Economics 30330: Statistics for Economics
Problem Set 8 - Suggested Solutions
University of Notre Dame
Instructor: Julio Garín
Spring 2012

Hypothesis Testing (80 Points)

1. Consider the following hypothesis test:

$$H_0 : \mu \geq 20$$

$$H_A : \mu < 20$$

A sample of 40 provided a sample mean of 19.4. The population standard deviation is 2.

(a) Create a 95% confidence interval for the mean.
We know σ, therefore we should use the $z-table$. This is a one-tailed (lower tail) test, so the 95% confidence interval will be given then by

$$\left[\bar{x} - z_{.05}\frac{\sigma}{\sqrt{n}}, \infty\right)$$

$$\left[19.4 - 1.65\frac{2}{\sqrt{40}}, \infty\right)$$

The 95% confidence interval is $\mu \in [18.878, \infty)$.

(b) What is the p-value?
The p-value is the area in the lower tail. First, we calculate the z-value:

$$z = \frac{19.4 - 20}{2/\sqrt{40}} = -1.9$$

Using the normal table with z = -1.9, p-value =.0287.

(c) At $\alpha = 0.01$, what is your conclusion?
p-value > .01, so we fail reject H_0 at the 99% level.

(d) What is the rejection rule using the critical value? What is your conclusion?
Reject H_0 at the 99% level if $z \leq z^c_\alpha$=-2.33. In this example, -1.9 > -2.33, so we fail to reject H_0 at the 99% level.

2. Consider the following hypothesis test:

$$H_0 : \mu = 15$$

$$H_A : \mu \neq 15$$

A sample of 50 provided a sample mean of 14.5. The population standard deviation is 3.

(a) Create a 95% confidence interval for the mean. *We know σ, therefore we should use the $z-$table. This is a two-tailed test, so the 95% confidence interval will be given then by*

$$\left(\bar{x} - z_{.025}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{.025}\frac{\sigma}{\sqrt{n}}\right)$$

$$\left(14.5 - 1.96\frac{3}{\sqrt{50}}, 14.5 + 1.96\frac{3}{\sqrt{50}}\right)$$

The 95% confidence interval is $\mu \in (13.67, 15.33)$.

(b) What is the p-value?
The p-value is the area in the upper and lower tails. First, we calculate the z-value:

$$z = \frac{14.5 - 15}{3/\sqrt{50}} = -1.18$$

Using the normal table with z = -1.18, p-value $= 2(.1190) = 0.238$.

(c) At $\alpha = 0.01$, what is your conclusion?
p-value > .01, so we fail reject H_0 at the 99% level.

(d) What is the rejection rule using the critical value? What is your conclusion?
Reject H_0 at the 99% level if $|z| \geq 2.58$. *In this example,* $1.18 < 2.58$, *so we fail to reject H_0 at the 99% level.*

3. Wall Street securities firms paid out record year-end bonuses of $125,000 per employee for 2005 (*Fortune*, February 6, 2006). Suppose we would like to take a sample of employees at the Garín & Munnich securities firm to see whether the mean year-end bonus is different from the reported mean of $125,000 for the population.

(a) State the null and alternative hypotheses you would use to test whether the year-end bonuses paid by Garín & Munnich were different from the population mean.

$$H_0 : \mu = 125000$$

$$H_A : \mu \neq 125000$$

(b) Suppose a sample of 40 Garín & Munnich employees showed a sample mean year-end bonus of $118,000. Assume a population standard deviation of $30,000 and compute the p-value.

$$z = \frac{118000 - 125000}{30000/\sqrt{40}}$$

Because we have a two tail test, the p-value is two times the lower tail area. Using the normal table with z = -1.48: p-value = 2(.0694) = 0.1388.

(c) With $\alpha = 0.05$ as the level of significance, what is your conclusion?
p-value $> .05$, *do not reject H_0. We cannot conclude that the year-end bonuses paid by Garín & Munnich differ significantly from the population mean of $125,000.*

(d) Repeat the preceding hypothesis test using the critical value.
Reject H_0 *if* $z \leq -1.96$ *or* $z \geq 1.96$. z = *-1.48; cannot reject* H_0.

4. During the 2004 election year, new polling results were reported daily. In an IBD/TIPP poll of 910 adults, 503 respondents reported that they were optimistic about the national outlook, and President Bush's leadership index jumped 4.7 points to 55.3 (*Investor's Business Daily* January 14, 2004).

