



Univerzita Hradec Králové  
Pedagogická fakulta

# Studijní opora k předmětu kombinované formy studia

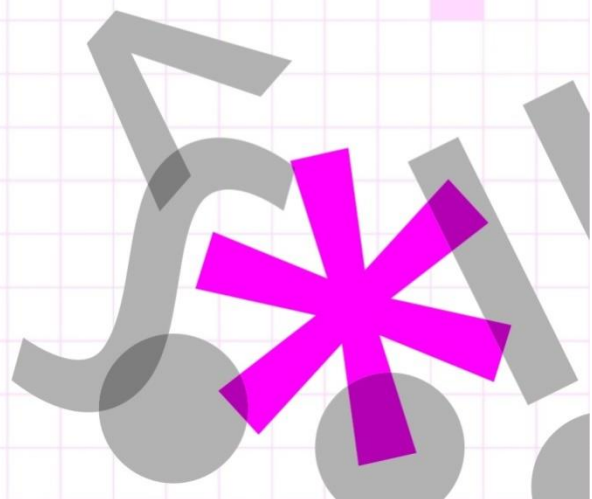
*předmět*

**VÝZKUMNÉ METODY A ZPRACOVÁNÍ DAT 2**

*obor*

**Sociální pedagogika**

Univerzita Hradec Králové  
Pedagogická fakulta



## Vysvětlivky k používaným symbolům



### ÚVOD KE STUDIU PŘEDMĚTU NEBO KONKRÉTNÍHO TÉMATU



### KONTROLNÍ OTÁZKY A ÚKOLY

*prověřují, do jaké míry student text a problematiku pochopil, zapamatoval si podstatné a důležité informace*



### SHRNUTÍ

*shrnutí tématu*



### POJMY K ZAPAMATOVÁNÍ

*vyzdvihuje důležité či nové termíny, nebo hlavní body, které by student při studiu tématu neměl opomenout*



### LITERATURA

*použitá ve studijním materiálu, pro doplnění a rozšíření poznatků*



### POŽADAVKY NA UKONČENÍ KURZU

*student se zde dozví, jakým způsobem bude kurz ukončen (test, seminární práce, projekt, ústní zkouška apod.)*



## ZÁKLADNÍ INFORMACE O PŘEDMĚTU

**Hodinová dotace přímé výuky:** 12 hodin

**Hodinová dotace samostudia:** 27 hodin

**Zařazení předmětu:** 2. semestr

**Prerekvizity:** Výzkumné metody a zpracování dat 2

**Pravidla komunikace s vyučujícím:** Při prvním setkání (řízené konzultaci) budou studentům poskytnuty základní informace o pojetí a zaměření předmětu a vysvětlení k využívání průvodce tématy k samostudiu. Dále budou seznámeni s možnostmi e-mailové komunikace, příp. osobní komunikace s vyučujícím v rámci konzultačních hodin, anebo diskusních bloků. Na závěrečném setkání budou diskutovány seminární práce, které bude potřebné odevzdat předem prostřednictvím studijního e-learnigového prostředí, ve kterém budou probíhat i diskusní bloky k vybraným tématům. Výukové metody – řízené konzultace, přednáška, panelové diskuse, závěrečná kolokvia k seminárním pracím.

## ÚVOD DO STUDIA PŘEDMĚTU (ANOTACE)

Předmět navazuje na Výzkumné metody 1 (jedná se o dvousemestrální předmět). Cílem je naučit studenty orientovat se a využívat základní statistické metody při vlastním šetření (závěrečná práce). Prakticky se seznámit s možnostmi zpracováním získaných dat z výzkumných šetření, statisticky je vyhodnotit a interpretovat, přehledně zpracovat a prezentovat s možností využití dostupných počítačových programů (SPSS, Excel, NCSS). Ukázky některých statistických šetření, kdy se za určitých předpokladů statisticky testují stanovené hypotézy apod.

## CÍLE PŘEDMĚTU:

Na konci tohoto předmětu bude student schopen zpracovat a vyhodnotit data získaná prostřednictvím kvantitativních metod a bude se dobře orientovat ve výzkumných metodách, ve způsobech základní analýzy a interpretaci získaných dat. Smyslem této disciplíny je prohloubení dosavadních vědomostí a dovedností studentů z oblasti zkoumání pedagogických jevů a posílení jejich samostatného metodologického uvažování.

## OSNOVA PŘEDMĚTU:

1. Pojem statistika, metody lidského poznání, fáze vědeckého výzkumu, měření v pedagogickém výzkumu, základní statistické pojmy – základní a výběrový soubor, znaky a jejich typy, zjišťování a zápis statistických dat.
2. Statistická šetření (deskriptivní, výběrová), formy sdělování výsledků (tabulky a sestavování tabulek četností, intervalové rozdělení četností, grafická znázornění statistického souboru, tvary rozdělení četností, atd.
3. Základní statistické charakteristiky – střední hodnoty (aritmetický průměr, Modus, medián), míry rozptýlenosti (rozptětí, průměrná a směrodatná odchylka, rozptyl, variační koeficient) a míry koncentrace (šikmost, špičatost a normální rozdělení).
4. Statistické metody užívané při testování platnosti hypotéz – věcné a statistické hypotézy ve výzkumu, zásady testování statistických hypotéz, příklady

statistických metod pro analýzu nominálních, ordinálních a metrických dat (chí-kvadrát, Fisherův kombinatorický test, znaménkový test, Wilcoxonův test, U-test, studentův t-test, atd.) – z části samostudium.

5. Možnosti určování těsnosti vztahu mezi proměnnými pomocí korelačních koeficientů (Spearmanův koeficient pořadové korelace, Kendallův koeficient shody, Pearsonův koeficient korelace) – z části samostudium.

6. Zpracování dat v MS Excel, SPSS, příp. ukázka možností NCSS. Zápis získaných dat, jejich uspořádání a příp. transformace, škálování, základní vyhodnocení dat v tabulkách prvního stupně třídění. Možnosti tabulek druhého a vyššího stupně třídění, pilotování tabulek. Grafické znázornění, možnosti grafů – rozptyl, .... – z části samostudium.



## DOPORUČENÁ LITERATURA A JINÉ ZDROJE

### ZÁKLADNÍ LITERATURA

- CHRÁSKA, M. *Metody pedagogického výzkumu*. Praha: Grada, 2007. 272 s.
- SKUTIL, M. a kol. *Základy pedagogicko-psychologického výzkumu pro studenty učitelství*. Praha: Portál, 2011. 256 s.

### DOPORUČENÁ LITERATURA

- HANOUSEK, J., CHARAMZA, P. *Moderní metody zpracování dat – matematická statistika pro každého*. Praha: Grada, 1992. 216 s.
- HENDL, J. *Kvalitativní výzkum*. Praha: Portál, 2005. 408 s.
- LAŠEK, J., MANĚNOVÁ, M. *Základy statistického zpracování pedagogicko-psychologického výzkumu*. Hradec Králové: Gaudeamus, 2003. 42 s.
- KOŘÍNEK, M. *Excel 2002*. České Budějovice: Kopp 2001. 246 s.
- ŠKALOUDOVÁ, A. *Statistika v pedagogickém a psychologickém výzkumu*. Praha: PdF UK Praha, 1998. 112 s.
- Studentům jsou k dispozici studijní materiály zveřejněné na portálu UHK a upravené metodické materiály k statistickému programu SPSS.



## POŽADAVKY NA UKONČENÍ KURZU

Předmět je ukončen zápočtem a zkouškou. Zápočet se sestává z obhajoby fiktivního výzkumného šetření zaměřeného k tématu závěrečné práce studenta. V projektu fiktivního šetření se předpokládá: vymezení problému a cíle; otázek či vstupních hypotéz šetření, včetně jejich zdůvodnění; zdůvodnění výběru metod pro sběr dat a vyhodnocení dat, charakteristika výzkumného vzorku (výstupem je podrobně rozpracovaný projekt výzkumu včetně event. realizovaného předvýzkumu). Dále, prezentace fiktivních dat, verifikace hypotéz na základě diskuse a vybraných statistických metod. Student též prokáže seznámení se základními dokumenty upravujícími tvorbu bakalářských/diplomových prací (Rozhodnutí děkana 17/2009 a v něm uváděné normy, Rektorský výnos č. 4/2009. Zkouška se skládá z písemného testu (3 příklady s cílem ověření znalostí z oblasti základních statistických charakteristik, statistických metod při testování platnosti hypotéz a možností určování těsnosti vztahů mezi proměnnými za pomoci vybraných/dostupných statistických programů. Zkouška je následně zakončena ústní rozpravou nad písemným testem s doplňujícími otázkami oblasti teorie (příklady viz ukázka možných otázek v závěru průvodce).



## ÚVOD A CÍL TÉMATU

### Pojem statistika

- slovu statistika je dáván nejrůznější význam – dotazníky a výkazy, číselné údaje, výsledky, organizace, teoretická disciplína (společenská X matematická)
- existuje množství definic, např. ze slovníku cizích slov „statistika = obor zabývající se sběrem, analýzou a zpracováním informací, které kvantitativně charakterizují zákonitosti jevů ve spojitosti s jejich kvalitativním obsahem“
- většina lidí statistiku ovládá na uživatelské úrovni, pro analýzy je lepší se spojit s odborníky
- existují, ale i lidé, kteří statistiku znevažují (dost často politici):
  - britský spisovatel a politik Benjamin Disraeli (1804 - 1881) údajně pronesl v parlamentu „Existují tři druhy lidí – prostá, odsouzeníhodná a statistika“
  - neméně slavný výrok řekl Winston Churchill (1874 – 1965) „Věřím, pouze té statistice, kterou si sám zfalšuji“.

- z latin. „*Status*“ (stav nebo stát) - původně k popisu státu = stav společnosti
- Anglie 16.-17. st. tzv. „*politická aritmetika*“
  - vzniká škola tzv. politických aritmetiků – politické a ekonomické argumenty musí pocházet z empirických dat (William Petty)
  - Graunt (pův. obchodník) napsal v r. 1662 demografickou studii „Přirozená a politická pozorování založená na seznamech zemědělců“
  - ze souboru různých lidských jedinců se vypočítává ideální střední jedinec (normální rozdělení, normální křivka, střední hodnota, rozptyl,...)
- 18. a 19. st. – pozorování a popis zákonitostí pozorovaných na tzv. hromadných jevech (bratři Bernoulliové, Langrange, Euler, Moivre, Gauss, Laplace, Bayes,...)
  - rozvoj pravděpodobnosti a variačního počtu
  - Gauss normální rozložení pravděpodobnosti
  - Poisson – rozdělení vhodné pro nízkou pravděpodobnost u velkého souboru
- až do poč. 20. st. tzv. *vyčerpávající šetření*
  - čím více čísel, tím lepší statistické výsledky
- 20. a 30. léta 20. st. *metody náhodného výběru* a dílčích šetření (Fischer, Person)
  - návrhy experimentů v biologii, lékařství a zemědělství – vznik biostatistiky a následně rozvoj statistiky v ekonomice - ekonometrie
  - rozpracování korelačních analýz, testů dobré shody
- Současnost – MS Excel, SPSS, NCSS, Statistica, STATGRAPHICS, ...
  - rozvoj počítačů (hlavně osobních), statistických programů, databází
  - přesnost, rychlost, grafické zobrazení dat i výsledků, možnost transformace některých veličin,...
  - v období socialismu se statistika u nás do značné míry ztotožňovala se zjišťování údajů o obyvatelstvu a k sestavování a kontrole plnění plánů

- Statistika = vědní obor, metoda, informace

### **Statistika v současnosti**

1. Vědní obor – deskriptivní + induktivní
2. Metoda sběru, zpracování a vyhodnocování dat
3. Informace

Předmětem zkoumání statistiky ve společenských vědách je člověk

### **Formulace hypotézy**

- pokusné
- předběžné
- prozatímní odpovědi na položené otázky (problémy)

### **Pravidla stanovení hypotézy (H) podle Gavory**

- H je tvrzení, v oznamovací větě (výzkumný problém je naopak lepší vyjádřit tázací větou)
- H musí vyjadřovat vztah mezi dvěma proměnnými – vždy je to o rozdílech, vztazích nebo následcích
- H musí být možno empiricky ověřitelné, proměnné musí být měřitelné
- H jsou vlastně predikcí o vztazích mezi proměnnými - málokdy je to důsledek jediného faktoru

### **Chyby při formulacích H**

- Nesprávná, neurčitá formulace
- Složitě souvětí
- Věcná hypotéza X statistická hypotéza

**Proměnné - xi** (je to jev nebo vlastnost; ve výzkumu se mění – věk, klasifikace)

Rozlišujeme:

- Nezávisle proměnné = jev, vlastnost, která je příčinou nebo podmínkou vzniku jiné vlastnosti, jevu
- Závisle proměnné = je vlastnost, jev, která je výsledkem působení nezávislé proměnné

**Testování / verifikace hypotézy** - prokazujeme pravdivost nebo nepravdivost hypotézy

Rozhodujeme na základě:

- třídění
- zpracování
- vyhodnocení shromážděných dat

Data shromažďujeme od „X“ respondentů - tj. od výzkumného vzorku

- základní soubor – populace
- výběrový soubor – výběr

Výběr prvků do výzkumných souborů = volba jedinců – situací, jejich počtu ...

**Druhy výběrů**

- Prostý náhodný výběr (náhodná čísla)  
Výběr s vracením  
Výběr bez vracení
- Skupinový výběr
- Stratifikovaný výběr
- Kontrolovaný výběr
- Vícenásobný výběr
- Záměrný výběr
- Mechanický výběr
- Spárované výběry

### Rozsah (velikost) výběru

Čím větší soubor pořídíme, tím více se blížíme skutečným vlastnostem základního souboru

Odhady rozsahu výběru

- u metrických dat:  $n = (t_{\alpha} \cdot s_x) / \Delta_x$
- u nominálních či ordinálních dat:  $n = [t_{\alpha} \cdot p \cdot (1 - p)] / d_x$

### Měření v pedagogickém výzkumu

„Měření v nejširším slova smyslu je přiřazování čísel předmětům nebo jevům podle pravidel“ (Stevens, 1951, s. 51)

Platí 3 postuláty

- Jestliže  $(a = b) \vee (a \neq b)$  ne však oboje
- Jestliže  $(a = b) \wedge (b = c) \Rightarrow (a = c)$
- Jestliže  $(a > b) \wedge (b > c) \Rightarrow (a > c)$

### Úrovně měření

- Nominální (*tj. označkování*)
- Ordinální (*pořadové*)
- Kardinální (*metrické*)  
Intervalové  
Poměrové

### Vlastnosti dobrého měření:

- Validita
- Reliabilita
- Praktičnost – jednoduchost, hospodárnost, ....

**Způsob výuky:** téma je přednášeno z části a je proto nutné samostudium



## SHRNUTÍ A TÉMATA KE STUDIU

Pojem statistika, metody lidského poznání, fáze vědeckého výzkumu, měření v pedagogickém výzkumu, základní statistické pojmy – základní a výběrový soubor,

znaky a jejich typy, zjišťování a zápis statistických dat.



### DOPORUČENÁ LITERATURA A JINÉ ZDROJE

- HANOUSEK, J., CHARAMZA, P. *Moderní metody zpracování dat – matematická statistika pro každého*. Praha: Grada, 1992. 216 s.
- CHRÁSKA, M. *Metody pedagogického výzkumu*. Praha: Grada, 2007. 272 s.
- SKUTIL, M. a kol. *Základy pedagogicko-psychologického výzkumu pro studenty učitelství*. Praha: Portál, 2011. 256 s.

Studentům jsou k dispozici studijní materiály zveřejněné na portálu UHK a upravené metodické materiály k statistickému programu SPSS.



### KONTROLNÍ OTÁZKY A ÚKOLY

1. Popište vývoj statistiky.
2. Jaké má významy statistika v současné době
3. Jaká jsou pravidla pro stanovení hypotéz?
4. Uveďte nejčastější chyby při formulaci hypotéz.
5. Jaké jsou vlastnosti dobrého měření?

*Prostor pro odpovědi:*

---

---

---

---

---

---

---

---

---

---





## ÚVOD A CÍL TÉMATU

### **Fáze klasického pedagogického výzkumu**

Ve výzkumu se řeší buď jeden nebo více (zpravidla spolu souvisejících) problémů. Řešení vědeckého problému potom představuje řadu navzájem propojených a na sobě závislých kroků a činností. Jednotlivé výzkumy se mohou navzájem lišit co do posloupností jednotlivých realizovaných činností, ale základní schéma postupu bývá následující:

- a) stanovení problému,
- b) formulace hypotézy,
- c) testování (verifikace, ověřování) hypotézy,
- d) vyvození závěrů a jejich prezentace.

### **Metody uspořádání a zpracování dat (tzv. popisná statistika)**

#### **Uspořádání a sestavování tabulek**

- Čárkovací metoda
- Interval - jeho hloubka a střed
- Zásady tvorby tabulek
- Četnost – absolutní, relativní, kumulativní
- Využití MS Excel – s přenosem dat do NCSS, SPSS, ....

#### **Četnostní tabulky - základní pojmy**

- Znak –  $x_i$
- Četnost –  $n_i$
- Relativní četnost –  $n_i/n$
- Kumulativní četnost  $n_1, n_1+n_2, \dots$
- Kumulativní relativní četnost  $p_1, p_1+p_2, \dots$

#### **Intervalové rozdělení četností**

Obor všech možných hodnot sledovaného znaku rozdělíme do vzájemně se vylučujících intervalů – tříd

Čím větší rozsah sledovaného souboru – tím větší počet intervalů (max. 15 – pro přehlednost)

**Výpočet intervalu** – pro diskrétní náhodnou veličinu  $h = 0,08 \times R$

#### **Grafické metody zobrazování dat**

- Histogramy četností (sloupcový graf)
- Polygony četností (spojnicový graf)
- Výšečové grafy
- Kartografy

## Histogram

- Sloupcový graf
- Osa x – jednotlivé naměřené hodnoty
- Osa y – četnosti hodnot (absolutní či relativní)

**Polygon** - četnosti spojujeme úsečkami ve středu jednotlivých intervalů

## Stromový graf

### Krabicový graf (kvantil k)

- 25% kvantil = dolní kvartil
- 50% kvantil = medián
- 75% kvantil = horní kvartil
- 10% kvantily = decily
- 100% kvantily = percentily

(Ukázky grafického zpracování získaných dat pomocí MS Excel je uvedeno podrobně v 6 kapitole tohoto podkladu)!

