Hypothesis Testing

• Test of hypothesis - Test whether a population parameter is less than, equal to, or greater than a specified value.
• Remember an inference without a measure of reliability is little more than a guess.

Example

• A contract between a city and a road construction company called for an asphalt road with an average thickness of 10 inches. The county thought that the construction company had defrauded the city by making the road less than 10 inches thick. The city then took 100 core samples 1 inch in diameter. The mean thickness of the samples was 9.5 inches. Did fraud occur?

Example

• Suppose building specifications for a city require that the average breaking strength of residential sewer pipe to be more than 2,400 lbs per foot of length. To sell pipe in the city a manufacturer must demonstrate that its product meets the specifications.
• Here we are interested not so much in the value of μ than we are in testing a hypothesis about its value. In other words, we are interested in whether the mean breaking strength of the pipe exceeds 2,400 lbs per foot.
Hypotheses Statements

- Two hypotheses:
  - **Null hypothesis** \( (H_0) \) - status quo to the party performing the experiment. The hypothesis will not be rejected unless the data provide convincing evidence that it is false.
  - **Alternative or research hypothesis** \( (H_a) \) - which will be accepted only if the data provide convincing evidence of its truth.

Hypotheses Statements Examples

- For pipe example:
  - \( (H_0) \) \( \mu \leq 2,400 \) (the pipe does not meet specifications)
  - \( (H_a) \) \( \mu > 2,400 \) (the pipe meets specifications)

- How can the city decide when enough evidence exists to conclude the pipe meets specifications?
  - When the sample mean \( \bar{x} \) convincingly indicates that the population mean exceeds 2,400 lbs per foot.

Evidence

- "Convincing" evidence in favor of the alternative exists when the value of \( \bar{x} \) exceeds 2,400 by an amount that cannot be readily attributed to sampling variability.
- We need to compute a "test statistic" which will be the z-value that measures the distance between the values of \( \bar{x} \) and the value of \( \mu \) specified in the null hypothesis.

Example

- The test statistic is:
  \[ z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{\bar{x} - 2,400}{\sigma / \sqrt{n}} \]

- A value of \( z=1 \) means that x-bar is 1 standard error above \( \mu = 2,400 \)
- A value of \( z=1.5 \) means that x-bar is 1.5 standard errors above \( \mu = 2,400 \)
Two Types of Errors: Type I Error

- Deciding that the null hypothesis is false when in fact it is true is called a Type I error. The probability of making a Type I error is only:
  - \( P(\text{Type I error}) \)  \( \alpha = 0.05 \)
  - \( = P(\text{reject } H_0 \text{ when in fact } H_0 \text{ is true}) \)
  - \( = P(z > 1.645 \text{ when in fact } \mu \leq 2,400) = 0.05 \)

The Test Itself

- Null and alternative hypothesis:
  - \( H_0: \mu \leq 2,400 \) \( H_a: \mu > 2,400 \)
- Test statistic:
  - \( z = \frac{\bar{x} - 2,400}{s} \)
- Rejection region:
  - \( z > 1.645 \) for \( \alpha = 0.05 \)
- Example: assume \( \bar{x} = 2,460 \) \( s = 200 \) \( n = 50 \)
- What is your decision?

Judicial Analogy

Two Types of Errors: Type II Error

- Not rejecting the null hypothesis when in fact it is false is called a Type II error. We denote the probability of committing a Type II error by \( \beta \).

- Usually difficult to determine so we simply say that the sample evidence is insufficient to reject the null hypothesis at the \( \alpha = 0.05 \) confidence level.
Chapter 9

Decisions and Consequences for a Test of Hypothesis

<table>
<thead>
<tr>
<th>Decision</th>
<th>True State of Nature</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ho True</td>
<td>Ha True</td>
<td></td>
</tr>
<tr>
<td>Do Not Reject Ho</td>
<td>Correct Decision</td>
<td>Type II Error</td>
</tr>
<tr>
<td>Reject Ho</td>
<td>Type I Error</td>
<td>Correct Decision</td>
</tr>
</tbody>
</table>

A Type I error has a probability of $\alpha$.  
A Type II error has a probability of $\beta$.

