HUMBER INSTITUTE OF TECHNOLOGY
AND ADVANCED LEARNING
(HUMBER COLLEGE)

ASSIGNMENT: **Individual Assignment 1**

Submitted by:

Last Name	First Name	Student Number
Sampson	Dwight	N0********

Grade/Comments

Submitted to:
Professor Nataliya Korchagina

Submission Date:

Data Cleaning

The first step taken in analyzing the dataset is to clean it by looking for errors that may be present. The table below was used to assess the minimum and maximum values of each categorical variable to assess whether impossible values exist in the dataset.

Statistics

		Gender	I am able to recognise my own strengths and weaknesses, needs and motives	I don't understand what other people think about me (R)	I recognise the effect of my own cultural background on my thinking and behaviour	I acknowledge my personal values and beliefs	I don't set goals for personal development (R)	I undertake activities to enhance my skills and competencies	I want to know more than is required for task accomplishment	I have no confidence in my own capabilities (R)	I have a positive approach toward work-related challenges	I am satisfied with the success I have acheived in my career	I am satisfied with the progress I have made toward meeting my overall career goals	I am satisfied with the progress I have made toward meeting my goals for income	I am satisfied with the progress I have made toward meeting my goals for advancement	I am satisfied with the progress I have made toward meeting my goals for the development of new skills
N	Valid	195	195	195	195	195	195	195	195	195	195	195	195	195	195	195
	Missing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Minimum		0	3	1	3	3	1	1	1	1	2	1	1	1	1	1
Maximum		1	7	7	7	7	6	7	7	5	7	7	6	6	6	14

The only categorial variable that contains a value outside the range of possible values each variable can hold is CS5. This variable represents the degree participants agree with the statement, “I am satisfied with the progress I have made towards meeting my goals for the development of new skills”. This variable (CS5) contains the value 14 which is an impossibility as the scale can only contain numbers between 1 and 7.

The next step in the cleaning process would be to find how many cases in CS5 have a value over 7.

I am satisfied with the progress I have made toward meeting my goals for the development of new skills

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	3	1.5	1.5	1.5
	2	4	2.1	2.1	3.6
	3	17	8.7	8.7	12.3
	4	36	18.5	18.5	30.8
	5	85	43.6	43.6	74.4
	6	49	25.1	25.1	99.5
	14	1	.5	.5	100.0
	Total	195	100.0	100.0	

Since there is only one value over 7 and we do not have access to the original questionnaire, the value of 14 (at ID 16) can be deleted from variable CS5.

After rechecking the range of all variable, all the data appears to be clean.

Statistics

		Gender	I am able to recognise my own strengths and weaknesses, needs and motives	I don't understand what other people think about me (R)	I recognise the effect of my own cultural background on my thinking and behaviour	I acknowledge my personal values and beliefs	I don't set goals for personal development (R)	I undertake activities to enhance my skills and competencies	I want to know more than is required for task accomplishment	I have no confidence in my own capabilities (R)	I have a positive approach toward work-related challenges	I am satisfied with the success I have acheived in my career	I am satisfied with the progress I have made toward meeting my overall career goals	I am satisfied with the progress I have made toward meeting my goals for income	I am satisfied with the progress I have made toward meeting my goals for advancement	I am satisfied with the progress I have made toward meeting my goals for the development of new skills
N	Valid	195	195	195	195	195	195	195	195	195	195	195	195	195	195	194
	Missing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Minimum		0	3	1	3	3	1	1	1	1	2	1	1	1	1	1
Maximum		1	7	7	7	7	6	7	7	5	7	7	6	6	6	6

