

Case Study 6: General linear hypothesis tests:

Toxicology experiment

Richard Wilkinson

The data

In a toxicology experiment, 28 mice were randomly assigned to one of six different treatment groups and a control group. The control group received no treatment. After the treatment period the liver weight of each mouse was measured. Note that the response is a continuous quantity, but the regressor is a discrete factor with 7 levels.

```
filepath <- "https://www.maths.nottingham.ac.uk/personal/pmzrdw/ToxicologyData.txt"
```

```
# Download the data from the internet
```

```
download.file(filepath, destfile = "ToxicologyData.txt", method = "curl")
```

```
ToxData <- read.table(file='ToxicologyData.txt', header=TRUE)
```

```
str(ToxData) # look at the data structure
```

```
## 'data.frame': 28 obs. of 2 variables:
```

```
## $ Weight : num 89.8 93.8 88.4 112.6 84.4 ...
```

```
## $ Treatment: Factor w/ 7 levels "Control","T1",...: 1 1 1 1 2 2 2 2 3 3 ...
```

```
# Don't worry about these commands - they just allow us to see the data
```

```
# structured in a nice way
```

```
library(reshape2)
```

```
ToxData$mouse = rep(1:4,7)
```

```
dcast(ToxData, mouse~Treatment, value.var = "Weight")
```

```
##   mouse Control   T1   T2   T3   T4   T5   T6
## 1      1   89.8  84.4 64.4 75.2 88.4 56.4 65.6
## 2      2   93.8 116.0 79.8 62.4 90.2 83.2 79.4
## 3      3   88.4  84.0 88.0 62.4 73.2 90.4 65.6
## 4      4  112.6  68.6 69.4 73.8 87.8 85.6 70.2
```

```
ToxData <- subset(ToxData, select = -c(mouse) )
```

- Response = organ weight (continuous).
- Input = treatment (discrete - a factor with 7 levels)

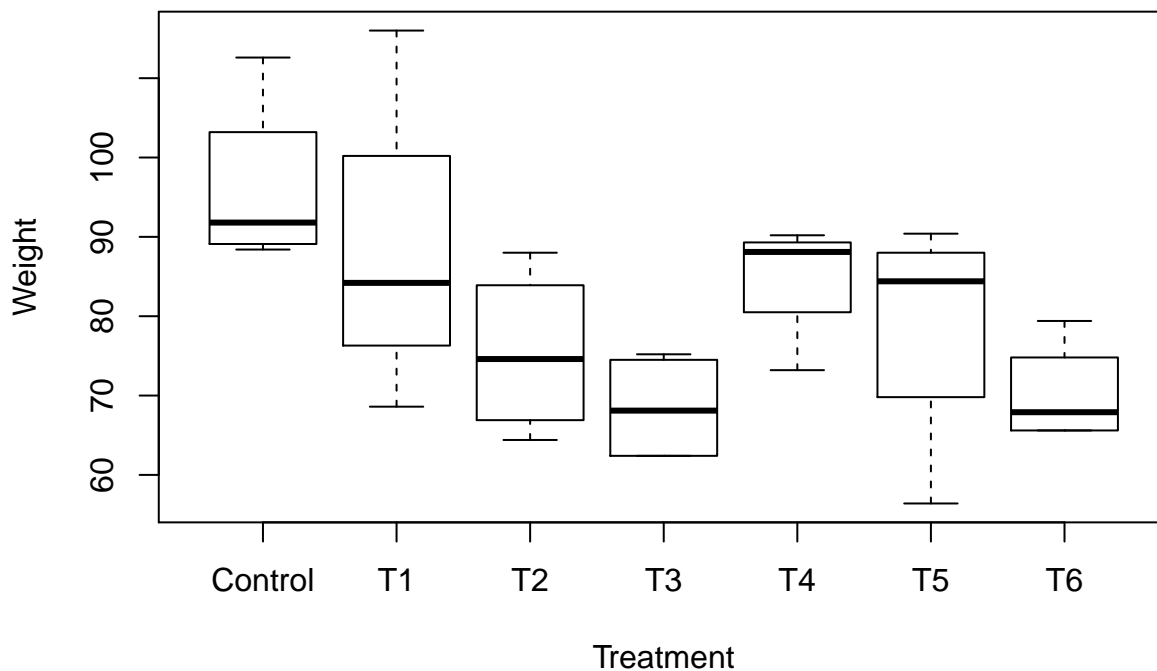
What questions might we want to answer about this data set?

- Is there evidence that any of the treatments is having an effect on liver weight?
- Is treatment 1 better than treatment 2?
- Are treatments 1-6 all equally effective?
- etc.

Visualise the data

Check the data and plot it

```
plot(Weight ~ Treatment, data=ToxData) # Plot the data
```



This informally suggests that the treatments are having an effect. But with only 4 mice in each group, are the differences statistically significant?

