Let's work with a made-up example. Suppose that we measure the class attendance and the outdoor temperature for the 38 lectures of psych 315 this quarter.

Here are some statistics for our made up data: (1) temperatures (x) are distributed with a mean of 60 and a standard deviation of 8 degrees, (2) attendance (y) is distributed with a mean of 65 and a standard deviation of 10 students, and (3) temperature and attendance correlates with a value of -0.6.

$\bar{x} = 60$, $s_x = 8$

$\bar{y} = 65$, $s_y = 10$ $\quad r = -0.60$

what is the regression line?

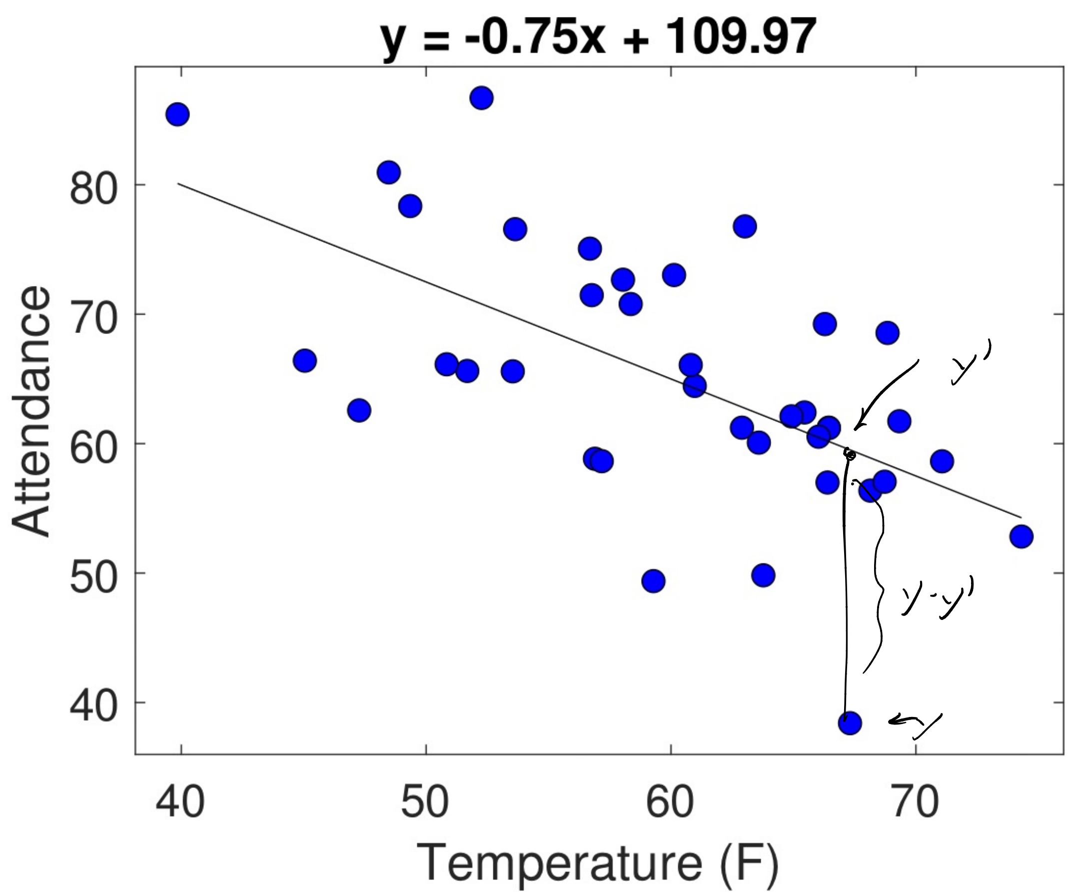
It goes through $(\bar{x}, \bar{y}) = (60, 65)$

has slope: $r \frac{s_y}{s_x} = -0.6 \frac{10}{8} = -0.75$

$$y' - \bar{y} = r \cdot \frac{s_y}{s_x} (x - \bar{x})$$

$$y' - 65 = -0.75 (x - 60)$$

$$y' = -.75x + (60)(.75) + 65 = -.75x + 109.97$$



How well does this line fit the data?

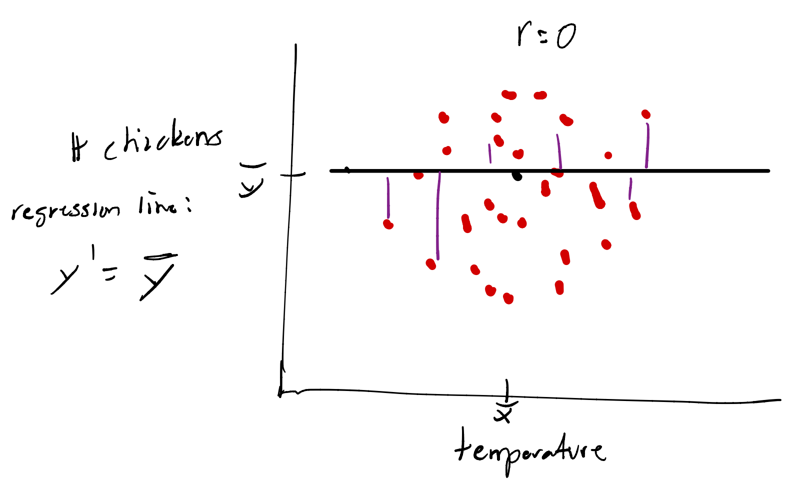
Standard error of the estimate: $s_{yx} = \frac{\sum (x'-y)^2}{n} = 8$

