Print

Quiz, Lesson 2: Regression Diagnostics and Remedial Measures

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. In least squares regression, which of the following is not a required assumption about the error term, ε?

   a. The expected value of the error term is 1.
   b. The variance of the error term is the same for all values of *x*.
   c. The values of the error term are independent.
   d. The error term is normally distributed.

   Your answer: a
   Correct answer: a

   The assumptions for linear regression are that the error terms are independent and normally distributed with mean of 0 and constant variance, σ^2, that is, $\varepsilon \sim$ iid $N(0,\sigma^2)$.

---

2. When the error terms exhibit constant variance, a plot of the residuals versus the independent variable *x* has a pattern that does which of the following?

   a. fans out
   b. funnels in
   c. shows random scatter pattern in a horizontal band around a reference line
   d. fans out but then funnels in

   Your answer: c
   Correct answer: c

   To check for constant variance, you can examine the plots of the residuals versus the predicted values and the plots of the residuals versus the independent variables. In the case of constant variance, these plots show no obvious trends or patterns and display a random scatter about the reference line.

---

3. If the Spearman rank correlation coefficient between the absolute value of the residuals and the predicted values is positive, then it implies which of the following?

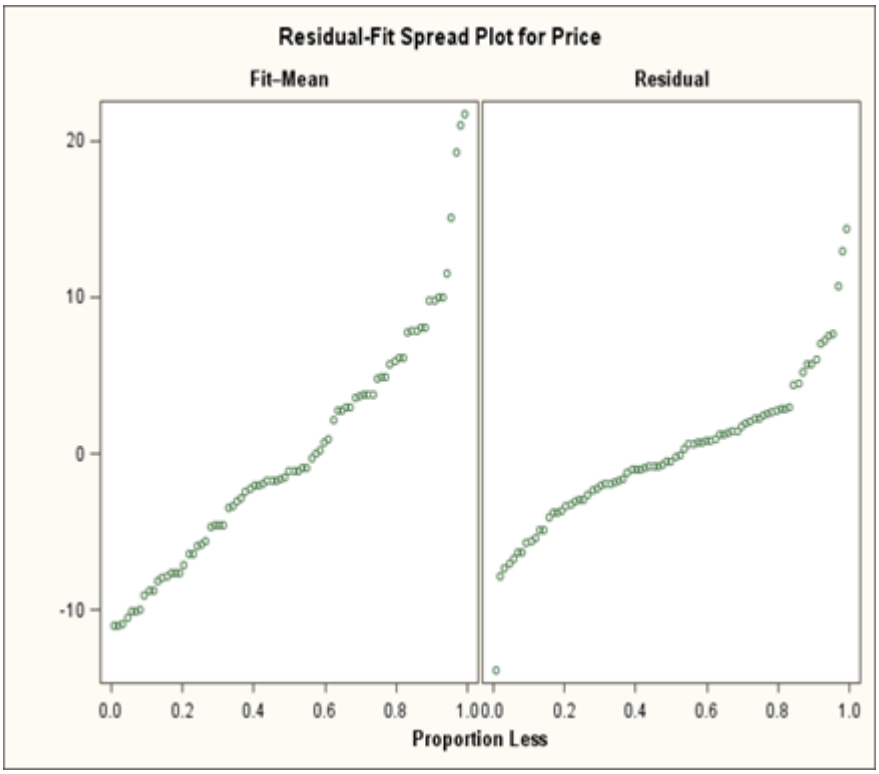
   a. The variances are approximately equal.
   b. The variance increases as the mean increases.
   c. The variance decreases as the mean increases.
   d. none of the above

