Statistics

Lecture 5

Feb 19-8:47 AM

Working with Ordered-Pairs (x, y)

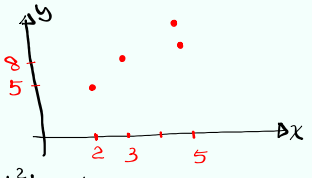
SG 9

ex: Consider the ordered-pairs below

$(2,5)$, $(3,8)$, $(5,10)$, $(5,12)$

Plot these Points

Scatter Plot



x	y	x^2	y^2	xy
2	5	4	25	10
3	8	9	64	24
5	10	25	100	50
5	12	25	144	60

$\sum x = 15$

$\sum x^2 = 63$

$n = 4$

$\sum y = 35$

$\sum y^2 = 333$

$\sum xy = 144$

Clear all lists

2nd | + | 4: Clear All Lists | Enter

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

STAT | Edit | 1: Edit

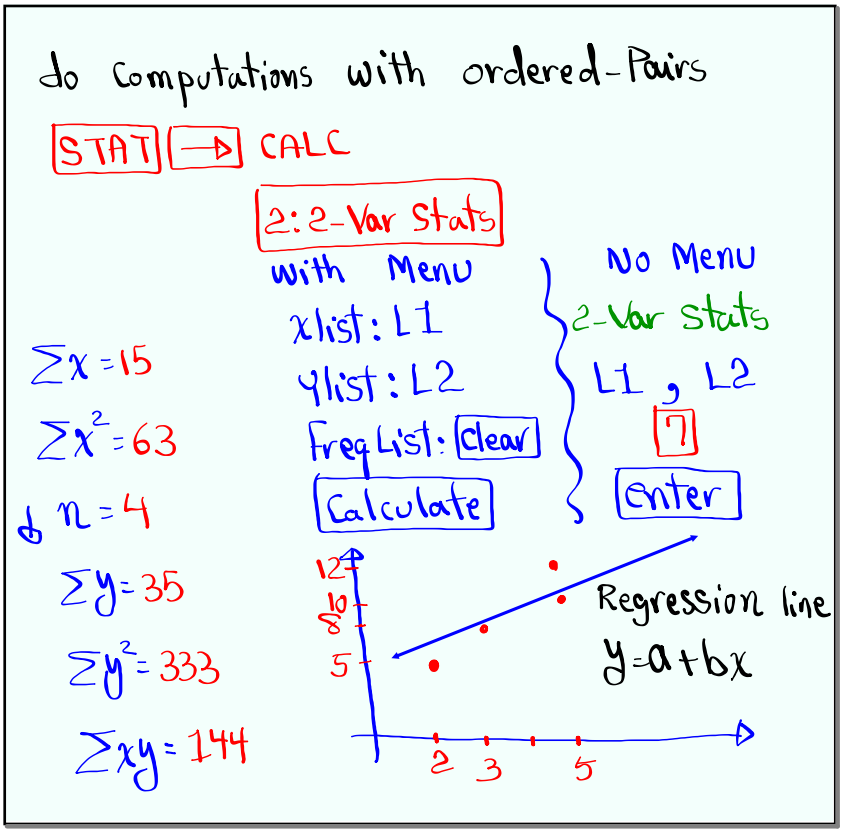
L1	L2
2	5
3	8
5	10
5	12

→

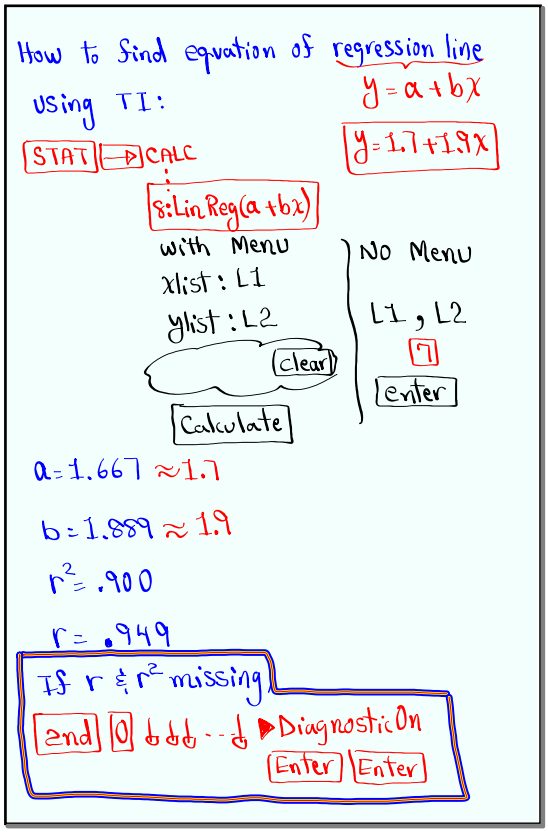
Quit & clear Screen

2nd | MODE | Clear

Sep 29-5:06 PM



Sep 29-5:20 PM



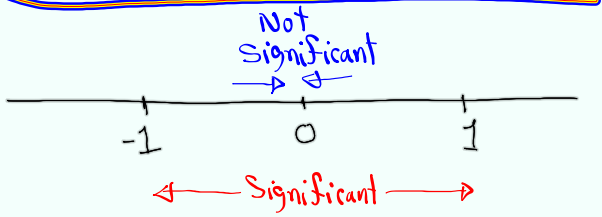
Sep 29-5:27 PM

r → Linear Correlation Coefficient

$-1 \leq r \leq 1$

When r is close to 1 or −1, Linear Correlation is Significant.

When r is close to 0, Linear Correlation is not Significant.



From Last example

$r = .949$ It is close to 1, Linear Correlation is Significant.

Sep 29-5:36 PM

r^2 → Coefficient of determination

Always in whole %.

r^2 tells us what % of y-values are explained by x-values.

From Last example

$r^2 = .900 \rightarrow r^2 \approx 90\%$

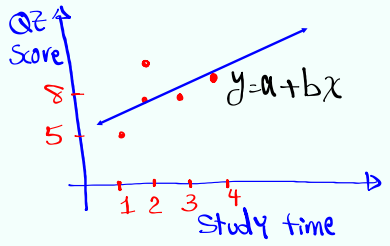
90% of y-values are explained by x-values.

