Hypothesis Testing

• Chapter 8
• Aplia (week 5 – Sampling distributions)

• Chapter 12
• Aplia (week 7)

• Chapter 9 - section 10
• Aplia (week 6 – Correlation, questions 6 and 7)
Hypothesis Testing

• Testing hypotheses with data

• The case of Madam X
  • professes an ability to predict the future
  • devise a means of testing her ability and deciding whether her claims are valid
Hypothesis Testing

• Logic of hypothesis testing

• Assume some hypothesis is true
  
  *Madam X has no special power*

• Establish alternative hypothesis
  
  *Madam X can predict better than chance*

• Collect a sample of data and determine whether the observed result is unlikely under the assumed hypothesis (*null hypothesis*; $H_0$)
  
  • if result is unlikely, then reject the null hypothesis and accept the *alternative hypothesis* ($H_1$)
  
  • if result is not unlikely, then …
Hypothesis Testing

• Null hypothesis

• coin flip predictions – series of 12 coin flips
  • distribution of outcomes (number of correct predictions) for a person who is just guessing

What kind of result would indicate special powers?

\[ p(6) = (924)(.5)^{12} = .2256 \]

\[ p(12) = (.5)^{12} = .0002 \]
Hypothesis Testing

• Normal approximation for the binomial test
  • probability distribution of possible outcomes approximates the normal distribution

• 50 coin-flip predictions
  \[ n = \text{number of flips} \]
  \[ p = p(\text{correct}) \]
  \[ q = p(\text{incorrect}) \]
  \[ \mu = np = 50(.5) = 25 \]
  \[ \sigma = \sqrt{npq} = \sqrt{50(.5)(.5)} = 3.536 \]
Hypothesis Testing

• Normal approximation for the binomial test
  • probability distribution of possible outcomes approximates the normal distribution

• 50 coin-flip predictions
  - suppose there are 33 correct predictions
  - how likely is this result, or anything more extreme?

\[ z = \frac{r - np}{\sqrt{npq}} = \frac{33 - 25}{3.536} = 2.26 \]

Area in smaller portion = .0119
Hypothesis Testing for a Single Mean

• 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 a Single Mean

• Example: ability to mentally represent spatial information---mental rotation test

• Standardized test provides $\mu$ and $\sigma$ for a population

• Sample of subjects given practice on an action video game
  • does this practice improve spatial skills?
Hypothesis Testing for a Single Mean

- Sample drawn from a population
  - consider distribution of all possible outcomes (sample means) when drawing a sample from a population
    - population: 3, 4, 5, 6 \( \mu = 4.50 \) \( \sigma^2 = 1.25 \)
    - all possible samples of \( N = 2 \) (with replacement)
Sample Testing for a Single Mean

<table>
<thead>
<tr>
<th>Sample</th>
<th>$M$</th>
<th>$(M - \mu)^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3, 3</td>
<td>3.0</td>
<td>$(3.0 - 4.5)^2$</td>
</tr>
<tr>
<td>3, 4</td>
<td>3.5</td>
<td>$(3.5 - 4.5)^2$</td>
</tr>
<tr>
<td>4, 3</td>
<td>3.5</td>
<td>$(3.5 - 4.5)^2$</td>
</tr>
<tr>
<td>3, 5</td>
<td>4.0</td>
<td>$(4.0 - 4.5)^2$</td>
</tr>
<tr>
<td>5, 3</td>
<td>4.0</td>
<td>$(4.0 - 4.5)^2$</td>
</tr>
<tr>
<td>4, 4</td>
<td>4.0</td>
<td>$(4.0 - 4.5)^2$</td>
</tr>
<tr>
<td>3, 6</td>
<td>4.5</td>
<td>$(4.5 - 4.5)^2$</td>
</tr>
<tr>
<td>6, 3</td>
<td>4.5</td>
<td>$(4.5 - 4.5)^2$</td>
</tr>
<tr>
<td>4, 5</td>
<td>4.5</td>
<td>$(4.5 - 4.5)^2$</td>
</tr>
<tr>
<td>5, 4</td>
<td>4.5</td>
<td>$(4.5 - 4.5)^2$</td>
</tr>
<tr>
<td>4, 6</td>
<td>5.0</td>
<td>$(5.0 - 4.5)^2$</td>
</tr>
<tr>
<td>6, 4</td>
<td>5.0</td>
<td>$(5.0 - 4.5)^2$</td>
</tr>
<tr>
<td>5, 5</td>
<td>5.0</td>
<td>$(5.0 - 4.5)^2$</td>
</tr>
<tr>
<td>5, 6</td>
<td>5.5</td>
<td>$(5.5 - 4.5)^2$</td>
</tr>
<tr>
<td>6, 5</td>
<td>5.5</td>
<td>$(5.5 - 4.5)^2$</td>
</tr>
<tr>
<td>6, 6</td>
<td>6.0</td>
<td>$(6.0 - 4.5)^2$</td>
</tr>
</tbody>
</table>

