Statistics 2000, Section 001, Midterm 1 (185 Points)

Friday, October 1, 2010

Part I: Text Answers

Your Name: _____

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

There are two major tests for readiness for college, the ACT and the SAT. ACT scores are reported on a scale from 1 to 36. The distribution of ACT scores for more than 1 million students in a recent high school graduating class was roughly Normal with mean $\mu = 20.7$ and standard deviation $\sigma = 4.7$. SAT scores are reported on a scale from 400 to 1600. The distribution of SAT scores for 1.4 million students in the same graduating class was roughly Normal with mean $\mu = 1022$ and standard deviation $\sigma = 212$.
Show your work!

1. (**10 Points**) Find the ACT equivalent: Paul scores 800 on the SAT. Assuming that both tests measure the same thing, what score on the ACT is equivalent to Paul's SAT score?

2. (**10 Points**) Compare a SAT with an ACT score: Liza scores 1380 on the SAT. John scores 24 on the ACT. Assuming that both tests measure the same thing, who has the higher score — Liza or John? Report the z–scores for both students and answer the question.

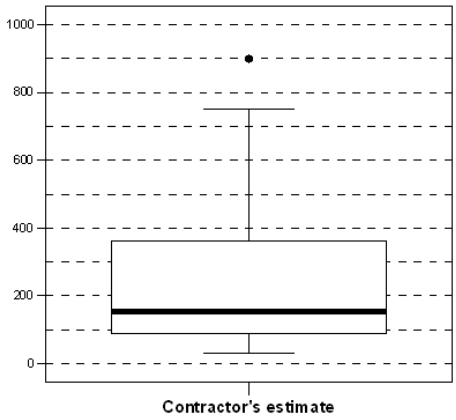
3. (**10 Points**) Top percentage: Jessica is hoping to qualify for a scholarship. Those are only awarded to applicants who have a SAT score among the top–10% of all SAT scores. Which score does she need at least to be sure to be awarded a scholarship?

4. (**10 Points**) Top scores: Mike scores 29 on the ACT. What is the percentage of ACT scores that is higher than his score?

5. (**10 Points**) Find the SAT percentile: Reports on a student's ACT or SAT usually give the percentile as well as the actual score. The percentile is just the cumulative proportion stated as a percent: the percent of all scores that were lower than this one. Melissa scores 1300 on the SAT. What is her percentile?

Question 2: Boxplots (**45 Points**)

The Michigan Department of Transportation (M–DOT) is working on a major project: 80% of the highways in Michigan need to be repaved. To speed completion of this project, many contractors will be working for M–DOT. Contractors are currently bidding on the next part of the project. To help make a decision about which contractor to hire, M–DOT collects many variables besides just the estimated cost. One of those variables is the contractor's estimate of the number of workdays required to finish the job. Twenty contractors have bid on the next job. The boxplot below represents their estimates of the number of work days required:



Answer the questions below, based on the boxplot above. Answers within ±20 days (or within ±5%) of the correct answer will all result in full points.

1. (**7 Points**) What is (approximately) the estimated median number of days?
   Answer: _____ days

2. (**7 Points**) What is (approximately) the estimated minimum number of days?
   Answer: _____ days

3. (**7 Points**) What is (approximately) the estimated maximum number of days?
   Answer: _____ days

4. (**7 Points**) What is (approximately) the estimated third quartile (Q3)?
   Answer: _____ days

5. (**7 Points**) What is (approximately) the interquartile range (IQR)?
   Answer: _____ days

6. (**7 Points**) What is (approximately) the percentage of contractors that estimated the number of days to be more than 100?
   Answer: _____ %

7. (**3 Points**) When we compare the mean with the median, **(i) the mean will be higher than the median, (ii) both will be about the same, or (iii) the median will be higher than the mean.** Just circle the correct answer.

