Statistics 1040, Section 004, Midterm 2 (200 Points)

November 7, 2003

Your Name: _____

based on: Stat 1040, Spring 2003, Midterm 2, March 28, 2003, Question 4 | -2 each calculation error

Question 1: Chances and Probabilities (**40 Points**)

A shelf contains 9 books: 5 novels, 3 books of poems, and a dictionary. **Show your work!**

1. (**8 Points**) If one book is picked at random from the shelf, what is the chance that it is a novel?

   chance first is novel: $\frac{5}{9} = 0.556 = 55.6\%$ (3) (3) (2)

2. (**8 Points**) If one book is picked at random from the shelf, what is the chance that it is a novel or a dictionary?

   chance first is novel: $\frac{5}{9}$ (1)

   chance first is dictionary: $\frac{1}{9}$ (3) > mutually exclusive (4) correct rule

   chance first is novel or dictionary: $\frac{5}{9} + \frac{1}{9} = \frac{6}{9} = \frac{2}{3} = 0.667 = 66.7\%$

3. (**8 Points**) If two books are picked at random from the shelf, what is the chance that they are both novels?

   chance first is novel: $\frac{5}{9}$ (1)

   chance second is novel, given that first is novel: $\frac{4}{8}$ (3) (4) correct rule

   chance both are novels: $\frac{5}{9} \cdot \frac{4}{8} = \frac{20}{72} = \frac{5}{18} = 0.278 = 27.8\%$

4. (**8 Points**) If two books are picked at random from the shelf, what is the chance that neither of them is a novel?

   chance first is not novel: $\frac{4}{9}$ (2)

   chance second is not novel, given that first is not novel: $\frac{3}{8}$ (2) (4) correct rule

   chance both are not novels: $\frac{4}{9} \cdot \frac{3}{8} = \frac{12}{72} = \frac{1}{6} = 0.167 = 16.7\%$

5. (**8 Points**) If two books are picked at random from the shelf, what is the chance that at least one of them is a novel?

   "at least one is a novel" is opposite from "neither is a novel". (2)

   opposite rule:

   chance at least one is novel: $1 - \frac{1}{6} = \frac{5}{6} = 0.833 = 83.3\%$ (6) correct rule

Question 2: Sample Size and SE% **(10 Points)**

A professor at a university wants to know what percentage of students visits the university Web page on a regular basis. This professor randomly selects and interviews 100 students (Study 1). Unknown to this professor, another professor at the same university has assigned the same study as a course project — however, only 25 randomly selected students will be interviewed during the course project (Study 2). Even without knowing the exact box and EV%, we can make a statement about the SE%'s for the two studies. Just circle the correct answer(s). (You don't have to provide any further explanation here.)

(0)

1. The SE% of Study 1 will be 4 times as big as the SE% of Study 2.
2. The SE% of Study 1 will be 2 times as big as the SE% of Study 2.
3. The SE% of Study 1 will be $\sqrt{2}$ times as big as the SE% of Study 2.
4. The SE% of Study 1 will be about the same as the SE% of Study 2.

(2) 5. The SE% of Study 2 will be 4 times as big as the SE% of Study 1.

(10) (6) The SE% of Study 2 will be 2 times as big as the SE% of Study 1.

(2) 7. The SE% of Study 2 will be $\sqrt{2}$ times as big as the SE% of Study 1.

(2) for "6." and one additional answer

(0) for ≥ 3 answers

Explanation (not required):

Study 1	Study 2	
100	25	sample size

$\frac{100}{25} = 4.25$

Note on p. 167 in the workbook says:

"Multiplying the sample size by some factor divides the SE% by the square root of the factor."

