Data: What do I do with it?

University of Edinburgh

Melanie I Stefan - mstefan@exseed.ed.ac.uk
Based on lecture slides by Iris Oren

December 2016

Getting help

- BMS Stats user forum
- Handbook of Biological Statistics
- Quabinet Wiki
- Lecture slides will be available on `http://melaniestefan.net/teaching.html`

Learning objectives

Think

- Know how a hpyothesis tests works
- Know examples of hypothesis tests and the ideas behind them
- Critically evaluate approaches to summarising, visualising, interpreting and sharing data

Learning objectives

Think

- Know how a hpyothesis tests works
- Know examples of hypothesis tests and the ideas behind them
- Critically evaluate approaches to summarising, visualising, interpreting and sharing data

Do

- Summarise and share your data
- Run a common hypothesis test
- Get help

Learning objectives

Think

- Know how a hpyothesis tests works
- Know examples of hypothesis tests and the ideas behind them
- Critically evaluate approaches to summarising, visualising, interpreting and sharing data

Do

- Summarise and share your data
- Run a common hypothesis test
- Get help

Feel

- Appreciate the importance of good statistical practice
- Gain confidence in handling datasets

The scientific process

The scientific process

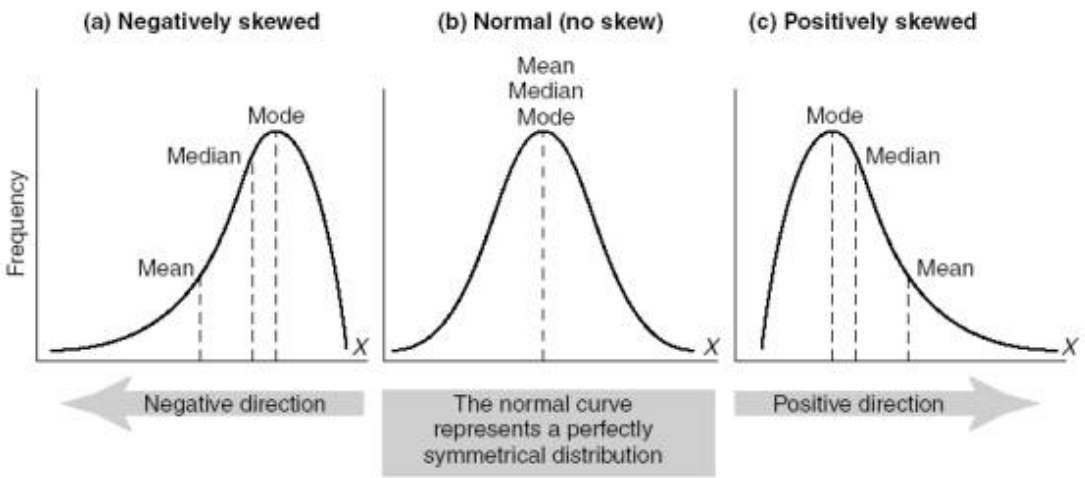
- Question
- Hypothesis
- Experimental design
- Data collection
- Data processing
- Data exploration
- Hypothesis testing
- Interpretation
- Presentation
- Publication

Outline

Data exploration: summary statistics

Describing your data: Measure of central tendency

- Mean: average (population mean, μ; sample mean, $\bar{x}$)
- Median: halfway of ranked data. Use to describe skewed distributions
- Mode: most common value(s). Use for multimodal data

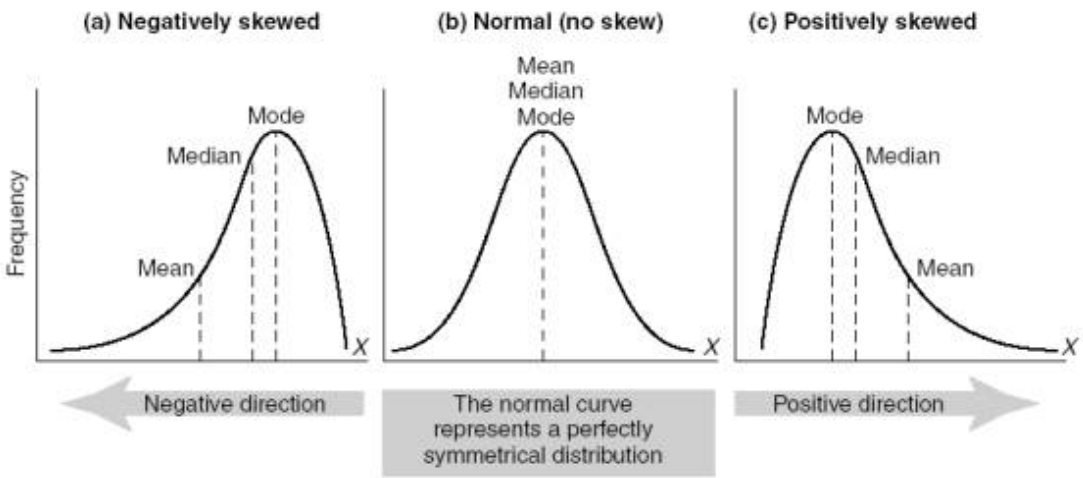


http://www.southalabama.edu/coe/bset/johnson/lectures/lec15.htm

Data exploration: summary statistics

Describing your data: Measure of central tendency

- Mean: average (population mean, μ; sample mean, $\bar{x}$)
- Median: halfway of ranked data. Use to describe skewed distributions
- Mode: most common value(s). Use for multimodal data



http://www.southalabama.edu/coe/bset/johnson/lectures/lec15.htm

Data exploration: summary statistics

Describing your data: Measure of variability

- **Range**
  - Smallest to largest values
  - Sensitive to extreme values

Data exploration: summary statistics

Describing your data: Measure of variability for median - IQR

Firing rate X- (Hz)
22.28
21
18.97
15.06
18.25
21.93

Sort

Firing rate X- (Hz)
15.06
18.25
18.97
21
21.93
22.28

Q1: 25% of data below

Median Q2=19.985

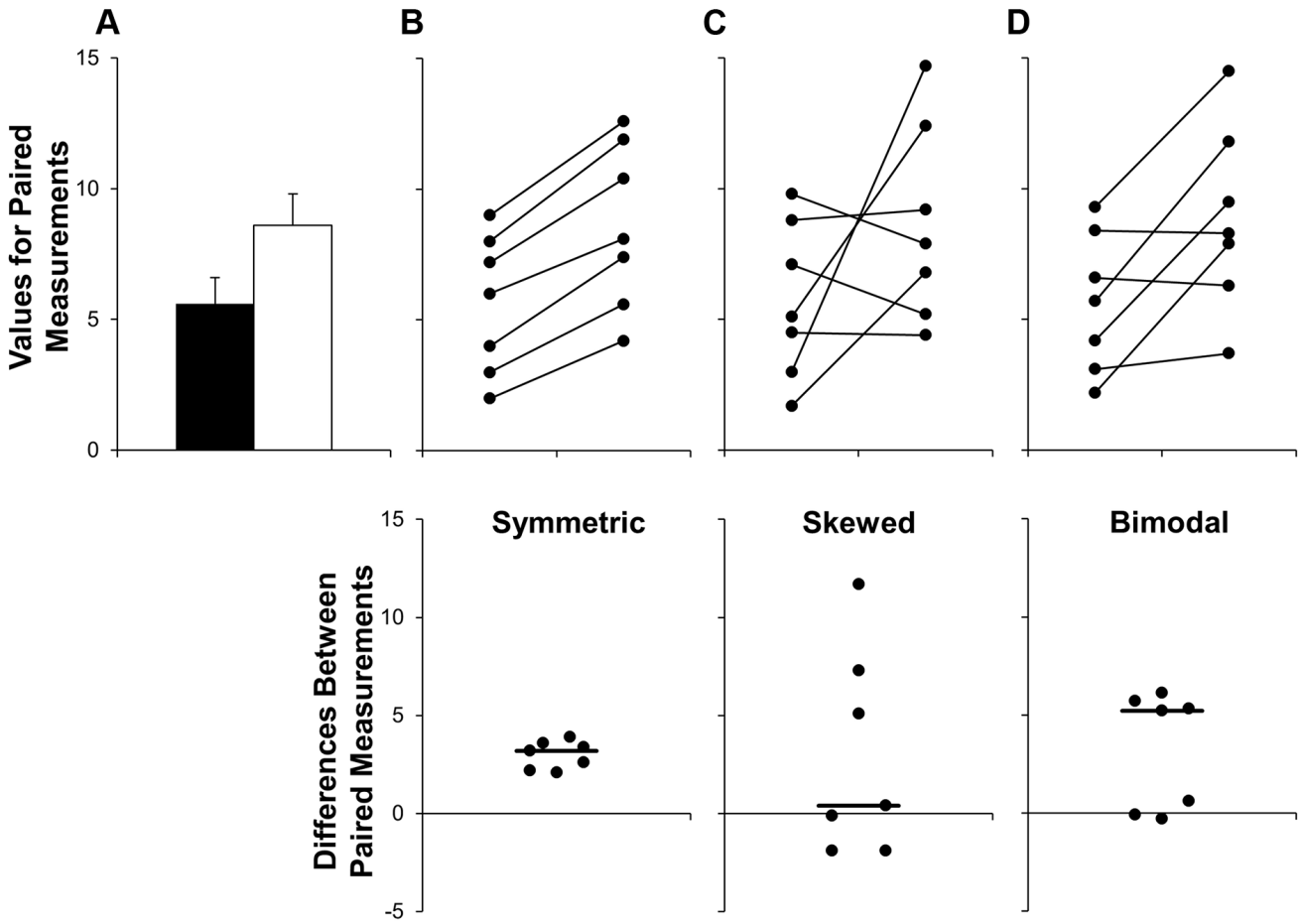
Q3: 25% of data above

Interquartile range (IQR): Q3-Q1

- Should be quoted with median (median is Q2)
- Independent of distribution
- Not necessarily symmetrical about median

Outline

How to represent data?

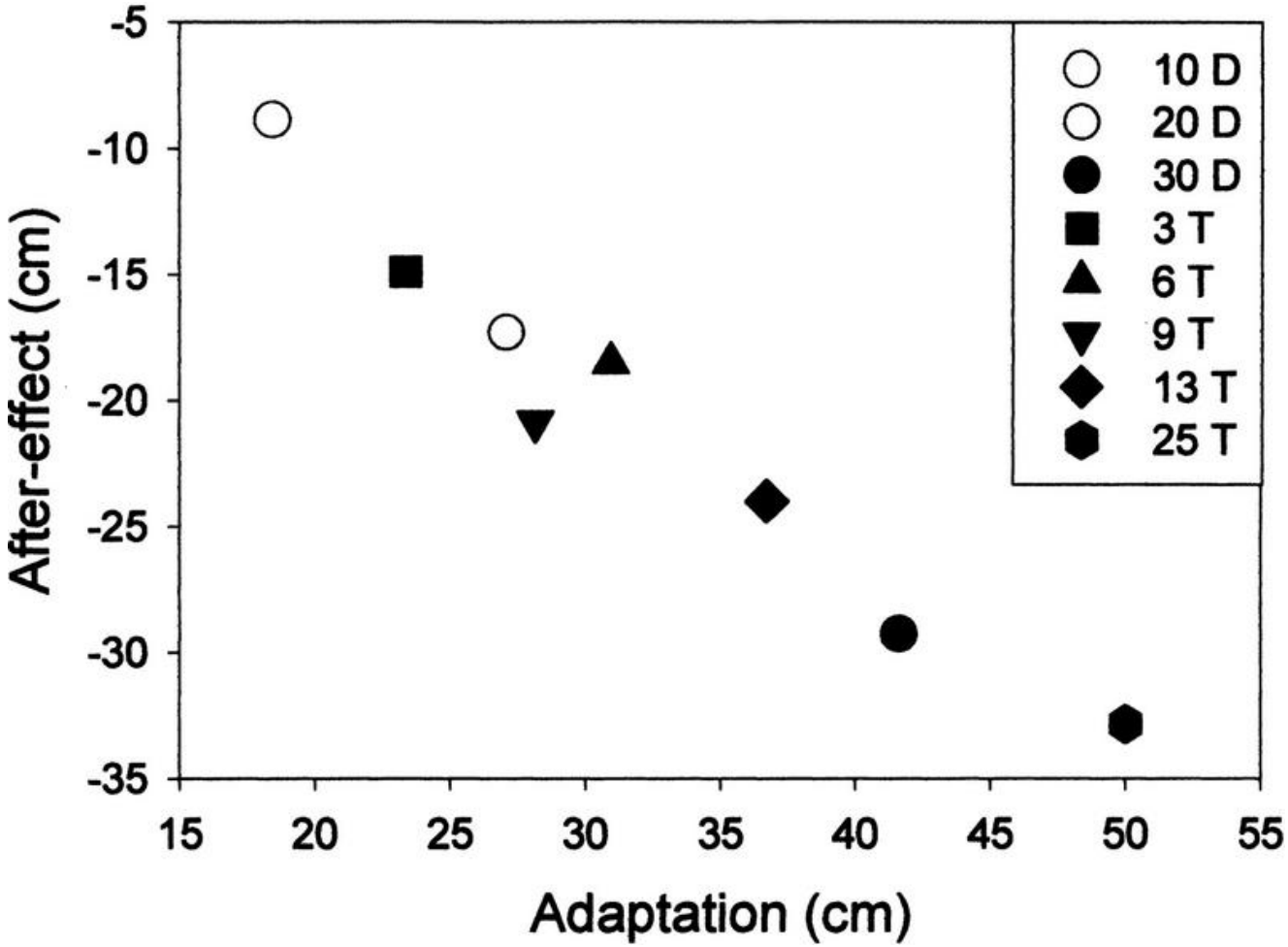


Weissgerber, 2015

Discussing graphics choices

Discussing graphics choices

- What do the axes show? What are the units? Where is the origin?
- How are colours and patterns used? For instance, what is the experimental condition, what is the control?
- What do symbols on the plot represent? If they represent aggregate data (e.g. mean, median, standard error), how many data points are represented by one symbol in the plot?
- What is the "message" you get from the plot?
- If the data is represented in a novel or unusual way, do you understand what it is you are looking at?
- Is there something about the representation you particularly like/dislike? Is there something that could be improved?



