Journal of Statistical Science and Application 5 (2017) 118-126

Test of Normality of Waist Measurement Data of Young Male and Female Adults based on the Quantile - Quantile Plot

F. Z. Okwonu[1] and J. N. Igabari[2]
[1,2]Department of Mathematics, Delta State University, Abraka, Nigeria.
e-mail: [1]fzokwonu_delsu@yahoo.com, [2]*jn_igabari@delsu.edu.ng*

This paper investigates the normality of some real data set obtained from waist measurements of a group of 49 young adults. The quantile - quantile (Q-Q) plot and the analysis of correlation coefficients for the Q-Q plot is used to determine the normality or otherwise of the data set. In this regards, the probabilities of the quantiles were computed, modified and plotted. Thereafter the correlation coefficients for the quantile – quantile plots were obtained. Results indicate that at 0.1 level of significance, the data for young adult males of the sample were not normally distributed, and had a mean value that is within the range of low risk, healthwise, whereas the distribution of the data for young female adults showed reasonable normality, but also with a mean value that is within the range of low risk in terms of health condition.

Keywords: Correlation coefficient, probability plot, quantile - quantile plot, test of normality, waist measurement, young adults.

Introduction

Recent research findings [1,2] have given credence to the notion that waist measure can be a clue to the health status of an individual. Such findings have continued to generate interest as studies have shown that in several instances, Waist to Hip Ratio has proved to be a simpler measurement for assessment of lifestyle health risk and overweight. It is even considered by many to be a more objective and stable measurement than Body Mass Index (BMI)[3]. In testing for normality by means of graphical procedures, specifically, normal probability plot or the quantile - quantile plot (Q-Q plot), a formal comparative analysis is performed between the cumulative distribution of the obtained data set and the cumulative distribution of the normal distribution[4] or the normal quanitle in the case of the Q-Q plot. In general, the obtained data set is plotted to compare with the expected straight line generated by the normal distribution plot. With this illustration and assuming the distribution is normal then a straight line in the diagonal is generated. If the data set of interest is normal it will reasonably approximate the straight line generated by the normal distribution. Often the case of univariate normality is what the bivariate or multivariate normality depends on[5].

This paper is organized in such a way that the quantile - quantile procedure is presented first, while discussion and analysis comes next, and then conclusions follow.

Methods of the quantile - Quantile plot

This procedure for testing normality of data set has been discussed in many papers [5,6]. The procedure involves arranging the data points in ascending order with their respective probabilities and quantile values.

Define the data point as: $X_m, m = 1, 2, 3, \dots, l$

with unique profile variable by ordering the data points as: $X_1 \leq X_2 \leq X_3 \leq \cdots \leq X_l$

Then $X_m, m = 1, 2, 3, \dots, l$ are the data point quantiles.

Define $l/_m$ as the corrected proportion of the data point denoted as $(l - 0.5)/_m$.

The quantiles μ_{ql} are given by

$$P(W \leq q_l) = \int_{-\infty}^{q_i} \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} dw = p_l = \frac{l-\alpha}{m}$$

$$\rightarrow P(W \leq q_l) = p_l = \frac{l-\alpha}{m}, (\alpha = 0.5) \tag{1}$$

Where α is the correction factor.

It has been suggested [6] that $p_l = \frac{l-\alpha}{m}$ be modified such that $\widetilde{p_i} = \frac{l-\beta}{m+\varepsilon}, \beta = \frac{3}{8}\ and\ \varepsilon = \frac{1}{4}$ (2)

