2 Intervalové rozložení četností

2.1 Jednorozměrné intervalové rozložení četností

Jestliže v jednorozměrném datovém souboru je počet variant znaku X blízký rozsahu souboru n, pak četnosti přiřazujeme nikoliv jednotlivým variantám, ale třídicím intervalům $(u_1; u_{j+1}\rangle$, …, $(u_r; u_{r+1}\rangle$. Hovoříme pak o intervalovém rozložení četností.

Označení:

- $(u_j; u_{j+1}\rangle$ – j-tý třídicí interval znaku X, $j = 1, \ldots, r$,
- $d_j = u_{j+1} – u_j$ – délka j-tého třídicího intervalu znaku X,
- $x_{[j]} = \frac{u_j + u_{j+1}}{2}$ – střed j-tého třídicího intervalu znaku X.

Třídicí intervaly volíme nejčastěji stejně dlouhé. Jejich počet určíme např. pomocí Sturgesova pravidla: $r \approx 1 + 3,3 \log_{10} n$, kde n je rozsah souboru.

Názvy četností jsou podobné jako v 1.2. Navíc se používá četnostní hustota j-tého třídicího intervalu: $f_j = \frac{p_j}{d_j}$. Četnosti a četnostní hustotu zapisujeme do tabulky rozložení četností (viz tabulka 1).

Tabulka 1: Tabulka rozložení četností znaku X

$(u_j; u_{j+1}\rangle$	d_j	$x_{[j]}$	n_j	p_j	N_j	F_j	f_j
$(u_1; u_2\rangle$	d_1	$x_{[1]}$	n_1	p_1	N_1	F_1	f_1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$(u_r; u_{r+1}\rangle$	d_r	$x_{[r]}$	n_r	p_r	N_r	F_r	f_r

Jednorozměrné intervalové rozložení četností graficky znázorníme pomocí histogramu. Nad třídicími intervaly sestrojíme soustavu obdélníků, jejichž plocha je rovna relativním četnostem (a výšky tedy četnostním hustotám) jednotlivých intervalů. Schodovitá čára shora omezující histogram je grafem hustoty četnosti:

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

Pomocí hustoty četnosti zavedeme intervalovou empirickou distribuční funkci: $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$.

Příklad 2.1. Řešený příklad
Načtěte datový soubor 16-anova-head.txt. Za předpokladu, že znak X popisuje délku hlavy v mm (head.L) u mužů (a) zjistěte hranice třídicích intervalů; (b) vytvořte tabulku rozložení četností; (c) nakreslete histogram, resp. graf intervalové empirické distribuční funkce $F(x)$.

Řešení příkladu 2.1
Datový soubor načteme příkazem read.delim(). Z načtených dat vybereme pomocí operátoru [] pouze řádky týkající se mužů a sloupec head.L. Z výběru odstraníme chybějící údaje příkazem na.omit(). Ke stanovení hranic třídicích intervalů potřebujeme znát rozsah datového souboru a minimální a maximální naměřenou hodnotu délky hlavy mužů. Rozsah datového souboru zjistíme pomocí příkazu length(), minimální a maximální naměřenou hodnotu pomocí příkazu range(). Optimální počet třídicích intervalů stanovíme na základě Sturgesova pravidla.

```
1 data <- read.delim('16-anova-head.txt', sep = '\t', dec = '.')
2 head.LM <- na.omit(data[data$sex == 'm', 'head.L'])
3 n <- length(head.LM) # 75
4 range(head.LM) # 180-214
5 r <- round(3.3 * log10(n) + 1) # 7
6 # 214 - 180 = 34 -> 35; 35 / 7 = 5 -> seq(179, 214, by = 5)
7 b.head.LM <- seq(179, 214, by = 5) # 179, 184, ..., 214
```

Na základě Sturgesova pravidla stanovíme optimální počet třídicích intervalů $r = 7$. Naměřené hodnoty délky hlavy se pohybují v rozmezí 180 až 214 mm. Vzdálenost mezi minimální naměřenou hodnotou sníženou o 1 a maximální naměřenou hodnotou je $214 - 179 = 35$ mm. Tato vzdálenost je beze zbytku dělitelná počtem třídicích intervalů, tj. sedmi. Optimální šířka jednoho třídicího intervalu $d = \frac{35}{7} = 5$ mm. Hranice třídicích intervalů tedy stanovíme jako posloupnost 179, 184, ..., 214 mm. Celkem získáme osm hranic definujících sedm třídicích intervalů o optimální šířce 5 mm.

