

Statistics

Lecture 4



Feb 19-8:47 AM

Class QZ 1

Consider the Sample below

6	7	8	6
5	4	9	10

Use Your Calc. to find

- 1) $\bar{x}$
- 2) S
- 3) S^2

Round to whole #
Reduced fraction

Clear all lists
2nd [] 4: Clear All Lists [Enter]

Reset all lists
[STAT] Edit [Enter]
[5: SetupEditor]

[STAT] Edit
[1: Edit]

L1
6
7
8
...
10

quit & clear Screen
[2nd] [Mode] [clear]

[STAT] [→] CALC
[1: 1-Var Stats]

With Menu
List: L1
Freq List: [clear]
[Calculate] [Enter]
[end] [1]

NO Menu
1-Var Stats
L1
[Enter]

$\bar{x} = 6.875 \approx 7$

$S = S_x = 2.031 \approx 2$

[VARS] [5: Statistics] [3: S_x]

χ^2 [Math] [1: $\frac{\Box}{\Box}$] [Enter]

$S^2 = \frac{33}{8}$

Sep 19-11:23 AM

$$\bar{x} = 7$$

By empirical rule

$$S = 2$$

Usual Range (95% Range)

$$\bar{x} \pm 2S$$

$$= 7 \pm 2(2) = 7 \pm 4 \Rightarrow \boxed{3 \text{ to } 11}$$

Find the Z-Score for data element 10

$$Z = \frac{x - \bar{x}}{S} = \frac{10 - 7}{2} = \frac{3}{2} = \boxed{1.5}$$

Find the data element with $Z = -2.5$

$$Z = \frac{x - \bar{x}}{S} \quad -2.5 = \frac{x - 7}{2}$$

Cross-Multiply $x - 7 = 2(-2.5)$

$$x = 7 - 5$$

$$\boxed{x = 2}$$

Sep 19-11:49 AM

I randomly Selected 25 Students, here are their ages

32 28 19 40 45

18 20 25 50 52

19 24 30 30 48

50 60 27 35 35

42 58 45 33 29

clear all lists

[2nd] [1] [4:clear...] [Enter]

Store this Sample

in L1 [STAT] Edit

[1:Edit]

Sort L1

[STAT] Edit

[2nd] [1] [Enter]

[2:SortA]

View L1 and make STEM Plot

[2nd] [1] [Enter]

1	899
2	045789
3	002355
4	02558
5	0028
6	0

Sep 19-11:54 AM

use Your Calc to find

$\bar{x} = 35.8$
 $S = S_x = 12.6$
 S^2

Round to 1-dec
 Reduced Fraction

STAT → CALC
 1:1-VarStats
 List: L1
 FreqList: Clear
 Calculate

use & Several times

Min = 18
 $Q_1 = 26$
 Med = 33
 $Q_3 = 46.5 \approx 47$
 Max = 60

VARs
 5: Statistics
 3: S_x
 χ^2
 Math 1: Frac
 Enter

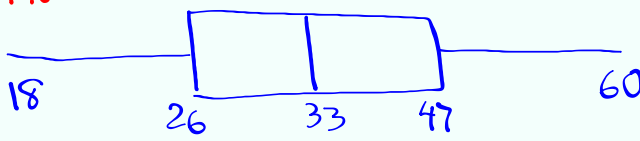
$S^2 = \frac{47257}{300}$

Sep 19-12:05 PM

5-Number Summary

18 26 33 47 60

Draw Box Plot



$IQR = Q_3 - Q_1 = 21$

Upper Fence = $Q_3 + 1.5(IQR) = 47 + 1.5(21) = 78.5$

Lower Fence = $Q_1 - 1.5(IQR) = 26 - 1.5(21) = -5.5$

Discuss outliers None

Sep 19-12:12 PM

1) $n=25$

2) Range = $60 - 18 = 42$

3) Mode 19, 30, 35, 45, 50
Multi modal

4) Estimate $S \approx \frac{\text{Range}}{4}$
 $= \frac{42}{4} = 10.5$

5) How many data elements are below 40?
15

6) What % of data elements are below 40?
15 is what % of 25? $15 = \frac{P}{100} \cdot 25$
 $\frac{15}{25} \cdot 100$ 60%

60% ——— 40% ———
40

Intro to Percentile
"Data must be Sorted"

Notation P_K $K\%$ $(100-K)\%$
 P_K

in our example
 $P_{60} = 40$

60% ——— 40% ———
 $P_{60} = 40$

Sep 19-12:20 PM

1) 899

2) 045789

3) 002355

4) 02558

5) 0028

6) 0

what % of data elements are below 30?

