Chapter 9  Tests of significance for one population mean

The Nature of Hypothesis Testing

Hypothesis is a statement about population parameters, like mean, proportion, two means or two proportions. In this chapter we will consider tests for one population mean, but the ideas we will introduce here can be transferred to many other situations.

For example we may be interested in answering following questions regarding one population:

1. Is mean monthly weight gain ($\mu$) of lambs fed a standard diet fortified with niacin larger than monthly mean weight gain of lambs fed a standard diet that is known to be 10 pounds per month?

2. Is mean age ($\mu$) of people in juvenile detention in Arizona in 2008 less than 16, the mean age in 2005.

3. The owner of a construction company would like to know if his current work crew has a mean time to build the deck on the house ($\mu$) different than his past crews that took on average about 28 hours.

A hypothesis test allows us to make a decision of judgment about the value of a parameter (or parameters)

A hypothesis test involves two hypothesis – Null hypothesis and Alternative (Research) hypothesis.

- **Null hypothesis ($H_0$):** A hypothesis to be tested. Generally states that there is no effect present, mean is not different than a given number.

- **Alternative hypothesis ($H_a$):** A hypothesis to be considered as an alternate to the null hypothesis. Often called “research “ hypothesis. Generally it states that there is an effect present, there is a difference.

The choice of the alternative hypothesis depends on and should reflect the purposes of the hypothesis test and depends on question asked in the problem.

- ex1 $H_0: \mu = 10$, $H_a: \mu > 10$ - right-tailed test
- ex2 $H_0: \mu = 16$, $H_a: \mu < 16$ - left-tailed test
- ex3 $H_0: \mu = 28$, $H_a: \mu \neq 28$ - two-tailed test

Hypothesis in ex1 and ex2 are also sometimes called directional and in ex3 nondirectional.
The logic of hypothesis testing

1. Take a random sample from one population (or two samples from 2 populations)
2. If the data do not provide enough evidence in favor of the alternative hypothesis, do not reject the null hypothesis.
3. If the data provide enough evidence in favor of the alternative, reject the null hypothesis.

Test statistic: The statistic is a value computed from a sample (s) used as a basis for deciding whether the null hypothesis should be rejected. Test statistics will have a following general format:

\[
\frac{\text{estimate} - \text{hypothesized value}}{\text{SD of the estimate}}
\]

The distribution of our test statistics (if null hypothesis is true) will be different for different tests, often it will be N(0,1) or t distribution.

Let us consider example#1: our hypotheses are:

\[ \begin{align*}
H_0 &: \text{ mean monthly weight gain on the niacin diet is not different than 10 lb }, \quad \mu = 10 \\
H_a &: \text{ mean monthly weight gain on the niacin diet is larger than 10 lb }, \quad \mu > 10
\end{align*} \]

Suppose SRS of lambs is fed a niacin diet for 3 months and the average monthly weight gain is estimated from the sample to be 12 lb. Suppose also that SD of the estimate is 15lb.

Our test statistics here is then \[ z = \frac{12 - 10}{15} = 0.13 \] and it will have N(0,1) distribution under \( H_0 \)

What is the conclusion of our test at 5% significance level?

Possible conclusions for a hypothesis test

Suppose a hypothesis test is conducted at a small significance level \( \alpha \) = probability of rejecting a true null hypothesis, \( \alpha \) is always selected in advance of seeing data.

1. If the null hypothesis is rejected, we conclude that there is evidence for alternative hypothesis and results are statistically significant at the level \( \alpha \)

2. If the null hypothesis is not rejected, we conclude that the data do not provide sufficient evidence to support the alternative hypothesis and results are not statistically significant at the \( \alpha \) level.

Hypothesis Tests for one Population Mean when \( \sigma \) is known
Obtaining critical values
Suppose a hypothesis test is to be performed at a specified significance level, \( \alpha \). Then the critical value(s) must be chosen so that, if the null hypothesis is true, the probability is \( \alpha \) that the test statistic will fall in the rejection region.

