Statistics

Lecture 15

Feb 19-8:47 AM

Consider the Sample below

x	y
1	8
2	5
5	2
4	3
4	2

1) $x \to L1$, $y \to L2$

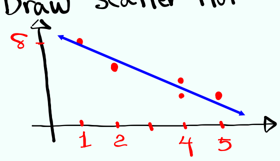
use $\boxed{\text{2-Var Stats}}$ with L1 & L2

$\sum x = 16$ $\qquad \sum y = 20$

$\sum x^2 = 62$ $\qquad \sum y^2 = 106$

$n = 5$ $\qquad \sum xy = 48$

Draw Scatter Plot



Use LinReg$(a+bx)$ with L1 & L2

$a = 8.741$ ✓

$b = -1.481$ ✓

$r^2 = .912$

$r = -.955$ ✓

Regression line

$\boxed{\hat{y} \approx 8.7 - 1.5x}$

$r^2 \approx 91\%$

91% of Y-values are explained by x-values.

r is close to -1

Linear Correlation is Significant

Sep 19-8:51 AM

How to find a & b using formula:

$\sum x = 16$ $\quad \sum y = 20$

$\sum x^2 = 62$ $\quad \sum y^2 = 106$

$n = 5$ $\quad \sum xy = 48$

$$\boxed{a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n \sum x^2 - (\sum x)^2}}$$

$$a = \frac{20 \cdot 62 - 16 \cdot 48}{5 \cdot 62 - 16^2} = \frac{472}{54} = \boxed{8.741}$$

$$\boxed{b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}}$$

$$b = \frac{5 \cdot 48 - 16 \cdot 20}{5 \cdot 62 - 16^2} = \frac{-80}{54} = \boxed{-1.481}$$

Sep 19-9:04 AM

$\sum x = 16$ $\quad \sum y = 20$

$\sum x^2 = 62$ $\quad \sum y^2 = 106$

$n = 5$ $\quad \sum xy = 48$

what about Formula for ?

$$\boxed{r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2}\sqrt{n \sum y^2 - (\sum y)^2}}}$$

$$r = \frac{5 \cdot 48 - 16 \cdot 20}{\sqrt{5 \cdot 62 - 16^2}\sqrt{5 \cdot 106 - 20^2}} = \frac{-80}{\sqrt{54}\sqrt{130}} = \frac{-80}{\sqrt{7020}} = \boxed{-.955}$$

80 $\boxed{\div}$ $\boxed{2nd}$ $\boxed{x^2}$ 7020 $\boxed{Enter}$

$\underbrace{.955 \ \boxed{x^2} \ \boxed{Enter}}_{r^2}$ $.912 \approx 91\%$

Sep 19-9:11 AM

Walk time	BS level
10	135
15	120
20	110
5	140
30	100
10	125

Walk time → x → L1

BS level → y → L2

Use LinReg(a+bx)

with L1 & L2

$a = 146.042 \approx 146$

$b = -1.625 \approx -2$

$r^2 = .932$

$r = -.965$

Linear Regression line

$y \approx 146 - 2x$

$r^2 \approx 93\%$

93% of my BS level are explained by walk time.

r is close to -1, Linear Correlation is significant.

Sep 19-9:18 AM

How to make Predictions:

If r is Significant, use the regression line

If r is not Significant, use $\bar{y}$

$\bar{y} = \frac{\sum y}{n}$ or $\boxed{\text{VARS}}$ $\boxed{\text{5: Statistics}}$ $\boxed{5: \bar{y}}$ $\boxed{\text{Enter}}$

$\bar{y} \approx 122$

Sep 19-9:26 AM

Given $n=10$, $\sum y=205$, $y=8+2.5x$

Predict y if $x=6$

1) Assume $\underbrace{r \text{ is Significant.}}$

Use regression line $y=8+2.5x$

$=8+2.5(6)=\boxed{23}$

2) Assume $\underbrace{r \text{ is not Significant}}$

Use $\bar{y}$ $\bar{y}=\frac{\sum y}{n}=\frac{205}{10}=\boxed{20.5}$

Sep 19-9:30 AM

# absences	Exam Score
1	88
0	95
2	85
4	70
3	80

absences → x → L1

Exam Score → y → L2

Use LinReg(a+bx) with L1 & L2

$a=95.2\approx 95$ $r^2=.963$

$b=-5.8\approx -6$ $r=-.982$

$y\approx 95-6x$

$r^2\approx 96\%$

96% of exam Scores are Explained by attendance.

r is close to -1, Linear Correlation is Significant.

Sep 19-9:34 AM

If a student misses 4 classes,

Predict his/her exam score

1) Assume $\underbrace{r \text{ is significant.}}$

Use Regression line $y = 95 - 6x$ $= 95 - 6(4) = \boxed{71}$

2) Assume $\underbrace{r \text{ is not significant.}}$

Use $\bar{y}$ $\boxed{\text{VARS}}$ $\boxed{\text{5: statistics}}$ $\boxed{5: \bar{y}}$ $\boxed{\text{Enter}}$ $\approx \boxed{84}$

Sep 19-9:41 AM