Question 1
L2ET-PQ1002-1410
LOS: LOS-6020
LOS: LOS-6030
LOS: LOS-6040
LOS: LOS-6050
Lesson Reference: Lesson 1: Multiple Linear Regression
Difficulty: medium

An analyst obtained the following regression results:

	Coefficient	**Standard Error**	***t*-Statistic**
Intercept (b_0)	240.3256	48.5896	4.9460
b_1	1.3920	0.1970	7.0660
b_2	−3.6480	9.6853	−0.3767

ANOVA	**df**	**SS**	**MS**	***F***
Regression	2	3,940.2860		
Residual	97	4,268.1851		
Total	99	8,208.4711		

Regression equation: $Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + e_i$

The analyst wants to test whether there is a negative relationship between X_2 and Y. Given a 5% significance level, the hypotheses used and the result of the hypothesis test are *most likely*:

	Hypotheses	**Conclusion**
A	$H_0: b_2 \geq 0$, versus $H_a: b_2 < 0$	Reject the null hypothesis and conclude that there is a negative relationship between X_2 and Y.
B	$H_0: b_2 \leq 0$, versus $H_a: b_2 > 0$	Reject the null hypothesis and conclude that the relationship between X_2 and Y is not negative.
C	$H_0: b_2 \geq 0$, versus $H_a: b_2 < 0$	Fail to reject the null hypothesis and conclude that the relationship between X_2 and Y is not negative.

○ Row A
○ Row B
● Row C

Rationale

✓ **This Answer is Correct**

The null hypothesis is the position the analyst is looking to reject. The alternate hypothesis is the position he is looking to validate. Therefore:

H_0: $b_2 \geq 0$,

H_a: $b_2 < 0$

t-stat = −0.3767

This is a one-tailed test. Given a 5% significance level, the critical *t*-value with 97 degrees of freedom is −1.660. Since the test-stat is not less than the negative critical *t*-value, the analyst cannot reject the null hypothesis. He concludes that the relationship between X_2 and Y is not negative.

Question 2

L2R10TB-AC007-1512

LOS: LOS-6120

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

An analyst wants to test for heteroskedasticity in a regression equation and assess the effects of it on regression, if it exists. The analyst will *most likely* use a:

● Bruesch-Pagan test and will be concerned if conditional heteroskedasticity is found.

○ Bruesch-Pagan test and will be concerned if conditional or unconditional heteroskedasticity is found.

○ robust standard errors test and will be concerned if conditional or unconditional heteroskedasticity is found.

Rationale

✓ **Bruesch-Pagan test and will be concerned if conditional heteroskedasticity is found.**

Heteroskedasticity is the result of error terms differing across observations. The analyst should use a Bruesch-Pagan test to determine if the regression error terms are heteroskedastic. But, the analyst should only be concerned if conditional heteroskedasticity is found, which occurs when the error variances are conditional on the values of the independent variables. Conditional heteroskedasticity results in the overstatement of the regression's overall significance and the significance of the individual regression coefficients. In contrast, unconditional heteroskedasticity does not cause any problems with regression inference.

Rationale

✗ **Bruesch-Pagan test and will be concerned if conditional or unconditional heteroskedasticity is found.**

Heteroskedasticity is the result of error terms differing across observations. The analyst should use a Bruesch-Pagan test to determine if the regression error terms are heteroskedastic. But, the analyst should only be concerned if conditional heteroskedasticity is found, which occurs when the error variances are conditional on the values of the independent variables. Conditional heteroskedasticity results in the overstatement of the regression's overall significance and the significance of the individual regression coefficients. In contrast, unconditional heteroskedasticity does not cause any problems with regression inference.

Rationale

✗ **robust standard errors test and will be concerned if conditional or unconditional heteroskedasticity is found.**

Heteroskedasticity is the result of error terms differing across observations. The analyst should use a Bruesch-Pagan test to determine if the regression error terms are heteroskedastic. But, the analyst should only be concerned if conditional heteroskedasticity is found, which occurs when the error variances are conditional on the values of the independent variables. Conditional heteroskedasticity results in the overstatement of the regression's overall significance and the significance of the individual regression coefficients. In contrast, unconditional heteroskedasticity does not cause any problems with regression inference.

Question 3

L2QM-TB0007-1412

LOS: LOS-6130

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

Which of the following potential issues with regression models is *least likely* to be directly related to the residuals of the model?

○ Heteroskedasticity.

● Multicollinearity.

○ Serial correlation.

Rationale

✓ **This Answer is Correct**

Multicollinearity refers to a linear relationship existing between the independent variables of the regression. Heteroskedasticity relates to a nonconstant variance of the residual (error term), while serial correlation refers to a relationship between an error term and a lagged version of itself.

Question 4
L2ET-PQ1003-1410
LOS: LOS-6020
LOS: LOS-6030
LOS: LOS-6040
LOS: LOS-6050
Lesson Reference: Lesson 1: Multiple Linear Regression
Difficulty: medium

An analyst obtained the following regression results:

	Coefficient	**Standard Error**	***t*-Statistic**
Intercept (b_0)	240.3256	48.5896	4.9460
b_1	1.3920	0.1970	7.0660
b_2	−3.6480	9.6853	−0.3767

ANOVA	**df**	**SS**	**MS**	***F***
Regression	2	3,940.2860		
Residual	97	4,268.1851		
Total	99	8,208.4711		

Regression equation: $Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + e_i$.

The analyst wants to test whether b_0 is different from zero. Given a 5% significance level, the hypotheses used and the result of the hypothesis test are *most likely*:

	Hypotheses	**Conclusion**
A	$H_0 : b_0 = 0$, versus $H_a : b_0 \neq 0$	Reject the null hypothesis and conclude that b_0 is not equal to zero.
B	$H_0 : b_0 \neq 0$, versus $H_a : b_0 = 0$	Reject the null hypothesis and conclude that b_0 is equal to zero.
C	$H_0 : b_0 = 0$, versus $H_a : b_0 \neq 0$	Fail to reject the null hypothesis and conclude that b_0 is equal to zero.

● Row A
○ Row B
○ Row C

Rationale

✓ **This Answer is Correct**

The hypothesis test will be structured as:

$H_0 : b_0 = 0$, versus $H_a : b_0 \neq 0$

test-stat = 4.9460

This is a two-tailed test. Given a 5% significance level, the critical *t*-values with 97 degrees of freedom are +1.984 and −1.984. Since the test-stat is greater than the upper critical value, the analyst can reject the null hypothesis and conclude that b_0 is significantly different from zero.

Question 5

L2R10TB-AC019-1512

LOS: LOS-6120

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

The *least likely* implication of heteroskedasticity in a regression is that:

- ● regression parameters are inconsistent.
- ○ the *F*-test for the whole regression is unreliable.
- ○ the *t*-statistics for individual parameters in the regression are unreliable.

Rationale

✓ **regression parameters are inconsistent.**

Heteroskedasticity does not affect the consistency of the parameters but creates problems for statistical inference. When errors are heteroskedastic, the *F*-test for the overall significance of the regression and the *t*-tests for the significance of individual regression coefficients are unreliable, because heteroskedasticity introduces bias into estimators of the standard error of regression coefficients.

Rationale

☒ **the *F*-test for the whole regression is unreliable.**

Heteroskedasticity does not affect the consistency of the parameters but creates problems for statistical inference. When errors are heteroskedastic, the *F*-test for the overall significance of the regression and the *t*-tests for the significance of individual regression coefficients are unreliable, because heteroskedasticity introduces bias into estimators of the standard error of regression coefficients.

Rationale

☒ **the *t*-statistics for individual parameters in the regression are unreliable.**

Heteroskedasticity does not affect the consistency of the parameters but creates problems for statistical inference. When errors are heteroskedastic, the *F*-test for the overall significance of the regression and the *t*-tests for the significance of individual regression coefficients are unreliable, because heteroskedasticity introduces bias into estimators of the standard error of regression coefficients.

Question 6
L2ET-PQ1031-1410
LOS: LOS-6140
Lesson Reference: Lesson 4: Errors in Specification and Qualitative Dependent Variables
Difficulty: medium
Which of the following is the *least likely* effect of model misspecification on the results of regression analysis?

○ Estimates of the regression coefficients would be biased and inconsistent.

○ Standard errors of estimates of regression coefficients would be inconsistent.

● Hypothesis tests on regression coefficients would be reliable.

Rationale

✓ **This Answer is Correct**

Due to model misspecification, regression coefficients and their standard errors become useless for statistical analysis.