Fernandez-Ruiz, 1999

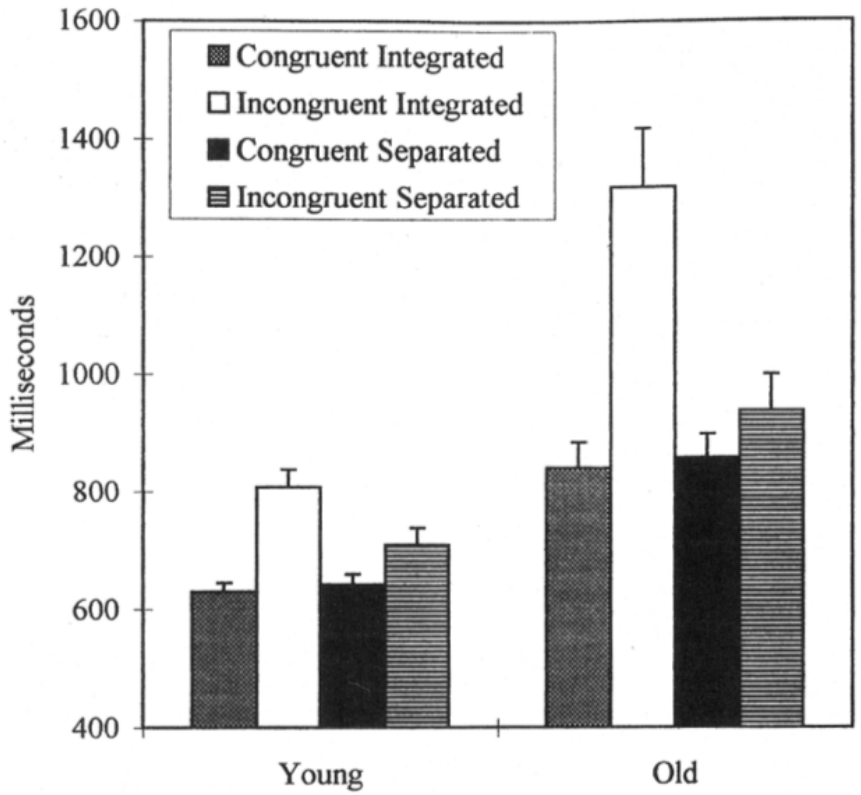
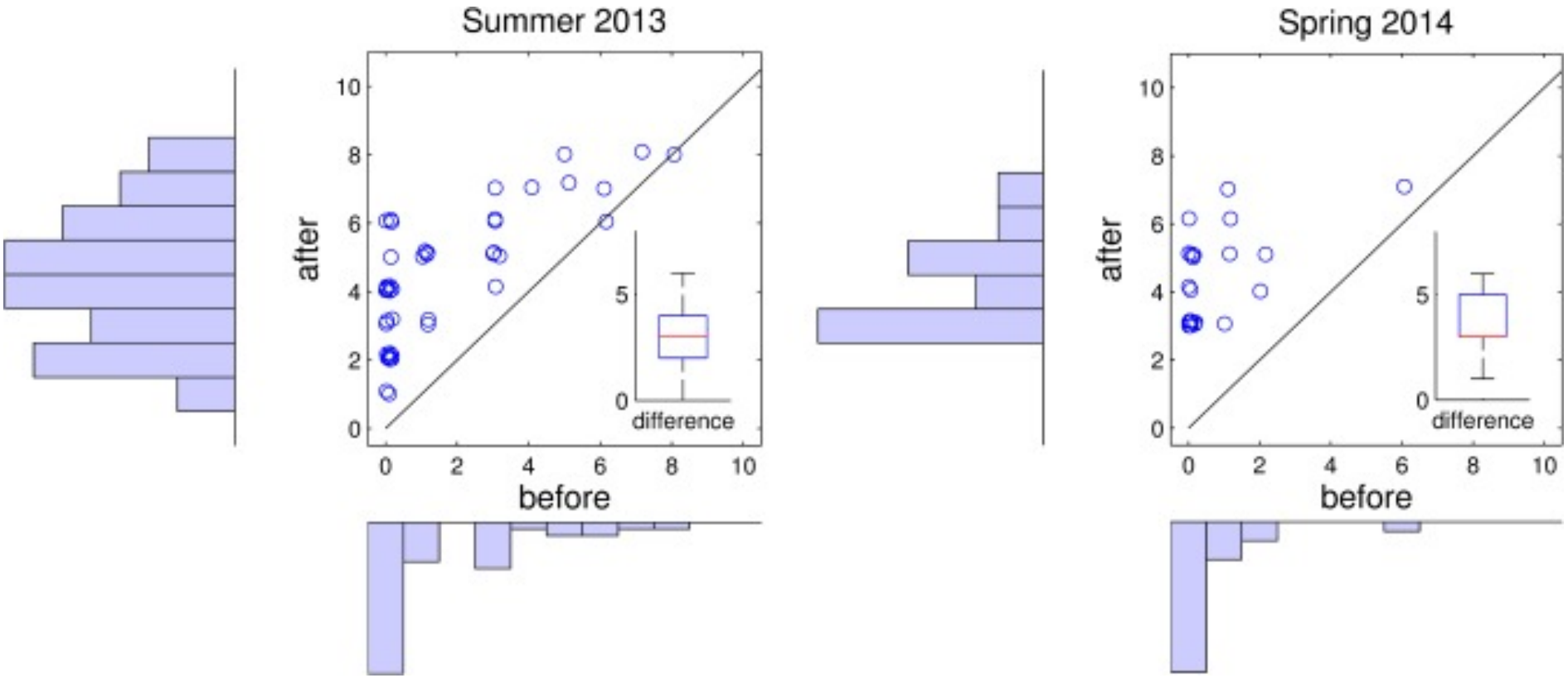
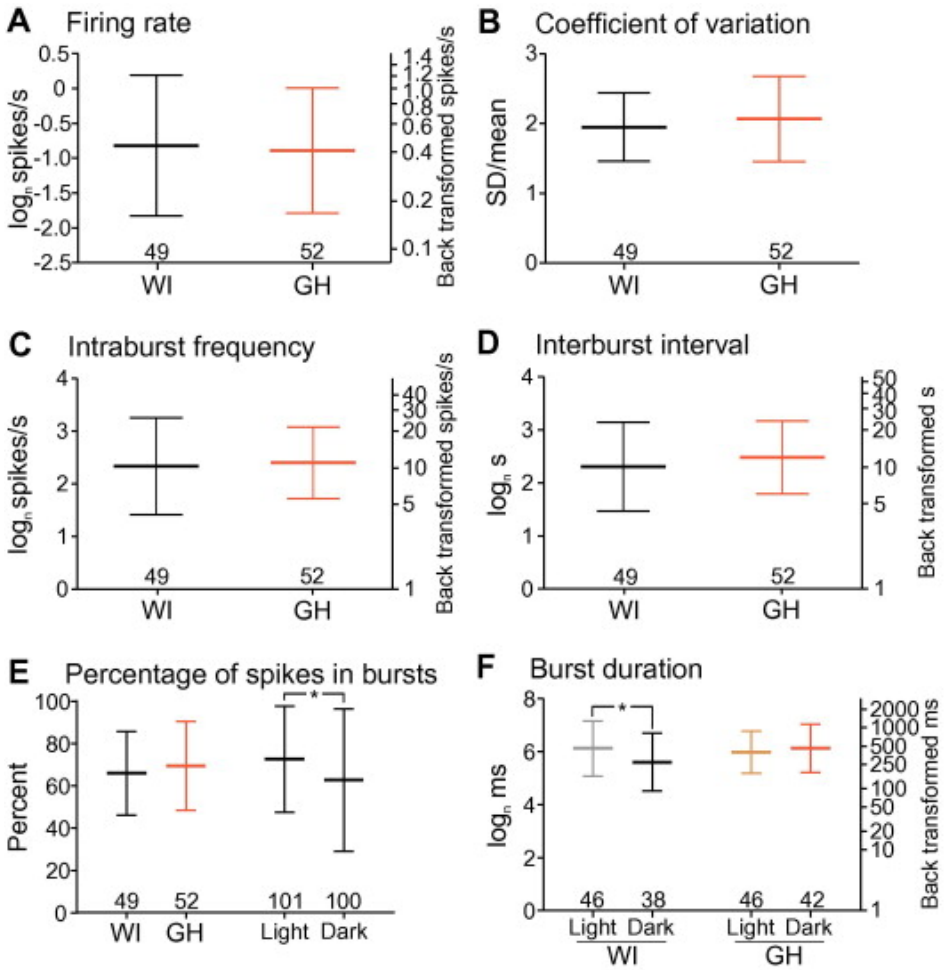


Figure 1. **Median response latencies in the Stroop task for young and old adults as a function of stimulus congruency and color–word integration. Error bars represent one standard error.**

