

6

Normal Linear Models

6.1 Introduction

This chapter is about models of the form

$$E(Y_i) = \mu_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad ; \quad Y_i \sim N(\mu_i, \sigma^2) \quad (6.1)$$

where Y_1, \dots, Y_N are independent random variables. The link function is the identity function, i.e., $g(\mu_i) = \mu_i$. This model is usually written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (6.2)$$

where

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_N \end{bmatrix}$$

and the e_i 's are independently, identically distributed random variables with $e_i \sim N(0, \sigma^2)$ for $i = 1, \dots, N$. Multiple linear regression, analysis of variance (ANOVA) and analysis of covariance (ANCOVA) are all of this form and together are sometimes called **general linear models**.

Estimation in Normal Linear Model

6.2.1 Maximum likelihood estimation

From Section 5.4.1, the maximum likelihood estimator of $\boldsymbol{\beta}$ is given by

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (6.3)$$

provided $(\mathbf{X}^T \mathbf{X})$ is non-singular. As $E(\mathbf{b}) = \boldsymbol{\beta}$, the estimator is unbiased. It has variance-covariance matrix $\sigma^2(\mathbf{X}^T \mathbf{X})^{-1} = \mathfrak{J}^{-1}$.

In the context of generalized linear models, σ^2 is treated as a nuisance parameter. However it can be shown that

$$\hat{\sigma}^2 = \frac{1}{N-p} (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b}) \quad (6.4)$$

is an unbiased estimator of σ^2 and this can be used to estimate \mathfrak{J} and hence make inferences about \mathbf{b} .

6.2.3 Deviance

From Section 5.6.1

$$\begin{aligned}
 D &= \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b}) \\
 &= \frac{1}{\sigma^2} (\mathbf{y}^T \mathbf{y} - 2\mathbf{b}^T \mathbf{X}^T \mathbf{y} + \mathbf{b}^T \mathbf{X}^T \mathbf{X} \mathbf{b}) \\
 &= \frac{1}{\sigma^2} (\mathbf{y}^T \mathbf{y} - \mathbf{b}^T \mathbf{X}^T \mathbf{y})
 \end{aligned} \tag{6.5}$$

because $\mathbf{X}^T \mathbf{X} \mathbf{b} = \mathbf{X}^T \mathbf{y}$ from equation (6.3).

6.2.4 Hypothesis testing

Consider a null hypothesis H_0 and a more general hypothesis H_1 specified as follows

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_q \end{bmatrix} \quad \text{and} \quad H_1 : \boldsymbol{\beta} = \boldsymbol{\beta}_1 = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

where $q < p < N$. Let \mathbf{X}_0 and \mathbf{X}_1 denote the corresponding design matrices, \mathbf{b}_0 and \mathbf{b}_1 the maximum likelihood estimators, and D_0 and D_1 the deviances. We test H_0 against H_1 using

$$\begin{aligned}
 \Delta D &= D_0 - D_1 = \frac{1}{\sigma^2} \left[(\mathbf{y}^T \mathbf{y} - \mathbf{b}_0^T \mathbf{X}_0^T \mathbf{y}) - (\mathbf{y}^T \mathbf{y} - \mathbf{b}_1^T \mathbf{X}_1^T \mathbf{y}) \right] \\
 &= \frac{1}{\sigma^2} (\mathbf{b}_1^T \mathbf{X}_1^T \mathbf{y} - \mathbf{b}_0^T \mathbf{X}_0^T \mathbf{y})
 \end{aligned}$$

13

F-test of H0 vs H1

$$F = \frac{D_0 - D_1}{p - q} \bigg/ \frac{D_1}{N - p} = \frac{(\mathbf{b}_1^T \mathbf{X}_1^T \mathbf{y} - \mathbf{b}_0^T \mathbf{X}_0^T \mathbf{y})}{p - q} \bigg/ \frac{(\mathbf{y}^T \mathbf{y} - \mathbf{b}_1^T \mathbf{X}_1^T \mathbf{y})}{N - p}$$

will have the central distribution $F(p - q, N - p)$ if H_0 is correct or F

Table 6.1 *Analysis of Variance table.*

Source of variance	Degrees of freedom	Sum of squares	Mean square
Model with β_0	q	$\mathbf{b}_0^T \mathbf{X}_0^T \mathbf{y}$	
Improvement due to model with β_1	$p - q$	$\mathbf{b}_1^T \mathbf{X}_1^T \mathbf{y} - \mathbf{b}_0^T \mathbf{X}_0^T \mathbf{y}$	$\frac{\mathbf{b}_1^T \mathbf{X}_1^T \mathbf{y} - \mathbf{b}_0^T \mathbf{X}_0^T \mathbf{y}}{p - q}$
Residual	$N - p$	$\mathbf{y}^T \mathbf{y} - \mathbf{b}_1^T \mathbf{X}_1^T \mathbf{y}$	$\frac{\mathbf{y}^T \mathbf{y} - \mathbf{b}_1^T \mathbf{X}_1^T \mathbf{y}}{N - p}$
Total	N	$\mathbf{y}^T \mathbf{y}$	

