|  |  |
| --- | --- |
| MS5105 Exercise Number: | MS5105 Handup Exercise1 |
| Student Number: | 19230487 |
| Student Name: | Jayakarthi Boovendran |
| Total Number of Pages: | 19 |

**Exercise Number: 4.2**

**Dataset Used**: [DataSet1] G:\MS5105 - Statistical Techniques for Business Analytics\Assignments\1\Datasets\Ch 04 - Exercise 02.sav

**Part A**

Descriptive statistics and a histogram with a normal curve for donated for the whole data set

* Frequencies

FREQUENCIES VARIABLES=donated

/FORMAT=NOTABLE

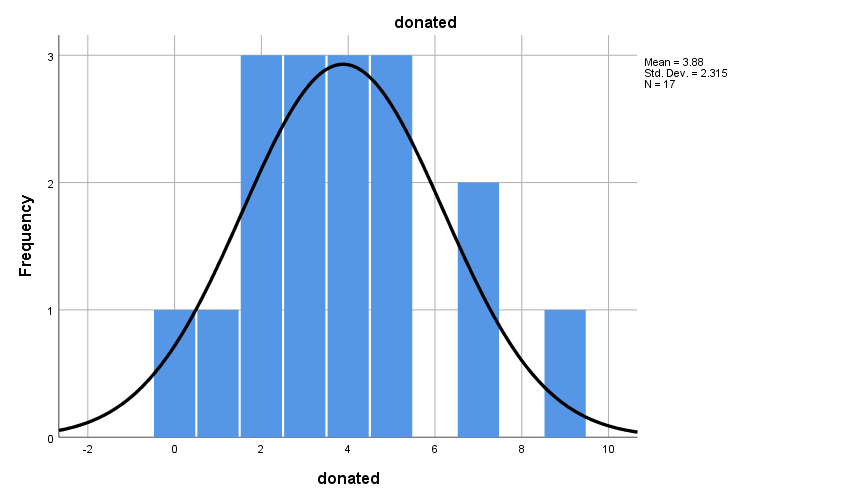
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM

/HISTOGRAM NORMAL

/ORDER=ANALYSIS

|  |  |  |
| --- | --- | --- |
| **Statistics** | | |
| donated | | |
| N | Valid | 17 |
| Missing | 0 |
| Mean | | 3.88 |
| Median | | 4.00 |
| Mode | | 2a |
| Std. Deviation | | 2.315 |
| Variance | | 5.360 |
| Range | | 9 |
| Minimum | | 0 |
| Maximum | | 9 |
| a. Multiple modes exist. The smallest value is shown | | |

