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

Friday, September 25, 2009

Your Name: _____

based on: HW 3, Exercise 2, Q 1.126 – 1.130

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 $\mu = 20.8$ and standard deviation $\sigma = 4.8$. 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 $\mu = 1026$ and standard deviation $\sigma = 209$.
Show your work!

– 2 for each calculation error

1. (**10 Points**) Compare a SAT with an ACT score: Wendy scores 1350 on the SAT. Jeremy scores 25 on the ACT. Assuming that both tests measure the same thing, who has the higher score — Wendy or Jeremy? Report the z–scores for both students.

Wendy's z-score = $\frac{1350 - 1026}{209} = 1.55$ (4)

Jeremy's z-score = $\frac{25 - 20.8}{4.8} = 0.875$ (4)

⇝ Wendy (2) has the higher score

2. (**10 Points**) Find the ACT equivalent: Rob scores 1420 on the SAT. Assuming that both tests measure the same thing, what score on the ACT is equivalent to Rob's SAT score?

Rob's z-score = $\frac{1420 - 1026}{209} = 1.885$ (5)

⇝ the equivalent ACT score is

$20.8 + 1.885 \cdot 4.8 \overset{(5)}{=} 29.85$ (~ 30 as ACT scores are integers)

3. **(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. Jessica scores 880 on the SAT. What is her percentile?

Jessica's z-score $= \frac{880 - 1026}{209} = -0.698 \approx -0.70$ (4)

area to the left of -0.70: 0.2420 (4)

$\rightsquigarrow$ this is roughly the 24th (2) percentile

4. **(10 Points)** Top scores: Allen scores 27 on the ACT. What is the percentage of ACT scores that is higher than his score?

Allen's z-score $= \frac{27 - 20.8}{4.8} = 1.29$ (5)

area to the right of 1.29 = area to the left of -1.29 $= 0.0985 = 9.85\%$ (5)

5. **(10 Points)** Top percentage: Melissa is hoping to qualify for a scholarship. Those are only awarded to applicants who have a SAT score among the top–5% of all SAT scores. Which score does she need at least to be sure to be awarded a scholarship?

to be in the top-5%, we need an area of 0.95 (95%) to the left; (3)

this is the case for $z \approx 1.645$ (3) (working with 1.64 or 1.65 is fine)

$\rightsquigarrow$ her SAT score should be

$1026 + 1.645 \cdot 209$ (4) $= 1369.8$ (~ 1370)

from: HW 2, Exercise 1, Q 1.34 (a)

Question 2: Histograms (**40 Points**)

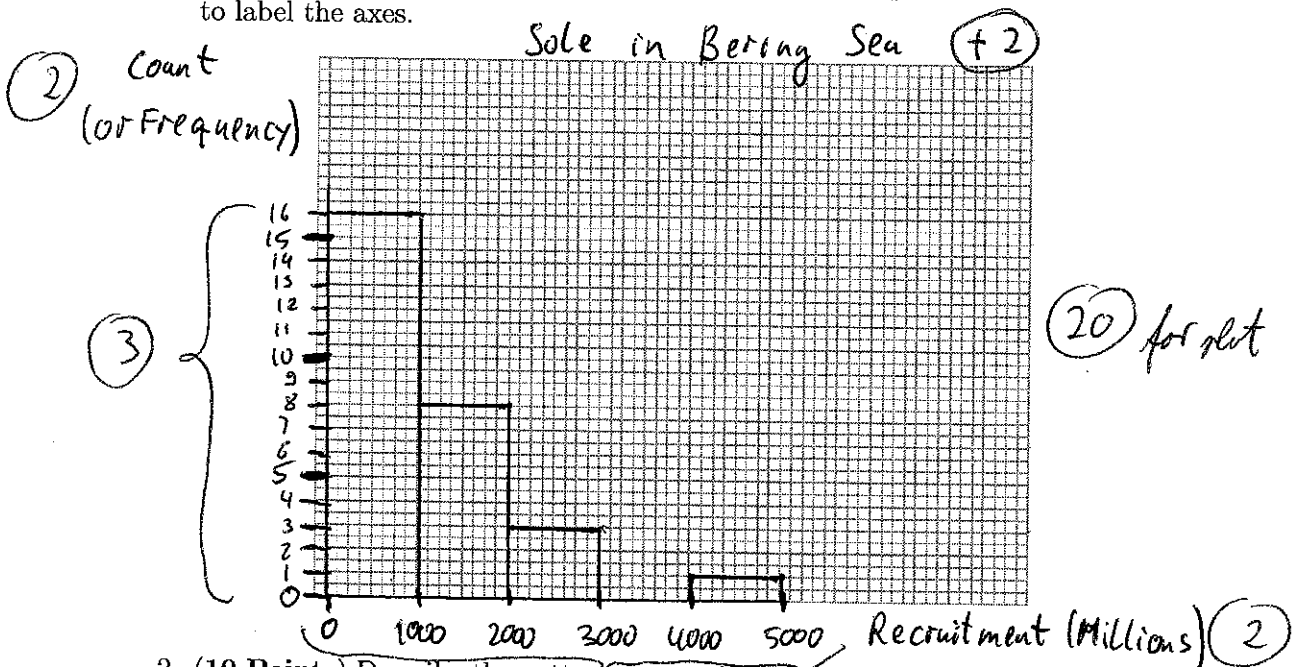
Fish in the Bering Sea. "Recruitment", the addition of new members to a fish population, is an important measure of the health of ocean ecosystems. Here are data on the recruitment of rock sole in the Bering Sea between 1973 and 2000:

Year	Recruitment (millions)	Year	Recruitment (millions)
1973	173	1987	4700
1974	234	1988	1702
1975	616	1989	1119
1976	344	1990	2407
1977	515	1991	1049
1978	576	1992	505
1979	727	1993	998
1980	1411	1994	505
1981	1431	1995	304
1982	1250	1996	425
1983	2246	1997	214
1984	1793	1998	385
1985	1793	1999	445
1986	2809	2000	676

Class	count	#
0- <1000	𝍸 𝍸 𝍸 |	16
1000- <2000	𝍸 |||	8
2000- <3000	|||	3
3000- <4000		0
4000- <5000	|	1
		28

+3

1. (**30 Points**) Draw a histogram to display the distribution of rock sole recruitment. You should work with 5 classes. Use the graph paper provided below. Make sure to label the axes.



2. (**10 Points**) Describe the pattern you see in your histogram and mention any striking deviations that you see. (3)

- the distribution is skewed to the right (5)
- there is a high outlier (4,700 millions) (5)

[- other members range from about 100 million to 2,800 millions] (+3)

3

Question 3: CrunchIt Output (**50 Points**)

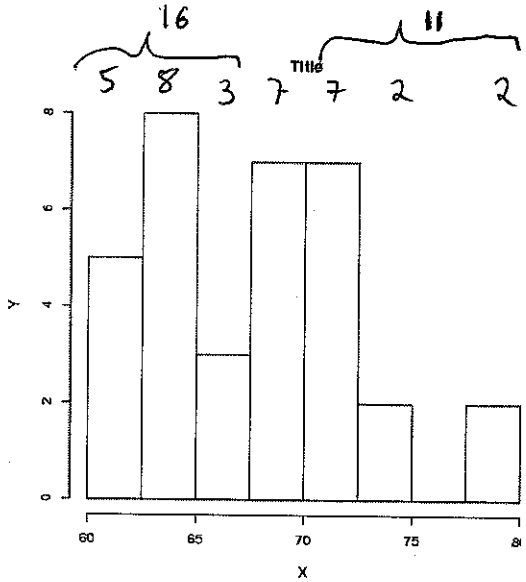
Shown below is output from CrunchIt based on Height and Weight of our Stat 2000 student data. Only 34 complete records have been considered.

Columns -- Selected Field: WEIGHT

Statistic	Result
n	34.0000
mean	159.8824
variance	1225.6827
std.dev	35.0098
std.err	6.0041
median	162.5000
range	180.0000
min	105.0000
max	285.0000
q1	130.0000
q3	180.0000
mode	no unique mode

Columns -- Selected Field: HEIGHT

Statistic	Result
n	34.0000
mean	67.9853
variance	21.8861
std.dev	4.6783
std.err	0.8023
median	68.0000
range	19.0000
min	61.0000
max	80.0000
q1	64.0000
q3	71.6250
mode	no unique mode



Linear Regression -- Select Fields: WEIGHT, HEIGHT

Estimates	Estimate	t value	Pr(>|t|)	Std. Error	
WEIGHT	0.0898	5.1302	0.0000	0.0175	
(Intercept)	53.6328	18.7397	8.1611e-19	2.8620	
ANOVA	Df	Sum Sq	Mean Sq	F value	Pr(>F)
WEIGHT	1.0000	325.9466	325.9466	26.3194	0.0000
Residuals	32.0000	396.2961	12.3843	NaN	NaN
R.sq	0.4513				
adjusted.R.sq	0.4342				
F.value	26.3194				

Correlation -- Select Fields: WEIGHT, HEIGHT

	WEIGHT	HEIGHT
WEIGHT	1.0000	0.6718
HEIGHT	0.6718	1.0000

Answer the questions on the next page based on the output provided above.

