Statistics

Lecture 6

Feb 19-8:47 AM

500 Students randomly Selected and below is the 5-Number Summary of their ages

$500 \div 4 = 125$

16 → Min

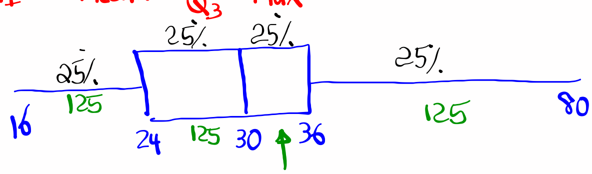
24 → Q_1

30 → Median

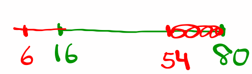
36 → Q_3

80 → Max

Box Plot



$IQR = Q_3 - Q_1 = 36 - 24 = 12$



Upper Fence $Q_3 + 1.5(IQR) = 36 + 1.5(12) = 54$

Lower Fence $Q_1 - 1.5(IQR) = 24 - 1.5(12) = 6$

Discuss outliers 54 – 80

How many students were at least 24 Yrs old?

75% of 500 = $\boxed{375}$

Jan 18-4:36 PM

500 students were randomly selected, their ages had a bell-shape dist. with $\bar{x} = 25$ and standard deviation 7.5 Yrs.

$\Rightarrow$ Symmetric

1) Usual Range

$\bar{x} \pm 2S = 25 \pm 2(7.5) = 25 \pm 15$

$\Rightarrow 10$ to 40

95% Range

2) How many of them were at least 40?

$100\% - 95\% = 5\%$

$5\% \div 2 = 2.5\%$

2.5% 95% 2.5%

10 40

2.5% of $500 = 12.5$ about 13

Jan 18-4:44 PM

Consider the Stem Plot below

1	0 5
2	0 3 5 8
3	0 2 4 5 5 9
4	2 6 7 8 8
5	3 4 7 7
6	0 5

1) $n = 23$

2) Find P_{30}

$L = \frac{30}{100} \cdot 23 = 6.9 \Rightarrow L = 7$

$P_{30} = 7\text{th element} = 30$

3) Find K such that $P_K = 55$

Below

$K = \frac{B}{n} \cdot 100 = \frac{19}{23} \cdot 100 = 82.60\ldots$

$\approx \boxed{83}$

83% 17%

$P_{83} = 55$

Jan 18-4:50 PM

Consider the Stem Plot below

6	58
7	0258
8	0035559
9	02589
10	05

1) $n = 20$

2) Range $= 105 - 65 = 40$

3) Estimate $S \approx \frac{\text{Range}}{4} = 10$

4) Find P_{50} Median

$L = \frac{50}{100} \cdot 20 = 10$

$P_{50} = \frac{10\text{th} + \text{Next one}}{2}$

$= \frac{85 + 85}{2} = 85$

5) Find K Such that $P_K = 90$

Below

$K = \frac{B}{n} \cdot 100 = \frac{13}{20} \cdot 100 = \boxed{65}$

$P_{65} = 90$

65% 35%

$P_{65} = 90$

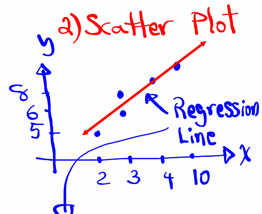
Jan 18-4:57 PM

Working with ordered-Pairs (x, y) SG 9

x	y	x^2	y^2	xy
2	5	4	25	10
3	8	9	64	24
3	6	9	36	18
4	9	16	81	36
5	10	25	100	50

1) $n = 5$

2) Scatter Plot



$y = a + bx$

3) $\sum x = 17$ $\sum y = 38$

$\sum x^2 = 63$ $\sum y^2 = 306$

$n = 5$ $\sum xy = 138$

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

Reset all lists $\boxed{\text{STAT}}$ Edit $\boxed{\text{5:SetupEditor}}$ $\boxed{\text{Enter}}$

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

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

L1	L2
2	5
3	8
3	6
4	9
5	10

$\boxed{\to}$

Quit & clear Screen

$\boxed{\text{2nd}}$ $\boxed{\text{MODE}}$ $\boxed{\text{Clear}}$

Jan 18-5:04 PM

How to find $\sum x, \sum x^2, \sum y, \sum y^2, \sum xy$

STAT → CALC

2:2-Var Stats

No Menu

2-Var Stats

L1 , L2 Enter

7

with Menu

Xlist: L1

ylist: L2

FreqList: clear

Calculate

$\sum x$ $\sum y$

$\sum x^2$ $\sum y^2$

$n = 5$ $\sum xy$

Jan 18-5:17 PM

How do we find the $\underbrace{\text{equation of the regression line}}$?

