Statistics 2000, Section 001, Midterm 1 (200 Points)

Friday, February 12, 2010

Your Name: _____

Question 1: z–Scores and Normal Distributions (**50 Points**)

The Graduate Record Examination (GRE) is a test taken by college students who intend to pursue a graduate degree in the United States. For around 210,000 female US citizens who took the General GRE Test in 2005–06, the mean for the quantitative ability portion of the exam was about 520 and the standard deviation was about 135 (`http://www.ets.org/Media/Tests/GRE/pdf/05-06_factors_final.%20pdf.pdf`). We can assume that the histogram follows a normal curve. **Show your work!**

1. (**15 Points**) The percentage of female US citizens who scored **more than 669** on the GRE test is roughly _____ %.

2. (**20 Points**) The percentage of female US citizens who scored **between 351 and 574** is about _____ %.

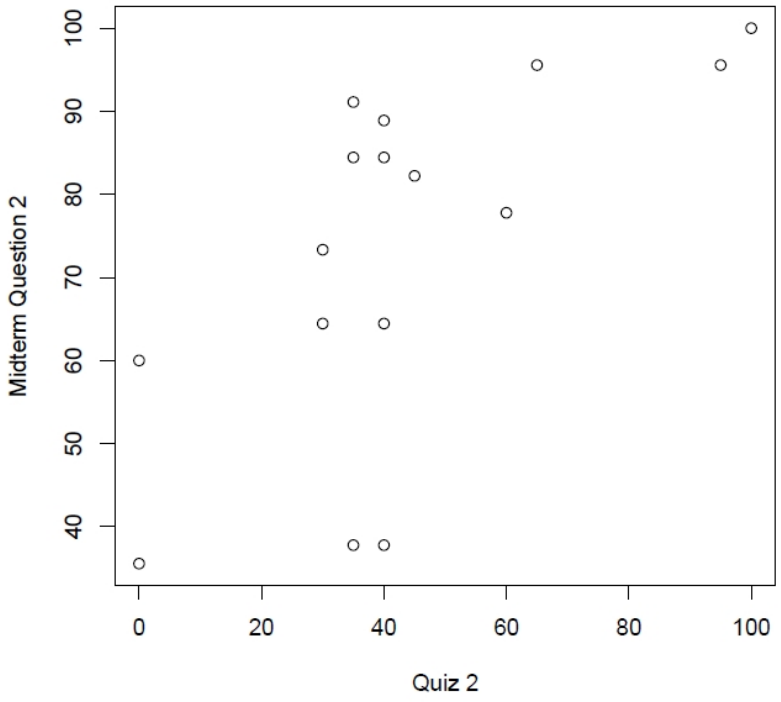
3. (**15 Points**) In order to be among the top 80% of all female US citizens, a student must have obtained a minimum GRE score of about _____.

Question 2: Regression (**50 Points**)

In a particular section of Stat 1040, students had to answer a review exercise of their textbook in Quiz 2. The result was anything but satisfactory, with the median score being an F. Detailed solutions were handed out, together with the graded quizzes. To determine whether students studied the solutions, the instructor reused the same question a few weeks later as Question 2 in Midterm 1. For a better comparability, the scores were adjusted to 100 points. In the table below, avg represents the average and SD the standard deviation.

Midterm Question 2 score:	avg = 73 points;	SD = 21 points;	
Quiz 2 score:	avg = 43 points;	SD = 27 points;	$r = 0.65$.

The scatterplot that shows the data is displayed below and can be assumed to be football–shaped.



Show your work!

1. (**20 Points**) Find the regression equation for predicting the Midterm Question 2 score from the Quiz 2 score.

2. (**10 Points**) Using your regression equation, estimate the Midterm Question 2 score for a student who had a Quiz 2 score of 80 points.

3. (**10 Points**) Can we safely estimate the Midterm Question 2 score for a student who had a Quiz 2 score of 10 points? **YES** or **NO**? **Circle your answer and provide a short explanation.**

4. (**10 Points**) As mentioned above, all scores were adjusted as if graded out of 100 points. However, the Quiz 2 scores were originally graded out of 20 points, that means, each individual Quiz 2 score was multiplied by 5 for this question. Therefore, we had an original average score of _____ points and an original SD of _____ points when grading out of 20 points.

Question 3: Sums and Order Notation (**40 Points**)

In statistics, we usually refer to x_1 as the first observation, x_2 as the second observation, etc., and x_n as the final observation when we write down our observations in the order they were obtained (where n represents the total number of observations).

