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

Friday, March 26, 2010

Your Name: _____

based on: HW 7, Exercise 1: 4.22

-2 for each calculation error (or no final result)

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?

$P(\text{both type O}) = 0.45 \cdot 0.35 = 0.1575 = 15.75\%$ (3) (4) (3)

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?

$P(\text{both same type}) = P(\text{both A}) + P(\text{both B}) + P(\text{both AB}) + P(\text{both O})$

$= 0.40 \cdot 0.27 + 0.11 \cdot 0.26 + 0.04 \cdot 0.12 + 0.45 \cdot 0.35$ (2) (1) (2) (1) (2) (1) (1)

$= 0.2989 = 29.89\%$

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?

$P(\text{at least one type O}) = 1 - P(\text{no type O})$

$= 1 - (1 - 0.45)^6 = 1 - 0.0277$ (5) (5)

$= 0.9723 = 97.23\%$

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?

$P(\text{at least one type O}) = 1 - P(\text{neither type O})$

$= 1 - (1 - 0.45) \cdot (1 - 0.35) = 1 - 0.3575 = 0.6425 = 64.25\%$ (5) (5)

OR: $P(\text{at least one type O}) = P(\text{American O or Chinese O})$

$= P(\text{American O}) + P(\text{Chinese O}) - P(\text{both O})$

$= 0.45 + 0.35 - 0.1575 = 0.6425 = 64.25\%$

-2 for each calculation error (or no final result)

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)?

$W = \frac{1}{16} X$

$\left[\sigma_W^2 = \left(\frac{1}{16}\right)^2 \sigma_X^2 = \left(\frac{1}{16}\right)^2 \cdot 1^2 \right]$

$\sigma_W = \frac{1}{16}\sigma_X = \frac{1}{16} \cdot 1 \overset{(16)}{=} \frac{1}{16} = 0.0625$ pound

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.)?

$Y = X_1 + X_2 + X_3 + X_4$

$\mu_Y = \mu_{X_1+X_2+X_3+X_4} = \mu_{X_1} + \mu_{X_2} + \mu_{X_3} + \mu_{X_4}$

$= 10 + 10 + 10 + 10 = 4 \cdot 10 \overset{(10)}{=} 40$ 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.)?

$Y = X_1 + X_2 + X_3 + X_4$

$\sigma_Y^2 = \sigma^2_{X_1+X_2+X_3+X_4} = \sigma^2_{X_1} + \sigma^2_{X_2} + \sigma^2_{X_3} + \sigma^2_{X_4}$

$= 1^2 + 1^2 + 1^2 + 1^2 = 4 \cdot 1^2 \overset{(5)}{=} 4\ \text{oz}^2$

$\sigma_Y = \sqrt{4} \overset{(5)}{=} 2$ 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.)?

$V = X_1 - X_2$

$\sigma_V^2 = \sigma^2_{X_1-X_2} \overset{!}{=} \sigma^2_{X_1} + \sigma^2_{X_2} = 1^2 + 1^2 \overset{(5)}{=} 2\ \text{oz}^2$

$\sigma_V = \sqrt{2} \overset{(5)}{=} 1.41$ oz

-2 for each calculation error (or no final result)

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!**

Gender	Year in School: Freshman	Sophomore	Junior	Senior	Graduate	Total
Male	1	2	9	17	2	31
Female	23	17	13	7	3	63
Total	24	19	22	24	5	94

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

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%. 10 × (1)

e.g., $\frac{1}{94} = 0.011 = 1.1\%$, etc.

Gender	Year in School: Freshman	Sophomore	Junior	Senior	Graduate	Total
Male	1.1%	2.1%	9.6%	18.1%	2.1%	33.0%
Female	24.5%	18.1%	13.8%	7.4%	3.2%	67.0%
Total	25.5%	20.2%	23.4%	25.5%	5.3%	100%

3. (**7 Points**) Add the **marginal distribution** of Year in School and the **marginal distribution** of Gender to the table above. 7 × (1)

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?

