This is a required activity and will count toward course completion.

Take the Quiz Again

Attempt History

	Attempt	Time	Score
KEPT	**Attempt 2**	2 minutes	6 out of 6
LATEST	**Attempt 2**	2 minutes	6 out of 6
	Attempt 1	12 minutes	4 out of 6

Answers will be shown after your last attempt

Score for this attempt: **6** out of 6
Submitted May 3 at 12:02pm
This attempt took 2 minutes.

Question 1

1 / 1 pts

What are the two methods that can determine significance? **Select all that apply.**

- [x] If p value is < 0.05
- [ ] If the confidence interval contains a 0
- [x] If the confidence interval does not contain a 0
- [ ] If p value is > 0.05

That is correct! A *p* value of less than 0.05 and a confidence interval that does not contain 0 are equivalent.

Question 2

1 / 1 pts

Which is the correct confidence interval?

```
OLS Regression Results
==============================================================================
Dep. Variable:                  sales   R-squared:
0.890
Model:                            OLS   Adj. R-squared:
0.867
Method:                 Least Squares   F-statistic:
38.68
Date:                Mon, 27 Jun 2022   Prob (F-statistic):
1.01e-10
Time:                        17:18:56   Log-Likelihood:
-81.394
No. Observations:                  30   AIC:
174.8
Df Residuals:                      24   BIC:
183.2
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
const        -90.1059     18.349     -4.911      0.000    -127.976
-52.236
score          0.2059      0.029      7.041      0.000       0.146
0.266
age            6.0880      0.939      6.482      0.000       4.150
8.026
stress         0.3003      0.484      0.621      0.541      -0.698
1.299
experience     0.0766      0.807      0.095      0.925      -1.589
1.742
gpa           -0.4549      1.786     -0.255      0.801      -4.141
3.231
==============================================================================
Omnibus:                        6.287   Durbin-Watson:
1.613
Prob(Omnibus):                  0.043   Jarque-Bera (JB):
4.723
Skew:                           0.929   Prob(JB):
0.0943
Kurtosis:                       3.570   Cond. No.
1.42e+03
```

==========

- Age: [4.150, 0.266]
- Stress: [-0.697, 1.3]
- Score [−4.150, 8.026]
- GPA: [-4.141, 3.231]

That is correct! The confidence interval for GPA is [-4.141, 3.231].

Question 3

1 / 1 pts

What does it mean when we say a variable is significant for 95% confidence? **Select all that apply.**

- [x] We are 95% confident that the confidence interval contains the true population parameter.
- [ ] The confidence interval includes a 0.
- [x] The confidence interval does not contain a 0.
- [ ] We are 95% confident that the confidence interval does not contain the true population parameter.

That is correct! A *p* value of less than 0.05 is associated with 95% confidence. This is equivalent to a confidence interval that does not cross 0. The difference is that the confidence interval provides more information by showing the range in which the true value will occur.

Question 4

1 / 1 pts

A college is seeking to model cafeteria sales based on several student predictor variables. How can you identify possible overfitting using the available information?

Please note the regression output and correlation table below.(Assume the correlation is moderate to strong at or above 0.65.)

```
OLS Regression Results
==============================================================================
========
Dep. Variable:                   sales   R-squared:
0.890
Model:                             OLS   Adj. R-squared:
0.867
Method:                  Least Squares   F-statistic:
38.68
Date:                 Mon, 27 Jun 2022   Prob (F-statistic):
1.01e-10
Time:                         17:18:56   Log-Likelihood:
-81.394
No. Observations:                   30   AIC:
174.8
Df Residuals:                       24   BIC:
183.2
Df Model:                            5
Covariance Type:             nonrobust
==============================================================================
========
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
--------
const        -90.1059     18.349     -4.911      0.000    -127.976
-52.236
score          0.2059      0.029      7.041      0.000       0.146
0.266
age            6.0880      0.939      6.482      0.000       4.150
8.026
stress         0.3003      0.484      0.621      0.541      -0.698
1.299
experience     0.0766      0.807      0.095      0.925      -1.589
```

```
1.742
gpa           -0.4549      1.786     -0.255      0.801      -4.141
3.231
==============================================================================
========
Omnibus:                        6.287   Durbin-Watson:
1.613
Prob(Omnibus):                  0.043   Jarque-Bera (JB):
4.723
Skew:                           0.929   Prob(JB):
0.0943
Kurtosis:                       3.570   Cond. No.
1.42e+03
==============================================================================
========

Correlation Matrix
                sales           score           age             stress
experience      gpa
sales           1.000000        0.673719        0.788887        -0.233
564     0.549834        0.621784
score           0.673719        1.000000        0.216618        -0.147
833     0.343515        0.333466
age             0.788887        0.216618        1.000000        -0.286
808     0.541611        0.675425
stress          -0.233564       -0.147833       -0.286808       1.0000
00      -0.276569       -0.131167
experience      0.549834        0.343515        0.541611        -0.276
569     1.000000        0.312129
gpa             0.621784        0.333466        0.675425        -0.131
167     0.312129        1.000000
```

- Look for a p value greater than 0.05.
- Look for two variables correlated above 0.65.
- Look for a p value of less than 0.05.
- Look for two variables correlated below 0.65.