Often, we prefer to work with data that are sorted from smallest to largest, e.g., when calculating the median, we need the data to be sorted. Obviously, we can simply reorder any given list of numbers. However, we often use the notation $x_{(1)}$ to refer to the smallest observation, $x_{(2)}$ to refer to the 2nd smallest observation, etc., and $x_{(n)}$ to refer to the largest observation.

Show your work!

For $x_1 = -3, x_2 = 5, x_3 = 3, x_4 = 4, x_5 = -2, x_6 = 20$, and $n = 6$, determine the following sums (**8 Points each**):

$$\sum_{i=1}^{n} x_i =$$

$$\sum_{i=2}^{n-2} x_i =$$

$$\sum_{i=2}^{n-2} x_{(i)} =$$

$$\sum_{i=2}^{n-1} x_1 =$$

$$\sum_{i=\frac{n}{2}}^{n^2-32} \frac{x_{n-i}}{x_{(i+1)}} =$$

Question 4: Multiple Choice Questions **(60 Points)**

Mark your answer for each multiple choice question in the table below. There is only one correct answer for each question. Each correct answer is worth 4 points.

Question	(a)	(b)	(c)	(d)	Question	(a)	(b)	(c)	(d)
1	◯	◯	◯	◯	11	◯	◯	◯	◯
2	◯	◯	◯	◯	12	◯	◯	◯	◯
3	◯	◯	◯	◯	13	◯	◯	◯	◯
4	◯	◯	◯	◯	14	◯	◯	◯	◯
5	◯	◯	◯	◯	15	◯	◯	◯	◯
6	◯	◯	◯	◯					
7	◯	◯	◯	◯					
8	◯	◯	◯	◯					
9	◯	◯	◯	◯					
10	◯	◯	◯	◯					

1. Data on the mileage of 20 randomly selected cars are listed below. The values are ordered for convenience.

   ```
   12 13 13 15 15 16 16 16 17 17
   18 20 22 23 24 26 26 27 27 29
   ```

   What is the median mileage for these 20 cars?

   (a) 17.5 miles per gallon

   (b) 19 miles per gallon

   (c) 19.5 miles per gallon

   (d) 20 miles per gallon

2. A company has 20 female workers whose average salary is \$43,000 and 30 male workers whose average salary is \$47,000. What is true about the average salary of all 50 workers?

   (a) It must be \$45,000.

   (b) It must be \$45,400.

   (c) It must be \$47,000 because we have more male workers than female workers.

   (d) It could be any number between \$43,000 and \$47,000.

3. A sample was taken of the salaries of 20 employees from a large company. The following are the salaries (in thousands of dollars) for this year (the data are ordered).

   ```
   28  31  34  35  37  41  42  42  42  47
   49  51  52  52  60  61  67  72  75  77
   ```

   Suppose each employee in the company receives a \$5,000 raise for next year (each employee's salary is increased by \$5,000). The interquartile range of the salaries will

   (a) increase by \$5,000.

   (b) be multiplied by 5000.

   (c) become 50 (i.e., \$50,000).

   (d) be unchanged.

4. A researcher wishes to determine whether the rate of water flow (in liters per second) over an experimental soil bed can be used to predict the amount of soil washed away (in kilograms). The researcher measures the amount of soil washed away for various flow rates, and from these data calculates the least–squares regression line to be

   ```
   amount of eroded soil = 0.4 + 1.3 * (flow rate).
   ```

   What do we know about the correlation r between amount of eroded soil and flow rate?

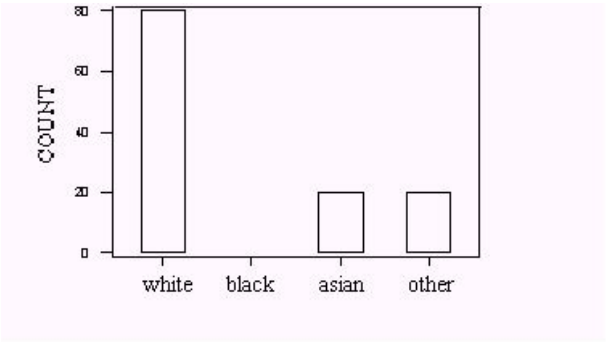
   (a) $r = 1/1.3$

   (b) $r = 0.4$

   (c) It would either be positive or negative. It is impossible to say anything about the correlation from the information given.

   (d) It would be positive, but we cannot determine the exact value.

5. A sample of 160 workers in the downtown area classified each worker by race. A bar graph of the results is given below, but the bar for blacks in the graph below has been omitted.



   Using the information provided, the *proportion* of black workers in the sample must be

   (a) 40%.

   (b) 25%.

   (c) 20%.

   (d) Just 40 (no units).

6. The time plot below is for motor vehicle deaths in the United States. The rate is the number of deaths per million miles driven and is plotted for the 18 years 1960, 1962, 1964, ..., 1992, 1994 (a data point every second year).

