IX. Tests for Means: One Sample problems

Scenario: You will have sample observations from an experimental population and census data from a control population. The observations will be appropriate for computing means and standard deviations. Observations on the experimental group have been taken only once, after exposure to the experimental treatment.

√ For the *Experimental Population* you will have only sample data:

	Parameters (unknown)	**Sample Estimates (known)**
mean	$\boldsymbol{\mu}_1$	$\bar{\boldsymbol{x}}$
standard deviation	$\boldsymbol{\sigma}_1$	$\boldsymbol{s}$
sample size	–	$\boldsymbol{n}$

Sometimes you will know the standard deviation of the control population; when this happens you do not need to know the standard deviation of the sample.

√ For the *Control Population* you will have census data:

	Parameter (known)
mean	$\boldsymbol{\mu}_0$
standard deviation	$\boldsymbol{\sigma}_0$

Sometimes you will not know the standard deviation of the control population. When this happens, you will need to use the standard deviation of the sample instead.

√ Research Objective: Determine if the observed relationship between the observed means of the experimental sample and the control populations is due to

- the experimental treatment; or

- random chance inherent in sampling.

Solution Template

Step 1. Make a dictionary assigning values to each of the variables:

sample size	$\boldsymbol{n}$
sample mean	$\bar{\boldsymbol{x}}$
standard deviation	either $\boldsymbol{s}$ or $\boldsymbol{\sigma}_0$
mean of control population	$\boldsymbol{\mu}_0$
significance level	$\boldsymbol{\alpha}$

(Your dictionary can either contain the sample standard deviation $\boldsymbol{s}$ or the standard deviation of the control population $\boldsymbol{\sigma}_0$ depending on the data given in the problem.)

Step 2. Write down the null and alternative hypotheses. The null hypothesis will always be:

$$\boldsymbol{H}_0 : \quad \boldsymbol{\mu}_1 = \boldsymbol{\mu}_0$$

while the alternative hypothesis will be one of the following:

$$\boldsymbol{H_A} : \ \boldsymbol{\mu}_1 < \boldsymbol{\mu}_0 \quad \textit{(a left tailed test)}$$
$$\boldsymbol{H_A} : \ \boldsymbol{\mu}_1 > \boldsymbol{\mu}_0 \quad \textit{(a right tailed test)}$$
$$\boldsymbol{H_A} : \ \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_0 \quad \textit{(a two tailed test).}$$

(The reason for the terms right, left and two tailed tests is given in step 5.)

Step 3. Compute the value of the test statistic:

$$\frac{\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0}{\boldsymbol{s}/\sqrt{\boldsymbol{n}}}.$$

This test statistic is normally distributed if the sample size is larger than 30; otherwise it has Student's $\boldsymbol{t}$-distribution. (If you know the value of $\boldsymbol{\sigma}_0$ then the test statistic is

$$\frac{\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0}{\boldsymbol{\sigma}_0/\sqrt{\boldsymbol{n}}}$$

substituting $\boldsymbol{\sigma}_0$ for $\boldsymbol{s}$ in the formula. If you use this second formula, the test statistic is normally distributed.)

$\boxed{\textit{Step 4.}}$ Find the cutoff(s) using the table in Appendix B.

- Find the significance level in the first column in the table.

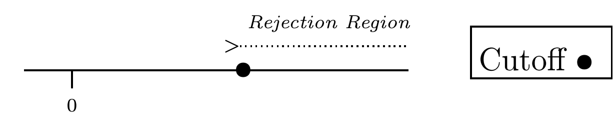
- Determine which kind of test you are running from the alternative hypothesis; find the corresponding column at the top of the table.

- The entry in the intersection of the above row and column is the cutoff corresponding to the test. If you are running a two tailed test, you will actually find two cutoffs.

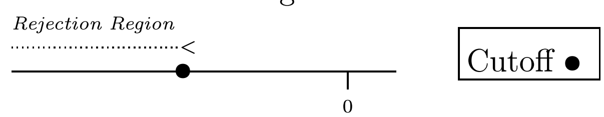
$\boxed{\textit{Step 5.}}$ Decision rules (it often helps to plot the cutoff(s) and the value of the test statistic on a number line):

- First plot the cutoff(s) and the rejection region(s) on a number line.