* Histogram with a normal curve for donated for whole data set



* **Tests of Normality**

EXAMINE VARIABLES=donated BY gender

/PLOT BOXPLOT HISTOGRAM NPPLOT

/COMPARE GROUPS

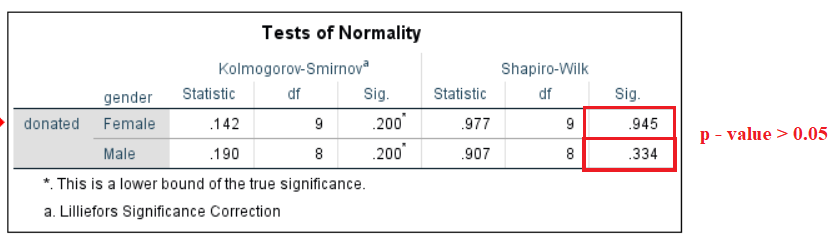
/STATISTICS DESCRIPTIVES

/CINTERVAL 95

/MISSING LISTWISE

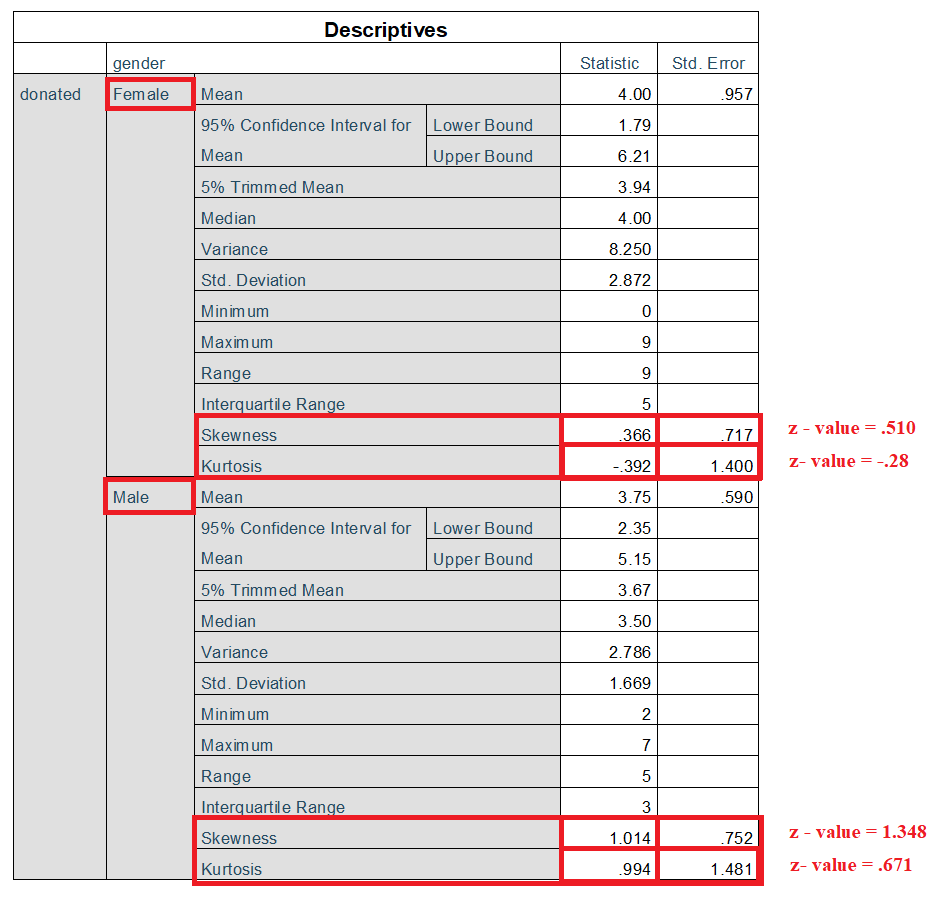
/NOTOTAL

**The Shapiro-Wilk Test**

****

The p-value for both male and female donors are above 0.05, thus according to Shapiro-Wilk test of normality is assumed that the data is **approximately normally distributed** (Shapiro & Wilk, 1965; Razali & Wah, 2011).

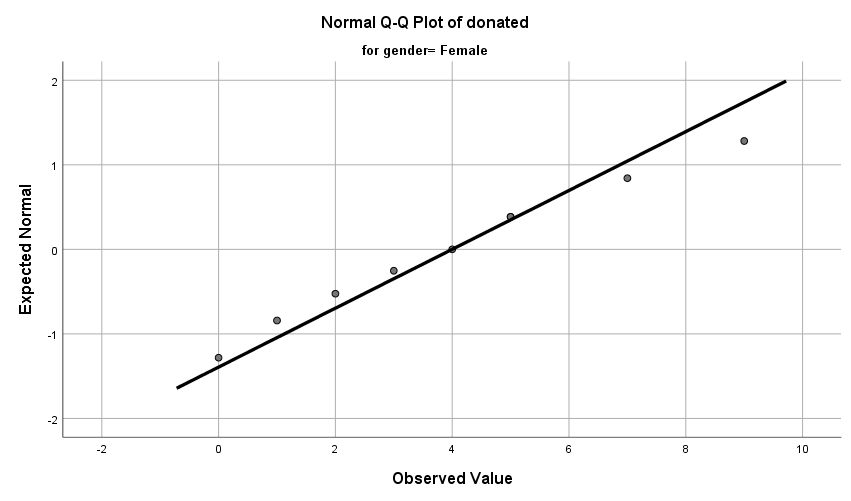
**Skewness and Kurtosis z-value Test**

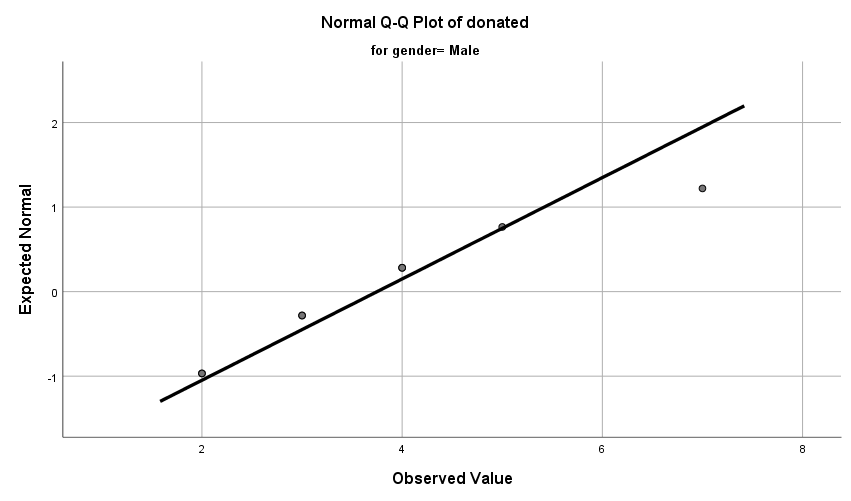
****

In all the cases, the z-values lie within the span of **-1.96 to +1.96**. Hence donation is **approximately normally distributed** in terms of skewness and kurtosis (Doane & Seward, 2011; Kim, 2013)

**Visual Interpretation of Histogram and Normal Q-Q Plot**

Through the visual interpretation of the histogram, it is evident that the data is normally distributed and the normal curve is fairly symmetrical, thus the assumption of normality is met. Furthermore, visual analysis of Normal Q-Q plot also proves that donation is **approximately normally distributed**.





**Documenting Normality Test Results**

The Shapiro-Wilk’s test (p>.05) (Shapiro & Wilk, 1965; Razali & Wah, 2011) and a visual inspection of their histograms and Normal Q-Q plots showed that the donations were approximately normally distributed for both males and females, with a skewness of 0.366 for males and a skewness of 1.014 for females (Doane & Seward, 2011; Kim, 2013)

**Discussion**

* Normality: The distribution of donation is reasonably normal for both male and female donors.

