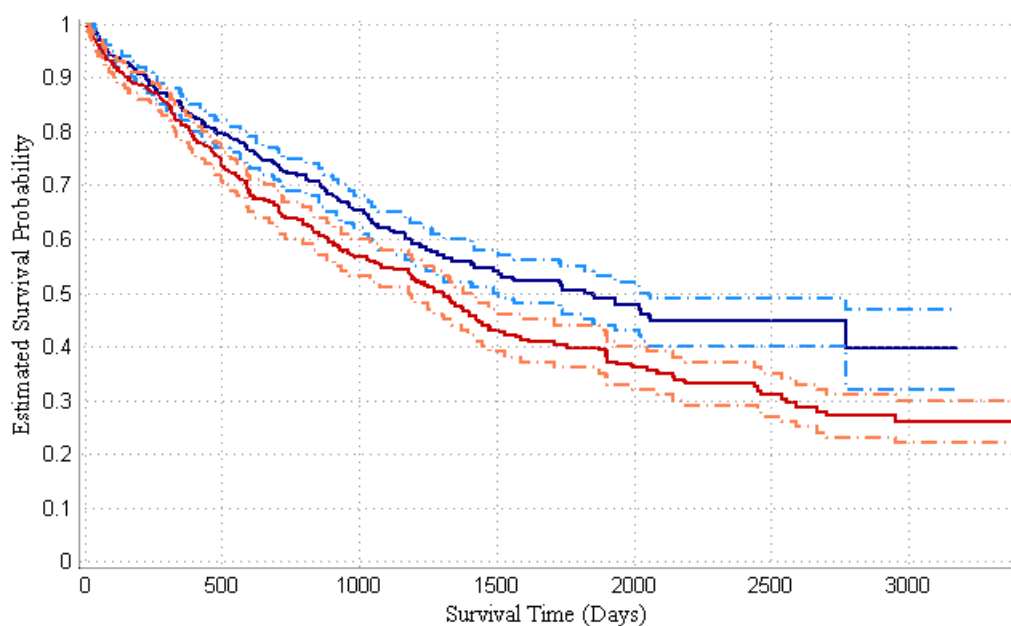


# Biostatistika

## Cvičení - pracovní listy



Martina Litschmannová, Kateřina Janurová  
5.května 2013

Jméno: .....

---

# 1. Kombinatorika

Kombinatorika je vstupní branou do teorie pravděpodobnosti. Zabývá se různými způsoby výběru prvků z daného souboru.

- **Kombinatorické pravidlo o součinu**

Počet všech uspořádaných  $k$ -tic, jejichž první člen lze vybrat  $n_1$  způsoby, druhý člen po výběru prvního členu  $n_2$  způsoby atd. až  $k$ -tý člen po výběru všech předcházejících členů  $n_k$  způsoby, je roven  $n_1 \cdot n_2 \cdot \dots \cdot n_k$ .

- **Kombinatorické pravidlo součtu**

Jsou-li  $A_1, A_2, \dots, A_n$  konečné množiny, které mají po řadě  $p_1, p_2, \dots, p_n$  prvků, a jsou-li každé dvě disjunktní, pak počet prvků množiny  $A_1 \cup A_2 \cup \dots \cup A_n$  je roven  $p_1 + p_2 + \dots + p_n$ .

- **Typy výběrů**

Uspořádané výběry	Bez opakování	Variace bez opakování	$V(n, k) = \frac{n!}{(n-k)!}$
		Permutace bez opakování	$P(n) = V(n, n) = n!$
	S opakováním	Variace s opakováním	$V^*(n, k) = n^k$
		Permutace s opakováním	$P^*(n_1, n_2, \dots, n_k) = \frac{n!}{n_1!n_2!\dots n_k!}$
Neuspořádané výběry	Bez opakování	Kombinace bez opakování	$C(n, k) = \frac{n!}{(n-k)!k!}$
	S opakováním	Kombinace opakováním	$C^*(n, k) = \frac{(n+k-1)!}{(n-1)!k!}$

- **Klasická definice pravděpodobnosti**

$$P(A) = \frac{m}{n}$$

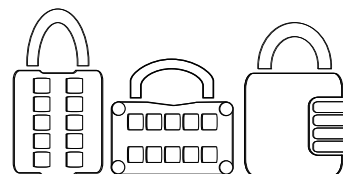
kde:  $m$  ... počet výsledků příznivých jevu  $A$

$n$  ... počet všech možných výsledků

## 1.1 Příklady

*Příklady 1-10 jsou upravené příklady z přednášky prof. Adama Plocki (Kombinatorika a pravděpodobnost kolem nás a ve škole; Seminář matematiky pro středoškolské profesory a učitele základních škol; Ostrava; 2014)*

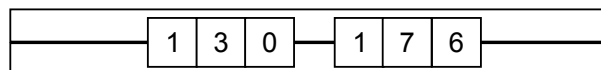
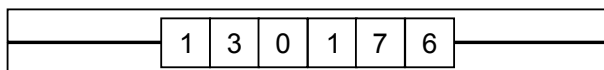
1. V prodejně mají k dispozici tři typy zámků. Pro otevření prvního zámku je nutno zmáčknout čtyři z deseti tlačítek označených číslicemi 0 až 9. (Na pořadí nezáleží - tlačítka zůstávají zmáčkuta.) Druhý zámek se otevře pokud zmáčkneš šest tlačítek z deseti. Pro otevření třetího zámku je nutno nastavit správnou kombinaci na čtyřech kotoučích. Který z těchto zámků nejlépe chrání před zloději?



Řešení:

[třetí]

2. V prodejně nabízejí dva druhy zamykání kufríku. První kufrík se zamyká šifrou, která se skládá z šesti číslic. Druhý kufrík se zamyká dvěma zámky, které se otvírají současně. Šifra každého z nich se skládá ze tří číslic. Určete pro každý kufrík pravděpodobnost otevření zlodějem při prvním pokusu. Který typ zámku je bezpečnější?



Řešení:

[oba jsou stejně bezpečné, pravděpodobnost otevření je  $10^{-6}$ ]

3. V urně je 40 koulí - 2 červené a 38 bílých. Z urny náhodně vytáhneme 2 koule. S jakou pravděpodobností budou obě červené?

Řešení:

[0, 0013]

4. Student si měl ke zkoušce připravit odpovědi na 40 otázek. Na dvě otázky, které mu dal zkoušející, neuměl odpovědět a tak řekl „To mám smůlu! To jsou jediné dvě otázky, na které neumím odpovědět.“ S jakou pravděpodobností mluví pravdu?

Řešení:

[0, 0013]

5. Test z chemie žák složí, pokud v seznamu 40 chemických sloučenin podtrhne jediné dva aldehydy, které v seznamu jsou. Jaká je pravděpodobnost, že test složí žák, který provede výběr sloučenin náhodně?

Řešení:

[0, 0013]

6. Ze zahraničí se vracela skupina 40 turistů a mezi nimi byli 2 pašeráci. Na hranici celník 2 turisty vyzval k osobní prohlídce a ukázalo se, že oba dva jsou pašeráci. Zbylí turisté na to reagovali: „Celník měl opravdu štěstí!“, „Pašeráky někdo udal!“, ... Jak se postavit k těmto výroky? Je oprávněné podezření, že pašeráky někdo udal?

Řešení:

[pravděpodobnost, že pašeráky někdo udal: 0, 9987 ]

7. Z urny se třemi koulemi, dvěma červenými a jednou bílou, budou současně vybrány dvě koule. Student a učitel uzavřou sázku. Pokud budou obě koule stejné barvy, vyhraje student. Pokud budou mít koule různou barvu, vyhraje učitel. Je hra férová? Jaké jsou pravděpodobnosti výhry učitele a studenta?

Řešení:

[hra není férová;  $S : 0, 33$ ;  $U : 0, 67$ ]

8. Hra popsaná v příkladu 7 nebyla férová. Jakou kouli (červenou nebo bílou) musíme do urny přidat, aby hra férová byla?

Řešení:

[červenou]

9. Chcete hrát Člověče nezlob se, ale ztratila se hrací kostka. Čím a jak lze nahradit hrací kostku, máte-li k dispozici hrací karty (balíček 32 karet) a 4 různobarevné kuličky?

Řešení:

10. Chcete hrát Člověče nezlob se, ale ztratila se hrací kostka. Jak lze nahradit hrací kostku, máte-li k dispozici 3 různobarevné kuličky?

Řešení:

11. V prodejně vozů Škoda mají v měsíci únoru prodejní akci. Ke standardnímu vybavení nabízejí 3 položky z nadstandardní výbavy zdarma. Nadstandardní výbava zahrnuje 7 položek:

- tempomat, vyhřívání sedadel, zadní airbagy, xenonová světla, stropní okénko, bezpečnostní zámek převodovky, speciální odolný metalízový lak.

Kolik možností má zákazník, jak zvolit 3 položky z nadstandardní výbavy?

Řešení:

[35]

12. Při zkoušce si do 5. řady sedlo 12 studentů. Zkoušející chce určit sám, jak tyto studenty v řadě rozesadit.

- a) Kolik je možností jak studenty rozesadit?

Řešení:

[479 001 600]

- b) Student Brahy žádá, aby mohl sedět na kraji a odejít dříve, aby stihl vlak. Kolik je možností jak studenty rozesadit, chce-li zkoušející vyhovět požadavku studenta Brahyho?

Řešení:

[79 833 600]

- c) Kolik je možností jak studenty rozesadit, nesmějí-li Pažout a Horáček sedět vedle sebe?

Řešení:

[399 168 000]

13. Kolik anagramů lze vytvořit ze slova STATISTIKA?

Řešení:

[75 600]

14. V Tescu dostali nové zboží – 6 druhů chlapeckých trik. Od každého druhu mají alespoň 7 kusů. Maminka chce synovi koupit 4 trika. Kolik je možností, jak je vybrat

- a) mají-li být všechna různá?

Řešení:

[15]

b) připouští-li, že mohou být všechna stejná?

Řešení:

[126]

15. Kolik hesel délky 5 můžeme vytvořit ze znaků abecedy

a) nejsou-li rozlišována velká a malá písmena?

Řešení:

[11 881 376]

b) jsou-li rozlišována velká a malá písmena?

Řešení:

[380 204 032]



---

## 2. Pravděpodobnost

**Náhodný pokus** je každý konečný děj, jehož výsledek není předem jednoznačně určen podmínkami, za nichž probíhá, a který je, alespoň teoreticky, neomezeně opakovatelný.

Množinu všech možných výsledků  $\{\omega\}$  daného pokusu označujeme pojmem **základní prostor** a značíme jako  $\Omega$

Prvky, popř. jednoprvkové podmnožiny, základního prostoru jsou **elementární jevy**.

**Jevem**  $A$  je libovolná podmnožina základního prostoru. Pro náhodné jevy platí algebraické zákony a rovnosti stejné jako pro množiny.

**Úplná množina vzájemně disjunktních jevů** je množina po dvou disjunktních jevů  $A_1, A_2, A_3, \dots, A_n$ , jejichž sjednocení tvoří množinu  $\Omega$ .

**Jevové pole**  $\mathcal{A}$  je neprázdný systém podmnožin základního prostoru uzavřený vůči doplňku a vůči sjednocení ( $\sigma$ -algebra na  $\Omega$ ).

**Vlastnosti pravděpodobnostního prostoru** jsou dány Kolmogorovovým axiomatickým systémem.

**Podmíněná pravděpodobnost** je pravděpodobnost výskytu jevu za podmínky, že nastal určitý jev, který není nemožný.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Jesliže platí  $P(A \cap B) = P(A) \cdot P(B)$  nebo  $P(B) = 0$ , pak nazýváme **jevy**  $A, B$  **nezávislé**.

**Věta o úplné pravděpodobnosti** nám dává návod, jak určit pravděpodobnost jevu  $A$ , o kterém je známo, že může nastat pouze současně s některým z jevů  $B_1, B_2, \dots, B_n$ , které tvoří úplnou množinu disjunktních jevů.

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i)$$

**Bayesova věta** nám umožňuje spočítat podmíněné pravděpodobnosti jednotlivých jevů této úplné množiny za předpokladu, že nastal jev  $A$ .

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^n P(A|B_i) \cdot P(B_i)}$$

**Vlastnosti pravděpodobnosti:**

- $0 \leq P(A) \leq 1$
- $P(\emptyset) = 0$
- $P(\bar{A}) = 1 - P(A)$
- $A \subset B \Rightarrow P(A) \leq P(B)$
- $P(B - A) = P(B) - P(A \cap B)$ , speciálně pro  $A \subset B \Rightarrow P(B - A) = P(B) - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ , speciálně pro  $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$
- $P(A \cup B) = 1 - P(\overline{A \cup B}) = 1 - P(\bar{A} \cap \bar{B})$ , z 1. de Morganova zákona
- $P(A \cap B) = P(A) \cdot P(B|A)$ , speciálně pro nezávislé jevy  $A, B \Rightarrow P(A \cap B) = P(A) \cdot P(B)$
- $P(A \cap B) = 1 - P(\overline{A \cap B}) = 1 - P(\bar{A} \cup \bar{B})$ , z 2. de Morganova zákona

## 2.1 Test z teorie

1. Určete, která z následujících tvrzení jsou pravdivá.
  - a) Klasická definice pravděpodobnosti vychází ze stability relativních četností.
  - b) Kolmogorovy axiomy pravděpodobnosti udávají návod ke stanovení pravděpodobnosti elementárních jevů.
  - c) Je-li pravděpodobnost jevu  $A$  rovna 0,75, pak pravděpodobnost podjevu jevu  $A$  je nejvýše 0,75.
  - d) Jestliže pravděpodobnosti dvou jevů z jevového pole  $\mathcal{A}$  jsou 0,7 a 0,5, pak tyto jevy nejsou disjunktní.
  - e) Pravděpodobnost, že při deseti hodech mincí padne desetkrát po sobě „panna“ je menší než pravděpodobnost, že při deseti hodech klasickou kostkou padne desetkrát po sobě sudé číslo.
2. Pravděpodobnost poruchy každé součástky je  $p$ . Předpokládejme, že součástky pracují nezávisle na sobě. Určete pravděpodobnost poruchy bloku složeného z 10 paralelně zapojených součástek. (Je-li funkční alespoň jedna součástka, blok funguje.)
  - a)  $p/10$
  - b)  $10 \cdot p$
  - c)  $10/p$
  - d)  $p^{10}$
  - e)  $1 - p^{10}$
  - f)  $(1 - p)^{10}$
  - g)  $1 - (1 - p)^{10}$
  - h)  $(1 - p)/10$
3. Pravděpodobnost poruchy každé součástky je  $p$ . Předpokládejme, že součástky pracují nezávisle na sobě. Určete pravděpodobnost poruchy bloku složeného z 10 sériově zapojených součástek. (Je-li porouchaná alespoň jedna součástka, blok nefunguje.)
  - a)  $p/10$
  - b)  $10 \cdot p$
  - c)  $10/p$
  - d)  $p^{10}$
  - e)  $1 - p^{10}$
  - f)  $(1 - p)^{10}$
  - g)  $1 - (1 - p)^{10}$
  - h)  $(1 - p)/10$
4. Podmíněná pravděpodobnost  $P(A|B)$  je rovna:
  - a)  $P(A \cap B) \cdot P(B)$
  - b)  $\frac{P(A \cap B)}{P(A)}, P(A) \neq 0$
  - c)  $P(A \cap B) \cdot P(A)$
  - d)  $\frac{P(A \cap B)}{P(B)}, P(B) \neq 0$

5. Mějme jevy  $A$  a  $B$ . Pravděpodobnost jevu  $A$  je  $P(A)$  a pravděpodobnost jevu  $B$  je  $P(B)$ . Pravděpodobnost sjednocení jevů  $A$  a  $B$  je rovna:
- a)  $P(A) + P(B)$
  - b)  $P(A) \cdot P(A)$
  - c)  $P(A) + P(B) - P(A \cap B)$
  - d)  $P(A|B) \cdot P(B)$
6. Mějme nezávislé jevy  $A$  a  $B$ . Pravděpodobnost jevu  $A$  je  $P(A)$  a pravděpodobnost jevu  $B$  je  $P(B)$ . Pravděpodobnost sjednocení jevu  $A$  a  $B$  je rovna:
- a)  $P(A) + P(B)$
  - b)  $P(A) \cdot P(A)$
  - c)  $P(A) + P(B) - P(A \cap B)$
  - d)  $P(A|B) \cdot P(B)$
7. Mějme disjunktní jevy  $A$  a  $B$ . Pravděpodobnost jevu  $A$  je  $P(A)$  a pravděpodobnost jevu  $B$  je  $P(B)$ . Pravděpodobnost průniku jevu  $A$  a  $B$  je rovna:
- a)  $P(A) + P(B)$
  - b)  $P(A) \cdot P(A)$
  - c)  $P(A) + P(B) - P(A \cap B)$
  - d) 0
8. Mějme jevy  $A$  a  $B$ . Jev  $C$  je průnikem jevů  $A$  a  $B$ . Pravděpodobnost jevu  $A$  je  $P(A)$  a pravděpodobnost jevů  $B$  je  $P(B)$ . Pravděpodobnost sjednocení jevu  $B$  a  $C$  vyjádřena pomocí pravděpodobností jevů  $A$  a  $B$  je rovna:
- a)  $P(A)$
  - b)  $P(B)$
  - c)  $P(B) \cdot (1 + P(A))$
  - d)  $P(B) \cdot (1 - P(A))$
  - e)  $P(B) \cdot (1 + P(A|B))$
  - f)  $P(B) \cdot (1 - P(A|B))$
9. Mějme nezávislé jevy  $A$  a  $B$ . Jev  $C$  je doplněk jevu  $A$ . Pravděpodobnost jevu  $A$  je  $P(A)$  a pravděpodobnost jevu  $B$  je  $P(B)$ . Pravděpodobnost průniku jevu  $B$  a  $C$  vyjádřena pomocí pravděpodobností jevů  $A$  a  $B$  je rovna:
- a)  $P(A)$
  - b)  $P(B)$
  - c)  $P(B) \cdot (1 + P(A))$
  - d)  $P(B) \cdot (1 - P(A))$
  - e)  $P(B) \cdot (1 + P(A|B))$
10. Vyberte 3 Kolmogorovovy axiomy pravděpodobnosti.
- a) Pravděpodobnost každého jevu  $A$  je nezáporné reálné číslo.
  - b) Pravděpodobnost každého jevu  $A$  je menší než 1.
  - c) Pravděpodobnost jistého jevu  $\Omega$  je rovna nule.
  - d) Pravděpodobnost jistého jevu  $\Omega$  je rovna jedné.
  - e) Pravděpodobnost sjednocení konečného počtu vzájemně disjunktních jevů je rovna součtu jejich pravděpodobností.
  - f) Pravděpodobnost sjednocení jevů je rovna součtu jejich pravděpodobností.

