Quiz 2

PSTAT 120B

Instructions: the quiz is open book and open note and has no strict time limit. You can use any course materials (and are encouraged to do so!). Complete the quiz on your own and do not share your work with anyone else until after the deadline has passed. Write your answers on a printed copy of the quiz or on blank paper and upload a PDF copy to Gradescope by the deadline. Please be sure to match the pages of your PDF to the outline.

1. Let $Y_1, ;Y_n \overset{iid}{\sim} \text{Bernoulli}(p)$. For this problem, you will assess the quality of the CLT approximation to the binomial distribution. You may find the following applets helpful:

   - https://homepage.divms.uiowa.edu/ mbognar/applets/bin.html
   - https://homepage.divms.uiowa.edu/ mbognar/applets/normal.html

   (a) Consider the standardized sample mean:

   $$\frac{\sqrt{n}(\bar{Y} - p)}{\sqrt{p(1-p}}$$

   Write an expression approximating the CDF:

   $$F(c) = P\left(\frac{\sqrt{n}(\bar{Y} - p)}{\sqrt{p(1-p}} \leq c\right)$$

   Denote the approximation by $\tilde{F}(c)$. What is $\tilde{F}(2)$? Calculate a specific number.

   (b) Find c^* in terms of c, n, p such that:

   $$F(c) = P\left(\frac{\sqrt{n}(\bar{Y} - p)}{\sqrt{p(1-p}} \leq c\right) = P(n\bar{Y} \leq c^*)$$

   How can you calculate $F(c)$ exactly? (Hint: what is the sampling distribution of $n\bar{Y}$?)

(c) For each of the values of n and p in the table below, calculate $F(2)$ exactly and fill in the first empty column. Calculate the absolute difference between the exact probability and the approximation you calculated earlier and fill in the second empty column. Round to three decimal places.

n	p	$F(2)$	$\lvert F(2) - \tilde{F}(2) \rvert$
16	0.01		
49	0.01		
144	0.01		
16	0.20		
49	0.20		
144	0.20		

(d) Often it is said that the approximation is good when n is "large enough". What do you notice about the approximation error as n increases?

(e) However, "large enough" varies in different contexts. What do you notice about the approximation error for larger sample sizes when p is small? Do you think that $n = 50$ is sufficient to achieve a good approximation if $p = 0.01$?

2. The WHO classifies body weight categories based on BMI as follows.

Underweight	less than 18.5
Healthy	18.5 - 24.99
Overweight	25 - 29.99
Obsese	30 or greater

Data on 134 Americans from the CDC's National Health and Nutrition Examination Survey indicated that average BMI was 28.8 kg/m^2. We will assume that the variance is 64. Let Y_i denote the BMI of the ith survey respondent, and assume that the data are collected from a random sample of Americans. Let $\bar{Y}$ denote the (random) sample mean and $\bar{y}$ denote the (observed) sample mean 28.8.

(a) Identify any given parameters and quantities other than $\bar{y}$ and write a transformation of $\bar{Y}$ that is approximately standard normal. (The transformation can include unknown parameters, but any known parameters should be replaced by their given values.)

(b) Suppose that the mean BMI were the midpoint of the healthy range. Under this assumption, calculate the probability that the sample mean exceeds 28.8.

(c) Do you think the data suggest Americans are overweight on average? Why or why not?