In this discuss, the focus is to compare the values of Equations (1) and (2) by modifying the equations such that their average is defined as follows:

$$\overline{p_l} = \frac{p_l + \widetilde{p_l}}{2} \tag{3}$$

The objective of this modification is to compare their numerical values and to determine the information which the probability plot presents. The respective values of these equations are based on their quantile values. The idea is to pair the quantile values and the data points, that is (u_{ql}, x_l) this is to enable the plotting of the quantile - quantile plot to determine if the data set are normally distributed or otherwise. Suppose the data set comes from a normal distribution, then (u_{ql}, x_l) maybe a straight line. The conclusion of this graphical presentation in some cases may cast doubt though depending on the sample size of the data set. The correlation coefficient for the quantile quantile plot is defined as:

$$C_R = \frac{\sum_{i=1}^{k}(x_l - \bar{x})(u_{ql} - x_l)}{\sqrt{\sum_{i=1}^{k}(x_l - \bar{x})^2 \sum_{i=1}^{k}(u_{ql} - x_l)^2}} \tag{4}$$

This is applied to validate the conclusion of the Q - Q plot. This equation will help to determine the straightness of the quantile quantile plot [6].

Data presentation and Analysis

Waist measurements were taken from a group of 49 young adult volunteers, consisting of 32 males and 17 females, all undergraduate students of the fourth year in a certain University in Nigeria.

Tables 1 and Table 3 contain the raw data X_i for the males and females respectively, and the results of the analyses of data based on equations (1), (2) and (3). Similarly, Tables 2 and 4 show the computed values and the critical values for the respective data. We observe that each equation has unique probabilities and quantile values and this strictly reflected on the computed values of the correlation coefficient. From Table 2, equation (3) is relatively robust compared to the values obtained from the other equations. Fig. 1 shows the quantile quantile plot, while Fig.2, Fig.3 and Fig.4 are the plots of probabilities against the quantile values of Table 1.

Table 1

Analysis of waist measurement data for selected young male adults

	$p_l = \frac{l-\alpha}{m}$		$\bar{p_l} = \frac{p_l + \tilde{p_l}}{2}$		$\tilde{p_i} = \frac{l-\beta}{m+\varepsilon}$	
X_l	PROB	QUANTILES	PROB	QUANTILE	PROB	QUANTILE
28	0.015625	-2.153875	0.0175024	-2.108302	0.0193798	-2.066729
28	0.046875	-1.67594	0.0486313	-1.658272	0.0503876	-1.641107
28	0.078125	-1.417797	0.0797602	-1.406687	0.0813953	-1.395747
28	0.109375	-1.229859	0.1108891	-1.221814	0.1124031	-1.213847
28	0.140625	-1.077516	0.1420179	-1.071297	0.1434109	-1.06512
28	0.171875	-0.946782	0.1731468	-0.941803	0.1744186	-0.936847
28	0.203125	-0.830511	0.2042757	-0.826446	0.2054264	-0.822394
28	0.234375	-0.724514	0.2354046	-0.721163	0.2364341	-0.71782
28	0.265625	-0.626099	0.2665334	-0.623331	0.2674419	-0.620568
29	0.296875	-0.53341	0.2976623	-0.531136	0.2984496	-0.528865
29	0.328125	-0.445097	0.3287912	-0.443254	0.3294574	-0.441412
29	0.359375	-0.36013	0.3599201	-0.358672	0.3604651	-0.357216
30	0.390625	-0.27769	0.3910489	-0.276586	0.3914729	-0.275482
30	0.421875	-0.197099	0.4221778	-0.196325	0.4224806	-0.195551
30	0.453125	-0.11777	0.4533067	-0.117311	0.4534884	-0.116853
30	0.484375	-0.039176	0.4844356	-0.039024	0.4844961	-0.038872
30	0.515625	0.0391761	0.5155644	0.0390242	0.5155039	0.0388722
30	0.546875	0.1177699	0.5466933	0.1173113	0.5465116	0.1168527
30	0.578125	0.1970991	0.5778222	0.1963252	0.5775194	0.1955515
30	0.609375	0.2776904	0.6089511	0.2765862	0.6085271	0.2754823
30	0.640625	0.3601299	0.6400799	0.3586725	0.6395349	0.3572158
31	0.671875	0.4450965	0.6712088	0.4432535	0.6705426	0.441412
32	0.703125	0.5334097	0.7023377	0.5311359	0.7015504	0.5288648
32	0.734375	0.626099	0.7334666	0.6233313	0.7325581	0.6205683
32	0.765625	0.7245144	0.7645954	0.7211632	0.7635659	0.7178201
32	0.796875	0.8305109	0.7957243	0.8264456	0.7945736	0.822394
32	0.828125	0.9467818	0.8268532	0.9418028	0.8255814	0.9368471
32	0.859375	1.0775156	0.8579821	1.0712971	0.8565891	1.0651198
32	0.890625	1.2298588	0.8891109	1.2218137	0.8875969	1.2138469
32	0.921875	1.4177971	0.9202398	1.4066866	0.9186047	1.395747
32	0.953125	1.6759397	0.9513687	1.6582719	0.9496124	1.6411071
32	0.984375	2.1538747	0.9824976	2.1083023	0.9806202	2.0667291

