Design of Experiment and Analysis of Variance

Repetition

Raju Rimal

29 Aug, 2017

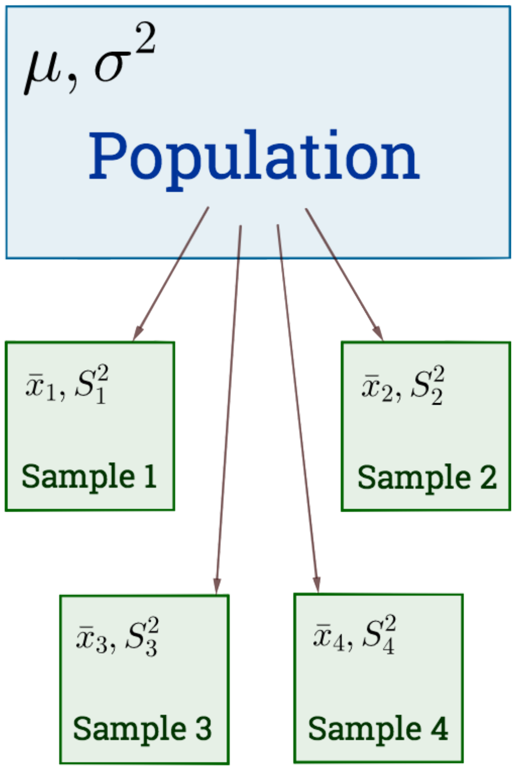
NMBU

Norges miljø- og biovitenskapelige universitet

Statistical Inference

Hypothesis Testing

Inference Steps



Steps

1. Make a hypothesis
2. Collect data
3. Calculate test-statistic
4. Compare test-statistic with theoretical distribution (T-statistic, F-statistic, χ^2 statistic)
5. Conclusion and decision

Remember

For an estimate we always use hat. For instance, Sample mean $\bar{x} = \hat{\mu}$ is an estimate of population mean μ.

Similarly, $\hat{\sigma}$ is an estimate of population standard deviation σ.

Hypothesis

Null Hypothesis

Exam questions

2011: 2(a)
2012: 2(b)

Use cases

One sample t-test : $H_0 : \mu = 5$

Two sample t-test : $H_0 : \mu_1 = \mu_2$

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

Random effect Model : $H_0 : \sigma_\tau^2 = 0$

Always write hypothesis in terms of ***population parameter***

Hypothesis

Alternative Hypothesis

Exam questions

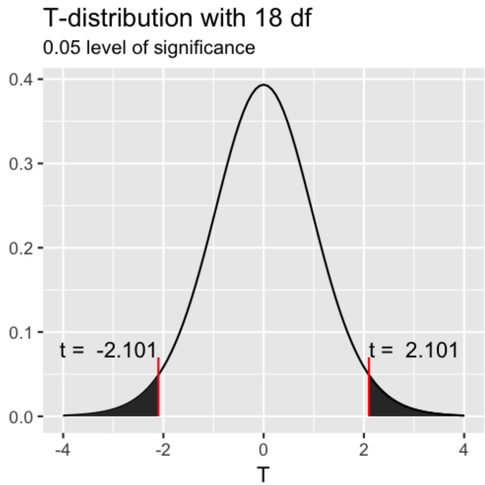
2011: 2(a)
2012: 2(b)

One sided vs Two sided

- Test if expected variance of x is greater than 11.
- Are the three soya groups significantly different?
- Can we conclude that the average of $soya$ diet gives higher protein than $non - soya$ diet?
- Does fat percent vary accross these randomly chosen farms?

T Statistic

T-test



The *darker region* under the curve is *rejection region*. If the calculated t-value lies in this region, we reject Null hypothesis.

