

**PRINT**

NAME _____

(Last name)

(First name)

KEY
(net ID, your UI email)**CIRCLE SECTION L1 (11am) or L2 (2 pm)**Write answers in appropriate blanks. When no blanks are provided **CIRCLE** your answers.**SHOW WORK** when requested, otherwise points will be deducted.

No notes or books are allowed. Calculators (including graphing ones) are allowed.

Do not use your own scrap paper. If you need some, ask a proctor.

Make sure you have all 5 pages (7 problems).**DO NOT WRITE BELOW THIS LINE** _____

The numbers written in each blank below indicate how many points you missed on each page. The numbers printed to the right of each blank indicate how many points each page is worth.

Page 1 _____ 15

Page 2 _____ 33

Page 3 _____ 28

Page 4 _____ 18

Page 5 _____ 14

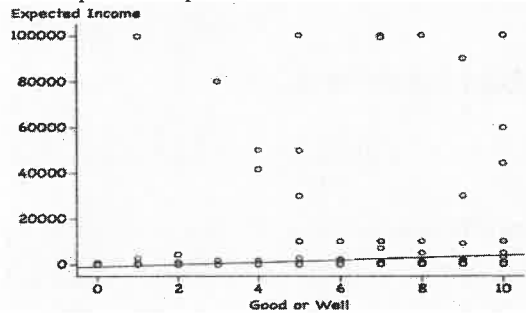
Score _____

Scores will be posted late tonight and exams returned in class on Thursday.

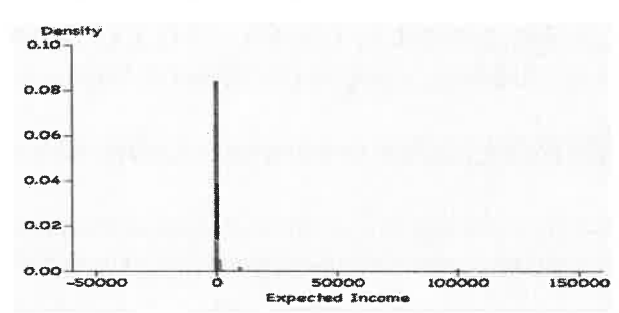
Question 1 (15 pts.) pertains to this semester's 1209 responses to these 2 survey questions:

1. (Good or Well) When you graduate are you more interested in "doing good" (helping others) or "doing well" (gaining material resources)? On a scale of 0-10 rate where you fall: "0" means "doing good" and "10" means "doing well"
2. (Expected Income) Twenty years down the road, how much money do you expect to make per year? (\$ in thousands)

Scatter plot of Expected Income (\$ in thousands) vs Good or Well Ratings



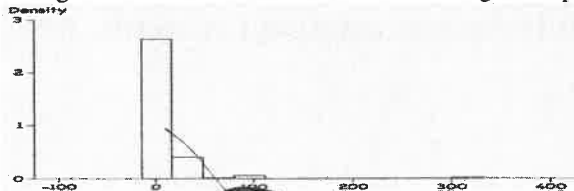
Histogram of Expected Income (\$ in thousands)



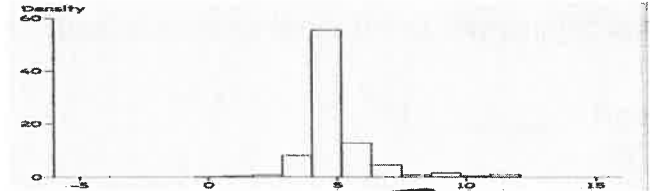
- a) (3 pts.) Look how the Y values are squished at the bottom of the plot and how skewed the histogram of Expected Income is. I want to transform Y to make its histogram more normal. Which transformations would definitely **not** work? Circle ALL that could NOT work.

1 pt i) Y^2 ii) Y^3 iii) e^Y iv) $\sqrt{Y}$ v) $\ln(Y)$ vi) $1/Y$

- b) (4 pts.) Below are histograms of the transformed Expected Income. Which of the 6 transformations do the histograms depict? Below each histogram circle the transformation the histogram represents.

