Probability and Statistics
Chapter 9
Hypothesis Testing

I. Introduction to Hypothesis Testing

A statistician's most important job is to draw inferences about populations based on samples taken from the population.

2 methods for drawing inferences about parameters:
1) Make decisions concerning the value of a parameter.
2) Actually estimate the value of the parameter. (Like in Chapter 8)

In this chapter we will make decisions concerning the value of a parameter obtained by hypothesis testing.

Null Hypothesis   A working hypothesis about the population parameter in question. Denoted by H_0
The value specified in this hypothesis is often an historical value, a claim, or a production specification.

Alternate Hypothesis   Any hypothesis that differs from the null hypothesis. These are constructed in a way that it is the one to be accepted when the null hypothesis must be rejected. Denoted by H_1

Example 1
A company manufactures ball bearings for precision machines. The average diameter of a certain type of ball bearing should be 6.0 mm. To check that the average diameter is correct, the company formulates a statistical test.
   (a) What should be used for H_0 ? (Hint: What is the company trying to test?)

   (b) What should be used for H_1 ? (Hint: An error either way, too small or too large, would be serious)

Answers.
   (a) If μ is the mean diameter of the ball bearings, the company wants to test $\mu = 6.0$ mm. Therefore, $H_0 : \mu = 6.0$
   (b) An error either way could occur, and it would be serious. Therefore, $H_1 : \mu \neq 6.0$ (μ is either smaller than or larger than 6.0)

Example 2

A package delivery service claims it takes an average of 24 hours to send a package from New York to San Francisco. An independent consumer agency is doing a study to test the truth of this claim. Several complaints have led the agency to suspect that the deliver time is longer than 24 hours.

    (a) What should be used for the null hypothesis?

    (b) Assuming that the delivery service does not underrate itself, what should be used for the alternate hypothesis?

Answers.

    (a) The claim $\mu = 24$ hours is in question, so we take $H_0 : \mu = 24$

    (b) If the delivery service does not underrate itself, then the only reasonable alternate hypothesis is $H_1 : \mu > 24$

Type I error -   Rejecting the null hypothesis when it is in fact true.

Type II error -   Do not reject the null hypothesis when it is in fact false.

Level of significance - The probability with which we are willing to risk a type I error.
                          Denoted by the Greek letter α ("alpha")
                          A is specified in advance before any samples are drawn so that results will not influence the choice for the level of significance.

The Greek letter β ("beta") denotes the probability of making a type II error.

Methods of hypothesis testing require us to choose α and β values to be as small as possible.

Power of a Test
- The quantity $1 - \beta$
- Represents the probability of rejecting H_0 when it is in fact false.

For a given level of significance, how much power can we expect from a test?

1) The power of a statistical test increases as the level of significance α increases. A test performed at the $\alpha = 0.05$ level has more power than one at $\alpha = 0.01$. This means that the less stringent we make our significance level α, the more likely we will reject the null hypothesis when it is false.

2) Using a larger value of α will increase the power, but it will also increase the probability of a type I error. Despite this fact, most business executives, administrators, social scientists, and scientists use *small* α values. This choice reflects the conservative nature of administrators and scientists, who are usually more willing to make an error by failing to reject a claim than to make an error by accepting another claim that is false.

Meaning of the Terms *Fail* to Reject H_0 and Reject H_0

Term	Meaning
Fail to Reject H_0	There is not enough evidence in the data (and the test being used) to justify a rejection of H_0 . this means that we retain H_0 with the understanding that we have not proved it to be true beyond all doubt.
Reject H_0	There is enough evidence in the data (and the test employed) to justify rejection of H_0. This means that we choose the alternative hypothesis H_1 with the understanding that we have not proved H_1 to be true beyond all doubt.

Example 3

Let's reconsider Example 1 in which we were considering the manufacturing specifications for ball bearings. The hypothesis were

H_0 : $\mu = 6.0$ mm   (manufacturer's specification)
H_1 : $\mu \neq 6.0$ mm (cause for adjusting process)

(a) Suppose the manufacturer requires a 1% level of significance. Describe a type I error, its consequence and probability

(b) Discuss a type II error and its consequence.

Answers.

(a) A type I error is caused when sample evidence indicates that we should reject H_0 when in fact the average diameter of the ball bearings being produced is 6.0 mm. A type I error will cause a needless adjustment and delay of the manufacturing process. The probability of such an error is 1% because $\alpha = 0.01$.

