Sec 9.3: Measures of Regression & Prediction Intervals

Three Types of Variation About a Regression Line

1. **TOTAL** variation
2. **EXPLAINED** variation
3. **UNEXPLAINED** variation

******Without** a regression line, the best predictor for y given a value for x is $\bar{y}$

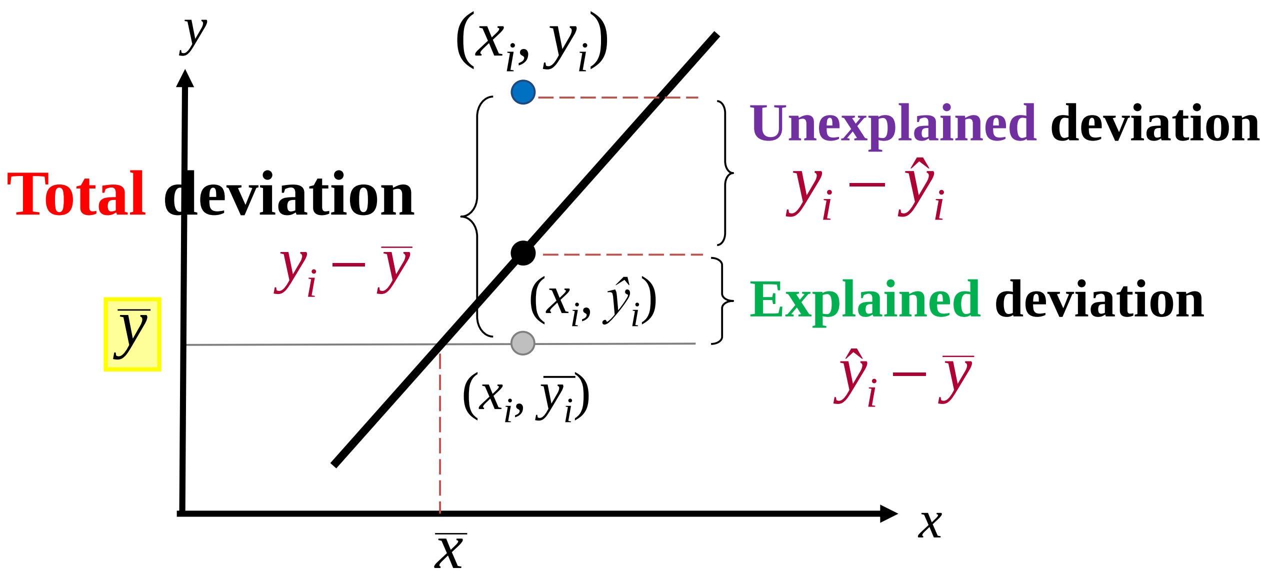
(*the mean of the y-values*)

