Print

# Quiz: Model Post-Fitting for Inference

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.

When you're finished, exit the lesson.

✔ 1. What is an influential observation?

*a.* an observation so far away from the rest of the data that it influences the slope of the regression line
*b.* an observation with a large residual
*c.* both **a** and **b**
*d.* neither **a** nor **b**

**Your answer: a**
**Correct answer: a**

An influential observation is an unusual data point that has a large effect on some part of the model, such as the slope, the intercept, the standard errors, or the predicted values. It can sometimes have a large residual compared to the rest of the points, but it's an observation so far away from the rest of the data that it single-handedly exerts influence on the regression model.

---

✔ 2. Which of these programs requests diagnostic statistics as well as diagnostic plots?

*a.*
```
proc reg data=stat1.bodyfat2 plots(only)=
             (RSTUDENTBYPREDICTED(LABEL)
             COOKSD(LABEL)
             DFFITS(LABEL)
             DFBETAS(LABEL));
    PREDICT: model PctBodyFat2 = Abdomen Weight Wrist Forearm / vif;
    id Case;
run;
quit;
```

*b.*
```
proc reg data=stat1.bodyfat2 plots(only)=
             (QQ RESIDUALBYPREDICTED RESIDUALS);
    PREDICT: model PctBodyFat2 = Abdomen Weight Wrist Forearm;
    id Case;
run;
```

```
quit;
```

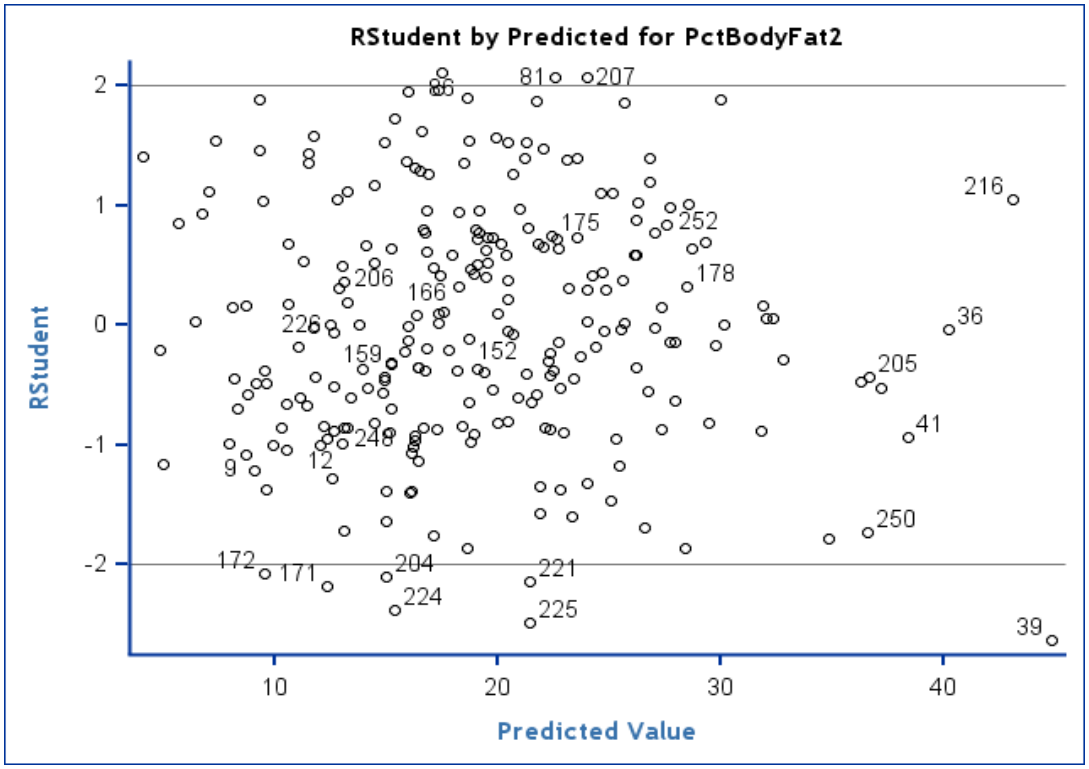
*c.* both **a** and **b**

**Your answer: a**
**Correct answer: a**

Program **a** specifies the VIF option, which requests diagnostic statistics to assess the magnitude of the collinearity problem.

---

3. Examine this plot of RSTUDENT residuals versus predicted values of **PctBodyFat2**. What does it indicate?




*a.* The model does not fit the data well.

*b.* The residuals have a cyclical shape, so the independence assumption is being violated.

*c.* Several observations exceed the cutoff values, so these observations might be influential.

*d.* All of the above.

