

Zápočtová úloha NMAI061/2018

Jméno a příjmení

Datum narození: 19700131

1. V hlavičce tohoto souboru správně uveďte svoje jméno (v příkazu `\author{...}`) a datum narození (`\date{...}`).
2. T_EXový kód zadání Vaší zápočtové úlohy získáte pomocí příkazu `Sweave("zapoctovka.Rnw")` v R. Přitom je nutné mít nainstalovanou knihovnu `lmtest`. Příkaz `Sweave("zapoctovka.Rnw")` zadávejte až po zadání svého data narození do souboru `zapoctovka.Rnw!`
3. Vzniklý T_EXový soubor `zapoctovka.tex` přejmenujte na `prijmeni.jmeno.tex` a přeložte do PDF pomocí `pdflatex`.
4. Odpovědi na otázky můžete psát přímo do Vašeho T_EXového souboru nebo do zvláštního souboru `prijmeni.jmeno-odpovedi.pdf`.
5. Úlohu můžete odevzdat elektronicky (buď soubor `prijmeni.jmeno.pdf` se zadáním i s odpověďmi nebo soubor `prijmeni.jmeno.pdf` se zadáním a soubor `prijmeni.jmeno-odpovedi.pdf` s odpověďmi).
6. Za každou správně vyřešenou úlohu získáte 1 bod. Pro získání zápočtu je nutné získat minimálně 8 bodů (počet bodů může být zohledněn i při zkoušce).
7. Zápočty budu zapisovat při zkoušce.

Data

Načtete datový soubor `ChickWeight` a nastudujte jeho popis v nápovědě.

```
> data(ChickWeight)
```

Analyzovat budeme data pouze pro dvě náhodně vybraná kuřata. Výsledná data jsou:

	weight	Time	Chick
144	41	0	13
145	48	2	13
146	53	4	13
147	60	6	13
148	65	8	13
149	67	10	13
150	71	12	13
151	70	14	13
152	71	16	13
153	81	18	13
154	91	20	13
155	96	21	13
281	42	0	26
282	48	2	26
283	57	4	26

284	74	6	26
285	93	8	26
286	114	10	26
287	136	12	26
288	147	14	26
289	169	16	26
290	205	18	26
291	236	20	26
292	251	21	26

Vaším úkolem bude prozkoumat závislost váhy těchto dvou kuřat na čase.

Úkoly

Úkol 1: Okomentujte následující popisné statistiky a grafy (mimo jiné můžete okomentovat, které kuře je těžší, vysvětlit význam jednotlivých popisných statistik, tj. co tyto statistiky odhadují nebo měří a jaké jsou jejich vztahy a vysvětlit proč jsou statistiky u proměnné Time stejné pro obě kuřata).

```
> sapply(kurata[, -3], tapply, Chick, mean)
```

```
      weight      Time
13 67.83333 10.91667
26 131.00000 10.91667
```

```
> sapply(kurata[, -3], tapply, Chick, median)
```

```
      weight Time
13   68.5   11
26  125.0   11
```

```
> sapply(kurata[, -3], tapply, Chick, sd)
```

```
      weight      Time
13 16.32529 7.076958
26 72.54716 7.076958
```

```
> sapply(kurata[, -3], tapply, Chick, var)
```

```
      weight      Time
13 266.5152 50.08333
26 5263.0909 50.08333
```

```
> par(mfrow=c(2,2))
```

