S1201, Summer 2022

100 points

Final Exam

Aug 10 6:15-7:50PM EST

Name: **UNI:**

Instructions:

(i) Please submit a single PDF file on courseworks. You may scan your written solutions or directly write the solutions on your tablet. Please write your answers on the exam paper.

(ii) The exam is scheduled on Aug 10, 6:15-7:50 PM EST. Unless you have the permission from Ye or Yizi, please stop writing at 7:50 PM. You should scan and upload your solutions on Courseworks by 8:00 PM.

(iii) Please **keep your video on** during the exam. Make sure your computer camera is angled so I can see you writing at all times.

(iv) You are allowed to use a hand-held calculator. You can't use the calculator on your smartphone/iPad/computer.

(v) This is a **close-book/slides** exam. You can't access any course materials during the exam, but you are allowed to use **two** double-sided pages of cheat sheet (of any size).

(vi) If you finish and want to submit your exam early, please message Ye first. Then you can submit your exam on Courseworks and quit from Zoom. I will mark the time stamp for your message and the submission.

(vii) During the exam, feel free to send a private chat to Ye for any questions!

(viii) Don't cheat! Cheating will lead to a straight zero, and will be reported to the university.

(ix) There are 100 points plus 10 bonus points, but the maximum score is 100.

(x) Good luck!

Please sign below to indicate your agreement with the Columbia College Honor Code, whether or not you are a student of Columbia College.

I affirm that I will not plagiarize, use unauthorized materials, or give or receive illegitimate help on assignments, papers, or examinations. I will also uphold equity and honesty in the evaluation of my work and the work of others. I do so to sustain a community built around this Code of Honor.

Signature:

1. **$2 \times 10 = 20$ points** TRUE/FALSE questions. No explanations are needed.

F (a) p-value is the probability that H_0 is true.

F (b) In all cases, it's possible to make Type-1 and Type-2 errors small at the same time.

F (c) In an estimation problem, based on the current data, the 99% confidence interval of parameter μ is $[1.2, 1.4]$. Then the probability that $[1.2, 1.4]$ covers the true value of μ is 99%.

T (d) All else equal, a 95% confidence interval will be wider than a 90% confidence interval.

T (e) Normal distribution is unimodal.

T (f) Exponential distribution and Geometric distribution have the memoryless property.

F (g) If we are drawing 5 balls **with replacement** (i.e. draw one each time and then put it back, then draw again) from a bag of 5 red balls and 5 black ones, then the number of black balls we draw follow a Hypergeometric distribution.

F (h) For a binomial random variable with n and p, $\mathbb{P}(X = 0) \neq \mathbb{P}(X = n)$ when $p = 1/2$.

T (i) According to Central Limit Theorem, we can approximate the binomial distribution by normal distribution when n and $n \min(p, 1-p)$ are large.

F (j) If X_1, X_2 and X_3 is a sample from some distribution with mean μ and variance σ^2, and they are independent with each other, then $\mathbb{E}\bar{X} = \mu$, and $\mathrm{Var}(\bar{X}) = 3\sigma^2$.

2. **$6 \times 3 = 18$ points** TV advertising agencies face increasing challenges in reaching audience members because viewing TV programs via digital streaming is gaining in popularity. The Harris poll reported on November 13, 2012, that 53% of 2343 American adults surveyed said they have watched digitally streamed TV programming on some type of device.

   (a) Calculate and interpret a confidence interval at the 99% confidence level for the proportion of all adult Americans who watched streamed programming up to that point in time.

   (b) Based on your confidence interval, do these results suggest that the proportion is less than 30%? Why?

   (c) If we want to narrow down our CI width while keeping the same significance level, what should we do?

Solution: (a) $\hat{p} = 0.53$, $n = 2343$.

$$99\% \text{ CI} = \left[\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\, z_{0.005} ,\ \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\, z_{0.005} \right]$$

$$= \left[0.53 - \sqrt{\frac{0.53 \times 0.47}{2343}} \times 3.3 ,\ 0.53 + \sqrt{\frac{0.53 \times 0.47}{2343}} \times 3.3 \right]$$

$$\approx [0.4960 ,\ 0.5640]$$

(b) No because 0.3 is below the CI.

(c) Increase sample size n. Q.

3. **5 × 4 = 20 points** Suppose a particular state allows individuals filing tax returns to itemize deductions only if the total of all itemized deductions is at least \$5000. Let X (in 1000s of dollars) be the total of itemized deductions on a randomly chosen form. Assume that X has the pdf

$$f(x) = \begin{cases} \alpha/x^{\cancel{2}4}, & x > 2, \\ 0, & \text{elsewhere.} \end{cases}$$

   (a) What's the value of α?
   Hint: When will f be a valid pdf?

   (b) Calculate $\mathbb{E}X$ and $\mathbb{E}(X^2)$.