Elements of a Test of Hypothesis

- 1. Null Hypothesis
- 2. Alternative Hypothesis
- 3. Test Statistic
- 4. Rejection Region
- 5. Assumptions
- 6. Sample and Calculate Test Statistic
- 7. Conclusion

Elements of a Test of Hypothesis

- 1. Null Hypothesis
- 2. Alternative Hypothesis
- 3. Test Statistic
- 4. Rejection Region
- 5. Assumptions
- 6. Sample and Calculate Test Statistic
- 7. Conclusion

Steps for Selecting the Null and Alternative Hypotheses

1. Select the alternative hypothesis - 3 types:
   - One-tailed, upper tail $H_a$: $\mu > 2400$
   - One-tailed, lower tail $H_a$: $\mu < 2400$
   - Two-tailed $H_a$: $\mu \neq 2400$

2. Select the null hypothesis as the status quo, which will be presumed true unless the sample evidence conclusively establishes the alternative hypothesis: $H_0$: $\mu \leq 2400$
Large-Sample Test of Hypothesis about a Mean

- Use the z-statistic.

- Example - A farmer hypothesizes that the mean yield of soybeans per acre is greater than 520 bushels. A sample of 36 1-acre plots results in a sample mean value of 543 with a standard deviation of 124. At a .025 level, can we conclude that the mean yield is above 520?

Small-Sample Test of Hypothesis about a Mean

- If n is small, we cannot assume that the sample standard deviation will provide a good approximation of the population standard deviation. We must use the t-distribution rather than the z-distribution to make inferences about the population mean.

- Test Statistic: \( t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \)

Example

- A tire company guarantees that a particular tire has a mean useful lifetime of 42,000 miles or more. A consumer testing agency, wishing to verify this claim, observed n=10 tires on a test wheel that simulated normal road conditions. The lifetimes (in thousands of miles) were as follows: 42, 36, 46, 43, 41, 35, 43, 45, 40, 39. Use these data to determine whether there is sufficient evidence to contradict the manufacturers claim at the .05 level.

P-Values

- For all the hypothesis testing examples so far, the value of \( \alpha \) and thus the rejection region are selected prior to conducting the test.

- A second method is to report the extent to which the statistic disagrees with the null hypothesis and letting the reader decide whether to reject the null hypothesis. This measure of disagreement is called the p-value.
Steps for Calculating the p-value

- Choose the maximum level of $\alpha$ you are willing to tolerate.
- Determine the value of the test statistic $z$ from the sample data. Look up the $z$-statistic and find the corresponding probability:
  - One-tailed test - the p-value = tail area beyond $z$ in the same direction as the alternative hypothesis.
  - Two-tailed test - the p-value = 2 times the tail area beyond the $z$ value in the direction of the sign of $z$.
- Reject the null hypothesis if the observed p-value is less than $\alpha$.

Example

- According to a national survey, the price of a new family home averages $115,000. You hypothesize that the average price in your city exceeds the national average. To test your belief, you record the prices of 20 new homes in your city and find the sample mean to equal $120,000 with a sample standard deviation of $30,000. If $\alpha = .05$, does the data support your belief? What is the p-value? What would be the result if the sample was 100 homes with the same mean and standard deviation above? What is the p-value?

Large Sample Test of Hypothesis about a Binomial Probability

- Inferences about proportions (or percentages) are often made in the context of the probability, $p$, of "success" for a binomial distribution.
- Previously we have calculated a confidence interval for $p$.
- For hypothesis tests on proportions, do not use the sample proportion to calculate the standard deviation. Instead, use the standard deviation based on the value of the null hypothesis.

Hypothesis Test

- $H_0$: $p = p_0$
- $H_a$: $p > p_0$  
  $p < p_0$  
  $p \neq p_0$
- Test statistic: $z = \frac{\hat{p} - p_0}{\sigma_p}$
- Rejection region: $z > z_\alpha$  
  $z < -z_\alpha$  
  $|z| > z_\alpha$
Example

• A rating service for local radio stations claims that 40% of the prime drive-time audience listens to WXIX. You are the advertising manager of a competing radio station and you doubt that WXIX’s market share is that large. Formulate and test the appropriate hypothesis at the 5% significance level if 380 people out of 1,000 surveyed say that they listen to WXIX. What is the p-value?

The Relationship Between Confidence Intervals and Hypothesis Tests

• A Confidence Interval is really a two-tailed Hypothesis Test (at the same $\alpha$-level):
  – For a confidence interval, we are (1-$\alpha$)% sure that the true value of the population parameter is between the lower confidence level (LCL) and the upper confidence level (UCL). We are $\alpha$% sure that the true value of the population parameter is not in this range.
  – For a hypothesis test, we either (1) fail to reject $H_0$ at the (1-$\alpha$)% level effectively creating a confidence interval or (2) reject $H_0$ in favor of $H_a$ which is the area outside of the confidence interval.

