Name: EID: Time of your class:

Midterm Exam #1, STA 371G, Statistics and Modeling, Spring 2016

- You may tear off the blank page at the end as a scratch paper. Please turn in the scratch paper with the exam. The exam is closed-book. You are allowed 1 page (letter/A4 size) of notes. You may use a calculator.
- Please answer all problems in the space provided on the exam. The full score is 100.
- Read each question carefully and clearly present your answers.
- You must show all your work and give a complete explanation. No credit will be given for only the answer without an explanation/equation.

$Z \sim N(0,1)$

x	P(Z<x)	x	P(Z<x)	x	P(Z<x)	x	P(Z<x)
-3	0.0013	-1.5	0.0668	0	0.5	1.5	0.9332
-2.95	0.0016	-1.45	0.0735	0.05	0.5199	1.55	0.9394
-2.9	0.0019	-1.4	0.0808	0.1	0.5398	1.6	0.9452
-2.85	0.0022	-1.35	0.0885	0.15	0.5596	1.65	0.9505
-2.8	0.0026	-1.3	0.0968	0.2	0.5793	1.7	0.9554
-2.75	0.003	-1.25	0.1056	0.25	0.5987	1.75	0.9599
-2.7	0.0035	-1.2	0.1151	0.3	0.6179	1.8	0.9641
-2.65	0.004	-1.15	0.1251	0.35	0.6368	1.85	0.9678
-2.6	0.0047	-1.1	0.1357	0.4	0.6554	1.9	0.9713
-2.55	0.0054	-1.05	0.1469	0.45	0.6736	1.95	0.9744
-2.5	0.0062	-1	0.1587	0.5	0.6915	2	0.9772
-2.45	0.0071	-0.95	0.1711	0.55	0.7088	2.05	0.9798
-2.4	0.0082	-0.9	0.1841	0.6	0.7257	2.1	0.9821
-2.35	0.0094	-0.85	0.1977	0.65	0.7422	2.15	0.9842
-2.3	0.0107	-0.8	0.2119	0.7	0.758	2.2	0.9861
-2.25	0.0122	-0.75	0.2266	0.75	0.7734	2.25	0.9878
-2.2	0.0139	-0.7	0.242	0.8	0.7881	2.3	0.9893
-2.15	0.0158	-0.65	0.2578	0.85	0.8023	2.35	0.9906
-2.1	0.0179	-0.6	0.2743	0.9	0.8159	2.4	0.9918
-2.05	0.0202	-0.55	0.2912	0.95	0.8289	2.45	0.9929
-2	0.0228	-0.5	0.3085	1	0.8413	2.5	0.9938
-1.95	0.0256	-0.45	0.3264	1.05	0.8531	2.55	0.9946
-1.9	0.0287	-0.4	0.3446	1.1	0.8643	2.6	0.9953
-1.85	0.0322	-0.35	0.3632	1.15	0.8749	2.65	0.996
-1.8	0.0359	-0.3	0.3821	1.2	0.8849	2.7	0.9965
-1.75	0.0401	-0.25	0.4013	1.25	0.8944	2.75	0.997
-1.7	0.0446	-0.2	0.4207	1.3	0.9032	2.8	0.9974
-1.65	0.0495	-0.15	0.4404	1.35	0.9115	2.85	0.9978
-1.6	0.0548	-0.1	0.4602	1.4	0.9192	2.9	0.9981
-1.55	0.0606	-0.05	0.4801	1.45	0.9265	2.95	0.9984
-1.5	0.0668	0	0.5	1.5	0.9332	3	0.9987

Problem 1 (10 points)

A men's clothing store is running a special promotion on its polo shirts and limits each customer to buy up to four polo shirts. Let us denote X as the random number of polo shirts purchased by a customer, then X is equal to 0, 1, 2, 3, or 4. The manager currently thinks that the probabilities for these five possible random outcomes are in the ratio of 1 to 3 to 3 to 2 to 1. That is to say, $X = 1$ is three times more likely than $X = 0$.

(a) (**2 points**) Find the probability distribution of X.

(b) (**2 points**) What is the expected number of polo shirts bought by a customer visiting this clothing store?

(c) (**2 points**) How much variability exists around the expected number of polo shirts bought by a customer visiting this clothing store? (Hint: calculate the variance/standard deviation)

(d) (**2 points**) Suppose 200 customers visit this store on the first promotion day. Find the mean and standard deviation of the total number of polo shirts sold on that day.

(e) (**2 points**) During the promotion, what's the probability for a customer visiting the store to buy two or more polo shirts?

