1 Bodové a intervalové rozdělení četností

Příklad 1.1. Načtěte soubor znamky.txt. Znakům X, Y, Z vytvořte návěští (X - známka z matematiky, Y - známka z angličtiny, Z - pohlaví studenta). Popište, co znamenají jednotlivé varianty (u znaků X a Y: 1 - výborně, 2 - chvalitebně, 3 - dobře, 4 - neprospěl, u znaku Z: 0 - žena, 1 - muž).

```
##     matematika   anglictina pohlavi
## 1  chvalitebne  chvalitebne    zena
## 2      vyborne        dobre     muz
## 3 nedostatecne        dobre     muz
## 4      vyborne      vyborne    zena
## 5      vyborne  chvalitebne     muz
## 6 nedostatecne nedostatecne     muz
```

Příklad 1.2. Vytvořte

a) variační řadu známek z matematiky a známek z angličtiny;

b) sloupkový diagram absolutních četností znaků X=Matematika a Y=Angličtina;

c) polygon absolutních četností znaků X=Matematika a Y=Angličtina.

a) Variační řada známek z matematiky

```
##              nj Nj   pj   Fj
## vyborne       7  7 0.35 0.35
## chvalitebne   3 10 0.15 0.50
## dobre         2 12 0.10 0.60
## nedostatecne  8 20 0.40 1.00
```

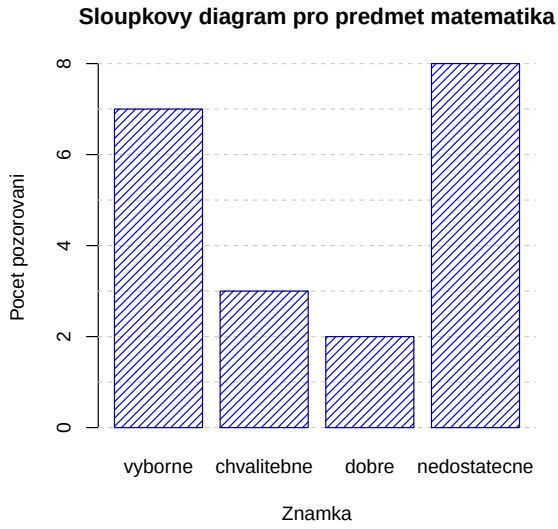
Variační řadu můžeme také získat použitím funkce variacni_rada(X, nazvy), která je naprogramovaná ve skriptu AS-funkce.

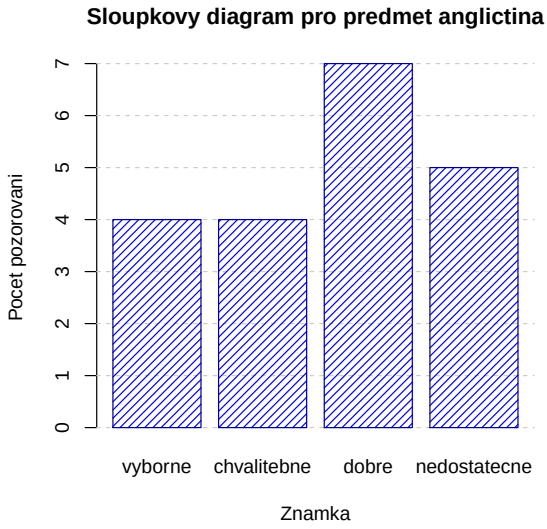
```
##              nj Nj   pj   Fj
## vyborne       7  7 0.35 0.35
## chvalitebne   3 10 0.15 0.50
## dobre         2 12 0.10 0.60
## nedostatecne  8 20 0.40 1.00
```

Variační řada známek z angličtiny

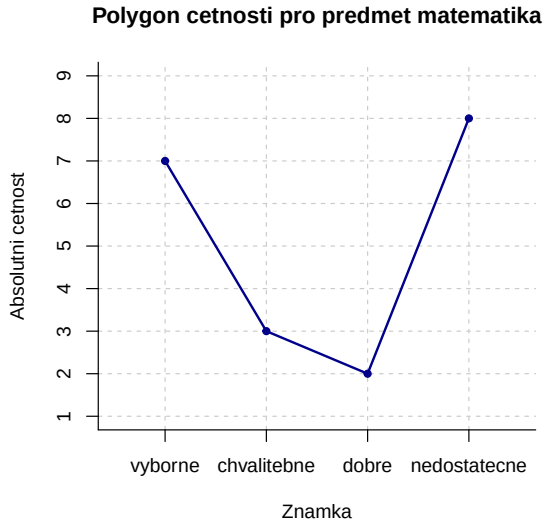
```
##              nj Nj   pj   Fj
## vyborne       4  4 0.20 0.20
## chvalitebne   4  8 0.20 0.40
## dobre         7 15 0.35 0.75
## nedostatecne  5 20 0.25 1.00
```

