Statistics 1040, Section 006, Midterm 2 (200 Points)

November 8, 2002

Your Name: _____

From: Freedman, Pisani, Purves, Chapter 17, page 305, question 7,

Question 1: The Expected Value and Standard Error **(40 Points)**

A hundred draws are made at random with replacement from the box

1	2	3	4	5	6

.

1. **(5 Points)** If the sum of the draws is 321, what is their average?

$$\text{avg} = \frac{\text{sum of draws}}{\#\text{draws}} = \frac{321}{100} = 3.21 \quad (5)$$

2. **(5 Points)** If the average of the draws is 3.78, what is the sum?

$$\text{sum of draws} = \text{avg} \cdot \#\text{draws} = 3.78 \cdot 100 = 378 \quad (5)$$

3. **(30 Points)** Estimate the chance that the average of the draws is between 3 and 4.
"Show your work." (Added on board before exam.)

Worklook:

(c) The average will be between 3 and 4 if the sum is between 300 and 400. The average of the box is 3.5, and the SD is 1.7, so the expected value of the sum is 350 and the standard error is 17. The chance is about 99.7%.

Additional details:

avg of 3 ⇝ sum of 300 (4)

avg of 4 ⇝ sum of 400 (4)

$$\text{box avg} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5 \quad (3)$$

$$\text{box SD} = \sqrt{\frac{(1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2}{6}}$$

$$= \sqrt{\frac{17.5}{6}} = \sqrt{2.917} = 1.7 \quad (3)$$

$$EV_{sum} = 100 \cdot 3.5 = 350 \quad (3)$$

$$SE_{sum} = \sqrt{100} \cdot 1.7 = 17 \quad (3)$$

$$\text{S.U.: } \frac{300-350}{17} = -2.94 \quad (3)$$

$$\frac{400-350}{17} = 2.94 \quad (3)$$

area between −2.95 to 2.95:

99.68% ≈ 99.7% (4)

From: Stat 1040, Fall 1999, Final Test, December 17, 1999. Question 6 a, b, c, d (extended)

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

In a box of 12 chocolates, three are mint, two are orange, five are caramel, and two are cherry. I choose two chocolates at random (without replacement!). **Answer each of the following questions separately.**

(-2) each calculation error

1. (**5 Points**) What is the chance that the first chocolate is mint?

chance first is mint: $\frac{3}{12} = \frac{1}{4} = 0.25 = 25\%$ (5)

2. (**10 Points**) What is the chance that the first two chocolates are both mint? (5) correct rule

chance first is mint: $\frac{3}{12}$ (1)

chance second is mint, given that first is mint: $\frac{2}{11}$ (4)

chance both are mint: $\frac{3}{12} \cdot \frac{2}{11} = \frac{6}{132} = \frac{1}{22}$

$= 0.04545 = 4.5\%$

3. (**10 Points**) What is the chance that the first chocolate is mint and the second is orange?

chance first is mint: $\frac{3}{12}$ (1)

chance second is orange, given that first is mint: $\frac{2}{11}$ (4)

(5) correct rule

chance first is mint & second is orange: $\frac{3}{12} \cdot \frac{2}{11} = 0.04545 = 4.5\%$ (as in 2.)

4. (**15 Points**) If I only like caramel, what is the chance that I choose at least one chocolate I like?

chance first is not caramel: $\frac{7}{12}$ (3)

chance second is not caramel, given that first is not caramel: $\frac{6}{11}$ (4)

(4) Correct rule

chance both are not caramel: $\frac{7}{12} \cdot \frac{6}{11} = \frac{42}{132} = \frac{7}{22}$

chance at least one is caramel: $1 - \frac{7}{22} = \frac{22}{22} - \frac{7}{22} = \frac{15}{22} = 0.6818 = 68.2\%$

(4) correct rule

Question 3: Regression **(50 Points)**

In a study between the amount of rainfall and the quantity of air pollution removed, the following data were collected:

x: Daily rainfall (mm): Average = 5, SD = 1.5

y: Air pollution removed (in $\mu g/m^3$ particles removed): Average = 122, SD = 9.5
Correlation = -0.98.

(-4) if x, y flipped

(-2) each calculation error

1. **(15 Points)** Find the regression equation for predicting the air pollution removal from the amount of daily rainfall.

slope $= r \cdot \frac{SD_y}{SD_x} = -0.98 \cdot \frac{9.5}{1.5} = -6.2$ (6)

intercept $=$ avg$_y$ $-$ slope $\cdot$ avg$_x$ $= 122 - (-6.2) \cdot 5 = 153$ (6)

equation: pollution removed $= 153 - 6.2 \cdot$ amount of daily rainfall (3)

or: $y = 153 - 6.2 \cdot x$

2. **(10 Points)** Using your regression equation, predict the air pollution removal when the daily rainfall is 4.8 mm.

