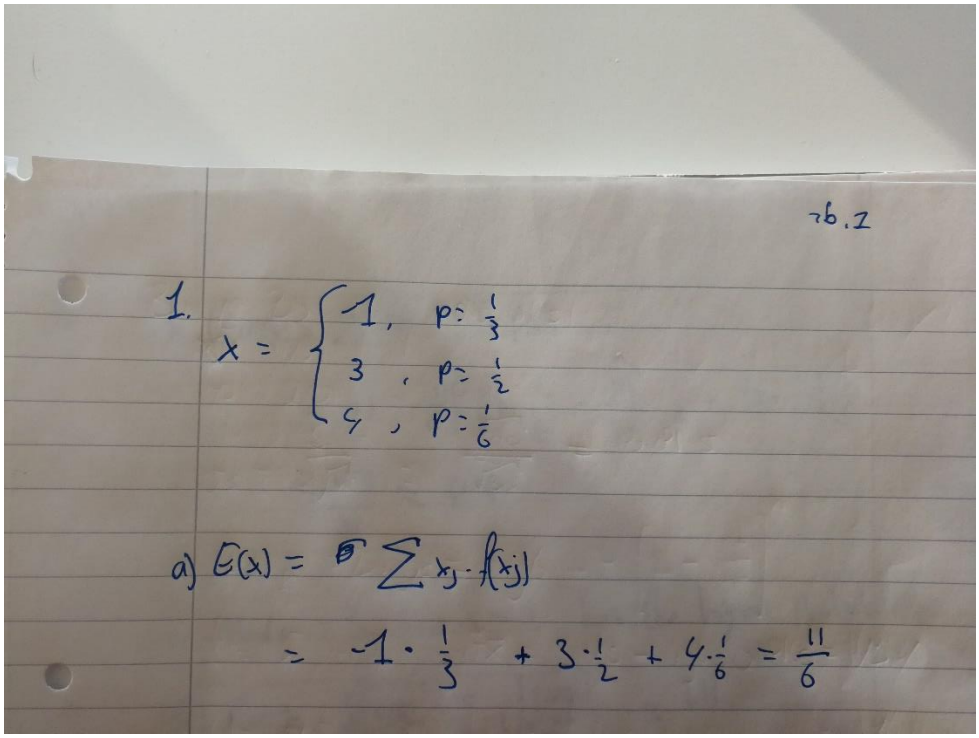
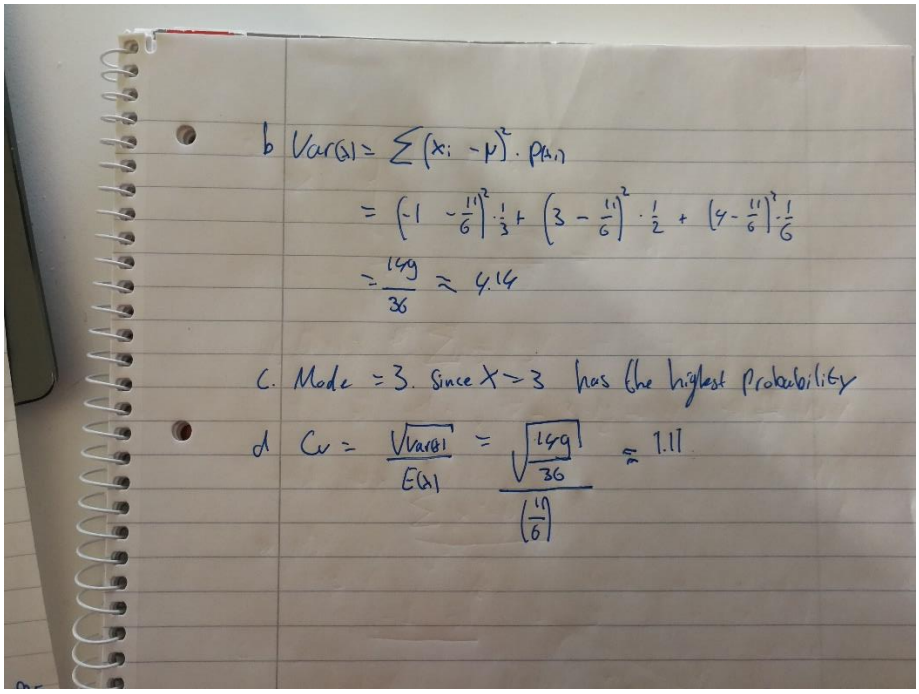
