Design of Experiment and Analysis of Variance

Repetition

Raju Rimal

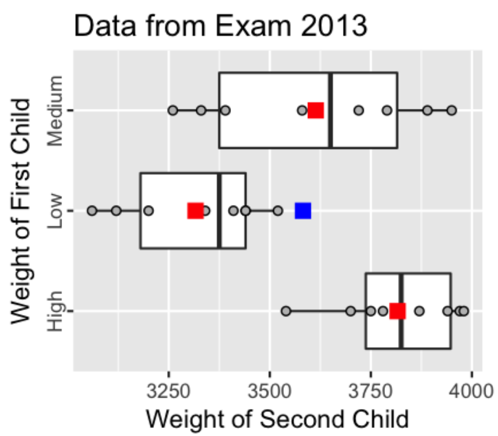
30 Aug, 2017

Norges miljø- og
biovitenskapelige
universitet

ANOVA Model

ANOVA Model

ANOVA table



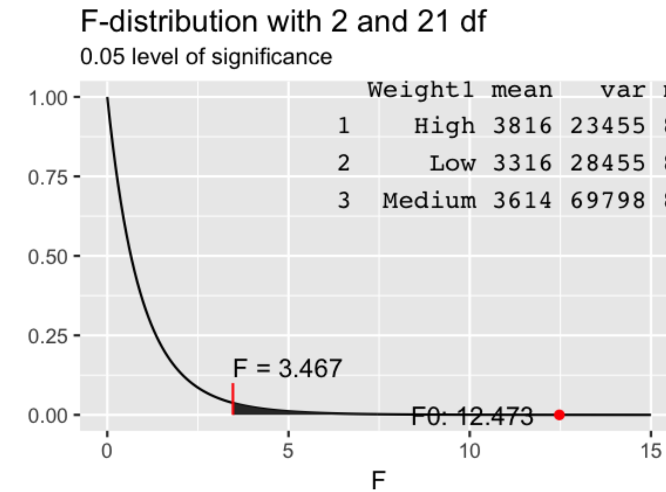

```
          Df  Sum Sq Mean Sq F value
Weight1    2 1012033  506017    12.5
Residuals 21  851962   40570
```

Hypothesis

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

Decision

From F-table, $F_0 > F_c$.
So we reject H_0 at 95% confidence level.



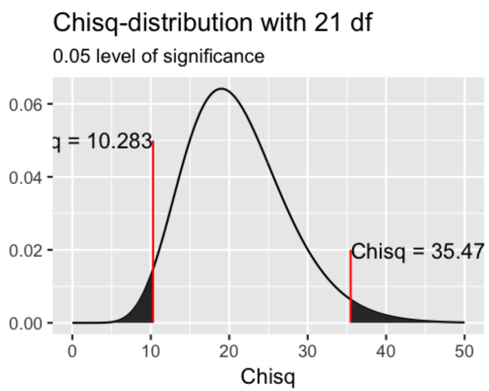
Questions

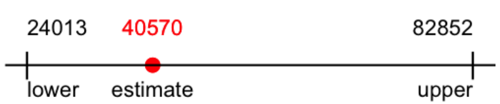
a) What is the 95% **confidence interval of error variance**?

b) We found that there is significant effect of first child's weight on second child, but *how different is the effect of `High` from `low` and `medium`*?

ANOVA Model

Confidence Interval






```
          Df  Sum Sq Mean Sq F value
Weight1    2 1012033  506017    12.5
Residuals 21  851962   40570
```

Confidence interval of error variance

$$\frac{\text{SSE}}{\chi^2_{\alpha/2,N-a}} \leq \sigma^2 \leq \frac{\text{SSE}}{\chi^2_{1-(\alpha/2),N-a}}$$

Exam 2013 Data

Using values in previous slide in above formula,

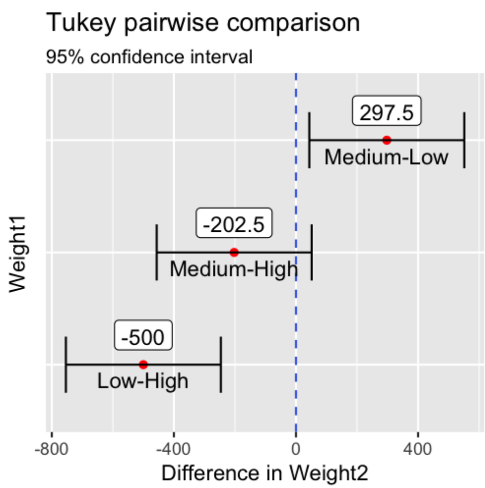
$$\left[\frac{851962.5}{35.479}, \frac{851962.5}{10.283}\right] = [24013.2, 82852.4]$$

At **95% confidence level**, we can say that the true error variance σ^2 lies between the interval (24013.2, 82852.4).

