MA50259: Statistical Design of Investigations

Lab sheet 7: Randomized complete block design

In this practical you will learn the basics of randomized complete block design.

Rats drug experiment data

Consider the following data from an experiment to determine the effect of the drug *d-amphetamine sulfate* on the behavior of rats. The behavior under study was the rate at which water-deprived rats pressed a lever to obtain water.

- **Response:** The lever press rate defined as the number of lever presses divided by the elapsed time of the session.
- **Treatment factor levels:** Five different dosages of the drug in milligrams per kilogram of body weight, including a control dosage consisting of saline solution. Levels are $\{0, 0.5, 1, 1.5, 2\}$.
- An experiment, or run, consisted of injecting a rat with a drug dosage, and after one hour the rat would receive water each time after a second lever was pressed.
- **Experimental unit:** The state of a single rat during one experiment or run, since an individual rat could be used in many experiments by repeatedly injecting it with different doses of the drug (after an appropriate washout period) and by observing the lever pressing behavior.
- **Blocking** Since there was wide variability in the lever pressing rate between rats, an RCB design was used, and a rat represented the blocking factor.
- **Randomization** Each rat received all five doses in a random order with an appropriate washout period in between.

The data can be downloaded as follows:

```r
library(tidyverse)
url_data<-"http://people.bath.ac.uk/kai21/MA50259/Data/drug_rats.txt"
rats<-url_data %>% read.table(header=TRUE)
rats
```

```
##    rat dose rate
## 1    1  0.0 0.60
## 2    1  0.5 0.80
## 3    1  1.0 0.82
## 4    1  1.5 0.81
## 5    1  2.0 0.50
## 6    2  0.0 0.51
## 7    2  0.5 0.61
## 8    2  1.0 0.79
## 9    2  1.5 0.78
## 10   2  2.0 0.77
## 11   3  0.0 0.62
## 12   3  0.5 0.82
## 13   3  1.0 0.83
## 14   3  1.5 0.80
## 15   3  2.0 0.52
## 16   4  0.0 0.60
## 17   4  0.5 0.95
## 18   4  1.0 0.91
## 19   4  1.5 0.95
## 20   4  2.0 0.70
## 21   5  0.0 0.92
## 22   5  0.5 0.82
## 23   5  1.0 1.04
## 24   5  1.5 1.13
## 25   5  2.0 1.03
## 26   6  0.0 0.63
## 27   6  0.5 0.93
## 28   6  1.0 1.02
## 29   6  1.5 0.96
## 30   6  2.0 0.63
## 31   7  0.0 0.84
## 32   7  0.5 0.74
## 33   7  1.0 0.98
## 34   7  1.5 0.98
## 35   7  2.0 1.00
## 36   8  0.0 0.96
## 37   8  0.5 1.24
## 38   8  1.0 1.27
## 39   8  1.5 1.20
## 40   8  2.0 1.06
## 41   9  0.0 1.01
## 42   9  0.5 1.23
## 43   9  1.0 1.30
## 44   9  1.5 1.25
## 45   9  2.0 1.24
## 46  10  0.0 0.95
## 47  10  0.5 1.20
## 48  10  1.0 1.18
## 49  10  1.5 1.23
## 50  10  2.0 1.05
```

```
rats$rat<-as.factor(rats$rat)
rats$dose<-as.factor(rats$dose)
```

1. Consider first the following linear model for the RCB design

$$y_{ij} = \mu + b_i + \tau_j + (b\tau)_{ij} + \epsilon_{ij}$$

where

- b_i is the effect of block $i \in \{1, \dots, b\}$

- $\tau_1, \ldots, \tau_t$ are the treatment effects
- $(b\tau)_{ij}$ are the effects of the interactions
- All $\{\epsilon_{ij}\}$ are $N(0, \sigma^2)$ and independent

