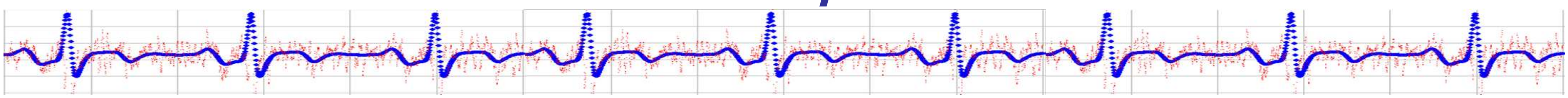
Empirical Research Methods in Information Science

IS 4800 / CS6350



Lecture 20

Outline

- T2 presentations
- Finish (quickly)
  - *t* tests
  - Power
- Work in teams for T3 – Experimental!

Last time

- *t* test for a single sample
- *t* test for dependent means

Reporting results

- Significant results

  $t(df)=\textit{tscore}$, $p<\textit{sig}$

  e.g., $t(38)=4.72$, $p<.05$ (two-tailed test)

  (If type of tail not noted, assume two-tailed; if one-sample t-test, note it (rare))

- Non-significant results

  e.g., $t(38)=4.72$, n.s.

- Note: usually report absolute value of t score and mean and SD of sample

Reporting results example

"As predicted, participants … rated substances with hard-to-predict names (M=4.12, SD=0.78) as more harmful than substances with easy-to-pronounce names (M=3.7, SD=0.74), t(19) = 2.41,P<.03"

Reporting results example

"During the eight test trials for gesture, dogs performed significantly above chance on at target trials: one-sample *t* test ($t(13) = 5.3$, $p < .01$)"

Z and t

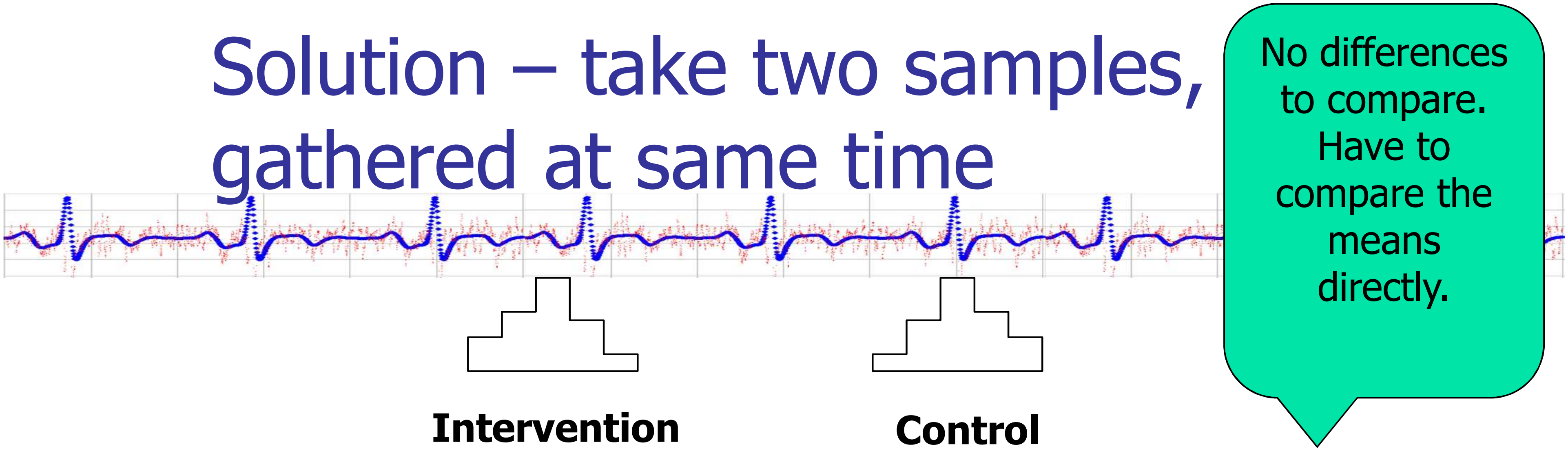
Table 7-9 Review of the *Z* Test, the *t* Test for a Single Sample, and the *t* Test for Dependent Means

