STATISTICS 200 EXAM 2

PRINT NAME______________________________ ______________________ net ID___________
(Last name) (First name)

Circle Section: L1: 11:00 am or L2: 2:00 pm

Write answers in appropriate blanks. When no blanks are provided **CIRCLE** your answers.
SHOW WORK when requested.

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 10 pages (10 problems) including the Z, t, F and Chi-Sq tables.

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 _____ 19

Page 2 ______16

Page 3______ 21

Page 4 ______17

Page 5 ______ 21

Page 6 ______ 6

Score _________

Scores will be posted on Compass Friday morning and returned after spring break.

Stat 200 Exam 2 March 16, 2017

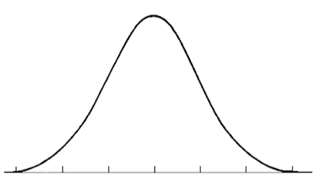
Question 1 *(15 points total)* We're trying to fit a simple linear regression model for the whole population: $Y=\beta_0 + \beta_1 X + \varepsilon$. (Assume ε are independent and normally distributed with constant variance). We draw a random sample of **n= 47** from the population and get a sample correlation **r = 0.2**. Compute the 4 test statistics for testing the **null H_0: $\beta_1=0$**. (same as testing **H_0: $r_{population}=0$.**) ***(Round your final answers to 3 decimal places, but don't round during intermediate steps.)***

a) *(1 pt.)* **R^2 = __________** **1-R^2 = ____**

b) *(12 pts.)* Now compute the 4 statistics below.

	Z	χ^2	t	F
Compute the values of the 4 test statistics. *Show work below your answers.*	Z= ______ *(1 pt.)*	χ^2 = ______ *(1 pt.)*	t = _____ *(1 pt.)*	F=_____ *(1 pt.)*

c) Compute the p-values for each statistic. Assume the alternative for the Z and t test is 1-sided: **H_A: $\beta_1 > 0$**, and assume the alternative for the χ^2 and F is 2-sided: **H_A: $\beta_1 \neq 0$**.

Z	χ^2	t	F
*(1 pt.)*p-value= ______% *(1 pt.)*Label Z on the normal curve below and shade the area representing the p-value. 	p-value=______% *(1 pt.)* How many degrees of freedom? _____ *(1 pt.)*	Choose closest one: i) 4.5% ii) 8% ii) 9% *(1 pt.)* How many degrees of freedom? _____ *(1 pt.)*	Choose closest one: i) 4% ii) 9% iii) 18% *(1 pt.)* How many degrees of freedom in numerator?___ in denominator?___ *(1 pt.)*

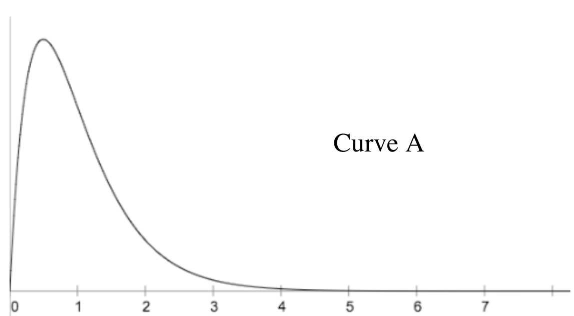
d) *(4 pts.)* Which of the following 4 quantities do you have enough information to calculate? If you have enough info to calculate write the correct number in the blank, if you don't have enough info leave it blank. (Round to 4 decimal places.)

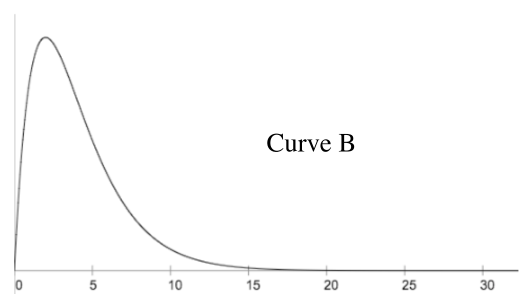
i) SST =_____ ii) SSM=_____ iii) SSM/SST =______ iv)SSM/SSE=_______

Question 2 You compute an F and χ^2 stat to test the null that all β's= 0 in the model : $\hat{\mathbf{y}} =\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$, n= 200.

i) (1 pt.) The χ^2 stat is about _________times the F stat. Choose closest answer: i) 1/4 ii) 1/2 iii) 1 iv) 2 v) 4

ii) (1 pt.) Here are the the F and χ^2 curves. Which is which? The F curve is …..***choose one:*** a) A b) B.