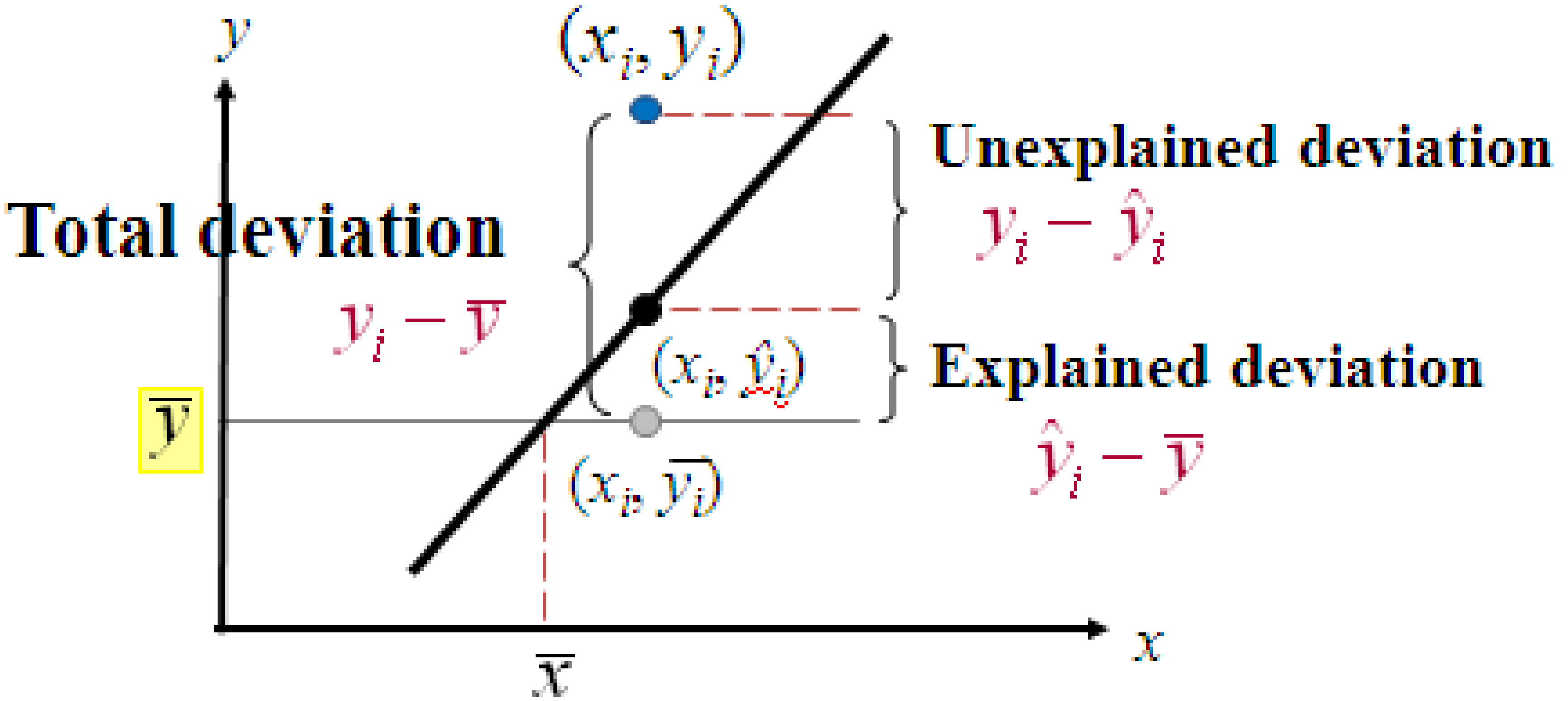
Variation About a Regression Line

TOTAL Deviation = $y_i - \bar{y}$

EXPLAINED Deviation= $\hat{y}_i - \bar{y}$ Can be explained by relationship between x & y

UNEXPLAINED Deviation= $y_i - \hat{y}_i$ **Cannot** by rel. between x & y …**chance or other variables**





TOTAL variation = EXPLAINED + UNEXPLAINED

Variation About a Regression Line

TOTAL variation: sum of the squares of the differences between the ***y*-value** of each ordered pair and the **mean** of ***y***.

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

EXPLAINED variation: sum of the squares of the differences between each **predicted** *y*-value and the **mean** of *y*.

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

UNEXPLAINED variation: sum of the squares of the differences between the ***y*-value** of each ordered pair & each corresponding **predicted** *y*-value.

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

TOTAL variation = EXPLAINED + UNEXPLAINED

Coefficient of Determination: The ratio of the explained variation to the total variation.

Denoted by r^2

$$r^2 = \frac{\text{Explained variation}}{\text{Total variation}}$$

Ex: Coefficient of Determination

The correlation coefficient for the Old Faithful data was r ≈ 0.979

Find the coefficient of determination.

$$r^2 = (0.979)^2$$

What does this mean?

$$r^2 = 0.958$$

- About **95.8%** of the variation in times is **EXPLAINED** by the variation in durations**…relationship between x & y!**

- About **4.2%** of the variation is **UNEXPLAINED**
…due to other factors or to sampling error.

Standard Error of Estimate: The standard deviation of the **observed** y_i -values about the **predicted** $\hat{y}$-value for a given x_i -value.

