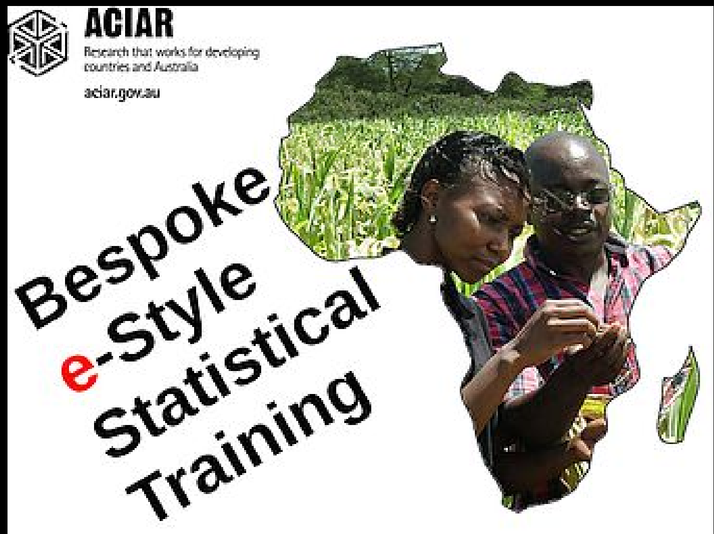
LOOKING AT SAMPLING FOR A T TEST



OUTLINE

- Just taking 2 values is not enough to do a statistical comparison – as we don't have any measures of variability for our comparison.
- For a simple 2 sample comparison we need a *sample* of each treatment
- We also have to understand what our experimental unit is
- Statistical testing
- Test statistic (t statistic)

STEPS FOR UNDERTAKING AN EXPERIMENT

- Develop the Research Question
- Remember the importance of random sampling
- Planning an experiment
- Treatments

- Collect data

REMEMBER THE THREE PRINCIPLES OF DESIGN

- Randomisation
- Replication
- Reduction of Experimental Error

WHY RANDOMISE ?

- Avoid bias.
- Important step and it must be undertaken
- We need to randomise every time if we repeat a particular experiment, say for two seasons.
- This requires practical knowledge of the context

THE RESEARCH QUESTION: COMPARE TWO VARIETIES WITHIN A RESEARCH STATION

- On the research station we could decide to compare two varieties.

- Variety A and variety B.
- How do we set up an experiment ?

- In this module we explored the principles of experimental design by replicate and randomise.

- Lets divide the land into 10 plots and plant a random 5 plots with variety A and the other 5 plots with variety B

