Basics of Hypothesis Testing

Two main activities in statistical inference are using sample data to:

1. estimate a population parameter—forming confidence intervals
2. **test a hypothesis or claim about a population parameter**

Hypothesis

A claim or statement about a property of a population

Hypothesis Test

A standard procedure for testing a claim about a property of a population

Example Hypotheses

- A newspaper headline makes the claim: "Most workers get their jobs through networking"
- Medical researchers claim: "The mean body temperature of healthy adults is not equal to 98.6^{o} F"
- The FAA claims: "The mean weight of an airline passenger with carry-on baggage is greater than the 185 lb that it was 20 years ago"

Testing will be done based on the:

Rare event rule for inferential statistics

If, **under a given assumption**, the probability of a particular observed event is exceptionally small, we conclude that the assumption is probably not correct

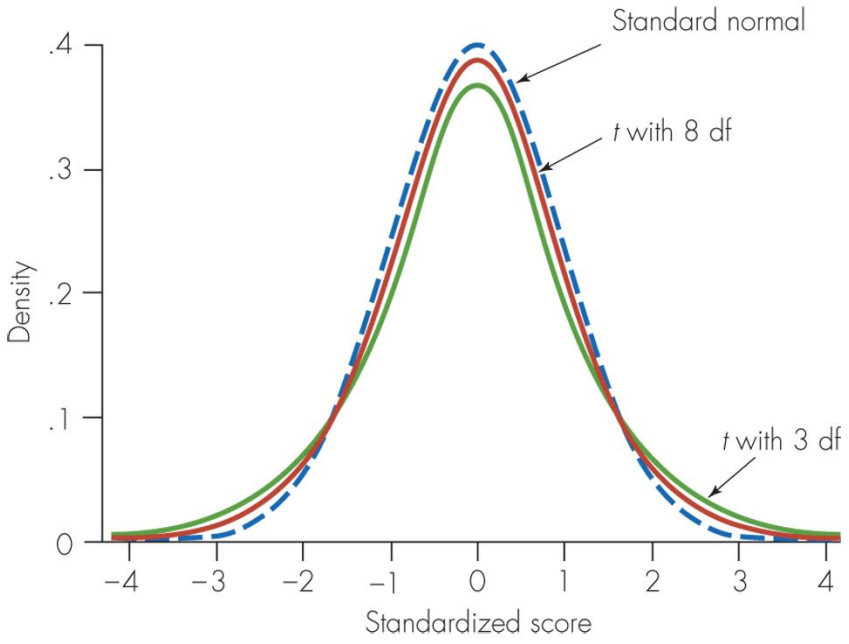
We analyze sample data in an attempt to distinguish between results that can ***easily occur by chance*** and results that are ***highly unlikely to occur by chance.***

If we observe a highly unlikely result one of two things has happened—either

1. a rare event has indeed occurred, or

2. the underlying assumption is not true

For the methods we will learn in this class, we will be using the normal distribution or *t*-distribution to make decisions about something being a rare event.



Components of a Formal Hypothesis Test

1. Hypotheses

Null hypothesis (H_0) is a statement that the value of a population parameter is ***equal to*** some claimed value

Examples include:

$$H_0: p = 0.5$$
$$H_0: \mu = 98.6$$
$$H_0: p_1 - p_2 = 0 \Rightarrow p_1 = p_2$$
$$H_0: \mu_1 - \mu_2 = 0 \Rightarrow \mu_1 = \mu_2$$

We test the null hypothesis directly ➔ assume it is true and reach a conclusion to either reject H_0 or fail to reject H_0

Alternative hypothesis (H_a or H_1) is a statement that the population parameter has a value that somehow differs from the null hypothesis.

Statements for alternative hypotheses will use one of these symbols: $<$ or $>$ or $\neq$.

Examples include:

$H_a: p \neq 0.5;\ H_a: p > 0.5$

$H_a: \mu < 98.6$

$H_a: p_1 - p_2 \neq 0 \Rightarrow p_1 \neq p_2$

$H_a: \mu_1 - \mu_2 < 0 \Rightarrow \mu_1 < \mu_2$

You will need to be able to specify at least one pair of null and alternative hypotheses for your research question.

