Stat 200 Exam 2 October 30, 2017

STATISTICS 200 EXAM 2

PRINT NAME_____ _____ _____

(Last name) (First name) Net ID (email, not UIN)

Circle Section: L1 Online

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

No notes or books are allowed. Calculators (except for ones connected to the internet) are allowed.
Do not use your own scrap paper. If you need some, ask me.

Make sure you have all 8 pages (10 problems) including the Normal and F 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_____15

Page 2 _____14

Page 3_____ 8

Page 4 _____28

Page 5 & 6_____ 34

Score _____

Scores will be posted on Compass Wed night and exams will be returned in class on Thursday. Online student can pick up their exams during office hours between 4-6 pm in 23 IH.

Stat 200 Exam 2 October 30, 2017

Question 1 *(16 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=20** from the population and get a sample correlation **r = 0.4**. Compute the 4 test statistics for testing the **null H_0: r=0** in the population. (same as testing **H_0: β_1=0.**)

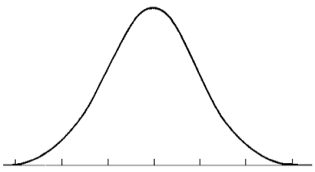
(Round your final answers to 3 decimal places, but don't round during intermediate steps.)

a) *(1 pt.)* First compute**: R^2 =** _____ (Don't round.) **1-R^2 =**_____ **(**Don't round).

b) Now compute the values of 4 test- statistics below. *Show work below your answers* (4 pts.)

Z	χ^2	t	F
Z= _____*(1 pt.)*	χ^2 = _____*(1 pt.)*	t = _____*(1 pt.)*	F=_____*(1 pt.)*

c) (7 pts.) Compute the p-values for each statistic.
Assume alternative is $H_A: r_{pop} \neq 0$ for all 4 tests.

Z	**χ^2**	**t**	**F**
i) p-value= _____% *(1 pt.)* ***ii)*** *Label Z on the normal curve below and shade the area representing the p-value.(1 pt.) 	p-value=_____% *(1 pt.)* No work necessary.	p-value for t is __ p-value for Z *(1 pt.) Circle one:* **a)** < **b)** > ***c)*** =	**i)** p-value for F is ____ p-value for t *(1 pt.) Circle one:* **a)** < **b)** > ***c)*** = **ii)** p-value for F is ____ p-value for χ^2 *(1 pt.) Circle one:* **a)** < **b)** > **c)** =

*If Z is between 2 lines in the table, choose the closest line. Don't round the given middle area given in the table.

d) (1 pt.) If you changed the alternative for the Z and t tests to $H_A: r_{pop} > 0$ how their p-values change? They would be …
i) divided by 2 ii) stay the same iii) multiplied by 2

e) To use the ANOVA chart to calculate F, you'd need to calculate SST. Suppose $SD_y = 5$, and n=20, and r=0.4.
i) *(2 pts.)* What is SST? *Show work.*

