University of California, Berkeley, Statistics 21: Probability and Statistics for Business

Michael Lugo, Spring 2012

Exam 2

Wednesday, April 4, 2012, 12:10 - 1:00 PM

Name: Student ID:

Section (circle one):

202 (Seenae Chong, TR 9) 204 (Jose Gutierrez, TR 10)
205 (Seenae Chong, TR 11) 206 (Jose Gutierrez, TR 11)

This exam consists of seven pages: this cover page; five pages containing problems; and a table of the normal distribution. You may use a calculator, and notes on one side of a standard 8.5-by-11-inch (or A4) sheet of paper which you have written by hand, yourself.

Each question indicates whether it is necessary to show work or not. You may use a calculator to do arithmetic but you should indicate what you are asking the calculator to do; you should not use built-in calculator functions to compute standard deviations or correlation coefficients.

Write your name at the top of each page.

The total number of possible points is 38. The number of points for each problem are indicated in brackets.

DO NOT WRITE BELOW THIS LINE

<table>
<thead>
<tr>
<th>Question number:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Points possible:</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Your score:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question number:</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Points possible:</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Your score:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1
For questions 1 through 3, **show your work.** I roll three six-sided, fair dice. What is the probability that:

1. [2] All three dice show the same number (such as 4, 4, 4 or 2, 2, 2)?
   **Solution:** $1/6^2$. The first die can show any number; the remaining two dice then must show the same number as the first. (Alternatively, $6/6^3$; there are $6^3$ possible outcomes, of which 6 have all three dice showing the same number.

2. [2] All three dice show different numbers (such as 6, 2, 3 or 5, 4, 1)?
   **Solution:** This is like the birthday problem; $6/6 \times 5/6 \times 4/6 = 120/216 = 5/9$.

3. [2] Two dice show one number, and the remaining die shows another number (such as 5, 1, 5 or 2, 2, 3)?
   **Solution:** The easiest way to do this one is to note that the answers to 1, 2, and 3 must add up to 1; so we have $1 - (6 + 120)/216 = 90/216 = 5/12$.

For questions 4 and 5, consider drawing three cards from a special deck of twenty-four cards. This deck of cards contains four cards labeled with each number from 1 through 6. **Circle the answer for each; no explanation is necessary.**

4. [1] The probability that all three cards show the same number is
   - less than
   - equal to
   - greater than
   the answer to question 1.
   We’re sampling without replacement, so duplication becomes less likely. (The exact probability is $(4/24)(3/23)(2/22)$.)

5. [1] The probability that all three cards show different numbers is
   - less than
   - equal to
   - greater than
   the answer to question 2.
   Same reason as 4; the exact probability is $(24/24)(20/23)(16/22)$.

For questions 6 through 9, **show your work.** Consider the following situation. A multiple-choice test has 50 questions. Each question has five choices. The test is scored as follows: for each question there is

- one correct choice, worth 2 points
- one “half-correct” choice, worth 1 point
A student taking this test guesses at random on each question.

6. [2] What is the expected value of their score?

Solution. The expected score on a single question is \((2+1+0+0+0)/5 = 0.6\) and so the expected overall score is \((0.6)(50) = 30\).

7. [2] What is the standard error of their score?

Solution. The standard error on a single question is \(\sqrt{(2^2 + 1^2 + 0^2 + 0^2 + 0^2)/5 - (0.6)^2} = \sqrt{1 - 0.6^2} = 0.8\) and so the standard error on 50 questions is \(0.8\sqrt{50} \approx 5.7\).

8. [3] What is the probability that the student scores between 35 and 40 points, inclusive?

Solution. This is the area under the normal curve between \((34.5 - 30)/5.7\) and \((40.5 - 30)/5.7\), or approximately 0.80 and 1.85; from tables this is about 0.18.

9. [2] Consider only the first five questions. What is the exact probability that exactly two are answered correctly? (“Correctly” does not include the half-correct answers.)

Solution. Using the binomial formula, \(\binom{5}{2}(0.2)^2(0.8)^3 = 0.2048\).

