Statistics 1040, Sections 002, 003 & 004, Midterm 2 (200 Points)

March 27, 2002

Your Name: ____________________

Question 1: Chances and Probabilities (**30 Points**)

I have a bag with 20 balls in it: 10 are red, 8 are blue, and 2 are green.

1. **If I draw one ball at random from the bag, what is the chance that I get a red ball or a green ball? (10 Points)**

(-2) each calculation error (in part 1, 2, & 3)

chance of red ball: $\frac{10}{20} = \frac{1}{2} = 0.5 = 50\%$ (3)

chance of green ball: $\frac{2}{20} = \frac{1}{10} = 0.1 = 10\%$ (3)

mutually exclusive

chance of red ball or green ball: $\frac{10}{20} + \frac{2}{20} = \frac{12}{20} = \frac{3}{5} = 0.6 = 60\%$

addition rule (4) correct rule

2. **If I draw two balls at random without replacement, what is the chance that I get a red ball, followed by a green ball? (10 Points)**

chance that first is red: $\frac{10}{20} = \frac{1}{2} = 0.5 = 50\%$ (3)

chance that second is green, given that first is red: $\frac{2}{19} = 0.105 = 10.5\%$ (3)

dependent

chance that first is red and second is green: $\frac{10}{20} \cdot \frac{2}{19} = \frac{1}{19} = 0.0526 \approx 5.3\%$

multiplication rule (4) correct rule

3. **If I draw three balls at random with replacement, what is the chance that I get at least one red ball? (10 Points)**

chance of red ball in 1 draw: $\frac{10}{20} = \frac{1}{2}$ (2)

chance of no red ball in 1 draw: $1 - \frac{1}{2} = \frac{1}{2}$ (2)

chance of no red ball in 3 draws: $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^3 = \frac{1}{8} = 0.125 = 12.5\%$ (3)

chance of at least a red ball in 3 draws: $1 - \frac{1}{8} = \frac{7}{8} = 0.875 = 87.5\%$ (3)

opposites rule

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

	avg	SD
x: height woman	66	2.0
y: height man	69	2.5

r = 0.57

A student wonders if people of similar heights tend to date each other. To find this out, she measured herself, her dormitory roommates, and all of her female classmates; then she measured the man each woman was currently dating. After making many measurements and analyzing the data, she found out that the women were on average 66 inches tall, with a standard deviation of 2.0 inches; their dates were on average 69 inches tall, with a standard deviation of 2.5 inches. The correlation coefficient between the women's and the men's heights was 0.57.

(-4) if x, y flipped

(-2) each calculation error (in part 1, 2, & 3)

1. **Find the regression equation for predicting the height of a woman's date based on her own height. (20 Points)**

slope: $r \cdot \frac{SD_y}{SD_x} = 0.57 \cdot \frac{2.5}{2.0} = 0.7125 \approx 0.71$ (8)

intercept $= avg_y - slope \cdot avg_x = 69 - 0.7125 \cdot 66 = 21.975 \approx 22.0$ (8)

equation:

height man $= 22 + 0.71 \cdot$ height woman (4)

or $y = 22 + 0.71 \cdot x$

2. **Using your regression equation, predict the height of a date for a woman who is 67 inches tall. (10 Points)**

$y = 22 + 0.71 \cdot 67 = 69.57''$

(-2) for old method, correct result

(-8) for old method, incorrect result

3. **Find the r.m.s. error for predicting the date's height from the woman's height. (10 Points)**

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

$= \sqrt{1-0.57^2} \cdot 2.5 = \sqrt{1-0.3249} \cdot 2.5$

$= \sqrt{0.6751} \cdot 2.5 = 0.82 \cdot 2.5 = 2.05$

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

4. **Does the slope of the regression line (that you found in Part 1) say that men, if they date taller women, will become taller? Why or why not? Explain! (10 Points)**

no – the data originates from an observational study;

the slope suggests that associated with each extra inch for a woman, there is an increase of 0.71 inches for a man;

but: association is not the same as causation

(7) for "no"

(3) for reasonable explanation

From: Stat 1040, Fall 1999, Final Test, Friday, December 17, 1999; Question 8

Question 3: Sampling (30 Points)

In Web polls, anyone who views a certain Web page is allowed to vote by clicking on their choice of button. In fact, there is nothing to stop someone voting as many times as they want. The results of one such poll suggest that almost 90% of the US population wants to ban firearm sales. The poll has a very large sample size (over 1 million).

1. Web based polls such as this are notoriously susceptible to bias. Give **three** possible sources of bias for this poll. **(21 Points)**

Workbook answer:

"1. People must own a computer - creates a bias towards younger, wealthier, more technologically - trained, etc, people.

2. People who feel strongly are more likely to vote + these people may vote multiple times.

3. The source will create a bias - what type of people read this web page? Maybe it's in a newspaper web page - only people who find it will have the chance to vote."

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

Additional explanation:

1. & 3. result in a <u>selection bias</u>,

2. is a typical problem of a <u>voluntary response survey</u> and leads to a bias

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

Workbook answer:

"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 <u>large</u> biased sample - we <u>cannot</u> ignore the problem."

