Final Exam

PSTAT 5A, Summer B, 2018

Name:

Perm #:

Section Time & TA Name (draw a circle):	A. Bernstein TW 2 pm	A. Bernstein TW 3 pm	K. Wang TW 5 pm

Instructions:

- You have 80 minutes to complete the exam.
- Read each question carefully and answer all questions.
- **Round numbers to 3 decimal places.**
- You must show your work clearly: NO WORK=NO CREDIT.
- Anyone found copying another students' work will be given an F for the course.
- You are **NOT ALLOWED** to consult any notes or textbook during this exam.
- You are **NOT ALLOWED** to consult any cellphones, smartphones, computers or electronic device of any form during this exam.
- All cellphones, smartphones and computers must be turned off.
- You may use a calculator. You cannot not use a phone as a calculator.

Questions	Points
1 (40 pts)	
2 (30 pts)	
3 (30 pts)	
TOTAL	

Good Luck!!!

Question 1

You are hired to study the rental market for UCSB students. The Housing Office did a survey of 43 apartments for rent in the area. They found that the average rent was \$1,873 per month with a sample standard deviation of \$800. To see how the size of the apartment affects the rent, they also recorded the overall square footage for each apartment which found an average of 750 square feet with a sample standard deviation of 343.4 square feet. The sample correlation between these measurements was 0.685.

a) (10 points) Identify the explanatory and response variables in this study and estimate a simple linear regression. Give an interpretation to the slope.

(2 points) Identify variables:
X= Size of the apartment (square feet)
Y= Monthly rent

(5 points) Estimate regression

$$\widehat{\beta_1} = r\frac{s_y}{s_x} = 0.685\times\frac{800}{343.4} = 1.5958 \approx 1.596$$

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x} = 1873 - 1.596\times(750) = 676$$

(3 points) Interpretation of the slope: $\widehat{\beta_1}$ means that every additional square feet of apartment size increases the rent in \$1.596 dollars per month

b) (5 points) What would be the rent for an 800 square feet apartment?

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x = 676 + 1.596\times 800 = 1952.8$$

c) (10 points) Find the coefficient of determination and give its interpretation in this study.

(5 points) The sample correlation is $r = 0.685$
Then the coefficient of determination is $r^2 = 0.685^2 = 0.469$

(5 points) Interpretation: The size of the apartment explains 46.9% of the variation in the monthly rent

d) (10 points) Find a 98% confidence interval for the population mean of explanatory variable. Assume it has a normal distribution.

(5 points)

$\bar{x} = 750$, $s_x = 343.4$, $n = 43$

For a 98% confidence level we have df=n-1=42 so we use $t = 2.418$

(5 points)

Hence, the 98% CI is:

$$\bar{x} \pm t\frac{s}{\sqrt{n}} = 750 \pm 2.418 \times \frac{343.4}{\sqrt{43}} = (623.37, 876.63)$$

e) (10 points) Suppose you dig more into the details of this same survey. You find out that 35% of the surveyed apartments are infested with cockroaches. But your boss believes that at most 25% of the apartments have this issue. Would you agree with him at a 10% significance level? Check any necessary conditions.

(8 points) We know: $\hat{p} = 0.35$, n=43

1) Hypothesis: $H_0: p \leq 0.25$ vs $H_A: p > 0.25$, that is $p_0 = 0.25$
2) test statistic: $z_{test} = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.35-0.25}{\sqrt{\frac{0.25(1-0.25)}{43}}} = 1.5143 \approx 1.51$
3) p-value: $H_A: p > p_0 \Rightarrow$ p-value $= P(Z > z_{test})$
$$= 1 - P(Z \leq z_{test}) = 1 - P(Z \leq 1.51)$$
$$= 1 - 0.9345 = 0.0655$$
4) Decision: p-value $< \alpha$ since $0.0655 < 0.10$. Hence, we reject H_0
5) Conclusion: We reject the boss claim that at most 25% of the apartments are infested with cockroaches.

(2 points) The necessary condition for this hypothesis test to hold:

$$n\hat{p} = 43 \times 0.35 = 15.05 > 10$$
$$n(1-\hat{p}) = 43 \times (1-0.35) = 27.95 > 10$$

Question 2

A scientist believes that the distribution of adult's height in North Dakota is well represented by a normal random variable with mean 70 inches and standard deviation 5 inches. And a study from the US Dept. of Health found a national average height of 69.15 inches.

a) (10 pts) What' the probability that a randomly chosen North Dakotan is taller than an average American given that is shorter than 70.3 inches?

