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

Friday, October 1, 2010

Part I: Text Answers

Your Name: _____

based on: HW 3, Exercise 2, Q 1.126 - 1.131

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

& Stat 2000, Fall 2009, Midterm 1, Q1

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!

- 2 for each calculation error (or no final result)

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?

Paul's z-score $= \frac{800-1022}{212} = -1.05$ (5)

$\rightsquigarrow$ the equivalent ACT score is

$20.7 + (-1.05) \cdot 4.7 = \underline{\underline{15.77}}$ (5) ($\sim$ 16 as ACT scores are integers)

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.

Liza's z-score $= \frac{1380-1022}{212} = \underline{1.69}$ (4)

John's z-score $= \frac{24-20.7}{4.7} = \underline{0.70}$ (4)

$\rightsquigarrow$ Liza (2) has the higher score

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?

To be in the top-10%, we need an area of 0.90 (90%) ③ to the left;

this is the case for $z \approx 1.28$ ③ (working with 1.29 is fine)

$\rightsquigarrow$ her SAT score should be at least

$1022 + 1.28 \cdot 212$ ④ $= \underline{1293.4}$ (~ 1293)

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

Mike's z-score $= \dfrac{29 - 20.7}{4.7}$ ⑤ $= 1.77$

area to the right of 1.77 = area to the left of -1.77

$= 0.0384 = \underline{3.84\%}$ ⑤

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?

Melissa's z-score $= \dfrac{1300 - 1022}{212}$ ④ $= 1.31$

area to the left of 1.31: 0.9049 ④

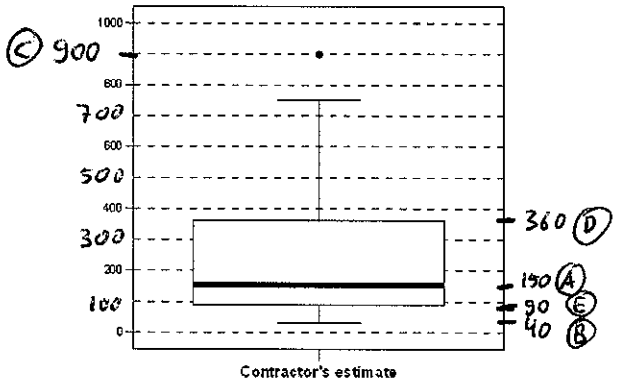
$\rightsquigarrow$ this is roughly the $\underline{90^{th}}$ ② percentile

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:

+7 for answer in acceptable interval

+1 for answer outside acceptable interval



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: 150 days (A) [130 to 170 OK]
2. (**7 Points**) What is (approximately) the estimated minimum number of days?
   Answer: 40 days (B) [20 to 60 OK]
3. (**7 Points**) What is (approximately) the estimated maximum number of days?
   Answer: 900 days (C) [880 to 920 OK]
4. (**7 Points**) What is (approximately) the estimated third quartile (Q3)?
   Answer: 360 days (D) [340 to 380 OK]
5. (**7 Points**) What is (approximately) the interquartile range (IQR)?
   Answer: 270 days (D) (E) [250 to 290 OK; IQR = Q3 − Q1 = 360 − 90 = 270]
6. (**7 Points**) What is (approximately) the percentage of contractors that estimated the number of days to be more than 100?
   Answer: 70 % [65% to 75% OK; 75% are above Q1 which is about 90; above 100 is a slightly smaller percentage]
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.

(i) [the data are skewed to the right; in particular, there is a high outlier of 900; this will considerably effect the mean]

	avg	SD	
x: ACT	21	5	$r = 0.82$
y: SAT	912	180	

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!

