

Statistics 1040, Section 008, Midterm 2 (200 Points)

Friday, March 28, 2008

Your Name: _____

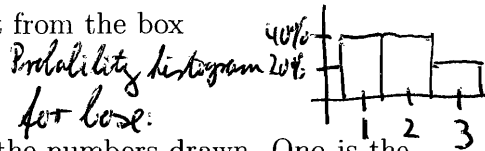
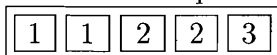
Instructions: Carefully check whether you have to provide an explanation or not. In case you have to provide an explanation, keep it short. Just 1 sentence (or 2 sentences at most) or a short calculation will be fine. If you do not have to provide an explanation, do not waste your time giving an unneeded explanation.

from: Stat 1040, Fall 2004; Midterm 2, Question 5

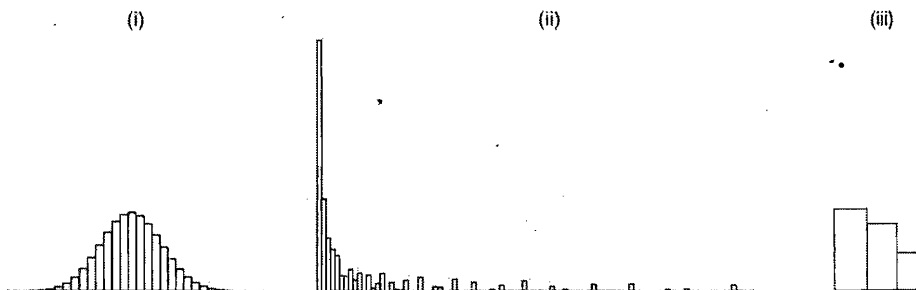
(Solutions → Course Web-Site)

Question 1: Normal Approximation for Probability Histograms (30 Points)

Twenty-five draws are made at random with replacement from the box



One of the graphs below is an (empirical) histogram for the numbers drawn. One is the probability histogram for the sum. And one is the probability histogram for the product. Which is which? **Explain!**



- An (empirical) histogram for the numbers drawn is (iii) (7).
 Explanation: *According to the probability histogram (see above), we should have 40% 1's, 40% 2's, and 20% 3's. The empirical histogram after 25 draws will somewhat resemble the probability histogram, but not too closely.* (3)
- The probability histogram for the sum is (i) (7).
 Explanation: *The probability histogram (see above) is not perfectly symmetric, but also not very asymmetric. The probability histogram for the sum will follow the normal curve even after only 25 draws.* (3)
- The probability histogram for the product is (ii) (7).
 Explanation: *The probability histogram for the product typically does not follow the normal curve.* (3)

based on: Stat 1040, Fall 2007, Final, Question 5

(Solutions → Course Web Site)

Question 2: Probability and Chance (40 Points)

For a road trip, a student places the following nine CDs into the glove compartment of his car:

- 5 modern rock CDs (Fallout Boy, Hawthorne Heights, The Used, Finger Eleven, Taking Back Sunday),
- 3 pop CDs (P!nk, Fergie, Gwen Stefani),
- 1 American Idol CD (Jordin Sparks).

On his trip, the student blindly grabs a CD from the glove compartment, listens to it, and places it on the back seat when finished. Then he blindly grabs a second CD from the glove compartment. You should NOT comment on the musical taste of this student, but answer each of the following questions separately. Show your work!

1. (8 Points) What is the chance that the FIRST CD will be a pop CD or the American Idol CD? The chance is 44.4 %

1st pop or (mutually exclusive) 1st Idol
 $\frac{3}{9} \text{ (3)} + \frac{1}{9} \text{ (3)} = \frac{4}{9} = \underline{0.444} = 44.4\%$

2. (8 Points) What is the chance that the SECOND CD will be a pop CD or the American Idol CD? The chance is 44.4 % (the same as in 1!)

2nd pop or (mutually exclusive) 2nd Idol
 $\frac{3}{9} \text{ (3)} + \frac{1}{9} \text{ (3)} = \frac{4}{9} = \underline{0.444} = 44.4\%$

3. (8 Points) What is the chance that he will listen to Jordin Sparks as one of his two selections? The chance is 22.2 %

(1st Jordin and 2nd other) or (mutually exclusive) (1st other and 2nd Jordin) 1st Jordin or (mutually exclusive) 2nd Jordin
 $\frac{1}{9} + \frac{8}{8} \text{ (3)} + \frac{8}{9} + \frac{1}{8} \text{ (3)} = \frac{1}{9} + \frac{1}{9}$

4. (8 Points) What is the chance that he will listen to none of the pop CDs? The chance is 41.7 %

1st not pop and 2nd not pop, given 1st not pop
 $\frac{6}{9} \text{ (3)} \cdot \frac{5}{8} \text{ (3)} = \frac{30}{72} = \underline{0.417} = 41.7\%$