**Způsob výuky:** téma je přednášeno v rámci výuky v celém rozsahu



## SHRNUTÍ A TÉMATA KE STUDIU

Statistická šetření (deskriptivní, výběrová), formy sdělování výsledků (tabulky a sestavování tabulek četností, intervalové rozdělení četností, grafická znázornění statistického souboru, tvary rozdělení četností, atd.



## DOPORUČENÁ LITERATURA A JINÉ ZDROJE

- HANOUSEK, J., CHARAMZA, P. *Moderní metody zpracování dat – matematická statistika pro každého*. Praha: Grada, 1992. 216 s.
- CHRÁSKA, M. *Metody pedagogického výzkumu*. Praha: Grada, 2007. 272 s.
- SKUTIL, M. a kol. *Základy pedagogicko-psychologického výzkumu pro studenty učitelství*. Praha: Portál, 2011. 256 s.

Studentům jsou k dispozici studijní materiály zveřejněné na portálu UHK a upravené metodické materiály k statistickému programu SPSS.



## KONTROLNÍ OTÁZKY A ÚKOLY

1. Pokuste se o ukázkou zpracování tabulky prvního a druhého stupně třídění.
2. Uveďte, pro která data je jaký typ grafů vhodný?
3. Využijte možností všech typů grafů a znázorněte v nich vhodná data.

*Prostor pro odpovědi:*

---

---

## Téma č. 3:

## Základní statistické charakteristiky



### ÚVOD A CÍL TĚMATU

#### **Základní statistické charakteristiky** (číselný popis dat)

- Střední hodnoty – charakteristiky polohy
- Míry rozptýlenosti - variability
- Míry koncentrace

#### **Charakteristiky polohy**

##### **Modus** (Mode)

- označení
- nejčastěji se vyskytující hodnota (nejčetnější)
- může odhalit nehomogenitu výběru
- neříká nic o extrémních hodnotách

##### **Medián** (Median)

- označení
- prostřední hodnota v řadě hodnot uspořádaných podle velikosti
- používá se jako charakteristika polohy, chceme-li odstranit vliv extrémních hodnot

##### **Aritmetický průměr** (Mean)

označení  
má velký význam, nelze však přeceňovat  
citlivý na extrémní hodnoty

V symetrickém rozdělení se modus, medián i aritmetický průměr shodují!

#### Další charakteristiky polohy

- Vážený průměr

- Useknutý průměr
- Harmonický průměr (Harmonic Mean)
- Geometrický průměr (Geometric Mean)

### Míry variability

#### Rozpětí (Range)

- označení **R**
- výpočet max. hodnota – min. hodnota
- značně ovlivněno extrémními hodnotami

#### Mezikvartilové rozpětí (Interquartile Range)

- výpočet horní kvartil – dolní kvartil
- délka obdélníka v krabicovém grafu
- není ovlivněno extrémními hodnotami

#### Střední kvadratická odchylka, rozptyl

- doplňuje průměr
- rozdělení se stejným průměrem může být více – liší se rozptylem

#### Směrodatná odchylka (Standard Deviation)

- spolu s rozptylem nejužívanější doplnění průměru
- kritérium věrohodnosti průměru

#### Variační koeficient (Coefficient of Variation)

- bezrozměrný
- pro porovnání variability hodnot měřených v různých jednotkách
- orientačně signalizuje případnou hrubou nesourodost dat

Další míry variability

#### Průměrná odchylka a Relativní průměrná odchylka

#### Míry koncentrace

**Šikmost** (angl. Skewness) označení  $S_m$

- $S_m = 0$  rozdělení symetrické
- $S_m > 0$  **zešikmené zprava** (kladné hodnoty šikmosti)
- $S_m < 0$  **zešikmené zleva** (záporné hodnoty šikmosti)

**Špičatost** (angl. Kurtosis – někdy také Exces) označení  $K_m$

- $K_m = 0$  normované normální rozdělení
- $K_m > 0$  **špičatost** (větší četnosti prostředních hodnot)
- $K_m < 0$  **plochost** (přibližně stejně velké četnosti prostředních a ostatních hodnot)

#### Normální rozdělení - Gaussova křivka normálního rozložení

pravděpodobnost výskytu hodnot

- V intervalu od  $-S$  do  $+S$  (kolem aritm. prům.) se nachází přibližně 2/3 (68,27%) všech hodnot
- V intervalu od  $-2S$  do  $+2S$  (kolem  $\bar{X}$ ) se nachází přibližně 19/20 (95,4%)
- V intervalu od  $-3S$  do  $+3S$  (kolem  $\bar{X}$ ) se nachází téměř všechny hodnoty (99,73%)

**Způsob výuky:** téma je přednášeno v rámci výuky částečně, je potřebné samostudium



## POJMY K ZAPAMATOVÁNÍ

- Střední hodnoty, míry rozptýlenosti, míry koncentrace
- Průměr, medián, modus
- Rozpětí
- Rozptyl, standardní směrodatná odchylka, variační koeficient
- Průměrná odchylka, relativní průměrná odchylka
- Šikmost, špičatost
- Gaussova křivka normálního rozložení



## SHRNUTÍ A TÉMATA KE STUDIU

Základní statistické charakteristiky – střední hodnoty (aritmetický průměr, Modus, medián), míry rozptýlenosti (rozpětí, průměrná a směrodatná odchylka, rozptyl, variační koeficient) a míry koncentrace (šikmost, špičatost a normální rozdělení).



## DOPORUČENÁ LITERATURA A JINÉ ZDROJE

- HANOUSEK, J., CHARAMZA, P. *Moderní metody zpracování dat – matematická statistika pro každého*. Praha: Grada, 1992. 216 s.
- HENDL, J. *Kvalitativní výzkum*. Praha: Portál, 2005. 408 s.
- CHRÁSKA, M. *Metody pedagogického výzkumu*. Praha: Grada, 2007. 272 s.
- LAŠEK, J., MANĚNOVÁ, M. *Základy statistického zpracování pedagogicko-psychologického výzkumu*. Hradec Králové: Gaudeamus, 2003. 42 s.
- SKUTIL, M. a kol. *Základy pedagogicko-psychologického výzkumu pro studenty učitelství*. Praha: Portál, 2011. 256 s.
- ŠKALOUDOVÁ, A. *Statistika v pedagogickém a psychologickém výzkumu*. Praha: PdF UK Praha, 1998. 112 s.



## KONTROLNÍ OTÁZKY A ÚKOLY

1. Následující tabulka uvádí odpovědi respondentů – žáků, kdy (v jaké třídě) jim poprvé byla nabídnuta cigareta. Vypočítejte průměr, modus, medián, průměrnou i standardní odchylku včetně jejich vyjádření v %. Zkuste načrtnout alespoň jedno grafické vyjádření.

třída	Počty žáků
1.	3
2.	6
3.	10
4.	33
5.	45
6.	29
7.	17

8.	4
9.	3

**Výsledky:**

modus je 5. třída

medián 5.tá třída

aritmetický průměr 5,0267 - tedy opět 5.tá třída

průměrná odchylka  $d = 1,14$  (tzn. , že odchylka činí  $\pm 1$  třídu okolo 5. té třídy)

relativní průměrná odchylka  $rd = 22,68 \%$

rozptyl  $S^2 = 2,37$

směrodatná odchylka  $S = 1,54$

2. Spočítejte průměrnou měsíční a čtvrtletní spotřebu, modus a medián, variační rozpětí, standardní odchylku včetně vyjádření v % (vždy vzhledem k měsíci i ke čtvrtletí). Zkuste načrtnout grafická vyjádření. Vycházejte z údajů získaných podle měsíců (1 = leden, 2 = únor, ...).

1 leden	15 litrů týdně
2	18
3	9
4	12
5	21
6	25
7	28
8	23
9	16
10	12
11	15
12 prosinec	11

**Výsledky:**

Po měsících:

modus 7 měsíc (28 litrů týdně)

medián 10 tí hodnota, opět 7 měsíc

aritmetický průměr 17,08 litru týdně

rozptyl  $S^2 = 122,77$

směrodatná odchylka  $S = 11,08$

variační koeficient  $V = 64,87\%$

Po čtvrtletích (začala jsem zimou – prosinec + leden + únor; atd.):

modus léto

medián léto

aritmetický průměr 51,25 litru týdně

rozptyl  $S^2 = 2370,29$

směrodatná odchylka  $S = 48,69$

variační koeficient  $V = 95,00 \%$

3. V okrese Jeseník byl sledován výskyt dopravních nehod podle čtvrtletí – viz. tabulka s počty dopravních nehod. Pohybuje se stabilita výskytu dopravních nehod ve sledovaných letech pod 40%?

	1.čtvrtletí	2. čtvrtletí	3. čtvrtletí	4.čtvrtletí
2005	195	115	92	129
2006	223	110	88	115
2007	214	107	105	130

### Výsledky:

2005 = 28,84% (S = 38,29)

2006 = 39,09% (S = 52,37)

2007 = 31,94% (S = 44,40)

Ano.

## Téma č. 4:

## Statistické metody užívané při testování platnosti hypotéz



### ÚVOD A CÍL TÉMATU

**Přehled používaných testů pro statistické testování hypotéz** - Využito podkladů a příkladů z publikace Skutíla (2011)

Hypotéza je slovo složené z latinské předpony *hypo*, označující „zmenšení“, a slova *thesis*, označujícího „výrok, tvrzení“.

**Hypotéza může vzniknout pouze třemi způsoby:**

1. Tvrzení na základě empirické zkušenosti: *Například známe tři nositelky jména Marie a všechny studují matematiku, postulujeme si tedy hypotézu, že všechny Marie jsou matematicky;*
2. Tvrzení odvozená z teorie: *Například teorie sociální atraktivity tvrdí, že si vybíráme protějšky na stejné úrovni atraktivity, kterou si myslíme, že máme sami. Marie je učitelka a připadá si průměrně atraktivní. Předpokládáme tedy, že její partner nebude milionář;*
3. Tvrzení odvozená z publikovaných výsledků jiných výzkumů: *Například výzkum tvrdí, že běžný Evropan utratí za vánoční dárky v přepočtu průměrně 10 000 korun, a Češi jsou Evropané, takže očekáváme, že Češi průměrně utratí stejnou částku.*

Všechna takto vzniklá tvrzení něco podstatného postrádají. Jsou to pouhé názory, které je žádoucí **ověřit**, otestovat je. Lze to učinit dvěma způsoby: buď prověříme celou zkoumanou populaci, tedy základní soubor, nebo se budeme snažit prověřit tyto názory na datech získaných z vybrané skupiny, tedy výběrového souboru, s vědomím toho, že konkrétní výběr může získané výsledky ovlivnit, čili že výběr nebude reprezentativní. Pokud budeme pracovat se základním souborem, stačí nám k ověření či vyvrácení hypotéz deskriptivní statistiky, které již byly popsány. Zde se budeme zabývat pouze situacemi, v nichž z výsledků výběru usuzujeme na situaci v celé populaci.

Při kvantitativním zpracování dat se hypotézy nejčastěji vztahují k dvojici proměnných a snaží se zjistit, zda „je mezi nimi určitá souvislost“, či zda „mezi nimi souvislost není“. Tvrzení jsme dali do uvozovek, protože je potřeba je precizovat. Je zvykem stanovovat dvě hypotézy: hypotézu nulovou a hypotézu alternativní, která je negací hypotézy nulové. Nulová hypotéza je obvykle tvrzení, které předpokládá, že rozdíl mezi skupinami je nulový (odtud název nulová hypotéza). Stanovuje se proto, že ji lze matematicky vyvrátit (falzifikovat).

*Příklad: Mějme dvě proměnné. První proměnná „pohlaví“ je nominální a druhá proměnná, bodový výsledek v určitém testu matematických znalostí, je kardinální s normálním rozdělením.*

Cílem je zjistit, zda chlapci mají v tomto matematickém testu odlišnou úspěšnost než dívky. To je nutné přeformulovat do podoby korektní testovatelné hypotézy (pomocí tzv. operacionalizace), která pro tyto dvě proměnné bude znít: “Průměrný dosažený výsledek je v obou skupinách shodný” (hypotéza nulová). Alternativní hypotézou je tvrzení, že průměrný dosažený výsledek se v obou skupinách liší.

Matematicky má příslušná procedura, v uvedeném příkladě nepárový t-test (viz dále), podobu **funkce**, jejímž **vstupem** jsou jednak **parametry** (počet mužů a počet žen v souboru), jednak **získaná data** (výsledky jednotlivých osob) a **výstupem** je **reálné číslo**, v našem příkladě označované  $t$ . Toto  $t$  je konkrétní číslo, které popisuje náš výběrový soubor.

Nyní potřebujeme (zjednodušeně řečeno) zjistit, jakých hodnot  $t$  může nabývat funkce (t-test) pro všechny možné výběry ze základního souboru a s jakou četností se jednotlivá teoretická  $t$  vyskytují. Předpokládáme přitom, že v základním souboru platí nulová hypotéza (tedy např. že pro všechny žáky v České republice je průměrný dosažený výsledek u chlapců i u dívek stejný), protože pro takovýto soubor lze matematicky odvodit četnosti jednotlivých teoretických  $t$  (a většina statistického softwaru má tento výpočet zabudován).

Teoretickou množinu všech možných  $t$  spolu s jejich četnostmi nazýváme **rozdělení** (v našem příkladu Studentovo rozdělení či t-rozdělení, viz Hendl, 2004) a lze jej popsat základními percentilovými statistikami, tedy mediánem, kvartily apod. Nyní budeme s tímto rozdělením, resp. s jeho percentily, porovnávat naše reálné  $t$ . Pro statistiky je významný 95% percentil označovaný  $t_{0,95}$ , tedy hodnota taková, že 95 % možných hodnot  $t$  je rovno nebo menších než  $t_{0,95}$ . Tato hodnota se označuje jako **kritická hodnota pětiprocentní hladiny významnosti** nebo jen stručněji **pětiprocentní hladina významnosti**. Analogicky je používán 99% percentil, označovaný jako jednoprocenní hladina významnosti. Představme si, že reálné  $t = 6,2$ , teoretické  $t_{0,95} = 3,5$ . V takovém případě můžeme říci, že nulovou hypotézu zamítáme na hladině významnosti 5 %.

Analogicky pro reálné  $t = 3,1$  a teoretické  $t_{0,95} = 3,5$  nulovou hypotézu zamítnout nelze. Vzhledem k uvedenému rozdělení můžeme určovat **pravděpodobnost  $p$ , že náš vybraný soubor dat pochází ze základního souboru, ve kterém platí nulová hypotéza**.

Hodnota  $p$  je určena jako  $(1 - q)$ , kde  $q$  je hodnota kvantilu odpovídající našemu nalezenému  $t$ . Některý software umožňuje určit hodnotu pravděpodobnosti přímo. Pravděpodobnost  $p$  může nabývat hodnoty  $0 - 1$ , a pokud je rovna nebo menší než  $0,05$  (resp.  $0,01$ ), lze nulovou hypotézu zamítnout na hladině významnosti 5 % (resp. 1 %). Podrobněji viz Chráska, 2007.

**Nulovou hypotézu můžeme zamítnout, ale nemůžeme ji potvrdit.** Je tomu tak proto, že nalezené  $t$  porovnááme s teoretickým rozdělením  $t$  pro základní soubor, ve kterém platí nulová hypotéza. O rozdělení  $t$  v základním souboru, v němž nulová hypotéza neplatí, nevíme „nic“. Naše nalezené  $t$  tak může vyhovovat nejen základnímu souboru s platnou nulovou hypotézou (v našem příkladě takovému, v němž mají stejný průměr muži i ženy), ale i jiným základním souborům (např. takovému, v němž ženy dosahují větších průměrných hodnot než muži). Výzkumník proto vítá, když může nulovou hypotézu zamítnout, a tak může **navrhnout hypotézu alternativní**.

#### **Postup statistického testování hypotéz**

Nejdůležitější úlohou výzkumníka v této části kvantitativního výzkumu je **stanovit**



**metody** (tzv. inferenční statistiky), **formulovat příslušné nulové hypotézy** a na jejich základě stanovit, jaké údaje je třeba od respondentů získat. Interval spolehlivosti obsahuje s předem danou pravděpodobností hodnotu nějakého parametru základního souboru (např. jeho průměru). Stanovená hodnota s danou pravděpodobností tedy leží uvnitř tohoto intervalu.

Čtyři základní metody, totiž test chí-kvadrát, t-test, analýzu rozptylu a korelační koeficient (koeficient korelace) budeme nyní demonstrovat na příkladech, ostatní lze dohledat v další literatuře – nejlépe u Hendla, 2004.

### **Chí-kvadrát (test nezávislosti)**

*Příklad: Intuitivní představou výzkumníka je, že respondenti z hornatých krajů preferují jiné sporty než respondenti z krajů rovinatých.*

Respondenti odpovídají v dotazníku na otázku: „Jaký typ sportu preferujete?“ s možnostmi odpovědí:

- senzomotorický (střelba),
- funkčně-mobilizační (běh na 100 m),
- anticipační (míčové sporty).