## 2.2 Příklady

1. Určete pravděpodobnost, že při hodu 20stěnnou spravedlivou (férovou) kostkou padne číslo větší než 14.

Řešení:

[0,3]

2. Určete pravděpodobnost, že při hodu 20stěnnou kostkou padne číslo větší než 14, víte-li, že sudá čísla padají 2x častěji než lichá.

Řešení:

[0,3]

3. Určete pravděpodobnost, že ve sportce uhodnete 4 čísla. (Ve sportce se losuje 6 čísel ze 49.)

Řešení:

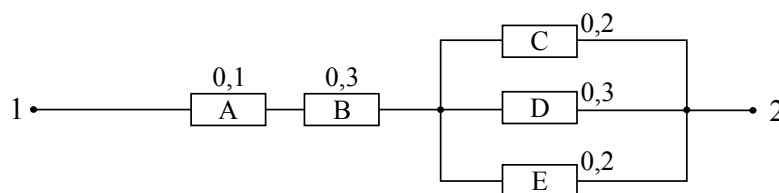
[0,001]

4. Z abecedního seznamu studentů zapsaných na dané cvičení vybere učitel prvních 12 a nabídne jim sázku: „Pokud se každý z Vás narodil v jiném znamení zvěrokruhu, dám každému z Vás 100 Kč. Pokud jsou však mezi Vámi alespoň dva studenti, kteří se narodili ve stejném znamení, dá mi každý z Vás 100 Kč.“ Vyplatí se studentům sázku přijmout? S jakou pravděpodobností studenti vyhrají?

Řešení:

$[5 \cdot 10^{-5}]$

5. Spočítejte pravděpodobnost toho, že z bodu 1 do bodu 2 bude protékat elektrický proud, je-li část el. obvodu včetně pravděpodobnosti poruch jednotlivých součástek vyznačen na následujícím obrázku. (Poruchy jednotlivých součástek jsou na sobě nezávislé.)



Řešení:

$[0,622]$

6. Ohrada má obdélníkový tvar, východní a západní stěna mají délku 40 m, jižní a severní pak 100 m. V této ohradě běhá kůň. Jaká je pravděpodobnost, že je k jižní stěně blíže než ke zbývajícím třem?

Řešení:

[0,4]

7. U pacienta je podezření na jednu ze čtyř vzájemně se vylučujících nemocí -  $N1, N2, N3, N4$  s pravděpodobnostmi výskytu  $P(N1) = 0,1, P(N2) = 0,2, P(N3) = 0,4, P(N4) = 0,3$ . Laboratorní zkouška  $A$  je pozitivní v případě první nemoci v 50 % případů, u druhé nemoci v 75 % případů, u třetí nemoci v 15 % případů a u čtvrté v 20 % případů. Jaká je pravděpodobnost, že výsledek laboratorní zkoušky bude pozitivní?

Řešení:

[0,32]

8. Telegrafické znaky se skládají ze signálů „tečka“, „čárka“. Je statisticky zjištěno, že se zkomolí 25 % sdělení „tečka“ a 20 % signálů „čárka“. Dále je známo, že signály se používají v poměru 3:2. Určete pravděpodobnost, že byl přijat správně signál, jestliže byl přijat signál „tečka“.

Řešení:

[0,849]

9. V jednom městě jezdí 85 % zelených taxíků a 15 % modrých. Svědek dopravní nehody vypověděl, že nehodu zavinil řidič modrého taxíku, který pak ujel. Testy provedené za obdobných světelných podmínek ukázaly, že svědek dobře identifikuje barvu taxíku v 80 % případů a ve 20 % případů se mýlí.

- a) Jaká je pravděpodobnost, že viník nehody skutečně řídil modrý taxík?

Řešení:

[0,414]

- b) Následně byl nalezen další nezávislý svědek, který rovněž tvrdí, že taxík byl modrý. Jaká je nyní pravděpodobnost, že viník nehody skutečně řídil modrý taxík?

Řešení:

[0,739]

- c) Ovlivní pravděpodobnost, že viník nehody skutečně řídil modrý taxík to, zda dva výše zmínění svědci vypovídali postupně nebo najednou?

Řešení:

[Ne]

Úlohu prezentovali psychologové Kahneman a Tversky

(Anděl; Matematika náhody; 2007)

10. Potřebujeme zjistit odpověď na určitou citlivou otázku (např. „Podváděl(a) jste u zkoušky?“, „Jste homosexuál (lesba)?“, apod.). Jak odhadnout, kolik procent dotazovaných na otázku odpoví ANO a přitom všem respondentům zaručit naprostou anonymitu? Jedním z řešení je tzv. **dvojitě anonymní anketa**:

Necháme respondenty hodit korunou a dvojkorunou a ti, kterým padl na koruně líc napíší na lísteček odpověď (ANO/NE) na citlivou otázku. Ostatní respondenti napíší, zda jim padl na dvojkoruně líc (ANO/NE). Jakým způsobem určíme podíl studentů, kteří na citlivou otázku odpověděli ANO?

Předpokládejme, že respondenti byli dotazováni, zda podváděli u zkoušky. Z anketních lístků se zjistilo, že „ANO“ odpovědělo 120 respondentů a „NE“ odpovědělo 200 respondentů. Kolik procent studentů podvádělo u zkoušky?

Řešení:

[přibližně 25 %]



---

## 3. Náhodná veličina

**Náhodná veličina** - funkce, která každému výsledku náhodného pokusu přiřadí reálné číslo. Je to matematický model popisující více či méně dobře realitu, který vytváříme, jestliže chceme zpracovávat výsledky náhodného pokusu.

**Hodnota náhodné veličiny** - výsledek náhodného pokusu vyjádřený reálným číslem.

**Rozdělení pravděpodobnosti** - pravidlo, které každé hodnotě (nebo intervalu hodnot) přiřazuje pravděpodobnost, že náhodná veličina nabude této hodnoty (nebo intervalu hodnot). Jestliže známe rozdělení pravděpodobnosti, je náhodná veličina z pravděpodobnostního hlediska úplně popsána. Nejčastějším způsobem popisu rozdělení pravděpodobnosti je distribuční funkce.

**Distribuční funkce** je reálná funkce definována jako  $F(x) = P(X < x)$ . Jde tedy o funkci, která každému reálnému číslu přiřazuje pravděpodobnost, že náhodná veličina nabývá hodnot menších než toto reálné číslo.

**Diskrétní náhodná veličina** je náhodná veličina, která může nabývat pouze konečného (výsledek hodu kostkou) nebo spočetně nekonečného (počet zákazníků snažících se dovolat do call centra během dne) množství hodnot. Diskrétní náhodnou veličinu popisujeme prostřednictvím **pravděpodobnostní funkce**, popř. distribuční funkce.

**Spojité náhodná veličina** je náhodnou veličinou, která má spojitou distribuční funkci. Pro popis spojitě náhodné veličiny používáme distribuční funkci a **hustotu pravděpodobnosti**.

Pravděpodobnost výskytu náhodné veličiny na nějakém intervalu určujeme na základě následujících vztahů.

$$P(X < a) = F(a)$$

$$P(X \geq b) = 1 - F(b)$$

$$P(a \leq X < b) = F(b) - F(a)$$

**Číselné charakteristiky diskrétní náhodné veličiny:**

- střední hodnota:  $E(X) = \mu = \sum_{(i)} x_i \cdot P(x_i)$
- rozptyl - výpočetní vztah:  $D(X) = E(X^2) - (E(X))^2$
- směrodatná odchylka:  $\sigma = \sqrt{D(X)}$
- modus:  $\hat{x}$  - taková hodnota DNV, v níž  $P(x_i)$  nabývá svého maxima

**Číselné charakteristiky spojitě náhodné veličiny:**

- střední hodnota:  $E(X) = \mu = \int_{-\infty}^{\infty} x \cdot f(x)$
- rozptyl - výpočetní vztah:  $D(X) = E(X^2) - (E(X))^2$
- směrodatná odchylka:  $\sigma = \sqrt{D(X)}$
- modus:  $\hat{x}$  - taková hodnota SNV, v níž  $f(x)$  nabývá svého maxima
- p-quantil:  $x_p$  - taková hodnota, že pravděpodobnost, že SNV nabude menších hodnot než  $x_p$  je 100p%:  $P(X < x_p) = p \Rightarrow F(x_p) = p$

### 3.1 Test z teorie

1. Vytvořte dvojice pojem - příklad.

- a) náhodný pokus      1) Doba přenosu testovacího datového souboru je delší než 30 s.
- b) náhodný jev        2) Měření doby přenosu testovacího datového souboru.
- c) náhodná veličina    3) Doba přenosu testovacího datového souboru.

2. Určete pravdivost následujících výroků.

- a) Náhodnou veličinu chápeme jako výsledek náhodného pokusu.
- b) Diskrétní náhodná veličina může nabývat konečného nebo spočetného množství hodnot.
- c) Distribuční funkce náhodné veličiny  $X$  v bodě  $t$  udává pravděpodobnost, že  $X$  nabývá hodnot menších než  $t$ .
- d) Má-li náhodná veličina spojitou distribuční funkci, je spojitá.
- e) Je-li  $X$  diskrétní náhodná veličina, pak  $\sum_i P(X = x_i) = 1$ .
- f) Oborem hodnot distribuční funkce jsou všechna reálná čísla.
- g) Medián je střední hodnota.
- h) Nabývá-li funkce  $f(x)$  hodnoty 1,3, nemůže jít o hustotu pravděpodobnosti.
- i) Rozdělení spojitě náhodné veličiny můžeme popsat distribuční funkcí a hustotou pravděpodobnosti.

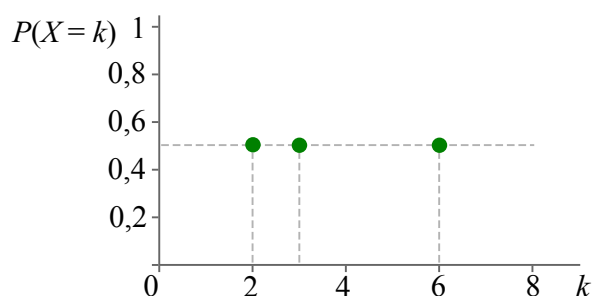
3. Určete, která ze zadaných funkcí nemůže představovat pravděpodobnostní funkci.

a) 
$$P(X = k) = \begin{cases} \frac{1}{k} & k \in \{2; 3; 6\} \\ 0 & k \notin \{2; 3; 6\} \end{cases}$$

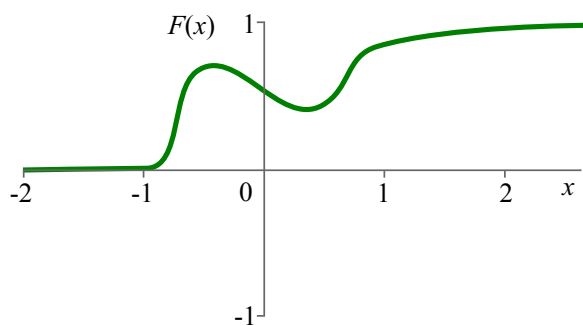
b)

$k$	2	3	6
$P(X = k)$	0,2	0,4	0,4

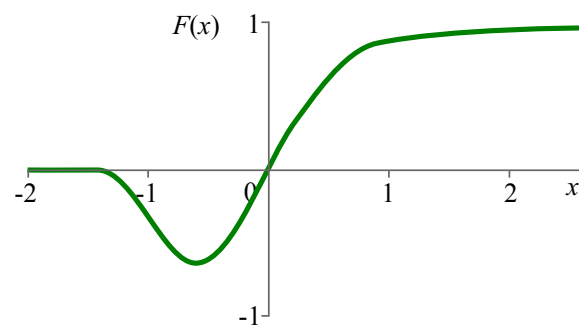
c)



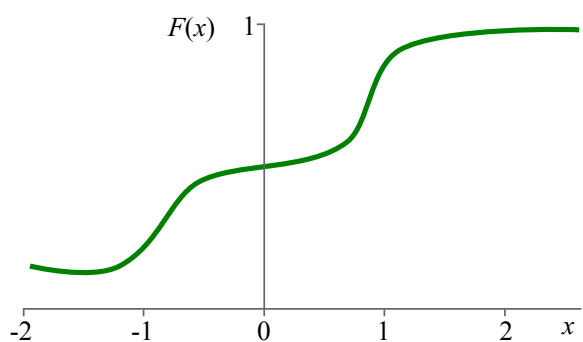
4. Určete, zda by grafy znázorněných funkcí mohly představovat distribuční funkci.



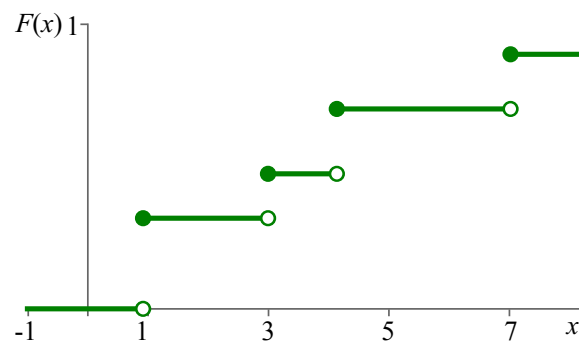
a)



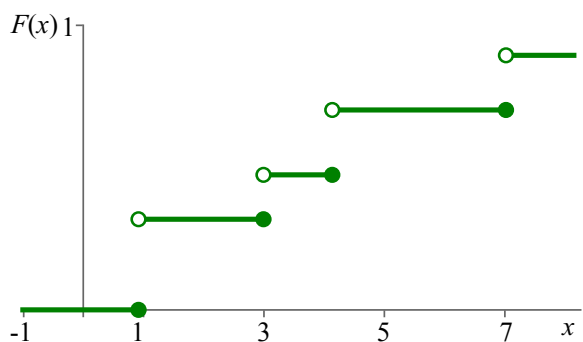
b)



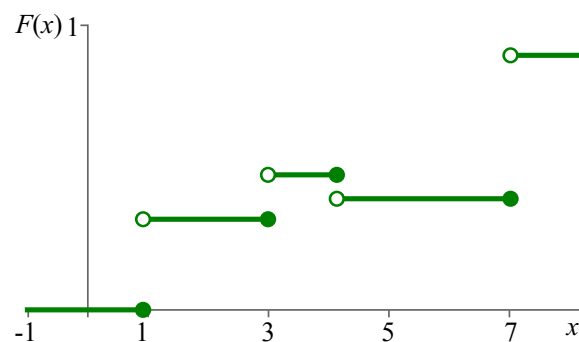
c)



d)

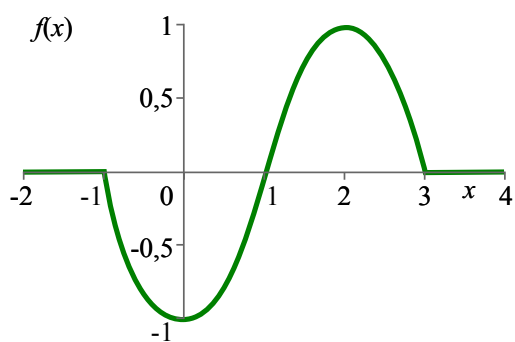


e)

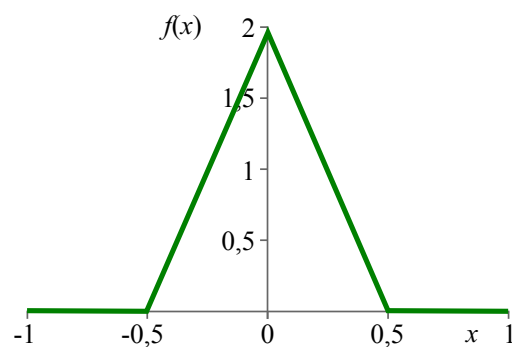


f)

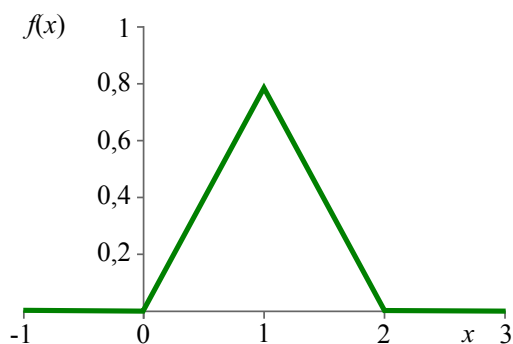
5. Určete, zda by grafy znázorněných funkcí mohly představovat hustotu pravděpodobnosti.



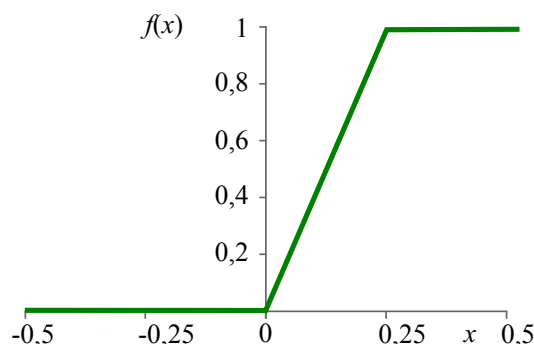
a)



b)



c)



d)

6. Necht náhodná veličina  $X$  představuje životnost (dobu do poruchy) monitorů na počítačové učebně E320. Určete pravdivost následujících výroků.
- $X$  je spojitou náhodnou veličinou.
  - Rozdělení  $X$  může být popsáno pravděpodobnostní funkcí.
  - Pro popis  $X$  lze použít intenzitu poruch.
7. Vyjádřete následující pravděpodobnosti pomocí distribuční funkce.
- $P(X < 10)$ ,
  - $P(X \geq 5)$ ,
  - $P(5 \leq X < 10)$ .
8. Necht  $X$  je diskrétní náhodná veličina. Vyjádřete co nejjednodušeji následující pravděpodobnosti pomocí  $P(X = 10)$ ,  $P(X < 10)$ ,  $P(X > 10)$ ,  $P(X = 5)$ ,  $P(X < 5)$ ,  $P(X > 5)$ .
- $P(X \leq 10)$ ,
  - $P(X \geq 10)$ ,
  - $P(5 < X \leq 10)$ ,
  - $P(5 \leq X \leq 10)$ .
9. Necht  $X$  je spojitá náhodná veličina. Vyjádřete co nejjednodušeji následující pravděpodobnosti pomocí  $P(X = 10)$ ,  $P(X < 10)$ ,  $P(X > 10)$ ,  $P(X = 5)$ ,  $P(X < 5)$ ,  $P(X > 5)$ .
- $P(X \leq 10)$ ,
  - $P(X \geq 10)$ ,
  - $P(5 < X \leq 10)$ ,
  - $P(5 \leq X \leq 10)$ .
10. Necht  $X$  je spojitá náhodná veličina. Vyjádřete následující pravděpodobnosti pomocí hustoty pravděpodobnosti.
- $P(X \leq 10)$ ,
  - $P(X \geq 10)$ ,
  - $P(5 < X \leq 10)$ ,
  - $P(5 \leq X \leq 10)$ .

