Design and Analysis of Experiment

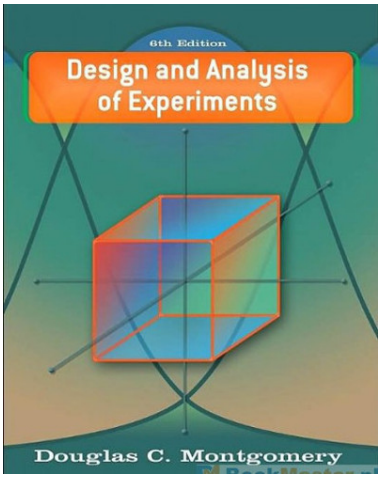
Trygve Almøy

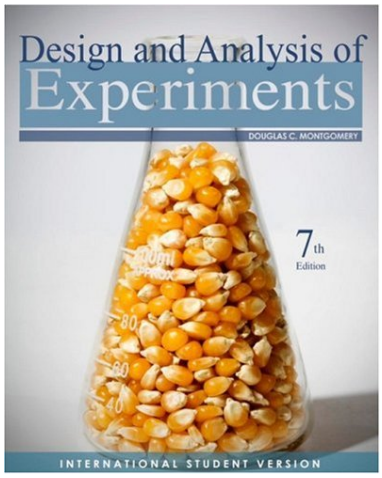
2017

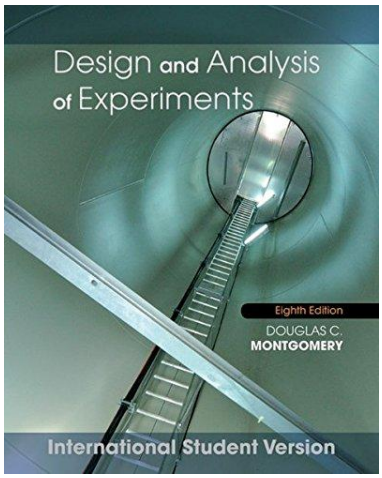
Practical Information

Note: Editions 6, 7 and 8 of Montgomery can all be used. Note, however, that the numbering of exercises and tables typically differs and information is given to clarify as needed. The data is in the fronter folder data, also in some cases where the student is asked to input the data.

Book we use in this course







Exercise and Page Number in Different Editions

Table.Exercise	Edition6	Edition7	Edition8
Table	2.5	2.50	2.60
Exercise	Missing	2.20	2.10
Exercise	Missing	2.30	2.30
Exercise	2.5	2.16	2.20
Exercise	2.13	2.21	2.29
Exercise	2.18	2.27	2.34

Week One

Exercise 1

a) Input the 'Portland cement' data (reproduced from Montgomery Table 2.1 in week1, slide 13) into R. Excel file called Tabl2

This can be done in several ways. One possibility is to enter data in excel and copy and then go R Commander: Data > Import data > From text … - Tick: Clipboard, Comma (If you use comma as Decimal point in excel) - Check that data looks right by ticking: View data set

You can also load it using `load` function if your dataset is in `.Rdata` format. In this case, if your dataset `Tabl21.Rdata` is in download folder under your home folder,

```
load("~/Downloads/Tabl21.Rdata")
```

From R-commander: Data > Load data > …

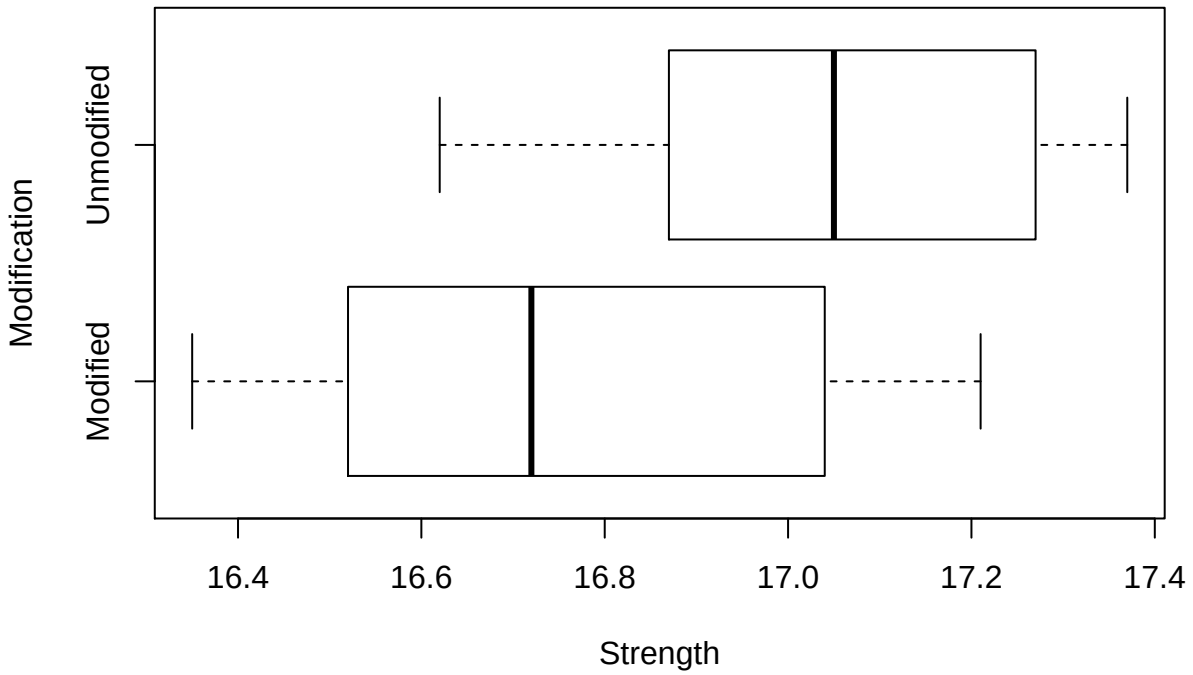
b) Stack the columns (Data > Active dataset> …)

```
PortlandStack <- stack(Portlandcement)
names(PortlandStack) <- c('Strength', 'Modification')
```

c) Use R to produce summary statistics (as produced by the summary command in R) for each of the groups.

```
   Modified       Unmodified
 Min.   :16.35   Min.   :16.62
 1st Qu.:16.54   1st Qu.:16.90
 Median :16.72   Median :17.05
 Mean   :16.77   Mean   :17.04
 3rd Qu.:17.02   3rd Qu.:17.23
 Max.   :17.21   Max.   :17.37
```

d) Make boxplot for each of the groups.



e) We would like to test if the two cement types are equal. Formulate the hypotheses formally and use R to perform the test. Formulate a conclusion.

Let Y_{ij} is strength in cement type i, sample j. We assume $Y_{ij} \sim N(\mu_i, \sigma^2)$, where all observations are independent. This is equal to assume this model:

$$Y_{ij} = \mu_i + \epsilon_{ij}, \text{ where } \epsilon_{ij} \sim N(0, \sigma^2), i = 1, 2 \text{ and } j = 1, 2, \ldots 9$$

Note that we assumed equal variances for both types. The hypothesis for testing if the two mortar formulations are equal is,

$$H_0 : \mu_{\text{modified}} - \mu_{\text{unmodified}} = 0$$
$$H_1 : \mu_{\text{modified}} - \mu_{\text{unmodified}} \neq 0$$

```
    Two Sample t-test

data:  Modified and Unmodified
t = -2.1767, df = 18, p-value = 0.04306
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.542393085 -0.009606915
sample estimates:
       mean of x        mean of y  pooled std.dev.
      16.7660000       17.0420000        0.2835293
```

Since p-value is smaller than 0.05, we reject null hypothesis (H_0) and conclude that there is significant different between two mortar formulation at 5 percent level.

f) Explain different ways of performing tests (hints: test statistic, p-values, and confidence intervals).

Although we have used p-value in previous question, we can use following ways to perform this test,

I) Test Statistics: The test-statistic is compared with the critical value from t-table at 5% level of significance. We reject H_0 if,

$$t_{\text{calculated}}(\text{test-statistic}) > t_{\alpha/2, n_1+n_2-2}$$

.

II) Confidence Interval Method: If 95% confidence interval for the estimate does not include zero, then we reject the null hypothesis H_0 at 5% level of significance. In our case, the 95% confidence interval (-0.5423931, -0.0096069) does not include zero so, we reject null hypothesis at 5% level of significance.

III) p-value Approach: If p-value is less than 0.05 then we reject H_0 at 5% level of significance.

g) Repeat the calculations for e) above, but now using only a calculator, summary output from R and a table of the t distribution.

The null hypotheses expresses equality of expected values, i.e.,

$$H_0 : \mu_1 = \mu_2$$

The test statistic is,

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\text{SEM}}$$

where,

$$\begin{aligned}
\bar{x}_1 &= 16.766 & \bar{x}_2 &= 17.042 \\
s_1^2 &= 0.0993156 & s_2^2 &= 0.0614622 \\
s_p^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = 0.2835293 \\
\text{SEM} &= s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = 0.1267982 \\
t &= -2.1766874
\end{aligned}$$

Here, We reject the null hypothesis since

$$|t| = 2.1766874 > t_{0.025;18} = 2.100922$$

Exercise 2

Two sorts of wheat, Zebra and RB07, were grown in 8 randomly chosen fields per sort. The protein content was measured for each sort and the data are given below.

Sort	Protein
Zebra	12.1
Zebra	12.8
Zebra	10.4
Zebra	11.9
Zebra	11.8
Zebra	11.6
Zebra	13.4
Zebra	13.3
RB07	18.3
RB07	19.5
RB07	12.7
RB07	14.7
RB07	15.3
RB07	16.1
RB07	15.4
RB07	16.8

The data are stored in a file called `Zebra` (Excel). Import the data to R by Data > Import data from > from excel file

a) Consider both wheat sorts. Formulate a model for the data analysis where you assume equal variances.

A model for the data:

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

where, $\epsilon_{ij} \sim N(0, \sigma^2)$ and the observations are indipendent for all Sorts $i = 1, 2$ and Portein Sample $j = 1, 2, \ldots, 8$

b) How would you estimate this variance?

We first estimate the variance in each group. Here are summary statistics from R:

```
         mean        sd       var length
RB07  16.1000 2.1226668 4.5057143      8
Zebra 12.1625 0.9898593 0.9798214      8
```

Combine the two estimates by the pooled variance estimate as,

$$
\begin{aligned}
S_p^2 &= \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \\
&= \frac{(8-1)4.51 + (8-1)0.98}{8+8-2} \\
&= 2.745
\end{aligned}
$$

c) How many parameters are included in the model? Give an interpretation of the parameters.

There are 3 parameters in this model:

- μ_1 is the expected protein level in Zebra wheat (in the whole population)
- μ_2 is the expected protein level in RB07 wheat (in the whole population)
- σ^2 is the variance in protein level within a wheat sort. This is assumed to be the same in both Zebra and RB07.

d) What is the estimated difference between the population means? What are the standard deviation and the standard error of this estimate?

```
        mean        sd       var length
RB07  16.1000 2.1226668 4.5057143      8
Zebra 12.1625 0.9898593 0.9798214      8
```

We estimate the expected difference $\mu_1\mu_2$ by the difference in sample means which is $16.112.1625 = 3.9375$. The standard deviation for this estimate is

$$\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_1}} = \frac{\sigma}{2}(\text{ since, } n_1 = n_2 = 8)$$

The standard error is the estimated standard deviation or

$$\hat{\sigma}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{\hat{\sigma}}{2} = \frac{S_{\text{pooled}}}{2} = 0.8280652$$

e) Test whether the expected protein content in Zebra is different from RB07.

The two sample t-test result from R is,

```
    Two Sample t-test

data:  RB07 and Zebra
t = 4.7551, df = 14, p-value = 0.0003075
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.161477 5.713523
sample estimates:
       mean of x        mean of y  pooled std.dev.
        16.10000         12.16250          1.65613
```

Since p-value $= 0.00031 << 0.05$, we reject H_0 and claim that the there is significant different in the expected protein content in two types of wheat at 5% level of significance.

f) Construct a 95 % CI for the true difference in protein between the two sorts. Is zero included in the interval? If not, what does this mean?

From the two-sample t-test output above, the 95% confidence interval is (2.1614768, 5.7135232). Here zero is not included in the interval, this also confirm our previous result that the expected protein content in Zebra and RB07 is significantly different.

NOTE:: The remaining questions are a bit harder, more theoretical …

g) If you are told that

$$(n_1 + n_2 2)\frac{S^2_{\text{pooled}}}{\sigma^2} \sim \chi^2_{n_1+n_2 2}$$

show that s^2_{pooled} is an unbiased estimate of σ^2.

From theory, we know,

$$\frac{k \cdot S^2}{\sigma^2} \sim \chi^2_k$$

where S^2 is a variance estimate based on k independent terms (that is k is the degrees of freedom associated with S^2). Hence, we have,

$$\frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}$$

For χ^2 distribution variables with k degrees of freedom we know that the expected value is k and the variance is $2k$. Therefore we can deduce:

$$E\left[\frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2}\right] = n_1 + n_2 - 2$$

$$\left[\frac{(n_1 + n_2 - 2)E\left(S_p^2\right)}{\sigma^2}\right] = n_1 + n_2 - 2$$

$$\text{Therefore, } E\left(S_p^2\right) = \sigma^2$$

h) Find the variance of S^2_{pooled}. Construct a 95% CI for σ^2. Explain the interval to a person without statistical knowledge.

Similarly using the fact that the variance of the χ^2 distributed variable with k degrees of freedom is $2k$:

$$\text{var}\left[\frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2}\right] = 2(n_1 + n_2 - 2)$$

$$\left[\frac{(n_1 + n_2 - 2)}{\sigma^2}\right]^2 \text{var}\left(S_p^2\right) = 2(n_1 + n_2 - 2)$$

$$\text{Therefore, var}\left(S_p^2\right) = \frac{2\sigma^4}{n_1 + n_2 - 2}$$

Further,

$$\text{sd}\left(S_p^2\right) = \sqrt{\frac{2\sigma^4}{n_1+n_2-2}}$$

and the standard error is the estimate of,

$$\text{se}\left(S_p^2\right) = \sqrt{\frac{2S_p^4}{n_1+n_2-2}}$$

Which by inserting all the known values gives,

$$\text{se}\left(S_p^2\right) = \sqrt{\frac{2\times 2.743^2}{8+8-2}} = 1.075$$

and the confidence interval is,

$$\left[\frac{(n_1+n_2-2)S_p^2}{\chi^2_{0.025,n_1+n_2-2}}, \frac{(n_1+n_2-2)S_p^2}{\chi^2_{0.975,n_1+n_2-2}}\right]$$

for $\alpha/2 = 0.025$, we find from the Table III in the Appendix over tail probabilities of the Chi-square distribution that $\chi^2_{0.025,14} = 26.119$ and $\chi^2_{0.975,14} = 5.629$. This gives the interval [1.47, 6.823]. This means that we are 95% certain that the true variance of protein yield in either Zebra or RB07 is expected to lie in this interval.

i) Test if the (population) variance is greater than 5. State the null hypothesis, the alternative and the level of significance, draw the conclusions.

Hypothesis for the test is,

$$H_0 : \sigma^2 = 5 \text{ vs } H_1 : \sigma^2 > 5$$

Under null hypothesis,

$$\chi^2 = \frac{(n_1+n_2-2)S_p^2}{\sigma_0^2} = \frac{38.402}{5} = 7.686$$

is distributed as $\chi^2_{n_1+n_2-2}$. We reject the null hypothesis at 5% level if $\chi^2 > \chi^2_{0.05,14} = 23.685$, which it is not. Thus, we retain the null hypothesis and cannot claim that the population variance in protein yield is larger than 5.

Exercise 3 (Relatively Theoritical)

In a pond there are thousands of fish. Two students want to estimate the average weight of all fish in the pond. Student 1 catches one fish of 1 kg, while student 2 catches 2 fishes, one with weight 1 kg, another weighing 1.2 kg. The following estimators for the unknown expectation are suggested

$$\hat{\mu}_1 = \bar{Y}_1 = 1$$
$$\hat{\mu}_2 = \bar{Y}_2 = 1.1$$
$$\hat{\mu}_3 = \frac{\bar{Y}_1 + \bar{Y}_2}{2} = 1.05$$
$$\hat{\mu}_4 = \frac{\bar{Y}_1 + 2\bar{Y}_2}{3} = 1.07$$

Show that all estimators are unbiased, calculate their variance, and give your vote to the one with smallest variance.

Remember that $E(\bar{Y}) = \mu$, thus the first two estimators are unbiased as they are the average value of each students fish weight. Further,

$$E(\hat{\mu}_3) = \frac{E(\bar{Y}_1) + E(\bar{Y}_2)}{2} = \frac{\mu + \mu}{2} = \mu$$
$$\text{and, } E(\hat{\mu}_4) = \frac{E(\bar{Y}_1) + E(2\bar{Y}_2)}{3} = \frac{\mu + 2\mu}{3} = \mu$$

Also, the variance for each of them are,

$$\text{var}(\hat{\mu}_1) = \sigma^2$$
$$\text{var}(\hat{\mu}_2) = \sigma^2/2$$
$$\text{var}(\hat{\mu}_3) = \frac{1}{4}\left(\sigma^2 + \frac{\sigma^2}{2}\right) = \frac{3}{8}\sigma^2$$
$$\text{var}(\hat{\mu}_4) = \frac{1}{9}\left(\sigma^2 + \frac{4\sigma^2}{2}\right) = \frac{3}{9}\sigma^2$$

In general: If student 1 has caught n_1 fishes and student 2 has caught n_2 fishes, the best estimator is

$$\hat{\mu} = \frac{n_1\bar{Y}_1 + n_2\bar{Y}_2}{n_1 + n_2}$$

Show that this is unbiased and find the variance.

Here,

$$\begin{aligned} E(\hat{\mu}) &= E\left(\frac{n_1\bar{Y}_1 + n_2\bar{Y}_2}{n_1 + n_2}\right) \\ &= \frac{n_1 E(\bar{Y}_1) + n_2 E(\bar{Y}_2)}{n_1 + n_2} \\ &= \frac{n_1\mu + n_2\mu}{n_1 + n_2} = \mu \end{aligned}$$

Thus, $\hat{\mu}$ is an unbiased estimator of μ. Further, the variance is,

$$\text{var}\left(\hat{\mu}\right) = \frac{1}{(n_1+n_2)^2}\left[n_1^2\text{var}(\bar{Y}_1) + n_2^2\text{var}(\bar{Y}_2)\right]$$

Exercise 4

Assume the following result (on scale 0-100) based on 8 randomly selected students from an exam in a basic course in statistics.

	S1	S2	S3	S4
Females	80	85	73	69
Males	90	60	76	67

a) Execute a test (choose the level of significance yourself) to investigate if there is larger variation in statistical ability among males compared to females.

N

Mean

Variance

Females

4

76.75

50.91667

Males

4

73.25

167.58333

Y_{ij} is result for person j belonging to group i.

The Model is:

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

, where $\epsilon_{ij} \sim N(0, \sigma_i^2)$, $i = 1, 2$ and $j = 1, 2, 3, 4$ and the observations are independent

$i = 1$ is females, otherwise males

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 < \sigma_2^2$$

The test statistics is,

$$F = \frac{S^2_{\text{males}}}{S^2_{\text{females}}} = 3.29$$

If we apply 5% level of significance, then we reject H_0 if $F > 9.28$.

Conclusion: We cannot reject.

b) Based on the result of the test in a) state a model and test if there is difference in statistical knowledge between males/females. The test gave a 0.65 as P-value. Explain in detail what this means.

Wa assume equal variances for males and females end state the following model:

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

where, $\epsilon_{ij} \sim N(0, \sigma^2)$

The hypothesis is,

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Assume H_0, and $T_{\text{calculated}} = 0.474$.

Very small T, it is impossible to reject H_0 also p-value = 0.653.

If males and felmale in average (population) obtain the same result, the probability of observing a difference (absolute value) between the sample means equal or greater than 0.474 is 0.653.

Exercise 5

Consider the data of Exercise 1 again (available as `Table21Stacked.RData` in fronter; you can go Data > Load … to load the data). If necessary stack the data.

a) Are the variances of the two groups different? Formulate a hypothesis test and perform the test using R (Statistics > Variances)

The hypothesis to test if the variance of the two groups different is,

$$H_0 : \sigma^2_{\text{modified}} = \sigma^2_{\text{unmodified}}$$
$$H_1 : \sigma^2_{\text{modified}} \neq \sigma^2_{\text{modified}}$$

The test result from R is,

```
    F test to compare two variances

data:  variable by factor
F = 1.6293, num df = 9, denom df = 9, p-value = 0.4785
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4046845 6.5593806
sample estimates:
ratio of variances
          1.629257
```

Here, p-value $= 0.478 > 0.05$, we cannot reject H_0 and could not claim that the variance between two groups are different.

b) Do the test by as explained in Week1, slide 39 using only the fact below

var

n

Modified

0.1001378

10

Unmodified

0.0614622

10

The observed test statistic is,

$$F = \frac{0.1001}{0.0614} = 1.63$$

Under the null hypothesis, $F \sim \text{Fisher}(9, 9)$ From table (or in R: Distributions > Continuous distributions > F distribution > F quantiles),

$$F_{0.025,9,9} = 4.026$$

This value is not exceeded (draw a figure) and therefore we cannot claim that variances differ.

Comment: Observe that we only know that the p-value > 0.05 as opposed to the exact p-value found in a) above. In this case we don't need the lower critical value as $F > 1$. However, this can be found as,

$$F_{0.975,9,9} = \frac{1}{F_{0.025,9,9}} = \frac{1}{4.03} = 0.25$$

c) Assume now the variances of the two groups to be equal and calculate a 95% confidence interval for the common standard deviation using the output in b) above.

