Logan Albiani

ECN 102

11/1/18

Prof. Siegler

Problem Set #2

1. *One Sample Univariate Hypothesis Testing of a Mean*

*Consider a random sample of 5 adults over the age of 25 from a large population, which is normally distributed, where E represents the total years of education completed:*

*𝐸 = (10, 12, 12, 16, 16)*

*Suppose that someone claims that the average person in the population is a college graduate (𝜇 = 16).*

1. ***What is the null hypothesis? What is the alternative hypothesis?***

Null Hypothesis: H0: 𝜇 = 16

Alternative Hypothesis: HA: 𝜇 ≠ 16

1. ***Can you reject the null hypothesis at the 10 percent level of significance? Can you reject the null hypothesis at the 5 percent level of significance? Use the critical value approach. You can use R for critical values, but you must show all of your calculations and explain. Use R, however, to check your work.***

Calculate the sample estimate (): = ; = 13.2

Calculate standard error: = 2.68

Calculate test statistic: = 2.33

T4,.1/2 = 2.132; 2.33 > 2.132, therefore we can reject the null hypothesis and accept the alternative with a significance level of 10%

T4,.05/2 = 2.78; 2.33 < 2.78, therefore we cannot reject the null and thus, fail to accept the alterative hypothesis at the 5% significance level.

1. *What is the 95% confidence interval for years of education? Provide a written interpretation explaining your answer.*

Confidence Interval: 13.2 ± (2.78 \* 1.2); CI = [9.86, 16.53]

What this information displays is that to reject the null hypothesis that the true mean of the population is equal to sixteen, we would need to have a test statistic that fell out of this range. That is, if the test statistic fell below 9.86 or above 16.53, we would have evidence to reject the null hypothesis and accept the alternative hypothesis.

1. *One Sample Hypothesis Testing with Proportions*

***For this question, show the results by hand, but you can use R to check your work. Suppose that the 4-year graduation rate at a large, public university is 70 percent. In an effort to increase graduation rates, the university randomly selects 200 incoming freshman to participate in a peer advising program. After 4 years, 154 of these students graduated. What are the null and alternative hypotheses? Can you conclude that this program was a success at the 5 percent level of significance? Can you conclude that the program is a success at the 1 percent level of significance? Show your work and explain. Since success is an increase in graduation rates this is a one tailed test.***

Null Hypothesis: H0: .7

Alternative Hypothesis: HA: > .7

Calculate z-score: z = = .032404 which has a corresponding z score of roughly 2.16. z = 2.16

Find the z-score of alpha: The 5 percent level of significance has a corresponding z score of 1.645. z = 1.645

Compare z – scores: Because the 2.16 > 1.645, we have evidence to reject the null and accept the alternative hypothesis.

Find z-score of alpha: The 1 percent level of confidence has a corresponding z score of 2.33. z = 2.33

Compare z-scores: Because 2.16 < 2.33, we have no evidence to reject the null, and therefore, fail to accept the alternative hypothesis.

1. *Test Driving Speeds with R*

***Surveyed students in a class who have driven an automobile were asked their highest speed. The result is attached in driving.csv. Consider this sample of students to representative of all UCD students. Suppose someone claims that the average UCD student has driven faster than 90mph.***

1. ***What are the null and alternative hypotheses?***

Null Hypothesis: H0: 𝜇 = 90

Alternative Hypothesis: HA: 𝜇 ≠ 90

1. ***What can you conclude at the 1 percent level of significance? Use R to conduct this test and intuitively explain how you arrived at your conclusion.***

Running a t – test in R results in a sample mean of 97.9 alongside a t statistic of 3.28. On top of that, the resulting p-value is .0007308. The corresponding t-value of 91 degrees of freedom and a significance level of 1% is about 2.46 due to the fact that a t-statistic with a large number of observations approaches the standard normal curve. From this, we can compare the test statistic and the t-value given, in which case we find 3.28 > 2.46. This is evidence to reject the null and support the alternative hypothesis. Furthermore, we can simply observe from the extremely low p-value that the given outcome randomly occurring has an extremely small probability.

