PSY 201: Statistics in Psychology

Lecture 27

Hypothesis testing for dependent sample means

Within is better than between.

Greg Francis

Purdue University

Fall 2023

DEPENDENT SAMPLES

- two samples of data are dependent when each score in one sample is paired with a specific score in the other sample
- e.g
  - testing the same set of subjects before and after treatment
  - matched subjects in two groups (match along IQ before treatment and test mathematics)

CORRELATED DATA

- when samples are dependent, the scores across samples may be correlated
- suggests that you can (partly) predict one from other
- removes some of the randomness from the samples
- generally a good thing (more control over variables)
- but requires slightly different analysis

VARIABLE

- the variable of interest for dependent groups is the difference scores

$$d_i = X_{1i} - X_{2i}$$

- where
  - the ith scores in each group are matched (same subject)
  - X_{1i} is the ith score in the first group
  - X_{2i} is the ith score in the second group
- note that for dependent groups $n_1 = n_2 = n$, so we can calculate n difference scores

HYPOTHESIS

- we can calculate the mean of the difference scores for the sample

$$\overline{d} = \frac{\Sigma d_i}{n}$$

- which is the same as

$$\overline{d} = \overline{X}_1 - \overline{X}_2$$

- we would like to know if the mean of difference scores for the **population** $(\mu_1 - \mu_2)$ is different from zero

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

- This is actually the same as a one-sample t test!

STANDARD ERROR

- we estimate the standard error with

$$s_{\overline{d}} = \frac{s_d}{\sqrt{n}}$$

- where s_d is the standard deviation of the difference scores

$$s_d = \sqrt{\frac{\Sigma\left(d_i - \overline{d}\right)^2}{n-1}}$$

- or in raw score form

$$s_d = \sqrt{\frac{\Sigma d_i^2 - \left(\Sigma d_i\right)^2 / n}{n-1}}$$

TEST STATISTIC

- the test statistic is the same in form as all those before

$$\text{Test statistic} = \frac{\text{Statistic} - \text{Parameter}}{\text{Standard Error}}$$

- for our specific situation it is

$$t = \frac{\overline{d} - (\mu_1 - \mu_2)}{s_{\overline{d}}}$$

- which is used to compute a p-value in a t-distribution with $n - 1$ degrees of freedom

EXAMPLE

- do your thoughts control your autonomic processes?
- relax and take your pulse for 30 seconds
  - write the number down (X_1)
- picture yourself running and then take your pulse for 30 seconds
  - write the number down (X_2)
- we want to know if the mean difference across the two measurements (samples) is different from zero

EXAMPLE

- the measurements are dependent because if you tend to have a high pulse rate, it will be high for both measurements
- but we are interested in the **difference**, so the overall rate is unimportant
- calculate the difference of your pulse rates

$$d_i = X_{1i} - X_{2i}$$

HYPOTHESIS

- (1) our null hypothesis is that there is no effect of imagination on pulse rate

$$H_0 : \mu_1 - \mu_2 = 0$$

- the alternative hypothesis is that there is an effect

$$H_a : \mu_1 - \mu_2 < 0$$

- note, this is a directional hypothesis because we suspect that thinking about exercise should **increase** heart rate
- we will use a level of significance of $\alpha = 0.05$

DATA

- take a sample of pulse rate differences from ten people
- with your sampled data calculate the sample mean

$$\overline{d} = \frac{\sum d_i}{10}$$

- and the sample standard deviation (you can use the on-line calculator)

$$s_d = \sqrt{\frac{\Sigma d_i^2 - \left(\Sigma d_i\right)^2 / 10}{9}}$$

- and the estimate of the standard error

$$s_{\overline{d}} = \frac{s_d}{\sqrt{10}}$$

TEST STATISTIC

- (2) now calculate the test statistic as

$$t = \frac{\overline{d} - (\mu_1 - \mu_2)}{s_{\overline{d}}}$$

- since we assume $\mu_1 - \mu_2 = 0$ this is

$$t = \frac{\overline{d}}{s_{\overline{d}}}$$

p VALUE

- (3) You can compute the corresponding p-value with the t-Distribution Calculator for a one-tailed test
- with your sample of 10 people you have

$$df = n - 1 = 9$$

- if

$$p < \alpha = 0.05$$

- you reject H_0

DECISION

- (4) if you reject H_0 that means there is evidence that imagination of exercise **does** affect heart rate
- if you do not reject H_0 that means there is no evidence that imagination of exercise affects heart rate
- if you reject, that means that if $\mu_1 - \mu_2 = 0$, then the probability of the observed (or a more extreme) sample mean $\overline{d}$ value is less than 0.05

SIGNIFICANCE VS. IMPORTANCE

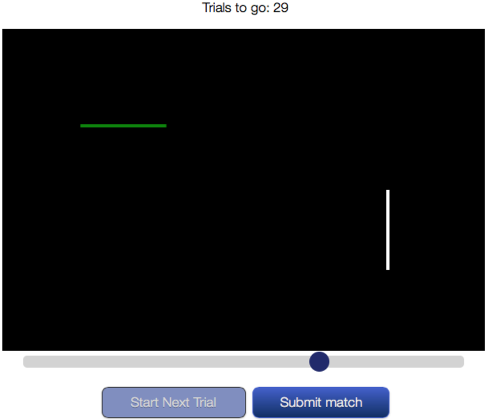
- if you failed to reject H_0, it may have been because you had too small a sample, n,
- or may have been because there was no real difference
- in principle, it is hard to believe that imagined running has **no effect at all** on pulse rate
  - Surely the brain uses energy differently during imagined running compared to not
- the effect might be very small, so small that our experiment cannot find it

SIGNIFICANCE VS. IMPORTANCE

- on the other hand
- probably everyone had a sample difference $\overline{d}$ that was non-zero
- but some people probably did not reject H_0
- we cannot just look at numbers like $\overline{d}$ and take them at face value
- the statistical procedures keep us from rushing to conclusions that are unwarranted

POWER

- The computation of power is very similar (actually, identical) to the one-sample t-test situation
- Consider the STATLAB Horizontal-Vertical illusion. Across the class, we have data from 26 students that reports the mean (for each student) matching length for a horizontal and a vertical line.