Here, we have,

$$n = n_1 + n_2 = 20, \text{df} = 20 - 2 = 18$$

Under the assumption of equal variance of the two groups, the pooled variance is calculated as,

$$\begin{aligned} S_p^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \\ &= \frac{s_1^2 + s_2^2}{2} = 0.0808 \end{aligned}$$

Also,

$$\chi^2_{0.975} = 8.231 \text{ and } \chi^2_{0.025} = 8.231$$

$$\begin{aligned} \left[\frac{\text{df} \times S_p^2}{\chi^2_{0.025}}, \frac{\text{df} \times S_p^2}{\chi^2_{0.975}}\right] &= \left[\frac{18 \times 0.0808}{31.53}, \frac{18 \times 0.0808}{8.23}\right] \\ &= [0.046, 0.0177] \end{aligned}$$

Exercise 6: Two sample t-test or One way ANOVA in R

Load the data svin.RData. The data are as described in Week1 slide 17. Final weights have been collected for 9 pigs in the soya group and 8 pigs in the non-soya group. The question addressed is: Do the different feeding strategies give different carcass weight (Norwegian: Slaktevekt)?

a) Produce relevant summary statistics and plots. There are several ways to summarise, e.g., here's default for R: Data > Numerical summaries split on groups

mean

sd

var

n

not.soya

77.75

3.370036

11.35714

8

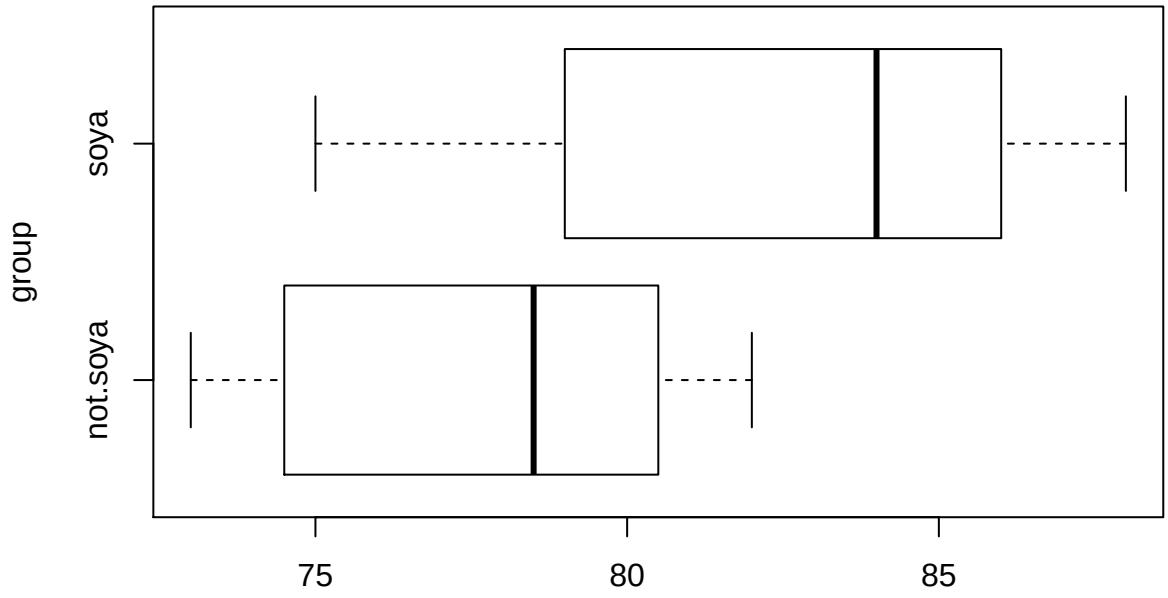
soya

82.88889

4.675587

21.86111

9



Boxplot:

b) Formulate the model (include assumptions, also equal variances in both groups) and hypotheses.

See Week 1, Exercise - 4 Answer

c) Perform the test and comment

As a two-sample test (R: Statistics > Means …)

```
    Two Sample t-test

data:  slaktevekt by group
t = -2.5681, df = 15, p-value = 0.02142
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
 -9.4040625 -0.8737153
sample estimates:
       mean of x        mean of y  pooled std.dev.
       77.750000        82.888889         4.118162
```

As an ANOVA

```
Analysis of Variance Table
          Df Sum Sq Mean Sq F value  Pr(>F)
group      1 111.85 111.846   6.595 0.02142 *
Residuals 15 254.39  16.959
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Comment: Same p-value. ANOVA only gives two sided test.

d) Calculate a 95% confidence interval for the difference in weights in the two groups.

See Week 1, Exercise - 6(c)

e) Repeat c) and d) above but now use a significance level of 0.01 and calculate 99% confidence intervals.

f) Explain the difference between one-sided and two-sided hypotheses.

Exercise 7: One way ANOVA

Three diets for pigs were investigated, two of the diets had different soya proteins (`S1` and `S2`) one diet had no soya included (`NONS`). The following weights were recorded.

NONS	S1	S2
80	82	82
84	84	86

NOTE::Do the following without any software or pocket calculator:

State the model that you would apply to this experiment.

The model is:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}, \text{ where } \epsilon_{ij} \sim N(0, \sigma^2),\ i = 1, 2, 3;\ \ j = 1, 2$$

and where the observations are independent

How many parameters are included in the model?

With the assumption of $\tau_i's$ sum to zero, i.e. $\sum_{i=1}^{3} \tau_i = 0$, there are 4 unknown paramters, μ, τ_1, τ_2 and σ^2.

State the null hypothesis and the alternative to test if there is any effect of diet.

The null and alternative hypothese for testing if there is any effect of diet is,

$$H_0 : \tau_1 = \tau_2 = \tau_3$$
$$H_1 : \tau_i \neq 0 \text{ for at least one } i$$

Find the rejection area if you test on level 0.05 or on level 0.1

- at 5% level of significance, we reject H_0, if F-statistic is larger than $F_{0.05,(2,3)}$, i.e. 9.552
- at 1% level of significance, we reject H_0, if F-statistic is larger than $F_{0.01,(2,3)}$, i.e. 30.817

Find SS_T, $SS_{\text{Treatments}}$, SS_E, and their degrees of freedom.
Let y_{ij} denotes the measurement for i^{th} treatment and j^{th} observation. So, $i = 1, 2, 3$ and $j = 1, 2$ (n)

$$\text{SS}_{\text{total}} = \sum_{i=1}^{3}\sum_{j=1}^{2}\left(y_{ij} - \mu\right)^2 = 22$$

$$\text{SS}_{\text{residual}} = \sum_{i=1}^{3}\sum_{j=1}^{2}\left(y_{ij} - \bar{y}_i\right)^2 = 18$$

$$\text{and, SS}_{\text{treatment}} = n\sum_{i=1}^{3}\left(\bar{y}_i - \mu\right)^2$$
$$= \text{SS}_{\text{total}} - \text{SS}_{\text{residual}} = 4$$

Also, the degree of freedom for treatment is $3 - 1 = 2$, for residual is is $6 - 3 = 3$ and for total is $6 - 1 = 5$
Find $MS_{\text{Treatments}}$ and MS_E and F_0. Write up the ANOVA table. Use R to get the pvalue. HINT: Distribution > Continuous > F …)

$$\text{MS}_{\text{Treatment}} = \frac{\text{SS}_{\text{Treatment}}}{\lceil\{_{\text{treatment}}} = 2$$
$$\text{MSE} = \text{MS}_{\text{residual}} = \frac{\text{SS}_{\text{residual}}}{\lceil\{_{\text{residual}}} = 6$$

Further,

$$F_0 = \frac{\text{MS}_{\text{Treatment}}}{\text{MSE}} = 0.333$$

Using R, the pvalue corresponding to F_0 is 0.7401
Using all these values, we can construct ANOVA table as,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	2	4	2	0.3333	0.7401
Residuals	3	18	6		

What is your conclusion on the test?
Calculated F_0 is too small; for this test only large values can lead to rejection and the critical value is $F_{0.05,2,3} = 9.55$. Therefore, We can not reject H_0.
Estimate all the parameters in the model.
The estimated parameters in the model are,

$$\begin{aligned}
\hat{\sigma}^2 &= \text{MSE} = 6 \\
\hat{\mu} &= \bar{Y}_{..} = 83 \\
\hat{\tau}_1 &= \hat{\mu}_1 - \hat{\mu} = -1 \\
\hat{\tau}_2 &= \hat{\mu}_2 - \hat{\mu} = 0 \\
\hat{\tau}_3 &= \hat{\mu}_3 - \hat{\mu} = 1
\end{aligned}$$

NOTE:: Use pocket calculator:

Find a 95% CI for σ^2

Use Chi-square table with 3 degrees of freedom to get,

$$\begin{aligned}
\chi^2_{0.025,3} &= 9.348 \\
\chi^2_{0.975,3} &= 0.216
\end{aligned}$$

The 95% CI for σ^2 is,

$$\left[\frac{\text{SSE}}{\chi^2_{0.025,3}}, \frac{\text{SSE}}{\chi^2_{0.975,3}}\right] = [1.925, 83.412]$$

Find a 99% CI for the difference between the two soya diets.

The CI for $\mu_i - \mu_j$ is same as CI for $\tau_i - \tau_j$, so, the CI for soya diets (τ_2 and τ_3) is,

$$\begin{aligned}
\bar{y}_{3\cdot} - \bar{y}_{2\cdot} - t_{\alpha/2,N-a}\sqrt{\frac{2\text{MSE}}{n}} &\leq \mu_3 - \mu_2 \\
&\leq \bar{y}_{3\cdot} - \bar{y}_{2\cdot} + t_{\alpha/2,N-a}\sqrt{\frac{2\text{MSE}}{n}} \\
\hat{\tau}_{3\cdot} - \hat{\tau}_{2\cdot} - (5.841) \times \sqrt{\frac{2\times 6}{2}} &\leq \mu_3 - \mu_2 \\
&\leq \hat{\tau}_{3\cdot} - \hat{\tau}_{2\cdot} + (5.841) \times \sqrt{\frac{2\times 6}{2}} \\
1 - (14.307) \leq &\mu_3 - \mu_2 \leq 1 + (14.307) \\
-13.307 \leq &\mu_3 - \mu_2 \leq 15.307
\end{aligned}$$

In otherwords, the 95% confidence limit for true difference between μ_3 and μ_2 lie in the interval $[-9.26, 7.26]$

Control as many answers as possible by using R

NOTE:: Without any software or pocket calculator:

Add 6 kg to all the weights in the soya group.

What influence will this have on the sum of squares?

If we add 6 to the soya groups then the difference between the group averages will be larger, but noting happens to variation inside the group. $SS_{\text{treatments}}$ will increase, but the noise (SSE) will be unaffected. Sum of squares for Treatment, Residual and Toatal before and after adding 6 to soya group is,

	before	after
treatment	4	76
residual	18	18
total	22	94

What is F_0. now? Comment!

Since F value is a ratio betwen MS for treatment and residuals, F_0 (F value) will be larger. This increases the chance of rejecting null hypothesis (no difference between groups).

Perform the test in this situation.

The ANOVA table in this situation is,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	2	76	38	6.3333	0.0838
Residuals	3	18	6		

Here, the p-value is much smaller than in the previous case, and even is significant at 10%. This shows that the difference between treatments (NONS, S1 and S2) has increased in this situation.

Go back to the original data, but add 6 kg to the heaviest pig in each group.

What influence will this have on the Sum of squares?

The differences between the means will be unaffected, but the variance inside the groups will be larger. Hence $SS_{\text{treatments}}$ will be unaffected, but the noise and SSE will increase. The ANOVA in this situation is,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	2	4	2	0.0455	0.9562
Residuals	3	132	44		

What is F_0 now? Comment!

Here, F_0 is even smaller since noise (variation within groups) has increases.

Perform the test in this situation.

Small F_0 leads to low probability of rejecting null hypothesis. From ANOVA table, high p-value shows that we can not reject null hypothesis and conclude that there is not significant difference between the effects of diet on pigs.

Exercise 8

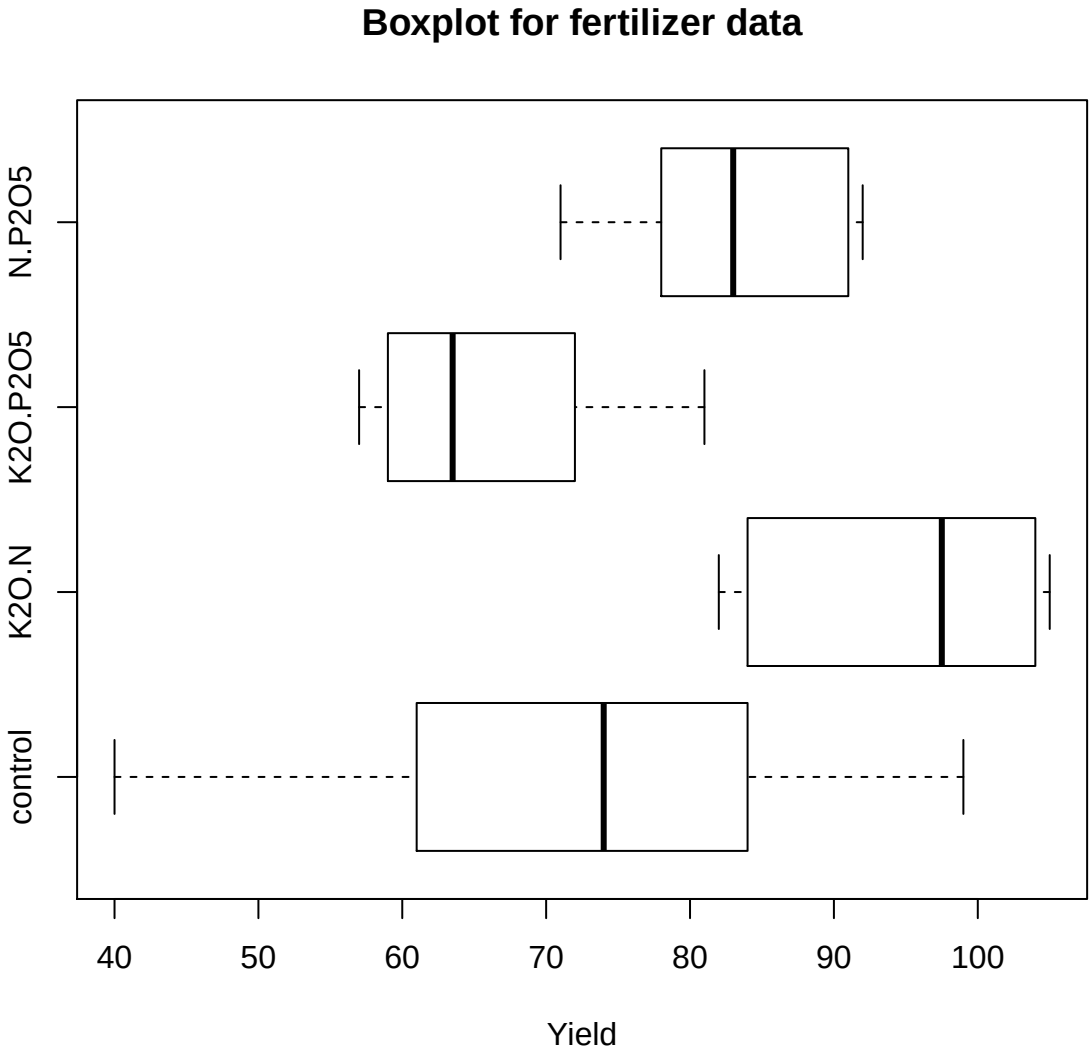
Load the data `fertilizer.RData`. The data have been discussed in class, different slides:

a) Produce relevant summary statistics and plots.

A summary statistics for the fertilizer data is,

	mean	sd	n	0%	25%	50%	75%	100%
control	72	20.169	6	40	63.75	74.0	82.00	99
K2O.N	95	9.879	6	82	87.00	97.5	102.75	105
K2O.P2O5	66	8.989	6	57	60.00	63.5	70.00	81
N.P2O5	83	8.319	6	71	78.25	83.0	90.00	92

Further, a boxplot can be a relevant plot to compare the effect of fertilizer,



b) Do the fertilizers give different yields? Formulate the hypotheses and do the ANOVA in R.

Let y_{ij} be the yield from i^{th} fertilizer and j^{th} replication.

The model:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \epsilon_{ij} \sim NID(0, \sigma^2)$$

Here, $i = 1, 2, 3, 4$ and $j = 1, 2, \ldots 6$

The hypothesis for testing if the fertiliizers give different yields is,

$$H_0 : \tau_i = 0 \text{ for all } i = 1, 2, 3, 4$$
$$H_1 : \tau_i \neq 0 \text{ for at least one } i$$

The ANOVA table we obtain is,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fertilizer	3	2940	980.0	5.99	0.004
Residuals	20	3272	163.6		

Here, p-value is smaller than 0.05 or even 0.01. So we reject null hypothesis H_0 at 5% and also at 1% level of significance.

c) Formulate a conclusion.

Since the null hypothesis is rejected, we conclude that fertilizers result significantly different yield.

d) There are four different fertilizers. How many pairwise comparisons can be made?

From four differnt fertilizer, we can make ${}^4C_2 = 6$ comparisons. The comparisons can be between:

control	control	control	K2O.N	K2O.N	K2O.P2O5
K2O.N	K2O.P2O5	N.P2O5	K2O.P2O5	N.P2O5	N.P2O5

e) Use Tukey's method to compute confidence intervals for the differences in means. Calculate p-values based on Tukey's method.

The result from Tukey's method is,

```
    Simultaneous Confidence Intervals and Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = y ~ fertilizer, data = fertilizer)

Quantile = 2.7989
Minimum significant difference = 20.6692
95% confidence level

Linear Hypotheses:
```

```
                    Lower  Center   Upper Std.Err t value   P(>t)
control-K2O.N    -43.669 -23.000  -2.331   7.385  -3.115 0.02582 *
control-K2O.P2O5 -14.669   6.000  26.669   7.385   0.812 0.84784
control-N.P2O5   -31.669 -11.000   9.669   7.385  -1.490 0.46190
K2O.N-K2O.P2O5     8.331  29.000  49.669   7.385   3.927 0.00427 **
K2O.N-N.P2O5      -8.669  12.000  32.669   7.385   1.625 0.38785
K2O.P2O5-N.P2O5  -37.669 -17.000   3.669   7.385  -2.302 0.13109
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

The formula for finding Tukey's confidence interval for same sample size for all groups:

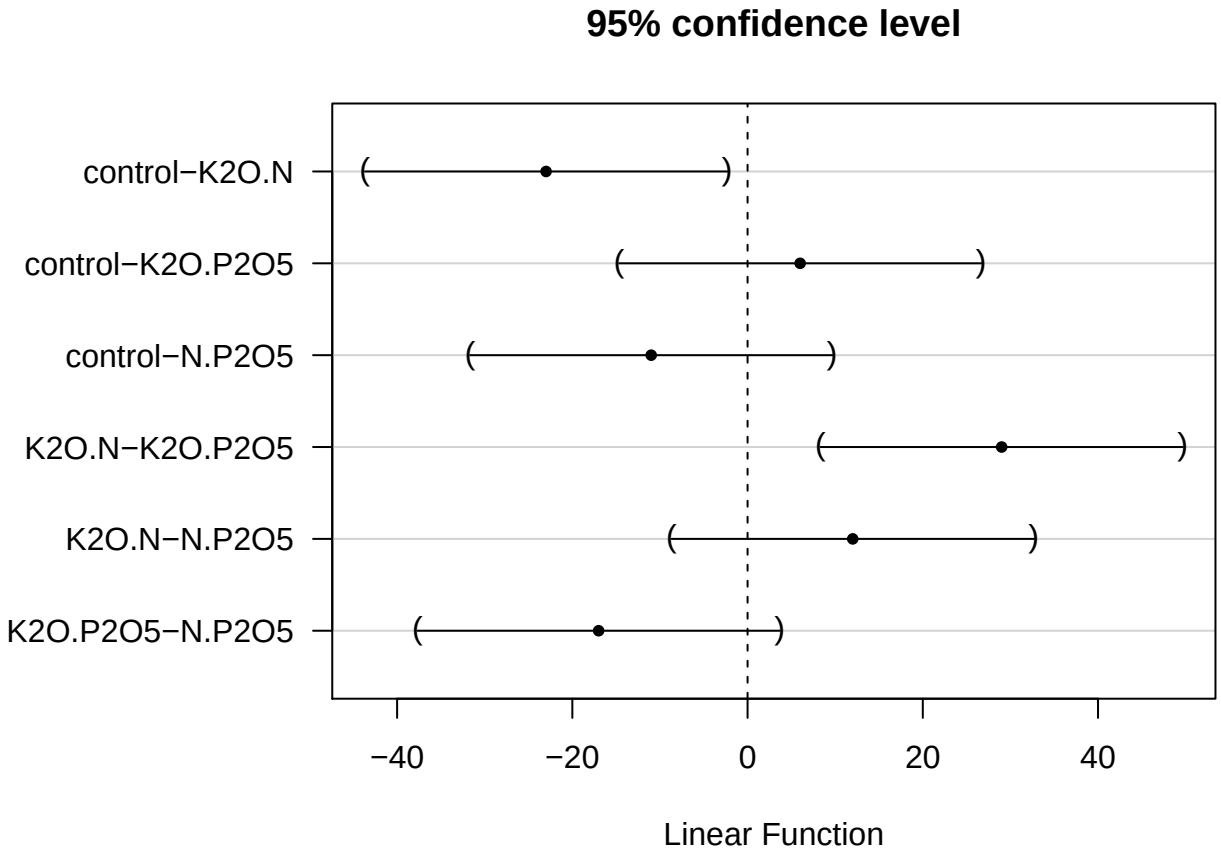
$$\left[\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \pm q_{\alpha(a,f)}\sqrt{\frac{\text{MSE}}{n}}\right]$$

Here, a = Total number of groups, f = degree of freedom associated with MSE and α = level of significance. The value of $q_{\alpha(a,f)}$ is on Studentized Range Statistic Table

f) Formulate a conclusion.

Tukey's output in previous answer shows that yield from fertilizer `K20.N` differs significantly from `control` (p-value: 0.02562) and `K20.P205` (p-value: 0.00439).

We can also use plot for Tukey test for some visualization,



g) Test if the 3 new fertilizers are better than the control by using a suitable contrast.

For testing the average effect of 3 new fertilizer with control, a contrast can be constructed with coefficients `c(-1, 1/3, 1/3, 1/3)`. So, the hypothesis in this situation will be,

$$H_0 : \frac{1}{3}(\mu_2 + \mu_3 + \mu_4) - \mu_1 = 0$$
$$H_1 : \frac{1}{3}(\mu_2 + \mu_3 + \mu_4) - \mu_1 \neq 0$$

In general form, the hypothesis can be written as,

$$H_0 : \sum_{i=1}^{a} c_i \bar{y}_{i\cdot} = 0$$
$$H_1 : \sum_{i=1}^{a} c_i \bar{y}_{i\cdot} \neq 0$$

In this case, $c_i = -1, 1/3, 1/3, 1/3$ for $i = 1, 2, 3, 4$ respectively

We can write contrast as,

$$C = \sum_{i=1}^{a} c_i \bar{y}_{i\cdot}$$

The Standard Error for C is,

$$\text{SE}(C) = \sqrt{\frac{\text{MSE}}{n} \sum_{i=1}^{a} c_i^2}$$

Thus, the test statistics for testing the hypothesis is,

$$t_0 = \frac{\sum_{i=1}^{a} c_i \bar{y}_{i\cdot}}{\sqrt{\frac{\text{MSE}}{n} \sum_{i=1}^{a} c_i^2}} \sim t_{\alpha/2, N-a}$$

Further, the Confidence Interval is given as,

$$\left[\sum_{i=1}^{a} c_i \bar{y}_{i\cdot} \pm t_{\alpha/2, N-a} \sqrt{\frac{\text{MSE}}{n} \sum_{i=1}^{a} c_i^2} \right]$$

The result from R-commander contrast test is,

```
                                    Estimate Std. Error  t value
fertilizer c=( 1 -0.33 -0.33 -0.33 ) -9.310175   6.014596 -1.54793
                                    Pr(>|t|) DF lower CI upper CI
fertilizer c=( 1 -0.33 -0.33 -0.33 ) 0.1373179 20 -21.8564 3.236053
```

From the output, we can see that we fail to reject Null hypothesis and thus conclude that the average yield on using new fertilizers is not significatly different from that obtained from `control`.