Dále odpovídají na druhou otázku: „Ze kterého kraje pocházíte?“

Obě zjištěné charakteristiky jsou nominální, proto použijeme metodu chí-kvadrát. Údaje zapíšeme do **kontingenční tabulky** (pomocí níž můžeme vizualizovat vztah dvou statistických znaků). Předpokladem pro provedení metody je četnost větší než 5 v každé možné dvojici nominálních hodnot (v každé buňce tabulky). Nulová hypotéza je pro tuto metodu taková, že obě zjištěné veličiny jsou nezávislé, tedy že spolu nesouvisí. V našem případě by byla formulována ve tvaru: „Preferovaný typ sportu nesouvisí s tím, z jakého kraje respondent pochází“. Alternativní hypotéza má tedy tvar: „Preferovaný

typ sportu souvisí s tím, ze kterého kraje respondent pochází.“ Výpočet probíhá pomocí rozdílů mezi skutečnými četnostmi a očekávanými četnostmi. Oba tyto údaje získáme pomocí počítačových programů (MS Excel, Statistica atd.).

Konkrétní postup bude uveden v tématu č. 6. Pro nás je podstatné, že pokud vyjde hodnota chí-kvadrát vypočtená z konkrétních dat větší než hodnota teoreticky vypočtená, pak můžeme nulovou hypotézu zamítnout. Žádnou podrobnější informaci tato metoda nedává, pomocí testu chí-kvadrát nelze např. vyčíst, zda jedna proměnná je nezávislá (příčina) a druhá na ní závisí (následek).

Postup výpočtu chí-kvadrátu:

- proměnné: (nominální, nominální),
- předpoklady: četnost každé kombinace příslušnosti k jednotlivým kategoriím je větší než 5,
- testovaná charakteristika:  $\chi^2$  (tedy chí-kvadrát),
- interval spolehlivosti:  $(0; \chi_{0,95}^2)$ , odvozen z teoretického rozdělení četností  $\chi^2$ ,
- nulová hypotéza: „rozdělení do skupin v obou proměnných je nezávislé“,
- test zamítnutí nulové hypotézy:  $\chi_{0,95}^2 < \chi^2$ .

### **T-test**

*Příklad: Intuitivní představou je, že ženy dosahují v běhu na 100 m odlišných výsledků než muži. Získáme tedy hodnoty proměnné „pohlaví“ a „výsledek v běhu na 100 m“.*

Pohlaví je proměnná nominální dichotomická a výsledek běhu proměnná kardinální.

Pro porovnání takových proměnných použijeme t-test.

Pro t-test platí:

- proměnné: (nominální dichotomická, kardinální),
- předpoklad: kardinální proměnná má normální rozdělení,
- testovaná charakteristika:  $t$ ,
- interval spolehlivosti:  $(-\infty; t_{0,95})$ , odvozen z teoretického rozdělení četností  $t$ ,
- nulová hypotéza: „průměrný dosažený výsledek je v obou skupinách shodný“,
- test zamítnutí nulové hypotézy:  $t_{0,95} < t$ .

Tuto metodu lze nalézt ve více variantách, námi popisovaná je takzvaná dvouvýběrová (nepárová) varianta s jednostranným intervalem spolehlivosti. Pro podrobnosti o jednotlivých typech viz Hendl, 2004.

### **Analýza rozptylu (ANOVA)**

*Příklad: Máme představu, že výsledek běhu na 100 m závisí na tom, jaký typ sportu respondenti preferují. Mohou si volit z těchto variant:*

- anaerobní – posilovna,
- aerobní – běh,
- rekreační – turistika.

Typ sportu je proměnná nominální polytomická a výsledek v běhu na 100 m je proměnná kardinální. Metodou vhodnou pro tyto typy proměnných je analýza rozptylu, pro kterou platí:

- proměnné: (nominální polytomická, kardinální),
- předpoklad: kardinální proměnná má normální rozdělení,
- testovaná charakteristika:  $F$ ,
- interval spolehlivosti:  $(0; F_{0,95})$
- nulová hypotéza: „průměr je ve všech skupinách shodný“, alternativní hypotéza: „alespoň ve dvou skupinách je rozdíl průměrů statisticky průkazný“,
- test zamítnutí nulové hypotézy:  $F_{0,95} < F$ .

### **Korelační koeficient**

*Příklad: Porovnáváme výšku u jednotlivých žáků a jejich výsledek v běhu na 100 m. Obě proměnné jsou kardinální, vhodnou procedurou bude korelační koeficient. Pro ten platí:*

- proměnné: (kardinální, kardinální),
- předpoklad: obě proměnné mají normální rozdělení,
- testovaná charakteristika:  $\gamma$ ,
- nulová hypotéza: „proměnné jsou nezávislé“,
- interval spolehlivosti: výpočet intervalu spolehlivosti a korektního odůvodnění zamítnutí nulové hypotézy je záležitostí přepočtu koeficientu  $r$  na koeficient  $t$  a výpočet teoretického  $t_{0,95}$ , který jsme popisovali u t-testu.

V sociálních vědách se však často používá přímo interpretace koeficientu korelace. Interpretace korelačního koeficientu (Chráška, 2007):

- $r = 1$  naprostá závislost (funkční závislost)
- $1,00 > r > 0,90$  velmi vysoká závislost
- $0,90 > r > 0,70$  vysoká závislost
- $0,70 > r > 0,40$  střední závislost

- $0,40 > r > 0,20$  nízká závislost
- $0,20 > r > 0,00$  velmi slabá závislost
- $r = 0$  naprostá nezávislost

Upozorníme ještě na fakt, že korelační koeficient může nabývat hodnot mezi  $-1$  a  $1$ . Pokud by v našem příkladě měl koeficient  $r$  hodnoty blízké  $1$ , pak by to znamenalo, že „čím větší je člověk, tím delšího času dosáhl v běhu“, což nepředpokládáme (má delší nohy, tedy mohl by běžet rychleji a mít kratší čas). Pokud je naopak hodnota  $r$  blízká  $-1$ , proměnné spolu opět souvisí, ovšem „obráceně“. V našem příkladě by to znamenalo, že „čím větší člověk, tím dosáhne kratšího času v běhu“, což asi bude naším předpokladem. Záporné hodnoty se tedy interpretují podobně. Hodnoty koeficientu blízké nule znamenají lineární nezávislost.

Jiný příklad na interpretaci záporného korelačního koeficientu může být následující.

*Příklad : Respondenti reagovali na výroky „Mám rád pokrmy z vepřového masa“ a „Mám rád zeleninová jídla“ na škále : „nesouhlasím  $-2$ ,  $-1$ ,  $0$ ,  $1$ ,  $2$  souhlasím“.*

Korelační koeficient  $r$  dosáhl hodnoty  $-0,92$ , tedy velmi výrazné negativní závislosti. Znamená to, že kdybychom změnili výrok „Mám rád zeleninová jídla“ na „Nemám rád zeleninová jídla“, vyšel by koeficient korelace  $+0,92$ , tedy velmi výrazná závislost. V obou případech je interpretace stejná – respondenti mají rádi pokrmy z vepřového masa zjevně nemají v oblibě zeleninová jídla.

Přehled základních metod pro jednotlivé typy proměnných dle Hendla (2004)

Typ proměnných	Nominální dichotomická	Nominální polytomická	Ordinální	Kardinální s normálním rozdělením
Nominální dichotomická	Chí-kvadrát	Chí-kvadrát	Wilcoxonův test	t-test (Studentův)
Nominální polytomická	Chí-kvadrát	Chí-kvadrát	Kruskalova-Wallisova analýza rozptylu	Analýza rozptylu (ANOVA)
Ordinální	Wilcoxonův test	Kruskalova-Wallisova analýza rozptylu	Kendalův koeficient korelace	Kendalův koeficient korelace
Kardinální s normálním rozdělením	t-test (Studentův)	Analýza rozptylu (ANOVA)	Kendalův koeficient korelace	Koeficient korelace (Pearsonův)

**Způsob výuky:** téma je přednášeno v rámci výuky v celém rozsahu



## POJMY K ZAPAMATOVÁNÍ

- Typy proměnných – nominální, ordinální, kardinální
- chí-kvadrát,
- Fisherův kombinatorický test,

- znaménkový test,
- Wilcoxonův test,
- U-test,
- studentův t-test



## SHRNUTÍ A TÉMATA KE STUDIU

Statistické metody užívané při testování platnosti hypotéz – věcné a statistické hypotézy ve výzkumu, zásady testování statistických hypotéz, příklady statistických metod pro analýzu nominálních, ordinálních a metrických dat (chí-kvadrát, Fisherův kombinatorický test, znaménkový test, Wilcoxonův test, U-test, studentův t-test, atd.) – z části samostudium.



## DOPORUČENÁ LITERATURA A JINÉ ZDROJE

- HENDL, J. *Kvalitativní výzkum*. Praha: Portál, 2005. 408 s.
- CHRÁSKA, M. *Metody pedagogického výzkumu*. Praha: Grada, 2007. 272 s.
- LAŠEK, J., MANĚNOVÁ, M. *Základy statistického zpracování pedagogicko-psychologického výzkumu*. Hradec Králové: Gaudeamus, 2003. 42 s.
- SKUTIL, M. a kol. *Základy pedagogicko-psychologického výzkumu pro studenty učitelství*. Praha: Portál, 2011. 256 s.
- ŠKALOUDOVÁ, A. *Statistika v pedagogickém a psychologickém výzkumu*. Praha: PdF UK Praha, 1998. 112 s.



## KONTROLNÍ OTÁZKY A ÚKOLY

1. Na základě údajů uvedených v tabulce rozhodněte, zda existuje závislost mezi roky a počtem firem ve Šluknovském výběžku, které zaměstnávaly osoby zdravotně postižené (dále jen OZP) a které nezaměstnávaly OZP.

Plnění povinného podílu	r.2004	v %	r.2005	v %	r.2006	v %
se zaměstnáváním OZP	94	76	97	87	88	86
bez zaměstnávání OZP	29	24	15	13	14	14
celkem firem	123	100	12	100	102	100

**Výsledek:**  $\chi^2 = 5,514 < 9,21 = \chi^2_{0,01} (2)$  ..... přijímáme nulovou H

2. Rozhodněte, zda existuje závislost mezi absolvováním rekvalifikačního kurzu a nástupem do zaměstnání (popř. rokem), pokud víme, že v roce 2004 nastoupilo do zaměstnání po kurzu rekvalifikace 24 lidí a v roce 2006 již 117. Bez rekvalifikace nastoupilo do zaměstnání v roce 2004 – 199 lidí a v roce 2006 jen 151 lidí.

	r.2004 nastoupilo do zaměstnání	r.2006 nastoupilo do zaměstnání	celkem
s rekvalifikací	24	117	141
bez revalifikace	199	151	350
celkem	223	268	491

**Výsledek:**  $\chi^2 = 64,339 > 6,635 = \chi^2_{0,01} (1)$  ..... přijímáme alternativní H

3. V regionu Kutné hory byli respondenti dotazováni na preference v oblasti zájmových činností provozovaných ve volném čase. Rozhodněte zda mezi preferencí zájmové činnosti a pohlavím existuje statisticky významná souvislost.

Zájmové činnosti	Pohlaví	
	chlapci	dívky
estetické	8	26
tělovýchovné	67	62
pracovní- technické	23	3
společensko- vědní	6	6
přírodovědné	14	8

**Výsledek:**  $\chi^2 = 26,07496 > 9,483 = \chi^2_{0,05} (4)$  ..... přijímáme alternativní H

4. Rozhodněte, zda mezi sledovanými celními úřady je v názorech na důležitost profesního vzdělávání významný vztah.

Celní úřad	Názor na důležitost vzdělání		
	Důležitý	Neím	Nedůležitý
TU	47	21	14
NA	66	45	23
PCE	71	35	6

**Výsledek:**  $\chi^2 = 11,062 > 9,483 = \chi^2_{0,05} (4)$  ..... přijímáme alternativní H

*Prostor pro odpovědi:*

---

---

---

---

---

---

---

---

---

---



## ÚVOD A CÍL TÉMATU

### Stupeň závislosti mezi jevy při ordinálním měření

- **Koeficient kontingence (C)** – lze použít jako míry závislosti v kontingenční tabulce
 
$$C = \sqrt{[\chi^2 : (n + \chi^2)]}$$
 n ... celková četnost v kontingenční tabulce  
 $\chi^2$  ... vypočítaná hodnota testového kritéria chí-kvadrát pro kont. tab.
  - koeficient kontingence C nabývá hodnot v intervalu od 0 do +1, přičemž platí, že čím vyšší je hodnota, tím vyšší je stupeň závislosti,
  - nevýhodou tohoto koeficientu je, že jeho hodnota závisí na počtu řádků a na počtu sloupců v kontingenční tabulce a tudíž dva koeficienty kontingence nelze srovnávat, pokud byly vypočítány z tabulek, které nemají stejné rozměry (tj. stejný počet řádků a sloupců).
- **Normovaný koeficient kontingence (C<sub>norm.</sub>)**
  - do jisté míry odstraňuje nevýhody předchozího koeficientu,
  - je definován jako poměr mezi dosaženou hodnotou koeficientu kontingence a největší možnou hodnotou C<sub>max.</sub>, které může koeficient dosáhnout.
 
$$C_{max.} = \sqrt{[(r - 1) : r]}$$
 r ... je menší z počtu řádků nebo sloupců
 
$$C_{norm.} = C : C_{max.}$$
    - tato hodnota nebývá příliš spolehlivá, zvláště u tabulek, kde je velký rozdíl mezi počtem řádků a sloupců,
    - i při porovnávání hodnot C norm. pro dvě kontingenční tabulky je potřeba, aby mezi celkovými četnostmi nebyl příliš velký rozdíl.
- **Čuprovův koeficient (K)**
  - v případě tabulky s rozdílným počtem řádků a sloupců bývá doporučován
 
$$K = \sqrt{\{\sqrt{\chi^2} : \sqrt{[n \cdot (r - 1) \cdot (s - 1)]}\}}$$

Těsnost vztahu pro jevy ve čtyřpolní tabulce se ověřuje pomocí koeficientů:

- **Fí - koeficient (r $\phi$ )**
  - je vhodný k vyjádření těsnosti vztahu mezi oběma alternativními znaky v čtyřpolní tabulce:

$$r\phi = (A \cdot D - B \cdot C) : \sqrt{[(A+B) \cdot (A+C) \cdot (B+D) \cdot (D+C)]}$$

- kde význam písmen je shodný s významem ve čtyřpolní tabulce,
- pokud máme pro čtyřpolní tabulku již vypočítáno testové kritérium  $\chi^2$ ,

můžeme tento koeficient vypočítat:

$$r\phi = \sqrt{(\chi^2 : n)}$$

- hodnota tohoto koeficientu se může pohybovat od -1 do +1, přičemž platí, že čím vyšší je vypočítaná absolutní hodnota, tím vyšší je závislost mezi znaky; znaménko vypočítaného koeficientu je dáno znaménkem výrazu  $A.D - B.C$  a nemá pro interpretaci výsledku význam,
- $\phi$ -koeficient nebývá vždy spolehlivým ukazatelem stupně závislosti mezi sledovanými jevy, vlastnostmi – jeho využití není vhodné v některých extrémních případech.

- **Yulův koeficient asociace (Q)**

- někdy bývá doporučován k ověření těsnosti vztahu mezi znaky v čtyřpolní tabulce:

$$Q = (A.D - B.C) : (A.D + B.C)$$

- kde význam písmen je shodný s významem ve čtyřpolní tabulce,
- Yulův koeficient může nabývat opět hodnot od 0 do  $\pm 1$ , ale na rozdíl od  $\phi$ -koeficientu zachycuje jen jednostrannou závislost mezi jevy a poskytuje tedy i jiné výsledky než  $\phi$ -koeficient (každý deficient je odvozen z jiných předpokladů)

### **Statistické metody pro analýzu ordinálních dat**

Jestliže ověřujeme vztah mezi jevy, které byly změřeny na úrovni ordinálního (pořadového) měření. Můžeme k tomu použít po vhodné kategorizaci všech postupů uvedených pro nominální data, ale můžeme využít také postupů, které byly vyvinuty speciálně pro data ordinální:

- Znaménkový test
- Wilcoxonův test (T)
- U – test Manna a Whitneyho (U, U')
  - U test pro velmi malé výběry (četnosti ve srovnávaných skupinách jsou menší než 8)
  - U test pro větší skupiny (četnosti ve srovnávaných skupinách jsou do 20)
  - U test při velkých četnostech
- Kolmogorovův – Smirnovův test (D)
- Kruskalův – Wallisův test (je zobecněním U testu) (H)

Pokud se potvrdí určitá souvislost mezi jevy, tzv. přijímáme  $H_A$ , zkoumáme stupeň závislosti mezi jevy pomocí koeficientů korelace:

- Spearmanův koeficient pořadové korelace
- Kendallův koeficient shody (W)

### **Statistické metody pro analýzu kardinálních (metrických) dat**

Při analýze intervalových nebo poměrových (kardinálních/metrických) dat je možno využívat všech postupů, které byly uvedeny v případě dat nominálních i ordinálních. Data „vyššího“ typu lze totiž vždy (zavedením vhodného typu kategorizace) převést na data „nižšího“ typu. Takový postup je však zpravidla doprovázen určitou ztrátou informace. Využívat můžeme:

- Funkční a statistická závislost mezi jevy
- Regresní a korelační analýza
- Pearsonův koeficient korelace ( $r_p$ )
- Bodová biseriální korelace ( $r_{bb}$ )
- Biseriální korelace ( $r_{bis}$ )
- Tetrachordický koeficient korelace ( $r_{tet}$ )
- Studentův t – test (t)
- Fisherův – Snedecorův F - test (F)
- Párový t – test
- Princip analýzy rozptylu
  - Jednoduchá analýza rozptylu, Duncanův test
  - Dvoufaktorová analýza rozptylu

### **Principy faktorové analýzy**

Faktorová analýza je metoda, která umožňuje určit základní proměnné (faktory), které ovlivňují provedená měření určitého objektu. Jestliže u určitého objektu provedeme několik různých měření, může se stát, že se některá měření budou navzájem více či méně podobat. Ta měření, která dávají podobné výsledky, je dobré studovat společně, protože vyjadřují něco společného (společný faktor). Faktorová analýza umožňuje rozhodnout, které základní faktory (proměnné) se provedených měřeních projevují, a tak umožňuje nahradit velký počet provedených měření několika základními faktory.

