### 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 effect 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. True
8. False (Multicollinearity prevents us to detect the individual contributions of the predictors.)
9. True
10. True
11. False
12. True

### 4. Model selection logic:

#### True-False

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.
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:

#### True-False

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 must try a model with more 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. False (If cross-validation is unsuccessful this means the model is overfitted, and we have to use fewer parameters.)
10. True