Multiple Linear Regression

A. Model and Estimation

1. Statistical model is $y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots \beta_p X_{pi} + e_i$
2. $X_{1i}, X_{2i} \ldots X_{pi}$ are unknown constants, and $\beta_0, \beta_1, \ldots \beta_p$ are known constants.
3. The e_i normally distributes with mean 0 and σ^2 variance.
4. In polynomial regression the model forms: $X_{1i} = X_i$, $X_{2i} = X_i^2$, ..., $X_{pi} = X_i^p$.
5. We can compare the importance of predictors by looking at their coefficients, for example, $y = 7 + 2*X1 + 5*X2$, which means X2 is more important than X1.
6. If we change the set of predictors, coefficients may change in magnitude but not in sign, because each predictor has a certain effect on the dependent variable positive or negative.

Result

1. True
2. False (Xs are known constants and βs are unknown constants
3. True
4. True
5. False (Relative importance of the predictors can't be determined from the size of their coefficients. Because coefficients depend on the scale of their predictors (X values), coefficients also depend on the correlations among predictors. (Multicollinearity)
6. False (Changing predictors can change the magnitude and sign of the coefficients.

B. TEST of FULL MODEL

1. When testing the full model, we want to accept $H_0: \beta_1 = \beta_2 = \ldots = \beta_p = 0$.
2. $E(MS_{res}) = \sigma^2$
3. If the null hypothesis above is true, then $E(MS_{reg}) = E(MS_{res})$ and F-obs is close to 1.

4. Degrees of freedom total = n-1, df of regression = p - 1, and df of residuals = n - p - 1.
5. When we sum the SS-regression and the SS-residuals, we get the SS-total.
6. The expectation of E(MSreg) = σ^2, if all the β's in the model are equal to zero.
7. If, H0: β1 = β2 = ... = βp = 0 then E(MSreg) > E(MSres) and F-obs > F-table.
8. When evaluating the quality of the predictions we want the adjusted R squared to be high.
9. When evaluating the quality of the model we want the coefficient of determination to be low.
10. When rejecting the H0, we want F-obs > F-table (p, n-p-1).

Result

1. False (We want to reject the H0 because we want at least one of them to be good at predicting y values.)
2. True
3. True
4. False (df total = n - 1, df regression = p, df residuals = n - p - 1)
5. True
6. True
7. False (To F-obs > F-table and reject the null hypothesis, at least one of the βs should be non-zero.)
8. True
9. False (The coefficient of determination is R^2 and we always want it to be large.
10. True

C. VARIABLE SELECTION

1- Partial F test:

1. We can draw conclusions based on the magnitude or sign of a regression coefficient.

2. A partial F test can be used to compare any two models as long as one is a sub-model of the other.

3. R squared is not always larger for the more complete model.

4. The F test can be interpreted as a test to see if R^2 is significantly larger.

5. We should apply F tests one predictor at a time to see the individual contributions of the predictors.

6. The expression $R(\beta 4, \beta 5 \mid \beta 1, \beta 2, \beta 3) = R(\beta 1, \beta 2, \beta 3, \beta 4, \beta 5) - R(\beta 1, \beta 2, \beta 3)$ represents the additional SSreg due to adding X4, and X5 to a model that already containing X1, X2, and X3.

7. The reduction sum of squares can be negative.

Result

1. False (Regression chefs (b's) may change for different models.
2. True
3. False (R^2 is always larger for a more complete model because SS reg/ SS res increases)
4. True
5. True
6. True
7. False (Reduction SS can never be negative because the full model has always higher SSreg.)

2- Variable selection with ordered predictors:

1. When we decide the order of predictors, we often consider factors such as logical priority, ease of measurement, cost of measurement, or practical importance.

2. In polynomial regression, higher-order terms should be included first when ordering the predictors.

3. The sequential SS is a way of partitioning SS reg.

4. When using Type I SS (sequential SS), the process starts by testing the highest-order term in the full model with an F-test. If the term is not significant, it is removed, and

the test is repeated. This continues until a significant term is found, and then we stop. The result becomes the simplest and most effective model.

