LECTURE 07

HYPOTHESIS TESTING - I

Outline of today's lecture:

I. Elements of Statistical Test

- Usually, the main objective of a statistical test is to test a hypothesis concerning the values of one or more population parameters.

- We generally have a theory (*a research hypothesis)* about the parameter(s) that we wish to support.
  - For example, suppose that a vice president in charge of sales for a large corporation claims that salespeople are averaging no more than 15 sales contacts per week. (He would like to increase this figure.)
  - If we do not believe vice president's claim, we might seek to support the research hypothesis that vice president's claim is incorrect.

- Support for this research hypothesis, also called the *alternative hypothesis*, is obtained by showing (using the sample data as evidence) that the converse of the alternative hypothesis, called the *null hypothesis* is false.
    - In a sense, it is a *proof by contradiction*.
- Because we seek support for the alternative hypothesis that vice president's claim is false, our alternative hypothesis is that μ, the mean number of sales contacts per week, is higher than 15.
- If we can show that the data support rejection of the null hypothesis $\mu=15$ (the maximum value needed to confirm vice present's argument) in favor of the alternative hypothesis $\mu>15$, we have achieved our research objective.
- Although it is common to speak of testing a null hypothesis, the research objective usually is to show support for the alternative hypothesis, if such support is warranted.

- How do we utilize that data to decide between the null hypothesis and the alternative hypothesis? Suppose that n=36 salespeople are randomly selected from the company and $\bar{Y}$, the average number sales contracts that salespeople sell in this sample, is recorded. If the average sales contract of the sample is 30 ($\bar{Y}$=30), what would you conclude about vice president's claim? If, in reality, the average sales contract that salespeople sell per week (population parameter, μ) is at most 15, it is not impossible to observe $\bar{Y}$=30 in a sample of size *n*=15, but it is *highly improbable*. It is much more likely that we would observe $\bar{Y}$=30 if the alternative hypothesis were true (μ>*15*). Thus we would reject the *null* hypothesis (μ=15) in favor of the *alternative* hypothesis (μ>*15*). If we observed $\bar{Y}$=25 (or any other value of $\bar{Y}$ very much higher than 15), the same type of reasoning would lead us to the same conclusion.

- Any statistical test of hypotheses works in exactly the same way and is composed of the same essential elements.

The Elements of a Statistical Test

1. Null hypothesis, H_0.
2. Alternative hypothesis, H_A.
3. Test statistic
4. Rejection region

- For our example the hypothesis to be tested, called the *null hypothesis* and denoted by H_0, is $\mu=15$.
- The *alternative (or research) hypothesis*, denoted as H_a, is the hypothesis to be accepted in case H_0 is rejected. The alternative hypothesis usually is the hypothesis we seek to support on the basis of the information contained in the sample; thus, in our example, H_a is $\mu>15$.

- The *rejection region*, which will henceforth be denoted by RR, specifies the values of the test statistic for which the null hypothesis is to he rejected in favor of the alternative hypothesis.
  - As previously indicated, for our example high values of $\bar{Y}$ would lead us to reject H_0. Therefore, one rejection region we might want to consider is the set of all values of $\bar{Y}$ higher than or equal to 20. We will use the notation $RR = \{\bar{y} : \bar{y} \geq 20\}$ or more simply $RR = \{\bar{y} \geq 20\}$ to denote this rejection region.
  - Here, we intuitively choose the rejection region as $RR = \{\bar{y} \geq k\}$. 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 $\{\bar{y} \geq k\}$.
- For any fixed rejection region (determined by a particular value of *k*), two types of errors can be made in reaching a decision.
  - We can decide in favor of H_a, when H_0 is true (make a *type I error*), or we can decide in favor of H_0 when H_a is true (make a *type II error*).

Type I and Type II Errors

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 level of significance).

- $\alpha = P$(type I error)=P(rejecting H_0 when H_0 is true)
- $\alpha = P$(value of test statistic is in RR when H_0 is true)

A *type II error* is made if H_a is rejected when H_a is true. The *probability of a type II error* is denoted by β.

- $\alpha = P$(type II error)=P(rejecting H_a when H_a is true)

For most real situations, incorrect decisions cost money, prestige, or time and imply a loss. Thus α and β, the probabilities of making these two types of errors, measure the risks associated with the two possible erroneous decisions that might result from a statistical test.

