2020 Summer STAT 22000 FINAL EXAM, Part 1

Important: read the following instructions carefully.

1. During the exam, you may refer to the textbook, slides, homework, and the solutions and other materials posted on Canvas. You can use R during the exam. However, you cannot use Google or other search engines during the exam. You must do the exam all by yourself. You cannot get assistance from other people.

2. If a question asks you do some calculations, you must **show your work to receive full credit**.

3. Please check Canvas or email regularly during the exam. Yibi might send out corrections or clarifications about exam problems that you don't want to miss.

4. If you are unsure of what a question is asking for, **you may send questions to Yibi by email**.

5. Whenever appropriate, parts of a question will be graded conditionally on how you answered the preceding part(s). For example, even if you get part (a) of a question wrong, you will still get credit for the rest of the question provided your answers to parts (b), (c), etc. are consistent with how you answered part (a).

1. The game of roulette involves spinning a wheel with 38 slots: 18 red, 18 black, and 2 green. A ball is spun onto the wheel and will eventually land in a slot, where each slot has an equal chance of capturing the ball. One popular bet is that it will stop on a red slot; such a bet has an 18/38 chance of winning.

   For part (a-b) below, suppose the gambler bets on red in 100 different spins.

   (1a) (4pts) How many times do you expect the gambler to win in the 100 spins? And with what standard deviation?

   (1b) (5pts) What is the probability that the gambler wins at least 55 times in the 100 spins? Please calculate using the normal approximation to the Binomial WITH the continuity correction. If you don't know how to do the continuity correction, you can also use normal approximation WITHOUT continuity correction and get 4 points if it is done correctly.

   (1c) (3pts) Which of the following events is more likely to happen, or they are equally likely to happen? Explain briefly.

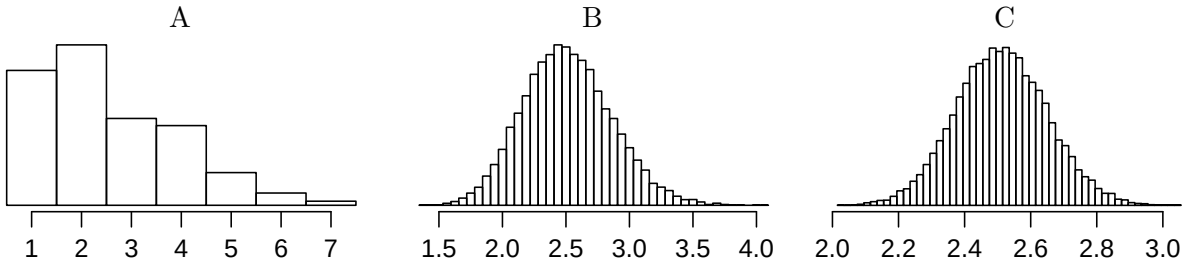
   - Event A: The gambler bets on red 20 times, and he wins at least half of the times.
   - Event B: The gambler bets on red 50 times, and he wins at least half of the times.

2. According to the 2019 *Current Population Reports* published by the U.S. Census Bureau, the distribution of household size in the U.S. has a mean of 2.52, median of 2, and standard deviation of 1.4. About 28.4% of U.S. households are of size 1. About 1.2% of the households are of size over 7.

   (2a) (4pts) Can we calculation (an approximate) probability that a randomly chosen household in the U.S. is of size 2 or more? If yes, please (i) explain why you can and (ii) calculate the probability. If no, explain why not.

   (2b) (4pts) Can we calculation an approximate probability that the mean size of 50 randomly chosen households is over 2.8 using the normal distribution? If yes, please (i) explain why you can and (ii) calculate the probability. If no, explain why not.

   (2c) (3pts) Please match the 3 histograms below (A,B,C) with the 3 descriptions of the histograms (I, II, III) and explain your reasoning.



I sampling distribution of the mean size of 16 randomly selected households

II sampling distribution of the mean size of 100 randomly selected households

III Histogram of the sizes of 500 randomly sampled households

3. To determine whether glaucoma (a disease of the eye) is related to the thickness of the cornea (the transparent tissue that covers the front of the eye), measurements were made of 8 people who had the disease in one eye, but not in the other eye. The corneal thicknesses (in microns) were as follows[1]:

Subject	1	2	3	4	5	6	7	8	Mean	SD
Corneal thickness of healthy eye	490	484	498	450	442	404	470	482	465	31.3141
Corneal thickness of diseased eye	488	478	480	426	440	410	458	460	455	27.6922
Difference = Healthy − Diseased	2	6	18	24	2	−6	12	22	10	10.7438

Carry out a test H_0: $\mu_{\text{healthy}} = \mu_{\text{diseased}}$ v.s. H_a: $\mu_{\text{healthy}} > \mu_{\text{diseased}}$, where μ_{healthy} and μ_{diseased} are the population mean corneal thicknesses of the healthy eye and that of the diseased eye respectively.

(3a) (4pts) Report a test statistic (with degrees of freedom if any). Please show your calculation.

(3b) (2pts) Report the P-value and make a conclusion in the context of the problem (not simply whether H_0 is rejected). Use a 5% significance level. Please include the R code you used in your work.

4. Ionizing radiation is being given increasing attention as a method for preserving horticultural products. The article "The Influence of Gamma-Irradiation on the Storage Life of Red Variety Garlic" (*J. Food Process. Preserv.*, 1983: 179-183) reports that 153 of 180 irradiated garlic bulbs were marketable (no external sprouting, rotting, or softening) 240 days after treatment, whereas only 119 of 180 untreated bulbs were marketable after this length of time.

