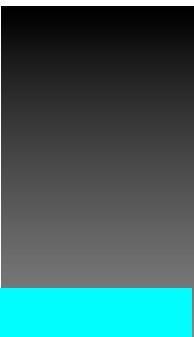
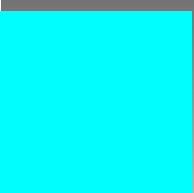
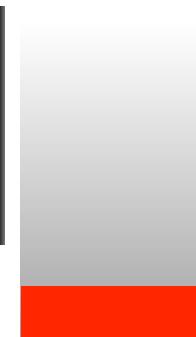
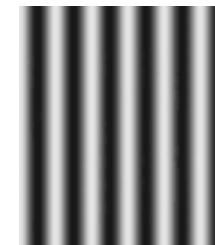


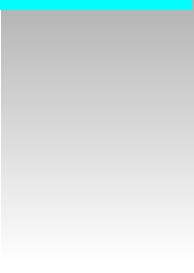
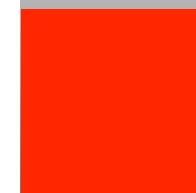
Smallest font



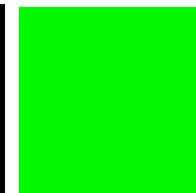
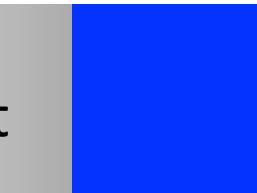
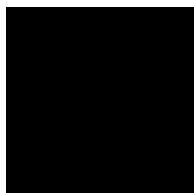
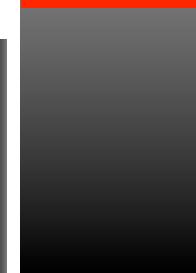
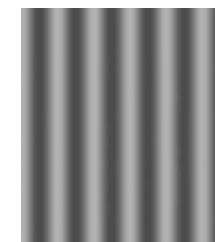
Welcome



Calibration slide



Stand by

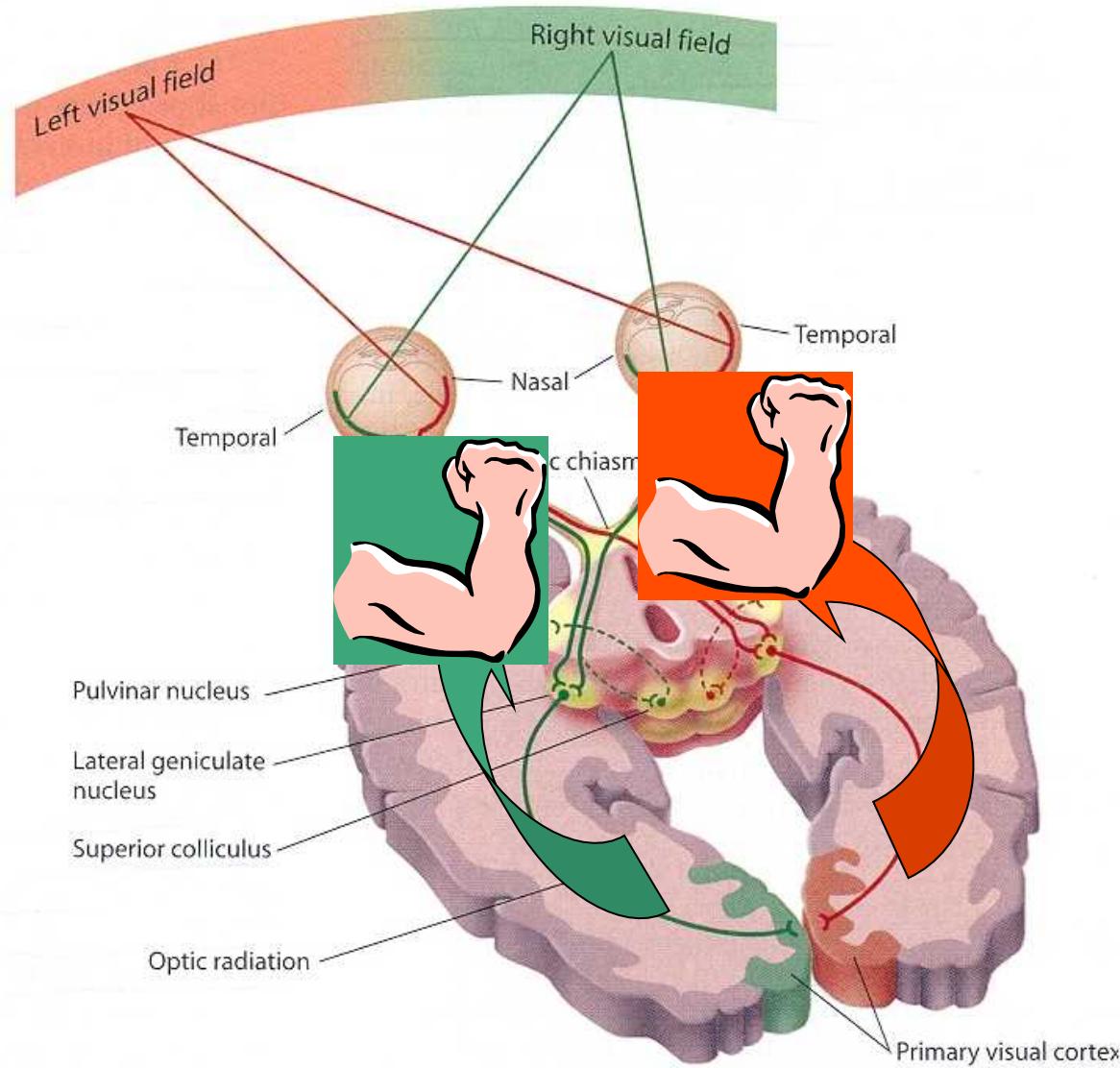


Smallest font

# Scientific Programming and Computing for the Behavioral Sciences



# The visual processing cascade



# The canonical data analysis cascade

**1 Loader** - This program puts data into Matlab format.

Analogous to transducer: Retina or cochlea.

