PAPER TITLE CENTERED AND IN CAPS LOCK

by Student's Name

Code + Course Name

Professor's Name

University Name

City, State

Date

Table of Contents

I. Task 1

The researcher used simple random sampling to select a sample of 200 households. In the opinion of the researcher, simple random sampling was the best sampling technique as it enabled that each household had an equal chance of selection.

II. Descriptive Statistics and Box-Whisker Plots

i)

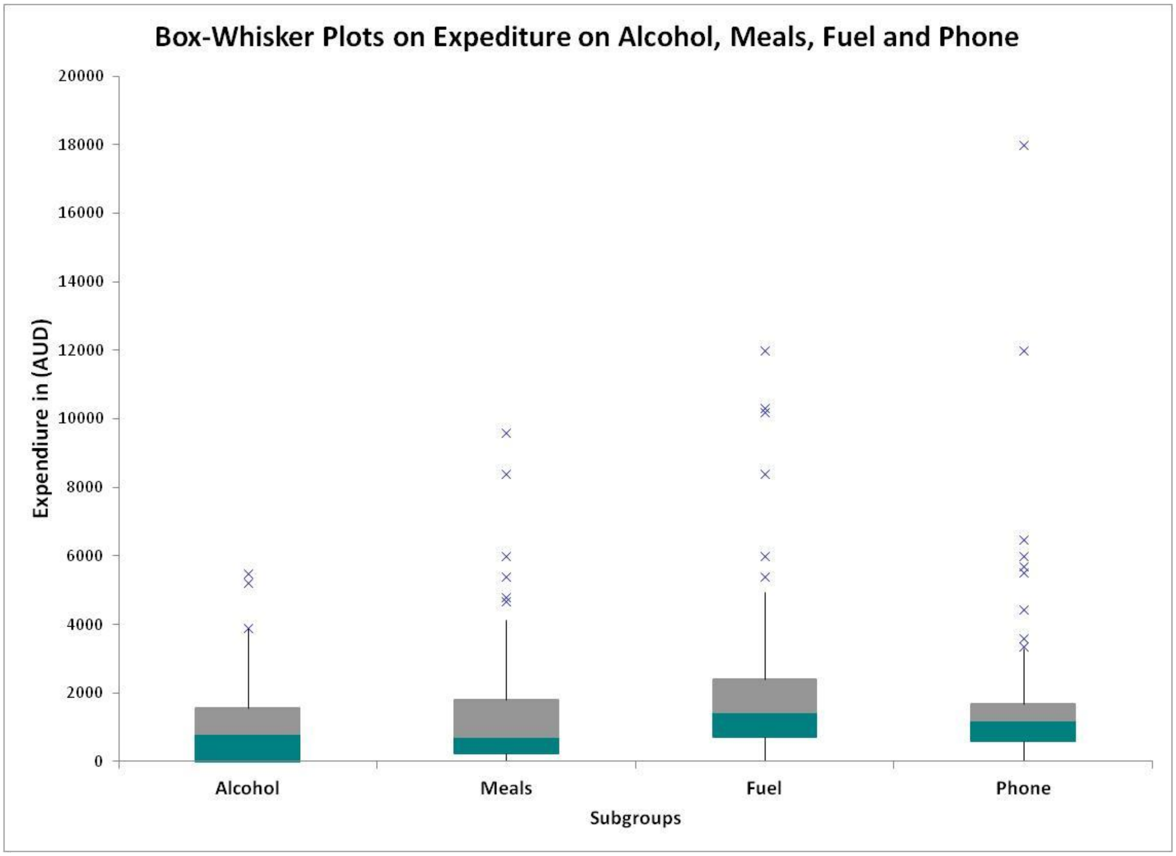

Figure 1: Box-Whisker Plots on Expenditure on Alcohol, Meals, Fuel, and Phone

Figure 1 above shows the box-whisker plots of the expenditures on alcohol, meals, fuel, and phone. The “x” marking above each of the box-whisker plots suggests the outliers. Outliers are abnormal observations in the data (King’oriah, 2004). By looking at the box-whisker plots above, it is clear that phone expenditure has more extreme observations than the other types of expenditure with the highest observation at 18,000 AUD. Alcohol recorded the least number of extreme values with a maximum expenditure of 9,600 AUD

C. To measure variation of each the expenditure s coefficient of variation was found to be appropriate. The coefficient of variation derived by dividing the standard deviation by the mean and multiplying by one hundred. For alcohol expenditure, the coefficient of variation is:

1270.45/1115.6 x 100= 113.88

The variability derived above is about 114 percent, which implies that expenditure in alcohol is expected to vary over 100 percent of the time. That is to say we would expect based on the sample different results from the population.