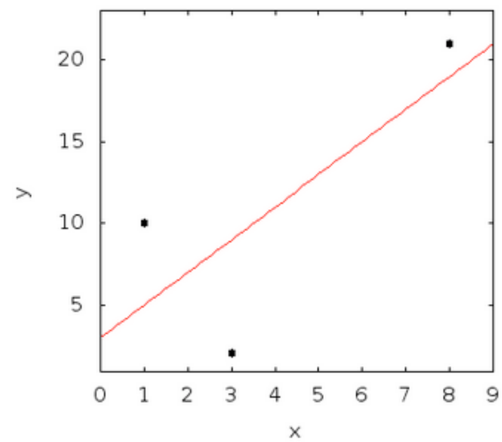
Question 3

(4 pts.) Suppose a random sample yields these 3 points: **(1, 10), (3, 2), (8, 21).**
$\bar{y} = 11$ **and SST= 182** .
Compute SSM and SSE. *Show work below each answer*.

Sample Scatter plot and regression equation:
$\hat{y} = 3 + 2x$

i) SSM= _____

ii) SSE= ____



Question 4 *(6 pts.)*
Suppose A and B are 2 drugs designed to prolong life in patients with terminal cancer. The numbers in the table indicate the mean survival time (in months) for those who received A alone, B alone, neither or both. Each table describes a different hypothetical study.
Fill in the 8 blanks in the tables so they match their equations and fill in the 4 blanks in the equation so it matches its table.

Months= 10A + 6B

	A=0	A=1
B=0		
B=1		

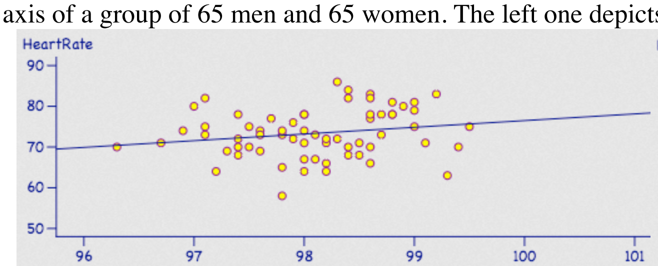
Months= 10A + 6B + 5AB

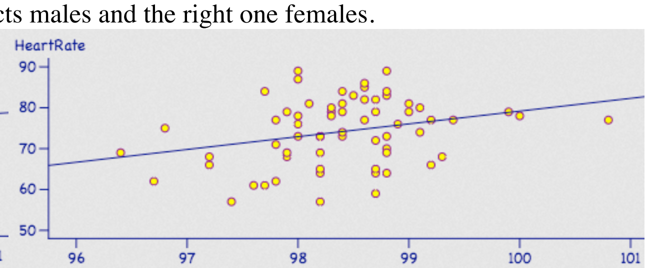
	A=0	A=1
B=0		
B=1		

Months = _____+_____A+ _____B +______AB

	A=0	A=1
B=0	4	14
B=1	10	17

Question 5 *(6 pts.)*
The scatter plots below show the body temperature (in Fahrenheit) on the X axis and the Heart Rate (beats per minute) on the Y axis of a group of 65 men and 65 women. The left one depicts males and the right one females.





Males: **Heart Rate = -88 + 1.6*Temperature**   Females: **Heart Rate = -234 + 3.1*Temperature**

a) *(4 pts.)* Translate the 2 separate male and female regression equations into the multiple regression equation below by filling in the correct numbers in the 4 blanks. Assume Gender is a 0-1 variable coded with Males=0 and Females=1.

Heart Rate = ________ + ___________ *Temperature + ________*Gender + ________Gender*Temperature

b) *(2 pt.)* If you switched the code to Males=1 and Females=0, which of the following would change?
Circle all that would change:
i) Scatter Plots pictured above **ii)** Multiple Regression Equation **iii)** Simple Regression Equations

Question 6 (21 pts)

To find out how education affects household income in a Midwestern city, a researcher collected data from **123** randomly selected Husband-Wife households in the city on the following 3 variables: Years of Education of Wife (E_w), Years of Education of Husband (E_h), Total household Income (I). Here's the multiple regression equation for predicting Household Income from husbands' and wives' education years. $\mathbf{\hat{I} = -\$9{,}580 + \$2{,}500E_w + \$4{,}000E_h}$

a) *(2 pts.)* The above equation describes the best fitting ______________ through all the points so as to minimize the sum of squared errors in _______.