Tabulku rozložení četností dopočítáme pomocí funkce trc() implementované v R-skriptu AS-sbirka-funkce.R, který je součástí této publikace. R-skript načteme příkazem source(). Vstupními argumenty funkce trc() jsou vektor naměřených hodnot x a hranice třídicích intervalů breaks. Výstupem funkce trc() je tabulka rozložení četností pro znak X.

```
source('AS-sbirka-funkce.R')
tab.rel.cet <- trc(head.LM, b.head.LM)
```

```
  dj     xj nj      pj Nj      Fj      fj
1  5 181,5  5 0,0667  5 0,0667 0,0133
2  5 186,5  6 0,0800 11 0,1467 0,0160
3  5 191,5 20 0,2667 31 0,4133 0,0533
4  5 196,5 24 0,3200 55 0,7333 0,0640
5  5 201,5 12 0,1600 67 0,8933 0,0320
6  5 206,5  6 0,0800 73 0,9733 0,0160
7  5 211,5  2 0,0267 75 1,0000 0,0053
```

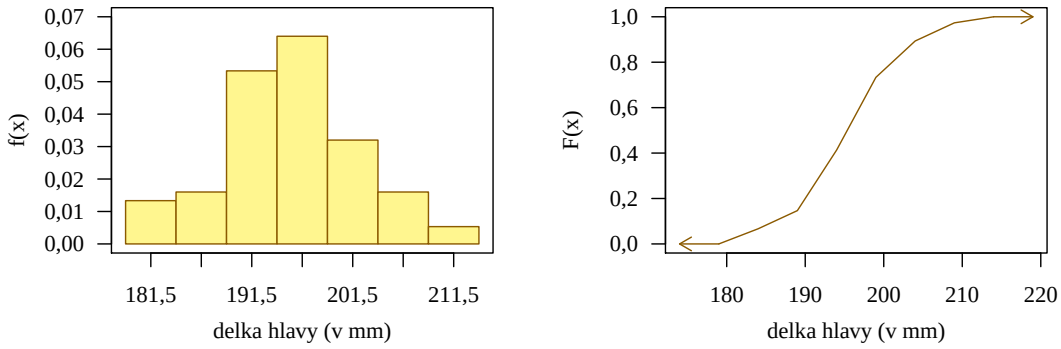
Pro ukázku si zde uvedeme interpretace vybraných hodnot ze třetího a pátého řádku tabulky rozložení četností: Celkem 20 mužů (26,67 %) v datovém souboru mělo délku hlavy v rozmezí 189 až 194 mm, celkem 31 mužů (41,33 %) v datovém souboru mělo délku hlavy nejvýše 194 mm. Celkem 12 mužů (16,00 %) v datovém souboru mělo délku hlavy v rozmezí 199 až 204 mm, celkem 67 mužů (89,33 %) v datovém souboru mělo délku hlavy nejvýše 204 mm.

Histogram vykreslíme pomocí příkazu hist() s argumentem prob = T. Tento argument zajistí, že obsahem sloupců histogramu budou relativní četnosti a výškou sloupců četnostní hustoty. Hranice třídicích intervalů nastavíme pomocí argumentu breaks. Okolo grafu dokreslíme rámeček příkazem box(). Osu x, resp. osu y vykreslíme samostatně pomocí příkazu axis() s argumentem side = 1, resp. side = 2. V souladu se zavedenou konvencí vykreslíme měřítko osy x ve středech třídicích intervalů. Středy třídicích intervalů můžeme získat buď ručním výpočtem, nebo jako výstup mids funkce hist() se specifikací argumentu plot = F. Výsledný histogram je zobrazen na obrázku 1 (vlevo).

```
h <- hist(head.LM, breaks = b.head.LM, plot = F)
hist(head.LM, breaks = b.head.LM, prob = T, include.lowest = F, ylim = c(0, 0.07),
     col = 'khaki1', border = 'orange4', xlab = 'delka hlavy (v mm)',
     ylab = 'f(x)', main = '', axes = F)
box(bty = 'o')
axis(side = 1, at = h$mids)
axis(side = 2, las = 1)
```

Graf intervalové empirické distribuční funkce $F(x)$ vykreslíme pomocí příkazu plot() s argumentem type = 'l'. Prvním vstupním argumentem příkazu bude vektor hodnot empirické distribuční funkce $F(x)$ vložený v tabulce rozložení četností tab.rel.cet ve sloupci Fj. Do grafu dále dokreslíme šipku značící nulovou hodnotu empirické distribuční funkce $F(x)$ ve všech hodnotách menších než dolní hranice prvního třídicího intervalu ($u_1 = 179$) a dále šipku značící jednotkovou hodnotu funkce $F(x)$ ve všech hodnotách větších než horní hranice posledního třídicího intervalu ($u_8 = 214$). Šipky vykreslíme pomocí funkce arrows(). Výsledný graf intervalové empirické distribuční funkce je zobrazen na obrázku 1 (vpravo).

```
Fj <- tab.rel.cet$Fj
plot(b.head.LM, c(0, Fj), type = 'l', las = 1, col = 'orange4', xlim = c(174, 219),
     xlab = 'delka hlavy (v mm)', ylab = 'F(x)')
arrows(179, 0, 174, 0, col = 'orange4', length = 0.1)
arrows(214, 1, 219, 1, col = 'orange4', length = 0.1)
```



Obrázek 1: Histogram (vlevo) a graf intervalové empirické distribuční funkce $F(x)$ (vpravo) pro znak X popisující délku hlavy mužů (v mm)

★

Příklad 2.2. Neřešený příklad
Načtěte datový soubor 11-two-samples-means-skull.txt. Za předpokladu, že znak X popisuje výšku lebky v mm (skull.H) u žen (a) zjistěte hranice třídicích intervalů; (b) vytvořte tabulku rozložení četností; (c) nakreslete histogram, resp. graf intervalové empirické distribuční funkce $F(x)$.
Výsledky: (a) optimální počet třídicích intervalů $r = 8$, optimální šířka jednoho třídicího intervalu $d = 3$, hranice třídicích intervalů: $u_1 = 114$, $u_2 = 117$, …, $u_9 = 138$; (b) tabulka rozložení četností viz tabulka 2; (c) histogram, resp. graf intervalové empirické distribuční funkce $F(x)$ viz obrázek 2.

