Section 8.2: Hypothesis Testing

In Chapter 7 and Section 8.1, we looked at point estimates and confidence intervals. Hypothesis testing generalizes this a bit. The idea behind hypothesis testing is that we can make "guesses" about a parameter, and then analyze how plausible those guesses are. To that end, a <u>statistical hypothesis</u> is a conjecture about a population parameter. It's important to remember that the hypothesis may be true or false, and we won't know for sure in many cases. All we can do is compute how probable it is that the hypothesis is true.

The <u>null hypothesis</u>, symbolized by H_0, is a statistical hypothesis that designates possible values for a parameter. That is, we guess what the parameter IS.

The <u>alternative hypothesis</u>, symbolized by H_A, is the "opposite" of the null hypothesis. No matter what the null hypothesis says the parameter IS, the alternative hypothesis will say that's what the parameter ISN'T.

We note that a null hypothesis and alternative hypothesis are always a pair. One determines the other.

The most common population parameter to make conjectures about is the mean.
For example
1) We could make the null hypothesis H_0: $\mu = 82$ which gives the alternative hypothesis H_A: $\mu \neq 82$
2) We could have null hypothesis H_0: $\mu \leq 82$, so that H_A: $\mu > 82$
3) Or, we could have null hypothesis H_0: $\mu \geq 82$ and alternative hypothesis H_A: $\mu < 82$

Note the null hypothesis always includes the equal sign and the alternative hypothesis never includes it.

$=$	$\geq$	$\leq$	These may only be **null hypothesis**
Is equal to Is exactly the same as Has not changed from Is the same as	Is greater than or equal to Is at least Is not less than	Is less than or equal to Is at most Is not more than	
$\neq$	$>$	$<$	These may ONLY be **alternative hypothesis**
Is not equal to Is different from Has changed from Is not the same as	Is greater than Is above Is higher than Is bigger than Is longer than Is increased	Is less than Is below Is lower than Is smaller than Is shorter than Is decreased from Is reduced from	

Example 1: State the null and alternative hypotheses for the following conjecture. A researcher thinks that if expectant mothers use vitamin pills, the birth weight of the babies will increase. The average birth weight of the population is 8.5 pounds.
 H_0: $\mu \leq 8.5$ and H_A: $\mu > 8.5$ (claim is H_A)

Example 2: State the null and alternative hypotheses for the following conjecture. An engineer hypothesizes that the mean number of defects will not decrease in a certain manufacturing process by using robots instead of humans for certain tasks. The mean number of defective disks per 1000 is 15.
H_0: $\mu \geq 15$ and H_A: $\mu < 15$ (claim is H_0)

As stated earlier, in most cases, we won't know for sure whether our hypotheses are true or false, but we instead want to know the probability that a hypothesis is true or false. But the hypothesis is being made about the actual underlying probability distribution, and all we have is a sample. So what we calculate is the probability that we would get the sample we got (or "worse") GIVEN that our null hypothesis is true. This gives us the following definition.

Definition of a *p*-Value
The *p-value* for a particular null hypothesis H_0 based on an observed sample is the probability of getting that sample or worse when H_0 is true. A "worse" data set is one in which H_0 is less plausible than it is for the actual observed data set.

Then when a *p*-value is small, this implies that it is unlikely that an experimenter would obtain this value when the null hypothesis is true, we would want to reject the null hypothesis. As a rule of thumb, we'll reject the null hypothesis whenever the *p*-value is less than 0.01.
When a *p*-value is relatively large (say, greater than 0.10), an experimenter usually assumes that enough evidence to reject the null has not been found so the null hypothesis is accepted as plausible. Note that "accepting" the null hypothesis does not mean that the null hypothesis is likely, just that it can't be ruled out. (It's sort of like the idea of being innocent until "proven" guilty beyond a reasonable doubt.)
When the *p*-value is between 0.01 and 0.10 it gets fuzzy, but generally, this case is considered inconclusive since the evidence is not overwhelming either way.

