

Introduction to Statistics

1. DATA PREPARATION

From the statistics table, we can see the presence of missing values for the variables **reserved**, **perseveres** and **efficient**. We have replaced these missing values as -1 and indicated this on SPSS for the missing value definition.

Statistics													
		What is your sex	What is your age?-years	talkative	does a thorough job	reserved	disorganized	quiet	perseveres	shy	efficient	lazy	careless
N	Valid	451	451	451	451	447	451	451	449	451	447	451	451
	Missing	0	0	0	0	4	0	0	2	0	4	0	0

When analyzing the other tables, we can observe that the variables: **does a thorough job**, **lazy** and **careless** have wrong values that are not in the defined scale. Hence, I will consider these values as missing values.

does a thorough job

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Totally disagree	7	1,6	1,6	1,6
	2	30	6,7	6,7	8,2
	3	58	12,9	12,9	21,1
	4	154	34,1	34,1	55,2
	Totally agree	201	44,6	44,6	99,8
	6	1	,2	,2	100,0
	Total	451	100,0	100,0	

lazy

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Totally disagree	122	27,1	27,1	27,1
	2	91	20,2	20,2	47,2
	3	92	20,4	20,4	67,6
	4	104	23,1	23,1	90,7
	Totally agree	39	8,6	8,6	99,3
	7	1	,2	,2	99,6
	33	2	,4	,4	100,0
	Total	451	100,0	100,0	

careless

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Totally disagree	90	20,0	20,0	20,0
	2	129	28,6	28,6	48,6
	3	99	22,0	22,0	70,5
	4	111	24,6	24,6	95,1
	Totally agree	21	4,7	4,7	99,8
	22	1	,2	,2	100,0
	Total	451	100,0	100,0	

After running again the frequency of variables, we can observe that the missing values are identify by -1 and the variables are grouped in the defined scale.

does a thorough job

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Totally disagree	7	1,6	1,6	1,6
	2	30	6,7	6,7	8,2
	3	58	12,9	12,9	21,1
	4	154	34,1	34,2	55,3
	Totally agree	201	44,6	44,7	100,0
	Total	450	99,8	100,0	
Missing	-1	1	,2		
Total		451	100,0		

reserved

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Totally disagree	92	20,4	20,6	20,6
	2	90	20,0	20,1	40,7
	3	102	22,6	22,8	63,5
	4	120	26,6	26,8	90,4
	Totally agree	43	9,5	9,6	100,0
	Total	447	99,1	100,0	
Missing	-1	4	,9		
Total		451	100,0		

perseveres

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Totally disagree	22	4,9	4,9	4,9
	2	40	8,9	8,9	13,8
	3	98	21,7	21,8	35,6
	4	148	32,8	33,0	68,6
	Totally agree	141	31,3	31,4	100,0
	Total	449	99,6	100,0	
Missing	-1	2	,4		
Total		451	100,0		

efficient

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Totally disagree	2	,4	,4	,4
	2	29	6,4	6,5	6,9
	3	82	18,2	18,3	25,3
	4	186	41,2	41,6	66,9
	Totally agree	148	32,8	33,1	100,0
	Total	447	99,1	100,0	
Missing	-1	4	,9		
Total		451	100,0		

		lazy			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Totally disagree	122	27,1	27,2	27,2
	2	91	20,2	20,3	47,5
	3	92	20,4	20,5	68,1
	4	104	23,1	23,2	91,3
	Totally agree	39	8,6	8,7	100,0
	Total	448	99,3	100,0	
Missing	-1	3	,7		
Total		451	100,0		

		careless			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Totally disagree	90	20,0	20,0	20,0
	2	129	28,6	28,7	48,7
	3	99	22,0	22,0	70,7
	4	111	24,6	24,7	95,3
	Totally agree	21	4,7	4,7	100,0
	Total	450	99,8	100,0	
Missing	-1	1	,2		
Total		451	100,0		

2. FACTOR ANALYSIS ON PERSONALITY VARIABLES (PERS 1 → PERS 10)

I would perform factor analysis to **reduce the variables** related to personality measures. The first thing is, we need to know if it is meaningful to perform a factor analysis. The KMO results in a value of **0.735** which is greater than 0.50 and Bartlett's Test results in a value of **0.001** which is less than 0.05. Therefore these results indicates that it makes sense to perform the factor analysis on this data.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.735
Bartlett's Test of Sphericity	Approx. Chi-Square	1123,349
	df	45
	Sig.	≤.001

The communalities show that the proposed factors solution explain in a good level the variance for each one of the variables under analysis, in this case each variable are at least explained in **53.9%** by the factor solution.