Verify that there are zero degrees of freedom for the error sum of squares ssE

```
mod1 <- lm(rate ~ rat*dose, data = rats )
anova(mod1)
```

```
## Warning in anova.lm(mod1): ANOVA F-tests on an essentially perfect fit are
## unreliable
```

```
## Analysis of Variance Table
##
## Response: rate
##           Df  Sum Sq  Mean Sq F value Pr(>F)
## rat        9 1.66846 0.185384
## dose       4 0.46021 0.115052
## rat:dose  36 0.30055 0.008349
## Residuals  0 0.00000
```

Solution: If we write the model in matrix form we have:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

and consider the full rank representation given by R via `model.matrix` as follows:

```
y<-rats$rate
X<-model.matrix(~rat*dose,rats)
dim(X)
```

```
## [1] 50 50
```

```
as.numeric(rankMatrix(X))
```

```
## [1] 50
```

Clearly, in this peculiar case, there is a unique solution $\widehat{\boldsymbol{\beta}}$ to the system

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}$$

so the residuals $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$ are exactly zero and therefore `ssE`$= \boldsymbol{r}^\top \boldsymbol{r} = 0$. The solution is given by

```
beta.est<-solve(X)%*%y
```

which is the same obtained by `lm`

```
S<-summary(mod1)
beta.lm<-coefficients(S)[,"Estimate"]
max(abs(beta.lm-beta.est))
```

```
## [1] 1.374225e-15
```

To see there are zero degrees of freedom we recall that the formula for the degrees of freedom associated with `ssE` is

$$\text{rank}(\boldsymbol{I}_n - \boldsymbol{H}_{\boldsymbol{X}}) = n - \text{rank}(\boldsymbol{H}_{\boldsymbol{X}}) = n - \text{rank}(\boldsymbol{X}) = 50 - 50 = 0$$

and therefore there are zero degrees of freedom for `ssE`.

The reason for this problem is there is only one replicate (r = 1) per cell if we see this design as a factorial design with 2 factors: one factor given by the blocking variable and the other being the treatment factor variable. The ANOVA sum of squares ssE is equal to zero and therefore there is no way to make F-tests on the treatment effects in the traditional way.

We would have obtained the same result if we use the less-than-full rank version of the model and then using a generalised inverse. This is because the associated matrix $\boldsymbol{X}$ will still be of rank 50 and the residuals would then be zero as is shown below:

```
inter<-interaction(rats$rat,rats$dose,sep=":")
X1<-model.matrix(~rat-1,rats)
X2<-model.matrix(~dose-1,rats)
X3<-model.matrix(~inter-1)
X<-cbind(1,X1,X2,X3)
rankMatrix(X)
```

```
## [1] 50
## attr(,"method")
## [1] "tolNorm2"
## attr(,"useGrad")
## [1] FALSE
## attr(,"tol")
## [1] 1.465494e-14
```

```
G<-ginv(t(X)%*%X)
residual<-y-X%*%G%*%t(X)%*%y
t(residual)
```

```
##                 1            2            3 4            5             6
## [1,] 4.32987e-15 1.887379e-15 1.554312e-15 0 2.775558e-16 -5.329071e-15
##                  7            8            9           10           11
## [1,] -3.330669e-16 3.330669e-16 2.220446e-15 2.220446e-15 3.885781e-15
##                 12            13           14            15            16
## [1,] -3.552714e-15 -1.776357e-15 9.992007e-16 -1.554312e-15 -6.661338e-16
##                17            18            19            20            21
## [1,] 3.885781e-15 -2.331468e-15 -1.110223e-15 -2.109424e-15 -2.775558e-15
##                 22            23            24            25           26
## [1,] -1.776357e-15 -2.664535e-15 -1.110223e-15 -3.330669e-15 3.330669e-16
##                 27           28            29           30            31
## [1,] -1.887379e-15 4.440892e-16 -1.776357e-15 2.664535e-15 -3.996803e-15
##                 32           33            34            35            36
## [1,] -4.218847e-15 9.992007e-16 -1.110223e-16 -1.776357e-15 -2.775558e-15
##                37            38           39            40           41
## [1,] -2.88658e-15 -2.220446e-15 -2.88658e-15 -4.440892e-16 2.664535e-15
##                 42            43 44            45           46           47
## [1,] -1.776357e-15 -3.330669e-15  0 -2.442491e-15 2.220446e-16 1.776357e-15
##                  48 49           50
## [1,] -2.220446e-16  0 3.330669e-15
```

