SELECTION OF BEST REGRESSION EQUATION AND TEST FOR NORMALITY

By:

Dr. Akash Asthana

Assistant Professor,

Dept. of Statistics,

University of Lucknow, Lucknow

SELECTING BEST REGRESSION EQUATION

- Let it is required to establish a regression equation of response variable (Y) on predictors X_1, X_2, …, X_k. Also Z1, Z2, …, Zr be all the functions of X_j's representing the complete set of variables from which the equation is to be chosen. Then it is required to:

1. To make the equation useful for the predictive purpose it is required to keep as many Z's as possible in the model to keep bias errors small.
2. To keep the variance of predictors reasonably small and cost involved for obtaining information on Z's it is required to keep as few Z's as possible in the model.

SELECTING BEST REGRESSION EQUATION

- For the selection of best regression equation following methods are used:

1. Using all possible regressions
2. Forward selection method
3. Backward elimination method
4. Stepwise regression

USING ALL POSSIBLE REGRESSIONS

- In this method every possible regression equation is fitted. If there are r predictors then 2^r regression equations are possible
- Among all possible regression equation that is said to be best for which:

1. Value of R^2 and adjusted R^2 are high with smaller no. of predictors.
2. Value of C_p statistic is near to the no. of predictors.

C_P STATISTIC

- This statistic was defined by C.L. Mallows thus also called as Mallows' C_p statistic.
- The Cp statistic is given by:

$$C_p = \frac{RSS_p}{s^2 - (n - 2p)}$$

Where RSSp is the residual sum of square consisting of p parameters including β0.

s2 is mean sum of square due to residuals for the model consisting of all the predictors.

- If model includes all the r predictors then the value of Cr+1 will be r+1

FORWARD SELECTION METHOD

- In this method variables are entered in the model one by one.
- In this first the regression model $Y = \beta 0$ is considered.
- Then a variable which is most significant enters to the model.
- Then most significant variable among the remaining predictors enters to the model.
- The process is repeated till no significant predictors remain, i.e., all the predictors will become insignificant for the model.

FORWARD SELECTION METHOD

- The significance of the predictors for the model is decided with the help of following F statistic:

$$F_{j+r} = \frac{SSE_j - SSE_{j+r}}{MSE_j}$$

Where SSE_j = Residual sum of square with j predictors in the model,

$SSE_{(j+r)}$ = Residual sum of square with j+1 predictors in the model when r^{th} predictor enters,

MSE_j = mean sum of square due to residuals with j predictors in the model.

FORWARD SELECTION METHOD

- This F statistic is compared with the tabulated value of $F_{1, j, \alpha}$.
- That predictor will enter to the model for which $F_{j+r} > F_{1, j, \alpha}$ and maximum.
- Then next model is fitted and for each predictor this F statistic is again calculated.
- The process is repeated till all the remaining predictors become insignificant for the model.

BACKWARD ELIMINATION METHOD

- It is just opposite of forward selection method.
- In this first the complete model is fitted.
- Then variables are removed one by one from the model depending upon their significance for the model.
- The significance of the predictor is judged with the help of following F-statistic.

$$F_{j-r} = \frac{SSE_j - SSE_{j-r}}{MSE_j}$$

BACKWARD ELIMINATION METHOD

Where SSE_j = Residual sum of square with j predictors in the model,

$SSE_{(j-r)}$ = Residual sum of square with j-1 predictors in the model when r^{th} predictor is removed,

MSE_j = mean sum of square due to residuals with j predictors in the model.

- If $F_{j-r} < F_{1,j,\alpha}$ then r^{th} variable is removed from the model.
- The process is repeated till all insignificant predictors are removed from the model.

STEPWISE METHOD

- It is a mixture of Forward selection method and Backward elimination method.
- In this the regression is started with $Y=\beta_0$.
- Then predictors are entered in the model one by one depending upon their relative significance in the model.
- If any predictor in the model become insignificant at any stage then it is removed from the model.
- A predictor removed from the model will never enter again in the model.
- For the purpose both F_{j+r} and F_{j-r} are computed at each stage

ASSUMPTION OF NORMALITY

- In regression model it is assumed that the error term should be normally distributed with mean zero and variance σ_e^2, or the dependent variable follows Normal distribution with mean $X\beta$ and variance-covariance matrix $\sigma_e^2 I$.
- If the distribution of error term in not normal then the estimation of parameters remain unaffected but the testing of different hypothesis is affected as the distribution of different statistic will not remain same as in case of Normal distribution.
- For detection of departure from assumption of normality following methods are used:

1. Kolmogorov-Smirnov test
2. Shapiro-Wilk's Test
3. Q-Q (P-P) plot

KOLMOGOROV-SMIRNOV TEST

- In this the null hypothesis under consideration is that the distribution of error term is normal against the alternative hypothesis the distribution is not normal.
- For testing following steps are used:

1. First the residuals are arranged in the ascending order of magnitude.
2. For each value of residual two different cumulative probabilities are computed one under the null hypothesis $F(r_j)$ and other for $F_n(r_j)$ empirical density function defined by

$$F_n(r_j) = \begin{cases} 0; \ if \ j < 1 \\ \frac{i}{n}; if \ i - 1 < j < i \\ 1; if \ j > n \end{cases}$$

3. Then the test statistic is obtained by $D = \max|F_n(r_j) - F(r_j)|$
4. If the calculated value of D is greater than the tabulated value null hypothesis is rejected.

