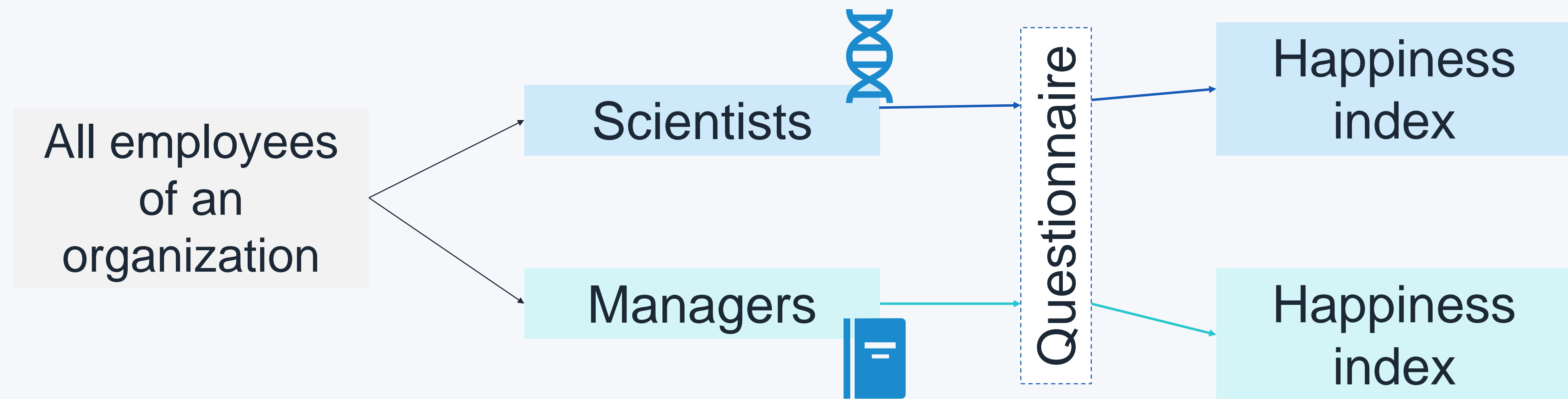# ANOVA Recap

Statistics Tutorial
Day 12

Prabesh Dhakal
2020 July 02

T-TEST

Comparison of sample mean between two groups/levels.

ANALYSIS OF VARIANCE

● ● ●

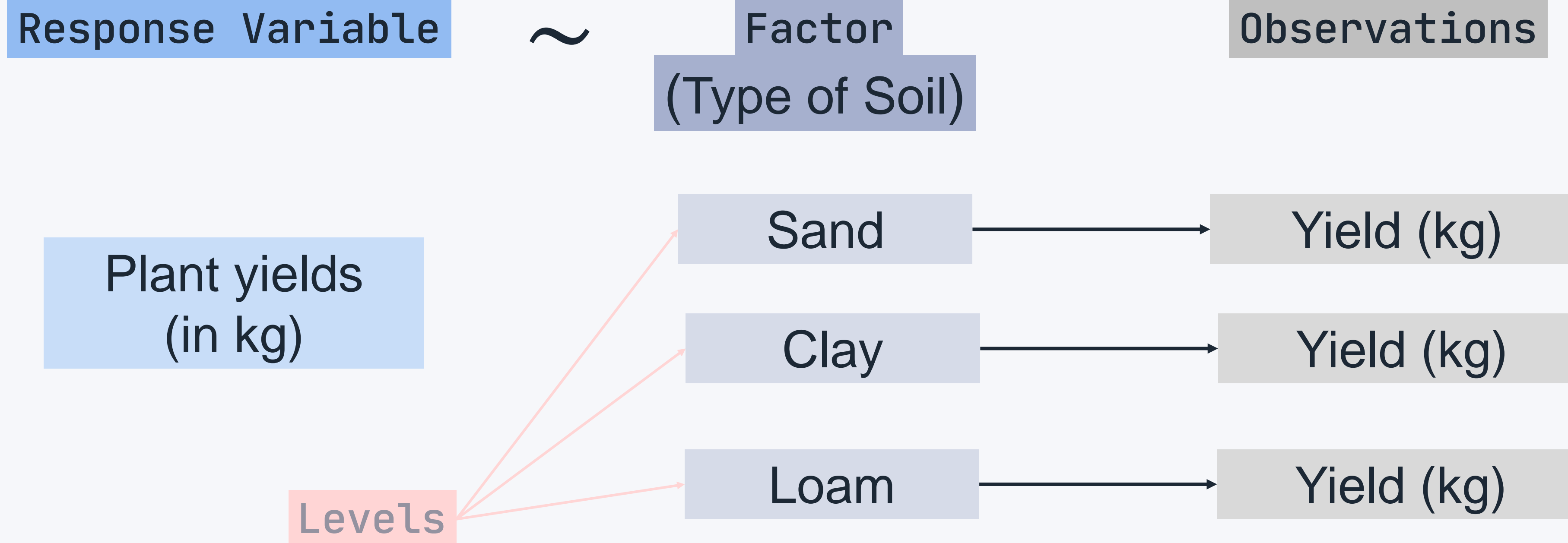ANOVA : **<u>An</u>**alysis **<u>o</u>**f **<u>Va</u>**riance

$$response \sim predictors \text{ (factors)}$$

- **Factors** = explanatory variables (aka Predictors)