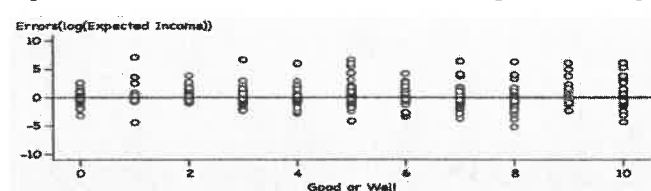
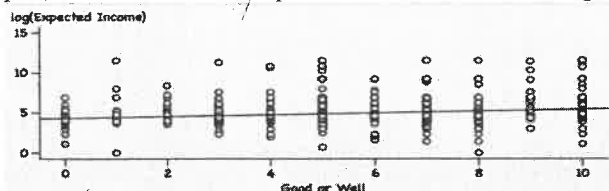


i) Y^2 ii) Y^3 iii) e^Y iv) $\sqrt{Y}$ v) $\ln(Y)$ vi) $1/Y$



i) Y^2 ii) Y^3 iii) e^Y iv) $\sqrt{Y}$ v) $\ln(Y)$ vi) $1/Y$

- c) (8 pts.) Below is the scatter plot of "Good or Well" ratings vs $\ln(\text{Expected Income})$ on the left and the residual plot on the right.



The regression equation is: **Predicted $\ln(\text{Expected Income}) = 4.3 + 0.11 * (\text{Good or Well rating})$** and $SD_{\text{errors}} = 1.3$

- i. (2 pts.) Use the above equation to estimate the $\ln(\text{Exp. Income})$ of a student with a "Good or Well" rating of = 5
 $4.3 + 0.11(5) =$ $\ln(\text{Expected Income}) = 4.85$ Don't round answer.

- ii. (2 pts.) Now estimate that students' expected-income (\$ in thousands). (Use the result from (i) to find ii.) **Show work.**

$e^{4.85}$ Expected Income = \$ 127.74 (in thousands). Round answer to 2 decimal places.

- iii. (2 pts.) Compute a 95% Confidence Interval for your estimate in part ii. (You may use $Z=2$ as the critical value for 95%) **Show work. Round to 2 decimal places.**

$e^{4.85 - 2.6}$ $e^{4.85 + 2.6}$

(\$ 9.49, \$ 1719.86) (\$ in thousands)

- iv. (2 pts.) A 95% Confidence Interval for the **expected Income** of another student is (\$5.47 to \$992.27) (\$ in thousands), what is his estimated Expected Income? **Show work. Round to 2 decimal places.**

$$\sqrt{5.47 \times 992.27} =$$

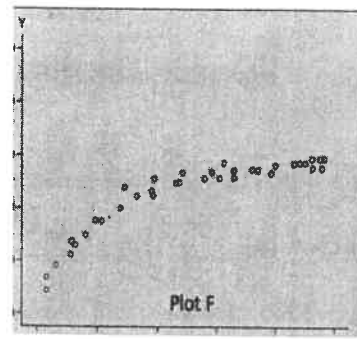
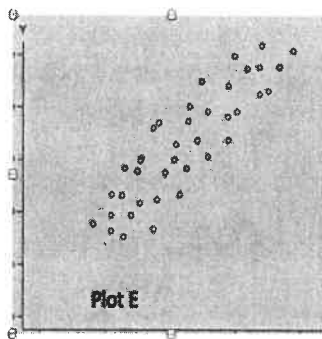
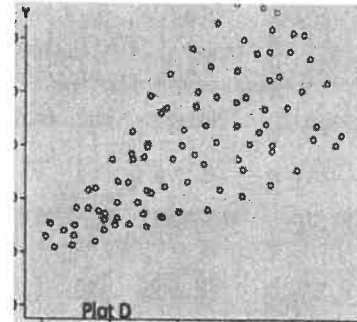
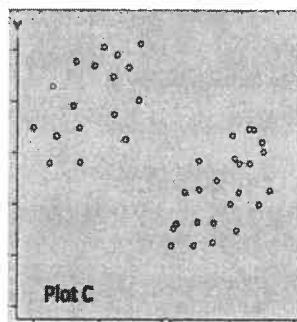
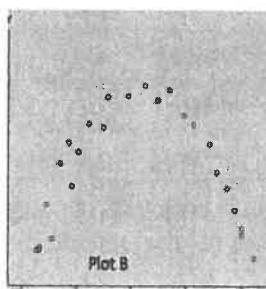
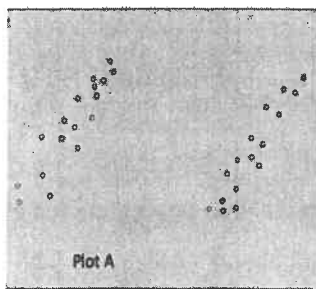
Estimated Expected Income = 73.67 (\$ in thousands)