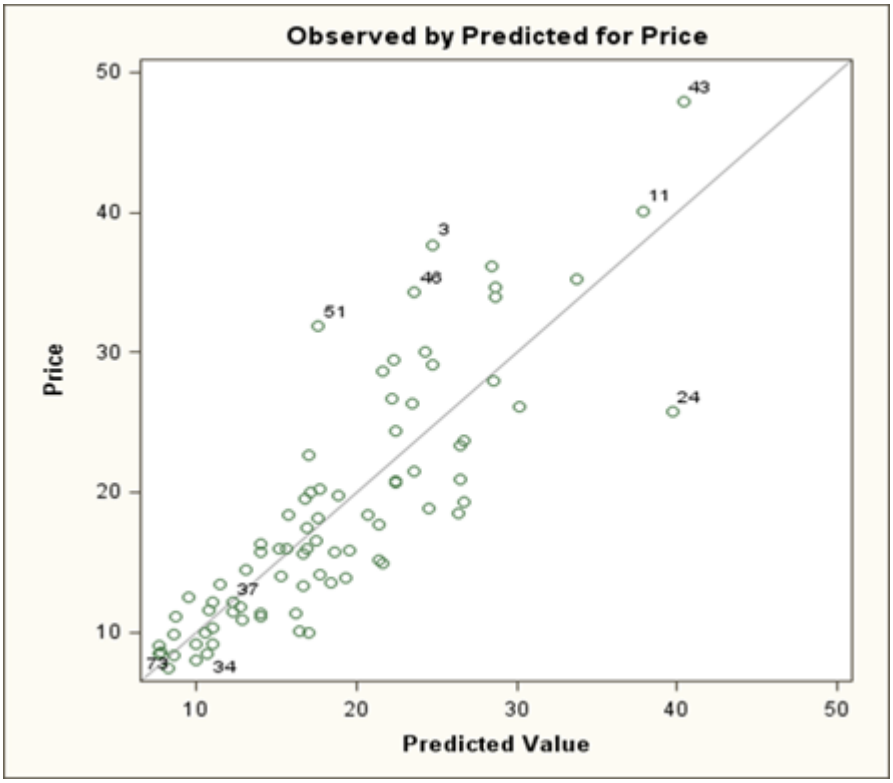
   Your answer: b
   Correct answer: b

To determine whether the constant variance assumption is met, you can compute the Spearman rank correlation coefficient between the absolute value of the residuals and the predicted values. If this quantity is close to zero, it means that there is no correlation between the size of the predicted value and the magnitude of the residual, which indicates that the variances are equal. A positive value suggests that the magnitude of the residual increases as the predicted values increase (that is, the variance increases as the mean increases). However, negative values indicate that the variance decreases as the mean increases.

---

4. The two plots (residual-fit spread plot and plot of observed values versus predicted values) below can be used for which of the following purposes?





a. to evaluate model assumptions
b. to identify influential observations
c. to identify multicollinearity
d. to evaluate model fit

Your answer: d
Correct answer: d

The residual-fit spread plot provides a visual summary of the amount of variability accounted for by a model. If the range of the fit-mean plot is substantially larger than the range of the residual plot, then the model explains most of the variability in the data (in other words, the model fits the data well). Similarly, plots of the observed values versus the predicted values serve as a visual tool for examining how close the fitted values are to the observed values.

5. Which of these programs request heteroscedasticity-consistent standard errors of the parameter estimates?

a.
```
proc reg data=data2;
   model price = s_hwympg s_hwympg_2 horsepower / hcc;
run;
quit;
```
b.
```
proc reg data=data2;
model price = s_hwympg s_hwympg_2 horsepower / hccmethod=3;
run;
quit;
```
c.
```
proc glmselect data=data2;
   model price = s_hwympg s_hwympg_2 horsepower  / stb;
run;
```
d.
```
proc glimmix data=data2;
   model price = s_hwympg s_hwympg_2 horsepower  / hcc;
run;
```

Your answer: a
Correct answer: a

An alternative way of dealing with nonconstant variance is to use heteroscedasticity-consistent standard errors. You can request heteroscedasticity-consistent standard errors of the parameter estimates with the HCC option in the MODEL statement in PROC REG.

6. Which of the following is not a plausible remedy for addressing multicollinearity among the variables?

a. Use principal component regression.
b. Drop redundant independent variables.
c. Take logarithms of each of the variables.
d. Redefine independent variables.

Your answer: c
Correct answer: c

Principal component regression combats multicollinearity by using less than the full set of principal components in the model. This technique works by transforming the original explanatory variables into a new set of explanatory variables that are constructed to be

orthogonal to one another. Another possible approach would be to drop one of the collinear predictor variables to mitigate the multicollinearity problem, although there might be other objections to doing this. Still another possible approach could be to redefine predictor variables. But taking logarithms of the variables will not help in remediating this multicollinearity issue.

7. Which of the following problems is Cook's D statistic used to identify?

   a. whether there is significant multicollinearity
   b. whether the overall regression model is significant
   c. whether there is significant first-order autocorrelation in the error terms
   d. influential observations in multiple regression analysis

Your answer: d
Correct answer: d

Cook's D statistic is used to identify influential observations. It is a measure of the simultaneous change in the parameter estimates when an observation is deleted from the analysis. An observation might have an adverse effect on the analysis if the Cook's D statistic is greater than 4/*n* (where *n* is the sample size).

8. Exponentiating the predicted values from a lognormal model produces unbiased estimates of which of the following?

   a. mean of the original data
   b. median of the original data
   c. geometric mean of the original data
   d. none of the above

Your answer: b
Correct answer: b

Back-transforming the predicted values from the lognormal model provides unbiased estimates of only the median rather than mean of the response variable on the original scale. To overcome this, you need to apply a low-bias adjustment factor (0.5 *σ^2, where σ^2 is the mean squared error from the regression model) to obtain low biased estimates of the means on the original scale.

9. Which of the following are the plausible approaches when the relationship between the dependent variable and one or more predictor variables does not follow a linear relationship?

   a. Transform the independent variables.
   b. Use biased regression techniques.
   c. Fit a nonparametric regression model.
   d. all of the above
   e. only a and c

Your answer: e
Correct answer: e

When the relationship between the dependent variable and one or more predictor variables does not follow a linear relationship, you might consider transforming the predictor variables to obtain the linearity. Also, when the parametric form of the relationship is difficult or impossible to define, you might want to fit a nonparametric regression model using PROC LOESS. Biased

regression techniques are used when there is multicollinearity.

10. When error terms are independent, which of the following characteristics do the plots of residuals versus time likely exhibit?

    a. a pattern of cyclical error terms over time
    b. a pattern of alternating error terms overtime
    c. a random pattern of error terms over time
    d. none of the above

    Your answer: c
    Correct answer: c

    You can plot residuals versus time or another ordering component to examine whether there seems to be any positive or negative autocorrelations. A pattern that is not random suggests a lack of independence.

Close
