

STATISTIKA

Sázíte-li ve Sportce, je to hazard.
Sázíte-li se, že vám v kartách přijdou tři
postupky po sobě, je to zábava.
Vsadíte-li se, že cena plynu stoupne o
10 %, je to podnikání. Vidíte ten rozdíl?

Martin Sebera, FSpS MU, 12.2.2014

Pravidla výzkumu z pohledu analýzy dat

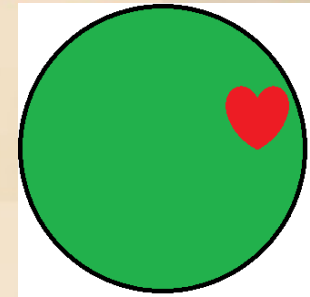
- 1. příprava výzkumného šetření je nejdůležitější část**
- 2. sběr a analýza dat slouží k zamítnutí/nezamítnutí předem stanovených úkolů práce a hypotéz (explorační vs. konfirmační přístup)**
- 3. vždy mít na paměti věcné hledisko výzkumu, zejména v souvislosti s interpretací statistických výsledků**

Role statistiky

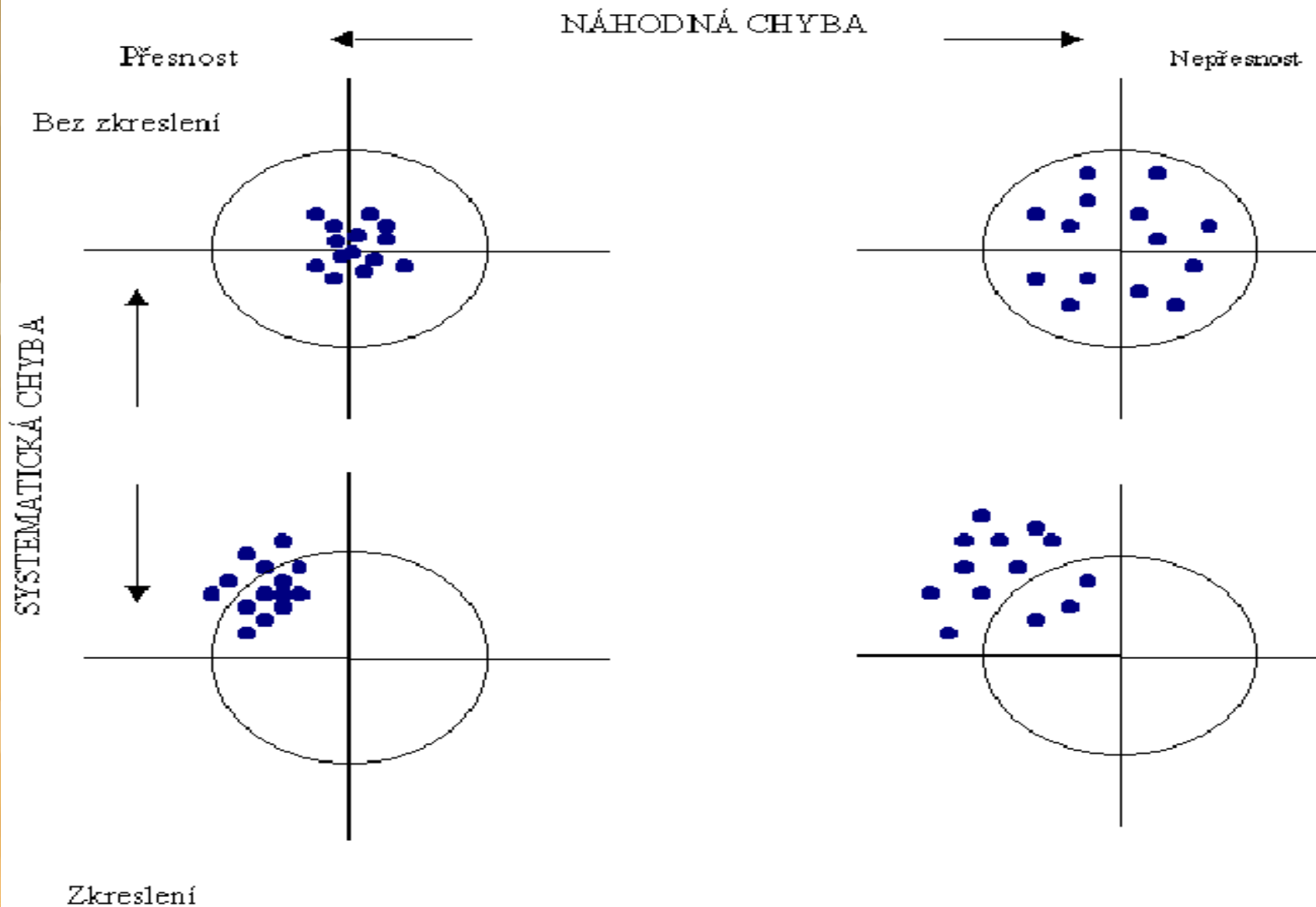
- Porozumění a zkoumání hromadných jevů
- Zjišťování zákonitostí
- V kvantitativním výzkumu (deduktivní princip) – pojitko mezi teorií a výzkumem
- Zpracování, popsání a analyzování dat

Základní pojmy

- **Základní** a **výběrový** soubor a jeho rozsah (N)
- **Výběr:**
 - náhodný (každý prvek má stejnou pravděpodobnost výběru - losování)
 - systematický (n-tý objekt, $n < N$)
 - stratifikovaný (náhodný výběr ve skupinách)



Náhodná a systematická chyba



Typy proměnných

- **Nominální (text, číselné kódy; hodnoty jsou různé; nelze provádět aritmetické operace)**
- **Ordinální (lze seřadit; většinou se převede na čísla),**
- **Intervalová (lze říct o kolik je hodnota větší)**
- **Poměrová (lze říct kolikrát je hodnota větší)**
- **Spojité X Diskrétní**

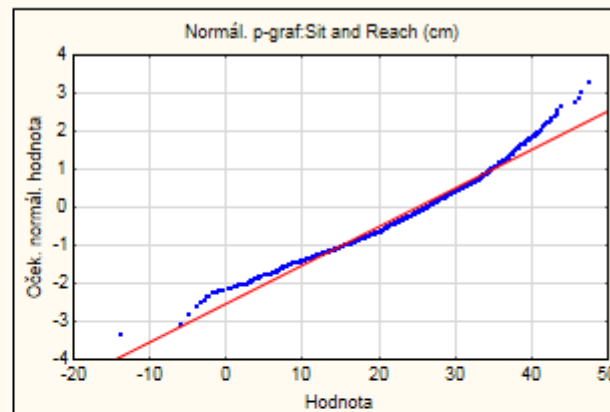
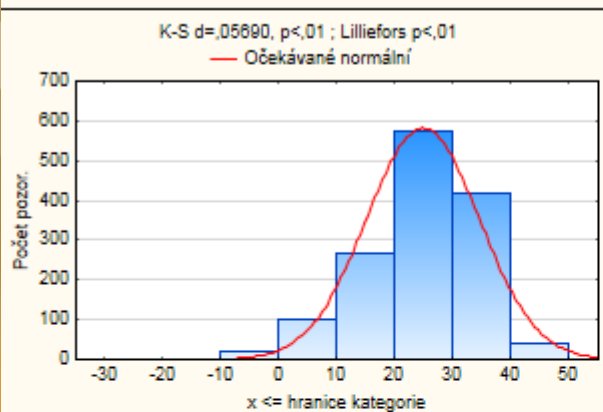
Škály (měřítka, stupnice)