$\frac{9}{25} \cdot 100 = 36\%$

36% ——— 64% ———
30

$P_{36} = 30$

Sep 19-12:30 PM

How to find P_K :

- 1) Make sure data is sorted
- 2) Find location $L = \frac{k}{100} \cdot n$

Sample Size

If L is decimal $\rightarrow$ Round-up to a whole #

$$P_K = L^{\text{th}} \text{ element}$$

If L is a whole # $\rightarrow$

$$P_K = \frac{L^{\text{th}} \text{ element} + \text{Next element}}{2}$$

Sep 19-12:32 PM

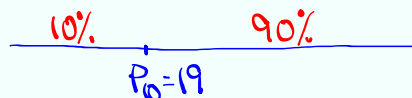
1	899
2	045789
3	002355
4	02558
5	0028
6	0

Find

$$P_{10}$$

$$L = \frac{10}{100} \cdot 25 = 2.5 \quad L=3$$

$$P_{10} = 3^{\text{rd}} \text{ element} = \boxed{19}$$

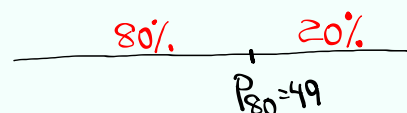


Find P_{80}

$$L = \frac{80}{100} \cdot 25 = 20$$

$$P_{80} = \frac{20^{\text{th}} \text{ element} + \text{Next element}}{2}$$

$$= \frac{48 + 50}{2} = \boxed{49}$$



Sep 19-12:36 PM

1	899	Find P_{50} ← Median
2	045789	
3	002355	
4	02558	
5	0028	
6	0	

$L = \frac{50}{100} \cdot 25 = 12.5$ $L = 13$
 $P_{50} = 13\text{th element}$
 $P_{50} = 33$

Doing Reverse

Find K such that $P_K = 50$

$K = \frac{B}{n} \cdot 100$ Round to whole %
 ↗ Sample Size ↖ Below

$K = \frac{20}{25} \cdot 100 = 80$ $P_{80} = 50$

Sep 19-12:40 PM

1	899	Find percentile ranking of 35.
2	045789	
3	002355	
4	02558	
5	0028	
6	0	

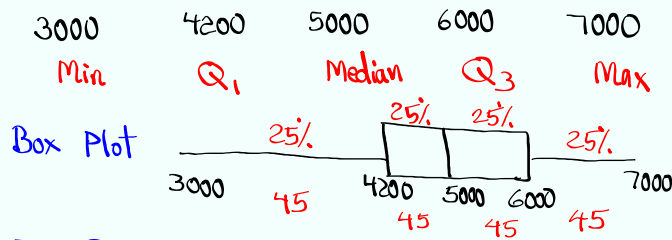
Find K such that $P_K = 35$

$K = \frac{B}{n} \cdot 100 = \frac{13}{25} \cdot 100$
 $= 52$

52% 48%
 —————
 $P_{52} = 35$

Sep 19-12:45 PM

Salaries of 180 randomly selected uber drivers had the following 5-Number Summary.



$$IQR = 6000 - 4200 = \boxed{1800}$$

Discuss outliers None

$$\text{Upper Fence} = Q_3 + 1.5(IQR) = 6000 + 1.5(1800)$$

$$= \boxed{8700}$$

Lower Fence

$$= Q_1 - 1.5(IQR) = 4200 - 1.5(1800)$$

$$= \boxed{1500}$$

Sep 19-1:04 PM

Complete the chart below

class limits	class MP	class F
18 - 28	23	5
29 - 39	34	12
40 - 50	45	8

1) How many classes?