Circle one for the 1st blank: **a)** line **b)** plane **c)** ellipsoid **d)** cube

Circle one for the 2nd blank: **a)** Income **b)** E_w **c)** E_h

b) *(2 pts.)* What does the \$4000 slope mean in the multiple regression equation above?

i) For those husbands with wives at the **same** educational level, each extra year of husband's education increases household income \$4000 on the average.

ii) For all husbands, regardless of how educated their wives are, each extra year of husband's education increases household income \$4000 on the average.

c) *(2 pts.)* Calculate the predicted Household Income for a married couple who each have only 10 years of education?

\$_________________

d) *(2 pt.)* Consider two couples: Couple A (Adam and Alice) and Couple B (Billy and Baby). Adam has 3 years more education than Billy has, and Alice has 8 years more education than Baby has. How much more household income would the regression equation predict for Couple A than for Couple B?

\$______________ more.

e) (4 pts) Look at the correlation matrix. Fill in the 4 missing blanks. Do you think the slopes predicting household income in the 2 simple regression equations: $I = b_0 + b_1E_w$ and $I = b_0 + b_1E_h$ would be the same as they are in the multiple regression equation?

i) Yes, they would still be \$2500 and \$4000.

ii) No, they'll both be larger in the simple regressions because they're all positively correlated

iii) No, they'll both be smaller in the simple regressions because they're fewer variables.

iv) It's impossible to know because they're all correlated with each other.

Fill in 4 blanks. (2 pts.)

	E_w	E_h	I
E_w	1.00	0.53	0.34
E_h	_______	1.00	_______
I	________	0.41	_______

f) (2 pts.) The multiple correlation is **R= 0.4** How was that calculated?

i) R is the correlation between the 3 variables once each variable has been controlled for the other two.

ii) Each of the **123** husband-wife pairs has a predicted income from the regression equation and an actual income. R is the correlation between those 2 sets of numbers.

iii) R is calculated by converting the 3 variables to Z scores and then taking the average of the product of their Z scores.

g) Do the χ^2 and F test for the overall regression effect for the model: $\mathbf{\hat{I} = -\$9{,}580 + \$2{,}500E_w + \$4{,}000E_h}$

F-test (4 pts)	χ^2 (3 pts.)
(1 pt.) Compute the F statistic. Show work. Circle answer	(1 pt.) Compute the χ^2 statistic. Show work. Cirlce answer.
(1 pt.) Look at the F table. What is F*(the critical value) at α=0.01 (1%)? F* = ______	(1 pt) Look at the χ^2 table. What is χ^{2*}(critical value) at at α=0.001 (0.1%) ? χ^{2*} = ______
(2 pts.) Our F test stat ___F* so our p-value is ______ than 1%. Fill in the two blanks with either <, >, or =	(1 pt) p-value is ….. a) < 0.1% b) between 0.1% and 1% c) > 1%

Question 6 cont. (9 pts.)

h) (2 pts.) Suppose you decided to reject the null, you'd conclude that …. ***Choose one:***
- **i)** Both slopes must be significant
- **ii)** The E_w slope must be significant
- **iii)** The E_h slope must be significant
- **iv)** The intercept must be significant
- **v)** Either the E_w or the E_h slope or both must be significant

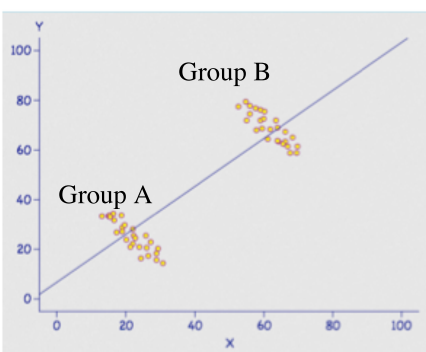
i) (2 pts.) I did a t-test and a Z-test to see if the E_w slope was significant and got a p-value just under 5% by one test and just over 5% by the other test, which p-value belongs with which test?
- **i)** The t-test must have given the bigger p-value since the t-curve has fatter tails.
- **ii)** The z-test must have given the bigger p-value since the Z statistic is bigger.
- **iii)** If done correctly the tests should have given exactly the same p-value.

