STATS 200 EXAM 2 April 3, 2019

PRINT Name _____ _____ **Circle Section:**
Last (family) name First name L1 ONL S1 S2 S3 S4 S5

Net ID _____ Signature _____

Instructions- This is a closed book, closed notes exam. You have 1.5 hours to complete it.

- **At the end of this exam, you must return this exam with all pages, and you must return your scantron sheet. Please circle all of your answers on this exam and fill in all of your answers on the scantron.**
- **If you do not turn in a complete exam and scantron form, you will receive the grade AB (Absent) for this exam.**
- Use a **#2 pencil**. **Each question has only *one* answer**. If you bubble in more than one answer it will automatically be marked wrong. Erase mistakes completely.
- This exam is **either Form A, B, C.** You don't know which test form you have so you MUST turn in your scantron with the exam so the TAs can correctly mark the test form box on your scantron sheet after the exam.

READ → How to fill out the Scantron form

- Print and bubble in your **LAST NAME with NO SPACES or DASHES** starting in the left most column. Print your **FIRST INITIAL** in the right-most column.
- Print and bubble in your Student ID number (UIN) **NO SPACES or DASHES** in the Student Number box.
- Print and bubble in the date in the Date box.
- **Print and bubble in your NET ID with NO SPACES or DASHES in the NETWORK ID box.**
- **Print and bubble in the Section Box. See section codes**. →
- *Write Stat 200* on the COURSE line.
- *Write your instructor's name* on the INSTRUCTOR line.
- *Write your section on the SECTION line.*

Section Codes:
L1 (Fireman TR at 9:30) = 00001
ONL (Fireman) = 00002
S1 (Liu MWF at 9am) = 00003
S2 (Wang MWF at 11am)= 00004
S3 (Yubai at 1pm)= 00005
S4 (Yang TR at 9:30) = 00006
S5 (Chen TR at 1:30) = 00007

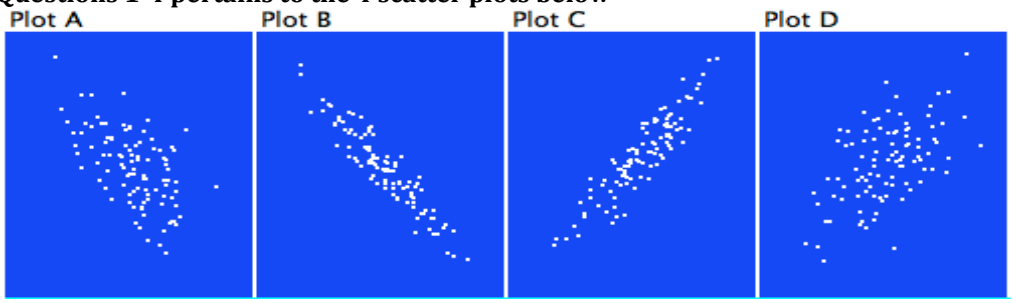
- Sign your name, and right underneath the student signature line PRINT your name

Warning -All Cheating including being caught with a non-permissible calculator or formula sheet will result in a 0 and an academic integrity violation on your University record.

CHECK NOW THAT YOU HAVE COMPLETED ALL OF THE STEPS. Before starting the exam, check to make sure that your test booklet is complete. You should have **9 pages** (**63 problems**), including **2 tables**: the normal table, and the chi-square table.

Formulas: $SD_{errors} = \sqrt{1-r^2} * SD_y$ $\quad SE_{slope} = \frac{SD_{errors}}{\sqrt{n} * SD_x} = \frac{\sqrt{1-r^2}}{\sqrt{n}} * \frac{SD_y}{SD_x}$

Questions 1-4 pertains to the 4 scatter plots below



For each plot write the correlation that is closest to it.