$y = 153 - 6.2 \cdot 4.8 = 123.24$

(-2) for old method, correct result

(-8) for old method, incorrect result

3. **(10 Points)** Find the r.m.s. error for predicting the air pollution removal from the amount of daily rainfall.

(-4) for each major mistake, e.g., SD_x instead of SD_y, $\sqrt{}$ of everything, r instead of r^2, etc.

r.m.s. error $= \sqrt{1 - r^2} \cdot SD_y$

$= \sqrt{1 - (-0.98)^2} \cdot 9.5 = \sqrt{1 - 0.9604} \cdot 9.5$

$= \sqrt{0.0396} \cdot 9.5 = 0.19899 \cdot 9.5 = 1.89 \approx 1.9$

4. **(15 Points)** Would you be surprised if you read that the daily rainfall of 4 mm had removed the amount of air pollution of 150 $\mu g/m^3$? Why or why not? Explain your reasoning, using the r.m.s. error.

predicted $y = 153 - 6.2 \cdot 4 = 128.2$ (5)

relate 150 to 128.2, using r.m.s. error:

$\frac{150 - 128.2}{1.9} = \frac{21.8}{1.9} = 11.47 \approx 11.5$ (5) [or any other reasonable explanation]

the observed value of 150 would be about 11.5 times the r.m.s. error above the predicted value of 128.2; yes, I would be really surprised!

3

(5)

Question 4: Sampling (**30 Points**)

Suppose you have a telephone directory for your local college from which you randomly select 100 names. To find out how students feel about a new pub on campus, you call the 100 numbers and interview the person who answers the phone.

1. (**21 Points**) Give **three** possible sources of bias for this survey.

- the person who answers the phone may not be the student intended for the interview but the spouse, children, family, roommates, friends, or visitors
- students that might go to a pub may be more difficult to reach via phone than students that usually stay at home
- when roommates share a phone number, it may happen that the same number is called more than once and the same persons again
- the person who answers the phone may not like to be interviewed and just hangs up
- student may not have phone but only cell phone (that is not listed ...)
- wording: pub, restaurant, bar?

(+7) for each reasonable source of bias
(-3) for repeating the same in different words
(-5) if not a real source of bias

2. (**9 Points**) Are the sources of bias you listed in Part 1. a problem even with a very large sample, or does a large sample size imply that these sources of bias can then be ignored? **Explain!**

correct idea (7)
specific explanation (2)

All a large sample does is repeat a mistake on a grand scale, if the data are subject to bias! So all we have is a large biased sample – we cannot ignore the problem.

From: Freedman, Pisani, Purves, Chapter 18, page 328, question 8

(+8) Correct answer

Question 5: Normal Approximation for Probability Histograms (**40 Points**)

(+2) correct explanation

A coin is tossed 100 times. True or false, and **explain**.
(**Answer and explain each of the following questions separately.**)

(+1) incorrect answer, some explanation

H: 1
T: 0

1. (**10 Points**) The expected value for the number of heads is 50. (True)/False?

box: [0] [1]
#draws: 100
box avg = $\frac{1}{2}$
box SD = $\sqrt{\frac{1}{2}\cdot\frac{1}{2}} = \frac{1}{2}$
$EV_{sum} = 100 \cdot \frac{1}{2} = 50$
$SE_{sum} = \sqrt{100} \cdot \frac{1}{2} = 5$

(-5) if missing

Is $EV_{sum} = 50$? Yes - see calculation to the left. $\Rightarrow$ True

2. (**10 Points**) The expected value for the number of heads is 50, give or take 5 or so. True/(False?)

The EV_{sum} is exactly 50, not give or take $\Rightarrow$ False

3. (**10 Points**) The number of heads will be 50. True/(False?)

The number of heads most likely will not be exactly 50, but it will be relatively close to 50. $\Rightarrow$ False

4. (**10 Points**) The number of heads will be around 50, give or take 5 or so. (True)/False?

The number of heads will be close to EV_{sum} (i.e., 50), give or take SE_{sum} (i.e., 5). See the calculation in 1. above where we obtained these numbers for EV_{sum} & SE_{sum}.

$\Rightarrow$ True