FST 2-2 Notes
TOPIC: Linear Models
GOAL
Discuss the idea of a line of fit and the use of the sum of squared residuals as a measure of that fit.

SPUR Objectives

B Compute residuals from observed and predicted values.

F Find and interpret linear models.

I Use scatterplots and residual plots to draw conclusions about models for data.

Vocabulary

linear function
linear model
interpolation
extrapolation
observed values
predicted values
residual
sum of squared residuals

SLOPE INTERCEPT FORM → **Linear function** – is a set of ordered pairs (x, y) satisfying an equation of the form $y = mx + b$ where the slope is m and the y-intercept is b.

+m   −m

m = Slope = RATE OF CHANGE ; $\frac{\text{RISE}}{\text{RUN}}$; $\frac{y_2 - y_1}{x_2 - x_1}$

y-intercept = POINT WHERE LINE CROSSES Y-AXES

Point-slope form = $y = y_1 + m(x - x_1)$

INTERPOLATION is predicting values between observed data.

EXTRAPOLATION is predicting values outside the range of observed data. It depends on an assumption that a relationship will continue past the known data.

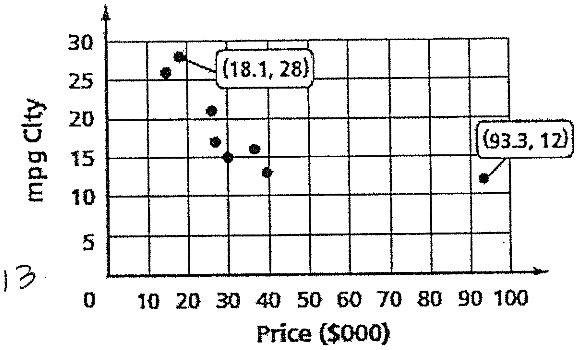
OBSERVED VALUES are data collected from sources such as experiments or surveys.

PREDICTED VALUES are values predicted by a model

Residual = OBSERVED value minus PREDICTED value.

2-2 p. 2

1) At the right is a scatterplot of the prices of selected 2008 vehicles and their estimated city mpg.



a) Use the identified data points to find an equation for a line to fit the data.

$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{12 - 28}{93.3 - 18.1} = \frac{-16}{75.2} = -.213$

$y = y_1 + m(x - x_1)$

$y = 28 - 0.213(x - 18.1)$

$y = 28 - 0.213x + 3.8553$

$\boxed{y = -0.213x + 31.8553}$

b) What does a negative slope of the line mean in this context?

MPG ↓ BY 0.213 FOR EVERY \$1 ↑ IN PRICE

c) If a vehicle cost \$32,000 how many city MPGs are expected? (interpolation)

$y = -0.213(32) + 31.8553$

$y = 25.0393$

$\boxed{25 \text{ MPG}}$

d) If a vehicle cost \$95,000 how many city MPGs are expected? (extrapolation)

$y = -0.213(95) + 31.8553$

$y = 11.6203$

$\boxed{12 \text{ MPG}}$

e) What is the expected price of a vehicle that gets 19 mpg?

$19 = -0.213x + 31.8553$

$-31.8553 \qquad -31.8553$

$\frac{-12.8553}{-0.213} = \frac{-0.213x}{-0.213}$

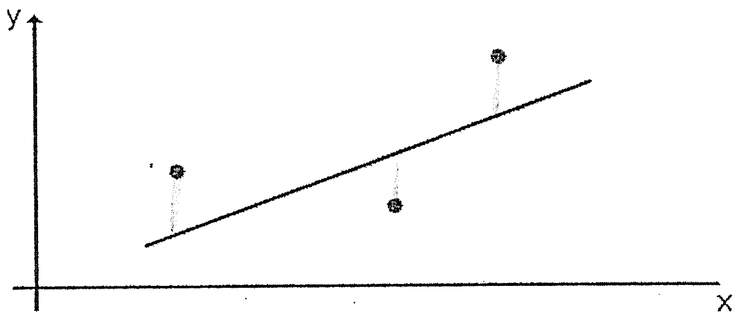
$x = 60.354$

$\boxed{\$60,354}$

2-2 p.3

How well did our ~~model~~ EQUATION fit the data?
Is it the best model?