Tabulka 2: Tabulka rozložení četností

	d_j	$x_{[j]}$	n_j	p_j	N_j	F_j	f_j
1	3,00	115,50	7	0,07	7	0,07	0,02
2	3,00	118,50	12	0,11	19	0,18	0,04
3	3,00	121,50	16	0,15	35	0,33	0,05
4	3,00	124,50	22	0,21	57	0,53	0,07
5	3,00	127,50	27	0,25	84	0,79	0,08
6	3,00	130,50	17	0,16	101	0,94	0,05
7	3,00	133,50	5	0,05	106	0,99	0,02
8	3,00	136,50	1	0,01	107	1,00	0,00

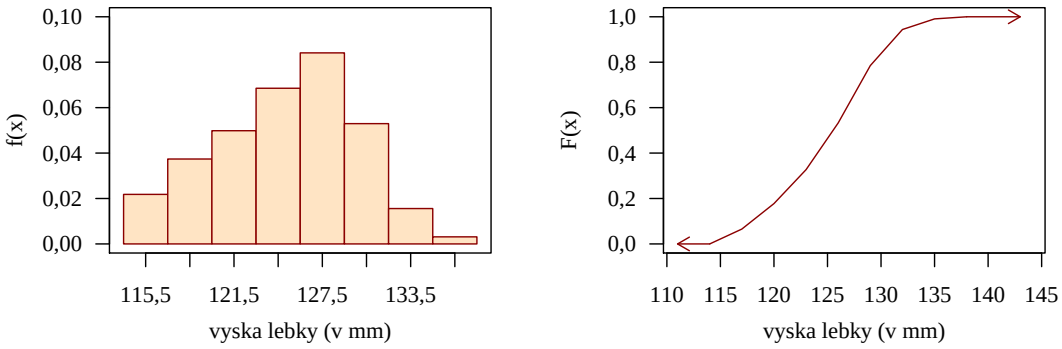
★

2.2 Dvourozměrné intervalové rozložení četností

Ve dvourozměrném datovém souboru o rozsahu n roztřídíme hodnoty znaku X do r třídicích intervalů $(u_j; u_{j+1}\rangle$, $j = 1, \ldots, r$ s délkami $d_1, \ldots, d_r$ a hodnoty znaku Y taktéž roztřídíme do r třídicích intervalů $(v_k; v_{k+1}\rangle$, $k = 1, \ldots, r$ s délkami $h_1, \ldots, h_r$. Obdélník $(u_j; u_{j+1}\rangle \times (v_k; v_{k+1}\rangle$ se nazývá (j, k)-tý dvourozměrný třídicí interval.
Názvy četností jsou podobné jako v 1.3. Navíc zavádíme simultánní a marginální četnostní hustoty:

- $f_{jk} = \frac{p_{jk}}{d_j h_k}$ – simultánní četnostní hustota v (j, k)-tém třídicím intervalu,
- $f_{j\cdot} = \frac{p_{j\cdot}}{d_j}$ – marginální četnostní hustota v j-tém třídicím intervalu pro znak X,
- $f_{\cdot k} = \frac{p_{\cdot k}}{h_k}$ – marginální četnostní hustota v k-tém třídicím intervalu pro znak Y.

Kteroukoliv ze simultánních četností zapisujeme do kontingenční tabulky.



Obrázek 2: Histogram (vlevo) a graf intervalové empirické distribuční funkce $F(x)$ (vpravo) pro znak X popisující výšku lebky žen (v mm)

Dvourozměrné intervalové rozložení četností graficky znázorníme pomocí stereogramu. Nad dvourozměrnými třídicími intervaly sestrojíme soustavu kvádrů, jejichž objem je roven relativním četnostem (a výšky tedy simultánním četnostním hustotám) jednotlivých intervalů. Schodovitá plocha shora omezující stereogram je grafem simultánní četnostní hustoty:

$$f(x,y) = \begin{cases} f_{jk} \text{ pro } u_j < x \le u_{j+1},\ v_k < y \le v_{k+1},\ j,k = 1,\dots,r, \\ 0 \text{ jinak.} \end{cases}$$

Marginální hustoty četnosti pro znaky X a Y odlišíme indexem takto:

$$f_1(x) = \begin{cases} f_{j\cdot} \text{ pro } u_j < x \le u_{j+1},\ j = 1,\dots,r, \\ 0 \text{ jinak,} \end{cases}$$

$$f_2(y) = \begin{cases} f_{\cdot k} \text{ pro } v_k < y \le v_{k+1},\ k = 1,\dots,r, \\ 0 \text{ jinak.} \end{cases}$$

Mezi simultánní hustotou četnosti a marginálními hustotami četnosti platí vztahy: $f_1(x) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}y$, $f_2(y) = \int_{-\infty}^{\infty} f(x,y)\mathrm{d}x$.

