

Statistics 2000, Section 001, Midterm 2 (200 Points)

Friday, October 30, 2009

Your Name: _____

Question 1: Short Answers (40 Points)

Please answer the following short answers. **Show your work!**

1. **(10 Points)** Universal blood donors: People with type O–negative blood are universal donors. That is, any patient can receive a transfusion of O–negative blood. About 6% of a particular population have O–negative blood. If 12 people appear at random to give blood, what is the probability that at least 1 of them is a universal donor?

2. **(10 Points)** Disappearing Internet sites: Internet sites often vanish or move, so the references to them can't be followed. In fact, for the *The Journal of the American Medical Association*, 9/42 (i.e., 21.4%) of Internet sites referenced in papers published in this journal are lost within about 26 months (slightly more than two years) after publication. If a paper in this journal contains 3 Internet references, what is the probability that all 3 are still good about 26 months after the publication? We assume that each site is independent of the others and that they can be considered as a random sample from the collection of all sites referenced in this journal.

3. **(20 Points)** Portfolio analysis. Here are the means, standard deviations, and correlations for the annual returns from three Fidelity mutual funds for the 10 years ending in February 2004. Because there are three random variables, there are three correlations. We use subscripts to show which pair of random variables a correlation refers to.

W = annual return on 500 Index Fund
X = annual return on Investment Grade Bond Fund
Y = annual return on Diversified International Fund

$$\begin{aligned}\mu_W &= 10.12\%, & \sigma_W &= 17.46\% \\ \mu_X &= 6.46\%, & \sigma_X &= 4.18\% \\ \mu_Y &= 11.10\%, & \sigma_Y &= 15.62\%\end{aligned}$$

Correlations:

$$\rho_{WX} = -0.22, \quad \rho_{WY} = 0.56, \quad \rho_{XY} = -0.12$$

Assume that a portfolio contains 70% Investment Grade Bond Fund (X) and 30% Diversified International Fund (Y) stocks. Calculate the mean and standard deviation of the returns for this portfolio.

Question 3: Two-Way Tables (50 Points)

Performance in Exams. In a recent Stat 2000 midterm, one question (Q2) seemed to have a high impact on the overall exam performance. A total of 39 students participated. Their grades have been combined as AB, C, and DF. The performance on Q2 was either well (> 30 points) or poor (< 10 points). Scores between 10 points and 30 points were not awarded for question Q2. **Show your work!**

Below is the table that summarizes grade and Q2 performance:

	AB	C	DF	Total
Q2Well	8	1	0	
Q2Poor	2	12	16	
Total				39

1. (5 Points) Calculate the row and column totals and add them to the table above.
2. (6 Points) Determine the **joint distribution** of Q2 performance and grade. Add the percentages that represent this distribution to the empty table cells below. Report your numbers as percentages rounded to one decimal digit, e.g., 40.8% or 2.7%. When all your roundings are done correctly, your percentages should sum up to exactly 100%.

	AB	C	DF	Total
Q2Well				
Q2Poor				
Total				100%

3. (5 Points) Add the **marginal distribution** of Q2 performance and the **marginal distribution** of grade to the table above.

Answer the following probability questions. When doing so, first translate the everyday language into probability statements, e.g., poor answer on Q2 and D or F grade should be translated into $P(\text{Q2poor and DF})$. Then read off the probabilities directly from the table or indicate any calculations you have to perform to obtain the final answer. Report your final answer as a percent with one decimal digit (as in the table above).

4. (5 Points) What is the probability for a randomly selected student to do well on Q2 and still only obtain a D or F grade?

5. **(5 Points)** What is the probability for a randomly selected student to do well on Q2 and obtain an A or B grade?

6. **(5 Points)** What is the probability for a randomly selected student to obtain a C grade?

7. **(6 Points)** Knowing that a randomly selected student does well on Q2, what is the probability for this student to obtain an A or B grade?

8. **(6 Points)** Knowing that a randomly selected student obtains an A or B grade, what is the probability for this student to do well on Q2?

9. **(6 Points)** What is the probability for two **different** randomly selected students to do well on Q2 and obtain an A or B grade each?

10. **(6 Points)** Can we say that doing well on Q2 **causes** a good overall grade?
Yes / No ? Circle your answer.
If **No**, what would be a better way to describe what we have seen in the tables above? Do not write more than two sentences here.

11. **(5 Points)** Indicate at least one lurking and/or confounding variable that effects the relationship we have seen in the tables above. Explain how this variable effects the observed relationship and draw an appropriate diagram (think of the diagrams used for causation, common response, and confounding).

Question 4: Multiple Choice Questions (60 Points)

Circle your answer. There is only one correct answer for each question. Each correct answer is worth 4 points.

1. A large company has been sued for sex discrimination. The case brought by the female managers said they were underrepresented in management. However, further analysis of the company by division found that females were actually more likely than males to be managers in each division. This is an example of
 - (a) aggregating data.
 - (b) Simpson’s Paradox.
 - (c) biased data collection.
 - (d) all of the above.