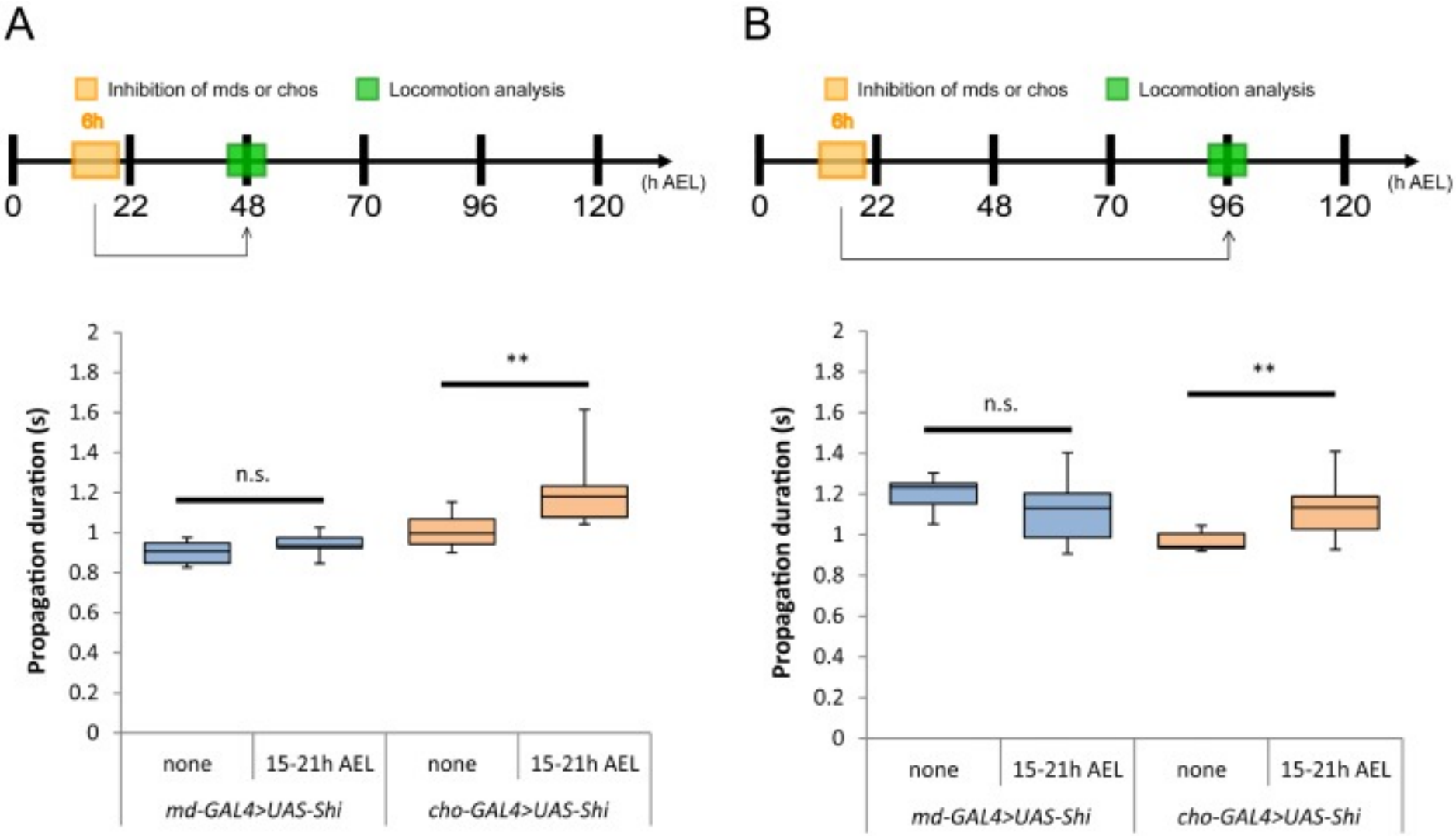
West, 1997


Summer 2013
after
before
difference
Spring 2014
after
before
difference

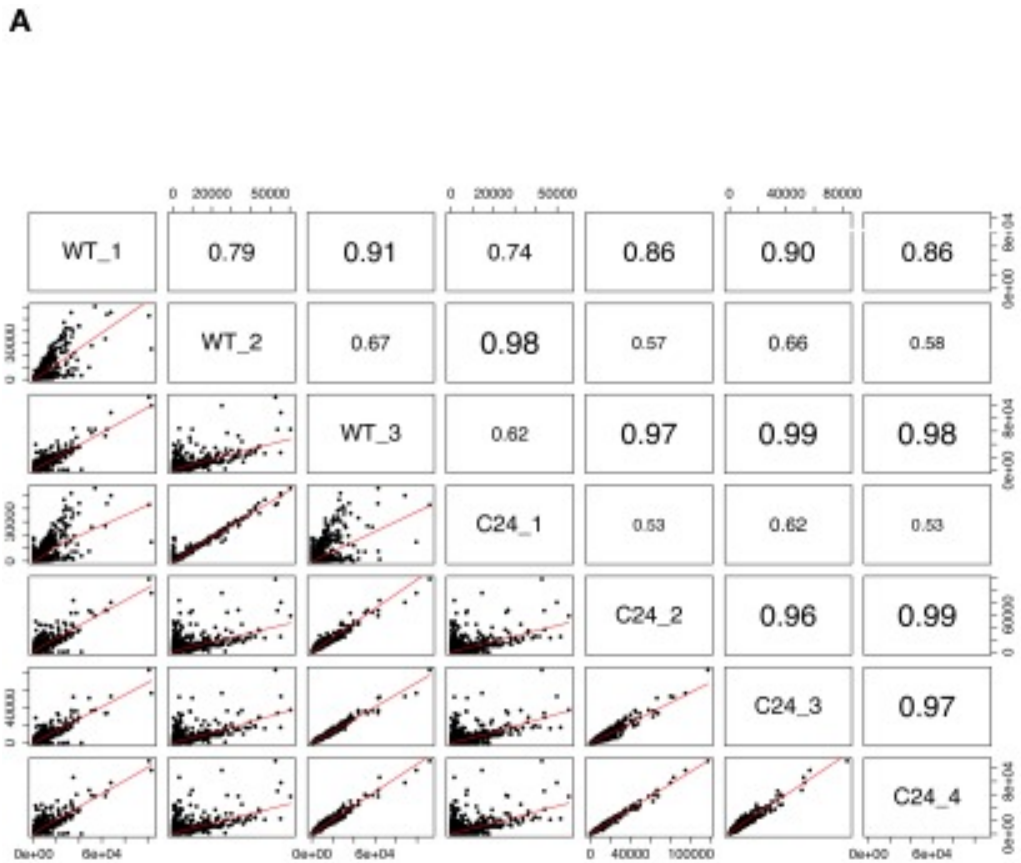
Stefan, 2015

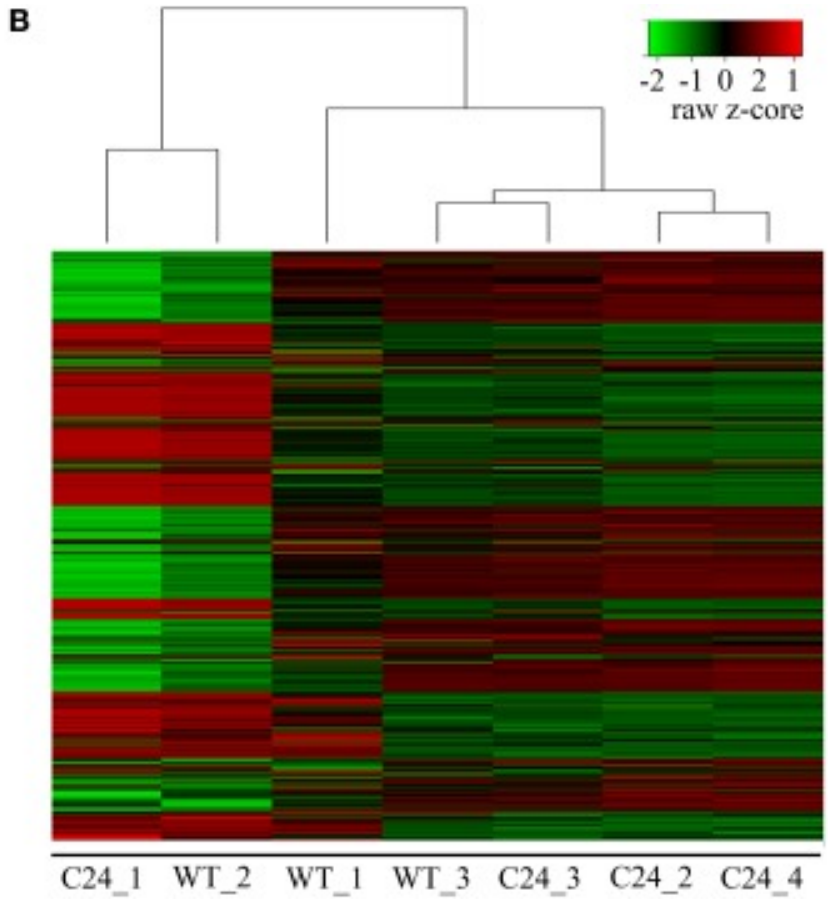

A
Firing rate
log$_n$ spikes/s
Back transformed spikes/s
49
52
WI
GH
B
Coefficient of variation
SD/mean
49
52
WI
GH
C
Intraburst frequency
log$_n$ spikes/s
Back transformed spikes/s
49
52
WI
GH
D
Interburst interval
log$_n$ s
Back transformed s
49
52
WI
GH
E
Percentage of spikes in bursts
Percent
49
52
101
100
WI
GH
Light
Dark
F
Burst duration
log$_n$ ms
Back transformed ms
46
38
46
42
Light
Dark
Light
Dark
WI
GH

Perk, 2015

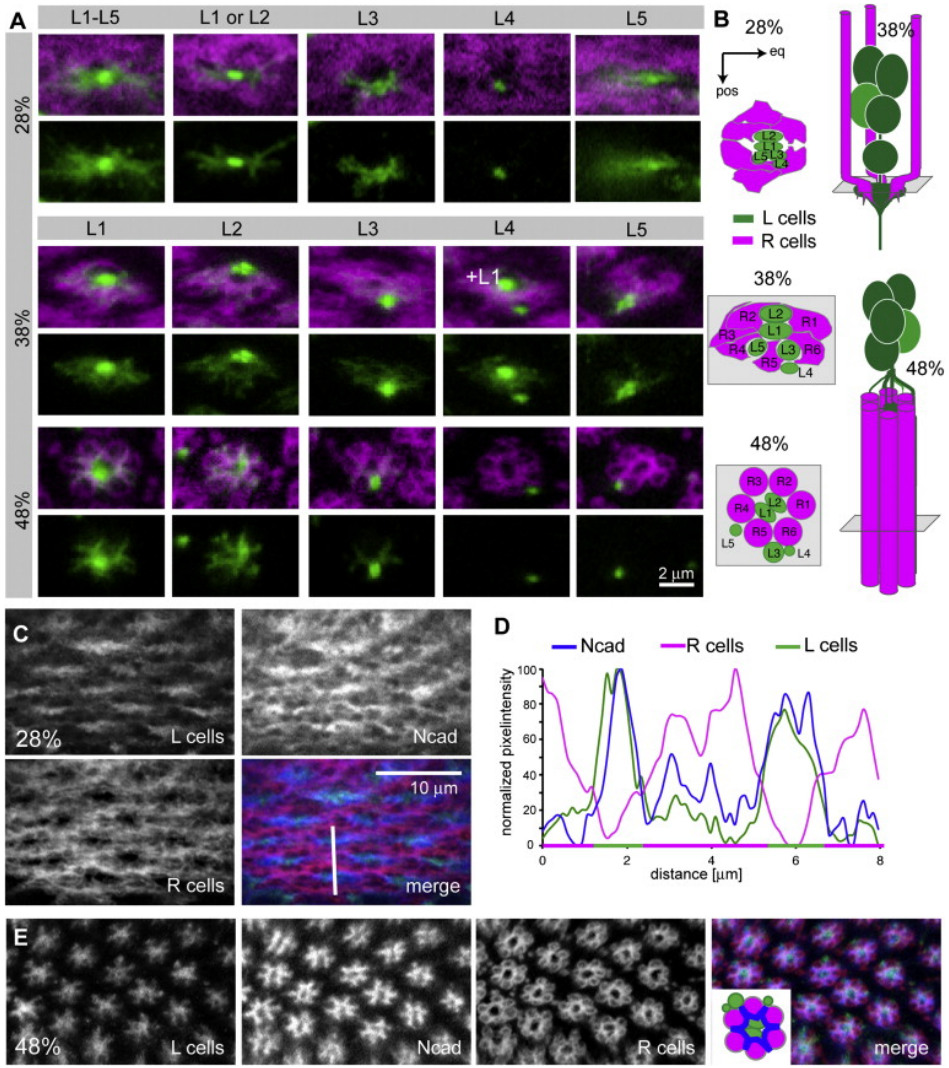


Fushiki, 2013


A
WT_1
0.79
0.91
0.74
0.86
0.90
0.86
WT_2
0.67
0.98
0.57
0.66
0.58
WT_3
0.62
0.97
0.99
0.98
C24_1
0.53
0.62
0.53
C24_2
0.96
0.99
C24_3
0.97
C24_4


B
-2 -1 0 2 1
raw z-core
C24_1
WT_2
WT_1
WT_3
C24_3
C24_2
C24_4

Amaral, 2014



Schwabe, 2014

Outline

All hypothesis tests are the same

1. Formulate the Null Hypothesis, alternative hypothesis
2. Design your experiment and collect data
3. Think about what you would expect if H0 was true.
4. Could your data be explained by the Null Hypothesis?
5. Determine the probability of your data given H0
6. Interpret your p value and make a decision
7. Be aware that hypothesis tests are not perfect.

All hypothesis tests are the same - example

Example

Example: Is there a difference in the incidence of knee injuries between men and women?

All hypothesis tests are the same - example

Example

Example: Is there a difference in the incidence of knee injuries between men and women?

1. Formulate the Null Hypothesis, alternative hypothesis

All hypothesis tests are the same - example

Example

Example: Is there a difference in the incidence of knee injuries between men and women?

1. Formulate the Null Hypothesis, alternative hypothesis
2. Design your experiment and collect data

All hypothesis tests are the same - example

Example

Example: Is there a difference in the incidence of knee injuries between men and women?

1. Formulate the Null Hypothesis, alternative hypothesis
2. Design your experiment and collect data
3. Think about what you would expect if H0 was true.

All hypothesis tests are the same - example

Example

Example: Is there a difference in the incidence of knee injuries between men and women?

1. Formulate the Null Hypothesis, alternative hypothesis
2. Design your experiment and collect data
3. Think about what you would expect if H0 was true.
4. Could your data be explained by the Null Hypothesis?

All hypothesis tests are the same - example

Example

Example: Is there a difference in the incidence of knee injuries between men and women?

1. Formulate the Null Hypothesis, alternative hypothesis
2. Design your experiment and collect data
3. Think about what you would expect if H0 was true.
4. Could your data be explained by the Null Hypothesis?
5. Determine the probability of your data given H0

All hypothesis tests are the same - example

Example

Example: Is there a difference in the incidence of knee injuries between men and women?

1. Formulate the Null Hypothesis, alternative hypothesis
2. Design your experiment and collect data
3. Think about what you would expect if H0 was true.
4. Could your data be explained by the Null Hypothesis?
5. Determine the probability of your data given H0
6. Interpret your p value and make a decision

All hypothesis tests are the same - example

Example

Example: Is there a difference in the incidence of knee injuries between men and women?

1. Formulate the Null Hypothesis, alternative hypothesis
2. Design your experiment and collect data
3. Think about what you would expect if H0 was true.
4. Could your data be explained by the Null Hypothesis?
5. Determine the probability of your data given H0
6. Interpret your p value and make a decision
7. Be aware that hypothesis tests are not perfect.

Outline

There are different hypothesis tests!

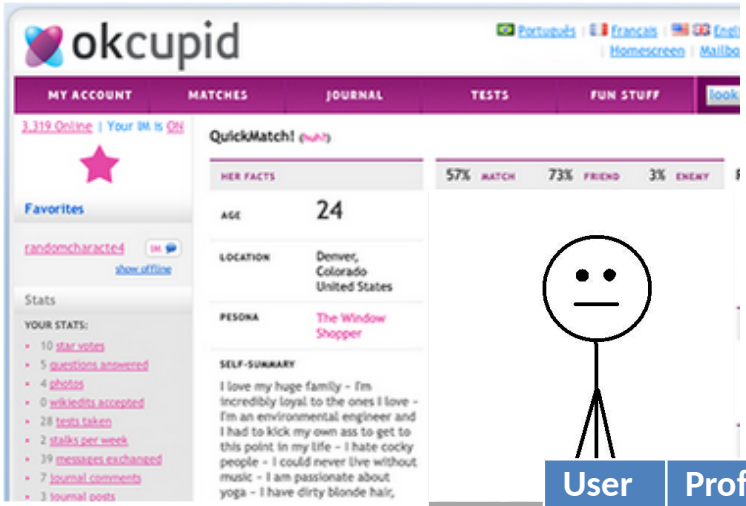
Why are there different hypothesis tests? And which one should you use?
It depends on . . .

There are different hypothesis tests!

Why are there different hypothesis tests? And which one should you use? It depends on . . .

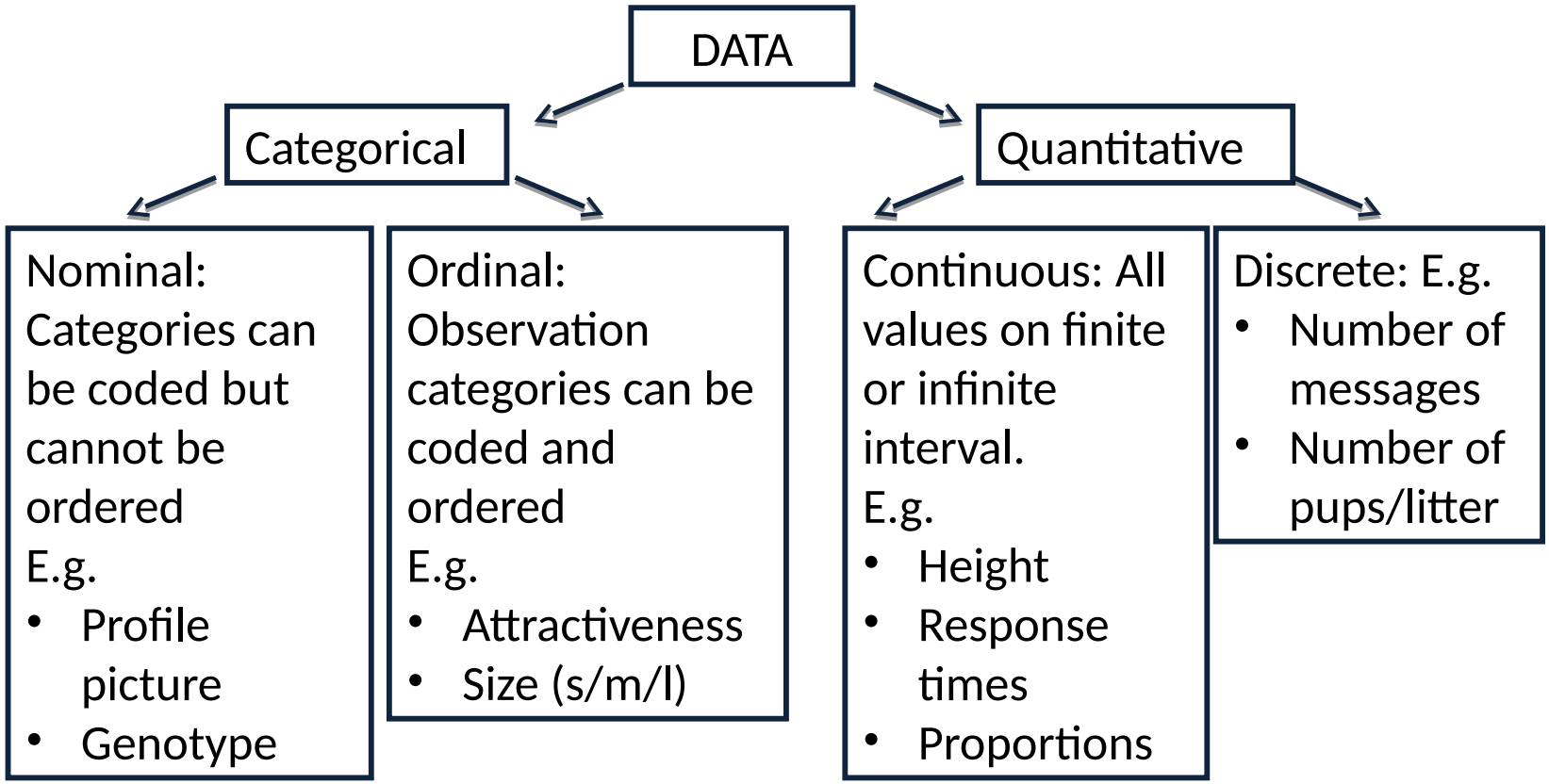
- The type of your data
- The question you ask
- Your sample size and distribution
- Other properties of your dataset
- Other factors that shouldn't matter (computer, skill, supervisor, . . .)

