# Mathematics 231

Lecture 22

Liam O'Brien

# Announcements

- Reading
  - Today  M&M 6.2  372-390
  - Next class  M&M 7.1  422-428

# Topics

- Hypothesis testing

# Introduction to Hypothesis Testing

- The goal of statistical inference is to draw some conclusion about a population parameter based on data from a sample.
- Confidence intervals are appropriate when the goal is to estimate a population parameter, e.g. population mean, $\mu$, or proportion, p.
- Another approach to statistical inference is hypothesis testing.

# Introduction to Hypothesis Testing

- Hypothesis testing allows us make a decision about the value of a population parameter based upon evidence obtained in a sample.
- There is a useful analogy with the legal system in the United States.
- We want to decide whether a person is innocent or not based upon evidence.

# Introduction to Hypothesis Testing

- Hypothesis testing assesses the evidence provided by the data in favor of some claim concerning the population parameter.
- For example, is the mean bank income "gain" positive or negative?
- Another example… is the mean birthweight of infants for mothers who smoke less than that for mother's who don't?
- Hypothesis testing answers the following question:  Are the data compatible with the claim (hypothesis)?

# Null and Alternative Hypotheses

- **Null hypothesis**: Claim that we (often) wish to find evidence **against**, usually denoted $H_0$.
- Null hypothesis is often a statement of "no effect" or "no difference," expressed in terms of the population parameter, e.g. $H_0$: $\mu = \mu_0$.
- **Alternative hypothesis**: Claim that we suspect may be true instead of $H_0$, usually denoted $H_A$.

# Example: Bank Incomes

- The net change in income from last year to this year was calculated for a sample of banks.
- We want to know if bank incomes have changed compared to last year.
- What is the null hypothesis?

# Example: Birthweights

- Birthweights are recorded for all children born in Norway for a given year.
- We want to know if birthweights are different from babies born to mothers who smoke.
- What is the null hypothesis?

# One- versus Two-Sided Alternatives

- It is not always clear whether $H_A$ should be one-sided or two-sided.
- One-sided: only concerned with deviations from $H_0$ in one direction (e.g., $H_A$: $\mu<100$).
- Two-sided: concerned with deviations from $H_0$ in either direction (e.g., $H_A$: $\mu\neq100$).
- We will always use a two-sided alternative unless we're told otherwise.

# One- versus Two-Sided Alternatives

- Should set up two hypotheses, $H_0$ and $H_A$, to cover all possibilities for $\mu$.

1. $H_0$: $\mu = \mu_0$, versus $H_A$: $\mu \neq \mu_0$ (two-sided alternative), or
2. $H_0$: $\mu = \mu_0$, versus $H_A$: $\mu > \mu_0$ (or $H_A$: $\mu < \mu_0$) (one-sided alternative)

# Examples

- For the bank example should we consider?

1. $H_0$: $\mu = 0$, versus $H_A$: $\mu \neq 0$ (two-sided alternative), or
2. $H_0$: $\mu = 0$, versus $H_A$: $\mu > 0$ (one-sided alternative)

- What about the birthweight example? (mean weight is 3500 grams)

# Test Statistic

- Is it likely that the sample came from a population with mean $\mu_0$?
- To evaluate whether $H_0$ is likely to be true, compare $\overline{x}$ to the hypothesized population mean $\mu_0$.
- Values for $\overline{x}$ that are far from $\mu_0$ provide evidence against $H_0$, i.e., $H_0$ is very unlikely.
- More formally, a test statistic measures the compatibility of the data and $H_0$ being true.

# Test Statistic

- To evaluate whether $H_0$ is likely to be true, compare $\overline{x}$ to hypothesized population mean $\mu_0$.
- Is $\overline{x}$ so far away from $\mu_0$ that it seems very unlikely that $H_0 : \mu = \mu_0$ could be true?
- How do we determine what is far away?
- How do we determine what is likely or unlikely?

