CE 311S: Final Exam

Thursday, May 1

1:00 – 3:00 PM

Name _____

Instructions:

- **SHOW ALL WORK** unless instructed otherwise. No shown work means no partial credit!
- If you require additional space, you may use the back of each sheet and/or staple additional pages to the end of the exam.
- If you need to make any additional assumptions, state them clearly.
- You may use a calculator and one regular-sized sheet of notes. No additional resources are permitted. Please turn in this sheet with your exam.
- The number of points associated with each part of each problem is indicated.

Problem	Points	Possible
1		20
2		20
3		20
4		20
5		20
TOTAL		100

Problem 1. (20 points). As you recall, at the end of Exam 2 you found Robin on the trail with a backpack full of old CE 311S exams. You confront Robin, who claims that they have the exams because they are trying to train a new AI app to solve probability and statistics questions. Robin invites you to be a partner in developing the app. You aren't sure if this is a good idea (most of Robin's ideas aren't) but decide to play along while investigating further.

You help Robin test the app by coming up with 10 true/false questions, and seeing if the app can answer the questions better than random chance; let p be the probability that the app answers a given true/false question correctly.

(a) (5) State your null and alternative hypotheses, both mathematically *and* in plain English without using any statistical jargon.

(b) (5) You decide to use the number of correct answers (out of 10) as your test statistic. Which of the following rejection regions is most appropriate, and why? $\{0, 1, 2\}$, $\{0, 1, 9, 10\}$, $\{4, 5, 6\}$, or $\{8, 9, 10\}$.

(c) (5) For the rejection region you chose, what is the probability of a Type I error?

(d) (5) It turns out that the app actually answers questions correctly 60% of the time. For the rejection region you chose, what is the probability of a Type II error?

Problem 2. (20 points). When you show the results, Robin gets defensive and insists that the poor performance on true/false questions is because Dr. Boyles doesn't ask that kind of question on his exams. You try again, looking at 5 exams from the past, all with the same number of questions. The following data set indicates the number of problems on each exam that were answered correctly:

$$2 \quad 3 \quad 0 \quad 3 \quad 2$$

For this question, write down **one-sided** intervals with 95% confidence, based on the idea that you need to answer a certain number of questions correctly to pass the class.

(a) (5) What are the sample mean and sample standard deviation?

(b) (5) Form an interval containing the mean number of questions answered correctly.

(c) (5) Form an interval containing the number of questions answered correctly on this year's final exam.

(d) (5) Form an interval containing the number of questions answered correctly on *every one of the next five years'* final exam.

Problem 3. (20 points). Over the current semester, Robin has five friends try the app. You track how many times someone logs on to your app (x), and their GPA (y), to see if there is a relationship between these variables. You find $\sum x_i = 83$, $\sum y_i = 11.6$, $\sum x_i^2 = 1395$, $\sum y_i^2 = 27.7$, and $\sum x_i y_i = 190.4$.

(a) (5) What is the best-fit linear regression line relating GPA to number of logins?

(b) (5) What is the R^2 value?

(c) (4) What is the t-statistic on your regression line (for the null hypothesis that there is no relationship between Robin's app and GPA)?

(d) (6) Use your answers to (a)–(c) to answer these three questions (explaining specifically why your findings lead to these answers): does there seem to be a positive or negative relationship between usage of your app and GPA? does this relationship seem to explain a lot of someone's GPA, or a little? are you confident that this relationship exists in reality, or is it just a coincidence from your sample?

Problem 4. (20 points). For each of the following random variables (A, B, C, and D), indicate (a) the name of the distribution that best describes that random variable; (b) what its expected value is; and (c) what its standard deviation is.

(a) (5) You are still suspicious of Robin's motives, so when you find their phone lying in the apartment, you decide to snoop and see if you can find out what's really going on. Luckily for you, Robin chooses weak passwords, so there is a 30% chance you can log into any given app on Robin's phone (independent of whether you can log into any other app). You try to log into 5 messaging apps; the random variable A is the number of these apps that you are able to log into.

(b) (5) You gain access to Robin's Signal account, and find a number of messages from a mysterious user named `sdbatx` about the 311S app. The number of messages seems to follow a Poisson distribution, with an average of 2 messages per day. The random variable B is the number of *hours* until the next message from `sdbatx` is received.

(c) (5) You study the messages from `sdbatx` to try to discover who this really is. You find the number of words in each message follows a negative binomial distribution with $m = 10$ and $p = 0.2$. The random variable C is the *average* number of words in the next 100 messages.

(d) (5) After receiving the next 7 messages from `sdbatx`, you have enough information to unmask their real identity. The random variable D is the time elapsed between now, and when the 7th message is received.

Problem 5. (20 points). After following the clues, you find the `sdbatx` username on a chess website and... shock horror! Robin has been communicating with Dr. Boyles all along!!! You realize that Dr. Boyles has been orchestrating all the events on this semester's exams as part of a master plan to trap Robin on academic dishonesty charges. He is part of a vigilante group of professors catching shady people who make apps to cheat on engineering courses. You discover that this group has tracked down a number of other cheaters; the time (in days) to catch the last 5 cheaters is given by:

$$60 \quad 40 \quad 30 \quad 30 \quad 40$$

(a) (10) What is the maximum likelihood estimate of λ, assuming that the time to catch a cheater is described by an exponential distribution?

(b) (10) Assuming your estimate of λ is correct, what is the median time to catch a cheater? What is the 90th percentile?