**Part B**

Descriptive statistics and a bar chart for gender for the whole data set

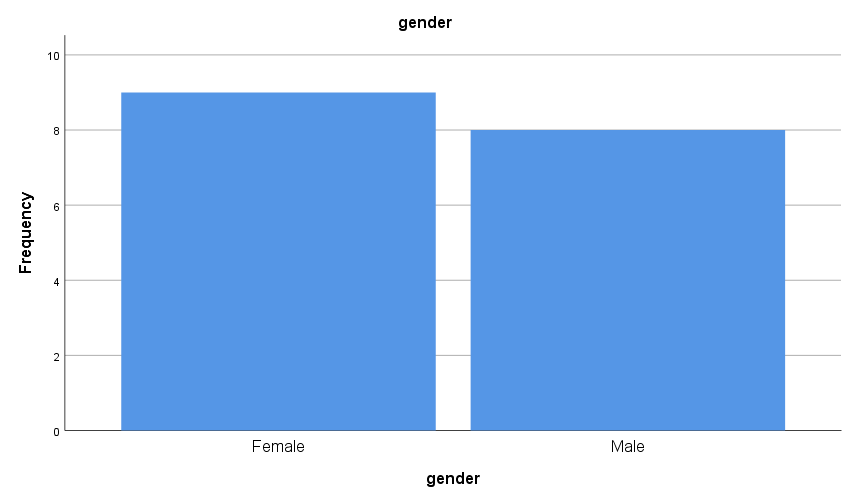
* Frequencies

FREQUENCIES VARIABLES=gender

/BARCHART FREQ

/ORDER=ANALYSIS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **gender** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Female | 9 | 52.9 | 52.9 | 52.9 |
| Male | 8 | 47.1 | 47.1 | 100.0 |
| Total | 17 | 100.0 | 100.0 |  |

* Bar chart for gender for the whole data set

**Discussion**

* The number of Female donors is slightly greater than the male donors

**Part C**

Descriptive statistics and a histogram with a normal curve for donated for women only

* Frequencies

USE ALL.

COMPUTE filter\_$=(gender=1).

VARIABLE LABELS filter\_$ 'gender=1 (FILTER)'.

VALUE LABELS filter\_$ 0 'Not Selected' 1 'Selected'.

FORMATS filter\_$ (f1.0).

FILTER BY filter\_$.

EXECUTE.

FREQUENCIES VARIABLES=filter\_$ donated

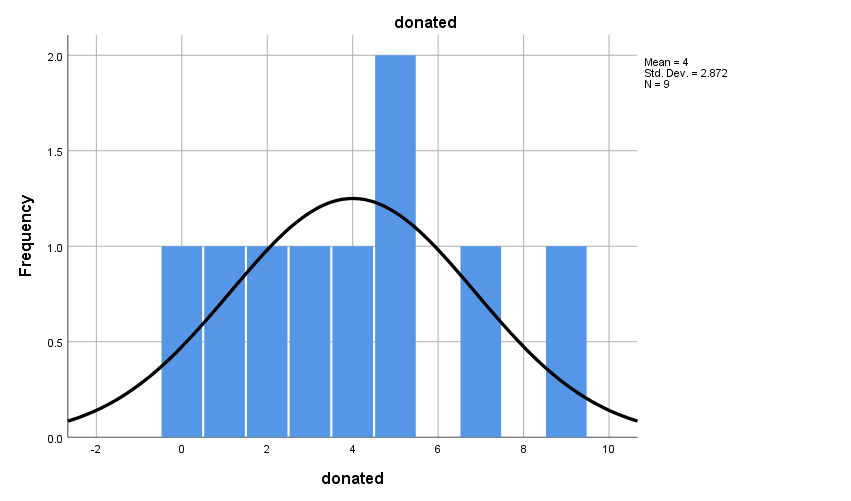
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE SUM

/HISTOGRAM NORMAL

/ORDER=ANALYSIS.

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | | | |
|  | | gender= 1 (FILTER) | donated |
| N | Valid | 9 | 9 |
| Missing | 0 | 0 |
| Mean | | 1.00 | 4.00 |
| Median | | 1.00 | 4.00 |
| Mode | | 1 | 5 |
| Std. Deviation | | .000 | 2.872 |
| Variance | | .000 | 8.250 |
| Range | | 0 | 9 |
| Minimum | | 1 | 0 |
| Maximum | | 1 | 9 |

* Histogram with a normal curve for donated for women only



**Discussion**

* Normality : The distribution of donations of female donors is reasonably normal (verified with normality tests)
* General Findings : The minimum value of donated ‘0’ indicates that a female has no prior donations and is donating blood for the very first time

**Part D**

Descriptive statistics and a histogram with a normal curve for donated for men only

* Frequencies

USE ALL.

COMPUTE filter\_$=(gender=2).

VARIABLE LABELS filter\_$ 'gender=2 (FILTER)'.

VALUE LABELS filter\_$ 0 'Not Selected' 1 'Selected'.

FORMATS filter\_$ (f1.0).

FILTER BY filter\_$.

EXECUTE.

