PŘÍKLADY ZE CVIČENÍ PSE

LS 2014/2015

MICHAL FRIESL

Na následujících stranách jsou sepsána zadání zhruba těch příkladů, které se budou počítat na mých cvičeních KMA/PSE v letním semestru školního roku 2014/2015. Smyslem je využít krátký čas, který je pro cvičení vyhrazen, efektivněji, tedy nezdržovat se opisováním zadání.

Jsou přidány i některé poznámky k teorii, ty ale v žádném případě nemají a nemohou nahradit výklad, který by si posluchači měli nejprve přečíst v příslušném studijním textu. Na poslední straně se pak nacházejí nejnutnější čísla ze statistických tabulek.

U zcela základních příkladů, bez jejichž pochopení snad ani nemá smysl pokoušet se psát zápočtový test, je připojen odkaz na WWW stránku. Tam lze procházet jejich podrobné řešení.

Informace k mým cvičením:

- http://home.zcu.cz/~friesl/Vyuka/Pse.html

Materiály, na něž je v jednotlivých částech odkazováno:

- B. Šedivá a kol.: Pravděpodobnost a statistika pro FEL, aktualizováno 17. března 2011, dostupné z mého archivu
  Text připravený k předmětu KMA/PSE.
- M. Friesl: Pravděpodobnost a statistika hypertextově, verze 2014-02-08, dostupné z http://home.zcu.cz/~friesl/hpsb/
  Stručný přehled pojmů v podobě miniencyklopedie.
- M. Friesl: Pravděpodobnost a statistika — podrobná řešení, pracovní verze, dostupné z http://home.zcu.cz/~friesl/kpse/
  Několik základních příkladů. Zároveň obsahuje seznam vzorečků.

1. Pravděpodobnost

V učebním textu k PSE: Cvičení 1
V hypertextově: na webu str. 7, 9–10 (ve verzi k tisku odst. 1.1, 1.3–1.4)

Co se hodí vědět:

- Pracujeme s množinou Ω výsledků náhodného pokusu (množina elementárních jevů). Náhodný jev $\equiv$ obyčejná podmnožina $A \subset \Omega$.
  $A \cup B$: A nebo B (aspoň jeden), $A \cap B$: A a B (oba zároveň),
  $A \subset B$: nastane-li A, pak i B, $A \cap B = \emptyset$: neslučitelné.

- Pravděpodobnost je funkce, která jevům přiřazuje čísla z $\langle 0, 1 \rangle$.
  - sjednocení $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$
  - přes doplněk $\mathrm{P}(A \cup B) = 1 - \mathrm{P}(\overline{A} \cap \overline{B})$

  Nelze vynechat (kromě neslučitelných)

- Pravděpodobnost říká „kolik místa“ (jakou část) jev A zabírá v Ω. Jsou-li všechny elementární jevy stejně pravděpodobné, pak

$$\mathrm{P}(A) = \frac{|A| \leftarrow \text{počet příznivých}}{|\Omega| \leftarrow \text{počet všech}} \quad \text{(„klasická pravděpodobnost“)}$$

Ne vždy ji lze použít!

Počty k-tic sestavovaných z n prvků:
- podmnožin (bez pořadí, kombinace): $\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \binom{n}{n-k}$
- uspoř. k-tic (variace): s opak. n^k, bez opak. $n \cdot \ldots \cdot (n-k+1)$

1.1. U jednotlivých příkladů náhodných pokusů zapište množinu reprezentující náhodný jev, který je vpravo vyjádřený slovně:

Náhodný pokus	Množina výsledků Ω	Náhodný jev $A \subset \Omega$
a) Hod kostkou	$\{1, 2, 3, 4, 5, 6\}$ konečná	padne sudé číslo
b) Doba životnosti stroje	$(0, \infty)$ nekonečná nespočetná	vydrží déle než rok
		vydrží přesně rok
c) Napokolikáté dá zápočet	$\{\overbrace{0}^{\text{nepokusil se}}, 1, 2, \ldots, \overbrace{\infty}^{\text{opakovaně ne}}\}$	zápočet udělá
d) Hod mincí	$\{L, R\}$ abstraktní	padne líc
e) Dva hody mincí	$\{(L, L), (L, R), (R, L), (R, R)\}$	padne líc
f) Ze 4 studentů losováni 2	$\{\{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}\}$	a vylosován

1.2. Zapište jako množiny další jevy a prozkoumejte vztahy mezi nimi (implikace, neslučitelnost): a) A: padne šestka, B: padne sudé, b) A: životnost aspoň rok, B: životnost aspoň půl roku.

1.3. Je vylosováno přirozené číslo. Jev A značí, že bude dělitelné 5, jev B, že 10. Zakreslete, vypočtěte a slovně vyjádřete jevy $A \cap B$, $A \cup B$, $\overline{A} \cap B$, $A \cap \overline{B}$, $A \cup \overline{B}$, $\overline{A \cap B}$.

1.4. Výrobek je podroben vždy třem zkouškám. A, B, C budiž jevy, že obstojí v 1., 2., resp. 3. zkoušce. Zapište množinově a zakreslete jevy a) obstojí ve všech třech, b) obstojí v 1. a 2., ale ne ve 3., c) obstojí aspoň v jedné, d) neobstojí v žádné, e) obstojí jen v 1., f) obstojí aspoň ve dvou, g) obstojí maximálně ve dvou.

1.5. Ze 4 studentů budou losování dva. Určete pravděpodobnosti, že mezi vylosovanými bude a) student b, b) a i b, c) a nebo b.

1.6. Pravděpodobnosti, že životnost zařízení (v letech) bude delší než t, jsou e^{-t}, $t > 0$. Určete pravděpodobnosti, že se porouchá a) během prvního roku, b) ve 2. pololetí 1. roku, c) přesně za rok.

1.7. Jaká je pravděpodobnost, že při hodu 2 kostkami bude součet 8?

1.8. Máme 16 lahví minerálek: 10 Šaratic a 6 Mlýnských pramenů. Náhodně vybereme 3 lahve. Jaká je pravděpodobnost, že vybereme 2 Šaratice a 1 Mlýnský pramen?

1.9. U výrobku se rozlišují dva druhy závad. Víme, že výrobek bude mít vyhovující rozměr s pravděpodobností 0,96 a vyhovující hmotnost s pravděpodobností 0,93, a dále, že některou ze závad má s pravděpodobností 0,09. Jaká je pravděpodobnost, že výrobek nebude mít obě závady současně?

1.10. Z 10 studentů je losována tříčlenná komise. Jaká je pravděpodobnost, že A **nebo** B budou mezi vylosovanými?

1.11. Ze 30 žáků ve třídě jich je 70% připraveno. Budou zkoušeni 3 žáci. Jaká je pravděpodobnost, že aspoň 2 z nich budou připraveni?

2. Nezávislost

V učebním textu k PSE: Cvičení 2
V hypertextově: na webu str. 11 (ve verzi k tisku odst. 1.5)

Co se hodí vědět:

- Nezávislost dvou jevů: A, B nezávislé $\equiv \mathrm{P}(\underbrace{A \cap B}_{\text{nastanou zároveň}}) = \mathrm{P}(A) \cdot \mathrm{P}(B)$

- Nezávislost více jevů: Když rovnosti uvedeného typu platí pro všechny dvojice, trojice, ... z nich vybrané. Např. tři jevy A_1, A_2, A_3 nezávislé $\equiv$ platí všechny rovnosti
$$\begin{aligned} \mathrm{P}(A_1 \cap A_2) &= \mathrm{P}(A_1) \cdot \mathrm{P}(A_2) \\ \mathrm{P}(A_1 \cap A_3) &= \mathrm{P}(A_1) \cdot \mathrm{P}(A_3) \\ \mathrm{P}(A_2 \cap A_3) &= \mathrm{P}(A_2) \cdot \mathrm{P}(A_3) \\ \mathrm{P}(A_1 \cap A_2 \cap A_3) &= \mathrm{P}(A_1) \cdot \mathrm{P}(A_2) \cdot \mathrm{P}(A_3) \end{aligned}$$

- Rovnosti neplatí automaticky! Pokud nevíme, zda jsou jevy nezávislé, je nutné rovnosti ověřit, nebo je nepoužívat.