Znaky X, Y jsou v daném výběrovém souboru četnostně nezávislé při daném dvourozměrném intervalovém rozložení četností, jestliže pro $\forall j,k = 1,\dots,r$ platí: $f_{jk} = f_{j\cdot} f_{\cdot k}$ neboli pro $\forall (x,y) \in \mathbb{R}^2$: $f(x,y) = f_1(x) f_2(y)$.

Příklad 2.3. Řešený příklad
Načtěte datový soubor 16-anova-head.txt. Za předpokladu, že znak X popisuje délku hlavy v mm (head.L) a znak Y popisuje šířku tváře v mm (bizyg.W) u mužů (a) roztřiďte data do dvourozměrných třídicích intervalů; (b) vytvořte kontingenční tabulku simultánních absolutních, resp. relativních četností; (c) nakreslete histogram pro znak X, resp. pro znak Y; (d) nakreslete stereogram a dvourozměrný tečkový diagram; (e) zjistěte, zda jsou znaky X a Y četnostně nezávislé při daném dvourozměrném intervalovém rozložení četností.

Řešení příkladu 2.3
Datový soubor načteme příkazem read.delim(). Z načtených dat vybereme pomocí operátoru [] pouze řádky týkající se mužů a sloupce head.L a bizyg.W. Z výběru odstraníme chybějící údaje příkazem na.omit(). Ke stanovení hranic třídicích intervalů potřebujeme znát rozsah dvourozměrného datového souboru a minimální a maximální naměřenou hodnotu každého znaku. Rozsah datového souboru stanovíme příkazem length(). Minimální a maximální naměřenou hodnotu znaku X, resp. znaku Y zjistíme příkazem range(). Optimální počet třídicích intervalů každého znaku stanovíme na základě Sturgesova pravidla.

```
30 data <- read.delim('16-anova-head.txt', sep = '\t', dec = '.')
31 data.M <- data[data$sex == 'm', c('head.L', 'bizyg.W')]
```

```r
data.M <- na.omit(data.M)
head.LM <- data.M$head.L
bizyg.WM <- data.M$bizyg.W
n <- length(head.LM) # 75
range(head.LM) # 180-214
range(bizyg.WM) # 113-155
r <- round(3.3 * log10(n) + 1) # 7
# head.LM: viz priklad 2.1
# bizyg.WM: 155 - 112 = 43 / 7 => 49 / 7 = 7 -> seq(110, 159, by = 7)
b.head.LM <- seq(179, 214, by = 5)
b.bizyg.WM <- seq(110, 159, by = 7)
```

Na základě Sturgesova pravidla stanovíme pro oba znaky X i Y optimální počet třídicích intervalů $r = 7$. Naměřené hodnoty znaku X, tj. délky hlavy, se pohybují v rozmezí 180 až 214 mm. Analogicky jako v příkladu 2.1 rozdělíme naměřené hodnoty tohoto znaku do 7 třídicích intervalů o optimální ekvidistantní šířce $d = 5$ mm, a to stanovením hranic 179, 184, ..., 214 mm. Naměřené hodnoty znaku Y, tj. šířky tváře, se pohybují v rozmezí 113 až 155 mm. Vzdálenost mezi minimální naměřenou hodnotou sníženou o 1 a maximální naměřenou hodnotou je $155 - 112 = 43$ mm. Nejbližší vyšší celé číslo dělitelné beze zbytku počtem třídicích intervalů, tj. sedmi, je 49. Optimální šířka jednoho třídicího intervalu $h = \frac{49}{7} = 7$ mm. Hranice třídicích intervalů pro šířku tváře tedy stanovíme jako posloupnost 110, 117, ..., 159 mm. Celkem získáme osm hranic definujících sedm třídicích intervalů o optimální šířce 7 mm.

Po stanovení hranic třídicích intervalů roztřídíme naměřené hodnoty každého znaku do příslušných třídicích intervalů, a to pomocí příkazu cut(). Příkaz přiřadí každému měření příslušný třídicí interval. Informace o třídicím intervalu přiřazeném každému měření znaku X, resp. znaku Y vložíme do proměnné head.LM.c, resp bizyg.WM.c. Kontingenční tabulku simultánních absolutních četností vypočítáme pomocí funkce table(), jejímiž vstupními argumenty budou vektory head.LM.c a bizyg.WM.c. Kontingenční tabulku simultánních relativních četností získáme příkazem prop.table().

```r
head.LM.c <- cut(head.LM, breaks = b.head.LM)
bizyg.WM.c <- cut(bizyg.WM, breaks = b.bizyg.WM)
KT.abs <- table(head.LM.c, bizyg.WM.c)
```

```
          (110;117] (117;124] (124;131] (131;138] (138;145] (145;152] (152;159]
(179;184]         0         0         0         3         1         1         0
(184;189]         1         0         0         0         4         0         1
(189;194]         0         2         2         7         7         2         0
(194;199]         0         0         0         5        12         6         1
(199;204]         0         0         1         2         6         2         1
(204;209]         0         0         1         0         2         3         0
(209;214]         0         0         1         1         0         0         0
```

```r
KT.rel <- prop.table(KT.abs)
```

