Statistics

Lecture 9

Feb 19-8:47 AM

Store the following in L1 & L2

x	y
2	7
3	10
4	10
5	12
6	15

$\boxed{\text{2nd}}$ $\boxed{+}$ $\boxed{\text{4: Clear all lists}}$ $\boxed{\text{Enter}}$

$\boxed{\text{STAT}}$ Edit $\boxed{\text{1: Edit}}$

L1	L2
2	7
3	10
4	10
5	12
6	15

$\boxed{\text{STAT}}$ $\boxed{\rightarrow}$ CALC $\boxed{\text{2: 2-Var Stats}}$

Xlist: L1

Ylist: L2

FreqList: $\boxed{\text{clear}}$

$\boxed{\text{Calculate}}$

$\sum x = 20$ $\quad$ $\sum y = 54$

$\sum x^2 = 90$ $\quad$ $\sum y^2 = 618$

$n = 5$ $\quad$ $\sum xy = 234$

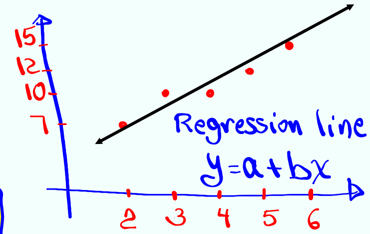
Mar 23-9:58 AM

$\sum x = 20$ $\quad$ $\sum y = 54$

$\sum x^2 = 90$ $\quad$ $\sum y^2 = 618$

$n = 5$ $\quad$ $\sum xy = 234$

Scatter Plot



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

$$a = \frac{54 \cdot 90 - 20 \cdot 234}{5 \cdot 90 - 20^2} = \frac{180}{50} = \boxed{3.6}$$

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

$$b = \frac{5 \cdot 234 - 20 \cdot 54}{5 \cdot 90 - (20)^2} = \frac{90}{50} = \boxed{1.8}$$

Regression line

$$\boxed{y = 3.6 + 1.8x}$$

Mar 23-10:05 AM

[STAT] [→] CALC

[8: LinReg(a+bx)]

xlist: L1

ylist: L2

[clear]

[Calculate]

$\boxed{\begin{array}{l} a = 3.6 \\ b = 1.8 \end{array}}$

$r^2 = .931$

$r = .965$

If r & r^2 missing:

[2nd] [0] ↓ ↓ ↓ ⋯ ↓ ▶DiagnosticOn

[Enter] [Enter]

Mar 23-10:15 AM

Formula for r:

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

$\sum x = 20$ $\quad\sum y = 54$

$\sum x^2 = 90$ $\quad\sum y^2 = 618$

$n = 5$ $\quad\sum xy = 234$

$$r=\frac{5\cdot 234-20\cdot 54}{\sqrt{5\cdot 90-20^2}\,\sqrt{5\cdot 618-54^2}}=\frac{90}{\sqrt{50}\,\sqrt{174}}$$

$$=\frac{90}{\sqrt{8700}}=\boxed{.965}$$

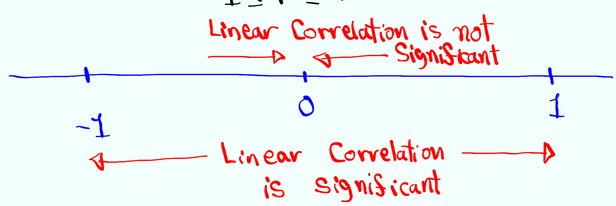
90 $\boxed{\div}$ $\boxed{2nd}$ $\boxed{x^2}$ 8700 $\boxed{\text{Enter}}$

Now find r^2

$r^2 = .965^2 \approx \boxed{.931}$

$r \rightarrow$ Linear Correlation Coefficient

$-1 \le r \le 1$



From Last example $r = .965$, it is close to 1 therefore Linear Correlation is Significant

$r^2 \rightarrow$ Coefficient of determination

Always convert to whole %.
it tells us what % of y-values are explained by x-values.

From Last example $r^2 \approx 93\%$

about 93% of y-values are explained by x-values.

Study time	Exam Score
8	85
7	80
10	95
10	100
6	70
5	65

Study time → x → L1

Exam Score → y → L2

$\boxed{\text{STAT}}$ $\boxed{\rightarrow}$ CALC

8: LinReg(a+bx)

use L1 & L2

$a = 32.1875 \approx 32$

$b = 6.5625 \approx 7$

$r^2 = .98$

$r = .990$

Regression line

$y \approx 32 + 7x$

Coef. of determination

$r^2 = 98\%$

98% of exam Scores are explained by Study time.

Linear Correlation

$r = .990$

It is close to 1,

It is significant.

Mar 23-10:35 AM

How to make predictions:

1) If r is significant,

use the regression line: Plug in x-value find Y-value

2) If r is not significant.

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

If Diego studies 8 hrs,

Predict his exam Score

1) r is significant

$y = 32 + 7x$

$= 32 + 7(8)$

$= 32 + 56 = \boxed{88}$

2) r is not significant

$\bar{y} = 82.5 \approx \boxed{83}$

Mar 23-10:44 AM

Walk time	Blood Sugar level
10	135
20	110
15	120
5	140
30	100

walk time → x → L1

BS level → y → L2

use LinReg(a+bx)

with L1 & L2

$a = 148.243 \approx 148$

$b = -1.703 \approx -2$

$r^2 = .958$

$r = -.979$

Regression line

$y \approx a + bx$

$\boxed{y \approx 148 - 2x}$

$r^2 \approx 96\%$

96% of BS level are explained by walking time.

r is close to -1

Linear Correlation is Significant

Mar 23-10:52 AM

Predict my BS level if I walk for 15 minutes:

1) Assume r is Significant

use $y \approx 148 - 2x$

$\approx 148 - 2(15) = 148 - 30 \approx \boxed{118}$

2) Assume r is not Significant.

use $\bar{y}$

121

VARS

5: Statistics

5: $\bar{y}$ Enter

SG 9

Mar 23-11:00 AM