Sec. 6.2: Tests of Significance (Hypothesis Testing)

Statistical inference provides methods for drawing conclusions about a population from sample data. Two of the most common types of statistical inference:

1) **Confidence intervals**
   Goal is to estimate a population parameter.

2) ______________________________
   Goal is to assess the evidence provided by the data about some claim concerning the population.

A ______________________________ is a formal procedure for comparing observed data with a __________________ whose truth we want to assess.

The __________________________ is a statement about the __________________ in a population.

The results of a test are expressed in terms of a __________________________ that measures how well the data and __________________ agree.

**Basic Idea:** an outcome that would __________ happen if a __________ was _______ is good evidence that the hypothesis is false.

**Example:** PCBs in the drinking water?
EPA regulations allow for up to ______ ppm of PCBs in drinking water. A small city is worried about the PCB levels in their water supply and takes _____ independent samples to be measured. They find an average PCB level of _____ ppm in these _____ samples with a standard deviation of_______ ppm.

Is this strong evidence that the city's water supply __________________ standards?
Let's look at the data and determine if this claim is true.

Assume the distribution is __________ l with $\sigma = 2$ (in Chapter 7, we'll see how to handle the case where we estimate $\sigma$ from the data but let's say we know it for now).

What is the probability of observing a sample mean of ________ or_________ if we assume the true PCB level is actually ________?

**Steps to Developing a __________________________**

1) **State __________________________**
   State your research question as two __________________, the _____, and the __________ hypotheses. Remember that these are written in terms of the __________________________!!
The **null hypothesis** \((H_0)\) is the statement being tested. This is assumed \(\text{_________}\) and compared to the data to see if evidence exists \(\text{_________}\) it. It is often a statement that there is \(\text{_________}\) new or unusual- or there is "\(\text{_________}\)."

Suppose we want to test the null hypothesis that \(\mu\) is some specified value, say \(\mu_0\). Then \(H_0\) is \(\text{_________}\):

\[ H_0: \quad \mu = \mu_0 \]

The **alternative hypothesis** \((H_a)\) is the statement about the population parameter that we \(\text{_________}\). \(H_a\) can be \(\text{_________}-sided\) \(\text{or}\) \(\text{_________}-sided\).

Q: What are \(H_0\) and \(H_a\) in the PCB example?

2) **Calculate** \(\text{_________}\) \(\text{statistic}\)

A **statistic** measures the compatibility between the null hypothesis and the data. It is used for the \(\text{_________}\) calculation needed for the test.

The test statistic is a \(\text{_________}\) variable with a \(\text{_________}\) distribution. Use the values for the parameters specified in the null hypothesis, \(H_0\).

In many situations, the test statistic has the form

\[
z = \frac{\text{estimate} - \text{hypothesized value}}{\text{standard deviation of the estimate}}
\]

When testing for a population mean \(\mu\), the test statistic is the **standardized** sample mean \(z\):

\[
z = \frac{\text{estimate}}{\sigma/\sqrt{n}}
\]

Q: What is the value of our test statistic in the previous PCB example?

3) Decide whether the test statistic is "large enough" to count as evidence against the null hypothesis. To do this, we compare \(z\) to a "critical value": \(z^*\). The form of the comparison depends on what the alternative hypothesis is. The rule is: Reject \(H_0\) when the observed value of \(z\), call it \(z_{\text{obs}}\), is in the Rejection Region, or "critical region" defined by:

For \(H_a: \mu > \mu_0\), we reject \(H_0\) when \(z_{\text{obs}} > z^*\)

For \(H_a: \mu < \mu_0\), we reject \(H_0\) when \(z_{\text{obs}} < z^*\)

For \(H_a: \mu \neq \mu_0\), we reject \(H_0\) when \(|z_{\text{obs}}| > z^*\), i.e.

\[ \text{when either } z_{\text{obs}} < -z^* \text{ or } z_{\text{obs}} > z^* \]
Since we expect $z$ to be near zero when $H_0$ is true, this rule rejects when $z$ is far from zero.

**How do we choose $z^*$?**

We want to reject the hypothesis only when we have observed a value that is "rare" if $H_0$ is true, i.e. a value that has a very small probability of occurring if $H_0$ is true. Choose a value, $\alpha$, called the "significance level" of the test, to represent what you would consider to be a "small probability." (Typical values are 0.05 and 0.01.) Now it is easy to choose $z^*$: 

$H_a$: $\mu > \mu_0$, $P(Z > z^*) = \alpha$.