**2 Pruner/Integrity checker/Filter** - This makes sure that the data to be analyzed is usable. Analogous to filter function of LGN or MGN. Thalamus. Choking off irrelevant data.

**3 Categorizer.** Analogous to V1. Format the data properly.

**4a/b/c Calculator** Analogous to extrastriate cortex. Do specialized calculations on the same data. Specialized streams.

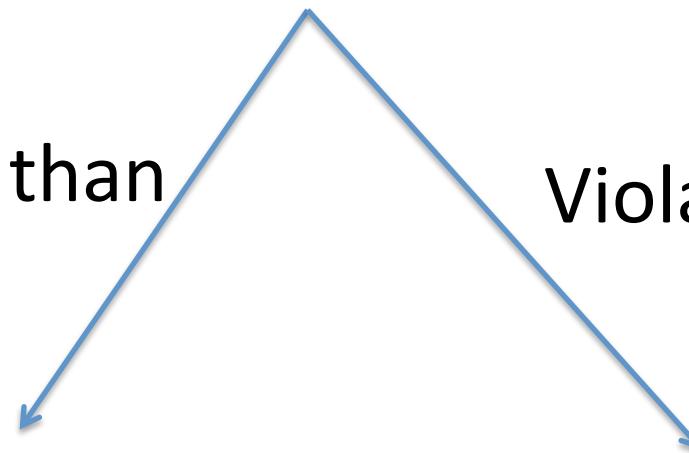
**5a/b/c Plotter** - Output. Analogous to motor cortex. Makes figures. Also saves files.

**6 Wrapper** - invoking the previous 5 programs, in the correct order. That would be the brain itself. Heavily commented.

# Null hypothesis significance testing

- How far – in units of SEM – are empirical sample means from each other?
- How likely is that to happen by chance (p)?
- Known population standard deviation → z-test
- Unknown population standard deviation → t-test.