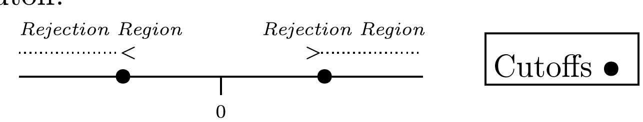
  - If $\boldsymbol{H_A}: \quad \boldsymbol{\mu}_1 > \boldsymbol{\mu}_0$: This is called a *right-tailed* test since the rejection region is to the right of the cutoff.



  - If $\boldsymbol{H_A}: \quad \boldsymbol{\mu}_1 < \boldsymbol{\mu}_0$: This is called a *left-tailed* test since the rejection region is to the left of the cutoff. Don't forget that the cutoff will be negative.



  - If $\boldsymbol{H_A}: \quad \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_0$: This is called a *two-tailed* test since there are two rejection regions: one to the left of the negative cutoff and one to the right of the positive cutoff.



- Plot the value of the test statistic on the number line. Reject the null hypothesis if the value of the test statistic is in the rejection region. (For two tailed tests, reject $\boldsymbol{H}_0$ if the test statistic falls in *either* of the two rejection regions.)

Of course, if you *reject* the null hypothesis then you *accept* the alternative hypothesis. If you *do NOT reject the null hypothesis* then you *accept* the null hypothesis and *reject* the alternative hypothesis.

End of Solution Template

Interpretation: Hypothesis tests let you choose between accepting and rejecting the null hypothesis. No matter which decision you make you have possibly made an error. The following table shows exactly when those errors can occur.

	$\boldsymbol{H}_0$ true	$\boldsymbol{H_A}$ true
accept $\boldsymbol{H}_0$	OK	Type II error
reject $\boldsymbol{H}_0$	Type I error	OK

If your decision is to reject the null hypothesis, then you have possibly made a Type I error. The significance level $\boldsymbol{\alpha}$ is the largest chance of Type I error which you are willing to tolerate. The researcher is responsible for determining what this level should be, but there are some generally accepted standards for what various levels of $\boldsymbol{\alpha}$ should mean:

$\boldsymbol{\alpha} > 5\%$	results are not statistically significant
$1\% < \boldsymbol{\alpha} \leq 5\%$	results are statistically significant
$\boldsymbol{\alpha} \leq 1\%$	results are highly significant

While the researcher is free to assign any value to $\boldsymbol{\alpha}$, deviation from the above standards should be carefully justified.

If your decision is to accept the null hypothesis you have potentially made a Type II error. This kind of error is not generally controlled in hypothesis testing. This is because hypothesis tests are designed to be conservative: we believe that the experimental treatment made no difference unless there is compelling evidence to the contrary. It also turns out to be much more difficult theoretically to compute the chances of Type II error.

Hints on how to determine the alternative hypothesis. Which form $\boldsymbol{H_A}$ takes is determined by your research objective. You can often guess which form is appropriate for $\boldsymbol{H_A}$ by looking at the relationship suggested by the data: suppose for example that $\bar{x} < \mu_0$. If $\boldsymbol{H_A}$ were $\mu_1 > \mu_0$ there would be no reason to run a complicated statistical test since the data do not even appear to support the alternative. For this reason, if $\bar{x} < \mu_0$ then $\boldsymbol{H_A}$ is probably $\mu_1 < \mu_0$. Similar reasoning applies if $\bar{x} > \mu_0$, suggesting $\boldsymbol{H_A}$ is probably $\mu_1 > \mu_0$.

Generally hypothesis tests involve an experimental treatment. The researcher performs the experiment because the researcher suspects that the treatment will increase or decrease the response measured by the mean. For this reason, in most cases the researcher will use one-sided rather than two sided alternatives.

Assumptions. You must assume that the control and experimental populations have the same standard deviation and that the sample size is at least 30.