2. Fill in the blank. A researcher is conducting a study on contact lenses. She finds that in a sample of contact lens-wearing adults, those who wear their contact lenses for longer periods of time have more irritation in their eyes. Those who wear their contact lenses for only a few hours a day have less irritation. However, those who wear their contacts for longer times also tend to spend long hours at the office. As explanations for having more irritation in the eyes, the variables “length of time contacts are worn each day” and “length of time spent at the office per day” are _____ variables.
 - (a) response
 - (b) independent
 - (c) confounding
 - (d) placebo

3. When exploring very large sets of data involving many variables, which of the following is true?
 - (a) Extrapolation is safe because it is based on a greater quantity of evidence.
 - (b) Associations will be stronger than would be seen in a much smaller subset of the data.
 - (c) A strong association is good evidence for causation because it is based on a large quantity of information.
 - (d) None of the above.

4. Researchers are studying the absorption of two drugs into the bloodstream. Each drug is to be injected at three dosages. There are 24 people in the study, and they are randomly divided into six groups. Each of the six groups is randomly assigned to a different drug and dosage combination. After 30 minutes the concentration of drugs in a subject's blood is determined. Which of the following is correct?
- (a) The type of drug is a factor.
 - (b) The dosage of the drug is a factor.
 - (c) Both (a) and (b) are correct.
 - (d) Neither (a) nor (b) are correct.
5. A sociologist wants to study the attitudes of American male college students toward marriage and husband–wife relations. She gives a questionnaire to 25 of the men enrolled in Sociology 101 at her college. All 25 complete and return the questionnaire. The sample in this situation is
- (a) all American male college students.
 - (b) all men taking a comparable sociology class.
 - (c) the 25 men who received and returned the questionnaire.
 - (d) all the men in the Sociology 101 class.
6. A statistics instructor wants to know which route will get her to school the fastest. Each day from October 2 to November 15, when she gets to the turn point she checks the odometer on her car. If it shows an even number she takes the freeway; if it shows an odd number, she takes the in–town route. She records the total time each day. What is the explanatory variable in this study?
- (a) Her speed.
 - (b) The route.
 - (c) The odometer reading.
 - (d) The time to school.
7. A researcher is studying the effects of a new drug on reducing high blood pressure. He recruits 250 men to test either the new active drug against a current standard. At the end of six weeks, the decrease in systolic blood pressure will be evaluated. He believes the drug will be more effective for black men than for white men. To properly test his belief, the experiment should be
- (a) Completely randomized.
 - (b) Blind.
 - (c) Double–blind.
 - (d) Blocked.

8. I want to take a survey of students currently enrolled in my statistics course. There are 250 of them, so I number them alphabetically from 001 to 250. Use the portion of the random number table below to select the numbers for first five to be interviewed.

69041 65817 87174 09514 8174 06423 93758 23612 17894

- (a) 69, 41, 65, 81, 78
(b) 690, 416, 581, 787, 174
(c) 174, 095, 148, 064, 239
(d) 174, 095, 148, 174, 064
9. A set of ten cards consists of five red cards and five black cards. The cards are shuffled thoroughly and I am given the first four cards. I count the number of red cards X in these four cards. The random variable X has which of the following probability distributions?
- (a) The binomial distribution with parameters $n = 10$ and $p = 0.5$.
(b) The binomial distribution with parameters $n = 5$ and $p = 0.5$.
(c) The binomial distribution with parameters $n = 4$ and $p = 0.5$.
(d) None of the above.
10. Which of the following might be reasonably modeled by the binomial distribution?
- (a) The number of customers that enter a store in a one-hour period, assuming customers enter independently.
(b) The number of questions you get correct on a 100-question multiple choice exam in which each question has four possible answers and exactly one of these answer is correct. Assume you have not studied at all for the test and have to guess each answer.
(c) A couple will have children until they have three girls or five children. X is the number of children in the family.
(d) None of the above.
11. A fair die is rolled 12 times. Let $X =$ the number of times an even number occurs on the 12 rolls. What is the appropriate distribution for the random variable X ?
- (a) A binomial distribution with a mean of 2.
(b) A binomial distribution with a mean of 12.
(c) A binomial distribution with a mean of 0.5.
(d) A binomial distribution with a mean of 6.

12. A fair coin is tossed, and a fair six-sided die is rolled. Suppose the outcomes for the coin and die are independent. The probability of getting a head and rolling a six is
- (a) $1/6$
 - (b) $1/12$.
 - (c) $2/3$.
 - (d) $1/3$.
13. A process is called random if
- (a) Individual outcomes are uncertain but happen in a predictable manner through time.
 - (b) One has no idea what will happen.
 - (c) Outcomes happen in a 50–50 split.
 - (d) I flip a coin and get 5 heads on my first 5 flips, I am very likely to get 5 tails on my next 5 flips.
14. Suppose we toss a penny and a nickel. Let A be the event that the penny is a head and B be the event that the nickel is a tail. The events A and B are
- (a) disjoint.
 - (b) complements.
 - (c) independent.
 - (d) (b) and (c).
15. At the end of a production run manufacturing rubber gaskets, items are sampled at random and inspected to determine if the item is Acceptable (A), or Defective (D). Suppose it is planned to select two items and determine if each is either A or D . What is the sample space S of the outcomes?
- (a) $S = \{A, D\}$.
 - (b) $S = \{AA, DD\}$.
 - (c) $S = \{AA, AD, DA, DD\}$.
 - (d) This depends upon the assignment of probability to the outcomes of the sampling.