Comparing more than  
2 sample means



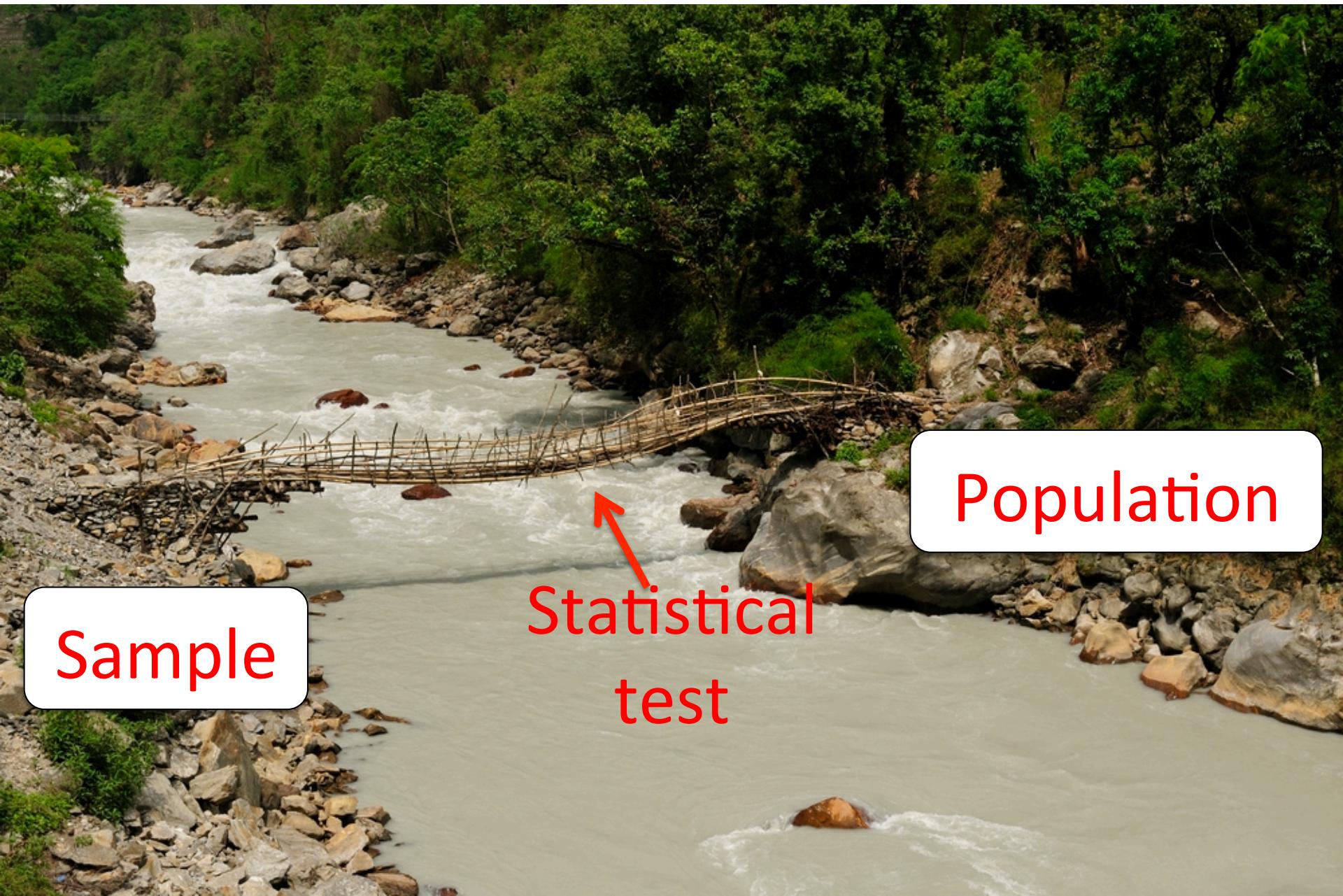
**ANOVA**

**Nonparametric tests**

# Inferential statistics

- We want to know general truths about populations
- We only ever have access to samples. In science, as in life.
- In order to close this fundamental gap as best as we can, we do inferential statistics.

# Statistical tests



Sample

Population

Statistical  
test

# Inferential statistics

- We want to know general truths about populations
- We only ever have access to samples. In science, as in life.
- In order to close this fundamental gap as best as we can, we do inferential statistics.
- We use some kind of test statistic to assess how likely the observed difference between an observed and expected sample mean is.
- Most simply:

$$z = \frac{\text{observed} - \text{expected}}{SE}$$

# Why can't we always just use z?

- It relies on the central limit theorem.
- In other words, it doesn't work for small samples (below n=20 or so).
- Also, it relies on a known standard deviation:

$$z = \frac{\text{observed} - \text{expected}}{SE}$$

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

- We don't always have that, but need to estimate it from the sample itself.
- So we use t instead.
- What is t and how is it distributed?

# Student's t-test

- William Gosset
- Employed by Guinness brewery.
- Not allowed to publish under his name.
- Published as “student”
- Worked out a test for small sample sizes and unknown population variance.
- Intended for quality control.



# The rationale

- Due to central limit theorem, sample means will distribute normally, given a large number of observations in the sample.
- It is not always possible to obtain large samples.
- In that case, the sample means will follow a t-distribution.
- That's basically what Gosset worked out.

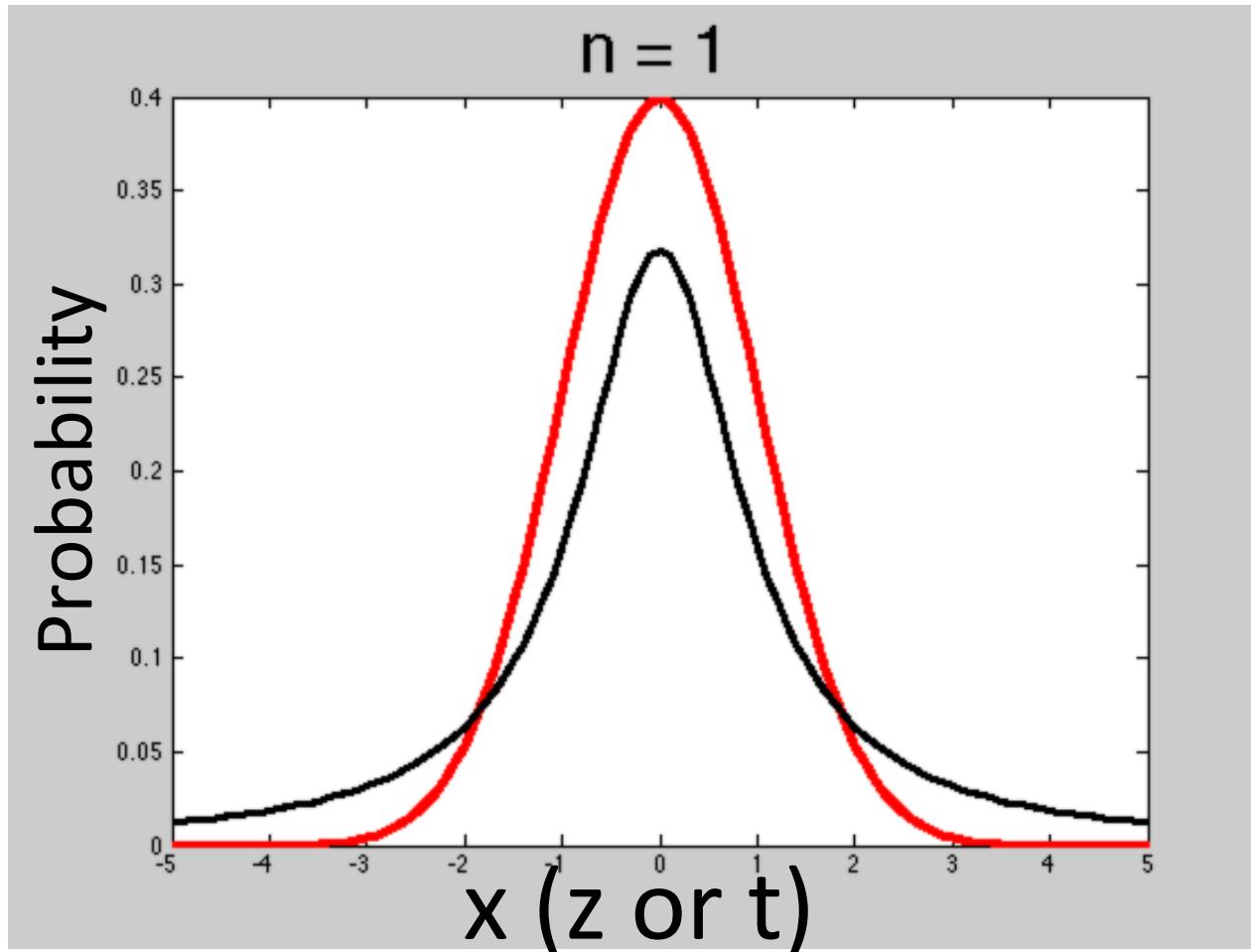
t:

**Normal:**

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$y = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

The t distribution approaches the normal distribution for large n



# What does this mean in practice?

- We can't just use
- because we don't know the population standard deviation.
- We have to estimate it from the sample itself.
- This introduces a degree of uncertainty.
- At small sample sizes, this matters.
- The SE will be larger than expected.
- To correct, we use a t distribution instead (and use  $s$  as an estimate of  $\sigma$ , but we lose degrees of freedom).

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

# Two kinds of t-test

- There are actually two kinds of t-test to compare sample means, depending on the number of independent groups involved.
- 2 independent groups: t-test for independent groups (or samples)
- 1 independent group: t-test for dependent groups (or paired samples or correlated groups).

# t Test for Independent Groups

- Suitable if one has a two-group design.
- Comparing means in group 1 vs. group 2.
- Either a treatment group vs. a control group (often placebo) or two treatment groups.
- It is a parametric test.
- degrees of freedom:  $df = n_1 + n_2 - 2$

# t Test for Correlated Groups (paired samples / dependent samples)

- The previous test was suitable for a “between-subjects-design” where there were different people in the two treatment groups (assigned randomly).
- This is not always the case. In some experiments, the \*same\* people are in both groups, inherent to the logic of a “within-subjects-design”. But in this case, the observations are no longer independent.
- So we need to use a test for correlated groups.
- Heuristic: If there is one measure of a given variable for each person in the group, use the t-test for independent groups. If you have two (like in repeated measures), use the one for correlated groups.

# Procedure

- The two scores per person must be converted into one score.
- Most straightforwardly by subtraction.
- So we work with difference scores.
- This has implications for the null hypothesis.
- Null hypothesis for correlated groups:
  - $H_0: \mu_1 - \mu_2 = 0$
  - In practice...  $H_0: \bar{D} = 0$

# Null hypothesis significance testing

- How far – in units of SEM – are empirical sample means from each other?
- How likely is that to happen by chance (p)?
- Known population standard deviation → z-test
- Unknown population standard deviation → t-test.

Comparing more than  
2 sample means

**ANOVA**

Violated assumptions

**Nonparametric tests**

# Analysis of Variance (ANOVA)

- Useful if we have more than 2 treatment groups.
- In real life, we will often have more than 2 treatment groups.
- For instance, testing the effect of caffeine on performance. Most of the time, we wouldn't just test "caffeine vs. no caffeine", as dose is likely to matter. Or gender.
- So we quickly get a lot of groups.

# Example 1: One-way ANOVA

Treatment

No caffeine

100 mg  
caffeine

200 mg  
caffeine

# Example 2: One-way ANOVA (more levels)

Treatment
No caffeine
50 mg
100 mg
150 mg
200 mg
250 mg
300 mg

# Example 3: Two-way ANOVA

Treatment	Male	Female
No caffeine		
50 mg		
100 mg		
150 mg		
200 mg		
250 mg		
300 mg		

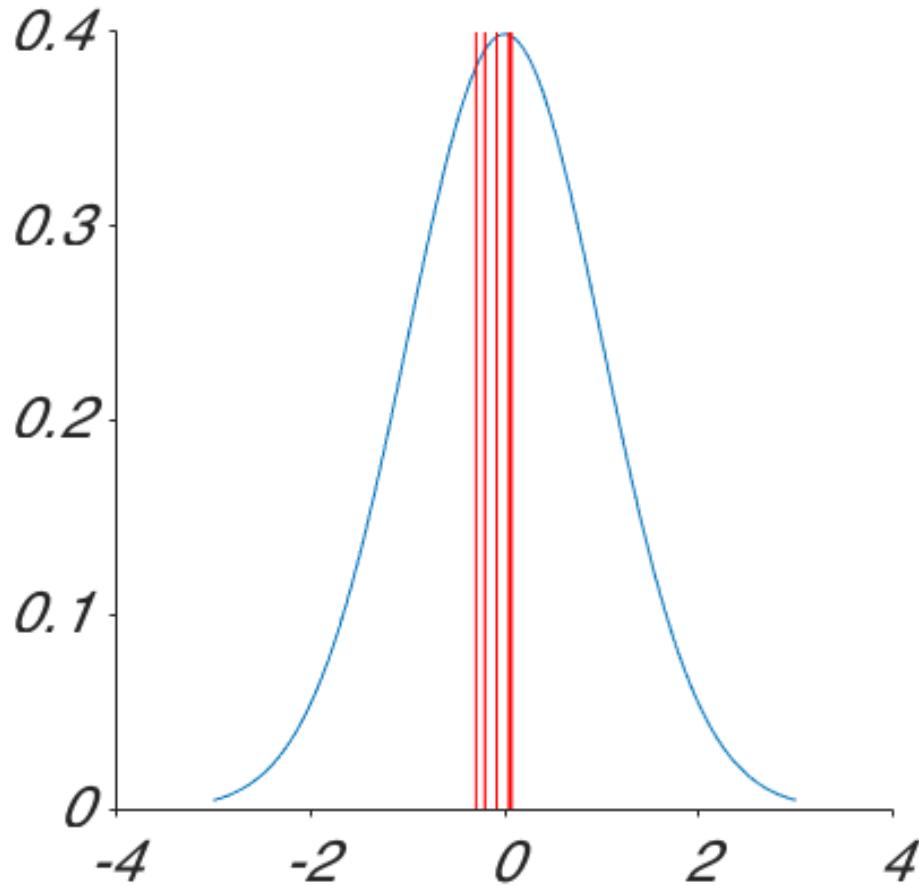
# Example 4: Three-way ANOVA

Age	Young	
	Old	
Treatment	Male	Female
No caffeine		
50 mg		
100 mg		
150 mg		
200 mg		

