PORTFOLIO PROJECT

Body Fat Estimation using Multi Regression & Factor Analysis in SPSS

Prepared by Afif Rifa’ie Bin Mohd Gazali

TABLE OF CONTENT

Multi Regression Analysis

Factor Analysis

(a) Problem Statement

Obesity has been on the rise all around the world. Obesity also leads to a variety of health conditions, including cardiovascular and noncommunicable diseases such as cancer and all-causes of death (mortality), high blood pressure, and Type 2 diabetes, stroke and coronary heart disease. (Tiwari & Balasundaram, 2023) In the long run, this will become a national issue because the government would have to increase its expenditures for the treatment of its citizens.

A precise body fat percentage evaluation is critical for understanding obesity and its implications, as well as understanding how to manage this disease. Relying on pricey and specialized machines that are only available in medical facilities is no longer the best solution. (Frankenfield et al., 2001)

Early interventions are better than treatment. A precise and quick assessment of bodyfat is critical in identifying, guiding, and tracking any individual with obesity issues. Overreliance on traditional methods prevent routine examination and resolution of the situation.

Thus, simple yet exact and dependable approaches for calculating fat percentage and its components such as Density, Age, Weight, Height, Neck, Chest, Abdomen, Hip, Thigh, Knee, Ankle, Biceps, Forearm, and Wrist are required.

(b) Introduction

Body fat measurement may be used to assess adiposity, such as BMI, and it is an important indicator of general health and fitness, providing more information than traditional weight assessment. (Akindele et al., 2016) Age, Weight, Height, Neck, Chest, Abdomen, Hip, Thigh, Knee, Ankle, Biceps, Forearm, and Wrist are all factors or independent variables that have a strong relationship with body fat.

A study conducted by the Malaysian Journal of Business to investigate the causes of obesity found that body fat is substantially influenced by physical activities and mental stress. (Ai et al., 2021) These studies, however, did not adequately investigate the independent variables that this study tries to address. In addition to that, a study (Pawan Kumar Jha1, Subhasis Mukherjee2, Purab Kalyan Modak3, Sharada Mayee Swain4, 2022) shows that height, weight, waist circumference, and hip circumference of an individual significantly predicted the outcome of body fat.

The purpose of this research is to apply multi regression analysis with a stepwise method to analyze body fat as a dependent variable and related independent variable such as Density, Age, Weight, Height, Neck, Chest, Abdomen, Hip, Thigh, Knee, Ankle, Biceps, Forearm, and Wrist with the main goal of developing a predictive model for body fat association and estimation.

(c) Research Objectives

Objective 1: To develop a Multiple Linear Regression model using stepwise regression to predict body fat percentage based on a set of metric independent variables.

Hypothesis 1.1: There is a significant relationship between the independent variables (Density, Age, Weight, Height, Neck, Chest, Abdomen, Hip, Thigh, Knee, Ankle, Biceps, Forearm, and Wrist) and body fat percentage.

Hypothesis 1.2: The MLR model, including the selected independent variables, will significantly predict body fat percentage.

Hypothesis 1.3: The stepwise regression method will result in a model that includes the most significant predictor variables for body fat percentage.

Objective 2: To assess the significance of individual predictor variables in the model and to identify which variables have the strongest impact on body fat.

Hypothesis 2.1: Each independent variable significantly contributes to the prediction of body fat percentage.

Hypothesis 2.2: Some variables will have a stronger impact on body fat percentage than others.

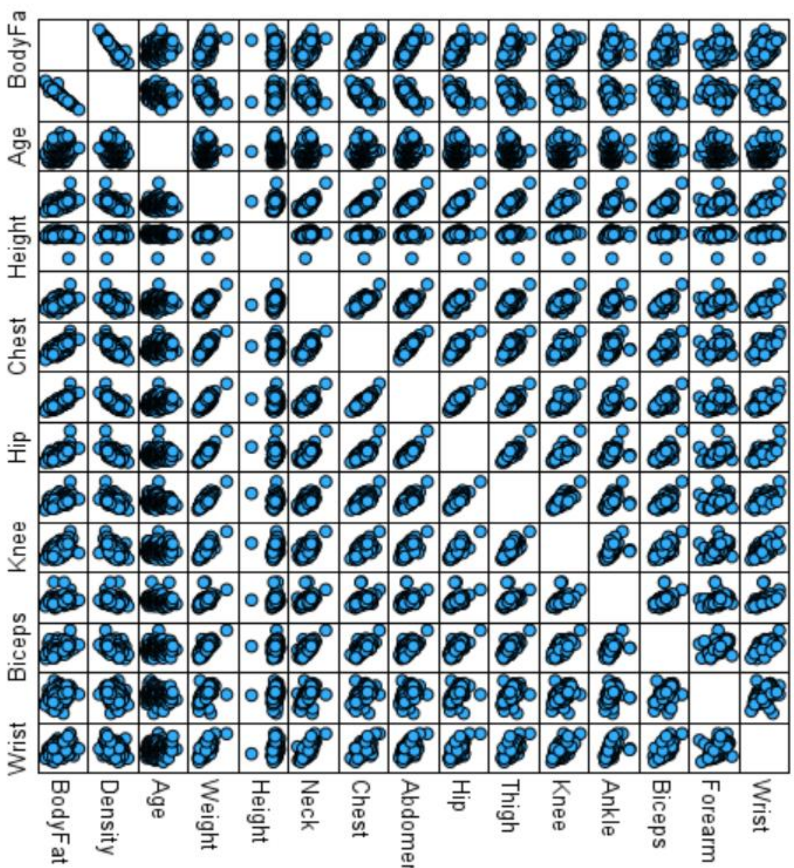
Hypothesis 2.3: The inclusion of specific variables (e.g., Abdomen, Weight) will significantly improve the predictive accuracy of the model.

[this page is intentionally left blank]

(d) Analysis Results and Interpretations

Assumption 1 – Linearity of the variables.

The first assumption is to use the scatter plot to examine the correlation between the independent and dependent variables. Based on the Scatter Plot 1 below, there is no linear relationship between body fat and age or height. The density indicates a negative association, but the other predictors show a positive association.



