Print

Quiz, Lesson 1: Multiple Linear Regression

Your Score: 100%

Congratulations! Your score of 100% indicates that you've mastered the topics in this lesson. If you'd like, you can review the feedback for each question.

When you're ready to start the next lesson, exit this lesson and begin the next one.

1. Multiple linear regression models are which of the following?

   a. linear in parameters and linear in variables
   b. linear in parameters but might not be linear in variables
   c. linear in variables but might not be linear in parameters
   d. models that might not be linear in both parameters and variables

Your answer: b
Correct answer: b

The regression model must be linear in parameters (a linear combination of parameters) although it might not be linear in variables.

---

2. Suppose you run a PROC GLMSELECT program to remove non-significant polynomial terms from a fourth-degree polynomial model using backward elimination and the significance level criterion. Which PROC GLMSELECT program below can produce the following output?

Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.442647	0.061884	39.47	<.0001
Amount^2	1	0.091622	0.018736	4.89	0.0001
Amount^3	1	-0.014623	0.003495	-4.18	0.0005

a.
```
proc glmselect data=sasuser.paper outdesign=d_paper;
   effect pamount=polynomial(amount / degree=4);
   model strength = p_amount / selection=backward select=sl slstay=0.05
         hierarchy=single showpvalues;
run;
```
b.
```
proc glmselect data=sasuser.paper outdesign=d_paper;
   effect p_amount=polynomial(amount / degree=4);
   model strength = p_amount / selection=none select=sl slstay=0.05
         showpvalues;
run;
```
c.
```
proc glmselect data=sasuser.paper outdesign=d_paper;
   effect p_amount=polynomial(amount / degree=4);
   model strength = p_amount / selection=none select=sl slstay=0.05
         hierarchy=single showpvalues;
run;
```

d.

```
proc glmselect data=sasuser.paper outdesign=d_paper;
   effect p_amount=polynomial(amount / degree=4);
   model strength = p_amount / selection=backward select=sl slstay=0.05
         hierarchy=none showpvalues;
run;
```

Your answer: d
Correct answer: d

To remove non-significant terms using backward elimination, you specify SELECTION=BACKWARD in the MODEL statement of PROC GLMSELECT. The HIERARCHY= option specifies whether or how the model hierarchy is applied. HIERARCHY=SINGLE specifies that only single terms can enter or leave the model at one time, following the principle of hierarchy. HIERARCHY=NONE specifies that the model hierarchy not be maintained. When HIERARCHY=NONE, any single effect can enter or leave the model at any given step of the selection process. The parameter estimate table shown above contains the **Amount^2** and **Amount^3** terms only but not the term **Amount**. This means that the model is not hierarchically well-formulated and has allowed any single effect to leave the model at any given step.

3. Suppose some observations in your design are replicated, and you obtain a significant test result for a lack-of fit test using PROC REG. Which of the following does this result indicate?

   a. The specified model is adequate.
   b. The specified model is inadequate.
   c. To perform this test, you should use PROC GLM instead of PROC REG.
   d. none of the above

Your answer: b
Correct answer: b

The test for lack of fit decomposes the residual error into two components. The first component is due to the variation of the replications around their mean value (the “pure” error). The second component is due to the variation of the mean values around the model prediction(the “bias” error). If the model is adequate, both components estimate the nominal level of error. However, if the bias component of error is much larger than the pure error, this constitutes evidence that there is significant lack of fit and the model is inadequate.

4. Which of the following programs requests collinearity diagnostic statistics in a multiple linear regression model?

a.

```
proc glmselect data=data1 plots=none;
   model Y=&_GLSMOD / vif collin collinoint;
run;
```

b.

```
proc glmselect data=data1 plots=all;
   model Y=&_GLSMOD / stats=all;
run;
```

c.

```
proc reg data=data1 plots(unpack) = (diagnostics (stats=all));
   model Y=&_GLSMOD /details;
run;
```

d.

```
proc reg data=data1;
   model Y=&_GLSMOD / vif collin collinoint;
run;
quit;
```

Your answer: d
Correct answer: d

Selected regression diagnostics (including collinearity diagnostics) are not yet available in PROC GLMSELECT, so PROC REG is used to perform these analyses. To request multicollinearity diagnostics like the variance inflation factor (VIF) and condition index values, you can specify the VIF, COLLIN, and COLLINOINT options in the MODEL statement of PROC REG.

---

5. Suppose you fit a polynomial regression model on the original variables, which generates the following ANOVA table. If you then center the variables by changing the location of data values (e.g., by subtracting the mean from each observation), what happens to the *F* value?

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	3	0.52986	0.17662	16.22	<.0001
Error	18	0.19605	0.01089		
Corrected Total	21	0.72591			

a. It increases.
b. It decreases.
c. It stays the same.
d. There is not enough information to answer this question.

Your answer: c
Correct answer: c

Changing only the location of data values by centering does not change the *F* value. The *F* value is calculated as the ratio of two mean sums of squares (variances), and the variance is not affected by changing the location of data values.

---

6. When you fit a polynomial regression model, which of the following do you need to do?

a. Use new estimation techniques because the OLS assumptions no longer apply.
b. Use the same techniques for estimation and inference as for multiple regression.
c. Use OLS estimation techniques, with the understanding that the *t*-statistics do not have an asymptotic normal distribution.
d. Change the critical values from the normal distribution to (1.96)2, (1.96)3, etc.

Your answer: b
Correct answer: b

A polynomial regression model is a special type of multiple linear regression. Therefore, the techniques for estimation and inference developed for multiple regression can be applied to polynomial regression models.

---

7. What does a residual plot display?

a. residuals of the explanatory variable versus residuals of the response variable
b. residuals of the explanatory variable versus the response variable
c. residuals versus predicted values of the response variable

d. predicted values versus the response variable

Your answer: c
Correct answer: c

To identify some common regression problems one can use residual plots: plotting the residuals versus the predicted values and versus the predictor variable, and doing a univariate analysis of the residuals.

---

8. Which of the following statements is false?

a. PROC REG supports a variety of model selection methods but not a CLASS statement.
b. PROC GLM supports a few model selection methods and a CLASS statement.
c. PROC GLMSELECT supports a variety of model selection methods and a CLASS statement.

Your answer: b
Correct answer: b

PROC GLM supports the CLASS statement but does not support any model selection methods. The statements about PROC REG and PROC GLMSELECT are true.

---

9. When you are calculating condition index values, what does a ratio of eigenvalues that has a very large magnitude indicate?

a. severe multicollinearity
b. little multicollinearity
c. the presence of influential observations
d. violation of regression assumptions

Your answer: a
Correct answer: a

The ratio of the eigenvalues can be useful for examining multicollinearity. Eigenvalues of relatively equal magnitudes result in a smaller ratio, which indicates that there is little multicollinearity. However, eigenvalues of widely varying magnitudes result in a large ratio, which indicates severe multicollinearity.

---

10. The assumption for GLMs that the errors are normally distributed is essential for which of the following?

a. estimation of model parameters
b. tests of significance of parameters
c. construction of confidence intervals of the parameters
d. both *b* and *c*

Your answer: d
Correct answer: d

The normality assumption is needed only for tests of significance and for constructing the confidence intervals of the parameters. The t-test, *F* test and chi-square test require the normality assumption of the residuals. The assumption that the errors are normally distributed is not necessary for estimation of the model parameters. The least squares estimates are still the best linear unbiased estimates (BLUE) if the other assumptions are met.

Close