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

Friday, March 26, 2010

Your Name: _____

Question 1: Short Probability Answers (**40 Points**)

Distribution of blood types: All human blood can be "ABO–typed" as one of O, A, B, or AB, but the distribution of the types varies a bit among groups of people. Below are the distributions of blood types for randomly chosen persons in the United States and in China. **Show your work!**

Blood Type	A	B	AB	O
U.S. Probability	0.40	0.11	0.04	0.45
China Probability	0.27	0.26	0.12	0.35

1. (**10 Points**) We choose an American and a Chinese at random, independently from each other. What is the probability that both have type O blood?

2. (**10 Points**) We choose an American and a Chinese at random, independently from each other. What is the probability that both have the same blood type?

3. (**10 Points**) If 6 Americans (and 0 Chinese) appear at random to give blood (independently from each other), what is the probability that at least one of them has type O blood?

4. (**10 Points**) If 1 American and 1 Chinese appear at random to give blood (independently from each other), what is the probability that at least one of them has type O blood?

Question 2: Means and Variances of Random Variables (**40 Points**)

The weight of medium–size tomatoes selected at random from a large bin at the local supermarket is a random variable X with mean $\mu_X = 10$ oz. and standard deviation $\sigma_X = 1$ oz. **Show your work!**

1. (**10 Points**) Let the random variable W be the weight of the tomatoes in pounds (1 pound = 16 oz). What is the standard deviation σ_W of the random variable W (in pounds)?

2. (**10 Points**) Suppose we pick four tomatoes from the bin at random and put them in a bag. Define the random variable Y as the weight of the content of the bag containing the four tomatoes. What is the mean μ_Y of the random variable Y (in oz.)?

3. (**10 Points**) Suppose we pick four tomatoes from the bin at random and put them in a bag. Define the random variable Y as the weight of the content of the bag containing the four tomatoes. What is the standard deviation σ_Y of the random variable Y (in oz.)?

4. (**10 Points**) Suppose we pick two tomatoes at random from the bin. Let the random variable V be the difference in the weights (in oz.) of the two tomatoes selected (i.e., the weight of the first tomato minus the weight of the second tomato). What is the standard deviation σ_V of the random variable V (in oz.)?

Question 3: Two–Way Tables (**60 Points**)

The 94 students in a statistics class are categorized by gender and by the year in school. The numbers obtained are displayed in the table below. **Show your work!**

	Year in School					
Gender	Freshman	Sophomore	Junior	Senior	Graduate	Total
Male	1	2	9	17	2	
Female	23	17	13	7	3	
Total						94

1. (**7 Points**) Calculate the row and column totals and add them to the table above.

2. (**10 Points**) Determine the **joint distribution** of Year in School and Gender. 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 roughly 100%.

	Year in School					
Gender	Freshman	Sophomore	Junior	Senior	Graduate	Total
Male						
Female						
Total						100%

3. (**7 Points**) Add the **marginal distribution** of Year in School and the **marginal distribution** of Gender to the table above.

Answer the probabiliy questions following on the next page. When doing so, first translate the everyday language into probability statements, e.g., Freshman and Female should be translated into P(Freshman and Female). 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. (**6 Points**) What is the probability for a randomly selected student to be a Sophomore?

5. (**6 Points**) What is the probability for a randomly selected student to be Male?

6. (**6 Points**) What is the probability for a randomly selected student to be Male and Sophomore?

7. (**8 Points**) Knowing that a randomly selected student is Male, what is the probability for this student to be a Sophomore?

8. (**10 Points**) What is the probability for two **different** randomly selected students to be Male and Sophomore each?

Question 4: Multiple Choice Questions **(60 Points)**

Mark your answer for each multiple choice question in the table below. There is only one correct answer for each question. Each correct answer is worth 4 points.

Question	(a)	(b)	(c)	(d)	Question	(a)	(b)	(c)	(d)
1	◯	◯	◯	◯	11	◯	◯	◯	◯
2	◯	◯	◯	◯	12	◯	◯	◯	◯
3	◯	◯	◯	◯	13	◯	◯	◯	◯
4	◯	◯	◯	◯	14	◯	◯	◯	◯
5	◯	◯	◯	◯	15	◯	◯	◯	◯
6	◯	◯	◯	◯					
7	◯	◯	◯	◯					
8	◯	◯	◯	◯					
9	◯	◯	◯	◯					
10	◯	◯	◯	◯					

1. A small store keeps track of X = the number of customers who make a purchase during the first hour that the store is open each day. Based on the records, X has the following probability distribution:

Value of X	0	1	2	3	4
Probability	0.1	0.1	0.1	0.1	0.6

What is the mean number of customers who make a purchase during the first hour that the store is open?

