil. S použitím kvartilů, decilů a percentilů se často setkáváme při prezentaci výsledků antropometrických studií.

Pro veličinu $X \sim N(\mu, \sigma^2)$ lze vypočítat hodnotu distribuční funkce $F(x) = P(X \leq x)$ pro libovolné x pomocí hustoty normálního rozdělení. Ve statistických tabulkách nalezneme vypočtené hodnoty distribuční funkce pouze pro normální rozdělení s průměrem 0 a rozptylem 1 (viz tabulku [A.2](#)). Veličinu s tímto normálním rozdělením označíme Z , tj. $Z \sim N(0, 1)$ a rozdělení nazýváme **standardizované (normované) normální rozdělení**. Distribuční funkci standardizovaného normálního rozdělení označíme $\Phi(z) = P(Z \leq z)$. Libovolnou veličinu $X \sim N(\mu, \sigma^2)$ můžeme transformovat na veličinu $Z = (X - \mu)/\sigma$, která má standardizované normální rozdělení, tj. $Z \sim N(0, 1)$.

Podle pravidel pro úpravu nerovností lze snadno odvodit, že

$$\begin{aligned} F(x) &= P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \\ &= P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right). \end{aligned}$$

Příklad 4.2 Výška v populaci chlapců ve věku 3,5-4 roky má normální rozdělení s průměrem $\mu = 102$ cm a směrodatnou odchylkou $\sigma = 4,5$ cm. Spočítejte, jaké procento chlapců v uvedeném věku má výšku menší nebo rovnou 93 cm.

Řešení: Pravděpodobnost, že výška nabude hodnoty menší nebo rovné 93 cm, je vyjádřena hodnotou distribuční funkce

$$\begin{aligned} F(93) &= P(X \leq 93) = P\left(Z \leq \frac{93 - 102}{4,5}\right) = \\ &= P(Z \leq -2) = \Phi(-2). \end{aligned}$$

V tabulce [A.2](#) najdeme $\Phi(-2) = 0,0228$. Tudiž pouze 2,28 % chlapců ve věku 3,5-4 roky má výšku menší nebo rovnou 93 cm. \diamond

Jak již bylo uvedeno, řada veličin v biologii a medicíně má Gaussovo normální rozdělení. Setkáváme se však také s rozděleními jiného typu, ale někdy lze takové veličiny vhodnou transformací převést opět k normálnímu rozdělení. Příkladem je logaritmická transformace. Řekneme, že veličina X má **logaritmicko-normální rozdělení** s parametry μ a σ , tj. $X \sim LN(\mu, \sigma^2)$, jestliže veličina $Y = \ln X$ má normální rozdělení s parametry μ a σ^2 , tj. $Y \sim N(\mu, \sigma^2)$. Ze zkušenosti víme, že logaritmicko-normální rozdělení mívá např. tělesná hmotnost, doba přežití po jedné dávce ozaření nebo minimální smrtelná dávka přípravku v homogenní skupině pokusných zvířat.

Již z dat můžeme usuzovat, zda je model normálního rozdělení vhodný. Velmi často jsou data zešikmena - buď kladně (dlouhý chvost napravo) nebo záporně (dlouhý chvost nalevo). V medicíně se s tím lze setkat při měření

hematologických, hormonálních nebo biologických veličin. Co tedy s takovými daty dělat?

Obvyklý přístup

1. Ztransformujeme původní veličinu (např. pomocí logaritmu, druhé odmocniny či převrácené hodnoty) na novou veličinu, pro kterou je model normálního rozdělení přijatelný.
2. Analýzu potom provedeme na transformované veličině.
3. Výsledky analýzy (např. průměry či intervaly spolehlivosti) lze pro účely prezentace výsledků zpětně transformovat.

Pokud nenajdeme vhodnou transformaci na normální rozdělení, nabízí statistika jiné přístupy založené např. na neparametrických metodách.

4.3 Binomické rozdělení

Normální rozdělení je vhodné pro *spojité* veličiny, ale často musíme pracovat s veličinami, které jsou ve své podstatě *nespojité*. Existuje několik typů diskrétních rozdělení, nejčastější z nich se nazývá *binomické*.

Uvažujeme n statisticky nezávislých pokusů. V každém pokusu může sledovaný jev buď nastat (= "úspěch") nebo nenastat (= "neúspěch"). Odpovídající pravděpodobnosti označíme π a $(1 - \pi)$ a jsou v každém pokusu stejné.

V tomto případě tedy π nemá význam Ludolfova čísla (3,141...)! Celkový počet úspěchů X v n nezávislých pokusech je *binomická veličina*. Tato náhodná veličina může nabývat pouze celočíselných hodnot od 0 do n .

Je-li v každém pokusu pravděpodobnost úspěchu π , potom pravděpodobnost, že v n nezávislých pokusech nastane přesně k úspěchů, je

$$P(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}, \quad \text{pro } k = 0, 1, 2, \dots, n. \quad (4.7)$$

Binomický koeficient $\binom{n}{k}$ je počet k -členných kombinací z n objektů a je definován následovně:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad (4.8)$$

kde *faktoriál* $n!$ znamená

(4.9)

$$n! = n(n-1)(n-2) \cdots 1.$$

Například binomický koeficient $\binom{3}{1}$ je

$$\binom{3}{1} = \frac{3!}{1!2!} = \frac{3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 1} = 3.$$

Poznámka: $1! = 1$; $0! = 1$.

Binomické rozdělení náhodné veličiny X symbolicky značíme $X \sim B_i(n, \pi)$, kde n je přirozené číslo a $0 < \pi < 1$.

Pravděpodobnosti lze počítat rovněž podle rekurzivního vzorce

$$P(X = k + 1) = \frac{n - k}{k + 1} \cdot \frac{\pi}{1 - \pi} P(X = k). \quad (4.10)$$

Jestliže spočteme pravděpodobnosti $P(X = k)$ pro všechny hodnoty k , snadno ověříme, že jejich součet je roven jedné. Podle klasické binomické věty dostaneme

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} \pi^k (1 - \pi)^{n-k} = (\pi + (1 - \pi))^n = 1.$$

Průměrný počet úspěchů ve výběru o rozsahu n se získá vynásobením počtu nezávislých pokusů (někdy nazývaných Bernoulliho pokusy) pravděpodobností úspěchu v jednom pokusu. Průměr binomické veličiny je tedy $n\pi$.

Rozptyl binomické veličiny je $n\pi(1 - \pi)$, kde $(1 - \pi)$ je pravděpodobnost neúspěchu. Směrodatná odchylka je druhá odmocnina z rozptylu, tedy $\sqrt{n\pi(1 - \pi)}$.

Speciálním případem binomického rozdělení $B_i(n, \pi)$ je rozdělení $B_i(1, \pi)$, které se nazývá *alternativní*.

Příklad 4.3 Předpokládejme, že pravděpodobnost narození dívky je 0,49. Jaká je pravděpodobnost toho, že mezi třemi dětmi v rodině je právě jedna dívka?

Tabulka 4.2: Parametry binomického rozdělení v příkladu 4.3

--	--	--	--	--	--

Pokus	Úspěch	Neúspěch	Pravděpodobnost úspěchu	Počet pokusů	Počet úspěchů
			π	n	k
narození dítěte	dívka	chlapec	0,49	počet dětí	počet dívek

Řešení: Jak je vidět z tabulky 4.2, počet narozených dívek v rodině je náhodná veličina s binomickým rozdělením. Pravděpodobnost, že mezi třemi dětmi je právě jedna dívka, tedy vypočteme jako

$$P(X = 1) = \binom{3}{1} 0,49^1 \cdot 0,51^2 = 3 \cdot 0,127 = 0,38. \diamond$$

Při použití tohoto postupu jsme předpokládali, že pokusy jsou nezávislé, tj. znalost výsledku prvního pokusu neovlivní pravděpodobnost v druhém pokusu atd. V našem příkladu se tedy nejedná o rodinu s dvojčaty či trojčaty.

Příklad 4.4 Jaká je pravděpodobnost, že v rodině s 8 dětmi jsou právě 3 dívky? Opět vylučujeme, že některé děti jsou z vícečetných těhotenství. (Znovu předpokládejme, že pravděpodobnost narození dívky je 0,49.)

Řešení:

$$P(k = 3) = \binom{8}{3} 0,49^3 \cdot 0,51^5 = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} \cdot 0,118 \cdot 0,035 = 0,23. \diamond$$

V praxi se lze vyhnout nepříjemnému počítání binomických koeficientů

1. vyhledáním koeficientu v tabulkách,
2. použitím kalkulačky,
3. použitím počítače.

4.4 Poissonovo rozdělení

Toto diskrétní rozdělení vznikne buď jako limitní případ binomického rozdělení nebo tehdy, když události nějakého druhu nastávají náhodně v čase či prostoru. Je-li pravděpodobnost (π) nějaké výjimečné události (např. určité mutace genu) relativně malá a rozsah výběru poměrně velký, pak Poissonovo rozdělení v podstatě splývá s binomickým, ale je mnohem výhodnější pro počítání. Například když n je příliš velké - řekněme v tisících - pak je velmi obtížné vypočítat binomický koeficient $\binom{n}{x} = \frac{n!}{x!(n-x)!}$.

Uvažujme náhodnou veličinu X , která představuje počet výskytů nějaké výjimečné události (např. mutace) v daném intervalu (natažení DNA). Veličina X tedy může nabývat celočíselných hodnot od 0 do nekonečna. Nechť λ je konstanta označující průměrný počet událostí v intervalu (času nebo prostoru). Potom

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad (4.11)$$

a X je náhodná veličina s Poissonovým rozdělením s parametrem λ , kde symbol e je základ přirozeného logaritmu, $e \doteq 2,71828$. Symbolicky Poissonovo rozdělení značíme $X \sim Po(\lambda)$, $\lambda > 0$.

Poissonovo rozdělení je charakterizováno následujícími vlastnostmi.

1. Pravděpodobnost výskytu jedné události v daném intervalu (času nebo prostoru) je úměrná délce tohoto intervalu.
2. Události se vyskytují nezávisle jak ve stejném intervalu, tak mezi po sobě jdoucími intervaly.

Připomeňme, že průměr binomické náhodné veličiny je roven $n\pi$ a rozptyl je $n\pi(1 - \pi)$. Když je π velmi malé, pak $1 - \pi$ je blízko 1 a $n\pi(1 - \pi)$ je přibližně rovno $n\pi$. V takové situaci je tedy průměr i rozptyl rozdělení totožný a může být nahrazen jediným parametrem λ . Skutečnost, že průměr je roven rozptylu (tj. parametru λ), je charakteristickou vlastností Poissonova rozdělení.

Příklad 4.5 Předpokládejme, že v určité populaci krys se vyskytuje albín s pravděpodobností $\pi = 0,001$, ostatní krysy jsou normálně pigmentované. Ve vzorku 100 krys náhodně vybraných z této populace určete pravděpodobnost, že vzorek a) neobsahuje albína, b) obsahuje právě jednoho albína.

Řešení: Pravděpodobnost výskytu albína je $\pi = 0,001$. Předpokládaný počet albínů ve výběru o rozsahu n je $\lambda = n\pi$ (průměr binomické náhodné veličiny), tj. v našem příkladu $\lambda = n\pi = 100 \cdot 0,001 = 0,1$. Počet řídkých událostí (tj. albínů) označme x . Potom podle vzorce (4.11)

$$\text{a) pro } x = 0 \text{ máme } \frac{e^{-0,1} \cdot 0,1^0}{0!} = \frac{e^{-0,1} \cdot 1}{1} = 0,9048,$$

$$\text{b) pro } x = 1 \text{ máme } \frac{e^{-0,1} \cdot 0,1^1}{1!} = \frac{e^{-0,1} \cdot 0,1}{1} = 0,09048.$$

Jak je vidět, pravděpodobnost, že ve vzorku 100 krys nebude žádný albín, je desetkrát vyšší než pravděpodobnost, že ve vzorku bude právě jeden albín. Pravděpodobnosti výskytu dvou a více albínů jsou již velmi malé. \diamond

[Další](#) [Výš](#) [Předchozí](#) [Obsah](#) [Rejstřík](#)

Další: [5. Výběr a popisné statistiky](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [3. Základy teorie pravděpodobnosti](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [6. Odhady populačních průměrů a pravděpodobností](#) **Výš:** [Základy statistiky pro biomedicínské obory](#)

Předchozí: [4. Náhodná veličina a rozdělení pravděpodobnosti](#)

Kapitola

- [5.1 Výběrové rozdělení](#)
- [5.2 Míry polohy](#)
- [5.3 Míry variability](#)

5. Výběr a popisné statistiky

Potíže, které má mnoho inteligentních lidí se sčítáním, jsou nekonečné.

M. Greenwood

5.1 Výběrové rozdělení

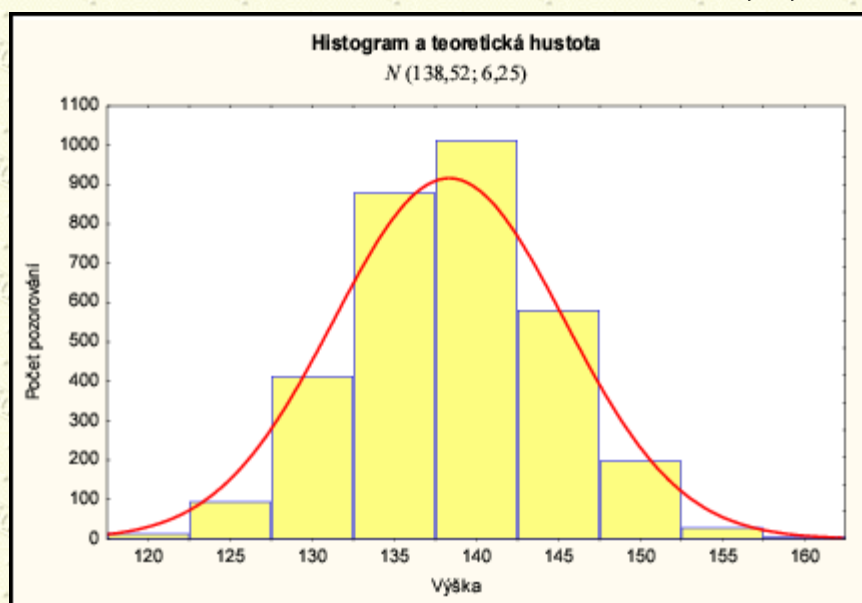
V předchozí kapitole jsme poznali základní teoretická rozdělení náhodných veličin. Na základě výsledků výběrového šetření můžeme teoretické rozdělení pouze odhadovat prostřednictvím výběrového rozdělení. Výběrové rozdělení se graficky znázorňuje pomocí histogramu nebo polygonu četností. Například výběrové rozdělení tělesné výšky v souboru 3231 chlapců při ekvidistantním třídění s délkou třídního intervalu 5 cm (viz tab. 5.1) je graficky znázorněno histogramem na obrázku 5.1. Kdybychom teoreticky zvětšovali velikost výběru nade všechny meze a současně zkracovali délku třídních intervalů, obrys histogramu by se postupně přibližoval spojitě křivce. Tříděním tělesné výšky nekonečně velkého souboru chlapců do nekonečně malých intervalů bychom dostali hladkou křivku, která znázorňuje teoretické rozdělení tělesné výšky chlapců v uvedeném věku, v tomto případě hustotu normálního rozdělení (viz obrázek 5.1).

Tabulka 5.1: Rozdělení chlapců ve věku 9,5-10 let podle tělesné výšky (délka třídního intervalu 5 cm)

Střed třídy x_i	Absolutní četnost n_i	Kumulativní absolutní četnost	Relativní četnost n_i/n	Kumulativní relativní četnost
120	13	13	0,0040	0,0040
125	95	108	0,0294	0,0334
130	414	522	0,1281	0,1615
135	880	1402	0,2724	0,4339
140	1013	2415	0,3135	0,7474

145	582	2997	0,1801	0,9275
150	199	3196	0,0616	0,9891
155	29	3225	0,0090	0,9981
160	6	3231	0,0019	1,0000
Celkem	3231	-	1,0000	-

Obrázek 5.1: Histogram výběrového rozdělení tělesné výšky 3231 chlapců ve věku 9,5-10 let (délka třídního intervalu 5 cm) a teoretická hustota normálního rozdělení tělesné výšky



5.2 Míry polohy

V souboru dat, např. výsledků nějakého pozorování, potřebujeme často určit hodnotu, kolem které se data soustřeďují, stanovit jakýsi jejich "střed". Pro tento účel byla zavedena celá řada popisných měř, které se nazývají *míry polohy*.

5.2.1 Průměr

Průměr (aritmetický průměr) \bar{x} používáme, když čísla můžeme opravdu počítat, tj. znaky jsou kvantitativní, měřené na číselné stupnici. Neměl by být používán pro ordinální znaky vzhledem k libovůli při volbě ordinální stupnice. Je rovněž velmi citlivý na odlehlé hodnoty. Průměr z hodnot ve výběru vypočítáme, jestliže součet všech hodnot dělíme rozsahem výběru (n). Máme-li tedy n pozorování: $x_1, x_2, x_3, \dots, x_n$, pak průměr počítáme následujícím způsobem. Součet pozorování se značí symbolem

$$\sum_{i=1}^n x_i = x_1 + x_2 + x_3 + \dots + x_n.$$

Počet pozorování je n . Průměrem je

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (5.1)$$

Příklad 5.1 Vypočítejte průměr následujících výsledků vyšetření: 39, 42, 73, 67, 24, 55.

Řešení: Součet pozorování je 300. Počet pozorování je 6. Průměrem je podíl $300/6 = 50$. \diamond

Pokud původní data nejsou již dostupná, odhadujeme někdy průměr z roztříděných dat. Jsou-li data roztříděna podle tabulky 5.1, hodnoty v každé třídě reprezentujeme pouze hodnotou středu třídy. Potom průměr odhadujeme jako **vážený průměr** z hodnot středů třídy (x_i) s přihlédnutím k počtu pozorování ve třídě (absolutní četnost n_i).

Nejprve tedy vynásobíme na každém řádku hodnoty x_i a n_i , tyto součiny sečteme a získaný součet dělíme rozsahem výběru (n), tj.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i, \quad (5.2)$$

kde k je počet tříd a $n = \sum_{i=1}^k n_i$.

Z tabulky 5.1 pro $n = 3231$, $k = 9$ a $\sum_{i=1}^k n_i x_i = 447570$, dostáváme $\bar{x} = 447570/3231 = 138,52$. Tabulka 5.1 uvádí

kromě absolutních četností (n_i) rovněž relativní četnosti (n_i/n) pro jednotlivé třídy. Kumulativní relativní četnost

např. v j -té třídě je pak dána jako $\sum_{i=1}^j \frac{n_i}{n}$.

Průměr \bar{x} vypočítaný z hodnot roztříděných do intervalů je zatížen chybou. Její nejvyšší hodnota je rovna polovině délky intervalu a vznikla by tehdy, kdyby všechny hodnoty zahrnuté do intervalů ležely buď na jejich horní nebo dolní hranici. V praxi bývá tato extrémní situace velmi vzácná, a proto vzniklá chyba je podstatně menší. Volbě délky intervalu při ekvidistantním třídění je však třeba věnovat náležitou pozornost i z tohoto hlediska.

5.2.2 Modus

Modus \hat{x} je hodnota, která se v souboru dat vyskytuje nejčastěji. Důležitý je pro kvalitativní, zejména nominální znaky.

Příklad 5.2 Co je modus v následujících výsledcích zjišťování krevních skupin: A, 0, 0, B, B, AB, A, A, 0, 0, 0, AB, B, 0, B, A, 0, AB, 0, 0, B, 0, A?

Řešení: V tabulce 5.2 jsou tyto výsledky přehledně zapsány.

Tabulka 5.2: Četnosti výskytu krevních skupin v příkladu 5.2

krevní skupina	četnost výskytu
A	5
B	5
AB	3
0	10

Modus našich pozorování je krevní skupina 0. \diamond

Pro kvantitativní znaky, kdy jsou data rozříděna do intervalů, určíme modální interval (má nejvyšší četnost) a modus stanovíme interpolací vzhledem k četnostem v sousedních intervalech, tj.

$$\hat{x} = x_D + h \frac{n_2}{n_1 + n_2}, \quad (5.3)$$

kde h je délka modálního intervalu, x_D jeho dolní hranice, n_1 četnost předchozího intervalu a n_2 četnost následujícího intervalu. Modus pro data v tabulce 5.1 je tedy

$$\hat{x} = 137,5 + 5 \cdot \frac{582}{880 + 582} = 139,49.$$

Modus není ovlivňován hodnotami všech prvků ve výběru. Využívá se hlavně pro vystižení nejtypičtější hodnoty znaku, např. délku trvání určité nemoci nebo věk, v němž se vyskytuje sledované onemocnění.

5.2.3 Medián

Máme-li pozorování uspořádána vzestupně nebo sestupně, potom **medián** \tilde{x} je ta hodnota, která rozdělí pozorování na dvě stejně velké skupiny. Přesněji řečeno, máme-li lichý počet uspořádaných pozorování, pak mediánem je prostřední z nich. U sudého počtu se mediánem rozumí obvykle průměr ze dvou prostředních pozorování. Medián využívá pouze informaci o pořadí hodnot, a proto ho má smysl používat pouze pro kvantitativní a ordinální veličiny.

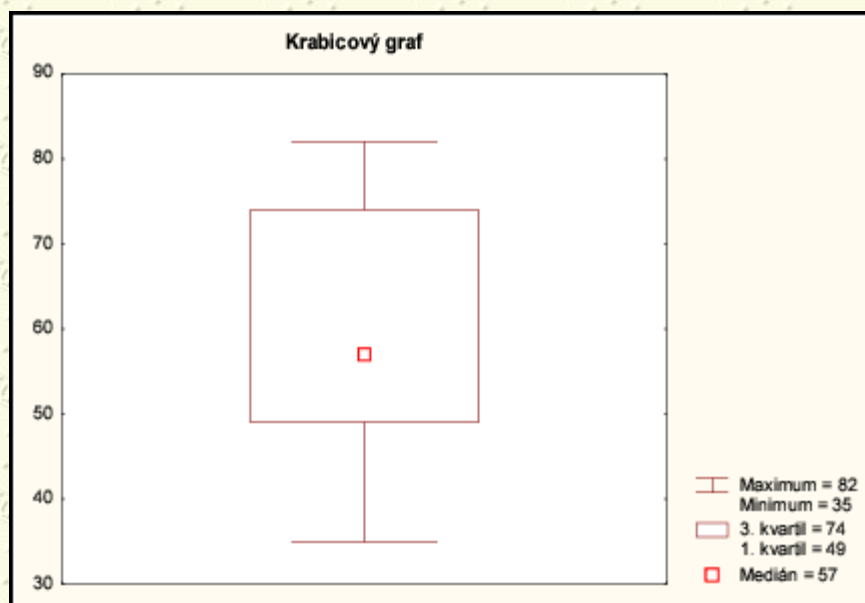
Příklad 5.3 Co je mediánem následujících výsledků hodnocení závažnosti průběhu onemocnění, přičemž A je nejlehčí a F je nejtěžší průběh: C, E, B, D, A, A, B, F, C, C, D?

Řešení: Uspořádejme pozorování např. vzestupně: A, A, B, B, C, C, C, D, D, E, F. Mediánem je stupeň C. \diamond

Příklad 5.4 Co je mediánem následujících výsledků vyšetření: 61, 49, 35, 74, 53, 82?

Řešení: Uspořádejme pozorování vzestupně: 35, 49, 53, 61, 74, 82. Mediánem je průměr z hodnot 53 a 61, tj. $(53 + 61)/2 = 57$. Rozdělení dat ve výběru pomocí mediánu, 1. a 3. kvartilu (viz odst. 5.3.3), minimální a maximální hodnoty se často znázorňuje na krabicovém grafu (viz obrázek 5.2). ◊

Obrázek 5.2: Zobrazení mediánu: krabicový graf (Box and Whisker Plot)



U souborů roztržiděných do ekvidistantních intervalů nalezneme nejprve interval, ve kterém leží prostřední prvek, tzv. mediánový interval. Medián (\tilde{x}) stanovíme interpolací jako

$$\tilde{x} = x_D + \frac{h}{n_*} \left(\frac{n}{2} - k \right), \quad (5.4)$$

kde x_D je dolní hranice, h délka a n_* četnost mediánového intervalu, k je kumulativní četnost intervalu před mediánovým intervalem a n je rozsah výběru. Medián pro data v tabulce 5.1 je tedy

$$\tilde{x} = 137,5 + \frac{5}{1013} \left(\frac{3231}{2} - 1402 \right) = 138,55.$$

Vidíme, že vypočtený medián se velmi dobře shoduje s průměrem \bar{x} , což je typické pro symetrická rozdělení četností. Výhodou mediánu (ve srovnání s průměrem) je to, že není ovlivněn velmi malými nebo velmi velkými hodnotami v souboru. Ve farmakologii se medián používá pod názvem 50% účinná dávka (ED50) nebo 50% smrtná dávka (LD50) jako charakteristika účinnosti přípravků.

5.2.4 Geometrický průměr

K dalším mírám polohy patří geometrický průměr, který se však nepoužívá tak často jako aritmetický průměr nebo medián. **Geometrický průměr** \bar{x}_G se vypočítá jako n -tá odmocnina ze součinu pozorování, tedy

$$\bar{x}_G = \sqrt[n]{x_1 x_2 \cdots x_n} = \sqrt[n]{\prod_{i=1}^n x_i}. \quad (5.5)$$

Geometrický průměr je vhodný tehdy, má-li smysl počítat součin pozorovaných hodnot. Obvykle se používá pro veličiny měřené na logaritmické stupnici. Povšimněme si, že když zlogaritmujeme obě strany ve výše uvedeném vzorci dostaneme

$$\log \bar{x}_G = \frac{1}{n} \sum_{i=1}^n \log x_i. \quad (5.6)$$

To znamená, že logaritmus geometrického průměru je roven průměru zlogaritmovaných pozorování.

Příklad 5.5 Spočítejte geometrický průměr z následujících pěti titrů protilátek 4, 8, 16, 16 a 64.

Řešení: Zlogaritmováním zjištěných hodnot dostaneme čísla 0,60, 0,90, 1,20, 1,20 a 1,81. Jejich aritmetický průměr je 1,142. Odlogaritmováním této hodnoty dostaneme hodnotu geometrického průměru jako 13,9. Vidíme, že geometrický průměr se značně liší od aritmetického průměru původních hodnot, který činí 21,6. \diamond

Geometrický průměr nepoužíváme, když jsou původní pozorování záporná nebo nulová. Je velmi vhodný pro použití v situacích, kdy měříme hodnoty v postupně nařazených roztocích nebo když je rozdělení hodnot asymetrické a logaritmická transformace jej opět vrací k symetrii.

5.3 Míry variability

Míry polohy samy o sobě neříkají nic o variabilitě dat. Grafické metody uvedené v kapitole 1, například histogram a polygon četností, jsou sice velmi vhodným nástrojem pro znázornění variability dat, ale často potřebujeme k tomuto účelu spíše sumární číselné charakteristiky. Máme tři hlavní metody jejich výpočtu. Buď pouze zjistíme rozpětí naměřených hodnot, nebo můžeme stanovit určité kvantily pomocí kumulativních relativních četností, či spočítat některé číselné charakteristiky popisující kolísání dat kolem průměru nebo jiné míry polohy.

5.3.1 Rozpětí

Jednou z měř variability je **rozpětí (variační šíře) R** , což je rozdíl mezi nejvyšší (x_{\max}) a nejnižší (x_{\min}) hodnotou v datech, tj.

$$R = x_{\max} - x_{\min}. \quad (5.7)$$

Je to celkem užitečná míra, ale má nevýhodu v tom, že závisí na extrémních hodnotách, takže může poskytnout velmi zavádějící obrázek (např. výběr může mít značně velké rozpětí, i když většina pozorování bude blízko u sebe).

Příklad 5.6 Sedm obyvatel malé obce A může mít stejný průměrný měsíční příjem jako sedm obyvatel malé obce B, ale rozdělení příjmů může být velmi odlišné, jak je vidět z tabulky 5.3.

Tabulka 5.3: Příjmy obyvatel v obcích A a B z příkladu 5.6

obec A	obec B
4 000 Kč	8 000 Kč
6 000 Kč	8 000 Kč
8 000 Kč	9 000 Kč
10 000 Kč	10 000 Kč
12 000 Kč	11 000 Kč
14 000 Kč	12 000 Kč
16 000 Kč	12 000 Kč
$\bar{x}_A = 10\,000$ Kč	$\bar{x}_B = 10\,000$ Kč

V obci A je průměrný příjem 10 000 Kč, ale rozdíl mezi nejvyšší hodnotou (16 000 Kč) a nejnižší hodnotou (4 000 Kč) příjmu je 12 000 Kč. V obci B je průměrný příjem také 10 000 Kč, ale rozdíl je mnohem menší, pouze 4 000 Kč. ◊

5.3.2 Rozptyl, směrodatná odchylka a variační koeficient

Pokud jsou pozorování soustředěna kolem svého průměru, je jejich variabilita malá. Pokud jsou naopak roztroušena ve značné vzdálenosti od průměru, pak je jejich variabilita velká. Variabilitu tedy často měříme právě pomocí odchylek pozorování od průměru.

Rozptyl s^2 je průměr čtverců odchylek od průměru. Když však počítáme výběrový rozptyl, nedělíme většinou součet čtverců odchylek výrazem n , ale $(n - 1)$, protože tím docílíme lepšího odhadu celkového rozptylu populace (σ^2).

Dělitel $(n - 1)$ se nazývá **počet stupňů volnosti** rozptylu. Obecný vzorec tedy vypadá takto:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (5.8)$$

Lze jej upravit i na tvar

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right). \quad (5.9)$$

Příklad 5.7 Spočítejte rozptyl s^2 z hodnot hmotností dětí uvedených v tabulce 5.4, jejichž průměr je $\bar{x} = 54,2$ kg.

Tabulka 5.4: Hmotnosti dětí v příkladu 5.7

Číslo	Hmotnost (v kg)	Absolutní odchylka od průměru	Čtverec odchylky
1	68	13,8	190,44
2	65	10,8	116,64
3	59	4,8	23,04
4	59	4,8	23,04
5	57	2,8	7,84
6	52	2,2	4,84
7	49	5,2	27,04
8	48	6,2	38,44
9	48	6,2	38,44
10	48	6,2	38,44
11	43	11,2	125,44
Součet	596	74,2	633,64

Rozptyl se počítá jako součet čtverců (633,64) dělený počtem stupňů volnosti ($n - 1$, tj. $11 - 1 = 10$), což se v tomto případě rovná 63,36.

V našem příkladu je tedy

$$s^2 = \frac{633,64}{10} \doteq 63,36 \text{ kg. } \diamond$$

Druhá odmocnina z rozptylu se nazývá **směrodatná odchylka** s . V příkladu 5.7 vyjde

$$s = \sqrt{63,36} \doteq 7,96.$$

Směrodatná odchylka je používána častěji než rozptyl. Poznamenejme, že směrodatná odchylka s je ve stejných jednotkách jako původní hodnoty.

Jsou-li naměřené hodnoty roztrženy podobně jako v tab. 5.1 nebo se stejné hodnoty v datech vícekrát opakují, má vzorec pro výpočet rozptylu tvar

$$s^2 = \frac{1}{n-1} \sum_{i=1}^k n_i (x_i - \bar{x})^2 \quad (5.10)$$

nebo po úpravě

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^k n_i x_i^2 - n \bar{x}^2 \right), \quad (5.11)$$

kde k je počet různých hodnot x_i (počet tříd) a n_i jsou četnosti hodnot x_i . U dat tříděných do intervalů hodnoty x_i ve vzorci vyjadřují středy třídních intervalů. Například z dat tabulky 5.1 dostáváme

$$s^2 = \frac{1}{3230} (13(120 - 138,52)^2 + 95(125 - 138,52)^2 + \dots + 6(160 - 138,52)^2) = 39,1046$$

a $s = 6,2534$.

Rozptyl s^2 vypočtený z hodnot roztržěných do ekvidistantních intervalů délky h je zatížen chybou, kterou korigujeme odečtením hodnoty $h^2/12$ (tzv. *Sheppardova korekce*). Z dat tabulky 5.1 dostáváme

$$\frac{h^2}{12} = \frac{25}{12} = 2,0833$$

a korigovaný rozptyl

$$\begin{aligned} s_{\text{kor}}^2 &= s^2 - \frac{h^2}{12} = \\ &= 39,1046 - 2,0833 = 37,0213. \end{aligned} \quad (5.12)$$

Korigovaná směrodatná odchylka je $s_{\text{kor}} = 6,0845$.

Spočítejme ještě rozptyl u souboru, který je rozdělen do k dílčích podsouborů. V i -tém podsouboru s počtem prvků n_i je průměr \bar{x}_i a rozptyl s_i^2 . Potom celkový rozptyl s^2 je

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^k (n_i - 1) s_i^2 + \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2 \right). \quad (5.13)$$

Celkový rozptyl s^2 závisí tedy nejen na počtu pozorování a rozptylech v dílčích souborech, ale i na rozdílech mezi dílčími průměry \bar{x}_i a celkovým průměrem \bar{x} .

Variační koeficient v je užitečnou mírou relativního rozptýlení dat. Počítá se jako podíl směrodatné odchytky k průměru v procentech, tj.

$$v = \frac{s}{\bar{x}} \cdot 100 \%. \quad (5.14)$$

Variační koeficient se často používá při statistické kontrole kvality laboratorních testů.

Příklad 5.8 V nemocnici bylo v určitém období hospitalizováno 150 osob (n_1) na chirurgickém oddělení s průměrnou délkou hospitalizace 19 dní (\bar{x}_1) a směrodatnou odchytkou 3,4 dne (s_1), 100 osob (n_2) na gynekologickém oddělení s průměrnou délkou hospitalizace 7 dní (\bar{x}_2) a směrodatnou odchytkou 2,4 dne (s_2) a na dětském oddělení 90 dětí (n_3) s průměrnou délkou hospitalizace 12 dní (\bar{x}_3) a směrodatnou odchytkou 3,8 dne (s_3).

Spočtěte průměrnou délku hospitalizace (\bar{x}) a směrodatnou odchytku (s) pro všech 340 pacientů.

