1 Bodové rozložení četností

1.1 Základní, výběrový a datový soubor

Na množině objektů $\{\epsilon_1, \dots, \epsilon_n\}$, která tvoří výběrový soubor rozsahu n a jejíž prvky byly vybrány ze základního souboru E, zjišťujeme hodnoty znaku X. Hodnota znaku X na objektu ϵ_i se značí x_i, $i = 1, \dots, n$. Tyto hodnoty zaznamenáme do jednorozměrného datového souboru $(x_1, \dots, x_n)^T$. Uspořádané hodnoty $x_1 \leq \cdots \leq x_n$ tvoří uspořádaný datový soubor $(x_{(1)}, \dots, x_{(n)})^T$ a navzájem různé uspořádané hodnoty $x_{[1]} < \cdots < x_{[r]}$, $r \leq n$, tvoří vektor variant $(x_{[1]}, \dots, x_{[r]})^T$.

1.2 Jednorozměrné bodové rozložení četností

Je-li počet variant r malý ve srovnání s rozsahem n výběrového souboru, přiřazujeme četnosti jednotlivým variantám a hovoříme o bodovém rozložení četností. Pro $j = 1, \dots, r$ definujeme:

- n_j = počet těch objektů ve výběrovém souboru, pro něž $X = x_{[j]}$ – absolutní četnost varianty $x_{[j]}$ ve výběrovém souboru,
- $p_j = \frac{n_j}{n}$ – relativní četnost varianty $x_{[j]}$ ve výběrovém souboru,
- $N_j = N(X \leq x_{[j]}) = n_1 + \cdots + n_j$ – absolutní kumulativní četnost prvních j variant ve výběrovém souboru,
- $F_j = \frac{N_j}{n} = p_1 + \cdots + p_j$ – relativní kumulativní četnost prvních j variant ve výběrovém souboru.

Tabulka 1.1 se nazývá variační řada (nebo též tabulka rozložení četností). Absolutní či relativní četnosti znázorňujeme graficky například pomocí sloupkového diagramu či polygonu četností.

Tabulka 1.1: Variační řada znaku X

$x_{[j]}$	n_j	p_j	N_j	F_j
$x_{[1]}$	n_1	p_1	N_1	F_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_{[r]}$	n_r	p_r	N_r	F_r

Pomocí relativních četností definujeme četnostní funkci:

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

Pomocí relativních kumulativních četností definujeme empirickou distribuční funkci:

$$F(x) = \begin{cases} 0 \text{ pro } x < x_{[1]}, \\ F_j \text{ pro } x_{[j]} \leq x < x_{[j+1]}, j = 1, \dots, r-1, \\ 1 \text{ pro } x \geq x_{[r]}. \end{cases}$$

Mezi četnostní funkcí a empirickou distribuční funkcí platí součtový vztah: $\forall x \in \mathbb{R} : F(x) = \sum_{t \leq x} p(t)$.

Příklad 1.1. Řešený příklad

Načtěte datový soubor 22-multinom-palmar-lines.txt obsahující údaje o zakončení tří hlavních dlaňových linií (Lo – nízké; Mi – střední; Hi – vysoké) a údaje o odstínu barvy vlasů (LiH – světlý; MH – střední; DaH – tmavý) u mužů a žen. Za předpokladu, že znak X popisuje zakončení tří hlavních dlaňových linií u mužů, (a) popište jednotlivé varianty znaku X; (b) vytvořte variační řadu; (c) nakreslete sloupkový diagram a polygon četností; (d) nakreslete graf četnostní funkce $p(x)$ a graf empirické distribuční funkce $F(x)$.

Řešení příkladu 1.1

Datový soubor načteme pomocí funkce read.delim(). Vstupními argumenty funkce jsou název datového souboru, argument sep = "\t" definující, že sloupce jsou od sebe odděleny tabulátorem, a argument row.names = 1 určující, že hodnoty vložené v prvním sloupci mají být vnímány jako názvy řádků načtené tabulky. Často se přidává také argument dec, kterým specifikujeme oddělovač desetinných míst použitý v datovém souboru.

```
data <- read.delim("22-multinom-palmar-lines.txt", sep = "\t", row.names = 1)
```

```
    Lo.m Mi.m Hi.m Lo.f Mi.f Hi.f
LiH    4    6    6    6    6    4
MH     7   15   20   10   10   18
DaH   12   12   18   12   22   12
```

Z načtených dat dále vybereme pomocí operátoru [ ] první tři sloupce tabulky, neboť právě ty obsahují údaje pro muže. Absolutní četnosti n_j jednotlivých variant znaku X získáme jako součty hodnot v jednotlivých sloupcích tabulky. Sloupcové součty vypočítáme pomocí funkce apply() s argumenty FUN = sum (součet) a MARGIN = 2 (po sloupcích). Nyní se zaměříme na výpočet variační řady. Příkazem source() nejprve načteme R skript AS1-sbirka-funkce.R, ve kterém je implementovaná funkce variacni.rada(). Dále si pomocí funkce rep() vytvoříme vektor s hodnotami 1, 2 a 3, kterých bude tolik, kolik vyšla absolutní četnost výskytu nízkého, středního a vysokého zakončení tří hlavních dlaňových linií. Vstupními argumenty funkce variacni.rada() bude potom tento vektor a argument row.names specifikující názvy řádků variační řady odpovídající variantám znaku X.

```
data.M <- data[, 1:3]
nj <- apply(data.M, MARGIN = 2, FUN = sum)
source("AS1-sbirka-funkce.R")
palmar.M <- rep(1:3, times = nj)
vr <- variacni.rada(palmar.M, row.names = c("nizke", "stredni", "vysoke"))
```

```
        nj   pj  Nj   Fj
nizke   23 0,23  23 0,23
stredni 33 0,33  56 0,56
vysoke  44 0,44 100 1,00
```

