Undergraduate internship 2019

Presented by Elly Huynh

Week 1:

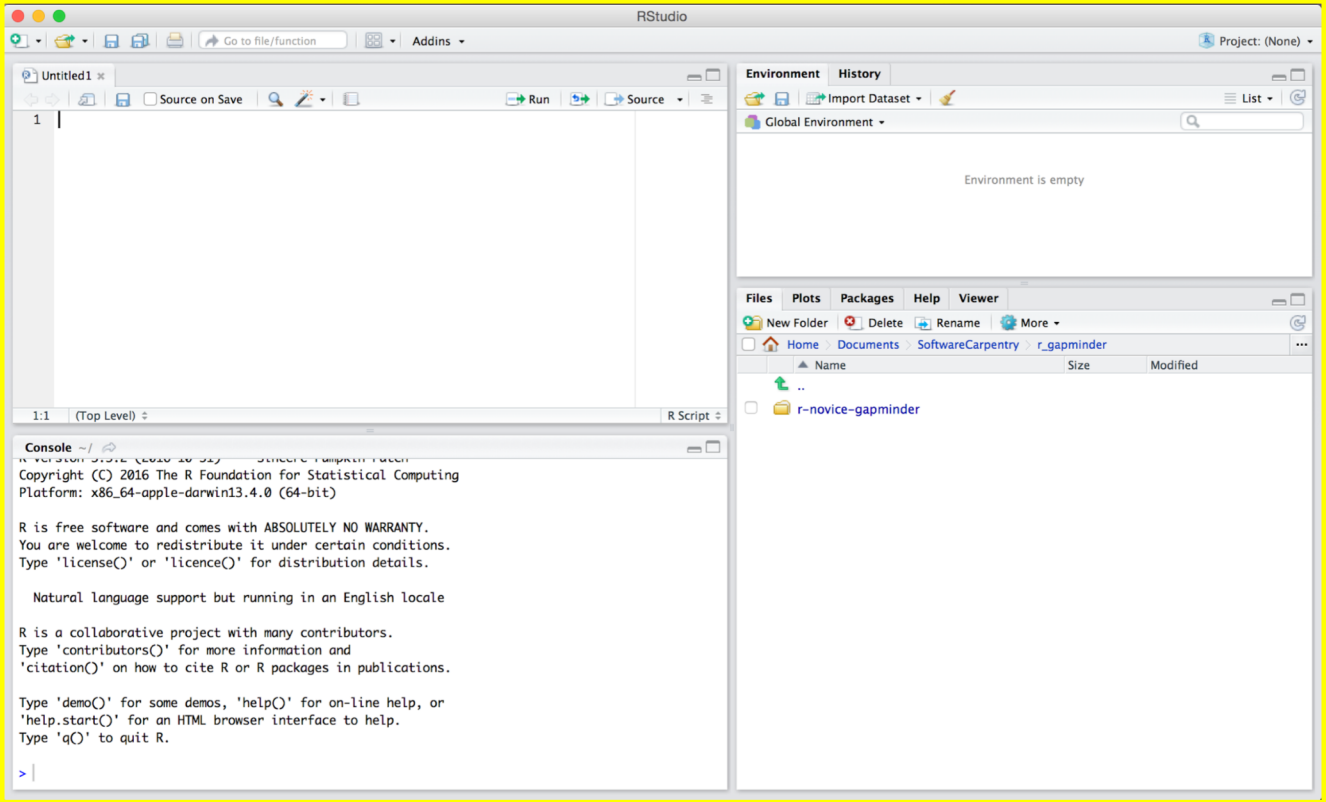
1. An Introduction to R and RStudio

2. Planning and Designing an agronomic experiment

3. Learning how to create ggplot

4. Attending Peter's presentation and a statistical meeting with Jing and Shiyu

1. An Introduction to R and RStudio



1. An Introduction to R and RStudio

- Install and load packages
- Search for help on functions
- R basics: simple calculations, variables, referencing, subsetting, practice
- Data management
- Data Frames
- Graphics in R

