MV011 Statistika I

7. Průzkumová analýza dat

Jan Koláček (kolacek@math.muni.cz)

Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita, Brno

Průzkumová analýza jednorozměrných dat
Exploratory data analysis

Průzkumová analýza dat je odvětví statistiky, které pomocí různých postupů odhaluje zvláštnosti v datech. Při zpracování dat se často používají metody, které jsou založeny na předpokladu, že data pocházejí z nějakého konkrétního rozložení, nejčastěji normálního. Tento předpoklad nemusí být vždy splněn, protože data

- mohou pocházet z jiného rozložení
- mohou být zatížena hrubými chybami
- mohou pocházet ze směsi několika rozložení.

Proto je důležité provést průzkumovou analýzu dat, abychom se vyvarovali neadekvátního použití statistických metod.

Funkcionální charakteristiky datového souboru

Označení

Na množině objektů $\{\varepsilon_1, \ldots, \varepsilon_n\}$ zjišťujeme hodnoty znaku X. Hodnotu znaku X na objektu ε_i označíme $x_i, i = 1, \ldots, n$. V teorii pravděpodobnosti se jim také říká **realizace** náhodné veličiny X. Tyto hodnoty zaznamenáme do jednorozměrného datového souboru:

$$\mathbf{x} = (x_1, \ldots, x_n)'.$$

Uspořádané hodnoty $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ tvoří uspořádaný datový soubor:

$$\mathbf{x}_{(\cdot)} = (x_{(1)}, \ldots, x_{(n)})'.$$

Vektor

$$\mathbf{x}_{[\cdot]} = (x_{[1]}, \ldots, x_{[r]})',$$

kde $x_{[1]} < \ldots < x_{[r]}$, $r \leq n$, jsou navzájem různé hodnoty znaku X, se nazývá **vektor variant** (**levels**).

Bodové rozložení četností

indikátor množiny:

$$I_B(x) = \begin{cases} 1 & x \in B, \\ 0 & x \notin B. \end{cases}$$

Pro datový soubor $\mathbf{x} = (x_1, \ldots, x_n)'$ definujeme následující pojmy

- **absolutní četnost** (**absolute frequency**) varianty $x_{[j]}$:

$$n_j = \sum_{i=1}^{n} I_{\{x_{[j]}\}}(x_i)$$

- **relativní četnost** (**relative frequency**) varianty $x_{[j]}$:

$$p_j = \frac{n_j}{n}$$

- **absolutní kumulativní četnost** prvních j variant:

$$N_j = n_1 + \ldots + n_j$$

- **relativní kumulativní četnost** prvních j variant:

$$F_j = \frac{N_j}{n} = p_1 + \ldots + p_j$$

- **četnostní funkce**:

$$p(x) = \begin{cases} p_j & \text{pro } x = x_{[j]},\ j = 1, \ldots, r \\ 0 & \text{jinak} \end{cases}$$

- **empirická distribuční funkce** (**empirical distribution function**):

$$F(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x>}(x_i)$$

Absolutní či relativní četnosti znázorňujeme graficky např. pomocí sloupkového diagramu či polygonu četností.

Příklad

U 30 domácností byl zjišťován počet členů.

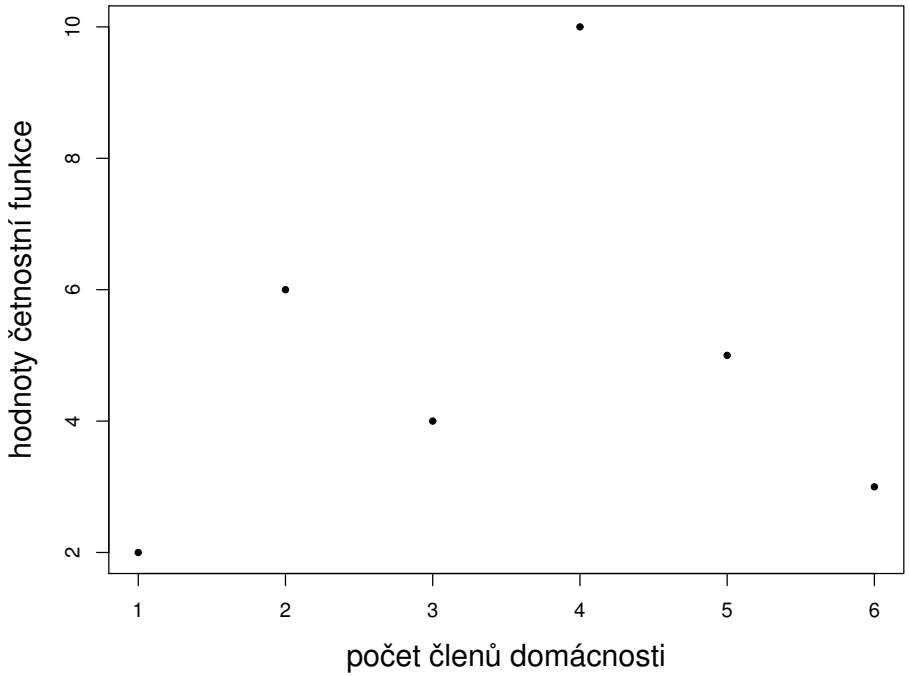
Počet členů	1	2	3	4	5	6
Počet domácností	2	6	4	10	5	3

Vytvořte tabulku rozložení četností. Nakreslete grafy četnostní funkce a empirické distribuční funkce. Dále nakreslete sloupkový diagram a polygon četností počtu členů domácnosti.

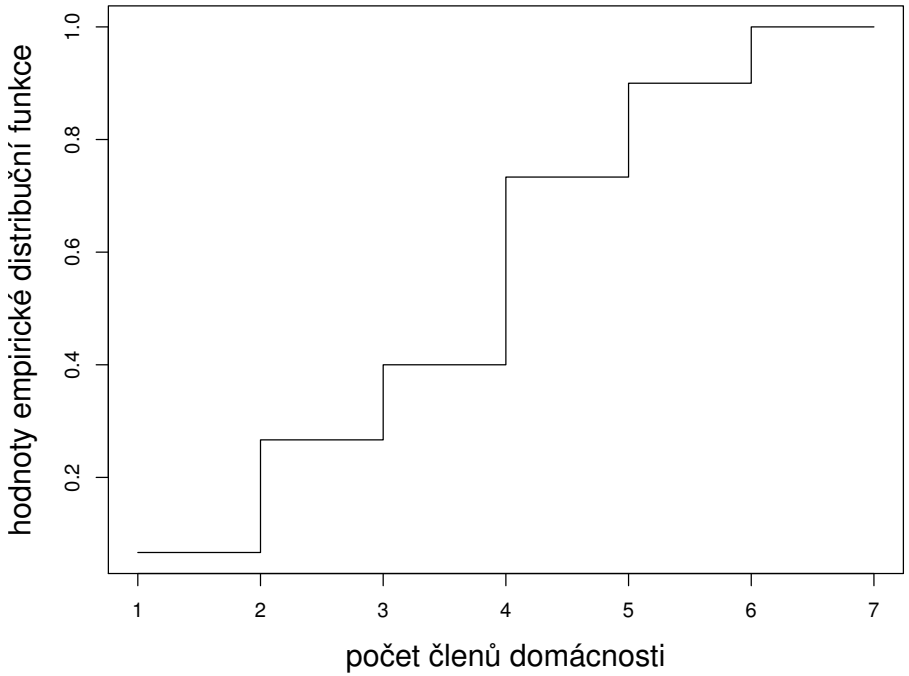
Řešení. Tabulka rozložení četností:

$x_{[j]}$	n_j	p_j	N_j	F_j
1	2	$2/30$	2	$2/30$
2	6	$6/30$	8	$8/30$
3	4	$4/30$	12	$12/30$
4	10	$10/30$	22	$22/30$
5	5	$5/30$	27	$27/30$
6	3	$3/30$	30	1

