MICHAL FRIESL

Pravděpodobnost a statistika

H Y P E R T E X T O V É

verze 2014-02-08

Obsahuje 88 hesel, v nich 638 křížových odkazů,
30 obrázků.

Pravděpodobnost a statistika hypertextově
verze 2014-02-08 PRINT

Michal Friesl
http://home.zcu.cz/~friesl/
friesl@kma.zcu.cz

Tento materiál je dostupný ve 3 variantách:

- HTML verze v podobě klikacích obrázků určená k online prohlížení přes internet,
- PDF verze pro obrazovku zaručující kvalitní zobrazení,
- PDF verze pro tisk (tato) s jinou grafickou úpravou.

Domácí stránka: http://home.zcu.cz/~friesl/hpsb/

Práce na tomto projektu nebyla podporována žádným grantem.

© Michal Friesl, 2002–2004, 2014

I když autor zpřístupňuje toto své dílo prostřednictvím internetu, **neuděluje** tím nikomu **žádná práva** k užití díla, včetně (ale nikoli výlučně) práva na rozmnožování, rozšiřování, a také zpřístupňování prostřednictvím počítačové sítě.

Užití tohoto díla pro osobní potřebu je možné **pouze** v rozsahu daném zákonem (volná užití).

OBSAH

Obsah 3
Úvod 5
1. Pravděpodobnost 7
1.1. Náhodný jev 7
1.2. Sigma-algebra 7
1.3. Pravděpodobnostní míra 7
1.4. Klasická definice pravděpodobnosti 7
1.5. Nezávislost jevů 8
1.6. Podmíněná pravděpodobnost 8
1.7. Věta o celkové pravděpodobnosti 8
1.8. Bayesova věta 9
2. Náhodná veličina 9
2.1. Náhodná veličina 9
2.2. Rozdělení 9
2.3. Distribuční funkce 9
2.4. Diskrétní rozdělení 10
2.5. Pravděpodobnostní funkce 10
2.6. Spojité rozdělení 10
2.7. Hustota 11
2.8. Intenzita poruch 11
2.9. Střední hodnota 11
2.10. Momenty 12
2.11. Rozptyl 12
2.12. Směrodatná odchylka 12
2.13. Medián 12
2.14. Kvantil 13
3. Náhodný vektor 13
3.1. Náhodný vektor 13
3.2. Sdružené rozdělení 13
3.3. Marginální rozdělení 13
3.4. Diskrétní sdružené rozdělení 13
3.5. Spojité sdružené rozdělení 14
3.6. Střední hodnota vektoru 14
3.7. Kovariance 14
3.8. Varianční matice 15
3.9. Korelační koeficient 15
3.10. Nezávislé náhodné veličiny 15
4. Některá rozdělení 15
4.1. Alternativní rozdělení 15
4.2. Binomické rozdělení 16
4.3. Poissonovo rozdělení 16
4.4. Aproximace binomického rozdělení Poissonovým 16
4.5. Exponenciální rozdělení 16
4.6. Normální rozdělení 17
4.7. Vícerozměrné normální rozdělení 17
4.8. Chí-kvadrát rozdělení 18
4.9. Studentovo t-rozdělení 18
4.10. Fisherovo-Snedecorovo F-rozdělení 18
5. Limitní věty 18
5.1. Zákon velkých čísel 18
5.2. Kolmogorovův zákon velkých čísel 19
5.3. Slabý zákon velkých čísel 19
5.4. Centrální limitní věta 19
5.5. Ljapunovova centrální limitní věta 20
5.6. Centrální limitní věta pro binomické rozdělení 20
6. Teorie odhadu 20
6.1. Náhodný výběr 20

6.2. Statistika (hodnota) 20
6.3. Odhad 20
6.4. Popisné statistiky 20
6.5. Výběrový průměr 21
6.6. Výběrový medián 21
6.7. Výběrový rozptyl 21
6.8. Výběrová směrodatná odchylka 21
6.9. Výběrová kovariance 22
6.10. Výběrový korelační koeficient 22
6.11. Nestrannost 22
6.12. Konzistence 22
6.13. Střední čtvercová chyba 22
6.14. Nejlepší odhad 23
6.15. Maximálně věrohodný odhad 23
6.16. Intervalové odhady 23
6.17. Intervaly spolehlivosti v normálním rozdělení 23
7. Testování hypotéz 24
7.1. Hypotéza 24
7.2. Test 24
7.3. Kritický obor 24
7.4. Chyba 1. a 2. druhu 24
7.5. Hladina významnosti 25
7.6. Síla testu 25
7.7. P-hodnota 25
7.8. Testy střední hodnoty normálního rozdělení 25
7.9. Test o střední hodnotě 26
7.10. Test o rozptylu normálního rozdělení 26
7.11. Párový test 27
7.12. Dvouvýběrový test 27
7.13. Chí-kvadrát test dobré shody 28
7.14. Chí-kvadrát test při neznámých parametrech 28
7.15. Test nezávislosti v kontingenční tabulce 29
7.16. Test nezávislosti v normálním rozdělení 29
8. Regrese 30
8.1. Regresní model 30
8.2. Lineární regrese 30
8.3. Metoda nejmenších čtverců 31
8.4. Reziduální součet čtverců 31
8.5. Odhady v regresi 31
8.6. Koeficient determinace 31
8.7. Test významnosti regrese 32
Seznam hesel 33
Seznam obrázků 35
Index 36

ÚVOD

Snad mohu říci, že na českém internetu není nadbytek materiálů z oblasti pravděpodobnosti a statistiky a mnoho z těch, které existují, nevyužívá dostatečně hypertextových možností formátů, ve kterých jsou prezentovány (pojem *hypertextově* používám, možná trochu nepřesně, ve smyslu přítomnosti křížových odkazů). *Pravděpodobnost a statistika hypertextově* je mým pokusem o vytvoření takového „pořádně hypertextového materiálu“.

Text sleduje výklad z úvodních kurzů pravděpodobnosti a statistiky, nemá nicméně nahrazovat klasický učební text. Obsahuje definice pojmů a tvrzení o nich (i když ne explicitně označené), ne už ale podrobnější vysvětlování a komentáře. Chybí také příklady (viz ovšem sbírku pseudořešených příkladů *Posbírané příklady z pravděpodobnosti a statistiky*). Text se svým uspořádáním spíše podobá jakési encyklopedii a může posloužit třeba těm, kteří si potřebují rychle připomenout některé z pojmů, vztahů, poznatků.

Oproti první verzi, zveřejněné v září 2002, jsem v roce 2003 opravil některé chyby, doplnil 22 hesel, přidal obrázky, v roce 2004 opravil další drobné chyby a doplnil jedno chybějící heslo, v této verzi jsem opravil další překlepy. Materiál je dostupný ve třech podobách — HTML verze v podobě klikacích obrázků, PDF verze pro obrazovku a PDF verze k tisku — spolu s dalšími informacemi z internetové adresy

http://home.zcu.cz/~friesl/hpsb/

Hlášení o chybách, připomínky i náměty samozřejmě rád přijmu.

Michal Friesl
Plzeň, září 2002 – únor 2014 friesl@kma.zcu.cz

1. PRAVDĚPODOBNOST

1.1. *Náhodný jev.* Je-li dána množina Ω (všech výsledků **náhodného pokusu**, tj. pokusu, jehož výsledek není jednoznačně určen podmínkami, za kterých je prováděn), pak **náhodným jevem** (v Ω) nazýváme každou podmnožinu množiny Ω.

Jev Ω nazýváme **jistý**, jev $\emptyset$ **nemožný**. Jednotlivé výsledky $\omega \in \Omega$ se nazývají **elementární jevy**.

Jev označuje určitou vlastnost výsledků pokusu. Pro jevy $A, B \subset \Omega$ interpretujeme

- $\omega \in A$ — jev A nastal,
- $\omega \in \Omega \setminus A$ — jev A nenastal,
- $\omega \in A \cup B$ — aspoň jeden z jevů (A nebo B) nastal,
- $\omega \in A \cap B$ — oba dva jevy (A i B) nastaly,
- $A \subset B$ — nastává-li A, nastává i B,
- $A \cap B = \emptyset$ — jevy A a B jsou **neslučitelné** (nenastávají současně).

Ne všechny jevy v Ω musí být pozorovatelné a ne všem musí být možné rozumně připsat PRAVDĚPODOBNOST[1.3].

◇ Viz též: SIGMA-ALGEBRA[1.2], PRAVDĚPODOBNOSTNÍ MÍRA[1.3], KLASICKÁ DEFINICE PRAVDĚPODOBNOSTI[1.4], NÁHODNÁ VELIČINA[2.1]

1.2. *Sigma-algebra.* Systém $\mathcal{A}$ podmnožin množiny Ω se nazývá **σ-algebra**, pokud

- $\Omega \in \mathcal{A}$,
- $A \in \mathcal{A} \implies \Omega \setminus A \in \mathcal{A}$,
- $A_1, A_2, \dots \in \mathcal{A} \implies \bigcup A_i \in \mathcal{A}$,

tedy pokud v $\mathcal{A}$ je zahrnuta množina Ω, s každou množinou je v $\mathcal{A}$ i její doplněk a s každými spočetně mnoha množinami i jejich sjednocení.

Můžeme si představovat, že σ-algebra představuje souhrn informací, které o výsledcích NÁHODNÉHO POKUSU[1.1] máme — ne všechny JEVY[1.1] v Ω musí být totiž pozorovatelné.

1.3. *Pravděpodobnostní míra.* Reálná funkce P definovaná na σ-ALGEBŘE[1.2] (přiměřeně bohatém systému) $\mathcal{A}$ JEVŮ[1.1] v Ω se nazývá **pravděpodobnostní míra**, pokud

- P má hodnoty v $\langle 0, 1\rangle$,
- $P(\Omega) = 1$,
- pro disjunktní množiny (neslučitelné jevy) $A_1, A_2, \dots \in \mathcal{A}$ je $P(\bigcup A_i) = \sum P(A_i)$.

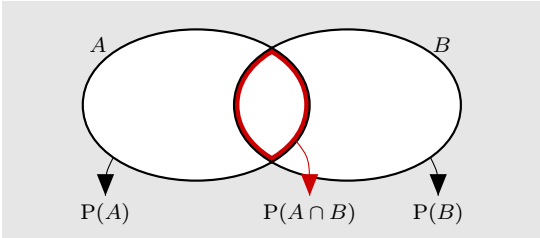
Trojice $(\Omega, \mathcal{A}, P)$ se nazývá **pravděpodobnostní prostor**.

Z těchto vlastností lze odvodit, že $P(\Omega \setminus A) = 1 - P(A)$, dále $P(\emptyset) = 0$. Pro libovolné jevy $A, B \in \mathcal{A}$ lze psát

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

pro NESLUČITELNÉ[1.1] je poslední člen nulový. Když $B \subset A$, máme $P(A \setminus B) = P(A) - P(B)$.

◇ Viz též: KLASICKÁ DEFINICE PRAVDĚPODOBNOSTI[1.4], ROZDĚLENÍ[2.2], ZÁKON VELKÝCH ČÍSEL[5.1]



Pravděpodobnost sjednocení

1.4. *Klasická definice pravděpodobnosti.* Je-li množina výsledků pokusu $\Omega = \{\omega, \dots \omega_n\}$ konečná (a neprázdná) a všechny elementární jevy $\omega_1, \dots, \omega_n$ mají stejnou pravděpodobnost, že nastanou (rovnou $1/n$), můžeme pro všechny JEVY[1.1] $A \subset \Omega$ použít **klasickou definici pravděpodobnosti**

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|}.$$

Ta splňuje požadavky kladené na PRAVDĚPODOBNOSTNÍ MÍRU[1.3], za σ-ALGEBRU[1.2] jevů v tomto případě můžeme vzít σ-algebru všech podmnožin Ω.

Klasickou definici nelze použít, pokud všechny elementární jevy nejsou stejně pravděpodobné nebo jich je nekonečně (spočetně, či dokonce nespočetně) mnoho.

1.5. *Nezávislost jevů.* Řekneme, že JEVY[1.1] ze systému $\mathcal{C}$ jsou **nezávislé**, jestliže pro PRAVDĚPODOBNOSTI[1.3] každého konečného počtu k jevů $\{A_1, \dots, A_k\} \subset \mathcal{C}$ platí

$$\mathrm{P}(\bigcap_i A_i) = \prod_i \mathrm{P}(A_i).$$

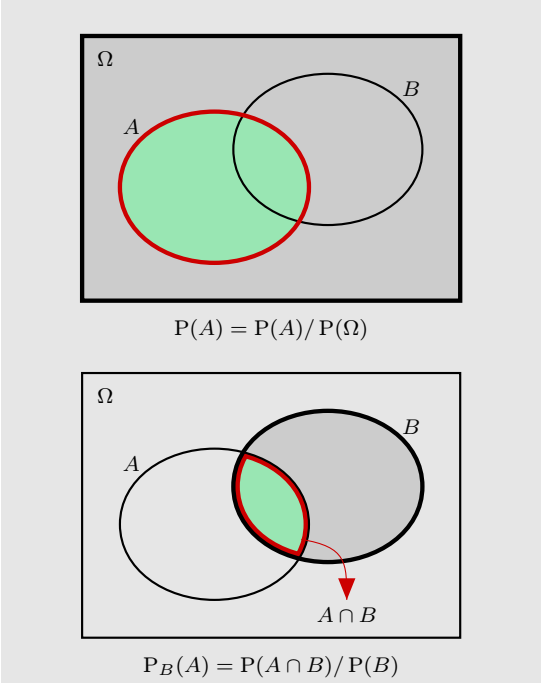
V opačném případě jevy nazýváme **závislé**.

Nezávislost jevů znamená, že výskyt jedněch neposkytuje žádnou dodatečnou informaci k pravděpodobnosti výskytu ostatních.

Dva jevy A a B jsou tedy nezávislé, pokud $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$. Jev C s $\mathrm{P}(C) = 0$ nebo $\mathrm{P}(C) = 1$ je nezávislý s kterýmkoli jiným jevem.

Nahradíme-li v $\mathcal{C}$ některé jevy jejich doplňky, nezávislost (závislost) zůstává zachována.

◇ Viz též: NEZÁVISLÉ NÁHODNÉ VELIČINY[3.10], PODMÍNĚNÁ PRAVDĚPODOBNOST[1.6]



Podmíněná a nepodmíněná pravděpodobnost

1.6. *Podmíněná pravděpodobnost.* **Podmíněná pravděpodobnost** jevu A za podmínky, že nastal jev B (s $\mathrm{P}(B) > 0$) je definována jako

$$\mathrm{P}_B(A) = \mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)},$$

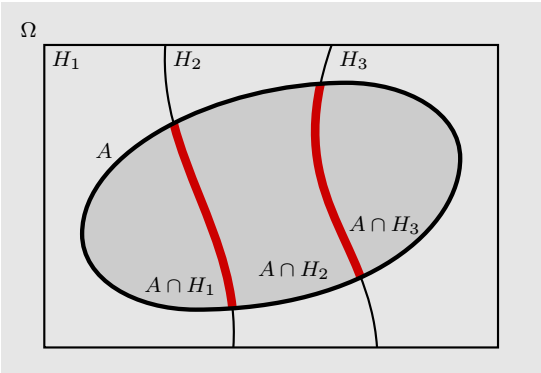
kde P značí (nepodmíněnou) PRAVDĚPODOBNOST[1.3].

Podmíněná pravděpodobnost P_B je opět PRAVDĚPODOBNOSTNÍ MÍROU[1.3].

Pro jevy A, B s kladnými pravděpodobnostmi je ekvivalentní

- A, B jsou NEZÁVISLÉ[1.5]
- $\mathrm{P}_B(A) = \mathrm{P}(A)$
- $\mathrm{P}_A(B) = \mathrm{P}(B)$.

◇ Viz též: BAYESOVA VĚTA[1.8], VĚTA O CELKOVÉ PRAVDĚPODOBNOSTI[1.7]



Rozklad pravděpodobnosti jevu

1.7. *Věta o celkové pravděpodobnosti.* Tvoří-li JEVY[1.1] $H_1, H_2, \dots$ rozklad pravděpodobnostního prostoru (jsou navzájem disjunktní, $\sum_i \mathrm{P}(H_i) = 1$, $\mathrm{P}(H_i) > 0$), pak pro každý jev A

$$\mathrm{P}(A) = \sum_i \underbrace{\mathrm{P}(A \mid H_i)\,\mathrm{P}(H_i)}_{\mathrm{P}(A \cap H_i)}.$$

Známe-li (PODMÍNĚNÉ[1.6]) pravděpodobnosti $\mathrm{P}(A \mid H_i)$ jevu A za hypotéz H_i a známe-li PRAVDĚPODOBNOSTI[1.3] $\mathrm{P}(H_i)$, s jakými tyto hypotézy nastávají, umíme spočítat celkovou pravděpodobnost jevu A.

◇ Viz též: BAYESOVA VĚTA[1.8], PODMÍNĚNÁ PRAVDĚPODOBNOST[1.6]

1.8. *Bayesova věta.* Tvoří-li JEVY[1.1] $H_1, H_2, \ldots$ rozklad pravděpodobnostního prostoru (jsou navzájem disjunktní, $\sum_i \mathrm{P}(H_i) = 1$, $\mathrm{P}(H_i) > 0$), a A je jev s $\mathrm{P}(A) > 0$, pak pro každé k

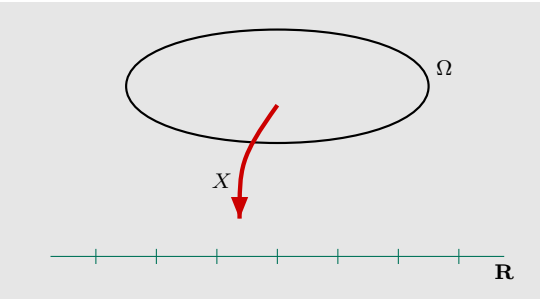
$$\mathrm{P}(H_k \mid A) = \frac{\mathrm{P}(A \mid H_k)\,\mathrm{P}(H_k)}{\sum_i \mathrm{P}(A \mid H_i)\,\mathrm{P}(H_i)}.$$

Nastal-li jev A, můžeme na základě této informace opravit (aktualizovat) apriorní PRAVDĚPODOBNOSTI[1.3] $\mathrm{P}(H_k)$ na aposteriorní $\mathrm{P}(H_k \mid A)$.

Ve jmenovateli zlomku je CELKOVÁ PRAVDĚPODOBNOST[1.7] jevu A.

