

1.(a)

Nulová hypotéza: druh kriminality a nejvyšší dosažené vzdělání jsou nezávislé náhodné veličiny

Alternativní: druh kriminality a nejvyšší dosažené vzdělání jsou závislé náhodné veličiny

Vytvořme textový soubor s daty jako

Kriminalita.txt

48 31 22

10 9 11

43 29 27

109 77 58

10 8 7

6 9 6

```
> mosaicplot(Kriminalita)
```

L, 29L, 77L, 8L, 9L, 22L, 11L, 27L, 58L, 7L, 6L), .Dim = c(6L



b)

```
> chisq.test(Kriminalita)
```

Pearson's Chi-squared test

data: Kriminalita

X-squared = 5.9465, df = 10, p-value = 0.8197

P-hodnota je velmi vysoká 0.8197. Nezamítáme nulovou hypotézu. Na hladině významnosti menší než 0.8197 bychom ještě zamítali nulovou hypotézu.

c) Předpoklady: chisq je pouze asymptotický test, potřebujeme mít dostatečný počet teoretických četností. Teoretické četnosti musí být alespoň 5 v 80% případů, jinak staší aby byly větší než 1. R vyhazuje warning pokud alespoň jedna s těch teoretických četností je méně než 5

2. a)

```
> cesta<-"C:/Users/Serhii/Documents/Aplikovana_statistika+_R/mladata.txt"
```

```
> mladata<-read.table(cesta, header = TRUE)
```

```
> summary(mladata)
```

matka dite

Min. :45.95 Min. :2.490

1st Qu.:56.22 1st Qu.:3.027

Median :59.94 Median :3.410

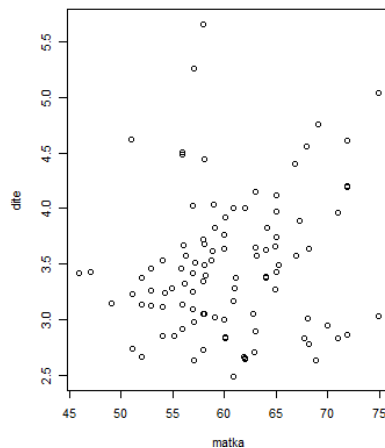
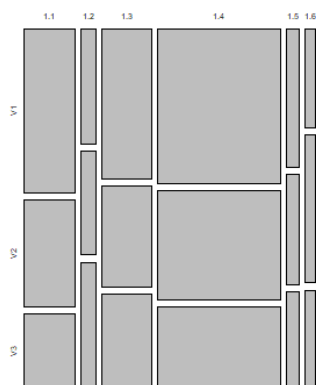
Mean :60.46 Mean :3.483

3rd Qu.:64.92 3rd Qu.:3.777

Max. :74.96 Max. :5.660

```
> plot(mladata)
```

L, 29L, 77L, 8L, 9L, 22L, 11L, 27L, 58L, 7L, 6L). .Dim = c(6L)



```
> cor(mladata, method = "pearson")
```

matka dite

matka 1.0000000 0.1617939

dite 0.1617939 1.0000000

```
> cor(mladata, method = "spearman")
```

matka dite

matka 1.0000000 0.1507756

dite 0.1507756 1.0000000

```
> cor(mladata, method = "kendall")
```

```
matka    dite
```

```
matka 1.0000000 0.1140486
```

```
dite 0.1140486 1.0000000
```

Korelační koeficient je kladný, vztah je spíš korespondující - to znamená, že ANO vyšší váha matky způsobuje vyšší váhu dítě a naopak

Nulová hypotéza: váha dítěte nezávisle na hmotě matky

Alternativní: váha dítěte přímo závisí na váze matky, čím větší váha matky, tím větší váha dítěte

```
> lregrese<-lm(mladata$dite~mladata$matka)
```

```
> lregrese
```

Call:

```
lm(formula = mladata$dite ~ mladata$matka)
```

Coefficients:

```
(Intercept) mladata$matka
```

```
2.49921    0.01626
```

```
> summary(lregrese)
```

Call:

```
lm(formula = mladata$dite ~ mladata$matka)
```

Residuals:

```
Min    1Q  Median    3Q   Max
```

```
-0.99989 -0.45360 -0.05815  0.28639  2.21744
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) 2.49921 0.60902  4.104 8.42e-05 ***
```

```
mladata$matka 0.01626 0.01002  1.623  0.108
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.6164 on 98 degrees of freedom
```

```
Multiple R-squared: 0.02618, Adjusted R-squared: 0.01624
```