Procedure 9.1: The one-sample z-test for a population mean
Critical Value Approach

ASSUMPTIONS
1. Normal population or large samples
2. \( \sigma \) is known
3. Sample is SRS

STEPS
1. Formulate the null hypothesis is \( H_0 : \mu = \mu_0 \) and select appropriate alternate hypothesis is one of the following:
   - \( H_a : \mu \neq \mu_0 \) (two-tailed)
   - \( H_a : \mu < \mu_0 \) (left-tailed)
   - \( H_a : \mu > \mu_0 \) (right-tailed)
2. Decide on the significance level, \( \alpha \).
3. Compute the value of the test statistic:
   \[ z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \], \( z \sim N(0,1) \) under \( H_0 \)
4. Choose the critical value(s) according to the alternate hypothesis:
   - \( \pm z_{\alpha/2} \) (two-tailed) Rejection Region: \( z \leq -z_{\alpha/2} \) or \( z \geq z_{\alpha/2} \)
   - \( -z_\alpha \) (left-tailed) Rejection Region: \( z \leq -z_\alpha \)
   - \( z_\alpha \) (right-tailed) Rejection Region: \( z \geq z_\alpha \)
5. Decide if null hypothesis should be rejected or not: If the value of the test statistic falls in the rejection region, reject \( H_0 \); otherwise, do not reject \( H_0 \).
6. Interpret the results of the hypothesis test, make sure to clearly answer question posed in the problem

NOTE: The hypothesis test is exact for normal populations and is approximately correct for large samples from non-normal populations.

Following illustrates the rejection regions for a z-test.
Back to our example #1. If we use 5% significance level, our rejection region is: 
\[ z \geq 1.645 \]. Our test statistics \( z = 0.13 \) does not fall into the rejection region, so we do not reject null hypothesis. There is no evidence for alternative hypothesis at \( \alpha = 0.05 \). We have no evidence that monthly weight gain \( (\mu) \) of lambs fed a standard diet fortified with niacin is larger than 10 pounds per month.

**Procedure 9.2: The one-sample z-test for a population mean (p-value approach)**

**ASSUMPTIONS**
1. Normal population or large sample
2. \( \sigma \) is known
3. Sample is SRS

**STEPS**
1. Formulate the null hypothesis is \( H_0 : \mu \neq \mu_0 \) and select appropriate alternative hypothesis is one of the following:
   
   \( H_a : \mu \neq \mu_0 \) (two-tailed)
   
   \( H_a : \mu < \mu_0 \) (left-tailed)
   
   \( H_a : \mu > \mu_0 \) (right-tailed)

2. Decide on the significance level, \( \alpha \).

3. Compute the value of the test statistic \( z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = z_0 \) and denote the value by \( z_0 \),

   \( z \sim N(0,1) \) under \( H_0 \)

4. Use z-tables to obtain the P-value
   
   (two-tailed) \( p\text{-value}= \text{area under } N(0,1) \text{ over following regions: } z \leq -|z_0| \text{ and } z \geq |z_0| \)
   
   (left-tailed) \( p\text{-value}= \text{area under } N(0,1) \text{ over following region: } z \leq z_0 \)
   
   (right-tailed) \( p\text{-value}= \text{area under } N(0,1) \text{ over following region: } z \geq z_0 \)

5. Decide if null hypothesis is rejected or not: If \( P \leq \alpha \) reject \( H_0 \); otherwise, do not reject \( H_0 \).

6. State the conclusion in words, make sure to clearly answer question posed in the problem.

The hypothesis test is exact for normal populations and is approximately correct for large samples from non-normal populations.

**Following illustrates p-value computations for a z-test**

![Diagram showing p-value computations for a z-test](image-url)

(a) Two-tailed  
(b) Left-tailed  
(c) Right-tailed
P-Value Approach to tests of significance.

**P-Value:** To obtain a P-value (P) of a hypothesis test, we compute, assuming the null hypothesis is true, the probability of observing a value of the test statistic as extreme or more extreme than the one observed. By extreme we mean far from what we would expect to observe if the null hypothesis were true.