For Meals expenditure the coefficient of variation is;

1460.07/1210.19 x 100 = 120.65

The coefficient of variation for meals was about 121 percent. This result also shows a high variability meaning that we expect the average expenditure on meals to change over 120 percent of the time.

The coefficient of variation for fuel was;

1782.17/1859.09 x 100 = 95.86

The variability of fuel was about ninety-six percent meaning variation from the values of fuel is likely to differ 96% of the time.

The coefficient of variation of phone expenditure was;

1786.42/1469.82 x 100 = 121.54

The variability of fuel expenditure was recorded as about 122% that meant we would expect fuel consumption to vary over one hundred and twenty percent of the time.

D.

Table 1: Descriptive Statistics for Alcohol Expenditure

Alcohol		*Meals*		*Fuel*		*Phone*	
Mean	1115.6	Mean	1210.19	Mean	1859.09	Mean	1469.82
Standard Error	89.83463983	Standard Error	103.2426654	Standard Error	126.0181528	Standard Error	126.3189647
Median	808.5	Median	720	Median	1440	Median	1194
Mode	0	Mode	0	Mode	1200	Mode	1200
Standard Deviation	1270.45366	Standard Deviation	1460.071776	Standard Deviation	1782.165808	Standard Deviation	1786.419931
Sample Variance	1614052.503	Sample Variance	2131809.592	Sample Variance	3176114.967	Sample Variance	3191296.168
Kurtosis	2.575335161	Kurtosis	8.215278832	Kurtosis	9.40534103	Kurtosis	42.35244383
Skewness	1.565454209	Skewness	2.403832623	Skewness	2.509999649	Skewness	5.482269623
Range	5475	Range	9600	Range	12000	Range	18000
Minimum	0	Minimum	0	Minimum	0	Minimum	0
Maximum	5475	Maximum	9600	Maximum	12000	Maximum	18000
Sum	223120	Sum	242038	Sum	371818	Sum	293964
Count	200	Count	200	Count	200	Count	200

According to table 1 above, the average expenditure for the sampled 200 households was 1115.6 AUD with a standard deviation from the mean of 1270.45 (M= 1115.6, SD= 1270.45, N= 200). The median was 808.5AUD while the maximum expenditure on alcohol was 5475 AUD. A skew is a longer tail of the frequency distribution caused by the existence of a number of very large or very small observations (King'oriah, 2004). Skewness depicts the variability of data and the direction of that variability. From the table above the analysis revealed a positive skew of 1.57 meaning that most observations were distributed below the mean. Kurtosis is a measure that shows or explains the shape of a random parameter probability distribution. The analysis above revealed a kurtosis of 2.5.

The average meals expenditure according to the results was 1210.19 AUD associated with a standard deviation of 1,460.07AUD (M= 1,210.19, SD= 1,460.07, N= 200). There was a slightly higher expenditure on meals than on alcohol. The highest expenditure on meals from the sample was 9,600 AUD with a median of 720 AUD. Meals expenditure had a positive skewness of 2.4 and kurtosis of 8.2.

By comparison of all the expenditures under review, fuel recorded the highest expenditure at 12,000 AUD with an average expenditure of 1,859. 09 (M= 1859.09, SD= 1,782.17, N= 200). The median was 1,440 AUD and mode of 1,200 AUD with a positive skew of about 2.51while kurtosis was 9.41.

Phone expenditure averaged 1469.82 AUD with a standard deviation of 1,786.41 (M= 1,469.82, SD= 1,789.41, N= 200). The median was 1,194 AUD and mode of 1,200 AUD. The skewness was 5.48 with kurtosis of 42.35. Phone expenditure was the second highest followed by meals and lastly alcohol.

II. Task 2

A.

Table 2: Frequency Distribution of Utilities Expenditures

Bin	*Frequency*	*Cumulative %*
400	22	11.00%
801	49	35.50%
1201	50	60.50%
1601	34	77.50%
2001	24	89.50%
2401	8	93.50%
2801	4	95.50%
3200	5	98.00%
6400	4	100.00%

B

1. According to the frequency distribution in table 2 above, the percentage of households that spent up to 1200 AUD on utilities was 35.5%

2. The percentage of households that spent between $ 1200 and $ 2400 from the frequency distribution above was thirty-nine percent.

3. The percentage of households that spent more than $ 2400 on utilities was six and a half percent.

C.