One sample mean

$$\text{t-statistic} = \frac{\bar{y}}{\mathrm{SE}(\bar{y})} = \frac{\bar{y}}{\hat{\sigma}/\sqrt{n}} \sim t_{\alpha/2,n-1}$$

Difference between two groups

$$\text{t-statistic} = \frac{\bar{y}_i - \bar{y}_j}{\mathrm{SE}(\bar{y}_i - \bar{y}_j)} = \frac{\bar{y}_i - \bar{y}_j}{S_{\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{\alpha/2,N-a}$$

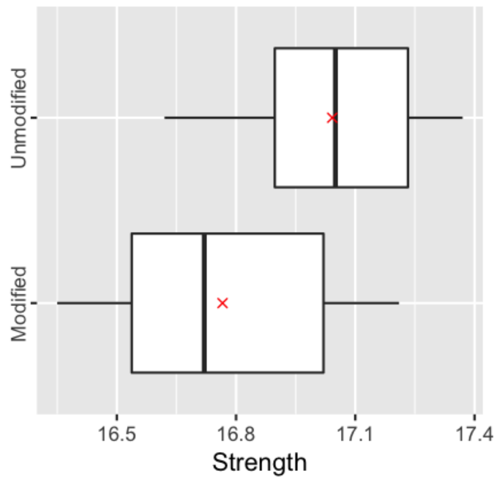
C.I. true difference between two groups

$$\left[\bar{y}_i - \bar{y}_j \pm t_{\alpha/2,N-a} \times S_{\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

Remember: Here $N = n_1 + n_2$ is total number of observation in all groups.

T Statistic

Example




```
  Modified Unmodified
1     16.9       16.6
2     16.4       16.8
```

Two sample t-test

```
    Two Sample t-test

data:  Modified and Unmodified
t = -2, df = 20, p-value = 0.04
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.54239 -0.00961
sample estimates:
       mean of x       mean of y  pooled std.dev.
          16.766          17.042            0.284
```