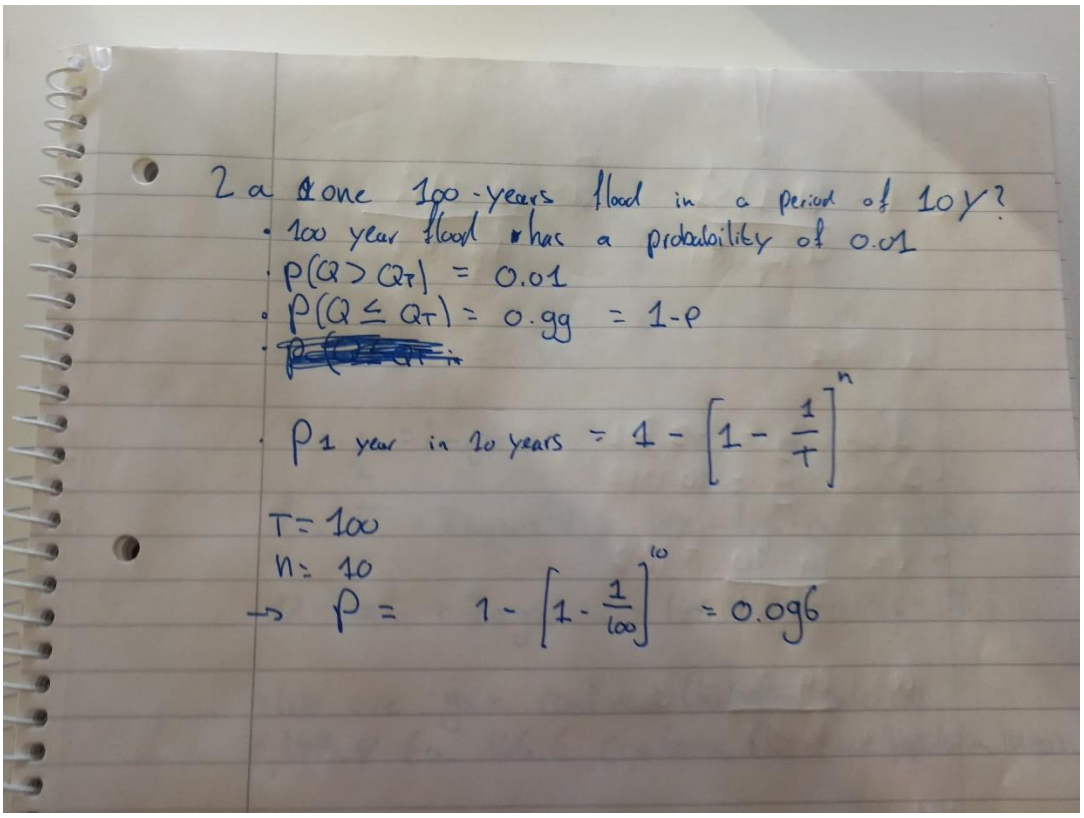
GEO4300 Exam

