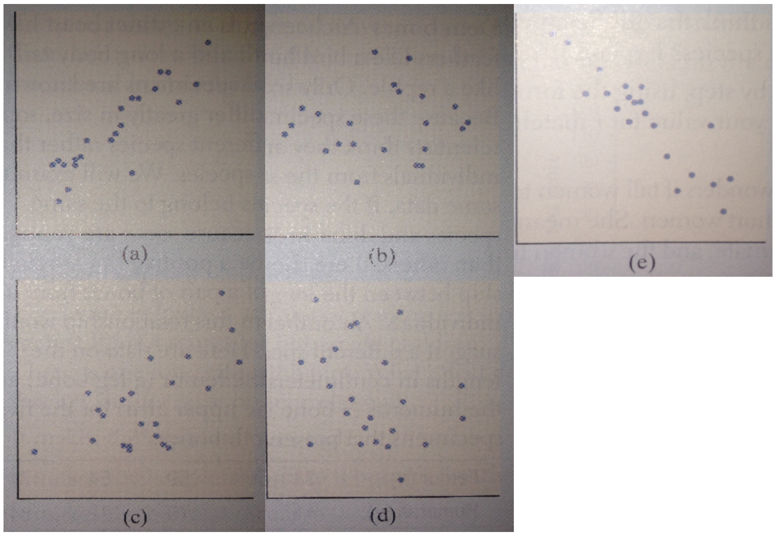
Section 1.5 Worksheet – Linear Regression Using Technology

MDM4U
Jensen

1) Match each of the following scatterplots to the r below that describes it. Then describe the direction and strength of the correlation. (Some r's will be left over)



$r = -0.9$ $\quad$ $r = -0.7$ $\quad$ $r = -0.3$

$r = 0$ $\quad$ $r = 0.3$ $\quad$ $r = 0.7$

$r = 0.9$

a) $r = 0.9$ **b)** $r = 0$ **c)** $r = 0.7$ **d)** $r = -0.3$ **e)** $r = -0.9$

2) Researchers studying acid rain measured the acidity of precipitation in a Colorado wilderness area for 150 consecutive weeks. Acidity is measured by pH. Lower pH values show higher acidity. The researchers observed a linear pattern over time. They reported that the regression line $\widehat{pH} = 5.43 - 0.0053(weeks)$ fit the data well.

a) Identify the slope of the line and explain what it means in this setting.

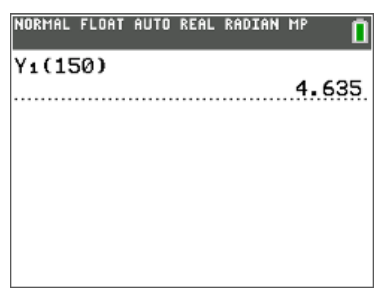
For every 1 unit increase in weeks, our model predicts a 0.0053 unit decrease in pH.

b) Identify the y-intercept of the line and explain what it means in this setting.

At 0 weeks, our model predicts an pH of 5.43

c) According to the regression line, what was the pH at the end of this study?

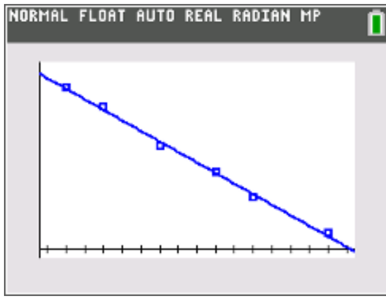
$\widehat{pH} = 5.43 - 0.0053(150) = 4.635$ *At the end of the study, the model predicts a pH of 4.635*

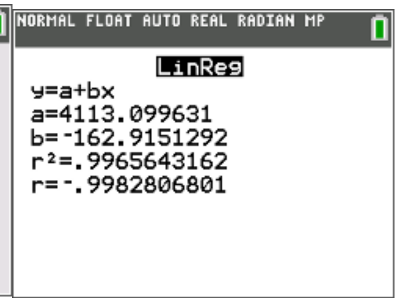


3) Market research has provided the following data on the monthly sales of a licensed T-shirt for a popular rock band.