Příklad – pokračování

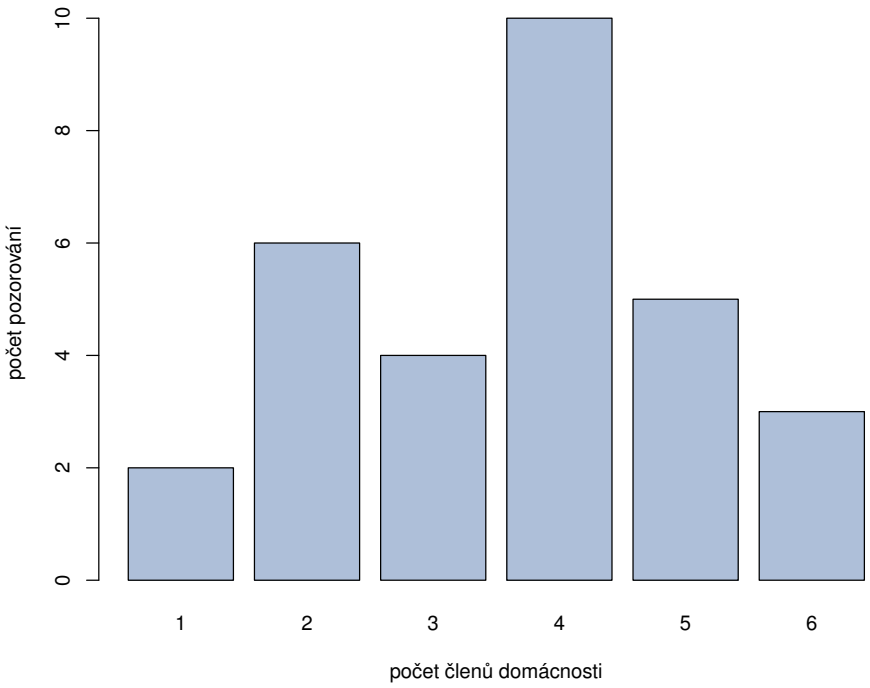


Obr.: Graf četnostní funkce

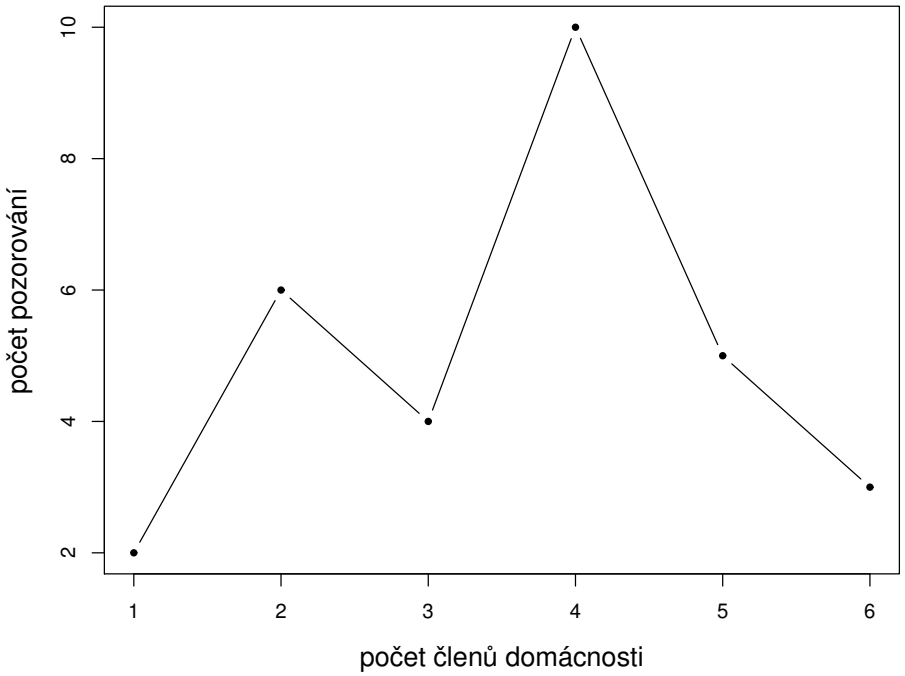


Obr.: Graf empirické distribuční funkce

Příklad – pokračování



Obr.: Sloupkový diagram



Obr.: Polygon četností

Intervalové rozložení četností

- třídicí intervaly $(u_1, u_2\rangle, \ldots, (u_r, u_{r+1}\rangle$
- doporučuje se volit r blízké $\sqrt{n}$.

Četnostní hustota j-tého třídicího intervalu je definována vztahem

$$f_j = \frac{p_j}{d_j}$$

kde $d_j = u_{j+1} - u_j$. Soustava obdélníků sestrojených nad třídicími intervaly, jejichž plochy jsou rovny relativním četnostem, se nazývá **histogram**.

- **hustota četnosti**:

$$f(x) = \begin{cases} f_j & \text{pro } u_j < x \leq u_{j+1},\ j = 1, \ldots, r \\ 0 & \text{jinak} \end{cases}$$

  (grafem hustoty četnosti je schodovitá čára shora omezující histogram)
- **Intervalová empirická distribuční funkce**:

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Příklad

U 70 domácností byly zjišťovány týdenní výdaje na nealkoholické nápoje (v Kč).

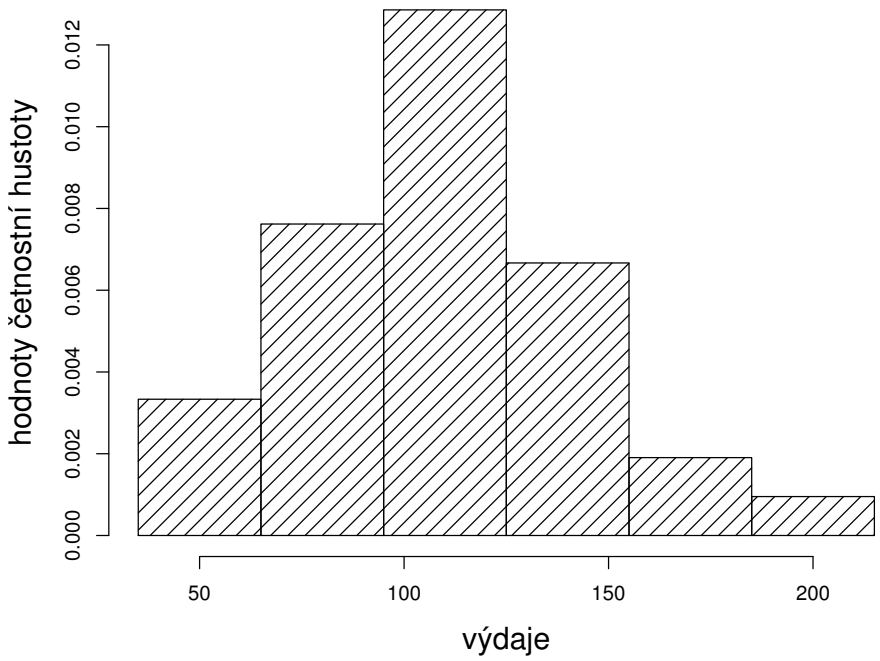
Výdaje	$\langle 35,65\rangle$	$\langle 65,95\rangle$	$\langle 95,125\rangle$	$\langle 125,155\rangle$	$\langle 155,185\rangle$	$\langle 185,215\rangle$
Počet domácností	7	16	27	14	4	2

Sestavte tabulku rozložení četností, nakreslete histogram a graf intervalové empirické distribuční funkce.

