

EXAM 2: Statistics 100**READ THE DIRECTIONS BELOW TWICE!****Cover Sheet Questions**

1) What's your **name**? _____
(Last name) (First name)

2) What's your **net ID** (email)? _____@illinois.edu

3) Which **section** are you in?

Circle one: i) L1 (Kelly Findley In-Person) ii) L2 (Karle Flanagan In-Person) iii) KF (Online)

This test is ALL multiple choice. Circle all answers on this exam and fill in the corresponding bubble on your scantron. All questions have exactly one answer. If you circle/bubble in more than one answer, you will automatically be marked wrong. Make sure to circle the answers on this test and fill out your scantron. If you don't do both, you will get a 0.

SCANTRON Form Directions

- Print and bubble in your **LAST NAME** with **no spaces** starting in the left most column. Print your **FIRST INITIAL** in the right-most column.
- Print and bubble in your Student ID number (UIN) in the Student Number box.
- Print and bubble in your **NET ID** with **no spaces** in the **NETWORK ID** box.
- *No need to bubble in anything for Section or Form.*

READ THIS: Failure to fill out your scantron correctly will result in a loss of 2 points on your exam!

WARNING- The exams look alike but you are sitting next to people who actually have a different version than you. Copying from anyone is equivalent to giving a signed confession.

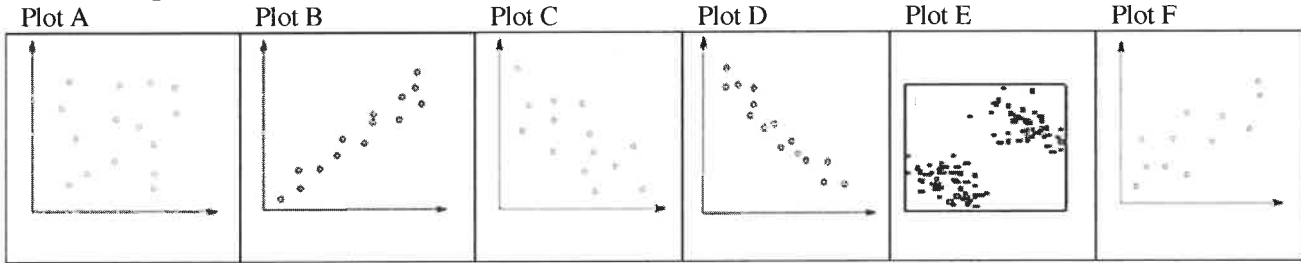
All cheating including being caught with a non-permissible calculator or formula sheet will result in a 0 and an academic integrity violation on your University record.

Make sure you have all 6 pages including the normal table (78 questions).

There is NO CLASS on Thursday or Friday!

Scores will be posted on Compass by Friday afternoon (the deadline to drop a class here at UIUC) and exams will be returned in class next week. Online students may pick up their exam in 23 Illini Hall during office hours next week.

Questions 1-6 pertain to the 6 scatter plots below:



Determine whether the correlation coefficient is appropriate for analyzing the plots. If so, choose the r which best represents the plot.

1. Plot A	<input checked="" type="radio"/> a) $r = 0$	<input type="radio"/> b) $r = +0.8$	<input type="radio"/> c) $r = +0.3$	<input type="radio"/> d) $r = -0.3$	<input type="radio"/> e) Not appropriate to use r
2. Plot B	<input type="radio"/> a) $r = 0$	<input checked="" type="radio"/> b) $r = +0.8$	<input type="radio"/> c) $r = +0.3$	<input type="radio"/> d) $r = -0.3$	<input type="radio"/> e) $r = -0.8$
3. Plot C	<input type="radio"/> a) $r = 0$	<input type="radio"/> b) $r = +0.8$	<input type="radio"/> c) $r = +0.3$	<input checked="" type="radio"/> d) $r = -0.3$	<input type="radio"/> e) $r = -0.8$
4. Plot D	<input type="radio"/> a) $r = 0$	<input type="radio"/> b) $r = +0.8$	<input type="radio"/> c) $r = +0.3$	<input type="radio"/> d) $r = -0.3$	<input checked="" type="radio"/> e) $r = -0.8$
5. Plot E	<input type="radio"/> a) $r = 0$	<input type="radio"/> b) $r = +0.8$	<input type="radio"/> c) $r = +0.3$	<input type="radio"/> d) $r = -0.3$	<input checked="" type="radio"/> e) Not appropriate to use r
6. Plot F	<input type="radio"/> a) $r = 0$	<input type="radio"/> b) $r = +0.8$	<input checked="" type="radio"/> c) $r = +0.3$	<input type="radio"/> d) $r = -0.3$	<input type="radio"/> e) $r = -0.8$

