HUMBEHV 3HB3
two-sample t-tests & statistical power

Week 9
Prof. Patrick Bennett

Concepts from previous lectures
• t distribution
• standard error of the mean
• degrees-of-freedom
• Null and alternative/research hypotheses (H0 vs H1)

Hand et al (1994)
• Experimental Question:
  - Is family therapy an effective treatment for anorexia?
• 17 girls participated in study
  - weighed before & after treatment
  - weights (in pounds given in Table 13.1)
• Statistical Question:
  - Does before/after weight differ?

t-test for 2 matched/related/dependent samples
Hand et al (1994)

- N=17
- Subjects weighed before & after family therapy
- diff = after - before
- Before & after measures are not independent because they come from the same subject

<table>
<thead>
<tr>
<th>subject</th>
<th>before</th>
<th>after</th>
<th>diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83.8</td>
<td>95.2</td>
<td>11.4</td>
</tr>
<tr>
<td>2</td>
<td>83.3</td>
<td>94.3</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>86</td>
<td>91.5</td>
<td>5.5</td>
</tr>
<tr>
<td>4</td>
<td>82.5</td>
<td>91.9</td>
<td>9.4</td>
</tr>
<tr>
<td>5</td>
<td>86.7</td>
<td>100.3</td>
<td>13.6</td>
</tr>
<tr>
<td>6</td>
<td>79.6</td>
<td>76.7</td>
<td>-2.9</td>
</tr>
<tr>
<td>7</td>
<td>76.9</td>
<td>76.8</td>
<td>0.1</td>
</tr>
<tr>
<td>8</td>
<td>94.2</td>
<td>101.6</td>
<td>7.4</td>
</tr>
<tr>
<td>9</td>
<td>73.4</td>
<td>94.9</td>
<td>21.5</td>
</tr>
<tr>
<td>10</td>
<td>80.5</td>
<td>75.2</td>
<td>-5.3</td>
</tr>
<tr>
<td>11</td>
<td>81.6</td>
<td>77.6</td>
<td>-3.8</td>
</tr>
<tr>
<td>12</td>
<td>82.1</td>
<td>95.5</td>
<td>13.4</td>
</tr>
<tr>
<td>13</td>
<td>77.6</td>
<td>90.7</td>
<td>13.1</td>
</tr>
<tr>
<td>14</td>
<td>83.5</td>
<td>92.5</td>
<td>9.0</td>
</tr>
<tr>
<td>15</td>
<td>89.9</td>
<td>93.8</td>
<td>3.9</td>
</tr>
<tr>
<td>16</td>
<td>86</td>
<td>91.7</td>
<td>5.7</td>
</tr>
<tr>
<td>17</td>
<td>87.3</td>
<td>98</td>
<td>10.7</td>
</tr>
</tbody>
</table>

Hand et al (1994)

- Before & after measures are correlated - not surprising because measures were taken on same subjects at different times
- T tests depend, in part, on N (e.g., df)
  - Should N total be N before + N after?
  - No, because the 2 sets of measures are not independent
  - Our analysis must take dependence into account
- Simple solution: analyze difference scores

<table>
<thead>
<tr>
<th>Weight Before Therapy (lbs.)</th>
<th>Weight After Therapy (lbs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>75</td>
</tr>
<tr>
<td>80</td>
<td>85</td>
</tr>
<tr>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>100</td>
<td>105</td>
</tr>
</tbody>
</table>

$r = 0.54$
$r_s = 0.61$

Hand et al (1994)

- Diff = after - before

- Before & after measures are correlated

<table>
<thead>
<tr>
<th>subject</th>
<th>before</th>
<th>after</th>
<th>diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83.8</td>
<td>95.2</td>
<td>11.4</td>
</tr>
<tr>
<td>2</td>
<td>83.3</td>
<td>94.3</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>86</td>
<td>91.5</td>
<td>5.5</td>
</tr>
<tr>
<td>4</td>
<td>82.5</td>
<td>91.9</td>
<td>9.4</td>
</tr>
<tr>
<td>5</td>
<td>86.7</td>
<td>100.3</td>
<td>13.6</td>
</tr>
<tr>
<td>6</td>
<td>79.6</td>
<td>76.7</td>
<td>-2.9</td>
</tr>
<tr>
<td>7</td>
<td>76.9</td>
<td>76.8</td>
<td>-0.1</td>
</tr>
<tr>
<td>8</td>
<td>94.2</td>
<td>101.6</td>
<td>7.4</td>
</tr>
<tr>
<td>9</td>
<td>73.4</td>
<td>94.9</td>
<td>21.5</td>
</tr>
<tr>
<td>10</td>
<td>80.5</td>
<td>75.2</td>
<td>-5.3</td>
</tr>
<tr>
<td>11</td>
<td>81.6</td>
<td>77.6</td>
<td>-3.8</td>
</tr>
<tr>
<td>12</td>
<td>82.1</td>
<td>95.5</td>
<td>13.4</td>
</tr>
<tr>
<td>13</td>
<td>77.6</td>
<td>90.7</td>
<td>13.1</td>
</tr>
<tr>
<td>14</td>
<td>83.5</td>
<td>92.5</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>89.9</td>
<td>93.8</td>
<td>3.9</td>
</tr>
<tr>
<td>16</td>
<td>86</td>
<td>91.7</td>
<td>5.7</td>
</tr>
<tr>
<td>17</td>
<td>87.3</td>
<td>98</td>
<td>10.7</td>
</tr>
</tbody>
</table>

Mean: $\bar{D} = 7.26$
Standard deviation: $s_D = 7.16$
Hand et al (1994)

