Hypothesis Tests: Two Related Samples

AKA Dependent Samples Tests
AKA Matched-Pairs Tests
Ψ320
Ainsworth

Major Points

- Related samples? Matched Samples?
- Difference scores?
- An example
- *t* tests on difference scores
- Advantages and disadvantages
- Effect size

Review: Hypothesis Testing

1. State Null Hypothesis
2. Alternative Hypothesis
3. Decide on α (usually .05)
4. Decide on type of test (distribution; ***z***, ***t***, etc.)
5. Find critical value & state decision rule
6. Calculate test
7. Apply decision rule

Related/Dependent Samples

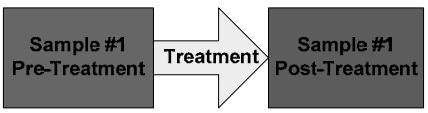
- Samples can be related for 2 basic reasons
- First, they are the same people in both samples
  - This is usually called either repeated measures or within subjects design

Related/Dependent Samples

- Samples can be related for 2 basic reasons
- Second, individuals in the two sample are so similar they are essentially the same person
  - Often called a matched-pairs design

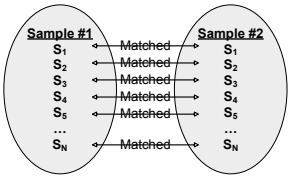
Related/Dependent Samples

- Repeated Measures
- The same participants give us data on two measures
  - e.g. Before and After treatment
  - IQ levels before IQPLUS, IQ levels after IQPLUS



Related/Dependent Samples

- Matched-Pairs Design
- Two-separate groups of participants; but each individual in sample 1 is matched (on aspects other than DV) with an individual in sample 2



Related/Dependent Samples

- With dependent samples, someone high on one measure is probably high on other.
- Scores in the two samples are highly correlated
  - Since they are correlated cannot treat them as independent (next chapter)
  - However the scores can be manipulated (e.g. find the differences between scores)

Difference Scores

- Calculate difference between first and second score
  - e. g. Difference = Before - After
- Base subsequent analysis on difference scores
  - Ignoring Before and After data

An Example

- Therapy for rape victims
  - Foa, Rothbaum, Riggs, & Murdock (1991)
- One group received Supportive Counseling
- Measured post-traumatic stress disorder symptoms before and after therapy

Hypotheses?

- H_0: $\mu_{symptoms/before} \leq \mu_{symptoms/after}$
- H_1: $\mu_{symptoms/before} > \mu_{symptoms/after}$

OR

- H_0: $\mu_{symptoms/before} - \mu_{symptoms/after} \leq 0$
- H_1: $\mu_{symptoms/before} - \mu_{symptoms/after} > 0$

OR

- H_0: $\mu_{(symptoms/before - symptoms/after)} \leq 0$
- H_1: $\mu_{(symptoms/before - symptoms/after)} > 0$

Supportive Therapy for PTSD

Person	Pre - Treatment	Post - Treatment
1	21	15
2	24	15
3	21	17
4	26	20
5	32	17
6	27	20
7	21	8
8	25	19
9	18	10
Mean	23.889	15.667
SD	4.197	4.243

Supportive Therapy for PTSD

- We want to compare the means to see if the mean after is significantly larger than the mean before
- However, we can't perform the test this way (reasons I'll explain in the next chapter)
- Since scores in the 2 conditions come from the same people we can use that to our advantage (subtract post from pre)

Calculating a difference score

Person	Pre - Treatment	Post - Treatment	Difference (Pre - Post)
1	21	15	6
2	24	15	9
3	21	17	4
4	26	20	6
5	32	17	15
6	27	20	7
7	21	8	13
8	25	19	6
9	18	10	8
Mean	23.889	15.667	8.222
SD	4.197	4.243	3.598

Supportive Therapy for PTSD

Person	Difference (Pre - Post)
1	6
2	9
3	4
4	6
5	15
6	7
7	13
8	6
9	8
Mean	8.222
SD	3.598

We now have a single sample problem identical to chapter 12.

These are change scores for each person.

Results

- The Supportive Counseling group decreased number of symptoms
- Was this enough of a change to be significant?
- Before and After scores are not independent.
  - See raw data (subjects high stayed high, etc.)
  - *r* = .637 (obviously dependent samples)

16

Results

- If no change, mean of differences should be zero
  - So, test the obtained mean of *difference* scores (we'll call *D*) against $\mu = 0$.
  - Then, use same test as in Chapter 12.
- We don't know σ, so use *s* and solve for *t*

17

t_D test

$\overline{D}$ and s_D are the mean and standard deviation of the difference scores.

Standard Error of D is $s_{\bar{D}} = s_D / \sqrt{n}$

$$t_D = \frac{\overline{D} - \mu}{s_{\bar{D}}} = \frac{8.222 - 0}{3.598/\sqrt{9}} = \frac{8.222}{1.199} = 6.86$$

- *df* = *n* - 1 = 9 - 1 = 8

18

t test

- 8 df, α = .05, 1-tailed $\rightarrow$ t_{crit} = 1.860
- We calculated *t* = 6.86
- Since 6.86 > 1.860, reject H_0
- Conclude that the mean number of symptoms after therapy was less than mean number before therapy.
- Supportive counseling seems to help reduce symptoms

SPSS Printout

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PRE	23.89	9	4.197	1.399
	POST	15.67	9	4.243	1.414

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	PRE & POST	9	.637	.065

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	PRE - POST	8.22	3.598	1.199	5.46	10.99	6.856	8	.000

Related/Dependent Samples

- Advantages
  - Eliminate subject-to-subject variability
  - Control for extraneous variables
  - Need fewer subjects
- Disadvantages
  - Order effects
  - Carry-over effects
  - Subjects no longer naive
  - Change may just be a function of time
  - Sometimes not logically possible

Effect Size Again

- We could simply report the difference in means.
  - Difference = 8.22
  - But the units of measurement have no particular meaning to us—Is 8.22 large?
- We could "scale" the difference by the size of the standard deviation.

Effect Size

$$d = \frac{\mu_1 - \mu_2}{\sigma} = \frac{\mu_{Before} - \mu_{After}}{\sigma_{Before}}$$

$$= \frac{23.89 - 15.67}{4.20} = \frac{8.22}{4.20} = 1.96$$

Note: This effect size *d* is not the same thing as *D* (difference)

It's called *d* here because it is in reference to Cohen's *d*

Effect Size

- The difference is approximately 2 standard deviations, which is very large.
- Why use standard deviation of Before scores?
- Notice that we substituted statistics for parameters.