- Note that α and β are inversely related.

II. Common Large-Sample Tests

Suppose that we want to test a set of hypotheses concerning a parameter θ based on a random sample $Y_1, Y_2, ..., Y_n$. In this section we will develop hypothesis-testing procedures that are based on an estimator that has an (approximately) normal sampling distribution with mean μ and standard error $\sigma_{\hat{\theta}}$. The large-sample estimator such as $\bar{Y}$ used for estimating a population mean μ satisfy these requirements. So do the estimator for the comparison of two means $(\mu_1 - \mu_2)$.

If θ_0 is a specific value of θ, we may wish to test $H_0: \theta = \theta_0$ versus $H_a: \theta > \theta_0$. the null and alternative hypotheses, the test statistic, and the rejection region are as follows:

$H_0: \theta = \theta_0$

$H_a: \theta > \theta_0$

Test statistic: $\hat{\theta}$

Rejection Region $RR = \{\hat{\theta} \geq k\}$ for some choice of *k*.

- The actual value of *k* in the rejection region *RR* is determined by fixing the type I error probability α (the level of the test) and choosing *k* accordingly (see Figure below).

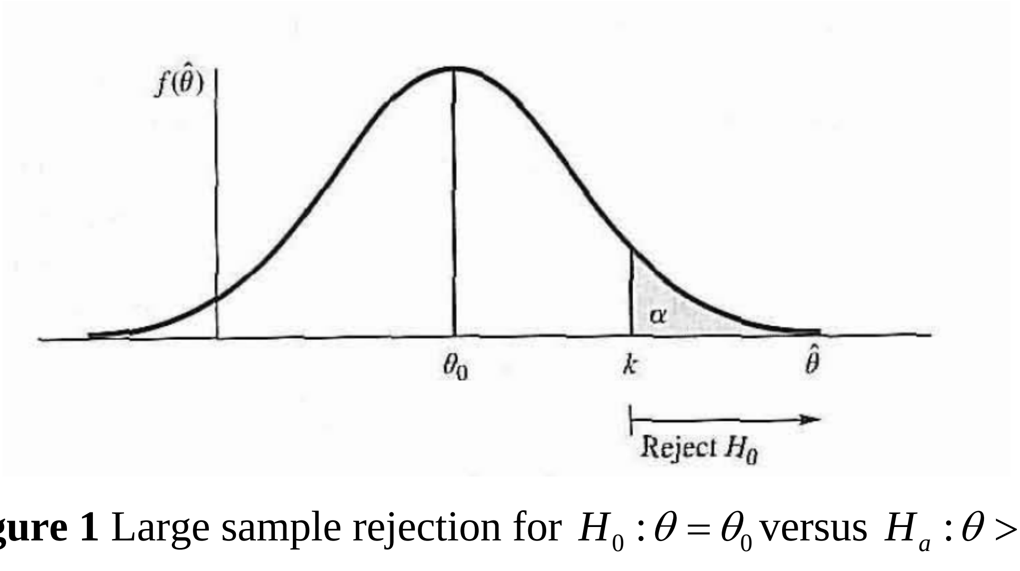

Figure 1 Large sample rejection for $H_0: \theta = \theta_0$ versus $H_a: \theta > \theta_0$

- If H_0 is true, $\hat{\theta}$ has an approximately normal distribution with mean θ_0 and standard error $\sigma_{\hat{\theta}}$. Therefore, if we desire an α-level test,

$$k = \theta_0 + z_\alpha \sigma_{\hat{\theta}}$$

is the appropriate choice for k [Z has a standard normal distribution such that $P(Z > z_\alpha) = \alpha$]

- We can write the Rejection Region as follows:

$$RR = \left\{\hat{\theta} : \hat{\theta} > \theta_0 + z_\alpha \sigma_{\hat{\theta}}\right\} = \left\{\hat{\theta} : \underbrace{\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}}_{z} > z_\alpha\right\}$$

Hence $Z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$ is the test statistic. So, the RR can also be written as :

$$RR = \left\{z > z_\alpha\right\}$$

- Consequently, an equivalent form of the test of hypothesis, with level α, is as follows:

$H_0 : \theta = \theta_0$
$H_a : \theta > \theta_0$
Test statistic: $Z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$
Rejection Region: $RR = \{z > z_\alpha\}$.