j) (1 pt.)How many degrees of freedom for the t-test above? __________

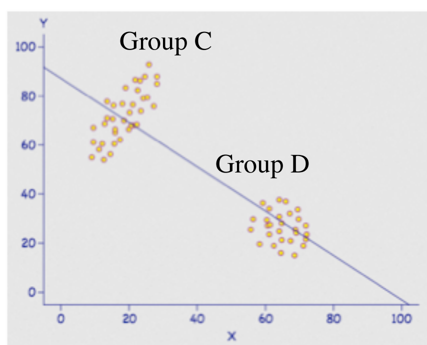
k) (2 pts.) If I delete E_w from the model will R^2 go up or down?
- **i)** It has to go down since E_w is correlated with Income.
- **ii)** It has to go up or stay the same.
- **iii)** It could go up, down or stay the same depending on whether it is significant.

l) (2 pt.)I decide to add a 3rd variable, either $\mathbf{X_{3a}}$ or $\mathbf{X_{3b}}$ to the full model since both look like good predictors of income on their own. I check the correlation matrix and see that X_{3a} has almost no correlation with either X variable already in the model, while $\mathbf{X_{3b}}$ has a correlation of 0.95 with husband's education. Which variable should I add to the full model?
- **i.** It's a toss up**--** the higher the correlation the better the fit will be so $\mathbf{X_{3b}}$ is a good candidate, but X_{3a} adds a completely new element to the mix.
- **ii.** $\mathbf{X_{3b}}$**,** there's no point in adding something that does not fit well with the other X's. The X's need to work together. No correlation is equivalent to no communication. Predictive power is lost.
- **iii.** Choose $\mathbf{X_{3a}}$, putting 2 variables that are highly correlated in the same model causes problems.

Question 7 pertains to the 2 scatter plots below *(8 pts.)*



Plot 1



Plot 2

a) Which group has a positive relation between X and Y? ***Circle one:***
i) A **ii)** B **iii)** C **iv)** D

b) Which scatter plot has an overall positive relation between X and Y? ***Circle one:***
i) Plot 1 only **ii)** Plot 2 only **iii)** Both **iv)** Neither

c) For which plot does there appear to be an interaction? ***Circle one:***
i) Plot 1 only **ii)** Plot 2 only **iii)** Both **iv)** Neither

d) Let's say we're trying to figure out if X causes Y, for which plot is the causal relation between X and Y confounded by group membership? ***Circle one:***
i) Plot 1 only **ii)** Plot 2 only **iii)** Both **iv)** Neither

Question 8 *(12 pts.)* On Survey 2 this semester **183** Stat 200 students responded to the survey question: "How many hours per week do you talk with either of your parents on the phone?" They also identified their ethnicity. Imagine the students were randomly sampled from a much larger population of all Stat 100 students. Here are the results:

	Ethnicity	Average hrs.(rounded)	n
Parent Phone time	White	5.5	55
Parent Phone time	East Asian	2.5	91
Parent Phone time	South Asia	4.2	17
Parent Phone time	Other	3.8	20

SD^+_{errors} = 4.85
(Use **SD^+_{errors}** as a pooled estimate for the SD of each group in the population)

a) *(9 pt.)* To test the null hypothesis that all 4 group means are the same in the population (same as testing that R_{pop}=0), compute the F-ratio by filling in the 9 missing numbers in the ANOVA table below.

Source	SS (rounded to nearest whole number)	df	MS (Round to 3 decimal places)	(Round to 3 decimal places)
Model	SSB=326	_______	________	F=________ The p-value is … **i)** < 1% **ii)** 1% to 5% **iii)** > 5%
Error	SSW=______	________	________	SD^+_{errors}= ________
Total	SST=4535	_______	Don't write anything here.	R^2 = _________

b) (1 pt.) Assuming you reject the null, what do you conclude? ***Choose one.***
- **i)** That all the group averages are significantly different from each other in the population.
- **ii)** That *at least one* of the group averages is significantly different from the others in the population.
- **iii)** That none of the group averages are significantly different from each other in the population.

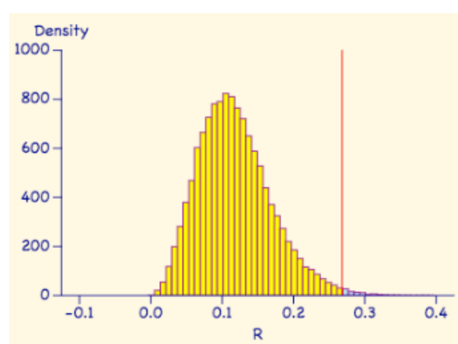
f) Compute the t-statistic to test whether the difference in group averages between Whites and East Asians is statistically significant at α=0.5.

i) *(2 pts.)*What is the $SE^+_{difference}$ ***Show work. Circle answer.*** *Round to 2 decimal places*.