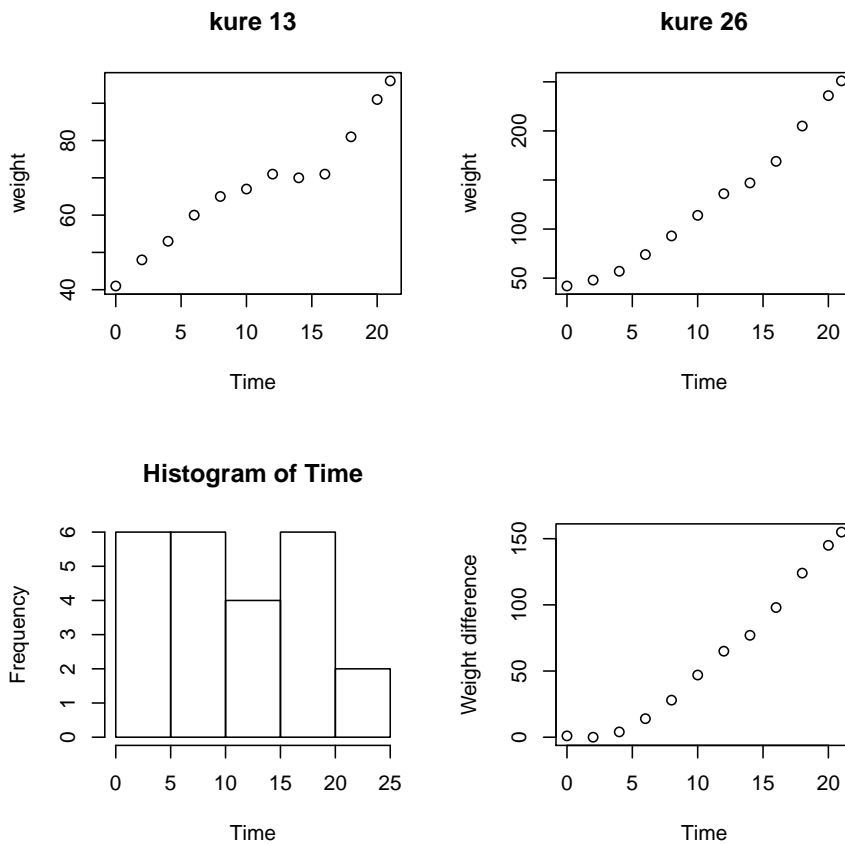
```
> plot(weight~Time, main=paste("kure", Chick[1]), subset=(Chick==Chick[1]))
```

```
> plot(weight~Time, main=paste("kure", Chick[13]), subset=(Chick==Chick[13]))
```

```
> hist(Time)
```

```
> poc=max(k)-1
```

```
> plot(weight[13:24]-weight[1:12]~Time[1:12], ylab="Weight difference", xlab="Time")
```



Úkol 1: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v $\text{T}_{\text{E}}\text{X}$ u) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Úkol 2: Vysvětlete, co přesně znamenají jednotlivé odhady koeficientů v následujícím lineárním modelu. Které koeficienty jsou významně odlišné od nuly?

```
> kure1.lm=lm(weight~Time+I(Time^2),subset=(Chick==Chick[k1s]))
> kure1.lm
```

Call:

```
lm(formula = weight ~ Time + I(Time^2), subset = (Chick == Chick[k1s]))
```

Coefficients:

```
(Intercept)      Time      I(Time^2)
   44.41964     1.91929     0.01491
```

```
> summary(kure1.lm)
```