Exercise: Exam 2011 1(b)

ANOVA Model

Post-hoc Test



Think what will happen if you increase the confidence level from 0.05 to 0.1.

Tukey Pairwise Comparison

For multiple testing, Tukey has suggested **studentized range statistic** as,

$$q = \frac{\bar{y}_{\max} - \bar{y}_{\max}}{\sqrt{\mathrm{MSE}/n}}$$

Compare q with $q_\alpha(a, f)$ found in table. We declare a pair to be significantly different if $q > q_\alpha(a, f)$.

Confidence Interval

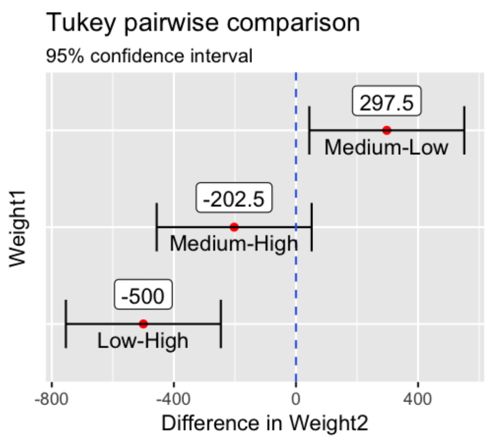
$$(\bar{y}_i - \bar{y}_j) \pm q_\alpha(a, f)\sqrt{\frac{\mathrm{MSE}}{n}}$$

$$(\bar{y}_i - \bar{y}_j) - q_\alpha(a, f)\sqrt{\frac{\mathrm{MSE}}{n}} \leq \mu_i - \mu_j \leq (\bar{y}_i - \bar{y}_j) + q_\alpha(a, f)\sqrt{\frac{\mathrm{MSE}}{n}}$$

You can also check if this interval contains zero.

ANOVA Model

Post-hoc Example




```
Weight1 mean n
   High 3816 8
    Low 3316 8
 Medium 3614 8
```

Exam 2013, 2(d)

From the Appendix 1 and Appendix 4, we have,

```
          Df  Sum Sq Mean Sq F value   Pr(>F)
Weight1    2 1012033  506017    12.5 0.000269
Residuals 21  851962   40570      NA       NA
```

Using studentized, T-distribution table $q_{0.05}(3, 21) = 3.565$. So,

$$T_{0.05} = q_{0.05}(3, 21)\sqrt{\frac{40569.643}{8}} = 253.845$$

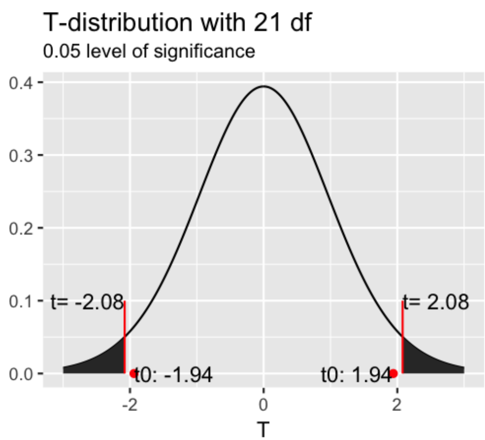
We have three possible pairs for comparison: `Low-High`, `Medium-High` and `Medium-Low`. The difference in their means are: `-500, -202.5, 297.5` respectively.

If compared with $T_{0.05}(3, 21)$, we see that at 95% confidence level, we find `Low-High` and `Medium-Low` differ significantly.

