Solutions for Assignment #1

(1) This example uses the intervals [$3392, $3644] and [$3491, $3749] for the convenience store sales. I used 100 observations for each simulated column.

Here is a table with my results:

(a) Subset selection, #1	3578	SD	29	n=11
(c) Subset selection, #2	3535		51	n=12
(b) Regression, #1	3570		42	
(d) Regression, #2	3558		45	

If you compare the results of the two procedures to the "exact" interval [$3480, $3658] or 3569 with SD≈44.5, you'll see that regression consistently was near the target mean and SD with both samples. The results from the subset selection (which here only use 10% of the data) are quite variable from sample to sample and do not give reliable estimates of the pooled results.

(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.

The intent of this question was simply to remind you that not all intervals are 95% intervals, as well as to convey the effect of combining a precise interval with a much wider interval. When "decoding" the SD from the interval for the third source, you need to simulate it using the formula 3580 + 5*?normal. The length of this interval is only 2 SDs since it's a 66.67% coverage interval, not a 95% interval. Combining this interval which is very, very narrow with the pooled interval gives the following result: a center of 3580 with an SD of 5. These are basically the same as the narrow interval alone.

Pooled = 3580.3 – 1.01322 S3-Poole

Summary of Fit

RSquare	0.98
RSquare Adj	0.98
Root Mean Square Error	5.06