Question 7

L2R10TB-AC027-1512

LOS: LOS-6150

Lesson Reference: Lesson 4: Errors in Specification and Qualitative Dependent Variables

Difficulty: medium

A probit model is *most likely* based on a:

- ● normal distribution and uses qualitative dependent variables.
- ○ normal distribution and uses quantitative dependent variables.
- ○ nonnormal distribution and uses quantitative independent variables.

Rationale

✓ **normal distribution and uses qualitative dependent variables.**

Both probit and logit models are based on qualitative dependent variables. The probit model uses a normal distribution, and the logit model uses a logistic distribution.

Rationale

✗ **normal distribution and uses quantitative dependent variables.**

Both probit and logit models are based on qualitative dependent variables. The probit model uses a normal distribution, and the logit model uses a logistic distribution.

Rationale

✗ **nonnormal distribution and uses quantitative independent variables.**

Both probit and logit models are based on qualitative dependent variables. The probit model uses a normal distribution, and the logit model uses a logistic distribution.

Question 8

L2R10TB-AC023-1512

LOS: LOS-6120

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

Serial correlation in a regression equation *most likely* indicates that the:

● residuals are correlated with each other.

○ independent variables are correlated with each other.

○ dependent variable is correlated with the independent variables.

Rationale

✓ **residuals are correlated with each other.**

Serial correlation (also known as autocorrelation) relates to the residuals.

Rationale

✗ **independent variables are correlated with each other.**

Serial correlation (also known as autocorrelation) relates to the residuals.

Rationale

✗ **dependent variable is correlated with the independent variables.**

Serial correlation (also known as autocorrelation) relates to the residuals.

Question 9

L2ET-PQ1018-1410

LOS: LOS-6120

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

Which of the following can correct for heteroskedasticity and autocorrelation?

○ Modifying the regression equation

○ Using generalized least squares

● Using Hansen’s method

Rationale

✓ **This Answer is Correct**

Hansen’s method corrects for both heteroskedasticity and autocorrelation (serial correlation).

Question 10
L2R10TB-AC013-1512
LOS: LOS-6060
Lesson Reference: Lesson 1: Multiple Linear Regression
Difficulty: medium

An analyst has constructed a linear regression equation linking stock returns to a company's size and return on equity. The results of the regression are as follows.

	Coefficient	**Standard Error**	***t*-Statistic**
Intercept	0.03	0.009	3.3333
Size	−0.20	0.099	−2.0202
ROE	0.50	0.214	2.3364

The sample size is 30 and the sample R^2 is 0.42.

STUDENT'S *t*-DISTRIBUTION (ONE-TAILED PROBABILITIES)

df	***p* = 0.05**	***p* = 0.025**	***p* = 0.005**
26	1.706	2.056	2.779
27	1.703	2.052	2.771
28	1.701	2.048	2.763
29	1.699	2.045	2.756
30	1.696	2.042	2.750
31	1.694	2.040	2.744
32	1.692	2.037	2.738
33	1.691	2.035	2.733
34	1.690	2.032	2.728

The 95% confidence interval for the population value of the size coefficient is *closest to*:

○ −0.3140 to −0.3686
○ 0.0028 to −0.4028
● 0.0031 to −0.4031

Rationale
⊗ **−0.3140 to −0.3686**
First, find the critical statistic (*t*-critical) to use for the interval. Based on a 95% interval (two-tailed test, so 0.025 in each tail) and 27 degrees of freedom, the *t*-critical is 2.052. The interval is now calculated as follows:

$$95\% \text{ confidence interval for size coefficient} = -0.20 \pm 2.052 \times 0.099 = 0.00315 \text{ to } -0.4035$$

Rationale
⊗ **0.0028 to −0.4028**
First, find the critical statistic (*t*-critical) to use for the interval. Based on a 95 confidence interval (two-tailed test, so 0.025 in each tail) and 27 degrees of freedom, the *t*-critical is 2.052. The interval is now calculated as follows:

$$95\% \text{ confidence interval for size coefficient} = -0.20 \pm 2.052 \times 0.099 = 0.00315 \text{ to } -0.4035$$

Rationale

✓ **0.0031 to –0.4031**

First, find the critical statistic (*t*-critical) to use for the interval. Based on a 95 confidence interval (two-tailed test, so 0.025 in each tail) and 27 degrees of freedom, the *t*-critical is 2.052. The interval is now calculated as follows:

$$95\% \text{ confidence interval for size coefficient} = -0.20 \pm 2.052 \times 0.099 = 0.00315 \text{ to } -0.4035$$

Question 11
L2R10TB-AC016-1512
LOS: LOS-6090
Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables
Difficulty: medium

The adjusted R^2 for a multiple linear regression equation will *most likely* be:

- ● less than the R^2.
- ○ greater than the R^2.
- ○ less than or greater than the R^2.

Rationale

✓ **less than the R^2.**

The adjusted R^2 is penalized by the addition of additional independent variables. As a result, the adjusted R^2 will always be less than the R^2 when $k \geq 1$.

Rationale

✗ **greater than the R^2.**

The adjusted R^2 is penalized by the addition of additional independent variables. As a result, the adjusted R^2 will always be less than the R^2 when $k \geq 1$.

Rationale

✗ **less than or greater than the R^2.**

The adjusted R^2 is penalized by the addition of additional independent variables. As a result, the adjusted R^2 will always be less than the R^2 when $k \geq 1$.

Question 12

L2ET-PQ1017-1410

LOS: LOS-6120

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

Which of the following is the preferred approach to correct for heteroskedasticity?

- ● Using White-corrected standard errors
- ○ Using generalized least squares
- ○ Step-wise regression

Rationale

✓ **This Answer is Correct**

Using White-corrected standard errors (also known as robust standard errors) is the preferred approach to correct for heteroskedasticity.

Question 13
L2ET-PQ1032-1410
LOS: LOS-6120
Lesson Reference: Lesson 3: Violations of Regression Assumptions
Difficulty: medium

An analyst uses the following regression equation to predict the excess return on an asset:

$$\mathrm{r_{i,t}} = \alpha_\mathrm{i} + \beta_\mathrm{i}\mathrm{r_{m,t}} + \epsilon_\mathrm{i,t}$$

Where:

$r_{i,t}$ = Excess return (return above the risk-free rate) to portfolio i in period t

$r_{m,t}$ = Excess return to the market as a whole in period t

The initial results of this regression suggest that excess returns on an asset are indeed explained by excess returns of the market as a whole. Before reaching a conclusion, the analyst wants to test whether the errors in the regression model are normally distributed and homoskedastic. He therefore runs a regression of the errors from the first regression against excess returns of the market and attains a value of 0.1645 as the coefficient of determination. Given that 240 observations were used in the analysis, what conclusion can the analyst make regarding conditional heteroskedasticity at the 5% significance level?

○ The analyst should reject the null hypothesis. He should conclude that there is no conditional heteroskedasticity.

○ The analyst should not reject the null hypothesis. He should conclude that there is conditional heteroskedasticity.

● The analyst should reject the null hypothesis. He should conclude that there is conditional heteroskedasticity.

Rationale

✓ **This Answer is Correct**

The BP test for heteroskedasticity is a one-tailed chi-squared test.

Chi-squared test stat: $nR^2 = 240 \times 0.1645 = 39.48$

The critical value with 1 degree of freedom (k = 1) and 5% significance level is 3.8415.

The null hypothesis of no conditional heteroskedasticity is rejected at the 5% significance level, because the *t*-statistic (39.48) from the Breusch-Pagan test is greater than the critical χ^2 value. The analyst should conclude that the data does suffer from conditional heteroskedasticity.

Question 14
L2R10TB-ITEMSET-AC001-1512
LOS: LOS-6020
LOS: LOS-6080
LOS: LOS-6040
Lesson Reference: Lesson 1: Multiple Linear Regression
Difficulty: N/A

Use the following information to answer the next 3 questions:

Michal Neuvirth, CFA, wants to forecast sales of CanStick, a hockey stick made from Canadian ash wood that offers strength and yet flexibility in shooting techniques. Neuvirth has developed the sales regression model shown in Exhibit 1 and supporting data found in Exhibits 2 through 5 to assist in his Canadian sales forecast of CanStick.

