THE LOGIC OF HYPOTHESIS TESTING

☐Hypothesis testing is a statistical procedure that allows researchers to use sample to draw inferences about the population of interest.

☐It is the most commonly used inferential procedures.

☐The general process of hypothesis testing remains constant from one situation to another.

☐**The logic behind the hypothesis testing is as follows:**

1)State a hypothesis about a population. The hypothesis usually concerns the value of a population parameter.

2)Before we actually select a sample, we use the hypothesis to predict the characteristics that the sample should have.

3)We obtain a random sample from the population

4)We compare the obtained sample data with the prediction that was made from the hypothesis.

•If the sample mean is consistent with prediction we conclude that hypothesis is reasonable.

•If the discrepancy between the data and the prediction is big, we decide that the hypothesis is wrong.

Hypothesis testing has a formal structure which is a four-step process.

1.STEP 1: STATE THE HYPOTHESIS

2.STEP 2: SET THE CRITERIA FOR A DECISION

3.COLLECT DATA AND COMPUTE SAMPLE STATISTIC

4.MAKE A DECISION

UNCERTAINTY AND ERRORS IN HYPOTHESIS TESTING

☐In hypothesis test, there are two different kinds of errors that can be made.

☐Type 1 Errors: The researcher concludes that a treatment does have an effect when in fact it has no effect.

☐Type 1 Error occurs when a researcher unknowingly obtains an extreme/unrepresentative sample

☐Alpha level determines the probability of a Type 1 Error.

☐Type 2 Errors mean that a treatment has a real effect, but the hypothesis test fails to detect it.

☐Type 2 error generally occurs when the effect of treatment is relatively small

Type 2 Error is represented by the symbol β

The alpha level for a hypothesis test serves two very important functions:

It helps determine the boundaries of the critical region by defining the concept of "very unlikely" outcomes.

It determines the probability of a Type 1 Error.

The primary concern when selecting an alpha level is to minimize the risk for a Type I error.

Conventionally, we use α=.05

However, to reduce the probability of Type I error, we can use more conservative alpha levels such as α=.01 or α=.001

Importantly, when minimize the risk for Type I error, the hypothesis test demands more evidence from research results.

FACTORS THAT INFLUENCE A HYPOTHESIS TEST

A large value for z is grounds for concluding that the treatment has a significant effect.

There are three factors that can influence the outcomes of a hypothesis test.

The size of the difference between sample mean and original population mean

Larger the difference betwen the sample mean and the population mean is, the larger the z-score will be, and the greater the likelihood that we will find a significant treatment effect.

The variability of scores. It influences the size of standard error.

Higher variability reduce the chances of finding a significant treatment effect.

The number of scores in the sample. It influences the size of standard error.

Increasing the sample size decrease standrad errro and increase the likelihood of finding a significant result.

ASSUMPTIONS FOR HYPOTHESIS TEST WITH Z-SCORE

Random Sampling: It ensures that our sample is representative of our population

Independent Observations: Two events are independent if the occurrance of the first event has no effect on the probability of the second event.

The value of σ is unchanged by the treatment

Normal Sampling Distribution.

ONE-TAILED AND TWO-TAILED HYPOTHESIS TESTS

□Previously, two-tailed hypothesis test was introduced.

□The term two-tailed test comes from the fact that the critical region is located in both tails of the distribution.

□In two-tailed hypothesis test, the researcher does not make a specific prediction about the direction of the treatment effect.

□There is an alternative form of hypothesis testing: One-tailed test

□In one-tailed test or directional hypothesis test, the statistical hypotheses specify either an increase or a decrease in the population mean score.

COMPARISON OF ONE-TAILED VERSUS TWO-TAILED TESTS

□Major distinction between one-tailed and two-tailed tests is in the criteria they use for rejecting null hypothesis. One tailed test requires relatively small amount of difference between the sample and population compared to two tailed test.

□Two tailed tests are more convincing than one-tailed ones. Two tailed tests demand more evidence to reject null hypothesis

□One tailed tests are more sensitive that is, they can detect a relatively small effect.

□One tailed test are more precise because they make directional prediction

MEASURING EFFECT SIZE

□Although the hypothesis testing is the most commonly used technique for evaluating and interpreting data, there is a variety of concerns about hypothesis testing procedure.

□The most important criticism is about the interpretation of a significant result.

□A significant treatment effect does not necessarily indicate a substantial treatment effect. In particular, statistical significance does not provide any real information about the absolute size of a treatment effect.

□When a treatment effect is small, it could be large enough to be significant.

□Whenever, we report a significant treatment effect, it is recommended that we should report effect size

STATISTICAL POWER

☐An alternative approach to determine size or strength of a treatment effect is to measure the power of statistical test.

☐Power of a statistical test is defined as *the probability* that the test will identify a treatment effect if one really exists.

☐Researchers usually calculate the power of a hypothesis test before they actually conduct the research study.

Factors affecting power

☐**Treatment Effect:** as the size of treatment effect increases, the power of the test also increases.

☐**Sample Size:** as the sample size increases, the power of the test also increases.

☐**Alpha Level:** as the alpha level decreases, the power of the test also decreases.

☐**One-tailed versus Two-tailed tests.** Changing from two-tailed test to a one-tailed test will increase the power of the hypothesis test.

HYPOTHESIS TESTING AND POWER OF TEST

Example: Hypothesis Test with z

A researcher begins with a known population—in this case, scores on a standardized test that are normally distributed with $\mu=65$ and $\sigma=15$. The researcher suspects that special training in reading skills will produce a change in the scores for the individuals in the population. Because it is not feasible to administer the treatment (the special training) to everyone in the population, a sample of n = 25 individuals is selected, and the treatment is given to this sample. Following treatment, the average score for this sample is M = 70. Is there evidence that the training has an effect on test scores? Use $\alpha = .05$

    i.   <u>State the hypothesis</u>

The null hypothesis states that the special training has no effect. In symbols,

H_0: $\mu=65$ (After special training, the mean is still 65.)