- V některých situacích naopak víme, že jevy jsou nezávislé, pak je možné použít libovolné z uvedených rovností. Tedy, jsou-li např. jevy $A_1, \ldots, A_n$ nezávislé, pak platí mj.
$$\mathrm{P}(\underbrace{A_1 \cap \cdots \cap A_n}_{\text{nastanou všechny}}) = \mathrm{P}(A_1) \ldots \mathrm{P}(A_n)$$
$$\mathrm{P}(\underbrace{A_1 \cup \cdots \cup A_n}_{\text{nastane aspoň jeden}}) = 1 - \underbrace{\big(1 - \mathrm{P}(A_1)\big)}_{\text{nenastane } A_1} \cdots \underbrace{\big(1 - \mathrm{P}(A_n)\big)}_{\text{nenastane } A_n}$$

!! **2.1.** Známe pravděpodobnosti $\mathrm{P}(A) = 0{,}3$, $\mathrm{P}(B) = 0{,}7$, $\mathrm{P}(A \cup B) = 0{,}8$. Určete pravděpodobnost $\mathrm{P}(A \cap B)$ a rozhodněte, zda jsou jevy A a B nezávislé. *(Viz Podrobně 1.1)*

!! **2.2.** Hodíme hrací kostkou. Nechť A značí jev, že padne sudé číslo, B jev, že padne číslo dělitelné 3, a C jev, že padne 3 nebo 4. Rozhodněte zda jevy A a B jsou nezávislé, podobně zda B a C. *(Viz Podrobně 1.2)*

2.3. U výrobku se rozlišují dva druhy závad. Víme, že výrobek bude mít vyhovující rozměr s pravděpodobností 0,96 a vyhovující hmotnost s pravděpodobností 0,93, a dále, že některou ze závad má s pravděpodobností 0,09. Jsou závada v rozměru a v hmotnosti nezávislé?

!! **2.4.** Zařízení se porouchá, jakmile u některé ze 7 součástek nastane porucha. Pravděpodobnosti poruch jsou 0,01. Jsou-li poruchy součástek nezávislé, jaká je pravděpodobnost, že se zařízení porouchá? *(Viz Podrobně 1.3)*

2.5. Zařízení je sestavené z nezávisle se chovajících součástek typu A (které pracují s pravděpodobností 0,9) a B (které pracují s pravděpodobností 0,8). S jakou pravděpodobností funguje, jestliže jsou součástky zapojené a) $A\mathrm{s}B$, b) $(A\mathrm{s}B)\mathrm{p}(A\mathrm{s}B)$, tj. zálohování celého zařízení, c) $(A\mathrm{p}A)\mathrm{s}(B\mathrm{p}B)$, tj. zálohování jednotlivých součástek?

2.6. V zařízení $A\mathrm{s}(B_1\mathrm{p}B_2)$ nastávají poruchy u jednotlivých součástek nezávisle, s pravděpodobnostmi 0,03, 0,2, 0,2. Jaká je pravděpodobnost, že zařízení bude mít poruchu?

2.7. Student skládá 5 zkoušek, na každou má 3 pokusy. Jaká je pravděpodobnost, že složí všechny zkoušky, má-li u každého pokusu pravděpodobnost úspěchu $p = 0{,}6$? Kolik by bylo potřeba povolit pokusů, aby tato pravděpodobnost byla aspoň 90 %?

3. Podmíněná pravděpodobnost

V učebním textu k PSE: Cvičení 2 a 3
V hypertextově: na webu str. 12–14 (ve verzi k tisku odst. 1.6–1.8)

Co se hodí vědět:

- Podmíněná pravděpodobnost jevu A za podmínky, že nastal jev B, říká, jak velkou část zabírají prvky množiny A v množině B (oproti nepodmíněné, kde v celém Ω)

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(\overbrace{A \cap B}^{\text{nastanou oba}})}{\mathrm{P}(B)} \begin{array}{l} \leftarrow \text{část } A \text{ ležící v } B \\ \leftarrow \text{ku } B \end{array}$$

Pří pevném B má všechny vlastnosti pravděpodobnosti.

- Souvislost s nezávislostí: U nezávislých jevů je $\mathrm{P}(A \mid B)$ stejná jako $\mathrm{P}(A \mid \overline{B})$ a rovna nepodmíněné $\mathrm{P}(A)$, u závislých se liší.

- Věta o celkové pravděpodobnosti: Známe-li jednotlivé podmíněné pravděpodobnosti $\mathrm{P}(A \mid B_i)$ pro všechna B_i z nějakého disjunktního rozkladu $B_1 \cup \cdots \cup B_n = \Omega$ množiny všech výsledků, pak z nich můžeme poskládat celkovou pravděpodobnost

$$\mathrm{P}(A) = \mathrm{P}(A \mid B_1)\,\mathrm{P}(B_1) + \cdots + \mathrm{P}(A \mid B_n)\,\mathrm{P}(B_n)$$

- Bayesova věta: Za situace v předchozím bodě lze dopočítat i „opačnou“ podmíněnou pravděpodobnost

$$\mathrm{P}(B_i \mid A) = \frac{P(A \mid B_i)\,\mathrm{P}(B_i)}{\mathrm{P}(A \mid B_1)\,\mathrm{P}(B_1) + \cdots + \mathrm{P}(A \mid B_n)\,\mathrm{P}(B_n)}$$

3.1. Ze 4 studentů budou losování dva. Je-li ve vylosované dvojici b, jaká je pravděpodobnost, že že je tam i a?

!! **3.2.** Hodíme hrací kostkou. Nechť A značí jev, že padne sudé číslo, B jev, že padne číslo dělitelné 3, a C jev, že padne 3 nebo 4. Určete podmíněné pravděpodobnosti $\mathrm{P}(A \mid B)$ a $\mathrm{P}(B \mid C)$. *(Viz Podrobně 1.2)*

3.3. Dvakrát hodíme kostkou. Jaká je pravděpodobnost, že součet přesahuje 10, víme-li, že mezi hozenými čísly je (aspoň jedna) šestka?

!! **3.4.** Ve skladu

- 10 % výrobků pochází od 1. dodavatele,
- 60 % výrobků pochází od 2. dodavatele,
- 30 % výrobků pochází od 3. dodavatele,

přičemž výrobek dodaný

- 1. dodavatelem je vadný s pravděpodobností 1 %,
- 2. dodavatelem je vadný s pravděpodobností 0,5 %,
- 3. dodavatelem je vadný s pravděpodobností 2 %.

S jakou pravděpodobností bude výrobek náhodně vybraný ze skladu vadný? Je-li takto vybraný výrobek vadný, s jakými pravděpodobnostmi pochází od 2. resp. 3. dodavatele? *(Viz Podrobně 1.4)*

3.5. Z balíčku 32 mariášových karet postupně vytáhneme dvě karty (bez vracení). Určete pravděpodobnosti jevů A_1 a A_2, že první, resp. druhá tažená karta bude eso.

3.6. Pravděpodobnost odhalení AIDS při testu je 0,999, pravděpodobnost správného otestování zdravého jedince je 0,99. Víme-li, že AIDS se vyskytuje u 0,006 lidí, jaká je pravděpodobnost, že člověk, u kterého byl test pozitivní, AIDS skutečně má?

4. Náhodná veličina

V učebním textu k PSE: Cvičení 4
V hypertextově: na webu str. 16–18 (ve verzi k tisku odst. 2.1–2.3)

Co se hodí vědět:

- Náhodná veličina je obyčejná funkce $X\colon \Omega \to \mathbf{R}$, každému elementárnímu jevu $\omega \in \Omega$ je přiřazeno nějaké reálné číslo $X(\omega)$.