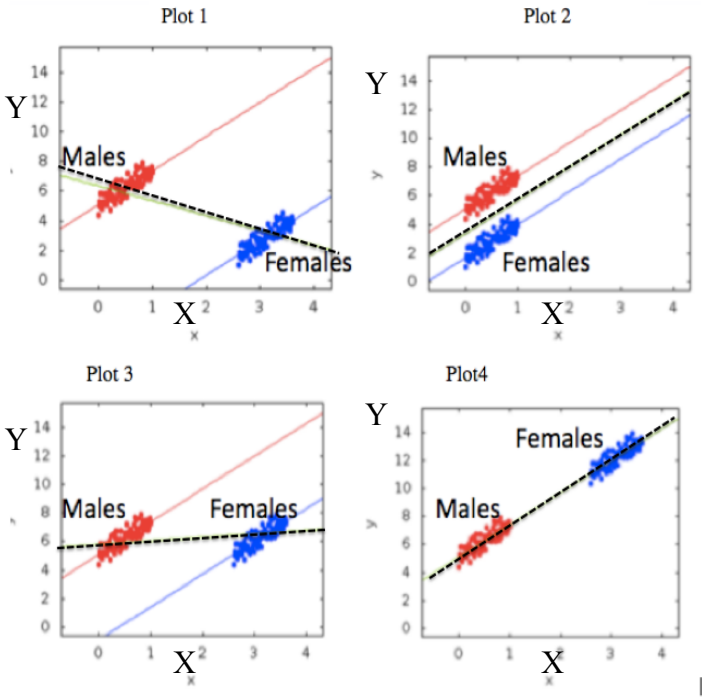
ii) *(2 pts.)*What is SSM? *Show work.*

Stat 200 Exam 2 October 30, 2017

Question 2 *(8 pts.)*

Let's say the 4 plots below depict data from 4 populations and we're trying to figure if X causes Y in these 4 populations. Each plot consists of 2 groups (males and females as marked).

.



a) *(2 pts.)* First let's focus on the relation between X and Y within each group.

Is there the same strong positive relation between X and Y for both males and females in each population?

i) No, because males and females have different Y levels in some of the populations.

ii)No because males and females have different X values in some of the populations.

iii) Yes because they all have the same slope

b) *(2 pts.)* Now, let's focus on the overall regression effect (indicated by the dashed line) in the 4 plots.

For which plots does the overall regression effect agree with the group regression effects?

i) Only Plot 4 since the overall slope and the overall intercept is the same as the group slopes and intercepts.

ii) Plots 2 and 4 only, since the overall slope is the same as the group slopes.

iii) None of them because men and women are clearly separate groups in all 4 plots.

c) *(2 pt.)* In which plot is there an interaction effect between Gender and X?

Circle all that apply: **i)** Plot 1 **ii)** Plot 2 **iii)** Plot 3 **iv)** Plot 4 v) None

d) *(2 pt.)* In which plots is the overall influence of X on Y confounded because of gender?

Circle all that apply: **i)** Plot 1 **ii)** Plot 2 **iii)** Plot 3 **iv)** Plot 4 v) None

Question 3 *(6 pts.)*

Suppose A and B are 2 drugs designed to help improve test scores.

The numbers in each table indicate the average number of points gained in 4 groups—those who received neither drug, those who received only Drug A, those who received only Drug B, and those who received both drugs. Each table describes a different hypothetical study.

Fill in the missing blanks so that the equation and the tables match. Fill in ALL 12 blanks.