Scatter Plot 1: Linearity of Dependent and Independent Variables

The Pearson Correlation Table can also be used to check the linearity. Table 1 below demonstrates that the predictors have a positive linear relationship. However, there is a negative linear relationship between body fat to density and height. Furthermore, Heigh with a significance value of 0.078 greater than p-value 0.05 showed that Heigh is not statistically significant.

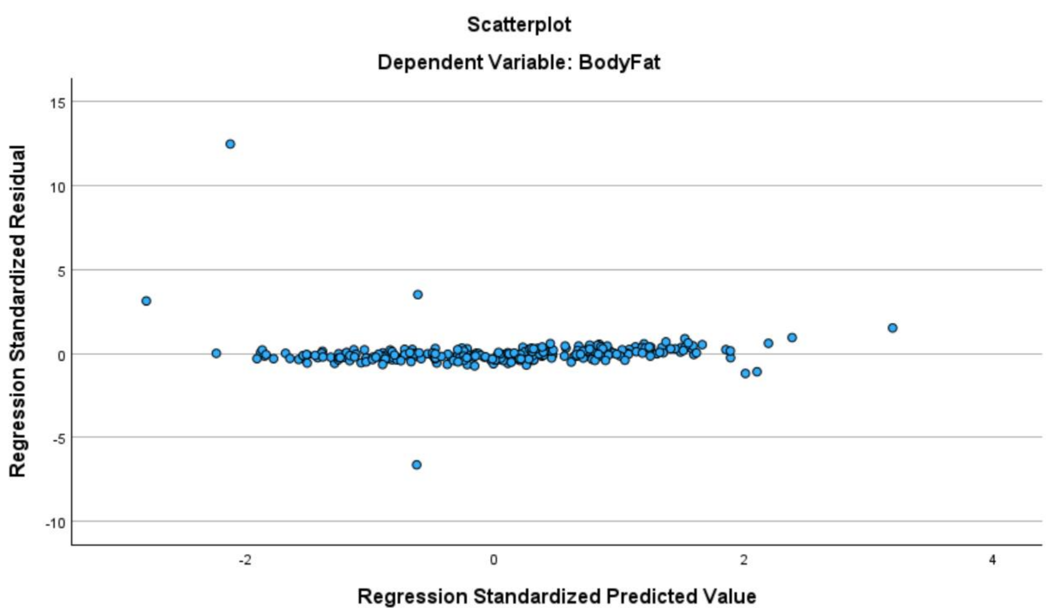
		BodyFat
Pearson Correlation	BodyFat	1.000
	Density	-.988
	Age	.291
	Weight	.612
	Height	-.089
	Neck	.491
	Chest	.703
	Abdomen	.813
	Hip	.625
	Thigh	.560
	Knee	.509
	Ankle	.266
	Biceps	.493
	Forearm	.361
	Wrist	.347
Sig. (1-tailed)	BodyFat	.
	Density	.000
	Age	.000
	Weight	.000
	Height	.078
	Neck	.000
	Chest	.000
	Abdomen	.000
	Hip	.000
	Thigh	.000
	Knee	.000
	Ankle	.000
	Biceps	.000
	Forearm	.000
	Wrist	.000

Table 1: snipped of correlation table

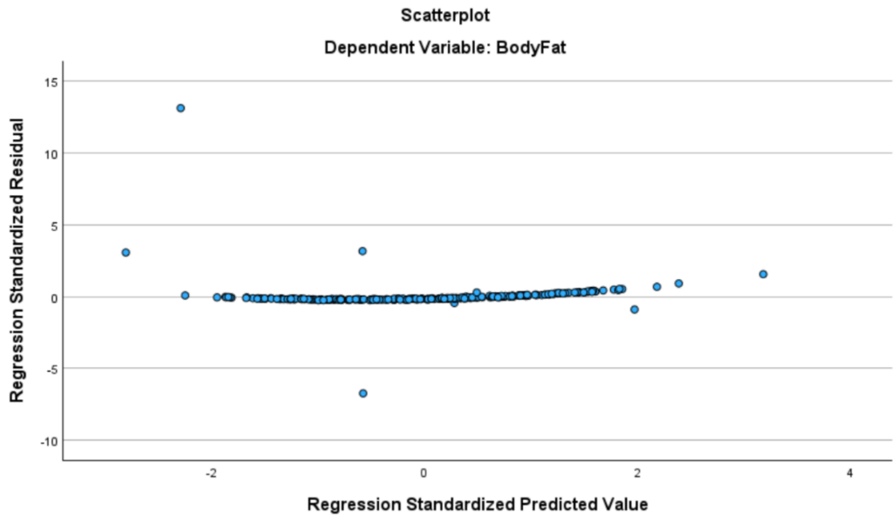
Based on the aforementioned results, the Age and Height variables have no linear correlation with the Body Fat variable, thus we exclude it from the analysis.

Assumption 2 – Constant Variance of Error Term

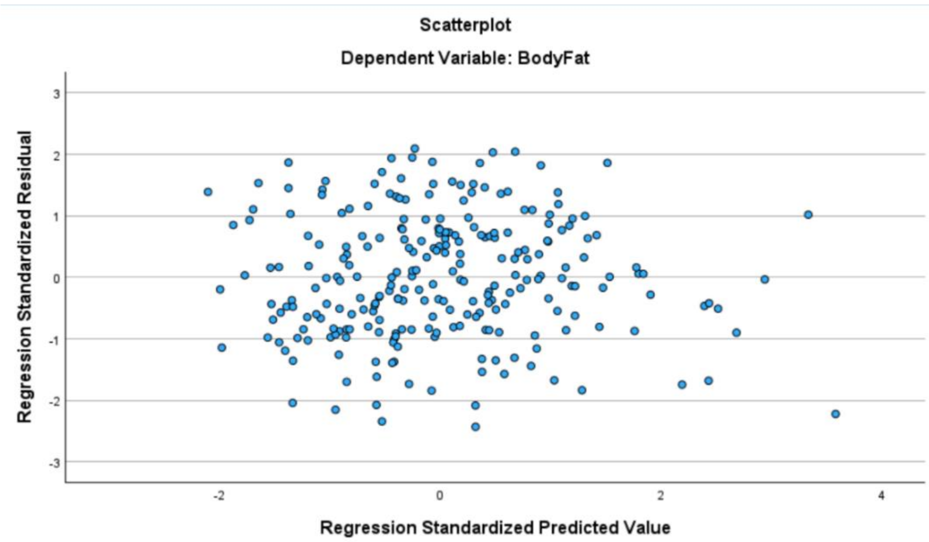
The second assumption is that the Constant Variance of Error Term, also known as homoscedasticity, is checked. This assumption implies that the variance of the errors, or the spread of the residuals, should be roughly constant for all predictor variable values. According to the scatter plot 2, the points do not fall inside the plus minus 3 range. As a result, it shows that there is a heteroscedasticity problem. We will then examine each predictor to see which independent variables contribute to the outcome. According to scatter plot 4, the error term has attained constant variance after removing the density variable. The points are now normally distributed and range between plus and minus 3. As a result, homoscedasticity is met.