- Hypotheses for scores:
  - H0: $\mu_{\text{Before}} = \mu_{\text{After}}$
  - H1: $\mu_{\text{Before}} \neq \mu_{\text{After}}$

- for difference scores:
  - H0: $(\mu_{\text{After}} - \mu_{\text{Before}}) = \mu_D = 0$
  - H1: $(\mu_{\text{After}} - \mu_{\text{Before}}) = \mu_D \neq 0$

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

mean: $\bar{D} = 7.26$

standard deviation: $s_D = 7.16$

$t = \frac{7.26 - 0}{\frac{7.16}{\sqrt{17}}} = \frac{7.26}{1.74} = 4.18$

df = $N - 1 = 17 - 1 = 16$

- $t$ observed is more extreme than $t$ critical
- reject H0 in favour of H1

Hand et al (1994)

- is family therapy an effective treatment for anorexia?
- 17 participants weighed before & after therapy
- $t$ test used to evaluate H0 of no change in weight
  - rejected H0 in favour of H1 (i.e., weight change $\neq$ zero)
  - direction of effect (after > before) means weight gain not loss
- conclude that family therapy is/was an effective treatment?
  - can you think of an alternative explanation of result?
Hand et al (1994)

- is family therapy an effective treatment for anorexia?
  - rejected H0 in favour of H1 (i.e., weight change ≠ zero)
  - but this is weak evidence for an effect of therapy
  - because simple alternative explanation exists:
    - weight gain was due to normal growth over time
- experiment needs a control group
  - in experimental studies, a group that does not receive the treatment/procedure of interest
  - in correlational studies, a group that is not exposed to or does not experience the variable of interest
- control group in Hand et al study:
  - set of individuals who do not receive therapy
  - ideally, individuals would be assigned randomly to therapy and no-therapy groups. Why?

Hand et al (1994)

- is family therapy an effective treatment for anorexia?
- original experiment did include a control group
  - experimental group: N=17, received family therapy
  - control group: N=26, did not receive family therapy
- new experimental hypothesis:
  - was weight gain different in the two groups?
    - let μFT & μC represent mean weight gain in 2 groups
      - H0: μFT = μC
      - H1: μFT ≠ μC

Hand et al. (1994)

- control (N=26):
  - mean = -0.45, sd = 7.99
- therapy (N=17):
  - mean = 7.26, sd = 7.16
- is difference between group means due to chance?

Sampling Distributions of Group Means

Scores
- control group:
  - mean = -0.42
  - sd = 1.56
  - ≈ 7.99/√26
- therapy group:
  - mean = 7.26
  - sd = 7.16
  - ≈ 7.16/√17

distributions of 10,000 simulated group means
Sampling Distribution of Group Difference

Control Group
mean = -0.45
sd = 7.99
n = 26

Therapy Group
mean = 7.26
sd = 7.16
n = 17

Frequency
0 5 10 15
Mean(Therapy) - Mean(Control)

Group Difference

mean = 7.66
sd = 2.34

Sampling distribution of difference between means

- Each mean has a sampling distribution
- Difference between means also has a sampling distribution
  - mean = μ1 - μ2; variance = VAR1 + VAR2
  - Variance of sum or difference of 2 independent variables equals the sum of their variances
  - if means are distributed normally, then difference (or sum) is distributed normally
    - Central Limit Theorem

Control Group
mean = -0.45
sd = 7.99
n = 26

Therapy Group
mean = 7.26
sd = 7.16
n = 17

mean = 7.66
\( \hat{\mu}_D = \bar{X}_1 - \bar{X}_2 \)

sd = 2.34
\( \hat{\sigma}_D = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \)

Is our observed difference between group means unusually large given the null hypothesis that the two groups do not differ?

We will try to answer this question using a t test.
t test for 2 independent samples

Convert difference between means to a t statistic

\[ t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \]

when \( H_0 \) is \( (\mu_1 - \mu_2) = 0 \)

\[ t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \]

estimating standard error of difference

\[ t = \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \]

\( \hat{\sigma} \bar{X}_1 - \bar{X}_2 = s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \)

when \( n_1 \neq n_2 \)

"pooled" variance estimate

\[ s^2_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \]

\( s^2_p \) is a weighted average of \( s_1^2 \) and \( s_2^2 \)

Hand et al. (1994)

Control Group mean = -0.45 sd = 7.99 n = 26

Therapy Group mean = 7.26 sd = 7.16 n = 17

\[ H_0: \mu_T = \mu_C \]

\[ H_1: \mu_T \neq \mu_C \]

Is the observed value of \( t = 3.22 \) unusual given that the Null hypothesis is true?

\[ t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{s^2_p}{n_1} + \frac{s^2_p}{n_2}}} \]

\[ s^2_p = \frac{25(7.99^2) + 16(7.16^2)}{26 + 17 - 2} = 58.9 \]

\[ t = \frac{7.26 - (-0.45)}{\sqrt{\frac{58.9}{26} + \frac{58.9}{17}}} = \frac{7.71}{2.39} = 3.22 \]

\[ df = 26 + 17 - 2 = 41 \]
Hand et al (1994)

- is family therapy an effective treatment for anorexia?
- measured weight gain in control and therapy groups
- difference between groups was significant
  - \( t(41) = 3.22, p < .05, \) 2-tailed
- reject null hypothesis that weight gain was the same in 2 groups
- result supports (i.e., is consistent with) hypothesis that family therapy is an effective treatment for anorexia
- next steps in research program?
  - replicate findings!
  - "Without replication, all results should be taken as preliminary." -- Gary Marcus