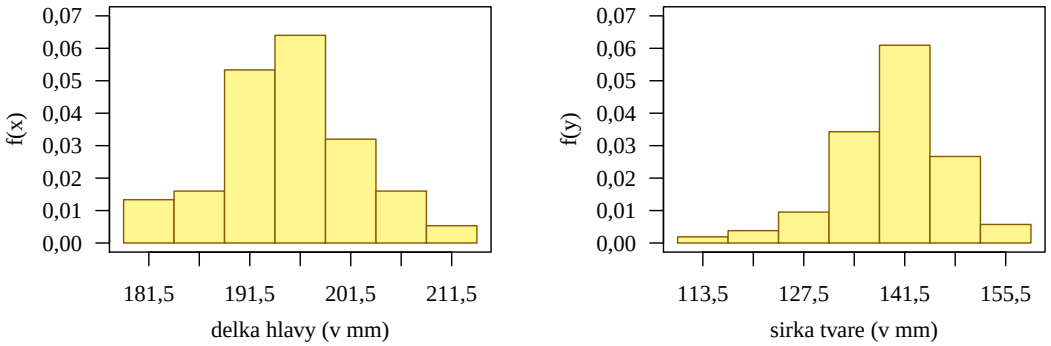
```
          (110;117] (117;124] (124;131] (131;138] (138;145] (145;152] (152;159]
(179;184]    0,0000    0,0000    0,0000    0,0400    0,0133    0,0133    0,0000
(184;189]    0,0133    0,0000    0,0000    0,0000    0,0533    0,0000    0,0133
(189;194]    0,0000    0,0267    0,0267    0,0933    0,0933    0,0267    0,0000
(194;199]    0,0000    0,0000    0,0000    0,0667    0,1600    0,0800    0,0133
(199;204]    0,0000    0,0000    0,0133    0,0267    0,0800    0,0267    0,0133
(204;209]    0,0000    0,0000    0,0133    0,0000    0,0267    0,0400    0,0000
(209;214]    0,0000    0,0000    0,0133    0,0133    0,0000    0,0000    0,0000
```

Pro ukázku si uvedeme interpretace vybraných hodnot z prvního, druhého a třetího sloupce kontingenční tabulky simultánních absolutních, resp. relativních četností. V datovém souboru se vyskytoval jeden muž (1,33 %), jehož délka lebky nabývala hodnoty z intervalu $(184; 189\rangle$ mm a jehož šířka tváře nabývala hodnoty z intervalu $(110; 117\rangle$ mm, dva muži (2,67 %), jejichž délka lebky nabývala hodnoty z intervalu $(189; 194\rangle$ mm a jejichž šířka tváře nabývala hodnoty z intervalu $(117; 124\rangle$ mm, dva muži (2,67 %), jejichž délka lebky nabývala hodnoty z intervalu $(189; 194\rangle$ mm

a jejichž šířka tváře nabývala hodnoty z intervalu $(124; 131\rangle$ mm, apod.

Histogram znaku X, resp. znaku Y vykreslíme pomocí příkazu hist() s argumentem prob = T. Hranice třídicích intervalů definujeme explicitně pomocí argumentu breaks. Okolo grafu vykreslíme rámeček příkazem box(). Měřítko osy x, resp. osy y dokreslíme samostatně pomocí příkazu axis() s argumentem side = 1, resp. side = 2. Měřítko osy x vykreslíme v hodnotách středů třídicích intervalů. Histogramy pro oba znaky jsou zobrazeny na obrázku 3.

```
63 h <- hist(head.LM, breaks = b.head.LM, plot = F)
64 hist(head.LM, breaks = b.head.LM, prob = T, ylim = c(0, 0.07), col = 'khaki1',
65      border = 'orange4', xlab = 'delka hlavy (v mm)', ylab = 'f(x)',
66      main = '', axes = F, include.lowest = F)
67 box(bty = 'o')
68 axis(side = 1, at = h$mids)
69 axis(side = 2, las = 1)
70
71 h <- hist(bizyg.WM, breaks = b.bizyg.WM, plot = F)
72 hist(bizyg.WM, breaks = b.bizyg.WM, prob = T, ylim = c(0, 0.07), col = 'khaki1',
73      border = 'orange4', xlab = 'sirka tvare (v mm)', ylab = 'f(y)',
74      main = '', axes = F, include.lowest = F)
75 box(bty = 'o')
76 axis(side = 1, at = h$mids)
77 axis(2, las = 1)
```

