

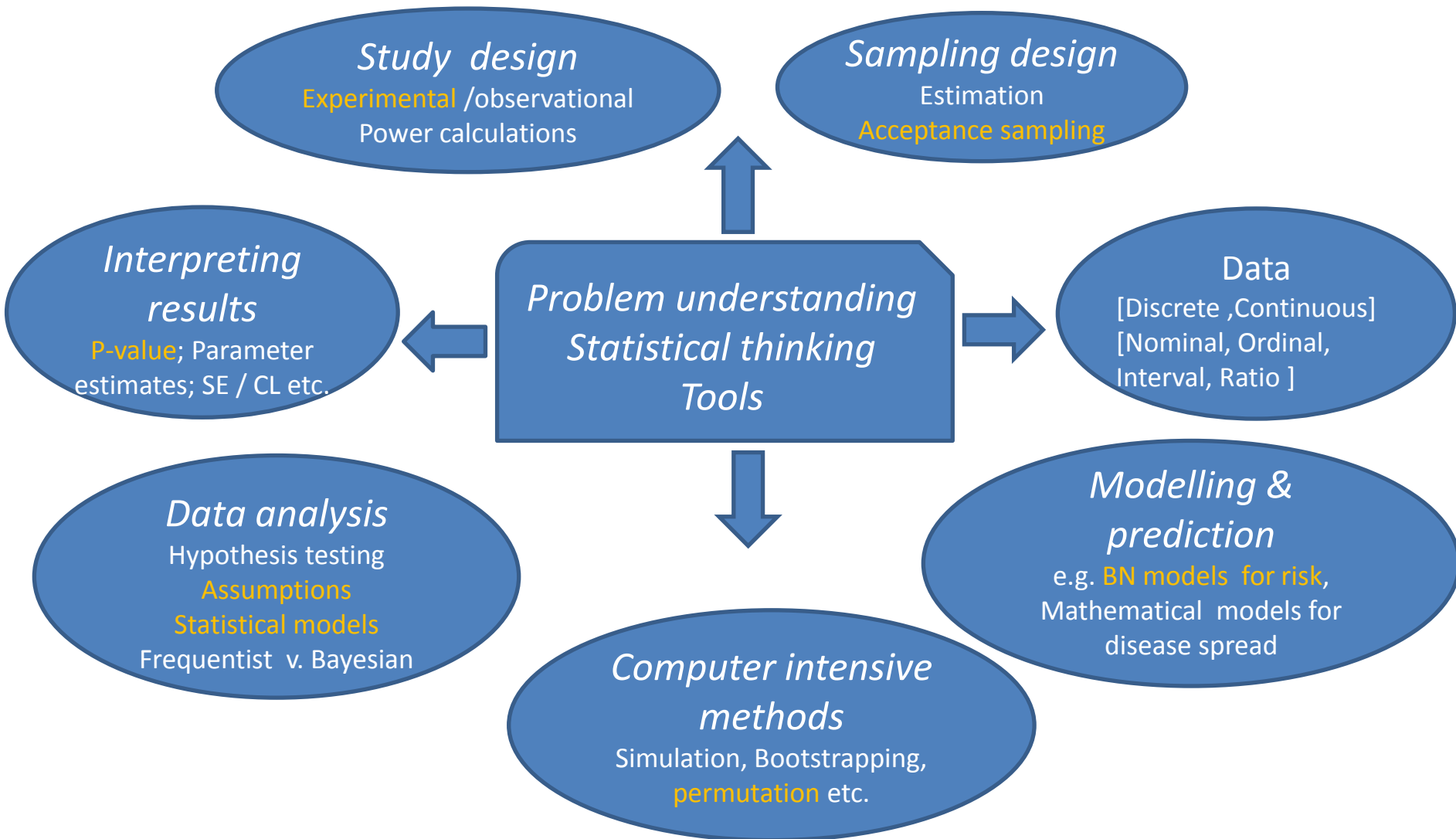
Biometrics in Plant protection science

Nihal De Silva, Mark Wohlers, and Patrick Connolly
Mt Albert Research Centre

Biometrics session, NZPPS conference
Taupo, Aug 2014



Mind map outline



Acceptance sampling: Gonick and Smith (1993), The cartoon guide to statistics

THEY PULL OUT A
SIMPLE RANDOM
SAMPLE OF 49 BOXES,
WEIGH EACH ONE, AND
DETERMINE THE
SAMPLE'S SUMMARY
STATISTICS:

$$\bar{x} = 15.90 \text{ oz.}$$

$$s = .35 \text{ oz.}$$

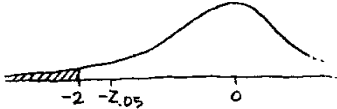
A LITTLE LIGHT—BUT
SIGNIFICANTLY SO?




THEY PLUG THE VALUES INTO THE TEST STATISTIC TO FIND

$$Z_{obs} = \frac{15.9 - 16}{.35 / \sqrt{49}} = -2$$

NOW THEY COMPUTE THE P-VALUE:

$$Pr(Z < -2 \mid H_0) = .0227$$


THIS BEING LESS THAN THE .05
SIGNIFICANCE LEVEL, GENUINE GROCERY
REJECTS THE NULL HYPOTHESIS, AND
THE SHIPMENT.



p -value is ...

- the probability under H_0 that sampling variation alone would produce an estimate equal to or more extreme than observed
 - $p\text{-value} = \Pr(T \geq T_{\text{obs}} | H_0)$
 - p -value is a function of the data
- a measure of evidence against H_0 in the traditional *frequentist hypothesis testing*
 - Smaller the p -value, the more evidence we have against H_0

p -value	Evidence against H_0
$p > 0.10$	No evidence
$0.05 < p < 0.10$	Weak evidence
$0.01 < p < 0.05$	Some evidence
$0.001 < p < 0.01$	Strong evidence
$p < 0.001$	Very strong evidence

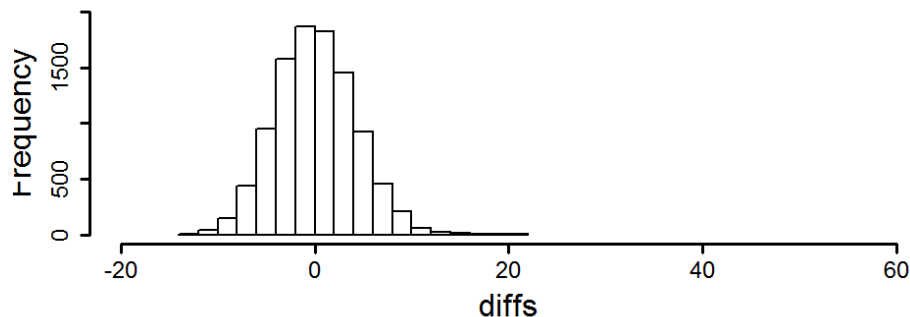
p -value is NOT ...

- evidence for H_0 .
 - a larger p -value implies no evidence against H_0 , but not evidence for H_0 .
- the *Type 1 error rate* (α) / false-positive error rate / *significance level*
 - Type 1 error rate is set at the beginning of an experiment (it is not a function of *data*). *Type I error rate* = Pr(rejecting a true null hypothesis)
 - When designing an experiment we want to decide on the number of replicates required for a given Type 1 error rate, detection level, and power
 - If $p\text{-value} \leq \alpha$ we reject H_0 and the test is said to be *statistically significant*, otherwise it is *non-significant*
- probability that H_0 is true