- **Nominální (temperament; národnost)**
- **Ordinální (školní známky, bodování v slopestyle; relace =, ≠, >, <,),**
- **Metrické**
 - **Intervalová (lze říct o kolik je hodnota větší)**
 - **Poměrová (lze říct kolikrát je hodnota větší)**
 - **Př. teplota, čas, hmotnost, ...**

První náhled na data – popisná statistika

- průměr, sm. odchylka, medián, kvartily aj.
- četnosti: absolutní, relativní, kumulativní
- grafy: krabicový, histogram

Souhrn: Sit and Reach (cm)



Souhrnné statistiky: Sit and Reach (cm)

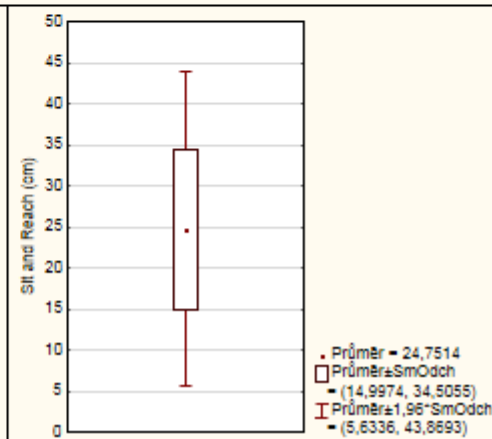
N platných=1423,000000

Průměr= 24,751441

Minimum=-14,000000

Maximum= 47,300000

Sm.odch.= 9,754011



Proč?

- chybná měření, extrémny
- homogenitu souboru
- chybějící data

Intervalové rozložení četností

BMI:

18 19 19 20 20 20 20 20 20
20 21 21 21 21 21 21 22 22 22

N – rozsah souboru

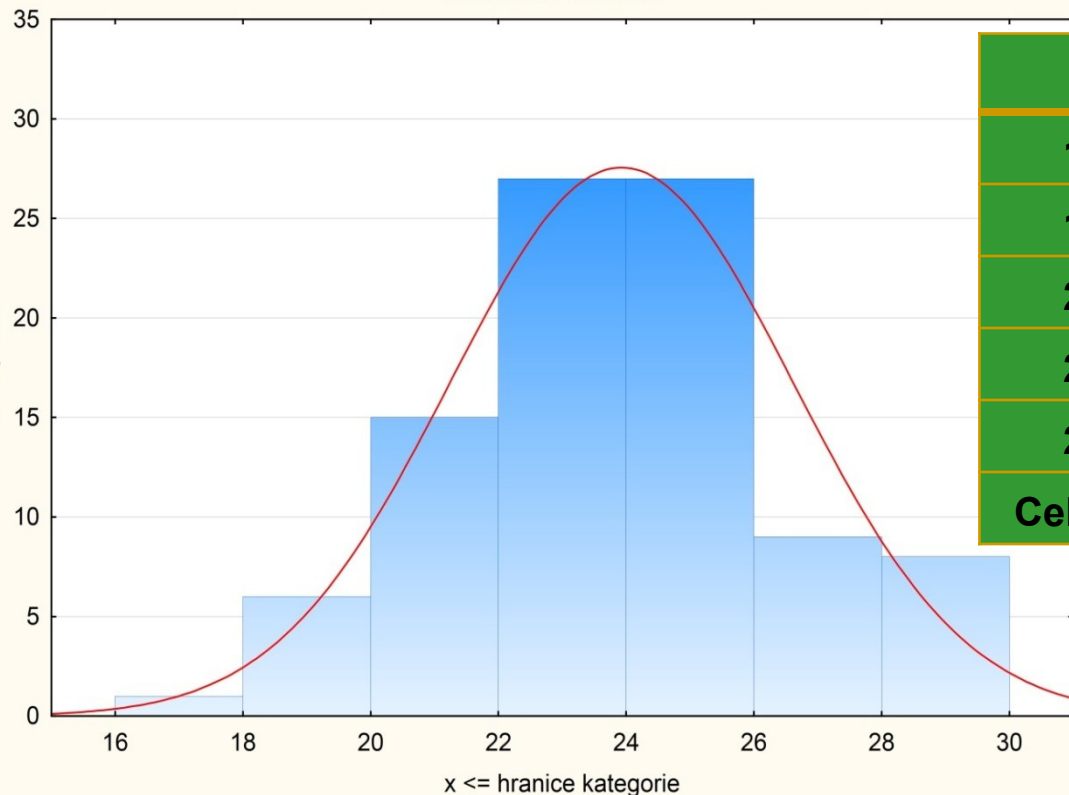
n_i – absolutní četnost

r_i – relativní četnost

N_i – kumulativní absolutní četnost

F_i – kumulativní relativní četnost

Histogram: BMI
 — Očekávané normální



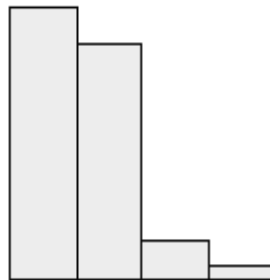
x	n_i	r_i	N_i	F_i
18	1	0,05 (= 1/20)	1	0,05
19	2	0,10 (= 2/20)	3	0,15
20	8	0,40 (= 8/20)	11	0,55
21	6	0,30 (= 6/20)	17	0,85
22	3	0,15 (= 3/20)	20	1,00
Celkem	20	1,00		