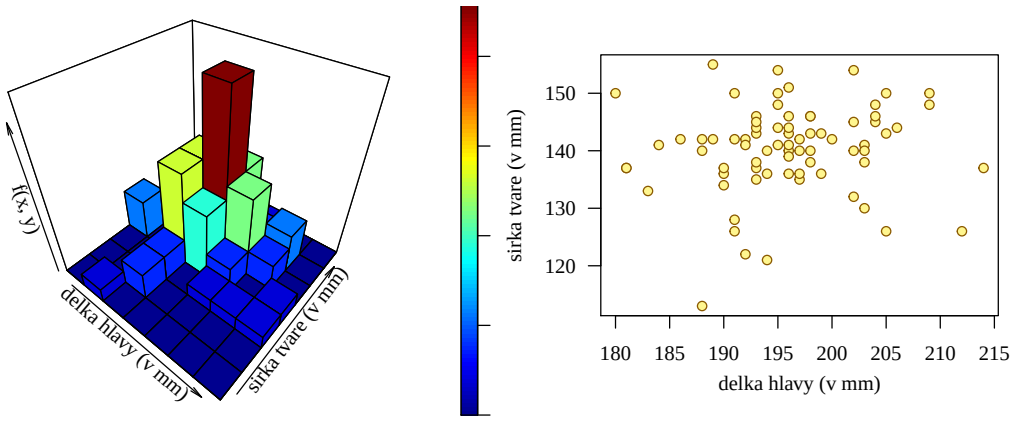


Obrázek 3: Histogram (a) pro znak X popisující délku hlavy mužů (vlevo); (b) pro znak Y popisující šířku tváře mužů (vpravo)

Stereogram vykreslíme pomocí funkce hist3D() z knihovny plot3D. Jediným povinným vstupním argumentem funkce je matice simultánních četnostních hustot f_{jk}. Simultánní četnostní hustoty dopočítáme pomocí vzorce $f_{jk} = \frac{p_{jk}}{d_j h_k}$, kde p_{jk} jsou simultánní relativní četnosti (viz proměnná KT.rel), $d_j = 5$ je délka j-tého třídicího intervalu znaku X a $h_k = 7$ je délka k-tého třídicího intervalu znaku Y, $j, k = 1, \ldots, 7$. Dvourozměrný tečkový diagram vykreslíme příkazem plot(). Stereogram i dvourozměrný tečkový diagram jsou zobrazeny na obrázku 4.

```
78 pjk <- KT.rel; dj <- 5; hk <- 7
79 fjk <- pjk / (dj * hk)
80 plot3D::hist3D(z = fjk, border = "black", xlab = 'delka hlavy (v mm)',
81                ylab = 'sirka tvare (v mm)', zlab = 'f(x, y)')
82 plot(head.LM, bizyg.WM, pch = 21, col = 'orange4', bg = 'khaki1', las = 1,
83      xlab = 'delka hlavy (v mm)', ylab = 'sirka tvare (v mm)')
```

Znaky X a Y jsou četnostně nezávislé, pokud $f_{jk} = f_{j.} f_{.k}$, kde f_{jk} jsou simultánní četnostní hustoty, $f_{j.}$ je marginální četnostní hustota j-tého třídicího intervalu znaku X a $f_{.k}$ je marginální četnostní hustota k-tého třídicího intervalu



Obrázek 4: (a) Stereogram (vlevo); (b) dvourozměrný tečkový diagram (vpravo) pro znaky X a Y

znaku Y, $j, k = 1, \ldots, 7$. Simultánní četnostní hustoty f_{jk} máme vloženy v proměnné fjk.

```
84 round(fjk, 4)
```

```
85 format(round(fjk, 4), scientific = F)
```

Marginální četnostní hustoty znaku X dopočítáme pomocí vzorce $f_{j.} = \frac{p_{j.}}{d_j}$, kde $p_{j.}$ jsou marginální relativní četnosti, které vypočítáme pomocí funkce apply() s argumenty MARGIN = 1 a FUN = sum aplikované na matici simultánních relativních četností pjk, a $d_j = 5$ je šířka j-tého třídicího intervalu znaku X, $j = 1, \ldots, 7$. Marginální četnostní hustoty znaku Y dopočítáme pomocí vzorce $f_{.k} = \frac{p_{.k}}{h_k}$, kde $p_{.k}$ jsou marginální relativní četnosti, které vypočítáme pomocí funkce apply() s argumenty MARGIN = 2 a FUN = sum aplikované na matici simultánních relativních četností pjk a $h_k = 7$ je šířka k-tého třídicího intervalu znaku Y, $k = 1, \ldots, 7$.

```
86 pj. <- apply(pjk, 1, sum)
87 p.k <- apply(pjk, 2, sum)
88 fj. <- pj. / dj
89 f.k <- p.k / hk
90 fj.f.k <- fj.%*% t(f.k)
```

```
          (110;117] (117;124] (124;131] (131;138] (138;145] (145;152] (152;159]  91
(179;184]    0,0000    0,0001    0,0001    0,0005    0,0008    0,0004    0,0001  92
(184;189]    0,0000    0,0001    0,0002    0,0005    0,0010    0,0004    0,0001  93
(189;194]    0,0001    0,0002    0,0005    0,0018    0,0033    0,0014    0,0003  94
(194;199]    0,0001    0,0002    0,0006    0,0022    0,0039    0,0017    0,0004  95
(199;204]    0,0001    0,0001    0,0003    0,0011    0,0020    0,0009    0,0002  96
(204;209]    0,0000    0,0001    0,0002    0,0005    0,0010    0,0004    0,0001  97
(209;214]    0,0000    0,0000    0,0001    0,0002    0,0003    0,0001    0,0000  98
```

