

Compulsory Assignment

Design and Analysis of Experiment (STAT210)

Raju Rimal

19 August 2016

Hide Answers

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).

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

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 $j = 1, \dots, 8$

τ_i : effect of brand i

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,

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		

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:

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

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	7e-04	3	21

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?

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

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

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

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

Here we have,

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

Further, this value is now compared 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 then we reject the null hypothesis and claim that the pair are significantly different from each other. The R output for this test is,

	Estimate	lwr	upr	Std. Error	z value	Pr(> z)	
merke2 - merke1	2.6000	1.2586	3.9414	0.5723	4.543	1.81e-05	***
merke3 - merke1	1.5875	0.2461	2.9289	0.5723	2.774	0.0153	*
merke3 - merke2	-1.0125	-2.3539	0.3289	0.5723	-1.769	0.1800	

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 overall 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,

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

Also, all random variables are independent

Use the output below:

```
          Df Sum Sq F value    Pr(>F)
besetning  2 13.545  35.213 9.521e-06 ***
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 $MS_{\text{treatment}}$ and MS_E as,

$$MS_{\text{treatment}} = \frac{SS_{\text{treatment}}}{df_{\text{treatment}}} = 6.773$$

$$\text{and, } MS_E = \frac{SS_E}{df_{\text{Error}}} = 0.192$$

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

$$\hat{\sigma}^2 = MS_E = 0.192$$

$$\hat{\sigma}_\tau^2 = \frac{MS_{\text{treatment}} - MS_E}{n}$$

$$= \frac{6.773 - 0.192}{5} = 1.316$$

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

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

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,

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

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

Here, σ_τ^2 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,

$$\frac{SS_E}{\chi_{\alpha/2, N-a}^2} \leq \sigma^2 \leq \frac{SS_E}{\chi_{1-\alpha/2, N-a}^2}$$

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

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

Therefore, using $SS_E = 2.308$, we can find,

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

$$\frac{2.308}{28.3} \leq \sigma^2 \leq \frac{2.308}{3.074}$$

$$0.082 \leq \sigma^2 \leq 0.751$$

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 *Model 2*? 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, $i = 1, \dots, 5$ (Fortype) and $j = 1, \dots, 5$ (Dommer)

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

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

Here $i = 1, \dots, 5$ (Fortype) and $j = 1, \dots, 5$ (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.,

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

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

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		

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		

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 consequently decreases F-value corresponding to Fortype.

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

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		

Since, the p-value corresponding to Fortype is $(3 \times 10^{-4}) \ll 0.05$, 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,

	Estimate	lwr	upr	Std. Error	z value	Pr(> z)	
Fortype2 - Fortype1	-1.250	-3.512	1.012	0.829	-1.51	0.5575	
Fortype3 - Fortype1	-4.000	-6.262	-1.738	0.829	-4.82	1.5e-05	***
Fortype4 - Fortype1	-2.750	-5.012	-0.488	0.829	-3.32	0.0081	**
Fortype5 - Fortype1	-5.250	-7.512	-2.988	0.829	-6.33	1.2e-09	***
Fortype3 - Fortype2	-2.750	-5.012	-0.488	0.829	-3.32	0.0082	**
Fortype4 - Fortype2	-1.500	-3.762	0.762	0.829	-1.81	0.3682	
Fortype5 - Fortype2	-4.000	-6.262	-1.738	0.829	-4.82	1.4e-05	***
Fortype4 - Fortype3	1.250	-1.012	3.512	0.829	1.51	0.5575	
Fortype5 - Fortype3	-1.250	-3.512	1.012	0.829	-1.51	0.5575	
Fortype5 - Fortype4	-2.500	-4.762	-0.238	0.829	-3.02	0.0217	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The result shows that at 95% confidence level, Fortype1 is significantly different than Fortype3, Fortype4 and Fortype5 while Fortype2 is significantly different from Fortype3 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,

$$t_0 = \frac{\sum_{i=1}^a c_i \hat{\tau}_i}{\sqrt{MS_E \sum_{i=1}^a \frac{c_i^2}{n_i}}} \sim t_{0.05, N-a}$$

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.