Table 2

Computed values and critical values (bold italics) of the coefficient of correlation

C_R (1)	C_R (2)	C_R (3)
0.9189 *(0.9715)*	0.9208 *(0.9715)*	0.9224 *(0.9715)*

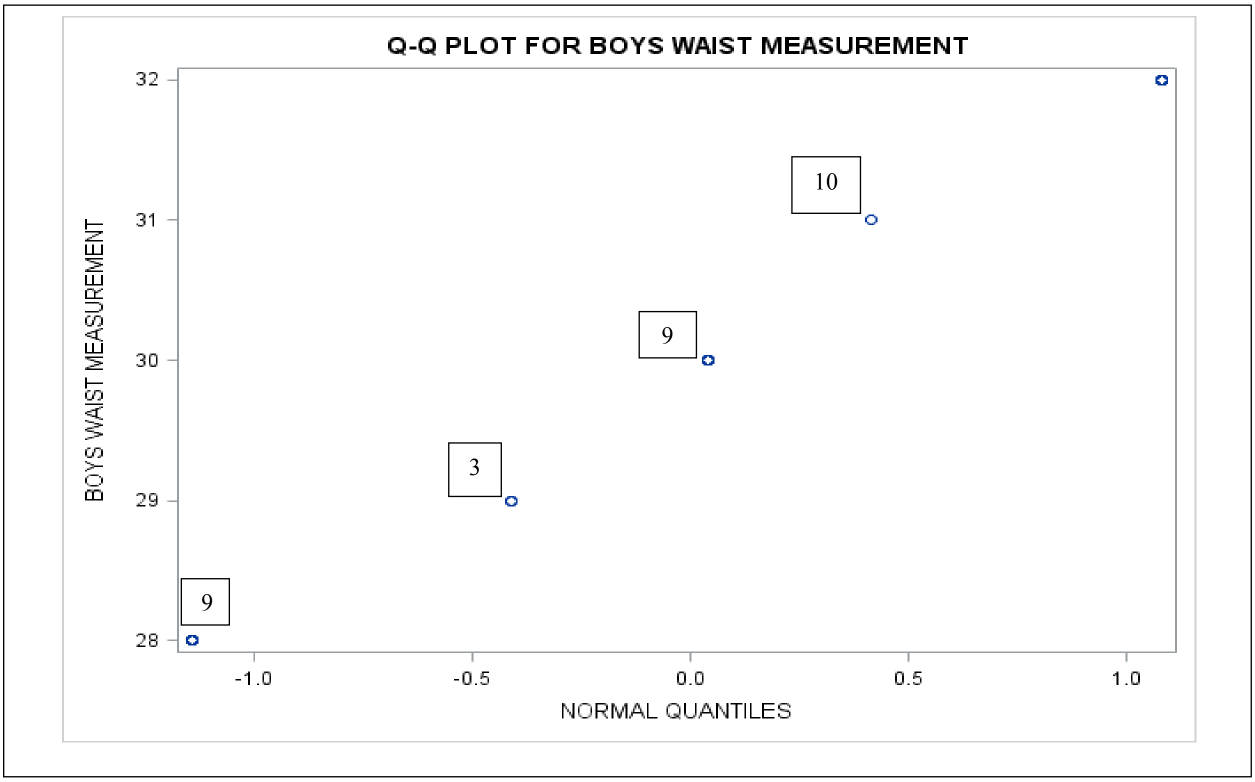

Figure 1. The Q - Q plot waist measurement of young adult males. (The values in the box indicate the frequency of the data points).

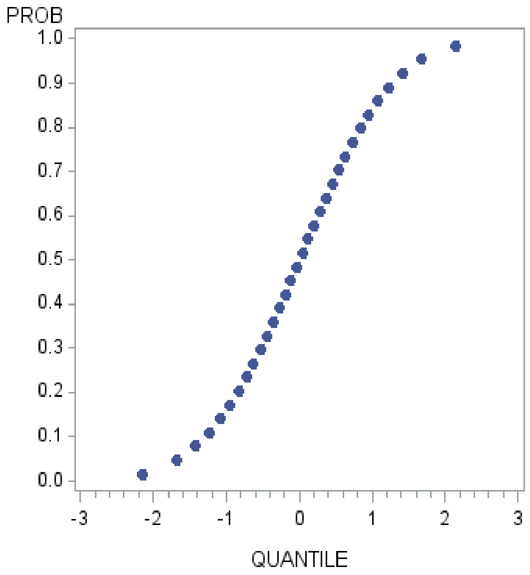

