Statistics

Lecture 5

Feb 19-8:47 AM

In-Person QZ 3 (open Notes)

Use the data below

35	28	30	20	25
40	42	38	18	25

Find

1) $\bar{x} = 30.1 \approx \boxed{30}$ ✓

2) $S = 8.373 \approx \boxed{8}$ ✓

} Round to whole #

3) $S^2 = \boxed{\frac{701}{10}}$ ✓ } Reduced fraction

68% Range $\bar{x} \pm S$

$= 30 \pm 8 \rightarrow \boxed{22 \text{ to } 38}$

95% Range $\bar{x} \pm 2S$

usual Range $= 30 \pm 2(8) \rightarrow \boxed{14 \text{ to } 46}$

99.7% Range $\bar{x} \pm 3S$

$= 30 \pm 3(8) \rightarrow \boxed{6 \text{ to } 54}$

Find z-Score for data element 42.

$z = \frac{x - \bar{x}}{S} = \frac{42 - 30}{8} = \frac{12}{8} = \boxed{1.5}$ usual data element

Find a data element with -1.5 z-Score.

$z = \frac{x - \bar{x}}{S}$ $\quad -1.5 = \frac{x - 30}{8}$

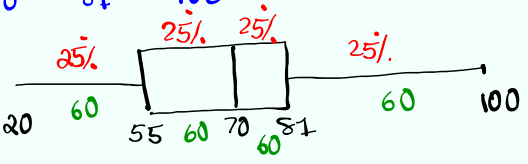
$x - 30 = 8(-1.5)$ $\quad x = 30 - 12 = \boxed{18}$

Jul 1-7:02 PM

I took a sample of 240 exams, 5-Number Summary of Scores were $240 \div 4 = 60$

20 55 70 81 100

1) Box Plot



2) $IQR = Q_3 - Q_1 = 81 - 55 = \boxed{26}$

3) Upper fence $= Q_3 + 1.5(IQR) = 81 + 1.5(26) = \boxed{120}$

Lower fence $= Q_1 - 1.5(IQR) = 55 - 1.5(26) = \boxed{16}$

4) what about outliers?

Nothing Below 16

Nothing above 120

$\Rightarrow$ No outliers

Jul 2-4:40 PM

I randomly selected 32 nurses, here are their ages:

2	48
3	0225558
4	0134666889
5	023558
6	0235558

$n = 32$

Range $= 68 - 24 = 44$

Estimated value for

$S \approx \frac{\text{Range}}{4} = \frac{44}{4} = \boxed{11}$

P_{20}

$L = \frac{20}{100} \cdot 32 = 6.4 \rightarrow L = 7$

$P_{20} = 7\text{th element} = 35$

P_{50}

$L = \frac{50}{100} \cdot 32 = 16$

$P_{50} = \frac{16\text{th element} + \text{Next element}}{2} = \frac{46 + 48}{2} = \boxed{47}$

Find K such that $P_K = 60$

$\boxed{P_{78} = 60}$

$K = \frac{B}{n} \cdot 100 = \frac{25}{32} \cdot 100 = 78.125 \approx 78$ (B: below)

Jul 2-4:45 PM

Clear all lists.

Quit 2nd MODE

Store the following in L1

43	25	18	50	60
40	28	20	55	62
19	24	35	35	38
42	39	30	60	65
58	48	52	42	18

Sort L1, then view it, and make STEM Plot

STAT Edit

2: SortA(L1 Enter

2nd 1 Enter

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

Find

$\bar{x} = 40.24 \approx 40.2$ } Round to 1-decimal

$S = 14.959 \approx 15.0$ }

$S^2 = \frac{1683}{75}$ } Red. Frac.

STAT → CALC

1: 1-Var Stats

VARS 5: Statistics

3: S_x x^2

Math 1: ▶Frac Enter

Jul 2-4:55 PM

How to find the 5-Number Summary:

Follow steps to find $\bar{x}$ & S,

Arrow down to get

5-Number Summary

STAT → CACL

1: 1-Var Stats

$\bar{x} = 40.24$

$S = S_x = 14.959$

$n = 25$

Min = 18

$Q_1 = 26.5$

Med = 40

$Q_3 = 53.5$

Max = 65

Jul 2-5:09 PM

Use the chart below

class MP	class F
15	5
31	9
47	13
63	17
79	6

1) How many classes?
5

2) Class width
31 − 15 = 16

3) Sample Size
$n = \Sigma f = 50$

Use class MP & class F to find

$\bar{x} = 50.2 \approx 50$ } Round to whole #

$s = 18.848 \approx 19$ } Round to whole #

$s^2 = \frac{17408}{49}$ } Reduced fraction

SG 7 & 8 ✓

class MP → L1
" f → L2

STAT → CALC
1: 1-Var Stats
List: L1
FreqList: L2
Calculate

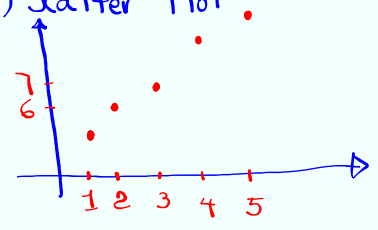
Jul 2-5:13 PM

SG 9

x	y
3	7
2	6
5	12
4	10
1	5

L1 L2

1) Scatter Plot



x → L1, y → L2 use 2-Var Stats to find

$\Sigma x = 15$ $n = 5$ $\Sigma y^2 = 354$

$\Sigma x^2 = 55$ $\Sigma y = 40$ $\Sigma xy = 138$

STAT → CALC

2: 2-Var Stats

Menu
Xlist: L1
Ylist: L2
FreqList: clear
Calculate

No Menu
L1 , L2 Enter
7

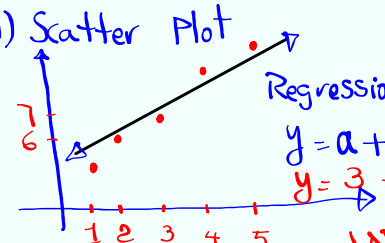
Jul 2-5:36 PM

SG 9

x	y
3	7
2	6
5	12
4	10
1	5

L1 L2

1) Scatter Plot



Regression line

$y = a + bx$

$y = 3 + 2x$

STAT → CALC

8: LinReg(a+bx)

