Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
0	23	1343.6	58.41739	820.607		
1	13	1234.8	94.98462	136.6497		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	11105.87	1	11105.87	19.17416	0.000108	4.130015
Within Groups	19693.15	34	579.2103			
Total	30799.02	35				

Ecology Lab Introduction

I thought this was ecology—not statistics

Modern ecology includes statistics

- Overkill or essential?

Canoco for Windows

PC-ORD for Windows

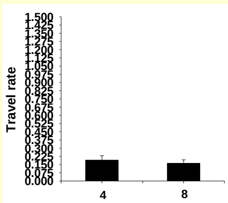
EstimateSWin750
4th Dimension
4D

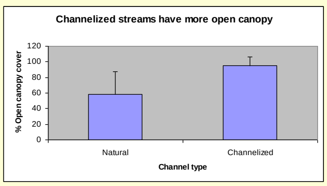
SPSS

The SAS System for Windows V8

Today's lab

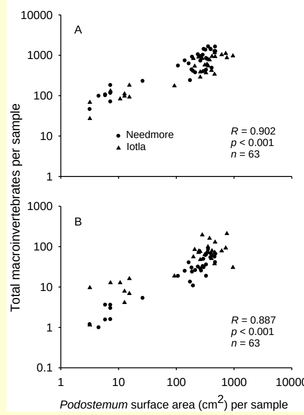
- Good graphical techniques
- Basic statistics
- A brief write-up





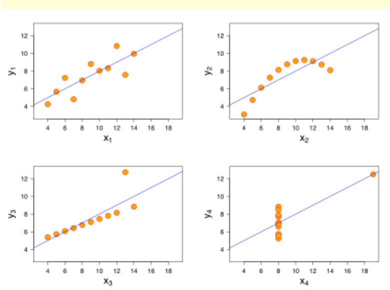
Three points to remember…

- First, **look** at the data!
- Second, see if you can apply the **'O test'**
- Third, **results come first**, statistics second



Anscombe's quartet

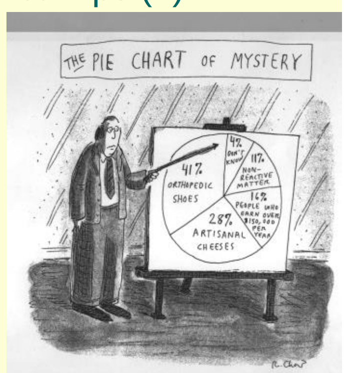
- How statistics can unintentionally lie to you
- Four very different datasets, all with the same basic statistics



Statistic	Value
X mean	9
X variance	11
Y mean	7.5
Y variance	4.122

Graphical tips (1)

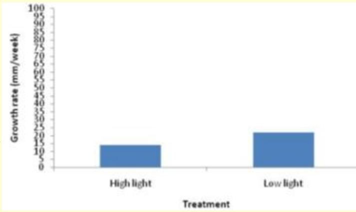
- Above all else, **show the data**
- Avoid distortion--don't lie with numbers
- For *written* reports, use a figure caption; for *oral* reports, use a descriptive title



Thanks to Chris Hill, Edward Tufte, and Jack Webster

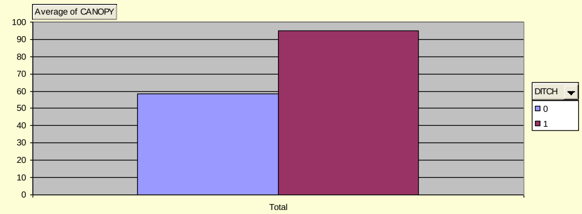
Graphical tips (2)

- Use space efficiently
- Draw axes to fit data
- Use 5 to 10 tick marks on the axes (fewer can sometimes be ok)
- Make sure the data points are big enough



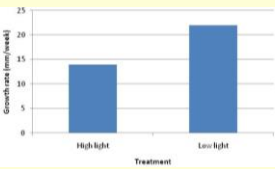
Graphical tips (3)

- Label axes clearly and legibly (units!)
- Use labels on the graph itself, rather than putting a legend in the margin



Graphical tips (4)

- Edward Tufte's tips
  - Maximize the data-ink ratio
    - Erase non-data ink
    - Erase redundant data ink
  - **Revise and edit**




PRINCETON UNIVERSITY Admission Office

December, 2000

Dear Sam,

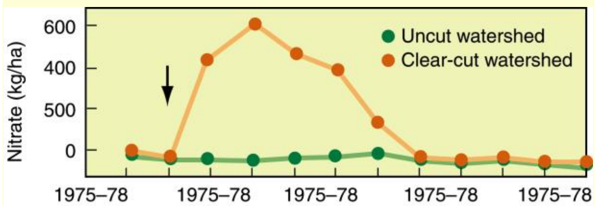
Yes!

