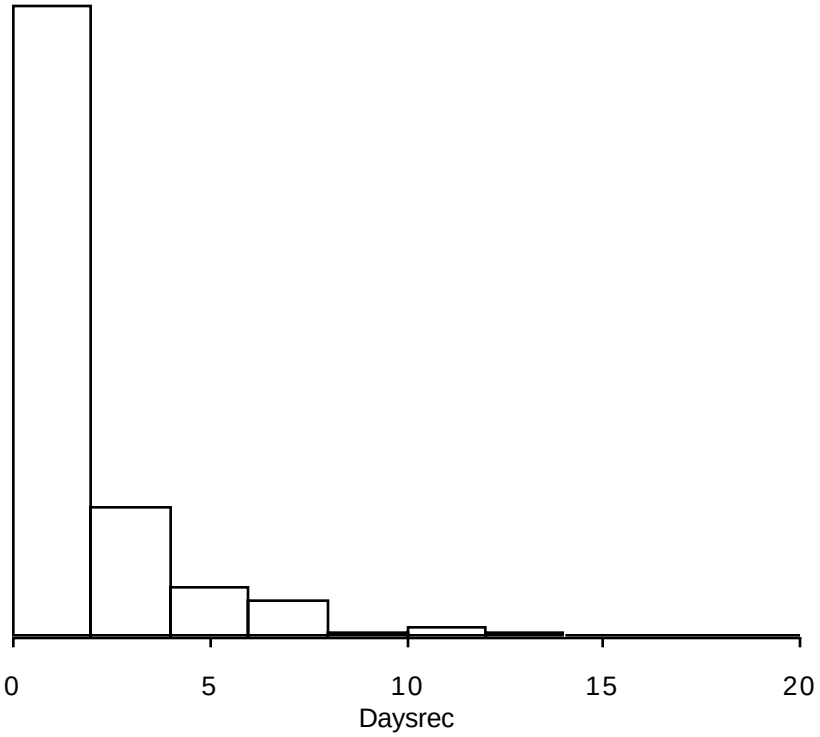
Midterm I
Stats M12

April 30

Name:
ID:
Section:

Instructions: Write neatly. Full credit is given for the correct answer AND a clear, concise explanation. The relative value of each problem (in terms of points) is given in parentheses before each section.



1. For unknown reasons, many dairy cows lie down after giving birth and never get up again. Others recover. A study of 500 dairy cows in New Zealand measured, among other things, the number of days they were recumbant (lying down). The histogram above shows this information.

a) (5) Describe the shape of the histogram. Be brief. (Hint: "brief" means one short sentence.)

Right-skewed. (Full credit.)

It's okay to say other things as long as they are true. And relevant.

Continued on next page

b) (5) The average cow was recumbant for 1.95 days. True or false and explain: the majority of the cows were recumbant for more than 1.95 days.

False. The distribution is right-skewed, and so the average > median. This means less than half are bigger than the average.

c) (5) Will the median be bigger or less than the average? Explain.

It will be less. The distribution is right-skewed, and so the average will be greater than the median.

d) (5) According to the empirical rule, how many cows in this sample should be within one SD of average? Do you think this is the case for this data set? Explain. (There were 500 cows in the sample. The SD for the number of days recumbant is 2.4.)

340 (= .68*500). The empirical rule applies to symmetric distributions, which this is not.

Also full credit if their explanation says that average minus 1SD is negative, and since their can be no negative cows...

2. A newspaper reported on a story with the headline "Tai-chi improves bone density of the senior citizens." The story explained that the researchers examined 50 senior citizens who had enrolled in a tai-chi course (a form of martial arts that emphasizes balance) and compared their bone density with 50 sedentary senior citizens. It found that the bone density of the tai-chi senior citizens was 5% denser than the sedentary subjects. (Note: dense bones are good. One of the curses of age is that bones become increasingly brittle and subject to fracture.)

a) (5) Is this an observational or controlled study? Explain.

Observational because the subjects chose which group they would be long to: tai-chi (treatment) or sedentary (control). For full credit it is not necessary to identify treatment and control groups, but they do exist and students should lose points for saying something contrary to this. Their explanation must make use of the fact that this is an observational study because the subjects

Continued on next page

choose their treatment group. The lack of placebos, confounders, or the lack of randomization are not reasons for concluding that it is an observational study.

(2, continued)
(b) (5) Is the headline correct? If yes, what features of the experimental design allow you to reach this conclusion? If no, describe the shortcomings in the experimental design and give a possible explanation for the outcome that the researchers observed.

The headline is not correct. (Actually, there's no way to tell from this study.) The study does not rule out any of a number of confounding factors that might explain the relationship. For example, active seniors are both more likely to have denser bones and to take a tai-chi class.