Questions 7-11 pertain to the following situation: For each of the following pairs of variables, check the box that best describes its correlation coefficient r . **READ THIS: Each column will be used exactly once!**

	Exactly +1	Between 0 and +1	About 0	Between 0 and -1	Exactly -1
7. Amount of exercise and percent of body fat	<input type="checkbox"/> a)	<input type="checkbox"/> b)	<input type="checkbox"/> c)	<input checked="" type="checkbox"/> d)	<input type="checkbox"/> e)
8. Ice cream sales and temperature	<input type="checkbox"/> a)	<input checked="" type="checkbox"/> b)	<input type="checkbox"/> c)	<input type="checkbox"/> d)	<input type="checkbox"/> e)
9. Weight in pounds and weight in kilograms	<input checked="" type="checkbox"/> a)	<input type="checkbox"/> b)	<input type="checkbox"/> c)	<input type="checkbox"/> d)	<input type="checkbox"/> e)
10. Number of siblings you have and GPA	<input type="checkbox"/> a)	<input type="checkbox"/> b)	<input checked="" type="checkbox"/> c)	<input type="checkbox"/> d)	<input type="checkbox"/> e)
11. Two variables that always add up to 1000.	<input type="checkbox"/> a)	<input type="checkbox"/> b)	<input type="checkbox"/> c)	<input type="checkbox"/> d)	<input checked="" type="checkbox"/> e)

Parts A, B, and C pertain to this situation. Compute the correlation coefficient (r) by filling in the table below.

The averages of $X=4$ and SD of $X=2$. The average of $Y=3$ and SD of $Y=2$.

Plot the points on the graph below.

Part A (Questions 12-15):

X	Y	Z-score for X	Z-score for Y	Products	Summary Statistics									
1	2	12. a) 1.5 <input checked="" type="radio"/> b) -1.5 c) -1 d) 1 e) 0.5	-0.5	0.75	<table border="1"> <tr> <td></td> <td>Average</td> <td>SD</td> </tr> <tr> <td>X</td> <td>4</td> <td>2</td> </tr> <tr> <td>Y</td> <td>3</td> <td>2</td> </tr> </table>		Average	SD	X	4	2	Y	3	2
	Average	SD												
X	4	2												
Y	3	2												
3	0	-0.5	-1.5	0.75										
4	4	0	0.5	13. a) 1.5 b) -1.5 <input checked="" type="radio"/> c) 0 d) 1 e) 0.5										
5	3	0.5	0	0										
7	6	1.5	14. <input checked="" type="radio"/> a) 1.5 b) -1.5 c) -1 d) 1 e) 0.5	2.25										
Totals		15. Total should = <u> </u> a) 1 <input checked="" type="radio"/> b) 0 c) 20 d) -1	Total should = <u>0</u>	Total = <u>3.75</u>										

Part B (Question 16): Find the correlation coefficient r . a) 2.25 b) 1 c) 0.5 d) 0.625 **e) 0.75**

Part C (Questions 17-20): X and Y below refer to the original X and Y values given in Part A. What happens to the correlation coefficient r if?