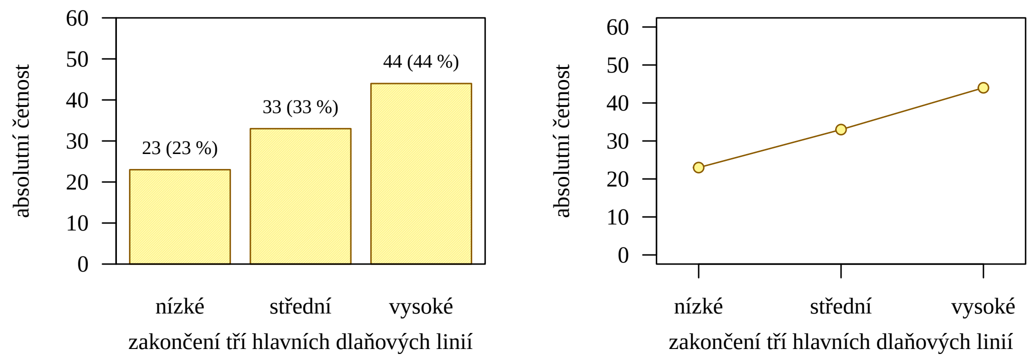
Znak X popisující zakončení tří hlavních dlaňových linií má tři varianty: (1) nízké zakončení, (2) střední zakončení, (3) vysoké zakončení. Z celkového počtu 100 mužů má 23 mužů (23,00 %) nízké zakončení, 33 mužů (33,00 %) střední zakončení a 44 mužů (44,00 %) vysoké zakončení tří hlavních dlaňových linií. Z celkového počtu 100 mužů má 23 (23,00 %) mužů nízké, 56 mužů (56,00 %) střední nebo nízké a 100 mužů (100 %) vysoké, střední nebo nízké zakončení tří hlavních dlaňových linií.

Sloupkový diagram vykreslíme pomocí funkce barplot(). V diagramu nastavíme jednotkovou šířku sloupců (width = 1), ekvidistantní mezeru mezi dvěma sousedícími sloupci (space = 0.2), barvu výplně sloupců (col = "khaki1"), barvu obrysu sloupců (border = "orange4"), hustotu šrafování výplně (density = 80), rozsah osy y od nuly do šedesáti (ylim = c(0, 60)), otočení popisků měřítka osy y o 90° (las = 1), popisek osy x (xlab), popisek osy y (ylab) a názvy variant znaku X (names). Rámeček okolo grafu doplníme příkazem box() s argumentem bty = "o". Popisky s údaji o absolutních a relativních četnostech všech tří variant znaku X vypíšeme příkazem text(). Vstupními argumenty příkazu budou vektor x-ových souřadnic popisků (argument x), vektor y-ových souřadnic popisků (argument y) a text každého popisku (argument labels).

```
nj <- vr$nj; pj <- vr$pj
barplot(nj, width = 1, space = 0.2, col = "khaki1", border = "orange4", density = 80,
    ylim = c(0, 60), las = 1, xlab = "zakončení tří hlavních dlaňových linií", ylab =
    "absolutní četnost", names = c("nízké", "střední", "vysoké"))
box(bty = "o")
text(x = seq(from = 0.7, to = 5, by = 1.2), y = nj + 5, labels = paste(nj, " (",pj *
    100, " %)", sep = ""))
```

Polygon četností vykreslíme příkazem plot() s argumentem type = "o". V polygonu nastavíme rozsah osy x (xlim = c(0.8, 3.2)), rozsah osy y (ylim = c(0, 60)), potlačení vykreslení os x a y (axes = F), jednotkovou šířku čar (lwd = 1), spojitý styl každé čáry (lty = 1), typ bodu ve tvaru kruhu s obrysem a výplní (pch = 21), barvu čar a obrysu bodů (col = "orange4"), barvu výplně bodů (bg = "khaki1"), popisek osy x (xlab) a popisek osy y (ylab). Názvy variant znaku X doplníme do grafu samostatným vykreslením osy x pomocí příkazu axis(). Zvolenými vstupními argumenty funkce definujeme vykreslení osy x (side = 1) s popisky na pozicích s x-ovými souřadnicemi 1, 2 a 3 (at = 1:3), přičemž texty popisků jsou specifikovány argumentem labels. Funkcí axis() s argumenty side = 2 a las = 1 potom vykreslíme osu y. Sloupkový diagram a polygon četností jsou zobrazeny na obrázku 1.1.

```
plot(nj, type = "o", xlim = c(0.8, 3.2), ylim = c(0, 60), axes = F, lwd = 1, lty = 1,
    pch = 21, col = "orange4", bg = "khaki1", xlab = "zakončení tří hlavních dlaňových
    linií", ylab = "absolutní četnost")
box(bty = "o")
axis(side = 1, at = 1:3, labels = c("nízké", "střední", "vysoké"))
axis(side = 2, las = 1)
```



Obrázek 1.1: Sloupkový diagram (vlevo); polygon četností (vpravo) pro zakončení tří hlavních dlaňových linií u mužů

Graf četnostní funkce $p(x)$ vykreslíme pomocí funkce plot() s argumentem type = "h". Do grafu doplníme body ve výšce odpovídající hodnotám četnostní funkce $p(x)$ příkazem points().

