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 $\langle u_1; u_2\rangle, \ldots, \langle u_r; u_{r+1}\rangle$. Hovoříme pak o intervalovém rozložení četností.

Označení:

- $\langle u_j; u_{j+1}\rangle$ – 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, jinak řečeno ekvidistantní. Jejich počet určíme například pomocí Sturgesova pravidla: $r \approx 1 + 3,3\log_{10}(n)$, kde n je rozsah souboru.

Názvy četností jsou podobné jako v sekci 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 2.1).

Tabulka 2.1: Tabulka rozložení četností znaku X

$\langle u_j; u_{j+1}\rangle$	d_j	$x_{[j]}$	n_j	p_j	N_j	F_j	f_j
$\langle u_1; u_2\rangle$	d_1	$x_{[1]}$	n_1	p_1	N_1	F_1	f_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\langle u_r; u_{r+1}\rangle$	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 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 obsahující údaje o délce hlavy v mm (head.L) u mladých dospělých mužů a žen. Pro znak X popisující délku hlavy u mužů (a) zjistěte hranice třídicích intervalů; (b) vytvořte tabulku rozložení četností; (c) nakreslete histogram a graf intervalové empirické distribuční funkce $F(x)$.

Řešení příkladu 2.1

Datový soubor načteme příkazem read.delim() s argumenty sep = "\t" a dec = ".". Pro náhled datového souboru si zobrazíme prvních pět řádků tabulky pomocí funkce head() s argumentem n = 5.

```
data <- read.delim("16-anova-head.txt", sep = "\t", dec = ".")
head(data, n = 5)
```

```
  sex sexor obra body.H head.L head.W bigo.W bizyg.W
1   m  <NA>   no   1782    180    159    109     150
2   m    op   no   1824    195    163    108     150
3   f  <NA>   no   1631    197    144    100     138
4   f    op   no   1576    186    148    105     127
5   m    op   no   1792    197    151    106     142
```

Z načtených dat vybereme pomocí operátoru [ ] pouze řádky týkající se mužů a sloupec head.L. Z výběru odstraníme řádky s chybějícími pozorováními příkazem na.omit(). Ke stanovení hranic třídicích intervalů potřebujeme znát rozsah datového souboru a minimální a maximální naměřenou hodnotu délky hlavy u mužů. Rozsah datového souboru zjistíme příkazem length(), minimální a maximální naměřenou hodnotu příkazem range(). Optimální počet třídicích intervalů stanovíme pomocí vzorce $r \approx 1 + 3,3 \log_{10}(n)$. Hodnotu logaritmu při základu 10, tj. $\log_{10}(n)$, vypočítáme příkazem log10().

```
head.LM <- na.omit(data[data$sex == "m", "head.L"])
n <- length(head.LM) # 75
range(head.LM) # 180; 214
r <- round(1 + 3.3 * log10(n)) # 7
```

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 délka jednoho třídicího intervalu $d = \frac{35}{7} = 5$ mm. Hranice třídicích intervalů tedy stanovíme jako posloupnost $u_1 = 179$, $u_2 = 184$, $\ldots$, $u_8 = 214$ mm. Celkem získáme osm hranic definujících sedm třídicích intervalů. Posloupnost hranic vytvoříme příkazem seq().

```
# 214 - 179 = 35; 35 / 7 = 5 -> seq(179, 214, by = 5)
b.head.LM <- seq(from = 179, to = 214, by = 5) # 179, 184, ..., 214
```