Exercise 9

Exam STAT 210 Sep 2012, Exercise 1

Question: The purpose of this exercise to determine if a new diet treatment or method (called 'M1') designed to help people loosing weight is better than two well known methods (called 'M2' and 'M3'). Two individuals were recruited in each group. The individuals were weighed at the beginning and the end of the study. For each individual the weight difference ('final weight'-'initial weight') was recorded. The data and some summary statistics are as follows:

```
   Treatment Observations Averages
1:        M1          0,2        1
2:        M2          1,3        2
3:        M3          5,7        6
```

Treatment	Observations	Averages
M1	0, 2	1
M2	1, 3	2
M3	5, 7	6

Total Average		3

We will use the model,

$$y_{ij} = \mu + \tau_i + \epsilon_{ij},\ \epsilon_{ij} \sim NID(0, \sigma^2)$$

Here,

$i = 1, 2, 3$ corresponding to M1, M2 and M3

$j = 1, 2$ corresponding to the two observations in each treatment group

y_{ij} = Weight Loss and $\sum_{i=1}^{3} \tau_i = 0$

a) Calculate $\text{SS}_{\text{Treatment}}$ and SS_{E} defined below:

$$\text{SS}_{\text{Treatment}} = 2 \sum_{i=1}^{3} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

$$\text{SS}_{\text{E}} = \sum_{i=1}^{3} \sum_{j=1}^{2} (y_{ij} - \bar{y}_{i\cdot})^2$$

b) Consider the null hypothesis

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0$$

Show that the test statistic is $F_0 = 7$ and use this to perform the test. Formulate a conclusion.

c) We would like to extend on the above pilot study and design a real study. How would you design this study? You are free to state your assumptions. You can mention words like randomization, blocking and replication. You are not supposed to write more than 10 full sentences.

Exercise 10

We will compare two sorts of barley cultivated at two different sites. Let $Y_{i,j}$ be yield (in kg) for sort i, site j, where $i = 1, 2$ and $j = 1, 2$ We have,

$$\begin{aligned} Y_{1,1} &= 1 \quad Y_{1,2} = 3 \\ Y_{2,1} &= 2 \quad Y_{2,2} = 6 \end{aligned}$$

Find and interpret:

$$\bar{Y}_{i\cdot} = \frac{1}{2}\sum_{j=1}^{2} Y_{ij} \text{ for } i = 1 \text{ and } 2$$

$$\bar{Y}_{\cdot j} = \frac{1}{2}\sum_{j=1}^{2} Y_{ij} \text{ for } i = 1 \text{ and } 2$$

$$\bar{Y}_{\cdot\cdot} = \frac{1}{4}\sum_{j=1}^{2}\sum_{j=1}^{2} Y_{ij}$$

Find:

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\left(Y_{ij} - \bar{Y}_{\cdot\cdot}\right)^2 \quad \sum_{i=1}^{2}\sum_{j=1}^{2}\left(Y_{ij} - \bar{Y}_{i\cdot}\right)^2$$

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\left(Y_{ij} - \bar{Y}_{\cdot j}\right)^2 \qquad \sum_{i=1}^{2}\left(\bar{Y}_{i} - \bar{Y}_{\cdot\cdot}\right)^2$$

Here we can write,

	Site 1	Site 2	Row Mean
Sort 1	$y_{11} = 1$	$y_{12} = 3$	$\bar{y}_{1\cdot} = 2$
Sort 2	$y_{21} = 2$	$y_{22} = 6$	$\bar{y}_{2\cdot} = 4$
Column Mean	$\bar{y}_{\cdot 1} = 1.5$	$\bar{y}_{\cdot 2} = 4.5$	$\bar{y}_{\cdot\cdot} = 3$

$$\bar{Y}_{i\cdot} = \frac{1}{2}\sum_{j=1}^{2} Y_{ij} = 2 \text{ and } 4 \text{ for } i = 1 \text{ and } 2 \text{ respectively}$$

These are the average of sort 1 and 2 over all the sites.

$$\bar{Y}_{\cdot j} = \frac{1}{2}\sum_{i=1}^{2} Y_{ij} = 1.5 \text{ and } 4.5 \text{ for } i = 1 \text{ and } 2 \text{ respectively}$$

These are the average of site 1 and 2 for all the sorts.

Further,

$$\bar{Y}_{..} = \frac{1}{4}\sum_{j=1}^{2}\sum_{j=1}^{2} Y_{ij} = 3$$

This is the overall cultivated barley for all sorts and sites.

Thus, we can also find,

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\left(Y_{ij} - \bar{Y}_{..}\right)^2 = 14 \quad \sum_{i=1}^{2}\sum_{j=1}^{2}\left(Y_{ij} - \bar{Y}_{i.}\right)^2 = 10$$

$$\sum_{i=1}^{2}\sum_{j=1}^{2}\left(Y_{ij} - \bar{Y}_{.j}\right)^2 = 5 \qquad \sum_{i=1}^{2}\left(\bar{Y}_{i} - \bar{Y}_{..}\right)^2 = 2$$

Exercise 11

From Five sires (bulls used for breeding are often referred to as "sires") have we recoded the annual milk production from 8 daughters which are randomly picked out from all daughter of those sires. The data: `Sires.Rdata` (fronter)

a) State a model with `milk` production as response and `sire` as factor. How would you interpret all the parameters in this model?

Let Y_{ij} be milk produced by daughter j of sires i.

Model:

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

Here, $\epsilon_{ij} \sim \text{NID}(0, \sigma^2)$ and $i = 1, 2, \ldots 5$ and $j = 1, 2, \ldots 8$

Here, we also have one restriction,

$$\sum_{i=1}^{5} \tau_i = 0$$

Interpretaion of Parameters:

μ	:	Average milk production for all daughters under the 5 sires
τ_i	:	The average profit in annually milk production of using only sire i
σ^2	:	The variance in milk production for all daughter under one particular sire

b) Test if there is effect of `sire` with respect to milk production. (R-commander: Statistics > Fit model > Linear model. Note that `sire` must be factor variable)

The hypothesis is,

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0$$
$$H_1 : \tau_i \neq 0 \text{ for at least one } i$$

ANOVA for this model is,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sire	4	1251009	312752.3	2.805913	0.0403794
Residuals	35	3901166	111461.9		

Here, sire effect is significant at 5% (p-value: 0.0404)

c) Estimate all parameters in the model (R-commander: Model > summarize model)

The summary of above fitted model is,

```
Call:
lm(formula = milk ~ sire, data = Sires)

Residuals:
    Min      1Q  Median      3Q     Max
-651.25 -247.37   -2.69  199.84  872.75

Coefficients:
            Estimate Std. Error t value            Pr(>|t|)
(Intercept)  6518.15      52.79 123.478 <0.0000000000000002 ***
sire(1)       171.10     105.58   1.621               0.1141
sire(2)      -236.15     105.58  -2.237               0.0318 *
sire(3)      -192.40     105.58  -1.822               0.0769 .
sire(4)       152.97     105.58   1.449               0.1562
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

s: 333.9 on 35 degrees of freedom
Multiple R-squared: 0.2428,
Adjusted R-squared: 0.1563
F-statistic: 2.806 on 4 and 35 DF,  p-value: 0.04038
```

Here, the estimates of unknown parameters are,

	estimates
$\hat{\mu}$	6518.1500
$\hat{\tau}_1$	171.1000
$\hat{\tau}_2$	-236.1500
$\hat{\tau}_3$	-192.4000
$\hat{\tau}_4$	152.9750
$\hat{\sigma}$	333.8591

d) (Use the general fact that in an one way ANOVA model with a levels on the factor and n replicates $^{SSE}/_{\sigma^2}$ is chi square distributed with naa degrees of freedom). Find a 95 % confidence interval for the error variance (the unexplained variance). Explain the interval to farmer. (use the general fact that in an one way ANOVA model with a levels on the factor and n replicates $^{SSE}/_{\sigma^2}$ is chi square distributed with naa degrees of freedom)

95% confidence interval for σ^2 is given by,

$$\left(\frac{\text{SSE}}{\chi^2_{35,0.975}}, \frac{\text{SSE}}{\chi^2_{35,0.025}}\right) = \left(\frac{3901165.8}{53.2033485}, \frac{3901165.8}{20.5693766}\right)$$
$$= (73325.5689144, 189658.9220973)$$

Therefore, 95% confidence limit for σ is (271, 435) kg of milk. This value revers that with 95% of confidence level, the true variation in annual milk production for all daughter under one particular sire lie between (271, 435) kg.

Exercise 12

Tree diets for pigs were test out, in an experiment with 2 replicates. The Response is carcass weight (slaktevekt in Norwegian). Let's assume 4students did the experiment and that they really obtained different results:

Experiment 1

Diet1	69	71
Diet2	79	81
Diet3	89	91

Experiment 2

Diet1	60	80
Diet2	70	90
Diet3	80	100

Experiment 3

Diet1	70	90
Diet2	71	91
Diet3	72	92

Experiment 4

Diet1	80	82
Diet2	82	84
Diet3	83	85

Without computing anything:

Describe all 4 experiments, have focus on large explained or unexplained variation.

Experiment 1: Large difference between diets, small difference inside diets

Experiment 2: Large difference between diets, large difference inside diets

Experiment 3: Small difference between diets, large difference inside diets

Experiment 4: Small difference between diets, small difference inside diets

In which experiment do you think you will have a large, small or moderate F value?

Experiment 1: Large F, small P-value

Experiment 2: Moderate F

Experiment 3: Small F, large P-value

Experiment 4: Moderate F

In which experiment do you think you will have a large, small or moderate P value?

See previous answer

In which experiment do you think you will prove a significant effect of diet?

Experiment 1 will prove a significant effect of diet due to large difference between diets and small difference inside diets.

Experiment 2 could results with significant effect of diet but may get influenced by the noise within diets.

Experiment 3 and Experiment 4 has very small difference between diets however Experiment 4 has small variation within diets which can help us to see the difference between diets.

Check your answers by R-commander.

The ANOVA table for four experiments from R-commander is,

Experiment 1:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
diet	2	400	200	100	0.0017965
Residuals	3	6	2		

Experiment 2:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
diet	2	400	200	1	0.464758
Residuals	3	600	200		

Experiment 3:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
diet	2	4	2	0.01	0.9900827
Residuals	3	600	200		

Experiment 4:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
diet	2	9.333333	4.666667	2.333333	0.2447778
Residuals	3	6.000000	2.000000		

Extra exercises from textbook with solutions

Exercise 3.1

An experimenter has conducted a single-factor experiment with six levels of the factor, and each factor level has been replicated three times. The computed value of the F-statistic is $F_0 = 4.89$. Find bounds on the P-value.

F~Fisher (5,12). $0.01 <$ Table P-value < 0.025 Computer P-value $= 0.0114$

Exercise 3.2

An experimenter has conducted a single-factor experiment with four levels of the factor, and each factor level has been replicated six times. The computed value of the F-statistic is $F_0 = 4.02$. Find bounds on the P-value.

F~Fisher (3,20). $0.01 <$ Table P-value < 0.025 Computer P-value $= 0.022$

Exercise 3.3

A computer ANOVA output is shown below. Fill in the blanks. You may give bounds on the P-value.

One-way ANOVA

Source	DF	SS	MS	F	P
Factor	?	?	246.93	?	?
Error	25	186.53	?		
Total	29	1174.24			

Completed table is: (Answer regarding P requires computer, with table in Montgomery one can only say $P < 0.01$)

One-way ANOVA

Source	DF	SS	MS	F	P
Factor	4	987.71	246.93	33.09	< 0.0001
Error	25	186.53	7.46		
Total	29	1174.24			

Exercise 3.4

A computer ANOVA output is shown below. Fill in the blanks. You may give bounds on the P-value.

One-way ANOVA

Source	DF	SS	MS	F	P
Factor	3	36.15	?	?	?
Error	?	?	?		
Total	19	196.04			

Completed table is: (Answer P requires computer)

One-way ANOVA

Source	DF	SS	MS	F	P
Factor	3	36.15	12.05	1.21	0.3395
Error	16	159.89	9.99		
Total	19	196.04			

Exercise 3.5

A regional opera company has tried three approaches to solicit donations from 24 potential sponsors. The 24 potential sponsors were randomly divided into three groups of eight, and one approach was used for each group. The dollar amounts of the resulting contributions are shown in the following table

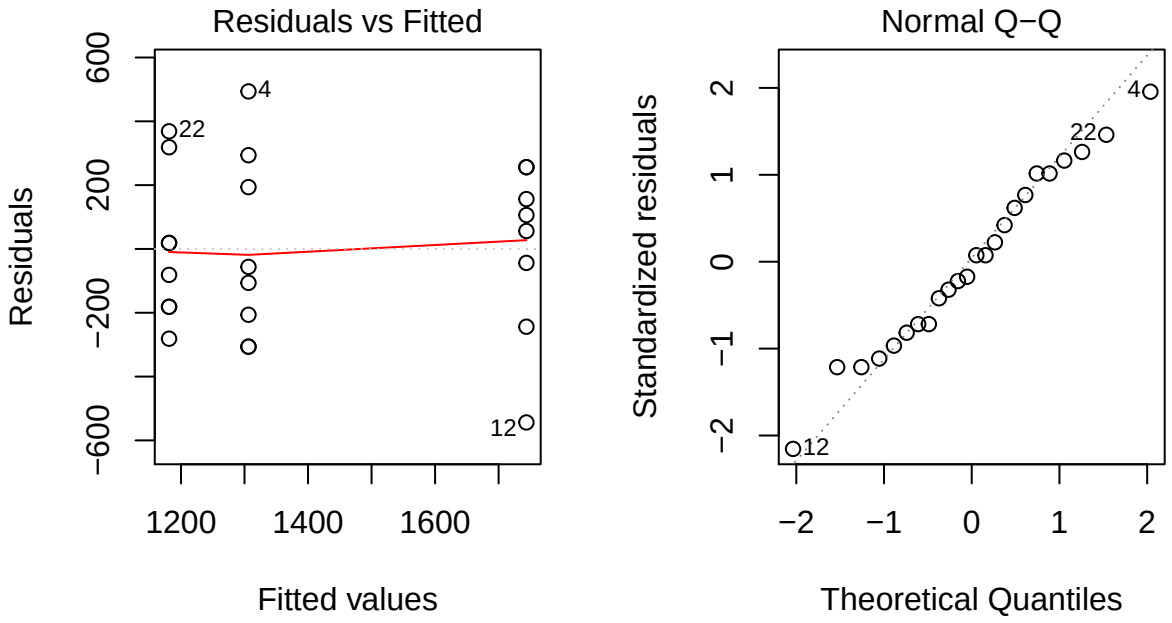
Approach	Contributions
1	1000, 1500, 1200, 1800, 1600, 1100, 1000, 1250
2	1500, 1850, 2000, 1200, 2000, 1700, 1800, 1900
3	900, 1000, 1200, 1500, 1200, 1550, 1000, 1100

The data is available as `Table35.Dollar.RData` and below R is used rather than Minitab used in Montgomery

(a) Do the data indicate that there is a difference in results obtained from the three different approaches? Use $\alpha = 0.05$.

```
Analysis of Variance Table

Response: Dollar
          Df  Sum Sq Mean Sq F value   Pr(>F)
Approach   2 1395833  697917  9.5851 0.001102 **
Residuals 21 1529062   72812
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


    Simultaneous Confidence Intervals and Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts


Fit: lm(formula = Dollar ~ Approach, data = Tabl35.Dollar)

Quantile = 2.5206
Minimum significant difference = 340.0727
95% confidence level

Linear Hypotheses:
        Lower  Center   Upper Std.Err t value   P(>t)
A1-A2 -777.57 -437.50  -97.43  134.92  -3.243 0.01043 *
A1-A3 -215.07  125.00  465.07  134.92   0.926 0.63000
A2-A3  222.43  562.50  902.57  134.92   4.169 0.00121 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)
```

(b) Analyze the residuals from this experiment and comment on the model adequacy.



There is nothing unusual about the residuals.

Extra exercises

NOTE:: only if time

Textbook ed 8 (numbers in ed 7 in parenthesis): 2.1 (2.2), 2.3 (2.3) ,2.20 (2.16), 2.29 (2.21).

The exercises, short solutions (and a few more problems) can also be found in `ExtraProblemsWithSolutionsch02ed7.pdf`.

Week Two

Exercise 1

Air pollution measured by chlorine content (measured in ppm, parts per million) for three cities, randomly selected during one year.

One year ago the cities were approximately equal polluted, but City 1 has the last year tried to reduce the pollution. Data is in a word file called Monday week 2 on fronter.

a) Stack the data. Use the ANOVA model to investigate if the chlorine content really differs. Check the model assumptions, which problems do you see?

   One of the way you can import data from word file into R is using clipboard. Just copy the data and import the data in clipboard into R (Rcommander: Data > Import

Data > from text file ...)

Let y_{ij} be the chlorine content (polution measurement) for city i. The ANOVA model to investigate if the chlorine content really differs is,

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \text{ where, } \epsilon_{ij} \sim \text{NID}\left(0, \sigma^2\right) \tag{1}$$

Since, city is randomly selected, we also have assumption that the treatment effect follows normal distribution with mean 0 and variance σ_τ^2, i.e. $\tau_i \sim \text{NID}\left(0, \sigma_\tau^2\right)$

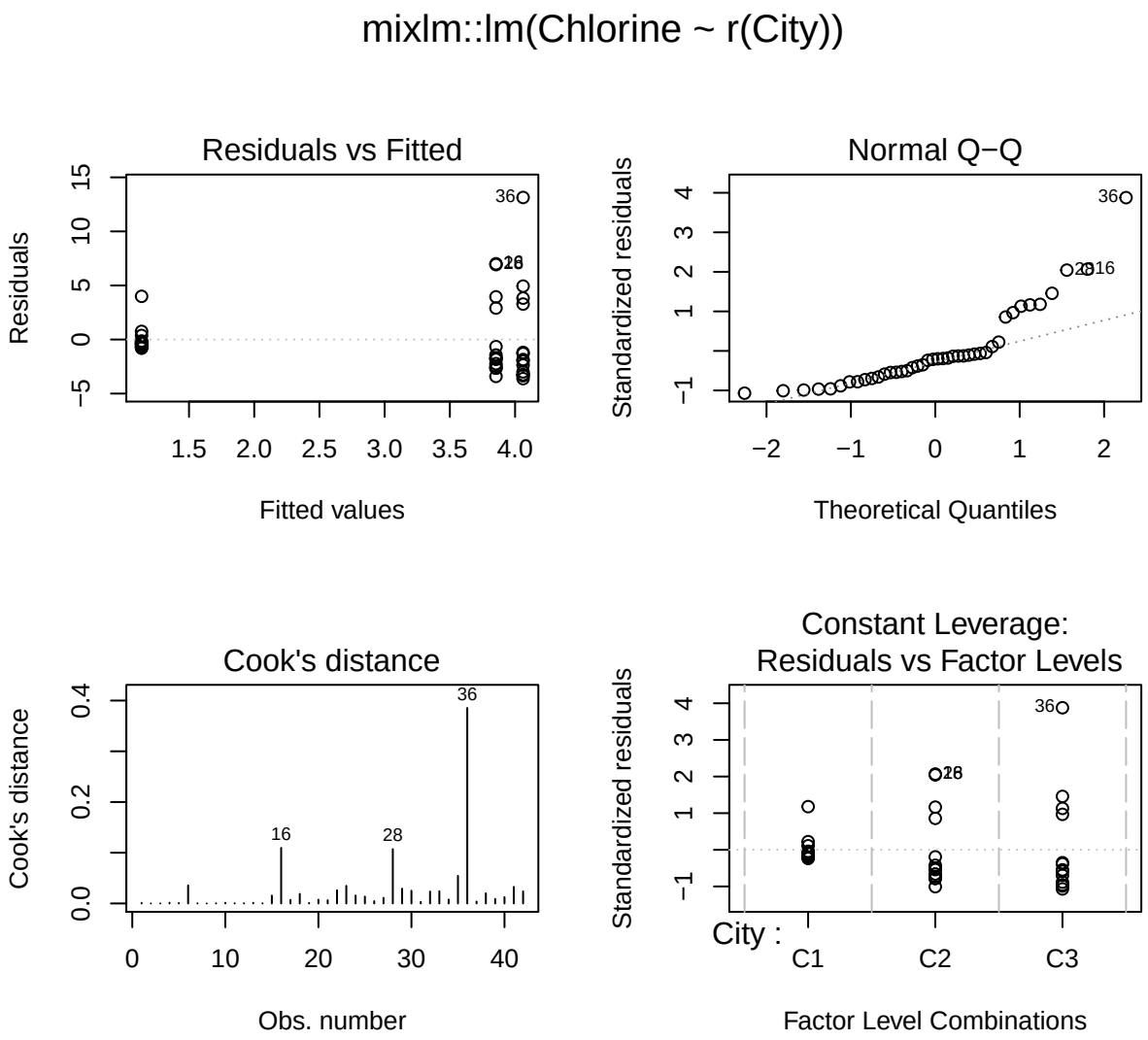
The ANOVA table for the Model ((1)) is,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
City	2	74.581	37.291	3.016	0.061
Residuals	39	482.230	12.365		

The p-value in ANVOA table is larger than 0.05, so can not reject following Hypothesis at 95% confidence level and stay on the fact that the polution level of different cities are same. However, at 90% confidence level the cities appears to be significantly different.

$$H_0 : \sigma_\tau^2 = 0$$
$$H_1 : \sigma_\tau^2 > 0$$

Following diagnostic plot (Rcommander: Model>Graph>Basic Diagnostic Plot) gives a picture of assumption wheather they hold true in this situation.



