Statistics 1040, Section 003, Midterm 2 (200 Points)

Friday, November 7, 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.**

based on: Quiz 9, Fall 2008, Question 1

(new values; solution outline provided in class!)

Question 1: Probability Histograms (**40 Points**)

Four hundred draws will be made at random with replacement from the box

1	3	5	7	9

. **Show your work!**

-2 for each calculation error

draws: 400

1. (**20 Points**) Estimate the chance that the sum of the draws will be more than 1,500.

   The chance is: about 100 %

box avg $= \frac{1+3+5+7+9}{5} = \frac{25}{5} = 5$ (3)

box SD $= \sqrt{\frac{(1-5)^2+(3-5)^2+(5-5)^2+(7-5)^2+(9-5)^2}{5}}$

$= \sqrt{8} = 2.83$ (3)

$EV_{sum} = 400 \cdot 5 = 2000$ (3)

$SE_{sum} = \sqrt{400} \cdot 2.83 = 56.6$ (3)

1,500 2000

-8.83 0 s.u.

s.u.: $\frac{1,500-2,000}{56.6} = -8.83$ (4)

area between -8.83 and 8.83: about 100%

area above -8.83: about 100% (4)

2. (**20 Points**) Estimate the chance that there will be fewer than 90 [3]'s.

   The chance is: 89.44 %

new box with: 1: [3]

0: [1], [5], [7], or [9]

1× [1]	4× [0]

(8)

draws: 400

box avg $= \frac{1}{5} = 0.2$ (2)

box SD $= \sqrt{\frac{1}{5} \cdot \frac{4}{5}} = \frac{2}{5} = 0.4$ (2)

$EV_{sum} = 400 \cdot 0.2 = 80$ (2)

$SE_{sum} = \sqrt{400} \cdot 0.4 = 8$ (2)

80 90

0 1.25 s.u.

s.u.: $\frac{90-80}{8} = 1.25$ (2)

area between -1.25 and 1.25: 78.87%

area below 1.25: $50\% + \frac{78.87\%}{2}$

$= 89.435\%$ (2)

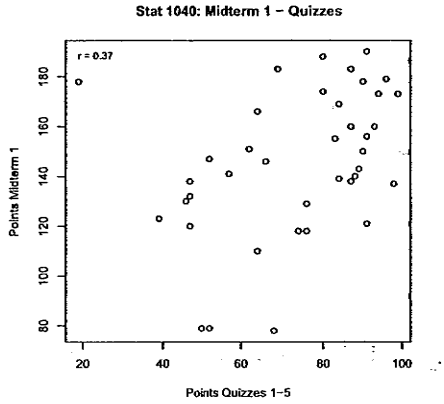
from: Stat 1040, Fall 2007, Midterm 2, Question 2 (Solutions: → Course Web Page)

Question 2: Regression (**50 Points**)

In a recent section of Stat 1040, the following scores for the sum of the first five quizzes and the first midterm were observed:

x Quiz 1-5 score: avg = 73 points; SD = 19 points;
y Midterm 1 score: avg = 145 points; SD = 29 points; $r = 0.37$.

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



Show your work!

−2 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 1 score from the Quiz 1-5 score.

$\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.37 \cdot \frac{29}{19} = 0.56$ (6)

$\text{intercept} = \text{avg}_y - \text{slope} \cdot \text{avg}_x = 145 - 0.56 \cdot 73 = 104.1$ (6)

regression equation: $\boxed{\text{Midterm 1 score} = 104.1 + 0.56 \cdot \text{Quiz 1-5 score}}$ (3)

or: $\boxed{y = 104.1 + 0.56 \cdot x}$

2. (**8 Points**) Using your regression equation, estimate the Midterm 1 score for a student who had a Quiz 1-5 score of 60 points.

predicted Midterm 1 score for someone with 60 points in the quizzes =

$104.1 + 0.56 \cdot 60 = \underline{137.7}$ points (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 1 score from the Quiz 1-5 score.

$$\text{r.m.s. error} = \sqrt{1-r^2} \cdot SD_Y$$
$$= \sqrt{1-0.37^2} \cdot 29$$
$$= \underline{\underline{26.9}} \text{ points} \quad (7)$$

−3 for each major mistake, e.g. SD_X instead of SD_Y, $\sqrt{}$ of everything, r instead of r^2 etc.

