Design of Experiment and Analysis of Variance

Repetition

Raju Rimal

30 Aug, 2017

Norges miljø- og
biovitenskapelige
universitet

ANOVA Model

ANOVA Model

Random Effect Model

Exam 2011: 1(c), 1(d)
Exam 2012: 1(c)
Exam 2013: 1(c)
Exam 2014: 1(e)
Exam 2015: 2(c)

Intraclass Correlation Coefficient

Proportion of variation between groups to total variation.

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

Using estimates of variance components, we can estimate **intraclass correlation coefficient**.

Confidence interval of overall mean

The $100(1-\alpha)$ level of confidence interval for overall mean μ in case of random effect model is,

$$\hat{\mu} \pm t_{\alpha/2, a-1}\sqrt{\frac{\text{MS}_{\text{treatment}}}{N}}$$

(Refer to Thore's Lecture on Random effect Model)

ANOVA Model

Random Effect Model

Intraclass Correlation

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

CI for overall mean

$$\hat{\mu} \pm t_{\alpha/2, a-1}\sqrt{\frac{\text{MS}_{\text{treatment}}}{N}}$$

Interpretation of Intraclass correlation coefficient

- *Proportaion of variation* between groups to total variation
- Correlation between the observation **within same group**
- In `besettning` and `fettprosent` example, if the correlation is 0.90 shows that the major variation in `fettprosent` is explained by besettning and thus the `cows` in each besettning is more identical and has correlation of 0.90.

Interpretation of Overall Mean

We can extend interpretation of overall mean for whole population

For example, if `besetning` (farms) is a random factor, than the overall mean can refer to the average `fettprosent` in the milk from the entile population of besetning.

ANOVA Model

Random Effect Model

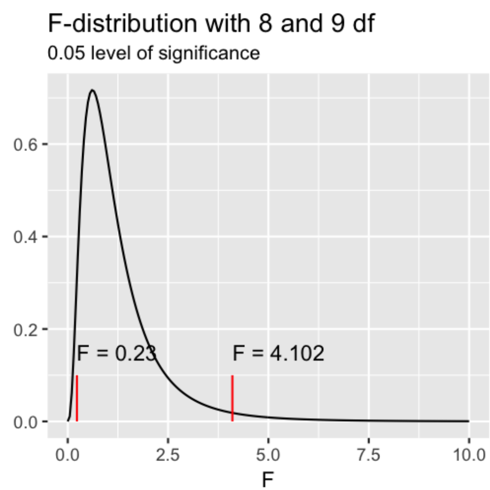


Table 4: Anova Output
F distribution table at 0.025 level

Confidene interval for correlation

L and U gives the confidence interval for σ_τ^2/σ^2.

$$L = \frac{1}{n}\left(\frac{\mathrm{MS}_{\mathrm{treatments}}}{\mathrm{MSE}}\frac{1}{F_{\alpha/2,a-1,N-a}} - 1\right)$$
$$U = \frac{1}{n}\left(\frac{\mathrm{MS}_{\mathrm{treatments}}}{\mathrm{MSE}}\frac{1}{F_{1-\alpha/2,a-1,N-a}} - 1\right)$$

So, the confidence interval for $\rho = \frac{\sigma_\tau^2}{(\sigma_\tau^2+\sigma^2)}$ is,

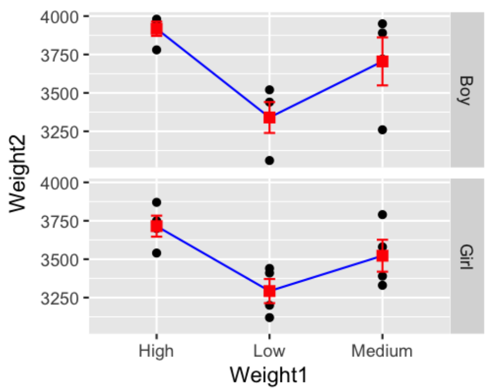
$$\frac{L}{1+L} \leq \frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2} \leq \frac{U}{1+U}$$

Exam 2014: 1(e)

Here, $L = 8.39$ and $U = 158.385$, so, the confidence interval for ρ is (0.893, 0.994).

ANOVA Model

Two factors



Do we need interaction?
How about Gender, is it significant?
What can we see if interaction is not significant?
Is blocking a two factor model?