Scatter Plot 2: Constant Variance of Error Term



Scatter Plot 3: Constant Variance of Error Term of Density Variable



Scatter Plot 4: Contant variance of Error Term after removal of Density Variable

Assumption 3 – Independence of error term

The third assumption is the independence of errors term. This assumption states that the error term for one observation is not correlated with the error term of any other observation. One common method to check for independence of errors is the Durbin-Watson test. Based on table 2 below, the Durbin-Watson value is 1.789 close to 2, which indicates that there is no significant autocorrelation. Hence, the error terms are independent of one another.

Model Summary[e]

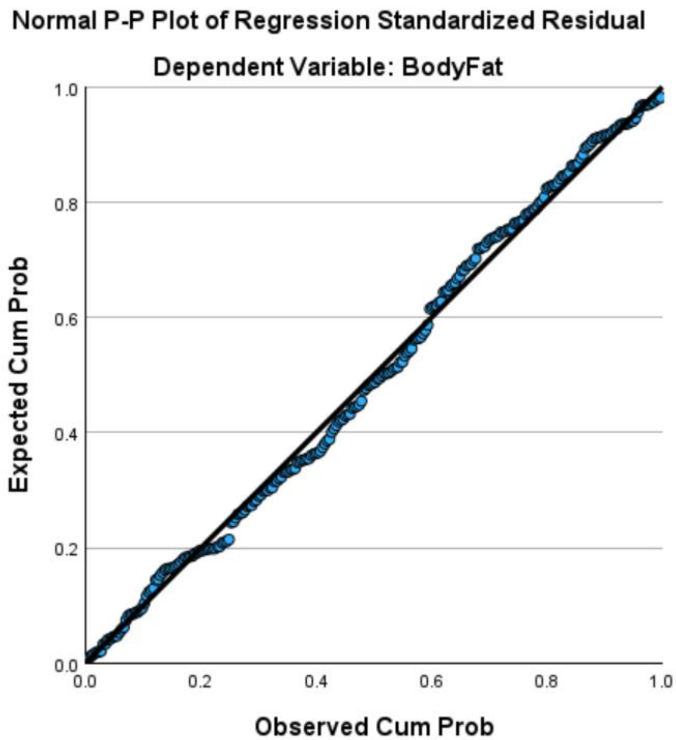
					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.813[a]	.662	.660	4.8775	.662	488.928	1	250	<.001	
2	.848[b]	.719	.717	4.4556	.057	50.584	1	249	<.001	
3	.853[c]	.728	.724	4.3930	.009	8.145	1	248	.005	
4	.857[d]	.735	.731	4.3427	.007	6.777	1	247	.010	1.798

a. Predictors: (Constant), Abdomen
b. Predictors: (Constant), Abdomen, Weight
c. Predictors: (Constant), Abdomen, Weight, Wrist
d. Predictors: (Constant), Abdomen, Weight, Wrist, Forearm
e. Dependent Variable: BodyFat

Table 2: Model Summary Table

Assumption 4 – Normality of error term

The fourth assumption is the normality of error terms. This assumption states that the errors (residuals) of the model are normally distributed. P-P plot and Normality test can be used to check for the normality of error term. The points drawn in P-P plot 1 fall along a straight diagonal line, indicating that residuals are regularly distributed. We utilize the Kolmogorov-Smirnov test for the Test of Normality because the sample size is more than 50. The significance value is 0.2, which is more than the P value of 0.05, hence fail to reject the null hypothesis and it indicates that the error terms are normally distributed.



P-P plot 1: Normality of Error Term

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	.042	252	.200*	.989	252	.061

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

a. Lilliefors Significance Correction

Table 3: Test of Normality

We can assume that the assumptions are met based on the Linearity, Constance Variance of Error Term, Independence of Error Term, and Normality of Error Term. Hence, we move on to the next step, analysis.

(d) Analysis Results and Interpretations 2

Based on the Model Summary Table, we may look at the R, R Square (R2), and R Square (R2) Change.

Model Summary[e]

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.813[a]	.662	.660	4.8775	.662	488.928	1	250	<.001	
2	.848[b]	.719	.717	4.4556	.057	50.584	1	249	<.001	
3	.853[c]	.728	.724	4.3930	.009	8.145	1	248	.005	
4	.857[d]	.735	.731	4.3427	.007	6.777	1	247	.010	1.798

a. Predictors: (Constant), Abdomen

b. Predictors: (Constant), Abdomen, Weight

c. Predictors: (Constant), Abdomen, Weight, Wrist

d. Predictors: (Constant), Abdomen, Weight, Wrist, Forearm

e. Dependent Variable: BodyFat

Table 4: Model Summary Table

R is the correlation coefficient, which measures the strength and direction of the linear relationship between the dependent variable and the independent variables. The final model shows 0.857 which close to 1, the predictors and Body Fat are positively associated.

R Square (R^2) or known as coefficient of determination, which measures the proportion of the variance in the dependent variable that can be explained by the independent variables. In the final model, it shows that 73.5% of the variation of Body fat, is explained by abdomen, weight, wrist and forearm.