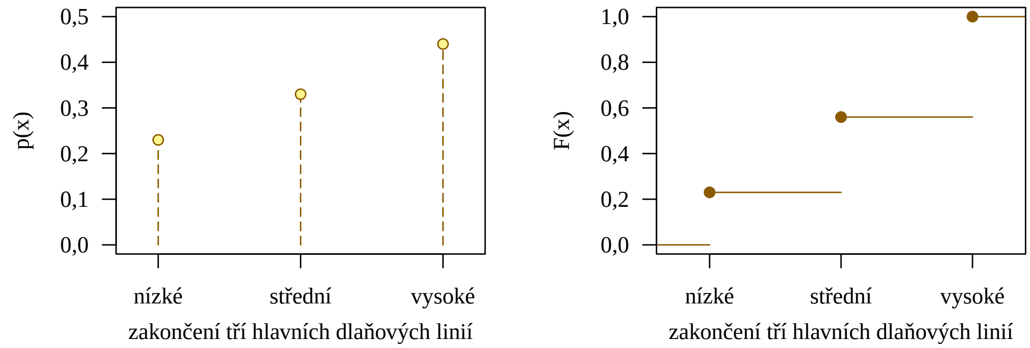
```
plot(pj, type = "h", lty = 5, xlim = c(0.8, 3.2), ylim = c(0, 0.50), axes = F, col =
    "orange4", xlab = "zakončení tří hlavních dlaňových linií", ylab = "p(x)")
points(x = 1:3, y = pj, pch = 21, col = "orange4", bg = "khaki1")
box(bty = "o")
axis(side = 1, at = 1:3, labels = c("nízké", "střední", "vysoké"))
axis(side = 2, las = 1)
```

Graf empirické distribuční funkce $F(x)$ vykreslíme pomocí funkce plot() aplikované na výstup funkce ecdf(), která vrací hodnoty empirické distribuční funkce $F(x)$. V tomto případě máme k dispozici speciální argumenty funkce plot(), které lze použít pouze tehdy, když aplikujeme funkci plot() na výstup funkce ecdf(). Jedním z nich je argument col.01line určující barvu čárkovaných horizontálních referenčních čar procházejících body $F(x) = 0$ a $F(x) = 1$. Zvolením hodnoty argumentu col.01line = NA naopak vykreslení těchto čar potlačíme. Nastavením argumentu main = "" také potlačíme vypsání nežádoucího nadpisu grafu. Poznamenejme, že argument main nepatří do množiny speciálních argumentů, jejichž použití je podmíněno funkcí ecdf(). Graf četnostní funkce $p(x)$ a graf empirické distribuční funkce $F(x)$ jsou zobrazeny na obrázku 1.2.

```
plot(ecdf(palmar.M), xlim = c(0.7, 3.3), axes = F, col.01line = NA, col = "orange4",
    xlab = "zakončení tří hlavních dlaňových linií", ylab = "F(x)", main = "")
box(bty = "o")
axis(side = 1, at = 1:3, labels = c("nízké", "střední", "vysoké"))
axis(side = 2, las = 1)
```

★

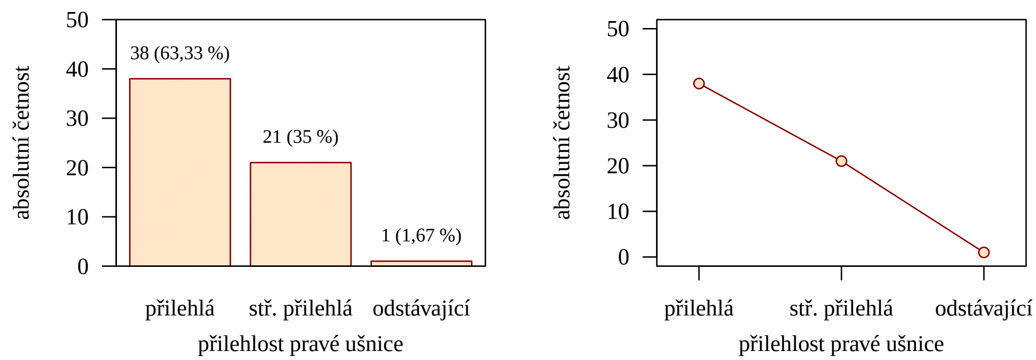
Příklad 1.2. Neřešený příklad
Načtěte datový soubor 23-multinom-earlobe.txt obsahující údaje o přilehlosti ušnic (1 – přilehlá; 2 – středně přilehlá; 3 – odstávající) z pravé a z levé strany u mužů a žen. Za předpokladu, že znak X popisuje přilehlost ušnice z pravé strany u žen, (a) popište jednotlivé varianty znaku X; (b) vytvořte variační řadu; (c) nakreslete sloupkový diagram a polygon četností; (d) nakreslete graf četnostní funkce $p(x)$ a graf empirické distribuční funkce $F(x)$.
Výsledky: (a) tři varianty znaku X: (1) přilehlá ušnice, (2) středně přilehlá ušnice, (3) odstávající ušnice; (b) variační řada viz tabulka 1.2; (c) sloupkový diagram a polygon četností viz obrázek 1.3; (d) graf četnostní funkce $p(x)$ a graf empirické distribuční funkce $F(x)$ viz obrázek 1.4. ★