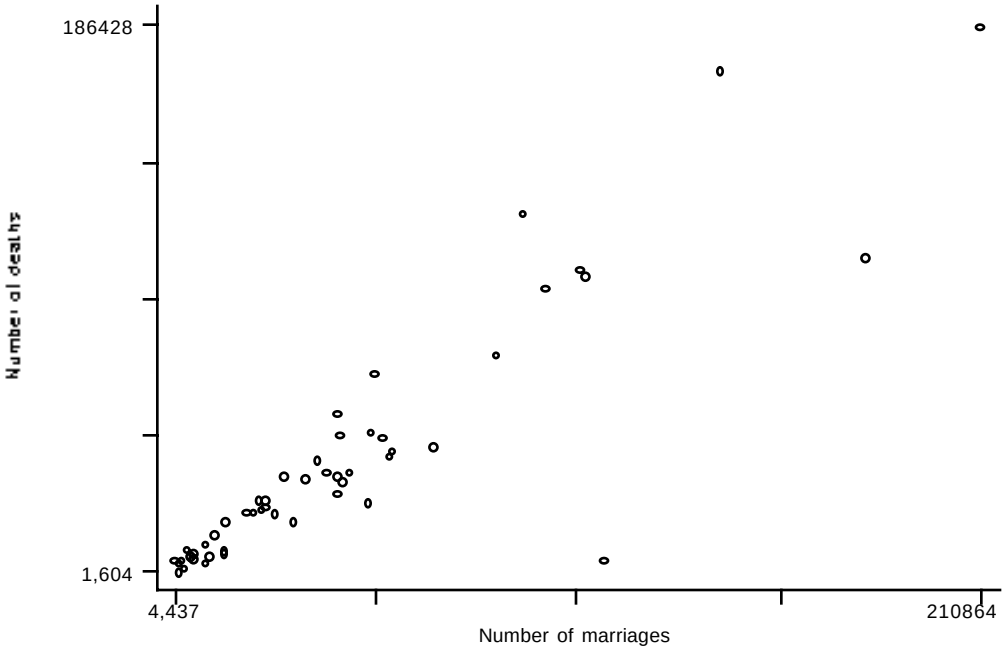
An answer to this must give details. It is not enough to say "because of confounding variables", they must give an example, too. (And the example must be good: it must explain both why the confounder affects the 'assignment' to treatment AND the response variable. An example of a wrong confounder might be 'age: the people in the tai-chi group might have been younger.' This is wrong because while it does explain why we see different values on the response variable, it does not tell us why the age is different in the two groups.)

It is not enough to say "Because there was no placebo." They must also give an explanation of why that would affect this study.

Full credit: for example of confounding variable.

3. The 1980 census collected data on the number of deaths and marriages in each state. (Of course, they collected much more than that.) These numbers for the 50 states are shown in the graph:.

Continued on next page



a) (5) Describe the relationship between the number of marriages and the number of deaths.

There is a strong positive association; states with greater numbers of marriages also have a higher number of deaths. Note that the basic unit here is a state, not a person. Full credit requires an "interpretation". E.g. not just "positive associtions" but "states with greater numbers ..."

Note that it is incorrect to say "as the number of marriages increases, the number of deaths increases,", since there was no state in which we observed an increase of marriages. At the very least, this is an ambiguous statement, because it doesn't make clear whether they mean that marriages increase from state to state or within a state.

b)(5) Someone looking at this might comment that it appears that we can cut down on deaths by preventing marriages. What is another explanation for this trend?

States with a great population will have higher numbers of deaths and marriages.

A common mistake was to say that increased marriages leads to increased populations which leads to increased deaths. But this graph shows a "snapshot" for a singleyear; it can't tell us about future deaths, or about past marriages.