(a) What is the sample proportion of respondents who are optimistic about the national outlook?

$$\bar{p} = 503/910 = 0.5527$$

(b) A campaign manager wants to claim that this poll indicates that the majority of adults are optimistic about the national outlook. Construct a hypothesis test so that the rejection of the null hypothesis will permit the conclusion that the proportion optimistic is greater than 50%.

$$H_0 : p \leq .50$$
$$H_A : p > .50$$

(c) Compute the p-value and explain to the manager what it means about the level of significance of the results.

$$z = \frac{0.5527 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{910}}}$$

The p-value is the area in the upper tail. Using normal table with z = *3.18: p-value* ≈ 0. *You can tell the manager that the observed level of significance is very close to zero and that this means the results are highly statistically significant. Any reasonable person would reject the null hypotheses and conclude that the proportion of adults who are optimistic about the national outlook is greater than .50.*

5. In a survey of 1,200 high school seniors in 1992, 27% answered yes to the question: Have you smoked at least one cigarette in the past 30 days? In a 1997 survey of 1100 students, 35% answered yes to the same question.

(a) Construct a 95% confidence interval for the change in the fraction of high school seniors who smoke.

$$n_1 = 1,200; \quad \bar{p}_1 = 0.27$$
$$n_2 = 1,100; \quad \bar{p}_2 = 0.35$$

$$H_0 : d = p_1 - p_2 = 0$$
$$H_A : d = p_1 - p_2 \neq 0$$

First we need to find $\bar{p}$

$$\begin{aligned}\bar{p} &= \frac{\bar{p}_1 n_1 + \bar{p}_2 n_2}{n_1 + n_2} \\ &= \frac{0.27(1,200) + 0.35(1,100)}{1,200 + 1,100} \\ &= 0.308\end{aligned}$$

Now we have all the information needed to construct the CI

$$\begin{aligned} d &= \Delta \pm z_{\alpha/2}\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \\ &= 0.35 - 0.27 \pm z_{.025}\sqrt{0.308(1-0.308)\left(\frac{1}{1,200}+\frac{1}{1,100}\right)} \\ &= 0.08 \pm 1.96(0.01927) \\ &= 0.08 \pm 0.0377 \end{aligned}$$

The 95% confidence interval is $(0.0423, 0.1177)$.

(b) Using a z-test, test null hypothesis that the fraction of high school seniors who smoked in the past 30 days has not changed over the 1992-1997 period.

$$\begin{aligned} z &= \frac{\Delta - d}{se(\Delta)} \\ &= \frac{\bar{p}_2 - \bar{p}_1 - 0}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \\ &= \frac{0.08}{0.01927} \\ &= 4.15 \end{aligned}$$

If $|z| > z_{0.025}$*, we reject the null at the 95% level.* $|4.15| > 1.96$*, so we reject the null hypothesis that the fraction of seniors who smoke did not change between 1992 and 1997.*

6. Listed below are sample characteristics from a 1987 survey that examines average hourly wage rates for union and non-union workers.

- Nonunion: $\bar{x}_n = 11.47$; $n_n = 1206$; $s_n = 6.58$
- Union: $\bar{x}_u = 12.19$; $n_u = 376$; $s_u = 4.77$:

(a) What is the average difference in hourly wages between union and nonunion workers?

$$\begin{aligned} \Delta &= \bar{x}_u - \bar{x}_n \\ &= 12.19 - 11.47 \\ &= 0.72 \end{aligned}$$

(b) Construct a 95% confidence interval around this difference.

$$d = \Delta \pm t_{\alpha/2,n-1} s_p \sqrt{\frac{1}{n_u}+\frac{1}{n_n}}$$

$$s_p^2 = \frac{s_u(n_u - 1) + s_n(n_n - 1)}{n_u + n_n - 2}$$
$$= \frac{6.58^2(1205) + 4.77^2(375)}{1206 + 376 - 2}$$
$$= 38.42$$
$$s_p = 6.198$$
$$d = 0.72 \pm 1.96(6.198)\sqrt{\frac{1}{1206} + \frac{1}{376}}$$
$$= 0.72 \pm 1.96(0.366)$$

The 95% confidence interval is $(0.0025, 1.4375)$.

(c) Test the null hypothesis that there is no difference in wages across the two groups.

$$H_0 : d = \mu_n - \mu_u = 0$$

$$H_A : d = \mu_n - \mu_u \neq 0$$

Since 0 is not in the 95% confidence interval, we reject the null hypothesis (barely) that there is no difference in wages across the two groups at the 95% level.

7. In a survey of 700 undergraduates (350 females and 350 males), 48% of males reported an episode of binge drinking in the past year (five or more drinks in a row on one occasion), whereas only 40% of females reported binge drinking.

(a) What is the point estimate of the difference between the two population proportions?

$$\Delta = \bar{p}_m - \bar{p}_f$$
$$= 0.48 - 0.40$$
$$= 0.08$$

(b) Construct a 95% confidence interval on the difference in binge drinking rates between males and females.