FREQUENCIES VARIABLES=filter\_$ donated

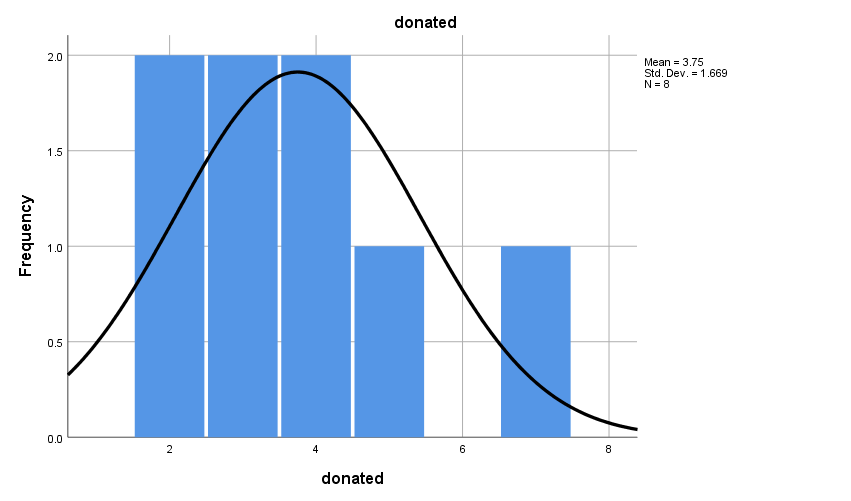
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE

/HISTOGRAM NORMAL

/ORDER=ANALYSIS.

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | | | |
|  | | gender= 2 (FILTER) | donated |
| N | Valid | 8 | 8 |
| Missing | 0 | 0 |
| Mean | | 1.00 | 3.75 |
| Median | | 1.00 | 3.50 |
| Mode | | 1 | 2a |
| Std. Deviation | | .000 | 1.669 |
| Variance | | .000 | 2.786 |
| Range | | 0 | 5 |
| Minimum | | 1 | 2 |
| Maximum | | 1 | 7 |

* Histogram with a normal curve for donated for men only



**Discussion**

* Normality: The distribution of donations of male donors is reasonably normal.
* Skewness: The distribution is slightly skewed towards right. (i.e., slightly positively-skewed distribution, where Mode < Median < Mean)

**Exercise Number: 4.4**

**Dataset Used**: [DataSet-1] G:\MS5105 - Statistical Techniques for Business Analytics\Assignments\1\Datasets\2\Ch 04 - Exercise 04.sav

**Part A**

Descriptive statistics and a histogram with a normal curve for atmsec for the whole data set

* Frequencies

FREQUENCIES VARIABLES=atmsec

/FORMAT=NOTABLE

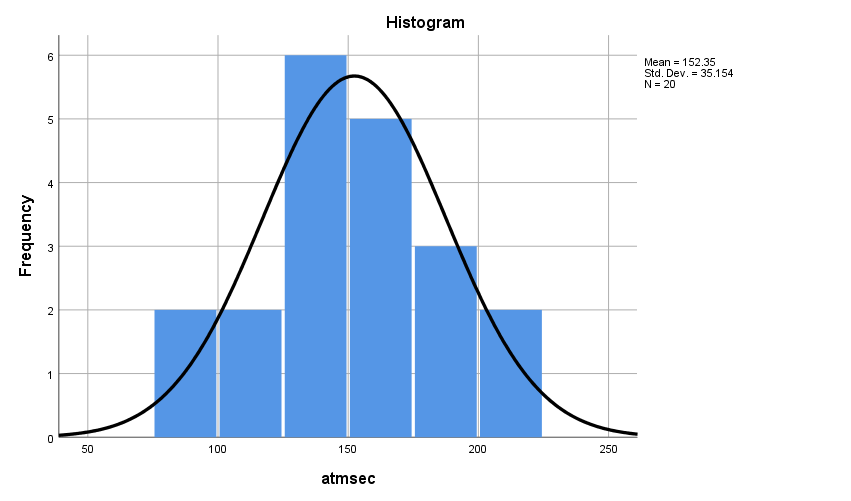
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE

/HISTOGRAM NORMAL

/ORDER=ANALYSIS

|  |  |  |
| --- | --- | --- |
| **Statistics** | | |
| atmsec | | |
| N | Valid | 20 |
| Missing | 0 |
| Mean | | 152.35 |
| Median | | 153.00 |
| Mode | | 142 |
| Std. Deviation | | 35.154 |
| Variance | | 1235.818 |
| Range | | 125 |
| Minimum | | 95 |
| Maximum | | 220 |
| Sum | | 3047 |

* Histogram with a normal curve for atmsec for the whole data set



**Normality Test Results**

A Shapiro-Wilk’s test (p>.05) (Shapiro & Wilk, 1965; Razali & Wah, 2011) and a visual inspection of their histograms and Normal Q-Q plots showed that the atmsec were approximately normally distributed for both male and female ATM users, with a skewness of 0.117 for females and a skewness of -0.463 for males (Doane & Seward, 2011; Kim 2013)

**Discussion**

* Normality : The distribution of atmsec for both male and female is reasonably normal