Types of variables



User	Profile picture	Attractiveness	Height (m)	Messages received/ week
0001	Outdoors	Med	1.60	10
0002	Selfy	Low	1.58	11
0003	With friends	High	1.69	26
.....				
0050	Outdoors	Med	1.79	15

Types of variables



- Independent variable/Explanatory variable
- Dependent variable/Response variable

There are different hypothesis tests

Choose your own adventure!

- Compare a mean to a proposed population mean ▸ Go
- Compare two population means ▸ Go
- Compare more than two population means ▸ Go
- Compare proportions of nominal variables ▸ Go
- Relationships between variables ▸ Go
- . . .

If in doubt, check this website

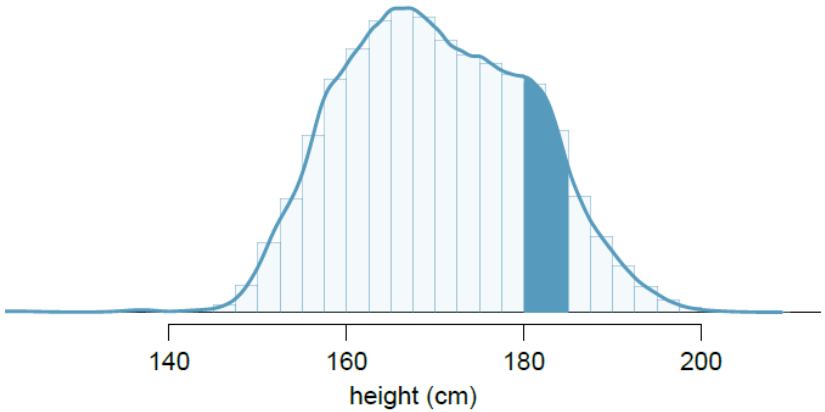
Done? ▸ Sample size and experimental design

Compare a mean to a proposed population mean

- In theory: **z-test**
- In real life: one-sample t-test
- If you can't do a t-test: permutation test

From histograms to probabilities

Height of 3 million randomly sampled Americans



Normalised smooth curve = probability density function

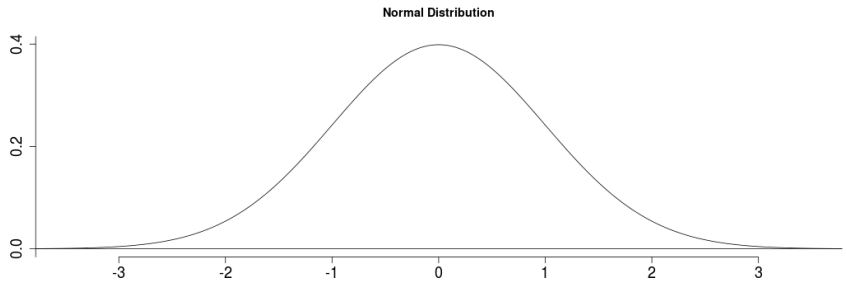
Total area of *pdf* = 1

Area under *pdf* gives the probability

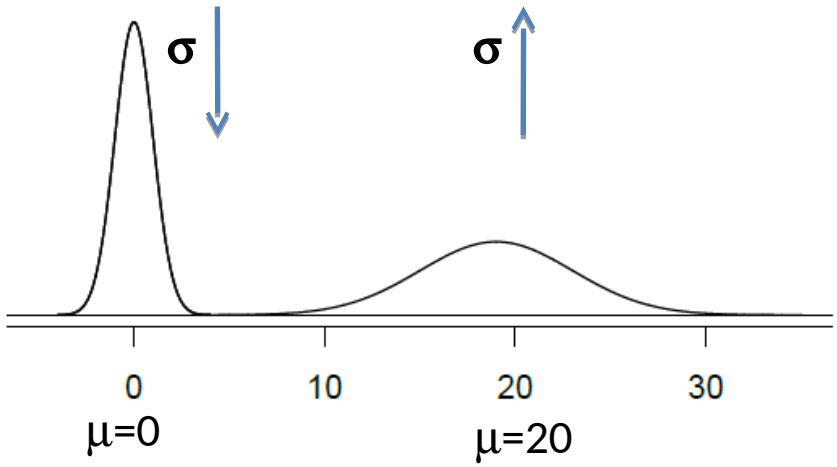
Probability that randomly sampled American is between 180 – 185 cm = 11.6%

Probability that randomly sampled American is between 130 – 210 cm ≈ 100%

The normal distribution



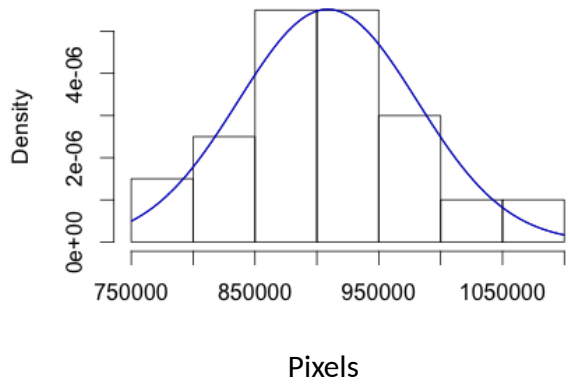
- Symmetrical
- Described by **mean (μ)** and **standard deviation (σ)**
- Area under curve = 1
- Distribution of population NOT sample



The normal distribution

Willerman et al. (1991) measured brain volume(pixel count) of 40 college students by MRI

Subject	Pixels
01	816932
02	1001121
03	1038437
04	965353
05	951545
06	928799
07	991305
...	...
39	930016
40	935863



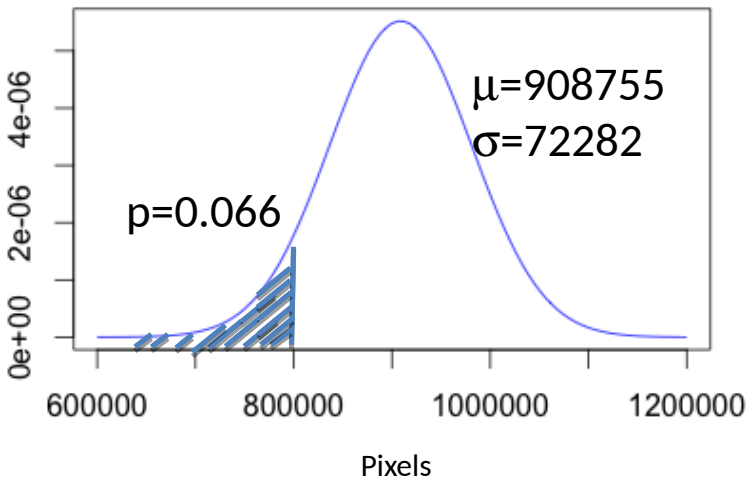
Mean=908755

Standard deviation=72282

Let's assume that these are the true population mean (μ) and standard deviation (σ)

The normal distribution

Assume that brain volumes of Edinburgh students follows the same distribution, what is the probability that your brain volume is less than 800000 pixels?



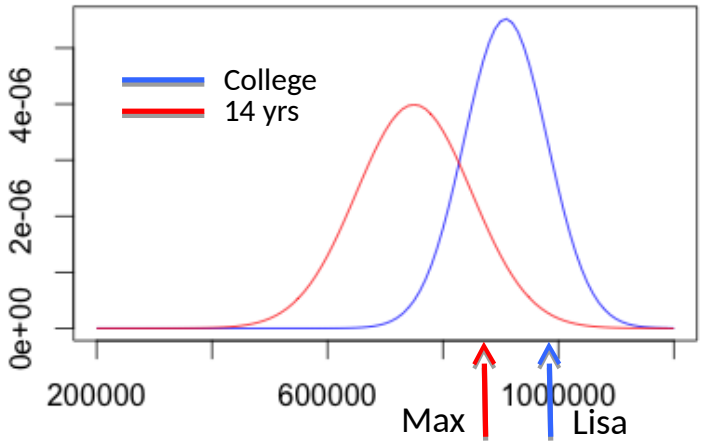
http://homepage.stat.uiowa.edu/~mbognar/applets/normal.html

There is a 6.6% probability that your brain volume is less than 800000 pixels

The normal distribution

Normalising the normal distribution: The z-statistic

The brains of college students are normally distributed with mean 908755 and standard deviation 72282
The brain volumes of 14 year old high school students are normally distributed with mean 750000 and standard deviation 100000



Max (14 yrs) has brain volume 850000 pixels. Lisa (21 yrs, college student) has brain volume 990500. Who has the bigger brain relative to their age?

The **Z score** = number of standard deviations an observation falls above or below the mean.

$$z = \frac{x - \mu}{\sigma}$$

$$z_{Max} = \frac{850000 - 750000}{100000}$$

$$= 1$$

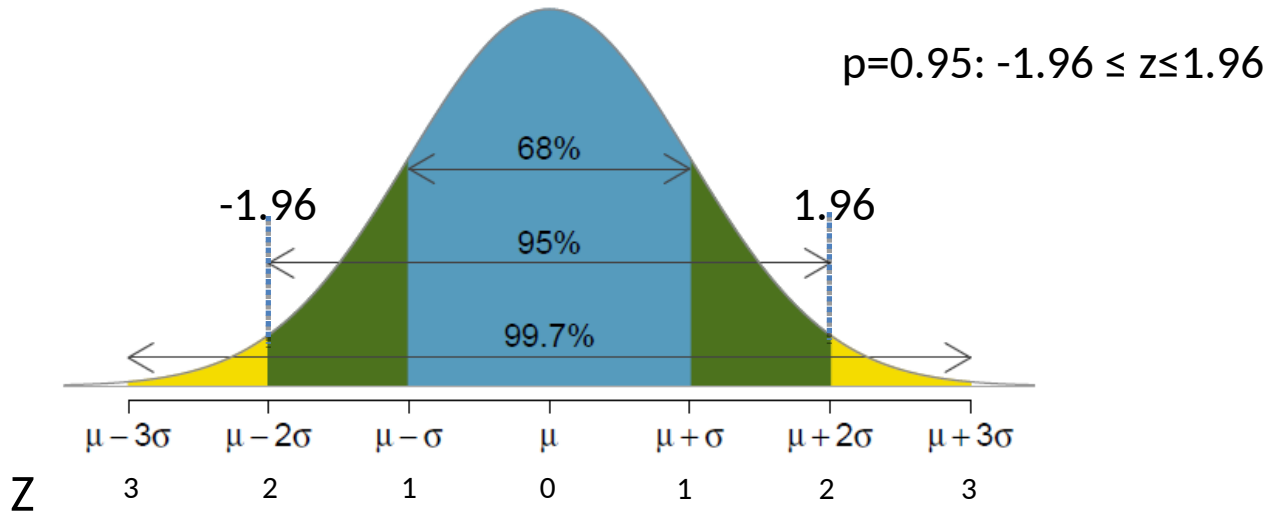
$$z_{Lisa} = \frac{990500 - 908755}{72282.05}$$