2. Planning and Designing an agronomic experiment

- **Terms and Definitions:**

  population, sample, treatments, replication, blocking, randomisation

  experimental and observational units, pseudoreplication,confounding

  heterogeneity, factors and levels, main effects a and interactions

- **Planning and designing an agronomic experiment:**

  Random numbers
  CRD, RCBD, Latin Square, Factorial, Split-plot
  Residual Degrees of Freedom

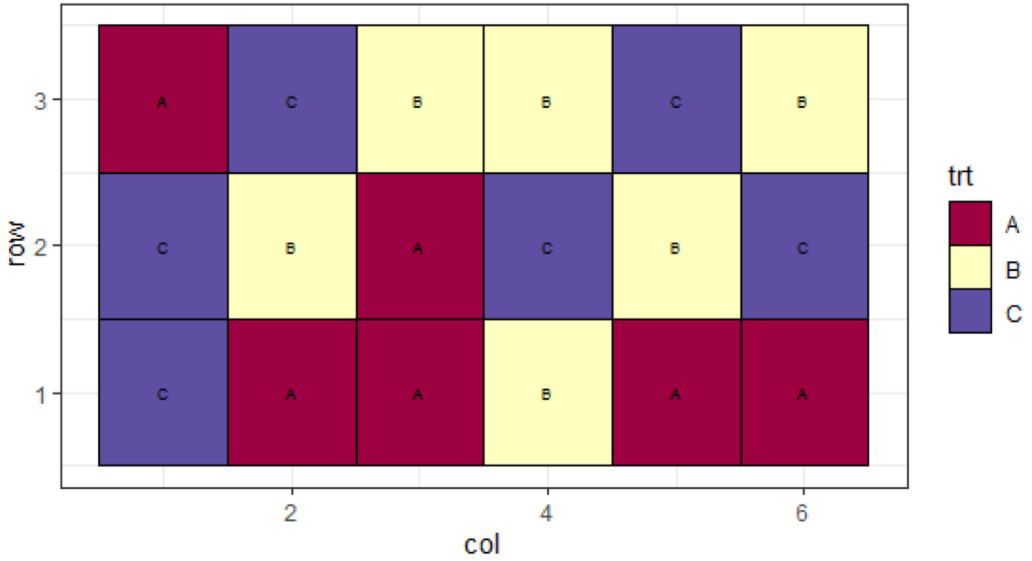


Fig. 1: Completely Randomised Design

Source of Variation	df
trt	2
Residual	15
Total	17

Fig. 2: The skeletal ANOVA table of CRD

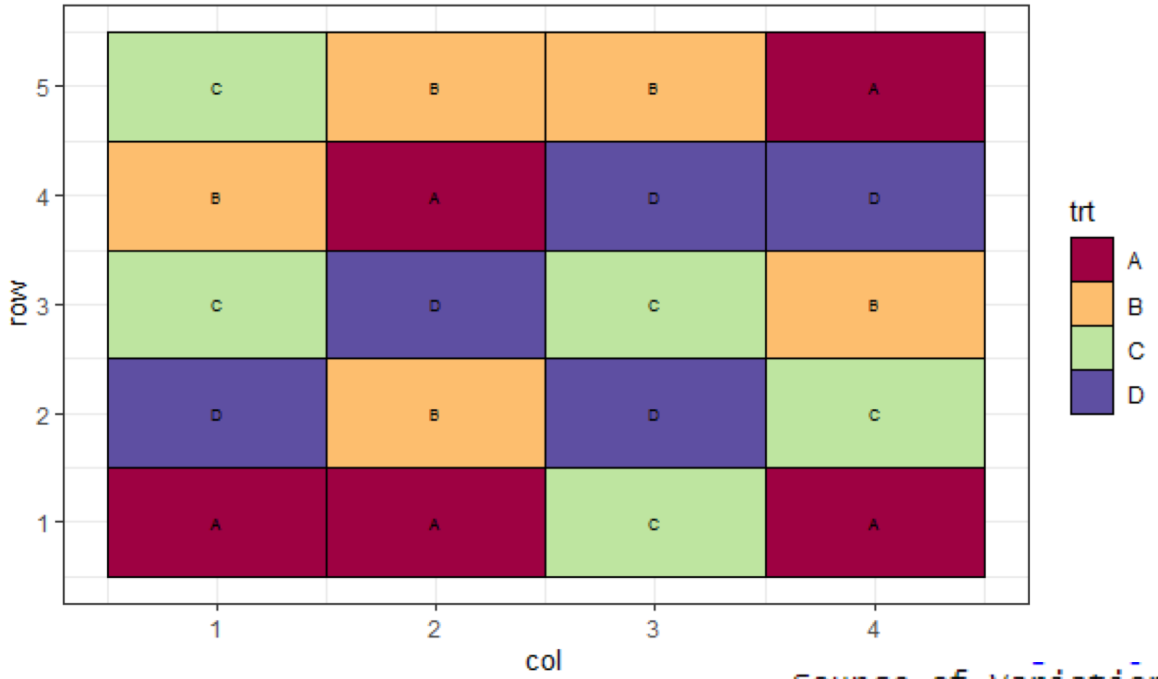


Fig.3: Randomised Complete Block Design

Source of Variation	df
Block stratum	4
trt	3
Residual	12
Total	19

Fig.4: The skeletal ANOVA table of RCBD

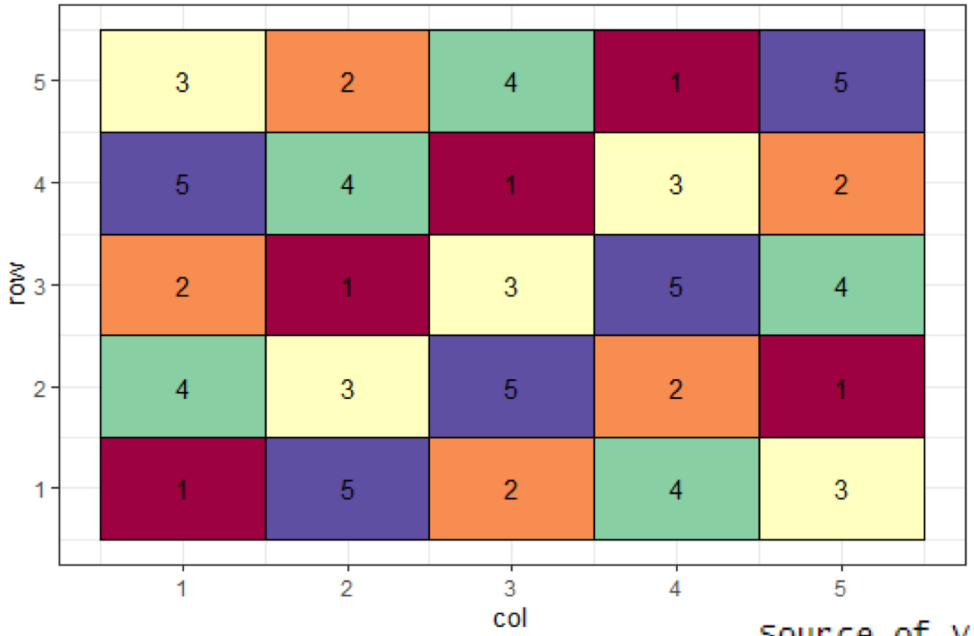
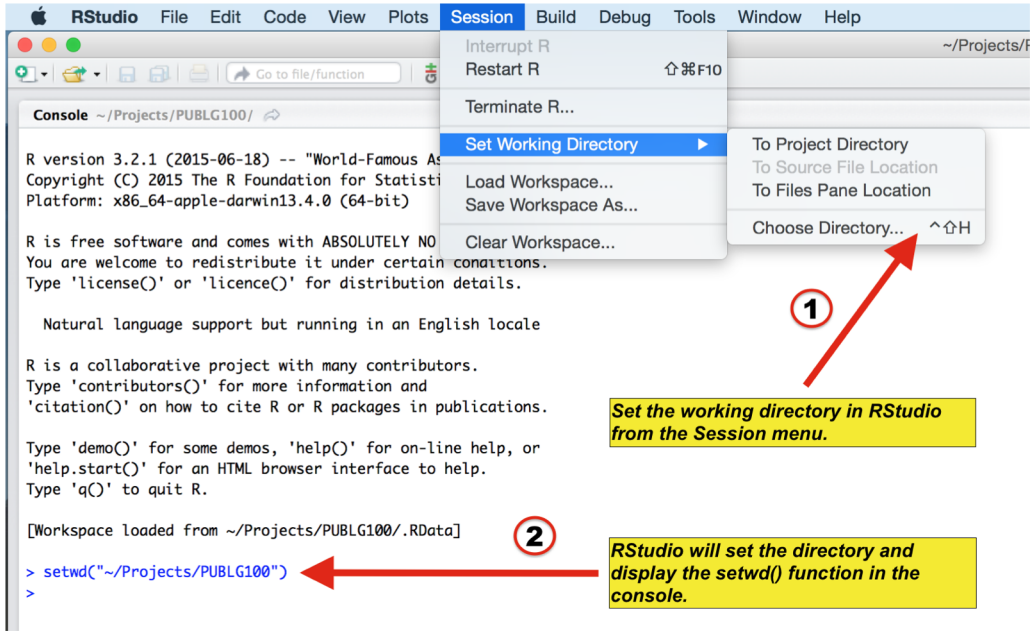