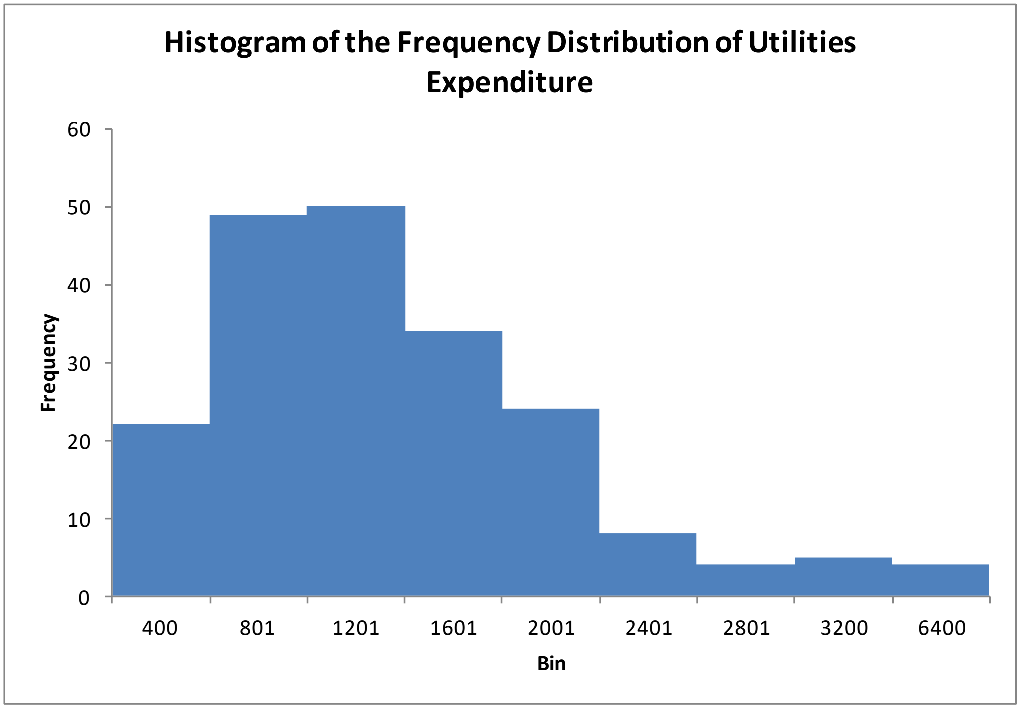

Figure 2: Histogram of the Frequency Distribution of Utilities Expenditure

Observing from the histogram in figure 2 above, it is clear that the utility expenditures are not normally distributed. On close examination if we were to draw a line attaching the mid-point of each bar we would get a positive skew meaning most of the observations would fall to the right side of the mean (M= 1,229.88). This suggests some abnormal observations from the samples data implying a number of outliers. It would be prudent to identify the outliers and eliminating them using a stem and leaf plot so as to normalize the data. However, it should be noted that perfect normal distribution of data is not possible in practice but only in theory. In any real data collected, a researcher would expect some slight skewness (negative or positive) and the best remedy would be to eliminate extreme cases or observations.

III. Task 3

A. To determine the top 10% and the bottom 10% of values of household is annual after tax income, the researcher made use of percentiles. Percentiles rank the nth percentage of a value that falls below a specified percentage. In this case to determine the top 10% of values of household's annual after tax income, the 90th percentile was used. Likewise, to determine the bottom 10% value of household's annual after tax income, the 10th percentile was used. Table 3 below shows the results

Table 3: Top 10% and Bottom 10% Values of household's Annual after Tax Income

Rank	Value	Formula
Top 10%	116809.1	=PERCENTILE(A2:A201,0.9)
Bottom 10%	16560	=PERCENTILE(A2:A201,0.1)

The value of $ 116,809.10 implies that 90% of the households' annual after tax income falls below $ 116,809.10 from the sampled data. Likewise, the value of $ 16,560 implies that 10% of the households' annual after tax income falls below $ 16560.

B. Since the variable ownhouse is categorical in nature it would be to derive any measure of central tendency such as the mean, median or mode. This is because there is no particular ranking of the values coded for either owning a house or not owning a house. The values provided are just to assist in coding the data for further analysis. To determine the number of households that own a house can be done by doing a frequency count of those who own a house and those who do not. In this case 146 (73%) households own a house while 54 (27%) of households do not.

C.

Table 4: Probability of a Household having a Family Size Equal to 5

	A	B
1	**Data**	**Description**
2	5	Value of distribution
3	1.2925	Mean
4	1.1225	Standard deviation
	Formula	**Description (Result)**
	=NORMDIST(A3,A4,A5,0)	0.00152

Table 4 above revealed that the probability of a household having a family size equal to five was 0.00152.

D.