Zakladatelem faktorové analýzy je anglický psycholog Charles Spearman, který se pomocí ní pokusil objasnit strukturu lidské inteligence. Zjistil, že základ inteligence tvoří několik faktorů, které bývají označovány např. jako verbální schopnosti, matematické schopnosti, abstraktní usuzování, prostorová představivost, paměť, atd.

### **Shluková analýza**

Shluková analýza je metoda, jejímž cílem je přiřadit jednotky analýzy (např. osoby, případy, události apod.) na základě podobnosti ke skupinám (shlukům). Přitom charakteristiky shluků ani jejich počet není předem znám – musí být odvozeny z výzkumných dat. Mírou podobnosti (resp. nepodobnosti) jednotek analýzy bývá často korelace nebo tzv. euklidovská distance D.

Lze využít i tzv. hierarchického shlukování a shlukování metodou k-průměrů.

### **Metaanalýza v pedagogickém výzkumu**

Jedním ze základních fenoménů dnešní doby je stále se zrychlující tempo narůstání vědeckých poznatků (v této souvislosti se často hovoří o informační explozi). Problém získávání syntetických informací z mnoha dílčích poznatků pomáhá řešit jedna z moderních oblastí statistiky, tzv. metaanalýza. Jejím cílem je shrnout výsledky dvou nebo více empirických výzkumů, které se zabývají stejným nebo podobným problémem (Hendl, 2004).

**Způsob výuky:** téma je přednášeno částečně, je potřebné doplnění – viz doporučené zdroje





## SHRNUTÍ A TÉMATA KE STUDIU

Možnosti určování těsnosti vztahu mezi proměnnými pomocí korelačních koeficientů (Spearmanův koeficient pořadové korelace, Kendallův koeficient shody, Pearsonův koeficient korelace) – z části samostudium.



## DOPORUČENÁ LITERATURA A JINÉ ZDROJE

- HENDL, J. *Kvalitativní výzkum*. Praha: Portál, 2005. 408 s.
- CHRÁSKA, M. *Metody pedagogického výzkumu*. Praha: Grada, 2007. 272 s.
- LAŠEK, J., MANĚNOVÁ, M. *Základy statistického zpracování pedagogicko-psychologického výzkumu*. Hradec Králové: Gaudeamus, 2003. 42 s.
- SKUTIL, M. a kol. *Základy pedagogicko-psychologického výzkumu pro studenty učitelství*. Praha: Portál, 2011. 256 s.

Téma č. 6:

**Zpracování dat ve vybraných statistických programech**



## ÚVOD A CÍL TÉMATU

*Využito podkladů a příkladů z publikace Skutila (2011)*

### **Zpracování dat v programu MS Excel 2007**

Základní zpracování dat lze provést v programu MS Excel, jehož součástí jsou i statistické funkce, které můžeme použít pro popis souboru a pro výpočty na jednorozměrné i dvourozměrné úrovni.

Na elementárním příkladu si ukážeme, jakým způsobem můžeme data prezentovat.

#### *Příklad 1*

Na základní škole na začátku roku v tělesné výchově proběhlo měření výšky žáků. Byly naměřeny následující údaje (v cm):

154, 156, 171, 168, 158, 166, 178, 159, 160, 166, 168, 149, 158, 164, 163, 171, 169, 172, 156, 174, 147, 167, 169, 173, 164, 157, 149, 153, 168.

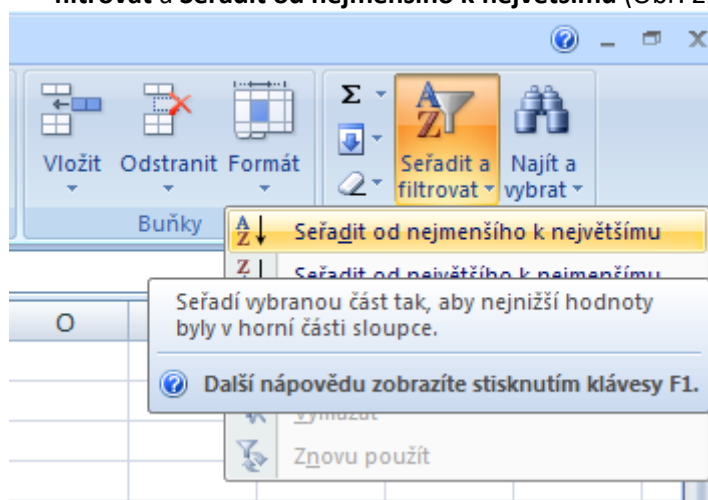
### **Četnostní tabulka**

V Excelu vytvoříme tabulku, kam zaznamenáme jednotlivá získaná data (co žák, to řádek).

A	B	C	D
	výška [cm]		
	154		
	156		
	171		
	168		
	158		
	166		
	178		
	159		
	160		
	166		
	167		
	149		
	158		
	164		
	163		
	171		
	169		
	172		
	156		
	171		

Obr. 1 Základní zaznamenání dat

Data seřadíme vzestupně (případně sestupně) tak, že data označíme a klikneme na **Seřadit a filtrovat** a **Seřadit od nejmenšího k největšímu** (Obr. 2).



Obr. 2 Seřazení dat

Vzhledem k variační šíři dat, je vhodnější zobrazit četnosti v určitých intervalech. Výpočet optimální hloubky intervalu uvádí např. Chráska, 2007. Procházíme jednotlivé intervaly a zaznamenáváme pomocí čárek výskyt dat v příslušném intervalu. Výsledek čárkování převedeme do tabulky (Obr. 3).

H	I	J	K
<b>Intervaly</b>		<b>Absolutní četnost</b>	
140-149	III	3	
150-159	IIIIIIII	8	
160-169	IIIIIIIIIIII	12	
170-179	IIIIII	6	

Obr. 3 Sestavení tabulky četností

Počet žáků, jejichž výška se nachází v daném intervalu, nazýváme absolutní četnost (uvádí se i pouze **četnost**). Celkový počet dat (součet) získáme kliknutím na tlačítko suma  $\Sigma$  a označením sloupce s absolutními četnostmi (Obr. 4).

Intervaly		Absolutní četnost
140-149	III	3
150-159	IIIIIIII	8
160-169	IIIIIIIIIIII	12
170-179	IIIIII	6
	$\Sigma$	29

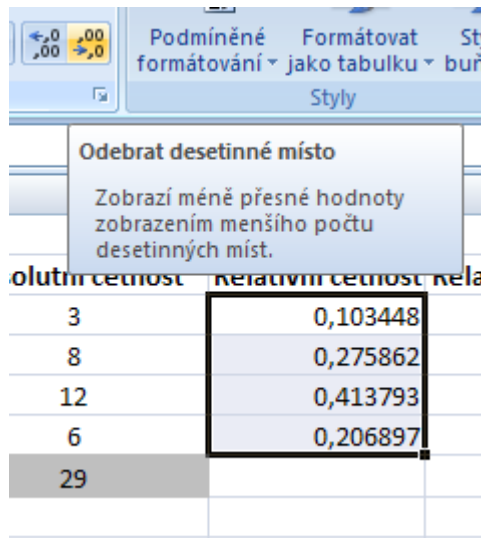
Obr. 4 Součet četností

Někdy bývá pro interpretaci výsledků vhodnější pracovat s **relativními četnostmi**, které vyjadřují, jak velká část z celkového počtu hodnot připadá na daný interval (hodnotu, kategorii). (Chráška, 2007) Relativní četnosti je možné vypočítat dle postupu uvedeném na následující obrázku (Obr. 5).

ČETNOSTI					
	A	B	C	D	E
1					
2	<b>Intervaly</b>		<b>Absolutní četnost</b>	<b>Relativní četnost</b>	
3	140-149	III	3	=C3/C7	
4	150-159	IIIIIIII	8		
5	160-169	IIIIIIIIIIII	12		
6	170-179	IIIIII	6		
7		$\Sigma$	29		
8					

Obr. 5 Relativní četnost

U výsledku upravíme počet desetinných míst (Obr. 6).



Obr. 6 Úprava počtu desetinných míst

Pokud získané údaje (relativní četnost) vynásobíme 100, dostaneme relativní četnost vyjádřenou v procentech (Obr. 7)

Intervaly	Absolutní četnost	Relativní četnost v %
140-149	3	10
150-159	8	28
160-169	12	41
170-179	6	21
	29	100

Obr. 7 Relativní četnost v %

Pokud máme jednotlivá data zapsaná v tabulce v Excelu pod sebou, můžeme **absolutní četnosti** určit přímo pomocí funkce v Excelu.

#### Příklad 2

Máme známky z matematiky jedné třídy na konci školního roku. Zajímají nás četnosti jednotlivých známek. Data (známku každého studenta) zapíšeme do tabulky a vedle sloupce s daty zapíšeme varianty (obměny) sledovaného znaku „známka“ (Obr. 8).


B	C	D
známka na vysvědčení		známka
3		1
4		2
2		3
3		4
2		5
3		
4		
1		
3		
2		
2		
3		
3		
4		
5		
1		
3		
3		
4		
3		
2		
4		
2		
3		
3		

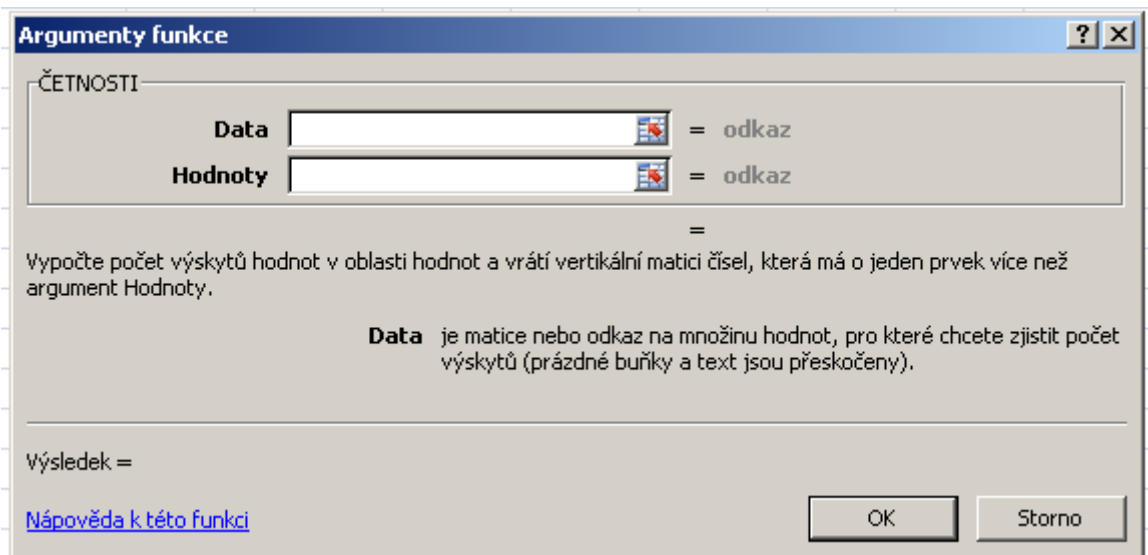
Obr. 8 Základní data a možnosti hodnot dat (známek)

Vedle sloupce **známka** označíme sloupec, kam přijdou vložit získané četnosti (Obr. 9).

známka	
1	
2	
3	
4	
5	

Obr. 9 Označení pole pro výsledek

Klikneme na tlačítko , zadáme Statistické a vybereme ČETNOSTI. Objeví se dialogové okno Argumenty funkce (Obr. 10).



Obr. 10 Okno Argumenty funkce

Klikneme do řádku **Data** a označíme pomocí kurzoru data ve sloupci **známka na vysvědčení**. Pak klikneme na řádek **Hodnoty** a označíme kurzorem čísla ve sloupci **známka**. Potvrdíme současným stiskem kláves **CTRL**, **SHIFT** a **ENTER**. Vedle sloupce známka získáme absolutní četnosti jednotlivých známek (Obr. 11).

fx {=ČETNOSTI(B2:B26;D2:D6)}

	B	C	D	E	F
známka na vysvědčení			známka		
	3		1	2	
	4		2	6	
	2		3	11	
	3		4	5	
	2		5	1	
	3				


Obr. 11 Výsledek výpočtu

Pokud chceme získat absolutní četnosti v určitých intervalech (viz příklad 1), musíme zadat horní hraniční hodnoty intervalů (Obr. 12).

C3 fx 149

	A	B	C	D	E
výška [cm]			intervaly		
	147		149		
	149		159		
	149		169		
	153		179		
	154				
	156				
	156				

Obr. 12 Intervalové rozložení četností

První interval je v rozsahu 0-149, druhý 150-159 atd. Dále je postup stejný, označíme sloupec, kam chceme umístit absolutní četnosti, klikneme na tlačítko , vybereme **Statistické a ČETNOSTI**. Označíme příslušný sloupec pro řádek **Data** a meze intervalů pro řádek **Hodnoty**. Potvrdíme současným stiskem kláves **CTRL, SHIFT** a **ENTER**. Zobrazí se nám absolutní četnosti v jednotlivých intervalech (Obr. 13).

D3      fx {=ČETNOSTI(A3:A31;C3:C6)}

A	B	C	D	E	F	G
výška [cm]		intervaly				
147		149	3			
149		159	8			
149		169	12			
153		179	6			
154						
156						
156						

Obr. 13 Výsledek výpočtu – četnosti dat v jednotlivých intervalech

#### Grafické znázornění dat

Získaná data lze prezentovat i graficky. Můžeme vytvořit:

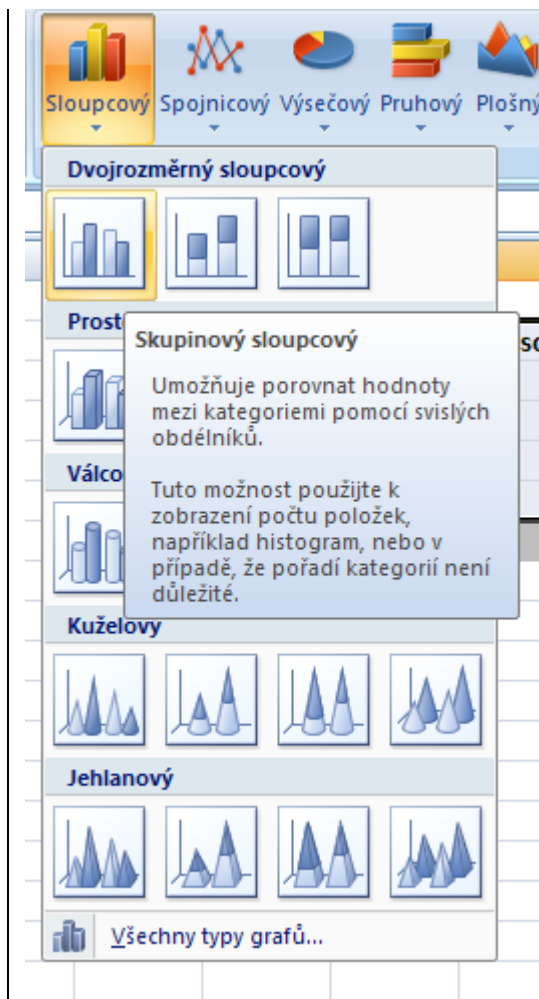
- Histogram četností
- Polygon četností
- Výsečový (koláčový) graf

Histogram četností je sloupcový graf, kde na vodorovné ose jsou zobrazeny jednotlivé naměřené hodnoty (intervaly) a na ose svislé pak četnosti (absolutní nebo relativní). V Excelu histogram vytvoříme tak, že vyjdeme z tabulky četností, kde označíme hodnoty (intervaly) a příslušné četnosti (Obr. 14)

Intervaly	Absolutní četnost
140-149	3
150-159	8
160-169	12
170-179	6
	29

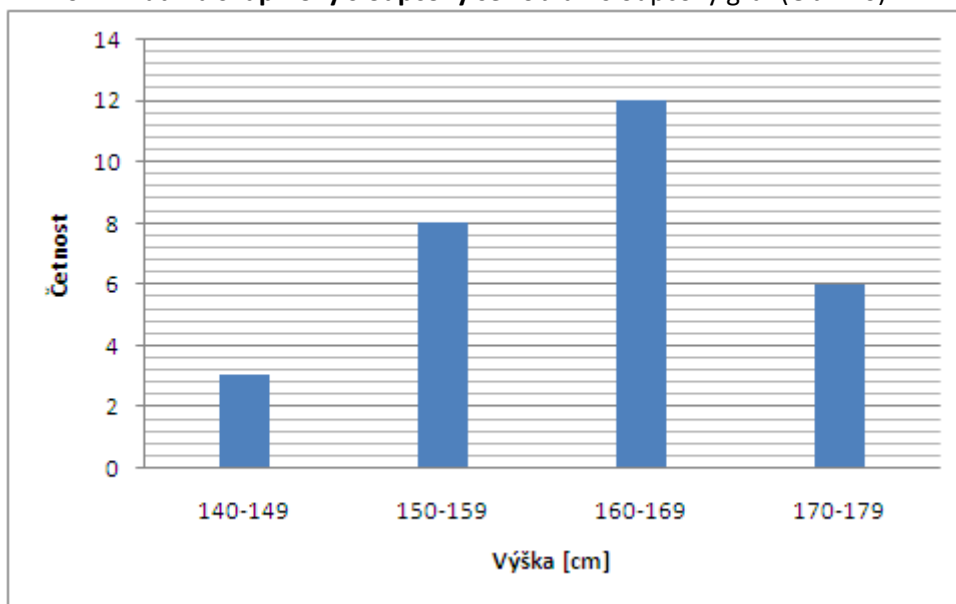
Obr. 14 Tabulka četností

Na záložce **Vložení** zvolíme nabídku **Sloupcový** a z nabídky sloupcových grafů vybereme tvar (Obr. 15).