Residuals - OBSERVED VALUE (DATA POINTS) − PREDICTED VALUES / LINE



data points
(observed values)

model of data - LINE

☆ vertical distance between data point and model

☆ used to determine the line of best fit

Residuals are (+) if observed value is above line

Residuals are (-) if observed value is below line

RESIDUALS OF ZERO ARE ON THE LINE

f) Explain why the value of the residual for (93.3, 12) [OBSERVED] is equal to the value of the residual for (18.1, 28). ★ USE EQN. TO FIND PREDICTED VALUES

OBSERVED

$y = -0.213x + 31.8553$
$= -0.213(93.3) + 31.8553$
$\doteq 12$ (predicted) ★ Little under 12 due to rounding

$R =$ obs − Pred
$= 12 - 12 = 0$

$\boxed{R=0}$

$y = -0.213(18.1) + 31.8553$
$y = 28$ (predicted)
$R =$ obs − Pred
$= 28 - 28 = 0$

$\boxed{R=0}$

★ THESE POINTS ARE ON THE LINE, SO BOTH ARE R=0.

g) One of the points on the scatterplot is (26.0, 21) [OBSERVED]. Calculate its residual.

$y = -0.213(26.0) + 31.8553$
$y = 26.3173$ (predicted)
$R =$ obs − pred
$= 21 - 26.3173$

$\boxed{R = -5.3173}$

★ Observed value is below the line.

2-2 p. 4

2) A diamond speculator used the line with equation $y = 2400x + 400$ to estimate the price of diamond rings.

Diamond Ring Prices by Weight of Diamond

Weight	Price (U.S. dollars)
0.15	484.50
0.16	507.00
0.18	702.00
0.25	963.00
0.27	1080.00
0.33	1417.50
0.23	829.50

a) What would the speculator predict for the price of the 0.25-carat diamond ring.

$y = 2400(.25) + 400$

$y = \$1000$ (predicted)

b) What is the residual for the 0.25-carat diamond ring.

$R = \text{obs} - \text{pred.}$

$= 963 - 1000$

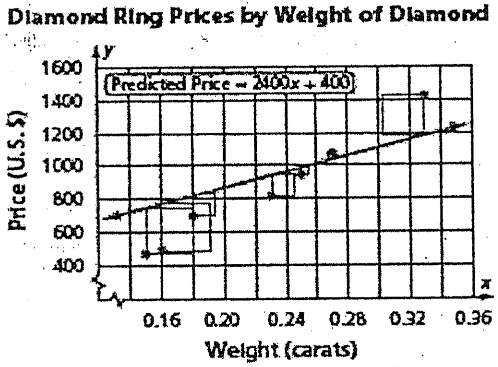
$\boxed{R = -37}$

Actual cost is \$37 less than the predicted cost

Linear Model 1

Squares are shown for a line that does not go through any data points.

Diamond Ring Prices by Weight of Diamond

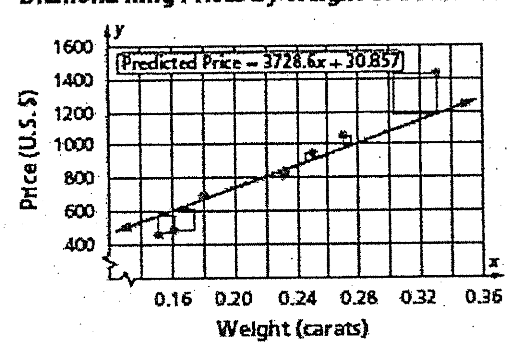


Total area of the squares ≈ 237,800

Linear Model 2

Squares are shown for a line through two of the data points.

Diamond Ring Prices by Weight of Diamond



Total area of the squares ≈ 59,870

The second line is a better model of the data because it has a smaller total area of the squares. The total area is the **sum of squared residuals.**

Definition of Sum of Squared Residuals

$$\textbf{Sum of squared residuals} = \sum_{i=1}^{n} (\text{observed } y_i - \text{predicted } y_i)^2$$

WE ARE MEASURING THE VARIATION BETWEEN THE DATA AND THE LINE. THE BETTER THE EQN, THE SMALLER THE SUM OF THE SQUARED RESIDUALS.