Exhibit 1: CanStick Sales Regression Model

Regression Equation: SALES (CA $ millions) = $8.530 million + $6.078 (POP) + 5.330 (INC) + 7.380 (ADV)

where:

POP = Canadian population in millions

INC = Average operating income (in CA $ millions) for Canadian NHL teams

ADV = Advertising spent (in CA $ millions) by company marketing CanStick

	Coefficient	Standard Error	*t*-statistic
Intercept	8.530	3.4365	2.482
POP	6.078	2.7256	2.230
INC	5.330	2.5261	2.110
ADV	7.380	2.6836	2.750

$n = 20$

$R^2 = 0.370$

Adjusted $R^2 = 0.252$

Sum of the squared errors (SSE) = 214.2

Regression sum of squares (RSS) =125.8

Durbin-Watson (DW) = 1.66

Neuvirth has made the following 20X5 estimates for the independent variables:

POP = 34.7 million
INC = $27.4 million
ADV = $8.2 million

In Exhibits 2 and 3 below, Neuvirth has also reproduced portions of the Student's *t*-Distribution table of critical values and *F*-Distribution table of critical values.

Exhibit 2: Student's *t*-Distribution

	Area in Upper Tail		
df	0.100	0.050	0.025
16	1.3368	1.7459	2.1199
17	1.3334	1.7396	2.1098
18	1.3304	1.7341	2.1009

Area in Upper Tail

19	1.3277	1.7291	2.0930
20	1.3253	1.7247	2.0860

Exhibit 3: Critical Values for the *F*-Distribution (right-hand tail area = 0.05)

Numerator: df1 and Denominator: df2

df2	df1					
	1	2	3	4	5	6
15	4.54	3.68	3.29	3.06	2.90	2.79
16	4.49	3.63	3.24	3.01	2.85	2.74
17	4.45	3.59	3.20	2.96	2.81	2.70

i.

Using the regression, Neuvirth's forecast sales for CanStick for 20X5 is *closest to*:

○ $342 million.

○ $417 million.

● $426 million.

Rationale

✓ This Answer is Correct

The calculation is as follows:

$$\begin{aligned}\text{Sales} &= \$8.530\text{ million} + \$6.078(34.7\text{ million}) + 5.330(\$27.4\text{ million}) + 7.38(\$8.2\text{ million})\\ &= \$426\text{million}\end{aligned}$$

Rationale

✓ This Answer is Correct

The calculation is as follows:

$$\begin{aligned}\text{Sales} &= \$8.530\text{ million} + \$6.078(34.7\text{ million}) + 5.330(\$27.4\text{ million}) + 7.38(\$8.2\text{ million})\\ &= \$426\text{million}\end{aligned}$$

Rationale

✓ This Answer is Correct

The calculation is as follows:

$$\begin{aligned}\text{Sales} &= \$8.530\text{ million} + \$6.078(34.7\text{ million}) + 5.330(\$27.4\text{ million}) + 7.38(\$8.2\text{ million})\\ &= \$426\text{million}\end{aligned}$$

ii.

Which of the following is *closest to* the *F*-statistic for a test concerning whether all the independent variables taken simultaneously have a significant relationship with the dependent variable and what is the appropriate conclusion for the test at a significance level of 0.05?

○ The *F*-statistic is 3.33 and the null hypothesis should be rejected.

○ The *F*-statistic is 3.13 and the null hypothesis should be rejected.

● The *F*-statistic is 3.13 and the null hypothesis should not be rejected.

Rationale

✖ **This Answer is Incorrect**

The calculation of the *F*-statistic is:

$$F = \frac{\text{RSS}/k}{\text{SSE}/[n-(k+1)]} = \frac{\text{Mean regression sum of squares}}{\text{Mean square error}} = \frac{\text{MSR}}{\text{MSE}} = \frac{125.8/3}{214.2/[20-(3+1)]} = 3.13$$

For the *F*-test, the null hypothesis is that all of the coefficients are all simultaneously equal to zero. Using Exhibit 3, with df1 = k = 3 and df2 = $n - (k + 1)$ = 20 – (3 + 1) = 16, we find the *F*-critical value is 3.24. Since the regression's *F*-statistic of 3.13 is less than 3.24, we cannot reject the null hypothesis.

Rationale

✖ **This Answer is Incorrect**

The calculation of the *F*-statistic is:

$$F = \frac{\text{RSS}/k}{\text{SSE}/[n-(k+1)]} = \frac{\text{Mean regression sum of squares}}{\text{Mean square error}} = \frac{\text{MSR}}{\text{MSE}} = \frac{125.8/3}{214.2/[20-(3+1)]} = 3.13$$

For the *F*-test, the null hypothesis is that all of the coefficients are all simultaneously equal to zero. Using Exhibit 3, with df1 = k = 3 and df2 = $n - (k + 1)$ = 20 – (3 + 1) = 16, we find the *F*-critical value is 3.24. Since the regression's *F*-statistic of 3.13 is less than 3.24, we cannot reject the null hypothesis.

Rationale

✖ **This Answer is Incorrect**

The calculation of the *F*-statistic is:

$$F = \frac{\text{RSS}/k}{\text{SSE}/[n-(k+1)]} = \frac{\text{Mean regression sum of squares}}{\text{Mean square error}} = \frac{\text{MSR}}{\text{MSE}} = \frac{125.8/3}{214.2/[20-(3+1)]} = 3.13$$

For the *F*-test, the null hypothesis is that all of the coefficients are all simultaneously equal to zero. Using Exhibit 3, with df1 = k = 3 and df2 = $n - (k + 1)$ = 20 – (3 + 1) = 16, we find the *F*-critical value is 3.24. Since the regression's *F*-statistic of 3.13 is less than 3.24, we cannot reject the null hypothesis.

iii.

Which of the coefficients for the regression's independent variables are statistically different from zero at the 0.05 level of significance?

○ INC only.

● POP and ADV only.

○ POP, INC, and ADV.

Rationale

✖ **This Answer is Incorrect**

The first step is to determine the *t*-critical using the Student's *t*-distribution table provided in Exhibit 2. The number of degrees of freedom to use is found as follows:

df = Number of observations – (Number of independent variables + 1) = $n - (k + 1) = 20 - (3 + 1) = 16$

The test is two-tailed, since we are testing whether the coefficient is different from zero. Thus, we have 0.025 in each tail. Using the df and 0.025 in one tail, the critical value of the *t*-statistic is 2.1199.

A regression coefficient is statistically different from zero if the calculated *t*-statistic is greater than the *t*-critical value determined above. The calculated *t*-statistics for POP, INC, and ADV are 2.230, 2.110, and 2.750, respectively. Both POP and ADV have calculated *t*-statistics greater than 2.1199, so they are considered statistically different from zero at a 0.05 significance level.

Rationale

✖ **This Answer is Incorrect**

The first step is to determine the *t*-critical using the Student's *t*-distribution table provided in Exhibit 2. The number of degrees of freedom to use is found as follows:

df = Number of observations – (Number of independent variables + 1) = $n - (k + 1) = 20 - (3 + 1) = 16$

The test is two-tailed, since we are testing whether the coefficient is different from zero. Thus, we have 0.025 in each tail. Using the df and 0.025 in one tail, the critical value of the *t*-statistic is 2.1199.

A regression coefficient is statistically different from zero if the calculated *t*-statistic is greater than the *t*-critical value determined above. The calculated *t*-statistics for POP, INC, and ADV are 2.230, 2.110, and 2.750, respectively. Both POP and ADV have calculated *t*-statistics greater than 2.1199, so they are considered statistically different from zero at a 0.05 significance level.

Rationale

✖ **This Answer is Incorrect**

The first step is to determine the *t*-critical using the Student's *t*-distribution table provided in Exhibit 2. The number of degrees of freedom to use is found as follows:

df = Number of observations – (Number of independent variables + 1) = $n - (k + 1) = 20 - (3 + 1) = 16$

The test is two-tailed, since we are testing whether the coefficient is different from zero. Thus, we have 0.025 in each tail. Using the df and 0.025 in one tail, the critical value of the *t*-statistic is 2.1199.

A regression coefficient is statistically different from zero if the calculated *t*-statistic is greater than the *t*-critical value determined above. The calculated *t*-statistics for POP, INC, and ADV are 2.230, 2.110, and 2.750, respectively. Both POP and ADV have calculated *t*-statistics greater than 2.1199, so they are considered statistically different from zero at a 0.05 significance level.

Question 15

L2ET-PQ1030-1410

LOS: LOS-6130

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

Which of the following is *least likely* a symptom of multicollinearity?

- ● Low standard errors
- ○ High R^2
- ○ Low t-statistic

Rationale

✓ **This Answer is Correct**

A relatively high R^2, low t-statistics for the individual coefficients, and a high F-stat are the major symptoms of multicollinearity.

Low standard errors would result in high t-stat.

Question 16

L2R10TB-AC015-1512

LOS: LOS-6090

Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables

Difficulty: medium

When increasing the number of independent variables in a regression equation, the $R2$ will *most likely* be:

● unchanged or increased.