Řešení: Průměrná délka hospitalizace je

$$\begin{aligned} \bar{x} &= \frac{1}{n} (n_1 \bar{x}_1 + n_2 \bar{x}_2 + n_3 \bar{x}_3) = \\ &= \frac{1}{340} (150 \cdot 19 + 100 \cdot 7 + 90 \cdot 12) = 13,6 \end{aligned}$$

a rozptyl je podle (5.13)

$$\begin{aligned} s^2 &= \frac{1}{339} (149 \cdot 3,4^2 + 99 \cdot 2,4^2 + 89 \cdot 3,8^2 + \\ &+ 150(19 - 13,6)^2 + 100(7 - 13,6)^2 + 90(12 - 13,6)^2) = \\ &= 36,99. \end{aligned}$$

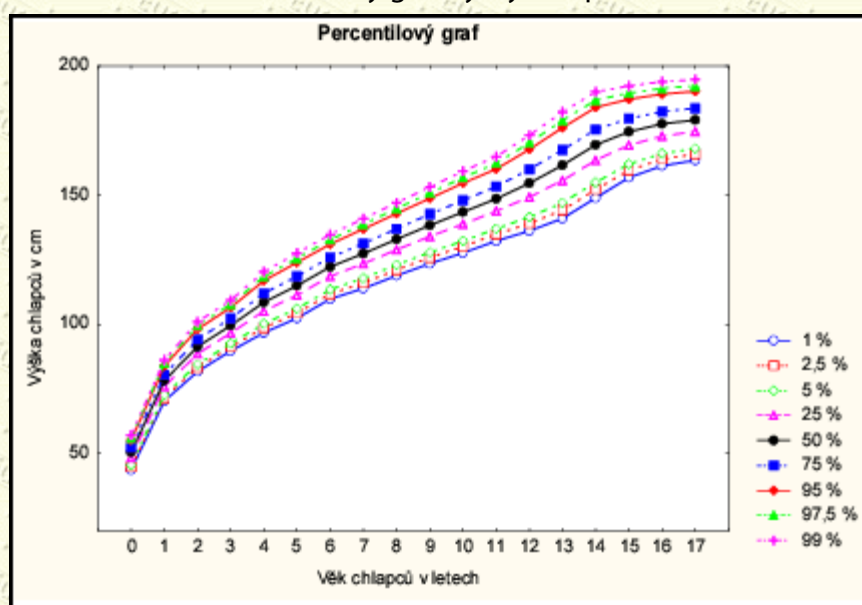
Směrodatná odchytko délky hospitalizace spočtená ze všech 340 pacientů je tedy $s = 6,08$ dne. Pomocí výběrového průměru (\bar{x}) a výběrové směrodatné odchytky (s) snadno spočteme variační koeficient v pro výběrový soubor podle (5.14)

$$v = \frac{6,08}{13,6} \cdot 100 \% = 44,7 \%. \diamond$$

5.3.3 Kvantily, kvartily, decily a percentily

Užitečným krokem při popisu kvantitativních znaků (veličin) je vytvoření tabulky rozdělení četností. Tato tabulka ukazuje, kolikrát byly pozorovány jednotlivé hodnoty či kolik pozorování padlo do určitých intervalů. U diskrétní veličiny mohou být v tabulce uvedeny četnosti výskytu buď pro každou možnou hodnotu zvlášť, nebo pro určité skupiny hodnot. U spojité veličiny je zpravidla vhodné vytvořit třídní intervaly.

Obrázek 5.3: Percentilový graf výšky chlapců ve věku 0-17 let



Z kumulativní relativní četnosti (empirické distribuční funkce) můžeme zjišťovat **kvantily**. 100P% kvantil je taková hodnota, kdy 100P% hodnot ve výběru má hodnotu menší nebo rovnou tomuto kvantilu. V souvislosti s kvantily používáme následující terminologie. Pokud oddělujeme postupně hodnoty po dvaceti pěti procentech, dostáváme **kvartily**. 1. kvartil je roven 25% kvantilu, 2. kvartil 50% kvantilu (neboli mediánu) a 3. kvartil 75% kvantilu. Podobně oddělujeme-li hodnoty postupně po deseti procentech, dostáváme **decily** (například 1. decil je 10% kvantil, 5. decil je medián neboli 2. kvartil), a oddělujeme-li hodnoty po jednom procentu, dostáváme **percentily** (například 1. percentil je 1% kvantil, 50. percentil je medián neboli 5. decil). Kvantily tak poskytují velmi cenný popis variability výběrového souboru. Percentily jsou často užívané pro srovnání individuálních hodnot se stanovenými normami. Velmi často jsou používány pro vytvoření a interpretaci růstových křivek u dětí (viz obrázek 5.3) a při měření schopností a inteligence. Často rovněž určují rozmezí normálních hodnot pro laboratorní nálezy, kdy obvykle je obor hodnot normálu stanoven mezi 2,5% a 97,5% percentilem, takže obor hodnot normálu pokrývá 95% hodnot uvažované populace. Pomocí kvartilů se počítá i **kvartilové rozpětí** jako rozdíl mezi 3. kvantilem (75% kvantilem) a 1. kvantilem (25% kvantilem).

Příklad 5.9 U každé ze 70 žen byl změřen hemoglobin s přesností 0,1 g/100 ml. Naměřené hodnoty jsou v tabulce 5.5.

Tabulka 5.5: Hladina hemoglobinu
v g/100 ml pro 70 žen

10,2	13,7	10,4	14,9	11,5	12,0	11,0
13,3	12,9	12,1	9,4	13,2	10,8	11,7
10,6	10,5	13,7	11,8	14,1	10,3	13,6
12,1	12,9	11,4	12,7	10,6	11,4	11,9
9,3	13,5	14,6	11,2	11,7	10,9	10,4
12,0	12,9	11,1	*8,8	10,2	11,6	12,5
13,4	12,1	10,9	11,3	14,7	10,8	13,3
11,9	11,4	12,5	13,0	11,6	13,1	9,7
11,2	*15,0	10,7	12,9	13,4	12,3	11,0
14,6	11,1	13,5	10,9	13,1	11,8	12,2

Nejprve zjistíme nejnižší a nejvyšší hodnotu, což nám pomůže určit, jaké intervaly máme použít. Pro hrubou představu se bere 5 až 12 intervalů v závislosti na počtu pozorování. Je-li zvolený interval příliš široký, ztratí se informace o detailech, zatímco je-li příliš úzký, tabulka bude nepřehledná. Intervaly volíme pokud možno stejně široké a vždy tak, aby se vzájemně nepřekrývaly, ale pokrývaly celý obor možných hodnot. Také by mělo být jasné, do kterého ze dvou sousedních intervalů patří pozorování, které padne na jejich hranici.

V tabulce 5.5 jsou extrémními hodnotami 8,8 a 15,0 g/100 ml (jsou označeny hvězdičkou). V tabulce rozdělení četností (tab. 5.6) jsme zvolili délku intervalů 1 g/100 ml, což dalo celkem 8 skupin. Skupiny byly označeny 8,0-8,9; 9,0-9,9 atd. Pokud bychom je označili 8-9; 9-10 atd., nebylo by jasné, do které skupiny patří např. pozorování 9,0.

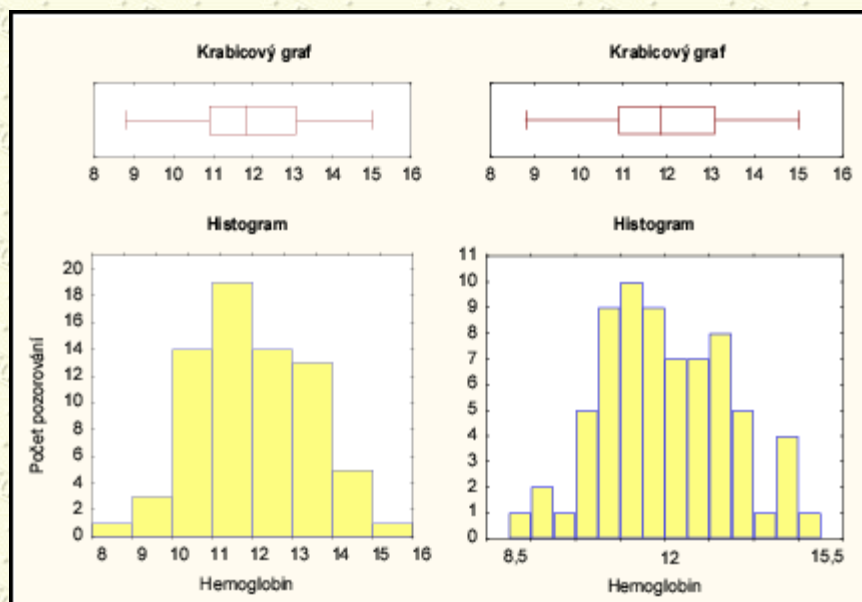
Tabulka 5.6: Rozdělení četností hladiny hemoglobinu v g/100 ml pro 70 žen

Hladina hemoglobinu v g/100 ml	Počet	Relativní četnost v %	Kumulativní rel. četnost v %
8,0-8,9	1	1,4	1,4
9,0-9,9	3	4,3	5,7
10,0-10,9	14	20,0	25,7
11,0-11,9	19	27,1	52,9
12,0-12,9	14	20,0	72,9
13,0-13,9	13	18,6	91,4
14,0-14,9	5	7,1	98,6
15,0-15,9	1	1,4	100,0

Celkem	70	100,0	-
--------	----	-------	---

Jak bylo již zmíněno, nejčastějším zobrazením rozdělení četností je histogram, na který lze pohlížet jako na sloupcový graf pro kvantitativní data. Histogramy na obrázku 5.4 ukazují rozdělení četností hladiny hemoglobinu pro ekvidistantní třídění do osmi skupin (viz tabulku 5.6) a pro ekvidistantní třídění do čtrnácti skupin. Jako doplněk a shrnutí vlastností histogramu je nakreslen také vodorovný krabicový graf, kde je znázorněn medián, 1. a 3. kvartil a maximální a minimální hodnota ve výběru. ◊

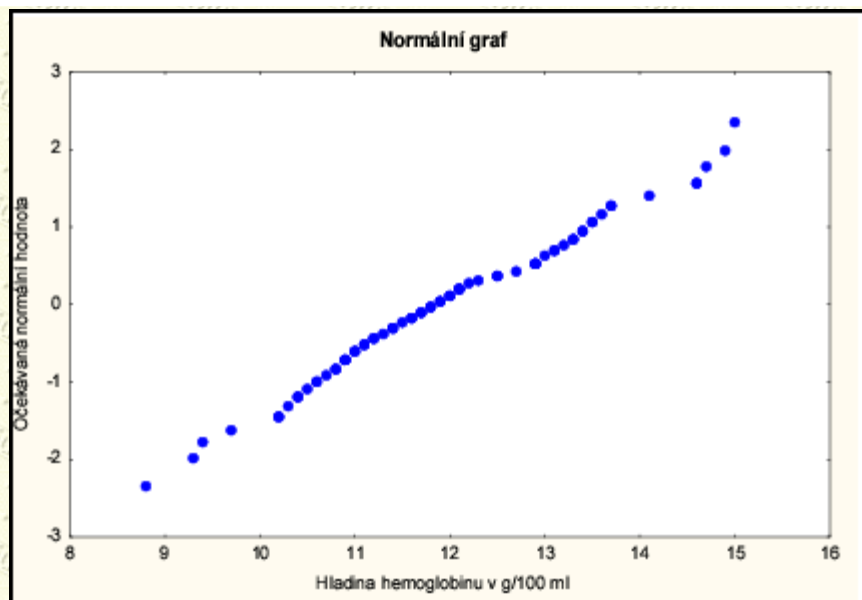
Obrázek 5.4: Histogramy a krabicový graf



Často potřebujeme zjistit, zda se model normálního rozdělení hodí na pozorovaná data. Statistika zná řadu testů normality. Zmíníme se zde o jednoduchém grafickém testu, který používá tzv. *pravděpodobnostní papír*. Tento papír je strukturovaný tak, že kumulativní relativní četnosti (kumulativní procenta) pro data z normálního rozdělení budou přibližně ležet na přímce.

Podobně tzv. *normální graf*, se kterým se často setkáváme při užití komerčních statistických programů, je zkonstruován tak, že na vodorovné ose x jsou zaneseny pozorované hodnoty a na ose y jsou udány jejich kumulativní relativní četnosti transformované do standardizovaného normálního rozdělení (očekávané normální hodnoty), viz obrázek 5.5. Pro data z normálního rozdělení budou body grafu opět ležet přibližně na přímce.

Obrázek 5.5: Normální graf dat z příkladu 5.9



[Další](#) [Výš](#) [Předchozí](#) [Obsah](#) [Rejstřík](#)

Další: [6. Odhady populačních průměrů a pravděpodobností](#) **Výš:** [Základy statistiky pro biomedicínské obory](#)

Předchozí: [4. Náhodná veličina a rozdělení pravděpodobnosti](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [7. Testování hypotéz](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [5. Výběr a popisné statistiky](#)

Kapitola

- [6.1 Rozdělení výběrového průměru](#)
- [6.2 Intervaly spolehlivosti](#)
- [6.3 Rozsah výběru](#)

6. Odhady populačních průměrů a pravděpodobností

Nemusíte sníst celého vola na to, abyste poznali, že maso je tuhé.

S. Johnson

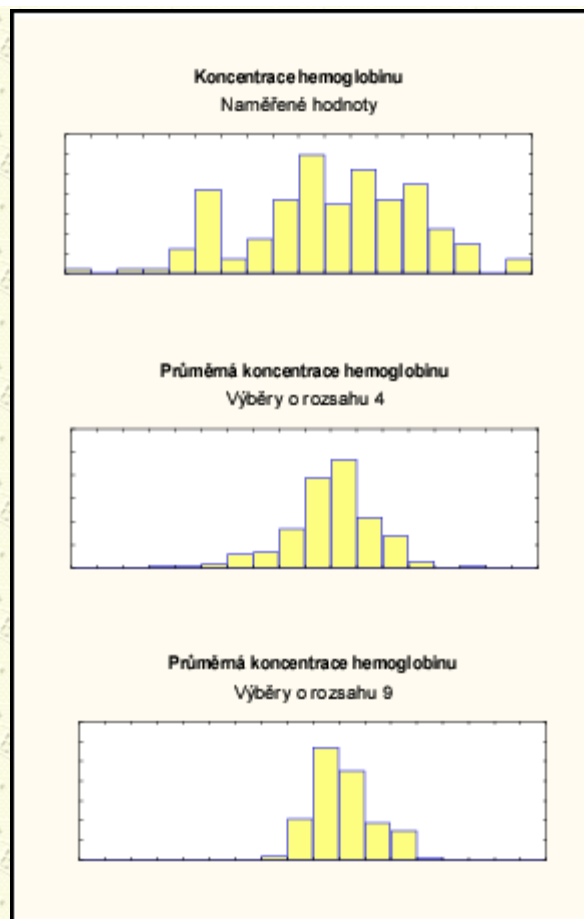
Obvyklým případem v medicíně (jakož i jinde) je, že máme poměrně malý počet pozorování nějaké veličiny a chceme učinit závěry o tom, co bychom obdrželi, kdybychom měli pozorování mnohokrát více, třeba tisíckrát. V této části si ukážeme, jak odhadnout populační průměr z výběru pomocí tzv. "intervalu spolehlivosti". K tomu však potřebujeme vědět něco o tom, jak kolísají výběrové průměry (\bar{x}) kolem populačního průměru (μ).

6.1 Rozdělení výběrového průměru

Viděli jsme, že normální rozdělení je přirozeným modelem jak pro některé znaky u lidí, tak pro chybu měření. Je zde však ještě třetí důvod, proč má toto rozdělení zásadní význam v analýze dat: když provedeme výběr a spočítáme průměry, pak se tyto průměry budou obvykle chovat tak, jako kdyby pocházely z normálního rozdělení. Tento důvod je o něco abstraktnější než první dva, takže abychom to trochu přiblížili, prohlédneme si výsledky experimentu provedeného před několika lety v Trinity College v Dublinu.

Smíšená třída 160 studentů přírodních věd a medicíny byla požádána o změření koncentrace hemoglobinu ve vlastní krvi. Studenti dali dohromady seznam všech získaných hodnot a dostali pokyny, jak z nich mají (pomocí náhodných čísel) provést náhodný výběr. Všichni provedli dva náhodné výběry, první o rozsahu 4 a druhý 9 pozorování, a spočítali průměrné hodnoty v každém z nich. Tyto výběrové průměry byly odevzdány třídnímu učiteli. K dispozici tedy bylo celkem 160 původních měření, 160 průměrů z výběrů o rozsahu 4 a 160 průměrů z výběrů o rozsahu 9. Pro všechny tři sady čísel byly nakresleny histogramy - viz obrázek [6.1](#).

Obrázek 6.1: Histogramy koncentrace hemoglobinu pro naměřené hodnoty a výběrové průměry při rozsahu výběru $n = 4$ a $n = 9$



Na tomto obrázku jsou pozoruhodné tři skutečnosti:

1. všechny tři histogramy kolísají kolem stejného středu,
2. čím větší rozsah výběru, tím užší rozdělení,
3. rozdělení průměrů pro $n = 4$ a $n = 9$ jsou podobnější normálnímu rozdělení než rozdělení původních dat.

První dva body jsou celkem pochopitelné - provedeme-li náhodný výběr, dostaneme většinou několik velkých a několik malých hodnot, takže výběrový průměr bude mít sklon padnout do blízkosti středu původního rozdělení. Rozdělení velkého počtu takových výběrových průměrů bude tedy užší než původní rozdělení a bude mít stejný střed. Je rozumné očekávat, že čím větší bude rozsah výběru, tím více se bude průměr výsledného rozdělení blížit středu původního rozdělení a výsledné rozdělení bude užší. Matematicky lze dokázat, že směrodatná odchylka rozdělení výběrových průměrů je rovna směrodatné odchylce původního rozdělení vydělené druhou odmocninou rozsahu výběru. Směrodatná odchylka výběrového rozdělení průměrů se nazývá **směrodatná chyba průměru** (nebo též **střední chyba průměru**).

Důvod, proč je rozdělení výběrových průměrů podobnější normálnímu rozdělení než rozdělení původní, je méně zřejmý. Je ale jasné, že výběrové průměry budou shluknuty kolem středu (je-li ve výběru přibližně stejný počet velkých i malých hodnot) a že stále méně hodnot bude od středu vzdáleno.

Rozdělení vzniklé opakovaným vybíráním a výpočtem výběrového průměru se nazývá **rozdělení výběrového**

průměru. Matematicky lze dokázat, že (téměř) bez ohledu na tvar původního rozdělení se rozdělení výběrového průměru blíží k normálnímu rozdělení pro rozsah výběru jdoucí do nekonečna. V našem příkladu vypadá rozdělení hodnot založených na výběru již o rozsahu 4 jako blízké normálnímu rozdělení. Ve skutečnosti bylo v naší skupině okolo 40 % žen, u nichž je koncentrace hemoglobinu jak známo nižší než u mužů. První histogram z původních dat odpovídá ještě směsi dvou normálních rozdělení, zatímco rozdělení výběrového průměru se už chová jako jediné normální rozdělení.

Zásadní význam toho všeho spočívá v tom, že naprostá většina rozhodnutí a odhadů populačních charakteristik je založena spíše na průměrných hodnotách více měření než na jednotlivých hodnotách. Tudiž když posuzujeme kvalitu takových odhadů (výběrových průměrů), opíráme se více o rozdělení výběrových průměrů než o původní rozdělení.

6.2 Intervaly spolehlivosti

Jen zřídka jsme v situaci, že vyšetříme skupinu věcí či osob a zajímáme se jen o tuto zkoumanou skupinu. Obvykle je tato skupina pokládána za výběr z mnohem větší populace, jejíž vlastnosti chceme zjistit. Například podáme-li lék skupině pacientů s vysokým krevním tlakem a měříme následný pokles tlaku, pak typickým úkolem takové studie je odhadnout průměrný pokles krevního tlaku, který bychom pozorovali po podání léku celé populaci těchto pacientů. Přirozeným odhadem by byl průměrný pokles v našem výběru. Víme ale, že tento odhad je ovlivněn náhodnými výkyvy závisujícími na tom, kteří lidé jsou pro měření vybráni, kdy jsou měření a jaké chyby při měření vzniknou. Proto místo jednoduchého bodového odhadu skutečné hodnoty v populaci dáváme přednost intervalu zkonstruovanému pomocí výběrového průměru, uvnitř kterého skutečná hodnota populačního průměru s vysokou pravděpodobností leží.

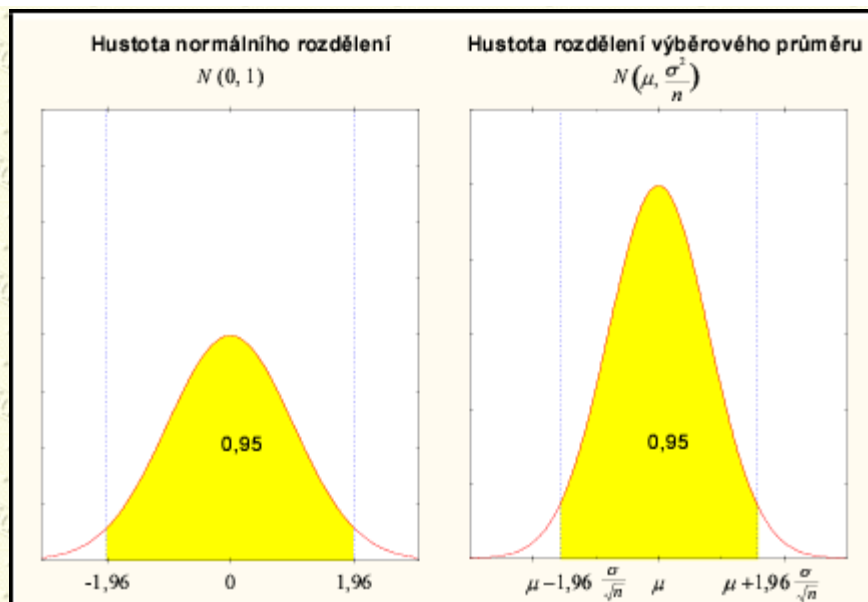
Použijme výše uvedené vlastnosti rozdělení výběrového průměru k sestavení takového intervalu.

6.2.1 Intervalový odhad průměru

Viděli jsme, že když nějaký znak (veličina) má rozdělení s průměrem μ a směrodatnou odchylkou σ (tj. s rozptylem σ^2), pak výběrový průměr bude mít normální rozdělení s průměrem μ a rozptylem σ^2/n . To se často zapisuje jako

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right). \quad (6.1)$$

Obrázek 6.2: Vlastnosti rozdělení výběrového průměru



Obrázek 6.2 ukazuje, že když provedeme náhodný výběr o rozsahu n a spočteme výběrový průměr (\bar{x}), pak jeho hodnota padne s pravděpodobností 0,95 do vzdálenosti menší než $1,96 \sigma / \sqrt{n}$ od populačního průměru μ . Tuto vlastnost lze vyjádřit i jinak: provedeme-li výběr o rozsahu n a spočteme \bar{x} , pak populační průměr μ leží s pravděpodobností 0,95 ve vzdálenosti menší než $1,96 \sigma / \sqrt{n}$ od \bar{x} , tj. interval s krajními body

$$\bar{x} \pm 1,96 \cdot \frac{\sigma}{\sqrt{n}} \quad (6.2)$$

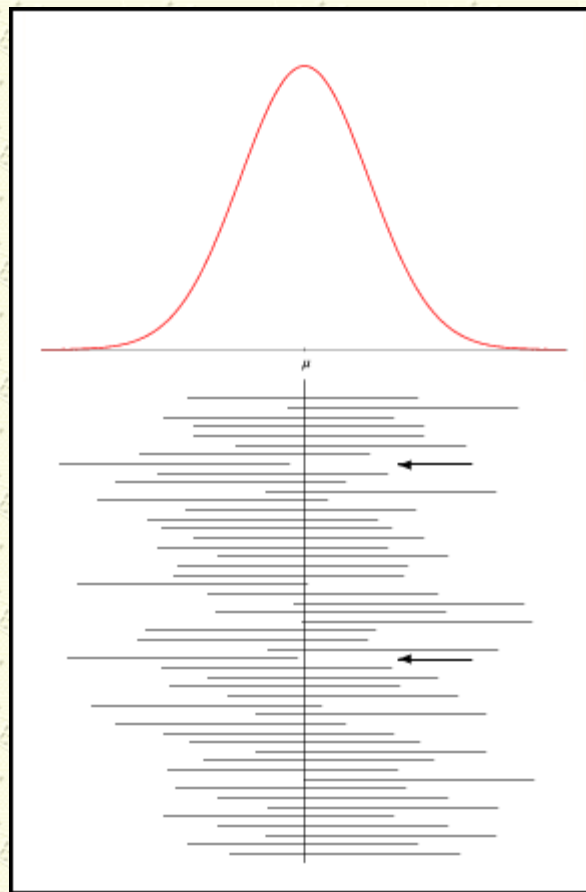
pokrývá populační průměr μ s pravděpodobností 0,95. Tento interval se nazývá **interval spolehlivosti pro průměr**.

Hodnota $z = 1,96$ je kritická hodnota standardizovaného rozdělení pro **koeficient spolehlivosti** $P = 0,95$ (tj. hladinu významnosti $\alpha = (1 - P) = 0,05$ - viz 7.1). Obecně platí, že pro koeficient spolehlivosti $P = (1 - \alpha)$ lze kritickou hodnotu z vyjádřit pomocí kvantilů standardizovaného normálního rozdělení jako $z = |z_{\frac{\alpha}{2}}| = z_{1 - \frac{\alpha}{2}}$. Z tabulky A.1 snadno zjistíme, že pro $\alpha = 0,05$ je kritická hodnota z opravdu 1,96.

Předvedme si nyní celý proces v praxi. Obrázek 6.3 ukazuje křivku rozdělení výběrového průměru se středem μ a směrodatnou odchylkou $\sigma_{\bar{x}} = \sigma / \sqrt{n}$. Z populace se známým průměrem bylo na počítači náhodně vygenerováno padesát výběrů o rozsahu $n = 10$ a pomocí vzorce (6.2) byly spočteny a nakresleny intervaly spolehlivosti. Z padesáti intervalů pouze dva nepokrývají populační průměr μ , který je znázorněn vertikální čarou (tyto dva intervaly jsou na obrázku označeny šipkami). Teorie předvídá, že 5 % takových intervalů nepokryje μ . Naše simulace je tedy v dobrém souladu s teorií.

Obrázek 6.3: Intervaly spolehlivosti pro 50 výběrů

o rozsahu 10



V praxi samozřejmě provedeme pouze jediný výběr o rozsahu n a spočteme intervalový odhad parametru μ .

Nevíme ovšem, zda náš interval pokryje μ , nebo ne. Protože ale známe vlastnosti výběrového rozdělení průměru, můžeme se značnou spolehlivostí předpokládat, že jsme nedostali jeden z těch neobvyklých intervalů, které μ nepokrývají. Proto také vypočtenému intervalu říkáme 95% interval spolehlivosti pro μ .

Chceme-li jiný koeficient spolehlivosti, stačí prostě nahradit číslo 1,96, které odpovídá 95% spolehlivosti, příslušným kvantilem standardizovaného normálního rozdělení, např. 1,645 pro 90% či 2,576 pro 99% spolehlivost (viz tabulku [A.1](#)). Jak se tyto odlišné hodnoty projeví, si ukážeme na jednoduchém příkladu.

Příklad 6.1 Úkolem je určit průměrnou hladinu cholesterolu v séru v určité populaci mužů. V náhodném výběru 25 mužů je výběrový průměr 6,3 mmol/l. Předpokládejme, že v podobné velké populaci byla zjištěna směrodatná odchylka $\sigma = 1,25$ mmol/l a že je možné tuto hodnotu použít i zde.

Řešení: Sestrojíme tři intervaly spolehlivosti odpovídající různým koeficientům pomocí vzorce:

$$\bar{x} \pm z \frac{\sigma}{\sqrt{n}}, \quad (6.3)$$

kde za z dosadíme příslušnou kritickou hodnotu standardizovaného normálního rozdělení, tj. kvantil $z_{1-\frac{\alpha}{2}}$.

Připomeňme, že výraz $\sigma_{\bar{x}} = \sigma / \sqrt{n}$ je směrodatná chyba průměru.

90% interval spolehlivosti ($\alpha = 0,1$):

$$6,3 \pm 1,645 \cdot \frac{1,25}{5} = 6,3 \pm 0,4 \rightsquigarrow (5,9; 6,7),$$

95% interval spolehlivosti ($\alpha = 0,05$):

$$6,3 \pm 1,960 \cdot \frac{1,25}{5} = 6,3 \pm 0,5 \rightsquigarrow (5,8; 6,8),$$

99% interval spolehlivosti ($\alpha = 0,01$):

$$6,3 \pm 2,576 \cdot \frac{1,25}{5} = 6,3 \pm 0,6 \rightsquigarrow (5,7; 6,9). \diamond$$

Všimněte si, že čím vyšší je koeficient spolehlivosti, tím delší - a tedy méně přesný - je výsledný interval! Je proto třeba nalézt kompromis mezi požadovanou spolehlivostí a přesností odhadu, tj. délkou intervalu.

Postup při neznámé směrodatné odchylce v populaci

Jestliže neznáme směrodatnou odchylku v populaci (σ), nahradíme ji výběrovou směrodatnou odchylkou (s). To ale přidá do výpočtu další nejistotu, kterou je třeba uvažovat. Proto místo kritické hodnoty standardizovaného normálního rozdělení použijeme odpovídající kritickou hodnotu tzv. [Studentova \$t\$ rozdělení](#). To je další teoretické rozdělení, podobné normálnímu. Podrobněji se o něm zmíníme v [7. kapitole](#). Místo hledání kritických hodnot v tabulkách normálního rozdělení použijeme tedy tabulky t rozdělení (viz tabulku [A.3](#)). K tomu potřebujeme znát ještě tzv. *počet stupňů volnosti*, což je rozsah výběru minus jedna, tj. $n - 1$. Pro velké výběry ($n > 100$) je t rozdělení velmi podobné normálnímu. Tomu odpovídá poslední řádek (∞) v tabulce [A.3](#), kde jsou hodnoty stejné jako u normálního rozdělení.

Pro neznámou směrodatnou odchylku v populaci (σ) má tedy interval spolehlivosti krajní body

$$\bar{x} \pm t \frac{s}{\sqrt{n}}, \quad (6.4)$$

kde t je příslušná kritická hodnota t rozdělení. Tu můžeme opět vyjádřit pomocí kvantilů Studentova t rozdělení o $n - 1$ stupních volnosti jako $t = t_{1-\frac{\alpha}{2}}(n - 1)$ - viz tabulku [A.3](#). Výraz $s_{\bar{x}} = s / \sqrt{n}$ je odhadem směrodatné chyby průměru $\sigma_{\bar{x}}$.

Příklad 6.2 Vezměme náhodný výběr 20 dospělých Čechů ($n = 20$) a spočtěme průměrnou hmotnost (kg) a

$$(\bar{x} = 70$$

výběrovou směrodatnou odchylku ($s = 5$ kg). Chceme spočítat 95% interval spolehlivosti.

Řešení: Počet stupňů volnosti je $n - 1 = 19$. Koefficient spolehlivosti 95 % odpovídá $\alpha = 0,05$ (tj. $1 - 0,95$) v tabulce [A.3](#). V ní najdeme hodnotu 2,093. Interval spolehlivosti tedy bude

$$70 \pm 2,093 \cdot \frac{5}{\sqrt{20}} \rightsquigarrow (67,7; 72,3). \diamond$$

6.2.2 Intervalový odhad populační pravděpodobnosti

Předpokládejme, že provedeme náhodný výběr o rozsahu n a zjišťujeme počet lidí s danou vlastností. Je-li jich r , pak [relativní četnost](#) výskytu vlastnosti ve výběru je $p = r/n$. Relativní četnost se také často udává v procentech.

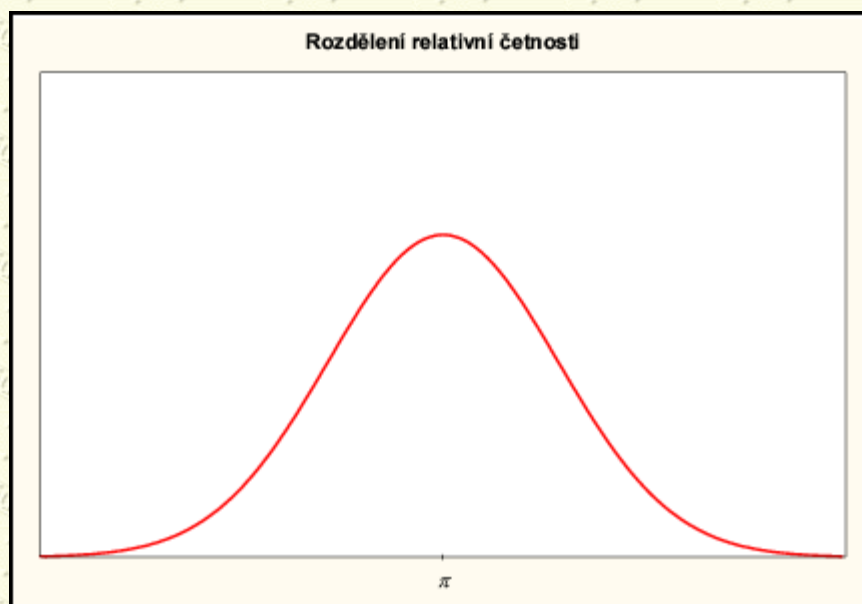
V takovém případě pak hovoříme o **procentu** výskytu dané vlastnosti.

Statistická teorie nám říká, že je-li $n > 100$ a p není příliš malé, pak relativní četnost výskytu vlastnosti má normální rozdělení s průměrem rovným pravděpodobnosti výskytu v celé populaci (π) a směrodatnou odchylkou

$$\sqrt{\pi(1 - \pi)/n}.$$

Graficky to můžeme znázornit následovně - viz obrázek [6.4](#).

Obrázek 6.4: Rozdělení výběrové relativní četnosti



95% interval spolehlivosti pro populační pravděpodobnost π má tedy krajní body

$$p \pm 1,96 \sqrt{\frac{\pi(1-\pi)}{n}}$$

Neznámou hodnotu π ve směrodatné odchylce nahradíme výběrovou relativní četností p a dostaneme 95% **interval spolehlivosti pro populační pravděpodobnost**

$$p \pm 1,96 \sqrt{\frac{p(1-p)}{n}}$$

Máme-li malý rozsah výběru ($n < 30$), použijeme speciální tabulky - viz například [11, tab. 26].

Příklad 6.3 Předpokládejme, že v náhodném výběru 200 mladých mužů má 120 z nich vyšší než doporučenou hladinu cholesterolu v séru. Spočítejte 95% interval spolehlivosti pro procento mladých mužů s vyšší hladinou cholesterolu v populaci.

Řešení: Relativní četnost v uvažovaném výběru je $p = 120/200 = 0,60$, tj. 60 %. 95% interval spolehlivosti pro populační pravděpodobnost π je tedy

$$0,6 \pm 1,96 \sqrt{\frac{0,6(1-0,6)}{200}} = 0,6 \pm 0,07 \rightsquigarrow (0,53; 0,67).$$

Tento odhad vyjádřený v procentech říká, že procento mladých mužů s vyšší než doporučenou hladinu cholesterolu v séru je s 95% spolehlivostí v rozmezí 53 % až 67 %.

6.3 Rozsah výběru

Ze vzorce (6.2) je zřejmé, že rozsah výběru ovlivňuje šířku intervalu spolehlivosti pro danou směrodatnou odchylku a daný koeficient spolehlivosti. Předpokládejme například, že chceme sestavit 95% interval spolehlivosti pro průměrnou hladinu cholesterolu v populaci s přesností 0,4 mmol/l, tj. tvaru: výběrový průměr $\pm 0,2$ mmol/l.

Směrodatná odchylka v populaci necht' je přitom stejně jako v příkladu 6.1 $\sigma = 1,25$. Vypočtený interval pak bude mít krajní body

$$\bar{x} \pm 1,96 \cdot \frac{1,25}{\sqrt{n}}$$

Chceme, aby

$$1,96 \cdot \frac{1,25}{\sqrt{n}} \leq 0,2,$$

tedy

$$n \geq \left(1,96 \cdot \frac{1,25}{0,2} \right)^2 \doteq 150.$$

Abychom dostali 95% interval spolehlivosti s přesností 0,4 mmol/l (tj. ve tvaru výběrový průměr $\pm 0,2$ mmol/l), potřebovali bychom výběr 150 mužů.