4. **(10 Points)** Can we use the regression equation to predict the Midterm 1 score for a student who had a Quiz 1-5 score of 19 points? **YES** or **NO**? **Circle your answer and provide a short explanation.** (4)

$$\text{S.U.} = \frac{19-73}{19} = -2.84$$

This is within 3 SDs of the average where 99.7% of the data are located; this is not a problem and also no extrapolation (and actually we have an observation with a Quiz 1-5 score of 19 points: see the scatterplot) (6)

5. **(10 Points)** Independently from your previous answer, let us assume that we can use the regression equation to predict the Midterm 1 score for a student who had a Quiz 1-5 score of 19 points. Would you be *surprised* that a student with 19 points in the quizzes got a score of 178 points in Midterm 1? **YES** or **NO**? **Circle your answer and provide a short explanation.** (4)

predicted Midterm 1 score for someone with 19 points in the quizzes:

$$104.1 + 0.56 \cdot 19 = 114.7 \text{ points} \quad (2)$$

observed Midterm 1 score = 178 points

$$\text{S.U.} = \frac{\text{observed} - \text{predicted}}{\text{r.m.s. error}} = \frac{178 - 114.7}{26.9} = 2.35$$

This is within the 3 r.m.s. error band of the regression line where 99.7% of the data are located; even though this point looks like an outlier in the scatterplot, the observed score of 178 is still not too surprising in the framework of the overall data (note the large r.m.s. error of almost 27 points!). By the way, a score of 196 points (or better) would have been really surprising for someone with 19 points in the quizzes. (4)

based on: Stat 1040, Spring 2008, Final, Question 7

(new values; solution outline → Course Web Page)

Question 3: Probability and Chance (**50 Points**)

A class of 28 fourth–graders has 16 boys and 12 girls. This class goes on a field trip. Two children are chosen at random to ride with the teacher.

Answer each of the following questions separately. Show your work!

-2 for each calculation error

1. (**10 Points**) What is the chance the first child is a boy?

   The chance is 57.14 %

   1st boy

   $\frac{16}{28} = 0.5714 = 57.14\%$ (10)

2. (**10 Points**) What is the chance the second child is a boy?

   The chance is 57.14 %

   2nd boy

   $\frac{16}{28} = 0.5714 = 57.14\%$ (10)

   [alternatively: (1st boy and 2nd boy) or (1st girl and 2nd boy)

   $\left(\frac{16}{28} \cdot \frac{15}{27}\right) + \left(\frac{12}{28} \cdot \frac{16}{27}\right)$

   $= \frac{432}{756} = 0.5714$ (as before)]

3. (**10 Points**) What is the chance both children are boys?

   The chance is 31.75 %

   1st boy and 2nd boy

   $\frac{16}{28}$ (2) $\cdot$ (3) $\frac{15}{27}$ (5) $= \frac{240}{756} = 0.3175 = 31.75\%$

4. (**10 Points**) What is the chance neither of the children are boys?

   The chance is 17.46 % = chance both are girls:

   1st girl and 2nd girl

   $\frac{12}{28}$ (3) $\cdot$ (3) $\frac{11}{27}$ (4) $= \frac{132}{756} = 0.1746 = 17.46\%$

5. (**10 Points**) What is the chance one of the children is a boy and the other is a girl?

   The chance is 50.79 %

   = 1 − both boys − both girls

   = 1 − 0.3175 − 0.1746 = 0.5079

   = 50.79 %

   [alternatively: (1st boy and 2nd girl) (2) (4) or (1st girl and 2nd boy) (4)

   $\left(\frac{16}{28} \cdot \frac{12}{27}\right) + \left(\frac{12}{28} \cdot \frac{16}{27}\right) = \frac{384}{756} = 0.5079$ (as above)]

from: FPP, Chapter 16, Review Exercise 6 & 9, p. 286 (Solutions: → Worklook)

Question 4: The Law of Averages (**20 Points**)

Answer the following two questions. **Just circle your correct answer!**

1. (**10 Points**) According to genetic theory, there is very close to an even chance that both children in a two–child family will be of the same sex. Here are two possibilities:

   (a) 15 couples have two children each. In 10 or more of these families, it will turn out that both children are of the same sex.

