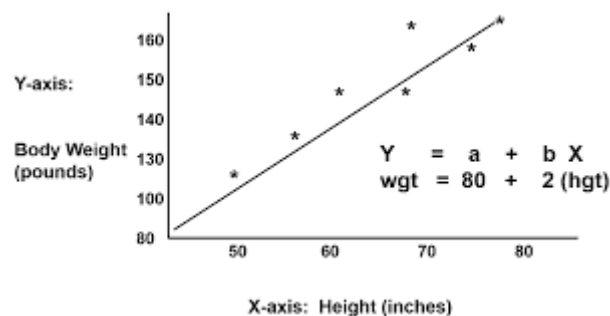


# Regression.

## - Univariate Regression:

- focuses on determining relationship between one independent (explanatory variable) variable and one dependent variable.
- **Equation of linear regression:**  $y_i = bx_i + a + e_i$   
y= Dependent var.  
x= independent var.  
b= Slope  
a= Intercept  
e= error value



- **Least-Square Method for finding values of the regression coefficients (slope):**

$$b = \frac{\sum(x - \bar{x}) * (y - \bar{y})}{\sum(x - \bar{x})^2}$$

$\bar{x}$  → Avg. / Mean of Independent features.

- **Performance Evaluation:** Evaluation metrics are a measure of how good a model performs and how well it approximates the relationship.

### ✓ MSE (Mean Squared Error):

- It is the average of the squared difference between the predicted and actual value.
- MSE penalizes large errors.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

✓ **MAE (Mean Absolute Error):**

- This is simply the average of the absolute difference between the target value and the value predicted by the model. Not preferred in cases where outliers are prominent.
- MAE does not penalize large errors.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

✓ **R-square:**

- R2 score is a metric that tells the performance of your model, not the loss in an absolute sense that how many wells did your model perform.
- R2 squared calculates how much regression line is better than a mean line.
- R2 squared is also known as Coefficient of Determination or sometimes also known as Goodness of fit

***R2 = Regression line error / Residual error (y - y-hat (predicted Val))***