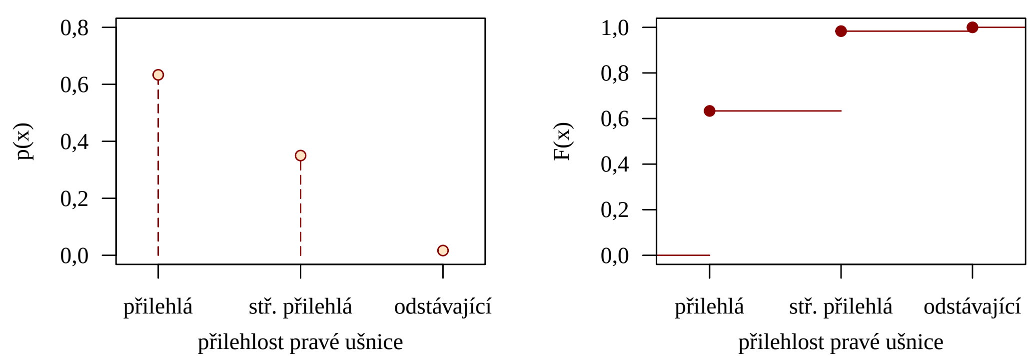
Obrázek 1.2: Graf četnostní funkce $p(x)$ (vlevo); graf empirické distribuční funkce $F(x)$ (vpravo) pro zakončení tří hlavních dlaňových linií u mužů

Tabulka 1.2: Variační řada pro přilehlost ušnice z pravé strany u žen

	n_j	p_j	N_j	F_j
přilehlá	38	0,63	38	0,63
středně přilehlá	21	0,35	59	0,98
odstávající	1	0,02	60	1,00



Obrázek 1.3: Sloupkový diagram (vlevo); polygon četností (vpravo) pro přilehlost ušnice z pravé strany u žen



Obrázek 1.4: Graf četnostní funkce (vlevo); graf empirické distribuční funkce (vpravo) pro přilehlost ušnice z pravé strany u žen

1.3 Dvourozměrné bodové rozložení četností

Máme dvourozměrný datový soubor $(x_1, y_1)^T, \ldots, (x_n, y_n)^T$, kde znak X má r variant a znak Y má s variant. Pak definujeme:

- n_{jk} = počet těch objektů ve výběrovém souboru, pro něž $X = x_{[j]}$ a současně $Y = y_{[k]}$ – simultánní absolutní četnost dvojice $(x_{[j]}, y_{[k]})^T$ ve výběrovém souboru,
- $p_{jk} = \frac{n_{jk}}{n}$ – simultánní relativní četnost dvojice $(x_{[j]}, y_{[k]})^T$ ve výběrovém souboru,
- $n_{j.} = N(X = x_{[j]}) = n_{j1} + \cdots + n_{js}$ – marginální absolutní četnost varianty $x_{[j]}$,
- $p_{j.} = \frac{n_{j.}}{n} = p_{j1} + \cdots + p_{js}$ – marginální relativní četnost varianty $x_{[j]}$,
- $n_{.k} = N(Y = y_{[k]}) = n_{1k} + \cdots + n_{rk}$ – marginální absolutní četnost varianty $y_{[k]}$,
- $p_{.k} = \frac{n_{.k}}{n} = p_{1k} + \cdots + p_{rk}$ – marginální relativní četnost varianty $y_{[k]}$.

Pro dvourozměrný datový soubor vytváříme kontingenční tabulku simultánních a marginálních absolutních četností (viz tabulka 1.3 vlevo), resp. kontingenční tabulku simultánních a marginálních relativních četností (viz tabulka 1.3 vpravo). Simultánní četnosti se v tabulkách nacházejí v bloku uprostřed, marginální četnosti se nacházejí v dolním a pravém bloku.

Tabulka 1.3: Kontingenční tabulka (a) simultánních a marginálních absolutních četností (vlevo); (b) simultánních a marginálních relativních četností (vpravo)

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

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

Pomocí simultánních relativních četností zavedeme simultánní četnostní funkci:

$$p(x, y) = \begin{cases} p_{jk} \text{ pro } x = x_{[j]}, y = y_{[k]}, j = 1, \ldots, r; k = 1, \ldots, s, \\ 0 \text{ jinak.} \end{cases}$$

Pomocí marginálních relativních četností zavedeme marginální četnostní funkce $p_1(x)$, $p_2(y)$:

$$p_1(x) = \begin{cases} p_{j.} \text{ pro } x = x_{[j]}, j = 1, \ldots, r, \\ 0 \text{ jinak,} \end{cases}$$

$$p_2(y) = \begin{cases} p_{.k} \text{ pro } y = y_{[k]}, k = 1, \ldots, s, \\ 0 \text{ jinak.} \end{cases}$$

Mezi simultánní četnostní funkcí a marginálními četnostními funkcemi platí vztahy: $p_1(x) = \sum_{y=-\infty}^{\infty} p(x, y)$, $p_2(y) = \sum_{x=-\infty}^{\infty} p(x, y)$.

