

HOMEWORK ASSIGNMENTS**Homework #1****Assigned: 10/27/09****Due: 10/30/09**

#1. A food product is supposed to contain 75 mg of sodium per serving. However, servings might contain more than 75 mg of sodium. The data below shows a random sample of 30 servings

72.5 72.4 72.9 73.6 73.3 73.2 73.9 74.1 74.3 74.1
74.3 74.6 74.6 75.1 75.2 75.2 75.1 75.2 74.9 74.4
75.5 76.2 76.5 76.3 75.6 76.7 77.3 77.2 77.3 77.2

Construct a 95% confidence interval to assess the situation. [5 points]

Homework #2**Assigned: 10/29/09****Due: 11/06/09**

#1 Exercise 6.8, page 333, of the PNC text book.

#2. Use the data on cherry trees (sasdata.trees) and the SAS software to draw scatterplots of all possible pairs of the three variables (volume, diameter and height). Use Pearson correlation coefficient to quantify the strength of the association between each pair. Which pair of variables has the largest correlation coefficient? [Use the multivariate option under ANALYSIS.]

#3. The following observations are obtained from a bivariate normal population.

x	100	120	119	130	135	125	130	126	121	123
y	60	75	70	80	81	76	79	90	73	69

- (i) Draw a scatterplot of y versus x and describe its main features.
- (ii) Find the Pearson correlation coefficient between x and y . Is this consistent with the scatterplot?
- (iii) Test the hypothesis of no association versus a positive association between x and y .

#4. A random sample of 103 people is taken from a population, and two variables are measured. The Pearson correlation coefficient (r) is calculated as $r = .80$.

- (i) Find the approximate p-value of the test that the population correlation coefficient (ρ) is larger than .75.
- (ii) Find an approximate 95% confidence interval for the population correlation coefficient (ρ).
- (iii) Explain the similarity in (i) and (ii).

Homework #3

Assigned: 11/06/09

Due: 11/13/09

#1. Consider the data set, sasdata.height, of a random sample of 58 children from age 2 months to 100 months.

- (i) Obtain the fitted regression straight line of height on age. [3 points]
- (ii) Obtain the fitted values and ordinary residuals for this fitted line. [4 points]
- (iii) Plot the ordinary residuals versus age, ordinary residuals versus observed height, and ordinary residuals versus fitted height. Discuss these plots. [8 points]

#2. A scientist fitted a SLR model to his data (x_i, y_i) , $i = 1, \dots, n > 2$. He did not present the n points; rather he summarized the data.

- (i) Let $\hat{y}_i$ and $\hat{e}_i$ denote the ordinary residuals and the fitted values for the n data points. Give a mathematical proof to show that $(\hat{y}_i, \hat{e}_i), i = 1, \dots, n$ are not correlated. [5 points]
- (ii) Suppose the summaries are $\bar{x} = 10$, $S_x^2 = 4$, $\bar{y} = 7$, $S_y^2 = 9$ and the Pearson sample correlation $r = .5$. Find the fitted regression line. Interpret the slope parameter $\hat{\beta}_1$. [10 points]

#3. Consider the data set, sasdata.height, of a random sample of 58 children from age 2 months to 100 months. Fit the SLR model and use the studentized deleted residuals to check the normality assumption in the SLR model. [5 points]

#4. (a) Suppose two random variables x and y are negatively correlated, and in the SLR model R -square is .36. Find the Pearson correlation coefficient r . [5 points]

(b) Consider the data set, sasdata.height, of a random sample of 58 children from age 2 months to 100 months. For the SLR model, write down the residual sum of squares and the regression sum of squares. For the C & E model write down the residual sum of squares. [All these quantities are in the SAS output of the SLR model fit.] Use these numbers to show that R -square in the SAS output is correct. [5 points]

Homework #4**Assigned: 11/13/09****Due: 11/18/09**

#1. Consider the data set, `sasdata.height`, of a random sample of 58 children from age 2 months to 100 months. Fit the SLR model as you did in HW #3, exercise #4 (b).

