Correlation & Regression

1) Formula: $\hat{y} = a + bx$, where

Slope: $$b = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2},$$

$y-$intercept: $$a = \frac{\left(\sum y\right)\left(\sum x^2\right) - \left(\sum x\right)\left(\sum xy\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

---

2) Linear Correlation Coefficient: r

1) Measures the strength of a linear relationship
2) $-1 \le r \le 1$

$$r = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

---

3) Coefficient of Determination: r^2

1) Measures the amount of variation in y that is explained by the linear relationship between x and y.
2) Write r^2 as percentage.

4) Standard error of estimate:

1) Measures the differences between the observed sample $y-$values and the predicted values $\hat{y}$ obtained by using the regression equation.

2) Find the equation of the regression line

3) Compute s_e:

$$s_e = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{n-2}}$$

---

5) Prediction Interval for an Individual y when a fixed value x_0 is given:

1) Find the equation of the regression line

2) Compute $\hat{y}$ for the given fixed value of x_0.

3) Compute the standard error of estimate s_e.

4) Find $t-$score with $n-2$ degrees of freedom for the required confidence level.

5) Compute $\overline{x}$.

6) Compute the margin of error E where

$$E = t \bullet s_e \bullet \sqrt{1+\frac{1}{n}+\frac{n(x_0-\overline{x})^2}{n\left(\sum x^2\right)-\left(\sum x\right)^2}}$$

Finally, the prediction interval for an individual y can be found by

$$\hat{y} - E < y < \hat{y} + E$$