AIM: Conducting Statistical Analysis on The Students' Mathematics subject Scores

A Statistical Analysis is performed on the StudentData.txt dataset from the UCI machine learning repository and the data has many categorical variables like school, sex, mother's job, father's job etc and we found some interesting relationships between different categorical variables by summarizing the categorical variables and then relating them to the test scores.

Graphical Methods used to identify key relationships

Math Test Score Dataset has a wide range of data in terms of the information related to the student. So, some categorical variables are taken into observation, and conclusions are drawn regarding the relationship between the given student's information and the test scores.

- Bar Plots displaying the relationship between Mother's Job and Father's Job (Mjob & Fjob) and Its effect on the math test scores
- Summarization of test scores based on mother's job and father's job
- Scatter plots displaying how math test scores (G1) depend on internet connectivity
- Relationships are drawn between the math test scores (G1, G2,G3) and the school they are studying in.
- Scatter Plots are created depicting the relationship between the test scores (G1, G2, G3) and the reason they joined the particular school

Datasets contain both numerical and categorical variables, so we analysed the information considering both graphical and numeric variables across graphical summaries:

Graphical and numeric summaries:

- The CONTENTS procedure
- The FREQ procedure
- The MEANS procedure
- BAR PLOTS
- SCATTER PLOTS

Exploring Relationships on Mother's Job, Father's Job, School, Internet Connectivity, Reason and School

- Our concerns are addressed in the following order. First, we wanted to understand if there is a strong relationship existed between the test scores and the jobs of their mother and father.
- We dive down further by analyzing how internet connectivity plays a major role in student's learning graph and eventually the test scores.
- Then we investigated the relationship between the student's test scores and the school they are studying in. We believe that the school's course, distance from home, and reputation drive the student's education environment and eventually the test scores.
- Finally, the reason for choosing the best school by their parents impacts the learning environment of the student and its effect on test scores.

Summary of Questions:

1. How do the parents' jobs correlate to the math test scores?
2. What is the structure of Bar plots for MJob and Fjob?
3. How is the parameter internet connectivity impact the student's test scores?
4. How does a choice of school impact the student's test scores?
5. What can be inferred from the relationship between the school and the reason behind choosing the particular school by their parents?

Analysis and Summary of Findings

The CONTENTS Procedure

Data Set Name	MYLIB.STUDENTDATA	Observations	886
Member Type	DATA	Variables	23
Engine	V9	Indexes	0
Created	02/06/2024 08:52:42	Observation Length	152
Last Modified	02/06/2024 08:52:42	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	862
Obs in First Data Page	886
Number of Data Set Repairs	0
Filename	/home/u63773409/sasuser.v94/Stdat
Release Created	9.0401M7
Host Created	Linux
Inode Number	84643827147
Access Permission	rw-r--r--
Owner Name	u63773409
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
17	Date	Num	8	BEST.		Date
8	Fees	Num	8	BEST.		Fees
10	Fyd	Char	8	$8.	$8.	Fyd
14	G1	Num	8	BEST.		G1
15	G2	Num	8	BEST.		G2
16	G3	Num	8	BEST.		G3
7	Marks	Num	8	BEST.		Marks
9	Mjyb	Char	8	$8.	$8.	Mjyb
6	Potatus	Char	1	$1.	$1.	Potatus
20	Wale	Num	8	BEST.		Wale
19	absences	Num	8	BEST.		absences
18	activities	Char	3	$3.	$3.	activities
3	address	Char	1	$1.	$1.	address
2	age	Num	8	BEST.		age
13	failures	Num	8	BEST.		failures
24	famrel	Num	8	BEST.		famrel
5	famsize	Char	3	$3.	$3.	famsize
12	famsup	Char	3	$3.	$3.	famsup
22	freetime	Num	8	BEST.		freetime
25	good	Num	8	BEST.		good
11	guardian	Char	8	$8.	$8.	guardian
26	health	Num	8	BEST.		health
21	higher	Char	3	$3.	$3.	higher
23	internet	Char	3	$3.	$3.	internet
28	nursery	Char	3	$3.	$3.	nursery
16	paid	Char	3	$3.	$3.	paid
11	reason	Char	10	$10.	$10.	reason
23	romantic	Char	3	$3.	$3.	romantic
1	school	Char	2	$2.	$2.	school

