

# Technologies for Data Analysis for Experimental Biologists

Melanie I Stefan, PhD

HMS

15 May 2014

## Before we start

Download files for today's class

- Go to <https://nanosandothercourses.hms.harvard.edu/> and log in
- Download all new files

# Outline

- 1 Review, homework
- 2 Sampling distributions
- 3 t test
- 4 Wilcoxon's rank sum test
- 5 Hypothesis testing considerations
- 6 Testing practise
- 7 Sample size and power
- 8 Correlations
- 9 Questions

# Review(Powerpoint Karaoke version)

Last week ...

- First steps with JMP
- Descriptive statistics
- Normality
- Z scores
- Sampling distributions and confidence intervals

# Homework

1. The file `bunnies.xls` contains a list of weights of 500 rabbits. Select the parameters that best describe your dataset, and plot them.

# Homework

2. The file dragons.xls contains a list of wingspans of 500 dragons. Select the parameters that best describe your dataset, and plot them.

# Homework

3. How do the means of both datasets compare? How do their standard deviations compare?

# Homework

4. Are your datasets normal? For each one, give three reasons as to why or why not.



# Outline

- 1 Review, homework
- 2 Sampling distributions**
- 3 t test
- 4 Wilcoxon's rank sum test
- 5 Hypothesis testing considerations
- 6 Testing practise
- 7 Sample size and power
- 8 Correlations
- 9 Questions

# Sampling distributions



# Sampling distributions

## Sampling distributions in JMP

- Go to [http://www.jmp.com/academic/learning\\_modules.shtml](http://www.jmp.com/academic/learning_modules.shtml)
- Under 2: Sampling Distribution of Sample Means, click Download Now (requires you to register)
- Or retrieve script from today's course materials:  
02Dist\_Sample\_Means.jsl\_.txt (remove “\_.txt” after download)
- The script can be opened from within JMP

# Sampling distributions - example

- Bunnies.xls
- Use Sampling Distribution of Sample Means script
- Under Population Shape, select My Data
- Select sample size and number of samples

# Sampling distributions - exercise

- Dragons.xls
- Use Sampling Distribution of Sample Means script
- Under Population Shape, select My Data
- Select sample size and number of samples

# Outline

- 1 Review, homework
- 2 Sampling distributions
- 3 t test**
- 4 Wilcoxon's rank sum test
- 5 Hypothesis testing considerations
- 6 Testing practise
- 7 Sample size and power
- 8 Correlations
- 9 Questions

# t test

## Idea

- Are my two populations really the same?
- Take two samples, compare them
- Difference between sample means should be 0(-ish)
- What counts as close enough to zero?
- Use t distribution to decide

# t test in JMP

## Problem

- Comparing wing spans of valley dragons and mountain dragons
- $H_0$ ?  $H_A$ ?
- One-tailed or two-tailed?
- What alpha?



## t test in JMP

- Select H0, HA, one-/two-tailed, alpha level
- Import tables, rename as appropriate
- Tables → Concatenate (select Create source column)
- Analyse → Fit X by Y
- t Test
- Try it yourself! (Field bunnies vs wood bunnies)

# Outline

- 1 Review, homework
- 2 Sampling distributions
- 3 t test
- 4 Wilcoxon's rank sum test**
- 5 Hypothesis testing considerations
- 6 Testing practise
- 7 Sample size and power
- 8 Correlations
- 9 Questions

# Wilcoxon's rank sum test

## Wilcoxon's rank sum test in JMP

- Your turn! (Field bunnies vs wood bunnies)

# Wilcoxon's rank sum test

## Wilcoxon's rank sum test in JMP

- Your turn! (Field bunnies vs wood bunnies)

## Steps

- Fit X by Y
- Select Wilcoxon

# Outline

- 1 Review, homework
- 2 Sampling distributions
- 3 t test
- 4 Wilcoxon's rank sum test
- 5 Hypothesis testing considerations**
- 6 Testing practise
- 7 Sample size and power
- 8 Correlations
- 9 Questions

# Order of a test

## Order of a test

- Select  $H_0$ ,  $H_A$
- Determine one-/two-tailed
- Determine alpha level
- (Ideally): Collect data
- Determine what test to use
- Analyse data
- Accept or reject hypothesis
- Report your p values!

# Choosing alpha

“But alpha is arbitrary!”

# Choosing alpha

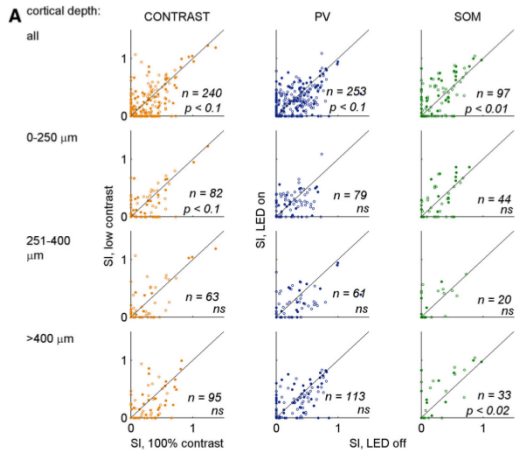
“But alpha is arbitrary!”

Discuss.