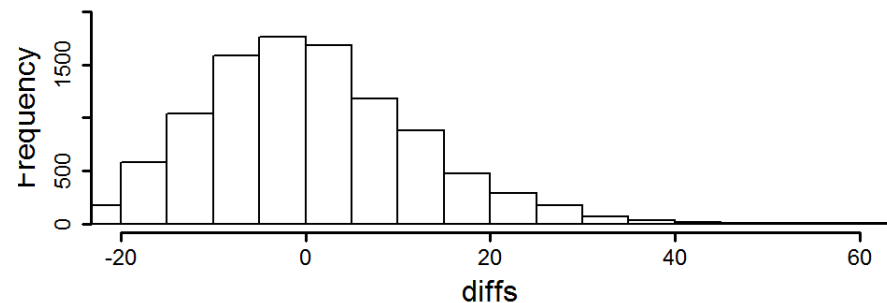
Every statistical test has assumptions...

- E.g. A *parametric* test statistic assumes the data come from a known distribution, e.g. normal
- *Central limit theorem*: when sample size gets large sampling distribution of the mean of any distribution tends to be normal
- ```
n1 <- 6; n2 <- 6; mean1 <- 25; mean2 <- 25;
diffs <- replicate(10000, mean(arexp(n1, 1/mean1)) -
mean(runif(n2, 0, 2*mean2))); hist(diffs)
```

