14.7: Prediction

One valuable application of the regression model is to make predictions about the value of the dependent variable if the independent variable is known.

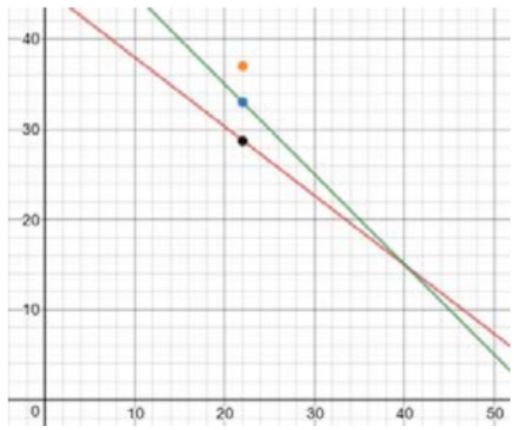
Consider the example about rainfall and sunglasses sales. Suppose we know that a city has 22 inches of rainfall. We can use the regression equation to predict the sales of sunglasses:

$\hat{Y} = 45.647 - .767X$

$\hat{Y}_{22} = 45.647 - .767(22) = 28.7$

For a city with 22 inches of annual rainfall, the model predicts sales of 28.7 per 1000 population.

To measure the **reliability** of this prediction, we can construct confidence intervals. However, we first have to decide what we are estimating. We could (1) be estimating the **expected** sales for a city with 22 inches of rainfall, or we could (2) be predicting the **actual** sales for a city with 22 inches of rainfall.



In the graph shown, the green line represents $Y = \beta_0 + \beta_1 X + \varepsilon$ the actual regression line which is unknown. The red line represents the least square equation, $\hat{Y} = 45.647 - .767X$, which is derived from the data. The black dot represents our prediction $Y_{22} = 28.7$. The green dot represents the correct population **expected** value of Y_{22}, while the yellow dot represents a possible value for the **actual** predicted value of Y_{22}. There is more uncertainty in predicting an actual value of Y_x than the expected value.

Confidence interval and Prediction interval

The **confidence interval** for the **expected** value of Y for a given value of X is given by:

$$\hat{Y}_X \pm t \cdot s_e \sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{SSX}}$$

Degrees of freedom for $t = n$-2

The **prediction interval** for the **actual** value of Y for a given value of X is given by:

$$\hat{Y}_X \pm t \cdot S_e \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{SSX}}$$

Degrees of freedom for $t = n$-2

Example: Rainfall sunglasses sales

1. Find a 95% confidence interval for the expected value of sales for a city with 22 inches of rainfall.
2. Find a 95% prediction interval for the value of sales for a city with 22 inches of rainfall.

Solution

1. **Confidence interval**

$$28.7 \pm 3.182 \cdot 3.586 \sqrt{\frac{1}{5} + \frac{(22-23)^2}{580}} = 28.7 \pm 5.1 \rightarrow (23.6, 33.8)$$

We are 95% confident that the expected annual sales of sunglasses for a city with 22 inches of annual rainfall is between 23.6 and 33.8 sales per 1000 population.

2. **Prediction interval**

$$28.7 \pm 3.182 \cdot 3.586 \sqrt{1 + \frac{1}{5} + \frac{(22-23)^2}{580}} = 28.7 \pm 12.5 \rightarrow (16.2, 41.2)$$

We are 95% confident that the actual annual sales of sunglasses for a city with 22 inches of annual rainfall is between 16.2 and 41.2 sales per 1000 population.