Statistical Inference in Multiple Regression

Chapter 9

Yi Zhou

2016-1-29

Outline

9.1 Preview

Q: The contributions of X_i to Y

- *Overall* test: F test
- Test for additional of a *single* variable: partial F test
- Test for additional of a *group* variable: multiple partial F test

F test in regression analysis context:

$$F = \frac{\hat{\sigma}_0^2 \; (H_0 \;\; is \;\; true)}{\hat{\sigma}^2 \; (MS \;\; residual)}$$

- Full model
- Reduced model

$$F = \frac{\hat{\sigma}_0^2 \; (reduced \;\; model)}{\hat{\sigma}^2 \; (full \;\; model)} \quad (??)$$

9.2 Test for significant overall regression

- H_0: all k independent variables considered together do not explain a significant amount of the variation in Y.
- H_0: there is no significant overall regression using all k independent variables in the model
- H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

$$F = \frac{MS\ \ Regression}{MS\ \ Residual} \sim F_{k,n-k-1} \quad (9.1)$$

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F_{k,n-k-1} \quad (9.2)$$

$$R^2 = \frac{SSY - SSE}{SSY}$$

- $(SSY - SSE)/k$, $MS\ \ Regression$: $\hat{\sigma}_0^2 = \hat{\sigma}^2$ under H_0 (reduced model).
- $SSE/(n-k-1)$, $MS\ \ Residual$: $\hat{\sigma}^2$ under the assumed model (full model).

9.3 Partial F test I

Partial F test: whether the *addition* of any specific independent variable (only one) is significantly contributes to the prediction of Y.

- H_0: X^* does not significantly add to the prediction of Y given that $X_1, X_2, \ldots, X_p$ are already in the model.
- H_0: $\beta^* = 0$ in the model
  $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \beta^* X^* + E$
- Full model: $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \beta^* X^* + E$
- Reduced model: $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + E$

$$F(X^*|X_1, \ldots, X_p) = \frac{MS(X^*|X_1, \ldots, X_p)}{MS(X_1, \ldots, X_p, X^*)} \sim F_{1,n-p-1} \quad (9.4)$$

- $F(X_2|X_1)$ and $F(X_3|X_1, X_2)$
- t test

9.3 Partial F test II

Application: the control of extraneous variable.

To work backward by deleting X^* variables, at one time, until a best model is obtained. (Chapter 16, P_{447})

9.4 Multiple partial F test I

Multiple partial F test: whether the simultaneously addition of any specific independent variables (more than one) are significantly contributes to the prediction of Y.

- H_0: $X_1^*, X_2^*, \ldots, X_s^*$ do not significantly add to the prediction of Y given that $X_1, X_2, \ldots, X_p$ are already in the model.
- H_0: $\beta_1^* = \beta_2^* = \cdots = \beta_s^* = 0$ in the model
  $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \beta_1^* X_1^* + \cdots + \beta_s^* X_s^* + E$

$$F(X_1^*, \ldots, X_s^* | X_1, \ldots, X_p) = \frac{MS(X_1^*, \ldots, X_s^* | X_1, \ldots, X_p)}{MS(X_1, \ldots, X_p, X_1^*, \ldots, X_s^*)} \sim F_{s, n-p-s-2}$$

9.4 Multiple partial F test II

Application: the simultaneous importance of particular subsets of a set of predictor variables.

Whether a chunk of variables (having some trait in common) is important when considered together.

9.5 Strategies

- Variables-add-in-odder tests
- Variables-add-last tests

9.6 Additional inference methods for multiple regression I

- Test involving the intercept: $H_0 : \beta_0 = 0$
- Confidence interval about regression coefficients:

$$\hat{\beta}^* \pm t_{n-k-1}(S_{\beta^*})$$

- Whether the mean value of Y equals to a hypothesized value at a given set of values for the independent variables. (9.8)
  - T test
  - CI
- Prediction of Y at a specific set of predictor values: PI
- Inference method for linear functions of regression coefficients: the joint effect of the predictors