

# Statistics 1040, Section 009, Midterm 2 (200 Points)

Friday, March 31, 2006

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) will be fine. If you do not have to provide an explanation, do not waste your time giving an unneeded explanation.

*from: FPP, Chapter 19, p. 351, Question 5 [Answer: → Workbook]*

## Question 1: Sampling (20 Points)

(Hypothetical) A survey is carried out by the finance department to determine the distribution of household size in a certain city. They draw a simple random sample of 1,000 households. After several visits, the interviewers find people at home in only 653 of the sample households. Rather than face such a high non-response rate, the department draws a new batch of households, and uses the first 347 completed interviews in the second batch to bring the sample up to its planned strength of 1,000 households. The department counts 3,087 people in these 1,000 households, and estimates the average household size in the city to be about 3.1 persons. Is this estimate likely to be too low, too high, or about right?

12

Circle your answer and explain clearly!

*Workbook: "Larger households are more likely to have someone home, so it's like the sample is substituting larger households for smaller ones."*

8

from: Stat 1040, Spring 2005, Midterm 2, Question 2, Part 1

**Question 2: Probability (40 Points)**

Professor J.S. recently experimented with homegrown chili plants. From 4 seeds that were planted, one seed germinated. Use this chance (25%) as the basis for all calculations in this question and assume that seeds germinate independently from each other.

Show your work!

-2 for each calculation error  
(or no final result in %)  
-10 if % > 100% or % < 0%

In a second experiment, J.S. planted 4 more seeds. Determine the following chances:

1. (15 Points) The chance that all 4 of these 4 seeds will germinate is about 0.39 %.

first germinates:  $\frac{1}{4}$   
second germinates:  $\frac{1}{4}$   
third germinates:  $\frac{1}{4}$   
fourth germinates:  $\frac{1}{4}$   
all 4 germinate:  $\frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \left(\frac{1}{4}\right)^4 = 0.0039 = \underline{\underline{0.39\%}}$

independent (8)  
multiplication rule (7)

-6 if not using  $\frac{1}{4}$  in this part

2. (15 Points) The chance that none of these 4 seeds will germinate is about 31.64 %.

first does not germinate:  $\frac{3}{4}$   
second does not germinate:  $\frac{3}{4}$   
third does not germinate:  $\frac{3}{4}$   
fourth does not germinate:  $\frac{3}{4}$   
all 4 do not germinate:  $\frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \left(\frac{3}{4}\right)^4 = 0.3164 = \underline{\underline{31.64\%}}$

independent (8)  
multiplication rule (7)

-6 if using  $\frac{1}{4}$  in this part

3. (10 Points) The chance that at least 1 of these 4 seeds will germinate is about 68.36 %.

"at least 1" is opposite of "none":

at least 1 germinates:  $1 - \left(\frac{3}{4}\right)^4 = 1 - 0.3164 = 0.6836 = \underline{\underline{68.36\%}}$

opposite rule (9)

from: Stat 1040, Fall 2001, Final Test, December 13, 2001, Question 2 b & c  
 & Stat 1040, Spring 2005, Mitterm 2, Question 3  
 Question 3: Regression (40 Points)

A selection of 65 varieties of cereal were tested for calories and sodium (in milligrams) for an one-cup serving. The results can be summarized as follows:

- ✓ Average sodium = 240 mg; SD = 131 mg;
- ✗ Average calories = 149 calories; SD = 62 calories;  $r = 0.53$ .

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

Show your work!

1. (10 Points) Find the equation of the regression line for predicting number of mg sodium in an one-cup serving of cereals from calories.

$$\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.53 \cdot \frac{131}{62} = 1.12 \quad (4)$$

$$\text{intercept} = \text{avg } y - \text{slope} \cdot \text{avg } x = 240 - 1.12 \cdot 149 = 240 - 166.88 = 73.12 \quad (4)$$

regression equation:  $\boxed{\text{sodium} = 73.12 + 1.12 \cdot \text{calories}}$  (2)  
 or  $\boxed{y = 73.12 + 1.12 \cdot x}$

2. (10 Points) Predict the number of mg sodium in an one-cup serving of cereals that has 200 calories per cup.

for 200 calories:  $\text{sodium} = 73.12 + 1.12 \cdot 200 = 73.12 + 224 = \underline{\underline{297.12}}$

-2 for old method, correct result  
 -8 for old method, incorrect result  
 -5 of result makes no sense at all

3. (10 Points) Find the r.m.s. error for predicting mg sodium from calories.

$$\begin{aligned} \text{r.m.s. error} &= \sqrt{1-r^2} \cdot SD_y &= \sqrt{0.7191} \cdot 131 && -4 \text{ for each major mistake,} \\ &= \sqrt{1-0.53^2} \cdot 131 &= 0.848 \cdot 131 && \text{e.g. } SD_x \text{ instead of } SD_y, \\ &= \sqrt{1-0.2809} \cdot 131 &= \underline{\underline{111.1}} && \sqrt{\text{of everything}}, r \text{ instead} \\ & & && \text{of } r^2 \text{ etc.} \end{aligned}$$

4. (10 Points) Explain why it would not be a good idea to use the information in the question to estimate the amount of sodium for a cereal with 350 calories per cup.

$$350 \text{ calories: } \frac{350-149}{62} = 3.2 \text{ s.u.}$$

350 is more than 3 s.u. above the average (of 149 calories);  
 this is extrapolation (5) and the result will be meaningless  
 (perhaps this cereal contains big chocolate chips that are high in calories,  
 but may be low on sodium)

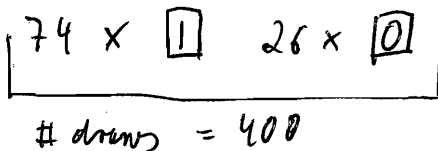
**Question 4: EV, SE, Normal Curve & Sampling (70 Points)**

Time reported in its March 20, 2006, issue on page 30: "74% Proportion of female college students and graduates who said women on spring-break trips use drinking as an excuse for behavior like public display of nudity and table dancing."