## 3.2 Diskrétní náhodná veličina - příklady

1. Majitel servisního střediska nabídl prodejně automobilů, která si zřídila autopůjčovnu své služby. Za každý automobil zapůjčený jeho prostřednictvím obdrží od autopůjčovny 500,- Kč. Zároveň se však zavázal, že každý den investuje do údržby zapůjčených automobilů 800,- Kč. Počet automobilů zapůjčených prostřednictvím servisního střediska za 1 den je popsán následující pravděpodobnostní funkcí:

$x_i$	0	1	2	3	4	5	6
$P(x_i)$	0,01	0,40	0,25	0,15	0,10		0,03

- a) Hodnota pravděpodobnostní funkce pro 5 automobilů byla špatně čitelná. Určete ji.

Řešení:

[0,06]

- b) Určete a zakreslete distribuční funkci náhodné veličiny  $X$ , která je definována jako počet zapůjčených automobilů.

Řešení:

- c) Určete střední hodnotu, rozptyl, směrodatnou odchylku a modus počtu zapůjčených automobilů během jednoho dne.

Řešení:

$$[E(X) = 2,23; D(X) = 1,96; \sigma(X) = 1,4; \hat{x} = 1]$$

- d) Určete pravděpodobnostní funkci a distribuční funkci náhodné veličiny  $Y$ , která je definována jako denní příjem majitele servisu.

Řešení:

- e) Určete střední hodnotu, směrodatnou odchylku a modus příjmu majitele servisu ze zapůjčených automobilů během jednoho dne.

Řešení:

$$[E(Y) = 1\,115 \text{ Kč}; \sigma(Y) = 700 \text{ Kč}; \hat{y} = 500 \text{ Kč}]$$

- f) Určete pravděpodobnost, že příjem majitele servisu (náhodná veličina  $Y$ ) z půjčování automobilů převýší jeho výdaje.

Řešení:

$$[0,59]$$

- g) Určete střední hodnotu, směrodatnou odchylku a modus náhodné veličiny  $Z$ , která je definována jako zisk majitele servisu ze zapůjčených automobilů během jednoho dne.

Řešení:

$$[E(Z) = 315 \text{ Kč}; \sigma(Z) = 700 \text{ Kč}; \hat{z} = -300 \text{ Kč}]$$

2. Pro distribuční funkci náhodné veličiny  $X$  platí:

$$F(x) = \begin{cases} 0 & x \leq -1 \\ 0,3 & -1 < x \leq 0 \\ 0,7 & 0 < x \leq 1 \\ 1 & x > 1 \end{cases}$$

- a) Určete pravděpodobnostní funkci náhodné veličiny  $X$ , její střední hodnotu a směrodatnou odchylku.

Řešení:

$$[E(X) = 0; \sigma(X) = 0,77]$$

- b) Náhodná veličina  $Y = 1 - 3X$ , určete  $P(y), F(y), E(Y), D(Y)$ .

Řešení:

$$[E(Y) = 1; D(Y) = 5,4]$$

- c) Náhodná veličina  $W = 3X^2$ , určete  $P(w), F(w), E(W), D(W)$ .

Řešení:

$$[E(W) = 1,8; D(W) = 2,16]$$

3. V dílně jsou dva stroje pracující nezávisle na sobě. Pravděpodobnost poruchy prvního stroje je 0,2, pravděpodobnost poruchy druhého stroje je 0,3. Náhodná veličina  $X$  je definována jako počet současně porouchaných strojů. Určete:

a) pravděpodobnostní funkci náhodné veličiny  $X$ ,

Řešení:

b) distribuční funkci náhodné veličiny  $X$ ,

Řešení:

c) střední hodnotu a rozptyl náhodné veličiny  $X$ .

Řešení:

$$[E(X) = 0,50; D(X) = 0,37]$$



### 3.3 Spojitá náhodná veličina - příklady

1. Náhodná veličina  $X$  má distribuční funkci

$$F(x) = \begin{cases} 0 & x \leq 0 \\ cx^2 & 0 < x \leq 1 \\ 1 & x > 1 \end{cases}$$

Jaké hodnoty může nabývat konstanta  $c$ ?

Řešení:

[1]

2. Rozdělení náhodné veličiny  $X$  je dáno hustotou

$$f(x) = \begin{cases} 2x + 2 & x \in \langle -1; 0 \rangle \\ 0 & x \notin \langle -1; 0 \rangle \end{cases}$$

Určete:

- a)  $F(x)$ ,

Řešení:

b)  $P(-2 \leq X \leq -0,5), P(-2 \leq X \leq -1), P(X > 0,5), P(X = 0,3)$

Řešení:

$$[0,25; 0; 0; 0]$$

c) střední hodnotu, rozptyl a směrodatnou odchylku náhodné veličiny  $X$ .

Řešení:

$$\left[ E(X) = -\frac{1}{3}; D(X) = \frac{1}{18}; \sigma(X) = \frac{\sqrt{2}}{6} \right]$$

d) modus  $\hat{x}$

Řešení:

$$[0]$$

e) medián  $x_{0,5}$

Řešení:

$$\left[ -1 + \frac{\sqrt{2}}{2} \right]$$

3. Náhodná veličina  $Y$  je definována jako:  $Y = 3X + 1$ , kde  $X$  je náhodná veličina z předcházejícího příkladu. Určete:

a)  $F_Y(y)$

Řešení:

b)  $f_Y(y)$

Řešení:

c)  $E(Y), D(Y), \sigma(Y)$

Řešení:

$$\left[ E(Y) = 0; D(Y) = \frac{1}{2}; \sigma(Y) = \frac{\sqrt{2}}{2} \right]$$

---

## 4. Náhodný vektor

**Náhodným vektorem** rozumíme vektor  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  složený z náhodných veličin, který je charakterizován sdruženým rozdělením pravděpodobnosti. Ze sdruženého rozdělení pravděpodobnosti můžeme snadno najít marginální rozdělení pravděpodobnosti charakterizující jednotlivé složky náhodného vektoru.

**Podmíněné rozdělení pravděpodobnosti** určuje rozdělení NV  $X$  za předpokladu, že NV  $Y$  nabyla hodnoty  $y$ . Chápeme ho jako podíl sdruženého a marginálního rozdělení pravděpodobnosti (má-li tento podíl smysl), v souladu s definicí podmíněné pravděpodobnosti.

Podmíněná pravděpodobnostní funkce má tvar:

$$P(x|y) = \frac{p(x, y)}{P_Y(y)}, P_Y(y) \neq 0$$

$$P(y|x) = \frac{p(x, y)}{P_X(x)}, P_X(x) \neq 0$$

**Nezávislost** složek náhodného vektoru se projevuje tím, že jeho sdružená distribuční funkce (sdružená pravděpodobnostní funkce, resp. sdružená hustota pravděpodobnosti) se dá matematicky vyjádřit jako součin marginálních distribučních funkcí (marginálních pravděpodobností, resp. marginálních hustot pravděpodobnosti) jednotlivých náhodných veličin.

**Střední hodnota** náhodného vektoru  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  je dána jako vektor marginálních středních hodnot  $E(\mathbf{X}) = (E(X_1), E(X_2), \dots, E(X_n))^T$ .

**Rozptyl** náhodného vektoru  $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$  je dán jako vektor marginálních rozptylů  $D(\mathbf{X}) = (D(X_1), D(X_2), \dots, D(X_n))^T$ . Obdobně pro ostatní číselné charakteristiky.

**Kovariance**  $cov(X, Y)$  je nejjednodušší ukazatel vztahu mezi dvěma náhodnými veličinami. Kladná hodnota kovariance znamená, že se zvětšením hodnoty  $X$  se pravděpodobně zvýší i hodnota  $Y$ . Oproti tomu záporná hodnota kovariance informuje o tom, že se zvětšením hodnoty  $X$  se pravděpodobně sníží hodnota  $Y$ . Kovarianci nelze použít jako míru závislosti, protože  $cov(X, Y) \in (-\infty; \infty)$ .

$$cov(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$$

V praxi se často setkáváme s **kovarianční maticí**

$$var(X) = \begin{pmatrix} D(X) & cov(X, Y) \\ cov(Y, X) & D(Y) \end{pmatrix}$$

**Korelační koeficient**  $\rho(X, Y)$  je mírou lineární závislosti dvou složek náhodného vektoru.

$$\rho(X, Y) = \begin{cases} \frac{cov(X, Y)}{\sqrt{D(X) \cdot D(Y)}}, & D(X), D(Y) \neq 0 \\ 0, & \text{jinak.} \end{cases}$$

**Korelační matice**

$$cor(X) = \begin{pmatrix} 1 & \rho(X, Y) \\ \rho(Y, X) & 1 \end{pmatrix}$$

## 4.1 Test z teorie

1. Určete, zda jsou pravdivé následující výroky.
  - a) Náhodný vektor je definován jako dvourozměrný vektor, jehož složkami jsou náhodné veličiny.
  - b) Sdružené rozdělení popisuje rozdělení náhodného vektoru.
  - c) Marginální rozdělení popisuje rozdělení jednotlivých složek náhodného vektoru.
  - d) Je-li  $\mathbf{X} = (X, Y)^T$ , pak  $E(\mathbf{X}) = E(XY)$ .
  - e)  $E(XY) = E(X) \cdot E(Y)$ .
  - f) Marginální charakteristiky náhodného vektoru popisují vztah mezi náhodnými veličinami, které tvoří jeho složky.
  - g) Kovariance je mírou závislosti náhodných veličin.
  - h) Je-li  $cov(X, Y) = 0$ , pak jsou náhodné veličiny  $X$  a  $Y$  nezávislé.
  - i) Je-li  $cov(X, Y) = 0$ , jsou náhodné veličiny  $X$  a  $Y$  nekorelované.
  - j) Je-li  $\rho(X, Y) = 0$ , pak jsou náhodné veličiny  $X$  a  $Y$  nekorelované.
  - k) Jsou-li náhodné veličiny  $X$  a  $Y$  nekorelované, jsou lineárně nezávislé.
  - l)  $cov(X, X) = 1$ .
  - m)  $cov(X, Y) = cov(Y, X)$ .
  - n)  $\rho(X, X) = 1$ .
  - o)  $cov(X, Y) = E(XY) - E(X) \cdot E(Y)$ .

## 4.2 Příklady

1. Náhodný vektor  $\mathbf{Z} = (Y; X)^T$  má pravděpodobnostní funkci zadanou tabulkou:

$X \backslash Y$	1	2	3	4
3	0,01	0,02	0,03	0,25
5	0,04	0,16	?	0,05
7	0,12	0,07	0,06	0,01

Určete:

- a) chybějící hodnotu sdružené pravděpodobnostní funkce,

**Řešení:**

b)  $p(2; 5)$ ,

Řešení:

[0,16]

c)  $p(5; 2)$ ,

Řešení:

[0]

d)  $F(2, 8; 7, 1)$ ,

Řešení:

[0,42]

e)  $P(Y > 2, 1; X < 5, 3)$ ,

Řešení:

[0,51]

f)  $P(Y > 2, 1 | X < 5, 3)$ ,

Řešení:

[0,69]

g) marginální rozdělení NV  $X$  (marginální pravděpodobnostní f-ci i marginální distribuční f-ci),

Řešení:

h) marginální rozdělení NV  $Y$  (marginální pravděpodobnostní f-ci i marginální distribuční f-ci),

Řešení:

i)  $F_X(5, 3),$

Řešení:

[0,74]

j)  $F_Y(5, 3),$

Řešení:

[1]

k)  $P(X = 5|Y = 1),$

Řešení:

[0,24]

l)  $P(Y = 5|X = 1),$

Řešení:

[není definováno]

m) podmíněnou pravděpodobnostní funkci  $P(x|y),$

Řešení:

n) podmíněnou pravděpodobnostní funkci  $P(y|x)$ ,

Řešení:

o) základní číselné charakteristiky NV  $X$ , tj.  $E(X), D(X), \sigma(X), \hat{X}$ ,

Řešení:

$$[E(X) = 4, 9; D(X) = 2, 3; \sigma_X = 1, 5; \hat{X} = 5]$$

p) základní číselné charakteristiky NV  $Y$ , tj.  $E(Y), D(Y), \sigma(Y), \hat{Y}$ ,

Řešení:



$$[E(Y) = 2, 7; D(Y) = 1, 2; \sigma_Y = 1, 1; \hat{Y} = 4]$$

q) základní číselné charakteristiky náhodného vektoru  $\mathbf{Z} = (Y; X)^T$ , tj.  $E(\mathbf{Z})$ ,  $D(\mathbf{Z})$ ,  $\sigma(\mathbf{Z})$ ,

Řešení:

$$[E(\mathbf{Z}) = (2, 7; 4, 9), D(\mathbf{Z}) = (1, 2; 2, 3), \sigma(\mathbf{Z}) = (1, 1; 1, 5)]$$

r)  $E(X|Y = 2)$ ,

Řešení:

$$[5, 4]$$

s) kovarianci  $cov(Y; X)$ ,

Řešení:

$$[-1, 1]$$

t) koeficient korelace  $\rho(Y; X)$ .

Řešení:

$$[-0, 7]$$

u) Jsou NV  $X$  a  $Y$  nezávislé?

Řešení:

[nejsou]

v) Jsou NV  $X$  a  $Y$  lineárně nezávislé ?

Řešení:

[nejsou]

2. Náhodný vektor  $\mathbf{Z} = (X; Y)^T$  nabývá hodnot  $(0; 1)^T$  s pravděpodobností  $1/2$ , hodnoty  $(0; 2)^T$  s pravděpodobností  $1/3$  a hodnoty  $(1; 1)^T$  s pravděpodobností  $1/6$ .

a) Určete korelační koeficient.

Řešení:

$[-0, 32]$

b) Rozhodněte, zda jsou NV  $X$  a  $Y$  nezávislé.

Řešení:

[nejsou]

c) Rozhodněte, zda jsou NV  $X$  a  $Y$  lineárně nezávislé.

Řešení:

[nejsou]

## 5. Vybraná rozdělení diskrétní náhodné veličiny

Název NV $X$	Popis	Pravděpodobnostní funkce	$E(X)$	$D(X)$
<b>Binomická</b> $Bi(n, \pi)$	počet úspěchů v $n$ Bernoulliho pokusech	$P(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}$	$n\pi$	$n\pi(1 - \pi)$
<b>Hypergeometrická</b> $H(N, M, n)$	počet úspěchů v $n$ <b>závislých</b> pokusech	$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$		
<b>Alternativní</b> $A(\pi)$	počet úspěchů v 1 pokusu	$P(X = 1) = \pi$ $P(X = 0) = 1 - \pi$	$\pi$	$\pi(1 - \pi)$
<b>Geometrická</b> $Ge(\pi)$	počet pokusů do 1. úspěchu (včetně)	$P(X = n) = \pi(1 - \pi)^{n-1}$	$\frac{1}{\pi}$	$\frac{1 - \pi}{\pi^2}$
<b>Negativně binomická (Pascalova)</b> $NB(k, \pi)$	počet pokusů do $k$ . úspěchu (včetně)	$P(X = n) = \binom{n-1}{k-1} \pi^k (1 - \pi)^{n-k}$	$\frac{k}{\pi}$	$\frac{k(1 - \pi)}{\pi^2}$
<b>Poissonova</b> $Po(\lambda t)$	počet událostí v Poissonově procesu v uzavřené oblasti (v čase, na ploše, v objemu)	$P(X = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$	$\lambda t$	$\lambda t$

Popis	Podmínky		Název NV $X$
počet úspěchů v $n$ pokusech	nezávislé pokusy	$n = 1$	<b>Alternativní</b> - $A(\pi)$
		$n \geq 1$	<b>Binomická</b> - $Bi(n, \pi)$
	závislé pokusy		<b>Hypergeometrická</b> - $H(N, M, n)$
počet pokusů do $k$ . úspěchu (včetně)	nezávislé pokusy	$k = 1$	<b>Geometrická</b> - $Ge(\pi)$
		$k \geq 1$	<b>Negativně binomická</b> - $NB(k, \pi)$
počet události v uzavřené oblasti (v čase, na ploše, v objemu)	ordinarita, stacionarita, beznáslednost procesu		<b>Poissonova</b> - $Po(\lambda t)$

## 5.1 Test z teorie

1. Určete pravdivost následujících tvrzení.
  - a) Rozdělení pravděpodobnosti diskrétní náhodné veličiny může být dáno výhradně pravděpodobnostní funkcí.
  - b) Posloupnost nezávislých pokusů majících pouze dva možné výsledky se stejnou pravděpodobností úspěchu nazýváme Bernoulliho pokusy.
  - c) Počet úspěchů v  $n$  pokusech lze popsat binomickou náhodnou veličinou.
  - d) Geometrické rozdělení je speciálním případem negativně binomického rozdělení.
  - e) Pascalovo rozdělení je pouze jiný název pro negativně binomické rozdělení.
  - f) Jistý supermarket má otevřeno 24h denně. Počet zákazníků v supermarketu během otevírací doby lze popsat náhodnou veličinou s Poissonovým rozdělením.
2. Charakterizujte rozdělení náhodné veličiny popisující
  - a) počet studentů, kteří úspěšně ukončí kurz STA1 v tomto semestru (z minulých let víme, že pravděpodobnost, že student úspěšně dokončí kurz STA1 je 0,63; do kurzu je v tomto semestru přihlášeno 248 studentů),
  - b) počet vadných mikroprocesorů na chipu (na chipu je průměrně 1 vadný mikroprocesor),
  - c) počet hodů poctivou kostkou nutných k padnutí šestky,
  - d) počet řidičů obslužených na čerpací stanici za půl hodiny (během 1h je na čerpací stanici obsluženo průměrně 72 řidičů),
  - e) počet řidičů obslužených do chvíle, kdy 1. řidič ujede bez placení (průměrně ujíždí bez placení 1 z 50 řidičů),
  - f) počet týdnů v roce (52 týdnů), v nichž neujede žádný řidič z čerpací stanice bez placení (během týdne je na čerpací stanici obsluženo průměrně 4 000 řidičů, z nichž cca 2 % ujedou bez placení),
  - g) počet řidičů obslužených do chvíle, kdy 4. řidič ujede bez placení (průměrně ujíždí bez placení 1 z 50 řidičů).