- Sledujeme, jakých hodnot a s jakými pravděpodobnostmi veličina X nabývá (rozdělení veličiny), např. pro daná čísla $x \in \mathbf{R}$

$$\mathrm{P}(X = x), \quad \mathrm{P}(X \leqq x), \quad \mathrm{P}(X > x), \quad \text{apod.}$$

- Rozdělení je určeno distribuční funkcí $F(x) = \mathrm{P}(X \leqq x)$, $x \in \mathbf{R}$,
  - s rostoucím x do jevu $X \leqq x$ přispívají další hodnoty veličiny
  - zprava spojitá, neklesající, $F(-\infty) = 0$, $F(\infty) = 1$.

- Rozlišujeme veličiny diskrétní (má nejvýše spočetně hodnot), spojité (viz dále), jinými se zabývat nebudeme.

4.1. Veličina X má hodnotu 1, když při hodu mincí padne líc, a 0, když rub. U veličiny Y je tomu opačně. Určete a nakreslete jejich distribuční funkce.

4.2. Pravděpodobnosti, že životnost zařízení (v letech) bude delší než t, jsou e^{-t}, $t \in (0, \infty)$. Veličina X udává dobu životnosti. Určete a nakreslete její distribuční funkci.

4.3. Náhodně vybereme reálné číslo z intervalu $(0, 1)$. Náhodná veličina udává jeho hodnotu po zaokrouhlení na desetiny. Určete a nakreslete její distribuční funkci.

... Diskrétní veličina

V učebním textu k PSE: Cvičení 4
V hypertextově: na webu str. 19–20 (ve verzi k tisku odst. 2.4-2.5)

Co se hodí vědět:

- Diskrétní veličina: Nejvýše spočetně hodnot, pro jejich pravděpodobnosti $p_k = P(X = k)$ platí $\sum_k p_k = 1$.

- Neplést pravděpodobnostní funkci $P(x) = \mathrm{P}(X = x)$ (pravděpodobnosti jednotlivých hodnot, grafem „tyčky“)
a distribuční funkci $F(x) = \mathrm{P}(X \leqq x)$ (postupně nasčítávané pravděpodobnosti, schodovitá).

- Na rozdíl od spojitých veličin:
  - jednotlivé hodnoty mají kladné pravděpodobnosti
  - je rozdíl mezi $\mathrm{P}(X < x)$ a $\mathrm{P}(X \leqq x)$, např. pro celočíselnou:

$$\left.\begin{aligned} \mathrm{P}(X < k) &= \cdots + p_{k-1} &&= F(k-1) \\ \mathrm{P}(X \leqq k) &= \cdots + p_{k-1} + p_k &&= F(k) \end{aligned}\right\} \text{liší se o } p_k = \mathrm{P}(X = k) = F(k) - F(k-1)$$

- Charakteristiky:
  - Střední hodnota $\mathrm{E}\,X$: v průměru (při opakovných pokusech) očekávaná hodnota, $\mathrm{E}\,X = \sum_k k \cdot p_k$ (vážený průměr).
  - Rozptyl $\mathrm{var}\,X$ (též $\mathrm{D}\,X$, $\sigma^2(X)$ apod.): jak moc kolísají hodnoty veličiny okolo její střední hodnoty,

$$\mathrm{var}\,X = \mathrm{E}(X - \mathrm{E}\,X)^2 = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2, \quad \text{kde } \mathrm{E}\,X^2 = \sum_k k^2 p_k.$$

  - Směrodatná odchylka: odmocnina z rozptylu.
  - Platí $\mathrm{E}(X + Y) = \mathrm{E}\,X + \mathrm{E}\,Y$, $\mathrm{E}(cX) = c\,\mathrm{E}\,X$.
    Jsou-li X, Y nezávislé, pak také $\mathrm{var}(X + Y) = \mathrm{var}\,X + \mathrm{var}\,Y$.

!! **4.4.** Náhodná veličina X udává, kolikrát při dvou hodech mincí padne líc. Odvoďte pravděpodobnostní funkci P veličiny X, její distribuční funkci F, střední hodnotu $\mathrm{E}\,X$ a rozptyl $\mathrm{var}\,X$. Pomocí distribuční funkce pak vyjádřete pravděpodobnosti, že líc padne aspoň jednou a že právě jednou. *(Viz Podrobně 2.1)*

4.5. Náhodně vybereme reálné číslo z intervalu $(0, 1)$. Náhodná veličina udává jeho hodnotu po zaokrouhlení na desetiny. Určete její pravděpodobnostní a distribuční funkci, spočtěte střední hodnotu a rozptyl.

4.6. V dílně pracují dva stroje. První má poruchu s pravděpodobností 0,3, druhý s pravděpodobností 0,4 a poruchy strojů jsou nezávislé. Veličina X udává počet porouchaných strojů. Určete její pravděpodobnostní a distribuční funkci, střední hodnotu a rozptyl.

!! **4.7.** Náhodná veličina X nabývá pouze hodnot $0, 1, 2, \ldots$ s pravděpodobnostmi $p_k = c(5/6)^k$, $k = 0, 1, 2, \ldots$. Určete konstantu c a pravděpodobnost $\mathrm{P}(X \geqq 2)$. *(Viz Podrobně 2.2)*

4.8. Student má 3 pokusy na zkoušku, u každého stejnou pravděpodobnost úspěchu $p = 60\,\%$ nezávisle na ostatních. Kolikrát v průměru jde na zkoušku?

5. Některá rozdělení

V učebním textu k PSE: Cvičení 5 a 6
V hypertextově: na webu str. 42–45 (ve verzi k tisku odst. 4.1-4.4)

Co se hodí vědět:

- Binomické rozdělení $\mathrm{Bi}(n,p)$ ($n \in \mathbf{N}, p \in (0,1)$):
  - obor hodnot $0, 1, \ldots, n$
  - $\mathrm{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$
  - $\mathrm{E}\, X = np$, $\mathrm{var}\, X = np(1-p)$
  - typicky počet výskytů jevu („úspěchů“) při n nezávisle opakovaných pokusech (za stejných podmínek, při každém s pravděpodobností výskytu p), výběry s vracením

- Hypergeometrické rozdělení $\mathrm{HG}(N,K,n)$ ($N, K, n \in \mathbf{N}$, $K < N$):
  - obor hodnot je podmnožinou $0, 1, \ldots, n$
  - $\mathrm{P}(X = k) = \binom{K}{k}\binom{N-K}{n-k} / \binom{N}{n}$, $0 \leqq k \leqq K, 0 \leqq n-k \leqq N-K$
  - $\mathrm{E}\, X = n\frac{K}{N}$
  - typicky počet „úspěchů“ při výběru bez vracení

- Poissonovo rozdělení $\mathrm{Po}(\lambda)$ ($\lambda > 0$):
  - obor hodnot $0, 1, 2, \ldots$ (nabývá hodnot shora neomezených)
  - $\mathrm{P}(X = k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$, $k = 0, 1, , 2, \ldots$
  - $\mathrm{E}\, X = \lambda$, $\mathrm{var}\, X = \lambda$
  - typicky počet výskytů jevu, který se objevuje v čase náhodně (s intenzitou λ) v daném období.

- Aproximace:
  - Je-li n vůči N malé, pak $\mathrm{HG}(N,K,n) \approx \mathrm{Bi}(n, K/N)$.
  - Je-li n velké (> 30) a p malé ($< 0{,}1$), pak $\mathrm{Bi}(n,p) \approx \mathrm{Po}(np)$.

!! **5.1.** Hodíme 5 kostkami. Jaká je pravděpodobnost, že padnou aspoň dvě šestky? *(Viz Podrobně 2.3)*

5.2. Hodíme desetkrát mincí. Jaká je pravděpodobnost, že padne stejně líců jako rubů?

5.3. Z 10 výrobků, mezi nimiž jsou 3 vadné, postupně vybereme 2 výrobky. Veličina X nechť udává počet vadných (mezi vybranými). Určete pravděpodobnostní funkci veličiny X, jestliže vybraný výrobek a) vracíme zpět, b) nevracíme zpět.