1)	Plot A	a) -0.94	b) -0.58	c) 0.54	d) 0.91
2)	Plot B	a) -0.94	b) -0.58	c) 0.54	d) 0.91
3)	Plot C	a) -0.94	b) -0.58	c) 0.54	d) 0.91
4)	Plot D	a) -0.94	b) -0.58	c) 0.54	d) 0.91

Questions 5-8

5) X and Y are two lists of numbers with correlation **r = 0.3.**
If all the Y values are multiplied by negative 2, the new correlation coefficient would be
a) 0.6 b) -0.6 c) -0.3 d) 0.3 e) Not enough information given

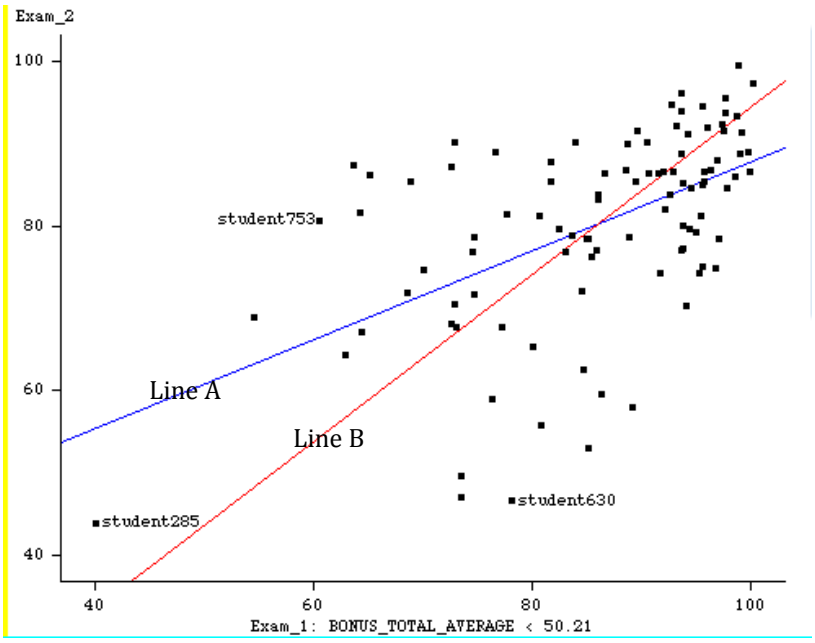
6) The regression line is the same as the SD line when ***Choose one:***
a) The correlation is 0
b) The correlation is perfect (1 or -1)
c) The average and SD of both variables are the same

7) The regression line is a horizontal line through the average of Y when… ***Choose one:***
a) The correlation is 0
b) The correlation is 1
c) The average and SD of both variables are the same

8) Outliers (extreme data points) can strongly influence the correlation coefficient. Which one of the statements below is true?

a) Removing outliers always causes the correlation to go down in absolute value.
b) Removing outliers always causes the correlation to go up in absolute value.
c) Removing outliers could cause the correlation to go up or down in absolute value depending on the scatter plot.

Questions 9-16 pertain to the exam 1 and exam 2 scores of 100 students in this class. (The ones who did less than about 50% of the bonus assignments.) The 5 summary statistics and the scatter plot are shown below:



	Average	*SD*
Exam 1	*86*	*12*
Exam 2	*80*	*12*

Correlation: r = 0.5

9) Two lines are shown. One is the regression line and one is the SD line. Which is the ***regression*** line?
a) Line A **b)** Line B

10) One student scored exactly average on both exams, which line does his data point lie on?
a) Regression Line only **b)** SD line only **c)** Both **d)** Neither

11) Which student did better on Exam 2 than predicted by the regression line?
a) Student 630 only
b) Student 753 only
c) Student 285 only
d) Both Student 285 and Student 753

12) If a student scored exactly 1 SD below average on Exam 1, what is the regression estimate for his Exam 2 score?
a) 0.5 SD's below average on Exam 2
b) 1 SD below average on Exam 2
c) Exactly Average on Exam 2

13) One student scored a 98 on Exam 1, what is the regression estimate for his Exam 2 score?
a) 98 b) 92 c) 83 d) 86 e) 80

14) What is the slope of the regression line for predicting Exam 2 scores from Exam 1 scores?

a) 0.5*86/80 **b)** 0.5*80/86 **c)** 0.5 **d)** 0.5*12/86 **e)** 0.5*12/80

15) What is the y-intercept for predicting Exam 2 scores from Exam 1 scores?
a) 0 **b)** 37 **c)** 55 **d)** 43 **e)** 46

16) The SD of the prediction errors (also known as the RMSE) when predicting Exam 2 scores from Exam 1 scores is
a) 12 **b)** 6 **c)** $\sqrt{1-0.5^2} * 86$ **d)** $\sqrt{1-0.5^2} * 80$ **e)** $\sqrt{1-0.5^2} * 12$

Questions 17-25 How do the number of hours students in a large Chemistry class studied for their Final correlate with their Final exam scores? To find out we randomly sampled 64 of the 3000 students enrolled in the class and got the 5 summary stats in the table below. (Assume the scatter plot follows a liner trend.)

