HW-2

MS -Business Intelligence & Analytics

Fall 2015

BIA – 654 A

January 31, 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: 01/31/2015___

Assignment – 2

1a)

The sample size (N) = 750

S = 0.80 and $\bar{x}$ = 1.86.

For a 95% confidence interval, $Z_{a/2}$ = 1.96.

Since the sample size is large enough, Confidence Interval is: $\bar{x} + Z_{a/2} * (\text{sigma} / (N)^{0.5})$

We can substitute the Standard deviation of the population with sample standard deviation.

Therefore, C.I = $1.86 \pm 1.96 * (0.80 / (750)^{0.5})$

= 1.86 ± 0.572

Average number of television sets lies between 1.288 and 1.917 with a 95% confidence interval.

1b)

Since sample size is less than 50, we need to know if the sample is coming from a population which is normally distributed. The assumption of normality in the population isn't mentioned in the question. Therefore we cannot calculate the confidence interval.

2a)

Here N=21

$\bar{x}$ = 5.92 and s = 1.413

Since the sample size is less, we need to confirm if this sample is coming from a population that is normally distributed. To check this, we run the shapiro-wilk test and inspect the Q-Q plot.

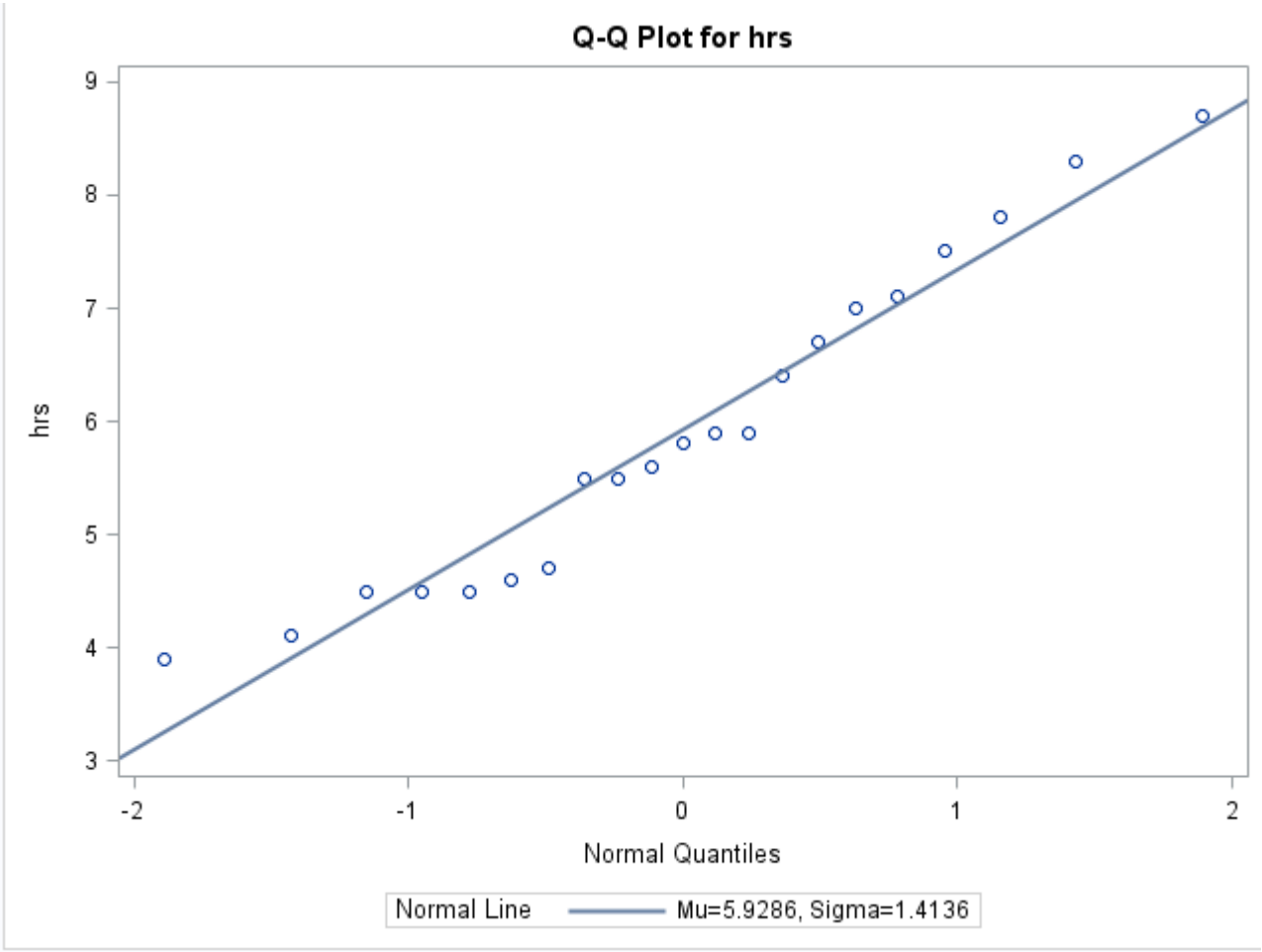
Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.948231	Pr < W	0.3153
Kolmogorov-Smirnov	D	0.140947	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.056311	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.377208	Pr > A-Sq	>0.2500

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 0.05 and the p-value is less than 0.05, then the null hypothesis that the data are normally distributed is rejected. If the p-value is greater than 0.05, then the null hypothesis is not rejected.

Since p-value is greater than 0.05, we cannot reject the null hypothesis for an alpha of 0.05. 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.



The t value at 95% confidence interval at 20 degrees of freedom = 1.725

C.I = 5.92 +- 1.725 * (1.413 / $21^{0.5}$)

= 5.92 +- 0.531

Therefore, the mean cleaning time is between 5.389 and 6.451 with a 95% confidence level. The original estimated time of 5.7 hours is a reasonable estimate.

2b)

Since the number of hours are normally distributed, we can use t-distribution.

Sample size = 8.

The t value at 95% confidence interval at 7 degrees of freedom = 1.895

C.I = 5.92 +- 1.895 * (1.413 / $8^{0.5}$)

= 5.92 +- 0.946

Therefore, the mean cleaning time is between 4.974 and 6.866 with a 95% confidence level.