

Statistické metody v biologii a medicíně II



Opakování

- deskriptivní statistika
 - prezentace dat
- statistická indukce
 - závěry o populacích na základě vzorků
- statistická inference (odhad)
 - testování hypotéz

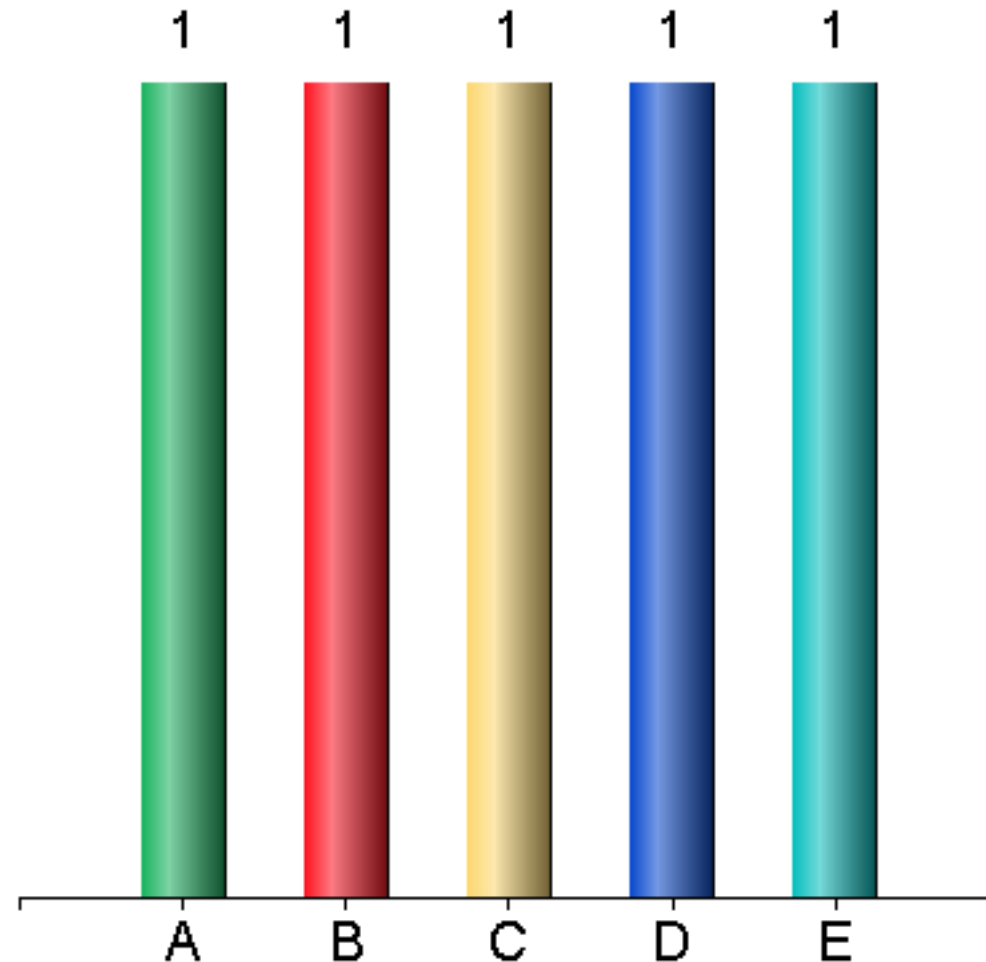
Opakování – druhy dat

- Spojitá (vždy kvantitativní) – parametr může teoreticky nabýt jakékoliv hodnoty v určitém intervalu (např. koncentrace glukózy: $0-\infty$; ejekční frakce: 0-100%)
 - Poměrová vs. intervalová data – lze definovat rozdíl, ale ne poměr mezi dvěma hodnotami (např. IQ)
- Kategoriální (diskrétní – obvykle kvalitativní) – parametr může nabývat pouze určených hodnot (např. krevní skupina: 0, A, B, AB; pohlaví: muž, žena; choroba je/není přítomna)
 - Ordinální data – jsou kategoriálního charakteru, ale kvantitativní (lze je seřadit – např. stupeň srdečního selhání – NYHA I-IV)
 - Diskrétní data – lze je seřadit a tvoří lineárně rostoucí řadu (mezi jednotlivými kategoriemi jsou stejné rozestupy – např. počet dětí v rodině: 0,1,2...) – často se testují jako spojitá data
 - Binární data – pouze dvě možnosti (zdravý/nemocný)



Stupeň dosaženého vzdělání představuje data...

- A. Ordinální
- B. Intervalová
- C. Binární
- D. Spojitá
- E. Kvalitativní



Formulace statistických hypotéz

- Výzkumná hypotéza (např. léčivo A má lepší efekt než léčivo B, krevní tlak při zavedené léčbě klesá, existuje vztah mezi pohlavím a tělesnou výškou atd...) – lze formulovat jak pro experiment, tak i v rámci pozorování
- Statistické testování výzkumné hypotézy – v podstatě důkaz sporem
- Pro účely testování definujeme **nulovou hypotézu H_0** (tj. např. mezi skupinami není rozdíl v průměrech, není rozdíl v rozptylech, mezi dvěma parametry není vztah, parametr se v čase nemění...resp. rozdíly jsou dány náhodou)
- Tuto nulovou hypotézu se snažíme vyvrátit (resp. prokázat, že je vysoce nepravděpodobná)
- V případě neplatnosti nulové hypotézy platí její negace – **alternativní hypotéza H_A**
- Výsledek testu hypotéz tedy zní:
 - A) nezamítnutí nulové hypotézy (na určité hladině statistické významnosti α)
 - B) zamítnutí nulové hypotézy ve prospěch alternativní

Chyby v testování hypotéz

	Skutečná povaha nulové hypotézy	
Statistické rozhodnutí	H_0 platí	H_0 neplatí
Zamítnutí H_0	chyba I. typu (α)	Správně ($1-\beta$)
Potvrzení H_0	Správně ($1-\alpha$)	chyba II. typu (β)

