HW-3

MS -Business Intelligence & Analytics
Fall 2015
BIA – 654 A

Feb 9, 2015

Mohit Ravi Ghatikar
CWID - 10405877

Ethics Statement

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

Signature _Mohit Ravi Ghatikar_______________ Date: 02/09/2015___

Assignment – 3

1a)

n = 2000, p = 0.56 (56%)

Assumptions: np>5 and n(1-p) > 5

: 2000*0.56 = 1120 >5 and 2000*0.44 = 880 > 5

Therefore, normal approximation for the binomial distribution is satisfied.

Confidence interval = p +/- $Z_{a/2}$ * $(p(1-p)/n)^{0.5}$

= 0.56 +/- 0.021

= 0.539 and 0.581 (Confidence interval lies between 53.9% and 58.1%)

1b)

The interval from 53.9% to 58.1% may or may not contain the true proportion, 95% of intervals formed in this manner from sample size of 2000 will contain the true proportion.

1c)

Margin of error (e) = $Z_{a/2}$ * $(p(1-p)/n)^{0.5}$

= 0.021 for 95% confidence interval.

1d)

Sample size, n = ($Z_{a/2}^2$ * Pi * (1-Pi)) / e^2

If we have no prior knowledge of pi, we assume pi = 0.5

= (1.96^2 * 0.5 * 0.5) / 0.02^2

= 2400

For a 99% confidence interval,

n = ($Z_{a/2}^2$ * Pi * (1-Pi)) / e^2

= (2.58^2 * 0.5 * 0.5) / 0.02^2

= 4160

2a)

Null Hypothesis: “flex-time” had no effect on absenteeism with mean >= 6.3 days off work.

Alternate Hypothesis: Absenteeism went down due to “flex-time” with mean < 6.3 days off work.

Level of significance = 95% with n = 100.

Since sample size is large enough, t-statistic and Z-statistic will be almost equal.

So we use a Z statistic, where X(bar) = 5.5 and s = 2.9

Zstat = (X(bar) – mu) / (s / root(n))

= -2.75

P(z < -2.75) = 0.00298

p-value = 0.00298

Since the p-value of 0.00298 < 0.05, we reject the null hypothesis with 95% confidence. The absenteeism went down due to flex time with a mean less than 6.3 days off work.

2b)

Null Hypothesis: "flex-time" had no effect on absenteeism with mean >= 6.3 days off work.

Alternate Hypothesis: Absenteeism went down due to "flex-time" with mean < 6.3 days off work.

Level of significance = 95% with n = 100.

Since sample size is large enough we t-statistic and Z-statistic will be almost equal.

So we use a Z statistic, where X(bar) = 5.9 and s = 2.9

Zstat = (X(bar) – mu) / (s / root(n))

= -1.37

P(z < -1.37) = 0.085

p-value = 0.085 = 0.085

Since the p-value of 0.085>0.05, we fail to reject the null hypothesis with 95% confidence. We can say that flex time had no effect on absenteeism with mean >= 6.3.

3)

Null Hypothesis: The temperature difference is not greater than 0. (u <= 0)

Alternate Hypothesis: The temperature difference is greater than 0. (u > 0)

Level of significance = 99% with n=20. We have to use the t-test since the sample size is less. Before that we need to examine if the given sample is coming from a normal distribution.

X(bar) = 1.12, s = 0.623.

Shapiro-wilk test:

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.926525	Pr < W	0.1324
Kolmogorov-Smirnov	D	0.196036	Pr > D	0.0429
Cramer-von Mises	W-Sq	0.085871	Pr > W-Sq	0.1675
Anderson-Darling	A-Sq	0.521319	Pr > A-Sq	0.1689

