Statistics 13 Elementary Statistics
Summer Session I 2012

Lecture Notes 6: Tests of Hypothesis

Suppose you wanted to determine whether the mean level of a driver’s blood alcohol exceeds the legal limit after two drinks, or whether the majority of registered voters approve of the president’s performance. In both cases, you are interested in making an inference about how the value of a parameter relates to a specific numerical value. Is it less than, equal to, or greater than the specified number? This type of inference, called a test of hypothesis.

6.1 The Elements of a Test of Hypothesis

The elements of the test:

1. Null hypothesis ($H_0$): A theory about the values of one or more population parameters. The theory generally represents the status quo, which we adopt until it is proven false. By convention, the theory is stated as $H_0$: parameter=value.

2. Alternative (research) hypothesis ($H_a$): A theory that contradicts the null hypothesis. The theory generally represents that which we will accept only when sufficient evidence exist to establish its truth.

3. Test statistic: A sample statistic used to decide whether to reject the null hypothesis.

4. Rejection region: The numerical values of the test statistic for which the null hypothesis will be rejected. The rejection region is chosen so that the probability is $\alpha$ that it will contain the test statistic when the null hypothesis is true, thereby leading to a Type I error. The value of $\alpha$ is usually chosen to be small (e.g., 0.01, 0.05, or 0.10) and is referred to as the level of significance of the test.

5. Assumptions: Clear statements of any assumptions made about the population(s) being sampled.

6. Experiment and calculation of test statistic: Performance of the sampling experiment and determination of the numerical value of the test statistic.

7. Conclusion:
   a. If the numerical value of the test statistic falls into the rejection region, we reject the null hypothesis and conclude that the alternative hypothesis is true. We know

---

1Last update: July 15, 2012
that the hypothesis-testing process will lead to this conclusion incorrectly (a Type I error) only $100\alpha\%$ of the time when $H_0$ is true.

b. If the test statistic does not fall into the rejection region, we do not reject $H_0$. Thus, we reserve judgement about which hypothesis is true. We do not conclude that the null hypothesis is true because we do not (in general) know the probability $\beta$ that our test procedure will lead to an incorrect acceptance of $H_0$ (a Type II error).

**Example 1** Suppose building specifications in a certain city require that the average breaking strength of residential sewer pipe be more than 2,400 pounds per foot of length (i.e., per linear foot). Each manufacturer that wants to sell pipe in that city must demonstrate that its product meets the specification. Note that we are interested in making an inference about the mean $\mu$ of a population. However, in this example, we are less interested in estimating the value of $\mu$ than we are in testing a hypothesis about its value. That is, we want to decide whether the mean breaking strength of the pipe exceeds 2,400 pounds per linear foot. Suppose we test 50 sections of sewer pipe and find the mean and standard deviation for the measurements are $\bar{x} = 2,460$ pounds per linear foot and $s = 200$ pounds per linear foot. Using $\alpha = 0.05$.

**Solution:**

1. **Null hypothesis** ($H_0$): $\mu \leq 2,400$ (i.e., the manufacturer’s pipe does not meet specifications)

2. **Alternative hypothesis** ($H_a$): $\mu > 2,400$ (i.e., the manufacturer’s pipe meets specifications)

3. **Test statistic:**

   \[
   z = \frac{\bar{x} - 2,400}{\sigma/\sqrt{n}} = \frac{\bar{x} - 2,400}{s/\sqrt{n}} \approx \frac{\bar{x} - 2,400}{s/\sqrt{n}} = \frac{2,460 - 2,400}{200/\sqrt{50}} = 2.12
   \]

4. **Rejection region:** $z > 1.645$, which corresponds to $\alpha = 0.05$.

5. **Conclusion:** The sample mean lies 2.12$\sigma_z$ above the hypothesized value of $\mu = 2,400$. Since this value of $z$ exceeds 1.645, it falls into the rejection region. That is, we reject the null hypothesis that $\mu = 2,400$ and conclude that $\mu > 2,400$. Thus, it appears that the company’s pipe has a mean strength that exceeds 2,400 pounds per linear foot.

Note: When the null hypothesis contains more than one value of $\mu$, as in this case ($H_0 : \mu \leq 2,400$), we use the value of $\mu$ closest to the values specified in the alternative hypothesis. The idea is that if the hypothesis that $\mu$ equals 2,400 can be rejected in favor of $\mu > 2,400$, then $\mu$ less than or equal to 2,400 can certainly be rejected.
Example 2  If the sample mean breaking strength to for the 50 sections of sewer pipe in Example 1 turned out to be \( \bar{x} = 2,430 \) pounds per linear foot. Assume that the sample standard deviation is still \( s = 200 \). Let’s repeat the test at \( \alpha = 0.05 \).