○ unchanged or decreased.

○ increased or decreased; it depends on its explanatory power.

Rationale

✓ **unchanged or increased.**

$R2$, which measures how much the change in the dependent variable is explained by the independent variables, will rise or stay the same when a new independent variable is added. The more independent variables added, the higher the $R2$, even if the latest independent variables have no explanatory power and are just capturing random noise. The solution is to use the adjusted $R2$, which does not automatically increase when another variable is added to a regression. In fact, it can decrease when a new independent variable is added. This is because it takes into account the explanatory power of the new independent variable weighed against the fact that the equation now has more variables.

Rationale

☒ **unchanged or decreased.**

$R2$, which measures how much the change in the dependent variable is explained by the independent variables, will rise or stay the same when a new independent variable is added. The more independent variables added, the higher the $R2$, even if the latest independent variables have no explanatory power and are just capturing random noise. The solution is to use the adjusted $R2$, which does not automatically increase when another variable is added to a regression. In fact, it can decrease when a new independent variable is added. This is because it takes into account the explanatory power of the new independent variable weighed against the fact that the equation now has more variables.

Rationale

☒ **increased or decreased; it depends on its explanatory power.**

$R2$, which measures how much the change in the dependent variable is explained by the independent variables, will rise or stay the same when a new independent variable is added. The more independent variables added, the higher the $R2$, even if the latest independent variables have no explanatory power and are just capturing random noise. The solution is to use the adjusted $R2$, which does not automatically increase when another variable is added to a regression. In fact, it can decrease when a new independent variable is added. This is because it takes into account the explanatory power of the new independent variable weighed against the fact that the equation now has more variables.

Question 17
L2ET-PQ1020-1410
LOS: LOS-6120
Lesson Reference: Lesson 3: Violations of Regression Assumptions
Difficulty: medium

Consider the following statements:

Statement 1: Negative serial correlation causes the *F*-stat to be inflated.

Statement 2: Positive serial correlation increases the chances of making Type-I errors when evaluating the significance of regression coefficients.

Which of the following is *most likely*?

○ Only Statement 1 is correct.
● Only Statement 2 is correct.
○ Both statements are correct.

Rationale

✓ **This Answer is Correct**

Negative serial correlation causes the *F*-stat to be **deflated**, because MSE will tend to overestimate the population error variance.

Statement 2 is correct. Positive serial correlation causes the standard errors for the regression coefficients to be underestimated, which results in larger *t*-values. Consequently, the analyst would reject some null hypotheses incorrectly, making Type-I errors.

Question 18

L2ET-PQ1016-1410
LOS: LOS-6120
Lesson Reference: Lesson 3: Violations of Regression Assumptions
Difficulty: medium

Consider the following statements:

Statement 1: Heteroskedasticity does not affect the consistency of the estimators of regression parameters.

Statement 2: Heteroskedasticity does not affect estimates of the regression coefficients.

Which of the following is *most likely*?

○ Only Statement 1 is correct.
○ Only Statement 2 is correct.
● Both statements are correct.

Rationale

✓ **This Answer is Correct**

Both statements are indeed correct.

Question 19

L2QM-TBX104-1502
LOS: LOS-6110
Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables
Difficulty: easy

An analyst wants to introduce dummy variables into his regression model to represent the different months of the year. How many dummy variables will the analyst need to introduce to the model?

● 11.
○ 12.
○ 13.

Rationale

✓ **This Answer is Correct**

In order to represent the 12 months of the year, the analyst will need to introduce a dummy variable to represent 11 of the months of the year. When all these dummy variables take a value of zero, this will represent the 12th month of the year.

Question 20
L2QM-TB0010-1412
LOS: LOS-6030
Lesson Reference: Lesson 1: Multiple Linear Regression
Difficulty: medium

An analyst derives the following output from a multiple linear regression when analyzing the relationship between wages, real GDP, and inflation.

	Name	**Estimate**	***p*-value**
Dependent variable	Wages	N/A	N/A
Independent variable 1	Real GDP	0.03	0.04
Independent variable 2	Inflation	0.88	0.08

Which of the variables is *most likely* to be statistically significant when tested at the 95% confidence level?

● Real GDP.
○ Inflation.
○ Both Real GDP and inflation.

Rationale
✓ **This Answer is Correct**
A coefficient that is statistically significant at the 95% confidence level will have a *p*-value that is lower than 0.05. This is true for real GDP but not inflation.

Question 21

L2R10TB-AC018-1512

LOS: LOS-6110

Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables

Difficulty: medium

When the errors in a regression are heteroskedastic, it *most likely* indicates that the:

○ errors are not normally distributed.

○ errors are correlated across observations.

● variance of the error term is not constant across observations.

Rationale

✗ **errors are not normally distributed.**

Linear regression assumes homoskedasticity, which is when the variance of the error term is constant across observations. When they are not constant across observations, then the errors are heteroskedastic.

Rationale

✗ **errors are correlated across observations.**

Linear regression assumes homoskedasticity, which is when the variance of the error term is constant across observations. When they are not constant across observations, then the errors are heteroskedastic.

Rationale

✓ **variance of the error term is not constant across observations.**

Linear regression assumes homoskedasticity, which is when the variance of the error term is constant across observations. When they are not constant across observations, then the errors are heteroskedastic.

Question 22

L2R10TB-AC025-1512

LOS: LOS-6130

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

When a regression has a high R^2 and the individual coefficients are all insignificant, it is *most likely* that the regression is exhibiting:

● multicollinearity.

○ serial correlation.

○ heteroskedasticity.

Rationale

✓ **multicollinearity.**

Multicollinearity is indicated by a high R^2 and beta coefficients that taken individually are insignificant.

Rationale

✗ **serial correlation.**

Multicollinearity is indicated by a high R^2 and beta coefficients that taken individually are insignificant.

Rationale

✗ **heteroskedasticity.**

Multicollinearity is indicated by a high R^2 and beta coefficients that taken individually are insignificant.

Question 23
L2ET-PQ1036-1410
LOS: LOS-6060
Lesson Reference: Lesson 1: Multiple Linear Regression
Difficulty: medium

An analyst formulates the following regression equation:

$Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + e_i$

Y_i = Natural logarithm of the ratio of the bid-ask spread-to-stock price for stock i

X_{1i} = Natural logarithm of the number of DJIA market makers for stock i

X_{2i} = Natural logarithm of the market capitalization (in millions of dollars) of company i

He obtains the following regression results:

	Coefficient	**Standard Error**	***t*-Statistic**
Intercept	−0.8247	0.1589	−5.1901
ln (Number of DJIA market makers)	−0.3120	0.0725	−4.3034
ln (Company’s market capitalization)	−0.6261	0.0396	−15.8106

Given that XYZ stock has 6 DJIA market makers and that the market capitalization of the company is $150 million, the ratio of the company’s bid-ask spread to its stock price is *closest to*:

○ 5.73%
○ −4.52%
● 1.09%

Rationale
✓ This Answer is Correct

Natural log of the number of DJIA market makers = ln 6 = 1.7918

Natural log of the company’s market capitalization (in millions of dollars) = ln 150 = 5.0106

$Y_i = -0.8247 + (-0.3120 \times 1.7918) + (-0.6261 \times 5.0106) = -4.5209$

Ratio of the bid-ask spread to the stock price = $e^{-4.5209}$ = 0.0109 or 1.09%

Question 24
L2QM-TBX103-1502
LOS: LOS-6090
Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables
Difficulty: easy

You have performed a multiple linear regression with two independent variables. The R^2 of the model is 0.276, and the adjusted R^2 is 0.243. You are considering adding a third independent variable to your model and wish to recalculate the adjusted R^2 with the new variable. Which of the following values is *least likely* to be observed for the adjusted R^2 with the third variable added?

○ –0.011
○ 0.251
● 0.281

Rationale

✓ **This Answer is Correct**

The adjusted R^2 can increase or decrease as new variables are added to the model, and can even turn negative. However, the adjusted R^2 cannot be greater than the original unadjusted R^2 of the model.

Question 25
L2ET-PQ1007-1410
LOS: LOS-6090
Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables
Difficulty: medium

An analyst obtained the following regression results:

	Coefficient	**Standard Error**	***t*-Statistic**
Intercept (b_0)	240.3256	48.5896	4.9460
b_1	1.3920	0.1970	7.0660
b_2	−3.6480	9.6853	−0.3767

ANOVA	**df**	**SS**	**MS**	***F***
Regression	2	3,940.2860		
Residual	97	4,268.1851		
Total	99	8,208.4711		

Regression equation: $Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + e_i$

The analyst adds another independent variable, X_3, to the regression model. As a result, RSS increases to 3,972.2967. The slope coefficient of X_3 is significantly different from zero at the 5% significance level. Should the third independent variable be added to the analysis?