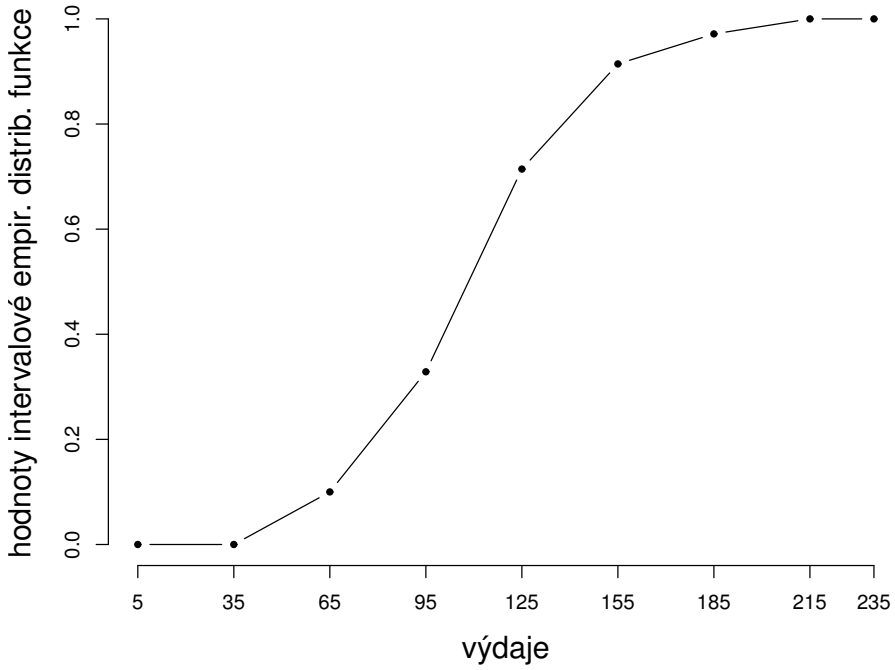
Řešení. Tabulka rozložení četností

$\langle u_j, u_{j+1}\rangle$	n_j	p_j	f_j	N_j	F_j
$\langle 35,65\rangle$	7	$7/70$	$7/2100$	7	$7/70$
$\langle 65,95\rangle$	16	$16/70$	$16/2100$	23	$23/70$
$\langle 95,125\rangle$	27	$27/70$	$27/2100$	50	$50/70$
$\langle 125,155\rangle$	14	$14/70$	$14/2100$	64	$64/70$
$\langle 155,185\rangle$	4	$4/70$	$4/2100$	68	$68/70$
$\langle 185,215\rangle$	2	$2/70$	$2/2100$	70	1

Příklad – pokračování



Obr.: Histogram



Obr.: Graf intervalové empirické distribuční funkce

Číselné charakteristiky datového souboru

Znaky nominálního typu

Nominální škála klasifikuje objekty do určitých předem vymezených tříd či kategorií. Hodnoty v nominální škále se dají vyjádřit slovně a mezi různými hodnotami není definováno žádné uspořádání. Pokud jsou hodnoty nominální škály někdy označovány číselně, mějme na paměti, že toto číslo je pouze jakousi zkratkou (kódem) slovní hodnoty. O znacích měřených v nominální škále hovoříme jako o znacích nominálního typu.

Příklady znaků nominálního typu mohou být např.:

- pohlaví (s možnými hodnotami mužské, ženské)
- barva očí (modrá, hnědá, černá)
- výsledek léčby (uzdraven, zemřel)
- národnost (česká, slovenská, polská, německá, ...)

Charakteristikou polohy je **modus** – nejčetnější varianta či střed nejčetnějšího intervalu. (Modus je jediná charakteristika polohy vhodná pro nominální veličiny).

Číselné charakteristiky datového souboru

Znaky ordinálního typu

Znaky ordinálního typu lze podle sledované vlastnosti nejen rozlišovat, ale také uspořádat ve smyslu vztahů „je větší“, „je menší“ nebo „předchází“, „následuje“, aniž bychom však byli schopni vyjádřit číselně vzdálenost mezi větším a menším či mezi předcházejícím a následujícím.

Znaky ordinálního typu mohou být např.:

- dosažené vzdělání (základní, střední, vysokoškolské)
- prospěch ve školním předmětu (výborně, velmi dobře, dobře, nevyhověl)
- stav pacienta (vyléčen, remise, recidiva)
- hodnocení funkce technických zařízení (stupně závažnosti poruchy jaderné elektrárny)
- hodnocení postojů v sociologických průzkumech (škála má hodnoty např. souhlasím, spíše souhlasím, spíše nesouhlasím, nesouhlasím)
- četnost výskytu (často, občas, zřídka, nikdy)

Vhodnou charakteristikou polohy je **α-kvantil**.
Je-li $\alpha \in (0;1)$, pak α-kvantil x_α je číslo, které rozděluje uspořádaný datový soubor na dolní úsek, obsahující aspoň podíl α všech dat a na horní úsek obsahující aspoň podíl $1-\alpha$ všech dat.

Číselné charakteristiky datového souboru

Pro výpočet α-kvantilu slouží algoritmus:

$$n\alpha = \begin{cases} \text{celé číslo } c & \Rightarrow x_\alpha = \frac{x_{(c)} + x_{(c+1)}}{2} \\ \text{necelé číslo} & \Rightarrow \text{zaokrouhlíme nahoru na nejbližší celé číslo} \\ & \qquad x_\alpha = x_{(c)} \end{cases}$$

Pro speciálně zvolená α užíváme názvů:

- $x_{0,50}$ – medián
- $x_{0,25}$ – dolní kvartil
- $x_{0,75}$ – horní kvartil
- $x_{0,1}, \ldots, x_{0,9}$ – decily
- $x_{0,01}, \ldots, x_{0,99}$ – percentily.

Jako charakteristika variability slouží **kvartilová odchylka**: $q = x_{0,75} - x_{0,25}$ (**interquartile range**).

Příklad

Během semestru se studenti podrobili písemnému testu z matematiky, v němž bylo možno získat 0 až 10 bodů. Výsledky jsou uvedeny v tabulce:

Počet bodů	0	1	2	3	4	5	6	7	8	9	10
Počet studentů	1	4	6	7	11	15	19	17	12	6	3

Zjistěte modus, medián, 1. decil, 9. decil a kvartilovou odchylku počtu bodů.

Řešení. Modus je nejčetnější varianta znaku, v tomto případě tedy 6. Vypočtěme rozsah datového souboru: $n = 1 + 4 + \cdots + 3 = 101$.
Výpočty uspořádáme do tabulky.

α	$n\alpha$	c	$x_\alpha = x_{(c)}$
0,50	50,5	51	6
0,10	10,1	11	2
0,90	90,9	91	8
0,25	25,25	26	4
0,75	75,75	76	7

Kvartilová odchylka: $q = 7 - 4 = 3$.

Znaky intervalového a poměrového typu

U znaků **intervalového typu** lze stanovit vzdálenost mezi hodnotami měřené veličiny. Je zde definována jednotka měření, avšak nula je definována pouze relativně. To nám dovoluje proto počítat s rozdíly naměřených hodnot, nikoliv s jejich podíly. Typickým příkladem je teplota, která se dá měřit v různých stupnicích (Celsiova, Fahrenheitova).
U znaků **poměrového typu** lze určit nejen rozdíly (intervaly) mezi hodnotami, ale i podíly hodnot, neboť tyto znaky mají nulu stanovenu absolutně a jednoznačně.
Charakteristiky **polohy**:

- **Aritmetický průměr** (**mean**) $\bar{x}$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1}$$

U poměrových znaků, které nabývají pouze kladných hodnot, lze použít

- **geometrický průměr** (**geometric mean**):

$$\sqrt[n]{x_1 \cdot \ldots \cdot x_n} \tag{2}$$

Znaky intervalového a poměrového typu

Charakteristiky **variability**:

- **rozptyl** (**variance**):

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{3}$$

- **směrodatná odchylka** (**standard deviation**):

$$s = \sqrt{s^2} \tag{4}$$

- **koeficient variace** (**variation**):

$$\frac{s}{\bar{x}} \tag{5}$$

Rozptyl se zpravidla počítá podle vzorce $s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$.