[Další](#) [Výš](#) [Předchozí](#) [Obsah](#) [Rejstřík](#)

Další: [7. Testování hypotéz](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [5. Výběr a popisné statistiky](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [8. Porovnání dvou skupin](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [6. Odhady populačních průměrů a pravděpodobností](#)

Kapitola

- [7.1 Nulová a alternativní hypotéza](#)
- [7.2 Statistická vs. klinická významnost](#)
- [7.3 Chyba I. a II. druhu a síla testu](#)
- [7.4 Testování hypotéz o průměru pro jeden výběr](#)

7. Testování hypotéz

Dobré odpovědi pocházejí z dobrých otázek, a ne z esoterické analýzy.

H. M. Schoolman

7.1 Nulová a alternativní hypotéza

Mnohdy je nejlepším způsobem analýzy a zveřejnění výsledků srovnávacích studií použití intervalů spolehlivosti, protože výsledky se tak netváří jako skutečné hodnoty bez náhodné odchylky, ale nejistota v měřené veličině je odhadnuta explicitně. Statistické testování hypotéz je ale stále ještě široce rozšířeno ve vědeckých pracích a téměř povinné v některých časopisech. Je proto vhodné, abychom se použitím takových testů zabývali.

Většina statistických analýz zahrnuje porovnání různých způsobů léčby, metod, či skupin případů. Číselná hodnota odpovídající tomuto porovnání se nazývá *efekt*. Tvrzení, že efekt je nulový, nazveme *nulovou hypotézou*. Znamená to například, že průměry bílkovinného séra jsou stejné pro muže i ženy nebo že dvě léčby bolesti hlavy jsou stejně efektivní.

Nulová hypotéza je obvykle opakem toho, co chceme výzkumem prokázat, když zahajujeme studii a začínáme sbírat data.

V našem příkladu by zkoumanou hypotézou mohla být existence rozdílu v průměru bílkovinného séra mezi muži a ženami. Té se říká *alternativní hypotéza*. Obecně tvrdí, že efekt není nulový.

Shrňme tedy, že pokud budeme testovat hypotézy statistickými metodami, musíme na začátku naší práce zformulovat dvě hypotézy: nulovou hypotézu a alternativní hypotézu. Alternativní hypotéza přesně vymezuje, do jaké situace se dostáváme, když nulová hypotéza neplatí.

Poté, co zformulujeme nulovou hypotézu a nasbíráme data, spočteme pravděpodobnost, s jakou bychom mohli obdržet pozorovaná data nebo data stejně, či ještě více odporující nulové hypotéze, *za předpokladu, že je nulová hypotéza pravdivá*. Tato pravděpodobnost se nazývá **dosažená hladina významnosti** a značí se p (p -value, p -level).

Čím menší je p , tím neudržitelnější čili méně důvěryhodná je nulová hypotéza.

Například v naší studii o bílkovinném séru bychom mohli zjistit, že průměrná hladina albuminu je u mužů vyšší než u žen, přičemž po výpočtu testové statistiky bychom dostali řekněme $p = 0,10$. To by znamenalo, že pokud by ve skutečnosti byla hladina albuminu u obou pohlaví stejná, pak s 10% pravděpodobností dostaneme rozdíl stejně velký, nebo ještě větší než ten námi pozorovaný. Můžeme z pozorovaného rozdílu usoudit na to, že je hladina albuminu u mužů a žen různá? Jak velké nebo malé by mělo být p , abychom zamítli nulovou hypotézu?

Nejprve zavedme **testovou statistiku (testové kritérium)** vhodnou při ověřování hypotéz o parametru polohy:

$$\text{testová statistika} = \frac{\text{pozorovaná hodnota} - \text{očekávaná hodnota}}{\text{směrodatná chyba pozorované hodnoty}}. \quad (7.1)$$

Testovou statistiku vyhodnotíme tak, že spočteme pravděpodobnost, že bychom mohli pozorovat námi zjištěnou, nebo ještě extrémnější (tj. méně pravděpodobnou) hodnotu, pokud by byla nulová hypotéza pravdivá.

Příklad 7.1 Předpokládejme, že máme výběr 216 pacientů a změřili jsme jejich bílkovinné sérum. Průměr vyšel 34,46 g/l se směrodatnou chybou 0,397 g/l. Chceme ověřit nulovou hypotézu, že průměrné bílkovinné sérum všech pacientů tohoto typu (populační průměr μ) je 33,5 g/l proti alternativě, že průměrné bílkovinné sérum se liší od hodnoty 33,5 g/l, tj.

$$H_0: \mu = 33,5,$$

$$H_1: \mu \neq 33,5 \quad (\text{oboustranná alternativa}).$$

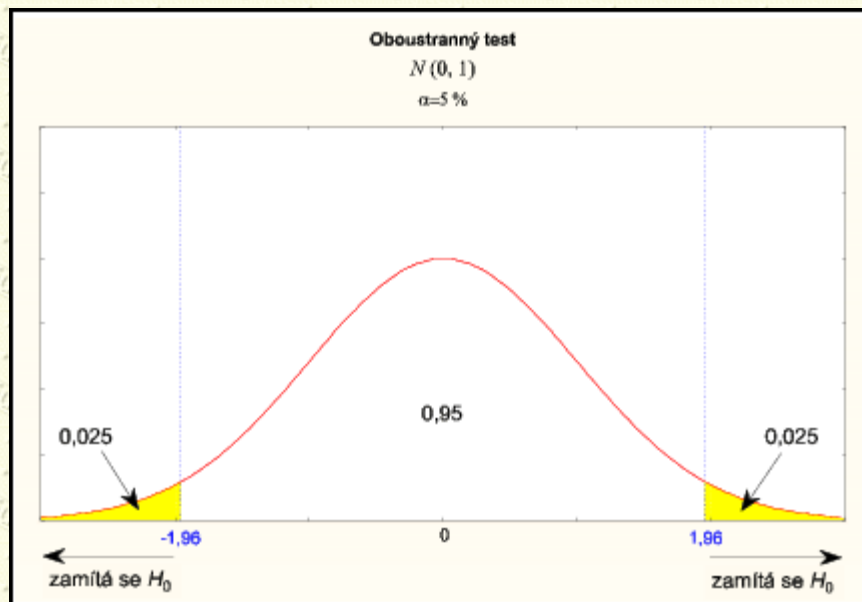
Řešení: Měříme-li spojitou veličinu a máme-li dostatečně velký rozsah výběru (větší než 60), můžeme použít normální rozdělení. V takovém případě se testová statistika označuje Z . Použitím vzorce (7.1) dostaneme testovou statistiku

$$\begin{aligned} Z &= \frac{\text{výběrový průměr} - \text{očekávaný průměr při } H_0}{\text{směrodatná chyba výběrového průměru}} = \\ &= \frac{34,46 - 33,5}{0,397} = 2,418. \end{aligned}$$

Hodnota $Z = 2,418$ leží mezi 1,960 a 2,576, což jsou kritické hodnoty standardizovaného normálního rozdělení odpovídající 95%, resp. 99% koeficientu spolehlivosti neboli 5%, resp. 1% hladině významnosti (α) pro oboustrannou alternativu, jak uvidíme později. (Nalezneme je jako kvantily $z_{1-\frac{\alpha}{2}}$ standardizovaného normálního rozdělení v tabulce A.1.) Můžeme tedy učinit závěr, že pravděpodobnost, že dostaneme stejně velkou, nebo ještě větší absolutní hodnotu testové statistiky, je-li nulová hypotéza pravdivá, je mezi 1 % a 5 % (tj. $0,01 < p < 0,05$). \diamond

Obecně, leží-li testová statistika v *oboru přijetí*, nezamítáme H_0 . Leží-li testová statistika v *kritickém oboru*, zamítáme H_0 a prohlásíme výsledek za statisticky významný - viz obrázek 7.1, kde je oborem přijetí interval $(-1,96; 1,96)$ a kritickým oborem intervaly $(-\infty; -1,96)$ a $(1,96; +\infty)$.

Obrázek: Oboustranný test H_0



Pojem "hladina významnosti" se hojně vyskytuje v medicínských (i jiných) výzkumných pracích, takže je nezbytné přesně rozumět tomu, co znamená, ale také tomu, co neznamená.

Jak jsme již uvedli, dosažená hladina významnosti p je pravděpodobnost, že *pozorujeme naše data nebo data stejně či více extrémní, když je nulová hypotéza pravdivá*.

Interpretaci dosažené hladiny významnosti (p) můžeme uvést tímto příkladem. Když porovnáme dvě léčby a dostaneme vysoké p , řekněme vyšší než 0,10 (neboli 10 %), pak můžeme tvrdit, že taková data, jako jsou naše, bychom mohli dostat celkem často i v případě, že platí nulová hypotéza. Nelze proto vyloučit možnost, že nulová hypotéza je pravdivá, tj. že obě léčby jsou stejně efektivní.

Naopak je-li p velmi malé, řekněme menší než 0,001 (neboli 0,1 %), pak se nulová hypotéza zdá být téměř nemožnou, protože naše data by mohla sotva kdy vzniknout pouze náhodou, kdyby platila nulová hypotéza. Můžeme tedy tvrdit se značnou spolehlivostí, že nulová hypotéza není pravdivá a jedna léčba je prokazatelně lepší než druhá.

Mezi těmito dvěma extrémny leží šedá oblast nejistoty. Postupuje se tak, že se zvolí jistá mezní hodnota α , říká se jí **hladina významnosti**, a pokud je p menší než tato mez, zamítáme nulovou hypotézu. Test nulové hypotézy tedy spočívá v tom, zda p překročí, nebo nepřekročí zvolenou mez. Za tuto mezní hodnotu se často bere 0,05 (neboli 5 %). Jinými slovy - výsledek, který bychom mohli za platnosti nulové hypotézy dostat méně než jednou z dvaceti případů, vede k zamítnutí nulové hypotézy.

Zamítáme-li nulovou hypotézu, přijmeme hypotézu alternativní, která v našem příkladu říká, že léčby nejsou stejně

účinné - jedna se ukazuje být prokazatelně lepší.

Je-li p menší než zvolená hladina významnosti, řekněme 0,05, pak výsledek je **statisticky významný** (je-li menší než nějaká nižší hodnota, jako např. 0,01, pak můžeme mluvit o *vysoké statistické významnosti*). Pokud je naopak p větší než 0,05, říkáme, že výsledek *není statisticky významný*.

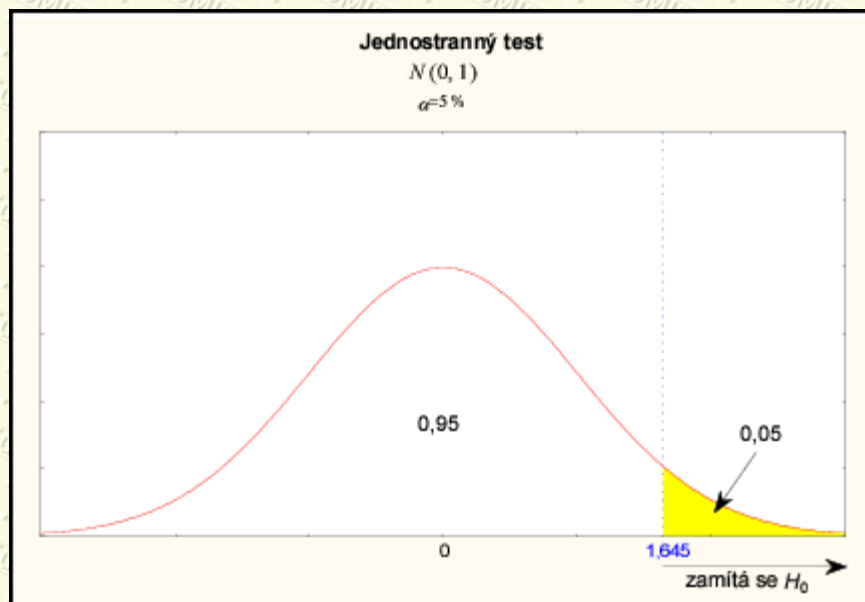
Příklad 7.2 Vraťme se znovu k situaci z příkladu 7.1, ale tentokrát budeme zjišťovat, zda je populační průměr μ větší než hodnota $\mu_0 = 33,5$ g/l. V tomto případě zformulujeme jednostrannou alternativní hypotézu

$$H_1: \mu > 33,5 \text{ g/l.}$$

Řešení: K zamítnutí nulové hypotézy nyní povede jen velká hodnota testové statistiky Z . Kritický obor a obor přijetí vypadají proto jinak než u oboustranné alternativy (viz obrázek 7.2). Testová statistika $Z = 2,418$ se nyní porovnává s kritickou hodnotou, která je na rozdíl od oboustranné alternativy rovna kvantilu $z_{1-\alpha}$ (nikoli tedy $z_{1-\frac{\alpha}{2}}$)

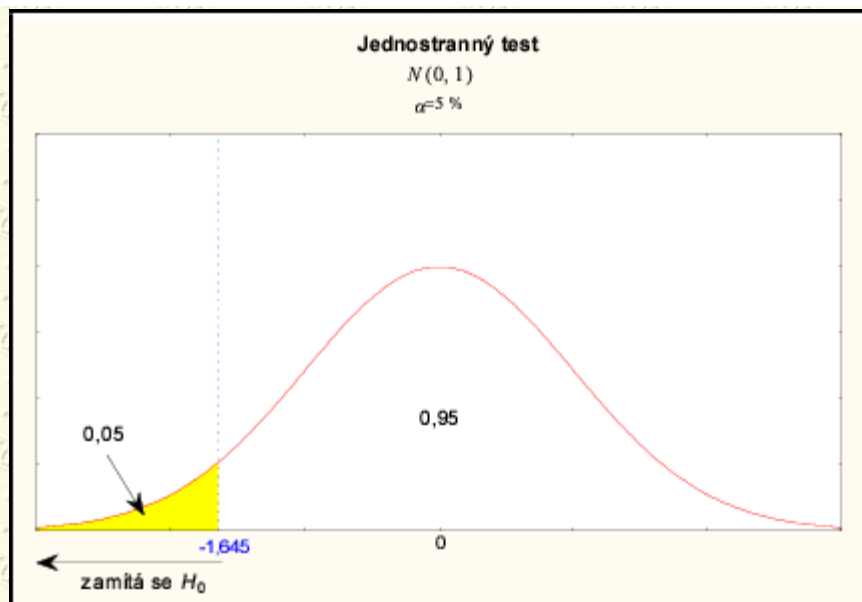
standardizovaného normálního rozdělení. Pro $\alpha = 0,05$ je $z_{1-\alpha} = 1,645$ a pro $\alpha = 0,01$ je $z_{1-\alpha} = 2,326$ - viz tabulku A.1. Protože $Z = 2,418 > 2,326$, vidíme, že volba jednostranné alternativy vede k zamítnutí nulové hypotézy i na 1% hladině významnosti. \diamond

Obrázek: Jednostranný test H_0 proti $H_1: \mu > \mu_0$



Pokud bychom měli jednostrannou alternativu $H_1: \mu < \mu_0$, pak by k zamítnutí nulové hypotézy vedla naopak malá hodnota testové statistiky Z . Pro hladinu významnosti $\alpha = 0,05$ by kritickou hodnotou byl kvantil $z_\alpha = -1,645$. Nulovou hypotézu bychom zamítli v případě, že $Z \leq z_\alpha$ - viz obrázek 7.3.

Obrázek: Jednostranný test H_0 proti $H_1: \mu < \mu_0$

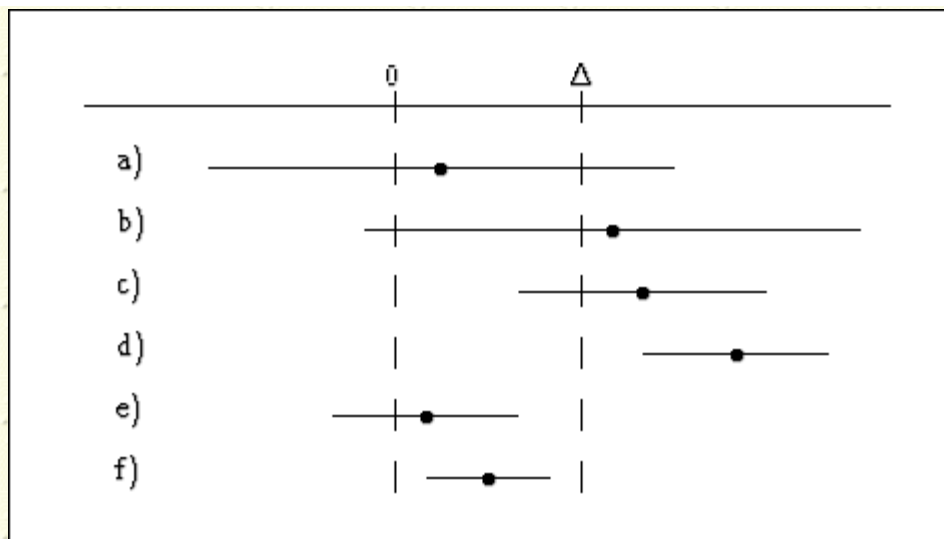


7.2 Statistická vs. klinická významnost

Použití slova "významnost" vede k mnoha nedorozuměním, zda je míněna významnost statistická, nebo klinická. Obvyklá praxe je pokládat statisticky významný výsledek za skutečný efekt, a v důsledku toho i za výsledek klinicky důležitý, a naopak. Tato interpretace však nemusí být nutně oprávněná. Např. při porovnávání krevního tlaku na levé a pravé ruce byl zjištěn průměrný rozdíl 1 mm Hg. Tento rozdíl je vysoce statisticky významný (díky velkému rozsahu výběru), ale není důležitý klinicky. Podobně není rozumné brát nevýznamný výsledek jako indikaci neexistence efektu jen proto, že nemůžeme zamítnout nulovou hypotézu. (Je možné, že experiment měl nedostatečnou sílu způsobenou příliš malým počtem pozorování - viz diskuse o síle níže.)

Pokud klinicky významný rozdíl populačních průměrů je dán překročením konstanty Δ , můžeme tuto skutečnost interpretovat například pomocí 95% intervalu spolehlivosti pro rozdíl populačních dat. Na obrázku [7.4](#) si nyní ukážeme možnosti, které mohou nastat. Jejich interpretace je rozebrána v tabulce [7.1](#).

Obrázek 7.4: Interval spolehlivosti



Tabulka 7.1: Interpretace intervalů spolehlivosti

Možnost	Statistická významnost	Klinická významnost
a)	ne	možná
b)	ne	možná
c)	ano	možná
d)	ano	ano
e)	ne	ne
f)	ano	ne

7.3 Chyba I. a II. druhu a síla testu

Rozhodování o nulové hypotéze na 5% nebo jiné hladině významnosti na základě náhodného výběru může vést ke vzniku dvou možných chyb.

Za prvé můžeme dostat významný výsledek (řekněme $p \leq 0,05$) a zamítnout nulovou hypotézu, i když ve skutečnosti platí. To se nazývá **chybou I. druhu**. Pravděpodobnost chyby I. druhu je v tomto případě 5 % a může být předem stanovena výzkumníkem. Za druhé můžeme dostat nevýznamný výsledek (řekněme $p > 0,05$), i když nulová hypotéza ve skutečnosti neplatí. V takovém případě se dopustíme **chyby II. druhu**.

Tabulka: Možné výsledky statistického testu

	Skutečnost

Rozhodnutí	H_0 platí	H_0 neplatí, platí H_1
Nemůžeme zamítnout H_0 (nevýznamný výsledek)	Správné rozhodnutí	Chyba II. druhu (s pravděpodobností β)
Zamítneme H_0 (významný výsledek)	Chyba I. druhu (s pravděpodobností α)	Správné rozhodnutí

Pravděpodobnosti chyby I. a II. druhu jsou označovány α (alfa) a β (beta). Hodnota β závisí jak na velikosti zkoumaného efektu, tak na rozsahu výběru. Častěji hovoříme o *síle testu*, což je $1 - \beta$. Přílišná šířka dříve zmíněného intervalu spolehlivosti svědčí o nízké síle. Lze také předem zajistit β volbou vhodného rozsahu výběru, jinými slovy, můžeme spočítat, jak velký rozsah výběru budeme potřebovat, abychom skutečný efekt dané velikosti odhalili s vysokou pravděpodobností.

7.4 Testování hypotéz o průměru pro jeden výběr

Předpokládejme, že máme náhodný výběr n lidí, u nichž měříme nějakou spojitou veličinu (např. hladinu albuminu, výšku atd.). Chceme testovat hypotézu, že průměr populace, z níž výběr pochází, je roven určitému číslu - tzv. očekávané hodnotě. Jak již víme z odstavce 7.1, testová statistika má obecný tvar

$$Z = \frac{\text{pozorovaná hodnota} - \text{očekávaná hodnota}}{\text{směrodatná chyba pozorované hodnoty}},$$

kde pozorovaná hodnota je průměr naměřených hodnot (\bar{x}), a směrodatná chyba pozorované hodnoty je tedy $\sigma_{\bar{x}}$.

Řekli jsme si také, že pro dostatečně velký rozsah výběru má statistika Z normální rozdělení $N(0,1)$. Ve většině případů však neznáme přímo směrodatnou chybu $\sigma_{\bar{x}}$, ale pouze její odhad $s_{\bar{x}}$ vypočtený z našeho výběru. V praxi proto obvykle místo testové statistiky Z počítáme statistiku

$$T = \frac{\bar{x} - \mu}{s_{\bar{x}}}. \quad (7.2)$$

Tím, že jsme nahradili směrodatnou chybu $\sigma_{\bar{x}}$ jejím odhadem $s_{\bar{x}}$, jsme do výpočtu vnesli ještě další nejistotu, kterou je třeba brát v úvahu. Ze statistické teorie vyplývá, že testová statistika T již nemá normální rozdělení, ale tzv. *Studentovo t rozdělení* o $n - 1$ stupních volnosti. Proto se také tomuto postupu říká *jednovýběrový t test*.

Historická poznámka: Roku 1908 vydal slavný statistik William Gosset (který strávil celý život prací v Guinness Brewery nejprve v Dublinu a potom v Londýně, kde se stal roku 1935 ředitelem pivovaru v Park Royal) pod

pseudonymem "Student" důležitý článek v časopise Biometrika. Podstatou bylo, že Gosset v tomto článku ukázal, že zatímco chování průměrů z rozsáhlých výběrů lze dobře popsat pomocí normálního rozdělení, u malých výběrů je třeba provést určitou modifikaci.

Je-li rozsah výběru "malý", mohou se změnit dvě věci:

1.

Výběrová směrodatná odchylka (s), která sama podléhá variabilitě výběru, již nemusí být spolehlivým odhadem populační směrodatné odchylky (σ).

2.

Není-li rozdělení populace normální, nemusí být normální ani rozdělení výběrových průměrů. Tento druhý efekt je důležitý pouze tehdy, je-li rozsah výběru velmi malý (obvykle menší než 15) a rozdělení populace se extrémně liší od normálního. (Důležitá matematická věta, známá pod názvem *centrální limitní věta*, totiž říká, že i když není náhodná veličina normálně rozdělena, rozdělení výběrového průměru se blíží, za splnění rozumných podmínek, normálnímu rozdělení).

Z tvrzení centrální limitní věty vyplývá, že použití normálního rozdělení při testování hypotéz či výpočtech intervalů spolehlivosti z malých výběrů znehodnocuje zejména první z uvedených bodů, tedy výběrová variabilita s . Pro takové výběry je proto nutné použít t rozdělení.

Pod pojmem *Studentovo t rozdělení* se ve skutečnosti skrývá řada rozdělení, z nichž každé je charakterizováno tzv. počtem stupňů volnosti, zkracovaným na df (degrees of freedom). Existuje tedy t rozdělení o jednom, dvou, třech atd. stupních volnosti. V případě jednovýběrových testů, kterými se nyní zabýváme, je počet stupňů volnosti roven rozsahu výběru bez jedné, tj. $n - 1$. Je-li tedy rozsah výběru 10, pak testovou statistiku porovnáme s kritickou hodnotou t rozdělení o $10 - 1 = 9$ stupních volnosti a tak podobně. Stejně jako normální rozdělení je i t rozdělení symetrické, jednovrcholové a zvonovité.

Pro oboustrannou alternativu

$$H_1: \mu \neq \mu_0,$$

kde μ_0 je konstanta (nejčastěji $\mu_0 = \bar{u}$), je kritická hodnota rovna kvantilu $t_{1-\frac{\alpha}{2}}(df)$ Studentova t rozdělení, kde α je hladina významnosti a df je počet stupňů volnosti (tj. $df = n - 1$) - viz tabulku [A.3](#). Nulovou hypotézu zamítáme, pokud je absolutní hodnota testové statistiky větší nebo rovna kritické hodnotě ($|T| \geq t_{1-\frac{\alpha}{2}}(df)$), tj.

$$T \geq t_{1-\frac{\alpha}{2}}(df) \text{ nebo } T \leq -t_{1-\frac{\alpha}{2}}(df).$$

Pro jednostrannou alternativu

$$H_1: \mu > \mu_0$$

je kritická hodnota rovna kvantilu $t_{1-\alpha}(df)$. Nulovou hypotézu zamítáme, pokud je hodnota testové statistiky větší nebo rovna kritické hodnotě, tj. $T \geq t_{1-\alpha}(df)$.

Pro jednostrannou alternativu

$$H_1: \mu < \mu_0$$

je kritická hodnota rovna kvantilu $t_{\alpha}(df)$. Ze symetrie t rozdělení plyne, že $t_{\alpha}(df) = -t_{1-\alpha}(df)$, takže hodnoty kvantilů $t_{\alpha}(df)$ snadno určíme z tabulky A.4. Nulovou hypotézu zamítáme, pokud je hodnota testové statistiky menší nebo rovna kritické hodnotě, tj. $T \leq t_{\alpha}(df)$.

Příklad 7.3 Výběr 100 pacientů s rakovinou plic léčených novým lékem má průměrnou dobu přežití 27,5 měsíce se směrodatnou odchylkou 25,0 měsíce. Z předchozích studií je známo, že průměrné přežití takových pacientů bez podávání nového léku je 22,2 měsíce. Lze na základě těchto dat usoudit, že nový lék prodlužuje přežití?

Řešení: Naše nulová hypotéza říká:

$$H_0: \mu = 22,2 \text{ měsíce.}$$

Alternativní hypotéza tvrdí, že doba přežití pacientů, kterým byl podáván nový lék, se prodlouží oproti době přežití neléčených pacientů (jednostranná alternativa):

$$H_1: \mu > 22,2 \text{ měsíce.}$$

Testová statistika Z by měla tvar

$$Z = \frac{\text{pozorovaná hodnota} - \text{očekávaná hodnota}}{\text{směrodatná chyba pozorované hodnoty}} = \frac{\bar{x} - 22,2}{\sigma_s}$$

Protože směrodatnou chybu σ_s neznáme, odhadneme ji jako

$$s_s = \frac{s}{\sqrt{n}} = \frac{25}{\sqrt{100}} = 2,5 \text{ měsíce.}$$

Místo Z tedy vypočteme testovou statistiku T

$$T = \frac{\bar{x} - 22,2}{s_s} = \frac{27,5 - 22,2}{2,5} = 2,120,$$

která má t rozdělení o $100 - 1 = 99$ stupních volnosti.

Protože provádíme jednostranný test (tj. test proti jednostranné alternativě), je kritická hodnota rovna kvantilu $t_{1-\alpha}(df)$ Studentova t rozdělení. Pro $\alpha = 0,05$ a $df = 99$ je tedy kritická hodnota rovna 1,660, jak zjistíme z tabulky

A.4 (místo $df = 99$ použijeme přibližnou hodnotu $df = 100$). Jelikož hodnota statistiky $T = 2,120$ překračuje kritickou hodnotu 1,660, řekneme, že můžeme zamítnout nulovou hypotézu. Rozdíl v době přežití je významný na hladině $p < 0,05$. \diamond

V předchozím příkladu bychom ovšem klidně mohli použít i kritické hodnoty normálního rozdělení, aniž bychom se tím dopustili nějaké velké chyby. Rozsah výběru (a tedy i počet stupňů volnosti) je totiž poměrně velký, takže odlišnost mezi t rozdělením a normálním rozdělením je prakticky zanedbatelná. Opravdu, 5% kritická hodnota normálního rozdělení, tj. kvantil $z_{1-\alpha} = 1,645$ (viz tabulku [A.1](#)), se příliš neliší od příslušné kritické hodnoty t rozdělení 1,660. V následujícím příkladu je ale rozsah výběru tak malý, že použití kritických hodnot normálního rozdělení by hrubě zkreslilo výsledek testu.

Příklad 7.4 Byla vypracována studie o dětech s vrozenou srdeční vadou, která zjišťovala, v jakém věku začínají děti s touto vadou chodit. Ve studii bylo 9 dětí, průměrný věk, kdy začínaly chodit, byl 12,8 měsíce a směrodatná odchylka byla 2,4 měsíce. Z předchozích výzkumů je známo, že v celé populaci dětí byl průměrný věk, kdy začínaly chodit, 11,4 měsíce. Je možné na základě této studie prohlásit, že srdeční vada negativně ovlivňuje věk, kdy začíná dítě chodit?

Řešení: V tomto příkladu se jedná o test nulové hypotézy

$$H_0: \mu = 11,4 \text{ měsíce}$$

proti jednostranné alternativě

$$H_1: \mu > 11,4 \text{ měsíce.}$$

Testová statistika

$$T = \frac{\bar{x} - 11,4}{s_{\bar{x}}},$$

kde

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2,4}{\sqrt{9}} = 0,8 \text{ měsíce.}$$

Máme tedy

$$T = \frac{12,8 - 11,4}{0,8} = 1,750.$$

Protože jde o jednostranný test, je kritická hodnota rovna kvantilu $t_{1-\alpha}(df)$, kde $df = 9 - 1 = 8$. Pro $\alpha = 0,05$ je tedy kritická hodnota rovna 1,860 (viz tabulku [A.4](#)). Protože statistika $T = 1,750$ nepřekročila kritickou hodnotu 1,860, nemůžeme zamítnout nulovou hypotézu. Na základě našeho výběru tedy nelze prokázat, že by vrozená srdeční vada měla negativní vliv na věk, v němž začíná dítě chodit. \diamond

Všimněte si, že pokud bychom v příkladu [7.4](#) použili kritickou hodnotu normálního rozdělení, tj. kvantil $z_{1-\alpha} = 1,645$, nulovou hypotézu bychom zamítli, přestože dosažená hladina významnosti je ve skutečnosti větší než 0,05.

[Další](#) [Výš](#) [Předchozí](#) [Obsah](#) [Rejstřík](#)

Další: [8. Porovnání dvou skupin](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [6. Odhady populačních průměrů a pravděpodobností](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [9. Neparametrické metody](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [7. Testování hypotéz](#)

Kapitola

- [8.1 Nezávislé výběry](#)
- [8.2 Párová pozorování](#)

8. Porovnání dvou skupin

Je málo způsobů jak správně provést výzkumnou studii, ale jsou jich tisíce jak ji provést špatně.

D. L. Sackett

8.1 Nezávislé výběry

Jedním z nejobvyklejších použití statistiky v lékařském výzkumu je srovnávání dvou souborů pozorování, z nichž každý se vztahuje k jiné skupině pacientů. Například v lékařském procesu může jedna skupina dostávat placebo, zatímco druhá je aktivně léčena. Cílem je prokázat rozdíl mezi účinky placeba a aktivní léčby.

8.1.1 Porovnání průměrů

V lékařském výzkumu se obvykle zajímáme o rozdíl mezi populačním průměrem sledované veličiny v ošetřované skupině (treatment mean) a populačním průměrem této veličiny v kontrolní skupině (control mean). Sledujme tedy rozdíl mezi dvěma výběrovými průměry.

Výběrové průměry se mění výběr od výběru. Viděli jsme, že jeden výběrový průměr z dané populace nemusí být stejný jako jiný výběrový průměr, takže se mění i rozdíly mezi výběrovými průměry (tj. rozdíl mezi prvním a druhým výběrem nemusí být nutně stejný jako rozdíl mezi prvním a třetím nebo druhým a třetím a tak dále).

Rozdělení rozdílů výběrových průměrů má nulovou střední hodnotu se směrodatnou chybou, která je určena směrodatnou odchylkou celé populace a rozsahem výběrů. Protože běžně neznáme směrodatnou odchylku celé populace, používáme obvykle směrodatné odchylky výběrů.

U nezávislých skupin pozorování nás zajímá rozdíl populačních průměrů ve skupinách, přičemž důležitou roli hraje rovněž variabilita uvnitř skupin. Jak interval spolehlivosti pro rozdíl mezi populačními průměry, tak tzv. dvouvýběrový *t* test (viz dále) jsou založeny na předpokladu, že každý soubor pozorování je výběrem z populace s normálním rozdělením a že rozptyly těchto dvou populací jsou shodné.

Příklad 8.1 Podívejme se na výzkum syndromu náhlého dětského úmrtí. Tento syndrom je hlavní příčinou úmrtí dětí v rozvinutém světě. V posledních dvaceti letech došlo ve výzkumu této oblasti k významnému pokroku. V naší studii se zaměříme na dvě skupiny dětí. První skupinu tvoří děti, které byly nalezeny téměř mrtvé bez známek života

a vyžadovaly resuscitaci. Všechna další lékařská vyšetření byla negativní a děti se z této příhody plně zotavily v průběhu následujících 48 hodin až osmi dnů, kdy probíhala další pozorování. Tuto skupinu dětí jsme nazvali "téměř ztracené". Často se o nich také hovoří jako o A.L.T.E. (Apparent Life Threatening Events - události, které očividně ohrožují život). Druhou skupinu tvoří normální děti. U každého dítěte byla za standardizovaných podmínek určena dlouhodobá proměnlivost tepové frekvence (LTV - long term variability). LTV byla definována jako rozdíl mezi minimálními a maximálními hodnotami novorozenecké tepové frekvence po dobu 512 R vln vyjádřených v úderech za minutu. Údaje jsou v tabulce [8.1](#).

Tabulka 8.1: Hodnoty LTV pro obě skupiny

Téměř ztracené		Normální
9,33	13,33	29,00
15,50	11,67	17,33
21,17	8,17	17,83
13,83	9,17	11,33
24,67	23,00	14,33
18,00	7,67	31,33
9,33	9,67	20,67
7,00	17,33	27,83
8,83	22,33	32,00
5,00	8,33	19,00
20,60	15,17	32,50
22,67		22,33
14,17		35,00
11,00		31,17
9,33		13,67

Pozorování pocházejí ze dvou různých skupin dětí - skupiny téměř ztracených dětí a skupiny normálních dětí. V každém výběru jsou pozorování nezávislá. Někdy můžeme mít u každého případu několik pozorování, např. trojí měření množství enzymu v krvi. V takové situaci uvažujeme pouze jedno měření u každého případu, protože tato tři měření by nebyla nezávislá. Například když je u určitého dítěte první naměřená hodnota vysoká, další dvě budou pravděpodobně také vysoké, jinými slovy, jedna ze tří naměřených hodnot nám poví něco i o dalších.

