Module 1
Introduction to Hypothesis Testing

Lesson 1: Elements of Hypothesis Testing

Lesson Summary

Hypothesis testing is a statistical procedure which uses data to decide whether to reject or not one of two competing statements about the parameter(s) of a distribution. The procedure includes (1) stating the null and alternative hypotheses, (2) calculating the test statistic based on sample data, (3) constructing the rejection rule, and (4) making a decision on the null hypothesis.

Challenge/Motivation

Suppose your colleague claims that the mean claim amount θ for a new class of customers is larger than the mean amount for the standard class of customers, known to be θ_0. How can we determine (statistically) if there is evidence to support this claim? Here, it makes sense to think of two competing hypotheses:

$$H_0 : \theta = \theta_0$$
$$\text{versus}$$
$$H_a : \theta > \theta_0$$

Here, H_0 says that the mean claim amount for the new class of customers, θ, is the same as the mean claim amount for the standard class, θ_0. While, H_a says that the mean claim amount for the new class of customers, θ, is larger than the mean claim amount for the standard class, θ_0, that is, your colleague's claim is correct. Based on a sample of claim amounts Y_1, Y_2, ..., Y_n from the new class of customers, how should we formally decide between H_0 and H_a? This question can be answered by performing a hypothesis test.

Learning Outcomes

At the end of the lesson, the students should be able to explain the different elements of hypothesis testing.

Instructional Materials and Equipment Needed

Learning module, calculator, PPT slides, Recorded Video of the Lecture

Teaching Strategies

Lecture/ Discussion, Demonstration

Discussion

A statistical hypothesis is a statement concerning the probability distribution of a random variable or population parameters that are inherent in a probability distribution (Ramachandan and Tsokos, 2009). In statistics, a hypothesis is a statement about a population parameter, where this statement typically is represented by some specific numerical value. A *test* of a statistical hypothesis is a rule or procedure for deciding whether to reject a hypothesis (Mood, Graybill, and Boes, 1974).

The first step in hypothesis testing is to set up two competing hypotheses. These hypotheses are the most important elements in hypothesis testing. If the hypotheses are incorrect, your conclusion will also be incorrect. The two hypotheses are named the null hypothesis and the alternative hypothesis. The **null hypothesis** is typically denoted as H_0. The null hypothesis states the "status quo". This hypothesis is assumed to be true until there is evidence to suggest otherwise.

On the other hand, the **alternative hypothesis** is typically denoted as H_a or H_1. This is the statement that one wants to conclude. It is also called the researcher's hypothesis because it is usually the hypothesis that we seek to support on the basis of the information contained in the sample (Wackerly, Mendenhall & Scheaffer, 2008).

To illustrate the above concepts, consider the following examples.

Example 1

Suppose your colleague claims that the mean claim amount θ for a new class of customers is larger than the mean amount for the standard class of customers, known to be $\theta_0 = 10$. The parameter of interest here is the mean θ, hence, the two competing hypotheses are:

$H_0 : \theta = 10$

versus

$H_a : \theta > 10$

Example 2 (From Wackerly, Mendenhall & Scheaffer, 2008)

A political candidate, Jones, claims that he will gain at least 50% of the votes in a city election and thereby emerge as the winner. For this situation the parameter of interest is the proportion (P) of voters in favor of Jones, thus, the null and alternative hypotheses are:

$H_0 : P \geq 0.50$

versus

$H_a : P < 0.50$

Example 3

A psychologist was interested in exploring whether or not male and female college students have different driving behaviors. There were a number of ways that she could quantify driving behaviors. She opted to focus on the fastest speed ever driven by an individual. Therefore, the particular statistical question she framed was as follows:

Is the mean fastest speed driven by male college students different than the mean fastest speed driven by female college students?