# Choosing alpha

Nienborg et al. • SOM or PV Neuron Role in Size Tuning in Mouse V1  
 • J. Neurosci., July 3, 2013 • 33(27):11145–11154



# Alternative Medicine and the Placebo effect

Schuessler cell salts uses: gull...

schuessler-cell-salts.com/healing/gullibility.htm

## Schuessler cell salts uses: gullibility

[Home](#)  
[List of cell salts](#)  
[Basic cell salts](#)  
[Additional cell salts](#)

[Diseases \(Uses\):](#)  
[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#)  
[N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [Z](#)

[Principle](#)  
[Potency](#)  
[Use](#)  
[Hot drink](#)  
[External use](#)  
[Facial diagnosis](#)  
[Biography](#)

[New](#)  
[Search](#)  
[Contact](#)

**Books**

### Cell salts internal

You can use the following cell salts for the treatment of gullibility :

- [Nr. 3. Iron phosphate](#)
- [Nr. 7. Magnesium phosphate](#)
- [Nr. 20. Alum](#)

Choose the right cell salt:  
 You can decide yourself wether you want to take all fitting cell salts together, only three salts at a time or only one salt at a time.

Use of the tablets:  
 3 until 6 times a day 1 - 3 tablets  
 High dosage: Every 1 to 10 minutes 1 tablet

Take the tablets one by one and let them dissolve in your mouth.

More informations about the use of the cell salts:

- [Use of the cell salts](#)

---

**Disclaimer**  
 Schuessler cell salts can't substitute the doctor. Call a doctor if you are seriously ill.

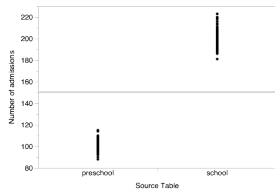
# Hospital study

## Study design

- A group of researchers wants to study the occurrence of a childhood infectious disease. They have permission to gather hospital admission data from all hospitals in a state and count the number of children admitted each month with the disease. The research question is whether the infection is likely to be spread in schools.
- They want to compare the number of admissions for the disease in preschool children (age 3-5) and school-age children (age 6-11). The method is to analyse all hospital records and count the number of children admitted in any given month over the period of the study. The hypothesis is that children in school are at a higher risk, because the school might be an environment where the disease can spread more easily.
- What is  $H_0$ ?

# Hospital study

Oneway Analysis of Number of admissions By Source Table

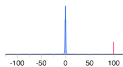


## t Test

school-preschool

Assuming unequal variances

Difference	99.542	t Ratio	118.0766
Std Err Dif	0.843	DF	200.8019
Upper CL Dif	101.204	Prob >  t	<.0001*
Lower CL Dif	97.879	Prob > t	<.0001*
Confidence	0.95	Prob < t	1.0000



## Time for a break?



**Sarah Hörst** @PlanetDr

14 Jan

We would've run rest of the experiments necessary to answer this question but the instrument belongs to lab next door

[#overlyhonestmethods](#)

Expand



**Elinor Amit** @AmitElinor

14 Jan

We had to cite those papers because the authors will probably be our reviewers. [#overlyhonestmethods](#)

Expand



**Paul Bretherton** @NeuroConsulting

14 Jan

Participants were undergrad psychology students because the general public are too expensive and ask too many questions

[#overlyhonestmethods](#)

Expand



**Liam Pomfret** @LiamPomfret

13 Jan

Thanks to paywalls, the only articles cited in this paper come from those I managed to find on the [#pdftribute](#) hashtag.

[#overlyhonestmethods](#)

Expand

# Outline

- 1 Review, homework
- 2 Sampling distributions
- 3 t test
- 4 Wilcoxon's rank sum test
- 5 Hypothesis testing considerations
- 6 Testing practise**
- 7 Sample size and power
- 8 Correlations
- 9 Questions

# Hypothesis testing practise

## Research question

- How has running speed changed over the years?
- Groups A: Use Chicago2013 and Chicago1996 datasets (randomly drawn runners)
- Groups B: Use Chicago13\_Top10 and Chicago96\_Top10

## How to proceed

- Select  $H_0$ ,  $H_A$
- One- or two-tailed?
- Alpha?
- Conclusion?

# Outline

- 1 Review, homework
- 2 Sampling distributions
- 3 t test
- 4 Wilcoxon's rank sum test
- 5 Hypothesis testing considerations
- 6 Testing practise
- 7 Sample size and power**
- 8 Correlations
- 9 Questions



# Sample size and power in JMP

- DOE (Design of Experiments) tab

# Outline

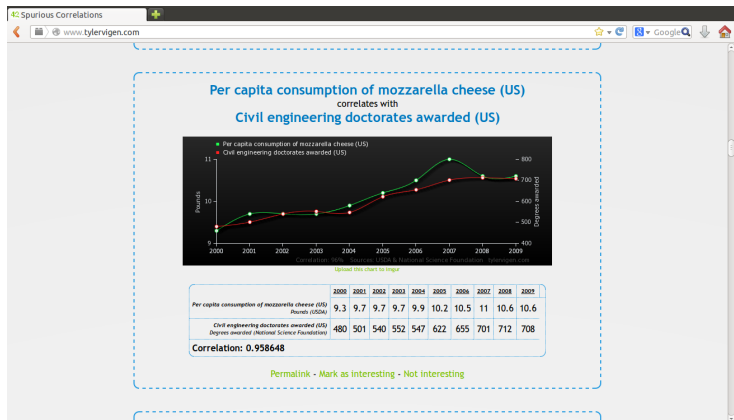
- 1 Review, homework
- 2 Sampling distributions
- 3 t test
- 4 Wilcoxon's rank sum test
- 5 Hypothesis testing considerations
- 6 Testing practise
- 7 Sample size and power
- 8 Correlations**
- 9 Questions

# Correlations in JMP

## Correlations exercise

- Open file correlation.xls
- Find a way to correlate the variables “Pound” and “Number” in JMP
- What sort of data might this be?

## Correlations in JMP



<http://www.tylervigen.com/>

# Outline

- 1 Review, homework
- 2 Sampling distributions
- 3 t test
- 4 Wilcoxon's rank sum test
- 5 Hypothesis testing considerations
- 6 Testing practise
- 7 Sample size and power
- 8 Correlations
- 9 Questions