Table of Contents

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
17	Date	Num	8	BEST.		Date
8	Fees	Num	8	BEST.		Fees
10	Fyd	Char	8	$8.	$8.	Fyd
14	G1	Num	8	BEST.		G1
15	G2	Num	8	BEST.		G2
16	G3	Num	8	BEST.		G3
7	Marks	Num	8	BEST.		Marks
9	Mjyb	Char	8	$8.	$8.	Mjyb
6	Potatus	Char	1	$1.	$1.	Potatus
20	Wale	Num	8	BEST.		Wale
19	absences	Num	8	BEST.		absences
18	activities	Char	3	$3.	$3.	activities
3	address	Char	1	$1.	$1.	address
2	age	Num	8	BEST.		age
13	failures	Num	8	BEST.		failures
24	famrel	Num	8	BEST.		famrel
5	famsize	Char	3	$3.	$3.	famsize
12	famsup	Char	3	$3.	$3.	famsup
22	freetime	Num	8	BEST.		freetime
25	good	Num	8	BEST.		good
11	guardian	Char	8	$8.	$8.	guardian
26	health	Num	8	BEST.		health
21	higher	Char	3	$3.	$3.	higher
23	internet	Char	3	$3.	$3.	internet
28	nursery	Char	3	$3.	$3.	nursery
16	paid	Char	3	$3.	$3.	paid
11	reason	Char	10	$10.	$10.	reason
23	romantic	Char	3	$3.	$3.	romantic
1	school	Char	2	$2.	$2.	school

Messages: 2 User: u63773409

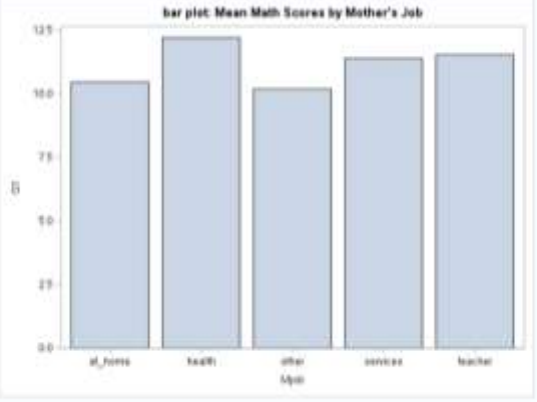
The FREQ Procedure

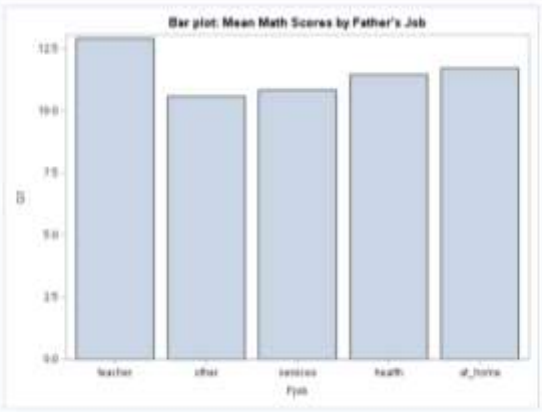
Mjob	Frequency	Percent	Cumulative Frequency	Cumulative Percent
at_home	59	14.94	59	14.94
health	34	8.61	93	23.54
other	141	35.70	234	59.24
services	103	26.08	337	85.32
teacher	58	14.68	395	100.00

Fjob	Frequency	Percent	Cumulative Frequency	Cumulative Percent
at_home	20	5.06	20	5.06
health	18	4.56	38	9.62
other	217	54.94	255	64.56
services	111	28.10	366	92.66
teacher	29	7.34	395	100.00