We are happy to offer you admission to Princeton University and are delighted to welcome you as a member of the class of 2005.

Sincerely,

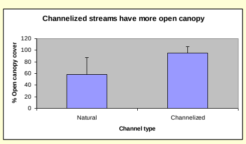
Steve LeMenager
Acting Dean of Admissions

A test!



Now, some stats…

- Descriptive statistics
- Comparing groups
- Examining relationships



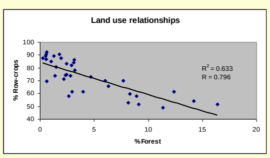
Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
0	23	1343.6	58.41739	820.607
1	13	1234.8	94.98462	136.6497

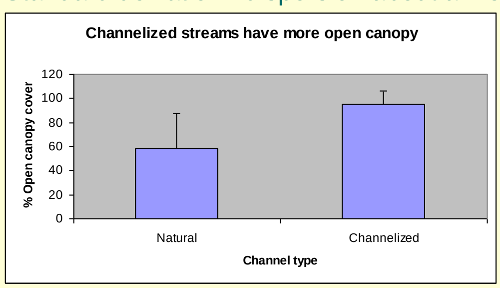
ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	11105.87	1	11105.87	19.17416	0.000108	4.130015
Within Groups	19693.15	34	579.2103			
Total	30799.02	35				



Descriptive statistics

- Mean or average: central tendency
- Standard deviation: dispersion about a mean



Comparing groups

- Comparing the means of <u>two</u> groups: **t-test**
- Comparing the means of <u>more than two</u> groups: analysis of variance (**ANOVA**)
- Recall, complete analysis = graphs AND stats…
  - Comparing groups: **bar charts** (= column charts in Excel)

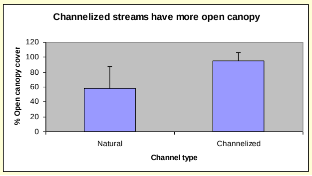
Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
0	23	1343.6	58.41739	820.607
1	13	1234.8	94.98462	136.6497

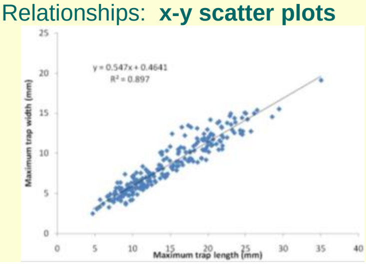
ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	11105.87	1	11105.87	19.17416	0.000108	4.130015
Within Groups	19693.15	34	579.2103			
Total	30799.02	35				



Examining relationships

- 'Cause-effect' type of relationship between two variables: **regression**
  - Also: prediction, explaining variation
- Recall, complete analysis = graphs AND stats…
  - Relationships: **x-y scatter plots**

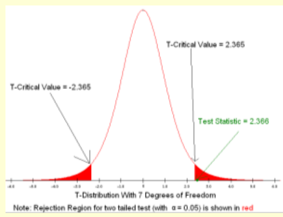


Significance

- In science, you <u>only</u> use the word '**significant**' if you have run statistical tests to see if your results are different than what would be expected from chance
- So how do you determine significance from statistical output?
  - Usually through the use of 1 or 2 pieces of statistical information:
    - **Calculated and critical test statistics**
    - **P-values**

Interpreting significance: test statistics

- A result is significant if:
  - *The calculated test statistic is **greater** than the critical (or theoretical) test statistic*
  - Example:
    - T_{calc} = 5.34, T_{crit} = 1.99; Because $T_{calc} > T_{crit}$, there is a significant difference between the two means
- The values for the critical test statistics depend on:
  - the test being run,
  - the sample size,
  - and the probability level being used (usually α = 0.05)



http://www.algebra.com/algebra/homework/Probability-and-statistics/Probability-and-statistics.faq.question.588398.html

Interpreting significance: P-values

- A result is significant if:
  - *The computer-calculated P-value is **less** than a pre-defined α-value (usually 0.05)*
  - Example:
    - P-value = 0.01; Because the P-value < 0.05, then there is a significant relationship between the two sets of data
- P-values, or probability values, describe the probability that observed or more extreme differences would be found if the *null hypothesis* is true
- The pre-defined α-value can vary depending on the needs of the researcher, but scientists often settle on 0.05 (i.e., reaching the wrong conclusion 5% of the time assuming the null hypothesis is true) as being an acceptable chance of error

Easier to observe significance if…

- You have a well-designed study
- You have many replicates
- You have *little variability within* the treatments or factors in your study
- You have large differences between the variables of interest (i.e., *more variability between* the treatments or factors)