Obr. 15 Okno s nabídkou grafů

Po kliknutí na **skupinový sloupcový** se zobrazí sloupcový graf (Obr. 16).

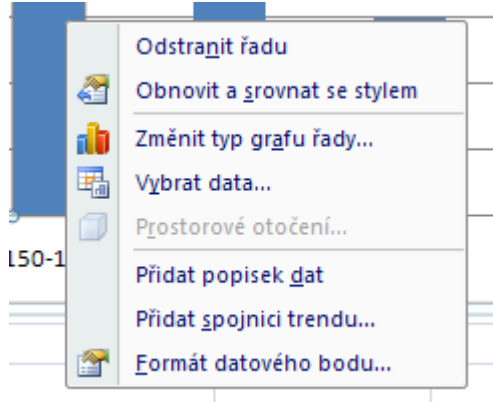


Obr. 16 Sloupcový graf

Sloupcový graf upravíme na histogram, když mezery mezi sloupci nastavíme na 0. Pravým

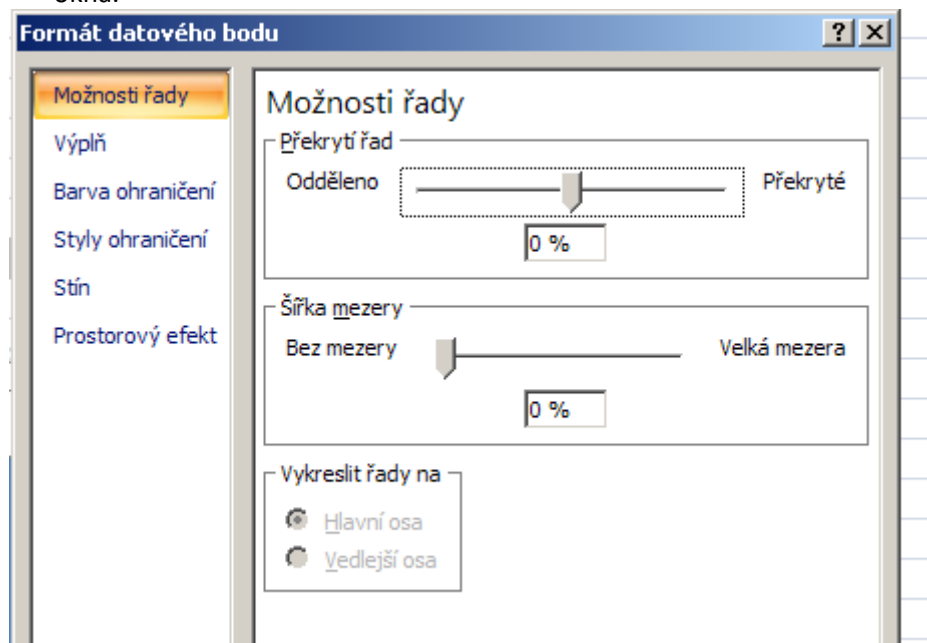


tlačítkem myši klikneme na jeden ze sloupců, objeví se nabídka (Obr. 17).

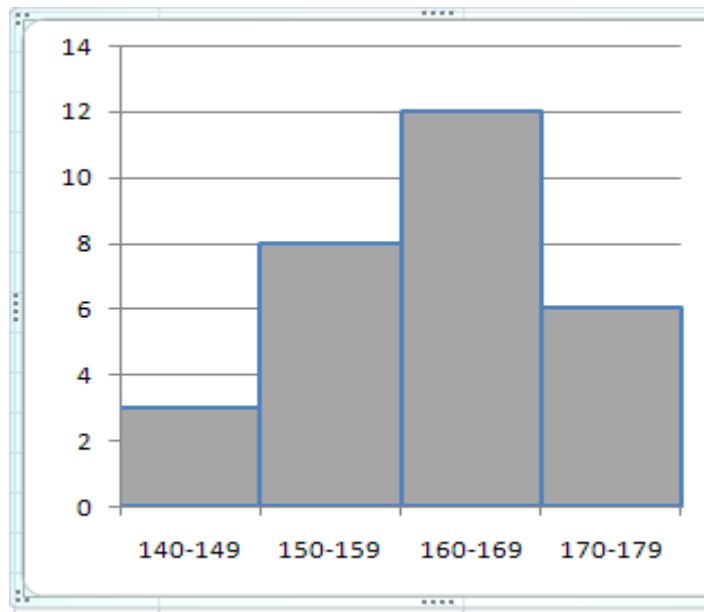


Obr. 17 Úprava sloupcového grafu

Zvolíme **Formát datového bodu** a v následujícím okně nastavíme nulovou šířku mezery (Obr. 18). Získáme histogram, který lze dále graficky upravit podle nabídky tohoto dialogového okna.



Obr. 18 Okno s nabídkou úprav grafu



Obr. 19 Výsledný histogram

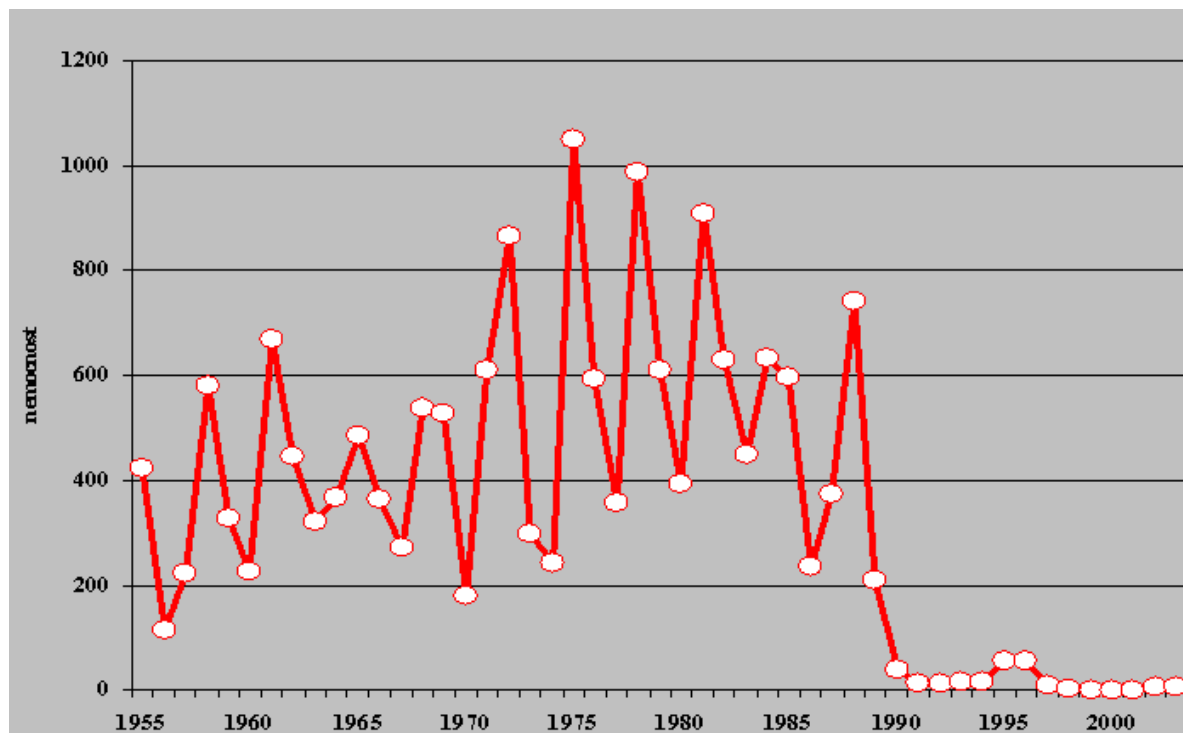
### Polygon četností

Polygon četností je spojnicový graf a lze použít pouze v případě, že hodnoty na ose x lze jednoznačně seřadit. V Excelu tento typ grafu vytvoříme tak, že označíme data a na záložce

**Vložení** zvolíme **Spojnicový** (Obr. 20) a vybereme vhodný typ (Obr. 21).



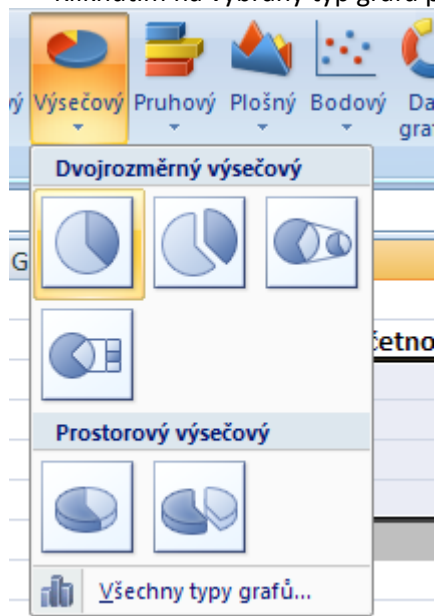
Obr. 20 Okno s nabídkou spojnicových grafů



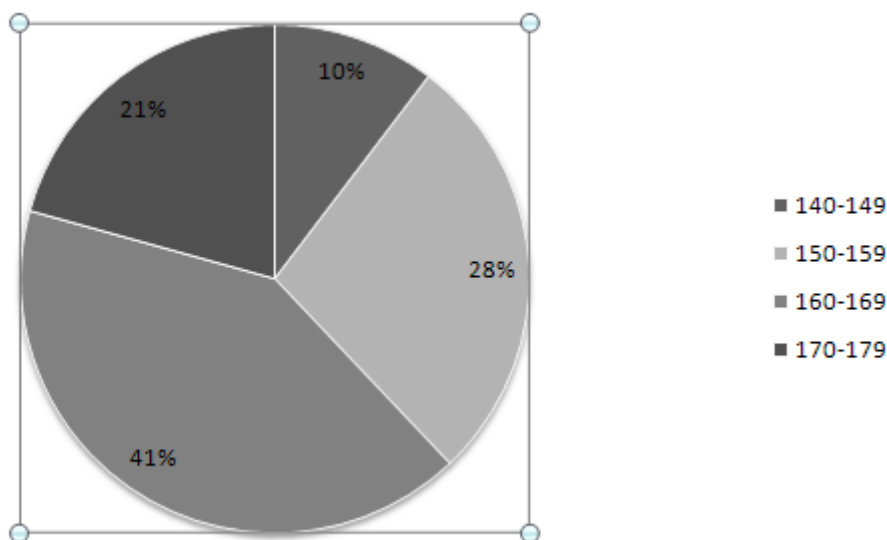
Obr. 21 Ukázka polygonu četnosti

### Výšečový graf

Výšečový graf se používá pro znázornění struktury složení výběrového souboru (Chráska, 2007), například rozložení dle pohlaví, věkové struktury apod. Postup vytvoření výšečového grafu je obdobný jako u sloupcového grafu, na záložce **Vložení** zvolíme **Výšečový** (Obr. 22). Kliknutím na vybraný typ grafu příslušný graf vytvoříme (Obr. 23).



Obr. 22 Okno s nabídkou výšečových grafů



Obr. 23 Výsledný výšečový graf

Další grafické úpravy pak můžeme provádět pomocí lišty (Obr. 24)



Obr. 24 Možnosti úprav výšečového grafu

### Základní veličiny popisné statistiky


Popis dat pomocí tabulek a grafů usnadňuje vizuální vnímání dat, jejich vizuální analýzu a celkové posouzení. Pro další zpracování a interpretaci se počítají číselné charakteristiky – popisné statistiky. Mezi nejznámější se řadí aritmetický průměr, modus, medián, směrodatná odchylka, maximum, minimum a rozpětí. (Hendl, 2004)

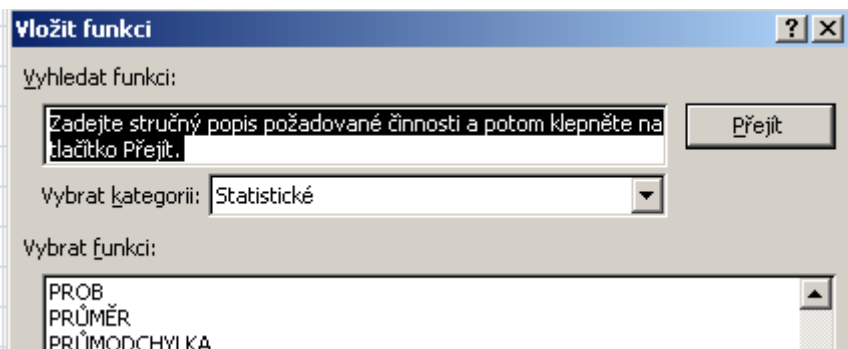
Všechny tyto veličiny popisné statistiky jsou implementovány do Excelu a dají se snadno vypočítat. Pro ukázkou postupu výpočtu použijeme příklad uvedený na začátku – výška žáků.

Vytvoříme tabulku, kde jsou zaznamenána jednotlivá data. Pod sloupec s hodnotami napíšeme do řádků pod sebou jednotlivé veličiny popisné statistiky, které budeme určovat u daného souboru (Obr. 25).

31		178
32	Průměr	
33	Směrodatná odchylka	
34	Modus	
35	Medián	
36	Maximum	
37	Minimum	
38		

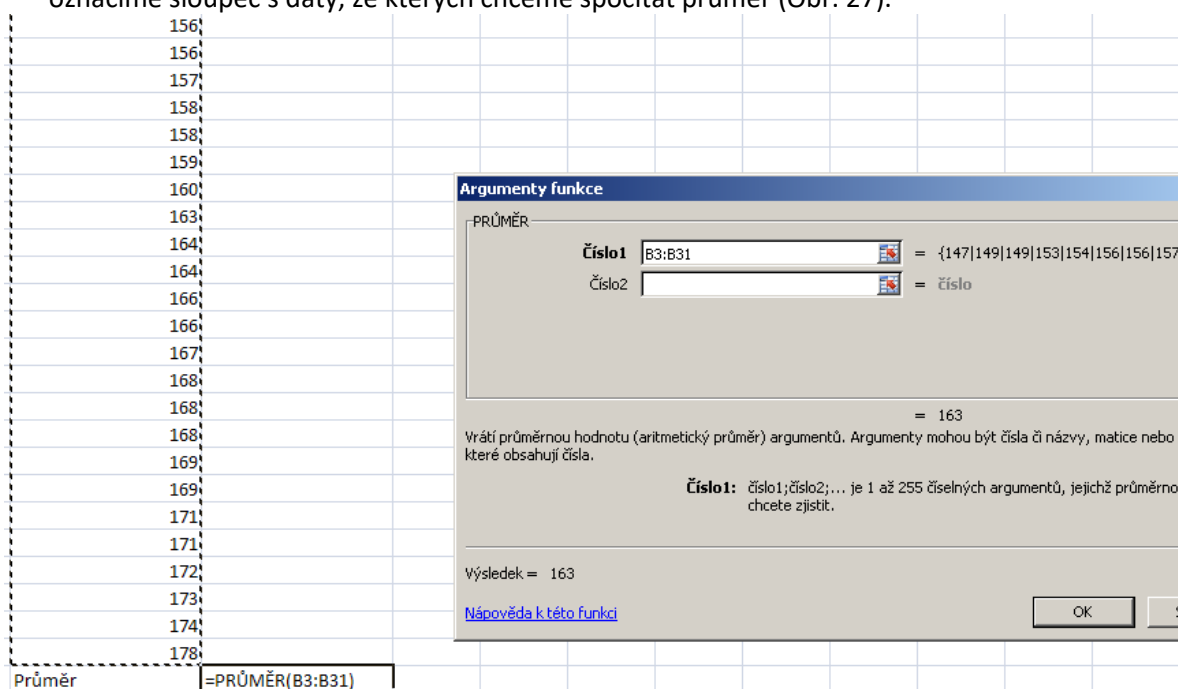
Obr. 25 Popisné statistiky

Začneme **aritmetickým průměrem**. Klikneme v tabulce na buňku vedle buňky **Průměr**, pak na řádek nad tabulkou na tlačítko  a objeví se dialogové okno s nabídkou vložení funkce, kde na řádku **Vybrat kategorii** zvolíme **Statistické** (Obr. 26).



Obr. 26 Výběr funkce průměr

Dále v tomto dialogovém okně v nabídce **Vybrat funkci** zvolíme **PRŮMĚR**. Po potvrzení se objeví dialogové okno **Argumenty funkce**, klikneme do řádku **Číslo 1** a pomocí kurzoru označíme sloupec s daty, ze kterých chceme spočítat průměr (Obr. 27).





Obr. 27 Zadání Argumentů funkce průměr

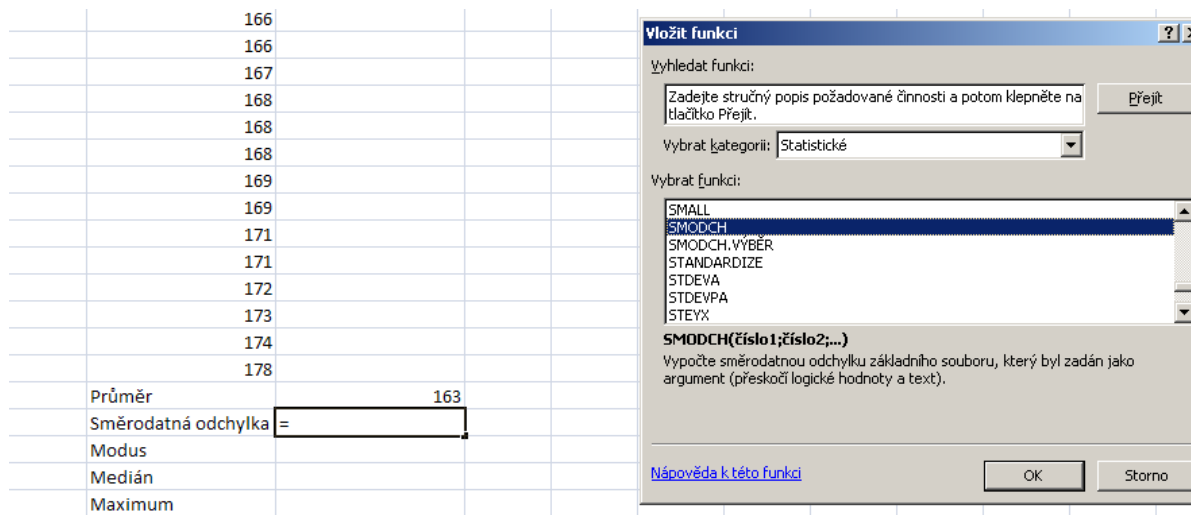
Po potvrzení dialogového okna se ve zvolené buňce tabulky objeví hodnota aritmetického průměru (Obr. 28).