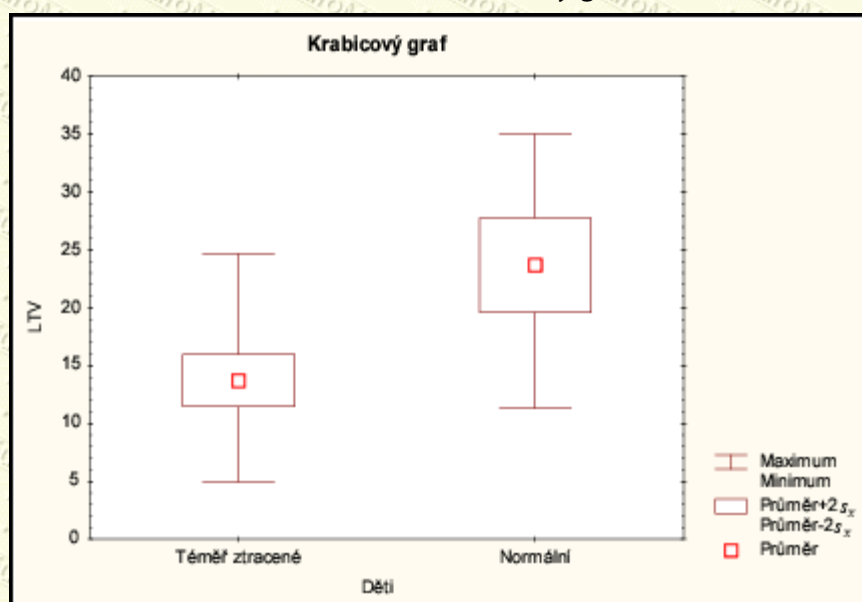
V tomto příkladu máme prostý náhodný výběr z jedné skupiny a jiný prostý náhodný výběr z jiné skupiny. Naše měření je kvantitativní a ze zjištěných údajů se chceme dozvědět něco o skupinových průměrech. Také by nás mohlo zajímat, zda jsou skupinové průměry stejné, nebo zda lze dokázat, že jsou odlišné.

Řešení: Nejprve zakreslíme údaje do grafu a odvodíme souhrnné statistiky. Vhodnou metodou je použití krabicového či bodového grafu - viz obrázky 8.1 a 8.2.

Obrázek 8.1: Bodový graf



Obrázek 8.2: Krabicový graf



Někdo by mohl namítnout, že máme v každé skupině příliš mnoho pozorování, než abychom mohli údaje zobrazit pomocí bodového grafu. Body se totiž vzájemně překrývají. V takovém případě nám dá více informací krabicový graf. Z grafu lze vyčíst, že průměr LTV je u normálních dětí vyšší než u dětí téměř ztracených. Ještě je však nutné ověřit, že se v žádné skupině nevyskytuje pozorování, které by se výrazně lišilo od ostatních. Pokud se tak stane,

musíme nejprve zkontrolovat vstupní údaje a zjistit, zda odlišnost pozorování nebyla způsobena jen chybou při záznamu či přenosu dat do počítače.

Dále spočteme souhrnné statistiky pro obě skupiny (viz tab. 8.2).

Tabulka 8.2: Souhrnné statistiky

Děti	n	\bar{x}	s
Téměř ztracené	26	13,7	5,82
Normální	15	23,7	7,94

Jak jsme se již zmínili, zajímá nás, zda po léčbě došlo k nějakému rozdílu. Zavedeme tedy nulovou a alternativní hypotézu následovně:

H_0 :

Neexistuje žádný rozdíl v průměrných hodnotách LTV mezi oběma skupinami, což se často zapisuje jako $\mu_1 = \mu_2$ nebo $\mu_1 - \mu_2 = 0$, kde μ_1 je populační průměr ve skupině "téměř ztracených" dětí a μ_2 je populační průměr ve skupině "normálních" dětí.

H_1 :

Existuje rozdíl v průměrných hodnotách LTV mezi oběma skupinami, tj. $\mu_1 \neq \mu_2$, resp. $\mu_1 - \mu_2 \neq 0$.

Jak bylo řečeno již v 7. kapitole, v praxi chceme většinou ukázat, že na základě dat je možné nulovou hypotézu zamítnout.

Obecný postup

Jako obvykle předpokládáme, že nulová hypotéza je pravdivá, a za tohoto předpokladu spočteme pravděpodobnost, s jakou dostaneme náš výsledek, nebo ještě extrémnější hodnotu. Abychom vypočítali tuto pravděpodobnost, potřebujeme vědět něco o rozdělení rozdílu průměrů obou výběrů. Jinými slovy potřebujeme vědět, co by se stalo, kdybychom opakovaně provedli náhodné výběry z obou skupin (nikdy to nedělejte v praxi) a spočítali pro každou dvojici výběrů rozdíl průměrů. Jak by vypadalo rozdělení rozdílu výběrových průměrů? Předpokládejme teď pro každou skupinu zvlášť, že výběrové průměry mají normální rozdělení. To je splněno v případě, kdy základní rozdělení skupiny bude dostatečně podobné normálnímu. Potom teorie říká, že testová statistika

$$T = \frac{\text{rozdíl výběrových průměrů} - \text{očekávaný rozdíl za platnosti } H_0}{\text{odhad směrodatné chyby rozdílu výběrových průměrů}}$$

má Studentovo t rozdělení o $n_1 + n_2 - 2$ stupních volnosti. Tento postup se nazývá **dvouvýběrový t test**.

Směrodatná chyba rozdílu výběrových průměrů je směrodatná odchylka rozdělení rozdílu výběrových průměrů, který označíme $\bar{d} = \bar{x}_1 - \bar{x}_2$, a počítá se následujícím způsobem:

$$\sigma_d = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad (8.1)$$

kde n_1 a n_2 jsou rozsahy výběrů a σ je směrodatná odchylka obou skupin, kterou odhadneme z našich výběrů. Budeme přitom předpokládat, že směrodatná odchylka je v obou skupinách, z nichž jsme výběry provedli, stejná.

Tuto směrodatnou odchylku σ potřebujeme nyní odhadnout. K dispozici již máme dva odhady, s_1 a s_2 . Tyto dva odhady nyní spojíme do jednoho. Často se mu říká *sdužený odhad směrodatné odchylky*. Nejprve vypočteme *sdužený odhad rozptylu*

$$s_d^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \quad (8.2)$$

kde s_1^2 a s_2^2 jsou výběrové rozptyly pro jednotlivé skupiny. Sdužený odhad směrodatné odchylky je potom odmocnina z s_d^2 . Směrodatnou chybu rozdílu výběrových průměrů σ_d tedy odhadneme výrazem

$$s_d = s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \quad (8.3)$$

Testovou statistiku nyní můžeme vyjádřit následovně:

$$T = \frac{\bar{d} - \mu}{s_d}. \quad (8.4)$$

Za platnosti H_0 je v tomto příkladu očekávaným rozdílem nula - naše nulová hypotéza tvrdila, že neexistuje žádný rozdíl mezi dvěma skupinovými průměry. Někdy ale můžeme chtít testovat jistý konkrétní rozdíl, který nemusí být nutně nulový. V takovém případě jednoduše nahradíme nulu v (8.4) touto hodnotou.

Nyní jsme již dostatečně teoreticky vyzbrojeni, abychom mohli rozhodnout o platnosti nulové hypotézy. Z kapitoly 7 víme, že existují dvě metody, jak to lze provést.

Nemáme-li k dispozici počítač, zvolíme předem hladinu významnosti, např. $\alpha = 0,05$, a v tabulce A.3 nalezneme příslušnou kritickou hodnotu jako kvantil $t_{1-\frac{\alpha}{2}}(df)$. K tomu potřebujeme znát ještě počet stupňů volnosti, který je v tomto případě $df = 26 + 15 - 2 = 39$ (místo $df = 39$ můžeme použít $df = 40$ jako přibližnou hodnotu).

Z tabulky A.3 zjistíme, že hledaná kritická hodnota je 2,021. To znamená, že vyjde-li nám hodnota testové statistiky T mezi $-2,021$ a $2,021$, pak nezamítneme (přijmeme) nulovou hypotézu, a vyjde-li menší nebo rovna $-2,021$, či větší nebo rovna $2,021$, pak nulovou hypotézu zamítneme.

Pro naše data je sdužený odhad rozptylu σ^2

$$s^2 = \frac{(26 - 1) \cdot 5,82^2 + (15 - 1) \cdot 7,94^2}{26 + 15 - 2} = 44,344.$$

Odhad směrodatné odchylky σ

$$s = \sqrt{44,344} = 6,659$$

je vždy někde mezi oběma výběrovými směrodatnými odchylkami s_1 a s_2 .

Odhad směrodatné chyby výběrového průměru vyjde

$$s_d = 6,659 \sqrt{\frac{1}{26} + \frac{1}{15}} = 2,159,$$

takže hodnota testové statistiky T je

$$T = \frac{(13,7 - 23,7) - 0}{2,159} = -4,632.$$

Protože $-4,632$ leží mimo interval $(-2,021; 2,021)$, zamítneme nulovou hypotézu.

Kdybychom zpracovávali data na počítači, mohli bychom přesně spočítat dosaženou hladinu významnosti (pravděpodobnost, s jakou dostaneme naši, nebo ještě extrémnější hodnotu testové statistiky, je-li nulová hypotéza pravdivá). V tomto případě bychom dostali hodnotu $p < 0,001$. Výsledky t testu se často uvádějí ve tvaru $(t = -4,632; df = 39; p < 0,001)$. \diamond

V předchozím příkladu jsme na základě testu nulovou hypotézu zamítli. Pokud ovšem na základě testu nulovou hypotézu nezamítneme, neznamená to ještě, že opravdu platí. Musíme se zeptat na sílu testu, jinými slovy, jak velkou možnost zamítnutí nulové hypotézy jsme připustili, pokud existuje významný rozdíl (v lékařském smyslu) mezi oběma skupinovými průměry.

Předpoklady pro použití předchozího postupu:

- **Nezávislost výběrů**

- nejsou-li výběry na sobě nezávislé, tj. oba např. obsahují měření od stejného člověka, pak uvedený postup nemůžeme použít!

- **Normální rozdělení**

- rozdělení sledované náhodné veličiny v obou skupinách by nemělo být příliš odlišné od normálního, abychom mohli říci, že i rozdělení průměrů je normální.

- **Prosté náhodné výběry**

- pokud by pozorování v některém výběru nebyla nezávislá, pak by mohlo dojít k neoprávněnému zmenšení odhadu rozptylu s^2 . Mohl by tedy vyjít významný rozdíl i v případě, kdy léčba nebyla dostatečně účinná, nebo naopak nevýznamný rozdíl, přestože léčba účinná byla.

● Shodné rozptyly ve skupinách

- pokud se rozptyly (resp. směrodatné odchylky) v obou skupinách výrazně liší, můžeme použít modifikaci dvouvýběrového t testu. V tomto případě dostaneme odlišný počet stupňů volnosti - nemusí se dokonce jednat ani o celé číslo. Shodnost rozptylů ve skupinách lze také testovat pomocí statistických testů.

Z předchozího výkladu jsme se dozvěděli, že lze dokázat, že průměrné hodnoty LTV jsou pro obě skupiny různé. Jinými slovy, co se týče průměrných hodnot LTV v obou skupinách, existuje statisticky významný rozdíl. Chtěli bychom tento rozdíl odhadnout. Abychom to mohli provést, vytvoříme interval spolehlivosti pro rozdíl mezi dvěma průměry.

8.1.2 Interval spolehlivosti pro rozdíl mezi dvěma průměry

Krajní body intervalu spolehlivosti se vypočítají pomocí rozdílu výběrových průměrů jako

$$\bar{d} \pm t s_{\bar{d}}, \quad (8.5)$$

kde za t dosadíme příslušnou kritickou hodnotu, tj. v tomto případě kvantil $t_{1-\frac{\alpha}{2}}(df)$ Studentova t rozdělení o

$df = n_1 + n_2 - 2$ stupňů volnosti, který nalezneme v tabulce [A.3](#). Chceme-li 95% interval spolehlivosti, použijeme hladinu $\alpha = 5\%$.

Příklad 8.2 V příkladu [8.1](#) je rozdíl výběrových průměrů $\bar{d} = 13,7 - 23,7$ ("téměř ztracené" - "normální") = $-10,0$, odhad směrodatné chyby $s_{\bar{d}} = 2,159$ a kvantil $t_{1-\frac{\alpha}{2}}(df)$ pro $\alpha = 0,05$ a $df = 39$ je 2,021. Interval spolehlivosti je tedy $-10,0 \pm 2,021 \cdot 2,159 \rightsquigarrow (-14,36; -5,64)$.

To znamená, že skutečný rozdíl v LTV mezi oběma skupinami leží někde mezi 5,64 až 14,36 údery za minutu. Vidíme, že interval spolehlivosti nezahrnuje nulu (neboli žádný rozdíl), což opět svědčí pro zamítnutí nulové hypotézy. \diamond

Příklad 8.3 Ve studii respiračních příznaků u školních dětí chceme vědět, zda děti, u kterých rodiče uvedli, že mají respirační příznaky, měly horší funkci plic než děti, u nichž nebyly příznaky uváděny. V případě 92 dětí bylo uvedeno, že mají kašel a průměrná hodnota jejich maximální expirační průtokové frekvence (PEFR) byla 294,8 l/min se směrodatnou odchylkou 57,1 l/min. Počet dětí, u kterých nebyly příznaky hlášeny, byl 1643 a jejich průměrná hodnota maximální expirační průtokové frekvence byla 316,6 l/min se směrodatnou odchylkou 55,2 l/min.

Máme tedy dva rozsáhlé výběry a můžeme použít předpoklad o normálním rozdělení. Zapišeme popisné statistiky do tabulky [8.3](#).

Tabulka 8.3: Souhrnné statistiky

Děti	n	\bar{x}	s
S kašlem	92	294,8	57,1
Normální	1643	313,6	55,2

Rozdíl průměrů pak je

$$\bar{d} = \bar{x}_1 - \bar{x}_2 = 294,8 - 313,6 = -18,8.$$

Je tento rozdíl $-18,8$ významný? Pokud by mezi průměry nebyl vůbec žádný rozdíl, dalo by se očekávat, že \bar{d} se nebude příliš lišit od nuly. Jinak řečeno, pokud neexistuje významný rozdíl, lze očekávat, že interval spolehlivosti pro rozdíl středních hodnot bude zahrnovat nulu. Pokud naopak statisticky významný rozdíl existuje, potom interval spolehlivosti pro rozdíly nulu zahrnovat nebude.

Směrodatnou odchylku pro rozdíl výběrových průměrů vypočteme podle vzorců (8.2) a (8.3) následovně:

$$s_d = \sqrt{\frac{(92-1)57,1^2 + (1643-1)55,2^2}{92+1643-2}} \sqrt{\frac{1}{92} + \frac{1}{1643}} = 5,92.$$

Počet stupňů volnosti je v tomto případě tak velký (1733), že t rozdělení prakticky splývá s normálním. Kritickou hodnotou pro $\alpha = 0,05$ je tedy kvantil $z_{1-\frac{\alpha}{2}} = 1,960$ - viz tabulku A.1 nebo poslední řádek tabulky A.3 ($df = \infty$).

95% interval spolehlivosti má tedy krajní body

$$-18,8 \pm 1,960 \cdot 5,92 \rightsquigarrow (-30,4; -7,2).$$

Vzhledem k tomu, že interval spolehlivosti neobsahuje nulu, dostáváme dobrý důkaz, že mezi těmito dvěma skupinami dětí je významný rozdíl v průměrné funkci plic. \diamond

V následujícím příkladu si ukážeme situaci, kdy interval spolehlivosti obsahuje nulu.

Příklad 8.4 Při průzkumu, který se týkal hmotností dětí daného školního věku, byla porovnávána skupina 70 dětí - jedináčků se skupinou 72 dětí z rodin s pěti a více dětmi, aby se zjistilo, zda je hmotnost v první skupině větší než ve druhé. V první skupině byla průměrná hmotnost 37 kg se směrodatnou odchylkou 4 kg. Ve druhé skupině byla průměrná hmotnost 36 kg se směrodatnou odchylkou 4,16 kg.

Souhrnné statistiky jsou v tabulce 8.4.

Tabulka 8.4: Souhrnné statistiky

Skupina	n	\bar{x}	s
Jedináčci	70	37	4,00

Z více dětí	72	36	4,16
-------------	----	----	------

Analogicky jako v předchozích příkladech spočítáme

$$\bar{d} = 1,$$

$$s_d = 0,685.$$

Za kritickou hodnotu můžeme opět vzít 1,960. 95% interval spolehlivosti má tedy krajní body

$$1 \pm 1,960 \cdot 0,685 \rightsquigarrow (-0,34; 2,34).$$

Protože interval spolehlivosti obsahuje nulu, závěr zní, že mezi skupinami není rozdíl v průměrné hmotnosti statisticky významný (na hladině 5 %). \diamond

Interval spolehlivosti dovoluje posoudit klinickou nebo praktickou významnost našeho nálezu tak, že nám dá představu o velikosti rozdílu průměrů v jednotlivých skupinách. Statistická významnost ale není všechno. Vždy je nutná porada s odborníkem, abychom určili, zda je zjištěný rozdíl důležitý i klinicky.

Poznámka: Máme-li porovnat více než dvě skupiny, musíme být opatrní. Řekněme například, že máme 5 skupin. Nemůžeme se slepě pustit do testování všech 10 možných porovnání. (Z 5 skupin lze vytvořit 10 různých dvojic.) V takové situaci se používá postup zvaný [analýza rozptylu](#) a testovací metody vícenásobného porovnávání.

8.1.3 Porovnání populačních pravděpodobností

Mnoho studií se týká porovnání pravděpodobností výskytu sledovaného jevu ve dvou různých populacích na základě dvou nezávislých výběrů. V podstatě se jedná o stejnou úlohu jako při srovnávání dvou průměrů, kterou jsme již rozebírali, ale je nutné provést některé technické změny. Statistická teorie (konkrétně centrální limitní věta) říká, že když vezmeme výběry o dostatečně velkých rozsazích n_1 a n_2 ze dvou různých populací a sledujeme relativní četnosti p_1 a p_2 dané vlastnosti, potom budou rozdělení těchto relativních četností následující:

$$p_1 \sim N\left(\pi_1, \frac{\pi_1(1-\pi_1)}{n_1}\right), \quad (8.6)$$

$$p_2 \sim N\left(\pi_2, \frac{\pi_2(1-\pi_2)}{n_2}\right), \quad (8.7)$$

kde π_1 a π_2 jsou populační pravděpodobnosti výskytu sledovaného jevu.

$\pi_1 \quad \pi_2$

Nejprve popíšeme statistický test vhodný pro testování nulové hypotézy o shodnosti populačních pravděpodobností (procent) na základě dvou nezávislých náhodných výběrů většího rozsahu. Řekněme, že v prvním výběru nastal sledovaný jev v r_1 případech a ve druhém výběru v r_2 případech. Vypočteme společný odhad relativní četnosti z obou výběrů

$$p = \frac{r_1 + r_2}{n_1 + n_2}. \quad (8.8)$$

Testové kritérium má tvar

$$Z = \frac{p_1 - p_2}{\sqrt{p(1-p)}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}. \quad (8.9)$$

Pro oboustrannou alternativu: "populační pravděpodobnosti se liší" je kritickou hodnotou kvantil $z_{1-\frac{\alpha}{2}}$, který najdeme v tabulce [A.1](#). Tento test používáme, platí-li podmínka $n_i p_i (1 - p_i) > 9$ pro oba výběry.

Příklad 8.5 Z dat populační studie (viz tabulku [8.5](#)) ověřte na 5% hladině významnosti, zda procento osob s antigenem HLA-DR4 je stejné u nemocných progresivní polyartritidou jako u zdravých kontrol.

Tabulka 8.5: Výskyt antigenu HLA-DR4 u nemocných progresivní polyartritidou a u zdravých kontrol

Skupina	HLA-DR4 pozitivní	HLA-DR4 negativní	Celkem
Progresivní polyartritida	46	28	74
Zdravé kontroly	50	184	234
Celkem	96	212	308

Řešení: Testujeme nulovou hypotézu H_0 : "Pravděpodobnost výskytu HLA-DR4 pozitivních je v obou skupinách stejná," proti alternativní hypotéze H_1 : "Pravděpodobnost výskytu HLA-DR4 pozitivních se ve skupinách liší." Z dat uvedených v tabulce [8.5](#) vypočteme relativní četnosti HLA-DR4 pozitivních osob $p_1 = 46/74 = 0,6216$, tj. 62,16 %, $p_2 = 50/234 = 0,2137$, tj. 21,37 %, a $p = 96/308 = 0,3117$, tj. 31,17 %. Vzhledem ke splnění podmínek použitelnosti testu: $n_1 p_1 (1 - p_1) = 17,4 > 9$ a $n_2 p_2 (1 - p_2) = 39,3 > 9$, vypočteme testové kritérium $Z = 6,604$. Toto číslo překračuje kritickou hodnotu 1,960, což vede k závěru, že pravděpodobnost výskytu HLA-DR4 pozitivních osob se ve skupinách významně liší na 5% hladině. \diamond

Také nás může zajímat výběrový rozdíl $p_1 - p_2$. Podobně jako u průměrů, předpokládáme nezávislost obou výběrů.

Rozdíl relativních četností $p_1 - p_2$ se bude chovat jako normální náhodná veličina s následujícími parametry:

$$(p_1 - p_2) \sim N \left(\pi_1 - \pi_2, \frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2} \right). \quad (8.10)$$

Směrodatnou odchylku rozdílu $p_1 - p_2$ odhadneme výrazem

$$s_{(p_1 - p_2)} = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}. \quad (8.11)$$

Krajní body 95% intervalu spolehlivosti pro rozdíl mezi populačními pravděpodobnostmi $\pi_1 - \pi_2$ jsou tedy

$$(p_1 - p_2) \pm 1,96s_{(p_1 - p_2)}. \quad (8.12)$$

Příklad 8.6 V přehledu důkazů, které se týkaly terapeutické hodnoty kyseliny askorbové (vitamin C) pro léčbu běžného nachlazení, popsal Pauling francouzskou studii z roku 1961, která zahrnovala 279 lyžařů ve dvou obdobích po 5-7 dnech. Jednalo se o tzv. dvojité slepý pokus. Jedna skupina (139 subjektů) dostávala 1 gram kyseliny askorbové denně, zatímco druhá skupina (140) dostávala placebo. V této studii nás zajímá relativní výskyt nachlazení v těchto dvou skupinách. Tabulka 8.6 obsahuje Paulingova data.

Tabulka 8.6: Paulingova data

Léčba	Nastydli	Nenastydli	Celkem
Vitamin C	17	122	139
Placebo	31	109	140

Relativní četnost nastydnutí byla tedy ve skupině lyžařů užívajících vitamin C $p_1 = 17/139 = 0,12$ a ve skupině lyžařů dostávajících placebo $p_2 = 31/140 = 0,22$. Samozřejmě nás zajímá, co by se stalo, kdybychom tyto dva lékařské postupy aplikovali na rozsáhlé počty (populaci) lidí, tj. jaký by byl rozdíl populačních pravděpodobností $\pi_1 - \pi_2$?

Řešení: Abychom mohli zodpovědět tuto otázku, musíme vytvořit 95% interval spolehlivosti pro $\pi_1 - \pi_2$. Použitím vzorce (8.11) dostaneme

$$s_{(p_1 - p_2)} = \sqrt{\frac{0,12(1 - 0,12)}{139} + \frac{0,22(1 - 0,22)}{140}} = 0,045,$$

takže krajní body 95% intervalu spolehlivosti jsou podle (8.12)

$$(0,12 - 0,22) \pm 1,96 \cdot 0,045 = -0,10 \pm 0,09 \rightsquigarrow (-0,19; -0,01).$$

Tento interval říká, že rozdíl populačních pravděpodobností je mezi 1 % a 19 %. Budeme-li ignorovat možný účinek placebo, potom výsledek naznačuje, že by takové dávky vitamínu C snižovaly výskyt nachlazení v podobných populacích (tj. v populacích podobných francouzským lyžařům!) o 1 % až 19 %. Tento příklad je převzat z [5], kde se uvádí, že lékařská hodnota studie byla zpochybněna z hlediska návrhu (konstrukce) studie. \diamond

Příklad 8.7 Během epidemie proběhl proces očkování proti chřipce. Celkem 20 z 240 osob ($p_1 = 0,083$ neboli 8,3 %), které dostaly skutečnou očkovací látku, onemocnělo chřipkou ve srovnání s 80 osobami z 220 ($p_2 = 0,364$ neboli 36,4 %), které dostaly placebo. Základní statistiky jsou shrnuty v tabulce 8.7. Je toto dostatečný důkaz, že byla očkovací látka účinná?

Tabulka 8.7: Základní statistiky

Očkování	n	p
Vakcína	240	0,083
Placebo	220	0,364

Řešení: Vypočteme opět 95% interval spolehlivosti pro $\pi_1 - \pi_2$. Použitím vzorce (8.11) dostaneme

$$s_{(p_1 - p_2)} = \sqrt{\frac{0,083(1 - 0,083)}{240} + \frac{0,364(1 - 0,364)}{220}} = 0,037,$$

takže krajní body 95% intervalu spolehlivosti jsou podle (8.12)

$$(0,083 - 0,364) \pm 1,96 \cdot 0,037 = -0,281 \pm 0,073 \rightsquigarrow (-0,354; -0,208).$$

Na základě tohoto zjištění můžeme učinit závěr, že pravděpodobnost onemocnění chřipkou se snižuje po očkování očkovací látkou o 20,8 % až 35,4 % ve srovnání s aplikací placebo. \diamond

8.2 Párová pozorování

Studie obsahující párovaná data se mohou zabývat pacienty, kteří jsou párováni tak, že se co nejvíce podobají s ohledem na nějakou charakteristiku, o které se soudí, že může mít vliv na výsledek studie. V klinickém procesu je jeden člen z každé dvojice náhodně zařazen do jedné ze dvou léčených skupin (z nichž jedna může být "placebo" neboli kontrolní). Podobně by párování mohlo být tzv. sebpárováním, které by se týkalo skupiny pacientů sledovaných ve dvou časových úsecích, mezi nimiž by proběhla nějaká léčba. Zde se budeme zabývat právě tímto druhým případem, ale metoda statistické analýzy je stejná jako pro prvně jmenovaný případ.

8.2.1 Rozdíl průměrů

Příklad 8.8 Začneme Coltonovým příkladem párového porovnávání. Ten ve své studii hovoří o účincích hydrochlorothiazidu na krevní tlak v prostém náhodném výběru 11 hypertoniků. Každý pacient dostal nejprve placebo a o měsíc později hydrochlorothiazid. Krevní tlak se měřil v obou případech stejným způsobem. Údaje uvedené v tabulce 8.8 představují systolický tlak (v mm Hg).

Tabulka 8.8: Systolický tlak (v mm Hg) u 11 pacientů

Placebo	Hydrochlorothiazid	Rozdíl
211	181	30
210	172	38
210	196	14
203	191	12
196	167	29
190	161	29
191	178	13
177	160	17
173	149	24
170	119	51
163	156	7

Výsledky ukazují, že rozdíly (placebo – hydrochlorothiazid) jsou vždy pozitivní. Je zřejmé, že mezi účinky placeba a hydrochlorothiazidu existuje systematický rozdíl. My si na těchto datech předvedeme, jak lze prokázat, že průměrný rozdíl je "statisticky významný".

Předpokládejme, že rozdíly jsou normálně rozdělené s nějakým průměrem μ a kladným rozptylem σ^2 . To nás nutně vede k závěru, že očekávaný dlouhodobý průměrný rozdíl mezi placebem a hydrochlorothiazidem bude mít ve skupině hypertoniků hodnotu μ - to se vysvětluje jako systematický účinek léku. Je-li toto rozumný předpoklad, potom vezmeme-li průměr z n takových rozdílů (říkajme mu \bar{d}), dojde k tomu, že \bar{d} bude mít průměrnou hodnotu μ a rozptyl σ^2/n (její odmocnině říkáme opět směrodatná chyba průměru).

Problém, zda lék má či nemá nějaký účinek, může být nyní formulován takto: "Jestliže lék nemá žádný účinek, potom očekávejme, že μ se bude rovnat nule."

Toto tvrzení označíme jako "nulovou hypotézu", tj. pokud je tato věta pravdivá, potom lék nemá žádný účinek. Pokud hypotéza neplatí, potom průměrný rozdíl μ může nabývat jakékoli hodnoty - to je naše alternativní hypotéza. Tedy:

$$\begin{aligned} H_0: \mu &= 0, \\ H_1: \mu &\neq 0. \end{aligned}$$

Zopakujme, že μ je průměrem párových rozdílů. Abychom mohli testovat nulovou hypotézu, vypočteme testovou statistiku

$$T = \frac{\text{pozorovaná hodnota} - \text{předpokládaná hodnota}}{\text{odhad směrodatné chyby}} = \frac{\bar{d} - 0}{s_d}.$$

V tomto případě je "pozorovanou hodnotou" jednoduše \bar{d} a "předpokládanou hodnotou" je 0; směrodatná chyba průměru \bar{d} se odhaduje jako $s_d = s/\sqrt{n}$, kde s je směrodatná odchylka ze sloupce rozdílů. Jako vždy provedeme statistický test tak, že spočteme pravděpodobnost, s jakou dostaneme námi vypočtenou hodnotu testové statistiky, či hodnotu ještě extrémnější (tj. méně pravděpodobnou), za předpokladu, že nulová hypotéza platí.

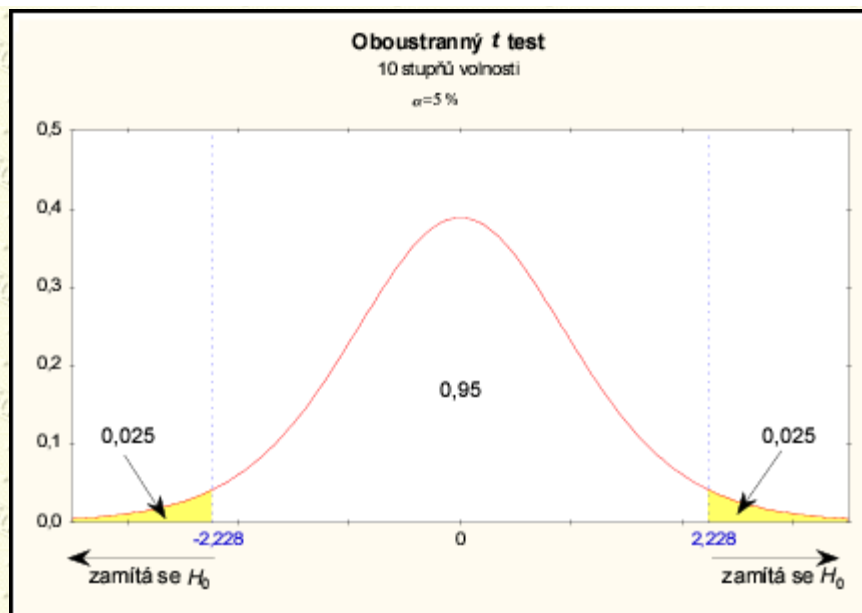
Testová statistika

$$T = \frac{\bar{d} - 0}{\frac{s}{\sqrt{n}}}$$

má za uvedeného předpokladu t rozdělení (neboť n je malé a σ neznámé) o $n - 1$ stupních volnosti (tj. v našem případě 10). Tento postup se proto nazývá **párový t test**.

U Studentova t rozdělení o 10 stupních volnosti leží 95 % plochy pod křivkou mezi $-2,228$ a $2,228$ (viz obrázek [8.3](#) a tabulku [A.3](#)). To znamená, že pokud platí nulová hypotéza, tj. průměrný rozdíl μ je roven nule, pak hodnota testového kritéria bude s 95% pravděpodobností mezi $-2,228$ a $2,228$.

Obrázek 8.3: Oboustranný t test



Pozorovaný rozdíl v měřeních průměrného krevního tlaku u placebo a hydrochlorothiazidu je 24 mm Hg se směrodatnou odchylkou 13,1 mm Hg. Naše testová statistika je tedy

$$T = \frac{24 - \bar{0}}{\frac{13,1}{\sqrt{11}}} = 6,076.$$

Tato hodnota je větší než výše uvedená horní mez 2,228. Při platnosti nulové hypotézy, že $\mu = \bar{0}$, je dosažení tak vysoké hodnoty velmi nepravděpodobné. Proto H_0 zamítneme a vyslovíme závěr, že byl prokázán rozdíl mezi účinky placebo a hydrochlorothiazidu na krevní tlak.

Říkáme, že pozorovaný rozdíl 24 mm Hg je "statisticky významný". \diamond

Testování hypotéz sestává z následujících kroků:

1. Stanovíme nulovou a alternativní hypotézu.
2. Určíme rozdělení pravděpodobnosti testové statistiky při nulové hypotéze. (V předchozím příkladu měla testová statistika T při $H_0: \mu = \bar{0}$ Studentovo t rozdělení o $n - 1$ stupních volnosti.)
3. Zvolíme velikost chvostu neboli hladinu významnosti testu α (je to doplněk koeficientu spolehlivosti P ; pro $P = 0,95$, tj. 95 %, je tedy $\alpha = 1 - 0,95 = 0,05$, tj. 5 %).
4. Na základě zvolené hladiny významnosti vypočteme tzv. kritické hodnoty (příslušného rozdělení pravděpodobnosti testové statistiky za platnosti nulové hypotézy). Tyto kritické hodnoty ohraničují tzv.

kritický obor testu, jenž obsahuje ty hodnoty testové statistiky, které jsou v rozporu s nulovou hypotézou.

5. Vypočítáme hodnotu testové statistiky. Pokud padne do kritického oboru, tj. překročí kritickou hodnotu, zamítneme nulovou hypotézu na hladině významnosti α .
6. V opačném případě učiníme závěr, že na základě zkoumaných dat nemůžeme zamítnout nulovou hypotézu na hladině významnosti α , takže případný pozorovaný rozdíl mohl vzniknout čistě náhodou. To ovšem ještě neznamená, že je nulová hypotéza pravdivá.

Poznámka: Pokud máme k hodnotě testové statistiky k dispozici také dosaženou hladinu významnosti p , potom nulovou hypotézu zamítneme, kdykoli vyjde $p \leq \alpha$. Výpočet kritické hodnoty podle bodu 4 není třeba provádět.

[Další](#) [Výš](#) [Předchozí](#) [Obsah](#) [Rejstřík](#)

Další: [9. Neparametrické metody](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [7. Testování hypotéz](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [10. Analýza kategoriálních dat](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [8. Porovnání dvou skupin](#)

Kapitola

- [9.1 Základní pojmy](#)
- [9.2 Kvantilový, mediánový a znaménkový test](#)
- [9.3 Wilcoxonův párový test](#)
- [9.4 Mannův-Whitneyův test](#)

9. Neparametrické metody

Žádné statistické techniky nevytvoří "dobré" výsledky z dat pochybné kvality.

