Homework Answers:

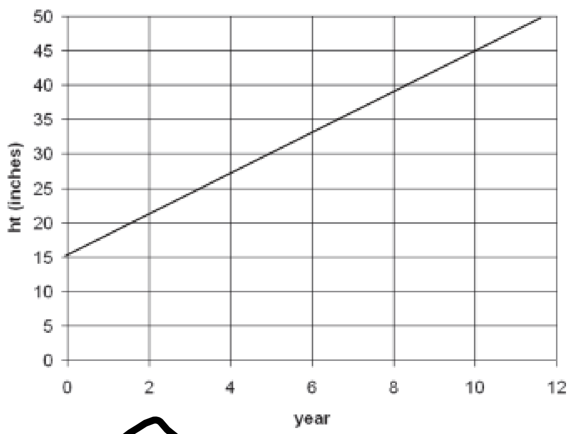
Statistics Chapter 7: Slopes and Intercepts – KEY

Use the following slope and y-intercept clues to graph and find equations.

context: $\left(\overset{x}{years\ since\ planting}, \overset{y}{height}\right)$ of a tree

***y*-intercept:** The tree was 15 inches tall when planted (year 0)

slope: The tree grew 3 inches per year (3 inches/yr)

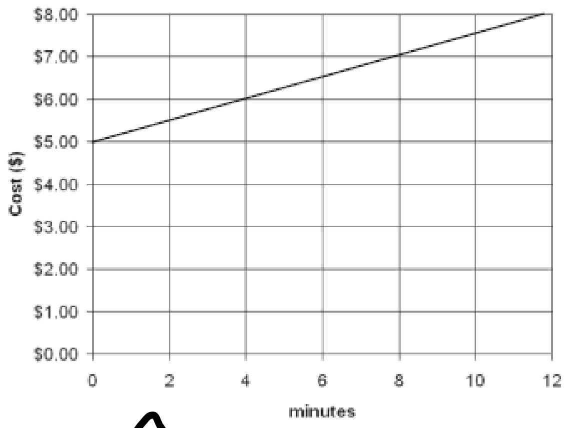


equation: $\widehat{height} = 15 + 3year$

context: $\left(\overset{x}{minutes}, \overset{y}{cost}\right)$ for using the wi-fi service at a local coffee shop.

***y*-intercept:** The cost for setting up the service is \$5 (minute 0)

slope: The cost per minute of use is \$0.25 (\$0.25/min)



equation: $\widehat{cost} = 5 + 0.25minute$

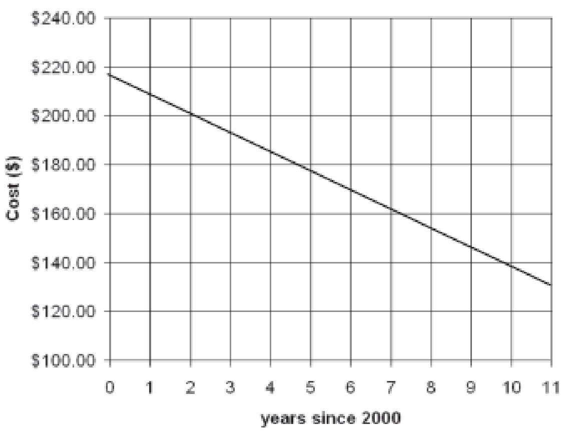
1) at year 0, the tree is 15" tall.
as the year goes up by 1, the tree height ↑ by 3"

2) as the # of minutes ↑ by 1, the cost ↑ by \$0.25

context: $(year,\ cost\ of\ a\ 19\ inch\ TV)$

y-intercept: In 2000, the average price of a 19 inch TV was \$219.

slope: The cost changed by –\$8/yr.



equation: $\widehat{cost} = 219 - 8year$

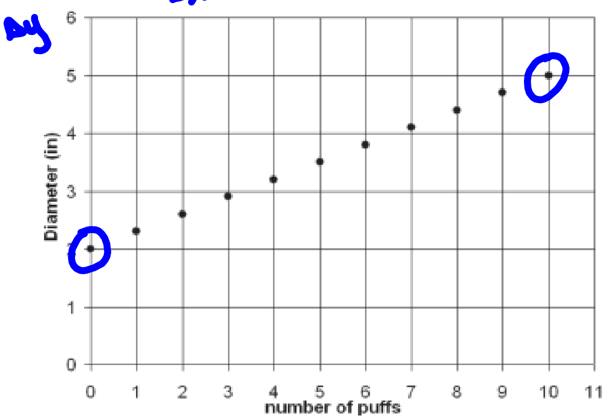
context: $(number\ of\ puffs,\ diameter)$ when blowing up a balloon

y-intercept: The balloon is 2 inches in diameter before any air is puffed into it.

slope: Every 10 puffs forces the diameter to increase by 3 inches

Δx

Δy



equation: $\widehat{diameter} = 2 + \frac{3}{10}puffs$

or .3

$\frac{5-2}{10-0}$

3) as the # of years after 2000 ↑ by 1, the average price of a 19" TV will ↓ by \$8

4)

- Last chapter we looked at data, generally linear in shape.
- In this chapter we will focus on making linear models from generally linear data.
- We will use the calculators to find the equations, and use these equations to make limited predictions.
- We will check the models for appropriateness using scatterplots.