No Interaction

Points= _____ + _____A + ____B

	A=0	A=1
B=0	2	5
B=1	6	_____

Only Interaction

Points= _____ + ____ AB

	A=0	A=1
B=0	2	____
B=1	_____	7

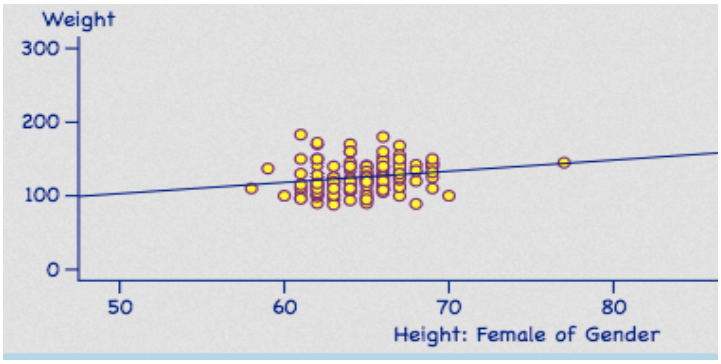
Points= 5A + 6B – 2AB

	A=0	A=1
B=0	____	____
B=1	_____	_____

Question 4 *(4 pts.)*

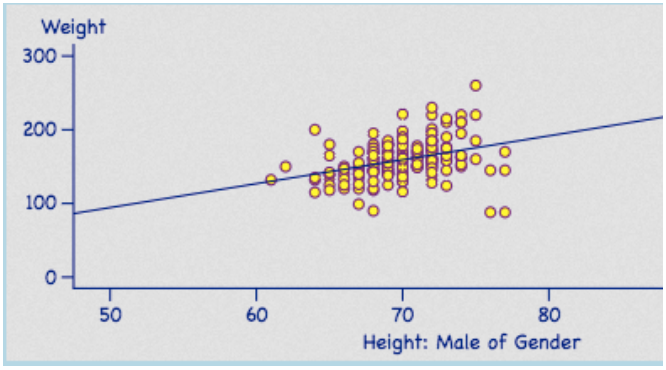
The scatter plots below show the Height (in inches) on the X axis and the Weight (in lbs.) on the Y axis of the 123 females and 165 males in this class who responded to Survey 1.

Females



Female: Weight = 27.18 + 1.613 × Height

Males



Male: Weight = -67.78 + 3.251 × Height

a) *(2 pts.)*Translate the 2 simple regression equations into the multiple regression equation below. Assume Gender is a 0-1 variable coded with Males=0 and Females=1.

Weight = _____ + _____***Height** + _____ **Gender** + _____ **Gender*Height**

—

b) *(2 pts.)* If you switched the code so that Males=1 and Females=0, what would the multiple regression equation be?

Weight = _____ + _____***Height** + _____ **Gender** + _____ **Gender*Height**

Question 5 *(4 pts.)* When the null hypothesis is true in a regression model with 8 parameters and large n, you'd expect your F stat to be about _____ and your χ^2 stat to be about _____ when the null is true. *Write a number in each blank.*

Question 6 (*28 pts.)* Our last survey yielded this multiple regression equation: $\hat{P}iercings = -0.01 + 1.34 * Gender + 0.6 * Tattoos$ based on 231 responses to questions asking: How many piercings do you have?, How many tattoos do you have? and What's your gender? Imagine the 231 students were randomly drawn from a much larger population of Stat 200 students.

a) *(2 pts.)* The above equation describes the best fitting _____ *Circle one:* **i)** line **ii)** plane **iii)** ellipsoid **iv)** cube through all the points so as to minimize the squared errors in the _____. *Circle one:* **a)** Piercings **b)** Tattoos **c)** Gender

b) *(2 pts.)* What does the 1.34 Gender slope mean in the equation above? *(Gender is coded as 1 for females and 0 for males.)*
i) For those with the same number of piercings, females have 1.34 more tattoos than males on the average.
ii) For those with the same number of tattoos, females have 1.34 more piercings than males on the average.
iii) Regardless of the number of tattoos, females have 1.34 more piercings than males on the average.

c) *(2 pts.)* The multiple regression equation for all students with Tattoos = 5 simplifies to: $\hat{P}iercings =$ _____ + _____*Gender
(Fill in the 2 blanks with numbers.)

d) *(1 pt.)* Comparing 2 people of the same sex who differ by 5 in their number of tattoos, how much does the regression equation estimate they'd differ in their number of piercings?_____

e) (2 pts.) Comparing 2 people of the opposite sex who differ by 5 in their number of tattoos, how much does the regression equation estimate they'd differ in their number of piercings …
i) if the female has 5 more tattoos than the male?_____ ii) if the male has 5 more tattoos than the female?_____

f) Do the F test for the overall regression effect for the model: $\hat{P}iercings = -0.01 + 1.34 * Gender + 0.6 * Tattoos$ given **R= 0.7**
i) *(2 pt.)* Compute the F statistic.
Show work. Circle answer.

ii) *(2 pt.)*Look at the F table. What is F* at $\alpha = 0.01$? F* = _____(If df is between 2 lines on table use line with the smaller df)
iii) *(2 pts.)* Our F test stat ____F^* so our p-value is _____ than 1%. *Fill in the two blanks with either* $<$, $>$, *or* $=$

g) *(2 pt.)* Suppose you decided to reject the null, you'd conclude that … *Circle one:*
i)The gender slope must be significant ii) The tattoo slope must be significant
iii) Both must be significant iv) Either the gender or tattoo slope or both must be significant.

