13 Chapter 9 - Large-Sample Tests of Hypothesis

A good empirical study requires three components:

1. A concise and sensible theoretical framework that is related to the questions to be asked,
2. Reasonably good data, and
3. An experiment or an event or a set of circumstances that give the data a chance to answer the questions asked. In short, the model needs to be identifiable from the data at hand.

Zvi Griliches

This section begins the discussion of using data to test research questions. Economics is a theory-driven discipline. Theoretical economists develop theories to describe the ways resources are distributed in the market and how uncertainty and market distortions affect these choices. Applied economists use data to test these theories. The work of theoretical and applied economists complement one another.

Here is an example that describes this relationship. A fundamental theory in consumer choice is the relationship between quantity demanded and income. This relationship is illustrated using an income expansion path (your intermediate micro book should have a discussion on income expansion paths).

Suppose you have a world consisting of two goods, Spam (canned meat) and eggs. Consumers have the same preferences, they differ only by income. There is a high income group and a low income group. Unit price for both goods is \$1.00. Income (I) is spent on Spam (S) and eggs (E). the proportion of income spent on Spam is (p=S/I). Holding prices and preferences constant, the differences between the proportion of income spent by the high and low income persons (p_{high}, p_{low}) tells you if Spam is a normal good, a luxury good or an inferior good. If

$$\begin{array}{ll} p_{low} - p_{high} < 0 & \text{Spam is a luxury good} \\ p_{low} - p_{high} = 0 & \text{Spam is a normal good} \\ p_{low} - p_{high} > 0 & \text{Spam is an inferior good} \end{array}$$

A research test can be developed that applies the predictions of the theory. If a good is inferior, its proportion of the consumer's bundle of a high income person (p_{high}) should be less than its proportion of the bundle of the lower income consumer(p_{low}). The difference in proportions in the consumption bundles of low and high income persons ($p_{low} - p_{high}$) should be greater than zero. If the good is not inferior, then ($p_{low} - p_{high} \leq 0$).

In this case, the researcher can conclude that good is inferior by eliminating alternative hypotheses and failing to reject the initial hypothesis. In this case, the researcher would hope that the data

- rejects the hypothesis that ($p_{low} - p_{high} \leq 0$) (rejects that it is a normal or luxury good) and
- fails to reject the hypothesis that ($p_{low} - p_{high} > 0$) (fails to reject that the good is inferior).

The researcher needs to conduct both steps. If the research is based on one step, it will not provide solid support. For instance, if the researcher only rejects the hypothesis that ($p_{low} - p_{high} \leq 0$) and fails to test the second step, the researcher has not eliminated the possibility that data would reject all hypotheses.

Note that the second step reads 'fails to reject the hypothesis'. In spite of what the book claims, failing to reject a hypothesis is not the same as accepting the hypothesis. The hypothesis (or research question) is based on a relationship drawn from one theory. Finding out that the implications of a theory are true does not necessary imply that the theory is true. Logicians refer to this as affirming the consequent. If data supports the hypothesis that the good is inferior, we can say that the economic theory developed to describe the relationship of inferior goods and income *may* be the true theory but it is also possible that any theory from economics or other disciplines arriving that the same relationship may be the true theory.

In this chapter, we will developed the tools needed to conduct hypothesis tests. The tools described in this chapter are used with large samples. In chapter 10, we will examine the tools used for small samples.

There is a standard method of in testing a hypothesis.

1. **Make Assumptions**
   (a) Type variable (categorical, interval)
   (b) Type of population (binomial, normal, normal given the Central Limit Theorem)
   (c) Type of analysis (univariate, bivariate)
   (d) Null hypothesis (H_0), and alternative hypothesis (H_a).
2. **Obtain a Sampling distribution** - for large samples, this is the z distribution. For small sample, we used the Student's t distribution (more on this in the next chapter)
3. **Determine if the test is one tailed or two tailed test. Choose a level of significance and critical region**.

The hypothesis determines whether or not a one tailed test is done. If the hypothesis is an inequality (e.g., $z > 0, z < 1$), we can use a one tail test. If we are testing if z is a specific value, the alternative hypothesis is that z is not this value and can be any value in the distribution. For this case, we use a two tail test.

The significance level and critical region helps us identify the outcomes that would allow us to reject the hypothesis (the critical region). The level is chosen prior to the computing the test. In determining the critical level, there are two conditions to keep in mind. First, it is important to know if the critical region include one or both tails of the distribution (e.g., is it a one or two tail test). Second, the researcher must determine the risks she is willing to take in making what is known as a Type I and Type II error.