1. **(10 Points)** List the values for the five number summary for Height and clearly name them (e.g., if variance is one of these numbers than indicate "variance: ..."). ① each

1.) Minimum : 61
2.) First Quartile (Q_1) : 64
3.) Median (m) : 68
4.) Third Quartile (Q_3) : 71.625
5.) Maximum : 80

−1 for different order

2. **(10 Points)** Read off the values (as stated in the CrunchIt output) that allow to predict Height (response) from Weight (explanatory) and combine them in the regression equation: ⑤ ⑤

predicted Height = 53.6328 + 0.0898 * Weight

3. **(15 Points)** Manually calculate the values that allow to predict Weight (response) from Height (explanatory).

Calculation: Height (x) : $\bar{x} = 67.9853$, $s_x = 4.6783$ correlation $r = 0.6718$
Weight (y) : $\bar{y} = 159.8824$, $s_y = 35.0098$

slope: $b_1 = r \cdot \frac{s_y}{s_x} = 0.6718 \cdot \frac{35.0098}{4.6783} = 5.027$ ⑥ | intercept: $b_0 = \bar{y} - b_1 \bar{x}$ ⑤
$= 159.8824 - 5.027 \cdot 67.9853$
$= -181.88$

Then combine them in the regression equation:

predicted Weight = −181.88 ② + 5.027 ② * Height

4. **(5 Points)** Predict the weight for someone who is 72 inches tall. Indicate which equation you use (either from 2. or 3.) and write down your calculation and the final result. ② ③

$\hat{y}$ (for 72") $= -181.88 + 5.027 \cdot 72 = 180.064$

5. **(10 Points)** One histogram is shown. Unfortunately, it is not labeled. Based on all other information, does this histogram relate to (i) Weight, (ii) Height, or (iii) none of these two variables? Circle your answer and explain. ⑥

- Based on the histogram, the median must be between 67.5 to 70; for Height, the median is 68.
- We have data for men & women, with men generally taller than women; therefore, we should expect to see two peaks in the histogram! ④ for valid explanation

- The horizontal axis is scaled from 60 to 80; this matches min = 61 and max = 80 for Height.

from: HW 4, Exercise 3: 10 (out of 15) questions directly taken from "Online Quizzes" for Chapters 1 & 2

Stat 2000, Midterm 1, Question 4 – Solutions

1. (c) The median is the average of the two observations in the middle of the ordered data, i.e., (19 + 20) / 2 = 19.5.

2. (d) is correct. The last number in the five number summary is the maximum and this will be increased from 92 to 95. Neither the quartiles, nor the median, nor the minimum will be affected by increasing the highest score.

3. (b) Area in square inches is a number - we can do arithmetic such as computing the average area of the windows ordered.

4. (a) To display a relationship between a categorical explanatory variable (in this case, gender) and a quantitative response (in this case, amount spent), we make a side-by-side comparison of the distributions of the response for each category. Boxplots are one way to display distributions of quantitative variables.

5. (d) Every other statement is incorrect.

6. (b) The simplest way is to first compute the total of the 50 salaries and then divide by 50. The total salary of the 20 female workers is (20)($43,000) = $860,000 and the total salary of the 30 male workers is (30)($47,000) = $1,410,000. The total of all the salaries is thus $2,270,000 and the average salary is $2,270,000 / 50 = $45,400.

7. (a) The effect of adding $3,000 to each observation is to increase the first quartile by $3,000 and to increase the third quartile by $3,000. The difference, which is the interquartile range, is thus unchanged (the $3,000s cancel).

8. (d) If we use the line for her, we find GPA = -6 + 0.15 * 200 = 24. But GPA is only on a 12-point scale, so this is meaningless. We have been doing an extrapolation - the line was not created from data on geniuses like her.

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

10. (c) It is y = 10 + 0.9 * 90 = 91.

11. (b) The correlation r measures association between two quantitative variables and both variables are quantitative in case (b), but not in case (a). We would expect the correlation to be positive, but it cannot be greater than 1.0 as in (d) and it does not contain any units as in (c).

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

13. (a) The standard deviation measures the spread of the scores. If all the scores were the same, there is no spread, so the standard deviation will be 0.

14. (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.

15. (d) We know that slope $b1 = r * sy / sx$. As sy and sx cannot be negative, b1 and r must have the same sign, i.e., be positive here. But, because sy and sx are not provided, we cannot calculate the exact value of r here.