!! **5.4.** Veličina X se řídí Poissonovým rozdělením pravděpodobnosti s parametrem $\lambda = 4$. Určete pravděpodobnosti $P(X = 2)$ a $P(X \leqq 2)$. *(Viz Podrobně 2.4)*

5.5. Na lince se vyskytují poruchy náhodně, v průměru 2 za týden. Jaká je pravděpodobnost, že v týdnu bude výroba přerušena více než třikrát?

5.6. Odběratel nepřijme dodávku 1000 výrobků, pokud mezi 30 namátkou vybranými zjistí více než 2 zmetky. S jakou pravděpodobností bude zamítnuta dodávka obsahující 50, resp. 100 zmetků?

6. Spojité veličiny

V učebním textu k PSE: Cvičení 7
V hypertextově: na webu str. 22–23 (ve verzi k tisku odst. 2.6–2.7)

Co se hodí vědět:

- Spojitá veličina: Nabývá hodnot z celých intervalů.

- Rozdělení lze popsat hustotou f (nezáporná, $\overbrace{\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x}^{\text{obsah plochy pod } f} = 1$).
Neplést f s distribuční funkcí F: $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$, $f(x) = F'(x)$.

- Pravděpodobnosti jsou rovny obsahům ploch pod grafem f,
$\mathrm{P}(X \in (a,b)) = \int_a^b f(x)\,\mathrm{d}x = F(b) - F(a)$ a stejně pro $\langle a,b)$,...

- Odlišnosti od diskrétní veličiny:
  - jednotlivé hodnoty mají nulové pravděpodobnosti
  - distribuční funkce F je spojitá
  - pravděpsti $P(X < x)$ a $\mathrm{P}(X \leqq x)$ jsou stejné, obě rovny $F(x)$

- V charakteristikách místo p_x máme $f(x)\,\mathrm{d}x$ a místo $\sum$ máme $\int$,
např. $\mathrm{E}\,X = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$, $\mathrm{E}\,X^2 = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x$

- p-kvantil: ta hodnota x_p, pro kterou $F(x_p) = p$
  - dělí plochu pod hustotou na p a $1-p$
  - speciálně $\underbrace{\text{medián}}_{50\,\%\text{ kvantil}}$, $\underbrace{\text{první a třetí}}_{25\,\%\text{ a }75\,\%\text{ kvantil}}$ (dolní a horní) kvartil,

6.1. Veličina X udává hodnotu náhodně vybraného čísla z intervalu $(0,1)$. Napište předpis její distribuční funkce a hustoty (načrtněte i grafy), spočtěte střední hodnotu a rozptyl.

!! **6.2.** Veličina X má hustotu $f(x) = c(2-2x)$ pro $x \in (-2,1)$ a nulovou jinde. Určete konstantu c, střední hodnotu $\mathrm{E}\,X$, předpis pro distribuční funkci $F(x)$ a pravděpodobnost $\mathrm{P}(X < 0)$. *(Viz Podrobně 2.5)*

6.3. Náhodní veličina má hustotu $f(x) = c - \frac{x}{2}$ pro $x \in (0,2)$ a nulovu jinde. Určete konstantu c.

6.4. Náhodná veličina X má hustotu $f(x) = \frac{a}{1+x^2}$ na $\mathbf{R}$ (Cauchy). Určete a, distribuční funkci, $\mathrm{P}(X > \sqrt{3})$, střední hodnotu a medián.

!! **6.5.** Náhodná veličina X má distribuční funkci $F(x) = 1 - \mathrm{e}^{-x/5}$ pro $x > 0$. Najděte hustotu $f(x)$ a pravděpodobnost $\mathrm{P}(5 < X < 6)$. Určete 5 % kvantil veličiny X. *(Viz Podrobně 2.6)*

6.6. Náhodná veličina X má distribuční funkci $x^2/4$ na $(0,2)$, nulovou pro $x < 0$ a jednotkovou pro $x > 2$. Najděte její hustotu, medián, střední hodnotu a $\mathrm{P})0{,}5 < X < 1{,}5)$.

7. Některá spojitá rozdělení

V učebním textu k PSE: Cvičení 8
V hypertextově: na webu str. 46 (ve verzi k tisku odst. 4.5)

Co se hodí vědět:

- Rovnoměrné rozdělení $\mathrm{Ro}(a,b)$ $(-\infty < a < b < \infty)$
  - hodnoty náhodně vybrané z omezeného intervalu (a,b),
  - hustota $f(x) = 1/(b-a)$ na (a,b) a 0 jinde, distribuční funkce $F(x) = 0$ pro $x \leqq a$, $\frac{x-a}{b-a}$ pro $a < x < b$, 1 pro $x \geqq b$,
  - $\mathrm{E}\,X = \frac{a+b}{2}$, $\mathrm{var}\,X = \frac{(b-a)^2}{12}$.

- Exponenciální rozdělení $\mathrm{Exp}(\delta)$ $(\delta > 0)$
  - nezáporná veličina
  - hustota $f(x) = (1/\delta)\mathrm{e}^{-x/\delta}$ pro $x > 0$ a 0 jinde distribuční funkce $F(x) = 0$ pro $x \leqq 0$, a $1 - \mathrm{e}^{-x/\delta}$ pro $x > 0$
  - $\mathrm{E}\,X = \delta$, $\mathrm{var}\,X = \delta^2$
  - doba životnosti zařízení, které „nestárne“ či doby mezi událostmi vyskytujícími se v čase náhodně
  - někdy místo δ parametr $\lambda = 1/\delta$ (intenzita poruch)

- Řada dalších
  - normální $\mathrm{N}(\mu, \sigma^2)$ — viz dále,
  - Weibullovo, gama, beta, logaritmicko-normální, extrémních hodnot, ...
  - ve statistice Studentovo t, Fisherovo-Snedecorovo F, chí-kvadrát

7.1. Náhodná veličina X má $\mathrm{Ro}(-2,0)$. Napište její hustotu a distribuční funkci a určete $\mathrm{P}(-1/2 < X < 0)$.

7.2. Doba čekání na dalšího zákazníka (v minutách) se řídí rozdělením s hustotou $f(x) = 0$, $x < 0$, $f(x) = 0{,}1\mathrm{e}^{-0{,}1x}$, $x > 0$. a) Jaká je pravděpodobnost, že zákazník dorazí přesně po 10 minutách? Jaká, že dorazí v průběhu 10. minuty? Jaká, že během hodiny nepřijde? b) Jak dlouho se bude na zákazníka v průměru čekat? Jaký je rozptyl doby čekání? c) Určete dobu, během níž s pravděpodobností 90 % přijde.

7.3. Doba do poruchy má exponenciální rozdělení s intenzitou poruch 0,02. Najděte střední dobu do poruchy a pravděpodobnost, že po dobu 80 hodin nedojde k poruše. Jaká je pravděpodobnnost tohoto jevu, pracuje-li zařízení už 50 hodin bez poruchy?

7.4. Nechť životnost výrobků se řídí exponenciálním rozdělením s distribuční funkcí $F(x) = 1 - e^{-x/5}$, $x > 0$. Jakou záruční dobu stanoví výrobce, nemá-li podíl reklamovaných výrobků překročit 10%?

... Normální rozdělení

V učebním textu k PSE: Cvičení 9
V hypertextově: na webu str. 47 (ve verzi k tisku odst. 4.6)
V hypertextově: na webu str. 56,58 (ve verzi k tisku odst. 5.4,5.6)

Co se hodí vědět:

- Normální rozdělení $\mathrm{N}(\mu, \sigma^2)$ ($\mu \in \mathbf{R}$, $\sigma^2 > 0$)
  - hustota symetrická a vysoká kolem μ (σ malé = více koncentrovaná, σ velké = hustota více do šířky)
  - parametry rovny momentům: $\mathrm{E}\,X = \mu$, $\mathrm{var}\,X = \sigma^2$,
  - lineární transformace mají opět normální rozdělení
  - typicky jako aproximace

- Výpočty převodem na normální normované $\mathrm{N}(0,1)$
  - $X \sim \mathrm{N}(\mu, \sigma^2)$ právě když $\frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$
  - tudíž $F_X(x) = \Phi(\frac{x-\mu}{\sigma})$ a pro kvantily $x_p = \mu + \sigma u_p$
  - kde $\Phi(x)$ a u_p (distribuční funkce a kvantily $\mathrm{N}(0,1)$) tabulky, plus symetrie $\Phi(-x) = 1 - \Phi(x)$, $u_{1-p} = -u_p$

- Centrální limitní věta: Rozdělení součtu n nezávislých stejně rozdělených veličin se středními hodnotami $\mathrm{E}\,X_i = \mu$ a rozptyly $\mathrm{var}\,X_i = \sigma^2$ je při velkém n přibližně normální

$$\sum_1^n X_i \approx \mathrm{N}(n\mu, n\sigma^2), \quad \mathrm{Bi}(n,p) \approx \mathrm{N}(np, np(1-p)),$$

Mají-li X_i normální rozdělení, platí přesně.