1. *Consider a random sample of 10 people, 5 of whom graudated with a Man Econ degree and the other 5 with a standard Econ degree. Several years later, their hourly earnings are reported.*

|  |  |
| --- | --- |
| *page2image3720144****Economics*** | ***Managerial Economics*** |
| *$45* | *$20* |
| *$65* | *$55* |
| *$35* | *$40* |
| *$45* | *$40* |
| *$85* | *$20* |

*Assume that the population distribution are normally distributed. At the 10 percent level of significance, can you conclude that there is a difference in hourly earnings between economics majors and managerial economics majors? What are the null and alternative hypotheses? What test statistic should you use for this test and why? Be sure to show all of your calculations and to interpret your findings. That is, while you can use R to check your work, you must show all calculations below by hand.*

Calculate sample mean of Economics: (45+65+35+$5+85)/5 = 55

Calculate sample mean of Man Econ: (20+55+40+40+20)/5 = 35

Calculate degrees of freedom: Man Econ (5 – 1 = 4) Econ (5 – 1 = 4)

Calculate Variance of Econ: (100+400+400+100+900)/5 = 380

Calculate Variance of Man Econ: (225+400+25+25+225)/5 = 180

Calculate pooled variance: (5-1)(380)+(5-1)(180)/(4+4) = 280

Calculate t: t = [(55 – 35) – 0]/() = 20/11.18 = 1.7889

Compare t-statistic to the level of confidence given. In this case, we can find that the corresponding t-statistic, t8,.1/2 = 1.39. Due to this, we can observe that 1.39 < 1.79. Given this, we have evidence to reject the null hypothesis and accept the alternative hypothesis that the two means are not equal.

Null Hypothesis: H0: 𝜇A = 𝜇B

Alternative Hypothesis: HA: 𝜇A ≠ 𝜇B

1. *Correlation and Simple Regression by Hand*

|  |  |
| --- | --- |
| ***Y (Hourly Earnings in Dollars)*** | *page3image3730960****X (Years of Education)*** |
| *$35* | *16* |
| *$15* | *14* |
| *$40* | *18* |
| *$30* | *12* |
| *$55* | *20* |

*Be sure to show all of your calculations and to interpret your findings. That is, while you can use R to check your work, you must show all of the calculations below by hand. Assume that the populations are normal.*

1. ***Compute the sample correlation coefficient, r, between hourly earnings and years of education for this random sample of 5 students in a class.***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x | ( | (2 | y | ( | (2 |
| 16 | 0 | 0 | 35 | 0 | 0 |
| 14 | -2 | 4 | 15 | -20 | 400 |
| 18 | 2 | 4 | 40 | 5 | 25 |
| 12 | -4 | 16 | 30 | -5 | 25 |
| 20 | 4 | 16 | 55 | 20 | 400 |
|  |  | Sx2=40 |  |  | Sy2=850 |

Calculate the correlation coefficient.

R = (0\*0)+(-2\*-20)+(2\*5)+(-4\*-5)+(4\*20)/(\*) = 150/184.391; r = .8135

1. ***Test the null hypothesis is Ho: rho = 0 versus the alternative hypothesis HA: rho ≠ 0. Can you reject the null hypothesis at the 5-percent level of significance? Can you reject the null at the 1 percent level of significance? Show your work and explain.***

Ho: rho = 0

HA: rho ≠ 0

Calculate standard error of the correlation coefficient:

se(r) = /(5-2) = = .469668

Calculate the test statistic: (r-p0)/se(r) = (.813-0)/.470 = 1.79

The corresponding t-value for the significance level (alpha): t3,.05=2.35

2.35> 1.73, therefore we’ve no evidence to reject the null hypothesis and accept the alternative hypothesis.

