

MASARYKOVA UNIVERZITA  
FAKULTA INFORMATIKY



# MV011 Statistika I

---

## Semestrální projekt

**Martin Jarmar (172981)**

## Obsah

Obsah .....	2
Zadání práce .....	3
Úkol 1.....	3
Úkol 2.....	3
Úkol 3.....	3
Úkol 4.....	3
Úkol 5.....	3
Úkol 6.....	3
Úkol 1.....	4
Řešení .....	4
Úkol 2.....	5
Řešení .....	5
Úkol 3.....	6
Řešení .....	6
Úkol 4.....	8
Řešení .....	8
Úkol 5.....	9
Řešení .....	9
Úkol 6.....	10
Řešení .....	10
Použitá literatura.....	11
Použitý software.....	11
Seznam tabulek.....	11
Seznam grafů.....	11

## Zadání práce

U 96 náhodně vybraných studentů VŠE v Praze byly zjištěny následující údaje:

- Pohlaví (0 – žena, 1 – muž)
- Výška (tělesná výška v cm)
- Hmotnost (tělesná hmotnost v kg)
- Znamka (znamka z matematiky v 1. semestru)

### Úkol 1.

Zjistěte absolutní a relativní četnosti proměnných Pohlaví a Znamka, přičemž pro proměnnou Znamka zjistěte též kumulativní absolutní a relativní četnosti. Pro proměnnou Pohlaví vytvořte sloupkový diagram, pro proměnnou Znamka polygon četností.

### Úkol 2.

Pro proměnné Pohlaví a Znamka sestavte kontingenční tabulky absolutních a relativních četností, sloupcově a řádkově podmíněných relativních četností. Kolik procent žen má z matematiky jedničku? Kolik procent studentů, kteří mají jedničku, jsou muži?

### Úkol 3.

Podle Sturgesova pravidla stanovte optimální počet třídících intervalů pro proměnné Výška a Hmotnost a nakreslete jejich histogramy, a to:

- a) pro celý soubor
- b) pro ženy
- c) pro muže

### Úkol 4.

Vypočítejte minimum, maximum, medián, průměr, směrodatnou odchylku, šikmost a špičatost proměnných Výška a Hmotnost:

- a) pro celý soubor
- b) pro ženy
- c) pro muže

### Úkol 5.

Vypočítejte a interpretujte Pearsonův koeficient korelace proměnných Výška a Hmotnost:

- a) pro celý soubor
- b) pro ženy
- c) pro muže

### Úkol 6.

Najděte rovnici regresní přímky vyjadřující závislost proměnné Hmotnost na proměnné Výška. Jaký je index determinace a co vyjadřuje? Jaká je predikovaná hodnota hmotnosti pro výšku 175 cm? Nalezenou regresní přímku zakreslete do dvourozměrného tečkového diagramu.

## Úkol 1.

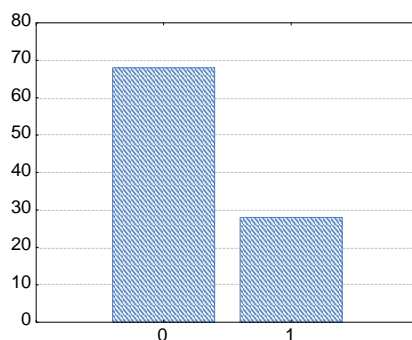
Zjistěte absolutní a relativní četnosti proměnných Pohlaví a Známká, přičemž pro proměnnou Známká zjistěte též kumulativní absolutní a relativní četnosti. Pro proměnnou Pohlaví vytvořte sloupkový diagram, pro proměnnou Známká polygon četností.

