UTAustinX: UT.7.01x Foundations of Data Analysis

KarenWest ▾

Courseware Course Info Discussion Syllabus Download R and RStudio R Tutorials Readings Contact Us

Progress Office Hours Community

PAIRED SAMPLES T-TEST

0:00 / 10:02 1.0x

There are times when comparing between two independent groups

might not be the best test.

Instead, it might make more sense to test the same groups of subjects

twice - once before some intervention or event, and then again after.

This is typically called a pre/post design,

and can not only answer some interesting questions,

Help

but uses a new (yet very familiar form) of the t-test.

I recently ran across a study that was interested in the idea of pain

reduction after surgery for children who received tonsillectomies.

This was prompted by the recent banning of codeine elixir

by the US FDA because of recognized possibilities of death.

The study observed 42 children who had recently undergone a tonsillectomy.

A day after surgery, children were asked if they were in any pain

and asked to point to a position on a Faces Pain Scale

that represented their pain.

If there were in pain, they were asked (with their parents)

if they wanted to receive acupuncture to help alleviate the pain.

31 participants opted to receive acupuncture

and were asked their pain score immediately

following the acupuncture procedure.

Let's look at some simulated data based on the study.

Notice, each row of data represents one

Help

subject,

and each subject has two scores: one immediately

before acupuncture (the pre score) and one immediately after (the post score).

Let's look at these two columns with a plot

- and include the Standard Error if we calculate each column's Standard

Error separately.

What do you see?

I see overlap of the 95% confidence intervals.

I don't see significance.

If we were to run an independent samples t-test, we'd be wrong on two counts.

First, we'd violate the assumption of independence;

the pre- and the post-columns are related to one another

through the subject.

Second, we wouldn't be using the correct errors.

And here's why.

If we use the errors in the graph, the Standard Error for each column,

we're effectively saying that the pre-measure

is unrelated to the post-measure.

Help

We could, if we want to, sort the columns independently of one another,

in any order we want, and still end up with the SAME mean and Standard

Deviation and therefore Standard Error for that particulate column.

BUT, we would be breaking the relationship between the two columns

- the fact that pre and post are related to one another through the subject.

So how do we capture this idea of related-ness

and come up the correct test?

We turn to difference scores.

Since the two measures of data are related to one another

though the subject, we can subtract one column from the other

and come up with a change score.

This SINGLE value now represents the change from pre

to post for all subjects.

If we take the average of this value, we come up with an AVERAGE of the change

- a single sample mean.

Does that sound familiar?

What if we wanted to know if that mean difference was significant?

Help

Well, if the treatment had no effect, then

the comparison value, the average of the difference, should be - that's right

- zero.

See where this is going?

This is basically a single sample t-test -

just using the DIFFERENCE scores as our single sample.

Let's formalize this: If we want to know if two means are different

form one another, and those two mean are NOT independent,

then we MUST use what's called the paired samples t-test

or dependent samples t-test.

We'll get into the process in a little bit,

but again, let's start with the assumptions for this particular test:

The sample of differences is random; Each difference score

is independent from another (SUBJECTS are independent);

and the distribution of differences in the population is normal.

Again, it's not surprising that these assumptions sound a lot like the single

sample t-test - just instead of a looking at a single raw score,

Help

we're looking at a difference between to related raw scores.

But it's still just one sample of values.

Moving on to our hypothesis test.

Step one: Our Null and Alternative Hypothesis statements.

Our Null Hypothesis states that the difference between the two

related score will be some value - usually zero.

Here, instead of mu, we have the greek symbol for delta

- representing a DIFFERENCE score.

Our Alternative Hypothesis is that the difference will not be equal to zero.

Again, this is a non-directional test (we

don't know what the effect of acupuncture

will have on the perceived pain for the subjects).

Step two: decide on our alpha level.

0.05 again is the standard.

Step three: collect the data and run the analysis.

Now, as promised, here's the full form of the paired samples t-test.

Again, notice how it looks a lot like the single sample t-test -

just instead of referencing a single sample

Help

of raw values,

it's referencing a single sample of differences.

The numerator compares the difference from our data,

to some Null Hypothesis value - usually zero.

The denominator is another Standard Error - this time the "estimated

Standard Error of mean difference."

Again, it's similar to the single sample t-test estimated Standard Error

of the mean - but this time it uses the Standard Deviation of the difference

score, with N being the number of difference values.

