

Final Assignment Biostatistics 2016

All calculations, unless specified otherwise, must be done using the derived equations presented in class. You cannot present your work based on using built-in linear regression functions. You may however use built-in linear regression functions to check that your manually derived results are correct.

Calculations may be done in Excel, a Calculator, Python, Matlab etc. Whatever you use, indicate what equations you used and present your code, screenshots (non-fuzzy) in the case of Excel, or a description of your workings if done by calculator.

1. State three assumptions made in unweighted linear regression?
2. Consider the following data that was collected over a period of 6 months. The data shows the price of a generic drug based on bulk quantities and appears to indicate a linear relationship between the two variables. Let the weight of generic drug purchased be the x variable and the spot price be the y variable.

Kg of drug purchased	30	28	32	25	25	25	22	24	35	40
Spot prices	25	30	27	40	42	40	50	45	30	25

Use the data to answer the following questions:

- a) Determine the best line fit through the data, show the plot and quote the best slope and intercept.
- b) Compute the SSE and MSE. Explain what the MSE measures.
- c) Determine the 95% confidence limits on the slope of the best line
- d) Test the hypothesis that there is no relationship between the two variables.
 - i) State H_0 and H_1
 - ii) Use a t-test and a significance level of 0.05 (95%)
 - iii) **Use a calculator or computer** to compute the exact p-value and show how you did this.
- e) Determine the 95% confidence limits on the y-intercept of the best line
- f) Compute the correlation coefficient of the line
- g) If 20 Kg of drug is purchased predict the corresponding spot price. Using error propagation and the standard errors of the slope and intercept computed previously, estimate the error in the predicted price.