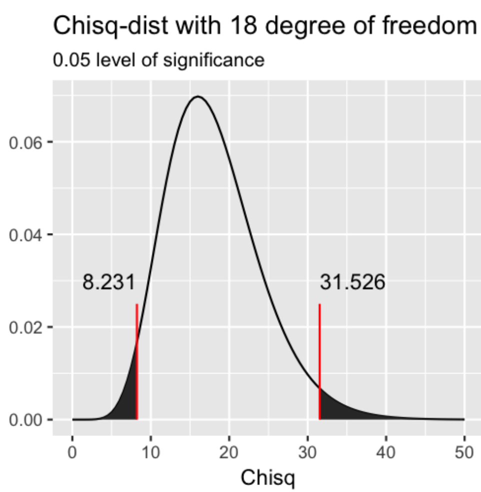
ANOVA test

```
Analysis of Variance Table
          Df Sum Sq Mean Sq F value Pr(>F)
key        1  0.381   0.381    4.74  0.043 *
Residuals 18  1.447   0.080
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pooled Variance

χ^2 Statistic

Chisq Test



The shaded region covers *5%* area of the curve.

Variance Test

$$\chi^2 \text{ statistic } = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{\alpha,n-1}$$

ANOVA: Confidence Interval of MSE ($\hat{\sigma}^2$)

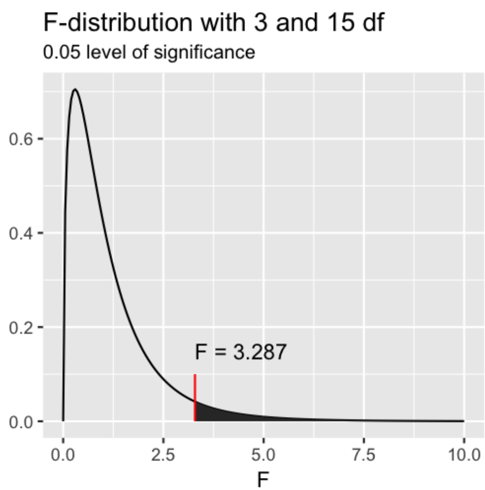
$$\left[\frac{(N-a)\text{MSE}}{\chi^2_{\alpha/2,N-a}}, \frac{(N-a)\text{MSE}}{\chi^2_{1-\alpha/2,N-a}}\right] = \left[\frac{\text{SSE}}{\chi^2_{\alpha/2,N-a}}, \frac{\text{SSE}}{\chi^2_{1-\alpha/2,N-a}}\right]$$

Things to remember

- $(N-a)$ is the degree of freedom for residual
- $(N-a)\text{MSE} = \text{SSE}$
- χ^2_{N-a} is unsymmetric unlike t-distribution

F Statistic

F test



The ratio of two chisq distribution is F-distribution

Testing difference in variablility of two groups

$$H_0 : \sigma_1^2 = \sigma_2^2 \text{ vs } H_1 : \sigma_1^2 \neq \sigma_2^2$$

F statistic

If $S_1^2 \sim \chi^2_{n_1-1}$ and $S_2^2 \sim \chi^2_{n_2-1}$ are sample mean of two groups respectively,

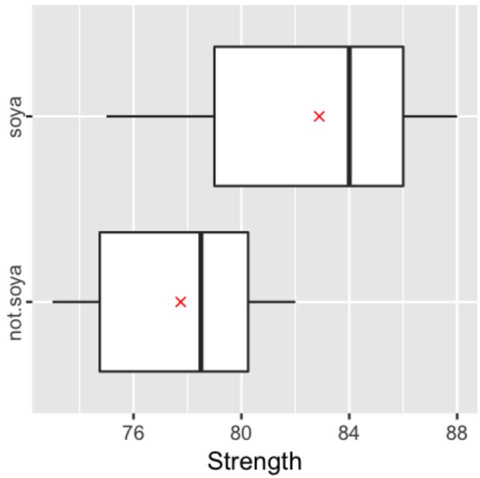
$$F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1}$$

Secret Trick

$$F_{(1-\alpha), n_1, n_2} = \frac{1}{F_{\alpha, n_2, n_1}}$$

F Statistic

Example

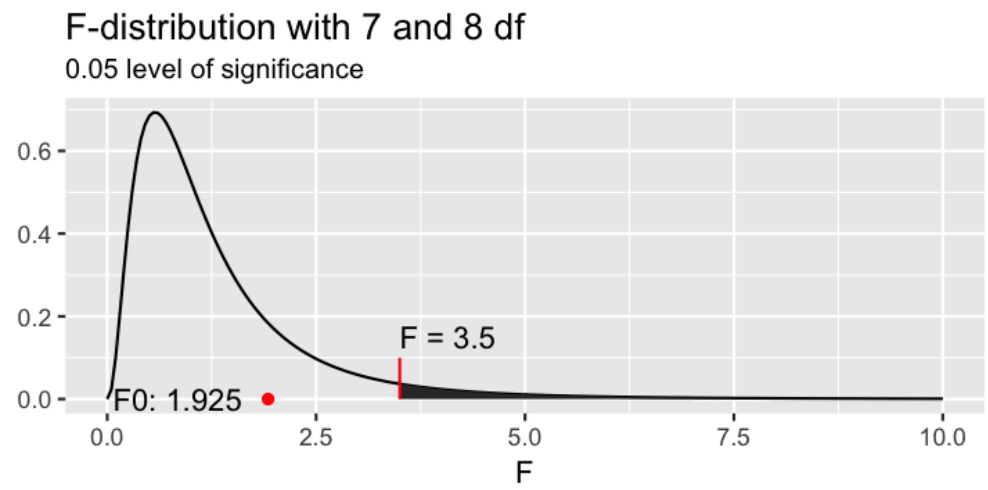