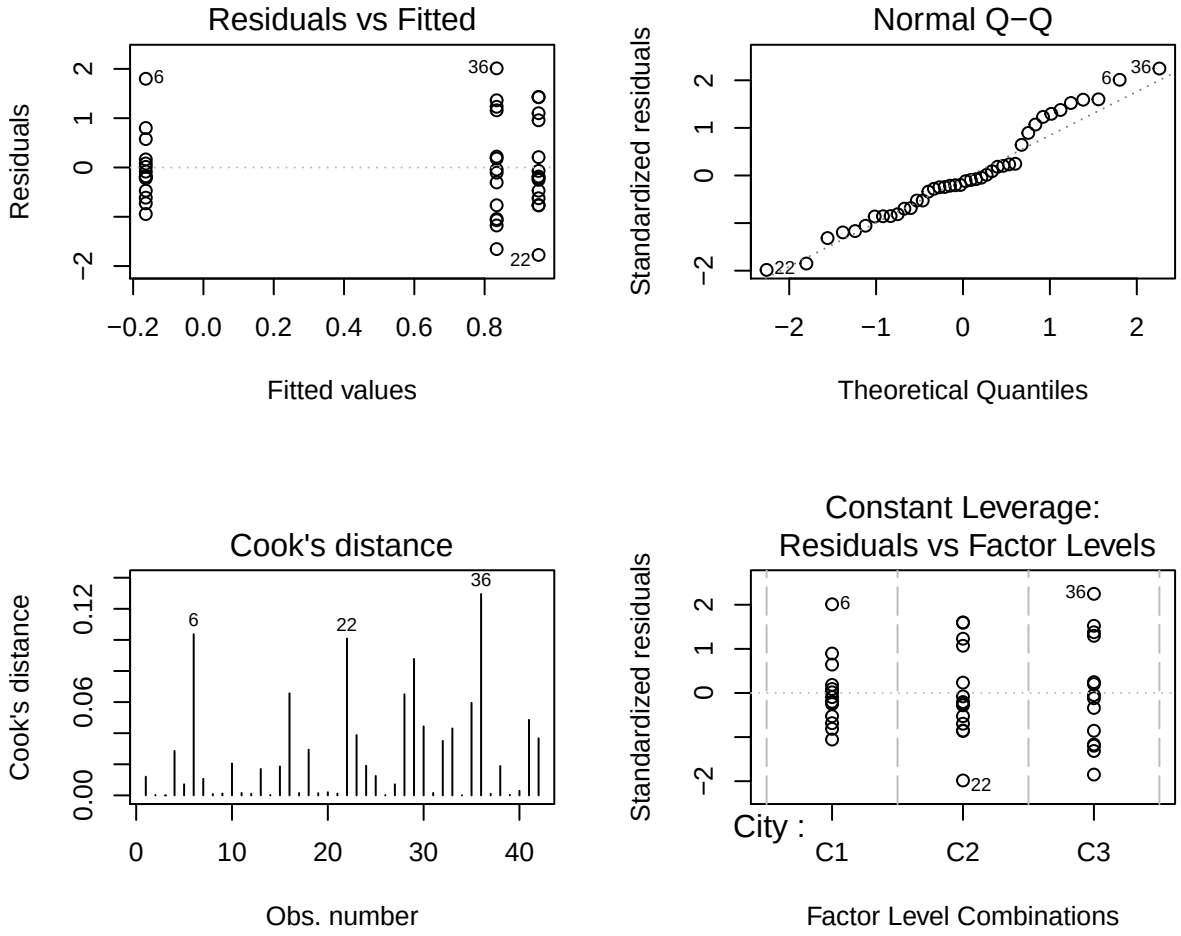
Try to interpreate the plot youself

b) Take the logarithm of the response. (Data > Manage variables in active dataset > compute new variables). Note that natural logarithm is done by `log(variable)`. Repeat the analysis, but on `log(chlorine)`. Check model assumptions once again. Investigate if city 1 has been able to reduce the pollution significantly compared the two others cities by a suitable contrast.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
City	2	10.5435	5.2718	6.1158	0.0049
Residuals	39	33.6177	0.8620		

The diagnostic plot can be obtained as in previous question,

mixlm::lm(‘log(Chlorine)‘ ~ r(City))



Interpreate the plot youself and find its differences from the previous diagnostic plot

To investigate if `city 1` has been able to reduce the pollution significantly compared the two others cities, a contrast with cofficient `c(1, -0.5, -0.5)`. Since,

$$\text{Contrast } (\Gamma) = \mu_1 - \frac{1}{2}(\mu_2 + \mu_3) = \tau_1 - \frac{1}{2}(\tau_2 + \tau_3)$$

The hypothesis for this test is,

$$H_0 : \Gamma = 0$$
$$H_1 : \Gamma < 0$$

In R-commander, we can test this hypothesis from Models > Test Contrasts in ANOVA. The result for the test is,

	Estimate	Std. Error	t value	Pr(>|t|)
City c=(1 -0.5 -0.5)	-1.057827	0.3039017	-3.480822	0.0012468

Since this is one sided hypothesis, the p-value for this test will be half the p-value we obtain in above table (See More). i.e.

$$\frac{0.0012}{2} = 0.0006$$

Since the p-value is very small, we reject the null hypothesis and claim that `City1` has significantly low chlorine level as compared to the other two cities.

Exercise 2

The following data are from an exploration of what kind of brand people in general prefer. Data could be found in the word file called `Monday week 2` on fronter. 6 recipes were tried out, each with 8 replicates. 48 respondents were randomly picked out and given one piece of bread (without information on recipe). They were asked to give points on a scale from 0 to 10 (10 is best).

R1	R2	R3	R4	R5	R6
4	7	6	4	3	7
4	6	6	3	5	8
6	5	8	5	5	6
5	3	10	4	6	4
10	5	7	6	6	10
8	6	10	7	9	6
9	9	8	3	2	4
7	4	8	3	2	6

R1, R2 and R3 are bread baked on coarse-grained flour (Norwegian grovmel), the others on fine.

R1, R3 and R5 is Swedish recipes, the others are Norwegian.

- State a model for this investigation.

  Let y_{ij} be the point for i^{th} recipe and j^{th} replication. The model is,

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \text{ where, } \epsilon_{ij} \sim \text{NID}\left(0, \sigma^2\right) \tag{2}$$

  In the model ((2)), we have parameters,

  μ : Overall average point obtained by all recipes

  τ_i : Effect of recipe i^{th} (point obtained by recipe i more(less) than the overall average)

  Here, we also assume that the overall effect of all recipes equals to zero.,

$$\sum_{i=1}^{5} \tau_i = 0$$

- Test if there is an effect of recipes.

  To fit the model from the given data, first we need to stack it. The hypothesis for testing if there is an effect of recipes is,

  $$H_0 : \tau_i = 0 \text{ for all } i$$
  $$H_1 : \tau_i \neq 0 \text{ for any } i$$

  The ANOVA table for model ((2)) in the previous answer is,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Recipe	5	66.938	13.388	3.517	0.01
Residuals	42	159.875	3.807		

  From the ANOVA table, the low p-value (0.01) shows that `Recipes` have significantly different preferences.

Answer the following questions by suitable contrasts.

- Is there positive effect of coarse-grained flour?

  Here, R1, R2 and R3 are bread baked on coarse-grained flour, so the contrast can be written with coefficients `c(1/3, 1/3, 1/3, -1/3, -1/3, -1/3)` as,

  $$\Gamma = \frac{1}{3}(\tau_1 + \tau_2 + \tau_3) - \frac{1}{3}(\tau_1 + \tau_2 + \tau_3)$$

  The hypothesis to test if this contrast is positive is,

  $$H_0 : \Gamma = 0$$
  $$H_1 : \Gamma > 0$$

  From the test output below, we reject H_0 (very small p-value, compare with half of p-value) and conclude that there is positive effect of coarse-grained flour.

	Estimate	Std. Error	t value	Pr(>|t|)
Recipe c=(1/3 1/3 1/3 -1/3 -1/3 -1/3)	1.5417	0.5632	2.7373	0.009

- Is there positive effect of coarse-grained flour, if we decide to use Norwegian recipes?

  Here only Norwegian recipes (R2, R4, R6) are considered among which only R2 is made from coarse-grained flour. Since we are testing R2 against average of R4 and R6, the coefficient for contrast in this situation is `c(0, 1, 0, -0.5, 0, -0.5)`. So,

$$\text{Contrast: } \Gamma = \tau_2 - \frac{1}{2}(\tau_4 + \tau_6)$$

The hypothesis for this test is,

$$H_0 : \Gamma = 0$$
$$H_1 : \Gamma > 0$$

From the test output below, we can not reject H_0 (high p-value) and conclude that there is not a positve effect of coarse-grained flour in Norwegian recipes. Here we compare p-value with half of its value since this is one-tailed test. Even in this situation, p-value is high enough not to reject H_0.

	Estimate	Std. Error	t value	Pr(>|t|)
Recipe c=(0 1 0 -0.5 0 -0.5)	0.25	0.8448	0.2959	0.7687

- Is there different effect of Norwegian and Swedish recipes if we decide to use coarse-grained flour?

  Among the course grained flour, R1 and R3 are Swedish recipes while R2 is Norwegian. Here,

  Contrast Coefficients: `c(-0.5, 1, -0.5, 0, 0, 0)`

  Contrast:

$$\text{Contrast: } \Gamma = \tau_2 - \frac{1}{2}(\tau_1 + \tau_3)$$

  Hypothesis:

$$H_0 : \Gamma = 0$$
$$H_1 : \Gamma \neq 0$$

  Test Result:

	Estimate	Std. Error	t value	Pr(>|t|)
Recipe c=(-0.5 1 -0.5 0 0 0)	-1.625	0.8448	-1.9235	0.0612

  Decision:

  Could not reject H_0 at 5% level of significance so there is no eveidence of Norwegian and Swedish recipes being different. However, at 10% percent, we conclude that there is significance difference between Norwegian and Swedish recipes.

Exercise 3

a) See Exercise 11 Week 1. Load the sire data once more.

- Explain what we mean by a residual.

  The part of response that your model could not expalin are residuals. Residuals are also termed as error terms or noise. You can obtain residuals as,

  $$\text{Error: } (\epsilon_{ij}) = y_{ij} - \hat{y}_{ij}$$

  For example in a one-way ANOVA model $y_{ij} = \mu + \tau_i + \epsilon_{ij},\ \epsilon \sim \text{NID}(0, \sigma^2)$, the estimated error terms are,

  $$\text{Error: } (\epsilon_{ij}) = y_{ij} - \hat{y}_{ij} = y_{ij} - (\hat{\mu} + \hat{\tau}_i)$$

- Calculate the residuals for all observations.

  In R-commander: Model > Add Observation to … and choose Residuals.

- What does it mean if a residual is positive or negative?

  Since $\epsilon_{ij} = y_{ij} - \hat{y}_{ij}$, when the true values are larger than the fitted values, residuals are positive. Similary, when the true values are smaller than the fitted values, residuals are negative. Large residual, either it is positive or negative indicates that the observation corresponding that residual is far away from the fitted model.

- Find the standardized residual. Do any residual have large absolute value?

  In R-commander: Model > Add Observation to … and choose standardized residuals. Following observation has largest absolute residual and standardized residuals.

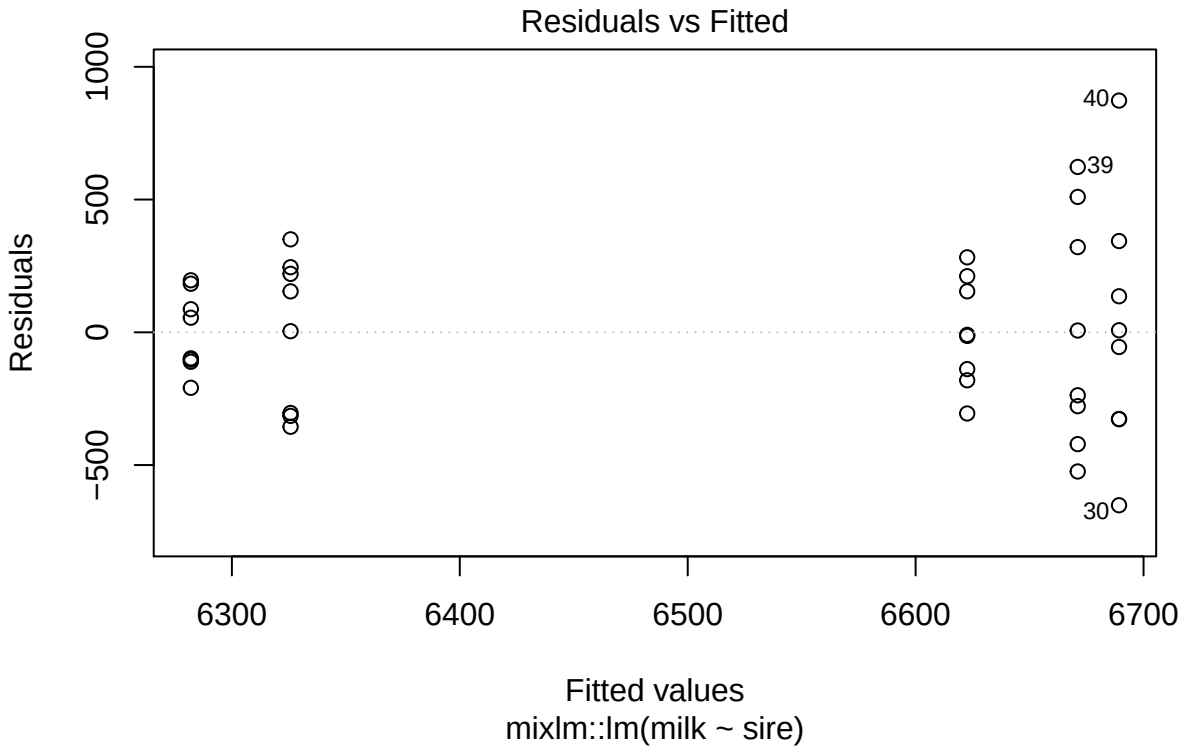
milk	sire	residual	std.residual
7562	1	872.75	2.794619

- Find the animal with the largest and the smallest residual.

  Sire 1 has both largest and smallest residuals. This also shows that there is large variation in the milk production from `sire1`.

milk	sire	residual	std.residual
6038	1	-651.25	-2.085358
7562	1	872.75	2.794619

- Plot the residuals against the fitted values, make comments



The plot shows that the residual terms have constant variance expcept for some larger fitted values. Observations 30, 39 and 40 have largest absolute residuals. These observations have distorted the assumption of constant variation of residual and could lead to poor predictions.

b) Assume now that sire is a random effect (discussed on lecture Tuesday).

If we are interested in the sire effect in general, explain why it is natural to assume sire as a random factor.

- Write down the model. Interpretate the 3 parameter in the model.

  The model with sire as random effect is,

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \text{ where, } \epsilon_{ij} \sim \text{NID}(0, \sigma^2) \text{ and } \tau_i \sim \text{NID}(0, \sigma_\tau^2) \tag{3}$$

  The three unknown parameters are,

  μ : Overall average annual milk production from all sires

  σ_τ^2 : Variability in annual milk production between different sires

  σ^2 : Variability in annual milk production within same sires

- Is there effect of sire? State the null hypothesis and the alternative. Estimate the variance components, and give an interpretation of these estimates. Estimate the correlation between cows with same father.

  Hypothesis for testing effect of sire is,

$$H_0 : \sigma_\tau^2 = 0$$
$$H_1 : \sigma_\tau^2 > 0$$

The ANOVA table for model ((3)) is,

	Mean Sq	Sum Sq	Df	F value	Pr(>F)
sire	312752.3	1251009	4	2.806	0.04
Residuals	111461.9	3901166	35		

Here, we have Mean Sum of Squares,

$$\text{MS}_{\text{treatment}} = 312752.34$$
$$\text{MS}_{\text{Error}} = 111461.88$$

The variance components σ^2 and σ_τ^2 referes to the variation within a factor and variation between factors respectively. The total variation present in y_{ij} is the sum of these two variances. The estimates of these variance components can be obtained as,

$$\hat{\sigma}^2 = 111462$$
$$\hat{\sigma}_\tau^2 = \frac{\text{MS}_{\text{treatment}} - \text{MS}_{\text{E}}}{n}$$
$$= \frac{201290.46}{8} = 25161.31$$

- How would you estimate the variance and the standard deviation for annual milk production in the whole population?

  Since the total variation present in y_{ij} is the sum of two variance components we have,

$$\text{Total Variation} = \sigma_\tau^2 + \sigma^2$$
$$= 25161.31 + 111461.88 = 136623.19$$

  Thus, the standard deviation for annual milk production in whole population is estimated as $\sqrt{136623.186} = 369.626$

- How would you estimate the expected annual milk production in the whole population? What is the standard error of this estimate? Construct a 95% CI for these expectation?

Exam Questions for exercise

a) Exam STAT 210 Sep 2012, Exercise 1. Problems and solutions available on fronter.
b) Exam STAT 210 Sep 2013, Exercise 3. Problems and solutions available on fronter (Tuesday).

Exercise 4

Assume the following result from a completely randomized block design (called experiment1) Where

		Block	
Treatment		1	2
	1	1	2
	2	3	5

- State the model.

Use pocket calculator (if necessary).

- State the model for the block experiment.
- Find SSTotal, SSTReatment, SSBlock and SSError.
- Estimate all the parameters in the model.
- Find the standard error of the estimates.
- Find the residuals.
- Find the fitted values.
- Split the fitted values into a general part, a part by treatment and a part by block.
- Make an ANOVA table and, test if there is treatment effect.

You can check your results with R-commander

If you instead obtained the following result (Experiment 2)

		Block	
Treatment		1	2
	1	1	2
	2	4	6

Answer the following without any calculations.

Experiment 2 gave:

- larger SSTR than Experiment 1?
- larger SSB than Experiment 1?
- larger SSE than Experiment 1?
- larger SST than Experiment 1?
- larger F or a larger p-value when testing for treatment effect than Experiment 1.

Analyze experiment 1 (incorrectly) by a one factor model.

- Why is SSE larger compared to the block design?
- Is the unexplained variance (σ^2) larger or small compared to the block design?

From textbook

Exercise 4.2

The ANOVA from a randomized complete block experiment output is shown below.

Source	DF	SS	MS	F	P
Treatment	4	1020.56	?	30.14	?
Block	?	?	64.765	?	?
Error	20	169.33	?		
Total	29	1513.71			

(a) Fill in the blanks. You may give bounds on the P-value.

Source	DF	SS	MS	F	P
Treatment	4	1020.56	255.1400	30.14000	< 0.0001
Block	5	323.82	64.7650	7.64956	0.00037
Error	20	169.33	8.4665		
Total	29	1513.71			

We can find p-value using computer or some advanced calculator but try to find a range of p-value from the F-table.

(b) How many blocks were used in this experiment?

One block factor with 5 levels is used in this experiment.

(c) What conclusions can you draw?

Problems and solutions for previous exams are found on fronter.

- Exam Stat 210 Sep 5, 2011.
- Exam Stat 210 Sep, 2012 Exercise 2.
- Exam Stat 210 Sep, 2013 Exercises 1 and 2.
- Exam Stat 210 Sep, 2014 Exercise 2.

Comment: Datasets (named `Exam*.RData`) are available for the exam problems for those who would like to reproduce output and check answers using R.

Exercise 5

will be discussed on lecture Thursday, but try yourself

Load the R data `blockwheat`. This is a block experiment where the response is `protein` in wheat. We have 3 sorts (Bastian, Berserk and Bjarne), each sort is tried once on 4 different fields (lock).

- Why is this regarded as a block experiment?

  Here, our primary interest is sorts irrespective of where it is tried (or grown). However, `field` factor can affect (increase noise) the analysis if it is not included in the model which consequently prevent us from discovering the effect of sort. Thus, a block experiment where `field` factor is also included in the model not as a primary subject of interest but rather to block unnecessary noises that it can create if it is not included in the model. Block experiment reduces noise so that we can investigate the factor effect that is of our interest.

- How many observations are there in total?

  There are 4 fields and 3 sorts where a protein value is measured for each of the combination. The total number of observation is 12.

- Do we have replicates?

  Only one measurement is made for each combination of field and sort, so there are not any replications. But if we remove field factor and consider one-way ANOVA model, we will have 4 replication of sort.

- State the model.

  The block experiment model we have is,

$$y_{ij} + \mu + \tau_i + \beta_j + \epsilon_{ij} \tag{4}$$

  Here the model follows the assumptions,

$$\epsilon_{ij} \sim \text{NID}\left(0, \sigma^2\right)$$
$$\sum_{i=1}^{3} \tau_i = 0$$
$$\sum_{j=1}^{4} \beta_j = 0$$

- Make an ANOVA table.

  ANOVA table for model ((4)) is,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sort	2	1.487	0.743	3.867	0.083
Field	3	10.907	3.636	18.913	0.002
Residuals	6	1.153	0.192		

- Does it seem smart to include field in the experiment (give reason for your answer?

  It is a better idea to include `field` in this experiment. Since `field` factor has significant effect and if it is removed, the variation present in protein measurement that it describes

will appear as a noise. Consequently, the probability of finding the effect of Sort which could be significant will decreases.

- Estimate the parameters in the model.