	Avg	*SD*
Final Score	60	18
# Study Hours	12	3

r = 0. 4

17) Our best estimate of the **slope** of the regression equation for predicting Final Scores from study hours for all 3000 students is β = _____pts/study hour **a)** 0.067 **b)** 0.08 **c)** 0.4 **d)** 2 **e)** 2.4

18) SE_{slope} = _____ pts/study hr. is closest to … **a)** 0.573 **b)** 0.687 **c)** 0.092 **d)** 4.58 **e)** 5.5

19) 92% Confidence Interval for β= sample slope ± _____ * SE_{slope}
a) 0.92 **b)** 0.8 **c)** 1.45 **d)** 1.65 **e)** 1.75

To test H_0: β=0 against the H_A: $\beta \neq 0$, Z, t, χ^2 and F statistics were computed.

20) Z stat = **a)** $\frac{0.4}{\sqrt{1-0.16}} * \sqrt{64}$ **b)** $\frac{0.4}{\sqrt{1-0.16}} * \sqrt{62}$ **c)** $\frac{0.16}{1-0.16} * 62$ **d)** $\frac{0.16}{1-0.16} * 64$

21) t stat = **a)** $\frac{0.4}{\sqrt{1-0.16}} * \sqrt{64}$ **b)** $\frac{0.4}{\sqrt{1-0.16}} * \sqrt{62}$ **c)** $\frac{0.16}{1-0.16} * 62$ **d)** $\frac{0.16}{1-0.16} * 64$

22) χ^2 stat = **a)** $\frac{0.4}{\sqrt{1-0.16}} * \sqrt{64}$ **b)** $\frac{0.4}{\sqrt{1-0.16}} * \sqrt{62}$ **c)** $\frac{0.16}{1-0.16} * 62$ **d)** $\frac{0.16}{1-0.16} * 64$

23) F stat = **a)** $\frac{0.4}{\sqrt{1-0.16}} * \sqrt{64}$ **b)** $\frac{0.4}{\sqrt{1-0.16}} * \sqrt{62}$ **c)** $\frac{0.16}{1-0.16} * 62$ **d)** $\frac{0.16}{1-0.16} * 64$

Suppose the **t stat** testing $\mathbf{H_0}$: β=0 against $\mathbf{H_A}$: $\beta \neq 0$ yielded a p-value= 0.1%,

24) then the F stat would yield a p-value _____ 0.1% **a)** < **b)** > **c)** =

25) and the Z stat would yield a p-value _____0.1%. **a)** < **b)** > **c)** =

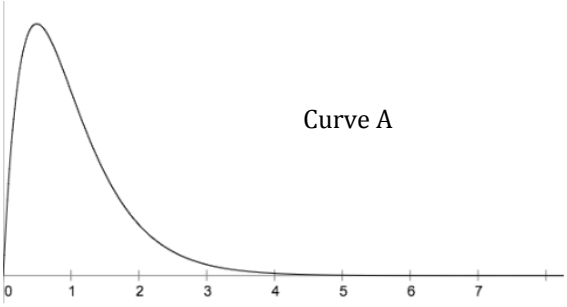
26) If we changed to a **1-sided $\mathbf{H_A}$**: $\beta > 0$, the **t stat** would give a p-value = **a)** 0.05% **b)** 0.1% **c)** 0. 2% **d)** 0.1% * $\sqrt{\frac{64}{62}}$

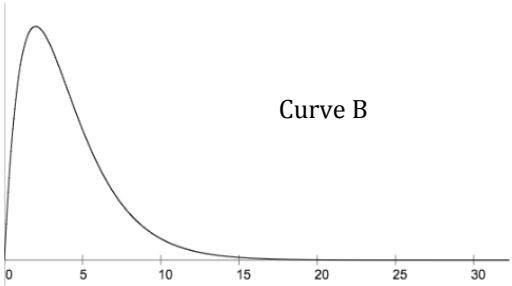
Questions 27 - 28

You compute F and χ^2 stats to test H_0: All pop slopes =0 in the model: $\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$ **n= 200**.