Price ($)	Number of Shirts Sold
10	2500
12	2200
15	1600
18	1200
20	800
24	250

a) Make a scatterplot of the data.





b) Find the equation of the regression line and interpret the slope and y-intercept in context.

$\widehat{shirts\ sold} = 4113.1 - 162.9(price)$

The slope of -162.9 tells us that for every $1 increase in price, our model predicts a decrease of 162.9 shirts sold.

The y-intercept of 4113.1 tells us that at a price of $0, our model predicts 4113.1 shirts would be sold.

c) Find and interpret correlation coefficient, *r*.

$r = -0.998$*; this tells us there is a strong negative linear correlation between price and shirt sales.*

d) Find the coefficient of determination, r^2. Interpret it in the context of this data.

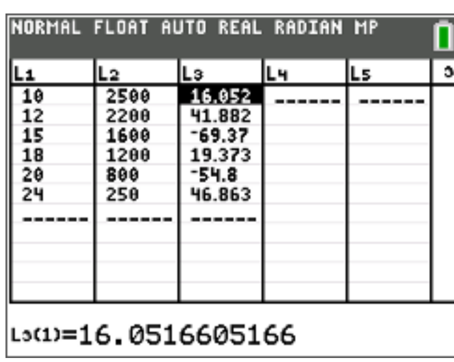
$r^2 = 0.9966$*; this tells us that about 99.66% of the variation in in shirts sold can be explained by the approximate linear relationship with price.*

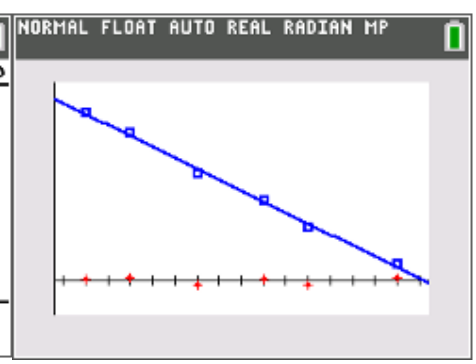
e) Predict the sales if the shirts are priced at $19.

$\widehat{shirts\ sold} = 4113.1 - 162.9(19) = 1018$*; our model predicts that 1018 shirts would be sold*

f) Calculate the residual values, record them and analyze them using the residual plot to help. Is a linear model a good fit?

There is no distinguishable pattern in the residual plot and the residual values are relatively small. This indicates that a linear model is an appropriate model for the variables and there is a strong correlation between them.

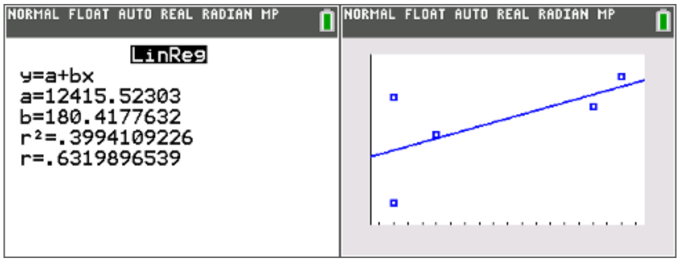




4) Average home attendance and number of home wins for the 2009 – 2010 NBA Pacific Division teams were as follows:

	Lakers	**Suns**	**Clippers**	**Warriors**	**Kings**
Home Wins, x	34	32	21	18	18
Average Attendance, y	18 997	17 648	16 343	18 027	13 254

a) Make a scatterplot of the data.



b) Find the equation of the regression line and interpret the slope and y-intercept in context.

$\widehat{attendance} = 12415.5 + 180.4(home\ wins)$

The slope of 180.4 tells us that for every 1 more win, our model predicts a 180.4 person increase in attendance.

The y-intercept of 12 415.5 tells us that with 0 home wins, our model predicts an attendance of 12 415.5 people.

c) Find and interpret correlation coefficient, *r*.

$r = 0.63$; *this tells us there is a moderate, positive, linear correlation between home wins and attendance.*

d) Find the coefficient of determination, r^2. Interpret it in the context of this data.