Tabulku rozložení četností dopočítáme pomocí funkce trc() implementované v R skriptu AS1-sbirka-funkce.R. Vstupními argumenty funkce jsou vektor naměřených hodnot x a hranice třídicích intervalů breaks. Výstupem funkce je tabulka rozložení četností, kterou si zobrazíme zaokrouhlenou na čtyři desetinná místa.

```
source("AS1-sbirka-funkce.R")
tab.rc <- trc(head.LM, b.head.LM)
round(tab.rc, 4)
```

```
          dj    xj nj     pj Nj     Fj     fj
(179;184]  5 181,5  5 0,0667  5 0,0667 0,0133
(184;189]  5 186,5  6 0,0800 11 0,1467 0,0160
(189;194]  5 191,5 20 0,2667 31 0,4133 0,0533
(194;199]  5 196,5 24 0,3200 55 0,7333 0,0640
(199;204]  5 201,5 12 0,1600 67 0,8933 0,0320
(204;209]  5 206,5  6 0,0800 73 0,9733 0,0160
(209;214]  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.

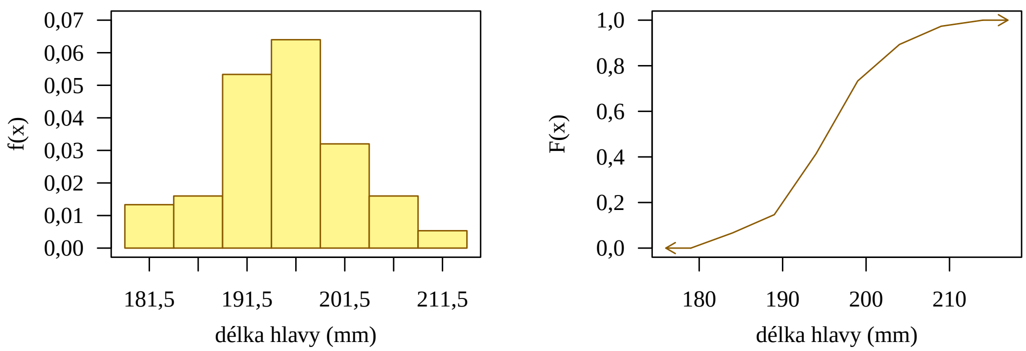
Histogram vykreslíme příkazem hist() s argumentem prob = T. Tento argument zajistí, že plocha obdélníků v histogramu bude odpovídat relativním četnostem a výšky obdélníků četnostním hustotám. V histogramu dále nastavíme hranice (breaks = b.head.LM) zleva otevřených (include.lowest = F) třídicích intervalů, rozsah osy y (ylim = c(0, 0.07)), potlačení vykreslení os x a y (axes = F), barvu výplně obdélníků (col), barvu obrysu obdélníků (border), popisek osy x (xlab), popisek osy y (ylab) a potlačení nadpisu (main = ""). Okolo grafu dokreslíme rámeček příkazem box(). Osu x, resp. osu y vykreslíme samostatně pomocí příkazu axis(). V souladu se zavedenou konvencí vykreslíme měřítko osy x ve středech třídicích intervalů. Středy intervalů můžeme získat buď ručním výpočtem nebo jako výstup mids funkce hist() se specifikací argumentu plot = F. Tento argument umožní provedení všech výpočtů potřebných k vytvoření histogramu, ale potlačí jeho vykreslení. Výsledný histogram je zobrazen na obrázku 2.1 vlevo.

```
hist(head.LM, prob = T, breaks = b.head.LM, include.lowest = F, ylim = c(0, 0.07), axes
    = F, col = "khaki1", border = "orange4", xlab = "délka hlavy (mm)", ylab = "f(x)",
    main = "", )
box(bty = "o")
c.head.LM <- hist(head.LM, breaks = b.head.LM, plot = F)$mids
axis(side = 1, at = c.head.LM)
```

```
axis(side = 2, las = 1)
```

K vykreslení grafu intervalové empirické distribuční funkce $F(x)$ potřebujeme zadefinovat její průběh. Funkce $F(x)$ je až do hodnoty 179 rovná nule, v bodech 184, 189, ..., 214 její hodnoty odpovídají hodnotám kumulativní distribuční funkce F_j a od hodnoty 214 je tato funkce rovná jedné. Graf funkce $F(x)$ vykreslíme příkazem plot() s argumentem type = "l". Prvním vstupním argumentem příkazu bude vektor hranic třídicích intervalů, druhým argumentem bude vektor hodnot funkce $F(x)$. Do grafu dále dokreslíme šipku značící nulovou hodnotu empirické distribuční funkce $F(x)$ ve všech hodnotách menších než 179, a dále šipku značící jednotkovou hodnotu funkce $F(x)$ ve všech hodnotách větších než 214. Šipky vykreslíme pomocí funkce arrows(). Prvními čtyřmi vstupními argumenty funkce jsou x-ová a y-ová souřadnice počátečního bodu (argumenty x0 a y0) a x-ová a y-ová souřadnice koncového bodu (argumenty x1 a y1). Při definování souřadnic počátečního a koncového bodu je třeba mít na paměti, že hrot šipky se vždy vykresluje v koncovém bodu. Dále specifikujeme barvu šipky (argument col) a délku hrotu šipky (argument length). Výsledný graf intervalové empirické distribuční funkce $F(x)$ je zobrazen na obrázku 2.1 vpravo.

