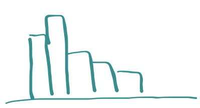
Reminders of some
Key ideas and points

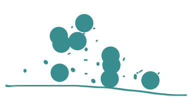
before tomorrow's Ch. 3 TEST

3. How to Describe a Scatter Plot...... **DUFS**

Distributions of 1 variable

SOCV



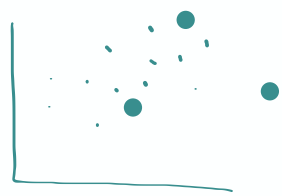


D irection

U nusual features

F orm

S trength



How to Describe a Scatterplot

To describe a scatterplot, make sure to address the following fo...
characteristics in the context of the data...

- **Direction:** A scatt...

CAUTION:

When ***describing*** the association shown in a scatterplot, write ***in the context of the problem.***

This means that you need to use both variable names in your description.

- ... quite a bit from the ...
- ...**features:** Look for outliers that fall outside the overall pattern and distinct clusters of points.

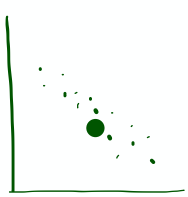
For a linear association between two quantitative variables, the **correlation *r*** measures the direction and strength of the association.

direction

strength

CAUTION:

It is only appropriate to use the correlation to describe strength and direction for a ***linear*** relationship.



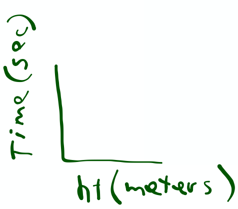
The correlation of $r = -.927$ confirms that the linear association between years since 1900 and 100-meter record time is strong and negative

strength direction

- Correlation doesn't imply causation
- Correlation doesn't measure form
- Correlation should only be used to describe a linear association
- Correlation isn't a resistant measure of strength
- Correlation is just a supplement to a scatterplot—don't start with correlation

$$r = \frac{1}{n-1}\sum\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)$$

$$= \frac{1}{n-1}\sum z_x z_y$$



If you change units of any of the variables, the correlation does not change. This is because it is calculated from standardized scores, which are independent of units.

1. Correlation requires that both variables be quantitative.
2. Correlation makes no distinction between explanatory and response variables.
3. r does not change when we change the units of measurement of x, y, or both.
4. The correlation r has no unit of measurement. It's just a number.

LSRL

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

y-intercept

slope

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

Big Ideas:

LT#1: Least squares regression line (LSRL)

- The line with the smallest sum of (residuals)^2.

$\bar{x}$ s_x

$\bar{y}$ s_y

r

$\sum (y_i - \hat{y})^2$

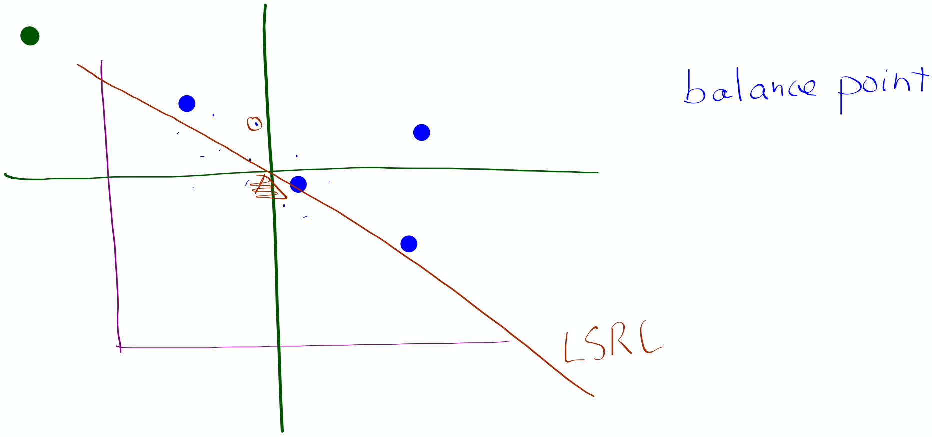
Want it small as possible

slope $b_1 = r \cdot \frac{s_y}{s_x}$

y-int $b_0 = \bar{y} - b_1\bar{x}$

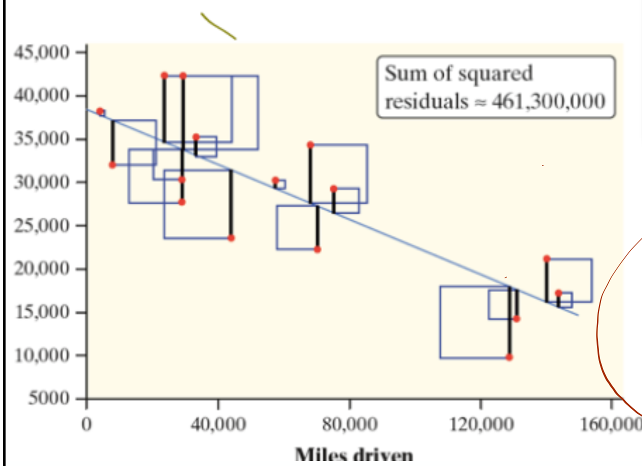
#5 The LSRL changes very slightly. Outliers in the vertical direction have less influence than the than the outliers in the horizontal direction

#6 Outliers can greatly affect the LSRL, with outliers in the horizontal direction having more influence than the vertical direction.



CAUTION:

A strong association between two variables is not enough to draw conclusions about cause and effect.



The **least-squares regression line** is the line that makes the sum of the squared residuals as small as possible.

6. Is the LSRL resistant to outliers

How do we know if a linear model is adequate?