Ize usuzovat na některé vlastnosti, záleží na počtu intervalů

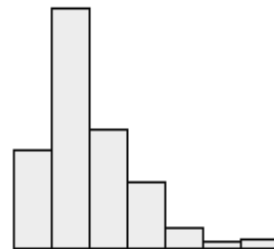
**Sturgersovo
pravidlo**



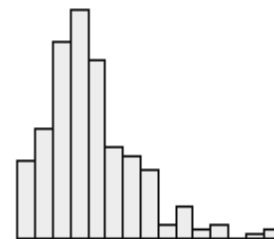
$$I = 1 + 3,3 \log n$$



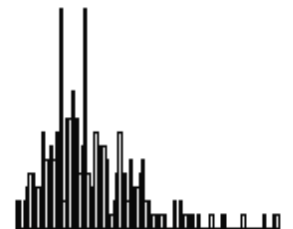
4 intervaly



8 intervalů



15 intervalů



100 intervalů

**Dixonovo
pravidlo**



$$I = 10 \log n$$



Základní statistické charakteristiky

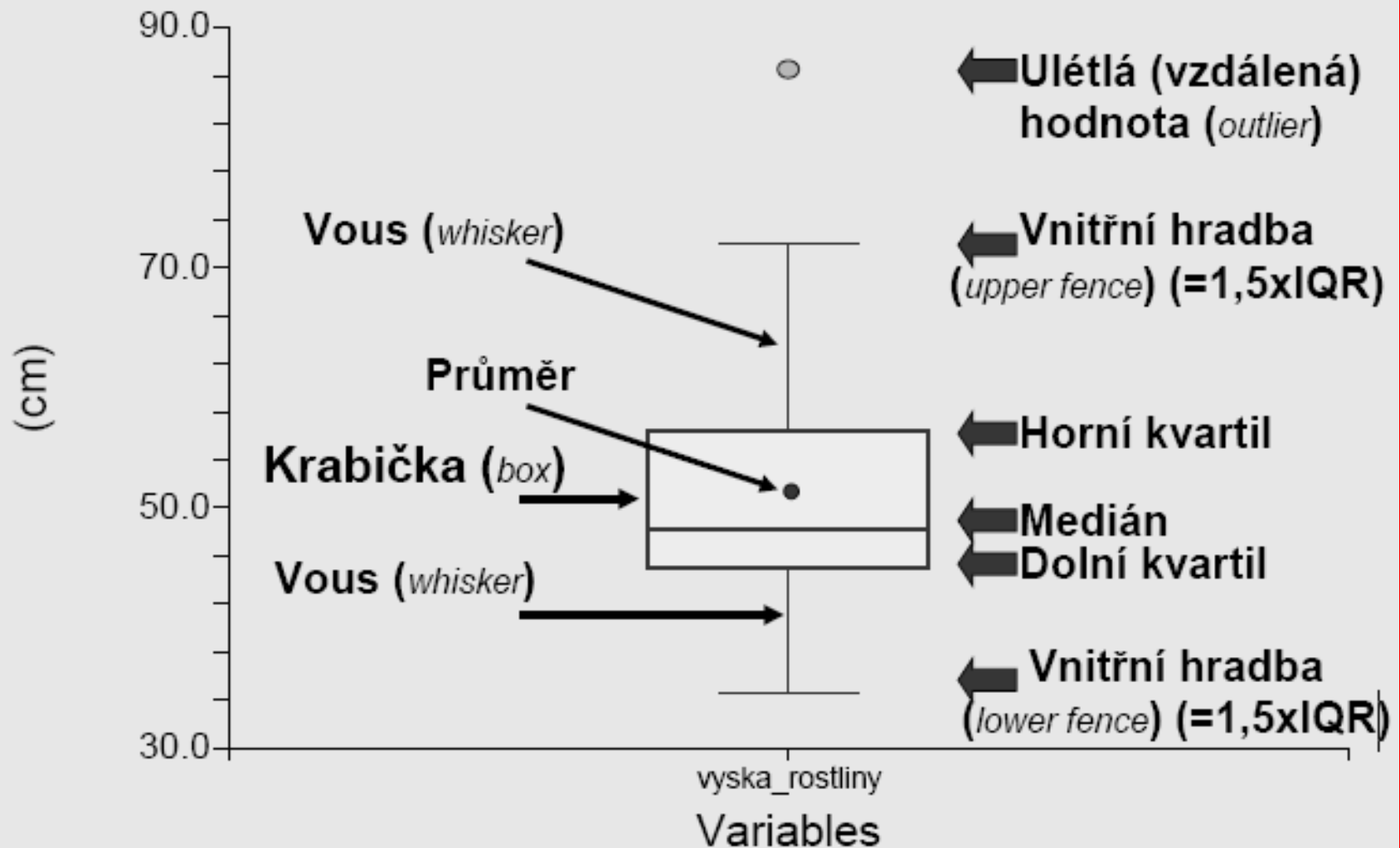
- **Míry střední hodnota**
 - Aritmetický a geometrický průměr, modus, medián
- **Míry variability**
 - variační rozpětí, kvantily, rozptyl, směrodatná odchylka, variační koeficient
- **ztrácíme mnoho cenných informací o původních datech**

– 1; 10; 22	průměr 11	SD 10,53	n = 3
– 11; 11; 11	průměr 11	SD 0	n = 3

Časté chyby při statistických výpočtech

- Uvedení **průměru** bez směrodatné odchyly **SD** a bez **N**
- Procenta
 - Regulovaná složka stoupla o 200 %, silová zlevnila o 20 %. Jak se změnila celková cena?
 - Regul: 100,- Kč → 300,- Kč **původní cena** **3100,- Kč**
 - Silová: 3000,- Kč → 2700,- Kč **nová cena:** **3000,- Kč**
 - Nejen procenta, ale i z jakých základů se počítají
- snížení platu o 30 % a jeho následné zvýšení o 30 %
 - při původním platu **100 Kč** je plat po snížení 70 Kč (-30 %), ale po následném zvýšení o 30 % pouze **91 Kč**.

Box Plot



Testování hypotéz, koncept věcné vs. statistické významnosti

Postup testování hypotéz → poměrně jasný a jednoduchý.

- Vytvoříme hypotézu H_0 , o které předpokládáme, že platí. Proti ní postavíme alternativu (H_A). Sesbíráme data. Najdeme věrohodný aparát, který konstatuje, zda domněnka platí nebo ne → statistický test.
- **chyba 1. druhu** se značí α a nazývá se **hladina významnosti**. Výraz $1 - \alpha$ se nazývá **spolehlivost**
- **chyba 2. druhu** se značí β . Výraz $1 - \beta$ se nazývá **síla testu**
- Obvyklé hodnoty spolehlivost: 0,95 nebo 0,99;
- síla testu např. 0,8 → volíme např. hladinu významnosti $\alpha = 0,05$ nebo 0,01.

