Question 01

H_0: The number of mortgages approved per week follows a Poisson distribution

H_1: The number of mortgages approved per week does not follow a Poisson distribution

Mean $= \frac{0 \text{ x } 13+ 1 \text{ x } 25+2 \text{ x } 32+3 \text{ x } 17+4 \text{ x } 9+5 \text{ x } 6+6 \text{ x } 1+7 \text{ x } 1}{104}$

No. Of Commercials	O_i	Probability	E_i
0	13	0.12246	12.74
1	25	0.25715	26.74
2	32	0.27002	28.08
3	17	0.18901	19.66
4	09	0.09923	10.32
5	06	0.04168	4.33
6	01	0.01459	1.52
7	01	0.00586	0.61

Since the Expected frequency(E_i) column contains values less than 5, rows should be merged.

No. Of Commercials	O_i	E_i
0	13	12.74
1	25	26.74
2	32	28.08
3	17	19.66
4	09	10.32
≥ 5	08	6.46

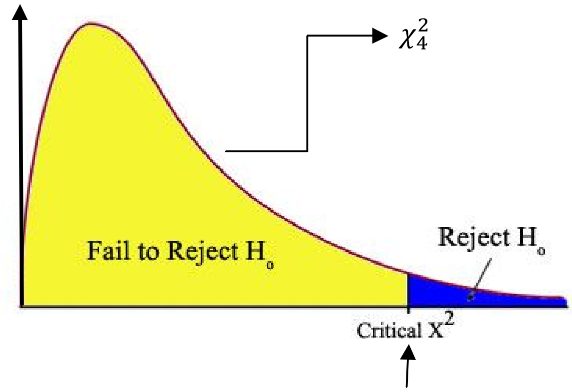
Test Statistics $= \sum \frac{(O_i - E_i)^2}{E_i}$

$$= \frac{(13-12.74)^2}{12.74} + \frac{(25-26.74)^2}{26.74} + \cdots + \frac{(8-6.46)^2}{6.46} = 1.56$$

$\chi^2_{tabulated}$ Value = $\chi^2_{4,0.05}$ = 9.48733.

i.e $\chi^2_{calculated}$ Value < $\chi^2_{tabulated}$ Value

Since p- value > 0.05 we do not reject the null hypothesis with 95% level of confidence. We have enough evidence to say that mortgages approved per week follows a Poisson distribution with 95% level of confidence.



Question 02

4	4	5	5	5	5	5	6	6	7
8	9	9	9	10	11	11	12	12	12
13	14	14	15	15	15	16	16	16	17
17	17	18	19	21	21	21	22	23	25
25	25	26	27	27	29	29	33	33	34

a) Mode = Most Common Value = 5.

b) Position of Median = $\frac{(n+1)}{2}$ th term

= 25.5

Median = (15 + 15)/2 = 15

c) Mean = $\frac{\sum_{i=1}^{n} x_i}{n}$

= 15.96

d) Variance = $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{(n-1)}$

Standard Deviation = $\sqrt{Variance}$

= 8.49

(e). Let us first examine whether the data follows a normal distribution using the normality test and the probability plot(p-p plot).Hypothesis for the normality test is as follows.

H_0: Data follow a normal distribution vs H_1: Otherwise

Since our sample size is 50 we can use the Kolmogorov Smirnov test for normality. If the sample is less than 50 we have to use the Shapiro Wilk test for the normality. Since the p value is greater than 0.05, the null hypothesis can be accepted. Therefore we can conclude that data follows a normal distribution. Since data follows a normal distribution it can be said that data follows the empirical formula.

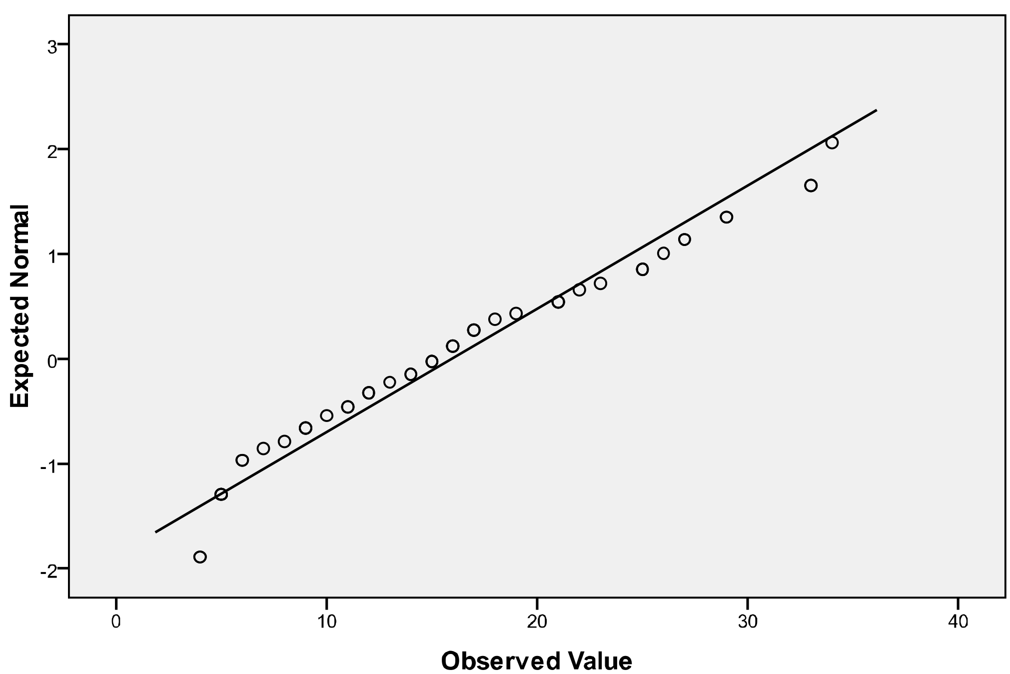
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
VAR00001	.091	50	.200*	.949	50	.030