   (b) 30 couples have two children each. In 20 or more of these families, it will turn out that both children are of the same sex.

   Choose the correct statement among the following statements:

   10 (i) Possibility (a) is more likely.

   (ii) Possibility (b) is more likely.

   (iii) Possibilities (a) and (b) are equally likely.

   Reason: $\frac{10}{15} = \frac{20}{30} = \frac{2}{3}$. Option (a) is like tossing a coin 15 times, and asking for $\frac{2}{3}$ or more heads. Option (b) ups the number of tosses to 30. With the bigger number of tosses, you are less likely to get $\frac{2}{3}$ or more heads.

2. (**10 Points**) A box contains red and blue marbles; there are more red marbles than blue ones. Marbles are drawn one at a time from the box, at random with replacement. You win a dollar if a red marble is drawn more often than a blue one. There are two choices:

   (a) 100 draws are made from the box.

   (b) 200 draws are made from the box.

   Choose one of the four options below:

   (i) Choice (a) gives a better chance of winning.

   10 (ii) Choice (b) gives a better chance of winning.

   (iii) Choices (a) and (b) give the same chance of winning.

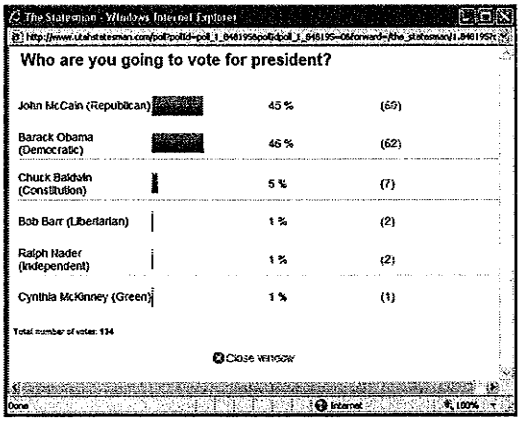
   (iv) Can't tell without more information.

   With more draws, the percentage of reds is likely to be closer to the percentage in the box, therefore, above 50%. (When the percentage of reds among the draws is above 50%, more reds are drawn than blues, and you win the dollar.

Next Question. (no solutions available)

Question 5: Chance Errors in Sampling / Accuracy of Percentages (**40 Points**)

The online edition of the Utah Statesman reported the following intermediate standings (on 10/31/08 at 10.15pm) regarding their question "Who are you going to vote for president?":



-2 if Boz listed (no bootstrap!)
-2 for each calculation error

1. (**20 Points**) For this question part only, assume that the 134 students who replied to this survey represent a simple random sample (SRS) of all USU students. Using the idea of bootstrap, one could estimate that the percentage of Obama voters (a total of 62 students in this sample) in the whole USU student population is about 46 %, give–or–take 4.3 %. **Show your work!**

$\text{prop. \%} = \text{sample \%} = \frac{62}{134} = 0.463 = 46.3\% \approx 46\%$ (5)

$SD_{Boz} = \sqrt{\frac{46}{100} \cdot \frac{54}{100}} = 0.498 \approx 0.5$ (5) (via bootstrap!)

$SE_{sum} = \sqrt{134} \cdot 0.5 = 5.78$ (5) | $SE_{\%} = \frac{5.78}{134} \cdot 100\% = 4.3\%$ (5) = give-or-take

2. (**5 Points**) Unfortunately, this survey is not based on a SRS and, therefore, it does not provide a statistically reliable answer (note that many surveys of this type indicate that the survey is not a scientific survey). This particular survey is one example of a voluntary response (5) survey. ((4) if just "voluntary")

3. (**15 Points**) Provide **three** reasons why a Web survey such as the Utah Statesman survey cited above is likely to provide invalid results that do **not** relate to the entire population (here: all USU students). (5) for each valid reason

- only students who read the online version of the Statesman can vote
- voters may include USU faculty and non-USU students who access this Web page by chance (and these do not belong to the target population of all USU students)
- people may vote more than once
- people with strong opinions (in favor of one of the candidates) are more likely to vote