```
Fx <- c(0, tab.rc$Fj)
plot(b.head.LM, Fx, type = "l", las = 1, col = "orange4", xlim = c(176, 217), xlab =
    "délka hlavy (mm)", ylab = "F(x)")
arrows(x0 = 179, y0 = 0, x1 = 176, y1 = 0, col = "orange4", length = 0.1)
arrows(x0 = 214, y0 = 1, x1 = 217, y1 = 1, col = "orange4", length = 0.1)
```



Obrázek 2.1: Histogram (vlevo); graf intervalové empirické distribuční funkce $F(x)$ (vpravo) pro délku hlavy u mužů

★

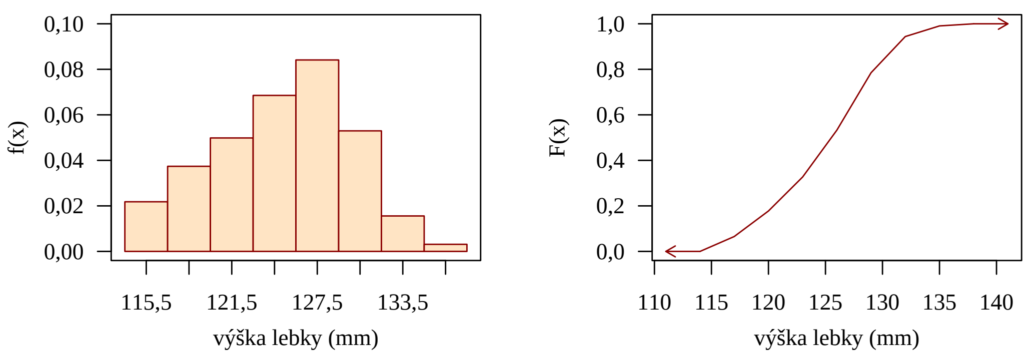
Příklad 2.2. Neřešený příklad

Načtěte datový soubor 11-two-samples-means-skull.txt obsahující údaje o výšce lebky v mm (skull.H) u mužů a u žen starověké egyptské populace. Pro znak X popisující výšku lebky u žen (a) zjistěte hranice třídicích intervalů; (b) vytvořte tabulku rozložení četností; (c) nakreslete histogram a graf intervalové empirické distribuční funkce $F(x)$.

Výsledky: (a) rozsah datového souboru $n = 107$, optimální počet třídicích intervalů $r = 8$, rozsah naměřených hodnot je 117 až 136 mm, optimální délka jednoho třídicího intervalu $d = 3$, hranice třídicích intervalů: $u_1 = 114$, $u_2 = 117, \ldots, u_9 = 138$; (b) tabulka rozložení četností viz tabulka 2.2; (c) histogram, resp. graf intervalové empirické distribuční funkce $F(x)$ viz obrázek 2.2.

Tabulka 2.2: Tabulka rozložení četností pro výšku lebky u žen

	d_j	$x_{[j]}$	n_j	p_j	N_j	F_j	f_j
$(114; 117\rangle$	3,00	115,50	7	0,07	7	0,07	0,02
$(117; 120\rangle$	3,00	118,50	12	0,11	19	0,18	0,04
$(120; 123\rangle$	3,00	121,50	16	0,15	35	0,33	0,05
$(123; 126\rangle$	3,00	124,50	22	0,21	57	0,53	0,07
$(126; 129\rangle$	3,00	127,50	27	0,25	84	0,79	0,08
$(129; 132\rangle$	3,00	130,50	17	0,16	101	0,94	0,05
$(132; 135\rangle$	3,00	133,50	5	0,05	106	0,99	0,02
$(135; 138\rangle$	3,00	136,50	1	0,01	107	1,00	0,00



Obrázek 2.2: Histogram (vlevo); graf intervalové empirické distribuční funkce $F(x)$ (vpravo) pro výšku lebky u žen

★

2.2 Dvourozměrné intervalové rozložení četností

Ve dvourozměrném datovém souboru o rozsahu n roztřídíme hodnoty znaku X do r ekvidistantních třídicích intervalů $(u_j; u_{j+1}\rangle$, $j = 1, \ldots, r$ s jednotnou délkou d a hodnoty znaku Y taktéž roztřídíme do r ekvidistantních třídicích intervalů $(v_k; v_{k+1}\rangle$, $k = 1, \ldots, r$ s jednotnou délkou h. 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 sekci 1.3. Simultánní absolutní četnosti, resp. simultánní relativní četnosti zapisujeme do kontingenční tabulky (viz tabulka 2.3 vlevo, resp. vpravo).

Tabulka 2.3: Kontingenční tabulka (a) simultánních absolutních četností (vlevo); (b) simultánních relativních četností (vpravo)

	Y		
X	$y_{[1]}$	$\ldots$	$y_{[s]}$
$x_{[1]}$	n_{11}	$\ldots$	n_{1s}
$\vdots$	$\vdots$	$\ldots$	$\vdots$
$x_{[r]}$	n_{r1}	$\ldots$	n_{rs}

	Y		
X	$y_{[1]}$	$\ldots$	$y_{[s]}$
$x_{[1]}$	p_{11}	$\ldots$	p_{1s}
$\vdots$	$\vdots$	$\ldots$	$\vdots$
$x_{[r]}$	p_{r1}	$\ldots$	p_{rs}

Navíc zavádíme simultánní a marginální četnostní hustoty:

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

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

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

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

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

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

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

Příklad 2.3. Řešený příklad

Načtěte datový soubor 16-anova-head.txt obsahující údaje o délce hlavy v mm (head.L) a šířce tváře v mm (bizyg.W) u mladých dospělých mužů a žen. Pro znak X popisující délku hlavy a znak Y popisující šířku tváře u mužů (a) zjistěte hranice 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 a z načtených dat vybereme pouze řádky týkající se mužů a sloupce head.L a bizyg.W. Z výběru odstraníme řádky s chybějícími pozorováními příkazem na.omit(). V dalším kroku si z výběru separujeme hodnoty délky hlavy, které vložíme do proměnné head.LM, a hodnoty šířky tváře, které vložíme do proměnné bizyg.WM. Ke stanovení hranic třídicích intervalů potřebujeme znát rozsah dvourozměrného datového souboru a minimální a maximální naměřenou hodnotu každého znaku. Rozsah datového souboru zjistíme příkazem length(), minimální a maximální naměřenou hodnotu příkazem range(). Optimální počet třídicích intervalů stanovíme na základě Sturgesova pravidla. Počet třídicích intervalů je stejný pro oba znaky, neboť závisí pouze na rozsahu datového souboru.

