PSY 201: Statistics in Psychology

Lecture 30

Hypothesis testing for two correlations

How careful are students?

Greg Francis

Purdue University

Fall 2023

CORRELATIONS

- you have two independent populations (tests with dependent samples are more complicated)
- each with two scores for which you can calculate a correlation coefficient. e.g.
  - male population of students
  - female population of students
- might want to compare correlations between verbal SAT and math SAT

HYPOTHESIS

- the null hypothesis to test is

$$H_0 : \rho_1 = \rho_2$$

- where ρ_1 is the correlation coefficient of scores in population 1
- and ρ_2 is the correlation coefficient of scores in population 2
- and the alternative hypothesis is

$$H_a : \rho_1 \neq \rho_2$$

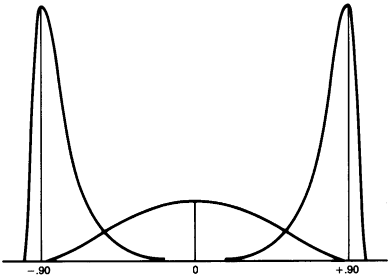
- or the same thing is:

$$H_0 : \rho_1 - \rho_2 = 0$$

$$H_a : \rho_1 - \rho_2 \neq 0$$

SAMPLING DISTRIBUTION

- to test H_0 we need to know the sampling distribution of the difference of correlation coefficients
- unfortunately, just like for the one-sample case, the sampling distribution changes as ρ changes



FISHER z TRANSFORMATION

$$z_r = \frac{1}{2} \log_e \left(\frac{1+r}{1-r} \right)$$

- which can be found by using the calculator in the textbook
- the sampling distribution of difference of z_r values is approximately normal and has a mean of

$$z_{\rho_1} - z_{\rho_2}$$

- which is zero, if H_0 is true

SAMPLING DISTRIBUTION

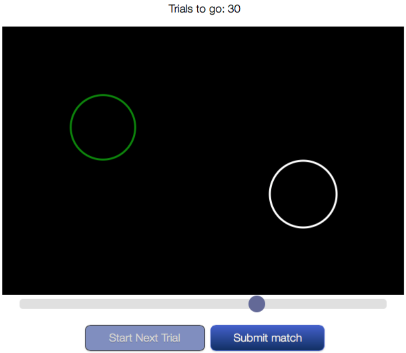
- so we know that the sampling distribution of the $z_{r_1} - z_{r_2}$ values is normally distributed and (if H_0 is true) has a mean of zero.
- all we need to know is the standard error of the difference between independent transformed correlation coefficients

$$s_{z_{r_1}-z_{r_2}} = \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}$$

- (note, you need more than three scores in each group)
- We just apply the same hypothesis testing approach as for other cases!

EXAMPLE

- In some tasks, correlation can be a measure of consistency or carefulness
- For example, in the STATLAB Weber's Law experiment, subjects adjust the size of a circle so it matches a target. There were two different target sizes, and we expect these matches to be correlated across subjects.



EXAMPLE

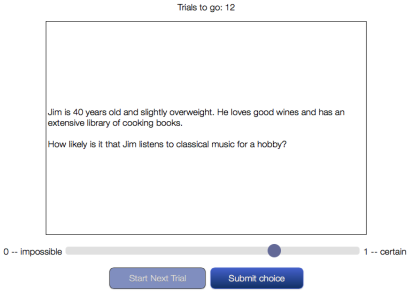
- Using the STATLAB data, we find that for the $n_1 = 30$ subjects who completed the Weber's Law lab, the correlation in matching sizes for the 10 pixel and 50 pixel targets is

$$r_1 = 0.1516$$

- In part this correlation reflects carefulness by the subject. If subjects are careless in their judgments, then they essentially add noise to their matching circles, and this will reduce the correlation

EXAMPLE

- In some tasks, correlation can be a measure of consistency or carefulness
- For example, in the STATLAB Typical Reasoning experiment, subjects rate the likelihood of certain characteristics of a described person. The descriptions were set up in a systematic way, so that some descriptions were expected to produce high ratings and other descriptions were expected to produce low ratings.



EXAMPLE

- Using the STATLAB data, we find that for the $n_2 = 29$ subjects who completed the Typical Reasoning lab, the correlation in likelihood ratings for the *Low typicality and two activities* and the *High typicality and two activities* is

$$r_2 = 0.1836$$

- In part this correlation reflects carefulness by the subject. If subjects are careless in their judgments, then they essentially add noise to their ratings, and this will reduce the correlation

EXAMPLE

- Are the correlations similar across the two tasks? They might seem like very different tasks, but to some extent, the correlations measure “effort” or “consistency” by the subjects.
- The overall strength of the correlation is less interesting than the similarity of the correlations. Some tasks may involve rather a lot of variability, so the correlation cannot be very large. Still, we can compare across tasks.
- Just looking at the correlations would not let us draw strong conclusions, but it could be part of a bigger argument.