Porovnáním matice simultánních četnostních hustot f_{jk} s příslušnými součiny marginálních četnostních hustot $f_{j.}f_{.k}$ vidíme, že znaky X a Y nejsou při daném intervalovém rozložení četností četnostně nezávislé. Například $f_{14} \neq f_{1.}f_{.4}$ $(0,0011 \neq 0,0005)$, $f_{36} \neq f_{3.}f_{.6}$ $(0,0008 \neq 0,0014)$, $f_{65} \neq f_{6.}f_{.5}$ $(0,0008 \neq 0,0010)$, apod. ★

Příklad 2.4. Neřešený příklad
Načtěte datový soubor 05-one-sample-correlation-skull-mf.txt. Za předpokladu, že znak X popisuje největší výšku mozkovny v mm (skull.pH) a znak Y popisuje morfologickou výšku tváře v mm (face.H) u žen (a) roztřiďte data do dvourozměrných třídicích intervalů; (b) vytvořte kontingenční tabulku simultánních absolutních, resp. relativních četností; (c) nakreslete histogram pro znak X, resp. pro znak Y; (d) nakreslete stereogram a dvourozměrný tečkový diagram; (e) zjistěte, zda jsou znaky X a Y četnostně nezávislé při daném dvourozměrném intervalovém rozložení četností.
Výsledky: (a) optimální počet třídicích intervalů $r = 7$; znak X: optimální šířka jednoho třídicího intervalu $d_j = 4$, $j = 1, \ldots, 7$, hranice třídicích intervalů: $u_1 = 117$, $u_2 = 121$, $\ldots$, $u_8 = 145$; znak Y: optimální šířka jednoho třídicího intervalu $h_k = 4$, $k = 1, \ldots, 7$, hranice třídicích intervalů: $u_1 = 93$, $u_2 = 97$, $\ldots$, $u_8 = 121\,\text{mm}$; (b) kontingenční tabulka simultánních absolutních, resp. relativních četností viz tabulka 3, resp. tabulka 4; (c) histogram znaku X, resp. znaku Y viz obrázek 5; (d) stereogram a dvourozměrný tečkový diagram viz obrázek 6; (e) mezi znaky X a Y neexistuje při daném dvourozměrném intervalovém rozložení četností četnostní nezávislost (viz porovnání simultánních četnostních hustot f_{jk} (tabulka 5) se součiny marginálních četnostních hustot $f_{j.}f_{.k}$ (tabulka 6)).

Tabulka 3: Tabulka simultánních absolutních četností n_{jk}

	$\langle 93; 97)$	$\langle 97; 101)$	$\langle 101; 105)$	$\langle 105; 109)$	$\langle 109; 113)$	$\langle 113; 117)$	$\langle 117; 121)$
$\langle 117; 121)$	0	1	0	0	1	0	0
$\langle 121; 125)$	1	0	3	2	1	1	1
$\langle 125; 129)$	0	1	8	4	6	2	1
$\langle 129; 133)$	2	2	4	7	6	3	0
$\langle 133; 137)$	0	2	1	3	6	0	3
$\langle 137; 141)$	0	1	2	2	0	0	0
$\langle 141; 145)$	0	0	0	0	0	0	1

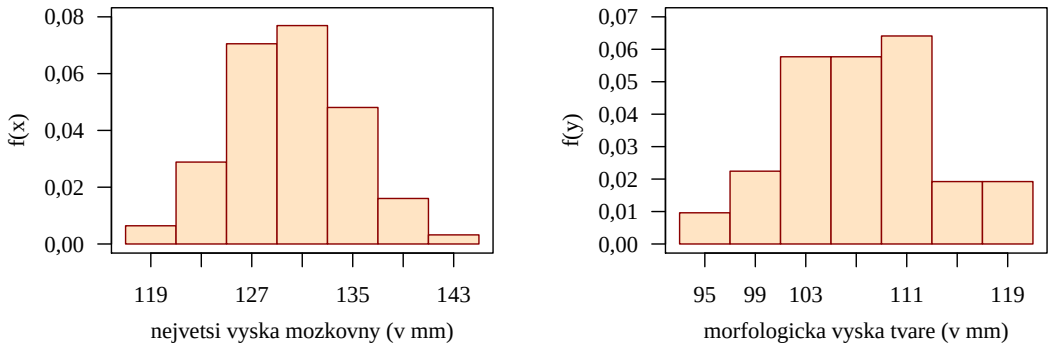
Tabulka 4: Tabulka simultánních relativních četností p_{jk}