POWER

- Suppose you want to test for a difference between matching lengths for a horizontal and vertical line.

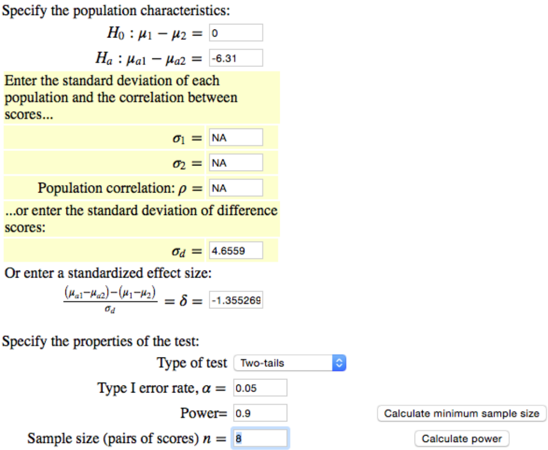
$$H_0 : \mu_1 - \mu_2 = 0$$

- How many subjects should you use to have 90% power?
- We can use the STATLAB data to estimate power and then compute an appropriate sample size
- From the STATLAB data we find:

$$\overline{X}_1 - \overline{X}_2 = 99.3213 - 105.6313 = -6.31$$

$$s_d = 4.655993$$

ONLINE CALCULATOR



Specify the population characteristics:

$H_0 : \mu_1 - \mu_2 =$ 0

$H_a : \mu_{a1} - \mu_{a2} =$ -6.31

Enter the standard deviation of each population and the correlation between scores...

$\sigma_1 =$ NA

$\sigma_2 =$ NA

Population correlation: $\rho =$ NA

...or enter the standard deviation of difference scores:

$\sigma_d =$ 4.6559

Or enter a standardized effect size:

$\frac{(\mu_{a1}-\mu_{a2})-(\mu_1-\mu_2)}{\sigma_d} = \delta =$ -1.355269

Specify the properties of the test:

Type of test: Two-tails

Type I error rate, $\alpha =$ 0.05

Power= 0.9

Sample size (pairs of scores) $n =$ 8

Calculate minimum sample size

Calculate power

ONLINE CALCULATOR

- You get exactly the same numbers using the one-sample calculator:

Specify the population characteristics:

$H_0 : \mu_0 =$ 0

$H_a : \mu_a =$ -6.31

$\sigma =$ 4.6560

Or enter a standardized effect size

$\frac{\mu_a - \mu_0}{\sigma} = \delta =$ -1.355240

Specify the properties of the test:

Type of test: Two-tails

Type I error rate, $\alpha =$ 0.05

Power= 0.9 [Calculate minimum sample size]

Sample size, $n =$ 8 [Calculate power]

ONLINE CALCULATOR

- Instead of using s_d (the standard deviation of the differences), you could use the standard deviation of each group and the correlation between scores:

Specify the population characteristics:

$H_0 : \mu_1 - \mu_2 =$ 0

$H_a : \mu_{a1} - \mu_{a2} =$ -6.31

Enter the standard deviation of each population and the correlation between scores...

$\sigma_1 =$ 3.5502

$\sigma_2 =$ 6.4328

Population correlation: $\rho =$ 0.7073

...or enter the standard deviation of difference scores:

$\sigma_d =$ 4.656027(

Or enter a standardized effect size:

$\frac{(\mu_{a1}-\mu_{a2})-(\mu_1-\mu_2)}{\sigma_d} = \delta =$ -1.355232

Specify the properties of the test:

Type of test: Two-tails

Type I error rate, $\alpha =$ 0.05

Power= 0.9

Sample size (pairs of scores) $n =$ 8

Calculate minimum sample size

Calculate power

INDEPENDENT TEST

- Suppose you wanted to do the experiment with different subjects assigned to different line orientations. How many subjects do you need?
- Independent Means Power Calculator

Specify the population characteristics:

$H_0 : \mu_1 - \mu_2 =$ 0

$H_a : \mu_{a1} - \mu_{a2} =$ -6.31

$\sigma_1 =$ 3.5502

$\sigma_2 =$ 6.4328

Or enter a standardized effect size

$\frac{(\mu_{a1}-\mu_{a2})-(\mu_1-\mu_2)}{\sigma} = \delta =$ NA

Specify the properties of the test:

Type of test: Two-tails

Type I error rate, $\alpha =$ 0.05

Power= 0.9 [Calculate minimum sample size]

Sample size for group 1, $n_1 =$ 16

Sample size for group 2, $n_2 =$ 16 [Calculate power]

- 4 times as many subjects for an independent means experiment!

WITHIN vs. BETWEEN

- A within subjects (dependent test) design is usually more powerful than a between subjects (independent test)
- This is because you are able to remove one source of variability from the standard error
  - The variability in overall score values
- Standard error reflects variability between conditions
- A between subjects calculation of variability includes variability between conditions and variability across subjects (more variability!)

CONCLUSIONS

- dependent samples
- very important for lots of interesting tests
- more powerful than independent tests

NEXT TIME

- two-sample case for independent proportions
- hypothesis testing
- confidence interval
- power

What is a "margin of error"?