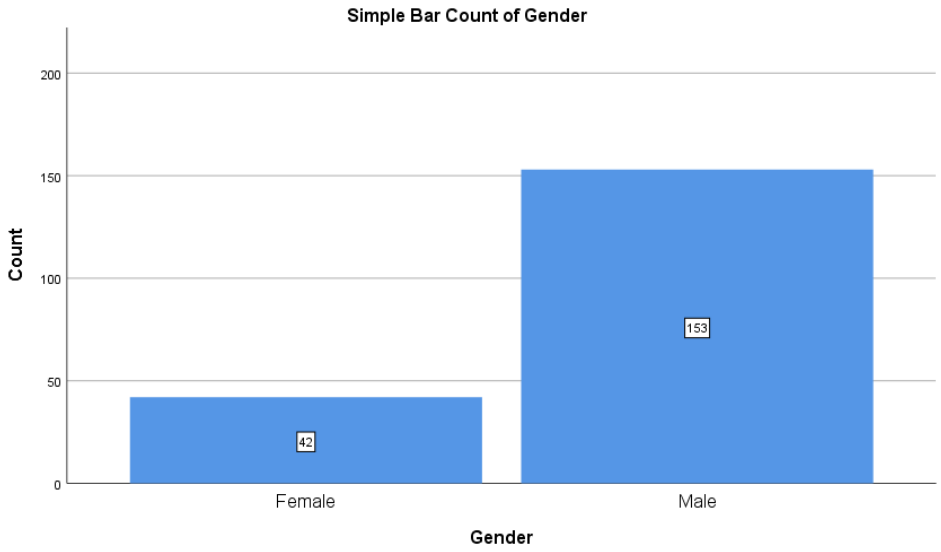
After cleaning categorical variables, it is necessary to clean continuous variables. The continuous variables in this dataset are age, years since graduation and number of promotions.

Descriptive Statistics

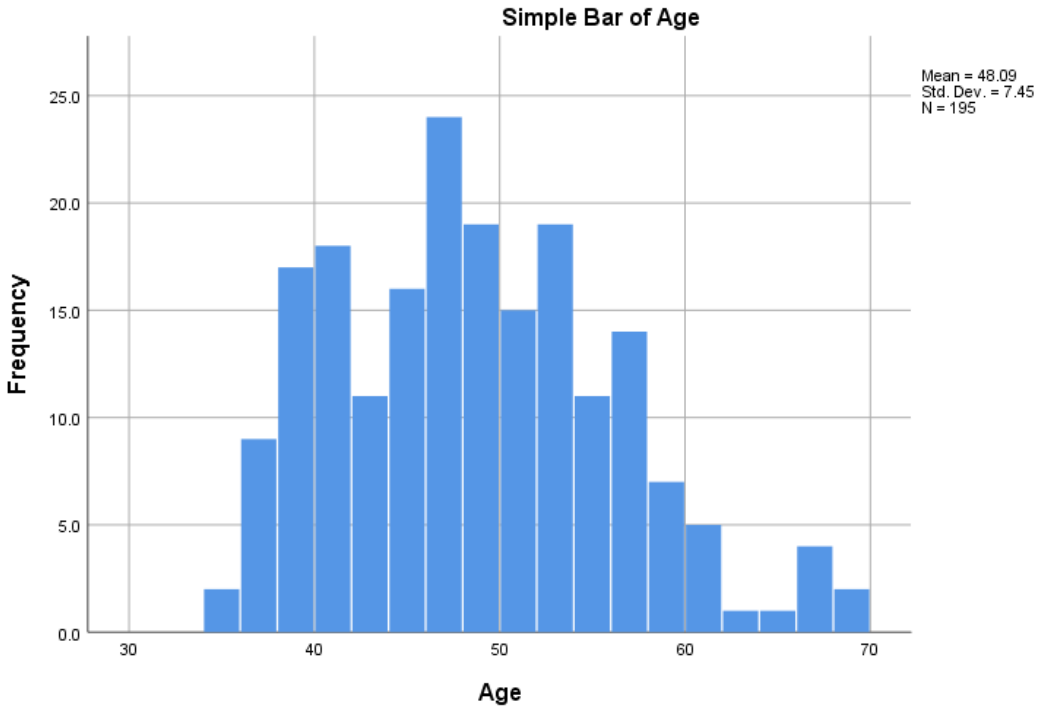
	N	Minimum	Maximum	Mean	Std. Deviation
Age	195	35	69	48.09	7.450
Years since graduation	193	1	31	9.82	6.503
Number promotions	189	0	7	1.70	1.640
Valid N (listwise)	187				

As seen in the table above, there are a few missing values in the variables: years since graduation and number of promotions. Otherwise, there are no quantifiable concerns for these variables as the range of values they can take on plausible. Though we should ask ourselves why there are only responses from people of age 35+ in the dataset.