After all, isn't it possible another type of model might be even better?

Determining if a Linear Model Is Appropriate: Residual Plots

A residual plot magnifies the deviations of the points from the line, making it easier to see unusual observations and patterns. If a regression model is appropriate:

- The residual plot should show no obvious patterns.
- The residuals should be relatively small in size.

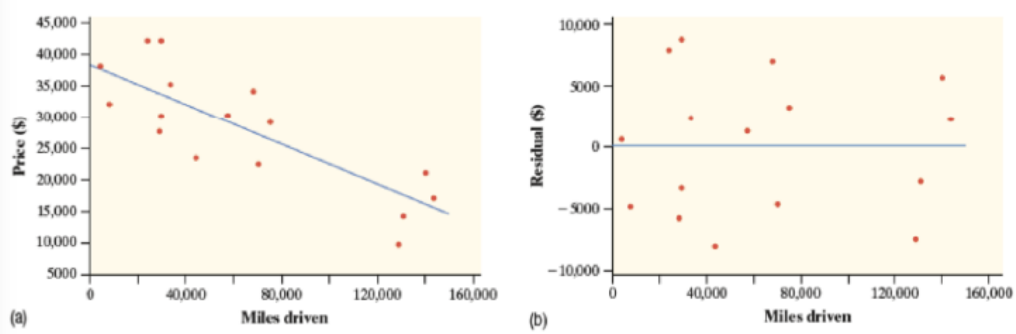


FIGURE 3.11 The (a) scatterplot and (b) residual plot for the relationship between price and miles driven for Ford F-150s.

If so, we can use

S and r^2

to determine how good the predictions will be.

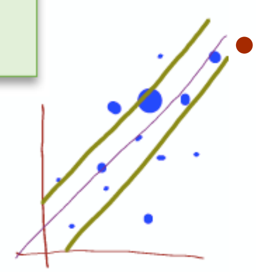
(How well does the line work)

S Standard Deviations of the residuals

r^2 Coefficient of Determination

The **standard deviation of the residuals**, ***s***, measures the size of a typical residual. That is, ***S*** measures the typical distance between the actual *y* values and the predicted *y* values.

y $\hat{y}$



S is sometimes called the typical prediction error.

The **standard deviation of the residuals *s*** measures the size of a typical residual. That is, ***s*** measures the typical distance between the actual *y* values and the predicted *y* values.

$$s = \sqrt{\frac{\sum residuals^2}{n-2}} = \sqrt{\frac{\sum (y_i - \hat{y})^2}{n-2}}$$

Most likely you will be given this value.

Coefficient of Determination

r^2 measures the fraction of the variability in the *y* variable that is accounted for by the LSRL using *x*.

2. Find and interpret *s*.

$s = 2.27$

The actual backpack weight is typically about 2.27 lb. away from the weight predicted by the LSRL with x = the body weight.

3. Find and interpret the value of r^2.

$r^2 = 0.632$

About 63.2% of the variability in backpack weight is accounted for by the LSRL with x = body weight.

Minitab

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

(1) (2)

(2) Slope

(1) y intercept

Predictor	Coef	SE Coef	T	P
Constant	$\boxed{38257}$	2446	15.64	0.000
Miles Driven	$\boxed{-0.16292}$	0.03096	-5.26	0.000

r^2

$\boxed{S = 5740.13}$ $\boxed{\text{R-Sq} = 66.4\%}$ R-Sq(adj) = 64.0%

r^2

Standard deviation of the residuals

Starnes, The Practice of Statistics, 6e | Student Resources By Type

Chapter Review Exercise Videos | Data Sets | Errata | Flashcards | FRAPPY! Student Samples | HP Prime Technology Corners | Statistical Applets | Extra Applets | TI-Nspire Technology Corners | TI-89 Technology Corners | Technology Corner Videos | Worked Example Videos | Worked Exercise Videos

Student Resources By Type

You can also view resources by chapter

Chapter Review Exercise Videos

- TPS6e_ReviewExercise_R1.1
- TPS6e_ReviewExercise_R1.2
- TPS6e_ReviewExercise_R1.3
- TPS6e_ReviewExercise_R1.4
- TPS6e_ReviewExercise_R1.5
- TPS6e_ReviewExercise_R1.6
- TPS6e_ReviewExercise_R1.7
- TPS6e_ReviewExercise_R1.8

Videos available for all Ch. Review Exercises.

Formulas

(I) Descriptive Statistics

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s_x = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2}$$

$$s_p = \sqrt{\frac{\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2}{\left(n_1 - 1\right) + \left(n_2 - 1\right)}}$$

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

$$b_1 = \frac{\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum\left(x_i - \bar{x}\right)^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$$b_1 = r\frac{s_y}{s_x}$$

$$s_{b_1} = \frac{\sqrt{\frac{\sum\left(y_i - \hat{y}_i\right)^2}{n-2}}}{\sqrt{\sum\left(x_i - \bar{x}\right)^2}}$$

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s_x = \sqrt{\frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2}$$

$$s_p = \sqrt{\frac{\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2}{\left(n_1 - 1\right) + \left(n_2 - 1\right)}}$$

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

$$b_0 = \bar{y} - b_1\bar{x}$$

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

$$b_1 = r\frac{s_y}{s_x}$$

Tomorrow

HW due, each with scores 0, 3, 5, 7

Total $\frac{}{42}$ 6 assignments

Two Options Today (can do both)
for in-class

A 3.2 Practice Quiz - Can Check Solutions
- Solutions stay in class (No photos)

B Frappy! + Model Solution + 2 student samples with scores

Assignment
Do all Review Problems pp 215-217