RenR 480/711

Hypothesis testing

One way Analysis of Variance (ANOVA)

Hypothesis testing

Lentil experiment

$\bar{x}_{Variety\ A} = 728$ kg/ha
$\bar{x}_{Variety\ B} = 514$ kg/ha

Farm1

H_0 *Null Hypothesis*: $\mu_{Variety\ A} = \mu_{Variety\ B}$
H_1 *Alternative Hypothesis*: $\mu_{Variety\ A} \neq \mu_{Variety\ B}$

All statistics can do is to **try to reject H_0**

Is random chance a plausible explanation for the observed difference?

$p = 0.2$	Yes, one out of 5 times
$p = 0.05$	Yes, one out of 20 times
$p = 0.0001$	Yes, one out of 10000 times

Hypothesis testing

<u>Type I error</u>

- The chance to reject H_0 when it is actually true
- Type I error = p
- Depends on the chosen α-level (Predetermined probability at which we make a decision)

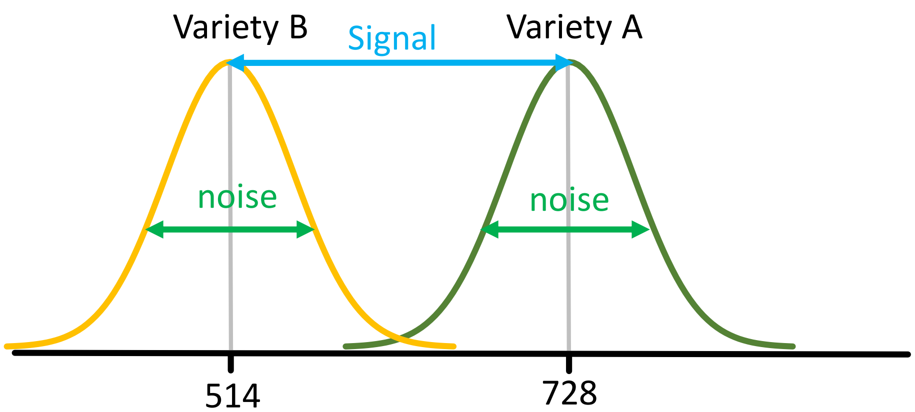
$$\alpha \downarrow_{0.01}^{0.05} \quad Type\ I\ error \downarrow$$

<u>Type II error</u>

- The chance to accept H_0 when it is actually false
- There is not enough evidence in the data to reject H_0

$$\alpha \downarrow_{0.01}^{0.05} \quad Type\ II\ error \uparrow$$

t-Test for two samples



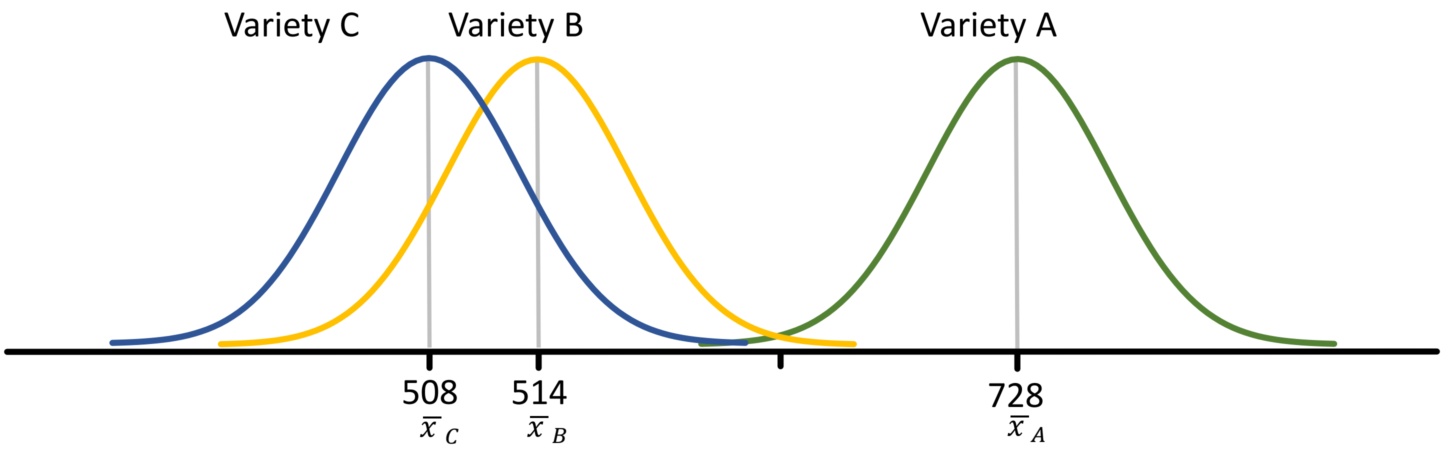
$$t = \frac{signal}{noise} \quad t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Pooled SE

