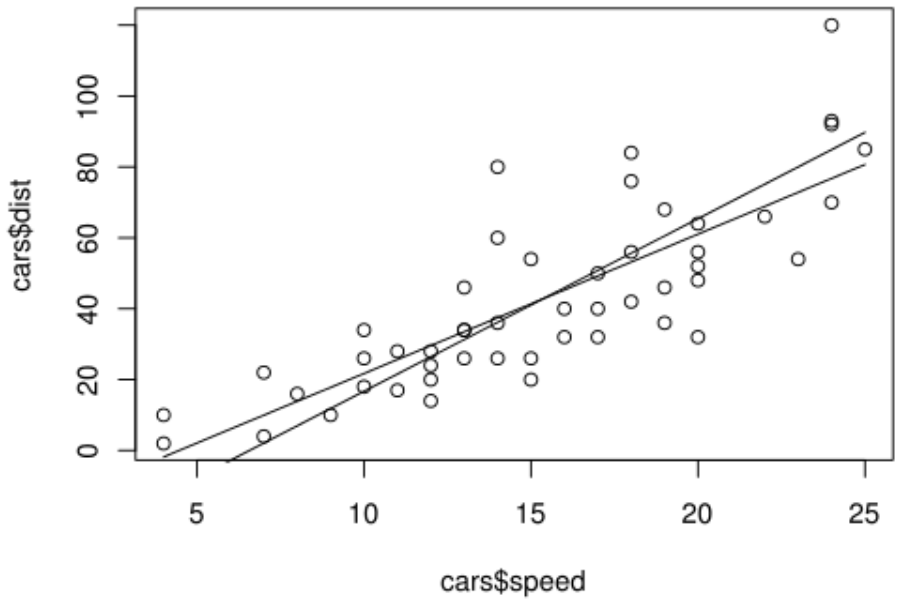
Math 207: Introduction to Statistics

Chapter 10: Regression



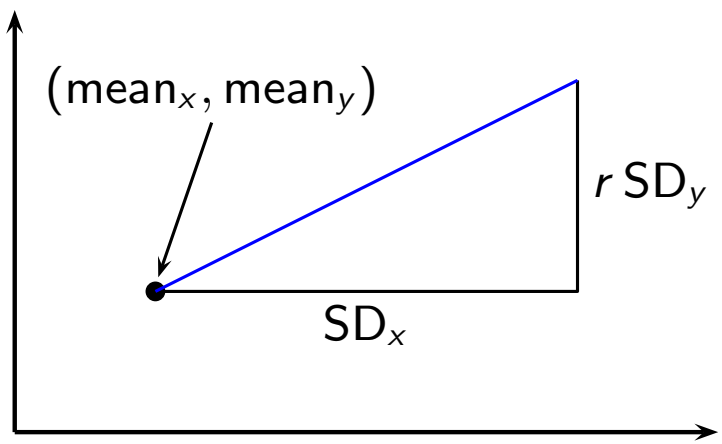
Dr. Ralph Wojtowicz

SHENANDOAH UNIVERSITY

The Regression Line

- The regression line for y on x estimates the average value for y corresponding to each value of x.
- Associated with each increase of one SD in x, there is an increase of only r SDs in y, on the average.
- To see why r is the right factor, consider the cases $r = 0$, $r = 1$ and $r = -1$.
- The regression line is a smoothed version of the graph of averages.
- The equation for the regression line is:

$$(y - \text{mean}_y) = r \frac{\text{SD}_y}{\text{SD}_x} (x - \text{mean}_x)$$



Example p. 165 (part I)

For the men age 18–24 in HANES5, the relationship between height and weight can be summarized as follows:

$$\begin{array}{llll} \text{average height} \approx 70 \text{ inches}, & \text{SD} \approx 3 \text{ inches}, & & \\ \text{average weight} \approx 180 \text{ pounds}, & \text{SD} \approx 45 \text{ pounds}, & r \approx 0.40 \end{array}$$

Example p. 165 (part I)

For the men age 18–24 in HANES5, the relationship between height and weight can be summarized as follows:

$$\begin{array}{llll} \text{average height} \approx 70 \text{ inches}, & \text{SD} \approx 3 \text{ inches}, & & \\ \text{average weight} \approx 180 \text{ pounds}, & \text{SD} \approx 45 \text{ pounds}, & r \approx 0.40 \end{array}$$

- Find the equation for the regression line to predict weight from height.