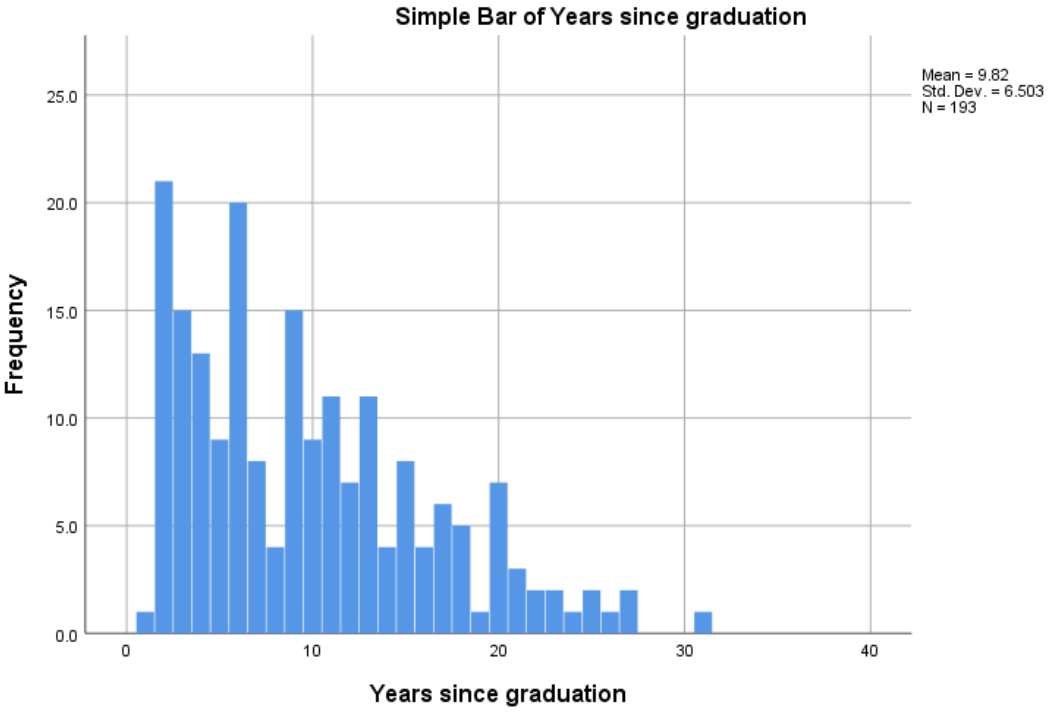
Exploratory Data Analysis



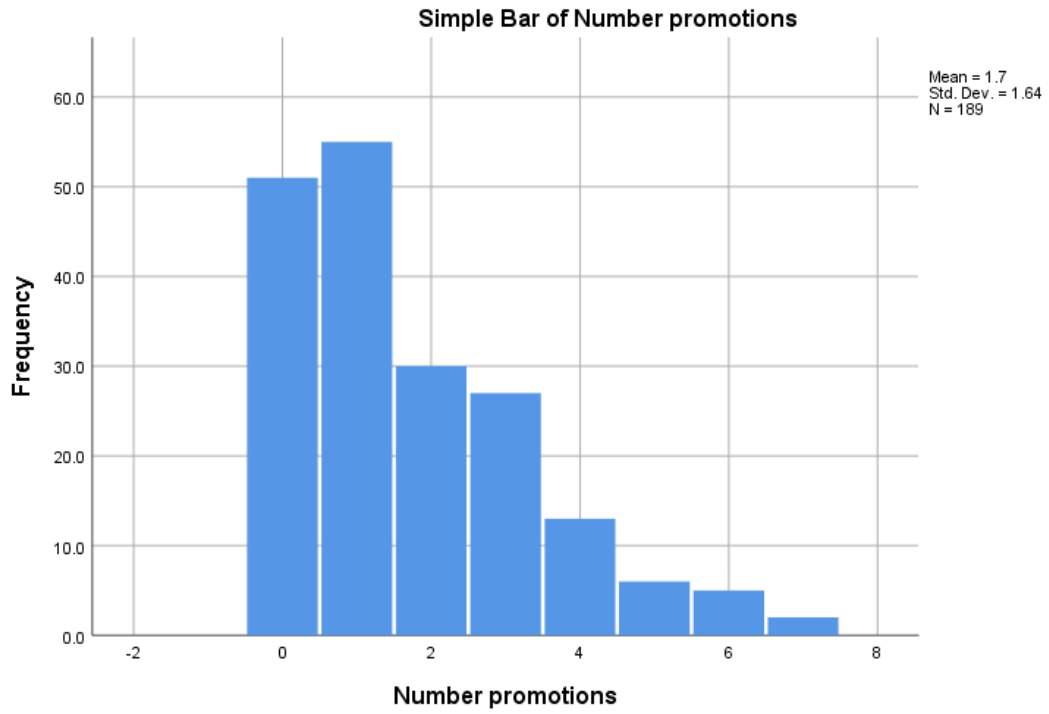
There are more males in the sample than females, almost 4 times as many.



