

Configuration Manual

M.Sc. Research Project

Data Analytics

SHEKHAR TOMAR Student ID: x22156861

School of Computing National College of Ireland

Supervisor: Dr. Arghir Nicolae Moldovan

National College of Ireland



Year:2023-2024

MSc Project Submission Sheet

School of Computing

Student Name: Shekhar Tomar

Student ID: x22156861

Programme: M.Sc. in Data Analytics

Module: M.Sc. Research Project

Supervisor: Dr. Arghir Nicolae Moldovan

Submission Due Date: 14 December 2023

Project Title: A STATISTICAL STUDY ON ANTHROPOMETRIC DATA OF CHILDREN

Word Count:

Page Count: 07

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

All internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: SHEKHAR TOMAR

Date: 14 DECEMBER 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	
Attach a Moodle submission receipt of the online project	
You must ensure that you retain a HARD COPY of the project, both	
for your own reference and in case a project is lost or mislaid. It is not	
sufficient to keep a copy on computer.	

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Configuration Manual

Shekhar Tomar X22156861

Introduction

Statistical analysis plays a crucial role in drawing meaningful insights from data, aiding in decision-making processes, and supporting research findings. In this document, we will focus on the statistical software IBM SPSS Statistics version 29 for data analysis. The system configuration, including software and hardware specifications, will be outlined, followed by steps for data generation. Finally, we will delve into the procedures for conducting various statistical tests, such as T-test, ANOVA, MANCOVA, Regression, Correlation, and Descriptive Summary.

System Configuration

2.1 Software Specification

SPSS Statistics 29 is a widely used statistical software that facilitates data analysis and interpretation. To install SPSS 29, follow these steps:

Open your web browser and go to your preferred search engine.

and using this link open the NCI moodle page to download the SPSS.

Step 1:-

https://www.ncirl.ie/Students/Current-Students-Hu



Step 2:-

	None International International			SharePoint			P Search this library					
â		L	IT Studer	nt Support								
0												
	년 Sha	re	Copy link	🗟 Add shortcut to One	Drive	\downarrow Download $\mathscr{P}_{\mathbb{R}}^{\mathbb{R}}$	Automate \sim 🗈 Copy to \cdots					
٥	_											
0	Docun	nents	SPSS									
۲		۵	Name \vee			Modified \vee	Modified By \vee					
		•	Linux		х	January 21, 2020	Joseph McDonnell					
	٢	•	Mac	x & 50		September 21, 2016	Frank Byrne					
		•	Windows		х	September 21, 2016	Frank Byrne					
		a l	Link to Key a	nd Install Guide.docx	х	August 20, 2021	Joseph McDonnell					

Step 3:-

	Normal Industry		SharePoint			₽ Search this library
© ⊕		1	T Student Support			
•	🖻 Shan	e	\odot Copy link $\ extsf{ }$ Add shortcut \lor $\ extsf{ }$	Do	wnload 🖓 Automate	e 🗸 🗈 Copy to 🛛 …
C	Docum	ents	> SPSS > Mac			
Ð		D	Name \vee		Modified \vee	Modified By \sim
			ReadMe V29.txt	×	December 6	Joseph McDonnell
	0	-	SPSS Statistics Version 2 $ earrow earrow$		June 8	Joseph McDonnell
		80	SPSS Statistics Version 29 MacOS Son	×	December 6	Joseph McDonnell
		-	SPSS_Statistics_25_mac.dmg	×	October 5, 2018	Joseph McDonnell
		-	SPSS_Statistics_25_mac_FP001.dmg	×	October 31, 2018	Joseph McDonnell
		-	SPSS_Statistics_26_mac.dmg	×	August 21, 2019	Joseph McDonnell
		-	SPSS_Statistics_Client_27_Mac.pkg	х	November 12, 2020	Joseph McDonnell
		-	Statistics 28 0 11 Mac Log4J.pkg	×	May 11, 2022	Joseph McDonnell

2.2 Hardware Specification for SPSS 29.0.1.0

2.2.1 macOS

Operating System: macOS SONOMA

Processor: M2 Pro

RAM: 16 GB RAM Hard Disk: At least 4 GB free space for installation, with additional space for temporary files during analysis.

Screen Resolution: 3024×1964

3.1 How to load the .sav files with datasets and output

Loading a .sav file:

1. Open SPSS:

- Launch the SPSS software on your computer.

- 2. Open or Create a New Syntax or Data File:
 - You can either open an existing data file or create a new one.
- 3. Import Data:
 - To import data, go to 'File > Open > Data...'.
 - Choose the .sav file you want to load and click "Open."
- 4. Explore the Data:
 - Once the data is loaded, you can explore it in the Data View or Variable View tabs.