$$= 1.13$$

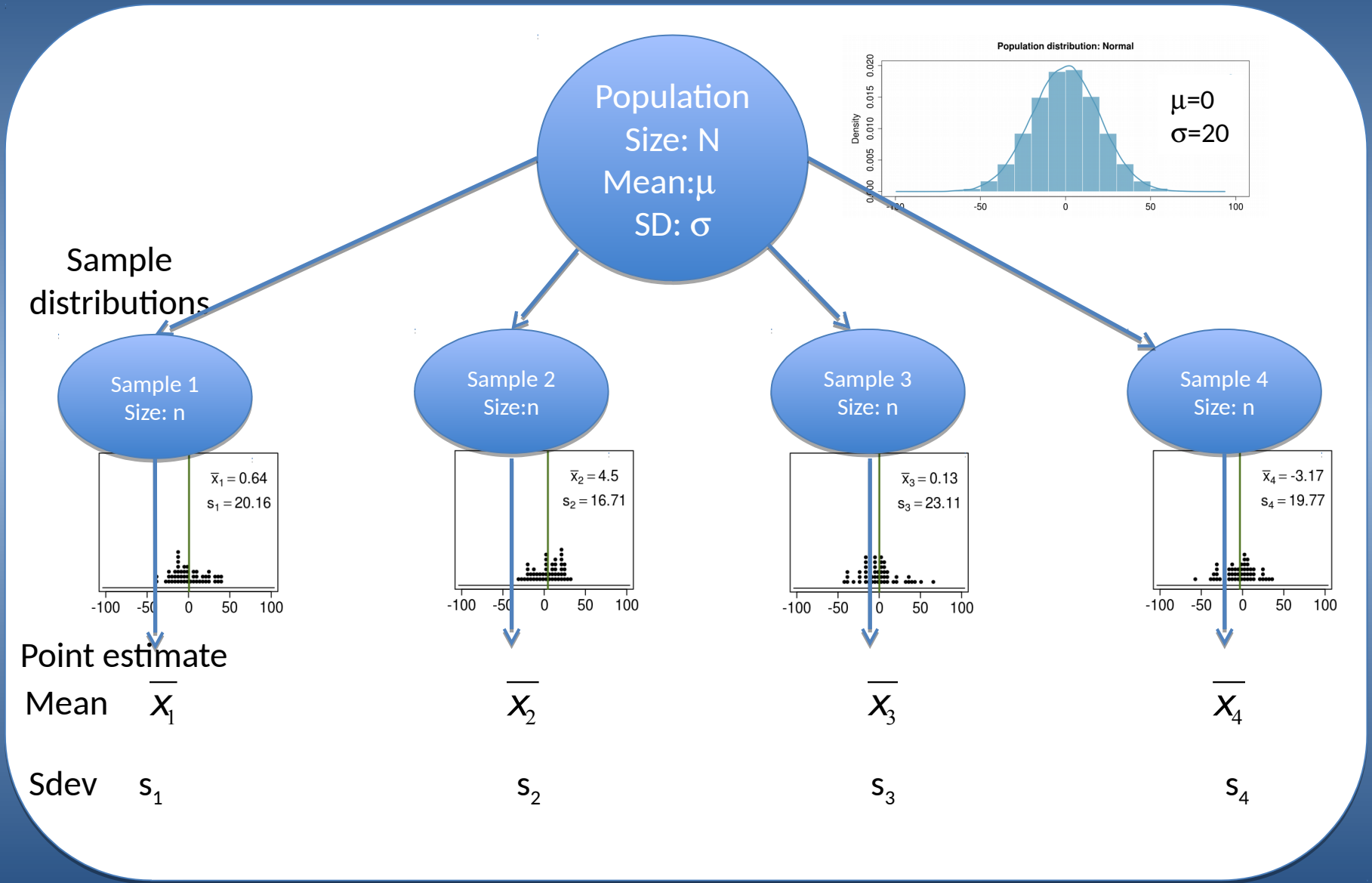
Larger brain

The normal distribution

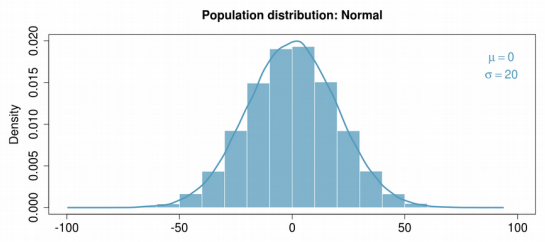
Probabilities of normal distribution



Sample statistics

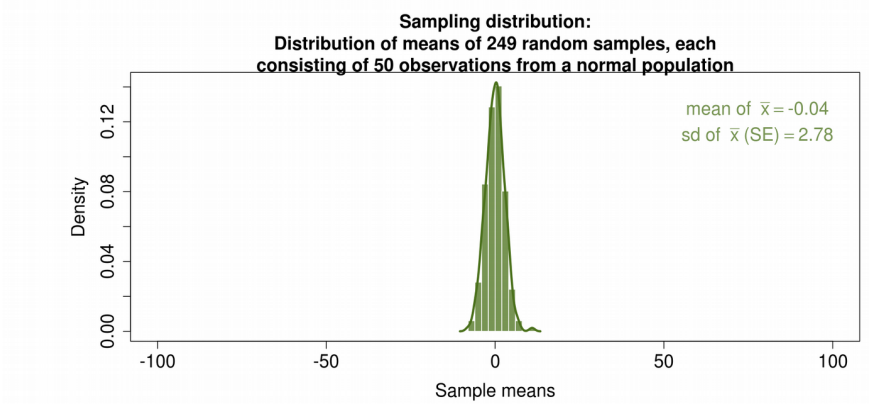


Sample statistics & CLT



$\overline{x}_1$ $\overline{x}_2$ $\overline{x}_3$ $\overline{x}_4$

Sampling distribution



Mean $(\overline{X}) \approx \mu$

S. dev. $(\overline{X}) < \sigma$

http://onlinestatbook.com/stat_sim/sampling_dist/

S. dev. $(\overline{X})$ = **Standard error of the mean (SEM)**

$$= \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$

↑n ↓ SEM

Sample statistics & CLT

Central Limit Theorem:

- The distribution of sample means (the sampling distribution) is nearly normal
- The mean of the sampling distribution is approximately equal to the population mean (μ)
- The standard error (the standard deviation of sample means) $= \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$
- SEM estimates variability in sample estimates of mean

Conditions:

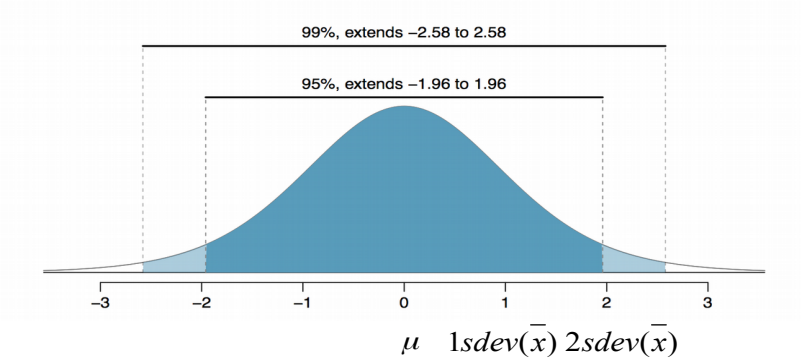
- Observations are independent
- Sample size is sufficiently large (>30)
- Distribution is not strongly skewed (but moderate skew allowed if large sample)

CLT also applies to other point estimates (variance, standard deviation)

Confidence interval

We can use the standard error of the mean to **define a confidence interval** for our estimate of the mean

Example
The mean brain volume for the 40 sampled college students was 908755. How confident can we be that our estimate represents the mean brain volume for the population of college students?



Since sampling distribution approx normal, 95% of sample means will be within ~1.96 x sdev($\bar{x}$) around the mean of the sampling distribution. More generally,

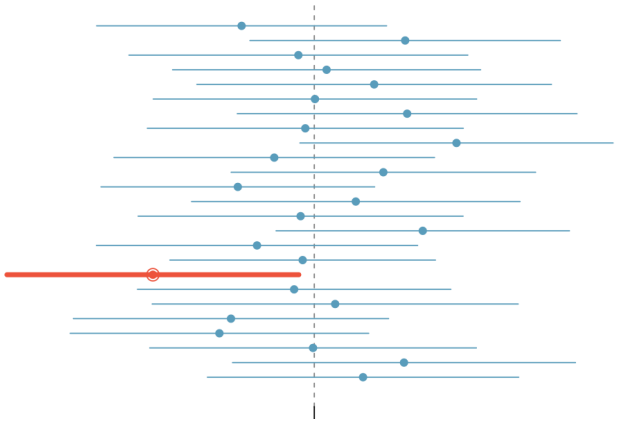
$$CI = \bar{x} \pm z^* \times SEM$$

Where the value of z^* depends on the confidence level (1.96 for 95%)

Confidence interval

$$CI = \bar{x} \pm (1.96 \times s/\sqrt{n})$$
$$= [886355, 931155]$$

You are 95% confident that the true population mean is between 886355 and 931155 pixels



True population mean
(unknown)

~95% of samples will contain the true population mean

The normal distribution & CLT

- Normal distribution is symmetric about the mean
- Area under the curve gives probability
- 68, 95, 99.7% rule
- Normalise standard distribution with $z = \frac{x - \mu}{\sigma}$

- Means of random samples from normal distribution are normally distributed around population mean with standard deviation $\frac{\sigma}{\sqrt{n}}$

- Confidence intervals tell you the probability with which you will find the **true population mean** in an interval

Assumptions for z test

When am I allowed to do a z test?

Assumptions for z test

When am I allowed to do a z test?

- The parameter you are testing should be normally distributed.
- It should be possible to estimate the standard deviation accurately.

Assumptions for z test

When am I allowed to do a z test?

- The parameter you are testing should be normally distributed.
- It should be possible to estimate the standard deviation accurately. (This means the sample should be "large enough")

Assumptions for z test

When am I allowed to do a z test?

- The parameter you are testing should be normally distributed.
- It should be possible to estimate the standard deviation accurately. (This means the sample should be "large enough")

What if those assumptions are not met?

Assumptions for z test

When am I allowed to do a z test?

- The parameter you are testing should be normally distributed.
- It should be possible to estimate the standard deviation accurately. (This means the sample should be "large enough")

What if those assumptions are not met?
For smaller samples: Do a t test.

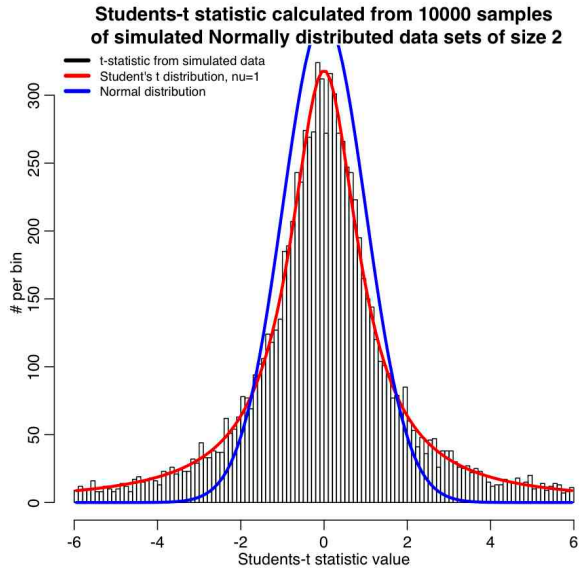
The t-distribution

Recall…

$$z_{Mean} = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$ but we don't generally know σ

Instead use the **student's t** statistic

$$t = \frac{\bar{x} - \mu}{SEM}$$



- *t*-distribution has fatter tail than *z*-distribution to make up for uncertainty in σ
- Shape of *t*-distribution depends on n. This defines **"degrees of freedom"** (df).

Using the t statistic

Recall: Willerman et al. (1991) measured brain volume (pixel count) of 40 college students by MRI
We had:
Mean=908 755 Standard deviation=72 282
Should have used t-statistic (instead of z-statistic).
For $n = 40$, $t = 2.02$

$$\text{CI} = \bar{x} \pm t \times SEM$$

$$\text{CI} = 908\,755 \pm \left(2.02 \times \frac{72282}{40}\right)$$

$$\text{CI} = [905\,104, 912\,405]$$

Back to overview

Compare two population means

Example: Are men and women of different height?

Compare two population means

Example: Are men and women of different height?

- Most often used: 2-sample t-test (same idea as 1-sample t-test, except you are comparing difference between means to zero)
- If conditions for t-test are not met: **Wilcoxon rank sum test** (Mann Whitney U test)
- You can always do: simulation-based test

Wilcoxon rank sum test

Idea: Look not at absolute values, but rank of data. Ask the question whether the ranking you see could come about under H_0

Wilcoxon rank sum test

Idea: Look not at absolute values, but rank of data. Ask the question whether the ranking you see could come about under H_0

Example: Are men and women *in graduate school* of different height?

- Collect data in both group.
- Rank *all* data.
- Compute sum of ranks in each group.
- Look smaller one up in a table.

Wilcoxon rank sum test

Table Critical values of the smallest rank sum for the Wilcoxon-Mann-Whitney test

n_1 = number of elements in the largest sample;

n_2 = number of elements in the smallest sample.

Level of significance α						Level of significance α					
Two-sided		0.20	0.10	0.05	0.01	Two-sided		0.20	0.10	0.05	0.01
One-sided		0.10	0.05	0.025	0.005	One-sided		0.10	0.05	0.025	0.005
n_1	n_2					n_1	n_2				
3	2	3	-	-	-	10	6	38	35	32	27
3	3	7	6	-	-	10	7	49	45	42	37
4	2	3	-	-	-	10	8	60	56	53	47
4	3	7	6	-	-	10	9	73	69	65	58
4	4	13	11	10	-	10	10	87	82	78	71
5	2	4	3	-	-	11	1	1	-	-	-
5	3	8	7	6	-	11	2	6	4	3	-
5	4	14	12	11	-	11	3	13	11	9	6
5	5	20	19	17	15	11	4	21	18	16	12

From `http://www.watpon.com/table/wilcoxonmannwhitney.pdf`

Wilcoxon rank sum test

Discussion

- What are the advantages/disadvantages of using this (rather than a t-test)?
- When is each of those tests appropriate?
- Can I be really cheeky and just do both?

Back to overview

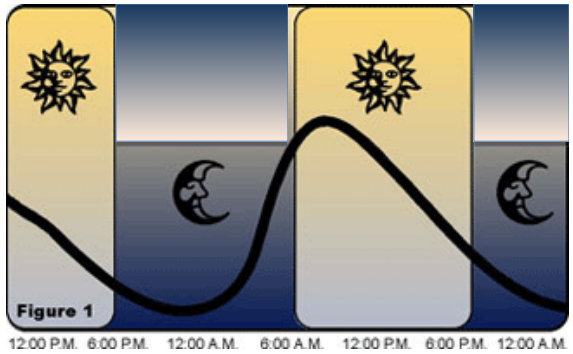
Compare more than two population means

- One explanatory variable: **One-way ANOVA**
- Multiple explanatory variables: Multifactorial ANOVA
- Assumptions not met: non-parametric alternatives
- You can always do: simulation-based test

Statistical inference and hypothesis testing

3. Compare more than 2 populations – one-way ANOVA

Example: A change in circadian clock (eg. Jet lag) is called a "phase shift", which can be measured by plasma melatonin levels. Campbell and Murphy (1998) reported that the human circadian clock can be reset by exposing the back of the knee to light. Wright and Czeisler (2002) re-examined the phenomenon. 22 participants were divided into three groups: no light (control), bright light to the back of the knee, and bright light to the eyes. Melatonin levels were measured 2 days later. A negative/positive number means a delay/advance of melatonin production compared to undisturbed cycle



How to compare mean phase shift in 3 groups?

Three t-tests??

http://www.afar.com/images/circadian.gif

Statistical inference and hypothesis testing

Multiple comparisons and error rate inflation

5% chance of false positive 5%chance of false positive

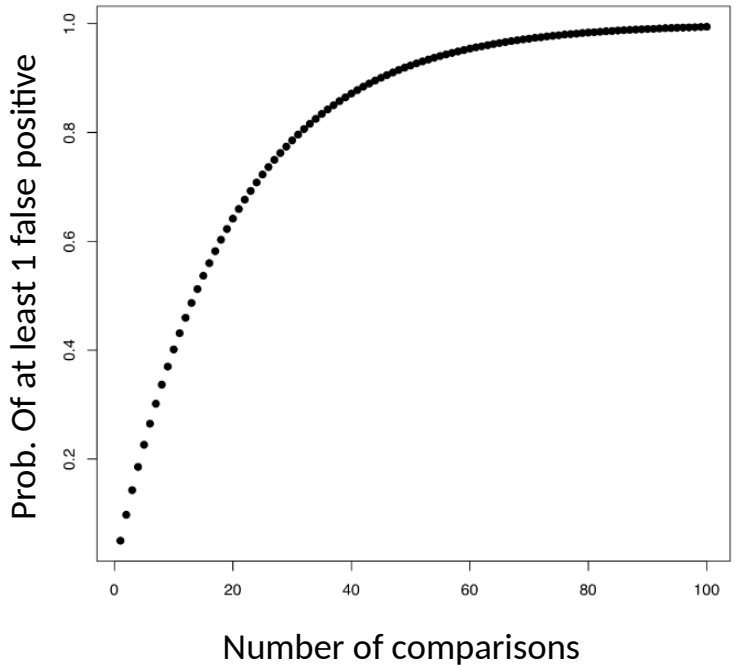
5% chance of false positive

Overall chance of false positive = $1-(1-0.05)^3 = 0.14$

Likelihood of false positive increases with number of comparisons!