- Pravděpodobnost chyby I. typu (α) – také **hladina statistické významnosti**
- α se definuje již před testováním – v biomedicíně je zvykem obvykle 0,05 (tj. pokud je H_0 zamítnuta, pak si můžeme být na 95% jisti, že skutečně neplatí a pozorovaný rozdíl/vztah je skutečný)
- $1-\beta$ – také **síla testu**
- Statistická významnost p – pravděpodobnost, že jsme pozorovaný výsledek získali při platnosti nulové hypotézy
- Je-li $p < \alpha$, zamítáme na dané hladině významnosti nulovou hypotézu a platí tedy hypotéza alternativní
- Říkáme, že rozdíl (efekt) je **statisticky signifikantní** (což samozřejmě nemusí znamenat, že je signifikantní – významný – i prakticky)

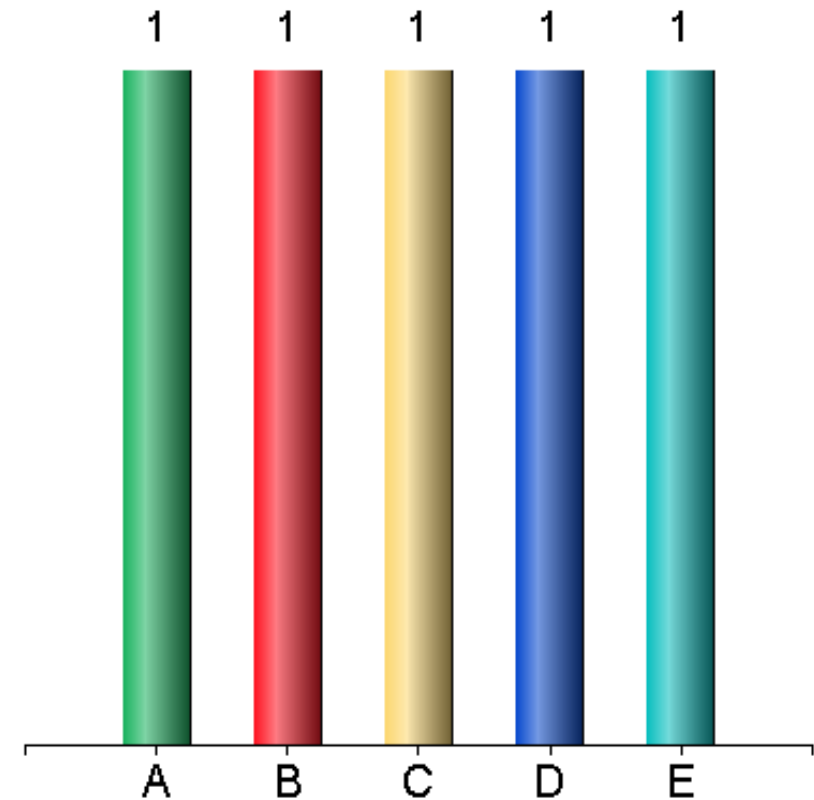
Statistické testy

- Pro různé druhy statistických hypotéz používáme různé statistické testy
- Použití jednotlivých testů se řídí:
 - počtem srovnávaných skupin
 - charakterem dat (kategorická vs. spojitá)
 - rozložením dat
 - vzájemnou závislostí dat



Síla testu...

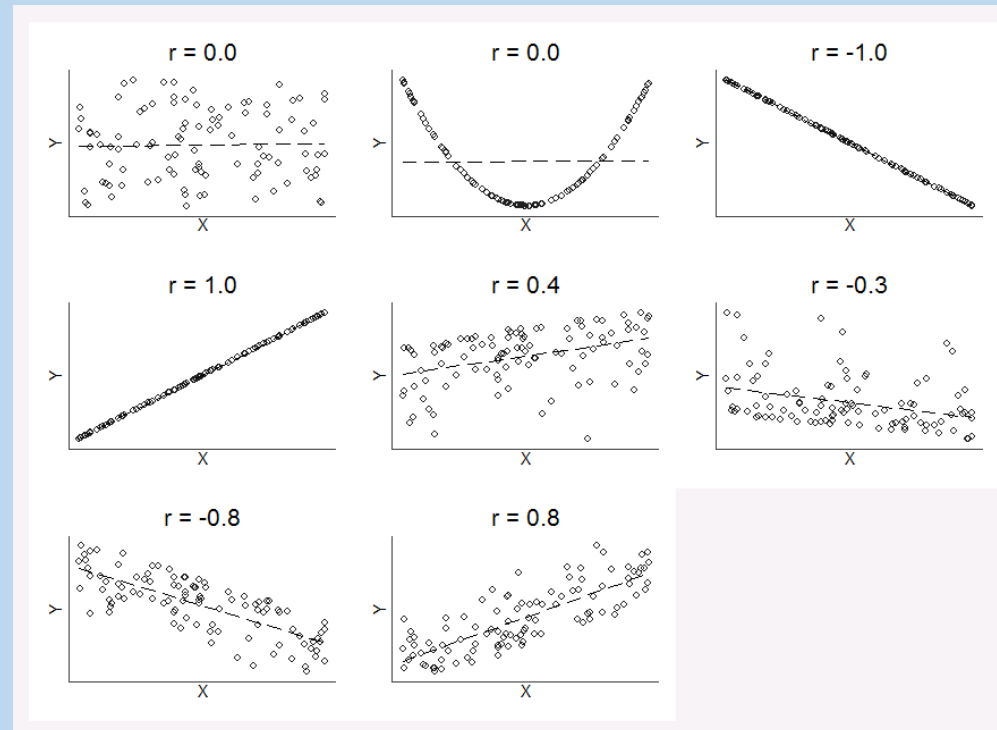
- A. Je vyjádřením jeho praktické (ne statistické) významnosti
- B. Roste se zvětšující se variabilitou dat
- C. Udává schopnost testu oprávněně zamítnout nulovou hypotézu
- D. Se značí písmenem p
- E. Je pravděpodobnost, že při zamítnutí nulové hypotézy bude platit hypotéza alternativní



Vzájemné vztahy parametrů u jednoho výběru

- Vztah dvou spojitých proměnných – korelace
- Vyjádřením je korelační koeficient (r)
- r obecně vyjadřuje míru závislosti obou proměnných
- Může dosahovat hodnot od -1 do 1, kde 0 značí žádnou závislost, 1 značí 100% kladnou závislost (s jedním faktorem stoupá druhý) a -1 naprostou zápornou závislost

- Pozor – některé nelineární vztahy se do korelačního koeficientu nemusí promítnout!