$x \to L1$

$y \to L2$

STAT → CALC

$y = a + bx$

~~4:LinReg(ax+b)~~

8: LinReg(a+bx)

No Menu

L1 , L2 Enter

7

$$\boxed{\begin{array}{l} y = a + bx \\ a = 1.846 \\ b = 1.692 \\ r^2 = .866 \\ r = .931 \end{array}}$$

with Menu

Xlist: L1

ylist: L2

FreqList: clear

Store...: clear

Calculate

$\boxed{y = 1.846 + 1.692x}$

Regression line

In Case r & r^2 missing:

2nd 0 ↓↓↓---↓ ▶DiagnosticOn Enter Enter

Jan 18-5:21 PM

What is r^2?

r^2 is the Coef. of determination.

Always express as whole %.

r^2 tells us what percent of Y-values are explained by x-Values.

in the last example $r^2 = .866 \approx 87\%$

87% of Y-values are explained by x-Values.

Jan 18-5:30 PM

What about r?

r is linear Correlation Coef.

$-1 \leq r \leq 1$

when r is close to ± 1,

$\Rightarrow$ Linear Correlation is Significant

when r is close to 0,

$\Rightarrow$ Linear Correlation is not Significant

from last example, $r = .931 \Rightarrow$ close to 1

$\Rightarrow$ Linear Correlation is Significant.

Jan 18-5:34 PM

Study time	Exam Score
5	75
6	82
6	90
5	80
10	95

Study time → x → L1

Exam Score → y → L2

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

8: LinReg(a+bx)

xlist : L1

ylist : L2

$\boxed{\text{clear}}$

$\boxed{\text{Calculate}}$

$y = a + bx$

$a = 63.488 \approx 63$

$b = 3.267 \approx 3$ $\Rightarrow \boxed{y = 63 + 3x}$

$r^2 = .714$ $\Rightarrow r^2 \approx 71\%$

$r = .845$

71% of exam Scores are explained by study time.

Linear Correlation Coef is close to 1

therefore linear Correlation is Significant.

Jan 18-5:38 PM

Recall There were two branches in Statistics:

1) Descriptive

2) Inferential

when we use data to make predictions

How To make predictions:

1) If r is Significant, use the regression line. Plug in the x-Value, find y-Value.

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

$\bar{y} = \frac{\sum y}{n}$ or $\boxed{\text{VARS}}$

$\boxed{\text{5: Statistics}}$

$\boxed{5: \bar{y}}$

$\boxed{\text{Enter}}$

84.4

≈ 84

From last example, Predict exam Score for Someone who studied 8 hrs. Assume r is Significant.

$y = 63 + 3x$

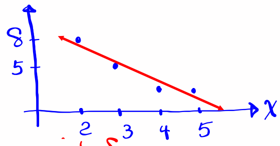
$= 63 + 3(8) = 63 + 24 = \boxed{87}$

Jan 18-5:46 PM

Complete the chart below

x	y	x^2	y^2	xy
2	8	4	64	16
3	5	9	25	15
4	2	16	4	8
5	2	25	4	10

Scatter Plot



what are the formula for a & b for the regression line $y = a + bx$?