Menu

Xlist: L1

Ylist: L2

No Menu

L1, L2

Enter

Clear

Calculate

If r & r^2 are missing,

2nd 0 ↓ ↓ ··· ↓

▶ Diagnostic On Enter Enter

$a = 2.6 \approx 3$

$b = 1.8 \approx 2$

$r^2 = .953$

$r = .976$

Jul 2-5:36 PM

How to find a & b using formula:

$$a = \frac{\Sigma y \, \Sigma x^2 - \Sigma x \, \Sigma xy}{n \Sigma x^2 - (\Sigma x)^2} = \frac{40 \cdot 55 - 15 \cdot 138}{5 \cdot 55 - 15^2}$$

$$= \frac{130}{50} = \boxed{2.6}$$

$$b = \frac{n \Sigma xy - \Sigma x \Sigma y}{n \Sigma x^2 - (\Sigma x)^2} \quad \frac{5 \cdot 138 - 15 \cdot 40}{5 \cdot 55 - 15^2}$$

$$= \frac{90}{50} = \boxed{1.8}$$

Jul 2-5:50 PM

Qz Score	Exam Score
8	85
7	80
6	68
9	88
10	98
10	90

Qz Score → x → L1

Exam Score → y → L2

STAT → CALC

8:LinReg(a+bx)

using L1 & L2

$a = 35.875 \approx 36$

$b = 5.875 \approx 6$

$r^2 = .890$

$r = .944$

Regression line

$y = a + bx$

$\boxed{y \approx 36 + 6x}$

Jul 2-5:57 PM

what is r?

Linear Correlation Coef.

$-1 \le r \le 1$

when r is close to ±1,

Linear Correlation is Significant.

when r is close to 0,

Linear Correlation is not Significant.

from Last example

$r = .944$ → It is fairly close to 1.

⇒ It is Significant.

Jul 2-6:03 PM

How to find r using formula:

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

use [2-Var Stats] with L1 & L2

$\Sigma x = 50$ $\quad \Sigma y = 509$

$\Sigma x^2 = 430$ $\quad \Sigma y^2 = 43697$

$n = 6$ $\quad \Sigma xy = 4320$

$$r = \frac{6 \cdot 4320 - 50 \cdot 509}{\sqrt{6 \cdot 430 - 50^2}\ \sqrt{6 \cdot 43697 - 509^2}}$$

$$= \frac{470}{\sqrt{80}\ \sqrt{3101}}$$

470 [÷] [2nd] [x^2] 248080 [Enter]

$$= \frac{470}{\sqrt{248080}}$$

$= .944$

$r^2 = (.944)^2 \Rightarrow r^2 \approx .890$

↑ Coef. of determination (%) $\approx 89\%$

Jul 2-6:07 PM

What does r^2 tell us?

It tells what % of Y-values are explained by x-values.

From last example $r^2 \approx 89\%$

89% of exam scores are explained by QZ-scores. 11% are unexplained

Jul 2-6:18 PM

Two Branches

1) Descriptive

2) Inferential

If r is Significant

⇒ use the regression line to make Predictions.

If r is not Significant

⇒ use $\bar{y}$ as prediction value

$\bar{y} = \frac{\Sigma y}{n}$ OR [VARS] [5: Statistics] [5: $\bar{y}$] [Enter]

Jul 2-6:34 PM

Study time	Exam Score
5	72
6	80
6	85
8	92
12	95

$\hat{y} = 64 + 3x$

73% of exam Scores are explained by Study time.

Study time → x → L1

Exam Score → y → L2

[STAT] [→] CALC

[8: LinReg(a+bx)]

Use L1 & L2

$a = 63.833 \approx 64$

$b = 2.833 \approx 3$

$r^2 = .731 \approx 73\%$

$r = .855$

Jul 2-6:37 PM

Predict exam score for someone who studied 8 hrs.

a) Assume r is Significant.

use Regression line

$y = 64 + 3x = 64 + 3(8) = \boxed{88}$

b) Assume r is not Significant.

use $\bar{y}$

VARS | 5: Statistics | 5: $\bar{y}$ | Enter

$\bar{y} = 84.8 \approx \boxed{85}$

Jul 2-6:44 PM

Walk time	BS level
30	120
20	135
10	140
25	125
40	100

Walk time → X → L1

BS level → y → L2

STAT → CALC

8: LinReg(a+bx)

Using L1 & L2

$a = 157.75$

$b = -1.35$

$r^2 = .939 \approx 94\%$

$r = -.969$

$\boxed{y = 157.75 - 1.35x}$

$y \approx 158 - x$

Jul 2-6:48 PM

Predict my BS level if I walk 25 minutes

1) Assume r is Significant
use Reg. line
$y = 158 - x = 158 - 25 = \boxed{133}$

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

VARS | 5:Statistics | 5: $\bar{y}$ | Enter → 124

Jul 2-6:54 PM

In-Person QZ 4

Use the chart below

MP	F
18	5
30	8
42	12
54	6

Find

1) $\bar{x} = 37.355 \approx \boxed{37}$
2) $S = 11.870 \approx \boxed{12}$
3) $n = \boxed{31}$
4) $s^2 = \boxed{\frac{4368}{31}}$

} Round to whole #
} Red. fraction

Jul 2-6:58 PM