

Statistics

Lecture 7

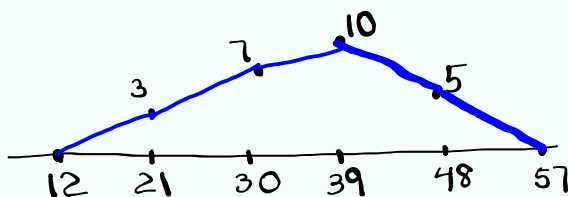


Feb 19-8:47 AM

Complete the Freq. table below

class limits	class BNDRS	class MP	class F	Cum. F	Rel. F	% F
17 - 25	16.5 - 25.5	21	3	3	.12	12%
26 - 34	25.5 - 34.5	30	7	10	.28	28%
35 - 43	34.5 - 43.5	39	10	20	.40	40%
44 - 52	43.5 - 52.5	48	5	25	.20	20%

4 classes , CW=9 , $n=25$, $Rel. F = \frac{f}{n} = \frac{f}{25}$



Mar 5-1:50 PM

Find \bar{x} & S of this group data.

clear All lists [2nd] + [4:ClearAllLists] [Enter]

Reset All lists [STAT] Edit [Enter]

5:SetupEditor

class MP \rightarrow L1

class F \rightarrow L2

[STAT] Edit [1:Edit]

L1	L2
21	3
30	7
39	10
48	5

[STAT] \rightarrow CALC

[1:1-Var Stats]

Menu List: L1 } No Menu

Freq List: L2 } L1, L2 [Enter]

[Calculate] [7]

$\bar{x} = 36.12$ Find S^2 in reduced fraction

$S = S_x = 8.506$ [VARS] [5:Statistics] [3:Stx]

$n = 25$ [x²] [MATH] [1:Div] [Enter]

$\bar{x} \approx 36, S \approx 9$ $S^2 = \frac{1809}{25}$

68% Range $\bar{x} \pm S = 36 \pm 9 \rightarrow [27 - 45]$

Usual Range $\bar{x} \pm 2S \rightarrow [18 - 54]$

95% Range

Mar 5-2:00 PM

I randomly selected 40 nurses. Here are their ages:

32	38	25	18	30
24	40	42	48	28
55	50	43	34	29
46	19	53	58	60
70	65	68	62	59
53	57	40	30	36
52	26	21	33	39
50	40	30	20	60

Store in L1

Sort L1

[STAT] Edit [2:SortA] [Enter]

View L1

[2nd] [1] [Enter]

{18 19 20 21 ...}

Make Stem Plot

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

find \bar{x} , S , and S^2 in reduced fraction.

$\bar{x} = 42.075$ $S^2 = \frac{340511}{1560}$

$S = 14.774$

Find 5-Number Summary

$n = 40$

Min = 18

$Q_1 = 30$

Med = 40

$Q_3 = 54$

Max = 70

IQR = $Q_3 - Q_1 = 24$

UF = $Q_3 + 1.5(IQR) = 90$

LF = $Q_1 - 1.5(IQR) = -6$

No outliers

Mar 5-2:13 PM

1 | 89
 2 | 01 456 89
 3 | 000 23 468 9
 4 | 000 23 68
 5 | 00 233 578 9
 6 | 00 258
 7 | 0

Find P_{20}
 $L = \frac{20 \cdot 40}{100} = 8$
 $P_{20} = \frac{8^{th} + 9^{th}}{2}$
 $= \frac{28 + 29}{2} = \boxed{28.5}$

Find P_{62}
 $L = \frac{62 \cdot 40}{100} = 24.8$ $L = 25$ $P_{62} = 25^{th} \text{ element}$
 $\rightarrow = \boxed{48}$

[2nd] [1] [25] [Enter]

Find k such that $P_k = 55$
 $k = \frac{B}{n} \cdot 100 = \frac{30}{40} \cdot 100 = 75$
 Round to whole%
 $P_{75} = 55$

Mar 5-2:32 PM

x	y	x ²	y ²	xy
2	5	4	25	10
4	10	16	100	40
3	8	9	64	24
5	12	25	144	60

$n=4$
 $x \rightarrow L1$
 $y \rightarrow L2$

STAT \rightarrow CALC
 2: 2-Var Stats No Menu
 Xlist: L1 L1, L2
 Ylist: L2
 Freqlist: clear [Enter]
 [Calculate]

STAT \rightarrow CALC
 8: Lin Reg(a+bx)
 Xlist: L1 L1, L2
 Ylist: L2 [Enter]
 [Diagnostic On]
 [Enter] [Enter] [Clear]
 [Calculate]

