BIOS 662, Fall 2018

Homework 2

Assigned: Tuesday, September 4

Due: Tuesday, September 11

1. In the "Datasets" sub-folder of the "Homework materials" folder under "Resources" on the Sakai web site for this course, there is a dataset "HW2_SBP.txt". The data are systolic blood pressures of 40 women who had an MI (myocardial infarction, i.e. heart attack) less than two years after their blood pressure was measured and of 160 women who did not have an MI within two years. Each row in the dataset corresponds to one woman. The first observation in a row is 1 for those who had an MI within two years and 0 otherwise. The second observation is the systolic blood pressure.
   (a) Use R or SAS to draw a histogram and boxplot of systolic blood pressure for all 200 women (that is, do not separate those who did and did not have an MI).
   (b) Using the definition of percentile from the class notes, compute the 25th, 50th (i.e., median) and 75th percentiles.
   (c) Determine the IQR.
   (d) Find the largest observation $\leq 75^{\text{th}}$ percentile + 1.5 IQR and the smallest observation $\geq 25^{\text{th}}$ percentile $-$ 1.5 IQR (i.e., the extent of the "whiskers"). Based on these results, does the computed boxplot appear to agree with the definition of a boxplot from our notes? If not, investigate the discrepancy and report your findings.
   (e) Use a plot to compare the distribution of systolic blood pressure in those who had an MI against that of those who did not. Do blood pressures in the two groups appear to differ? If so, in what direction?

2. The dataset "HW2_PGE.txt" in the "Datasets" sub-folder contains the data in Table 3.20 of the textbook. The last value in each row is 1 for the patients with hypercalcemia and 0 otherwise.
   (a) Obtain the mean and standard deviation of plasma iPGE separately for patients with and without hypercalcemia. Do you think there is enough evidence to conclude that the means of the two groups differ? (Later in this course we will study more formal ways to compare the two means.)
   (b) Do part (c) of Problem 3.15 of the textbook.
   (c) The values for one patient appear to be particularly anomalous. Identify this patient. Suppose it was determined that there had been an

error in measuring the patient's serum calcium. Suggest a value for serum calcium that would be more consistent with the patient's plasma iPGE value and the pattern in the rest of the data.

(d) Without re-doing any of your calculations, what effect do you think changing the serum calcium value to the one you suggested would have on the means and standard deviations in the first part of this problem?