Features	Type of Test		
	***Z* Test**	***t* Test for a Single Sample**	***t* Test for Dependent Means**
Population variance is known	Yes	No	No
Population mean is known	Yes	Yes	No
Number of scores for each participant	1	1	2
Shape of comparison distribution	*Z* distribution	*t* distribution	*t* distribution
Formula for degrees of freedom	Not applicable	$df = N - 1$	$df = N - 1$
Formula	$Z = (M - \mu_M)/\sigma_M$	$t = (M - \mu)/S_M$	$t = (M - \mu)/S_M$

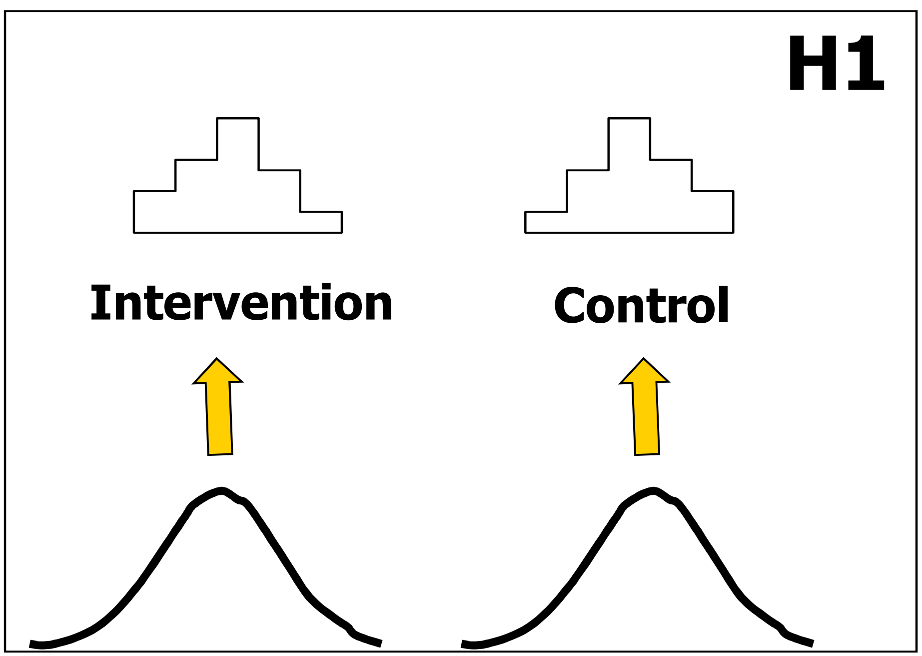
t-test for independent means

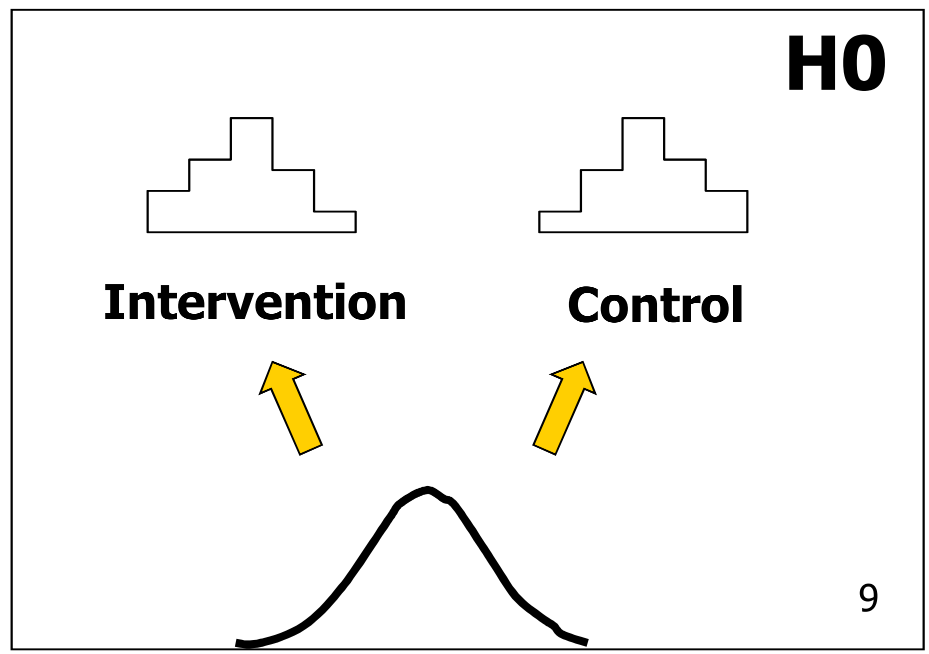
- Two *unrelated* samples
  - E.g., as obtained in a between-subjects experiment
- No other information about comparison distribution
  - Do not know population variances (so t-test situation, where they are estimated)