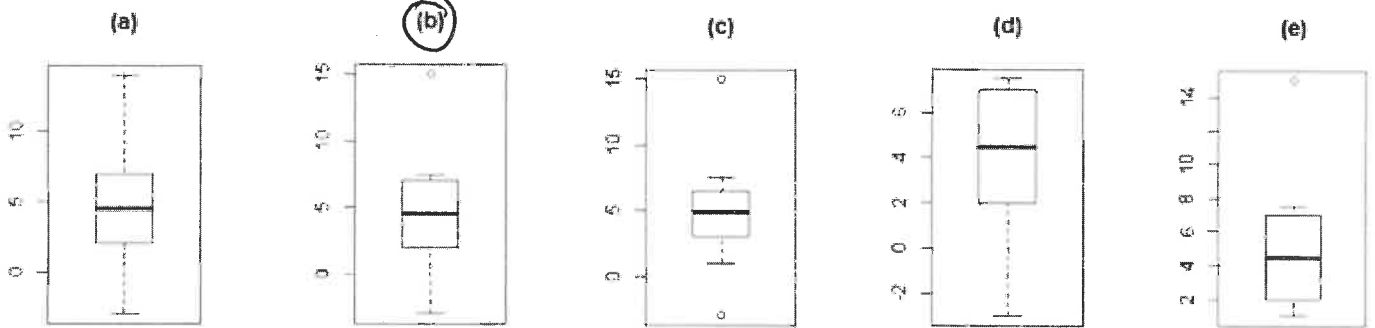
- 17. We add 5 to each Y value. a) r changes signs b) r changes, but we aren't sure how **c) r stays the same**
- 18. We multiply each X value by 2. a) r changes signs b) r changes, but we aren't sure how **c) r stays the same**
- 19. We divide each X value by -4. **a) r changes signs** b) r changes, but we aren't sure how c) r stays the same
- 20. We switch all X and Y values. a) r changes signs b) r changes, but we aren't sure how **c) r stays the same**

Questions 21-26 pertain to this situation about boxplots: Consider the following set of numbers: -3, 1, 3, 3, 6, 6.5, 7.5, 15

- 21. The median for this set of numbers is... a) 3 b) 4 **c) 4.5** d) 5 e) 5.75
- 22. The 1st quartile (Q1) is... a) -3 **b) 2** c) 3 d) 4 e) 7
- 23. The 3rd quartile (Q3) is... a) 3 b) 5 c) 6.5 **d) 7** e) 15
- 24. The IQR is... a) 2 b) 4 **c) 5** d) 7 e) 18
- 25. Are there any outliers?
 a) No b) Yes, a low outlier **c) Yes, a high outlier** d) Yes, a high and low outlier

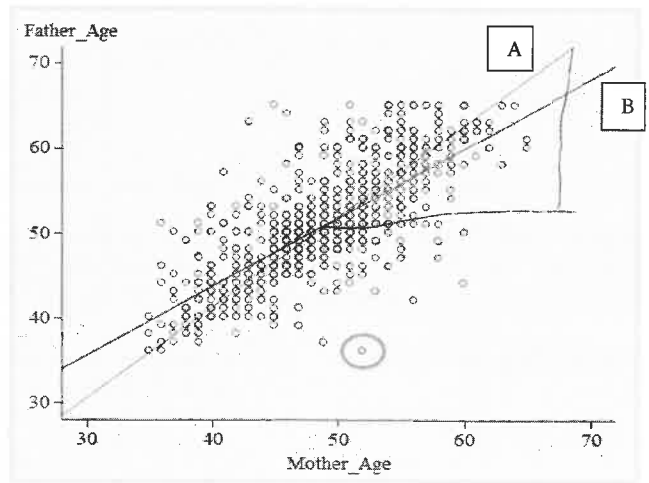
low outliers $< 2 - 1.5(5) = -5.5$ high outliers $> 7 + 1.5(5) = 14.5$

26. Which of the following looks like an appropriate box model for this set of numbers?



Questions 27-34 pertain to the following scenario: In the very first survey of the year, we asked students "How old is your mother?" and "How old is your father?" The scatter plot of mother's and father's ages is shown below along with the summary statistics.

- 27. The SD Line and the Regression Line are shown. Which is the SD Line? **a) Line A** b) Line B
- 28. The correlation is closest to a) 0.2 b) 0.4 **c) 0.8** d) -0.2 e) -0.6
- 29. About what is the average Mother's Age? a) 40 **b) 50** c) 55 d) 60 e) 65
- 30. About what is the average Father's Age? a) 40 **b) 50** c) 55 d) 60 e) 65
- 31. We think that the point circled on the scatter plot is an outlier. What does removing this outlier do to the correlation coefficient?
 a) Outliers have no effect on the correlation coefficient.
b) Removing this outlier raises the correlation coefficient.
 c) Removing this outlier lowers the correlation coefficient.