## 5.2 Příklady

1. Bridž se hraje s 52 bridžovými kartami, které se rozdají mezi 4 hráče. Vždy 2 hráči hrají spolu. Při rozdávání (13 karet) jste dostali do rukou 2 esa. Jaká je pravděpodobnost, že váš partner bude mít zbývající dvě esa?

Řešení:

[0,11]

2. Pokusy se zjistilo, že radioaktivní látka vyzařuje během 7,5 s průměrně 3,87  $\alpha$ -částice. Určete pravděpodobnost toho, že za 1 sekundu vyzáří tato látka alespoň jednu  $\alpha$ -částici.

Řešení:

[0,40]

3. Kamarád vás pošle do sklepa, abyste donesl(a) 4 lahvová piva - z toho dvě desítky a dvě dvanáctky. Nevíte, kde rozsvítit, proto vezmete z basy poslepu 4 láhve. S jakou pravděpodobností jste vyhověl(a), víte-li, že v base bylo celkem 10 desítek a 6 dvanáctek?

Řešení:

[0,37]

4. V jednom mililitru určitého dokonale rozmíchaného roztoku se v průměru nachází 15 určitých mikroorganismů. Určete pravděpodobnost, že při náhodném výběru vzorku o objemu  $1/2$  mililitru bude ve zkumavce méně než 5 těchto mikroorganismů.

Řešení:

[0,13]

5. Na stůl vysypeme 15 mincí. Jaká je pravděpodobnost, že počet mincí ležících lícem nahoře, je od 8 do 15?

Řešení:

[0,50]

6. Pravděpodobnost, že se dovoláme do studia rozhlasové stanice, která právě vyhlásila telefonickou soutěž je 0,08. Jaká je pravděpodobnost, že se dovoláme nejvýše na 4. pokus?

Řešení:

[0,28]

7. V továrně se vyrobí denně 10 % vadných součástek. Jaká je pravděpodobnost, že vybereme-li třicet součástek z denní produkce, tak nejméně dvě budou vadné?

Řešení:

[0,82]

8. Ve skladu je 200 součástek. 10 % z nich je vadných. Jaká je pravděpodobnost, že vybereme-li ze skladu třicet součástek, tak nejméně dvě budou vadné?

Řešení:

[0,84]

9. V určité firmě bylo zjištěno, že na 33 % počítačů je nainstalován nějaký nelegální software. Určete pravděpodobnostní a distribuční funkci počtu počítačů s nelegálním softwarem mezi třemi kontrolovanými počítači.

Řešení:

10. Sportka je loterijní hra, v níž sázející tipuje šest čísel ze čtyřiceti devíti, která očekává, že padnou při budoucím slosování. K účasti ve hře je nutné zvolit alespoň jednu kombinaci 6 čísel (vždy 6 čísel na jeden sloupec) a pomocí křížků tato čísla označit na sázence společnosti Sazka a.s. do sloupců, počínaje sloupцем prvním. Sázející vyhrává v případě, že uhodne alespoň tři čísla z tažené šestice čísel. Jaká je pravděpodobnost, že proto, aby sázející vyhrál, bude muset vyplnit:

- a) právě tři sloupce,

Řešení:

[0,02]

- b) alespoň 5 sloupců,

Řešení:

[0,93]

- c) méně než 10 sloupců,

Řešení:

[0,16]

- d) více než 5 a nejvýše 10 sloupců?

Řešení:

[0,08]



## 6. Vybraná rozdělení spojité náhodné veličiny

Rozdělení NV	Popis	Hustota pravděpodobnosti Distribuční funkce Intenzita poruch	$E(X)$	$D(X)$
<b>Rovnoměrné</b> $Ro(a, b)$	$f(x)$ je na $(a; b)$ konstantní, jinde nulová	$f(x) = \begin{cases} \frac{1}{b-a} & x \in \langle a; b \rangle \\ 0 & \text{jinde} \end{cases}$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$
<b>Exponenciální</b> $Exp(\lambda)$	doba do 1. události, doba mezi událostmi (pouze v období stabilního života)	$\begin{aligned} f(t) &= \lambda \cdot e^{-\lambda t}; t > 0; \lambda > 0 \\ F(t) &= 1 - e^{-\lambda t}; t > 0; \lambda > 0 \\ \lambda(t) &= \lambda = \text{konst.}; t > 0; \lambda > 0 \end{aligned}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
<b>Erlangovo</b> $Erlang(k; \lambda)$	doba do k-té události	$\begin{aligned} f(t) &= \lambda \cdot e^{-\lambda t} \cdot \frac{(\lambda t)^{k-1}}{(k-1)!}; t > 0 \\ F(t) &= 1 - e^{-\lambda t} \cdot \sum_{j=0}^{k-1} \frac{(\lambda t)^j}{j!} \\ \lambda(t) &= \frac{\lambda}{(k-1)! \sum_{j=0}^{k-1} \frac{1}{(k-1-j)! (\lambda t)^j}} \end{aligned}$	$\frac{k}{\lambda}$	$\frac{k}{\lambda^2}$
<b>Weibullovo</b> $Wb(\Theta, \beta)$	doba do 1. události (poru- chy) (vhodná volba $\beta$ umožňuje použití v libovolném období intenzity poruch)	$\begin{aligned} f(t) &= \frac{\beta}{\Theta} \left( \frac{t}{\Theta} \right)^{\beta-1} e^{-\left(\frac{t}{\Theta}\right)^\beta} \\ F(t) &= 1 - e^{-\left(\frac{t}{\Theta}\right)^\beta} \\ \lambda(t) &= \frac{\beta}{\Theta} \left( \frac{t}{\Theta} \right)^{\beta-1} t > 0; \Theta > 0; \beta > 0 \end{aligned}$		
<b>Normované normální</b> $N(0, 1)$	hodnoty distribuční funkce jsou tabelovány, hustota pravděpodobnosti je sudá funkce - Gaussův klobouk	$\begin{aligned} \phi(x) &= \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}; -\infty < x < \infty \\ \Phi(x) &= \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \end{aligned}$	0	1
<b>Normální</b> $N(\mu; \sigma^2)$	distribuční funkci určíme pomocí standardizace normální náhodné veličiny	$\begin{aligned} f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\left(\frac{x-\mu}{\sqrt{2\sigma}}\right)^2} \\ F(x) &= \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^x e^{-\left(\frac{t-\mu}{\sqrt{2\sigma}}\right)^2} dt \end{aligned}$	$\mu$	$\sigma^2$
<b>Logaritmicko- normální</b> $LN(\mu; \sigma^2)$	distribuční funkci určíme převodem na dis. funkci norm. normálního rozdělení	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\left(\frac{\ln(x)-\mu}{\sqrt{2\sigma}}\right)^2}; x > 0$	$e^{\mu+\frac{\sigma^2}{2}}$	$e^{2\mu+\sigma^2} \cdot (e^{\sigma^2}-1)$

## 6.1 Test z teorie

1. Určete pravdivost následujících tvrzení.

- a) Intenzita poruch (hazardní funkce) je neklesající funkce.
- b) Exponenciální rozdělení používáme k modelování životnosti výrobků nacházejících se v období stárnutí.
- c) Exponenciální rozdělení je speciálním případem Weibullova rozdělení.
- d) Weibullovo rozdělení lze použít k modelování životnosti výrobků nacházejících se v libovolném období života.
- e) Normální rozdělení má právě jeden parametr.
- f) Hustota pravděpodobnosti normální náhodné veličiny je sudá funkce.
- g) Distribuční funkce normální náhodné veličiny je tabelována.
- h) Má-li náhodná veličina normální rozdělení, pak (střední hodnota = medián = modus).
- i) Má-li náhodná veličina normální rozdělení se střední hodnotou  $\mu$  a sm. odchylkou  $\sigma$ , pak přibližně 5 % hodnot náhodné veličiny leží mimo interval  $\langle \mu - 3\sigma; \mu + 3\sigma \rangle$ .

2. Doplňte:

- a) Intenzitu poruch lze použít k popisu ..... spojitých náhodných veličin.
- b) Exponenciální rozdělení používáme k modelování životnosti výrobků nacházejících se v období .....
- c) Pro modelování životnosti výrobku, který má lineárně rostoucí intenzitu poruch lze použít Weibullovo rozdělení s parametrem tvaru  $\beta = \dots\dots\dots$
- d) Gaussova křivka je grafem ..... normálního rozdělení.

## 6.2 Příklady

1. Výška v populaci chlapců ve věku 3,5-4 roky má normální rozdělení se střední hodnotou 102 cm a směrodatnou odchylkou 4,5 cm. Určete, jaké procento chlapců v uvedeném věku má výšku menší nebo rovnou 93 cm.

Řešení:

[2,28 %]

2. Průměrná životnost strojní součástky je 30 000 hodin. Předpokládejme, že součástka je v období stabilního života. Určete:

- a) pravděpodobnost, že součástka nevydrží více než 2 000 hodin,

Řešení:

[0,06]

- b) pravděpodobnost, že součástka vydrží více než 35 000 hodin,

Řešení:

[0,31]

- c) dobu, do níž se porouchá 95 % součástek.

Řešení:

[89 872 h]

3. Výrobní zařízení má poruchu v průměru jednou za 2000 hodin. Veličina  $Y$  představující dobu čekání na poruchu má exponenciální rozdělení. Určete dobu  $T_0$  tak, aby pravděpodobnost, že přístroj bude pracovat delší dobu než  $T_0$ , byla 0,99.

Řešení:

[20,1 h]

4. Výsledky měření jsou zatíženy jen normálně rozdělenou chybou s nulovou střední hodnotou a se směrodatnou odchylkou 3 mm. Jaká je pravděpodobnost, že při 3 měřeních bude alespoň jednou chyba v intervalu (0 mm; 2,4mm)?

Řešení:

[0,64]

5. Ve velké počítačové síti se průměrně přihlašuje 25 uživatelů za hodinu. Určete pravděpodobnost, že:

a) se nikdo nepřihlásí během 14:30 - 14:36,

Řešení:

[0,08]

b) do dalšího přihlášení uplynou 2-3 minuty.

Řešení:

[0,15]

c) Určete maximální délku časového intervalu tak, aby pravděpodobnost, že se nikdo nepřihlásí byla alespoň 0,90.

Řešení:

[0,25 min]

6. Náhodná veličina  $X$  má normální rozdělení  $N(\mu; \sigma^2)$ . Určete:

a)  $P(\mu - 2\sigma < X < \mu + 2\sigma)$ ,

Řešení:

[0,955]

b)  $k \in \mathbb{Z}$  tak, aby  $P(\mu - k\sigma < X < \mu + k\sigma) > 0,99$ .

Řešení:

[2,58]

7. Na prohlídce výstavy je promítán doprovodný film o životě autora vystavovaných děl. Jeho projekce začíná každých 20 minut. Určete pravděpodobnost, že pokud náhodně přijdete do promítacího sálu,

a) nebudete na začátek filmu čekat víc než 5 minut,

Řešení:

[0,25]

b) budete čekat mezi 5 a 10 minutami,

Řešení:

[0,25]

c) střední hodnotu a směrodatnou odchylku doby čekání na začátek filmu.

Řešení:

$[E(X) = 10,0 \text{ min}; \sigma = 5,8 \text{ min}]$

8. Při kontrole jakosti přebíráme součástku pouze tehdy, jestliže se její rozměr pohybuje v mezích 26-27 mm. Rozměry součástek mají normální rozdělení se střední hodnotou 26,4 mm a směrodatnou odchylkou 0,2 mm. Jaká je pravděpodobnost, že rozměr součástky náhodně vybrané ke kontrole bude v požadovaných mezích?

Řešení:

[0,98]

9. Délka skoků sportovce Jakuba měřená v cm má normální rozdělení  $N(\mu_1; \sigma_1^2)$ , kde  $\mu_1 = 690$  a  $\sigma_1 = 10$ . Délka skoků sportovce Aleše měřená v cm má také normální rozdělení  $N(\mu_2; \sigma_2^2)$ , kde  $\mu_2 = 705$  a  $\sigma_2 = 15$ . Na závody se kvalifikuje ten, kdo ze dvou skoků alespoň jednou skočí více než 700 cm.

a) S jakou pravděpodobností se oba dva kvalifikují na závody?

Řešení:

[0,25 ]

b) S jakou pravděpodobností se kvalifikuje Aleš, ale Jakub ne?

Řešení:

[0,61]



## 7. Explorační analýza dat

Kvantitativní - Numerická proměnná

- Míry polohy

Průměr	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
Kvantily	dolní kvartil, horní kvartil, medián

- Míry variability

Variační rozpětí	$x_{max} - x_{min}$
Interkvartilové rozpětí	$IQR = x_{0,75} - x_{0,25}$
Výběrový rozptyl	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$
Výběrová směrodatná odchylka	$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$
Variační koeficient	$V_x = \frac{s}{\bar{x}}, \text{ popř. } V_x = \frac{s}{\bar{x}} \cdot 100 (\%)$

- Míry šikmosti a špičatosti

Výběrová šikmost	$\alpha = \frac{n}{(n-1)(n-2)} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{s^3}$
Výběrová špičatost	$\beta = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \cdot \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{s^4} - 3 \frac{(n-1)^2}{(n-2)(n-3)}$

- Identifikace odlehlých pozorování

Vnitřní hradby	dolní mez: $h_D = x_{0,25} - 1,5 \cdot IQR$ horní mez: $h_H = x_{0,75} + 1,5 \cdot IQR$
Z – souřadnice	$z\text{-skóre} = \frac{x_i - \bar{x}}{s}$
Mediánová souřadnice	$x_{0,5}\text{-skóre} = \frac{x_i - x_{0,5}}{1,483 \cdot MAD}$

## 7.1 Test z teorie

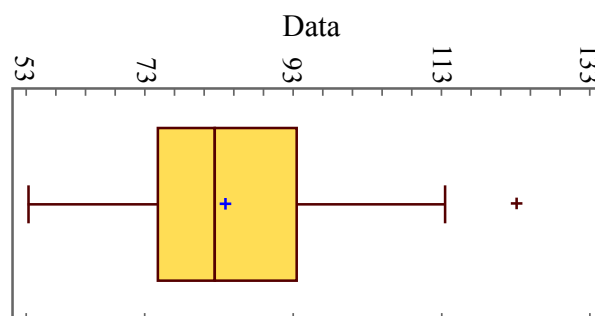
1. Test ze Statistiky píše velké množství studentů. Představte si, že každý z nich odpoví správně přesně na polovinu otázek. V tomto případě bude směrodatná odchylka počtu správných odpovědí
  - a) rovna průměru,
  - b) rovna mediánu,
  - c) rovna nule,
  - d) směrodatnou odchylku nelze určit bez dalších informací.
2. Největší kumulativní absolutní četnost v množině čísel se rovná
  - a) součtu všech absolutních četností,
  - b) 1,
  - c) dvojnásobku průměru,
  - d) dvojnásobku mediánu,
  - e) dvojnásobku módu.
3. Několik studentů píše test ze Statistiky s 10 otázkami. Nejhorší výsledek jsou 3 správné odpovědi, nejlepší výsledek je 10 správných odpovědí. Jakou hodnotu má medián?
  - a)  $7 = (10-3)$ ,
  - b)  $6,5 = \frac{3+10}{2}$ ,
  - c) medián nelze určit, pokud neznáme konkrétní výsledky jednotlivých žáků.
4. Představte si, že jste absolvovali normovaný test (např. SCIO test) a že Vám sdělili, že patříte do 91. percentilu. To znamená, že
  - a) 90 žáků, kteří se podrobili stejnému testu, dosáhlo vyšších výsledků než vy,
  - b) 90 žáků, kteří se podrobili stejnému testu, dosáhlo nižších výsledků než vy,
  - c) 90 % žáků, kteří se podrobili stejnému testu, dosáhlo vyšších výsledků než vy,
  - d) 90 % žáků, kteří se podrobili stejnému testu, dosáhlo nižších výsledků než vy.
5. Průměrná mzda je 60% kvantil mzdy. Lze tedy říci, že
  - a) medián mzdy je vyšší než průměrná mzda,
  - b) medián mzdy je nižší než průměrná mzda,
  - c) medián mzdy je stejný jako průměrná mzda,
  - d) o vztahu mezi mediánem mzdy a průměrnou mzdou nelze rozhodnout.
6. Průměrná mzda je 60% kvantil mzdy. Lze tedy říci, že
  - a) mzdy mají kladnou šikmost,
  - b) mzdy mají zápornou šikmost,
  - c) mzdy mají kladnou špičatost,
  - d) mzdy mají zápornou špičatost,
  - e) vztah mezi průměrem a 60% kvantilem nevypovídá nic o šikmosti ani o špičatosti dat.

7. Lékař Petře sdělil, že patří do 3. percentilu ohledně BMI (Body mass index - poměr váhy (kg) ke kvadrátu výšky (m)). Petra má pravděpodobně
- a) podváhu,
  - b) normální váhu,
  - c) nadváhu,
  - d) bez dalších informací nelze usuzovat na Petřinu váhu.
8. Představte si, že jste absolvovali normovaný test (např. SCIO test). Měl(a) jste lepší výsledek než 85 studentů ze 100. To znamená, že
- a) patříte do 99. decilu,
  - b) patříte do 95. decilu,
  - c) patříte do 10. decilu,
  - d) patříte do 9. decilu,
  - e) patříte do 2. kvartilu.
9. Pro srovnání variability váhy a výšky je možné použít
- a) průměr,
  - b) rozptyl,
  - c) směrodatnou odchylku,
  - d) variační koeficient,
  - e) šikmost.
10. Zvýšíme-li každému zaměstnanci ve firmě plat o 100,- Kč, průměrný plat ve firmě se zvýší
- a) o 100,- Kč,
  - b) o 1 000,- Kč,
  - c) průměrný plat se nezmění.
11. Zvýšíme-li každému zaměstnanci ve firmě plat dvojnásobně, průměrný plat ve firmě se zvýší
- a) dvojnásobně,
  - b) čtyřnásobně,
  - c) průměrný plat se nezmění.
12. Zvýšíme-li každému zaměstnanci ve firmě plat o 20%, průměrný plat ve firmě se zvýší
- a) o 20%,
  - b) o 400%,
  - c) o 40%,
  - d) o 44%,
  - e) průměrný plat se nezmění.
13. Zvýšíme-li každému zaměstnanci ve firmě plat o 100,- Kč, rozptyl platů ve firmě se zvýší
- a) o 100,- Kč,
  - b) o 1 000,- Kč,
  - c) rozptyl platů se nezmění.