The corresponding t-value for the significance level (alpha): t3,.1=4.54

4.54>1.73, therefore, we’ve no evidence to reject the null and accept the alternative hypothesis

1. ***Use cor.test to check results.***

data: earnings and years.of.education

t = 2.4227, df = 3, p-value = 0.09394

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.2436439 0.9872164 sample estimates:

cor = 0.8134892

1. ***Use OLS to estimate the following regression model by hand:***

***Ŷ = b1+b2Xi***

b2 = 150/40 = 3.75

b1=35-3.5(16): b1=-25

Ŷ = -25 + 3.75(xi)

1. ***Use R to replicate and report the stargazer results:***

===============================================

Dependent variable:

---------------------------

earnings

-----------------------------------------------

years.of.education 3.750\*

(1.548)

Constant -25.000

(25.150)

-----------------------------------------------

Observations 5

R2 0.662

Adjusted R2 0.549

Residual Std. Error 9.789 (df = 3)

F Statistic 5.870\* (df = 1; 3)

===============================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

1. ***Is the slope coefficient statistically different from zero at the 10 percent level of significance? is it at the 5 percent level? Show your work and explain***

Ho; b2 = 0

HA; b2≠ 0

Calculate RSS: RSS = (35-31)2+(15-24)2+(40+38)2+(30+17)2+(55+45)2 = 370

Calculate se(b2): = 1.756

Calculate the test statistic: 3.75/1.756: t = 2.136

Find the corresponding t-value for the 5 and 10 percent levels of significance:

t3,.1= 1.648. 2.136 > 1.648, therefore, the result is statistically significant and we have evidence to reject the null hypothesis and accept the alternative.

t3,.05= 2.353. 2.353 > 2.136, therefore, we fail to reject the null hypothesis

1. ***Precisely interpret the exact meaning of the estimated slope coefficient in this particular case.***

For every singular increase in the number of years of education, we can expect to see an increase in hourly wage of $3.75.

1. ***What is the predicted hourly wage of someone with 13 years of education based on the regression results. Show your work.***

Ŷ = -25 + 3.75(13) = $23.75

1. *Regression toward the mean:*

*The term “regression” was popularized by Sir Francis Galton with the publication of his paper, “Regression towards Mediocrity in Hereditary Stature” (The Journal of the Anthropological Institute of Great Britain and Ireland 15 (1886), 246‐263). In this paper, he showed that extreme characteristics of parents do not fully manifest themselves in their children. In fact, these characteristics regress toward “mediocrity”, which is regression towards the population mean. That is, very tall parents tend to have children that are not as tall as they are, and very short parents tend to have children taller than they are.*

*The question uses Galton’s original data set and it is attached as galton.csv.*

1. ***Use R to estimate the following model using OLS and report the results with stargazer.***

===============================================

Dependent variable:

---------------------------

father

-----------------------------------------------

son 0.326\*\*\*

(0.021)

Constant 46.135\*\*\*

(1.412)

-----------------------------------------------

Observations 928

R2 0.210

Adjusted R2 0.210

Residual Std. Error 1.589 (df = 926)

F Statistic 246.839\*\*\* (df = 1; 926)

===============================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Ŷ = 46.13535 + .32565 (xi)

1. ***Use OSL to test whether or not the correlation coefficient regresses to the mean:***

Null Hypothesis: H0: 2 1

Alternative Hypothesis: HA: 2 < 1

T = (.32565 – 1)/se(b2)

R Code:

##Problem 1##

edu<-c(10,12,12,16,16)

t.test(edu, mu=16, conf.level=.9)

t.test(edu, mu=16, conf.level=.95)

##Problem 2##

prop.test(154,200,.7,alternative="greater", correct=FALSE)

prop.test(154,200,.7,alternative="greater", correct=FALSE, conf.level=.99)

##Problem 3##

setwd("~/Desktop/POL 51 R Folder")

speed<-read.csv("driving.csv")

attach(speed)

t.test(speed, mu=90, alternative="greater", conf.level=.99)

##Problem 4##

ecn<-c(45,65,35,45,85)

are<-c(20,55,40,40,20)

t.test(ecn,are,var.equal = TRUE, mu=0)

##Problem 5##

years.of.education<-c(16,14,18,12,20)

earnings<-c(35,15,40,30,55)

cor.test(earnings,years.of.education)

qt(.95,3)

qt(.99,3)

install.packages("stargazer")

library(stargazer)

regmod1<-lm(earnings~years.of.education)

summary(regmod1)

stargazer(regmod1, type="text")

confint(regmod1, level = .9)

confint(regmod1, level = .95)

qt(.9,3)

qt(.95,3)

##Problem 6##

heights<-read.csv("galton.csv")

regmod2<-lm(father~son, data=heights)

summary(regmod2)

stargazer(regmod2, type="text")