Call:

```
lm(formula = weight ~ Time + I(Time^2), subset = (Chick == Chick[k1s]))
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-7.945 -2.953  1.033   2.555  4.700
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	44.41964	3.19636	13.897	2.19e-07	***
Time	1.91929	0.69830	2.749	0.0225	*
I(Time^2)	0.01491	0.03138	0.475	0.6460	

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

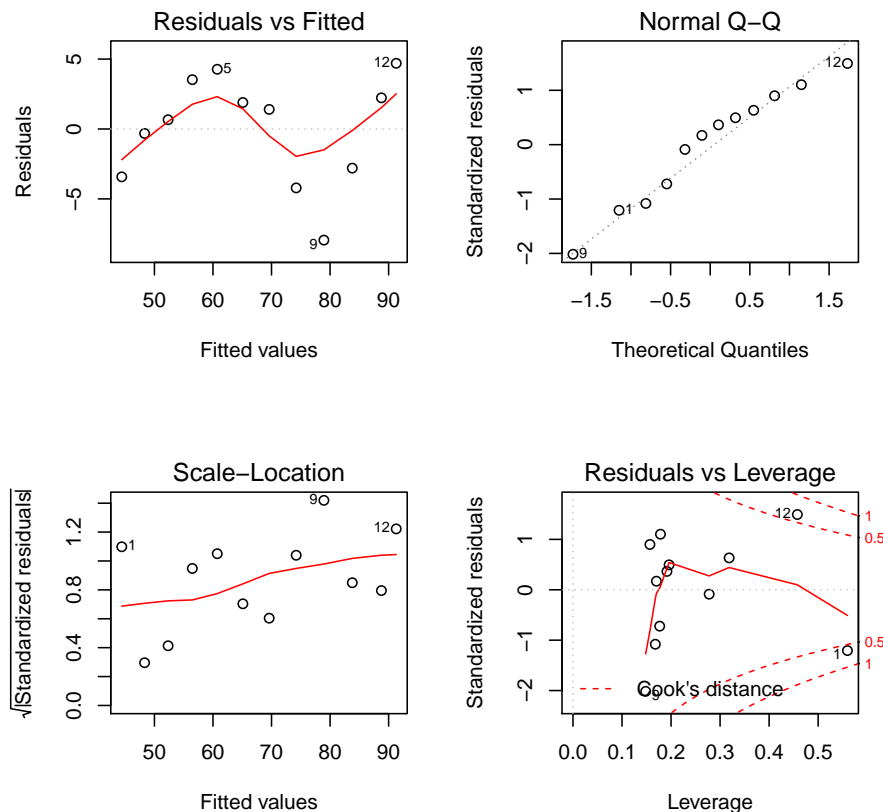
Residual standard error: 4.272 on 9 degrees of freedom

Multiple R-squared: 0.944, Adjusted R-squared: 0.9315

F-statistic: 75.82 on 2 and 9 DF, p-value: 2.332e-06

```
> par(mfrow=c(2,2))
```

```
> plot(kure1.lm)
```



Úkol 2: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v \TeX u) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Úkol 3: Vysvětlete, co přesně znamenají jednotlivé odhady koeficientů v následujícím lineárním modelu. Které koeficienty jsou významně odlišné od nuly?

```
> kure2.lm=lm(weight~Time+I(Time^2),subset=(Chick==Chick[k2s]))  
> kure2.lm
```

```
Call:
lm(formula = weight ~ Time + I(Time^2), subset = (Chick == Chick[k2s]))
```

```
Coefficients:
(Intercept)      Time      I(Time^2)
      40.222      4.066      0.281
```

```
> summary(kure2.lm)
```

```
Call:
lm(formula = weight ~ Time + I(Time^2), subset = (Chick == Chick[k2s]))
```

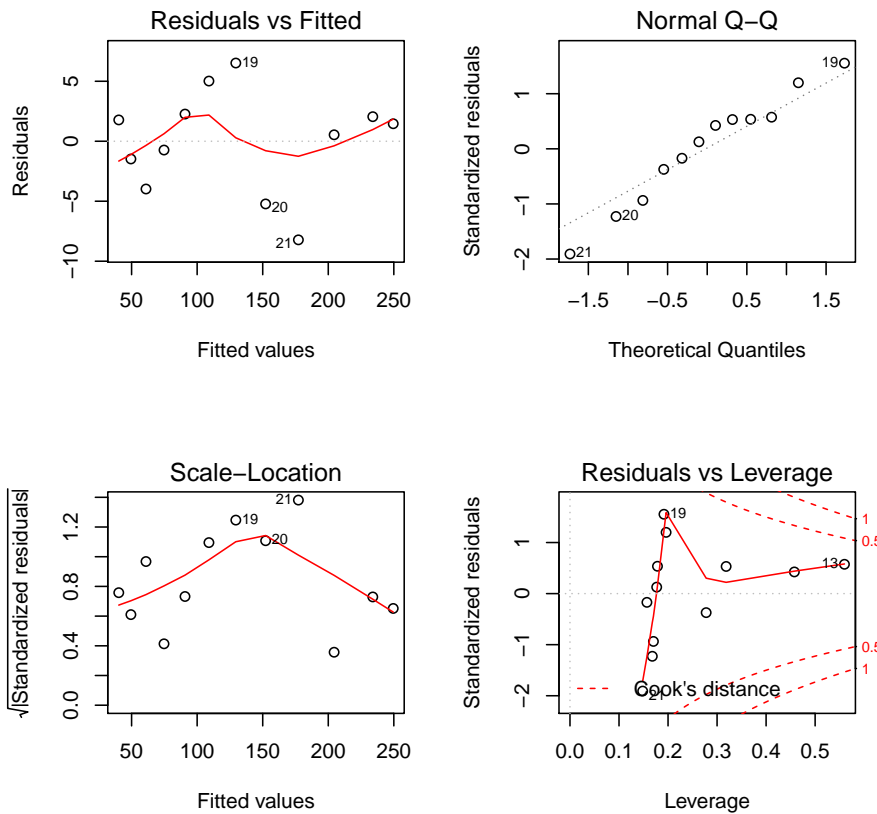
```
Residuals:
      Min       1Q   Median       3Q      Max
-8.2184 -2.1027  0.9995  2.1031  6.5212
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  40.22164     3.49188   11.519 1.09e-06 ***
Time          4.06550     0.76286    5.329 0.000475 ***
I(Time^2)    0.28105     0.03428    8.198 1.82e-05 ***
```

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