ANOVA model with two factors

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

where, $\varepsilon_{ijk} \sim N(0, \sigma^2)$, $i = 1, 2, \ldots, a(3)$, $j = 1, 2, \ldots, b(2)$ and $k = 1, 2, \ldots n(4)$ When μ is a overall mean, we will have,

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

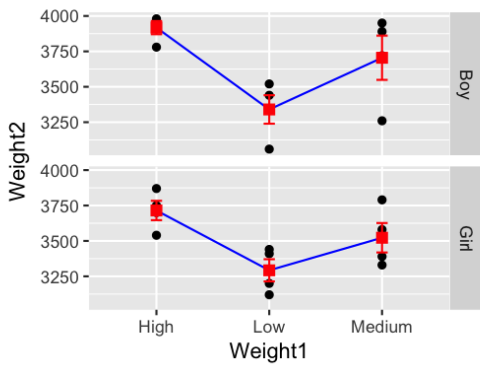
Exam 2013: 2(a)

```
Analysis of Variance Table

Response: Weight2
               Df  Sum Sq Mean Sq F value  Pr(>F)
Weight1         2 1012033  506017   13.03 0.00032 ***
Gender          1  124704  124704    3.21 0.08992 .
Weight1:Gender  2   28433   14217    0.37 0.69842
Residuals      18  698825   38824
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA Model

Two factors



	Estimate
(Intercept)	3582.1
Weight1(High)	234.2
Weight1(Low)	-265.8
Gender(Boy)	72.1
Weight1(High):Gender(Boy)	29.2
Weight1(Low):Gender(Boy)	-48.3

Prediction

Exam 2013: 2(b) wants us to predict the weight of second child (girl) if the first child has `High` weight.

The Model:

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

where, $\varepsilon_{ijk} \sim N(0, \sigma^2)$, $i = 1, 2, 3$, $j = 1, 2$ and $k = 1, 2, 3, 4$ So, the predicted weight for `Girl` child whose first sibling has `High` weight is,

$$\hat{y}_{\texttt{High, girl}} = \hat{\mu} + \hat{\tau}_{\texttt{High}} + \hat{\beta}_{\texttt{Girl}} + \widehat{(\tau\beta)}_{\texttt{High, Girl}}$$

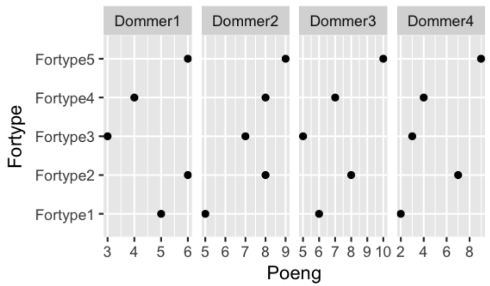
	factor1	coef
1	Weight1(High)	234