The MEANS Procedure

Mjob	Fjob	N Obs	Mean	Std Dev	Minimum	Maximum
at_home	at_home	7	13.8571429	3.8321651	6.0000000	18.0000000
	health	2	12.0000000	3.0000330	10.0000000	15.0000000
	other	33	9.7272727	2.7809420	5.0000000	17.0000000
	services	15	10.6666667	3.5730957	6.0000000	18.0000000
	teacher	2	7.0000000	2.8284271	5.0000000	9.0000000
health	health	6	12.0333333	2.7868746	9.0000000	17.0000000
	other	17	12.0000000	2.6220321	7.0000000	16.0000000
	services	10	13.4000000	4.1100132	8.0000000	19.0000000
	teacher	1	10.0000000	.	10.0000000	10.0000000
other	at_home	5	8.4000000	2.3021720	7.0000000	13.0000000
	health	3	11.0000000	2.1213203	10.0000000	13.0000000
	other	100	10.1442308	3.2132891	4.0000000	18.0000000
	services	24	10.0833333	3.1811317	5.0000000	15.0000000
	teacher	6	11.3333333	3.1411251	6.0000000	15.0000000
services	at_home	6	11.1666667	3.3115958	6.0000000	14.0000000
	health	4	9.7500000	2.6299656	6.0000000	12.0000000
	other	42	11.4761905	3.4037791	6.0000000	18.0000000
	services	43	10.8604651	3.0625507	6.0000000	17.0000000
	teacher	8	14.7500000	3.0165151	7.0000000	19.0000000
teacher	at_home	2	11.0000000	7.7781746	6.0000000	17.0000000
	health	4	10.0000000	2.0199113	8.0000000	14.0000000
	other	21	11.0000000	2.8100932	6.0000000	16.0000000
	services	19	11.0000000	2.8468012	7.0000000	16.0000000
	teacher	12	13.0000667	3.5700144	7.0000000	18.0000000

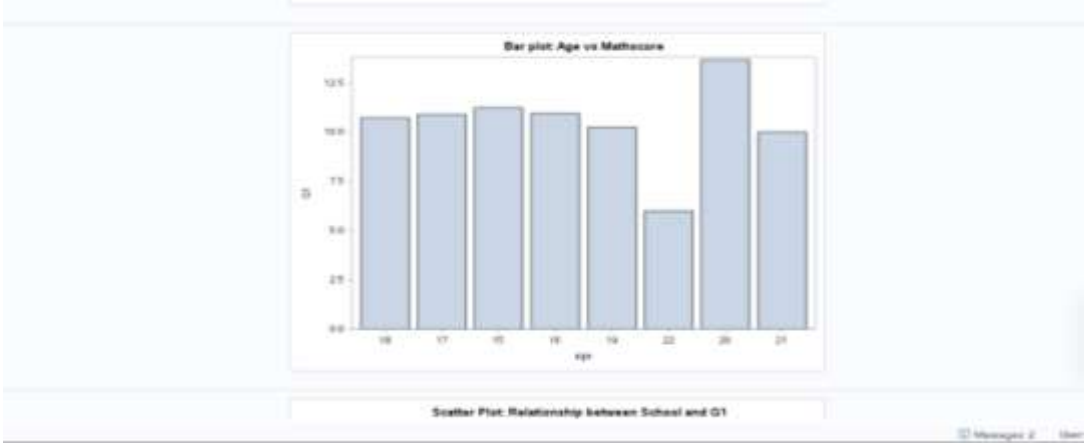



Bar plot: Mean Math Scores by Father's Job
G3
12.5
10.0
7.5
5.0
2.5
0.0
teacher
other
services
health
at_home
Fjob


Scatter Plot: Internet vs G1
G1
20
15
10
5
no
yes
internet


Scatter Plot: Internet vs G2
G2
20
15
10
5
0
no
yes
internet

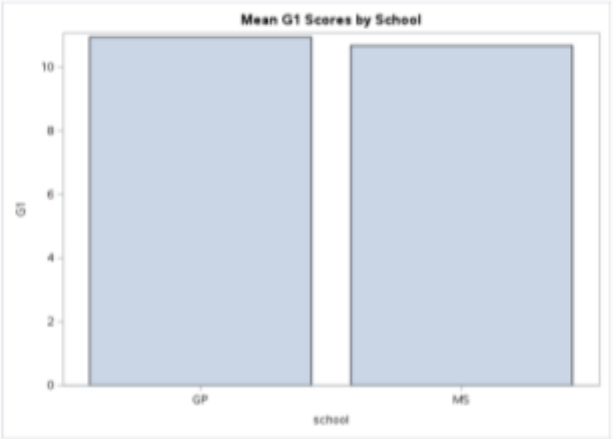

Scatter Plot: Internet vs G3
G3
internet
Bar plot: Age vs Mathscore

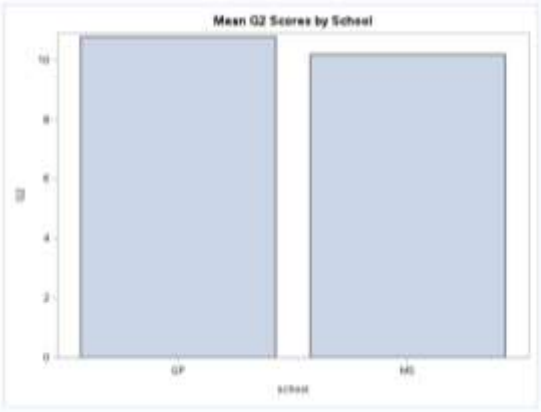

Bar plot: Age vs Mathscore
G3
age
Scatter Plot: Relationship between School and G1