- Též průměr $\frac{\sum_1^n X_i}{n} \approx \mathrm{N}(\mu, \sigma^2/n)$, či $\mathrm{Po}(\lambda) \approx \mathrm{N}(\lambda, \lambda)$ při $\lambda > 9$.

!! **7.5.** Délky výrobků v mm se řídí normálním rozdělením $\mathrm{N}(68{,}3; 0{,}04)$. S jakou pravděpodobností bude mít výrobek délku mezi 68 a 69 mm? Určete čísla a a b, pro která $P(X < a) = 0{,}1$ a $\mathrm{P}(X < b) = 0{,}9$. *(Viz Podrobně 2.7)*

7.6. Měsíční výdaje (v Kč) domácností na určité zboží mají $\mathrm{N}(900, 19\,600)$. Jaká je pravděpodobnost, že výdaje jedné náhodně vybrané domácnosti překročí 1000 Kč?

7.7. Jaká je pravděpodobnost, že po 200 hodinách provozu budou fungovat aspoň 3 výrobky z 5, jestliže jejich životnost v hodinách má $\mathrm{N}(180, 400)$?

7.8. Výsledky radarového měření jsou zatíženy normálně rozdělenou náhodnou chybou s nulovou střední hodnotou, která s pravděpodobností 0,95 nepřesahuje ± 20 m. Určete směrodatnou odchylku měření.

!! **7.9.** Zatížení letadla s 64 místy nemá překročit 6000 kg. Jaká je pravděpodobnost, že při plném obsazení bude tato hodnota překročena, mají-li hmotnosti cestujících střední hodnotu 90 kg a směrodatnou odchylku 10 kg? *(Viz Podrobně 2.8)*

7.10. Hmotnost pomerančů v dodávce má $\mathrm{N}(170, 144)$ (v gramech). Jaká je pravděpodobnost, že síťka s 8 pomeranči bude vážit více než 1,5 kg?

7.11. Měsíční výdaje (v Kč) domácností na určité zboží mají $\mathrm{N}(900, 19\,600)$. Jaká je pravděpodobnost, že průměrné výdaje 5 náhodně vybraných domácností překročí 1000 Kč?

8. Odhady parametrů

V učebním textu k PSE: Cvičení 10
V hypertextově: na webu str. 60–68 (ve verzi k tisku odst. 6.1–6.9)
V hypertextově: na webu str. 75–76 (ve verzi k tisku odst. 6.16–6.17)

Co se hodí vědět:

- Statistika: Rozdělení neznáme, závěry o populaci činíme na základě opakovaných pozorování $X_1, \dots, X_n$ (výběrový soubor).

- Pozorování předpokládáme nezávislá (zamyslet se) a stejně rozdělená $\equiv$ náhodný výběr rozsahu n.

- Bodové odhady — Odhad:
  - výběrový průměr $\overline{X} = \sum_1^n X_i/n$ — ... stř. hodnoty
  - výběrový rozptyl $s^2 = \frac{1}{n-1}(\sum_1^n X_i^2 - n\overline{X}^2)$ — ... rozptylu
  - relativní četnost $\frac{n_A}{n}$ $\leftarrow$počet výskytů hodnot z A $\leftarrow$počet pozorování — ... $\mathrm{P}(X_i \in A)$
  - výběrový medián: prostřední (či průměr dvou prostředních) v posloupnosti pozorování seřazených podle velikosti

- Intervalový odhad parametru θ se spolehlivostí $1-\alpha$: interval s takovými (náhodnými) krajními body, že

$$\mathrm{P}(D(X_1, \dots, X_n) < \theta < H(X_1, \dots, X_n)) = 1 - \alpha$$

Pro stř. hodnotu μ při výběru z $\mathrm{N}(\mu, \sigma^2)$ (nebo přibližný z jiného)

- $\left(\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}, \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right)$
- $\left(\overline{X} - \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1), \overline{X} + \frac{s}{\sqrt{n}} t_{1-\alpha/2}(n-1)\right)$

Pro parametr p binomického rozdělení $S \sim \mathrm{Bi}(n, p)$, $\widehat{p} = S/n$,

- $\left(\overline{X} - \frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}} u_{1-\alpha/2}, \overline{X} + \frac{\sqrt{\widehat{p}(1-\widehat{p})}}{\sqrt{n}} u_{1-\alpha/2}\right)$

!! **8.1.** U 26 náhodně vybraných studentů byly naměřeny výšky 190, 175, 168, 182, 180, 179, 200, 191, 156, 180, 178, 191, 185, 202, 182, 187,5, 166, 182, 178, 177, 176, 182, 185, 175, 182, 185 cm. Zakreslete histogram četností a odhadněte průměrnou výšku. *(Viz Podrobně 3.1)*

!! **8.2.** Z 42 náhodně vybraných účastníků sportovního odpoledne bylo 16 dívek a 26 chlapců. Odhadněte podíl dívek mezi účastníky. *(Viz Podrobně 3.2)*

8.3. Data z ankety PSE. Zakreslete histogram odpovědí na otázku, jak posluchači studují, a odhadněte (jednak bodově, jednak intervalově se spolehlivostí 90 % a 95 %) podíl studentů, kteří se připravují pravidelně před každým cvičením.

8.4. Data z ankety PSE. Zakreslete histogramy odpovědí na otázky, kolik bodů z písemky posluchači čekali před písemkou a kolik po písemce. Dále pro počet bodů, které studenti v průměru očekávají před a po písemce, sestrojte bodové odhady a 90 % a 95 % intervaly spolehlivosti.

8.5. Deset balíčků mouky pocházejících z balícího stroje mělo hmotnosti v gramech: 987, 1 001, 993, 994, 993, 1 005, 1 007, 999, 995, 1 002. Sestrojte 95% interval spolehlivosti pro střední hodnotu a rozptyl hmotnosti (předpokládejte normální rozdělení).

8.6. Při kontrole ze 100 vozidel 24 překročilo rychlost 50 km/h, průměrná rychlost byla 55 km/h, směrodatná odchylka 7 km/h. Sestrojte 95% interval spolehlivosti pro průměrnou rychlost vozidel a pro podíl vozidel překračujících rychlost.

9. Testování hypotéz

V učebním textu k PSE: Cvičení 11
V hypertextově: na webu str. 78–86 (ve verzi k tisku odst. 7.1–7.9)
V hypertextově: na webu str. 89 (ve verzi k tisku odst. 7.12)

Co se hodí vědět:

- Hypotéza je nějaké tvrzení o rozdělení základního souboru, proti sobě H_0 základní, H_1 alternativní.

- Test: na základě pozorovaných hodnot hypotézu H_0
  - buď H_0 zamítne (ve prospěch H_1), když pozorované hodnoty jsou za platnosti H_0 „málo pravděpodobné“ (kritický obor),
  - anebo H_0 nezamítne (připouštíme, že H_0 platí).