Suppose we drew a histogram of these 18 death rates using class intervals 1 – 1.9, 2 – 2.9, 3 – 3.9, 4 – 4.9 and 5 – 5.9. Using the histogram, we would

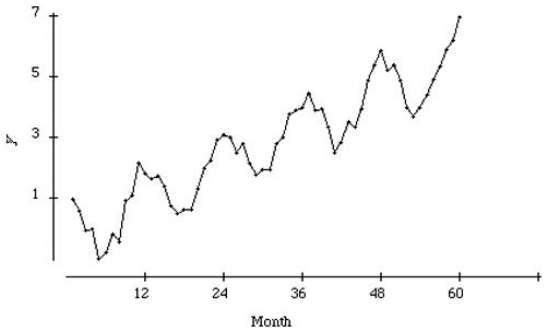
(a) lose all information about trends over time.

(b) be able to compute the number of years in this period for which the death rate was 5 or higher.

(c) both of the above.

(d) none of the above.

7. Consider the following time plot of the values of variable Y for each month of a 60–month period.



This plot displays

(a) seasonal variation.

(b) a trend.

(c) both of the above.

(d) an extreme outlier for month 60.

8. A distribution has a mean of 100 and a median of 120. The shape of this distribution is most likely

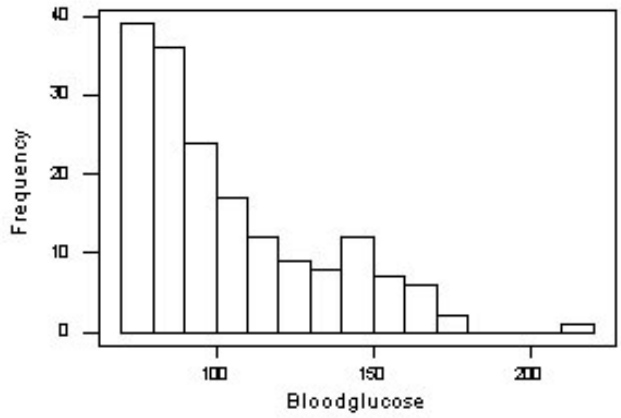
   (a) skewed to the left.

   (b) skewed to the right

   (c) symmetric.

   (d) with two peaks (one around 100 and one around 120).

9. High levels of glucose in the blood are indications of diabetes, which is becoming more prevalent in the United States. Diabetes can lead to many complications such as blindness and heart disease. A random sample of 180 individuals had their blood sugar level measured. The results are displayed in the graph below.



   Blood glucose levels above 130 are considered indicative of diabetes. Based on the histogram, about what percent of those tested in this sample may have diabetes?

   (a) about 80%.

   (b) about 50%.

   (c) about 44%.

   (d) about 25%.

10. Volunteers for a research study were divided into three groups. Group 1 listened to Western religious music, Group 2 listened to Western rock music, and Group 3 listened to Chinese religious music. The blood pressure of each volunteer was measured before and after listening to the music, and the change in blood pressure (blood pressure before listening minus blood pressure after listening) was recorded. What could we do to explore the relationship between type of music and change in blood pressure?

    (a) See if blood pressure decreases as type of music increases by examining a scatterplot.

    (b) Make a histogram of the change in blood pressure for all of the volunteers.

    (c) Make side–by–side boxplots of the change in blood pressure, with a separate boxplot for each group.

    (d) Do all of the above.

11. A study found a correlation of $r = -0.61$ between the gender of a worker and his or her income. Determine which of the following conclusions, based on this correlation coefficient, is true. Note that there is just one true statement!

    (a) Women earn more than men on the average.

    (b) Women earn less than men on the average.

    (c) An arithmetic mistake was made. Correlation must be positive.

    (d) This measurement makes no sense; r can only be measured between two quantitative variables.

12. In a study of 1991 model cars, a researcher computed the least–squares regression line of price (in dollars) on horsepower. He obtained the following equation for this line:

    ```
    price = -6677 + 175 * horsepower.
    ```

    Based on the least–squares regression line, what would we predict the cost of a 1991 model car with horsepower equal to 200 to be?