Znaky intervalového a poměrového typu

Známe-li absolutní či relativní četnosti variant $x_{[1]}, \ldots, x_{[r]}$, můžeme spočítat

- **vážený průměr** (**weighted mean**):

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{r} n_j x_{[j]} \tag{6}$$

nebo

- **vážený rozptyl** (**weighted variance**):

$$s^2 = \frac{1}{n} \sum_{j=1}^{r} n_j (x_{[j]} - \bar{x})^2 \tag{7}$$

Vážený rozptyl se zpravidla počítá podle vzorce $s^2 = \frac{1}{n} \sum_{j=1}^{r} n_j x_{[j]}^2 - \bar{x}^2$.

Znaky intervalového a poměrového typu

Aritmetický průměr a rozptyl jsou speciální případy tzv. momentů. V následující definici obecně zavedeme k-tý počáteční a centrální moment.

- **k-tý počáteční moment** (k^{th} **moment**):

$$m'_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k, \quad \text{kde } k = 1, 2, \ldots \tag{8}$$

- **k-tý centrální moment**:

$$m_k = \frac{1}{n}\sum_{i=1}^{n} (x_i - m)^k, \quad \text{kde } k = 1, 2, \ldots \tag{9}$$

Znaky intervalového a poměrového typu

Pomocí 3. a 4. centrálního momentu se definuje šikmost a špičatost:

- **šikmost** (**skew**):

$$\alpha_3 = \frac{m_3}{s^3} \tag{10}$$

  Šikmost měří nesouměrnost rozložení četností kolem průměru.

- **špičatost** (**kurtosis**):

$$\alpha_4 = \frac{m_4}{s^4} - 3 \tag{11}$$

  Špičatost měří koncentraci rozložení četností kolem průměru.

Příklad

Pro údaje z příkladu o domácnostech vypočtěte průměr a rozptyl počtu členů domácnosti.

Řešení

$$\bar{x} = \frac{1}{30}(1 \cdot 2 + 2 \cdot 6 + 3 \cdot 4 + 4 \cdot 10 + 5 \cdot 5 + 6 \cdot 3) = \frac{109}{30} = 3,6\bar{3}$$

$$s^2 = \frac{1}{30}\left(1^2 \cdot 2 + 2^2 \cdot 6 + 3^2 \cdot 4 + 4^2 \cdot 10 + 5^2 \cdot 5 + 6^2 \cdot 3\right) - \left(\frac{109}{30}\right)^2 = \frac{1769}{900} = 1,96\bar{5}$$

Příklad

Nechť $\bar{x}$ je průměr a s_1^2 rozptyl hodnot $x_1, \ldots, x_n$. Nechť a, b jsou reálné konstanty. Položme $y_i = a + bx_i, i = 1, \ldots, n$. Vypočtěte průměr $\bar{y}$ a rozptyl s_2^2 hodnot $y_1, \ldots, y_n$.

Řešení

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} (a + bx_i) = a + b\frac{1}{n}\sum_{i=1}^{n} x_i = a + b\bar{x},$$

$$s_2^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n} (a + bx_i - a - b\bar{x})^2 = b^2\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 = b^2 s_1^2.$$

Diagnostické grafy

Krabicový diagram (Box plot)

Umožňuje posoudit symetrii a variabilitu datového souboru a existenci odlehlých či extrémních hodnot. Můžete se setkat i z názvem **box plot**.

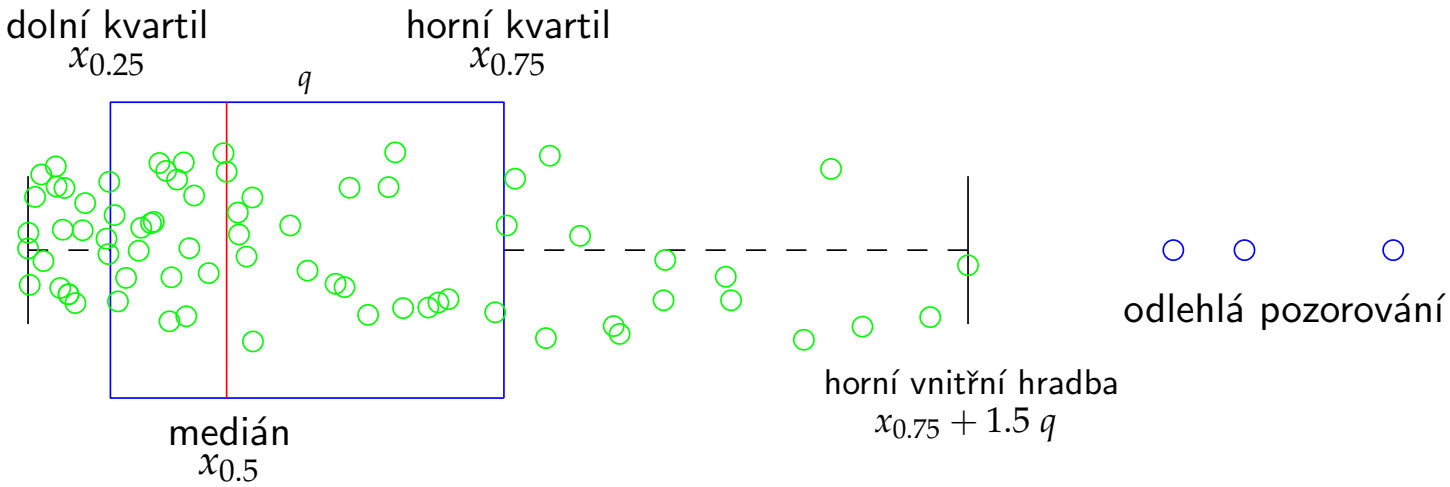
Krabicový diagram je specifikován těmito pojmy:

- **Dolní vnitřní hradba**:
$$x_{0,25} - 1,5q$$
- **Horní vnitřní hradba**:
$$x_{0,75} + 1,5q$$
- **Dolní vnější hradba**:
$$x_{0,25} - 3q$$
- **Horní vnější hradba**:
$$x_{0,75} + 3q$$

Odlehlá hodnota (**Outlier**) je hodnota, která leží mezi vnitřními a vnějšími hradbami. **Extrémní hodnota** (**Extreme value**) je hodnota, která leží za vnějšími hradbami.

Diagnostické grafy

Způsob konstrukce krabicového diagramu:



Příklad

Pro data z příkladu o domácnostech sestrojte krabicový diagram.

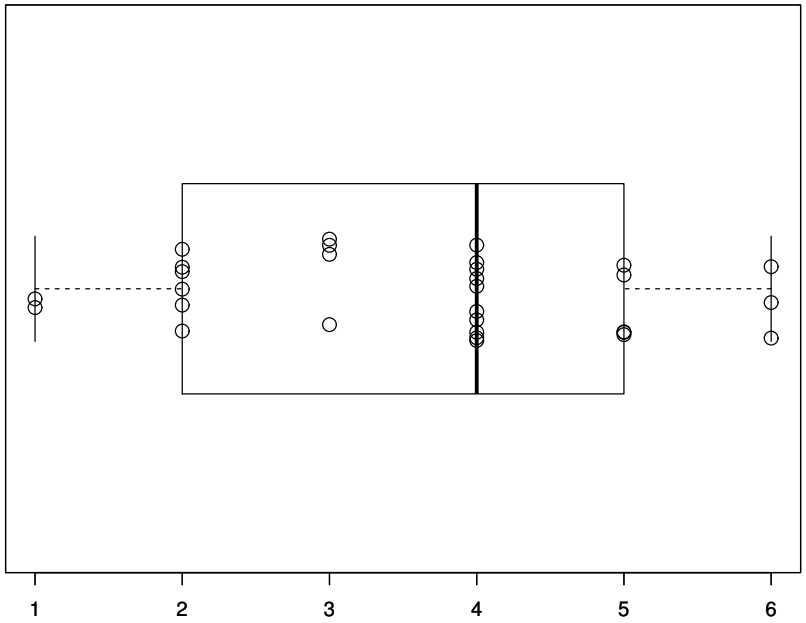
Počet členů	1	2	3	4	5	6
Počet domácností	2	6	4	10	5	3

Řešení.
Rozsah souboru $n = 30$. Výpočty potřebných kvantilů uspořádáme do tabulky.

