CE 311S: Final Exam
Friday, May 15
2:00 – 5:00 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.

- The number of points associated with each part of each problem is indicated.

Problem	Points	Possible
1		15
2		30
3		30
4		15
5		10
TOTAL		100

Please copy the following statement (based on UT's honor code and the ASCE code of ethics) in your own handwriting, and sign it. For the purposes of this statement, academic dishonesty includes (but is not limited to) sharing with or receiving information from others about the exam, by any mode of communication.

"As a student of The University of Texas at Austin and as a civil engineer, I certify that I have not and will not participate in any acts of academic dishonesty related to this exam. If I witness any acts of academic dishonesty, I will report them to the instructor."

Problem 1. (15 points). You are a Technology Evangelist working for Microsoft (they really do have people with this title). Your job is to visit potential business customers in Austin and convince them that Microsoft products are superior to whatever they are currently using. Assume that there are currently 70,000 small businesses in the region and this constitutes your potential client base. Your primary performance is based on the proportion of clients you have that use MS products. As long as 30% of your potential client base is using MS products, you will receive satisfactory evaluations. You call a sample of 35 businesses and ask them if they currently use MS products. You find that 15 of the businesses are current MS clients.

(a) (5) Is this sample size large enough for the central limit theorem to apply? Explain your answer.

(b) (10) Regardless of your answer to (a), assume the sample size is large enough and conduct a hypothesis test (with 5% significance) to see if you can state that you are meeting the goal of more than 30% of clients using MS products: state your null and alternative hypotheses, and whether you can reject the null hypothesis or not.

Problem 2. (30 points). Not trusting your judgment (or statistical abilities), Microsoft performs an extensive study and finds that precisely 25,000 of your potential client base currently uses MS products. Further, they have selected 5 specific businesses at random for you to focus your efforts on (evidently they thought you needed a smaller job). Let X denote the number of these 5 businesses that currently use MS products.

(a) (5) What is the complete PMF of X?

(b) (2) What is the probability that more than two of the businesses currently use MS products?

(c) (5) What are the mean and standard deviation of X?

(d) (6) If every business that currently uses MS products spends $3000/yr on them, what are the mean and standard deviation of the amount of money currently collected from these five businesses?

(e) (12) For any company not currently using MS products, there is an independent 25% chance you will get them to switch to MS each time you visit them. How many visits do you expect to make to get all 5 companies using MS products?

Problem 3. (30 points) To better convince your potential clients that they should switch to MS, you conduct a study to relate the use of MS products with business revenues. You use 3 different levels of MS products ($1 =$ none, $2 =$ basic, $3 =$ premium) which you denote as X. While generated revenue is clearly a continuous value you create discrete categories of revenue for the analysis due to painful memories regarding continuous joint distributions previously in your life. Denote by Y the level of generated revenue ($1 =$ poor, $2 =$ good, $3 =$ excellent). You conduct an exhaustive study of your potential client base and find the following data for the joint distribution of X and Y:

		Y		
		1 (poor)	2 (good)	3 (excellent)
	1 (none)	0.04	0.06	0.10
X	2 (basic)	0.09	0.09	0.12
	3 (premium)	0.25	0.10	0.15

(a) (5) What are the mean values of X and Y?

(b) (5) Are X and Y independent? (Explain your answer.)

(c) (10) What is the correlation coefficient between X and Y?

(d) (4) If a given client is not using MS products, what is the probability that their revenue is excellent? What is the probability that a client chosen at random will not be using an MS product and have excellent revenues?

(e) (6) Looking only at the clients not using MS products, what average revenue do you see? Repeat this for clients using MS products at the basic and premium levels.

Problem 4. (15 points). Now struggling with your bosses, you try to build a case for the stability (or lack thereof) of your products with linear regression. You collect data on the number of MS products a company uses (denoted x) versus the number of complains they generate in a week (denoted y), producing the following data points:

MS products	1	3	4	5	8	10	12
Complaints per week	2	4	8	14	20	24	28

(a) (5) State the linear regression model for this data.

(b) (5) What is R^2 for this model?

(c) (5) If a company uses 9 MS products, what is the probability they generate more than 30 complaints in a week?

Problem 5. (10 points). Having been fired by Microsoft, you consider opening up your own Linux distribution and support company. Before going forward you decide to gauge the desire of current MS customers to this possibility. Due to your dwindling funds and time constraints, you can survey only 6 potential clients and ask them how much they would be willing to spend on Linux licenses and support per month. You get the following data for revenues (assume that it behaves as a normal distribution):

$$680 \quad 570 \quad 460 \quad 710 \quad 960 \quad 420$$

(a) (5) Set up a hypothesis test to see if you can earn at least $400 on average from your clients. State your null and alternative hypotheses, whether you can reject the null hypothesis or not, **and** state whether this means you will open the business or not.

(b) (5) As your are running low on funds, it is important that you begin earning money very quickly. To get an idea of your near-term revenue, create a 95% confidence interval for the revenue from your very first client.