2. Verify that the model model without interaction

$$y_{ij} = \mu + b_i + \tau_j + \epsilon_{ij}$$

is able to estimate the effect of ssE from the effect of the sum of squares due to the interaction. Also verify that we clearly reject the null hypothesis of equality of mean dose effects, that is

$$H_0 \, : \, \tau_0 = \tau_{0.5} = \tau_1 = \tau_{1.5} = \tau_2 = 0 \qquad \text{vs} \qquad H_a \, : \, \tau_i \neq 0 \quad \text{for some i } \in \{0, 0.5, 1, 1.5, 2\}$$

using $\alpha = 0.05$.

```
mod2 <- aov( rate ~ rat + dose, data = rats )
summary(mod2)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## rat          9 1.6685 0.18538   22.20 3.75e-12 ***
## dose         4 0.4602 0.11505   13.78 6.53e-07 ***
## Residuals   36 0.3006 0.00835                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that basically, the `ssE` is the same now as the sum of squares due to the interaction in the previous question (model with interactions) and therefore also has the same value and and the same $36 = (10-1)(5-1)$ degrees fo freedom.

3. Verify that we obtain a different (and wrong) conclusion if we use the model for a CRD, that is

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

By using this model, what assumption we are actually making about the randomization process and what we would be assuming experimental units were?

```
mod3 <- aov( rate ~ dose, data = rats )
summary(mod3)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)  
## dose         4 0.4602 0.11505   2.629 0.0466 *
## Residuals   45 1.9690 0.04376                 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

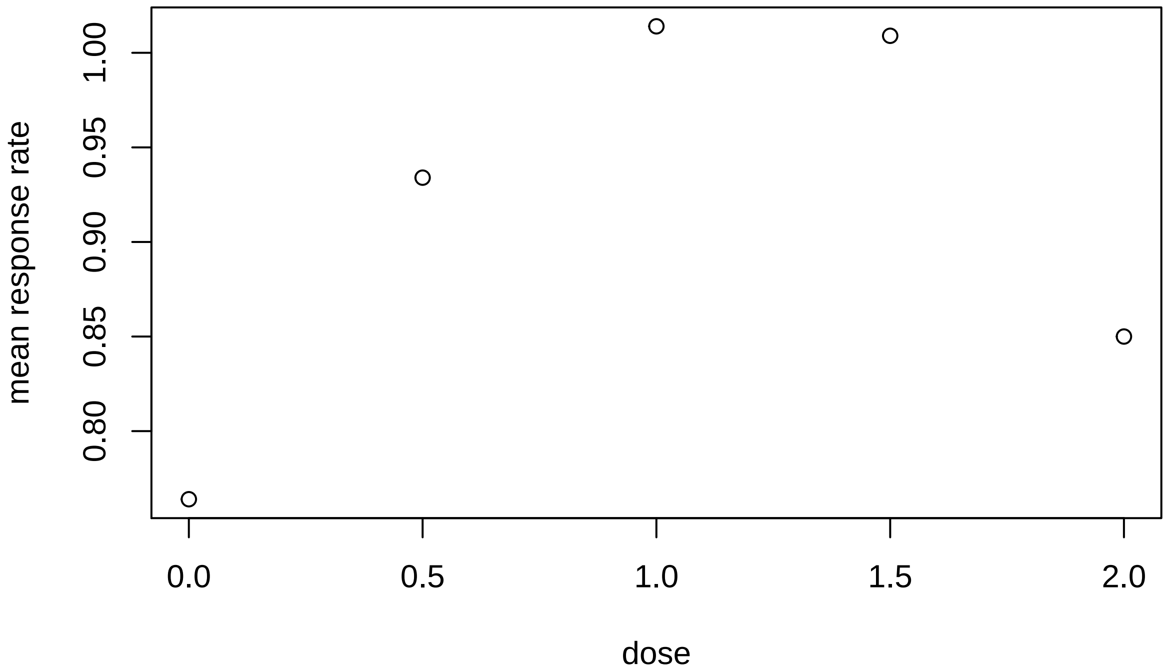
Solution: In this model we are assuming the experimental units are the same (rats in a particular run) but we are further assuming all of these are homogeneous even for different rats which does not seem tenable! We would also be assuming that the randomization was done by randomly allocating the whole set of 50 experimental units to the 5 levels of the treatment so that each level has 10 experimental units which is clearly not what has been done in this experiment where the treatment where allocated randomly withing each rat to the 5 levels of treatment! We can see the conclusion we still reject the null hypothesis (but only just!) if $\alpha = 0.05$ but we would have not rejected the null hypothesis had we set $\alpha = 0.01$ before carrying out the experiment! This obviously would have been the wrong conclusion! The main reason is that the estimated variance is equal to the 0.04376 using the CRD model which is basically 5 fimes larger than the estimated variance of 0.00835 under the correct RCB model. This demonstrates the effectiveness of blocking by rat in the experiment where blocking has reduced the estimated variance substantially!

