

Statistical Thinking in Biology Research

Terry Neeman and Timothee Bonnet

November 16, 2018

Acknowledgements and warning

Key ideas for today

- Statistics in biology is the study of biological variation

Key ideas for today

- Statistics in biology is the study of biological variation
- Statistical ideas about biological variation inform the design of experiments

Key ideas for today

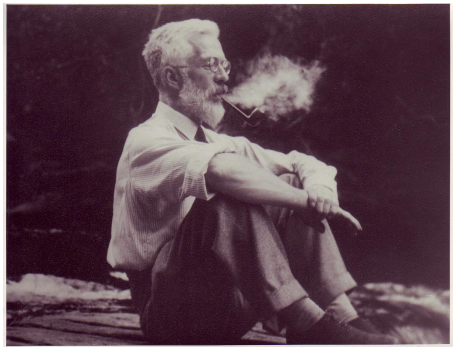
- Statistics in biology is the study of biological variation
- Statistical ideas about biological variation inform the design of experiments
- Statistical ideas about biological variation inform the analysis of experiments

Key ideas for today

- Statistics in biology is the study of biological variation
- Statistical ideas about biological variation inform the design of experiments
- Statistical ideas about biological variation inform the analysis of experiments
- Statistical thinking is an essential component of scientific thinking

A bit of history of statistical methods

R.A. Fisher: 1890-1962



Statistical Principles for Research Workers (1925)

A bit of history of statistical methods

R.A. Fisher: 1890-1962

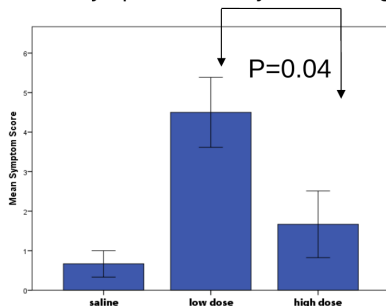


Statistical Principles for Research Workers (1925)

Cautionary tales from the front

Message 1: A small p-value is not always evidence of a treatment effect

Mean symptom score by treatment group



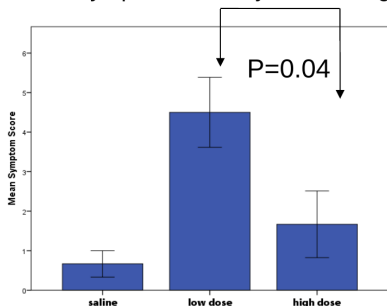
Vaccine challenge experiment:

- 6 mice/group (saline/low dose/high dose)
- All mice challenged with *Shigella*
- Followed for 14 days
- Outcome: Symptom score average Days 2 - 8

One-way ANOVA (post-hoc Bonferroni) $p=0.04$

Message 1: A small p-value is not always evidence of a treatment effect

Mean symptom score by treatment group



Vaccine challenge experiment:

- 6 mice/group (saline/low dose/high dose)
- All mice challenged with *Shigella*
- Followed for 14 days
- Outcome: Symptom score average Days 2 - 8

One-way ANOVA (post-hoc Bonferroni) $p=0.04$

Do you think the vaccine works? What is strange?

Message 1: A small p-value is not always evidence of a treatment effect

Message 1: A small p-value is not always evidence of a treatment effect

Experimental design

The observed difference in outcome could be the result of:

- Cage effects
- Mouse strain effects

These effects are **CONFOUNDED** with treatment effect



Cage 1:
saline



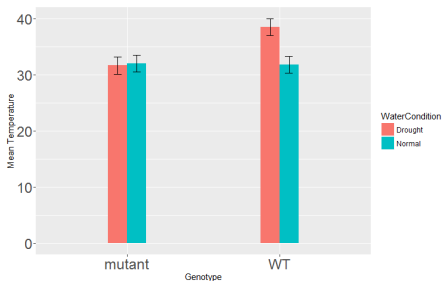
Cage 2:
Low Dose



Cage 3:
High Dose

Message 2: p-values from simple comparisons cannot tell us when differences are “different”

Message 2: p-values from simple comparisons cannot tell us when differences are “different”



Are temperature mechanisms modified in a genetically modified tomato plant?