To compute *p*-values, we'll again use *t*-statistics. We'll start with a null hypothesis of the form $\mu = \mu_0$. Note that then the alternative hypothesis is $\mu \neq \mu_0$, which includes both the cases in which $\mu > \mu_0$ and in which $\mu < \mu_0$. For this reason, we call this a "two-sided test."

Two-Sided *t*-test (*p*-Value method)
The *p*-value for the two-sided hypothesis test H_0: $\mu = \mu_0$ versus H_A: $\mu \neq \mu_0$
based on a data set of n observations with a sample mean of $\bar{x}$ and a sample standard deviation of s, is
p-value $= 2 \times P(X \geq |t|)$, where In this case the random variable X has a *t*-distribution with n-1
degrees of freedom, and $t = \dfrac{\sqrt{n}(\bar{x} - \mu_0)}{s}$ which is known as the *t*-statistic.

Example 3: Suppose for a certain population, we get a sample with n = 20 observations, sample mean $\bar{x} = 13.6$, and sample standard deviation $s = 2.15$. Do a two-sided t-test to test the hypothesis $\mu = 15$.

Answer: Note that the null hypothesis is $H_0 : \mu = 15$, and the alternative hypothesis is $H_a : \mu \neq 15$.

We compute the t-statistic, $t = \dfrac{\sqrt{n}(\bar{x} - \mu_0)}{s} = \dfrac{\sqrt{20}(13.6 - 15)}{2.15} = -2.912.$

Then $|t| = 2.912$. We want to use the t-table to compute (or at least estimate) $P(X \geq 2.912)$ when X has a t-distribution with 19 degrees of freedom. We have to use the table in reverse for this. That is, we look at the row for 19 degrees of freedom and try to find something close to 2.912 in that row. We see the value 2.861 in that row under the column $\alpha = .005$, and we see the value 3.579 in that row under the column $\alpha = .001$. Since 2.912 is between those values, we can conclude that $P(X \geq 2.912)$ is between .001 and .005. From this, we can estimate the p-value, which is defined as $2 \times P(X \geq |t|)$. Thus, the p-value is between .002 and .01. In particular, since the p-value is less than .01, we reject the null hypothesis as implausible.

In the above example, we had to estimate rather than find the exact value of the p-value. Although some statistical software will be able to compute exact values of the p-value, this method of estimation will be good enough for our purposes.

If the null hypothesis is of the form $\mu \geq \mu_0$ or $\mu \leq \mu_0$, then the alternative hypothesis is only one-sided, so we can change our t-test slightly.

One-Sided t-Tests (p-Value method)
Based on a data set of n observations with a sample mean $\bar{x}$ and a sample standard deviation of s:

The p-value for the one sided testing hypothesis testing problem H0: $\mu \leq \mu_0$ versus HA: $\mu > \mu_0$ is given by p-value $= P(X \geq t)$.

The p-value for the one sided testing hypothesis testing problem H0: $\mu \geq \mu_0$ versus HA: $\mu < \mu_0$ is given by p-value $= P(X \leq t)$.

In this case the random variable X has a t-distribution with n - 1 degrees of freedom, and

$t = \dfrac{\sqrt{n}(\bar{x} - \mu_0)}{s}$

Significance Levels

So far, we have used the default rule of rejecting the null hypothesis if the *p*-value is less than 0.01, and accepting the null hypothesis as plausible if the *p*-value is more than 0.10. But instead, we can set a custom threshold to use. This threshold is supposed to be a measure of how improbable or probable it is that the null hypothesis is true, but it's important to remember that this is never a certainty no matter how small the *p*-value is. That is we have that the following 4 situations could all happen.

You reject the null hypothesis correctly (the null hypothesis is false).
You reject the null hypothesis incorrectly (the null hypothesis is true).
You do not reject the null hypothesis, but the null hypothesis is false so you should have rejected it.
You do not reject the null hypothesis, but you are correct since the null hypothesis is true.