◇ Viz též: PODMÍNĚNÁ PRAVDĚPODOBNOST[1.6], VĚTA O CELKOVÉ PRAVDĚPODOBNOSTI[1.7]

2. NÁHODNÁ VELIČINA



Náhodná veličina

2.1. *Náhodná veličina.* Náhodné veličiny jsou funkce definované na nějakém prostoru, obvykle PRAVDĚPODOBNOSTNÍM[1.3].

Přesněji, funkce $X\colon \Omega \to \mathbf{R}$ definovaná na prostoru Ω se σ-ALGEBROU[1.2] $\mathcal{A}$ se nazývá (reálná) **náhodná veličina**, jestliže

$$X^{-1}(B) = \{\omega; X(\omega) \in (a, b\rangle\} \in \mathcal{A}$$

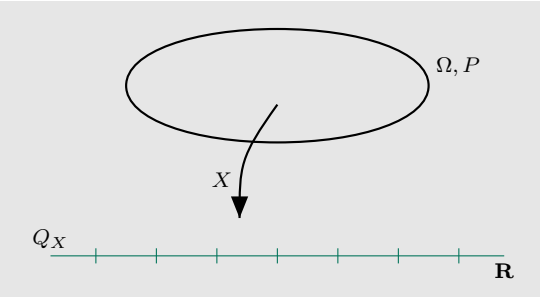
pro libovolné $-\infty \leqq a < b \leqq \infty$. Tento technický požadavek — **měřitelnost** funkce X — zajišťuje, že vzorům rozumných množin z $\mathbf{R}$ bude možné přiřadit PRAVDĚPODOBNOST[1.3].

Používáme označení

$$[X \in B] := \{\omega \in \Omega; X(\omega) \in B\} = X^{-1}(B)$$

pro vzor množiny $B \subset \mathbf{R}$ při zobrazení X.

◇ Viz též: NÁHODNÝ VEKTOR[3.1], ROZDĚLENÍ[2.2]



Přenos pravděpodobnosti

2.2. *Rozdělení.* **Rozdělením** reálné NÁHODNÉ VELIČINY[2.1] X definované na PRAVDĚPODOBNOSTNÍM PROSTORU[1.3] $(\Omega, \mathcal{A}, P)$ nazýváme PRAVDĚPODOBNOSTNÍ MÍRU[1.3] Q

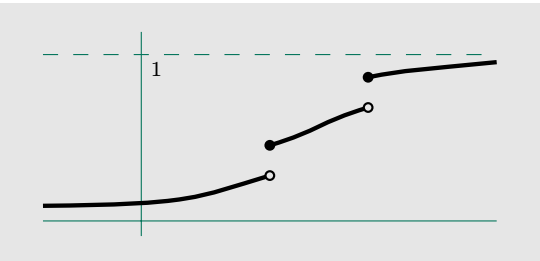
$$Q(B) = P[X \in B], \quad B \in \mathcal{B}(\mathbf{R}),$$

působící na vhodném systému podmnožin $\mathbf{R}$, tzv. **Borelovské σ-algebře** $\mathcal{B}(\mathbf{R})$. Jde o nejmenší σ-ALGEBRU[1.2] v $\mathbf{R}$ obsahující všechny intervaly.

Náhodná veličina X přenáší pravděpodobnost P z Ω na $\mathbf{R}$, Q je „obrazem“ míry P. Ke každé pravděpodobnostní míře Q na $\mathbf{R}$ existuje náhodná veličina, jejímž je rozdělením.

Rozdělení popisuje pravděpodobnostní chování náhodné veličiny (při dané pravděpodobnostní míře P), nikoliv veličinu samotnou. Můžeme tak mít dvě různé veličiny, na stejném nebo na různých pravděpodobnostních prostorech, které mají stejné rozdělení. Změníme-li P, může se rozdělení veličiny X změnit.

◇ Viz též: SPOJITÉ ROZDĚLENÍ[2.6], DISKRÉTNÍ ROZDĚLENÍ[2.4], SDRUŽENÉ ROZDĚLENÍ[3.2], MARGINÁLNÍ ROZDĚLENÍ[3.3], CHÍ-KVADRÁT TEST DOBRÉ SHODY[7.13], NĚKTERÁ ROZDĚLENÍ[4]



Distribuční funkce

2.3. *Distribuční funkce.* Funkce $F\colon \mathbf{R} \to \langle 0, 1\rangle$ se nazývá **distribuční**, je-li neklesající, zprava spojitá a s limitami $F(-\infty) = 0$, $F(\infty) = 1$.

Distribuční funkcí ROZDĚLENÍ[2.2] Q na $\mathbf{R}$, resp. (reálné) NÁHODNÉ VELIČINY[2.1] X, rozumíme funkci

$$F(x) = Q((-\infty, x\rangle), \qquad \text{resp.} \quad F(x) = \mathrm{P}[X \leqq x], \quad x \in \mathbf{R}.$$

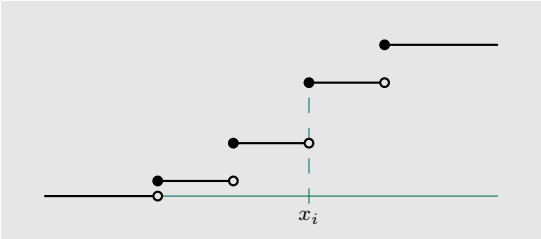
Pomocí distribuční funkce můžeme zapisovat PRAVDĚPODOBNOST[1.3], že hodnoty veličiny X padnou do intervalu $(a, b\rangle \subset \mathbf{R}$, resp. že X nabude právě hodnoty $a \in \mathbf{R}$, jako

$$\mathrm{P}[X \in (a, b\rangle] = F(b) - F(a), \qquad \text{resp.} \quad \mathrm{P}[X = a] = F(a) - F(a-),$$

kde $F(a-)$ značí limitu F v bodě a zleva.

Každá distribuční funkce popisuje nějaké rozdělení na $\mathbf{R}$ a je distribuční funkcí nějaké náhodné veličiny.

◇ Viz též: PRAVDĚPODOBNOSTNÍ FUNKCE[2.5], HUSTOTA[2.7], INTENZITA PORUCH[2.8], KVANTIL[2.14], SDRUŽENÉ ROZDĚLENÍ[3.2]



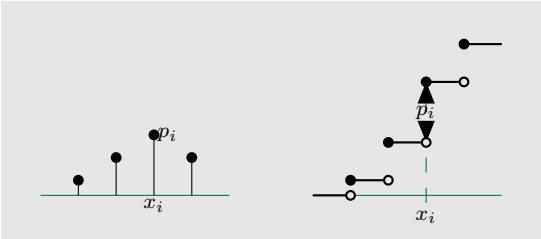
Distribuční funkce diskrétního rozdělení

2.4. *Diskrétní rozdělení.* NÁHODNÁ VELIČINA[2.1] X má **diskrétní rozdělení**, existuje-li spočetná množina $M = \{x_1, x_2, \dots\} \subset \mathbf{R}$ (nosič rozdělení) taková, že

$$\mathrm{P}[X \in M] = 1.$$

DISTRIBUČNÍ FUNKCE[2.3] takového rozdělení je schodovitá, se skoky v bodech x_i, $i = 1, 2, \dots$

◇ Viz též: PRAVDĚPODOBNOSTNÍ FUNKCE[2.5], DISKRÉTNÍ SDRUŽENÉ ROZDĚLENÍ[3.4], NĚKTERÁ ROZDĚLENÍ[4]



Pravděpodobnostní a distribuční funkce

2.5. *Pravděpodobnostní funkce.* **Pravděpodobnostní funkcí** rozumíme nezápornou funkci nulovou všude s výjimkou spočetně bodů $x_1, x_2, \dots \in \mathbf{R}$, v nichž nabývá postupně hodnot $p_1, p_2, \dots \in (0, 1\rangle$, přičemž $\sum p_i = 1$.

Je-li dáno DISKRÉTNÍ ROZDĚLENÍ[2.4] Q na $\mathbf{R}$, jeho pravděpodobnostní funkcí rozumíme funkci

$$P(x) = Q(\{x\}) = \mathrm{P}[X = x] = F(x) - F(x-) = \begin{cases} p_i & x = x_i, \quad i = 1, 2, \dots, \\ 0 & x \neq x_i, \end{cases}$$

kde F je DISTRIBUČNÍ FUNKCE[2.3] rozdělení Q, x_i jsou body skoků F a p_i velikosti skoků v těchto bodech.

Pro NÁHODNOU VELIČINU[2.1] X s takovým rozdělením je

$$\mathrm{P}[X = x_i] = p_i, \quad i = 1, 2, \dots$$
$$\mathrm{P}[X \in B] = \sum_{i, x_i \in B} p_i, \quad B \subset \mathbf{R}.$$

Každou pravděpodobnostní funkcí (tj. dvojicemi (x_i, p_i), $i = 1, 2, \dots$, kde $x_i \in \mathbf{R}$, $p_i \in (0, 1\rangle$, $\sum p_i = 1$) je popsáno nějaké diskrétní rozdělení. Pro jeho distribuční funkci platí

$$F(x) = \sum_{i; x_i \leqq x} p_i, \qquad x \in \mathbf{R}.$$

◇ Viz též: HUSTOTA[2.7]

2.6. *Spojité rozdělení.* NÁHODNÁ VELIČINA[2.1] X s DISTRIBUČNÍ FUNKCÍ[2.3] F má **spojité rozdělení**, existuje-li funkce $f(x)$ taková, že

$$F(x) = \mathrm{P}[X \leqq x] = \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \quad x \in \mathbf{R}.$$

Distribuční funkce takového rozdělení je (v řeči matematické analýzy) absolutně spojitá.

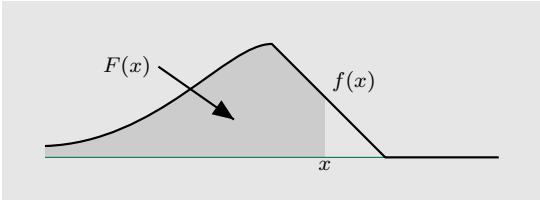
◇ Viz též: HUSTOTA[2.7], DISKRÉTNÍ ROZDĚLENÍ[2.4], SPOJITÉ SDRUŽENÉ ROZDĚLENÍ[3.5], NĚKTERÁ ROZDĚLENÍ[4]

2.7. *Hustota.* Jako **hustotu** označujeme každou funkci $f(x)$ s vlastnostmi

$$f(x) \geqq 0, \quad x \in \mathbf{R}, \qquad \int_{\mathbf{R}} f(x)\,\mathrm{d}x = 1.$$

Funkci f nazýváme hustotou SPOJITÉHO ROZDĚLENÍ[2.6] s DISTRIBUČNÍ FUNKCÍ[2.3] F, pokud

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \quad x \in \mathbf{R}.$$



Hustota rozdělení

Taková f má vlastnosti hustoty a je $f(x) = F'(x)$ pro „skoro všechna“ $x \in \mathbf{R}$ (a např. v bodech, kde F' neexistuje, můžeme f volit libovolně).

Pro VELIČINU[2.1] X s hustotou f platí

$$\mathrm{P}[X = x] = 0, \quad x \in \mathbf{R},$$

a je-li I interval s krajními body $a < b$ (ať už uzavřený, polouzavřený, nebo otevřený), pak

$$\mathrm{P}[X \in I] = F(b) - F(a) = \int_a^b f(x)\,\mathrm{d}x.$$

Každá hustota f je hustotou nějakého ROZDĚLENÍ[2.2], příslušná distribuční funkce má tvar

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t, \quad x \in \mathbf{R}.$$

◇ Viz též: PRAVDĚPODOBNOSTNÍ FUNKCE[2.5], SPOJITÉ SDRUŽENÉ ROZDĚLENÍ[3.5]

2.8. *Intenzita poruch.* Pro nezápornou NÁHODNOU VELIČINU[2.1] X se SPOJITÝM ROZDĚLENÍM[2.6] definujeme **intenzitu poruch**

$$\lambda(x) = \lim_{h \to 0+} \frac{\mathrm{P}[X \in (x, x+h\rangle \mid X > x]}{h} = \frac{f(x)}{1 - F(x)}$$

tam, kde $F(x) < 1$, přičemž f a F jsou HUSTOTA[2.7] a DISTRIBUČNÍ FUNKCE[2.3] X.

Představuje-li X dobu do poruchy nějakého zařízení, intenzita $\lambda(x)$ vyjadřuje, že pokud do času x nedošlo k poruše, tak PRAVDĚPODOBNOST[1.3], že k ní dojde v následujícím okamžiku malé délky h, je přibližně $\lambda(x)h$.

Intenzita charakterizuje ROZDĚLENÍ[2.2] nezáporné náhodné veličiny, lze psát

$$F(x) = 1 - \exp\Big(-\int_0^x \lambda(t)\,\mathrm{d}t\Big), \quad x > 0.$$

◇ Viz též: EXPONENCIÁLNÍ ROZDĚLENÍ[4.5]

2.9. *Střední hodnota.* **Střední hodnota** NÁHODNÉ VELIČINY[2.1] X, resp. $g(X)$, kde $g : \mathbf{R} \to \mathbf{R}$ je (MĚŘITELNÁ[2.1]) funkce, je

- pro X DISKRÉTNĚ ROZDĚLENOU[2.4] s PRAVDĚPODOBNOSTNÍ FUNKCÍ[2.5] P definována jako

$$\mathrm{E}\,X = \sum_i x_i P(x_i),$$

$$\text{resp.} \quad \mathrm{E}\,g(X) = \sum_i g(x_i) P(x_i),$$

- pro X SPOJITĚ ROZDĚLENOU[2.6] s HUSTOTOU[2.7] f jako

$$\mathrm{E}\,X = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x,$$

$$\text{resp.} \quad \mathrm{E}\,g(X) = \int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x,$$

mají-li výrazy napravo smysl.

Střední hodnota $\mathrm{E}\,X$ představuje průměrnou hodnotu (ve smyslu váženého průměru), které veličina X nabývá.

Střední hodnota je lineární operátor. Jsou-li X, Y náhodné veličiny definované na stejném PRAVDĚPODOBNOSTNÍM PROSTORU[1.3] Ω a $a, b \in \mathbf{R}$, platí

$$\mathrm{E}\,a = a, \qquad \mathrm{E}(aX + bY) = a\,\mathrm{E}\,X + b\,\mathrm{E}\,Y.$$

Operátor E také zachovává monotonii. Je-li $X \leqq Y$ (tj. $X(\omega) \leqq Y(\omega)\ \forall_{\omega\in\Omega}$), pak $\mathrm{E}\,X \leqq \mathrm{E}\,Y$.

◇ Viz též: MOMENTY[2.10], ROZPTYL[2.11], STŘEDNÍ HODNOTA VEKTORU[3.6], MEDIÁN[2.13], VÝBĚROVÝ PRŮMĚR[6.5], TEST O STŘEDNÍ HODNOTĚ[7.9]

2.10. *Momenty.* **Momenty** ROZDĚLENÍ[2.2] jsou STŘEDNÍ HODNOTY[2.9] mocnin příslušné NÁHODNÉ VELIČINY[2.1].

$\mathrm{E}\,X^k$ je k-tý (necentrální) moment rozdělení veličiny X, $\mathrm{E}(X - \mathrm{E}\,X)^k$ je k-tý centrální moment ($k \in \mathbf{N}$).

◇ Viz též: ROZPTYL[2.11]

2.11. *Rozptyl.* **Rozptyl** (ROZDĚLENÍ[2.2]) VELIČINY[2.1] X je definován jako druhý CENTRÁLNÍ MOMENT[2.10] rozdělení X,

$$\operatorname{var} X = \mathrm{E}(X - \mathrm{E}\,X)^2 = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2,$$

pokud je $\mathrm{E}\,X^2$ konečná.

Pro $c \in \mathbf{R}$ platí

$$\operatorname{var} c = 0, \quad \operatorname{var} cX = c^2 \operatorname{var} X.$$

Rozptyl vyjadřuje variabilitu rozdělení kolem jeho STŘEDNÍ HODNOTY[2.9]. Je-li $\mathrm{E}\,X^2 < \infty$, **Čebyševova nerovnost** odhaduje (velice hrubě) PRAVDĚPODOBNOST[1.3] odchylky X od její střední hodnoty

$$\mathrm{P}[|X - \mathrm{E}\,X| \geqq \varepsilon] \leqq \frac{\operatorname{var} X}{\varepsilon^2}, \quad \varepsilon > 0.$$

◇ Viz též: VARIANČNÍ MATICE[3.8], VÝBĚROVÝ ROZPTYL[6.7], TEST O ROZPTYLU NORMÁLNÍHO ROZDĚLENÍ[7.10]

2.12. *Směrodatná odchylka.* **Směrodatná odchylka** je druhá odmocnina z ROZPTYLU[2.11].

◇ Viz též: ROZPTYL[2.11], VÝBĚROVÁ SMĚRODATNÁ ODCHYLKA[6.8]

2.13. *Medián.* **Medián** ROZDĚLENÍ[2.2] veličiny X je libovolné číslo $\widetilde{x} \in \mathbf{R}$ splňující

$$\mathrm{P}[X < \widetilde{x}] \leqq 1/2 \leqq \mathrm{P}[X \leqq \widetilde{x}].$$

Jde o jakousi „prostřední“ hodnotu rozdělení, podobně jako STŘEDNÍ HODNOTA[2.9] je měřítkem polohy rozdělení.

Pro SPOJITĚ ROZDĚLENOU[2.6] veličinu s DISTRIBUČNÍ FUNKCÍ[2.3] F je mediánem $\widetilde{x}$ číslo splňující $F(\widetilde{x}) = 1/2$.

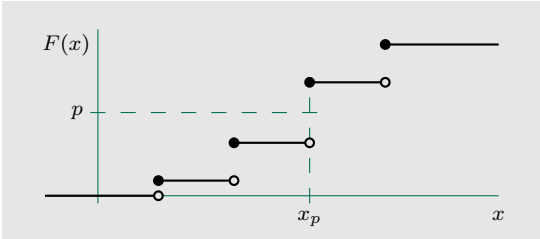
◇ Viz též: KVANTIL[2.14], VÝBĚROVÝ MEDIÁN[6.6]

2.14. *Kvantil.* Pro $p \in (0,1)$ je **p-kvantil** x_p ROZDĚLENÍ[2.2] s DISTRIBUČNÍ FUNKCÍ[2.3] F definován jako číslo splňující

$$F(x_p-) \leqq p \leqq F(x_p).$$

Je-li F spojitá, p-kvantilem jsou hodnoty $F^{-1}(p)$.

