

All about Regression

Sections 10.3-10.4

- Objectives
 - Compute the equation of the regression line
 - Make a prediction using the regression line

Section 10-4 regression

- Once we know that a linear relationship exists, we need to describe that relationship between two variables algebraically and graphically.
- Algebraically, we want to find the LINEAR equation that "best fits" the relationship (This is called the **Regression Equation**)
- Graphically, we graph the linear equation on the coordinate plane (This is called the **Regression Line**)

Regression

Regression Equation: $y' = a + bx$
 $= mx + y_{\text{int}}$

$$b = m = \text{slope} = \frac{n \left(\sum xy \right) - \left(\sum x \sum y \right)}{n \left(\sum x^2 \right) - \left(\sum x \right)^2}$$

$$a = y_{\text{int}} = \bar{y} - b_1 \bar{x}$$

Round the values for m and b to three decimal places

- The sign of the correlation coefficient, r , and the sign of the slope, m , of the regression line will always be the same.
- The regression equation can be used to make predictions for the dependent variable.
- The regression line will always pass through the point:
 (\bar{x}, \bar{y})

Quick checks

- For valid predictions, the following assumptions must be met:
 - For any specific value of the independent variable, x , the value of the dependent variable, y , must be normally distributed about the regression line
 - The standard deviation of each of the dependent variables must be the same for each value of the independent variable
 - The value of the correlation coefficient, r , must be significant.

Prediction assumptions

Linear Correlation Exists, r is significant

- Use y' (regression equation)
- Substitute given x -value and solve

Linear Correlation does NOT exist, r is not significant

- Use \bar{y}

Remember that when predictions are made, they are based on present conditions or on the premise that present trends will continue, which may or may not prove true in the future.

Predictions

- Listed below are the weights (in pounds) and the highway fuel consumption (in mi/gal) of randomly selected cars. Determine the regression equation that models the fuel consumption based on the weight of the vehicle. What is the predicted fuel consumption for a vehicle weighing 3000 pounds?

Weight (lbs)	3175	3450	3225	3985	2440	2500	2200
Fuel Consumption (mpg)	27	29	27	24	37	34	37

- Example: #6 p. 557
 - Construct a scatterplot
 - What type of relationship does the scatterplot suggest? Does the relationship, if any, appear to be positive or negative?
 - Calculate the linear correlation coefficient.
 - Based on the value of r , does there appear to be significant linear correlation in the data? Explain. (Hint: Use $\alpha = 0.05$)
 - Determine the regression equation.
 - Which value, y' or \bar{y} would be a better legitimate prediction?
 - What is the predicted diastolic pressure for a vegetarian that consumes 7.5 grams of protein each day?
 - Can we conclude that a change in the number of protein grams consumed changes the diastolic blood pressure?

Putting it all together