Residuals

- The model won't be perfect, regardless of the line we draw.
- As George Box said "All models are wrong, but some models are useful."
- Some points will be above the line and some will be below.
- The estimate made from a model is the predicted value (denoted as $\hat{y}$).

Residuals (cont.)

- The difference between the observed (true) value and the line's predicted value is called the residual.
- To find the residuals, we always subtract the predicted value from the observed one:

comes from the model

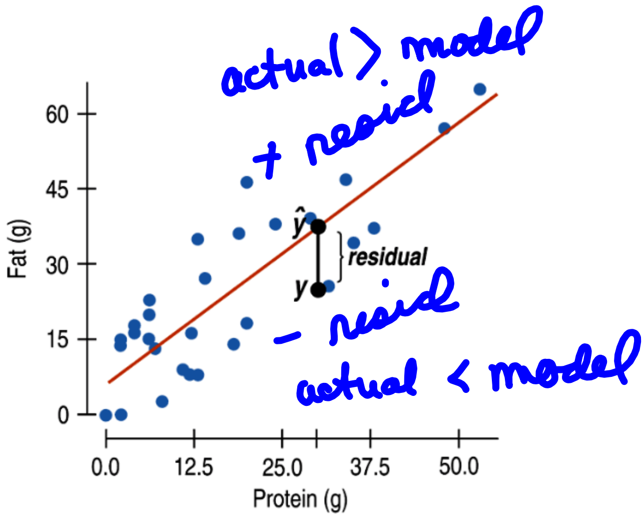
$$\boxed{* \ residual = observed - predicted = y - \hat{y}}$$

(more with these tomorrow...)

ALWAYS LEARNING PEARSON Chapter 7, Slide 8

Residuals (cont.)

- A negative residual means the predicted value's too big (an overestimate).
- A positive residual means the predicted value's too small (an underestimate).
- In the figure, the estimated fat of the BK Broiler chicken sandwich is 36 g, while the true value of fat is 25 g, so the residual is –11 g of fat.



"Best Fit" Means Least Squares

- Some residuals are positive, others are negative, and, on average, they cancel each other out.
- So, we can't assess how well the line fits by adding up all the residuals.
- Similar to what we did with deviations, we square the residuals and add the squares.
- The smaller the sum, the better the fit.
- The line of best fit is the line for which the sum of the squared residuals is smallest, we call it the least squares line.

"model"

Calculator does this for us

 Chapter 7, Slide 10

The Least Squares Line

- In our model, we have a slope (*b*):
  - The slope is built from the correlation and the standard deviations:

$$b = r\frac{s_y}{s_x}$$

  - Our slope is always in units of *y* per unit of *x*.

S_x and S_y are the standard deviations of the lists.
r is the linear correlation coefficient.

 Chapter 7, Slide 11

The Least Squares Line (cont.)

- In our model, we also have an intercept (*a*).
    - The intercept is built from the means and the slope:

$$a = \bar{y} - b\bar{x}$$

    - Our intercept is always in units of *y*.

$\bar{x}$ and $\bar{y}$ are the means of the two lists

Technology

- We almost always use technology to find the equation of the regression line.
- Your calculator has a Linear Regression tool.
- Computer software makes a table for regression.

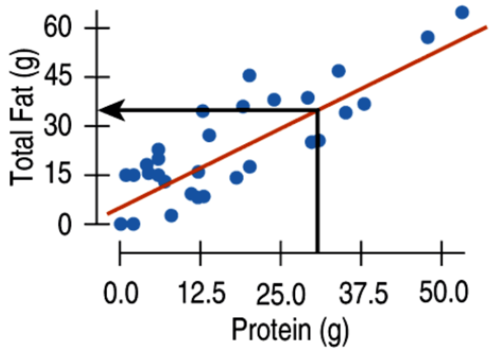
Stat-Calc-8:LinReg(a+bx) L1, L2

 Chapter 7, Slide 13

Fat Versus Protein: An Example

- The regression line for the Burger King data fits the data well:
  - The equation is

$$\widehat{fat} = 6.8 + 0.97\, protein.$$



The *predicted fat* content for a BK Broiler chicken sandwich (with 30 g of protein) is 6.8 + 0.97(30) = 35.9 grams of fat.