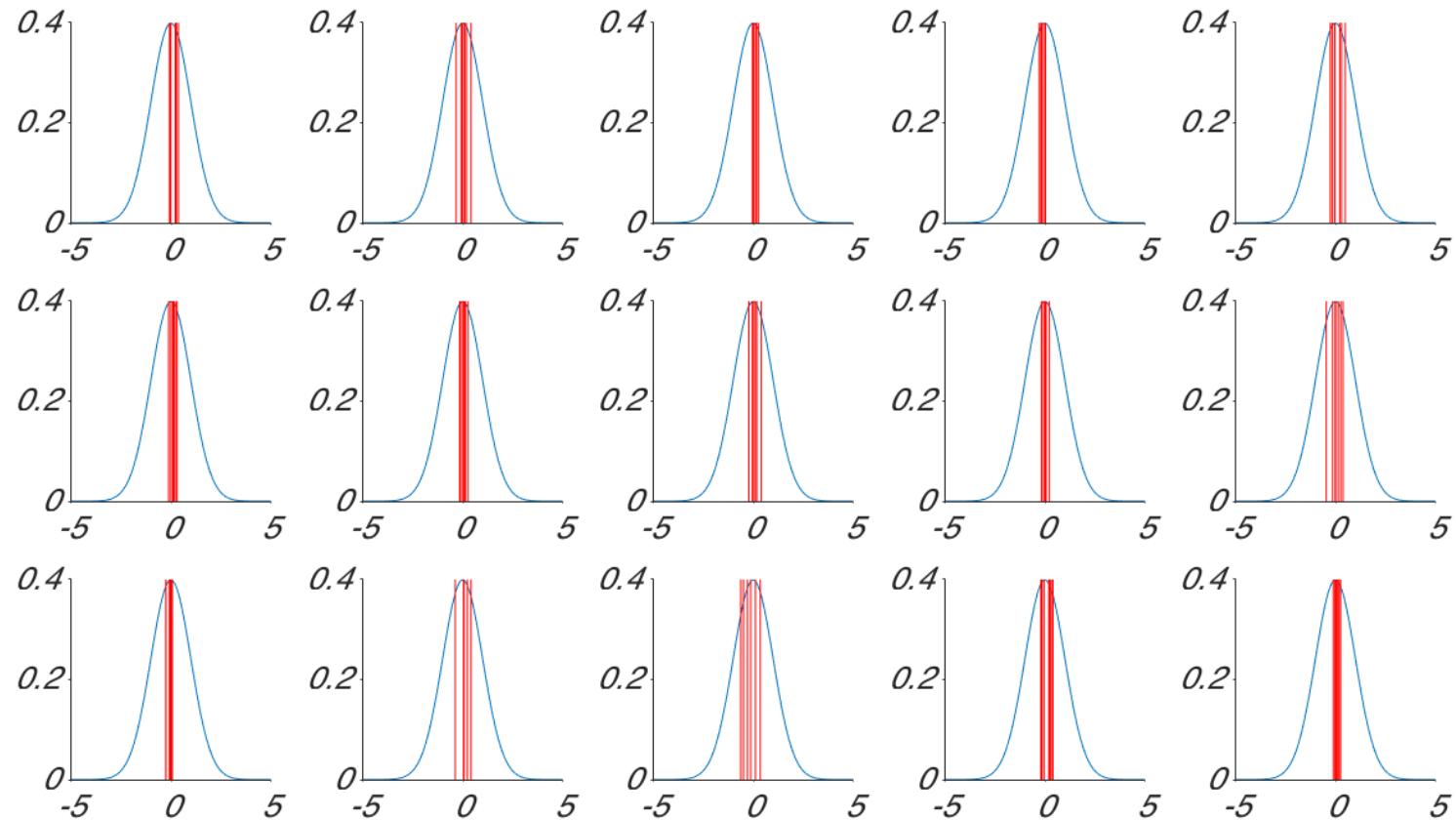
# What is the logic of the ANOVA?

- Same as the t-test, basically.
- But with more groups.
- Big idea: Due to the central limit theorem, **sample means** will cluster around the population mean and be distributed as a normal distribution with known variation.
- If the empirical distribution of sample means deviates from this, the samples are unlikely to be drawn from the same population.
- This is as true for 2 samples as it is for n samples.