4. The treatment variable (dose levels) can actually be treated as a quantitative variable instead of a factor. Plot the mean response effect (when averaging over all rats) of the dose levels and verify the effect is close to being quadratic.

```
dosage<-as.numeric(levels(rats$dose))
summary.means <- rats %>% group_by(dose) %>% summarize(mean=mean(rate))
summary.means
```

```
## # A tibble: 5 x 2
##   dose   mean
##   <fct> <dbl>
## 1 0     0.764
## 2 0.5   0.934
## 3 1     1.01
## 4 1.5   1.01
## 5 2     0.85
```

```
plot(summary.means$mean~dosage,summary.means,xlab="dose",ylab="mean response rate")
```



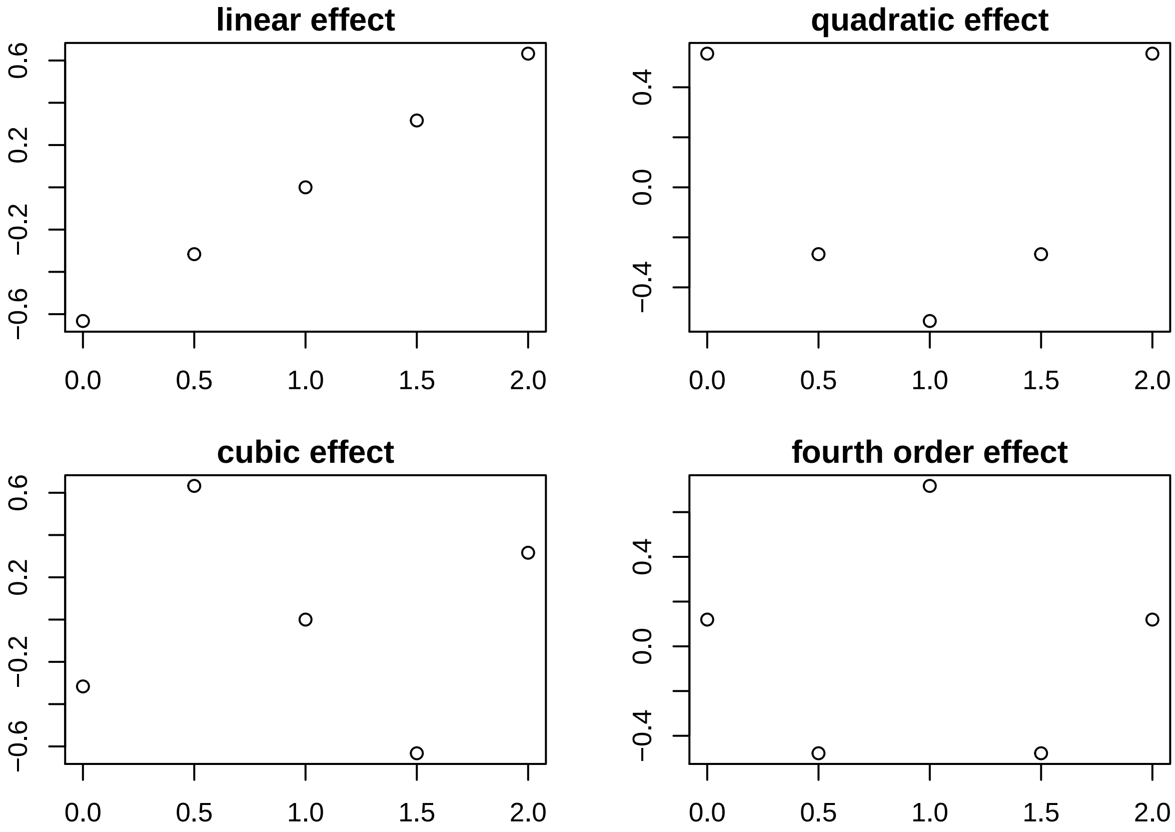
We can see that a quadratic relation (parabolic type openning downwards and with a maximum) between the dose level and the response rate seems quite plausible!