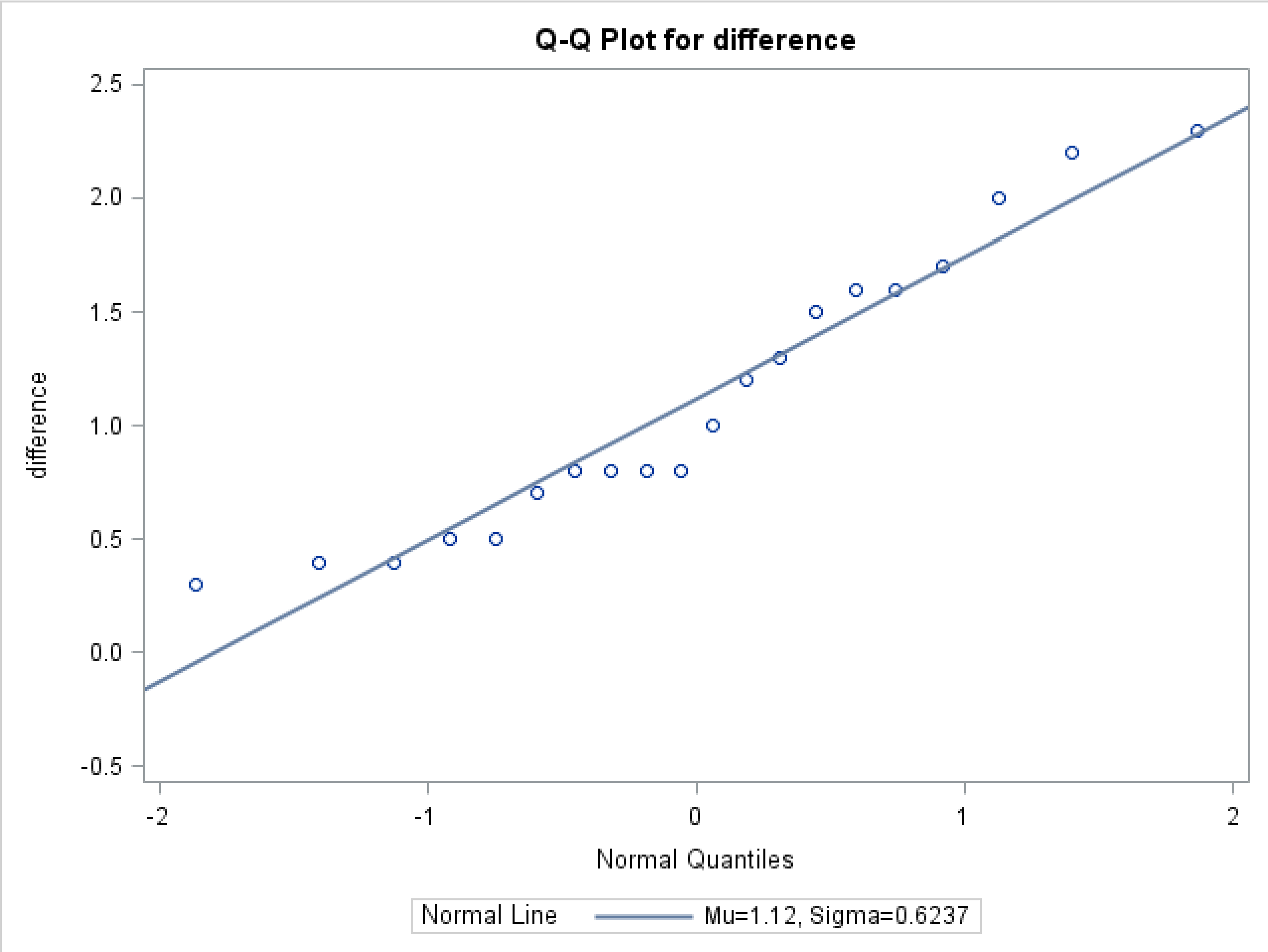
The underlying assumptions in the shapiro-wilk test are:

The null hypothesis for this test is that the data are normally distributed. The Prob < W value listed in the output is the p-value. If the chosen alpha level is 0.01 and the p-value is less than 0.01, then the null hypothesis that the data are normally distributed is rejected. If the p-value is greater than 0.01, then the null hypothesis is not rejected.

Since p-value of 0.132 is greater than 0.01, we cannot reject the null hypothesis for an alpha of 0.01.

Therefore the sample came from a population that is normally distributed.

This can also be seen from the Q-Q plot. Most of the points are close to straight line, which suggests that the sample is close to a normal distribution.



Tstat = (X(bar) – mu) / (s / root(n))

= (1.12 – 0) / (0.623 / 4.47)

= 8.03

p-value at 19 degrees of freedom = 0.00001

Since the p-value is less than 0.01, we can reject the null hypothesis at 99% confidence. We have sufficient evidence to prove that the temperature difference is greater than 0.

4)

False

To use a T-test with n=5, we need to use 4 degrees of freedom.