Making our substitutions, and solving for t, we get a t-statistic of -7.99.

Step four: we draw our conclusion.

Is this t-statistic one that indicates a difference between the two samples?

To answer that we again turn to our t-distribution,

and our degrees of freedom.

It should be no surprise that the paired samples

t-test uses the same idea of degrees of freedom

Help

as the single sample t-test of n-1, with n here

being the number of difference values in the data.

Now, using our t-distribution, 30 degrees of freedom gives us a critical

value of +/-2.04.

Since our t-statistic is outside this critical value,

and into the critical region, we Reject our Null Hypothesis:

the difference between pre- and post-pain scores are not in fact

equal to zero.

Examining the means of the two scores, we

see that the pain significantly decreases after acupuncture.

If we run this in software using the t.test() in R,

we see that we have a paired t-test p-value of less than 0.001 -

lower than our alpha level of 0.05, so again, we Reject our Null Hypothesis.

If we want to visualize, the dependent samples

should be represented by a line graph to indicate the relationship between pre

and post, with a 95% Confidence Interval using

Help

the method proposed by Morey in 2005.

And here's another cool thing about the paired samples t-test:

it isn'y JUST for pre/post data.

Although pre/post data is its primary use,

it can also be used to study twins.

You may have recently heard of Scott and Mark Kelly.

Now they're astronauts that are also identical twins.

In 2015, one will go to space for a year, and the other will stay on earth.

If NASA had more twin pairs, then each row would be an identical twin set,

and the two columns would represent each twin - one for space and one for earth.

The paired samples t-test can come into play

because we can make the argument that the twins are identical,

and thus the genetic variation is the same for each twin within the pair.

So, learn statistics, maybe help an astronaut.

Help

1. For each scenario below, identify whether the groups are independent or dependent:

(1 point possible)

1a. A researcher wants to know if reaction time is the same for men and women. She measures the reaction time of 30 female and 30 male drivers.

Independent

Hide Answer

(1 point possible)

1b. A researcher wants to know if reaction time changes after drinking an alcoholic beverage. He measures the reaction time of 40 middle-aged drivers before and after consuming 8 ounces of beer.

Dependent

Hide Answer

(1 point possible)

1c. A researcher wants to know if students that take AP Calculus will score higher on their first college math exam than students who do not take AP Calculus. She compares exam scores on two groups of students, where the only difference between groups is whether they took AP Calculus or not.

Independent

Help

Hide Answer

2. Does chewing gum make you less accurate while target shooting? Each of the following subjects shot at a target (to earn a maximum score of 100) in random order, once while chewing gum and once while not chewing. Assume all test conditions are met.

Participant	Chewing Gum	No Gum
1	79	80
2	95	94
3	85	87
4	82	84

(1 point possible)

2a. What is the null hypothesis for this test? (difference score = accuracy while chewing - accuracy without gum)

- $\delta \neq 0$
- $\delta = 0$ ✔
- $\delta \leq 0$
- $\delta \geq 0$

Help

Hide Answer

(1 point possible)

2b. What is the average difference score for this sample?

- -1 ✔
- -2
- +1
- -5

Hide Answer

(1 point possible)

2c. What is the standard error for this test?

- 5.990
- 0.879
- 1.412
- 0.707 ✔

Help

Hide Answer

(1 point possible)

2d. What is the value of the t-statistic and the appropriate conclusion for the test, assuming $\alpha = 0.05$

- t = 2.871; Chewing gum did result in a decrease in accuracy.
- t = -1.414; Chewing gum did not result in a decrease in accuracy. ✔
- t = 1.837; Chewing gum did not result in a decrease in accuracy.
- t = 1.414; Chewing gum did not result in a decrease in accuracy.

Hide Answer

Help

EdX offers interactive online classes and MOOCs from the world's best universities. Online courses from MITx, HarvardX, BerkeleyX, UTx and many other universities. Topics include biology, business, chemistry, computer science, economics, finance, electronics, engineering, food and nutrition, history, humanities, law, literature, math, medicine, music, philosophy, physics, science, statistics and more. EdX is a non-profit online initiative created by founding partners Harvard and MIT.

Terms of Service and Honor Code

Privacy Policy (Revised 10/22/2014)

About edX

About

News

Contact

FAQ

edX Blog

Donate to edX

Jobs at edX

Follow Us

Twitter

Facebook

Meetup

LinkedIn

Google+