Assume that 74% indeed is the true percentage of all female college students and graduates in the US who share this opinion. Suppose an independent researcher wants to do some follow-up study and draws a simple random sample of 400 female college students and graduates in the US.

Show your work!

1. (10 Points) Indicate the box model.



-2 for each calculation error  
 -3 for slightly incorrect number of 1's/0's in box  
 -5 if box given as 1 0 etc.

1: shares this opinion

0: does not share this opinion

-7 if box contains something else than 1/0's

-2 if # draws missing or incorrect

2. (20 Points) The expected number of these 400 females in the follow-up study who share the opinion given above is 296 with a standard error of about 8.8.

$$\text{box avg} = \text{fraction of 1's} = \frac{74}{100} = 0.74$$

$$\text{box SD} = \sqrt{\text{fraction of 1's} \cdot \text{fraction of 0's}} = \sqrt{\frac{74}{100} \cdot \frac{26}{100}} = \sqrt{0.74 \cdot 0.26} = \sqrt{0.1924} = 0.44$$

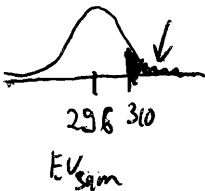
$$EV_{\text{sum}} = 400 \cdot 0.74 = 296$$

$$SE_{\text{sum}} = \sqrt{400} \cdot 0.44 = 20 \cdot 0.44 = 8.8$$

-2 for each minor mistake

-5 for each major mistake (or step missing)

3. (20 Points) The chance that at least 310 of these 400 females in the follow-up study who share the opinion given above is about 5.48 %.



$$s.u.: \frac{310 - 296}{8.8} = \frac{14}{8.8} = 1.59$$

$$\text{area between } -1.60 \text{ and } 1.60: 89.04\%$$

$$\text{area above } 1.60: \frac{100\% - 89.04\%}{2} = 5.48\%$$

-4 for each incorrect curve parameter (i.e., anything else than  $EV_{\text{sum}}$  &  $SE_{\text{sum}}$ )

-4 for incorrect s.u.  
 -4 for incorrect standard value

-4 for incorrect area under the curve

4. (10 Points) (Hypothetical) Suppose *Time* would have asked the female college students and graduates in their study: "On your spring-break trips, do you use drinking as an excuse for behavior like public display of nudity and table dancing?"

Other things being equal, the percentage of women who would have answered yes to this question would have been (a) higher than 74%, (b) about 74%, or

6

(c) much less than 74%.

Circle your answer and explain clearly!

*This new question uses strong personal wording. While it is easy to assume other women show such a behavior, hardly any woman would actually admit that she behaves this way.*

4

5. (10 Points) (Hypothetical) Suppose that you were asked by the independent researcher to conduct this follow-up study for him. It is up to you to determine how to draw a sample of 400 females that are representative for all female college students and graduates in the US and obtain their opinion regarding the question originally asked by *Time*. If you have the choice, the best possible way to draw this sample is:

- (a) Travel to South Padre Island in Texas at the start of Spring Break, go to the beach, ask women (as they arrive on the beach) whether they are college students or graduates, and then ask the first 400 of those for their opinion.
- (b) To avoid travel, get a list of all current female USU students and female USU graduates, draw a simple random sample of 400 female USU students/graduates, contact them by phone, cell phone, or at home, and ask them for their opinion.
- (c) Get lists of current female students and female graduates from all colleges (and universities) in the US, draw a simple random sample of 400 female students/graduates, contact them by phone, cell phone, or travel across the country if necessary to reach them at home (this may take weeks!), and ask them for their opinion.

6

Circle your answer and explain clearly!

*The only valid SRS that represents all female college students and graduates in the US is described in (c). Option (a) is biased towards female college students/graduates that travel to one particular resort, while option (b) is only representing the opinions of female USU students/graduates (and these are not representative for all female US students/graduates).*

4

P.S.: If you have never been to South Padre Island or a similar location during Spring Break, take a look at Web pages such as <http://www.spadre.com/springbreak.htm> and think again whether the percentage reported in *Time* might be true or whether it is totally far-fetched...

from: FPP Chapter 20, p. 372, Question 7  
 & Stat (040, Spring 2004, Midterm 2, Question 5)  
 Question 5: Chance Errors in Sampling (30 Points)

[Answers → Workbook]

Calculation: (not required)

$$\text{box avg} = \frac{20,000}{80,000} = \frac{1}{4} = 0.25 (= 25\%)$$

$$\text{box SD} = \sqrt{\frac{20,000}{80,000} \cdot \frac{60,000}{80,000}} = \sqrt{\frac{1}{4} \cdot \frac{3}{4}}$$

$$= \sqrt{\frac{3}{16}} = 0.433$$

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

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

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

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

Five hundred draws are made at random from the box

60,000 ×  0    20,000 ×  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:

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

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 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 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:

(box avg = fraction of 1'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 1's in the population (i.e., box) (which is 25% - no give or take)