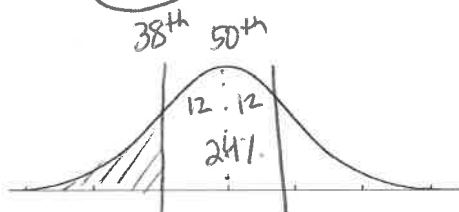
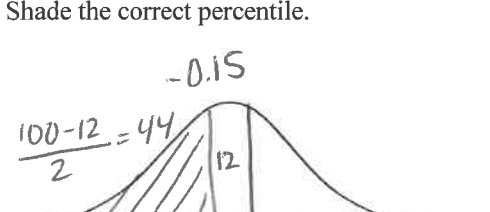
32. Points above the regression line have prediction errors (residuals) that are: a) negative b) positive c) equal to zero
33. The average of all the prediction errors is always _____. a) different depending on the data b) 1 c) 0 d) r
34. Which must be true if you lie on the SD line?
 a) You have the same z-scores for mother age and father age
 b) Your prediction error is equal to 0.
 c) You are exactly average.

Questions 35-39 pertain to the following scenario.

Suppose people's scores on a Math skills test and a Social skills test follow the normal curve but have different correlations in different countries. Consider 5 populations where the correlation coefficients between people's Math and Social skills are as given in the table below. If a child is in the 25th percentile in math, estimate her percentile in social skills in each country.

Math Skills Percentile	r	Social Skills Percentile
25 th	-1	35. a) 25 th b) 37 th c) 50 th d) 63 rd e) 75 th
25 th	1	36. a) 25 th b) 37 th c) 50 th d) 63 rd e) 75 th
25 th	0	37. a) 25 th b) 37 th c) 50 th d) 63 rd e) 75 th
25 th	-0.5	38. a) 25 th b) 37 th c) 50 th d) 63 rd e) 75 th
25 th	0.5	39. a) 25 th b) 37 th c) 50 th d) 63 rd e) 75 th

Questions 40-43 pertain to the following scenario. If someone is in the 38th percentile in Math where $r = 0.5$, what percentile would you estimate for his social skills? Solve by filling in the table below. Round the middle areas to the nearest whole numbers.

Math Skills Percentile	Math Z	r = 0.5	Social Z	Social Skills Percentile
38 th Percentile	41. Z = -0.3	r = 0.5	42. Z = -0.15	43. Social Skills Percentile = 44 th
40. What middle area on the table should you look up to find the Z score? a) 12% b) 24% c) 31% d) 38% e) 62%	a) 0.5 b) -0.5 c) 0.3 d) -0.3 e) -0.15		a) 0.25 b) -0.25 c) 0.5 d) 0.15 e) -0.15	Hint: Mark the Z score on the graph below. Shade the correct percentile.
				

Questions 44-48 pertain to a roadside drunk driving test. Suppose only 10% of those who get stopped and tested for drunk driving are really drunk. 70% of the drunk drivers will correctly fail the test. But 30% of the sober drivers will also fail the test (due to nervousness or other reasons). Fill in the table for 100 drivers total who get stopped and test.

	Fail Test	Pass Test	Total
Drunk Drivers	44. a) 1 b) 3 c) 5 d) 7 e) 9	3	10
Sober Drivers	0.3(90) 27	63	45. a) 10 b) 30 c) 50 d) 70 e) 90
Total	34	46. a) 66 b) 30 c) 50 d) 70 e) 90	100

47. If a driver fails the test, what's the chance he or she is really sober? a) 21% b) 27% c) 30% d) 79% e) 95%
48. If a driver passes the test, what is the chance that he or she is really drunk? a) 3% b) 4.5% c) 30% d) 79% e) 11%

Questions 49-54 pertain to the survey questions: "What is the fastest you've ever driven in mph?" and "What is the most you have ever drank in a 24-hour period?" To the right are the 5 summary statistics.

	Average	SD
Speed	96mph	20mph
Drinks	13	10

Correlation: $r = 0.4$

Make regression estimates for Students A and B by filling in the table below. For student A, you're given speed and asked to predict drinks using the 3-step process. For student B, you're given drinks and asked to predict speed using the 3-step process so start with drinks.