	$\langle 93; 97)$	$\langle 97; 101)$	$\langle 101; 105)$	$\langle 105; 109)$	$\langle 109; 113)$	$\langle 113; 117)$	$\langle 117; 121)$
$\langle 117; 121)$	0,0000	0,0128	0,0000	0,0000	0,0128	0,0000	0,0000
$\langle 121; 125)$	0,0128	0,0000	0,0385	0,0256	0,0128	0,0128	0,0128
$\langle 125; 129)$	0,0000	0,0128	0,1026	0,0513	0,0769	0,0256	0,0128
$\langle 129; 133)$	0,0256	0,0256	0,0513	0,0897	0,0769	0,0385	0,0000
$\langle 133; 137)$	0,0000	0,0256	0,0128	0,0385	0,0769	0,0000	0,0385
$\langle 137; 141)$	0,0000	0,0128	0,0256	0,0256	0,0000	0,0000	0,0000
$\langle 141; 145)$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0128

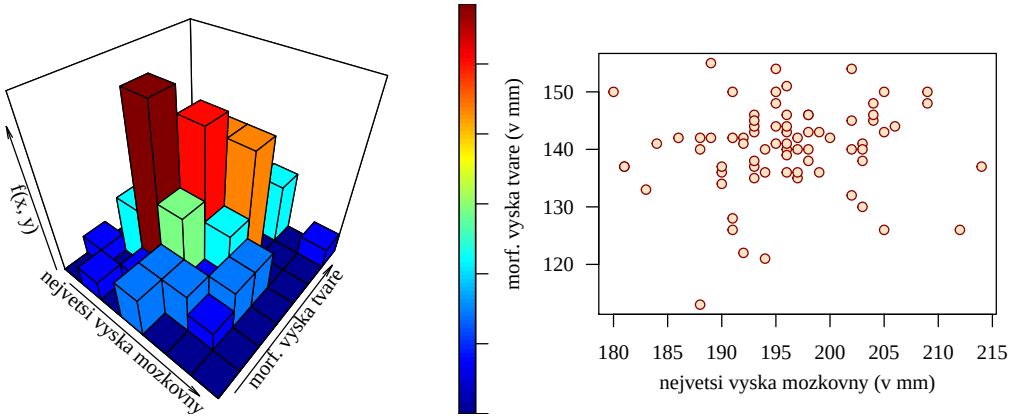
Tabulka 5: Tabulka simultánních četnostních hustot f_{jk}

	$\langle 93; 97)$	$\langle 97; 101)$	$\langle 101; 105)$	$\langle 105; 109)$	$\langle 109; 113)$	$\langle 113; 117)$	$\langle 117; 121)$
$\langle 117; 121)$	0,0000	0,0004	0,0000	0,0000	0,0004	0,0000	0,0000
$\langle 121; 125)$	0,0004	0,0000	0,0011	0,0007	0,0004	0,0004	0,0004
$\langle 125; 129)$	0,0000	0,0004	0,0029	0,0015	0,0022	0,0007	0,0004
$\langle 129; 133)$	0,0007	0,0007	0,0015	0,0026	0,0022	0,0011	0,0000
$\langle 133; 137)$	0,0000	0,0007	0,0004	0,0011	0,0022	0,0000	0,0011
$\langle 137; 141)$	0,0000	0,0004	0,0007	0,0007	0,0000	0,0000	0,0000
$\langle 141; 145)$	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0004

★



Obrázek 5: Histogram (a) pro znak X popisující největší výšku mozkovny žen (vlevo); (b) pro znak Y popisující morfologickou výšku tváře žen (vpravo)



Obrázek 6: (a) Stereogram (vlevo); (b) dvourozměrný tečkový diagram (vpravo) pro znaky X a Y

Tabulka 6: Tabulka součinů marginálních četnostních hustot $f_{j.}f_{.k}$

	$\langle 93;97\rangle$	$\langle 97;101\rangle$	$\langle 101;105\rangle$	$\langle 105;109\rangle$	$\langle 109;113\rangle$	$\langle 113;117\rangle$	$\langle 117;121\rangle$
$\langle 117;121\rangle$	0,0000	0,0001	0,0002	0,0002	0,0002	0,0001	0,0001
$\langle 121;125\rangle$	0,0001	0,0003	0,0008	0,0008	0,0008	0,0003	0,0003
$\langle 125;129\rangle$	0,0003	0,0007	0,0019	0,0019	0,0021	0,0006	0,0006
$\langle 129;133\rangle$	0,0003	0,0008	0,0020	0,0020	0,0023	0,0007	0,0007
$\langle 133;137\rangle$	0,0002	0,0005	0,0013	0,0013	0,0014	0,0004	0,0004
$\langle 137;141\rangle$	0,0001	0,0002	0,0004	0,0004	0,0005	0,0001	0,0001
$\langle 141;145\rangle$	0,0000	0,0000	0,0001	0,0001	0,0001	0,0000	0,0000