14

Data: Birthweight and gestational age (file="Table.2.3.txt")

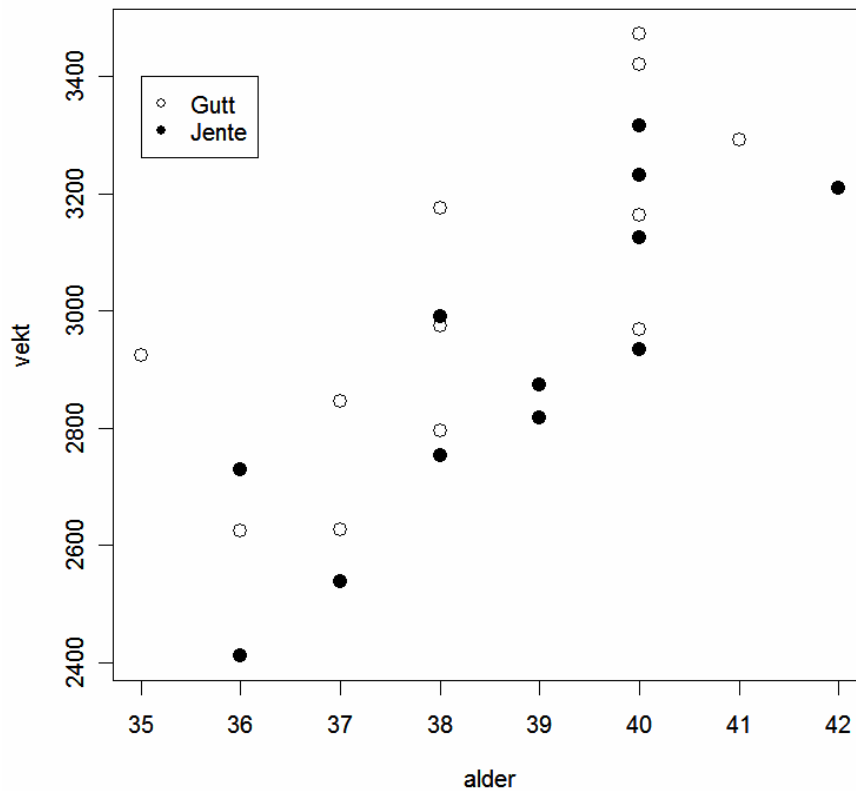
alder	vekt	kjonn
40	2968	1
38	2795	1
40	3163	1
35	2925	1
36	2625	1
37	2847	1
41	3292	1
40	3473	1
37	2628	1
38	3176	1
40	3421	1
38	2975	1
40	3317	2
36	2729	2
40	2935	2
38	2754	2
42	3210	2
39	2817	2
40	3126	2
37	2539	2
36	2412	2
38	2991	2
39	2875	2
40	3231	2

15

R-code for Example: Birthweight and gestational age

```
> Birth <- read.table(file="Table.2.3.txt",header=T)
> Birth <- data.frame(Birth)
> Birth
  alder vekt kjonn
1     40 2968     1
2     38 2795     1
3     40 3163     1
....
22    38 2991     2
23    39 2875     2
24    40 3231     2
> sapply(Birth,class)
  alder      vekt      kjonn
"integer" "integer" "integer"
> Birth <- transform(Birth,kjonn=factor(kjonn,labels =
  c("Gutt","Jente")))
> plot(vekt ~ alder, data=Birth,pch=c(1,16)[kjonn],cex=1.5)
> legend(35,3400,legend=c("Gutt","Jente"),pch=c(1,16))
> M1 <- glm(vekt ~ kjonn + alder + kjonn:alder, data = Birth)
> summary(M1)
```

16



17

```
Call:
glm(formula = vekt ~ kjonn + alder + kjonn:alder, data = Birth)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-246.69	-138.11	-39.13	176.57	274.28

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1268.67	1114.64	-1.138	0.268492
kjonnJente	-872.99	1611.33	-0.542	0.593952
alder	111.98	29.05	3.855	0.000986 ***
kjonnJente:alder	18.42	41.76	0.441	0.663893

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

(Dispersion parameter for gaussian family taken to be 32621.23)

Null deviance: 1829873 on 23 degrees of freedom
 Residual deviance: 652425 on 20 degrees of freedom
 AIC: 323.16

Number of Fisher Scoring iterations: 2

18