Example

• Our previous farmer example:
  – A farmer hypothesizes that the mean yield of soybeans per acre is not equal to 520 bushels. Sampling info of 36 one-acre plots finds a sample mean of 543 with a standard deviation of 124.
  – At a 0.10 level, can we conclude that the mean yield is not equal to 520?
  – Construct a 90% confidence interval using the above information.
Two Population Means, \( \mu_1 \) and \( \mu_2 \)

- **Independent Samples:**
  - Mean income of households in two different neighborhoods in Fairfax County, VA: Incomes in one neighborhood are unrelated (independent) to incomes in the other neighborhood.
- **Dependent Samples:**
  - Mean daily sales of two restaurants located on the same block in New York City: Is there a difference in daily sales volume between the two restaurants?

### Independent Sampling

- Consider the Sampling Distribution of \( \bar{X}_1 - \bar{X}_2 \)
  - The Mean of the sampling distribution of \( \mu_1 - \mu_2 \) is \( \bar{X}_1 - \bar{X}_2 \)
  - Since the two samples are independent, the standard deviation of the sampling distribution (standard error of \( \bar{X}_1 - \bar{X}_2 \)) is:
    \[
    \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}
    \]
  - The sampling distribution of \( \sigma(\bar{X}_1 - \bar{X}_2) \) is approximately normal for large samples (Central Limit Theorem) so use a z-statistic.

### Large Sample Hypothesis Test for \( \mu_1 - \mu_2 \) - Independent Sampling

- **One-tailed Test:**
  - H\(_0\): \( \mu_1 - \mu_2 \) = \( \mu_0 \) vs H\(_a\): \( \mu_1 - \mu_2 \) > \( \mu_0 \)
  - H\(_0\): \( \mu_1 - \mu_2 \) = \( \mu_0 \) vs H\(_a\): \( \mu_1 - \mu_2 \) < \( \mu_0 \)
- **Two-tailed Test:**
  - H\(_0\): \( \mu_1 - \mu_2 \) = \( \mu_0 \) vs H\(_a\): \( \mu_1 - \mu_2 \) not equal to \( \mu_0 \)
- **Test statistic:**
  \[
  z = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sigma_{\bar{X}_1 - \bar{X}_2}}
  \]
  \[
  \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}
  \]

### Small Sample Confidence Interval for \( \mu_1 - \mu_2 \) - Independent Sampling

- Both populations must be approximately normally distributed with equal variances (these variances are usually unknown):
  \[
  \left( \bar{X}_1 - \bar{X}_2 \right) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}
  \]
  \[
  \sigma^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}
  \]
Chapter 9

Small Sample Hypothesis Test for $(\mu_1 - \mu_2)$ - Independent Sampling

• One-tailed Test:
  - $H_0: (\mu_1 - \mu_2) = D_0$ (or 0) vs $H_a: (\mu_1 - \mu_2) > D_0$ (or 0)
  - $H_0: (\mu_1 - \mu_2) = D_0$ (or 0) vs $H_a: (\mu_1 - \mu_2) < D_0$ (or 0)

• Two-tailed Test:
  - $H_0: (\mu_1 - \mu_2) = D_0$ (or 0) vs $H_a: (\mu_1 - \mu_2) \neq D_0$ (or 0)

• Test statistic:
  $t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
  $s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

Dependent Sampling

• We are unable to use the two-sample $(\mu_1 - \mu_2)$ method for performing hypothesis tests and calculating confidence intervals if the samples are dependent on some external factor.

• Instead, use Paired Difference method which takes sample dependence into account when performing hypothesis tests and calculating confidence intervals.

Paired Difference Method

• The test statistic becomes:
  $t = \frac{\bar{D} - D_0}{\frac{s_D}{\sqrt{n_D}}}$

  – Assumptions:
    • The distribution of differences is approximately normally distributed.
    • The sample differences are random.

Paired Difference Hypothesis Test

• One-tailed Test:
  - $H_0: (\mu_1 - \mu_2) = D_0$ (or 0) vs $H_a: (\mu_1 - \mu_2) > D_0$ (or 0)
  - $\mu_D = D_0$ (or 0)
  - $\mu_D > D_0$ (or 0)
  - $H_0: (\mu_1 - \mu_2) = D_0$ (or 0) vs $H_a: (\mu_1 - \mu_2) < D_0$ (or 0)
  - $\mu_D = D_0$ (or 0)
  - $\mu_D < D_0$ (or 0)

• Two-tailed Test:
  - $H_0: (\mu_1 - \mu_2) = D_0$ (or 0) vs $H_a: (\mu_1 - \mu_2) \neq D_0$ (or 0)
  - $\mu_D = D_0$ (or 0)
  - $\mu_D \neq D_0$ (or 0)

• Test statistic:
  $t = \frac{\bar{D} - D_0}{\frac{s_D}{\sqrt{n_D}}} \text{ where } t \text{ has } (n_D - 1) \text{ degrees of freedom}$
Hypothesis Testing using Excel