  The coefficient estimates are,

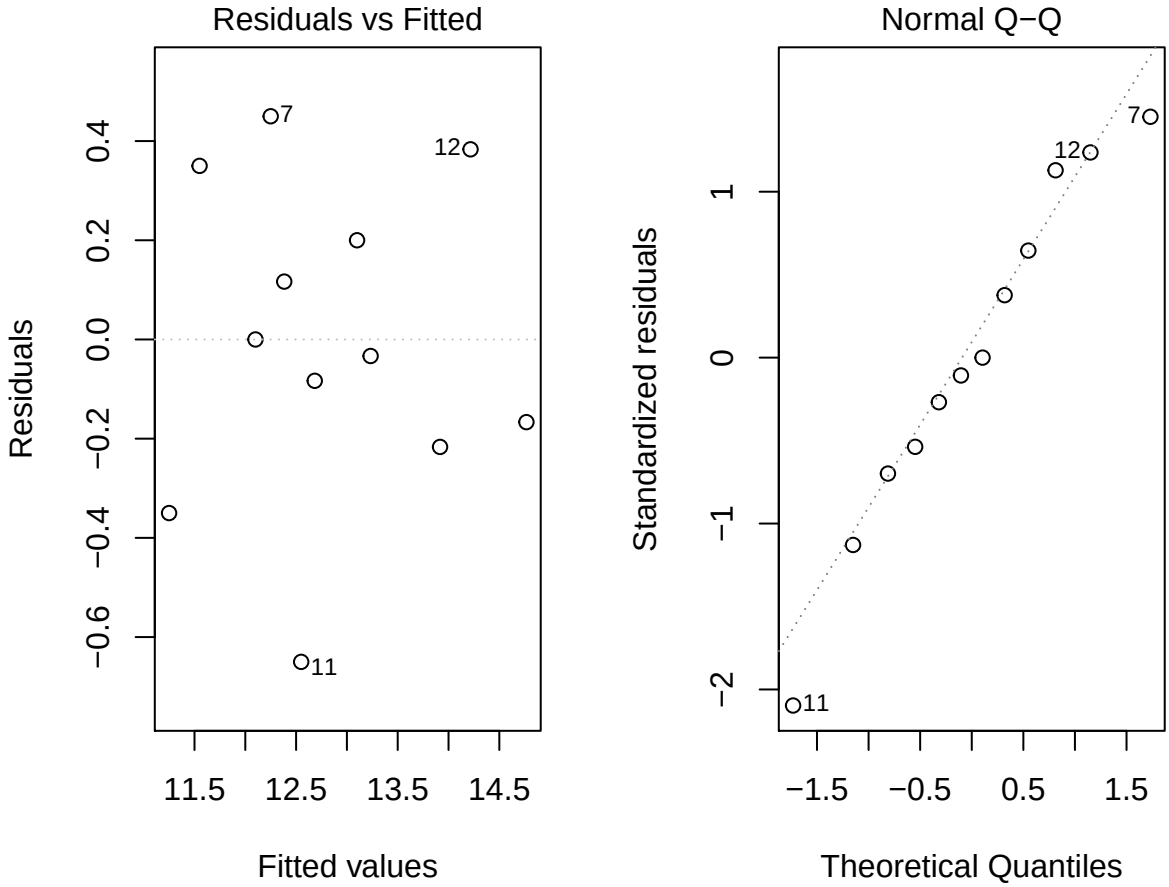
(Intercept) $\hat{\mu}$	Sort(Bastian) $\hat{\tau}_1$	Sort(Berserk) $\hat{\tau}_2$	Field
12.833	0.467	-0.383	

  From the sum-to-zero assumptions, we can get $\hat{\tau}_3 = -(\hat{\tau}_1 + \tau_2) = -0.083$ and $\beta_4 = -(\beta_1 + \beta_2 + \beta_3) = 1.467$

  Similarly, the estimate for error variance is given by MS_E which is 0.192

- Can you prove sort effect?
- Find the fitted values and the residuals for all observations.
- Check the model assumptions.

  We can check model assumption from basic diagnostic plots for the fitted model.

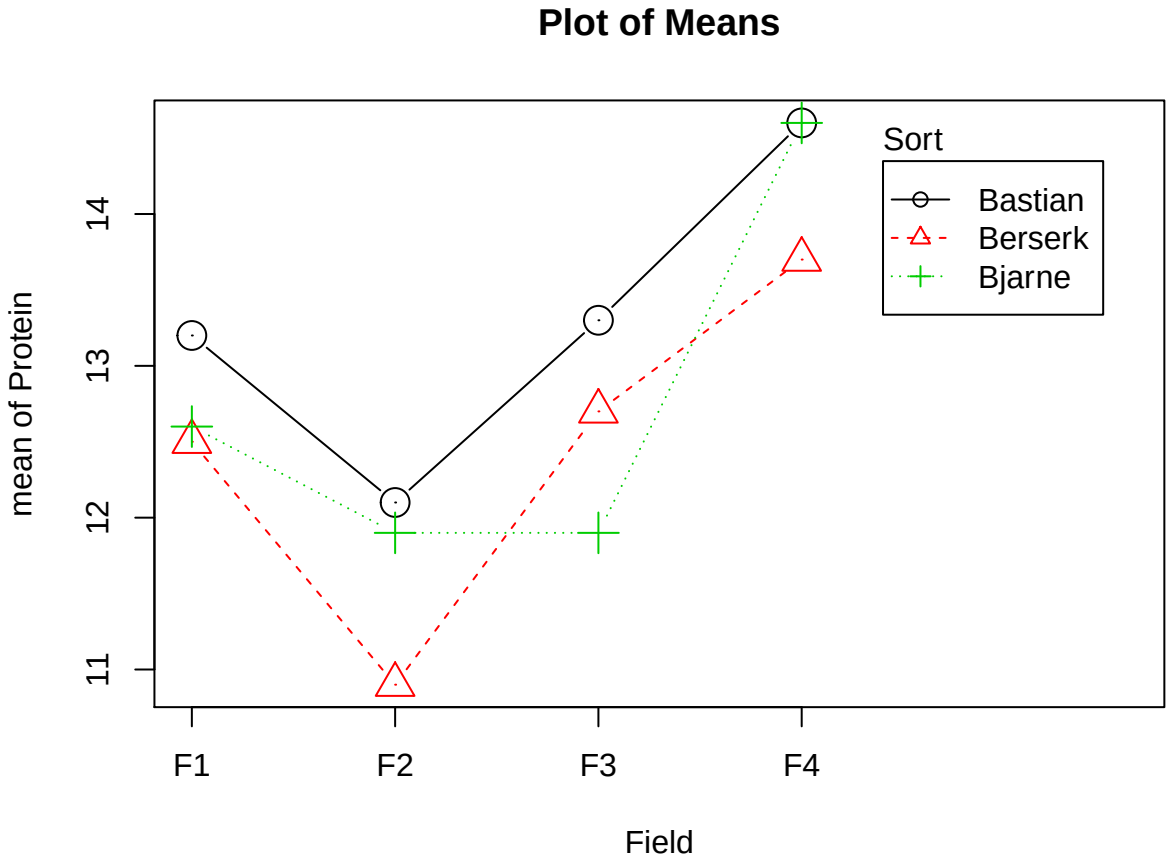


Plot shows that the residual terms are scattered randomly and constant over different fitted values. This means the assumption of constant residual variance holds true. Further, in the normal q-q plot, points are aligned with the q-q line which fairly suggest that error terms are normally distributed.

- Explain the residual to a person without knowledge in statistics.

Use the model to show that the expected differences between two sorts are independent on the block. (In block experiments it is unusual to include interaction.)

Does the plot below (Graphs > Plot of means) support the statement above?



The model for a block experiment is,

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

Here, the following assumption for a fixed effect models holds.

$$\epsilon_{ij} = \text{NID}\left(0, \sigma^2\right)$$
$$\sum_{i=1}^{a} \tau_i = 0 \text{ and } \sum_{j=1}^{b} \beta_j = 0$$

Since, $\text{E}(y_{ij}) = \mu + \tau_i + \beta_j$, the expected difference between two sorts can be written as,

$$\begin{aligned} \text{E}\left(y_{ij} - y_{kj}\right) &= \mu + \tau_i + \beta_j - \mu - \tau_k + \beta_j \\ &= \tau_i - \tau_k \end{aligned}$$

which is independent of block. In other words, the expected difference between sorts is same irrespective of which block they belong. In the following plot sorts `Bastian` and `Berserk` holds this statement, however it is not very true for sort `Bjarne`.

Assume incorrectly a One Way Analysis of Variance model (sort as factor).

- Why is this incorrect?
- Compare the SSError, the degrees of freedom and the estimate of the variance for this model and the model including blocks.

If you had only the output from the block model, could you find the SSError from the One Way Analysis of Variance model.

Exercise 6

Load `sortsoil.Rdata`. The dataset discussed on friday's lecture,

- Calculate the mean in each cell (by hand)

  The average in each cell along with column mean, row mean and overall mean is,

Sort	Clay	Sand	(all)
A1	515.0	685.0	600
A2	560.0	400.0	480
(all)	537.5	542.5	540

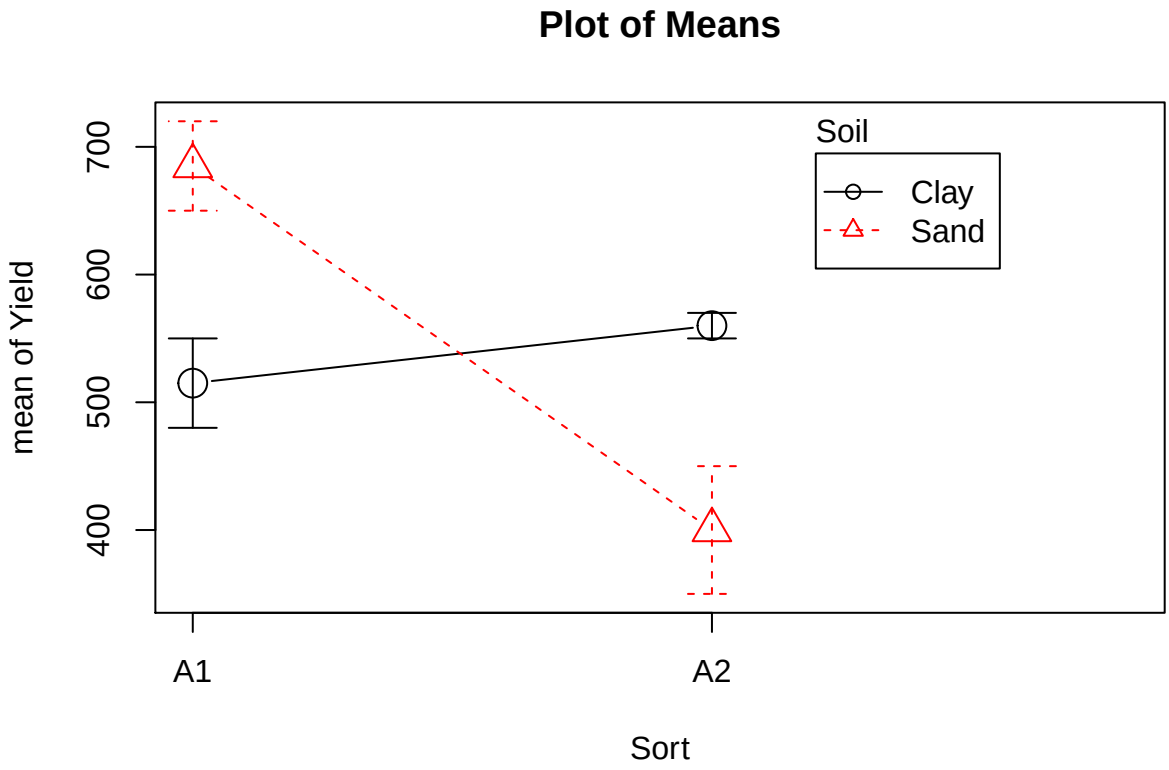
  Calculate it by and hand, answer above are just for checking

- Regard the 4 means, do those support that interaction is present?

  Averages in previous answer suggests interaction is present. Here, if you choose sort `A2` instead of sort `A1`, there is an increase in `Yield` for soil type `clay` while `Yield` decreases for soil type `Sand`. Similary, if you planned to cultivate a sort in `sand` rather than `clay`, the change in average `Yield` is different for Sort `A1` and `A2`.

- Have we constructed a sand-sort and a clay-sort?

  The averages can also be visualized by the mean plot below:



Apply a model with main effect and interaction effect, call this Model 1.

The model can be written as,

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ij}, \text{ where, } \epsilon_{ij} \sim \text{NID}(0, \sigma^2)$$

Since this is two-factor fixed effect model, the assumption of sum-to-zero effect are,

$$\sum_{i=1}^{a} \tau_i = 0 \quad \sum_{j=1}^{a} \beta_j = 0 \quad \sum_{i=1}^{a} (\tau\beta)_{ij} = \sum_{j=1}^{b} (\tau\beta)_{ij} = 0$$

- Test if interaction is present.

  A test hypothesis for testing interaction effect is,

  $$H_0 : (\tau\beta)_{ij} = 0, \text{for all } i \text{ and } j$$
  $$H_1 : \text{at least one } (\tau\beta)_{ij} \neq 0$$

  The ANOVA table for Model 1 is,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sort	1	28800	28800	11.405941	0.0278557
Soil	1	50	50	0.019802	0.8948935
Sort:Soil	1	54450	54450	21.564356	0.0097068
Residuals	4	10100	2525		

Low p-value corresponding to `Sort:Soil` rejects the null hypothesis, thus interaction is significant and is present in the model.

- Estimate all expectation parameters in the model.

  The coefficient estimates for Model 1 are,

(Intercept) $(\hat{\mu})$	Sort(A1) $(\hat{\tau}_1)$	Soil(Clay) $(\hat{\beta}_1)$	Sort(A
540	60	-2.5	

  However, the output does not contain estimates for τ_2, β_2, $(\tau\beta)_{12}$, $(\tau\beta)_{21}$, $(\tau\beta)_{22}$. These estimates can be obtained using sum-to-zero assumption for fixed effect model. These estimates are,

$$\begin{aligned}
\hat{\tau}_2 &= -\hat{\tau}_1 = -60 \\
\hat{\beta}_2 &= -\hat{\beta}_1 = 2.5 \\
\left(\hat{\tau}\hat{\beta}\right)_{12} &= -\left(\hat{\tau}\hat{\beta}\right)_{11} = 82.5 \\
\left(\hat{\tau}\hat{\beta}\right)_{21} &= -\left(\hat{\tau}\hat{\beta}\right)_{12} = \left(\hat{\tau}\hat{\beta}\right)_{11} = -82.5 \\
\left(\hat{\tau}\hat{\beta}\right)_{22} &= -\left(\hat{\tau}\hat{\beta}\right)_{21} = -\left(\hat{\tau}\hat{\beta}\right)_{11} = 82.5
\end{aligned}$$

- Estimate and give an interpretation of σ^2

  The estimate of σ^2 is MSE, i.e.

$$\hat{\sigma}^2 = \mathrm{MS}_{\mathrm{E}} = 2525$$

  Interpretation: MSE is the estimate of variance of error terms ϵ_{ij}. It is the expected variation present in response (`Yield`) for any given `sort` and `soil`.

- Find the fitted values and the residuals.

  Fitted values are obtained as, $\hat{y_{ij}} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j + \left(\hat{\tau}\hat{\beta}\right)_{ij}$. It is the average yield for each combination of `sort` and `soil`.

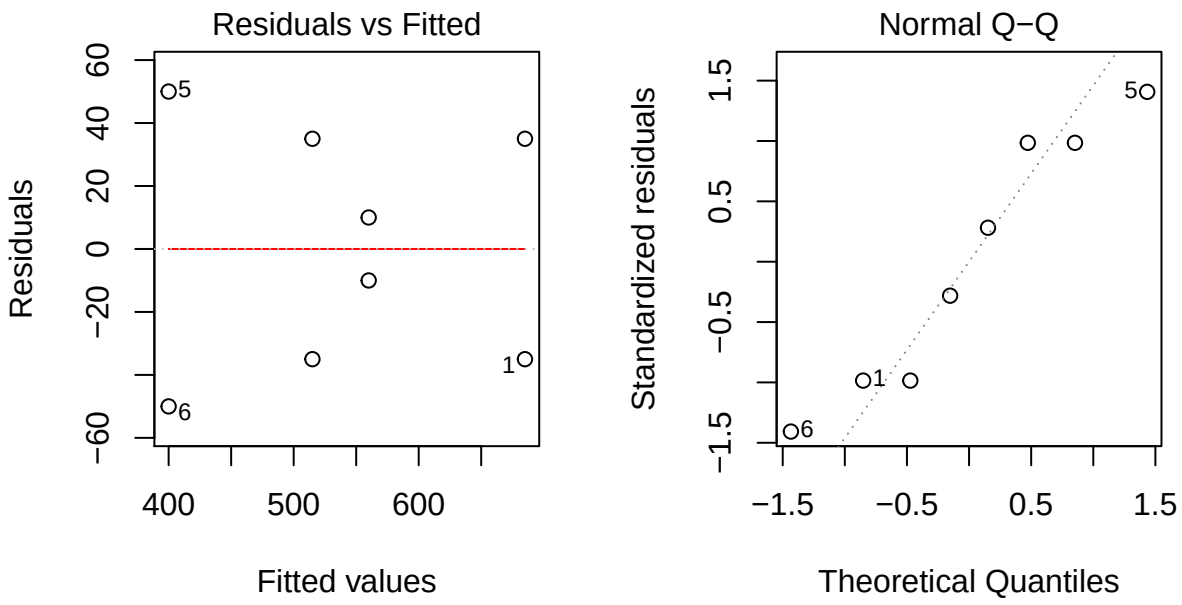
  Residuals are variation present in data that our model is not able to explain. It is the difference between the actual value and the fitted value. i.e, $e_{ij} = y_{ij} - \hat{y}_{ij}$

  Rcomander:model > Add observation statistics to data

- Find the standardized residuals, are some too large (absolute value).

- Check the model by suitable graphics.

  We can check model assumptions using basic diagnostic plots (Rcomander: Model>Graphs>Basic Diagnositic Plot). The first two plots from Basic diagnostic plot for Model 1 is as follows,



Interpreate the plot yourself

Assume a model without interaction, call this Model 2

- Compare SSE in both models.

	Model 1	Model2
SSE	10100	64550

Sum of square of residuals in Model 2 (without interaction term) is very high compared to Model 1. This also show us that without interaction term model error increases significantly. In addition, this also supports the significance of interaction term.

- Find the fitted values and the residuals in Model 2 and compare the models.

  See how fitted values and residuals are dependent on how you specify your model. Both of these values depends on your model.

- Check the model assumptions in model 1, can you see any problems. (Yes, there is a problem).

Below you will find data from an identical experiment, but with one large difference. Carefully inspect the data, what is the difference? Repeat the analysis with these data. Why do the p-values change? Data is saved with the name `sortsoil2.rdata`.

Yield	sort	soil
369	1	1
359	1	2
340	2	1
492	2	2
369	1	1
361	1	2
336	2	1
497	2	2
369	1	1
360	1	2
338	2	1
493	2	2

The ANOVA table for `sortsoil2` data is,

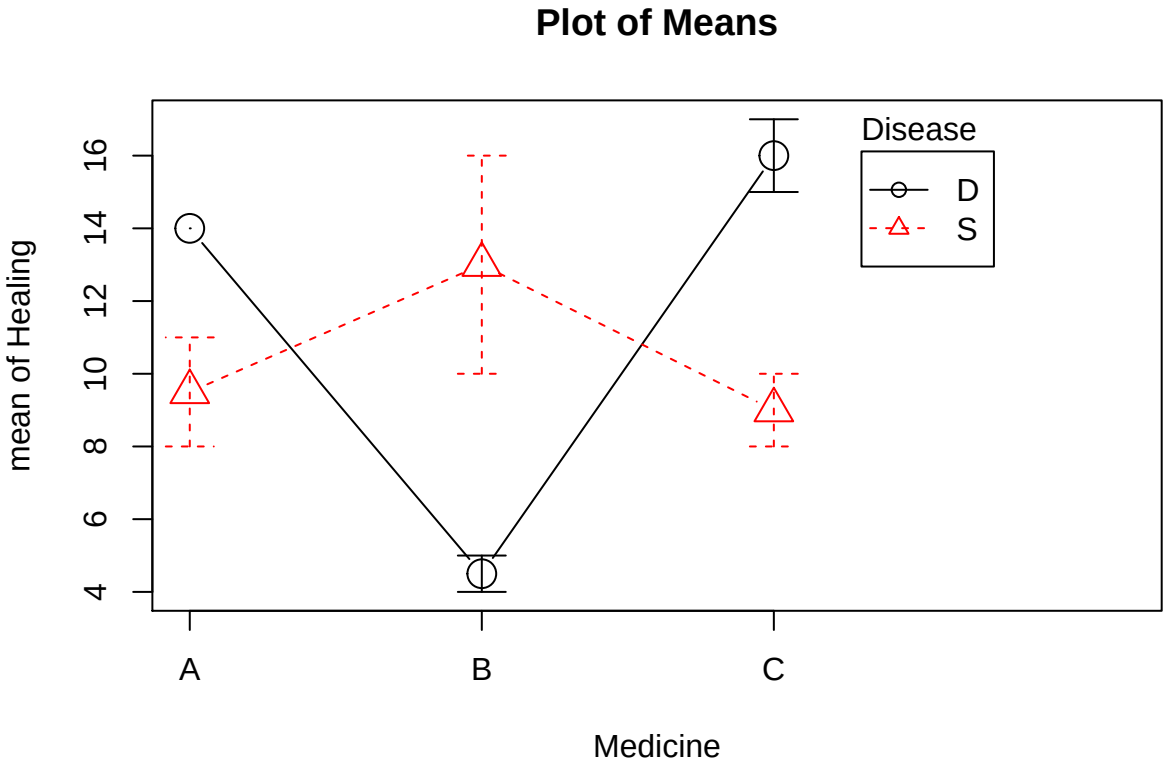
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sort	1	7956.75	7956.75	2652.25	0
soil	1	16206.75	16206.75	5402.25	0
sort:soil	1	20418.75	20418.75	6806.25	0
Residuals	8	24.00	3.00		

It is clear that all factor including interaction terms are highly significant in this model. Noise (measured by MSE) is very small as compared to Model 1, so that the variation present in different factors and interactions are clearly visible and thus they are highly significant.

Week Three

Exercise 1 (Hospital Data)

- Load the `Hospital.Rdata`
- Construct an interaction effect plot. (Graphs > Plot of means)



- Apply a model with Healing as response, Disease, and Medicine as factor where interaction is included, call this `Model1`.
- Test if interaction is present.
- Estimate all parameters.
- Split the first observation into 5 parts: constant, (sample) effect of disease, (Sample) effect of medicine, (Sample) effect of interaction and residual Show that the sum of the 4 effects gives the sample mean in the level combination the observation belongs.
- Repeat the analysis without including interaction, call this `Model2`.
- Find and store the fitted value and the residuals for both model and compare.
- Compare the Sum of squares for the two different models.
- If you should test on main effects for in `Model2` without use of computer, show how you can use the results from the `Model1` relatively easy.
- Plot the residuals (y-axis) against the fitted value (x-axis) for both models and show how this plot reveals model problems in `Model2`.
- A student that had not heard about 2-factor design tried to investigate the effect of medicine by a One Way ANOVA model (`Model3`). Explain why this is incorrect for this data.