   (c) Calculate $\text{Var}(X)$.
   Hint: Use the shortcut formula and the numbers in (b).

   (d) What's the cdf of X?

Solution:

(a) $\int_2^{+\infty} f(x)\,dx = 1 \Rightarrow \alpha \cdot \left(-\frac{1}{3x^3}\right)\Big|_2^{+\infty} = 1 \Rightarrow \alpha = 24$

(b) $\mathbb{E}X = \int_2^{+\infty} x f(x)\,dx = \int_2^{+\infty} \frac{24}{x^3}\,dx = -\frac{12}{x^2}\Big|_2^{+\infty} = 3$

$\mathbb{E}X^2 = \int_2^{+\infty} x^2 f(x)\,dx = \int_2^{+\infty} \frac{24}{x^2}\,dx = -\frac{24}{x}\Big|_2^{+\infty} = 12$

(c) $\text{Var}(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 12 - 9 = 3$

(d) $F(x) = \int_{-\infty}^{x} f(t)\,dt$

$$= \begin{cases} 0, & \text{if } x \le 2 \\ \int_2^x \frac{24}{t^4}\,dt, & \text{if } x > 2 \end{cases}$$

$$= \begin{cases} 0, & \text{if } x \le 2 \\ 1 - \frac{8}{x^3}, & \text{if } x > 2. \end{cases}$$

□.

4. **$6 \times 3 = 18$ points** To determine whether the pipe welds in a nuclear power plant meet specifications, a random sample of welds is selected, and tests are conducted on each weld in the sample. Weld strength is measured as the force required to break the weld. Suppose the specifications state that mean strength of welds should exceed 100 (lb/in^2). As part of the inspection team, answer the following questions:

   (a) How will you formulate the hypothesis testing problem? More specifically speaking, what are the null and alternative hypotheses?

   (b) In the context of this situation, describe type I and type II errors.

   (c) Which type of error would you consider more serious? Explain.

Solution: (a) H_0 : the mean strength ≤ 100 lb/in^2 (or "=")

H_a : - - - - - - - - - > 100 lb/in^2

(b) Type-1 error: the mean strength ≤ 100 lb/in^2 but we claim that it > 100 lb/in^2

Type-2 error: the mean strength > 100 lb/in^2 but we claim that it ≤ 100 lb/in^2

(c) Type-1 error. □.

5. **8 × 3 = 24 points** Suppose only 75% of all drivers in a certain state regularly wear a seat belt. A random sample of 500 drivers is selected.

   (a) What's the distribution of the number of the drivers in the sample who regularly wear a seat belt?

   (b) Can we use normal distribution to approximate the distribution in (a)? If yes, explain why and write down the normal distribution. If no, explain why.

   Hint: Check the conditions before using the normal approximation.

   (c) What is the probability that between 360 and 400 (inclusive) of the drivers in the sample regularly wear a seat belt?

   Hint: Don't forget to use the continuity correction.

Solution: (a) Bin(n = 500, p = 0.75).

(b) Yes.

Check: $n > 30$

$n \cdot \min\{p, 1-p\} = 125 > 30.$

$N(np, np(1-p)) = N(125, 93.75)$

(c) Denote the number of drivers in the sample who regularly wear a seat belt as X.

$\Rightarrow$ By (b), $X \overset{d}{\approx} N(375, 93.75)$

$$\Rightarrow P(360 \le X \le 400) = P\left(\frac{360-375-0.5}{\sqrt{93.75}} \le \frac{X-375}{\sqrt{93.75}} \le \frac{400-375+0.5}{\sqrt{93.75}}\right)$$
$$= P(-1.60 \le Z \le 2.63)$$
$$= \Phi(2.63) - \Phi(-1.60)$$
$$= 0.9957 - 0.0548$$
$$= 0.9409 \quad \square.$$

6. **Extra question with 10 bonus points** Read the following story from the memoir of a famous statistician George E P Box, "An Accidental Statistician."

   "While I was at University College, there was a day set aside for visiting relatives. My mother came, and I showed her the Statistics Department. Next door to it was the Genetics Department. There I showed her some Ishihara charts that have colored dots arranged so that if you had normal vision, you would see one number, and if you were red-green colorblind, you would see another. I showed one to my mother and said, "If you were colorblind, you would see a six." She said, "Well it is a six." Dr. Kalmus, the professor of genetics, happened to be passing by and asked, "Is someone colorblind?" I replied, "Yes, my mother." "I don't think so," he said. Of course it turned out that *I* was colorblind. The professor explained that only 0.4% of women are colorblind, as opposed to 7% of men. At this point, I was 26 years old, had been in the Army, and through the war without a clue about this."

   Question: Do you think the professor's explanation is convincing? Why? (Any reasonable answers)

Any reasonable answer.