Solution:

1. Null hypothesis \((H_0)\): \( \mu \leq 2,400 \) (i.e., the manufacturer’s pipe does not meet specifications)

2. Alternative hypothesis \((H_a)\): \( \mu > 2,400 \) (i.e., the manufacturer’s pipe meets specifications)

3. Test statistic:

\[
    z = \frac{\bar{x} - 2,400}{\sigma/\sqrt{n}} \approx \frac{\bar{x} - 2,400}{s/\sqrt{n}} = \frac{2,430 - 2,400}{200/\sqrt{50}} = 1.06
\]

4. Rejection region: \( z > 1.645 \), which corresponds to \( \alpha = 0.05 \).

5. Conclusion: The sample mean \( \bar{x} = 2,430 \) is only 1.06 standard deviations above the null hypothesized value of \( \mu = 2,400 \). This value does not fall into the rejection region \( (z > 1.645) \). Therefore, we know that we cannot reject \( H_0 \) if we use \( \alpha = 0.05 \). Even though the sample mean exceed the specification by enough to provide convincing evidence that the population mean exceeds 2,400.

Note: Concluding that the null hypothesis is true (the pipe does not meet specifications) when in fact it is false (the pipe does meet specifications) is the **Type II decision error**.

### Conclusions and Consequences for a Test of Hypothesis

<table>
<thead>
<tr>
<th>Conclusion</th>
<th>True State of Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0 ) True</td>
<td>( H_0 ) True</td>
</tr>
<tr>
<td>( H_a ) True</td>
<td>( H_a ) True</td>
</tr>
</tbody>
</table>

Note: Be careful not to “accept \( H_0 \)” when conducting a test of hypothesis, since the measure of reliability, \( \beta = P(\text{Type II error}) \), is almost always unknown. If the test statistic does not fall into the rejection region, it is better to state the conclusion as “insufficient evidence to reject \( H_0 \)”.
6.2 Large-Sample Test of Hypothesis about a Population Mean

The null and alternative hypotheses may take one of several forms, a one-tailed (or one-sided) statistical test and a two-tailed (or two-sided) hypothesis.

Steps for Selecting the Null the Alternative Hypotheses

1. Select the alternative hypothesis as that which the sampling experiment is intended to establish. The alternative hypothesis will assume on of three forms:
   a. One tailed, upper tailed
      Example: $H_a: \mu > 2,400$
   b. One tailed, lower tailed
      Example: $H_a: \mu < 2,400$
   a. Two tailed
      Example: $H_a: \mu \neq 2,400$

2. Select the null hypothesis as the status quo – that which will be presumed true unless the sampling experiment conclusively establishes the alternative hypothesis. The null hypothesis will be specified as that parameter value closest to the alternative in one-tailed tests and as the complementary (or only unspecified) value in two-tailed tests.
   Example: $H_0: \mu = 2,400$

Rejection Regions for Common Values of $\alpha$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Lower Tailed</th>
<th>Alternative Hypotheses</th>
<th>Upper Tailed</th>
<th>Two Tailed</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>$z &lt; -1.28$</td>
<td>$z &gt; 2.8$</td>
<td>$z &lt; -1.645$ or $z &gt; 1.645$</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>$z &lt; -1.645$</td>
<td>$z &gt; 1.645$</td>
<td>$z &lt; -1.96$ or $z &gt; 1.96$</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>$z &lt; -2.33$</td>
<td>$z &gt; 2.33$</td>
<td>$z &lt; -2.575$ or $z &gt; 2.575$</td>
<td></td>
</tr>
</tbody>
</table>

Large-Sample Test of Hypothesis about $\mu$

One-Tailed Test

$H_0: \mu = \mu_0$

$H_a: \mu < \mu_0$ (or $H_a: \mu > \mu_0$)

Test statistic: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Rejection region: $z < -z_{\alpha}$

(when $H_a: \mu > \mu_0$)

$P(z < -z_{\alpha}) = \alpha$

Two-Tailed Test

$H_0: \mu = \mu_0$

$H_a: \mu \neq \mu_0$

Test statistic: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$

Rejection region: $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

$P(z < -z_{\alpha/2}) = \alpha/2$

Note: $\mu_0$ is the symbol for the numerical value assigned to $\mu$ under the null hypothesis.

