

**Math 110**  
Test 2 Sample Solutions  
April 1, 2010

Be sure to provide explanations for your answers as indicated.

1. (20 pts.) Short answer.

(a) Figure 1 contains a residual plot of a data set.

i. Based on the residual plot, would you say the correlation coefficient is about (choose one):

- .05    .3    .8    .95    can't tell

Since the residuals are spread but not too much,  $r$  should be closer to 1 than 0. Since the data is not very tightly clustered,  $r$  should be approximately .8. (In fact an explicit calculation from the data shows that  $r = .829$ .)

ii. Based on the residual plot, would you say the slope of the regression line is about (circle one):

- .05    .3    .8    .95    can't tell

Can't tell. The residual plot is the difference between the true value and the regression line estimate. The slope of the regression line can't be determined from this information.

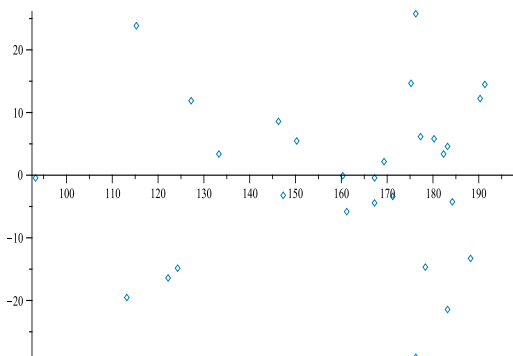


Figure 1:

(b) A club with 15 members has to choose from its membership an executive board consisting of 4 club members. How many ways are there to select the executive board?

Use the binomial coefficient  $\frac{15!}{4!11!} = 1365$ .

(c) Consider a chance process represented by a box model. Fill in the blanks:

- i. On average, the absolute value of the observed error for the sum of the draws will **increase** as the number of draws increases.
- ii. On average, the value for the observed error as a percentage of the sum of the draws will **decrease** as the number of draws increases.

This is the law of averages.

2. (20 pts.) (Hypothetical) A mathematics department at a small New England College tracks student SAT math scores and first semester calculus grades on a 4.0 scale. For the students who complete first semester calculus, the average SAT math score was 570 with an SD of 100 and the average calculus grade was 2.9 with an SD of .5. The correlation coefficient was found to be  $r = 0.6$ .

- (a) For a student with an SAT math score of 500, use the regression line to estimate his or her calculus grade on a 4.0 scale.

Use the formula for the regression line  $r \frac{SD_y}{SD_x}(x - \mu_x) + \mu_y = .6 \frac{.5}{100}(x - 570) + 2.9$ .  
When  $x = 500$ ,  $y = 2.69$ .

- (b) Find the r.m.s. error for the regression line of calculus test grades on SAT math scores.

The r.m.s. error is  $\sqrt{1 - r^2} \cdot SD_y = \sqrt{1 - .6^2} \cdot .5 = .4$ .

- (c) Among first semester calculus students, 25 students scored 500 on their SAT math test. Use the r.m.s. error and regression line to approximate how many of the 25 students obtained a grade of 2.5 or better in their first semester of calculus.

Use a normal curve with mean 2.69 and  $SD = .4$  to estimate this number. The  $z$ -value is  $z = \frac{2.69 - 2.5}{.4} \approx .475$ , which corresponds to a symmetric area of 36%. The percentage of students scoring 2.5 or above is then  $\frac{1}{2}36 + 50 = 68\%$ . The answer is 68% of 25, or 17.

3. (20 pts.) The thirteen hearts from a deck of cards are placed in an upside down hat. (Note: The face cards are the king, queen, and jack.)

- (a) If four cards are drawn at random from the hat with replacement, what are the chances that none of the cards are face cards?

$\frac{10}{13} \cdot \frac{10}{13} \cdot \frac{10}{13} \cdot \frac{10}{13} \approx .35$  or 35%.

- (b) If four cards are drawn at random from the hat without replacement, what are the chances that none of the cards are face cards?

$\frac{10}{13} \cdot \frac{9}{12} \cdot \frac{8}{11} \cdot \frac{7}{10} \approx .29$  or 29%.

- (c) If four cards are drawn at random from the hat without replacement, what are the chances that only the first and fourth cards are face cards?

$\frac{3}{13} \cdot \frac{10}{12} \cdot \frac{9}{11} \cdot \frac{2}{10} \approx .031$  or 3.1%.

4. (20 pts.) A gambler at the Bellagio casino is playing roulette. The gambler's favorite bet is to place \$5 on four adjoining numbers. This bet pays 8 to 1. (See the attached roulette table.)

- (a) On 100 plays of this bet, what is the gambler's chance of winning more than \$20? (*Hint:* What are the box model, EV, and SE for this bet played 100 times?)

The box should contain 38 tickets, 4 labeled \$40 for the winning bet of \$5 paying 8-to-1 and 34 labeled -\$5 for a losing bet. The expected value is  $100 \times \text{Avg}_{\text{box}} = 100 \times \frac{-5}{19} \approx -26.3$ . The SE for this bet is  $\sqrt{100} \times (40 - (-5))\sqrt{\frac{4}{38} \frac{34}{38}} \approx 138.1$ . Compute the  $z$ -value  $z = \frac{20 - (-26.3)}{138.1} \approx .335$ . This corresponds to an area of 26%. We want the area of the associated right tail,  $\frac{1}{2}(100 - 26) = 37\%$ . So the chance of winning \$20 playing this bet 100 times is 37%.

- (b) On 100 plays of this bet, what is the gambler's chance of winning 15 or more times? (*Hint*: What are the box model, EV, and SE for winning in 100 plays?)

Use a box containing 38 tickets, 4 labeled 1 and the rest labeled 0. The EV is  $100 \times \frac{4}{38} = 10.5$ . The SE is  $\sqrt{100} \times (1 - 0)\sqrt{\frac{4}{38} \frac{34}{38}} \approx 3.07$ . The  $z$ -value is  $z = \frac{10.5 - 15}{3.07} \approx -1.46$ . The corresponding symmetric area is 86%. The area of the right tail is 7%, calculated as in (a), which is the answer.

5. (20 pts.) (Hypothetical) A large statistics class has 400 students with 100 students from each of the four years, first-year through senior year. The instructor takes a simple random sample of size  $N=50$  from the class.

- (a) What are the EV and SE for the percentage of seniors in the this sample? (*Hint*: What is the box model for this question?)

Use a box containing 400 tickets, 100 labeled 1 and 300 labeled 0. The EV for percent is 25%, the percentage of 1's in the box. The SE for percent is  $(1 - 0)\sqrt{\frac{1}{4} \frac{3}{4}} / \sqrt{50} \times 100\% \approx 6.1\%$ . Since this is a simple random sample, draws are made without replacement. So we must use the correction factor to obtain the correct SE,  $\sqrt{\frac{400-100}{400-1}} 6.1 \approx 5.7\%$ .

- (b) What are the chances that the percentage of seniors in the sample will be less than 20%?

The  $z$ -value is  $\frac{25-20}{5.7} \approx .877$ , which corresponds to an area of 62%. The area of the left tail is  $\frac{1}{2}(100 - 62) = 19\%$ , which is the answer.