Statistical inference and hypothesis testing

Multiple comparisons and error rate inflation

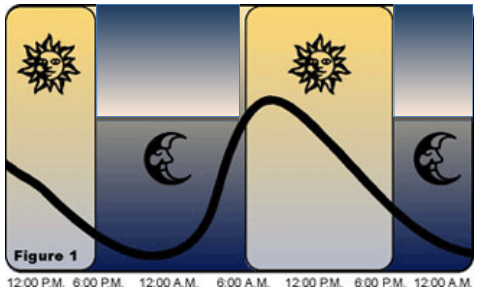


http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture10.pdf

Statistical inference and hypothesis testing

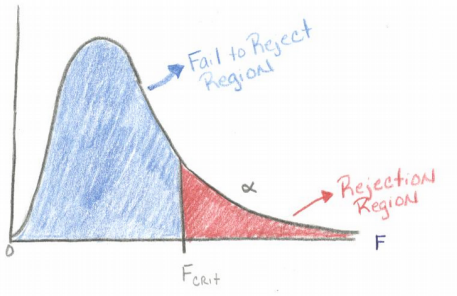
3. Compare more than 2 populations – one-way ANOVA

Example: 22 participants were divided into three groups: no light (control), bright light to the back of the knee, and bright light to the eyes. Melatonin levels were measured 2 days later. A negative/positive number means a delay/advance of melatonin production compared to undisturbed cycle



ANOVA (Analysis of Variance)

- Compares variances between and within group
- Calculates F-statistic. Always a one-tailed probability distribution.
- Shape defined by 2 degrees of freedom F(m,n)
- Result tells you likelihood of at least one difference between the groups but **not** which/how many groups



H_0: There is no difference in phase shift between all three groups

H_A: There is a difference in phase shift in at least one of the three groups

http://mtweb.mtsu.edu/stats/dictionary/delatog.htm

Statistical inference and hypothesis testing

3. Compare more than 2 populations – one-way ANOVA

Assumptions

i. No outliers
ii. Measurements in all groups are a random sample from the corresponding population
iii. The variable is normally distributed
iv. The variance between all groups is equal

Also consider

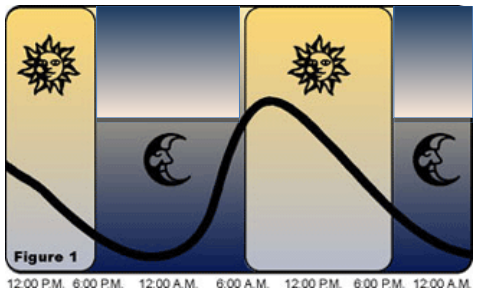
- **Fixed effect –** predefined groups which are of direct interest (e.g. treatment, genotype)
- **Random effect** – the groups are randomly sampled from a population (e.g. we want to know if different animal handlers, picked at random, affect animal stress. "Animal handler" is a random effect)
- **Repeated measures** – are measurements related (e.g. before and after drug administration)

Note: For 2 groups, an unpaired t-test and one-way ANOVA will give equivalent results

Statistical inference and hypothesis testing

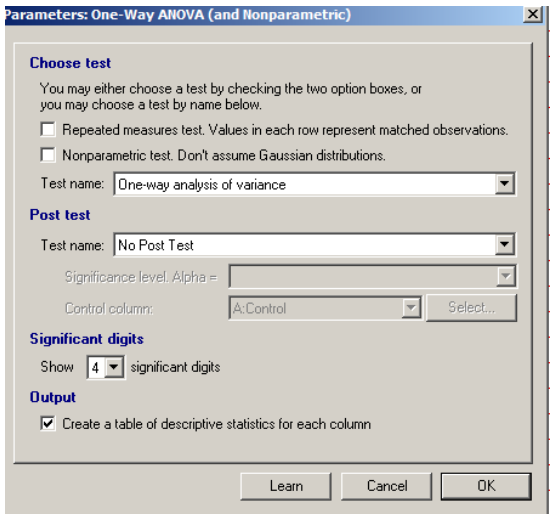
3. Compare more than 2 populations – one-way ANOVA

Example: 22 participants were divided into three groups: no light (control), bright light to the back of the knee, and bright light to the eyes. Melatonin levels were measured 2 days later. A negative/positive number means a delay/advance of melatonin production compared to undisturbed cycle



1) **State hypothesis**
2) **Test assumptions**
3) **Perform test**

A	B	C
Control	Knees	Eyes
Y	Y	Y
0.53	0.73	-0.78
0.36	0.31	-0.86
0.20	0.03	-1.35
-0.37	-0.29	-1.48
-0.60	-0.56	-1.52
-0.64	-0.96	-2.04
-0.68	-1.61	-2.83
-1.27		



Statistical inference and hypothesis testing

3. Compare more than 2 populations – one-way ANOVA

Example: 22 participants were divided into three groups: no light (control), bright light to the back of the knee, and bright light to the eyes. Melatonin levels were measured 2 days later. A negative/positive number means a delay/advance of melatonin production compared to undisturbed cycle

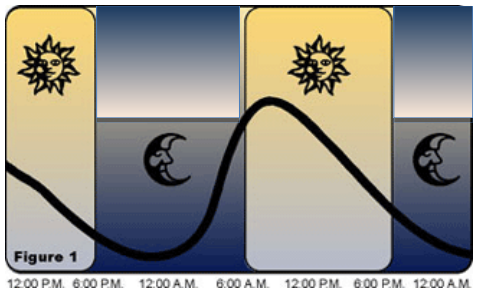


Table Analyzed	Data 1		
One-way analysis of variance			
P value	0.0045		
P value summary	**		
Are means signif. different? (P < 0.05)	Yes		
Number of groups	3		
F	7.289		
R squared	0.4342		
Bartlett's test for equal variances			
Bartlett's statistic (corrected)	0.3702		
P value	0.8310		
P value summary	ns		
Do the variances differ signif. (P < 0.05)	No		
ANOVA Table	SS	df	MS
Treatment (between columns)	7.224	2	3.612
Residual (within columns)	9.415	19	0.4955
Total	16.64	21	

R squared: fraction of variation in measure explained by group differences

F-statistic (ratio of errors)

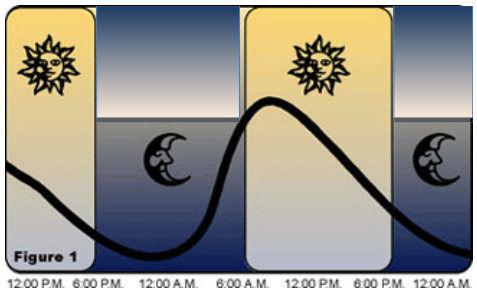
Group mean square: variation between groups

Error mean square: variation within groups

Statistical inference and hypothesis testing

3. Compare more than 2 populations – one-way ANOVA

Example: 22 participants were divided into three groups: no light (control), bright light to the back of the knee, and bright light to the eyes. Melatonin levels were measured 2 days later. A negative/positive number means a delay/advance of melatonin production compared to undisturbed cycle



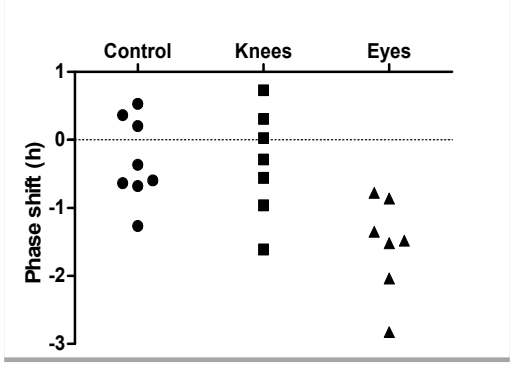
~~H_0~~: *There is no difference in phase shift between all three groups*

H_A: *There is a difference in phase shift in at least one of the three groups*

Which group/s differ? Follow significant ANOVAs with post-hoc tests

To compare all groups use **post-hoc tests**

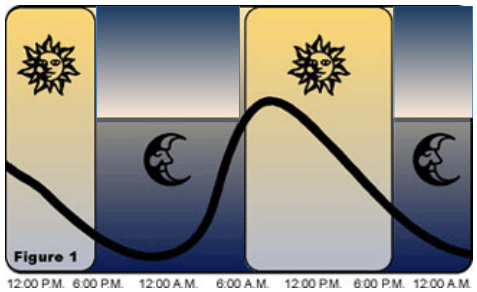
- Various available
- Tukey-Kramer good for equal variances

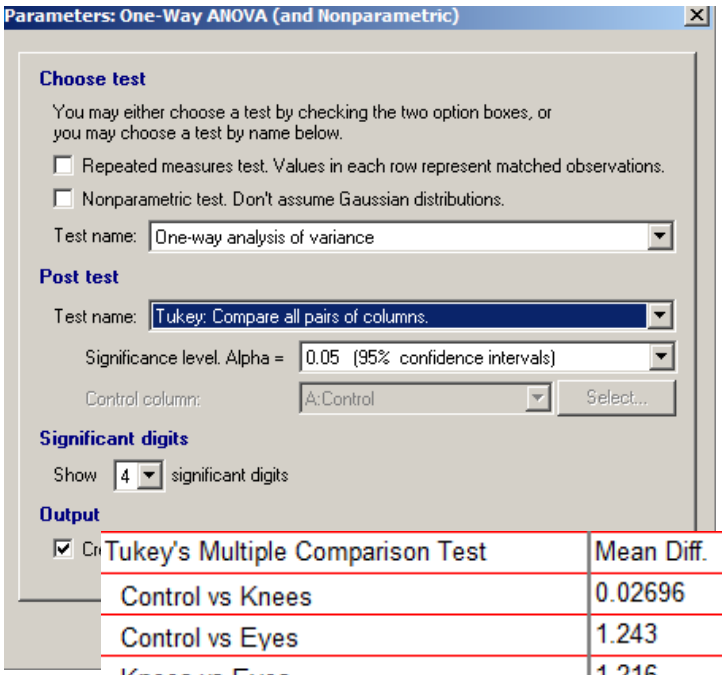


Statistical inference and hypothesis testing

3. Compare more than 2 populations – one-way ANOVA

Example: 22 participants were divided into three groups: no light (control), bright light to the back of the knee, and bright light to the eyes. Melatonin levels were measured 2 days later. A negative/positive number means a delay/advance of melatonin production compared to undisturbed cycle





Tukey's Multiple Comparison Test	Mean Diff.	q	Significant? P < 0.05?	Summary	95% CI of diff
Control vs Knees	0.02696	0.1047	No	ns	-0.8987 to 0.9526
Control vs Eyes	1.243	4.824	Yes	**	0.3171 to 2.168
Knees vs Eyes	1.216	4.569	Yes	*	0.2597 to 2.172

Statistical inference and hypothesis testing

3. Compare more than 2 populations – one-way ANOVA

Example: 22 participants were divided into three groups: no light (control), bright light to the back of the knee, and bright light to the eyes. Melatonin levels were measured 2 days later. A negative/positive number means a delay/advance of melatonin production compared to undisturbed cycle

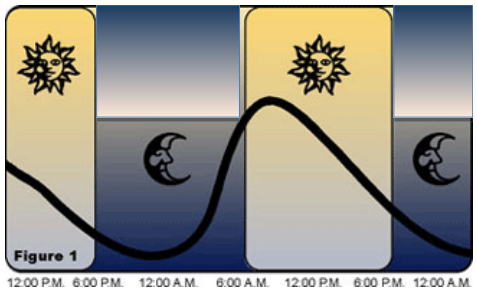


Table Analyzed	Data 1
One-way analysis of variance	
P value	0.0045
P value summary	**
Are means signif. different? (P < 0.05)	Yes
Number of groups	3
F	7.289
R squared	0.4342
Bartlett's test for equal variances	
Bartlett's statistic (corrected)	0.3702
P value	0.8310
P value summary	ns
Do the variances differ signif. (P < 0.05)	No
ANOVA Table	SS
Treatment (between columns)	7.224
Residual (within columns)	9.415
Total	16.64

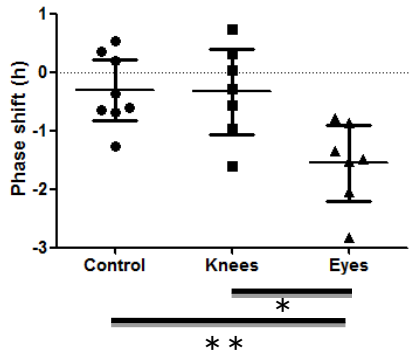
ANOVA Table	df	MS
Treatment (between columns)	2	3.612
Residual (within columns)	19	0.4955
Total	21	

Tukey's Multiple Comparison Test	Mean Diff.	q	Significant? P < 0.05?	Summary	95% CI of diff
Control vs Knees	0.02696	0.1047	No	ns	-0.8987 to 0.9526
Control vs Eyes	1.243	4.824	Yes	**	0.3171 to 2.168
Knees vs Eyes	1.216	4.569	Yes	*	0.2597 to 2.172

Interpretation and presentation

There was a significant difference in phase shift between groups ($p=0.005$, $F(2,19)=7.3$), with the eye group showing a greater phase delay than both other groups (Tukey-Kramer post-hoc test).

*Fig 1: Phase shifts after light exposure in 3 groups. Points represent individuals. Mean and 95% CI error bars. *, $p<0.05$, **, $p<0.01$*



Summary

When can we use ANOVA? What does it tell us? How is it related to t tests?

▸ Back to overview

Compare proportions of nominal variables

You are studying development in fruit flies and are interested in a gene that affects wing growth. The gene has two alleles, s (small wings) and W (wildtype). The wildtype allele is dominant. Adult flies carrying two copies of the s have shorter wings than normal flies. This phenotype is well described and easy to measure. However, you suspect that the gene also has a function in early development and that early embryonic development is often disrupted in individuals homozygous for the s allele, meaning many of them never make it to the adult stage.
Here is data from a recent experiment:

Phenotype	number
normal wing length:	318
short wings:	82

Compare proportions of nominal variables

- Most often used: χ^2 **("Chi square") test**
- Better for small numbers: Fisher's exact test
- Always works: Simulation-based test

Chi-Square test

Remember all hypothesis tests are the same!