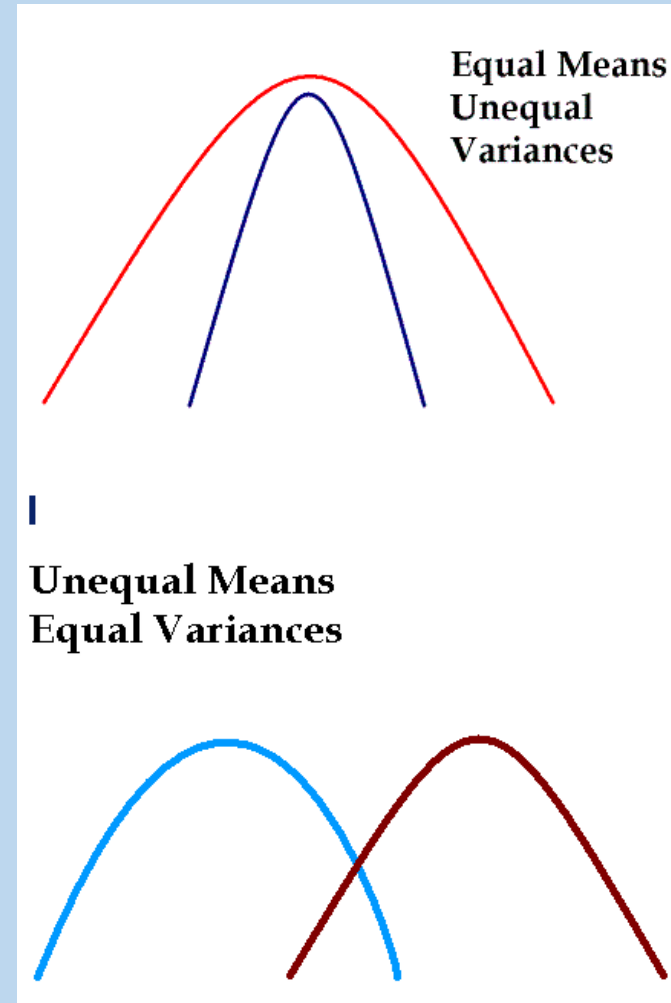
- Vztahy kategoriálních proměnných – viz srovnávání více výběrů (jednotlivé výběry jsou v podstatě kategoriálními proměnnými definovány)

Příklady korelačních koeficientů

- Pearsonův koeficient (parametrický) – hodnotí lineární závislost mezi proměnnými
 - Hlavní podmínkou je přibližně normální rozložení dat
- Spearmanův koeficient (neparametrický) – hodnotí závislost v pořadí hodnot obou proměnných
- Ani jeden z koeficientů ovšem neodhalí závislost např. u křivky ve tvaru U
- Mimo korelačního koeficientu lze určit i odpovídající p-hodnotu (čili pravděpodobnost, že je pozorovaná korelace dána náhodou)

Porovnávání spojité proměnné u dvou a více výběrů

- H_0 – mezi oběma výběry není rozdíl v hodnotě spojité proměnné (resp. je dán náhodou - např. koncentrace glykovaného hemoglobinu u léčených a neléčených diabetiků se neliší)
- Obecně lze testovat shodu polohy (častěji, viz dále) nebo variability (např. F-test, Levenův test)



Parametrické vs. neparametrické testy

Parametrické

- Pracují s hodnotami
- Mají vyšší sílu, ale pouze za splnění předpokladů (hl. normální rozložení dat v každém výběru)
- Není-li rozložení dat normální, můžeme se pokusit je transformovat (normalizovat)

Neparametrické

- Pracují s pořadím hodnot
- Síla je poněkud nižší (ovšem u velkých souborů se rozdíl stírá)
- Jsou „robustnější“ – nejsou na rozložení dat tolik závislé
- Lze je použít i u diskretních dat

Normalitu je možno testovat testy normality (např. Kolmogorov-Smirnov, Shapiro-Wilks – porovnávají skutečné rozložení s normálním) a „okometrickým“ zhodnocením, zda histogramy odpovídají Gaussově křivce, u malých souborů vhodnější normal probability plot

Párové vs. nepárové testy

Párové

- Používané tehdy, když každé hodnotě z výběru A můžeme přiřadit hodnotu z výběru B, která se liší právě jen příslušností k výběru (např. platy v nemocnici: ředitel A – ředitel B; primář A – primář B... až uklízečka A – uklízečka B)
- **Nejčastěji pro vývoj souboru v čase** (hmotnost pacientů teď a po 5 letech: pacient XY teď je stejná osoba jako po 5 letech, stejně jako další pacienti – výběry se liší pouze časovým údajem)
- Hodnotí rozdíly mezi oběma výběry (nebo jejich pořadí)

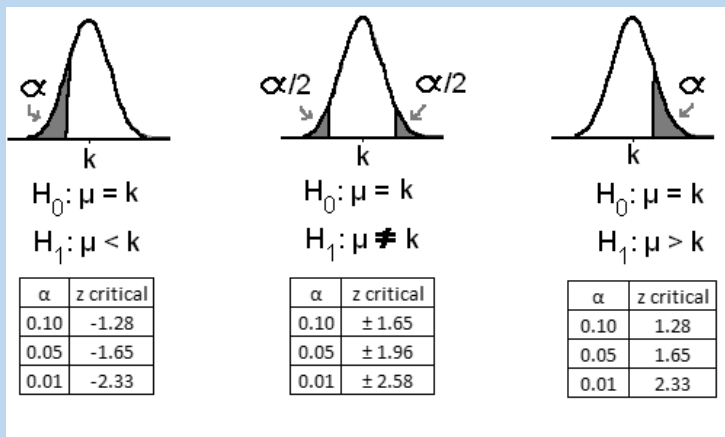
Nepárové

- Používané u na sobě nezávislých skupin (mohou být i různě velké)
- Porovnávají skutečné hodnoty proměnných (nebo jejich pořadí) mezi skupinami
- O párovém či nepárovém designu je třeba rozhodnout už před začátkem studie (párování je organizačně náročné, ale párové testy mají vyšší sílu)

Jednostranné vs. oboustranné testy

Jednostranné (one-tailed)