ii) *(2 pt.)* What is the t-statistic? ***Show work. Circle answer.*** *Round to 2 decimal places*. How many df? ______(1 pt)

g)*(2 pt.)* The p-value is about 0.02%. The Bonferroni correction would _______________that p-value by_______. **Fill in the first blank with either "multiply" or "divide" and the second blank with a number.**

h) (2 pts) Look back at part (a). I tested the same null using the re-randomization test. The histogram to the right shows the distribution of R's resulting from 60,000 randomizations. The vertical line falls between 0.2 and 0.3, give its exact location to 4 decimal places. *Show work. Circle your answer.*



i) True or False? The area to the right of the vertical line represents …
a) (1 pt)The percentage of randomized R's as big or bigger than the R we obtained from our survey sample. True False
b) (1 pt)The p-value given by the randomization test. True False
c) (1 pt) The p-value given by the F-test. True False

Question 9
9 numbers are divided into 3 groups as shown below.

Group 1	Group 2	Group 3	
0	4	5	
2	6	7	
4	8	9	
Mean = 2	Mean = 6	Mean = 7	Overall Mean = 5

SST= 66
a) Compute SSB (same as SSM)

b) Compute SSW (same as SSE)

Question 10 *(2 pts.)*
If the χ^2 test doesn't yield significant results, is it possible the F test still would? (Assume you're testing the same null hypothesis using the same sample data)

- **i)** Yes, since the F test yields slightly more precise tests.
- **ii)** Yes, if the sample size is relatively small, the F test results could yield significantly different results.
- **iii)** No, the p-value for the F test will always be greater so it could never yield more significant results.
- **iv)** It's impossible to know since F is centered at 1 when the null is true and the χ^2 is centered at its degrees of freedom making comparisons of results statistically meaningless.

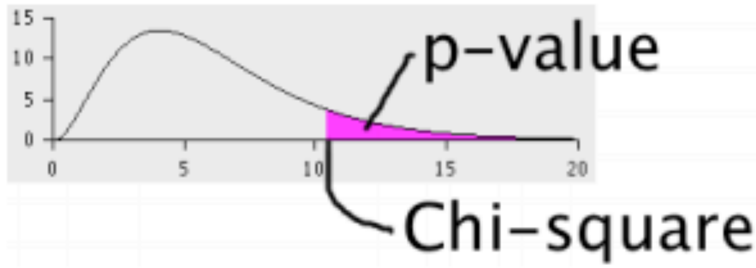
F Distribution critical values for P=0.05

Denominator DF	Numerator DF 1	2	3	4	5	7	10
1	161.45	199.50	215.71	224.58	230.16	236.77	241.88
2	18.513	19.000	19.164	19.247	19.296	19.353	19.396
3	10.128	9.5522	9.2766	9.1172	9.0135	8.8867	8.7855
4	7.7086	6.9443	6.5915	6.3882	6.2560	6.0942	5.9644
5	6.6078	5.7862	5.4095	5.1922	5.0504	4.8759	4.7351
7	5.5914	4.7375	4.3469	4.1202	3.9715	3.7871	3.6366
10	4.9645	4.1028	3.7082	3.4780	3.3259	3.1354	2.9782
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7066	2.5437
20	4.3512	3.4928	3.0983	2.8660	2.7109	2.5140	2.3479
30	4.1709	3.3159	2.9223	2.6896	2.5336	2.3343	2.1646
60	4.0012	3.1505	2.7581	2.5252	2.3683	2.1666	1.9927
120	3.9201	3.0718	2.6802	2.4473	2.2898	2.0868	1.9104
500	3.8601	3.0137	2.6227	2.3898	2.2320	2.0278	1.8496
1000	3.8508	3.0047	2.6137	2.3808	2.2230	2.0187	1.8402

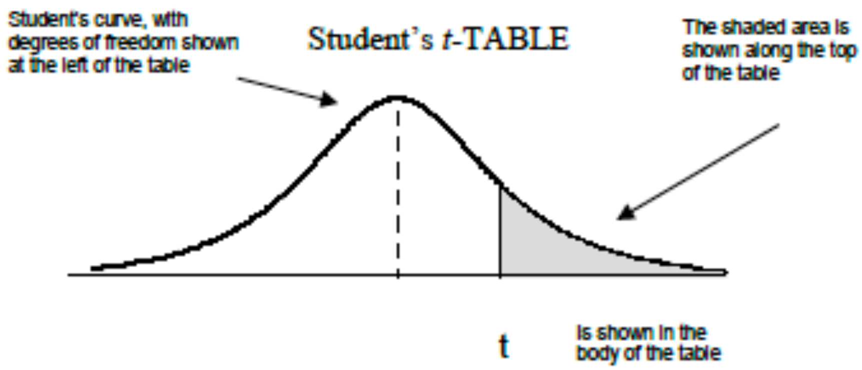
F Distribution critical values for P=0.01