(2 points) From the question we get: $X \sim N(\mu = 70, \sigma = 5)$

(6 points)

We solve for: $P(X > 69.15 | X < 70.3) = \frac{P(X>69.15 \text{ and } X<70.3)}{P(X<70.3)} = \frac{P(X<70.3)-P(X<69.15)}{P(X<70.3)}$

$$P(X < 69.15) = P\left(Z < \frac{69.15 - 70}{5}\right) = P(Z < -0.17) = 0.4325 \approx 0.433$$

$$P(X < 70.3) = P\left(Z < \frac{70.3 - 70}{5}\right) = P(Z < 0.06) = 0.5239 \approx 0.524$$

(2 points) Hence: $P(X > 69.15 | X < 70.3) = \frac{0.524-0.433}{0.524} = 0.17366 \approx 0.174$

b) (10 points) If I randomly chose five North Dakotans, what is the probability that at most 2 of them are shorter than 70.4 inches?

(3 points) First, the probability of success:

$X \sim N(\mu = 70, \sigma = 5)$

$$P(X < 70.4) = P\left(Z < \frac{70.4 - 70}{5}\right) = P(Z < 0.08) = 0.5319 \approx 0.532$$

(5 points) Then work with a Binomial: Let $Y \sim Bin(n = 5, p = 0.532)$

We calculate: $P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2)$

$$P(Y = 0) = \binom{5}{0} 0.532^0(1 - 0.532)^{5-0} = 0.02245067 \approx 0.022$$

$$P(Y = 1) = \binom{5}{1} 0.532^1(1 - 0.532)^{5-1} = 0.1276042 \approx 0.128$$

$$P(Y = 2) = \binom{5}{2} 0.532^2(1 - 0.532)^{5-2} = 0.2901087 \approx 0.290$$

(2 points) Compute $P(Y \leq 2) \approx 0.022 + 0.128 + 0.29 = 0.439$

c) (10 points) You want to make inference only on those North Dakotans taller than the national average. How many North Dakotans do you need to survey to get a 97% confidence interval with a margin of error of 0.15 for this task?

$$n_{MIN} = \left(\frac{z}{E}\right)^2 p(1-p)$$

(5 points) For a 97% CI we need $z = z_{98.5} = 2.1$ .We use p=0.5 always

(3 points) Hence:

$$n_{MIN} = \left(\frac{z_{98.5}}{E}\right)^2 p(1-p) = \left(\frac{2.17}{0.15}\right)^2 \times 0.5 \times (1-0.5) = 52.32$$

(2 points) We need a minimum of 53 North Dakotans

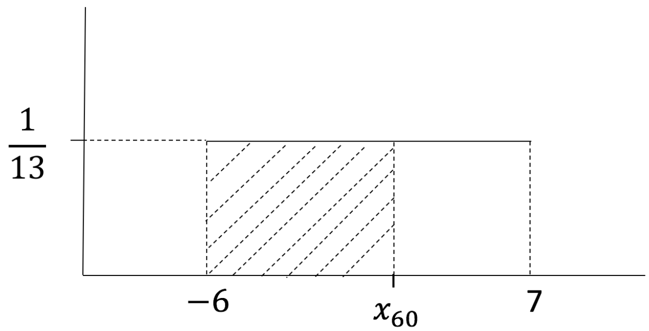
Question 3

a) (10 pts) Let $\boldsymbol{X} \sim Unif(-6,7)$. What is the 60th percentile of $\boldsymbol{X}$?

(4 points)

The 60th percentile is given by: $P(X \leq x_{60}) = 0.6$

(4 points)



$$P(X \leq x_{60}) = \left(x_{60} - (-6)\right) \times \frac{1}{13} = \frac{x_{60} + 6}{13}$$

(2 points)

Solve for x_{60}:

$$\frac{x_{60} + 6}{13} = 0.6$$
$$\Rightarrow x_{60} = 13 \times 0.6 - 6$$
$$= 1.8$$

b) (10 pts) You have two trucks: one full of oranges and another full of blueberries. You take a random sample of 150 oranges and 200 blueberries. You find that 78 and 122 of them respectively were in contact with fruit flies. Test the claim that the proportion of blueberries minus the proportion of oranges in touch with fruit flies is greater than 5%. Use a 10% significance level. Check any necessary conditions.

