

STATISTICAL ANALYSIS USING EXCEL

(STUDENTS PERFORMANCE ANALYSIS)



INDEX

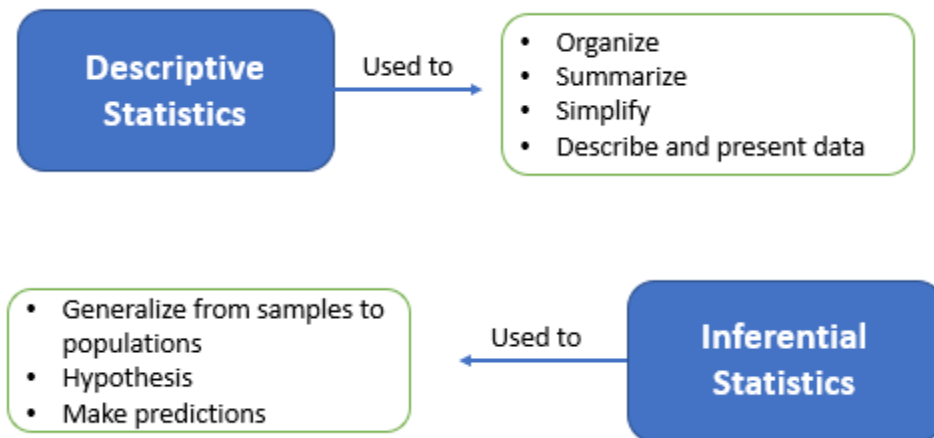
S NO	TOPIC	PAGE NO
1	Introduction	3
2	Descriptive Statistics	4
3	Inferential Statistics	4
4	Summary	8
5	References	9

INTRODUCTION

What is statistics?

Statistics is the science concerned with developing and studying methods for collecting, analyzing, interpreting and presenting empirical data.

Types of statistics:



About dataset:

Here I have taken the dataset about reading, writing and math performance of students.

Descriptive Statistics

- Descriptive statistics summarizes or describes the characteristics of a data set.

Summary Math Score	Column1
Mean	66.089
Standard Error	0.47949869
Median	66
Mode	65
Standard Deviation	15.1630801
Sample Variance	229.918998
Kurtosis	0.27496406
Skewness	-0.2789351
Range	100
Minimum	0
Maximum	100
Sum	66089
Count	1000

Summary Reading Score	Column1
Mean	69.169
Standard Error	0.46169861
Median	70
Mode	72
Standard Deviation	14.6001919
Sample Variance	213.165605
Kurtosis	-0.0682655
Skewness	-0.2591045
Range	83
Minimum	17
Maximum	100
Sum	69169
Count	1000

Summary Writing Score	Column1
Mean	68.054
Standard Error	0.48052887
Median	69
Mode	74
Standard Deviation	15.195657
Sample Variance	230.907992
Kurtosis	-0.0333646
Skewness	-0.289444
Range	90
Minimum	10
Maximum	100
Sum	68054
Count	1000

Inferential Statistics

- Inferential statistics helps to suggest explanations for a situation or phenomenon. It allows you to draw conclusions based on extrapolation.
- Hypothesis:**
 - Hypothesis testing is used to assess the plausibility of a hypothesis by using sample data.
 - For all hypothesis testing I have set alpha value as 0.05.
- Null hypothesis:**
 - The main purpose of a null hypothesis is to verify/ disprove the proposed statistical assumptions.
- Alternative Hypothesis:**
 - An alternative hypothesis is a statement that describes that there is a relationship between two selected variables in a study.