Figure 2. probability plot for table 1 with $p_l = \frac{l-\alpha}{m}$.

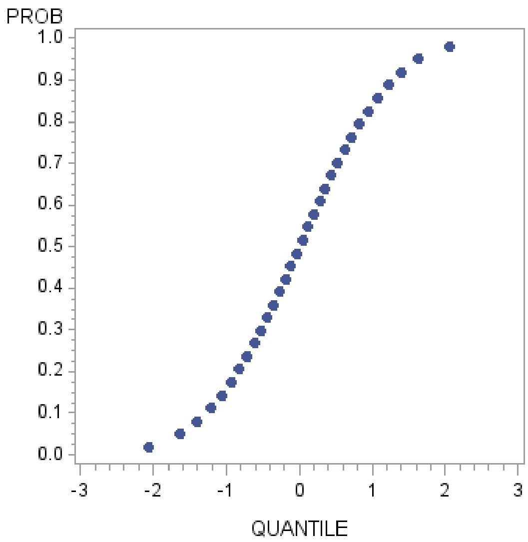

Figure 3. probability plot for table 1 with $\overline{p_l} = \frac{p_l + \widetilde{p_l}}{2}$.

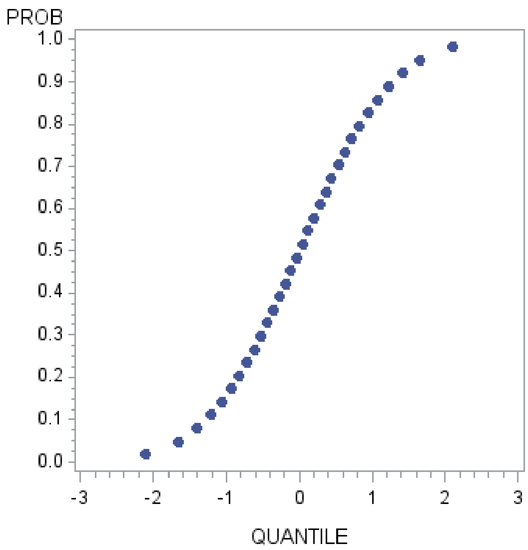

Figure 4. probability plot for table 1 with $\widetilde{p_i} = \frac{l-\beta}{m+\varepsilon}$.