Testování hypotéz		výsledek testu	
		hypotéza H_0 platí	hypotéza H_A platí
reálná situace	hypotéza H_0 platí	správné rozhodnutí	chyba 1. druhu značí se α
	hypotéza H_A platí	chyba 2. druhu značí se β	správné rozhodnutí

Koncept věcné významnosti

Alternativou k statistické významnosti je posuzování tzv. věcné významnosti (effect size). Lze ji stanovit jako:

- minimální hodnotu v absolutních hodnotách znamenající věcnou významnost
- minimální vysvětlené procento rozptylu (relativní zhodnocení podílu ostatních faktorů – koeficient ω^2)

Pro jednotlivé testy lze v literatuře nalézt mnoho tzv. koeficientů věcné významnosti. **Jednou z výhod konceptu věcné významnosti je nezávislost na počtu měření N.**

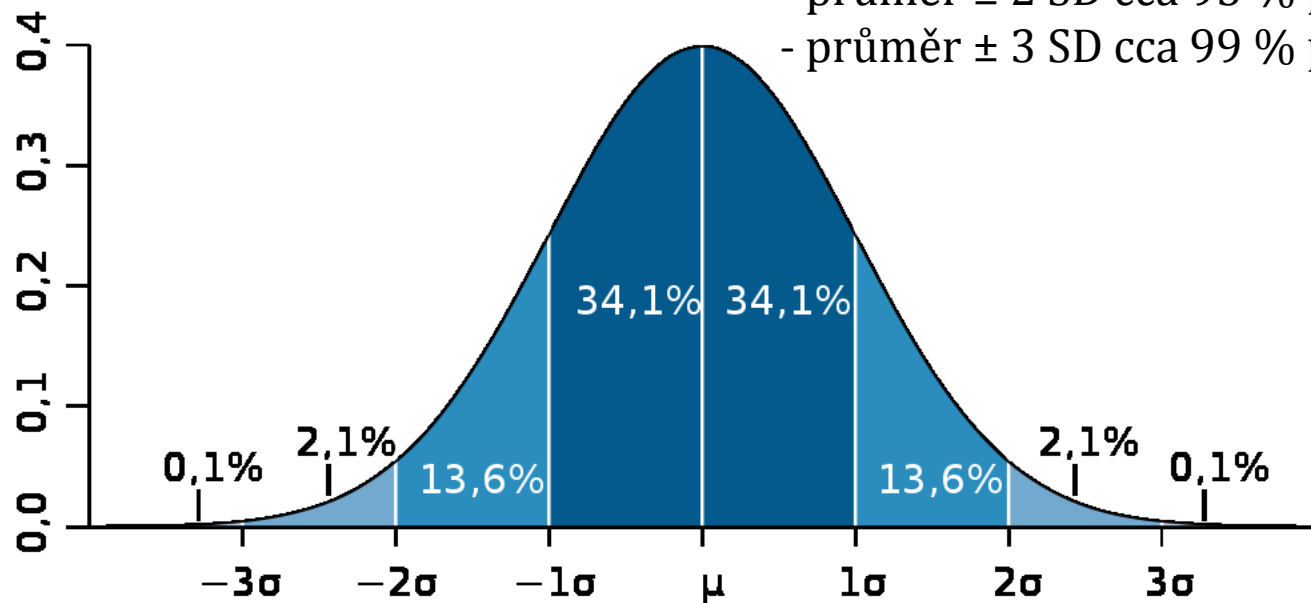
	koeficient	hodnocení efektu
Korelační koeficient r	r^2 koeficient determinace	malý (nízký) efekt: $r = 0,10 - 0,30$ střední efekt: $r = 0,31 - 0,70$ velký (výrazný) efekt: $r = 0,71 - 1$
t-test	Cohenovo d	$d = 0,20$ malý efekt $d = 0,50$ střední efekt $d = 0,80$ velký efekt

Normalita

- Kolmogorov-Smirnov a Shapiro-Wilks test
- **Proč?**
rozhodnutí, zda použít parametrické nebo neparametrické testy

Pro normální rozložení platí:

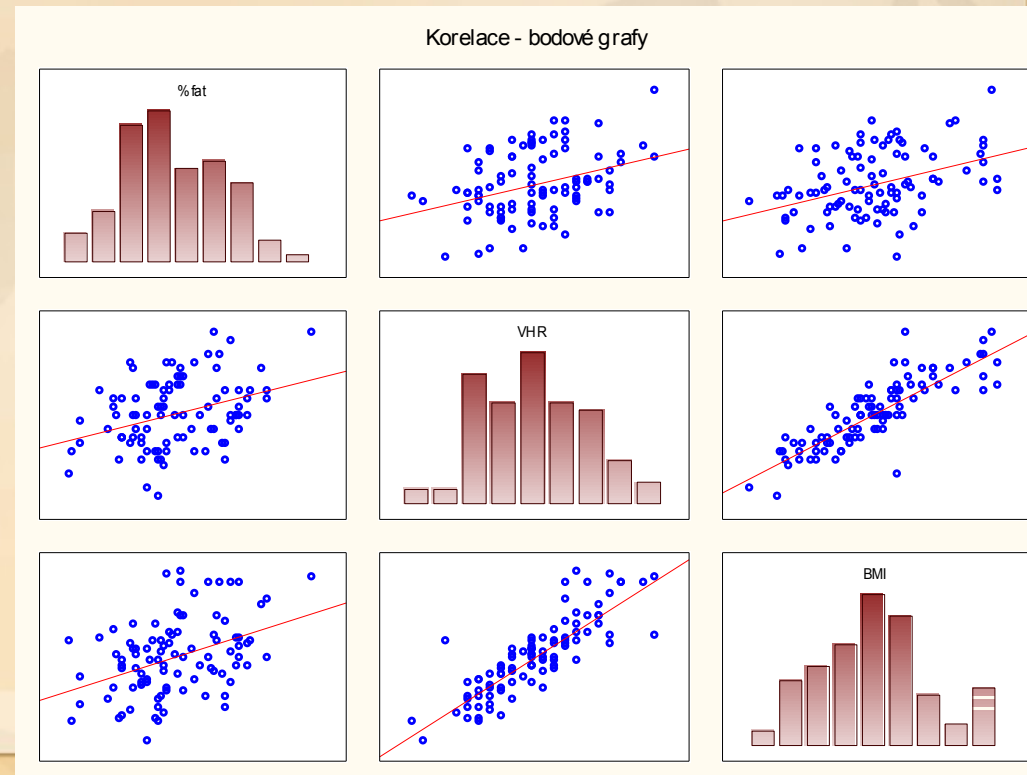
- průměr \pm 1 SD cca 68 % případů
- průměr \pm 2 SD cca 95 % případů
- průměr \pm 3 SD cca 99 % případů



