Homework 6 - STAT 231

Due in class, Thursday, Oct 17

Some of the following problems are from the Devore textbook.

1. Chapter 5: Problem 46
   For part (c), random variables with smaller spreads will be closer to their means.

2. Chapter 5: Problem 54

3. Chapter 7: Problem 3

4. Chapter 7: Problem 4

5. Chapter 7: Problem 12

6. Chapter 7: Problem 14

7. For this problem, you should turn in plots along with your responses to the questions. You do not need to turn in the spreadsheet of simulated values.

   (a) Using JMP, create 3 columns, labeled X1, X2, X3. Add 1,000 rows for each column and simulate 1000 values from a Uniform(0,1) distribution in each column. (Click the red upside-down triangle on the left to add 1000 rows to column X1. Then, in the column title space, right click to choose *formula* and then *Random*; select *Random Uniform* and enter 0 & 1 and click ok). In column 4, calculate the sample means of the X1,X2,X3 values for each row (by right clicking, choosing *formula* and typing in (X1+X2+X3)/3). Plot the distribution of these resulting sample means, based on $n = 3$.

      i. Describe the shape of the distribution. What is the mean and standard deviation? (Click "Analyze" in the menu and select "Distribution" to get a histogram and these numerical summaries.)

      ii. What proportion of the 1000 sample means were between 0.4 and 0.6? (Adding the CDF plot and using the crosshairs tool will be helpful. You can make a CDF by clicking the red upside-down triangle in the output window for the sample mean (i.e., column 4) from part(i) and selecting 'CDF'. The crosshairs are given by an icon at the top of the output window.)

      iii. Recall that a Uniform(0,1) random variable has mean 0.5 and variance 1/12. Calculate the mean and standard deviation for the sample average/mean based on a sample of size 3. How do your answers compare with the mean and standard deviation you observed in your simulation?

   (b) Repeat (a) using a sample of size 10 (i.e., 10 columns) *and* explain why and how your answer for part (ii) changes (the proportions)?

8. Create 5 columns in JMP, labeled X1, X2, X3, X4, X5. Add 1,000 rows for each column and, for each column, simulate 1,000 observations from a normal distribution with mean 30 and standard deviation 4.

   Save your spreadsheet. We may return to it next week.

   (a) Create new columns containing the following information.

   - A new column X6 containing the sample mean of X1, X2, X3, X4, X5.

- A column X7 containing the lower endpoint of a 90% (2-sided) confidence interval for μ.
- A column X8 containing the upper endpoint of a 90% (2-sided) confidence interval for μ.

(b) Create a new column containing a 1 if the interval actually does contain the true value of $\mu = 30$ and 0 if it does not. You can do this by right clicking, choosing *formula*, and using the

```
a<b<=c
```

function from the *Comparison* menu.

(c) How many of your 1,000 confidence intervals contained the true value of $\mu = 30$? Is this number close to what you would expect?

9. This problem using JMP refers to the Cars2015 dataset available on Canvas/Homework Materials page. For this problem, assume that the 110 cars in the dataset are a random sample of all cars produced in 2015. (Whether this is really a valid assumption is unclear.)

   You should turn in your answers along with any images from JMP that help explain you got them. The confidence intervals desired below can be produced under "Analyze→Distributions" and by clicking the red upside-down triangle in the resulting output window.

   (a) Find a 95% confidence interval for the mean highway gas mileage (HwyMPG) for all cars produced in 2015.

   (b) Find a 95% confidence interval for the proportion of 2015 cars that were midsized. (Refering to the Size variable in the dataset.)