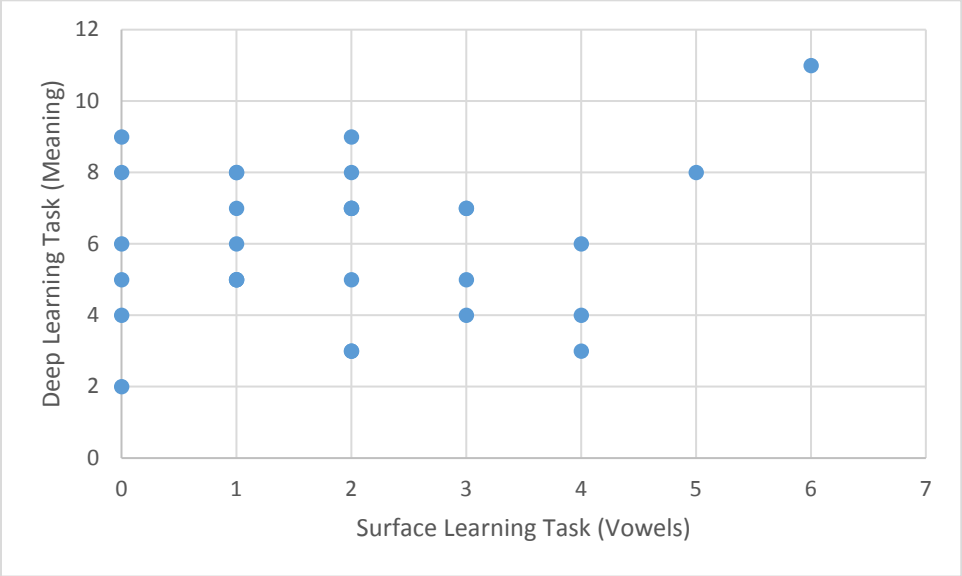
How People Learn

Exercises Week 7

Question 1

(a) Scatter plots are used to represent the relationship between two continuous variables. Represent this data on a scatter plot comparing each student's performance on each test. Make sure to include the features that are required in all charts (number title, axis titles etc.)

Chart 1: Comparison of incidental learning scores on meaningful and superficial tasks



(b) Numerically, the relationship between two continuous variables can be described using the correlation co-efficient (which can be calculated used the CORREL function in Excel). What is the correlation between these two variables? How would you interpret this figure (is the correlation positive or negative? Is it weak, moderate or strong?).

r=0.18 (rounded to two decimal places)

Correlation coefficients can normally be rounded to two decimal places.

The correlation is positive (i.e. as the x-value goes up, so, in general, does the y-value).

It is a pretty weak correlation. For reference, Andy Field's textbook (p. 32) suggests that a correlation of magnitude r=.10 is small/weak, r=.30 is medium/moderate and an r=.50 is large/strong. You will find variations on this in other text books. In many areas of psychology and sociology a correlation of .50 is so strong as to be almost unheard of.

(c) Correlation, (like mean average) can be distorted by outlier cases. Using the scatterplot, identify if there are any outlier (i.e. unusual) cases here. If so, remove the outlier and calculate the correlation between the two variables. Interpret your finding.

If a single outlier (case 15) is removed, the correlation becomes r=-0.05. This is negative but so weak as to be effectively zero.

How do we interpret this data? For a start, we might have expected that there would be a correlation between the two things because we might expect that people who are good at learning are good at learning in any situation. However the data does not support this contention. There is a weak correlation between the two which is effectively an artifact of a single outlier. Effectively, this means that their incidental learning under the two different circumstances seems unrelated.

Question 2

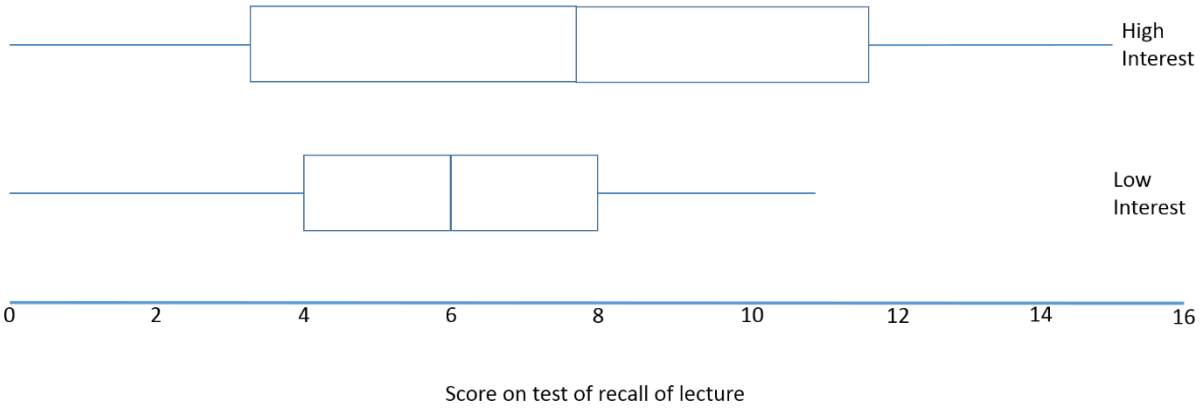
In this case we have data in matched pairs: that is the data in the two variables are collected from the same person. Therefore we will use a two sample t-test with matched pairs. H_0: The mean score on the superficial incidental learning task is equal to the mean score on the meaningful incidental learning task in the wider population.

H_1: The mean score on the superficial incidental learning task is less than the mean score on the meaningful incidental learning task in the wider population.

Because of the alternative hypothesis, this is a 1-tailed test. If I do this in excel I get a value of $p = 1.2545 \times 10^{-10}$, which I will report as p<0.001. This means there is very strong evidence against the null hypothesis and we should accept the alternate hypothesis that the meaningful incidental learning score is higher than the superficial incidental learning score.

Question 3

Chart 2: Scores of low interest and high interest groups on sports history test



Question 4

In this case, we do not have matched pairs, i.e., the students in the two groups are not the same people. We do not have two samples which are larger than 25 (rule of thumb from earlier in the term). Therefore we will use a t-test.

To know which t-test we will look at the variances of the two samples. The variance of the low interest group is 7.17. For the experimental group it is 22.53. There look quite different but we need to have some rule as to whether or not they are too different that we can pool them or not. One common rule of thumb is that is the ratio of the two variances is less than 3, then we can assume they are equal. In this case the ratio of the two variances is 7.17:22.53 or 1:3.14. We therefore do not use pooled variances.

H_0: The score of the high interest group is equal to the score of the low interest group.

H_1: The score of the high interest group is greater than the score of the low interest group.

Running a one tailed t-test with unequal variances assumed in excel gives a p value of p=0.13. This is greater than the critical value of 0.05 and so in this case we cannot reject the null hypothesis that the scores of the two groups in a wider population are equal.