Next section

3.2 four days

Mrs. Wierhman
tomorrow

Section 3.2
Learning Targets

- Make predictions using regression lines, keeping in mind the dangers of extrapolation.
- Calculate and interpret a residual.
- Interpret the slope and *y*-intercept of a least-squares regression line.

Experience first
|
Formalize later

Start by working on

AP Stats Class Notes Section 3.2 Day 1

How good are the predictions for Mickey M?

1 - 4

A class performed the "Mickey Mouse Bungee" activity. In this activity, students made a chain of rubber bands, connecting them one at a time to Mickey's feet and then measuring the distance that Mickey travels on his bungee jump. The distance is measured from the edge of the jumping platform to the lowest point that Mickey's body reaches.

Here is the data from one of the groups. The group forgot to record their measurement for 5 rubber bands.

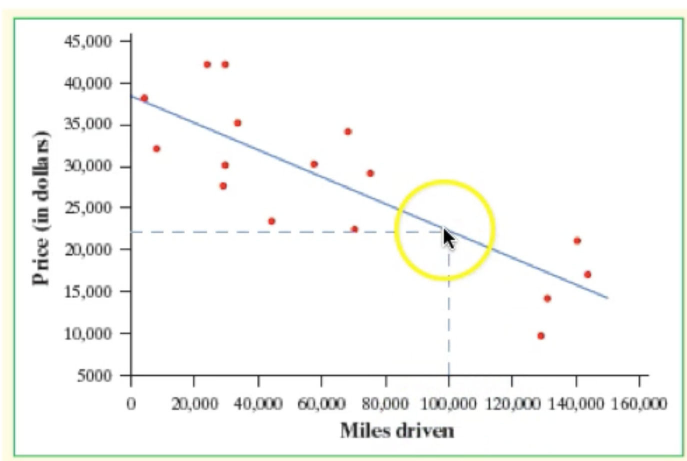
Number of rubber bands	0	1	2	3	4	5	6	7
Distance Mickey traveled (cm)	25	32	41	49	55	?	69	78

1. **Use your Graphing Calculator to make a scatterplot. Just look at it but you don't need to sketch it. Then Calculate "Least-squares regression line"** *using 8: LinReg(a+bx)*. **This is the line that best models the data. Write the equation below.**

Your answer: Lucky

AP answer: individuals to write their answers

- We use regression lines to predict the value of a response variable for a particular value of the explanatory variable



1. **Use your Graphing Calculator to make a scatterplot. Just look at it but you don't need to sketch it. Then Calculate "Least-squares regression line"** *using 8: LinReg(a+bx)*. **This is the line that best models the data. Write the equation below.**

Your answer:

$y = 7.46x + 25.33$

AP answer:

"y-hat" → $\hat{y} = 25.333 + 7.464x$

means predicted y

$\underbrace{\widehat{\text{Distance}} = 25.333 + 7.464\,(\text{Rubber bands})}_{\text{Use context when writing out regression equation}}$

Why do statisticians prefer ?

A **regression line** is a line that describes how a response variable *y* changes as an explanatory variable *x* changes. Regression lines are expressed in the form $\hat{y} = b_0 + b_1 x$ where $\hat{y}$ (pronounced "y-hat") is the predicted value of *y* for a given value of *x*.

$$\hat{y} = a + bx$$

y-intercept

Real World

There are often more than one explanatory variables that can help predict the response variable. (x_1, x_2, x_3, etc)

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3$$

the y-intercept is the starting point for making a prediction

process called multiple regression

textbook $\hat{y} = b_0 + b_1 x$

New in 2019 $\longrightarrow$ $\hat{y} = a + bx$

f September 30, 2019

2. Use the regression line to predict the distance Mickey travels for 5 rubber bands. Show work.

Your work: $7.46(5) + 25.33 = 62.6$

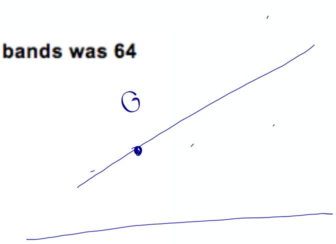
AP format:

$$\widehat{\text{distance}} = 25.333 + 7.464(5)$$

$$= 62.653 \text{ cm}$$

$$62.7 \text{ cm}$$

3. One of the group members later found the measurement for 5 rubber bands was 64 cm. Was the prediction from #2 too high or too low? How far off?



Your work: to high 1.5 off

AP format:

Residual (error) = Actual (data) − Predicted (LSRL)

$64 - 62.653 = 1.347$

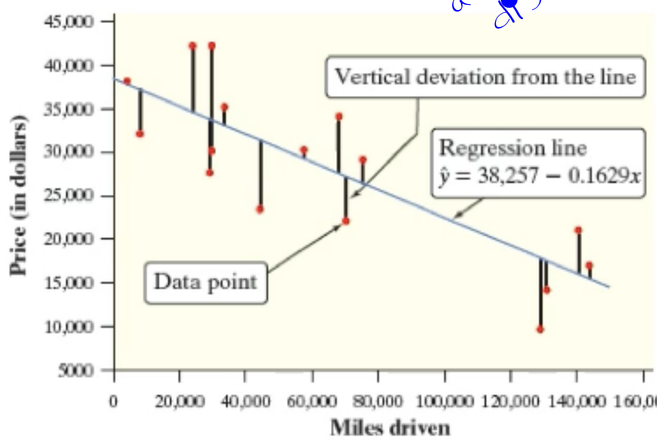
AP!!!