Parameterisation

How can we model these data? A simple model is to assume that each treatment group has a mean weight, but that this mean might be different for different treatment groups.

Model 1: Let y_{ij} be the j^{th} mouse weight in treatment group i .

$$y_{ij} = \begin{cases} \mu_C + \epsilon_{ij}, & \text{if } i = \text{'Control'} \\ \mu_1 + \epsilon_{ij}, & \text{if } i = 1 \\ \mu_2 + \epsilon_{ij}, & \text{if } i = 2 \\ \mu_3 + \epsilon_{ij}, & \text{if } i = 3 \\ \mu_4 + \epsilon_{ij}, & \text{if } i = 4 \\ \mu_5 + \epsilon_{ij}, & \text{if } i = 5 \\ \mu_6 + \epsilon_{ij}, & \text{if } i = 6 \end{cases}$$

This model allows the mean responses μ_i 's to be different for different treatments. This is an Analysis of Variance (ANOVA) model.

What is Z here? We need to decide on a way to stack the y_{ij} into an observation vector. If we set

$$y = \begin{bmatrix} 89.8 \\ 93.8 \\ 88.4 \\ 112.6 \\ 84.4 \\ \vdots \\ 70.2 \end{bmatrix}, \quad x = \begin{bmatrix} C \\ C \\ C \\ C \\ 1 \\ \vdots \\ 6 \end{bmatrix}, \quad \beta = \begin{bmatrix} \mu_C \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_6 \end{bmatrix}$$

then the model in matrix form is

$$y = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ & 1 & 0 & 0 & \vdots \\ & 1 & 0 & 0 & \\ & 0 & 1 & 0 & \\ & 0 & 1 & 0 & \\ & 0 & 1 & 0 & \\ & 0 & 1 & 0 & \\ & 0 & 0 & 1 & \vdots \\ & 0 & 0 & 1 & \dots & 0 \\ & \vdots & & & \vdots \\ & 0 & 0 & 0 & \dots & 1 \\ & 0 & 0 & 0 & \dots & 1 \\ & 0 & 0 & 0 & \dots & 1 \\ & 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu_C \\ \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{bmatrix} + \epsilon.$$

So, Z is a (28×7) matrix of 0's and 1's (4 replicates of 7 treatments).

One can show that the least squares estimators are

$$\hat{\mu}_i = \bar{y}_{i\cdot} = \frac{1}{4} \sum_j y_{ij}$$

(the mean of the observations from group i).

Model 2: It is possible to parameterise this model in a different way. Consider:

$$y_{ij} = \begin{cases} \mu + \epsilon_{iC} & i = \text{'Control'} \\ \mu + \alpha_j + \epsilon_{ij} & i = 1, \dots, 6. \end{cases}$$

This is still a 7 parameter model with parameters $\mu, \alpha_1, \alpha_2, \dots, \alpha_6$. Now

$$Z = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

We can show that $\hat{\mu} = \bar{y}_{+C}$ $\hat{\alpha}_j = \bar{y}_{+j} - \bar{y}_{+C}$, and thus the residuals are $y_{ij} - \bar{y}_{+j}$, the same as in model 1. In this case model 2 is simply a *reparametrisation* of model 1, i.e.,

$$\begin{aligned}\mu &= \mu_C \\ \mu + \alpha_j &= \mu_j.\end{aligned}$$

Fitting the models in R

We can fit model 1 in R using the command

```
fit <- lm(Weight ~ Treatment - 1, data = ToxData)
fit

##
## Call:
## lm(formula = Weight ~ Treatment - 1, data = ToxData)
##
## Coefficients:
## TreatmentControl      TreatmentT1      TreatmentT2      TreatmentT3
##           96.2           88.2           75.4           68.4
##      TreatmentT4      TreatmentT5      TreatmentT6
##           84.9           78.9           70.2
```

The formula used contains a -1 as otherwise R automatically assumes an intercept term. If we don't use a -1 then we fit model 2 instead. Try

```
fit2 <- lm(Weight ~ Treatment, data=ToxData)
```

Note that we can see the design matrix used by lm by typing

```
model.matrix(fit)
```

A simple ANOVA test

Lets consider testing the two models

$$\begin{aligned}M_0 : y_{ij} &= \mu + \epsilon_{ij} \\ \text{vs } M_1 : y_{ij} &= \mu + \alpha_i + \epsilon_{ij}\end{aligned}$$

which is equivalent to testing

$$\begin{aligned}H_0 : \alpha_1 &= \dots = \alpha_6 = 0 \\ \text{vs } H_1 : \alpha_i &\text{ arbitrary}\end{aligned}$$

