STAT3015/7030:
Generalised Linear Modelling
Analysis of Covariance

Bronwyn Loong

Semester 2 2014

References

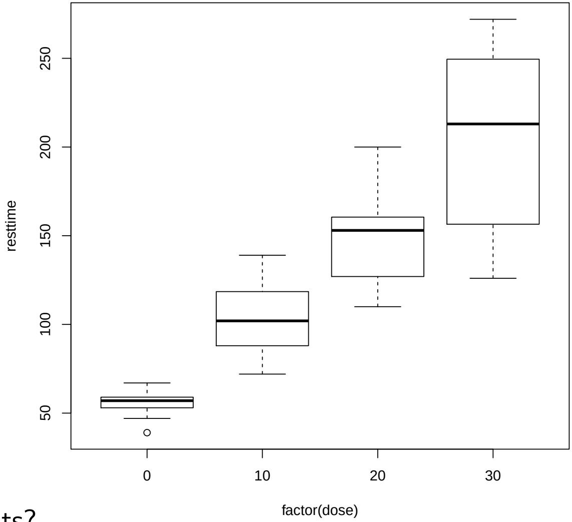
Ch 13 - Faraway, Linear models with R

Case Study I

An experiment was conducted to test the effects of the dose of a drug (in mg) on the induction of lethargy in rats, as measured by the number of minutes that the rat spends resting or sleeping in a 4 hour period. Rats of different ages (in months) are used.

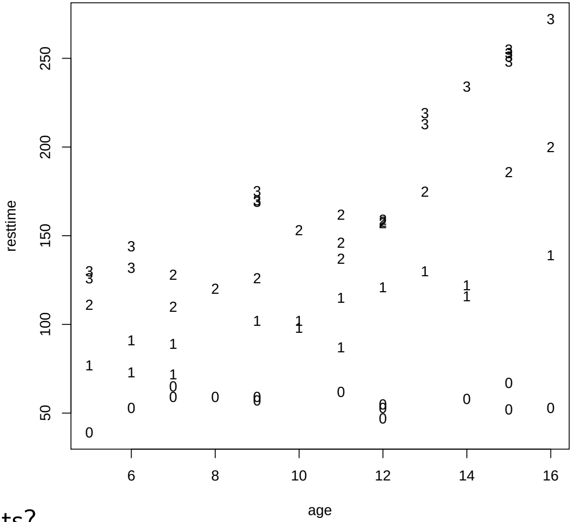
Q: Does the drug have a significant effect on rest-time after contolling for the different ages of the rats?

Case Study I



Comments?

Case Study I



Comments?

Case Study I

```
>by(lethargy,lethargy$dose,summary)
lethargy$dose: 0
      dose          age              resttime
 Min.   :0   Min.   : 5.00   Min.   :39.00
 1st Qu.:0   1st Qu.: 7.50   1st Qu.:53.00
 Median :0   Median :11.00   Median :57.00
 Mean   :0   Mean   :10.53   Mean   :55.87
 3rd Qu.:0   3rd Qu.:13.00   3rd Qu.:59.00
 Max.   :0   Max.   :16.00   Max.   :67.00

      dose           age              resttime
 Min.   :10   Min.   : 5.00   Min.   : 72.0
 1st Qu.:10   1st Qu.: 7.00   1st Qu.: 88.0
 Median :10   Median :10.00   Median :102.0
 Mean   :10   Mean   :10.07   Mean   :102.3
 3rd Qu.:10   3rd Qu.:12.50   3rd Qu.:118.5
 Max.   :10   Max.   :16.00   Max.   :139.0
```

Case Study I

```
>by(lethargy,lethargy$dose,summary)
lethargy$dose: 20
      dose            age              resttime
 Min.    :20    Min.    : 5.0    Min.    :110.0
 1st Qu.:20    1st Qu.: 8.5    1st Qu.:127.0
 Median :20    Median :11.0    Median :153.0
 Mean    :20    Mean    :10.6    Mean    :148.5
 3rd Qu.:20    3rd Qu.:12.0    3rd Qu.:160.5
 Max.    :20    Max.    :16.0    Max.    :200.0

lethargy$dose: 30
      dose            age              resttime
 Min.    :30    Min.    : 5.0    Min.    :126.0
 1st Qu.:30    1st Qu.: 7.5    1st Qu.:156.5
 Median :30    Median :13.0    Median :213.0
 Mean    :30    Mean    :11.0    Mean    :199.4
 3rd Qu.:30    3rd Qu.:15.0    3rd Qu.:249.5
 Max.    :30    Max.    :16.0    Max.    :272.0
```

Analysis of Covariance

Analysis of covariance (ANCOVA) refers to regression problems where there is a mixture of quantitative and qualitative predictors.