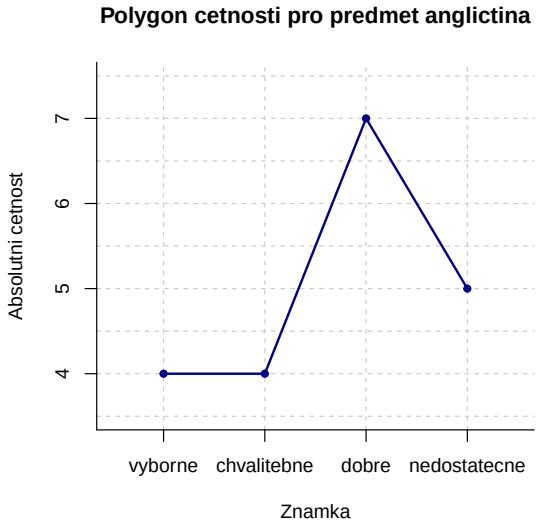
b) Sloupkový diagram absolutních četností znaků X=Matematika a Y=Angličtina





c) Polygon četností





Příklad 1.3. Vytvořte variační řady známek z matematiky a angličtiny pouze

a) pro ženy,

b) pro muže.

a) Variační řada známek z matematiky pro ženy

```
##              nj Nj  pj  Fj
## vyborne       5  5 0.5 0.5
## chvalitebne   2  7 0.2 0.7
## dobre         1  8 0.1 0.8
## nedostatecne  2 10 0.2 1.0
```

Variační řada známek z angličtiny pro ženy

```
##              nj Nj  pj  Fj
## vyborne       4  4 0.4 0.4
## chvalitebne   2  6 0.2 0.6
## dobre         1  7 0.1 0.7
## nedostatecne  3 10 0.3 1.0
```

b) Variační řada známek z matematiky pro muže

```
##              nj Nj  pj  Fj
## vyborne       2  2 0.2 0.2
## chvalitebne   1  3 0.1 0.3
## dobre         1  4 0.1 0.4
## nedostatecne  6 10 0.6 1.0
```

Variační řada známek z angličtiny pro muže

```
##              nj Nj  pj  Fj
## vyborne       0  0 0.0 0.0
## chvalitebne   2  2 0.2 0.2
## dobre         6  8 0.6 0.8
## nedostatecne  2 10 0.2 1.0
```

Příklad 1.4. Nadále budeme pracovat s celým datovým souborem. Vytvoříme kontingenční tabulku simultánních absolutních četností znaků X a Y.

```
##              vyborne chvalitebne dobre nedostatecne suma
## vyborne            4           1     2            0    7
## chvalitebne        0           2     1            0    3
## dobre              0           0     1            1    2
## nedostatecne       0           1     3            4    8
## suma               4           4     7            5   20
```

Vidíme, že ve výběrovém souboru byli 4 studenti, kteří měli z obou předmětů "výborně", jeden student, který měl z matematiky "výborně"a z angličtiny "chvalitebně"atd. až 4 studenti, kteří z obou předmětů neprospěli.

Příklad 1.5. Vytvořte kontingenční tabulku řádkově a sloupcově podmíněných relativních četností znaků X=Matematika a Y=Angličtina.

```
##               anglictina
## matematika     vyborne chvalitebne dobre nedostatecne
##   vyborne        0.571       0.143 0.286        0.000
##   chvalitebne    0.000       0.667 0.333        0.000
##   dobre          0.000       0.000 0.500        0.500
##   nedostatecne   0.000       0.125 0.375        0.500
```

Interpretace např. 2. sloupce ve 4. řádku: V souboru bylo 8 studentů, kteří neprospěli z matematiky. Mezi nimi byl jeden, který měl velmi dobře z angličtiny, což představuje $1/8 = 12.5\%$.

```
##               anglictina
## matematika     vyborne chvalitebne dobre nedostatecne
##   vyborne        1.000       0.250 0.286        0.000
##   chvalitebne    0.000       0.500 0.143        0.000
##   dobre          0.000       0.000 0.143        0.200
##   nedostatecne   0.000       0.250 0.429        0.800
```

Interpretace např. 4. řádku ve 2. sloupci: V souboru byli 4 studenti, kteří měli velmi dobře z angličtiny. Mezi nimi byl jeden, který neprospěl z matematiky, což představuje $1/4 = 25\%$.