- For all 3 models find R-square, and see how this increases with complexity.

$$\text{R-square} = \frac{\text{SSModel}}{\text{SS}_{\text{Total}}} = 1\frac{\text{SSE}}{\text{SS}_{\text{Total}}}$$

Exercise 2 (Mussels Data)

- Load the mussels data.
- Assume a model where eatable is the response size and season are the factors.
- Assume no interaction. Store the residuals and the fitted values.
- Plot residual against fitted values and make comments.
- Give an interpretation of the replicate variance (σ^2). Estimate σ^2 and construct a 95% confidence interval for the true value.
- Assume a model including interaction. Store the residuals and the fitted values.
- Plot residual against fitted values and make comments.
- Give an interpretation of the replicate variance (σ^2) and construct a 95% confidence interval for the true value.

Exercise 3

Load the mussels data. The data is from a commercial farming of Blue Mussels where response is eatable food after boiling in percent of total weight. The mussels are sorted after size (Large and small) and season they were harvested (summer or autumn).

Assume a model where eatable is the response, size and season are the factors and interaction is included. Analyze the model.

The model we are analysing is,

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \tag{5}$$

Where,

i	=	1 (small), 2 (large) corresponding to size
j	=	1 (autumn), 2 (summer) corresponding to season
k	=	$1, 2, \ldots, 6$ corresponding to replicates
y_{ijk}	=	eatable percentage of total weight of k^{th} mussels of size i on farmed during j season
ϵ_{ijk}	$\sim$	NID$(0, \sigma^2)$, $\sum_{i=1}^{2} \tau_i = 0$ and $\sum_{j=1}^{2} \beta_j = 0$

In addition,

$$\sum_{i=1}^{2}(\tau\beta)_{ij} = \sum_{j=1}^{2}(\tau\beta)_{ij} = 0$$

The ANOVA table for model in equation ((5)) is,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
season	1	294	294.0	122.5	0
size	1	96	96.0	40.0	0
season:size	1	294	294.0	122.5	0
Residuals	20	48	2.4		

Here, both factors `season` and `size` along with their interactions are significant (low p-value). Thus, we can conclude the model contains interaction between `size` and `season`.

The farm decide to harvest in summer, and they wanted to find out if there are significant positive effect of harvesting large mussels compare to small mussels. Answer the question by a t-test. State the null hypothesis and the alternative hypothesis.

Here we have average percent of total weigth of eatable mussels by size and summer is,

season	liten	stor	(all)
host	24	13	18.5
sommer	24	27	25.5
(all)	24	20	22.0

The hypotheis to test if there is positive effect of harvesting large mussels than small mussels in summer season is,

$$\begin{aligned} H_0 &: \mu_{\text{large:summer}}\ (\mu_{22}) = \mu_{\text{small:summer}}\ (\mu_{12}) \\ H_1 &: \mu_{\text{large:summer}}\ (\mu_{22}) > \mu_{\text{small:summer}}\ (\mu_{12}) \end{aligned}$$

Note that:

$$\begin{aligned} \mu_{12} - \mu_{22} &= \mu + \tau_2 + \beta_2 + (\tau\beta)_{22} - \mu - \tau_1 - \beta_2 - (\tau\beta)_{12} \\ &= \tau_2 - \tau_1 + (\tau\beta)_{22} - (\tau\beta)_{12} \end{aligned}$$

Due to interaction term, the difference in mean is dependent on both of the factors. We can use the difference of their respective estimates, i.e, $\bar{y}_{12} - \bar{y}_{22} = 27 - 24 = 3$.

Further,

$$\begin{aligned}\text{var}\left(\bar{y}_{12}-\bar{y}_{22}\right) &= \frac{\sigma^2}{6}+\frac{\sigma^2}{6}=\frac{\sigma^2}{3}\\ \text{SE}\left(\bar{y}_{12}-\bar{y}_{22}\right) &= \sqrt{\frac{\hat{\sigma}^2}{3}}\\ &= \sqrt{\frac{\text{MSE}}{3}}=0.894\end{aligned}$$

Therefore, the test-statistics,

$$t=\frac{\left(\bar{y}_{12}-\bar{y}_{22}\right)}{\text{SE}\left(\bar{y}_{12}-\bar{y}_{22}\right)}=\frac{3}{0.894}=3.354\sim t_{\alpha,N-a}$$

At 95% confidence level, $t_{\alpha,N-a}=t_{0.05,2.4}=2.614$

Since, $3.354 > 2.614$, we reject Null hypothesis and conclude that the farm can make more profit by farming larger mussels than in smaller mussels in summer.

Exercise 4

Load hospital data.

Apply a model with Healing as response, Disease, and Medicine as factors and interaction is included. Test if interaction is present.

The model is,

$$y_{ijk}=\mu+\tau_i+\beta_j+(\tau\beta)_{ij}+\epsilon_{ijk}$$

where,

i	=	1(D), 2(S) corresponding to `Disease`
j	=	1(A), 2(B), 3(C) corresponding to `Medicine`
i	=	1, 2 corresponding to replication
y_{ijk}	=	Healing of `Disease` i due to `Medicine` j
ϵ_{ijk}	$\sim$	NID $\left(0,\sigma^2\right)$

The ANOVA table for the model is,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Medicine	2	31.5	15.75	3.500	0.098
Disease	1	3.0	3.00	0.667	0.445
Medicine:Disease	2	138.5	69.25	15.389	0.004
Residuals	6	27.0	4.50		

The table shows that the interaction is significant (low p-value corresponding the interaction). Thus we can conclude that the interaction is present in the model.

Execute Tukey tests on the pairs of medicine for patients having Depression and for patients having schizophrenia.

The average of factors and their combinations are as follows,

Disease	Medicine:A	Medicine:B	Medicine:C	Medicine:(all)
D	14.00	4.50	16.0	11.5
S	9.50	13.00	9.0	10.5
(all)	11.75	8.75	12.5	11.0

Thus, we have the difference in mean between the pairs of Medicine i.e, AB, AC and BC for patients having Depression and Schizophrenia as,

Schizophrenia

$$\begin{aligned}\hat{\mu}_{A:S} - \hat{\mu}_{B:S} &= -3.5\\ \hat{\mu}_{A:S} - \hat{\mu}_{C:S} &= 0.5\\ \hat{\mu}_{B:S} - \hat{\mu}_{C:S} &= 4\end{aligned}$$

Depression

$$\begin{aligned}\hat{\mu}_{A:D} - \hat{\mu}_{B:D} &= 9.5\\ \hat{\mu}_{A:D} - \hat{\mu}_{C:D} &= -2\\ \hat{\mu}_{B:D} - \hat{\mu}_{C:D} &= -11.5\end{aligned}$$

Further, the Tukey test criteria is,

$$T = q_{\alpha,(a,f)}\sqrt{\frac{\text{MSE}}{n}} = q_{0.05,(3,6)}\sqrt{\frac{4.5}{3}} = 4.339 \times \sqrt{1.5} = 5.314$$

Medicine A and C are new medicine. Test if those are better than B for patients suffering from Depression.

Exercise 5

A student wanted to study how a medicament influenced feed consumption on rats. She included sex and the fact that some of the rats were sterilized in her study. In addition half of the animals were given placebo. In total 40 animals were available, and the design is balanced.

Sex	Sterilized	Medicament	Mean	Placebo	Mean
Male	No	21.46, 23.92, 22.56, 16.12, 21.48	21.108	25.64, 28.84, 26.00, 26.02, 23.24	25.948
Male	Yes	15.44, 23.54, 23.52, 17.96, 19.02	19.896	22.50, 24.48, 25.52, 24.76, 20.62	23.576
Female	No	18.58, 15.44, 16.12, 16.88, 17.58	16.920	17.82, 15.76, 12.96, 15.00, 19.54	16.216
Female	Yes	18.20, 14.56, 15.54, 16.82, 14.56	15.936	19.74, 17.48, 16.46, 16.44, 15.70	17.164

This model was applied:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl}$$

Where,

i	=	1 (Active Medicament) or 2(Placebo)
j	=	1 (Male) or 2 (Female)
k	=	1 (Non sterilizes) or 2 (Sterilized)
l	=	1, 2, ..., 5 corresponding to replications

a) Give the Common assumptions concerning ϵ_{ijkl}. Explain why these are necessary in the model.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Sex	1	368.813	368.813	69.703	0.000
Sterilized	1	8.190	8.190	1.548	0.222
Medicament	1	51.121	51.121	9.662	0.004
Sex:Sterilized	1	7.868	7.868	1.487	0.232
Sex:Medicament	1	39.960	39.960	7.552	0.010
Sterilized:Medicament	1	0.372	0.372	0.070	0.792
Sex:Sterilized:Medicament	A	B	C	D	0.296
Residuals	E	169.318	5.291		
Total	F	651.618			

`S = 2.30026 , R-Sq = G%`

b) Find the missing values `A`, `B`, `C`, `D`, `E`, `F` and `G`.

A: Degree of freedom of 3 factor interaction = 1

B: Sum of square of 3 factor interaction

$$\text{SS}_{\text{Total}} - \text{ sum of all other degree of freedom } = 5.975$$

C: Mean sum of square of 3 factor interaction

$$\text{MS}_{\alpha\beta\gamma} = \frac{\text{SS}_{\alpha\beta\gamma}}{df_{\alpha\beta\gamma}} = 5.975$$

D: F statistic corresponding to 3 factor interaction

$$\frac{\text{MS}_{\alpha\beta\gamma}}{\text{MS}_{\text{residual}}} = 1.129$$

E: Error Degree of freedom $= abc(n-1) = 32$
F: Total Degree of freedom $= (N-1) = 39$
G: R-squared (variation explained by the model)

$$\text{R-sq} = 1 - \frac{\text{SS}_{\text{E}}}{\text{SS}_{\text{Total}}} = 74.02$$

---

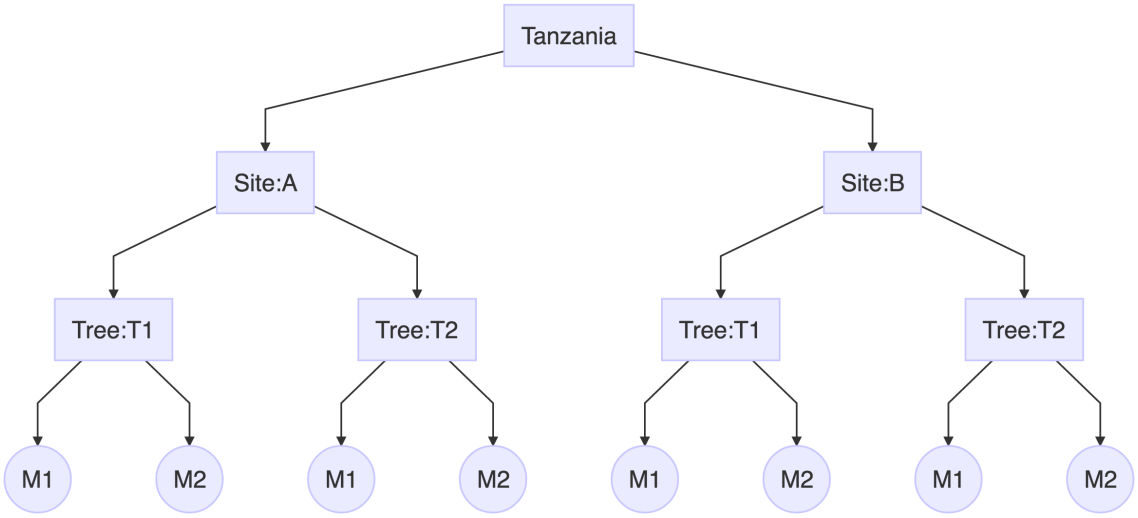
c) What important conclusions can be detected?

Exercise 6

Mango is an excellent source for vitamin C. Assume content of vitamin C in mango are measured on 2 sites in Tanzania, A in the highland, B in the lowland. In each site 2 trees are randomly chosen and 2 fruit on each tree are also randomly chosen. Data is stored in the file called `Mango.Rdata`

- Explain why this must be treated as a nested design.

  Here, trees are choosen after the sites are selected. It is impossible to see the effect of same trees in both lowland and highland. The trees selected at lowland is different from the trees selected at highland. and the measurements are made on mangos form each of the trees (replicated twice). Thus this must be a nested design with the following hierarchical structure.



Here, `Tree:T1` from `Site:A` is different from `Tree:T1` from `Site:B` and so on. It is not possible observe the effect of `Tree:T1` of `Site:A` in `Site:B`.

- Explain why it is natural to regard tree as a random effect.

  Two specific trees can not be an interest of a research, rather a population of tree can be a subject of interest in the research. This research is more interested to see if there is any variation in vitamin C content in mango found in different places in Tanzania. So, it is natural to regard tree as a random effect.

Assume site to be a fixed effect State the model.

- Test if there is site effect, and if there is tree effect. Write down both the null hypothesis ant the alternative hypothesis.

  A model is,

$$y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{(ij)k}$$

  where, $i = 1, 2$ for site, $j = 1, 2$ for tree nested in site i and $k = 1, 2$ for two replication of mangos.

  The hypothesis to test the effect of tree is,

$$H_0 : \sigma^2_\beta = 0$$
$$H_1 : \sigma^2_\beta > 0$$

	Mean Sq	Sum Sq	Df	F value	Pr(>F)
site	8.0	8	1	1.000	0.423
site:tree	8.0	16	2	1.455	0.335
Residuals	5.5	22	4		

Here, neither the site or the trees nested on site are significant. There is no evidence of tree effect in this model.

- Estimate both variance parameters.

  From the fitted model, we have

```
            Err.term(s) Err.df VC(SS)
1 site              (2)      2  fixed
2 site:tree         (3)      4   1.25
3 Residuals           -      -   5.50
(VC = variance component)
```

  From the output above, the estimates of variance parameters σ^2_β and σ^2 are 1.25 and 5.5 respectively.

  In addition, we can also find variance components using mean sum of squares as,

$$\begin{aligned}\hat{\sigma}^2 &= \mathrm{MS}_E = 5.5\\ \hat{\sigma}^2_\beta &= \frac{\mathrm{MS}_{\mathrm{B(A)}} - \mathrm{MS}_E}{n}\\ &= \frac{8 - 5.5}{2} = 1.25\end{aligned}$$

- The researchers concluded that the variance in Vitamin C is larger inside the trees than between the trees. Do you agree?

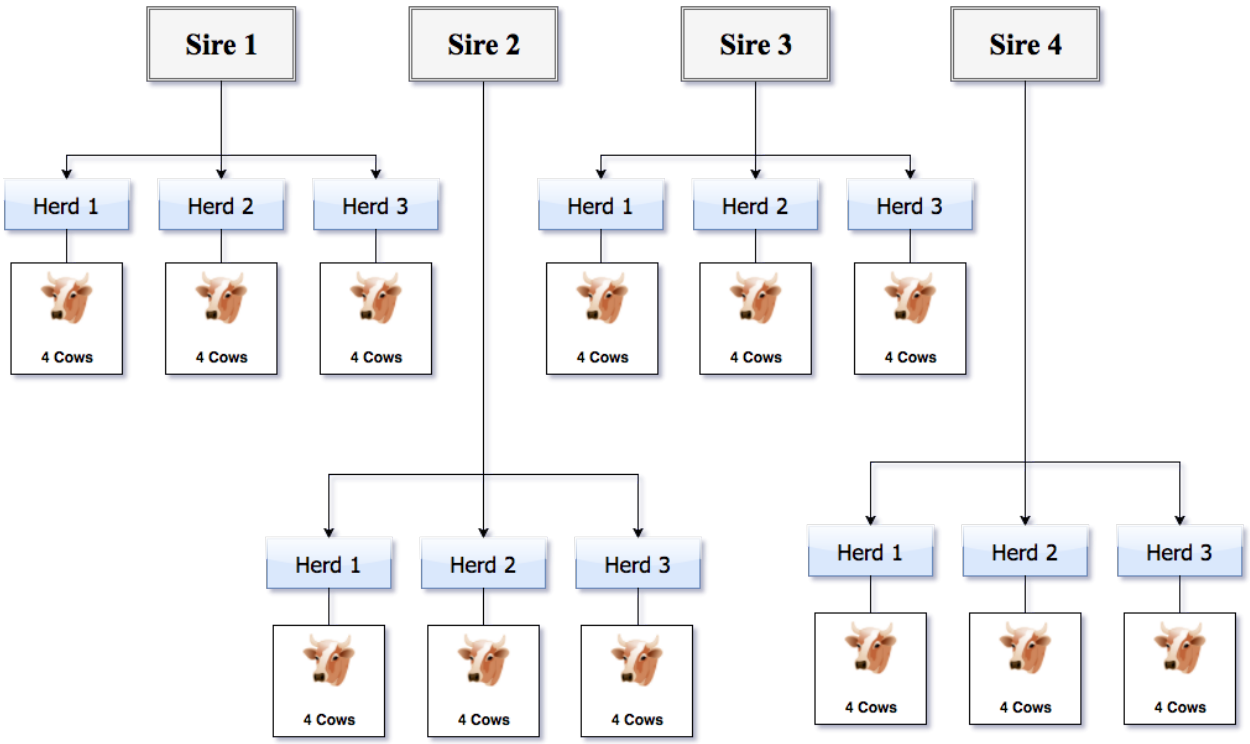
  The estimate of variance components $\hat{\sigma}^2 = 5.5$ refers to the variance within the trees and $\hat{\sigma}^2_\tau = 1.25$ refers to the variance between the trees. Here the variance witin the tree is larger than between the tree, the conclusion of the researcher is true. However statistical testing of this difference is out of the scope of this course.

Exercise 7

In an experiment 4 sires were used. Each sire was tried in 3 herds (different herds for each sire). From each herd 4 cows were randomly picked out. Data are saved as `Nestedsires.Rdata`

If we assume a nested design (herd nested in sire) draw a figure of the design. How many herds are included in the experiment? How many observations do we have?

The nested design in this experiment can be visualized as,



Here, `Herd` is nested in `Sire` so, `Herd1` in `Sire1` is different from `Herd1` in `Sire2`. Therefore, there are 12 herd included in the experiment. Since 4 cows are randomly selected from each herd, there are 48 observations in this experiment.

- State the model you will use for this experiment.

  The model in this experiment is,

$$y_{ijk} = \mu + \tau_i + \beta_{i(j)} + \epsilon_{(ij)k}$$

  Where,

i	=	1, 2, ..., 4 corresponding to `sire`
$j(i)$	=	1, 2, 3 corresponding to `herd` nested in `sire` i
$\epsilon_{(ij)k}$	$\sim$	NID $\left(0, \sigma^2\right)$ where $k = 1, 2, 3, 4$ for each replication (`cow`)

- How do you explain the deviations (ϵ's) in this experiment?

  Deviations or Error ($\epsilon_{(ij)k}$) is the difference between observed response and expected response. i.e., $\epsilon_{(ij)k} = y_{ijk} - \mathrm{E}\left(y_{(ij)k}\right)$. Here the expected response is the value you expect to observe in population for the given model. The residuls, on the other hand, is fitted deviations and are the difference between the observed values and the fitted values for the given sample.

- Assume sire is a fixed factor and herd a random factor. How is the F value calculated

for testing Herd effect and for testing Sire effect?

When Factor A is fixed and Factor B is random, we have following expression,

Now Use R-commander:

- Test for sire and for herd effect.
- Estimate the variance components.
- Explain to the farmers what these estimates are saying.

The farmers are sometimes asked to follow courses in the purpose of producing more milk. If the variance component for herd was extremely large, compare to the variance component for the error, would this support the opinion that we need better courses for the farmers obtaining the poorest results?

Repeat, but assume sire to be a random effect. Estimate all 3 variance components. Which variance seems to be most important? How do you estimate the variance in milk production among the population of cattle's?

Exercise 8 (Cheese Production)

Data for Model 1 is called *cheese1.Rdata* and for Model 2 is called *cheese2.Rdata*

Bacterial fermentation is an important part of cheese making and ripening, and specific starter cultures (bacteria) are used. However, some other bacteria may also be added in the milk, they are not starters, but can have huge influence on the final quality of the cheese.

We want to study the effect of two types of non-starters, we call them R50 and R21.

Both factors (bacteria type) have 2 levels, absent (level 1) or present (level 2). The response is the total content of free amino acid which we want as high as possible. Each combination was replicated 3 times.

The first model we use is:

$$\text{Model 1: } y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}$$

τ_i is the effect of R50 on level i, β_j is the effect of R21 of level j, $(\tau\beta)_{ij}$ is the effect of interaction between R50 and R21 on combination ij. Where, $i = 1,2 \;\; j = 1,2 \;\; k = 1,2,3$

The ϵ_{ijk}'s are independent and normally distributed with expectation 0 and variance σ^2. Also assume,

$$\sum_{i=1}^{2} \tau_i = \sum_{j=1}^{2} \beta_j = \sum_{i=1}^{2} (\tau\beta)_{ij} = \sum_{j=1}^{2} (\tau\beta)_{ij} = 0$$

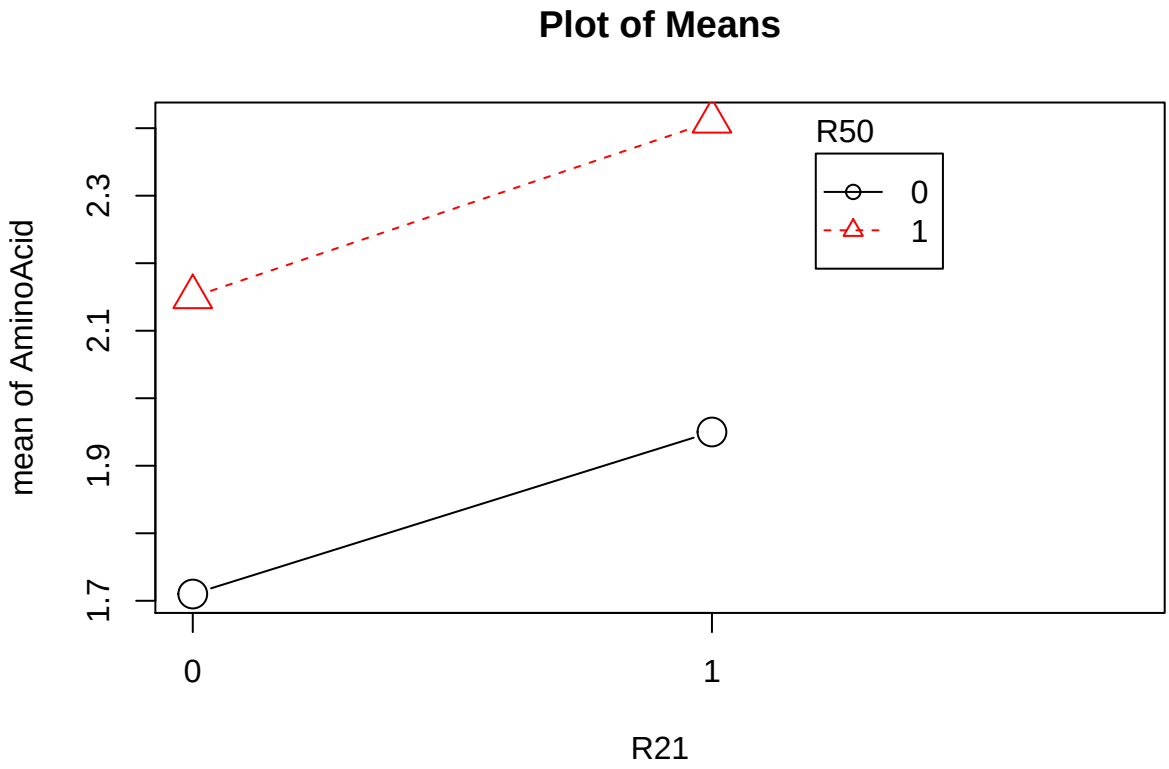
a) Use this experiment to discuss the difference between fixed and random factors.

