Hypothesis Testing for Means

• Chapter 12 (omit 12.7 - 12.11)

• Testing hypotheses about the value of a population mean by using a sample mean

• Research based on a sample of subjects

• Sample mean is used to test hypotheses
  • need to know what value(s) to expect for the sample mean, assuming $H_0$ vs. $H_1$ is true
  • what happens when a sample is randomly drawn from a population?
Hypothesis Testing for Means

• Sample drawn from a normally distributed population
  • consider population of intelligence test scores
    \( \mu = 100 \quad \sigma = 15 \)
  • draw many samples of \( N = 20 \)

• R code

```r
means = NULL
means = replicate(100000, c(means, mean(rnorm(20, 100, 15))))
hist(means, 100)
mean(means)
sd(means)
```

• try with uniform distribution and vary \( N \)
  ```r
  runif(N, min, max)
  ```
Hypothesis Testing for Means

• Distribution of sample means
  • approximates normal distribution as sample size increases (especially 30 or more), no matter what distribution the original population of scores has
  • always normal if original population is normal
  • mean = population mean $\mu_M = \mu$
  • variance = population variance/sample size

\[
\sigma_M^2 = \frac{\sigma^2}{N} \quad \sigma_M = \frac{\sigma}{\sqrt{N}} \quad (standard \ error \ of \ the \ mean)
\]

• Influence of sample size on variance of sample means
Hypothesis Testing for Means

• Using distribution of sample means to test hypotheses about a population mean
  • mean of random sample taken from a population (distribution of sample means = distribution of all possible outcomes)

• Logic of testing hypotheses about a single population mean using a random sample
  • \( H_0 \) specifies a value for \( \mu \)
  • is the obtained sample mean an unlikely value?
    • among the least likely in the distribution of sample means based on \( H_0 \)?
Hypothesis Testing for Means

• Example
  • research hypothesis: children with no siblings have higher intelligence than the population mean
  • in the general population intelligence scores are normally distributed with \( \mu = 100 \) and \( \sigma = 15 \)
  • hypotheses for only-child population:
    \[ H_0: \mu = 100 \quad H_1: \mu \neq 100 \]
  • draw random sample of 20 only-child children
  • define a result as unlikely if it is among the 5% least likely outcomes
  • suppose for this sample, \( M = 105.6 \)
Hypothesis Testing for Means

• Where is $M = 105.6$ in the distribution of sample means? (This does not refer to the distribution of raw scores!)

• under $H_0$, $\mu = 100$

\[
\sigma_M = \frac{\sigma}{\sqrt{N}} = \frac{15}{\sqrt{20}} = 3.35
\]

\[
Z = \frac{M - \mu}{\sigma_M} = \frac{105.6 - 100}{3.35} = 1.67
\]
Hypothesis Testing for Means

- Summary of hypothesis testing situation
  - one of two possible true states
    - $H_0$ is true or $H_1$ is true
  - one of two possible decisions
    - reject $H_0$ or not reject $H_0$

<table>
<thead>
<tr>
<th>Real Situation</th>
<th>$H_0$ true</th>
<th>$H_1$ true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject $H_0$</td>
<td>Error (Type I) $\alpha$</td>
<td>Correct decision</td>
</tr>
<tr>
<td>Not reject $H_0$</td>
<td>Correct decision</td>
<td>Error (Type II) $\beta$</td>
</tr>
</tbody>
</table>
Hypothesis Testing for Means