```

> confint(M1)
Waiting for profiling to be done...
              2.5 %      97.5 %
(Intercept)  -3453.32354  915.9787
kjonnnJente  -4031.14470  2285.1562
alder        55.05424   168.9113
kjonnnJente:alder -63.42266  100.2571
> param <- coef(summary(M1))
> ci <- confint(M1)
> u <- cbind(param[, "Estimate"], ci)
> colnames(u) <- c("Estimat", "2.5%", "97.5%")
> u
              Estimat          2.5%          97.5%
(Intercept) -1268.67241 -3453.32354  915.9787
kjonnnJente  -872.99425 -4031.14470  2285.1562
alder        111.98276   55.05424   168.9113
kjonnnJente:alder 18.41724  -63.42266  100.2571

```

19

Connection to model

$Y = \text{f\oedselsvekt}$
 $x_1 = \begin{cases} 0 & \text{hvis gutt} \\ 1 & \text{hvis jente} \end{cases}$
 $x_2 = \text{svangerskapsalder}$

$$E(Y) = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$$

(Intercept)	μ	Intercept of "Gutt"
kjonnnJente	β_1	Difference in intercept "Jente" and "Gutt"
alder	β_2	slope for "Gutt"
kjonnnJente:alder	β_{12}	Difference in slope for "Jente" and "Gutt"

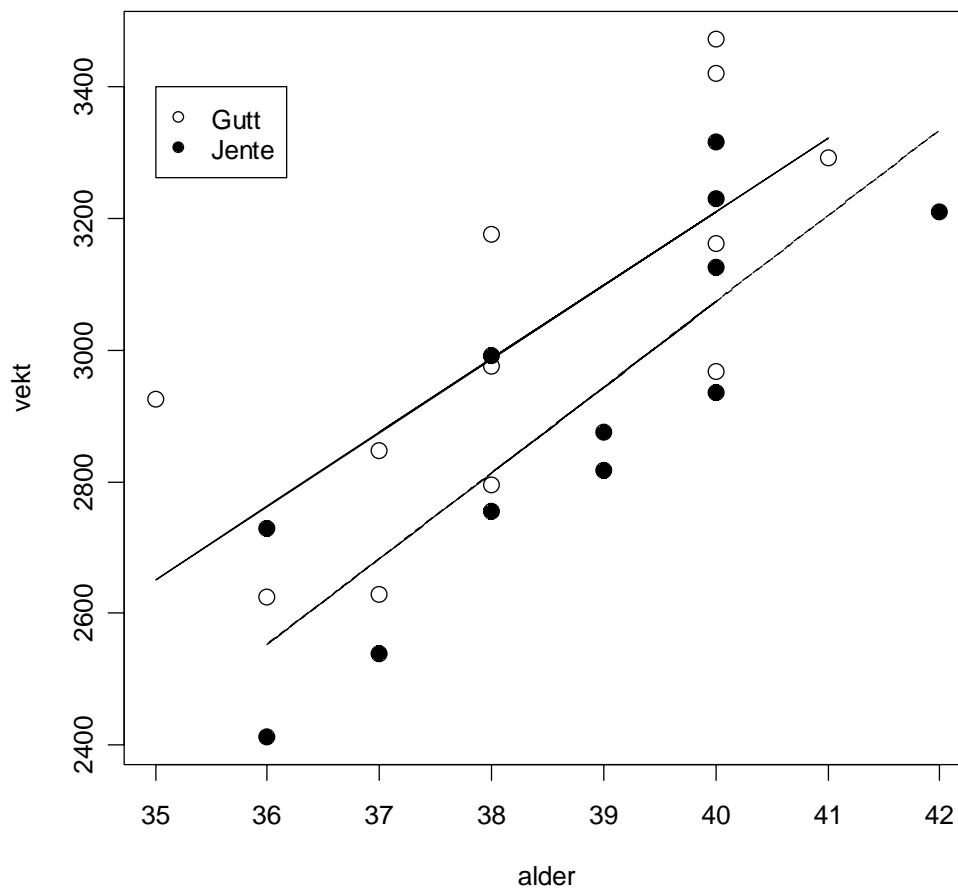
20

```

> Birth$fit <- predict(M1)
> lines(fit ~ alder, data=subset(Birth, kjonn=="Gutt"),lty=1)
>
> lines(fit ~ alder, data=subset(Birth, kjonn=="Jente"),lty=2)
> Birth
  alder vekt kjonn      fit
1     40 2968  Gutt 3210.638
2     38 2795  Gutt 2986.672
3     40 3163  Gutt 3210.638
4     35 2925  Gutt 2650.724
...
21    36 2412 Jente 2552.733
22    38 2991 Jente 2813.533
23    39 2875 Jente 2943.933
24    40 3231 Jente 3074.333

```

21



22

Analysis of deviance

Comparison of models

$$Y = \text{fødselsvekt}$$
$$x_1 = \begin{cases} 0 & \text{hvis gutt} \\ 1 & \text{hvis jente} \end{cases}$$
$$x_2 = \text{svangerskapsalder}$$

Model 1: $E(Y) = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$

Model 0: $E(Y) = \mu + \beta_1 x_1 + \beta_2 x_2$