## Řešení

Tabulka 1. Tabulka četností proměnné Pohlaví

	Četnost	Rel. četnost
0	68	70,83333
1	28	29,16667

Obrázek 1. Sloupkový diagram proměnné Pohlaví

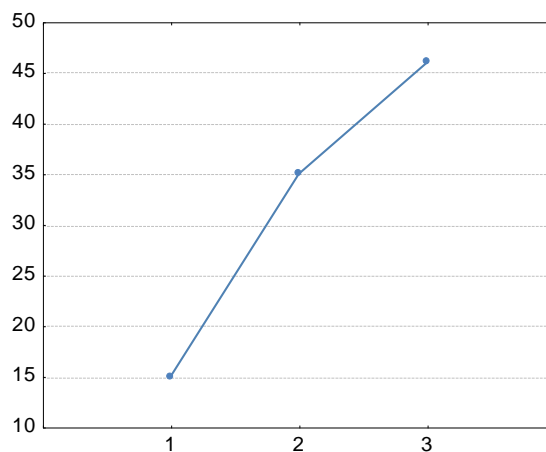


**Komentář:** Z tabulky a sloupkového diagramu je vidět, že zastoupení dívek oproti chlapcům je ve výběrovém souboru více než dvounásobné. Z 96 náhodně vybraných studentů VŠE v Praze je dívek 68 a chlapců 28.

Tabulka 2. Tabulka četností proměnné Známká

	Četnost	Kumulativní (četnost)	Rel. četnost	Kumulativní (rel. četnost)
1	15	15	15,62500	15,6250
2	35	50	36,45833	52,0833
3	46	96	47,91667	100,0000

Obrázek 2. Polygon četností proměnné Známká



**Komentář:** Téměř polovina (47,9 %) studentů z výběrového souboru získalo z matematiky v 1. semestru známku 3 – celkem 46. Dvojku pak mělo 35 studentů (36,5 %) a jedničku získalo 15 studentů (15,6 %).

## Úkol 2.

Pro proměnné Pohlaví a Znamka sestavte kontingenční tabulky absolutních a relativních četností, sloupcově a řádkově podmíněných relativních četností. Kolik procent žen má z matematiky jedničku? Kolik procent studentů, kteří mají jedničku, jsou muži?

### Řešení

Tabulka 3. Kontingenční tabulka absolutních a relativních četností pro proměnné Pohlaví a Znamka

	Pohlaví \ Znamka	1	2	3	Řádk. (součty)
Četnost	0	11	27	30	68
Celková četn.		11,46%	28,13%	31,25%	70,83%
Četnost	1	4	8	16	28
Celková četn.		4,17%	8,33%	16,67%	29,17%
Četnost	V. skup.	15	35	46	96
Celková četn.		15,63%	36,46%	47,92%	

**Komentář:** Nejvíce jsou zastoupeny ve výběrovém souboru dívky, které dostaly z matematiky trojku – 31,25 %. Dívek s dvojkou je jen o něco málo méně – 28,13 %. Nejméně, 4,17 %, jsou zastoupeni chlapci s jedničkou.

Tabulka 4. Kontingenční tabulka řádkově podmíněných relativních četností pro proměnné Pohlaví a Znamka

	Pohlaví \ Znamka	1	2	3	Řádk. (součty)
Četnost	0	11	27	30	68
Řádk. četn.		16,18%	39,71%	44,12%	
Četnost	1	4	8	16	28
Řádk. četn.		14,29%	28,57%	57,14%	
Četnost	V. skup.	15	35	46	96

**Komentář:** Ze studentů ženského pohlaví je 16,18 % těch, kteří dostali z matematiky jedničku.

Tabulka 5. Kontingenční tabulka sloupcově podmíněných relativních četností pro proměnné Pohlaví a Znamka

	Pohlaví \ Znamka	1	2	3	Řádk. (součty)
Četnost	0	11	27	30	68
Sloupc. četn.		73,33%	77,14%	65,22%	
Četnost	1	4	8	16	28
Sloupc. četn.		26,67%	22,86%	34,78%	
Četnost	V. skup.	15	35	46	96

**Komentář:** Z těch studentů, kteří mají z matematiky jedničku, je 26,67 % mužského pohlaví.

### Úkol 3.

Podle Sturgesova pravidla stanovte optimální počet třídících intervalů pro proměnné Výška a Hmotnost a nakreslete jejich histogramy, a to:

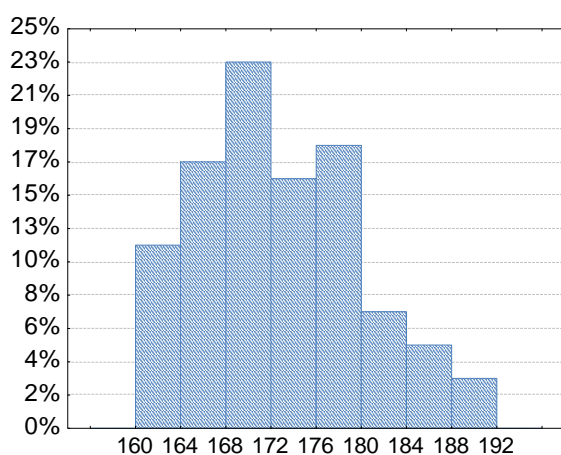
- pro celý soubor
- pro ženy
- pro muže

### Řešení

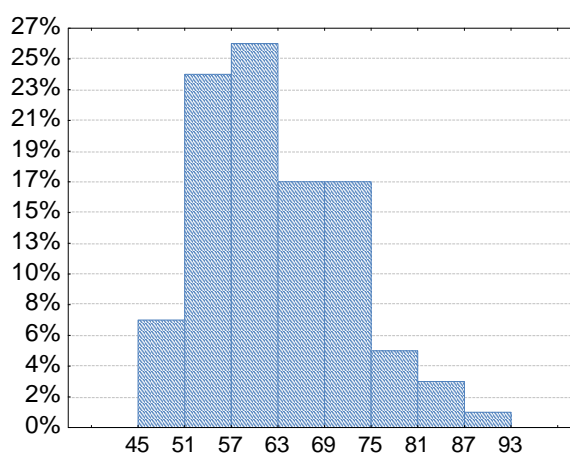
Sturgesovo pravidlo pro určení optimálního počtu třídících intervalů se řídí rovnicí:  $r = 1 + 3,3 \cdot \log_{10}(n)$ , kde  $n$  je rozsah souboru.

Celý soubor má rozsah  $n = 96$ , tedy podle Sturgesova pravidla je optimální počet třídících intervalů roven  $r = 1 + 3,3 \cdot \log_{10}(96) \approx 8$ .

Obrázek 3. Histogram proměnné Výška podle třídících intervalů (pro celý soubor)



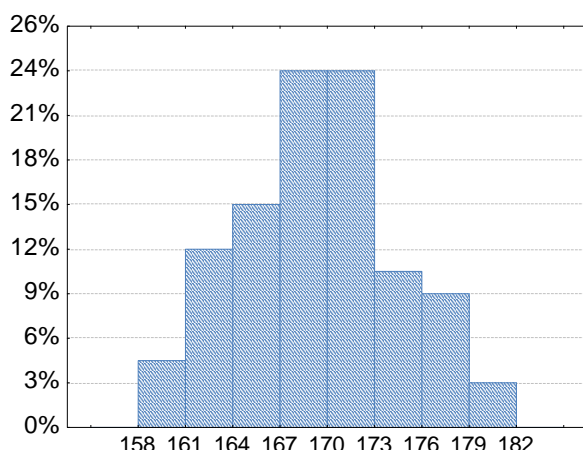
Obrázek 4. Histogram proměnné Hmotnost podle třídících intervalů (pro celý soubor)



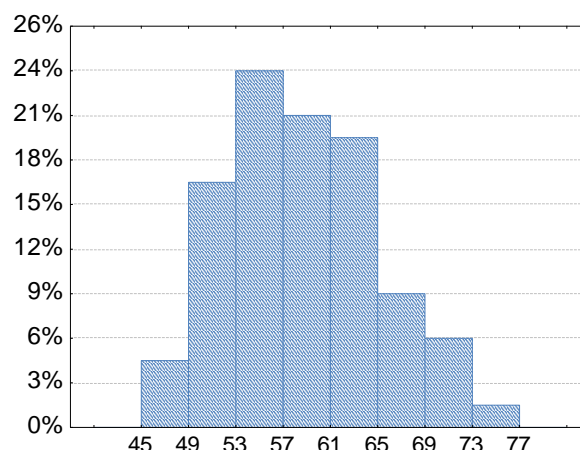
**Komentář:** Z histogramu pro proměnnou Výška je patrné, že nejvíce studentů má výšku z třídícího intervalu 168 až 172 centimetrů (23 %). Naopak nejvyšší studenti jsou zastoupeni nejméně. Podobně dopadl i histogram pro proměnnou Hmotnost. Nejvíce jsou zastoupeni studenti s váhou od 57 do 63 kilogramů (26 %). Nejméně pak studenti s nejvyšší váhou. Oba histogramy pak vykazují kladně zešikmené rozložení, tj. častější jsou nižší hodnoty.