14. Zvýšíme-li každému zaměstnanci ve firmě plat dvojnásobně, rozptyl platů ve firmě se zvýší
- dvojnásobně,
  - čtyřnásobně,
  - průměrný plat se nezmění.
15. Zvýšíme-li každému zaměstnanci ve firmě plat o 20%, rozptyl platů ve firmě se zvýší
- o 20%,
  - o 400%,
  - o 40%,
  - o 44%,
  - průměrný plat se nezmění.
16. Největší kumulativní relativní četnost se rovná
- dvojnásobku průměru,
  - dvojnásobku mediánu,
  - dvojnásobku módu,
  - součtu všech jednotlivých hodnot absolutních četností,
  - 1.
17. Určete, zda jsou následující tvrzení pravdivá.
- Geometrický průměr je definován pro proměnné, které nabývají pouze kladných hodnot.
  - Jedna čtvrtina hodnot je větší než 25% kvantil, zatímco tři čtvrtiny hodnot jsou menší.
  - Mají-li dvě proměnné stejný průměr a stejný rozptyl, mají stejný variační koeficient.
  - Mzdy v ČR mají kladnou šikmost. (V ČR mají zhruba 2/3 lidí podprůměrný plat.)
  - Nejčtenější hodnota v souboru se nazývá medián.
  - Rozptyl má vždy kladnou hodnotu.

18. V grafu na obrázku, modrý křížek označuje

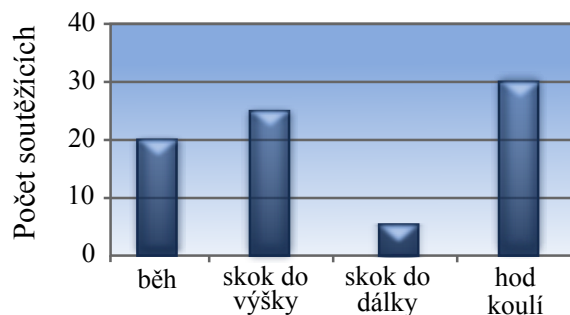
- medián,
- průměr,
- modus,
- interkvartilové rozpětí (IQR).



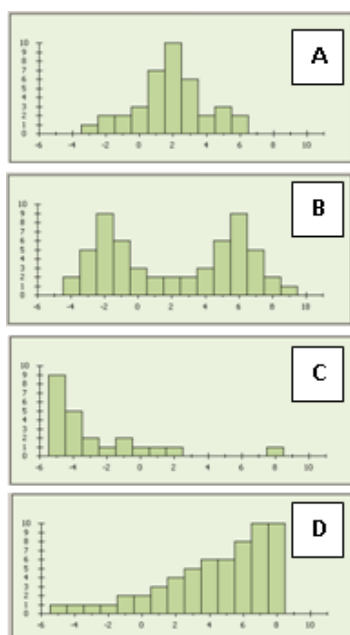
19. Určete, zda jsou následující tvrzení pravdivá. Proměnná znázorněna na předcházejícím obrázku
- neobsahuje odlehlá pozorování,
  - má kladnou šikmost,
  - je kladná,
  - má více než polovinu hodnot větších než 83.

20. Na atletických závodech mládeže žáci soutěžili ve 4 kategoriích. Určete pravdivost výroků.

- Na obrázku je znázorněn histogram a nejméně soutěžících bylo ve skoku do dálky.
- Celkem ve čtyřech kategoriích soutěžilo 80 žáků.
- Modus = hod koulí.
- Modus = 30.



21. Spárujte histogramy s odpovídajícími krabicovými grafy.



## 7.2 Příklady

1. V jistém supermarketu byla ve stejné chvíli na 8 pokladnách měřena doba, během které pokladní ověří platnost platební karty zákazníka v bance. U pěti zákazníků trvalo ověření 2 minuty, u zbývajících tří to byly 3 minuty. Určete průměrnou dobu potřebnou k ověření platnosti karty.

Řešení:

[2,29 minut]

2. Při sledování proměnné  $x$  byl určen aritmetický průměr 110 a rozptyl 800. Dodatečně byly zjištěny chyby u dvou údajů. Místo 85 mělo být správně 95 a místo 120 má být 150. Ostatních 18 údajů bylo správných. Opravte vypočítané charakteristiky (průměr a rozptyl).

Řešení:

a) Průměr:

[112]

b) Rozptyl:

[853,7]

3. Ze čtyřiceti hodnot byl vypočítán aritmetický průměr 7,50 a rozptyl 2,25. Při kontrole bylo zjištěno, že chybí dvě hodnoty proměnné:  $-3,8$  a  $7$ . Opravte uvedené charakteristiky.

Řešení:

- a) Průměr:

[7,2]

b) Rozptyl:

[5,5]

4. V důsledku výstavby satelitního městečka poklesl průměrný věk obyvatel vesnice o 19%, rozptyl věku vzrostl o 21%. Jak se změnil variační koeficient?

Řešení:

[35,8 %]



## 8. Výběrové charakteristiky

Mějme náhodný výběr  $\mathbf{X}$  z normálního rozdělení:

$$\mathbf{X} = (X_1, \dots, X_n), \forall i = 1, \dots, n : X_i \rightarrow N(\mu, \sigma).$$

Výběrová charakteristika	Rozdělení pravděpodobnosti	Poznámka
$\frac{\bar{X} - \mu}{\sigma} \sqrt{n}$	$N(0; 1)$	viz CLV
$\frac{\bar{X} - \mu}{S} \sqrt{n}$	$t_{n-1}$	viz vlastnosti Studentova rozdělení
$\frac{S^2}{\sigma^2} (n-1)$	$\chi_{n-1}^2$	viz vlastnosti $\chi^2$ - rozdělení
$\frac{p - \pi}{\sqrt{\pi(1-\pi)}} \sqrt{n}$	$N(0; 1)$	viz vlastnosti reativní četnosti, předpoklad: $n > \frac{9}{p(1-p)}$

Mějme dva nezávislé výběry z normálního rozdělení:

$$\forall i = 1, 2, \dots, n_1, \text{ kde } n_1 \text{ je rozsah prvního výběru: } X_{1i} \rightarrow N(\mu_1, \sigma_1^2)$$

$$\forall j = 1, 2, \dots, n_2, \text{ kde } n_2 \text{ je rozsah druhého výběru: } X_{1j} \rightarrow N(\mu_2, \sigma_2^2)$$

Výběrová charakteristika	Rozdělení pravděpodobnosti	Poznámka
$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$N(0; 1)$	viz CLV
$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_1^2(n_1-1) + S_2^2(n_2-1)}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$	$t_{n_1+n_2-2}$	viz vlastnosti Studentova rozdělení předpoklad: $\sigma_1^2 = \sigma_2^2$
$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$t_\nu$ $\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^2}{n_1}\right)^2 \frac{1}{n_1+1} + \left(\frac{S_2^2}{n_2}\right)^2 \frac{1}{n_2+1}} - 2$	viz vlastnosti Studentova rozdělení předpoklad: $\sigma_1^2 \neq \sigma_2^2$
$\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}}$	$F_{n_1-1; n_2-1}$	viz vlastnosti Fisher-Snedecorova rozdělení
$\frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}}$	$N(0; 1)$	viz CLV předpoklad: $\left(n_1 > \frac{9}{p_1(1-p_1)}\right) \wedge$ $\left(n_2 > \frac{9}{p_2(1-p_2)}\right)$

## 8.1 Test z teorie

1. Střední hodnota pevně zvolené náhodné veličiny je
  - a) náhodná veličina,
  - b) konstanta,
  - c) náhodný jev,
  - d) výběrová charakteristika.
2. Výběrový průměr je
  - a) náhodná veličina,
  - b) konstanta,
  - c) náhodný jev,
  - d) výběrová charakteristika.
3. S rostoucím rozsahem výběru se obvykle rozptýl průměru
  - a) snižuje,
  - b) zvyšuje,
  - c) nemění.
4. Statistická indukce je
  - a) experiment,
  - b) metoda, která umožňuje odhadnout vlastnosti výběru na základě znalostí vlastností populace,
  - c) zobecnění statistických výsledků získaných zpracováním výběru na celou populaci,
  - d) metoda sběru dat.
5. Zákon velkých čísel v důsledku říká, že při dostatečném rozsahu výběru
  - a) má průměr normální rozdělení,
  - b) má průměr Studentovo rozdělení,
  - c) se střední hodnota přibližuje teoretické hodnotě průměru,
  - d) se relativní četnost přibližuje teoretické hodnotě pravděpodobnosti.
6. Pro modelování průměru výběru dostatečně velkého rozsahu je vhodné použít rozdělení
  - a) normální,
  - b) Pearsonovo ( $\chi^2$ ),
  - c) Studentovo,
  - d) Fisherovo-Snedecorovo.
7. Pro modelování průměru výběru malého rozsahu je vhodné použít rozdělení
  - a) normální,
  - b) Pearsonovo ( $\chi^2$ ), Studentovo,
  - d) Fisherovo-Snedecorovo.

8. Pro modelování relativní četnosti ve výběru o dostatečném rozsahu je vhodné použít rozdělení
- a) normální,
  - b) Pearsonovo ( $\chi^2$ ),
  - c) Studentovo,
  - d) Fisherovo-Snedecorovo.
9. Pro modelování rozptylu výběru z normálního rozdělení je vhodné použít rozdělení
- a) normální,
  - b) Pearsonovo ( $\chi^2$ ),
  - c) Studentovo,
  - d) Fisherovo-Snedecorovo.
10. Pro modelování poměru rozptylů dvou výběrů z normálního rozdělení je vhodné použít rozdělení
- a) normální,
  - b) Pearsonovo ( $\chi^2$ ),
  - c) Studentovo,
  - d) Fisherovo-Snedecorovo.

## 8.2 Příklady

1. Zatížení letadla s 64 místy nemá překročit 6 000 kg. Jaká je pravděpodobnost, že při plném obsazení bude tato hodnota překročena, má-li hmotnost cestujícího střední hodnotu 90 kg a směrodatnou odchylku 10 kg?

Řešení:

[0,001]

2. Zásilka obsahuje 300 výrobků určitého typu. Je známo, že pravděpodobnost zhotovení vadného výrobku tohoto typu je 0,04.
  - a) Odhadněte pravděpodobnost, že absolutní odchylka podílu vadných výrobků v zásilce od pravděpodobnosti vyrobení vadného výrobku bude menší než 1 %.

Řešení:

[0,623]

- b) Jak se změní výsledek, jestliže zásilka bude obsahovat 3 000 výrobků?

Řešení:

[0,994]

3. Cestující pravidelně jezdí do zaměstnání a zpět MHD. Je známo, že doba čekání na příjezd MHD se pohybuje v mezích od 0 do 3 minut. Jaká je pravděpodobnost, že celková doba čekání zaměstnance na příjezd MHD během 23 pracovních dnů bude kratší než 80 minut?

Řešení:

[0,970]

4. Předpokládejme, že průměrná spotřeba elektrické energie domácností v určitém městě v lednu je 120 kWh a směrodatná odchylka spotřeby je 100 kWh. Určete pravděpodobnost, že průměrná spotřeba 100 náhodně vybraných domácností bude větší než 140 kWh.

Řešení:

[0,023]

5. Společnost Acme Battery Company vyvinula nový typ baterie mobilních telefonů. V průměru vydrží baterie 60 minut na jedno nabití. Směrodatná odchylka této doby je 4 minuty. Předpokládejme, že výrobní oddělení po 6 měsících spustí test kontroly kvality. Provedli dva náhodné výběry o rozsahu 10 baterií a v obou zjistili směrodatnou odchylku výdrže baterií větší než 6 minut. S jakou pravděpodobností takový výsledek mohli očekávat?

Řešení:

[0, 000 256]

6. Z úmrtnostních tabulek vyplývá pravděpodobnost 0,99, že se 35 - letý muž dožije dalšího roku. Roční pojistné této věkové skupiny činí 2 000 Kč, v případě úmrtí pojišťovna vyplatí 100 000 Kč. Jaká je pravděpodobnost, že zisk z 500 pojištěných mužů ve věku 35 let bude alespoň 500 000 Kč? (Řešte dvěma způsoby - pomocí binomického rozdělení a pomoci aproximace binomického rozdělení rozdělením normálním.)

a) Řešení:

[0,616]

b) Řešení:

[0,589]

7. Předpokládejme, že v populaci má přibližně 60 % mladých mužů vyšší než doporučenou hladinu cholesterolu v séru. S jakou pravděpodobností bude mít v náhodném výběru 200 mladých mužů více než 120 z nich vyšší než doporučenou hladinu cholesterolu v séru?

a) Řešení:

[0,470]

b) Řešení:

[0,5]

---

## 9. Intervalové odhady

viz applet

### 9.1 Test z teorie

1. Chceme-li najít nejlepší možný odhad směrodatné odchylky vybrané vlastnosti nekonečné populace, měli bychom
  - a) použít co možná největší výběrový soubor,
  - b) použít co možná nejmenší výběrový soubor,
  - c) zjistit hodnotu sledované vlastnosti u všech prvků populace,
  - d) použít výběrový soubor o rozsahu nejvýše 10 000 prvků populace.
2. Chceme-li najít nejlepší možný odhad směrodatné odchylky vybrané vlastnosti populace o rozsahu 50 000 jednotek (prvků), pak by rozsah výběru neměl překročit
  - a) 49 999 jednotek,
  - b) 10 000 jednotek,
  - c) 5 000 jednotek,
  - d) 2 500 jednotek,
  - e) 1 000 jednotek.
3. Doplňte:
  - a) Průměr je (*náhodná veličina, konstanta*).
  - b) Střední hodnota je (*výběrová, populační*) charakteristika.
  - c) Odhadujeme-li populační charakteristiku jedním číslem, hovoříme o (*bodovém, intervalovém*) odhadu.
  - d) Řekneme, že odhad je (*nestranný, vydatný, konzistentní*), jestliže se jeho střední hodnota rovná hledanému parametru.
  - e) Nestranný odhad, jehož rozptyl je (*nejmenší, největší*) mezi rozptyly všech nestranných odhadů příslušného parametru, se nazývá nejlepší nestranný odhad.
  - f) Mějme náhodný výběr. S rostoucí spolehlivostí odhadu  $1 - \alpha$  se obvykle intervalové odhady populačních parametrů (*zuzují, rozšiřují*).
  - g) S rostoucí spolehlivostí odhadu  $1 - \alpha$  (*roste, klesá*) hladina významnosti  $\alpha$ .
  - h) Při dané spolehlivosti odhadu  $1 - \alpha$  se obvykle intervalové odhady populačních parametrů s rostoucím rozsahem výběru (*zuzují, rozšiřují*).
  - i) V technické praxi se obvykle volí spolehlivost odhadu  $1 - \alpha$  rovna (*0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01*).
  - j) V technické praxi se obvykle volí hladina významnosti  $\alpha$  rovna (*0,80; 0,90; 0,95; 0,99; 0,20; 0,10; 0,05; 0,01*).
  - k) Horní mez pravostranného intervalového odhadu je (*stejná, menší než, větší než*) horní mez příslušného oboustranného odhadu.



4. Výběrová charakteristika (náhodná veličina), která nabývá hodnot „blízkých“ hledanému parametru, se nazývá
  - a) bodový odhad hledaného parametru,
  - b) nestranný odhad hledaného parametru,
  - c) konzistentní odhad hledaného parametru,
  - d) vydatný odhad hledaného parametru.
5. Interval, v němž skutečná hodnota hledaného parametru leží s pravděpodobností  $1 - \alpha$  ( $\alpha$  je hladina významnosti), se nazývá
  - a) interval spolehlivosti,
  - b) intervalový odhad.
6. Hladina významnosti  $\alpha$  je pravděpodobnost toho, že skutečná hodnota hledaného parametru
  - a) leží uvnitř intervalu spolehlivosti,
  - b) neleží uvnitř intervalu spolehlivosti.
7. Spolehlivost odhadu  $1 - \alpha$  je pravděpodobnost toho, že skutečná hodnota hledaného parametru
  - a) leží uvnitř intervalu spolehlivosti,
  - b) neleží uvnitř intervalu spolehlivosti.

## 9.2 Příklady

1. Při kontrolních zkouškách 16 žárovek byl stanoven odhad střední hodnoty  $\bar{x}_{16} = 3\,000$  hodin a směrodatné odchylky  $s_{16} = 20$  hodin jejich životnosti. Za předpokladu, že životnost žárovky má normální rozdělení, určete 90% intervalový odhad pro parametry  $\mu$  a  $\sigma$ .

*Odhadované parametry:*

*Ověření předpokladů:*

*Intervalový odhad parametru 1:*

*Intervalový odhad parametru 2:*

$[(2\,991; 3\,009) \text{ h}, (15; 29) \text{ h}]$

2. Hloubka moře se měří přístrojem, jehož systematická chyba je rovna nule a náhodné chyby mají normální rozdělení se směrodatnou odchylkou 20 m. Kolik nezávislých měření je třeba provést, aby s pravděpodobností 95 % stanovila hloubku s chybou menší než 10 m?

*Odhad:*

*Ověření předpokladů:*

*Minimální počet nez. měření nutných pro stanovení střední hloubky moře s chybou menší než 10 m:*

3. Úkolem je určit průměrnou hladinu cholesterolu v séru v určité populaci mužů. V náhodném výběru (pocházejícím z normálního rozdělení) 25 mužů je výběrový průměr 6,3 mmol/l a výběrová směrodatná odchylka 1,3 mmol/l.

*Odhadovaný parametr:*

*Ověření předpokladů:*

*Intervalový odhad parametru:*

[(5,8; 6,8) mmol/l]

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

*Odhadovaný parametr:*

*Ověření předpokladů:*

*Intervalový odhad parametru (Waldův – dle CLV):*

*Intervalový odhad parametru (Clopperův - Pearsonův – Statgraphics):*

[Waldův odhad: (53,2 %; 66,8 %), C-P odhad: (52,8 %; 66,8 %)]

5. V rámci výzkumné studie pracujeme s náhodným výběrem 70 žen z české populace. U každé z žen byl změřen hemoglobin s přesností 0,1 g/100 ml. Naměřené hodnoty jsou v uvedeny v souboru [Hemoglobin.xls](#). Nalezněte 95% intervalové odhady směrodatné odchylky a střední hodnoty hemoglobinu v populaci českých žen. (Normalitu ověřte na základě exploračních grafů.)

*Odhadované parametry:*

*Ověření předpokladů:*

*Intervalový odhad parametru 1:*

*Intervalový odhad parametru 2:*

$[(11, 7; 12, 3) \text{ g/100 ml}, (1, 2; 1, 7) \text{ g/100 ml}]$

6. Jaký musí být počet pozorování, jestliže chceme s pravděpodobností 0,95 stanovit průměrnou hodnotu hemoglobinu u novorozenců s chybou nejvýše 1,0 g/l. Populační rozptyl hodnot se odhaduje hodnotou 46,0  $\text{g}^2/\text{l}^2$ .