R Square Change (R^2) is how much additional variance in the dependent variable is explained by adding a new set of independent variables to the model. In the final model, R square change increased from 0.660 to 0.735. The sig F change is 0.010 lower than P value 0.05, thus it is a significant change.

		BodyFat
Pearson Correlation	BodyFat	1.000
	Weight	.612
	Neck	.491
	Chest	.703
	Abdomen	.813
	Hip	.625
	Thigh	.560
	Knee	.509
	Ankle	.266
	Biceps	.493
	Forearm	.361
	Wrist	.347
Sig. (1-tailed)	BodyFat	.
	Weight	.000
	Neck	.000
	Chest	.000
	Abdomen	.000
	Hip	.000
	Thigh	.000
	Knee	.000
	Ankle	.000
	Biceps	.000
	Forearm	.000
	Wrist	.000

Table 5: Correlation Table

According to the Table 5, the significance values are 0.000, which is less than p-value 0.05, indicating that there is a statistically significant correlation between the dependent and independent variables.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	11631.527	1	11631.527	488.928	<.001[b]
	Residual	5947.463	250	23.790		
	Total	17578.990	251			
2	Regression	12635.745	2	6317.872	318.242	<.001[c]
	Residual	4943.245	249	19.852		
	Total	17578.990	251			
3	Regression	12792.936	3	4264.312	220.965	<.001[d]
	Residual	4786.054	248	19.299		
	Total	17578.990	251			
4	Regression	12920.754	4	3230.189	171.279	<.001[e]
	Residual	4658.236	247	18.859		
	Total	17578.990	251			

a. Dependent Variable: BodyFat

b. Predictors: (Constant), Abdomen

c. Predictors: (Constant), Abdomen, Weight

d. Predictors: (Constant), Abdomen, Weight, Wrist

e. Predictors: (Constant), Abdomen, Weight, Wrist, Forearm

Table 6: Correlation Table

The significance value of the final model in the ANOVA table is <0.001, which is less than p-value 0.05. It implies that at least one of the independent variables explains Body Fat well.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	-39.280	2.660		-14.765	<.001	-44.520	-34.041					
	Abdomen	.631	.029	.813	22.112	<.001	.575	.688	.813	.813	.813	1.000	1.000
2	(Constant)	-45.952	2.605		-17.640	<.001	-51.083	-40.822					
	Abdomen	.990	.057	1.275	17.447	<.001	.878	1.101	.813	.742	.586	.211	4.729
	Weight	-.148	.021	-.520	-7.112	<.001	-.189	-.107	.612	-.411	-.239	.211	4.729
3	(Constant)	-27.930	6.817		-4.097	<.001	-41.357	-14.503					
	Abdomen	.975	.056	1.256	17.368	<.001	.865	1.086	.813	.741	.575	.210	4.767
	Weight	-.114	.024	-.402	-4.841	<.001	-.161	-.068	.612	-.294	-.160	.159	6.281
	Wrist	-1.245	.436	-.139	-2.854	.005	-2.104	-.386	.347	-.178	-.095	.464	2.157
4	(Constant)	-34.854	7.245		-4.811	<.001	-49.124	-20.584					
	Abdomen	.996	.056	1.283	17.760	<.001	.885	1.106	.813	.749	.582	.206	4.864
	Weight	-.136	.025	-.476	-5.480	<.001	-.184	-.087	.612	-.329	-.180	.142	7.041
	Wrist	-1.506	.443	-.168	-3.401	<.001	-2.377	-.634	.347	-.212	-.111	.440	2.273
	Forearm	.473	.182	.114	2.603	.010	.115	.831	.361	.163	.085	.558	1.793

a. Dependent Variable: BodyFat

Table 7: Coefficient Table

Unstandardized B is a measure that is used to describe the change in the dependent variable for a one-unit change in an independent variable while holding all other variables constant. According to the model, increasing the abdomen by one unit increases body fat by 0.996. A one-unit increase in weight results in a 0.136-unit decrease in body fat. A one-unit increase in wrist reduces body fat by 1.506. Body fat increases by 0.473 units for every unit increase in forearm.

The 95% Confidence Interval for B coefficient indicates the range of values within which the true coefficient is most likely to reside with 95% confidence. Based on the table 7, we can conclude that; as 95% confidence interval does not include value 0, Hence, it can be concluded that abdomen, weight, wrist and forearm are significant independent variables.

The Variance Inflation Factor (VIF) analysis measures the degree of multicollinearity among independent variables. Based on the table 7, Multicollinearity does not appear to be a major issue for any of the four variables in the model. Specifically, the VIF for abdomen is 4.8, indicating a moderate level of multicollinearity, while the VIF for weight is slightly higher at 7.0, suggesting some potential multicollinearity. However, both are below the threshold of 10. The VIF values for wrist and forearm are even lower, at 2.2 and 1.7 respectively, indicating that multicollinearity is not a concern for these variables.

The Tolerance value is the degree of variance in an independence variable that cannot be explained by the other independent variables. Multicollinearity does not appear to be a significant concern for any of the four variables in the model based on the tolerance values, as all values abdomen 0.206, weight 0.142, wrist 0.440, forearm 0.558 are above the 0.1 threshold, indicating that the proportion of variance not explained by other variables is satisfactory.

We can conclude that the regression model is appropriate and significant based on these findings. The model explains a considerable percentage of the variance in Body Fat, and the predictors are significant but not significantly linked with one another.

(e) Conclusion and Recommendation

The goal of this study was to use a Multiple Linear Regression model with a stepwise method to measure body fat percentage and its associated independent variables. The goal of the study was to create a prediction model for body fat association and quantification based on a set of metric independent variables: density, age, weight, height, neck, chest, abdomen, hip, thigh, knee, ankle, biceps, forearm, and wrist.

The regression model was found to be significant, with a R Square value of 73.5%, implying that the four relevant predictors: abdomen, weight, wrist, and forearm, explain approximately 73.5% of the variation in body fat. The positive correlation coefficient (R) of 0.857 indicates a strong positive association between the predictor variables and body fat. The ANOVA test further confirmed the significance of the model.