```
data <- read.delim("16-anova-head.txt", sep = "\t", dec = ".")
data.M <- data[data$sex == "m", c("head.L", "bizyg.W")]
data.M <- na.omit(data.M)
head.LM <- data.M$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
```

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 ekvidistantních třídicích intervalů o optimální délce $d = 5$ mm, a to stanovením hranic $u_1 = 179$, $u_2 = 184$, …, $u_8 = 214$ 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 délka jednoho třídicího intervalu $h = \frac{49}{7} = 7$ mm. Hranice třídicích intervalů pro šířku tváře tedy stanovíme jako posloupnost $v_1 = 110$, $v_2 = 117$, …, $v_8 = 159$ mm.

```
# head.LM: viz priklad 2.1
# bizyg.WM: 155 - 112 = 43 => 49 / 7 = 7 -> seq(110, 159, by = 7)
b.head.LM <- seq(from = 179, to = 214, by = 5)
b.bizyg.WM <- seq(from = 110, to = 159, by = 7)
```

V dalším kroku roztřídíme naměřené hodnoty každého znaku do příslušných třídicích intervalů, a to pomocí příkazu cut(). Vstupními argumenty příkazu jsou vektor naměřených hodnot a vektor hranic třídicích intervalů (argument breaks). 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 právě vektory head.LM.c a bizyg.WM.c.

