

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

Friday, October 30, 2009

Your Name: _____

based on: HW 7, Exercise 1, Q 4.34 & 4.35

Question 1: Short Answers (40 Points)

Please answer the following short answers. Show your work!

-2 for each calculation error

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?

$$P(\text{none are O-negative}) = (1 - 0.06)^{12} \stackrel{(5)}{=} 0.4759$$

$$\Rightarrow P(\text{at least 1 is O-negative}) = 1 - 0.4759 \stackrel{(5)}{=} 0.5241 = \underline{\underline{52.41\%}}$$

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.

$$P(\text{site still good}) = 1 - 0.214 \stackrel{(5)}{=} 0.786$$

$$\Rightarrow P(\text{all 3 still good}) = 0.786^3 \stackrel{(5)}{=} 0.4856 = \underline{\underline{48.56\%}}$$

based on: HW8, Exercise 1, Q 4.91

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

$$\mu_W = 10.12\%, \quad \sigma_W = 17.46\%$$

$$\mu_X = 6.46\%, \quad \sigma_X = 4.18\%$$

$$\mu_Y = 11.10\%, \quad \sigma_Y = 15.62\%$$

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.

$$R = 0.7X + 0.3Y \quad (2)$$

$$\begin{aligned} \mu_R &= 0.7\mu_X + 0.3\mu_Y \\ &= 0.7 \cdot 6.46\% + 0.3 \cdot 11.10\% = \underline{7.852\%} \end{aligned} \quad (8)$$

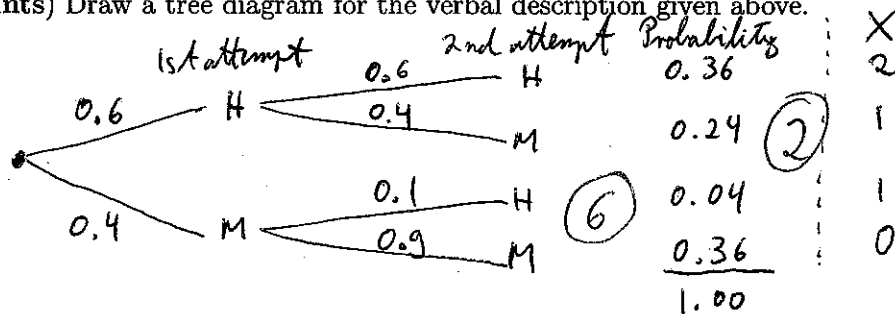
$$\begin{aligned} \sigma_R &= \sqrt{(0.7\sigma_X)^2 + (0.3\sigma_Y)^2 + 2 \cdot \rho_{XY} \cdot (0.7\sigma_X) \cdot (0.3\sigma_Y)} \\ &= \sqrt{0.7^2 \cdot 4.18^2 + 0.3^2 \cdot 15.62^2 + 2 \cdot (-0.12) \cdot 0.7 \cdot 4.18 \cdot 0.3 \cdot 15.62} \% \\ &= \sqrt{27.23} \% \\ &= \underline{5.22\%} \end{aligned} \quad (10)$$

Question 2: General Probability Rules (40 Points)

-2 for each calculation error

A highly nervous basketball player has two free throw attempts. The player will make a hit with a probability of 0.6 in the first attempt. The outcome of the second attempt highly depends on the outcome of the first attempt: If the first attempt was a hit, the player will also make a hit with a probability of 0.6 in the second attempt. However, if the first attempt was a miss, the player will also miss in the second attempt with a probability of 0.9. Show your work!

1. (8 Points) Draw a tree diagram for the verbal description given above.



2. (4 Points) Let X be a random variable that counts the hits in the two free throw attempts. Indicate the corresponding sample space S .

$$S = \{0, 1, 2\} \quad (4)$$

3. (10 Points) Using your tree diagram from above, fill in the following table.

Value of X	0	1	2
Probability	0.36 (3)	0.24+0.04 = 0.28 (4)	0.36 (3)

4. (8 Points) Calculate the mean (expected value) μ_X of the random variable X .

$$\mu_X = x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 = 0 \cdot 0.36 + 1 \cdot 0.28 + 2 \cdot 0.36 = \underline{\underline{1.00}} \quad (8)$$

5. (10 Points) Calculate the standard deviation σ_X of the random variable X .

$$\begin{aligned} \sigma_X^2 &= (x_1 - \mu_X)^2 \cdot p_1 + (x_2 - \mu_X)^2 \cdot p_2 + (x_3 - \mu_X)^2 \cdot p_3 \\ &= (0 - 1)^2 \cdot 0.36 + (1 - 1)^2 \cdot 0.28 + (2 - 1)^2 \cdot 0.36 \quad (7) \\ &= 1 \cdot 0.36 + 0 \cdot 0.28 + 1 \cdot 0.36 \\ &= 0.72 \\ \Rightarrow \sigma_X &= \sqrt{0.72} = \underline{\underline{0.85}} \quad (3) \end{aligned}$$

Question 3: Two-Way Tables (60 Points)

-1 for each calculation error

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	9
Q2Poor	2	12	16	30
Total	10	13	16	39

5 x (1)

- (5 Points) Calculate the row and column totals and add them to the table above.
- (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%. *e.g., $8/39 = 0.205 = 20.5\%$*

	AB	C	DF	Total
Q2Well	20.5%	2.6%	0%	23.1%
Q2Poor	5.1%	30.8%	41.0%	76.9%
Total	25.6%	33.4%	41.0%	100%

6 x (1)

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

e.g., $20.5\% + 2.6\% + 0\% = 23.1\%$ 5 x (1)

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).