  - Have different "**levels**" where each level can be thought of as

    a category

- **Response Variable**

  - Numeric (integer or decimal point numbers)

ANALYSIS OF VARIANCE

Comparison of sample mean between 3 or more groups/levels.

Response Variable	∼	Factor (Type of Soil)	Observations

Plant yields (in kg)

Levels

Sand	→	Yield (kg)
Clay	→	Yield (kg)
Loam	→	Yield (kg)

HYPOTHESES IN ANOVA

H_0 : the group means are not different

H_1 : the group means are different

H_1 does not specify which groups vary from each other.

2 MAIN TYPES OF ANOVA

One-way ANOVA

One factor with more than 3 levels

Type of soil affects yield

Factor	Soil
Levels	Different types of soil (sand, loam, clay)
Response	Crop Yield (kg)

Two-way ANOVA

Two factors with multiple levels

Type of soil AND type of fertilizer affects yield

Factors	Soil and Fertilizer
Levels	Different types of soil (sand, loam, clay) **&** Different types of fertilizers (x, y, z)
Response	Crop Yield (kg)

KEY ASSUMPTIONS OF ANOVA

● ● ●

The data is normally distributed.

Variance is homogeneous.

Sample size is balanced.

One-way ANOVA Intuition

Total Variation
(SSY)

=

Between-groups Variation
(SSA)