Use this experiment to explain what we mean by interaction.

Use this experiment to explain what we mean by a balanced design.

How would you interpret σ^2?

b) Draw an interaction plot, (see Table 11).



In Table 12 are parts of an ANOVA table, finish the table and test if interaction is present.

The complete ANOVA table is,

	Sum Sq	Df	F value
R50	0.610	1	10.043
R21	0.189	1	3.102
R50:R21	0.000	1	0.003
Residuals	0.486	8	

Why is it natural to exclude interaction from Model 1?

Sum of square corresponding interaction term `R50:R21` is almost zero so is the F-value. This consequently suggest that the interaction term is not significant and thus it is natural to exclude interaction from Model 1.

Name the model without interaction Model 1b. Construct the ANOVA table for Model 1b.

The model with out interaction term is,

Table 11: Sample means for each combination.

	0	1
0	1.71	2.15
1	1.95	2.41

Table 12: Parts of ANOVA table Model 1

	Sum Sq	Df	F value
R50	0.610		
R21	0.189		
R50:R21	0.000		
Residuals	0.486		

Model 1b: $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$

ANOVA table for this model is,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
R50	1	0.610	0.610	11.294	0.008
R21	1	0.189	0.189	3.489	0.095
Residuals	9	0.486	0.054		

Here, sum of square of Residuals has not increased than in Model 1, i.e. the additional noise in the model due to removal of interaction term is very small (almost zero).

Later a student realized that amino acid varied within each cheese, so she divided each cheese into 3 parts and regarded each part as replicate. The student had 4 different bacteria cultures in her experiment. The following nested model was applied:

Model 2: $y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \epsilon_{ijk}$

Where, $i = 1, 2, 3, 4 \;\; j = 1, 2, 3 \;\; k = 1, 2, 3$

τ_i is effect of bacteria group i, $\beta_{j(i)}$ is effect of cheese j nested in group i.

The $\beta_{j(i)}$'s are independent and normally distributed with expectation 0 and variance σ^2_β.

The ϵ_{ijk}'s are independent and normally distributed with expectation 0 and variance σ^2.

The $\beta_{j(i)}$'s and ϵ_{ijk}'s are independent.

In addition,

Table 13: Some results from **Model 2**

	Mean Sq	Sum Sq	Df	F value	Pr(>F)
Bacteria	0.086	0.259	3		
Bacteria:Cheese	0.068	0.541	8		
Residuals	0.010	0.242	24		

$$\sum_{i=1}^{4} \tau_i = 0$$

c) Use this experiment to explain the difference between designs with nested and crossed factors.

Why is it logical to assume cheese as a random factor?

How would you explain σ^2_β and σ^2 to the scientist?

```
                Expected mean squares
Bacteria         (3) + 3 (2) + 9 Q[1]
Bacteria:Cheese  (3) + 3 (2)
Residuals        (3)
```

d) Estimate both variance components in Model 2 (Table 13). Test if there are effects of cheese and of bacteria group.

e) The student was told that it is better to use the average of the 3 parts of the cheese as replicate.

Reformulate Model 5 for this situation.

What is the response now?

Test if there is effect of bacteria group.

Exam 2011 (Sep 5, 2011)

The problems are based on the simulated data of Appendix 1. There are 24 observations and the variables are as follows:

- For Three types of feed: `For1`, `For2` og `For3`.
- Frittgående The variable is `Ja` for freely moving chickens and otherwise `No`.
- Besetning Lifestock. The variable indicates which of 6 lifestock (or farms for that matter) the chicken comes from.
- Vekt Weight in gram (g).

Exercise 1

Only variables Besetning and Vekt are used in this exercise. Consider,

$y_{ij} = \mu + \tau_i + \epsilon_{ij}$ where,
$i = 1, 2, \ldots, 6$ corresponding to `Bes1,Bes2,...,Bes6`
$j = 1, 2, \ldots, 4,$ corresponding to the four observations of each Besetning
$y_{ij} =$ '*Vekt*', $\tau_i \sim \mathrm{N}(0, \sigma_\tau^2)$, $\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$

The random variables on the right hand side of the model equation are assumed to be independent.

a) Use the output in Appendix 2 to test at 5% significance level the null hypothesis

$$H_0 : \sigma_\tau^2 = 0$$

Formulate a conclusion.

The p-value is 0.176 and therefore we cannot reject the null hypothesis. In other words, we cannot claim a significant effect of Besetning.

b) Use the output in Appendix 2 to estimate σ^2. Construct a 95% confidence interval for σ^2. What is the 95% confidence interval for the standard deviation (σ)? Interpret the confidence interval for σ.

Here, we have, $\hat{\sigma}^2 = \mathrm{MSE} = 605.722$ and degree of freedom corresponding to residual is 18. Therefore $\mathrm{SSE} = \mathrm{MSE} \times df = 10902.996$

From χ^2 table, we have $\chi^2_{0.975,18} = 8.231$ and $\chi^2_{0.025,18} = 31.526$

Thus, the 95% confidence interval for σ^2 is,

$$\left[\frac{\mathrm{SSE}}{\chi^2_{0.025,18}}, \frac{\mathrm{SSE}}{\chi^2_{0.975,18}}\right] = [345.84, 1324.67]$$

Further, 95% confidence interval for standard deviation σ can be obtained by taking square root of confidence interval for σ^2, i.e. 18.6, 36.4.

Interpretation: We are 95% sure that the standard deviation is between 18.6 and 36.4.

c) The correlation coefficient for observations from the same Besetning is

$$\rho = \frac{\sigma_\tau^2}{\sigma_\tau^2 + \sigma^2}$$

Use the output in Appendix 2 to estimate ρ. Comment.

The estimate of correlation coefficient ($\hat{\rho}$) is obtained as,

$$\hat{\rho} = \frac{\hat{\sigma}_\tau^2}{\hat{\sigma}_\tau^2 + \hat{\sigma}^2} = \frac{112.3}{112.3 + 605.7} = 0.156$$

There are several ways to comment on this:

i) 15.6414899% of the total variability is between Besetning. Generally this number is between 0 and 100%. The value 0 would mean that the variable Besetning is irrelevant whereas the value 100% would imply no error, i.e., all variance explained by Besetning.

ii) The correlation in this model is always between 0 and 1 (or equivalently between 0 and 100%); here the correlation for animals from the same Besetning is estimated to 0.1564149

Regarding i) and ii) It is difficult to say that whether 0.1564149 or 15.6414899% is small or large and there is no definitive answer. Most would say that the correlation is rather low and one could add that numerator σ_τ^2 is not significantly greater than 0.

d) Mean `Vekt` is 925.6. Use this and the output of Appendix 2 to estimate μ and to calculate a 95% confidence interval for μ.

We have, $\hat{\mu} = 925.6$

Here, $a = 6$ and from t-table, $t_{0.025,a-1} = t_{0.025,5} = 2.571$. Hence, the 95% confidence interval for μ is,

$$\begin{aligned}
\hat{\mu} - t_{0.025,5}\sqrt{\frac{\text{MS}_{\text{Treatment}}}{N-a}} \leq \mu \leq \hat{\mu} + t_{0.025,5}\sqrt{\frac{\text{MS}_{\text{Treatment}}}{N-a}} \\
925.6 - 2.571\sqrt{\frac{1054.97}{24}} \leq \mu \leq 925.6 + 2.571\sqrt{\frac{1054.97}{24}} \\
908.56 \leq \mu \leq 942.64
\end{aligned}$$

Exercise 2

We now use the variables For, Frittgående and Vekt and consider the model,

$$
\begin{aligned}
y_{ijk} &= \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk} \text{ where,} \\
i &= 1,2,3 \text{ corresponding to } \texttt{For1,For2, and For3} \\
j &= 1,2, \text{ corresponding to } \texttt{Ja and Nei} \\
k &= 1,2,\ldots,4 \text{ corresponding to replications,} \\
y_{ijk} &= \texttt{Vekt, } \epsilon_{ijk} \sim \mathrm{N}(0,\sigma^2), \\
\sum_{i=1}^{3} \tau_i = 0 \; \sum_{j=1}^{2} \beta_j = 0 \; \sum_{i=1}^{3} (\tau\beta)_{ij} &= 0 \; \sum_{j=1}^{2} (\tau\beta)_{ij} = 0
\end{aligned}
$$

We assume ϵ_{ijk} to be independent.

a) Use the output in Appendix 3 to determine if there is a significant interaction between For and Frittgående by stating a null hypothesis and formulating a conclusion. Use 5% significance level. Use the output in Appendix 3 to determine if there is a main effect of the variable Frittgående by stating a null hypothesis and formulating a conclusion. Use 5% significance level.

   The null hypothesis of no interaction can be stated as,

$$H_0 : (\tau\beta)_{ij} = 0 \text{ for all} i \text{ and } j$$

   We do not reject as the p-value is greater than 0.05. A significant interaction would have meant that the effect of Frittgående depends on For and vice versa; there is no significant interaction.

   Next, the null hypothesis of no Frittgående effect is

$$H_0 : \beta_1 = \beta_2 = 0$$

   We do not reject as the p-value is greater than 0.05. Therefore, we cannot claim that there is a significant effect of Frittgående.

b) For the remaining part we use the model

$$
\begin{aligned}
y_{ij} &= \mu + \tau_i + \epsilon_{ij}, \text{ where,} \\
i &= 1,2,3, \text{ corresponding to } \texttt{For1,For2, and For3} \\
j &= 1,2,\ldots,8 \text{ corresponding to the eight observation for each } \texttt{For} \\
y_{ij} &= \texttt{Vekt, } \epsilon_{ij} \sim \mathrm{N}(0,\sigma^2),
\end{aligned}
$$

   We assume ϵ_{ij} to be independent.

   Use the output in Appendix 4 to show that there are significant differences between For at 5% significance level. Between which types of For are there differences? Use Tukey's test and the output of Appendix 4 for your answer.

The p-value is below 0.05 and we therefore reject. In other words, we can claim that there is a significant difference depending on For. Tukey demonstrates a significant effect only between `For3` og `For1`, p-value=0.03.

c) Use the output in Appendix 4 to estimate μ, τ_1, τ_2 and τ_3

Using Appendix 4, we have,

$\hat{\mu}$	=	925.58
$\hat{\tau}_1$	=	18.67
$\hat{\tau}_1$	=	-5.08

Using the assumption of sum-to-zero effect, i.e. $\sum_{i=1}^{3} \tau_i = 0$, we have,

$$\tau_1 + \tau_2 + \tau_3 = 0 \Rightarrow \tau_3 = -13.59$$

d) We would like to compare fortype `For1` against the average of `For2` and `For3`. Formulate a suitable contrast, compute a 95% CI and comment on the result.

The contrast can be formulated as,

$$\Gamma = \tau_1 - \frac{1}{2}(\tau_2 + \tau_3), \text{ i.e., } c = \left(1, -\frac{1}{2}, -\frac{1}{2}\right)$$

The estimated difference is,

$$\hat{\Gamma} = 18.67 - 0.5 \times -18.67 = 28.005$$

The variance of $\hat{\Gamma}$ is,

$$\text{var}\left(\hat{\Gamma}\right) = \frac{\sum_{i=1}^{3} c_i^2 \times \hat{\sigma}^2}{n} = \frac{836.25}{8} = 104.53$$

Therefore, standard error is,

$$\text{SE}\left(\hat{\Gamma}\right) = \sqrt{104.53} = 10.2239914$$

From t-table we have, $t_{0.025,21} = 2.08$, so,

95% confidence interval for Γ is,

$$\hat{\Gamma} \pm 2.08 \times \text{SE}\left(\hat{\Gamma}\right) = [6.74, 49.27]$$

Comment: The difference between `For1` and the average of the others is estimated to 28.005. There is 95% probability that the true difference is in the interval from 6.74 to 49.27.

e) Calculated predicted `Vekt` for a chicken fed on `For1`. Also calculate the residual for the first observation of Appendix 1.

The predicted `Vekt` and residual for the first observation is,

$$\text{Predicted} : \hat{\mu} + \hat{\tau}_1 = 944.25$$
$$\text{Residual} : 910 - 944.25 = -34.25$$

f) Describe the assumptions for the model in b) above. Would you say that the assumptions are met? Your answer is supposed to be brief and you can use the figures of Appendix 5.

We have to assume independence, normality and constant variance. The two last assumptions appear OK based on the figures: There are no striking deviations from the straight line and no striking pattern in the plot of fitted values against the residuals.

Independence is not tested formally, but Exercise 1 does not indicate strong deviation from independence.

Appendix 1

For	Frittgående	Besetning	Vekt
For1	Ja	Bes1	910
For1	Ja	Bes1	938
For1	Ja	Bes2	933
For1	Ja	Bes2	915
For1	Nei	Bes1	963
For1	Nei	Bes1	935
For1	Nei	Bes2	969
For1	Nei	Bes2	991
For2	Ja	Bes3	908
For2	Ja	Bes3	911
For2	Ja	Bes4	935
For2	Ja	Bes4	919
For2	Nei	Bes3	939
For2	Nei	Bes3	906
For2	Nei	Bes4	948
For2	Nei	Bes4	898
For3	Ja	Bes5	925
For3	Ja	Bes5	905
For3	Ja	Bes6	919
For3	Ja	Bes6	919
For3	Nei	Bes5	920
For3	Nei	Bes5	880
For3	Nei	Bes6	878
For3	Nei	Bes6	950

Appendix 2

```
Analysis of variance (unrestricted model)
Response: Vekt
          Mean Sq   Sum Sq Df F value Pr(>F)
Besetning 1054.97  5274.83  5    1.74 0.1761
Residuals  605.72 10903.00 18       -      -

            Err.term(s) Err.df VC(SS)
1 Besetning         (2)     18    112
2 Residuals           -      -    606
(VC = variance component)
```

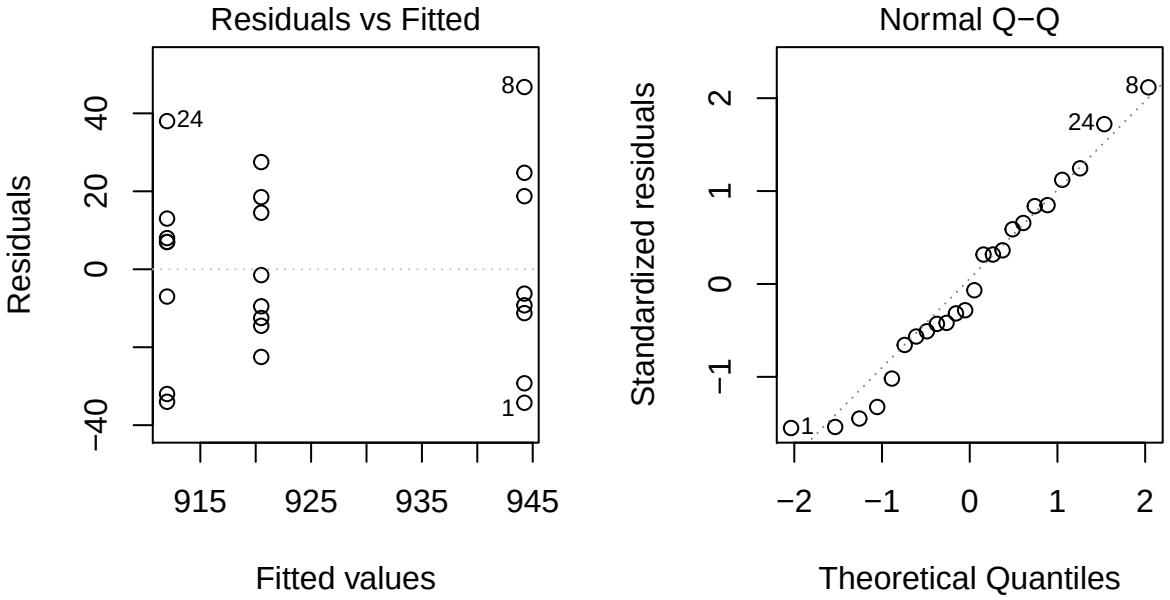
Appendix 3

```
Anova Table (Type II tests)

Response: Vekt
                Sum Sq Df F value  Pr(>F)
For             4470.3  2  4.9146 0.01981 *
Frittgående      816.7  1  1.7956 0.19691
For:Frittgående 2704.3  2  2.9731 0.07662 .
Residuals       8186.5 18
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix 4

```
Analysis of Variance Table

Response: Vekt
          Df  Sum Sq Mean Sq F value  Pr(>F)
For        2  4470.3  2235.2  4.0093 0.03351 *
Residuals 21 11707.5   557.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            Estimate Std. Error t value Pr(>|t|)
(Intercept)  925.583      4.820 192.043   <2e-16 ***
For(For1)     18.667      6.816   2.739   0.0123 *
For(For2)     -5.083      6.816  -0.746   0.4641
---

  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = mdl4)

$For
            diff       lwr       upr     p adj
For2-For1 -23.75 -53.50714  6.007144 0.1340322
For3-For1 -32.25 -62.00714 -2.492856 0.0321413
For3-For2  -8.50 -38.25714 21.257144 0.7545317
```

Appendix 5



Exam 2012 (Sep 3, 2012)

Each sub-question will be given the same weight in the evaluation of the exam. You may answer in English or Norwegian (or "Scandinavian"). There are three exercises.

Exercise 1

The purpose of this exercise to determine if a new diet treatment or method (called `M1`) designed to help people loosing weight is better than two well known methods (called `M2` and `M3`). Two individuals were recruited in each group. The individuals were weighed at the beginning and the end of the study. For each individual the weight difference (`final weight-initial weight`) was recorded. The data and some summary statistics are as follows:

Treatments	Observations	Averages
M1	0, 2	1
M2	1, 3	2
M3	5, 7	6
Total Average		3

We will use the model,

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

Where,

$i = 1, 2, 3$ Corresponding to `M1`, `M2` and `M3`

$j = 1, 2$ corresponding to the two observations in each treatment group

y_{ij} = Weight loss and

$\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$, and $\sum_{i=1}^{3} \tau_i = 0$

a) Calculate $\mathrm{SS}_{\text{Treatments}}$ and SS_{E} defined below :

$$\mathrm{SS}_{\text{Treatments}} = 2 \sum_{i=1}^{3} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

$$\mathrm{SS}_{\mathrm{E}} = \sum_{i=1}^{3} \sum_{j=1}^{2} \left(\bar{y}_{ij} - \bar{y}_{i\cdot}\right)^2$$

$$\begin{aligned}
\mathrm{SS}_{\text{Treatments}} &= 2 \sum_{i=1}^{3} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 \\
&= 2 \left((1-3)^2 + (2-3)^2 + (6-3)^2 \right) \\
&= 28 \\
\mathrm{SS}_{\mathrm{E}} &= \sum_{i=1}^{3} \sum_{j=1}^{2} \left(\bar{y}_{ij} - \bar{y}_{i\cdot}\right)^2 \\
&= (0-1)^2 + (2-1)^2 + (1-2)^2 + (3-2)^2 + (5-6)^2 + (7-6)^2 \\
&= 6
\end{aligned}$$

b) Consider the null hypothesis

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0$$

Show that the test statistic is $F_0 = 7$ and use this to perform the test. Formulate a conclusion.

We have, degree of freedom for treatment is $3 - 1 = 2$ and degree of freedom for error is $6 - 3 = 3$. So,

$$\mathrm{MS}_{\text{treatment}} = \frac{\mathrm{SS}_{\text{treatment}}}{df_{\text{treatment}}} = \frac{28}{2} = 14$$

$$\mathrm{MS}_{\mathrm{E}} = \frac{\mathrm{SS}_{\mathrm{E}}}{df_{\mathrm{E}}} = \frac{6}{3} = 2$$

Thus,

$$F_0 = \frac{\text{MS}_{\text{treatment}}}{\text{MS}_{\text{E}}} = \frac{14}{2} = 7$$

Further, $f_{0.05,2,3} = 9.552$

Since, $7 < 9.552$, we cannot reject H_0 and cannot claim a significance difference between treatments `M1`, `M2` and `M3`.

c) We would like to extend on the above pilot study and design a real study. How would you design this study? You are free to state your assumptions. You can mention words like randomization, blocking and replication. You are not supposed to write more than 10 full sentences.

We should have a sample from the population for which we would like to generalize to. For instance, if we would like to generalize to adults aged 18 to 60, our sample should be individuals from 18 to 60. We could block, for instance wrt to sex. In this case it would be reasonable to have a random sample of females and a random sample of males. Normally we would choose these samples to be equally large aiming for a balanced study. The individuals should be assigned M1,M2 and M3 based on randomization. It is important to have a sufficiently large study and one could do power calculations to secure that the sample size is adequate.

Comment: power calculations are mentioned in the course, but he the students have not been taught how these are done.

Exercises 2 and 3 below are based on the simulated data reproduced in Appendix 1. The response variable is weight of pigs. There are 24 observations. The variables are

Race	This variable indicates one of three possible races and is coded: `Race1`, `Race2` and `Race3`
Farm	This variable indicates which of six farms the pigs come from and is coded `Farm1`, …, `Farm6`
Modern	This variable is `Yes` if the farm is considered to be run in a modern way and otherwise `No`
Weight	The final weight (English: carcass weight, Norwegian: slaktevekt) of pigs in kg

Exercise 2

In this exercise we will only be using the variables Farm and Weight. We consider the following random effect model:

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

Where,

i	=	$1, 2, \ldots, 6$ corresponding to `Farm 1`, `Farm 2`, ..., `Farm 6`
j	=	$1, 2, \ldots, 4$ corresponding to the four observations for each farm,
y_{ij}	=	`Weight`, $\tau_i \sim \mathrm{N}(0, \sigma_\tau^2)$ and $\epsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$

a) We would like to find out if the weights differ between farms, i.e., if we can reject,

$$H_0 : \sigma_\tau^2 = 0$$

Use the output in Appendix 2 to perform the test at 5% significance level. Formulate a conclusion.

The p-value (0.1388) is above 0.05 and therefore we cannot reject. In other words, we cannot claim that there is significant variability between farms.

b) Use the output of Appendix 2 to estimate σ^2. Calculate a 95% confidence interval for σ^2 and interpret the answer.

The estimate of σ^2 is,

$$\hat{\sigma}^2 = \mathrm{MS_E} = 4.01$$

A 95% confidence interval is,

$$\left[\frac{\mathrm{SS_E}}{\chi^2_{0.025,18}}, \frac{\mathrm{SS_E}}{\chi^2_{0.975,18}} \right] = \left(\frac{72.09}{31.53}, \frac{72.09}{8.23} \right) = (2.3, 8.8)$$

Interpretation: We are 95% sure that the true variation is in the interval $(2.3, 8.8)$.

Exercise 3

In this exercise we will be using the variables Race, Modern and Weight. We use the model,

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ij}$$

Where,

i	=	$1, 2, 3$ corresponding to `Race1`, `Race2` and `Race3`
j	=	$1, 2$ corresponding to `Yes` and `No`
k	=	$1, 2, \ldots, 4$ corresponding to replications
y_{ijk}	=	`Weight` and $\epsilon_{ijk} \sim \mathrm{N}(0, \sigma^2)$

$$\sum_{i=1}^{3}\tau_i = 0 \quad \sum_{j=1}^{2}\beta_i = 0 \quad \sum_{i=1}^{3}(\tau\beta)_{ij} = 0 \quad \sum_{j=1}^{2}(\tau\beta)_{ij} = 0$$

The random variables ϵ_{ijk} are assumed to be independent.

a) Use the output of Appendix 3 to determine if there is an interaction between Race and Modern. Formulate the hypotheses and a conclusion. Use 5% significance level.

The null hypothesis is,

$$H_0 : (\alpha\beta)_{ij} = 0 \text{ for all } i \text{ and } j$$

We can not reject as the p-value is 0.09 which does not exceed 0.05. In other words, there is no significant interaction. (A significant interaction would have meant that the effect of Modern depends significantly on Race and vice versa.)

b) For the remaining part of the exercise we will be using the reduced model:

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

Where,

i	=	$1, 2, 3$ corresponding to the $a = 3$ groups `Race1`, `Race2`, `Race3`
j	=	$1, 2, \ldots, 8$ corresponding to the eight $(n = 8)$ observations for each Race
y_{ij}	=	`Weight` and $\epsilon_{ij} \sim \text{NID}(0, \sigma^2)$ and $\sum_{i=1}^{3}\tau_i = 0$

The random variables ϵ_{ij} are assumed to be independent.

Use the output of Appendix 4 to show that there significant differences depending on race at 5% significance level. Between which raises are there significant differences? Use Tukey's test.

The p-value is below 0.05 and therefore we reject and claim that there is a statistically significant difference between the races.

From Tukey, we see that only Race1 and Race3 differ significantly as the p-value is 0.02 while the other p-values exceed 0.05.

c) Use Appendix 5 to estimate τ_1, τ_2 and τ_3

From Appendix 5, we can find,

$$\hat{\tau}_1 = \hat{\mu}_1 - \mu = 76.825 - 78.004 = -1.179$$
$$\hat{\tau}_1 = \hat{\mu}_1 - \mu = 77.55 - 78.004 = -0.454$$
$$\hat{\tau}_1 = \hat{\mu}_1 - \mu = 79.638 - 78.004 = 1.634$$

d) Calculate the predicted value for a `Race1` pig. Calculate the residual corresponding to the first observation.

The predicted value for a `Race1` is the average of such pigs as given in Appendix 4, i.e.,

$$\text{Predicted value } = 76.825$$
$$\text{Residual } = 77.9 - 76.825 = 1.075$$

Appendix 1

Race	Modern	Farm	Weight
Race1	Yes	Farm1	77.9
Race1	Yes	Farm1	76.2
Race1	Yes	Farm2	77.4
Race1	Yes	Farm2	77.4
Race1	No	Farm1	77.5
Race1	No	Farm1	74.2
Race1	No	Farm2	74.0
Race1	No	Farm2	80.0
Race2	Yes	Farm3	76.5
Race2	Yes	Farm3	76.8
Race2	Yes	Farm4	78.8
Race2	Yes	Farm4	77.4
Race2	No	Farm3	79.1
Race2	No	Farm3	76.3
Race2	No	Farm4	79.8
Race2	No	Farm4	75.7
Race3	Yes	Farm5	77.5
Race3	Yes	Farm5	79.0
Race3	Yes	Farm6	78.6
Race3	Yes	Farm6	77.1
Race3	No	Farm5	81.1
Race3	No	Farm5	78.8
Race3	No	Farm6	81.6
Race3	No	Farm6	83.4

Appendix 2