Many of the participants in the survey are middle aged or nearing to retirement.



Many participants have recently graduated within the last 10 years.



The majority of MBA graduates seem to have very few promotions after graduating from the MBA program. Some of the positive skew in the graph may be attributed to large amount of recent graduates in the dataset.

Analyzing Scales

Career Satisfaction Scale:

To assess career satisfaction, a new variable that accounts for all existing the career satisfaction variables must be created. In this case all Career Satisfaction variable will be average to create the new variable Average Career Satisfaction (AVG_CS).

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
average Career Satisfaction (avg of CS1 to CS5)	195	100.0%	0	0.0%	195	100.0%

From the table above, it is evident that we have used all cases in the Average Career Satisfaction variable.

Descriptives

			Statistic	Std. Error
average Career Satisfaction (avg of CS1 to CS5)	Mean		4.5090	.07968
	95% Confidence Interval for Mean	Lower Bound	4.3518	
		Upper Bound	4.6661	
	5% Trimmed Mean		4.5749	
	Median		4.8000	
	Variance		1.238	
	Std. Deviation		1.11264	
	Minimum		1.00	
	Maximum		6.00	
	Range		5.00	
	Interquartile Range		1.20	
	Skewness		-.843	.174
	Kurtosis		.240	.346

On average participants are neutral about their career satisfaction, this evident as the mean score is 4.5. At first look, there is evidence of a normal distribution which can be seen in the difference between the mean and the 5% trimmed mean being arbitrary.

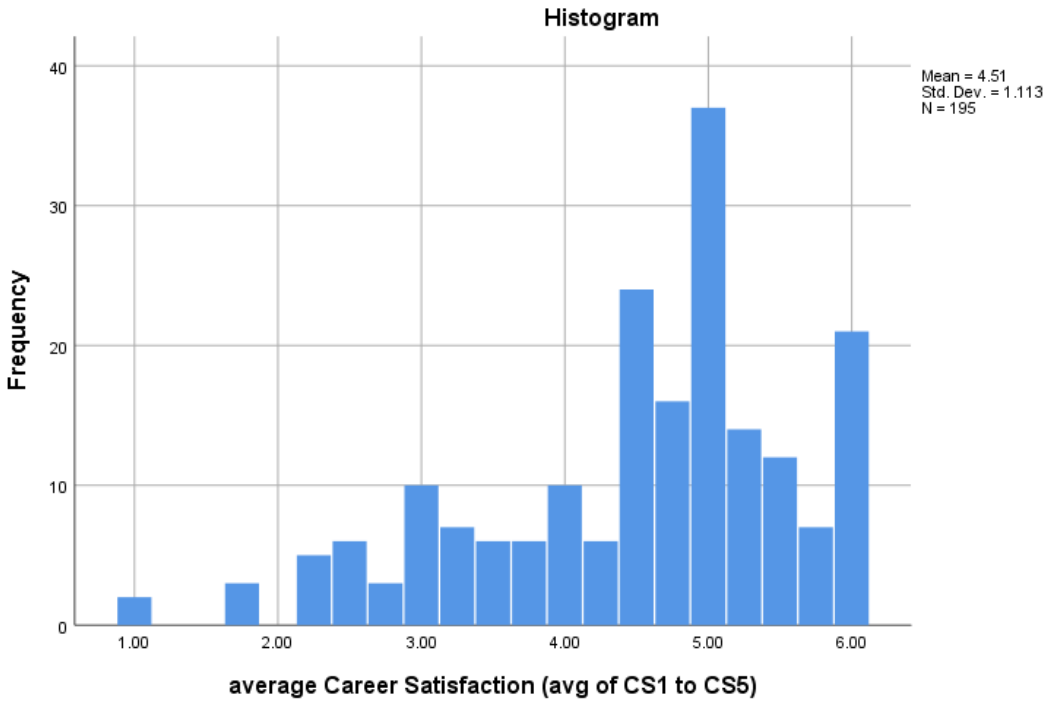
Though when testing Normality through Kolmogorov- Smirnov Test, it is clear that the distribution is not normal. This can be seen in the p-value being less than 0.05. Furthermore, since the skewness is negative, we can extrapolate that the distribution has a small left tail. Moreover, since the kurtosis is close to positive, we can expect the distribution to be relatively peaked.