Example. Patients with health insurance will remain in a certain hospital an average of 8.6 days after a heart attack. You have selected 68 patients who were hospitalized in this hospital for a heart attack and who had no insurance coverage. These patients had an average hospital stay of 8.3 days with a standard deviation of 1.11 days.

The state hospital board suspects that this hospital is discriminating against patients without insurance by discharging them sooner than patients with insurance. Do the above data support this suspicion at the $\alpha = 4\%$ significance level?

Solution.

Step 1. We first need to identify the parameters:

sample size	n	68
sample mean	$\bar{x}$	8.3
standard deviation	s	1.11
mean of control population	μ_0	8.6
significance level	α	4%

Step 2. Next write down the null and alternative hypotheses:

$$
\begin{aligned}
H_0 &: \mu_1 = 8.6 \\
H_A &: \mu_1 < 8.6
\end{aligned}
$$

Note that the hospital is innocent until proven guilty; thus we will support $\boldsymbol{H_A}$ only if there is compelling evidence to do so.

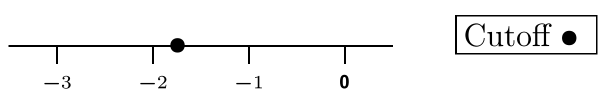
Step 3. Next compute the test statistic:

$$\begin{aligned}
\text{test statistic} &= \frac{\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0}{(\boldsymbol{s}/\sqrt{\boldsymbol{n}})} \\
&= \frac{8.3 - 8.6}{(1.11/\sqrt{68})} \\
&= \frac{-0.3}{(1.11/8.25)} \\
&= \frac{-0.3}{0.1345} \\
&= -2.23
\end{aligned}$$

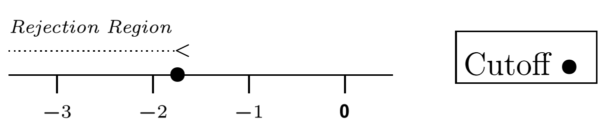
and hence the test statistic is -2.23.

Step 4. This is a 4% left tailed test, so the cutoff is -1.750.

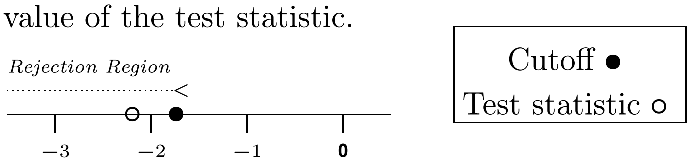
Step 5. To help make the decision whether or not to reject the null hypothesis, *first draw a number line* and plot the cutoff from step 4.



Next find the rejection region.



Finally, plot the value of the test statistic.



Since the test statistic is to the left of the cutoff, we *reject* the null hypothesis and *accept* the alternative hypothesis. This means that we believe that the hospital is in discriminating against those without insurance. The chance that this belief is in error is 4%.

Remark on p-values. Reexamining column three of the table in Appendix B, we see that we could have specified a significance level as low as 2% since the corresponding cutoff (-2.053) would have still resulted in rejecting the null hypothesis. However, a significance level of 1% would have resulted in *accepting* the null hypothesis. This means that 2% is a "threshold" value for $\boldsymbol{\alpha}$, found by comparing the actual value of the test statistic with values in the table. This threshold thus contains more information than just the "accept" or "reject" decision and is often reported as the "$\boldsymbol{p}$" value of the test.

Question 1. Are there other uncontrolled factors which may have influenced the length of the hospital stay and thus the conclusions reached?

Question 2. Can you conclude that the hospital is discriminating against uninsured patients suffering from other ailments?

Remark on small samples. The above methods will work with only minor change for small samples provided that the underlying experimental and control populations are normally distributed. The only difference in the methodology is in step four which becomes:

Step 4. Find the value of the cutoff by looking in the appropriate row and column of the table in Appendix C. The precise way you do this depends on the alternative hypothesis and the sample size.

