Mathematical Statistics II

Statistical Computing Activity: Module 8

One purpose of these Statistical Computing Activities (SCAs) is to give you a chance to explore statistics using the computer. Another purpose is to give you more skills in thinking about the randomness that is life. Usually, like here, these SCAs will have a theme and several problems dealing with that theme or purpose. The reason for that extra layer of complexity is to tie what we do in the class with what we can use these techniques for in our lives as statisticians and/or consultants and/or full members of a democratic society.

The purpose of this activity is to give you some practice in collecting and analyzing data from a Bayesian perspective. It will also allow you to compare the analysis results from three Bayesian analyses and a frequentist analysis.

⋆⋆⋆

The Procedure

The original Skittles candies come in five "flavors": green, orange, purple, red, and yellow. Your job is to draw conclusions about the proportion of Skittles that are green-flavored. To do this, you will count the number of green Skittles and total number of Skittles in your bag. Then, you will perform the usual frequentist analysis to create a confidence interval. Then, you will select three prior distributions for the proportion of green Skittles, determine the likelihood, calculate the posterior distribution, and formulate the credible intervals.

The Data. Here is a nice place to keep track of your data:

Green Skittles: _____

Total Skittles: _____

The Frequentist Analysis. Given these results (this data), calculate a 95% confidence interval for π, the population proportion of green-flavored Skittles.

Confidence Interval: from _____ to _____

The Bayesian Analysis.

The Prior Distribution. The first step in Bayesian analysis is to formulate the prior distribution for the parameter of interest, π. For this SCA, I want you to create three legitimate prior distributions.

Remember that one can calculate the two parameters of a Beta distribution using the following formulas:

$$\alpha = \frac{\mu^2(1-\mu)}{\sigma^2} - \mu$$
$$\beta = \frac{\mu(1-\mu)^2}{\sigma^2} - (1-\mu)$$

Here, μ is the expected value you determine, and σ is the standard deviation. So, if you think that π is 0.75 ± 0.20, you would use $\mu = 0.75$ and $\sigma = 0.20$.

Prior Distribution 1: _____

Prior Distribution 2: _____

Prior Distribution 3: _____

The Likelihood. The likelihood is a reflection of the probability distribution of the data-generating process. Clearly, we are dealing with Binomial data. Thus, the likelihood function is

Likelihood function: _____

The Posterior Distribution. Now that you have the prior distribution and the likelihood, calculate their product and determine the posterior distributions (one for each prior distribution):

Posterior Distribution 1: _____

Posterior Distribution 2: _____

Posterior Distribution 3: _____

The Credible Interval. Those are the posterior distributions. The last thing to do is to calculate the endpoints of the 95% credible intervals:

Credible Interval 1: from _____ to _____

Credible Interval 2: from _____ to _____

Credible Interval 3: from _____ to _____

⋆⋆⋆

The Questions

Please fill in the information on this activity. Then, include it at the end of your submissions. In addition to this, please answer the following questions clearly and concisely. Make sure you include the script you used to answer these questions in your analysis.

(1) Test the null hypothesis $H_o : \pi \geq 0.20$. How different are the four p-values? Given the p-values, do you think we should accept the null hypothesis or not?

(2) Test the null hypothesis $H_o : \pi = 0.20$ using all four methods. How do the two methods provide two different interpretations of the p-value? Which actually gives the probability that the null hypothesis is correct? Explain.

(3) Test the null hypothesis $H_o : 0.19 \leq \pi \leq 0.21$ using the three Bayesian results. Why did I not ask you to do this with the frequentist method?

(4) Provide graphics for each of the Bayesian results. Each graphic should have the prior distribution and the posterior distribution shown. The π-limits should be from 0 to 1.