or $\ln(5.47)$ $\ln(992.27)$
 1.7 4.3 6.9

(accept 73.6 - 73.7)

$e^{4.3}$

Question 2 (8 pts.) Each plot will be the answer to exactly one of the questions.



a) Which plot(s) are well suited to linear regression analysis as they are? A B C D E F

b) Which plot(s) show a curvilinear trend?

A B C D E F

c) Which plot(s) should be broken into subgroups since they show 2 distinct groups? A B C D E F

d) Which plot(s) are linear but would not be suited to using the SD_{errors} for building confidence intervals around our estimates? A B C D E F

Question 3 (10 pts.) For which of following would it be appropriate to use either simple or multiple logistic regression? Circle Yes if appropriate, No if not.

- 2 pt i) Predicting whether you get A, B, or C on a test from hours studying. a) Yes b) No
- 2 pt ii) Predicting whether you get an A on a test from hours of studying. a) Yes b) No
- 2 pt iii) Predicting surviving breast cancer 5 years after diagnosis based on type of treatment. a) Yes b) No
- 2 pt iv) Predicting years of survival after breast cancer diagnosis based on type of treatment. a) Yes b) No
- 2 pt v) Predicting $\ln(\text{Income})$ from years of education. a) Yes b) No

Question 4: (15 pts.) For the following problems p is defined as the probability of "success" and $1-p$ is the probability of "failure". Fill out the 15 missing values in the table below.

p	0.2	4/9	0.5	5/9	0.8
1-p	0.8	5/9	0.5	4/9	0.2
odds	0.25	4/5	1	5/4	4
$\ln(\text{odds})$	-1.39	-0.22	0	0.22	1.39

Express p, 1-p and odds as fractions for 2nd and 4th columns.

← Round $\ln(\text{odds})$ to 2 decimal places.

Question 5 (28 pts.)

On our last survey, 1427 students anonymously answered these 3 questions:

1. Would you date someone who is transgender? (No = 0, Yes = 1)
2. How many children would you ideally like to have?
3. Are you male or female? (Males=0, Females=1)

	Slopes	SE	Z	p-values
Intercept	-1.604	0.2169	-7.397	< .005%
Gender	1.002	0.1907	5.252	< .005%
children	-0.3823	0.06538	-5.848	< .005%

Suppose the 1427 students were randomly selected from all 40,000 UI undergrads. Above is the output of the logistic regression model used for predicting the probability of dating a transgender from gender and ideal number of children:

- a) (2 pts.) Are Males more or less likely than Females to date a transgender given the same # of children they want? **Choose one:**
 i) More ii) Less iii) Same iv) Not enough info

- b) (2 pts.) The more children desired the _____ the probability of dating a transgender given the same sex? **Choose one:**
 i) Higher ii) Lower iii) Same iv) Not enough info

- c) (2 pts.) Fill in the 3 missing blanks in the ln(odds) regression equation:

$$\text{Predicted } \ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = \underline{-1.604} + \underline{1.002} * \text{Gender} + \underline{-0.3823} * \text{Children}$$

- d) (4 pts.) What's the estimated odds ratio for Gender? 2.72 for Children? 0.68 Round to 2 decimal places. Show work.

$$e^{1.002} \quad e^{-0.3823}$$

- e) (12 pts.) Use the regression equation to predict the ln(odds), odds, and probability of dating a transgender for the individuals in the table below. Show work and round all answers to 3 decimal places. You may use rounded answers in your calculations.