Speed (in mph)	Speed Z-score	r	Drinks Z-score	Number of Drinks
Student A: 106 mph $z = \frac{106-96}{20} = 0.5$	49. Z = 0.5 a) 0.5 b) -0.5 c) -1 d) 1 e) 0.4	$r = 0.4$	50. Z = 0.2 a) 1 b) 0.5 c) -1 d) 0.2 e) 0.4	51. Drinks = 15 a) 100 b) 11 c) 15 d) 18 e) 23 $13 + 0.2(10)$
54. 112 mph	53. Z = 0.8 a) 1 b) 5 c) 2.5 d) 0.4 e) 0.8	$r = 0.4$	52. Z = 2 a) 1 b) 2 c) -1 d) -2 e) 0.4	Student B: 33 Drinks $z = \frac{33-13}{10} = 2$

Question 55

In 2010, millions of students nationwide took the Math and Verbal SAT tests. Within each state, the average verbal and the average math SAT scores were calculated. The correlation between these 50 pairs of averages was 0.97. Does that mean that the correlation between the millions of individual students' Verbal and Math SAT scores would also be 0.97?

- a) Yes, since the state averages are computed from the individual scores, the correlation for individual Verbal and Math scores must also be 0.97.
- b) No, the correlation for individuals is likely to be lower than 0.97 since the individuals within each state would add more scatter so the points wouldn't follow a straight line so closely.
- c) No, the correlation for individuals is likely to be higher than 0.97 since it's based on millions of data points instead of just 50.

Questions 56-61 pertain to the following situation

A large study gave the following results on the heights of fathers and the heights of their daughters. Assume the scatter plot is football-shaped and round to two decimal places.

	Average	SD
Father's Height	69"	2.5"
Daughter's Height	64"	1.5"

Correlation: $r = 0.7$

Find the regression equation for predicting father's heights from daughter's heights. $Y = \text{_____} \bar{X} + \text{_____}$ (round the slope and y intercept to two decimal places).

56. First calculate the slope. Round your answer to two decimal places!

- a) 1.67
- b) 0.60
- c) 0.70
- d) 0.42
- e) .17

$r \times \frac{SD_y}{SD_x} = 0.7 \times \frac{2.5}{1.5} = 1.17$

57. Next, calculate the y-intercept (choose the closest answer).

- a) -5.88
- b) -16.73
- c) 42.12
- d) 35.02
- e) 0.92

$69 = 1.17(64) + b$
 $b = -5.88$

58. Use the regression equation to predict the height of a father whose daughter is 62 inches tall. Round to the nearest whole number.

- a) 20 inches
- b) 42 inches
- c) 67 inches
- d) 70 inches
- e) None of the above

$y = 1.17(62) - 5.88$

59. Suppose our regression line predicts that a father's height is 64 inches, but his actual height is 70 inches. What is his residual?

- a) -6
- b) 8
- c) -8
- d) 1.78
- e) 6

$70 - 64$

60. What is the SD of the prediction errors (the RMSE) when predicting father's heights from daughter's heights?

- a) 2.5
- b) 1.5
- c) $\sqrt{1-0.7^2} \times 1.5$
- d) $\sqrt{1-0.7^2} \times 2.5$
- e) 0

61. Of course, the prediction we made in question 58 isn't going to be exactly correct. Instead there's a range of father's heights, with about 95% of them falling within ...

- a) 1 SDerrors
- b) 2 SDerrors
- c) 3 SDerrors
- d) Not enough info

Questions 62-66 pertain to the table below which shows the survey responses of 389 students who identified themselves as "white" & the 201 students who identified themselves as "non-white" to the question "Who do you plan to vote for in the 2012 election?"

