

Engineering Statistics II, HW 3

Due Start of class, Friday, Sept. 19

Instructions: This homework is due at the beginning of class Friday, Sept. 19. A subset of these problems will be graded.

1. Late hw's are not accepted.
2. Pages are to be stapled. Hw's with unstapled pages will not be accepted.
3. All work is to be placed in order in your hw, i.e., all work for problem 1 should be at the front, followed by all work for problem 2, etc.
4. You must turn in hardcopy; e-mailed HW will NOT be accepted.
5. You are to include all related Minitab output, especially graphics, in your hw. **All work and output for a given problem are to be placed together.** To save paper, please cut-and-paste multiple Minitab graphs onto one Word page before printing them out.
6. Non-Minitab work, e.g., mathematical calculations and derivations, may be neatly handwritten.
7. You are encouraged to work together on homeworks but the finished product must be in your own words/style, e.g., if three of you do the HW together, three identical copies (except for the name) will not be acceptable.
8. Be sure to put your mailbox number on your HW.

If you have *any* questions about the preceding or the questions below, please contact me as soon as possible.

0: Be sure to read the IID assumption handout plus sections 11.1-11.2, 11-7, 12.1.1 and 12.1.3 before doing the following.

1. If you have not already done so, enable macros in your version of Minitab and install the macros `tsplot` and `lagplot`. Then open the Minitab worksheet for this problem on the course website and do the following:

i. Construct a time series plot and a lag plot of the data in column C1 by issuing the following commands after the "MTB>" prompt in the top (session) window:

```
%tsplot c1.  
%lagplot c1.
```

ii. Include copies of both graphs in your hw. To save paper, copy and paste these two graphs into Word and resize them to fit on one page.

iii. See next page ...

- iii. Are there any trends in the data with respect to location or variability? Are the data identically distributed?
- iv. Do the observations appear to be independent or dependent?
- v. Based on your answers to iii and iv, do the data appear to satisfy the IID assumption?
- vi. Repeat parts i-v for the data in column c2.
- vii. Repeat parts i-v for the data in column c3.

2. Open the reaction time data set at the bottom of the course website and assess whether or not student 6's reaction times (column c7) are stationary by construction and analyzing a time series plot/run chart and a lag plot using the corresponding Minitab macros.

3. In your introductory statistics class you should have learned that a simple random sample of size n is a *sample of size n collected in such a way that each subset of size n of the population sampled has an equal chance of being the sample*. In the card game Red-Black each player is dealt a 4 card hand in the following manner. The deck is split in half according to color with the 26 red cards in one stack and the 26 black cards in the other. Both stacks of cards are thoroughly shuffled then the first player's hand is dealt by drawing the first two cards off the red card stack and then the black card stack. This procedure is repeated for the remaining players so that, initially, each player's hand consists of two red and two black cards.

- i. Is the hand dealt to the first player a simple random sample? (yes/no)
- ii. Justify your answer in part 1.

4: Fit the simple linear regression line $y = \beta_0 + \beta_1 x + \epsilon$ to the following (x, y) data points

x:	1	2	3
y:	2	5	8

by manually computing - no Maple - the least squares estimates of β_0 and β_1 using the matrix formula $\vec{\beta} = (X'X)^{-1}X'Y$ as follows. Show all your work for full credit.

- i. Determine the design matrix X for this data set.
- ii. Compute $(X'X)^{-1}$, $X'\vec{y}$, and then multiply them to get $\hat{\vec{\beta}} = (X'X)^{-1}X'\vec{y}$.
- iii. Check your values for $\hat{\beta}_0$ and $\hat{\beta}_1$ by fitting the regression model using Minitab. Be sure to include a copy of your Minitab output in your hw - either by copy-and-past or literal cut-and-paste.