- H_0 je postavena asymetricky, např. léčivo A není lepší než léčivo B – ale už nás nezajímá, jestli je nebo není horší
- Mají větší sílu



Oboustranné (two-tailed)

- H_0 je symetrická: není rozdíl mezi léčivem A a léčivem B (tj. A není ani lepší, ani horší než B)
- Odhalují tedy odchylky na obě strany
- Obvykle vhodnější – a priori nevíme, jak pokus dopadne, a zajímají nás oba možné efekty

Příklady testů pro spojitá data – 2 výběry

Test	parametrický	neparametrický
Párový	Párový (závislý) t-test	Wilcoxonův párový test Znaménkový test
Nepárový	Nepárový (nezávislý) t-test	Mann-Whitneyův U-test * Kolmogorov-Smirnovův test

- * má téměř stejnou sílu, jako t-test, ale předpokládá alespoň přibližně podobnou variabilitu obou výběrů (stejně jako t-test)

Příklady testů pro spojitá data – více výběrů

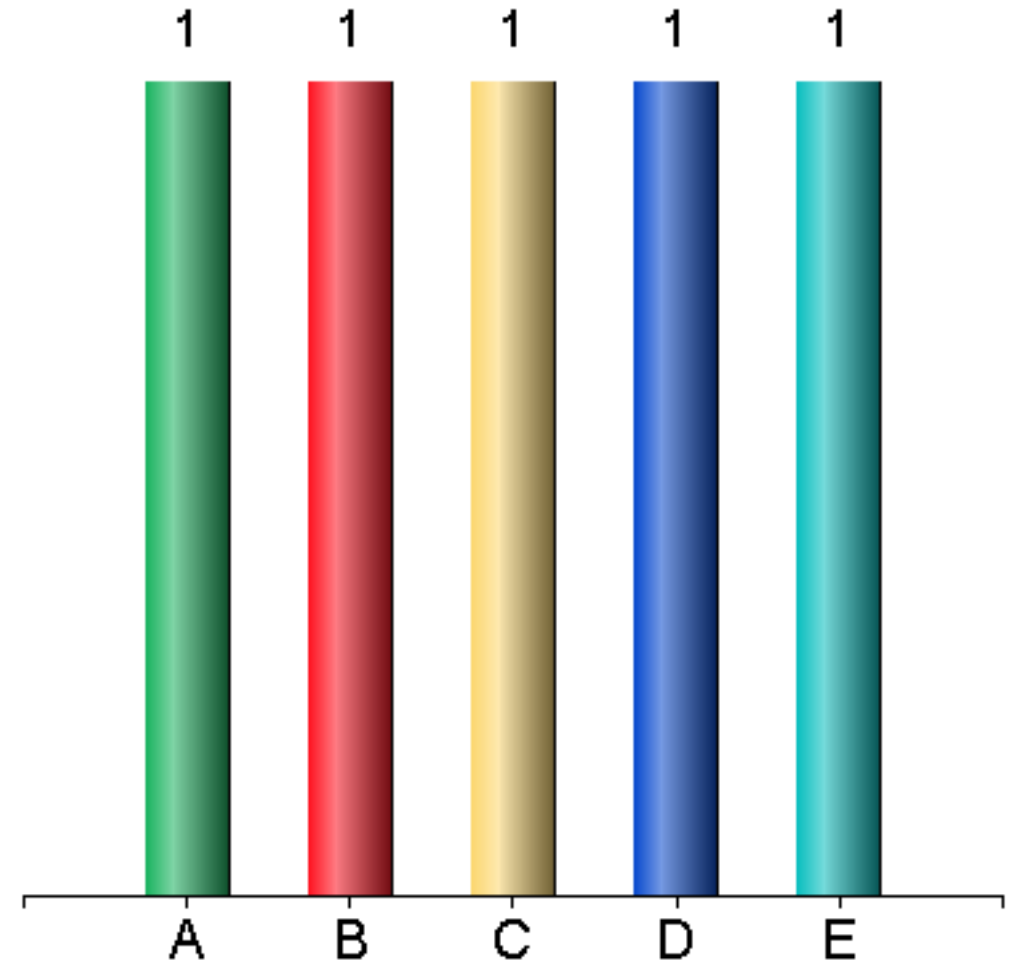
Test	parametrický	neparametrický
Párový	ANOVA (analýza rozptylu) pro opakovaná měření (RMANOVA)	Friedmanův test („ANOVA“)
Nepárový	Jednocestná ANOVA (a její varianty)	Kruskal-Wallisův test („ANOVA“)

- Vyloučí-li ANOVA H_0 , je třeba se ptát, mezi kterými konkrétními výběry je rozdíl – post hoc testy

Vyberte nejvhodnější test

V rámci studie dostávají pacienti nový lék proti epilepsii nebo placebo. Studie je randomizovaná (lék je náhodně losován). Zařazeni jsou pacienti, kteří mají nejméně jeden a nejvýše deset záchvatů za tři měsíce. Hodnoceným parametrem je počet epileptických záchvatů v prvním roce.

- A. Párový t-test
- B. Nepárový t-test
- C. Mann-Whitney U-test
- D. Znaménkový test
- E. ANOVA pro opakovaná měření



Problém mnohonásobného srovnání

- Provedeme-li více testů najednou, zvyšuje se pravděpodobnost, že některý z nich dá signifikantní výsledek pouze díky náhodě (tj. chyba I. typu – neoprávněné zamítnutí H_0) – problém zejm. při post hoc testech u ANOVA
- Např. při deseti testech a $\alpha = 0,05$ je pravděpodobnost, že při platnosti H_0 ani jeden test nedá signifikantní výsledek rovna $(1-\alpha)^{10} = 60\%$, tj. ve 40% dojde k neoprávněnému zamítnutí H_0 .
- Proto různými korekcemi (Bonferroni, Benjamini-Hochberg...) hladinu α snižujeme (a tedy zpřísňujeme kritéria pro zamítnutí H_0).
- Bonferroniho korekce: původní α se dělí počtem testů (nebo jsou p-hodnoty testů násobené počtem testů při nezměněné α).
 - velmi „konzervativní“.