Fig.5: Latin Square Design

Fig.6: The skeletal ANOVA table of Latin Square

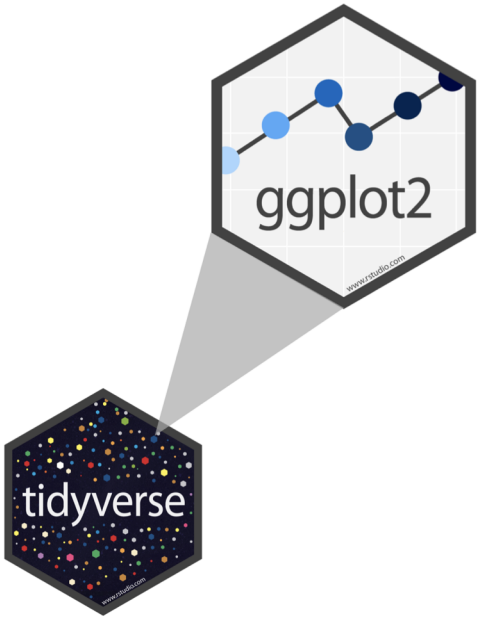
Source of Variation	df
Row	4
Column	4
trt	4
Residual	12
Total	24

Set the working directory



write.csv(des.out , "design file name.csv" , row.names = FALSE)