Obr. 28 Výsledná hodnota průměru

Obdobným způsobem vypočítáme i ostatní veličiny.

**Směrodatnou odchylku** vypočítáme, když klikneme do buňky vedle buňky s názvem směrodatná odchylka, poté klikneme na tlačítko  a vybereme kategorii **Statistické** a na řádku **Vybrat funkci** zvolíme **SMODCH** (Obr. 29). Objeví se okno **Argumenty funkce**, klikneme do řádku **Číslo 1**, označíme sloupec s daty a klikneme na tlačítko OK. V označené buňce se objeví výsledek – hodnota směrodatné odchylky. Počet desetinných míst ve výsledku upravíme tlačítky .



Obr. 29 Výběr funkce směrodatná odchylka

**Modus** vypočítáme analogickým postupem jako průměr a směrodatnou odchylku, v řádku **Vybrat funkci** zvolíme **MODE**.

Pro **medián** vybíráme funkci **MEDIAN**, pro **maximum** funkci **MAX**, pro **minimum** funkci **MIN**. **Rozpětí** je pak rozdíl maxima a minima.

Výsledné hodnoty veličin popisné statistiky našeho příkladu jsou uvedeny na obrázku 30.

Průměr	163
Směrodatná odchylka	8,0
Modus	168
Medián	164
Maximum	178
Minimum	147
Rozpětí	31

Obr. 30 Výsledné popisné statistiky

### Test nezávislosti chí-kvadrát pro kontingenční tabulku

Pokud máme rozhodnout, zda existuje závislost mezi dvěma kvalitativními jevy, které byly zaznamenány na úrovni nominálního (případně ordinálního) měření, je vhodné použít test nezávislosti chí-kvadrát. Kontingenční tabulka v následujícím příkladu je čtyřpolní (o dvou sloupcích a dvou řádcích této variantě, kdy sledujeme vztah mezi dvěma alternativními znaky, se říká asociační tabulka), obecně můžeme kontingenční tabulku použít pro větší počet sloupců a řádků.

#### Příklad 3

Studenti učitelství pro 1. stupeň základní školy v rámci projektu zkoumali, zda se liší názory na tentýž problém podle pohlaví. Dotazovali celkem 75 mužů a 75 žen. Výsledky jsou uvedeny v tabulce 1 (Elifson, Runyon, Haber, 1990).

Tab. 1 Odpovědi respondentů

	muži	ženy
souhlas	30	50
nesouhlas	45	25

V tomto případě se jedná o data na nominální úrovni měření. Zajímá nás, zda názory na tentýž problém jsou závislé na pohlaví. Stanovíme nulovou hypotézu.

$H_0$ : Četnosti odpovědí na daný problém nejsou závislé na pohlaví.

Pro testování hypotézy bude následující postup:

1. Zaznamenáme data (pozorované četnosti),
2. Vypočítáme očekávané četnosti,
3. Vypočítáme hodnotu testovací statistiky (chí test),
4. Vypočítáme nebo dle tabulek určíme kritickou hodnotu testovací statistiky,
5. Porovnáme vypočítanou hodnotu testovací statistiky s kritickou hodnotou.

ad 1) Tabulku s **pozorovanými četnostmi** jsme zaznamenali do Excelu (Obr. 31) a doplnili je o součty v řádcích a sloupcích.

2	Pozorované četnosti			
3		muž	žena	
4	souhlas	30	50	80
5	nesouhlas	45	25	70
6		75	75	150
7				

Obr. 31 Tabulka Pozorovaných četností

ad 2) Do druhé tabulky vypočítáme **očekávané četnosti**. Očekávanou četnost pro příslušné pole kontingenční tabulky vypočítáme tak, že násobíme marginální (součtovou) četnost v příslušném sloupci s marginální (součtovou) četností v příslušném řádku a součin dělíme celkovou četností. Příklad (Obr. 32) ukazuje výpočet očekávané četnosti pro případ souhlasu u mužů.


Očekávané četnosti		
	=75*80/150	
	35	35

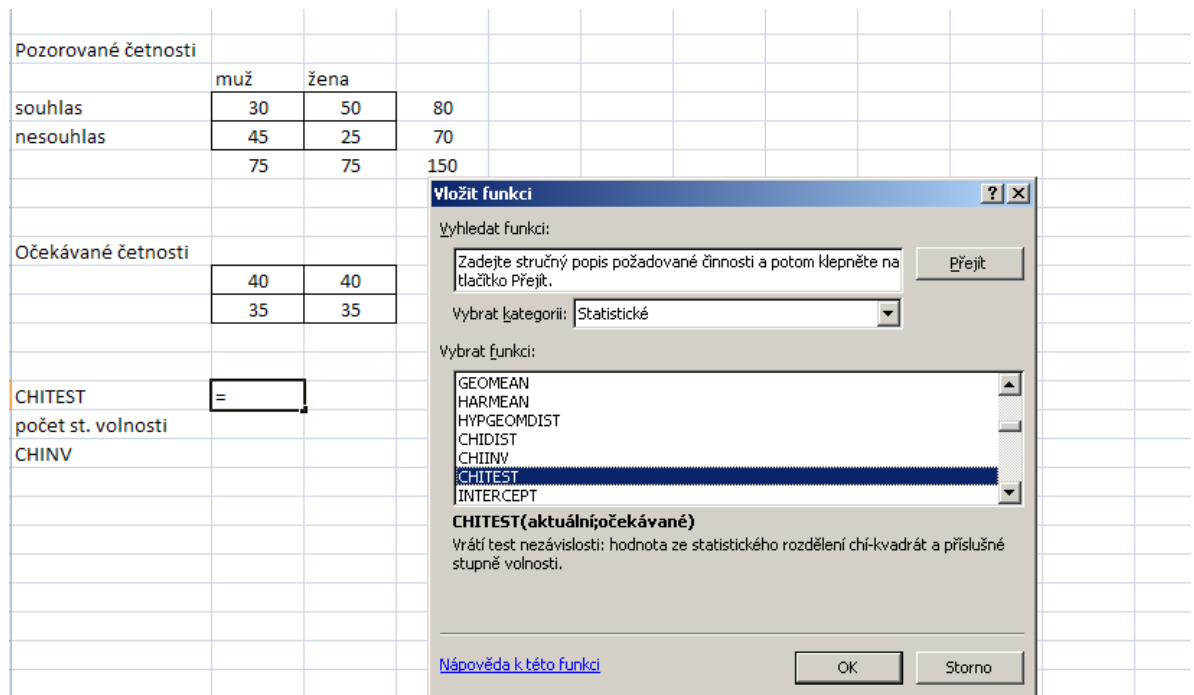
Obr. 32 Výpočet očekávané četnosti

Po provedení všech výpočtů získáme tabulku očekávaných četností (Obr. 33).

Očekávané četnosti		
	40	40
	35	35

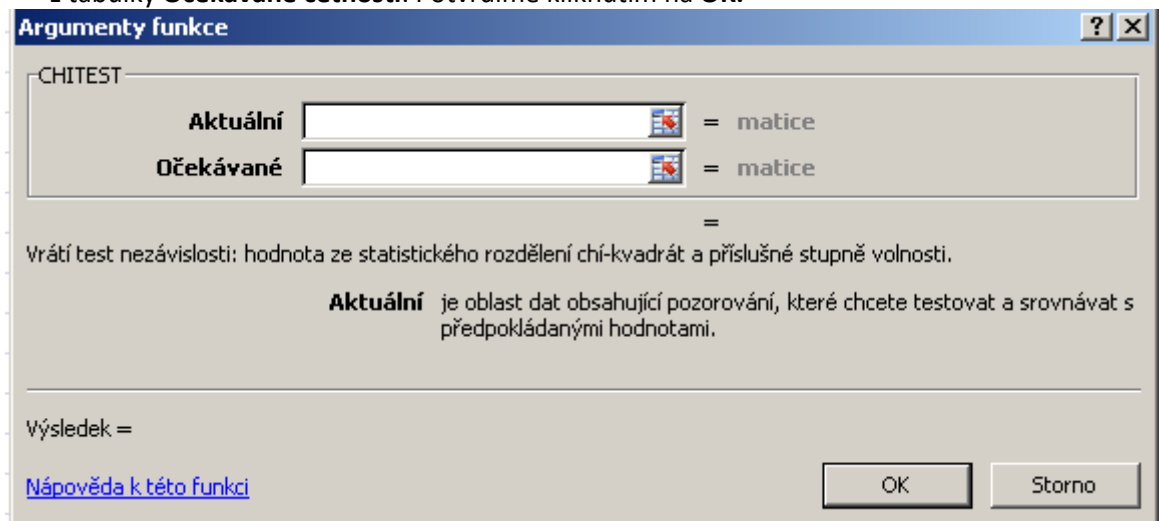
Obr. 33 Očekávané četnosti

ad 3) Pro výpočet konkrétní hodnoty testovací statistiky  $\chi^2$  klikneme do buňky (nejlépe pod tabulkou), kde chceme zobrazit výsledek vedlejší buňku označíme a následně na tlačítko , vybereme kategorii **Statistické** a funkci **CHITEST** (Obr. 34), výběr potvrdíme **OK**.



Obr. 34 Volba funkce CHITEST

Ukáže se dialogové okno **Argumenty funkce** (Obr. 34), kde do řádku **Aktuální** označíme kurzorem hodnoty v tabulce **Pozorované četnosti** a do řádku **Očekávané** označíme hodnoty z tabulky **Očekávané četnosti**. Potvrdíme kliknutím na **OK**.




Obr. 35 Dialogové okno pro zadání argumentů funkce

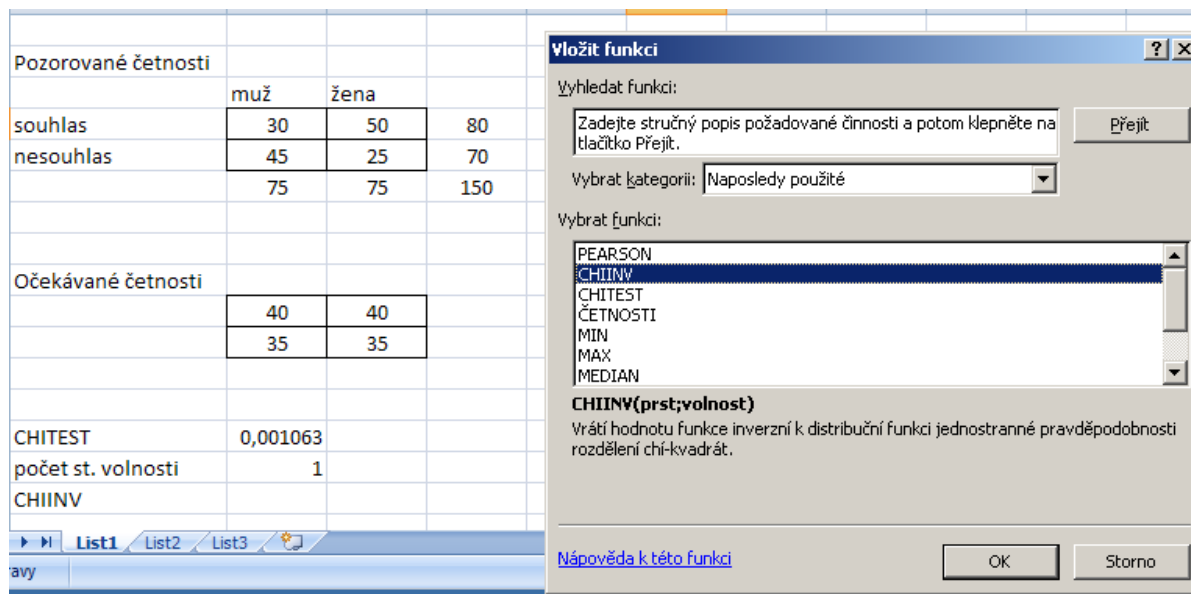
V příslušné buňce se objeví výsledek (jedná se o pravděpodobnost). Dále určíme počet stupňů volnosti podle vztahu:

$$f = (r - 1) * (s - 1)$$

kde  $r$  je počet řádků a  $s$  počet sloupců. Počet stupňů volnosti zadáme do tabulky (je vhodné zadat pod CHITEST).

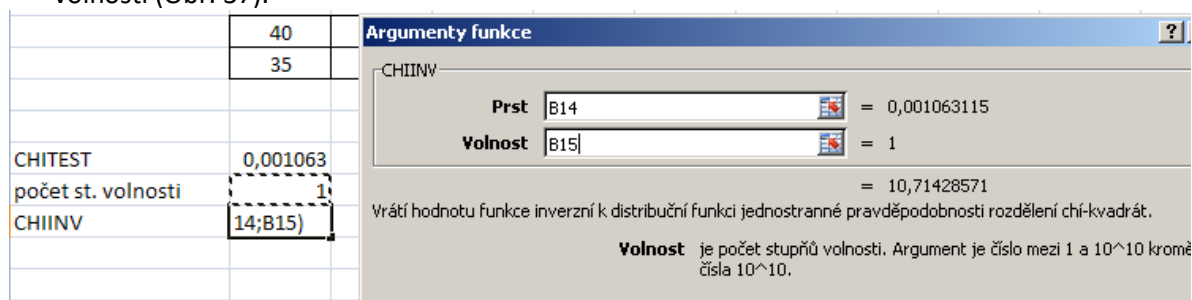
Posledním krokem vypočítáme hodnotu testového kritéria. Vybereme buňku, kam chceme umístit výsledek (vedlejší buňku označíme CHIINV). Kliknutím na , zvolíme kategorii Statistické a vybereme funkci CHIINV (Obr. 36).





Obr. 36 Volba funkce CHIINV

Výběr potvrdíme **OK**. V následujícím dialogovém okně **Argumenty funkce** klikneme do řádku **Prst** a kurzorem označíme buňku, kde je výsledek CHITEST, pak klikneme do řádku **Volnost** a dopíšeme počet stupňů volnosti nebo klikneme na příslušnou buňku s hodnotou počtu stupňů volnosti (Obr. 37).



Obr. 37 Zadání argumentů funkce

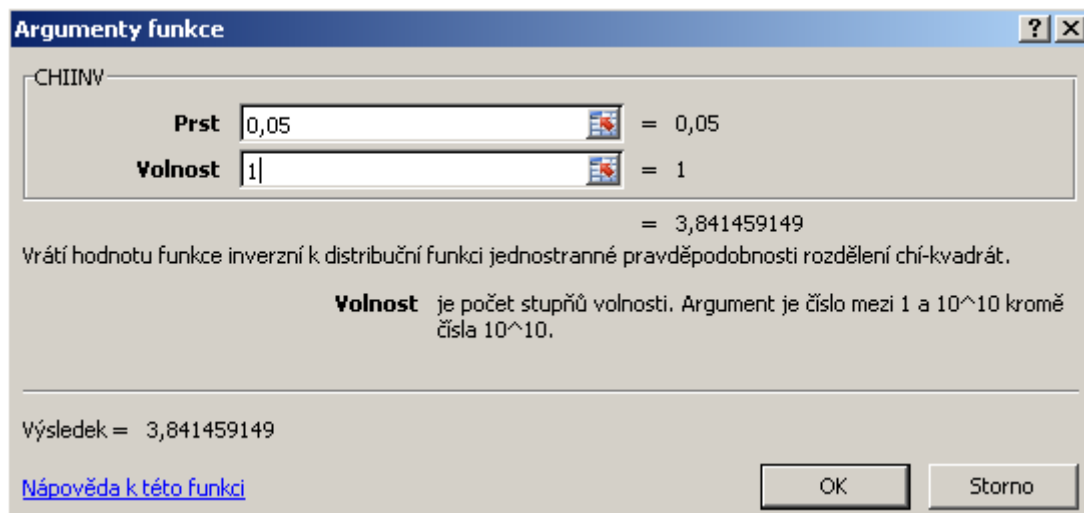
Po potvrzení tlačítkem **OK** dostáváme hodnotu testového kritéria. Tuto hodnotu musíme porovnat s tabulkovou hodnotou pro odpovídající počet stupňů volnosti. Pokud vypočítaná hodnota testového kritéria bude menší než tabulková hodnota, přijímáme nulovou hypotézu. Pokud hodnota testového kritéria bude větší než tabulková hodnota, přijímáme hypotézu alternativní.

P

ro náš příklad:

Hodnota testového kritéria (vypočítaná v Excelu)  $\chi^2 = 10,71$

ad 4) Tabulková hodnota (tzv. kritická hodnota) na hladině významnosti 0,05 pro 1 stupeň volnosti  $\chi^2_{(0,95)} = 3,84$ . Tabulkovou hodnotu určíme pomocí programu MS Excel, pokud hodnotu hladiny významnosti  $\alpha$  dosadíme do řádku **Prst** a do řádku **Volnost** dosadíme vypočítaný počet stupňů volnosti (Obr. 38).



Obr. 38 Výpočet tabulkové hodnoty  $\chi^2$

ad 5) Platí tedy, že  $10,71 > 3,84$ , zamítáme nulovou hypotézu a přijímáme alternativní hypotézu, že četnosti odpovědí na daný problém (souhlas či nesouhlas) jsou závislé na pohlaví respondentů.