$$(y - 180) = 0.4 \cdot \frac{45}{3}(x - 70)$$

$$(y - 180) = 6\,(x - 70)$$

Example p. 165 (part I)

For the men age 18–24 in HANES5, the relationship between height and weight can be summarized as follows:

$$\begin{array}{llll} \text{average height} \approx 70 \text{ inches}, & \text{SD} \approx 3 \text{ inches}, & & \\ \text{average weight} \approx 180 \text{ pounds}, & \text{SD} \approx 45 \text{ pounds}, & r \approx 0.40 \end{array}$$

- Find the equation for the regression line to predict weight from height.

$$(y - 180) = 0.4 \cdot \frac{45}{3}(x - 70)$$

$$(y - 180) = 6\,(x - 70)$$

- Predict the weight of a subject who was 6'2".

$$y = 180 + 6\,(74 - 70) = 180 + 24 = 204 \text{ pounds}$$

Example p. 165 (part I)

For the men age 18–24 in HANES5, the relationship between height and weight can be summarized as follows:

$$\begin{array}{lll} \text{average height} \approx 70 \text{ inches}, & \text{SD} \approx 3 \text{ inches}, & \\ \text{average weight} \approx 180 \text{ pounds}, & \text{SD} \approx 45 \text{ pounds}, & r \approx 0.40 \end{array}$$

- Find the equation for the regression line to predict weight from height.

$$(y - 180) = 0.4 \cdot \frac{45}{3}(x - 70)$$

$$(y - 180) = 6\,(x - 70)$$

- Predict the weight of a subject who was 6'2".

$$y = 180 + 6\,(74 - 70) = 180 + 24 = 204 \text{ pounds}$$

- Predict the weight of a subject who was 5'6".

$$y = 180 + 6\,(66 - 70) = 180 - 24 = 156 \text{ pounds}$$

Example p. 165 (part I)

For the men age 18–24 in HANES5, the relationship between height and weight can be summarized as follows:

$$\begin{array}{llll} \text{average height} \approx 70 \text{ inches}, & \text{SD} \approx 3 \text{ inches}, & & \\ \text{average weight} \approx 180 \text{ pounds}, & \text{SD} \approx 45 \text{ pounds}, & r \approx 0.40 \end{array}$$

- Find the equation for the regression line to predict weight from height.

$$(y - 180) = 0.4 \cdot \frac{45}{3}(x - 70)$$

$$(y - 180) = 6\,(x - 70)$$

- Predict the weight of a subject who was 6'2".

$$y = 180 + 6\,(74 - 70) = 180 + 24 = 204 \text{ pounds}$$

- Predict the weight of a subject who was 5'6".

$$y = 180 + 6\,(66 - 70) = 180 - 24 = 156 \text{ pounds}$$

- What was the average weight of all subjects who were 6'2" tall?
  Answer: same as the previous answer: 204 pounds.

Example p. 165 (part II)

For the men age 18–24 in HANES5, the relationship between height and weight can be summarized as follows:

$$\begin{array}{llll} \text{average height} \approx 70 \text{ inches}, & \text{SD} \approx 3 \text{ inches}, & & \\ \text{average weight} \approx 180 \text{ pounds}, & \text{SD} \approx 45 \text{ pounds}, & r \approx 0.40 \end{array}$$

Example p. 165 (part II)

For the men age 18–24 in HANES5, the relationship between height and weight can be summarized as follows:

$$\begin{array}{llll} \text{average height} \approx 70 \text{ inches}, & \text{SD} \approx 3 \text{ inches}, & & \\ \text{average weight} \approx 180 \text{ pounds}, & \text{SD} \approx 45 \text{ pounds}, & r \approx 0.40 \end{array}$$

- Find the equation for the regression line to predict height from weight.

$$(y - 70) = 0.4 \cdot \frac{3}{45}(x - 180)$$

$$(y - 70) = \frac{2}{75}(x - 180)$$

Example p. 165 (part II)

For the men age 18–24 in HANES5, the relationship between height and weight can be summarized as follows:

$$\begin{array}{llll} \text{average height} \approx 70 \text{ inches}, & \text{SD} \approx 3 \text{ inches}, & & \\ \text{average weight} \approx 180 \text{ pounds}, & \text{SD} \approx 45 \text{ pounds}, & r \approx 0.40 \end{array}$$

- Find the equation for the regression line to predict height from weight.

$$(y - 70) = 0.4 \cdot \frac{3}{45}(x - 180)$$

$$(y - 70) = \frac{2}{75}(x - 180)$$

- Predict the height of a subject who weighed 204 pounds.

