Name ______________________________ October 5, 2012

Math 445 - Statistical Methods Long Exam I

Instructions: *For p-values, write exactly what you get from R (do not round it off). For confidence intervals, all lower and upper limits should be round off to 2 decimal places.*

1. Consider the Health Exam data from the U.S. Department of Health and Human Services, National Center for Health Statistics, Third National Health and Nutrition Examination Survey. It has a total of 80 cases (40 males and 40 females) with each case having values for 14 variables. These variables are: Gender, Age (in years), Height (in inches), Weight (in pounds), Waist (circumference in cm.), Pulse (pulse rate in beats per minute), SysBP (systolic blood pressure in mmHg), DiasBP (diastolic blood pressure in mmHg), Cholesterol (in mg), BodyMass (body mass index), Leg (upper leg length in cm), Elbow (elbow breadth in cm), Wrist (wrist breadth in cm), Arm (arm circumference in cm).

a. What is the mean age of the 80 people in the sample? [3]

b. Based on the boxplot for Age, determine if there is an outlier. If there are, list them. [3]

c. Looking at the scatterplot between Age and Height, would you say that they are linearly related? [3]

d. Determine if cholesterol level is linearly related to weight by computing the coefficient of determination, R^2. Interpret that meaning of the value of R^2 in the context of this problem. Would it be appropriate to use linear regression to estimate the cholesterol level of a person based on the weight of the person? [5]

e. Looking at the scatterplot between Waist and Weight, would you say that they are linearly related? Compute and interpret the coefficient of determination, R^2. [5]

f. If we decide to model the weight of a person based on his/her waistline, determine the (least-squares) regression line. Does β_0 have a meaningful interpretation in this context? Briefly explain. [6]

g. Using the regression line you obtained in part (f), predict the weight of a person with a waistline of 100 cm. Construct a 95% prediction interval for Weight for a person with waistline of 100 cm. [5]

h. Construct and interpret a 99% confidence interval for the mean weight of people with waistline of 100 cm. [5]

i. Based on the regression line you obtained in part (f), what would you expect to happen to someone's weight when his/her waistline increases by 1 cm? [4]

j. Based on the regression line you obtained in part (f), test the null hypothesis that the slope is zero versus the alternative hypothesis that the slope is not equal to zero. Give the observed value of the t-statistic, of SE_{b1}, and the corresponding p-value. Using a level of significance of 0.05, write a conclusion in the context of this problem. [6]

k. Repeat part (j) using the ANOVA approach. Construct that ANOVA table. What is the p-value? Write your conclusion. [6]

l. Repeat part (j), but this time, test the null hypothesis that the slope is equal to one versus the alternative hypothesis that the slope is greater than one. Give the observed value of the t-statistic and the corresponding p-value. Using a level of significance of 0.05, write a conclusion in the context of this problem. [6]

m. Estimate the Pearson's correlation coefficient between Waist and Weight. Construct and interpret a 95% confidence interval for the true correlation coefficient between Waist and Weight. [5]

n. Compute the Spearman's correlation and Kendall tau between Waist and Weight. [4]

o. In the model of part (f), give an estimate of the standard deviation of the random error, ϵ. [3]

p. What assumptions about the random error (ϵ) are required for the t-test that you used in part (j) to test the null hypothesis that the slope is zero? [4]

q. Use the appropriate formal test to check the normality of the residuals of the linear model you used in part (f). Give the observed value of the statistic, the corresponding p-value, and write the appropriate conclusion. Use $\alpha = 0.05$. [4]

2. Using the method of least-squares, derive the formula for the slope of the regression line. That is, show that $b_1 = \frac{SS_{xy}}{SS_{xx}}$. [20]