### Korelační koeficient

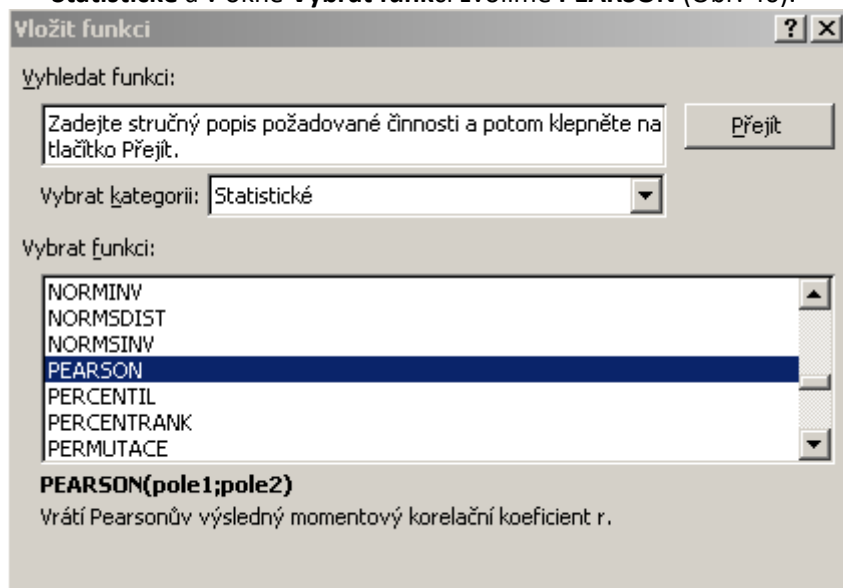
Příklad 4

Žáci 5. třídy v tělesné výchově běhali 100 m, současně byla změřena jejich výška. Výsledky měření jsou uvedeny na obrázku 38. Učitelka chtěla zjistit, zda existuje vztah mezi výškou žáků a jejich výkonem v běhu na 100 m.

výška [cm]	čas [s]
147	15,1
149	15,3
149	16,1
153	15,1
154	14,9
156	15
156	14,6
157	14,9
158	14,9
158	15
159	14,8
160	14,5
163	14,1
164	14,3
164	13,9
166	14,6
166	14,7
167	14,5
168	13,2
168	14,7

Obr. 39 Zjištěné údaje výšky žáků a jejich čas v běhu na 100 m

Vypočítáme hodnotu korelačního koeficientu. Konkrétně pomocí Excelu můžeme vypočítat Pearsonův korelační koeficient. Otevřeme dialogové okno **Vložit funkci**, vybereme kategorii **Statistické** a v okně **Vybrat funkci** zvolíme **PEARSON** (Obr. 40).



Obr. 40 Výběr funkce PEARSON (korelačního koeficientu)

Po potvrzení výběru se objeví dialogové okno Argumenty funkce, kde do řádku Pole1 zadáme hodnoty prvního sloupce (výška), do řádku Pole2 pak hodnoty ve druhém sloupci (čas). Na řádku Výsledek je vypočítána hodnota korelačního koeficientu (- 0,744 po zaokrouhlení) – viz obrázek 41.

výška [cm]	čas [s]
147	15,1
149	15,3
149	16,1
153	15,1
154	14,9
156	15
156	14,6
157	14,9
158	14,9
158	15
159	14,8
160	14,5
163	14,1
164	14,3
164	13,9
166	14,6
166	14,7
167	14,5
168	13,2
168	14,7

Obr. 41 Výpočet korelačního koeficientu

Korelační koeficient může nabývat pouze hodnot v intervalu  $<-1; 1>$  (viz předcházející kapitola). Čím větší je hodnota korelačního koeficientu v absolutní hodnotě, tím těsnější je

vztah mezi zkoumanými jevy.

Ve výše uvedeném příkladě byla vypočítána hodnota korelačního koeficientu – 0,744. Můžeme tedy konstatovat, že mezi výškou žáků a jejich časem v běhu na 100 m je negativní (opačný, inverzní) vztah a to poměrně vysoký. Čím je žák vyšší, tím je rychlejší v běhu.

## NCSS 2004 (Number Crunching Statistical Systems)

*Trial verze a web společnosti: <http://www.ncss.com>*

### **Práce s datovým polem**

Program začíná práci otevřením prázdného listu datového pole. data importovat .

Je možné pokračovat některou z vybraných možností:

- Začít vytvářet svoje vlastní nová data zápisem do otevřeného prázdného listu.
- Otevřít (přes **File - Open**) již existující soubor vytvořený programem NCSS (název souboru \*.s0) , nebo data načíst ze seznamu naposledy otevřených dat. Data - Open
- Importovat (**File - Import**) data (možnost cca 32 formátů databázových, spreadsheetových nebo statistických) **například** z formátů

\*.txt, volném nebo pevném formátu ASCII, \*.dbf formátu DBase, \*.xls formátu Excelu, \*.wk\* formátu z Lotusu, \*.wkq formátu Quattro, \*.dat formátu SOLO (starší verze NCSS pro DOS), \*.sav formátu SPSS for Windows, \*.sys formátu SYSTAT (statistický software, používaný často pro ekonomické aplikace), \*.STA formátu STATISTICA, \*.DTA (S Plus), a dále: Minitab, Matlab, GAUSS, atd.

### **Vytvoření nového datového souboru**

Otevřený list má tvar dvourozměrné tabulky (flat file). Ve sloupcích jsou zaznamenány hodnoty jednotlivých měřených proměnných (měřené veličiny nebo jejich transformace). V každém řádku jsou pozorování pro jeden objekt (pro jeden prvek souboru).

Na ukázce jsou v souboru zaznamenány výsledky bodového hodnocení písemného testu u 10 studentů. Sledují se 4 proměnné: jméno, skóre v testu, doba, která byla potřebná na vyplnění (v sekundách) a kód pohlaví (1=muž, 2=žena).

Každá proměnná používá pro identifikaci svůj název nebo číslo. Standardně (default) nastavené označení C1, C2, ... lze změnit na listu **Variable Info**. List **Sheet1** je datový list.

**Název proměnné** musí začínat písmenem. Může obsahovat pouze písmena, číslice nebo znak podtržení, žádné jiné znaky ani mezeru. Názvem se ve všech výpočetních procedurách odvoláváme na proměnnou, pro kterou se mají výpočty provést. Je proto praktické, když neobsahuje velký počet znaků.

Pro kódovaná data (v našem příkladě pohlaví studenta) je užitečné zaznamenat také informaci s vysvětlením použitých kódů. Kódy a jim odpovídající slovní popisy se zaznamenají do listu dat jako dvě proměnné. Protože toto označení narušuje do jisté míry integritu vlastní databáze, je vhodnější vynechat mezi daty a popisy kódů prázdné sloupce pro případné nově dodané proměnné do databáze (doplnění nově změřených proměnných nebo transformace již zaznamenaných hodnot).

Do listu **Variable Info** se musí ve sloupci Value Label u příslušné proměnné zaznamenat název proměnné, která obsahuje vysvětlení kódů.

Stejná proměnná kódů může být přiřazena několikrát, různým datovým proměnným (například kódy: 1=nízká, 2=střední, 3=vysoká mohou být použity u několika datových proměnných, vysvětlení kódů postačí jediné. Sloupec Value Label tedy může stejnou proměnnou obsahovat vícekrát.

Sloupec Label (listu **Variable Info**) je určen pro podrobnější popis významu dané proměnné.

Užití v jednotlivých procedurách: List Format – Value Label – vybrat Both (chceme-li kódy i

popis hodnot)

### ***Uložení datového souboru***

Po zápisu do listu dat jsou hodnoty uloženy v dočasném souboru přístupném z NCSS, ale nikoliv na disku. Pokud by nastala nějaká nehoda (vypnutí proudu, porucha v síti apod.), budou data ztracena. Proto je užitečné uložit do souboru nejen nově napsaná data, ale během jejich přípravy také pamatovat na jejich průběžné ukládání.

Databáze vytvořená v NCSS má vždy koncovku **.SO**. Soubor, který je otevřen v ukázce, je uložen na lokálním disku v adresáři OBRÁZKY a má název POKUSNÁ.SO.

**Kromě souboru \*.SO vytvoří program NCSS vždy další soubor, který má koncovku \*.S1** (v našem případě POKUSNÁ.S1). V něm jsou uloženy názvy proměnných a další pomocné informace. Oba soubory musí být vždy přítomny v adresáři, ve kterém se pracuje. Při kopírování je potřebné na toto pamatovat.

Pro tisk vybraných proměnných uložených v databázi je vhodný příkaz *Data Report* v činnosti Vytisknutá sestava bude zobrazena v novém výstupním, **Output** okně.

### ***Okna programu NCSS***

Program NCSS je možné ovládat pomocí tří oken:

**NCSS Data** - zobrazuje aktuálně otevřený datový soubor, který je možné analyzovat. V tomto okně lze prohlížet, upravovat a ukládat datové soubory. Je to hlavní okno programu. **Jeho uzavřením se program ukončí.**

**NCSS Procedure** jsou okna, ve kterých se zadávají nastavení pro provádění analýzy dat nebo pro grafické zobrazování. Příkladem okna typu NCSS Procedure je také výše zmíněné okno **Data Report** používané pro výpis datového pole. Uzavření okna nezpůsobí ukončení programu NCSS.

**NCSS Output** je okno pro zobrazení výsledků výpočetních a grafických procedur. Má charakter mini-word procesoru, takže je možné v něm editovat, prohlížet, ukládat nebo tisknout výstupní sestavu. Výstupní soubor lze ukládat ve formátu **.RTF**

Pro přepínání mezi jednotlivými okny je nejvýhodnější používat toolbar systému Windows.

### ***Procedura pro výpočet základních statistických charakteristik***

Statistické procedury se spouštějí z činnosti **Analysis, Descriptive Statistics** (Popisné statistiky), příkazem *Descriptive Stats*.

### ***Krabicové grafy - Box Plots.***

Charakterizují polohu znaku a variabilitu hodnot v souboru. Dávají informaci o odlehlých pozorováních. Kromě krabicových grafů dává podobnou informaci také stromkový graf (stem-leaf plot), histogram, křivka hustoty a pravděpodobnostní graf.

Horní a dolní strana obdélníku: 1. a 3. kvartil, prostřední 2. kvartil. Takže délka obdélníku vyjadřuje kvartilové rozpětí (IQR), které se používá jako míra variability. Místo obdélníku se používá někdy kosočtverec nebo elipsa, nebo speciální útvar, který usnadňuje rozhodnutí při porovnávání různých souborů.

Úsečky vyjadřují rozmezí přiléhajících hodnot (Adjacent Values), mají délku (1,5 IQR) a jsou kresleny od 1. a 3. kvartilu. Dolní přilehlá hodnota vyjadřuje pozici nejmenšího pozorování z těch, která jsou větší nebo rovna P25 - (1,5 IQR). Analogicky horní.

Odlehlé hodnoty (outside values) jsou takové, které jsou mimo rozmezí přilehlých hodnot. Odlehlé hodnoty (mild outliers) jsou hodnoty mimo rozmezí přilehlých hodnot, ale ve vzdálenosti do 3 IQR. Výrazně odlehlé hodnoty (severe outliers) jsou mimo rozmezí (3 IQR) od 3. a 1. kvartilu.

Krabicový graf má výhodné použití pro vícenásobná porovnání (multiple comparisons). Pokud je zejména porovnáván střed rozdělení, lze vycházet z grafu se zářezy (notched box plot),

kteře jsou umístěny podle vztahu: medián 1,57 (IQR)/n

Jestliže soubor je rozdělen na podskupiny podle třídícího znaku, potom skupiny, pro které se zářezy kolem mediánu (jsou to 95 % intervaly spolehlivosti) navzájem nepřekřývají, lze považovat za výběry z populací se shodnými mediány (nejsou statisticky významně odlišné). Moderní verze grafu ještě modifikuje graf tak, že jsou vynechány spojnice mezi konci intervalů spolehlivosti a hranicemi 1. a 3. kvartilu, takže pro porovnání mediánů a jejich intervalů spolehlivosti jsou výhodnější vizuální podmínky.

Struktura dat. Graf se kreslí pro jednu kvantitativní proměnnou. Další proměnnou lze použít pro rozdělení souboru na podskupiny, potom se zobrazí krabicové grafy v rozdělených souborech.

### **Violine plot**

Krabicový graf je doplněn o histogram (jehož zobrazení nahrazuje krabici v grafu), takže z grafu je patrný i tvar rozdělení.

**Pro správnou funkci procedury Box plot** je nutné nastavit desetinnou tečku v prostředí Windows přes Ovládací panely – místní a jazykové nastavení.

### **Popisné statistiky - Descriptive statistics**

Výpočet statistických charakteristik a některé typy grafických zobrazení. Pro každou zadanou proměnnou se počítají (jednorozměrné) charakteristiky polohy, variability a rozdělení (šikmosti, špičatosti, testy normality).

Volba výpočtu:

#### **Analysis Descriptive Statistics Descriptive Stats**

**Příklad:** Data **World95.s0**, popisné charakteristiky pro střední délku života mužů (LIFEXPM).

Ana-lýza je provedena odděleně pro skupiny států, vymezené jejich lokalizací státu nebo economic-kým zařazením. Kódy jsou na proměnné REGION:

1 OECD

2 East Europe

3 Pacific\_Asia

4 Middle East

5 Latn.America

### **Ukázka výstupu pro země OECD (Data World95):**

*Popisné statistiky:*

rozsah souboru střední hodnota směrodatná odchylka směrodatná odchylka průměru (standardní chyba) rozpětí

95 % LCL - 95% mez intervalu spolehlivosti průměru (dolní) = Lower Confidence Limit

95 % UCL - horní Upper

Pro meze spolehlivosti se předpokládá, že populační směrodatná odchylka není známá a že data mají přibližně normální rozdělení.

95% meze spolehlivosti pro medián

Modus není definován (stejně četnosti)

Jednovýběrový test hypotézy o průměru (hypotéza - nulový průměr): testuje se hypotéza, že vzorek byl vybrán z populace s nulovým průměrem. Test je možné použít pouze při výběru z normálního rozdělení.

95 % meze spolehlivosti pro rozptyl (variance) a směrodatnou odchylku (založeny na 2 rozdělení)

Nevychýlený odhad směrodatné odchylky z výběru, založený na použití Gama rozdělení

Kvartilové rozpětí je rozdíl mezi hodnotami 3. a 1. kvartilu

Tak zvané useknuté charakteristiky (s % useknutí). Ze souboru se oddělí příslušné procento mini-málních a maximálních pozorování a pro zbylý soubor se počítají tzv. useknutý průměr a

useknu-tá směrodatná odchylka. Není vhodné při malých počtech pozorování. Porovnání průměru s například 5% useknutým průměrem je důležité pro posouzení vlivu odleh-lých pozorování.

Pod označením count je počet zbylých pozorování (pro výpočet vzata celá část).

Šikmost - nulová hodnota indikuje symetrické rozdělení. Pokud je nula obsažena v rozmezí přibliž-ně: Šikmost  $2\text{StdError}$  (tj.  $-0,2420,418$ ), nelze hypotézu o symetrii zamítnout (platí asymptoticky).

Špičatost vzhledem k normálnímu rozdělení (které má špičatost 3)

Variační koeficient -použití pro porovnání variability v souborech s různými průměry

Alternativní míry šikmosti a špičatosti (Fisherovy míry)

Koeficient disperze -  $\text{COD} = (x_i - \text{medián})/n/\text{medián}$  také relativní míra variability (použití pro analýzy daňových výměrů a oceňování nemovitostí).

Kvartily a jejich 95 % meze (dolní a horní) spolehlivosti. Výpočty intervalů spolehlivosti nepředpo-kládají, že data byla vybrána z normálního rozdělení. Pouze se předpokládá náhodný výběr ze spojitého rozdělení.

*Testy normality.*

Údajně nejlepší vlastnosti má test *D Agostinův omnibusový a Martinezův - Iglewiczův*.

Testy jsou málo citlivé na odchylky od normálního rozdělení, pokud soubory mají méně než 100 pozorování. Pro menší rozsahy lze pouze učinit závěr, že hypotézu o normalitě nelze zamítnout (včetně rizika, že ji nelze zamítnout z důvodu malého počtu pozorování a výsledek je málo průkaz-ný).

Test Martineze a Iglewitze je založen na mediánu a robustním odhadu jeho rozptylu.

DAGostinův omnibusový test spočívá v testování normality na základě dvou charakteristik, šik-mosti a špičatosti.

Počet třídících intervalů v histogramu (lze zvolit) může ovlivnit tvar rozdělení.

Stopa hustoty - vznikne vyrovnaním histogramu, počet třídících intervalů potom neovlivňuje vzhled rozdělení.

Normal probability plot (viz také Probability plots nebo Q-Q plots. Pokud jsou data vybrána z nor-málního rozdělení, leží všechny body na přímce. Na vvislou osu grafu se zobrazují naměřené hod-noty v pořadí podle velikosti, na vodorovné ose je inverzní hodnota, která odpovídá dané proporci pozorování ze standardního normálního rozdělení.

Pásky spolehlivosti slouží jako referenční meze pro posouzení odchylky od normálního rozdělení. Když jsou některé body umístěny mimo pásky spolehlivosti, data nemají normální rozdělení. Pokud je pouze výjimečně některý bod mimo pásky spolehlivosti, může se jednat o odlehlé pozorování. Graf není vhodný při vzorcích do 30 pozorování.

Jednotlivé percentily a jejich 95% meze spolehlivosti (při jejich výpočtu se nepředpokládá normali-ta rozdělení znaku v populaci, ze které byl vzorek vybrán). Program počítá interval spolehlivosti, s chybou která je menší než specifikovaná hladina. Přesná hladina významnosti je v posledním sloupci.