# Test for a Population Mean (σ known)

- Given a SRS of size n, we want to test $H_0 : \mu = \mu_0$ against $H_A : \mu \neq \mu_0$ (two-sided alternative).
- Base test on sample mean and use test statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

- When $H_0 : \mu = \mu_0$ is true, what is the distribution of this test statistic?

# Example: IQ and Neonatal Stroke

- $\mathrm{H}_0 : \mu = 0; \ H_A : \mu \neq 0$. (Assume known $\sigma$=26.4)
- A simple random sample of 110 banks shows an average gain of 8.1%.
- Is this gain large enough to say that there was a significant increase?
- If $\mathrm{H}_0 : \mu = 0$ is true, is a mean of $\bar{x}$=8.1 unusual?

# Test Statistic and P-Values

- Note:  Test statistic is a random variable.
- A test of significance assesses evidence against $H_0$ in terms of probability (p-value).
- It finds the probability of getting an outcome as or more extreme than the actual outcome, given the null hypothesis is true.
- Here, "extreme" means far from what we might expect if $H_0$ were in fact true.

# The P-Value

- The p-value is one of the most misinterpreted elements of a statistical analysis.
- It is not an icon to be worshipped, or the net worth of your research.
- It is simply a conditional probability.
- Is tells you the probability that you would have obtained a test statistic as extreme, or more extreme, than the one you obtained *given* the null hypothesis is true.

# The P-value

- Why do we assume that the null hypothesis is true when interpreting the p-value?
- Test statistics have distributions, and these distributions have different centers, spreads (and possibly shapes) depending on what we assume about the true underlying parameters.
- Our goal is to see whether the data that we have obtained from our sample is consistent with the null hypothesis.

# What Does the P-Value Tell Us?

- The p-value is the determining factor in establishing statistical significance.
- The smaller the p-value, the more evidence we have against the null hypothesis.
- How much evidence is enough evidence to say that the null hypothesis should be rejected?
- This is highly subjective and depends only on the predetermined significance level of the test.

# The Significance Level

- The significance level of the hypothesis test is denoted by $\alpha$.
- A common choice for $\alpha$ is 0.05.
- If the p-value is less than $\alpha$, then we often say there is sufficient evidence against the null hypothesis and reject it in favor of the alternative (we achieve statistical significance).
- The significance level must be determined *a priori* as the decisive value as which significance is achieved.

# The Significance Level

- How do we choose the significance level?
- We will cover this in more detail later, but for now will just state that the lower the level of the test, the more conservative it is (i.e., harder to reject the null).
- The vast majority of tests use $\alpha = 0.5$, but some typically use $\alpha = 0.10$ or higher.
- **Statistical significance, and scientific (or practical) significance are not synonymous.**

# Test for a Population Mean (σ known)

- Given a SRS of size n, we want to test $H_0 : \mu = \mu_0$
- Base test on sample mean and use test statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

- When $H_0 : \mu = \mu_0$ is true, this statistic has a standard normal distribution.
- Calculate the p-value = 2P(Z>|z|) when $H_A$ is two sided;
  or p-value = P(Z>z) for $H_A : \mu > \mu_0$;
  or p-value = P(Z<z) for $H_A : \mu < \mu_0$.

# Example: Bank Incomes

- A sample of 110 banks shows an average gain of 8.1% from last year to this year (assume $\sigma$=26.4%).
- $H_0: \mu = 0$; $H_A: \mu \neq 0$.
- What is the test statistic?
- P-value < 0.05, therefore we have enough evidence to reject the null hypothesis at the $\alpha$=0.05 significance level. There is a significant net increase in bank incomes over the past year.

# Example: Birthweights

- A sample of 50 mothers shows an average birthweight of 3200 grams (assume $\sigma$=430 g).
- $H_0 : \mu = 3500;\ H_A : \mu \neq 3500.$
- What is the test statistic? z = -4.93
- P-value < 0.05, therefore we have enough evidence to reject the null hypothesis at the $\alpha$=0.05 significance level. There is a significant decrease in birthweights among mothers who smoke.