α	$n\alpha$	c		x_α
0,25	7,5	8	$x_{(c)} = x_{(8)}$	2
0,50	15	15	$\frac{x_{(15)}+x_{(16)}}{2}$	4
0,75	22,5	23	$x_{(c)} = x_{(23)}$	5

$q = 5 - 2 = 3$
Dolní vnitřní hradba: $x_{0,25} - 1,5q = 2 - 1,5.3 = -2,5$
Horní vnitřní hradba: $x_{0,75} + 1,5q = 5 + 1,5.3 = 9,5$

Příklad



Obr.: Krabicový diagram

Diagnostické grafy

Normal probability plot (N-P plot)

N-P plot konstruujeme tak, že na vodorovnou osu vynášíme uspořádané hodnoty $x_{(1)} \leq \cdots \leq x_{(n)}$ a na svislou osu kvantily normálního rozdělení u_{α_j}, kde

$$\alpha_j = \frac{3j-1}{3n+1}.$$

Jsou-li některé hodnoty $x_{(1)} \leq \cdots \leq x_{(n)}$ stejné, pak za j bereme průměrné pořadí odpovídající takové skupince.

- Pocházejí-li data z normálního rozložení, pak budou všechny dvojice $\left(x_{(j)}, u_{\alpha_j}\right)$ ležet na přímce.
- Pro data z rozložení s kladnou šikmostí se budou dvojice $\left(x_{(j)}, u_{\alpha_j}\right)$ řadit do konkávní křivky.
- Pro data z rozložení se zápornou šikmostí se budou dvojice $\left(x_{(j)}, u_{\alpha_j}\right)$ řadit do konvexní křivky.

Diagnostické grafy

Quantile - quantile plot (Q-Q plot)

Q-Q plot konstruujeme tak, že na svislou osu vynášíme uspořádané hodnoty $x_{(1)} \leq \cdots \leq x_{(n)}$ a na vodorovnou osu kvantily $K_{\alpha_j}(X)$ vybraného rozložení, kde

$$\alpha_j = \frac{j - r_{adj}}{n + n_{adj}},$$

přičemž r_{adj} a n_{adj} jsou korigující faktory $\leq 0,5$. Implicitně se klade $r_{adj} = 0,375$ a $n_{adj} = 0,25$. Pokud vybrané rozložení závisí na nějakých parametrech, pak se tyto parametry odhadují z dat, nebo se volí na základě teoretického modelu. Body $(K_{\alpha_j}(X), x_{(j)})$ se metodou nejmenších čtverců proloží přímka. Čím méně se body odchylují od této přímky, tím lepší je soulad mezi empirickým a teoretickým rozložením.
Jsou-li některé hodnoty $x_{(1)} \leq \cdots \leq x_{(n)}$ stejné, pak za j bereme průměrné pořadí odpovídající takové skupince.

Příklad

Desetkrát nezávisle na sobě byla změřena jistá konstanta. Výsledky měření:
$2 \quad 1,8 \quad 2,1 \quad 2,4 \quad 1,9 \quad 2,1 \quad 2 \quad 1,8 \quad 2,3 \quad 2,2.$
Pomocí N-P plotu a Q-Q plotu ověřte, zda se tato data řídí normálním rozložením.

Řešení

usp.hodnoty	1,8	1,8	1,9	2	2	2,1	2,1	2,2	2,3	2,4
pořadí	1	2	3	4	5	6	7	8	9	10
průměrné pořadí	1,5	1,5	3	4,5	4,5	6,5	6,5	8	9	10

- **N-P plot:**

  $j = (1,5;\ 3;\ 4,5;\ 6,5;\ 8;\ 9;\ 10)$

  $\alpha_j = \frac{3j-1}{3n+1} = (0,1129;\ 0,2581;\ 0,4032;\ 0,5968;\ 0,7419;\ 0,8387;\ 0,9355)$

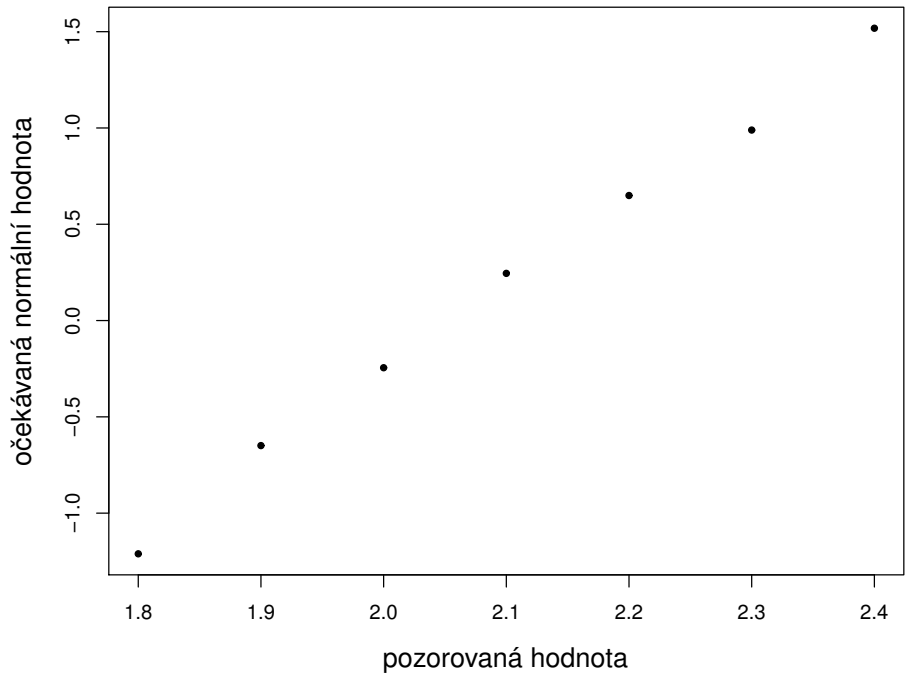
  $u_{\alpha_j} = (-1,2112;\ -0,6493;\ -0,245;\ 0,245;\ 0,6493;\ 0,9892;\ 1,5179)$