Tests of Normality

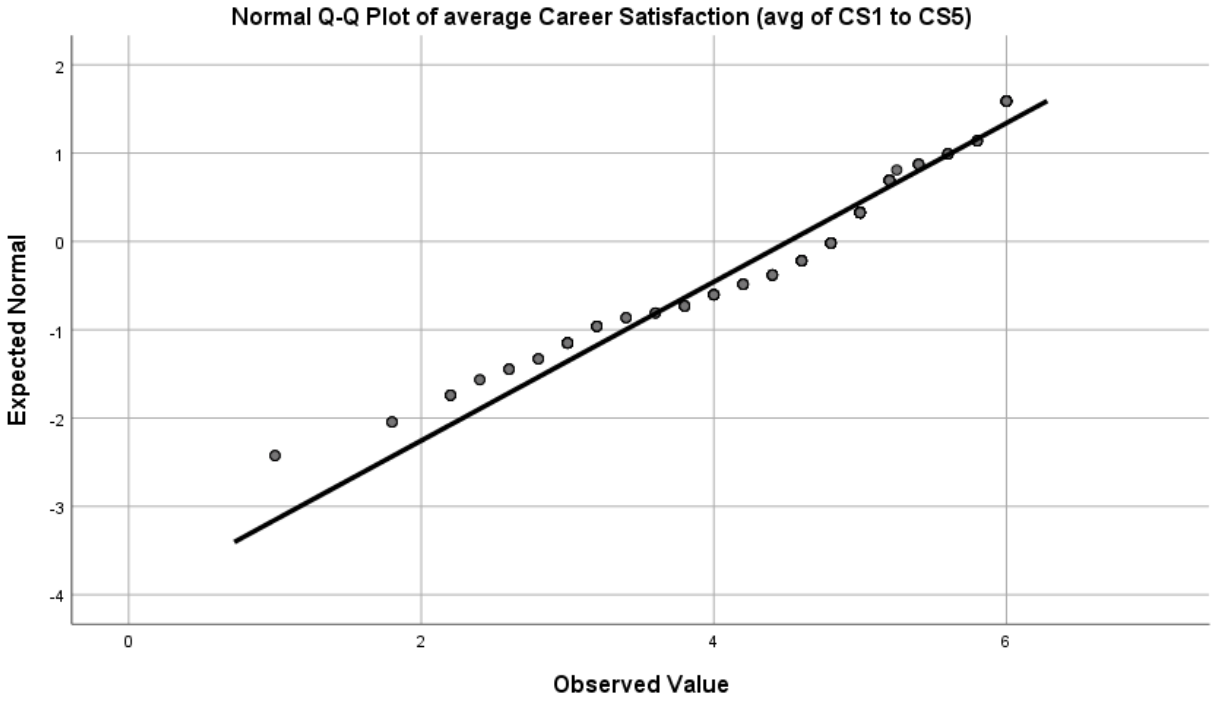
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
average Career Satisfaction (avg of CS1 to CS5)	.158	195	.000	.924	195	.000

a. Lilliefors Significance Correction

This conclusion of a non-normal distribution is supported by the histogram of the distribution of Average Career Satisfaction.



Moreover, when the qq-plot is examined, there is further evidence of a non-normal distribution. This can be seen in the fact that the dots in the plot, representing observation in the dataset, do not align with the fitted line which represents the expected data points for a normal distribution.



Thus, we can conclude as a whole that the distribution of Average Career Satisfaction is not normal.

Knowing Why Scale:

To assess a person's personal values and work interests (Jokinen, 2008), a new variable that accounts for all existing the Knowing Why variables must be created. In this case some of the questions in the Knowing Why scale was negatively worded, such as the questions represented by KW2. KW5, KW8. As a result, we will be average all KW and KW_rev except for KW2. KW5, KW8 to create the Average Knowing Why score (AVG_KW).

