LECTURE: HYPOTHESIS TESTING (I)

Welcome to the magical world of hypothesis testing, which is using statistics to verify if a claim is true or not

1. Hypothesis Testing

Definition:

A **hypothesis** is a claim about one or more parameters of a population of interest

For example, we might claim that a parameter equals to 10 or is between 20 and 25

Example 1:

Suppose you are a pollster who is interested in the preferences of strawberry ice cream in the US.

Your hypothesis is that "people in cold areas are less likely to prefer strawberry ice cream"

Parameter of Interest: $p_1 - p_2$ the difference in the proportion of strawberry supporters in the hot vs cold areas.

Hypothesis: $p_1 - p_2 > 0$

Estimator: $\hat{p}_1 - \hat{p}_2$ where you sample people from both populations.

Using statistics, you can accept or reject your hypothesis based on the value of $\hat{p}_1 - \hat{p}_2$

Example 2:

You are the principal investigator for a trial of a new Peyamphetamine drug to treat high blood pressure.

You claim that your new drug will reduce systolic blood pressure by 10 mmHg compared to a placebo pill.

Parameter of Interest: $\mu_1 - \mu_2$ the difference of mean blood pressure between people who got the **placebo** vs. those who got the drug

Hypothesis: $\mu_1 - \mu_2 \geq 10$

Estimator: $\overline{Y_1} - \overline{Y_2}$ where you do a double-blind study and randomly assign 100 patients to receive the placebo pill and 100 patients to receive the drug.

Upshot: Hypothesis testing uses the ideas we learned in the previous sections (sampling, estimators, and confidence intervals) to make inferences about a population using samples from that population, and to quantify how confident we are with these inferences.

2. Elements of a Hypothesis Test

There are four steps in a hypothesis test

Example 3:

Suppose we are interested in the chocolate vs. vanilla ice cream preferences in the US

STEP 1: Formulate the alternative hypothesis

Definition:

The **alternative hypothesis** is a hypothesis that we would like to support

For example, our alternative hypothesis could be "More than 50% of the people in the US prefer chocolate ice cream"

Mathematically, this can be written as:

Alternative Hypothesis: $p > 0.5$ where p is the proportion of people who prefer chocolate ice cream.

STEP 2: Reject the null hypothesis

Definition:

The **null hypothesis** is the opposite of the alternative hypothesis

In this case, the null hypothesis is $p \leq 0.5$ but there is a more useful way to state this:

Null Hypothesis: $p = 0.5$

This is because if we reject $p = 0.5$ then we will reject $p \leq 0.5$ as well. But it'll be easier to deal with equalities than with inequalities.

STEP 3: Test statistic/Estimator

Definition:

A **test statistic** is something we can measure to either reject or accept the null hypothesis.

In general, it will be one of our common estimators, such as $\bar{Y}$ or $\hat{p}$

In this case, we will use $\hat{p} = \frac{Y}{n}$ for our estimator

STEP 4: Rejection Region

Definition:

A **rejection region** (RR) specifies the values of the test statistic for which the null hypothesis will be rejected.

In this case we'll reject the null hypothesis if $\hat{p}$ is high, so the the rejection region looks like:

Rejection Region: $\hat{p} \geq k$ where k is TBA

In general, it is you who chooses the value of k, it depends on how much you're willing to tolerate.

A lower value of k (think $k = 0.2$) means that you are more likely to reject the null hypothesis even though it's actually be true. This is called a **false positive error**

A higher value of k (think $k = 0.8$) means that we are more likely to accept the null hypothesis even though it's actually false. This is called a **false negative error**

Summary: Elements of a Hypothesis Test:

(1) Alternative hypothesis, H_a

(2) Null hypothesis, H_0

(3) Test statistic

(4) Rejection region (RR)

Example 4:

Chocolate Ice Cream Example

(1) Alternative hypothesis, $H_a : p > 0.5$

(2) Null hypothesis, $H_0 : p = 0.5$

(3) Test statistic, $\hat{p} = Y/n$

(4) Rejection region (RR), $\{\hat{p} > k\}$

Example 5:

Peyamphetamine Example

(1) Alternative hypothesis, $H_a : \mu_1 - \mu_2 > 10$

(2) Null, $H_0 : \mu_1 - \mu_2 = 10$

(3) Test statistic, $\bar{Y}_1 - \bar{Y}_2$

(4) Rejection region (RR), $\{\bar{Y}_1 - \bar{Y}_2 > k\}$

Example 6:

Suppose you're designing a ball bearing machine which is supposed to produce ball bearings that are 5 mm in diameter, but your suspect/claim that there is something wrong with the machine.

The parameter of interest is μ, the average ball bearing diameter.

(1) Alternative hypothesis, $H_a : \mu \neq 5$

(2) Null hypothesis, $H_0 : \mu = 5$

(3) Test statistic, $\bar{Y}$

(4) Rejection region (RR) $|\bar{Y} - 5| > k$

The first two examples are called **one-tailed hypothesis tests** since we reject the null hypothesis if our parameter of interest is above a certain value. The last example is a **two-tailed hypothesis test** since we reject it if our parameter is above *or* below a certain value.

Error: There are two types of errors from a hypothesis test.

Definition:

A **type I (false positive) error** α is made if the null hypothesis is rejected when it is in fact true

$$\begin{aligned}\alpha &= P(\text{reject null hypothesis when it's actually true}) \\ &= P(\text{test statistic lies in RR when null hypothesis is true})\end{aligned}$$

Definition:

A **type II (false negative) error** β is made if the null hypothesis is accepted when it's in fact false

$$\begin{aligned}\beta &= P(\text{accept null hypothesis when it's actually false}) \\ &= P(\text{test statistic lies outside RR when } \textbf{alt} \text{ hypothesis is true})\end{aligned}$$

Note: Some statisticians will use the term **power** which is $1 - \beta$

3. Large Sample Hypothesis Tests

Goal: Test a hypothesis about the mean μ or proportion p of a population, assuming that the sample size is large

In that case, we can assume that (by the Central Limit Theorem) the test statistic is normally distributed

Setting: The parameter of interest if θ

(1) Alternative hypothesis: $H_a : \theta > \theta_0$

(2) Null hypothesis, $H_0 : \theta = \theta_0$

(3) Test statistic, $\hat{\theta}$, some estimator

(4) Rejection region (RR) $\hat{\theta} > k$, where k is TBA

Problem: How to find k?

Here we're given the type I error α that we're willing to accept, think like a tolerance