- Notice that the preceding formula for Z is simply:

$$Z = \frac{\text{estimator for the parameter-value for the parameter given by } H_0}{\text{standard error of the estimator}}$$

Example 10.5 Recall our previous example that a vice president in charge of sales for a large corporation claims that salespeople are averaging no more than 15 sales contacts per week. (He would like to increase this figure.) As a check on his claim, *n=36* salespeople are selected at random, and the number of contacts made by each is recorded for a single randomly selected week. The mean and variance of the 36 measurements were 17 and 9, respectively. Does the evidence contradict the vice president's claim? Use a test with level $\alpha = 0.05$.

Solution:

We are interested in the research hypothesis that the vice president's claim is incorrect. This can be formally written as $H_a : \mu > 15$, where μ is the mean number of sales contacts per week. Thus, we are interested in testing

$$H_0 : \mu = 15 \quad \text{against} \quad H_a : \mu > 15$$

We know that for large enough *n*, the sample mean $\overline{Y}$ is a point estimator of μ that is approximately normally distributed with $\mu_{\overline{Y}} = \mu$ and $\sigma_{\overline{Y}} = \sigma / \sqrt{n}$. Hence, our test statistic is:

$$Z = \frac{\overline{Y} - \mu_0}{\sigma_{\overline{Y}}} = \frac{\overline{Y} - \mu_0}{\sigma / \sqrt{n}}$$

The rejection region, with $\alpha = 0.05$ is given by $\{z > z_{0.05} = 1.645\}$. The population variance σ^2 is not known, but it can be estimated very accurately (because *n*=36 is sufficiently large) by the sample variance $s^2 = 9$.

Thus, the observed value of the test statistic is approximately:

$$Z = \frac{\overline{y} - \mu}{s / \sqrt{n}} = \frac{17 - 15}{3 / \sqrt{36}} = \frac{2}{1/2} = 4$$

Because the observed value of *Z* lies in the rejection region (because *z*=4 exceeds $z_{0.05} = 1.645$), we reject $H_0 : \mu = 15$.

Thus, at the $\alpha = 0.05$ level of significance, the evidence is sufficient to indicate that the vice president's claim is incorrect and that the average number of sales contacts per week exceeds 15.

- Testing $H_0 : \theta = \theta_0$ against $H_a : \theta < \theta_0$ is done in an analogolis manner, except that we now reject H_0 for values of $\hat{\theta}$ that are much smaller than θ_0. The test statistic remains $Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$ but for a fixed level a we reject the H_0 when $z < -z_\alpha$.
  - Because we reject H_0 in favor of H_a when z falls far enough into the lower tail of the standard normal distribution, we call $H_a : \theta < \theta_0$ a *lower tail alternative* and $RR = \{z < -z_\alpha\}$ *lower tail rejection region.*
- In testing $H_0 : \theta = \theta_0$ against $H_a : \theta \neq \theta_0$, we reject H_0 if $\hat{\theta}$ is either much smaller or much larger than θ_0. The test statistic is still *Z*, as before, but the *rejection region is located symmetrically* in the two tails of the probability distribution for *Z*. Thus, we reject H_0 if either $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$. Equivalently, we reject H_0 if $|z| > z_{\alpha/2}$. This test is called a *two-tailed test*, as opposed to the one-tailed tests used for the alternatives $\theta < \theta_0$ and $\theta > \theta_0$.

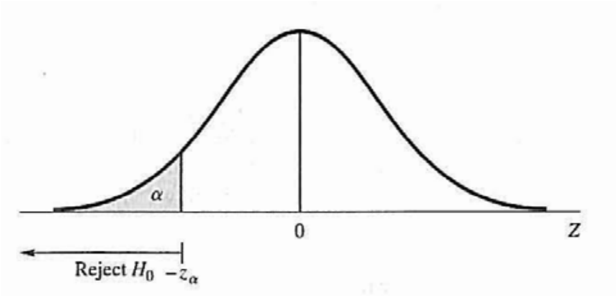

Figure 2 Large sample rejection for $H_0 : \theta = \theta_0$ versus $H_a : \theta < \theta_0$