First we need to find $\bar{p}$

$$\bar{p} = \frac{\bar{p}_m n_m + \bar{p}_f n_f}{n_m + n_f}$$
$$= \frac{0.48(350) + 0.40(350)}{700}$$
$$= 0.44$$

We can proceed to create the CI:

$$d = \Delta \pm z_{\alpha/2}\sqrt{\bar{p}(1 - \bar{p})\left(\frac{1}{n_m} + \frac{1}{n_f}\right)}$$
$$= 0.08 \pm 1.96\sqrt{0.44(1 - 0.44)\left(\frac{1}{350} + \frac{1}{350}\right)}$$
$$= 0.08 \pm 1.96(0.0375)$$

The 95% confidence interval is $(0.0065, 0.1535)$.

(c) Can you reject the null hypothesis that at the 95% confidence level males and females have the same binge drinking rates?
Since 0 is not in the 95% confidence interval, we reject the null hypothesis that males and females have the same binge drinking rates at the 95% level.

(d) How does your answer change if you increase the confidence interval to 99%?

$$\begin{aligned} d &= \Delta \pm z_{0.01/2}\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_m}+\frac{1}{n_f}\right)} \\ &= 0.08 \pm 2.57\sqrt{0.44(1-0.44)\left(\frac{1}{350}+\frac{1}{350}\right)} \\ &= 0.08 \pm 2.57(0.0375) \end{aligned}$$

The 99% confidence interval is $(-0.0163, 0.1763)$. *Now 0 is in the confidence interval, so we fail to reject the null hypothesis that males and females have the same binge drinking rates at the 99% level.*

(e) What are the types of errors you can commit in this particular example? Explain.
Type I error: reject the null when it is true (false positive). In this example, that means that we concluded that males and females do not have the same binge drinking rates when in fact they do.
Type II error: fail to reject the null when it is false (false negative). In this example, that means that we concluded that males and females have the same binge drinking rates when they do not.

(f) What is the p-value for this example?
The p-value is the area in the upper and lower tails. First, we calculate the z-value:

$$z = \frac{\Delta - d}{se(\Delta)} = \frac{0.08}{0.0375} = 2.139$$

Using the normal table with $z = 2.14$, *p-value* $= 2(1 - .9838) = 0.0324$.

8. Economists Joe Price and Justin Wolfers studied discrimination among NBA referees in a recent paper and argued that more personal fouls were awarded against players when they are officiated by an opposite-race officiating crew than when officiated by an own-race refereeing crew.[1] Specifically, using a sample of 266,984 observations, they find that a player earns 0.197 fewer fouls per 48 minutes played when facing three referees of his own race than when facing three opposite race referees. In other words, $\hat{\beta}_1 = .197$, with a standard error of 0.061, as shown in row 1, column 1 of Table 4 (see next page). [2]

   (a) What is $\hat{\beta}_1$ measuring?
   $\hat{\beta}_1$ measures the number additional fouls per 48 minutes a player receives when facing three opposite race referees than when facing three referees of his own race (or, alternatively, the number fewer fouls per 48 minutes when facing three referees of his own race than when facing three opposite race referees).

   (b) What is the null hypothesis that the authors are testing? (Write out the equations and explain in your own words).

$$H_0 : \beta_1 = 0$$
$$H_A : \beta_1 \neq 0$$

   The authors are testing whether players earn more fouls when facing referees of the opposite race than when they face referees of the same race. In other words, they are testing whether $\hat{\beta}_1$ is statistically different from zero.

   (c) Using the estimate of $\hat{\beta}_1$ and its standard error, create a 95% confidence interval for $\hat{\beta}_1$. (Note that the number of observations is listed in the footnote of the table).
   We do not know σ, and we are not working with proportions, therefore we should use the $t-$ table. The 95% confidence interval will be given then by

$$\left(\hat{\beta}_1 - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \hat{\beta}_1 + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

   We are given values of $\hat{\beta}_1$ and the standard error, so the 95% confidence interval is

$$(0.197 - 1.96(0.061), 0.197 + 1.96(0.061))$$

   Therefore, $\mu \in (0.0774, 0.3166)$ with 95% probability.

   (d) What is the p-value?
   To calculate the p-value, we first need to calculate the t-test (because we are working with a sample, and $\hat{\beta}_1$ does not represent a proportion):

$$t = \frac{0.197}{0.061} = 3.23$$

   Using the t-table with $t = 3.23$, for a two-tailed test, p-value $= 2(0.001) = 0.002$.

   (e) At $\alpha = 0.01$, what is your conclusion about discrimination among NBA referees?
   $p < 0.01$ therefore $\hat{\beta}_1$ is statistically different from zero at the 1% level. Based on this estimate, there appears to be discrimination among NBA referees.

[1] The paper is available at http://bpp.wharton.upenn.edu/jwolfers/Papers/NBARace.pdf.
[2] For now, just ignore the stars in the table.