Statistics 21
Fall 2002

Mr. Purves

Midterm I

Print your name ____________________

Sign your name ____________________

TA's Name I ____________________

Lab time M1 ____________________

To get full credit, you must give reasons and/or show work.

Score:

1: 5
2: 5
3: 10
4: 5
5: 10
6: 10
7: 4
Total 49

(5) 1. The quotation below comes from an article in the New York Times of Tuesday, September 24, 2002.

TRAUMA

Putting a Premium on Helmets

Motorcyclists who do not wear helmets have much higher hospital costs if they crash than those who wear them, a new study has found.

The findings, the authors say, suggest a public health motive for states to require riders to wear helmets. States that allow motorcyclists to ride unprotected should require them to pay higher insurance premiums, the study says.

The study, reported in the current issue of The Journal of Trauma, examined the cases of 218 motorcyclists admitted to the University of Michigan Health System from 1996 to 2000.

Although Michigan requires riders to wear helmets, 19 percent of the injured motorcyclists had not worn them, the researchers found. The helmetless riders were also less likely to be insured, the study found.

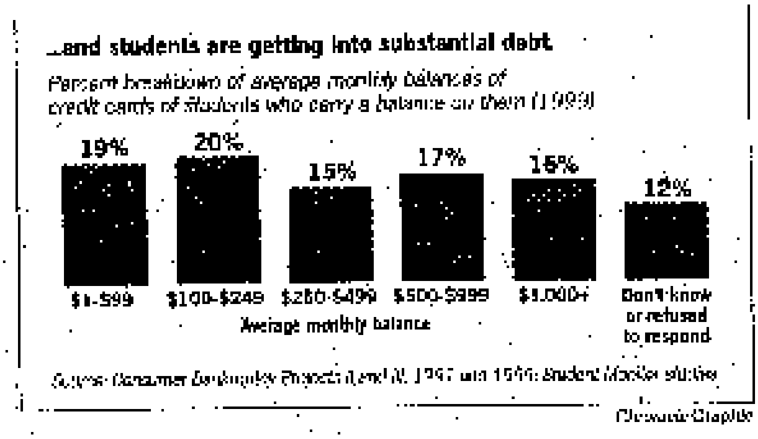
The hospital costs for those not wearing helmets averaged $37,317 — more than $8,000 higher than the average for patients who had worn helmets. The cost of in-patient treatment at rehabilitation centers was also higher for those who had not worn helmets.

True or false, and explain: From the study (at least as described in the article), it would be correct to conclude that the 19% who had not worn helmets incurred the higher hospital costs because they lacked the protection of a helmet.

5

False. Since this is an observational study, not a randomized controlled experiment, there could be confounding factors at work. ~~that could~~ People who ride without helmets could differ from helmeted riders in other crucial ways. I would assume that those who ride without helmets would tend to have riskier behavior and ~~~~ drive faster in general. ~~Thus.~~ Higher speed accidents would lead to higher hospital costs with or without a helmet.

(5) 2. Consumer organizations are concerned about the amount of debt University students carry on their credit cards. An article in the San Francisco Chronicle of December 23, 2001, presented the following graph:



True or false, and explain: The graph shows that if you compare a smaller average balance and a larger one, for example $150 and $750, there is not that much difference between the percentage of students carrying the smaller amount and the percentage carrying the larger one. both

5

False. The graph shows that there is not much difference between the percent of students with $100-$249 in debt and those with $500-$999 in debt. However, this graph is not a histogram. The density of those with $150 in debt is far greater than those with $750 since the interval from $100-$249 is only an interval of $150 (indicating an average density of $\frac{20}{150}$ = 0.13% per dollar) whereas the interval from $500-$999 is an interval of $500 (indicating an average density of $\frac{17}{500}$ = 0.034% per dollar.

(10) 3. At the end of the first month of a six month training program, the participants were given a written examination. The average score (out of 100) was 62 and the SD was 11.8. The scores followed the normal curve closely.

(a) About what percent of the scores were between 75 and 85?

$\frac{75-62}{11.8} = +1.1\,SDs \quad \frac{85-62}{11.8} = +1.95\,SDs$

This percent is the % of scores between +1.1 and +1.95 SDs

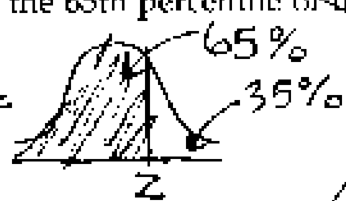
$= \frac{1}{2}\Big($  $\Big)$

$= \frac{1}{2}(94.88 - 72.87)$

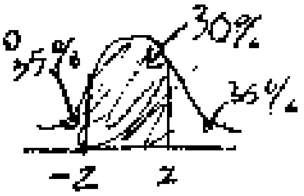
$\boxed{= 11.0\%}$

4

(b) Find, approximately, the 65th percentile of the scores.