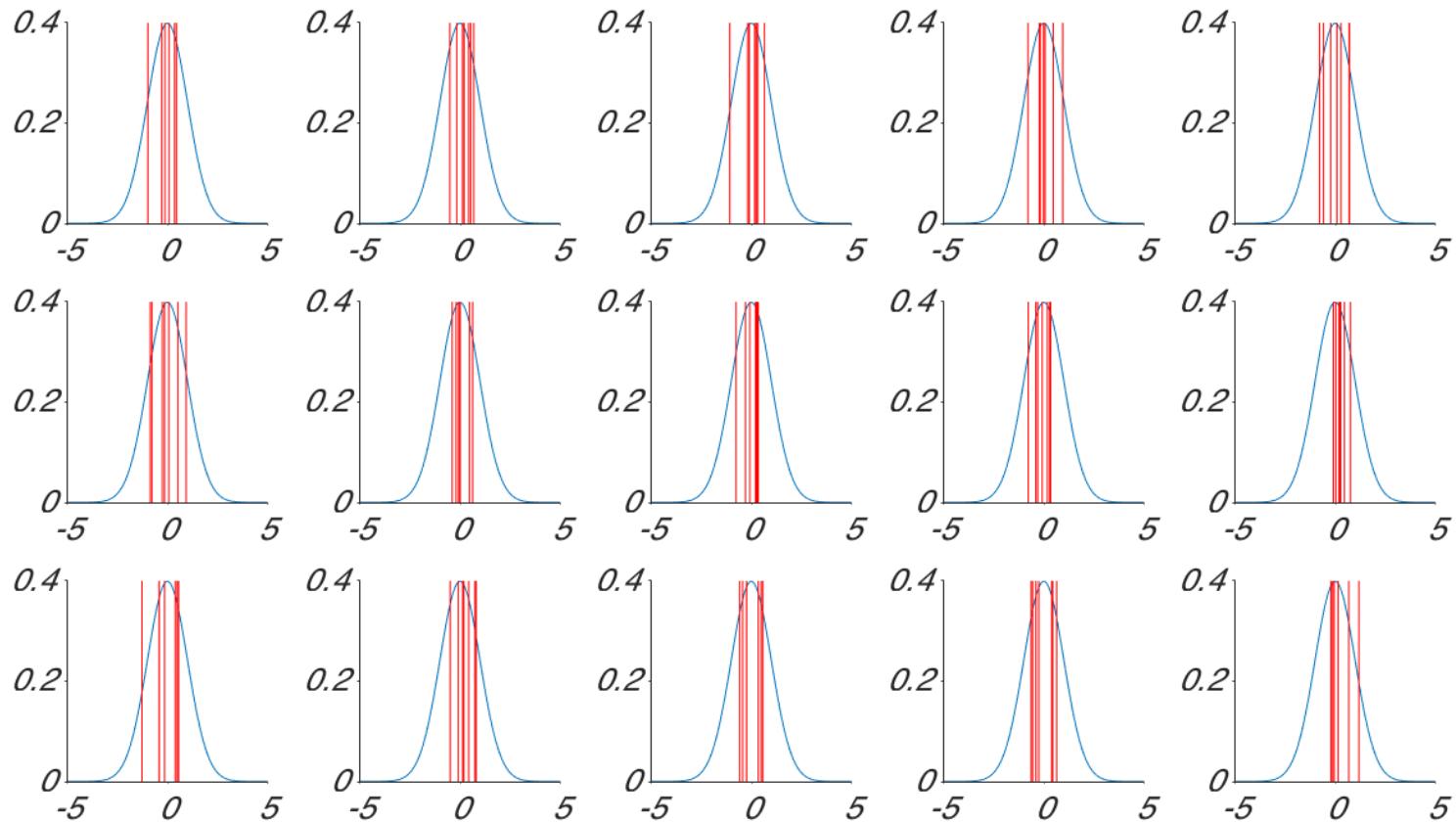
Illustrated – 7 conditions, 20 participants each, 1 draw



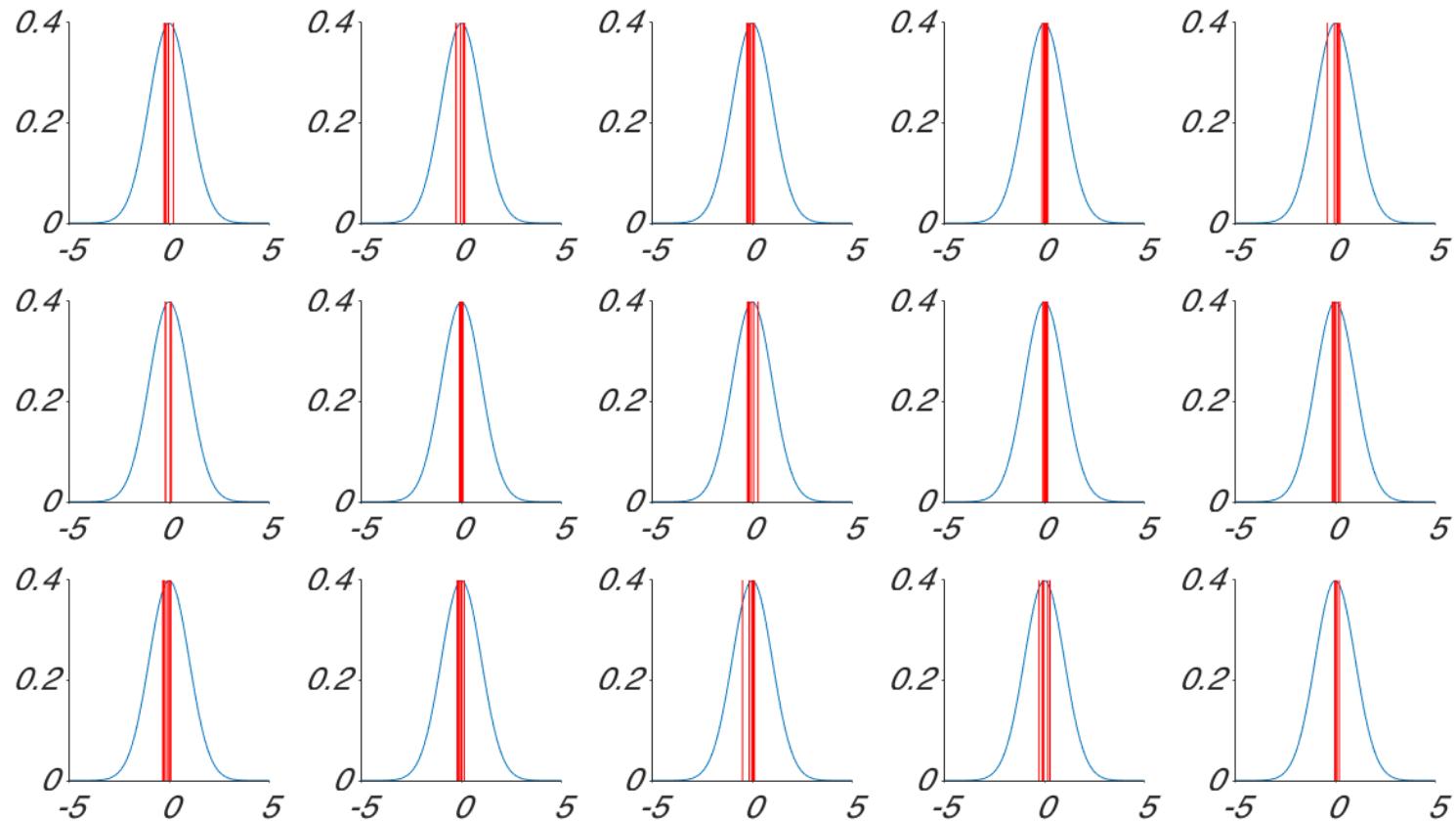
# Illustrated – 7 conditions, 20 participants each, 15 draws