○ Yes, because it increases R^2.
○ Yes, because its slope coefficient is significantly different from zero.
● No, because adjusted R^2 falls when the third independent variable is added.

Rationale
✓ **This Answer is Correct**

Before adding the third independent variable:

R^2 = RSS / SST = 3940.2860 / 8208.4711 = 0.480027 or 48.0027%

Adjusted R^2 = 1 − (n − 1 / n − k − 1) × (1 − R^2)]

Adjusted R^2 = 1 − (100 − 1 / 100 − 2 − 1) × (1 − 0.480027)]

Adjusted R^2 = 0.469306 or 46.9306%

After adding the third independent variable:

R^2 = RSS / SST = 3972.2967 / 8208.4711 = 0.483927 or 48.3927%

Adjusted R^2 = 1 − (100 − 1 / 100 − 3 − 1) × (1 − 0.483927)]

Adjusted R^2 = 0.467799 or 46.7799%

Since the adjusted R^2 after adding X_3 (46.78%) is less than the adjusted R^2 of the two-independent-variable regression (46.93%), it should not be added to the analysis.

Question 26
L2R10TB-AC009-1512
LOS: LOS-6130
Lesson Reference: Lesson 3: Violations of Regression Assumptions
Difficulty: medium

An analyst has examined the correlations among the independent variables as pairs (pairwise correlations) of a regression and found them to be low. The analyst can *most likely*:

○ be certain that there is multicollinearity.

○ be certain that there is no multicollinearity.

● be unsure whether there is multicollinearity.

Rationale

✗ **be certain that there is multicollinearity.**

While often used, pairwise correlations are insufficient in identifying the presence of multicollinearity. The analyst should examine whether there is a high R^2 and significant F-statistic even though the t-statistics on the estimated slope coefficients are not significant.

Rationale

✗ **be certain that there is no multicollinearity.**

While often used, pairwise correlations are insufficient in identifying the presence of multicollinearity. The analyst should examine whether there is a high R^2 and significant F-statistic even though the t-statistics on the estimated slope coefficients are not significant.

Rationale

✓ **be unsure whether there is multicollinearity.**

While often used, pairwise correlations are insufficient in identifying the presence of multicollinearity. The analyst should examine whether there is a high R^2 and significant F-statistic even though the t-statistics on the estimated slope coefficients are not significant.

Question 27
L2ET-ITEMSET-PQ1033-1410
LOS: LOS-6090
Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables
Difficulty: medium

Use the following information to answer the next three questions:

An analyst performs a regression with two independent variables and attains an R^2 of R_0^2 and an adjusted R^2 of adjusted R_0^2. He then adds a third independent variable to the analysis and obtains an R^2 of R_1^2 and an adjusted R^2 of adjusted R_1^2.

i.
Which of the following is *least likely*?

● $R_1^2 < R_0^2$

○ $R_1^2 = R_0^2$

○ $R_1^2 > R_0^2$

Rationale
✓ **This Answer is Correct**

If the new (third) variable has explanatory power, RSS will increase relative to SST and R_1^2 will be greater than R_0^2.

If the new (third) variable has zero explanatory power, RSS will remain the same and R_1^2 will equal R_0^2.

It is not possible for R_1^2 to be less than R_0^2, as the R^2 of the second regression would at least account for the variation in the dependent variable that is explained by the two independent variables.

ii.
Which of the following is *most likely*?

○ $R_0^2 <$ adjusted R_0^2 and $R_1^2 <$ adjusted R_1^2

○ $R_0^2 >$ adjusted R_0^2 and $R_1^2 <$ adjusted R_1^2

● $R_0^2 >$ adjusted R_0^2 and $R_1^2 >$ adjusted R_1^2

Rationale
✓ **This Answer is Correct**
If k is greater than 1, R^2 will always be greater than adjusted R^2.

iii.
Which of the following is *most likely*?

● Adjusted R_0^2 can be greater than, less than, or equal to adjusted R_1^2.

○ Adjusted R_0^2 can only be less than or equal to adjusted R_1^2.

○ Adjusted R_0^2 can only be greater than or equal to adjusted R_1^2.

Rationale

✓ **This Answer is Correct**

Adjusted $R_0{}^2$ can be greater than, less than, or equal to adjusted $R_1{}^2$.

- If the third independent variable has significant explanatory power, adjusted $R_1{}^2$ will be greater than adjusted $R_0{}^2$.
- If the third independent variable has zero explanatory power, adjusted $R_1{}^2$ will be less than adjusted $R_0{}^2$.
- It is mathematically possible that adjusted R^2 remains the same after incorporating an additional independent variable into the analysis.

Question 28

L2R10TB-AC008-1512

LOS: LOS-6120

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

An analyst has performed a Durbin-Watson test and the results show a Durbin-Watson statistic of 0. The analyst can *most likely* conclude that:

○ no serial correlation is present.

● positive serial correlation is present.

○ negative serial correlation is present.

Rationale

✗ **no serial correlation is present.**

The closer the Durbin-Watson statistic is to 0, the greater the positive serial correlation. If the DW statistic is 0, then the serial correlation is 1.

Rationale

✓ **positive serial correlation is present.**

The closer the Durbin-Watson statistic is to 0, the greater the positive serial correlation. If the DW statistic is 0, then the serial correlation is 1.

Rationale

✗ **negative serial correlation is present.**

The closer the Durbin-Watson statistic is to 0, the greater the positive serial correlation. If the DW statistic is 0, then the serial correlation is 1.

Question 29

L2R10TB-AC024-1512

LOS: LOS-6130

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

The *least likely* impact of multicollinearity is that the:

● R^2 of the regression will be understated.

○ standard errors of the coefficients will be inflated.

○ estimates of the regression coefficients become unreliable.

Rationale

✓ **R^2 of the regression will be understated.**

Multicollinearity deals with correlations between the independent variables and results in imprecise estimates of the coefficients and inflated standard errors.

Rationale

☒ **standard errors of the coefficients will be inflated.**

Multicollinearity deals with correlations between the independent variables and results in imprecise estimates of the coefficients and inflated standard errors.

Rationale

☒ **estimates of the regression coefficients become unreliable.**

Multicollinearity deals with correlations between the independent variables and results in imprecise estimates of the coefficients and inflated standard errors.

Question 30

L2ET-PQ1013-1410

LOS: LOS-6150

Lesson Reference: Lesson 4: Errors in Specification and Qualitative Dependent Variables

Difficulty: medium

An analyst wants to predict whether a company will go bankrupt based on its debt-to-equity ratio and its interest-coverage ratio. Which of the following models should *least likely* be used for this analysis?

○ Discriminant analysis

● Multiple regression with dummy variables

○ Probit model

Rationale

✓ **This Answer is Correct**

Since the test has a qualitative dependent variable, multiple regression is not the right technique to estimate it. The appropriate models are probit or logit models and discriminant analysis.

Question 31

L2R10TB-ITEMSET-AC004-1512
LOS: LOS-6090
LOS: LOS-6120
Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables
Difficulty: N/A

Use the following information to answer the next 3 questions:

Steve Ott works as a junior analyst for Convexity Associates, a quantitative modeling firm. Ott is required to review the output from regressing the returns of AsiaBus on the return on the bond market index (BOND), the return on the HSE Index (HSE), and the appreciation of Yen currency against the U.S. dollar (YEN CURR).

Selected data from the regression output and other additional data are contained in Exhibits 1 through 4 below.

Exhibit 1: Regression Analysis

	Coefficient	Standard Error	*t*-statistic	*p*-Value
Intercept	5.456600	2.476750	2.203129	0.0426
BOND	0.091658	0.168159	0.545067	0.5932
HSE	0.549225	0.143887	3.817058	0.0015
YEN CURR	−0.387119	0.139876	−2.767587	0.0137

Exhibit 2: Analysis of Variance (ANOVA)

	Sum of Squares	Mean Square	*F*	Significance *F*
Regression	1756.30	585.4333	8.5900	0.0013
Residual	1090.45	68.1531		
Total	2846.75			

Exhibit 3: Other Regression Data

R^2 (unadjusted)	0.61695
Standard error of estimate	8.25548
Durbin-Watson	0.13910
Number of observations	20

Exhibit 4: Correlation Matrix

	BOND	HSE	YEN CURR
BOND	1.000		
HSE	0.327	1.000	
YEN CURR	−0.139	0.039	1.000

i.
Which of the following statements is *most accurate* if Ott increases the number of independent variables in the model?