Suppose in an experiment, we wish to test the effect of a drug on blood pressure. If the drug and control groups differ with respect to age, we cannot consider this problem as a simple 2-sample t-test. We need to account for the effect of age as well.

Fortunately, we can use the regression framework we have already learnt to adjust for quantitative predictors, in determining the effect on the response of a qualitative treatment.

Analysis of Covariance

Let Y_{ij} denote the response at level i of factor α (i=1,...,I) for the j^{th} observed outcome at level i. Data is also available on a continuous predictor x.

$$y_{ij} = \mu + \alpha_i + \beta x_{ij} + \epsilon_{ij} \ OR \ y_{ij} = \mu_i + \beta x_{ij} + \epsilon_{ij}$$

$\epsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$, (and subject to constraints (see one-way ANOVA) to avoid over parameterisation)

Using indicator variables, we have

$$Y_{ij} = \beta_0 + \beta_1 z_{1,ij} + \beta_2 z_{2,ij} + ... + \beta_{(\mathrm{I}-1)} z_{(\mathrm{I}-1),ij} + \beta_{\mathrm{I}} x_{ij} + \epsilon_{ij}$$

(assuming $\alpha_1 = 0$)

Parameter estimation

Least squares estimation (and maximum likelihood) yield

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

.....we can show

$$\hat{\beta}_0 = \bar{Y}_1 - \hat{\beta}_{\mathrm{I}} \bar{x}_1$$

$$\hat{\beta}_1 = (\bar{Y}_2 - \bar{Y}_1) - \hat{\beta}_{\mathrm{I}} (\bar{x}_2 - \bar{x}_1)$$

.....

$$\hat{\beta}_{I-1} = (\bar{Y}_{\mathrm{I}} - \bar{Y}_1) - \hat{\beta}_{\mathrm{I}} (\bar{x}_{\mathrm{I}-1} - \bar{x}_1)$$

$$\hat{\beta}_{\mathrm{I}} = \frac{\sum_{i=1}^{\mathrm{I}} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(Y_{ij} - \bar{Y}_i)}{\sum_{i=1}^{\mathrm{I}} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}$$

Parameter estimation

So the treatment effect estimate is:

$$
\begin{aligned}
\hat{\mu}_i &= \hat{\beta}_0 + \hat{\beta}_{i-1} \\
&= \bar{Y}_1 - \hat{\beta}_{\mathrm{I}}\bar{x}_1 + (\bar{Y}_i - \bar{Y}_1) - \hat{\beta}_{\mathrm{I}}(\bar{x}_i - \bar{x}_1) \\
&= \bar{Y}_i - \hat{\beta}_{\mathrm{I}}\bar{x}_i \text{ for } i = 2, ..\mathrm{I}
\end{aligned}
$$