```
   group  mean   var n
not.soya 77.75 11.36 8
    soya 82.89 21.86 9
```

Soya diet experiment

$$\text{F-value} = \frac{S_1^2}{S_2^2} = \frac{21.861}{11.357} = 1.925 \sim F_{7,8}$$

F-value from Table is 3.726

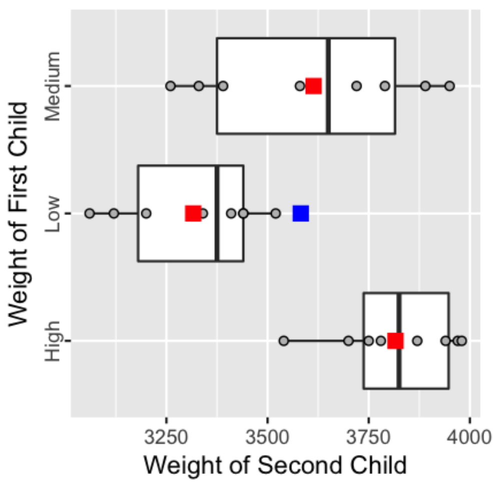


Reject H_0: Variation between two groups significantly differs.

ANOVA Model

ANOVA Model

Assumptions



1. Errors are random and independent
2. Errors are normally distributed with mean 0 and constant variance σ^2

Mean Model

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

Here, μ_i is mean of group i; $i = 1, 2, \ldots a$ and $j = 1, \ldots n$

Effect Model

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

Here, We split μ_i (group mean) into overall mean (μ) and effect (τ_i) of group i as, $\mu_i = \mu + \tau_i$. In this case we will have $\sum_{i=1}^{a} \tau_i = 0$, i.e, $\tau_1 + \tau_2 + \tau_3 = 0$

Assumptions

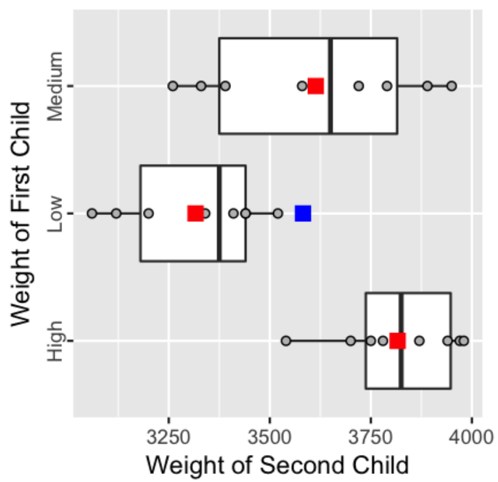
In ANVOA model, we assume $\varepsilon_{ij} \sim \text{NID}(0, \sigma^2)$

ANOVA Model

ANOVA table

■ **TABLE 3.3**
The Analysis of Variance Table for the Single-Factor, Fixed Effects Model

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Between treatments	$SS_{\text{Treatments}} = n\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2$	$a-1$	$MS_{\text{Treatments}}$	$F_0 = \frac{MS_{\text{Treatments}}}{MS_E}$
Error (within treatments)	$SS_E = SS_T - SS_{\text{Treatments}}$	$N-a$	MS_E	
Total	$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{..})^2$	$N-1$		



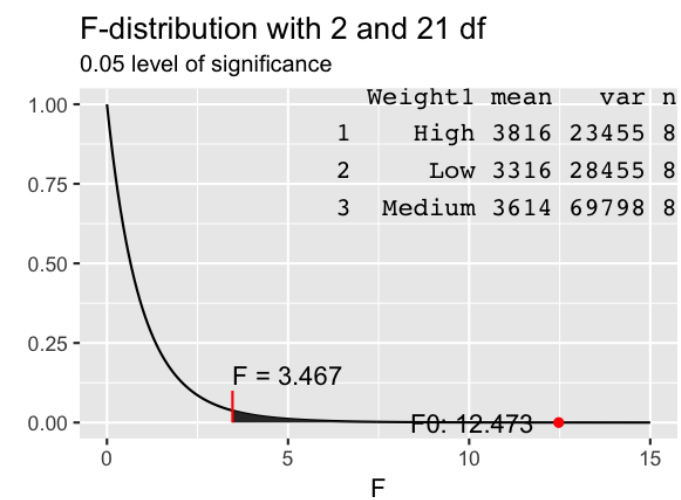

```
          Df  Sum Sq Mean Sq F value
Weight1    2 1012033  506017    12.5
Residuals 21  851962   40570
```

Hypothesis

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

Decision

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



Exam Questions

ANOVA Model

Exam Questions

2012: 2(b), 2(c)

2012: Appendix 4-5

Question 2(b)

Hypothesis:

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0$$
$$H_1 : \text{At least one } \tau_i \neq 0, i = 1, 2, 3$$

We have p-value in ANOVA table given in Appendix 4. **If p-value is *less than* the level of significance α, we reject H_0**

Question 2(c)

Since, $\mu_i = \mu + \tau_i$. All group means μ_1, μ_2, μ_3 and overall mean μ are given, you can find τ_1, τ_2 and τ_3.