Znaky X a Y jsou v daném výběrovém souboru četnostně nezávislé, právě když pro $\forall j = 1, \ldots, r$, $\forall k = 1, \ldots, s$ platí: $p_{jk} = p_{j.} p_{.k}$ neboli pro $\forall (x, y)^T \in \mathbb{R}^2$: $p(x, y) = p_1(x) p_2(y)$.

Sloupcově podmíněná relativní četnost varianty $x_{[j]}$ za předpokladu $y_{[k]}$: $p_{j(k)} = \frac{n_{jk}}{n_{.k}}$. Řádkově podmíněná relativní četnost varianty $y_{[k]}$ za předpokladu $x_{[j]}$: $p_{(j)k} = \frac{n_{jk}}{n_{j.}}$. Podmíněné relativní četnosti potom zapisujeme do kontingenční tabulky řádkově podmíněných relativních četností (viz tabulka 1.4 vlevo), resp. do kontingenční tabulky sloupcově podmíněných relativních četností (viz tabulka 1.4 vpravo).

Tabulka 1.4: Kontingenční tabulka (a) řádkově podmíněných relativních četností (vlevo); (b) sloupcově podmíněných relativních četností (vpravo)

X	Y			
	$y_{[1]}$	$\dots$	$y_{[s]}$	1
$x_{[1]}$	$p_{(1)1}$	$\dots$	$p_{(1)s}$	1
$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
$x_{[r]}$	$p_{(r)1}$	$\dots$	$p_{(r)s}$	1

X	Y		
	$y_{[1]}$	$\dots$	$y_{[s]}$
$x_{[1]}$	$p_{1(1)}$	$\dots$	$p_{1(s)}$
$\vdots$	$\vdots$	$\dots$	$\vdots$
$x_{[r]}$	$p_{r(1)}$	$\dots$	$p_{r(s)}$
$p_{.k}$	1	$\dots$	1

Dvourozměrné rozložení četností graficky znázorníme například grafem simultánní četnostní funkce $p(x, y)$ nebo grafy marginálních četnostních funkcí $p_1(x)$ a $p_2(y)$.

Příklad 1.3. Řešený příklad

Načtěte datový soubor 22-multinom-palmar-lines.txt. Za předpokladu, že znak X popisuje odstín barvy vlasů a znak Y popisuje zakončení tří hlavních dlaňových linií u mužů, (a) popište jednotlivé varianty znaku X, resp. znaku Y; (b) vytvořte kontingenční tabulku simultánních absolutních, resp. relativních četností; (c) vytvořte kontingenční tabulku řádkově, resp. sloupcově podmíněných relativních četností; (d) ověřte četnostní nezávislost znaků X a Y; (e) nakreslete graf simultánní četnostní funkce $p(x, y)$; (f) nakreslete graf marginální četnostní funkce znaku X, tj. $p_1(x)$, resp. znaku Y, tj. $p_2(y)$.

Řešení příkladu 1.3

Datový soubor načteme příkazem read.delim(). Pomocí operátoru [ ] vybereme z datového souboru pouze sloupce týkající se mužů. Výsledná tabulka je rovnou kontingenční tabulkou simultánních absolutních četností. Příkazem row.names() akorát přejmenujeme její řádky a příkazem names() přejmenujeme její sloupce. Kontingenční tabulku simultánních relativních četností získáme příkazem prop.table().

```
data <- read.delim("22-multinom-palmar-lines.txt", sep = "\t", row.names = 1)
```

```
    Lo.m Mi.m Hi.m Lo.f Mi.f Hi.f
LiH    4    6    6    6    6    4
MH     7   15   20   10   10   18
DaH   12   12   18   12   22   12
```

```
KT.abs <- data[, 1:3]
row.names(KT.abs) <- c("svetly", "stredni", "tmavy")
names(KT.abs) <- c("nizke", "stredni", "vysoke")
```

```
        nizke stredni vysoke
svetly      4       6      6
stredni     7      15     20
tmavy      12      12     18
```

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

```
        nizke stredni vysoke
svetly   0,04    0,06   0,06
stredni  0,07    0,15   0,20
tmavy    0,12    0,12   0,18
```

Znak X popisující odstín barvy vlasů u mužů má tři varianty: (1) světlý odstín, (2) střední odstín, (3) tmavý odstín. Znak Y popisující zakončení tří hlavních dlaňových linií u mužů má rovněž tři varianty: (1) nízké zakončení, (2) střední zakončení, (3) vysoké zakončení. V datovém souboru se vyskytují čtyři muži (4 %) se světlým odstínem barvy vlasů a s nízkým zakončením tří hlavních dlaňových linií, šest mužů (6 %) se světlým odstínem barvy vlasů a se středním zakončením tří hlavních dlaňových linií, sedm mužů (7 %) středním odstínem barvy vlasů a s nízkým zakončením tří hlavních dlaňových linií, apod.