Solution – take two samples, gathered at same time

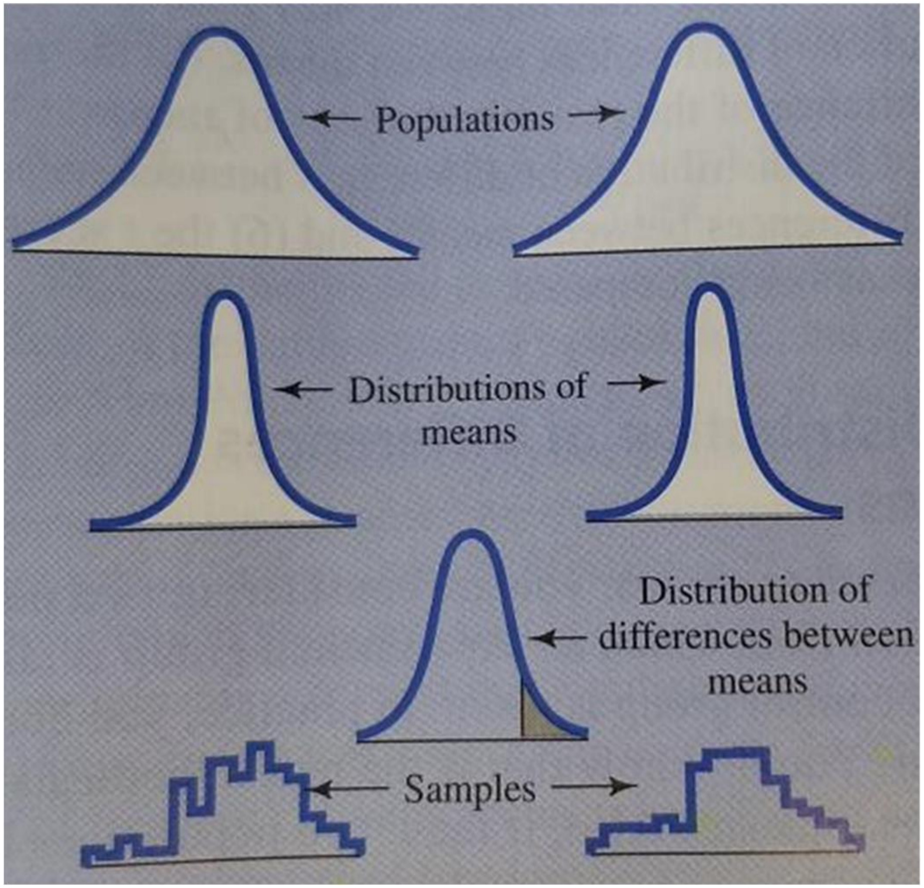


The big question: which is correct?

