Statistics Project

Introduction

This project is basically the analysis of the vocabularies of people based on their medium of schooling , their reading habits and the various newspapers they read. We conducted three tests for every individual namely verbal speed, how fast you are and verbal responsiveness.

Basically the first test checked the verbal speed of an individual in determining if 2 adjacent words are either same, opposite or different. They had to solve this test in 3 minutes for 30 words.

The 2nd test comprised of jotting down words beginning from 'D' within 3 minutes except nouns and tenses of a single word.

The third test was to write synonyms or approximately same meaning word of the given word mandatorily starting with alphabet 'P'. This one had no time limit. We collected about 150 samples in our college ageing from 17-24 years.

Result Table

Table displaying Summary Statistics -

Test-1		Test-2		Combining test1 and test2		Test-3	
Mean	**56.52174**	Mean	**33.69565**	Mean	**90.21739**	Mean	**7.695652**
Standard Error	**3.919077**	Standard Error	**2.98576**	Standard Error	**4.905009**	Standard Error	**0.651647**
Median	**50**	Median	**25**	Median	**75**	Median	**7**
Mode	**50**	Mode	**25**	Mode	**75**	Mode	**6**
Standard Deviation	**18.79524**	Standard Deviation	**14.3192**	Standard Deviation	**23.5236**	Standard Deviation	**3.125188**
Sample Variance	**353.2609**	Sample Variance	**205.0395**	Sample Variance	**553.3597**	Sample Variance	**9.766798**
Kurtosis	**0.072466**	Kurtosis	**1.411298**	Kurtosis	**-0.893199**	Kurtosis	**1.872025**
Skewness	**0.217922**	Skewness	**1.457427**	Skewness	**0.193846**	Skewness	**1.366518**
Range	**75**	Range	**50**	Range	**75**	Range	**13**
Minimum	**25**	Minimum	**25**	Minimum	**50**	Minimum	**3**
Maximum	**100**	Maximum	**75**	Maximum	**125**	Maximum	**16**
Count	**23**	Count	**23**	Count	**23**	Count	**23**

Interpretation from the result table:

We thus conclude that the since there is a great variance the vocabulary varies from individual to individual.
Here in the cases of test1 , test2 and their combined summary statistics we notice
A.M. $>M_O \geq M_E$ therefore the distributions are positively skewed.

But only in test 3
A.M.$>M_E>M_O$, hence it denotes that the distribution of scores in test 3 is perfectly positively skewed.

Kurtosis for the distributions of all the three tests are less than 3 i.e. hence all the distributions are platykurtic.

Correlation between age and score of test 1 is :
Corr(age, test1)= -0.1261 this reveals that the correlation is negligible since mod r< 0.3