A Type I error consists of rejecting a hypothesis when in fact, it is true. A Type II error, on the other hand, involves failing to reject a hypothesis when, in fact, one should reject it. From the sampling probabilities, one can determine the exact probabilities that certain outcomes will occur if the assumptions are actually true.

The probability of making a Type I error is the sum of the probabilities of each of the outcomes falling within the critical region. For example, suppose we toss ten pennies. The probabilities of getting a success (head) is represented in a binomial distribution. The probability of no heads is equal to or 0.001. This is the same probability of getting all heads. Suppose the critical region of the distribution consists of zero successes or all successes. The probability of a Type I error is the sum of these probabilities or 0.002. The probability of making a Type I error is also known as a significance level (or α level). The higher the significance level, the higher the probability of making a Type I error.

The probability of a Type II error is β. The value of β is computed in the following manner (this describes the conditions for a two tail test). Suppose your hypothesis test is that

$$\text{HO: } \mu_0 = A$$

You want to compute a power test to determine the probability of rejecting HO when the alternative mean $\mu_a =$ C.

(a) Compute the two confidence interval values. The book uses the margin of error values but the example (9.8) assumes that the significance level is 5%. My instructions apply for all significance levels. These values are the endpoints of the Type II region. The formula for the confidence interval is

$$\mu_0 \pm z_{\frac{\alpha}{2}} SE$$

Note, function uses μ_0. From this you have the left boundary value (LBV) and right boundary value (RBV).

(b) Draw the two graphs. The left and right boundary points are points around μ_0. Determine where μ_a is relative to these boundaries and determine the rejection area of the new distribution overlapping the acceptance region of the old distribution.

(c) Compute z scores for the two values using the following functions.

$$\text{z(left boundary)} = \frac{LBV - \mu_a}{SE}$$

$$\text{z(right boundary)} = \frac{RBV - \mu_a}{SE}$$

Note: this set of functions uses μ_a.

(d) Given the drawing above, determine the p(accepting h_a when $\mu = \mu_a$). The power of the test or the probability of correctly rejecting H_0 given that μ is μ_a is (1-β).

There are several important relationships between the Type I & II probabilities and power test.

(a) Increasing the significance level reduces the confidence interval thus increasing the probability of a Type II error and reducing the power of the test.

(b) Increasing the sample size decreases the standard error. This decreases the probability of a Type II error and increases the power of the test.

(c) If μ_a is very close to μ_0, it weakens the power test.

4. **Test statistic and p-value**. The test statistic varies by distribution, sample size and type of analysis.

   P-value - actual probability measure of the score computed by the test statistic. If the p value is less than or equal to the level of significance α, one can reject the hypothesis. If the p-value is greater than α, one fails to reject the hypothesis.

5. **Make a decision** What does the analysis support, how strong is this support, has our understanding of market improved by the results of this analysis.

14 Chapter 10 - Inference from Small Samples

In Chapter 9, we examined tests that work with large samples. It was assumed that these samples were normally distributed. For a portion of these samples the property of normality was the result of an application of the Central Limit Theorem.

For small samples, we are restricted to those samples that are normal or approximately normal. Because of this, small sample test of proportion is not part of this discussion.

1. Small Sample testing of Means

   With small sample testing, we used a set of techniques modified to work with smaller samples. The major difference has to do with the choice of sampling distribution. For small sample testing of means, we use a Student's t distribution.

   Properties of Student's t distribution.

   (a) Shape is similar to a standard normal distribution. Shape is symmetric around the mean of the distribution t=0.

   (b) Distribution is described as having "heavier tails" than a standard normal. As shown in Figure 10.1 of your book, the t distribution is a bit flatter than the standard normal.

   (c) Shape of the t distribution depends on the sample size (n). As n approaches 30 (the minimum sample size for 'large' samples), the shape of the distribution approaches the standard normal.

   To work with t distributions, we need to understand the concept of **degrees of freedom**. Previously, the distribution were based on a set of values that were not constrained by any properties. The t distribution is an approximation of the standard normal. To accomplish this, we need to constrain or restrict a number of values. For the t distribution, we restrict one value. This means that n-1 of the sample can be made of a set of unknown values. The number of observations that make up this set of unknowns is the degree of freedom