*Odhad:*

*Ověření předpokladů:*

*Minimální počet novorozenců, které musíme vyšetřit, abychom stanovili střední hodnotu hemoglobinu u novorozenců s přesností na 1,0 g/l:*

7. V průběhu experimentu sledujeme vliv chlazení (skupina 1 - žádné, skupina 2 - chlazení vodou) okolních struktur na největší rozměr poškození tkáně slinivky břišní [slinivka.xls](#). Kvantifikujte efekt vlivu chlazení a určete jeho 95% intervalový odhad.

*Odhadované parametry:*

*Ověření předpokladů:*

*Intervalový odhad středního poškození tkáně pro skupinu 1:*

*Intervalový odhad středního poškození tkáně pro skupinu 2:*

*Intervalový odhad rozdílu středních poškození tkáně pro skupinu 1 a skupinu 2:*

$[(24, 83; 25, 19), (21, 29; 22, 39), (2, 59; 3, 75)]$

---

# 10. Testování hypotéz - základní terminologie

## 10.1 Test z teorie

Doplňte

- a) Statistická hypotéza je výrok o .....
- b) Rozhodovací proces, který používáme k učinění závěru o rozdělení náhodné veličiny na základě výběrového souboru a hypotéz se nazývá .....
- c) Při testování hypotéz se rozhodujeme mezi ..... a ..... hypotézou.
- d) Obor hodnot testové statistiky (testového kritéria) lze rozdělit na dvě disjunktní množiny nazývané ..... a .....
- e) Kritický obor se stanovuje tak, aby pravděpodobnost, že hodnota testové statistiky padne do kritického oboru byla v případě platnosti nulové hypotézy rovna .....
- f) Pravděpodobnost chyby I. druhu i chyby II. druhu lze snížit, zvýšíme-li .....
- g) Přístup k testování hypotéz, který je založen na rozhodování pomocí kritického oboru bývá nazýván .....
- h) Přístup k testování hypotéz, který je založen na rozhodování pomocí p-hodnoty bývá nazýván .....
- i) Při testování hypotéz je možno učinit dvě rozhodnutí - ..... nebo .....  
.....
- j) Je-li  $p\text{-hodnota}=0,03$ , pak ..... nulovou hypotézu se spolehlivostí 0,95.

# 11. Jednovýběrové testy hypotéz o parametrech populace

Typ proměnné	Požadovaný typ analýzy	Předpoklady	Testy, resp. intervalové odhady
Spojitá proměnná	Ověření variability	Normalita	Test o rozptylu (test o směr. odchylce)
			Intervalový odhad rozptylu (směr. odchylky)
	Ověření polohy	Normalita	Studentův $t$ -test, (test o střední hodnotě)
			Intervalový odhad střední hodnoty
		Výběr většího rozsahu	Znaménkový test (test o mediánu)
		Symetrické rozdělení	Wilcoxonův test (test o mediánu)
Dichotomická proměnná (0-1)	Ověření shody relativní četnosti s očekávanou pravděpodobností	$n > \frac{9}{p(1-p)}$	Test o parametru $\pi$ binomického rozdělení
			Intervalový odhad parametru $\pi$ binomického rozdělení

## 11.1 Test z teorie

- Zamítneme-li na základě  $t$ -testu nulovou hypotézu, pak lze tvrdit, že rozdíl mezi testovanou hodnotou a průměrem výběrového souboru
  - je na dané hladině významnosti statisticky významný,
  - není na dané hladině významnosti statisticky významný.
- Označte všechny parametrické testy, tj. testy vyžadující znalost rozdělení populace
  - test o střední hodnotě ( $t$ -test),
  - test o rozptylu,
  - mediánový test nebo Wilcoxonův test,
  - test o parametru  $\pi$  binomického rozdělení.
- Neparametrické testy
  - nevyžadují splnění žádných předpokladů,
  - nevyžadují znalost rozdělení populace,
  - vyžadují znalost rozdělení populace.
- Neparametrické testy mají
  - větší sílu testu než jejich parametrické protějšky,
  - menší sílu testu než jejich parametrické protějšky.

5. Předpokladem pro použití testu o parametru  $\pi$  binomického rozdělení je
- a) normalita výběru,
  - b) výběr ze spojitého symetrického rozdělení,
  - c) dostatečný rozsah výběru ( $n > 9/(p(1-p))$ ), kde  $p$  je relativní četnost výskytu sledovaného jevu.
6. Předpokladem pro použití Wilcoxonova testu je
- a) normalita výběru,
  - b) výběr ze spojitého symetrického rozdělení,
  - c) dostatečný rozsah výběru ( $n > 9/(p(1-p))$ ), kde  $p$  je relativní četnost výskytu sledovaného jevu.
7. Chceme-li ověřit, zda lze výrobcem udávanou spotřebu 8,8 l/100km považovat za pravdivou (bylo testováno 11 automobilů, normalita výběru byla zamítnuta), použijeme
- a) test o střední hodnotě ( $t$ -test),
  - b) test o rozptylu,
  - c) mediánový test nebo Wilcoxonův test,
  - d) test o parametru  $\pi$  binomického rozdělení.
8. Chceme-li ověřit, zda lze očekávat, že v prodejně je více než 5 % konzerv s prošlou záruční lhůtou (v kontrolním vzorku 100 konzerv bylo nalezeno 7 konzerv s prošlou záruční lhůtou), použijeme
- a) test o střední hodnotě ( $t$ -test),
  - b) test o rozptylu,
  - c) mediánový test nebo Wilcoxonův test,
  - d) test o parametru  $\pi$  binomického rozdělení.
9. Chceme-li ověřit, zda je průměrná výška dospělé populace v ČR větší než 170 cm (rozsah výběru je 120, byla ověřena normalita výběru), použijeme
- a) test o střední hodnotě ( $t$ -test),
  - b) test o rozptylu,
  - c) mediánový test nebo Wilcoxonův test,
  - d) test o parametru  $\pi$  binomického rozdělení.
10. Pro bavlněnou přízi je předepsaná horní mez variability pevnosti vlákna. Rozptyl pevnosti (která má normální rozdělení) nemá překročit 0,36. Chceme-li ověřit, zda je důvod k podezření na vyšší variabilitu než je stanoveno, použijeme
- a) test o střední hodnotě ( $t$ -test),
  - b) test o rozptylu,
  - c) mediánový test nebo Wilcoxonův test,
  - d) test o parametru  $\pi$  binomického rozdělení.



## 11.2 Příklady

1. Máme výběr 216 pacientů a změřili jsme jejich bílkovinné sérum [bilik\\_serum.xlsx](#). Ověřte, zda se průměrné bílkovinné sérum (Albumin) všech pacientů tohoto typu (populační průměr  $\mu$ ) statisticky významně liší od hodnoty 35 g/l.

*Testovaný parametr:*

*Možné testy:*

*Předpoklady testů:*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Pozorovaná hodnota  $x_{OBS}$ :*

*p-hodnota:*

*Rozhodnutí:*

[p-hodnota  $\ll 0,001$ ]

2. V souboru [preziti.xls](#) jsou uvedeny doby přežití pro 100 pacientů s rakovinou plic léčených novým lékem. Z předchozích studií je známo, že průměrné přežití takových pacientů bez podávání nového léku je 22,2 měsíce. Lze na základě těchto dat usoudit, že nový lék prodlužuje přežití?

*Testovaný parametr:*

*Možné testy:*

*Předpoklady testů:*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Pozorovaná hodnota  $x_{OBS}$ :*

*p-hodnota:*

*Rozhodnutí:*

[p-hodnota = 0,758]

3. Automat vyrábí pístové kroužky o daném průměru. Výrobce udává, že směrodatná odchylka průměru kroužku je 0,05 mm. K ověření této informace bylo náhodně vybráno 80 kroužků a vypočtena směrodatná odchylka jejich průměru 0,04 mm. Lze tento rozdíl považovat za statisticky významný ve smyslu zlepšení kvality produkce? Ověřte čistým testem významnosti. Předpokládejte, že průměr pístových kroužků má normální rozdělení.

*Testovaný parametr:*

*Možné testy:*

*Předpoklady testů:*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Pozorovaná hodnota  $x_{OBS}$ :*

*p-hodnota:*

*Rozhodnutí:*

[p-hodnota = 0,006]

4. Automat vyrábí pístové kroužky o daném průměru. Výrobce udává, že směrodatná odchylka průměru kroužku je 0,05 mm. K ověření této informace bylo náhodně vybráno 80 kroužků a byl změřen jejich průměr [krouzky.xls](#). Lze zjištěné výsledky považovat za statisticky významné ve smyslu zlepšení kvality produkce? Ověřte čistým testem významnosti.

*Testovaný parametr:*

*Možné testy:*

*Předpoklady testů:*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Pozorovaná hodnota  $x_{OBS}$ :*

*p-hodnota:*

*Rozhodnutí:*

[p-hodnota << 0,001]

5. Firma TT udává, že 1% jejich rezistorů nesplňuje požadovaná kritéria. V testované dodávce 1000 ks bylo nalezeno 15 nevyhovujících rezistorů. Potvrzuje tento výsledek tvrzení TT? Ověřte čistým testem významnosti.

*Testovaný parametr:*

*Možné testy:*

*Předpoklady testů:*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Pozorovaná hodnota  $x_{OBS}$ :*

*p-hodnota:*

*Rozhodnutí:*

[p-hodnota = 0,08]

## 12. Dvouvýběrové testy o shodě parametrů dvou populací

Typ proměnné	Požadovaný typ analýzy	Předpoklady		Testy, resp. intervalové odhady
Dvě nezávislé spojité proměnné	Ověření shody rozptylů (homoskedasticity)	Normalita		$F$ -test (test shody rozptylů)
				Intervalový odhad <i>poměru</i> rozptylů, resp. směr. odchylek
		—		Leveneho test
	Ověření shody měr polohy (středních hodnot, resp. mediánů)	Normalita	Shoda rozptylů (homoskedasticita)	Dvouvýběrový Studentův $t$ -test (test shody stř. hodnot)
				Intervalový odhad rozdílu stř.hodnot
			Různé rozptyly (heteroskedasticita)	Aspinové-Welchův test (test shody stř. hodnot)
				Intervalový odhad rozdílu stř.hodnot
		—		Mannův-Whitneho test test shody mediánů
Párová (spojitá) data	Ověření shody úrovně párových dat	Normalita		Párový studentův $t$ -test
				Intervalový odhad střední hodnoty rozdílů
		Výběry většího rozsahu		Párový znaménkový test
		Symetrické rozdělení		Wilcoxonův párový test
Dvě dichotomické proměnné	Ověření shody pravděpodobností	$n_i > \frac{9}{p_i(1-p_i)}, i = 1, 2$		Test homogeneity dvou binomických rozdělení
				Intervalový odhad rozdílu parametru binomických rozdělení

## 12.1 Test z teorie

1. Označte všechny neparametrické (robustní) testy.
  - a) dvouvýběrový  $t$ -test,
  - b) párový  $t$ -test,
  - c) Aspinové-Welchův test,
  - d) Mannův-Whitneyho test,
  - e) znaménkový test nebo párový Wilcoxonův test,
  - f) dvouvýběrový  $F$ -test (test o shodě rozptylů),
  - g) test homogenity dvou binomických rozdělení.
2. Předpokladem pro použití Mannova-Whitneyho testu je
  - a) normalita obou výběrů,
  - b) normalita a homoskedasticita obou výběrů ,
  - c) normalita a heteroskedasticita obou výběrů ,
  - d) spojitost a stejný tvar rozdělení obou výběrů,
  - e) dostatečný rozsah obou výběrů ( $n_i > 9/(p_i(1 - p_i))$ ,  $i = 1, 2$ ), kde  $p_i$  je relativní četnost sledovaného jevu v  $i$ -tém výběru.
3. Předpokladem pro použití párového  $t$ -testu je
  - a) normalita obou výběrů,
  - b) normalita a homoskedasticita obou výběrů ,
  - c) normalita a heteroskedasticita obou výběrů ,
  - d) spojitost a stejný tvar rozdělení obou výběrů,
  - e) dostatečný rozsah obou výběrů ( $n_i > 9/(p_i(1 - p_i))$ ,  $i = 1, 2$ ), kde  $p_i$  je relativní četnost sledovaného jevu v  $i$ -tém výběru.
4. Předpokladem pro použití Aspinové-Welchova testu je
  - a) normalita obou výběrů,
  - b) normalita a homoskedasticita obou výběrů,
  - c) normalita a heteroskedasticita obou výběrů,
  - d) spojitost a stejný tvar rozdělení obou výběrů,
  - e) dostatečný rozsah obou výběrů ( $n_i > 9/(p_i(1 - p_i))$ ,  $i = 1, 2$ ), kde  $p_i$  je relativní četnost sledovaného jevu v  $i$ -tém výběru.
5. Neparametrickým protějškem Aspinové-Welchova testu je
  - a) dvouvýběrový  $t$ -test,
  - b) párový  $t$ -test,
  - c) Mannův-Whitneyův test,
  - d) znaménkový test nebo párový Wilcoxonův test,
  - e) dvouvýběrový  $F$ -test (test o shodě rozptylů),
  - f) test homogenity dvou binomických rozdělení.

6. Neparametrickým protějškem párového  $t$ -testu je
- a) dvouvýběrový  $t$ -test,
  - b) Aspinové-Welchův test,
  - c) Mannův-Whitneyho test,
  - d) znaménkový test nebo párový Wilcoxonův test,
  - e) dvouvýběrový  $F$ -test (test o shodě rozptylů),
  - f) test homogenity dvou binomických rozdělení.
7. Neparametrickým protějškem dvouvýběrového  $t$ -testu je
- a) párový  $t$ -test,
  - b) Aspinové-Welchův test,
  - c) Mannův-Whitneyho test,
  - d) znaménkový test nebo párový Wilcoxonův test,
  - e) dvouvýběrový  $F$ -test (test o shodě rozptylů),
  - f) test homogenity dvou binomických rozdělení.
8. Tabáková firma TAB prohlašuje, že jejich cigarety mají nižší obsah nikotinu než cigarety NIK. Obsah nikotinu byl změřen ve 100 cigaretách TAB a 100 cigaretách NIK. Na základě obou výběrů byla ověřena homoskedasticita obsahů nikotinu v cigaretách TAB a NIK. Bylo ověřeno, že obsah nikotinu v cigaretách má normální rozdělení. Chceme-li ověřit, zda lze tvrzení firmy TAB prohlásit za nepravdivé, použijeme
- a) dvouvýběrový  $t$ -test,
  - b) párový  $t$ -test,
  - c) Aspinové-Welchův test,
  - d) Mannův-Whitneyho test,
  - e) znaménkový test nebo párový Wilcoxonův test,
  - f) dvouvýběrový  $F$ -test (test o shodě rozptylů),
  - g) test homogenity dvou binomických rozdělení.
9. Při testování ojetí (mm) pneumatik 11 automobilů určité značky byla zamítnuta normalita ojetí pneumatik (mm). Chceme-li ověřit, zda se pravé a levé přední pneumatiky automobilů této značky ojíždějí srovnatelně, použijeme
- a) dvouvýběrový  $t$ -test,
  - b) párový  $t$ -test,
  - c) Aspinové-Welchův test,
  - d) Mannův-Whitneyho test,
  - e) znaménkový test nebo párový Wilcoxonův test,
  - f) dvouvýběrový  $F$ -test (test o shodě rozptylů),
  - g) test homogenity dvou binomických rozdělení.

10. Bylo ověřeno, že hmotnost balení cukru má normální rozdělení. Testujeme-li, zda seřazením výrobní linky došlo ke snížení kolísavosti hmotnosti balení cukru, použijeme
- a) dvouvýběrový  $t$ -test,
  - b) párový  $t$ -test,
  - c) Aspinové-Welchův test,
  - d) Mannův-Whitneyho test,
  - e) znaménkový test nebo párový Wilcoxonův test,
  - f) dvouvýběrový  $F$ -test (test o shodě rozptylů),
  - g) test homogenity dvou binomických rozdělení.
11. Určete, zda jsou následující tvrzení pravdivá.
- a) Při neparametrickém testu homogenity dvou binomických rozdělení nemusíme ověřovat žádné předpoklady o výběrech.
  - b) Mannův-Whitneyho test se používá pro ověření shody úrovně ve dvou závislých výběrech.
  - c) Každý test hypotézy  $H_0 : \mu_1 = \mu_2$ , tj. hypotézy o shodě dvou středních hodnot, je testem párovým.

## 12.2 Příklady

1. Data v souboru [cholesterol2.xls](#) udávají hladinu cholesterolu v krvi mužů dvou různých věkových skupin (20-30 letých a 40-50 letých). Ověřte na hladině významnosti 0,05 hypotézu, zda se hladina cholesterolu v krvi starších mužů neliší od hladiny cholesterolu v krvi mladších mužů.

*Testované parametry:*

*Možné testy:*

*Předpoklady testů:*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Pozorovaná hodnota  $x_{OBS}$ :*

*p-hodnota:*

*Rozhodnutí:*

[p-hodnota  $\ll 0,001$ ]



2. Údaje v souboru [deprese.xls](#) představují délku remise ve dnech z prostého náhodného výběru ze dvou různých skupin pacientů (pacienti s endogenní depresí a pacienti s neurotickou depresí). Ověřte, zda je pozorovaný rozdíl mezi průměrnou délkou remise u těchto dvou skupin pacientů statisticky významný.

*Testované parametry:*

*Možné testy:*

*Předpoklady testů:*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Pozorovaná hodnota  $x_{OBS}$ :*

*$p$ -hodnota:*

*Rozhodnutí:*

[ $p$ -hodnota = 0,62]

3. Sledujeme osmolalitu moči na lůžkové stanici v 08:00 hodin a v 11:00 hodin u 16 mužů. Na základě výsledků uvedených v souboru [osmolalita.xls](#) ověřte, zda se osmolalita statisticky významně zvýšila.

*Testované parametry:*

*Možné testy:*

*Předpoklady testů:*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Pozorovaná hodnota  $x_{OBS}$ :*

*p-hodnota:*

*Rozhodnutí:*

[p-hodnota  $\ll 0,001$ ]

4. Byly testovány polovodičové součástky dvou výrobců - MM a PP. MM prohlašuje, že její výrobky mají nižší procento vadných kusů. Pro ověření tohoto tvrzení bylo z produkce MM náhodně vybráno 200 součástek, z nichž 14 bylo vadných. Podobný experiment byl proveden u firmy PP s výsledkem 10 vadných ze 100 náhodně vybraných součástek.

- a) Otestujte tvrzení firmy MM čistým testem významnosti.

*Testované parametry:*

*Možné testy:*

*Předpoklady testů:*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Pozorovaná hodnota  $x_{OBS}$ :*

*p-hodnota:*

*Rozhodnutí:*

[p-hodnota = 0,18]

- b) Otestujte tvrzení firmy MM prostřednictvím intervalového odhadu na hladině významnosti 0,05.