5. (8 Points) What is the chance that he will listen to at least one of the modern rock CDs? The chance is 83.3 %

opposite rule 1st not modern rock and 2nd not modern rock, given 1st not modern rock
 $1 - \frac{4}{9} \text{ (2)} - \frac{3}{8} \text{ (2)} = \frac{72}{72} - \frac{12}{72} = \frac{60}{72} = \underline{0.833} = 83.3\%$

from: Stat 1040, Fall 2004, Midterm 2, Question 2

(Solutions → Course Website)

Question 3: EV, SE, and Normal Curve (50 Points)

During the 2004 presidential elections, Kerry needed to win the state of Ohio to become the next president. Early on Nov 3, the day after Election Day, Bush had a 51% to 49% lead over Kerry, which related to about 140,000 more votes for Bush in Ohio. However, there were possibly up to 250,000 uncounted provisional ballots at that time. If Kerry could have gotten 140,000 of those, plus 1/2 of the remaining 110,000, plus 1, i.e., a total of 195,001, he would have won Ohio and would have been the next president. However, Kerry eventually conceded to Bush later on Nov 3 (even with many of the provisional ballots still being uncounted) because Kerry's advisors figured out that it was *statistically impossible* for Kerry to win Ohio and thus the election. **Show your work!**

1. (10 Points) Assume you are a highly optimistic advisor of Kerry, assuming that he might win up to 70% of the uncounted provisional ballots because a huge majority of these votes come from a population group close to the Democrats. Find the box model.

$$\boxed{70 \times \boxed{1} \quad 30 \times \boxed{0}}$$

draws: 250,000

Kerry
O: Bush
-2 if slightly incorrect numbers of 0/1's
-4 if box given as $\boxed{0} \boxed{1}$, etc.
-6 if box contains something else than 0/1's
-2 if # draws missing or incorrect

2. (15 Points) The expected number of votes for Kerry from the uncounted provisional ballots is 175,000 with an SE of 229.

$$\text{box avg} = \frac{70}{100} = 0.7$$

$$\text{box SD} = \sqrt{\frac{70}{100} \cdot \frac{30}{100}} = \sqrt{0.21} = 0.458$$

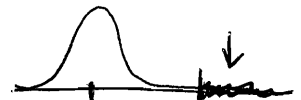
$$EV_{\text{sum}} = 250,000 \cdot 0.7 = \underline{175,000}$$

$$SE_{\text{sum}} = \sqrt{250,000} \cdot 0.458 = 500 \cdot 0.458 = \underline{229}$$

-2 for each calculation error
-2 for each minor mistake
-4 for each major mistake (or step missing)

3. (20 Points) The chance that at least 195,001 of the uncounted provisional ballots are in favor of Kerry is about 0.0 %.

$$\text{s.u.: } \frac{195,001 - 175,000}{229} = \frac{20,001}{229} = 87.34 \text{ (far off the table!)}$$



area between -4.45 and 4.45: 99.9991%

area between -87.34 and 87.34: almost exactly 100%

area above 87.34: basically 0.0%

-2 for each calculation error
-4 for each incorrect curve parameter, i.e., anything else than EV & SE
-4 for incorrect s.u.
-4 for incorrect table value (i.e., not 100%)
-4 for incorrect area under the curve (i.e., not 0%)

4. (5 Points) So, do you agree that it was statistically impossible for Kerry to win Ohio and thus the election? Yes No (5) - needs to match part 3.

Yes, we can really say that such an outcome is statistically impossible; moreover, our assumption that 70% of these votes may go to Kerry is already highly optimistic, conceding was the best he could do from a statistical point of view.

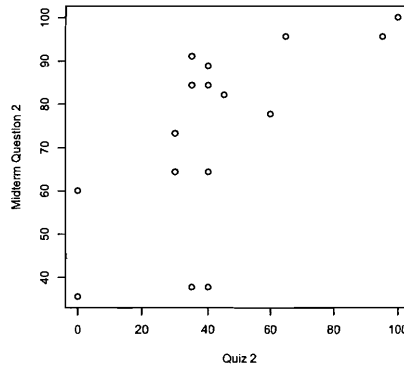
New Question!

Question 4: Regression (50 Points)

In a particular section of Stat 1040, students had to answer Review Exercise 2 from Chapter 3 of their textbook in Quiz 2. The result was anything but satisfactory, with the median score being an F. Detailed solutions were handed out, together with the graded quizzes. To determine whether students studied the solutions, the instructor basically reused the same question (with some part added) a few weeks later as Question 2 in Midterm 1. For a better comparability, the scores below were adjusted to 100 points.

- ✓ Midterm Question 2 score: avg = 73 points; SD = 21 points;
- ✗ Quiz 2 score: avg = 43 points; SD = 27 points; $r = 0.65$.

The scatterplot that shows the data is displayed below and can be assumed to be football-shaped.



Show your work!