Problem 2 (8 points)

Suppose $X \sim \mathcal{N}(4, 16)$, i.e., X is normally distributed with mean 4 and variance 16. Compute:

(a) (2 points) $P(X = 4)$

(b) (2 points) $P(X > 8)$

(c) (2 points) $P(X \geq 8)$

(d) (2 points) $P(-4 < X < 8)$

Problem 3 (10 points)

It is known that 76% of the UT 2015 BBA graduates chose to work in Texas. For the UT 2015 BBA graduates who chose to work in Texas, 26%, 35%, 38%, and 1% of them chose to work in Austin, Dallas, Houston, and other parts of Texas, respectively. Suppose John and Laura are both 2015 BBA graduates, and whether John chose to work in Texas and whether Laura chose to work in Texas were two independent events.

(a) (2 points) If we randomly choose a UT 2015 BBA graduate, what is the probability that this person did not choose to work in Texas.

(b) (2 points) If we randomly choose a UT 2015 BBA graduate, what is the probability that this person chose to work in Austin, TX.

(c) (3 points) What's the probability that both John and Laura chose to work in Texas?

(d) (3 points) What' the probability that either John or Laura (not both of them) chose to work in Texas?

Problem 4 (10 points)

For a "Yes/No" question, suppose that the proportion of people in the population that would answer the question with "Yes" is p. If we survey n people randomly selected from a large population with this "Yes/No" question, where the population size is considerably larger than n, then the number of "Yes" from a random sample of n people can be considered as a binomial random variable

$$X \sim \text{Binomial}(n, p).$$

If n is not too small and p is not too close to 0 or 1, then $X \sim \text{Binomial}(n, p)$ can be further approximated with a normal random variable as

$$X \sim \mathcal{N}(np, np(1-p)),$$

where np is mean and $np(1-p)$ is the variance.

For Fall 2015, there were 6,968 applicants to McCombs BBA programs and 1,776 were admitted. Answer Question (a).

(a) (5 points) If you randomly select 70 applicants that applied for Fall 2015 McCombs BBA programs, what would be the 95% Confidence Interval of the number of admitted applicants among these randomly selected 70 applicants?

Suppose it is found in a recent survey that 30 out of 100 randomly selected McCombs BBA students have participated in the BBA studying abroad programs. Answer Question (b).

(b) (5 points) Based on this survey, find the 95% Confidence Interval for the true proportion of McCombs BBA students who have participated in the BBA studying abroad programs.

Problem 5 (20 points)

The figure below shows the 2015 McCombs BBA Salary Survey, which is based on 588 voluntary reports.

BBA Salary Survey 2015	Average	Median	Standard Deviation
Full-Time Overall Salaries	$61,419	$60,000	$13,538

Suppose the annual salary for a 2015 BBA graduate follows a normal distribution, whose mean is equal to the "Average" reported in the Survey and whose standard deviation is equal to the "Standard Deviation" reported in the Survey. Answer Questions (a)-(e).

(a) (4 points) Find the 95% Confidence Interval for the annual salary of a 2015 BBA graduate.

(b) (4 points) Find the probability for a 2015 BBA graduate to have an annual salary that is between \$47,881 and \$88,495.

(c) (4 points) If a 2015 BBA graduate decided to accept a job offer immediately if he/she received an offer with an annual salary that was among the top 16%, what would be the least amount of annual salary for him/her to accept the offer immediately?

(d) (4 points) Describe the distribution of the average annual salary of 16 randomly selected 2014 BBA graduates.

The average annual salary of \$61,419 reported in the survey provides an estimate of the TRUE average annual salary, which can only be obtained if all 2015 BBA graduates reported their salaries.

(e) (4 points) Does the average annual salary reported in this survey provide an accurate estimation of the true average annual salary of all 2015 BBA graduates? If Yes, provide your explanations and find the 95% Confidence Interval of the true average salary of all 2015 BBA graduates. If No, provide your explanations and give suggestions on how to improve the estimation accuracy.

Problem 6 (20 points)

The table below shows the fuel economy of four randomly selected cars. The weight of a car is measured in thousands of pounds and the fuel economy is measured in MPG (miles per gallon). Let X denote Weight and Y denote Fuel Economy (Note that four data points are usually far from enough in practice. We choose four points only for illustration purpose).

Weight (X)	2.9	4.0	3.4	3.7
Fuel Economy (Y)	31	21	29	23

(a) (2 points) Calculate the sample means of X and Y.

(b) (2 points) Calculate the sample standard deviations of X and Y.

