STA261: Problems 10

Alex Stringer

August, 2018

This assignment is not for credit. Complete the questions as preparation for the quiz in tutorial 10 on August 13th. The questions on the quiz will be very similar to the questions on the assignment.

1. The professor who taught me introductory statistics (see youtube channel: jbstatistics) went out and bought $n = 16$ boxes of Raisin Bran and measured their weights, with the goal of testing Kellogs' claim that the mean weight of cereal present in boxes of Raisin Bran is 755g (family size). I don't have these data but here are some synthetic datapoints representing the weights (in grams) of some randomly purchased boxes of family size raisin bran:

$$782.1, 765, 708.9, 768.6, 751.5, 732, 763.3, 766.5, 768.7, 749.2, 750.5, 748.4, 781.8, 779, 775.2, 746.9$$

Answer the following questions; my answers are included.

(a) Find a 95% confidence interval for the mean weight of boxes of raisin bran sampled in this way (My answer: 747.77, 769.43)

(b) Test the hypothesis that the mean weight of boxes of raisin bran sampled in this way is actually 755g, at the 0.05 significance level (My answer: $p = 0.46$)

(c) How many boxes of raisin bran would you need to buy such that observing a sample with this same mean and standard deviation would cause you to reject the null hypothesis? You won't get a closed form answer for n. Show that this sample size must satisfy

$$\frac{t_{n-1,1-\alpha/2}}{\sqrt{n}} < 0.1841496$$

You can use `R` or Excel to find n numerically by brute force; I got $n = 116$

2. With thanks to Jerry Brunner: A random sample of size $n = 150$ yields a sample mean of $\bar{X} = 8.2$ (unless otherwise stated). Give an estimate and an approximate 95% CI for
   (a) θ, if $X_i \sim Bern(\theta_0)$; use $\bar{X} = 0.82$
   (b) λ, if $X_i \sim Poission(\lambda_0)$
   (c) θ, if $X_i \sim Exponential(\theta_0)$, with the parametrization having $E(X) = \theta$
   (d) θ, if $X_i \sim Exponential(\theta_0)$, with the parametrization having $E(X) = 1/\theta$. There are (at least) two acceptable answers to this one.
   (e) θ, if $X_i \sim Unif(0, \theta_0)$. Note that the MLE doesn't depend on $\bar{X}$, and the regularity conditions aren't satisfied. Try using the Method of Moments estimator.

3. With thanks to Jerry Brunner: That $Unif(0, \theta_0)$ confidence interval is messy. It also doesn't make sense: we know that θ_0 has to be greater than $max(X_i)$, so we should put the lower bound of our confidence interval at $X_{(n)} = max(X_i)$. You have seen part of the below questions before.
   (a) Show that
   $$X/\theta_0 \sim Unif(0, 1)$$
   (b) Show that
   $$X_{(n)}/\theta_0 \sim Beta(n, 1)$$
   (c) Find a $1 - \alpha$ confidence interval for θ of the form
   $$(X_{(n)}, qX_{(n)})$$
   I.e. find the constant q such that $P(X_{(n)} < \theta_0 < qX_{(n)}) = 1 - \alpha$

(d) For the following random sample from a $Unif(0, \theta_0)$ distribution, evaluate both your confidence interval from this, and the previous question. Which would you prefer to use for making inferences about θ, and why?

$$4.12, 6.42, 10.51, 4.1, 11.57, 13.35, 5.26, 8.5, 0.98, 4.04$$

4. With thanks to Jerry Brunner: In question 1, you derived a $1-\alpha$ CI for θ from a $Bern(\theta_0)$ distribution, by using the CLT for the MLE and plugging in the estimate $j(\hat{\theta})$ for $i(\theta_0)$ in the Fisher Information. Can we improve this confidence interval, by not replacing θ_0 with an estimate?

(a) Use the CLT to show that

$$P\left(\left(\frac{\sqrt{n}(\bar{X}-\theta_0)}{\sqrt{\theta_0(1-\theta_0)}}\right)^2 < z^2_{\alpha/2}\right) \approx 1-\alpha$$

(b) Keep going; show that

$$P\left((n+z^2_{\alpha/2})\theta_0^2 - (2n\bar{X}+z^2_{\alpha/2})\theta_0 + n\bar{X}^2 < 0\right) \approx 1-\alpha$$

(c) Because $(n+z^2_{\alpha/2}) > 0$, these random parabolae all open upward. That means that the endpoints of the confidence interval we want are the points at which this parabola (as a function of θ) intersect the x-axis, because between these points, the parabola is less than zero, which is the event that happens with probability $1-\alpha$. Use the quadratic formula and simplify to show that the confidence interval is

$$\frac{2n\bar{X}+z^2_{\alpha/2}}{2(n+z^2_{\alpha/2})} \pm \frac{\sqrt{4nz^2_{\alpha/2}\bar{X}(1-\bar{X})+z^4_{\alpha/2}}}{2(n+z^2_{\alpha/2})}$$

(d) Compare the interval calculated in this way to the interval calculated in question 1. Do you notice much difference for $n=150$? Try it for smaller values of n, say $n=5,10,15,30$. How does replacing $i(\theta_0)$ by $j(\hat{\theta})$ affect the width of the interval?