3. ggplot



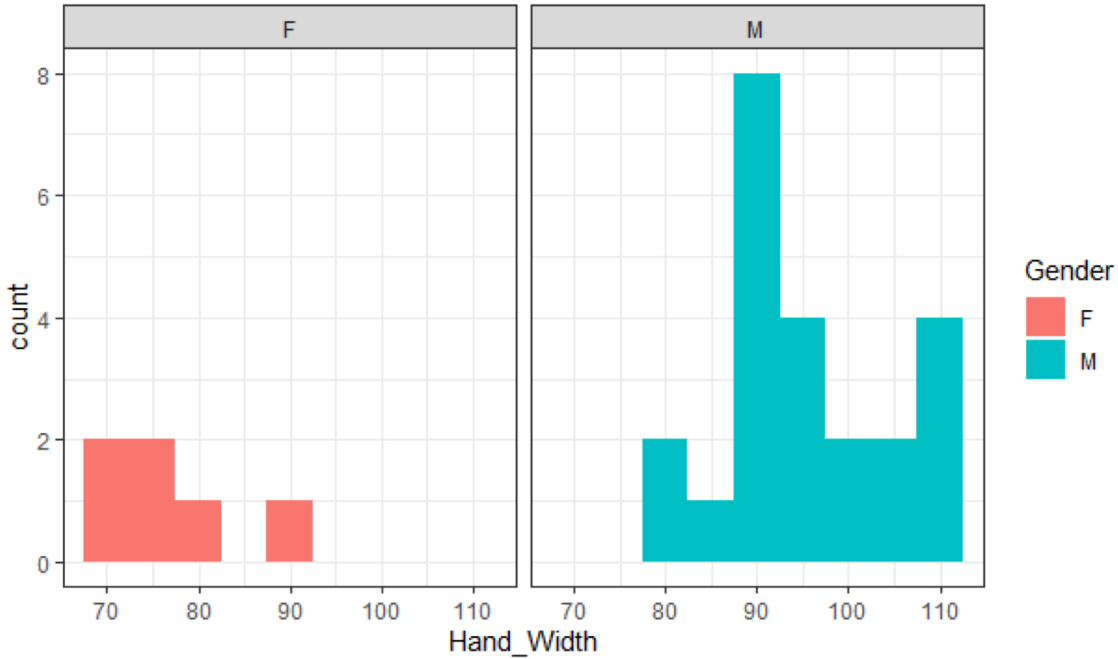
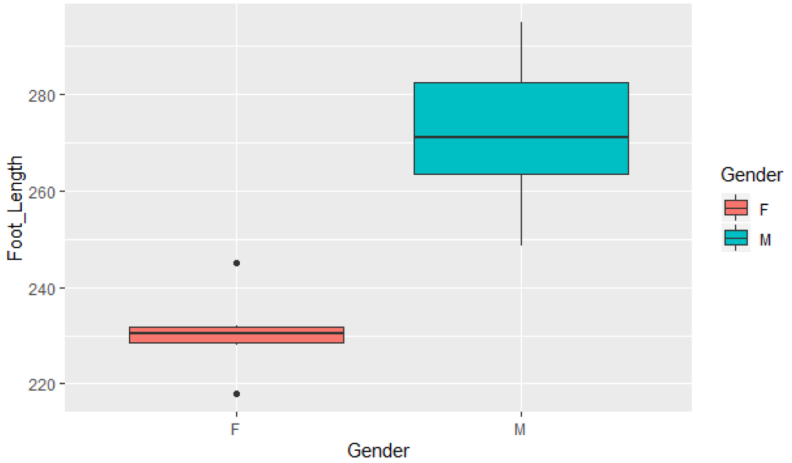
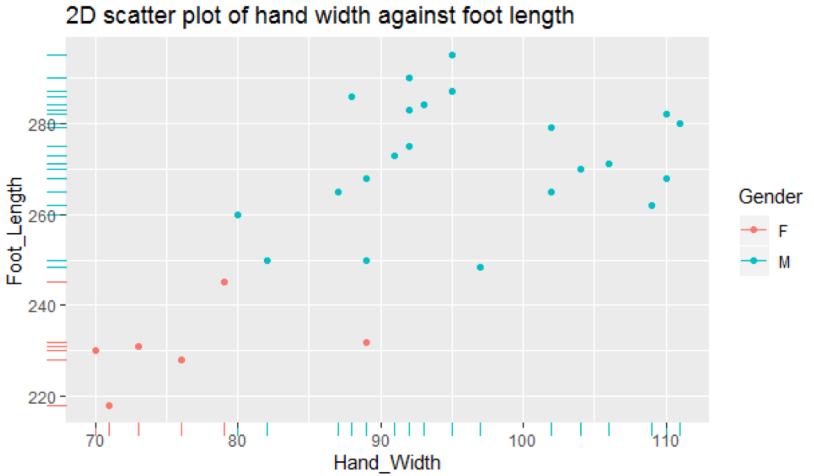
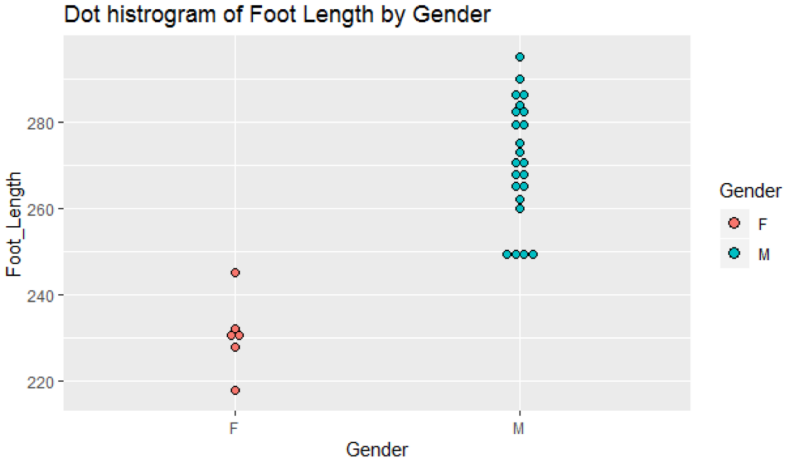
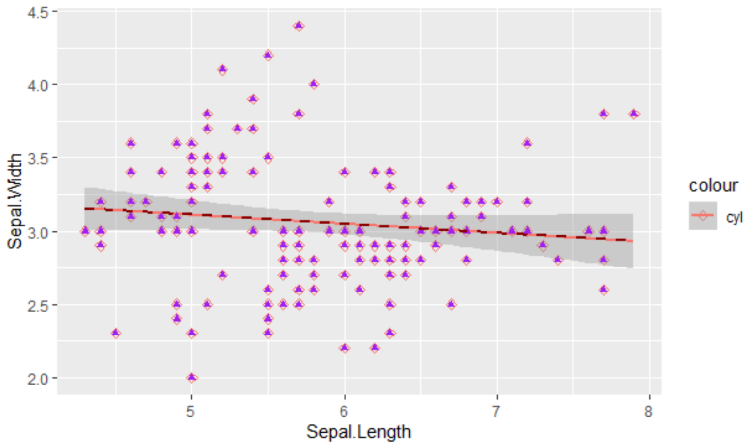