Based on the output, the hypothesis in objective one and two are met.

Objective 1: The Multiple Linear Regression model was successfully developed using stepwise regression to predict body fat percentage with selected variables (Abdomen, Weight, Wrist, Forearm).

Hypothesis 1.1: Supported, as significant relationships were found between body fat percentage and the independent variables in the model.

Hypothesis 1.2: Supported, as the model significantly predicts body fat percentage with a significant F value and R^2 value of 73.5%.

Hypothesis 1.3: Supported, as stepwise regression resulted in a model that includes significant predictor variables for body fat percentage.

Objective 2: The significance and impact of individual predictor variables in the model were assessed.

Hypothesis 2.1: Partially supported, as not all original independent variables were included in the final model, indicating that not all significantly contribute to the prediction of body fat percentage.

Hypothesis 2.2: Supported, as standardized coefficients (Beta) indicate varying strengths of impact on body fat percentage.

Hypothesis 2.3: Supported, as the inclusion of specific variables like Abdomen and Weight significantly improved the predictive accuracy of the model.

The findings of this study have important implications for understanding and managing obesity. The prediction model can help healthcare practitioners and individuals make educated decisions about obesity prevention and treatment.

However, there are some limitations on this study. Secondary data were used in the study, which means they may not be 100% accurate or full and the model suggests that the predictor variables and the dependent variable are linked in a straight line, but this might not always be the case.

(f) Appendix and References

Ai, T. Y., Yee, T. S., Tieng, S. T. S., & Jie, Y. X. (2021). FACTORS THAT CAUSE THE PROBLEM OF OBESITY AMONG ADULTS IN TAMAN SENTOSA, KLANG. *Malaysian Journal of Business and Economics (MJBE)*, *8*(1), Article 1. https://doi.org/10.51200/mjbe.vi.3320

Akindele, M. O., Phillips, J. S., & Igumbor, E. U. (2016). The Relationship Between Body Fat Percentage and Body Mass Index in Overweight and Obese Individuals in an Urban African Setting. *Journal of Public Health in Africa*, *7*(1), 515. https://doi.org/10.4081/jphia.2016.515

Frankenfield, D. C., Rowe, W. A., Cooney, R. N., Smith, J. S., & Becker, D. (2001). Limits of body mass index to detect obesity and predict body composition. *Nutrition*, *17*(1), 26–30. https://doi.org/10.1016/S0899-9007(00)00471-8

Pawan Kumar Jha1, Subhasis Mukherjee2, Purab Kalyan Modak3, Sharada Mayee Swain4. (2022). Establishing and testing a multiple regression equation to predict body fat from height, weight, waist circumference, and hip circumference. *National Journal of Physiology, Pharmacy and Pharmacology*, *12*(05), 698.

Tiwari, A., & Balasundaram, P. (2023). Public Health Considerations Regarding Obesity. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK572122/

Factor Analysis

(a) In view of your problem statement in Question 1, elaborate briefly the purpose of performing factor analysis on your set of metric independent variables.

The goal of factor analysis is to perform data summarization and data reduction. Data summarization is the process of reducing the amount of data in a dataset so that it is easier to understand. This is done in factor analysis by finding the underlying latent factors that show how the measured variables are related to each other.

For example, Density, Body Fat, Age, Weight, Height, Neck, Chest, Abdomen, Hip, Thigh, Knee, Ankle, Biceps, Forearm, and Wrist are all metric factors that can be changed on their own. These elements might be able to be summed up into a few factors, like "Body Size," "Body Proportion," and "Limb Size."

Data reduction means keeping as much information as possible while cutting down on the number of factors in the dataset. This is done by finding a smaller group of factors that account for most of the variation in the original variables. If factor analysis extracts three factors: "Body Size," "Body Proportion," and "Limb Size," we can reduce the fourteen original variables to these three factors. These factors capture the essence of the original variables and can be used for further analysis, such as predictive modeling. For example, to predict body fat percentage, we could use the three factors that were found through factor analysis instead of all fourteen variables.

(b) Discuss briefly why the non-metric independent variable was excluded from the factor analysis.

The basic in designing the Factor Analysis is to have the metric variables. Furthermore, Factor Analysis uses Pearson correlation to assess the similarity between each pair of variables, and Pearson correlation is a numerical value that can only be calculated using metric variables, while non-metric independent variables were not included in the analysis. The correlation matrix with metric variables is shown in Table 1 below.

Correlation Matrix

		Density	BodyFat	Age	Weight	Height	Neck	Chest	Abdomen	Hip	Thigh	Knee	Ankle	Biceps	Forearm	Wrist
Correlation	Density	1.000	-.988	-.278	-.594	.098	-.473	-.683	-.799	-.609	-.553	-.495	-.265	-.487	-.352	-.326
	BodyFat	-.988	1.000	.291	.612	-.089	.491	.703	.813	.625	.560	.509	.266	.493	.361	.347
	Age	-.278	.291	1.000	-.013	-.172	.114	.176	.230	-.050	-.200	.018	-.105	-.041	-.085	.214
	Weight	-.594	.612	-.013	1.000	.308	.831	.894	.888	.941	.869	.853	.614	.800	.630	.730
	Height	.098	-.089	-.172	.308	1.000	.254	.135	.088	.170	.148	.286	.265	.208	.229	.322
	Neck	-.473	.491	.114	.831	.254	1.000	.785	.754	.735	.696	.672	.478	.731	.624	.745
	Chest	-.683	.703	.176	.894	.135	.785	1.000	.916	.829	.730	.719	.483	.728	.580	.660
	Abdomen	-.799	.813	.230	.888	.088	.754	.916	1.000	.874	.767	.737	.453	.685	.503	.620
	Hip	-.609	.625	-.050	.941	.170	.735	.829	.874	1.000	.896	.823	.558	.739	.545	.630
	Thigh	-.553	.560	-.200	.869	.148	.696	.730	.767	.896	1.000	.799	.540	.761	.567	.559
	Knee	-.495	.509	.018	.853	.286	.672	.719	.737	.823	.799	1.000	.612	.679	.556	.665
	Ankle	-.265	.266	-.105	.614	.265	.478	.483	.453	.558	.540	.612	1.000	.485	.419	.566
	Biceps	-.487	.493	-.041	.800	.208	.731	.728	.685	.739	.761	.679	.485	1.000	.678	.632
	Forearm	-.352	.361	-.085	.630	.229	.624	.580	.503	.545	.567	.556	.419	.678	1.000	.586
	Wrist	-.326	.347	.214	.730	.322	.745	.660	.620	.630	.559	.665	.566	.632	.586	1.000

Table 1 Correlation Table

(c) Perform factor analysis on your metric independent variables by using SAS or SPSS. Present and label all the SPSS outputs for ease of reference.