Kontingenční tabulku řádkově podmíněných relativních četností získáme příkazem prop.table() s argumentem margin

= 1. Prvním vstupním argumentem příkazu bude kontingenční tabulka simultánních absolutních četností ve tvaru matice. Proměnnou typu data.frame převedeme na proměnnou typu matrix příkazem as.matrix().

```
KTP.r <- prop.table(as.matrix(KT.abs), margin = 1)
```

```
        nizke stredni vysoke
svetly  0,2500  0,3750 0,3750
stredni 0,1667  0,3571 0,4762
tmavy   0,2857  0,2857 0,4286
```

Ze všech mužů, kteří měli světlý odstín barvy vlasů, mělo 25,00 % nízké, 37,50 % střední a 37,50 % vysoké zakončení tří hlavních dlaňových linií. Ze všech mužů, kteří měli střední odstín barvy vlasů, mělo 16,67 % nízké, 35,71 % střední a 47,62 % vysoké zakončení tří hlavních dlaňových linií. Ze všech mužů, kteří měli tmavý odstín barvy vlasů, mělo 28,57 % nízké, 28,57 % střední a 42,86 % vysoké zakončení tří hlavních dlaňových linií.

Kontingenční tabulku sloupcově podmíněných relativních četností získáme příkazem prop.table() s argumentem margin = 2.

```
KTP.s <- prop.table(as.matrix(KT.abs), margin = 2)
```

```
        nizke stredni vysoke
svetly  0,1739  0,1818 0,1364
stredni 0,3043  0,4545 0,4545
tmavy   0,5217  0,3636 0,4091
```

Ze všech mužů, kteří měli nízké zakončení tří hlavních dlaňových linií, mělo 17,39 % světlý, 30,43 % střední a 52,17 % tmavý odstín barvy vlasů. Ze všech mužů, kteří měli střední zakončení tří hlavních dlaňových linií, mělo 18,18 % světlý, 45,45 % střední a 36,36 % tmavý odstín barvy vlasů. Ze všech mužů, kteří měli vysoké zakončení tří hlavních dlaňových linií, mělo 13,64 % světlý, 45,45 % střední a 40,91 % tmavý odstín barvy vlasů.

Četnostní nezávislost ověříme porovnáním simultánních relativních četností p_{jk} (viz kontingenční tabulka relativních četností KT.rel) se součiny marginálních relativních četností $p_{j.}p_{.k}$. Tabulka součinů je jedním z výstupů funkce chisq.test() pojmenovaném expected.

```
pj.p.k <- chisq.test(KT.rel)$expected
```

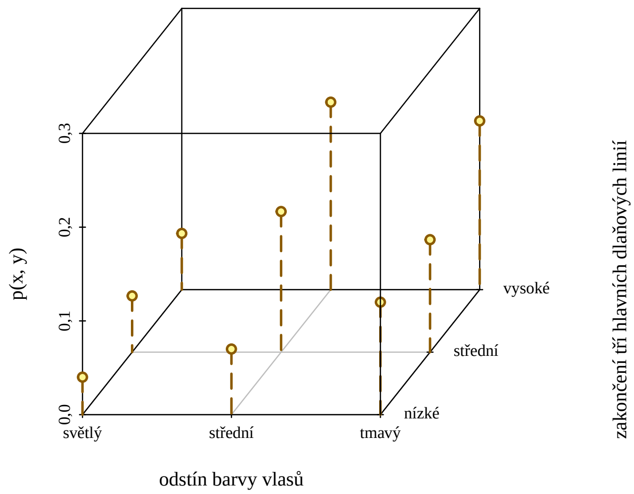
```
        nizke stredni vysoke
svetly  0,0368  0,0528 0,0704
stredni 0,0966  0,1386 0,1848
tmavy   0,0966  0,1386 0,1848
```

Na základě porovnání simultánních relativních četností p_{jk} se součiny marginálních relativních četností $p_{j.}p_{.k}$ docházíme k závěru, že mezi znaky X a Y neexistuje četnostní nezávislost. Například $p_{11} \neq p_{1.}p_{.1}$ $(0,04 \neq 0,0368)$, $p_{23} \neq p_{2.}p_{.3}$ $(0,20 \neq 0,1848)$, $p_{31} \neq p_{3.}p_{.1}$ $(0,12 \neq 0,0966)$, apod.

Graf simultánní četnostní funkce $p(x,y)$ vykreslíme příkazem scatterplot3d() z knihovny scatterplot3d. Vstupními argumenty příkazu je vektor x-ových souřadnic x, vektor y-ových souřadnic y a vektor simultánních relativních četností z příslušných každé dvojici souřadnic $[x_j, y_k]$, $j = 1,2,3$, $k = 1,2,3$. V grafu nastavíme vykreslení horizontálních čar zakončených body (type = "h"), úhel 60° mezi osami x a y (angle = 60), typ bodu ve tvaru kruhu s obrysem a výplní (pch = 21), silnější šířku čar (lwd = 2), delší čárkovaný typ čar (lty.hplot = 5), barvu čar a obrysu bodů (color = "orange4"), barvu výplně bodů (bg = "khaki1"), vzdálenost mezi pravou hranou krychle a popiskem osy y (y.margin.add = 0.4), dimenzi podstavy, tj. celkem 2×2 polí (lab = c(2, 2)), popisek osy x (xlab), popisek osy y (ylab), popisek osy z (zlab), popisky variant znaku X (x.ticklabs) a popisky variant znaku y (y.ticklabs). Graf simultánní četnostní funkce $p(x,y)$ je zobrazen na obrázku 1.5.

```
x <- rep(1:3, 3)
y <- rep(1:3, rep(3, 3))
z <- c(as.matrix(KT.rel))
```

```
scatterplot3d::scatterplot3d(x, y, z, type = "h", angle = 60, pch = 21, lwd = 2,
    lty.hplot = 5, color = "orange4", bg = "khaki1", y.margin.add = 0.4, lab = c(2, 2),
    xlab = "odstín barvy vlasů", ylab = "zakončení tří hlavních dlaňových linií", zlab
    = "p(x, y)", x.ticklabs = c("světlý", "střední", "tmavý"), y.ticklabs = c("nízké",
    "střední", "vysoké"))
```