Question 1



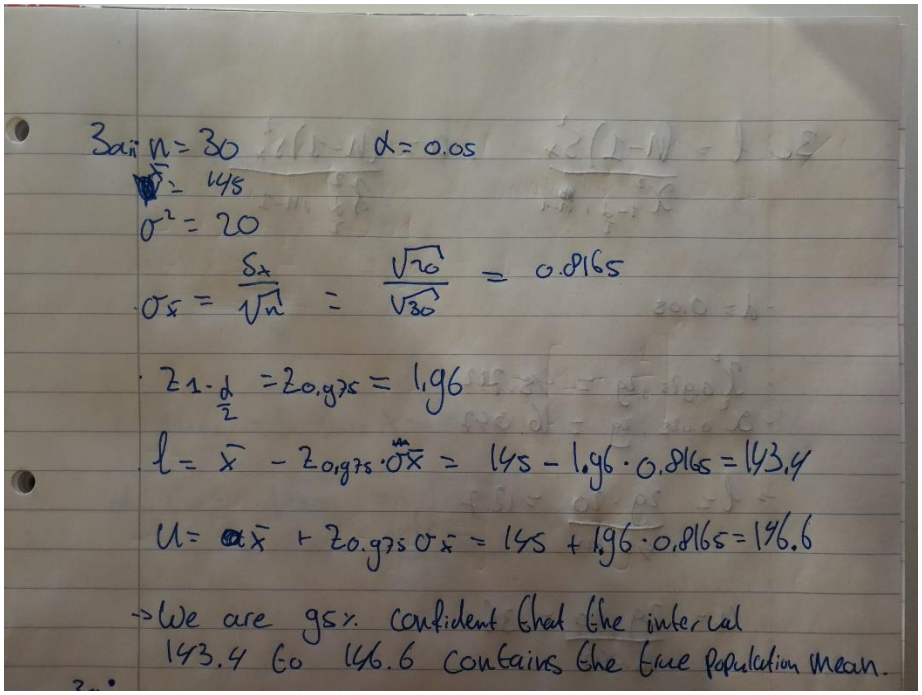


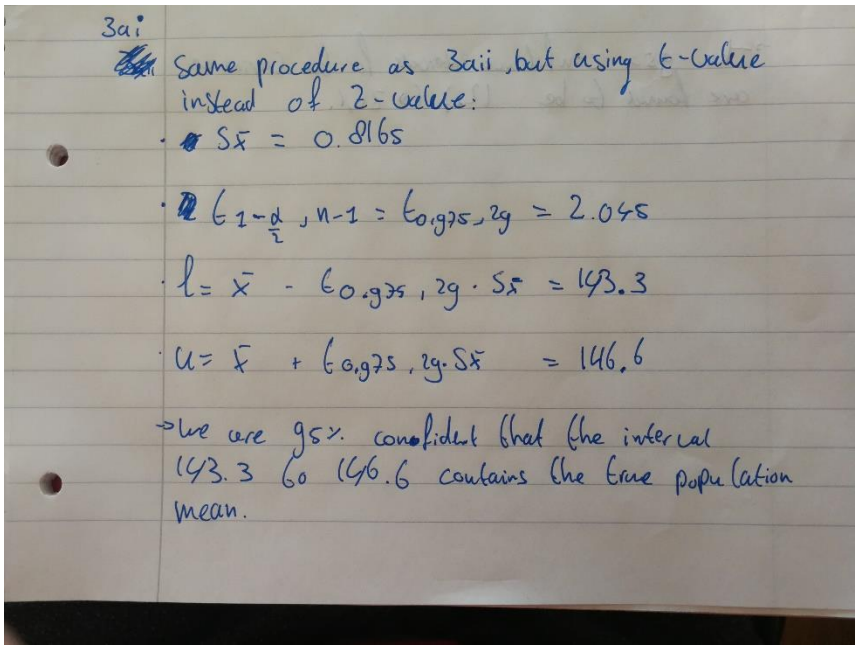
Question 2



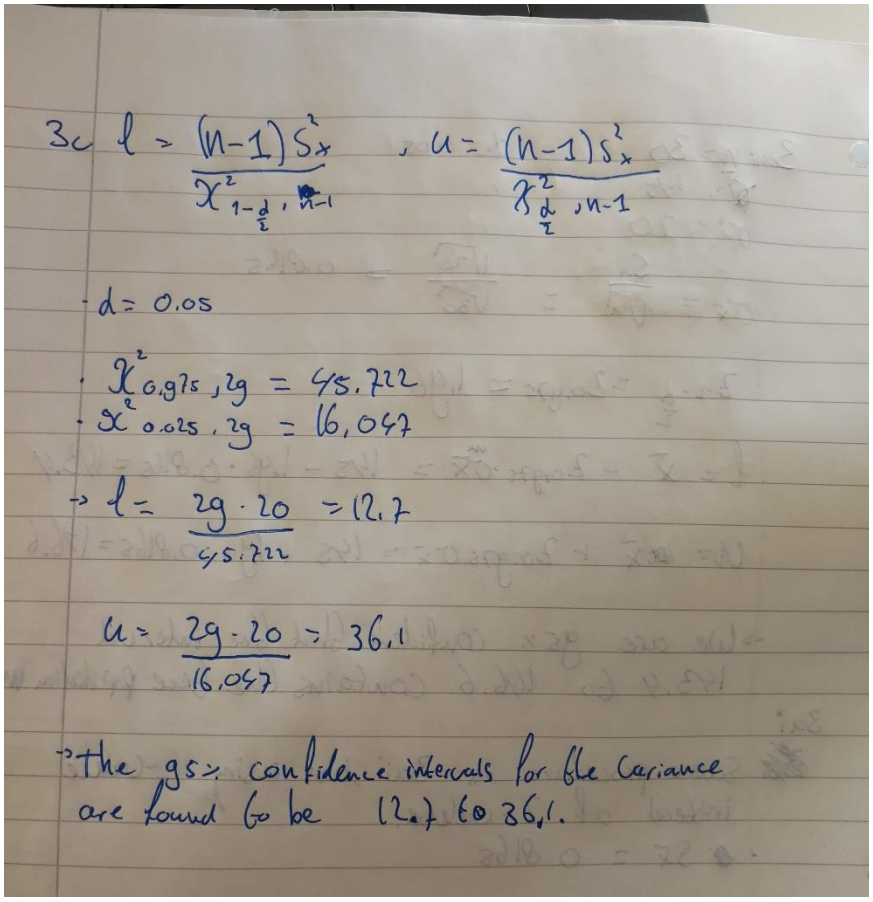
2b) The assumption of normality is violated in this analysis. The data have more extreme values than would be expected from a normal distribution. The use of a lognormal distribution (or another heavy-tail distribution) instead of a normal distribution can improve the analysis.

Question 3 (I did 3aii before 3ai, therefore the referring to 3aii in 3ai)





3b) The reason for the difference is that in part i we assume the true variance is not known and therefore assume a t-distribution with n-1 degrees of freedom instead of the z-value that we use in part ii with infinite degrees of freedom.



4a) In machine learning, a model is trained on the data in the training set. The best possible fit for this data is constructed. The prediction error is estimated with a loss function (i.e. the mean squared error). This error is the training error. Then the model is used to predict the outcome of the test set. The test set is NOT USED in order to construct the model. Again the prediction error (now for the test set) is estimated. This is the test error. The test error is a good estimation of how the model performs in practice (because the data is not used to construct a learner).

4b) We want to control model complexity because a too complex model might lead to overfitting. This means the model predicts the training data very well (a low training error), but performs poorly on the test data (high test error). And since the goal is to get a model with a low test error, we need to control the model complexity.

5a) We can fit a linear regression to the time series and then perform a t-test to test if the slope, beta, of the regression is significantly different from 0. The null hypothesis is that there is no trend. The alternative hypothesis is that there is a trend. The test statistic is:

$$t = \frac{\beta - 0}{S_\beta}$$

, and the critical value is $t_{1-\alpha/2,n-2}$

Sb is the standard deviation of beta. The null hypothesis is rejected it if |t| > $t_{1-\alpha/2,n-2}$

5b) The times series is not a perfect sine or cosine wave, so it is not C. From the times series there seems to be a strong periodicity of around 5 s. 1/5 = 0.2 hz. Graph A shows a peak at f = 0.2 hz, so graph A shows the Fourier transform of X_t.