Počet dívek v souboru je  $n_0 = 68$ . Optimální počet třídících intervalů je tedy  $r = 1 + 3,3 \cdot \log_{10}(68) \approx 8$ .

**Obrázek 5. Histogram proměnné Výška podle třídících intervalů (pro dívky)**



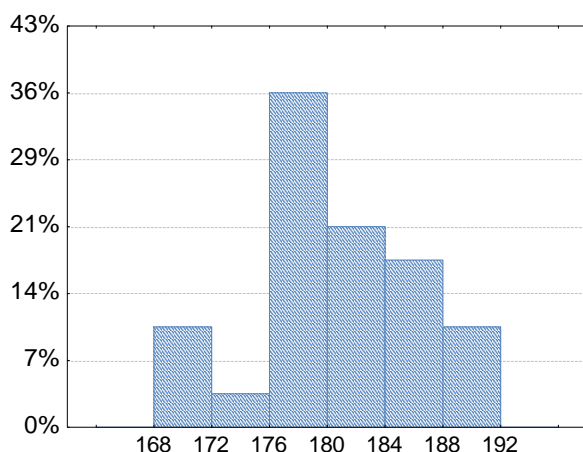
**Obrázek 6. Histogram proměnné Hmotnost podle třídících intervalů (pro dívky)**



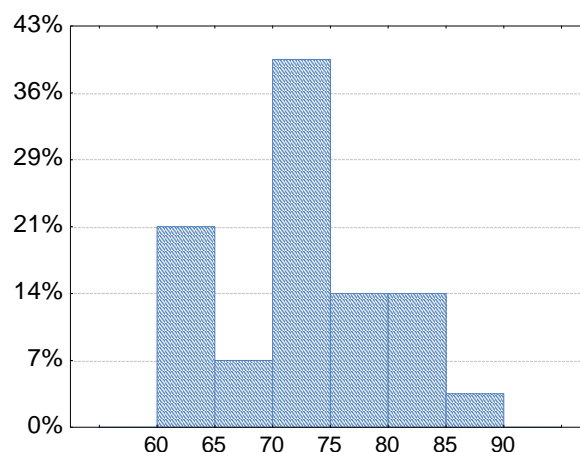
**Komentář:** Nejvíce dívek má výšku z třídících intervalů 167 až 170 a 170 až 173 centimetrů (shodně 24 %). Okrajové hodnoty jsou zastoupeny nejméně. Rozložení proměnné Výška je u dívek v podstatě symetrické, na rozdíl od proměnné Hmotnost, která se jeví kladně zešikmená. Častější jsou menší hodnoty. Nejvíce dívek má hmotnost v třídícím intervalu 53 až 57 kilogramů (24 %). Nejméně jsou pak zastoupeny okrajové hodnoty.

Chlapců je v souboru celkem  $n_1 = 28$ , tedy optimální počet třídících intervalů podle Sturgesova pravidla je  $r = 1 + 3,3 \cdot \log_{10}(28) \approx 6$ .

**Obrázek 7. Histogram proměnné Výška podle třídících intervalů (pro chlapce)**



**Obrázek 8. Histogram proměnné Hmotnost podle třídících intervalů (pro chlapce)**



**Komentář:** Je vidět, že nejčastěji se vyskytují chlapci s výškou z třídícího intervalu 176 až 180 centimetrů (36 %). Nejméně často se pak vyskytuje výška od 172 do 176 centimetrů (3,5 %). Z histogramu hmotnosti lze vyčíst, že nejčastěji jsou zastoupeni chlapci vážící 70 až 75 kilogramů (39 %). Nejméně často se vyskytují chlapci s hmotností 85 až 90 kilogramů (3,5 %). Rozložení hmotnosti se jeví jako lehce kladně zešikmené. Výška je pak, na rozdíl od předchozích, záporně zešikmená – vyšší hodnoty jsou častější než nižší.