H_0 *Null Hypothesis*: $\mu_{Variety\ A} = \mu_{Variety\ B}$

One way ANOVA

For 3 or more sample groups

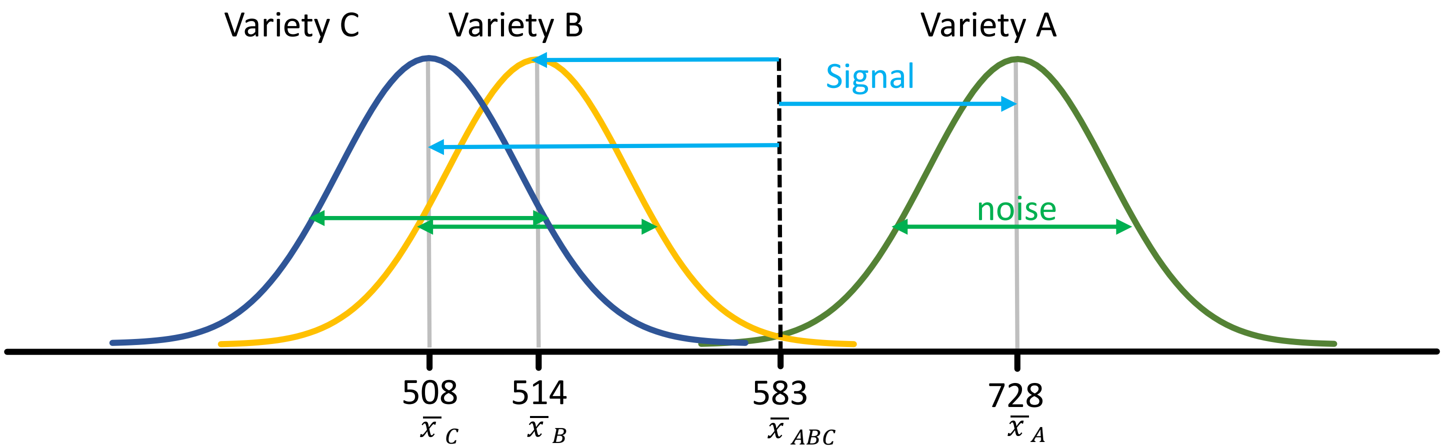


H_0 *Null Hypothesis*: $\mu_{Variety\ A} = \mu_{Variety\ B} = \mu_{Variety\ c}$

H_0 is rejected when the probability for any difference between groups being due to chance is less than the alpha-level.

