

NAME (Please Print): \_\_\_\_\_

HONOR PLEDGE (Please Sign): \_\_\_\_\_

Statistics 101

**Midterm 1**

This is a closed book exam. You may use your calculator and a single page of notes.

The room is crowded. Please be careful to look only at your own exam.

Report all numerical answers to at least two correct decimal places or (when appropriate) write them as a fraction.

1. You want to predict the metabolic rate (in average heartbeats/minute) of an animal from its body mass (in kilograms). You draw a sample of 50 animals and find that the estimated slope is  $-0.4$  and the estimated intercept is 200. The mean body mass is 150 kg, and the SD of the body mass is 80 kg. The regression line you fit explains 80% of the variation in metabolic rate as a function of body mass.

Answer the following questions. If a question cannot be answered, mark "NA".

\_\_\_\_\_ What is the correlation coefficient?

\_\_\_\_\_ Would you expect elephants to have fast heartrates?

\_\_\_\_\_ A seal weighs 100 kilograms. What is its estimated heart rate?

\_\_\_\_\_ What is the dependent variable?

\_\_\_\_\_ A gerbil's heart beats 180 times a minute. What is its estimated mass?

\_\_\_\_\_ What is the residual for a mule who weighs 200 kilograms and whose heart beats 80 times a minute?

\_\_\_\_\_ What is the mean heart rate?

\_\_\_\_\_ Is this an ecological correlation?

\_\_\_\_\_ Assume that the sum of the squared residuals is 2000. About 60% of the the data will fall within what vertical distance of the regression line?

\_\_\_\_\_ Which of the following is not a potential problem in regression?

- A. an outlier.
- B. homoscedasticity.
- C. extrapolation.
- D. none of the above.

\_\_\_\_\_ 2. Some studies find an association between exercise and longevity. However, diet is a confounding variable. This means that:

- A: good diet causes longer life.
- B: people who exercise tend to eat healthy diets.
- C: exercise causes longevity and is associated with a good diet.
- D: a good diet increases lifespan and is associated with exercise.
- E: none of the above.

\_\_\_\_\_ 3. In the analysis of the clofibrate trial of heart disease:

- A: people who adhered to the clofibrate treatment had better survival.
- B: people who adhered to the placebo treatment had better survival.
- C: The treatment was not more effective than the control.
- D: all of the above.
- E: none of the above.

4. True or False:

\_\_\_\_\_ The regression line for  $y$  on  $x$  estimates the average value for  $x$  corresponding to each value of  $y$ .

\_\_\_\_\_ According to the regression method, associated with each increase of one SD in  $x$  there is an increase of only  $r$  SDs in  $y$ , on the average.

\_\_\_\_\_ A correlation of .4 implies a stronger association than a correlation of -.6.

\_\_\_\_\_ 5. Which of the following is true?

- A. The *Belmont Report* requires animal studies to conform to the principles of Respect, Beneficence, and Justice.
- B. All human experiments at Duke must be approved by the Institutional Review Board.
- C. All federal surveys with more than nine respondents must be approved by the Census Bureau.
- D. All of the above.

6. California is evaluating a new program to rehabilitate prisoners before their release; the object is to reduce the recidivism rate—the percentage who will be back in prison within two years of release. The program involves several months of “boot camp”—military-style basic training with very strict discipline. Admission to the program is voluntary. According to a prison spokesman, “Those who complete boot camp are less likely to return to prison than other inmates.”

\_\_\_\_\_ Was there an association between recidivism and the program?

\_\_\_\_\_ Was this double-blind?

\_\_\_\_\_ Is the prison spokesman’s comparison based on an observational study (OS) or a randomized controlled experiment (RCE)?

\_\_\_\_\_ 7. Subgroup analysis

- A: is a kind of randomization.
- B: is double-blind.
- C: can control for a confounding factor.
- D: all of the above.
- E: none of the above.

8. Suppose you have the following sample:

1, -5, 22, -4, 11

\_\_\_\_\_ What is the mean?

\_\_\_\_\_ What is the sd?

\_\_\_\_\_ What is the IQR?

\_\_\_\_\_ Suppose the mean is  $m$ . Multiply each number by  $a$  and subtract  $b$ . What is the new mean?

9. The side-by-side box-plots below show the number of letters printed in 15 seconds by men and women, using their non-dominant hand.

\_\_\_\_\_ What is the median number of letters printed by women?

\_\_\_\_\_ About 75% of men printed more than how many letters?

\_\_\_\_\_ Which gender contains an outlier?

\_\_\_\_\_ What is the IQR for men?

10. Among freshman at Emory, scores on the math SAT followed the normal curve with an average of 500 and an sd of 100.

\_\_\_\_\_ A student who scored 350 on the math SAT was in what percentile of the score distribution for Emory?

\_\_\_\_\_ A quarter of the class got higher scores on the math SAT than what value?

\_\_\_\_\_ 11. A pair of dice are rolled 36 times. What is the probability of getting at least one double-ace (i.e., a pair of ones)?

\_\_\_\_\_ 12. In how many ways can one arrange 5 books in a row?

13. A standard deck of 52 cards is shuffled; three cards are dealt (without replacement).

\_\_\_\_\_ What is the probability that the first card is a King or a club?

\_\_\_\_\_ What is the probability that the third card is a seven, given that the first two are Queens?

\_\_\_\_\_ What is the probability that all three cards are face cards (Aces, Kings, Queens, or Jacks)?

14. Abelard is running from 12 henchmen of canon Fulbert. He is faster than 70% of the people in the world. In order to be arrested he must be overtaken by at least two people (if just one catches him, Abelard will knock him down and keep on running).

\_\_\_\_\_ What is the probability that exactly one henchman catches up to him?

\_\_\_\_\_ What is the probability that Abelard is caught?

\_\_\_\_\_ 15. An art museum has three paintings by Picasso, five by Matisse, and 6 by Rembrandt. In how many ways can they put on a show with two paintings from each artist?

16. The American Embassy in London decides to have a party with several foreign visitors. At the party the assistant to the American Ambassador meets a very charming foreign visitor but she can't remember which country he was from. She knows that only visitors from Spain, Italy, and France attended the party; half the visitors were from Spain, a third were from Italy, and the rest were from France. Half of the Spainards were male, 20% of the Italians were male, and 75% of the French were male.

\_\_\_\_\_ What is the probability that the charming visitor was French?

\_\_\_\_\_ What is the probability that a randomly chosen person at the party is male?

17. Dr. Goog has a new ‘color’ diet which he believes will help people lose weight. He enrolls 2000 people in a trial and randomly assigns half to the color diet and half to a control diet. Dr. Goog weighs everyone at the beginning of the study and after 6 months and records whether each person has lost 20 or more pounds. Here are the results:

	lost weight	did not lose weight
control	290	710
color	305	695

\_\_\_\_\_ What is the odds ratio of losing weight on the color diet compared to the control diet?

\_\_\_\_\_ What is the relative risk of losing weight on the control diet compared to the color diet?

\_\_\_\_\_ Which diet appears to be more successful at helping people lose weight?

\_\_\_\_\_ Could the result you found be due to Simpson’s Paradox?

\_\_\_\_\_ 18. Which of the following is a good residual plot for regression?