MATH 243 (PHILLIPS, SPRING 2008): MIDTERM 1 (WHITE) SOLUTIONS

Multiple choice: 3 points/part; 21 points total. Circle the letter of the best answer.

MC1. As part of an investigation of the mean Math SAT score of University of Oregon students, the Registrar's office selects 100 students from the University of Oregon, and looks up their Math SAT scores in their application materials. What is the sample?

a. The Registrar's office.
b. All University of Oregon students.
[✓ c.] The 100 selected students.
d. The Math SAT scores of the 100 selected students.
e. The Math SAT scores of all University of Oregon students.
f. Impossible to tell from the information given.
g. None of the above.

MC2. Wang's Widgets Inc. has 20 employees. The mean of their annual salaries is \$50,000, the median is \$40,000, the standard deviation of the annual salaries is \$24,000, and the range of the annual salaries is \$75,000. The total payroll of Wang's Widgets Inc. is:

a. \$1,500,000
[✓ b.] \$1,000,000
c. \$800,000
d. \$480,000
e. Impossible to tell from the information given.
f. None of the above.

(Multiply the mean by the number of employees.)

MC3. A researcher wants to study the relationship between color and a certain genetic mutation in Rocky Mountain spotted frogs. He chooses a stream which looks like good frog habitat and collects for study all the frogs he can net on this stream. The frogs he collected form:

a. A population.
b. A simple random sample.
c. A stratified random sample.
[✓ d.] A convenience sample.
e. A voluntary response sample.
f. A systematic random sample.

Date: 2 May 2008.

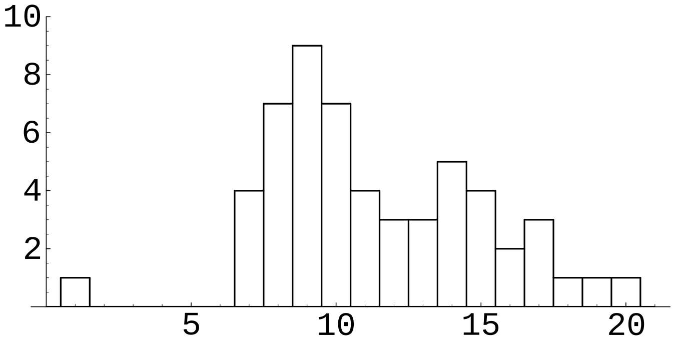
MC4. A magician has a coin which is biased in such a way that the probability of getting heads when the coin is tossed is 1/3. Which of the following statement gives a valid interpretation of this probability?

a. If the coin is tossed a very large number of times, the proportion of the tosses which are heads will be very close to $\underline{\text{1/2.}}$
b. Every third toss of this coin will be heads.
c. In any three consecutive tosses of this coin, $\underline{\text{exactly one}}$ will be heads.
d. In any three consecutive tosses of this coin, $\underline{\text{at least one}}$ will be heads.
$\boxed{\text{e.}}$ If the coin is tossed a very large number of times, the proportion of the tosses which are heads will be very close to $\underline{\text{1/3.}}$
f. The magician is intending to cheat in a game involving coin tosses.

MC5. Two variables in a study are said to be confounded if what?

$\boxed{\text{a.}}$ One cannot separate their effects on a response variable.
b. They are highly correlated.
c. They do not have a normal distribution.
d. One of them is a placebo.
e. Both are lurking variables.
f. The statistician conducting the study is incompetent.

MC6. Consider the following histogram (made from integer data):



The distribution of the data plotted is:

a. Roughly symmetric with $\underline{\text{no}}$ outliers.
b. Roughly symmetric with $\underline{\text{one or more}}$ outliers.
c. Skewed to the $\underline{\text{right}}$ with $\underline{\text{no}}$ outliers.
$\boxed{\text{d.}}$ Skewed to the $\underline{\text{right}}$ with $\underline{\text{one or more}}$ outliers.
e. Skewed to the $\underline{\text{left}}$ with $\underline{\text{no}}$ outliers.
f. Skewed to the $\underline{\text{left}}$ with $\underline{\text{one or more}}$ outliers.

It is skewed in the direction of the long tail.

MC7. Let μ be the mean IQ of children in the Hicksville school system. Using a simple random sample of Hicksville school children, the following 95% confidence interval for the mean μ is computed: $103 \leq \mu \leq 112$. Which of the following statements gives a valid interpretation of this interval?

a. 95% of $\underline{\text{the sample of}}$ Hicksville school children have IQs between 103 and 112.

b. 95% of all Hicksville school children have IQs between 103 and 112.