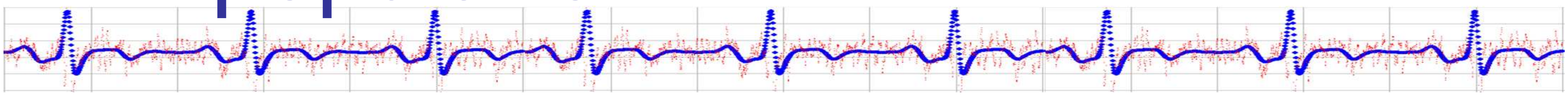




Now two steps removed from population

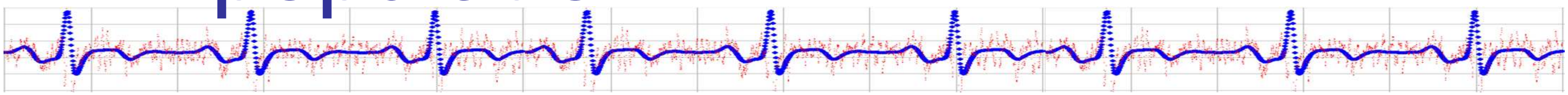


Now two steps removed from population



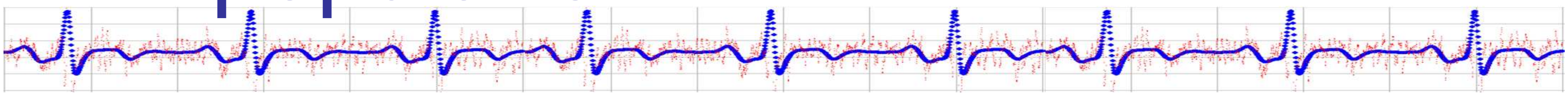
- 1st: Distribution of means from each population
- 2nd: Distribution of differences between each pair of means, one of each pair of means taken from it's particular distribution of means

Now two steps removed from population



- Don't know the actual mean, but do know for null hypothesis: $\mu_1 = \mu_2$ (so two distributions of means equal as well … and differences between means of samples would balance out to 0)

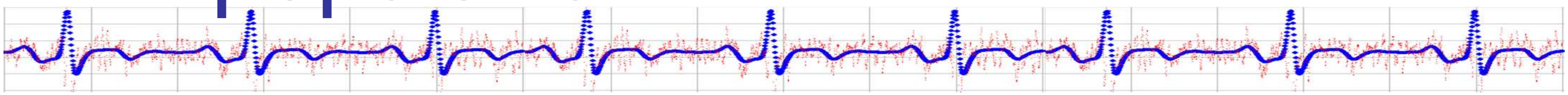
Now two steps removed from population



- Estimate the variances of the population based on samples
  - Rarely identical
  - Since supposed to be the same, get the pooled estimate of the population variance ($S^2{}_{Pooled}$) (if same size)
  - If not the same size, take weighted average of variance based on dfs (N-1)

$$S^2{}_{Pooled} = \frac{df_1}{df_{Total}}\left({S_1}^2\right) + \frac{df_2}{df_{Total}}\left({S_2}^2\right)$$

Now two steps removed from population

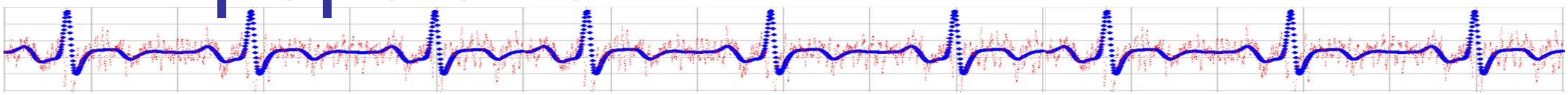


- If samples not the same size, the distributions of means do not have the same variance

$$S^2{}_{M1} = \frac{S^2{}_{Pooled}}{N_1}$$ (also compute $S^2{}_{M2}$)

(Variance of distribution of means is the population variance divided by the sample size)

Now two steps removed from population



- Variance of the distribution of the differences between means:

$$S^2_{Difference} = S^2_{M_1} + S^2_{M_2}$$

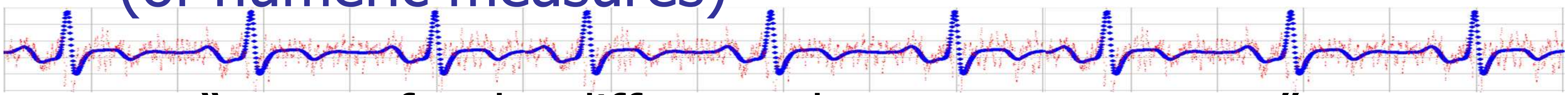
(In a difference between two numbers, variation in each contributions to the overall variations in their differences)