+

Within-groups Variation
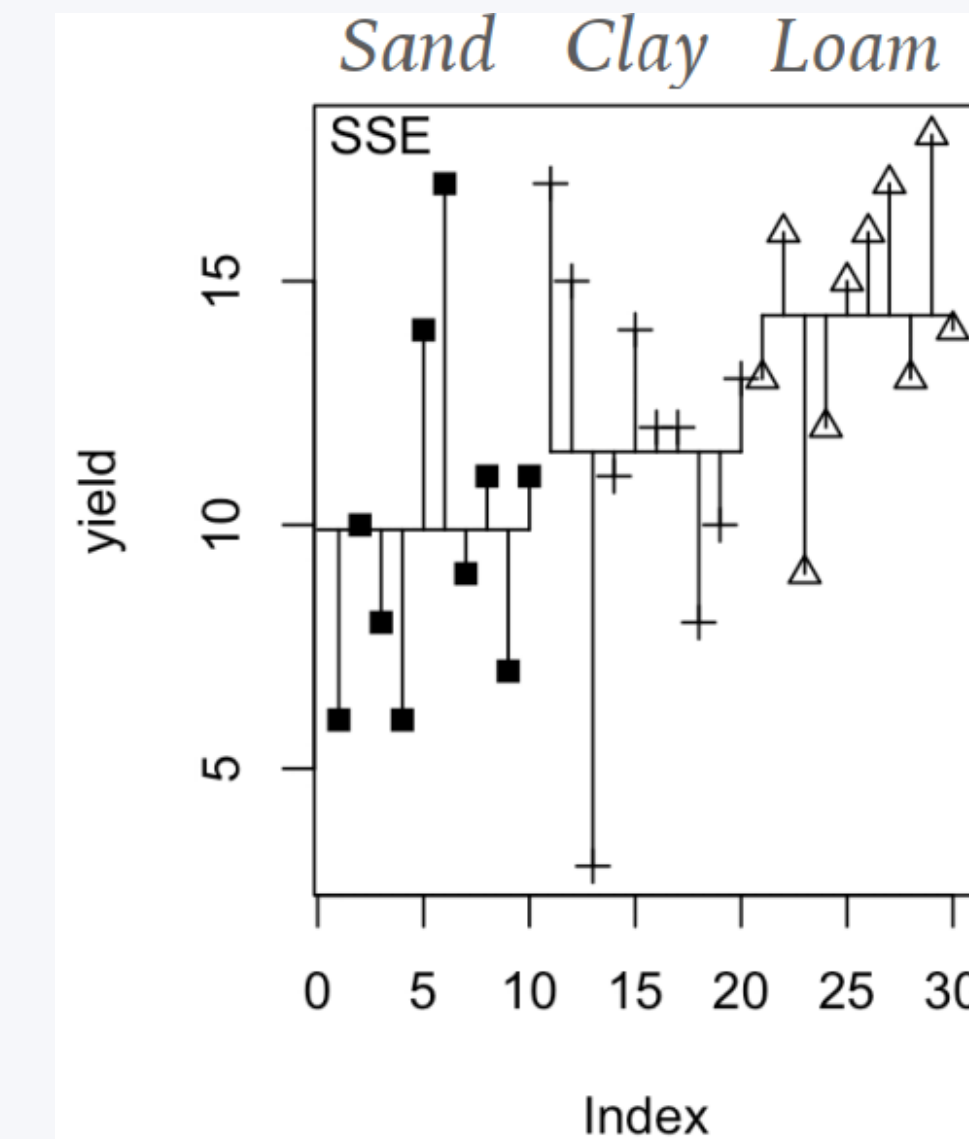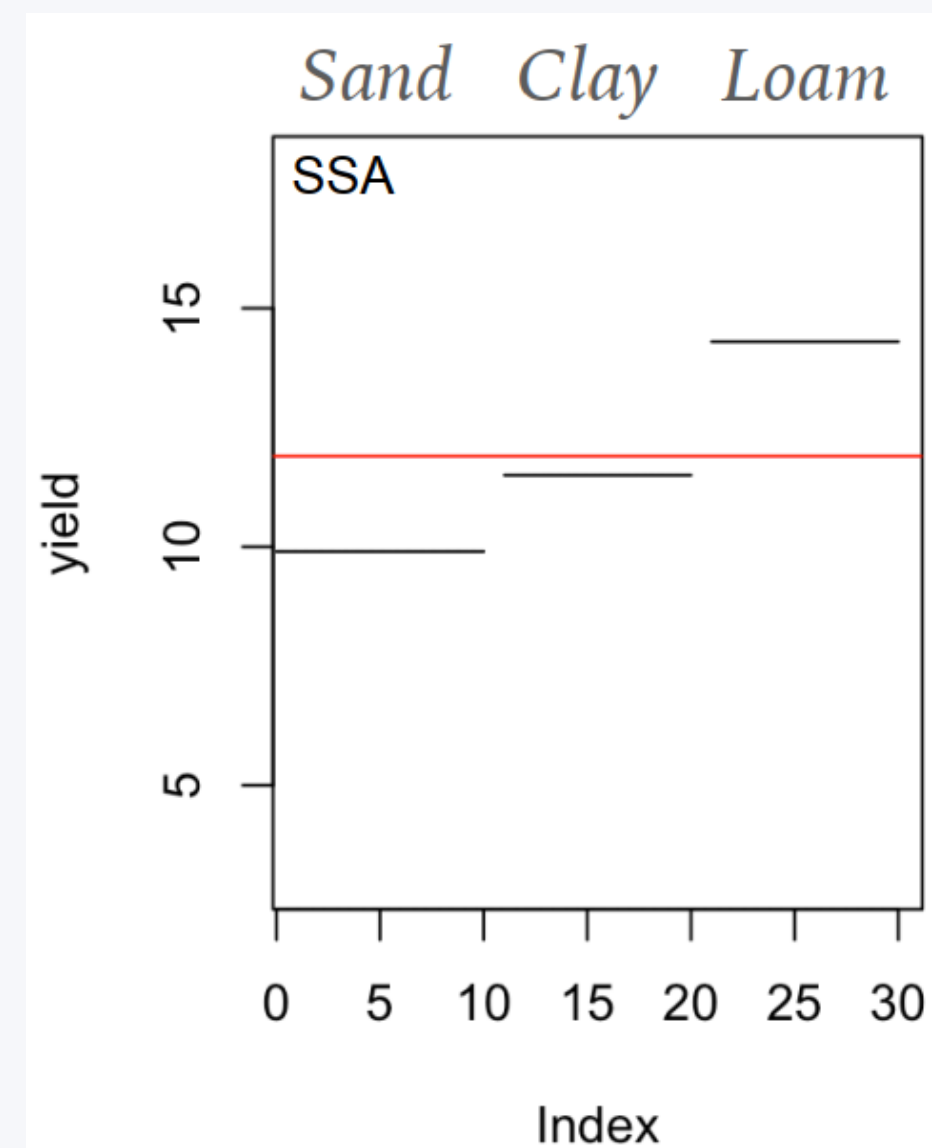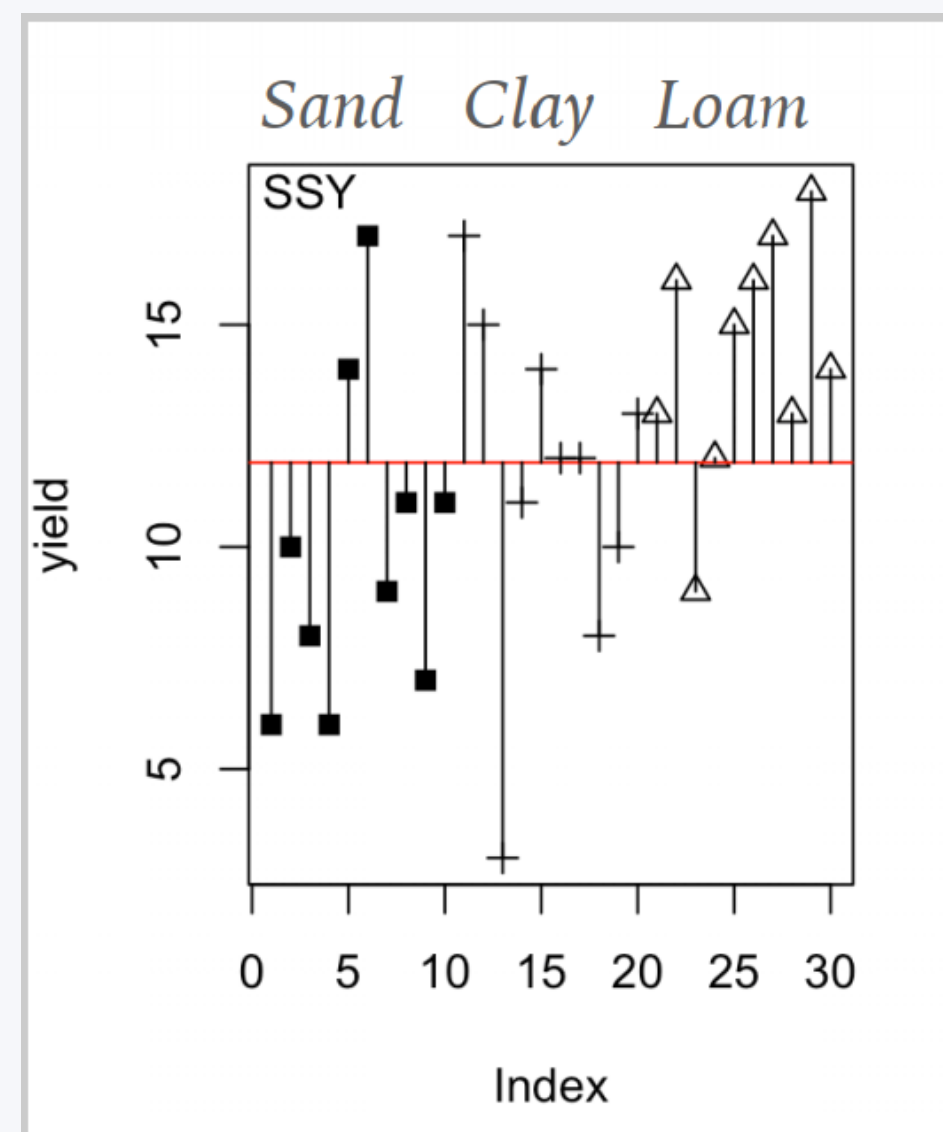(SSE)

