

FST 2-2 Notes

TOPIC: Linear Models

GOAL (EQUATIONS)

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

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 .

Slope = $\frac{\text{rise}}{\text{run}}$, Steepness of line, rate of change, $\frac{y_2 - y_1}{x_2 - x_1}$

y-intercept = point where line crosses y-axis

Point-slope form =

$y = y_1 + m(x - x_1)$ where (x_1, y_1) is a point on line

interpolation is predicting ^{values} between observed ^{data} values.

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 ^{equation}

*** Residual = observed value minus predicted value.

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.

$$\text{Slope} = \frac{y_2 - y_1}{x_2 - x_1} = \frac{12 - 28}{93.3 - 18.1} = \frac{-16}{75.2} = -0.213$$

$$y = y_1 + m(x - x_1) \quad \text{pt } (18.1, 28)$$

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

$$y = 28 + -0.213x + 3.8553$$

$$y = -0.213x + 31.855$$

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

Estimated city MPG's for a vehicle ↓
by 0.213 for every \$1 ↑ in price of car

c) If a vehicle cost \$32,000 how many city MPG's are expected? (interpolation) $32,000 \rightarrow 32$

$$y = -0.213x + 31.855$$

$$y = -0.213(32) + 31.855$$

$$y = 25.039 \rightarrow \boxed{25 \text{ mpg}}$$

d) If a vehicle cost \$95,000 how many city MPG's are expected? (extrapolation) $95,000 \rightarrow 95$

$$y = -0.213x + 31.855$$

$$y = -0.213(95) + 31.855$$

$$y = 11.62 \rightarrow \boxed{12 \text{ mpg}}$$

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

$$y = -0.213x + 31.855$$

$$19 = -0.213x + 31.855$$

$$-31.855 \quad -31.855$$

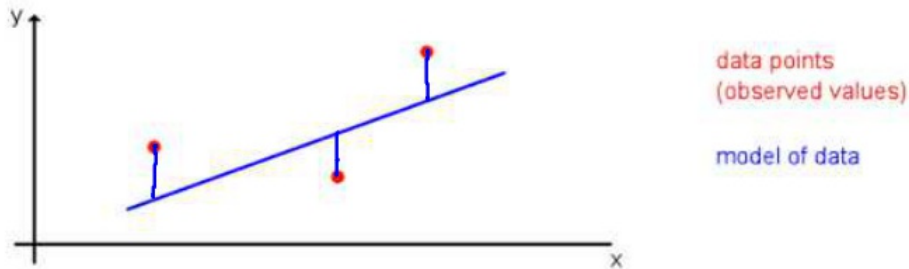
$$\frac{-12.855}{-0.213} = \frac{-0.213x}{-0.213}$$

$$60.352 = x$$

$$\boxed{\$60,352}$$

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

Residuals - Observed values minus predicted values



★ 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 are zero if observed value is on line

observed
(93.3, 12)

f) Explain why the value of the residual for (93.3, 12) is equal to the value of the residual for (18.1, 28).