$\sum x = 14$
 $\sum x^2 = 54$
 $\sum y = 35$
 $\sum y^2 = 333$
 $\sum xy = 134$

$a = .7$ ✓
 $b = 2.3$ ✓
 $r^2 = .989$
 $r = .994$

Scatter Plot

Mar 5-2:39 PM

$\sum x = 14$
 $\sum x^2 = 54$
 $\sum y = 35$
 $\sum y^2 = 333$
 $\sum xy = 134$

Regression Line $y = a + bx$

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

$$= \frac{35 \cdot 54 - 14 \cdot 134}{4 \cdot 54 - 14^2} = \frac{14}{20} = 0.7$$

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

$$= \frac{4 \cdot 134 - 14 \cdot 35}{4 \cdot 54 - 14^2} = \frac{46}{20} = 2.3$$

Mar 5-3:02 PM

Linear Correlation Coefficient r

$-1 \leq r \leq 1$

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

$\sum x = 14$
 $\sum x^2 = 54$
 $\sum y = 35$
 $\sum y^2 = 333$
 $\sum xy = 134$

$$r = \frac{4 \cdot 134 - 14 \cdot 35}{\sqrt{4 \cdot 54 - 14^2} \sqrt{4 \cdot 333 - 35^2}}$$

$$= \frac{46}{\sqrt{20} \sqrt{101}} = \frac{46}{\sqrt{2010}}$$

46 \div $\sqrt{2010}$ \approx 1.994

when r is close to ± 1 ,
Linear Correlation is Significant

when r is close to 0,
Linear Correlation is not Significant

Mar 5-3:08 PM

what about r^2 ?

Always write r^2 in %.

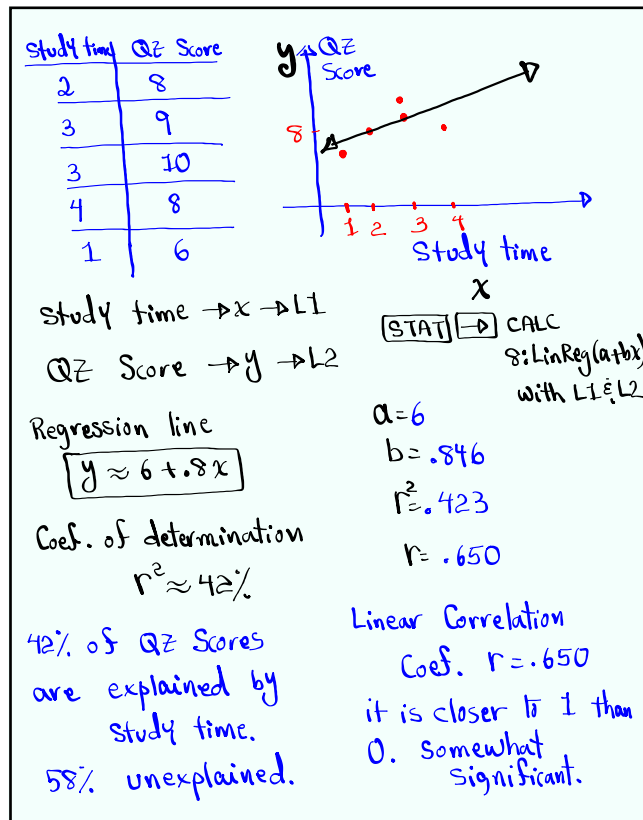
Coef. of determination r^2

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

$r^2 \approx 98.9\% \approx 99\%$

99% of y -values are explained by x -values

Mar 5-3:18 PM



Mar 5-3:23 PM

How to make Predictions:

when r is significant

use the regression line

when r is not significant

use \bar{y}

$$\bar{y} = \frac{\sum y}{n}$$

VARS

5: Statistics

5: \bar{y} Enter

Mar 5-3:32 PM

Predict QZ Score when study time is 4hrs

1) Assume r is significant

$$y = 6 + .8x$$

$$= 6 + .8(4) = 6 + 3.2 = \boxed{9.2} \approx 9$$

2) Assume r is not significant.

$$\text{use } \bar{y} = 8.2 \approx \boxed{8}$$

Mar 5-3:36 PM

QZ Score	Exam Score
8	85
7	80
9	92
10	90
6	75

QZ Score $\rightarrow X \rightarrow L1$
 Exam Score $\rightarrow Y \rightarrow L2$
 use LinReg(a+bx)
 $a = 50.8 \approx 51$
 $b = 4.2 \approx 4$

Regression line
 $y \approx 51 + 4x$
 $r^2 = .895 \approx 89\%$
 $r = .946$
 89% of exam scores are explained by QZ Scores.
 r is close to 1
 Linear Correlation is significant.

Predict exam score for QZ score 8.

1) Assume r is significant
 $y \approx 51 + 4x = 51 + 4(8) = 51 + 32 = 83$

2) Assume r is not significant
 use $\bar{y} = 84.4 \approx 84$

VARs
 5: Statistics 5: \bar{y} Enter

Mar 5-3:39 PM