Post hoc testy u ANOVA

- každý s každým („fotbalové zápasy“)
 - Bonferroniho korekce $\alpha / [n(n - 1) / 2]$
 - Tukey, Scheffé (ANOVA)
 - Dunn (Kruskal-Wallis)
 - Neményi (Friedman)
- s kontrolní skupinou
 - Bonferroniho korekce $\alpha / (n - 1)$
 - A priori nás nezajímá srovnání dalších skupin mezi sebou
 - Dunnett (ANOVA)
 - Dunnett rank sum (neparametrické testy)

„Manuální“ korekce na mnohonásobné srovnání

- Tam, kde neexistují standardizované post hoc testy jako součást statistického software
 - např. genetické testy – parametr u řady kandidátních polymorfismů, srovnání kategoriálních dat u více skupin
- Bonferroni: α se vydělí počtem testů (k)
- Bonferroni-Holm: každý test má jinou α -hodnotu. U testu s nejmenší p -hodnotou je $\alpha(\text{corr})$ rovna α/k , u druhého $\alpha/(k-1)$, u třetího $\alpha/(k-2)$... až u posledního je rovna α
- Benjamini-Hochberg (FDR): každý test má jinou α -hodnotu. U testu s nejmenší p -hodnotou je $\alpha(\text{corr})$ rovna α/k , u druhého $\alpha/(k/2)$, u třetího $\alpha/(k/3)$... až u posledního je rovna α
- Narazíme-li na $p > \alpha(\text{corr})$, jsou výsledky dalších testů statisticky nevýznamné

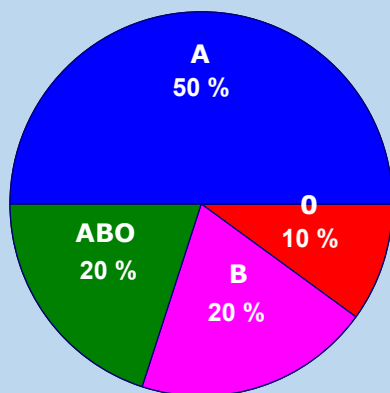
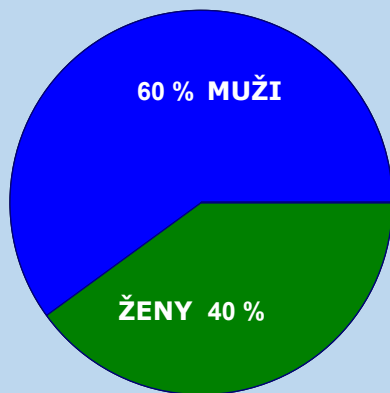
Kontingenční tabulky

- Vztah dvou kategoriálních proměnných lze vyjádřit kontingenční tabulkou $n \times n$ resp. $n \times m$ (příklad pro tabulku 3×2)

Počet z Císlu	Skupina		
Genotyp ADRB1	control	TTC	Celkový součet
CC	15	15	30
CG	12	10	22
GG		1	1
Celkový součet	27	26	53

Testy pro kategoriální data

- binární proměnná – kategoriální proměnná s pouze 2 hodnotami
 - 1/0, ano/ne, ...
- Někdy je výhodný převod větších tabulek na tabulku 2x2 [lze zejména u ordinálních dat – např. kategorie srdečního selhání NYHA I-IV lze převést na binární data jako lehké selhání (NYHA I+II) a těžké selhání (NYHA III+IV)]
- Z kontingenční tabulky lze spočítat její pravděpodobnost při platnosti H_0 (tedy p-hodnotu) i velikost sledovaného efektu – v našem příkladu např. vztah mezi mutací a nemocí (ukazatele RR – relativní riziko; OR – poměr šancí)
- U tabulek s binárními proměnnými lze využít i párový design (typicky zdraví a nemocní v čase)
- Vždy je nutno předem určit, která proměnná je nezávislá a která závislá



	nemoc	zdraví
mutace	50	2
ne	4	48

Relativní riziko a poměr šancí u tabulek 2x2

		Disease	
		+	-
Exposure	+	a	b
	-	c	d

Relative Risk =
Incidence of disease among those exposed
= $(a/a+b)$ $355/(355+3140) = 1.92$
Incidence of disease among those not exposed
($c/c+d$) $140/(140+2507)$

Odds Ratio =
Odds of people with disease being exposed
= (a/c) $355/140 = 2.02$
Odds of people without disease being exposed
(b/d) $3140/2507$

Example:
MI

		MI	
		+	-
Smoking	+	355	3140
	-	140	2507

www.mdedge.com

- pravděpodobnost vs. šance
- RR je vhodné pro prospektivní studie, u OR na designu nezáleží
- Pokud je závislá (modelovaná) proměnná vždy stejná (např. nemoc v tabulce vlevo), jsou hodnoty RR ($a/(a+b)$) a OR (a/b) podobné u nízké incidence jevu a
- RR je intuitivněji pochopitelné, OR je univerzálnější, oblíbené např. v logistické regresi

Příklady testů pro kategoriální data

Test	Kontingenční tabulky 2x2	Více kategorií/více měření ‡
Párový	McNemarův test	Cochranův Q test (více měření, binární data) Znaménkový test (dvě měření, ordinální data)
Nepárový	Chi-kvadrát (χ^2) test * Fisher exact test	Chi-kvadrát (χ^2) test * Cochran-Armitage test (tabulky 3x2, ordinální data)

* předpokládá jisté minimální četnosti v každém poli kontingenční tabulky
(cca $n \geq 5$)

‡ při vyloučení H_0 nutno doplnit testy pro 2 x 2 tabulky s korekcí
při mnohonásobném srovnání

Příklad

Cílem studie bylo testování vztahu mezi krevní skupinou v ABO systému (A, B, AB a 0) a přítomností akutních komplikací krevní transfúze. Kolik polí má příslušná kontingenční tabulka?

Ranking	Response	Votes
Correct Answer		
1		
2		
3		
4		
Others		

Příklad

V předchozím případě bylo při testování χ^2 testem zjištěno $p < 0.05$ a následovala série post hoc testů pro tabulky 2x2 „každý s každým“. V jednom z nich byl zjištěn vyšší výskyt komplikací u pacientů s krevní skupinou AB oproti skupině A, $p = 0,05$ (5 %). Jak se p-hodnota změní po Bonferroniho korekci (korigujeme zde p , ne α)? Výsledek uveďte v celých procentech (přirozené číslo), případně na ně zaokrouhlete.

