

MATH 1681.100 SECOND EXAM

Print Name: Stephen Muir

Sign Name:

15 pts

(2) Calculate the 5 statistic summary for this paired data set.

x	y
1	2
3	3
5	6
7	6

6 pts $\rightarrow A_x = \frac{1+3+5+7}{4} = 4$; $A_y = \frac{2+3+6+6}{4} = 4.25$

6 pts $\rightarrow SD_x = \sqrt{\frac{(1-4)^2 + (3-4)^2 + (5-4)^2 + (7-4)^2}{4}}$
 $= \sqrt{\frac{9+1+1+9}{4}} = \sqrt{\frac{20}{4}} = \sqrt{5} \approx 2.24$

$SD_y = \sqrt{\frac{(2-4.25)^2 + (3-4.25)^2 + (6-4.25)^2 + (6-4.25)^2}{4}}$

$= \sqrt{\frac{(2.25)^2 + (1.25)^2 + (1.75)^2 + (1.75)^2}{4}}$

$= \sqrt{\frac{1}{4} \left(\frac{81}{16} + \frac{25}{16} + \frac{49}{16} + \frac{49}{16} \right)} = \frac{1}{8} \sqrt{204}$

3 pts $A_{x \cdot y} = \frac{2+9+30+42}{4} = \frac{83}{4} = 20.75$ $= \frac{1}{4} \sqrt{51} \approx 1.79$

$r = \frac{20.75 - (4)(4.25)}{(1.79)(2.24)} = .935$

5 pts

- (4) Write the generic regression equation for y on x , using the symbols A_x , A_y , SD_x , SD_y , and r for for the average x value, the average y value, the standard deviation of the x values, the standard deviation of the y values, and the correlation coefficient, respectively.

$$y = \frac{r \cdot SD_y}{SD_x} (x - A_x) + A_y$$

(Note - this predicts the average y value for a given input x)

5 pts

- (5) Write the generic regression equation for x on y , using the symbols A_x , A_y , SD_x , SD_y , and r for for the average x value, the average y value, the standard deviation of the x values, the standard deviation of the y values, and the correlation coefficient, respectively.

$$x = \frac{r \cdot SD_x}{SD_y} (y - A_y) + A_x$$

(Note - this predicts the average x value for a given input y)

For the remaining questions, we use the data from Galton's study of the relationship between heights of fathers and heights of their sons.

Let x be the height of a father, and y the height of his son. He observed 1000 father-son pairs and obtained the 5 statistic summary: $A_x = 68''$, $SD_x = 2.7''$, $A_y = 69''$, $SD_y = 2.7''$, $r = 0.5$.

10 pts

- (6) Use regression to predict the average height of a son whose father is 66" tall.

$$A(y | x = 66'') = \frac{(0.5)(2.7)}{2.7} (66 - 68) + 69$$

$$= 68''$$

10pts

- (9) Use regression to predict the average height of a father whose son is 68" tall.

$$A(x|y=68'') = \frac{(0.5)(2.7)}{2.7} (68 - 69) + 68$$
$$= 67.5''$$

10pts

- (10) Among the pairs where the son was 68" tall, estimate the percent of fathers that were taller than their sons, i.e. taller than 68". (Use the normal approximation in the "horizontal strip" of pairs with $y = 68$ ". Be sure to use an RMS error for the standard deviation.)

- Use Average: $67.5'' = A(x|y=68'')$
- Use SD: $(\sqrt{1 - 0.5^2})(2.7) \approx 2.34''$
(RMS error for x)

• Standardize: $z = \frac{68 - 67.5}{2.34} \approx +.21$

- From table: ~~52~~ 38% to left
i.e. shorter.
- Thus 41.68% to right
i.e. taller.