We use an F-test to do this ($q = 6$ as we must constrain 6 parameters to go from M_1 to M_0). The anova command is the easiest way to do this in R (although you could use linearHypothesis if you wished).

```
fit <- lm(Weight ~ Treatment-1, data=ToxData)
fit0 <- lm(Weight~1, data=ToxData) # fit the null model
anova(fit0, fit)
```

```
## Analysis of Variance Table
##
## Model 1: Weight ~ 1
## Model 2: Weight ~ Treatment - 1
##   Res.Df  RSS Df Sum of Sq    F Pr(>F)
## 1      27 5479
## 2      21 3063  6      2416 2.76 0.039 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

So there is some evidence (p-value = 0.039) to reject H_0 in favour of H_1 .

Can you manually calculate this F statistic?

Single parameter hypothesis test

Suppose we want to test $H_0 : \alpha_1 = 0$ vs $H_1 : \alpha_1$ arbitrary.

We can use a simple t-test here (we can use an F-test, but as $q = 1$, and $F_{1,n-p} = (t_{n-p})^2$ they are equivalent). The test statistic is

$$T = \frac{\mathbf{a}^T \hat{\beta} - \mathbf{a}^T \beta}{s \sqrt{\mathbf{a}^T (Z^T Z)^{-1} \mathbf{a}}}$$

with $\mathbf{a}^T = (0, 1, 0, 0, 0, 0, 0)$.

Our T statistic is therefore,

$$T = \frac{\hat{\alpha}_1}{\text{std.error}(\hat{\alpha}_1)},$$

where,

$$\text{std.error}(\hat{\alpha}_1) = s \sqrt{d_{ii}}.$$

R tells us that $s = 12.076$, $\hat{\alpha}_1 = -7.9$, and $d_{ii} = 0.5$.

```
coef(fit2)
```

```
## (Intercept) TreatmentT1 TreatmentT2 TreatmentT3 TreatmentT4 TreatmentT5
##      96.15      -7.90      -20.75      -27.70      -11.25      -17.25
## TreatmentT6
##      -25.95
```

```
fit.sum <- summary(fit2)
fit.sum$sigma
```

```
## [1] 12.08
```

```
solve(t(model.matrix(fit2))%*%model.matrix(fit2))
```

```
##           (Intercept) TreatmentT1 TreatmentT2 TreatmentT3 TreatmentT4
## (Intercept)      0.25      -0.25      -0.25      -0.25      -0.25
## TreatmentT1     -0.25       0.50       0.25       0.25       0.25
## TreatmentT2     -0.25       0.25       0.50       0.25       0.25
## TreatmentT3     -0.25       0.25       0.25       0.50       0.25
## TreatmentT4     -0.25       0.25       0.25       0.25       0.50
## TreatmentT5     -0.25       0.25       0.25       0.25       0.25
## TreatmentT6     -0.25       0.25       0.25       0.25       0.25
##           TreatmentT5 TreatmentT6
## (Intercept)     -0.25     -0.25
## TreatmentT1       0.25       0.25
## TreatmentT2       0.25       0.25
## TreatmentT3       0.25       0.25
## TreatmentT4       0.25       0.25
## TreatmentT5       0.50       0.25
## TreatmentT6       0.25       0.50
```

Hence,

$$T = \frac{-7.9}{12.0763 \times \sqrt{0.5}} = -0.925.$$

This is not significant, as $t_{21}(0.975) = 2.0796$.

```
qt(0.975, df=21)
```

```
## [1] 2.08
```

NB: there are $28 - 7 = 21$ degrees of freedom for this hypothesis test. Note that all of the 7 groups are contributing to the estimation of σ^2 , whereas in the two-sample unpaired t -test (comparing group 1 with the control group) only two groups would be used, and so the number of degrees of freedom would be $n_5 + n_6 - 2 = 6$.

Note that the details above are just for illustrative purposes. To perform this test you can simply type

