

LAST NAME (Please Print): _____

FIRST NAME (Please Print): _____

HONOR PLEDGE (Please Sign): _____

Statistics 111

Midterm 2

- This is a closed book exam.
- You may use your calculator and a single page of notes.
- The room is crowded. Please be careful to look only at your own exam. Try to sit one seat apart; the proctors may ask you to randomize your seating a bit.
- Report all numerical answers to at least two correct decimal places or (when appropriate) write them as a fraction.
- All question parts count for 1 point.

Some possibly useful formulæ:

The Gamma(α, β) distribution has mean α/β , variance α/β^2 , and density function

$$f(x) = \frac{\beta^\alpha}{(\alpha - 1)!} x^{\alpha-1} \exp(-\beta x) \text{ for } 0 \leq x \text{ and } \alpha > 0, \beta > 0.$$

The Beta(α, β) has mean $\alpha/(\alpha + \beta)$, variance $\alpha\beta/[(\alpha + \beta)^2(\alpha + \beta + 1)]$, and density function

$$f(x) = \frac{(\alpha + \beta - 1)!}{(\alpha - 1)!(\beta - 1)!} x^{\alpha-1} (1 - x)^{\beta-1} \text{ for } 0 \leq x \leq 1 \text{ and } \alpha > 0, \beta > 0.$$

1. Assume the number of phone calls you receive in an hour has a Poisson distribution with parameter λ . Let X be the number of calls you receive between noon and 1 p.m., and let Y be the number of calls you receive between noon and 3:00 p.m.

_____ What is the correlation between X and Y ?

2. You own 2 shares of Apple, 3 shares of Google, and 4 shares of Alcoa. The annual return on investment for an Apple share is normally distributed with mean 1 and sd 2, or $N(1,2)$. The annual share-wise ROI for Google is $N(2,3)$ and for Alcoa is $N(3,4)$. Since Apple and Google are in the same sector, their ROIs are have correlation r .

_____ If $r = 0.5$, what is the covariance in shares of Apple and Google?

_____ Suppose the covariance between Apple and Google is 2. What is your expected total ROI?

_____ Suppose the covariance between Apple and Google is 2. What is the sd on your total ROI?

_____ What is the probability that your total ROI exceeds 15.8?

3. Suppose x_1, \dots, x_n are a random sample from an exponential distribution with parameter λ . You do not know λ but you have a prior distribution for it. Your prior is also exponential, with parameter 2.

What is your posterior distribution for λ ? _____

Under what circumstances would the median of your posterior distribution be a good guess?

4. Let 0.8 and 0.75 be a random sample from the Beta distribution with unknown α . However, you know that $\beta = 1$.

_____ What is the maximum likelihood estimate (MLE) of α ?

_____ What is the MLE of the mean of this Beta distribution?

5. You believe that your probability p of answering True or False questions correctly on a statistics test has a Beta distribution with mean 0.5 and variance 0.1. When your test comes back, you see that you got only 5 of the 15 T/F questions.

What is your new distribution for p ? _____

What is your best one-number guess about p under squared error loss? _____

6. Suppose you take 32 courses during your time at Duke. In each class, the grade you get is a random variable. An A+ counts for 12 points, an A counts for 11, and so forth. Assume your probability of getting a B is 0.25, your probability of a B+ is 0.2, your probability of a B- is 0.2, and all other grades are equally likely.

_____ What grade point average do you expect?

_____ What is the standard error in your average grade?

_____ What is the (approximate) probability that your average grade at graduation is greater than 7.5?

7. Durham has 25 fast-food chain restaurants. You visit 10 of them, and find that the average sanitation score is 90.2 with a sample sd of 5.

_____ Set a one-sided lower 95% confidence interval on the average sanitation score in such restaurants.

8. Jed Bartlett is running for president. His team polls 100 voters in Durham, and finds that 55 plan to vote for him. Before pulling their ad funding for Durham, his two top strategists set confidence intervals on his support. Josh Lymon sets a two-sided 95% interval; Toby Zeigler sets a one-sided lower 95% interval.

Josh's interval: L = _____ U = _____

Toby's interval: L = _____

Which advisor is correct and why?

9. Assume X_1, \dots, X_n are a random sample from the $\text{Gamma}(\alpha, 3)$ distribution.

_____ Find an unbiased estimate of α .

_____ What is the standard error of your estimate?

10. You estimate the probability of Heads on a coin by tossing it n times, counting the number of Heads, and dividing by $n - 1$. What is the bias, variance, and mean squared error in your estimator?

bias = _____ variance = _____

mean squared error = _____

11. Consider the linear congruential generator $X_n \equiv 7X_{n-1} \pmod{5}$. Set the seed for the random number generator to be 11.

_____ What is X_2 ?

12. You think that women may be more likely to vote Democrat than men. You sample 100 women in NC and find that 60 support the Democrat Kay Hagan. You sample 200 men in NC and find that 150 support Thom Tillis, the Republican. Set an approximate two-sided confidence 90% interval on the difference between the proportion of female Democrats and male Democrats in NC (subtract men from women).

L = _____ U = _____

13. The Medicare program maintains a sample of 5% of its participants. When a person is selected into the sample, they are followed until their death, and Medicare tracks the services they use and the associated costs. When someone dies, or when the number of enrollees increases (as has happened each year since the program began, due to the Baby Boomers), additional people are enrolled, and these are chosen at random from all current Medicare patients not currently in the sample. In what way is the sample not representative of the population of Medicare patients?

14. True or False. List only the true statements. _____
- A.** There was a stomach cancer hotspot centered on Pittsburgh.
 - B.** Response bias occurs if the wording of a survey question affects the answer.
 - C.** Respondent bias occurs if people who refuse to answer are unlike those who do.
 - D.** Holding all else constant, as your confidence level increases, the width of your confidence interval increases.
 - E.** Holding all else constant, as your variance increases, the width of your confidence interval increases.
 - F.** Holding all else constant, as your sample size increases, the width of your confidence interval increases.
 - G.** Mean squared error is the sum of the variance and the bias.
 - H.** Linear combinations of normal random variables are normally distributed.
 - I.** As the sample size increases, maximum likelihood estimates have minimum MSE.
 - J.** John Snow stopped a smallpox outbreak by removing the handle from the Broad Street pump.
 - K.** The population pyramid in a developing nation tends to be short and wide.