-2 for each calculation error
 -2 if x, y flipped
 -2 if x, y not specified

1. (15 Points) Find the regression equation for predicting the Midterm Question 2 score from the Quiz 2 score.

$$\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.65 \cdot \frac{21}{27} = 0.51 \quad (6)$$

$$\text{intercept} = \text{avg}_y - \text{slope} \cdot \text{avg}_x = 73 - 0.51 \cdot 43 = 51.1 \quad (6)$$

$$\text{regression equation: } \boxed{\text{Midterm Q2 score} = 51.1 + 0.51 \cdot \text{Quiz 2 score}} \quad (3)$$

$$\text{or } \boxed{y = 51.1 + 0.51 \cdot x}$$

2. (8 Points) Using your regression equation, estimate the Midterm Question 2 score for a student who had a Quiz 2 score of 80 points.

predicted Midterm Q2 score for someone with 80 points in Quiz 2 =

$$51.1 + 0.51 \cdot 80 = \underline{91.9} \text{ points} \quad (8)$$

-2 for old method, correct result
 -7 for old method, incorrect result
 -5 if result makes no sense at all

3. (7 Points) Find the r.m.s. error for predicting the Midterm Question 2 score from the Quiz 2 score.

$$\begin{aligned} \text{r.m.s. error} &= \sqrt{1-r^2} \cdot SD_y \\ &= \sqrt{1-0.65^2} \cdot 21 \\ &= 0.76 \cdot 21 \\ &= \underline{15.96} \text{ points } (7) \end{aligned}$$

-3 for each major mistake,
e.g. SD_x instead of SD_y ,
 $\sqrt{\quad}$ of everything,
 r instead of r^2 , etc.

4. (10 Points) Would you be surprised if a student with 85 points in Quiz 2 would have obtained a Midterm Question 2 score of 40 points? YES or NO? Circle your answer and provide a short explanation. (4)

predicted Midterm Q2 score for someone with 85 points in Quiz 2 =
 $51.1 + 0.51 \cdot 85 = 94.45$ points (2)

observed Midterm Q2 score = 40 points

$$S.E. = \frac{\text{observed} - \text{predicted}}{\text{r.m.s. error}} = \frac{40 - 94.45}{15.96} = -3.41$$

This is more than the 3 r.m.s. error band away from the regression line; so, yes, this would be quite surprising. (4)

5. (10 Points) As mentioned above, all scores were adjusted as if graded out of 100 points. However, the Quiz 2 scores were originally graded out of 20 points, that means, each individual Quiz 2 score was multiplied by 5 for this question. Therefore, we had an original average score of 8.6 points and an original SD of 5.4 points when grading out of 20 points. (4)

As each point score initially was multiplied by 5, we now have to divide by 5. Therefore:

$$\text{original avg} = \frac{\text{reported avg}}{5} = \frac{43}{5} = \underline{8.6} \text{ points}$$

$$\text{original SD} = \frac{\text{reported SD}}{5} = \frac{27}{5} = \underline{5.4} \text{ points}$$

From: FPP, Chapter 20, p. 372, Review Exercise 7
 & Stat 1040, Spring 2006, Midterm 2, Question 5
 Question 5: Chance Errors in Sampling (30 Points)

(Solutions → Workbook)
 & → Course Website

Calculation:
 (not required)

$$\text{box avg} = \frac{20,000}{80,000} = \frac{1}{4}$$

$$\text{box SD} = \sqrt{\frac{20,000 \cdot 20,000}{80,000} + \frac{60,000 \cdot 60,000}{80,000}}$$

$$= 0.433$$

Five hundred draws are made at random from the box

$$60,000 \times \boxed{0} \quad 20,000 \times \boxed{1}$$

True or false? Circle your answers. No explanation is needed.

1. (5 Points) True / false: The expected value for the percentage of 1's among the draws is exactly 25%.

see calculation:

$$EV_{\%} = 25\%$$

$$EV_{\text{sum}} = 500 \cdot \frac{1}{4} = 125$$

$$SE_{\text{sum}} = \sqrt{500} \cdot 0.433 = 9.68$$

$$EV_{\%} = 25\%$$

$$SE_{\%} = \frac{9.68}{500} \cdot 100\% = 1.94\% \approx 2\%$$

2. (5 Points) True / false: The expected value for the percentage of 1's among the draws is around 25%, give or take 2% or so.

we know exactly the expected value for the percentage of $\boxed{1}$'s among the draws (which is 25% - no give or take)

3. (5 Points) True / false: The percentage of 1's among the draws will be around 25%, give or take 2% or so.

see calculation:
 (close to $EV_{\%} = 25\%$, but give or take of about $SE_{\%} = 2\%$)

4. (5 Points) True / false: The percentage of 1's among the draws will be exactly 25%.

the percentage of $\boxed{1}$'s most likely will not be exactly 25% (but it will be relatively close to 25%)

5. (5 Points) True / false: The percentage of 1's in the box is exactly 25%.

see calculation:

$$\text{box avg} = \text{fraction of } \boxed{1}\text{'s} = 25\%$$

6. (5 Points) True / false: The percentage of 1's in the box is around 25%, give or take 2% or so.

we know exactly the percentage of $\boxed{1}$'s in the population (i.e., box) (which is 25% - no give or take)