Table 3

Analysis of waist measurement data for selected young female adults.

	$p_l = \frac{l-\alpha}{m}$		$\widetilde{p_i} = \frac{l-\beta}{m+\varepsilon}$		$\overline{p_l} = \frac{p_l + \widetilde{p_l}}{2}$	
X_l	PROB	QUANTILE	PROB	QUANTILE	PROB	QUANTILE
27	0.0294118	-1.88951	0.0328218	-1.840849	0.0362319	-1.796193
27	0.0882353	-1.351702	0.0912191	-1.333285	0.0942029	-1.31531
28	0.1470588	-1.049131	0.1496164	-1.03808	0.1521739	-1.027154
28	0.2058824	-0.820792	0.2080136	-0.813333	0.2101449	-0.805918
28	0.2647059	-0.628904	0.2664109	-0.623704	0.2681159	-0.618521
28	0.3235294	-0.457852	0.3248082	-0.454295	0.326087	-0.450744
29	0.3823529	-0.299307	0.3832055	-0.297073	0.384058	-0.29484
29	0.4411765	-0.147987	0.4416027	-0.146907	0.442029	-0.145827
30	0.5000000	-1.15E-17	0.5000000	-1.15E-17	0.500000	-1.15E-17
30	0.5588235	0.1479871	0.5583973	0.146907	0.557971	0.145827
30	0.6176471	0.2993069	0.6167945	0.2970728	0.615942	0.2948402
30	0.6764706	0.4578519	0.6751918	0.4542952	0.673913	0.4507442
31	0.7352941	0.6289042	0.7335891	0.6237043	0.7318841	0.6185211
32	0.7941176	0.8207921	0.7919864	0.8133328	0.7898551	0.8059185
32	0.8529412	1.0491314	0.8503836	1.0380802	0.8478261	1.0271543
32	0.9117647	1.3517022	0.9087809	1.3332853	0.9057971	1.3153098
32	0.9705882	1.88951	0.9671782	1.8408493	0.9637681	1.7961934

Table 4

Computed values and critical values (bold italics) of the coefficient of correlation