M. Buyse

9.1 Základní pojmy

Dosud jsme se zabývali statistickými metodami, které zahrnovaly předpoklady o rozdělení dat. Existuje ještě jiná třída statistických metod, které nepředpokládají konkrétní rozdělení - nazývají se neparametrické metody. Nejvíce se užívají k analýze údajů, které nevyhovují požadavkům na rozdělení v parametrických metodách, například párovém a dvouvýběrovém t testu. Zopakujeme, že v případě párového t testu jsme předpokládali, že rozdíly jsou normálně rozdělené. Někdy je jasné, že rozdělení nemůže být normální, už pouhým pohledem na normální graf. Zde nám při rozhodování, zda měření sledují normální rozdělení, opět pomůže odborná znalost oboru. Určování na základě malého rozsahu výběru je zpravidla naprosto nespolehlivé. Druhým důvodem, proč používat neparametrické testy, je fakt, že jsou často jedinou dostupnou metodou pro určité typy údajů, například pro ordinální data, pořadí nebo četnosti. Na druhou stranu mnohé testy, které jsou považované za neparametrické, zahrnují parametry a rozdělení. To je proto, že se tyto metody netýkají rozdělení testových statistik, ale mohou být používány na výběry pocházející z široké třídy rozdělení určené pouze velmi obecnými předpoklady, např. symetrií rozdělení. Nemusí mít nějaké konkrétní rozdělení, jako např. normální.

V posledním desetiletí došlo v oblasti neparametrických metod ke značnému rozvoji, hlavně díky dostupnosti počítačů. Stalo se tak především v oblasti exaktních testů a intervalů spolehlivosti. Podrobný výčet této oblasti viz [\[15\]](#).

V každém případě, slabší předpoklady, které neodmyslitelně patří k neparametrickým testům, způsobují, že tyto metody nejsou tak silné, jako jejich parametrické protějšky. A konečně, až do poslední doby se neparametrické metody zaměřovaly hlavně na testování hypotéz, a kladly malý důraz na odhad.

Zavedeme nyní pojem, který se v neparametrických metodách často vyskytuje. Mějme dána reálná čísla uspořádaná podle velikosti x_1, x_2, \dots, x_n . Jsou-li všechna tato čísla vzájemně různá, je pořadí čísla x_i dáno hodnotou indexu i .

Pořadí R_i udává počet čísel x_1, x_2, \dots, x_n , která jsou menší nebo rovna číslu x_i . Příklad přiřazení pořadí vzájemně různým hodnotám je uveden v tabulce [9.1](#).

Tabulka 9.1: Přiřazení pořadí vzájemně různým hodnotám

Vzestupně uspořádané hodnoty x_i	-2	0	5	7	18
Pořadí R_i	1	2	3	4	5

Někdy se stane, že čísla x_1, x_2, \dots, x_n nejsou různá, ale některá z nich jsou si rovna a vytvářejí tzv. **shody**. Číslům, která tvoří určitou shodu, se pak přiřazuje průměrné pořadí odpovídající takové skupince. Příklad na přiřazení pořadí hodnotám, které nejsou různé, je uveden v tabulce [9.2](#).

Tabulka 9.2: Přiřazení pořadí hodnotám v případě shody

Vzestupně uspořádané hodnoty x_i	-5	-5	0	0	0	10	21	21
Očíslování hodnot x_i	1	2	3	4	5	6	7	8
Pořadí R_i	1,5	1,5	4	4	4	6	7,5	7,5

Uspořádat však lze nejen hodnoty kvantitativního znaku, ale i hodnoty kvalitativního ordinálního znaku. Proto i tam, kde je vyjádření složitějších vlastností členů souboru jedním nebo několika kvantitativními znaky obtížné nebo zcela nemožné, lze často uspořádání provést prostřednictvím vhodného ordinálního znaku, například při posuzování některých sportovních výkonů, školní klasifikace, hodnocení závažnosti onemocnění nebo posuzování soběstačnosti handicapovaných osob.

9.2 Kvantilový, mediánový a znaménkový test

V **kvantilovém testu** vycházíme z nulové hypotézy

$$H_0: x_q = c,$$

že $100q\%$ kvantil základního souboru x_q je roven dané konstantě c . Na základě náhodného výběru rozsahu n

zjistíme počet členů m v náhodném výběru, u nichž hodnota znaku x je menší než konstanta c , tj. $x < c$. Zpravidla se před touto operací všechna pozorování rovná hodnotě c z výběru odstraní a hodnota n se příslušně zmenší.

Statistický test pro ověřování odchylek od nulové hypotézy je založen na veličině

$$Z = \frac{m - nq}{\sqrt{nq(1-q)}}, \quad (9.1)$$

která má při platnosti H_0 standardizované normální rozdělení. V případě oboustranné alternativní hypotézy

$$H_1: x_q \neq c$$

je kritickou hodnotou kvantil $z_{1-\frac{\alpha}{2}}$ standardizovaného normálního rozdělení, který nalezneme v tabulce [A.1](#).

Nulovou hypotézu zamítneme, pokud $|Z| \geq z_{1-\frac{\alpha}{2}}$.

Pro jednostrannou alternativu

$$H_1: x_q > c$$

je kritickou hodnotou kvantil z_α , který opět nalezneme v tabulce [A.1](#). Nulovou hypotézu zamítneme, pokud

$$Z \leq z_\alpha.$$

Pro jednostrannou alternativu

$$H_1: x_q < c$$

je kritickou hodnotou kvantil $z_{1-\alpha}$ a nulovou hypotézu zamítneme, pokud $Z \geq z_{1-\alpha}$.

Popsaný test lze použít, jestliže $n > 30$ a $0,10 < q < 0,90$. Speciálním případem kvantilového testu je **mediánový test** při $q = 0,50$. Dosazením do obecného vzorce testového kritéria pro kvantilový test dostáváme, že mediánový test je založen na veličině

$$Z = \frac{2m - n}{\sqrt{n}}. \quad (9.2)$$

Jestliže sledovaná veličina má spojitě symetrické rozdělení (např. normální rozdělení), je medián zároveň průměrem, takže jde o test hypotézy o průměru náhodné veličiny.

Příklad 9.1 Ve skupině 49 chlapců ve věku 9,5-10 let dispenzarizovaných v roce 1960 po dobu nejméně čtyř let pro jisté onemocnění bylo nalezeno 27 chlapců menších než 138,5 cm, kde 138,5 cm je zjištěný průměr tělesné výšky

v populaci chlapců stejného věku při celostátním šetření. Ověřte na 5% hladině významnosti, zda u nemocných dětí je průměrná výška menší než v odpovídající věkové skupině zdravých dětí.

Řešení: Nulová hypotéza předpokládá, že se dispenzarizovaní chlapci v průměru neliší výškově od stejně starých chlapců v populaci, tj.

$$H_0: \mu_{0,50} = 138,5.$$

Alternativní hypotéza

$$H_1: \mu_{0,50} < 138,5$$

vyjadřuje, že dispenzarizovaní chlapci jsou v průměru menší. Vypočteme hodnotu testové statistiky

$$Z = \frac{2 \cdot 27 - 49}{\sqrt{49}} = \frac{5}{7} = 0,714.$$

Kritická hodnota pro jednostrannou alternativu a hladinu významnosti $\alpha = 0,05$ je kvantil $z_{1-\alpha} = 1,645$ (viz tabulku [A.1](#)), takže na 5% hladině významnosti nelze nulovou hypotézu zamítnout. Naše pozorování statisticky neprokázalo, že onemocnění brzdí růst dětí. \diamond

V případě, že hodnoty sledované veličiny jsou rozdíly párových pozorování, užíváme pro mediánový test název **znaménkový test**.

Příklad 9.2 Při kontrole pyrogenity infuzních roztoků se zjišťovala ve skupině 81 králíků změna tělesné teploty po injekci. U 47 králíků se tělesná teplota zvýšila. Ověřte na 10% hladině významnosti, zda vzestup teploty po injekci je statisticky významný.

Řešení: U každého králíka jsme změřili dvě hodnoty, před injekcí a po injekci, které jsou párově závislé. Reakce na podaný roztok je vyjádřena pomocí difference x obou párových hodnot, ve zjednodušeném případě pouze znaménkem této difference. Kdyby byl injekční roztok zcela nepyrogenní, teplota zvířat by kolísala jen nahodile, takže zhruba v polovině případů by mělo být znaménko difference kladné.

Zvolíme tedy nulovou hypotézu

$$H_0: x_{0,50} = 0, \text{ tj. roztok je nepyrogenní,}$$

kteřou testujeme vůči jednostranné alternativě

$$H_1: x_{0,50} > 0,$$

kteřá vyjadřuje pyrogenitu roztoku pomocí zvyšování teploty. Dosazením do vzorce pro výpočet testové statistiky dostaneme

$$Z = \frac{2 \cdot 34 - 81}{\sqrt{81}} = -\frac{13}{9} = -1,444.$$

Pro hladinu významnosti $\alpha = 0,10$ je $z_\alpha = -1,282$, jak zjistíme z tabulky [A.1](#). Protože $Z = -1,444 < -1,282$, zamítneme nulovou hypotézu a dojdeme k závěru, že na 10% hladině významnosti je vzestup teploty významný a roztok je pyrogenní. Vzhledem k tomu, že pyrogenita je vážnou závadou roztoku, zvolili jsme vyšší, 10% hladinu významnosti, abychom tak snáze zachytili podezřelé roztoky. \diamond

Poznámka: Při popsané kontrolní zkoušce jsme nerespektovali možnost, že vzestup teploty může způsobit i jiná příčina než pyrogenita roztoku, např. stres při vpichu jehly, různé pokusné podmínky při obou měřeních teploty apod. Chceme-li vyloučit jiné příčiny zvýšení teploty, musíme použít standardního nepyrogenního roztoku v kontrolní skupině pokusných zvířat. Potom srovnáváme procento zvýšení teploty v pokusech s nepyrogenním placebem v kontrolní skupině s kontrolovaným přípravkem v pokusné skupině. Při menším počtu pozorování (v praxi většinou pro $n < 20$) hledáme kritické hodnoty znaménkového testu ve speciálních tabulkách. Znaménkový test je příkladem párového testu. V klinickém výzkumu se dá výhodně využít v situaci, kdy záměrně vytváříme páry vzájemně podobných jedinců. V každém takovém homogenním páru náhodně vylosujeme kontrolního jedince. Výsledky zjištěné u kontrolního a pokusného jedince vytvářejí párová pozorování. Jedná-li se navíc o slepý pokus, kdy není experimentátorovi známe, který jedinec je kontrolní, hodí se znaménkový test i k posuzování složitých vlastností, pokud je experimentátor dokáže ve dvojicích vzájemně porovnat.

Nyní představíme některé neparametrické testy, například neparametrickou verzi párového a dvouvýběrového t testu.

9.3 Wilcoxonův párový test

Příklad 9.3 Tabulka [9.3](#) udává počet T_4 a T_8 buněk/ m^3 ve vzorcích krve 10 pacientů v remisi Hodgkinovy nemoci. Zajímá nás srovnání průměrného počtu T_4 a T_8 buněk.

Tabulka: Počty T_4 a T_8 buněk v 1 m^3 krve u 10 pacientů

Počet T_4 buněk	Počet T_8 buněk
396	236
568	786
1212	311
171	449
554	811

1104	686
257	412
435	286
295	336
397	936

Řešení: Z parametrických metod bychom použili párový t test, neboť oba znaky zjišťujeme na stejných pacientech, takže se nejedná o dva nezávislé výběry. Jedním z předpokladů tohoto testu je, že rozdíly (počet T_8 buněk – počet T_4 buněk) u jednotlivých pacientů jsou normálně rozdělené. Pokud nechceme učinit takový předpoklad, mohli bychom použít neparаметrický test zvaný *Wilcoxonův párový test*.

V párovém t testu se nulová a alternativní hypotéza vztahovaly k průměru rozdílů. U Wilcoxonova párového testu se hypotézy týkají mediánu rozdílů:

H_0 : Medián rozdílů je nulový.

H_1 : Medián rozdílů je různý od nuly.

Abychom mohli provést náš statistický test, musíme nejprve vypočítat rozdíly pro každou osobu a pořadí absolutních hodnot rozdílů. Potom připojíme znaménko rozdílu k pořadí. Výsledky ukazuje tabulka 9.4.

Tabulka 9.4: Rozdíly a pořadí rozdílů

Počet T_4 buněk	Počet T_8 buněk	Rozdíl	Pořadí
396	236	-160	-4
568	786	218	5
1212	311	-901	-10
171	449	278	7
554	811	257	6
1104	686	-418	-8
257	412	155	3
435	286	-149	-2
295	336	41	1
397	936	539	9

Pokud je nulová hypotéza pravdivá, tj. mediány se shodují, pak součet záporných pořadí by měl být teoreticky roven součtu kladných pořadí. Pokud je pravdivá alternativní hypotéza, pak toto neplatí.

V našem příkladu je součet kladných pořadí ve výběru $5 + 7 + 6 + 3 + 1 + 9 = 31$ a součet záporných pořadí je $4 + 10 + 8 + 2 = 24$.

Součty záporných a kladných hodnot jsou vzájemně svázané, takže stačí vzít v úvahu jen jeden z nich. Pro naše účely uvažujeme pouze nižší z obou hodnot, v tomto případě 24. Nazvěme ji S .

Nyní určíme hladinu významnosti $\alpha = 5\%$ (oboustranný test) a v tabulce [A.7](#) zjistíme příslušnou kritickou hodnotu.

Pro $n = 10$ je daná kritická hodnota 8. Pokud je S menší než tato hodnota, můžeme zamítnout nulovou hypotézu.

Pro náš příklad je $S = 24 > 8$, a proto nemáme dostatečný důkaz pro zamítnutí nulové hypotézy. Na základě zkoumaných dat nelze učinit závěr, že u pacientů s Hodgkinovou nemocí je medián skupiny T_4 buněk odlišný od mediánu skupiny T_8 buněk. U větších výběrů může být použita normální aproximace testové statistiky S . \diamond

9.4 Mannův-Whitneyův test

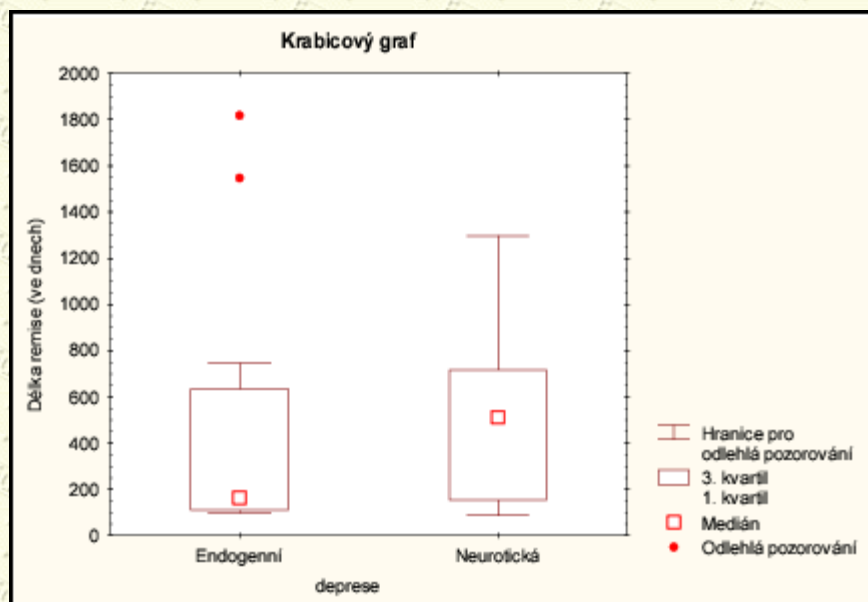
Příklad 9.4 Údaje v tabulce [9.5](#) představují délku remise ve dnech z prostého náhodného výběru ze dvou různých skupin pacientů - s endogenní depresí a s neurotickou depresí. Opět předpokládáme, že pozorování v rámci skupiny jsou nezávislá - každý pacient představuje jednu hodnotu. Zajímá nás, zda existuje rozdíl v těchto dvou rozděleních.

Tabulka 9.5: Délka remise ve dnech u pacientů s depresí

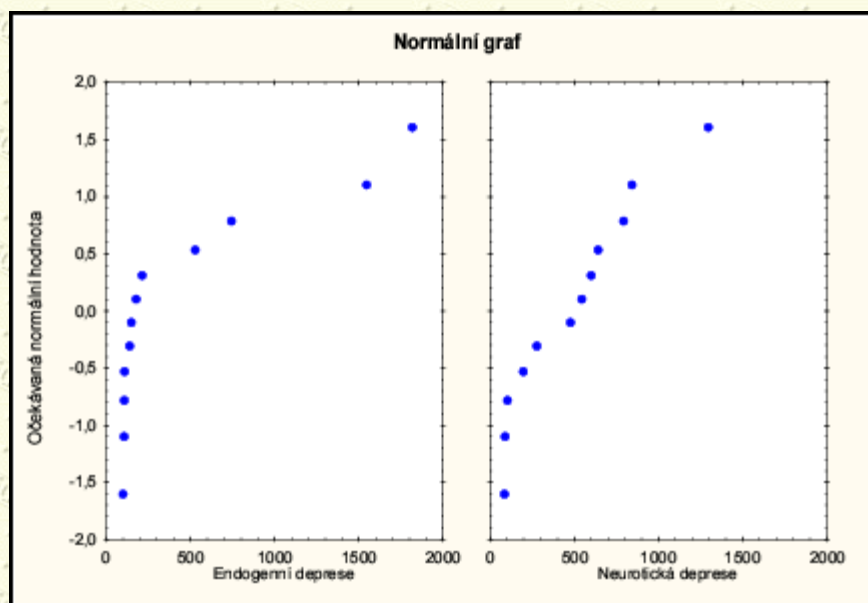
Pacienti s endogenní depresí	Pacienti s neurotickou depresí
109	546
214	844
1818	602
140	87
179	794
744	643
108	199
101	91
107	105
1547	479
529	1296
140	279

Řešení: Obrázek 9.1 ukazuje krabicový graf porovnávající délku remise (ve dnech) pro obě skupiny.

Obrázek 9.1: Délka remise pro obě skupiny



Obrázek 9.2: Normální graf pro obě skupiny



Obrázek 9.2 ukazuje normální graf (normal probability plot) pro obě skupiny. Je z něj vidět, že body na normálních grafech neleží na přímce, ale tvoří spíše silně zahnutou křivku. Předpoklad normálního rozdělení pro délku remise je proto těžko udržitelný. K otestování rozdílů ve skupinových průměrech proto nepoužijeme dvouvýběrový t test, ale raději neparametrický test zvaný *Mannův-Whitneyův* nebo také *Wilcoxonův dvouvýběrový test*. Nulová a alternativní hypotéza jsou vyjádřeny následujícím způsobem:

H_0 : Rozdělení obou skupin je shodné.

H_1 : Rozdělení obou skupin se liší.

Test pokračuje kombinováním obou výběrů, seřazením pozorování od nízkých hodnot po vysoké a potom stanovením pořadí pro každé pozorování. Výsledky jsou uvedeny v tabulce 9.6.

Tabulka 9.6: Pořadí pacientů s depresí

Endogenní deprese	Celkové pořadí	Neurotická deprese	Celkové pořadí
109	(7)	546	(16)
214	(12)	844	(21)
1818	(24)	602	(17)
140	(8)	87	(1)
179	(10)	794	(20)
744	(19)	643	(18)
108	(6)	199	(11)
101	(3)	91	(2)
107	(5)	105	(4)
1547	(23)	479	(14)
529	(15)	1296	(22)
150	(9)	279	(13)

Dále spočteme součet pořadí pro 12 pacientů s endogenní depresí a pro 12 pacientů s neurotickou depresí.

Endogenní deprese: součet pořadí $S_1 = 141$, rozsah výběru $n_1 = 12$.

Neurotická deprese: součet pořadí $S_2 = 159$, rozsah výběru $n_2 = 12$.

Nyní spočteme testové statistiky U_1 a U_2 , kde

$$U_1 = S_1 - \frac{n_1(n_1 + 1)}{2} = 141 - \frac{12 \cdot 13}{2} = 63, \quad (9.3)$$

$$U_2 = S_2 - \frac{n_2(n_2 + 1)}{2} = 159 - \frac{12 \cdot 13}{2} = 81. \quad (9.4)$$

Mezi těmito dvěma statistikami opět existuje vzájemný vztah. Stačí nám proto spočítat pouze jednu z nich, např. U_1 , a druhou vypočteme ze vzorce

$$U_2 = n_1 n_2 - U_1 = 144 - 63 = 81, \quad (9.5)$$

což jsme dostali i výše.

Nyní zvolíme hladinu významnosti $\alpha = 5\%$ a v tabulce [A.8](#) nalezneme kritickou hodnotu pro rozsahy výběrů n_1 a n_2 . Nulovou hypotézu zamítneme, pokud menší z čísel U_1 a U_2 je menší než kritická hodnota.

Pro náš příklad ($n_1 = 12$, $n_2 = 12$, $\alpha = 0,05$) je nalezená kritická hodnota 37 (viz tabulku [A.8](#)). Menší z čísel U_1 a U_2 je 63. Protože $63 > 37$, nemůžeme zamítnout nulovou hypotézu. Nemáme tedy žádný důvod pro tvrzení, že mezi střední délkou remise v těchto dvou skupinách je nějaký rozdíl.

Tabulky [A.8](#) a [A.9](#) obsahují kritické hodnoty pro 5% a 1% hladinu významnosti pro rozsah většího výběru $n_1 \leq 30$ a rozsah menšího výběru $n_2 \leq 20$. Pro větší hodnoty n_1 a n_2 můžeme použít normální aproximaci

$$Z = \frac{U_1 - \frac{1}{2}n_1 n_2}{\sqrt{\frac{1}{12}n_1 n_2(n_1 + n_2 + 1)}}. \quad (9.6)$$

Pro zvolenou hladinu významnosti α je kritickou hodnotou kvantil $z_{1-\frac{\alpha}{2}}$ standardizovaného normálního rozdělení, který nalezneme v tabulce [A.1](#). \diamond

Uvedené dva testy jsou příklady neparametrických testů. Podrobnější informace o výběrech větších rozsahů než 20 (použití normální aproximace) a pojednání o shodách lze nalézt například v [\[15\]](#). Mnohé z novějších statistických programů obsahují i část s neparametrickými statistickými metodami, kterými se v tomto dílu sbírky podrobněji nezabýváme.

Závěrem lze říci, že neparametrické testy se používají na data, u nichž lze snadno určit uspořádané hodnoty, pořadí nebo četnosti. Používají se také k analýze dat, která neodpovídají rozdělení předpokládanému v parametrických metodách, jako je například párový nebo dvouvýběrový t test. Ačkoli nepředpokládají normální rozdělení, nejsou úplně bez předpokladů. A konečně, některé neparametrické metody vyžadují pouze minimální informaci. Předpokládejme, že chceme testovat, mají-li váhy předmětů v jistém vzorku určitý předem daný medián M . Z provozních důvodů přitom nelze získat přesná měření, ale je možné určit, zda jsou předměty těžší, či lehčí než M . U tohoto typu údajů můžeme s výhodou použít právě neparametrický mediánový test.

[Další](#) [Výš](#) [Předchozí](#) [Obsah](#) [Rejstřík](#)

Další: [10. Analýza kategoriálních dat](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [8. Porovnání dvou skupin](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [11. Měření závislosti](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [9. Neparametrické metody](#)

Kapitola

- [10.1 \$\chi^2\$ test dobré shody](#)
- [10.2 Kontingenční tabulky](#)

10. Analýza kategoriálních dat

Explorace dat je obdivuhodná, ale výzkumník by měl vědět, že data prohledává, a nedělá potvrzující experiment.

P. A. Lachenbruch

10.1 χ^2 test dobré shody

Často si klademe otázku, zda pozorovaná data neodporují našim předpokladům o použitém matematickém modelu pro popis dané situace. Například celá řada statistických testů může být aplikována jen za předpokladu, že sledovaná veličina má rozdělení pravděpodobnosti určitého typu.

Příklad 10.1 Sledujeme počet alel *Se* v genotypu 639 potomků heterozygotních vylučovatelů (viz příklad [4.1](#)). Z toho 159 potomků má dvě alely *Se* (homozygotní vylučovatelé), 321 potomků jednu alelu *Se* (heterozygotní vylučovatelé) a 159 potomků nemá žádnou alelu *Se* (nevylučovatelé). Ověřte na 5% hladině významnosti, že uvedená data neodporují hypotéze o stejné pravděpodobnosti výskytu čtyř možných genotypů potomka, přičemž heterozygotní vylučovatel je popsán dvěma možnými genotypy, kdy dominantní gen pochází od otce či od matky.

Řešení: Uvedme pozorované četnosti pro počet alel *Se* v genotypu potomků heterozygotních vylučovatelů a dle předpokládaného modelu dopočítejme očekávané četnosti. Počet alel *Se* nabývá tří ($k = 3$) možných hodnot 0, 1 a 2 (viz tabulku [10.1](#)).

Tabulka 10.1: Pozorované a očekávané četnosti v jednotlivých genotypech

Počet alel <i>Se</i>	2 (<i>Se, Se</i>)	1 (<i>Se, se</i>)	0 (<i>se, se</i>)	Celkem
Pozorované četnosti	159	321	159	639
Očekávané četnosti	159,75	319,5	159,75	639

Očekávané četnosti vypočítáme tak, že vynásobíme celkový počet pozorování $n = 639$ pravděpodobností výskytu počtu alel Se podle předpokládaného modelu dědičnosti alel. Například potomek, který je homozygotní vylučovatel (Se, Se) získá s pravděpodobností 0,5 alelu Se od otce a s pravděpodobností 0,5 od matky. Při nezávislé kombinaci genů se tak homozygotní vylučovatel vyskytne s pravděpodobností $0,5 \cdot 0,5 = 0,25$. Z 639 potomků heterozygotních vylučovatelů je tedy očekávaný počet homozygotních vylučovatelů $639 \cdot 0,25 = 159,75$.

Testová statistika tzv. χ^2 testu se v tomto případě vypočte jako součet tří ($k = 3$) výrazů typu (pozorovaná četnost – očekávaná četnost)²/(očekávaná četnost) pro všechny možné počty alel Se (0, 1, 2) v genotypu potomka heterozygotních vylučovatelů. Obecně je

$$\chi^2 = \sum \frac{(\text{pozorovaná četnost} - \text{očekávaná četnost})^2}{\text{očekávaná četnost}} \quad (10.1)$$

Za předpokladu, že geny se dědí nezávisle s pravděpodobností 0,5 od každého z rodičů (nulová hypotéza H_0), má testová statistika χ^2 rozdělení o $df = k - 1 = 2$ stupních volnosti. V našem případě je tedy

$$\begin{aligned} \chi^2 &= \frac{(159 - 159,75)^2}{159,75} + \frac{(321 - 319,5)^2}{319,5} + \frac{(159 - 159,75)^2}{159,75} = \\ &= 0,01. \end{aligned}$$

Testujeme-li uvedenou nulovou hypotézu H_0 proti alternativě H_1 , že model dědičnosti je jiný, než popisuje nulová hypotéza, porovnáme vypočtenou testovou statistiku s kritickou hodnotou, která je rovna kvantilu $\chi^2_{1-\alpha}(df)$ rozdělení χ^2 o $df = k - 1 = 2$ stupních volnosti. Pro hladinu významnosti $\alpha = 0,05$ je $\chi^2_{1-\alpha}(2) = 5,99$ - viz tabulku [A.5](#). Protože testová statistika $\chi^2 = 0,01$ je menší než kritická hodnota 5,99, nezamítáme nulovou hypotézu. Nenalezli jsme tedy statisticky významnou odchylku od předpokládaného modelu dědičnosti na 5% hladině významnosti. \diamond

Shrňme tedy obecný postup χ^2 testu dobré shody. Test používáme pro náhodné veličiny nominální, ordinální či diskrétní. Počet jejich možných hodnot (kategorií) označme k . Dále lze χ^2 test dobré shody použít i pro spojité veličiny, které kategorizujeme, neboli rozdělíme obor jejich hodnot do k vzájemně se nepřekrývajících intervalů. Symbolem π_i ($i = 1, 2, \dots, k$) označíme pravděpodobnosti, s nimiž náhodná veličina X nabývá hodnot z i -té kategorie. Pro náhodný výběr o rozsahu n spočteme pozorované četnosti n_i pro jednotlivé kategorie, tj. počet hodnot ve výběru, které padly do i -té kategorie, $i = 1, 2, \dots, k$. Očekávané četnosti při daných pravděpodobnostech π_i spočteme jako $n\pi_i$. Anglický statistik K. Pearson ukázal, že výběrová funkce

(10.2)

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - n\pi_i)^2}{n\pi_i}$$

má pro výběry velkých rozsahů přibližně χ^2 rozdělení s $k - 1$ stupni volnosti. Na této skutečnosti je založen test dobré shody (rozumí se shody relativních četností n_i/n a hypotetických pravděpodobností π_i). V případě, že náhodná veličina X má předpokládané pravděpodobnostní rozdělení dané pravděpodobnostmi π_i (nulová hypotéza), bude hodnota testové statistiky χ^2 ve výrazu (10.2) malá. Naopak, bude-li hodnota statistiky χ^2 příliš velká, nebudeme věřit, že hodnoty hypotetických pravděpodobností π_i jsou správné, a nulovou hypotézu zamítneme. Hranici mezi důvěrou a nedůvěrou ve správnost nulové hypotézy určuje tzv. kritická hodnota, která je pro zvolenou hladinu významnosti α rovna kvantilu $\chi^2_{1-\alpha}(df)$ rozdělení χ^2 o $df = k - 1$ stupních volnosti - viz tabulku A.5.

Často používáme tento test pro ověření předpokladu, že rozdělení pravděpodobnosti veličiny X je určitého typu, například normální rozdělení, Poissonovo rozdělení apod. V případě, že nulová hypotéza specifikuje pouze typ pravděpodobnostního rozdělení, ale ne hodnoty parametrů rozdělení, musí být parametry předem z výběru odhadnuty. Na každý takový odhad se počet stupňů volnosti testové statistiky snižuje o jedničku. Ukážeme si tento postup na příkladu ověřování nulové hypotézy o Poissonově rozdělení.

Příklad 10.2 Studujme rozdělení počtu pacientů, kteří přijdou na zubní pohotovost ve všední den. Ordinační dobu rozdělíme do půlhodinových intervalů a v každé půlhodině zjistíme počet pacientů, kteří se během ní na zubní pohotovost dostavili. Z tabulky 10.2 vidíme, že bylo 79 půlhodin, ve kterých se nedostavil žádný pacient, 188 půlhodin, kdy přišel jeden pacient, 282 půlhodin, během kterých přišli dva pacienti atd. Ověřte na 5% hladině významnosti, zda je přijatelný předpoklad o Poissonově rozdělení počtu pacientů.

Tabulka 10.2: Rozdělení počtu pacientů, kteří se během půlhodiny dostavili na stomatologické pracoviště

Číslo kategorie	Počet pacientů	Pozorovaná četnost	Očekávaná četnost
i	x_i	n_i	$n\pi_i$
1	0	79	72,97
2	1	188	204,32
3	2	282	286,05
4	3	275	266,98
5	4	196	186,89
6	5	114	104,66
7	6	45	48,84

8	7	10	19,54
9	8	7	6,84
10	9	3	2,13
11	10	1	0,78
12	11 a více	0	0,00
Celkem	-	1200	1200,00

Řešení: Formulujeme nulovou hypotézu H_0 : "Počet příchodů pacientů během 30 minut má Poissonovo rozdělení," a alternativní hypotézu H_1 : "Počet příchodů pacientů během 30 minut nemá Poissonovo rozdělení."

V případě platnosti nulové hypotézy se může pravděpodobnost příchodu určitého počtu pacientů x vypočítat jako

$$P(x) = e^{-\lambda} \frac{\lambda^x}{x!}.$$

Parametr λ neznáme a odhadneme jej jako vážený výběrový průměr

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \frac{1}{1200} (79 \cdot 0 + 188 \cdot 1 + \dots + 0 \cdot 11) = \frac{3364}{1200} = 2,80.$$

Při hodnotě parametru $\lambda = 2,80$ vypočítáme postupně pravděpodobnosti

$$P(x_1 = 0) = \pi_1, P(x_2 = 1) = \pi_2, \dots, P(x_{12} = 11 \text{ a více}) = \pi_{12}.$$

Očekávané četnosti $n\pi_i$ získáme vynásobením vypočtených pravděpodobností π_i celkovým počtem půlhodin $n = 1200$. Očekávané četnosti jsou uvedeny v tabulce 10.2. Vzhledem k tomu, že krajní očekávané četnosti jsou menší než 5, spojíme okrajové třídy před výpočtem χ^2 testové statistiky (viz tabulku 10.3).

Tabulka 10.3: Rozdělení počtu pacientů, kteří se během půlhodiny dostavili na stomatologické pracoviště po sloučení tříd s malou očekávanou četností

Číslo kategorie	Počet pacientů	Pozorovaná četnost	Očekávaná četnost
i	x_i	n_i	$n\pi_i$
1	0	79	72,97
2	1	188	204,32

3	2	282	286,05
4	3	275	266,98
5	4	196	186,89
6	5	114	104,66
7	6	45	48,84
8	7	10	19,54
9	8 a více	11	9,75
Celkem	-	1200	1200,00

Vypočtená hodnota testové statistiky χ^2 podle vzorce (10.2) je 8,50. Počet stupňů volnosti df spočteme z počtu kategorií k a počtu parametrů m , které jsem odhadli z výběru, jako $df = k - m - 1$. V našem případě je $k = 9$ (po sloučení okrajových kategorií) a $m = 1$. Tedy počet stupňů volnosti $df = 7$. Kritická hodnota pro 7 stupňů volnosti a 5% hladinu významnosti je 14,07 (viz tabulku A.5). Protože testová statistika $\chi^2 = 8,50 < 14,07$, nepokládáme odchylku od hypotézy o Poissonově rozdělení příchodů pacientů na stomatologické oddělení za statisticky významnou na 5% hladině. \diamond

10.2 Kontingenční tabulky

χ^2 test (neparametrická metoda) se používá k zjištění, zda mezi dvěma znaky existuje prokazatelný výrazný vztah.

Znaky mohou být:

- kvalitativní,
- diskrétní kvantitativní,
- spojitě kvantitativní, ale s hodnotami sloučenými do skupin.

Máme-li dva takové znaky, uspořádáme data do **kontingenční tabulky**. Kategorie jednoho znaku určují řádky a kategorie druhého znaku sloupce (viz příklad 10.3). Jednotlivá pozorování jsou zařazena do příslušné buňky kontingenční tabulky podle hodnot daných dvou znaků. Pokud jeden ze znaků má r kategorií a druhý znak má s kategorií, dostáváme kontingenční tabulku typu $r \times s$. Kontingenční tabulku typu 2×2 nazýváme **čtyřpolní tabulka**.

Kontingenční tabulky umožňují testování různých hypotéz. Uvedeme dále tři obvykle testované hypotézy:

1. [hypotézu o shodnosti struktury](#),
2. [hypotézu o nezávislosti](#),
- 3.

[hypotézu o symetrii.](#)

10.2.1 Test hypotézy o shodnosti struktury

V kontingenční tabulce často testujeme nulovou hypotézu o shodnosti struktury jednoho ze sledovaných znaků za různých podmínek, které vyjadřují kategorie druhého znaku. Například nás může zajímat, zda věková struktura hospitalizovaných pacientů je ve dvou nemocnicích stejná. Pokud uvažujeme čtyři věkové kategorie (věk do 25 let, 26–45 let, 46–65 let, 66 let a více), budeme do tabulky zapisovat počty hospitalizovaných pacientů v uvedených věkových kategoriích pro každou z obou nemocnic.