Ranking	Response	Votes
Correct Answer		
1		
2		
3		
4		
Others		

Regresní modely

- „Regrese k průměru“ (Galton) – ale metody již Friedrich Gauss
- Cílem je odhadnout hodnotu modelované veličiny (závislá proměnná = regresand) pomocí jiných známých parametrů (faktorů = regresorů – kategoriální a spojité proměnné)
- Lze odhadnout míru přispění jednotlivých faktorů samostatně (univariační modely) i společně ve vzájemné interakci (multivariační modely)
- Předpoklad: faktory jsou **nezávislé**
- Nejčastěji
 - Lineární regrese (závislá proměnná je spojitá)
 - Logistická regrese (závislá proměnná je binární)
 - Coxova regrese (závislou proměnnou je přežití – endpoint a čas přežití)

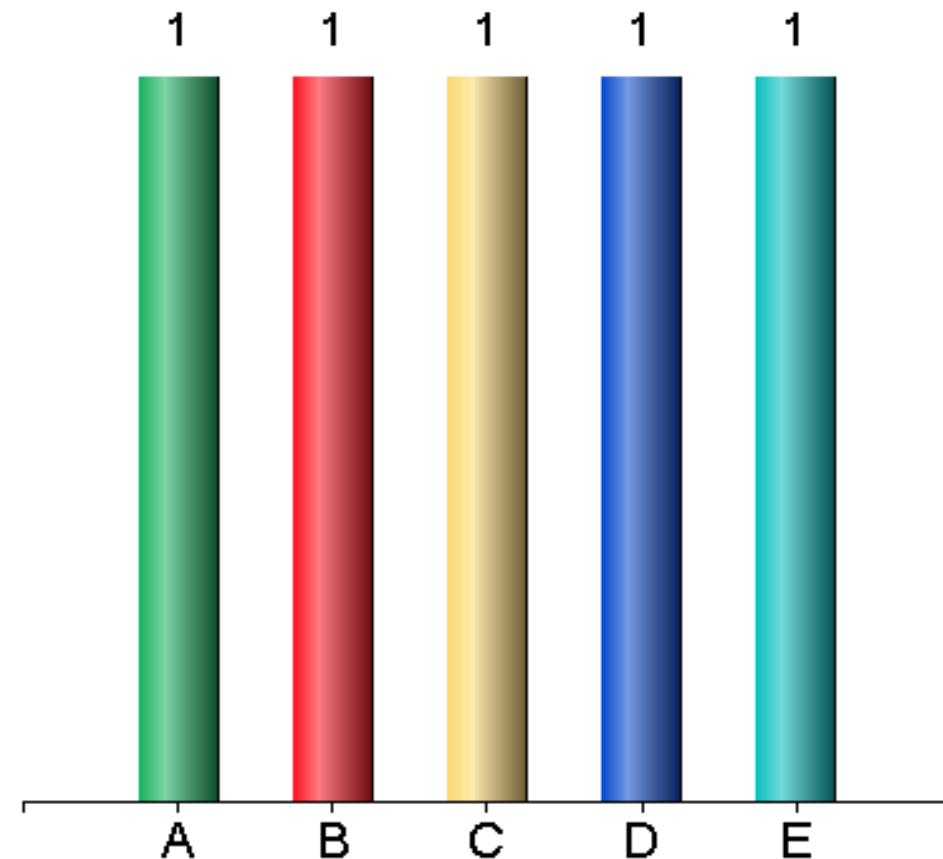
Vyjádření příspěvku jednotlivých faktorů

- Lineární regrese – regresní koeficient β (standardizovaný, nestandardizovaný) a 95% interval spolehlivosti (CI) – tj. odhad, kde je koeficient s 95% pravděpodobností
 - Na rozdíl od korelace záleží na tom, která proměnná je závislá a která nezávislá
 - Je-li regresor kategoriální, jedná se fakticky o ANOVu
- Logistická regrese – OR a 95% interval spolehlivosti
- Coxova regrese – poměr rizik (HR) a 95% interval spolehlivosti
- Zahrnuje-li $\beta \pm 95\%$ CI hodnotu 0, není příspěvek daného faktoru statisticky významný (tj. nelze rozhodnout, jestli výslednou hodnotu zvyšuje nebo snižuje)
- U OR a HR platí totéž, zahrnuje-li 95% CI hodnotu 1 (pod 1 snižuje pravděpodobnost události, nad 1 zvyšuje)
- 95% CI tak mohou nahradit p-hodnotu
- Je-li nezávislá proměnná kategoriální, je nutno jednu kategorii určit jako referenční a regresní koeficienty / OR / HR patří zvlášť každé další kategorii
- Je-li nezávislá proměnná spojitá, odpovídá vždy β / OR / HR 1 jednotce (např. 1 roku věku – předpokládá se lineární efekt (jinak lepší kategorizace))

Vyberte správné tvrzení

V průřezové studii v populaci 700 hospitalizovaných pacientů mezi 80 – 90 lety byly u 40 % zjištěny známky kognitivní dysfunkce. Asociace s kandidátními rizikovými faktory (věk, hypertenze, diabetes) byla hodnocena univariační logistickou regresí. Přítomnost kognitivní dysfunkce byla postupně asociována s: věkem (pro každý další rok OR = 1.20; 95 % CI = 1.12 – 1.40), hypertenzí (OR 1.40; 95 % CI 1.20 – 1.78) i diabetem (OR 2.80; 95 % CI 2.00 – 6.40)

- A. faktor věku není pro kognitivní dysfunkci statisticky významný
- B. Pravděpodobnost vzniku kognitivní dysfunkce je dvakrát větší u diabetiků než u hypertoniků
- C. Věk, diabetes i hypertenze jsou navzájem nezávislé rizikové faktory
- D. P-hodnota je ve všech případech < 0.05
- E. Lze usoudit na příčinnou souvislost mezi jednotlivými faktory a kognitivní dysfunkcí