27) The χ^2 stat is about _____times the F stat. Choose closest answer: a) ¼ b) ½ c) 1 d) 2 e) 4

28) Here are the F and χ^2 curves. Which is which? The F curve is …..***choose one:*** a) A b) B.

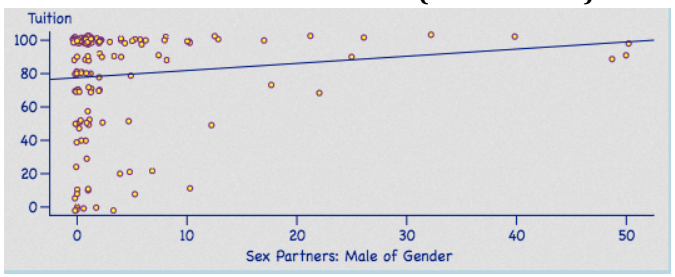




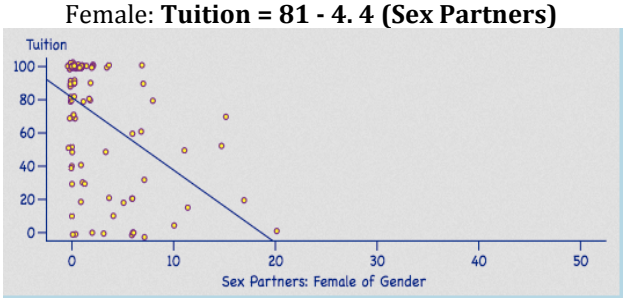
Questions 29-30

The scatter plots below show the survey responses of 100 females and 140 males in Stat 200 to these 2 questions: "How many people have you slept with in your life?" and "What percent of your tuition are your parents paying for?"

Male: **Tuition = 79 + 0. 4 (Sex Partners)**



Female: **Tuition = 81 - 4. 4 (Sex Partners)**



29) Translate the male and female regression equations above into an equivalent multiple regression.

Code Males=0 and Females=1 for the Gender variable.

a) Tuition = 79 + 0.4 (Sex Partners) -2 (Gender) - 4.4 (Sex Partners*Gender)
b) Tuition = 81 - 4.4 (Sex Partners) + 2 (Gender) + -4.8(Sex Partners*Gender)
c) Tuition = 79 + 0.4 (Sex Partners) + 2 (Gender) - 4.8(Sex Partners*Gender)
d) Tuition = 81 - 4.4 (Sex Partners) - 2 (Gender) + 4.4(Sex Partners*Gender)
e) None of the above since the 2 groups have different slopes

30) If you switched the code to Females=0 and Males=1, which one of the following would change?

a) Scatter Plots pictured above **b)** Multiple Regression Equation **c)** Simple Regression Equations

Question 31-32

Suppose A and B are 2 drugs designed to prolong life in patients with terminal cancer. The numbers in the tables indicate the mean survival time (in months) from those who received A alone, B alone, neither or both. Each table describes a different hypothetical study.

31) Which model matches the data in the table?

a) Months = 10A + 17B
b) Months = 4 + 10A + 6B + 17AB
c) Months= 4 + 6A +2B + 11AB
d) Months= 4 + 6A +2B + 5AB
e) None of the above

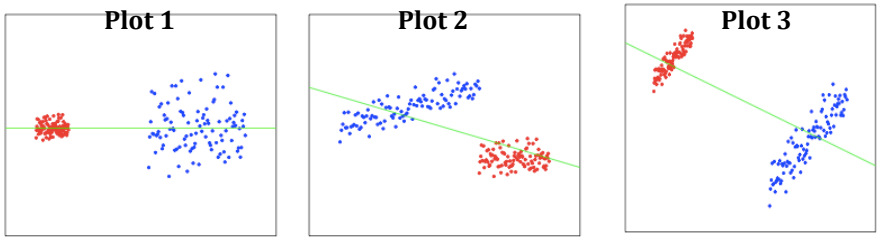
	A=0	A=1
B=0	4	10
B=1	6	17

32) Fill in the missing cell in the table to match a model with no interaction term.

a) 0 **b)** 3 **c)** 5 **d)** 8 **e)** 13

	A=0	A=1
B=0	2	10
B=1	5	

Question 33-34 pertains to the 3 scatter plots, each displaying 2 separate groups. As usual, X is the horizontal axis and Y is the vertical.



