

Statistics 2000, Section 001, Midterm 2 (185 Points)

Friday, March 25, 2011

Part I: Text Answers

Your Name: Solutions

Question 1: Two-Way Tables (48 Points)

In a recent Stat 2000 midterm, one question (Q2) seemed to have a high impact on the overall exam performance. A total of 40 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.

Below is the table that summarizes grade and Q2 performance:

	AB	C	DF	Total
Q2Well	8	2	0	10
Q2Poor	2	12	16	30
Total	10	14	16	40

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 (e.g., 10.5%).

Show your work!

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

1. (8 Points) What is the probability for a randomly selected student to do well on Q2 and still only obtain a D or F grade? Answer: 0.0 %

$$P(\text{Q2 well and DF}) = 0/40 = \underline{0\%} \quad (\text{does not happen})$$

2. (8 Points) What is the probability for a randomly selected student to do well on Q2 and obtain an A or B grade? Answer: 20.0 %

$$P(\text{Q2 well and AB}) = 8/40 = 0.2 = \underline{20\%}$$

3. (8 Points) What is the probability for a randomly selected student to obtain an A or B grade? Answer: 25.0 %

$$P(AB) = 10/40 = 0.25 = \underline{25\%}$$

4. (8 Points) Knowing that a randomly selected student does well on Q2, what is the probability for this student to obtain a C grade? Answer: 20.0 %

$$P(C | \text{Q2 well}) = 2/10 = 0.2 = \underline{20\%}$$

5. (8 Points) Knowing that a randomly selected student obtains a C grade, what is the probability for this student to do well on Q2? Answer: 14.3 %

$$P(\text{Q2 well} | C) = 2/14 = 0.143 = \underline{14.3\%}$$

6. (8 Points) What is the probability for two different randomly selected students to do well on Q2 and obtain an A or B grade each? Answer: 3.6 %

$$\begin{aligned} & P(\text{1st: Q2 well and AB and 2nd: Q2 well and AB}) \\ &= P(\text{1st: Q2 well and AB}) \cdot P(\text{2nd: Q2 well and AB} | \text{1st: Q2 well and AB}) \\ &= \frac{8}{40} \cdot \frac{7}{39} = \frac{56}{1560} = 0.036 = \underline{3.6\%} \end{aligned}$$

-2 for each calculation error
 (or no final result)
 +2 if correct formula only
 (but no values)

Question 2: Means and Variances of Random Variables (37 Points) (but no values)

Typographical and spelling errors can be either "nonword errors" or "word errors." A nonword error is not a real word, as when "the" is typed as "teh." A word error is a real word, but not the right word, as when "lose" is typed as "loose." When undergraduates are asked to write a 250-word essay (without spell-checking), the number of nonword errors (N) has the following distribution:

N :

Errors	0	1	2	3	4
Probability	0.2	0.2	0.3	0.2	0.1

$\mu_N = 1.8$ (from 1.)
 $\sigma_N = 1.249$ (given)

The number of word errors (W) has this distribution:

W :

Errors	0	1	2	3
Probability	0.5	0.2	0.2	0.1

$\mu_W = 0.9$ (given)
 $\sigma_W = 1.044$ (from 2.)

Show your work!

1. (9 Points) The average (mean) number μ_W of word errors in an essay of 250 words is 0.9. What is the average (mean) number μ_N of nonword errors in an essay of 250 words? Round your answer to 1 decimal digit (such as 0.9 for μ_W).

Answer: 1.8

$$\mu_N = \sum_{i=1}^5 x_i p_i = 0 \cdot 0.2 + 1 \cdot 0.2 + 2 \cdot 0.3 + 3 \cdot 0.2 + 4 \cdot 0.1 = \underline{1.8}$$
 (9)

2. (9 Points) The standard deviation σ_N of nonword errors in an essay of 250 words is about 1.249. What is the standard deviation σ_W of word errors in an essay of 250 words? Round your answer to 3 decimal digits (such as 1.249 for σ_N).

Answer: 1.044

$$\sigma_W^2 = \sum_{i=1}^4 (x_i - \mu_W)^2 \cdot p_i = (0 - 0.9)^2 \cdot 0.5 + (1 - 0.9)^2 \cdot 0.2 + (2 - 0.9)^2 \cdot 0.2 + (3 - 0.9)^2 \cdot 0.1$$

$$= 1.09$$
 (3)

$$\Rightarrow \sigma_W = \sqrt{1.09} = \underline{1.044}$$

3. (9 Points) Find the mean of the total number of errors (nonword errors plus word errors) in an essay of 250 words. Assume that it is not known here whether W and N are independent or not. Answer: 2.7

in any case:

$$\mu_{N+W} = \mu_N + \mu_W = 1.8 + 0.9 = \underline{2.7}$$
 (9)

4. (10 Points) Find the standard deviation of the total number of errors (nonword errors plus word errors) in an essay of 250 words. Let us assume that students who make many nonword errors also tend to make many word errors, so that the correlation between the two error counts is 0.4. Answer: 1.922

$$\sigma_{N+W}^2 = \sigma_N^2 + \sigma_W^2 + 2 \cdot \rho \cdot \sigma_N \cdot \sigma_W$$

$$= \underbrace{1.249^2 + 1.044^2}_{(2)} + \underbrace{2 \cdot 0.4 \cdot 1.249 \cdot 1.044}_{(5)} = 3.693$$

$$\Rightarrow \sigma_{N+W} = \sqrt{3.693} = \underline{1.922}$$
 (6)

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

Question 3: Probability (40 Points)

A box contains 3 red marbles and 7 blue marbles. For parts (1.) and (2.), assume we draw **WITH** replacement from the box. For parts (3.) and (4.), assume we draw **WITHOUT** replacement from the box. We draw a total of 2 marbles. Report your final answer as a percent with one decimal digit (e.g., 10.5%).