If test is **left tailed**, extreme means equal or much smaller than test statistics observed, if test is **right tailed**, extreme means equal or much larger than test statistics observed, if test is **two tailed**, extreme means either equal or larger than absolute value of test statistics observed or equal or smaller than negative of absolute value of it.

P-value: referred to as **observed significance level** or probability value.

**How to decide if \( H_0 \) should be rejected or not:**

If \( p \)-value \( \leq \alpha \), \( H_0 \) should be rejected. The smaller is the \( p \)-value, the stronger is evidence against null hypothesis.

**Now let us go back again to example #1 and consider \( p \)-value approach this time:**

p-value for our test = probability of our test statistics to be 0.13 or larger if null hypothesis is true = area right of our test statistics under \( N(0,1) = 0.4483 > 0.05 \)

We have have close to 45% chance to receive the test statistics as large or larger than the one we computed from our data, if null hypothesis was true. That is a large chance, our test statistics is not unusual if null is true, so we have no evidence against \( H_0 \)

**Conclusion:** \( H_0 \) not rejected, we have no evidence that average weight on niacin diet exceeds 10 pounds. Results are not significant at 5% significance level

**Guidelines for using the \( P \)-value to assess the evidence against \( H_0 \)**

<table>
<thead>
<tr>
<th>P-value</th>
<th>Evidence against ( H_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P &gt; 0.10 )</td>
<td>Weak or none</td>
</tr>
<tr>
<td>( 0.05 &lt; P \leq 0.10 )</td>
<td>Moderate</td>
</tr>
<tr>
<td>( 0.01 &lt; P \leq 0.05 )</td>
<td>Strong</td>
</tr>
<tr>
<td>( P \leq 0.01 )</td>
<td>Very Strong</td>
</tr>
</tbody>
</table>
Comparison of the Critical-value and p-value approaches

<table>
<thead>
<tr>
<th>CRITICAL-VALUE APPROACH</th>
<th>P-VALUE APPROACH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. State $H_0$ and $H_a$</td>
<td>1. State $H_0$ and $H_a$</td>
</tr>
<tr>
<td>2. Decide on $\alpha$</td>
<td>2. Decide on $\alpha$</td>
</tr>
<tr>
<td>3. Compute test statistic</td>
<td>3. Compute test statistic</td>
</tr>
<tr>
<td>4. Determine the critical value(s)</td>
<td>4. determine the P-value</td>
</tr>
<tr>
<td>5. If test statistic falls in rejection region, reject $H_0$; otherwise, do not reject $H_0$.</td>
<td>5. If $P \leq \alpha$ reject $H_0$; otherwise, do not reject $H_0$.</td>
</tr>
<tr>
<td>6. Interpret the result of the hypothesis test.</td>
<td>6. Interpret the result of the hypothesis test.</td>
</tr>
</tbody>
</table>

Let us consider two other examples from the beginning of the section:

**Ex2.** Is mean age ($\mu$) of people in juvenile detention in Arizona in 2008 less than 16, the mean age in 2005. Test at 5% significance level.

Suppose we know that standard deviation of all ages of people in juvenile detention in 2008 in AZ is $\sigma = 3.5$ years. SRS of 24 individuals had mean $\bar{X} = 15.3$ years.

Our hypothesis are:

$H_0 : \mu = 16 \quad H_a : \mu < 16$, need to assume normal population.

notice that $\mu$ is defined in the problem, otherwise student has to define it.
**STP226 Brief Class Notes, Instructor: Ela Jackiewicz**

\[ z = \frac{15.3 - 16}{3.5/\sqrt{24}} = -.98 = z_0 \]

**Critical value approach:** CV = \(-z_{0.05} = -1.645\), -0.98 is not in the rejection region, so \(H_0\) is not rejected, no evidence for \(H_a\) at 5% significance level.

**P-value:** P value is the area under \(N(0,1)\) left of -0.98, from the tables p-value = 0.1635 (calculator: \(\text{normalcdf}(-10^6, -0.98)\)). Since .1635 > .05, \(H_0\) is not rejected, no evidence for alternative hypothesis at 5% significance level.

Conclusion: Mean age of people in juvenile detention in AZ in 2008 is not greater than 16.