- Může nastat chyba
  - 1. druhu: H_0 platí a my ji zamítneme
  - 2. druhu: H_0 neplatí a my ji nezamítneme

  Chce se, aby za $H_0 \ldots$ P(ch. 1. druhu) $\leqq \alpha \leftarrow$ hladina významnosti

- Testy o střední hodnotě při výběru z $\mathrm{N}(\mu, \sigma^2)$ (či přibližné při výběru z jiného), $H_0\colon \mu = \mu_0$ proti $H_1 : \mu \neq \mu_0$
  - $U = \frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}$, zamítáme H_0, když $|U| > u_{1-\alpha/2}$,
  - $T = \frac{\overline{X}-\mu_0}{s}\sqrt{n}$, zamítáme H_0, když $|T| > t_{1-\alpha/2}(n-1)$,

  či $H_0\colon p = p_0$ vs. $H_1\colon p \neq p_0$ v binomickém $S \sim \mathrm{Bi}(n,p)$, $\widehat{p} = S/n$
  - $U = \frac{\widehat{p}-p_0}{\sqrt{\widehat{p}(1-\widehat{p})}}\sqrt{n}$, zamítáme H_0, když $|U| > u_{1-\alpha/2}$.

- Dvouvýběrový t-test: test o rozdílu mezi středními hodnotami dvou nezávislých výběrů
  - $X_1, \ldots, X_n$ z $\mathrm{N}(\mu_1, \sigma^2)$ a $Y_1, \ldots, Y_m$ z $\mathrm{N}(\mu_2, \sigma^2)$,
  - $H_0\colon \mu_1 - \mu_2 = d_0$ proti $H_1\colon \mu_1 - \mu_2 \neq d_0$ zamítá H_0, když

  $$T = \frac{\overline{X}-\overline{Y}-d_0}{\sqrt{(n-1)s_X^2+(m-1)s_Y^2}}\sqrt{\frac{nm(n+m-2)}{n+m}} \ldots |T| > t_{1-\alpha/2}(n+m-2)$$

!! **9.1.** U 26 náhodně vybraných studentů byly naměřeny výšky 190, 175, 168, 182, 180, 179, 200, 191, 156, 180, 178, 191, 185, 202, 182, 187,5, 166, 182, 178, 177, 176, 182, 185, 175, 182, 185 cm. Lze vyvrátit, že průměrná výška studentů je 185 cm? *(Viz Podrobně 3.3)*

9.2. Spotřeba téhož auta byla testována u 11 řidičů s výsledky 8,8, 8,9, 9,0, 8,7, 9,3, 9,0, 8,7, 8,8, 9,4, 8,6, 8,9 (l/100 km). Je pravdivá výrobcem udávaná spotřeba 8,8 l/100 km? Můžeme popřít tvrzení, že rozptyl získaných údajů je 0,1?

9.3. Při 200 hodech mincí byl rub zaznamenán 90krát. Je důvod se domnívat, že rub nepadá stejně často jako líc?

9.4. Z 42 náhodně vybraných účastníků sportovního odpoledne bylo 16 dívek a 26 chlapců. Je podíl dívek v rozporu s předpokladem, že dívky sportují stejně jako chlapci? (Víme určitě, že ne více).

9.5. Bodové zisky dosažené v písemce PSE studenty 1. a z 2. kroužku. Lze říci, že mezi studenty navštěvujícími 1. a 2. cvičení není z hlediska průměrného bodového zisku rozdíl?

9.6. Data z ankety PSE. Lze k zodpovězení otázky, zda je rozdíl mezi tím, kolik bodů studenti v průměru čekají před písemkou a kolik po písemce, použít dvouvýběrový t-test?

10. Chí-kvadrát test dobré shody

V učebním textu k PSE: Cvičení 12
V hypertextově: na webu str. 90–91 (ve verzi k tisku odst. 8.13–8.14)

Co se hodí vědět:

- Chí-kvadrát test o rozdělení náhodného výběru
  - každé pozorování zařazeno do některé z k tříd
  - z výběru rozsahu n pozorované četnosti $n_1, \dots, n_k$
  - H_0: rozdělení do tříd se řídí pravděpodobnostmi $p_1, \dots, p_k$
    H_1: jinými

- Jsou-li všechny očekávané četnosti $o_i = np_i > 5$
  - testová statistika $\chi^2 = \sum_i \frac{(n_i - o_i)^2}{o_i}$ ($\approx \chi^2_{k-1}$ při velkých n)
  - H_0 zamítáme, když $\chi^2 > \chi^2_{1-\alpha}(k-1)$ ← kritický obor jednostranný!

- Obecněji: Závisí-li rozdělení (tj. pravděpodobnosti p_i) na m neznámých parametrech... vhodný odhad parametrů... s nimi odhad očekávaných četností... počet stupňů volnosti $k - 1 - m$

!! **10.1.** Z 63 studentů získalo jednotlivé známky

známka	1	2	3	4
počet	7	25	21	10

Odpovídá to předpokladu, že jedničku získává 20 % studentů, dvojku a trojku po 30 % a čtyřku 20 % studentů? *(Viz Podrobně 3.4)*

10.2. Z 42 náhodně vybraných účastníků sportovního odpoledne bylo 16 dívek a 26 chlapců. Je podíl dívek v rozporu s předpokladem, že dívky sportují stejně jako chlapci?

10.3. Pro určitou celočíselnou náhodnou veličinu byly při 35 pozorováních zjištěny četnosti

Hodnota	0	1	2	3	4	5	více než 5
Počet	2	5	13	6	7	2	0

Může jít o veličinu s Poissonovým rozdělením?

11. Vícerozměrná veličina

V učebním textu k PSE: Cvičení 12 a odst. 11.1.6
V hypertextově: na webu str. 31–40 (ve verzi k tisku odst. 3.1–3.10)
V hypertextově: na webu str. 68–69 (ve verzi k tisku odst. 6.9–6.10)
V hypertextově: na webu str. 92–93 (ve verzi k tisku odst. 8.15–8.16)
V hypertextově: na webu str. 88 (ve verzi k tisku odst. 8.11)

Co se hodí vědět:

- Sledujeme veličiny X, Y, které spolu mohou souviset
  - sdružené rozdělení určují např. $\mathrm{P}(X \leqq x, Y \leqq y)$, $x, y \in \mathbf{R}$
  - Nezávislost: zda hodnoty X jakkoli ovlivňují hodnoty Y, u diskrétních $\equiv \forall_{ij}\, \mathrm{P}(X = i, Y = j) = \mathrm{P}(X = i)\,\mathrm{P}(Y = j)$
  - Korelační koeficient $\rho \in \langle -1, 1 \rangle$, míra „lineární“ závislosti: $\rho = 0$ není, $\rho = \pm 1$ naprostá ($Y = a + bX$ s $b \gtrless 0$)

  Platí: X, Y nezávislé $\Rightarrow \rho = 0$ (tj. $\rho \neq 0 \Rightarrow$ závislé). Opačně „$\Leftarrow$“ ne.

- Test nezávislosti v kontingenční tabulce
  - každé z n pozorování (X, Y) zařazeno do některé z $r \times s$ buněk
  - $H_0\colon X, Y$ nezávislé proti $H_1\colon X, Y$ závislé
  - χ^2-test dobré shody, pozorované n_{ij}, očekávané $o_{ij} = \frac{n_{i.}n_{.j}}{n} > 5$ (řádkové a slopcové součty)
  - $\chi^2 = \sum_{ij} \frac{(n_{ij} - o_{ij})^2}{o_{ij}}$, zamít. H_0, když $\chi^2 > \chi^2_{1-\alpha}((r-1)(s-1))$.

- Dvourozměrné normální rozdělení
  - lineární kombinace mají $N(.,.)$, navíc $\rho_{XY} = 0 \Leftrightarrow X, Y$ nezáv.
  - párový t-test o rozdílu mezi středními hodnotami veličin X a Y, $H_0\colon \mu_1 - \mu_2 = d_0$ proti $H_1\colon \mu_1 - \mu_2 \neq d_0$, $Z_i = X_i - Y_i$,
    $T = \frac{\overline{Z} - d_0}{s_Z}\sqrt{n}$, zamítáme H_0, když $|T| > t_{1-\alpha/2}(n-1)$,
  - test o koeficientu korelace $H_0\colon \rho = 0$ proti $H_1\colon \rho \neq 0$
    $T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$, zamítáme H_0, když $|T| > t_{1-\alpha/2}(n-2)$,
    kde $r = \frac{s_{XY}}{s_X s_Y}$ a $s_{XY} = \frac{1}{n-1}(\sum_{ij} X_i Y_j - n\overline{X}\overline{Y})$.