Terminology

Table 8-1 Summary of Different Types of Variance Used for the *t* Test for Independent Means

Type of Variance	Symbol
Estimated population variance for the first population	S_1^2
Estimated population variance for the second population	S_2^2
Pooled estimate of the population variance	S_{Pooled}^2
Variance of the distribution of means for the first population (based on an estimated population variance)	$S_{M_1}^2$
Variance of the distribution of means for the second population (based on an estimated population variance)	$S_{M_2}^2$
Variance of the distribution of differences between means	$S_{\text{Difference}}^2$

Stepping back: Wanted: a statistic to measure how similar two samples are
(of numeric measures)



- "t score for the difference between two means"

$$t = \frac{M_1 - M_2}{S_?}$$

- If samples are identical, t=0
- As samples become more different, t increases.

- What is the comparison distribution?
  - Want to compute probability of getting a particular t score IF the samples actually came from the same distribution.

Why t?

- In this situation, you do not know the population parameters; they must be estimated from the samples.
- When you have to estimate a comparison population's variance, the resulting distribution is not normal – it is a "t distribution".
  - Looks normal, but has thicker tails (need more extreme Z score for significance)
  - As df increases, t becomes normal
- The particular kind of t distribution we are using in this case is called a "distribution of the difference of means".

All things t

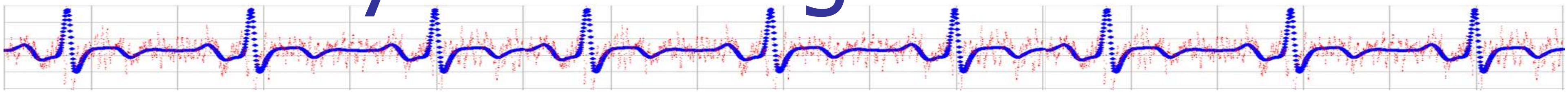
- t distribution shape is parameterized by “degrees of freedom”
- For a distribution of the difference of means,

$$df = df_1 + df_2 = (N_1 - 1) + (N_2 - 1)$$

Assumptions for t

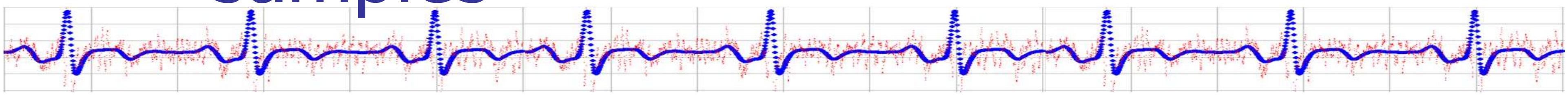
- Scores are sampled randomly from the population
- The sampling distribution of means is normal
- Variances of the two populations (whether they are the same or different) are the same.
  - *There are other forms of the statistic that do not make this assumption.*