5: Fit the simple linear regression line $y = \beta_0 + \beta_1 x + \epsilon$ to the following (x, y) data points

x:	-2	-1	1	2
y:	-2	-1	1	4

Analyze this data manually - no Maple - as follows, showing your work for full credit:

i. Determine the design matrix X for this data set.

ii. Compute $(X'X)^{-1}$, $X'\vec{y}$, and then multiply them to get $\hat{\vec{\beta}} = (X'X)^{-1}X'\vec{y}$.

iii. Using your values for $\hat{\beta}_0$ and $\hat{\beta}_1$, compute the residuals

$$\begin{aligned}e_i &= y_i - \hat{y}_i \\ &= y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i), \quad i = 1, \dots, n\end{aligned}$$

iv. Compute the least squares estimate of σ^2 ,

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n e_i^2$$

v. Compute the standard deviations (standard errors) of $\hat{\beta}_0$ and $\hat{\beta}_1$ (recall that these are the square roots of the diagonal entries of $\hat{\sigma}^2(X'X)^{-1}$).

vii. Verify your values of the least squares coefficient estimates ($\hat{\beta}$'s) and their standard deviations by fitting the regression model using Minitab. Include a copy of the relevant Minitab output in your hw.

6: Duplicate the book's regression analysis of the oxygen purity data by opening the data set for this problem on the course website and then doing the following:

i. Analyze the data using Minitab's regression procedure to fit the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

(Equation 11-1, page 403 in your book.) Be sure to request that residuals be computed by selecting **Residuals** under the **Storage** menu. Here we want to model the effect of hydrocarbons in the main condenser of the distillation unit on the purity of oxygen produced. Thus oxygen purity (y) is the response variable and hydrocarbon level (x) is the predictor variable; see page 403 for more details.

ii. Compare your output with the book's Minitab output (Table 11-2, page 409). Although the formatting is different due to Minitab 16 vs. 17, your results should agree. Include a copy of your regression output in your hw. Clearly identify the least squares estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ plus their standard errors. Also identify the estimate of σ - the standard deviation of the random errors.

iii. See next page ...

- iii.** Check the first two regression random error assumptions by making a scatter plot of e_i vs. x_i ; your plot should resemble figure 11-12 on page 429. Include a copy of this plot in your hw.
1. $\mu_{\epsilon_i} = 0$: If this assumption is met then the residuals when plotted vs. the predictor variable should exhibit no systematic trends about 0. Is this assumption met?
 2. $\sigma_{\epsilon_i} = \sigma$ (homoscedasticity): If this assumption is met then the residuals (when plotted vs. the predictor variable) should exhibit no systematic trends in spread. Is this assumption met?
- iv.** The third regression random error assumption is that the random errors are uncorrelated. If the observations were in order - they are not - we could check this by making a lag plot. Since we don't know the order of the observations can we say anything about this assumption? Often we can assess the reasonableness of this assumption by considering how the data are collected. Since we don't have that information, we can't do that for this data.
- v.** The fourth assumption is that the random errors are normally distributed. Check this assumption by making a normal probability plot of the residuals. Your plot should resemble figure 11-10, page 428. What do you conclude about this assumption?

7: The data for this problem consists of three variables: the monthly steam consumption in lbs (steamlb), the average monthly temperature in degrees Fahrenheit (avgtemp), and the month (month). The data was collected over a period of 25 consecutive months. The objective of this analysis is to construct a regression model to predict monthly steam consumption using average monthly temperature. Analyze the data using Minitab as follows:

- i.** Download the data from the course website and construct a scatter plot of the data.
- ii.** Fit the simple linear regression model

$$\text{steamlb}_i = \beta_0 + \beta_1 \text{avgtemp} + \epsilon_i$$

and compute the residuals. Be sure to include a copy of your Minitab regression output in your hw.

- iii.** Analyze the residuals and assess the four regression assumptions. Note that since the order of the data is known, you can check for lack of independence/autocorrelation by constructing a lag plot. Do the regression assumptions appear satisfied? Be sure to include all relevant Minitab output (graphs, p-values) in your hw.
- v.** What are the least squares estimates of β_0 and β_1 , i.e., $\hat{\beta}_0$ and $\hat{\beta}_1$, for your model?
- vi.** What are the standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$?
- vii.** What is the least squares estimate of σ_ϵ ?
- viii.** Compute the regression model's estimate of steam consumption (in steam lbs) for an average monthly temperature of 50 degrees.