Statistics:

- Covariance between two variables in a population: $cov(x,y) = \frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})$
- $cov(a_1 x + b_1, a_2 y + b_2) = a_1 a_2 cov(x,y)$
- $var(ax + by) = a^2 var(x) + b^2 var(y) + 2ab cov(x,y)$
- Formula for the difference in means of two independent samples: $var(\bar{X}_1 - \bar{X}_2) = var(\bar{X}_1) + var(\bar{X}_2)$
- When y is a binary variable then: $var(y) = p(1-p)$
- Pooled estimated proportion for two independent binary samples: $\hat{p}_{pooled} = \frac{n_1 \cdot \hat{p}_1}{n_1 + n_2} + \frac{n_2 \cdot \hat{p}_2}{n_1 + n_2}$

OLS:

Simple linear regression:

- $\hat{\beta}_1 = \frac{cov(x,y)}{var(x)}$
- $var(\hat{\beta}_1) = \frac{\hat{\sigma}_u^2}{SST_x}$ where $\hat{\sigma}_u^2 = \frac{\sum_i \hat{u}_i^2}{n-k-1}$ (k represents the number of regressors)

Multiple linear regression:

- $var(\hat{\beta}_j) = \frac{\hat{\sigma}_u^2}{SST_x(1-R_j^2)}$

Other:

- $SST = \sum_{i=1}^n (y_i - \bar{y})^2$
- $SSE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$
- $SSR = \sum_{i=1}^n \hat{u}_i^2$

Test Stats:

- General test statistic for population parameter θ: $\frac{\hat{\theta} - \theta_0}{se(\hat{\theta})}$
- General confidence interval for population parameter θ: $[\hat{\theta} - c \cdot se(\hat{\theta}),\ \hat{\theta} + c \cdot se(\hat{\theta})]$
- F statistic for q restrictions in a regression done with n observations and k exogenous variables:

$$F = \frac{(R_{UR}^2 - R_R^2)/q}{(1 - R_{UR}^2)/(n-k-1)} \sim F(q, n-k-1)$$

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n-k-1)} \sim F(q, n-k-1)$$