

Statistics

Lecture 5



Feb 19-8:47 AM

class QZ 5

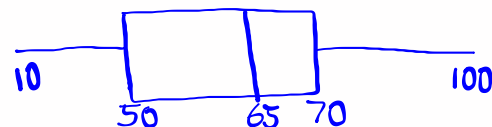
Consider the following

5 - Number Summary				
10	50	65	70	100

3) upper fence &
Lower fence.

Ranges for outliers
10 - 20

1) Draw Box Plot



$$2) IQR = Q_3 - Q_1 = 70 - 50 = 20 \checkmark$$

$$Upper\ Fence = Q_3 + 1.5(IQR) = 70 + 1.5(20) = 100 \checkmark$$

$$Lower\ Fence = Q_1 - 1.5(IQR) = 50 - 1.5(20) = 20 \checkmark$$

Feb 27-9:14 PM

(Pg 9)

Working with ordered pairs

x	y	x ²	y ²	xy
1	3	1	9	3
2	5	4	25	10
3	8	9	64	24
4	8	16	64	32
5	10	25	100	50

$\sum x = 15$ $\sum y = 34$ $\sum xy = 119$
 $\sum x^2 = 55$ $\sum y^2 = 262$ $n = 5$ # of ordered Pairs

Clear all lists [2nd] [+] [4: ClearAll Lists] [Enter]

x → L1 [STAT] Edit L1 | L2
 y → L2 [1: Edit] 1 | 3
 2 | 5
 3 | 8
 4 | 8
 5 | 10

Let's quit [end] [MODE]

[STAT] → [CALC] [2: 2-Var Stats]

With Menu No Menu
 XList: L1 L1, L2
 YList: L2 []
 Freq List: [clear] [Enter]
 [Calculate]

$\sum x = 15$ $\sum y = 34$
 $\sum x^2 = 55$ $\sum y^2 = 262$
 $n = 5$ $\sum xy = 119$

Mar 5-6:53 PM

Consider the chart below

x	y
2	8
3	10
5	14
5	16
6	18
8	25

1) Draw Scatter Plot

Clear all lists, x → L1, y → L2 quit [2nd] [MODE]

use 2-Var Stats to find

[STAT] → [CALC] [2: 2-Var Stats]

$\sum x = 29$ $\sum y = 91$
 $\sum x^2 = 163$ $\sum y^2 = 1565$
 $n = 6$ $\sum xy = 504$

Mar 5-7:07 PM

How to find the equation of the regression line using TI:

With Menu **NO Menu**

Xlist: L1 L1, L2

Ylist: L2 \downarrow

Enter

Freq. List: **clear**

Store Reg Eq: **clear**

Calculate

STAT \rightarrow **CALC**

8: LinReg(a+bx)

$a = 1.584 \approx 2$

$b = 2.810 \approx 3$

$r^2 = .976$

$r = .988$

$\rightarrow y = 2 + 3x$

If r^2 & r are missing,

2nd **0** $\downarrow \downarrow \downarrow \dots \downarrow$ **DiagnosticOn**

Enter **Enter**

now redo last steps

Mar 5-7:16 PM

I randomly selected 5 students, chart below shows Study time (hrs) and QZ Scores

Study time	QZ Score
3	8
2	5
2	6
4	8
5	10

1) Draw Scatter Plot

Regression Line
 $y = a + bx$
 $y = 2.9 + 1.4x$

Study time $\rightarrow x \rightarrow L1$

QZ Score $\rightarrow y \rightarrow L2$

STAT \rightarrow **CALC**

8: LinReg(a+bx)

$a = 2.882 \rightarrow a \approx 2.9$

$b = 1.412 \rightarrow b \approx 1.4$

$r^2 = .892$

$r = .944$

r^2 (Coef. of determination) \rightarrow Always round to whole%

$r^2 \approx 89\%$

r (Linear Correlation Coef.) $\rightarrow -1 \leq r \leq 1$

$r = .944$

Mar 5-7:25 PM

Consider the chart below

x	y
3	8
4	12
5	14
8	20

1) Draw Scatter Plot

$x \rightarrow L1, y \rightarrow L2$
 use 2-Var Stats

$\sum x = 20$ $\sum y = 54$
 $\sum x^2 = 114$ $\sum y^2 = 804$
 $n = 4$ $\sum xy = 302$

use Lin Reg(a+bx) to find

Regression Line $y = 2.1 + 2.3x$ $a = 2.071 \approx 2.1$
 $b = 2.286 \approx 2.3$