33) For which plot does there appear to be an interaction*:* **a)** Plot 1 **b)** Plot 2 **c)** Plot 3

34) Let's say we're trying to figure out if X causes Y, for which plot is the causal relation between X and Y NOT confounded by group membership? **a)** Plot 1 **b)** Plot 2 **c)** Plot 3

Questions 35-45 On Survey 1 **238** Stat 200 students reported their ACT scores, how many hours they study per week, and their GPA. Assume the students were chosen from a much larger population of all Stat 200 students. Here is the multiple regression equation predicting GPA from ACT and study hours. $\hat{G}PA=2.3 + 0.04(\text{Study Hours}) + 0.03(\text{ACT})$

35) The above equation describes the best fitting _____ **a)** line **b)** plane **c)** ellipsoid **d)** cube

36) through all the points so as to minimize the sum of squared errors in _____ **a)** Study Hours **b)** ACT **c)** GPA

37) Students A and B study the same number of hours per week but differ by 6 points in their ACT scores, the regression equation predicts they would differ by _____ pts in GPA.
a) 0.03 **b)** 0.04 **c)** 0.18 **d)** 0.24 **e)** 2.48

Look at the correlation matrix describing r between the 3 variables. **Fill in the 2 missing blanks in the top row.**

38) Blank (A) is **a)** 0 **b)** 0.17 **c)** 0.22 **d)** 0.25 **e)** 1

39) Blank (B) is **a)** 0 **b)** 0.17 **c)** 0.22 **d)** 0.25 **e)** 1

	Study Hrs	ACT	GPA
Study Hrs	__A__	__B__	0.22
ACT	0.17	_____	_____
GPA	_____	0.25	1.00

40) The slope for ACT in the multiple regression equation is 0.03. The slope for ACT in the simple regression equation (ACT predicing GPA) would be _____ 0.03. **a)** < **b)** > **c)** = **d)** not enough info.

.

41) The multiple correlation coefficient is **R= 0.31** How was that calculated?
a) R is the correlation between the 3 variables once each variable has been controlled for the other two.
b) R is the correlation between the students' predicted GPA's and their actual GPA's.
c) R is calculated by converting the 3 variables to Z scores, then taking the average of the product of the Z's.

42) To test H_o: Both β's = 0? (Both slopes in the population = 0) a χ^2 and F test stat were both computed: 12.49 and 25.3, but I forgot which is which? Which is the χ^2 stat? **a)** 12.49 **b)** 25.3 **c)** not enough info

43) Look at the χ^2 table. What is χ^2*(critical value) at α=0.001 (0.1%) ?
a) 5.99 **b)** 9.21 **c)** 11.34 **d)** 13.82 **e)** 16.27

44) The p-value is **a)** < 0.1% **b)** between 0.1% and 1% **c)** > 1%

45) Is the p-value using the F-test larger than, smaller than or the same as the p-value using the χ^2 test?
a) larger **b)** smaller **c)** same

Questions 46-51 continue the example on the previous page where we found $\hat{G}PA=2.3 + 0.04(\text{Study Hours}) + 0.03(\text{ACT})$ to be the best fit for our sample **n=238** with **R=0.31**.

46) Suppose you decided to reject the null, you'd conclude that
- **a)** Both slopes must be significant
- **b)** The ACT slope must be significant
- **c)** The Study Hours slope must be significant
- **d)** The intercept must be significant
- **e)** Either the ACT or the Study Hours slope or both must be significant

47) To see which slope is significant in the multiple regression equation: $\hat{G}PA=2.3 + 0.04(\text{Study Hours}) + 0.03(\text{ACT})$ the computer ran a Z test and a t-test. Which test would yield smaller p-value for the slopes?

a) t-test **b)** Z-test **c)** they'd be exactly the same **d)** Not enough information