(c) (2 points) Calculate the sample covariance between X and Y.

(d) (2 points) Calculate the sample correlation between X and Y.

(e) (2 points) Suppose we use simple linear regression to describe how the Fuel Economy changes as a linear function of the Weight. Calculate the least squares estimates of the intercept and slope.

(f) (2 points) What are the units of the intercept and slope.

(g) (2 points) Suppose we change the units of Weight from 1000 pounds to 1000 kilograms (1000 pounds = 453.59 kilograms), what would be the new intercept and slope?

(h) (2 points) Suppose we not only change the units of Weight from 1000 pounds to 1000 kilograms, but also change the units of Fuel Economy from MPG (miles per gallon) to KPL (kilometers per litter), where 1 MPG = 0.4251 KPL, what would be the new intercept and slope?

(i) (2 points) Calculate the coefficient of determination R^2 and explain its meaning.

(j) (2 points) Based on this analysis, briefly describe your understanding of the relationship between the Weight and Fuel Economy of a car.

Problem 7 (22 points)

The federal Class III milk price, although not the same as, is closely related to the California mailbox price that a milk farmer in California receives for his milk. Based on the monthly milk price data from January 2004 to May 2007, one can run a simple linear regression model to regress the federal Class III milk price on the California mailbox price. The milk price is measured with \$/cwt, where cwt is a unit of measurement that is roughly 100 pound of milk.

The simple linear regression results are presented in the table below:

	A	B	C	D	E	F	G	H
1	**Linear Regression**							
2								
3	**Regression Statistics**							
4	*R*							
5	*R Square*	**B5=?**						
6	*Adjusted R Square*	0.92495						
7	*Standard Error*	0.59885						
8	*Total Number Of Cases*	41						
9	**Class III Price = b0 + b1 * Mailbox Price**							
10								
11	**ANOVA**							
12		*d.f.*	*SS*					
13	*Regression*	1.	**C13=?**					
14	*Residual*	39.	13.99					
15	*Total*	40.	191.15					
16								
17		*Coefficients*	*Standard Error*	*LCL*	*UCL*	*t Stat*	*p-level*	*H0 (5%) rejected?*
18	**Intercept**	-1.56	0.70445	-2.98209	-0.13231	-2.21052	0.033	*Yes*
19	**Mailbox**	1.18	0.05319	1.0746	1.28977	22.2262	0.E+0	*Yes*
20	*T (5%)*	2.02269						
21	*LCL - Lower value of a reliable interval (LCL)*							
22	*UCL - Upper value of a reliable interval (UCL)*							

Based on the results presented in the table, answers Questions (a)-(e).

(a) (2 points) Suppose the estimated simple linear regression line is expressed as

$$\text{Class III Price} = b_0 + b_1 \times \text{Mailbox Price},$$

what's the values of b_0 and b_1?

(a) (2 points) What's the units of b_0 and b_1?

(c) (2 points) What's the value in cell C13?

(d) (2 points) What's the value in cell B5?

(e) (2 points) We choose the values of b_0 and b_1 to minimize which value in the table?

Consider the regression model

$$\text{ClassIII}_t = \beta_0 + \beta_1 \text{Mailbox}_t + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2),$$

where ClassIII_t represents the milk price in month t for the federal Class III milk and Mailbox_t represents the California mailbox price in month t. Supposing it is true that $\beta_0 = b_0$, $\beta_1 = b_1$ and $\sigma = 0.60$, answers Questions (f)-(h).

(f) (4 points) Suppose the California Mailbox Price is \$10/cwt in June 2014, what's the 95% Prediction Interval for the price of the federal Class III milk in that month?

(g) (4 points) Suppose the California Mailbox Price is \$10/cwt in August 2014, what's the probability that the federal Class III milk in that month will be greater than \$16/cwt ?

(h) (4 points) In order to hedge the risk of low milk price in California, in February 2014, a California milk farmer purchased an August 2014 put option on the federal Class III milk with a strike price of \$14/cwt. The payoff from the put option is zero, if the strike price is lower than or equal to the Class III milk price, and is equal to the strike price of the put option MINUS the Class III milk price if the strike price is higher than the Class III milk price.

Suppose the California Mailbox Price is \$10/cwt in August 2014, and the total cost of purchasing and trading the August 2014 put option is \$0.76/cwt, what's the probability that this farmer will make a net revenue (mailbox price PLUS payoff from the put option MINUS cost of purchasing and trading the put option) of more than \$13/cwt for his milk in August 2014?

Name: EID:

Please use this page as your Scratch Paper. This page must be turned in with the exam.