Population 1: $n_1 = 200$ and $\hat{p}_1 = \frac{122}{200} = 0.61$ for the blueberries.

Population 2: $n_2 = 150$ and $\hat{p}_2 = \frac{78}{150} = 0.52$ for the oranges.

1) Hypothesis: The problem says: *Test the claim that the difference between proportions of blueberries and oranges is greater than 5%.* So:
   $H_0: p_1 - p_2 \leq 0.05$ vs $H_A: p_1 - p_2 > 0.05$, that is $\Delta = 0.05$

2) test statistic: $z_{test} = \frac{(\hat{p}_1 - \hat{p}_2) - \Delta}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$
   where $\bar{p} = \frac{78+122}{150+200} = \frac{200}{350} = 0.5714 \approx 0.571$
   then:
   $$z_{test} = \frac{(0.61 - 0.52) - 0.05}{\sqrt{0.571(1-0.571)\left(\frac{1}{200}+\frac{1}{150}\right)}} \approx \frac{0.04}{0.053} = 0.754$$

3) p-value: $H_A: p_1 - p_2 > \Delta \Rightarrow \text{p-value} = P(Z > z_{test}) = 1 - P(Z \leq z_{test})$
   $$= 1 - P(Z \leq 0.754) \approx 1 - P(Z < 0.75)$$
   $$= 1 - 0.7734 = 0.2266 \approx 0.223$$

4) Decision: $\text{p-value} > \alpha$ since $0.223 > 0.10$. Hence, we can't reject H_0
5) Conclusion: We conclude that there is not enough evidence to reject the hypothesis that the difference between proportions of blueberries and oranges in touch with fruit flies is smaller or equal than 5%.

The conditions for this hypothesis test to hold:

$$n_1\hat{p}_1 = 200 \times 0.61 = 122 > 10$$
$$n_1(1-\hat{p}_1) = 200 \times (1-0.61) = 78 > 10$$
$$n_2\hat{p}_2 = 150 \times 0.52 = 50 > 10$$
$$n_2(1-\hat{p}_2) = 150 \times (1-0.52) = 72 > 10$$

c) Let $\boldsymbol{Z}$ be a standard normal random variable. You sample this variable and you record its value each time. Based on the value of $\boldsymbol{Z}$, you assign a value to $\boldsymbol{W}$ as follows:

$$\boldsymbol{W} = \begin{cases} 5 & \textbf{if } \boldsymbol{Z} > \ -\mathbf{0.6} \\ 10 & \textbf{if } \boldsymbol{Z} < -\mathbf{2} \\ 20 & \text{otherwise} \end{cases}$$

Find the mean of $\boldsymbol{W}$ and compute $P(\boldsymbol{W} > 15|\boldsymbol{W} > 5)$

(5 points)

$$\boldsymbol{P(W = 5) = P(Z > \ -0.6) = 1 - P(Z \leq \ -0.6)}$$
$$\boldsymbol{= 1 - 0.2743 = 0.7257 \approx 0.726}$$
$$\boldsymbol{P(W = 10) = P(Z < \ -2) = 0.0228}$$
$$\boldsymbol{P(W = 20) = P(Z = 1) = 1 - P(Z > \ -0.6) - P(Z < \ -2)}$$
$$\boldsymbol{= 1 - 0.726 - 0.0228 = 0.2512}$$

Then the mean of $\boldsymbol{W}$ **is:**

$$\mu = 5\times\boldsymbol{P(W = 5) + 10\times P(W = 10) + 20\times P(W = 20)}$$
$$= 5\times\boldsymbol{0.726 + 10\times0.0228 + 20\times0.2512}$$
$$= \boldsymbol{8.88}$$

(5 points)

Finally,

$$P(W > 15|W > 6) = \frac{P(W > 15 \text{ and } W > 6)}{P(W > 6)}$$
$$= \frac{P(W = 20)}{P(W = 10) + P(W = 20)} = \frac{0.2512}{0.0228 + 0.2512} = 0.9167 \approx 0.917$$