# Linear Regression Formulae

CODEMATHS.IN

# Actual relationship between $x$ and $y$



$$y = f(x) + \epsilon$$

Estimate of $y_i = \hat{y}_i$



$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

# Estimate of $\beta_1 = \hat{\beta}_1$



$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

# Estimate of $\beta_0 = \hat{\beta}_0$

CODEMATHS.IN

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

# Degrees of Freedom



Degrees of Freedom = (Number of observations) - (Number of parameters in the model of the means)

# Standard Deviation Estimate

CODEMATHS.IN

$$\hat{\sigma} = \sqrt{\frac{Sum\ Of\ All\ Squared\ Residuals}{Degrees\ of\ Freedom}}$$

# Variance of $x$

CODEMATHS.IN

$$s_x^2 = \frac{\sum(x_i - \overline{x})^2}{n-1}$$

# Standard Error of $\hat{\beta}_1$

CODEMATHS.IN

$$\mathrm{SE}(\hat{\beta}_1) = \hat{\sigma}\sqrt{\frac{1}{(n-1){s_X}^2}}, \quad \text{d.f.} = n-2$$

$$SE(\hat{\beta}_1) = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(n-2)\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

# Standard Error of $\hat{\beta}_0$



$$\mathrm{SE}(\hat{\beta}_0) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\overline{X}^2}{(n-1){s_X}^2}}, \quad \text{d.f.} = n-2$$

$$SE(\hat{\beta}_0) = SE(\hat{\beta}_1)\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

# t-Statistic of parameter estimate



$$t - statistic = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

$$t - statistic = \frac{\hat{\beta}_0}{SE(\hat{\beta}_0)}$$

$$t - statistic = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)}$$

# Confidence interval of parameter estimate

$$\hat{\beta}_1 \pm t_{0.025,\, df} \times SE(\hat{\beta}_1)$$

$$\hat{\beta}_1 - t_{0.025,\, df} \times SE(\hat{\beta}_1)$$

$$\hat{\beta}_1 + t_{0.025,\, df} \times SE(\hat{\beta}_1)$$

$$\hat{\beta}_0 \pm t_{0.025,\, df} \times SE(\hat{\beta}_0)$$

$$\hat{\beta}_0 - t_{0.025,\, df} \times SE(\hat{\beta}_0)$$

$$\hat{\beta}_0 + t_{0.025,\, df} \times SE(\hat{\beta}_0)$$

# Analysis of Variance - Mean Square Error

CODEMATHS.IN

$$MSE = \frac{RSS\,(ResidualSumOfSquares)}{DFE\,(DegreesOfFreedomResiduals)}$$

$$RSS\left(ResidualSumOfSquares\right) = \sum(y_i - \hat{y}_i\right)^2$$

$$DFE\,(DegreesOfFreedomResiduals) = n - 2$$

# Analysis of Variance - Mean Square due to Regression

CODEMATHS.IN

$$MSR = \frac{SSR\,(SumOfSquaresRegression)}{DFR\,(DegreesOfFreedomRegression)}$$

$$SSR\,(SumOfSquaresRegression) = \sum \left(\hat{y}_i - \bar{y}\right)^2$$

$$DFR\,(DegreesOfFreedomRegression) = 1$$

# Analysis of Variance - Mean Square Error

CODEMATHS.IN

$$F = \frac{\text{Mean Square due to Regression (MSR)}}{\text{Mean Squared Error (MSE)}}$$

# $R^2$ Statistic

CODEMATHS.IN

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