Show your work!

1. (10 Points) When we draw **WITH** replacement, what is the chance that both of the marbles are blue? Answer: 49.0 %

$$\begin{aligned} P(\text{both blue}) &= P(\text{1st blue and 2nd blue}) \\ &= \frac{7}{10} \cdot \frac{7}{10} = \frac{49}{100} = 0.49 = \underline{\underline{49\%}} \end{aligned}$$

(3) (2) (5)

2. (10 Points) When we draw **WITH** replacement, what is the chance that one of the marbles is blue and the other is red? Answer: 42.0 %

$$\begin{aligned} P(\text{one blue and one red}) &= P(\text{1st blue and 2nd red}) + P(\text{1st red and 2nd blue}) \\ &= \frac{7}{10} \cdot \frac{3}{10} + \frac{3}{10} \cdot \frac{7}{10} = \frac{21}{100} + \frac{21}{100} = \frac{42}{100} = 0.42 \\ &= \underline{\underline{42\%}} \end{aligned}$$

(3) (2) (5)

3. (10 Points) When we draw **WITHOUT** replacement, what is the chance that both of the marbles are blue? Answer: 46.7 %

$$\begin{aligned} P(\text{both blue}) &= P(\text{1st blue and 2nd blue}) \\ &= \frac{7}{10} \cdot \frac{6}{9} = \frac{42}{90} = 0.467 = \underline{\underline{46.7\%}} \end{aligned}$$

(1) (2) (7)

4. (10 Points) When we draw **WITHOUT** replacement, what is the chance that one of the marbles is blue and the other is red? Answer: 46.7 %

$$\begin{aligned} P(\text{one blue and one red}) &= P(\text{1st blue and 2nd red}) + P(\text{1st red and 2nd blue}) \\ &= \frac{7}{10} \cdot \frac{3}{9} + \frac{3}{10} \cdot \frac{7}{9} \\ &= \frac{21}{90} + \frac{21}{90} = \frac{42}{90} = 0.467 = \underline{\underline{46.7\%}} \end{aligned}$$

(3) (2) (5)

Statistics 2000, Section 001, Midterm 2 (185 Points)

Friday, March 25, 2011

Part II: Multiple Choice Questions

Your Name: _____

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	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	11	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	12	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	13	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	14	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
5	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	15	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
6	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>					
7	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>					
8	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>					
9	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>					
10	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>					

Stat 2000, Midterm 2, Question 4 — Solutions

1. (c) The correct numbers would be 5, 7, 4, 1 (this takes care of the men assigned to the treatment — Lewis, Simpson, Howard, Adams), and then 1, 3 (this takes care of the women assigned to the treatment — Braun, Miller).
2. (d) None of the above — all problems still exist for very large data sets.
3. (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.
4. 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.
5. (d) Confounding means that effects of two or more variables are mixed up.
6. (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.
7. (a) She wants to know which route is fastest, so this is the explanatory variable, or factor.
8. (c) Response variables are the outcomes of interest. Since she believes the driving time to school will depend on the route (which is the explanatory variable), driving time to school is the response variable.
9. (c) In an experiment, treatments are deliberately imposed. Here, she deliberately chose between the two routes. Experiments also have control: the comparison of two or more treatments (the routes); randomization (which route was selected based on the odometer); and replication (she repeated over 6 weeks). Therefore, this was an experiment. In an observational study, she would have decided herself which route to take (rather than using randomization based on the odometer reading).
10. (d) Yes! The purpose of randomization (choosing the route using the odometer) and replication (repeating for several weeks) is to eliminate any effects of lurking variables. Statistical significance means that we observed an effect so large that it would rarely occur by chance. Experiments give good evidence for causation, in this case that the difference in driving time to school is due to the route.
11. (d) In this study, the outcome, i.e., the response variable, is the accuracy the subjects drove the obstacle course.
12. (b) We have three different conditions. Those are called (factor) "levels".
13. (b) Condition (1) is the control. Younger drivers and elderly drivers probably won't get 10's for their driving due to lack of driving experience or age-related driving problems. So, the outcome for the control group tells us what is a common (average) performance without the additional task of texting.

14. (d) Participants with better driving skills will drive better and participants with better texting skills (e.g., faster and with one hand rather than two hands) will drive better — so both are lurking variables.
15. (d) Only when we properly randomize, we will see the effect of texting on the driving accuracy. In all other setups, any observed difference between the three groups easily could be contributed to lurking variables.