$\Sigma (M - \mu)^2 = 10$

$\mu_M = \mu = 4.50$

$\sigma_M^2 = \frac{10}{16} = 0.625$

$\sigma^2 = \frac{\sigma^2}{N} = \frac{1.25}{2} = 0.625$
Hypothesis Testing for a Single Mean

• 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

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

• Influence of sample size on variance of sample means
Hypothesis Testing for a Single Mean

• 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 a Single Mean

• Example
  • research hypothesis: experience of a traumatic event influences level of neuroticism
  • in non-traumatized population neuroticism scores are normally distributed with a mean of 12.6 and $\sigma = 3.2$
  • hypotheses for traumatized population:
    \[ H_0: \mu = 12.6 \quad H_1: \mu \neq 12.6 \]
  • draw random sample of 20 traumatized people
  • define a result as unlikely if it is among the 5% least likely outcomes
  • suppose for this sample, $M = 13.8$
Hypothesis Testing for a Single Mean

• Where is \( M = 13.8 \) in the distribution of sample means? (This does not refer to the distribution of raw scores!)

• under \( H_0, \ \mu = 12.6 \)

\[
\sigma_M = \frac{\sigma}{\sqrt{N}} = \frac{3.2}{\sqrt{20}} = 0.72
\]

\[
z = \frac{M - \mu}{\sigma_M} = \frac{13.8 - 12.6}{0.72} = 1.67
\]
Hypothesis Testing for a Single Mean

- 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 a Single Mean

- $z$ 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

$H_0: \mu = 40$
Hypothesis Testing for a Single Mean

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

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

- estimate $\sigma_M$ using $s$ -> $s_M = \frac{s}{\sqrt{N}}$
Hypothesis Testing for a Single Mean

• 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 a Single Mean

• Distribution of sample variances – $\sigma^2 = 25$, $N = 10$

![Histogram showing the distribution of sample variances with a median of $s^2 = 23.2$ and a mean of $s^2 = 25.0$.](image)
Hypothesis Testing for a Single Mean

- 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
Review of Essential Concepts

• Evolution of the $t$ distribution

Normal

$t \ (df = 40)$

$t \ (df = 10)$
Review of Essential Concepts

• Evolution of the $t$ distribution
Hypothesis Testing for a Single Mean

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

\[
t = \frac{M - \mu}{S_M}
\]

\[
df = N - 1
\]
Hypothesis Testing for a Single Mean

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

$$t = \frac{M - \mu}{\frac{S}{\sqrt{n}}}$$

$df = 9$

$t$-distribution with $df = 9$
Hypothesis Testing for a Single Mean

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

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

$t = 1.660$ for $df = 100$ at $0.05$ level of significance.
$t = 2.364$ for $df = 100$ at $0.01$ level of significance.
Hypothesis Testing for a Single Mean

• Example: level of dopamine is different in individuals with schizophrenia
  
  • mean level among unaffected individuals is 36 units
    
    \[ H_0: \mu = 36 \quad H_1: \mu \neq 36 \quad (\alpha = .05) \]
  