Pro $p = 1/2$, resp. $p = 1/4, 3/4$ hovoříme o MEDIÁNU[2.13], resp. dolním a horním kvartilu.



Kvantil

3. NÁHODNÝ VEKTOR

3.1. *Náhodný vektor.* **Náhodným vektorem** rozumíme vektor složený z NÁHODNÝCH VELIČIN[2.1] definovaných na stejném PRAVDĚPODOBNOSTNÍM PROSTORU[1.3], $\boldsymbol{X} = (X_1, \dots, X_n)$.

Náhodný vektor $\boldsymbol{X}$ můžeme chápat také jako náhodnou veličinu s hodnotami v $\mathbf{R}^n$.

◇ Viz též: SDRUŽENÉ ROZDĚLENÍ[3.2]

3.2. *Sdružené rozdělení.* **Sdružené rozdělení** NÁHODNÉHO VEKTORU[3.1] $\boldsymbol{X} = (X_1, \dots, X_n)$ s hodnotami v $\mathbf{R}^n$ je charakterizováno **sdruženou distribuční funkcí**

$$F(\boldsymbol{x}) = \mathrm{P}[X_1 \leqq x_1, \dots, X_n \leqq x_n], \qquad \boldsymbol{x} = (x_1, \dots, x_n) \in \mathbf{R}^n.$$

◇ Viz též: MARGINÁLNÍ ROZDĚLENÍ[3.3], DISKRÉTNÍ SDRUŽENÉ ROZDĚLENÍ[3.4], SPOJITÉ SDRUŽENÉ ROZDĚLENÍ[3.5]

3.3. *Marginální rozdělení.* **Marginálním rozdělením** NÁHODNÉHO VEKTORU[3.1] $\boldsymbol{X} = (X_1, \dots, X_n)$ rozumíme ROZDĚLENÍ[2.2] jeho složek.

Má-li vektor $\boldsymbol{X}$ SDRUŽENOU DISTRIBUČNÍ FUNKCI[3.2] $F(x_1, \dots, x_n)$, pak marginální distribuční funkce j-té složky X_j je

$$\begin{aligned} F_j(x_j) &= \mathrm{P}[X_j \leqq x_j] \\ &= \lim_{x_i \to \infty, i \neq j} F(x_1, \dots, x_n), \quad x_j \in \mathbf{R}, \end{aligned}$$

tj. v $F(x_1, \dots, x_n)$ všechny ostatní složky kromě j-té posíláme k ∞.

Obecněji, marginální rozdělení podvektoru $(X_{i_1}, \dots, X_{i_k})$, $i_1, \dots, i_k \in \{1, \dots, n\}$, $k < n$, má distribuční funkci

$$\begin{aligned} F_{i_1,\dots,i_k}(x_{i_1}, \dots, x_{i_k}) &= \mathrm{P}[X_{i_1} \leqq x_{i_1}, \dots, X_{i_k} \leqq x_{i_k}] \\ &= \lim_{x_i \to \infty, i \neq i_1,\dots,i_k} F(x_1, \dots, x_n), \quad x_{i_1}, \dots, x_{i_k} \in \mathbf{R}. \end{aligned}$$

◇ Viz též: NEZÁVISLÉ NÁHODNÉ VELIČINY[3.10]

3.4. *Diskrétní sdružené rozdělení.* NÁHODNÝ VEKTOR[3.1] $\boldsymbol{X}$ má **diskrétní rozdělení**, pokud existuje spočetná množina bodů $M = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots\}$ taková, že $\mathrm{P}[\boldsymbol{X} \in M] = 1$.

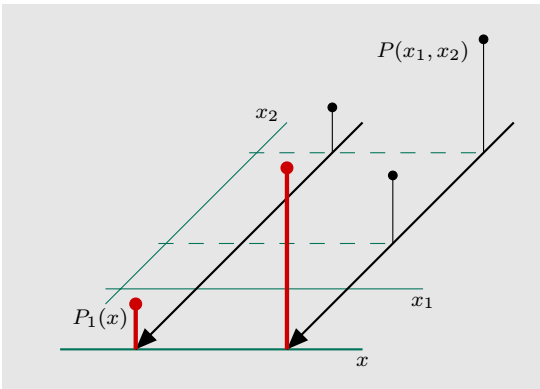
Jako pravděpodobnostní funkci takového vektoru označujeme funkci

$$P(\boldsymbol{x}) = \begin{cases} \mathrm{P}[\boldsymbol{X} = \boldsymbol{x}], & \boldsymbol{x} \in M \\ 0, & \boldsymbol{x} \notin M. \end{cases}$$

MARGINÁLNÍ ROZDĚLENÍ[3.3] diskrétně rozděleného vektoru $\boldsymbol{X}$ jsou DISKRÉTNÍ[2.4], PRAVDĚPODOBNOSTNÍ FUNKCE[2.5] j-té složky je

$$P_j(x) = \mathrm{P}[X_j = x] = \sum_{\boldsymbol{x} \in M; x_j = x} \cdots \sum P(\boldsymbol{x}), \quad x \in \mathbf{R},$$

tj. při daném x sčítáme sdruženou pravděpodobnostní funkci



Sdružená a marginální pravděpodobnostní funkce

přes všechny hodnoty ostatních složek.

◇ Viz též: Spojité sdružené rozdělení[3.5]

3.5. *Spojité sdružené rozdělení.* Náhodný vektor[3.1] $\boldsymbol{X} = (X_1, \dots, X_n)$ se sdruženou[3.2] distribuční funkcí $F(x_1, \dots, x_n)$ má **spojité rozdělení**, pokud existuje nezáporná funkce n proměnných f, že

$$F(x_1, \dots, x_n) = \int_{(-\infty, x_1\rangle \times \dots \times (-\infty, x_n\rangle} \cdots \int f(t_1, \dots, t_n)\, \mathrm{d}t_1 \dots \mathrm{d}t_n$$

pro všechna $(x_1, \dots x_n) \in \mathbf{R}$ Tato funkce se nazývá **sdružená hustota** vektoru $\boldsymbol{X}$.

Marginální rozdělení[3.3] spojitě rozděleného vektoru $\boldsymbol{X}$ jsou spojitá[2.6], složka X_j má hustotu[2.7]

$$f_j(x_j) = \int_{\mathbf{R}^{n-1}} \cdots \int f(x_1, \dots, x_n) \times \mathrm{d}x_1 \dots \mathrm{d}x_{j-1}\, \mathrm{d}x_{j+1} \dots \mathrm{d}x_n,$$

tj. pro dané x_j integrujeme hustotu sdruženého rozdělení přes všechny hodnoty ostatních složek.

◇ Viz též: Diskrétní sdružené rozdělení[3.4], Vícerozměrné normální rozdělení[4.7]

3.6. *Střední hodnota vektoru.* **Střední hodnotou** náhodného vektoru[3.1] $\boldsymbol{X} = (X_1, \dots, X_n)$ rozumíme vektor středních hodnot[2.9] jeho složek

$$\mathrm{E}\,\boldsymbol{X} = (\mathrm{E}\,X_1, \dots, \mathrm{E}\,X_n).$$

Střední hodnotu náhodné veličiny[2.1] $g(\boldsymbol{X})$, kde $g\colon \mathbf{R}^n \to \mathbf{R}$, lze pro vektor s diskrétním rozdělením[3.4] s pravděpodobnostní funkcí[2.5] P počítat jako

$$\mathrm{E}\,g(\boldsymbol{X}) = \sum_{\boldsymbol{x} \in M} g(\boldsymbol{x}) P(\boldsymbol{x})$$

a pro vektor se spojitým rozdělením[3.5] s hustotou[3.5] f jako

$$\mathrm{E}\,g(\boldsymbol{X}) = \int_{\mathbf{R}^n} \cdots \int g(\boldsymbol{x}) f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}.$$

◇ Viz též: Kovariance[3.7], Párový test[7.11], Dvouvýběrový test[7.12]

3.7. *Kovariance.* **Kovariance** mezi složkami X_i a X_j náhodného vektoru[3.1] $\boldsymbol{X}$ je

$$\begin{aligned}\operatorname{cov}(X_i, X_j) &= \mathrm{E}(X_i - \mathrm{E}\,X_i)(X_j - \mathrm{E}\,X_j) \\ &= \mathrm{E}\,X_i X_j - \mathrm{E}\,X_i\,\mathrm{E}\,X_j.\end{aligned}$$

Je-li $i = j$, dostáváme $\operatorname{cov}(X_i, X_i) = \operatorname{var} X_i$, tedy rozptyl[2.11].

S pomocí kovariance můžeme psát pro rozptyl lineární kombinace veličin[2.1] X a Y s konstantami $a, b \in \mathbf{R}$

$$\begin{aligned}&\operatorname{var}(aX + bY) \\ &\quad = a^2 \operatorname{var} X + 2ab \operatorname{cov}(X, Y) + b^2 \operatorname{var} Y.\end{aligned}$$

◇ Viz též: Varianční matice[3.8], Korelační koeficient[3.9]

3.8. *Varianční matice.* **Varianční matice** NÁHODNÉHO VEKTORU[3.1] $\boldsymbol{X} = (X_1, \dots, X_n)$ je matice složená z KOVARIANCÍ[3.7] mezi jeho složkami,

$$\operatorname{var} \boldsymbol{X} = \big(\operatorname{cov}(X_i, X_j)\big)_{i,j=1,\dots,n}.$$

Varianční matice je symetrická pozitivně semidefinitní, s ROZPTYLY[2.11] složek na diagonále.

$$V = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & & \sigma_{2n} \\ \vdots & & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix}$$

Varianční matice

◇ Viz též: VÍCEROZMĚRNÉ NORMÁLNÍ ROZDĚLENÍ[4.7]

3.9. *Korelační koeficient.* **Korelační koeficient** mezi složkami X_i a X_j NÁHODNÉHO VEKTORU[3.1] $\boldsymbol{X}$ je definován jako

$$\rho_{ij} = \frac{\operatorname{cov}(X_i, X_j)}{\sqrt{\operatorname{var} X_i \operatorname{var} X_j}}.$$

Tato hodnota z intervalu $\langle -1, 1\rangle$ je mírou lineární závislosti mezi X_i a X_j.

Při $\rho = 0$ mluvíme o nekorelovaných veličinách, krajní hodnoty $\rho_{ij} = \pm 1$ znamenají lineární závislost mezi X_i a X_j. Při $\rho_{ij} = \pm 1$ totiž existují konstanty $a, b \in \mathbf{R}$ (znaménko b je přitom stejné jako u ρ_{ij}), že s pravděpodobností 1 je $X_j = a + bX_i$.

◇ Viz též: NEZÁVISLÉ NÁHODNÉ VELIČINY[3.10]

3.10. *Nezávislé náhodné veličiny.* NÁHODNÉ VELIČINY[2.1] $X_1, X_2, \dots,$ definované na stejném PRAVDĚPODOBNOSTNÍM PROSTORU[1.3], se nazývají **nezávislé**, jestliže pro každé $n \in \mathbf{N}$ je

$$F_{(X_1,\dots,X_n)}(x_1, \dots, x_n) = F_1(x_1) \cdots F_n(x_n)$$

pro všechna $(x_1, \dots, x_n) \in \mathbf{R}^n$, tedy pokud SDRUŽENÁ[3.2] distribuční funkce vektoru $(X_1, \dots, X_n)$ je součinem MARGINÁLNÍCH[3.3] distribučních funkcí jednotlivých veličin. Obdobně definujeme nezávislost NÁHODNÝCH VEKTORŮ[3.1].

Pro SPOJITĚ ROZDĚLENÉ[3.5] $(X_1, \dots, X_n)$ je tato shoda ekvivalentní rovnosti HUSTOT[2.7]

$$f_{(X_1,\dots,X_n)}(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

pro všechna $x_1, \dots, x_n \in \mathbf{R}$.

Jsou-li $X_1, X_2, \dots$ nezávislé, pak pro libovolné (MĚŘITELNÉ[2.1]) funkce $g_i \colon \mathbf{R} \to \mathbf{R}$, $i = 1, 2, \dots,$ jsou nezávislé i veličiny $g_1(X_1), g_2(X_2), \dots$.

Jsou-li veličiny X_i a X_j nezávislé, pak

$$\mathrm{E}\, X_i X_j = \mathrm{E}\, X_i \,\mathrm{E}\, X_j, \quad \operatorname{cov}(X_i, X_j) = \rho_{ij} = 0.$$

Nezávislé veličiny jsou tedy NEKORELOVANÉ[3.9]. Opačné tvrzení obecně neplatí.

Jsou-li NÁHODNÉ VELIČINY[2.1] X a Y nezávislé, pak vzorec pro ROZPTYL[2.11] součtu, resp. rozdílu neobsahuje KOVARIANCI[3.7],

$$\operatorname{var}(X \pm Y) = \operatorname{var} X + \operatorname{var} Y.$$

◇ Viz též: TEST NEZÁVISLOSTI V KONTINGENČNÍ TABULCE[7.15], TEST NEZÁVISLOSTI V NORMÁLNÍM ROZDĚLENÍ[7.16]

4. NĚKTERÁ ROZDĚLENÍ

4.1. *Alternativní rozdělení.* **Alternativní rozdělení** $\mathrm{Alt}(p)$ s parametrem $p \in (0, 1)$ je DISKRÉTNÍ ROZDĚLENÍ[2.4] soustředěné na dvoubodové množině $\{0, 1\}$ s PRAVDĚPODOBNOSTNÍ FUNKCÍ[2.5]

$$P(1) = p, \quad P(0) = 1 - p$$

a nulovou jinde. NÁHODNÁ VELIČINA[2.1] s alternativním rozdělením, $X \sim \mathrm{Alt}(p)$, má STŘEDNÍ HODNOTU[2.9] a ROZPTYL[2.11]

$$\mathrm{E}\, X = p, \qquad \operatorname{var} X = p(1 - p).$$

◇ Viz též: Binomické rozdělení[4.2]

4.2. *Binomické rozdělení.* **Binomické rozdělení** s parametry $n \in \mathbf{N}$ a $p \in (0,1)$, značení např. $\mathrm{Bi}(n,p)$, je diskrétní rozdělení[2.4] na množině $\{0,1,\dots,n\}$ s pravděpodobnostní funkcí[2.5]

$$P(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k = 0,1,2,\dots,n \\ 0 & \text{jinak.} \end{cases}$$

Náhodná veličina[2.1] s binomickým rozdělením, $X \sim \mathrm{Bi}(n,p)$, má střední hodnotu[2.9] a rozptyl[2.11]

$$\mathrm{E}\,X = np, \qquad \mathrm{var}\,X = np(1-p).$$

Jsou-li $Y_1, Y_2, \dots, Y_n$ nezávislé[3.10] náhodné veličiny s alternativním rozdělením[4.1] $\mathrm{Alt}(p)$, pak $S = \sum_i Y_i \sim \mathrm{Bi}(n,p)$.

Binomickým rozdělením můžeme modelovat počet úspěchů nastávajících v posloupnosti n nezávisle[1.5] opakovaných náhodných pokusů[1.1], z nichž každý končí úspěchem s pravděpodobností[1.3] p a neúspěchem s pravděpodobností $1-p$.

◇ Viz též: Aproximace binomického rozdělení Poissonovým[4.4], Centrální limitní věta pro binomické rozdělení[5.6]

4.3. *Poissonovo rozdělení.* **Poissonovo rozdělení** s parametrem $\lambda > 0$, značení např. $\mathrm{Po}(\lambda)$, je diskrétní rozdělení[2.4] na množině celých nezáporných čísel s pravděpodobnostní funkcí[2.5]

$$P(k) = \begin{cases} \mathrm{e}^{-\lambda}\lambda^k/k!, & k = 0,1,2,\dots, \\ 0 & \text{jinak.} \end{cases}$$

Náhodná veličina[2.1] s Poissonovým rozdělením, $X \sim \mathrm{Po}(\lambda)$, má střední hodnotu[2.9] a rozptyl[2.11] rovny parametru rozdělení,

$$\mathrm{E}\,X = \mathrm{var}\,X = \lambda,$$

hodnoty distribuční funkce[2.3] bývají tabelovány.

◇ Viz též: Aproximace binomického rozdělení Poissonovým[4.4]

4.4. *Aproximace binomického rozdělení Poissonovým.* Při velkých hodnotách n a malých hodnotách p je možné pravděpodobnostní[2.5], resp. distribuční[2.3] funkci binomického rozdělení[4.2] $\mathrm{Bi}(n,p)$ **aproximovat** pravděpodobnostní, resp. distribuční funkcí Poissonova rozdělení[4.3] $\mathrm{Po}(\lambda)$ s parametrem $\lambda = np$.

Aproximace je považována za dobrou pro $n > 30$ a $p < 0{,}1$.

◇ Viz též: Centrální limitní věta pro binomické rozdělení[5.6]

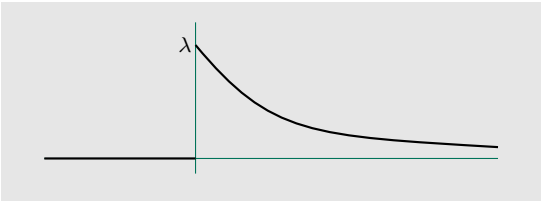
4.5. *Exponenciální rozdělení.* **Exponenciální rozdělení** s parametrem $\lambda > 0$, značení např. $\mathrm{Exp}(\lambda)$, je spojité rozdělení[2.6] na množině kladných čísel s distribuční funkcí[2.3]

$$F(x) = \begin{cases} 1 - \mathrm{e}^{-\lambda x}, & x > 0, \\ 0 & \text{jinak,} \end{cases}$$

tj. s hustotou[2.7]

$$f(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x > 0, \\ 0 & \text{jinak.} \end{cases}$$

Toto rozdělení má konstantní intenzitu poruch[2.8], $\lambda(x) \equiv \lambda$, $x > 0$.



Hustota exponenciálního rozdělení

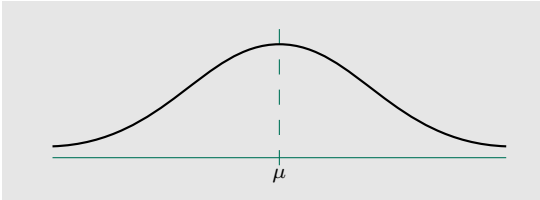
NÁHODNÁ VELIČINA[2.1] s exponenciálním rozdělením, $X \sim \mathrm{Exp}(\lambda)$, má STŘEDNÍ HODNOTU[2.9] a ROZPTYL[2.11]

$$\mathrm{E}\,X = 1/\lambda, \qquad \operatorname{var} X = 1/\lambda^2.$$

Někdy se jako parametr exponenciálního rozdělení udává nikoli intenzita poruch λ, ale její převrácená hodnota, tj. střední hodnota rozdělení $\theta = 1/\lambda$.