**Part B**

Descriptive statistics and a bar chart for gender for the whole data set

* Frequencies

FREQUENCIES VARIABLES=gender

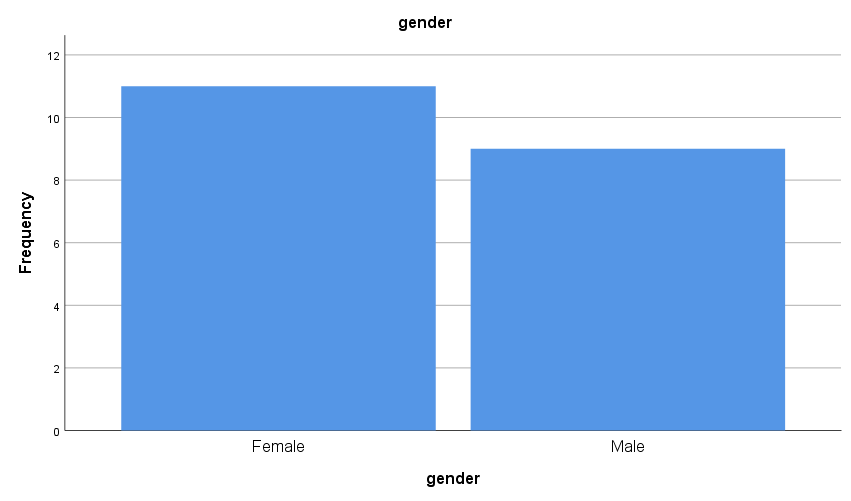
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE

/BARCHART FREQ

/ORDER=ANALYSIS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **gender** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Female | 11 | 55.0 | 55.0 | 55.0 |
| Male | 9 | 45.0 | 45.0 | 100.0 |
| Total | 20 | 100.0 | 100.0 |  |

* Bar chart for gender for the whole data set



**Discussion**

* The number female ATM users is greater than the male users in the given survey

**Part C**

Descriptive statistics and a histogram with a normal curve for atmsec for women only

* Frequencies

USE ALL.

COMPUTE filter\_$=(gender=1).

VARIABLE LABELS filter\_$ 'gender=1 (FILTER)'.

VALUE LABELS filter\_$ 0 'Not Selected' 1 'Selected'.

FORMATS filter\_$ (f1.0).

FILTER BY filter\_$.

EXECUTE.

FREQUENCIES VARIABLES=filter\_$ atmsec

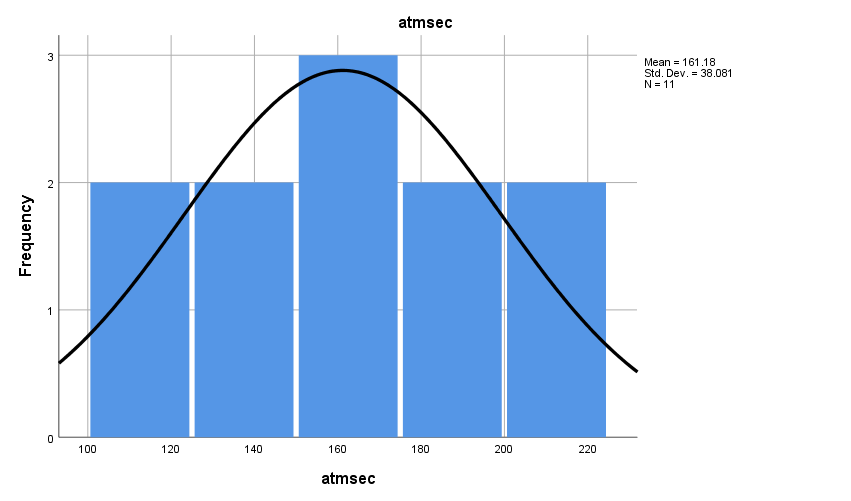
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE

/HISTOGRAM NORMAL

/ORDER=ANALYSIS.

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | | | |
|  | | gender=1 (FILTER) | atmsec |
| N | Valid | 11 | 11 |
| Missing | 0 | 0 |
| Mean | | 1.00 | 161.18 |
| Median | | 1.00 | 163.00 |
| Mode | | 1 | 107a |
| Std. Deviation | | .000 | 38.081 |
| Variance | | .000 | 1450.164 |
| Range | | 0 | 113 |
| Minimum | | 1 | 107 |
| Maximum | | 1 | 220 |
| Sum | | 11 | 1773 |
| a. Multiple modes exist. The smallest value is shown | | | |

* Histogram with a normal curve for atmsec for women only



**Discussion**

* Normality : The distribution for atmsec for female is reasonably normal

**Part D**

Descriptive statistics and a histogram with a normal curve for atmsec for men only

* Frequencies

USE ALL.

COMPUTE filter\_$=(gender=2).

VARIABLE LABELS filter\_$ 'gender=2 (FILTER)'.

VALUE LABELS filter\_$ 0 'Not Selected' 1 'Selected'.

FORMATS filter\_$ (f1.0).

FILTER BY filter\_$.

EXECUTE.

FREQUENCIES VARIABLES=filter\_$ atmsec

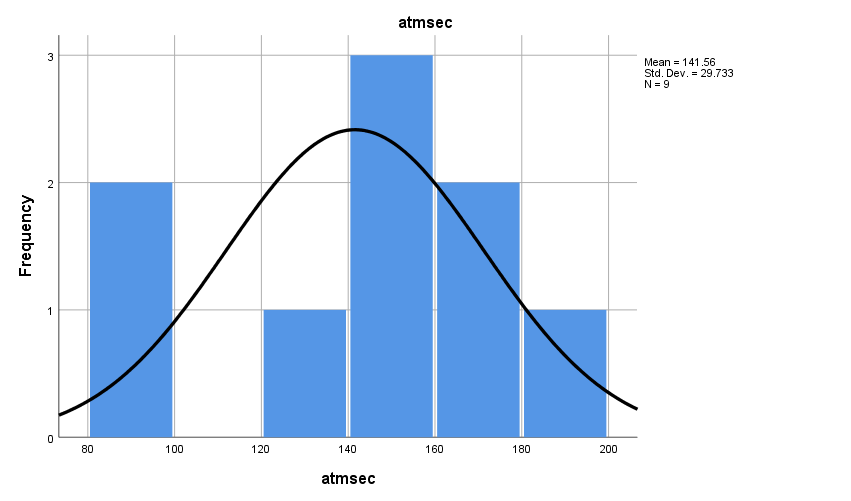
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE

/HISTOGRAM NORMAL

/ORDER=ANALYSIS.