$\sum x = 14$ $\sum y = 17$

$\sum x^2 = 54$ $\sum y^2 = 97$

$n = 4$ $\sum xy = 49$

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

$$= \frac{17 \cdot 54 - 14 \cdot 49}{4 \cdot 54 - (14)^2} = \frac{232}{20} = \boxed{11.6}$$

$$b = \frac{n \sum xy - \sum x \cdot \sum y}{n \sum x^2 - (\sum x)^2} = \frac{4 \cdot 49 - 14 \cdot 17}{4 \cdot 54 - 14^2} = \frac{-42}{20} = \boxed{-2.1}$$

$$y = a + bx \qquad \boxed{y = 11.6 - 2.1x}$$

Using TI

$x \to L1$

$y \to L2$

STAT → CALC

8: LinReg(a+bx)

L1 & L2

$a = 11.6$

$b = -2.1$

$r^2 = .891$

$r = -.944$

Jan 18-6:06 PM

Now formula for r:

$\sum x = 14$ $\sum y = 17$

$\sum x^2 = 54$ $\sum y^2 = 97$

$n = 4$ $\sum xy = 49$

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

$$= \frac{4 \cdot 49 - 14 \cdot 17}{\sqrt{4 \cdot 54 - (14)^2} \sqrt{4 \cdot 97 - 17^2}} = \frac{-42}{\sqrt{20} \sqrt{99}}$$

$$= \frac{-42}{\sqrt{1980}} = \boxed{-.944}$$

42 ÷ 2nd x^2 1980 Enter

what about r^2?

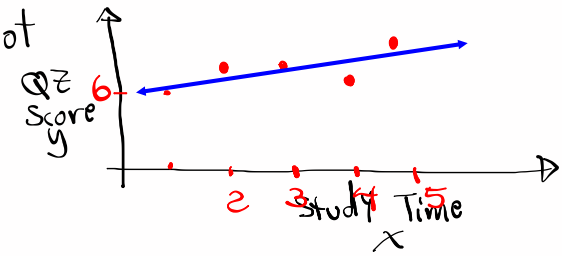
$$(-.944)^2 = \boxed{.891}$$

Jan 18-6:20 PM

Chart below shows Study time & QZ Scores

Study time (x)	QZ Score (y)	x^2	y^2	xy
1	6	1	36	6
2	8	4	64	16
3	8	9	64	24
4	7	16	49	28
5	10	25	100	50

Scatter Plot



$\sum x = 15$ $\sum y = 39$

$\sum x^2 = 55$ $\sum y^2 = 313$

$n = 5$ $\sum xy = 124$

Jan 18-6:28 PM

$\sum x = 15$ $\sum y = 39$

$\sum x^2 = 55$ $\sum y^2 = 313$

$n = 5$ $\sum xy = 124$

use Formula to find the eqn of the regression line

$y = a + bx$

$$a = \frac{\sum y \cdot \sum x^2 - \sum x \cdot \sum xy}{n \sum x^2 - (\sum x)^2} = \frac{39 \cdot 55 - 15 \cdot 124}{5 \cdot 55 - 15^2} = \frac{285}{50} = \boxed{5.7}$$

$$b = \frac{n \sum xy - \sum x \cdot \sum y}{n \sum x^2 - (\sum x)^2} = \frac{5 \cdot 124 - 15 \cdot 39}{5 \cdot 55 - 15^2} = \frac{35}{50} = \boxed{.7}$$

$$\boxed{y = 5.7 + .7x}$$

Jan 18-6:34 PM

Find r using Formula

$\sum x = 15$ $\sum y = 39$

$\sum x^2 = 55$ $\sum y^2 = 313$

$n = 5$ $\sum xy = 124$

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

$$= \frac{5 \cdot 124 - 15 \cdot 39}{\sqrt{5 \cdot 55 - 15^2}\ \sqrt{5 \cdot 313 - 39^2}} = \frac{35}{\sqrt{50}\ \sqrt{44}} = \frac{35}{\sqrt{2200}}$$

35 $\boxed{\div}$ $\boxed{\text{2nd}}$ $\boxed{x^2}$ 2200 $\boxed{\text{Enter}}$ $= \boxed{.746}$

$r^2 = (.746)^2$

≈ 55.7

$\% \approx 56\%$

Jan 18-6:39 PM

Given : $n = 10$, $\sum y = 825$, $y = 65 + 5x$

Predict y for $x = 4$

1) Assume r is Significant → use regression line

$y = 65 + 5x$

$= 65 + 5(4) = \boxed{85}$

2) Assume r is not Significant. → use $\bar{y}$

$$\bar{y} = \frac{\sum y}{n} = \frac{825}{10} = \boxed{82.5}$$

Jan 18-6:45 PM