In this situation the parameters of interest are the mean driving speed of male (μ_M) and the mean driving speed female (μ_F) students. Thus, the null and alternative hypotheses are:

$H_0 : \mu_M = \mu_F$

versus

$H_a : \mu_M \neq \mu_F$

Example 4

The Director of CHED Eastern Visayas wishes to verify if there is data to support the report that percentage of passers in the Veterinary Medicine Board Exam of University 1 is higher than that of University 2. In this example, the parameters of interest are the proportions of board passers for University 1 (P_1) and 2 (P_2), respectively. Therefore, the null and alternative hypotheses can be stated as

$H_0 : P_1 = P_2$

$H_a : P_1 > P_2$

Generally, the alternative hypothesis can be of two types: *one-sided* (*directional*) or *two-sided* (*non-directional*). The alternative hypotheses in Examples 1, 2, and 4 are one-sided while the alternative hypothesis in Example 3 is two-sided.

The goal of hypothesis testing is to see if there is enough evidence against the null hypothesis. This evidence is contained in the form of a test statistic. A **test statistic** is a random variable that is calculated from sample data and used in a hypothesis test. That is, the test statistic is used to determine whether to reject the null hypothesis. The test statistic compares your data with what is expected under the null hypothesis (Wackerly, Mendenhall & Scheaffer, 2008).

If the observed value of the test statistic is consistent with its sampling distribution under H_0, then this is not evidence for H_a, hence, H_0 is not rejected. On the other hand, if the observed value of the test statistic is not consistent with its sampling distribution under H_0, and it is more consistent with the sampling distribution under H_a, then this is evidence for H_a, and therefore, H_0 is rejected.

The sampling distribution of the test statistic is divided into two regions: *rejection* and *non-rejection*. The **rejection region**, denoted by RR, specifies the values of the test statistic for which H_0 is rejected. The rejection region is usually located in the tails of the sampling distribution of the test statistic computed under H_0. This is why we take H_0 to be sharp (*using the "=" sign*) so that we can construct a single sampling distribution. Thus, in Example 2, we may instead use $H_0 : P = 0.50$.

Example 5.

In Example 2, a test statistic that we might consider is the total number of persons ($T = \sum_{i=1}^{15} Y_i$) in a given sample of some size, say *n=15*, who are in favor of Jones and a possible rejection region would be all values of T less than or equal to, say, 2. That is, $RR = \{t : t \le 2\}$. Why the RR is like that? Because we seek a region that supports the alternative hypothesis and will lead to the rejection of H_0.

For any fixed rejection region (determined by a particular value of *k*), two types of errors can be made in reaching a decision on H_0. A *Type I error* is made if H_0 is rejected when H_0 is true. The *probability of a Type I error* is denoted by α. The value of α is called the *level* of the test or *size* of the rejection region. A *Type II error* is made if H_0 is not rejected when H_a is true. The *probability of a Type II error* is denoted by β.

The probabilities α and β measure the risks associated with the two possible erroneous decisions that might result from a statistical test. As such, they provide a very practical way to measure the goodness of a test (Wackerly, Mendenhall, and Scheaffer, 2008).

Example 6

In Example 1, a Type I error is committed when one concludes that the mean claim amount is greater than 10 when I fact it is equal to 10. While a Type II error is committed when we conclude that the mean claim amount is equal to 10 when in fact it is greater than 10.

Challenge:

Describe the Type I and Type II errors for Examples 2 thru 4.

Example 7

In Example 2, we wish to test $H_0: P = 0.50$ against $H_a: P < 0.50$ using the test statistic $T = \sum_{i=1}^{15} Y_i$, the total number of persons in a random sample (*n=15*) who are in favor of Jones with the possible rejection region $RR = \{t : t \le 2\}$. The probability of wrongly rejecting H_0 when in fact it is true is

$$\begin{aligned}
\alpha &= P(\text{Type I error}) \\
&= P(\text{value of test statistic is in RR when } H_0 \text{ is true}) \\
&= P(T \le 2 \text{ when } P = 0.50)
\end{aligned}$$