**Ex3.** The owner of a construction company would like to know if his current work crew takes on average different time to build a deck on a house than his past crews. He knows that the average time it took his past crews to build a deck is 28 hours. A random sample of 19 decks that the current crew has built resulted in a sample mean of building time of 31 hours. Assume standard deviation of times to build the deck for all jobs of the new crew is 6 hours. Conduct the appropriate hypotheses test using the \( \alpha = 0.05 \). Assume that times it takes his current crew to build the decks are normally distributed.

Our hypothesis are below, \( \mu \) mean time for a new crew, need to assume normal population. Notice that \( \mu \) needed to be defined.

\[ H_0 : \mu = 28 \quad H_a : \mu \neq 28 \quad z = \frac{31 - 28}{6/\sqrt{19}} = 2.18 = z_0 \]

**Critical value approach:** CV-s = \(\pm z_{0.025} = \pm 1.96\), 2.18 is in the rejection region, so \(H_0\) is rejected, we have evidence for \(H_a\) at 5% significance level.

**P-value:** P value is the area under \(N(0,1)\) left of -2.18 and right of 2.18, from the tables: p-value = 2*0.0146 = 0.0292 (calculator: 2* \(\text{normalcdf}(2.18, 10^6)\) = .0296). Since \(p\)-value < .05, \(H_0\) is rejected, there is evidence for alternative hypothesis at 5% significance level.

Conclusion: Mean time for new crew to build the deck is different from previous crews.

**The Relation Between Hypothesis Tests and Confidence Intervals**

If Null hypothesis \( H_0 : \mu = \mu_0 \) against two tailed alternative is rejected at significance level \( \alpha \), then (1- \( \alpha \))\(*100\%\) CI for \( \mu \) will not contain \( \mu_0 \), otherwise (if not rejected) \( \mu_0 \) will be inside of CI.

In Ex 3. 95% CI for \( \mu \) is (28.3, 33.7) and does not contain 28, consistent with rejecting \(H_0\).
When to use the z-test

1. For small samples, say, of size less than 15, the z-test should be used only when the variable under consideration is normally distributed or very close to being so.

2. For moderate size samples, say, between 15 and 30, the z-test can be used unless the data contains outliers or the variable under consideration is far from being normally distributed.

3. For large samples, say, of size 30 or more, the z-test can be used essentially without restriction. However, if outliers are present and their removal is not justified, the effect of the outliers on the hypothesis test should be examined; that is, the hypothesis test should be performed twice, once with the outliers retained and once with them removed. If the conclusion remains the same either way, we may be content to take that as our conclusion and close the investigation. But if the conclusion is affected, it is probably wise to make the more conservative, use a different procedure, or take another sample.

4. If outliers are present but their removal is justified and results in a data set for which the z-test is appropriate, then the procedure can be used.

Cautions about hypothesis tests:

**Statistical significance does not necessary imply practical significance!**

If we discover that a small difference is statistically significant, which may happened if we use a very large n, it will not mean that that difference has any practical meaning, for example a difference of 1 lb between weight gain (over 3 months period) on two different diets for lambs may show significance if large samples were used, but does it mean that one of the diets is significantly better?

On the other hand **lack of statistical significance should not always be ignored.** Suppose the difference does not show to be significant at 5% (p-value=0.052) but if we used only samples of size 5, this p-value is very close to 5%, maybe with larger sample we will be able to show that there is also a statistically significant difference.

Not all data sets produce valid results. **We must have randomization** of samples or experiments. If subjects in an experiment are not randomly assigned to treatments our statistical inference will not be valid.

**Observations in our samples must be independent,** lack of independence will invalidate our conclusions. For example observations will not be independent if we use as subjects offsprings of the same mother (mice from the same litter for example)

**We should not blindly search for significance.** The inference works best if we decide in advance what effect we are seeking and design experiments to search for it. If we just perform 1000 random experiments, just by chance 5% of them may show statistical significance where in fact there is no effect.
9.5 Hypothesis Tests for one population mean when $\sigma$ is unknown

If a variable $x$ of a population is normally distributed with mean $\mu$, then the studentized version of $\bar{x}$, $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$, has t-distribution with n-1 degrees of freedom.