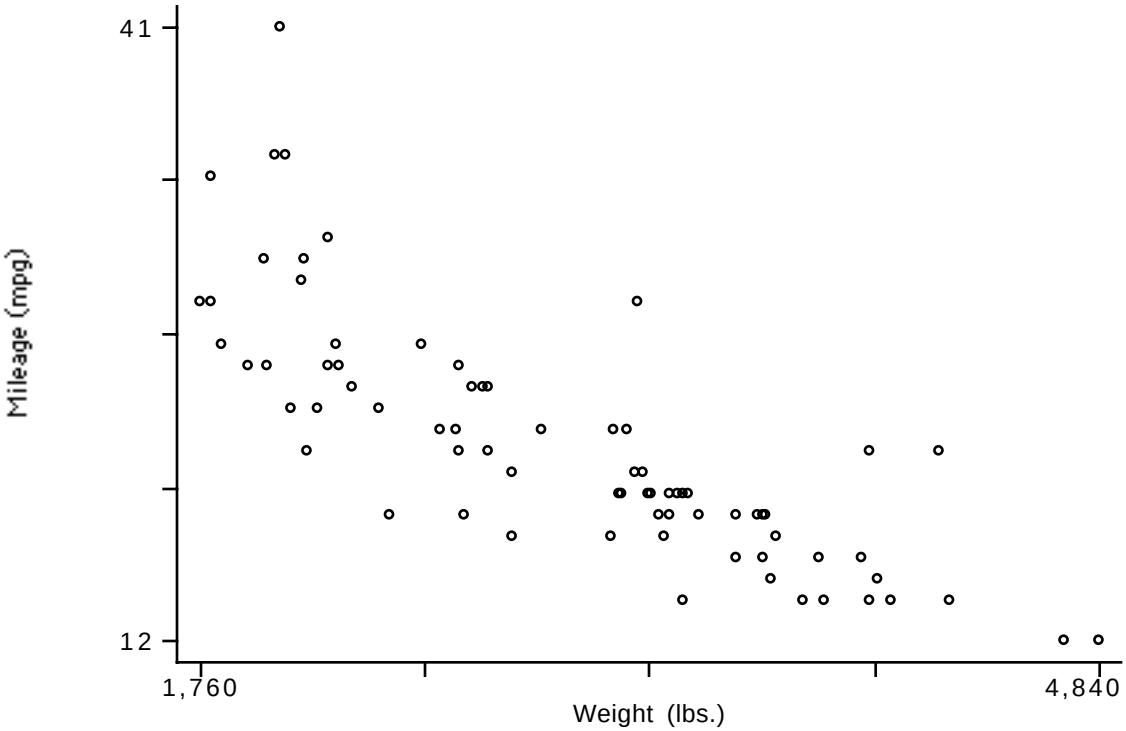
Continued on next page

c) (5) Which of the following is closest to the correlation coefficient of these data? Circle your answer and explain:

-1.0, -0.9, -0.1, 0, 0.1, 0.9, 1.0

They should circle 0.9. It is a positive association, and is too strong to have a correlation of 0.1. But the data do not lie perfectly along a line, so 1.0 is out, too.

4. When the cost of gasoline gets high, consumers begin to worry about the fuel efficiency of their cars. Fuel efficiency is measured by the mileage, which is measured in units of miles per gallon, or mpg. This is the average number of miles a car can go on a gallon of gas. Of course a number of factors affect the mileage, including the type of engine and the shape of the body of the car, to name two. But probably one of the biggest factors is the weight of the car. Shown below is a plot illustrating the relation between weight (pounds) and fuel efficiency (mpg) for 74 cars of various makes and models from 1978.



a) (5) Describe the relation between weight and mileage. Be brief.

Moderately strong negative association. (Non-linear, but full credit without this. However, less than full credit if they say linear.) Heavier cars have lower mileage. Must describe in the context of the data for full credit.

Continued on next page

A common mistake is to say "as weight increases, mileage decreases." This does not make it clear that weight is not increasing. Instead, we are comparing heavier cars.

Another common mistake was to say that "there is a strong negative correlation." While factually true, this begs the question; so what does a strong negative correlation tell us about the relationship? And besides, the exact value of the correlation is given in the next question, so this answer is not very descriptive.

b) (10) Here are some summary statistics from Stata:

Variable	Obs	Mean	Std. Dev.	Min	Max
mpg	74	21.2973	5.785503	12	41
weight	74	3019.459	777.1936	1760	4840

r = -0.8072

Find the equation of the regression line.

b1 = r(SDy)/SDx = (-0.8072* 5.785503)/777.1936 = -0.006 mpg/pound
b0 = ybar - b1*xbar = 21.2973 - (-0.006)*3019.459 = 39.41 mpg

yhat = 39.41 - 0.006x

For full credit, must give complete equation, not just the estimates.

c) (5) According to your regression line, what is the typical mileage for a car that weights 4000 pounds?

yhat = 39.41 - 0.006(4000) = 15.41 mpg.
Note: If their regression line is wrong, they can still get full credit for this as long as they do the calculation correctly (even though it is with the wrong regression line.)

d) (5) What does the value of the slope of your regression line tell us about the relation between weight and mileage? (In other words, interpret the slope.) Don't feel obligated to fill all of the space below. One or two sentences will suffice.

Cars that weight 1 pound more tend to get, on average, 0.006 mpg LESS.

Continued on next page

A common mistake: "as the weight increases by 1 pound, the mpg decrease by 0.006." This is wrong for two reasons: first, changing the weight of a car will affect that car's mileage, but that is not what this study examined, and so we don't know how it will affect the mileage. This study compares different cars, not the same car at different weights. So it is incorrect to say what happens as the weight changes. Second, the mpg does not go down by 0.006 for every type of car. Some even go up slightly. What does go down is the average mpg.

Another common mistake was to ignore the value of the slope and give a generic interpretation for any slope. One thing the slope tells us that the correlation does not is the amount the average y value changes with respect to x.

Finally, note that the slope tells us nothing about the strength of a linear relationship. The correlation does, but not the slope.

Continued on next page

e) (5) Do you think the regression line is a good fit for this data? Why or why not?

No. The data are non-linear and a regression line will not be a good fit.

(Not necessary for full credit, but still true: the line will tend to overestimate most of the mid-range of values.)

End