Exam 2015, 1(e)

ANOVA Model

Contrast test



Use cases

- Comparison of one group with average of other group
- Comparing treatments with control

Exam 2014: 2(d)

Hypothesis:

$$H_0 : \frac{1}{2}(\mu_1 + \mu_2) - \frac{1}{2}(\mu_3 + \mu_4) = 0 \text{ vs } H_1 : \frac{1}{2}(\mu_1 + \mu_2) - \frac{1}{2}(\mu_3 + \mu_4) \neq 0$$

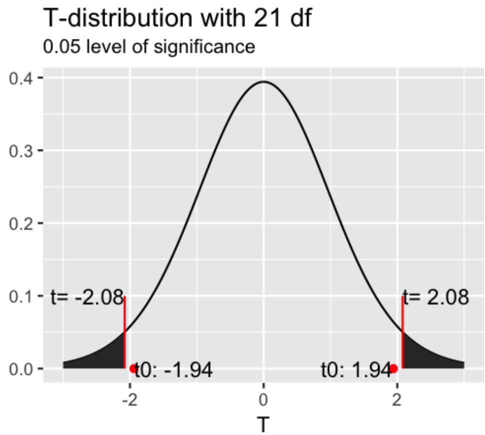
Decision:

From Table 8, p-value $(0.065) > 0.05$.

- *What should we conclude?*

ANOVA Model

Contrast Calculation



Continue on Exam 2014: 2(d)

Before any compution, **formulate a hypothesis**.

Estimate of Contrast

$$\Gamma = \sum_{i=1}^{a} c_i \mu_i = \frac{1}{2}(\mu_1 + \mu_2) - \frac{1}{2}(\mu_3 - \mu_4) = -5.5$$

Here the contrast coefficients c_i are: $1/2, 1/2, -1/2, -1/2$. Contrast coefficients sum to zero. For estimation of Γ, use respective sample mean. Also use Table 6 for following calculations.

Standard Error and test statistic

$$\mathrm{SE}(\hat{\Gamma}) = \sqrt{\frac{\mathrm{MSE}}{n} \sum_{i=1}^{a} c_i^2} = 2.82 \qquad t = \frac{\hat{\Gamma}}{\mathrm{S.E}(\hat{\Gamma})} \sim t_{\alpha/2, N-a}$$

ANOVA Model

Random Effect Model

Random factor

$$\tau_i \sim \mathrm{N}(0, \sigma_\tau^2)$$

Fixed factor

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

The Model

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \text{ where, } \varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2) \text{ and } \tau_i \sim \mathrm{N}(0, \sigma_\tau^2)$$

Why Random effect model

Specific levels are not of interest. General variation is more important. Levels of a factor are **randomly selected**. For example, if `besetning` (farm) is not of interest, we randomly sample such farms from a population of farm.

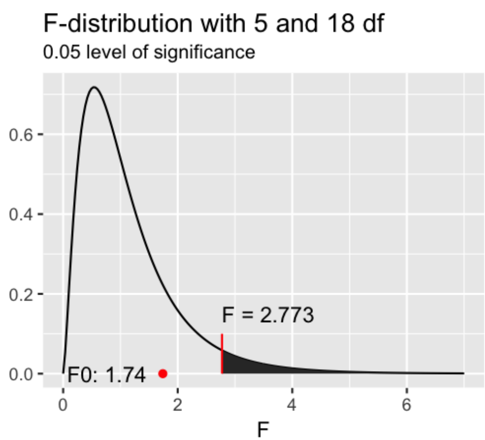
Usually, **blocks** are taken as random factor.

A comparison with Fixed Effect Model

Specific levels of a factor is important than their general variation. These levels are **specifically chosen**. For instance, comparison of new and old drug where a particular drug is used for comparison.

ANOVA Model

Random Effect Model



Exam 2011: 1(a)

Hypothesis:

$$H_0 : \sigma_\tau^2 = 0 \text{ vs } H_1 : \sigma_\tau^2 > 0$$

From Appendix 2, we see p-value is larger than $\alpha = 0.05$, we can not reject the null hypothesis. So, we claim that **at 95% confidence level** we claim that there is *not significant variation* between the farms.

Estimates of Variance Components

$$\text{total variation } \mathrm{var}(y_{ij}) = \sigma_\tau^2 + \sigma^2$$

We estimate them as,

$$\hat{\sigma}^2 = \text{MSE and } \hat{\sigma_\tau}^2 = \frac{\text{MS}_{\text{treatment}} - \text{MSE}}{n}$$

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,

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

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