C_c (1)	C_c (2)	C_c (3)
0.9579 *(0.9543)*	0.9594 *(0.9543)*	0.9608 *(0.9543)*

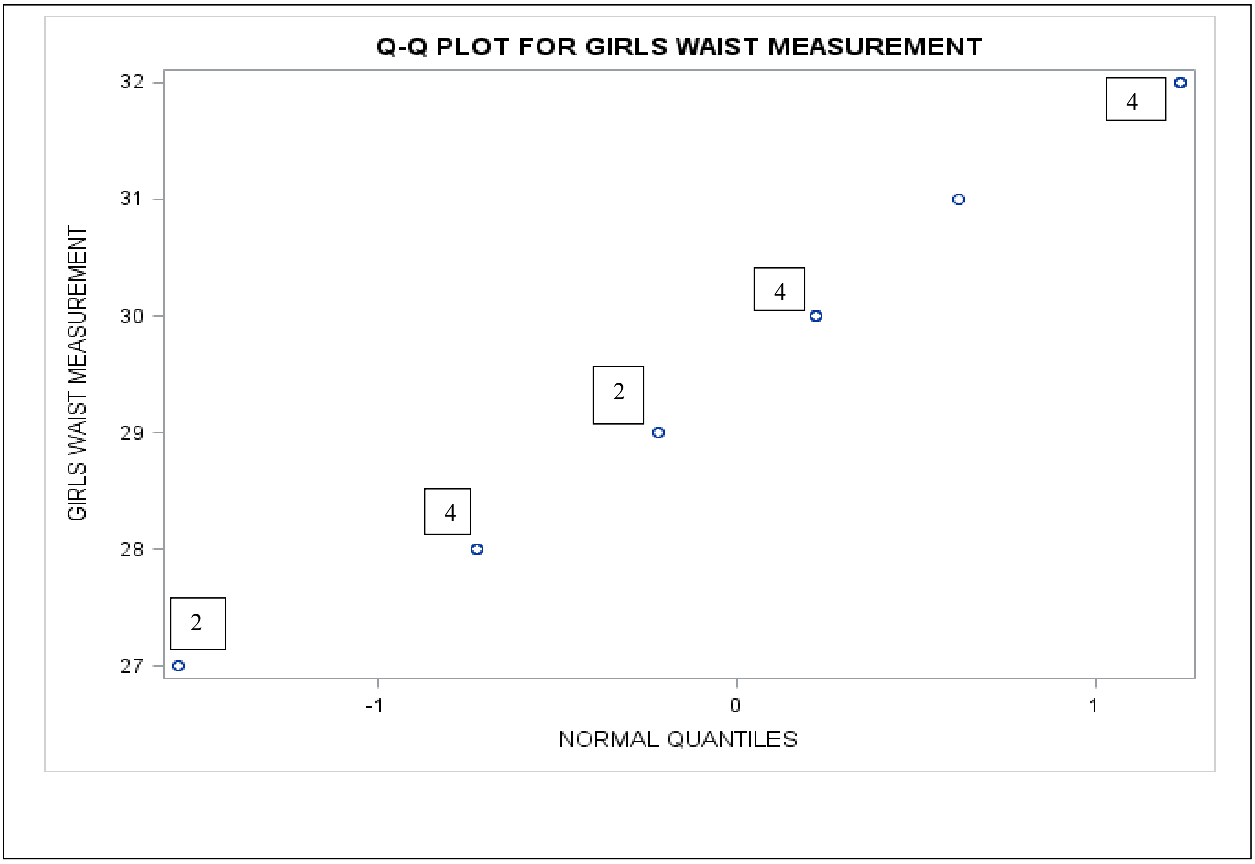

Figure 5. Q-Q plot for waist measurement of young adult females.

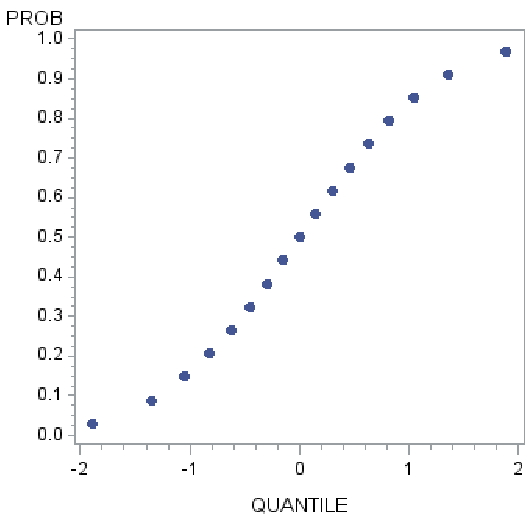

Figure 6 .probability plot for table 2 with $p_l = \frac{l-\alpha}{m}$.

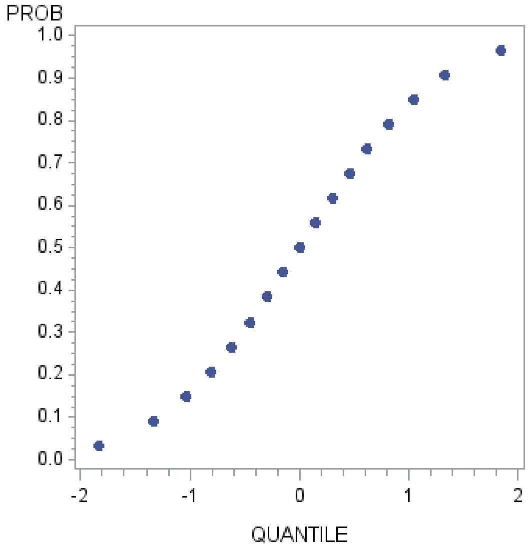

Figure 7. probability plot for table 2 with $\widetilde{p}_i = \frac{l-\beta}{m+\varepsilon}$.