4.6. *Normální rozdělení.* **Normální (Gaussovo) rozdělení** se STŘEDNÍ HODNOTOU[2.9] $\mu \in \mathbf{R}$ a ROZPTYLEM[2.11] σ^2 ($\sigma > 0$ je SMĚRODATNÁ ODCHYLKA[2.12]), značení $\mathrm{N}(\mu, \sigma^2)$, je SPOJITÉ ROZDĚLENÍ[2.6] s HUSTOTOU[2.7]

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbf{R}.$$



Hustota normálního rozdělení

Normální rozdělení s nulovou střední hodnotu a jednotkovým rozptylem, $\mathrm{N}(0,1)$, se nazývá standardní (normované) normální, jeho hustota se značí

$$\varphi(x) = (1/\sqrt{2\pi})\mathrm{e}^{-x^2/2}, \quad x \in \mathbf{R}.$$

DISTRIBUČNÍ[2.3] a KVANTILOVÁ[2.14] funkce rozdělení $\mathrm{N}(0,1)$,

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)\,\mathrm{d}t, \quad u_p = \Phi_{-1}(p), \quad p \in (0,1),$$

bývají tabelovány, vzhledem k symetrii $\mathrm{N}(0,1)$ platí

$$\Phi(-x) = 1 - \Phi(x), \quad u_{1-p} = -u_p.$$

Při lineární transformaci zůstává normalita rozdělení zachována. Je-li $X \sim \mathrm{N}(\mu, \sigma^2)$ a $a, b \in \mathbf{R}$, pak $aX + b \sim \mathrm{N}(a\mu + b, b^2\sigma^2)$. Pro distribuční funkci F_X VELIČINY[2.1] $X \sim \mathrm{N}(\mu, \sigma^2)$ můžeme tedy psát

$$F_X(x) = \Phi\Big(\frac{x-\mu}{\sigma}\Big), \quad x \in \mathbf{R}.$$

◇ Viz též: VÍCEROZMĚRNÉ NORMÁLNÍ ROZDĚLENÍ[4.7], CENTRÁLNÍ LIMITNÍ VĚTA[5.4]

4.7. *Vícerozměrné normální rozdělení.* n**-rozměrné normální** rozdělení $\mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{V})$, kde $\boldsymbol{\mu}$ je sloupcový vektor z $\mathbf{R}^n$ a $\boldsymbol{V}$ pozitivně definitní matice řádu n, je ROZDĚLENÍ[3.2] s HUSTOTOU[3.5]

$$f(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{V}}} \exp(-(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{V}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})/2),$$

$\boldsymbol{x} = (x_1, \dots, x_n)' \in \mathbf{R}^n$.

Parametr $\boldsymbol{\mu}$ je STŘEDNÍ HODNOTOU[3.6] a $\boldsymbol{V}$ VARIANČNÍ MATICÍ[3.8] tohoto rozdělení.

Má-li NÁHODNÝ VEKTOR[3.1] $\boldsymbol{X}$ rozdělení $\mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{V})$, jeho složky mají také NORMÁLNÍ ROZDĚLENÍ[4.6], $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, kde σ_i^2 je i-tý diagonální prvek ve $\boldsymbol{V}$. Normálně rozdělené jsou i lineární kombinace složek.

Jsou-li složky X_i a X_j náhodného vektoru se sdruženým normálním rozdělením NEKORELOVANÉ[3.9], pak X_i a X_j jsou NEZÁVISLÉ[3.10]. Ve sdruženém normálním rozdělení je tedy nekorelovanost ekvivalentní nezávislosti (což obecně neplatí).

◇ Viz též: TEST NEZÁVISLOSTI V NORMÁLNÍM ROZDĚLENÍ[7.16]

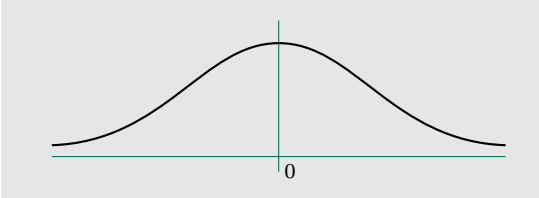
Hustota chí-kvadrát rozdělení

4.8. *Chí-kvadrát rozdělení.* **Chí-kvadrát rozdělení** o $n \in \mathbf{N}$ stupních volnosti, χ^2_n, je SPOJITÉ ROZDĚLENÍ[2.6] na množině kladných čísel s HUSTOTOU[2.7]

$$f(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} \mathrm{e}^{-x/2}, & x > 0, \\ 0 & \text{jinak.} \end{cases}$$

Takové ROZDĚLENÍ[2.2] má například NÁHODNÁ VELIČINA[2.1] $Y = X_1^2 + \cdots + X_n^2$, když $X_1, \ldots, X_n$ jsou NEZÁVISLÉ[3.10] náhodné veličiny s rozdělením NORMÁLNÍM[4.6] $\mathrm{N}(0,1)$.

◇ Viz též: STUDENTOVO T-ROZDĚLENÍ[4.9], FISHEROVO-SNEDECOROVO F-ROZDĚLENÍ[4.10], TEST O ROZPTYLU NORMÁLNÍHO ROZDĚLENÍ[7.10], CHÍ-KVADRÁT TEST DOBRÉ SHODY[7.13]



Hustota t-rozdělení

4.9. *Studentovo t-rozdělení.* **Studentovo t-rozdělení** o $n \in \mathbf{N}$ stupních volnosti, t_n, je SPOJITÉ ROZDĚLENÍ[2.6] s HUSTOTOU[2.7]

$$f(x) = \frac{\Gamma\big((n+1)/2\big)}{\Gamma(n/2)\sqrt{\pi n}} \Big(1 + \frac{x^2}{n}\Big)^{-(n+1)/2}, \quad x \in \mathbf{R}.$$

Takové ROZDĚLENÍ[2.2] má například NÁHODNÁ VELIČINA[2.1] $T = U/\sqrt{Y/n}$, když U a Y jsou NEZÁVISLÉ[3.10] náhodné veličiny s rozdělením NORMÁLNÍM[4.6] $\mathrm{N}(0,1)$ a CHÍ-KVADRÁT[4.8] χ^2_n.

◇ Viz též: FISHEROVO-SNEDECOROVO F-ROZDĚLENÍ[4.10], TESTY STŘEDNÍ HODNOTY NORMÁLNÍHO ROZDĚLENÍ[7.8], PÁROVÝ TEST[7.11], DVOUVÝBĚROVÝ TEST[7.12]

4.10. *Fisherovo-Snedecorovo F-rozdělení.* **Fisherovo-Snedecorovo F-rozdělení** o $m \in \mathbf{N}$ a $n \in \mathbf{N}$ stupních volnosti, $F_{m,n}$, je SPOJITÉ ROZDĚLENÍ[2.6] na množině kladných čísel s HUSTOTOU[2.7]

$$f(x) = \begin{cases} C \cdot x^{m/2-1} \Big(1 + \frac{m}{n} x\Big)^{-(m+n)/2}, & x > 0, \\ 0 & \text{jinak,} \end{cases}$$

kde

$$C = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \Big(\frac{m}{n}\Big)^{m/2}$$

je normovací konstanta.

Takové ROZDĚLENÍ[2.2] má například NÁHODNÁ VELIČINA[2.1] $Z = (X/m)/(Y/n)$, když X a Y jsou NEZÁVISLÉ[3.10] náhodné veličiny s CHÍ-KVADRÁT[4.8] rozděleními χ^2_m a χ^2_n.

◇ Viz též: TEST VÝZNAMNOSTI REGRESE[8.7]

5. LIMITNÍ VĚTY

5.1. *Zákon velkých čísel.* Jako **zákon velkých čísel** označujeme tvrzení o konvergenci průměrů v posloupnosti NÁHODNÝCH VELIČIN[2.1]. Nastává-li tato konvergence s PRAVDĚPODOBNOSTÍ[1.3] 1, hovoříme o **silném** zákonu velkých čísel.

Jsou-li $X_1, X_2, \ldots$ NEZÁVISLÉ[3.10] a stejně ROZDĚLENÉ[2.2] náhodné veličiny s konečnou STŘEDNÍ HODNOTOU[2.9] $\mu = \mathrm{E}\, X_1$, pak pro $S_n := \sum_{i=1}^n X_i$, $n \in \mathbf{N}$, platí

$$S_n/n \xrightarrow{n\to\infty} \mu \quad \text{s pravděpodobností 1.}$$

Pro S_n označující počet výskytů JEVU[1.1] A v prvních n pokusech v posloupnosti nezávisle se opakujících NÁHODNÝCH POKUSŮ[1.1], tj. když $X_i = 1$ nebo 0 podle toho, zda v i-tém

pokusu jev A nastal nebo nenastal, dostáváme s pravděpodobností 1 **statistickou definici pravděpodobnosti** založenou na limitě relativních četností,

$$S_n/n \xrightarrow{n\to\infty} \mathrm{E}\,X_1 = \mathrm{P}[\text{výskyt } A \text{ v daném pokusu}].$$

◇ Viz též: Kolmogorovův zákon velkých čísel[5.2], Slabý zákon velkých čísel[5.3]

5.2. *Kolmogorovův zákon velkých čísel.* **Kolmogorovův zákon velkých čísel** je zákon velkých čísel[5.1] tohoto znění:

Jsou-li náhodné veličiny[2.1] $X_1, X_2, \ldots$ nezávislé[3.10], s konečnými rozptyly[2.11], a je-li $\sum_{i=1}^{\infty} \operatorname{var} X_i/b_i < \infty$ pro nějakou posloupnost čísel $b_1 < b_2 < \ldots$ s $\lim b_n = \infty$ při $n \to \infty$, pak

$$\frac{1}{b_n}\sum_{i=1}^{n}(X_i - \mathrm{E}\,X_i) \xrightarrow{n\to\infty} 0 \quad \text{s pravděpodobností[1.3] } 1.$$

Není třeba, aby $X_1, X_2, \ldots$ byly stejně rozdělené[2.2].

◇ Viz též: Slabý zákon velkých čísel[5.3]

5.3. *Slabý zákon velkých čísel.* **Slabý zákon velkých čísel** postihuje — oproti silnému[5.1] — slabší formu konvergence (totiž jen pro pravděpodobnosti[1.3] odchylek průměru[6.5] od jeho střední hodnoty[2.9]), ovšem také za slabších předpokladů na nezávislost[3.10] v posloupnosti náhodných veličin[2.1].

Jsou-li $X_1, X_2, \ldots$ nekorelované[3.9] náhodné veličiny s konečnými rozptyly[2.11], $S_n := \sum_{i=1}^{n} X_i$ a je-li $\operatorname{var} S_n/n^2 = \sum_{i=1}^{n} \operatorname{var} X_i/n^2 \to 0$ při $n \to \infty$, pak

$$\mathrm{P}\Big[\Big|\frac{S_n - \sum_{i=1}^{n}\mathrm{E}\,X_i}{n}\Big| \leqq \varepsilon\Big] \xrightarrow{n\to\infty} 1 \quad \text{pro každé } \varepsilon > 0.$$

◇ Viz též: Zákon velkých čísel[5.1], Kolmogorovův zákon velkých čísel[5.2], Centrální limitní věta[5.4]

5.4. *Centrální limitní věta.* **Centrální limitní věty** popisují limity pravděpodobností[1.3] odchylek náhodných veličin[2.1] od jejich střední hodnoty[2.9].

Pro stejně rozdělené[2.2] náhodné veličiny platí **Lindebergova-Lévyova** centrální limitní věta: Jsou-li $X_1, X_2, \ldots$ nezávislé[3.10] stejně rozdělené[2.2] náhodné veličiny s konečným rozptylem[2.11], pak pro každé $x \in \mathbf{R}$

$$\mathrm{P}\Big[\frac{\sum_{i=1}^{n} X_i - n\,\mathrm{E}\,X_1}{\sqrt{n \operatorname{var} X_1}} < x\Big] \to \Phi(x) \quad \text{při } n \to \infty,$$

kde Φ je distribuční funkce[2.3] standardního normálního rozdělení[4.6] $N(0,1)$.

To znamená, že je-li $S_n = \sum_{i=1}^{n} X_i$ součtem velkého množství nezávislých náhodných veličin, můžeme její rozdělení považovat za přibližně normální,

$$S_n \mathrel{\dot\sim} \mathrm{N}(n\mu, n\sigma^2),$$

kde $\mu = \mathrm{E}\,X_1$ a $\sigma^2 = \operatorname{var} X_1$.

◇ Viz též: Ljapunovova centrální limitní věta[5.5], Centrální limitní věta pro binomické rozdělení[5.6], Slabý zákon velkých čísel[5.3]

5.5. *Ljapunovova centrální limitní věta.* **Ljapunovova centrální limitní věta** nevyžaduje, na rozdíl od LINDEBERGOVY-LÉVYOVY[5.4], stejné ROZDĚLENÍ[2.2] u jednotlivých NÁHODNÝCH VELIČIN[2.1].

Jsou-li $X_1, X_2, \ldots$ NEZÁVISLÉ[3.10] náhodné veličiny, $s_n^2 = \operatorname{var} \sum_1^n X_i = \sum_1^n \mathrm{E}(X_i - \mathrm{E}\, X_i)^2$ a je-li

$$\frac{\sum_1^n \mathrm{E}\,|X_i - \mathrm{E}\, X_i|^{2+\delta}}{s_n^{2+\delta}} \xrightarrow{n\to\infty} 0 \quad \text{pro nějaké } \delta > 0,$$

potom pro každé $x \in \mathbf{R}$

$$\mathrm{P}\Big[\frac{\sum_1^n X_i - \sum_1^n \mathrm{E}\, X_i}{\sqrt{s_n^2}} < x\Big] \to \Phi(x) \quad \text{při } n \to \infty,$$

kde Φ je DISTRIBUČNÍ FUNKCE[2.3] standardního NORMÁLNÍHO ROZDĚLENÍ[4.6] $N(0,1)$.

◇ Viz též: CENTRÁLNÍ LIMITNÍ VĚTA[5.4]

5.6. *Centrální limitní věta pro binomické rozdělení.* Uplatněním CENTRÁLNÍ LIMITNÍ VĚTY[5.4] na NÁHODNÉ VELIČINY[2.1] S_n s BINOMICKÝM ROZDĚLENÍM[4.2] $\mathrm{Bi}(n,p)$, chápané jako součty NEZÁVISLÝCH[3.10] veličin s ALTERNATIVNÍM ROZDĚLENÍM[4.1], dostaneme

$$S_n \mathrel{\dot\sim} \mathrm{N}(np, np(1-p)) \quad \text{při } n \to \infty.$$

Tuto aproximaci lze považovat za dobrou, když $\operatorname{var} S_n = np(1-p) > 9$.

◇ Viz též: APROXIMACE BINOMICKÉHO ROZDĚLENÍ POISSONOVÝM[4.4]

6. TEORIE ODHADU

6.1. *Náhodný výběr.* Posloupnost NÁHODNÝCH VELIČIN[2.1] (příp. NÁHODNÝCH VEKTORŮ[3.1]) $X_1, \ldots, X_n$ se nazývá **náhodný výběr** rozsahu n z ROZDĚLENÍ[2.2] s DISTRIBUČNÍ FUNKCÍ[2.3] F, jsou-li tyto veličiny NEZÁVISLÉ[3.10] a všechny mají stejné rozdělení charakterizované distribuční funkcí F.

Taková posloupnost veličin odpovídá výběru ze základního souboru s vracením.

6.2. *Statistika (hodnota).* **Statistikou** rozumíme (MĚŘITELNOU[2.1]) funkci pozorovaných hodnot, kterými je obvykle NÁHODNÝ VÝBĚR[6.1].

◇ Viz též: POPISNÉ STATISTIKY[6.4]

6.3. *Odhad.* **Odhadem** parametrické funkce $g(\theta)$ na základě pozorování $X_1, \ldots, X_n$ nazýváme libovolnou STATISTIKU[6.2] $S(X_1, \ldots, X_n)$, která nezávisí na θ.

◇ Viz též: POPISNÉ STATISTIKY[6.4], INTERVALOVÉ ODHADY[6.16], MAXIMÁLNĚ VĚROHODNÝ ODHAD[6.15]

6.4. *Popisné statistiky.* **Popisné statistiky** zachycují některé charakteristiky NÁHODNÉHO VÝBĚRU[6.1]. Jedná se o výběrové protějšky obdobných charakteristik ROZDĚLENÍ[2.2], jako např. VÝBĚROVÝ PRŮMĚR[6.5], VÝBĚROVÝ ROZPTYL[6.7], VÝBĚROVÝ SMĚRODATNÁ ODCHYLKA[6.8], VÝBĚROVÁ KOVARIANCE[6.9], VÝBĚROVÝ KORELAČNÍ KOEFICIENT[6.10], VÝBĚROVÝ MEDIÁN[6.6].

6.5. *Výběrový průměr.* **Výběrovým průměrem** z NÁHODNÉHO VÝBĚRU[6.1] $X_1, \dots, X_n$ rozumíme VELIČINU[2.1]

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Jde o výběrový protějšek STŘEDNÍ HODNOTY[2.9].

Výběrový průměr $\overline{X}$ je NESTRANNÝM[6.11] a KONZISTENTNÍM[6.12] odhadem střední hodnoty $\mathrm{E}\, X_i$ (existuje-li tato). Je také NEJLEPŠÍM[6.14] odhadem mezi lineárními (tj. tvaru $\sum_i a_i X_i$) nestrannými odhady $\mathrm{E}\, X_i$, v BINOMICKÉM[4.2], POISSONOVĚ[4.3], NORMÁLNÍM[4.6], či EXPONENCIÁLNÍM[4.5] rozdělení dokonce mezi všemi odhady.