● The unadjusted R^2 for the new model could increase, but it cannot decrease.

○ The new slope coefficients will definitely be statistically significant if the unadjusted R^2 increases for the new model.

○ The adjusted R^2 takes into account the number of independent variables in the new model and will be equal to or greater than the unadjusted R^2.

Rationale

✓ **This Answer is Correct**

The inclusion of an additional variable can never reduce the unadjusted R^2 (at worst, it can be totally irrelevant). The unadjusted R^2 can increase if the additional independent variable explains only a very small amount of the previously unexplained variation of the model, even though its coefficient may be statistically insignificant. The adjusted R^2 cannot be greater than the unadjusted R^2 and adjusted R^2 may increase as new independent variables are introduced, if those variables have high explanatory power.

Rationale

✓ **This Answer is Correct**

The inclusion of an additional variable can never reduce the unadjusted R^2 (at worst, it can be totally irrelevant). The unadjusted R^2 can increase if the additional independent variable explains only a very small amount of the previously unexplained variation of the model, even though its coefficient may be statistically insignificant. The adjusted R^2 cannot be greater than the unadjusted R^2 and adjusted R^2 may increase as new independent variables are introduced, if those variables have high explanatory power.

Rationale

✓ **This Answer is Correct**

The inclusion of an additional variable can never reduce the unadjusted R^2 (at worst, it can be totally irrelevant). The unadjusted R^2 can increase if the additional independent variable explains only a very small amount of the previously unexplained variation of the model, even though its coefficient may be statistically insignificant. The adjusted R^2 cannot be greater than the unadjusted R^2 and adjusted R^2 may increase as new independent variables are introduced, if those variables have high explanatory power.

ii.

The regression data that Ott is reviewing *most likely* suggests the presence of:

○ multicollinearity.

● serial correlation.

○ an omitted important variable.

Rationale

✗ **This Answer is Incorrect**

The very low Durbin-Watson is most likely indicating positive serial correlation. While the F-statistic is significant, two of the three independent variables also appear to be significant. Therefore, we cannot conclude that serial correlation is present. Finally, there is no way to know whether an important variable has been omitted.

Rationale

✗ **This Answer is Incorrect**

The very low Durbin-Watson is most likely indicating positive serial correlation. While the F-statistic is significant, two of the three independent variables also appear to be significant. Therefore, we cannot conclude that serial correlation is present. Finally, there is no way to know whether an important variable has been omitted.

Rationale

✗ **This Answer is Incorrect**

The very low Durbin-Watson is most likely indicating positive serial correlation. While the F-statistic is significant, two of the three independent variables also appear to be significant. Therefore, we cannot conclude that serial

correlation is present. Finally, there is no way to know whether an important variable has been omitted.

iii.

If Ott is worried about the presence of heteroskedasticity, he *most likely* should perform a:

○ Hanson test to diagnose conditional heteroskedasticity.

○ Hanson test to diagnose unconditional heteroskedasticity.

● Bruesch-Pagan test to diagnose conditional heteroskedasticity.

Rationale

✖ **This Answer is Incorrect**

Two types of heteroskedasticity can exist: unconditional and conditional. Unconditional heteroskedasticity is not a major problem for statistical inference, but conditional heteroskedasticity does cause problems for statistical inference. So he should use a Bruesch-Pagan test to diagnose conditional heteroskedasticity.

Rationale

✖ **This Answer is Incorrect**

Two types of heteroskedasticity can exist: unconditional and conditional. Unconditional heteroskedasticity is not a major problem for statistical inference, but conditional heteroskedasticity does cause problems for statistical inference. So he should use a Bruesch-Pagan test to diagnose conditional heteroskedasticity.

Rationale

✖ **This Answer is Incorrect**

Two types of heteroskedasticity can exist: unconditional and conditional. Unconditional heteroskedasticity is not a major problem for statistical inference, but conditional heteroskedasticity does cause problems for statistical inference. So he should use a Bruesch-Pagan test to diagnose conditional heteroskedasticity.

Question 32
L2ET-ITEMSET-PQ1008-1410
LOS: LOS-6120
Lesson Reference: Lesson 3: Violations of Regression Assumptions
Difficulty: medium

Use the following information to answer the next three questions:

An analyst performs a one-independent-variable time-series regression based on monthly observations from January 1996 to December 2010 (both inclusive) and finds that the coefficient on the independent variable is different from zero at the 5% significance level. The Durbin-Watson statistic for the regression is 1.7194.

1% table for k = 1 to 10 and n = 100,150,200.

5% table for k = 1 to 10 and n = 100,150,200.

i.
The sample correlation between the regression residuals from one period and those from the previous period is *closest to*:

○ 0.1632.
○ 0.8597.
● 0.1403.

Rationale
✓ **This Answer is Correct**
$\mathrm{DW} = 2 \times (1 - \mathrm{r}) \rightarrow 1.7194 = 2 \times (1 - \mathrm{r}) \rightarrow \mathrm{r} = 0.1403$

ii.
Which of the following statements is *most likely*?

○ The analyst can reject the null hypothesis and conclude that there is no positive serial correlation between the error terms.
○ The analyst cannot reject the null hypothesis and conclude that there is positive serial correlation between the error terms.
● The analyst can reject the null hypothesis and conclude that there is positive serial correlation between the error terms.

Rationale
✓ **This Answer is Correct**
Given one independent variable, d_l and d_u for 180 observations are approximately 1.758 and 1.779, respectively. Since the Durbin-Watson statistic of 1.7194 is less than d_l, the analyst can reject the null hypothesis of no positive serial correlation and conclude that there is positive serial correlation between the standard errors of the original regression.

iii.
Based on your findings in the previous question, the standard errors of the original regression are *most likely*:

○ Too large
● Too small
○ Suitable for conducting hypothesis tests

Rationale
✓ **This Answer is Correct**

Positive serial correlation implies that the standard errors of the original regression are too small, which results in inflated t-stats. The data cannot be used to conduct hypothesis tests unless it is corrected for serial correlation.

Question 33

L200-PQ0003-1412

LOS: LOS-6140

Lesson Reference: Lesson 4: Errors in Specification and Qualitative Dependent Variables

Difficulty: medium

In the following multiple regression analysis equation of Y = A + 1.5B + 1.65C + e, the C variable is correlated to the B variable. Which of the following statements regarding the estimates of regression coefficients is untrue if we remove the B variable from the equation?

○ Biased

● Unbiased

○ Inconsistent

Rationale

✓ **This Answer is Correct**

The estimates of regression coefficients would not become unbiased if we removed the B variable from the equation, even if it happened to be correlated to the A variable.

Question 34
L2ET-PQ1019-1410
LOS: LOS-6120
Lesson Reference: Lesson 3: Violations of Regression Assumptions
Difficulty: medium

Consider the following statements:

Statement 1: Negative serial correlation occurs when a positive error for one observation increases the chances of a negative error for another.

Statement 2: Serial correlation does not affect the consistency of the estimators of regression parameters.

Which of the following is *most likely*?

- ○ Only Statement 1 is correct.
- ○ Only Statement 2 is correct.
- ● Both statements are correct.

Rationale

✓ **This Answer is Correct**

Both statements are indeed correct.

Question 35
L2ET-ITEMSET-PQ1026-1410
LOS: LOS-6110
Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables
Difficulty: medium

Use the following information to answer the next 4 questions:

An analyst is trying to evaluate the seasonality of a company's annual sales. Management believes that the sales in the fourth quarter are significantly different from sales in the other three quarters. Therefore, he uses the fourth quarter as the reference point in his regression analysis. Results of the regression based on the sales data for the last 10 years are given below:

Results from Regressing Sales on Quarterly Dummy Variables

	Coefficient	**Standard Error**	***t*-Statistic**	
Intercept	5.35	0.95	5.6316	
Sales_{Q1}	−3.225	0.77	−4.1883	
Sales_{Q2}	−0.668	0.77	−0.8675	
Sales_{Q3}	−3.011	0.77	−3.9104	
ANOVA	**df**	**SS**	**MS**	***F***
Regression	3	41.258	13.753	14.8086
Residual	36	33.433	0.929	
Total	39	74.691		
Standard error	0.9380			
R^2	0.5524			
Observations	40			

i.
Average sales in the fourth quarter are *closest to*:

○ 5.63
● 5.35
○ 0.95

Rationale
✓ **This Answer is Correct**
The intercept term in the regression indicates the average value of the omitted category, which in this case is Q4.

ii.
Which of the following is *most likely* regarding the average difference between first and third quarter sales?