```
Residual standard error: 4.667 on 9 degrees of freedom
Multiple R-squared:  0.9966,    Adjusted R-squared:  0.9959
F-statistic: 1325 on 2 and 9 DF,  p-value: 7.649e-12
```

```
> par(mfrow=c(2,2))
> plot(kure2.lm)
```



Úkol 3: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v $\text{T}_{\text{E}}\text{X}_{\text{u}}$) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Úkol 4: Pomocí následujícího lineárního modelu rozhodněte, jestli obě kuřata mohla být při narození stejně těžká a jestli obě kuřata rostla stejně rychle. Které kuře roste rychleji? Je pozorovaný rozdíl rychlosti růstu statisticky signifikantní? Při interpretaci parametrů zvažte i matici modelu získanou příkazem `model.matrix(kure.lm)`.

```
> kure.lm=lm(weight~Time+Chick+Time:Chick)
> summary(kure.lm)
```

Call:

```
lm(formula = weight ~ Time + Chick + Time:Chick)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-15.1514	-6.2903	0.3968	3.3082	21.2928

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	43.3844	5.2336	8.290	6.70e-08	***
Time	2.2396	0.4073	5.498	2.21e-05	***
Chick26	-22.6772	7.4015	-3.064	0.00613	**
Time:Chick26	7.8636	0.5761	13.651	1.35e-11	***

```

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

Residual standard error: 9.561 on 20 degrees of freedom
Multiple R-squared:  0.9784,    Adjusted R-squared:  0.9752
F-statistic: 302.4 on 3 and 20 DF,  p-value: < 2.2e-16

> model.matrix(kure.lm)

  (Intercept) Time Chick26 Time:Chick26
1           1     0       0           0
2           1     2       0           0
3           1     4       0           0
4           1     6       0           0
5           1     8       0           0
6           1    10       0           0
7           1    12       0           0
8           1    14       0           0
9           1    16       0           0
10          1    18       0           0
11          1    20       0           0
12          1    21       0           0
13          1     0       1           0
14          1     2       1           2
15          1     4       1           4
16          1     6       1           6
17          1     8       1           8
18          1    10       1          10
19          1    12       1          12
20          1    14       1          14
21          1    16       1          16
22          1    18       1          18
23          1    20       1          20
24          1    21       1          21
attr("assign")
[1] 0 1 2 3
attr("contrasts")
attr("contrasts")$Chick
[1] "contr.treatment"

> anova(kure.lm2<-lm(weight~Time),kure.lm)

Analysis of Variance Table

Model 1: weight ~ Time
Model 2: weight ~ Time + Chick + Time:Chick
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1     22 42801
2     20  1828  2    40973 224.12 2.021e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Úkol 4: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v $\text{T}_{\text{E}}\text{X}$ u) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Úkol 5: Je některé z následujících studentizovaných reziduí příliš velké?