5. We can estimate the effect of the treatment effects by using the so-called orthogonal polynomial contrasts. The following code plots these contrasts against the dosage levels. The assumption is that the actual effect of the dosage on the response rate (after the mean level has been fitted) will be a linear combination of these 4 effects.

```
orth.contrasts<-contr.poly(5)
orth.contrasts
```

```
##              .L         .Q            .C         ^4
## [1,] -0.6324555  0.5345225 -3.162278e-01  0.1195229
## [2,] -0.3162278 -0.2672612  6.324555e-01 -0.4780914
## [3,]  0.0000000 -0.5345225 -4.095972e-16  0.7171372
## [4,]  0.3162278 -0.2672612 -6.324555e-01 -0.4780914
## [5,]  0.6324555  0.5345225  3.162278e-01  0.1195229
```

```
par(mfrow=c(2,2))
par(mar=c(3,3,1.5,1.5))
plot(dosage,orth.contrasts[,".L"],ylab="rate",main="linear effect")
plot(dosage,orth.contrasts[,".Q"],ylab="rate",main="quadratic effect")
plot(dosage,orth.contrasts[,".C"],ylab="rate",main="cubic effect")
plot(dosage,orth.contrasts[,"^4"],ylab="rate",main="fourth order effect")
```



We can see that the plots are defined around zero on the vertical (response rate) axis. This is because the (linear, quadratic, etc) effects are defined over and above the mean level defined by μ. This

6. Using the contrasts described above, fit the model

$$y_{ij} = \mu + b_i + \alpha_1\,\text{dose}_{lin} + \alpha_2\,\text{dose}_{quad} + \alpha_3\,\text{dose}_{cub} + \alpha_4\,\text{dose}_{fourt}$$

and obtain the corresponding ANOVA table using the following commands:

```
rats$dose.lin<-rep(orth.contrasts[,".L"],times=10)
rats$dose.quad<-rep(orth.contrasts[,".Q"],times=10)
rats$dose.cub<-rep(orth.contrasts[,".C"],times=10)
rats$dose.fourth<-rep(orth.contrasts[,"^4"],times=10)

mod3<-lm(rate~rat+dose.lin+dose.quad+dose.cub+dose.fourth,rats)
summary(aov(mod3))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## rat          9 1.6685  0.1854  22.205 3.75e-12 ***
## dose.lin     1 0.0610  0.0610   7.308   0.0104 *
## dose.quad    1 0.3943  0.3943  47.232 4.83e-08 ***
## dose.cub     1 0.0041  0.0041   0.491   0.4882
## dose.fourth  1 0.0008  0.0008   0.094   0.7613
## Residuals   36 0.3006  0.0083
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see that the fitted model is given by