- Genotypes: WT/mutant
- Water condition: Normal/Drought
- Leaf temperature measured

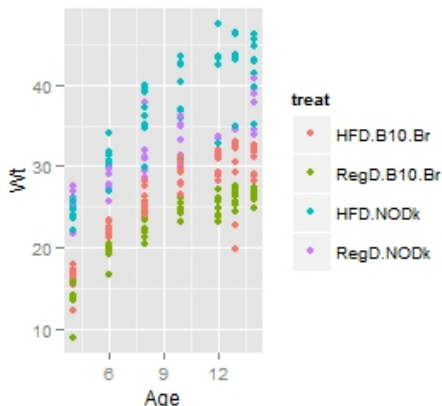
Comparisons made using t-tests

Evidence of difference + No evidence of difference \neq Evidence that differences are different.

Message 3: Interpreting experimental results needs more than t-tests

Message 3: Interpreting experimental results needs more than t-tests

Research question: Are mice susceptible to obesity when exposed to a high fat diet?



Experimental set-up:

- 37 mice: 16 NODk /21 WT
- Randomised to either regular or high fat diet
- Monitored for 14 weeks
- Outcome measure: Body weight (g)
- Experimental factors: Diet (2), Strain (2), Time (8)

Acknowledgements: Ainy Hussain, PhD student 2013

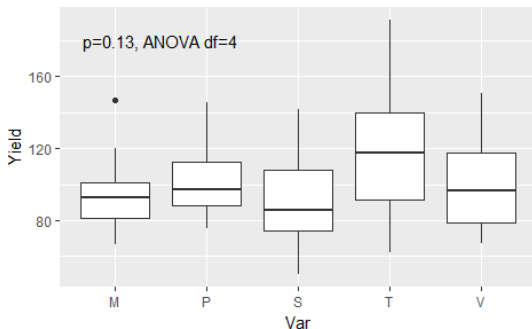
Message 4: Knowing how to combine information across subgroups can improve inference

Message 4: Knowing how to combine information across subgroups can improve inference

Comparing yield in five barley varieties (1930s)

Experimental factors: 5 varieties of barley, 6 locations, 2 time points.

Outcome measure: yield



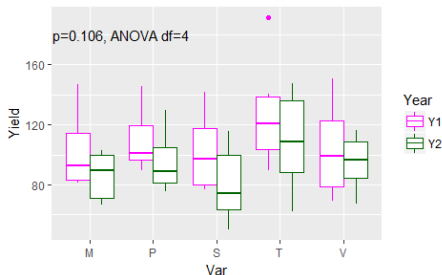
Acknowledgements: MASS R-package

Message 4: Knowing how to combine information across subgroups can improve inference

Comparing yield in five barley varieties (1930s)

Experimental factors: 5 varieties of barley, 6 locations, 2 time points.

Outcome measure: yield



Controlling for other sources of variation:

- Controlling for year = comparing yield **WITHIN** years and combining these

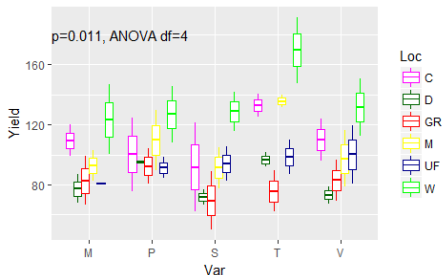
Acknowledgements: MASS R-package

Message 4: Knowing how to combine information across subgroups can improve inference

Comparing yield in five barley varieties (1930s)

Experimental factors: 5 varieties of barley, 6 locations, 2 time points.

Outcome measure: yield

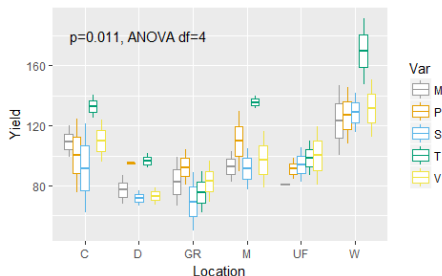


Controlling for other sources of variation:

- Control for year = compare yield WITHIN years and combine these
- Control for location = compare yield WITHIN locations and combine these

Acknowledgements: MASS R-package

Message 4: Knowing how to combine information across subgroups can improve inference



Controlling for other sources of variation:

- Control for year = compare yield WITHIN years and combine these
- Control for location = compare yield WITHIN locations and combine these

Acknowledgements: MASS R-package