MATH 1070 - Chapter 17 Notes

Tests of Significance: The Basics

Recall: A confidence interval is constructed when your goal is to estimate a population parameter.

The second common type of inference, called a **significance test**, has a different goal: to assess the evidence provided by data about some claim concerning a population.

Example: Suppose a basketball player claimed to be an 80% free-throw shooter. To test this claim, we have him attempt 50 free-throws. He makes 32 of them. His sample proportion of made shots is $32/50 = 0.64$.

What can we conclude about the claim based on this sample data?

We can use software to simulate 400 sets of 50 shots assuming that the player is really an 80% shooter.

You can assess the strength of the evidence against the player?s claim is by computing the probability that he would make as few as 32 out of 50 free throws if he really makes 80% in the long run (i.e., if indeed his long run probability of a successful throw is 0.8).
We find that under the assumption of the actual rate (parameter value) of 0.8, the value of the observed statistic (0.64) is highly unlikely. In other words, the probability that an 80% shooter would out of random chance throw 64% or less is very small.

This is then treated as evidence against the player's claim.

1. Stating Hypothesis

Definitions:

(1) The claim tested by a statistical test is called the **null hypothesis** (H_0).

(2) The claim about the population that we are trying to find evidence for is the **alternative hypothesis** (H_A).

The alternative is **one-sided** if it states that a parameter is *larger* or *smaller* than the null hypothesis value.

The alternative is **two-sided** if it states that a parameter is different from the null hypothesis value.

Write the null and alternative hypotheses in the following examples:

- Harvard claims that the average GRE score of an incoming grad student is 1575. You want to test this as your friend who scored 1500 made it in the first round.

- A math instructor would like to test if students on average do equally well on proofs and problems. Let μ_o and μ_r be the average scores (each out of 100) on proofs and problems respectively. The hypothesis can be written as:

- The Department of Education claims that 9 out of 10 students who manage to graduate repay their student loans on time. You want to test this.

- United Airlines claims that the average delay on a United flight is 12 minutes. You want to test this as you think it is an understatement.

- Philips advertizes that their bulbs last an average of 1500 hours. You want to test this as the Philips bulbs at home seem to burn out very quickly.

  Let μ_P be the average life in hours of Philips bulbs. The hypothesis can be written as:

2. How do we test a hypothesis?

Idea: Let us assume for a moment that the null hypothesis or H_0 is true.

Question being asked: Under this assumption, what is the probability of observing a value of the sample statistic as the one on hand or something even more extreme?

If the probability computed is ”small”, then,

If the probability computed is ”large”, then, How do you compute this probability? ____________________.

2.1. p**-value:** Under reasonable assumptions (such as normality of the population), a **test statistic** (such as a z-score) calculated from the sample data measures the probability of observing the sample statistic value under the null hypothesis H_0 (or assuming for a moment that the H_0 were true).

Extreme values of the test statistic indicate that the information from the sample does not support the H_0.

Definition: The probability computed assuming H_0 is true, that the statistic would take a value as extreme as or more extreme than the one actually observed is called the p-**value** of the test.

The smaller the p-value (or the probability), ______________________________________

____.

The larger the p-value (or the probability), ____

____.

Steps involved in Hypothesis Testing:

(1) State the null and alternative hypotheses.

(2) Check assumptions.

(3) Compute test statistic.

(4) Find p-value.

(5) State conclusion **in terms of the null hypothesis or H_0.**

Example: (Tire Life) Goodyear advertises that the average life of their snow tires in Salt Lake City is 60 months.

The publisher Consumer Reports, conducts a survey on 100 randomly selected people who bought the tires several years ago. The mean reported life of the tires of the 100 respondents is 46 months.

Assuming that the life of snow tires is normally distributed with a standard deviation of 5 months, test hypothesis that Goodyear tires last an average of 60 months.

Example: (Job Satisfaction) Does the job satisfaction of assembly-line workers differ when their work is machine-paced rather than self-paced?

One study chose 18 subjects at random from a company with over 200 workers who assembled electronic devices. Half of the workers were assigned at random to each of two groups. Both groups did similar assembly work, but one group was allowed to pace themselves while the other group used an assembly line that moved at a fixed pace.

