

EBBRMS Statistics Assignment – Resit

This assignment is worth 30% of the mark for EBBRMS for those resitting EBBRMS. There is one question in 3 parts: 1a, 1b, 1c.

NOTE: for Q1a you are expected to follow and report on each stage of the **Data Analysis Algorithm** shown on page 1 of Workbook 3, and ensure you cover both confidence interval and hypothesis testing approaches. Make sure that you address the clinical relevance of the results, as well as their statistical significance.

Q1

In a study of 251 men, investigators were interested in exploring the relationship between percentage of body fat and several simple body measurements. Their objective was to ascertain which of the measurements could be used to predict percentage of body fat. Here we will consider one of these measurements: weight in kilograms.

The data are in the dataset **bodyfat.xlsx**.

Variables in bodyfat.xlsx:

pcf	Percentage body fat
kg	Body weight in kilograms.

a) Investigate whether and how body weight is related to the percent bodyfat. Specifically, carry out a linear **regression** analysis with pcf as the response and kg as the predictor.

[21 marks]

[Either Minitab or SPSS can be used to do this.

HINTS: Ensure that you –

- remember that you are carrying out a regression analysis and therefore need to look carefully at the lecture notes, workbook and workbook answers on Regression and Correlation;
- give a subjective impression with the help of suitable graphs and, if you wish, an appropriate summary statistic [4 marks];
- check assumptions underlying the regression in your model [9 marks];
- give and interpret the results from your model. Also calculate and interpret the 95% confidence interval for the coefficient of the weight in your model (use a calculator [even though you can get these from Minitab and SPSS]) [8 marks].

NB: the t-distribution critical value you will need for your confidence interval will be = 1.97]

b) What are the degrees of freedom for the t-distribution used to calculate the confidence interval for weight's coefficient in part a) above?

[1 mark]

c) Produce a scatterplot of the data showing the fitted regression line, as well as the confidence intervals and prediction intervals for that line. Comment on the suitability of using bodyweight to predict percentage bodyfat.

[3 marks]

[It's easier to do this in Minitab: Stat > Regression > Fitted Line Plot, put in the appropriate Response and Predictor, choose a Linear regression model; then click on Options and tick on Display confidence interval and Display prediction interval.

It can be done in SPSS by creating the scatterplot and then editing the graph, but you will only be able to do one set of intervals at a time. Once you have created the scatterplot, double click it to edit it, choose Elements > Fit Line At Total; then select a Linear fit and the Mean confidence interval to get the confidence interval for the line and the Individual one to get the Prediction Interval.]