SHAPIRO–WILK'S TEST

- This test was defined by Shapiro SS and Wilk M in 1965.
- In this the null hypothesis is considered as the sample is selected form the Normal distribution.
- The testing procedure have following steps:

1. Let $Y_1, Y_2, \ldots, Y_n$ be n observation. These are arranged in the ascending order as $Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(n)}$.
2. Calculate sum of square using:

$$SS = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} \left(x_{(i)} - \bar{x}\right)^2$$

3. If n is even then take m=n/2 and if n is odd then take m=(n-1)/2
4. Calculate b using:

$$b = \sum_{i=1}^{m} a_i (x_{n-i+1} - x_i); \;\; a_i \; is \; obtained \; from \; the \; table \; for \; it$$

5. The Test statistic is given by: $W = b^2/SS$
6. If tabulated value of test statistic is larger than the calculated value null hypothesis can't be rejected.

Q-Q PLOT

- The Q-Q plot is also called as quantile-quantile plot.
- In this plot two different quantile are plotted against each other.
- For the purpose first observed values of the variable under consideration is arranged in ascending order.
- If there are n observations then the area of curve of distribution (Normal) is divided into n+1 parts. It will give n quantiles for the distribution under consideration.
- The observations arrange in ascending order served as second type of quantiles.
- These two quantiles are plotted against each other.
- If the points lies on the Line of 45^{o} then the variable is said to follow the hypothetical (Normal) distribution.

P-P PLOT

- It is also called as probability-probability plot.
- In this the empirical cumulative distribution is plotted against the theoretical cumulative distribution.
- For the purpose first the observations are arranged in ascending order.
- Then the cumulative probabilities under theoretical distribution (Normal) is computed.
- Also the cumulative probabilities for empirical distribution are computed using:

$$F(x_i) = i/n$$

- These two cumulative probabilities are plotted against each other.
- If all the point of the plot lies on the straight line making an angle of 45^o then variable is said follow theoretical distribution.

BOX–COX TRANSFORMATION

- If the distribution of dependent variable is not Normal then it is required to transform it into Normally distributed variable.
- For the purpose several transformations are used among which Box-Cox transformation is used.
- It is a unique transformation which give rise to different type of transformations.
- The Box-Cox transformation of variable Y is given by:

$$Y' = \begin{cases} \dfrac{Y^{\lambda} - 1}{\lambda}; if\ \lambda \neq 0 \\ ln(Y);\ \ if\ \lambda = 0 \end{cases}$$

BOX–COX TRANSFORMATION

- A disadvantage of transformation mentioned in previous slide is that as the size of λ varies Y' can change enormously which leads to the problem in analysis.
- To overcome this it is preferable to use another form of this transformation, which is given by:

$$Y'' = \begin{cases} \dfrac{Y^{\lambda} - 1}{\lambda \dot{Y}^{\lambda-1}} ; if\ \lambda \neq 0 \\ \dot{Y} ln(Y) \end{cases}$$

Where Ý is the geometric mean of the observations over variable under consideration

BOX–COX TRANSFORMATION

- Now a problem arise what should be the appropriate value of λ.
- For the purpose following steps are used:

1. Select different values of λ in a given range. Generally the range (-2, 2) is considered.
2. For each chosen value of λ fit the regression model $Y'' = X\beta + \epsilon$ and compute the residual sum of square $S(\lambda, Y'')$.
3. Plot $S(\lambda, Y'')$ against λ. Draw a smooth curve through these points. The value of λ for which the curve attends lowest value is called as Maximum Likelihood estimate of λ.

BOX–COX TRANSFORMATION

- For different values of λ Box-Cox give rise to different transformations called as Power transformations. Which are defined as follows:

λ	**Transformed data**
-2	Y^{-2}
-1	Y^{-1}
-0.5	$1/\sqrt{Y}$
0	ln (Y)
0.5	$\sqrt{Y}$
1	No transformation
2	Y^{2}

REFRENCES

1. Gujrati DN, Basic Econometrics, 4th edition (2004) , The McGraw-Hill Companies.
2. Draper NR & Smith H, Applied Regression Analysis, 3rd edition (1998), John Wiley & Sons Inc.
3. Johnston J & Dinardo J, Econometric Methods, 4th edition (1997), McGraw-Hill Companies.