[✓ c.] If the procedure were repeated many times, approximately 95% of the resulting confidence intervals would contain the mean IQ of all Hicksville school children.

d. If the procedure were repeated many times, approximately 95% of the resulting confidence intervals would contain the mean IQ of the sample of Hicksville school children.

e. The mean IQ of all Hicksville school children is 95.

f. The statistician must be wrong, since the mean IQ of all Hicksville school children is obviously less than 95.

Choices (a) and (b) are not related to the meaning of a confidence interval. If a large sample is used, the confidence interval will contain much less than 95% of the data. Choice (d) is incorrect because *every* confidence interval, not just 95% of them, contains the mean of the sample it was derived from.

1. (3 points/part; 18 points total. Work need not be shown.) The five number summary of the final exam scores in Professor Zhang's Math 315 is:

4 45 60 75 98.

The questions below are about the original data, which consists of one number for each of the 40 students in the class. For each of the following statements, circle "A" if the statement must be true, circle "S" if the statement might or might not be true, and circle "N" if the statement cannot be true. In other words, under the stated circumstances the statement is **A**lways true, **S**ometimes true, or **N**ever true.

A [✓ S] N (a) At least one student received a score of 60.

Since the number of scores is even, the median is half way between the two middle scores. Maybe the 20th best score was 62 and the next one down was 58.

A [✓ S] N (b) The score 4 is an outlier.

The 1.5 · IQR rule is only a test for *suspected* outliers.

[✓ A] S N (c) The median score is 60.

By definition, the middle number of the five number summary is the median.

A S [✓ N] (d) About 30 students had scores of 75 or better.

At least half the students had scores of 60 or worse.

A [✓ S] N (e) The standard deviation of the scores is 15.

You can't determine the standard deviation from the median and quartiles (unless you know the distribution is normal).

$\boxed{\text{A}}$ S N (f) If Professor Zhang multiplies all the scores by 2 (since the final exam counts twice as much as the midterm), the new first quartile will be 90.

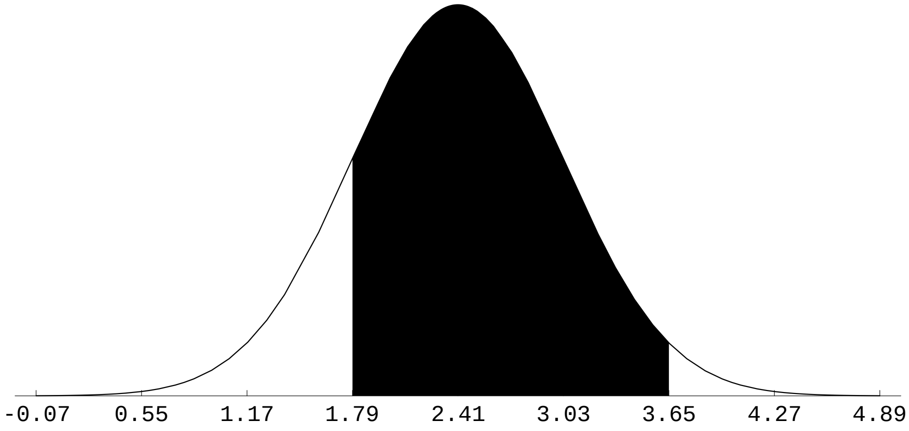
> Doubling all the scores also doubles all the numbers in the five number summary.

2. (4 points/part; total 8 points.) The weights of zucchinis harvested in a certain California valley are approximately normally distributed with mean 2.41 pounds and standard deviation 0.62 pounds. A truckload of zucchinis is taken to market.

In each of the following parts, show your work, and draw a picture of the appropriate normal curve with the relevant points on the horizontal axis clearly marked and with the appropriate area shaded and clearly identified. The curve must look reasonable.

a. Approximately what percentage of the zucchinis weigh between 1.79 and 3.65 pounds? (You must make your method clear and show 4 significant digits.)

Solution: Here is the picture:



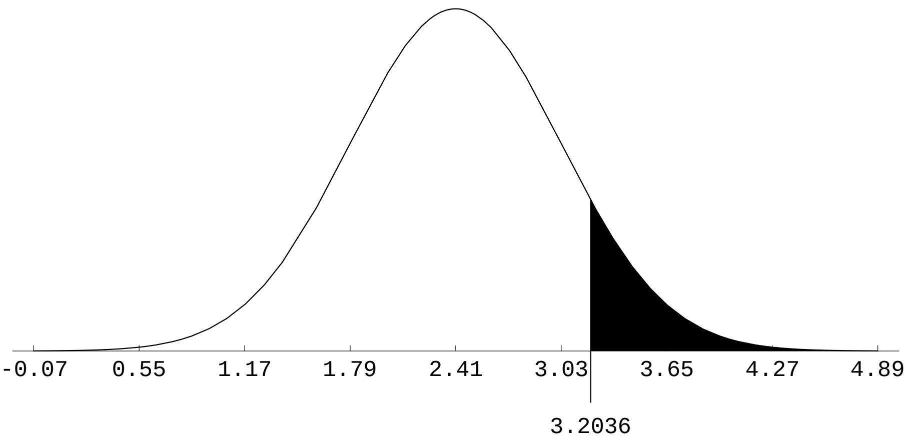
The labels on the horizontal axis are integer multiples of the standard deviation away from the mean. The shaded area corresponds to weights between 1.79 and 3.65 pounds. The rule of thumb is applicable, because 1.79 is one standard deviation below the mean and 3.65 is two standard deviations above the mean. Accordingly, the unshaded area is about $\frac{1}{2}(1 - 0.68) + \frac{1}{2}(1 - 0.95) = 0.185$, and the desired area is about $1 - 0.185 = 0.815$, or 81.5%.