The predicted distance was 1.347 cm too low.

a vertical distance

A residual is the difference between the actual value of *y* and the predicted value of *y* for a particular value of *x*.

If a residual is positive, the actual value is greater than the predicted value. If it is negative, the actual value is less than the predicted value.



4. **Predict the distance that Barbie would travel if the group used 20 rubber bands. Would you trust this prediction more or less than the prediction you made in #2?**

Your work: 175

AP format:

$\widehat{distance} = 25.333 + 7.464(20) = 174.613$ cm

We would trust this prediction less because it is an extrapolation

Extrapolation is the use of a regression line for prediction far outside the interval of *x* values used to obtain the line. Such predictions are often not accurate.

Time to go back and formalize

Now stop and wait ! $\hat{y} = 25.333 + 7.464x$

5. What is the y-intercept of the equation of the regression line? What does it mean?

25.333 cm

6. What is the slope of the equation of the regression line? What does it mean?

Now stop and wait !

5. **What is the y-intercept of the equation of the regression line? What does it mean?**

(0, y-int)

When we use 0 rubber bands the predicted distance travelled is 25.333 cm.

6. **What is the slope of the equation of the regression line? What does it mean?**

$\text{slope} = \frac{\Delta y}{\Delta x} = \frac{\text{change in } y}{\text{change in } x}$

7.464

The predicted distance goes up by 7.464 cm for each additional rubber band

Big Ideas:

Big Ideas:

$\hat{y} = a + b_0 x$

y-int

slope

careful of extrapolation

Residuals

Resid = Actual − Pred

R = A − P

The actual y-context was resid higher/lower than predicted for x=

y-int. when x=0 context the predicted y-context is y-int

slope

Check Your Understanding:

1. Some data were collected on the weight of a male white laboratory rat for the first 25 weeks after its birth. A scatterplot of *y* = weight (in grams) and *x* = time since birth (in weeks) shows a fairly strong, positive linear relationship. The regression equation $\hat{y} = 100 + 40x$ models the data fairly well.

Big Ideas:

$\hat{y} = a + b_0 x$

y-int

slope

careful of extrapolation

Residuals

Resid = Actual − Pred

R = A − P

The actual y-context was resid higher/lower than predicted for x=

y-int. when x=0 context the predicted y-context is y-int

slope the predicted y-context goes up/down for each increase of x-context

Check Your Understanding:

1. Some data were collected on the weight of a male white laboratory rat for the first 25 weeks after its birth. A scatterplot of *y* = weight (in grams) and *x* = time since birth (in weeks) shows a fairly strong, positive linear relationship. The regression equation $\hat{y} = 100 + 40x$ models the data fairly well.

Check for Understanding

try to be more precise
with notation and language
the the experience portion

1. Some data were collected on the weight of a male white laboratory rat for the first 25 weeks after its birth. A scatterplot of y = weight (in grams) and x = time since birth (in weeks) shows a fairly strong, positive linear relationship. The regression equation $\hat{y} = 100 + 40x$ models the data fairly well.

a. Interpret the slope of the regression line.

The predicted weight of the rat increases by 40 grams
every week since birth

b. Does the value of the y intercept have meaning in this context? If so, interpret the y intercept. If not, explain why.

when time since birth is equal to zero, the predicted
weight is 100g.

predicted birth weight

1. Some data were collected on the weight of a male white laboratory rat for the first 25 weeks after its birth. A scatterplot of y = weight (in grams) and x = time since birth (in weeks) shows a fairly strong, positive linear relationship. The regression equation $\hat{y} = 100 + 40x$ models the data fairly well.

a. Interpret the slope of the regression line.

The predicted weight goes up by 40 grams for each increase of one week.

b. Does the value of the y intercept have meaning in this context? If so, interpret the y intercept. If not, explain why.

Yes. When a rat is 0 weeks old, the predicted weight is 100 grams.

i.e... birth weight!

c. Predict the rat's weight at 16 weeks old.

d. Calculate and interpret the residual if the rat weighed 700 grams at 16 weeks old

c. Predict the rat's weight at 16 weeks old.

$\widehat{\text{Weight}} = 100 + 40(16)$
$= 740$ grams

d. Calculate and interpret the residual if the rat weighed 700 grams at 16 weeks old

Residual $= 700 - 740 = -40$ grams

The actual weight is 40 grams lower than predicted at 16 weeks old.

e. Should you use this line to predict the rat's weight at 2 years old? Use the equation to make the prediction and discuss your confidence in the result. (There are 454 grams in a pound.)

Nope. There are 104 weeks in 2 years. and our data is only from the first 25 weeks. This is an extrapolation.

CAUTION:

Don't make predictions using values of *x* that are much larger or much smaller than those that actually appear in your data.

LCQ

Assignment

3.2 ....37, 39, 41, 43, 45

I encourage you to read/study pp. 176-182