From the table below it is clear that all cases from the original Knowing Why Scale have a calculated value associated with the Average Knowing Why score.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
average Knowing Why Score (avg of KW and KW_rev except for KW2, KW5 and KW8)	195	3.44	7.00	5.7527	.73331
Valid N (listwise)	195				

Moreover, a mean of 5.75 and median of 5.89 indicate that on average, participants of the study had "self-awareness about personal values, work interests and their capabilities" (Jokinen, 2008). These findings are similar to the studies of Tiina Jokinen who found people with international work experience had scored above 4 on the knowing why scale, indicate the presence of self-awareness in work activities (Jokinen, 2008).

When assessing the normality of the Average Knowing Why score, the arbitrary difference between the mean and 5% trim mean gives us some indication that the distribution is normally distributed. However, the negative skewness indicates a left tail, and the positive kurtosis indicates a more peaked distributed than the normal distribution.

Descriptives

			Statistic	Std. Error
average Knowing Why Score (avg of KW and KW_rev except for KW2, KW5 and KW8)	Mean		5.7527	.05251
	95% Confidence Interval for Mean	Lower Bound	5.6491	
		Upper Bound	5.8563	
	5% Trimmed Mean		5.7930	
	Median		5.8889	
	Variance		.538	
	Std. Deviation		.73331	
	Minimum		3.44	
	Maximum		7.00	
	Range		3.56	
	Interquartile Range		1.00	
	Skewness		-.810	.174
	Kurtosis		.512	.346

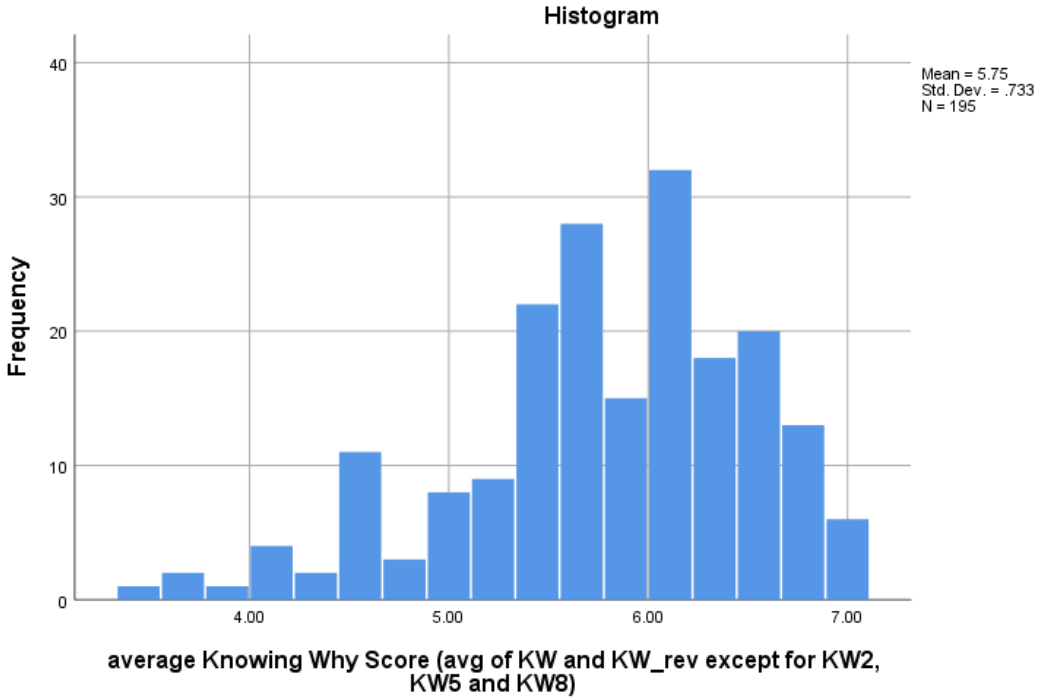
When testing normality using the Kolmogorov-Smirnov Test, it is even more clear that the distribution of the Average Knowing Why score is not normal. This is due to the significant result of the test.