2. Test Statistic

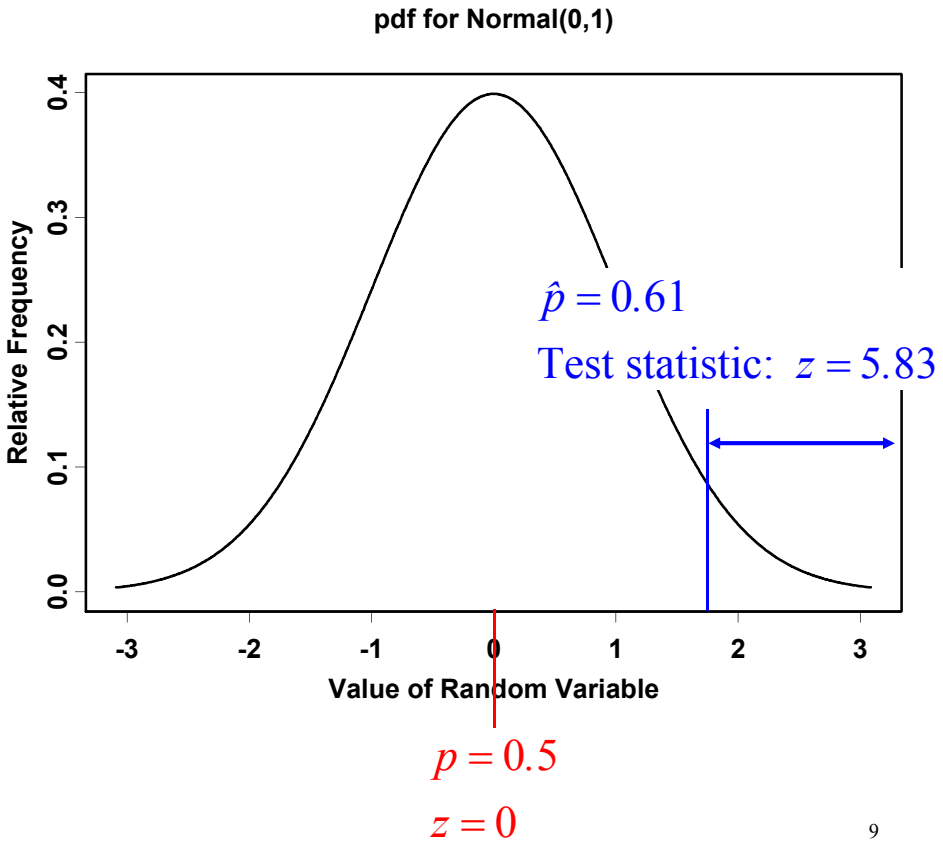
- Calculated from sample data
- Convert a sample proportion $\hat{p}$, or a sample mean $\bar{x}$ to a *z*- or *t*-score **with the assumption that the null hypothesis is true**

  Calculation of a z- or t-score uses the value of the population parameter proposed under the null hypothesis, along with a measure of the sampling variability for the particular statistic.

You will use computer software to calculate your test statistic(s).

3. Critical Region

The set of all values of the test statistic that cause us to reject the null hypothesis.



4. Significance Level

- Probability that the test statistic will fall in the critical region when the null hypothesis is actually true
- Denoted as α
- If the test statistic falls in the critical region, we will reject the null hypothesis--α is the probability of making a mistake of rejecting the null hypothesis when it is true
- Must be set ***a priori***—prior to running your hypothesis test—you will use $\alpha = 0.05$ for your projects

5. *P*-Value

- Called the observed significance level
- Is the probability of getting a value of the test statistic that is ***at least as extreme*** as the one representing the sample data, ***assuming the null hypothesis is true***
- The null hypothesis is rejected if the *P*-value is very small, such as 0.05 or less

The *P*-value will be calculated by statistical software:

- The larger the *z*- or *t*-statistic, the smaller the *P*-value
- Must differentiate between a two-tailed test and a one-tailed test
  - Two-tailed tests follow from alternative hypotheses that use ≠
  - The significance level, α, is divided equally between the two tails that constitute the critical region

6. **Conclusions**--Should be stated in the context of the problem. For example:

   Suppose we tested $H_0 : p = 0.5$ versus $H_a : p > 0.5$ for the proportion of red cards in a standard deck of playing cards. Based on our sample, $z = 0.67$.

   Fail to reject the claim that the population proportion of red cards in a deck is equal to 0.5 ($P = 0.2514$).

   Until stronger evidence is obtained, continue to assume that the population proportion of red cards in the deck is equal to 0.5 ($P = 0.2514$)

   Now suppose a different sample from a different deck gave $z = 2.03$—

   There is sufficient sample evidence to ***reject the claim*** that the population proportion of red cards in the deck is equal to 0.5 ($P = 0.0212$).

   There is sufficient sample evidence to support the claim that the population proportion of red cards in the deck is equal greater than 0.5 ($P = 0.0212$).

You will have to make the decisions and write the conclusions!

Identifying H_0 and H_a

1. More than 25% of Internet users pay bills online
2. Most households have telephones
3. The mean weight of women who won Miss America titles is equal to 121 lb.
4. The percentage of workers who got a job through their college is no more than 2%.
5. Plain M&M candies have a mean weight that is at least 0.8535 g.
6. The success rate with surgery is better than the success rate with splinting.
7. Unsuccessful job applicants are from a population with a greater mean age than the mean age of successful applicants.

Written Reports

1. **Introduction**

Contains a short write-up (2-3 paragraphs) of background information for your chosen research topic.

- Motivation for research
- Identification of your question

2. **Literature Review**

- Other work that has been done
- Must cite your references

3. **Methods**

This has mostly been provided for you. You will need to add a summary paragraph about the data.

- Number of records
- Combined responses from three years, if applicable
- Describe how analysis will be done—what tests you used
- Cite Excel and/or SPSS statistical software—include version of software and operating system

4. **Results**—will include

➢ Data summary section

- At least one appropriate graphic—label all of your Figures and add captions
- Figure captions go at the bottom of the Figures.
- Summary statistics—probably best to use a table—label all your Tables and add captions
- Table captions go at the top of the Tables.
- At least one confidence interval for the main parameter of interest
- Reference all Figures and Tables in your writing.

➢ Hypothesis Testing Section

- Hypotheses stated in both words and symbols
- Value of any calculated test statistics and corresponding *P*-values
- Basic decisions—reject or fail to reject the null

5. **Discussion/Conclusion**

Use decisions to restate conclusions in the context of the question—discuss your findings/observations

How does this relate to what other researchers have found?

Does this lead you to think about some new questions?

Discuss any problems that arose during your analysis.

What might you do differently next time?

6. **Literature Cited**

Use APA style guide