Coeff. of determination $r^2 \approx 98\%$ $r^2 = .975 \approx 98\%$
 $r = .988$

Linear Correlation Coef. $r = .988$

Mar 5-7:36 PM

How to find a & b using formulas:

$$a = \frac{\sum y \cdot \sum x^2 - \sum x \cdot \sum xy}{n \sum x^2 - (\sum x)^2} = \frac{54 \cdot 114 - 20 \cdot 302}{4 \cdot 114 - 20^2}$$

$$= \frac{116}{56} \approx 2.071 \approx \boxed{2.1}$$

$$b = \frac{n \sum xy - \sum x \cdot \sum y}{n \sum x^2 - (\sum x)^2} = \frac{4 \cdot 302 - 20 \cdot 54}{4 \cdot 114 - 20^2} = \frac{128}{56}$$

$$\approx 2.286 \approx \boxed{2.3}$$

Mar 5-7:46 PM

How to find r using formula

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

$\sum x = 20$
 $\sum x^2 = 114$
 $n = 4$
 $\sum y = 54$
 $\sum y^2 = 804$
 $\sum xy = 302$

$$r = \frac{4 \cdot 302 - 20 \cdot 54}{\sqrt{4 \cdot 114 - (20)^2} \sqrt{4 \cdot 804 - 54^2}} = \frac{128}{\sqrt{56} \sqrt{300}}$$

$$= \frac{128}{\sqrt{16800}} \approx \boxed{.988} \leftarrow \text{Linear Correlation Coef.}$$

128 ÷ `[end]` `[x2]` 16800 `[Enter]`

$$r^2 = .988^2 = .976 \approx \boxed{98\%} \leftarrow \text{Coef. of determination}$$

Mar 5-7:54 PM

More on Linear Correlation Coef. r

- 1) $-1 \leq r \leq 1$
- 2) when r is close to ± 1 :
Linear Correlation is Significant.
- 3) when r is close to 0:
Linear Correlation is not Significant.

Making Prediction

- 1) **If r is significant**
Use the regression line, plug in x-value, to find y-value
- 2) **If r is not significant**
use \bar{y} as prediction value
 $\bar{y} = \frac{\sum y}{n}$ or `VARs`
`[5: Statistics]`
`[5: 5]` `[Enter]`

Mar 5-8:14 PM

Chart below is for study time & exam score

Time	Score
5	75
6	82
6	85
8	90
9	98

Time $\rightarrow x \rightarrow L1$
 Score $\rightarrow y \rightarrow L2$
 use **Lin Reg(a+bx)**
 $a = 51.370 \approx 51$
 $b = 5.093 \approx 5 \rightarrow y = 51 + 5x$
 $r^2 = .940 \rightarrow r^2 \approx 94\%$
 $r = .969 \rightarrow r = .969$

Predict exam score for someone with study time 4.

1) Assume r is significant $\rightarrow y = 51 + 5x$
 $= 51 + 5(4) = 71$

2) Assume r is not significant
 \rightarrow use \bar{y} **VARS** **5: Statistics** **5: \bar{y}**
Enter
86

Mar 5-8:22 PM

I randomly selected 7 day, below is my walk time and next day blood sugar level

Time	B.S. level
20	130
15	145
25	120
30	100
30	120
10	140
20	125

Time $\rightarrow x \rightarrow L1$
 BS level $\rightarrow y \rightarrow L2$
 use **Lin Reg(a+bx)**
 $a = 162.872 \approx 163$
 $b = -1.734 \approx -2$
 $r^2 = .764$
 $r = -.874$

Regression line **$y = 163 - 2x$**

Coef. of determination $r^2 \approx 76\%$
 Linear Correlation Coef. $r = -.874$

Mar 5-8:33 PM

Predict my BS level if I walk for 25 minutes

1) Assume r is significant
 use Regression line
 $y = 163 - 2x \Rightarrow 163 - 2(25) = \boxed{113}$

2) Assume r is not significant
 use \bar{y} VARS 5: Statistics 5: \bar{y} Enter
 $\bar{y} = 125.714 \approx \boxed{126}$

what does r^2 tell us?
 $r^2(\%)$ tells us what% of Y -values are explained by x -values.
 using this last example, 76% of my BS level are explained by my walking time.
 24% are unexplained.

Mar 5-8:42 PM

x	y
3	7
4	10
5	10
8	14
10	18

$x \rightarrow L1, y \rightarrow L2$
 use Lin Reg ($a+bx$) to find
 $a = 3.153 \approx 3.2$
 $b = 1.441 \approx 1.4 \Rightarrow y = 3.2 + 1.4x$
 $r^2 = .970 \Rightarrow r^2 \approx 97\%$
 $r = .985 \Rightarrow r = .985$

Predict y when $x = 5$

1) Assume r is significant
 use regression line
 $y = 3.2 + 1.4(5) = \boxed{10.2}$

2) Assume r is not significant
 use \bar{y}
 $\bar{y} = 11.8$

$r^2 = \boxed{.97}$ of Y -values are explained by x -values.
 3% are unexplained.

SG 9 ✓

Mar 5-8:51 PM

Class QZ 6

Use the chart below

x	y
5	12
6	15
6	18
7	20
4	10
8	20

Find

$$a = -.967 \approx \boxed{-1} \quad \left. \vphantom{a} \right\} \text{Round to whole \#}$$

$$b = 2.8 \approx \boxed{3}$$

$$r^2 = .883 \approx \boxed{88\%} \quad \left. \vphantom{r^2} \right\} \text{Round to whole \%}$$

$$r = \boxed{.939} \quad \left. \vphantom{r} \right\} \text{Round to 3-decimal}$$

Mar 5-9:05 PM