$H_a$: $\mu < \mu_0$, $P(Z < z^*) = \alpha$ (this makes $z^*$ negative)

$H_a$: $\mu \neq \mu_0$, $P(|Z| > z^*) = \alpha$, i.e. $P(Z > z^*) = \frac{\alpha}{2}$

Just as for confidence intervals, we find $z^*$ from Table D.

In Table D the relation between confidence level $C$ and $\alpha$ is as follows.

<table>
<thead>
<tr>
<th>$H_a$: $\mu$</th>
<th>$\mu_0$: $C$ =</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_a$: $\mu$</td>
<td>$\mu_0$: $C$ =</td>
</tr>
<tr>
<td>$</td>
<td>z_{obs}</td>
</tr>
</tbody>
</table>

Area = $C$

Area = $\frac{1-C}{2}$

Area = $\frac{1-C}{2}$
What is our conclusion in the PCB example at the 0.05 $\alpha$-level? At the 0.01 $\alpha$-level?
Recall that the alternative is $H_a: \mu > 5$.
For, $\alpha = 0.05$, $z^* =$ __________, and since __________, __________ $H_0 :$ conclude that the EPA standard is __________ t.
For, $\alpha = 0.01$, $z^* =$ __________, and since __________, __________ $H_0 :$ the evidence is __________ enough to conclude that the EPA standard is ______________.

AN ALTERNATIVE WAY TO DO THE TEST
3) Calculate __________ -value
Assuming $H_0$ is __________, the __________ that the test statistic would take a value as extreme or more extreme than that actually observed is the __________-value of the test.
The smaller the __________-value the stronger the evidence __________ $H_0$. A __________ __________-value
says that if $H_0$ is __________, then the observed result is __________ to occur just by chance.
The P-value is calculated based on the form of $H_a$:

$$H_a : \mu > \mu_0 \text{ is } P(Z \geq z)$$

$$H_a : \mu < \mu_0 \text{ is } P(Z \leq z)$$

$$H_a : \mu \neq \mu_0 \text{ is } 2P(Z \geq |z|)$$

These __________-values are __________ if the population distribution is normal and are __________ correct for large n otherwise.

Q: What is the __________-value in the PCB example?

4) State your conclusion.
Prior to the actual testing, choose a __________ level $\alpha$ that you consider evidence __________ $H_0$.
If the P-value __________ $\alpha$ we say the data is __________ at level $\alpha$.
If P-value __________ $\alpha$, __________ $H_0$ __________ $H_a$.
If P-value __________ $\alpha$, __________ $H_0$.__________ $H_a$. (""")
Note that when we """" $H_0$ we are not so much claiming $H_0$ is __________ as we are concluding there is __________ evidence to __________.
Typical $\alpha$-levels used are 0.05 and 0.01.

Q: What is our conclusion in the PCB example at the 0.05 $\alpha$-level?, at the 0.01 $\alpha$-level?

The hypothesis test based on rejecting $H_0$ when $P$ – value $< \alpha$ is equivalent to the test based on rejecting $H_0$ when the observed value of $z$, call it $z_{obs}$, is in the Rejection Region, or "critical region" provided that the $P$ – value and "critical region" are defined as shown in the table below.

<table>
<thead>
<tr>
<th>$H_a$</th>
<th>P - value</th>
<th>&quot;Critical region&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu &gt; \mu_0$</td>
<td>$P(Z &gt; z_{obs})$</td>
<td>$z_{obs} &gt; z^<em>$, where $z^</em>$ is chosen so that $P(Z &gt; z^*) = \alpha$</td>
</tr>
<tr>
<td>$\mu &lt; \mu_0$</td>
<td>$P(Z &lt; z_{obs})$</td>
<td>$z_{obs} &lt; -z^<em>$, where $z^</em>$ is chosen so that $P(Z &gt; z^*) = \alpha$</td>
</tr>
<tr>
<td>$\mu \neq \mu_0$</td>
<td>$2P(Z &gt;</td>
<td>z_{obs}</td>
</tr>
</tbody>
</table>

Because of this equivalence we have the following conclusions:

(i) ____ represents $P$(Reject $H_0$) when $H_0$ is true (an error probability). When you specify ____, you are saying that you are willing to ______ a ______ $H_0$ ________ of the time. It is also referred to as the "___________" for the hypothesis test. Just as __________ of confidence intervals ________ cover the true parameter value, __________ of significance tests that are done when $H_0$ is __________, will __________ $H_0$ and conclude that there is a ____________ between $\mu$ and $\mu_0$. Thus, when $\alpha = 0.05$, a ___________ will be announced one ___________ even when it is ___________ (i.e., $H_0$ is ___________).

