Homework 7 - STAT 231

Due in class, Thursday, Oct 24

Some of the following problems are from the Devore textbook.

1. Chapter 7: Problem 19
2. Chapter 7: Problem 22
   The generic interpretation of a confidence interval is "We can be XX% confidence that the true (*insert parameter name here for the particular problem*) lies in (*put confidence interval here*)." So, in (b), "We can be 95% confident that true proportion of hips that develop squeaking lies in...."
3. Chapter 7: Problem 29
4. Chapter 7: Problem 33(c) only
5. Chapter 7: Problem 35
6. Chapter 7: Problem 38
   For (a), you can use JMP to make a histogram and normal QQ plot (under "Analyze→Distribution").
7. Chapter 7: Problem 42
8. Chapter 7: Problem 44
9. This problems involve a JMP dataset "CIsim.JMP" found in Canvas (Homework folder). "CIsim.JMP" was found by completing Problem 8 of Homework 6. (Some of the JMP functions used there will be helpful for this problem). The file contains the following:
   - columns $X1$-$X5$ each contain 1,000 simulated values from normal distributions with $\mu = 30$ and $\sigma = 4$. We will treat each row as a sample of size 5. Thus, we have 1,000 samples.
   - Column 6, titled "Mean" contains the sample means of the 5 values $X1$-$X5$.
   - Columns 7 and 8, titled "Lower" and "Upper" contains the lower and upper end-points for a 90% 2-sided confidence interval calculated using the formula $\bar{X} \pm z^* \frac{\sigma}{\sqrt{n}}$. This interval assumes that we know that $\sigma = 4$. (As we did on Problem 8 of Homework 6.)
   - The column "Contains_mu" contains a 1 if the true value of $\mu = 30$ is in the confidence interval and 0 if it is not.

   Note that double-clicking on the column headings allows you to see the formulas used to create these columns, which might be helpful when you create additional formulas.

   (a) Create a new column called "S" containing the standard deviation of the sample $X1$-$X5$. Use the formula `Std Dev(X1, X2, X3, X4, X5)`. `Std Dev` can be found using the Statistical Functions menu under formula.

   (b) Imagine the true value of σ was not known. For each sample, we could use the sample standard deviation S to estimate σ. Create columns titled "Lower2" and "Upper2" which contain the lower and upper endpoints of a 90% 2-sided confidence interval formed using the formula $\bar{X} \pm z^* \frac{S}{\sqrt{n}}$. Then add a column titled "Contains_mu2" with 0's and 1's indicating whether the true value of $\mu = 30$ was in the interval. (Note this is not the correct kind of confidence interval to use in this situation.)

(c) From (b), how does the proportion of intervals containing the true value of $\mu = 30$ compare the the 90% coverage rate we would expect? What is wrong with using the confidence interval formula from (b) here?

(d) Create three more columns "Lower3", "Upper3", and "Contains_mu3" that contain the lower and upper endpoints, and a 0 or 1 indicating whether the interval contains μ for a confidence interval using the formula $\bar{X} \pm t^* \frac{S}{\sqrt{n}}$, where t^* is the appropriate t-value for a sample of size 5.

(e) How does the width of the t-interval compare to the width of the z-interval? How do the proportions of intervals containing the true value of μ compare?

(f) Create columns "PredLower" and "PredUpper" with the lower and upper endpoints of a 90% prediction interval for a single new observation.

(g) How does the width of the prediction interval compare to the width of the confidence interval?

(h) Create a new column called "Xnew" that contains 1,000 values simulated from a normal distribution with mean 30 and standard deviation 4 (just as $X1$-$X5$ were).

(i) Create a column called "Containing_pred" indicating whether the new value, "Xnew" is contained in your 90% prediction interval.

(j) What proportion of the prediction intervals contained the new value? Is this what you would expect?