```
> rstudent(kure.lm)
```

1	2	3	4	5	6
-0.29110449	0.01582717	0.07356274	0.34789741	0.39886576	0.13003082
7	8	9	10	11	12
0.07894489	-0.51290911	-0.91802898	-0.30334878	0.32964241	0.67320116
13	14	15	16	17	18
3.22748706	0.83706240	-0.46365260	-0.81323783	-0.93721275	-0.83986813
19	20	21	22	23	24
-0.64056830	-1.75588275	-1.54969472	0.27384609	1.64626451	2.48632760

Úkol 5: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v \TeX) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Úkol 6: Rozhodněte, jestli je některá z následujících hodnot „leverage“ „podezřelá“.

```
> hatvalues(kure.lm)
```

1	2	3	4	5	6	7
0.29965209	0.22765088	0.17017093	0.12721222	0.09877477	0.08485857	0.08546362
8	9	10	11	12	13	14
0.10058993	0.13023748	0.17440629	0.23309635	0.26788686	0.29965209	0.22765088
15	16	17	18	19	20	21
0.17017093	0.12721222	0.09877477	0.08485857	0.08546362	0.10058993	0.13023748
22	23	24				
0.17440629	0.23309635	0.26788686				

```
> influence.measures(kure.lm)
```

Influence measures of

lm(formula = weight ~ Time + Chick + Time:Chick) :

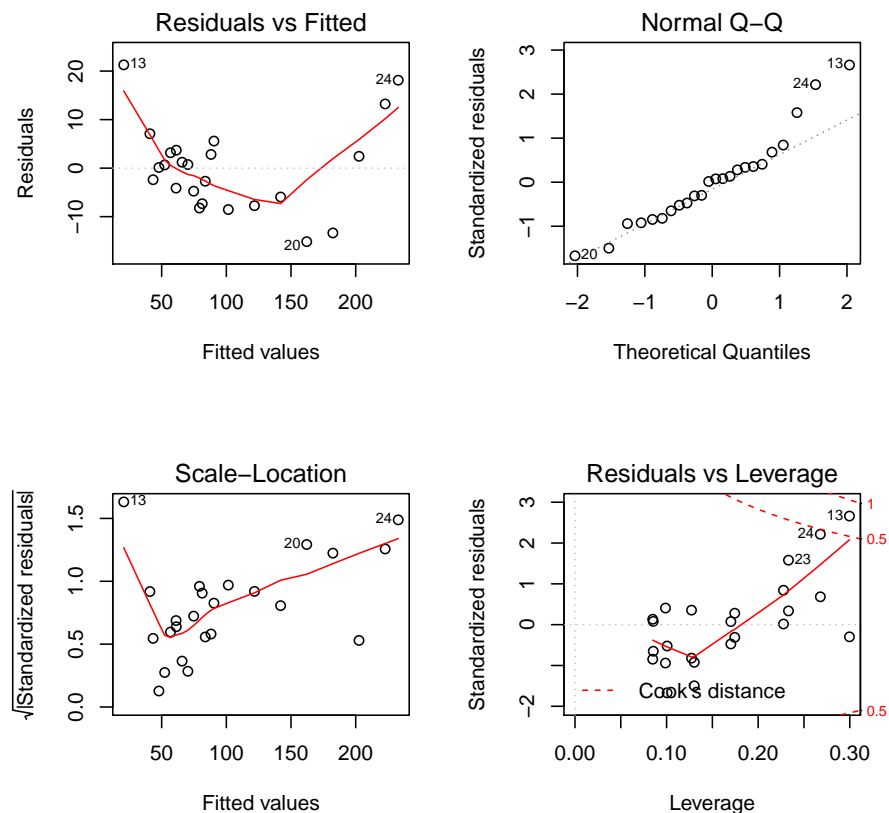
	dfb.1_	dfb.Time	dfb.Ch26	dfb.T.C2	dffit	cov.r	cook.d	hat	inf
1	-1.90e-01	1.62e-01	0.13464	-0.11440	-0.19041	1.722	9.50e-03	0.2997	*
2	8.55e-03	-6.84e-03	-0.00605	0.00484	0.00859	1.590	1.94e-05	0.2277	
3	3.25e-02	-2.38e-02	-0.02299	0.01683	0.03331	1.478	2.92e-04	0.1702	
4	1.23e-01	-7.80e-02	-0.08695	0.05516	0.13282	1.371	4.61e-03	0.1272	
5	1.08e-01	-5.22e-02	-0.07660	0.03692	0.13205	1.318	4.55e-03	0.0988	
6	2.52e-02	-5.31e-03	-0.01782	0.00375	0.03960	1.337	4.12e-04	0.0849	
7	9.33e-03	3.81e-03	-0.00660	-0.00269	0.02413	1.341	1.53e-04	0.0855	
8	-2.20e-02	-7.10e-02	0.01553	0.05024	-0.17153	1.292	7.64e-03	0.1006	
9	3.13e-02	-2.13e-01	-0.02212	0.15075	-0.35524	1.187	3.18e-02	0.1302	
10	3.48e-02	-1.01e-01	-0.02459	0.07124	-0.13942	1.459	5.09e-03	0.1744	
11	-6.65e-02	1.46e-01	0.04700	-0.10301	0.18174	1.565	8.64e-03	0.2331	
12	-1.67e-01	3.38e-01	0.11837	-0.23900	0.40722	1.526	4.26e-02	0.2679	