```
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
```

Kontingenční tabulku simultánních relativních četností získáme příkazem prop.table().

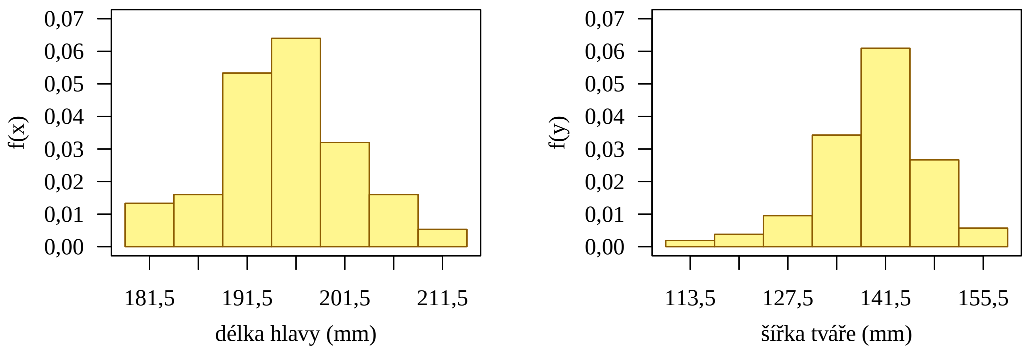
```
KT.rel <- prop.table(KT.abs)
```

```
           (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 hlavy 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 hlavy 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 hlavy 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 funkce hist() s argumentem prob = T. Hranice třídicích intervalů definujeme explicitně pomocí argumentu breaks. Měřítko osy x vykreslíme v hodnotách středů třídicích intervalů příkazem axis(). Histogramy pro oba znaky jsou zobrazeny na obrázku 2.3.

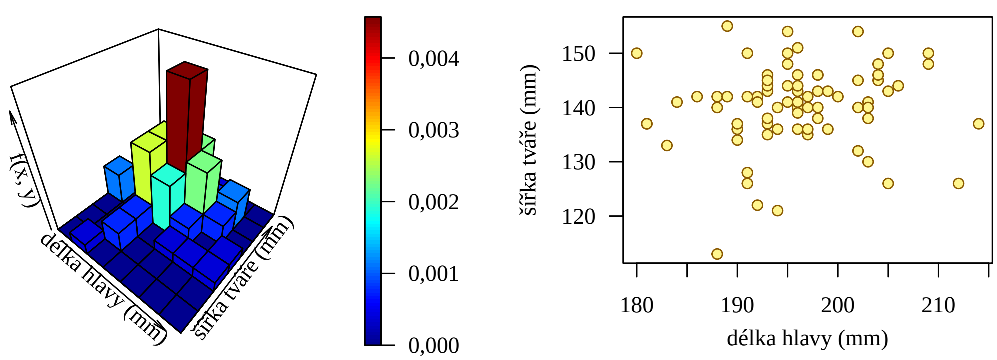
```
hist(head.LM, prob = T, breaks = b.head.LM, include.lowest = F, ylim = c(0, 0.07), axes
    = F, col = "khaki1", border = "orange4", xlab = "délka hlavy (mm)", ylab = "f(x)",
    main = "")
box(bty = "o")
c.head.LM <- hist(head.LM, breaks = b.head.LM, plot = F)$mids
axis(side = 1, at = c.head.LM)
axis(side = 2, las = 1)

hist(bizyg.WM, breaks = b.bizyg.WM, prob = T, ylim = c(0, 0.07), col = "khaki1", border
    = "orange4", xlab = "šířka tváře (mm)", ylab = "f(y)", main = "", axes = F,
    include.lowest = F)
box(bty = "o")
c.bizyg.WM <- hist(bizyg.WM, breaks = b.bizyg.WM, plot = F)$mids
axis(side = 1, at = c.bizyg.WM)
axis(2, las = 1)
```



Obrázek 2.3: Histogram pro délku hlavy u mužů (vlevo); histogram pro šířku tváře u mužů (vpravo)