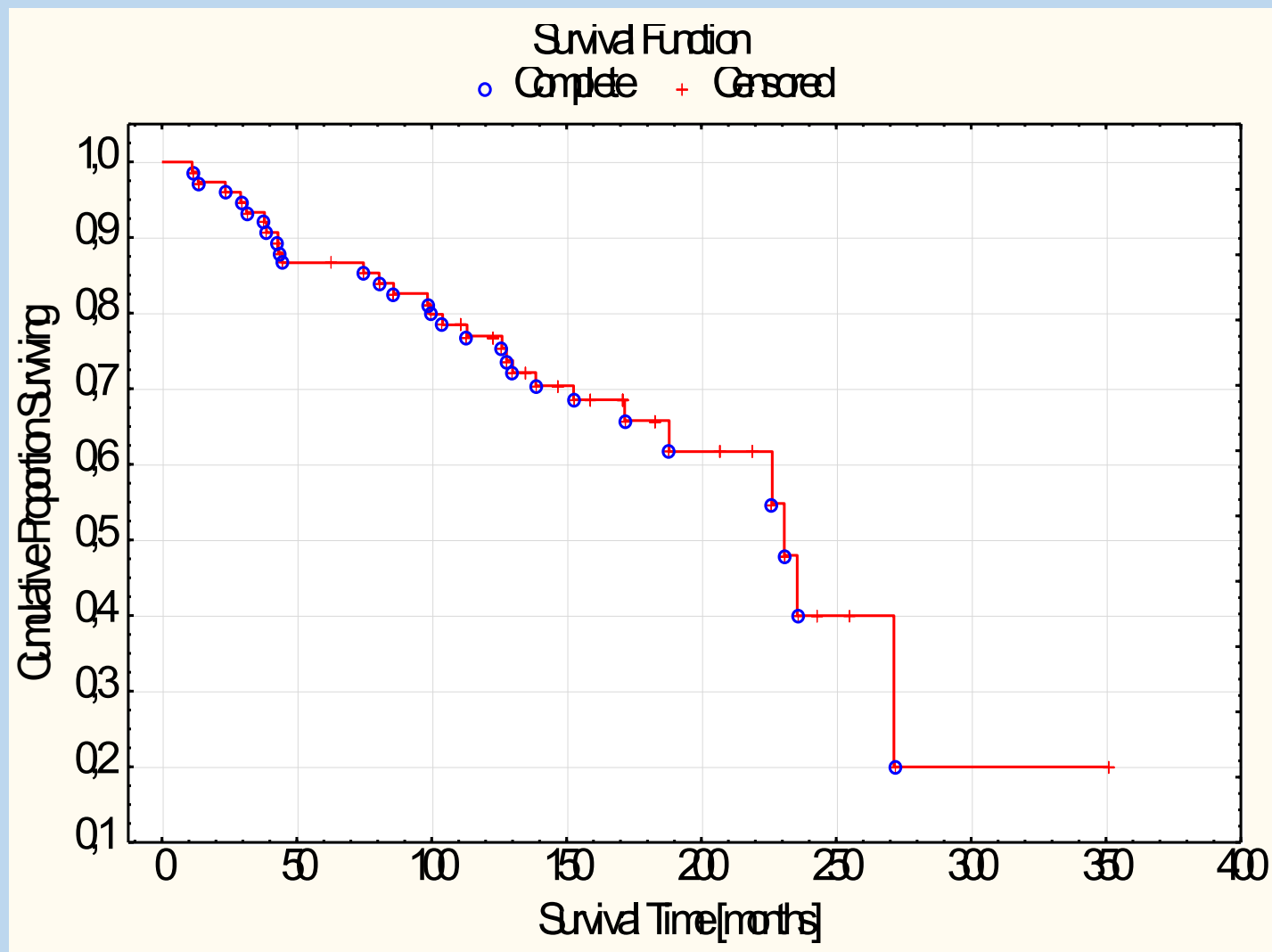
Co s ordinálními daty?

- Testy pro kategoriální data, ANOVA (ale: ignorujeme seřazení)
- Neparametrické testy (je-li hodně kategorií)
- Dichotomizace a testy pro binární data (v medicíně asi nejčastěji)
- Speciální testy – Cochran-Armitage (typicky genetika), znaménkový test (ale: ignoruje hodnoty, kde nedošlo ke změně)

Analýza přežití

- Četnost události (koncový bod, endpoint) obvykle klesá s klesajícím počtem účastníků studie („přeživší“)
- Cenzorovaná data
 - před koncem studie nedošlo k události
 - ztraceni ze studie
 - úmrtí z jiné příčiny
- Kaplan-Meierovy křivky
- Log rank test

