

Scatter Plots

Linear Regression – Line of Best Fit

Coefficient of Correlation

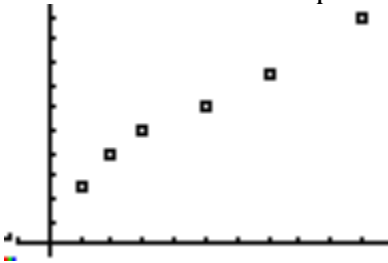
When given a set of ordered pairs, it is often useful to discover a relationship that can be used to predict outcomes that may not be directly given in the set.

Example: Given the set of data.

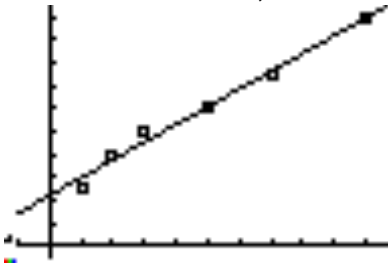
$\{(1, 5), (3, 10), (5, 12), (2, 8), (10, 20), (7, 15)\}$

By what method would you predict a good approximation for the outcome is the input was 50?

We will build a scatter plot of the data.



On the Scatter Plot, we draw a line that best fits the data.



We can pick points on the line and Write the Equation. This is what we can call "Linear Regression".

A possible approximate equation might be $y = \frac{3}{2}x + 4$

We can get a better equation when we use a calculator.

To Start with a TI-83

1. Press STAT
2. From the EDIT List Choose Edit
3. Under L1, Enter the x-values
4. Under L2, Enter the y-values

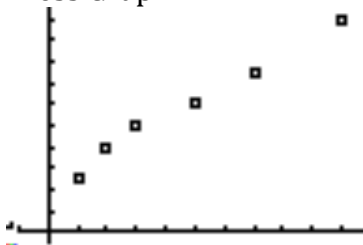
L1	L2	L3
1	5	
2	10	
3	12	
4	8	
10	20	
7	15	

L3(1)=

5. Press 2nd QUIT (MODE)
 6. Press CLEAR
 7. Press STAT
 8. Select the CALC Menu
 9. Select LinReg(ax+b)
 10. Press ENTER
- ```
LinReg
y=ax+b
a=1.558139535
b=4.395348837
```

11. Set the Window to the Following
- ```
WINDOW
Xmin=-1
Xmax=11
Xscl=1
Ymin=-1
Ymax=21
Yscl=2
Xres=1
```

12. To turn on the Plot, Press 2nd STAT PLOT (Y=)
13. Press ENTER
14. Press ENTER again to Highlight ON
15. Press Graph



16. Press Clear
17. Press STAT

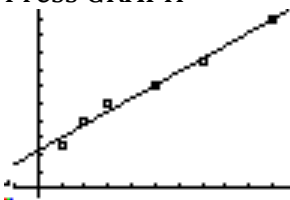
18. Select the CALC Menu
19. Select LinReg(ax+b)
20. Press VARS
21. Select the Y-VARS Menu
22. Select Function
23. Select Y1
24. Press ENTER
25. Press Y1

```

2nd 2nd Plot2 Plot3
} \Y1 1.5581395348
837X+4.395348837
2097
\Y2=
\Y3=
\Y4 X^
\Y5=

```

26. This is the Calculator Generated Line of Best Fit Equation $y = 1.558x + 4.395$
27. Press GRAPH



28. Now, using the generated function, we can predict the y-coordinate when any x-coordinate is given.

What y-value goes with the x-value of 52 ?

1. Press CLEAR
2. Press VARS
3. Select the Y-VARS Menu
4. Select Function
5. Select Y1
6. Inside Parentheses Place 52
7. Press ENTER

```

Y1(52)
85.41860465

```

8. We Get the Ordered Pair: (52, 85.41860465)

The value r is the Coefficient of Correlation. How do we find it?

1. Press 2nd Catalog (0)
2. Scroll Down to DiagnosticsOn and Select it. And Press ENTER Again.
3. Press 2nd QUIT (MODE)
4. Press CLEAR
5. Press STAT
6. Select the CALC Menu
7. Select LinReg(ax+b)
8. Press ENTER

```
LinReg
y=ax+b
a=1.558139535
b=4.395348837
r2=.9848617815
r=.9924020261
```

The correlation coefficient, denoted by r , is a number from -1 to 1 that measures how well a line fits a set of data pairs (x, y) . When r is near 1 , the points lie close to a line with a positive slope. When r is near -1 , the points lie close to a line with a negative slope. When r is near 0 , the points do not lie close to any line.

How do we interpret the relevance of the coefficient of correlation?

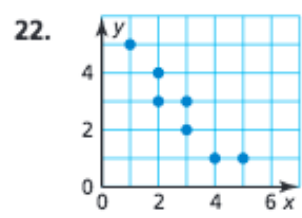
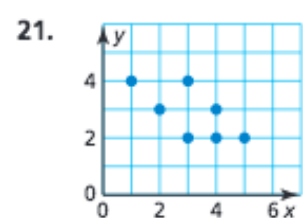
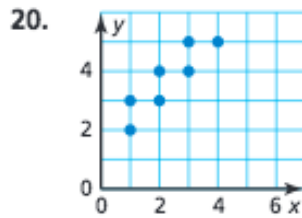
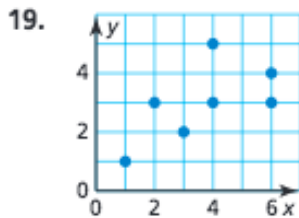
- | | | |
|----|-----------------|---|
| 1. | Exactly -1.00 | A perfect downhill (negative) linear relationship |
| 2. | -0.70 | A strong downhill (negative) linear relationship |
| 3. | -0.50 | A moderate downhill (negative) relationship |
| 4. | -0.30 | A weak downhill (negative) linear relationship |
| 5. | 0.00 | No linear relationship |
| 6. | $+0.30$ | A weak uphill (positive) linear relationship |
| 7. | $+0.50$ | A moderate uphill (positive) relationship |
| 8. | $+0.70$ | A strong uphill (positive) linear relationship |
| 9. | Exactly $+1.00$ | A perfect uphill (positive) linear relationship |

Use the TI-83 or TI-84

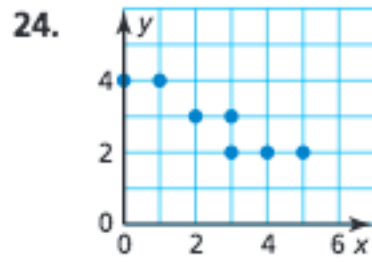
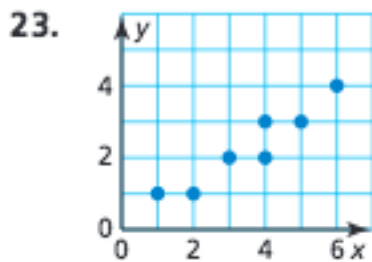
18. The table shows the numbers of tickets sold for a concert when different prices are charges. Write an equation of a line of fit for the data. Does it seem reasonable to use your model to predict the number of tickets sold when the ticket price is \$85? Explain.

Ticket price (dollars), x	17	20	22	26
Tickets sold, y	450	423	400	395

In Exercises 19-24, use the linear registration feature on a graphing calculator to find an equation of the line of best fit for the data. Find and interpret the correlation coefficient.



--	--	--	--



--	--