```
summary(fit2)
```

```
##
## Call:
## lm(formula = Weight ~ Treatment, data = ToxData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.50  -6.05  -1.18   5.69  27.75
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)    96.15      6.04   15.92 3.4e-13 ***
## TreatmentT1    -7.90      8.54   -0.93  0.3654
## TreatmentT2   -20.75      8.54   -2.43  0.0242 *
## TreatmentT3   -27.70      8.54   -3.24  0.0039 **
## TreatmentT4   -11.25      8.54   -1.32  0.2019
## TreatmentT5   -17.25      8.54   -2.02  0.0563 .
## TreatmentT6   -25.95      8.54   -3.04  0.0062 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.1 on 21 degrees of freedom
## Multiple R-squared:  0.441, Adjusted R-squared:  0.281
## F-statistic: 2.76 on 6 and 21 DF, p-value: 0.0387
```

and read off the answer.

More complex General Linear Hypothesis Tests

Now lets consider the parameterization

$$y_{ij} = \mu_i + \epsilon_{ij},$$

for $i = C, 1, \dots, 6$, $j = 1, \dots, 4$ and $\beta = (\mu_C, \mu_1, \dots, \mu_6)^T$. It is easy to show that $Z^T Z = 4I_7$.

The general linear hypothesis test can test any hypothesis of the form $\mathbf{A}\beta = \mathbf{c}$.

We use the R command `linearHypothesis` in the package `car`. Read about this function in the manual pages (type `?linearHypothesis` in R).

Here are some examples:

Example 1 Is there any difference between treatments 3 and 6?

$$H_0 : \mu_3 = \mu_6$$

vs

$$H_1 : \beta \text{ arbitrary.}$$

Here $\mathbf{A} = [0, 0, 0, 1, 0, 0, -1]_{1 \times 7}$ and $\mathbf{c} = [0]_{1 \times 1}$. This test can equivalently be done with a t -test as there is only one constraint (as in the previous section).

$(Z^T Z)^{-1} = 0.25I_7$ then after some multiplication...

$$\begin{aligned} Q_H - Q &= (\mathbf{A}\hat{\beta} - \mathbf{c})^T \left[\mathbf{A} (Z^T Z)^{-1} \mathbf{A}^T \right]^{-1} (\mathbf{A}\hat{\beta} - \mathbf{c}) \\ &= -1.75 \times \left(\frac{2}{4} \right)^{-1} \times (-1.75) \\ &= 6.125. \end{aligned}$$

but, $Q = 3062.6$

```
deviance(fit)
```

```
## [1] 3063
```

and so

$$F = \frac{(Q_H - Q)}{Q/21} = \frac{6.126}{3062.6/21} = 0.042,$$

which is not significant as

$$F_{1,21}(0.95) = 4.32$$

and so there is no evidence to reject the hypothesis that $\mu_3 = \mu_6$.

```
qf(0.95, df1=1, df2=21)
```

```
## [1] 4.325
```

In R, we simply do the following:

```
library(car)
A <- c(0,0,0,1,0,0,-1)
c = 0
hyp1 <- linearHypothesis(fit, A, 0)
hyp1
```

```
## Linear hypothesis test
##
## Hypothesis:
## TreatmentT3 - TreatmentT6 = 0
##
## Model 1: restricted model
## Model 2: Weight ~ Treatment - 1
##
##   Res.Df  RSS Df Sum of Sq    F Pr(>F)
## 1      22 3069
## 2      21 3063  1      6.12 0.04  0.84
```

Make sure you know what all the numbers mean and how to calculate them yourself.

Example 2 $H_0 : \mu_3 = \mu_6 = 70$

vs

$H_1 : \beta$ arbitrary.

This time,

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{2 \times 7} \text{ and } \mathbf{c} = \begin{bmatrix} 70 \\ 70 \end{bmatrix}_{2 \times 1}$$

```
A <- rbind(c(0,0,0,1,0,0,0),c(0,0,0,0,0,0,1))
c=c(70,70)
hyp2 <- linearHypothesis(fit, A,c)
hyp2
```

```
## Linear hypothesis test
##
## Hypothesis:
```



```
## TreatmentT3 = 70
## TreatmentT6 = 70
##
## Model 1: restricted model
## Model 2: Weight ~ Treatment - 1
##
##   Res.Df  RSS Df Sum of Sq    F Pr(>F)
## 1      23 3072
## 2      21 3063  2      9.77 0.03  0.97
```

Not significant at the 5% level as the p-value > 0.05

Example 3 $H_0 : \mu_3 + \mu_6 = 2\mu_1$, $\mu_1 = 80$ and $\mu_2 = 60$

vs

$H_1 : \beta$ arbitrary.

Now

$$\mathbf{A} = \begin{bmatrix} 0 & 2 & 0 & -1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}_{3 \times 7} \quad \text{and } \mathbf{c} = \begin{bmatrix} 0 \\ 80 \\ 60 \end{bmatrix}_{3 \times 1}$$

```
A <- rbind(c(0,-2,0,1,0,0,1),c(0,1,0,0,0,0,0), c(0,0,1,0,0,0,0))
c=c(0, 80,60)
hyp3 <- linearHypothesis(fit, A, c)
hyp3
```

```
## Linear hypothesis test
##
## Hypothesis:
## - 2 TreatmentT1 + TreatmentT3 + TreatmentT6 = 0
## TreatmentT1 = 80
## TreatmentT2 = 60
##
## Model 1: restricted model
## Model 2: Weight ~ Treatment - 1
##
##   Res.Df  RSS Df Sum of Sq    F Pr(>F)
## 1      24 5195
## 2      21 3063  3      2133 4.87  0.01 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```