*Testované parametry:*

*Předpoklady pro použití intervalového odhadu:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Rozhodnutí:*

[ $P(-0,095 < \pi_{MM} - \pi_{PP} < 0,035) = 0,95$ ]

# 13. Vybrané vícevýběrové testy parametrických hypotéz

Typ proměnné	Požadovaný typ analýzy	Předpoklady		Testy
Alespoň tři nezávislé spojité proměnné	Ověření shody rozptylů (homoskedasticity)	Normalita	Vyvážené třídění	Cochranův test
				Hartleyův test
			Nevyvážené třídění	Bartlettův test
		—		Leveneův test
	Ověření shody měr polohy (středních hodnot, resp. mediánů)	Normalita	Shoda rozptylů (homoskedasticita)	ANOVA (Analýza rozptylu = test shody stř. hodnot (Poznámka: V případě zamítnutí $H_0$ je vhodné provést post hoc analýzu, např. Schéffeho metodou.)
		Symetrické rozdělení		Kruskalův-Wallisův test (test shody mediánů, resp. shody distribucí; použití i pro pořadová data. Poznámka: V případě zamítnutí $H_0$ je vhodné provést post hoc analýzu, např. Dunnové metodou.)
Alespoň tři závislé (spojité) proměnné	Ověření shody úrovně závislých dat	—		Friedmanův test (Poznámka: V případě zamítnutí $H_0$ je vhodné provést post hoc analýzu Friedmanovou metodou.)

Tabulka ANOVA					
Zdroj variability	Součet čtverců	Počet stupňů volnosti	Rozptyl (prům. součet čtverců)	$F$ -poměr	$p$ -hodnota
Skupinový faktor	$SS_B = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2$	$df_B = k - 1$	$MS_B = \frac{SS_B}{df_B}$	$\frac{MS_B}{MS_e}$	$1 - F_0(x_{OBS})$
Reziduální faktor	$SS_e = \sum_{i=1}^k (n_i - 1) s_i^2$	$df_e = n - k$	$MS_e = \frac{SS_e}{df_e}$	—	—
Celkový faktor	$SS_T = \sum_{i=1}^k (X_{ij} - \bar{X})^2$	$df_T = n - k1$	—	—	—

## 13.1 Test z teorie

1. Určete, zda jsou následující tvrzení pravdivá.
  - a) Analýza rozptylu (ANOVA) je test shody rozptylů ve více než dvou výběrech.
  - b) Jedním z předpokladů analýzy rozptylu je alespoň přibližná shoda rozptylů v jednotlivých skupinách.
  - c) Reziduální rozptyl (v analýze rozptylu) lze určit jako aritmetický průměr rozptylů v jednotlivých skupinách.
  - d) Post hoc analýza znamená, že stanovíme nejprve hypotézy  $H_0$ ,  $H_A$ , a „následně“ provedeme řešení.
  - e) Kruskalův-Wallisův test se nazývá rovněž neparametrická ANOVA.
  - f) Hartleyův test homoskedasticity lze použít pouze v případě vyváženého třídění.
  - g) Jediným předpokladem Leveneova testu je nezávislost výběrů.
  - h) Bartlettův test je neparametrickým protějškem Leveneova testu.
  - i) Friedmanův test je neparametrickou obdobou Kruskalova-Wallisova testu.

## 13.2 Příklady

1. Testujeme nulovou hypotézu  $\mu_1 = \mu_2 = \mu_3$ . Bylo zjištěno, že data, která máme k dispozici jsou výběry z normálního rozdělení splňující předpoklad homoskedasticity (shody rozptylů). Na základě údajů získaných explorační analýzou doplňte tabulku ANOVA a vyplývající závěry.

Faktor	Rozsah výběru	Průměr	Výběrová směrodatná odchylka
Skupina 1	40	300	33
Skupina 2	40	290	34
Skupina 3	42	310	31
Celkem	122		—

Zdroj variability	Součet čtverců	Počet stupňů volnosti	Rozptyl (prům. součet čtverců)	$F$ -poměr	$p$ -hodnota
Skupinový					
Reziduální					
Celkový					

Řešení:

Nulová hypotéza  $H_0$ :

Alternativní hypotéza  $H_A$ :

$p$ -hodnota:

Rozhodnutí:

Odhady skupinových efektů (tj. jak se průměry jednotlivých skupin liší od celkového průměru):

[ $p$ -hodnota = 0,02]

2. 122 pacientů, kteří podstoupili operaci srdce, bylo náhodně rozděleno do tří skupin.

**Skupina 1:**

Pacienti dostali 50 % oxidu dusného a 50 % kyslíkové směsi nepřetržitě po dobu 24 hodin.

**Skupina 2:**

Pacienti dostali 50 % oxidu dusného a 50 % kyslíkové směsi pouze během operace.

**Skupina 3:**

Pacienti nedostali žádný oxid dusný, ale dostali 35-50 % kyslíku po dobu 24 hodin.

Data v souboru [kyselina listova.xls](#) odpovídají koncentracím soli kyseliny listové v červených krvinkách ve všech třech skupinách po uplynutí 24 hodin ventilace. Ověřte, zda pozorované rozdíly mezi koncentracemi soli kyseliny listové jsou statisticky významné, tj. zda existuje vliv složení směsi na sledovaný parametr.

*Možné testy:*

*Předpoklady testů:*

*Vizuální posouzení:*

*Exaktní posouzení (testování hypotéz):*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*p-hodnota:*

*Rozhodnutí:*

*Odhady skupinových efektů (tj. jak se průměry jednotlivých skupin liší od celkového průměru):*

*Post-hoc analýza (vícenásobné porovnávání):*

[p-hodnota  $\ll 0,001$ ]

3. Na farmě jsou chována tři plemena králíků. Byl proveden pokus [kralici.xls](#), jehož cílem bylo zjistit, zda i když chováme a vykrmujeme všechny králíky po stejnou dobu a za stejných podmínek, existuje statisticky významný (průkazný) rozdíl mezi plemeny v hmotnostech králíků. Ověřte.

*Možné testy:*

*Předpoklady testů:*

*Vizuální posouzení:*

*Exaktní posouzení (testování hypotéz):*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*p-hodnota:*

*Rozhodnutí:*

*Odhady skupinových efektů (tj. jak se průměry jednotlivých skupin liší od celkového průměru):*

*Post-hoc analýza (vícenásobné porovnávání):*

[p-hodnota  $\ll 0,001$ ]



4. Soutěž o nejlepší jakost výrobků obeslali čtyři výrobci A, B, C, D celkem 66 výrobky. Porota sestavila pořadí (uvedeno pouze pořadí výrobku od nejlepšího k nejhoršímu), jež je uvedené v souboru [jakost.xls](#). Na základě uvedených údajů posudte, zda původ výrobků má vliv na jeho jakost.

*Možné testy:*

*Předpoklady testů:*

*Vizuální posouzení:*

*Exaktní posouzení (testování hypotéz):*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*p-hodnota:*

*Rozhodnutí:* 1cm

*Odhady skupinových efektů (tj. jak se mediány jednotlivých skupin liší od celkového mediánu):*

*Post-hoc analýza (vícenásobné porovnávání):*

5. Byl sledován vliv tří preparátů na srážlivost krve. Kromě jiných ukazatelů byl zjišťován tzv. trombinový čas. Údaje o 42 sledovaných osobách jsou zaznamenány v souboru [trombin.xls](#). Závisí velikost trombinového času na tom, jaký byl použit preparát?

*Možné testy:*

*Předpoklady testů:*

*Vizuální posouzení:*

*Exaktní posouzení (testování hypotéz):*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*p-hodnota:*

*Rozhodnutí:*

*Odhady skupinových efektů (tj. jak se mediány jednotlivých skupin liší od celkového mediánu):*

*Post-hoc analýza (vícenásobné porovnávání):*

[p-hodnota  $\ll 0,001$ ]

6. Byl sledován vliv tří preparátů na srážlivost krve. Kromě jiných ukazatelů byl zjišťován tzv. trombinový čas. U každé osoby byl stanoven nejprve kontrolní údaj (K), který udává trombinový čas před zahájením pokusu. Pak byly aplikovány preparáty A, B, C, a to každý dostatečně dlouho po odeznění účinku těch předchozích. Údaje o 15 sledovaných osobách jsou uvedeny v souboru [trombin2.xls](#). Závisí velikost trombinového času na tom, jaký byl použit preparát?

*Možné testy:*

*Předpoklady testů:*

*Vizuální posouzení:*

*Exaktní posouzení (testování hypotéz):*

*Zvolený test:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*p-hodnota:*

*Rozhodnutí:*

*Odhady skupinových efektů (tj. jak se průměry jednotlivých skupin liší od celkového průměru):*

*Post-hoc analýza (vícenásobné porovnávání):*

[p-hodnota  $\ll 0,001$ ]

# 14. Vybrané testy neparametrických hypotéz

## Testy dobré shody

Název testu	Předpoklady testu	Testová statistika
$\chi^2$ test dobré shody	Očekávané četnosti $\geq 2$ , alespoň 80% očekávaných četností $> 5$	$G = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$

## Analýza závislosti v kontingenční tabulce

Název testu	Předpoklady testu	Testová statistika
Analýza závislosti v kontingenční tabulce	Očekávané četnosti $\geq 2$ , alespoň 80% očekávaných četností $> 5$	$K = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

- Koeficient korelace  $CC = \sqrt{\frac{K}{K+n}}$  (pro čtvercové kontingenční tabulky)
- Korigovaný koeficient kontingence  $CC_{cor} = \frac{CC}{CC_{max}}$ , kde  $CC_{max} = \sqrt{\frac{\min(r;s)-1}{\min(r;s)}}$  (pro obdélníkové kontingenční tabulky)
- Cramerův koeficient  $V = \sqrt{\frac{K}{n(\min(r;s) - 1)}}$ .

Tyto koeficienty se mohou vyskytovat v intervalu (0;1). Čím blíže jsou 1, tím je závislost mezi  $X$  a  $Y$  těsnější.

## Analýza závislosti v asociační tabulce

- Odhad poměru šancí:  $\widehat{OR} = \frac{ad}{bc}$
- Intervalový odhad  $OR$ :  $\left\langle \widehat{OR} \cdot e^{-\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \cdot z_{1-\frac{\alpha}{2}}}; \widehat{OR} \cdot e^{\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \cdot z_{1-\frac{\alpha}{2}}} \right\rangle$
- Odhad relativního rizika:  $\widehat{RR} = \frac{a(c+d)}{c(a+b)}$
- Intervalový odhad  $RR$ :  $\left\langle \widehat{RR} \cdot e^{-\sqrt{\frac{b}{a(a+b)} + \frac{d}{c(c+d)}} \cdot z_{1-\frac{\alpha}{2}}}; \widehat{RR} \cdot e^{\sqrt{\frac{b}{a(a+b)} + \frac{d}{c(c+d)}} \cdot z_{1-\frac{\alpha}{2}}} \right\rangle$

## 14.1 Test z teorie

1. Lze Kolmogorovův-Smirnovův test použít pro testování normality? (*ANO*, *NE*)
2. Použijeme-li  $\chi^2$  test dobré shody pro ověření toho, zda je klasická šestistěnná hrací kostka „férová“, pak má v případě platnosti nulové hypotézy testová statistika  $\chi^2$  rozdělení se
  - a) 4 stupni volnosti,
  - b) 5 stupni volnosti,
  - c) 6 stupni volnosti.
3. Jak postupujeme v případě, kdy při použití  $\chi^2$  testu dobré shody vyšlo po rozdělení dat do dvaceti tříd 7 očekávaných třídních četností rovných 1?
  - a) V tomto případě nelze  $\chi^2$  testem dobré shody rozhodnout.
  - b) Prohlásíme předpoklady testu za splněné.
  - c) Sloučíme příslušné sousední třídy.
4. Empirická distribuční funkce je funkce
  - a) diskrétní,
  - b) spojitá,
  - c) zleva spojitá,
  - d) zprava spojitá.
5. Čím je mozaikový graf členitější, tím je pozorovaná závislost mezi veličinami v kontingenční tabulce
  - a) slabší,
  - b) silnější.
6. Analyzujeme-li závislost v kontingenční tabulce, která má 4 řádky a 5 sloupců, pak  $\chi^2$  test nezávislosti můžeme použít, pokud alespoň
  - a) 4,
  - b) 10,
  - c) 16,
  - d) 20,

očekávaných četností je větších než 5 a ostatní jsou rovny alespoň

  - a) 0,
  - b) 1,
  - c) 2.
7. Cramerovo  $V(K)$ 
  - a) nabývá hodnot z intervalu  $(-\infty; \infty)$ ,
  - b) nabývá hodnot z intervalu  $(0; \infty)$ ,
  - c) nabývá hodnot z intervalu  $\langle 0; 1 \rangle$ .
8. (*Kontingenční*, *Asociační*) tabulka je speciálním případem (*kontingenční*, *asociační*) tabulky.

9. Je-li odhad relativního rizika  $\widehat{RR} = 1,2$ , pak
- a) mezi znaky v asociační tabulce existuje závislost,
  - b) mezi znaky v asociační tabulce neexistuje závislost,
  - c) o závislosti znaků v asociační tabulce musí rozhodnout test.
10. Je-li odhad poměru šancí  $\widehat{OR} = 10,2$ , pak je v daném výběru u exponované populace
- a) nižší šance výskytu nemoci než u neexponované populace,
  - b) vyšší šance výskytu nemoci než u neexponované populace,
  - c) stejná šance výskytu nemoci jako u neexponované populace.
11. Odhadujeme-li se spolehlivostí 0,95, že relativní riziko  $RR \in (0,87; 1,45)$ , pak na hladině významnosti 0,05 (tj. se spolehlivostí 0,95)
- a) nezamítáme hypotézu o nezávislosti znaků  $X$  a  $Y$ ,
  - b) zamítáme hypotézu o nezávislosti znaků  $X$  a  $Y$ ,
  - c) nelze o nezávislosti znaků  $X$  a  $Y$  rozhodnout.

## 14.2 Příklady

1. Hodilo se 6 000 krát hrací kostkou a zaznamenaly se počty padlých ok.

$x_i$ (číslo které padlo)	1	2	3	4	5	6
$n_i$ (četnost jeho výskytu)	979	1 002	1 015	980	1 040	984

Je možné na základě příslušného testu na hladině významnosti 0,05 spolehlivě tvrdit, že kostka je „falešná“, tj. že pravděpodobnosti všech čísel na kostce nejsou stejné?

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Zvolený test:*

*Předpoklady testu:*

$x_i$ (číslo které padlo)	1	2	3	4	5	6	Celkem
Pozorované četnosti	979	1 002	1 015	980	1 040	984	
Očekávané pravděpodobnosti	$\frac{1}{6}$						
Očekávané četnosti							—

*Počet stupňů volnosti:*

$x_{OBS}$ :

$p$ -hodnota:

*Rozhodnutí:*

[ $p$ -hodnota = 0,711]

2. Výrobní firma odhaduje počet poruch určitého zařízení během 100 hodin pomocí Poissonova rozdělení s parametrem 1,2. Zaměstnanci zaznamenali pro kontrolu skutečné počty poruch celkem ve 150 100hodinových intervalech (výsledky jsou uvedeny v tabulce). Ověřte čistým testem významnosti, zda má počet poruch daného zařízení během 100 hodin skutečně Poissonovo rozdělení s parametrem  $\lambda t = 1, 2$ .

$x_i$ (počet poruch během 100 hodin provozu)	0	1	2	3	4
$n_i$ (počet pozorování)	52	48	36	10	4

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Zvolený test:*

*Předpoklady testu:*

$x_i$ (počet poruch během 100 hodin provozu)	0	1	2	3	4	Celkem
$n_i$ (počet pozorování)	52	48	36	10	4	
Očekávané pravděpodobnosti						
Očekávané četnosti						—

*Počet stupňů volnosti:*

$x_{OBS}$ :

$p$ -hodnota:

*Rozhodnutí:*

[ $p$ -hodnota = 0,590]

3. Zaměstnanci zaznamenali pro kontrolu počty poruch celkem ve 150 100hodinových intervalech (výsledky jsou uvedeny v tabulce). Ověřte čistým testem významnosti, zda má počet poruch daného zařízení během 100 hodin skutečně Poissonovo rozdělení.

$x_i$ (počet poruch během 100 hodin provozu)	0	1	2	3	4
$n_i$ (počet pozorování)	52	48	36	10	4

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Odhad parametru*

*Poissonova rozdělení:*

*Zvolený test:*

*Předpoklady testu:*



$x_i$ (počet poruch během 100 hodin provozu)	0	1	2	3	4	Celkem
$n_i$ (počet pozorování)	52	48	36	10	4	
Očekávané pravděpodobnosti						
Očekávané četnosti						—

*Počet stupňů volnosti:*

$x_{OBS}$ :

*p-hodnota:*

*Rozhodnutí:*

[ $p$ -hodnota = 0,480]

4. Na dálnici byly v průběhu několika minut měřeny časové odstupy (s) mezi průjezdy jednotlivých vozidel. Zjištěné hodnoty těchto odstupů jsou zaznamenány v souboru [dalnice.xlsx](#).

(viz prezentace [přednášky](#))

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*Odhad parametrů  
normálního rozdělení:*

*Zvolený test:*

*Počet stupňů volnosti:*

*p-hodnota:*

*Rozhodnutí:*

[ $p$ -hodnota  $\ll 0,001$ ]

5. Rozhodněte na základě datového souboru [experimentovani-s-telem.xls](#) (Dudová, J. - Experimentování s tělem (výsledky průzkumu), 2013. Dostupné online na <http://experimentovani-s-telem.vyplnto.cz>), zda existuje souvislost mezi pohlavím respondentů a tím, zda mají tetování. Pro posouzení míry kontingence použijte Cramerovo  $V$ .

*Populace:*

*Předpoklad pro použití metod statistické indukce:*

*Vizuální posouzení:*

*Míra kontingence:*

*Zvolený test:*

*Předpoklady testu:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*$p$ -hodnota:*

*Rozhodnutí:*

[ $p$ -hodnota = 0,003]

6. Pro diferencovaný přístup v personální politice potřebuje vedení podniku vědět, zda spokojenost v práci závisí na tom, jedná-li se o pražský závod či závody mimopražské. Výsledky šetření jsou v následující tabulce. Zobraďte data pomocí mozaikového grafu a na základě testu nezávislosti v kombinační tabulce rozhodněte o závislosti spokojenosti v zaměstnání na umístění podniku. Pro posouzení míry kontingence použijte Cramerovo  $V$ .