Therefore:

$$\text{SE\% of Study 1} = \frac{\text{SE\% of Study 2}}{\sqrt{4}} = \frac{\text{SE\% of Study 2}}{2}$$

or $\text{SE\% of Study 2} = 2 \cdot \text{SE\% of Study 1}$

Question 3: Regression (**60 Points**)

-2 each calculation error

The average height of married American women in their early twenties is about 64.5 inches and the standard deviation is about 2.5 inches. The average height of married American men of the same age is about 68.5 inches, with a standard deviation of about 2.7 inches. The correlation between husbands heights and wives heights is about $r = 0.5$. **Show your work!**

-5 if x, y flipped
-2 if x, y not specified

	avg	SD
x: wife's height	64.5	2.5
y: husband's height	68.5	2.7
	$r = 0.5$	

1. (**20 Points**) Find the regression equation for predicting the height of a husband based on the wife's height in young couples.

$$\text{slope} = r \cdot \frac{SD_y}{SD_x} = 0.5 \cdot \frac{2.7}{2.5} = 0.54 \quad (8)$$

$$\text{intercept} = \text{avg}_y - \text{slope} \cdot \text{avg}_x = 68.5 - 0.54 \cdot 64.5 = 68.5 - 34.83 = 33.67 \quad (8)$$

equation: husband's height $= 33.67 + 0.54 \cdot$ wife's height (4)

(or: $y = 33.67 + 0.54 \cdot x$)

2. (**10 Points**) Using your regression equation, predict the height of a husband of a woman who is 67 inches tall. The predicted height of the husband is 69.85".

predicted height of a husband of a 67" tall woman $= 33.67 + 0.54 \cdot 67$
$= 33.67 + 36.18$
$= 69.85$

-2 for old method, correct result
-8 for old method, incorrect result

3. (**15 Points**) Find the r.m.s. error for predicting the husband's height from the wife's height. The r.m.s. error is 2.34".

$$\text{r.m.s. error} = \sqrt{1 - r^2} \cdot SD_y = \sqrt{1 - 0.5^2} \cdot 2.7 = \sqrt{1 - 0.25} \cdot 2.7$$

-5 for each major mistake, e.g.: SD_x instead of SD_y, $\sqrt{}$ of everything, r instead of r^2, etc.

$$= \sqrt{0.75} \cdot 2.7 = 0.866 \cdot 2.7 = 2.34$$

4. (**15 Points**) Would you be surprised to find out that a women who is 67 inches tall has a husband who is 77 inches tall? Why or why not? Explain!

from 2.) predicted height of a husband of a 67" tall woman: 69.85 (3)

S.U.: $\frac{77 - 69.85}{2.34} = \frac{7.15}{2.34} = 3.06$ (6) [or any other reasonable explanation]

77" is more than 3 r.m.s. errors above the predicted value of 69.85";
so this is very surprising (6)

from: FPP, p. 371-372, Review Exercises, Exercise 3 & 4 -2 each calculation error

Question 4: Box Model, Sums, Percentages, and the Normal Curve (**70 Points**)

A group of 50,000 tax forms has an average gross income of \$37,000 with an SD of \$20,000. Furthermore, 20% of the forms have a gross income over \$50,000.
A group of 900 forms is chosen at random for audit. Answer the questions below. Where necessary, find the box model, box average and box SD. **Show your work!**

1. (**35 Points**) Find the chance that between 19% and 21% of the forms chosen for audit have gross incomes over \$50,000. The chance is 54.67 %.

(2) 1: income over \$50k
0: income less than \$50k

box: [20 × 1, 80 × 0] (3) or [10,000 × 1, 40,000 × 0]
draws: 900 (1) # draws: 900

box avg = fraction of 1's = $\frac{20}{100} = 0.2$ (3)

box SD = $\sqrt{\text{fraction of 1's} \cdot \text{fraction of 0's}} = \sqrt{0.2 \cdot 0.8} = \sqrt{0.16} = 0.4$ (4)

[$EV_{sum} = 900 \cdot 0.2 = 180$ - not needed]

$SE_{sum} = \sqrt{900} \cdot 0.4 = 30 \cdot 0.4 = 12$ (5)

EV% = % of 1's in box = 20% (4)

$SE\% = \frac{SE_{sum}}{\#\text{draws}} \cdot 100\% = \frac{12}{900} \cdot 100\% = 1.33\%$ (5)

Normal curve:

s.u.: $\frac{19\% - 20\%}{1.33\%} = -0.75$ (2) $\frac{21\% - 20\%}{1.33\%} = 0.75$ (2)

area between -0.75 and 0.75: 54.67% (4)

2. (**35 Points**) Find the chance that the total gross income of the audited forms is over \$33,000,000. The chance is 69.15 %.