!! **11.1.** Lze z údajů o 91 059 manželstvích uzavřených v roce 1957 prokázat závislost mezi stavem ženicha a stavem nevěsty při vstupu do manželství?

	svob-á	ovd-á	rozv-á	
svob-ý	75 564	824	3 463	79 851
ovdov-ý	1 370	904	798	3 072
rozv-ý	4 603	590	2 943	8 136
	81 537	2 318	7 204	91 059

(Viz Podrobně 3.5)

11.2. Byla zjišťována souvislost mezi hladinou alkoholu v krvi (nízká, střední, vysoká) a rychlostí reakce (dobrá, špatná) u 100 náhodně vybraných lidí. Existuje souvislost?

	dobrá	špatná	
nízká	53	12	65
střední	5	15	20
vysoká	2	13	15
	60	40	100

11.3. Data z ankety PSE. Závisí to, zda posluchačům způsob výuky vyhovuje na období, kdy studují materiály?

!! **11.4.** Výrobní operace může být provedena dvěma postupy. Doba jejich trvání byla vyzkoušena u 9 pracovníků:

Pracovník	1	2	3	4	5	6	7	8	9
Postup A	214	253	276	215	238	221	210	229	269
Postup B	281	279	260	230	267	253	205	265	299

Jsou postupy stejně rychlé? *(Viz Podrobně 3.6)*

!! **11.5.** Data z předchozího příkladu. Existuje závislost mezi dobou trváním postupu A a postupu B? *(Viz Podrobně 3.7)*

11.6. Data z ankety PSE. Je souvislost mezi tím, kolik bodů studenti čekají před písemkou a kolik po písemce? Čekají studenti po písemce v průměru stejně bodů jako před písemkou?

11.7. Bodové zisky dosažené studenty v 1. a 2. písemce PSE. Je rozdíl mezi tím, kolik bodů studenti získávají v 1. a ve 2. písemce? Závisí bodový zisk ve 2. písemce na bodovém zisku v 1. písemce?

12. Regrese

V učebním textu k PSE: Cvičení 13
V hypertextově: na webu str. 95–100 (ve verzi k tisku odst. 8.1–8.6)

Co se hodí vědět:

- Zkoumá se, jakým způsobem hodnoty veličiny Y závisí na hodnotách proměnné(-ých) x. Regresní model: n pozorování

$$Y_i = f_{\beta_1,\ldots,\beta_p}(x_i) + \underset{\text{xxx}}{e_i}, \quad i = 1, \ldots, n$$

např. „přímka“ $Y_i = a + bx_i + e_i$, „exponenciela“ $Y_i = a\mathrm{e}^{bx_i} + e_i$

- Odhad metodou nejmenších čtverců: ta $\widehat{\beta}_1, \ldots \widehat{\beta}_p$, pro něž součet

$$S(\beta_1, \ldots, \beta_p) = \sum_{i=1}^{n} (Y_i - f_{\beta_1,\ldots,\beta_p}(x_i))^2 \quad \text{minimální,}$$

tj. extrémy funkce více proměnných... soustava $\partial S/\partial \beta_i = 0 \; \forall_i$ atd.

- Lineární model: speciální případ, kdy f je lineární v parametrech
  - $Y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i$
  - maticově $Y = X\beta + e$
  - odhady MNČ: $\widehat{\beta} = (X^T X)^{-1} X^T Y$ (má-li X lin.nezáv. sloupce)

Zahrnuje i parabolu, nepřímou úměru, linearizovanou exponencielu...

$$Y_i = a + bx_i + cx_i^2 + e_i, \; Y_i = a + b/x_i + e_i, \; \ln Y_i = \ln a + bx_i + e_i.$$

12.1. Byl sledován počet rozvodů za rok na 1000 obyvatel v Čechách v letech 1960–1970 (roky 0–10):

Rok	0	1	2	3	4	5	6	7	8	9	10
Rozvodů	1,34	1,45	1,47	1,52	1,48	1,66	1,77	1,76	1,89	2,08	2,19

Odhadněte parametry předpokládané kvadratické (resp. lineární, resp. exponenciální) závislosti počtu rozvodů na čase a odhadněte počet rozvodů na 1000 obyvatel v roce 1971.

12.2. V 10 podnicích se sledovala pracovní neschopnost (v procentech), průměrný věk pracovníků a podíl žen na počtu pracovníků:

neschop	věk	žen
3,1	35	40
4,0	33	44
3.5	42	40
3,0	34	38
1,9	40	30
2,0	36	32
2,5	40	35
3,0	38	36
3,5	32	40
2,5	40	35

Odhadněte parametry lineární závislosti procenta pracovní nechopnosti na průměrném věku a podílu žen.

12.3. Data z ankety PSE. Odhadněte model závislosti bodů očekávaných po písemce na bodech očekávaných před písemkou.

... Tabulky

1. Distribuční funkce Poissonova rozdělení

V tabulkách jsou hodnoty $\mathrm{P}[\mathrm{Po}(\lambda) \leqq x]$.

Pro $\lambda = 0, 1, \ldots, 0,9$.

x \ λ	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0	0,905	0,819	0,741	0,670	0,607	0,549	0,497	0,449	0,407
1	0,995	0,982	0,963	0,938	0,910	0,878	0,844	0,809	0,772
2	1,000	0,999	0,996	0,992	0,986	0,977	0,966	0,953	0,937
3	1,000	1,000	1,000	0,999	0,998	0,997	0,994	0,991	0,987
4	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,999	0,998
5	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Pro $\lambda = 1, \ldots, 9$.

x \ λ	1	2	3	4	5	6	7	8	9
0	0,368	0,135	0,050	0,018	0,007	0,002	0,001	0,000	0,000
1	0,736	0,406	0,199	0,092	0,040	0,017	0,007	0,003	0,001
2	0,920	0,677	0,423	0,238	0,125	0,062	0,030	0,014	0,006
3	0,981	0,857	0,647	0,433	0,265	0,151	0,082	0,042	0,021
4	0,996	0,947	0,815	0,629	0,440	0,285	0,173	0,100	0,055
5	0,999	0,983	0,916	0,785	0,616	0,446	0,301	0,191	0,116
6	1,000	0,995	0,966	0,889	0,762	0,606	0,450	0,313	0,207
7	1,000	0,999	0,988	0,949	0,867	0,744	0,599	0,453	0,324
8	1,000	1,000	0,996	0,979	0,932	0,847	0,729	0,593	0,456
9	1,000	1,000	0,999	0,992	0,968	0,916	0,830	0,717	0,587
10	1,000	1,000	1,000	0,997	0,986	0,957	0,901	0,816	0,706
11	1,000	1,000	1,000	0,999	0,995	0,980	0,947	0,888	0,803
12	1,000	1,000	1,000	1,000	0,998	0,991	0,973	0,936	0,876
13	1,000	1,000	1,000	1,000	0,999	0,996	0,987	0,966	0,926
14	1,000	1,000	1,000	1,000	1,000	0,999	0,994	0,983	0,959
15	1,000	1,000	1,000	1,000	1,000	0,999	0,998	0,992	0,978
16	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,996	0,989
17	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,998	0,995
18	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,999	0,998
19	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	0,999
20	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Pro $\lambda = 10, 12, 15, 20$.

x \ λ	10	12	15	20
0	0,000	0,000	0,000	0,000
1	0,000	0,000	0,000	0,000
2	0,003	0,001	0,000	0,000
3	0,010	0,002	0,000	0,000
4	0,029	0,008	0,001	0,000
5	0,067	0,020	0,003	0,000
6	0,130	0,046	0,008	0,000
7	0,220	0,090	0,018	0,001
8	0,333	0,155	0,037	0,002
9	0,458	0,242	0,070	0,005
10	0,583	0,347	0,118	0,011
11	0,697	0,462	0,185	0,021
12	0,792	0,576	0,268	0,039
13	0,864	0,682	0,363	0,066
14	0,917	0,772	0,466	0,105
15	0,951	0,844	0,568	0,157
16	0,973	0,899	0,664	0,221
17	0,986	0,937	0,749	0,297
18	0,993	0,963	0,819	0,381
19	0,997	0,979	0,875	0,470
20	0,998	0,988	0,917	0,559
21	0,999	0,994	0,947	0,644
22	1,000	0,997	0,967	0,721
23	1,000	0,999	0,981	0,787
24	1,000	0,999	0,989	0,843
25	1,000	1,000	0,994	0,888
26	1,000	1,000	0,997	0,922
27	1,000	1,000	0,998	0,948
28	1,000	1,000	0,999	0,966
29	1,000	1,000	1,000	0,978
30	1,000	1,000	1,000	0,987

2. Distribuční funkce a kvantily $\mathrm{N}(0,1)$

Distribuční funkce.