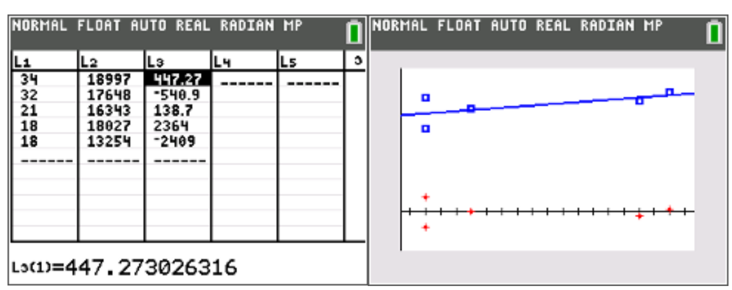
$r^2 = 0.3994$; *this tells us that about 39.94% of the variation in attendance can be explained by the approximate linear relationship with home wins.*

e) Predict the average attendance for a team with 25 home wins.

$\widehat{attendance} = 12415.5 + 180.4(25) = 16\,925.5$; *Our model predicts an attendance of about 16 926.*

f) Calculate the residual values, record them and analyze them using the residual plot to help. Is a linear model a good fit?

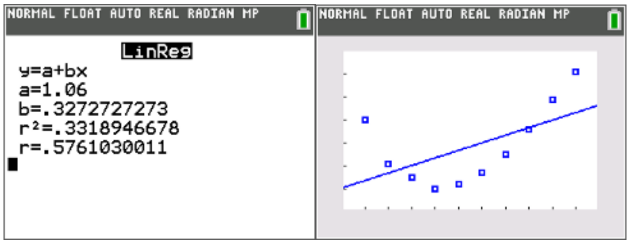
There is no distinguishable pattern in the residual plot. This tells us that the linear regression is a good model for the data.



5) Suppose the drying time of a paint product varies depending on the amount of a certain additive it contains.

Additive (oz), *x*	1	2	3	4	5	6	7	8	9	10
Drying Time (hr), *y*	4	2.1	1.5	1	1.2	1.7	2.5	3.6	4.9	6.1

a) Make a scatterplot of the data.



b) Find the equation of the regression line and interpret the slope and y-intercept in context.

$\widehat{drying\ time} = 1.06 + 0.327(additive)$

The slope of 0.327 tells us that for every 1 ounce increase in additive, our model predicts a 0.327 hour increase in drying time.

The y-intercept of 1.06 tells us that with 0 ounces of additive, our model predicts a drying time of 1.06 hours.

c) Find and interpret correlation coefficient, *r*.

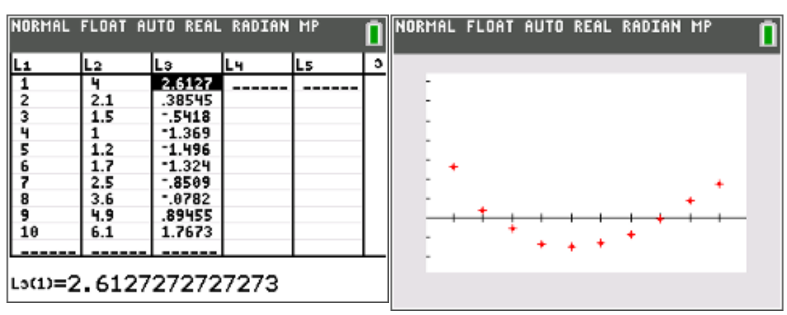
$r = 0.576$*; this tells us there is a moderate, positive, linear correlation between drying time and amount of additive.*

d) Find the coefficient of determination, r^2. Interpret it in the context of this data.

$r^2 = 0.3319$*; this tells us that about 33.19% of the variation in drying time can be explained by the approximate linear relationship with amount of additive.*

e) Calculate the residual values, record them and analyze them using the residual plot to help.

The pattern in the residual plot indicates that the linear regression is not a good model for the relationship between drying time and amount of additive.



6) Sketch the residual plot for the following scatterplot. Explain what it shows about the linear model.