Tests of Normality

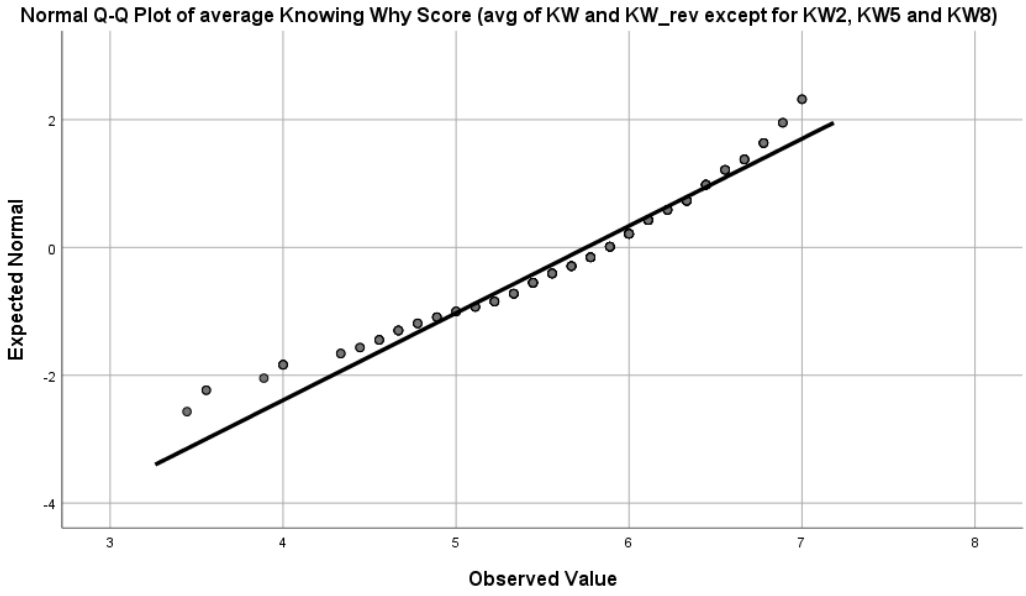
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
average Knowing Why Score (avg of KW and KW_rev except for KW2, KW5 and KW8)	.107	195	.000	.953	195	.000

a. Lilliefors Significance Correction

The conclusion of a non-normal distribution for the Knowing Why Scale is further supported by the shape of the histogram for the scale. It visualizes our extrapolation of skewness and kurtosis.



Lastly, the qq-plot indicates further proof of a non-normal distribution as observations don't align with the expected values of a normal distribution.



Reliability

Career Satisfaction Scale:

The reliability tests below do not use 1 of the rows from the dataset since, one value was deleted from CS5 for being an error. Otherwise, all 6 variables (CS1 to CS5 and AVG_CS) were included in Cronbach's Alpha Coefficient Test and the correlation analysis. The high Cronbach Coefficient of 0.975 indicates high internal consistency of the Career Satisfaction Scale.

Case Processing Summary

		N	%
Cases	Valid	194	99.5
	Excluded[a]	1	.5
	Total	195	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.957	.958	6

Moreover, the inter-item correlation values are all positive indicating that there aren't any negatively worded questions present in the scale. Meaning, all measurements are using the same reference points.

Inter-Item Correlation Matrix

	I am satisfied with the success I have acheived in my career	I am satisfied with the progress I have made toward meeting my overall career goals	I am satisfied with the progress I have made toward meeting my goals for income	I am satisfied with the progress I have made toward meeting my goals for advancement	I am satisfied with the progress I have made toward meeting my goals for the development of new skills	average Career Satisfaction (avg of CS1 to CS5)
I am satisfied with the success I have acheived in my career	1.000	.871	.800	.821	.641	.930
I am satisfied with the progress I have made toward meeting my overall career goals	.871	1.000	.759	.862	.703	.941
I am satisfied with the progress I have made toward meeting my goals for income	.800	.759	1.000	.749	.532	.870
I am satisfied with the progress I have made toward meeting my goals for advancement	.821	.862	.749	1.000	.702	.928
I am satisfied with the progress I have made toward meeting my goals for the development of new skills	.641	.703	.532	.702	1.000	.789
average Career Satisfaction (avg of CS1 to CS5)	.930	.941	.870	.928	.789	1.000

