## STAT 305 D Homework 3

## Due February 7, 2012 at 12:40 PM in class

1. Here is dataset giving the number of printer jams per day of a receipt printer in a supermarket for 17 days.

$$
43,100,500,23,89,89,89,89,72,72,72,21,32,41,39,47,56
$$

- Identify the sample.
- Identify the population.
- Calculate the sample mean.
- Calculate the median.
- Calculate the mode.
- Calculate the sample variance.
- Calculate the sample standard deviation.
- Calculate the range.

2. Chapter 3 section 2 part of exercise 1 (page 92 ):

The following are data (from Introduction to Contemporary Statistical Methods by L.H.Koopmans) on the impact strength of sheets of insulating material cut in two different ways. (The values are in ft lb .)

| Lengthwise Cuts | Crosswise Cuts |
| :---: | :---: |
| 1.15 | .89 |
| .84 | .69 |
| .88 | .46 |
| .91 | .85 |
| .86 | .73 |
| .88 | .67 |
| .92 | .78 |
| .87 | .77 |
| .93 | .80 |
| .95 | .79 |

For the lengthwise sample, $Q(0.25)=0.870, Q(0.5)=0.895$, and $Q(0.75)=0.930$. For the crosswise sample, $Q(0.25)=0.690, Q(0.5)=0.775$, and $Q(0.75)=0.800$.
a. Find the IQR.
b. Draw (to scale) carefully labeled side-by-side boxplots for comparing the two cutting methods. Discuss what these show about the two methods.
c. Make and discuss the appearance of a Q-Q (quantile-quantile) plot for comparing the shapes of these two data sets.
3. Below is a normal quantile plot of the yield data from problem 1 (Chapter 3 Section 1 Exercise 1, page 77). What does this plot tell you about the distributional shape of the yield data? Justify your answer based on the shape of the points in the plot.

4. You work for one of the main manufactures of fire rescue harnesses in the United States. In the news one day, you read about a firefighter who was using one of your companys harnesses in a rescue and suffered an accident because a clasp on his harness came undone. You contact the firefighter, obtain the harness that failed, and determine that the bad clasp was originally faulty when it came out of the plant in Gary, IN. Concerned that other harnesses may have faulty clasps, you consult your supervisor. He informs you that he is already investigating the problem, and he shows you the following bar chart:

Harnesses Produced in Jan. 2012 by the Plant in Gary, IN


1. Is the above plot an honest representation of the data? Why or why not? If there is there anything misleading about the plot, what is it?
2. Should the company arrange a recall of the harnesses produced in Jan. 2012 by the Gary plant, or is a recall unnecessary?
3. Vardeman and Jobe chapter 4 section 1 problem 3 parts a-c (page 140). Data can be found at http:
//www.will-landau.com/stat305/data/csv/polypolyols.csv.You're welcome to use a spreadsheet program to do this problem, but please show your calculations. If you use Excel formulas, for example, please write them down. The article Polyglycol Modified Poly (Ethylene Ether Carbonate) Polyols by Molecular Weight Advancement by R. Harris (Journal of Applied Poly- mer Science, 1990) contains some data on the effect of reaction temperature on the molecular weight of resulting poly polyols. The data for eight experimental runs at temperatures $165^{\circ} \mathrm{C}$ and above are as follows. Here, $x$ is pot temperature $\left({ }^{\circ}\right)$ and $y$ is average molecular weight.

| x | y |
| ---: | ---: |
| 165.00 | 808.00 |
| 176.00 | 940.00 |
| 188.00 | 1183.00 |
| 205.00 | 1545.00 |
| 220.00 | 2012.00 |
| 235.00 | 2362.00 |
| 250.00 | 2742.00 |
| 260.00 | 2935.00 |

a. What fraction of the observed raw variation in y is accounted for by a linear equation in x ?
b. Fit a linear relationship $y \approx b_{0}+b_{1} x$ to these data via least squares. About what change in average molecular weight seems to accompany a $1^{\circ} \mathrm{C}$ increase in pot temperature (at least over the experimental range of temperatures)?
c. Compute and plot residuals from the linear relationship fit in (b). Discuss what they suggest about the appropriateness of that fitted equation. (Plot residuals versus $x$ and residuals versus $\widehat{y}$.)
6. Vardeman and Jobe chapter 4 section 1 problem 4 (page 140). Data can be found at http://www.will-landau.com/stat305/data/csv/tools.csv.

Upon changing measurement scales, nonlinear relationships between two variables can sometimes be made linear. The article "The Effect of Experimental Error on the Determination of the Optimum Metal-Cutting Conditions" by Ermer and Wu (The Journal of Engineering for Industry, 1967) contains a data set gathered in a study of tool life in a turning operation. The data here are part of that data set.

| Cutting Speed, $x(\mathrm{sfpm})$ | Tool Life, $y(\mathrm{~min})$ |
| :---: | :--- |
| 800 | $1.00,0.90,0.74,0.66$ |
| 700 | $1.00,1.20,1.50,1.60$ |
| 600 | $2.35,2.65,3.00,3.60$ |
| 500 | $6.40,7.80,9.80,16.50$ |
| 400 | $21.50,24.50,26.00,33.00$ |

(a) Plot $y$ versus $x$ and calculate $R^{2}$ for fitting a linear function of $x$ to $y$. Does the relationship $y \approx \beta_{0}+\beta_{1} x$ look like a reasonable explanation of tool life in terms of cutting speed?
(b) Take natural logs of both $x$ and $y$ and repeat part (a) with these $\log$ cutting speeds and $\log$ tool lives.
(c) Using the logged variables as in (b), fit a linear relationship between the two variables using least squares. Based on this fitted equation, what tool life would you predict for a cutting speed of 550 ? What approximate relationship between $x$ and $y$ is implied by a linear approximate relationship between $\ln (x)$ and $\ln (y)$ ? (Give an equation for this relationship.) By the way, Taylor's equation for tool life is $y x^{\alpha}=C$.
7. Weekly feedback. You get full credit as long as you write something.

1. Is there any aspect of the subject matter that you currently struggle with? If so, what specifically do you find difficult or confusing? The more detailed you are, the better I can help you.
2. Do you have any questions or concerns about the material, class logistics, or anything else? If so, fire away.
3. EXTRA CREDIT. Using the normal quantile plot below, draw the approximate distributional shape of the original data.