**Procedure 9.3:** The one-sample $t$-test for a population mean

**Critical Value Approach**

**ASSUMPTIONS**
1. Normal population or large samples
2. $\sigma$ is unknown
3. Sample is SRS

**STEPS**
1. Formulate the null hypothesis is $H_0 : \mu = \mu_0$ and select appropriate alternate hypothesis is one of the following:

   $H_a : \mu \neq \mu_0$ (two-tailed)
   $H_a : \mu < \mu_0$ (left-tailed)
   $H_a : \mu > \mu_0$ (right-tailed)

2. Decide on the significance level, $\alpha$.

3. Compute the value of the test statistics $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$, $t$-distribution with n-1 df under $H_0$

4. Choose the critical value(s) according to the alternate hypothesis:

   $\pm t_{\alpha/2}$ (two-tailed) Rejection Region: $t \leq -t_{\alpha/2}$ or $t \geq t_{\alpha/2}$
   $-t_{\alpha}$ (left-tailed) Rejection Region: $t \leq -t_{\alpha}$
   $t_{\alpha}$ (right-tailed) Rejection Region: $t \geq t_{\alpha}$

5. Decide if null hypothesis should be rejected or not: If the value of the test statistic falls in the rejection region, reject $H_0$; otherwise, do not reject $H_0$.

6. Interpret the results of the hypothesis test, make sure to clearly answer question posed in the problem.

**NOTE:** The hypothesis test is exact for normal populations and is approximately correct for large samples from non-normal populations.
Procedure 9.4: The one-sample t-test for a population mean (p-value approach)

ASSUMPTIONS
1. Normal population or large sample
2. $\sigma$ is unknown
3. Sample is SRS

STEPS
1. Formulate the null hypothesis is $H_0: \mu = \mu_0$ and select appropriate alternative hypothesis is one of the following:
   - $H_a: \mu \neq \mu_0$ (two-tailed)
   - $H_a: \mu < \mu_0$ (left-tailed)
   - $H_a: \mu > \mu_0$ (right-tailed)
2. Decide on the significance level, $\alpha$.
3. Compute the value of the test statistic $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = t_0$ and denote the value by $t_0$,
   - $t \sim t$ distribution with n-1 df under $H_0$
4. Use t-tables to obtain the P-value
   - (two-tailed) $p$-value= area under $t$ curve over following regions: $t \leq -|t_0|$ and $t \geq |t_0|$
   - (left-tailed) $p$-value= area under $t$ curve over following region: $t \leq t_0$
   - (right-tailed) $p$-value= area under $t$ curve over following region: $t \geq t_0$
5. Decide if null hypothesis is rejected or not: If $P \leq \alpha$ reject $H_0$; otherwise, do not reject $H_0$.
6. State the conclusion in words, make sure to clearly answer question posed in the problem.

The hypothesis test is exact for normal populations and is approximately correct for large samples from non-normal populations.

---

**Note:**

If $P \leq \alpha$, reject $H_0$; otherwise, do not reject $H_0$. The hypothesis test is exact for normal populations and is approximately correct for large samples from non-normal populations.
STP226 Brief Class Notes, Instructor: Ela Jackiewicz

P-value for a t-test

- Two-tailed test: The P-value is the probability of observing a value of the test statistic $t$ at least as large in magnitude as the value actually observed, which is the area under the t-curve that lies outside the interval from $-|t_0|$ to $|t_0|$.

- Left-tailed test: The P-value is the probability of observing a value of the test statistic $t$ as small as or smaller than the value actually observed, which is the area under the t-curve that lies to the left of $t_0$.

- Right-tailed test: The P-value is the probability of observing a value of the test statistic $t$ as large as or larger than the value actually observed, which is the area under the t-curve that lies to the right of $t_0$.

Estimating the P-value of a t-test from the t-table

Example 1: Consider a right-tailed t-test with $n=15$, $\alpha=0.05$, and the value of the test statistic of $t=3.458$.