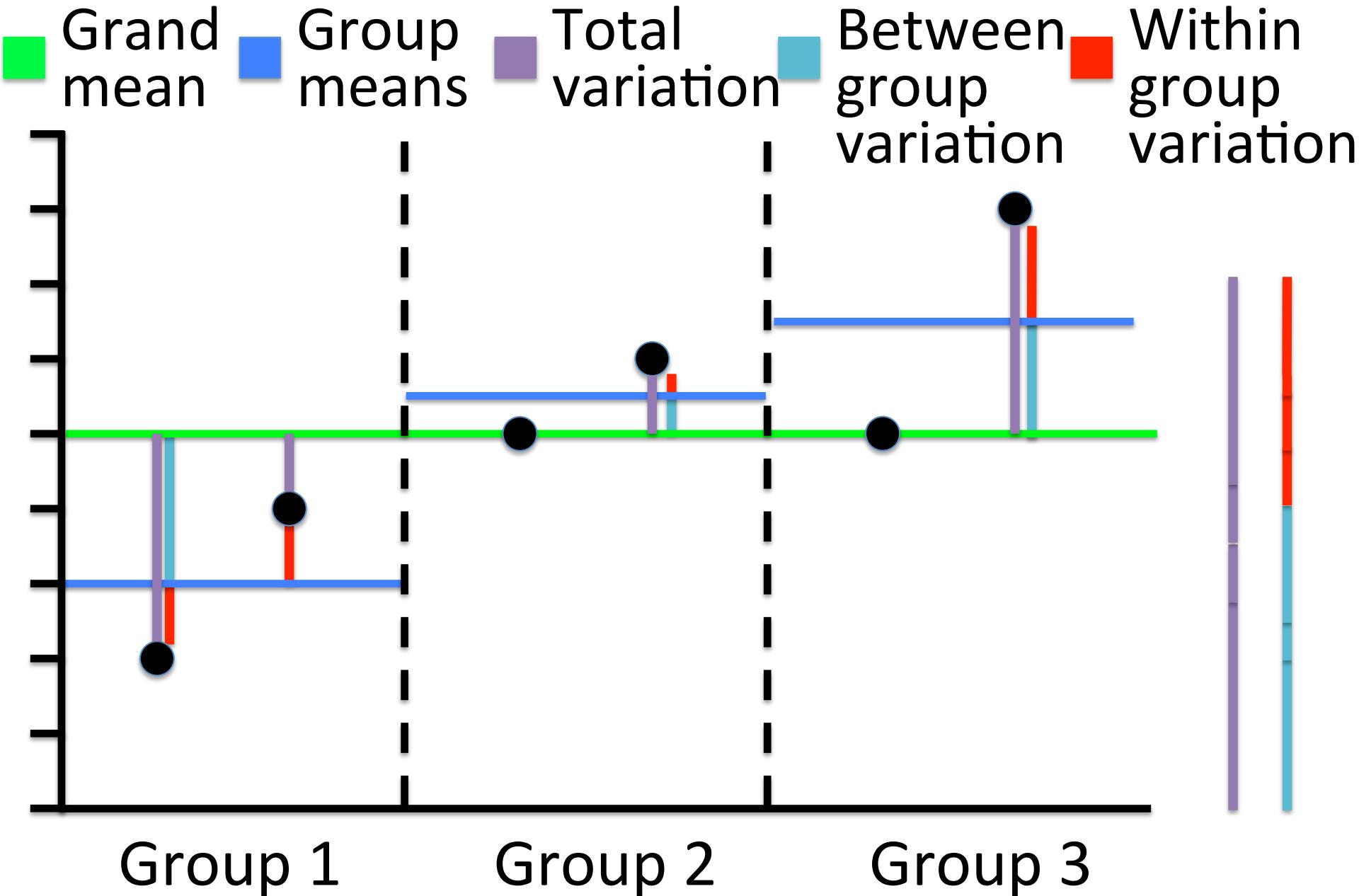
# Illustrated – 7 conditions, 5 participants each, 15 draws