3.2 Data Generation Steps

Statistical Analysis Procedures

T-test

🖨 🙇 🤌) 🛄 📭 🔟	Re <u>p</u> orts D <u>e</u> scriptive Statistics	*	, 🔳 🖸	•		
	GET FILE='C:\Users	Bayesian Statistics	*	rean_daras	et.sav'.		
Title Notes	DATASET NAME Dat GET FILE='C:\Users	Compare Means General Linear Model	•	Means	le T Test		
Descriptive Stat	DATASET NAME Dat DESCRIPTIVES VAR	Generalized Linear Models Mixed Models	*	Independe	n <u>t</u> -Samples T Test		
scriptives] Title h Notes	/STATISTICS=ME		*	H Summary I	ndependent-Samples T Test nples T Test		
Active Dataset	Descriptives(Ir	Loglinear Neural Networks	*	Qne-Way ANOVA			
ा rrelations ो Title	[DataSet2] C:\Us	Classify	*	\indian_da	taset.sav		
Notes Correlations		Sc <u>a</u> le Nonnarametric Tests	*	Mean	Std. Deviation		
nparametric Corr] Title	Age Height	Forecasting	•	13.04 4.8655	1.990 .58189		
Correlations	Weight Best Grip R	Multiple Response	*	36.7492 16.81197815	9.70339 7.678623216		
rrelations	Best Grip L BMI	Multiple Imputation	•	16.55297304 15.66052968	7.469701196 3.825294560		
Active Dataset Correlations	Valid N (listwise)	Simulation					
의 nparametric Corr]] Title	DATASET ACTIVATE DESCRIPTIVES VAR	Quality Control Spatial and Temporal Modeling	P 	st_L heigh	t_cm weight_kg BMI		
h Notes	/ SIALISIICS-ME	Direct Marketing	P				

- The t-test is used to compare the means between two groups. To perform a t-test in SPSS 26:
- Select Analyze > Compare Means > Independent Samples T-Test for two independent groups or Paired-Samples T-Test for dependent groups.
- Choose the variables of interest and set the grouping variable.

• Review the output for statistical significance.

ANOVA (Analysis of Variance)



- ANOVA is employed when comparing means across more than two groups. To conduct ANOVA in SPSS:
- Navigate to Analyze > Compare Means > One-Way ANOVA.
- Specify the dependent variable and factor variable.
- Examine the ANOVA table for significance.

MANCOVA (Multivariate Analysis of Covariance)

📕 🚔 🗟 🧈	🛄 🗠 🗠	Reports	•		•		
	•	Descriptive Statistics	•				
Dutput 🚔	GET	Bayesian Statistics					
	FILE='C:\Users	Tables		rean_daras	et.sav'.		
Descriptives	DATASET NAME Dat	Compare Heaps					
- Notes	GET	Compare means					
Active Dataset	FILE='C:\Users	<u>G</u> eneral Linear Model	- P	With the second			
Descriptive Stat	DATASET NAME Dat	Generalized Linear Models	•	Multivariate			
Log	DESCRIPTIVES VAR	Mixed Models					
E Descriptives	/STATISTICS=ME	Correlate		Repeated I	Measures		
📺 Title				Variance C	omponents		
- Rotes	Descriptives(Ir	Regression	•	· · · · · · · · · · · · · · · · · · ·			
Active Dataset		L <u>og</u> linear	•				
La Descriptive Sta	[DataSat2] Call	Neural Networks		\			
Correlations	[DataSet2] C:(US	Classify		(Indian_da	casec.sav		
Title		Dimension Deduction					
- Notes		Dimension Reduction					
Correlations		Sc <u>a</u> le	•	Maan	Std Deviation		
Log		Nonparametric Tests	•	mean	ota. Deviation		
🖲 Nonparametric Corr	Age	Forecasting		13.04	1.990		
Title	Height	Suprival		4.8655	.58189		
Notes	Weight	Sulvival		36.7492	9.70339		
	Best Grip R	Multiple Response	P	16.81197815	7.678623216		
Correlations	Best Grin I	🔣 Missing Value Analysis		16 55297304	7 469701196		
Title	DML	Multiple Imputation		5.66052060	2.025204560		
- Notes	DIVI	Complex Samples		5.00052908	3.825294560		
Active Dataset	Valid N (listwise)						
Correlations		er Simulation					
🖆 Log	DATASET ACTIVATE	Quality Control	•				
🖲 Nonparametric Corr	DESCRIPTIVES VAR	Spatial and Temporal Modeling		st. L heigh	t om weight ko	BMT	
Title	/STATISTICS=ME	Direct Marketing					
Correlations		Direct margeary]			

- MANCOVA is an extension of ANOVA that considers multiple dependent variables. To perform MANCOVA:
- Go to Analyze > General Linear Model > Multivariate.

