

# 9 Statistical Inference in Multiple Regression

Applied regression analysis and other multivariate methods

Yi Zhou

January 29, 2016

**Applied Regression Analysis And Other Multivariable Methods**  
*Kleinbaum/ Kupper/ Nizam/ Rosenberg*

## 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:

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

## 9.1 Preview

- ▶ Full model
- ▶ Reduced model

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

## 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
- ▶  $H_0$ :  $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

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

## 9.2 Test for significant overall regression

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

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

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

## 9.3 Partial F test

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  $X_1, X_2, \dots, X_p$  are already in the model.
- ▶  $H_0$ :  $\beta^* = 0$  in the model

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \beta^* X^* + E$$

$$F(X^*|X_1, \dots, X_p) = \frac{SS(X^*|X_1, \dots, X_p)}{MS(X_1, \dots, X_p, X^*)} \sim F_{1, n-p-2} \quad (9.4)$$

$$SS(X^*|X_1, \dots, X_p)$$

= SS Regression (full model) - SS Regression (reduced model)

## 9.3 Partial F test

- ▶ Full model:  $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \beta^* X^* + E$
- ▶ Reduced model:  $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + E$

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

- ▶  $F(X_2|X_1)$ ,  $F(X_3|X_1, X_2)$ , and  $MS \text{ Residual}(X_1, X_2)$
- ▶  $T$  test alternative
- ▶ 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

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

- ▶  $H_0$ :  $X_1^*, X_2^*, \dots, X_s^*$  do not significantly add to the prediction of  $Y$  given that  $X_1, X_2, \dots, X_p$  are already in the model.

- ▶  $H_0$ :  $\beta_1^* = \beta_2^* = \dots = \beta_s^* = 0$  in the model

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \beta_1^* X_1^* + \dots + \beta_s^* X_s^* + E$$

$$F(X_1^*, \dots, X_s^* | X_1, \dots, X_p) = \frac{SS(X_1^*, \dots, X_s^* | X_1, \dots, X_p) / s}{MS(X_1, \dots, X_p, X_1^*, \dots, X_s^*)}$$

$$\sim F_{s, n-p-s-1} \quad (9.6)$$

## 9.4 Multiple partial F test

$$F(X_1^*, \dots, X_s^* | X_1, \dots, X_p) = \frac{MS(X_1^*, \dots, X_s^* | X_1, \dots, X_p)}{MS(X_1, \dots, X_p, X_1^*, \dots, X_s^*)}$$

$$\begin{aligned} & SS(X_1^*, \dots, X_s^* | X_1, \dots, X_p) \\ &= SS \text{ Reg}(\text{full model}) - SS \text{ Reg}(\text{reduced model}) \\ &= SSE(\text{reduced model}) - SSE(\text{full model}) \end{aligned}$$

- ▶ 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

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

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

- ▶ Whether the mean value of Y equals to a hypothesized value at a given set of values for the independent variables.
  - ▶ T test (9.8)
  - ▶ 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