- z test depends on unrealistic situations in which \( \sigma^2 \) is known

- Implications of not knowing \( \sigma^2 \)
  - variability of sample means is not known
  - not known where to place observed sample mean in the distribution of sample means

\[ H_0: \mu = 40 \]

\[ M = 46 \]

\[ \sigma_M = ? \]

\[ N = 25 \]

\[ \mu = 40 \]
Hypothesis Testing for Means

• $\sigma^2$ can be estimated from the sample data
• variance of sample scores is an estimate of $\sigma^2$
• importance of using an unbiased estimate of $\sigma^2$

$$s^2 = \frac{\sum(X - M)^2}{N - 1}$$
Hypothesis Testing for Means

- Use sample standard deviation to estimate standard error of the mean
- Distribution of sample means is normal or approximates normal with sufficient $N$
- Estimate $\sigma_M$ using $s \rightarrow s_M = \frac{s}{\sqrt{N}}$

How do we build a useful translation of the distribution of sample means?

\[
Z = \frac{M - \mu}{\sigma_M}
\]
Hypothesis Testing for Means

• Rather than converting to z scores using $\sigma_M$, each sample mean would be converted using its own estimate of $\sigma_M$:

$$s_M = \frac{S}{\sqrt{N}}$$

$$t = \frac{M - \mu}{s_M}$$

• Unlike the z score conversion, this conversion will not generate a normal distribution of $t$, even though distribution of $M$ is normal

• source of the problem: $s^2$ varies across samples and its distribution is a positive skew
Hypothesis Testing for Means

- Distribution of sample variances – \( \sigma^2 = 25, \ N = 10 \)

![Histogram showing the distribution of sample variances with the median of \( s^2 = 23.2 \) and the mean of \( s^2 = 25.0 \).]
Hypothesis Testing for Means

• Distribution of sample variance is positively skewed
  • $s^2$ is smaller than $\sigma^2$ (and $s$ is smaller than $\sigma$) for most samples, especially when $N$ is small

• Converting sample means to $t$ values based on sample-specific $s$ generates a non-normal distribution of $t$

$$ t = \frac{M - \mu}{s_M} $$

More area in the tails of the $t$ distribution--upper .05 of $t$ is further out from the mean
Hypothesis Testing for Means

• Evolution of the $t$ distribution

Normal

$z$

$t$ ($df = 40$)

$t$ ($df = 10$)
Hypothesis Testing for Means

• Evolution of the $t$ distribution
Hypothesis Testing for Means

• What does this mean?
  • a completely new table (Table E.6)
  • a separate $t$ distribution for each value of $df$

$$ t = \frac{M - \mu}{S_M} $$

$df = N - 1$

<table>
<thead>
<tr>
<th>df</th>
<th>.20</th>
<th>.10</th>
<th>.05</th>
<th>.025</th>
<th>.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
<td>31.821</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.886</td>
<td>2.920</td>
<td>4.303</td>
<td>6.965</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.541</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.476</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>1.323</td>
<td>1.721</td>
<td>2.080</td>
<td>2.518</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>1.321</td>
<td>1.717</td>
<td>2.074</td>
<td>2.508</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>1.319</td>
<td>1.714</td>
<td>2.069</td>
<td>2.500</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>1.318</td>
<td>1.711</td>
<td>2.064</td>
<td>2.492</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1.316</td>
<td>1.708</td>
<td>2.060</td>
<td>2.485</td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis Testing for Means

• What does this mean?
  • a completely new table
  • a separate $t$ distribution for each value of $df$

$$t = \frac{M - \mu}{s_M}$$

<table>
<thead>
<tr>
<th>$df$</th>
<th>.05</th>
<th>.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>1.833</td>
<td>2.822</td>
</tr>
</tbody>
</table>
Hypothesis Testing for Means

• What does this mean?
• a completely new table
• a separate $t$ distribution for each value of $df$

$$t = \frac{M - \mu}{s_M}$$

$df = 100$

$0.05$ and $0.01$ critical values for $t$ distribution with $df = 100$.
Hypothesis Testing for Means