Denominator DF	Numerator DF 1	2	3	4	5	7	10
1	4052.2	4999.5	5403.4	5624.6	5763.6	5928.4	6055.8
2	98.503	99.000	99.166	99.249	99.299	99.356	99.399
3	34.116	30.817	29.457	28.710	28.237	27.672	27.229
4	21.198	18.000	16.694	15.977	15.522	14.976	14.546
5	16.258	13.274	12.060	11.392	10.967	10.455	10.051
7	12.246	9.5467	8.4513	7.8466	7.4605	6.9929	6.6201
10	10.044	7.5594	6.5523	5.9944	5.6363	5.2001	4.8492
15	8.6831	6.3588	5.4169	4.8932	4.5557	4.1416	3.8049
20	8.0960	5.8489	4.9382	4.4306	4.1027	3.6987	3.3682
30	7.5624	5.3903	4.5098	4.0179	3.6990	3.3046	2.9791
60	7.0771	4.9774	4.1259	3.6491	3.3388	2.9530	2.6318
120	6.8509	4.7865	3.9490	3.4795	3.1736	2.7918	2.4720
500	6.6858	4.6479	3.8210	3.3569	3.0539	2.6751	2.3564
1000	6.6603	4.6264	3.8012	3.3379	3.0356	2.6571	2.3387

Chi-square table

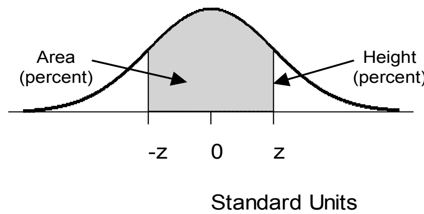


Degrees of freedom ↓	30%	10%	5%	1%	0.1%	← p-value
1	1.07	2.71	3.84	6.63	10.83	
2	2.41	4.61	5.99	9.21	13.82	
3	3.66	6.25	7.81	11.34	16.27	
4	4.88	7.78	9.49	13.28	18.47	
5	6.06	9.24	11.07	15.09	20.52	
6	7.23	10.64	12.59	16.81	22.46	
7	8.38	12.02	14.07	18.48	24.32	
8	9.52	13.36	15.51	20.09	26.12	
9	10.66	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	
12	14.01	18.55	21.03	26.22	32.91	← Chi-square
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	



Degrees of freedom	*25%*	*10%*	*5%*	*2.5%*	*1%*	*0.5%*
1	1.00	3.08	6.31	12.71	31.82	63.66
2	0.82	1.89	2.92	4.30	6.96	9.92
3	0.76	1.64	2.35	3.18	4.54	5.84
4	0.74	1.53	2.13	2.78	3.75	4.60
5	0.73	1.48	2.02	2.57	3.36	4.03
6	0.72	1.44	1.94	2.45	3.14	3.71
7	0.71	1.41	1.89	2.36	3.00	3.50
8	0.71	1.40	1.86	2.31	2.90	3.36
9	0.70	1.38	1.83	2.26	2.82	3.25
10	0.70	1.37	1.81	2.23	2.76	3.17
11	0.70	1.36	1.80	2.20	2.72	3.11
12	0.70	1.36	1.78	2.18	2.68	3.05
13	0.69	1.35	1.77	2.16	2.65	3.01
14	0.69	1.35	1.76	2.14	2.62	2.98
15	0.69	1.34	1.75	2.13	2.60	2.95
16	0.69	1.34	1.75	2.12	2.58	2.92
17	0.69	1.33	1.74	2.11	2.57	2.90
18	0.69	1.33	1.73	2.10	2.55	2.88
19	0.69	1.33	1.73	2.09	2.54	2.86
20	0.69	1.33	1.72	2.09	2.53	2.85
21	0.69	1.32	1.72	2.08	2.52	2.83
22	0.69	1.32	1.72	2.07	2.51	2.82
23	0.69	1.32	1.71	2.07	2.50	2.81
24	0.68	1.32	1.71	2.06	2.49	2.80
25	0.68	1.32	1.71	2.06	2.49	2.79

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