Následující příklad uvádí test hypotézy o shodnosti struktury (test homogenity) pro případ čtyřpolní tabulky.

Příklad 10.3 Vraťme se k údajům z příkladu 8.7. Očkování proti chřipce se zúčastnilo 460 dospělých, z nichž 240 dostalo očkovací látku proti chřipce (očkováná skupina) a 220 dostalo placebo (kontrolní skupina). Na konci experimentu onemocnělo 100 lidí chřipkou. 20 z nich bylo z očkové skupiny a 80 z kontrolní skupiny. Pozorované četnosti jsou v tabulce 10.4.

Tabulka 10.4: Pozorované četnosti v příkladu 10.3

Chřipka	Očkování		Placebo		Celkem	
Ano	20	8,3 %	80	36,4 %	100	21,7 %
Ne	220	91,7 %	140	63,6 %	360	78,3 %
Celkem	240	100,0 %	220	100,0 %	460	100,0 %

Je to dostatečný důkaz, že očkovací látka byla účinná, nebo mohly rozdíly mezi očkovanou a kontrolní skupinou vzniknout náhodou? Zvolme velmi přísnou hladinu významnosti $\alpha = 0,001$.

Řešení: Nulová hypotéza o shodnosti struktury ve čtyřpolní tabulce říká, že procento výskytu chřipky je v očkové a kontrolní skupině stejné.

Dále najdeme pro každé políčko v tabulce tzv. **očekávanou četnost**. To je četnost, kterou bychom měli očekávat, kdyby nulová hypotéza platila.

Při pohledu na naši tabulku vidíme, že ze 460 dospělých, kteří podstoupili experiment dostalo chřipku 100 jedinců (tj. něco přes 20 %). Pokud měla očkovací látka i placebo stejný účinek, potom bychom očekávali, že procento těch, kteří onemocněli chřipkou, bude v obou skupinách stejné (tj. něco přes 20 % z léčené skupiny a něco přes 20 % z kontrolní skupiny). Jinými slovy, neexistují-li žádné rozdíly mezi očkovací látkou a placebem, potom $(100/460) \cdot 240 = 52,17$ dospělých v léčené skupině a $(100/460) \cdot 220 = 47,83$ dospělých v kontrolní skupině by mělo dostat chřipku.

Podobně bychom očekávali, že pokud neexistuje žádný rozdíl mezi očkovací látkou a placebem, pak $(360/460) \cdot 240 = 187,83$ dospělých v léčené skupině a $(360/460) \cdot 220 = 172,17$ dospělých v kontrolní skupině by chřipce uniklo.

Vidíme tedy, že očekávané četnosti můžeme pro každé políčko kontingenční tabulky spočítat podle obecného vztahu

$$\text{očekávaná četnost} = \frac{\text{součet v řádku} \times \text{součet ve sloupci}}{\text{celkový počet pozorování}} \quad (10.3)$$

Proto by naše očekávané četnosti měly vypadat jako v tabulce 10.5 (všimněte si, že součty očekávaných četností v řádcích i sloupcích jsou stejné jako u pozorovaných četností uvedených výše).

Tabulka 10.5: Očekávané četnosti v příkladu 10.3

Chřipka	Očkování		Placebo		Celkem	
Ano	52,17	21,7 %	47,83	21,7 %	100,00	21,7 %
Ne	187,83	78,3 %	172,17	78,3 %	360,00	78,3 %
Celkem	240,00	100,0 %	220,00	100,0 %	460,00	100,0 %

Nyní můžeme porovnat pozorované a očekávané četnosti výskytu. Pokud mezi očkováním a chřipkou neexistuje žádný vztah, potom by se pozorované a očekávané četnosti výskytu měly navzájem blížit a případná odchylka by byla způsobena pouze náhodou. Nejlepší způsob, jak zjistit rozdíly mezi pozorovanými a očekávanými četnostmi výskytu, je spočítat χ^2 statistiku podle obecného vzorce (10.1), kde tentokrát sčítáme přes všechna políčka

v tabulce.

Pro náš příklad vyjde:

$$\begin{aligned} \chi^2 &= \sum \frac{(\text{pozorovaná četnost} - \text{očekávaná četnost})^2}{\text{očekávaná četnost}} = \\ &= \frac{(20 - 52,17)^2}{52,17} + \frac{(80 - 47,83)^2}{47,83} + \frac{(220 - 187,83)^2}{187,83} + \\ &+ \frac{(140 - 172,17)^2}{172,17} = 19,84 + 21,64 + 5,51 + 6,01 = 53,0. \end{aligned}$$

Pro interpretaci této χ^2 statistiky potřebujeme znát ještě počet stupňů volnosti (df). Pro kontingenční tabulky platí:

$$df = (\text{počet řádků} - 1)(\text{počet sloupců} - 1).$$

V našem příkladu jsou dva řádky ($r = 2$) a dva sloupce ($s = 2$), takže

$$df = (2 - 1)(2 - 1) = 1.$$

Vypočtenou testovou statistiku χ^2 porovnáme s kritickou hodnotou, kterou je kvantil $\chi^2_{1-\alpha}(df)$ rozdělení χ^2 o df stupních volnosti. Z tabulky [A.5](#) je vidět, že hodnota 53,0 je větší než 10,83, což je kritická hodnota pro $df = 1$ a hladinu významnosti 0,1 % ($\alpha = 0,001$). To znamená, že pravděpodobnost, že by tak velké pozorované procentuální rozdíly ve výskytu chřipky mohly vzniknout jen náhodou, pokud by neexistovaly žádné skutečné rozdíly mezi očkovací látkou a placebem, je menší než 0,1 %. Proto můžeme učinit závěr, že očkovací látka je účinnější než placebo. \square

Označíme-li pozorované četnosti v obecné čtyřpolní tabulce a, b, c, d (viz tabulku [10.6](#)), přičemž $n = a + b + c + d$ je rozsah výběru, potom testovou statistiku χ^2 testu lze upravit do zjednodušené formy

$$\chi^2 = n \frac{(ad - bc)^2}{(a+b)(a+c)(c+d)(b+d)}. \quad (10.5)$$

Tabulka 10.6:

Obecná čtyřpolní tabulka

a	b	$a + b$
c	d	$c + d$
$a + c$	$b + d$	n

Podle vzorce ([10.5](#)) je pro náš příklad

$$\chi^2 = 460 \cdot \frac{(20 \cdot 140 - 80 \cdot 220)^2}{100 \cdot 240 \cdot 360 \cdot 220} = 53,0.$$

Použití χ^2 testu se neomezuje pouze na nominální, ordinální a diskrétní znaky, ale lze jej použít i na spojité veličiny, které jsou kategorizovány pomocí intervalů.

Příklad 10.4 Kontingenční tabulka [10.7](#) ukazuje výsledky velmi slavného lékařského experimentu ze čtyřicátých let, který se zabýval účinkem streptomycinu při léčbě plicní tuberkulózy. Údaje z radiologického hodnocení po 6

měsících byly porovnány s tím, zda pacient patřil do léčené, nebo kontrolní skupiny. Existuje vztah mezi léčbou a výsledkem?

Tabulka 10.7: Léčba plicní tuberkulózy MRC streptomycinem

Radiologické hodnocení	Streptomycin	Kontrolní	Celkem
Významné zlepšení	28	4	32
Střední / malé zlepšení	10	13	23
Beze změn	2	3	5
Střední / malé zhoršení	5	12	17
Významné zhoršení	6	6	12
Smrt	4	14	18
Celkem	55	52	107

Řešení: Pokud mezi skupinami není žádný rozdíl, dostaneme očekávané četnosti, které jsou uvedeny v závorkách v tabulce 10.8.

Tabulka 10.8: Pozorované a očekávané četnosti

Radiologické hodnocení	Streptomycin	Kontrolní	Celkem
Významné zlepšení	28 (16,45)	4 (15,55)	32
Střední / malé zlepšení	10 (11,82)	13 (11,18)	23
Beze změn	2 (2,57)	3 (2,43)	5
Střední / malé zhoršení	5 (8,74)	12 (8,26)	17
Významné zhoršení	6 (6,17)	6 (5,83)	12
Smrt	4 (9,25)	14 (8,75)	18
Celkem	55 (55,00)	52 (52,00)	107

Dále spočteme testovou statistiku

$$\chi^2 = 26,96, \quad df = (6 - 1)(2 - 1) = 5.$$

Pro 5 stupňů volnosti je 1% kritická hodnota rovna kvantilu $\chi^2_{1-0,01}(5) = 15,09$ (viz tabulku A.5). Hodnota

vypočtené testové statistiky $\chi^2 = 26,96$ značně překračuje kritickou hodnotu 15,09, a proto můžeme učinit závěr, že na hladině významnosti $\alpha = 0,01$ existuje dostatečný důkaz vztahu mezi léčbou a rentgenovým výsledkem. \diamond

10.2.2 Test hypotézy o nezávislosti

Na prvcích jediného výběru sledujeme dva kategorizované znaky. Naším cílem je testovat nulovou hypotézu o nezávislosti sledovaných znaků. Uvedme následující příklad testu nezávislosti na čtyřpolní tabulce.

Příklad 10.5 V náhodném výběru padesáti obézních dětí ve věku 6-14 let byla u každého dítěte zjištěna obezita u matky a obezita u otce (viz tabulku [10.9](#)). Zajímá nás, zda obezita rodičů spolu souvisí.

Tabulka 10.9: Pozorované četnosti
v příkladu [10.5](#)

	Otec		
Matka	Obézní	Neobézní	Celkem
Obézní	15	9	24
Neobézní	7	19	26
Celkem	22	28	50

Řešení: Nulová hypotéza pro tuto čtyřpolní tabulku říká, že obezita rodičů spolu nesouvisí, alternativní hypotéza je, že obezita rodičů spolu souvisí. Testovou statistiku tedy spočteme podle vzorce ([10.5](#)) jako

$$\chi^2 = 50 \cdot \frac{(15 \cdot 19 - 9 \cdot 7)^2}{24 \cdot 22 \cdot 26 \cdot 28} = 6,41.$$

Z tabulky [A.5](#) zjistíme, že kritickou hodnotou χ^2 testu pro $df = 1$ stupeň volnosti a hladinu $\alpha = 0,05$ je 3,84. Vidíme tedy, že nulovou hypotézu zamítneme na 5% hladině, takže obezita rodičů spolu významně souvisí. \diamond

Testová statistika χ^2 má χ^2 rozdělení (tj. test lze korektně provést) za předpokladu, že očekávané četnosti jsou dostatečně "velké". Čím menší jsou očekávané četnosti, tím pochybnější je test.

Obvykle se použití χ^2 testu doporučuje, pokud je nejméně 80 % očekávaných četností větších než 5 a všechny očekávané četnosti výskytu jsou větší než 1.

Čtyřpolní tabulka je speciálním případem kontingenční tabulky se dvěma řádky a dvěma sloupci, tj. $r = s = 2$. Na datech uspořádaných do čtyřpolní tabulky proto můžeme ověřovat jak hypotézu o nezávislosti sledovaných znaků, tak hypotézu o shodnosti dvou struktur. Test shodnosti dvou struktur je pro čtyřpolní tabulku ekvivalentní s testem hypotézy o shodnosti dvou populačních pravděpodobností. Jestliže některá očekávaná četnost ve čtyřpolní tabulce

je menší než pět, je popsán χ^2 test velmi nepřesný. K hodnocení čtyřpolní tabulky se někdy užívají korigované χ^2 testy, například *Yatesova korekce*, kterou spočteme následovně

$$\chi^2 = n \frac{(|ad - bc| - \frac{n}{2})^2}{(a+b)(a+c)(c+d)(b+d)} \quad (10.6)$$

Efektlem této korekce je zmenšení hodnoty testového kritéria, což znamená, že je obtížnější zamítnout nulovou hypotézu, tj. test je konzervativnější. Protože se snižuje riziko chyby prvního druhu (tj. zamítnutí nulové hypotézy, která je správná), zvyšuje se současně riziko chyby druhého druhu (tj. nezamítnutí nulové hypotézy, která je chybná). Někteří statistici doporučují užívat korigovaných hodnot (Yates, 1984), jiní argumentují proti jejich užití (Grizzle, 1967). Pro čtyřpolní tabulku však již pomocí počítače není problémem vypočítat exaktní Fisherův test, jehož podrobný popis naleznete například v [2].

Protože χ^2 testy pro čtyřpolní tabulku se snadno spočítají, užívá se jich často i tenkrát, kdy je vhodnější jiná

statistická metoda. Jinou chybou je, že sledované spojité veličiny jsou kategorizovány do ordinální, či dokonce dichotomické veličiny, aniž to vyžadují další okolnosti (například velmi malý počet pozorování). Tímto způsobem se ztrácí značná část informace obsažená v původních měřeních, kterou by jiná statistická metoda dokázala při volbě vhodného statistického testu plně využít.

10.2.3 Test hypotézy o symetrii

Testem symetrie pro čtyřpolní tabulku je *McNemarův test*. Jeho užití si ukážeme na následujícím příkladu.

Příklad 10.6 Předpokládejme, že máme studii se srovnatelnými skupinami nebo s podvojnými případy, kde výstupní hodnota má dvě kategorie - úspěch/neúspěch. Například pozorujeme náhodný výběr 18 pacientů, kteří byli léčeni dvěma různými antihypertenzívy A a B. Každý pacient dostával po dobu jednoho měsíce lék A a po odeznění jeho případných účinků dostával po dobu jednoho měsíce lék B.

Výsledek byl klasifikován jako úspěch nebo neúspěch. Za úspěch byl pokládán pokles krevního tlaku alespoň o 15 mm Hg. Každý jiný výsledek byl považován za neúspěch. Bylo by zásadně nesprávné uspořádat výsledky do tabulky způsobem uvedeným v tabulce 10.10 a provést výpočet testové statistiky χ^2 podle (10.5).

Tabulka 10.10: Výsledky experimentu - nesprávné uspořádání

	Lék A	Lék B	Celkem
Úspěch	10	4	14
Neúspěch	8	14	22

Celkem	18	18	36
--------	----	----	----

Řešení: Nyní nemůžeme provést χ^2 test jako v čtyřpolní tabulce, protože těchto 36 pozorování není nezávislých.

Každá osoba přispívá hodnotami do dvou políček. Abychom mohli analyzovat údaje, potřebujeme uspořádat výsledky do jiné podoby - viz tabulku [10.11](#).

Tabulka 10.11: Výsledky experimentu -
správné uspořádání

	Lék B		
Lék A	Úspěch	Neúspěch	Celkem
Úspěch	1	3	4
Neúspěch	9	5	14
Celkem	10	8	18

Zformulujme nulovou a alternativní hypotézu:

H_0 : Procenta úspěšnosti jsou u obou léků shodná.

H_1 : Procenta úspěšnosti nejsou u obou léků shodná.

Předpokládáme, že nulová hypotéza je pravdivá, a ptáme se, co bychom očekávali, že dostaneme. Zaměříme se na 12 pacientů, kteří vykazovali u každého léku jiné výsledky.

Pokud by byly léky stejně účinné, očekávali bychom, že těchto 12 pacientů bude rozděleno shodně (6 u každého léku) na ty, kteří byli neúspěšní při použití léku A a byli úspěšní při použití léku B, a na ty, kteří byli neúspěšní při použití léku B a úspěšní při A.

V našem případě jsme v těchto dvou políčkách pozorovali hodnoty 9 a 3 a očekávali 6 a 6. Potom χ^2 test můžeme sestavit pro pozorované a očekávané četnosti jako:

$$\chi^2 = \sum \frac{(\text{pozorované} - \text{očekávané})^2}{\text{očekávané}} = \frac{(9 - 6)^2}{6} + \frac{(3 - 6)^2}{6} = 3,00,$$

přičemž $df = 1$.

Zvolíme-li hladinu významnosti $\alpha = 0,05$, potom příslušná kritická hodnota je rovna kvantilu $\chi^2_{1-\alpha}(1) = 3,84$ (viz tabulku [A.5](#)).

Protože $3,00 < 3,84$, nemáme dostatečný důkaz pro zamítnutí nulové hypotézy. Učiníme tedy závěr, že na základě zkoumaných dat nelze prokázat rozdíl v působení obou léků. \diamond

Tabulka 10.12: Obecné uspořádání dat při McNemarově testu

	Léčba II		
Léčba I	+	-	Celkem
+	a	b	$a + b$
-	c	d	$c + d$
Celkem	$a + c$	$b + d$	n

Obecně lze testovou statistiku χ^2 pro McNemarův test spočítat pro data uspořádaná do tabulky [10.12](#) podle vzorce

$$\chi^2 = \frac{(b - c)^2}{b + c}. \quad (10.7)$$

Pro data z příkladu [10.6](#) tedy dostáváme

$$\chi^2 = \frac{(3 - 9)^2}{3 + 9} = \frac{36}{12} = 3,00,$$

což jsme zjistili již dříve. Vypočtenou hodnotu statistiky χ^2 pak porovnáme s kvantilem $\chi^2_{1-\alpha}(df)$, kde $df = 1$.

Poznámka: Speciálním případem je situace, kdy znak "léčba I" nahradíme znakem "před léčbou" a znak "léčba II" znakem "po léčbě". V tomto případě je nulová hypotéza formulována jako H_0 : "Procento pozitivního výsledku před léčbou je stejné jako po léčbě." Je to tedy podobná situace jako u párového t testu, ovšem místo dvojice spojených veličin měřených u každého pacienta sledujeme pouze výskyt nějaké vlastnosti před léčbou a po léčbě.

Další: [12. Analýza rozptylu](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [10. Analýza kategoriálních dat](#)

Kapitola

- [11.1 Obecné principy](#)
- [11.2 Lineární regrese a korelace](#)
- [11.3 Závislost kvalitativních znaků, Spearmanův korelační koeficient](#)

11. Měření závislosti

Jedinečnou funkcí statistiky je, že umožňuje vědci číselně vyjádřit nejistotu v jeho závěrech.

G. W. Snedecor

11.1 Obecné principy

V biomedicínském výzkumu často u pozorovaných osob nebo jiných objektů zjišťujeme současně řadu znaků, které mohou být vzájemně závislé. Například u pacientů můžeme zaznamenat tělesnou výšku, tělesnou hmotnost, výsledky laboratorních testů či jiné údaje. Předmětem statistické analýzy je stanovit sílu a druh sledovaných závislostí. *Sílu závislosti* (korelaci) vyjadřujeme prostřednictvím různých měr statistické závislosti, ke kterým patří i korelační koeficienty. Obecně požadujeme, aby absolutní hodnota míry statistické závislosti ležela v uzavřeném intervalu od nuly do jedné. Přitom dolní hranice (nuly) by měla dosahovat při nezávislosti a horní hranice (jedné) při funkční závislosti sledovaných veličin.

Statistická závislost však zdaleka nemusí znamenat kauzalitu. Vysoký stupeň závislosti (korelace) často odráží příčinný vztah, ale nemusí tomu tak být vždy. Příčinné souvislosti čistě empirickými prostředky neodhalíme. Ke statistickým výsledkům je třeba přidat odborné znalosti, praktické zkušenosti a účelně kombinovat deduktivní a induktivní způsob uvažování. V nejjednodušším případě je souvislost mezi sledovanými znaky zcela jednoznačná. Například hmotnost předmětů, které jsou vytvořeny ze stejného materiálu, lze spočítat pomocí jejich objemu. Závislost tohoto druhu se nazývá *funkční závislost*. Předmětem statistiky je však hodnocení jiných závislostí, kdy neexistuje zcela jednoznačný vztah mezi sledovanými znaky. Při měření závislosti dvou kvantitativních znaků můžeme druh a sílu závislosti orientačně posoudit z bodového grafu, v němž je každá dvojice údajů graficky znázorněna jedním bodem v rovině. Druh závislosti odhadujeme pomocí křivky, která se dobře hodí k napozorovaným hodnotám. Podle typu křivky rozeznáváme závislost lineární, logaritmickou, exponenciální a další. Jedním z úkolů statistické analýzy dat je i vyjádření síly závislosti mezi sledovanými znaky, tj. stanovení, do jaké míry je hodnota jednoho znaku předurčena hodnotou druhého znaku.

V následujících částech této kapitoly si všimneme problému měření závislosti pro nejjednodušší případ dvou znaků. Problematikou měření závislosti v případě většího počtu znaků se budeme zabývat v dalších dílech této edice.

11.2 Lineární regrese a korelace

Nejprve se budeme zabývat měřením závislosti mezi dvěma kvantitativními znaky neboli veličinami.

Často chceme prozkoumat vztah mezi dvěma veličinami, kde jedna z nich, tzv. *nezávisle proměnná* X , má řídit druhou, tzv. *závisle proměnnou* Y . Předpokládá se, že obě veličiny jsou spojitě. Prvním krokem ve zkoumání by mělo být zakreslení dat do grafu. V řadě případů tento krok napoví mnohé o tom, co nás zajímá: Existuje vztah mezi oběma proměnnými (veličinami)? Pokud ano, pak rostou či klesají obě v jednom směru, nebo jedna klesá, když druhá roste? Je přímka vhodným modelem pro vyjádření vztahu mezi těmito dvěma veličinami? Chceme-li se dostat dále za tuto intuitivní úroveň analýzy, je lineární regrese často užitečným nástrojem. Tato metoda zahrnuje proložení přímky daty a analýzu statistických vlastností takovéto přímky.

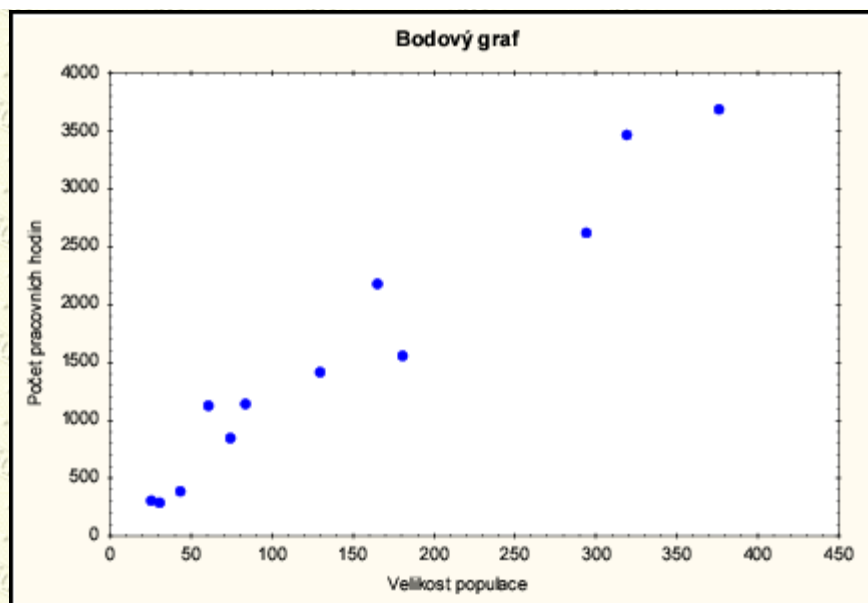
Příklad 11.1 Máme analyzovat data o počtu pracovních hodin za měsíc spojených s provozováním anesteziologické služby v závislosti na velikosti spádové populace nemocnice - viz tabulku 11.1. Údaje byly získány ve 12 nemocnicích ve Spojených státech.

Tabulka 11.1: Údaje z 12 amerických nemocnic

Nemocnice	Celkový počet pracovních hodin	Spádová populace (v tisících osob)
1	304,37	25,5
2	2616,32	294,3
3	1139,12	83,7
4	285,43	30,7
5	1413,77	129,8
6	1555,68	180,8
7	383,78	43,4
8	2174,27	165,2
9	845,30	74,3
10	1125,28	60,8
11	3462,60	319,2
12	3682,33	376,2

Nejprve bychom měli zakreslit data do grafu. Je pravidlem vynášet závisle proměnnou Y (zde pracovní doba) na svislou osu a nezávisle proměnnou X (zde velikost spádové populace) na vodorovnou osu.

Obrázek 11.1: Závislost počtu pracovních hodin na velikosti populace



Graf na obrázku 11.1 naznačuje lineární vztah mezi pracovní dobou a velikostí populace, a tak budeme pokračovat kvantifikací tohoto vztahu. Regresní přímku proložíme na základě nasbíraných dat, obecný zápis bude:

$$\text{pracovní doba} = \alpha + \beta \cdot \text{velikost populace} + \text{náhodná složka},$$

kde α je **absolutní člen** - ten lze interpretovat jako bod, v němž regresní přímka protíná svislou osu, β je **směrnice** ("sklon") přímky a **náhodná složka** je ta část, kterou nelze vysvětlit lineárním vztahem. Je zřejmé, že všechny body neleží přesně na přímce.

V daném bodovém grafu existuje velmi mnoho možností, jak proložit přímku. V řadě případů stačí pro prvotní orientaci přímka nakreslená od ruky. Přejeme-li si však vytvořit přímku nějakým obecně platným způsobem, který bude možno opakovat, potom musíme použít odpovídající kritérium, odpovídající metodu. Nejčastěji používaná metoda se nazývá **metoda nejmenších čtverců**.

Pro kteroukoli přímku můžeme spočítat hodnoty **reziduí**, což jsou vzdálenosti (rovnoběžné se svislou osou) od bodů k přímce. Označíme-li si a absolutní člen a b směrnici naší přímky, pak v bodě x_i bude y -souřadnice bodu na přímce

$$\hat{y}_i = \alpha + b x_i. \quad (11.1)$$

Reziduem je tedy rozdíl mezi skutečnou (pozorovanou) hodnotou y_i a hodnotou na přímce \hat{y}_i , tj.

$$y_i - \hat{y}_i. \quad (11.2)$$

Když tyto vzdálenosti umocníme na druhou a sečteme je, dostaneme hodnotu

$$(11.3)$$

$$S_e = \sum_i (y_i - \hat{y}_i)^2,$$

kterou nazýváme **reziduální součet čtverců**. Přímkou, pro kterou dostáváme nejmenší možnou hodnotu S_e , nazveme přímkou vyhovující kritériu nejmenších čtverců nebo přímkou získanou metodou nejmenších čtverců. V praxi nemusíme počítat řadu hodnot S_e , protože poměrně jednoduchými matematickými prostředky získáme vzorce jak pro výpočet směrnice, tak absolutního členu, které odpovídají nejmenší možné hodnotě S_e .

Podíl reziduálního součtu čtverců S_e a příslušného počtu stupňů volnosti - v případě regresní přímky $df = n - 2$ - nazýváme **reziduální rozptyl** a značíme jej s^2 , tj.

$$s^2 = \frac{S_e}{n - 2}. \quad (11.4)$$

Regresní koeficienty vypočtené metodou nejmenších čtverců mají následující obecný tvar. Směrnici b spočteme podle vzorce

$$b = \frac{s_{xy}}{s_x^2} \quad (11.5)$$

a absolutní člen a podle vzorce

$$a = \bar{y} - b\bar{x}, \quad (11.6)$$

přičemž \bar{x} a \bar{y} jsou výběrové průměry

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (11.7)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (11.8)$$

Podobně

$$s_x^2 = \frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (11.9)$$

$$s_y^2 = \frac{1}{n - 1} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (11.10)$$

jsou odhady rozptylů a

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (11.11)$$

je odhad tzv. **kovariance** sledovaných veličin X a Y .

Odhad absolutního členu (a) a směrnice (b) regresní přímky a jejich statistické charakteristiky jsou uvedeny v tabulce [11.2](#). Směrodatná chyba koeficientu je výběrová směrodatná odchylka odhadovaného parametru. (Ve statistických programech je obvykle označována anglicky jako Standard Error.)

Tabulka 11.2: Statistické charakteristiky lineární regrese

Parametr	Koeficient	Směrodatná chyba koeficientu	t	p
Absolutní člen (a)	180,658	128,381	1,407	0,1897
Směrnice (b)	9,429	0,681	13,846	< 0,0001

Výsledky regresní analýzy jsou často na výstupu z počítače prezentovány ve formě [tabulky analýzy rozptylu](#) (viz tabulku [11.3](#)), kterou se podrobněji zabývá kapitola [12](#). Tam je také vysvětleno zde použité značení.

Tabulka 11.3: Tabulka analýzy rozptylu pro lineární regresi pracovní doby na velikosti spádové populace

Zdroj variability	Součet čtverců SS	df	$MS = SS/df$	F	c
Regresní	14346071	1	14346071,0	191,7	< 0,0001
Reziduální	748192	10	74819,2		
Celkový	15094263	11			

Používáme-li model regresní analýzy na statistické zpracování našich dat, je dobré ověřit předpoklady, ze kterých model vychází. Shrňme je v následujících třech bodech.

1. Závisle proměnná Y má normální rozdělení pro každou hodnotu nezávisle proměnné $X = x$.
2. Rozptyl závisle proměnné Y je stejný pro každou hodnotu nezávisle proměnné $X = x$.
3. Závislost veličiny Y na X je lineární.

Pro tuto chvíli předpokládejme, že pro náš příklad jsou tyto předpoklady splněny. Z tabulky [11.2](#) tedy dostáváme:

$$\text{pracovní doba} = 180,658 + 9,429 \cdot \text{velikost populace.}$$

To je třeba interpretovat jako odhad průměrné hodnoty počtu pracovních hodin pro populaci s danou velikostí. Očekáváme, že na každých dalších 1 000 lidí stoupne za měsíc počet pracovních hodin o 9,429, což je směrnice regresní přímky. Uvědomte si, že absolutní člen (180,658) značí průměrný počet pracovních hodin, když je populace rovna nule. To zřejmě nedává smysl a mělo by nám to připomenout (pokud takové připomenutí je zapotřebí), že model by se měl používat pouze v tom rozmezí obou veličin, v němž se pohybovaly pozorované hodnoty. V tomto případě to znamená X od 26 do 370. Je ovšem pravda, že dosažená hladina významnosti pro absolutní člen je přibližně 0,19, a nelze tedy říci, že by se absolutní člen (a) významně lišil od nuly.

Připomeňme, že tyto výsledky jsme spočetli pro náhodný výběr 12 nemocnic. Kdybychom teď zvolili jiný náhodný výběr 12 nemocnic, dostali bychom odlišný odhad směrnice a absolutního členu. Pokusíme se teď odhadnout směrnici "pravé" přímky. Musíme určit interval spolehlivosti pro směrnici přímky, stejně jako když jsme odhadovali průměr, rozdíl dvou průměrů nebo jiné parametry.

Vztah určující krajní body $100P\%$ **intervalu spolehlivosti pro směrnici** je tvaru:

odhadnutá směrnice $\pm t_{1-\frac{\alpha}{2}}(df) \cdot$ směrodatná chyba směrnice,
tj.

$$\hat{b} \pm t_{1-\frac{\alpha}{2}}(df) s_b, \quad (11.12)$$

kde **směrodatná chyba směrnice** s_b se vypočte podle vzorce

$$s_b = \frac{s}{s_x \sqrt{n-1}} \quad (11.13)$$

(nalezneme ji ve sloupci "Směrodatná chyba koeficientu" v tabulce [11.2](#)), $\alpha = 1 - P$ a počet stupňů volnosti (df) pro kvantil t rozdělení je určen počtem stupňů volnosti reziduálního součtu čtverců (S_e). Pro regresní přímku to bude vždy $df = n - 2$, kde n je počet pozorování (v našem příkladu je $n = 12$, takže $df = 10$). Dva "ztracené" stupně volnosti vyplývají z nutnosti odhadnout dva parametry - směrnici a absolutní člen. Kvantil $t_{1-\frac{\alpha}{2}}(n-2)$ pro $\alpha = 0,05$ nalezneme v tabulce [A.3](#). Výsledný 95% interval spolehlivosti má tedy podle vzorce ([11.12](#)) krajní body

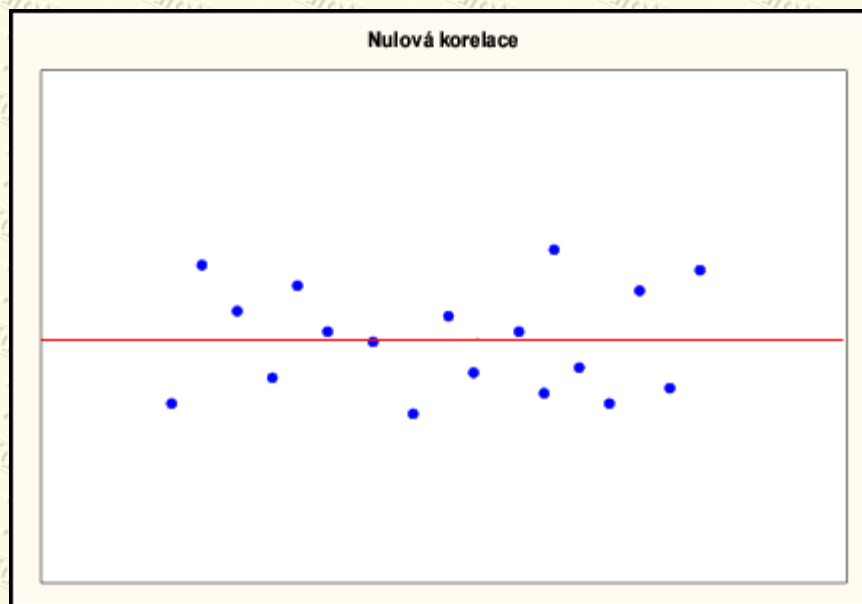
$$9,429 \pm 2,228 \cdot 0,681 = 9,429 \pm 1,517 \rightsquigarrow (7,912; 10,946).$$

Náš nejlepší odhad změny průměrného počtu pracovních hodin v závislosti na změně velikosti populace (v tisících) je ten, že leží mezi 7,912 a 10,946 pracovními hodinami za měsíc.

V našem příkladu jsou charakter závislosti (linearita) i síla závislosti dosti zřejmé již z bodového grafu. Někdy však graf naznačuje, že sice může jít o lineární závislost, avšak strmost stoupání (či klesání) přímky není příliš přesvědčivá. Existenci lineárního vztahu mezi našimi dvěma veličinami zjišťujeme tak, že se formálně ptáme, zda je směrnice

rovna nule. Pokud je odpověď na tuto otázku kladná, znamená to, že pozorovaná směrnice 9,429 se liší od nuly pouze náhodně a v novém pokusu by se mohla stejně snadno rovnat $-9,429$, tj. mohlo by se zdát, že pracovní hodiny klesají s velikostí populace (sotva pravděpodobné). Jeden příklad téměř nulové směrnice je znázorněn na obrázku 11.2. Nezdá se, že by na tomto obrázku existoval přímý vztah mezi X a Y .

Obrázek 11.2: Příklad nulové směrnice



Zformulujeme tedy nulovou a alternativní hypotézu.

H_0 : Směrnice je nula, tj. $\beta = 0$.

H_1 : Směrnice není nula, tj. $\beta \neq 0$.

Pokud nulová hypotéza platí, pak má statistika

$$t = \frac{\text{odhadnutá směrnice}}{\text{směrodatná chyba odhadnuté směrnice}} = \frac{\hat{\beta}}{s_{\hat{\beta}}} \quad (11.14)$$

Studentovo t rozdělení o $n - 2 = 10$ stupňů volnosti. Za platnosti nulové hypotézy by měla hodnota statistiky t s 95% pravděpodobností ležet v intervalu $(-2,228; 2,228)$.

