

Assignment #1

This assignment is due in class next Thursday, January 28, 1999. I expect that you will talk to your classmates about the assignment, but I remind each of you that it is important to do the work yourself.

(1) This question asks to you repeat some of the things done in Lecture 4 when combining intervals. Using the data on sales at convenience stores and the interval sampling JMP file, repeat the analysis two times with 100 observations each time (assuming no correlation and no bias). Make sure that you obtain a new random sample defined by the interval ranges. Your submitted solution should contain enough output to show that you got new samples. (You will have to compute each column using the “Evaluate” button in the formula dialog. Anyone doing this in Excel?)

(a) Compute the pooled interval using subset selection. That is, use only those observations for which the difference $\text{Source}_1 - \text{Source}_2$ is small. Record the interval that you obtain.

(b) Compute the pooled interval using the same data as in “a”, but now via regression. Record the RMSE of the regression as well as the prediction interval.

(c) Repeat “a” with a new sample.

(d) Repeat “b” using this second sample as well.

(e) Compare the results of the two procedures to the “exact” interval [\$3480, \$3658] obtained using the merged samples (pooling the data to form one sample which has 200 observations).

(f) Which of the two simulation methods appears preferable? Why?

(2) As the time to make your decision approaches, a third source of information comes on the scene. This source of information at the chain of convenience stores has access to other information and said

“A 66.67% interval for the average sales of the outlet is [\$3575, \$3585].”

Combine this new information with the pooled interval [\$3480, \$3658] based on the 200 pooled observations.