-2 for each calculation error (or no final result)
-2 if x, y flipped

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

slope: $b_1 = r \cdot \frac{s_y}{s_x} = 0.82 \cdot \frac{180}{5} = 29.52$ (8)

intercept: $b_0 = \bar{y} - b_1 \bar{x} = 912 - 29.52 \cdot 21 = 292.08$ (8)

regression equation: $\hat{y} = 292.08 + 29.52 \cdot x$ (4)

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

$\hat{y}$ (for 23) $= 292.08 + 29.52 \cdot 23 = \underline{971.04}$ (~ 971), (10)

[i.e., the estimated SAT score would be about 971 for someone who had an ACT score of 23]

Note: Statistics is not about punching numbers in a calculator, but about interpreting data. A result that makes no sense, but is left without interpretation, results in a "−4" point deduction. As we are not extrapolating here (23 is in the middle of the point cloud), the predicted SAT score should be within $912 \pm 3 \cdot 180 = 372$ to 1452. Moreover, from Q1, we know that SAT scores range from 400 to 1,600. When you get an implausible result (such as −340 or 6,745), you must comment that something went wrong! 4

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	○	●	○	○					

based on:

- 5 questions based on "Online Quizzes" for Chapters 1 & 2
- 5 questions based on previous versions of Stat 2000, Midterm 1 (Fall 2009 & Spring 2010)

Stat 2000, Midterm 1, Question 4 — Solutions

1. (c) We obtain the correct answer by the process of elimination: For 25 observations, the median is the 13th observation, the first quartile (Q1) is the mean of the 6th and 7th observation, and the third quartile (Q3) is the mean of the 19th and 20th observation. Now, notice from the histogram on top that the 13th observation falls into the 30 to 35 class — thus the median can't be less than 30. The 6th and 7th observations fall into the 20 to 25 class and the 19th and 20th observations fall into the 35 to 40 class. As the IQR is calculated as Q3 − Q1, this difference can be at most $40 - 20 = 20$ — thus the IQR cannot exceed 30.

2. (d) The IQR measures the spread of the middle half (the distance between Q1 and Q3) so any outliers will not affect this measure.

3. (c) We lose all information about trends over time as we ignore the time information entirely, but we gain additional information how many observations fall into each class and thus we are able to easily determine the number of years for which the death rate was 5 or higher.

4. (a) We cannot do any arithmetic with this variable, but with all other variables.

5. (a) Females always mate with males that are .75 years younger, so the relationship is perfectly linear, and as females get older so will the males they mate with, so the correlation is 1.

6. (a) The closer the points in a scatterplot are lying along a straight line, the closer the correlation is to $+1$ or -1. If the straight line has a positive slope (slopes up from left to right), the correlation is positive, and if the straight line has a negative slope (slopes down from left to right), the correlation is negative. In this case, the points in the plot fall roughly along a line that slopes up from left to right, so we expect the correlation to be positive and have a value closer to 1 (i.e., 0.95) than to 0 (i.e., 0.10).

7. (c) $z = 0.84$ (from Table A) is the 80th percentile for a standard normal curve. Adjust this to $0.84 \cdot 11 + 78 = 87.24$ as a DBP value.

8. (c) We are dealing with one quantitative variable (change in blood pressure) and one categorical variable (type of music). (a) is impossible because we need two quantitative variables for a scatterplot. When we create a single histogram as suggested in (b), we lose the information about the type of music. (c) makes most sense as each boxplot summarizes the data for each type of music and we can use these boxplots to explore the relationship between type of music and change in blood pressure.

9. (d) r measures the strength of the linear relationship between Y and X. Since the plot shows a relationsship that is clearly not linear, r has no meaning here.

10. (b) Since two points always lie exactly on a straight line the correlation must be either 1.0 or -1.0 (or 0.0 if the x or y values are *exactly* the same — but this is not the case here). So, r is -1.0, since the larger height goes with the smaller weight. The line connecting these two points has a negative slope.

11. (b) For the mean to be smaller than the median, we need a long left tail, i.e., some unusually small values. This will be the case in an exam where most score perfectly while a few do very poorly.

12. (b) As 4 is the mean of the original three observations (3, 4, and 5), an additional value of 4 won't change the mean. Each additional value that equals the mean that gets added to any data set will reduce the variance (as we now devide the sum that remains the same by a larger number of observations). If this explanation makes no sense, directly calculate the variance for both scenarios.

13. (a) Calculate $5 + (-3) + (-5) + (-1) = -4$.

14. (c) Resort the date from smallest to largest:

    ```
    -5    -3    -1    2    5    10
    ```

    Then calculate $-3 + (-1) + 2 + 5 = 3$.

15. (d) We are summing up 6 times the same value (2), so $2+2+2+2+2+2 = 6 \cdot 2 = 12$.