Midterm (75 minutes)

No calculators allowed. Just set up your answers, for example, P = 49/52. BE SURE TO EXPLAIN YOUR REASONING. If you want extra time, you can buy time at a price of 1 point a minute; for example, if your test is handed in 10 minutes after the scheduled finish time, 10 points will be subtracted from the test score.

- 1. The popular advice columnist, Ann Landers, once asked her readers, "If you had it to do over again, would you have children?" She received 10,000 responses, of which 70% said no, they would not. Why should we interpret this poll cautiously?
- 2. As a statistician, what is the error in this reasoning?

An Austrian physician is reputed to have estimated vulnerability of soldiers in World War I from hospital records, and to have concluded that injuries to limbs were most dangerous and injuries to the head least dangerous because so many hospitalized soldiers were wounded in the limbs and so few in the head.

3. The X-axis in the figure below groups the authors of tweets by the number of followers; for example, the 0-200 range encompasses authors with 0 to 200 followers. The clear red bars show the average number of tweets sent by authors in each range, while the shaded blue bars show the average number of retweets. (There can be more retweets than followers because a tweet can be retweeted by people who see a retweet and are not followers.) Why, despite the label, is the figure not a histogram?



- 4. There are 100 marbles in a black bag—60 red and 40 blue. What is the probability of picking 3 red marbles and 2 blue marbles if you randomly draw 5 marbles, one at a time, from the bag and
 - a. put each marble back in the bag after it is drawn?
 - b. leave each selected marble outside the bag?

- 5. The *Wall Street Journal* and *Washington Post* both reported the results of a 1991 study that estimated the probability that a 40-year-old, sober, seat-belted person driving a heavier-than-average car would have a fatal accident while making a 600-mile automobile trip. The authors of this study calculated this probability by multiplying the overall driver fatality rate by four risk factors. For example, the probability that a heavier-than-average car will have a fatal accident is 0.77 times the probability that a car of average weight will have a fatal accident. So the overall driver fatality rate was multiplied by 0.77. This adjusted number was then multiplied by 0.68 because the probability that a 40-year-old will have a fatal accident is 0.68 times the probability that a driver of average age will have a fatal accident. Similar adjustments were made for being sober and wearing a seat belt. What is wrong with these calculations?
- 6. Suppose that 10% of student essay assignments are written by the ChatGPT text generator and that a ChatGPT detection program is 90% accurate: it correctly flags 90% of all essays written by ChatGPT as having been written by ChatGPT and it correctly flags 90% of all essays written by humans as having been written by humans. What percent of the essays flagged as ChatGPT were in fact written by humans?

- 7. A study divided occupations into 5 categories and found that women had higher unemployment rates than men in each of the individual categories but, overall, had a lower unemployment rate than men. How is this possible?
- 8. Do you agree or disagree with the following statements?
 - a. If two sets of data have the same mean and standard deviation, their histograms are identical \Box yes \Box no

□yes □no

□yes □no

□yes □no

- b. The median is never outside the interquartile range
- c. If A and B are independent, then P[A or B] = P[A] + P[B] $\Box \text{ yes } \Box \text{ no}$
- d. A normal distribution is symmetrical about its median
- e. The expected value of a probability distribution is the most likely outcome
- 9. On Wednesday morning, Charlie realizes that there is a final examination beginning in 10 minutes in a class Charlie has not attended all semester. (Don't ask why.) The test consists of 10 true-false questions, each graded as follows: 1 point for giving the right answer, 0 points for not answering the question, and minus 2 points for giving a wrong answer. A score of 6 will be an A grade; 1-5 is a B; 0 is a C; −1 to −5 is a D; and anything lower is an F. Charlie doesn't know any of the answers and must decide how many questions to answer with random guesses.
 - a. What is the expected value of a guess on any question?
 - b. What strategy maximizes Charlie's chances of a B grade? If Charlie follows this strategy, what is the probability of a B grade?