## Úkol 4.

Vypočítejte minimum, maximum, medián, průměr, směrodatnou odchylku, šikmost a špičatost proměnných Výška a Hmotnost:

- pro celý soubor
- pro ženy
- pro muže

## Řešení

Tabulka 6. Číselné charakteristiky proměnných Výška a Hmotnost pro celý soubor

	Průměr	Medián	Minimum	Maximum	Sm. odch.	Šikmost	Špičatost
Výška	173,1875	172,0000	160,0000	192,0000	7,336014	0,427816	-0,299328
Hmotnost	63,3750	62,5000	48,0000	90,0000	9,327097	0,587740	-0,134992

**Komentář:** Směrodatná odchylka je u proměnné Hmotnost cca o dvě jednotky vyšší než u proměnné Výška. Obě proměnné dále vykazují kladně zešikmené rozložení (levostranná asymetrie), které je dle hodnot špičatosti ploché (podnormální špičatost). U výšky platí, v porovnání s hmotností, rozložení méně zešikmené a více ploché.

Tabulka 7. Číselné charakteristiky proměnných Výška a Hmotnost pro ženy

	Průměr	Medián	Minimum	Maximum	Sm. odch.	Šikmost	Špičatost
Výška	169,9412	170,0000	160,0000	181,0000	5,063426	0,016957	-0,637971
Hmotnost	59,2647	58,0000	48,0000	75,0000	6,391734	0,322420	-0,445733

Tabulka 8. Číselné charakteristiky proměnných Výška a Hmotnost pro muže

	Průměr	Medián	Minimum	Maximum	Sm. odch.	Šikmost	Špičatost
Výška	181,0714	181,0000	168,0000	192,0000	5,893679	-0,249187	0,088572
Hmotnost	73,3571	73,0000	61,0000	90,0000	7,670117	0,171732	-0,532916

**Komentář:** Chlapci vykazují v průměru daleko vyšší hodnoty u proměnných Výška a Hmotnost než dívky. Stejně tak maximální a minimální hodnoty i směrodatná odchylka jsou u chlapců vyšší. V porovnání s celým souborem je však směrodatná odchylka nižší jak u dívek, tak u chlapců.

Výška u dívek je v podstatě symetrická, avšak s plochým rozložením. Ploché rozložení lze také pozorovat u hmotnosti, která je zároveň kladně zešikmená.

Výška chlapců je na rozdíl od dívek, či v porovnání s celým souborem, záporně zešikmená s rozložením téměř normálním. Hmotnost je pak levostranně asymetrická s plochým rozložením.



## Úkol 5.

Vypočtete a interpretujte Pearsonův koeficient korelace proměnných Výška a Hmotnost:

- pro celý soubor
- pro ženy
- pro muže

## Řešení

Tabulka 9. Pearsonův koeficient korelace proměnných Výška a Hmotnost pro celý soubor

	Výška	Hmotnost
Výška	1,00	0,74
Hmotnost	0,74	1,00

Tabulka 10. Pearsonův koeficient korelace proměnných Výška a Hmotnost pro ženy

	Výška	Hmotnost
Výška	1,00	0,51
Hmotnost	0,51	1,00

Tabulka 11. Pearsonův koeficient korelace proměnných Výška a Hmotnost pro muže

	Výška	Hmotnost
Výška	1,00	0,49
Hmotnost	0,49	1,00

**Komentář:** U dívek a u chlapců existuje středně silná kladná korelace mezi proměnnými Výška a Hmotnost. To znamená, že čím vyšší (resp. nižší) hodnota jedné proměnné, tím vesměs vyšší (resp. nižší) hodnota druhé proměnné. V celém souboru je pak kladná korelace mezi proměnnými Výška a Hmotnost ještě silnější.

