6: Introduction to Hypothesis

The general idea of statistical inference is to use sample data to draw conclusions about the entire population.

Recall: Two Types of Inference

Estimation (prior chapter)
- Confidence intervals to estimate population parameters

Hypothesis testing (this chapter)
- Test statistics to judge claim about population
The General Idea

Research question → hypothesis ("claim") → collect data → test claim

Two Philosophies of Statistical Testing
- Evidence-weighing ("significance testing")
- Decision-making ("fixed-level testing")

This chapter focuses on decision-making approach
- Future chapters move toward evidence-weighing

Procedure
Decision-Making

(A) Convert research question to statistical hypotheses
   - Null ($H_0$): no difference in populations
   - Alternative ($H_1$): difference in populations
   - Assume $H_0$ is true until proved otherwise

(B) Set error threshold

(C) Calculate test statistic

(D)
   - Convert test statistic to probability
   - Reject or retain $H_0$
Potential Consequences of Decision

p. 6.3

<table>
<thead>
<tr>
<th>Decision</th>
<th>$H_0$ true</th>
<th>$H_0$ false</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retain $H_0$</td>
<td>OK retention</td>
<td>Type II error</td>
</tr>
<tr>
<td>Reject $H_0$</td>
<td>Type I error</td>
<td>OK rejection</td>
</tr>
</tbody>
</table>

P Errors
- Type I error = false rejection of $H_0$
- Type II error = false retention of $H_0$

P Error probabilities
- $\alpha = \text{Probability(Type I error)}$
- $\beta = \text{Probability(Type II error)}$

P Anti-error probabilities
- $1 - \alpha = \text{"confidence"}$
- $1 - \beta = \text{"power"}$

Z test, One-Sided Alternative

p. 6.4

$\sigma$ must be known to use $z$ test
- Otherwise, use $t$ test

Tests can be one-sided or two-sided
- We start with one-sided tests
- Which are simpler
- But . . .
- are less common in practice
One-Sample, Z Test

(A) $H_0: \mu = \mu_0$, where $\mu_0$ represents parameter if $H_0$ true ("null value")
- Alternative either
  - $H_1: \mu < \mu_0$ (looking for values to the left of $\mu_0$)
  - $H_1: \mu > \mu_0$ (looking for values to the right of $\mu_0$)

(B) Set $\alpha$ (Commonly .10, .05, .01)
- $\alpha$ set by researcher (not data)

(C) Test statistic

(D) Conclusion: $p$ value and decision

Illustrative Example (IQ Scores)

PWelscher IQ scores are normally distributed with $\mu = 100$ and $\sigma = 15$

PDo children from a particular school (population) have higher than average IQ scores?

PSample:
- $n = 9$
- sample mean = 112.8
- SEM = $15 / \sqrt{9} = 5$
Illustrative Example: IFF $H_0$ True

(p. 6.6)

# If $H_0$ true, then
Q$\mu = 100$
QSample means would vary from sample-to-sample
QSDM would be normal

# $\therefore$ SDM would be normal, centered on 100 with
SEM = 5
Qe.g., 97.5% of sample means will less than 110
QSeeing a value greater than 110 would occur only 2.5%

Illustrative Example (cont.)

p. 6.6

(A) $H_0: \mu = 100$ vs. $H_1: \mu > 100$
(B) Let $\alpha = .025$ (or whatever)
(C) Test statistic: Formula 6.1 $z_{stat} = \frac{\bar{x} - \mu_0}{SEM}$

where:
“xbar” $\equiv$ sample mean
$\mu_0 \equiv$ "null value"
SEM = $\sigma / \sqrt{n}$

$$z_{stat} = \frac{112.8 - 100}{5} = 2.56$$
Focus on $z_{\text{stat}}$

$$z_{\text{stat}} = \frac{\bar{x} - \mu_0}{SEM}$$

- Components of $z_{\text{stat}}$
  - Numerator = observed mean − expected mean
  - Denominator = standard error of mean
- $P_{z_{\text{stat}}} =$ "standard deviations" above expectation if $H_0$ were true
- $P$-value = probability test statistic is more extreme than observed assuming $\mu_0$ true = area under curve beyond $z_{\text{stat}}$
- Look-up probability with help of $z$ table

Converting $z_{\text{stat}}$ to Probability

$p$-value = probability test statistic is more extreme than observed assuming $\mu_0$ true = area under curve beyond $z_{\text{stat}}$
**Decision**

p. 6.6

If \( p \leq \alpha \rightarrow \text{reject } H_0 \\
\text{If } p > \alpha \rightarrow \text{retain } H_0\\

Illustrative example (IQ scores)