T-test:

- A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features.
- There are 3 types of T-test such as :
 - Performs a paired t-test.
 - Two-sample assuming equal variance t-test.
 - Two-sample assuming unequal variance t-test.
- Here I have performed "Two-sample assuming equal variance t-test".

t-Test: Two-Sample Assuming Equal Variances

	<i>Reading Score</i>	<i>Writing Score</i>
Mean	69.169	68.054
Variance	213.1656046	230.907992
Observations	1000	1000
Pooled Variance	222.0367983	
Hypothesized Mean Difference	0	
df	1998	
t Stat	1.67319821	
P(T<=t) one-tail	0.047222412	
t Critical one-tail	1.64561663	
P(T<=t) two-tail	0.094444825	
t Critical two-tail	1.961152015	

- The value of p is lesser than alpha so we need to reject null hypothesis.

ANOVA Test:

- The ANOVA test allows a comparison of more than two groups at the same time to determine whether a relationship exists between them.
- Types of ANOVA test are:
 - Anova: Single Factor
 - Anova Two Factor with Replication
 - Anova Two Factor without Replication

Anova: Single Factor

SUMMARY

<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Math Score	1000	66089	66.089	229.919
Reading Score	1000	69169	69.169	213.1656
Writing Score	1000	68054	68.054	230.908

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	4863.616667	2	2431.808	10.82419	2.07019E-05	2.998729
Within Groups	673318.602	2997	224.6642			
Total	678182.2187	2999				

- The value of p is lesser than alpha so we need to reject null hypothesis.

Correlation:

- Correlation means association - more precisely it is a measure of the extent to which two variables are related.
- There are three possible results of a correlational study: a **positive correlation**, a **negative correlation**, and **no correlation**.

	<i>Math Score</i>	<i>Reading Score</i>	<i>Writing Score</i>
Math Score	1		
Reading Score	0.817579664	1	
Writing Score	0.802642046	0.954598077	1

Covariance:

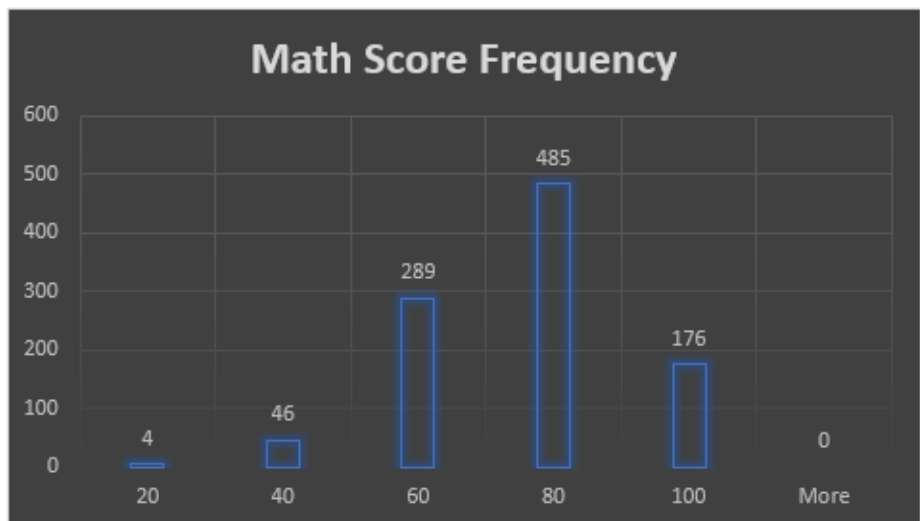
- Covariance can have both positive and negative values. Based on this, it has two types:
 - Positive Covariance
 - Negative Covariance

	<i>Math Score</i>	<i>Reading Score</i>	<i>Writing Score</i>
Math Score	229.689079		
Reading Score	180.817959	212.952439	
Writing Score	184.754194	211.574874	230.677084

Histogram:

- The histogram is a popular graphing tool. It is used to summarize discrete or continuous data that are measured on an interval scale.