K vykreslení stereogramu potřebujeme znát hodnoty simultánních četnostních hustot f_{jk}. Ty dopočítáme pomocí vzorce $f_{jk} = \frac{p_{jk}}{dh}$, kde p_{jk}, $j,k = 1,\ldots,7$ jsou simultánní relativní četnosti (viz proměnná KT.rel), $d = 5$ je ekvidistantní délka třídicích intervalů znaku X a $h = 7$ je ekvidistantní délka třídicích intervalů znaku Y. Stereogram potom vykreslíme pomocí funkce hist3D() z knihovny plot3D. Vstupními argumenty funkce budou matice simultánních četnostních hustot (z = as.matrix(fjk)), barva obrysu sloupců (border = "black") a popisky os x, y a z (xlab, ylab a zlab). Dvourozměrný tečkový diagram vykreslíme příkazem plot(). Stereogram je zobrazen na obrázku 2.4 vlevo. Barevné spektrum přecházející od odstínů modré po červenou vizuálně rozlišuje hodnoty simultánních četnostních hustot. Dvourozměrný tečkový diagram je zobrazen na obrázku 2.4 vpravo.

```
pjk <- KT.rel; d <- 5; h <- 7
fjk <- pjk / (d * h)
plot3D::hist3D(z = as.matrix(fjk), border = "black", xlab = "délka hlavy (mm)", ylab =
    "šířka tváře (mm)", zlab = "f(x, y)")
plot(head.LM, bizyg.WM, pch = 21, col = "orange4", bg = "khaki1", las = 1, xlab =
    "délka hlavy (mm)", ylab = "šířka tváře (mm)")
```



Obrázek 2.4: Stereogram (vlevo); dvourozměrný tečkový diagram (vpravo) pro délku hlavy a šířku tváře u mužů

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 v j-tém třídicím intervalu pro znak X a $f_{.k}$ je marginální četnostní hustota v k-tém třídicím intervalu pro znak Y, $j,k = 1,\ldots,7$. Simultánní četnostní hustoty f_{jk} máme vloženy v proměnné fjk. Tuto proměnnou si zobrazíme zaokrouhlenou na čtyři desetinná místa pomocí funkce round() s argumentem digits = 4.

```
round(fjk, digits = 4)
```

```
         (110;117] (117;124] (124;131] (131;138] (138;145] (145;152] (152;159]
(179;184]    0,0000    0,0000    0,0000    0,0011    0,0004    0,0004    0,0000
(184;189]    0,0004    0,0000    0,0000    0,0000    0,0015    0,0000    0,0004
(189;194]    0,0000    0,0008    0,0008    0,0027    0,0027    0,0008    0,0000
(194;199]    0,0000    0,0000    0,0000    0,0019    0,0046    0,0023    0,0004
(199;204]    0,0000    0,0000    0,0004    0,0008    0,0023    0,0008    0,0004
(204;209]    0,0000    0,0000    0,0004    0,0000    0,0008    0,0011    0,0000
(209;214]    0,0000    0,0000    0,0004    0,0004    0,0000    0,0000    0,0000
```

Marginální četnostní hustotu $f_{j.}$ dopočítáme pomocí vzorce $f_{j.} = \frac{p_{j.}}{d}$, kde $p_{j.}$ je marginální relativní četnost k-té varianty znaku X a d je ekvidistantní délka j-tého třídicího intervalu. Marginální relativní četnosti $p_{j.}$, $j = 1, \ldots, 7$ vypočítáme pomocí funkce apply() s argumenty MARGIN = 1 a FUN = sum aplikované na matici pjk. Marginální četnostní hustotu $f_{.k}$ dopočítáme pomocí vzorce $f_{.k} = \frac{p_{.k}}{h}$, kde $p_{.k}$ je marginální relativní četnost k-té varianty znaku Y a h je ekvidistantní délka k-tého třídicího intervalu. Marginální relativní četnosti $p_{.k}$, $k = 1, \ldots, 7$ vypočítáme pomocí funkce apply() s argumenty MARGIN = 2 a FUN = sum aplikované na matici pjk. Matici součinů $f_{j.}f_{.k}$ nakonec vypočítáme maticovým vynásobením vektoru fj. s transponovaným vektorem f.k. Operaci maticového násobení zadefinujeme operátorem %*%, transpozici vektoru f.k provedeme pomocí funkce t(). Matici součinů si opět zobrazíme zaokrouhlenou na čtyři desetinná místa.

```r
pj. <- apply(pjk, MARGIN = 1, FUN = sum)
p.k <- apply(pjk, MARGIN = 2, FUN = sum)
fj. <- pj. / d
f.k <- p.k / h
fj.f.k <- fj. %*% t(f.k)
round(fj.f.k, digits = 4)
```