Observe that T is a binomial random variable with $n = 15$. If H_0 is true, $P = 0.50$ and

$$\begin{aligned}
\alpha &= P(T=0) + P(T=1) + P(T=2) \\
&= \binom{15}{0} 0.50^0 (0.50)^{15} + \binom{15}{1} 0.50^1 (0.50)^{14} \binom{15}{2} 0.50^2 (0.50)^{13} \\
&= 0.003692627 \\
&\cong 0.004
\end{aligned}$$

Note that this probability can be obtained using the R command *pbinom(2, 15, 0.5, lower.tail=T).*

Example 8

Is our test equally good in protecting us from concluding that Jones is a winner if in fact he will lose? Suppose that he will receive 30% of the votes *(P* = 0.30*)*. What is the probability β that the sample will erroneously lead us to conclude that H_0 is true and that Jones is going to win?

By definition,

$$\begin{aligned}
\beta &= P(\text{Type II error}) \\
&= P(\text{not rejecting } \mathrm{H_0} \text{ when } \mathrm{H_a} \text{ is true}) \\
&= P(\text{value of test statistic is not in RR when } \mathrm{H_a} \text{ is true}) \\
\beta &= P(T > 2 \text{ when } P = 0.30) \\
&= 0.8731723 \\
&\cong 0.873
\end{aligned}$$

Example 9

Suppose that industrial accidents occur according to a Poisson process with mean $\theta = 20$ per site per year. New safety measures have been put in place to decrease the number of accidents at industrial sites. Suppose that after implementation of the new measures, we will observe the number of accidents for a sample of $n = 10$ sites. Denote these data by $Y_1, Y_2, ..., Y_{10}$. We are interested in testing $H_0: \theta = 20$ versus $H_a: \theta < 20$.

To perform the test, suppose we use the test statistic $T = \sum_{i=1}^{15} Y_i$ and the rejection region $RR = \{t : t \le 175\}$.

Questions:

a) What is the distribution of T when H_0 is true?
b) What is $\alpha = P$(Type I Error) for this RR?
c) Suppose that $\theta = 18$, that is, H_a is true. What is the probability of Type II Error when using this RR?

Answers:

a) Recall that the sum of n iid Poisson(θ) random variables is distributed as Poisson($n\theta$). Therefore, when H_0: θ = 20 is true, $T \sim$ Poisson(200).

b) The probability of Type I error is

$$\begin{aligned}\alpha &= P(\text{Type I error}) \\ &= P(\text{Reject } H_0 \mid H_0 \text{ is true}) \\ &= P(T \le 175 \mid \theta = 20) \\ &= \sum_{t=0}^{175} \frac{200^t e^{-200}}{t!} \\ &= 0.03940167 \\ &\cong 0.039\end{aligned}$$

This time this probability was obtained using the *ppois(175, 200, lower.tail=T)* command in R.

c) First, note that if θ = 18 then $T \sim$ Poisson(180). Therefore, the probability of Type II Error when θ=18 is

$$\begin{aligned}\beta &= P(\text{Type II error}) \\ &= P(\text{Do not reject } H_0 \mid \theta = 18) \\ &= P(T > 175 \mid \theta = 18) \\ &= \sum_{t=176}^{\infty} \frac{180^t e^{-180}}{t!} \\ &= 0.6271594 \\ &\cong 0.627\end{aligned}$$

This probability was obtained using *ppois(175, 180, lower.tail=F)* or *1-ppois(175, 180, lower.tail=F)* command in R.

The above calculations show three (3) things:

1. For the rejection region $RR = \{t : t \le 175\}$, the probability of Type I Error is small, $\alpha \cong 0.039$. This assures us that if H_0 is really true (and that the new safety measures, in fact, did not work), then we are not likely to reject H_0.

2. However, if implementing the safety measures did work and the mean number of accidents per site/per year actually decreased to θ = 18 (i.e., H_a is true), then we are still likely not to reject H_0 since $\beta = \beta(18) \cong 0.627$.