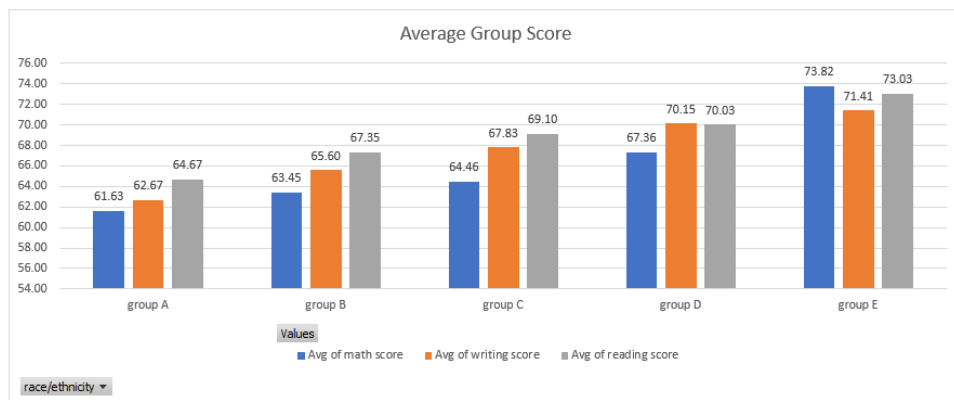
<i>Bin</i>	<i>Frequency</i>
20	4
40	46
60	289
80	485
100	176
More	0



Pivot Table:

- A PivotTable is an interactive way to quickly summarize large amounts of data.

Row Label	Avg of math score	Avg of writing score	Avg of reading score
group A	61.63	62.67	64.67
group B	63.45	65.60	67.35
group C	64.46	67.83	69.10
group D	67.36	70.15	70.03
group E	73.82	71.41	73.03
Grand Total	66.09	68.05	69.17



Rank and Percentile:

- Percentile** is a value below which a percentage of data falls.
- Ranking** is the data transformation in which numerical or ordinal values are replaced by their rank when the data are sorted.

Math Score				Reading Score				Writing Score			
Point	Math Score	Rank	Percent	Point	Reading Score	Rank	Percent	Point	Writing Score	Rank	Percent
150	100	1	99.30%	107	100	1	98.30%	107	100	1	98.60%
452	100	1	99.30%	115	100	1	98.30%	115	100	1	98.60%
459	100	1	99.30%	150	100	1	98.30%	166	100	1	98.60%
624	100	1	99.30%	166	100	1	98.30%	180	100	1	98.60%
626	100	1	99.30%	180	100	1	98.30%	378	100	1	98.60%
917	100	1	99.30%	382	100	1	98.30%	404	100	1	98.60%
963	100	1	99.30%	459	100	1	98.30%	459	100	1	98.60%
115	99	8	99.00%	547	100	1	98.30%	567	100	1	98.60%
264	99	8	99.00%	567	100	1	98.30%	686	100	1	98.60%
307	99	8	99.00%	595	100	1	98.30%	904	100	1	98.60%
105	98	11	98.70%	713	100	1	98.30%	917	100	1	98.60%
713	98	11	98.70%	887	100	1	98.30%	958	100	1	98.60%
935	98	11	98.70%	904	100	1	98.30%	963	100	1	98.60%
35	97	14	98.10%	917	100	1	98.30%	971	100	1	98.60%
180	97	14	98.10%	958	100	1	98.30%	595	99	15	98.20%
287	97	14	98.10%	963	100	1	98.30%	626	99	15	98.20%
540	97	14	98.10%	971	100	1	98.30%	713	99	15	98.20%

SUMMARY

All the statistical analysis is performed on the student performance dataset. From descriptive analysis we can see the average score of students are lies above 60. Comparing all three scores shows that reading score is highest than other two. Math score is lesser than the writing score. In both t-test and ANOVA tests, it depicts that there is a significant difference between the values. All the scores are positively relative to each other and in case of covariance also it is positively varies each other. The frequency values of math score is shown in this file all remaining test and charts are available in the excel file. It is uploaded in the github.

REFERENCE

- www.kaggle.com