```
summary(mod3)
```

```
##
## Call:
## lm(formula = rate ~ rat + dose.lin + dose.quad + dose.cub + dose.fourth,
##     data = rats)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.18780 -0.04155 -0.00380  0.05670  0.15620
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.706000   0.040862  17.278  < 2e-16 ***
## rat2        -0.014000   0.057788  -0.242  0.80995
## rat3         0.012000   0.057788   0.208  0.83667
## rat4         0.116000   0.057788   2.007  0.05227 .
## rat5         0.282000   0.057788   4.880 2.17e-05 ***
## rat6         0.128000   0.057788   2.215  0.03318 *
## rat7         0.202000   0.057788   3.496  0.00127 **
## rat8         0.440000   0.057788   7.614 5.23e-09 ***
## rat9         0.500000   0.057788   8.652 2.56e-10 ***
## rat10        0.416000   0.057788   7.199 1.80e-08 ***
## dose.lin     0.078108   0.028894   2.703  0.01041 *
## dose.quad   -0.198575   0.028894  -6.873 4.83e-08 ***
## dose.cub    -0.020239   0.028894  -0.700  0.48815
## dose.fourth -0.008845   0.028894  -0.306  0.76129
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09137 on 36 degrees of freedom
## Multiple R-squared:  0.8763, Adjusted R-squared:  0.8316
## F-statistic: 19.61 on 13 and 36 DF,  p-value: 1.59e-12
```

where we confirm that the estimated coefficient corresponding to the quadratic term is negative as expected from the parabola plot in question 4.

7. Verify in that the sum of the sums of squares attributed to each of the four dose effects (linear, quadratic, cubic and fourt order) equals the sum of squares due to the factor dose effect in Question 2.

```
sum.aov<-summary(aov(mod3))
sum(sum.aov[[1]][c("dose.lin","dose.quad","dose.cub","dose.fourth"),"Sum Sq"])
```

```
## [1] 0.460208
```

```
sum.aov<-summary(aov(mod2))
sum(sum.aov[[1]][c("dose"),"Sum Sq"])
```

```
## [1] 0.460208
```

This is because any sum of squares can always be decomposed into the sum of orthogonal components ! Also note that the 4 degrees of freedom of the sum of squares due to the dose in the ANOVA table in question 3 has also been decomposed into 4 individual sum of squares, each with its own single degree of freedom. If we now fit a model only with linear and quadratic constrasts then the remaining 2 degrees of freedom of the discarded cubic and fourth order terms are passed onto the degrees of freedom of the `ssE`. The actual value of the `ssE` increases slightly as expected.

```
mod4<-lm(rate~rat+dose.lin+dose.quad,rats)
summary(aov(mod4))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## rat          9 1.6685  0.1854   23.07 8.69e-13 ***
## dose.lin     1 0.0610  0.0610    7.59  0.00896 ** 
## dose.quad    1 0.3943  0.3943   49.06 2.43e-08 ***
## Residuals   38 0.3054  0.0080                     
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

Finally we note that due to the 2 degrees of freedom that have been passed onto the ssE we can now verify that the sum of the sums of squares attributed to the two dose effects (linear and quadratic) plus the ssE equals the sum of squares due to the factor dose plus the ssE effect in Question 2.

```
sum.aov<-summary(aov(mod4))
sum(sum.aov[[1]][c("dose.lin","dose.quad","Residuals"),"Sum Sq"])
```

```
## [1] 0.76076
```

```
sum.aov<-summary(aov(mod2))
sum(sum.aov[[1]][c("dose","Residuals"),"Sum Sq"])
```

```
## [1] 0.76076
```