(ii) The________ can be interpreted as the ___________ level of the test, i.e., the value of $\alpha$ you would be using if the "critical region" was defined to have ____________

(iii) Defining the $P$ – value the way we have done, as the probability of seeing a value of $Z$ at least as far away from 0 as we have seen, is the only way to make the tests agree and to assure that you would be rejecting $H_0$ for all values of $z_{obs}$ that are "too large."

(iv) The $P$ – value is a better indicator than $z_{obs}$ is by itself of how unusual or extreme an observed value of $\overline{x}$ is. For example, when $H_0$ is: $\mu > \mu_0$, and $\alpha$ is 0.05, so that $z^*$ is 1.64 suppose that $z_{obs}$ is 2.58. That may not seem to be extremely far above $z^*$, but the $P$-value would be 0.005, smaller than $\alpha$ by a factor of 10, showing how truly rare it is to see an observation this large.

(v) Statistical software gives _____-values, and the critical region approach requires finding _____ from tables, so it is considered ___________. People who use it often announce_______ and their ___________ without publishing __ so the reader cannot judge whether he/she ___________ with the conclusion.

(vi) If we publish the $P$ – value, another researcher who has a different preference for $\alpha$ can draw his/her own conclusion about rejecting $H_0$.

Because of (iv), (v) and (vi), most statisticians use the $P$ – value approach.
Example: The mean yield of corn in the U.S. has been about 120 bushels per acre. Is the mean yield, \( \mu \), of this year’s crop the same or different?
A survey of 40 farmers this year gives a sample mean yield of \( \bar{x} = \) _______ bushels per acre.
Is this good evidence that the national mean this year is __________ bushels per acre?
Assume the farmers surveyed is a ______ from the population of all commercial corn growers and the standard deviation of this population is \( \sigma = 10 \).
What is your conclusion at the 0.05 level? The 0.01 level? Is this conclusion valid even if the original population was somewhat nonnormal?

State hypotheses:

Calculate ____________________________.

Calculate___________-value.

Conclusion:

Note that for the one-sided \( H_a: \mu > 120 \), \( P = \) __________ and we would __________ at the 0.01 level.

**Beware of letting the data _______________________________!!**

The ___-value shows the strength of evidence ______________ the null hypothesis. Some statisticians feel that publishing the \( P \)-value is all you should do. Some feel that you should also show a conclusion – i.e. choose a level \( \alpha \).

**Confidence Intervals and Two-Sided Tests.**

The following works only when the alternative hypothesis has the form \( H_a: \mu \neq \mu_0 \) !!!!!

<table>
<thead>
<tr>
<th>Suppose the hypotheses being tested are</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0: \mu = \mu_0 ) vs ( H_a: \mu \neq \mu_0 ).</td>
</tr>
</tbody>
</table>

If \( \mu_0 \) ________________ of a level _____________________________ for \( \mu \)

then ________________ \( H_0 \) at level ____, otherwise ________________ \( H_0 \).

Example: Cough Syrup

Suppose the concentration of the active ingredient in a name brand cough syrup is _______ (or _______%). _____ measurements were made to determine the concentration of the same active ingredient in a competing generic brand of cough syrup. The question is whether the concentration of the active ingredient in the generic brand of cough syrup is __________ than in
the name brand. The mean was $\bar{x} =$___________. Assume the standard deviation of the measuring process is $\sigma =$ ______. Does the concentration of the active ingredient in the generic brand of cough syrup appear ________________ than the name brand at the $\alpha =$______ level?

Example: IQ Scores: The mean IQ test scores of ____ seventh-grade girls in a Midwest school district was __________. Assume that these______ girls are a __________ of all 7th graders in the school district and that $\sigma =$.

(a) Give a _______________________________ for the mean IQ score $\mu$ in the population.

(b) Is there _____________________________ at the ______ level that the mean IQ score in the population ________________________? State your hypotheses and only use the answer to part (a) to answer the question.

Example: Nicotine level
To determine whether the mean nicotine content of a brand of cigarettes is __________ than the advertised value of ________, a health advocacy group tests

$H_0: \mu = \mu_0$ vs $H_a: \mu \neq \mu_0$.

The value of their test statistic was $z =$.  
A group of researchers at a university repeated the experiment and produced a test statistic of $z =$.

What conclusions does each group make at the $\alpha =$_______________ level?
Moral of the Story
Avoid using _____-value ______ as the golden rule for determining what is __________. The criteria for ___________ H₀ vary from problem to problem and are a bit subjective. There is no clear __________ between “______________” and “__________________”. It is good practice to report the ____________ and let __________________ determine for themselves if the __________________________ is sufficiently ______________.