(b) A type II error occurs if the sample evidence tells us not to reject the null hypothesis H_0: $\mu = 6.0$ mm when in fact the average diameter of the ball bearing is either too large or too small to meet specifications. Such an error would mean that the production process would not be adjusted when it really needed to be adjusted. This could possibly result in a large production of ball bearings that do not meet specifications.

Chapter 9
Hypothesis Testing

II.   Tests Involving the Mean μ (Large Samples)

General procedure for hypothesis testing:
1) Establish the null hypothesis H_0 .
2) Establish the alternate hypothesis H_1.
3) Use the level of significance α and the alternate hypothesis to determine the critical region
4) Find the critical values that form the boundaries of the critical region.
5) Use the sample evidence to draw a conclusion regarding whether to reject the null hypothesis H_0 or not.

(Fig 9-1)   The Null and Alternate Hypothesis

Null Hypothesis	Alternate Hypothesis and Type of Test		
Claim about μ or historical value of μ $H_0 : \mu = k$	You believe μ is less than value stated in H_0 $H_1:\ \mu < k$ Left-tailed test	You believe μ is more than value stated in H_0 $H_1: \mu > k$ Right-tailed test	You believe μ is different from value stated in H_0 $H_1: \mu \neq k$ Two-tailed test

$$z = \frac{\overline{x} - \mu}{\sigma_{\overline{x}}} = \frac{\overline{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$$

where μ = mean specified in H_0

   σ = standard deviation of the population

   n = sample size

   $\overline{x}$ = sample mean

Example 4

A research meteorologist has been studying wind patterns over the Pacific Ocean. Based on these studies, a new route is proposed for commercial airlines going from San Francisco to Honolulu. The new route is intended to take advantage of existing wind patterns to reduce flying time. It is known that for the old route the distribution of flying times for a large four-engine jet has mean $\mu = 5.25$ hours with standard deviation $\sigma = 0.6$ hours. 36 flights on the new route have yielded a mean flying time of $\bar{x} = 4.90$ hours. Does this indicate that the average flying time for the new route is less than 5.25 hours? Use a 5% level of significance.

(a) What is H_0 ?

(b) What is H_1?

(c) What type of critical region must be used?
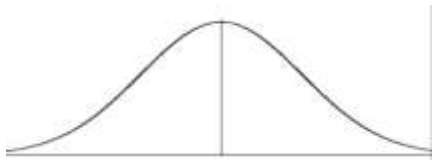
(d) What is the critical value?

(e) What is the z value corresponding to the sample mean $\bar{x} = 4.90$?

(f) Do we reject or not reject H_0 ?

Probability and Statistics
Chapter 9
Hypothesis Testing

Answers.

(a) The average time for the old route is $\mu = 5.25$ hours, so we will set up the null hypothesis

$H_0 : \mu = 5.25$

In words, we are saying that the mean flying time on the new route is the same as that on the old route.

(b) We want to see if the average flying time for the new route is less than 5.25 hours, so the alternate hypothesis is $H_1 : \mu < 5.25$

(c) Since the $<$ symbol is used in the alternate hypothesis , Figure 9-1 tells us to use a left-tailed test.

(d) Using the invnorm function on your graphing calculator(or table IV in your book) find the critical value for a left-tailed test with $\alpha = 0.05$. $z = -1.64$ (or -1.65)

Reject
Region   NonReject
          Region
z = -1.64 (the critical value)
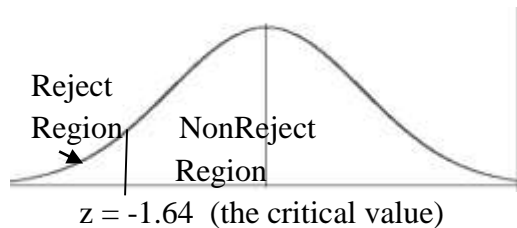
(e) We use the formula $z = \dfrac{\overline{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{4.90 - 5.25}{\dfrac{0.6}{\sqrt{36}}} = -3.50$ to find the test statistic.

(f) Compare the test statistic with the critical value. Since -3.50 < -1.64, it falls in the reject region. Therefore, we reject H_0, and choose H_1 at the $\alpha = 0.05$ level. This means that there is enough evidence to suggest that the flying time on the new route is less than the flying time on the old route.