Obrázek 1.5: Graf simultánní četnostní funkce $p(x, y)$ pro odstín barvy vlasů a zakončení tří hlavních dlaňových linií u mužů

K vykreslení grafů marginálních četnostních funkcí $p_1(x)$ a $p_2(y)$ potřebujeme znát hodnoty marginálních relativních četností. Ty vypočítáme pomocí funkce apply() aplikované na řádky (MARGIN = 1), resp. sloupce (MARGIN = 2) kontingenční tabulky simultánních relativních četností. Samotné grafy marginálních četnostních funkcí vykreslíme příkazem plot() s argumentem type = "h". Popisek osy y obsahující dolní index zadefinujeme pomocí funkce expression(). Do grafu dokreslíme body příkazem points(), okolo grafu vykreslíme rámeček příkazem box() a osy x a y dokreslíme zvlášť příkazem axis(). Grafy marginálních četnostních funkcí $p_1(x)$ a $p_2(y)$ jsou zobrazeny na obrázku 1.6.

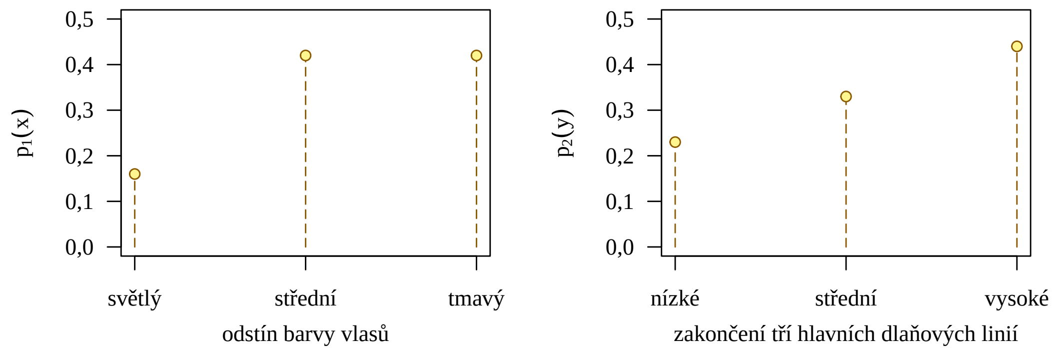
```
pj. <- apply(KT.rel, MARGIN = 1, FUN = sum)
p.k <- apply(KT.rel, MARGIN = 2, FUN = sum)
plot(pj., type = "h", lty = 5, ylim = c(0, 0.5), axes = F, col = "orange4", xlab =
    "odstín barvy vlasů", ylab = expression(p[1](x)))
points(x = 1:3, y = pj., pch = 21, col = "orange4", bg = "khaki1")
box(bty = "o")
axis(side = 1, at = 1:3, labels = c("světlý", "střední", "tmavý"))
axis(side = 2, las = 1)

plot(p.k, type = "h", lty = 5, ylim = c(0, 0.5), axes = F, col = "orange4", xlab =
    "zakončení tří hlavních dlaňových linií", ylab = expression(p[2](y)))
points(x = 1:3, y = p.k, pch = 21, col = "orange4", bg = "khaki1")
box(bty = "o")
axis(side = 1, at = 1:3, labels = c("nízké", "střední", "vysoké"))
axis(side = 2, las = 1)
```

★

Příklad 1.4. Neřešený příklad

Načtěte datový soubor 20-more-samples-probabilities-pubis.txt obsahující údaje o původu žen (european – evropský; african – africký; inuits – inuitský) a o míře změn kostního reliéfu na vnitřní straně stydké kosti v blízkosti stydké spony (absence – nepřítomnost změn; trace.to small – stopové až malé změny; moderate.to.large – střední až výrazné



Obrázek 1.6: Graf marginální četnostní funkce $p_1(x)$ pro odstín barvy vlasů u mužů (vlevo); graf marginální četnostní funkce $p_2(y)$ pro zakončení tří hlavních dlaňových linií u mužů (vpravo)