Communalities

	Initial	Extraction
talkative	1,000	,614
does a thorough job	1,000	,610
reserved	1,000	,681
disorganized	1,000	,539
quiet	1,000	,775
perseveres	1,000	,550
shy	1,000	,590
efficient	1,000	,538
lazy	1,000	,706
careless	1,000	,698

Extraction Method: Principal Component Analysis.

Using the Total Variance table and scree plot, we can decide about the number of factors which we can include in our model.

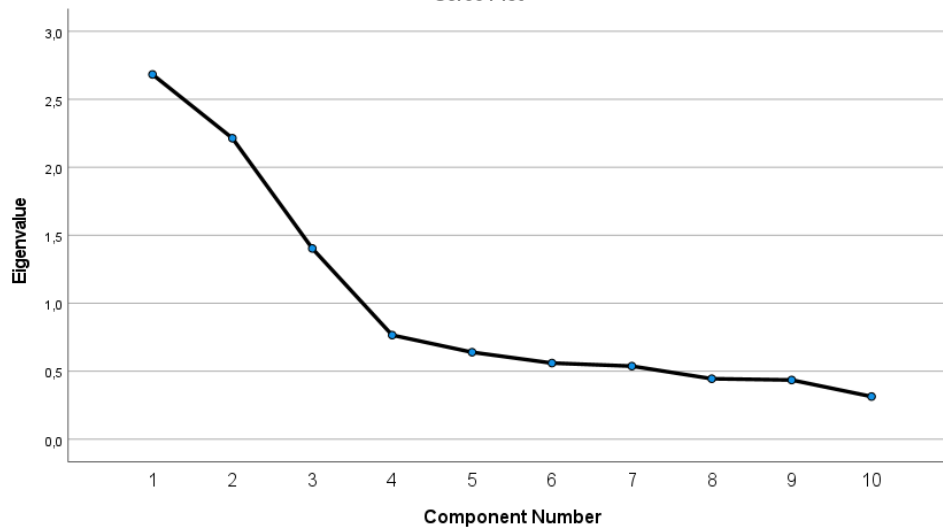
Total Variance Explained

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	2,683	26,832	26,832	2,683	26,832	26,832	2,596	25,956	25,956
2	2,214	22,142	48,974	2,214	22,142	48,974	2,283	22,830	48,786
3	1,403	14,035	63,009	1,403	14,035	63,009	1,422	14,222	63,009
4	,766	7,659	70,667						
5	,640	6,400	77,067						
6	,560	5,603	82,670						
7	,538	5,378	88,048						
8	,445	4,452	92,500						
9	,436	4,358	96,858						
10	,314	3,142	100,000						

Extraction Method: Principal Component Analysis.

The total variance explain suggests that we should consider **three** factors, based on the initial Eigenvalues, the first three are the only ones greater than 1. These three components explain the **63.009%** of the total variance in the dataset.

Scree Plot



From the above scree plot, we can say that the elbow is at 4, and you consider elbow – 1 factors, which means **three factors**. Both metrics shows the same conclusion about the inclusion of three factors in the analysis.

Linking variables to factors

By using the Rotated Component Matrix, we can decide about which variables to put in each factor.

Rotated Component Matrix^a

	Component		
	1	2	3
quiet	,880	,003	-,030
reserved	,813	,105	,091
shy	,766	-,018	-,064
talkative	-,747	,236	,008
does a thorough job	-,038	,776	-,077
efficient	,038	,732	-,017
disorganized	,003	-,732	-,055
perseveres	-,112	,731	-,061
lazy	,010	,015	,840
careless	-,024	-,087	,831

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

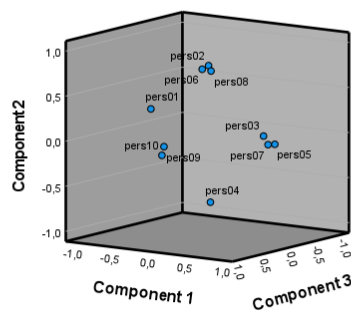
a. Rotation converged in 4 iterations.