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

Porovnáním matice simultánních četnostních hustot f_{jk} s příslušnými součiny marginálních četnostních hustot $f_{j.}f_{.k}$ vidíme, že znaky X a Y nejsou četnostně nezávislé při daném intervalovém rozložení četností. 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 obsahující údaje o největší výšce mozkovny v mm (skull.pH) a morfologické výšce tváře v mm (face.H) u mužů a žen starověké egyptské populace. Pro znak X popisující největší výšku mozkovny a znak Y popisující morfologickou výšku tváře u žen (a) zjistěte hranice 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.

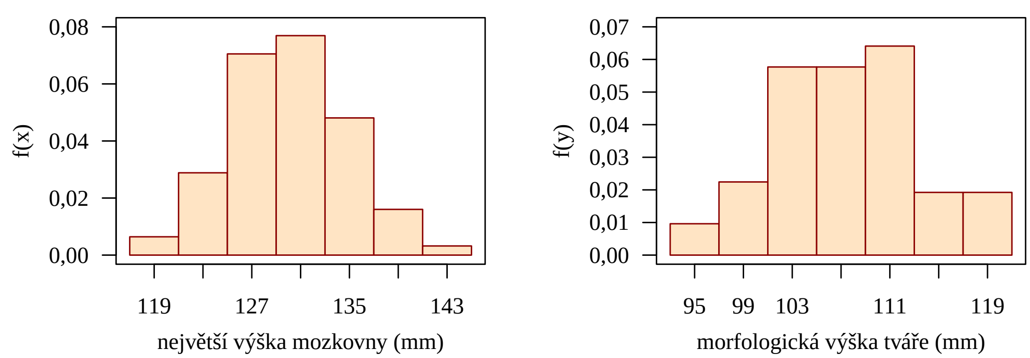
Výsledky: (a) rozsah náhodného výběru $n = 78$, optimální počet třídicích intervalů $r = 7$; znak X: rozsah naměřených hodnot je 120 až 142 mm, optimální délka jednoho třídicího intervalu $d = 4$, hranice třídicích intervalů: $u_1 = 117$, $u_2 = 121$, …, $u_8 = 145$; znak Y: rozsah naměřených hodnot je 95 až 119 mm, optimální délka jednoho třídicího intervalu $h = 4$, hranice třídicích intervalů: $u_1 = 93$, $u_2 = 97$, …, $u_8 = 121$ mm; (b) kontingenční tabulka simultánních absolutních, resp. relativních četností viz tabulka 2.4, resp. tabulka 2.5; (c) histogram znaku X, resp. znaku Y viz obrázek 2.5; (d) stereogram a dvourozměrný tečkový diagram viz obrázek 2.6; (e) mezi znaky X a Y neexistuje četnostní nezávislost při daném dvourozměrném intervalovém rozložení četností (viz porovnání simultánních četnostních hustot f_{jk} (tabulka 2.6) se součiny marginálních četnostních hustot $f_{j.}f_{.k}$ (tabulka 2.7)).

Tabulka 2.4: Tabulka simultánních absolutních četností n_{jk} pro největší výšku mozkovny a morfologickou výšku tváře u žen

	⟨93; 97⟩	⟨97; 101⟩	⟨101; 105⟩	⟨105; 109⟩	⟨109; 113⟩	⟨113; 117⟩	⟨117; 121⟩
⟨117; 121⟩	0	1	0	0	1	0	0
⟨121; 125⟩	1	0	3	2	1	1	1
⟨125; 129⟩	0	1	8	4	6	2	1
⟨129; 133⟩	2	2	4	7	6	3	0
⟨133; 137⟩	0	2	1	3	6	0	3
⟨137; 141⟩	0	1	2	2	0	0	0
⟨141; 145⟩	0	0	0	0	0	0	1

Tabulka 2.5: Tabulka simultánních relativních četností p_{jk} pro největší výšku mozkovny a morfologickou výšku tváře u žen