Solution: For $df=15-1=14$, the t-value 3.458 is larger than any t-value in Table IV, the largest one being $t_{0.005} = 2.977$. So the P-value $<0.005$.

Example 2: Consider a left-tailed t-test with $n=25$, $\alpha=0.05$, and the value of the test statistic of $t=-1.156$.

Solution: We need to estimate area right of positive $t=1.156$, since t-table only showe right tails. For $df=25-1=24$, the t-value is smaller than any t-value in Table IV, the smallest one being $t_{0.10} = 1.318$. So the P-value $>0.10$.

Example 3: Consider a two-tailed t-test with $n=20$, $\alpha=0.05$, and the value of the test statistic of $t=2.311$.

Solution: We need to estimate area right of both tails, left of $-2.31$ and right of 2.31, since t-table only showe right tails, we estimate first right tail and then we will double it. For $df=20-1=19$, the t-value is between $t_{0.025} = 2.093$ and $t_{0.01} = 2.539$, so $0.01 < \frac{p-value}{2} < 0.025$ and $0.02 < p-value < 0.05$.

Example t-test. The amount of vitamine C (in milligrams per 100 grams) was measured in 8 samples of CSB (corn-soy blend).

Our data is: 26, 31, 23, 22, 11, 22, 14, 31

We are interested to know if mean vitamine C content in CSB is equal to the required 30 milligrams per 100 grams or not. Let us use 5% significance level. Our test is then:
\[ H_0 : \mu = 30 \quad H_a : \mu \neq 30 \]

\[ t = \frac{22.5 - 30}{7.19 / \sqrt{8}} = -2.95 \]

Using **p-value approach**: From the table \(0.01 < \frac{1}{2}(p\text{-value}) < 0.025\), so \(0.02 < p\text{-value} < 0.05\) \((df=7)\)

Null hypothesis is rejected, we have evidence for alternative hypothesis. Mean vitamine content in CSB is different than required 30 milligrams per 100 grams (implication is that it is less than 30)

Using **critical value approach**, our cv-s are: \(\pm t_{0.025} = \pm 2.365\), -2.95 is clearly in the rejection region, so we have the same conclusion as above.

**Type I Error & Type II Errors**

<table>
<thead>
<tr>
<th>(H_0) is True</th>
<th>(H_0) is False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do not reject (H_0)</td>
<td>Correct Decision</td>
</tr>
<tr>
<td>Reject (H_0)</td>
<td>Type I error</td>
</tr>
</tbody>
</table>

- **Type I error**: Rejecting the null hypothesis when it is in fact true.
- **Type II error**: Not rejecting the null hypothesis when it is in fact false.

**Probabilities of Type I and Type II Errors**

- **Significance level** \(\alpha\): The probability of making a Type I error (rejecting a true null hypothesis), always selected in advance of seeing data.

- **Power** of a hypothesis test: \(\text{Power} = 1 - \text{P(Type II error)} = 1 - \beta = 1 - \text{The Probability of rejecting a false null hypothesis}\)
  
  Power near 0: the hypothesis test is not good at detecting a false null hypothesis.
  
  Power near 1: the hypothesis is extremely good at detecting a false \(H_0\)

**Relation between Type I and Type II error probabilities**

For a fixed sample size, the smaller we specify the significance level, \(\alpha\), the larger will be the probability, \(\beta\), of not rejecting a false null hypothesis

**Tests on the calculator:**

**Z test procedure:**

\(\sigma\) is known, normal population or a large sample)
use STAT menu then TESTS
option 1 is Z-test

Ex Test $H_0: \mu = 15$ vs $H_a: \mu > 15$, use $\alpha = .05$,
data gives $\bar{X} = 17$, $n=10$ and $\sigma = 4$, use

Z-test
Inpt: Stats
$\mu_0 = 15$
$\sigma: 4$
$\bar{X}: 17$
n: 10
$\mu: > \mu_0$

hit Enter, you get $z=1.58$, $p=.056$ (p-value), do not reject $H_0$ since $p > .05$

(You can use Data options in a similar way you use it in Z-Interval procedures, see examples in ch8 notes)