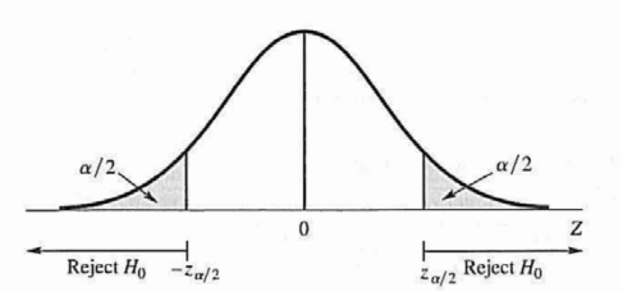

Figure 3 Large sample rejection for $H_0 : \theta = \theta_0$ versus $H_a : \theta \neq \theta_0$

Summary of Large Sample α-level Hypothesis Tests

$H_0 : \theta = \theta_0$

$$H_a : \begin{cases} \theta > \theta_0 & \text{(upper tail alternative)} \\ \theta < \theta_0 & \text{(lower tail alternative)} \\ \theta \neq \theta_0 & \text{(two-tailed alternative)} \end{cases}$$

Test statistic: $Z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$

Rejection Region: $RR : \begin{cases} \{z > z_\alpha\} & \text{(upper-tail RR)} \\ \{z < -z_\alpha\} & \text{(lower-tail RR)} \\ \{|z| > z_{\alpha/2}\} & \text{(two-tailed RR)} \end{cases}$.

How do we decide which alternative to use for a test? The answer depends on the hypothesis that we seek to support. If we are interested only in detecting an increase in the average sales contracts sell per salesperson, we should locale the rejection region in the upper tail of the standard normal distribution. On the other hand, if we wish to detect a change in μ either above or below μ =15, we should locate the rejection region in both tails of the standard normal distribution and employ a two-tailed test.

III. Calculating Type II Error Probabilities and Finding the Sample Size for the Z Test

- Calculating β can be very difficult for some statistical tests, but it is easy for the large-sample tests that we have seen in the previous section.

- For the test $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$, we can calculate type II error probabilities only for specific values for θ in H_0. Suppose that the experimenter *has in mind a specific a1temative*, say, $\theta = \theta_a$ (where $\theta_a > \theta_0$). Because the rejection region is of the form:

$$RR = \left\{ \hat{\theta} \geq k \right\}$$

the probability β of a *type II error* is

$\beta = P(\hat{\theta}$ is not in RR when H_a is true)

$$\beta = P(\hat{\theta} \le k \text{ when } \theta = \theta_a) = \text{P}\left(\frac{\hat{\theta} - \theta_a}{\sigma_{\hat{\theta}}} \le \frac{k - \theta_a}{\sigma_{\hat{\theta}}} \text{ when } \theta = \theta_a \right)$$

If θ_a is the true value of θ, then $\frac{\hat{\theta} - \theta_a}{\sigma_{\hat{\theta}}}$ has approximately a standard normal distribution. Consequently, the probability β can be determined (approximately) by finding a corresponding area under a standard normal curve.

- If θ_a is far from θ_0, the true value is relatively easy to detect, and β is considerably smaller.

- For a specified value of α, probability β of a *type II error* can both be made smaller by choosing a large sample size *n*.

Example 10.8 Suppose that the vice president in Example 10.5 wants to be able to detect a difference equal to one sale in the mean number of sale contracts per week. That is, he wishes to test $H_0 : \mu = 15$ against $H_0 : \mu = 16$. With the data as given in Example 10.5, find β for this test.

Note: The same question can be asked as follows: *suppose the true value of* μ *is 16, then find* β.

Solution

The rejection region for a 0.05 level test was given by

$$\frac{\bar{y}-\mu_0}{\sigma/\sqrt{n}} > 1.645$$

$$\bar{y}-\mu_0 > 1.645\frac{\sigma}{\sqrt{n}}$$

Substituting the values, μ_0=15, n=36 and s^2=9 (using s to approximate σ) yields

$$\bar{y}-15 > 1.645\frac{3}{\sqrt{36}}$$

$$\bar{y} > 0.8225+15$$

$$\Rightarrow \bar{y} > 15.8225$$

which is the rejection region for our case.