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | | | |
|  | | gender=2 (FILTER) | atmsec |
| N | Valid | 9 | 9 |
| Missing | 0 | 0 |
| Mean | | 1.00 | 141.56 |
| Median | | 1.00 | 142.00 |
| Mode | | 1 | 142 |
| Std. Deviation | | .000 | 29.733 |
| Variance | | .000 | 884.028 |
| Range | | 0 | 90 |
| Minimum | | 1 | 95 |
| Maximum | | 1 | 185 |
| Sum | | 9 | 1274 |

* Histogram with a normal curve for atmsec for men only



**Discussion**

* Normality : The distribution for atmsec for men are reasonably normal

**Exercise Number: 4.6**

**Dataset Used**: [DataSet1] G:\MS5105 - Statistical Techniques for Business Analytics\Assignments\1\Datasets\3\Ch 04 - Exercise 06.sav

**Part A**

Descriptive statistics and a histogram with a normal curve for bill for the whole data set

* Frequencies

FREQUENCIES VARIABLES=bill

/FORMAT=NOTABLE

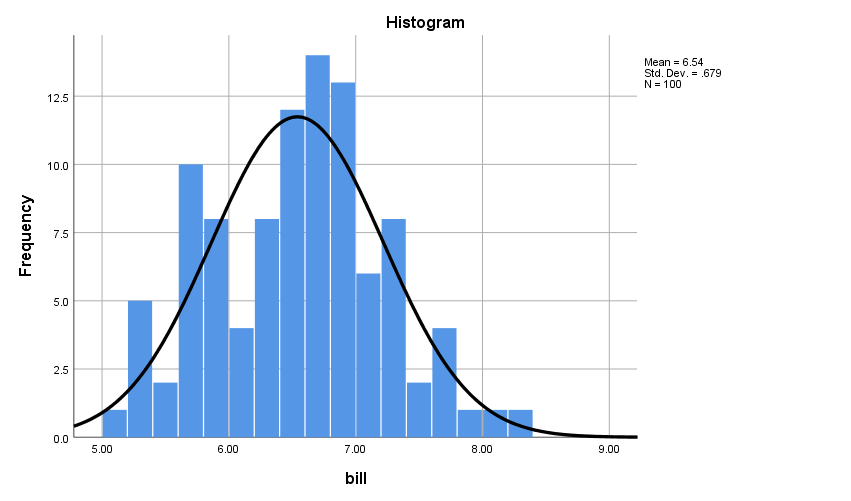
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE

/HISTOGRAM NORMAL

/ORDER=ANALYSIS

|  |  |  |
| --- | --- | --- |
| **Statistics** | | |
| bill | | |
| N | Valid | 100 |
| Missing | 0 |
| Mean | | 6.5403 |
| Median | | 6.6050 |
| Mode | | 6.70 |
| Std. Deviation | | .67939 |
| Variance | | .462 |
| Range | | 3.23 |
| Minimum | | 5.06 |
| Maximum | | 8.29 |
| Sum | | 654.03 |

* Histogram with a normal curve for bill for the whole data set



**Normality Test Results**

The Shapiro-Wilk’s test (p>.05) (Shapiro & Wilk, 1965; Razali & Wah, 2011) and a visual inspection of their histograms and Normal Q-Q plots showed that the bills were approximately normally distributed for nurses, doctors and others, with a skewness of 0.135 for nurses and a skewness of -0.047 for doctors and a skewness of-0.125 for others (Doane & Seward, 2011; Kim, 2013)

**Discussion**

* Normality: The bills are approximately normally distributed for nurses, doctors and others
* The data are distributed with focus to the mean (µ) hence results in a low standard deviation [σ =.679]

**Part B**

Descriptive statistics and a bar chart for profrole for the whole data set

* Frequencies

FREQUENCIES VARIABLES=profrole

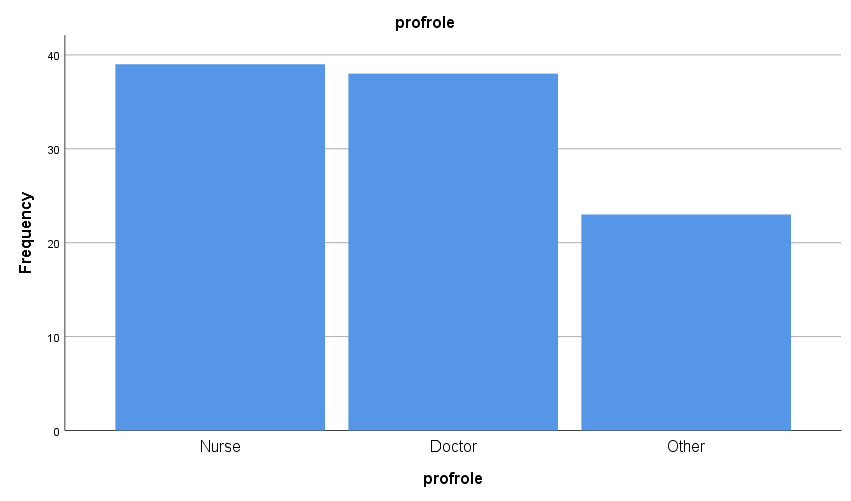
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE

/BARCHART FREQ

/ORDER=ANALYSIS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **profrole** | | | | | |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Nurse | 39 | 39.0 | 39.0 | 39.0 |
| Doctor | 38 | 38.0 | 38.0 | 77.0 |
| Other | 23 | 23.0 | 23.0 | 100.0 |
| Total | 100 | 100.0 | 100.0 |  |

* Bar chart for profrole for the whole data set



**Discussion**

* The doctors and nurses majorly used the hospital cafeteria.

**Part C**

Descriptive statistics and a histogram with a normal curve for bill for nurse only

* Frequencies

USE ALL.

COMPUTE filter\_$=(profrole=1).

VARIABLE LABELS filter\_$ 'profrole=1 (FILTER)'.

VALUE LABELS filter\_$ 0 'Not Selected' 1 'Selected'.

FORMATS filter\_$ (f1.0).

FILTER BY filter\_$.

EXECUTE.

FREQUENCIES VARIABLES=filter\_$ bill

/FORMAT=NOTABLE

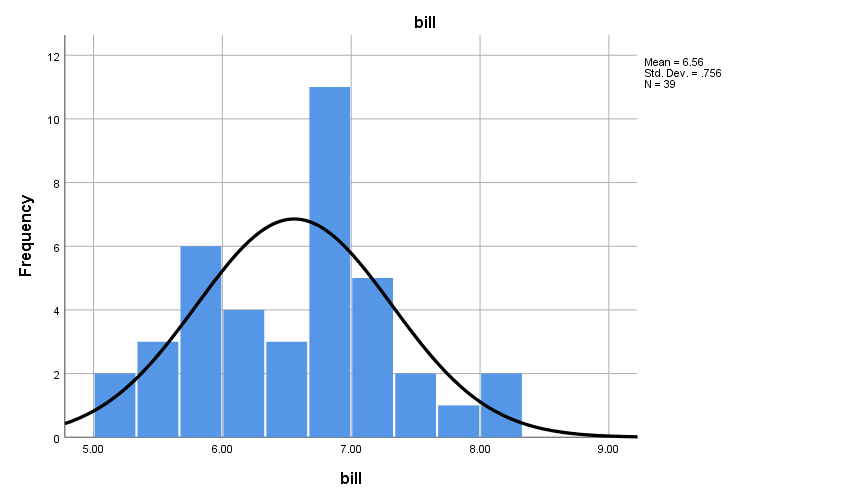
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE

/HISTOGRAM NORMAL

/ORDER=ANALYSIS.

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | | | |
|  | | profrole=1 (FILTER) | bill |
| N | Valid | 39 | 39 |
| Missing | 0 | 0 |
| Mean | | 1.00 | 6.5577 |
| Median | | 1.00 | 6.6800 |
| Mode | | 1 | 5.72a |
| Std. Deviation | | .000 | .75645 |
| Variance | | .000 | .572 |
| Range | | 0 | 3.04 |
| Minimum | | 1 | 5.25 |
| Maximum | | 1 | 8.29 |
| Sum | | 39 | 255.75 |
| a. Multiple modes exist. The smallest value is shown | | | |

* Histogram with a normal curve for bill for nurse only



**Discussion**

* Normality: The distribution of bill for nurses is reasonably normal.
* Skewness: The distribution of bills of nurses are slightly skewed towards right (slightly positively-skewed)

**Part D**

Descriptive statistics and a histogram with a normal curve for bill for doctor only

* Frequencies

USE ALL.

COMPUTE filter\_$=(profrole=2).

VARIABLE LABELS filter\_$ 'profrole=2 (FILTER)'.

VALUE LABELS filter\_$ 0 'Not Selected' 1 'Selected'.

FORMATS filter\_$ (f1.0).

FILTER BY filter\_$.

EXECUTE.

FREQUENCIES VARIABLES=filter\_$ bill

/FORMAT=NOTABLE

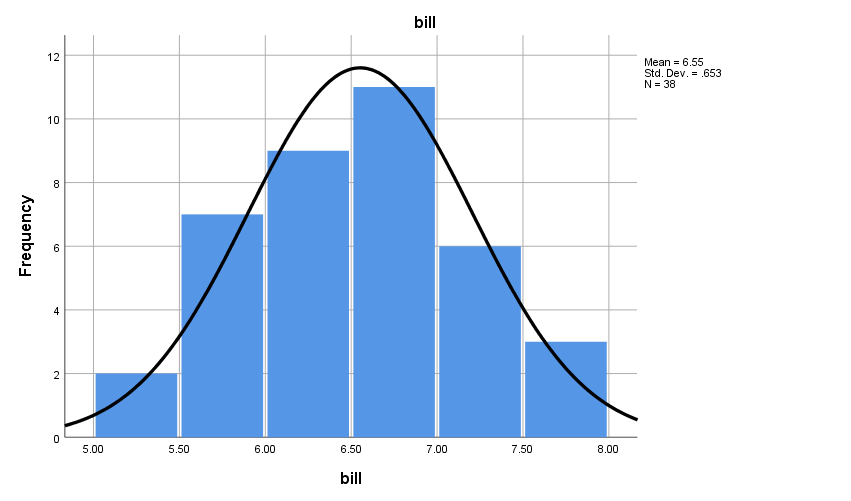
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE

/HISTOGRAM NORMAL

/ORDER=ANALYSIS.