Gender	Children	ln(odds)	Odds	p
Male	0	$-1.604 + 0 + 0 = -1.604$	$e^{-1.604} = 0.201$	$\frac{0.201}{1.0201} = 0.167$
Female	0	$-1.604 + 1.002 - 0 = -0.602$	$e^{-0.602} = 0.548$	$\frac{0.548}{1.548} = 0.354$
Male	3	$-1.604 + 0 + -0.3823(3) = -2.751$	$e^{-2.751} = 0.064$	$\frac{0.064}{1.064} = 0.060$
Female	3	$-1.604 + 1.002 - 0.3823(3) = -1.749$	$e^{-1.749} = 0.174$	$\frac{0.174}{1.174} = 0.148$

- f) (2 pts.) Kate wants 3 more children than Sarah, compare their **odds** of dating a transgender.

The odds of Kate dating a transgender is _____ times the **odds** of Sarah dating a transgender.

- i) 0.68×3 ii) 0.68^3 iii) 1.47×3 iv) 1.47^3 v) Not enough info

- g) (2 pts.) Kate wants 3 more children than Sarah, compare their **probability** of dating a transgender.

The probability of Kate dating a transgender is _____ times the **probability** of Sarah dating a transgender.

- i) $\frac{\text{answer to f above}}{1 + \text{answer to f above}}$ ii) $\frac{1}{0.6177}$ iii) $\frac{0.68}{1.68}$ iii) Not enough info

- h) (2 pts.) The odds of a female dating a transgender are the same as the odds of a male dating a transgender when the female wants more children than the male. Show work. Round answer to 2 decimal places. (The answer won't be a whole number.)

1 pt for correct answer

$$-1.604 + 1.002(1) + (-0.3823)G_F = -1.604 + 1.002 - 0.3823G_F$$

$$1.002 - 0.3823G_F = -0.3823G_M$$

$$1.002 = 0.3823G_F - 0.3823G_M$$

$$2.62 = G_F - G_M$$

Question 6 (18 pts.) A fourth question asked on the same survey was: Are you a US citizen? (No=0 and Yes=1)

Let's add Citizen as a predictor variable to the example in Question 5. Below is the output from the logistic regression model predicting the probability of dating a transgender from Gender (Female=1, Male=0), Citizen (No=0, Yes=1) and # Children wanted.

Y	R	n	# X's	Chi-square	df	p-value
Date Transgender?	0.2538	1427	3	76.59		< .005%

a) (2 pts.) A Chi Square test was done for the overall regression effect. If the sample size were smaller would it have been appropriate to do an F test instead? Circle one: i) Yes ii) No

b) (2 pts.) How many degrees of freedom for the Chi-square test? 3

	Slopes	SE	Z	p-values
Intercept	-2.136	0.2829	-7.550	< .005%
Gender	0.9997	0.1916	5.217	< .005%
citizen	0.7664	0.2414		
children	-0.4360	0.06759	-6.451	< .005%

c) (2 pts.) Calculate the Z stat to test $H_0: \text{Slope}_{\text{citizen}} = 0$ $H_A: \text{Slope}_{\text{citizen}} \neq 0$

$Z = \frac{0.7664}{0.2414} = 3.17$
 (Round to 2 decimals.) Is the p-value < 5%? Yes

d) (2pts.) If the sample size were smaller would it have been appropriate to do a t test for the slope instead of a z test? Circle one: i) Yes ii) No

e) (2 pts.) To test whether the individual slopes are significant the program does 2-sided Z tests:

H_0 : Slope = 0 and the alternative hypothesis is H_A : Slope $\neq 0$. How would the p-values change if we did one-sided Z tests?

i) They would double. ii) They'd be halved. iii) They'd stay the same

f) (2 pts.) Compute a 95% Confidence Interval for the Gender slope. (You may use $Z=2$ as the critical value for 95%.) Show work.

$$0.9997 \pm 2(0.1916) = 0.6165 \text{ to } 1.3829$$

(0.62 to 1.38) Put lower number first and round to 2 decimal places.

g) (2 pts.) Compute a 95% Confidence Interval for the Odds Ratio for Gender. Show work.