Korelace ANEB korelace není kauzalita

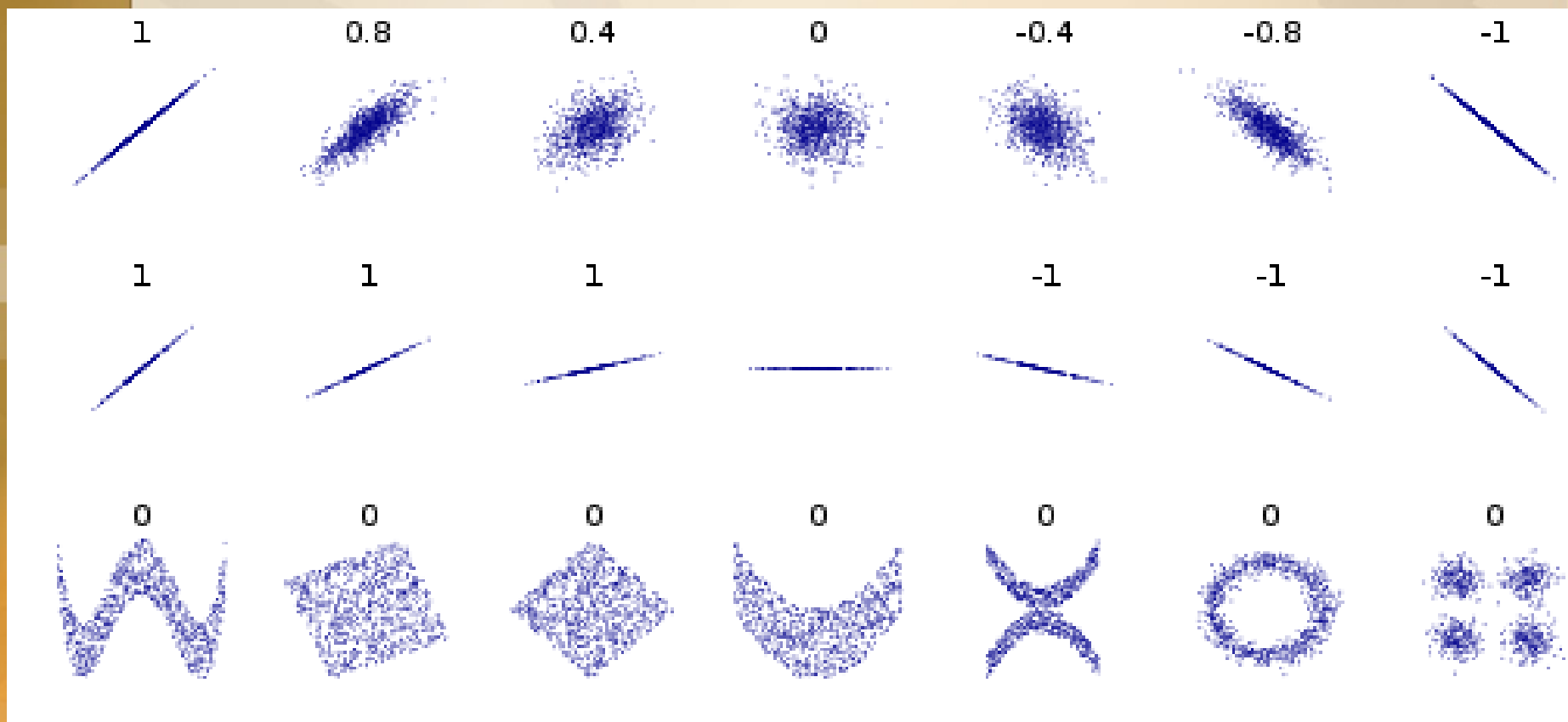
- = vzájemný vztah mezi veličinami proměnnými, jevy (dostatečně velký rozsah)
- Úkol: zjistit závislost a popsat ji
- Příklad 3 proměnné:

- BMI
- % fat
- WHR



Korelační koeficient

- **R : <-1 do 1>**
- **Omezení:**
 - předpokládá 2-rozměrné norm.rozdělení
 - měří pouze vztahy lineární
 - nerozeznává, která proměnná je závislá a která nezávislá. Nelze rozhodnout o příčinnosti vztahu mezi proměnnými
- **interpretace → dodatečné koeficienty, např. index determinace r^2**
- ***Pearsonův, neparametrický Spearmonův***
- ***jednoduchý, parciální, mnohonásobný***



Příklad

	% fat	WHR	BMI
% fat	1	0,36	0,41
WHR	0,36	1	0,85
BMI	0,41	0,85	1

Nejvyšší **jednoduchý** korelační koeficient je mezi proměnnými BMI a WHR a to 0,85. Celkem vysvětluje 72,2 % procent celkové variability mezi těmi to proměnnými. K číslu 72,2 % jsme dospěli pomocí koeficientu determinace ($r^2 = 0,85^2 = 0,722$).

Korelace vs. kauzalita

Michal Bozděch

Příklad 1: pozitivní korelace (vítr a větrná elektrárna)

Příklad 1: pozitivní korelace (vítr a větrná elektrárna)

- ▶ Krásným příkladem pozitivní korelace je vztah mezi větrem a otáčením lopatek větrné elektrárny. Když fouká silný vítr, lopatky se rychleji otáčejí, což způsobuje zvýšenou produkci elektřiny. Nicméně nemůžeme tvrdit, že lopatky elektrárny jsou příčinou větru. Tady pozorujeme pouze pozitivní korelaci, nikoliv však kauzální vztah.

Příklad 2: nesprávná kauzalita (videohry a násilí)

Příklad 2: nesprávná kauzalita (videohry a násilí)

- ▶ Často se můžete setkat s tvrzením, že násilí ve videohrách či televizi způsobuje, že děti se stávají více násilnými. Což je kauzální vztah. Avšak můžeme také tvrdit, že děti s přirozeným sklony k násilnému chování mají tendenci vyhledávat hry a pořady s násilnou tematikou. Proto nemůžeme přesně určit, zda A způsobuje B nebo B způsobuje A. V tomto případě se jedná pouze o korelaci, nikoliv o kauzalitu.

Příklad 3: prodej zmrzliny a počet lidí s úpalem

Příklad 3: prodej zmrzliny a počet lidí s úpalem

- Krásným příkladem omylu třetí proměnné (Third-Cause Fallacy) je situace, kdy pozorujeme pozitivní korelaci mezi vyššími tržbami zmrzlinového stánku a zvýšeným výskytem úpalu u lidí. Na první pohled bychom mohli usoudit, že vyšší tržby způsobují úpal. Avšak ve skutečnosti je zde třetí proměnná, která ovlivňuje oba tyto faktory a lépe vysvětluje pozorovaný jev – počasí. V teplém počasí se lidé více snaží ochladit a vyhledávají zmrzlinové stánky, což vede k vyšším tržbám. Zároveň je v teplém počasí větší riziko úpalu. Tedy počasí je klíčovou proměnnou, která má vliv na oba tyto jevy.

Příklad 4: Další příklad třetí proměnné (vši a zdraví)

Příklad 4: Další příklad třetí proměnné (vši a zdraví)