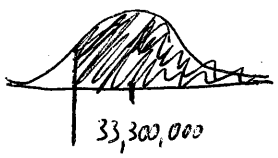
new box: tickets in box show gross incomes; exact numbers on tickets are unknown, but it is known that

box avg = \$37,000 (4)
box SD = \$20,000 (4) & # draws: 900 (2)

$EV_{sum} = 900 \cdot 37{,}000 = 33{,}300{,}000$ (5)

$SE_{sum} = \sqrt{900} \cdot 20{,}000 = 30 \cdot 20{,}000 = 600{,}000$ (5)

Normal curve:



33,000,000

s.u.: $\frac{33{,}000{,}000 - 33{,}300{,}000}{600{,}000} = -0.5$ (5)

area between -0.5 and 0.5: 38.29% (5)

area above -0.5:

$50\% + \frac{38.29\%}{2} = 50\% + 19.145\%$

$= 69.145\%$

$\approx 69.15\%$ (5)

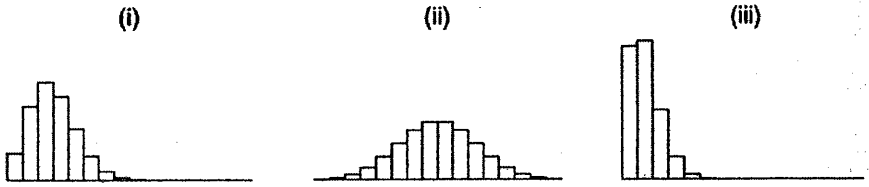
from: FPP, p. 324, Exercise Set C, Exercise 5 & 6

Question 5: Probability Histograms (**20 Points**)

1. (**10 Points**) Twenty-five draws are made at random with replacement from each of the boxes below:

A) $\boxed{0\ \ 1}$ B) $\boxed{9 \times \boxed{0}\quad 1 \times \boxed{1}}$ C) $\boxed{24 \times \boxed{0}\quad 1 \times \boxed{1}}$

The probability histograms for the sums are shown below, in scrambled order. Match the histogram with the boxes. **Briefly explain your choices.**

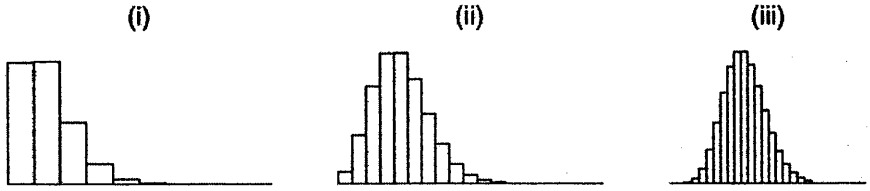


(i) goes with box B) (3) Textbook, p. A-77:

(ii) goes with box A) (3) "The more lopsided the box, the more skewed the histogram."

(iii) goes with box C) (3) [see book for additional "Comment"] (1)

2. (**10 Points**) Shown below are probability histograms for the sum of (a) 100, (b) 400, and (c) 900 draws from the box $\boxed{99 \times \boxed{0}\quad 1 \times \boxed{1}}$. Which histogram is which? **Explain briefly.**



(i) goes with sum (a) 100 (3) Textbook, p. A-78:

(ii) goes with sum (b) 400 (3) "The histograms get closer to the normal curve

(iii) goes with sum (c) 900 (3) as the number of draws goes up." (1)