(a) 2.0

(b) 2.5

(c) 3.0

(d) 4.0

2. There are four people in a family — a father, a mother, and two children — and they have won two tickets to go to Disneyland for a week. They decide to select a sample of two people for the trip as follows: The mother and father flip a coin to see which of the two of them will go, and then they flip a coin to see which of the two children will go. This is

(a) a simple random sample of size two from the family since two coins were flipped.

(b) a probability sample from the family since each member of the family has a known chance of being selected to go on the trip.

(c) not a probability sample since the mother and father can't go together.

(d) a voluntary response sample.

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.

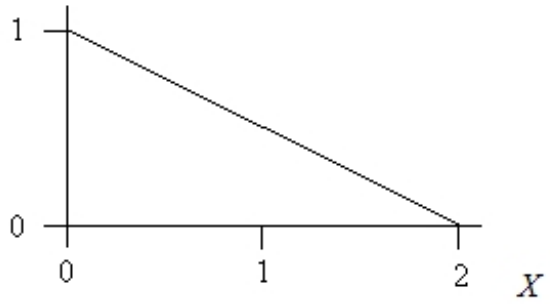
(c) A strong association is good evidence for causation because it is based on a large quantity of information.

(d) None of the above.

4. Let X be a random variable with mean $\mu_X = 25$ and standard deviation $\sigma_X = 6$ and let Y be a random variable with mean $\mu_Y = 30$ and standard deviation $\sigma_Y = 4$. It is known that X and Y are independent random variables. Suppose the random variables X and Y are added together to create a new random variable W, i.e., $W = X + Y$. What is the standard deviation σ_W of W?

   (a) $\sigma_W = 7.2$
   (b) $\sigma_W = 10$
   (c) $\sigma_W = 52$
   (d) $\sigma_W = 100$

5. Suppose X is a continuous random variable taking values between 0 and 2 and having the probability density curve below:



   What is $P(1 \leq X \leq 2)$?

   (a) 0.50
   (b) 0.33
   (c) 0.25
   (d) 0.00

6. A particular city is serviced by three airlines for its passenger traffic. Airline A carries 50% of the passengers, Airline B 30%, and Airline C the remaining 20%. Each of the airlines is responsible for handling its security. The probabilities that a passenger who is carrying some type of weapon will be detected by Airlines A, B, and C are 0.9, 0.5, and 0.4, respectively. What is the overall probability that a passenger who is carrying a weapon will be detected?

   (a) 0.15
   (b) 0.32
   (c) 0.45
   (d) 0.68

7. An experiment was conducted and the results involving the effect of a particular experimental variable was widely reported. Later a critic commented that, in fact, nothing could usefully be learned from the study because the observed effect was confounded. What did the critic mean by this?

   (a) The experimenter really wasn't clear about what was desired so the experiment was not well defined.

   (b) The results were so badly described that the critic could not figure out exactly what was being concluded.

   (c) The critic felt there was confusion as to whether or not the study was an observational study or was a real experiment.

   (d) Because of the way the experiment was conducted the effect observed in the experiment, ascribed to a particular variable, was mixed up with other influences.

8. In a controversial election district, **73%** of registered voters are Democrats. A random survey of 500 voters had **68%** Democrats. Are the bold numbers parameters or statistics?

   (a) Both are statistics.

   (b) 73% is a parameter and 68% is a statistic.

   (c) 73% is a statistic and 68% is a parameter.

   (d) Both are parameters.

9. Assuming the population is large, which sample size will give the smallest standard deviation to the statistic?

   (a) $n = 100$.

   (b) $n = 500$.

   (c) $n = 1000$.

   (d) It depends on the distribution, i.e., whether it is symmetric or skewed.

10. 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

11. A researcher is interested in the cholesterol levels of adults in the city she lives in. A free cholesterol screening program is set up in the downtown area during the lunch hour. Individuals can walk in and have their cholesterol determined for free. One hundred and seventy–three people use the service and their average cholesterol is 217.8. The sample obtained is an example of

    (a) simple random sample since the experimenter did not know beforehand which individuals would come to the screening.

    (b) a stratified sample of high and low cholesterol individuals.

    (c) a sample probably containing bias and undercoverage.

    (d) a stratified sample of men and women.

12. A game consists of drawing three cards at random (without replacement) from a deck of playing cards. You win \$3 for each red card that is drawn. It costs \$2 to play. For one play of this game, the sample space S for the net amount you win (after deducting the cost of play) is

    (a) $S = \{\$0, \$1, \$2, \$3\}$.

    (b) $S = \{\$0, \$3, \$6, \$9\}$.

    (c) $S = \{-\$2, \$1, \$4, \$7\}$.

    (d) $S = \{(Red, Red, Red), (Red, Red, Black), \ldots, (Black, Black, Black)\}$.

13. 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).

14. 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.

15. 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.