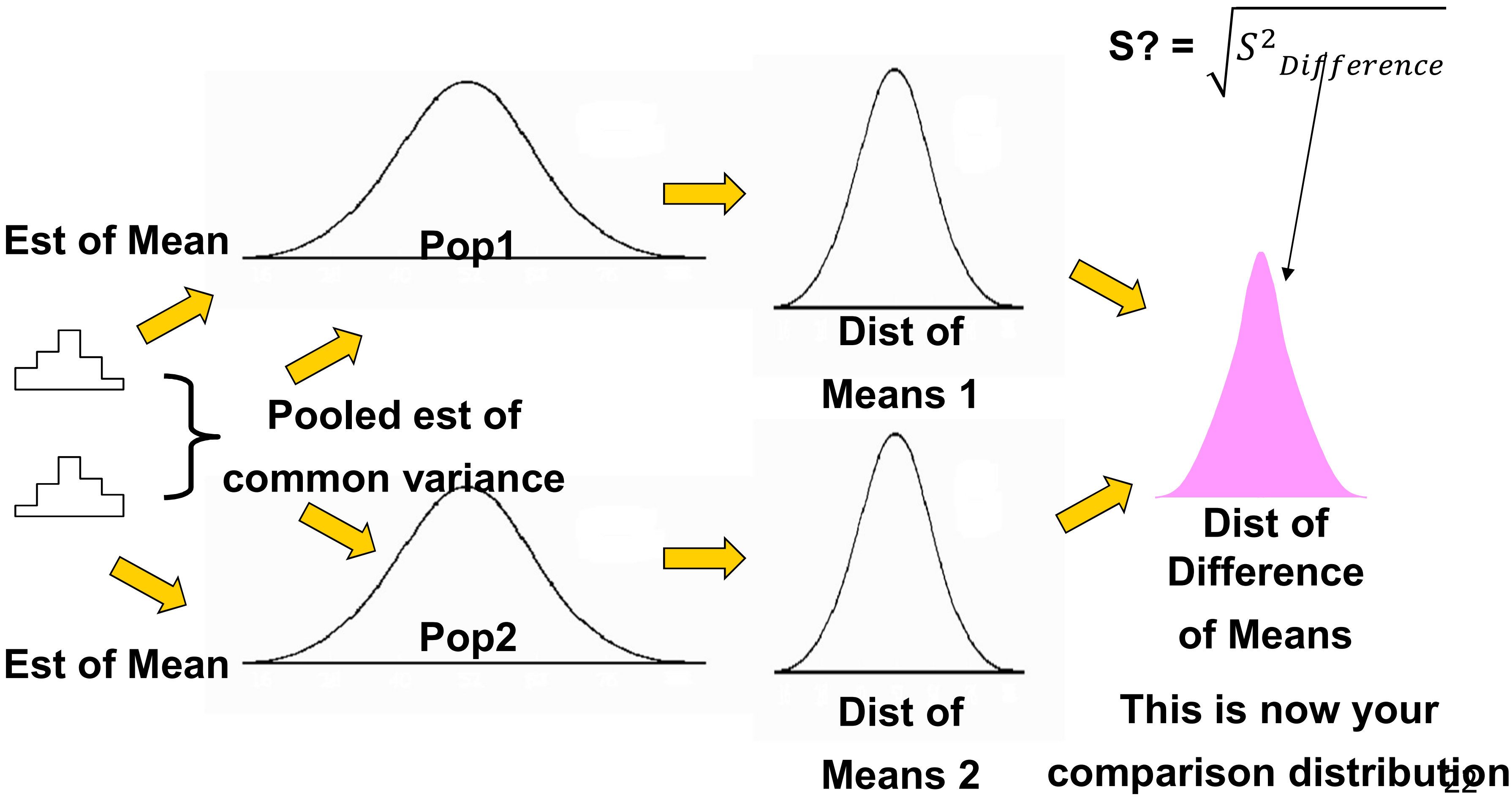
Only remaining loose end…



$$t = \frac{M_1 - M_2}{S_?}$$

Finally – the t test for independent samples

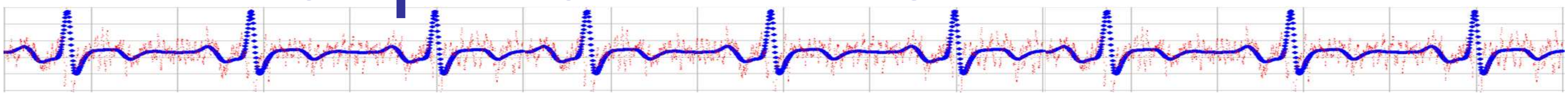




Assumptions of *t* test for independent means

- Population distributions normal (but holds up well UNLESS you have reason to think the two populations are dramatically skewed in opposite directions)
- Two populations have same variance (homogeneity of variance) (but holds up well even for fairly large differences; usually assumed) (tested via simulations)

Assumptions of *t* test for independent means



- Scores in samples suggest:
  - Populations approximately normal
  - Variances approximately same