- (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?

$P(\text{Q2Well and DF}) = 0\%$ (does not happen!)

9. (continued): $\star = \frac{8}{39} \cdot \frac{7^{\textcircled{5}}}{38} = \frac{56}{1482} = 0.038 = \underline{\underline{3.8\%}}$

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

$$P(\text{Q2 Well and AB})^{\textcircled{1}} = \underline{\underline{20.5\%}}^{\textcircled{4}}$$

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

$$P(C)^{\textcircled{1}} = \underline{\underline{33.4\%}}^{\textcircled{4}}$$

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?

$$P(AB | \text{Q2 Well})^{\textcircled{1}} = \frac{P(AB \text{ and Q2 Well})^{\textcircled{5}}}{P(\text{Q2 Well})} = \frac{20.5}{23.1} = 0.887 = \underline{\underline{88.7\%}}$$

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? [= 8/9!]

$$P(\text{Q2 Well} | AB)^{\textcircled{1}} = \frac{P(\text{Q2 Well and AB})^{\textcircled{5}}}{P(AB)} = \frac{20.5}{25.6} = 0.801 = \underline{\underline{80.1\%}}$$

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? [= 8/10!]

$$P(\text{each Q2 Well and AB})^{\textcircled{1}} = P(\text{first Q2 Well and AB}) \cdot P(\text{second Q2 Well and AB} | \text{first Q2 Well and AB})$$

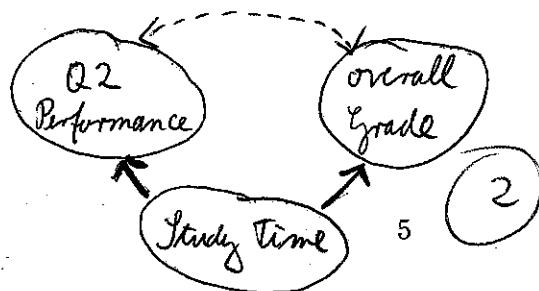
10. (6 Points) Can we say that doing well on Q2 causes a good overall grade? Yes / No $\textcircled{3}$ Circle your answer. = $\textcircled{\star}$

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.

There is an obvious association between doing well on Q2 and obtaining a good overall grade, but association is not causation! Most likely, other variables also contribute to obtaining a good grade! $\textcircled{3}$

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).

Study time $\textcircled{2}$. Someone who has studied long enough to answer Q2 well also should be able to answer many other questions. $\textcircled{1}$



This is the "Common Response" situation in Fig. 2.28(b)

from: HW8, Exercise 2: 10 (out of 15) questions directly taken from "Online Quizzes" for chapters 2 to 5

Stat 2000, Midterm 2, Question 4 – Solutions

1. (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.
2. (c) confounding
3. (d) None of the above – all problems still exist for very large data sets.
4. (c) The explanatory variables in an experiment are often called factors. Many experiments study the joint effects of several factors. This experiment is studying the effect of two factors, type of drug and dosage of drug on the response, which is concentration in the blood. Each treatment is formed by combining the specific values (levels) of each of the factors. Since there are two drugs and three dosages for each, there are $3 * 2 = 6$ treatments in this experiment, which is why the subjects were divided into six groups, one for each possible treatment.
5. (c) The sample is the part of the population we actually observe.
6. (b) She wants to know which route is fastest, so this is the explanatory variable, or factor.
7. (d) In a blocked experiment, the subjects are randomly assigned treatments within their blocks (like subgroups). This ensures that each treatment is replicated within the blocks and will allow us to make conclusions about the effectiveness of each treatment in each block.
8. (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.
9. (d) The distribution is not binomial as the four conditions given in the text for the binomial setting are not satisfied. In particular, the four observations are not independent.
10. (b) As you are guessing, this can be modeled as a $B(100, 0.25)$ distribution.
11. (d) As the die rolls are independent, this can be modeled as a binomial distribution with $n = 12$. The probability p equals $\frac{1}{2}$ to get an even number, so the mean is $n * p = 12 * \frac{1}{2} = 6$.

12. (b) Because the coin is fair, the probability of heads is $1/2$. Because the die is fair, the probability of a six is $1/6$. Because the outcomes are independent, the multiplication rule tells us that the probability of heads and a six is the product of $1/2$ and $1/6$, namely $1/12$.
13. (a) This is the definition of randomness.
14. 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.
15. (c) We observe pairs of outcomes – and there are four different pairs.