Pro naše data vyjde:

$$t = \frac{9,429}{0,681} = 13,846.$$

Tato hodnota je mnohem větší než 2,228 (dosažená hladina významnosti p je téměř nula), což představuje silný důkaz proti nulové hypotéze, že neexistuje žádný lineární vztah. Všimněte si, že výsledná tabulka [11.2](#) z počítačového výstupu udává hodnotu t statistiky a dosaženou hladinu významnosti p . Nezapomeňte, že dosažená hladina významnosti je pravděpodobnost, že dostaneme náš, nebo ještě extrémnější výsledek za předpokladu platnosti nulové hypotézy (tj. že skutečná směrnice je nulová). Za pozornost stojí také fakt, že test je vlastně zopakován v tabulce [11.3](#), neboť $F = t^2$.

Jak jsme již vypočítali, interval spolehlivosti pro směrnici je $(7,912; 10,946)$. Neobsahuje tedy nulu, což odpovídá závěru získanému z t statistiky - počet pracovních hodin skutečně lineárně závisí na rozsahu spádové populace. \diamond

Korelace

Pro měření síly *lineární* závislosti mezi dvěma spojitými náhodnými veličinami se používá tzv. **(Pearsonův) korelační koeficient**. Počítá se podle vzorce

$$r = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}}, \quad (11.15)$$

kde s_x^2 a s_y^2 jsou výběrové rozptyly a s_{xy} je kovariance - viz [\(11.9\)](#)-[\(11.11\)](#).

Korelační koeficient r a směrnice regresní přímky b jsou tedy svázány následujícími vztahy

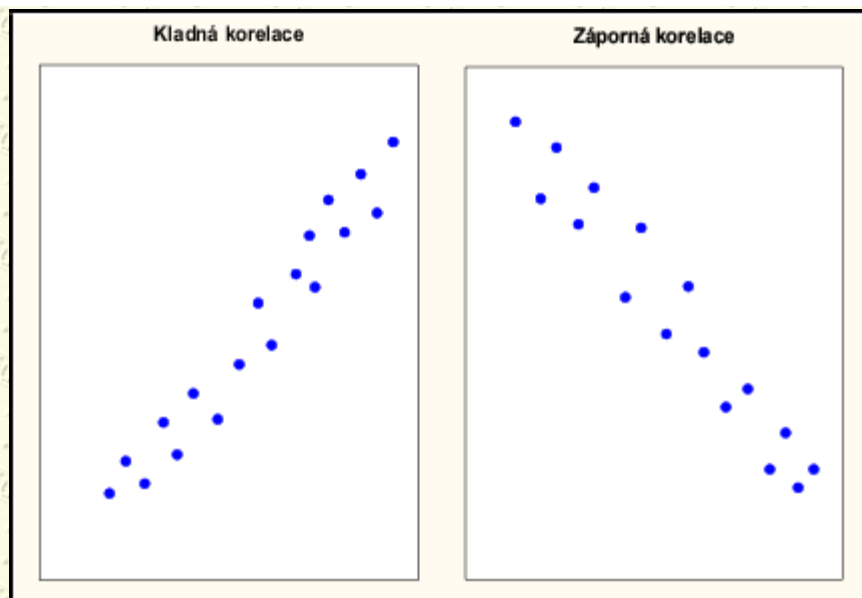
$$b = r \frac{s_y}{s_x}, \quad r = b \frac{s_x}{s_y},$$

kde s_x je směrodatná odchylka veličiny X a s_y směrodatná odchylka veličiny Y .

Hodnota korelačního koeficientu se pohybuje od -1 do 1 . Hodnoty ± 1 nabývá tehdy, pokud všechny body $[x_i, y_i]$ leží na přímce. Nule je roven v případě, že veličiny jsou nezávislé. Korelační koeficient však může být nulový i v případě, že veličiny jsou funkčně závislé, ale závislost *není lineární*. Proto je při užití Pearsonova korelačního koeficientu vždy třeba posoudit, zda je jeho aplikace vhodná. Při měření lineární závislosti je znaménko korelačního koeficientu kladné, když obě veličiny X a Y zároveň rostou nebo obě zároveň klesají, a záporné, když jedna z veličin roste, zatímco druhá klesá.

Na obrázku [11.3](#) jsou znázorněny typické případy dat s kladným a záporným korelačním koeficientem.

Obrázek 11.3: Příklady dat s kladným a záporným korelačním koeficientem



Při pohledu na hodnotu korelačního koeficientu by nás tedy mělo zajímat znaménko a velikost (absolutní hodnota). Znaménko udává směr, zatímco velikost udává, jak blízko jsou body nashromážděny kolem přímky. Často se setkáte s tím, že je korelace uváděna s dosaženou hladinou významnosti p . Ta se obvykle vztahuje k testu nulové hypotézy, že korelační koeficient (dvourozměrné) populace je nula, tedy že mezi veličinami X a Y neexistuje žádný lineární vztah.

Zde se je ale třeba zamyslet nad praktickým významem - pro rozsáhlé výběry se bude i velmi malá absolutní hodnota korelačního koeficientu statisticky významně lišit od nuly. Opět, jako pro jakoukoli jinou statistiku, můžeme vytvořit intervaly spolehlivosti pro populační koeficient korelace. Dále je důležité si uvědomit, že *korelace neznamena příčinnost*. Všechny tyto poznámky a úvahy shrneme do několika hlavních pravidel:

1. Pearsonův korelační koeficient je užitečný pro určení síly lineární závislosti mezi dvěma spojitými veličinami.
2. Vždy si nakreslete bodový graf, získáte rozumnou vizuální představu.
3. Podívejte se na znaménko korelačního koeficientu.
4. Podívejte se na velikost korelačního koeficientu.
5. Korelace neznamena příčinnost.
6. Hodnoty dosažené hladiny významnosti p uvedené spolu s hodnotami korelačních koeficientů berte kriticky - pro velký rozsah výběru může i *prakticky nevýznamná* korelace být *významná statisticky*.

Pokud nás zajímá pouze síla lineární závislosti, používáme místo korelačního koeficientu r spíše jeho druhou mocninu r^2 , které se říká *koeficient determinace*. (V literatuře a v počítačových výstupech se také často značí R^2 .)

Někdy nemáme jasně určeno, která veličina je nezávislá a která závislá. Lineární regrese X na Y nedává stejnou regresní přímku jako regrese Y na X . To proto, že regresní analýza předem očekává přímý "řídící" vztah, tj.

předpokládá se, že X ovlivňuje Y , a ne naopak. Přesto bude koeficient determinace r^2 získaný z obou regresí stejný, neboť měří jen velikost lineárního vztahu mezi X a Y bez ohledu na to, která veličina je nezávislá a která závislá.

Koeficient determinace je i měrou vhodnosti modelu. Udává tu část variability Y , kterou lze pomocí modelu vysvětlit. Jinými slovy, pouze $(1 - r^2) \cdot 100$ % variability Y nelze vysvětlit variabilitou X . Přesněji řečeno, tzv. **celkový součet čtverců**

$$S_T = \sum_i (y_i - \bar{y})^2 \quad (11.16)$$

lze rozložit na reziduální součet čtverců S_e - viz (11.3) a **regresní součet čtverců**

$$S_R = \sum_i (\hat{y}_i - \bar{y})^2, \quad (11.17)$$

kde \hat{y}_i je y -souřadnice i -tého bodu na přímce získané metodou nejmenších čtverců. Tedy

$$S_T = S_R + S_e. \quad (11.18)$$

Lze dokázat, že koeficient determinace lze vyjádřit i ve tvaru

$$r^2 = \frac{S_R}{S_T}, \quad (11.19)$$

z něhož plyne výše uvedené tvrzení o procentu vysvětlené variability. V příkladu 11.1 nalezneme jednotlivé součty čtverců ve sloupci "Součet čtverců SS" (Sum of Squares) tabulky 11.3. Podle vzorce (11.19) tedy snadno vypočteme koeficient determinace $r^2 = 0,95$. Znamená to, že velikost populace (X) vysvětluje 95 % variability celkového počtu pracovních hodin (Y).

Předpoklady modelu a jejich ověření

Jako u každé statistické metody, i v případě regresní analýzy mohou být získané závěry značně chybné, pokud nejsou splněny předpoklady, na nichž je model založen. Proto se vyplatí provést nějakou jednoduchou kontrolu těchto předpokladů. Bez použití počítače by to bylo dost pracné, ale pokud máte k dispozici slušný statistický program, vyžaduje to jen malé úsilí.

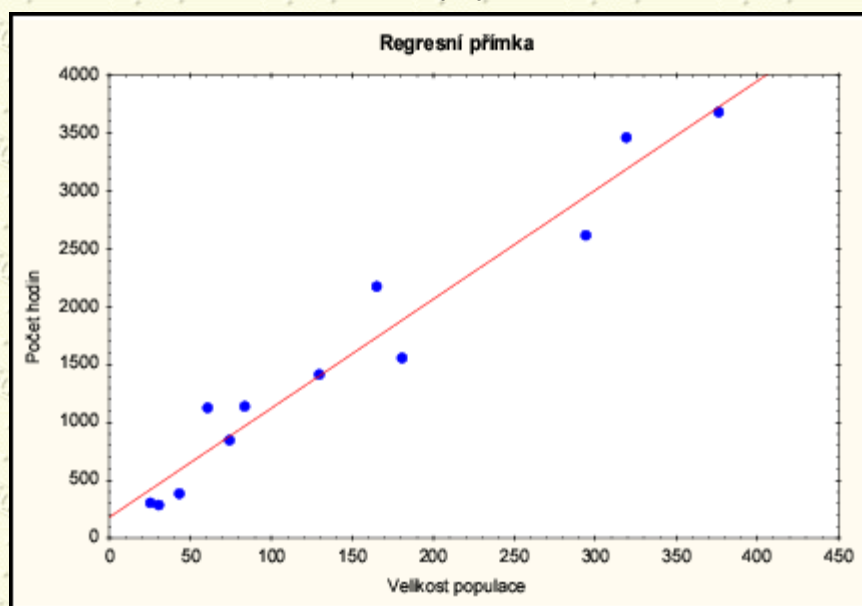
Statistický model, na kterém je založena lineární regrese, předpokládá, že pro kteroukoli danou hodnotu prediktoru (nezávisle proměnné X) má odpovídající veličina (závisle proměnná Y) střední hodnotu (průměr), která je dána příslušnou hodnotou na regresní přímce. Opakovaná pozorování Y při daném $X = x$ budou nabývat hodnot kolem střední hodnoty a jejich rozdělení by mělo být normální s nějakým (neznámým) rozptylem σ^2 . Tento rozptyl by měl být bez ohledu na hodnotu $X = x$ stejný. (Této vlastnosti se říká *homoskedasticita*.) Rozptyl σ^2 se odhaduje

reziduálním rozptylem s^2 - viz (11.4).

V příkladu 11.1 vyjde reziduální rozptyl $s^2 = 74819,2$. Toto číslo nalezneme v tabulce 11.3 ve sloupci "MS" (Mean Square) a řádku "reziduální".

Důsledkem tohoto modelu je, že celá systematická variabilita Y je zahrnuta v odhadu přímky a žádný další systematický rozptyl (přesněji řečeno, systematický trend) by data neměla obsahovat. Při prohlédnutí našeho bodového grafu a v něm zanesené regresní přímky (viz obrázek 11.4) nejsou vidět žádné zřetelné a systematické odchylky bodů od přímky.

Obrázek 11.4: Regresní přímka závislosti "počtu hodin" na "velikosti populace"



Obvyklým postupem při ověřování předpokladů modelu je nejprve z dat eliminovat efekt regresní přímky (tedy lineární funkce) a potom hledat systematické chování reziduí. Ta, jak už víme, získáme jako rozdíl pozorované hodnoty y_i a odpovídající hodnoty na regresní přímce \hat{y}_i .

V tabulce 11.4 jsou zobrazeny předpovězené a pozorované hodnoty pro každou z dvanácti nemocnic.

Tabulka 11.4: Pozorované a předpovězené hodnoty a rezidua

Nemocnice	Velikost populace	Pracovní hodiny pozorované	Pracovní hodiny předpovězené	Reziduum
1	25,5	304,37	421,10	-116,73
2	294,3	2616,32	2955,61	-339,29
3	83,7	1139,12	969,86	169,26

4	30,7	285,43	470,13	-184,70
5	129,8	1413,77	1404,54	9,23
6	180,8	1555,68	1885,42	-329,74
7	43,4	383,78	589,88	-206,10
8	165,2	2174,27	1738,33	435,94
9	74,3	845,30	881,23	-35,93
10	60,8	1125,28	753,94	371,34
11	319,2	3462,60	3190,39	272,21
12	376,2	3682,33	3727,84	-45,51

Například pro nemocnici 1 je $x_1 = 25,5$, což znamená rozsah spádové populace 25500 lidí. Předpovězená pracovní doba je tedy

$$180,658 + 9,429 \cdot 25,5 = 421,10$$

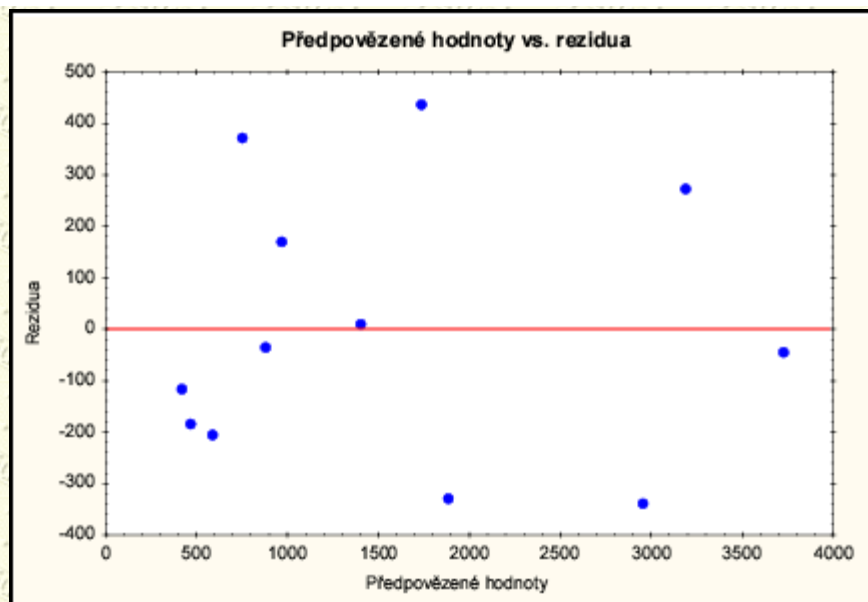
a residuum

$$304,37 - 421,10 = -116,73.$$

Reziduální rozptyl s^2 , resp. jeho odmocnina $s = 273,5$ představuje opět míru rozptýlenosti těchto reziduí.

Porovnání reziduí s čímkoli dalším (pozorovanými hodnotami, odhadnutými hodnotami, hodnotami X) by nemělo ukázat žádné systematické závislosti. Nejužitečnější je v takovém případě často graf reziduí a předpovězených hodnot - viz obrázek [11.5](#).

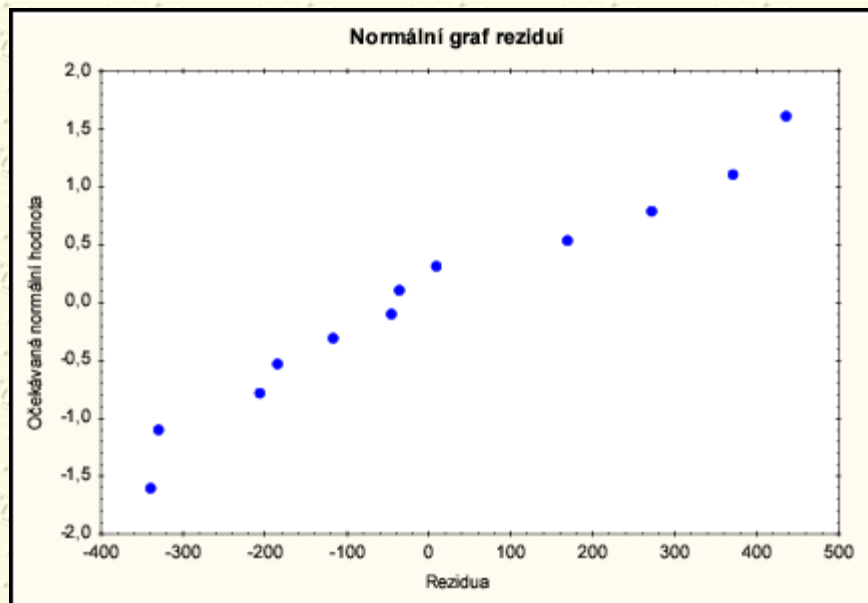
Obrázek 11.5: Graf reziduí a předpovězených hodnot



Rezidua jsou náhodně rozmístěna kolem nuly a nemají žádný zřejmý vztah k předpovídaným hodnotám: ani se systematicky nezvyšují ani se systematicky nesnižují spolu s rostoucími předpovídanými hodnotami a není zde ani náznak nelineárního vztahu.

Protože předpokládáme, že kolísání hodnot závisle proměnné kolem regresní přímky je dáno normálním rozdělením, rezidua by se měla chovat alespoň přibližně jako výběr z normálního rozdělení s nulovou střední hodnotou. Normální graf reziduí by tedy měl být přibližně přímkou - viz obrázek [11.6](#).

Obrázek 11.6: Normální graf reziduí



V tomto normálním grafu tvoří data zhruba přímkou, což tedy nevede ke zpochybnění předpokladu normality dat, která je základem již provedených testů a výpočtů intervalů spolehlivosti.

11.3 Závislost kvalitativních znaků, Spearmanův korelační koeficient

Již v 9. kapitole jsme ukázali, jakým způsobem lze jednoduše hodnotit data uspořádaná podle velikosti.

[Neparametrické metody](#) založené na pořadích rozšíříme nyní o metodu, která využívá pořadí při zjišťování závislosti dvou znaků. Dobrým ukazatelem takové závislosti je **Spearmanův korelační koeficient**, založený na pořadích jedinců uspořádaných podle velikosti vzhledem ke dvěma sledovaným veličinám. Každému jedinci se přiřadí dvojice pořadí Q (pořadí podle první veličiny X) a R (pořadí podle druhé veličiny Y). Kdyby s rostoucími hodnotami X vzrůstaly i hodnoty Y , byla by zřejmě pořadí obou veličin shodná, tj. $Q = R$ pro každého jedince. Jestliže s rostoucími hodnotami X klesají hodnoty Y , jsou pořadí obou veličin právě opačná. Při nezávislosti jsou pořadí zpřeházená zcela náhodně. Pro n pozorovaných dvojic ve výběru se Spearmanův korelační koeficient počítá pomocí diferencí pořadí $d_i = Q_i - R_i$ jako

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}. \quad (11.20)$$

Při shodném pořadí dosahuje koeficient r_s maximální hodnoty 1, při opačném pořadí minimální hodnoty -1 .

V ostatních případech je $-1 < r_s < 1$. Hodnoty korelačního koeficientu blízké nule naznačují, že pořadí jsou náhodně zpřeházená, a mezi sledovanými veličinami tedy není závislost. Při platnosti nulové hypotézy o nezávislosti obou veličin jsou odchylky Spearmanova korelačního koeficientu od nuly jen náhodné. Když tedy absolutní hodnota Spearmanova korelačního koeficientu $|r_s|$ překročí 5% nebo 1% kritickou hodnotu, zamítá se nulová hypotéza o nezávislosti na příslušné hladině významnosti. Kritické hodnoty Spearmanova korelačního koeficientu pro oboustrannou alternativu jsou uvedeny v tabulce [A.10](#) pro $n \leq 30$. Pro větší rozsah výběru můžeme použít testové statistiky

$$Z = r_s \sqrt{n-1}, \quad (11.21)$$

kteřá má za platnosti nulové hypotézy standardizované normální rozdělení. Jestliže tedy vyjde $|Z| \geq z_{1-\frac{\alpha}{2}}$, zamítáme nulovou hypotézu ve prospěch alternativy na zvolené hladině významnosti α .

Příklad 11.2 Na interním oddělení bylo léčeno na určitou chorobu deset pacientů. Za čtyři týdny po ukončení hospitalizace byli pozváni ke kontrolní prohlídce, při níž se mimo jiných údajů sledovala též sedimentace červených krvinek. Závažnost klinického průběhu všech deseti pacientů lze zhruba vyjádřit na základě uspořádání podle závažnosti zjištěných klinických příznaků do řady, jejíž členy označíme jedničkou (nejlehčí průběh) až desítkou (nejtěžší průběh). Označme pacienty velkými písmeny A, B, C atd. Jejich hodnocení podle závažnosti klinického průběhu, zjištěné hodnoty sedimentace a pořadí podle výše sedimentace jsou uvedeny v tabulce [11.5](#). Ověřte na 5% hladině významnosti, zda hodnota sedimentace nezávisí na klinickém průběhu onemocnění.

Tabulka 11.5: Uspořádání deseti pacientů podle závislosti klinického průběhu onemocnění a podle výše sedimentace

Pacient	A	B	C	D	E	F	G	H	I	K
Pořadí podle klinického průběhu (Q)	4	6	1	5	10	2	7	3	9	8
Sedimentace	19	22	26	30	33	23	28	21	59	39
Pořadí podle sedimentace (R)	1	3	5	7	8	4	6	2	10	9
Diference pořadí ($d = Q - R$)	3	3	-4	-2	2	-2	1	1	-1	-1
Čtverec diference (d^2)	9	9	16	4	4	4	1	1	1	1

Řešení: U každého pacienta vypočteme rozdíl mezi oběma pořadími, tj. $d = Q - R$, tyto rozdíly umocníme na druhou a sečteme. Dostáváme $\sum d_i^2 = 50$, takže

$$r_s = 1 - \frac{6 \cdot 50}{10(100 - 1)} = 1 - \frac{300}{990} = 0,6970.$$

Porovnáním vypočtené hodnoty korelačního koeficientu $r_s = 0,6970$ s příslušnou 5% kritickou hodnotou 0,6364 (viz tabulku [A.10](#), kde $n = 10$ a $\alpha = 0,05$) dojdeme k závěru, že zamítáme hypotézu o nezávislosti na 5% hladině významnosti. Prokázali jsme tak statistickou závislost výše sedimentace na klinickém průběhu onemocnění. \diamond

[Další](#) [Výš](#) [Předchozí](#) [Obsah](#) [Rejstřík](#)

Další: [12. Analýza rozptylu](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [10. Analýza kategoriálních dat](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [13. Statistické tabulky](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [11. Měření závislosti](#)

Kapitola

- [12.1 Porovnání průměrů pro tři a více skupin](#)
- [12.2 Ověřování předpokladů modelu](#)
- [12.3 Mnohonásobná porovnávání](#)
- [12.4 Obecné poznámky k analýze rozptylu](#)

12. Analýza rozptylu

Dobrá zpráva je, že statistická analýza se stává snazší a levnější.

Špatná zpráva je, že statistická analýza se stává snazší a levnější.

C. F. Hofacker

12.1 Porovnání průměrů pro tři a více skupin

V předchozích částech jsme viděli jak porovnávat průměry dvou skupin. V mnoha situacích ale potřebujeme porovnat průměry více skupin. V této části se budeme zabývat analýzou tří nebo více *nezávislých* skupin (výběrů).

Příklad 12.1 Dvacet dva pacientů, kteří podstoupili operaci srdce, bylo náhodně rozděleno do tří skupin.

Skupina 1:

Pacienti dostali 50 % oxidu dusného a 50 % kyslíkové směsi nepřetržitě po dobu 24 hodin;

Skupina 2:

Pacienti dostali 50 % oxidu dusného a 50 % kyslíkové směsi pouze během operace;

Skupina 3:

Pacienti nedostali žádný oxid dusný, ale dostali 35-50 % kyslíku po dobu 24 hodin.

Tabulka [12.1](#) ukazuje koncentraci soli kyseliny listové v červených krvinkách ve všech třech skupinách po uplynutí 24 hodin ventilace, dále počet pozorování (n), průměr (\bar{x}) a směrodatnou odchylku (s) ve všech třech skupinách.

Tabulka 12.1: Koncentrace soli kyseliny listové v červených krvinkách

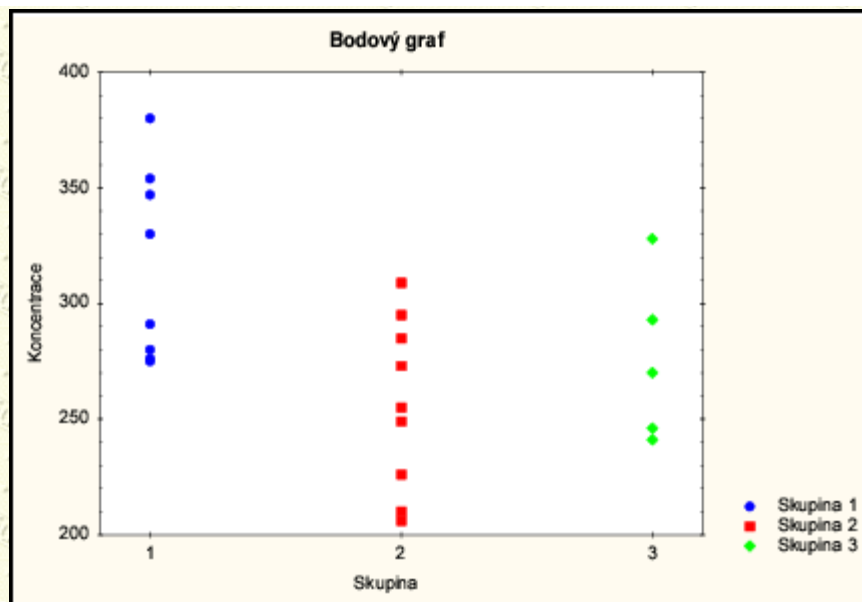
	Skupina 1	Skupina 2	Skupina 3

1	276	206	241
2	280	210	246
3	275	226	270
4	291	249	293
5	347	255	328
6	354	273	
7	380	285	
8	330	295	
9		309	
n	8	9	5
\bar{x}	316,6	256,4	275,6
s	41,2	37,1	35,9

Řešení: Opět předpokládáme, že v rámci jedné skupiny jsou pozorování navzájem nezávislá. V tomto případě jsou nezávislá všechna, protože pocházejí od různých pacientů. Pokud bychom ale například měli u každé osoby hodnoty naměřené třikrát, bylo by nutno vzít průměrnou hodnotu trojice měření, protože původní hodnoty by při opakovaných měřeních nebyly nezávislé.

Prvním krokem při analýze takovýchto dat je přezkoumat je vizuálně, tedy nakreslit si vhodný graf. (Říkáme také, že data jsou vynesena do grafu.) To provedeme vytvořením bodového grafu (viz obrázek [12.1](#)). Datový soubor je dostatečně malý, abychom mohli zobrazit jednotlivé body. Pro větší soubory dat je obvykle lepší zobrazovat krabicové grafy, protože body se navzájem mohou překrývat a jen obtížně vidíme, kde jsou data soustředěna. V našem případě se nezdá, že by soubor obsahoval nějaká odlehlá pozorování, a vypadá to, že mezi skupinami je rozdíl ve výši hodnot koncentrace soli kyseliny listové v červených krvinkách.

Obrázek 12.1: Bodový graf dat z příkladu [12.1](#)



První otázka zní, proč nemůžeme provést všechny možné dvouvýběrové t testy nezávislosti, tedy testy pro následující dvojice:

Skupina 1 vs. Skupina 2,
 Skupina 1 vs. Skupina 3,
 Skupina 2 vs. Skupina 3.

Důvodem, proč nemůžeme analýzu provést tímto způsobem, je, že toto opakované porovnávání významnosti neoprávněně zvyšuje pravděpodobnost chyby prvního druhu. U každého testu je řekněme 5% možnost chybného pozitivního výsledku (tedy chyby prvního druhu), pokud neexistuje žádný rozdíl. Tato pravděpodobnost, jak víme, se nazývá pravděpodobnost chyby prvního druhu neboli hladina významnosti α . Máme-li tři skupiny a provedeme všechny tři testy, pravděpodobnost, že dostaneme nejméně jeden chybný pozitivní výsledek (chybu prvního druhu) je větší než 5%. Abychom se tomuto problému vyhnuli, použijeme k testování hypotézy metodu analýzy rozptylu. Než vysvětlíme způsob použití této metody, podíváme se na matematický model, na kterém je založena.

Analýza rozptylu (ANOVA) je založena na předpokladu, že každý z výběrů pochází z populace s normálním rozdělením se stejnou směrodatnou odchylkou. Zajímá nás, zda střední hodnoty (průměry) skupin jsou všechny shodné, nebo zda se navzájem liší. Jako v předešlém případě předpokládáme, že máme náhodný výběr z každé populace (skupiny) a že všechna pozorování jsou nezávislá.

Použitý matematický model můžeme zapsat podobně jako pro regresi:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad (12.1)$$

kde y_{ij} je j -té pozorování z i -té skupiny. Například y_{32} je druhé pozorování z třetí skupiny, tj. v našem případě 246 (viz tabulku [12.1](#)).

To znamená, že každé pozorované y je funkcí nějaké celkové hodnoty μ , skupinového efektu α_i , a blíže nspecifikované náhodné chyby ε_{ij} . Z toho plyne, že střední hodnota i -té skupiny je rovna $\mu_i = \mu + \alpha_i$. V analýze

rozptylu chceme zjistit, zda jsou skupinové efekty důležité, tj. zda existuje nějaký rozdíl mezi průměry jednotlivých skupin.

Výpočty analýzy rozptylu jsou dosti komplikované. Myšlenkou této techniky je, že rozdělíme celkový rozptyl závisle proměnné (všech 22 hodnot koncentrace soli kyseliny listové v červených krvinkách) do dvou částí, na variabilitu uvnitř skupin a variabilitu mezi skupinami.

Variabilita uvnitř skupin popisuje variabilitu v jednotlivých skupinách. V podstatě se tedy jedná o to, jak se každá hodnota ve skupině liší od skupinového průměru. Při pohledu na data z první skupiny zjistíme, že to znamená porovnat hodnoty 276, 280, 275, 291, 347, 354, 380 a 330 s hodnotou 316,6. Podobně porovnáme data v ostatních dvou skupinách. Protože máme tři skupiny, odvodíme celkový odhad variability v rámci skupiny podobným způsobem jako v případě sdruženého odhadu rozptylu v dvouvýběrovém t testu.

Variabilita mezi skupinami je funkcí, která ukazuje, jak se navzájem liší skupinové průměry. Zahrnuje porovnání všech tří skupinových průměrů 316,6, 256,4 a 275,6 s průměrem ze všech 22 pozorování, kterému se obvykle říká celkový průměr (v našem případě je roven 282,7).

Pokud skutečně neexistuje žádný rozdíl mezi skupinovými průměry těchto tří skupin, pak variabilita mezi skupinami i variabilita v rámci skupiny popisují stejný jev - stejný populační rozptyl. Toto porovnání variability v rámci skupiny a mezi skupinami se provádí pomocí **F testu**. Než přikročíme k dalšímu výkladu, podívejme se, jak může vypadat počítačový výstup této analýzy.

Tabulka 12.2: Počítačový výstup analýzy rozptylu

Zdroj variability	SS	df	$MS = SS/df$	F	p
Mezi skupinami	15663,48	2	7831,74	5,3005	0,0148
Uvnitř skupin (reziduální)	28073,30	19	1477,54		
Celkový	43736,78	21			

Tabulka 12.2 je typickou **tabulkou analýzy rozptylu**. Podívejme se na součet čtverců v prvním sloupečku (SS). Ve sloupci "zdroj variability" figurují tři položky. Pro účely analýzy rozptylu využijeme jako zdroj variability příslušnost ke skupině. Celková variabilita dat se zjišťuje pomocí celkového součtu čtverců S_T , zde tedy 43736,8. Celkový součet čtverců se spočte tak, že od každé z 22 hodnot odečteme celkový průměr (v našem případě 282,7), výsledné rozdíly umocníme na druhou a sečteme všech 22 takto získaných čtverců. Tedy

$$S_T = (276 - 282,7)^2 + (280 - 282,7)^2 + \dots + (328 - 282,7)^2 = 43636,78.$$

Tuto celkovou variabilitu rozdělíme na dvě části - variabilitu uvnitř skupin (reziduální) a variabilitu mezi skupinami. Součet čtverců pro variabilitu *mezi skupinami* (S_A) se vypočte jako vážený součet druhých mocnin rozdílů každého

skupinového průměru a celkového průměru. Reziduální součet čtverců (S_e), který je mírou variability *uvnitř* skupin je součtem druhých mocnin rozdílů hodnot a příslušného skupinového průměru. Platí tedy

$$S_T = S_A + S_e.$$

(Opravdu, $43736,78 = 15663,48 + 28073,30$.) Počet stupňů volnosti df se počítá následovně:

df_T pro celkový součet S_T je počet pozorování $- 1$,

df_A pro skupinový součet S_A je počet skupin $- 1$,

df_e pro reziduální součet S_e je počet pozorování $-$ počet skupin.

V našem případě je tedy $df_T = 22 - 1 = 21$, $df_A = 3 - 1 = 2$ a $df_e = 22 - 3 = 19$.

Vždy je dobré zkontrolovat pečlivě počty stupňů volnosti df , protože to zaručuje kontrolu získaných dat a také to, zda byl program použit správně (např. nelze mít nulový počet stupňů volnosti pro reziduální složku modelu).

Podíly $MS = SS / df$ jsou součty čtverců dělené odpovídajícím počtem stupňů volnosti. Tyto veličiny jsou mírou variability pro jednotlivé zdroje a ve statistických programech jsou označovány anglicky jako Mean Square (průměrné čtverce). Například pro variabilitu mezi skupinami vyjde:

$$MS = \frac{S_A}{df_A} = \frac{15663,48}{2} = 7831,74.$$

Testovou statistiku F dostaneme jako následující poměr

$$F = \frac{MS \text{ (mezi skupinami)}}{MS \text{ (uvnitř skupin)}} = \frac{S_A / df_A}{S_e / df_e}. \quad (12.2)$$

F je testová statistika, podobně jako t nebo χ^2 . K této statistice přísluší dva počty stupňů volnosti - pro čítelel, v tomto případě 2, a pro jmenovatel, v tomto případě 19.

Jako vždy předpokládáme, že nulová hypotéza tvrdí, že mezi populačními průměry není žádný rozdíl. Alternativou je, že aspoň dva populační průměry se od sebe liší. Za zmínku stojí, že nespecifikujeme, které průměry to mají být.

V případě platnosti hypotézy (všechny populační průměry shodné) teorie říká, že čítelel F statistiky bude (zhruba) stejný jako jmenovatel, který se podobně jako u regrese nazývá reziduální rozptyl. Pak by tedy hodnota F statistiky byla přibližně rovna jedné. Nyní musíme použít statistické tabulky, abychom se ujistili, zda hodnota F je významně větší než 1, což by ukazovalo, že MS mezi skupinami je významně větší než MS uvnitř skupin, a tedy že se průměry skupin liší. (Pokud by F statistika byla menší než 1, pak to znamená, že variabilita mezi skupinami může být dokonce menší než uvnitř skupin, a tedy tím spíše není důvod zamítnat nulovou hypotézu.) K výpočtu příslušných kritických hodnot i dosažených hladin významnosti lze využít i různé statistické programy.

Abychom mohli z tabulek vyčíst kritickou hodnotu pro F , potřebujeme znát oba počty stupňů volnosti, tedy v našem

případě:

ν_1 = počet stupňů volnosti pro čísel = 2,

ν_2 = počet stupňů volnosti pro jmenovatel = 19.