3. We have only about a 37% chance of correctly concluding that the safety measures actually worked since
$$P(\text{Do not reject } H_0 \mid \theta = 18) \cong 0.627 \Leftrightarrow P(\text{Reject } H_0 \mid \theta = 18) \cong 0.373$$

Finding a good rejection region for a statistical test is an interesting problem that merits further attention (Wackerly, Mendenhall, and Scheaffer, 2008). In Examples 7 and 9, we intuitively choose the rejection region of the form $RR = \{t : t \le k\}$. But what value should we choose for k? More generally, we seek some objective criteria for deciding which value of k specifies a good rejection region of the form

$\{t \leq k\}$. Is the rejection region always of the form $\{t \leq k\}$? We shall answer these questions in the following examples.

Example 10

Suppose that $Y_1, Y_2, ..., Y_{25}$ is an iid sample of n = 25 observations from a $N(\theta, 100)$ distribution. We would like to test $H_0: \theta = 75$ versus $H_a: \theta > 75$. To perform the test, suppose we use the test statistic

$$\bar{Y} = \frac{1}{25}\sum_{i=1}^{25} Y_i$$

and the rejection region $RR = \left\{\bar{y} : \bar{y} \geq k\right\}$, where k is a constant.

Questions:

a) What is the distribution of $\bar{Y}$ when H_0 is true?
b) Find the value of k that provides a level α = 0.10 test.
c) Suppose that θ = 80, that is, H_a is true. What is the probability of Type II Error when using this RR?

Answers:

a) Recall that, generally, $\bar{Y} \sim N\left(\theta, \dfrac{\sigma^2}{n}\right)$. Thus, if H_0 is true, that is, $\theta = 75$, then $\bar{Y} \sim N(75, 4)$.

b) To find the value of k, we set $\alpha = 0.10$. That is,

$$\begin{aligned}\alpha = 0.10 &= P(\text{Type I error}) \\ &= P(\text{Reject } H_0 | H_0 \text{ is true}) \\ &= P(\bar{Y} > k \mid \theta = 75) \\ &= P\left(Z > \frac{k-75}{2}\right)\end{aligned}$$

where $Z \sim N(0,1)$. Looking at the Z table for $\alpha = 0.10$, the z score is 1.28, this means that

$$1.28 = \frac{k-75}{2} \Leftrightarrow k = 77.56.$$

The rejection region $RR = \left\{\bar{y} : \bar{y} \geq 77.56\right\}$ confers a Type I error probability of $\alpha = 0.10$.

c) First, note that when θ = 80, $\bar{Y} \sim N(80, 4)$. Therefore, the probability of Type II error, when $\theta = 80$ is

$$\begin{aligned}\beta = \beta(80) &= P(\text{Do not reject } H_0 \mid \theta = 80) \\ &= P(\bar{Y} \leq 77.56 \mid \theta = 80) \\ &= P\left(Z \leq \frac{77.56-80}{2}\right) \\ &= P(Z \leq -1.22) \\ &= 0.1112 \text{ (from the Z table)}\end{aligned}$$

You can obtained this probability using the *pnorm(-1.22, 0, 1)* command in R.

Example 11

Suppose that $Y_1, Y_2, ..., Y_n$ is an iid Bernoulli(p) sample, where n = 100. We would like to test $H_0: P = 0.10$ versus $Ha: P < 0.10$. To perform the test, suppose we use the test statistic

$$T = \sum_{i=1}^{100} Y_i$$

and the rejection region $RR = \{t : t \le k\}$, where k is a constant.

Questions:

a) What is the distribution of T when H_0 is true?
b) Is it possible to find an exact level $\alpha = 0.05$ rejection region?
c) With $k = 5$, find the probability of Type II Error when $P = 0.05$.