- Select the dependent variables and covariate(s).
- Review the Pillai's Trace or other multivariate tests for significance.

Regression

Data	Iranst	orm <u>I</u> nsert F <u>o</u> rmat	Analyze Graphs Utilities	Extensions	<u>Window H</u> elp
		🛄 🗠 🍽	Reports Descriptive Statistics	*	
		CRE	Bayagian Statistics		
		FILF='C:\Users		, r	rean daraset sav'.
tives		DATASET NAME Dat	i a <u>p</u> ies		
9		GET	Co <u>m</u> pare Means	•	
ve Dataset		FILE='C:\Users	<u>G</u> eneral Linear Model	•	dian_dataset.sav'.
scriptive St	at	DATASET NAME Dat	Generalized Linear Models	•	
		DESCRIPTIVES VAR	Mixed Models	•	pR BestGripL BMI
tives		/ STATISTICS-ME	<u>C</u> orrelate	•	
es		Descriptives/Ir	Regression		Automatic Linear Modeling
ve Datase		Descriptives(ii	L <u>og</u> linear	•	Linear
scriptive St	a	[DataSet2] C:\Us	Neural Net <u>w</u> orks	•	Curve Estimation
ions			Classify	•	Partial Least Squares
9			Dimension Reduction	•	
es relations			Sc <u>a</u> le	•	Binary Logistic
			Nonparametric Tests	•	Multinomial Logistic
ametric Co	n	Age	Forecasting	•	Section 2015 Secti
e		Height	Survival	•	🔛 Probit
relations		Weight	– Multiple Response	*	Nonlinear
		Best Grip R	💯 Missing Value Analysis		Weight Estimation
ions		Best Grip L	Multiple Imputation	•	2-Stage Least Squares
es		Valid N (listwise)	Complex Samples	•	Quantile
ve Dataset			Bimulation		Ontimal Scaling (CATREG)
1 5160 0115		DATASET ACTIVATE	Quality Control		
ametric Co	n	DESCRIPTIVES VAR	Spatial and Temporal Model	ling 🕨	st L height cm weight kg BMI
e		/STATISTICS=ME	Direct Marketing		
relations					
te Analysis		Descriptives/K	orean)		

- Regression analysis assesses the relationship between a dependent variable and one or more independent variables. To conduct regression in SPSS:
- Choose Analyze > Regression > Linear.
- Specify the dependent and independent variables.
- Interpret the regression coefficients and significance levels.

Correlation

<u>E</u> dit <u>V</u> iew <u>D</u> ata <u>T</u> rans	form <u>I</u> nsert F <u>o</u> rmat	Analyze Graphs Utilities Exten	sions	<u>W</u> indow	<u>H</u> elp			
) 🔲 合 🗟 🔊		Reports	•					
		Descriptive Statistics	•		<u>></u>			
Output	GET	Bayesian Statistics	•					
Log	FILE='C:\Users	Tables		rean_daras	et.sav'.			
E Descriptives	DATASET NAME Dat	Compare Means						
R Notes	GET	General Linear Model						
Active Dataset	FILE= C: USETS		ĺ.	dian_datas	lian_dataset.sav'.			
Ling Descriptive Stat	DESCRIPTIVES VAR		, r	pR BestGri	DL BMI			
E Descriptives	/STATISTICS=ME	Mixed Models	P					
m Title		<u>C</u> orrelate	•	Bivariate				
- Rotes	Descriptives/Ir	Regression	•	artial				
Active Dataset	Descriptives(ii	L <u>og</u> linear	•	M Distances				
Leg Descriptive Stat	[DataSat2] C.)IIa	Neural Networks						
Correlations	[Databet2] 0.105	Classify		E Canonical	Correlation			
👘 Title		Dimension Reduction	*					
- R Notes		Scale						
Correlations		Nonporometrio Testa		Mean	Std. Deviation			
Log	Age	Nonparametric rests	r.	13.04	1.990	-		
Title	Height	Forecasting		4.8655	.58189			
- Rotes	Weight	Survival	•	36,7492	9.70339			
Correlations	Best Grip R	M <u>u</u> ltiple Response	•	16 81197815	7 678623216	m		
Log	Best Grip I	🌠 Missing Value Analysis		16 55297304	7 469701196	m		
E Conelations	BMI	Multiple Imputation		15 66052968	3 825294560	-		
- Rotes	Valid N (listwise)	Complex Samples	•	0.00002000	0.020201000			
Active Dataset	vanu iv (listwise)	Bimulation				-		
Leg Correlations		Quality Control						
Nonparametric Corr	DATASET ACTIVATE							
- m Title	DESCRIPTIVES VAR	Spatial and Lemporal Modeling		st_L heigh	t_cm weight_	kg BMI		
- Notes	/STATISTICS=ME	Direct Marketing	•					

- Correlation examines the strength and direction of relationships between two continuous variables. To perform correlation analysis:
- Navigate to Analyze > Correlate > Bivariate.
- Select the variables of interest.
- Examine the correlation matrix for significance.