h) *(2 pt.)* To see which slope is significant in the multiple regression equation the computer ran a Z test and a t-test. Which test would yield higher p-values for the slopes? ***Circle one:*** **i)** t-test **ii)** Z-test **iii)** they'd be exactly the same **iv)** Not enough info

i) *(1 pt.)* How many degrees of freedom for the t-test? _____

j) (2 pts.) The computer ran 2-sided significance tests for each slope and found they were both significant at α=0.02. Translate these results into Confidence Intervals that give exactly the same information by filling in the 2 blanks with the correct numbers.

_____% Confidence Intervals for both slopes do not include _____. Fill in both blanks with numbers.

k) *(2 pt.)* Another variable is added to the model that is negatively correlated with piercings. Will R^2 go up or down? Circle one:
i) Up ii) Down iii) Stay the same iv) Either down or stay the same v) not enough info

l) *(2 pt.)* Let's say a 3rd variable that's correlated with piercings is added to the multiple regression model and the slopes for gender and tattoos stay the same. You can conclude the 3rd variable must be
i) correlated with either gender or tattoos. **ii)** uncorrelated with both gender and tattoos. **iii)**correlated with both gender and tattoos.
iv) negatively correlated with gender and positively correlated with tattoos (or vice versa) so their effects cancel out.

m) *(2 pt.)* How was the multiple correlation, **R= 0.7** calculated?
- **i.** It's the correlation between the 3 variables.
- **ii.** It's product of the 3 off-diagonal r's given in the correlation matrix.
- **iii.** It's the correlation between the students' actual number of piercings and their predicted number of piercings from the multiple regression equation.

Question 7 *(20 pts.)* On the last survey **231** students rated how strongly they believed in God on a scale of 0-10 (with 0 meaning not at all and 10 meaning absolutely). Below are the (rounded) results classified by ethnic groups. Imagine the 231 students were randomly drawn from a much larger population of all Stat 200 students.

	Group	Mean	SD	n
Belief in God	White	5.3	3.5	44
Belief in God	East Asian	3.7	3.2	133
Belief in God	South Asian	6.6	2.8	27
Belief in God	Other	6.1	4	27

$SD^{+}_{errors} = 3.34$

Use SD^{+}_{errors} as a pooled estimate of the SD for each group.

a) *(9 pts.)* First, to assess whether *any* of the differences between the group averages are significant in the population, fill out the 9 missing numbers in the ANOVA table below.

Source	SS (Rounded to nearest whole number)	df	MS (Round to 2 decimal places)	(Round to 3 decimal places)
Model	SSB= 299	_____	_____	F=_____
Error	SSW=_____	_____	_____	SD^{+}_{errors}= _____
Total	SST= 2835	_____	Don't write anything in this box.	R^2 =_____

b) *(1 pt.)* The p-value is closest to … i) < 1% ii) 1% to 5% iii) 5% to 10%

c) *(1 pt.)* What do you conclude? **Choose one.**

- **i)** That all of 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.

d) Now focus on the difference between the White and East Asian group averages.

i) *(2 pts.)* What is the $SE_{difference}$? ***Show work. Circle answer.*** *Round to 3 decimal place*.

ii) *(2 pts.)* What is the t-statistic? ***Show work. Circle answer.*** *Round to 2 decimal places*.