Intervalové rozdělení četností

Práci s intervalovým rozložením četností si ukážeme na datovém souboru lebky.txt.

Popis datového souboru: Máme k dispozici údaje o rozměrech lebek staroegyptské populace. Jedná se o 216 mužů a 109 žen. Znak X . . . největší délka mozkovny v mm (tj. přímá vzdálenost kraniometrických bodů glabella a opisthocranion) Znak Y . . . největší šířka mozkovny v mm (tj. přímá vzdálenost kraniometrických bodů euryon dx a euryon sin) Znak Z . . . pohlaví osoby (1–muž, 0–žena)

Příklad 1.6. Načtěte soubor lebky.txt. Podle Sturgersova pravidla najděte optimální počet třídicích intervalů pro znaky X a Y a vhodně stanovte meze třídicích intervalů, a to zvlášť pro muže a zvlášť pro ženy.

```
##   delka sirka pohlavi
## 1   188   145     muž
## 2   172   139     muž
## 3   176   138     muž
## 4   184   128     muž
## 5   183   139     muž
## 6   177   143     muž
```

Muži

```
## [1] 9
```

Protože mužů je 216, podle Sturgersova pravidla je optimální počet třídicích intervalů 9. Musíme zjistit minimum a maximum, abychom vhodně stanovili meze třídicích intervalů.

Délka lebky

```
## [1] 164 199
## [1] 35
## [1] 4
```

Pro znak X = Délka lebky je minimum 164 a maximum 199, rozsah těchto hodnot je 35 a ideální délka jednoho třídicího intervalu vyšla jako $\frac{199-164}{9} \approx 4$. Jeví se vhodné dolní mez prvního třídicího intervalu zvolit 163, horní mez posledního třídicího intervalu 199. Celkem třídicí intervaly pro znak X budou: $\langle 163, 167\rangle$, $\langle 167, 171\rangle$, . . . , $\langle 195, 199\rangle$.

Šířka lebky

```
## [1] 124 149
## [1] 25
## [1] 3
```

Pro znak Y = šířka lebky je minimum 124 a maximum 149, rozsah těchto hodnot je 25 a ideální délka jednoho třídicího intervalu vyšla jako $\frac{149-124}{9} \approx 3$. Jeví se vhodné dolní mez prvního třídicího intervalu zvolit 123, horní mez posledního třídicího intervalu 150. Celkem třídicí intervaly pro znak X budou: $\langle 123, 126\rangle$, $\langle 126, 129\rangle$, . . . , $\langle 147, 150\rangle$.

Ženy

```
## [1] 8
```

Protože žen je 109, podle Sturgersova pravidla je optimální počet třídicích intervalů 8. Postup je analogický jako u mužů.

Délka lebky

```
## [1] 157 188
## [1] 31
## [1] 4
```

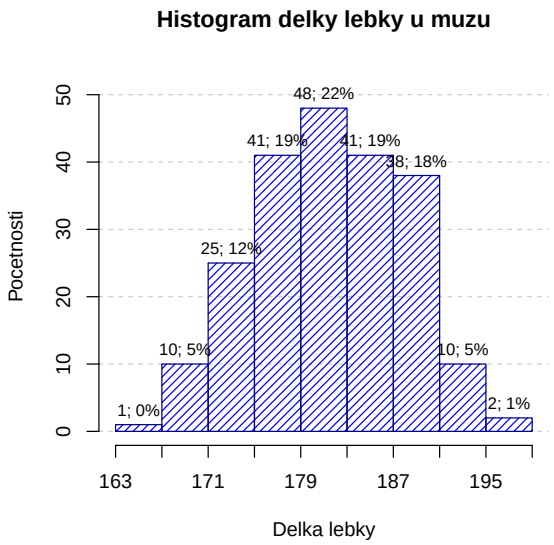
Pro znak X = Délka lebky je minimum 157 a maximum 188, rozsah těchto hodnot je 31 a ideální délka jednoho třídicího intervalu vyšla jako $\frac{188-157}{8} \approx 4$. Jeví se vhodné dolní mez prvního třídicího intervalu zvolit 156, horní mez posledního třídicího intervalu 188. Celkem třídicí intervaly pro znak X budou: $\langle 156, 160\rangle$, $\langle 160, 164\rangle$, ..., $\langle 184, 188\rangle$.