**Your answer: c**
**Correct answer: c**

The gray horizontal lines mark the +2 and -2 cutoff values of the RSTUDENT residuals. Several observations fall outside these lines, so these observations might be influential.

---

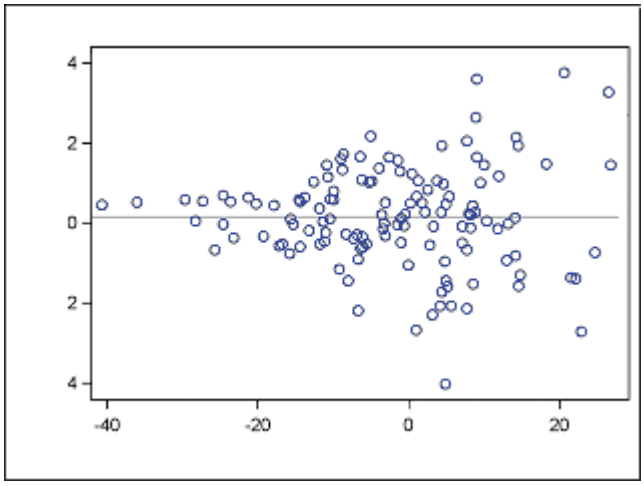
4. Which of the following can you use to detect outliers?

   *a.* DFFITS statistics
   *b.* Cook's *D* statistics
   *c.* STUDENT residuals
   *d.* RSTUDENT residuals

**Your answer: c**
**Correct answer: c**

You can use STUDENT residuals to detect outliers. To detect influential observations, you can use DFFITS and Cook's *D* statistics, and RSTUDENT residuals.

5. Suppose you have a residuals plot that shows a funnel shape for the residuals. Which assumption of linear regression is being violated?




   *a.* the linearity assumption
   *b.* the independence assumption
   *c.* both the linearity and the independence assumptions
   *d.* the equal variance assumption
   *e.* both the linearity and the equal variance assumptions

**Your answer: d**
**Correct answer: d**

When a residual plot displays a funnel shape, it indicates that the variance of the residual is not constant. That is, the variance increases toward the wide end of the "funnel." This shows you that your model violates the equal variance assumption.

6. Which program correctly saves statistics from the output statistics ODS output object to an output data set?

*a.*
```
proc reg data=stat1.bodyfat2;
    PREDICT: model PctBodyFat2 = Abdomen Weight Wrist Forearm / r influence;
    id Case PctBodyFat2 Abdomen Weight Wrist Forearm;
run;
quit;
```

*b.*
```
ods output RSTUDENTBYPREDICTED=Rstud
           COOKSDPLOT=Cook
           DFFITSPLOT=Dffits
           DFBETASPANEL=Dfbs;
proc reg data=stat1.bodyfat2 plots=none;
    PREDICT: model PctBodyFat2 = Abdomen Weight Wrist Forearm;
    id Case PctBodyFat2 Abdomen Weight Wrist Forearm;
title;
run;
quit;
```

*c.*
```
ods output RSTUDENTBYPREDICTED=Rstud
           COOKSDPLOT=Cook
           DFFITSPLOT=Dffits
           DFBETASPANEL=Dfbs;
proc reg data=stat1.bodyfat2 plots=
             (RSTUDENTBYPREDICTED(label)
              COOKSD(label)
              DFFITS(label)
              DFBETAS(label));
    PREDICT: model PctBodyFat2 = Abdomen Weight Wrist Forearm;
    id Case PctBodyFat2 Abdomen Weight Wrist Forearm;
title;
run;
quit;
```

*d.*
```
ods output outputstatistics;
proc reg data=stat1.bodyfat2 plots(only)=
             (RSTUDENTBYPREDICTED(LABEL)
              COOKSD(LABEL)
              DFFITS(LABEL)
              DFBETAS(LABEL));
    PREDICT: model PctBodyFat2 = Abdomen Weight Wrist Forearm;
    id Case PctBodyFat2 Abdomen Weight Wrist Forearm;
run;
```

**Your answer: c**
**Correct answer: c**

Program **b** is almost correct, but the images must be created for the data set to be saved. Program **c** creates the images and saves them into their own data set.

---

✔ 7. View this PROC REG output. What does the output indicate about the model?

*a.* The *p*-value for the overall model is not significant.

*b.* The model does not fit the data well.

*c.* The *p*-values for the parameter estimates indicate that collinearity is present in the model.

*d.* The variance inflation factors indicate that collinearity is present in the model.

*e.* None of the above.

**Your answer: d**
**Correct answer: d**

Several variance inflation factors are above 10. This indicates that collinearity among the predictor variables is present in the model.