Stupeň spokojenosti	Místo	
	Praha	Venkov
Velmi spokojen	15	40
Spíše spokojen	50	130
Spíše nespokojen	25	10
Velmi nespokojen	10	20

*Vizuální posouzení:*

*Míra kontingence:*

*Zvolený test:*

*Předpoklady testu:*

*Nulová hypotéza  $H_0$ :*

*Alternativní hypotéza  $H_A$ :*

*p-hodnota:*

*Rozhodnutí:*

[p-hodnota  $\ll 0,001$ ]

7. V letech 1965 až 1968 bylo v kohortové studii kardiovaskulárních onemocnění v rámci „Honolulu Heart Program“ zahájeno sledování 8 006 mužů, z nichž 7 872 nemělo při zahájení studie v anamnéze mrtvici (apoplexii). Z tohoto počtu bylo 3 435 kuřáků a 4 437 nekuřáků. Při jejich sledování po dobu 12 let dostalo mrtvici 171 mužů ve skupině kuřáků a 117 mužů ve skupině nekuřáků. (Zdroj: Malý, M., Zvárová, M., *Statistické metody v epidemiologii*, Praha, 2003, ISBN: 8 024 607 654)

- a) Zapište zjištěné výsledky do asociační tabulky.

Řešení:

- b) Na základě vizuálního posouzení odhadněte vliv kouření na výskyt kardiovaskulárních onemocnění.

Řešení:

- c) Určete absolutní riziko vzniku kardiovaskulárních onemocnění u kuřáků a nekuřáků.

Řešení:

[0,050; 0,026]

- d) Určete relativní riziko (včetně 95% intervalového odhadu) vzniku kardiovaskulárních onemocnění u kuřáků a nekuřáků. Vysvětlete praktický význam zjištěných výsledků.

Řešení:

$$[\widehat{RR} = 1,89; P(RR \in \langle 1,50; 2,38 \rangle) = 0,95]$$

- e) Určete absolutní šance vzniku kardiovaskulárních onemocnění u kuřáků a nekuřáků.

Řešení:

$$[\text{kuřáci: } 0,052; \text{nekuřáci: } 0,027]$$

- f) Určete relativní šanci (včetně 95% intervalového odhadu) vzniku kardiovaskulárních onemocnění u kuřáků a nekuřáků. Vysvětlete praktický význam zjištěných výsledků.

Řešení:

$$[\widehat{OR} = 1,93; P(OR \in \langle 1,52; 2,46 \rangle) = 0,95]$$

- g) Rozhodněte na hladině významnosti 0,05 o závislosti výskytu kardiovaskulárních chorob na kouření.

Řešení:

# 15. Korelační a regresní analýza

## Analýza závislosti v normálním rozdělení

Pearsonův (výběrový) korelační koeficient:  $r = \frac{s_{XY}}{s_X \cdot s_Y}$ , kde  $s_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$ ,  $s_X(s_Y)$  je výběrová směrodatná odchylka proměnné  $X$  ( $Y$ ).

Název testu	Testované parametry	Předpoklady testu	Testová statistika $T(X, Y)$	Nulové rozdělení
Test nulovosti korelačního koeficientu	$\rho$	normalita	$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$	$t_{n-2}$

## Analýza závislosti ordinálních veličin

Spearmanův korelační koeficient:  $r_s = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^n (R_{X_i} - R_{Y_i})^2$

Název testu	Testované parametry	Předpoklady testu	Testová statistika $T(X, Y)$	Kritický obor
Test nulovosti korelačního koeficientu	$\rho$	—	$T =  r_s $	$W = \{T : T \geq r_s^*(\alpha)\}$ (Tabuka 15)

## Doporučený postup při korelační a regresní analýze

1. Explorační analýza korelačního pole (případný odhad typu regresní funkce, identifikace vlivných bodů)
2. Odhad koeficientů regresní funkce (aplikace vyrovnávacího kritéria - např. metody nejmenších čtverců)
3. Verifikace modelu, tj. ověření předpokladů lineárního modelu
  - a) Celkový  $F$ -test - testujeme, zda hodnota vysvětlované proměnné závisí na lineární kombinaci vysvětlujících proměnných, tj. testujeme nulovou hypotézu  $H_0 : \beta_1 = \dots = \beta_k$  vůči alternativě  $H_A : \overline{H_0}$ . Pokud bychom nulovou hypotézu nezamítli, znamenalo by to, že model je chybně specifikován.
  - b) Dílčí  $t$ -testy - umožňují testovat oprávněnost setrvání vysvětlující proměnné v regresním modelu. Testujeme (postupně pro jednotlivá  $i$ ) nulovou hypotézu ve tvaru  $H_0 : \beta_i = 0$  vůči alternativě  $H_A : \beta_i \neq 0$  pro  $i = 0, 1, \dots, k$ . Pokud pro konkrétní  $i$  nelze zamítnout nulovou hypotézu, je třeba zvážit setrvání příslušné vysvětlující proměnné v modelu.
  - c) Analýza reziduí - ověřujeme předpoklady pro použití lineárního regresního modelu.
    - ověření normality reziduí - testy dobré shody,
    - ověření nulovosti střední hodnoty - vizuálně na základě grafu reziduí a odhadovaných hodnot závisle proměnné (rezidua musí kolísat kolem nuly) + dvouvýběrový  $t$  test,
    - ověření homoskedasticity - vizuálně na základě grafu reziduí a odhadovaných hodnot závisle proměnné (rezidua se systematicky nezvyšují ani se systematicky nesnižují spolu s rostoucími odhadovanými hodnotami),

- ověření autokorelace reziduí - vizuálně na základě grafu reziduí a odhadovaných hodnot závisle proměnné (autokorelace projeví tak, že se rezidua systematicky snižují nebo zvyšují, resp. můžeme mezi rezidui a předpovídanými hodnotami pozorovat nelineární závislost) + Durbinova-Watsonova statistika.
- d) Multikolinearita - v případě vícenásobné regrese musíme ověřit, zda neexistuje multikolinearita mezi regresory.
- e) Ověření kvality modelu - index determinace  $R^2$  (udává kolik procent vysvětlované proměnné bylo vysvětleno modelem), *koeficient korelace  $r$*  (míra korelace mezi závisle proměnnou a regresorem v případě přímkové regrese), *koeficient vícenásobné korelace  $r_{Y \cdot x_1, x_2, \dots, x_k}$*  (míra korelace mezi závisle proměnnou na lineární kombinaci regresorů  $x_1, x_2, \dots, x_k$ ), *koeficienty parciální korelace*, např.  $r_{Y, x_1 \cdot x_2, \dots, x_k}$  (míra korelace mezi závisle proměnnou a jedním z regresorů při vyloučení vlivu ostatních regresorů).
4. Využití verifikovaného modelu k predikci - odhad střední hodnoty závisle proměnné při daných hodnotách regresorů (pás spolehlivosti), odhad individuální hodnoty závisle proměnné při daných hodnotách regresorů (pás predikce). Pozor na extrapolaci!

## 15.1 Test z teorie

1. Kvalita 50 různých výukových materiálů byla dvěma odborníky hodnocena na stupnici od 1 do 5. Vhodnou mírou závislosti mezi hodnocením jednotlivých odborníků je
  - a) Spearmanův korelační koeficient,
  - b) Pearsonův korelační koeficient,
  - c) korigovaný koeficient kontingence,
  - d) Cramerovo  $V$ .
2. Hodnoty Pearsonova korelačního koeficientu blízké nule vypovídají o tom, že
  - a) sledované veličiny  $X$  resp.  $Y$  nenesou prakticky žádnou informaci o  $Y$  resp.  $X$ ,
  - b) mezi sledovanými veličinami  $X$  a  $Y$  existuje silná lineární závislost,
  - c) mezi sledovanými veličinami  $X$  a  $Y$  neexistuje silná lineární závislost,
  - d) sledované veličiny  $X$  a  $Y$  jsou nezávislé.
3. Hodnoty Pearsonova korelačního koeficientu blízké -1 vypovídají o tom, že
  - a) sledované veličiny  $X$  resp.  $Y$  nenesou prakticky žádnou informaci o  $Y$  resp.  $X$ ,
  - b) na měřených objektech jsou nízké hodnoty veličiny  $X$  doprovázeny spíše vysokými hodnotami veličiny  $Y$ ,
  - c) na měřených objektech jsou nízké hodnoty veličiny  $X$  doprovázeny spíše nízkými hodnotami veličiny  $Y$ .
4. Regresní a korelační analýza umožňuje získat informace o
  - a) tvaru a síle závislosti mezi kvalitativními proměnnými,
  - b) tvaru a síle závislosti mezi kvantitativními proměnnými,
  - c) tvaru a síle závislosti mezi kvantitativními proměnnými, mezi nimiž je lineární vztah.
5. V případě, že jsou splněny předpoklady lineárního regresního modelu, pak metoda nejmenších čtverců umožňuje nalézt
  - a) přesný funkční předpis hledané regresní funkce,
  - b) index determinace,
  - c) nejlepší odhad koeficientů hledané regresní funkce.
6. Lze metodami lineární regrese nalézt regresní funkci ve tvaru mocninné funkce  $f : y = ax^b$ , kde  $a, b \in \mathbb{R} \setminus \{0\}$ ?
  - a) Ano, tato funkce je lineární vzhledem k parametrům.
  - b) Ano, tuto funkci můžeme linearizovat logaritmováním funkčního předpisu.
  - c) Ne, tuto funkci nelze použít k vyjádření regresní funkce.
  - d) Ne, toto lze řešit metodami nelineární regrese.
7. Lze metodami lineární regrese nalézt regresní funkci ve tvaru mocninné funkce:  $y = a_0 + a_1x_1 + \dots + a_kx_k$ , kde  $a_i \in \mathbb{R}, i = 0, 1, \dots, k$ ?
  - a) Ano, tato funkce je lineární vzhledem k parametrům.
  - b) Ano, tuto funkci můžeme linearizovat logaritmováním funkčního předpisu.
  - c) Ne, tuto funkci nelze použít k vyjádření regresní funkce.
  - d) Ne, toto lze řešit metodami nelineární regrese.

8. Lze metodami lineární regrese nalézt regresní funkci ve tvaru mocninné funkce  $f : y = a_0 + a_1 e^{a_2 x^2}$ , kde  $a_i \in \mathbb{R}, i = 0, 1, 2$ ?
- a) Ano, tato funkce je lineární vzhledem k parametrům.
  - b) Ano, tuto funkci můžeme linearizovat logaritmováním funkčního předpisu.
  - c) Ne, tuto funkci nelze použít k vyjádření regresní funkce.
  - d) Ne, toto lze řešit metodami nelineární regrese.
9. Koeficienty regresní funkce jsou
- a) konstanty,
  - b) náhodné veličiny.
10. Index determinace může nabývat hodnot z intervalu
- a)  $\langle -1; 1 \rangle$ ,
  - b)  $\langle 0; 1 \rangle$ ,
  - c)  $\langle 0; \infty \rangle$ .
11. Rezidua jsou odchylky
- a) pozorovaných a odhadovaných hodnot závislé proměnné,
  - b) pozorovaných a odhadovaných hodnot nezávislé proměnné,
  - c) pozorovaných a odhadovaných regresních funkcí.
12. S rostoucím rozptylem reziduí se odhad rozptylu odhadů regresních koeficientů
- a) zvyšuje,
  - b) snižuje.
13. S rostoucím rozptylem jednotlivých regresorů se rozptyl odhadů regresních koeficientů
- a) zvyšuje,
  - b) snižuje.
14. K ověření, zda hodnota vysvětlované proměnné závisí na lineární kombinaci všech vysvětlujících proměnných, používáme
- a) celkový  $F$ -test,
  - b) dílčí  $t$ -testy,
  - c) analýzu reziduí,
  - d) index determinace.
15. K testování oprávněnosti setrvání jednotlivých vysvětlujících proměnných v regresním modelu používáme
- a) celkový  $F$ -test,
  - b) dílčí  $t$ -testy,
  - c) analýzu reziduí,
  - d) index determinace.



16. Rezidua považujeme za nekorelované, pokud Durbin-Watsonova statistika leží v intervalu
- a)  $\langle -1; 1 \rangle$ ,
  - b)  $\langle 0; 1 \rangle$ ,
  - c)  $\langle 0; \infty \rangle$ ,
  - d)  $\langle 1, 4; 2, 6 \rangle$ ,
  - e)  $\langle 1, 4; 2, 6 \rangle$  nebo  $\langle 2, 6; \infty \rangle$ .
17. Pojmem multikolinearita označujeme
- a) lineární závislost mezi vysvětlovanou proměnnou a jednotlivými vysvětlujícími proměnnými,
  - b) lineární závislost mezi vysvětlujícími proměnnými,
  - c) lineární závislost mezi vysvětlovanými proměnnými,
  - d) lineární závislost mezi jednotlivými regresními funkcemi.
18. Pás spolehlivosti (odhad střední hodnoty závisle proměnné při daných hodnotách regresorů) je
- a) stejně široký jako,
  - b) širší než,
  - c) užší než,
  - d) pás predikce.
19. Odhad závisle proměnné pro hodnoty regresorů ležící mimo interval pozorovaných hodnot označujeme jako
- a) interpolaci,
  - b) extrapolaci,
  - c) korelaci.
- pás predikce.
20. Doplňte:
- a) K ověření, zda hodnota vysvětlované proměnné závisí na lineární kombinaci vysvětlujících proměnných, používáme .....
  - b) K testování oprávněnosti setrvání jednotlivých vysvětlujících proměnných v regresním modelu používáme .....
  - c) Při analýze reziduí ověřujeme ....., ....., ....., .....
  - d) Pojmem multikolinearita označujeme .....

## 15.2 Příklady

1. Máme k dispozici výsledky prvního a druhého zápočtového testu deseti studentů. Na hladině významnosti 0,05 testujte hypotézu, že výsledky zápočtových testů jsou kladně korelované.

$X_i$ (1.test)	7	8	10	4	14	9	6	2	13	5
$Y_i$ (2.test)	9	7	12	6	15	6	8	4	11	8

*Vizuální posouzení závislosti:*

*Korelační koeficienty pro spojitá data:*

*Ověření normality:*

*Zvolený korelační koeficient:*

*Bodový odhad korelačního koeficientu:*

*Test, zda lze korelační koeficient považovat za kladný:*

[ $p$ -hodnota = 0,021]

2. V níže uvedené tabulce je zaznamenána spotřeba alkoholu a úmrtnost na cirhózu jater a alkoholismus ve vybraných zemích. Určete, zda úmrtnost na cirhózu jater a alkoholismus závisí na spotřebě alkoholu. (Zadání příkladu bylo převzato z [1]).

země	spotřeba alkoholu [l/osoba]	úmrtnost na cirhózu jater a alkoholismus [počet zemřelých na 100 000 obyvatel]
Finsko	3,9	3,6
Norsko	4,2	4,3
Irsko	5,6	3,4
Holandsko	5,7	3,7
Švédsko	6,0	7,2
Anglie	7,2	3,0
Belgie	10,8	12,3
Rakousko	10,9	7,0
SRN	12,3	23,7
Itálie	15,7	23,6
Francie	24,7	46,1

*Vizuální posouzení závislosti:*

*Korelační koeficienty pro spojitá data:*

*Ověření normality:*

*Zvolený korelační koeficient:*

*Bodový odhad korelačního koeficientu:*

*Test, zda lze korelační koeficient považovat za kladný:*

[ $p$ -hodnota = 0,015]

3. Procentuální obsah křemíku v surovém železe  $K$  závisí na teplotě strusky  $T$  ([kremik.xls](#)). Navrhnete pro tuto závislost regresní model, verifikujte jej a posuďte jeho kvalitu. Na základě nalezeného modelu odhadněte (včetně příslušného intervalu spolehlivosti) průměrný procentuální obsah křemíku v surovém železe, bude-li při jeho výrobě používaná struska o teplotě 1350 °C. (*Součástí řešení příkladu musí být závěr obsahující: nalezený model, hodnocení jeho věrohodnosti a kvality (slovně) + konkrétní vysvětlení výsledku predikce, včetně komentáře k její věrohodnosti!!!.*)

a) *Vizuální posouzení závislosti:*

b) *Primárně zvolený regresní model:*

c) *Ověření normality závislé proměnné i regresoru:*

d) *Dílčí  $t$ -testy:*

e) *Optimalizovaný regresní model:*

f) *Analýza reziduí:*

*Ověření normality:*

*Ověření nulové střední hodnoty:*

*Ověření nulové autokorelace:*

*Ověření homoskedasticity:*

g) *Index determinace:*

h) *Závěr:*

i) *Predikce:*

*Bodový odhad:*

*Intervalový odhad:*

*Komentář k věrohodnosti predikce:*

4. Proč nestačí při regresní analýze pouze najít odhad regresní funkce a index determinace? Srovnajte výsledky regresní analýzy pro tzv. Anscombův kvartet - [Anscombe.xls](#). Které předpoklady pro použití lineárního regresního modelu jsou porušeny u jednotlivých sad dat?

*Sada 1:*

*Sada 2:*

*Sada 3:*

*Sada 4:*

5. Byla vyšetřována výška 20 18letých mladíků  $y$  a výška jejich rodičů a prarodičů, žijících izolovaně v horské vesnici po několik generací a hledaná lineární závislost mezi závisle proměnnou  $y$  a nezávisle proměnnými  $x_1$  až  $x_7$  (podrobnější popis regresorů - viz [Mladici.xls](#)). Navrhněte pro tuto závislost regresní model, verifikujte jej a posuďte jeho kvalitu. Na základě nalezeného modelu predikujte výšku 18letého mladíka z dat jeho rodičů a prarodičů:  $x_1 = 51$  cm,  $x_2 = 152$  cm,  $x_3 = 183$  cm,  $x_4 = 155$  cm,  $x_5 = 180$  cm,  $x_6 = 157$  cm,  $x_7 = 178$  cm. (*Součástí řešení příkladu musí být závěr obsahující: nalezený model, hodnocení jeho věrohodnosti a kvality (slovně) + konkrétní vysvětlení výsledku predikce, včetně komentáře k její věrohodnosti!!!*)

a) *Vizuální posouzení:*

b) *Identifikace vlivných bodů:*

c) *Primárně zvolený regresní model:*

d) *Ověření normality závislé proměnné i regresorů:*

e) *Multikolinearita:*

f) *Celkový  $F$ -test:*

g) *Dílčí  $t$ -testy:*

h) *Optimalizovaný regresní model:*

i) *Analýza reziduí:*

*Ověření nulové střední hodnoty:*

*Ověření nulové autokorelace:*

*Ověření homoskedasticity:*

j) *Index determinace:*

*Hodnocení kvality modelu:*

k) *Závěr:*

l) *Predikce:*

*Bodový odhad:*

*Intervalový odhad:*

*Komentář k věrohodnosti predikce:*