$$y = -0.213x + 31.855$$

$$y = -0.213(93.3) + 31.855$$

$$y = 12 \text{ (predicted)}$$

$$\text{Resid} = \text{obs} - \text{pred}$$

$$12 - 12 = \boxed{0}$$

$$y = -0.213x + 31.855$$

$$y = -0.213(18.1) + 31.855$$

$$y = 28 \text{ (predicted)}$$

$$\text{Resid} = 28 - 28 = \boxed{0}$$

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

$$y = -0.213x + 31.855$$

$$y = -0.213(26.0) + 31.855$$

$$y = 26.3 \text{ (predicted)}$$

$$\text{Resid} = \text{obs} - \text{pred} = \boxed{-5.3}$$

$$21 - 26.3 = \boxed{-5.3}$$

observed value is below the line

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

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

Observed data

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

Use Model (Equation) to predict

$$y = 2400x + 400$$

$$y = 2400(.25) + 400$$

$$y = \$1000 \text{ (predicted value)}$$

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

$$\text{Residual} = \text{observed} - \text{predicted}$$

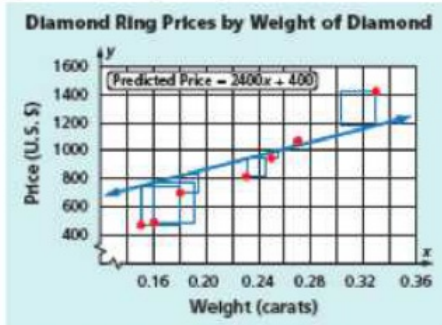
$$= 963 - 1000$$

$$= -37$$

Actual price was \$37 less than predicted

Linear Model 1

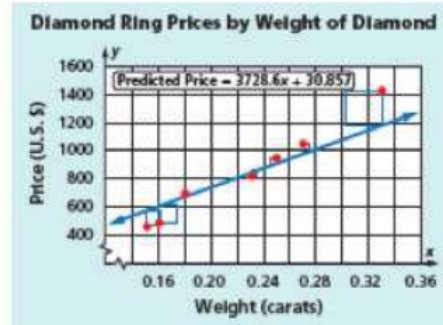
Squares are shown for a line that does not go through any data points.



Total area of the squares $\approx 237,800$

Linear Model 2

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



Total area of the squares $\approx 59,870$

✶✶ 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

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

We are measuring the variation between the data and the line. The better the model (Equation) the smaller the sum of squared residuals

3) model Equation is used to calculate predicted value

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
1.0	-0.4
2.5	1.2
9.0	3.0

$$\begin{aligned} \text{Pred} &= 4.2(1.0) - 5.7 \\ &= -1.5 \end{aligned}$$

Residual = Observed - predicted

$$-0.4 = X - (-1.5)$$

$$-0.4 = X + 1.5$$

$$\begin{array}{r} -1.5 \\ \hline -1.5 \end{array}$$

Observed $(-1.9) = X$

$$\begin{aligned} \text{Pred} &= 4.2(2.5) - 5.7 \\ &= 4.8 \end{aligned}$$

$$R = O - P$$

$$1.2 = X - 4.8$$

$$\begin{array}{r} +4.8 \\ \hline +4.8 \end{array}$$

Observed $(6.0) = X$

$$\begin{aligned} \text{Pred} &= 4.2(9.0) - 5.7 \\ &= 32.1 \end{aligned}$$

Resid = Obs - Pred

$$3.0 = X - 32.1$$

$$\begin{array}{r} +32.1 \\ \hline +32.1 \end{array}$$

Observed $(35.1) = X$

FST 2-2 Additional Notes

1) The equation for the scatterplot of these data is $y = 2400x + 400$.

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

$$y = 2400(.18) + 400$$

$$y = 832 \text{ (predicted)}$$

Resid = Obs - Pred
 $702 - 832$
 $R = -130$

observed values
Diamond Prices by Weight

Weight	Price \$
0.18	702.00
0.25	963.00
0.27	1080.00
0.33	1417.50

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

$$y = 2400(.27) + 400$$

$$y = 1048 \text{ (predicted)}$$

$$R = O - P$$

$$R = 1080 - 1048$$

$$R = 32$$

$$R^2 = 32^2 = 1024$$

c) Calculate the sum of squared residuals for the diamonds.

$$y = 2400(.18) + 400$$

$$832$$

$$y = 2400(.25) + 400$$

$$1000$$

$$y = 2400(.27) + 400$$

$$1048$$

$$y = 2400(.33) + 400$$

$$1192$$

$$\text{Obs} - \text{Pred} = \text{Resid} = R^2$$

$$702 - 832 = -130 = 16,900$$

$$963 - 1000 = -37 = 1369$$

$$1080 - 1048 = 32 = 1024$$

$$1417.50 - 1192 = 225.5 = 50,850.25$$

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

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

Extrapolation .40 is outside the range (.18 - .33)

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

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