```
> M0 <- glm(vekt ~ kjonn + alder, data=Birth)
> anova(M0,M1,test = "F")
Analysis of Deviance Table
```

Model 1: vekt ~ kjonn + alder

Model 2: vekt ~ kjonn + alder + kjonn:alder

	Resid. Df	Resid. Dev	Df	Deviance	F	Pr(>F)
1	21	658771				
2	20	652425	1	6346	0.1945	0.6639

23

Sequential ANOVA table

```
> anova(M1,test="F")
Analysis of Deviance Table
```

Model: gaussian, link: identity

Response: vekt

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	F	Pr(>F)
NULL			23	1829873		
kjonn	1	76163	22	1753711	2.3348	0.1422
alder	1	1094940	21	658771	33.5653	1.142e-05 ***
kjonn:alder	1	6346	20	652425	0.1945	0.6639

Computation of F:

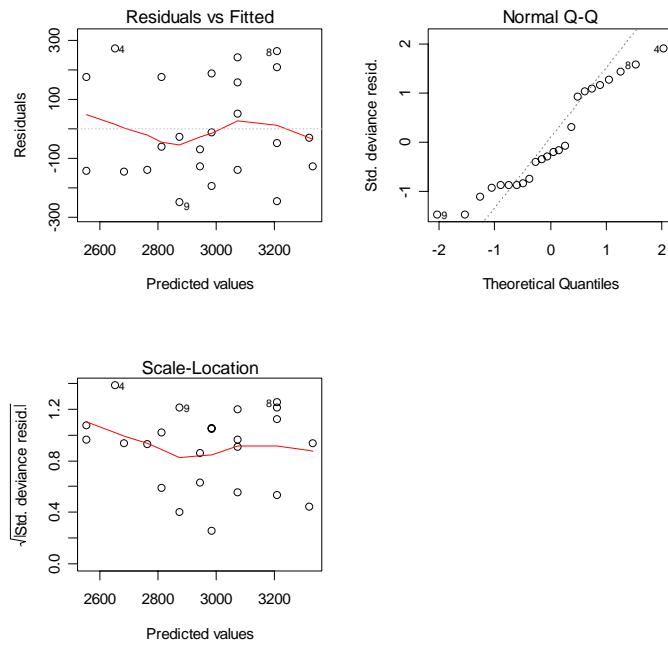
$$F(\text{kjonn}) = \frac{\frac{76163}{1}}{\frac{652425}{20}}$$

Thus: Denominator of F always uses the full model.

24

Residuals for Model 1

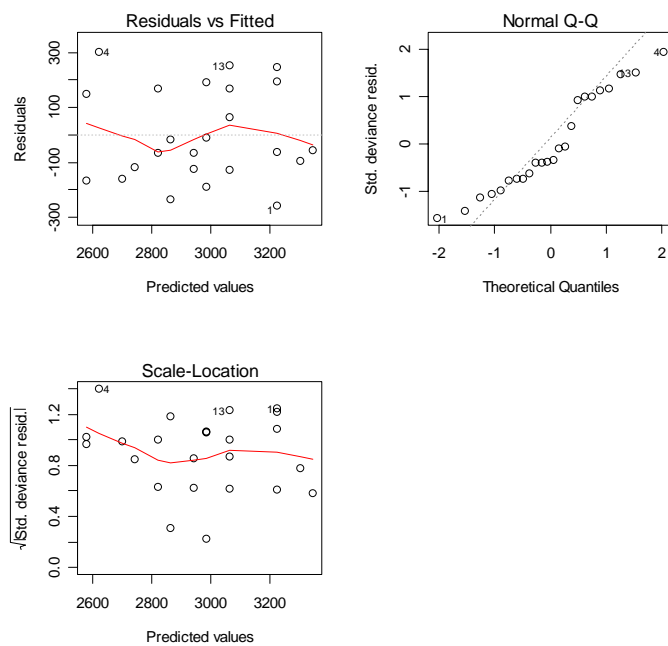
```
> par(mfrow=c(2,2))  
> plot(M1,which=c(1,2,3))
```



25

Residuals for Model 0

```
> par(mfrow=c(2,2))  
> plot(M0,which=c(1,2,3))
```



26