$$y = 70 + \frac{2}{45}(204 - 180) = 70 + 0.64 = 70.64 \text{ inches}$$

Example p. 165 (part II)

For the men age 18–24 in HANES5, the relationship between height and weight can be summarized as follows:

$$\begin{array}{llll} \text{average height} \approx 70 \text{ inches}, & \text{SD} \approx 3 \text{ inches}, & & \\ \text{average weight} \approx 180 \text{ pounds}, & \text{SD} \approx 45 \text{ pounds}, & r \approx 0.40 \end{array}$$

- Find the equation for the regression line to predict height from weight.

$$(y - 70) = 0.4 \cdot \frac{3}{45}(x - 180)$$

$$(y - 70) = \frac{2}{75}(x - 180)$$

- Predict the height of a subject who weighed 204 pounds.

$$y = 70 + \frac{2}{45}(204 - 180) = 70 + 0.64 = 70.64 \text{ inches}$$

- Predict the height of a subject who weighed 156 pounds.

$$y = 70 + \frac{2}{45}(156 - 180) = 70 - 0.64 = 69.36 \text{ inches}$$

Example p. 165 (part II)

For the men age 18–24 in HANES5, the relationship between height and weight can be summarized as follows:

$$\begin{array}{llll} \text{average height} \approx 70 \text{ inches}, & \text{SD} \approx 3 \text{ inches}, & & \\ \text{average weight} \approx 180 \text{ pounds}, & \text{SD} \approx 45 \text{ pounds}, & r \approx 0.40 \end{array}$$

- Find the equation for the regression line to predict height from weight.

$$(y - 70) = 0.4 \cdot \frac{3}{45}(x - 180)$$

$$(y - 70) = \frac{2}{75}(x - 180)$$

- Predict the height of a subject who weighed 204 pounds.

$$y = 70 + \frac{2}{45}(204 - 180) = 70 + 0.64 = 70.64 \text{ inches}$$

- Predict the height of a subject who weighed 156 pounds.

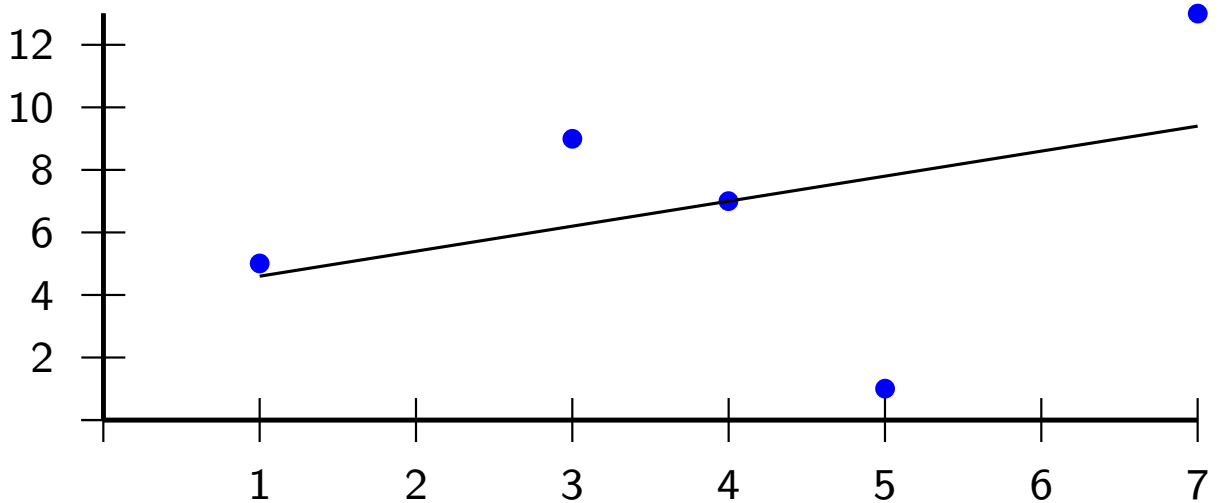
$$y = 70 + \frac{2}{45}(156 - 180) = 70 - 0.64 = 69.36 \text{ inches}$$

- What was the average height of all subjects who weighed 204 pounds?
  Answer: same as the previous answer: 70.64 inches.