For questions 10 through 13, show your work. In a certain town, there are 10,000 registered voters, of whom 2,000 are Democrats. A survey organization is about to take a simple random sample of 400 registered voters.

10. [1] What is the expected value for the percentage of Democrats in the sample?

Solution. \(2000/10000 = 0.20\), or 20 percent.

11. [2] What is the standard error for the percentage of Democrats in the sample, if the sample is taken with replacement?

Solution. \(\sqrt{(0.2)(0.8)/400} = 0.02\), or 2 percent.

12. [2] What is the chance that between 19 percent and 22 percent of the voters in the sample are Democrats? (Again, sampling with replacement.)

Solution. The area under the normal curve between \((19 - 20)/2 = -0.5\) and \((22 - 20)/2 = 1\); from tables, about 0.53.

13. [2] What is the standard error for the percentage of Democrats in the sample, if sampling is done without replacement?

Solution. Using the “correction factor”, we multiply the answer from 11 by \(\sqrt{(10000 - 400)/(10000 - 1)} \approx 0.98\) to get 1.96 percent.

14. [2] A nationwide telephone survey used random digit dialing. Out of 1,507 respondents, 3 percent said they had been homeless at some point in the last five years. The chance that the proportion of the entire national
population who have been homeless in the last five years is less than three percent is
much greater than one half  about equal to one half  much less than one half
Circle one, and explain briefly.

Those who have been homeless in the last five years have a greater than average chance of being homeless now, and even if they have homes are likely to be quite poor and may not have phones. Therefore we expect that more than three percent of the population is homeless; so the chance that less than three percent is homeless is less than one half.

15. [2] A simple random sample of 1,000 people in each state will be taken in order to estimate the proportion of the population which have college degrees. The accuracy to be expected in Massachusetts (population 6.6 million) is
about the same as  quite a bit higher than  quite a bit lower than
the accuracy to be expected in Connecticut (population 3.5 million).
Circle one, and explain briefly.
Sample accuracy depends quite strongly on the population size and very little on the sample size.

(A few people asked during the exam if the demographics of the two states were similar; they are, and I meant for you to make this assumption, but it’s a good question.)

16. [3] What is the approximate probability that a fair coin comes up heads 60 or more times when it is flipped 100 times? Show your work.

Solution. The number of heads has expected value $100/2 = 50$ and standard error $\sqrt{100(1/2)(1/2)} = 5$. So we need the area under the normal curve to the right of $(59.5 - 50)/5 = 1.9$, which is about 0.0287.

17. [2] I flip a fair coin 100 times and it comes up heads 60 times. I then flip the same coin 900 more times, for a total of 1,000. What is your best guess of the total number of heads obtained in all 1,000 flips? Explain briefly.

(The word “fair” wasn’t in the text originally but the correction was given at the start of the exam.)

Solution. We expect 450 more heads in the 900 coin flips; we guess $60 + 450 = 510$ heads.

18. [2] The average number of children born to each woman in some population, over the course of her lifetime, is 2.5. (This is the so-called total fertility rate). A simple random sample of this population is taken, and each person is asked how many children their mother had, including themselves.
(For example, an only child would say 1.) The average of the numbers given will be

much less than 2.5  about equal to 2.5  much greater than 2.5

Circle one, and explain briefly.

The larger families are overrepresented, so we expect the average of the numbers given by the children to be larger than the average of the numbers given by the mothers.

(But some people mentioned that depending on who you’re asking, their mothers may still be fertile and not done having children, which would work the other way.)

19. [3] In a certain class, twenty percent of students scored at least 90 on the midterm exam, and twenty percent of students scored at least 90 on the final exam. What proportion of students scored at least 90 on both exams?

(a) less than 4%  (b) 4%  (c) between 4% and 20%  (d) 20%  (e) more than 20%

Circle one, and explain briefly.

We expect there to be a positive correlation between the scores on the two exams. Four percent would score at least 90 on both exams if the two scores were independent, so we expect greater than four percent. But the scores are not perfectly correlated, so we expect that it’s not exactly the same students who broke 90 on the midterm and the final; so we expect less than twenty percent.