Conditions Required for a Valid Large-Sample Hypothesis Test for $\mu$

1. A random sample is selected from the target population.

2. The sample size $n$ is large (i.e., $n \geq 30$). (Due to the central limit theorem, this condition guarantees that the test statistic will be approximately normal regardless of the shape of the underlying probability distribution of the population.)
Possible Conclusions for a Test of Hypothesis

1. If the calculated test statistic falls into the rejection region, reject \( H_0 \) and conclude that the alternative hypothesis \( H_a \) is true. State that you are rejecting \( H_0 \) at the \( \alpha \) level of significance. Remember that the confidence is in the testing process, not the particular result of a single test.

2. If the test statistic does not fall into the rejection region, conclude that the sampling experiment does not provide sufficient evidence to reject \( H_0 \) at the \( \alpha \) level of significance. (Generally, we will not “accept” the null hypothesis unless the probability \( \beta \) of a Type II error has been calculated.)

Example 3 The effect of drugs and alcohol on the nervous system has been the subject of considerable research. Suppose a research neurologist is testing the effect of a drug on response time by injecting 100 rats with a unit dose of the drug, subjecting each rat to a neurological stimulus, and recording its response time. The mean and standard deviations for the 100 records are \( \bar{x} = 1.05 \) and \( s = 0.5 \) respectively. The neurologist knows that the mean response time for rats not injected with the drug (the “control” mean) is 1.2 seconds. She wishes to test whether the mean response time for drug-injected rats differs from 1.2 seconds. Set up the test of hypothesis for this experiment, using \( \alpha = 0.01 \).

Solution:

1. \( H_0: \mu = 1.2 \)
2. \( H_a: \mu \neq 1.2 \)
3. Test statistic:

\[
z = \frac{\bar{x} - 1.2}{\sigma_{\bar{x}}} = \frac{\bar{x} - 1.2}{s/\sqrt{n}} \approx \frac{\bar{x} - 1.2}{s/\sqrt{n}} = \frac{1.05 - 1.2}{0.5/\sqrt{100}} = -3.0
\]

4. Rejection region: \( z < -2.575 \) or \( z > 2.575 \), which corresponds to \( \alpha = 0.01 \).

5. Conclusion: This sampling experiment provides sufficient evidence to reject \( H_0 \) and conclude, at the \( \alpha = 0.01 \) level of significance, that the mean response time for drug-injected rats differs from the control mean of 1.2 seconds. It appears that the rats receiving an injection of the drug have a mean response time that is less than 1.2 seconds.

Note: Since the sample size of the experiment is large enough \( (n > 30) \), the central limit theorem will apply, and no assumptions need be made about the population of response time measurement. The sampling distribution of the sample mean response of 100 rats will be approximately normal, regardless of the distribution of the individual rats’ response times.
6.3 Observed Significance Levels: p-Values

A second method of presenting the results of a statistical test reports the extent to which the test statistic disagrees with the null hypothesis and leaves to the reader the task of deciding whether to reject the null hypothesis. This measure of disagreement is called the observed significance level (or \textit{p-value}) for the test.

**Definition 6.1**  The observed significance level, or \textit{p-value}, for a specific statistical test is the probability (assuming that \(H_0\) is true) of observing a value of the test statistic that is at least as contradictory to the null hypothesis, and supportive of the alternative hypothesis, as the actual one computed from the sample data.

**Steps for Calculating the \textit{p}-Value for a Test of Hypothesis**

Determine the value of the test statistic \(z\) corresponding to the result of the sampling experiment.

1. If the test is one-tailed, the \textit{p}-value is equal to the tail area beyond \(z\) in the same direction as the alternative hypothesis. Thus, if the alternative hypothesis is of the form \(>\), the \textit{p}-value is the area to the right of, or above, the observed \(z\)-value. Conversely, if the alternative is of the form \(<\), the \textit{p}-value is the area to the left of, or below, the observed \(z\)-value.

2. If the test is two tailed, the \textit{p}-value is equal to twice the tail area beyond the observed \(z\)-value in the direction of the sign of \(z\). That is, if \(z\) is positive, the \textit{p}-value is twice the area to the right of, or above, the observed \(z\) value. Conversely, if \(z\) is negative, the \textit{p}-value is twice the area to the left of, or below, the observed \(z\)-value.

**Reporting Test Results as \textit{p}-values: How to Decide Whether to Reject \(H_0\)**

1. Choose the maximum value of \(\alpha\) that you are willing to tolerate.

2. If the observed significance level (\textit{p}-value) of the test is less than the chosen value of \(\alpha\), reject the null hypothesis. Otherwise, do not reject the null hypothesis.
6.4 Small-Sample Test of Hypothesis about a Population Mean

Refer to section 5.3, when we are faced with making inferences about a population mean from the information in a small sample, two problems emerge:

1. The normality of the sampling distribution for $\bar{x}$ does not follow from the central limit theorem when the sampling size is small. We must assume that the distribution of measurements from which the sample was selected is approximately normally distributed.