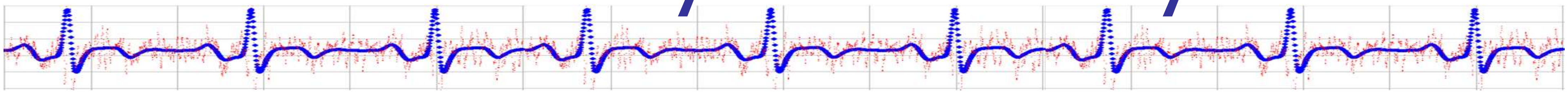
  (otherwise, there are alternative methods)
- Scores are independent of each other (otherwise need multilevel modeling)

E.g., clusters of people in one sample who might have similarities not found in other sample (e.g., seeing the same therapist in a study evaluating therapy interventions)

t tests in studies

- Suppose you do a large number of *t* tests for the same study (e.g., compare groups on each of 17 measures)
- Chance of any one of them coming out significant is greater than 5%
- Everyone agrees if only a few significant results, interpret cautiously

Controversy: Too many *t* tests



- What is “only a few”?
- For 100 comparisons, would have 5 potential false positives. Is that a few?
- PROBLEM: Comparisons are not independent (so might get clusters of significant results)
  (e.g., people who differ on assertiveness might also differ on self-confidence)

Controversy: Too many *t* tests

- Solution: Active area of research
- Relates to: p-hacking and “replication crisis” in psychology
- Some methods (e.g., Bonferroni procedure)
  - Use a more stringent significance level for each test
  - General idea: divide desired significance by number of tests

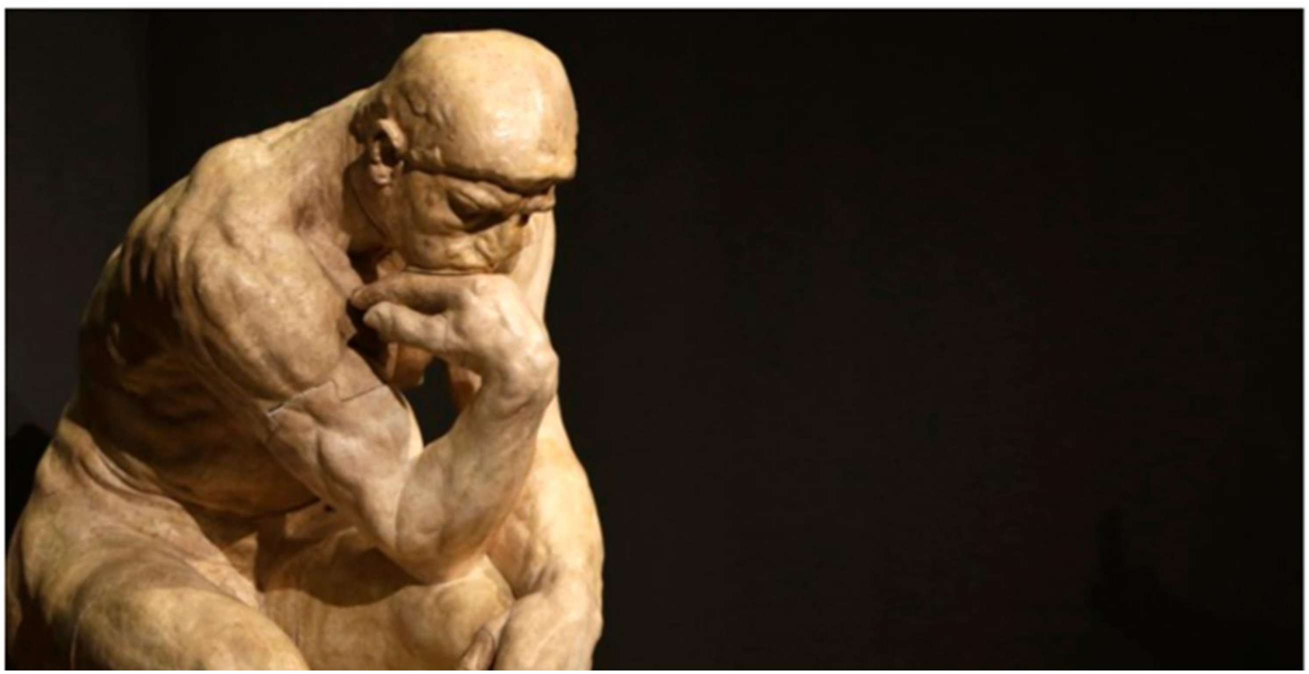
The Atlantic | Popular | Latest | Sections | Magazine | More | Subscribe

SCIENCE

Psychology's Replication Crisis Is Running Out of Excuses

Another big project has found that only half of studies can be repeated. And this time, the usual explanations fall flat.