BEHAVIOUR

JOB PERFORMANCE

COMMITMENT

Component Plot in Rotated Space



We can conclude that the variables quiet, reserved, shy and talkative are in the same factor. Variables: does a thorough job, efficient, disorganized, perseveres are in the same factor and finally lazy and careless belong to the last factor which gives us three factors in total.

I will define the variables in the following way:

Behaviour : This refers to how people act towards their relationship with others which include measure about being quiet, reserved, shy and talkative.

Job performance: It describe the variables related to how people do their job and include measures of does a thorough job, efficient; perseveres, disorganized.

Commitment: This factor explains the people who are lazy or careless. It includes measures of Lazy and careless.

From the results, we observe that it is necessary to perform a recoding in the variables **talkative** and **disorganized** because they have the minus symbol which means that they are not in the same direction as the other variables within the factor and before creating the new variables, we need to test the reliability of the variables in each factor. We will use Cronbach's Alpha for testing the variables inside the first two factors,

Factor 1: Behaviour

We run the reliability analysis for the variables inside the first factor; the Cronbach's Alpha value is **0.815** which is greater than 0.70. This indicates that the variables inside the factor 1 are consistent.

Reliability Statistics

Cronbach's Alpha	N of Items
,815	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
quiet	8,49	9,878	,747	,711
reserved	8,60	11,178	,636	,767
shy	8,21	11,384	,584	,790
talkative	9,05	11,300	,576	,795

We can also understand from the above table that there is no improvement in Cronbach's Alpha value if any of the variables has been removed from this factor.

Factor 2: Job Performance

We run the reliability analysis for the variables inside the second factor; the Cronbach's Alpha value is **0.715** which is greater than 0.70. This indicates that the variables inside the factor 2 are consistent.

Reliability Statistics

Cronbach's Alpha	N of Items
,715	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
does a thorough job	11,20	6,954	,548	,633
efficient	11,34	7,402	,516	,655
disorganized	11,91	5,605	,494	,679
perseveres	11,58	6,542	,505	,652

We can also understand from the above table that there is no improvement in Cronbach's Alpha value if any of the variables has been removed from this factor.

Factor 3: Commitment

For the third factor we only have two variables and hence Cronbach's alpha cannot be calculated. I will perform Pearson correlation in order to verify the consistency within variables from factor 3.

Correlations			
		lazy	careless
lazy	Pearson Correlation	1	,418**
	Sig. (2-tailed)		<,001
	N	448	448
careless	Pearson Correlation	,418**	1
	Sig. (2-tailed)	<,001	
	N	448	450

** . Correlation is significant at the 0.01 level (2-tailed).

In this case, we can observe that $r = 0.418$, with a significance level of $p < .001$. Hence, we can conclude that we can aggregate the two variables belonging to the third factor.

Now we can summate all variables because it satisfies all the **three conditions**:

- All questions need to be measured on the **same scale**
- All questions need to be scaled in the **same direction**
- The new variable should contain only variables that measure the same construct (Check **Cronbach's Alpha**)

After summated the three new variables, we have the following statistics for the total data set:

Statistics				
		behaviour	JobPerformance	Commitment
N	Valid	451	451	450
	Missing	0	0	1
Mean		2,8640	3,8343	2,6567
Median		3,0000	4,0000	2,5000
Mode		3,25	5,00	2,00
Std. Deviation		1,06666	,82048	1,05812

Splitting by age (mean=31), below are the statistics results:

Age >= 31:

Statistics				
		behaviour	JobPerformance	Commitment
N	Valid	165	165	164
	Missing	0	0	1
Mean		2,9470	3,8652	2,5915
Median		3,0000	4,0000	2,5000
Mode		3,25	4,25	3,00
Std. Deviation		1,09805	,81361	,99886

Age <31:

Statistics				
		behaviour	JobPerformance	Commitment
N	Valid	286	286	286
	Missing	0	0	0
Mean		2,8161	3,8164	2,6941
Median		2,8750	4,0000	2,5000
Mode		3,00 ^a	4,00	2,00
Std. Deviation		1,04707	,82532	1,09060

a. Multiple modes exist. The smallest value is shown