## Úkol 6.

Najděte rovnici regresní přímky vyjadřující závislost proměnné *Hmotnost* na proměnné *Výška*. Jaký je index determinace a co vyjadřuje? Jaká je predikovaná hodnota hmotnosti pro výšku 175 cm? Nalezenou regresní přímku zakreslete do dvourozměrného tečkového diagramu.

### Řešení

**Tabulka 12. Výsledky regrese se závislou proměnnou: Hmotnost**  
 $R = ,73970198$   $R^2 = ,54715901$   $\text{Upravené } R^2 = ,54234156$   
 $F(1,94) = 113,58$   $p < ,00000$   $\text{Směr. chyba odhadu} : 6,3098$

	b*	Sm. chyba (z b*)	b	Sm. chyba (z b)	t(94)	p-hodn.
Abs. člen			-99,5020	15,29667	-6,50481	0,000000
Výška	0,739702	0,069408	0,9405	0,08825	10,65732	0,000000

**Rovnice regresní přímky:**  $\text{Hmotnost} = -99,5 + 0,94 \cdot \text{Výška}$

Zvýší-li se výška o jeden centimetr, zvýší se hmotnost v průměru o 0,94 kilogramu.

**Index determinace:**  $ID^2 = 0,547$

Model regresní přímky vysvětluje variabilitu proměnné *Hmotnost* z 54,7 %.

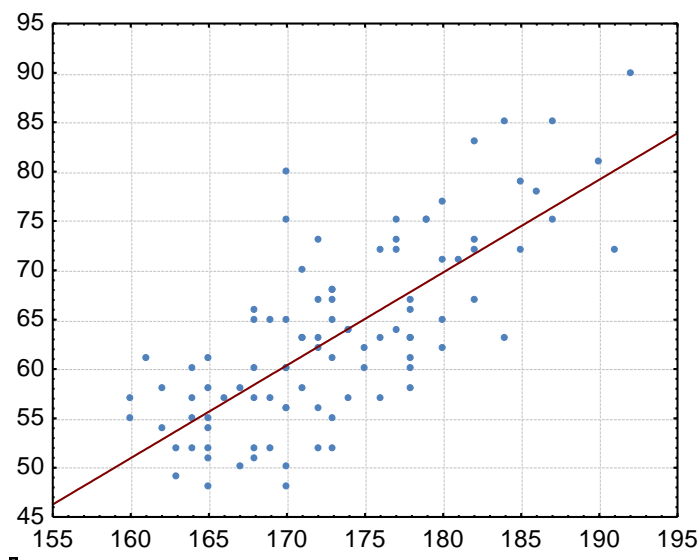
**Tabulka 13. Tabulka predikce**

	b-váha	Hodnota	b-váha (* Hodnot)
Výška	0,940466	175,0000	164,5816
Abs. člen			-99,5020
Předpověď			65,0796
-95,0%LS			63,7621
+95,0%LS			66,3971

**Komentář:** Tabulka udává, že predikovaná hodnota hmotnosti pro výšku 175 cm je:

$\text{Hmotnost} = -99,5 + 0,94 \cdot \text{Výška} = 65,08 \text{ kg}$

**Obrázek 9. Dvourozměrný tečkový diagram se zakreslenou regresní přímkou**



**Komentář:** Jednotlivé body jsou více méně rozmístěny kolem regresní přímky, tedy pro modelování dané závislosti je uvedená regresní přímka vhodná.

## Použitá literatura

**Budíková, Marie, Mikoláš, Štěpán a Osecký, Pavel. 2007.** *Popisná statistika*. Brno : Masarykova univerzita, 2007. ISBN: 978-80-210-4246-9.

## Použitý software

**Trial verze programu STATISTICA Cz 9.0.** Copyright StatSoft, Inc. 1984-2009. (<http://www.statsoft.com/>)

→ Pro účely výpočtů tabulek a grafů.

**MS Office 2007.** Copyright Microsoft Corporation 2007. (<http://www.microsoft.com/>)

→ Pro účely vizuální prezentace.