13	1.43e-15	-5.01e-16	1.49280	-1.26835	2.11114	0.305	7.58e-01	0.2997	*
14	1.96e-16	-6.86e-17	0.31992	-0.25586	0.45445	1.375	5.24e-02	0.2277	
15	-1.32e-16	3.51e-17	-0.14490	0.10606	-0.20996	1.414	1.15e-02	0.1702	
16	-1.87e-16	7.92e-17	-0.20325	0.12894	-0.31048	1.227	2.45e-02	0.1272	
17	-1.73e-16	8.18e-17	-0.17998	0.08675	-0.31027	1.137	2.42e-02	0.0988	
18	-1.12e-16	5.06e-17	-0.11511	0.02424	-0.25575	1.160	1.66e-02	0.0849	
19	-2.06e-17	-1.09e-17	-0.05353	-0.02186	-0.19582	1.233	9.88e-03	0.0855	

20	2.60e-18	-5.66e-17	-0.05318	-0.17198	-0.58721	0.748	7.81e-02	0.1006
21	1.20e-16	-9.85e-17	0.03734	-0.25447	-0.59967	0.877	8.40e-02	0.1302
22	-1.83e-17	1.40e-17	-0.02220	0.06431	0.12586	1.464	4.15e-03	0.1744
23	-4.75e-16	3.89e-16	-0.23471	0.51442	0.90761	0.939	1.90e-01	0.2331
24	-4.41e-16	1.08e-16	-0.43719	0.88270	1.50399	0.543	4.49e-01	0.2679

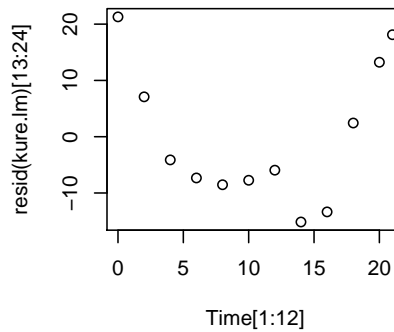
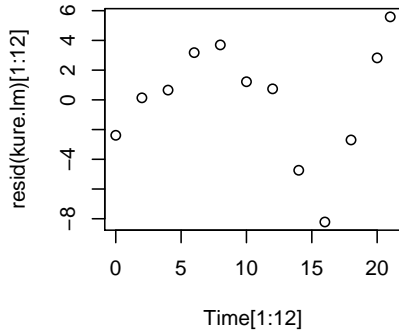
Úkol 6: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v $\text{T}_\text{E}_\text{X}$ u) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Úkol 7: Jsou podle Vašeho názoru splněny všechny předpoklady normálního lineárního modelu? Jak byste na následujících grafech poznali např. porušení předpokladu homoskedasticity (tj. konstantního rozptylu)? Jaké důsledky by mělo porušení tohoto předpokladu?

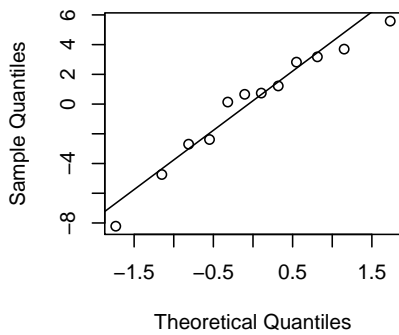
```
> par(mfrow=c(2,2))
> plot(kure.lm)
```



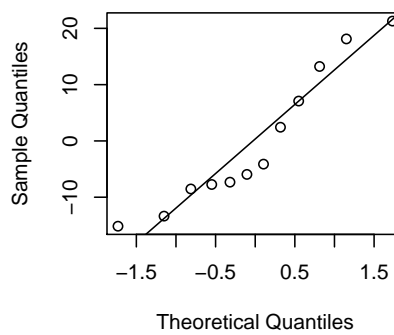
```
> par(mfrow=c(2,2))
> plot(resid(kure.lm)[1:12]~Time[1:12])
> plot(resid(kure.lm)[13:24]~Time[1:12])
> qqnorm(resid(kure.lm)[1:12])
> qqline(resid(kure.lm)[1:12])
> qqnorm(resid(kure.lm)[13:24])
> qqline(resid(kure.lm)[13:24])
```



Normal Q-Q Plot



Normal Q-Q Plot



Úkol 7: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v \TeX) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Úkol 8: Shapiro-Wilkův test se používá k ověřování normality. Okomentujte výsledek tohoto testu použitého na rezidua modelu pro váhy kuřat.