2-2 p. 5

3)

A situation is modeled by the equation $f(x) = 4.2x - 5.7$. Residuals for certain values for x are given at the right. What are the observed values?

	x	Residual
1)	1.0	–0.4
2)	2.5	1.2
3)	9.0	3.0

☆ LINEAR MODEL (EQUATION) IS USED TO CALCULATE PREDICTED VALUES.

$\boxed{R = \text{Obs} - \text{Pred}}$

① $4.2(1.0) - 5.7 = -1.5$ (PREDICTED)

$-0.4 = y + +1.5$

$-1.5 \qquad -1.5$

$\boxed{y = -1.9}$ observed

② $4.2(2.5) - 5.7 = 4.8$ (PREDICTED)

$1.2 = y - 4.8$

$+4.8 \qquad +4.8$

$\boxed{6.0 = y}$ Observed

③ $4.2(9.0) - 5.7 = 32.1$ (PREDICTED)

$3.0 = y - 32.1$

$+32.1 \qquad +32.1$

$\boxed{35.1 = y}$ Observed

FST 2-2 Additional Notes

Diamond Prices by Weight

Weight	Price $
0.18	702.00
0.25	963.00
0.27	1080.00
0.33	1417.50

Observed

1) The equation for the scatterplot of these data is $y = 2400x + 400$. → Use this to predict

a) What is the residual for a 0.18-carat diamond?

R = Obs − Pred

R = 702 − 832

$\boxed{R = -130}$

$y = 2400(.18) + 400$

$y = \$832$ predicted

Diamond is $130 below predicted price.

b) What is the squared residual for a 0.27-carat diamond?

R = obs − Pred

R = 1080 − 1048

R = 32

$y = 2400(.27) + 400$

$y = \$1048$ predicted

Diamond is $32 above predicted price.

$\boxed{R^2 = (32)^2 = 1024}$

c) Calculate the sum of squared residuals for the diamonds. ΣR^2

Observed	Predicted $y = 2400x + 400$	$\Sigma(\text{Obs} - \text{Pred})^2$
(0.18, 702)	$y = 2400(.18) + 400 = \$832$	$(702 - 832)^2 = 16,900$
(0.25, 963)	$y = 2400(.25) + 400 = \$1000$	$(963 - 1000)^2 = 1369$
(0.27, 1080)	$y = 2400(.27) + 400 = \$1048$	$(1080 - 1048)^2 = 1024$
(0.33, 1417.50)	$y = 2400(.33) + 400 = \$1192$	$(1417.50 - 1192)^2 = 50,850.25$

$\boxed{\text{Sum} = 70,143.25}$

* d + e next page

d) The diamond speculator uses the equation to predict the price of 0.40-carat diamond. Is this interpolation or extrapolation?

0.40 carat is outside the weight values in the data.

e) What is the diamond speculator prediction from part d?

$y = 2400(.40) + 400 = \boxed{\$1360}$

FST 2-3 Notes
TOPIC: Linear Regression & Correlation
GOAL
Discuss data which, when graphed, shows a roughly linear pattern of growth. Explain how to use technology to find an equation for the line of best fit and to determine the closeness of fit, as measured by the linear correlation coefficient.

SPUR Objectives

D Identify properties of regression lines and of the correlation coefficient.

F Find and interpret linear regression and models.

I Use scatterplots and residual plots to draw conclusions about linear models for data.

Vocabulary

method of least squares
line of best fit, least squares line, regression line
center of mass
correlation coefficient
perfect correlation
strong correlation
weak correlation

Linear Regression

Refers to finding the LINE OF BEST FIT by using the method of least squares.
- LEAST SQUARES LINE
- REGRESSION LINE

Properties

- Only 1 line of best fit for data set
- Contains the center of mass of the data $(\bar{x}, \bar{y})$ whose coordinates are the mean of the x-values and the mean of the y-values
- Slope & y-intercept computed from the data points