- Find the degrees of freedom. If the sample size is larger than 150 *OR* if the population standard deviation $\boldsymbol{\sigma}_0$ is known, use $\boldsymbol{\infty}$ for the degrees of freedom. If the sample size is 150 or smaller *AND* if the population standard deviation $\boldsymbol{\sigma}_0$ is unknown, then

$$\text{degrees of freedom} = \text{sample size} - 1$$

Locate the row in the table in Appendix B which corresponds to the degrees of freedom.

- If $\boldsymbol{H_A}: \quad \boldsymbol{\mu}_1 > \boldsymbol{\mu}_0$, select the column by identifying the significance level at the *top* of the table. The cutoff is the entry in the row and column just identified.

- If $\boldsymbol{H_A}: \quad \boldsymbol{\mu}_1 < \boldsymbol{\mu}_0$, select the column by identifying the significance level at the *top* of the table. The cutoff is the negative of the entry in the row and column just identified.

- If $H_A: \quad \mu_1 \neq \mu_0$, select the column by identifying the significance level at the *bottom* of the table. The cutoffs are the positive and negative values of the entry in the row and column just identified.

While it is generally accepted that the table in Appendix B can be used for samples larger than 30, there remains a fairly substantial difference in the cutoffs until the sample size reaches about 150. For this reason we have made Appendix C sufficiently complete to handle samples of up to size 150 before necessitating a resort to the normal cutoffs given in Appendix B.

Problems

1. In a sample of size 57, the sample mean is 93 and the sample standard deviation is 17. Test the null hypothesis that

$$H_0: \ \mu = 89$$

against the alternative that

$$H_A: \ \mu > 89$$

using a signficance level of 5%.

2. In a sample of size 83, the sample mean is 58.9 and the sample standard deviation is 9.3. Test the null hypothesis that

$$H_0: \ \mu = 56.8$$

against the alternative that

$$H_A: \ \mu > 56.8$$

using a signficance level of 5%.

3. In a sample of size 283, the sample mean is 5.7 and the sample standard deviation is 2.3. Test the null hypothesis that

$$H_0: \ \mu = 5.9$$

against the alternative that

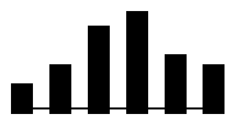
$$H_A: \ \mu < 5.9$$

using a signficance level of 3%.

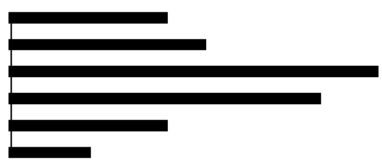
4. Patients suffering from moderate hypertension at a particular clinic are given a standard counselling therapy (dealing with exercise, stress reduction and diet). A researcher suspects that pet ownership may have a beneficial impact on these patients as well. To test this hypothesis, she selects 112 patients and, in addition to the standard therapy, gives each of them a pet to hold for 20 minutes a day.

Patients receiving the standard therapy will, on average, experience a reduction of 20.3 mgHG in their blood pressure readings at the end of six months. Patients in our researcher's sample experienced a reduction of 21.8 mg HG with a standard deviation of 9.4 mg HG. Has the researcher gather convincing evidence (at a significance level of 4%) that her "pet theory" regimen assists in the reduction of hypertension?

5. Histograms are conventionally presented with bars running up and down and arranged left to right across the page like so:



A researcher suspects that drawing histograms with the bars running horizontally and arranged from top to bottom on the page like this



will improve comprehension of the reader.

To test this, the researcher redraws the histograms on a standardized test of mathematical reasoning so that the bars run horizontally and are arranged from top to bottom on the page. The researcher then administers the test to 212 randomly selected subjects. In the sample, the average score was 62.8 with a standard deviation of 12.3. On the test with conventionally drawn histograms, the average score is 64.8.

(i) Has the researcher gathered evidence (at the 4% level) to support the original hypothesis?

(ii) Another researcher claims that this experiment actually proves that the conventional method of drawing histograms improves reader comprehension. Does the experiment support the second researcher's conjecture at the 5% level?

Video Assignment.

View the following program(s) from the series *Against All Odds*:

Program	*Title*
20	*Significance Tests*
21	*Inference for one Mean*