

# [None](https://assignbuster.com/none-essay-samples/)

[Science](https://assignbuster.com/essay-subjects/science/), [Statistics](https://assignbuster.com/essay-subjects/science/statistics/)

Statistics XL10: Project Brief introduction of the project The project seeks to understand the scores obtained by students in the U. S based on gender and State in which the student comes from. The dataset comprises of 4 variables namely; gender of the student, State in which the student comes from, the number of hours the student takes to prepare for a Mathematics test and the scores obtained by the student in a Mathematics test. In total, 30 students took part in this survey. Data was collected both from the exam registry and one on one with the students. The registry provided us with the students scores while the students were asked to state the number of hours they took to prepare for the test and the state in which they come from.
Qualitative and Quantitative Variables
a) Histograms
i) Histogram for the number of hours taken to read for the exam
ii) Histogram for the score in Mathematics
b) Measures of central tendency
Statistics
hrs\_read
score
N
Valid
30
30
Missing
0
0
Mean
5. 3000
61. 5000
Median
5. 0000
63. 5000
Std. Deviation
1. 64317
13. 1037
Variance
2. 700
171. 707
Range
7. 00
69. 00
Based on the above histogram, the most important metrics to study would be the mean and also looking at the parametric tests since from the histogram, the variables shows that they follow a normal distribution
c) Bar graph and pie charts of the qualitative variables
i) Bar graph for the gender
ii) Pie chart for gender
iii) A bar graph for the states
iv) A pie chart for the States
Clearly from the charts presented above it is clear that in terms of gender more male respondents (students) took part in the survey as compared to the female respondents. 57% (N= 17)of those who took part were the male respondents while the female respondents were 43% (N= 13).
In terms of the states, Alabama State had the highest number of the respondents while California had the least number of students interviewed in this survey. 30% (N= 9) of those who took part in the survey were from Alabama, 27% (N= 8) were from Arizona, 20% (N= 6) were from California while 23% (N= 7) were from the State of Illinois.
d) Boxplots
i) Box plot for the number of hours read
ii) Box plot for the score
Part Three
a) Z-score table for the quantitative variables
z\_score\_hrs
z\_score\_grade
z\_score\_hrs
z\_score\_grade
z\_score\_hrs
z\_score\_grade
1
-2. 00831
-1. 25919
11
1. 034586
0. 2671
21
-3. 06292
-4. 69333
2
-1. 39973
0. 419729
12
-0. 18257
-1. 25919
22
-3. 99179
-4. 69333
3
-0. 79115
-0. 57236
13
-0. 79115
-3. 09073
23
-5. 10643
-4. 69333
4
-0. 18257
-0. 49604
14
-0. 79115
-0. 41973
24
-3. 48091
-4. 69333
5
-0. 79115
-1. 18287
15
-0. 79115
0. 419729
25
-2. 97003
-4. 69333
6
-0. 79115
1. 259186
16
0. 426006
1. 259186
26
-2. 45916
-4. 69333
7
-0. 79115
2. 174958
17
0. 426006
-0. 57236
27
-3. 5738
-4. 69333
8
-0. 18257
1. 259186
18
0. 426006
0. 2671
28
-3. 06292
-4. 69333
9
0. 426006
0. 419729
19
1. 034586
0. 038157
29
-3. 20225
-4. 69333
10
1. 034586
0. 419729
20
1. 643165
-0. 19079
30
-3. 34158
-4. 69333
b) Histograms for the z-score transformed
c) Justification of outliers
Based on the first histogram (transformed z-score on number of hours), it is clear that the variable is free from the outliers however, the second histogram (transformed z-score on the score) we can clearly see by visualization that there are some elements of outlier in the variable.