

Multiple Linear Regression Analysis

MLRS: Variation of dependent variable explained by multiple independent variables.

Multiple linear regression models expressed as

$$Y = b_0 + b_1x + b_2x + b_3x + \dots + \varepsilon$$

Hypothesis testing for regression coefficients is

$$t = \frac{\hat{b}_1 - b_1}{s_{b_1}} = \frac{\text{Estimated regression Coefficient} - \text{Hypothesized Value}}{\text{Coefficient of Std. Error}}$$

Degrees of freedom for t-statistic are $n-k-1$.

p-Value: Smallest level of significance that null value can be rejected.

If p-value is less than significance level null can be rejected, however if p-value is greater than significance value null can not be rejected.

Confidence Interval for regression coefficient is: $b_1 \pm \text{std. Err}$

F-statistic: estimates variation of dependent variable explained by group of independent variables.

$$F = \frac{MSR}{MSE} = \frac{RSS/k}{SSE/n-k-1}$$

Coefficient of determination $R^2 = \frac{\text{Expained Variation (RSS)}}{\text{Total Variation (SST)}}$

$$\text{Adjusted } R^2 = 1 - \left[\left\{ \frac{n-1}{n-k-1} \right\} * (1 - R^2) \right]$$

ANOVA Table:

ANOVA			
Source of Variation	DOF	Sum Squares	Mean Sum Squares
Regression (Explained)	K	RSS	MSR = (RSS/k)
Error (Unexplained)	n-k-1	SSE	MSE = SSE/(n-k-1)
Total	n-1	SST	

Independent variables binary in nature (0 or 1) are known as dummy variables. Slope coefficient is change in dependent variable when the dummy variable is one.

Three violations of multiple regression are:

Heteroskedasticity.
Serial (Auto) Correlation.
Multicollinearity.

Heteroskedasticity occurs when variance of residue term is not constant across all observation.

Unconditional Heteroskedasticity: it does not change with change in independent variables, does not cause major issue.

Conditional Heteroskedasticity: variance of residue increases as independent variable increases. Causes large error term reduces t-statistic and cannot reject the null. F tests are unreliable.

Use Breusch-Pagan chi square test to detect heteroskedasticity.

BP chi-square test = $n * R^2$ with **k** degrees of freedom

k = independent variable (not samples)

It's a one tail test.

Correcting Heteroskedasticity done by calculating robust *standard errors (White Corrected Errors)* and by recalculating t-statistics.

IF the residual term is correlated to each other it is called **serial correlation**.

Positive error in one term increase the probability of positive error term in the next period is called **Positive serial correlation**.

Positive error in one term increase the probability of negative error term in the next period is called **Negative serial correlation**.

Residual plots and Durbin-Watson are two methods are used to find the autocorrelation.

Durbin-Watson (DW) statistic is approximated as $2*(1-r)$.

When there is no auto correlation $DW = 2$ (since $r= 0$), $DW = 0$ when $r= 1$ that is when there is positive correlation, $DW = 4$ when $r= -1$ that is when there is perfect negative correlation.

DW rule:

0 to d_l : Positive Correlation

d_l to d_u : Inconclusive

d_u to $4-d_u$: Don't reject H_0

$4-d_u$ to $4-d_l$: Inconclusive

$4-d_l$ to 4: Negative serial correlation.

Correcting serial correlation:

1. Using Hansen method; this also corrects heteroskasticity, however use white space method for correcting heteroskasticity. If both problems persist then use Hansen method.
2. Improve specification method.

Multicollinearity: When two or more independent variables correlate to each other, causing incorrect conclusion of variables not significant (Type II error).

Detecting multicollinearity: Individual coefficients of independent variables are not significantly different from zero but F-test and R^2 are significant.

Correction: Dropping one the correlated variables.

Model misspecifications:

1. Important variables are omitted.
2. Variable should be transformed.
3. Data improperly pooled.
4. Using lagged dependent variable as independent variable.
5. Forecasting Past
6. Measuring error of independent variable (ex: quality of input)