◇ Viz též: TEST O STŘEDNÍ HODNOTĚ[7.9]

6.6. *Výběrový medián.* **Výběrovým mediánem** z NÁHODNÉHO VÝBĚRU[6.1] $X_1, \dots, X_n$ rozumíme

- pro n liché veličinu

$$\widetilde{X} = X_{((n+1)/2)},$$

- pro n sudé veličinu

$$\widetilde{X} = (X_{(n/2)} + X_{(n/2+1)})/2,$$

kde $X_{(1)} \leqq \cdots \leqq X_{(n)}$ značí hodnoty $X_1, \dots, X_n$ uspořádané podle velikosti. Výběrový medián je prostřední hodnotou (co do velikosti), resp. průměrem dvou prostředních hodnot, v náhodném výběru.

Jde o výběrový protějšek MEDIÁNU[2.13].

◇ Viz též: VÝBĚROVÝ PRŮMĚR[6.5]

6.7. *Výběrový rozptyl.* **Výběrovým rozptylem** z NÁHODNÉHO VÝBĚRU[6.1] $X_1, \dots, X_n$, $n > 1$, rozumíme VELIČINU[2.1]

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

kde $\overline{X}$ je VÝBĚROVÝ PRŮMĚR[6.5]. Jde o výběrový protějšek ROZPTYLU[2.11].

Výběrový rozptyl s^2 je jakožto ODHAD[6.3] rozptylu $\operatorname{var} X_i$ (existuje-li tento) NESTRANNÝ[6.11] a KONZISTENTNÍ[6.12].

Pro odhad rozptylu lze použít také STATISTIKU[6.2]

$$\frac{1}{n} \sum_{i} (X_i - \overline{X})^2 = \frac{n-1}{n} s^2,$$

která má menší rozptyl než s^2, na rozdíl od něj je však VYCHÝLENÁ[6.11].

◇ Viz též: VÝBĚROVÁ SMĚRODATNÁ ODCHYLKA[6.8], VÝBĚROVÁ KOVARIANCE[6.9], TEST O ROZPTYLU NORMÁLNÍHO ROZDĚLENÍ[7.10]

6.8. *Výběrová směrodatná odchylka.* **Výběrová směrodatná odchylka** s je odmocninou z VÝBĚROVÉHO ROZPTYLU[6.7] s^2.

6.9. *Výběrová kovariance.* **Výběrovou kovariancí** mezi složkami NÁHODNÉHO VÝBĚRU[6.1]

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \dots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}, \quad n > 1,$$

je VELIČINA[2.1]

$$\begin{aligned} s_{XY} &= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y} \right), \end{aligned}$$

kde $\overline{X}$ a $\overline{Y}$ jsou VÝBĚROVÉ PRŮMĚRY[6.5] složek pozorování. Jde o výběrový protějšek KOVARIANCE[3.7].

◇ Viz též: VÝBĚROVÝ KORELAČNÍ KOEFICIENT[6.10]

6.10. *Výběrový korelační koeficient.* **Výběrový korelační koeficient** mezi složkami NÁHODNÉHO VÝBĚRU[6.1] $\binom{X_1}{Y_1}, \dots, \binom{X_n}{Y_n}$, $n > 1$ je definován pomocí VÝBĚROVÉ KOVARIANCE[6.9] s_{XY} a VÝBĚROVÝCH ROZPTYLŮ[6.7] s_X^2 a s_Y^2 jako

$$r_{XY} = \frac{s_{XY}}{\sqrt{s_X^2 s_Y^2}} = \frac{\sum X_i Y_i - n\overline{X}\overline{Y}}{\left(\sum X_i^2 - n\overline{X}^2\right)\left(\sum Y_i^2 - n\overline{Y}^2\right)},$$

kde $\overline{X}$ a $\overline{Y}$ jsou VÝBĚROVÉ PRŮMĚRY[6.5] složek pozorování. Jde o výběrový protějšek KORELAČNÍHO KOEFICIENTU[3.9].

◇ Viz též: TEST NEZÁVISLOSTI V NORMÁLNÍM ROZDĚLENÍ[7.16]

6.11. *Nestrannost.* ODHAD[6.3] $S_n(X_1, \dots, X_n)$ parametrické funkce $g(\theta)$ se nazývá **nestranný** (nevychýlený), jestliže

$$\mathrm{E}_\theta\, S_n(X_1, \dots, X_n) = g(\theta) \quad \text{pro všechna } \theta.$$

V opačném případě se nazývá **vychýlený**.

VÝBĚROVÝ PRŮMĚR[6.5] a VÝBĚROVÝ ROZPTYL[6.7] coby odhady STŘEDNÍ HODNOTY[2.9] a ROZPTYLU[2.11] jsou nestranné (existují-li příslušné momenty).

6.12. *Konzistence.* ODHAD[6.3] $S_n(X_1, \dots, X_n)$ parametrické funkce $g(\theta)$ se nazývá (silně) **konzistentní**, jestliže

$$\mathrm{P}_\theta[\lim_{n\to\infty} S_n(X_1, \dots, X_n) = g(\theta)] = 1 \quad \text{pro všechna } \theta.$$

VÝBĚROVÝ PRŮMĚR[6.5] a VÝBĚROVÝ ROZPTYL[6.7] coby odhady STŘEDNÍ HODNOTY[2.9] a ROZPTYLU[2.11] jsou konzistentní (existují-li příslušné momenty).

6.13. *Střední čtvercová chyba.* **Střední čtvercová chyba** (MSE, mean squarred error) ODHADU[6.3] $S(X_1, \dots, X_n)$ reálné parametrické funkce $g(\theta)$ je definována jako

$$\mathrm{MSE}_\theta\, S \;= \mathrm{E}_\theta (S(X_1, \dots, X_n) - g(\theta))^2,$$

pokud tato STŘEDNÍ HODNOTA[2.9] existuje.

Udává, jak se v průměru liší odhad od odhadované hodnoty, měříme-li odlišnost pomocí druhé mocniny jeho odchylky od správné hodnoty. Na základě MSE můžeme odhady porovnávat, určovat, který je LEPŠÍ[6.14].

◇ Viz též: NEJLEPŠÍ ODHAD[6.14]

6.14. *Nejlepší odhad.* Řekneme, že ODHAD[6.3] S_1 je (dle STŘEDNÍ ČTVERCOVÉ CHYBY[6.13], MSE) **lepší** než S_2, pokud

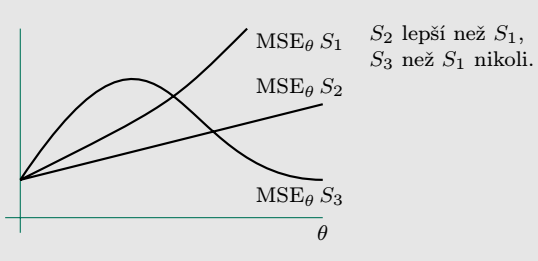
$$\mathrm{MSE}_\theta\, S_1 \leqq \mathrm{MSE}_\theta\, S_2 \quad \text{pro všechna } \theta.$$

Odhad S nazveme **nejlepším** odhadem $g(\theta)$ (ve smyslu MSE), pokud pro libovolný jiný odhad S' je

$$\mathrm{MSE}_\theta\, S \leqq \mathrm{MSE}_\theta\, S' \quad \text{pro všechna } \theta$$

Nejlepší odhad mezi všemi odhady nemusí existovat, často se proto omezujeme jen na určitou třídu odhadů.

◇ Viz též: ODHADY V REGRESI[8.5], VÝBĚROVÝ PRŮMĚR[6.5], VÝBĚROVÝ ROZPTYL[6.7]



Porovnání odhadů

6.15. *Maximálně věrohodný odhad.* Máme-li NÁHODNÝ VÝBĚR[6.1] $\boldsymbol{X} = (X_1, \ldots, X_n)$ s HUSTOTOU[2.7] $f_\theta(\boldsymbol{x})$, odhadem metodou **maximální věrohodnosti** rozumíme tu hodnotu θ, pro niž (při daných $\boldsymbol{X}$) funkce $L(\theta; \boldsymbol{X}) = f_\theta(\boldsymbol{X})$ nabývá maxima,

$$\widehat{\theta}(\boldsymbol{X}) = \arg\max_\theta (L(\theta; \boldsymbol{X})).$$

Maximálně věrohodné odhady jsou (za předpokladu regularity rozdělení) asymptoticky optimální (asymptoticky NESTRANNÉ[6.11] a asymptoticky eficientní).

6.16. *Intervalové odhady.* Je-li počet n pozorování $X_1, \ldots, X_n$ malý, může se často bodový ODHAD[6.3] dost lišit od odhadovaného parametru. Zavádějí se proto intervalové odhady.

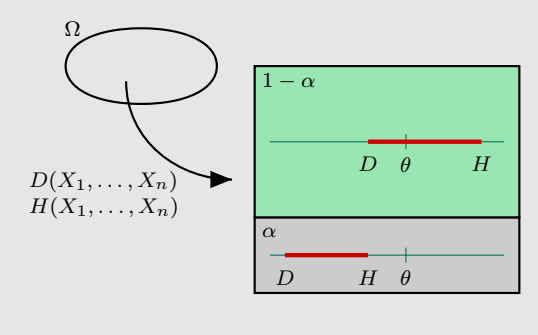
Interval $(D(X_1, \ldots, X_n), H(X_1, \ldots, X_n))$ nazveme **intervalovým odhadem** θ o spolehlivosti $1-\alpha$ nebo $(1-\alpha)$-**intervalem spolehlivosti** pro θ ($\alpha \in (0,1)$), jestliže

$$\mathrm{P}_\theta[D(X_1, \ldots, X_n) < \theta < H(X_1, \ldots, X_n)] = 1 - \alpha.$$

Je-li $H(X_1, \ldots, X_n) = \infty$, resp. $D(X_1, \ldots, X_n) = -\infty$, mluvíme o dolním, resp. horním, intervalovém odhadu.

Krajní body intervalu spolehlivosti jsou funkcemi NÁHODNÉHO VÝBĚRU[6.1], jsou to tedy NÁHODNÉ VELIČINY[2.1]. Definiční rovnost vyjadřuje, že intervalový odhad „pokrývá“ skutečnou hodnotu θ s pravděpodobností $1 - \alpha$. Tím není tvar odhadu definován jednoznačně, obvykle hledáme nejkratší interval s takovou vlastností.

◇ Viz též: INTERVALY SPOLEHLIVOSTI V NORMÁLNÍM ROZDĚLENÍ[6.17]



Pokrytí parametru intervalovým odhadem

6.17. *Intervaly spolehlivosti v normálním rozdělení.* Je-li $X_1, \ldots, X_n$ NÁHODNÝ VÝBĚR[2.1] z NORMÁLNÍHO ROZDĚLENÍ[4.6] $\mathrm{N}(\mu, \sigma^2)$, pak

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

resp.

$$\Big(\frac{(n-1)\sigma^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)\sigma^2}{\chi^2_{n-1,\alpha/2}}\Big),$$

kde $\overline{X}$ je VÝBĚROVÝ PRŮMĚR[6.5], s^2 VÝBĚROVÝ ROZPTYL[6.7] a u_p, $t_p(n-1)$, resp. $\chi^2_p(n-1)$ postupně p-KVANTILY[2.14] rozdělení $\mathrm{N}(0,1)$, t_{n-1}[4.9] a χ^2_{n-1}[4.8], jsou $1-\alpha$ INTERVALY SPOLEHLIVOSTI[6.16] pro μ (první použijeme, když σ^2 známe, druhý, když neznáme), resp. pro σ^2.

Jednostranné intervaly dostaneme, když jeden krajní bod nahradíme $\pm\infty$, resp. ∞ nebo 0, a u druhého použijeme α místo $\alpha/2$.

Díky CENTRÁLNÍ LIMITNÍ VĚTĚ[5.4] a KONZISTENCI[6.12] s^2 můžeme stejný tvar odhadu jako pro μ použít i při výběrech z jiných ROZDĚLENÍ[2.2] s konečným ROZPTYLEM[2.11] σ^2. Interval

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

případně podobně s s místo σ, pak představuje přibližně $1-\alpha$ intervalový odhad STŘEDNÍ HODNOTY[2.9] $\mathrm{E}\,X_1$.

◇ Viz též: TEST O STŘEDNÍ HODNOTĚ[7.9], TESTY STŘEDNÍ HODNOTY NORMÁLNÍHO ROZDĚLENÍ[7.8], TEST O ROZPTYLU NORMÁLNÍHO ROZDĚLENÍ[7.10]

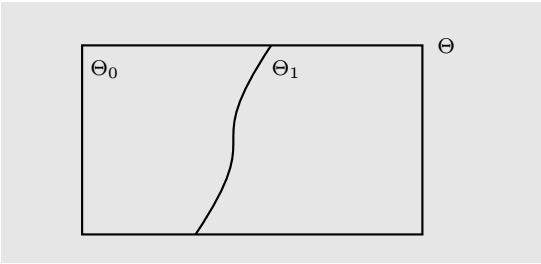
7. TESTOVÁNÍ HYPOTÉZ

7.1. *Hypotéza.* **Hypotézou** rozumíme tvrzení o ROZDĚLENÍ[2.2] pozorované NÁHODNÉ VELIČINY[2.1] (např. o rozdělení NÁHODNÉHO VÝBĚRU[6.1]).

Jsou-li rozdělení, která přicházejí v úvahu, parametrizována pomocí parametru $\theta \in \Theta$, považujeme za hypotézu také tvrzení o hodnotě tohoto parametru.

Jednoduchou hypotézu představuje hypotéza $H\colon \theta = \theta_0$, že parametr θ je roven danému číslu $\theta_0 \in \Theta$. Hypotéza $H'\colon \theta \in \Theta_0$, že parametr θ pochází z dané (víceprvkové) množiny $\Theta_0 \subset \Theta$, je **hypotézou složenou**.

◇ Viz též: TEST[7.2]



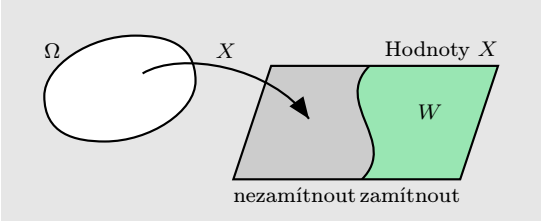
Nulová a alternativní hypotéza

7.2. *Test.* **Test hypotézy** je pravidlo, které pozorované hodnotě NÁHODNÉ VELIČINY[2.1] (obvykle realizaci NÁHODNÉHO VÝBĚRU[6.1]) přiřadí rozhodnutí HYPOTÉZU[7.1] zamítnout nebo nezamítnout.

Obvykle proti **nulové hypotéze** $H_0\colon \theta \in \Theta_0$, že hodnota parametru θ pochází z dané množiny $\Theta_0 \subset \Theta$, kde Θ je množina možných hodnot parametru, stavíme **alternativní hypotézu** $H_1\colon \theta \in \Theta_1$, kde $\Theta_1 = \Theta \setminus \Theta_0$, že tomu tak není.

Zamítnutí H_0 pak fakticky znamená rozhodnout se pro H_1.

◇ Viz též: KRITICKÝ OBOR[7.3], CHYBA 1. A 2. DRUHU[7.4]

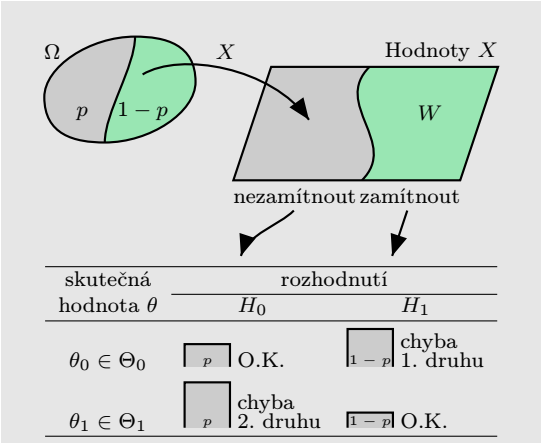


Kritický obor

7.3. *Kritický obor.* Je-li TEST[7.2] založen na NÁHODNÉ VELIČINĚ[2.1] (pozorování) X s hodnotami v $\mathbf{R}^n$, pak jeho **kritický obor** je množina $W \subset \mathbf{R}^n$ těch hodnot X, pro něž test HYPOTÉZU[7.1] zamítá.

Kritický obor lze často popsat prostřednictvím kritického oboru pro vhodnou (jednorozměrnou) STATISTIKU[6.2] pozorovaných hodnot.

◇ Viz též: CHYBA 1. A 2. DRUHU[7.4]



Chyby 1. a 2. druhu a jejich pravděpodobnosti

7.4. *Chyba 1. a 2. druhu.* Rozhodnutí TESTU[7.2] o HYPOTÉZE[7.1] nemusí být vždy správné. **Chyba 1. druhu** nastává, když je hypotéza zamítnuta, přestože platí, **chyba 2. druhu**, když hypotéza zamítnuta není, přestože neplatí.

Kvalita testu je dána pravděpodobnostmi, s jakými tyto chyby nastávají. Je-li test hypotézy $H_0\colon \theta \in \Theta_0$ proti ALTERNATIVĚ[7.2] $H_1\colon \theta \in \Theta_1$ o parametru θ založený na hodnotách NÁHODNÉ VELIČINY[2.1] X s KRITICKÝM OBOREM[7.3] W, pak pravděpodobnost chyby 1. druhu je

$$\mathrm{P}_\theta[X \in W], \quad \theta \in \Theta_0$$

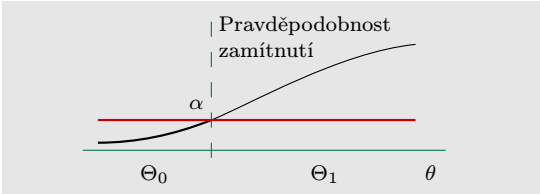
a chyby 2. druhu

$$\mathrm{P}_\theta[X \notin W], \quad \theta \in \Theta_1,$$

kde P_θ označuje PRAVDĚPODOBNOSTNÍ MÍRU[1.3], při níž má X ROZDĚLENÍ[2.2] s parametrem θ.

Při daném rozsahu VÝBĚRU[6.1] obvykle nelze současně minimalizovat pravděpodobnosti obou druhů chyb.

◇ Viz též: HLADINA VÝZNAMNOSTI[7.5], SÍLA TESTU[7.6]

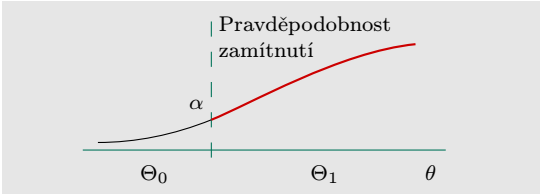


Hladina významnosti

7.5. *Hladina významnosti.* TESTEM[7.2] na **hladině významnosti** α (α-testem), $\alpha \in (0,1)$, rozumíme test, u nějž pravděpodobnost CHYBY 1. DRUHU[7.4] nepřekračuje hodnotu α.