65th percentile = 

6

equivalent to 

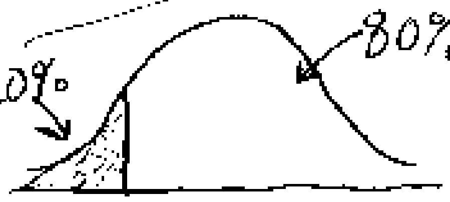
z = 0.39 (interpolating between 0.35 and 0.40)

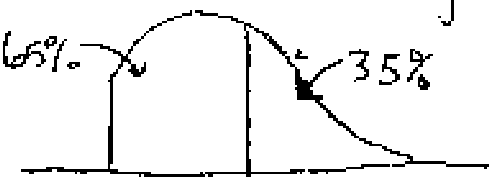
the 65th percentile is at +0.39 SDs

0.39 SDs · 11.8 = 4.6 points above average

4.6 + 62 = $\boxed{66.6 \text{ points}}$

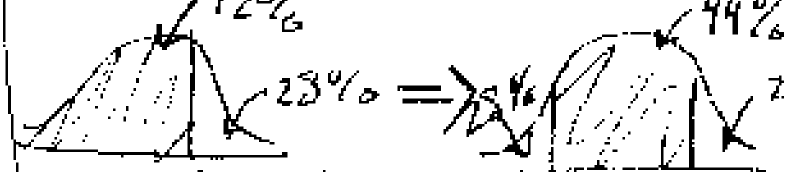
(5) 4. (See the preceding problem.) Anyone who scored in the bottom 20% on the written examination was dropped from the program. Find, approximately, the 65th percentile of the scores of the people remaining in the program.

A)


$\Rightarrow$ B)
new class score graph


However the 65th percentile on graph B is not the 65th percentile on A) since B is the upper 80% of A.

top 35% on B) is $0.8 \cdot 35 = 28\%$ top 28% / 72nd percentile on A). So I need to find the 72nd percentile of A



$z = 0.595$ SDs by interpolating between 0.55 and 0.6

looking for 0.59 SDs above average

$0.59 \cdot 11.8 = 7$ pts above average

$7 + 62 = \boxed{69 \text{ points}}$ ✓

7

(10) 5. (a) Find the correlation coefficient between x and y for the data below.

x	y
1	3
2	3
2	3
2	4
3	4
3	5
3	6
8	4

x (standard units)	y (standard units)	x · y (standard units)
−1	−1	1
−.5	−1	0.5
−.5	−1	0.5
−.5	0	0
0	0	0
0	1	0
0	2	0
2.5	0	0

sum = 2

$av(x) = \frac{1+3(2)+3(3)+8}{8} = 3$

$av(y) = \frac{2(3)+3(4)+5+6}{8} = 4$

$r(x,y) = \frac{2}{8}$

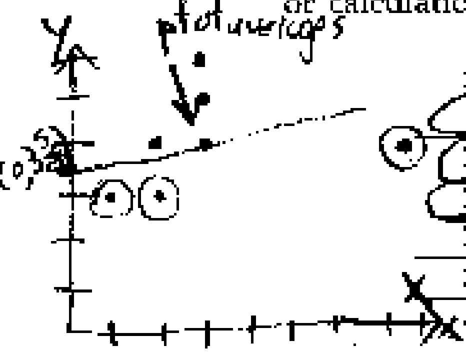
$SD(x) = \sqrt{\frac{(1-3)^2+3(2-3)^2+3(3-3)^2+(8-3)^2}{8}} = \sqrt{\frac{4+3+0+25}{8}} = 2$

$\boxed{r(x,y) = 0.25}$

$SD(y) = \sqrt{\frac{3(3-4)^2+3(4-4)^2+(5-4)^2+(6-4)^2}{8}} = \sqrt{\frac{3+0+1+4}{8}} = 1$

3

(b) In the table below, circle the points on the scatter diagram that lie below the regression line for predicting y from x. Explain your choices by reasoning and/or calculation.



x	y
1	3
2	3
2	3
2	4
3	4
3	5
3	6
8	4

I drew the regression line for predicting y from x to the left and plotted the point unit. You can then see visually which points lie below the regression line.

Slope of regression line to predict y from x $= \frac{r(x,y)SD(y)}{SD(x)} = \frac{0.25 \cdot 1}{2} = 0.125$

(10) 6. A growth study for men led to the following results:

average height at age 6 = 46 inches, SD = 1.7 inches
average height at age 18 = 70 inches, SD = 2.5 inches
correlation coefficient = 0.85

The scatter diagram is football shaped.

One man in the study was 48 inches tall at age 6 and 71 inches tall at age 18.