observation – overall mean

treatment mean – overall mean

treatment observation –

treatment means

ANOVA – SOURCES OF VARIATION

Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F-Ratio
Between (SSA)	$\displaystyle\sum_i^k n_i(\bar{x}_i - \bar{x})^2$	$k - 1$	$MSA = \dfrac{SSA}{k-1}$	$\dfrac{MSA}{MSE}$
Within (SSE)	$\displaystyle\sum_i^k \sum_j^l (x_j - \bar{x}_i)^2$	$N - k$	$MSE = \dfrac{SSE}{N-k}$	
Total (SSY)	$\displaystyle\sum_i^n (x_i - \bar{x})^2$	$N - 1$		

ONE-WAY ANOVA RESULT FROM R

```
> model <- lm(yield ~ soil)
> anova(model)
Analysis of Variance Table
Response: yield
          Df   Sum Sq   Mean Sq F value Pr(>F)
soil       2   99.2     49.600  4.2447  0.02495 *
Residuals 27   315.5    11.685  ---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

$$F - ratio \ = \ \frac{SS_{between\_groups}}{df_{treatment}} \ \div \ \frac{SS_{within\_groups}}{df_{residuals}}$$

test statistic

(Variation because of
Soil Types)

(Overall Variation)

Questions?

PLAN FOR NEXT WEEK

• • •

That's it for today! :-)

Next week, we are going to discuss:

- Linear Regression + Final Q&A

If you want to reach me, mail me at:
`prabesh.dhakal@stud.leuphana.de`