Descriptive Summary

ile <u>E</u> d	lit <u>V</u> iew <u>D</u> ata	ransf	orm <u>I</u> nsert F <u>o</u> rmat	<u>Analyze</u> <u>G</u> raphs <u>U</u> tilities	Extensions	Window H	<u>H</u> elp
	- 🗛 🗟	J		Regorts	۲.		
				Descriptive Statistics	•	Erequencie	s
■ <u>6</u> 0	utput Malaa		GET	Bayesian Statistics	•	Descriptive	S
- ē	Descriptives		FILE='C:\Users	Ta <u>b</u> les		A Explore	•
	Title		GET	Co <u>m</u> pare Means		Crosstabs	
	Active Dataset		FILE='C:\Users	General Linear Model		TURE Anal	vsis
	🖵 🗃 Descriptive Sta		DATASET NAME Dat	Generalized Linear Models	•	Ratio	,
	Log		/STATISTICS=ME	Mixed Models	•	D D Diete	
-	Title			Correlate	•	E-F Flots	
	- Rotes		Descriptives(Ir	Regression	•	u-u Plots	
	Active Dataset			L <u>og</u> linear	•		
-4	Log		[DataSet2] C:\Us	Neural Networks	•	\indian_da	taset.sa
₽ -{	Correlations			Classify	•		
	Notes			Dimension Reduction	*		
	Gorrelations			Sc <u>a</u> le	*	Mean	Std. Deviat
	Log Nonnarametric Cor		Age	Nonparametric Lests		13.04	1.9
1	Title		Height	Forecasting		4.8655	.5818
	Notes		Weight	Survival		36.7492	9.7033
-0	Log		Best Grip R	Multiple Response	,	16.81197815	7.678623216
į	Correlations		Best Grip L	Missing Value Anal <u>v</u> sis		16.55297304	7.46970119
	Title		BMI	Multiple Imputation	*	15.66052968	3.825294
	Active Dataset		Valid N (listwise)	Complex Samples	•		
	Correlations			Simulation			
	🖬 L0g 🖬 Nonnarametric Cor		DATASET ACTIVATE	Quality Control			
T.	Title		DESCRIPTIVES VAR	Spatial and Temporal Modelli Direct Marketics	1g P	st_L heigh	t_cm weigh
			/ SIATISTICS-ME	Direct Marketing	•		

- Descriptive statistics provide a summary of key characteristics of a dataset. To generate descriptive statistics in SPSS:
- Choose Analyze > Descriptive Statistics > Descriptives.
- Select the variables to analyse.
- Review the output for measures such as mean, standard deviation, and skewness.

Histogram and Line Chart

Data Transform Analyze	e <u>G</u> raphs <u>U</u> tilities Ex	tensions <u>W</u> indow <u>H</u> elp						
) 🛄 🛌 🧃								
	🚛 Chart Builder				×			Visible: 7
🛷 Age 🛛 💰 Gender	<u>V</u> ariables:	Chart preview uses example data	[÷]	Element Properties Chart Appearance	Options	var	var	var
10	A 400	[Edit Properties of:				
10	Gender				X			
10	A Height							
10	I Weight	Drag a Gallery chart here to use it as your						
10	Best Grip R BestGri	starting point		·				
10	BMI	OR			-			
10					-			
10		Click on the Basic Elements tab to build a cha element by element	art		-			
10	No setemarias (asolo				-			
10	variable)				-			
10								
10								
10								
10	Gallery Basic Elements	Groups/Point ID Titles/Footnotes						
10	<u>C</u> hoose from:							
10	Favorites							
10	Bar							
10	Area							
10	Pie/Polar							
10	Scatter/Dot 🛛 🍄 🧕	ΦφΦΦ						
10	Histogram	ΦΦ						
10	Boxplot				-			
10	Dual Axes				-			
10								
10		OK Paste Reset Cancel Help						

• Navigate to Graphs > Histogram.

- Choose the variable for which you want to create the histogram.
- Go to Graphs > Line.
- Choose Line and then Multiple