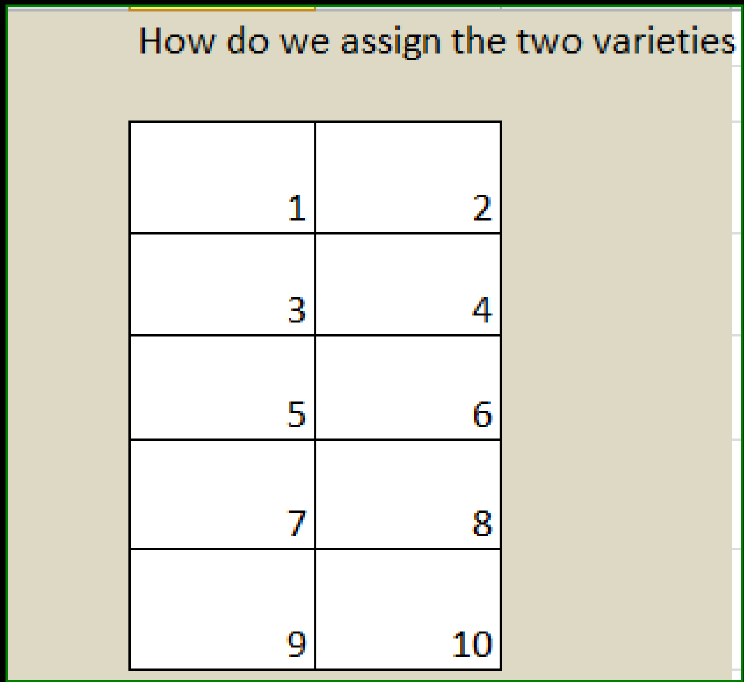
RANDOM PLOTS ON THE RESEARCH STATION

Ten plots on the research station

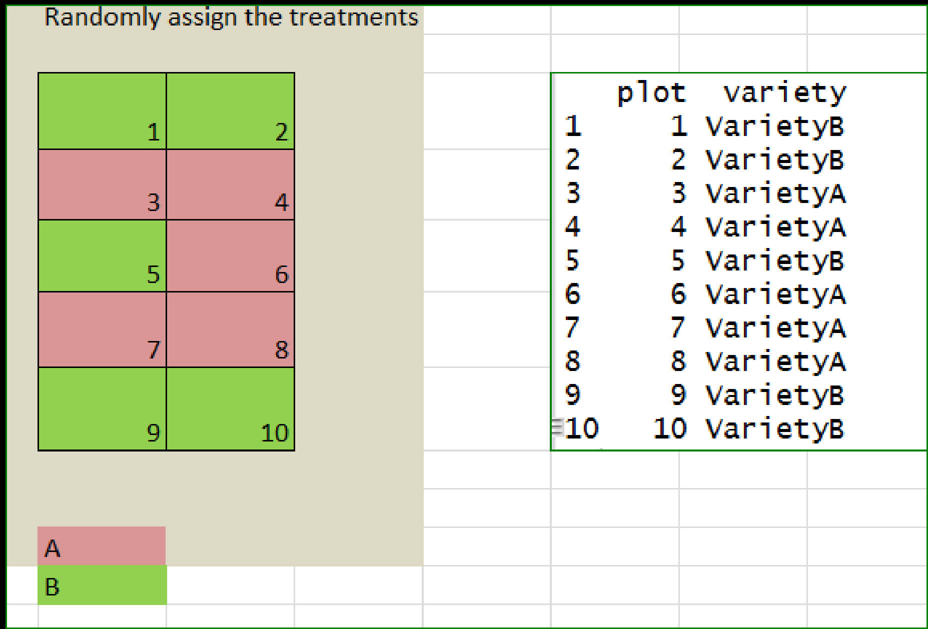
1	2
3	4
5	6
7	8
9	10

They are all adjacent plots

WE NEED TO RANDOMISE THE TREATMENT ALLOCATION TO THE TEN PLOTS (LABELLED 1 TO 10)



ASSIGN TO THE FILED. THIS IS A COMPLETELY RANDOMISED DESIGN



RANDOMISE USING R CODE, OUTPUT A SIMPLE CSV FILE

```
########################################################
##   Remember everything with a hash in front is a comment
##
##   Example of Simple randomisation using R
##
##   Our Experiment
##   We have two treatments (Variety A and Variety B)
##   and we have 10 plots of land on the research station
##
##   in a completely randomised design with two treatments
##   It will be analysed as a two sample t test
#########################################################
## The steps in this code:
## set a random seed
## specify the factor
## specify the replications
## sample the factor (fac) object ten times
## set experimental units (eu) to 10
## make a fieldplan of a randomisation
## save the random treatment association to a csv.file
## the fieldplan results are printed in the Console

set.seed(7638)
f <- as.factor( rep( c("VarietyA","VarietyB"), each = 5))
fac   <-   sample(f, 10)
eu   <- 1:10
fieldplan   <- data.frame(plot=eu, variety=fac)
write.csv(fieldplan, file= "fieldplan.csv", row.names =FALSE)

## print the randomised plan to the console
fieldplan
```

THE STATISTICAL QUESTION

- By considering exactly what the question is in terms of the measurement and our interest we can develop am hypothesis to test. The Greek letter mu (μ)stands for our mean over the replicates for that particular treatment (eg for variety A)
- In a simple comparison we only have one hypothesis
- The null hypothesis is the situation where there is no difference in our varieties
- The alternate is that there is a difference, or in other words they are not equal.
- We normally use symbols to define these two options:
  - H_o is written formally as $\mu 1 = \mu 2$?
  - H_a is written formally as $\mu 1 \neq \mu 2$?

STATISTICAL QUESTIONS

For a single sample

- Is the mean equal to a particular value ?
- Is $\mu = 0$?

For two samples

- Are the two means the same ?
- Is $\mu 1 = \mu 2$?

STATISTICAL QUESTIONS

For two samples

- Are the two means the same ?
- Is $\mu 1 = \mu 2$?

The way we set up the test is dependent on our design.
We have two samples in a completely randomised design

WHAT STATISTIC DO WE USE?

- For one sample
  - Z statistic
  - t statistic
- For two sample
  - t statistic
- The tests are called Z tests and t tests.
- Critical values depend upon the normal distribution and the t distribution.