accept $e^{0.62}$ to $e^{1.38}$

$$1.86 \text{ to } 3.97$$

(1.85 to 3.99) Put lower number first and round to 2 decimal places.

h) (4 pts.) In general, if the 2 sided p-value for the slope of X in the $\ln(\text{odds})$ equation < 5% then ...

i. a 95% Confidence interval for the slope of X Choose one:

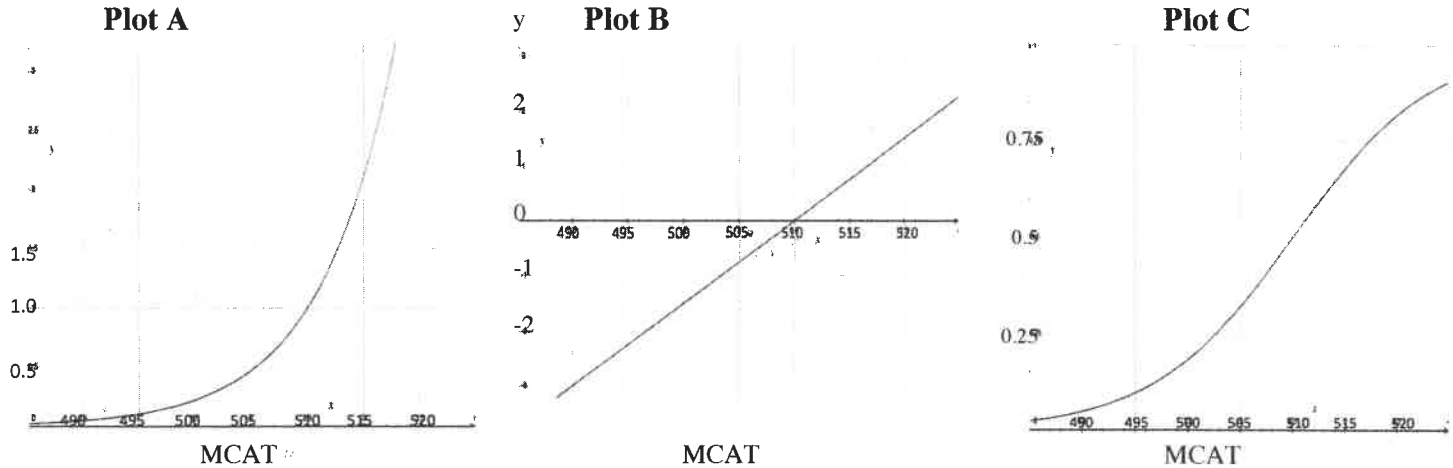
a. Includes 0 b. Doesn't include 0 c. Includes 1 d. Doesn't include 1 e. Includes 1.96 f. Doesn't include 1.96

ii. and a 95% Confidence Interval for the Odds Ratio of X Choose one:

a. Includes 0 b. Doesn't include 0 c. Includes 1 d. Doesn't include 1 e. Includes 1.96 f. Doesn't include 1.96

Question 7 (14 pts.)

Let's say we're predicting acceptance to medical school based on MCAT scores. All the graphs below depict MCAT on the X-axis. The Y-axis is either the probability, the odds or the $\ln(\text{odds})$ of acceptance to med school. Which is which?



- a) (6 pts.) Plot B depicts $\ln(\text{odds})$; Plot A depicts **odds**; and Plot C depicts **probabilities**.
Fill in the 3 blanks with A, B, or C.

- b) (4 pts.) The equation for Plot C is of the form $y = \frac{e^{b_0 + b_1 \text{MCAT}}}{1 + e^{b_0 + b_1 \text{MCAT}}}$ and the equation for Plot A is of the form $y = e^{b_0 + b_1 \text{MCAT}}$.
Fill in the 2 blanks with A, B, or C.

- c) (2 pts.) Judging from the plots, what MCAT score gives a 50-50 chance of getting into Med School? 510

- d) (2 pts.) The score that gives a 50-50 chance of getting acceptance has a y-value= 1 in Plot A and a y-value= 0 in Plot B.
Fill in the 2 blanks with the correct numbers.