$P(\text{Sophomore}) = \frac{19}{94} = \underline{20.2\%}$ (1) (5)

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

$P(\text{Male}) = \frac{31}{94} = \underline{33.0\%}$ (1) (5)

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

$P(\text{Male and Sophomore}) = \frac{2}{94} = \underline{2.1\%}$ (1) (5)

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

$P(\text{Sophomore} \mid \text{Male}) = \frac{2}{31} = 0.065 = \underline{6.5\%}$ (2) (6)

OR: $P(\text{Sophomore} \mid \text{Male}) = \frac{P(\text{Sophomore and Male})}{P(\text{Male})} = \frac{2/94}{31/94} = \frac{2}{31} = 0.065$

$= \underline{6.5\%}$ (as above!)

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

$P(\text{each Male and Sophomore})$ (2)

$= P(\text{first Male and Sophomore}) \cdot P(\text{second Male and Sophomore} \mid \text{first Male and Sophomore})$

$= \frac{2}{94} \cdot \frac{1}{93} = \frac{2}{8742} = 2.3 \cdot 10^{-4} = \underline{0.023\%}$

(3) (2) (3)

based on: • 5 questions directly taken from "Online Quizzes" for Chapters 2, 3 & 4 (see HW 8, Exercise 2 (EC))

• 5 questions adapted from Stat 2000, Fall 2009, Midterm 2

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	○	○	●	○					

Stat 2000, Midterm 2, Question 4 – Solutions

1. (c) mean = 0 * 0.1 + 1 * 0.1 + 2 * 0.1 + 3 * 0.1 + 4 * 0.6 = 3.0

2. (b) The chance of each person in the family going is 1/2 (the chance that their coin comes up favoring them), so that each person in the population has a known chance to be selected. This is the definition of a probability sample.

3. (d) None of the above – all problems still exist for very large data sets.

4. (a) standard deviation of W = sqrt($6^2 + 4^2$) = sqrt(52) = 7.2

5. (c) P = area of triangle from 1 to 2 = ½ * width * height = ½ * 1 * ½ = ¼

6. (d) P = 0.5 * 0.9 + 0.3 * 0.5 + 0.2 * 0.4 = 0.68

7. (d) Confounding means that effects of two or more variables are mixed up.

8. (b) Here, 73% is the percentage of all registered voters in the district, so it is a parameter. 68% was obtained from the sample of 500 voters, so it is a statistic.

9. (c) To reduce the variability of a sample statistic, increase the sample size, so the largest sample size will have the smallest variability.

10. (c) Since we have students numbered 001 to 250, all numbers selected must be within that range. We ignore any groups of three digits which result in numbers greater than 250 (or 000), as well as any duplicates.

11. (c) Individuals "volunteered" for the screening. Individuals who knew their cholesterol was fine may not have bothered to have it checked, which would cause the average from the screening to be too high. The downtown area was not equally accessible to all adults in the population - those that work in the downtown area are more likely to be in the sample while other groups may not have known about it or may have been unable to reach the screening center, which would lead to undercoverage.

12. (c) The amounts you would "win" if there were no cost to play are \$0 if there are no red cards drawn, \$3 if there were one red card drawn, \$6 if there were two red cards drawn, and \$9 if there were three red cards drawn. But, you must subtract the cost to play from each of these to determine the net amount you would win. If three red cards were drawn, your net profit would be \$9 - \$2 = \$7, etc.

13. Just (c). Since the tosses of the penny and nickel do not influence each other, any event concerning the outcome for the penny is independent of any event concerning the outcome for the nickel. So the event A, the penny shows a head, is independent of the event B, the nickel shows a tail.

14. (c) We observe pairs of outcomes – and there are four different pairs.

15. (b) Simpson's paradox refers to situations which may arise whereby the relationship seen in a two-way table may change direction when a third variable (division in this case) is considered.