We want to test whether ionizing radiation is **beneficial** as far as marketability is concerned.

(4a) (2pts) Please state the null and the alternative hypotheses H_0 and H_a with appropriate symbol(s). Please define the symbol(s) properly.

(4b) (4pts) Please report test statistic. Show your calculation.

(4c) (2pts) Please find P-value and make a conclusion in the context of the problem (not simply whether H_0 is rejected) using a 1% significance level. Please show the R code you used to find the P-value.

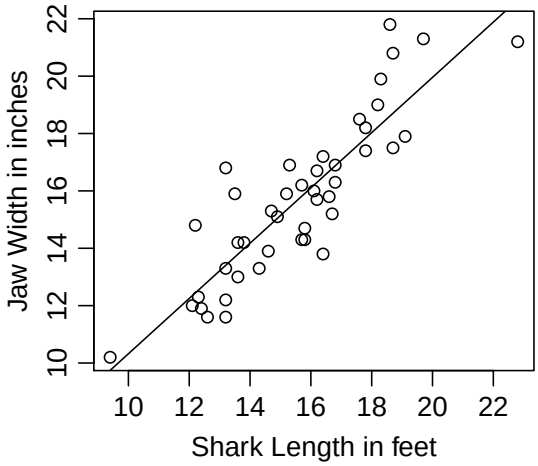
(4d) (2pts) Please check all the necessary condition(s) required to perform the test above.

(4e) (4pts) Calculate a 99% confidence interval for the change in marketable proportions after ionizing radiation. Show your calculation.

(4f) (4pts) Calculate a 99% confidence interval for the population proportion of garlic bulbs that were marketable 240 days after the ionizing radiation treatment. Show your calculation.

[1] N. Ehlers, "On corneal thickness and introcular pressure," *Acta Opthalmologica*, **48**, pp.1107-1112. Reprinted in *Small Data Sets, p.127.*

5. Physical characteristics of sharks are of interest to surfers and scuba divers as well as to marine researchers. Because it is difficult to measure jaw width in living sharks, researchers would like to determine whether it is possible to estimate jaw width from body length, which is more easily measured. The following regression output is for predicting jaw widths in inches from length of sharks in feet for 44 sharks. Suppose the 44 sharks were randomly sampled from some population.




```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   XXXXXXX    1.29905   XXXXX    0.599
shark.length  XXXXXXX    0.08228   XXXXX 8.22e-15 ***
```

The following is a summary of the data:

	Mean	SD	
Shark Length (ft.)	15.59	2.55	Correlation $r \approx 0.875$
Jaw Width (in.)	15.70	2.81	

(5a) (4pts) Write down the equation of the least square regression line for predicting the jaw width in inches of a shark from its body length in feet. Specifically, what are the values of the intercept and the slope?

(5b) (2pts) Interpret the slope of regression line you found in part (a).

(5c) (2pts) Which of the following statements about correlation $r = 0.875$ is TRUE? Circle ONE answer. No explanation is required.

(i) The sample variance of the predicted jaw widths of sharks is $r = 0.875 = 87.5\%$ of the sample variance of the observed jaw widths.

(ii) The sample SD of the predicted jaw widths of sharks is $r^2 = (0.875)^2 = 76.56\%$ of the sample SD of the observed jaw widths.

(iii) The sample SD of the residuals is $1 - r = 1 - 0.875 \approx 12.5\%$ of the sample SD of jaw widths.

(iv) The sample variance of the residuals is $1 - r^2 = 1 - (0.875)^2 \approx 23.44\%$ of the sample variance of jaw widths.

(5d) (4pts) Calculate a 95% confidence interval for the slope of the regression line for predicting the jaw width of a shark from its body length. No interpretation is required.

(5e) (2pts) If we were to predict the body length (in feet) of a shark from its jaw width (in inches) instead, the correlation coefficient would be ... (Choose one of the following. No explanation is required.)

(i) 0.875 (ii) -0.875 (iii) $1 - 0.875$ (iv) $1/0.875$
(v) More information is needed to determine.

(5f) (3pts) Predict the body length of a shark (in feet) with a jaw width of 20 inches, using the least square method

6. (3pts) Suppose the Career Advising Center in a certain university conducted a survey and found 97 of 100 randomly sampled recent graduates got a job before or within one year after graduation. True or False and explain: the 95% confidence interval for the proportion of the recent graduates who got a job before or within one year from graduation can be calculated as $0.97 \pm 1.96\sqrt{(0.97 \times 0.03)/100}$.

7. (3pts) A geography test was given to a simple random sample of 250 high-school students in a certain large school district. One question involved an outline map of Europe, with the counties identified only by number. The students were asked to pick out Great Britain and France. As it turned out, 176 students could find Great Britain, compared to 164 for France. True or False and explain: the 95% confidence interval for the difference of the proportion of high-school students in this school district that can pick out Great Britain from the map and the proportion that can pick out France from the map can be calculated as

$$\widehat{p}_B - \widehat{p}_F \pm 1.96\sqrt{\frac{\widehat{p}_B(1-\widehat{p}_B)}{250} + \frac{\widehat{p}_F(1-\widehat{p}_F)}{250}}, \quad \text{where } \widehat{p}_B = \frac{176}{250}, \text{ and } \widehat{p}_F = \frac{164}{250}.$$

Explain briefly.