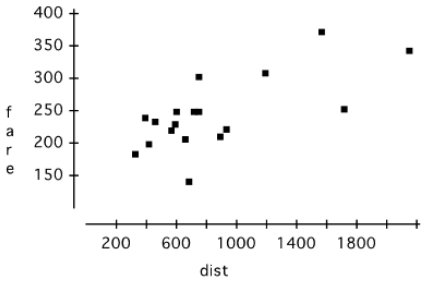
# Homework Answers:

***Statistics Chapter 7: Distance and Ticket Price – KEY***

This graph shows the relationship between distance (miles) and airfare ($) for flights out of Atlanta, GA..




L1 L2

| Atlanta to: | Distance | Fare |
|---|---|---|
| Baltimore | 568 | 219 |
| Boston | 933 | 222 |
| Dallas | 720 | 249 |
| Denver | 1190 | 308 |
| Detroit | 602 | 249 |
| Kansas City | 683 | 141 |
| Las Vegas | 1719 | 252 |
| Miami | 589 | 229 |
| Memphis | 327 | 183 |
| Minneapolis | 894 | 209 |
| New Orleans | 419 | 199 |
| NY | 749 | 248 |
| Okla City | 749 | 301 |
| Orlando | 392 | 238 |
| Philadelphia | 657 | 205 |
| St Louis | 461 | 232 |
| Salt Lake | 1565 | 371 |
| Seattle | 2150 | 343 |
| *Mean* | *853.7* | *244.3* |
| *St Dev* | *497.8* | *56.37* |
| *Correlation* | *0.694* | |

Stat - Calc - 8

1. Find the slope of the regression line. $\left(b = \frac{rs_x}{s_y}\right)$ $b = \frac{(0.694)(56.37)}{497.8} = 0.0786$

b = 0.0786

2. Find the $y$-intercept of the regression line. $(a = y - bx)$

a = 177.21

~~a = 343 - (0.0786)2150 = 174.07~~

3. Explain what the $y$-intercept means in this context.

177.21

For a flight that travels nowhere, we can expect the far to be ~~$174.07~~. This is not practical and so it just serves as the base for our model.

4. Explain what the slope means in this context.

For every mile traveled, we can expect the cost of the far to increase by about $0.08.

5. Write the equation of the linear model.

~~$\widehat{Fare} = 174.07 + 0.0786 Distance$~~

$\widehat{Fare} = 177.21 + .0786 Dist$

6. The fare to fly to Los Angeles, 1719 miles from Atlanta, is $212. Find the residual.

177.21 ~~$\widehat{Fare} = 174.07 + 0.0786(1719) = \$309.18$~~ $312.32. The residual is $212 – ~~$309.18~~ = ~~$97.18~~ above the actual price. Note that the residual is ~~-97.18~~. -100.32

7. Estimate the fare for a 200-mile flight.

177.21 ~~$\widehat{Fare} = 174.07 + 0.0786(200) = \$189.79$~~ 192.93

8. In general, a positive residual means…

the predicted fare is less than the actual fare. Likewise, the actual fare is greater than the predicted far.

9. Estimate the fare for a 2000-mile flight.

177.21 ~~$\widehat{Fare} = 174.07 + 0.0786(2000) = \$331.27$~~ 334.41

10. In general, a negative residual means…

the predicted fare is greater than the actual fare. Likewise, the actual value is less than the predicted value.

give ↓ find using ↓ eqn

Resid = actual − pred.

# More about Regression lines...

Remember:

- 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*.

## Correlation and the Line (cont.)

- Put generally, moving any number of standard deviations away from the mean in *x* moves us *r* times that number of standard deviations away from the mean in *y*.

# Example: Text pg. 172

Suppose you were told that a new male student was about to join the class, and you were asked to guess his height in inches. What would you guess? A safe guess would be the mean height of male students. Now suppose you are also told that this student has a grade point average (*GPA*) of 3.9—about 2 SDs above the mean *GPA*. Would that change your guess? Probably not. The correlation between *GPA* and *height* is near 0, so knowing the *GPA* value doesn't tell you anything and doesn't move your guess. (And the equation tells us that as well, since it says that we should move $0 \times 2$ SDs from the mean.)