Question 3: Regression (**30 Points**)

Many high school students take either the SAT or the ACT. However, some students take both. Data were collected from 60 students who took both college entrance exams. The average SAT score was 912 with a standard deviation of 180. The average ACT score was 21 with a standard deviation of 5. The correlation between the two variables equals 0.82.

The scatterplot (not reproduced here) shows that there is indeed a linear relationship between the two variables.
Show your work!

1. (**20 Points**) Find the regression equation for predicting the SAT score from the ACT score.

2. (**10 Points**) Using your regression equation, estimate the SAT score for a student who had an ACT score of 23 points.

Statistics 2000, Section 001, Midterm 1 (185 Points)

Friday, October 1, 2010

Part II: Multiple Choice Questions

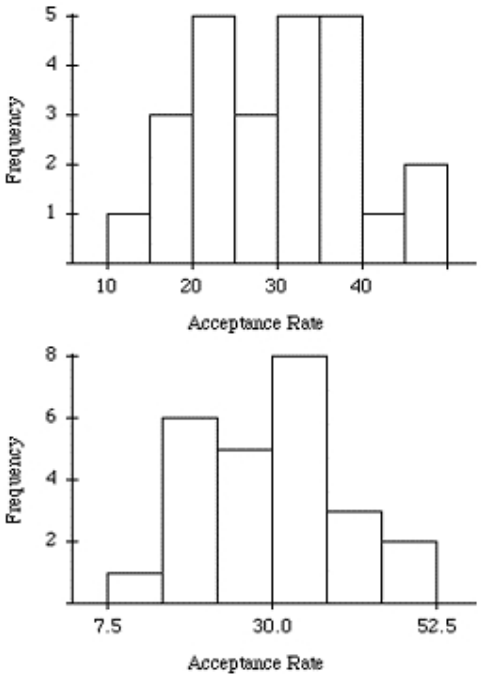
Your Name: _____

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. The accompanying two histograms represent the distribution of acceptance rates (percent accepted) among 25 business schools in 1995. The histograms use different class intervals, but are based on the same data. In each class interval, the left endpoint is included but not the right.



Which statement is true?

(a) The median rate must be less than 30.

(b) The interquartile range exceeds 30.

(c) Neither of the above.

(d) Both of the above.

2. Which of the following measures are not affected by outliers?

(a) The mean.

(b) The standard deviation.

(c) The correlation.

(d) The IQR.

3. 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.

4. Researchers are conducting a state–wide survey for the U.S. Postal Service. The survey records many different variables of interest. Which of the following variables is categorical? There is just one.

(a) County of residence.

(b) Number of people, both adults and children, living in the household.

(c) Total household income, before taxes, in 2003.

(d) Age of respondent.

5. If females of a certain species of lizard always mate with males that are .75 years younger than they are, what would the correlation between the ages of these male and female lizards be?

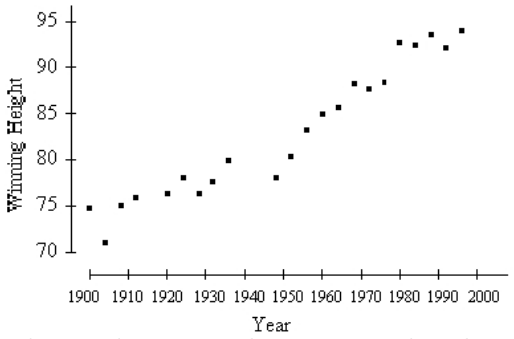
(a) 1.

(b) -0.75.

(c) -1.

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

6. Below is a plot of the Olympic gold medal winning performance in the high jump (in inches) for the years 1900 to 1996.



From this plot, the correlation between the winning height and year of the jump is

   (a) about 0.95.
   (b) about 0.10.
   (c) about -0.50.
   (d) about -0.95.

