2510-001: QUIZ WEEK 10 (ZINTERVAL, TINTERVAL, 1-PROPZINT, OR 2-PROPZINT?)

COLTON GRAINGER
ALL SAINTS' DAY 2019

Your name (print clearly in capital letters): Solutions (2019-11-04)

Read Me Carefully

- This is a 72 hour take-home quiz.
- Please **answer each problem**, and return this quiz to me on **Monday, November 4th at 8:00am**.
- This quiz is *open note and open internet*. The *Week 10 study guide* should help to answer these questions.
- I encourage y'all to collaborate!
- Who (if anyone) did you collaborate with? Fall 2018 Exam 2
- Note: each person does need to turn in their own copy of this quiz for credit.

Graded Questions

(1) (2 points) A marketing research company is estimating the average yearly compensation of famous European rappers. In 2016, data were randomly collected from 18 rappers and the 90% confidence interval for the mean was calculated to be (218126, 583618) (in Euros). If the same random sample were to be used to compute a %95 confidence interval, how would this new %95 confidence interval compare to the original %90 interval?

(a) The interval would get wider.
(b) The interval would get narrower.
(c) There would be no change in the width of the interval.
(d) It is impossible to tell until the 95% interval is constructed.

Your answer: A

At 17 degrees of freedom, $t_{0.9} = 1.7396$ and $t_{0.95} = 2.1098$ are the critical values. Since the conf. int. is of the form $\bar{x} \pm (t_c)\left(\frac{s_{\bar{x}}}{\sqrt{n}}\right)$, and $t_{0.9} < t_{0.95}$, we have (A).

(2) (2 points) Find numerically the %95 confidence interval for the 2016 famous European rapper yearly compensation data given in the previous problem.

Your answer: $\bar{x} \pm t_{0.975}(s_x/\sqrt{n}) = (179235.5, 622508.5)$

③ $\bar{x} = \frac{583618 + 218126}{2}$ ③ $E = 583618 - \bar{x}$ ⑤ $s_x = E\left(\frac{\sqrt{n}}{t_{0.9}}\right)$

② $n = 18$ ④ $E = (t_{0.9})\left(\frac{s_x}{\sqrt{n}}\right)$ ⑥ $t_{0.975} = \text{invT}(0.975)$

(3) (2 points) If $10^4 = 1000$ simple random samples each of size $n = 50$ are drawn from the population of American voters, and 1000 corresponding %90 confidence intervals

$\hat{p}_i \pm z_{0.9} \times$ standard error from the ith sample (one each for the samples $i = 1$ to $i = 1000$)

are created from this data to estimate a *population proportion* p, how many of these intervals are expected to contain p?

Your answer: $900 = 1000 \cdot 0.90 = E(X)$

$E(X)$ = the expected # of successful conf. intervals containing p

$X \sim \text{Bin}(0.9, 1000)$

(4) An educator wants to estimate the proportion of school children in Toronto who are living with only one parent. The report is to be published, thus they want a reasonably accurate estimate. However, their funding is limited so they do not want to collect a larger sample than necessary. They hope to use a sample size such that, with a confidence level of 0.99, the margin of error will not exceed 0.038.

(a) (1/2 point) What relevant mathematical information is as *known* in this problem?[1]

Your answer: $c = 0.99$, $E \leq 0.038$, $z_c \approx 2.5758$

(b) (1/2 point) What relevant mathematical parameters are *unknown* in this problem?[2]

Your answer: $\hat{p}$ the sample prop, p the popn prop, n the samp. size

(c) (1/2 point) What mathematical relation,[3] if any, exists between the unknown parameters and the known information?

Your answer: Find the least positive integer n such that $z_c \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq z_c \cdot \sqrt{\frac{(1/2)(1/2)}{n}} \leq 0.038$.

Note $E = z_c \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, so $n = \left(\frac{z_c}{E}\right)^2 (\hat{p}(1-\hat{p}))$. Also $\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) \geq (q)(1-q)$ for all $0 \leq q \leq 1$.

(d) (1/2 point) Exploit the relationship between the known information to determine the unknown parameters, then provide a numerical answer to the question: What sample size will ensure that this researcher's margin of error is bounded by 0.038, regardless of what sample proportion value occurs once they gather the sample?

Your answer: $n \geq \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(z_{0.99} / 0.038\right)^2 = 1148.701$

$n = 1149$ (note that some students may use lower precision for z_c and have approx answers, eg if $z_{0.99} = 2.58$, then $n = 1153$)

(5) A data scientist in charge of a charity organization that distributes free kits that help smokers quit. There are two towns that still need a shipment, Onett and Twoson. She knows that both towns have approximately 30,000 people in them and wants to determine how many kits to send to each town.

Here's the header[4] for the data:

(`town_name`, `sample_size`, `number_of_smokers_in_sample`)

Here is the data corresponding to the above header.

(`Onett`, 1780, 215)

(`Twoson`, 1751, 150)

(2 points) Before the researcher can report her results, her supervisor informs her that it is believed that there is a significant difference and asks her to estimate the difference of population proportions. Create a 95% confidence interval for the difference of population proportions based on the previously mentioned polling results.

Your answer: $(0.0151, 0.0551)$

2-PropZInt

[1] Hint: What is known should help you determine whether it is more appropriate to use the standard normal distribution or one of Student's t-distributions to determine the optimal sample size.

[2] Hint: Draw and label the probability distribution that is most appropriate to use. What should the labels be?

[3] For example, a "mathematical relationship" could be in the form of an equation or an approximation.

[4] In SQL, a header like this declares the names of the columns as *field* names. The rows of a SQL table are called *entries*.