Probability and Statistics
Chapter 9
Hypothesis Testing

Example 5

     A machine makes twist-off caps for bottles. The machine is adjusted to make caps of diameter 1.85 cm and with a standard deviation of $\sigma = 0.05$ cm. During production, an inspector checks the diameters of caps to see if the machine has slipped out of adjustment. A random sample of 64 caps is taken. If the mean diameter for this sample is $\overline{x} = 1.87$ cm, does this indicate that the machine has slipped out of adjustment and the average diameter of caps is no longer $\mu = 1.85$ cm? Use a 1% level of significance.

(a) What is the null hypothesis?

     $H_0 : \mu = 0.05$
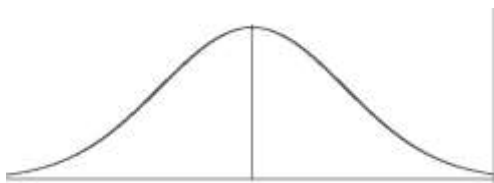 $H_0 : \mu = 1.85$
     $H_0 : \mu = 1.87$

(b) An error either way would be serious. Therefore, we want to test the null hypothesis against the hypothesis that the mean diameter is *not* 1.85 cm. Therefore, the alternate hypothesis is

     $H_1 : \mu < 1.85$
     $H_1 : \mu \neq 1.85$
     $H_1 : \mu > 1.85$

(c) Should we use a one- or two-tailed test? What are the critical values?

(d) What is the value of the sample mean? Convert it to a z-value (a test statistic)

(e) Show the critical value(s), rejection region(s) and the z value (test statistic) on the normal curve. Does the test statistic fall in the critical region or not? Do we reject or fail to reject H_0? At the 1% level of significance, can we say that the machine needs adjustment?

Answers.

(a) $H_0:$ $\mu = 1.85$ Be sure to use the hypothesized parameter, *not* the sample mean, $\overline{x}$, in the null hypothesis.

(b) $H_1:$ $\mu \neq 1.85$

(c) Because the alternate hypothesis uses the $\neq$ symbol, we use a two-tailed test. Since the sample size n = 64 is large, the central limit theorem tells us that the sampling distribution of $\overline{x}$ is approximately normal. Therefore, we can use a z-score for our critical values. For $\alpha = 0.01$, $z = \pm 2.58$

Reject
Region

NonReject
Region

Reject
Region

z = -2.58

z = 2.58
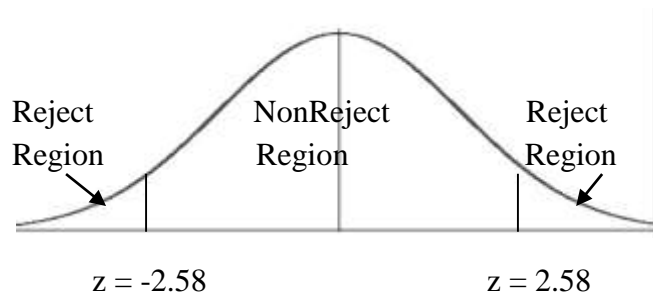
(d) n = 64, $\mu = 1.85$ cm, $\overline{x} = 1.87$ cm so $z = \dfrac{1.87 - 1.85}{\dfrac{0.05}{\sqrt{64}}} = \mathbf{3.20}$

(e) Compare the test statistic with the critical value. Since $3.20 > 2.58$, it falls in the reject region. Therefore, we reject H_0, and choose H_1 at the $\alpha = 0.01$ level of significance. This means that there is enough evidence to say that the machine needs adjustment.

III.   The *P* Value in Hypothesis Testing

The level of significance α in a test of hypothesis is the probability of making a type I error. (The probability of rejecting the null hypothesis when it is really true.)

For the distribution described by the null hypothesis, the *P value* is the smallest level of significance for which the observed test statistic tells us to reject H_0. In other words, it is the smallest level of significance at which the given null hypothesis is rejected.

Consequently if,
> *P value* $\leq \alpha$, then we reject H_0
> *P value* $> \alpha$, then we do not reject H_0

P values are the *areas* in the tail or tails of a probability distribution beyond the observed test statistic.

"If P is low, H_0 has to go"

Example 6

      Last year the average age of students attending Fremont College was 21.3 years. This semester, in order to meet the needs of alder students, more classes were scheduled during evening and weekend hours. Has the average age of students at Fremont College increased this semester? To answer this question, a random sample of 64 students enrolled this semester was studied. The average age of these students was 22.1 years with population standard deviation σ = 2.7 years. Use the *P value* of the test statistic to test the hypothesis that the average age of students this semester is higher than it was last year. Use $\alpha = 0.01$.