```
Analysis of Variance Table

Response: Weight
          Df Sum Sq Mean Sq F value Pr(>F)
Farm       5 38.677  7.7354  1.9314 0.1388
Residuals 18 72.093  4.0051
```

Appendix 3

```
            Df Sum Sq Mean Sq F value Pr(>F)
Race         2  34.12  17.058   5.586 0.0130 *
Modern       1   4.95   4.950   1.621 0.2191
Race:Modern  2  16.74   8.368   2.740 0.0914 .
Residuals   18  54.97   3.054
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix 4

```
            Df Sum Sq Mean Sq F value Pr(>F)
Race         2  34.12   17.06   4.673  0.021 *
Residuals   21  76.65    3.65
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = mdl4)

$Race
              diff        lwr     upr     p adj
Race2-Race1 0.7250 -1.6828302 3.13283 0.7316436
Race3-Race1 2.8125  0.4046698 5.22033 0.0202992
Race3-Race2 2.0875 -0.3203302 4.49533 0.0971376
```

Appendix 5

```
      mean weight  n
Race1        76.8  8
```

Table 19: Numerical Summary of 'frystemp' data

	Mean	SD	N
merke1	-16.163	1.309	8
merke2	-13.562	1.283	8
merke3	-14.575	0.755	8
Overall	-14.767	1.546	24

```
Race2        77.5  8
Race3        79.6  8
Total        78.0 24
```

Compulsory Assignment

Please upload your solution to the fronter folder Compulsory paper Stat 210 2016. Each student is asked to hand in their own private solution. Please hand in only one file. The name of the file should start with your family name. It's not important to type mathematical symbols nicely. You can, if you prefer, write longhand (by hand), scan and upload a pdf.

Please limit computer output as much as possible in your paper, and any computer output should be commented.

Exercise 1

You may use R or other statistical software for this exercise. The data for this exercise is called `frystemp` and can be loaded from the file `comp1.frystemp.RData` (all data sets are in the data folder of fronter). A lab has investigated the freezing temperature (called `frystemp` in the dataset reproduced below) for three different brands (called `merke` in the dataset).

a) Calculate mean and standard deviation for each brand. Calculate the overall mean (mean for all observations).

b) Consider the following model:

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \quad \text{where, } \sum_{i=1}^{3} \tau_i = 0$$

Here,

y_{ij} : freezing temperature for each brand, $i = 1, 2, 3$ and $j = 1, \ldots, 8$

τ_i : effect of brand i

Table 20: ANOVA table for ‘frystemp‘ model

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
merke	2	27.48083	13.740417	10.48791	0.0006947
Residuals	21	27.51250	1.310119		

Table 21: Coefficient Estimate of Linear Model: ‘frystemp‘ ‘merke‘

term	estimate	std.error	statistic	p.value
$\hat{\mu}$	-14.767	0.234	-63.202	0.000
$\hat{\tau_1}$	-1.396	0.330	-4.224	0.000
$\hat{\tau_2}$	1.204	0.330	3.644	0.002

State the standard assumptions of the model. Use ANOVA to test if there is a difference between the brands. Formulate the hypothesis, carry out a test and formulate a conclusion. Give the ANOVA table.

The standard assumption of the model is,

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

It is also assumed that the y_{ij} are independent. The hypothesis for testing if there is a difference between the brands is,

$$H_0 : \tau_i = 0 \text{ for all } i$$
$$H_1 : \tau_i \neq 0 \text{ for at least one } i$$

The ANOVA table for the model is,

Here, p-value (0.001) is much smaller than 0.05, so we reject H_0 and claim that there is significant difference between `merke` at 95% confidence level.

c) Estimate all parameters in the model.

The estimated parameters in the model can be obtained from model summary as below:

From the output, we can estimate τ_3 from the assumption that the overall mean sum to zero. i.e. $\hat{\tau}_3 = -(\tau_1 + \tau_2) = -(-1.4 + 1.2) = 0.192$

d) Calculate confidence intervals for $\tau_1 - \tau_2$, $\tau_1 - \tau_3$ and $\tau_2 - \tau_3$ using Tukey's method. Are there any significant differences?

Table 22: Summary of Linear Model: ‘frystemp‘ ‘merke‘

r.squared	adj.r.squared	sigma	statistic	p.value	df	df.residual
0.4997	0.4521	1.1446	10.4879	0.0007	3	21

The confidence intervals for Tukey's method can be computed as,

$$\begin{aligned}\bar{y}_{i\cdot} - \bar{y}_{j\cdot} - q_\alpha(a,f)\sqrt{\frac{\mathrm{MS_E}}{n}} &\leq \mu_i - \mu_j \\ &\leq \bar{y}_{i\cdot} - \bar{y}_{j\cdot} + q_\alpha(a,f)\sqrt{\frac{\mathrm{MS_E}}{n}} \quad \leq \mu_i - \mu_j, i \neq j\end{aligned}$$

since, $\mu_i - \mu_j$ is same as $\tau_i - \tau_j$, the conficence interval for group i and j can also be written as,

$$\begin{aligned}\hat{\tau}_{i\cdot} - \hat{\tau}_{j\cdot} - q_\alpha(a,f)\sqrt{\frac{\mathrm{MS_E}}{n}} &\leq \mu_i - \mu_j \\ &\leq \hat{\tau}_{i\cdot} - \hat{\tau}_{j\cdot} + q_\alpha(a,f)\sqrt{\frac{\mathrm{MS_E}}{n}} \quad \leq \mu_i - \mu_j, i \neq j\end{aligned}$$

Here we have,

$$\begin{aligned}q_\alpha(a,f)\sqrt{\frac{\mathrm{MS_E}}{n}} &= q_{0.05}(3,21)\sqrt{\frac{1.310119}{8}} \\ &= 1.4425278\end{aligned}$$

Further, this value is now compated with the absolute difference between the estimates. The difference between the estimates of various group-pairs are,

merke1 - merke2	merke1 - merke3	merke2 - merke3
-2.6	-1.5875	1.0125

If the absolute difference between two groups are larger than the value 1.443 than we reject the null hypothesis and claim that the pair are significantly different from each other. The R output for this test is,

```
                Lower  Center   Upper Std.Err t value  P(>t)
merke1-merke2 -4.0425 -2.6000 -1.1575  0.5723  -4.543 0.0005 ***
merke1-merke3 -3.0300 -1.5875 -0.1450  0.5723  -2.774 0.0294 *
merke2-merke3 -0.4300  1.0125  2.4550  0.5723   1.769 0.2042
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Above output shows the difference between `merke2` and `merke1` whose interpretation will be same as the difference between `merke1` and `merke2` except the sign of Estimate and interval will switch. For instance, in this situation, the estimate will be -2.6 and the lower and upper confidence interval at 95% confidence level will be -4.043 and -2.600 respectively.

The result shows that `marke1` is significantly different than `marke2` and `marke3` at 95% confidence level.

Exercise 2 (Topic discussed at lecture August 23)

You are not supposed to use R or other statistical software for this exercise. (But you should use R to check. The dataset is 'fett' contained in the file `comp2.fett.RData`).

We would like to investigate the fat concentration in milk (`fettprosent` in the below data set). Three farms have been randomly selected and 5 cows from each farm (`besetning`) is recruited to the study. The data is as given below:

a) Explain what is meant by a random effect and why `besetning` (farm) reasonably can be modeled as a random effect.

In many situations, factors have infinitely many levels and a researcher randomly choose some of the levels and make inference about the whole population. Models with such factors are called Random effect models. In case of random effect model, the interest is on the population distribution of factor rather than some specific chosen levels.

For example, An experiment where some specific drugs are to be tested for their efficacy. A randomly chosen drug will barely be an interest in any experiment. Here, the experiment is oriented in finding the effect of those specific drugs and thus, this is the case of fixed effect model. While in another experiment where researcher is interested in finding if there is any differences between farm in Akershus area of Norway in the context of milk production. There can be many farms and the research is not interested on some specific farm but rather is interested on overal population of farm. So some farms are randomly choosen and are used to construct a model. This is the case of random effect model.

b) Consider the model,

$$\begin{aligned} y_{ij} &= \mu + \tau_i + \epsilon_{ij} \\ \tau_i &\sim \text{NID}\left(0, \sigma_\tau^2\right) \\ \epsilon_{ij} &\sim \text{N}\left(0, \sigma^2\right) \\ i &= 1,2,3 \; j = 1,2,\ldots 5 \; N = 15 \end{aligned}$$

Also, all random variables are independent

Use the output below:

```
           Df Sum Sq F value      Pr(>F)
besetning   2 13.545  35.213 0.000009521 ***
Residuals 12  2.308
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Estimate σ_τ^2 and σ^2 and the correlation for animals from the same farm.

Here we can find $\text{MS}_{\text{treatment}}$ and MS_{E} as,

$$\text{MS}_{\text{treatment}} = \frac{\text{SS}_{\text{treatment}}}{df_{\text{treatment}}} = 6.773$$
$$\text{and, } \text{MS}_{\text{E}} = \frac{\text{SS}_{\text{E}}}{df_{\text{Error}}} = 0.192$$

Therefore, we can find the estimates of variance components $\hat{\sigma}^2_\tau$ and $\hat{\sigma}^2$ as,

$$\begin{aligned}\hat{\sigma}^2 &= \text{MS}_{\text{E}} = 0.192 \\ \hat{\sigma}^2_\tau &= \frac{\text{MS}_{\text{treatment}} - \text{MS}_{\text{E}}}{n} \\ &= \frac{6.773 - 0.192}{5} = 1.316\end{aligned}$$

Finally, the correlation for animals from the same farm is,

$$\text{cor}(y_{ij}, y_{ik}) = \hat{\rho} = \frac{\hat{\sigma}^2_\tau}{\hat{\sigma}^2_\tau + \hat{\sigma^2}} = \frac{1.32}{1.508} = 0.872$$

Thus the correlation for animals from the same farm is 87.2 percent.

c) Is there a significant effect of `besetning`? Formulate the hypothesis and do the test.

To test the significance of the random factor `besetning`, the hypothesis can be written as,

$$H_0 : \sigma^2_\tau = 0$$
$$H_1 : \sigma^2_\tau > 0$$

Here, σ^2_τ is the variation between farms and the null hypothesis states that there is no variation between farms. Given ANOVA table shows that the p-value corresponding to `besetning` is very small and thus we reject null hypothesis and claim that there is significant difference between farms. As this is a random effect model, the inference is made on just those specific farms but population of farms in general.

d) Calculate a 99% confidence interval for σ^2

The confidence interval for σ^2 at α level of significance is,

$$\frac{\text{SS}_E}{\chi^2_{\alpha/2,N-a}} \leq \sigma^2 \leq \frac{\text{SS}_E}{\chi^2_{1-\alpha/2,N-a}}$$

From Chi-square table, we can find chi-square values for $\alpha = 0.01$ and $N - a = 12$ as,

$$\chi^2_{0.005,12} = 28.3 \text{ and } \chi^2_{0.995,12} = 3.074$$

Therefore, using $\text{SS}_E = 2.308$, we can find,

$$\begin{aligned}
\frac{\text{SS}_E}{\chi^2_{\alpha/2,N-a}} \leq \sigma^2 \leq \frac{\text{SS}_E}{\chi^2_{1-\alpha/2,N-a}} \\
\frac{2.308}{28.3} \leq \sigma^2 \leq \frac{2.308}{3.074} \\
0.082 \leq \sigma^2 \leq 0.751
\end{aligned}$$

Thus at 99% confidence level, the true error variance (σ^2) lie between the interval $[0.082, 0.751]$.

Exercise 3 (Lectures August 24)

You are supposed to use R or other statistical software for this exercise. The data is `comp3.dommere.RData`.

The purpose of the data of this exercise is to investigate if 5 different types of feeding (`fortype` in the below data set) lead to differently tasting milk. Four judges (corresponding to `Dommer` in the data) were asked to taste and rate on a scale from 1 to 10, 10 being best.

Dommer	Fortype1	Fortype2	Fortype3	Fortype4	Fortype5
Dommer1	6	6	3	4	3
Dommer2	9	8	7	8	3
Dommer3	10	8	5	7	6
Dommer4	9	7	3	4	1

It is natural to let `Dommer` be regarded as a block effect.

Assume two different model, one includes the blocks (Model 1), and the other without blocking (Model 2).

a) Describe the models and state the standard assumptions. Would you prefer Model 1 or Model2? Give reasons for your answer.

Model 1: Let y_{ij} be the `poeng` of i^{th} fortype given by j^{th} dommer.

$$y_{ij} = \mu + \tau_i + \gamma_j + \epsilon_{ij}, \text{ where, } \sum_{i=1}^{5} \tau_i = 0 \text{ and } \sum_{j=0}^{4} \gamma_j = 0$$

Here, $i = 1, \ldots, 5$ (Fortype) and $j = 1, \ldots, 5$ (Dommer)

Model 2: Let y_{ij} be the j^{th} replication of `poeng` for i^{th} fortype.

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}, \text{ where, } \sum_{i=1}^{a} \tau_i = 0$$

Table 23: ANOVA for **Model 1**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Fortype	4	70.30	17.5750	12.7818	0.0003
Dommer	3	31.75	10.5833	7.6970	0.0039
Residuals	12	16.50	1.3750		

Table 24: ANOVA for **Model 2**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Fortype	4	70.30	17.5750	5.4637	0.0064
Residuals	15	48.25	3.2167		

Here $i = 1, \ldots, 5$ (Fortype) and $j = 1, \ldots, 5$ (Replication) Since in this model dommer is not considered as blocking factor so, the points are considered as replications.

The error terms ϵ_{ij} in both Model1 and Model2 are independent and follows normal distribution with mean 0 and constant variance σ^2. i.e.,

$$\epsilon_{ij} \sim \text{NID}\left(0, \sigma^2\right)$$

ANOVA output from Model 1 and Model 2 are as follows,

From the ANOVA tables above, we can see that the block factor `Dommer` has significant effect in Model 1 which indicates that the analysis will be affected if it is removed from the model. In Model 2 the MSE has increased that consiquently decreases F-value corresponding to Fortype.

b) Show that there is a significant effect of `Fortype` in Model 1.

Since, the p-value corresponding to `Fortype` is $(0.0003) << 0.05$, there is significant effect of `Fortype`.

c) Which types of feeding (`Fortype`) differ significantly Use Model 1?

Since there is significant effect of `Fortype`, it is desirable to perform pairwise comparison between different `Fortype`. This can be done using Tukey's pariwise comparison. The output from Tukey's test is,

Table 25: ANOVA table for **Model 1**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Fortype	4	70.30	17.5750	12.7818	0.0003
Dommer	3	31.75	10.5833	7.6970	0.0039
Residuals	12	16.50	1.3750		

	Lower	Center	Upper	Std.Err	t value	P(>t)
Fortype1-Fortype2	-1.393	1.250	3.893	0.829	1.508	0.58
Fortype1-Fortype3	1.357	4.000	6.643	0.829	4.824	0.00
Fortype1-Fortype4	0.107	2.750	5.393	0.829	3.317	0.04
Fortype1-Fortype5	2.607	5.250	7.893	0.829	6.332	0.00
Fortype2-Fortype3	0.107	2.750	5.393	0.829	3.317	0.04
Fortype2-Fortype4	-1.143	1.500	4.143	0.829	1.809	0.41
Fortype2-Fortype5	1.357	4.000	6.643	0.829	4.824	0.00
Fortype3-Fortype4	-3.893	-1.250	1.393	0.829	-1.508	0.58
Fortype3-Fortype5	-1.393	1.250	3.893	0.829	1.508	0.58
Fortype4-Fortype5	-0.143	2.500	5.143	0.829	3.015	0.07

The result shows that at 95% confidence level, `Fortype1`is significantly different than `Fortype3`, `Fortype4` and `Fortype5` while `Fortype2` is significantly different from `Frotype3` and `Fortype5`. In addition at 90% confidence level, `Fortype4` and `Fortype5` are also significantly different.

d) Fortype 1 and 2 are based on feed concentrates (Norwegian: Kraftfor), while Fortype 3 and 4 is based on rutabaga (Norwegian Kålrot).

Construct a contrast that measures the difference between these two feeding regimes, and test if the concentrates types taste significantly better than the other.

A contrast to test the average of `Fortype1` and `Fortype2` with average of `Fortype3` and `Fortype4` can be written as,

$$\text{Contrast: } (\Gamma) = \frac{1}{2}(\tau_1 + \tau_2) - \frac{1}{2}(\tau_3 + \tau_4)$$

The above hypothesis is written in terms of effects τ_i rather than μ_i since $\mu_1 + \mu_2 = \mu_3 + \mu_4$ is same as $\tau_1 + \tau_2 = \tau_3 + \tau_4$. Further, the coefficient of contrast is $c_i = (0.5, 0.5, -0.5, -0.5, 0)$. The hypothesis to test if the difference between concentrates types (`Fortype1` and `Fortype2`) is better than Rutabaga (`Foretype3` and `Fortype4`) is,

$$H_0 : \Gamma = 0 \text{ vs } H_1 : \Gamma > 0$$

The test statistic for this hypothesis is,

$$t_0 = \frac{\sum_{i=1}^{a} c_i \hat{\tau}_{i\cdot}}{\sqrt{\text{MS}_E \sum_{i=1}^{a} \frac{c_i^2}{n_i}}} \sim t_{0.05, N-a}$$

Using the test statistics above we can test the hypothesis. Test output obtained from R is as follows,

	Estimate	Std. Error	t value	Pr(>|t|)
Fortype c=(0.5 0.5 -0.5 -0.5 0)	2.75	0.586302	4.690416	0.0005228

From the test output, we can see that the p-value is very small (smaller than 0.05, level of significance). The p-value here is for two-sided test and for one-sided test, this p-value will be half of the p-value in output which is even smaller. So, we reject the null hypothesis and conclude that the concentrations types are significantly better than the rutabaga type.

References