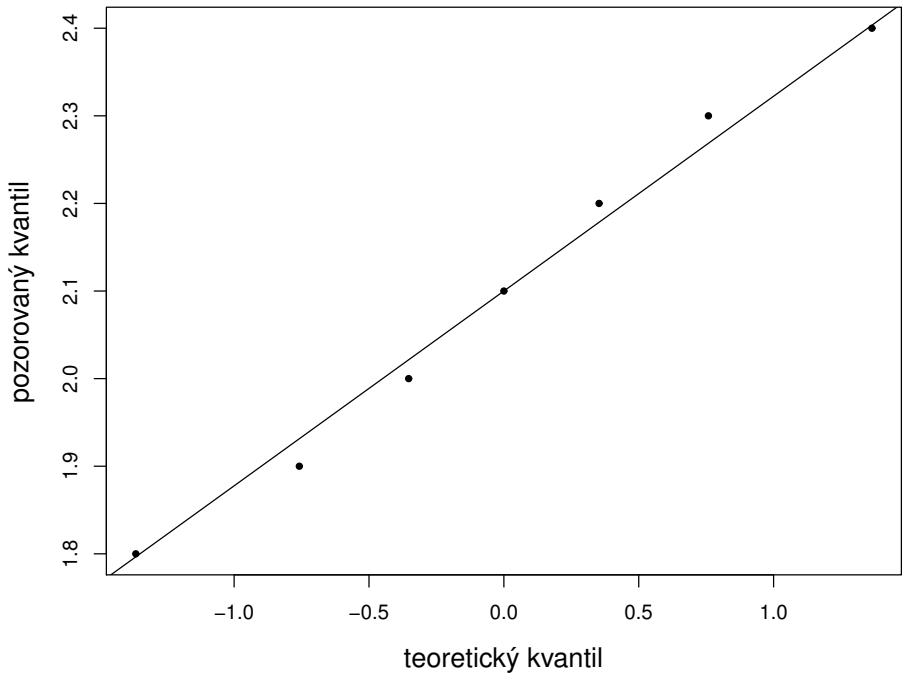
Příklad

- **Q-Q plot:**
  $j = (1,5;\ 3;\ 4,5;\ 6,5;\ 8;\ 9;\ 10)$

  $\alpha_j = \frac{j-0,375}{n+0,25} = (0,1098;\ 0,2561;\ 0,4024;\ 0,5976;\ 0,7439;\ 0,8415;\ 0,939)$

  $u_{\alpha_j} = (-1,2278;\ -0,6554;\ -0,247;\ 0,247;\ 0,6554;\ 1,0005;\ 1,566)$





Diagnostické grafy

Probability - probability plot (P-P plot)

Spočtou se standardizované hodnoty

$$z_{(j)} = \frac{x_{(j)} - \bar{x}}{s}, \quad j = 1, \ldots, n.$$

Na vodorovnou osu se vynesou hodnoty teoretické distribuční funkce $\Phi(z_{(j)})$ a na svislou osu hodnoty empirické distribuční funkce $F(z_{(j)}) = j/n$. Pokud se body $(\Phi(z_{(j)}), F(z_{(j)}))$ řadí kolem hlavní diagonály čtverce $\langle 0,1\rangle \times \langle 0,1\rangle$, lze usuzovat na dobrou shodu empirického a teoretického rozložení.
Jsou-li některé hodnoty $x_{(1)} \leq \ldots \leq x_{(n)}$ stejné, pak za j bereme průměrné pořadí odpovídající takové skupince.

Diagnostické grafy

Histogram

Umožňuje porovnat tvar hustoty četnosti s tvarem hustoty pravděpodobnosti vybraného teoretického rozložení. Např. normálního, Pearsonova, Studentova a jiných.

Diagnostické grafy

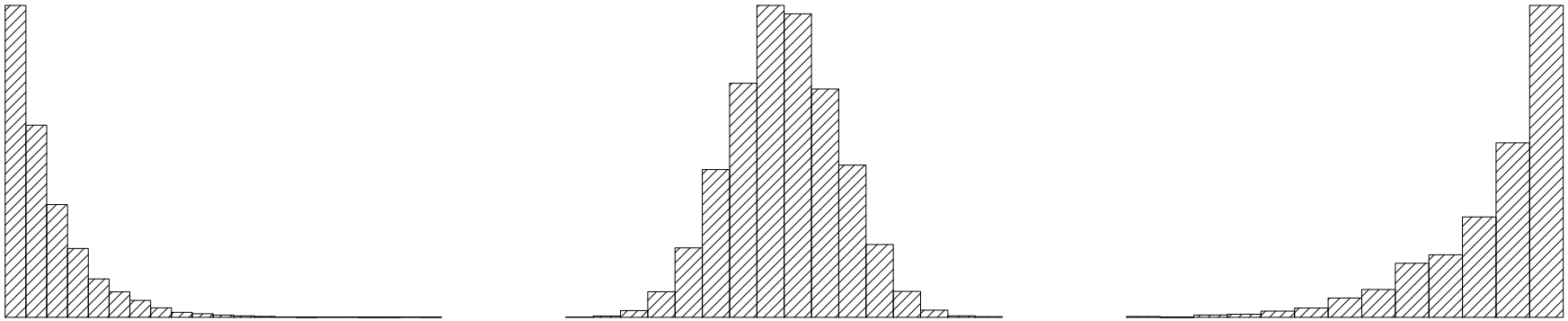
Vzhled diagnostických grafů pro rozložení s různou šikmostí

Vlastnosti rozložení četností datového souboru se projeví ve vzhledu histogramu, N-P plotu a krabicového diagramu, jak ukazují následující obrázky:

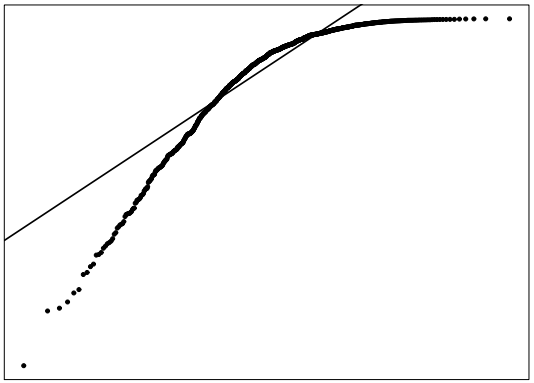
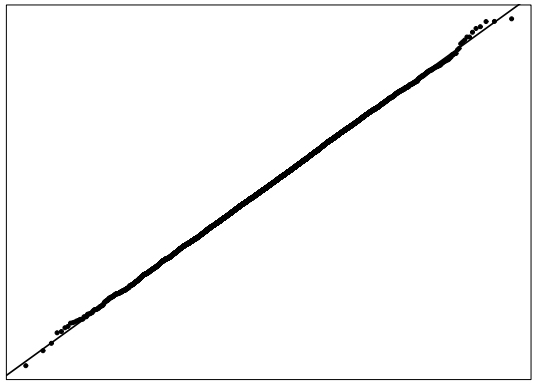
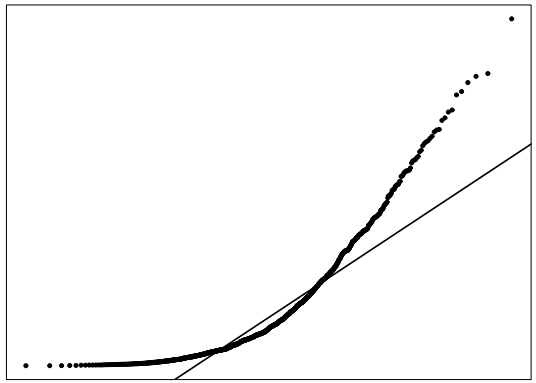
Rozložení s kladnou šikmostí

Normální rozložení

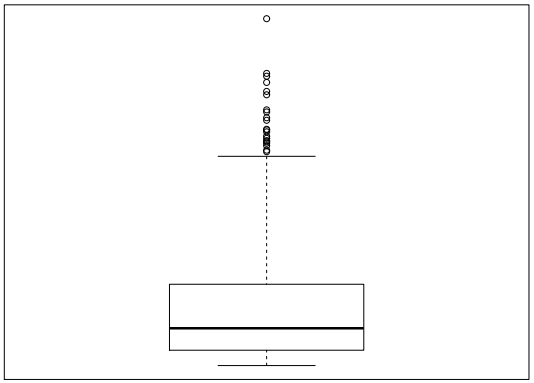
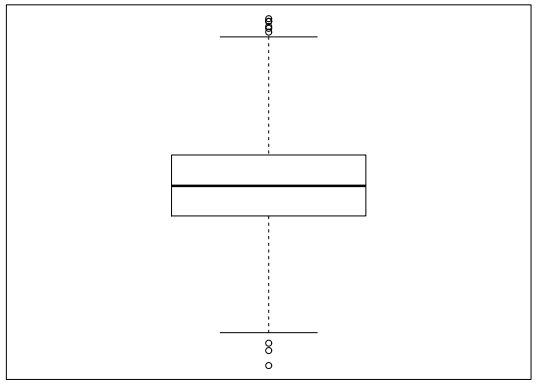
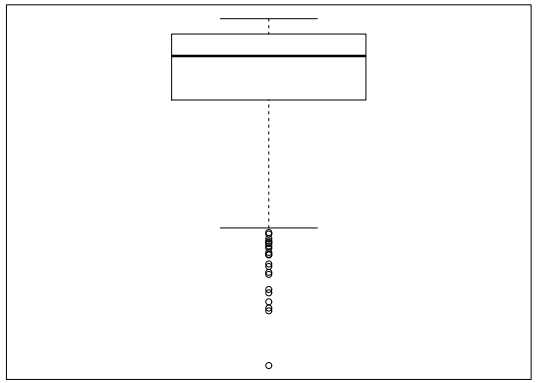
Rozložení se zápornou šikmostí