Also

$$Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

$$Var(c^T \hat{\beta}) = \sigma^2 c^T (X^T X)^{-1} c$$

Case Study I

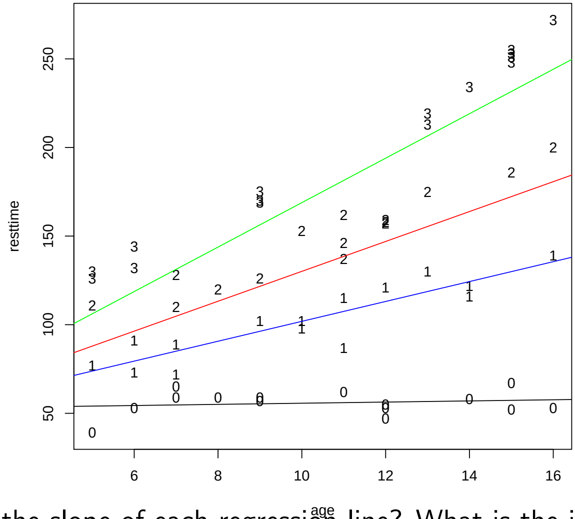
ANCOVA can disentangle the effects of age and dose on resttime.

```
> g <- lm(resttime ~ age+factor(dose)+age:factor(dose),lethargy)
> summary(g)

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)         52.4461     6.5917   7.956 1.49e-10 ***
age                  0.3247     0.5952   0.546    0.588    
factor(dose)10      -6.6609     9.3430  -0.713    0.479    
factor(dose)20       6.6850    10.1200   0.661    0.512    
factor(dose)30       8.9955     8.8873   1.012    0.316    
age:factor(dose)10   5.2860     0.8636   6.121 1.24e-07 ***
age:factor(dose)20   8.1094     0.9176   8.837 6.20e-12 ***
age:factor(dose)30  12.2169     0.7833  15.596  < 2e-16 ***

Residual standard error: 7.89 on 52 degrees of freedom
Multiple R-squared: 0.9857, Adjusted R-squared: 0.9837
F-statistic: 510.6 on 7 and 52 DF,  p-value: < 2.2e-16
```

Case Study I



What is the slope of each regression line? What is the intercept of each regression line?

Case Study I

(Compare the ANCOVA model fit to the one-way ANOVA model fit)

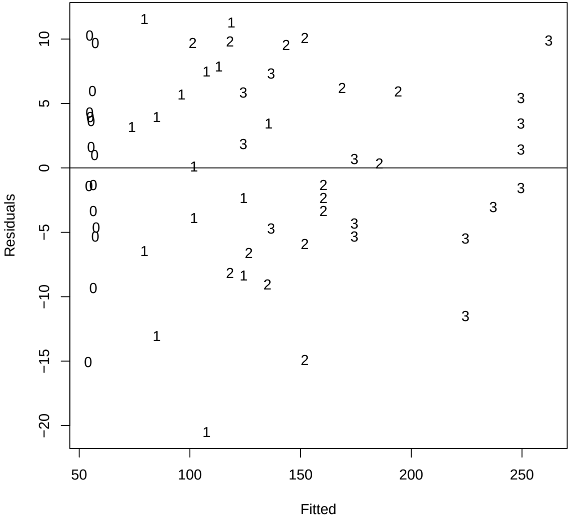
```
> h<-lm(resttime~factor(dose),lethargy)
> summary(h)

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       55.867      8.098   6.899 5.09e-09 ***
factor(dose)10    46.400     11.452   4.052 0.000158 ***
factor(dose)20    92.667     11.452   8.091 5.47e-11 ***
factor(dose)30   143.533     11.452  12.533  < 2e-16 ***

Residual standard error: 31.36 on 56 degrees of freedom
Multiple R-squared: 0.756, Adjusted R-squared: 0.7429
F-statistic: 57.82 on 3 and 56 DF,  p-value: < 2.2e-16
```

Case Study I

Residual diagnostic plot



Case Study I

Test a linear combination of regression coefficients. Example, suppose we want to test if the effect of age on resttime at does=30mg is 3 times the effect of age on resttime at dose =10mg

$$H_0 : 3(\beta_{age} + \beta_{age,dose=10}) - (\beta_{age} + \beta_{age,dose=30}) = 0$$

```
> h<-c(0,2,0,0,0,3,0,-1)
> Xmat <- model.matrix(g)
> XtXi <- solve(t(Xmat)%*%Xmat)
> est <- t(h)%*%coefficients(g)
> sd<-summary(g)$sigma*sqrt(t(h)%*%XtXi%*%h)
> upper <- est + (qt(0.975,df.residual(g))*sd)
>  lower <- est - (qt(0.975,df.residual(g))*sd)
>  c(lower,est,upper)
[1] 0.3869416 4.2905481 8.1941546
```

Confidence interval does not contain zero $\rightarrow$ reject null $\rightarrow$ the effect age on resttime does not increase linearly with dose level.