(-7) if no clear statement that sample is not the problem here

From: Stat 1040, Summer 2000, Test 2, July 17, 2000, Question 4

Question 4: Box Model, EV_{sum}, SE_{sum}, and the Normal Curve **(50 Points)**

Suppose it is known that 10% of all people in Utah have the blood type AB. Suppose we take a random sample of 500 Utahns and want to determine how many of them have the blood type AB.

1: blood type AB
0: other blood type

1. Find the box model for this situation. **(10 Points)**

$1 \times \boxed{1} \quad 9 \times \boxed{0}$ or $10 \times \boxed{1} \quad 90 \times \boxed{0}$ etc.

number of draws: 500

(-7) box with numbers other than $\boxed{0}/\boxed{1}$
(-3) box with $\boxed{0}/\boxed{1}$, but incorrect number
(-3) number of draws missing

2. How many people in our sample do you expect to have the blood type AB? **(10 Points)**

$\text{box average} = \frac{1}{10} = 0.1 = 10\%$ (5)

$EV_{sum} = 500 \cdot \frac{1}{10} = \underline{\underline{50}}$ (5)

(-2) each calculation error (in part 2, 3., & 4.)

3. What is the corresponding standard error? **(10 Points)**

$\text{box SD} = \sqrt{\frac{1}{10} \cdot \frac{9}{10}} = \sqrt{\frac{9}{100}} = \frac{3}{10} = 0.3$ (5)

$SE_{sum} = \sqrt{500} \cdot 0.3 = 22.36 \cdot 0.3 = \underline{\underline{6.7}}$ (5)

4. What is the chance that fewer than 40 people in our sample have the blood type AB ? **(20 Points)**

$\text{s.u.}: \frac{40 - EV_{sum}}{SE_{sum}} = \frac{40-50}{6.7} = \frac{-10}{6.7} = -1.49 \approx -1.5$

(5) correct curve parameters
(5) correct s.u. calculation
(5) correct table value
(5) correct side

area from -1.5 to 1.5: 86.64%

area below -1.5: $\frac{100\% - 86.64\%}{2} = \frac{13.36\%}{2} = \underline{\underline{6.68\%}}$

i.e, chance that fewer than 40 people in sample have blood type AB is 6.68%

From: Freedman, Pisani, Purves, Chapter 16, Review Exercise 4, page 285

Question 5: Law of Averages **(40 Points)**

A die will be rolled some number of times. Which is better: 60 rolls or 600 rolls in the situations listed below? Circle the number and briefly **explain** your answer for each situation:

(8) correct number
(1) correct explanation
(1) incorrect number, some explanation

1. You win \$1 if it shows ⚀ more than 20% of the times 60 / 600 **(10 Points)**

Workbook answer:
"To win, you need a large percentage error, and that is more likely in 60 rolls."

2. You win the dollar if the percentage of ⚀'s is more than 15%. 60 / 600 **(10 Points)**

Workbook answer:
"Now you want a small percentage error."

3. You win the dollar if the percentage of ⚀'s is between 15% and 20%. 60 / 600 **(10 Points)**

Workbook answer:
"You want a small percentage error."

4. You win the dollar if the percentage of ⚀'s is exactly $16\frac{2}{3}\%$. 60 / 600 **(10 Points)**

Workbook answer:
"60 rolls because to get exactly the expected value means getting exactly zero chance error, and that is more likely with fewer rolls."

Question 5:

Additional explanation:

Although only a verbal explanation was required (e.g., speaking of percentage error or law of averages), we can answer this question more precisely using a box model and working with $EV_{\%}$ and $SE_{\%}$:

1: "⚀" $\rightsquigarrow$ box: $\boxed{1\times\boxed{1}\quad 5\times\boxed{0}}$

6: anything else

number of draws: 60 or 600

box avg: $\frac{1}{6} = 0.167$

box SD $= \sqrt{\frac{1}{6}\cdot\frac{5}{6}} = \sqrt{\frac{5}{36}} = 0.37$

for 60 draws:

$EV_{sum} = 60\cdot\frac{1}{6} = 10$

$SE_{sum} = \sqrt{60}\cdot 0{,}37 = 2.87$

$EV_{\%} = 16\,^2/_3\,\%$

$SE_{\%} = \frac{2.87}{60}\cdot 100\% = 4.78\%$

$\xleftrightarrow{\text{the same}}$

for 600 draws:

$EV_{sum} = 600\cdot\frac{1}{6} = 100$

$SE_{sum} = \sqrt{600}\cdot 0.37 = 9.06$

$EV_{\%} = 16\,^2/_3\,\%$

$SE_{\%} = \frac{9.06}{600}\cdot 100\% = 1.51\%$

Note that $SE_{\%}$ is smaller for 600 draws than for 60 draws. So we will be closer to $EV_{\%} = 16\,^2/_3\%$ after 600 draws than after 60 draws.

We also can calculate the exact chances (e.g., more than 20% of "1"s or more than 15% of "1"s) using the normal curve.