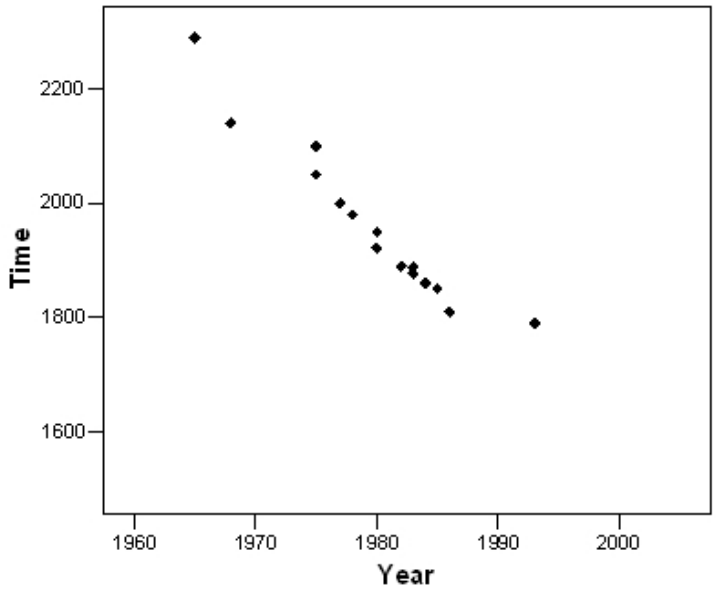
    (a) \$41,677.

    (b) \$35,000.

    (c) \$28,323.

    (d) \$13,354.

13. Below is a scatterplot of the world–record time for women in the 10,000–meter run versus the year in which the record was set. Note that time is in seconds and the data are for the period 1965 to 1995.



Based on this plot, what conclusion can we draw?

(a) By 2010 the world–record time for women will be well below 1500 seconds.

(b) About every decade, the world–record time will decrease by at least 100 seconds.

(c) About every decade, the world–record time will decrease by about 50 seconds.

(d) None of the above.

14. If females of a certain species of lizard always mate with males that are .50 years younger than they are, what would the correlation between the ages of these male and female lizards be?

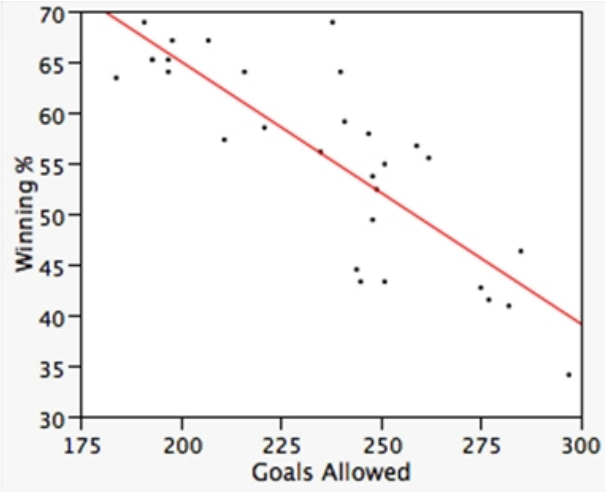
(a) –1.00.

(b) –0.50.

(c) 1.00.

(d) This cannot be answered without knowledge of the actual data.

15. In the National Hockey League a good predictor of the percentage of games won by a team is the number of goals the team allows during the season. Data were gathered for all 30 teams in the NHL and the scatterplot of their **Winning Percentage** against the number of **Goals Allowed** in the 2006/2007 Season with a fitted least–squares regression line is provided:



The least–squares regression line was calculated to be

```
Winning % = 116.95 - 0.26 *  (Goals Allowed)
```

Which of the following provides the best interpretation of the slope of the regression line?

(a) If the Winning % increases by 1% then the number of Goals Allowed decreases by 0.26.

(b) If Goals Allowed increases by one goal, the Winning % increases by 0.26%.

(c) If the Winning % increases by 1% then the number of Goals Allowed increases by 0.26.

(d) If Goals Allowed increases by one goal, the Winning % decreases by 0.26%.

Table entry for z is the area under the standard normal curve to the left of z.

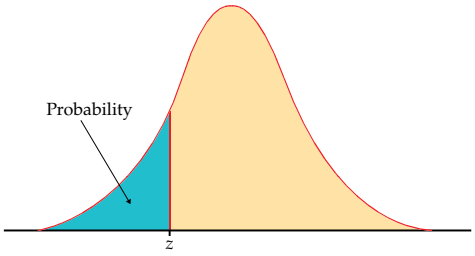


TABLE A

Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Table entry for z is the area under the standard normal curve to the left of z.

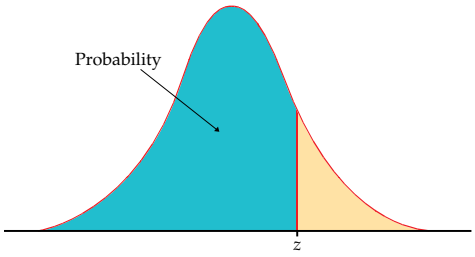


TABLE A

Standard normal probabilities (continued)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998