```
> shapiro.test(resid(kure.lm))

Shapiro-Wilk normality test

data:  resid(kure.lm)
W = 0.95455, p-value = 0.3389

> shapiro.test(rstandard(kure.lm))

Shapiro-Wilk normality test

data:  rstandard(kure.lm)
W = 0.94024, p-value = 0.1651

> shapiro.test(rstudent(kure.lm))

Shapiro-Wilk normality test

data:  rstudent(kure.lm)
W = 0.9155, p-value = 0.04649
```

Úkol 8: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v \TeX) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Úkol 9: Durbin-Watsonův test se používá k ověřování nezávislosti po sobě jdoucích pozorování. Okomentujte výsledek tohoto testu použitého na rezidua modelu pro váhy kuřat.

```
> library(lmtest)
> dwtest(kure.lm)
```

Durbin-Watson test

```
data: kure.lm
DW = 0.65529, p-value = 4.609e-07
alternative hypothesis: true autocorrelation is greater than 0
```

Úkol 9: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v \TeX) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Úkol 10: Jak byste pomocí parametrů předchozích lineárních modelů zformulovali hypotézu, že obě kuřata byla ve věku 16 dní stejně těžká? Jak byste tuto hypotézu otestovali?

Úkol 10: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v \TeX) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Úkol 11: Jak byste otestovali hypotézu, že obě kuřata byla po celou dobu sledování (tj. od narození do 20 dní) stejně těžká?

Úkol 11: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v \TeX) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Úkol 12: Stručně a výstižně shrňte výsledky získané analýzou lineárního modelu `kure.lm`. Vysvětlete, které předpoklady normálního lineárního modelu by v tomto případě mohly být porušené a jak by bylo možné model `kure.lm` vylepšit.

Úkol 12: Svoji odpověď můžete buď vepsat přímo do tohoto boxu (v \TeX) nebo můžete každé řešení označit číslem úkolu a vše odevzdat ve zvláštním dokumentu.

Jenom pro kontrolu: datum narození je 19700131 a vybrali jsme kuřata 26 a 13.