Correlation Matrix

		Density	BodyFat	Age	Weight	Height	Neck	Chest	Abdomen	Hip	Thigh	Knee	Ankle	Biceps	Forearm	Wrist
Correlation	Density	1.000	-.988	-.278	-.594	.098	-.473	-.683	-.799	-.609	-.553	-.495	-.265	-.487	-.352	-.326
	BodyFat	-.988	1.000	.291	.612	-.089	.491	.703	.813	.625	.560	.509	.266	.493	.361	.347
	Age	-.278	.291	1.000	-.013	-.172	.114	.176	.230	-.050	-.200	.018	-.105	-.041	-.085	.214
	Weight	-.594	.612	-.013	1.000	.308	.831	.894	.888	.941	.869	.853	.614	.800	.630	.730
	Height	.098	-.089	-.172	.308	1.000	.254	.135	.088	.170	.148	.286	.265	.208	.229	.322
	Neck	-.473	.491	.114	.831	.254	1.000	.785	.754	.735	.696	.672	.478	.731	.624	.745
	Chest	-.683	.703	.176	.894	.135	.785	1.000	.916	.829	.730	.719	.483	.728	.580	.660
	Abdomen	-.799	.813	.230	.888	.088	.754	.916	1.000	.874	.767	.737	.453	.685	.503	.620
	Hip	-.609	.625	-.050	.941	.170	.735	.829	.874	1.000	.896	.823	.558	.739	.545	.630
	Thigh	-.553	.560	-.200	.869	.148	.696	.730	.767	.896	1.000	.799	.540	.761	.567	.559
	Knee	-.495	.509	.018	.853	.286	.672	.719	.737	.823	.799	1.000	.612	.679	.556	.665
	Ankle	-.265	.266	-.105	.614	.265	.478	.483	.453	.558	.540	.612	1.000	.485	.419	.566
	Biceps	-.487	.493	-.041	.800	.208	.731	.728	.685	.739	.761	.679	.485	1.000	.678	.632
	Forearm	-.352	.361	-.085	.630	.229	.624	.580	.503	.545	.567	.556	.419	.678	1.000	.586
	Wrist	-.326	.347	.214	.730	.322	.745	.660	.620	.630	.559	.665	.566	.632	.586	1.000

Table 2 Correlation Matrix

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.906
Bartlett's Test of Sphericity	Approx. Chi-Square	4981.034
	df	105
	Sig.	<.001

Table 3 KMO and Bartlett's Test

Communalities

	Initial	Extraction
Density	1.000	.879
BodyFat	1.000	.891
Age	1.000	.927
Weight	1.000	.953
Height	1.000	.548
Neck	1.000	.773
Chest	1.000	.861
Abdomen	1.000	.929
Hip	1.000	.890
Thigh	1.000	.887
Knee	1.000	.765
Ankle	1.000	.510
Biceps	1.000	.724
Forearm	1.000	.536
Wrist	1.000	.823

Extraction Method: Principal Component Analysis.

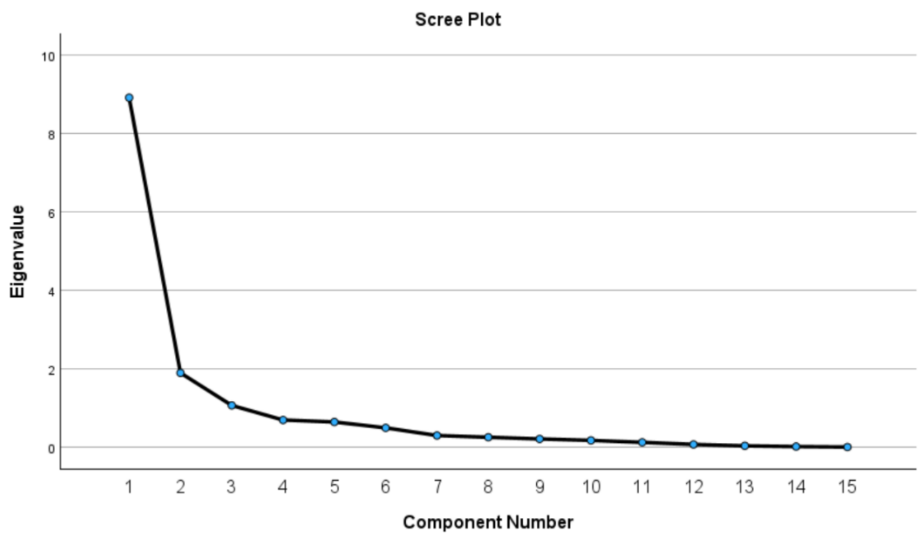
Table 4 Communalities Table

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.921	59.473	59.473	8.921	59.473	59.473	5.829	38.859	38.859
2	1.903	12.683	72.157	1.903	12.683	72.157	4.788	31.922	70.782
3	1.073	7.154	79.311	1.073	7.154	79.311	1.279	8.529	79.311
4	.701	4.671	83.981						
5	.650	4.336	88.317						
6	.501	3.340	91.657						
7	.306	2.037	93.694						
8	.262	1.744	95.438						
9	.218	1.451	96.888						
10	.181	1.206	98.095						
11	.132	.877	98.971						
12	.077	.513	99.485						
13	.042	.283	99.768						
14	.023	.156	99.924						
15	.011	.076	100.000						

Extraction Method: Principal Component Analysis.

