Midterm Exam #1 Name: EID:

STA 371G, Statistics and Modeling, Spring 2014
12:30-1:50 PM (session 1), 2:00-3:20PM (session 2)

- 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.
- The exam is closed-book. You are allowed one page of notes. You may use a calculator.

Z ~ 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 (5 points)

Suppose that Jennifer has a 90% probability to eat breakfast and David has a 30% probability to eat breakfast. Moreover, assume whether Jennifer eats breakfast is not related to whether David eats breakfast.

(a) (2 points) What's the probability that both Jennifer and David eat breakfast tomorrow morning?

$$90\% \times 30\% = 27\%$$

(b) (3 points) What's the probability that either Jennifer or David (not both of them) eats breakfast tomorrow morning?

$$90\% \times (1 - 30\%) + (1 - 90\%) \times 30\% = 66\%$$

Problem 2 (25 points)

The figure below shows the 2013 McCombs BBA Intern Salary Survey, which is based on 634 voluntary reports.

BBA Intern Salary Statistics

BBA INTERN SALARY SURVEY 2013

Internship Overall Salaries		
Average (Monthly): \$3,301	Median (Monthly): \$3,466	Standard Deviation (Monthly): \$1,297

Assume the monthly salary for a BBA intern in 2013 follows a normal distribution, whose mean is equal to the Average Monthly Salary reported in the Survey and whose standard deviation is equal to the Standard Deviation reported in the Survey. Answer Questions (a)-(e).

(a) (5 points) Find the 95% Confidence Interval for the monthly salary of a BBA intern in 2013.

Since $X \sim \mathcal{N}(3301, 1297^2)$, we have the 95% Confidence Interval as

$$3301 \pm 2 \times 1297$$

$$(707, 5895)$$

(b) (5 points) Find the probability for a BBA intern in 2013 to have a monthly salary that is between \$3,301 and \$5,895.

$$P(3301 < X < 5895) = P(X < 5895) - P(X \leq 3301) = P(Z < \frac{5895 - 3301}{1297}) - P(Z \leq 0)$$
$$= P(Z < 2) - P(Z \leq 0) = 0.9775 - 0.5 = 0.4775$$

Or because 5895 is two standard deviation from the mean, $P(3301 < X < 5895) \approx 95\%/2 = 47.5\%$

(c) (5 points) If a BBA student decides to accept an intern offer immediately if he/she receives an offer with a monthly salary that is among the top 16% of all accepted intern offers in 2013, what would be the least amount of monthly salary that will make him/her accept the offer immediately?

Suppose the least amount of monthly salary is x, then

$$P(X \geq x) = 0.16$$

$$P(X < x) = 0.84$$

$$P(\frac{X - 3301}{1297} < \frac{x - 3301}{1297}) = 0.84$$

$$P(Z < \frac{x - 3301}{1297}) = 0.84$$

$$\frac{x - 3301}{1297} = 1$$

$$x = 3301 + 1297 = 4598$$

(d) (5 points) If you take a random sample of 16 BBA interns in 2013, would the distribution of their average monthly salary be the same as the distribution of the monthly salary of a BBA intern? If Yes, provide your explanations. If No, describe the distribution of their average monthly salary.

No. $\bar{X} \sim \mathcal{N}(3301, (1297/\sqrt{16})^2)$, or $\bar{X} \sim \mathcal{N}(3301, (324.25)^2)$, or $\bar{X} \sim \mathcal{N}(3301, 105138.1)$

The Average Monthly Salary of $3,301 reported in the survey provides an estimate of the TRUE average monthly salary, which can only be obtained if all BBA interns in 2013 reported their salaries.

(e) (5 points) Does the Average Monthly Salary reported in this survey provide an accurate estimation of the true average salary of all BBA interns? If Yes, provide your explanations and find the 95% Confidence Interval of the true average salary of all BBA interns. If No, provide your explanations and give suggestions on how to improve the estimation accuracy.

Yes. $3301 \pm 2 \times \frac{1297}{\sqrt{634}}$, or 3301 ± 103, or $(3198, 3404)$

Comment: if your answers differed from mine but you provided your arguments for these yes/no questions, you were given partial or even full credits.

Problem 3 (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 2013, there were 7,495 applicants to McCombs BBA programs and 1,628 were admitted. Answer Question (a).

(a) (5 points) If you randomly select 100 applicants that applied for Fall 2013 McCombs BBA programs, what would be the 95% Confidence Interval of the number of admitted applicants among these randomly selected 100 applicants?

Since $p = 1628/7495 = 0.217$, so approximately

$$X \sim \mathcal{N}(100 * 0.217, 100 * 0.217 * (1 - 0.217)) = \mathcal{N}(21.7, 4.1^2)$$

Thus the 95% Confidence Interval is

$$21.7 \pm 2 \times 4.1$$

$$\text{or } [13, 30]$$

Suppose it is found in a recent survey that 32 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 percentage of McCombs BBA students who have participated in the BBA studying abroad programs.

Since $\hat{p} = 32/100 = 0.32$, we have the 95% Confidence Interval for the true percentage as

$$0.32 \pm 2 \times \sqrt{\frac{0.32 \times (1 - 0.32)}{100}} = 0.32 \pm 0.093$$

$$(0.227, 0.413)$$

Problem 4 (20 points)

The table below shows the MPG (miles per gallon) of four randomly selected cars. The weight of a car is measured in thousands of pounds. Let X denote Weight and Y denote MPG.