ED YONG NOV 19, 2018



MORE STORIES

Science Is Getting Less Bang for Its Buck
PATRICK COLLISON AND MICHAEL NIELSEN

Why Rich Kids Are So Good at the Marshmallow Test
JESSICA MCCRORY CALARCO

Psychology's Replication Crisis Can't Be Wished Away
ED YONG

PLOS | BIOLOGY
FIFTEENTH ANNIVERSARY

BROWSE PUBLISH ABOUT SEARCH

advanced search

OPEN ACCESS

PERSPECTIVE

The Extent and Consequences of P-Hacking in Science

Megan L. Head, Luke Holman, Rob Lanfear, Andrew T. Kahn, Michael D. Jennions

Published: March 13, 2015 • https://doi.org/10.1371/journal.pbio.1002106

Article | Authors | Metrics | Comments | Media Coverage

Download PDF | Print | Share

Abstract

A focus on novel, confirmatory, and statistically significant results leads to substantial bias in the scientific literature. One type of bias, known as "p-hacking," occurs when researchers collect or select data or statistical analyses until nonsignificant results become significant. Here, we use text-mining to demonstrate that p-hacking is widespread throughout science. We then illustrate how one can test for p-hacking when performing a meta-analysis and show that, while p-hacking is probably common, its effect seems to be weak relative to the real effect sizes being measured. This result suggests that p-hacking probably does not drastically alter scientific consensuses drawn from meta-analyses.

Check for updates

Included in the Following Collection

Meta-Research: Sources of Bias

Example reporting

- “The mean level of self-reported health in the expressive writing group was 79.00 (SD=9.72), and the mean for the control writing group was 68.00 (SD=10.55); $t(18) = 2.42$, $p<.05$, two-tailed.

How many participants?

Example reporting

- "Participants in the current sample reported spending an average of 39.93 min on Facebook each day (SD=32.13) and had between 25 and 1,000 Facebook friends (M = 296.19, SD = 173.04). […] Women, M = 40.57, SD = 26.76, in our sample spent significantly more time on Facebook than men, M=29.83, SD=23.73; $t(305)=-3.32$, $p<0.01$, and women, M=3.29, SD=1.24, score significantly higher on Facebook jealously than men, M=2.81, SD=1.09; $t(305)=-3.32$, $p<0.01$."

Table format

Variable	Owners	Nonowners	$t(215)$
Well-being measures			
Depression	30.00	31.72	1.29
Loneliness	38.64	41.64	1.79†
Self-esteem	34.27	32.21	2.59*
Physical illnesses and symptoms	3.98	4.21	0.45
Subjective happiness	5.20	5.06	0.66
Exercise and fitness	4.40	3.94	2.64**
Personality factors			
Openness	4.06	3.98	0.81
Agreeableness	3.98	3.88	1.02
Conscientiousness	4.03	3.68	3.52**
Extraversion	3.52	3.25	2.13*
Neuroticism	2.16	2.23	0.67
Attachment style endorsement			
Secure	4.62	4.50	0.40
Fearful	3.16	3.72	1.77†
Preoccupied	2.53	3.06	1.90†
Dismissing	3.64	3.10	2.64**

† $p < .08$. * $p < .05$. ** $p < .01$.

Source: McConnell, A. R., Brown, C. M., Shoda, T. M., Stayton, L. E., & Martin, C. E. (2011). Friends with benefits: On the positive consequences of pet ownership. *Journal of Personality and Social Psychology, 101*, 1239–1252. Reproduced by permission of the American Psychological Association.