Result

1. True
2. False (In polynomial regression, lower-order terms should be included first and higher-order terms such as $\beta 2\ X^2$, $\beta 3\ X^3$ should only be included if needed.)
3. True
4. True

3- Variable selection with unordered predictors:

1. Multicollinearity occurs when two or more predictor variables in a regression model are highly correlated with each other.

2. In the presence of multicollinearity, the individual contribution of each predictor variable to the model can be easily determined.

3. Orthogonal predictors are designed so that the correlation between them, $r(X1; X2)$ is zero.

4. The partial sum of squares (SS) tests the value of each predictor given all the other predictors currently in the model.

5. Multicollinearity doesn't affect the value of predictors in partial SS.

6. Since the partial SS tests the value of each predictor, the variable selection routines rely on tests of the partial sum of squares.

7. When there is high multicollinearity $R(\beta 1 \mid \beta 2, \ldots) = 0$ and $R(\beta 2 \mid \beta 1, \ldots) = 0$ happens.

Result

1. True
2. False (Multicollinearity prevents us from detecting the individual contributions of the predictors.)
3. True
4. True

5. False
6. True
7. True

4- Model selection logic:

1. R squared is ok for comparing models of the same size, but not adequate for models of different sizes.

2. When comparing models of different sizes, the one with a bigger R-squared has also bigger adjusted-R-squared all the time.

3. When comparing models we want a small adjusted-R-squared.

4. When comparing models we want a small MS res.

5. When comparing models we want a large Cp. (Cp = (RSS + 2p*sigma^2) / sigma^2)

6. MS res can't be used for comparing models at different sizes.

7. In underfitting models, regression coefficients (b's) are biased estimates of population coefficients (β's.)

8. The underfitting model has a high degree of multicollinearity due to including too many predictors.

9. The underfitting model has too few predictors and regression coefficients are biased estimates of population coefficients.

10. Overfitting models have wide CIs for β's and thus have wide prediction intervals.

Result

1. True
2. False (It's true when comparing models of the same size)
3. False (We always want large adjusted R squared)
4. True
5. False (When Cp is small, the model is neither overfit nor underfit. We want the Cp small.)
6. False (MS res can be used to compare models at different sizes like adjusted R squared.)

7. True
8. False (The overfitting model has multicollinearity because of having too many predictors)
9. True
10. True

5- Explain computer algorithms for forward selection:

1. Different model selection approaches will always yield the same terminal models.
2. Using alfa = 0.5 can be appropriate to ensure that any predictor which can contribute to the prediction of “y” has a chance to be included in the model.
3. In forward stepwise routines the model will be underfit in the first few steps, thus we should use a much less stringent alfa level to prevent the routine from stopping early with an underfit model.
4. In the underfit model MSres is too small, and Fobs is too large.
5. SS total is independent of the model and never changes.
6. Each time we change the model, we change SSreg, and SSres change in the opposite direction.
7. Sample size (n) in multiple linear regression should be greater than 4p.
8. In variable selection routines (n) should be greater than 7p.
9. If cross-validation is unsuccessful, we can try a model with fewer parameters.
10. The sum of squared errors shouldn't be too large in the checking sample than in the fitting sample.

Result

1. False (Different model selections usually yield different terminal models.)
2. True
3. True
4. False (MS res is too large, Fobs is too small)
5. True
6. True

7. False :) (3p)
8. True
9. True (If cross-validation is unsuccessful this means the model is overfitted, and we have to use fewer parameters.)
10. True

Bonus: Outliers

1. Leverage measures how unusual the X value is for an observation (not influenced by the Y value)
2. An observation considered unusual if: “ Leverage > 2 p / N ”.
3. Standardized residual measures how unusual the Y value is given the X value of the observation.
4. An observation considered unusual if: “ |standardized residual| > 2.0 “.
5. Cook’s D combines the information in leverage and standardized residual.
6. Cook’s D measured how different the estimated regression coefficients would be if this observation deleted.
7. An observation is unusual if: “ Cook’s D > F (p, n-p).50
8. We should remove the data point only trusting our analysis.

Result

1. True
2. True
3. True
4. True
5. True
6. True
7. True
8. False (We first have to look for a cause for the outliers before removing them.)