---

8. Which of the following is suggested for developing good regression models?

*a.* getting to know your data by performing preliminary analyses

*b.* identifying good candidate models

*c.* checking and validating your assumptions using residual plots and other statistical tests

*d.* identifying any influential observations or collinearity

*e.* revising the model if needed

*f.* validating the model with data that's not used to build the model

*g.* all of the above

*h.* **a**, **c**, and **d** only

**Your answer: g**
**Correct answer: g**

All six steps are important for developing good regression models. You might need to perform some steps iteratively to produce the best possible model.

---

9. Collinearity decreases the variance of the parameter estimates, and increases the prediction error of the model.

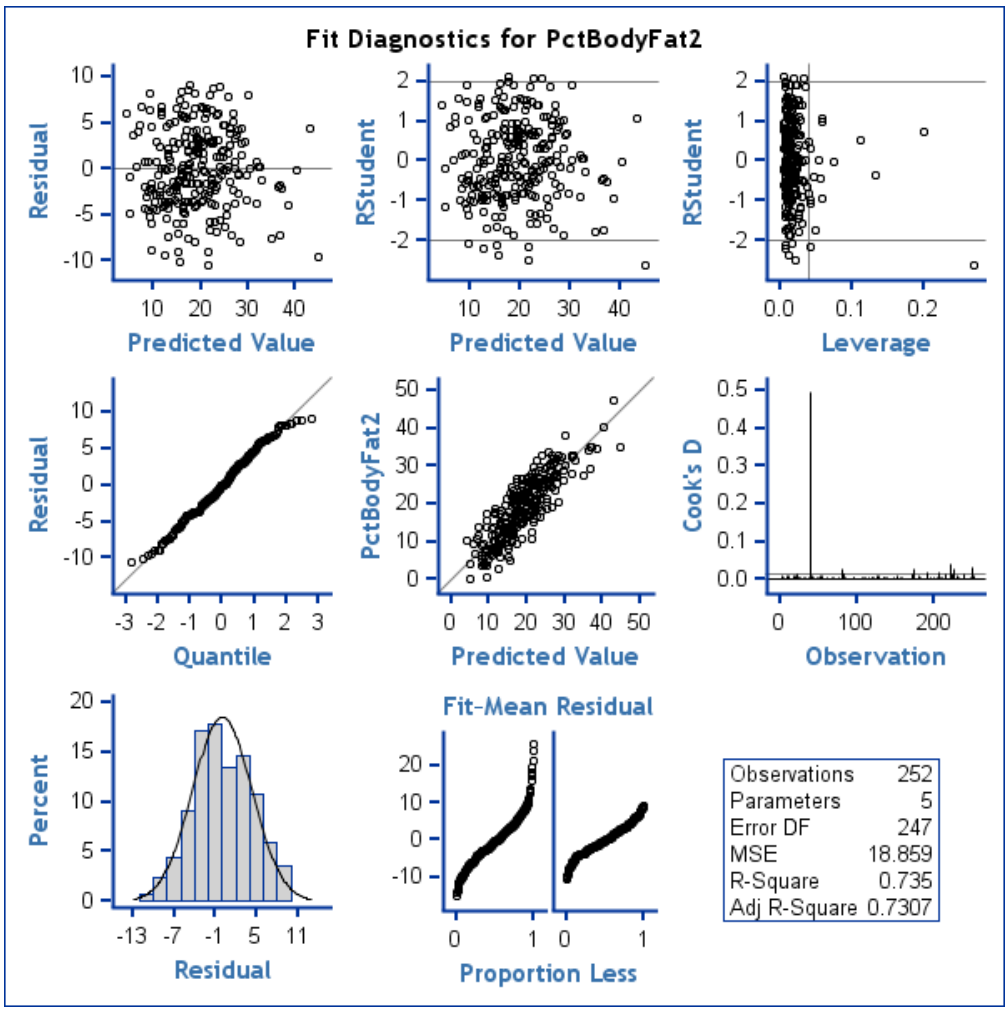
*a.* true

*b.* false

**Your answer: b**
**Correct answer: b**

Collinearity increases the variance of the parameter estimates, which also increases the prediction error of the model.

10. In the diagnostic plots below, what does the Residual versus Quantile plot indicate about the model?




*a.* that the errors are normally distributed

*b.* that the data set contains many influential observations

*c.* that the model is inadequate because the spread of the residuals is less than the spread of the centered fit

*d.* that the model is inadequate because patterns occur in the spread around the reference line

**Your answer: a**
**Correct answer: a**

The Residual versus Quantile plot is a normal quantile plot of the residuals. Using this plot, you can verify that the errors are normally distributed, which is one of your assumptions. Here, the residuals follow the normal reference line closely, so you can conclude that the errors are normally distributed.

Close