Matched T-Test

Cohen Chapter 11

For EDUC/PSY 6600

"…we are suffering from a plethora of surmise, conjecture, and hypothesis. The difficulty is to detach the framework of fact – of absolute undeniable fact – from the embellishments of theorists and reporters."

Sherlock Holmes

Silver Blaze

MOTIVATING EXAMPLES

Dr. Filburn wishes to assess the effectiveness of a leadership workshop for 60 middle managers. The 60 managers are rated by their immediate supervisors on the Leadership Rating Form (LRF), ***before and after*** *the workshop.*

Dr. Clarke is interested in determining if workers are more concerned with job security or pay. He gains the cooperation of 30 individuals who work in ***different settings*** *and asks each employee to rate his or her concern about 1) salary level and 2) job security on a scale from 1 to 10.*

Dr. Gale questions whether husbands or wives with infertility problems feel equally anxious. She recruits 24 infertile couples and then administers the Infertility Anxiety Measure (IAM) to both ***the husbands and the wives.***

PAIRED-SAMPLES DESIGNS

- Comparing means of 2 groups
  - Assumption of independence has been violated resulting in a dependency across groups
    - E.g., Members of same family, class, group, litter, twinship
  - Variance of DV smaller as groups consist of same or closely matched cases
- Paired-samples *t*-test also known as…
  - Matched-, Related-, Correlated-, Dependent-, or Non-independent samples *t*-test
  - Repeated-measures *t*-test

Experimental

Matching groups on some variable(s)

E.g., sex, age, education

↓ potential confounds on IV-DV relationship or when cases cannot receive both conditions

Naturalistic

Samples naturally related, correlated, dependent

REPEATED-MEASURES DESIGNS

- **Successive designs:**

2 measurements, conditions, or sets of stimuli are applied to cases sequentially

- Before-and-after (or longitudinal) designs
  - Pre- / post-test, time 1 / time 2
- Cross-over designs
  - Order effects? Need to counterbalance order
    - Random subset of cases → A then B
    - Another random subset of cases → B then A
  - Counterbalancing may not eliminate carry-over effects
  - Wash-out period

Simultaneous designs:

2 varying conditions or sets of stimuli inter-mixed w/in study and all cases receive both

No concern for order effects or temporality

Order is generally random

HYPOTHESES: 'DIRECT DIFFERENCE' METHOD

- Same as Independent-samples *t*-test
  - H_0: $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$ or $\mu_1 - \mu_2 = 0$
  - H_1: $\mu_1 \neq \mu_2$ or $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$

- H_0: $\mu_1 - \mu_2 = 0$ → H_0: $\mu_D = 0$
  - **Compute difference score for each subject**
    - $X_{i1} - X_{i2} = D$
    - H_0: $\mu_D = 0$ and H_1: $\mu_D \neq 0$

- **Now equivalent to 1-sample *t*-test**
  - Mean of difference scores compared w/ H_0: $\mu_D = 0$

CALCULATIONS

Mean of difference scores

Hypothesized population difference

$$t = \frac{\overline{D} - \mu_D}{s_{\overline{D}}} = \frac{\overline{D} - 0}{\frac{s_D}{\sqrt{N}}}$$

df = *N* - 1

Number of difference scores (pairs) - 1

$$s_D = \sqrt{\frac{\sum_{i=1}^{n} (D_i - \overline{D})^2}{N - 1}}$$

= *SD* of difference scores

ASSUMPTIONS

1. Independence of pairs of observations

2. Normality of sampling distribution of difference scores in population

3. Equal *ns*
   - Pair deleted when 1 member missing data

PAIRED-SAMPLES *T*-TEST AND CORRELATION

- Paired-samples *t*-test almost always more powerful than independent-samples *t*-test
  - More likely to reject H_0 when false
  - Requires fewer subjects
- Degree of correlation (*r*) between scores on 2 groups related to size of difference between paired- and independent-samples *t*-statistics
  - Larger correlation → larger difference

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right) - \frac{2rs_1s_2}{n}}}$$

PAIRED-SAMPLES *T*-TEST AND CORRELATION

- Paired-samples *t*-test almost always more powerful than independent-samples *t*-test
  - More likely to reject H_0 when false
  - Requires fewer subjects
- Degree of correlation (r) between scores on 2 groups related to size of difference between paired- and independent-samples *t*-statistics
  - Larger correlation → larger difference

Paired-samples *t*-test calculated as a function of r

- When $r = 0$,
  - equation reduces to independent-samples *t*-test
- When $r > 0$
  - denominator reduces, leading to larger *t*-statistic
- When $r < 0$
  - denominator increases, leading to smaller *t*-statistic
  - *Rare to have a negative correlation with paired-data*

CONFIDENCE INTERVALS

95% *CI* around μ_D

Rewrite: $$t = \frac{\overline{D} - \mu_D}{\frac{s_D}{\sqrt{N}}}$$

As: $$CI_{1-\alpha} = \overline{D} \pm t_{\alpha/2} * \frac{s_D}{\sqrt{N}}$$

Are paired sample means significantly different?