n=50



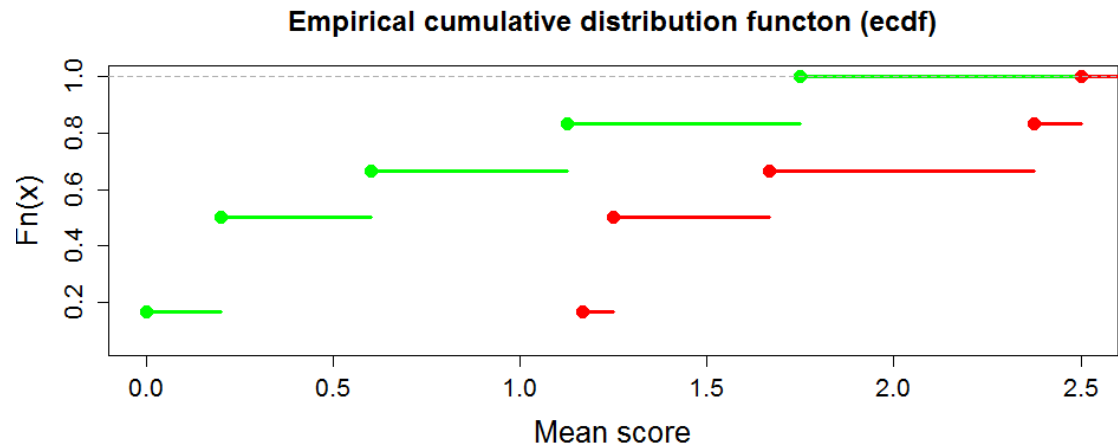
n=6



# To make a simple conclusion the analysis doesn't have to be always simple!

## • Psa treatment screening study

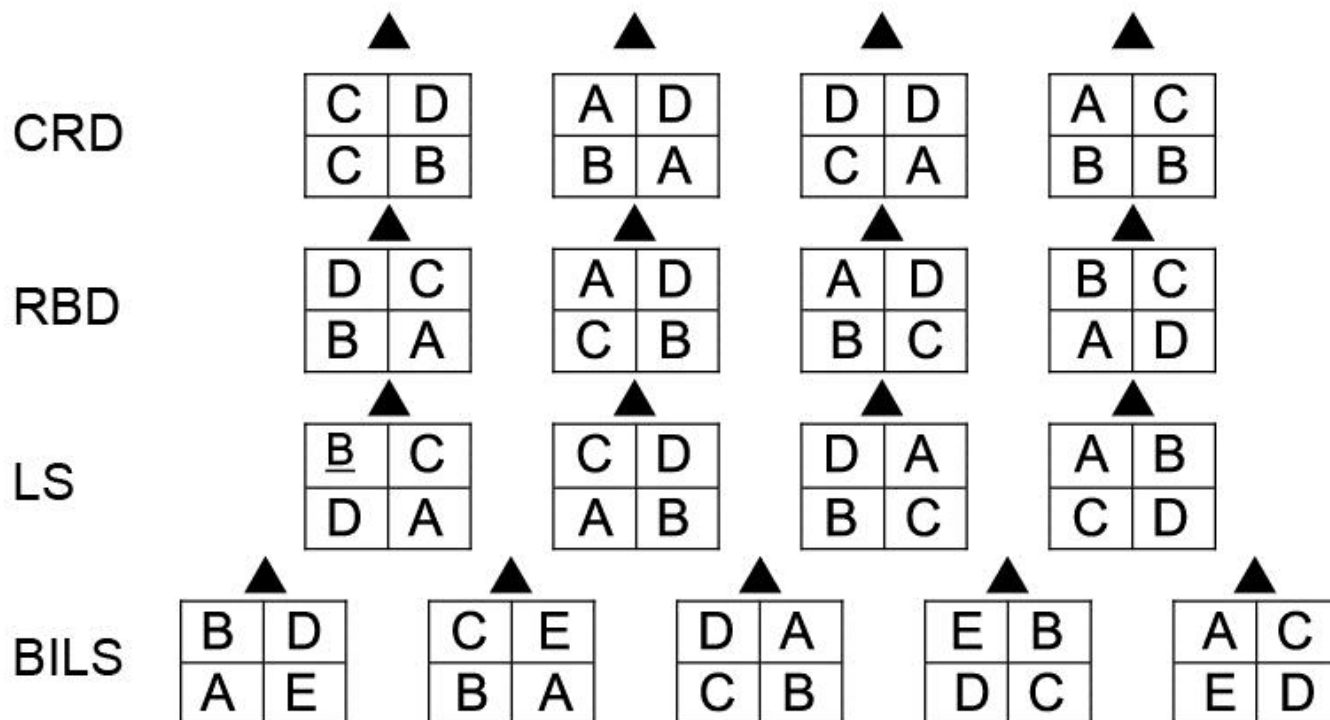
|      | Mean disease score |         |
|------|--------------------|---------|
|      | +Control           | Treated |
|      | 1.17               | 1.75    |
|      | 1.25               | 1.13    |
|      | 1.67               | 0.20    |
|      | 1.25               | 0.00    |
|      | 2.38               | 0.60    |
|      | 2.50               | 0.20    |
| Mean | 1.70               | 0.65    |
| SE   | 0.24               | 0.27    |



- `>t.test(tmt, con, alternative = "l")`  $t = -2.8723$ ,  $df = 9.862$ , **p-value = 0.008414**
- `>perm.test(tmt, con, paired=FALSE, exact=TRUE, alternative="l")`  $T = 18$ , **p-value = 0.008658**
- `>tmt <- tmt+ runif(6,0, 0.1)` `>con <- con+runif(6,0, 0.1)`
- `>wilcox.test(tmt, con, alternative = "l", paired=FALSE)`  $W = 5$ , **p-value = 0.02056**
- `>ks.test(tmt, con, alternative = "g", exact=NULL)` **p-value = 0.06948**

# Expt. design concepts: a hypothetical example

- Experimental layouts for comparing the wear and tear of different brands of tyres





# Statistical models and their assumptions

- LM – Linear model

$$y_{ij} = \mu + T_i + B_j + \beta_i x_{ij} + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma^2)$$

- LMM – Linear mixed model

$$B_j \sim N(0, \sigma_B^2)$$

- GLM – Generalised linear model, e.g. Logistic regression

$$y_{ij} \sim \text{Bin}(n_{ij}, \pi_{ij})$$

$$\log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = T_i + B_j + \beta x_{ij}$$

- GLMM – Generalised linear mixed model

- Non-linear models

# Bayesian Network (BN) models

A general BN model for fruit import risk assessment DRAFT V3, 25/7/2013