- \( \alpha = .025 \) (set by researcher)
- \( p = .0052 \) (calculated from data)
- Since \( p < \alpha \rightarrow \text{reject } H_0 \\
- \therefore \text{ difference is “significant”}

---

**One-sample Z test: two-sided } H_1**

p. 6.4

Similar to one-sided test but without prior specification of direction of difference

More common than one-sided test

Same four steps

- A: \( H_0: \mu = \mu_0 \) vs. \( H_1: \mu \neq \mu_0 \)
- B: Let \( \alpha = \) something
- C: Test stat = Formula 6.1
- D: Two-sided \( p \) value = 2 \times one-sided \( p \) value
**Illustrative Example (IQ scores)**

Do IQs differ (without specifying direction)

Procedure
- **A:** $H_0: \mu = 100$ vs. $H_1: \mu \neq 100$
- **B:** Let $\alpha = .025$ (or whatever)
- **C:** $z_{stat} = 2.56$ (same as before)
- **D:** Two-sided $p = 2 \times .0052 = .0104$; reject $H_0$ since $p < \alpha$

```
<table>
<thead>
<tr>
<th>IQ</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>85</td>
<td>-3</td>
</tr>
<tr>
<td>90</td>
<td>-2</td>
</tr>
<tr>
<td>95</td>
<td>-1</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>105</td>
<td>1</td>
</tr>
<tr>
<td>110</td>
<td>2</td>
</tr>
<tr>
<td>115</td>
<td>3</td>
</tr>
</tbody>
</table>
```

![Normal distribution with z-scores and critical values]

**One sample t test**

Use $t$ tests when $\sigma$ is unknown (calculates)

Test can be one-sided or two-sided

Illustrative example (\% IDEAL)
- Variable = % of ideal body weight
  - e.g., value of 100 represents 100% of ideal body weight
- Sample: 18 diabetics
- Question: Is $\mu$ different than 100?
One-Sample Test

(A) 
- $H_0$: $\mu = \mu_0$
- Either
  - $H_1$: $\mu < \mu_0$ (looking for values to the left of $\mu_0$)
  - $H_1$: $\mu > \mu_0$ (looking for values to the right of $\mu_0$)
  - $H_1$: $\mu \neq \mu_0$ (two-sided)

(B) Set $\alpha$

(C) Test statistic (Formula 6.2) 
$$t_{stat} = \frac{\bar{x} - \mu_0}{sem}$$
- with df = $n - 1$
- you lose 1 df b/c you have to estimate $\mu$ with xbar

(D)

Illustrative example (cont.)

- $H_0$: $\mu = 100$ vs. $H_1$: $\mu \neq 100$ (two-sided)
- Let $\alpha = .05$ (or whatever)
- Test statistic
  - $\bar{x} = 112.778$, $s = 14.242$, $n = 18$
  - $sem = 14.242 / \sqrt{18} = 3.40$
  - $t_{stat} = (112.778 - 100) / 3.40 = 3.76$
  - df = $18 - 1 = 17$
Convert $t_{\text{stat}}$ to $p$ value

- By computer – use SPSS or StaTable
- By hand – use $t$ table “wedgie technique”
  - Put $t_{\text{stat}}$ between two landmarks from $t$ table
  - Determine tail areas of landmarks
  - $p$ value less between these probabilities
  - Double probability if test is two-sided

Illustrative

Figure on p. 6.x

$t_{\text{stat}} = 3.76$

- One tail less than .001 and more than .0005
Double tail areas for two-sided tests

\[ \begin{align*}
&\begin{array}{c}
\frac{1}{2}p \\
-3.76
\end{array} \\
&\begin{array}{c}
.001 \\
3.76
\end{array}
\end{align*}\]

\[ \begin{align*}
&.001 < p < .002
\end{align*}\]

\[ \therefore \text{Two tails less than .002 and more than .001} \]

\[ \therefore 0.001 < p < 0.002 \]

Fallacies of Statistical Testing

p. 6.2

\begin{enumerate}
\item Failure to reject \( H_0 \) = accept \( H_0 \)
  \begin{itemize}
  \item WRONG!
  \item Failure to reject \( H_0 \) = insufficient evidence for rejection
  \end{itemize}
\item \( p \) value = probability \( H_0 \) is incorrect
  \begin{itemize}
  \item WRONG!
  \item \( p \) value = probability of data assuming \( H_0 \) is correct
  \end{itemize}
\end{enumerate}
Fallacies of Statistical Testing (cont.)

p. 6.2

Statistical significance implies importance

- WRONG!
- WRONG!
- WRONG!
- Statistical significance does not address causality or size of effect