(a) What are the null and alternate hypothesis?

(b) Convert $\bar{x}$ to z. Remember to round z to 2 decimals.

(c) Sketch a diagram showing the *P* value (the area) of the test statistic.

(d) Compute the *P value*

(e) Compare the *P value* to $\alpha = 0.01$. Do we reject H_0 or not? Has the average age at the college increased?
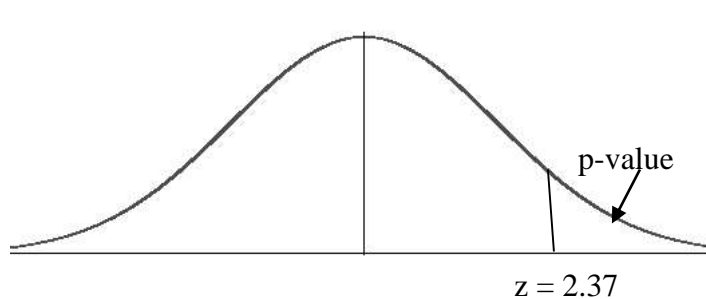
Probability and Statistics
Chapter 9
Hypothesis Testing

Answers.

(a) $H_0 : \mu = 21.3$ v. $H_1 : \mu > 21.3$

(b) Using $\mu = 21.3$ from H_0, n = 64, and s = 2.7, we get $z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{22.1 - 21.3}{\dfrac{2.7}{\sqrt{64}}} = \mathbf{2.370}$

(c)

p-value

z = 2.37

(d) To find the *P value*, use the normalcdf function on your graphing calculator. (Or use table IV in your book). The lower bound is 2.37 and the upper bound is ∞. df = n - 1. *P value* = 0.0089

(e) The *P* value = 0.0089 is the smallest level of significance for which we reject H_0.
   Since $0.0089 < 0.01$ (p is less than α), we reject H_0. At the 1% significance level, there is enough evidence to suggest that the average age of the students enrolled this semester has increased.

Probability and Statistics

Chapter 9

Hypothesis Testing

Decision Procedure for Testing

(1) If *P value* $\leq \alpha$, then we reject H_0 and we say that the data are *statistically significantly different* or *statistically different.*

(2) If *P value* $> \alpha$, then we do not reject H_0 and we say that the data are *not statistically significantly different* or *not statistically different*

See page 397 in your book.

Example 7

Compare the given *P value* and the level of significance α to conclude each test.

(a) Given: H_0: $\mu = 30$ v. H_1: $\mu \neq 30$

$\qquad \alpha = 0.01$

$\qquad P \ value = 0.0213$

Do we reject or not reject H_0?

(b) Given: H_0: $\mu = 0.15$ v. H_1: $\mu < 0.15$

$\qquad \alpha = 0.05$

$\qquad P \ value = 0.0171$

Do we reject or not reject H_0?

Probability and Statistics
Chapter 9
Hypothesis Testing

Answers.

(a) Since $\alpha = 0.01$ and *P value* = 0.0213, we see that *P value* > α, so we do not reject H_0.

(b) Since $\alpha = 0.05$ and *P value* = 0.0171, we see that *P value* < α, so we reject H_0.

Summary

The *P value* (probability of chance) is the area of the sampling distribution that lies beyond the observed test statistic. It tells us the probability that a sample statistic will be more extreme than the observed value. We compute the *P values* as follows:

1)   When H_1 indicates a right-tailed test,
                P value = area to the right of the observed test statistic

2)   When H_1 indicates a left-tailed test,
                P value = area to the left of the observed test statistic

3)   When H_1 indicates a two-tailed test,
                P value = sum of the areas in the two tails

To conclude the test, we compare the level of significance α with the *P value*.
        1) If the *P value* is less than or equal to α, reject H_0.

        2) If the *P value* is greater than α, do not reject H_0.

IV.   Tests Involving the Mean, μ, population standard deviation unknown

$$t = \frac{\overline{x} - \mu}{s_{\overline{x}}} = \frac{\overline{x} - \mu}{\dfrac{s}{\sqrt{n}}}$$

Where $\overline{x}$ = sample mean

   n = sample size

   s = sample standard deviation

Example 8

   Use table V in your book or the invT function on your calculator to find the critical value(s) t and the critical region(s) for the test described based on a sample of size n = 8. Include an illustration.
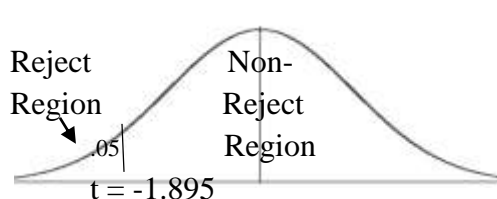
   (a) A *left-tailed* test of μ with a level of significance α = 0.05

   (b) A *two-tailed* test of μ with a significance level of α = 0.05.