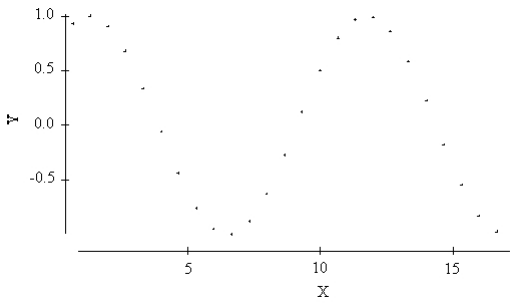
7. In 1998 the World Health Organization reported the findings of a major study on the quality of blood pressure monitoring around the world. In its report it stated that for Canada the results for diastolic blood pressure (DBP) had a mean of 78 mmHg and a standard deviation of 11 mmHg. Assuming that diastolic blood pressure measurements are Normally distributed, the DBP reading that represents the 80th percentile of the distribution is

   (a) about 85.4.
   (b) about 93.6.
   (c) about 87.2.
   (d) about 86.8.

8. 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.

9. The researcher notices that there is a distinct curved pattern in the plot. It would be appropriate to conclude:



   (a) r is small.

   (b) r is approximately $-2/3$ since Y decreases as X increases in approximately $2/3$ of the plot.

   (c) A decreasing straight line would describe the relationship between X and Y quite well.

   (d) r is meaningless here.

10. I wish to determine the correlation between the height (in inches) and weight (in pounds) of 21–year–old males. To do this I measure the height and weight of two 21–year–old men. The measured values are

	Male #1	Male #2
Height	70	69
Weight	160	164

The correlation r computed from the measurements on these two males is

   (a) 1.0.

   (b) -1.0.

   (c) Near 0, since the heights and weights of the men are similar.

   (d) Can't say without doing the actual calculations for r.

11. Which of the following is likely to have a mean that is **smaller** than the median?

    (a) The salaries of all National Football League players.

    (b) The scores of students (out of 100 points) on a very easy exam in which most score perfectly, but a few do very poorly.

    (c) The prices of homes in a large city.

    (d) The scores of students (out of 100 points) on a very difficult exam on which most score poorly, but a few do very well.

12. There are three children aged 3, 4, and 5 in a room. If another 4–year–old child enters the room, what will happen to the mean and variance?

    (a) The mean will stay the same but the variance will increase.

    (b) The mean will stay the same but the variance will decrease.

    (c) The mean and variance will both stay the same.

    (d) The mean and variance will both decrease.

Use the following to answer questions 13, 14, and 15:

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.

13. For $x_1 = 2, x_2 = 5, x_3 = -3, x_4 = -5, x_5 = -1, x_6 = 10$, and $n = 6$, the sum

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

equals

(a) -4.
(b) -1.
(c) 3.
(d) 8.

14. For $x_1 = 2, x_2 = 5, x_3 = -3, x_4 = -5, x_5 = -1, x_6 = 10$, and $n = 6$, the sum

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

equals

(a) -4.
(b) -1.
(c) 3.
(d) 8.

15. For $x_1 = 2, x_2 = 5, x_3 = -3, x_4 = -5, x_5 = -1, x_6 = 10$, and $n = 6$, the sum

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

equals

(a) 2.
(b) 8.
(c) 10.
(d) 12.

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

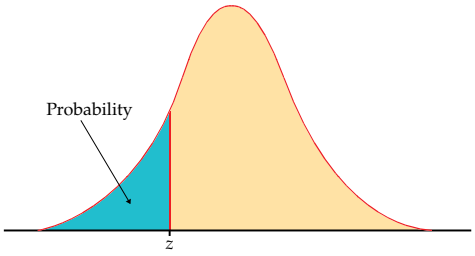


TABLE A

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.

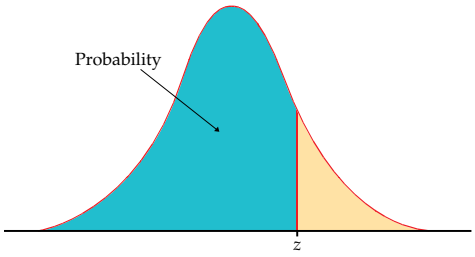


TABLE A

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