By definition

$\beta = P\left(\bar{Y} \leq 15.8225 \text{ when } \mu=\mu_a\right)$. Thus when $\mu_a = 16$, we have:

$$\beta = P\left(\frac{\overline{Y}-\mu_a}{\sigma/\sqrt{n}} \le \frac{15.8225-\mu_a}{\sigma/\sqrt{n}}\right)$$

$$\beta = P\left(Z \le \frac{15.8225-16}{3/\sqrt{36}}\right)$$

$$\beta = P\left(Z \le \frac{-0.1775}{1/2}\right)$$

$$\Rightarrow \ \beta = P(Z \le -0.355) \approx 0.3594$$

The <u>large value</u> of β tells us that samples of size *n*=36 <u>*frequently will fail lo detect a difference of one unit from the hypothesized means*</u>.

- *One can reduce the value of β by increasing the sample size n.*

Suppose that you want to test $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$. If you specify the desired values of α and β (where β is evaluated when $\mu = \mu_a$ when $\mu_a > \mu_0$), any further adjustment of the lest must involve two remaining quantities: the sample size, *n*, and the point at which the rejection region begins, *k*.

Thus,

- $\alpha = P\left(\overline{Y} > k \ \ when \ \ \mu = \mu_0\right)$

$$\alpha = P\left(\underbrace{\frac{\overline{Y}-\mu_0}{\sigma/\sqrt{n}}}_{Z} > \underbrace{\frac{k-\mu_0}{\sigma/\sqrt{n}}}_{z_\alpha} \ when \ \ \mu = \mu_0\right) = P\left(Z > z_\alpha\right)$$

- $\beta = P\left(\bar{Y} \le k \ \ when \ \ \mu = \mu_a\right)$

$$\beta = P\left(\underbrace{\frac{\bar{Y}-\mu_a}{\sigma/\sqrt{n}}}_{Z} \le \underbrace{\frac{k-\mu_a}{\sigma/\sqrt{n}}}_{-z_\beta} \ when \ \ \mu = \mu_a\right) = P\left(Z \le -z_\beta\right)$$

In the equations above, we have:

$$z_\alpha = \frac{k-\mu_0}{\sigma/\sqrt{n}} \Rightarrow k = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \quad \text{and}$$

$$-z_\beta = \frac{k-\mu_a}{\sigma/\sqrt{n}} \Rightarrow k = \mu_a - z_\beta \frac{\sigma}{\sqrt{n}}$$

Eliminating *k* from these two equations gives

$$\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} = \mu_a - z_\beta \frac{\sigma}{\sqrt{n}}$$

$$z_\alpha \frac{\sigma}{\sqrt{n}} + z_\beta \frac{\sigma}{\sqrt{n}} = \mu_a - \mu_0$$

$$\frac{\left(z_\alpha + z_\beta\right)\sigma}{\sqrt{n}} = \mu_a - \mu_0$$

$$\frac{\left(z_\alpha + z_\beta\right)\sigma}{\left(\mu_a - \mu_0\right)} = \sqrt{n}$$

$$\Rightarrow \quad n = \frac{\left(z_\alpha + z_\beta\right)^2 \sigma^2}{\left(\mu_a - \mu_0\right)^2}$$

which is the sample size for *upper-tail* α*-level* test.

The method just employed can be used to develop a similar formula for sample size for any one-tailed, hypothesis-testing problem that satisfies the conditions of large-sample test given before.

Example 10.9 Suppose that the vice president wanys to test $H_0 : \mu = 15$ against $H_a : \mu = 16$ with $\alpha = \beta = 0.05$. Find the sample size that will ensure this accuracy. Assume that σ^2 is approximately 9.

Solution

Because $\alpha = \beta = 0.05$, it implies that $z_\alpha = z_\beta = z_{0.05} = 1.645$
Then.

$$n = \frac{(1.645 + 1.645)^2 9}{(16 - 15)^2} = \frac{(3.29)^2 \times 9}{1} = 10.8241 \times 9 = 97.4169$$

Hence, n=98 observations should be used to meet the requirement that $\alpha \approx \beta \approx 0.05$ for the vice president's test.

IV. Another Way to Report the Results of a Statistical Test: *p*-value

Definition 10.2 if W is a test statistic, the p-value, or attained significance level, is the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected.

It may not be possible to find exact p-values from the usual Z tables. For example, if a test result is statistically significant for α=0.05 but *not* for α=0.25. Then, we will rep