Answers.

   (a) In this case, the area in the rejection region is 0.05 and $df = n – 1 = 7$.
       Therefore, t = -1.895
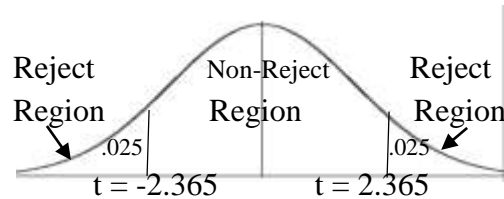

Reject
Region

.05

t = -1.895

Non-
Reject
Region

(b) For a two tailed test, the area in each rejection region is $.05/2 = .025$ and $df = n - 1 = 7$, so $t = \pm 2.365$

Reject Region
Non-Reject Region
Reject Region
.025
.025
t = -2.365
t = 2.365

Example 9a

Suppose that a random sample of eight engines was used and the mean fuel consumption was $\bar{x} = 5880$ pounds (for the first 15 seconds). The standard deviation for the sample was 250 lbs. Use a 1% level of significance to test the claim that the population average fuel consumption *exceeds* 5550 pounds. Use the critical method.
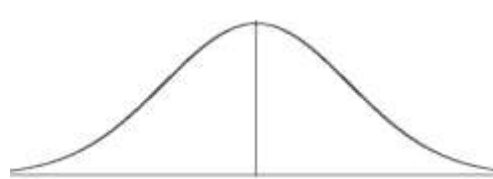
(a) We will use $H_0 : \mu = 5500$. What should we use for H_1 ?

$H_1 : \mu < 5500$

$H_1 : \mu \neq 5500$

$H_1 : \mu > 5500$

(b) Find the critical value t.

(c) Convert the sample mean to a t value.

(d) Compare the test statistic and the critical value. Do we reject H_0 or not at the 1% level of significance?
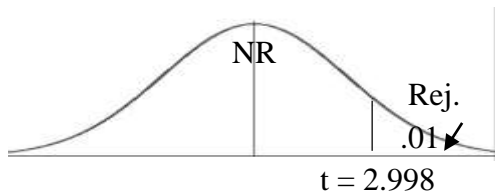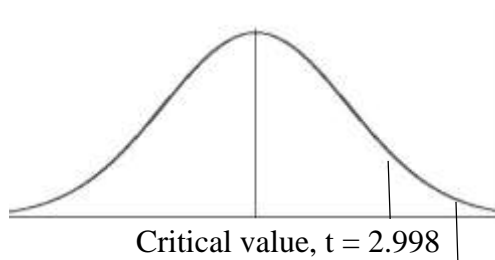
Answers.

(a) $H_1 :\ \mu > 5500$ Because we want to test the claim that the average fuel consumption *exceeds* 5500 pounds.

(b) Because the sample standard deviation is known, we need a t value. Use table V in your book or invT on your graphing calculator and $df = 7$. T = 2.998

NR

Rej.
.01

t = 2.998

(c) We use $\mu = 5500$, n = 8, and s = 250. $\quad t = \dfrac{5880 - 5500}{\dfrac{250}{\sqrt{8}}} = 4.299$

Critical value, t = 2.998

4.299 = t observed value or test statistic

(d) $4.299 > 2.998$. Since the test statistic falls in the critical region, we reject H_0. There is enough evidence to show that the average fuel consumption exceeds 5500pounds during the first 15 seconds of operation.
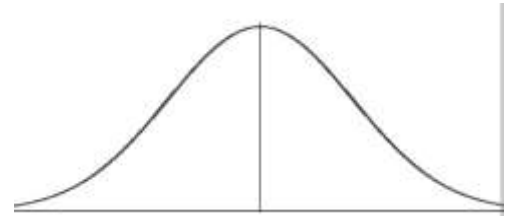
Probability and Statistics
Chapter 9
Hypothesis Testing