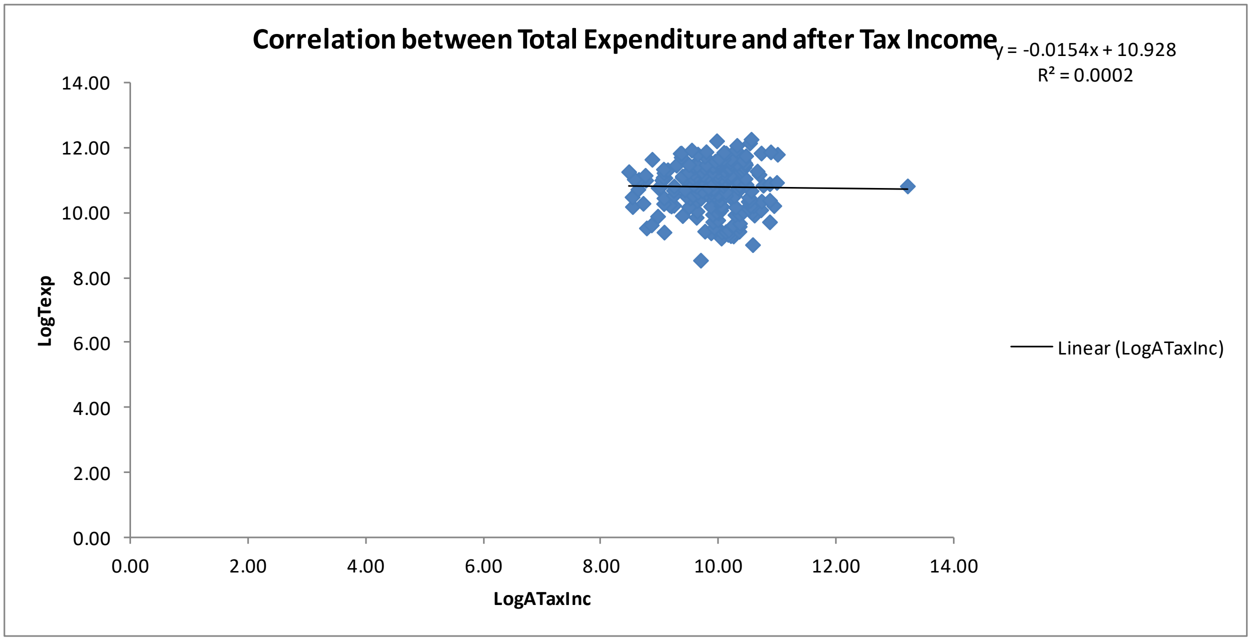

Figure 3: Correlation between Total Expenditure and after Tax Income

According to figure 3 above shows a coefficient of determination of 0.0002 (0.02%). This implies that a change in total expenditure can be explained by only 0.02% changes in after tax income. The coefficient of correlation is derived from finding the square root of the coefficient of determination (0.0002 x 0.0002 = 0.01414).This gives a very weak positive correlation coefficient of 0.01414 (1.414%).

IV. Task 4

Table 5: Contingency Table between the gender and the level of education

% of Highest Degree	B	I	M	P	S	Grand Total
F	53.66%	46.67%	36.59%	47.50%	43.75%	45.50%
M	46.34%	53.33%	63.41%	52.50%	56.25%	54.50%
Grand Total	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**
Frequency of Highest Degree	**B**	**I**	**M**	**P**	**S**	**Grand Total**
F	22	14	15	19	21	91
M	19	16	26	21	27	109
Grand Total	**41**	**30**	**41**	**40**	**48**	**200**

A. From the contingency table above, it is evident that more males (63.41%) than females (36.59%) have attained master's degree. However, more female (53.66%) than male (46.34%) have bachelor's degree. From the above comparison, it is clear that male and females differ in their highest level of education.

B. Probability that head of household is female and level of education is intermediate can be expressed as:

P [A and B] = P (A) x P (B)

Where P (A)= head of household is female

P (B) = Intermediate level of education

P [A and B] = 91/200 x 14/30 = 0.456 x 0.467 = 0.213

C. Probability that head of household is male and has Bachelor degree is expressed as;

P [A and B] = P (A) x P (B)

Where P (A)= head of household is male

P (B) = bachelor level of education

P [A and B] = 109/200 x 19/41 = 0.545 x 0.463 = 0.252

D. The proportion of females with secondary education as the highest level of education is 21 households (43.75%) out of forty-eight households that recorded secondary education as highest education level.

The event of gender of household being male and with master degree is two independent events, as one event does not necessarily affect the other. They are said to be mutually exclusive events

V. References

King'oriah, GK 2004, *Fundamentals of applied statistics.* Nairobi, Kenya: The Jomo Kenyatta Foundation