Je žádoucí, aby test měl nízkou hladinou významnosti, rozhodnutí o zamítnutí HYPOTÉZY[7.1] pak interpretujeme ve smyslu, že hypotéza neplatí. Obvykle volené hodnoty pro hladinu α bývají 0,05, 0,01 nebo 0,001.

◇ Viz též: SÍLA TESTU[7.6], P-HODNOTA[7.7]



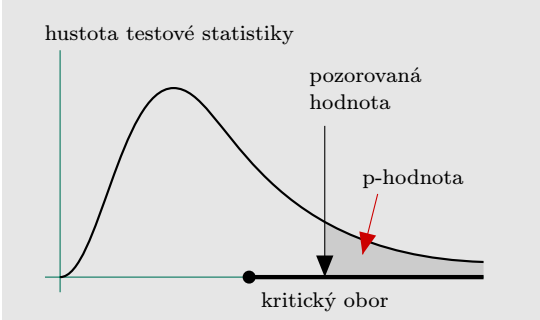
Síla testu

7.6. *Síla testu.* Má-li TEST[7.2] (s KRITICKÝM OBOREM[7.3] W) pravděpodobnost CHYBY 2. DRUHU[7.4] $\mathrm{P}_\theta[X \notin W]$, pak hodnota

$$\mathrm{P}_\theta[X \in W] = 1 - \mathrm{P}_\theta[X \notin W]$$

se nazývá **síla testu** proti ALTERNATIVĚ[7.2] $\theta \in \Theta_1$.

Při hledání optimálního testu stanovíme HLADINU VÝZNAMNOSTI[7.5] α (omezíme pravděpodobnost CHYBY 1. DRUHU[7.4]) a mezi α-testy hledáme ten s největší sílou (nejmenší chybou 2. druhu).



P-hodnota

7.7. *P-hodnota.* **P-hodnota** TESTU[7.2] je u testů, kde má tato definice smysl, pravděpodobnost, s jakou testovací STATISTIKA[6.2] nabývá hodnot „horších“ (více svědčících proti testované HYPOTÉZE[7.1]), než je pozorovaná hodnota statistiky.

P-hodnota je obvyklým výstupem počítačových programů na testování hypotéz, udává mezní HLADINU VÝZNAMNOSTI[7.5], při které bychom hypotézu ještě zamítali.

Hypotézu H_0 zamítáme na hladině α, právě když p-hodnota je menší než α.

7.8. *Testy střední hodnoty normálního rozdělení.* Pro testování HYPOTÉZY[7.1] o STŘEDNÍ HODNOTĚ[2.9] μ veličin $X_1, \ldots, X_n$ NÁHODNÉHO VÝBĚRU[6.1] z NORMÁLNÍHO ROZDĚLENÍ[4.6] $\mathrm{N}(\mu, \sigma^2)$

$$H_0\colon \mu = \mu_0 \quad \text{proti} \quad H_1\colon \mu \neq \mu_0,$$

kde $\mu_0 \in \mathbf{R}$ je dané číslo, používáme TESTY[7.2] založené na STATISTIKÁCH[6.2]

$$U = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}, \quad \text{resp.} \quad T = \frac{\overline{X} - \mu_0}{s/\sqrt{n}},$$

kde $\overline{X}$ je VÝBĚROVÝ PRŮMĚR[6.5] a s^2 VÝBĚROVÝ ROZPTYL[6.7]. Statistiku volíme podle toho, zda ROZPTYL[2.11] σ^2 je známý, nebo ne, druhý z testů se nazývá **t-test**.

Hypotézu H_0 zamítáme na HLADINĚ VÝZNAMNOSTI[7.5] α, pokud

$$|U| > u_{1-\alpha/2}, \quad \text{resp.} \quad |T| > t_{1-\alpha/2}(n-1),$$

kde $u_{1-\alpha/2}$, resp. $t_{1-\alpha/2}(n-1)$ jsou $(1-\alpha/2)$ KVANTILY[2.14] rozdělení $\mathrm{N}(0,1)$, resp. t_{n-1}[4.9].

Za platnosti H_0 mají U, resp. T, ROZDĚLENÍ[2.2] $N(0,1)$, resp. t-rozdělení o $n-1$ stupních volnosti, KRITICKÝ OBOR[7.3] tvoří hodnoty U, resp. T, blízké $\pm\infty$.

Pro test H_0 proti jednostranné alternativě

$$H_1' : \mu > \mu_0, \quad \text{resp.} \quad H_1'' : \mu < \mu_0,$$

na hladině významnosti α použijeme v případě známého σ^2 kritický obor

$$U > u_{1-\alpha}, \quad \text{resp.} \quad U < -u_{1-\alpha},$$

při neznámém σ^2 pak

$$T > t_{1-\alpha}(n-1), \quad \text{resp.} \quad T < -t_{1-\alpha}(n-1)$$

(kritický obor představují hodnoty statistik blízké $+\infty$ u H_1', resp. $-\infty$ u H_1'').

◇ Viz též: Intervaly spolehlivosti v normálním rozdělení[6.17], Dvouvýběrový test[7.12], Párový test[7.11], Test o střední hodnotě[7.9], Test o rozptylu normálního rozdělení[7.10], Test nezávislosti v normálním rozdělení[7.16]

7.9. *Test o střední hodnotě.* Pro přibližné testování hypotézy[7.1] o střední hodnotě[2.9] $\mu = \mathrm{E}\, X_i$ veličin v náhodném výběru[6.1] $X_1, \dots, X_n$

$$H_0 \colon \mu = \mu_0 \quad \text{proti} \quad H_1 \colon \mu \neq \mu_0,$$

kde $\mu_0 \in \mathbf{R}$ je dané číslo, použijeme, podobně jako u t-testu[7.8] v normálním rozdělení[4.6] $\mathrm{N}(\mu, \sigma^2)$, statistiku[6.2]

$$T = \frac{\overline{X} - \mu_0}{s/\sqrt{n}},$$

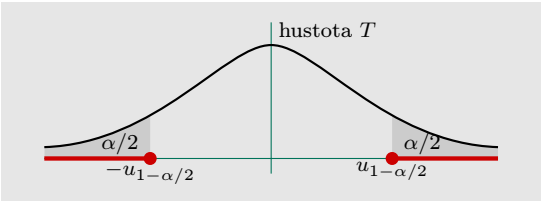
kde $\overline{X}$ je výběrový průměr[6.5] a s^2 výběrový rozptyl[6.7].

Hypotézu H_0 zamítneme na hladině významnosti[7.5] přibližně α v případě, že

$$|T| > u_{1-\alpha/2},$$

kde u_p značí p-kvantil[2.14] rozdělení $N(0,1)$[4.6].

Za H_0 má T díky centrální limitní větě[5.4] a konzistenci[6.12] výběrového rozptylu asymptoticky rozdělení $\mathrm{N}(0,1)$, kritický obor[7.3] představují hodnoty T blízké $\pm\infty$.



Kritický obor oboustranného testu

7.10. *Test o rozptylu normálního rozdělení.* Nechť $X_1, \dots, X_n$ je náhodný výběr[6.1] z normálního rozdělení[4.6] $\mathrm{N}(\mu, \sigma^2)$. Pro testování hypotézy[7.1] o hodnotě rozptylu[2.11] σ^2

$$H_0 \colon \sigma^2 = \sigma_0^2 \quad \text{proti} \quad H_1 \colon \sigma^2 \neq \sigma_0^2,$$

kde $\sigma_0^2 > 0$ je dané číslo, používáme test[7.2] založený na statistice[6.2]

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2},$$

kde s^2 je výběrový rozptyl[6.7].

Hypotézu H_0 zamítneme na hladině významnosti[7.5] α, když

$$\chi^2 > \chi^2_{1-\alpha/2}(n-1), \quad \text{nebo} \quad \chi^2 < \chi^2_{\alpha/2}(n-1),$$

kde $\chi^2_p(n-1)$ značí p-kvantil[2.14] rozdělení χ^2_{n-1}[4.8].

Za H_0 má statistika χ^2 rozdělení χ^2_{n-1}, kritický obor[7.3] tvoří hodnoty χ^2 blízké 0 nebo ∞.

◇ Viz též: Intervaly spolehlivosti v normálním rozdělení[6.17], Testy střední hodnoty normálního rozdělení[7.8], Test nezávislosti v normálním rozdělení[7.16]

7.11. *Párový test.* **Párový t-test** je testem HYPOTÉZY[7.1] o hodnotě rozdílu mezi STŘEDNÍMI HODNOTAMI[2.9] μ_1 a μ_2 složek NÁHODNÝCH VEKTORŮ[3.1] v NÁHODNÉM VÝBĚRU[6.1] $(X_1, Y_1), \dots, (X_n, Y_n)$ z DVOUROZMĚRNÉHO NORMÁLNÍHO ROZDĚLENÍ[4.7],

$$\begin{array}{cccc} X_1 & X_2 & & X_n \\ | & | & \dots & | \\ Y_1 & Y_2 & & Y_n \end{array}$$

Pozorování pro párový test

$$H_0\colon \mu_1 - \mu_2 = d \quad \text{proti} \quad H_1\colon \mu_1 - \mu_2 \neq d,$$

kde $d \in \mathbf{R}$ je dané číslo (pro $d = 0$ to znamená shodu středních hodnot složek $\mu_1 = \mu_2$).

Tento TEST[7.2] zamítá H_0 na HLADINĚ VÝZNAMNOSTI[7.5] α, pokud

$$\left| \frac{\overline{Z} - d}{s_Z/\sqrt{n}} \right| > t_{1-\alpha/2}(n-1),$$

kde $\overline{Z}$ a s^2 jsou VÝBĚROVÝ PRŮMĚR[6.5] a VÝBĚROVÝ ROZPTYL[6.7] náhodného výběru složeného z rozdílů

$$Z_i = X_i - Y_i, \quad i = 1, \dots, n,$$

a $t_p(n)$ značí p-KVANTIL[2.14] rozdělení t_n[4.9].

Ve skutečnosti provádíme t-TEST[7.8] o střední hodnotě $\mu_1 - \mu_2$ výběru NORMÁLNĚ ROZDĚLENÝCH[4.6] veličin $Z_1, \dots, Z_n$.

Párový test můžeme používat pouze pokud je výběr tvořen dvojicemi měření.

◇ Viz též: DVOUVÝBĚROVÝ TEST[7.12], TEST NEZÁVISLOSTI V NORMÁLNÍM ROZDĚLENÍ[7.16]

7.12. *Dvouvýběrový test.* **Dvouvýběrový t-test** je testem HYPOTÉZY[7.1] o hodnotě rozdílu STŘEDNÍCH HODNOT[2.9] μ_1 a μ_2 ve dvou NEZÁVISLÝCH[3.10] náhodných výběrech

$$X_1, \dots, X_n \quad \text{z NORMÁLNÍHO ROZDĚLENÍ[4.6] } \mathrm{N}(\mu_1, \sigma^2)$$

a

$$Y_1, \dots, Y_m \quad \text{z } \mathrm{N}(\mu_2, \sigma^2)$$

se stejným (i když neznámým) ROZPTYLEM[2.11] σ^2,

$$X_1, X_2, \dots\dots, X_n$$
$$Y_1, Y_2, \dots, Y_m$$

Pozorování pro dvouvýběrový test

$$H_0\colon \mu_1 - \mu_2 = d \quad \text{proti} \quad H_1\colon \mu_1 - \mu_2 \neq d,$$

kde $d \in \mathbf{R}$ je dané číslo (pro $d = 0$ to znamená shodu středních hodnot složek $\mu_1 = \mu_2$).

Použijeme-li k testování STATISTIKU[6.2]

$$T = \frac{\overline{X} - \overline{Y} - d}{\sqrt{(n-1)s_X^2 + (m-1)s_Y^2}} \sqrt{\frac{nm(n+m-2)}{n+m}},$$

kde s_X^2 a s_Y^2 jsou příslušné VÝBĚROVÉ ROZPTYLY[6.7], zamítneme H_0 na HLADINĚ VÝZNAMNOSTI[7.5] α, pokud

$$|T| > t_{1-\alpha/2}(n+m-2),$$

kde $t_p(n)$ je označení pro p-KVANTIL[2.14] rozdělení t_n[4.9].

Za platnosti H_0 má statistika T rozdělení t_{n+m-2}, KRITICKÝ OBOR[7.3] tvoří hodnoty T blízké $\pm\infty$.

Předpoklad nezávislosti testovaných výběrů (které mohou být různého rozsahu a měly by, alespoň přibližně, mít stejný rozptyl) je podstatný.

◇ Viz též: PÁROVÝ TEST[7.11]

výsledek č.	1	...	k
pozorované četnosti	n_1	...	n_k
očekávané četnosti	np_1	...	np_k

Chí-kvadrát test dobré shody

7.13. *Chí-kvadrát test dobré shody.* Pro testování HYPOTÉZY[7.1] o hodnotách PRAVDĚPODOBNOSTÍ[1.3] $p_1, \dots, p_k$, se kterými nastávají jednotlivé výsledky NÁHODNÉHO POKUSU[1.1] (o k možných výsledcích), na základě pozorování četností $n_1, \dots, n_k$ výsledků při n NEZÁVISLÝCH[1.5] opakováních pokusu

$$H_0\colon p_1 = p_1^0, \dots, p_k = p_k^0$$

proti

$$H_1\colon p_j \neq p_j^0 \quad \text{alespoň pro jedno } j,$$

kde $p_1^0, \dots, p_k^0 \in (0,1)$, $\sum_j p_j^0 = 1$, jsou daná čísla, se používá TEST[7.2] založený na STATISTICE[6.2]

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - o_j)^2}{o_j} = \sum_{i=1}^{k} \frac{n_j^2}{o_j} - n,$$

kde $o_j = np_j^0$.

Za předpokladu, že $o_j > 5$ pro všechna $j = 1, \dots, k$, zamítáme hypotézu H_0 přibližně na HLADINĚ VÝZNAMNOSTI[7.5] α, pokud

$$\chi^2 > \chi^2_{1-\alpha}(k-1),$$

kde $\chi^2_{1-\alpha}(k-1)$ je $(1-\alpha)$-KVANTIL[2.14] rozdělení χ^2_{k-1}[4.8].

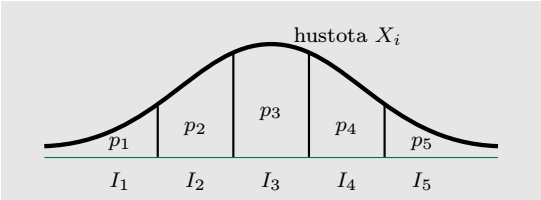
Není-li podmínka $o_j > 5 \quad \forall j$ splněna, je možné nejdříve některé výsledky (tj. četnosti a pravděpodobnosti) sloučit.

Za platnosti H_0 jsou o_j očekávanými četnostmi $\mathrm{E}\, X_j$ a statistika χ^2 má asymptoticky pro $n \to \infty$ rozdělení χ^2_{k-1}. KRITICKÝ OBOR[7.3] je tvořen hodnotami χ^2 blízkými ∞.

Pro testování hypotézy, že NÁHODNÝ VÝBĚR[6.1] $X_1, \dots, X_n$ pochází z daného ROZDĚLENÍ[2.2] (se známými parametry) rozdělíme obor hodnot VELIČIN[2.1] X_i na k intervalů $I_1, \dots, I_k$ a za výsledek i-tého pokusu považujeme zařazení hodnoty X_i do příslušného intervalu. Testujeme pak hypotézu o pravděpodobnostech

$$p_j = \mathrm{P}[X_i \in I_j], \quad j = 1, \dots, k,$$

na základě počtů pozorování $n_j = \#\{i; X_i \in I_j\}$, $j = 1, \dots, k$, v jednotlivých intervalech.



Testování shody s rozdělením

◇ Viz též: CHÍ-KVADRÁT TEST PŘI NEZNÁMÝCH PARAMETRECH[7.14]

7.14. *Chí-kvadrát test při neznámých parametrech.* Pro testování HYPOTÉZY[7.1] o hodnotách PRAVDĚPODOBNOSTÍ[1.3] $p_1, \dots, p_k$, se kterými nastávají jednotlivé výsledky NÁHODNÉHO POKUSU[1.1] (o k možných výsledcích) z pozorování $n_1, \dots, n_k$ četností výsledků při n NEZÁVISLÝCH[1.5] opakováních pokusu

$$H_0\colon p_j = p_j(\theta), \quad j = 1, \dots, k$$

proti

$$H_1\colon p_j \neq p_j(\theta) \quad \text{alespoň pro jedno } j,$$

kde $p_j(\theta)$ jsou dané funkce a $\theta \in \mathbf{R}^m$, $m < k-1$, neznámý parametr, se používá TEST[7.2] založený na STATISTICE[6.2] analogické té z testu PŘI ZNÁMÝCH PARAMETRECH[7.13],

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - o_j)^2}{o_j} = \sum_{i=1}^{k} \frac{n_j^2}{o_j} - n,$$

kde $o_j = np_j(\widehat{\theta})$ a $\widehat{\theta}$ je řešením soustavy rovnic

$$\sum_j \frac{n_j}{p_j(\widehat{\theta})} \frac{\partial p_j(\widehat{\theta})}{\partial \widehat{\theta}_i} = 0, \quad i = 1, \dots, m,$$

čili $\widehat{\theta}$ je MAXIMÁLNĚ VĚROHODNÝM[6.15] odhadem θ z pozorování $n_1, \dots, n_k$.

Předpokládá se, že v nějakém okolí skutečné hodnoty θ (včetně) jsou funkce $p_1, \dots, p_k$ kladné, $\sum_j p_j = 1$, a zobrazení $(p_1, \dots, p_k)\colon \mathbf{R}^m \to \mathbf{R}^k$ je dvakrát spojitě diferencovatelné a regulární. Statistika χ^2 pak má asymptoticky při $n \to \infty$ rozdělení χ^2_{k-m-1}.

Hypotézu H_0 zamítáme přibližně na HLADINĚ VÝZNAMNOSTI[7.5] α, pokud

$$\chi^2 > \chi^2_{1-\alpha}(k-m-1),$$

$\chi^2_p(n)$ značí p-KVANTIL[2.14] rozdělení χ^2_n[4.8].