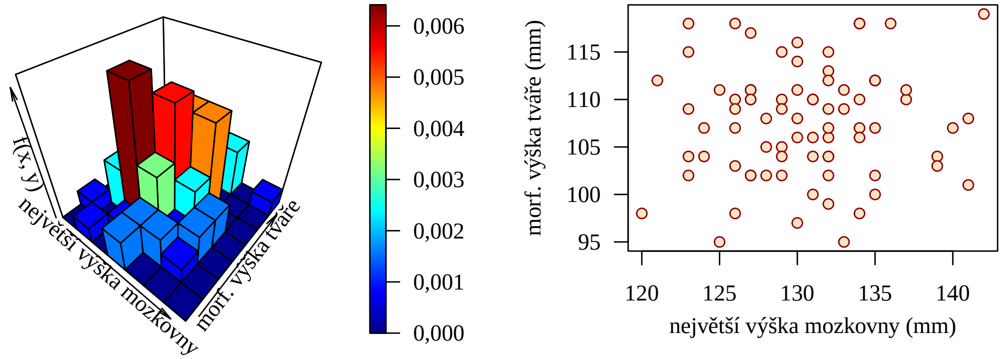
	⟨93; 97⟩	⟨97; 101⟩	⟨101; 105⟩	⟨105; 109⟩	⟨109; 113⟩	⟨113; 117⟩	⟨117; 121⟩
⟨117; 121⟩	0,0000	0,0128	0,0000	0,0000	0,0128	0,0000	0,0000
⟨121; 125⟩	0,0128	0,0000	0,0385	0,0256	0,0128	0,0128	0,0128
⟨125; 129⟩	0,0000	0,0128	0,1026	0,0513	0,0769	0,0256	0,0128
⟨129; 133⟩	0,0256	0,0256	0,0513	0,0897	0,0769	0,0385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



Obrázek 2.5: Histogram pro největší výšku mozkovny u žen (vlevo); histogram pro morfologickou výšku tváře u žen (vpravo)

Tabulka 2.6: Tabulka simultánních četnostních hustot f_{jk} pro největší výšku mozkovny a morfologickou výšku tváře u žen

	⟨93; 97⟩	⟨97; 101⟩	⟨101; 105⟩	⟨105; 109⟩	⟨109; 113⟩	⟨113; 117⟩	⟨117; 121⟩
⟨117; 121⟩	0,0000	0,0008	0,0000	0,0000	0,0008	0,0000	0,0000
⟨121; 125⟩	0,0008	0,0000	0,0024	0,0016	0,0008	0,0008	0,0008
⟨125; 129⟩	0,0000	0,0008	0,0064	0,0032	0,0048	0,0016	0,0008
⟨129; 133⟩	0,0016	0,0016	0,0032	0,0056	0,0048	0,0024	0,0000
⟨133; 137⟩	0,0000	0,0016	0,0008	0,0024	0,0048	0,0000	0,0024
⟨137; 141⟩	0,0000	0,0008	0,0016	0,0016	0,0000	0,0000	0,0000
⟨141; 145⟩	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0008



Obrázek 2.6: Stereogram (vlevo); dvourozměrný tečkový diagram (vpravo) pro největší výšku mozkovny a morfologickou výšku tváře u žen

Tabulka 2.7: Tabulka součinů marginálních četnostních hustot $f_{j.}f_{.k}$ pro největší výšku mozkovny a morfologickou výšku tváře u žen

	$\langle 93; 97)$	$\langle 97; 101)$	$\langle 101; 105)$	$\langle 105; 109)$	$\langle 109; 113)$	$\langle 113; 117)$	$\langle 117; 121)$
$\langle 117; 121)$	0,0001	0,0001	0,0004	0,0004	0,0004	0,0001	0,0001
$\langle 121; 125)$	0,0003	0,0006	0,0017	0,0017	0,0018	0,0006	0,0006
$\langle 125; 129)$	0,0007	0,0016	0,0041	0,0041	0,0045	0,0014	0,0014
$\langle 129; 133)$	0,0007	0,0017	0,0044	0,0044	0,0049	0,0015	0,0015
$\langle 133; 137)$	0,0005	0,0011	0,0028	0,0028	0,0031	0,0009	0,0009
$\langle 137; 141)$	0,0002	0,0004	0,0009	0,0009	0,0010	0,0003	0,0003
$\langle 141; 145)$	0,0000	0,0001	0,0002	0,0002	0,0002	0,0001	0,0001

★