Another definition of s_{yx}

$$s_{yx} = s_y\sqrt{1-r^2} = 10\sqrt{1-(-.6)^2} = 8$$

if $r=1$ or $r=-1$, $r^2=1$, $\sqrt{1-r^2}=0$, $s_{yx}=0$

if $r=0$, $r^2=0$, $\sqrt{1-r^2}=1$, $s_{yx}=s_y$



slope $= r \cdot \frac{s_y}{s_x} = 0$

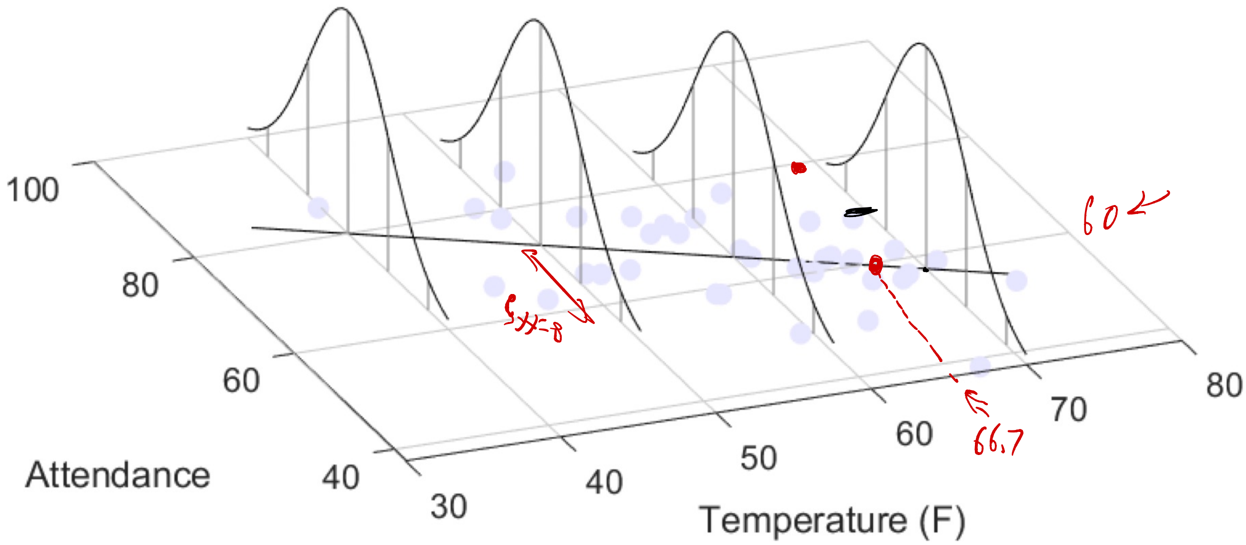
goes through $(\bar{x}, \bar{y})$

$$s_{yx} = \frac{\sum (y'-y)^2}{n}$$

$$s_{yx} = \frac{\sum (\bar{y}-y)^2}{n} = s_y$$

s.d of y.

Homoscedasticity



with homoscedasticity, the data points are distributed around the regression line as a normal distribution with mean of y' and a standard deviation of s_{yx}.

$s_{yx} = 8$

Example 1: What is the expected attendance when the outdoor temperature is 70 degrees?

$y' = -.75x + 109.97$

$y' = (-.75)(70) + 109.97 = 57.51$ students

Example 2: What is the temperature for which the expected attendance is 60?

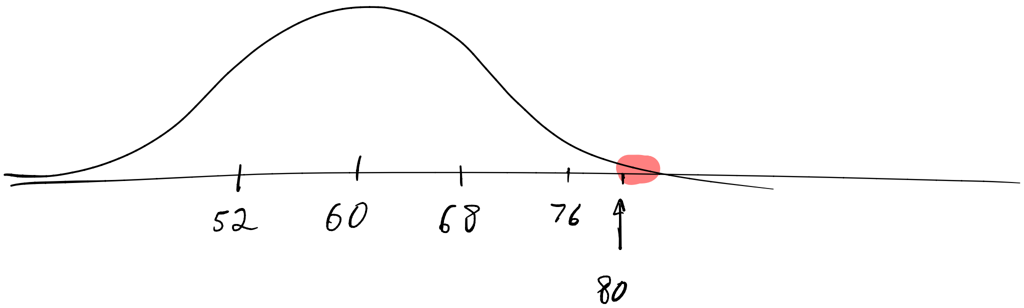
$y' = 60$ what is x?

$60 = -.75x + 109.97$

$x = \dfrac{60 - 109.97}{-.75} = 66.7$ degrees

Example 3: On a day that the outdoor temperature is 66.7, what is the probability that 80 or more students will show up for class?

we know that the number of students that come to class will be normally distributed with the mean of $y' = 60$ and a standard deviation of $s_{yx} = 8$



Z score for 80 with $\mu = 60$ $\sigma = 8$

$$Z = \frac{x - \mu}{\sigma} = \frac{80 - 60}{8} = 2.5$$

find the area above $Z = 2.5$ is 0.0062

dont worry about example 4 in the tutorial