2. If the population standard deviation $\sigma$ is unknown, as is usually the case, then we cannot assume that $s$ will provide a good approximation for $\sigma$ when the sample size is small. Instead, we must use the $t$-distribution rather than the standard normal $z$-distribution to make inferences about the population mean $\mu$.

### Small-Sample Test of Hypothesis About $\mu$

<table>
<thead>
<tr>
<th>One-Tailed Test</th>
<th>Two-Tailed Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0 : \mu = \mu_0$</td>
<td>$H_0 : \mu = \mu_0$</td>
</tr>
<tr>
<td>$H_a : \mu &lt; \mu_0$ (or $H_a : \mu &gt; \mu_0$)</td>
<td>$H_a : \mu \neq \mu_0$</td>
</tr>
<tr>
<td>Test statistic: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$</td>
<td>Test statistic: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$</td>
</tr>
<tr>
<td>Rejection region: $t &lt; -t_\alpha$ (or $t &gt; t_\alpha$ when $H_a : \mu &gt; \mu_0$)</td>
<td>Rejection region: $t &lt; -t_{\alpha/2}$ or $t &gt; t_{\alpha/2}$</td>
</tr>
</tbody>
</table>

where $t_\alpha$ and $t_{\alpha/2}$ are based on $(n - 1)$ degrees of freedom.

### Conditions Required for a Valid Small-Sample Hypothesis Test for $\mu$

1. A random sample is selected from the target population.

2. The population from which the sample is selected has a distribution that is approximately normal.

6.5 Large-Sample Test of Hypothesis about a Population Proportion

### Large-Sample Test of Hypothesis about $p$

<table>
<thead>
<tr>
<th>One-Tailed Test</th>
<th>Two-Tailed Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0 : p = p_0$</td>
<td>$H_0 : p = p_0$</td>
</tr>
<tr>
<td>$H_a : p &lt; p_0$ (or $H_a : p &gt; p_0$)</td>
<td>$H_a : p \neq p_0$</td>
</tr>
<tr>
<td>Test statistic: $z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$</td>
<td>Test statistic: $z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$</td>
</tr>
<tr>
<td>where $p_0$= hypothesized value of $p$, $\sigma_{\hat{p}} = \sqrt{p_0q_0/n}$, and $q_0 = 1 - p_0$</td>
<td></td>
</tr>
<tr>
<td>Rejection region: $z &lt; -z_\alpha$ (or $z &gt; z_\alpha$ when $H_a : p &gt; p_0$)</td>
<td>Rejection region: $z &lt; -z_{\alpha/2}$ or $z &gt; z_{\alpha/2}$</td>
</tr>
</tbody>
</table>
Conditions Required for a Valid Large-Sample Hypothesis Test for \( p \)

1. A random sample is selected from a binomial population.
2. The sample size \( n \) is large. (This condition will be satisfied if \( np_0 \) and \( nq_0 \) are both at least 15.)

**Example 4** The reputations (and hence sales) of many businesses can be severely damaged by shipments of manufactured items that contain a large percentage of defectives. For example, a manufacturer of alkaline batteries may want to be reasonably certain that less than 5% of its batteries are defective. Suppose 300 batteries are randomly selected from a very large shipment each is tested and 10 defective batteries are found. Does this outcome provide sufficient evidence for the manufacturer to conclude that the fraction defective in the entire shipment is less than 0.05? Use \( \alpha = 0.01 \).

**Solution:**

1. \( H_0: p = 0.05 \)
2. \( H_a: p < 0.05 \)
3. Test statistic:

\[
    z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{\hat{p} - 0.05}{\sqrt{p_0q_0/n}} = \frac{0.03333 - 0.05}{\sqrt{(0.05)(0.95)/300}} = -1.32
\]

4. Rejection region: \( z < -z_{0.01} = -2.33 \), which corresponds to \( \alpha = 0.01 \).

5. **Conclusion:** The calculated \( z \)-value does not fall into the rejection region. Therefore, there is insufficient evidence at the 0.01 level of significance to indicate that the shipment contains less than 5% defective batteries.

**Note:**

1. Before conducting the test, we should check to determine whether the sample size is large enough to use the normal approximation to the sampling distribution of \( \hat{p} \). Since \( np_0 = (300)(0.05) = 15 \) and \( nq_0 = (300)(0.95) = 285 \) are both at least 15, the normal approximation will be adequate.

2. We use \( p_0 \) to calculate \( \sigma_{\hat{p}} \) because, in contrast to calculating \( \sigma_{\hat{p}} \) for a confidence interval, the test statistic is computed on the assumption that the null hypothesis is true – that is, \( p = p_0 \).