Line of Best Fit

PROJECT

Median – Median Line

1. Organize the data with the *x*-coordinate from least to greatest.
2. Separate the data into three (3) groups of equal size. If not divisible by 3, groups 1 and 3 must be the same size.
3. Find the medians of the *x*-coordinates and the *y*-coordinates.
4. Calculate the slope using the medians from group 1 and the medians from group 3. This is the slope of the median-median line.
5. Find the sum of the three medians of the *y*-values.
6. Find the sum of the three medians of the *x*-values.
7. Calculate the *y*-intercept of the median-median line using the formula:

$$y\text{-intercept} = \frac{Sum\ of\ medians\ of\ y - slope \cdot (sum\ of\ medians\ of\ x)}{3}$$

8. Write the equation of the median-median line in slope-intercept form.

Example: Given the data

1	2	7	5	2	4	8	6	7
48	42	93	69	50	45	96	44	82

STEP 1: *Sort by x*

(1, 48), (2, 42), (2, 50), (4, 45), (5, 69), (6, 44), (7, 82), (7, 93), (8, 96)

STEP 2: *Separate into groups*

Group 1: (1, 48), (2, 42), (2, 50) *x*: 1, 2, 2 *y*: 42, 48, 50

Group 2: (4, 45), (5, 69), (6, 44) *x*: 4, 5, 6 *y*: 44, 45, 69

Group 3: (7, 82), (7, 93), (8, 96) *x*: 7, 7, 8 *y*: 82, 93, 96

STEP 3: *Find the medians*

Group 1 medians: *x*: 2 *y*: 48

Group 2 medians: *x*: 5 *y*: 45

Group 3 medians: *x*: 7 *y*: 93

STEP 4: *Calculate group 1-group 3 slope*

$\frac{93-48}{7-2} = \frac{45}{5} = 9$

STEP 5: *Sum of y-medians*

48 + 45 + 93 = 186

STEP 6: *Sum of x-medians*

2 + 5 + 7 = 14

STEP 7: *y-intercept value*

$\frac{186-9(14)}{3} = 20$

STEP 8: *Write the line*

$y = 9x + 20$

Correlation Coefficient

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

Using the example above, we compute the following

$\Sigma XY = (1)(48) + (2)(42) + (7)(93) + (5)(69) + (2)(50) + (4)(45) + (8)(96) + (6)(44) + (7)(82) = 3014$

$\Sigma X = 1 + 2 + 7 + 5 + 2 + 4 + 8 + 6 + 7 = 42$

$\Sigma Y = 48 + 42 + 93 + 69 + 50 + 45 + 96 + 44 + 82 = 569$

$\Sigma X^2 = 1^2 + 2^2 + 7^2 + 5^2 + 2^2 + 4^2 + 8^2 + 6^2 + 7^2 = 248$

$\Sigma Y^2 = 48^2 + 42^2 + 93^2 + 69^2 + 50^2 + 45^2 + 96^2 + 44^2 + 82^2 = 39879$

$n = 9$

$$r = \frac{3014 - \frac{(42)(569)}{9}}{\sqrt{\left(248 - \frac{(42)^2}{9}\right)\left(39879 - \frac{(569)^2}{9}\right)}} = 0.7958806296$$

Line of Best Fit

PROJECT

Project: Scatterplot and Correlation Coefficient Data below are the heights of mothers (X) and their daughters (Y) in inches.

X	63	67	64	60	65	67	59	60	58	72	63
Y	59	65	65	61	65	67	61	63	60	71	62

Source: U.S. Department of Health and Human Services, National Center for Health Statistics

1. Use a computer program (Google sheets, Excel, *etc*.) to generate a scatter plot of the data.
2. Describe the relationship between the data.
3. Calculate the correlation coefficient by hand and compare it to the correlation coefficient generated by a calculator or computer.
4. Look at the data and create a trend line by inspection.
5. Use the median-median method to create a trend line for the data.
6. Use a calculator to generate the trend line for the data.
7. Compare the three regression lines generated in steps 4 – 6.
8. Can we conclude that taller mothers will have taller daughters?