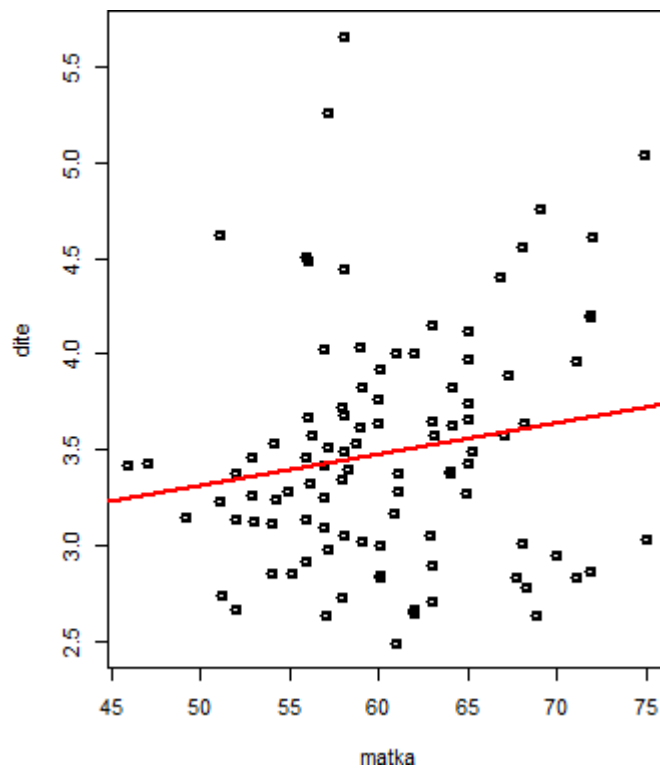
```
F-statistic: 2.634 on 1 and 98 DF, p-value: 0.1078
```

0.01626 -> rostoucí lineární funkce

> residuals(lregrese)

1	2	3	4	5	6	7	8	9
-0.546045425	-0.392725301	1.290477465	0.296825620	-0.268731942	0.370196695	0.076052011	-	
0.267666240	0.996136165							
10	11	12	13	14	15	16	17	18
-0.101912063	0.305344807	1.070129524	-0.828788555	1.137874337	0.813982116	2.217437346	-	
0.112629824	-0.688413073							
19	20	21	22	23	24	25	26	27
0.443443987	-0.770656172	-0.212170188	-0.335322003	-0.439478010	0.166187117	-0.797273774	-	
0.488894590	-0.646230718							
28	29	30	31	32	33	34	35	36
0.531194461	0.954952341	1.101430706	0.151683169	-0.230100294	-0.684648767	-0.070319554		
0.288384926	0.162799057							
37	38	39	40	41	42	43	44	45
0.183258694	-0.636230718	0.156063333	-0.045978254	-0.319242530	-0.857947010	-0.148131538	-	
0.112495483	-0.807829653							
46	47	48	49	50	51	52	53	54
-0.140919193	-0.824492546	0.173427768	0.940706518	0.413096046	-0.623561185	0.285721054		
0.521519756	-0.999893121							
55	56	57	58	59	60	61	62	63
-0.005322003	0.046786756	0.492540933	0.126113520	0.054487043	-0.393538540	-0.204323471		
0.052406592	-0.596348840							
64	65	66	67	68	69	70	71	72
-0.160964484	-0.711098825	0.258990991	-0.687577190	0.572473762	0.626438815	-0.471121470		
0.081587692	-0.284789534							
73	74	75	76	77	78	79	80	81
0.088872869	-0.475417480	-0.447761717	-0.092472838	1.322562813	0.563258694	-0.098799113		
1.831912987	-0.812747947							
82	83	84	85	86	87	88	89	90
0.032187331	-0.100335773	-0.846808477	-0.867296420	-0.009108189	-0.590986364	0.035513881	-	
0.989035357	0.277762642							
91	92	93	94	95	96	97	98	99
-0.150313893	0.100550297	-0.527340945	0.594840645	0.166046349	0.237112051	0.509618936	-	
0.174834060	0.103909285							
100								
-0.127229249								