○ Q1 sales are greater than Q3 sales by 0.668.
● Q1 sales are less than Q3 sales by 0.214.
○ Q1 sales are less than Q3 sales by 3.225.

Rationale
✓ **This Answer is Correct**

Q4 sales are greater than Q1 sales by 3.225.

Q4 sales are greater than Q3 sales by 3.011.

Therefore, Q3 sales are greater than Q1 sales by 0.214.

iii.

Sales in which of the following quarters are *least likely* different from Q4 sales?

○ Q1
● Q2
○ Q3

Rationale

✓ **This Answer is Correct**

The *t*-stat for Q2 sales is the lowest among the three quarters, which means that Q2 sales are least likely to be different from Q4 sales.

iv.

The analyst wants to evaluate the null hypothesis that, jointly, all the slope coefficients in the regression equal zero. At the 5% significance level, the result of the appropriate hypothesis test is *most likely* that:

● He will reject the null hypothesis that all the slope coefficients in the regression jointly equal zero.
○ He will fail to reject the null hypothesis that all the slope coefficients in the regression jointly equal zero.
○ Only the slope coefficient on Q2 sales is not significantly different from zero.

Rationale

✓ **This Answer is Correct**

H_0: $b_1 = b_2 = b_3 = 0$

H_A: at least one of the slope coefficients $\neq 0$

The *F*-test is used to determine whether the slope coefficients all jointly equal zero.

The *F*-stat is given as 14.8086. With 40 observations, the degrees of freedom for the numerator and denominator, respectively, are 3 and 36. Therefore, at the 5% level of significance, the F_{crit} value lies somewhere between 2.84 and 2.92. Since the *F*-stat is greater than the critical value, we reject the null hypothesis and conclude that the slope coefficients in the regression do not jointly equal zero.

Question 36
L2R10TB-AC021-1512
LOS: LOS-6120
Lesson Reference: Lesson 3: Violations of Regression Assumptions
Difficulty: medium
Unconditional heteroskedasticity is *most likely*:

○ a violation of the assumptions of multiple linear regression and creates a major problem for statistical inference.

● a violation of the assumptions of multiple linear regression, but does not create a major problem for statistical inference.

○ not a violation of the assumptions of multiple linear regression and does not create a major problem for statistical inference.

Rationale
✗ **a violation of the assumptions of multiple linear regression and creates a major problem for statistical inference.**
All types of heteroskedasticity are a violation of the assumptions for regression analysis, but only conditional heteroskedasticity creates a major problem for statistical inference.

Rationale
✓ **a violation of the assumptions of multiple linear regression, but does not create a major problem for statistical inference.**
All types of heteroskedasticity are a violation of the assumptions for regression analysis, but only conditional heteroskedasticity creates a major problem for statistical inference.

Rationale
✗ **not a violation of the assumptions of multiple linear regression and does not create a major problem for statistical inference.**
All types of heteroskedasticity are a violation of the assumptions for regression analysis, but only conditional heteroskedasticity creates a major problem for statistical inference.

Question 37
L2ET-PQ1021-1410
LOS: LOS-6130
Lesson Reference: Lesson 3: Violations of Regression Assumptions
Difficulty: medium

Consider the following statements:

Statement 1: Multicollinearity does not affect the consistency of the estimators of regression parameters.

Statement 2: Multicollinearity results in *t*-stats becoming too small.

Which of the following is *most likely*?

○ Only Statement 1 is correct.
○ Only Statement 2 is correct.
● Both statements are correct.

Rationale
✓ **This Answer is Correct**
Both statements are indeed correct. Regarding Statement 2, multicollinearity results in the standard errors for the regression coefficients becoming inflated, which results in *t*-stats becoming too small and less powerful (in terms of their ability to reject null hypotheses).

Question 38

L2R10TB-AC012-1512

LOS: LOS-6070

Lesson Reference: Lesson 1: Multiple Linear Regression

Difficulty: medium

The *least likely* underlying assumption of a multiple linear regression model is that the:

○ error term is normally distributed.

○ expected value of the error term is zero.

● error term is correlated across observations.

Rationale

✗ **error term is normally distributed.**

The assumptions with respect to the error term are that it is normally distributed, has an expected value of zero, and that it is *uncorrelated* across observations.

Rationale

✗ **expected value of the error term is zero.**

The assumptions with respect to the error term are that it is normally distributed, has an expected value of zero, and that it is *uncorrelated* across observations.

Rationale

✓ **error term is correlated across observations.**

The assumptions with respect to the error term are that it is normally distributed, has an expected value of zero, and that it is *uncorrelated* across observations.

Question 39

L2ET-PQ1014-1410

LOS: LOS-6120

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

After correcting for positive serial correlation using Hansen's method, the Durbin-Watson stat for the regression will *most likely*:

○ Increase toward a value of 2.

○ Decrease toward a value of 2.

● Stay the same.

Rationale

✓ **This Answer is Correct**

After correcting for positive serial correlation using Hansen's method, the robust standard errors are larger than they were originally, but the Durbin-Watson statistic remains the same.

Question 40

L2R10TB-AC014-1512
LOS: LOS-6060
Lesson Reference: Lesson 1: Multiple Linear Regression
Difficulty: medium

When predicting the value of a dependent variable using a multiple linear regression equation, it is *least likely* that the reliability of the estimate will be adversely affected by:

○ the standard error of the estimate.
○ errors in the beta coefficients estimated from the sample data.
● forecasting values inside the range of data on which the model was estimated.

Rationale

✗ **the standard error of the estimate.**

Predictions based on values of the independent variables *outside* of the range of the sample data is more likely to be a concern than when the values are inside the range of data.

Rationale

✗ **errors in the beta coefficients estimated from the sample data.**

Predictions based on values of the independent variables *outside* of the range of the sample data is more likely to be a concern than when the values are inside the range of data.

Rationale

✓ **forecasting values inside the range of data on which the model was estimated.**

Predictions based on values of the independent variables *outside* of the range of the sample data is more likely to be a concern than when the values are inside the range of data.

Question 41

L2R10TB-AC017-1512
LOS: LOS-6110
Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables
Difficulty: medium

When it is necessary to distinguish between three qualitative characteristics in a multiple linear regression equation, the number of dummy variables required is:

○ 1
● 2
○ 3

Rationale

✗ **1**

One fewer dummy variable is needed than the number of characteristics being distinguished, as the final characteristic is identified when all the dummy variables for the other characteristics equal zero.

Rationale

✓ **2**

One fewer dummy variable is needed than the number of characteristics being distinguished, as the final characteristic is identified when all the dummy variables for the other characteristics equal zero.

Rationale

✗ **3**

One fewer dummy variable is needed than the number of characteristics being distinguished, as the final characteristic is identified when all the dummy variables for the other characteristics equal zero.

Question 42

L2R10TB-AC026-1512

LOS: LOS-6130

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

The way to correct for multicollinearity is *most likely* to:

○ compute robust standard errors.

● omit one of the independent variables.

○ use generalized least squares regression.

Rationale

✗ **compute robust standard errors.**

By omitting one of the correlated independent variables, the problem may be rectified.

Rationale

✓ **omit one of the independent variables.**

By omitting one of the correlated independent variables, the problem may be rectified.

Rationale

✗ **use generalized least squares regression.**

By omitting one of the correlated independent variables, the problem may be rectified.

Question 43

L2R10TB-AC022-1512

LOS: LOS-6120

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

The *least* appropriate method of correcting for heteroskedasticity is to:

○ calculate robust standard errors.

○ use generalized least squares regression.

● re-specify the original equation in terms of the independent variables.

Rationale

✗ **calculate robust standard errors.**

Either robust standard errors (which correct the original underestimated standard errors) or generalized least squares (which modify the original equation to account for heteroskedasticity) is appropriate.

Rationale

✗ **use generalized least squares regression.**

Either robust standard errors (which correct the original underestimated standard errors) or generalized least squares (which modify the original equation to account for heteroskedasticity) is appropriate.

Rationale

✓ **re-specify the original equation in terms of the independent variables.**

Either robust standard errors (which correct the original underestimated standard errors) or generalized least squares (which modify the original equation to account for heteroskedasticity) is appropriate.

Question 44

L2QM-TB0008-1412
LOS: LOS-6030
Lesson Reference: Lesson 1: Multiple Linear Regression
Difficulty: medium

A multiple linear regression model is specified as follows:

$$Y_i = 7.5 + 2.3X_{1i} - 0.78X_{2i} + \varepsilon_i$$

Which of the following statements regarding this model is *most accurate?*

● The expected movement in Y for a unit increase in variable X_{1i} while holding X_{2i} constant is a positive move of 2.3.