Obr.: Histogramy

Obr.: N-P plot

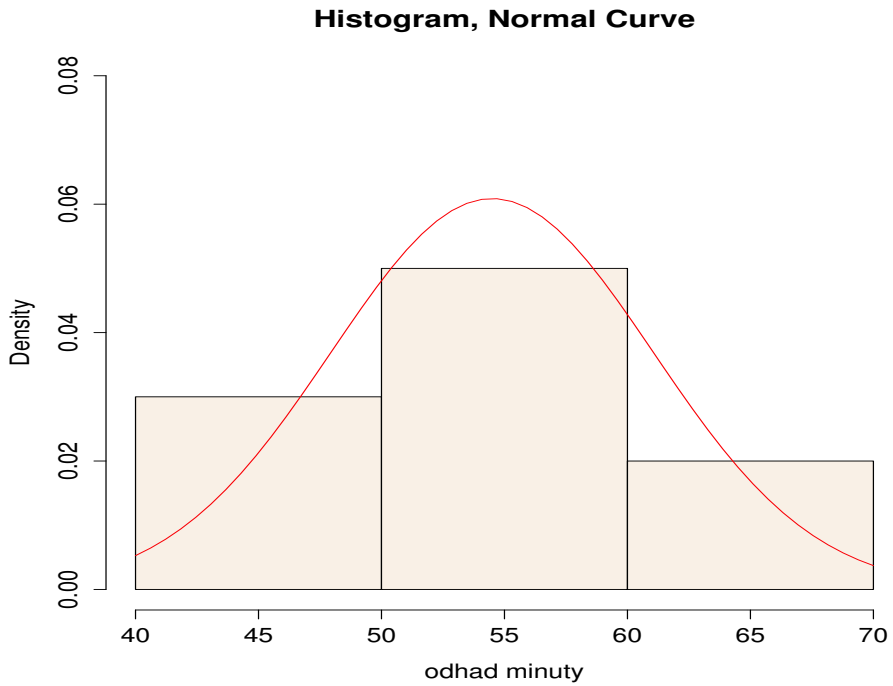
Obr.: Box plot

Příklad

Příklad 1

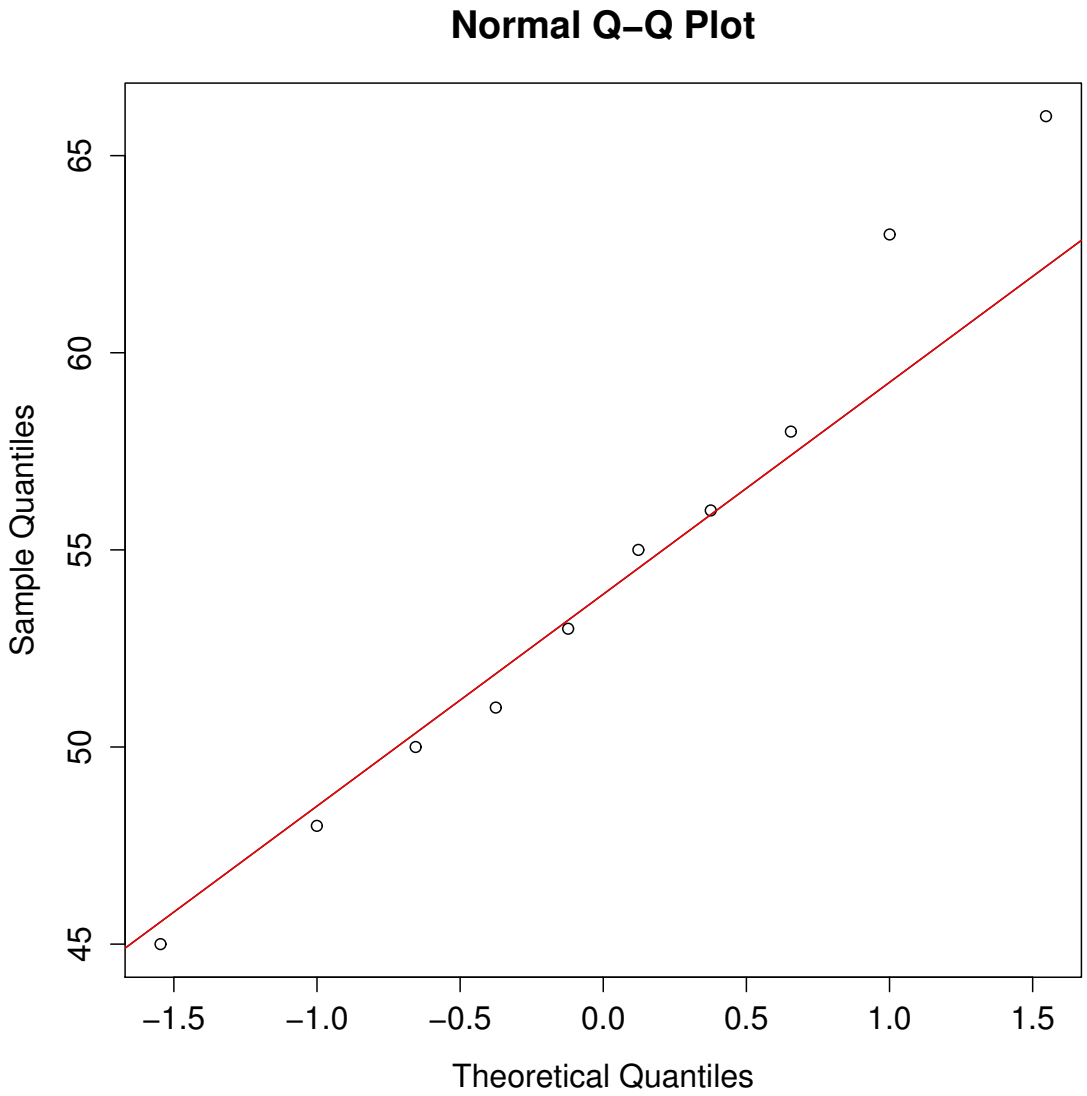
Deset pokusných osob mělo nezávisle na sobě bez předchozího nácviku odhadnout, kdy od daného signálu uplyne jedna minuta. Výsledky pokusu jsou uloženy v souboru „`minuta.RData`“. Testujte graficky, zda se jedná o výběr z normálního rozdělení.

