Hypothesis Tests in Multiple Regression Analysis

Multiple regression model: \( Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_{p-1}X_{p-1} + \varepsilon \) where \( p \) represents the total number of variables in the model.

I. Testing for significance of the overall regression model.

Question of interest: Is the regression relation significant? Are one or more of the independent variables in the model useful in explaining variability in \( Y \) and/or predicting future values of \( Y \)?

Null Hypothesis: The initial assumption is that there is no relation, which is expressed as: \( H_0: \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0 \).

Alternative Hypothesis: At least one of the independent variables IS useful in explaining/predicting \( Y \), expressed as: \( H_1: \) At least one \( \beta_i \) is \( \neq 0 \).

Test Statistic: \( F = \frac{SSR \,(p - 1)}{SSE \,(n - p)} = \frac{MSR}{MSE} \) which is found on any regression printout.

Sampling Distribution: Under the null hypothesis the statistic follows an \( F \)-distribution with \( p - 1 \) and \( n - p \) degrees of freedom. Reject in the upper tail of this distribution.

Interpreting Results: If we reject \( H_0 \) we conclude that the relation is significant/does have explanatory or predictive power. If we fail to reject, we conclude that there isn't any evidence of explanatory power, which suggests that there is no point in using this model.

II. Testing for the Significance/Contribution of a Single Independent Variable in the Model

Question of interest: Suppose we have a significant multiple regression model. In this model, does a single independent variable of interest, say \( X_j \), contribute to explaining/predicting \( Y \)? Or, would the model be just as useful without the inclusion of this variable?

Null Hypothesis: The initial assumption is that the variable does not contribute in this model, which is expressed as: \( H_0: \beta_j = 0 \).

Alternative Hypothesis: The alternative is that the variable does contribute and should remain in the model: \( H_1: \beta_j \neq 0 \).

Test Statistic: \( t = \frac{b_j - 0}{s_{b_j}} \) which is found on any regression printout.
Sampling Distribution: Under the null hypothesis the statistic follows a \( t \)-distribution with \( n - p \) degrees of freedom. Reject in the upper or lower tail of this distribution.

Interpreting Results: If we reject \( H_0 \) we conclude that the independent variable \( X_j \) does have explanatory or predictive power in our model. Note that conclusion is \textit{model-specific}, in that it might change if the model included a different set of independent variables. If we fail to reject, we conclude that there isn't any evidence of explanatory power. That suggests that there is no point in having \( X_j \) \textit{in this model} and we should consider dropping it and re-running the regression analysis.

III. A General Test for the Value of \( \beta_j \)

Question of interest: Does \( \beta_j \) equal a specified value of interest, say \( \beta_j^* \)? Or, do we have evidence to state that it is not equal to \( \beta_j^* \)? (A two-sided test situation is assumed. Make the obvious adjustment for a one-sided test.)

Null Hypothesis: \( H_0: \beta_j = \beta_j^* \)

Alternative Hypothesis: \( H_1: \beta_j \neq \beta_j^* \)

Test Statistic: \( t = \frac{b_j - \beta_j^*}{s_{b_j}} \) which is NOT found on the regression printout. You will, however, find \( b_j \) and \( s_{b_j} \) on the printout.

Sampling Distribution: Under the null hypothesis the statistic follows a \( t \)-distribution with \( n - p \) degrees of freedom. Reject in the upper or lower tail of this distribution, making the appropriate adjustment for one-sided tests.

Interpreting Results: If we reject \( H_0 \) we conclude that we have evidence that \( \beta_j \) is not equal to the specified \( \beta_j^* \) value (we can refute the claim that \( \beta_j = \beta_j^* \)). Otherwise, we can't refute the claim.
IV. Testing for the significance/contribution of a subset of independent variables in the regression model.

Question of interest: In the multiple regression model:

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_{g-1} X_{g-1} + \beta_g X_g + \ldots + \beta_{p-1} X_{p-1} + \varepsilon$$ (full model)

does the subset of independent variables $X_g, \ldots, X_{p-1}$ contribute to explaining/predicting $Y$? Or, would be do just as well if these variables were dropped and we reduced the model to

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_g X_{g-1} + \varepsilon$$ (reduced model).

Null Hypothesis: The initial assumption is that the subset does not contribute to the model's explanatory power, which is expressed as: $H_0$: $\beta_g = \ldots = \beta_{p-1} = 0$.

Alternative Hypothesis: At least one of the independent variables in the subset IS useful in explaining/predicting $Y$, expressed as: $H_1$: At least one $\beta_i$ is $\neq 0$, $i = g$ to $p-1$.

Test Statistic: You need to run two regressions, one for the full model and one for the reduced model as described above. Then calculate:

$$F = \frac{(SSR_{Full} - SSR_{Reduced})/(p - g)}{MSE_{Full}} = \frac{\text{Change in SSR / Number Variables Dropped}}{MSE_{Full}}$$

Sampling Distribution: Under the null hypothesis the statistic follows an $F$-distribution with $p - g$ and $n - p$ degrees of freedom. Reject in the upper tail of this distribution.

Interpreting Results: If we reject $H_0$ we conclude that at least one independent variable in the subset $(X_g, \ldots, X_p)$ does have explanatory or predictive power, so we don't reduce the model by dropping out this subset. If we fail to reject, we conclude we have no evidence that inclusion of the subset of independent variables in the model contributes to explanatory power. This suggests that you may as well drop them out and re-run the regression using the reduced model.

Comment: If $p - g = 1$, i.e. if the subset consists of a single independent variable, then this $F$-test is equivalent to the two-sided $t$-test presented in Part II. In fact, $t^2 = F$. You might recall a similar result from simple regression analysis.