  • random sample of 30 individuals with schizophrenia
    
    \[ M = 40.2, \quad s = 10.2 \]

\[
s_M = \frac{s}{\sqrt{N}} = \frac{10.2}{\sqrt{30}} = 1.86
\]

\[
t = \frac{M - \mu}{s_M} = \frac{40.2 - 36}{1.86} = 2.26
\]

Reject \( H_0 \)

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

• 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 \( H_0 \) 
\[ t(25) = -1.76, \quad p < .05 \]
Confidence Intervals

• Estimating a population mean
  • point estimate: sample mean, $M$
  • interval estimate: confidence interval
    • based on the fact that a known percentage of sample means fall within a specific distance of $\mu$

Probability that a sample mean will be within one $\sigma_M$ of $\mu$?

Within two $\sigma_M$ of $\mu$?
Confidence Intervals

• Confidence interval
  • consider an interval extending one $\sigma_M$ on each side of $M$
  • does this interval contain $\mu$?

With a random sample, what is the probability that $M$ is within one $\sigma_M$ of $\mu$?
Confidence Intervals

- Confidence interval of arbitrary size
- 95% confidence interval
Confidence Intervals

• Confidence interval of arbitrary size
• 95% confidence interval

Confidence interval is

\[ M \pm z_{\text{crit}}(\sigma_M) = M \pm 1.96(\sigma_M) \]
Confidence Intervals

• Example
  • confidence interval for estimating mean level of neuroticism in population of individuals who have suffered a trauma
    • in population of non-traumatized people, $\sigma = 3.2$
    • draw random sample of 20 traumatized people

$$\sigma_M = \frac{3.2}{\sqrt{20}} = 0.72 \quad M = 13.9$$

95% CI = $M \pm 1.96(\sigma_M) = 13.9 \pm 1.96(0.72)$

= $13.9 \pm 1.41$

95% CI lower limit = 12.49, upper limit = 15.31
Confidence Intervals

- Various levels of confidence

\[ \sigma_M = \frac{3.2}{\sqrt{20}} = 0.72 \]

99% CI = \[ M \pm 2.58(\sigma_M) = 13.9 \pm 1.86 \Rightarrow 12.04 - 15.76 \]

95% CI = \[ M \pm 1.96(\sigma_M) = 13.9 \pm 1.41 \Rightarrow 12.49 - 15.31 \]

90% CI = \[ M \pm 1.65(\sigma_M) = 13.9 \pm 1.19 \Rightarrow 12.71 - 15.09 \]

Principle: higher confidence => wider interval

Distribution of sample means

Principle: higher confidence => wider interval
Confidence Intervals

• Impact of sample size on CI

\[ \sigma_M = \frac{3.2}{\sqrt{20}} = 0.72 \]
\[ \sigma_M = \frac{3.2}{\sqrt{40}} = 0.51 \]
\[ \sigma_M = \frac{3.2}{\sqrt{60}} = 0.41 \]

\( N = 20: \) 95% CI = \( M \pm 1.96(\sigma_M) = 13.9 \pm 1.41 \)
\( N = 40: \) 95% CI = \( M \pm 1.96(\sigma_M) = 13.9 \pm 1.00 \)
\( N = 60: \) 95% CI = \( M \pm 1.96(\sigma_M) = 13.9 \pm 0.80 \)

Principle: larger \( N \) \( \Rightarrow \) smaller interval
Confidence Intervals