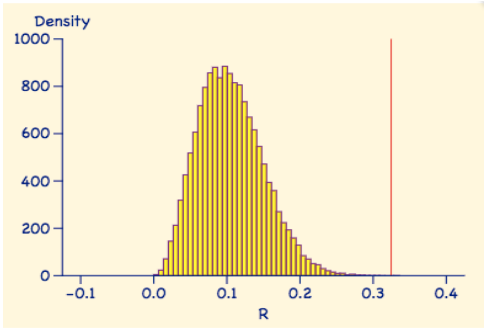
iii) *(1 pt.)* How many degrees of freedom?_____

iv) *(2 pts.)* The p-value is about 0.58%. The Bonferroni correction would change the p-value to _____%.

v) (2 pts.) The Bonferroni correction lowers the probability of making a… ***Choose one.***

i) Type 1 error **ii)** Type 2 error **iii)** Both

Question 7 cont. (6 pts.) Now let's test the same null ($R_{pop}= 0$) against the same alternative ($R_{pop} \neq 0$) using the randomization test. The histogram below shows the distribution of R's resulting from 60,000 randomizations.



e) (2 pts) How was each randomization done?

i. The 231 actual ratings given by the students were randomly distributed among the 4 ethnic groups.
ii. The 231 students were randomly assigned numbers between 0 and 10 and randomly assigned to the 4 ethnic groups.
iii. Random numbers between 0 and 10 were assigned to the 4 ethnic groups.
iv. The 151 students were randomly assigned numbers between 1 and 60,000 and randomly assigned to the 3 ethnic groups.

f) (2pts.) The vertical line falls between 0.3 and 0.4, give its exact location to 3 decimal points.

Show work.

g) (2 pts.)What does the area of the histogram to the right of the vertical line represent?

Circle all that are true.

i. The percentage of randomized R's as big or bigger as the R we obtained from our survey sample.
ii. The p-value given by the F-test.
iii. The p-value given by the randomization test.

Question 8 *(4 pts.)* 30 numbers are divided into 3 groups of 10 numbers each.
The 3 group means are: 1, 4 and -5, and the overall mean of the 30 numbers is 0. Which of the following do you have enough information to calculate?
Circle one: **i)** SST **ii)** SSB **iii)** SSW **iv)** None ***Calculate the one you circled***. If you circled none, leave blank. *Show work.*

Questions 9 *(2 pts.)* In an ANOVA analysis comparing group means, if SSB = 0, then what ***must*** be true:

a) All group means = 0
b) The overall mean = 0
c) All group means are the same.
d) All of the above
e) None of the above

Question 10 *(2 pts.)* In an ANOVA analysis comparing group means, if SSW= 0, then what ***must*** be true:

a) The numbers within each group are the same.
b) The numbers within each group must sum to 0.
c) All group means = 0
d) All of the above
e) None of the above

Critical Values for F distribution at p = 5% and p = 1%

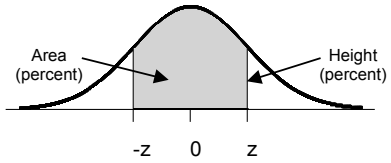
F Distribution critical values for P=0.05