EXAMPLE

- We use the on-line calculator to do the computations:

Sample correlation for group 1, $r_1 =$ 0.1516

Sample size for group 1, $n_1 =$ 30

Sample correlation for group 2, $r_2 =$ 0.1836

Sample size for group 2, $n_2 =$ 29

Specify hypotheses:

$H_0 : \rho_1 - \rho_2 =$ 0

H_a : Two-tails

$\alpha =$ 0.05

Run Test

Test summary	
Null hypothesis	$H_0 : \rho_1 - \rho_2 = 0$
Difference of null Fisher z transforms	$z_{\rho_1} - z_{\rho_2} = 0.0000$
Alternative hypothesis	$H_a : \rho_1 - \rho_2 \neq 0$
Type I error rate	$\alpha = 0.05$
Label for group 1	
Sample size for group 1	$n_1 = 30$
Sample correlation for group 1	$r_1 = 0.1516$
Fisher z transform of r_1	$z_{r_1} = 0.1528$
Label for group 2	
Sample size for group 2	$n_2 = 29$
Sample correlation for group 2	$r_2 = 0.1836$
Fisher z transform of r_2	$z_{r_2} = 0.1857$
Sample standard error	$s_{z_r} = 0.274770$
Test statistic	$z = -0.119839$
p value	$p = 0.904611$
Decision	Do not the reject null hypothesis

MISSING?

- It is possible to compute a confidence interval for a difference of independent correlations.
- It is also possible to compute a hypothesis test for a difference of dependent correlations.
- However, these methods require some new ideas that we do not have time to go into (lots of special cases).

POWER

- Computing power for a test of independent correlations is conceptually similar to other power calculations
- however, the use of the Fisher z transform of correlations makes it difficult to have good intuition into how sample size relates to power
- Since we take the Fisher z transform of each correlation and then take the difference, a fixed difference of correlations does not necessarily produce a fixed difference of Fisher z transform values.

POWER

- For example,

$$r_1 - r_2 = 0.3 - 0.2 = 0.1$$

- corresponds to

$$z_{r_1} - z_{r_2} = 0.203 - 0.100 = 0.103$$

- but

$$r_1 - r_2 = 0.5 - 0.4 = 0.1$$

- corresponds to a larger value:

$$z_{r_1} - z_{r_2} = 0.549 - 0.424 = 0.125$$

POWER

- these differences mean that testing for a specific alternative

$$H_a : \rho_1 - \rho_2 = 0.5 - 0.4$$

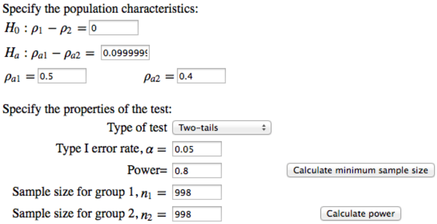
- is easier than testing for

$$H_a : \rho_1 - \rho_2 = 0.2 - 0.1$$

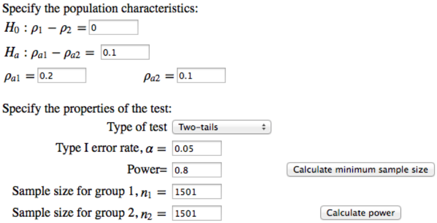
- Suppose you wanted 80% power for each test

POWER

$$H_a : \rho_1 - \rho_2 = 0.5 - 0.4$$



$$H_a : \rho_1 - \rho_2 = 0.2 - 0.1$$



- a difference of 1006 subjects across the two groups!

POWER

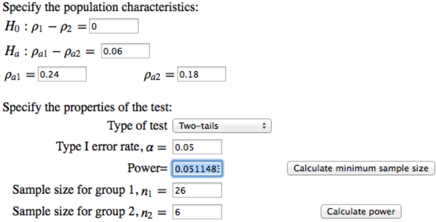
- Let's look at a specific example
- Height is correlated with economic success (income, wealth). Taller people are more successful
- Combining data across multiple studies suggests a difference in the correlation for men compared to women
  - Men: $r_1 = 0.24$
  - Women: $r_2 = 0.18$

POWER

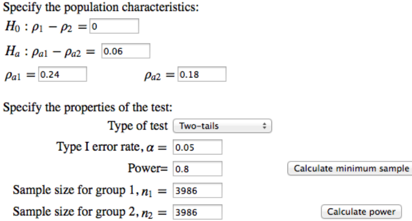
- Suppose you want to test this difference in correlations for Purdue engineering technology graduates. You will look at starting salaries and height.
- Engineering Technology degrees are given by Purdue Polytechnic, and each year it typically distributes BS degrees to 18 women and 78 men.
- You think you can get starting salary and height data for a third of the graduates. If the Purdue graduates are similar to the general population, what is the power of your study?

POWER

- there's
  no point in running such a study!



- You need a total of nearly 8000 subjects to have 80% power



- You simply cannot test for these kinds of small differences in correlations for relatively small populations (like Purdue graduates).

CONCLUSIONS

- two-sample case for correlations
- power
- trust the numbers!

NEXT TIME

- More than two comparisons
- Multiple testing

Error is sneaky.