(a) Use the regression method to predict his height at 18 from his height at 6.

at age 6

$\frac{48-46}{1.7} = 1.176$ SDs above average

so $1.176 \cdot r = 1.176 \cdot 0.85 = 1$ SD above average at age 18

$1 \cdot 2.5 = 2.5''$ above average at age 18

$2.5 + 70 = \boxed{72.5''}$

(b) The prediction in (a) is off by 1.5 inches. (error: actual - predicted = -1.5")

(c) Use the regression method to predict his height at 6 from his height at 18.

at age 18

$\frac{71-70}{2.5} = 0.4$ SDs above average

$0.4 \cdot r = 0.4 \cdot 0.85 = 0.34$ SDs above average at age 6

$0.34 \cdot 1.7 = .578''$ above average at age 6

$46 + 0.578 \boxed{= 46.578''}$

48-46.578 (error: actual - predicted = 1.422")

(d) The prediction in (d) is off by 1.422 inches.

10

(5) 7. A large class has one midterm and a final examination. The scatter diagram for the two exams was football shaped and showed a moderate positive association.

Two percentiles are described below:

(i) The 90th percentile of the final examination scores for the whole class.

(ii) The 90th percentile of the final examination scores for those in the class who scored at the 50th percentile on the midterm.

It would be reasonable to guess that

____ (i) is equal to (ii).

____ (i) is smaller than (ii).

_✓_ (i) is larger than (ii).

Or

____ there is not enough information to make a reasonable guess.

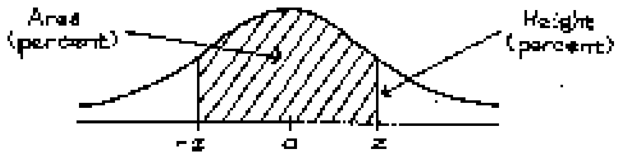
Check (√) one of the four options above, and then back up your choice with reasoning and or calculation.

Those students whose scores were at the 50th percentile on the midterm will still average 50th percentile on the final like the class as a whole. However, assuming that the correlation between midterm and final exam scores is not zero, the group that scored at 50th percentile on the midterm will score closer together around the average on the final than the class as a whole (ie, the SD for the group is less than the SD for the class)*. Since the 50th percentile group's SDs do not "reach" as far as the classes SDs but the mean for both groups is the same, the 90th percentile rank will be lower for those who scored at the 50th on the midterm than the class as a whole.

usually a safe assumption in a case like this

* The SD for the 50th percentile group is $\sqrt{1-r^2} \cdot$ class SD where r = corr between midterm & final scores

Table



A NORMAL TABLE

z	Height	Area	z	Height	Area	z	Height	Area
0.00	39.89	0	1.50	12.95	86.64	3.00	0.443	99.730
0.05	39.84	3.99	1.55	12.00	87.89	3.05	0.381	99.771
0.10	39.69	7.97	1.60	11.09	89.04	3.10	0.327	99.806
0.15	39.45	11.92	1.65	10.23	90.11	3.15	0.279	99.837
0.20	39.10	15.85	1.70	9.40	91.09	3.20	0.238	99.863
0.25	38.67	19.74	1.75	8.63	91.99	3.25	0.203	99.885
0.30	38.14	23.58	1.80	7.90	92.81	3.30	0.172	99.903
0.35	37.52	27.37	1.85	7.21	93.57	3.35	0.146	99.919
0.40	36.83	31.08	1.90	6.56	94.26	3.40	0.123	99.933
0.45	36.05	34.73	1.95	5.96	94.88	3.45	0.104	99.944
0.50	35.21	38.29	2.00	5.40	95.45	3.50	0.087	99.953
0.55	34.29	41.77	2.05	4.88	95.96	3.55	0.073	99.961
0.60	33.32	45.15	2.10	4.40	96.43	3.60	0.061	99.968
0.65	32.30	48.43	2.15	3.96	96.84	3.65	0.051	99.974
0.70	31.23	51.61	2.20	3.55	97.22	3.70	0.042	99.978
0.75	30.11	54.67	2.25	3.17	97.56	3.75	0.035	99.982
0.80	28.97	57.63	2.30	2.83	97.86	3.80	0.029	99.986
0.85	27.80	60.47	2.35	2.52	98.12	3.85	0.024	99.988
0.90	26.61	63.19	2.40	2.24	98.36	3.90	0.020	99.990
0.95	25.41	65.79	2.45	[illegible]	[illegible]	3.95	0.016	99.992
1.00	24.20	68.27	2.50	1.75	98.76	4.00	0.013	99.9937
1.05	22.99	70.63	2.55	1.54	98.92	4.05	0.011	99.9949
1.10	21.79	72.87	2.60	1.36	99.07	4.10	0.009	99.9959
1.15	20.59	74.99	2.65	1.19	99.20	4.15	0.007	99.9967
1.20	19.42	76.99	2.70	1.04	99.31	4.20	0.006	99.9973
1.25	18.26	78.87	2.75	0.91	99.40	4.25	0.005	99.9979
1.30	17.14	80.64	2.80	0.79	99.49	4.30	0.004	99.9983
1.35	16.04	82.30	2.85	0.69	99.56	4.35	0.003	99.9986
1.40	14.97	83.85	2.90	0.60	99.63	4.40	0.002	99.9989
1.45	13.94	85.29	2.95	0.51	99.68	4.45	0.002	99.9991