• Confidence intervals when \( \sigma^2 \) is not known
  • variability of distribution of sample means estimated using \( s_M \)

\[ t = \frac{M - \mu}{s_M} \]

\[ N = 30 \]
\[ df = 29 \]

| Level of Significance for One-Tailed Test |
|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| \( \cdot .15 \)   | \( .10 \)         | \( .05 \)        | \( .025 \)        | \( .01 \)        |

|LEVEL OF SIGNIFICANCE FOR TWO-TAILED TEST|
|-----------------|-----------------|-----------------|-----------------|-----------------|
| \( df = .30 \)  | \( .20 \)       | \( .10 \)       | \( .05 \)       | \( .02 \)       |

| \( df = 26 \)  | \( 1.058 \)  | \( 1.315 \)  | \( 1.706 \)  | \( 2.056 \)  | \( 2.479 \)  |
| \( df = 27 \)  | \( 1.057 \)  | \( 1.314 \)  | \( 1.703 \)  | \( 2.052 \)  | \( 2.473 \)  |
| \( df = 28 \)  | \( 1.056 \)  | \( 1.313 \)  | \( 1.701 \)  | \( 2.048 \)  | \( 2.467 \)  |
| \( df = 29 \)  | \( 1.055 \)  | \( 1.311 \)  | \( 1.699 \)  | \( 2.045 \)  | \( 2.462 \)  |
| \( df = 30 \)  | \( 1.055 \)  | \( 1.310 \)  | \( 1.697 \)  | \( 2.042 \)  | \( 2.457 \)  |
Confidence Intervals

- Confidence intervals when $\sigma^2$ is not known
  - variability of distribution of sample means estimated using $s_M$

Sample data: $N = 30$, $M = 78.3$, $s = 17.8$

$$s_M = \frac{17.8}{\sqrt{30}} = 3.25$$

95% CI: $M \pm t_{crit} (s_M)$

$$= 78.3 \pm 2.045(3.25)$$

$$= 78.3 \pm 6.65$$

71.65 to 84.95
Confidence Intervals

• A note about interpreting confidence intervals
  • a procedure for generating intervals such that 95% of them will contain $\mu$
  • once an interval is constructed from an obtained sample, we cannot make valid claims about how likely it is that $\mu$ is in that particular interval
Significance Test for $r$

- Section 9.10 in the Howell text book

- Population parameter $\rho$ null hypothesis: $\rho = 0$
  - even if $\rho$ is 0, it is likely that a (random) sample will produce $r \neq 0$

- Consider a small population of scores with $\rho = 0$
Significance Test for $r$

• Section 9.10

• Population parameter $\rho$ null hypothesis: $\rho = 0$
  • even if $\rho$ is 0, it is likely that a (random) sample will produce $r \neq 0$

• Consider a small population of scores with $\rho = 0$
Significance Test for $r$

- Population parameter $\rho$ null hypothesis: $\rho = 0$
  - is the observed value of $r$ among the least likely expected under the null hypothesis?

- what does the distribution of sample values of $r$ look like under the null hypothesis?
  - roughly normal, depending on sample size
  - as sample size increases, variability of $r$ values decreases

![Graphs showing $N = 8$ and $N = 40$](image)
Significance Test for $r$

- Table of critical values for $r$ (Table E.2 in textbook)
  - depends on sample size and significance level
    - degrees of freedom: $df = N - 2$
    - significance level usually .05 (least likely values)
  - two-tailed $p$ values

<table>
<thead>
<tr>
<th>df</th>
<th>$p = .10$</th>
<th>$p = .05$</th>
<th>$p = .025$</th>
<th>$p = .01$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.805</td>
<td>0.878</td>
<td>0.924</td>
<td>0.959</td>
</tr>
<tr>
<td>4</td>
<td>0.729</td>
<td>0.811</td>
<td>0.868</td>
<td>0.917</td>
</tr>
<tr>
<td>5</td>
<td>0.669</td>
<td>0.755</td>
<td>0.817</td>
<td>0.875</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.211</td>
<td>0.250</td>
<td>0.285</td>
<td>0.325</td>
</tr>
<tr>
<td>120</td>
<td>0.150</td>
<td>0.178</td>
<td>0.203</td>
<td>0.232</td>
</tr>
<tr>
<td>200</td>
<td>0.116</td>
<td>0.138</td>
<td>0.158</td>
<td>0.181</td>
</tr>
<tr>
<td>500</td>
<td>0.073</td>
<td>0.088</td>
<td>0.100</td>
<td>0.115</td>
</tr>
</tbody>
</table>
Significance Test for \( r \)