2	Weight1(Low)	-266
3	Weight1(Medium)	NA

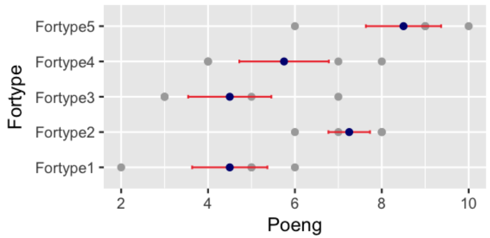
	factor2	Weight1(High)	Weight1(Low)	Weight1(Medium)
1	Gender(Boy)	29.2	-48.3	NA
2	Gender(Girl)	NA	NA	NA

$$\hat{y}_{\texttt{High, girl}} = 3582.083 + (234.167) + (-72.083) + (-29.167) = 3715 \text{ gram}$$

ANOVA Model

Two factors





Reducing a two factors Model

In compulsory assignment 3(c), you are asked to choose between two models.

- What happened when `Dommer` is removed from `Model 1`?

Interaction Term and Degree of freedom

With only one observation for each combination of `Fortype` and `Dommer` we cannot include *interaction term* in the model.

No *degree of freedom* left for residuals. So, we will only be able to find the estimate, but can not perform any kind of test for there significance.

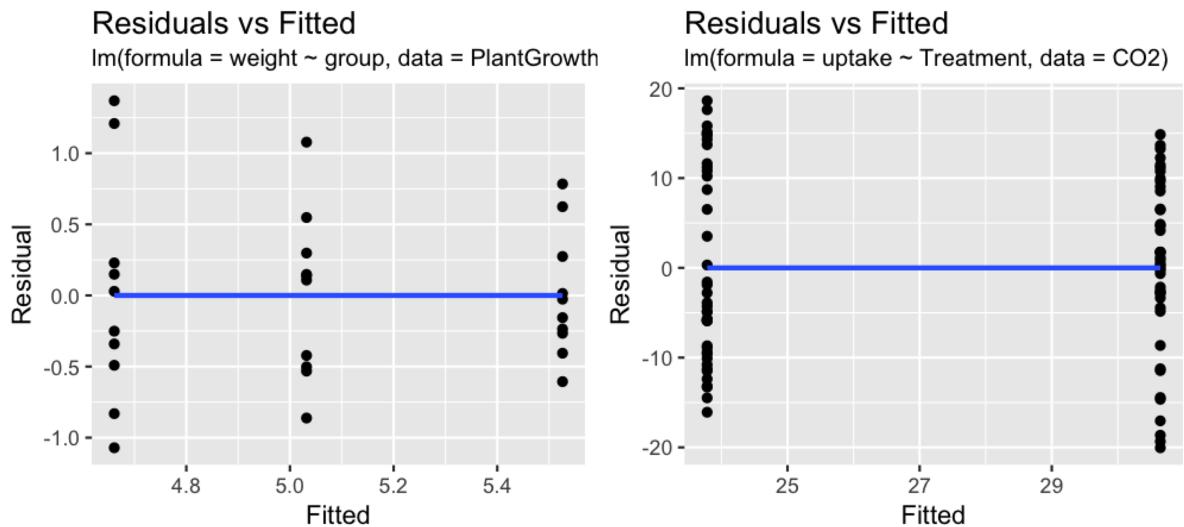
ANOVA Model

Assumption of random error with constant variance

Model Assessment

Error should be random, i.e. free from any kind of pattern

Error should be have constant variation for all the groups



Normality of Error term

ANOVA Model

Assumption of random error with constant variance

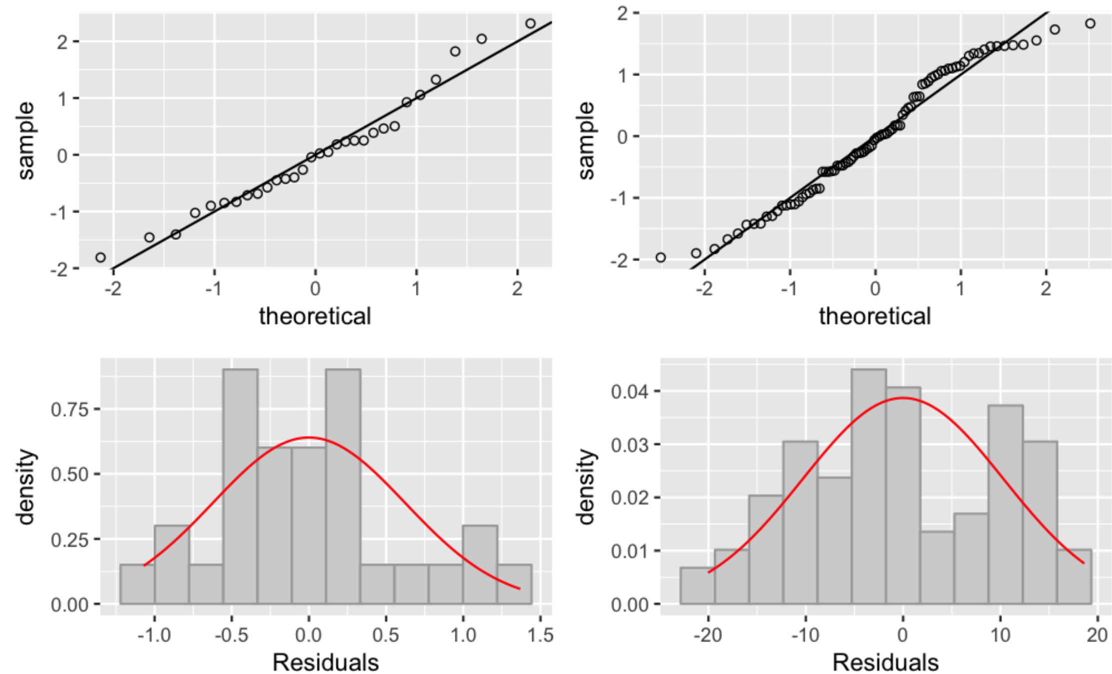
Model Assessment

Normality of Error term

Error (Residuals) should be randomly distribution

All the error should align with Normal Q-Q plot

You can also see histogram and/or density plot and compare with normal distribution plot



Best of Luck

Lykke til

Norges miljø- og
biovitenskapelige
universitet