## Seznam tabulek

TABULKA 1. TABULKA ČETNOSTÍ PROMĚNNÉ POHLAVÍ .....	4
TABULKA 2. TABULKA ČETNOSTÍ PROMĚNNÉ ZNÁMKA.....	4
TABULKA 3. KONTINGENČNÍ TABULKA ABSOLUTNÍCH A RELATIVNÍCH ČETNOSTÍ PRO PROMĚNNÉ POHLAVÍ A ZNÁMKA .....	5
TABULKA 4. KONTINGENČNÍ TABULKA ŘÁDKOVĚ PODMÍNĚNÝCH RELATIVNÍCH ČETNOSTÍ PRO PROMĚNNÉ POHLAVÍ A ZNÁMKA .....	5
TABULKA 5. KONTINGENČNÍ TABULKA SLOUPCOVĚ PODMÍNĚNÝCH RELATIVNÍCH ČETNOSTÍ PRO PROMĚNNÉ POHLAVÍ A ZNÁMKA .....	5
TABULKA 6. ČÍSELNÉ CHARAKTERISTIKY PROMĚNNÝCH VÝŠKA A HMOTNOST PRO CELÝ SOUBOR.....	8
TABULKA 7. ČÍSELNÉ CHARAKTERISTIKY PROMĚNNÝCH VÝŠKA A HMOTNOST PRO ŽENY.....	8
TABULKA 8. ČÍSELNÉ CHARAKTERISTIKY PROMĚNNÝCH VÝŠKA A HMOTNOST PRO MUŽE .....	8
TABULKA 9. PEARSONŮV KOEFICIENT KORELACE PROMĚNNÝCH VÝŠKA A HMOTNOST PRO CELÝ SOUBOR .....	9
TABULKA 10. PEARSONŮV KOEFICIENT KORELACE PROMĚNNÝCH VÝŠKA A HMOTNOST PRO ŽENY .....	9
TABULKA 11. PEARSONŮV KOEFICIENT KORELACE PROMĚNNÝCH VÝŠKA A HMOTNOST PRO MUŽE .....	9
TABULKA 12. VÝSLEDKY REGRESE SE ZÁVISLOU PROMĚNNOU: HMOTNOST $R = ,73970198$ $R^2 = ,54715901$ UPRAVENÉ $R^2 = ,54234156$ $F(1,94) = 113,58$ $p < ,00000$ SMĚROD. CHYBA ODHADU : 6,3098 .....	10
TABULKA 13. TABULKA PREDIKCE.....	10

## Seznam grafů

OBRÁZEK 1. SLOUPKOVÝ DIAGRAM PROMĚNNÉ POHLAVÍ .....	4
OBRÁZEK 2. POLYGON ČETNOSTÍ PROMĚNNÉ ZNÁMKA.....	4
OBRÁZEK 3. HISTOGRAM PROMĚNNÉ VÝŠKA PODLE TŘÍDÍCÍCH INTERVALŮ (PRO CELÝ SOUBOR).....	6
OBRÁZEK 4. HISTOGRAM PROMĚNNÉ HMOTNOST PODLE TŘÍDÍCÍCH INTERVALŮ (PRO CELÝ SOUBOR) .....	6
OBRÁZEK 5. HISTOGRAM PROMĚNNÉ VÝŠKA PODLE TŘÍDÍCÍCH INTERVALŮ (PRO DÍVKY) .....	7
OBRÁZEK 6. HISTOGRAM PROMĚNNÉ HMOTNOST PODLE TŘÍDÍCÍCH INTERVALŮ (PRO DÍVKY) .....	7
OBRÁZEK 7. HISTOGRAM PROMĚNNÉ VÝŠKA PODLE TŘÍDÍCÍCH INTERVALŮ (PRO CHLAPCE).....	7
OBRÁZEK 8. HISTOGRAM PROMĚNNÉ HMOTNOST PODLE TŘÍDÍCÍCH INTERVALŮ (PRO CHLAPCE) .....	7
OBRÁZEK 9. DVOUROZMĚRNÝ TEČKOVÝ DIAGRAM SE ZAKRESLENOU REGRESNÍ PŘÍMKOU.....	10