Weight (X)	3.5	2.7	3.6	2.2
MPG (Y)	17	19	13	31

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

$\bar{x} = 3$, $\bar{y} = 20$

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

$s_x = 0.668$, $s_y = 7.746$

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

$Cov(X, Y) = -4.733$

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

$r_{xy} = Corr(X, Y) = \frac{-4.733}{0.668 \times 7.746} = -0.915$

(e) (2 points) Suppose we use simple linear regression to describe how the MPG changes as a linear function of the Weight. Calculate the least squares estimates of the intercept and slope.
$b_1 = r_{xy} \frac{s_y}{s_x} = -0.91 \times 7.75/0.67 = -10.61$, $b_0 = \bar{y} - b_1 \bar{x} = 20 - 3 \times (-10.61) = 51.83$

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

Intercept: MPG

Slope: MPG/1000 pounds

(g) (2 points) Suppose we change the units of Weight from 1000 pounds to 1000 kilograms (1000 pounds = 453.6 kilograms), what would be the intercept and slope?

Intercept: 51.83 MPG

Slope:

-10.61 MPG/1000 pounds $= -10.61 \times \frac{\text{1000 pounds}}{\text{453.6 kilograms}}$ MPG/1000 pounds $= -23.39$ MPG/1000 kilograms

(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 = 0.425 KPL, what would be the intercept and slope?

Intercept: 51.83$\times$ 0.425 KPL = 22.03 KPL

Slope:

-23.21×0.425 KPL/1000 kilograms $= -9.94$ KPL/1000 kilograms

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

$R^2 = (-0.915)^2 = 0.837$: the proportion of variation in Y explained by X

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

The MPG tends to decrease by -10.61 as the weight increases by 1000 pounds.

Problem 5 (25 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)-(d).

(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?

$$b_0 = -1.56$$

$$b_1 = 1.18$$

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

191.15-13.99=177.16

(c) (3 points) What's the value in cell B5?

$177.16/191.15 = 0.927$

(d) (3 points) We choose b_0 and b_1 to minimize which value in the table?

$C14$ or $B7$

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 (e)-(g).

(e) (5 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?

$$(-1.56 + 1.18 \times 10) \pm 2 \times 0.6$$

$$10.24 \pm 1.2$$

(f) (5 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 ?

Since given $X = 10$, we have $Y \sim \mathcal{N}(10.24, 0.6^2)$ and hence

$$P(Y > 16) = P(Z > \frac{16 - 10.24}{0.6}) = P(Z > 9.6) \approx 0$$

(g) (5 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?

Since given $X = 10$, we have $Y \sim \mathcal{N}(10.24, 0.6^2)$ and hence

$$P(10 + (14 - Y) - 0.76 > 13) = P(Y < 10.24) = 50\%$$

(Note that if $Y > 14$ when $X = 10$, there is no chance to have a net revenue of \$13/cwt or more.)

Problem 6 (15 points)

Suppose that 60% of the students at UT are football fans, 10% of the students at UT are from McCombs Business School, and 80% of the McCombs students are football fans.

(a) (3 points) If we choose a student at random from UT, what is the probability that this student is a football fan from McCombs Business School?

Let $F = 1$ if the student is a football fan and $F = 0$ otherwise, and $M = 1$ if the student is from McCombs and $M = 0$ otherwise. Since $P(F = 1) = 0.60$, $P(M = 1) = 0.10$ and $P(F = 1|M = 1) = 0.80$, we have

$$P(F = 1, M = 1) = P(M = 1)P(F = 1|M = 1) = 0.10 \times 0.80 = 8\%$$

(b) (3 points) If we choose a student at random from UT, what is the probability that this student is from McCombs Business School but *not* a football fan?

$$P(F = 0, M = 1) = P(M = 1) - P(F = 1, M = 1) = 0.10 - 0.08 = 2\%$$

(c) (3 points) If we choose a student at random from UT, what is the probability that this student is neither a football fan nor from McCombs Business School?

$$P(F = 0, M = 0) = P(F = 0) - P(F = 0, M = 1) = (1 - 0.6) - 0.02 = 38\%$$

(d) (3 points) If we choose a football fan student at random from UT, what is the probability that this football fan student is from McCombs Business School?

$$P(M = 1|F = 1) = \frac{P(F = 1, M = 1)}{P(F = 1)} = \frac{0.08}{0.60} = 13.3\%$$

(e) (3 points) If we choose a student who is *not* from McCombs Business School at random, what is the probability for this non-McCombs student to be a football fan?

$$P(F = 1|M = 0) = \frac{P(F = 1, M = 0)}{P(M = 0)} = \frac{P(F = 1) - P(F = 1, M = 1)}{1 - P(M = 1)} = \frac{0.6 - 0.08}{1 - 0.1} = 57.8\%$$

Comment: one easy way to do this kind of problem is using a table shown below.

	McCombs (M=1)	Not McCombs (M=0)	
Football Fan (F=1)	0.08	0.52	0.60
Not Football Fan (F=0)	0.02	0.38	0.40
	0.10	0.90	1