One way ANOVA

For 3 or more sample groups



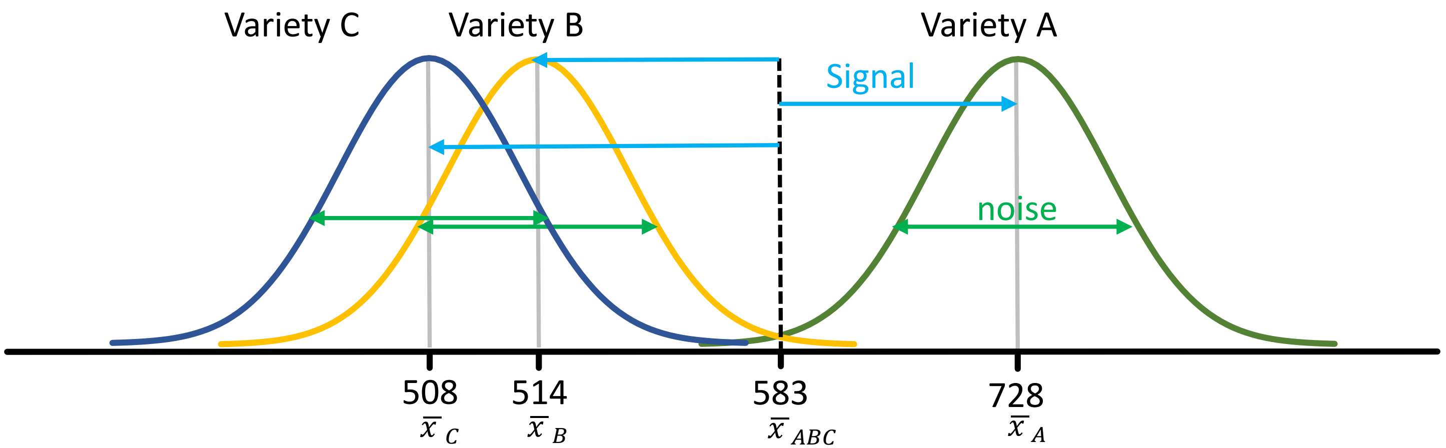
$$F = \frac{signal}{noise} = \frac{variance_{between}}{variance_{within}}$$

$$variance_{between} = \frac{\sum_i^n (\bar{x}_i - \bar{x}_{All})^2}{n-1} \times r$$