# Illustrated – 7 conditions, 40 participants each, 15 draws



# ANOVA logic illustrated



# What can contribute to variation in individual scores?

- Individual variation
- Measurement error (of the dependent variable)
- Variation in delivery (of the independent variable)
- Effectiveness of the treatment

# Number of terms of full models

- 1:  $Y = I + A + e$  (2)
- 2:  $Y = I + A + B + AB + e$  (4)
- 3:  $Y = I + A + B + C + AB + AC + BC + ABC + e$  (8)
- 4:  $Y = I + A + B + C + D + AB + AC + AD + BC + \dots$   
 $BD + CD + ABC + ABD + ACD + BCD + ABCD + e$  (16)
- 5:  $Y = I + A + B + C + D + E + AB + AC + AD + AE + BC\dots$   
 $+ BD + BE + CD + CE + DE + ABC + ABD + ABE + ACD\dots$   
 $+ ACE + ADE + BCD + BCE + BDE + CDE + ABCD + \dots$   
 $ABCE + ABDE + ACDE + BCDE + ABCDE + e$  (32)
- Explosive:  $2^k$  terms
- Number of terms in 10 factor model: 1024
- 20 factors: >1,000,000

# The ANOVA table

## Tests of Between-Subjects Effects

Dependent Variable: Assault

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	865.833 <sup>a</sup>	5	173.167	7.791	.000
Intercept	3648.100	1	3648.100	164.129	.000
Light	190.678	1	190.678	8.579	.004
Protein	673.400	2	336.700	15.148	.000
Light * Protein	1.756	2	.878	.039	.961
Error	1867.067	84	22.227		
Total	6381.000	90			
Corrected Total	2732.900	89			

a. R Squared = .317 (Adjusted R Squared = .276)

Intercept: Tests whether mean of all scores is different from zero = (mean score)<sup>2</sup>\*n

Error: Within group variation

Corrected Model: SS Model (Factors + Interactions)

Corrected Total: SS Total (Factors + Interactions + Error)

"Total": SS Total plus intercept (Factors + Interactions + Error + Intercept)

$\eta^2$  minus  
penalty for  
multiple  
factors

# Null hypothesis significance testing

- How far – in units of SEM – are empirical sample means from each other?
- How likely is that to happen by chance (p)?
- Known population standard deviation → z-test
- Unknown population standard deviation → t-test.

Comparing more than  
2 sample means

**ANOVA**

Violated assumptions

**Nonparametric tests**

# Assumptions of parametric tests

- They assume a normal (bell-shaped) distribution.
- They assume that mean and standard deviation are meaningful (assume that data are better than ordinal scale level).
- Independent groups t-test: Observations are independent.
- “Homogeneity of variance”: The variation in each underlying population is the same.

# Violating homogeneity of variance

High mean wealth,  
high variation in wealth

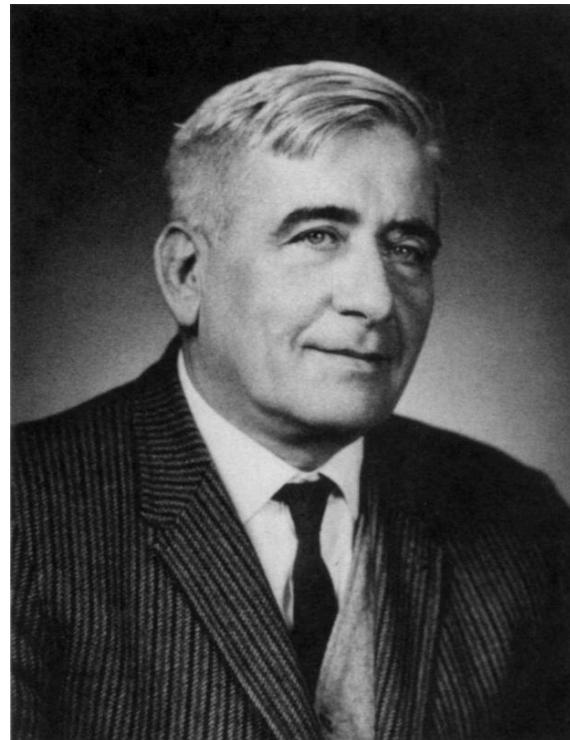


Low mean wealth,  
low variation in wealth



# The Mann-Whitney U test

- Aka Wilcoxon rank-sum test



Frank Wilcoxon

Henry Mann

Ransom Whitney

# Mann-Whitney-Wilcoxon

- Analogy to the t-test (which tests whether two samples have the same mean).
- Tests whether two samples come from distributions with the same median.
- This allows it to be used for data measured on ordinal scale and – as the median is robust to outliers – coming pretty much from any distribution.
- Test statistic: U or rank sums

# The basic idea

- Let's say we have two samples with values.
- If we have at least data on ordinal level, we arrange them all in rank order from smallest to largest value, regardless of which sample they came from.
- If both samples come from the same underlying distribution, there should be random mixing and a similar rank sum.
- $U$  quantifies probability of obtaining a given difference in rank sums.
- Similarly: Kruskal-Wallis test is a nonparametric analogue to ANOVA.