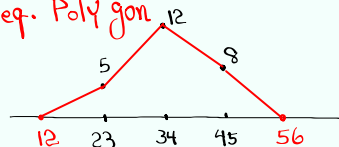
3

2) class width

11

3) Sample Size $n = \sum f = 25$

4) Draw Freq. Poly gon



class MP $\rightarrow$ L1

class F $\rightarrow$ L2

L1	L2
23	5
34	12
45	8

use 1-Var stats with L1 & L2 to find

$$\bar{x} = 35.32$$

$$s = 7.983$$

$$s^2 = \frac{9559}{150}$$

$$n = 25$$

STAT

EDIT CALC

(1-Var Stats)

List: L1

FreqList: L2

VAR

5: Statistics

3: Sx

χ^2

Math 1: frac

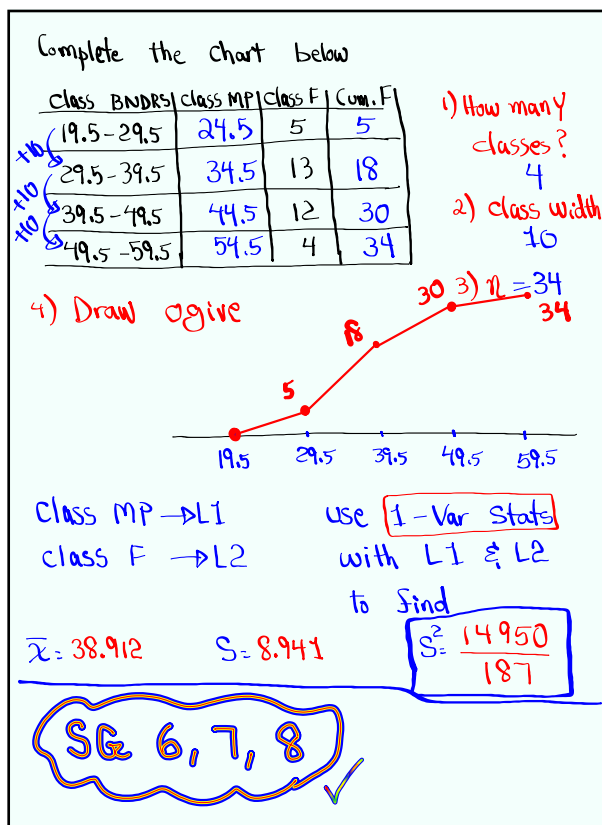
Enter

Calculate

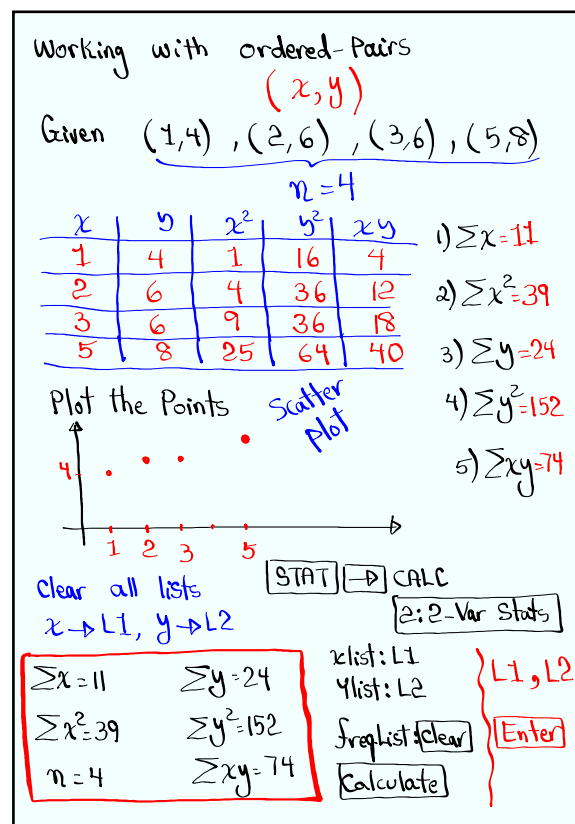
No Memo

L1, L2 Enter

Sep 19-1:12 PM



Sep 19-1:25 PM



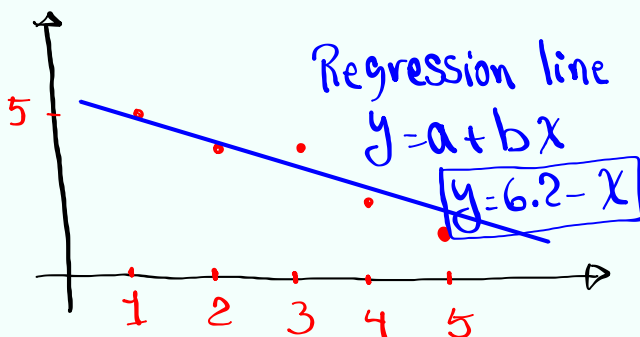
Sep 19-1:36 PM

Consider the ordered pairs below

(1,5) (2,4) (3,4) (4,2) (5,1)

Draw Scatter Plot

$x \rightarrow L1$, $y \rightarrow L2$



Use 2-Var Stats
to Find

$$\begin{aligned} \sum x &= 15 & \sum y &= 16 \\ \sum x^2 &= 55 & \sum y^2 &= 62 \\ n &= 5 & \sum xy &= 38 \end{aligned}$$

Sep 19-1:47 PM

[STAT] $\rightarrow$ CALC

8: LinReg(a+bx)

xlist: L1

ylist: L2

No Menu

L1, L2

[Enter]

[Calculate]

$a = 6.2$

$b = -1$

$r^2 = .926$

$r = -.962$

If r & r^2 are missing,

[2nd] [0] [down arrow] [DiagnosticOn]

[Enter]

[Enter]