48) How many degrees of freedom for the t-test? a) 2 b) 3 c) 235 d) 236 e) 237

49) The computer ran 2-sided significance tests for each slope and found they were both significant at $\alpha=0.01$. Which of the following gives equivalent information?
- **a)** 99% CI's do not include 0
- **b)** 99.5% CI's do not include 0
- **c)** 98% CI's do not include 0
- **d)** 99% CI's do include 0
- **e)** None of the above

50) Another variable is added to the model that is negatively correlated GPA. Will R^2 go up or down? Circle one:
a) Up b) Down c) Stay the same d) Either down or stay the same e) not enough info

51) Let's say a 3rd variable that's correlated with GPA is added to the multiple regression model and the slopes for Study Hours and ACT stay the same. You can conclude the 3rd variable must be …
- a) correlated with either study hours or ACT
- b) correlated with both study hours and ACT
- c) uncorrelated with both study hours and ACT.
- d) negatively correlated with study hours and positively correlated with ACT (or vice versa) so their effects cancel out

Questions 52-53

The numbers 1-15 are divided into 3 groups as shown below. **SST = 280.**

	Group 1	Group 2	Group 3	
	1	6	11	
	2	7	12	
	3	8	13	
	4	9	14	
	5	10	15	
	Mean=3	Mean=8	Mean=13	Overall Mean=8

52) SSB= a)30 b) 50 c)150 d) 250

53) SSW= a)30 b) 50 c)150 d) 250

Questions 54-60

The table displays the survey responses of **182** Stat 200 students to the question: "How many children would you ideally like to have?". The students also identified their ethnicity. Imagine the **182** students were randomly sampled from a much larger population of all Stat 200 students.

	Ethnicity	Average (rounded)	SD (rounded)	n
# of Children	White	3.0	1.6	55
# of Children	Asian	1.9	0.9	104
# of Children	Other	2.6	1.3	23

Questions **54-56** ask you to fill in 3 of the missing cells in the ANOVA table is below.

54) How many degrees of freedom for the model (SSB)? **a)** 2 **b)** 3 **c)** 179 **d)** 180 **e)** 18

55) The F-stat for testing the null hypothesis that all group means are the same in the population is:
a) 0.128 **b)** 10.215 **c)** 13.18 **d)** 6.81 **e)** Not enough info

56) R^2 is closest to ….**a)** 0.128 **b)** 0. 21 **c)** 10.215 **d)** 13.18 **e)** Not enough info

Source	SS (Sum of Squares)	df	Mean Square	F Statistic	P-value
Model	SSB=40.86	df=_____	MSB= 20.43	F=_____	< 0.005%
Error	SSW=_____	df=_____	MSW= 1.55	SD^+_{errors} = _____	
Total	SST=318.38	df= 181		R^2 = _____	

57) What do you conclude?
a) That all the group averages are significantly different from each other.
b) That at least one of the group averages is significantly different than the others.
c) That none of the group averages are significantly different from each other.

Compute the t-statistic to test whether the difference between Other and Asian is significant.

58) What is the $SE^+_{difference}$? Use $SD^+_{errors} = 1.245$. **a)** $1.245\sqrt{\frac{1}{104}+\frac{1}{23}}$ **b)** $1.245\sqrt{\frac{1}{1.3}+\frac{1}{0.9}}$ **c)** $\sqrt{1.245}\sqrt{\frac{1}{55}+\frac{1}{23}}$

59) What is the t-statistic? a) $\frac{0.4}{SE^+_{difference}}$ b) $\frac{0.4}{SE^+_{difference}} * \sqrt{179}$ c) $\frac{0.7}{SE^+_{difference}} * 182$ d) $\frac{0.7}{SE^+_{difference}}$

60) The p-value is 2.23%. The Bonferroni correction would _____ the p-value by ____
a) multiply, 2 b) multiply, 3 c) multiply, 6 d) divide, 2 e) divide, 3

Questions 61-63 pertain to correctly filling out your cover sheet and Scantron form.