Denoted by $\boldsymbol{s_e}$

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

n is the number of ordered pairs in the data set

- The closer the observed *y*-values are to the predicted *y*-values, the **smaller** the standard error of estimate is

The Standard Error of Estimate

In Words	***In Symbols***
1. Make a table that includes the column heading shown.	$x_i,\ y_i,\ \hat{y}_i,\ (y_i - \hat{y}_i),\ (y_i - \hat{y}_i)^2$
2. Use the regression equation to calculate the predicted *y*-values.	$\hat{y} = mx_i + b$
3. Calculate the sum of the squares of the differences between each observed *y*-value and the corresponding predicted *y*-value.	$\Sigma(y_i - \hat{y}_i)^2$
4. Find the standard error of estimate.	$s_e = \sqrt{\dfrac{\Sigma(y_i - \hat{y}_i)^2}{n-2}}$

The regression equation for the advertising expenses and company sales data as calculated in section 9.2 is: $\hat{y} = 50.729x + 104.061$

Find the standard error of estimate. You will be doing this by hand on the quiz and using the computer for the project.

x	y	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$
2.4	225	225.81	$(225 - 225.81)^2 = 0.6561$
1.6	184	185.23	$(184 - 185.23)^2 = 1.5129$
2.0	220	205.52	$(220 - 205.52)^2 = 209.6704$
2.6	240	235.96	$(240 - 235.96)^2 = 16.3216$
1.4	180	175.08	$(180 - 175.08)^2 = 24.2064$
1.6	184	185.23	$(184 - 185.23)^2 = 1.5129$
2.0	186	205.52	$(186 - 205.52)^2 = 381.0304$
2.2	215	215.66	$(215 - 215.66)^2 = 0.4356$
			$\Sigma = 635.3463$

UNEXPLAINED VARIATION

$n = 8, \quad \Sigma(y_i - \hat{y}_i)^2 = 635.3463$

$$s_e = \sqrt{\frac{\Sigma(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{635.3463}{8-2}} \approx 10.290$$

The **standard error of estimate** of the company sales for a specific advertising expense is about **$10,290**

Ex# 2

Team Detroit concludes that there is a significant relationship between the amount of radio advertising time (in minutes per week) and the weekly sales of a product (in hundreds of dollars.)

A) Find the regression equation
B) Use the regression equation to find the predicted $\hat{y}$ -values
C) Calculate the sum of the squared differences of each observed y-value and the corresponding predicted y-value
D) Calculate Se
E) INTERPRET the results

Radio ad time	*Weekly sales*
15	26
20	32
20	38
30	56
40	54
45	78
50	80
60	88

Linear regression equation:

$$\hat{y} = 1.4054x + 7.3108$$

x	y	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$
15	26	28.39	
20	32	35.41	
20	38	35.41	
30	56	49.46	
40	54	63.51	
45	78	70.54	
50	80	77.56	
60	88	91.61	

Linear regression equation:

$$\hat{y} = 1.4054x + 7.3108$$

x	y	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$
15	26	28.39	$(26 - 28.39)^2 = 5.7121$
20	32	35.41	$(32 - 35.41)^2 = 11.6281$
20	38	35.41	$(38 - 35.41)^2 = 6.7081$
30	56	49.46	$(56 - 49.46)^2 = 42.7716$
40	54	63.51	$(54 - 63.51)^2 = 90.4401$
45	78	70.54	$(78 - 70.54)^2 = 55.6516$
50	80	77.56	$(80 - 77.56)^2 = 5.9536$
60	88	91.61	$(88 - 91.61)^2 = 13.0321$
			$\Sigma = 231.8973$

UNEXPLAINED variation

$$n = 8,\ \Sigma(y_i - \hat{y}_i)^2 = 231.8973$$

$$s_e = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{231.8973}{6}} = \sqrt{38.64955} =$$ 6.217 in hundreds of dollars

The **standard error of estimate** of the weekly sales for a specific radio ad time is about **$621.70**.

9.3 Work

p. 531 #1-7all, **11, 13 need FULL chart showing work!**