7.15. *Test nezávislosti v kontingenční tabulce.* Pro testování HYPOTÉZY[7.1] o NEZÁVISLOSTI[3.10] složek X, Y NÁHODNÉHO VEKTORU[3.1] s DISKRÉTNÍM ROZDĚLENÍM[3.4], kde složky nabývají hodnot $x_1, \dots, x_r$ a $y_1, \dots, y_s$, $r, s \geqq 2$,

$$H_0\colon X, Y \text{ nezávislé} \quad \text{proti} \quad H_1\colon X, Y \text{ závislé},$$

se používá χ^2-TEST DOBRÉ SHODY PŘI NEZNÁMÝCH PARAMETRECH[7.14] vycházející z pozorovaných četností n_{ij}, $i = 1, \dots, r$, $j = 1, \dots, s$, dvojic hodnot (x_i, y_j) v NÁHODNÉM VÝBĚRU[6.1] rozsahu n z rozdělení (X, Y). (Tyto četnosti zapisujeme do tzv. **kontingenční tabulky**.) TEST[7.2] pracuje se STATISTIKOU[6.2]

	y_1	$\dots$	y_s	$\sum$
x_1	n_{11}	$\dots$	n_{1s}	$n_{1.}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
x_r	n_{r1}	$\dots$	n_{rs}	$n_{r.}$
$\sum$	$n_{.1}$	$\dots$	$n_{.s}$	n

Kontingenční tabulka

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(n_{ij} - o_{ij})^2}{o_{ij}} = n \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{n_{ij}^2}{n_{i.}n_{.j}} - n,$$

kde $o_{ij} = n_{i.}n_{.j}/n$, přičemž

$$n_{i.} = \sum_{j=1}^{s} n_{ij}, \quad n_{.j} = \sum_{i=1}^{r} n_{ij}.$$

jsou řádkové a sloupcové součty v tabulce.

Za předpokladu, že $o_{ij} > 5$ pro všechna i, j, zamítáme hypotézu H_0 přibližně na HLADINĚ VÝZNAMNOSTI[7.5] α, když

$$\chi^2 > \chi^2_{1-\alpha}((r-1)(s-1)),$$

$\chi^2_p(n)$ je označení pro p-KVANTIL[2.14] rozdělení χ^2_n[4.8].

Jsou-li některá $o_{ij} < 5$, je možné nejdříve sloučit příslušné řádky nebo sloupce se sousedními, počet řádků i sloupců se však nesmí zredukovat na jeden.

Za H_0 představují o_{ij} odhady očekávaných četností dvojic (x_i, y_j) a statistika χ^2 má asymptoticky (pro $n \to \infty$) rozdělení $\chi^2_{(r-1)(s-1)}$.

◇ Viz též: TEST NEZÁVISLOSTI V NORMÁLNÍM ROZDĚLENÍ[7.16]

7.16. *Test nezávislosti v normálním rozdělení.* Pro testování NEZÁVISLOSTI[3.10] složek NÁHODNÝCH VEKTORŮ[3.1] $(X_1, Y_1), \dots, (X_n, Y_n)$ v NÁHODNÉM VÝBĚRU[6.1] z DVOUROZMĚRNÉHO NORMÁLNÍHO ROZDĚLENÍ[4.7] vektoru (X, Y),

$$H_0\colon X, Y \text{ nezávislé} \quad \text{proti} \quad H_1\colon X, Y \text{ závislé},$$

se používá test založený na STATISTICE[6.2]

$$T = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2},$$

kde r je VÝBĚROVÝ KORELAČNÍ KOEFICIENT[6.10].

HYPOTÉZU[7.1] H_0 zamítáme na HLADINĚ VÝZNAMNOSTI[7.5] α, pokud

$$|T| > t_{1-\alpha/2}(n-2),$$

kde pravou stranu tvoří $(1-\alpha/2)$-KVANTIL[2.14] rozdělení t_{n-2}[4.9].

Hypotéza H_0 je díky normalitě ekvivalentní nulovosti KORELAČNÍHO KOEFICIENTU[3.9] mezi X a Y. Za platnosti H_0 má statistika T rozdělení t_{n-2}[4.9], KRITICKÝ OBOR[7.3] tvoří hodnoty T blízké $\pm\infty$.

◇ Viz též: TEST NEZÁVISLOSTI V KONTINGENČNÍ TABULCE[7.15], PÁROVÝ TEST[7.11]

8. REGRESE

8.1. *Regresní model.* **Regresním modelem** rozumíme model

$$Y_i = f(x_{i1}, \dots, x_{i\ell}; \beta_1, \dots, \beta_k) + e_i, \quad i = 1, \dots, n,$$

kde

- Y_i jsou NÁHODNÉ VELIČINY[2.1], jejichž hodnoty pozorujeme,
- f je známá funkce,
- x_{ij}, $i = 1, \dots, n$, $j = 1, \dots \ell$, známá čísla,
- $\beta_1, \dots, \beta_k$ parametry (které většinou neznáme) a
- $e_1, \dots, e_n$ náhodné chyby (náhodné veličiny s nulovou STŘEDNÍ HODNOTOU[2.9]).

Modelem se snažíme pomocí menšího počtu parametrů $\beta_1, \dots, \beta_k$ popsat vliv regresních konstant x_{ij} (**vysvětlující, nezávislé proměnné**) na hodnoty většího počtu náhodných veličin $Y_1, \dots, Y_n$ (**vysvětlované, závislé proměnné**).

Příkladem může být závislost lineární, $Y_i = \alpha + \beta x_i + e_i$, nebo exponenciální, $Y_i = \alpha e^{\beta x_i} + e_i$.

◇ Viz též: LINEÁRNÍ REGRESE[8.2], METODA NEJMENŠÍCH ČTVERCŮ[8.3]

8.2. *Lineární regrese.* **Lineární regresí** rozumíme REGRESNÍ MODEL[8.1] s regresní funkcí, která je lineární kombinací parametrů modelu,

$$Y_i = x_{i1}\beta_1 + \dots + x_{ik}\beta_k + e_i, \quad i = 1, \dots, n,$$

tj. zapsáno maticově

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e},$$

kde

$$\boldsymbol{Y}_{n\times 1} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \boldsymbol{X}_{n\times k} = \begin{pmatrix} x_{11} & \dots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{nk} \end{pmatrix},$$

$$\boldsymbol{\beta}_{k\times 1} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \boldsymbol{e}_{n\times 1} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$$

jsou vektor VYSVĚTLOVANÝCH POZOROVÁNÍ[8.1], regresní matice, vektor parametrů a chybový vektor s $\mathrm{E}\,\boldsymbol{e} = 0$.

Předpokládáme, že počet parametrů je menší než počet pozorování, $k < n$, a že hodnost matice $\boldsymbol{X}$ je k (tj. $\boldsymbol{X}$ má lineárně nezávislé sloupce — nezavádíme nadbytečné VYSVĚTLUJÍCÍ PROMĚNNÉ[8.1]).

Pokud některý sloupec matice $\boldsymbol{X}$ tvoří jednotkový vektor, hovoříme o modelu **s absolutním členem** (jímž je příslušná složka vektoru $\boldsymbol{\beta}$).

◇ Viz též: ODHADY V REGRESI[8.5], TEST VÝZNAMNOSTI REGRESE[8.7]

8.3. *Metoda nejmenších čtverců.* ODHADEM[6.3] parametrů REGRESNÍHO MODELU[8.1] $Y_i = f(x_{i1}, \dots, x_{i\ell}; \beta_1, \dots, \beta_k) + e_i$, $i = 1, \dots, n$, **metodou nejmenších čtverců** nazýváme ty hodnoty $\boldsymbol{b} = (b_1, \dots, b_k)'$ parametrů $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$, které minimalizují **součet čtverců**

$$S(\beta_1, \dots, \beta_k) = \sum_{i=1}^{n} (Y_i - f(x_{i1}, \dots, x_{i\ell}; \beta_1, \dots, \beta_k))^2.$$

Za předpokladu hladkosti S je tento odhad řešením soustavy **normálních rovnic** $\partial S / \partial \beta_i = 0$, $i = 1, \dots, k$.

Hodnoty $\widehat{Y}_i = f(x_{i1}, \dots, x_{i\ell}; b_1, \dots, b_k)$, $i = 1, \dots, n$, označujeme jako **vyrovnané hodnoty** pozorování Y_i, $i = 1, \dots, n$.

◇ Viz též: ODHADY V REGRESI[8.5], REZIDUÁLNÍ SOUČET ČTVERCŮ[8.4]

8.4. *Reziduální součet čtverců.* **Reziduálním součtem čtverců** S_e rozumíme minimální hodnotu SOUČTU ČTVERCŮ[8.3] S při odhadování parametrů REGRESNÍHO MODELU[8.1] pomocí METODY NEJMENŠÍCH ČTVERCŮ[8.3],

$$S_e = (\boldsymbol{Y} - \widehat{\boldsymbol{Y}})'(\boldsymbol{Y} - \widehat{\boldsymbol{Y}}),$$

kde $\boldsymbol{Y}$ a $\widehat{\boldsymbol{Y}}$ jsou pozorované, resp. VYROVNANÉ[8.3] hodnoty ZÁVISLÉ PROMĚNNÉ[8.1] $\boldsymbol{Y}$.

S_e udává modelem nevysvětlenou část z původního **celkového součtu čtverců**

$$S_t = (\boldsymbol{Y} - \overline{Y})'(\boldsymbol{Y} - \overline{Y}),$$

kde $\overline{Y}$ je aritmetický průměr složek $\boldsymbol{Y}$.

◇ Viz též: KOEFICIENT DETERMINACE[8.6]

8.5. *Odhady v regresi.* V modelu LINEÁRNÍ REGRESE[8.2] $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ má ODHAD[6.3] parametrů $\boldsymbol{\beta}$ METODOU NEJMENŠÍCH ČTVERCŮ[8.3] tvar

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{Y}$$

a REZIDUÁLNÍ SOUČET ČTVERCŮ[8.4] je

$$S_e = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}) = \boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{b}'\boldsymbol{X}'\boldsymbol{Y}.$$

Jsou-li složky vektoru chyb $\boldsymbol{e}$ NEKORELOVANÉ[3.9] a se stejným ROZPTYLEM[2.11] σ^2, tj. $\mathrm{E}\,\boldsymbol{e} = 0$, $\operatorname{var}\boldsymbol{e} = \sigma^2 I$, je

$$\mathrm{E}\,\boldsymbol{b} = \boldsymbol{\beta}, \quad \operatorname{var}\boldsymbol{b} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}, \quad \mathrm{E}\,\frac{S_e}{n-k} = \sigma^2,$$

tedy $\boldsymbol{b}$ a $s^2 = S_e/(n-k)$ jsou NESTRANNÝMI[6.11] odhady $\boldsymbol{\beta}$ a σ^2. Odhad $\boldsymbol{b}$ je dokonce NEJLEPŠÍ[6.14] mezi lineárními (ve složkách vektoru $\boldsymbol{Y}$) nestrannými odhady $\boldsymbol{\beta}$.

◇ Viz též: TEST VÝZNAMNOSTI REGRESE[8.7]

8.6. *Koeficient determinace.* **Koeficientem determinace** rozumíme veličinu

$$R^2 = 1 - \frac{S_e}{S_t},$$

kde S_e je REZIDUÁLNÍ[8.4] a S_t CELKOVÝ[8.4] součet čtverců.

V modelu LINEÁRNÍ REGRESE[8.2] s ABSOLUTNÍM ČLENEM[8.2] leží hodnota R^2 v intervalu $\langle 0, 1 \rangle$ a udává, jaký podíl ROZPTYLU[6.7] v pozorování ZÁVISLÉ PROMĚNNÉ[8.1] se podařilo REGRESÍ[8.1] vysvětlit (větší hodnoty znamenají větší úspěšnost regrese).

◇ Viz též: TEST VÝZNAMNOSTI REGRESE[8.7]

8.7. *Test významnosti regrese.* Předpokládejme, že v modelu LINEÁRNÍ REGRESE[8.2] má chybový vektor $(e_1, \dots, e_n)$ NORMÁLNÍ ROZDĚLENÍ[4.7] $\mathrm{N}_n(0, \sigma^2 I)$.

V modelu s ABSOLUTNÍM ČLENEM[8.2] β_1, kde $\beta_2, \dots, \beta_k$ jsou ostatní parametry, lze pro test HYPOTÉZY[7.1]

$$H_0 \colon \beta_2 = \dots = \beta_k = 0$$

proti

$$H_1 \colon \beta_j \neq 0 \text{ pro některé } j = 2, \dots, k$$

použít STATISTIKU[6.2]

$$F = \frac{R^2}{1 - R^2} \frac{n-k}{k-1},$$

kde R^2 je KOEFICIENT DETERMINACE[8.6]. Hypotézu H_0 zamítneme na HLADINĚ VÝZNAMNOSTI[7.5] α, když

$$F > F_{1-\alpha}(k-1, n-k),$$

kde $F_{1-\alpha}(k-1, n-k)$ je $(1-\alpha)$-KVANTIL[2.14] rozdělení $F_{k-1,n-1}$[4.10] (což je ROZDĚLENÍ[2.2] statistiky F za platnosti H_0). Zamítnutí znamená, že vliv VYSVĚTLUJÍCÍCH PROMĚNNÝCH[8.1] je významný.

Existují i TESTY[7.2] pro HYPOTÉZY[7.1] o nulovosti jednoho nebo několika parametrů $\beta_1, \dots, \beta_k$, či jejich lineární kombinace.

SEZNAM HESEL

Abecední seznam číslovaných hesel.

Alternativní rozdělení[4.1]
Aproximace binomického rozdělení Poissonovým[4.4]
Bayesova věta[1.8]
Binomické rozdělení[4.2]
Centrální limitní věta[5.4]
Centrální limitní věta pro binomické rozdělení[5.6]
Diskrétní rozdělení[2.4]
Diskrétní sdružené rozdělení[3.4]
Distribuční funkce[2.3]
Dvouvýběrový test[7.12]
Exponenciální rozdělení[4.5]
Fisherovo-Snedecorovo F-rozdělení[4.10]
Hladina významnosti[7.5]
Hustota[2.7]
Hypotéza[7.1]
Chí-kvadrát rozdělení[4.8]
Chí-kvadrát test dobré shody[7.13]
Chí-kvadrát test při neznámých parametrech[7.14]
Chyba 1. a 2. druhu[7.4]
Intenzita poruch[2.8]
Intervalové odhady[6.16]
Intervaly spolehlivosti v normálním rozdělení[6.17]
Klasická definice pravděpodobnosti[1.4]
Koeficient determinace[8.6]
Kolmogorovův zákon velkých čísel[5.2]
Konzistence[6.12]
Korelační koeficient[3.9]
Kovariance[3.7]
Kritický obor[7.3]
Kvantil[2.14]
Lineární regrese[8.2]
Ljapunovova centrální limitní věta[5.5]
Marginální rozdělení[3.3]
Maximálně věrohodný odhad[6.15]
Medián[2.13]
Metoda nejmenších čtverců[8.3]
Momenty[2.10]
Náhodná veličina[2.1]
Náhodný jev[1.1]
Náhodný vektor[3.1]
Náhodný výběr[6.1]
Nejlepší odhad[6.14]
Nestrannost[6.11]
Nezávislé náhodné veličiny[3.10]
Nezávislost jevů[1.5]
Normální rozdělení[4.6]
Odhad[6.3]
Odhady v regresi[8.5]
P-hodnota[7.7]
Párový test[7.11]
Podmíněná pravděpodobnost[1.6]
Poissonovo rozdělení[4.3]
Popisné statistiky[6.4]
Pravděpodobnostní funkce[2.5]
Pravděpodobnostní míra[1.3]
Regresní model[8.1]

Reziduální součet čtverců[8.4]
Rozdělení[2.2]
Rozptyl[2.11]
Sdružené rozdělení[3.2]
Sigma-algebra[1.2]
Síla testu[7.6]
Slabý zákon velkých čísel[5.3]
Směrodatná odchylka[2.12]
Spojité rozdělení[2.6]
Spojité sdružené rozdělení[3.5]
Statistika (hodnota)[6.2]
Střední čtvercová chyba[6.13]
Střední hodnota[2.9]
Střední hodnota vektoru[3.6]
Studentovo t-rozdělení[4.9]
Test[7.2]
Test nezávislosti v kontingenční tabulce[7.15]
Test nezávislosti v normálním rozdělení[7.16]
Test o rozptylu normálního rozdělení[7.10]
Test o střední hodnotě[7.9]
Test významnosti regrese[8.7]
Testy střední hodnoty normálního rozdělení[7.8]
Varianční matice[3.8]
Věta o celkové pravděpodobnosti[1.7]
Vícerozměrné normální rozdělení[4.7]
Výběrová kovariance[6.9]
Výběrová směrodatná odchylka[6.8]
Výběrový korelační koeficient[6.10]
Výběrový medián[6.6]
Výběrový průměr[6.5]
Výběrový rozptyl[6.7]
Zákon velkých čísel[5.1]

SEZNAM OBRÁZKŮ

Abecední seznam obrázků.