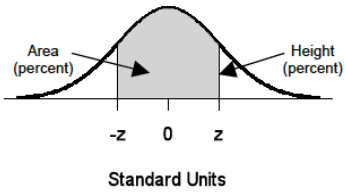
61) Cover Sheet -Did you circle your section on the cover sheet? a) Yes b) No

62) Scantron -Did you print and bubble in your NET ID with no spaces or dashes in the NETWORK ID box? a) Yes b)No

63) Scantron- Did you print and bubble in the *correct* section code (given on cover sheet)? a) Yes b) No

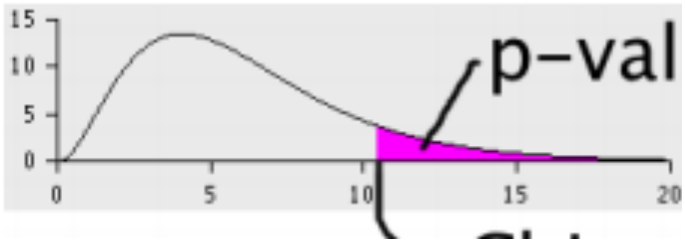
This is the end of the test. Carefully review all your answers and make sure you answered all the questions. Then bring both the text booklet and your Scantron to the proctors.

STANDARD NORMAL TABLE



z	*Area*		*z*	*Area*		*z*	*Area*
0.00	0.00		1.50	86.64		3.00	99.730
0.05	3.99		1.55	87.89		3.05	99.771
0.10	7.97		1.60	89.04		3.10	99.806
0.15	11.92		1.65	90.11		3.15	99.837
0.20	15.85		1.70	91.09		3.20	99.863
0.25	19.74		1.75	91.99		3.25	99.885
0.30	23.58		1.80	92.81		3.30	99.903
0.35	27.37		1.85	93.57		3.35	99.919
0.40	31.08		1.90	94.26		3.40	99.933
0.45	34.73		1.95	94.88		3.45	99.944
0.50	38.29		2.00	95.45		3.50	99.953
0.55	41.77		2.05	95.96		3.55	99.961
0.60	45.15		2.10	96.43		3.60	99.968
0.65	48.43		2.15	96.84		3.65	99.974
0.70	51.61		2.20	97.22		3.70	99.978
0.75	54.67		2.25	97.56		3.75	99.982
0.80	57.63		2.30	97.86		3.80	99.986
0.85	60.47		2.35	98.12		3.85	99.988
0.90	63.19		2.40	98.36		3.90	99.990
0.95	65.79		2.45	98.57		3.95	99.992
1.00	68.27		2.50	98.76		4.00	99.9937
1.05	70.63		2.55	98.92		4.05	99.9949
1.10	72.87		2.60	99.07		4.10	99.9959
1.15	74.99		2.65	99.20		4.15	99.9967
1.20	76.99		2.70	99.31		4.20	99.9973
1.25	78.87		2.75	99.40		4.25	99.9979
1.30	80.64		2.80	99.49		4.30	99.9983
1.35	82.30		2.85	99.56		4.35	99.9986
1.40	83.85		2.90	99.63		4.40	99.9989
1.45	85.29		2.95	99.68		4.45	99.9991

Chi-Square Table



Degrees of freedom ↓	30%	10%	5%	1%	0.1%	← p-value
1	1.07	2.71	3.84	6.63	10.83	
2	2.41	4.61	5.99	9.21	13.82	
3	3.66	6.25	7.81	11.34	16.27	
4	4.88	7.78	9.49	13.28	18.47	
5	6.06	9.24	11.07	15.09	20.52	
6	7.23	10.64	12.59	16.81	22.46	
7	8.38	12.02	14.07	18.48	24.32	
8	9.52	13.36	15.51	20.09	26.12	
9	10.66	14.68	16.92	21.67	27.88	
10	11.78	15.99	18.31	23.21	29.59	
11	12.90	17.28	19.68	24.72	31.26	←Chi-square
12	14.01	18.55	21.03	26.22	32.91	
13	15.12	19.81	22.36	27.69	34.53	
14	16.22	21.06	23.68	29.14	36.12	
15	17.32	22.31	25.00	30.58	37.70	
16	18.42	23.54	26.30	32.00	39.25	
17	19.51	24.77	27.59	33.41	40.79	
18	20.60	25.99	28.87	34.81	42.31	
19	21.69	27.20	30.14	36.19	43.82	
20	22.77	28.41	31.41	37.57	45.31	
21	23.86	29.62	32.67	38.93	46.80	
22	24.94	30.81	33.92	40.29	48.27	
23	26.02	32.01	35.17	41.64	49.73	
24	27.10	33.20	36.42	42.98	51.18	