On the other hand, suppose you were told that, measured in centimeters, the student's height was 2 SDs above the mean. There's a perfect correlation between *height in inches* and *height in centimeters,* so you'd know he's 2 SDs above mean height in inches as well. (The equation would tell us to move $1.0 \times 2$ SDs from the mean.)

Pg. 29

What if you're told that the student is 2 SDs above the mean in *shoe size*? Now you might guess that he's taller than average, since there's a positive correlation between *height* and *shoe size*. But would you guess that he's 2 SDs above the mean? When there was no correlation, we didn't move away from the mean at all. With a perfect correlation, we moved our guess the full 2 SDs. Any correlation between these extremes should lead us to move somewhere between 0 and 2 SDs above the mean. (To be exact, the equation tells us to move $r \times 2$ standard deviations away from the mean.)

$r = .85$

Shoe size $= +2$

Height $= .85(+2) = +1.70$

Ht $= 1.7$ SD. above the mean

## How Big Can Predicted Values Get?

- *r* cannot be bigger than 1 (in absolute value), so each predicted *y* tends to be closer to its mean (in standard deviations) than its corresponding *x* was.
- This property of the linear model is called regression to the mean; the line is called the regression line.

We are always multiplying by a number less than 1, so the prediction gets smaller (closer to zero standard deviations which is the mean).

| | Height ($x$) | Weight ($y$) |
|---|---|---|
| Mean | $\bar{x} = 68$ | $\bar{y} = 150$ |
| St. Dev. | $s_x = 4$ | $s_y = 25$ |
| Correlation | $r = 0.7$ | |

$\#S_y = r \cdot \#S_x$

a) How many standard deviations below the mean is this student's height of 60"? 8" below mean $= \frac{8}{4} = 2$ S.D. below

b) Using the regression factor $r$, how many standard deviations below the mean weight should we expect this student's weight to be? $\#S_{wt} = r \cdot \#S_{HT} = .7(2) = 1.4$ S.D. below mean wt.

c) How many pounds is that? $1.4(25) = 35$ lbs. below average

d) What's the predicted weight? $150 - 35 = 115$ lbs.

Homework:

Pg. 187 #18 - 21

**18. Real estate redux** For *Price* and *Size* of homes in Albuquerque as described in Exercise 16, the correlation is $r = 0.845$.

a) What would you predict about the *Price* of a home 1 standard deviation above average in *Size*?

b) What would you predict about the *Price* of a home 2 standard deviations below average in *Size*?

**19. Another ride** The correlation between the *Duration* of a roller coaster ride and the height of its initial *Drop*, described in Exercise 17, is $r = 0.39$.

a) What would you predict about the *Duration* of the ride on a coaster whose initial *Drop* was 1 standard deviation below the mean *Drop*?

b) What would you predict about the *Duration* of the ride on a coaster whose initial *Drop* was 3 standard deviations above the mean *Drop*?

**20. More real estate** Consider the Albuquerque home sales from Exercise 16 again. The regression analysis gives the model $\widehat{Price} = 47.82 + 0.061 Size$.

Price is in $1000
Size is per sq ft

a) Explain what the slope of the line says about housing prices and house size.

b) What price would you predict for a 3000-square-foot house in this market?

c) A real estate agent shows a potential buyer a 1200-square-foot home, saying that the asking price is $6000 less than what one would expect to pay for a house of this size. What is the asking price, and what is the $6000 called?

**21. Last ride** Consider the roller coasters described in Exercise 17 again. The regression analysis gives the model $\widehat{Duration} = 64.232 + 0.180 Drop$.

Dur. in sec.
Drop in ft.

a) Explain what the slope of the line says about how long a roller coaster ride may last and the height of the coaster.

b) A new roller coaster advertises an initial drop of 200 feet. How long would you predict the rides last?

c) Another coaster with a 175-foot initial drop advertises a 90-second ride. Is this longer or shorter than you'd expect? By how much? What's that called?