a. Lilliefors Significance Correction

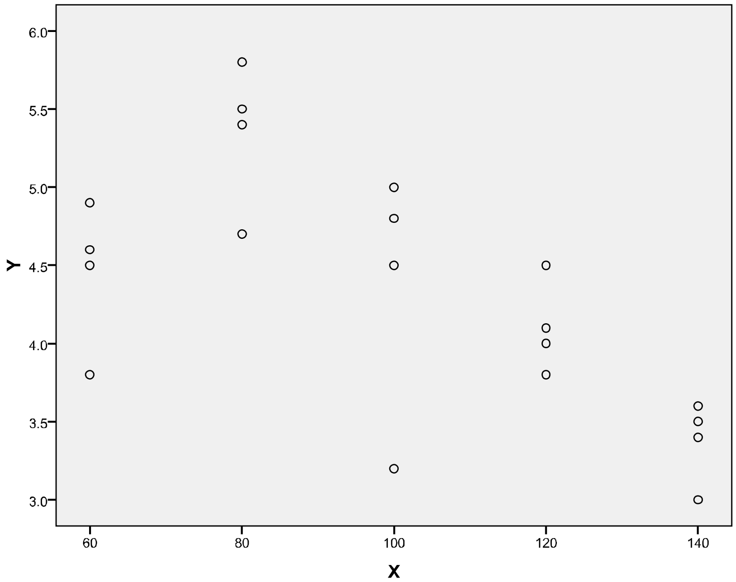
*. This is a lower bound of the true significance.

Normal Q-Q Plot of VAR00001



Question 03

3) a)



- Does there appear to be a relation?

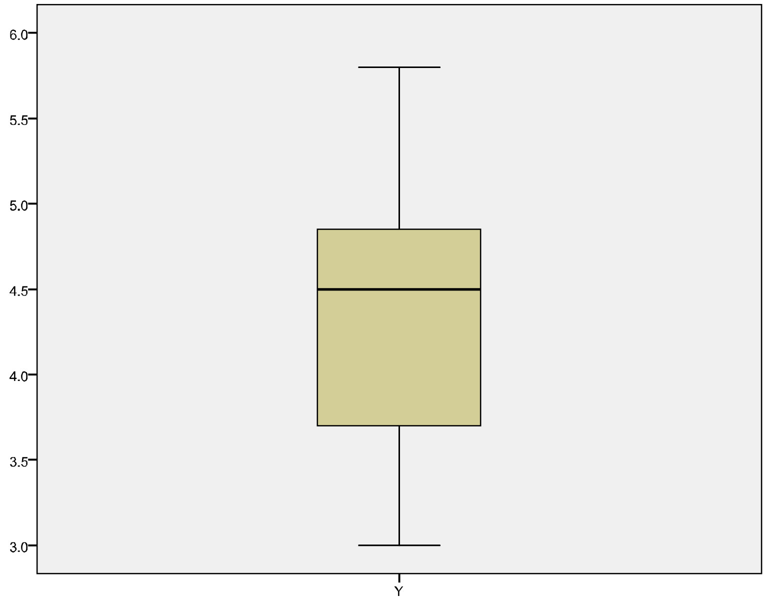
  It seems to have a relationship between the variables according to the scatter plot.

- Does it appear to be linear?

  The relationship seems to be linear according to the scatter plot.

Sample correlation coefficient is -0.625. There seems to have a correlation between the variables which is negative.

b)



The box plot in the above diagram shows no outliers. Therefore we can say that the data set is free of outliers.

C) Linear regression model is Y= a + b X + ∈

Assumption :

Random errors are normally distributed with zero mean and constant variance(σ^2).

(d). Using the calculator

A = 6.030

B = -0.17

Therefore the model is Y= 6.03 – 0.17 X

(i.e. If you increase the rotation speed by 1 then the life time of the drill will be decreased by 0.17.)

e)$H_{0:}\ \rho = 0 \quad Vs \quad H_{1:}\ \rho \neq 0$

<u>Test Statistics</u>

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$t = -0.625\frac{\sqrt{18}}{\sqrt{1-0.391}}$$

$$= -3.399$$

Tabulated value = $t_{18,0.05}$ = -1.734

Since $t_{calculated} < t_{table}$ we reject Ho with 90% confidence interval level of confidence. Therefore we have enough confidence to say that two variables are linearly correlated with 10% level of significance.

(e). Confidence Interval for Regression Slope parameter

The confidence interval for the parameter can be computed from the estimate b using the computed standard deviations and the appropriate critical value of the t_{n-2}distribution.
The confidence interval is given by $(b \ + \ t \times std.dev)$.

$$\text{Confidence Interval} = (-0.17 - 1.3678, -0.17 + 1.3678) = (-1.5378, 1.1978)$$

Question 4

$$P(X > 315) = P(X \geq 315.5)$$

$$= P(\frac{X - 300}{\sqrt{120}} \geq \frac{315.5 - 300}{\sqrt{120}})$$

$$= P(Z \geq 1.41) = 0.07927$$ **or** $$P(Z \geq 1.42) = 0.07780$$

$$P(X < 290) = P(X \leq 289.5)$$

$$= P(\frac{X - 300}{\sqrt{120}} \leq \frac{289.5 - 300}{\sqrt{120}}) = P(Z \leq -0.96) = 0.16853$$

Probability of not getting what they ask for = 0.16853 + 0.07780 = 0.244633