Kritická hodnota je rovna kvantilu $F_{1-\alpha}(\nu_1, \nu_2)$ rozdělení F o ν_1 a ν_2 stupních volnosti, který nalezneme v tabulce

[A.6](#). Pro $\alpha = 0,05$ je $F_{1-\alpha}(2, 19) = 3,52$.

Hodnoty větší než 3,52 naznačují, že nulová hypotéza neplatí, a hodnoty menší než 3,52 ukazují, že nemáme dostatečný důkaz, abychom mohli zamítnout nulovou hypotézu. Naše hodnota F statistiky se rovná 5,30, což dává dostatečný důkaz pro zamítnutí nulové hypotézy.

V počítačovém výstupu dostaneme dosaženou hladinu významnosti, tj. pravděpodobnost toho, že vyjde naše, nebo ještě extrémnější (významnější) hodnota testové statistiky za předpokladu, že nulová hypotéza platí. Tato dosažená hladina se nachází v posledním sloupečku tabulky [12.2](#). Protože $p = 0,0148$, což je méně než $\alpha = 0,05$, můžeme zamítnout nulovou hypotézu a učinit závěr, že průměrná hladina soli kyseliny listové není ve všech třech skupinách stejná. Výsledky analýzy rozptylu se v textu často zapisují následujícím způsobem: $F = 5,30$; $\nu_1 = 2$;

$\nu_2 = 19$; $p = 0,0148$.

Klíčovým bodem, o kterém se musíme zmínit, je, že tento výsledek nám neříká, které průměry se navzájem liší. Můžeme se podívat na skupinové průměry a zjistit, že skupina 1 má vyšší průměr než obě ostatní skupiny. V tuto chvíli ale nemůžeme říci, že tento průměr je *významně* vyšší. Musíme data analyzovat dále použitím metod mnohonásobného porovnávání, abychom zjistili, které průměry - pokud vůbec nějaké - se navzájem významně liší. Tyto metody budeme probírat podrobněji v části [12.3](#).

12.2 Ověřování předpokladů modelu

Podobně jako v případě regresní analýzy musíme provést testování předpokladů modelu. Většina statistických programů takový postup umožňuje. Můžeme vytvořit normální graf reziduí a graf předpovídaných hodnot proti reziduí. Abychom pochopili, jak získáme rezidua a předpovídané hodnoty, je nutné si připomenout matematický model, z něhož metoda analýzy rozptylu vychází. Hodnoty pozorovaných veličin můžeme vyjádřit takto:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

kde ε_{ij} jsou náhodné chybové složky, navzájem nezávislé.

Je zapotřebí si uvědomit, že toto je model pro populaci a my můžeme pouze odhadovat koeficienty (skutečné parametry) za použití dat, která máme ve výběru. Pro lepší pochopení modelu odhadněme každou z jeho složek. Následující postup je obvyklým, ne však jediným způsobem odvození odhadů. Stříška nad symbolem parametru znamená, že se jedná o odhad.

$$\begin{aligned}\hat{\mu} &= \text{celkový průměr} = 282,7, \\ \hat{\alpha}_1 &= \text{průměr první skupiny} - \text{celkový průměr} = \\ &= 316,6 - 282,7 = 33,9, \\ \hat{\alpha}_2 &= \text{průměr druhé skupiny} - \text{celkový průměr} = \\ &= 256,4 - 282,7 = -26,3, \\ \hat{\alpha}_3 &= \text{průměr třetí skupiny} - \text{celkový průměr} = \\ &= 275,6 - 282,7 = -7,1.\end{aligned}$$

Hodnoty označené $\hat{\alpha}_i$ jsou *odhady skupinových efektů* - tedy toho, jak se každý průměr liší od celkového průměru.

Předpídaná hodnota pro pozorování z i -té skupiny je průměr i -té skupiny $\hat{\mu}_i = \hat{\mu} + \hat{\alpha}_i$.

Například průměrná hodnota koncentrace soli kyseliny listové v červených krvinkách se v první skupině rovná $282,7 + 33,9 = 316,6$.

Rezidua pro každé pozorování spočteme jako rozdíl mezi pozorovanou hodnotou a předpídanou hodnotou.

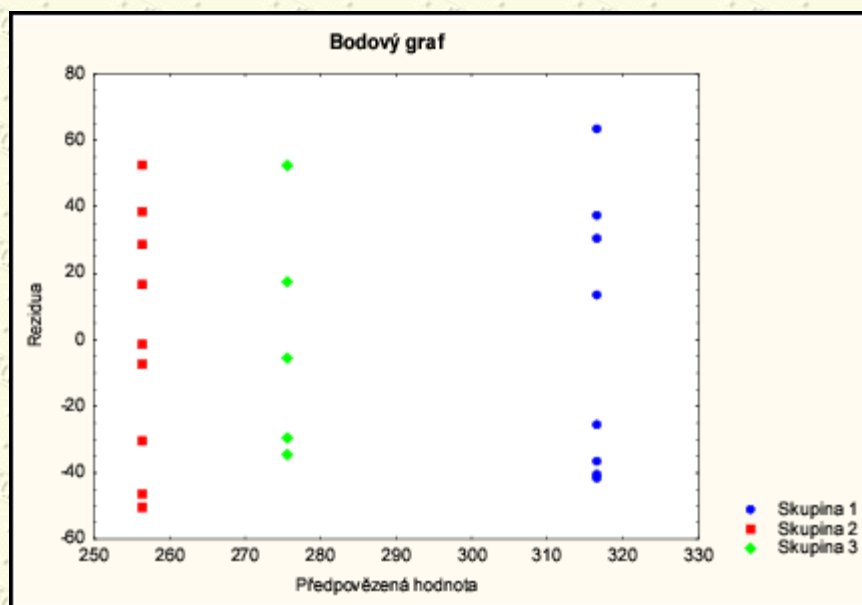
Například pro osm pozorovaných hodnot ve skupině 1 jsou rezidua následující (viz tabulku [12.3](#)).

Tabulka 12.3: Rezidua ve skupině 1

Pozorováno	Předpovězeno	Reziduum
276	316,6	-40,6
280	316,6	-36,6
275	316,6	-41,6
291	316,6	-25,6
347	316,6	+ 30,4
354	316,6	+ 37,4
380	316,6	+ 63,4
392	316,6	+ 75,4

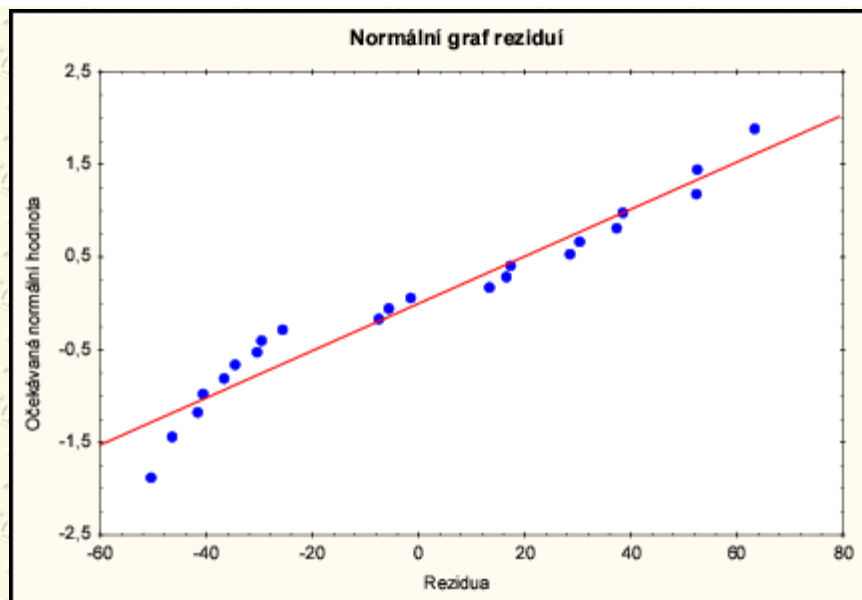
Pro testování předpokladů vytvoříme tytéž grafy jako při regresi. Graf předpovědovaných hodnot vs. pozorovaných hodnot (viz obrázek 12.2) by měl ukazovat "náhodnost" v tom smyslu, že nebude vidět žádná závislost mezi veličinami.

Obrázek 12.2: Předpovědané vs. pozorované hodnoty



Mají-li rezidua normální rozdělení, měl by normální graf vytvořit přímku. V našem případě (viz obrázek 12.3) je podoba normálního grafu přijatelná. Přítomnost jakýchkoli velkých odchylek by mohla znamenat doporučení transformace dat před provedením analýzy nebo nutnost provedení neparametrické verze testu. Všechny uvedené výpočty a grafy lze snadno získat pomocí počítače a vhodného programu.

Obrázek 12.3: Normální graf



12.3 Mnohonásobná porovnávání

Připomeňme si, že analýza rozptylu nám pouze říká, že průměry nejsou stejné. Je třeba provést další analýzu, abychom zjistili, jak se liší. Jednou z možností je porovnat každou dvojici průměrů, nebo dvojice, které nás zajímají. Jak jsme uvedli již dříve, problémem je skutečnost, že mnohonásobné testování významnosti dává vysokou pravděpodobnost, že bude nalezen významný rozdíl pouze náhodou. Řekněme například, že test má 5% možnost chybného pozitivního výsledku - to je hladina významnosti α , kterou určíme ještě před provedením analýzy. To znamená, že při opakovaném testování bychom chybně zamítli nulovou hypotézu v 5 % případů. Kdybychom měli čtyři skupiny a porovnali je navzájem tak, že bychom provedli všech šest testů, potom by pravděpodobnost, že dostaneme nejméně jednu chybný pozitivní výsledek (chyba prvního druhu), byla mnohem větší než 5 %.

Pro řešení tohoto problému mnohonásobného porovnávání existuje několik metod, jako například Bonferroniho, Tukeyova, Newman-Keulsova, Duncanova, Fisherovo LSD (nejmenší významný rozdíl - Least Significant Difference) a Scheffého. Úkolem každé metody je udržet danou hladinu pravděpodobnosti chyby prvního druhu (5 %) a v podstatě ji rozdělit mezi všechna porovnání. Některé z těchto testů jsou velmi konzervativní. Může se stát, že F test zamítne hypotézu o rovnosti průměrů, a přitom žádná dvojice průměrů se od sebe podle výsledků metod mnohonásobného porovnávání navzájem významně neliší!

Před vlastním provedením testu předpokládáme, že skupiny nejsou uspořádány. Příkladem uspořádané skupiny je situace, kdy je skupina definována věkem nebo množstvím léku (např. 5 mg, 10 mg, 20 mg). Nejprve bychom se ještě před provedením analýzy měli rozhodnout, o které skupiny se z hlediska porovnávání zajímáme. To bude záležet na tom, jaký problém výzkum řeší. Čím menší počet porovnání, tím lépe. Při určování toho, jaké skupiny budete porovnávat, se nerozhodujte až při pohledu na výsledky!

Ukážeme si **Bonferroniho metodu**, protože je nejsnadněji pochopitelná. Pro ta porovnání, která nás zajímají, provedeme modifikované t testy s *upravenou hladinou významnosti*. Tu získáme tak, že hladinu α jednoduše vydělíme celkovým počtem porovnání, která chceme provést. Tato hodnota pak bude naší hladinou významnosti

pro každý t test. Řekněme, že pro náš příklad chceme provést všechna možná porovnání - pro tři skupiny existují tři. Naše hladina významnosti pro každé porovnání nebude tedy 5 %, ale $(5/3) \% = 1,67 \%$. Nulová a alternativní hypotéza jsou stejné jako pro obyčejný t test.

V tomto případě se t testy nepatrně liší od těch, o kterých jsme již hovořili. Připomeňme si, že testová statistika t testu se počítala následujícím způsobem:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

kde \bar{x}_1, \bar{x}_2 jsou výběrové průměry v první a druhé skupině a s^2 je sdružený odhad rozptylu

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$$

kde n_1 a n_2 jsou rozsahy výběrů a s_1^2 a s_2^2 výběrové rozptyly v první a druhé skupině.

V případě upraveného testu zůstává čitatel t statistiky stejný a mění se jen jmenovatel. Namísto počítání sdruženého rozptylu ze dvou skupin použijeme sdruženou verzi ze všech skupin, včetně těch, které nepoužíváme při porovnávání. Odhad sdruženého rozptylu ze všech skupin je dán reziduálním rozptylem v tabulce analýzy rozptylu. Upravená statistika pro t test má pak tedy podobu

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_e}{df_e} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (12.3)$$

a za platnosti nulové hypotézy ($\mu_1 = \mu_2$) má t rozdělení s df_e (tj. v našem případě 19) stupni volnosti. Kritická hodnota pro t je také jiná. Uvědomme si, že pro každé porovnání se hladina významnosti rovná 1,67 % a nikoli 5 %. K výpočtu kritické hodnoty potřebujeme (pokud nepoužíváme počítač) velmi dobré tabulky tohoto rozdělení. Pokud použijeme počítač, lze spočítat dosaženou hladinu významnosti neboli pravděpodobnost, že získáme naše, nebo ještě silnější výsledky za předpokladu platnosti nulové hypotézy. Potom můžeme tuto hodnotu porovnat s 1,67 % neboli 0,0167. Je-li dosažená hladina významnosti p menší než 0,0167, zamítneme hypotézu o rovnosti μ_1 a μ_2 .

Výsledky srovnání jednotlivých skupin jsou následující.

1.

Srovnání první a druhé skupiny:

$$t = \frac{316,6 - 256,4}{\sqrt{1477,54 \left(\frac{1}{8} + \frac{1}{9} \right)}} = 3,22, \quad p < 0,0167.$$

2. Srovnání první a třetí skupiny:

$$t = \frac{316,6 - 275,6}{\sqrt{1477,54\left(\frac{1}{8} + \frac{1}{5}\right)}} = 1,87, \quad p > 0,0167.$$

3. Srovnání druhé a třetí skupiny:

$$t = \frac{256,4 - 275,6}{\sqrt{1477,54\left(\frac{1}{9} + \frac{1}{5}\right)}} = -0,90, \quad p > 0,0167.$$

Výše uvedené výsledky ukazují, že existuje významný rozdíl mezi průměry skupin 1 a 2, ale ne mezi ostatními porovnávanými průměry. Jinými slovy pacienti, kteří dostávali směs 50 % oxidu dusného a 50 % kyslíku nepřetržitě po dobu 24 hodin, měli v průměru významně vyšší hladiny koncentrace soli kyseliny listové v červených krvinkách než pacienti, kteří tuto směs dostávali pouze během operace. ◇

12.4 Obecné poznámky k analýze rozptylu

Postup analýzy rozptylu můžeme shrnout následovně:

1. Vytvořme graf dat - tím si pomůžeme zkontrolovat přítomnost odlehklých hodnot.
2. Vytvořme souhrnné (popisné) statistiky - vhodné pro orientaci (které průměry se mohou lišit, zda platí předpoklad o shodnosti rozptylů a podobně).
3. Spočtme analýzu rozptylu (ANOVA).
4. Zkontrolujme předpoklady - nechme si vykreslit graf předpovídaných hodnot proti pozorovaným hodnotám a normální graf reziduí.
5. Pokud je F statistika v analýze rozptylu významná, provedme mnohonásobná porovnávání.

Regrese a analýza rozptylu

Regrese a analýza rozptylu by neměly být považovány za dvě různé techniky. ANOVA je v podstatě součástí (speciálním případem) regrese. Všechny výsledky analýzy rozptylu lze získat regresní analýzou. V obou situacích model vytváříme a upravujeme podle dat, která máme k dispozici. V regresi se zabýváme následujícím modelem *lineární závislosti Y na X*:

$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (12.4)$$

Náš model analýzy rozptylu má tvar:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}.$$

Podle tohoto modelu je každé pozorování součtem celkového průměru, specifického efektu dané skupiny a chybové složky.

V analýze rozptylu tedy vyšetřujeme závislost veličiny Y na nominálním znaku. V regresi zpravidla vyšetřujeme závislost Y na spojitě veličině X - v našem případě uvažujeme lineární závislost. Ilustrujme tento pojem na příkladu.

Příklad 12.2 Tabulka [12.4](#) ukazuje průměr a směrodatnou odchylku hladin koncentrace trypsinu v séru u zdravých jedinců. Kategorizace je provedena podle věku.

Tabulka 12.4:
Koncentrace trypsinu
v séru

Věk	n	\bar{y}	s
10-19	32	128	50,9
20-29	137	152	58,5
30-39	38	194	49,3
40-49	44	207	66,3
50-59	16	215	60,0

Analýza rozptylu by zjišťovala, zda existuje významný rozdíl mezi průměrnou hladinou trypsinu v séru mezi jednotlivými věkovými skupinami, ale nezavedla by žádný model pro závislost na věku. Model lineární regrese ([12.4](#)) předpokládá lineární vztah mezi věkem a hladinou trypsinu v séru. Z teorie by mohlo např. plynout, že hladina roste s věkem. Pokud je tato úvaha správná, byl by model vytvořený regresí mnohem informativnější a silnější, než řešení pomocí ANOVA. Pro regresní analýzu by byly použity jednotlivé hodnoty věku. \diamond

V případě oboustranné alternativy je při porovnávání pouze dvou skupin analýza rozptylu zcela shodná s t testem. Platí totiž, že $F = t^2$.

Neparametrickým ekvivalentem analýzy rozptylu je Kruskalův-Wallisův test. Ten bychom použili buď v případě výrazné nenormality dat, nebo tehdy, pokud bychom měli dobrý důvod předpokládat přítomnost hrubých chyb v datech. Pak by totiž mohly být velmi vychýlené samotné průměry a výrazně (a přitom "neodůvodněně") by vzrostla variabilita uvnitř skupin, což by analýzu rozptylu podstatným způsobem ovlivnilo.

V příkladu [12.1](#) byly případy klasifikovány pouze podle jediné veličiny - příslušnosti ke skupině. Případy lze samozřejmě klasifikovat podle dvou (nebo i více) veličin a vytvořit tak analýzu rozptylu, která zjišťuje účinek dvou

veličin na objasnění celkového rozptylu závislé veličiny. Této metodě se říká analýza rozptylu dvojného třídění. Rovněž lze analyzovat data, kdy po čase jednotlivá měření zopakujeme. Technika pro analýzu takových dat se nazývá analýza rozptylu opakovaných pozorování, přičemž výpočet se provádí jiným způsobem než u analýzy rozptylu jednoduchého třídění.

[Další](#) [Výš](#) [Předchozí](#) [Obsah](#) [Rejstřík](#)

Další: [13. Statistické tabulky](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [11. Měření závislosti](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [Literatura](#) **Výš:** [Základy statistiky pro biomedicinské obory](#) **Předchozí:** [12. Analýza rozptylu](#)

13. Statistické tabulky

Statistika je fyzika čísel.

P. Diaconis

Tabulka A.1: Kvantily
standardizovaného normálního
rozdělení $N(0,1)$

α	0,1	0,05	0,01	0,001
$z_{1-\frac{\alpha}{2}}$	1,645	1,960	2,576	3,290
$z_{1-\alpha}$	1,282	1,645	2,326	3,090
z_{α}	-1,282	-1,645	-2,326	-3,090

Tabulka A.2: Distribuční funkce
standardizovaného normálního rozdělení
 $N(0,1)$

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
-3,50	0,0002	-1,00	0,1587	1,10	0,8643
-3,40	0,0003	-0,90	0,1841	1,20	0,8849
-3,30	0,0005	-0,80	0,2119	1,30	0,9032
-3,20	0,0007	-0,70	0,2420	1,40	0,9192
-3,10	0,0010	-0,60	0,2743	1,50	0,9332
-3,00	0,0013	-0,50	0,3085	1,60	0,9452
-2,90	0,0019	-0,40	0,3446	1,70	0,9554
-2,80	0,0026	-0,30	0,3821	1,80	0,9641

-2,70	0,0035	-0,20	0,4207	1,90	0,9713
-2,60	0,0047	-0,15	0,4404	2,00	0,9772
-2,50	0,0062	-0,10	0,4602	2,10	0,9821
-2,40	0,0082	-0,05	0,4801	2,20	0,9861
-2,30	0,0107	0,00	0,5000	2,30	0,9893
-2,20	0,0139	0,05	0,5199	2,40	0,9918
-2,10	0,0179	0,10	0,5398	2,50	0,9938
-2,00	0,0228	0,15	0,5596	2,60	0,9953
-1,90	0,0287	0,20	0,5793	2,70	0,9965
-1,80	0,0359	0,30	0,6179	2,80	0,9974
-1,70	0,0446	0,40	0,6554	2,90	0,9981
-1,60	0,0548	0,50	0,6915	3,00	0,9987
-1,50	0,0668	0,60	0,7257	3,10	0,9990
-1,40	0,0808	0,70	0,7580	3,20	0,9993
-1,30	0,0968	0,80	0,7881	3,30	0,9995
-1,20	0,1151	0,90	0,8159	3,40	0,9997
-1,10	0,1357	1,00	0,8413	3,50	0,9998

Tabulka A.3: Kvantily $t_{1-\frac{\alpha}{2}}(df)$
Studentova t rozdělení o df stupních
volnosti

$t_{1-\frac{\alpha}{2}}(df)$	α		
	0,05	0,01	0,001
df			
1	12,706	63,656	636,578
2	4,303	9,925	31,600
3	3,182	5,841	12,924
4	2,776	4,604	8,610
5	2,571	4,032	6,869
6	2,447	3,707	5,959

13. Statistické tabulky

7	2,365	3,499	5,408
8	2,306	3,355	5,041
9	2,262	3,250	4,781
10	2,228	3,169	4,587
11	2,201	3,106	4,437
12	2,179	3,055	4,318
13	2,160	3,012	4,221
14	2,145	2,977	4,140
15	2,131	2,947	4,073
16	2,120	2,921	4,015
17	2,110	2,898	3,965
18	2,101	2,878	3,922
19	2,093	2,861	3,883
20	2,086	2,845	3,850
21	2,080	2,831	3,819
22	2,074	2,819	3,792
23	2,069	2,807	3,768
24	2,064	2,797	3,745
25	2,060	2,787	3,725
30	2,042	2,750	3,646
35	2,030	2,724	3,591
40	2,021	2,704	3,551
45	2,014	2,690	3,520
50	2,009	2,678	3,496
60	2,000	2,660	3,460
70	1,994	2,648	3,435
80	1,990	2,639	3,416
90	1,987	2,632	3,402
100	1,984	2,626	3,390
∞	1,960	2,576	3,290

Tabulka A.4: Kvantily $t_{1-\alpha}(df)$

Studentova t rozdělení o df
stupních volnosti

$t_{1-\alpha}(df)$	α		
	0,05	0,01	0,001
1	6,314	31,821	318,289
2	2,920	6,965	22,328
3	2,353	4,541	10,214
4	2,132	3,747	7,173
5	2,015	3,365	5,894
6	1,943	3,143	5,208
7	1,895	2,998	4,785
8	1,860	2,896	4,501
9	1,833	2,821	4,297
10	1,812	2,764	4,144
11	1,796	2,718	4,025
12	1,782	2,681	3,930
13	1,771	2,650	3,852
14	1,761	2,624	3,787
15	1,753	2,602	3,733
16	1,746	2,583	3,686
17	1,740	2,567	3,646
18	1,734	2,552	3,610
19	1,729	2,539	3,579
20	1,725	2,528	3,552
21	1,721	2,518	3,527
22	1,717	2,508	3,505
23	1,714	2,500	3,485
24	1,711	2,492	3,467

25	1,708	2,485	3,450
30	1,697	2,457	3,385
35	1,690	2,438	3,340
40	1,684	2,423	3,307
45	1,679	2,412	3,281
50	1,676	2,403	3,261
60	1,671	2,390	3,232
70	1,667	2,381	3,211
80	1,664	2,374	3,195
90	1,662	2,368	3,183
100	1,660	2,364	3,174
∞	1,645	2,326	3,090

Tabulka A.5: Kvantily $\chi^2_{1-\alpha}$

rozdělení χ^2 o df stupních volnosti

$\chi^2_{1-\alpha}(df)$	α		
	0,05	0,01	0,001
df			
1	3,84	6,63	10,83
2	5,99	9,21	13,82
3	7,81	11,34	16,27
4	9,49	13,28	18,47
5	11,07	15,09	20,51
6	12,59	16,81	22,46
7	14,07	18,48	24,32
8	15,51	20,09	26,12
9	16,92	21,67	27,88
10	18,31	23,21	29,59
11	19,68	24,73	31,26

12	21,03	26,22	32,91
13	22,36	27,69	34,53
14	23,68	29,14	36,12
15	25,00	30,58	37,70
16	26,30	32,00	39,25
17	27,59	33,41	40,79
18	28,87	34,81	42,31
19	30,14	36,19	43,82
20	31,41	37,57	45,31
21	32,67	38,93	46,80
22	33,92	40,29	48,27
23	35,17	41,64	49,73
24	36,42	42,98	51,18
25	37,65	44,31	52,62
30	43,77	50,89	59,70
35	49,80	57,34	66,62
40	55,76	63,69	73,40
45	61,66	69,96	80,08
50	67,50	76,15	86,66
60	79,08	88,38	99,61
70	90,53	100,43	112,32
80	101,88	112,33	124,84
90	113,15	124,12	137,21
100	124,34	135,81	149,45

Tabulka A.6: 95% kvantily F rozdělení o ν_1 (čítatel) a ν_2 (jmenovatel) stupních volnosti

$F_{0,95}(\nu_1, \nu_2)$	ν_1							
ν_2	1	2	3	4	5	6	8	10

13. Statistické tabulky

1	161,45	199,50	215,71	224,58	230,16	233,99	238,88	241,88
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,85	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,85	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,77	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,37	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,34	2,24
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,16
35	4,12	3,27	2,87	2,64	2,49	2,37	2,22	2,11
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,08
45	4,06	3,20	2,81	2,58	2,42	2,31	2,15	2,05
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	2,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,10	1,99

70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,97
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,95
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,94
100	3,94	3,09	2,70	2,46	2,31	2,19	2,03	1,93

Tabulka A.7: Kritické hodnoty párového
Wilcoxonova testu

n	α		n	α		n	α	
	0,05	0,01		0,05	0,01		0,05	0,01
6	0	-	26	98	75	46	361	307
7	2	-	27	107	83	47	378	322
8	3	0	28	116	91	48	396	339
9	5	1	29	126	100	49	415	355
10	8	3	30	137	109	50	434	373
11	10	5	31	147	118	51	453	390
12	13	7	32	159	128	52	473	408
13	17	9	33	170	138	53	494	427
14	21	12	34	182	148	54	514	445
15	25	15	35	195	159	55	536	465
16	29	19	36	208	171	56	557	484
17	34	23	37	221	182	57	579	504
18	40	27	38	235	194	58	602	525
19	46	32	39	249	207	59	625	546
20	52	37	40	264	220	60	648	567
21	58	42	41	279	233	61	672	589
22	65	48	42	294	247	62	697	611
23	73	54	43	310	261	63	721	634
24	81	61	44	327	276	64	747	657
25	89	68	45	343	291	65	772	681

Tabulka A.8: 5% kritické hodnoty Mannova-Whitneyova testu

	n_2																		
n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
4	-	-	0																
5	-	0	1	2															
6	-	1	2	3	5														
7	-	1	3	5	6	8													
8	0	2	4	6	8	10	13												
9	0	2	4	7	10	12	15	17											
10	0	3	5	8	11	14	17	20	23										
11	0	3	6	9	13	16	19	23	26	30									
12	1	4	7	11	14	18	22	26	29	33	37								
13	1	4	8	12	16	20	24	28	33	37	41	45							
14	1	5	9	13	17	22	26	31	36	40	45	50	55						
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64					
16	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75				
17	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87			
18	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99		
19	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	
20	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
21	3	8	15	22	29	36	43	50	58	65	73	80	88	96	103	111	119	126	134
22	3	9	16	23	30	38	45	53	61	69	77	85	93	101	109	117	125	133	141
23	3	9	17	24	32	40	48	56	64	73	81	89	98	106	115	123	132	140	149
24	3	10	17	25	33	42	50	59	67	76	85	94	102	111	120	129	138	147	156
25	3	10	18	27	35	44	53	62	71	80	89	98	107	117	126	135	145	154	163
26	4	11	19	28	37	46	55	64	74	83	93	102	112	122	132	141	151	161	171
27	4	11	20	29	38	48	57	67	77	87	97	107	117	127	137	147	158	168	178
28	4	12	21	30	40	50	60	70	80	90	101	111	122	132	143	154	164	175	186
29	4	13	22	32	42	52	62	73	83	94	105	116	127	138	149	160	171	182	193
30	5	13	23	33	43	54	65	76	87	98	109	120	131	143	154	166	177	189	200

Tabulka A.9: 1% kritické hodnoty Mannova-Whitneyova testu

	n_2																			
n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
5	-	-	-	0																
6	-	-	0	1	2															
7	-	-	0	1	3	4														
8	-	-	1	2	4	6	7													
9	-	0	1	3	5	7	9	11												
10	-	0	2	4	6	9	11	13	16											
11	-	0	2	5	7	10	13	16	18	21										
12	-	1	3	6	9	12	15	18	21	24	27									
13	-	1	3	7	10	13	17	20	24	27	31	34								
14	-	1	4	7	11	15	18	22	26	30	34	38	42							
15	-	2	5	8	12	16	20	24	29	33	37	42	46	51						
16	-	2	5	9	13	18	22	27	31	36	41	45	50	55	60					
17	-	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70				
18	-	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81			
19	0	3	7	12	17	22	28	33	39	45	51	57	63	69	74	81	87	93		
20	0	3	8	13	18	24	30	36	42	48	54	60	67	73	79	86	92	99	105	
21	0	3	8	14	19	25	32	38	44	51	58	64	71	78	84	91	98	105	112	
22	0	4	9	14	21	27	34	40	47	54	61	68	75	82	89	96	104	111	118	
23	0	4	9	15	22	29	35	43	50	57	64	72	79	87	94	102	109	117	125	
24	0	4	10	16	23	30	37	45	52	60	68	75	83	91	99	107	115	123	131	
25	0	5	10	17	24	32	39	47	55	63	71	79	87	96	104	112	121	129	138	
26	0	5	11	18	25	33	41	49	58	66	74	83	92	100	109	118	127	135	144	
27	1	5	12	19	27	35	43	52	60	69	78	87	96	105	114	123	132	142	151	
28	1	5	12	20	28	36	45	54	63	72	81	91	100	109	119	128	138	148	157	
29	1	6	13	21	29	38	47	56	66	75	85	94	104	114	124	134	144	154	164	
30	1	6	13	22	30	40	49	58	68	78	88	98	108	119	129	139	150	160	170	

Tabulka A.10: Kritické hodnoty Spearmanova korelačního koeficientu

n	α		n	α		n	α	
	0,05	0,01		0,05	0,01		0,05	0,01
			11	0,6091	0,7545	21	0,4351	0,5545
			12	0,5804	0,7273	22	0,4241	0,5426
			13	0,5549	0,6978	23	0,4150	0,5306
			14	0,5341	0,6747	24	0,4061	0,5200
5	0,9000	-	15	0,5179	0,6536	25	0,3977	0,5100
6	0,8286	0,9429	16	0,5000	0,6324	26	0,3894	0,5002
7	0,7450	0,8929	17	0,4853	0,6152	27	0,3822	0,4915
8	0,6905	0,8571	18	0,4716	0,5975	28	0,3749	0,4828
9	0,6833	0,8167	19	0,4579	0,5825	29	0,3685	0,4744
10	0,6364	0,7818	20	0,4451	0,5684	30	0,3620	0,4665

[Další](#)
[Výš](#)
[Předchozí](#)
[Obsah](#)
[Rejstřík](#)

Další: [Literatura](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [12. Analýza rozptylu](#)

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Další: [Rejstřík](#) **Výš:** [Základy statistiky pro biomedicínské obory](#) **Předchozí:** [13. Statistické tabulky](#)

Literatura

- 1 Altman, D. G.: *Practical Statistics for Medical Research*. 2nd edition, Chapman & Hall, London, 1994
- 2 Anděl, J.: *Statistické metody*. 2. vydání, Matfyzpress, Praha, 1998
- 3 Colton, T.: *Statistics in Medicine*. Little, Brown and Company, Boston, 1974
- 4 Dawson-Saunders, B., Trapp, R. G.: *Basic and Clinical Biostatistics*. Prentice-Hall, International Inc., London, 1994
- 5 Fienberg, S. E.: *The Analysis of Cross-classified Categorical Data*. M.I.T., 1976
- 6 Havránek, T.: *Statistika pro biologické a lékařské vědy*. Academia, Praha, 1993
- 7 Josíčko, M.: *Pravděpodobnost a matematická statistika pro biology*. SPN, Praha, 1969
- 8 Komenda, S.: *Biometrie*. Vydavatelství Univerzity Palackého v Olomouci, Olomouc, 1994
- 9 Kubánková, V., Hendl, J.: *Statistika pro zdravotníky*. Avicenum, Praha, 1986
- 10 Lepš, J.: *Biostatistika*. Jihočeská univerzita České Budějovice, České Budějovice, 1996
- 11 Likeš, J., Laga, J.: *Základní statistické tabulky*. SNTL, Praha, 1978
- 12 Likeš, J., Machek, J.: *Matematická statistika*. SNTL, Praha, 1983
- 13 O'Regan, M., Kelly, A.: *Introduction to Medical Statistics*. In: Statistics in Clinics and Epidemiology I., EuroMISE Center of Charles University and Academy of Sciences, Prague, 1997
- 14 Skalská, H., Stránský, P.: *Základy biostatistiky*. Univerzita Karlova, Praha, 1994

- 15 Sprent, P.: *Applied Nonparametric Statistical Methods*. 2nd edition, Chapman & Hall, London, 1993
- 16 Swoboda, H.: *Moderní statistika*. Nakladatelství Svoboda, Praha, 1977
- 17 Victor, N., Holle, R.: *Basic Principles of Testing in Clinical Research*. In: Statistics in Clinics and Epidemiology II., EuroMISE Center of Charles University and Academy of Sciences, Prague, 1997
- 18 *Výběrové šetření o zdravotním stavu české populace HIS CR 93*. Ústav zdravotnických informací a statistiky České republiky, 1995
- 19 Wassertheil-Smoller, S.: *Biostatistics and Epidemiology*. Springer Verlag, New York, 1995
- 20 *Zdraví pro všechny do roku 2000*. Ústav zdravotnických informací a statistiky České republiky, 1994
- 21 Zvára, K., Prášková, Z.: *Pravděpodobnost a matematická statistika*. SPN, Praha, 1989
- 22 Zvára, K., Štěpán, J.: *Pravděpodobnost a matematická statistika*. Matfyzpress, Praha, 1997
- 23 Zvárová, J.: *Statistické metody v sociálním lékařství*. SPN, Praha, 1982

Tato stránka byla naposledy změněna dne 5. ledna 1999.

Výš: [Základy statistiky pro biomedicínské obory](#) Předchozí: [Rejstřík](#)

O tomto dokumentu ...

Tento dokument byl vytvořen překladačem [LaTeX₂HTML](#) verze 98.1p1 release (March 2nd, 1998).

Copyright © 1993, 1994, 1995, 1996, 1997, [Nikos Drakos](#), Computer Based Learning Unit, University of Leeds.

Tato stránka byla naposledy změněna dne 5. ledna 1999.