- Example: level of dopamine is different in individuals with schizophrenia
  - mean level among unaffected individuals is 36 units
  - \( H_0: \mu = 36 \quad \text{H}_1: \mu \neq 36 \quad (\alpha = .05) \)
  - random sample of 30 individuals with schizophrenia
    \( M = 40.2, \quad s = 10.2 \)

\[
\begin{align*}
\hat{s}_M &= \frac{s}{\sqrt{N}} = \frac{10.2}{\sqrt{30}} = 1.86 \\
t &= \frac{M - \mu}{s_M} = \frac{40.2 - 36}{1.86} = 2.26
\end{align*}
\]

Reject \( H_0 \)

\[
t(29) = 2.26, \quad p < .05
\]
Hypothesis Testing for Means

• Hypothesis: Victims of bullying develop lower than normal levels of self esteem
  • mean score among general population is 84
    \[ H_0: \mu = 84 \quad H_1: \mu < 84 \quad (\alpha = .05) \]
  • random sample of 26 individuals with a history of victimization
    \[ M = 78.3, \quad s = 16.5 \]

\[
s_M = \frac{s}{\sqrt{N}} = \frac{16.5}{\sqrt{26}} = 3.24
\]

\[
t = \frac{M - \mu}{s_M} = \frac{78.3 - 84}{3.24} = -1.76
\]

Reject \( H_0 \)

\[ t(25) = -1.76, \quad p < .05 \]
$t$ test for Related Samples

• Chapter 13 (omit sections 13.5 – 13.7)

• $t$ test for difference between means from related populations
  • repeated measures (Stroop task)
  • each subject has two scores
  • null hypothesis: no difference between the two conditions, so mean of difference scores \textit{in the population} is 0
  • test $H_0$ using one-sample $t$ test, but based on difference scores
$t$ test for Related Samples

Condition 1

Condition 2

Difference scores $(X_1 - X_2)$

$\sigma^2_{X_1 - X_2} = \sigma^2_1 + \sigma^2_2 - 2\rho \sigma_1 \sigma_2$
**t test for Related Samples**

- **Model of all possible results**

  Difference scores ($X_1 - X_2$)  
  
  Sample means ($M_D$)

  ![Diagram showing normal distributions](image)

  When $\sigma_D$ is unknown, use $s_D$ and $t$ distribution with  
  $df = N - 1$

  Normal shape, standard deviation  
  $\sigma_{M_D} = \frac{s_D}{\sqrt{N}}$
$t$ test for Related Samples

- Sample of subjects with alcohol dependency
  - hypothesize a heightened sensitivity to alcohol-related concepts
  - this sensitivity can be revealed in a Stroop task

- Stroop color-naming task with color carried by neutral vs. alcohol-related words

\textbf{WINE} \quad \textbf{DESK} \quad \textbf{LOG} \quad \textbf{BEER} \quad \textbf{CANDLE} \quad \textbf{VODKA}

- obtain mean color-naming time for neutral and for alcohol-related words for each subject and compute difference score (alcohol-related – neutral)

$H_0: \mu_D = 0 \quad H_1: \mu_D > 0 \quad \alpha = .05$

- random sample of 12 alcohol-dependent subjects
t test for Related Samples

• Color-naming times (milliseconds)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>689</td>
<td>701</td>
<td>−12</td>
</tr>
<tr>
<td>2</td>
<td>743</td>
<td>694</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>859</td>
<td>793</td>
<td>66</td>
</tr>
<tr>
<td>4</td>
<td>597</td>
<td>592</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>712</td>
<td>784</td>
<td>−72</td>
</tr>
<tr>
<td>6</td>
<td>852</td>
<td>648</td>
<td>204</td>
</tr>
<tr>
<td>7</td>
<td>634</td>
<td>594</td>
<td>40</td>
</tr>
<tr>
<td>8</td>
<td>821</td>
<td>793</td>
<td>28</td>
</tr>
<tr>
<td>9</td>
<td>932</td>
<td>872</td>
<td>60</td>
</tr>
<tr>
<td>10</td>
<td>743</td>
<td>720</td>
<td>23</td>
</tr>
<tr>
<td>11</td>
<td>840</td>
<td>721</td>
<td>119</td>
</tr>
<tr>
<td>12</td>
<td>750</td>
<td>702</td>
<td>48</td>
</tr>
</tbody>
</table>