Chi-Square test

Remember all hypothesis tests are the same!

- What is H_0 ? What is H_A ?
- What data would you expect to see if H_0 was true?
- How likely are you to see your data under H_0?

Chi-square test: Idea

- Table of observed and expected values:

Phenotype	number observed	number expected
normal wing length:	318	300
short wings:	82	100

Chi-square test: Idea

- Table of observed and expected values:

Phenotype	number observed	number expected
normal wing length:	318	300
short wings:	82	100

- Compute the χ^2 statistic
  $\sum \frac{(O-E)^2}{E}$

Chi-square test: Idea

- Table of observed and expected values:

Phenotype	number observed	number expected
normal wing length:	318	300
short wings:	82	100

- Compute the χ^2 statistic
  $\sum \frac{(O-E)^2}{E}$
- Determine degrees of freedom (basically, number of categories - 1)

Chi-square test: Idea

- Table of observed and expected values:

Phenotype	number observed	number expected
normal wing length:	318	300
short wings:	82	100

- Compute the χ^2 statistic
  $\sum \frac{(O-E)^2}{E}$
- Determine degrees of freedom (basically, number of categories - 1)
- Consult a χ^2 table to get your p-value.

Chi-square test: Idea

- Table of observed and expected values:

Phenotype	number observed	number expected
normal wing length:	318	300
short wings:	82	100

- Compute the χ^2 statistic
  $\sum \frac{(O-E)^2}{E}$
- Determine degrees of freedom (basically, number of categories - 1)
- Consult a χ^2 table to get your p-value. In our case: 0.043

Back to overview

Relationships between variables

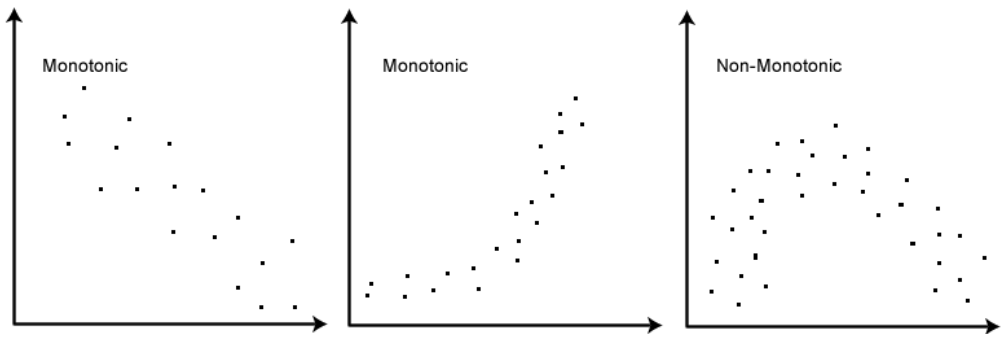
Answers the questions: Is there a relationship between variables A and B?

Statistical inference and hypothesis testing

5. Relationships between variables

You want to know if two quantitative variables are associated

1) Look at the data



Scatterplot for non-linearity/ non-monotonic / outliers

Statistical inference and hypothesis testing

5. Relationships between variables

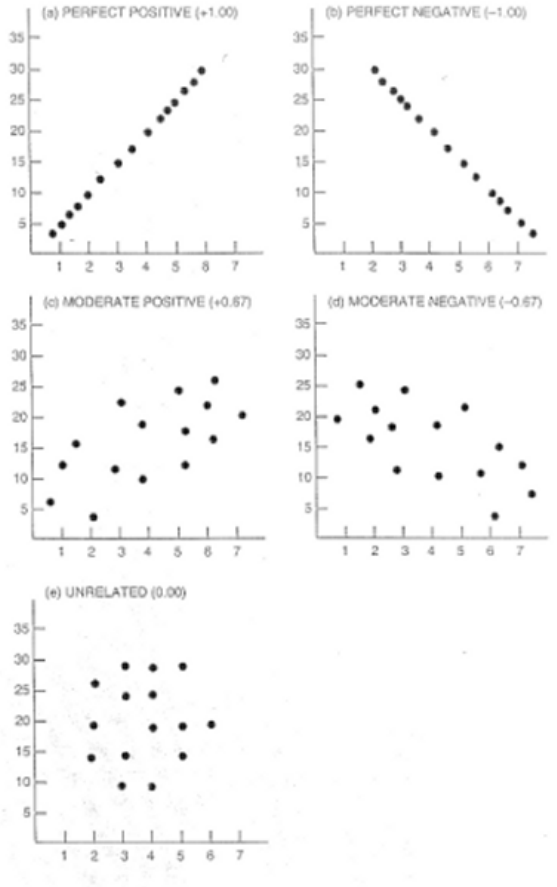
Which test?

Correlation	Linear Regression
Degree of linear relationship between two variables	Fits a straight line
Correlation is not causality	Assumes causality
X and Y values can be randomly sampled	Can be used to predict value of Y given X

http://www.biostathandbook.com/linearregression.html

Statistical inference and hypothesis testing

5. Relationships between variables



Pearson's correlation (Pearson's product moment) - parametric

- Correlation coefficient, r $(-1 \leq r \leq 1)$
- r^2 quantifies the amount of variability in one variable accounted for by correlation with the other
- Assumptions:
  1. Both variables normally distributed (Very sensitive to outliers!)
  2. Both variables continuous
  3. Monotonic relationship

Spearman's rank correlation – non-parametric

- OK for non-normal distributions, discrete variables
- Calculates correlation coefficient $-1 \leq \rho \leq 1$
- Uses ranked data

Statistical inference and hypothesis testing

5. Relationships between variables

Example: Huber et al (2004) asked whether there was an associated between improvements in task performance and the amount of slow wave sleep after learning the task. The data show % for time spend in SWS and % improvement on task from baseline

sleep	improvement
Y	Y
8	8
14	3
13	0
15	0
17	8
18	15

Data normal, continuous and monotonic

H_0: $r=0$

H_A: $r \neq 0$

Test: Parametric (Pearson's Correlation)

Number of XY Pairs	10
Pearson r	0.86
95% confidence interval	0.51 to 0.97
P value (two-tailed)	0.0013
P value summary	**
Is the correlation significant? (alpha=0.05)	Yes
R squared	0.75

Correlation coefficient

75% of variability in variables is accounted for by other variable

Presentation

The authors found a significant correlation between %SWS and % improvement (r=0.86, p=0.001, n=10, 95% CI(0.51, 0.97), Pearson's correlation)

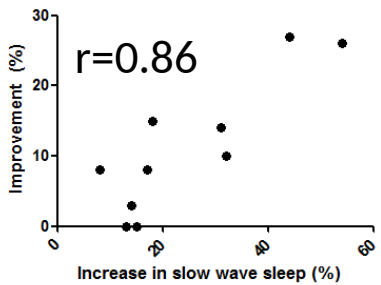


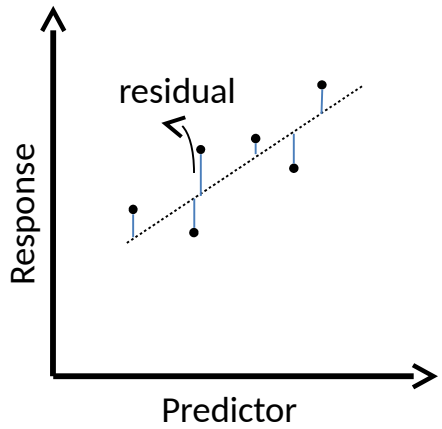
Fig 1. Relationship between %SWS and % Improvement. Points represent individual subjects

Statistical inference and hypothesis testing

5. Relationships between variables

Linear Regression

- Finds best-fitting straight line through data
- Minimises residuals to get best fit
- Output of regression
  - Model: $y = b_0 + b_1 x$, intercept and slope constants with p-values
  - Coefficient of determination r^2 : amount of variation explained by regression



- Assumptions:
  1. Predictor (x) variables error free
  2. Linearity
  3. Constant variance (plot residuals vs x – available in most stats packages)
  4. Errors of response uncorrelated
  5. Normality (but this is a weak requirement). If non-normal, try to transform data (eg. Logarithms)

Statistical inference and hypothesis testing

5. Relationships between variables

Example: rattlesnakes increase their body temperature after a meal. Tattersall et al (2004) asked whether the snakes could control their body temperature after a meal without extrinsic factors such as moving to a sunny spot. They measured body temperature changes after feeding the snakes meals of different sizes

Linear regression assumptions OK

meal.size	tempchange
Y	Y
0.0	0.3
0.0	0.2
0.0	
12.0	
14.5	
23.3	
27.1	
29.3	
20.8	
22.2	
23.4	
36.3	
34.7	
31.6	
29.4	
30.9	
50.4	

Best-fit values	
Slope	0.02738 ± 0.003447
Y-intercept when X=0.0	0.3196 ± 0.09098
X-intercept when Y=0.0	-11.67
1/slope	36.52
95% Confidence Intervals	
Slope	0.02004 to 0.03473
Y-intercept when X=0.0	0.1257 to 0.5134
X-intercept when Y=0.0	-24.95 to -3.716
Goodness of Fit	
r^2	0.8080
Sy.x	0.1914
Is slope significantly non-zero?	
F	63.11
DFn, DFd	1.000, 15.00
P value	< 0.0001
Deviation from zero?	Significant

Slope of straight line

CI of slope

Amount of variation accounted for by regression

Significance of relationship

Statistical inference and hypothesis testing

5. Relationships between variables

Example: rattlesnakes increase their body temperature after a meal. Tattersall et al (2004) asked whether the snakes could control their body temperature after a meal without extrinsic factors such as moving to a sunny spot. They measured body temperature changes after feeding the snakes meals of different sizes

meal.size	tempchange
Y	Y
0.0	0.3
0.0	
0.0	
12.0	
14.5	
23.3	
27.1	
29.3	
20.8	
22.2	
23.4	
36.3	
34.7	
31.6	
29.4	
30.9	
50.4	

Linear regression assumptions OK

Best-fit values	
Slope	0.02738 ± 0.003447
Y-intercept when X=0.0	0.3196 ± 0.09098
X-intercept when Y=0.0	-11.67
1/slope	36.52
95% Confidence Intervals	
Slope	0.02004 to 0.03473
Y-intercept when X=0.0	0.1257 to 0.5134
X-intercept when Y=0.0	-24.95 to -3.716
Goodness of Fit	
r^2	0.8080
Sy.x	0.1914
Is slope significantly non-zero?	
F	63.11
DFn, DFd	1.000, 15.00
P value	< 0.0001
Deviation from zero?	Significant

Presentation

Meal size significantly predicted temperature change in rattlesnakes (b=0.027, 95% CI[0.020, 0.034], F(1,15) =63.1), and accounted for 81% of the variance in temperature change

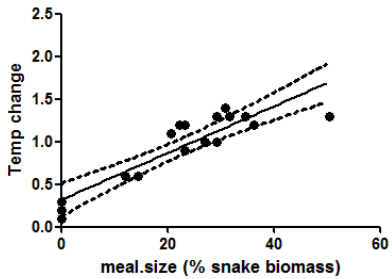


Fig 1. Temperature change was dependent on meal size. Points represent individual snakes. Regression line (solid) and 95% CI lines (dotted) indicate.

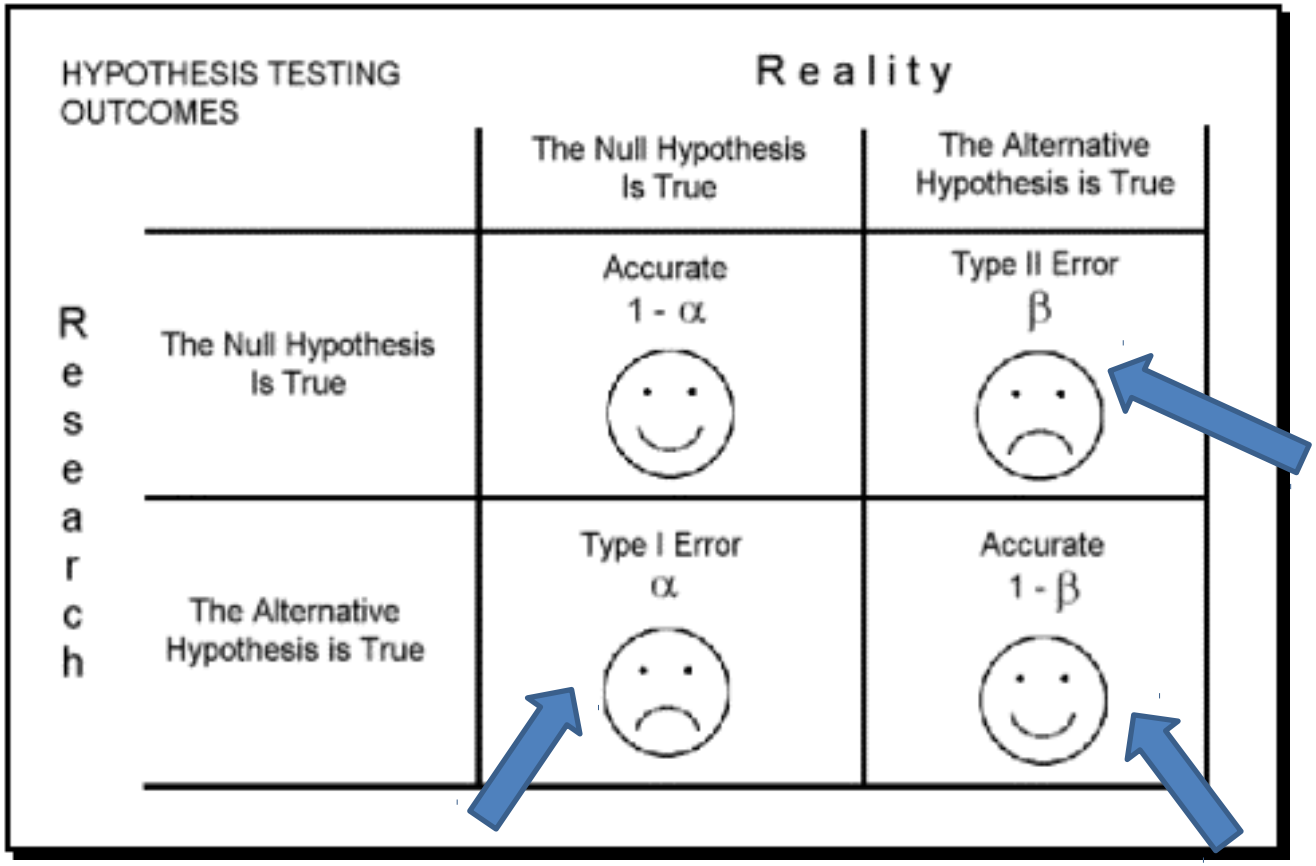
Summary

What do these tests tell us? What don't they tell us?