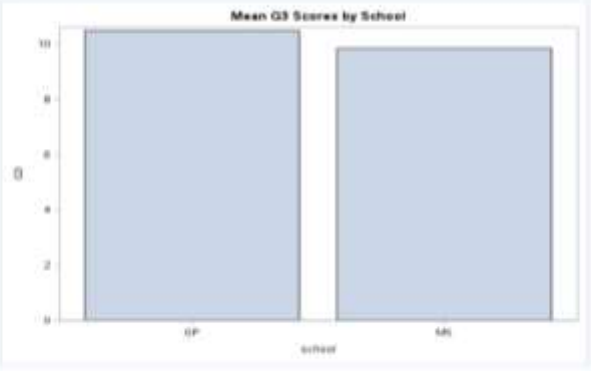

Scatter Plot: Relationship between School and G1
G1
Scatter Plot: Relationship between School and G2


Scatter Plot: Relationship between School and G2
20
15
10
5
0
G2
GP
MS
school

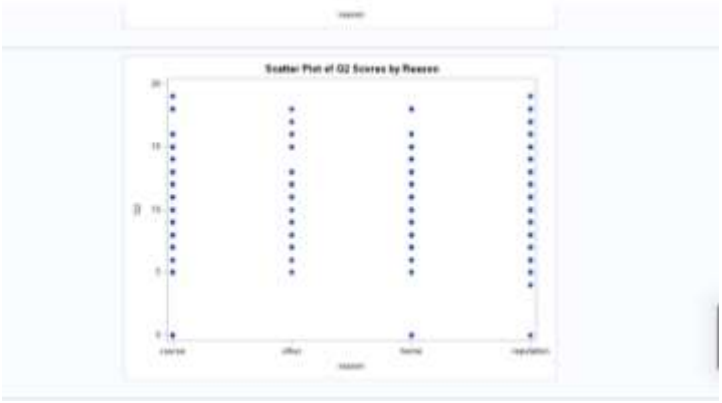

Scatter Plot: Relationship between School and G3
20
15
10
5
0
G3
GP
MS
school

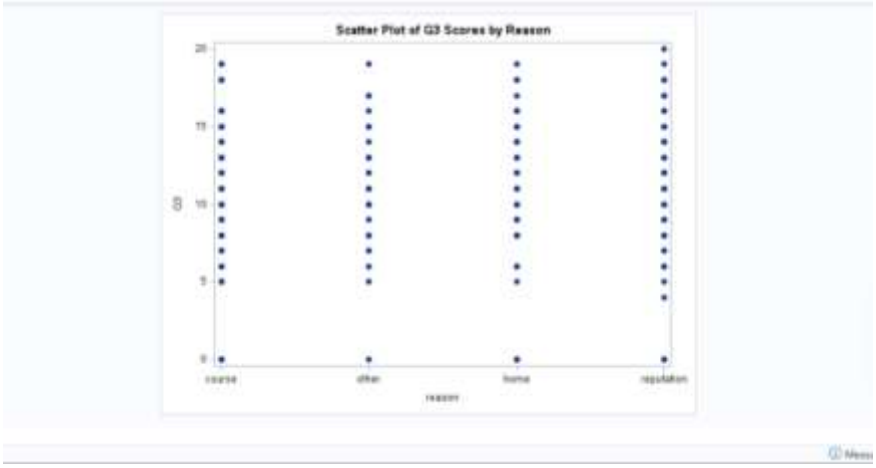

Mean G1 Scores by School
10
8
6
4
2
0
G1
GP
MS
school


Mean G2 Scores by School
G2
0
2
4
6
8
10
GP
MS
school


Mean G3 Scores by School
G3
0
2
4
6
8
10
GP
MS
school


Scatter Plot of G1 Scores by Reason
G1
0
5
10
15
20
course
other
home
reputation
reason


Scatter Plot of G2 Scores by Reason
G2
reason
course
other
home
reputation


Scatter Plot of G3 Scores by Reason
G3
reason
course
other
home
reputation