↓ Denominator	Numerator DF													
DF	**1**	**2**	**3**	**4**	**5**	**7**	**10**	**15**	**20**	**30**	**60**	**120**	**500**	**1000**
1	161.45	199.50	215.71	224.58	230.16	236.77	241.88	245.95	248.01	250.10	252.20	253.25	254.06	254.19
2	18.513	19.000	19.164	19.247	19.296	19.353	19.396	19.429	19.446	19.462	19.479	19.487	19.494	19.495
3	10.128	9.5522	9.2766	9.1172	9.0135	8.8867	8.7855	8.7028	8.6602	8.6165	8.5720	8.5493	8.5320	8.5292
4	7.7086	6.9443	6.5915	6.3882	6.2560	6.0942	5.9644	5.8579	5.8026	5.7458	5.6877	5.6580	5.6352	5.6317
5	6.6078	5.7862	5.4095	5.1922	5.0504	4.8759	4.7351	4.6187	4.5582	4.4958	4.4314	4.3985	4.3731	4.3691
7	5.5914	4.7375	4.3469	4.1202	3.9715	3.7871	3.6366	3.5108	3.4445	3.3758	3.3043	3.2675	3.2388	3.2344
10	4.9645	4.1028	3.7082	3.4780	3.3259	3.1354	2.9782	2.8450	2.7741	2.6996	2.6210	2.5801	2.5482	2.5430
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7066	2.5437	2.4035	2.3275	2.2467	2.1601	2.1141	2.0776	2.0718
20	4.3512	3.4928	3.0983	2.8660	2.7109	2.5140	2.3479	2.2032	2.1241	2.0391	1.9463	1.8962	1.8563	1.8498
30	4.1709	3.3159	2.9223	2.6896	2.5336	2.3343	2.1646	2.0149	1.9317	1.8408	1.7396	1.6835	1.6376	1.6300
60	4.0012	3.1505	2.7581	2.5252	2.3683	2.1666	1.9927	1.8365	1.7480	1.6492	1.5343	1.4672	1.4093	1.3994
120	3.9201	3.0718	2.6802	2.4473	2.2898	2.0868	1.9104	1.7505	1.6587	1.5544	1.4289	1.3519	1.2804	1.2674
500	3.8601	3.0137	2.6227	2.3898	2.2320	2.0278	1.8496	1.6864	1.5917	1.4820	1.3455	1.2552	1.1586	1.1378
1000	3.8508	3.0047	2.6137	2.3808	2.2230	2.0187	1.8402	1.6765	1.5811	1.4705	1.3318	1.2385	1.1342	1.1096

F Distribution critical values for P=0.01

↓ Denominator	Numerator DF													
DF	**1**	**2**	**3**	**4**	**5**	**7**	**10**	**15**	**20**	**30**	**60**	**120**	**500**	**1000**
1	4052.2	4999.5	5403.4	5624.6	5763.6	5928.4	6055.8	6157.3	6208.7	6260.6	6313.0	6339.4	6359.5	6362.7
2	98.503	99.000	99.166	99.249	99.299	99.356	99.399	99.433	99.449	99.466	99.482	99.491	99.497	99.498
3	34.116	30.817	29.457	28.710	28.237	27.672	27.229	26.872	26.690	26.504	26.316	26.221	26.148	26.137
4	21.198	18.000	16.694	15.977	15.522	14.976	14.546	14.198	14.020	13.838	13.652	13.558	13.486	13.474
5	16.258	13.274	12.060	11.392	10.967	10.455	10.051	9.7222	9.5526	9.3793	9.2020	9.1118	9.0424	9.0314
7	12.246	9.5467	8.4513	7.8466	7.4605	6.9929	6.6201	6.3143	6.1554	5.9920	5.8236	5.7373	5.6707	5.6601
10	10.044	7.5594	6.5523	5.9944	5.6363	5.2001	4.8492	4.5582	4.4055	4.2469	4.0818	3.9964	3.9303	3.9195
15	8.6831	6.3588	5.4169	4.8932	4.5557	4.1416	3.8049	3.5223	3.3719	3.2141	3.0471	2.9594	2.8906	2.8796
20	8.0960	5.8489	4.9382	4.4306	4.1027	3.6987	3.3682	3.0880	2.9377	2.7785	2.6078	2.5167	2.4446	2.4330
30	7.5624	5.3903	4.5098	4.0179	3.6990	3.3046	2.9791	2.7002	2.5486	2.3859	2.2078	2.1108	2.0321	2.0192
60	7.0771	4.9774	4.1259	3.6491	3.3388	2.9530	2.6318	2.3522	2.1978	2.0284	1.8362	1.7264	1.6328	1.6169
120	6.8509	4.7865	3.9490	3.4795	3.1736	2.7918	2.4720	2.1914	2.0345	1.8600	1.6557	1.5330	1.4215	1.4015
500	6.6858	4.6479	3.8210	3.3569	3.0539	2.6751	2.3564	2.0746	1.9152	1.7353	1.5175	1.3774	1.2317	1.2007
1000	6.6603	4.6264	3.8012	3.3379	3.0356	2.6571	2.3387	2.0564	1.8967	1.7158	1.4953	1.3513	1.1947	1.1586

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