Answers:

a) In general, we recall that if $Y_1, Y_2, \ldots, Y_n$ is an iid Bernoulli(p) sample, then the (sufficient) statistic

$$T = \sum_{i=1}^{n} Y_i \sim B(n, p).$$

Therefore, when H_0 is true and $n = 100$, we have $T = \sum_{i=1}^{100} Y_i \sim B(100, 0.10)$.

b) The value of k is chosen so that

$$\begin{aligned}\alpha &= P(\text{Reject } H_0 \mid H_0 \text{ is true}) \\ &= P(T \le k \mid P = 0.10) \\ &= \sum_{t=0}^{k} \binom{100}{t} 0.10^t (0.90)^{100-t}\end{aligned}$$

Using the R command *pbinom(k, 100, 0.10)* we obtained the following results:

$$k = 3 \Leftrightarrow \alpha \cong 0.008$$
$$k = 4 \Leftrightarrow \alpha \cong 0.024$$
$$k = 5 \Leftrightarrow \alpha \cong 0.058$$
$$k = 6 \Leftrightarrow \alpha \cong 0.117$$

This means that there is no value of k that will give us an exact level $\alpha = 0.05$ rejection region of the form $RR = \{t : t \le k\}$.

c) If k=5 and P=0.05, the level $\alpha = 0.058$ rejection region is of the form $RR = \{t : t \le 5\}$ and $T \sim B(100, 0.05)$. Therefore,

$$\begin{aligned}\beta = \beta(0.05) &= P(\text{Do not reject } H_0 \mid P = 0.05) \\ &= P(T > 5 \mid P = 0.05) \\ &= \sum_{t=6}^{100} \binom{100}{t} 0.05^t (0.95)^{100-t} \\ &\cong 0.384\end{aligned}$$

This probability was obtained using the R command *1-pbinom(5,100,0.05).*

Evaluation

Part I: In each of the following scenarios, determine the parameter of interest and the null and alternative hypotheses. Indicate also the type of alternative hypothesis. Answers are provided.

1. When debating the budget appropriation for Visayas State University (VSU), the following question is asked: "Are the majority of students at VSU from Eastern Visayas?"

   Answer:

2. A consumer test agency wants to see the whether the mean lifetime of a brand of tires is less than 42,000 miles. The tire manufacturer advertises that the average lifetime is at least 42,000 miles.

   Answer:

Part II: Answer the following as indicated.

An experimenter has prepared a drug dosage level that she claims will induce sleep for 80% of people suffering from insomnia. After examining the dosage, we feel that her claims regarding the effectiveness of the dosage are inflated. In an attempt to disprove her claim, we administer her prescribed dosage to 20 insomniacs and we observe *Y*, the number for whom the drug dose induces sleep. We wish to test the hypothesis $H_0: P = 0.80$ versus the alternative, $H_a: P < 0.80$. Assume that the rejection region $RR = \{y : y \le 12\}$ is used.

a) In terms of this problem, what is a Type I error?
b) Find α.
c) In terms of this problem, what is a Type II error?
d) Find β when P = 0.60.
e) Find the rejection region of the form $RR = \{y : y \le c\}$ so that $\alpha \approx 0.01$.

Answer

a)

b)

c)

d)

e)

References

MENDENHALL, W., SCHEAFFER, R. L., AND WACKERLY, D. D. (2008). *Mathematical Statistics with Applications*, 7^{th} ed. Brooks/Cole, Cenage Learning.

MOOD, A. M., GRAYBILL, F. A. and BOES, D. C. (1976). *Introduction to the Theory of Statistics*, 3^{rd} Ed. McGraw-Hill, Inc.

RAMACHANDRAN, K. M. AND TSOKOS, C. P. (2009). *Mathematical Statistics with Applications*. Elsevier Inc.

https://online.stat.psu.edu/stat414/node/266/

https://online.stat.psu.edu/stat414/node/274/

https://online.stat.psu.edu/stat500/lesson/6a/6a.3