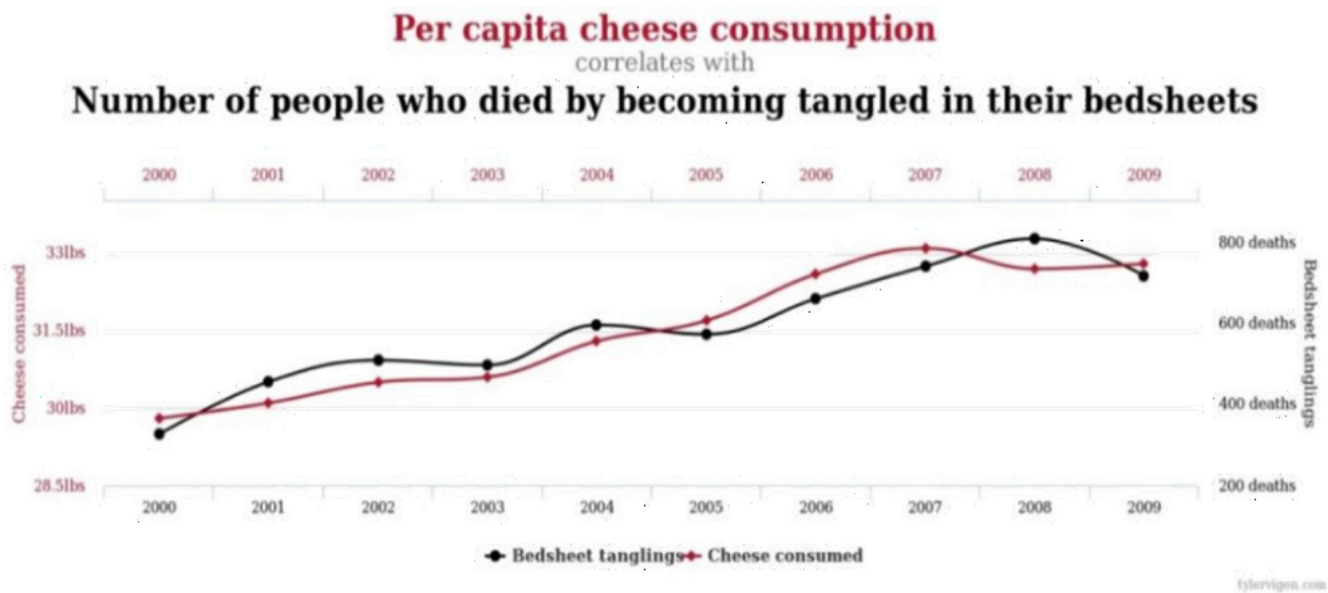
- ▶ Ve středověku si lidé všimli, že jedinci, kteří mají vši, jsou zdravější, zatímco nezdraví jedinci málokdy mají vši. Na základě tohoto pozorování došli k závěru, že množství vší pozitivně koreluje se zdravotním stavem, a předpokládali, že vši způsobují lepší zdraví. Avšak toto zjištění bylo ovlivněno dalším faktorem. Konkrétně, vši jsou citlivé na změny teploty lidského těla, které jsou reakcí na přítomnost viru v těle. Proto nemocní neměli vši.

**Příklad 5: Falešná korelace (konzumace sýra
a úmrtí z důvodu zamotání se do
prostěradla)**

Příklad 5: Falešná korelace (konzumace sýra a úmrtí z důvodu zamotání se do prostěradla)

- ▶ Falešná korelace (false correlation) je situace, kdy se zdá, že existuje vztah nebo spojitost mezi dvěma proměnnými, avšak ve skutečnosti není mezi nimi žádná příčinná souvislost. Jedná se o náhodnou nebo náhodně vzniklou korelaci, která není podložena žádným skutečným vztahem mezi studovanými proměnnými.
- ▶ Jako příklad můžeme uvést situaci, kdy autoři publikovali studii, ve které zjistili silnou a pozitivní korelaci mezi roční konzumací sýra a počtem úmrtí způsobených zamotáním do prostěradla. Tento zdánlivý vztah je však pouhým náhodným jevem a nemá žádný logický základ. Neexistuje žádná přímá příčinná souvislost mezi konzumací sýra a úmrtími způsobenými zamotáním do prostěradla. Jedná se pouze o falešnou korelaci, která nemá žádný skutečný význam nebo vysvětlení.

- Obrázek 10: Ukázka Falešné korelace ($r = .947$)



Hádanka: s rostoucí produkcí CO₂ se zvyšuje i počet lidí s obezitou

- Zde vzniká otázka, zda CO₂ způsobuje obezitu nebo zda je obezita zodpovědná za vyšší úroveň detekce CO₂, například způsobenou větší plynatostí u lidí s obezitou. Mohlo by se také jednat o situaci, kdy populace konzumuje více potravin a produkuje více CO₂, což je důsledek dnešní konzumní společnosti, a tak bohatství a je tou třetí vysvětlující proměnou. Musíme si uvědomit, že realita je často složitější, a i když přirozenou lidskou vlastností je rychle informace kategorizovat do přehledných „schránek“, tak existuje mnoho faktorů, které mohou ovlivňovat dané jevy. Ve vědeckém zkoumání je důležité pečlivě analyzovat data a provést další studie, abychom mohli identifikovat skutečnou příčinnou souvislost mezi proměnnými. Je zapotřebí přesněji rozlišovat mezi kauzalitou a pouhým korelačním vztahem, abychom mohli přijít ke správným závěrům.

Hádanka: s rostoucí produkcí CO₂ se zvyšuje i počet lidí s obezitou

- ▶ Zde vzniká otázka, zda CO₂ způsobuje obezitu nebo zda je obezita zodpovědná za vyšší úroveň detekce CO₂, například způsobenou větší plynatostí u lidí s obezitou. Mohlo by se také jednat o situaci, kdy populace konzumuje více potravin a produkuje více CO₂, což je důsledek dnešní konzumní společnosti, a tak bohatství a je tou třetí vysvětlující proměnou. Musíme si uvědomit, že realita je často složitější, a i když přirozenou lidskou vlastností je rychle informace kategorizovat do přehledných „schránek“, tak existuje mnoho faktorů, které mohou ovlivňovat dané jevy. Ve vědeckém zkoumání je důležité pečlivě analyzovat data a provést další studie, abychom mohli identifikovat skutečnou příčinnou souvislost mezi proměnnými. Je zapotřebí přesněji rozlišovat mezi kauzalitou a pouhým korelačním vztahem, abychom mohli přijít ke správným závěrům.
- ▶ **Bohatství je správná odpověď na hádanku.**

T-testy

- **Testy o rovnosti středních hodnot dvou výběrů**
- **Jaký konkrétní t-test vybrat?**
- **varianta testu bude**
 - **parametrická (závislé, nezávislé soubory)**
 - **neparametrická (Wilcoxonův - závislé, Mann-Whitneyův test nezávislé hodnoty)**
- **Statistická vs. věcná významnost**