	Obama	Romney	Undecided/Other	Won't Vote	Totals
White	207	82	77	23	389
Non-White	124	9	25	43	201
Totals	331	91	102	66	590

Suppose you randomly draw a student from those who answered this survey:

62. What is the chance that you'll get a student voting for Romney?
 a) $9/91$ b) $82/389$ c) $9/201$ d) $82/91$ e) $91/590$
63. What is the chance that you'll get a student voting for Romney if you draw only from white students?
 a) $91/389$ b) $91/590$ c) $82/389$ d) $82/91$ e) $9/91$
64. What is the chance of getting a white student if you draw only from Romney voters?
 a) $91/389$ b) $91/590$ c) $82/389$ d) $9/201$ e) $82/91$
65. What is the chance that a randomly selected student is either voting for Obama or white?
 a) $513/590$ b) $331/590 + 389/590$ c) $331/590 + 207/590$ d) $331/590 + 389/590$ e) $207/590$
66. Draw 3 students *with replacement*. What is the chance that at least one of them is a Romney supporter?
 a) $91/590$ b) $(91/590)^3$ c) $1 - (499/590)^3$ d) $(499/590)^3$ e) $1 - (91/590)^3$

Questions 67-70 pertain to a well-shuffled deck of 52 cards.

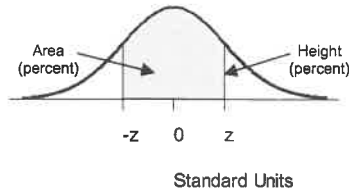
There are 52 total cards, 13 diamonds, 13 spades, 13 clubs, 4 jacks, 4 aces, 4 queens, and 1 queen of clubs in a deck.

67. Draw 2 cards *without replacement*. What is the chance that they are both diamonds?
 a) $26/52$ b) $13/52 * 12/51$ c) $16/52$ d) $13/52 * 13/52$ e) $4/52 * 3/51$
68. Draw 2 cards *with replacement*. What is the chance that the both cards are spades?
 a) $26/52$ b) $13/52 * 12/51$ c) $16/52$ d) $13/52 * 13/52$ e) $4/52 * 3/51$
69. Draw *one* card at random. What is the chance that it's *either* a Jack *or* an Ace?
 a) $8/52$ b) $13/52 * 12/51$ c) $17/52$ d) $13/52 * 13/52$ e) $16/52$
70. Draw *one* card at random. What is the chance that it's *either* a Queen *or* a Club?
 a) $16/52$ b) $17/52$ c) $8/52$ d) $13/52 * 13/52$ e) $26/52$

Questions 71-78 pertain to rolling fair 6 sided dice.

71. Two dice are rolled. What is the chance that the sum of the spots is 3? $2, 1$
 a) $1/36$ b) $2/36$ c) $3/36$ d) $4/36$ e) $2/6$
72. One die is rolled 4 times. What is the chance of getting *no* 2's?
 a) $(5/6)^4$ b) $(1/6)^4$ c) $1 - (5/6)^4$ d) $1 - (1/6)^4$ e) $4/6$
73. One die is rolled 4 times. What is the chance of *at least one* 2?
 a) $(5/6)^4$ b) $(1/6)^4$ c) $1 - (5/6)^4$ d) $1 - (1/6)^4$ e) $4/6$
74. One die is rolled 3 times. What is the chance of getting *all* 2's?
 a) $(5/6)^3$ b) $3/6$ c) $1 - (5/6)^3$ d) $1 - (1/6)^3$ e) $(1/6)^3$
75. One die is rolled 3 times. What is the chance of getting *not all* 2's?
 a) $(5/6)^3$ b) $(1/6)^3$ c) $1 - (5/6)^3$ d) $1 - (1/6)^3$ e) $3/6$
76. One die is rolled twice. What is the chance that the first roll is a 4 *or* the second roll is a 5?
 a) $1/6 * 1/6$ b) $1/6 + 1/6$ c) $4/6 + 5/6$ d) $1/6 + 1/6 - 1/36$
77. One die is rolled twice. What is the chance that the first roll is a 4 *and* the second roll is a 5?
 a) $1/6 * 1/6$ b) $1/6 + 1/6$ c) $4/6 + 5/6$ d) $1/6 + 1/6 - 1/36$
78. A die is rolled once. What is the chance of getting either a 4 or a 5?
 a) $1/6 * 1/6$ b) $1/6 + 1/6$ c) $4/6 + 5/6$ d) $1/6 + 1/6 - 1/36$

STANDARD NORMAL TABLE



<i>z</i>	<i>Area</i>		<i>z</i>	<i>Area</i>		<i>z</i>	<i>Area</i>
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

