

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

Friday, March 25, 2011

Part I: Text Answers

Your Name: _____

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!

This page intentionally left blank!

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: _____ %

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: _____ %

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

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: _____ %

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: _____ %

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: _____ %

Question 2: Means and Variances of Random Variables (37 Points)

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:

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

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

Errors	0	1	2	3
Probability	0.5	0.2	0.2	0.1

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: _____

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: _____

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: _____

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: _____

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: _____ %

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: _____ %

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

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: _____ %

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

1. Twelve people who suffer from chronic fatigue syndrome volunteer to take part in an experiment to see if shark fin extract will increase one's energy level. Eight of the volunteers are men and four are women. Half of the volunteers are to be given shark fin extract twice a day and the other half a placebo twice a day. We wish to make sure that 4 men and 2 women are assigned to each of the treatments, so we decide to use a block design with the men forming one block and the women the other.

The names of the men and women are given in the table below and each name is given a numerical label. Use the list of random digits below the table to assign 4 men and 2 women to the shark fin treatment. Read the random digits from left to right, first selecting the 4 men and then the 2 women. Use the numerical labels given in the table left of the names to match names and random digits.

	Men		Women
1	Adams	1	Braun
2	Barnett	2	Coleman
3	Charles	3	Miller
4	Howard	4	Smith
5	Lewis		
6	Monaghan		
7	Simpson		
8	Taylor		

Random Digits

95754 11893 98734 02189

The people assigned to the shark fin treatment are

- (a) 5, 7, 5, 4, 1, 1.
 - (b) Lewis, Simpson, Howard, and Braun. We must give both Lewis and Braun a double dose since they are selected twice.
 - (c) Lewis, Simpson, Howard, Adams — Braun, Miller.
 - (d) Smith, Braun — Adams, Taylor, Charles, Simpson.
2. 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.

3. 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.

4. 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).

5. 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.

6. 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.

Use the following to answer questions 7, 8, 9, and 10:

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 driving time to school each day.

7. What is the explanatory variable in this study?
 - (a) The route.
 - (b) The odometer reading.
 - (c) The driving time to school.
 - (d) Her average driving speed.

8. What is the response variable in this study?
 - (a) The route.
 - (b) The odometer reading.
 - (c) The driving time to school.
 - (d) Her average driving speed.

9. This study is a(n) _____.
 - (a) Observational study.
 - (b) Sample.
 - (c) Experiment.
 - (d) Voluntary response survey.

10. At the end of the study, if she detects a significant difference in driving time to school, can she conclude it is due to the route?
 - (a) Definitely not. There are lurking variables such as traffic, road construction, etc.
 - (b) No. The result is meaningless because it is based on only 1 participant, the statistics instructor. There is no replication.
 - (c) Maybe. There are lurking variables.
 - (d) Yes.

Use the following to answer questions 11, 12, 13, 14, and 15:

To investigate whether or not sending text messages while driving impacts driving ability, we have 100 participants (50 men and 50 women) drive an obstacle course under one of the following conditions:

- 1) No texting while driving.
- 2) Sending five (5) text messages while driving.
- 3) Sending ten (10) text messages while driving.

We measure the accuracy the subjects drove the obstacle course on a scale from 1 (= poor) to 10 (= excellent).

11. What is the response variable in this study?
 - (a) Age of the participants.
 - (b) Gender of the participants.
 - (c) Actual number of text messages sent while driving the obstacle course.
 - (d) Accuracy rating from driving the obstacle course.
12. We are testing the effect of sending text messages at three different _____.
 - (a) treatments.
 - (b) levels.
 - (c) factors.
 - (d) none of the above.
13. This study includes which of the following?
 - (a) Blinding.
 - (b) Control.
 - (c) Lack of realism.
 - (d) Placebo.
14. What is a possible lurking variable in this experiment?
 - (a) Driving skill level of participants.
 - (b) Skill level at sending text messages.
 - (c) None of the above.
 - (d) Both of the above.

15. The best approach to assign participants to one of the three groups (where Group 1 represents condition 1, Group 2 represents condition 2, and Group 3 represents condition 3) is:
- (a) Assign all elderly participants (70 years of age and above) to Group 1 as those rarely send text messages anyway, assign participants aged 25 to 70 to Group 2 as they send text messages only occasionally, and assign participants younger than 25 years to Group 3 as they text most frequently.
 - (b) Ask participants in advance about how many miles they drove last year. Then use this estimate and assign the 33 participants who drove most to Group 3 (they are used to all kinds of distractions and can compensate), the 33 participants who drove least to Group 1 (they are unexperienced and better shouldn't be distracted by texting), and the remaining 34 participants to Group 2.
 - (c) Let the participants decide themselves how many text messages (0, 5, or 10) they feel comfortable to send — then let them do so while driving the obstacle course.
 - (d) Randomly split subjects into three groups, e.g., by using a random digit table. Then assign each group to one of the three conditions, i.e., whether they will have to send 0, 5, or 10 text messages while driving.