The Least Squares Line (cont.)

- Since regression and correlation are closely related, we need to check the same conditions for regressions as we did for correlations:
  - Quantitative Variables Condition
  - Straight Enough Condition
  - Outlier Condition

text example pg. 175-177

Step-by-Step Example REGRESSION

Even if you hit the fast-food joints for lunch, you should first have a good breakfast. Nutritionists, concerned about "empty calories" in breakfast cereals, recorded facts about 77 cereals, including their *Calories* per serving and *Sugar* content (in grams).

QUESTION: How are calories and sugar content related in breakfast cereals?

THINK ➡ Plan State the problem and determine the role of the variables.

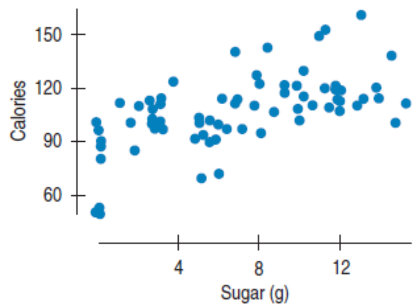
I am interested in the relationship between sugar content and calories in cereals. I'll use *Sugar* to estimate *Calories*.

Variables Name the variables and report the W's.

x y

✓ **Quantitative Variables Condition:** I have two quantitative variables, *Calories* and *Sugar* content per serving, measured on 77 breakfast cereals. The units of measurement are calories and grams of sugar, respectively.

Check the conditions for regression by making a picture. Never fit a regression line without looking at the scatterplot first.



✓ **Outlier Condition:** There are no obvious outliers or groups.

✓ The **Straight Enough Condition** is satisfied; I will fit a regression model to these data.

SHOW ➡ Mechanics If there are no clear violations of the conditions, fit a straight line model of the form $\hat{y} = a + bx$ to the data.

To use *Sugar* to estimate *Calories*, let x = *Sugar* and y = *Calories*

Calories

$\bar{y} = 107$ calories

$s_y = 19.5$ calories

Sugar

$\bar{x} = 7$ grams

$s_x = 4.4$ grams

Correlation

$r = 0.564$

Find the slope.

$$b = \frac{rs_y}{s_x} = \frac{0.564(19.5)}{4.4}$$

$= 2.50$ calories per gram of sugar.

Find the intercept.

$a = \bar{y} - b\bar{x} = 107 - 2.50(7) = 89.5$ calories.

So the least squares line is

Write the equation, using meaningful variable names.

$\hat{y} = 89.5 + 2.50\,x$ or

$\widehat{Calories} = 89.5 + 2.50$ Sugar.

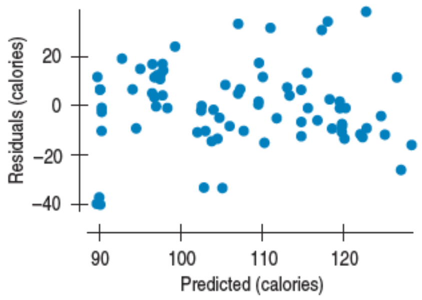
Stat-Calc 8

TELL ➡ Conclusion Describe what the model says in words and numbers. Be sure to use the names of the variables and their units.	The scatterplot shows a positive, linear relationship and no outliers. The slope of the least squares regression line suggests that cereals have about 2.50 more *Calories* per additional gram of *Sugar*.
The key to interpreting a regression model is to start with the phrase "*b* *y*-units per *x*-unit," substituting the estimated value of the slope for *b* and the names of the respective units. The intercept is then a starting or base value.	The intercept predicts that sugarless cereals would average about 89.5 calories.

Think AGAIN ➡ Check Again Even though we looked at the scatterplot *before* fitting a linear model, a plot of the residuals is essential to any regression analysis because it is the best check for additional patterns and interesting quirks in the data.

TI-*nspire*

Residuals plots. See how the residuals plot changes as you drag points around in a scatterplot.



The residuals show random scatter and a shapeless form so our linear model appears to be appropriate.

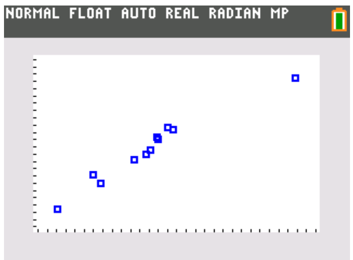
Classwork/Homework:

Calculating the Linear Model:

Step by Step Practice (Pkt Pg. 9)

Think:

1. How is CO amt. related to amt. of tar?
2. CO, Tar (both in mg)
3. Quantitative ✓
   Straight Enough? ✓
   Outliers? ✓



NORMAL FLOAT AUTO REAL RADIAN MP

LinReg
y=a+bx
a=3.831840265
b=0.7050179732
r²=0.955848465
r=0.9776750304

Show:

1.