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,05	0,519 9	0,80	0,788 1	1,6	0,945 2
0,10	0,539 8	0,85	0,802 3	1,7	0,955 4
0,15	0,559 6	0,90	0,815 9	1,8	0,964 1
0,20	0,579 3	0,95	0,828 9	1,9	0,971 3
0,25	0,598 7	1,00	0,841 3	2,0	0,977 2
0,30	0,617 9	1,05	0,853 1	2,1	0,982 1
0,35	0,636 8	1,10	0,864 3	2,2	0,986 1
0,40	0,655 4	1,15	0,874 9	2,3	0,989 3
0,45	0,673 6	1,20	0,884 9	2,4	0,991 8
0,50	0,691 5	1,25	0,894 4	2,5	0,993 8
0,55	0,708 8	1,30	0,903 2	2,7	0,996 5
0,60	0,725 7	1,35	0,911 5	2,9	0,998 1
0,65	0,742 2	1,40	0,919 2	3,1	0,999 0
0,70	0,758 0	1,45	0,926 5	3,3	0,999 5
0,75	0,773 4	1,5	0,933 2	3,5	0,999 8

Platí $\Phi(-x) = 1 - \Phi(x)$.

Kvantily.

p	u_p	p	u_p	p	u_p
0,5	0	0,80	0,841 6	0,95	1,644 9
0,52	0,050 2	0,81	0,877 9	0,955	1,695 4
0,54	0,100 4	0,82	0,915 4	0,960	1,750 7
0,56	0,151 0	0,83	0,954 2	0,965	1,811 9
0,58	0,201 9	0,84	0,994 5	0,970	1,880 8
0,60	0,253 3	0,85	1,036 4	0,975	1,960 0
0,62	0,305 5	0,86	1,080 3	0,980	2,053 7
0,64	0,358 5	0,87	1,126 4	0,985	2,170 1
0,66	0,412 5	0,88	1,175 0	0,990	2,326 3
0,68	0,467 7	0,89	1,226 5	0,995	2,575 8
0,70	0,524 4	0,90	1,281 6	0,996	2,652 1
0,72	0,582 8	0,91	1,340 8	0,997	2,747 8
0,74	0,643 3	0,92	1,405 1	0,998	2,878 2
0,76	0,706 3	0,93	1,475 8	0,999	3,090 2
0,78	0,772 2	0,94	1,554 8	1	∞

Platí $u_{1-p} = -u_p$.

3. Kvantily Studentova t-rozdělení

95 % a 97,5 % kvantily $t_p(n)$.

n	$t_{0,95}(n)$	$t_{0,975}(n)$	n	$t_{0,95}(n)$	$t_{0,975}(n)$
1	6,314	12,71	16	1,746	2,120
2	2,920	4,303	17	1,740	2,110
3	2,353	3,182	18	1,734	2,101
4	2,132	2,776	19	1,729	2,093
5	2,015	2,571	20	1,725	2,086
6	1,943	2,447	21	1,721	2,080
7	1,895	2,365	22	1,717	2,074
8	1,860	2,306	23	1,714	2,069
9	1,833	2,262	24	1,711	2,064
10	1,812	2,228	25	1,708	2,060
11	1,796	2,201	26	1,706	2,056
12	1,782	2,179	27	1,703	2,052
13	1,771	2,160	28	1,701	2,048
14	1,761	2,145	29	1,699	2,045
15	1,753	2,131	30	1,697	2,042

Platí $t_{1-p}(n) = -t_p(n)$.

4. Kvantily χ^2-rozdělení

95 % a 97,5 % kvantily $\chi^2_p(n)$.

n	$\chi^2_{0,95}(n)$	$\chi^2_{0,975}(n)$	n	$\chi^2_{0,95}(n)$	$\chi^2_{0,975}(n)$
1	3,841	5,024	16	26,30	28,85
2	5,991	7,378	17	27,59	30,19
3	7,815	9,348	18	28,87	31,53
4	9,488	11,14	19	30,14	32,85
5	11,07	12,83	20	31,41	34,17
6	12,59	14,45	21	32,67	35,48
7	14,07	16,01	22	33,92	36,78
8	15,51	17,53	23	35,17	38,08
9	16,92	19,02	24	36,42	39,36
10	18,31	20,48	25	37,65	40,65
11	19,68	21,92	26	38,89	41,92
12	21,03	23,34	27	40,11	43,19
13	22,36	24,74	28	41,34	44,46
14	23,68	26,12	29	42,56	45,72
15	25,00	27,49	30	43,77	46,98

5 % a 2,5 % kvantily $\chi^2_p(n)$.

n	$\chi^2_{0,05}(n)$	$\chi^2_{0,025}(n)$	n	$\chi^2_{0,05}(n)$	$\chi^2_{0,025}(n)$
1	0,004	0,001	16	7,962	6,908
2	0,103	0,051	17	8,672	7,564
3	0,352	0,216	18	9,390	8,231
4	0,711	0,484	19	10,12	8,907
5	1,145	0,831	20	10,85	9,591
6	1,635	1,237	21	11,59	10,28
7	2,167	1,690	22	12,34	10,98
8	2,733	2,180	23	13,09	11,69
9	3,325	2,700	24	13,85	12,40
10	3,940	3,247	25	14,61	13,12
11	4,575	3,816	26	15,38	13,84
12	5,226	4,404	27	16,15	14,57
13	5,892	5,009	28	16,93	15,31
14	6,571	5,629	29	17,71	16,05
15	7,261	6,262	30	18,49	16,79

Pro $n > 30$ lze použít $\chi^2_p(n) \doteq \frac{1}{2}(\sqrt{2n-1} + u_p)^2$.

5. Kvantily Fisherova F-rozdělení

95% kvantily $F_{0,95}(m,n)$.

m \ n	1	2	3	4	5	6	8	10	15	20	25	30	60	∞
1	161	18,5	10,13	7,71	6,61	5,99	5,32	4,96	4,54	4,35	4,24	4,17	4,00	3,84
2	199	19,0	9,55	6,94	5,79	5,14	4,46	4,10	3,68	3,49	3,39	3,32	3,15	3,00
3	216	19,2	9,28	6,59	5,41	4,76	4,07	3,71	3,29	3,10	2,99	2,92	2,76	2,60
4	225	19,2	9,12	6,39	5,19	4,53	3,84	3,48	3,06	2,87	2,76	2,69	2,53	2,37
6	234	19,3	8,94	6,16	4,95	4,28	3,58	3,22	2,79	2,60	2,49	2,42	2,25	2,10
10	242	19,4	8,79	5,96	4,74	4,06	3,35	2,98	2,54	2,35	2,24	2,16	1,99	1,83
20	248	19,4	8,66	5,80	4,56	3,87	3,15	2,77	2,33	2,12	2,01	1,93	1,75	1,57
30	250	19,5	8,62	5,75	4,50	3,81	3,08	2,70	2,25	2,04	1,92	1,84	1,65	1,46
∞	254	19,5	8,53	5,63	4,36	3,67	2,93	2,54	2,07	1,84	1,71	1,62	1,39	

Platí $F_p(n,m) = 1/F_{1-p}(m,n)$.