These 4 situations may be represented by the following chart:

	H_0 True	H_0 False
Reject H_0	Type I **Error**	Correct Decision
Do NOT reject H_0	Correct Decision	Type II **Error**

In other words
A <u>Type I error</u> occurs if you reject the null hypothesis when it is true.
A <u>type II error</u> occurs if you do not reject the null hypothesis when it is false.

So we can define our threshold is terms of these errors.

The <u>significance level</u> or <u>size</u> of a hypothesis test is the maximum probability of committing a type I error. Thus, it is the maximum probability that you reject the null hypothesis when it is true. This probability is symbolized by α. That is, P(type I error) $= \alpha$.

This turns out to be related to the α used in the last chapter. Just as the confidence level desired was given to us in the last chapter, in these problems a desired significance level α may be given to us. And we will still use the same critical values of the *t*-distribution that we used in 8.1. In particular, we'll use $t_{\alpha/2, \, n-1}$ for two-sided tests and $t_{\alpha, \, n-1}$ for one-sided tests.

We'll summarize what we mean on the next page.

Hypothesis Testing for a population mean

Case 1: H_0: $\mu = \mu_0$

A size α test for the two-sided hypotheses H_0: $\mu = \mu_0$ versus H_A: $\mu \neq \mu_0$

rejects the null hypothesis H_0 if the test statistic $|t|$ falls in the <u>rejection region</u> $|t| > t_{\alpha/2,\, n-1}$

and accepts the null hypothesis H_0 if the test statistic $|t|$ falls in the <u>acceptance region</u> $|t| \leq t_{\alpha/2,\, n-1}$

In other words, the value μ_0 is contained within a $1 - \alpha$ level two-sided confidence interval

$$\left(\bar{x} - t_{\alpha/2,\, n-1} \cdot \frac{s}{\sqrt{n}} \,,\, \bar{x} + t_{\alpha/2,\, n-1} \cdot \frac{s}{\sqrt{n}} \right)$$ if and only if the hypothesis test with size α accepts the

null hypothesis. Thus μ_0 is not in the interval if and only if the hypothesis test with size α rejects the null hypothesis.

Case 2: H_0: $\mu \leq \mu_0$

A size α test for the one-sided hypotheses H_0: $\mu \leq \mu_0$ versus H_A: $\mu > \mu_0$

rejects the null hypothesis H_0 if the test statistic t falls in the <u>rejection region</u> $t > t_{\alpha,\, n-1}$

and accepts the null hypothesis H_0 if the test statistic t falls in the <u>acceptance region</u> $t \leq t_{\alpha,\, n-1}$

In other words, the value μ_0 is contained with a $1 - \alpha$ level one-sided confidence interval

$$\left(\bar{x} - t_{\alpha,\, n-1} \cdot \frac{s}{\sqrt{n}} \,,\, \infty \right)$$ if and only if the right-sided size α hypothesis test accepts the null

hypothesis.

Case 3: H_0: $\mu \geq \mu_0$

A size α test for the one-sided hypotheses H_0: $\mu \geq \mu_0$ versus H_A: $\mu < \mu_0$

rejects the null hypothesis H_0 if the test statistic t falls in the <u>rejection region</u> $t < t_{\alpha,\, n-1}$

and accepts the null hypothesis H_0 if the test statistic t falls in the <u>acceptance region</u> $t \geq t_{\alpha,\, n-1}$

In other words, the value μ_0 is contained with a $1 - \alpha$ level one-sided confidence interval

$$\left(-\infty \,,\, \bar{x} + t_{\alpha,\, n-1} \cdot \frac{s}{\sqrt{n}} \right)$$ if and only if the left-sided size α hypothesis test accepts the null

hypothesis.

Don't forget that in all these cases, we use the *t*-statistic given by $t = \dfrac{\sqrt{n}(\bar{x} - \mu_0)}{s}$.

Power Levels

While we've noted that the probability of a Type I error determines the significance level, we can also get information using the probability of a Type II error. This information is called the *power* of a hypothesis test.

> The *power* of a hypothesis test is defined to be 1 - (probability of a Type II error).
> That is, the power is the probability that the null hypothesis is true OR we reject it OR both.