* M 764.3 718.8 46.5

$r = .74$
**t test for Related Samples**

- Color-naming times (milliseconds)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>689</td>
<td>701</td>
<td>-12</td>
<td></td>
</tr>
<tr>
<td>743</td>
<td>694</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>859</td>
<td>793</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>597</td>
<td>592</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>712</td>
<td>784</td>
<td>-72</td>
<td></td>
</tr>
<tr>
<td>852</td>
<td>648</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td>634</td>
<td>594</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>821</td>
<td>793</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>932</td>
<td>872</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>743</td>
<td>720</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>840</td>
<td>721</td>
<td>119</td>
<td></td>
</tr>
<tr>
<td>750</td>
<td>702</td>
<td>48</td>
<td></td>
</tr>
</tbody>
</table>

For difference scores:

\[ M_D = 46.50, \ s_D = 67.93 \]

\[ s_{M_D} = \frac{s_D}{\sqrt{N}} = \frac{67.93}{\sqrt{12}} = 19.61 \]

\[ t = \frac{M_D - 0}{s_{M_D}} = \frac{46.50 - 0}{19.61} = 2.37 \]

Critical \( t \) ratio for a one-tailed test:

\( t_{\text{crit}}(11) = 1.796 \)

\( t(11) = 2.37, \ p < .05 \)
t test for Related Samples

• Using R for a related-samples t test

> dat=read.table(file.choose(new=T),header=T)
> mean(dat$alc)
[1] 764.3333
> mean(dat$neut)
[1] 717.8333
> t.test(dat$alc,dat$neut,paired=T)

Paired t-test
data:  dat$alc and dat$neut
t = 2.3713, df = 11, p-value = 0.03706
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:  
  3.340329  89.659671
sample estimates: 
mean of the differences
               46.5
**t test for Related Samples**

- Analysis can be extended to cases in which pairs of related subjects are tested (matched pairs)
  - twins
  - pairs of subjects matched on some relevant variable e.g., match on intelligence when testing the effect of an educational program
$t$ test for Related Samples

- Relevance of correlation between pairs of scores

<table>
<thead>
<tr>
<th>Drug</th>
<th>Placebo</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>

$M_D = 4.0, s_D = 0.82$

$s_{M_D} = \frac{s_D}{\sqrt{N}} = \frac{0.82}{\sqrt{4}} = 0.41$

$t = \frac{M_D - 0}{s_{M_D}} = \frac{4.0 - 0}{0.41} = 9.76$

$t_{crit}(3) = \pm 3.182$

$r = .97$
### t test for Related Samples

- Relevance of correlation between pairs of scores

<table>
<thead>
<tr>
<th></th>
<th>Drug</th>
<th>Placebo</th>
<th>Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>8</td>
<td>-2</td>
</tr>
</tbody>
</table>

**M**

- **Drug:** 8.5
- **Placebo:** 4.5
- **Diff.:** 4.0

**Correlation Coefficient**

\[ r = -0.13 \]

**Hypothesis Testing**

- **H₀:** \( \mu_D = 0 \)
- **H₁:** \( \mu_D \neq 0 \)

**Sample Mean and Standard Deviation**

- \( M_D = 4.0 \)
- \( s_D = 4.24 \)

**Standard Error of the Mean Difference**

\[ s_{M_D} = \frac{s_D}{\sqrt{N}} = \frac{4.24}{\sqrt{4}} = 2.12 \]

**t-Statistic**

\[ t = \frac{M_D - 0}{s_{M_D}} = \frac{4.0 - 0}{2.12} = 1.89 \]

**Critical t-Value**

\[ t_{\text{crit}}(3) = \pm 3.182 \]