Example 9b - Using the p-value method

Suppose that a random sample of eight engines was used and the mean fuel consumption was $\bar{x} = 5880$ pounds (for the first 15 seconds). The standard deviation for the sample was 250 lbs. Use a 1% level of significance to test the claim that the population average fuel consumption *exceeds* 550 pounds. Use the p-value method

a) What should the null hypothesis be?

b) What should the alternate hypothesis be?

c) Draw a normal curve. Will this be a right-tailed, left-tailed, or two-tailed test?

d) Calculate the test statistic, t_o, and put it on your diagram.

e) Calculate the *p-value*. Compare to the level of significance. Write your conclusion.
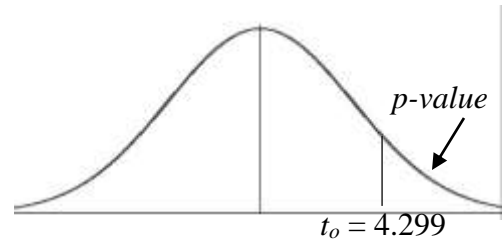
Probability and Statistics
Chapter 9
Hypothesis Testing

Answers.

a) $H_0 : \mu = 5500$

b) $H_1 : \mu > 5500$

c) Right-tailed test?

p-value

$t_o = 4.299$

d) $t = \dfrac{5880 - 5500}{\dfrac{250}{\sqrt{8}}} = 4.299$

e) tdcf: lower bound = 4.299, upper bound = infinity , $df = 8 \;\rightarrow\; p\text{-}value = .0018$
compare p to α: .0018 < .01. since .0018 < .01, we will reject the null hypothesis. There is enough evidence to say that the mean fuel consumption exceeds 5500 pounds.

Probability and Statistics
Chapter 9
Hypothesis Testing

V.   Tests Involving a Proportion

This section studies tests involving proportions. These tests are essentially the same as those in sections II and IV. The difference is that in this section, we are working with a distribution of proportions instead of means.

$$z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$$

where $\hat{p}$ is the sample proportion (sample statistic),
     n = sample size, and p is the population proportion (as specified in H_0)

Example 10

A botanist has produced a new variety of hybrid wheat thtat is better ble to withstand drought than other varieties. The botanist knows that for the parent plants the proportion of seed germinating is 80%. To test this claim, 400 seeds from the hybrid plant are tested and it is found that 312 germinated. Use a 5% level of significance to test the claim that the proportion germinating for the hybrid is 80%.

(a) Let p be the proportion of hybrid seeds that will germinate. What is the null hypothesis?

(b) Because we have no prior knowledge about germination proportion for the hybrid plant, what would be a good choice for the alternate hypothesis?

(c) Should our test be a one-tailed or two0tailed test? For $\alpha = 0.05$, what is (are) the critical value(s)?
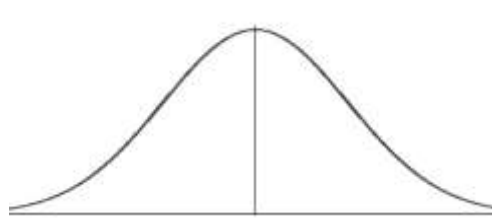
(d) Calculate the test statistic for $\hat{p}$. Based on our choice for H_0, what value should we use for p in our formula?
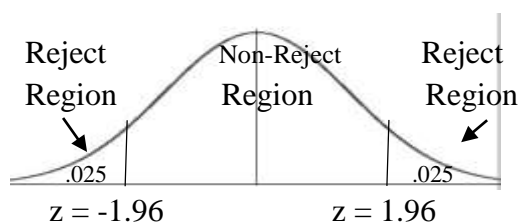
(e) Do we reject H_0 or not at the 5% level of significance?

Answers:

(a) H_0 : $p = 0.80$

(b) H_1 : $p \neq 0.80$

(c) Since the alternate hypothesis is $\neq$, this is a two-tailed test. Each critical region has area $= \alpha/2$ or 0.025. Using invnorm (.025) gives critical values of ± 1.96

Reject Region   Non-Reject Region   Reject Region

.025   .025
z = -1.96   z = 1.96

(d) $\hat{p} = \dfrac{312}{400} = 0.78$ so our test statistic is $z = \dfrac{.78 - .80}{\sqrt{\dfrac{(.80)(.20)}{400}}} = -1.00$

(e) Since $-1.96 < -1.00 < 1.96$, we do not reject the null hypothesis. There is not sufficient evidence to suggest that the population proportion is not 80%.