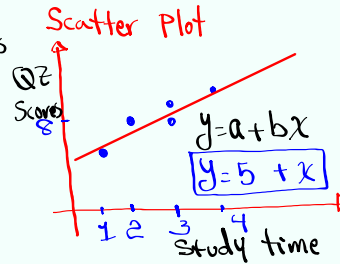
Sep 19-1:53 PM

I randomly selected 5 students.

Table below is for Study time (hrs) &

QZ Scores

Study time	QZ Scores
2	8
3	9
3	8
4	10
1	6



Study time $\rightarrow x \rightarrow L1$

QZ Score $\rightarrow y \rightarrow L2$

use LinReg(a+bx)
with L1 & L2

$$a = 5$$

$$b = 1.231 \approx 1$$

$$r^2 = .895$$

$$r = .946$$

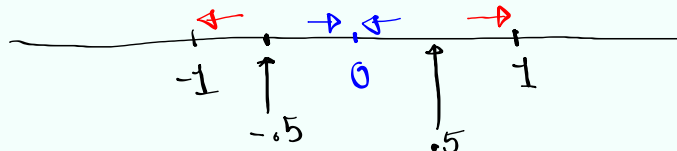
Sep 19-2:00 PM

$r \rightarrow$ 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.



Sep 19-2:08 PM

$r^2 \rightarrow$ Coefficient of determination
 $\rightarrow$ Always in whole %.

r^2 tells us what % of Y-values
 are explained by X-values.

From Last example $r^2 = .895 \approx 90\%$
 90% of QZ Scores are explained
 by study time.

Sep 19-2:13 PM

Walk(Mins)	Blood Sugar level	
10	130	walk $\rightarrow$ X $\rightarrow$ L1
15	120	BS level $\rightarrow$ Y $\rightarrow$ L2
10	125	Use LinReg(a+bx)
20	100	with L1 & L2
30	110	$a \approx 135$
Regression Line		$b \approx -1$
$y \approx 135 - x$		$r^2 = 54\%$
54% of my BS level are explained by how long I walk.		$r = -.732$
		r is close to -1
		Linear Correlation is significant.

Sep 19-2:17 PM