If you use Table A instead, look up $z = 2$ to get 0.9772 for the area to the left of 3.65, and look up $z = -1$ to get 0.1587 for the area to the left of 1.79. Subtract to get 81.91%.

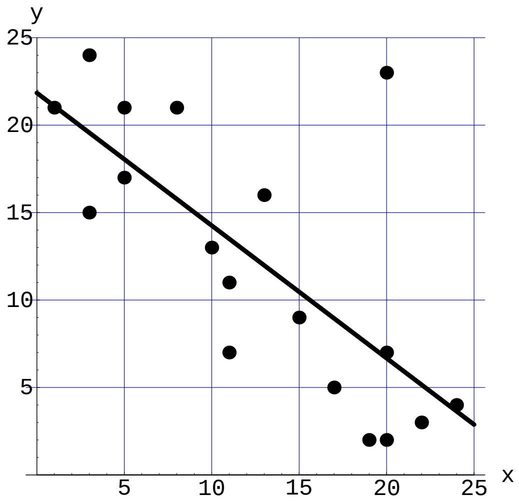
b. Approximately 10% of the zucchinis weigh more than ________ pounds.

Solution: I want the number x of pounds such that 90% of the zucchinis weigh less than x pounds. The closest number to 0.90 which occurs in Table A is 0.8997, for $z = 1.28$. So $x \approx 2.41 + (0.62)(1.28) = 3.2036$.

Here is the picture. The shaded area corresponds to weights above 3.2036 pounds, and has area about 0.1.



3. (4 points/part; total 12 points.) Consider the following scatterplot and regression line.



a. What is the approximate correlation between the variables? Circle one.

2.00 1.00 0.70 0 $\boxed{-0.70}$ -1.00 -2.00

Solution: The correlation is clearly negative and not close to zero, but the points are not so close to the regression line that it can be close to -1 either. (The points plotted actually all have integer coordinates, and the correlation is -0.721937, giving $r^2 \approx 0.521193$.)

b. Give the approximate coordinates of any outliers. If there are none, write "NONE".

Solution: $(20, 23)$. (Note that $(23, 20)$ is not correct.)

c. The residual for the observation with approximate coordinates $(17, 5)$ is closest to: (circle one)

-12 -8 $\boxed{-4}$ 0 4 8 12

Solution: The corresponding point on the regression line is about $(17, 9)$, so the residual, which is the observed value minus the predicted value, is about $5-9 = -4$.

4. (2 points/part; 10 points total.) For a biology project, a student measures the tail length (in centimeters) and the weight (in grams) of each of 12 mice of the same variety. What units of measurement do each of the following have? (Write “none” if appropriate.)

a. The mean of the lengths of the tails.

Solution: Centimeters.

b. The standard deviation of the weights.

Solution: Grams.

c. The z-score of the weight of the seventh mouse.

Solution: No units. (We also accept “standard deviations”.)

d. The correlation between weight and tail length.

Solution: No units.

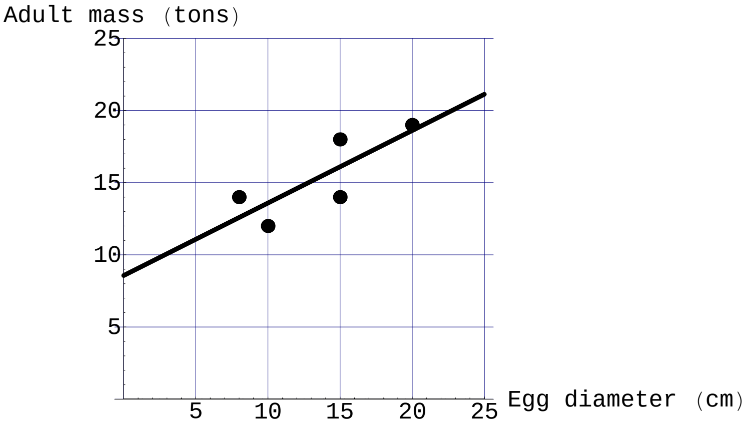
e. The damages awarded in a lawsuit after the student was bitten by a mouse with rabies.