(a) Obtain 95% confidence intervals for β_0 and β_1 . [10 points]

(b) Obtain 95% prediction intervals for (i) the mean height of all children age 110 months, and (ii) the height of a child age 110 months. [10 points]

#2. A surgeon wants to determine whether administration of a certain drug can reduce the incidence of pulmonary embolism in patients undergoing high-risk surgery. Fifty patients were selected for the study, with 20 getting the drug and 30 the standard treatment. The counts are shown in the table below. Write down the statistical hypothe-

	P.E.	No P.E.
Drug	5	15
No Drug	12	18

Table 1: *Incidence of Pulmonary Embolism (P.E.)*

ses, null and alternative, of interest. [2 points] Calculate the Pearson residuals. [3 points] Use the Pearson residuals to obtain a chi-square test to help the doctor understand this phenomenon. [4 points] Comment on your answer. [1 point]

#3. Exercise 7.20, page 427, of the PNC text book. [5 points]

Homework #5**Assigned: 11/18/09****Due: 12/4/09**

#1. Consider the problem on fire endurance, described in Exercise 8.11, page 495, of the PNC text book. The data is in [sasdata.fire5](#).

(a) Draw graphs to show dependence among the variables. Compute Pearson correlations for all pairs of variables. Say whether the correlation compare well with your graphs. [10 points]

(b) Fit an appropriate multiple regression model.[FIRE.END is related to three other variables.] Be sure to draw the diagnostic plots we discussed in class. [10 points]

(c) Make a table of R-squares and adjusted R-squares for all possible combinations of the regressors. [Begin with the additive model, and add all possible interactions.] Which is the best model? Explain. [10 points]

#2. In a MLR model with 25 observations and 3 regressors, $SSR = 100$ and $SSE = 10$. Compute the following

(a) R-square and adjusted R-sqaure; [5 points]

(b) P-value for the F-test with the null hypothesis $\beta_k = 0$, $k = 1, 2, 3$; [5 points]

(c) Compare $\hat{\sigma}^2$ for the C & E model and this MLR model. [5 points]

#3. Collar size (C) and shoe size (S) are used to predict body weight (W). For a random sample of 24 college-age men, the response surface is

$$\hat{w} = 0.167 + 3.09 \times C + 11.533 \times S.$$

The standard errors of $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, are 50.981, 3.134, 2.346. Also $\hat{\sigma}^2 = 18.738$

(a) Find 95% confidence intervals for β_0 , β_1 and β_2 . [5 points]

(b) Find the p-value for the $H_O : \beta_2 = 3$ versus $H_O : \beta_2 > 3$. [5 points]

(c) We need the predictive weight for colar size of 15 and shoe size of 11.5. The standard error for this prediction for a single person is 5.238 for these measurements. Find the 95% prediction interval for this person, and for the mean response. [5 points]

#4. Consider the data in Exercise 6.3, page 332, of the PNC text book. Interest is on the difference between the two shipments.

(a) Explain the assumptions you need and how to check that they hold. [5 points]

(b) Find the p-value of a hypothesis test to check if the first shipment has more carbon than the second. [5 points]

(c) Use a 95% confidence interval to estimate the difference btween the carbon contents of the two shipments.[5 points]

Homework #6**Assigned: 12/4/09****Due: 12/15/09**

#1. Exercise 9.3, page 565, of the PNC text book. [5 points]

#2. Exercise 9.5, page 566, of the PNC text book. [15 points]

#3. A scientist performed an experiment to compare 4 treatments and she used 4, 5, 3, 6 subjects for treatments 1, 2, 3, 4 respectively. The observed data gave sample means 10, 15, 12, 16 and variances 3, 1, 2, .90 respectively. Assume that the oneway means model holds.

- (a) Find a 95% confidence interval for the difference between treatments 1 and 2 if you consider only the data from these two treatments. [5 points]
- (b) Find a 95% confidence interval for the difference between treatments 1 and 2 if you consider the data from all 4 treatments. [5 points]
- (c) Find a 95% simultaneous confidence interval for the difference between treatments 1 and 2 using Bonferroni and Tukey's method. [5 points]
- (d) Comment on your answers in (a), (b) and (c). [5 points]

#4. Exercise 9.11, page 568, of the PNC text book. [20 points]

#5. Exercise 9.12, page 568, of the PNC text book. [5 points]

#6. A food product is supposed to contain 75 mg of sodium per serving. However, servings might contain more than 75 mg of sodium. The data below shows a random sample of 30 servings

72.5 72.4 72.9 73.6 73.3 73.2 73.9 74.1 74.3 74.1
74.3 74.6 74.6 75.1 75.2 75.2 75.1 75.2 74.9 74.4
75.5 76.2 76.5 76.3 75.6 76.7 77.3 77.2 77.3 77.2

Consider testing $H_0 : \theta = 75$ versus $H_1 : \theta > 75$, where θ is the median of the population of all servings.

Use SAS to perform the sign test and the Wilcoxon signed rank test procedures. Perform the the large sample sign test and Wilcoxon signed rank test using your own calculations. [10 points]