Sep 29-5:41 PM

I randomly selected 5 students, chart below shows Study time in hours and QZ Scores.

Study time	QZ Score
3	8
4	9
2	8
1	5
2	10

Scatter Plot



Clear all lists

Study time → X → L1

QZ Scores → Y → L2

STAT → CALC → 2:2-Var...

use 2-Var Stats with L1 & L2 to find

$\Sigma x = 12$ $\Sigma y = 40$

$\Sigma x^2 = 34$ $\Sigma y^2 = 334$

$n = 5$ $\Sigma xy = 101$

Sep 29-5:44 PM

use LinReg(a+bx) with L1 & L2 to

STAT → CALC → LinReg(a+bx)

Find

$a = 5.692 \approx 6$

$b = .962 \approx 1$

Linear Regression

$y \approx 6 + x$

$r^2 = .343$

$r = .586$

r^2 Coef. of determination

$r^2 \approx 34\%$

34% of QZ Scores are explained by study time.

66% are unexplained

r Linear Correlation Coef.

$r = .586$

it is somewhat close to 1, Linear Correlation is significant

Sep 29-5:54 PM

How to find a & b using formula

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

$\sum x = 12$ $\quad \sum y = 40$

$\sum x^2 = 34$ $\quad \sum y^2 = 334$

$n = 5$ $\quad \sum xy = 101$

$$= \frac{40 \cdot 34 - 12 \cdot 101}{5 \cdot 34 - 12^2}$$

$$= \frac{148}{26}$$

$$= 5.692 \approx \boxed{6}$$

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

$$= \frac{5 \cdot 101 - 12 \cdot 40}{5 \cdot 34 - 12^2}$$

$$= \frac{25}{26}$$

$$= .962 \approx \boxed{1}$$

Sep 29-6:03 PM

Now formula for r:

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

$\sum x = 12$ $\quad \sum y = 40$

$\sum x^2 = 34$ $\quad \sum y^2 = 334$

$n = 5$ $\quad \sum xy = 101$

$$= \frac{5 \cdot 101 - 12 \cdot 40}{\sqrt{5 \cdot 34 - 12^2}\ \sqrt{5 \cdot 334 - 40^2}}$$

$$= \frac{25}{\sqrt{26}\ \sqrt{70}}$$

$$= \frac{25}{\sqrt{1820}} = .586$$

25 ÷ end x^2 1820 enter

Find r^2 in whole %

$r^2 = (.586)^2 = .343 \approx \boxed{34\%}$

Sep 29-6:11 PM

Given: $n=5$, $\Sigma x=16$, $\Sigma x^2=66$,

$\Sigma y=41$, $\Sigma y^2=373$, $\Sigma xy=154$

1) Find eqn of the regression line

$$a=\frac{\Sigma y\Sigma x^2-\Sigma x\Sigma xy}{n\Sigma x^2-(\Sigma x)^2}=\frac{41\cdot 66-16\cdot 154}{5\cdot 66-16^2}=\frac{242}{74}$$

$$=3.270\approx \boxed{3}$$

$$b=\frac{n\Sigma xy-\Sigma x\cdot\Sigma y}{n\Sigma x^2-(\Sigma x)^2}=\frac{5\cdot 154-16\cdot 41}{5\cdot 66-16^2}=\frac{114}{74}$$

$$=1.541\approx \boxed{2}$$

Regression line

$$\boxed{y\approx 3+2x}$$

Sep 29-6:19 PM

Find linear Correlation Coef.

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

$$=\frac{5\cdot 154-16\cdot 41}{\sqrt{5\cdot 66-16^2}\ \sqrt{5\cdot 373-41^2}}=\frac{114}{\sqrt{74}\ \sqrt{184}}$$

114 ÷ [2nd] [x^2] 13616 [enter]

$$=\frac{114}{\sqrt{13616}}=\boxed{.977}$$

It is close to 1, Linear Correl. is significant

$r^2=(.977)^2=.9545\ldots\approx \boxed{95\%}$

95% of Y-values are explained by x-values.

Sep 29-6:28 PM

How to make predictions

When r is Significant,

Use the regression line to make Prediction, Plug in x, find y

When r is not Significant,

Use $\bar{y}$ $\bar{y} = \frac{\Sigma y}{n}$

Sep 29-6:36 PM

Given $y = 5 + 3x$, $n = 4$, $\Sigma y = 30$

Predict y when $x = 1$.

1) Assume r is Significant

$y = 5 + 3(1) = 5 + 3 = \boxed{8}$

2) Assume r is not Significant.

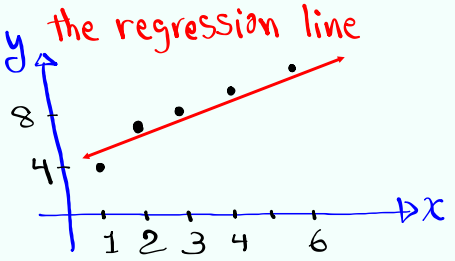
$\bar{y} = \frac{\Sigma y}{n} = \frac{30}{4} = \boxed{7.5}$

Sep 29-6:39 PM

Use the chart below

x	y
1	4
3	8
2	7
4	10
6	12

1) Draw Scatter Plot and the regression line



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

use LinReg(a+bx) with L1 & L2 to find

$a = 3.270 \approx 3$ $\to \boxed{y \approx 3 + 2x}$

$b = 1.541 \approx 2$

$r^2 = .954$ $r^2 \approx 95\%$ 95% of y-values are explained by x-values.

$r = .977$ close to 1 Linear Correlation is Significant

Sep 29-6:43 PM

Predict y when $x = 5$

1) Assume r is Significant

$y \approx 3 + 2(5) \approx \boxed{13}$

2) Assume r is not Significant

Use $\bar{y}$

$\bar{y} = \frac{\Sigma y}{n} \to \bar{y} = 8.2$ $\approx \boxed{8}$

VARS

5: Statistics

5: $\bar{y}$ Enter

SG 9

Sep 29-6:51 PM

Some Review

class MP	class F
12	5
20	10
28	12
36	3

Class MP → L1

Class F → L2

STAT → CALC

1:1-Var Stats

Find

$\bar{x} = 23.467 \approx 23$

$S = S_x = 7.181 \approx 7$

$n = 30$

$S^2 = \frac{22432}{435}$

VARS | 5:Statistics

3:S_x | x^2 | Math | 1:▶Frac

Enter

Sep 29-6:57 PM