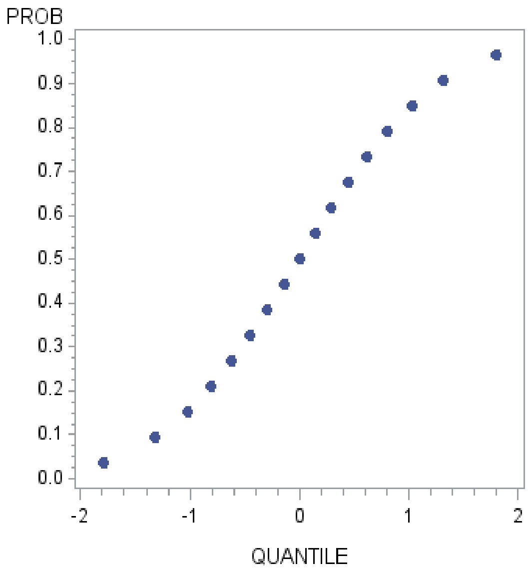

Figure 8. probability plot for table 2 with $\overline{p}_l = \frac{p_l+\widetilde{p}_l}{2}$.

Discusssion

The analysis based on the procedures applied showed that the data set obtained from the waist measurement of young adult males is not normally distributed. This conclusion is based on the fact that the computed value of the correlation coefficient for each of the procedures is less than the respective critical value at 10% level of significance. The contrary is the case for the waist measurement data obtained from young adult females at the same 10% level of significance. While we accepted that the data set for the females is normally distributed, however, we note that a sample size of $n = 17$ may not have been large enough to give detailed information required to determine normality.

On the health risk implication, the data for young adult males were considered not normally distributed and had a mean value that is within the range of low health risk. In the same vein, the data obtained from young adult females showed reasonable normality, and had a mean value that is within the range of low health risk.

The quantile - quantile plot clearly validated the conclusions about the normality of both groups. The plot of the probabilities against the quantiles has unique patterns. In general, high frequency of data points is observed in both measurement categories as shown in Fig.1 and Fig.5, respectively. The values in the square indicate the frequency of the data point.

Conclusion

This paper investigated the normality of data obtained from waist measurements of a group of young male and female adults. The quantile - quantile (Q-Q) plot and the analysis of correlation coefficients for the Q-Q plot was used to determine the normality or otherwise of the data set. Results indicated that the data for young adult males of the sample were not normally distributed, whereas the distribution of the data for young adult females showed reasonable normality. Both groups have mean values that are considered to be within the range of low health risk.

References

W. Blahd, *Calculating your waist circumference* (WebMD Medical Reference) WebMD LLC, 2016.

Q. Qiao and R. Nyamdorj, Is the association of type II diabetes with waist circumference or waist to hip ratio stronger than that with body mass index? *European Journal of Clinical nutrition, 64(1), 2009, 30 - 34.*

L.M. Browning, S.D. Hsieh, and M.A. Ashwell, A systematic review of waist to height ratio as a screening tool for the prediction of cardiovascular diseases and diabetes. *Nutr. Res. Rev. 23(2), 2010, 247 – 269.*

J.F. Hair, F.E. Anderson, R.L. Tatham, and W.C. Black, (1998). *Multivariate data analysis,* (Prentice Hall, New Jersey, 1998).

S. Ramzan, F.M. Zahid, and S. Ramzan, Evaluating multivariate normality: A graphical approach. *Middle -East Journal of Scientific Research.* 13, 2013, 254-263.

R.A Johnson, and D.W. Wichern, *Applied multivariate statistical analysis*, (Prentice -Hall International Inc. New Jersey, 2008).