- Reporting significance test for \( r \)
  - \( N = 70 \quad r = .30 \)
  - \( r(68) = .30, \ p < .05 \)

<table>
<thead>
<tr>
<th>( df )</th>
<th>( p = .10 )</th>
<th>( p = .05 )</th>
<th>( p = .025 )</th>
<th>( p = .01 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.805</td>
<td>0.878</td>
<td>0.924</td>
<td>0.959</td>
</tr>
<tr>
<td>4</td>
<td>0.729</td>
<td>0.811</td>
<td>0.868</td>
<td>0.917</td>
</tr>
<tr>
<td>5</td>
<td>0.669</td>
<td>0.755</td>
<td>0.817</td>
<td>0.875</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>60</td>
<td>0.211</td>
<td><strong>0.250</strong></td>
<td>0.285</td>
<td>0.325</td>
</tr>
<tr>
<td>120</td>
<td>0.150</td>
<td>0.178</td>
<td>0.203</td>
<td>0.232</td>
</tr>
<tr>
<td>200</td>
<td>0.116</td>
<td>0.138</td>
<td>0.158</td>
<td>0.181</td>
</tr>
<tr>
<td>500</td>
<td>0.073</td>
<td>0.088</td>
<td>0.100</td>
<td>0.115</td>
</tr>
</tbody>
</table>
Significance Test for $r$

- **Using R to test significance of $r$**

```r
> dat=read.table(file.choose(new=T),header=T)
> plot(dat)
> head(dat)
   X  Y
1  33 81
2  38 64
3  30 76
4  41 74
5  38 66
6  38 63
```
Significance Test for $r$

- Using R to test significance of $r$

```r
> cor.test(dat$X, dat$Y)

Pearson's product-moment correlation

data:  dat$X and dat$Y
t = -3.1535, df = 38, p-value = 0.003147
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:  
  -0.6716504  -0.1677122
sample estimates:  
  cor
-0.4554282
```
Significance Test for $r$

- Test of $r$ by conversion to $t$ ratio
  - degrees of freedom: $df = N - 2$

$$t = r \sqrt{\frac{N - 2}{1 - r^2}}$$
Significance Test for $r$

- $X =$ treatment, $Y =$ cog. latency
- cell phone use & simulated driving

$X$  $Y$
1  484
1  512
1  402
1  387
1  431
1  469
0  352
0  401
0  358
0  336
0  379
0  397

$r = .73$

$t = r \sqrt{\frac{N - 2}{1 - r^2}}$

$= .73 \sqrt{\frac{12 - 2}{1 - .73^2}}$

$= 3.38$

$t_{crit}(10) = \pm 2.228$

$H_0: \rho = 0$

$H_1: \rho \neq 0$
Significance Test for $r$

- $X = $ treatment, $Y = $ cog. latency

$H_0$: $\rho = 0$

$H_1$: $\rho \neq 0$

$N = 12$

$r = .73$

From Table E.2

$df = 10$, critical $r = \pm .576$

**Table E.2**

Significant Values of the Correlation Coefficient

<table>
<thead>
<tr>
<th>$df$</th>
<th>$p = .10$</th>
<th>$p = .05$</th>
<th>$p = .025$</th>
<th>$p = .01$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.805</td>
<td>.878</td>
<td>.924</td>
<td>.959</td>
</tr>
<tr>
<td>9</td>
<td>.521</td>
<td>.602</td>
<td>.667</td>
<td>.735</td>
</tr>
<tr>
<td>10</td>
<td>.498</td>
<td>.576</td>
<td>.640</td>
<td>.708</td>
</tr>
<tr>
<td>11</td>
<td>.476</td>
<td>.553</td>
<td>.616</td>
<td>.684</td>
</tr>
</tbody>
</table>

$r(10) = .73$, $p < .01$