▸ Back to overview

Outline

Experimental design & sample size

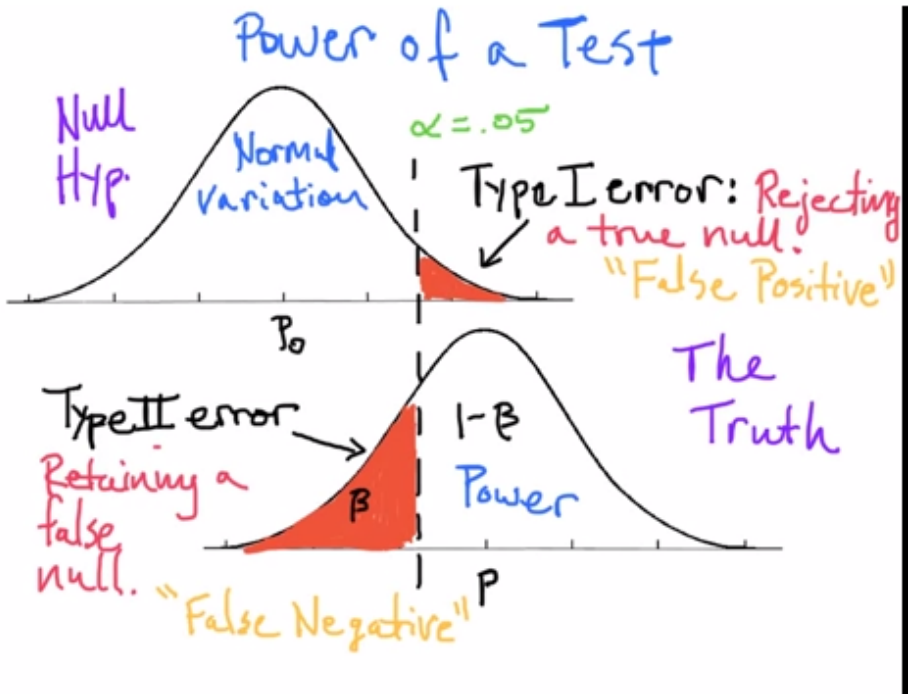


False negative

Probability of false positive
Eg. $\alpha = 0.05$
Statistical tests use this

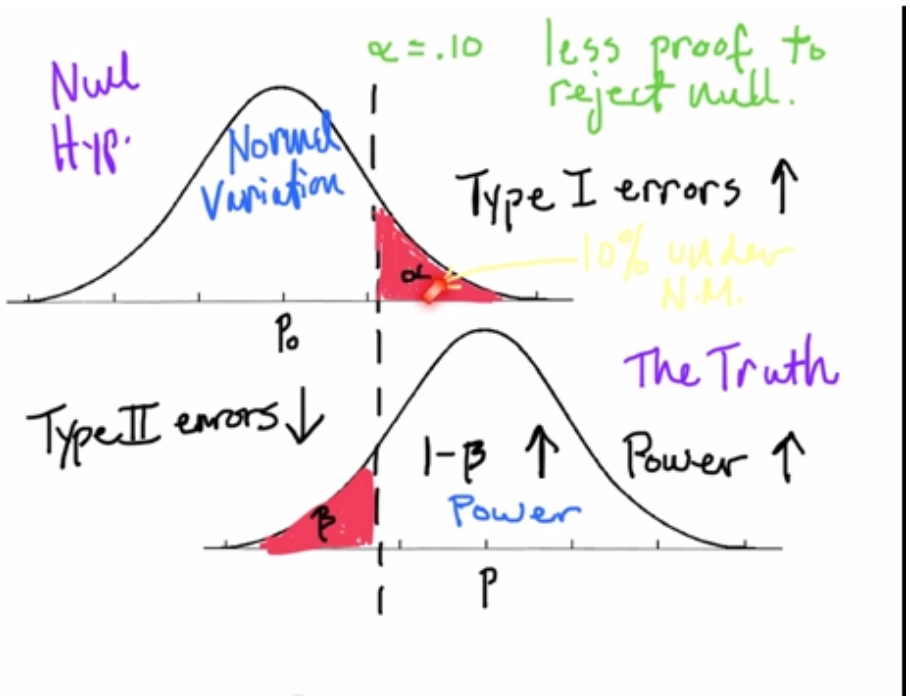
Power
Needs to be considered in experimental design stage

Experimental design & sample size



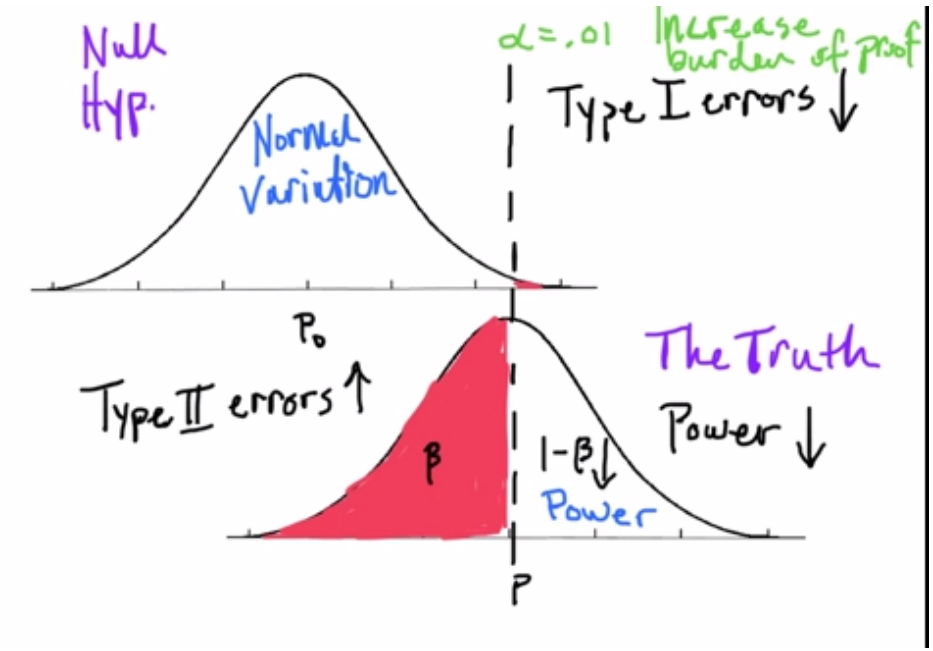
http://www.youtube.com/watch?v=0R5Zscd_Mp8

Experimental design & sample size



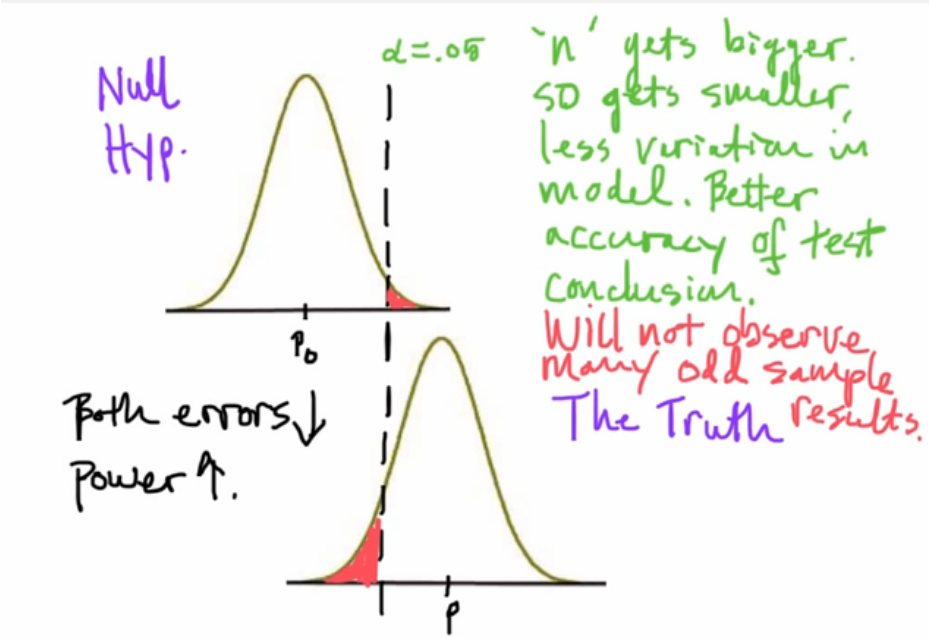
http://www.youtube.com/watch?v=0R5Zscd_Mp8

Experimental design & sample size



http://www.youtube.com/watch?v=0R5Zscd_Mp8

Experimental design & sample size



How big should n be?

Too many subjects: expensive, unethical

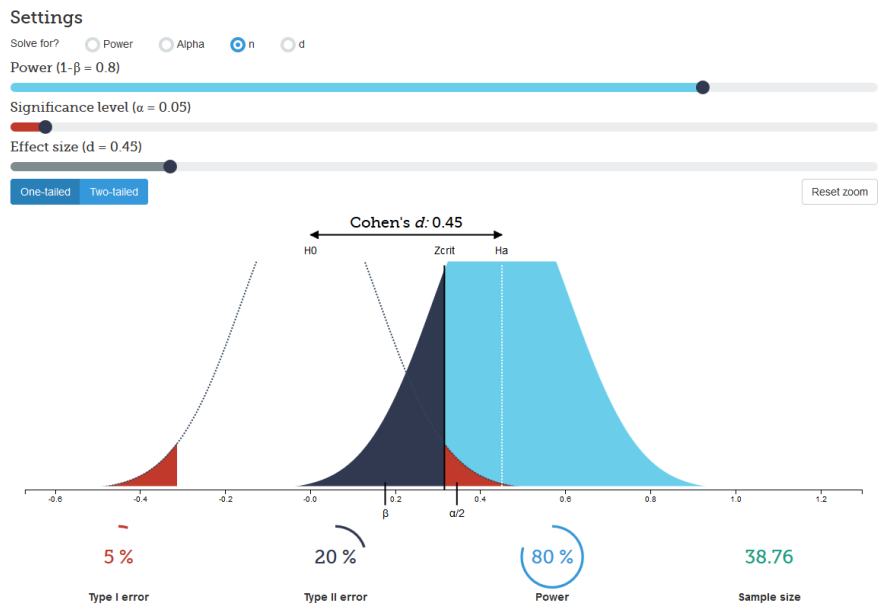
Too few subjects: increased likelihood of false negative (type II) error

Experimental design & sample size

Requirements for samples size calculations:

1. Significance level (α, usually 0.05)
2. Power: probability of detecting an existing effect. Usually >80
3. Effect size

Specific formula for sample size calculation depends on test used (distribution)



http://rpsychologist.com/d3/NHST/

See

- Dell et al., (2002) **Sample size determination.** ILAR J. 43(4): 207–213.
- Power calculation software: G*Power (http://www.gpower.hhu.de/en.html)

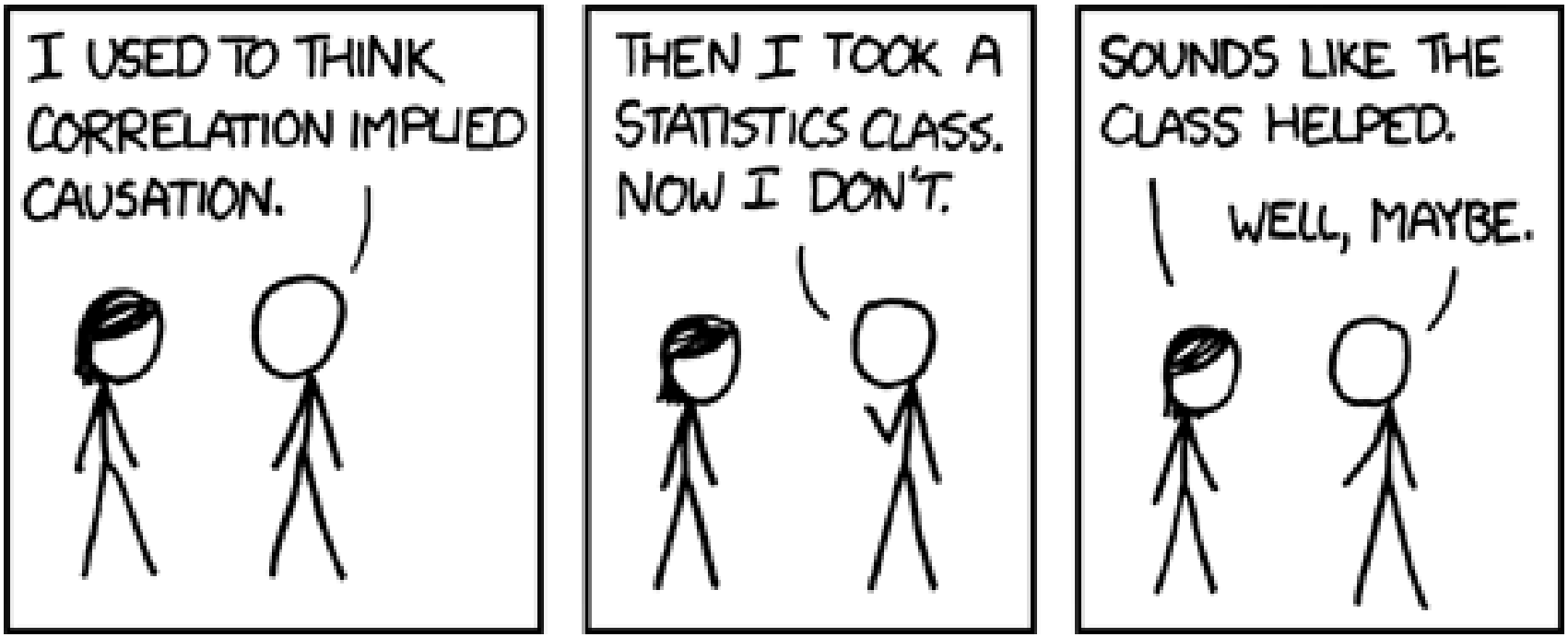
Outline

Sharing your results - good practice

Sharing your results - good practice

- Share your raw data (e.g. by using a repository)
- Be clear about tests used, assumptions made etc.
- Use scripting rather than point-and-click software
- Share your analysis code
- Make data analysis plan before conducting experiments
- ...

Thank you



https://xkcd.com/552/