The alternative hypothesis states that the treatment does have an effect.

H_1: $\mu \neq 65$ (After training, the mean is different from 65.)

    ii.   <u>Locate the critical region</u>

With a $\alpha = .05$, the critical region consists of sample means that correspond to z-scores beyond the critical boundaries of z = ±1.96.

    iii.   <u>Obtain the sample data, and compute the test statistic.</u>

For this example, the distribution of sample means, according to the null hypothesis, is normal with an expected value of μ=65 and a standard error of

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{25}} = 3$$

In this distribution, our sample mean of M = 70 corresponds to a z-score of

$$z = \frac{M - \mu}{\sigma_M} = \frac{70 - 65}{3} = 1.67$$

iv. <u>Make a decision about H₀, and state the conclusion.</u>

The z-score we obtained is not in the critical region. This indicates that our sample mean of M = 70 is not an extreme or unusual value to be obtained from a population with μ=65. Therefore, our statistical decision is to fail to reject H_0. Our conclusion for the study is that the data do not provide sufficient evidence that the special training changes test scores.

Effect Size Using Cohen's d

We will compute Cohen's d using the research situation and the data from above example Again, the original population mean was μ=65 and, after treatment (special training), the sample mean was M = 70. Thus, there is a 5-point mean difference. Using the population standard deviation, σ=15, we obtain an effect size of

$$Cohen's\ d = \frac{mean\ difference}{standard\ deviation} = \frac{M - \mu}{\sigma} = \frac{70 - 65}{15} = 0.33$$

this is a medium treatment effect.

EXERCISES

1. As the power of a test increases, what happens to the probability of a Type II error?

2. For a 5-point treatment effect, a researcher computes power of p = 0.50 for a two-tailed hypothesis test with a $\alpha = .05$.

a. Will the power increase or decrease for a 10-point treatment effect?

b. Will the power increase or decrease if alpha is changed to $\alpha = .05$ to $\alpha = .01$?

c. Will the power increase or decrease if the researcher changes to a one-tailed test?

3. How does sample size influence the power of a hypothesis test?

4. A researcher administers a treatment to a sample of n = 16 individuals selected from a normal population with μ=60 and σ=12. If the treatment increases

scores by 4 points, what is the power of a two-tailed hypothesis test with $\alpha = .05$?

ANSWERS

1. As power increases, the probability of a Type II error decreases.

2.

a. The hypothesis test is more likely to detect a 10-point effect, so power will be greater.

b. Decreasing the alpha level also decreases the power of the test.

c. Switching to a one-tailed test will increase the power.

3. Increasing sample size increases the power of a test

4. With standard error of 3 points, the critical boundary of z = 1.96 corresponds to a sample

mean of M = 65.88. With a 4-point effect, the distribution of sample means has μ= 64 and

a sample mean of M= 65.88 corresponds to z = 0.63. Power = p(z > 0.63) = 0.2643.

5. A researcher selects a sample from a population with $\mu=70$ and $\sigma=12$. After administering a treatment to the individuals in the sample, the researcher computes Cohen's d= 0.25. What is the mean for the sample?

Answer

5. There is a 3-point difference between the sample mean and μ=70, so the sample mean is either 73 or 67.

6.The value of the z-score in a hypothesis test is influenced by a variety of factors. Assuming that all other variables are held constant, explain how the value of z is influenced by each of the following:

a. An increase in the difference between the sample mean and the original population mean.

b. An increase in the population standard deviation.

c. An increase in the number of scores in the sample

7.Childhood participation in sports, cultural groups, and youth groups appears to be related to improved self-esteem for adolescents (McGee, Williams, Howden-Chapman, Martin, & Kawachi, 2006). In a representative study, a sample of n = 100 adolescents with a history of group participation is given a standardized self-esteem questionnaire. For the general population of adolescents, scores on this questionnaire form a normal distribution with a mean of $\mu=50$ and a standard deviation of $\sigma=15$. The sample of group-participation adolescents had an average of M = 53.8.

a. Does this sample provide enough evidence to conclude that self-esteem scores for these adolescents are significantly different from those of the general population? Use a two-tailed test with $\alpha=.05$.

b. Compute Cohen's d to measure the size of the difference.

8.Briefly explain how increasing sample size influences each of the following. Assume that all other factors are held constant.

a. The size of the z-score in a hypothesis test.

b. The size of Cohen's d.

c. The power of a hypothesis test.

9.A researcher is investigating the effectiveness of a new medication for lowering blood pressure for individuals with systolic pressure greater than 140. For this population, systolic scores average $\mu=160$ with a standard deviation of $\sigma=20$, and the scores form a normal-shaped distribution. The researcher plans to select a sample of n = 25 individuals, and measure their systolic blood pressure after they take the medication for 60 days. If the researcher uses a two-tailed test with $\alpha= .05$,

a. What is the power of the test if the medication has a 5-point effect?

b. What is the power of the test if the medication has a 10-point effect?

10.A researcher is evaluating the influence of a treatment using a sample selected from a normally distributed population with a mean of $\mu=80$ and a standard deviation of $\sigma=20$. The researcher expects a 12-point treatment effect and plans to use a two-tailed hypothesis test with $\alpha= .05$.

a. Compute the power of the test if the researcher uses a sample of n = 16 individuals.

b. Compute the power of the test if the researcher uses a sample of n = 25 individuals.