Table 5 Total Variance Explained Table



Scree Plot 1

Component Matrix[a]

	Component 1	Component 2	Component 3
Density	-.705	.598	.158
BodyFat	.719	-.595	-.138
Age	.078	-.632	.722
Weight	.969	.115	-.026
Height	.233	.614	.343
Neck	.851	.098	.198
Chest	.915	-.148	.049
Abdomen	.923	-.278	-.009
Hip	.925	.036	-.181
Thigh	.874	.131	-.326
Knee	.860	.158	-.001
Ankle	.622	.348	.037
Biceps	.836	.150	-.053
Forearm	.687	.250	.037
Wrist	.759	.207	.452

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Table 6 Component Matrix Table

Rotated Component Matrix[a]

	Component		
	1	2	3
Density	-.144	-.914	-.152
BodyFat	.163	.914	.168
Age	-.056	.202	.940
Weight	.789	.568	-.092
Height	.636	-.380	-.011
Neck	.765	.419	.113
Chest	.625	.677	.105
Abdomen	.539	.790	.120
Hip	.660	.648	-.186
Thigh	.626	.606	-.358
Knee	.740	.458	-.090
Ankle	.679	.164	-.148
Biceps	.699	.467	-.131
Forearm	.673	.270	-.101
Wrist	.841	.193	.280

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Table 7 Rotated Component Matrix Table

Component Transformation Matrix

Component	1	2	3
1	.757	.654	-.012
2	.563	-.660	-.497
3	.333	-.370	.867

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Table 8 Component Transformation Matrix Table

(d) By using relevant SPSS outputs in part (c), group your metric independent variables into factors.

To retain the number of factors, the eigenvalue must be higher than 1. As per Table 5 below, the total value that is higher than 1 is component 1, component 2, and component 3. Hence, 3 factor is retained.

In addition to that, number of factors to retain can be look at percentage of variance. Factor 1 explains 59.5% of the variation in the 15 items. Factor 2 explains 12.7 % of the variation in the 15 items. Factor 3 explains 7.2 % of the variation in the 15 items.

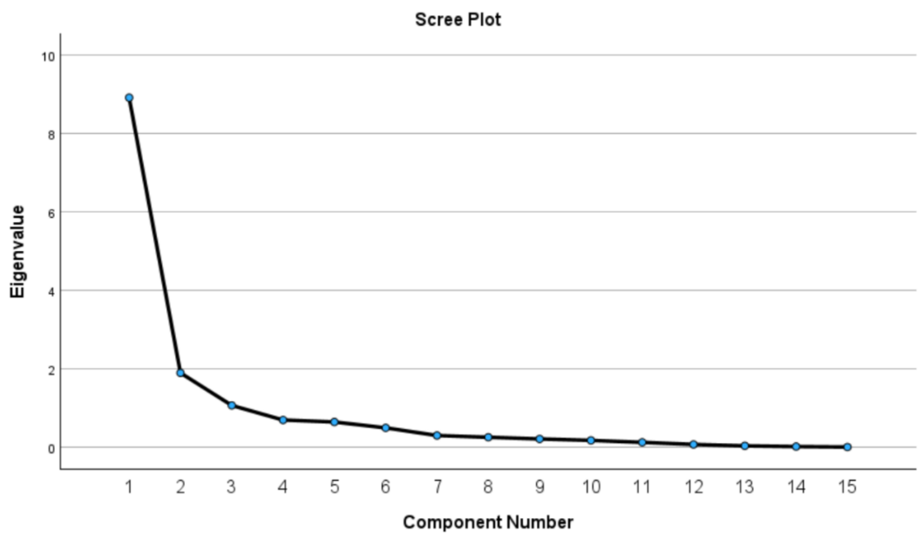
Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.921	59.473	59.473	8.921	59.473	59.473	5.829	38.859	38.859
2	1.903	12.683	72.157	1.903	12.683	72.157	4.788	31.922	70.782
3	1.073	7.154	79.311	1.073	7.154	79.311	1.279	8.529	79.311
4	.701	4.671	83.981						
5	.650	4.336	88.317						
6	.501	3.340	91.657						
7	.306	2.037	93.694						
8	.262	1.744	95.438						
9	.218	1.451	96.888						
10	.181	1.206	98.095						
11	.132	.877	98.971						
12	.077	.513	99.485						
13	.042	.283	99.768						
14	.023	.156	99.924						
15	.011	.076	100.000						

Extraction Method: Principal Component Analysis.

Table 5 Total Variance Explained Table

The Scree Test Criterion, scree plot 1 shows a scree line starts to flattening after third factor, it suggests that the first three factors explain a significant amount of variance in the items, and the remaining factors explain only a small additional amount of variance. Therefore, we would retain the first three factors.



Scree Plot 1

To know which variables has been assigned to which factor can be looked at the component Matrix. The values in the Component Matrix table can range from -1 to 1, where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no correlation. The variable that has higher value will be retained in that particular factor.

Component Matrix[a]

	Component		
	1	2	3
Density	-.705	.598	.158
BodyFat	.719	-.595	-.138
Age	.078	-.632	.722
Weight	.969	.115	-.026
Height	.233	.614	.343
Neck	.851	.098	.198
Chest	.915	-.148	.049
Abdomen	.923	-.278	-.009
Hip	.925	.036	-.181
Thigh	.874	.131	-.326
Knee	.860	.158	-.001
Ankle	.622	.348	.037
Biceps	.836	.150	-.053
Forearm	.687	.250	.037
Wrist	.759	.207	.452

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

Table 6 Component Matrix Table

(e) (i) Interpret the meaning of communality.

Communality represents the proportion of a variable's total variance that is accounted for by the common factors and the communality values range from 0 to 1.

Based Table 4 Communalities Table, all items indicate a communality value close to 1. Which indicates that most of the variance in the item is well-represented by the factors.