Fig.7 : histogram shows hand width of male and female


Foot_Length
280
260
240
220
F
M
Gender
Gender
F
M


2D scatter plot of hand width against foot length
Foot_Length
280
260
240
220
70
80
90
100
110
Hand_Width
Gender
F
M


Dot histrogram of Foot Length by Gender
Foot_Length
280
260
240
220
F
M
Gender
Gender
F
M


Sepal.Width
4.5
4.0
3.5
3.0
2.5
2.0
5
6
7
8
Sepal.Length
colour
cyl

4. Statistical meetings

A professional meeting with Peter:

computing program, estimate yield

A meeting with Lachlan :

Digit recognition by MINIST

A meeting with Jing and Shiyu:

CSA project, project-WGCNA analysis

Week 2:

1. Meeting with senior biometricians:

Principles of experimental designs. (Helena)

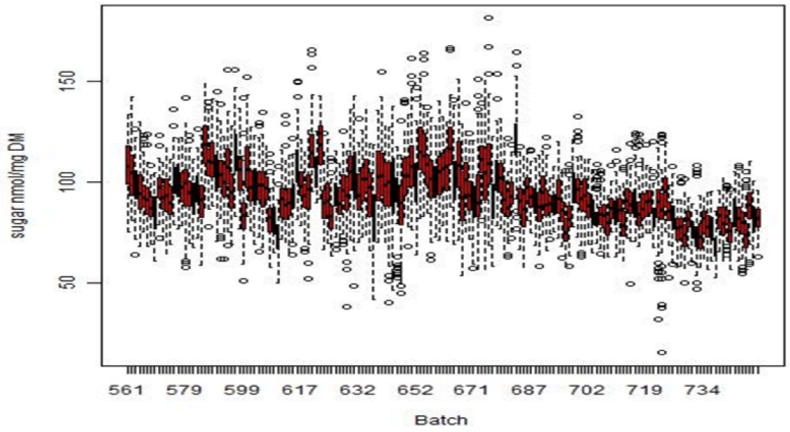
Principles of statistical inference in practical applications. (Richard)

Principles of sampling designs.(Pete)