The power of a hypothesis test is analogous to the length of a confidence interval. In particular, we have that the sample size, significance level, and power all work together in hypothesis testing in the same way that the sample size, confidence level, and length of the confidence interval work together. If we know two of these pieces of information, we can find the third piece. For a fixed significance level, the power increases as the sample size increases. For a fixed power, the significance level increases as the sample size increases. And for a fixed sample size, the significance level decreases as the power increases.

When σ is known

Finally, if we already know the actual standard deviation of the underlying probability distribution, then just as in 7.3, we can upgrade our *t*-tests to *z*-tests and use the standard normal table, with z_α being the critical point such that $P(X \geq z_\alpha) = \alpha$, where X is a standard normal random variable. We summarize these results in the next three boxes.

> **Two-Sided z-test (σ is known)**
> The *p*-value for the two-sided hypothesis test H_0: $\mu = \mu_0$ versus H_A: $\mu \neq \mu_0$
> based on a data set of n observations with a sample mean of $\bar{x}$ and an assumed *known* population standard deviation of σ, is
>
> *p*-value $= 2 \times \Phi(-|z|)$ where $\Phi(x)$ is the standard normal cumulative distribution function and
>
> $$z = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma}$$ is known as the z-statistic.
>
> A size α test rejects the null hypothesis H_0 if the test statistic $|z|$ falls in the rejection region $|z| > z_{\alpha/2}$ and accepts the null hypothesis H_0 if the test statistic $|z|$ falls in the acceptance region $|z| \leq z_{\alpha/2}$.
>
> The $1 - \alpha$ level two-sided confidence interval $\mu \in \left(\bar{x} - z_{\alpha/2,} \cdot \frac{\sigma}{\sqrt{n}} , \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$
>
> consists of the values of μ_0 for which this hypothesis testing problem has a *p*-value larger than α, that is, the values of μ_0 for which the size α hypothesis test accepts the null hypothesis.

Right-sided z-test (σ is known): H$_0$: $\mu \leq \mu_0$ versus H$_A$: $\mu > \mu_0$)

The p-value for the one sided testing hypothesis testing problem H$_0$: $\mu \leq \mu_0$ versus H$_A$: $\mu > \mu_0$ based on a data set of n observations with a sample mean of $\bar{x}$ and an assumed *known* population standard deviation of σ, is p-value $= 1 - \Phi(z)$. This testing procedure is called a <u>one-sided z-test</u>.

A size α test rejects the null hypothesis H$_0$ when $z > z_\alpha$ and accepts the null hypothesis when $z \leq z_\alpha$.

The $1 - \alpha$ level one-sided confidence interval $\mu \in \left(\bar{x} - z_\alpha \cdot \dfrac{\sigma}{\sqrt{n}} , \infty \right)$

consists of the values of μ_0 for which this hypothesis testing problem has a p-value larger than α, that is, the values of μ_0 for which the size α hypothesis test accepts the null hypothesis.

Left-sided z-test (σ is known): H$_0$: $\mu \geq \mu_0$ versus H$_A$: $\mu < \mu_0$)

The p-value for the one sided testing hypothesis testing problem H$_0$: $\mu \geq \mu_0$ versus H$_A$: $\mu < \mu_0$ based on a data set of n observations with a sample mean of $\bar{x}$ and an assumed *known* population standard deviation of σ, is p-value $= \Phi(z)$. This testing procedure is called a <u>one-sided z-test.</u>

A size α test rejects the null hypothesis when $z < - z_\alpha$ and accepts the null hypothesis when $z \geq -z_\alpha$.

The $1 - \alpha$ level one-sided confidence interval $\mu \in \left(-\infty, \bar{x} + z_\alpha \cdot \dfrac{\sigma}{\sqrt{n}} \right)$

consists of the values of μ_0 for which this hypothesis testing problem has a p-value larger than α, that is, the values of μ_0 for which the size α hypothesis test accepts the null hypothesis.

This marks the end of the material that we'll cover in Chapter 8.