ASSIGNMENT 2 — Regression

For all questions below, provide all programming code and plots in the report. Unless stated otherwise, assume $\alpha = 0.05$

Regression

1. Find the partial derivatives ($\frac{\partial SS}{\partial B_i}$) for $\hat{y}_i = B_0 + B_1 \cdot x_{i1} + B_2 \cdot x_{i2}$. Show your work (you can take a picture of your hand writing with your phone) (3 marks).

2. Evans et al. examined the effect of velocity on ground reaction forces (GRF) in dogs with lameness from a torn cranial cruciate ligament. The dogs were walked and trotted over a force platform and the GRF recorded (in newtons) during the stance phase. The following table contains 22 measurements of force expressed as the mean of five force measurements per dog when walking (X) and the mean of five force measurements per dog when trotting (Y). 8 Marks.

GRF-Walk	GRF-Trot	GRF-Walk	GRF-Trot
31.5	50.8	24.9	30.2
33.3	43.2	33.6	46.3
32.3	44.8	30.7	41.8
28.8	39.5	27.2	32.4
38.3	44.0	44.0	65.8
36.9	60.1	28.2	32.2
14.6	11.1	24.3	29.5
27.0	32.3	31.6	38.7
32.8	41.3	29.9	42.0
27.4	38.2	34.3	37.6
31.5	50.8	24.9	30.2

Source: Data provided courtesy of Richard Evans, Ph.D.

a. Plot the data. (1 mark)

b. Report B_0 and B_1. (1 mark)

c. Plot the line of best fit. (1 mark)

d. Calculate the Standard Error of Estimate. (1 mark)

e. Find the Correlation Coefficient. (1 mark)

f. Find the Coefficient of Determination. (1 mark)

g. Find the p-value. (1 mark)

h. Interpret the findings. (1 mark)

3. Let's use an example from the probability assignment since linear regression follows a bivariate normal distribution. The average height and weight of the 2016 San Antonio Spurs is 78.8 (SD = 3.668) inches and 211 (SD = 25.904) lbs, respectively. The correlation between height and weight is $r = 0.81$. TIP: Make sure to define the covariance matrix using the bivariate normal distribution Σ. 8 Marks.

a. Sample 1000 data points drawn from a bivariate (joint) normal distribution and show the data with a scatter plot. (1 mark)

b. Plot the data. (1 mark)

c. Report B_0 and B_1. (1 mark)

d. Plot the line of best fit. (1 mark)

e. Find the Correlation Coefficient. (1 mark)

f. Find the Coefficient of Determination. (1 mark)

g. Find the p-value. (1 mark)

h. Interpret the findings. (1 mark)

4. You are asked to determine if there is a relationship between the average amount of hours of sun per day and the risk of skin cancer. Here are the average hours spent outside [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] and respective probability of skin cancer [3.27, 6.24, 7.56, 2.89, 2.92, 7.89, 9.85, 18.94, 16.93, 26.65]. 4 Marks.

a. Plot the data. (1 mark)

b. Which bivariate test should you use to model this data? Justify your decision. (1 mark)

c. Find the p-value. (1 mark)

d. Interpret the findings. (1 mark)

5. Let’s assume we want to develop a model that predicts a players free throw percentage (FTP) based on the other 6 variables [Games: number of games played in previous season; PPM: average points scored per minute; MPG average minutes played per game; HGT: height of player (centimetres); FGP: field-goal percentage; AGE: age of player (years)]. First, read in the data from online using the following code (8 Marks):

```
bball_data <- read.csv("https://raw.githubusercontent.com/joshcash9/Statistics_BME/master/bball.csv")
```

a. Plot the data. (1 mark)

b. Print the correlation coefficient matrix. (1 mark).

c. What is the single best predictor of FTP. (1 mark)

d. What is the best model to predict FTP. (1 mark)

e. Does the best model significanlty explain FTP. (1 mark)

f. Interpret the results. (1 mark)

g. Write the equation of the best fit model. (1 mark)

h. We have a new recruit with the following stats: GAMES=64, PPM=0.4, MPG=20, HGT=200, FGP=60, AGE=23. What is our best guess of what his FTP will be. (1 mark)