3: HYPOTHESIS TESTING

Often, we hope to use our data to draw specific conclusions, or make decisions. We want to decide whether or not the sample data support some belief or hypothesis about the population.

- Is Rogaine effective?

- Does a market analyst have any skill at forecasting the Dow?

- Is the climate changing?

- Can people tell the difference between Diet Coke and Diet Pepsi?
Do McDonalds "Quarter Pounders" actually weigh less than 1/4 pound, on the average?

Hypothesis testing provides answers to questions like these.

It must be emphasized, however, that our conclusions may be wrong, because of sampling variability.

Here, we focus attention on methods of making decisions about a population mean, \( \mu \).

**Eg:** Based on a sample of 50 "Quarter Pounders," we want to decide if the mean weight \( \mu \) of all
"Quarter Pounders" is actually less than .25 pounds.

- Statistical hypothesis tests are set up like criminal court cases: the defendant is presumed innocent until proven guilty beyond a reasonable doubt.

Thus, we will start with the assumption that $\mu$ is in fact .25. This is called the null hypothesis, and is written as $H_0 : \mu = .25$.

- In most applications, it is hoped that the data will provide enough evidence to allow us to reject the null hypothesis in favor of the alternative hypothesis, which is in this case $H_1 : \mu < .25$. 
The evidence is provided by a single number called a **test statistic**, obtained from the observed data.

Based on the value of the test statistic, we will decide either to reject, or not to reject, the null hypothesis. Specifically, we reject $H_0$ whenever the test statistic lies in a prespecified range of values called the **rejection region**.

As in the courtroom analogy, there are two different kinds of incorrect decisions, or errors.

- **Type I Error**: Rejecting $H_0$ when $H_0$ is true
**Type II Error:** Not rejecting $H_0$ when $H_0$ is false.

Using statistical theory, we can control the probability of making a Type I error. This probability is denoted by $\alpha$ and called the **significance level** of the test.

- We are free to choose the value of $\alpha$. Clearly it is desirable for $\alpha$ to be small, but there is a tradeoff here: the smaller $\alpha$ is, the larger the probability of a Type II error will be.

- It is customary to use either $\alpha = .05$ or $\alpha = .01$. 
Examples: What are $H_0$, $H_1$, Type I errors and Type II errors in the following situations:

- Murder Trials
- Smoke Alarms
- Pregnancy Tests
- Boy Who Cried Wolf.

Format For Hypothesis Tests

(1) Before examining the data, formulate $H_0$, $H_1$, and select $\alpha$.

(2) Examine the data, evaluate the test statistic, and draw your conclusion.
To ensure the validity of the test results, $H_0$ and $H_1$ must be formulated \textit{before} the data are examined. The null hypothesis always takes the form $H_0: \mu = \mu_0$, so that $\mu_0$ is the value of the population mean if $H_0$ is true.

Depending on what we are trying to prove, we take $H_1$ as either:

- $H_1: \mu \neq \mu_0$ (A \textit{two-sided} alternative)

- $H_1: \mu < \mu_0$ (A \textit{one-sided} alternative)

- $H_1: \mu > \mu_0$ (A \textit{one-sided} alternative).

Suppose for now that $\sigma$ is known.
Then the test statistic is

\[ Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}. \]

The \( Z \) statistic provides a standardized measure of the difference between the hypothesized value of \( \mu \) and the actual value of the sample mean \( \bar{X} \).

If \( H_0 \) is true, then \( Z \) is (approximately) standard normal. (Why?)

In hypothesis testing, we simply check whether \( Z \) is "too large" or "too small" to have plausibly come from a standard normal distribution. If so, then we reject \( H_0 \).
How To Perform The Test

• If $Z$ lies in the rejection region (given below), reject $H_0$. Otherwise, do not reject $H_0$.

• The rejection region for a level $\alpha$ test depends on the nature of the alternative hypothesis.

If $H_1$ is two-sided, we get a two-tailed test.

If $H_1$ is one-sided, we get a one-tailed test.

(There are two kinds of one-tailed tests).

\[
\begin{array}{c|c}
H_1 & \text{Rejection Region} \\
\mu \neq \mu_0 & |Z| > z_{\alpha/2} \\
\mu < \mu_0 & Z < -z_{\alpha} \\
\mu > \mu_0 & Z > z_{\alpha}
\end{array}
\]
For example, to test $H_0: \mu = \mu_0$ against the two-sided alternative $H_1: \mu \neq \mu_0$, we reject $H_0$ if either $Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$. (The dividing lines, in this case $-z_{\alpha/2}$ and $z_{\alpha/2}$, are called critical values.)

This does give a level $\alpha$ test, since

\[
\text{Prob(Reject } H_0 \mid H_0 \text{ is True}) = \text{Prob(Type I Error)} = \text{Prob}(\text{Reject } H_0 \mid H_0 \text{ is True})
\]

\[
= \text{Prob}(\mid Z \mid > z_{\alpha/2} \mid H_0 \text{ is true}) = \text{Prob}(\mid \text{Std Normal RV} \mid > z_{\alpha/2}) = \alpha
\]

- Explain why the above 1-tailed test criteria are also correct.
Eg 1: For the "Quarter Pounders" example, test $H_0: \mu = .25$ against $H_1: \mu < .25$ at the 5% level of significance, assuming that $n = 50$, $\sigma = .035$ and $\bar{x} = .24$. 