After two weeks, all the workers took a test of job satisfaction. Then they switched work setups and took the test again after two more weeks. The response variable is the difference in satisfaction scores, self-paced minus machine-paced.

What is the parameter of interest in this problem?

State appropriate hypotheses for performing a significance test.

In the matched pairs study described earlier, data from 18 workers gave a sample mean score of 17 as the difference in satisfaction scores, self-paced minus machine-paced. Perform a significance test to address whether we have significant evidence for a difference in the mean job satisfaction. Assume that the difference in job satisfaction scores follows a Normal distribution with standard deviation $\sigma = 60$.

Test the hypothesis that there is no difference in job satisfaction between self-paced and machine-paced workers.

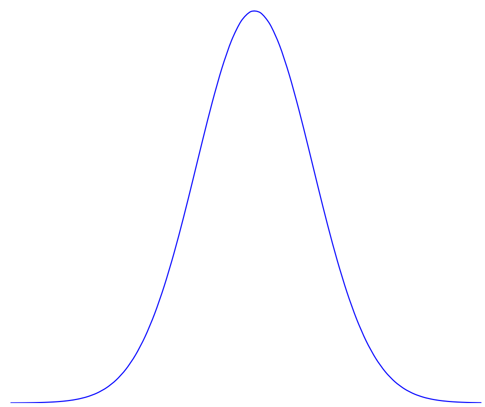
3. Statistical significance

We have been making conclusions based on whether the p-value was *small* or *large*. This begs the question:

How small is *small* and how large is *large*?

3.1. **Significance level:** In general, there is no universal rule for the size of p-values. It depends on the specific situation and is a matter of judgement.

But, we may use a **significance level** or a general standard for evidence against the H_0 (denoted by the letter α) to compare with the p-value.



If the p-value is smaller than the significance level or α, we say that the test result is **statistically significant** and reject H_0.

Finding **statistical significance** may be interpreted to mean that the result is *has a low likelihood of happening just by chance.*

$\alpha = 0.05$ implies that the chance of

In summary,

If p-value $< \alpha \implies$ we reject H_0.

If p-value $\geq \alpha \implies$ we fail to reject H_0.

Note:

(1) Significance level depends on the nature of the alternative hypothesis, or H_A. How is the p-value for a one-sided test related to the p-value for a two-sided test?

(2) Sample size could affect significance of a test. If n is small, then,

(3) Conducting multiple hypothesis tests simultaneously

3.2. **How small a p-value is really convincing?** Is α always equal to 0?

- If H_0 represents an assumption that people have believed in for years, strong evidence (very small p-value) will be needed to persuade them otherwise.

- If the consequences of rejecting H_0 are great (such as making an expensive or difficult change from one procedure or type of product to another), then strong evidence as to the benefits of the change will be required.

- Although $\alpha = 0.05$ is a common cut-off for the p-value, there is no set border between *significant* and *insignificant*, only increasingly strong evidence against H_0 (in favor of H_A) as the p-value gets smaller.

4. Cautions and Assumptions

Recall: The assumptions that go into Significance or Hypothesis Testing are:

(1) Simple Random Sample (SRS)
- Different experimental designs lead to different tests.
- The z score is applicable only when we have an SRS. Recall that $z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ only follows because it is initially assumed that $\bar{X} \sim N(\mu_0, \frac{\sigma}{\sqrt{n}})$. This assumption may not hold if the sample is not a SRS even if the population is Normal.

(2) Normality of population
- Skewness and outliers make the z procedures untrustworthy unless the sample is large.
- In practice, the z procedures are reasonably accurate for samples of at least moderate size from a fairly symmetric distribution.

(3) σ is known
- Rarely true in practice.
- In practice, the z procedures are reasonably accurate for samples of at least moderate size from a fairly symmetric distribution.
- Chapters 20 and 21 will introduce procedures for when σ is unknown.

Chapter 17 Objectives:

- Describe the reasoning of tests of significance
- Describe the parts of a significance test
- State hypotheses
- Define p-value and statistical significance
- Conduct and interpret a significance test for the mean of a Normal population