Yes: H_0 value not w/in *CI*

No: H_0 value within *CI*

EXAMPLE

- Same example from independent-samples *t*-test lecture
- But suppose participants were carefully matched into pairs based on their level of depression prior to initiation of study
- One member of each pair was randomly assigned to drug group, other to placebo group
- After 6 months, level of depression was measured by a psychiatrist
- Need to conduct paired-samples *t*-test due to matching

Group 1 - Drug	Group 2 - Placebo
11	11
1	11
0	5
2	8
0	4

R CODE: FIRST APPROACH

```
df <- read.csv(“drug_paired.csv”)

## do some plots and summaries

df %>%
  tidyr::pivot_longer(cols = group1:group2) %>%
  t.test(value ~ group,
         data = .,
         paired = TRUE)
```

Get the data into R

Plot and summaries

Reshape the data into long form

Paired samples t-test

Group 1 - Drug	Group 2 - Placebo
11	11
1	11
0	5
2	8
0	4

R CODE: FIRST APPROACH

```
      Paired t-test


data:  value by group
t = -3.1009, df = 4, p-value = 0.03619
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9.4768832 -0.5231168
sample estimates:
mean of the differences 
                     -5 
```

R CODE: SECOND APPROACH

Group 1 - Drug	Group 2 - Placebo
11	11
1	11
0	5
2	8
0	4

```
df <- read.csv("drug_paired.csv")
```
Get the data into R

```
## do some plots and summaries
```
Plot and summaries

```
df %>%
  dplyr::mutate(group_diff = group2 – group1) %>%
  dplyr::pull(group_diff) %>%
  t.test(mu = 0)
```
Create group difference variable

Single samples t-test of the group difference

R CODE: SECOND APPROACH

```
        One Sample t-test

data:  .
t = 3.1009, df = 4, p-value = 0.03619
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.5231168 9.4768832
sample estimates:
mean of x
        5
```

EFFECT SIZE

- *__Cohen's *d*__ (same as in 1-sample *t*-test)

  - $d = \frac{\overline{D}}{s_D}$ or $\frac{t}{\sqrt{N}}$

- **Eta squared** (η^2)

  - $\eta^2 = \frac{N * \overline{D}^2}{N * \overline{D}^2 + (N-1) * s_{\overline{D}}^{\,2}}$ or $\frac{t^2}{t^2 + N - 1}$

POWER ANALYSIS

Post hoc

With Cohen's *d* estimate and # pairs, compute delta to obtain power of study

$$\delta = d\sqrt{\frac{N}{2}}$$

A Priori

With desired power, compute delta and combine with estimated Cohen's *d* to obtain # pairs (*N*)

$$N = \left(\frac{\delta}{d}\right)^2$$

	One-Tailed Test (α)			
	.05	.025	.01	.005
	Two-Tailed Test (α)			
δ	.10	.05	.02	.01
0.5	.14	.08	.03	.02
0.6	.16	.09	.04	.02
0.7	.18	.11	.05	.03
0.8	.21	.13	.06	.04
0.9	.23	.15	.08	.05
1.0	.26	.17	.09	.06
1.1	.29	.20	.11	.07
1.2	.33	.22	.13	.08
1.3	.37	.26	.15	.10
1.4	.40	.29	.18	.12
1.5	.44	.32	.20	.14
1.6	.48	.36	.23	.16
1.7	.52	.40	.27	.19
1.8	.56	.44	.30	.22
1.9	.60	.48	.33	.25
2.0	.64	.52	.37	.28
2.1	.68	.56	.41	.32
2.2	.71	.60	.45	.35
2.3	.74	.63	.49	.39
2.4	.77	.67	.53	.43
2.5	.80	.71	.57	.47
2.6	.83	.74	.61	.51
2.7	.85	.77	.65	.55
2.8	.88	.80	.68	.59
2.9	.90	.83	.72	.63

WEAKNESSES

- Reduction in *df* for critical value
- Lack of a control group (sometimes)
- If samples are not truly matched, results will be spurious

ALTERNATIVES

- Violation of normality
  - Matched-pairs Wilcoxon Test

- Binomial Sign Test for Two Dependent Samples

- Sample Re-use methods
  - Exact tests
  - Randomization and permutation tests