Distribuční funkce[2.3]
Distribuční funkce diskrétního rozdělení[2.4]
Hladina významnosti[7.5]
Hustota exponenciálního rozdělení[4.5]
Hustota chí-kvadrát rozdělení[4.8]
Hustota normálního rozdělení[4.6]
Hustota rozdělení[2.7]
Hustota t-rozdělení[4.9]
Chí-kvadrát test dobré shody[7.13]
Chyby 1. a 2. druhu a jejich pravděpodobnosti[7.4]
Kontingenční tabulka[7.15]
Kritický obor[7.3]
Kritický obor oboustranného testu[7.9]
Kvantil[2.14]
Náhodná veličina[2.1]
Nulová a alternativní hypotéza[7.2]
P-hodnota[7.7]
Podmíněná a nepodmíněná pravděpodobnost[1.6]
Pokrytí parametru intervalovým odhadem[6.16]
Porovnání odhadů[6.14]
Pozorování pro dvouvýběrový test[7.12]
Pozorování pro párový test[7.11]
Pravděpodobnost sjednocení[1.3]
Pravděpodobnostní a distribuční funkce[2.5]
Přenos pravděpodobnosti[2.2]
Rozklad pravděpodobnosti jevu[1.7]
Sdružená a marginální pravděpodobnostní funkce[3.4]
Síla testu[7.6]
Testování shody s rozdělením[7.13]
Varianční matice[3.8]

INDEX

Stránky s výskytem položek

alternativní hypotéza, $24^{7.4}$, $\mathbf{24}^{7.2}$, $25^{7.6}$
aproximace binomického rozdělení Poissonovým , $\mathbf{16}^{4.4}$

Bayesova věta, $\mathbf{9}^{1.8}$

centrální limitní věta, $\mathbf{19}^{5.4}$, $20^{5.6}$, $24^{6.17}$, $26^{7.9}$
  Lindebergova-Lévyova, $\mathbf{19}^{5.4}$, $20^{5.5}$
  Ljapunovova, $\mathbf{20}^{5.5}$

Čebyševova nerovnost, $\mathbf{12}^{2.11}$

diskrétní rozdělení, $10^{2.5}$, $\mathbf{10}^{2.4}$, $11^{2.9}$, $13^{3.4}$, $\mathbf{13}^{3.4}$, $14^{3.6}$, $15^{4.1}$, $16^{4.2}$, $16^{4.3}$, $29^{7.15}$
distribuční funkce, $\mathbf{9}^{2.3}$, $10^{2.4}$, $10^{2.5}$, $10^{2.6}$, $11^{2.7}$, $11^{2.8}$, $12^{2.13}$, $13^{2.14}$, $16^{4.4}$, $16^{4.5}$, $16^{4.3}$, $17^{4.6}$, $19^{5.4}$, $20^{5.5}$, $20^{6.1}$
  sdružená, $13^{3.3}$, $\mathbf{13}^{3.2}$, $14^{3.5}$, $15^{3.10}$

hladina významnosti, $25^{7.7}$, $25^{7.6}$, $25^{7.8}$, $\mathbf{25}^{7.5}$, $26^{7.9}$, $26^{7.10}$, $27^{7.12}$, $27^{7.11}$, $28^{7.13}$, $29^{7.14}$, $29^{7.15}$, $29^{7.16}$, $32^{8.7}$
hustota, $11^{2.8}$, $\mathbf{11}^{2.7}$, $12^{2.9}$, $14^{3.5}$, $15^{3.10}$, $16^{4.5}$, $17^{4.6}$, $18^{4.10}$, $18^{4.8}$, $18^{4.9}$, $23^{6.15}$
  sdružená, $14^{3.6}$, $\mathbf{14}^{3.5}$, $17^{4.7}$
hypotéza, $24^{7.4}$, $24^{7.3}$, $24^{7.2}$, $\mathbf{24}^{7.1}$, $25^{7.5}$, $25^{7.7}$, $25^{7.8}$, $26^{7.9}$, $26^{7.10}$, $27^{7.12}$, $27^{7.11}$, $28^{7.14}$, $28^{7.13}$, $29^{7.15}$, $29^{7.16}$, $32^{8.7}$

chyba
  1. druhu, $\mathbf{24}^{7.4}$, $25^{7.5}$, $25^{7.6}$
  2. druhu, $\mathbf{24}^{7.4}$, $25^{7.6}$

intenzita poruch, $\mathbf{11}^{2.8}$, $16^{4.5}$
interval spolehlivosti, $23^{6.17}$, $\mathbf{23}^{6.16}$
intervalový odhad, $\mathbf{23}^{6.16}$

jednoduchá hypotéza, $\mathbf{24}^{7.1}$
jev
  elementární, $\mathbf{7}^{1.1}$
  jistý, $\mathbf{7}^{1.1}$
  náhodný, $7^{1.4}$, $7^{1.3}$, $7^{1.2}$, $\mathbf{7}^{1.1}$, $8^{1.7}$, $8^{1.5}$, $9^{1.8}$, $18^{5.1}$
  nemožný, $\mathbf{7}^{1.1}$

koeficient determinace, $\mathbf{31}^{8.6}$, $32^{8.7}$
kontingenční tabulka, $\mathbf{29}^{7.15}$
konzistentní, $21^{6.5}$, $21^{6.7}$, $\mathbf{22}^{6.12}$, $24^{6.17}$, $26^{7.9}$
korelační koeficient, $15^{3.10}$, $\mathbf{15}^{3.9}$, $17^{4.7}$, $19^{5.3}$, $22^{6.10}$, $30^{7.16}$, $31^{8.5}$
kovariance, $\mathbf{14}^{3.7}$, $15^{3.10}$, $15^{3.8}$, $22^{6.9}$
kritický obor, $24^{7.4}$, $\mathbf{24}^{7.3}$, $25^{7.6}$, $25^{7.8}$, $26^{7.9}$, $26^{7.10}$, $27^{7.12}$, $28^{7.13}$, $30^{7.16}$
kvantil, $\mathbf{13}^{2.14}$, $17^{4.6}$, $23^{6.17}$, $25^{7.8}$, $26^{7.9}$, $26^{7.10}$, $27^{7.12}$, $27^{7.11}$, $28^{7.13}$, $29^{7.14}$, $29^{7.15}$, $29^{7.16}$, $32^{8.7}$

lepší, $\mathbf{23}^{6.14}$

marginální rozdělení, $13^{3.4}$, $\mathbf{13}^{3.3}$, $14^{3.5}$, $15^{3.10}$
medián, $\mathbf{12}^{2.13}$, $13^{2.14}$, $21^{6.6}$
měřitelnost, $\mathbf{9}^{2.1}$, $11^{2.9}$, $15^{3.10}$, $20^{6.2}$
moment, $12^{2.11}$, $\mathbf{12}^{2.10}$

náhodná veličina, $9^{2.3}$, $9^{2.2}$, $\mathbf{9}^{2.1}$, $10^{2.4}$, $10^{2.5}$, $10^{2.6}$, $11^{2.9}$, $11^{2.7}$, $11^{2.8}$, $12^{2.10}$, $12^{2.11}$, $13^{3.1}$, $14^{3.7}$, $14^{3.6}$, $15^{4.1}$, $15^{3.10}$, $16^{4.2}$, $16^{4.3}$, $17^{4.5}$, $17^{4.6}$, $18^{4.10}$, $18^{4.8}$, $18^{4.9}$, $18^{5.1}$, $19^{5.4}$, $19^{5.2}$, $19^{5.3}$, $20^{5.6}$, $20^{5.5}$, $20^{6.1}$, $21^{6.5}$, $21^{6.7}$, $22^{6.9}$, $23^{6.17}$, $23^{6.16}$, $24^{7.1}$, $24^{7.4}$, $24^{7.3}$, $24^{7.2}$, $28^{7.13}$, $30^{8.1}$
náhodný
  pokus, $7^{1.2}$, $\mathbf{7}^{1.1}$, $16^{4.2}$, $18^{5.1}$, $28^{7.14}$, $28^{7.13}$
náhodný vektor, $13^{3.4}$, $13^{3.3}$, $13^{3.2}$, $\mathbf{13}^{3.1}$, $14^{3.7}$, $14^{3.6}$, $14^{3.5}$, $15^{3.9}$, $15^{3.10}$, $15^{3.8}$, $17^{4.7}$, $20^{6.1}$, $27^{7.11}$, $29^{7.15}$, $29^{7.16}$
náhodný výběr, $20^{6.4}$, $20^{6.2}$, $\mathbf{20}^{6.1}$, $21^{6.6}$, $21^{6.5}$, $21^{6.7}$,

$22^{6.10}$, $22^{6.9}$, $23^{6.16}$, $23^{6.15}$, $24^{7.1}$, $24^{7.2}$, $25^{7.4}$, $25^{7.8}$, $26^{7.9}$, $26^{7.10}$, $27^{7.11}$, $28^{7.13}$, $29^{7.15}$, $29^{7.16}$
nejlepší odhad, $21^{6.5}$, $22^{6.13}$, $\mathbf{23}^{6.14}$, $31^{8.5}$
neslučitelné jevy, $7^{1.3}$, $\mathbf{7}^{1.1}$
nestranný, $21^{6.5}$, $21^{6.7}$, $\mathbf{22}^{6.11}$, $23^{6.15}$, $31^{8.5}$
nezávislé, $\mathbf{15}^{3.10}$, $16^{4.2}$, $17^{4.7}$, $18^{4.10}$, $18^{4.8}$, $18^{4.9}$, $18^{5.1}$, $19^{5.4}$, $19^{5.2}$, $19^{5.3}$, $20^{5.6}$, $20^{5.5}$, $20^{6.1}$, $27^{7.12}$, $29^{7.15}$, $29^{7.16}$
nezávislost
  jevů, $8^{1.6}$, $\mathbf{8}^{1.5}$, $16^{4.2}$, $28^{7.14}$, $28^{7.13}$
normální rovnice, $\mathbf{31}^{8.3}$
nulová hypotéza, $\mathbf{24}^{7.2}$

odhad, $\mathbf{20}^{6.3}$, $21^{6.7}$, $22^{6.12}$, $22^{6.13}$, $22^{6.11}$, $23^{6.16}$, $23^{6.14}$, $31^{8.3}$, $31^{8.5}$
  metodou maximální věrohodnosti, $\mathbf{23}^{6.15}$, $29^{7.14}$
  metodou nejmenších čtverců, $31^{8.5}$, $31^{8.4}$, $\mathbf{31}^{8.3}$

p-hodnota, $\mathbf{25}^{7.7}$
popisné statistiky, $\mathbf{20}^{6.4}$
pravděpodobnost
  celková, $\mathbf{8}^{1.7}$, $9^{1.8}$
  klasická, $\mathbf{7}^{1.4}$
  podmíněná, $8^{1.7}$, $\mathbf{8}^{1.6}$
  statistická definice, $\mathbf{19}^{5.1}$
pravděpodobnostní funkce, $\mathbf{10}^{2.5}$, $11^{2.9}$, $13^{3.4}$, $14^{3.6}$, $15^{4.1}$, $16^{4.4}$, $16^{4.2}$, $16^{4.3}$
pravděpodobnostní míra, $7^{1.1}$, $\mathbf{7}^{1.3}$, $8^{1.7}$, $8^{1.4}$, $8^{1.5}$, $8^{1.6}$, $9^{1.8}$, $9^{2.1}$, $9^{2.2}$, $10^{2.3}$, $11^{2.8}$, $12^{2.11}$, $16^{4.2}$, $18^{5.1}$, $19^{5.4}$, $19^{5.2}$, $19^{5.3}$, $25^{7.4}$, $28^{7.14}$, $28^{7.13}$
pravděpodobnostní prostor, $\mathbf{7}^{1.3}$, $9^{2.1}$, $9^{2.2}$, $12^{2.9}$, $13^{3.1}$, $15^{3.10}$

regrese, $30^{8.2}$, $\mathbf{30}^{8.1}$, $31^{8.6}$, $31^{8.3}$, $31^{8.4}$
  lineární, $\mathbf{30}^{8.2}$, $31^{8.6}$, $31^{8.5}$, $32^{8.7}$
  s absolutním členem, $\mathbf{30}^{8.2}$, $31^{8.6}$, $32^{8.7}$
rozdělení, $9^{2.3}$, $\mathbf{9}^{2.2}$, $11^{2.7}$, $11^{2.8}$, $12^{2.13}$, $12^{2.10}$, $12^{2.11}$, $13^{2.14}$, $13^{3.3}$, $18^{4.10}$, $18^{4.8}$, $18^{4.9}$, $18^{5.1}$, $19^{5.4}$, $19^{5.2}$, $20^{5.5}$, $20^{6.1}$, $20^{6.4}$, $24^{7.1}$, $24^{6.17}$, $25^{7.4}$, $25^{7.8}$, $28^{7.13}$, $32^{8.7}$
  alternativní, $\mathbf{15}^{4.1}$, $16^{4.2}$, $20^{5.6}$
  binomické, $16^{4.4}$, $\mathbf{16}^{4.2}$, $20^{5.6}$, $21^{6.5}$
  exponenciální, $\mathbf{16}^{4.5}$, $21^{6.5}$
  F, $\mathbf{18}^{4.10}$, $32^{8.7}$
  chí-kvadrát, $18^{4.10}$, $18^{4.9}$, $\mathbf{18}^{4.8}$, $23^{6.17}$, $26^{7.10}$, $28^{7.13}$, $29^{7.14}$, $29^{7.15}$
  normální, $17^{4.7}$, $\mathbf{17}^{4.6}$, $18^{4.8}$, $18^{4.9}$, $19^{5.4}$, $20^{5.5}$, $21^{6.5}$, $23^{6.17}$, $25^{7.8}$, $26^{7.9}$, $26^{7.10}$, $27^{7.12}$, $27^{7.11}$
    vícerozměrné, $\mathbf{17}^{4.7}$, $27^{7.11}$, $29^{7.16}$, $32^{8.7}$
  Poissonovo, $16^{4.4}$, $\mathbf{16}^{4.3}$, $21^{6.5}$
  t, $\mathbf{18}^{4.9}$, $23^{6.17}$, $25^{7.8}$, $27^{7.12}$, $27^{7.11}$, $29^{7.16}$, $30^{7.16}$
rozptyl, $12^{2.12}$, $\mathbf{12}^{2.11}$, $14^{3.7}$, $15^{4.1}$, $15^{3.10}$, $15^{3.8}$, $16^{4.2}$, $16^{4.3}$, $17^{4.5}$, $17^{4.6}$, $19^{5.4}$, $19^{5.2}$, $19^{5.3}$, $21^{6.7}$, $22^{6.12}$, $22^{6.11}$, $24^{6.17}$, $25^{7.8}$, $26^{7.10}$, $27^{7.12}$, $31^{8.5}$

sdružené rozdělení, $\mathbf{13}^{3.2}$, $17^{4.7}$
sigma-algebra, $7^{1.3}$, $\mathbf{7}^{1.2}$, $8^{1.4}$, $9^{2.1}$, $9^{2.2}$
  borelovská, $\mathbf{9}^{2.2}$
síla testu, $\mathbf{25}^{7.6}$
složená hypotéza, $\mathbf{24}^{7.1}$
směrodatná odchylka, $\mathbf{12}^{2.12}$, $17^{4.6}$
součet čtverců, $31^{8.4}$, $\mathbf{31}^{8.3}$
  celkový, $31^{8.6}$, $\mathbf{31}^{8.4}$
  reziduální, $31^{8.6}$, $31^{8.5}$, $\mathbf{31}^{8.4}$
spojité rozdělení, $\mathbf{10}^{2.6}$, $11^{2.7}$, $11^{2.8}$, $12^{2.9}$, $12^{2.13}$, $14^{3.6}$, $14^{3.5}$, $\mathbf{14}^{3.5}$, $15^{3.10}$, $16^{4.5}$, $17^{4.6}$, $18^{4.10}$, $18^{4.8}$, $18^{4.9}$
statistika, $20^{6.3}$, $\mathbf{20}^{6.2}$, $21^{6.7}$, $24^{7.3}$, $25^{7.7}$, $25^{7.8}$, $26^{7.9}$, $26^{7.10}$, $27^{7.12}$, $28^{7.14}$,

$28^{7.13}$, $29^{7.15}$, $29^{7.16}$, $32^{8.7}$
střední čtvercová chyba, $\mathbf{22}^{6.13}$, $23^{6.14}$
střední hodnota, $\mathbf{11}^{2.9}$, $12^{2.13}$, $12^{2.10}$, $12^{2.11}$, $14^{3.6}$, $15^{4.1}$, $16^{4.2}$, $16^{4.3}$, $17^{4.5}$, $17^{4.6}$, $18^{5.1}$, $19^{5.4}$, $19^{5.3}$, $21^{6.5}$, $22^{6.12}$, $22^{6.13}$, $22^{6.11}$, $24^{6.17}$, $25^{7.8}$, $26^{7.9}$, $27^{7.12}$, $27^{7.11}$, $30^{8.1}$
  vektoru, $\mathbf{14}^{3.6}$, $17^{4.7}$

t-test, $\mathbf{25}^{7.8}$, $26^{7.9}$, $27^{7.11}$
test, $24^{7.4}$, $24^{7.3}$, $\mathbf{24}^{7.2}$, $25^{7.5}$, $25^{7.7}$, $25^{7.6}$, $25^{7.8}$, $26^{7.10}$, $27^{7.11}$, $28^{7.14}$, $28^{7.13}$, $29^{7.15}$, $32^{8.7}$
  chí-kvadrát, $28^{7.14}$, $\mathbf{28}^{7.13}$
    při neznámých parametrech, $\mathbf{28}^{7.14}$, $29^{7.15}$
  nezávislosti
    v normálním rozdělení, $\mathbf{29}^{7.16}$
  párový, $\mathbf{27}^{7.11}$
  střední hodnoty
    dvouvýběrový, $\mathbf{27}^{7.12}$
    normálního rozdělení, $\mathbf{25}^{7.8}$
test o rozptylu, $\mathbf{26}^{7.10}$
varianční matice, $\mathbf{15}^{3.8}$, $17^{4.7}$
výběrová kovariance, $20^{6.4}$, $22^{6.10}$, $\mathbf{22}^{6.9}$
výběrová směrodatná odchylka, $20^{6.4}$, $\mathbf{21}^{6.8}$
výběrový korelační koeficient, $20^{6.4}$, $\mathbf{22}^{6.10}$, $29^{7.16}$
výběrový medián, $20^{6.4}$, $\mathbf{21}^{6.6}$
výběrový průměr, $19^{5.3}$, $20^{6.4}$, $21^{6.7}$, $\mathbf{21}^{6.5}$, $22^{6.12}$, $22^{6.11}$, $22^{6.10}$, $22^{6.9}$, $23^{6.17}$, $25^{7.8}$, $26^{7.9}$, $27^{7.11}$
výběrový rozptyl, $20^{6.4}$, $21^{6.8}$, $\mathbf{21}^{6.7}$, $22^{6.12}$, $22^{6.11}$, $22^{6.10}$, $23^{6.17}$, $25^{7.8}$, $26^{7.9}$, $26^{7.10}$, $27^{7.12}$, $27^{7.11}$, $31^{8.6}$
vychýlený, $21^{6.7}$, $\mathbf{22}^{6.11}$
vyrovnané hodnoty, $31^{8.4}$, $\mathbf{31}^{8.3}$
vysvětlující proměnná, $30^{8.2}$, $\mathbf{30}^{8.1}$, $32^{8.7}$

zákon velkých čísel, $\mathbf{18}^{5.1}$, $19^{5.2}$
  Kolmogorovův, $\mathbf{19}^{5.2}$
  silný, $\mathbf{18}^{5.1}$, $19^{5.3}$
  slabý, $\mathbf{19}^{5.3}$
závislá proměnná, $30^{8.2}$, $\mathbf{30}^{8.1}$, $31^{8.6}$, $31^{8.4}$
závislost
  jevů, $\mathbf{8}^{1.5}$