T TEST RANDOMISED IN RSTUDIO

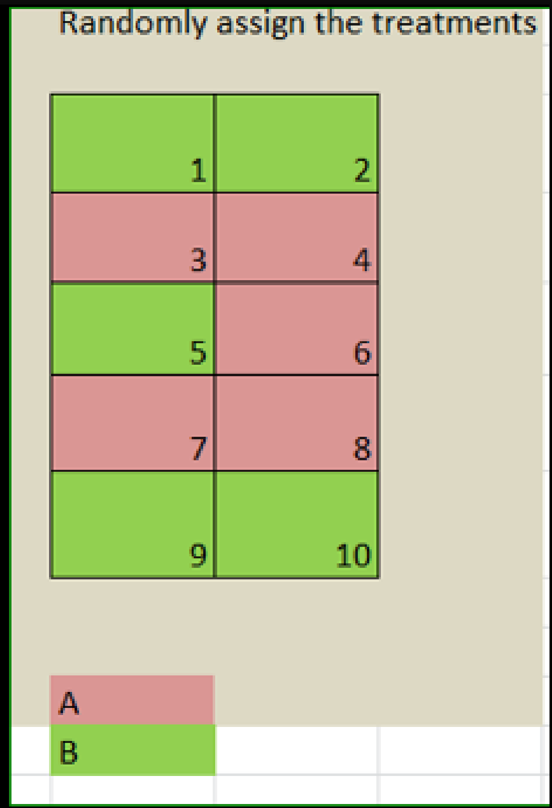
```r
## The steps in this code:
## set a random seed
## specify the factor
## specify the replications
## sample the factor (fac) object ten times
## set experimental units (eu) to 10
## make a fieldplan of a randomisation
## save the random treatment association to a csv.file
## the fieldplan results are printed in the Console

set.seed(7638)
f <- as.factor( rep( c("VarietyA","VarietyB"), each = 5))
fac  <-  sample(f, 10)
eu  <- 1:10
fieldplan  <- data.frame(plot=eu, variety=fac)
write.csv(fieldplan, file= "fieldplan.csv", row.names =FALSE)

## print the randomised plan to the console
fieldplan
```

FIELD LAYOUT FOR THE EXAMPLE

```
plot,variety
1,VarietyB
2,VarietyB
3,VarietyA
4,VarietyA
5,VarietyB
6,VarietyA
7,VarietyA
8,VarietyA
9,VarietyB
10,VarietyB
```



THE TRIAL IS RUN AND DATA ON YIELD COLLECTED

```
plot,variety,Yield
1,VarietyB,194
2,VarietyB,205.5
3,VarietyA,190.7
4,VarietyA,233.5
5,VarietyB,199.2
6,VarietyA,203.5
7,VarietyA,216.5
8,VarietyA,222.5
9,VarietyB,172.4
10,VarietyB,184
```

(this is showing you the summary of the yield as it appears in a csv file)

THE TRIAL IS RUN AND DATA ON YIELD COLLECTED

```
################################################
## The steps in this code:
# read in data  from the .csv file
# with the plan with the yields
# look  at the data
# remind yourself of the variable names
# and the number or rows in the variable "Yield"
###############################################

planY <- read.csv("planyield.csv")
View(planY)
head(planY)
names(planY)
length(planY$Yield)
hist(planY$Yield)
plot(planY$Yield)
## look at the structure of the data
## Is your variety a factor ?
## In this case it has come in as a factor
## as it was a string or text variable
str(planY)
# Plot all the data
boxplot(planY$Yield)

# Plot by variety
boxplot(planY$Yield~planY$variety)
```

ANALYSING A SIMPLE TWO SAMPLE T TEST

```
## Undertake a t test
## where y is numeric and x is a binary factor
##    t.test(y~x)

## The data is not structured, so there are no blocks
t.test(planY$Yield~planY$variety)


## What do you conclude ?
```

EXAMINE AND INTERPRET THE OUTPUT

```
## The data is not structured, so there are no blocks
t.test(planY$Yield~planY$variety)
```

```
##
##  Welch Two Sample t-test
##
## data:  planY$Yield by planY$variety
## t = 2.3585, df = 7.567, p-value = 0.04781
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.278033 44.361967
## sample estimates:
## mean in group VarietyA mean in group VarietyB
##                 213.34                 191.02
```

SUMMARY

- Design using randomisation and replication
- Run the code and write out the field layout
- Run the experiment
- Collect the data
- Look at the data, check and plot
- Run an appropriate analysis (in this case a two sample t test)