Using R to Find a Regression Line

- Example from pages 132–133.

```
> x <- c(1, 3, 4, 5, 7)
> y <- c(5, 9, 7, 1, 13)
> linearModel <- lm(y ~ x)
> summary(linearModel)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    3.800      4.733   0.803    0.481
x              0.800      1.058   0.756    0.505
Residual standard error:  4.733 on 3 degrees of freedom
Multiple R-squared: 0.16, Adjusted R-squared: -0.12
F-statistic: 0.5714 on 1 and 3 DF, p-value: 0.5046
```

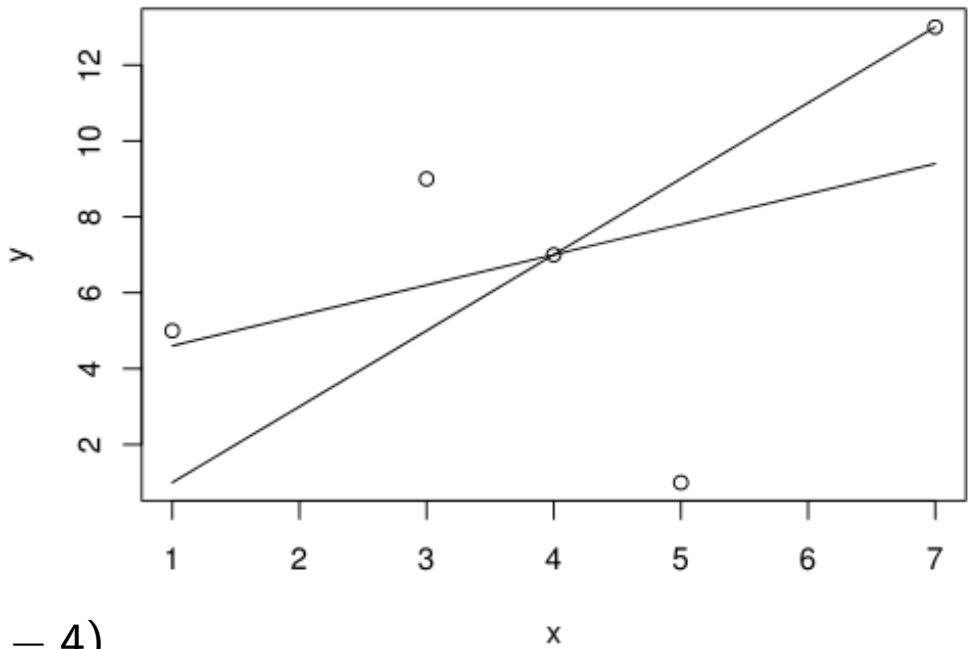


Plot the SD and Regression Lines

- Example from pages 132–133

```
> x <- c(1, 3, 4, 5, 7)
> y <- c(5, 9, 7, 1, 13)
```

```
> mean(x)
[1] 4
> mean(y)
[1] 7
> sd(y)/sd(x)
[1] 2
```



The SD line is $(y - 7) = (+1) \cdot 2\,(x - 4)$

```
> cor(x, y) [1] 0.4
```

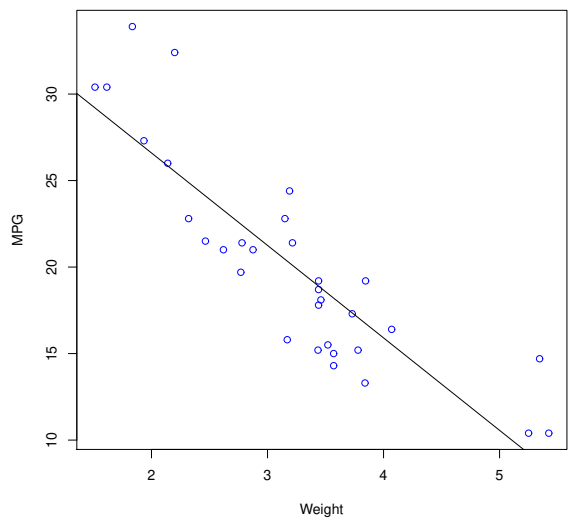
The regression line is $(y - 7) = 0.4 \cdot 2\,(x - 4)$

```
> plot(x, y)
> lines(x, 2*x - 1, type="l")
> lines(x, 0.8*x + 3.8, type="l")
```

Cars Data

- Use the following to get the equation for the regression lines for the mtcars data:

```
model = lm(mtcars$mpg ~ mtcars$wt)
summary(model)
plot(mtcars$wt, mtcars$mpg, col='blue', xlab='Weight', ylab='MPG')
abline(model)
```



- Try some other combinations of variables.