Kaplan-Meierova křivka přežití

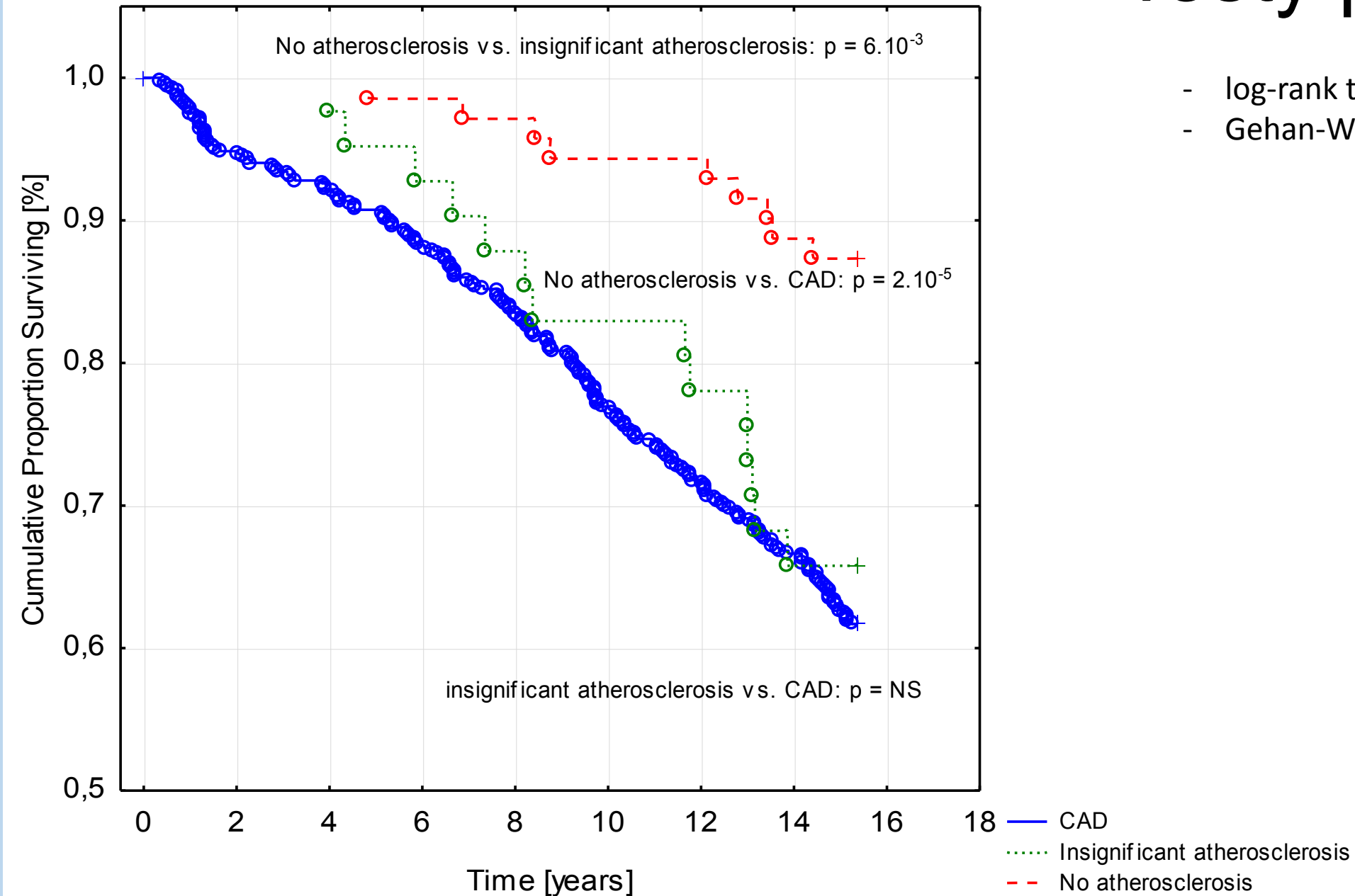


Cumulative Proportion Surviving (Kaplan-Meier)

○ Complete + Censored

Testy přežití

- log-rank test
- Gehan-Wilcoxonův test



Vyberte správnou odpověď...

Do studie, zabývající se rizikem recidivy infarktu myokardu (endpoint), se přihlásili čtyři pacienti. V následujících letech postupně došlo k následujícímu vývoji: jeden pacient se odstěhoval do Argentiny a ztratil se tak ze studie, poté jeden dostal infarkt a příští měsíc zemřel při autonehodě, následně jeden zemřel na nádor plic a poslední se ve zdraví dožil konce studie. Poslední bod Kaplan-Meierovy křivky je na hodnotě:

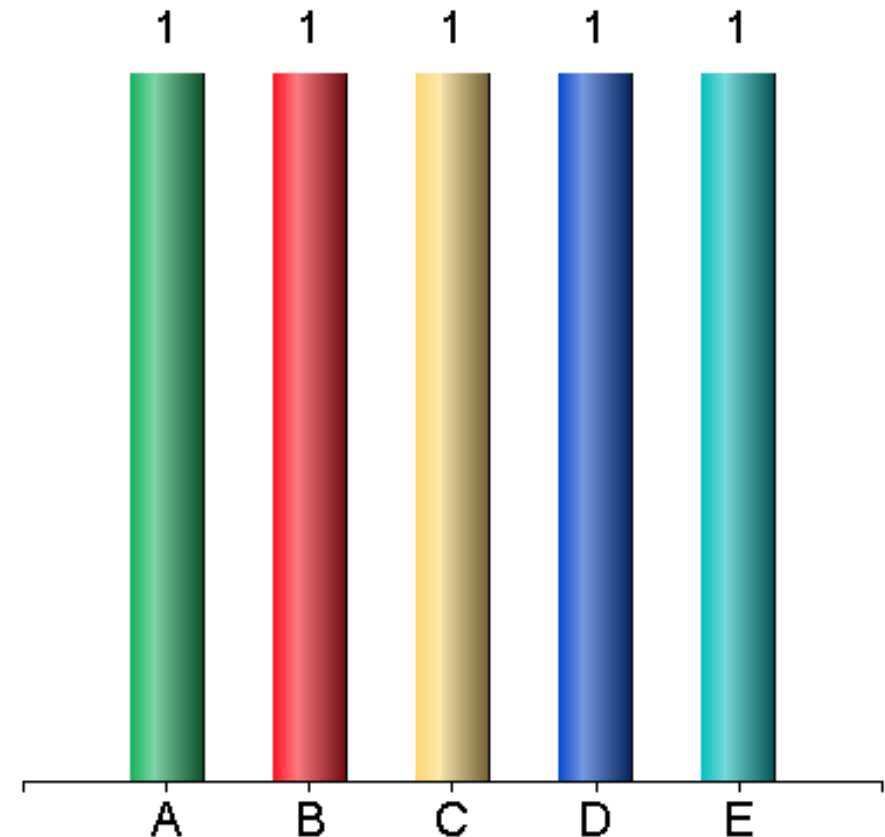
A. 66,6%

B. 50%

C. 33,3%

D. 25%

E. 0%



Shluková (clusterová) analýza

- vícerozměrná analýza
 - měření vzdálenosti
 - řadící algoritmus
 - standardizace dat
-
- k means clustering (předem známý počet clusterů)
 - hierarchický strom (dendrogram)

Vyberte správnou odpověď...

Na opuštěný ostrov se dostanou antropologové, kteří zde objeví neznámé lebky. Pomocí shlukové (clusterové) analýzy je budou chtít přiřadit k některé z okolních žijících populací. Vedle genetických markerů stanoví i kraniální index (v procentech, průměr = 85, sm. odch. = 10), faciální index (v procentech, průměr = 80, sm. odch. = 5) a objem mozkovny (v cm^3 , průměr = 1500, sm. odch. = 200). Co se stane, nebudou-li data před analýzou standardizována?

- A. Nic, standardizace se používá pro přehlednost.
- B. Objem mozkovny nebude z hlediska analýzy relevantní.
- C. Clusterová analýza nebude technicky možná.
- D. Zařazení do clusteru bude záviset především na objemu mozkovny.
- E. Vzroste vzájemná korelace kraniálního a faciálního indexu.