That is correct! Multicollinearity is a high correlation between independent variables. Two highly correlated independent variables may be measuring the same thing, dramatically altering the coefficients and p values of each variable and making them unreliable. This adds to the complexity of the model, which can lead to overfitting.

Question 5

1 / 1 pts

Using the data in the OLS regression results, how would you address overfitting in this case?

```
OLS Regression Results
==============================================================================
========
Dep. Variable:                   sales   R-squared:
0.890
Model:                             OLS   Adj. R-squared:
0.867
Method:                  Least Squares   F-statistic:
38.68
Date:                 Mon, 27 Jun 2022   Prob (F-statistic):
1.01e-10
Time:                         17:18:56   Log-Likelihood:
-81.394
No. Observations:                   30   AIC:
174.8
Df Residuals:                       24   BIC:
183.2
Df Model:                            5
Covariance Type:             nonrobust
==============================================================================
========
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
--------
const        -90.1059     18.349     -4.911      0.000    -127.976
-52.236
score          0.2059      0.029      7.041      0.000       0.146
0.266
age            6.0880      0.939      6.482      0.000       4.150
8.026
stress         0.3003      0.484      0.621      0.541      -0.698
1.299
experience     0.0766      0.807      0.095      0.925      -1.589
1.742
gpa           -0.4549      1.786     -0.255      0.801      -4.141
3.231
==============================================================================
========
Omnibus:                        6.287   Durbin-Watson:
1.613
Prob(Omnibus):                  0.043   Jarque-Bera (JB):
4.723
Skew:                           0.929   Prob(JB):
0.0943
Kurtosis:                       3.570   Cond. No.
1.42e+03
==============================================================================
========

Correlation Matrix
               sales           score           age             stress
experience      gpa
sales          1.000000        0.673719        0.788887        -0.233
```

```
564     0.549834        0.621784
score           0.673719        1.000000        0.216618        -0.147
833     0.343515        0.333466
age             0.788887        0.216618        1.000000        -0.286
808     0.541611        0.675425
stress          -0.233564       -0.147833       -0.286808       1.0000
00      -0.276569       -0.131167
experience      0.549834        0.343515        0.541611        -0.276
569     1.000000        0.312129
gpa             0.621784        0.333466        0.675425        -0.131
167     0.312129        1.000000
```

- (●) Drop one variable at a time and rerun the model each time until all of the coefficients are significant.
- ( ) Drop the one non-significant variable and rerun the model.
- ( ) Drop one variable at a time and rerun the model each time until most of the coefficients are significant.
- ( ) Drop the one non-significant variable, but do not rerun the model.

That is correct! To develop a final equation with which you can infer to the population and apply to unseen data, all of the coefficients must be significant (the *p* value is below 0.05 for a standard model). As all of the *p* values will change every time you drop a variable and rerun the model, this process must be repeated until all of the coefficients are significant.

Question 6

1 / 1 pts

A college is seeking to model cafeteria sales based on several student predictor variables. The following output from the latest run of a regression model has been handed to you for interpretation. Which of the variables is/are not statistically significant? Select all that apply.

```
OLS Regression Results
==============================================================================
========
Dep. Variable:                   sales   R-squared:
0.890
Model:                             OLS   Adj. R-squared:
0.867
Method:                  Least Squares   F-statistic:
38.68
Date:                 Mon, 27 Jun 2022   Prob (F-statistic):
1.01e-10
Time:                         17:18:56   Log-Likelihood:
-81.394
No. Observations:                   30   AIC:
174.8
Df Residuals:                       24   BIC:
183.2
Df Model:                            5
Covariance Type:             nonrobust
==============================================================================
========
                 coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
--------
const        -90.1059     18.349     -4.911      0.000    -127.976
-52.236
score          0.2059      0.029      7.041      0.000       0.146
0.266
age            6.0880      0.939      6.482      0.000       4.150
8.026
stress         0.3003      0.484      0.621      0.541      -0.698
1.299
experience     0.0766      0.807      0.095      0.925      -1.589
1.742
gpa           -0.4549      1.786     -0.255      0.801      -4.141
3.231
==============================================================================
========
Omnibus:                        6.287   Durbin-Watson:
1.613
Prob(Omnibus):                  0.043   Jarque-Bera (JB):
4.723
Skew:                           0.929   Prob(JB):
0.0943
Kurtosis:                       3.570   Cond. No.
1.42e+03
==============================================================================
========
```

- [x] GPA
- [ ] Age
- [x] Experience
- [ ] Score

- [x] Stress

That is correct! Stress, experience, and GPA contain a p-value greater than 0.05.

Quiz Score: **6** out of 6

Top Questions

It's all empty here!

If you have any questions ask one