$$variance_{within} = \frac{\sum_i^n (variance_i)}{n}$$

One way ANOVA

For 3 or more sample groups

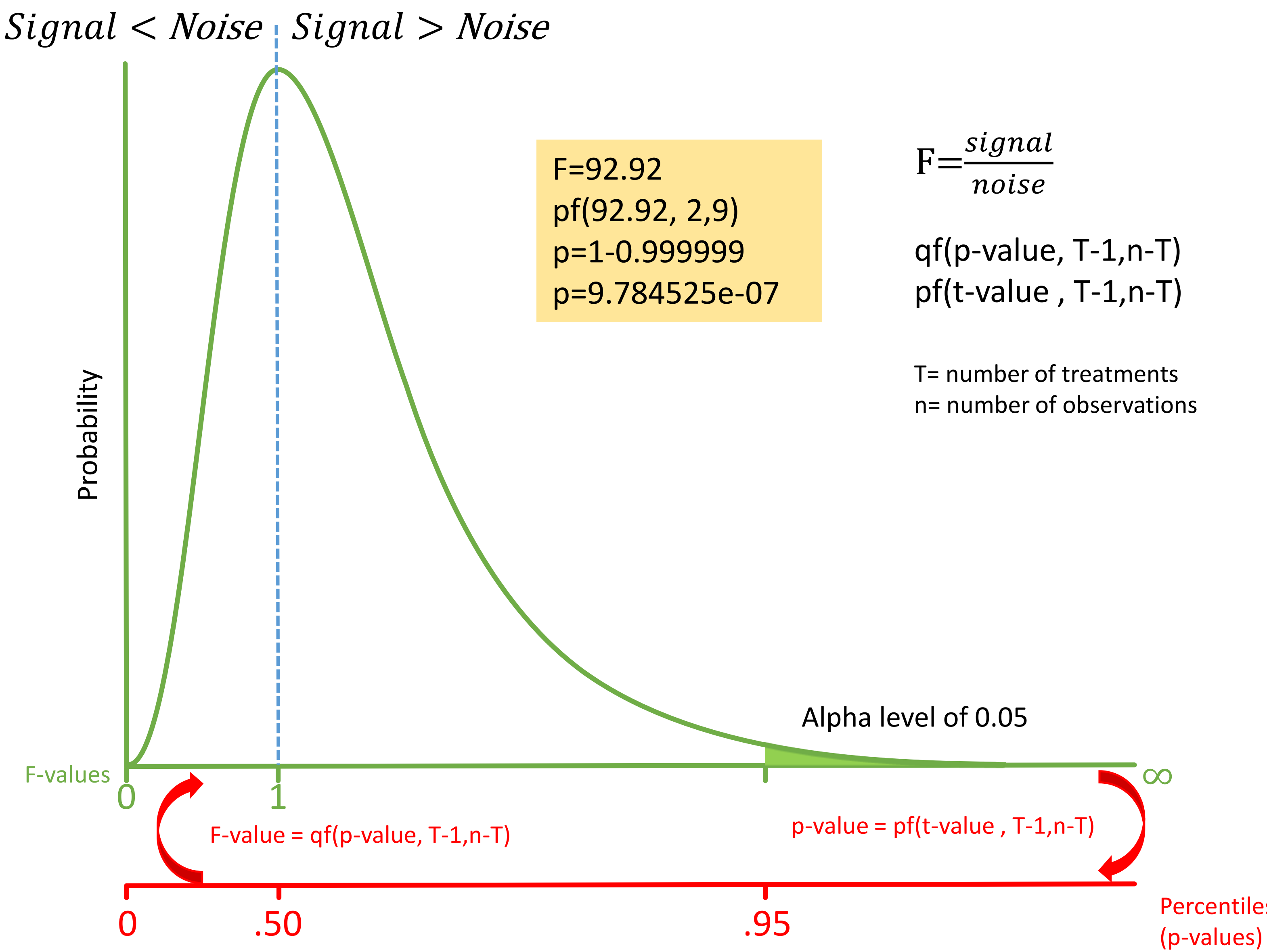


$$variance_{between} = \frac{\sum_i^n (\bar{x}_i - \bar{x}_{All})^2}{n-1} \times r = \frac{(728-583)^2 + (514-583)^2 + (508-583)^2}{3-1} \times 4 = \mathbf{62463}$$

$$variance_{within} = \frac{\sum_i^n (variance_i)}{n} = \frac{891.67 + 819.33 + 305.58}{3} = \mathbf{672}$$

$$F = \frac{signal}{noise} = \frac{variance_{between}}{variance_{within}} = \frac{62463}{672} = 92.92$$

F-distribution



ANOVA table

```
> anova(out1)
Analysis of Variance Table

Response: YIELD
          Df Sum Sq Mean Sq F value    Pr(>F)
VARIETY    2 124927   62463  92.924 9.783e-07 ***
Residuals  9   6050     672
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Source of variation	df	Sum of squares	Mean Squares	F-value	p-value
Variety	2	124927	62463	92.924	9.783e-07
Residuals	9	6050	672		

df1=Number of groups -1

df2= Number of observations - number of groups

MS1 * df1

MS2 * df2

$variance_{between}$

$variance_{within}$

$$F = \frac{variance_{between}}{variance_{within}}$$

pt(F,df1,df2)

ANOVA vocabulary

Treatment

Predictor variable e.g. Variety, Fertilisation, Irrigation, Age-class, water quality-class

Treatment level

Value of predictor variable e.g. A,B,C / C,IxN,2xN / 1,2,3

Covariate

Undesired, uncontrolled predictor variable, confounding, e.g. soil moisture-class of plots

F-value

$$F=\frac{signal}{noise}=\frac{variance_{between}}{variance_{within}}=\frac{mean\ squares_{treatment}}{mean\ squares_{error}}$$

p-value

Probability that the observed difference in means in due to random chance