Since all variables have high corrected inter-item correlation, we can say with confidence that the Career score is measuring exacting what we intend it to measure.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
I am satisfied with the success I have acheived in my career	22.5412	30.614	.896	.	.946
I am satisfied with the progress I have made toward meeting my overall career goals	22.5361	30.422	.913	.	.944
I am satisfied with the progress I have made toward meeting my goals for income	22.6598	30.859	.805	.	.957
I am satisfied with the progress I have made toward meeting my goals for advancement	22.6289	30.444	.892	.	.946
I am satisfied with the progress I have made toward meeting my goals for the development of new skills	22.2629	34.445	.714	.	.964
average Career Satisfaction (avg of CS1 to CS5)	22.5258	31.038	1.000	.	.936

Knowing Why Scale:

The reliability test below uses all cases from the dataset and all the positive worded questions represented as variables, are included as items in the Cronbach Alpha Coefficient. A coefficient of 0.918 indicates high internal consistency of the Knowing Why Scale.

Case Processing Summary

		N	%
Cases	Valid	195	100.0
	Excluded[a]	0	.0
	Total	195	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.918	.926	10

Since the inter-item correlation values are all positive, we know for sure that we correctly flipped all negatively worded question in the scale.

Inter-Item Correlation Matrix

	I am able to recognise my own strengths and weaknesses, needs and motives	I understand what other people think about me (reverse of KW2)	I recognise the effect of my own cultural background on my thinking and behaviour	I acknowledge my personal values and beliefs	I set goals for personal development (reverse of KW5)	I undertake activities to enhance my skills and competencies	I want to know more than is required for task accomplishment	I have no confidence in my own capabilities (reverse of KW8)	I have a positive approach toward work-related challenges	average Knowing Why Score (avg of KW and KW_rev except for KW2, KW5 and KW8)
I am able to recognise my own strengths and weaknesses, needs and motives	1.000	.561	.594	.700	.456	.426	.483	.621	.520	.771
I understand what other people think about me (reverse of KW2)	.561	1.000	.484	.456	.425	.478	.452	.454	.390	.694
I recognise the effect of my own cultural background on my thinking and behaviour	.594	.484	1.000	.677	.394	.342	.405	.554	.435	.705
I acknowledge my personal values and beliefs	.700	.456	.677	1.000	.512	.465	.464	.626	.523	.786
I set goals for personal development (reverse of KW5)	.456	.425	.394	.512	1.000	.758	.618	.474	.431	.781
I undertake activities to enhance my skills and competencies	.426	.478	.342	.465	.758	1.000	.672	.421	.423	.769
I want to know more than is required for task accomplishment	.483	.452	.405	.464	.618	.672	1.000	.447	.406	.757
I have no confidence in my own capabilities (reverse of KW8)	.621	.454	.554	.626	.474	.421	.447	1.000	.711	.767
I have a positive approach toward work-related challenges	.520	.390	.435	.523	.431	.423	.406	.711	1.000	.704
average Knowing Why Score (avg of KW and KW_rev except for KW2, KW5 and KW8)	.771	.694	.705	.786	.781	.769	.757	.767	.704	1.000

Since all variables have high corrected inter-item correlation, we can say with confidence that the Knowing Why scale is measuring exacting what we intend it to measure.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
I am able to recognise my own strengths and weaknesses, needs and motives	51.7168	45.496	.722	.	.910
I understand what other people think about me (reverse of KW2)	52.2296	44.747	.614	.	.915
I recognise the effect of my own cultural background on my thinking and behaviour	51.7732	45.158	.633	.	.913
I acknowledge my personal values and beliefs	51.6450	44.344	.733	.	.908
I set goals for personal development (reverse of KW5)	52.0399	41.412	.702	.	.911
I undertake activities to enhance my skills and competencies	51.8655	42.430	.695	.	.911
I want to know more than is required for task accomplishment	51.7271	42.479	.678	.	.912
I have no confidence in my own capabilities (reverse of KW8)	51.5271	44.884	.712	.	.910
I have a positive approach toward work-related challenges	51.4450	45.015	.631	.	.914
average Knowing Why Score (avg of KW and KW_rev except for KW2, KW5 and KW8)	51.7744	43.557	1.000	.	.898

Works Cited

Jokinen, T., Brewster, C., & Suutari, V. (2008). Career capital during international work experiences: contrasting self-initiated experiences and assigned expatriation. *The International Journal of Human Resource Management*, 987-989. Retrieved from https://doi.org/10.1080/09585190802051279