Solution: My best guess is millions of dollars.

5. (4 points/part; total 12 points.) A scientist studying the fire-breathing monsters of the planet Yuggxth wants to know if egg size is a good predictor of adult mass for these creatures. At considerable personal risk, he has managed to obtain the following data on five individuals.

Individual	Egg diameter (cm)	Adult mass (tons)
A	20	19
B	10	12
C	15	18
D	8	14
E	15	14

a. Draw a scatterplot on the axes provided. Be sure to label your axes, and make an appropriate choice of which variable to put on the horizontal axis.



b. Find the equation of the least squares regression line and plot it on the graph above.

Solution: My computer gave $\widehat{y} \approx 8.56951 + 0.502242\,x$. See the graph above for the plot.

c. What is the correlation between egg diameter and adult mass? What percentage of the variation in adult mass is explained by egg diameter?

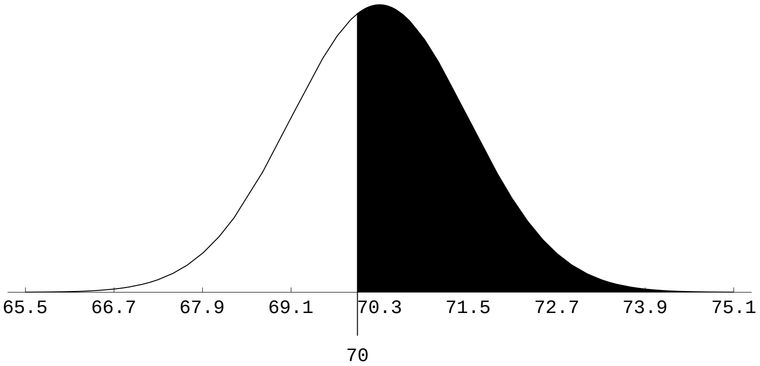
Solution: My computer gave $r \approx 0.799511$. So $r^2 \approx 0.639217$, which means that about 63.92% of the variation in adult mass is explained by egg diameter.

6. (7 points/part) Repeated measurements of the concentration of helium in a gas with a particular scientific instrument vary normally with standard deviation 3.6 milligrams/liter (mg/l) and with mean equal to the true concentration.

a. The true concentration of helium in one gas is 70.3 mg/l. If the concentration is measured 9 times with this instrument, what is the probability that the mean of the measurements is at least 70 mg/l? Illustrate your answer with a graph in which the appropriate areas are labelled.

Solution: The sampling distribution is normal with mean 70.3 mg/l (the same as the mean of the original distribution), and standard deviation $3.6/\sqrt{9} = 1.2$ mg/l. To use Table A, compute the z-score of 70 mg/l in the sampling distribution: $z = (70 - 70.3)/1.2 = -0.25$. Look up in Table A to get about 0.4013. Since we want the probability that $\overline{x} \geq 70$, the correct answer is about $1 - 0.4013 = 0.5987$. (My computer gave about 0.598706.)

Here is the graph. The labels are at integer numbers of sampling standard deviations above and below the mean. The shaded area represents the required probability.



b. For another gas, four measurements of the concentration of helium with this instrument gave $\overline{x} = 47$ mg/l. Construct a 95% confidence interval for the true concentration of helium in this gas. Be sure to show the critical value you used, and how you used it.

Solution: The population standard deviation is known, so we use the one sample z procedure. The confidence interval is $\overline{x} \pm z^*\sigma/\sqrt{n}$. We have $\overline{x} = 47$, $\sigma = 3.6$, $n = 4$, and the appropriate z^* for 95% confidence is 1.960 (from Table C). So the confidence interval is

$$47 \pm (1.960)(3.6)/\sqrt{4} = 47 \pm 3.528,$$

in mg/l.

The confidence interval can also be rewritten as (43.472, 50.528), in mg/l.

7. (5 points.) In 2002, Jane Smith got a score of 690 on the mathematics part of the SAT, and John Doe got a score of 28 on the ACT Assessment mathematics test. SAT math scores in 2002 were normally distributed with mean 516 and standard deviation 114, and ACT math scores in 2002 were normally distributed with mean 20.6 and standard deviation 5.0. Find the standardized scores for both students. Assuming both tests measure the same kind of ability, who did better?

Solution: Jane Smith's z-score is $z_1 = (690 - 516)/114 \approx 1.52632$. John Doe's z-score is $z_2 = (28 - 20.6)/5.0 = 1.48$. Since $1.52632 > 1.48$, Jane Smith is a larger number of standard deviations above the mean, and thus did better.