PSTAT 100: Spring 2024 Instructor: Ethan P. Marzban

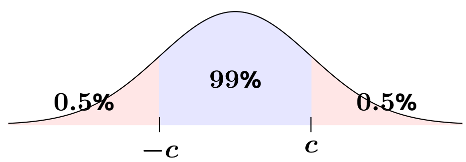
Solutions to Selected Exercises from ICA02

1. (**Q10 on VA; Q4 on VB**) Which of the following correctly gives a 99% confidence interval for the intercept of the linear model that was fit?

Solution: Most everybody correctly identified that we need to use the t distribution with $(n-2) = 98$ degrees of freedom. However, people seemed to struggle a bit with which percentile to use. A 99% confidence interval for β_0, based on $\widehat{\beta}_0$, will be of the form

$$\widehat{\beta}_0 \pm c_{t_{98}} \cdot \mathsf{sd}\left(\widehat{\beta}_0\right)$$

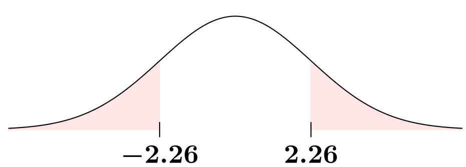
where $c_{t_{98}}$ is chosen to ensure that the interval covers the true value of β_0 with probability 99%. What this means is that we require the tail probabilities to be 0.05% each:



So, we actually want $c_{t_{98}}$ to be the 99.5$^{\text{th}}$ percentile, not the 99$^{\text{th}}$. Please see Lecture 10 for more details; specifically, slide 10 provides a formula for computing the confidence coefficient (though the slide lists it in the case of a normally-distributed statistic, we need only to replace $\Phi^{-1}(\cdot)$ with the quantile function of the t distribution to obtain the analog for this problem.)

2. (**Q21 on Both Versions**) Suppose the observed value of the test statistic is approximately 2.26. Let W be a random variable that follows the sampling distribution of the test statistic under the null. Which of the following correctly gives the p-value of the observed value of the test statistic?

Solution: Again, a picture is worth a thousand words! Recall that a $p-$value is the probability of observing something as or more extreme as what was observed, in the direction of the alternative. Hence, for an observed value of 2.26 and a random variable W following the sampling distribution of the test statistic under the null, our $p-$value can be computed pictorially by



We can see that there are several (equivalent) ways of writing this probability mathematically:

- $\mathbb{P}(W < 2.26) + \mathbb{P}(W > 2.26)$
- $2\mathbb{P}(W > 2.26)$
- $2\mathbb{P}(W < -2.26)$
- $\mathbb{P}(|W| > 2.26)$

Note that the fourth answer choice is Answer Choice (A) on the exam, and is hence the correct answer. (As practice, I encourage you to go through and sketch which regions the remaining answer choices on the exam correspond to.)

Estimation Question:

Consider an i.i.d. (independent and identically distributed) sample of random variables $X_1, \cdots, X_n$ (for $n \geq 2$) drawn from the distribution with density function given by

$$f_X(x) = \frac{2\theta^2}{x^3} \cdot \mathbb{1}_{\{x \geq \theta\}} = \begin{cases} \frac{2x^2}{\theta^3} & \text{if } x \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

Additionally, consider using the sample minimum as an estimator for θ that is, define

$$\widehat{\theta}_n := \min_{1 \leq i \leq n} \{X_i\} = \min\{X_1, \cdots, X_n\}$$

to be an estimator for θ. You may use, without proof, the fact that the sampling distribution of $\widehat{\theta}_n$ is given by

$$f_{\widehat{\theta}_n}(t) = \frac{2n\theta^{2n}}{t^{2n+1}} \cdot \mathbb{1}_{\{t \geq \theta\}}$$

3. (**Q16**) Which of the following correctly gives an expression for $\mathbb{E}[\widehat{\theta}_n]$ the expected value of $\widehat{\theta}_n$? (Yes, this is a PSTAT 120A-type exercise, but we have seen problems like this in Lecture and on Homework!)

Solution: By definition,

$$\begin{aligned} \mathbb{E}\left[\widehat{\theta}_n\right] &:= \int_{-\infty}^{\infty} t f_{\widehat{\theta}_n}(t) \, \mathrm{d}t = \int_{-\infty}^{\infty} t \cdot \frac{2n\theta^{2n}}{t^{2n+1}} \cdot \mathbb{1}_{\{t \geq \theta\}} \, \mathrm{d}t \\ &= 2n\theta^{2n} \cdot \int_{\theta}^{\infty} t^{-2n} \, \mathrm{d}t = 2n\theta^{2n} \cdot \frac{1}{-2n+1} \left[t^{-2n+1}\right]_{t=\theta}^{t=\infty} \\ &= \frac{2n}{2n-1} \cdot \cancel{\theta^{2n}} \cdot \theta^{\cancel{-2n}+1} = \boxed{\left(\frac{2n}{2n-1}\right)\theta} \end{aligned}$$

4. (**Q17**) Suppose the correct answer to question 16 above is (C) [which is NOT to say this is the ACTUAL correct answer for question 16!] Which of the following correctly computes the bias of using $\widehat{\theta}_n$ as an estimator for θ?

Solution: By definition,

$$\begin{aligned} \text{Bias}\left(\widehat{\theta}_n\,,\,\theta\right) &:= \mathbb{E}\left[\widehat{\theta}_n\right] - \theta = \frac{\theta}{2n-1} - \theta \\ &= \left(\frac{1}{2n-1} - 1\right)\theta = \boxed{\left(\frac{2-2n}{2n-1}\right)\theta} \end{aligned}$$

5. (**Q18**) Suppose the correct answer to question 16 above is (B) [which is NOT to say this is the ACTUAL correct answer for question 16!] Additionally, suppose a particular sample of 23 observations had an observed sample minimum value of 6.21. What would be an "ideal" estimate of θ? (Specifically, think about how you can construct an unbiased estimator of θ that is a function of $\widehat{\theta}_n$ based on your answer to question 16.)

Solution: If we assume $\mathbb{E}[\widehat{\theta}_n] = \theta/(2n)$, then $(2n)\widehat{\theta}_n$ will be an unbiased estimator for θ. Thus, since the observed value of the sample maximum is 6.21, an observed instance of our unbiased estimator will be $(2 \cdot 23) \cdot 6.21 = \boxed{285.66}$.

