9.1: The Nature of Hypothesis Testing

Suppose a manufacturer claims on the label that a package contains 8 ounces of potato chips. A customer (or a FDA analyst) buys five bags of chips, weighs them on a high-accuracy scale, and obtains weights of 7.5, 7.7, 8.1, 7.2, and 6.9 oz. Is the manufacturer guilty of falsifying the label information?

Questions such as this can be addressed through a statistical process called *hypothesis testing*.

Null and alternative hypotheses:

In science, a hypothesis is a statement which can be tested through experimentation or systematic observation.

In statistics, a *hypothesis* is a statement regarding the value of a parameter in one or more populations. Hypothesis testing involves two hypotheses:

Null hypothesis, denoted H_0: A statement of equality (so it uses $=$). The null hypothesis indicates that any apparent effect (difference) is due to chance.

→ H_1 or H_a

Alternative hypothesis, denoted H_1: A statement of inequality, that uses $\neq$, $<$, or $>$.
The alternative hypothesis indicates that any apparent effect (difference) is NOT due to chance.

Note: You can think of the null hypothesis and alternative hypothesis as complements of one another.

Suppose the population mean is our characteristic of interest. Here are the possible pairs of null and alternative hypotheses:

In all of these, μ_0 is a number (μ_0 is the "benchmark" we're comparing against)

$$\left.\begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{array}\right\}$$ This is called a two-tailed test.

$$\left.\begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{array}\right\}$$ This is a one-tailed test. It assumes that $\mu > \mu_0$ is not possible or is of zero interest.

$$\left.\begin{array}{l} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{array}\right\}$$ This is a one-tailed test. It assumes that $\mu < \mu_0$ is not possible or is of zero interest.

Note: For one-tailed tests, many books write the null hypothesis as $\mu \geq \mu_0$ (or $\mu \leq \mu_0$) instead of $\mu = \mu_0$. Under the assumption that $\mu > \mu_0$ (or $\mu < \mu_0$) is impossible, this is equivalent to the null hypotheses used in our book ($\mu = \mu_0$).

Note: Two-tailed tests are much more common than one-tailed tests. A researcher who wishes to use a one-tail test must present a solid rationale for doing so.

Example 1: A snack food company claims that its bags of potato chips weigh 8.0 ounces. A customer wants to determine whether this claim is true. State the null and alternative hypotheses.

Null: $H_0: \mu = 8$ oz

Alternative: $H_a: \mu \neq 8$ oz

(two-tailed test)

Example 2: The normal human body temperature is widely accepted to be $98.6°$ F. A medical researcher wants to know whether a certain population of Native Alaskans has a mean body temperature of $98.6°$ F. State the null and alternative hypotheses.

Null: $H_0: \mu = 98.6°F$

Alternative: $H_a: \mu \neq 98.6°F$

(two-tailed test)

Example 3: The average amount of lead in the blood of young children is 2 micrograms per deciliter (mcg/dL). A city has recently changed its water supply, and there have been widespread reports of increased lead levels in the water. A concerned doctor wants to dig into the city's medical records to find out whether the children in the city have blood lead levels above 2 mcg/dL. State the null and alternative hypotheses.
https://www.health.ny.gov/publications/2526.pdf

Null: $H_0: \mu = 2$ mcg/dL

Alt: $H_a: \mu > 2$ mcg/dL

(one-tailed test)

Logic of hypothesis testing:

We start with an assumption that the null hypothesis is true. We examine the evidence provided by the sample(s), and determine whether there is sufficient evidence to reject the null hypothesis.

If there is sufficient evidence to reject the null hypothesis, then we conclude that the alternative hypothesis is likely to be true. However, we cannot "prove" that the alternative hypothesis is true.

If there is not sufficient evidence to reject the null hypothesis, then it is still not appropriate to say we have "proven" or "accepted" the null hypothesis. All we can conclude is that this sample provided insufficient evidence to reject it.

In summary, we reach one of the following conclusions:

1) We reject the null hypothesis H_0.
2) We fail to reject the null hypothesis H_0.

Types of inference errors:

		What Really Happened	
		$\mu = \mu_0$ **H_0 is TRUE**	**H_0 is FALSE (So H_1 is true)** $\mu \neq \mu_0$ (2-tailed test)
Our Conclusion	**Do not reject H_0**	Correct inference	β Type II error
	Reject H_0	Type I error α	Correct inference

we missed an "interesting difference" that does exist

found an "interesting difference" that doesn't really exist.

<u>Important</u>: In any hypothesis test, any of these four results can happen!

The researcher decides what level of risk of making a Type I error he or she is willing to accept. by determining the α . The α is called the *level of significance*. It must be chosen <u>in advance</u>, before the sample is analyzed.

α is the conditional probability of (incorrectly) rejecting H_0 given that H_0 is true.

Common choice for α are:

0.10 (corresponds to 90% confidence interval)
0.05 (corresponds to 95% confidence interval)
0.01 (corresponds to 99% confidence interval)

Example 4: Suppose a veterinarian wants to learn whether wild mustangs have a front hoof angle of 45°, which for many years was considered the most desirable front hoof angle for domestic horses. Describe how a Type I and a Type II error would manifest themselves in this situation.

$H_0: \mu = 45°$

$H_a: \mu \neq 45°$

Type I error: we conclude that the mustangs' hooves differed from 45°, when in reality, their hooves had a mean of 45°.

Type II error: the hooves differed from 45°, but we did not pick up on this difference from our sample. (we missed it)

Example 5: It is known that the student body of Lone Star College – North Harris is composed of 61% women and 39% men. An administrator wants to find out whether the percentage of women in evening classes is also 61%. State the null and alternative hypotheses, and describe how a Type I and a Type II error would manifest themselves in this situation.

$H_0: p = 0.61$ $\qquad$ p = proportion of women

$H_a: p \neq 0.61$

Type I error: (we found a difference that doesn't exist)
the proportion of women is 0.61, but we concluded it differed from 0.61.

Type II error: (there is a difference, but we missed it)
The proportion of women differs from 0.61, but we concluded it was 0.61.