Řešení Histogram a teoretická hustota



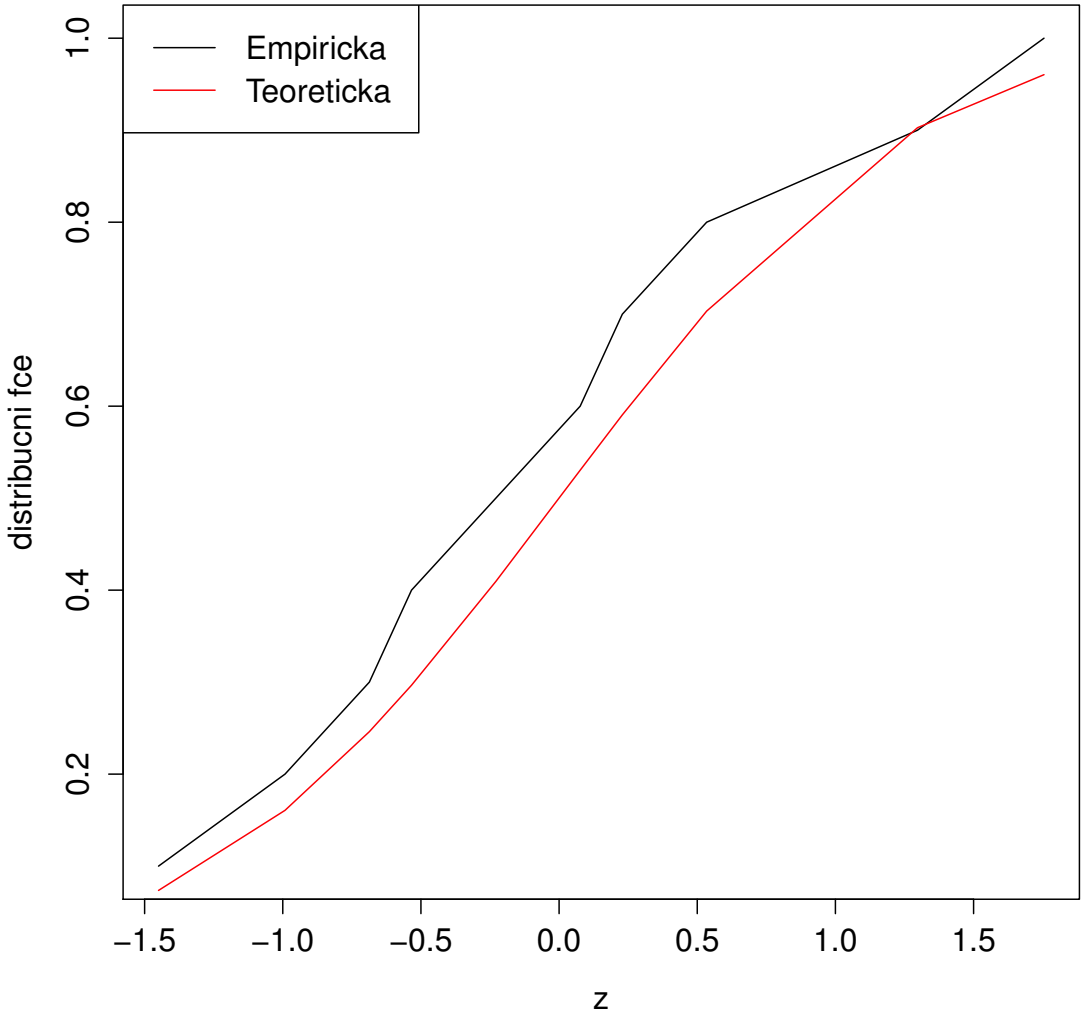
Řešení

Q–Q plot



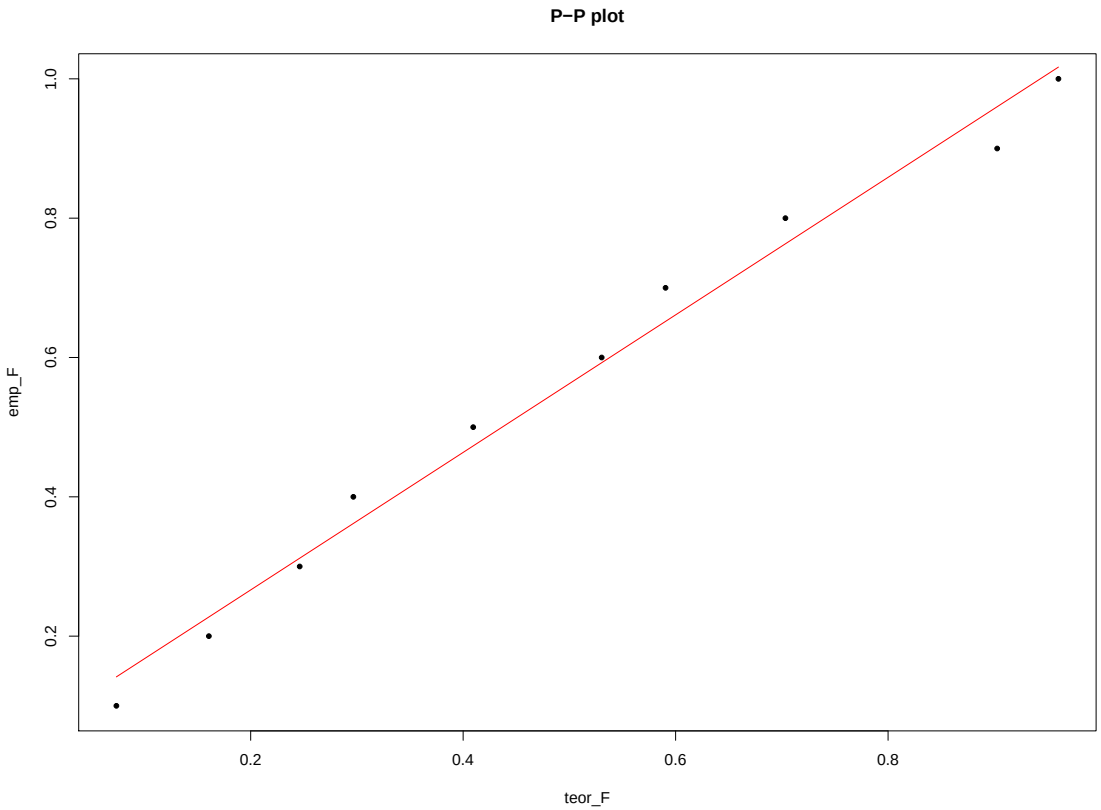
Řešení

Výběrová distribuční funkce



Řešení

P–P plot



Úlohy k procvičení

Příklad 2.1

U 20 studentů 1. ročníku byla zjišťována známka z matematiky na prvním zkušebním termínu.

Známka	*1*	*2*	*3*	*4*
Počet studentů	*7*	*3*	*2*	*8*

Vytvořte tabulku rozložení četností. Nakreslete grafy četnostní funkce a empirické distribuční funkce. Dále nakreslete sloupkový diagram a polygon četností známek.

Úlohy k procvičení

Příklad 2.2

U 60 vzorků oceli byla zjišťována mez plasticity.

Mez plasticity	$\langle 30,50\rangle$	$\langle 50,70\rangle$	$\langle 70,90\rangle$	$\langle 90,110\rangle$	$\langle 110,130\rangle$	$\langle 130,150\rangle$	$\langle 150,170\rangle$
Počet vzorků	*8*	*4*	*13*	*15*	*9*	*7*	*4*

Sestavte tabulku rozložení četností, nakreslete histogram a graf intervalové empirické distribuční funkce.

Příklad 2.3

Pro údaje z příkladu 2.2 vypočtěte průměr a rozptyl meze plasticity.

$[\bar{x} = 96,67,\ s^2 = 1148,89]$

Úlohy k procvičení

Příklad 2.4

V datovém souboru, z něhož byl vypočten průměr 110 a rozptyl 800, byly zjištěny 2 chyby: místo 85 má být 95 a místo 120 má být 150. Ostatních 18 údajů je správných. Opravte průměr a rozptyl.

[$\bar{x} = 112$, $s^2 = 851$*]*

Příklad 2.5

Pro údaje z příkladu 2.1 sestrojte krabicový diagram.

[$x_{0,50} = 2,5$, $x_{0,25} = 1$, $x_{0,75} = 4$, $q = 3$, *dolní vnitřní hradba* $= -3,5$, *horní vnitřní hradba* $= 8,5$*]*