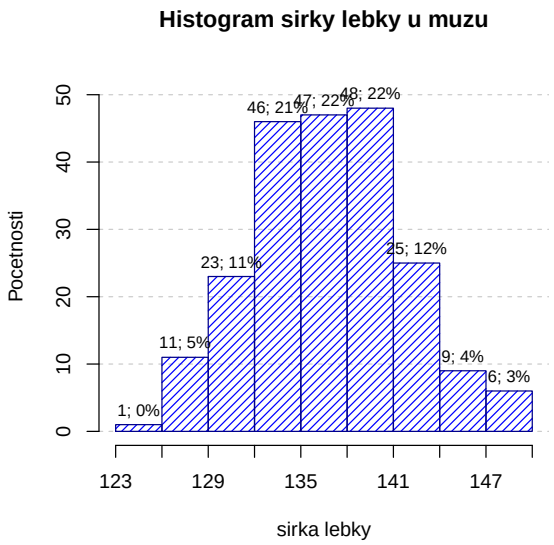
Šířka lebky

```
## [1] 118 146
## [1] 28
## [1] 4
```

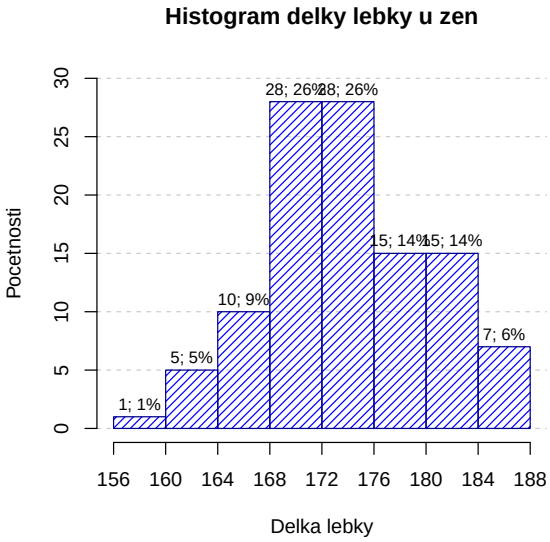
Pro znak Y = šířka lebky je minimum 118 a maximum 146, rozsah těchto hodnot je 28 a ideální délka jednoho třídicího intervalu vyšla jako $\frac{146-118}{8} \approx 4$. Jeví se vhodné dolní mez prvního třídicího intervalu zvolit 116, horní mez posledního třídicího intervalu 148. Celkem třídicí intervaly pro znak X budou: $\langle 116, 120\rangle$, $\langle 120, 124\rangle$, ..., $\langle 144, 148\rangle$.

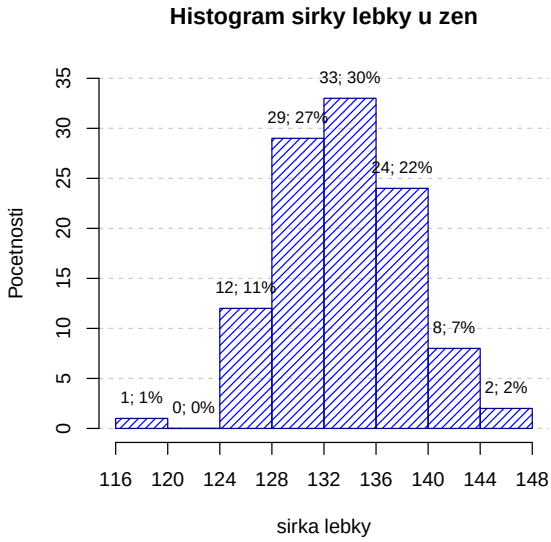
Příklad 1.7. Vytvořte histogram pro X a pro Y (s uvedenými absolutními a relativními četnostmi jednotlivých třídicích intervalů), a to zvlášť pro muže a zvlášť pro ženy.





Pro ženy je postup analogický jako pro muže.


Histogram delky lebky u zen
Pocetnosti
0
5
10
15
20
25
30
1; 1%
5; 5%
10; 9%
28; 26%
28; 26%
15; 14%
15; 14%
7; 6%
156
160
164
168
172
176
180
184
188
Delka lebky


Histogram sirky lebky u zen
Pocetnosti
0
5
10
15
20
25
30
35
1; 1%
0; 0%
12; 11%
29; 27%
33; 30%
24; 22%
8; 7%
2; 2%
116
120
124
128
132
136
140
144
148
sirka lebky