Homework 2

Due May 30, 2017 in class

Please show all work for full credit. Print and staple your assignment together and submit by end of class the due date. If you cannot attend class on the due date, please arrange to submit your homework prior to the due date.

1. [Ch. 3.1 Exercise 3, pg. 77] Osborne, Bishop, and Klein collected manufacturing data on the torques required to loosed bolts holding an assembly on a piece of heavy machinery. The accompanying table (also available on the website as `bolts.csv`) shows part of their data concerning two particular bolts. The torques recorded (in ft lb) were taken from 15 different pieces of equipment as they were assembled.

   a) Make a scatterplot of these paired data. Are there any obvious patterns in the data?

   b) A trick often employed in the analysis of pared data such as these is to reduce the pairs to differences by subtracting the values of one of the variables from the other. Compute differences (top bolt - bottom bolt) here. Then make and interpret a dot diagram for these values.

piece	top_bolt	bottom_bolt
1	110	125
2	115	115
3	105	125
4	115	115
5	115	120
6	120	120
7	110	115
8	125	125
9	105	110
10	130	110
11	95	120
12	110	115
13	110	120
14	95	115
15	105	105

2. [Ch 3, Exercise 3, pg. 114] The accompanying data (also available on the website as `manganese.csv`) are three hypothetical samples of size 10 that are supposed to represent measured manganese contents in specimens of 1045 steel (the units are points, or .01%). Suppose that these measurements were made on standard specimens having "true" manganese contents of 80, using three different analytical methods. (Thirty specimens were involved.)

method_1	87	74	78	81	78	77	84	80	85	78
method_2	86	85	82	87	85	84	84	82	82	85
method_3	84	83	78	79	85	82	82	81	82	79

   a) Make (on the same coordinate system) side by side boxplots that you can use to compare the three analytical methods.
   b) Discuss the apparent effectiveness of the three methods in terms of the appearance of your diagram from a) and in terms of the concepts of accuracy and precision discussed in Section 1.3 of the notes.
   c) An alternative method of comparing two such analytical methods is to use both methods of

analysis once on each of (say) 10 different specimens (10 specimens and 20 measurements). The the terminology of Section 1.2, what kind of data would be generated by such a plan? If one simply wishes to compare the average measurements produced by two analytical methods, which data collection plan (20 specimens and 20 measurements, or 10 specimens and 20 measurements) seems to you most likely to provide the better comparison? Explain.

3. [Ch 3, Exercise 8, pg. 116] The accompanying data are the times to failure (in millions of cycles) of high-speed turbine engine bearings made out of two different compounds. These were taking from "Analysis of Single Classification Experiments Based on Censored Samples from the Two-parameter Weibull Distribution" by J. I. McCool (*The Journal of Statistical Planning and Inference*, 1979).

compound_1	3.03	5.53	5.60	9.30	9.92	12.51	12.95	15.21	16.04	16.84
compound_2	3.19	4.26	4.47	4.53	4.67	4.69	5.78	6.79	9.37	12.75

a) Find the .84 quantile of the Compound 1 failure times.
b) Give the coordinates of the two lower-left points that would appear on a normal plot of the Compound 1 data.
c) Make back-to-back stem-and-leaf plots for comparing the life length properties of bearings made from Compounds 1 and 2.
d) Make (to scale) side-by-side boxplots for comparing the life length for the two compounds. Mark numbers on the plots indicating the locations of their main features.
e) Compute the sample means and standard deviations of the two sets of lifetimes.
f) Describe what your answers to parts c), d), and e) above indicate about the life lengths of these turbine bearings.

4. [Ch 3, Exercise 17, pg. 119] The data in the accompanying table are measurements of the latent heat of fusion of ice taken from *Experimental Statistics* (NBS Handbook 91) by M. G. Natrella. The measurements were made (on specimens cooled to $-.072$°C) using two different methods. The first was an electrical method, and the second was a method of mixtures. The units are calories per gram of mass.

electrical	79.98	80.04	80.02	80.04	80.03	80.03	80.04	79.97	80.05	80.03	80.02	80.00	80.02
mixtures	80.02	79.94	79.98	79.97	79.97	80.03	79.95	79.97					

a) Make side-by-side boxplots for comparing the two measurement methods. Does there appear to be any important difference in the precision of the two methods? Is it fair to say that at least one of the methods must be somewhat inaccurate? Explain.

b) Compute and compare the sample means and the sample standard deviations for the two methods. How are the comparisons of these numerical quantities already evident on your plot in a)?

5. [Ch. 4.1 Exercise 3, pg. 140] The article "Polyglycol Modified Poly (Ethylene EtherCarbonate) Polyols by Molecular Weight Advancement" by R. Harris (*Journal of Applied Polymer 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 temperature 165°C and above are as follows (see website for `polyols.csv`):

Pot temperature (°C)	Average molecular weight
165	808
176	940
188	1183
205	1545
220	2012
235	2362
250	2742

Pot temperature (°C)	Average molecular weight
260	2935

Use a statistical package (JMP or `R`) to help you complete the following (plots and computation):

a) What fraction of the observed raw variation in molecular weight of resulting poly polyols (y) is accounted for by a linear equation in reaction temperature (x)?
b) Fit a linear relationship $y \approx \beta_0 + \beta_1 x$ to these data via least squares. About what change in average molecular weight seems to accompany a 1°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.
d) These data came from an experiment where the investigator managed the value of x. There is a fairly glaring weakness in the experimenter's data collection efforts. What is is?
e) Based on your analysis of these data, what average molecular weight would you predict for an additional reaction run at 188°C? At 200°C? Why would or wouldn't you be willing to make a similar prediction of average molecular weight if the reaction is run at 70°C?

6. [Ch. 4.1 Exercise 4, pg. 140] Upon changing measurement scales, nonlinear relationships between two variables can sometimes be made linear. The article "The Effect of Experimental Error on 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 (and on the website as `tool_life.csv`) are part of that data set.

Cutting speed (sfpm)	Tool life (min)
400	21.5, 24.5, 26, 33
500	6.4, 7.8, 9.8, 16.5
600	2.35, 2.65, 3, 3.6
700	1, 1.2, 1.5, 1.6
800	1, 0.9, 0.74, 0.66

a) Plot tool life (y) by cutting speed (x) and calculate R^2 fir 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 the relationship.) As an aside, Taylor's equation for tool life is $yx^{\alpha} = C$.