Example-Ch10: One-way ANOVA

The example dataset comes from a study of bold coagulation times: 24 animals were randomly assigned to four different diets and the samples were taken in a random order.

```
>library(faraway)
>data(coagulation)
>coagulation
coag diet
1     62     A
2     60     A
3     63     A
4     59     A
5     63     B
6     67     B
……
21    63     D
22    64     D
23    63     D
24    59     D
```

$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

$i = 1, 2, 3, k$

Subjective Impression (guess)

exploratory data analysis

summary statistics

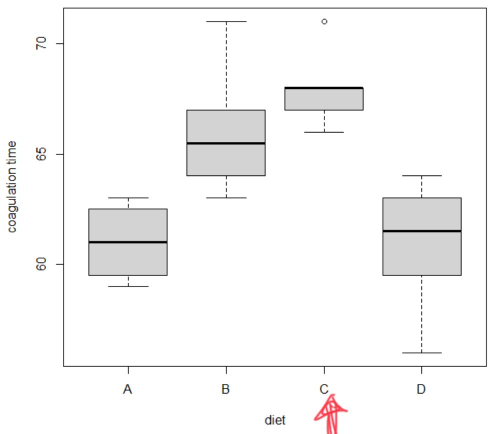
— min Q_1 Q_2 Q_3 max

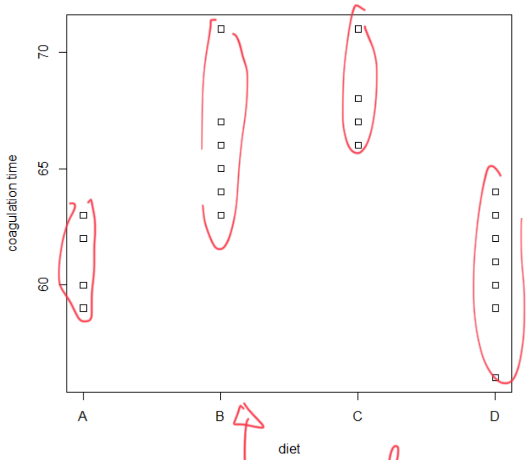
mean. s.d.

— box-plot, scatter plot

The first step is to plot the data by comparing boxplots and strip plots:

```
>plot(coag~diet,coagulation,ylab="coagulation time")
>with(coagulation,stripchart(coag~diet,vertical=TRUE, methods="stack",
xlab="diet", ylab="coagulation time" ))
```





n_{ij} — large

n_{ij} — small

$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$,

$\alpha_1 = 0$

treatment effects

Fit the model using a linear regression model (the default set)

```
>g<-lm(coag~diet,coagulation)
>summary(g)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 6.100e+01  1.183e+00  51.554  < 2e-16 ***
dietB       5.000e+00  1.528e+00   3.273 0.003803 **
dietC       7.000e+00  1.528e+00   4.583 0.000181 ***
dietD       2.991e-15  1.449e+00   0.000 1.000000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.366 on 20 degrees of freedom
Multiple R-squared:  0.6706,     Adjusted R-squared:  0.6212
F-statistic: 13.57 on 3 and 20 DF,   p-value: 4.658e-05
```

A 61

B 61+5=66

C 61+7=68

D 61+0=61

We can fit the model without an intercept term as

$\mu = 0$

$y_{ij} = \alpha_i + \varepsilon_{ij}$

```
> gi<-lm(coag~diet -1,coagulation)
> summary(gi)
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
dietA  61.0000     1.1832   51.55   <2e-16 ***
dietB  66.0000     0.9661   68.32   <2e-16 ***
dietC  68.0000     0.9661   70.39   <2e-16 ***
dietD  61.0000     0.8367   72.91   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.366 on 20 degrees of freedom
Multiple R-squared:  0.9989,    Adjusted R-squared:  0.9986
F-statistic:  4399 on 4 and 20 DF,   p-value: < 2.2e-16
```

treatment effects

$R^2 = \frac{SSR}{SST}$

$y_i = \beta_0 + \underbrace{\beta_1 x_{i1} + \cdots + \beta_k x_{ik}}_{SSR} + \varepsilon_i$

meaningless

Note that the R-squared and F-test are not correctly calculated. To generate usual test, we may use the following way:

```
> gnull<-lm(coag~1,coagulation)
> anova(gnull, gi)
Analysis of Variance Table

Model 1: coag ~ 1
Model 2: coag ~ diet - 1
  Res.Df RSS Df Sum of Sq       F    Pr(>F)
1     23 340
2     20 112  3       228 13.571 4.658e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$y_{ij} = \mu + \varepsilon_{ij}$

$y_{ij} = \alpha_i + \varepsilon_{ij}$

We can also use a sum coding:

```
> options(contrasts=c("contr.sum","contr.poly"))
> gs<-lm(coag~diet,coagulation)
> summary(gs)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  64.0000     0.4979 128.537  < 2e-16 ***
diet1        -3.0000     0.9736  -3.081 0.005889 **
diet2         2.0000     0.8453   2.366 0.028195 *
diet3         4.0000     0.8453   4.732 0.000128 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.366 on 20 degrees of freedom
Multiple R-squared:  0.6706,    Adjusted R-squared:  0.6212
F-statistic: 13.57 on 3 and 20 DF,  p-value: 4.658e-05
```

$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$

$\sum \alpha_i = 0$

treatment effects

A $64 - 3 = 61$

B $64 + 2 = 66$

C $64 + 4 = 68$

D $64 + \alpha_4 = 64 - 3 = 61$

$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0$

$\Rightarrow \alpha_4 = -3$