změny). Za předpokladu, že znak X popisuje původ žen a znak Y popisuje míru změny kostního reliéfu u těchto žen, (a) popište jednotlivé varianty znaku X, resp. znaku Y; (b) vytvořte kontingenční tabulku simultánních absolutních, resp. relativních četností; (c) vytvořte kontingenční tabulku řádkově, resp. sloupcově podmíněných relativních četností; (d) ověřte četnostní nezávislost znaků X a Y; (e) nakreslete graf simultánní četnostní funkce $p(x, y)$; (f) nakreslete graf marginální četnostní funkce znaku X, tj. $p_1(x)$, resp. znaku Y, tj. $p_2(y)$.
Výsledky: (a) tři varianty znaku X: (1) evropský původ, (2) africký původ, (3) inuitský původ; tři varianty znaku Y: (1) nepřítomnost změn; (2) stopové až malé změny; (3) střední až výrazné změny; (b) kontingenční tabulka simultánních absolutních, resp. relativních četností viz tabulka 1.5, resp. tabulka 1.6; (c) kontingenční tabulka řádkově, resp. sloupcově podmíněných relativních četností viz tabulka 1.7, resp. tabulka 1.8; (d) mezi znaky X a Y neexistuje četnostní nezávislost viz porovnání tabulky 1.6 s tabulkou 1.9; (e) graf simultánní četnostní funkce $p(x, y)$ viz obrázek 1.7; (f) graf marginální četnostní funkce znaku X, tj. $p_1(x)$, resp. znaku Y, tj. $p_2(y)$ viz obrázek 1.8.

Tabulka 1.5: Kontingenční tabulka simultánních absolutních četností n_{jk} pro původ a míru změny kostního reliéfu u žen

	nepřítomnost	stopové až malé	střední až výrazné
evropský původ	30	20	10
africký původ	56	37	17
inuitský původ	16	6	13

Tabulka 1.6: Kontingenční tabulka simultánních relativních četností p_{jk} pro původ a míru změny kostního reliéfu u žen

	nepřítomnost	stopové až malé	střední až výrazné
evropský původ	0,15	0,10	0,05
africký původ	0,27	0,18	0,08
inuitský původ	0,08	0,03	0,06

Tabulka 1.7: Kontingenční tabulka řádkově podmíněných relativních četností $p_{(j)k}$ pro míru změny kostního reliéfu podmíněnou původem žen

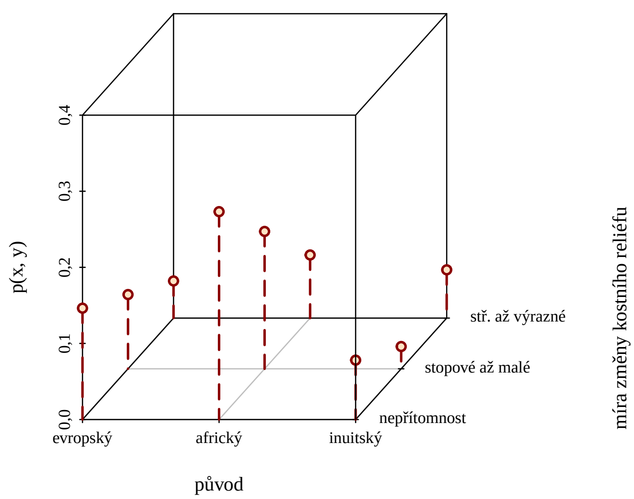
	nepřítomnost	stopové až malé	střední až výrazné
evropský původ	0,50	0,33	0,17
africký původ	0,51	0,34	0,15
inuitský původ	0,46	0,17	0,37

Tabulka 1.8: Kontingenční tabulka sloupcově podmíněných relativních četností $p_{j(k)}$ pro původ podmíněný mírou změny kostního reliéfu u žen

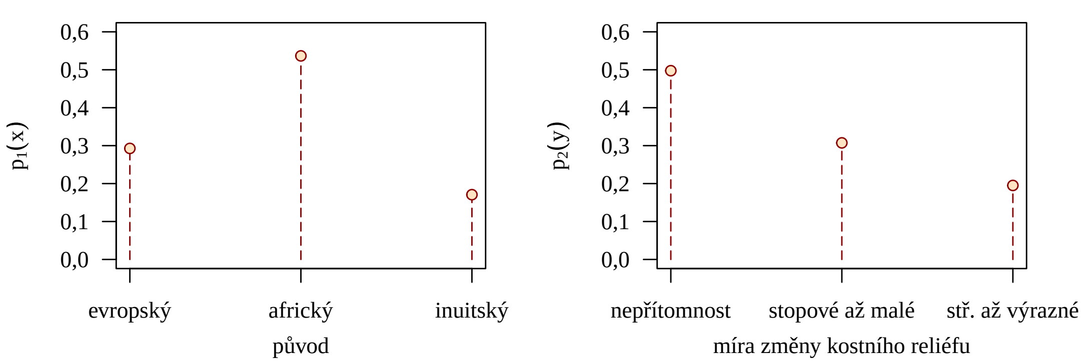
	nepřítomnost	stopové až malé	střední až výrazné
evropský původ	0,29	0,32	0,25
africký původ	0,55	0,59	0,42
inuitský původ	0,16	0,10	0,32

Tabulka 1.9: Kontingenční tabulka součinů marginálních relativních četností $p_{j.}p_{.k}$ pro původ a míru změny kostního reliéfu u žen

	nepřítomnost	stopové až malé	střední až výrazné
evropský původ	0,15	0,09	0,06
africký původ	0,27	0,16	0,10
inuitský původ	0,08	0,05	0,03



Obrázek 1.7: Graf simultánní četnostní funkce $p(x, y)$ pro původ a míru změny kostního reliéfu u žen



Obrázek 1.8: Graf marginální četnostní funkce $p_1(x)$ pro původ žen (vlevo); graf marginální četnostní funkce $p_2(y)$ pro míru změny kostního reliéfu u žen (vpravo)

★