T-test

DVA NÁHODNÉ VÝBĚRY

NEZÁVISLÉ

ZÁVISLÉ

t-test pro
nezávislé výběry

t-test pro
závislé výběry

F-test

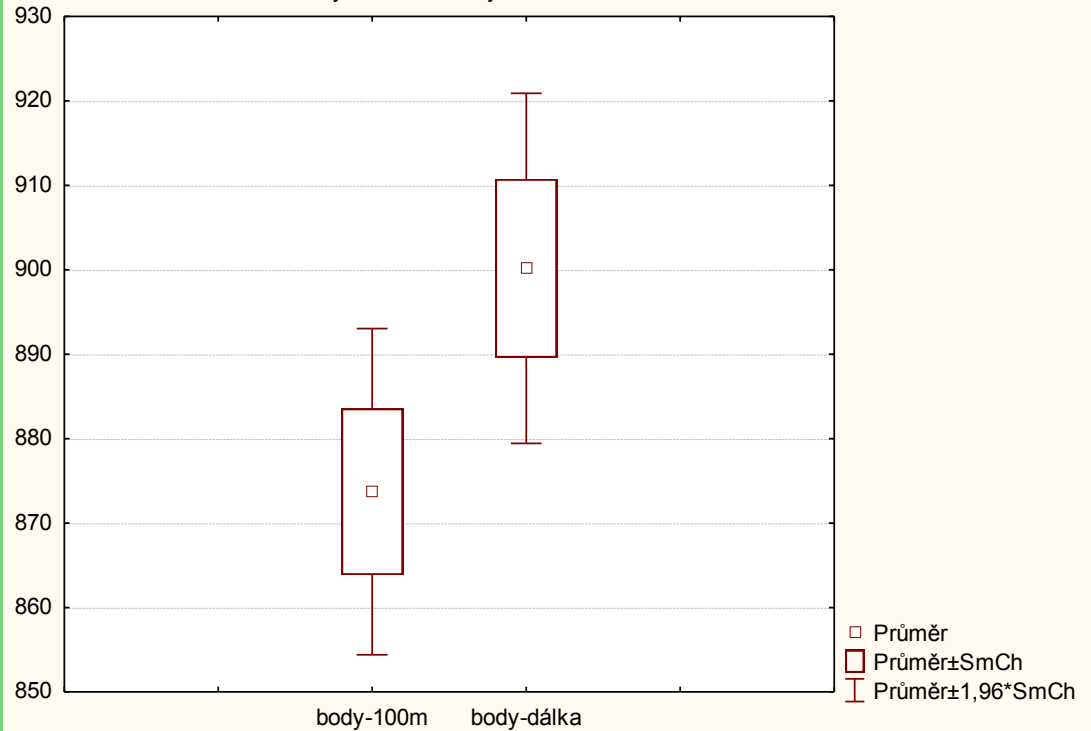
homogenní
rozptyl
 $s_1^2 = s_2^2$

heterogenní
rozptyl
 $s_1^2 \neq s_2^2$

t-test pro
homogenní
rozptyl

t-test pro
heterogenní
rozptyl

Krabicový graf
body-100m vs. body-dálka



T-test - příklad

Proměnná	t-test pro závislé vzorky (Tabulka1) Označ. rozdíly jsou významné na hlad. p < ,05000							
	Průměr	Sm.odch.	N	Rozdíl	Sm.odch. rozdílu	t	sv	p
body-100m	873,71	58,32						
body-dálka	900,17	62,56	35	-26,46	66,16	-2,37	34	0,02

- **Cohenovo d**

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}$$

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

- $d > 0,8 \rightarrow$ velký efekt
- d z intervalu $0,5 - 0,8 \rightarrow$ střední efekt
- $d < 0,2 \rightarrow$ malý efekt

- **$d = 0,44$**

- **rozdíl mezi oběma disciplínami je i věcně i statisticky významný.**

Statistický rozcestník ANEB co s daty

1. příprava výzkumného šetření je nejdůležitější část
2. sběr a analýza dat slouží k zamítnutí/nezamítnutí předem stanovených úkolů práce a hypotéz
3. vždy mít na paměti věcné hledisko výzkumu, zejména v souvislosti s interpretací statistických výsledků

CHCI S DATY PROVĚST	ZPŮSOB	UMOŽNÍ MI ZJISTIT
První náhled na data	<p>Základní popisná statistika</p> <ul style="list-style-type: none"> • průměr, směrodatná odchylka, rozptyl, N, medián, <u>kvartily</u> a další míry polohy a variability • tabulky četností: absolutní, relativní, kumulativní • grafy: krabicový, histogram, bodový 	<ul style="list-style-type: none"> • chybná měření, extrémny • homogenitu souboru • chybějící data • trend v datech
Otestovat normalitu	<ul style="list-style-type: none"> • <u>Kolmogorov-Smirnov test</u>, <u>Shapiro-Wilks test</u> 	<ul style="list-style-type: none"> • rozhodnutí, zda použít parametrické nebo neparametrické testy
Zjistit, zda výběry/skupiny jsou shodné nebo ne	<ul style="list-style-type: none"> • 2 skupiny/proměnné: t-testy • 3 a více skupin/proměnných: Analýza rozptylu (ANOVA) <ul style="list-style-type: none"> ○ T-testy i ANOVA má svou parametrickou i <u>neparametrickou variantu!</u> 	<ul style="list-style-type: none"> • konstatovat statisticky nebo věcně (<u>size of effect</u>) významný rozdíl Př. <u>došlo ke zlepšení výbušné síly po intervenci?</u> (<u>pretest-posttest</u>) Př. <u>kteřá ze dvou tréninkových metod je úspěšnější?</u> Př. <u>mezi kterými skupinami je statisticky významný rozdíl</u> Př. <u>byl zkoumán výsledný čas v motorickém testu v závislosti na typu <u>suplementace sportovce (faktor A)</u> a na <u>způsobu tréninku (faktor B)</u></u>
Zjistit závislost více proměnných (spojité)	<ul style="list-style-type: none"> • korelace, index determinace • faktorová analýza 	<ul style="list-style-type: none"> • těsnost lineárního vztahu mezi proměnnými • může existovat jasný vztah ale nelineární, který nezachytíme pomocí korelace nebo faktorové analýzy • korelace neznamená kauzalitu!!! Př. <u>závisí výkon v běhu na 100 m s výkonem do skoku do dálky?</u> Př. <u>závisí ekonomika běhu na povrchu?</u>
Zjistit závislost více proměnných (kategoriální-např. dotazník)	<ul style="list-style-type: none"> • test nezávislosti chí-kvadrát v kontingenční tabulce • vícerozměrné kontingenční tabulky - asociační stromy • shluková analýza • regresní a klasifikační stromy (CART, CHAID) 	<ul style="list-style-type: none"> • sílu a směr vztahu Př. <u>závisí bolestivost zad na věku a způsobu zaměstnání?</u> Př. <u>mezi kterými proměnnými z dotazníku existuje nejsilnější vazba?</u>
Redukovat velký počet vstupních dat	<ul style="list-style-type: none"> • faktorová analýza • analýza hlavních komponent 	<ul style="list-style-type: none"> • zda za naměřenými daty není nějaká latentní struktura (POZOR na interpretaci) Př. <u>lze 10 disciplín desetiboje popsat menším počtem faktorů?</u>
Vysvětlit závislou proměnnou několika nezávislými, provést předpověď	<ul style="list-style-type: none"> • lineární regrese • regresní a klasifikační stromy (CART, CHAID) • časové řady • neuronové sítě 	<ul style="list-style-type: none"> • příspěvek jednotlivých nezávislých proměnných k popisu proměnné závislé Př. <u>Popsat trend výkonnosti v atletických disciplínách a provést předpověď výkonů na olympiádě v Riu 2016</u>

Praktický postup

Nejprve se naučíme básničku:

„Je-li vypočítané $p \leq$ než α , zamítáme nulovou hypotézu H_0 “.