**Stromkový graf** umožňuje vidět detailně hodnoty v datech a jejich rozdělení. Lze jej využít i pro hledání chyb v datech a pro úvahy o tvaru rozdělení veličiny v souboru.

**Transformace dat**

**Transformace proměnných**

Pomocí vzorce transformací lze vytvářet nové hodnoty (kombinace proměnných, procentní body, odvozené proměnné apod.). Transformaci lze zapsat do transformačního okna, které lze vyvolat dvojitým stisknutím tlačítka myši na listu Variable Info ve sloupci Transformation.

Možné funkce:

Operátory sčítání, odčítání, násobení, dělení, exponent, logické: , , =, =,  
Datumové funkce, funkce vyplnění, matematické funkce, pravděpodobnostní funkce,  
Náhodná čísla  
Přeuspořádání hodnot: sloučení nebo rozdělení proměnných  
Překódování  
Statistické funkce  
Trigonometrické funkce  
Textové funkce  
Indikátorové proměnné

## IBM SPSS Statistics Base

datový soubor a práce s výstupy – viz příručka Statistic Base 16.0

**Způsob výuky:** téma je přednášeno částečně, je potřebné doplnění samostudiem doporučené literatury



### SHRNUTÍ A TÉMATA KE STUDIU

Zpracování dat v MS Excel, SPSS, příp. ukázka možností NCSS. Zápis získaných dat, jejich uspořádání a příp. transformace, škálování, základní vyhodnocení dat v tabulkách prvního stupně třídění. Možnosti tabulek druhého a vyššího stupně třídění, pilotování tabulek. Grafické znázornění, možnosti grafů – rozptyl, .... – z části samostudium.



### DOPORUČENÁ LITERATURA A JINÉ ZDROJE

- HENDL, J. *Kvalitativní výzkum*. Praha: Portál, 2005. 408 s.
- LAŠEK, J., MANĚNOVÁ, M. *Základy statistického zpracování pedagogicko-psychologického výzkumu*. Hradec Králové: Gaudeamus, 2003. 42 s.
- KOŘÍNEK, M. *Excel 2002*. České Budějovice: Kopp 2001. 246 s.
- SKUTIL, M. a kol. *Základy pedagogicko-psychologického výzkumu pro studenty učitelství*. Praha: Portál, 2011. 256 s.



### KONTROLNÍ OTÁZKY A ÚKOLY

1. *Vyzkoušejte si zpracování dat ze svého seminárního úkolu v jednom ze statistických programů*



### SEZNAM POŽADAVKŮ K ABSOLVOVÁNÍ PŘEDMĚTU

Předmět je ukončen zápočtem a zkouškou. Zápočet se sestává z obhajoby fiktivního výzkumného šetření zaměřeného k tématu závěrečné práce studenta. V projektu fiktivního šetření se předpokládá: vymezení problému a cíle; otázek či vstupních hypotéz šetření, včetně jejich zdůvodnění; zdůvodnění výběru metod pro sběr dat a vyhodnocení dat, charakteristika výzkumného vzorku (výstupem je podrobně rozpracovaný projekt výzkumu včetně event. realizovaného předvýzkumu). Dále, prezentace fiktivních dat, verifikace hypotéz na základě diskuse a vybraných statistických metod. Student též prokáže seznámení se základními dokumenty upravujícími tvorbu bakalářských/diplomových prací (Rozhodnutí děkana 17/2009 a v něm uváděné normy, Rektorský výnos č. 4/2009. Zkouška se skládá z písemného testu (3 příklady s cílem ověření znalostí z oblasti



základních statistických charakteristik, statistických metod při testování platnosti hypotéz a možností určování těsnosti vztahů mezi proměnnými za pomoci vybraných/dostupných statistických programů. Zkouška je následně zakončena ústní rozpravou nad písemným testem s doplňujícími otázkami oblasti teorie (příklady viz ukázka možných otázek v závěru průvodce). Termíny zápočtů a zkoušek budou stanoveny po dohodě s vyučujícím na konci semestru a vypsány ve Fakultním informačním systému.

## SEZNAM DÍLČÍCH OTÁZEK K ÚSTNÍ ZKOUŠCE

1. *Co je vědecký výzkum?*
2. *Jaké cíle si vědecký výzkum klade?*
3. *Čím se zabývá výzkum v pedagogických, sociálních vědách?*
4. *Čím se liší pedagogický, sociologický výzkum od jiných typů vědeckého výzkumu? (Uveďte příklady)*
5. *Na jakých předpokladech je založen empirický výzkum?*
6. *Co znamená redukce informací a proč se jí zabýváme?*
7. *Vysvětlete pojem reliability.*
8. *Vysvětlete pojem validity.*
9. *Jak byste vysvětlili pojem „teorie“?*
10. *Jak teorie vznikají?*
11. *Co je empirie?*
12. *Co je kvantitativní výzkum? Popište obecně.*
13. *Co je kvalitativní výzkum? Popište obecně.*
14. *Jaké jsou hlavní rozdíly mezi kvantitativní a kvalitativní metodologií?*
15. *Vysvětlete pojmy dedukce a indukce.*
16. *Co je teoretické nasycení výzkumného problému?*
17. *Co je triangulace ve výzkumu?*
18. *Kdy se používá kvantitativní výzkum?*
19. *Kdy se používá kvalitativní výzkum?*
20. *Jakým způsobem se rozhodujeme o zkoumaném vzorku a jeho velikosti v kvantitativním výzkumu?*
21. *Jakým způsobem se rozhodujeme o zkoumaném vzorku a jeho velikosti v kvalitativním výzkumu?*
22. *Co je reprezentativnost vzorku?*
23. *Jaké základní typy výběru při kvantitativním výzkumu znáte?*
24. *Čím může být výběr vzorku v kvantitativním výzkumu zkreslen?*
25. *Vyjmenujte typy náhodného čili pravděpodobnostního výběru.*
26. *Popište prostý náhodný výběr.*
27. *Popište stratifikovaný náhodný výběr.*
28. *Vysvětlete typy nenáhodného čili nepravděpodobnostního výběru*
29. *Popište kvótní výběr.*
30. *Popište výběr úsudkem a vysvětlete, proč a kdy se používá.*
31. *Kterou techniku sběru dat považujete pro svou další práci za nejužitečnější a proč?*
32. *Jak vzniká výzkumný nápad při kvalitativním výzkumu?*
33. *Která témata jsou vhodná pro techniku experimentu? Uveďte příklad(y).*
34. *Co znamená, že je proměnná dichotomická? Uveďte na takovou proměnnou příklady.*
35. *K čemu se používá pretest a posttest v experimentu a proč?*
36. *Uveďte postup při jednoduchém klasickém experimentu.*
37. *Uveďte, kde se používají kontrolní skupiny v různých typech experimentu; proč se tam použijí?*
38. *Čemu říkáme účinek badatele (např. při interview)?*
39. *Jaké výhody má náhodný (pravděpodobnostní) výběr do skupin v experimentu?*

40. *Jaké nevýhody má náhodný (pravděpodobnostní) výběr do skupin v experimentu?*
41. *Jak se provádí randomizace výběru do skupin v experimentu? Čím se liší od náhodného výběru?*
42. *Co znamená párování do skupin v experimentu a jak se provádí? Jak se liší v důsledcích od náhodného a od randomizovaného výběru do skupin.*
43. *Jak se provádí výběr do skupin experimentu regulací frekvenčních distribucí? Jaké to má důsledky proti náhodnému, párovanému a randomizovanému výběru do skupin?*
44. *Vysvětlete pojem rozptyl hodnot souboru. Víte, co znamená strmost a špičatost křivky rozdělení hodnot souboru?*
45. *Jaké je schéma Solomonova projektu dvou kontrolních skupin?*
46. *Na čem konkrétně a stručně, významně závisí reliabilita experimentu?*
47. *Na čem stručně závisí validita experimentu?*
48. *Vyjmenujte, které vlivy snižují validitu experimentu?*
49. *Vysvětlete pojem „úmrtnost v experimentu.“*
50. *Co to znamená „kontaminace kontrolní skupiny experimentu?“*
51. *Vysvětlete rozdíly v chybném ovlivnění experimentu: kompenzací, kompenzačním soupeřením, demoralizací.*
52. *Výhody a nevýhody experimentu.*
53. *Stručně vysvětlete pojem „respondent.“ Uveďte některá synonyma pojmu respondent.*
54. *Stručně vysvětlete pojem „tazatel“.*
55. *Stručně vysvětlete pojem „výzkumné šetření“.*
56. *Co to znamená „longitudinální studie“?*
57. *Čím se liší technika interview u kvantitativního a kvalitativního výzkumu?*
58. *Jakým způsobem interview připravujeme?*
59. *Co je seznam otázek interview?*
60. *Jaké typy otázek může seznam otázek interview zahrnovat?*
61. *Které chyby v otázkách se při interview nejčastěji objevují?*
62. *Jak postupujete při provádění interview? Uveďte alespoň 5 hlavních zásad postupu.*
63. *Která témata jsou vhodná pro výzkum dotazníkem?*
64. *Co je to anketa, jaké má výhody a jaké nevýhody?*
65. *Čím se liší anketa od dotazníku?*
66. *Která témata jsou vhodná pro výzkum pomocí interview?*
67. *Ve kterých všech dotazovacích technikách používáme seznam otázek?*
68. *Krátce popište tři základní formáty otázek a uveďte na ně příklady.*
69. *Jaké znáte výhody uzavřených otázek?*
70. *Jaké znáte nevýhody uzavřených otázek?*
71. *Jaké znáte výhody otevřených otázek?*
72. *Jaké znáte nevýhody otevřených otázek?*
73. *Jaké znáte výhody polootevřených otázek?*
74. *Jaké znáte nevýhody polootevřených otázek?*
75. *Popište filtrační otázky a čím bývají doprovázeny. Uveďte svůj příklad filtrační otázky.*
76. *Co to jsou kontingenční otázky a co jim nutně předchází?*
77. *Popište výčtové otázky, jejich varianty a jejich výhody. Uveďte svůj příklad.*
78. *Co rozumíme pod pojmem „baterie otázek“? Kdy se používá?*
79. *Jaké znáte typy chybně utvořených otázek? Uveďte příklady.*
80. *Vysvětlete, čím se liší dvojitá a nejednoznačná otázka a uveďte příklady.*
81. *Jaké druhy nesrozumitelných slov se mohou objevit při formulování otázek? Uveďte příklady všech druhů.*
82. *Uveďte své příklady na nesprávně formulovanou otázku intelektuálně formulovanou nad průměrnou úroveň populace.*
83. *Uveďte příklad příliš komplexní otázky.*
84. *Uveďte příklady sugestivních otázek.*
85. *Uveďte příklady citlivých otázek.*

86. *Jaké zásady bychom měli dodržovat při používání citlivých otázek?*
87. *Jaké zásady dodržujeme v pořadí otázek při vypracování seznamu otázek?*
88. *Které všechny parametry rozhodujeme při konstrukci dotazníku?*
89. *Jaký počet otázek a jaký celkový počet kategorií bychom, podle Vás, neměli v dotazníku překračovat?*
90. *Jaké způsoby pro vyznačení odpovědi na uzavřenou otázku znáte?*
91. *Uvedte příklady nominální proměnné.*
92. *Uvedte příklady ordinální proměnné.*
93. *Uvedte příklady na intervalové proměnné.*
94. *Uvedte příklady poměrné proměnné.*
95. *Jaká jsou možná řešení, když je v dotazníku v uzavřené otázce příliš mnoho alternativ odpovědí?*
96. *Uvedte příklad škály s lichým počtem stupňů u odpovědi na otázku s ordinální proměnnou.*
97. *Uvedte příklad škály se sudým počtem stupňů u odpovědi na otázku s ordinální proměnnou.*
98. *Jaká je výhoda a nevýhoda škály s lichým počtem stupňů u odpovědi na otázku s ordinální proměnnou?*
99. *Jaký význam má zařazení kategorie „Nevím“ do možných odpovědí na otázku v dotazníku?*
100. *Co má obsahovat průvodní dopis k dotazníku?*
101. *Jaké další písemnosti je třeba zaslat současně s dotazníkem?*
102. *Které faktory ovlivňují návratnost zasílaných dotazníků?*
103. *Jak zkoumáme reliabilitu jednotlivých otázek dotazníku a reliabilitu celého dotazníku?*
104. *Jak můžeme zkoumat validitu odpovědí v dotazníku?*
105. *Jaké znáte výhody a nevýhody dotazníku?*
106. *Jakými chybami může být kvalita dat ovlivněna?*
107. *Vysvětlete pojmy primární a sekundární skupina.*
108. *Jmenujte tři typy individuálního interview (face to face).*
109. *Popište standardizované interview a určete jeho výhody (event. nevýhody).*
110. *Jaká jsou hlavní pravidla při provádění interview?*
111. *Vyjmenujte alespoň 7 z 10 rad pro vedení interview.*
112. *Které charakteristiky tazatele mohou ovlivnit kvalitu dat?*
113. *Jak se projevuje rozdílný sociální status tazatele a respondenta?*
114. *Jak má působit na respondenta oblečení tazatele?*
115. *Jaká pravidla má tazatel dodržet při zahájení interview s respondentem?*
116. *Jaké jsou možné důvody respondenta k odmítnutí interview a jak jim může tazatel čelit?*
117. *Jaká tři základní pravidla musí dodržovat tazatel při kladení otázek při interview?*
118. *Vysvětlete co je to sonda, kdy se používá; uveďte několik svých příkladů.*
119. *Jaké má možnosti a taktiky tazatel pro upřesnění nedostatečné odpovědi respondenta?*
120. *Jaké jsou výhody a nevýhody interview?*
121. *V čem jsou výhody telefonického. interview pomocí náhodně zvolených čísel?*
122. *Co víte o validitě a reliabilitě telefonického interview?*
123. *Jaké jsou výhody telefonického interview?*
124. *Jaké jsou nevýhody telefonického interview?*
125. *Co víte o získávání dat s využitím počítače?*
126. *Co víte o osobním dotazování pomocí počítače?*
127. *Co víte o dotazování elektronickou poštou?*
128. *Co víte o reliabilitě a validitě dotazovacích technik jako celku?*
129. *Co víte o výhodách a nevýhodách dotazovacích technik jako celku?*
130. *Co to je focus group (skupinové interview): stručně popište hlavní rysy.*
131. *Kdy a jak technika focus group vznikla?*
132. *Kdy používáme focus group (skupinové interview)? Vyjmenujte alespoň 3 důvody pro použití focus group. j*
133. *Co vše tvoří finanční náklady na focus group (skupinové interview)?*
134. *Kdo je to moderátor a jaké má úkoly?*

135. Jak určujeme počet skupin u focus group (skupinového interview) a co z toho vyplývá?
136. Které tři prvky musíme předem rozhodnout před volbou techniky focus group (skupinového interview)?
137. Určete 4 kritéria pro efektivní vedení focus group (skupinového interview).
138. Podle jakých zásad probíhá skupinová diskuse u focus group (skupinového interview)?
139. Jaká tři základní pravidla oznámí moderátor účastníkům před provedením focus group (skupinového interview)?
140. Kdy je nevhodné použít techniku focus group (skupinového interview)?
141. Jaké výhody má technika focus group (skupinové interview) při svém použití?
142. Co to je synergický efekt při použití focus group?
143. Definujte co to je obsahová analýza a jaké je její postavení v oblasti metodologické?
144. Jaké (čtyři) vlastnosti obsahové analýzy znáte?
145. Popište postup obsahové analýzy v několika krocích.
146. Jaké typy jednotek obsahové analýzy znáte?
147. V čem spočívá analýza při obsahové analýze?
148. Popište techniku pozorování.
149. Která výzkumná témata jsou vhodná pro techniku pozorování?
150. Jaké znáte druhy pozorování?
151. Jaké znáte přednosti techniky pozorování?
152. Jaké znáte nevýhody techniky pozorování?
153. Popište techniku studia dokumentů.
154. Jaké jsou přednosti studia dokumentů?
155. Jaké jsou nevýhody studia dokumentů?
156. Jaké druhy (typy) dokumentů znáte pro techniku studia dokumentů?
157. Vysvětlete pojem historiografie.
158. Jaké zdroje dat v historickém výzkumu rozeznáváme?
159. Vysvětlete, co je externí kritika dokumentu.
160. Vysvětlete, co je interní kritika dokumentu.
161. Popište archivální výzkum a jeho přednosti.
162. Jaké typy dokumentů známe a jak s nimi pracujeme?
163. Co je studium materiálních stop lidí?
164. Jaké jsou výhody studia dokumentů?
165. Jaké jsou nevýhody studia dokumentů?
166. Jaké náměty jsou vhodné pro evaluační výzkum?
167. Co je to sociální intervence?
168. Které proměnné mimo experimentální nezávislé a výstupní závislé proměnné musíme při evaluačním výzkumu sledovat?
169. Co víte o projektu evaluačního výzkumu?
170. Definujte co je to sociometrie a čím se liší kvazisociometrie?
171. Co je to sociometrický test a co určuje?
172. Jaké tři charakteristiky je třeba předem určit u sociometrického testu?
173. Jakými typy analýzy zpracováváme sociometrická data?
174. Jaké má sociometrie použití v sociálních vědách?
175. Vysvětlete termíny „sociometrická hvězda“, „outsider“.
176. Vysvětlete termíny „klika“, „izolát“.
177. Od kdy zhruba proniklo do sociologického výzkumu i použití her a simulací?
178. Která témata jsou vhodná pro simulace a hry?
179. Co víte o sociologických počítačových simulacích?
180. Co to je kódování a proč a jak se provádí?
181. Vysvětlete postupy kódování
182. Jak píšeme závěrečnou zprávu z kvalitativního výzkumu a co musí obsahovat?
183. Jmenujte co vše má obsahovat závěrečná zpráva z výzkumu.