Communalities

	Initial	Extraction
Density	1.000	.879
BodyFat	1.000	.891
Age	1.000	.927
Weight	1.000	.953
Height	1.000	.548
Neck	1.000	.773
Chest	1.000	.861
Abdomen	1.000	.929
Hip	1.000	.890
Thigh	1.000	.887
Knee	1.000	.765
Ankle	1.000	.510
Biceps	1.000	.724
Forearm	1.000	.536
Wrist	1.000	.823

Extraction Method: Principal Component Analysis.

Table 4 Communalities Table

(e) (ii) Interpret the meaning of eigenvalue.

Eigenvalue is a measure of the variance in the item that is accounted for by a factor. It represents the total variance explained by the factor. Eigenvalues are used to determine the number of factors to retain and factors with eigenvalues greater than 1 are to be retained.

Based on the table 5 below, the total Eigen value that is higher than 1 is component 1, component 2, and component 3. Hence, 3 factor is retained.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.921	59.473	59.473	8.921	59.473	59.473	5.829	38.859	38.859
2	1.903	12.683	72.157	1.903	12.683	72.157	4.788	31.922	70.782
3	1.073	7.154	79.311	1.073	7.154	79.311	1.279	8.529	79.311
4	.701	4.671	83.981						
5	.650	4.336	88.317						
6	.501	3.340	91.657						
7	.306	2.037	93.694						
8	.262	1.744	95.438						
9	.218	1.451	96.888						
10	.181	1.206	98.095						
11	.132	.877	98.971						
12	.077	.513	99.485						
13	.042	.283	99.768						
14	.023	.156	99.924						
15	.011	.076	100.000						

Extraction Method: Principal Component Analysis.

Table 5 Total Variance Explained Table

(f)(i) Explain briefly how the factorability of your dataset can be improved prior to the task of factor analysis.

Factorability can be improved by checking the Multicollinearity. High multicollinearity among variables can improve factorability and it can be checked by examining the correlation matrix. Pairs of variables with correlations above 0.3 indicate potential shared variance that can be captured by factors.

Correlation Matrix

		Density	BodyFat	Age	Weight	Height	Neck	Chest	Abdomen	Hip	Thigh	Knee	Ankle	Biceps	Forearm	Wrist
Correlation	Density	1.000	-.988	-.278	-.594	.098	-.473	-.683	-.799	-.609	-.553	-.495	-.265	-.487	-.352	-.326
	BodyFat	-.988	1.000	.291	.612	-.089	.491	.703	.813	.625	.560	.509	.266	.493	.361	.347
	Age	-.278	.291	1.000	-.013	-.172	.114	.176	.230	-.050	-.200	.018	-.105	-.041	-.085	.214
	Weight	-.594	.612	-.013	1.000	.308	.831	.894	.888	.941	.869	.853	.614	.800	.630	.730
	Height	.098	-.089	-.172	.308	1.000	.254	.135	.088	.170	.148	.286	.265	.208	.229	.322
	Neck	-.473	.491	.114	.831	.254	1.000	.785	.754	.735	.696	.672	.478	.731	.624	.745
	Chest	-.683	.703	.176	.894	.135	.785	1.000	.916	.829	.730	.719	.483	.728	.580	.660
	Abdomen	-.799	.813	.230	.888	.088	.754	.916	1.000	.874	.767	.737	.453	.685	.503	.620
	Hip	-.609	.625	-.050	.941	.170	.735	.829	.874	1.000	.896	.823	.558	.739	.545	.630
	Thigh	-.553	.560	-.200	.869	.148	.696	.730	.767	.896	1.000	.799	.540	.761	.567	.559
	Knee	-.495	.509	.018	.853	.286	.672	.719	.737	.823	.799	1.000	.612	.679	.556	.665
	Ankle	-.265	.266	-.105	.614	.265	.478	.483	.453	.558	.540	.612	1.000	.485	.419	.566
	Biceps	-.487	.493	-.041	.800	.208	.731	.728	.685	.739	.761	.679	.485	1.000	.678	.632
	Forearm	-.352	.361	-.085	.630	.229	.624	.580	.503	.545	.567	.556	.419	.678	1.000	.586
	Wrist	-.326	.347	.214	.730	.322	.745	.660	.620	.630	.559	.665	.566	.632	.586	1.000

Table 2 Correlation Matrix

In addition to that, we can use Kaiser-Meyer-Olkin (KMO) Measure. The KMO measure is a statistic that indicates the proportion of variance among variables that might be common variance. A KMO value close to 1 indicates that factor analysis is likely to be useful, while a value below 0.5 indicates the opposite. Table 3 below shows the KMO of 0.906 which closer to 1.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.906
Bartlett's Test of Sphericity	Approx. Chi-Square	4981.034
	df	105
	Sig.	<.001

<u>Table 3 KMO and Bartlett's Test</u>

(f)(ii) Explain what is meant by factor cross-loading.

Cross-loading refers to the situation where a variable has significant loadings on two or more factors. This means that the variable is associated with more than one underlying factor. In the output table 7 below. Chest, Hip and thigh show cross-loading when the variable has significant loadings in factor 1 and factor 2.

Rotated Component Matrix[a]

	Component 1	2	3
Density	-.144	-.914	-.152
BodyFat	.163	.914	.168
Age	-.056	.202	.940
Weight	.789	.568	-.092
Height	.636	-.380	-.011
Neck	.765	.419	.113
Chest	.625	.677	.105
Abdomen	.539	.790	.120
Hip	.660	.648	-.186
Thigh	.626	.606	-.358
Knee	.740	.458	-.090
Ankle	.679	.164	-.148
Biceps	.699	.467	-.131
Forearm	.673	.270	-.101
Wrist	.841	.193	.280

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Table 7 Rotated Component Matrix Table

(f)(iii) Give suggestions on how the problem of cross-loading can be reduced.

The problem of cross-loading can be reduced by increase the sample size. Cross-loadings happens due to sample size issues. Increasing the sample size can provide more stable and reliable estimates of factor loadings.

Other suggestions are to implement rotation method such as Orthogonal or Oblique rotation. Rotation can make it easier to interpret and understand the factors. This rotation method is to see the clearer factor structure with fewer cross-loadings. If the cross loading still happening, it is best to drop the variable and re-run the analysis.