2. R versus Genstat

3. Shiny app

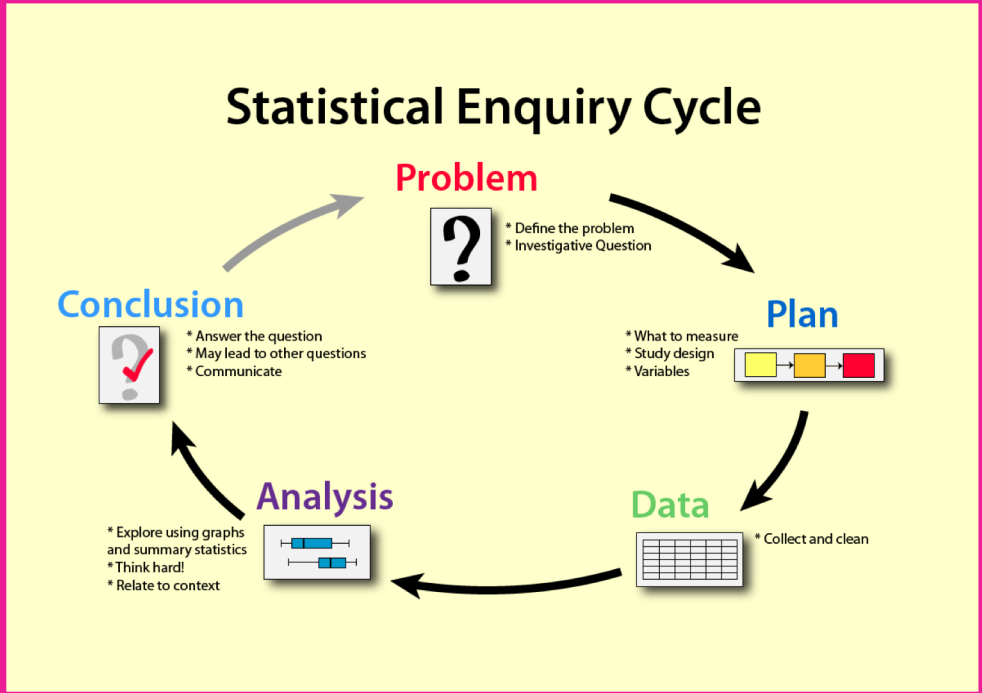
Principles of experimental designs. (Helena)



BATCH VARIATION

- Project: Biofuel potential of barley straw-the quest to find important genes controlling sugar release.
- Spatial Row-Column Design
- randomization: the field and the lab
- The purpose: minimize the chance that a particular variety occurs in the same column or row more than one time, the variation between batches
- higher variation in the lab

Principles of statistical inference in practical applications. (Richard)



Blocking: tighter result, minimise the variability of samplings.

Principles of sampling designs. (Peter)

- Why we sampling?
- Why we randomise in sampling?
- Infinite population & Finite Population

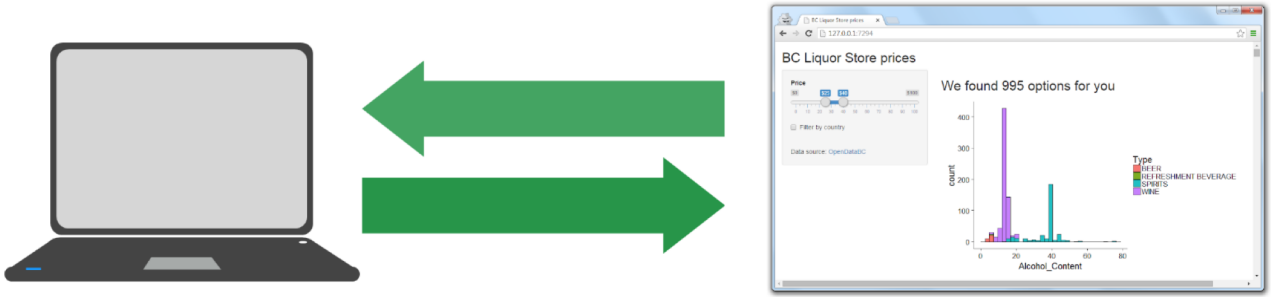
Three types of sampling:

1. Design based
2. Model-based
3. Model assistance

- Descriptive sampling
- Analytical sampling
- Pattern sampling
- Edge effects in sampling

Central limit theorem

Shiny App



Server code

User interface (UI)

open source R package

Week 3:

1. LME4 Tutorial: popularity data

2. Practical guide using the optimal design R package (OD)

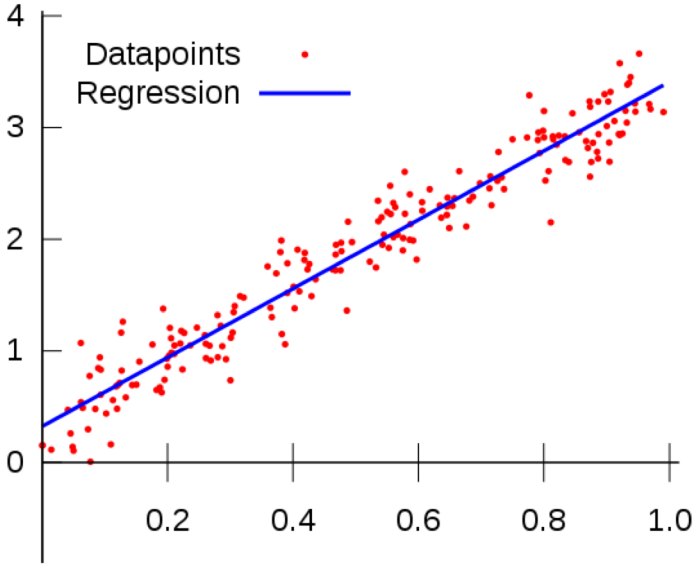
3. Basic statistical terms

1. LME4 Tutorial: popularity data

- **LME4 versus ASREML**

LME4	ASREML
open source (free, accessible for people use R)	close source (cost money and is normally used by the biometricians)
can’ t specify residual correlation structure	can specify residual correlation structure
can’t incorporate market based relationship matric	can incorporate market based relationship matric
limited variance structure available for random effects	flexible variance structure available for random effects

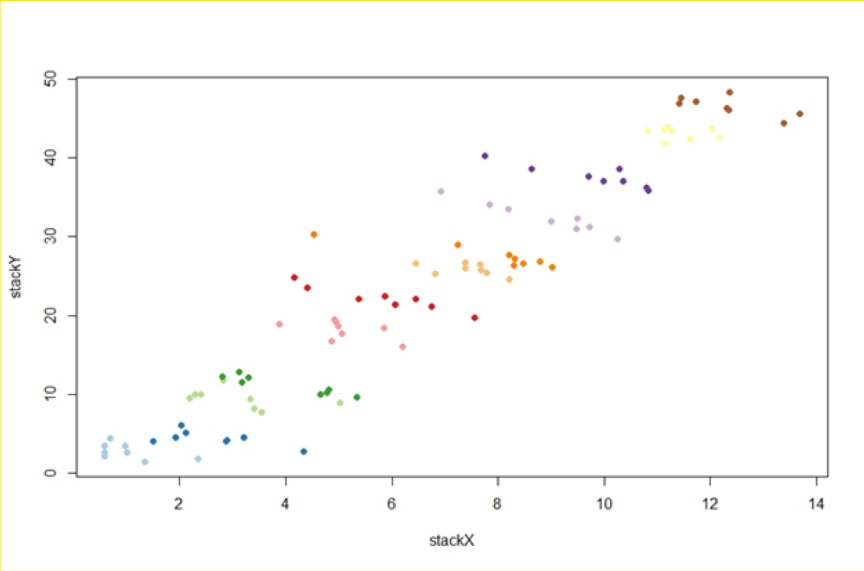
Linear models



Equation for a linear model:

$$Y = X\beta + \epsilon$$

Linear fixed effects model

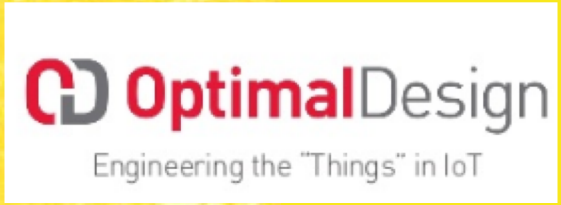


Equation for Linear fixed effects model:

$$\mathbf{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\epsilon}$$

Optimal design

a class of experimental designs that are optimal with respect to some statistical criterion.



allow parameters to be estimated without bias and with minimum variance, reduce the costs of experimentation

3. Basic statistical terms

- Type of variable and measurement:

  sample & population

  population variance, sample variance

  standard error, sum of squares, df

- Basic sata summary:

1. Discrete data
2. Continuous data
3. Quartiles and Ranges

The End


Why Some Ideas Survive and Others Die
MADE
to
STICK
Chip Heath & Dan Heath