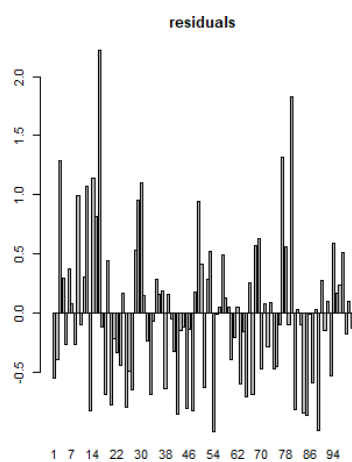
```
> abline(lregrese,col="red", lwd = 2)
```



podminky pro regresni model jsou

1. nulova stredni hodnota rezidui:

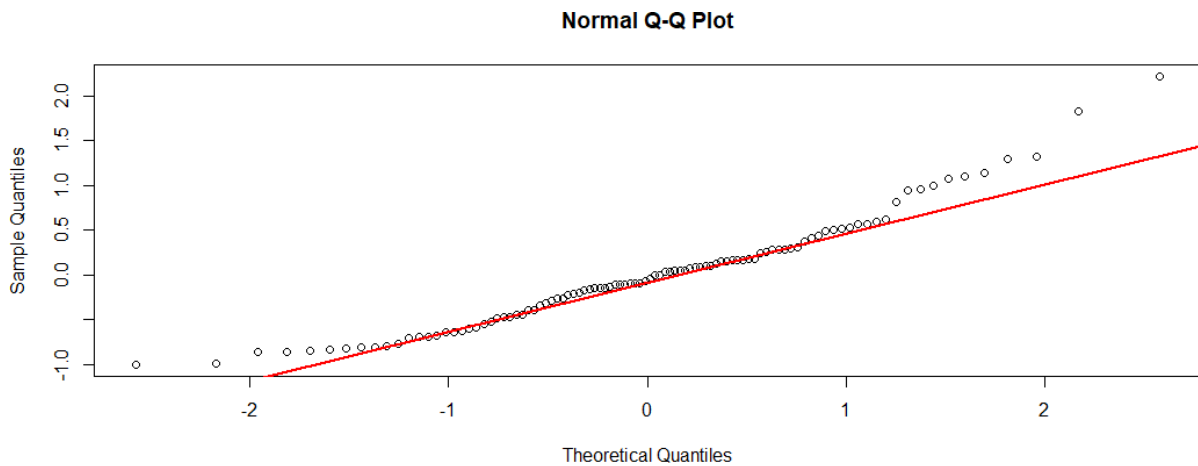
```
> barplot(residuals(lregrese), main="residuals")
```



Kolísaji kolem nuly

ostatni podminky na rezidua: nekorelovanost jednotlivych rezidui a stejny rozptyl.

```
> qqnorm(residuals(lregrese))  
> qqline(residuals(lregrese),col="red", lwd = 2)
```



Kromě konečků vypadá víceméně dobře

Navic potrebujeme normalitu rezidui

```
> shapiro.test(residuals(lregrese))
```

Shapiro-Wilk normality test

data: residuals(lregrese)

W = 0.94993, p-value = 0.0008203

Malé p-value znamená, že nemáme normalitu rezidui

Nemuzeme sestavit PAS SPOLEHLIVOSTI A PREDIKCNI PAS

```
> summary(lregrese)
```

Call:

```
lm(formula = mladata$dite ~ mladata$matka)
```

Residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

-0.99989 -0.45360 -0.05815 0.28639 2.21744

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.49921 0.60902 **4.104 8.42e-05 *****

mladata\$matka 0.01626 0.01002 1.623 **0.108**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6164 on 98 degrees of freedom

Multiple R-squared: 0.02618, Adjusted R-squared: 0.01624

F-statistic: 2.634 on 1 and 98 DF, p-value: 0.1078

4.104 8.42e-05 - je velmi malá pravděpodobnost, že koeficient, který je tam přítomen, bude zanedbatelný resp se bude rovnat nule

0.108 = 10,8% - hodnota, při které by byla stále zamítnuta nulová hypotéza, při které by tento prediktor nevstupoval do modelu této závislé proměnné

d) 10,8%

e) Ano, dá se předpokládat, že velká váha matky povede spíše k vyšší váze dítěte než k váze nižší.