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | | | |
|  | | profrole=2 (FILTER) | bill |
| N | Valid | 38 | 38 |
| Missing | 0 | 0 |
| Mean | | 1.00 | 6.5524 |
| Median | | 1.00 | 6.5350 |
| Mode | | 1 | 6.70 |
| Std. Deviation | | .000 | .65319 |
| Variance | | .000 | .427 |
| Range | | 0 | 2.82 |
| Minimum | | 1 | 5.06 |
| Maximum | | 1 | 7.88 |
| Sum | | 38 | 248.99 |

* Histogram with a normal curve for bill for doctor only



**Discussion**

* Normality: The distribution of bill for doctors is reasonably normal.

**Part E**

Descriptive statistics and a histogram with a normal curve for bill for other only

* Frequencies

USE ALL.

COMPUTE filter\_$=(profrole=3).

VARIABLE LABELS filter\_$ 'profrole=3 (FILTER)'.

VALUE LABELS filter\_$ 0 'Not Selected' 1 'Selected'.

FORMATS filter\_$ (f1.0).

FILTER BY filter\_$.

EXECUTE.

FREQUENCIES VARIABLES=filter\_$ bill

/FORMAT=NOTABLE

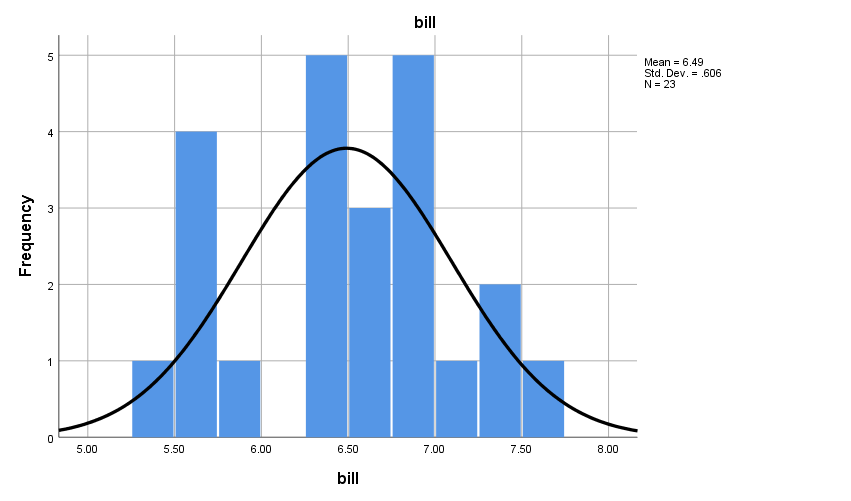
/STATISTICS=STDDEV VARIANCE RANGE MINIMUM MAXIMUM MEAN MEDIAN MODE

/HISTOGRAM NORMAL

/ORDER=ANALYSIS.

|  |  |  |  |
| --- | --- | --- | --- |
| **Statistics** | | | |
|  | | profrole=3 (FILTER) | bill |
| N | Valid | 23 | 23 |
| Missing | 0 | 0 |
| Mean | | 1.00 | 6.4909 |
| Median | | 1.00 | 6.5300 |
| Mode | | 1 | 6.34 |
| Std. Deviation | | .000 | .60643 |
| Variance | | .000 | .368 |
| Range | | 0 | 2.16 |
| Minimum | | 1 | 5.46 |
| Maximum | | 1 | 7.62 |
| Sum | | 23 | 149.29 |

* Histogram with a normal curve for bill for other only



**Discussion**

* Normality: The distribution of bill for others is reasonably normal.

**References**

* Shapiro, S.S., & Wilk, M.B (1965). An Analysis of Variance Test for Normality (Complete Samples), Biometrika ,Vol. 52, No. 3/4 (Dec., 1965), pp. 591-611

Available at : <https://www.jstor.org/stable/2333709>

* Razali, N.M., & Wah, Y.B.(2011) Power Comparisions of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. Journal of Statistical Modelling and Analytics 2(1) 21-33. Available at : <https://www.researchgate.net/publication/267205556_Power_Comparisons_of_Shapiro-Wilk_Kolmogorov-Smirnov_Lilliefors_and_Anderson-Darling_Tests>
* Doane, D.P., & Seward, L.E (2011). Measuring Skewness: A Forgotten Statistic? Journal of Statistics Education Volume 19, Number 2(2011).

Available at : www.amstat.org/publications/jse/v19n2/doane.pdf

* Kim, H.Y., (2013). Assessing Normal Distribution (2) through skewness and kurtosis.

Available at : <https://www.rde.ac/Synapse/Data/PDFData/2185RDE/rde-38-52.pdf>