p hledáme ve vygenerovaných výsledcích; α je obvykle 0,05

Nulové hypotézy:

- Normalita - H_0 : data pocházejí z normálního rozdělení
- Korelace - H_0 : korelační koeficient je nulový
- T-test - H_0 : střední hodnoty dvou výběrů jsou shodné
- ANOVA (analýza rozptylu) - H_0 : střední hodnoty výběrů jsou shodné

Data ze studijních materiálů „Lekce1-Inbody“

SW Tibco Statistica

Normalita: Statistika → Základní statistiky → Tabulky četností → Normalita

Proměnná	Testy normality (lekce_1_inbody (1))					
	N	max D	K-S p	Lilliefors p	W	p
Fitness Score	1411	0,054043	p < ,01	p < ,01	0,947784	0,000000
Visceral fat area	1411	0,205259	p < ,01	p < ,01	0,354495	0,000000
Weight	1411	0,047451	p < ,01	p < ,01	0,973893	0,000000
kosterní svalstvo	1411	0,101266	p < ,01	p < ,01	0,961272	0,000000
% fat	1411	0,042703	p < ,05	p < ,01	0,986423	0,000000
WHR	1411	0,066478	p < ,01	p < ,01	0,794258	0,000000
BMI	1411	0,125418	p < ,01	p < ,01	0,531985	0,000000

Protože všechny p < 0,05; proto zamítáme nulovou hypotézu o normalitě. Proto budeme používat neparametrické testy.

Korelace: Statistika → Neparametrické statistiky → Korelace

Proměnná	Spearmanovy korelace (lekce_1_inbody (1))						
	Fitness Score	Visceral fat area	Weight	kosterní svalstvo	% fat	WHR	BMI
Fitness Score	1,000000	-0,238034	0,209298	0,615880	-0,761108	-0,143368	-0,100002
Visceral fat area	-0,238034	1,000000	0,697437	0,323742	0,455617	0,921003	0,790271
Weight	0,209298	0,697437	1,000000	0,814258	0,018833	0,750010	0,789797
kosterní svalstvo	0,615880	0,323742	0,814258	1,000000	-0,527791	0,369612	0,383703
% fat	-0,761108	0,455617	0,018833	-0,527791	1,000000	0,445227	0,474905
WHR	-0,143368	0,921003	0,750010	0,369612	0,445227	1,000000	0,916666
BMI	-0,100002	0,790271	0,789797	0,383703	0,474905	0,916666	1,000000

„červené“ výsledky jsou statisticky významné. Nejvyšší hodnota je 0,92

Data ze studijních materiálů „Lekce1-Inbody“

SW Tibco Statistica

T-test: Statistiky → Neparametrické statistiky → Porovnání dvou nezávislých vzorků → Mann-Whitneyův test → grupovací proměnná „Sex“ → závislá proměnná „kosterní svalstvo“

Mann-Whitneyův U Test (w/oprava na spojitost) (lekce_1_inbody (1)) Dle proměn. Sex Označené testy jsou významné na hladině $p < ,05000$									
Proměnná	Sčt poř. women	Sčt poř. men	U	Z	p-hodnot	Z upravené	p-hodnot	platných women	platných men
kosterní svalstvo	297150,0	699016,0	13269,00	-30,7076	0,00	-30,7076	0,00	753	658

Protože všechny $p < 0,05$; proto zamítáme nulovou hypotézu o shodě středních hodnot dvou výběrů, v našem případě mužů a žen. Rozdíl v proměnné „kosterní svalstvo“ je statisticky významný. Podle aritmetického průměru či mediánu bychom též poznali, že v prospěch mužů.

ANOVA: Statistiky → Neparametrické statistiky → Porovnání více nezávislých vzorků → Kruskal-Wallis → grupovací proměnná „Age“ → závislá proměnná „kosterní svalstvo“. Poté „vícenásobné porovnání p-hodnot“

Kruskal-Wallisova ANOVA založ. na poř.; kosterní svalstvo Nezávislá (grupovací) proměnná : Age Kruskal-Wallisův test: $H (2, N= 1411) =55,41275 p = ,0000$				
Závislá: kosterní svalstvo	Kód	Platný N	Součet pořadí	Průměr Pořadí
18-39	101	779	595064,5	763,8825
40-59	102	408	281024,5	688,7855
>60	103	224	120077,0	536,0580

Vícenásobné porovnání p hodnot (oboustr.); kosterní svalstvo Nezávislá (grupovací) proměnná : Age Kruskal-Wallisův test: $H (2, N= 1411) =55,41275 p = ,0000$			
Závislá: kosterní svalstvo	18-39 R:763,88	40-59 R:688,79	>60 R:536,06
18-39		0,007688	0,000000
40-59	0,007688		0,000020
>60	0,000000	0,000020	

$p=0,00$, což je $< 0,05$, proto zamítáme hypotézu o rovnosti středních hodnot 3 věkových kategorií. Kde je rozdíl nám řekne vícenásobné porovnání p-hodnot. Zde jsou všechny hodnoty „červené“, takže statisticky významné rozdíly jsou mezi všemi kombinacemi věkových kategorií.

Data ze studijních materiálů „Lekce1-Inbody“

SW Tibco Statistica

Cohenovo d: Nástroje → Marko → Makra → Otevřít → c:\Program Files\TIBCO\Statistica\Examples\Macros\Analysis Examples\T test with Cohen D.svb → Spustit

Proměnná	t-testy; Grupovací: Sex (lekce_1_inbody (1)) Skup. 1: women Skup. 2: men										
	Průměr women	Průměr men	Hodnota	SV	p	platných women	platných N men	sm.odch. women	sm.odch. men	F-poměr Rozptyly	p Rozptyly
kosterní svalstvo	25,73603	38,00693	-51,6413	1409	0,00	753	658	3,846161	5,058506	1,729775	0,000000

Statistica Visual Basic

Cohen's D is 2,73086714121566.

OK

Hodnota Cohenova d je 2,73, což podle doporučené interpretace a norem pro d značí výrazný efekt.

Tedy rozdíl v proměnné „kosterní svalstvo“ mezi muži a ženami je nejen statisticky, ale i věcně významný.

Zdroje:

- Cyhelský, L., Kahounová, J., & Hindls, R. (2001). *Elementární statistická analýza*. (2. dopl. vyd., 318 s.) Praha: Management Press.
- Hendl, J. (2006). *Přehled statistických metod zpracování dat: analýza a metaanalýza dat*. (Vyd. 2., opr., 583 s.) Praha: Portál.
- Meloun, M., & Militký, J. (1998). *Statistické zpracování experimentálních dat*. (2. vyd., xxi, 839 s.) Praha: East Publishing.
- Sebera, M. *Vícerozměrné statistiky*, 2013
- Zvonař, M., Pavlík, J., Sebera, M., Vespalec, T. & Štochl, J. *Vybrané kapitoly z antropomotoriky*. Brno: Masarykova univerzita, 2010.