1. **Simple Linear Regression**

Pearson's correlation coefficient: $r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}$

Regression Model:

$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad E(\epsilon_i) = 0, \quad Var(\epsilon_i) = \sigma^2, \quad Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$

Regression Coefficients:

$\hat{\beta}_1 = \frac{n\sum_{i=1}^{n} x_i \; y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$

Analysis of Variance:

SST = SSR+SSE where SST $= \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$ and

SSR $= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \hat{\beta}_1\left[\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)\right]$

$R^2 = \frac{SSR}{SST}, \quad s^2 = \frac{SSE}{n-2}$

Significance of Model:

$F = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/n-2}$; Reject H_0 at the α level of significance if $f > F_{\alpha;1,n-2}$ (significant model)

Inference on the Model:

$\hat{\beta}_i \pm t_{\alpha/2;n-2}\, s(\hat{B}_i)$, where $s(\hat{B}_1) = s\sqrt{\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$ and $s(\hat{B}_0) = s\sqrt{\frac{\sum x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}}$

Tests for β_1:

$H_0 : \beta_1 = \beta_1^0$ vs $H_1 : \beta_1$ ($\neq$ or $<$ or $>$) β_1^0 (*test for the slope*)

test whether $t = \frac{\hat{\beta}_1 - \beta_1^0}{s(\hat{B}_1)} \sim t_{n-2}$ at a certain significance level α

Tests for β_0:

$H_0 : \beta_0 = \beta_0^0$ vs $H_1 : \beta_0$ ($\neq$ or $<$ or $>$) β_0^0

test whether $t = \frac{\hat{\beta}_0 - \beta_0^0}{s(\hat{B}_0)} \sim t_{n-2}$ at a certain significance level α

Use of Model:

$\hat{y}_0 = b_0 + b_1 x_0$

Confidence Interval for $E(Y_0)$: $\hat{y}_0 \pm t_{\alpha/2;n-2}\; s\; \sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$

Prediction Interval for a new observation of Y: $\hat{y}_0 \pm t_{\alpha/2;n-2}\; s\; \sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$

2. **Multiple Linear Regression**

Regression Model:

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} + \epsilon_i,$$
$$E(\epsilon_i) = 0, \quad Var(\epsilon_i) = \sigma^2, \;\; Cov(\epsilon_i, \epsilon_j) = 0 \text{ for } i \neq j$$

Regression Coefficients: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

Analysis of Variance:

SST=SSR+SSE

Source	**df**	**SS**	**MS**
Regression	k	SSR	MSR= SSR/k
Error	$n-k-1$	SSE	MSE=SSE/$(n-k-1)$
Total	$n-1$	SST	

$R^2 = \frac{SSR}{SST}, \quad s^2 = \frac{SSE}{n-k-1} = \frac{SSE}{n-p}$

Significance of Model: $F = \frac{MSR}{MSE}$

Inference on the model:

- Confidence intervals: $\left[\hat{\beta}_j - t_{a/2} S\sqrt{c_{jj}}\,,\; \hat{\beta}_j + t_{a/2} S\sqrt{c_{jj}}\right]$, and c_{ii} are the diagonal elements of $(\mathbf{X}'\mathbf{X})^{-1}$
- Tests Statistics: $t = \frac{\hat{\beta}_j - \beta_{j(0)}}{S\sqrt{c_{jj}}}$

Model selection:

- $AIC = n\left[\log\left(2\pi SSE/n\right) + 1\right] + 2\left(p+1\right)$
- $BIC = n\left[\log\left(2\pi SSE/n\right) + 1\right] + \left(p+1\right)\log n$
- $R^2_{adj} = 1 - \frac{MSE}{MST} = 1 - \frac{SSE/(n-p)}{SST/(n-1)} = 1 - \frac{n-1}{n-p}\frac{SSE}{SST}$
- C_p–Mallows
- $F = \frac{(SSE_0 - SSE_1)/q}{SSE_1/(n-k-1)}$, where SSE_0, SSE_1 are the error sums of squares for the reduced (smaller) model and for the full (larger) model, respectively.

Use of Model:

$\hat{y}_0 = b_0 + b_1 x_1^0 + b_2 x_2^0 + \cdots + b_k x_k^0$

Confidence Interval for $E(Y_0)$: $\hat{y}_0 \pm t_{(1-\alpha/2);n-k-1}\; s(\hat{y}_0)$

Prediction Interval for a new observation of Y: $\hat{y}_0 \pm t_{(1-\alpha/2);n-k-1}\; s(\hat{y}_0 - y_0)$