○ The expected movement in Y for a unit increase in variable X_{1i} while holding X_{2i} constant is a positive move of 9.8.

○ The expected movement in Y for a unit increase in variables X_{1i} and X_{2i} together is a positive move of 3.11.

Rationale

✓ **This Answer is Correct**

The slope coefficient for the variable X_{1i}, namely, the value 2.3, represents the expected movement in Y for a unit increase in the variable X_{1i} while holding all other independent variables constant.

Question 45

L2R10TB-AC010-1512

LOS: LOS-6060

Lesson Reference: Lesson 1: Multiple Linear Regression

Difficulty: medium

Details of a regression analysis to determine earnings per share (EPS) are as follows.

	Coefficient	**Standard Error**	***t*-Statistic**
Intercept	4.21	1.9118	2.202
Slope A	0.43	0.2098	2.050
Slope B	–0.78	0.2502	–3.118
Slope C	0.66	0.3213	2.054

An analyst has estimated that independent variables A, B, and C have values of 2.0, 4.0, and 3.0, respectively. The analyst's estimate of the dependent variable's value, EPS, is *closest* to:

○ –0.28

● 3.93

○ 10.17

Rationale

✗ **–0.28**

The calculation is as follows: EPS = 4.21 + (0.43 × 2.0) – (0.78 × 4.0) + (0.66 × 3.0) = 3.93.

Rationale

✓ **3.93**

The calculation is as follows: EPS = 4.21 + (0.43 × 2.0) – (0.78 × 4.0) + (0.66 × 3.0) = 3.93.

Rationale

✗ **10.17**

The calculation is as follows: EPS = 4.21 + (0.43 × 2.0) – (0.78 × 4.0) + (0.66 × 3.0) = 3.93.

Question 46

L2R10TB-AC020-1512

LOS: LOS-6120

Lesson Reference: Lesson 3: Violations of Regression Assumptions

Difficulty: medium

For regressions based on financial data, the *most likely* impact of heteroskedasticity is that:

● *t*-statistics will be inflated.

○ standard errors will be overestimated.

○ regression parameters will be overestimated.

Rationale

✓ ***t*-statistics will be inflated.**

Standard errors are underestimated, meaning that *t* statistics are inflated. Remember, heteroskedasticity does not affect the consistency of the parameters.

Rationale

☒ **standard errors will be overestimated.**

Standard errors are underestimated, meaning that *t* statistics are inflated. Remember, heteroskedasticity does not affect the consistency of the parameters.

Rationale

☒ **regression parameters will be overestimated.**

Standard errors are underestimated, meaning that *t* statistics are inflated. Remember, heteroskedasticity does not affect the consistency of the parameters.

Question 47

L2QM-TB0006-1412

LOS: LOS-6100

Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables

Difficulty: medium

An analyst is using a multiple linear regression to attempt to explain the liquidity of different stock exchanges. The analyst uses the number of market makers, average size of listed company, and age of exchange as independent variables across 48 different exchanges. In this analysis, how many degrees of freedom are relevant for the unexplained sum of squares for the residual term?

○ 3.

● 44.

○ 47.

Rationale

✓ **This Answer is Correct**

This regression has three independent variables (k) and 48 observation (n). The degrees of freedom for the unexplained sum of squares of the residual is $n - k - 1 = 48 - 3 - 1 = 44$. The degrees of freedom of the regression is $k = 3$ and the total degrees of freedom of the independent variable is $n - 1 = 47$.

Question 48

L2ET-PQ1004-1410
LOS: LOS-6080
Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables
Difficulty: medium

An analyst obtained the following regression results:

	Coefficient	**Standard Error**	***t*-Statistic**
Intercept (b_0)	240.3256	48.5896	4.9460
b_1	1.3920	0.1970	7.0660
b_2	−3.6480	9.6853	−0.3767

ANOVA	**df**	**SS**	**MS**	***F***
Regression	2	3,940.2860		
Residual	97	4,268.1851		
Total	99	8,208.4711		

Regression equation: $Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + e_i$

The *F*-stat is *closest to*:

○ 0.2232

○ 525.0173

● 44.7740

Rationale

✓ **This Answer is Correct**

F-stat = MSR / MSE = (RSS / k) / [SSE / (n-k-1)]

= (3940.2860 / 2) / [4268.1851 / (100-2-1)]

= 44.7740

Question 49
L2QM-TB0009-1412
LOS: LOS-6090
Lesson Reference: Lesson 2: The F-stat, R2, ANOVA, and Dummy Variables
Difficulty: medium

Joe Browne, CFA, is head of quantitative research for Alef-Nought Partners, a buy-side firm specializing in quantitative strategies for managed accounts of high-net-worth individuals. Browne has been presented with the outputs of three multiple linear regressions and immediately notices inconsistencies in the outputs. The outputs are as follows:

Regression	*R*-squared	Adjusted *R*-squared	*F*-ratio
A	0.72	0.55	89.766
B	0.54	0.62	62.454
C	−0.22	0.00	16.998

Which of the regressions above is *least likely* to contain inconsistencies?

● Regression A.
○ Regression B.
○ Regression C.

Rationale

✓ **This Answer is Correct**

The adjusted *R*-squared should always be smaller than the *R*-squared for a regression since it will reduce the *R*-squared every time an independent variable is added to the regression. This rules out regressions B and C as having consistent outputs. It can also be noted that a negative *R*-squared is impossible; hence, this rules out regression C.

Question 50
L2R10TB-AC011-1512
LOS: LOS-6040
Lesson Reference: Lesson 1: Multiple Linear Regression
Difficulty: medium

Details of a regression analysis to determine earnings per share are as follows.

	Coefficient	Standard Error	*t*-Statistic
Intercept	4.21	1.9118	2.202
Slope A	0.43	0.2098	2.050
Slope B	–0.78	0.2502	–3.118
Slope C	0.66	0.3213	2.054

The sample size is 30 and the sample R^2 is 0.57.

The analyst wants to determine whether independent variables B and C are useful for explaining the EPS.

STUDENT'S *t*-DISTRIBUTION (ONE-TAILED PROBABILITIES)

df	$p = 0.05$	$p = 0.025$	$p = 0.005$
26	1.706	2.056	2.779
27	1.703	2.052	2.771
28	1.701	2.048	2.763
29	1.699	2.045	2.756
30	1.696	2.042	2.750
31	1.694	2.040	2.744
32	1.692	2.037	2.738
33	1.691	2.035	2.733
34	1.690	2.032	2.728

Based on a 0.05 significance level, the analyst will *most likely* conclude that independent variable B is:

○ useful in explaining EPS and variable C is useful in explaining EPS.
● useful in explaining EPS, but variable C is not useful in explaining EPS.
○ not useful in explaining EPS, but variable C is useful in explaining EPS.

Rationale
☒ **useful in explaining EPS and variable C is useful in explaining EPS.**
The test is a two-tailed test because the analyst is testing whether either variable B or C is significantly different from 0, which means it does help explain EPS. Using the Student's *t*-Distribution table, the critical statistic for 26 ($n - k - 1$ = 30 – 3 – 1) degrees of freedom with 0.025 in each tail (0.05 significance level divided by 2 because it is a two-tailed test) is ±2.056. The *t*-statistic for independent variable B is –3.118 and for independent variable C, it is 2.054. Since variable B's *t*-statistic falls outside the *t*-critical range of ±2.056, the analyst can conclude at a 0.05 significance level that it is useful in explaining EPS. But, because variable C's *t*-statistic does not fall outside the *t*-critical range of ±2.056, the analyst should conclude at a 0.05 significance level that it is not useful in explaining EPS.

Rationale
✓ **useful in explaining EPS, but variable C is not useful in explaining EPS.**

The test is a two-tailed test because the analyst is testing whether either variable B or C is significantly different from 0, which means it does help explain EPS. Using the Student's *t*-Distribution table, the critical statistic for 26 ($n - k - 1$ = 30 – 3 – 1) degrees of freedom with 0.025 in each tail (0.05 significance level divided by 2 because it is a two-tailed test) is ±2.056. The *t*-statistic for independent variable B is –3.118 and for independent variable C, it is 2.054. Since variable B's *t*-statistic falls outside the *t*-critical range of ±2.056, the analyst can conclude at a 0.05 significance level that it is useful in explaining EPS. But, because variable C's *t*-statistic does not fall outside the *t*-critical range of ±2.056, the analyst should conclude at a 0.05 significance level that it is not useful in explaining EPS.

Rationale

✗ **not useful in explaining EPS, but variable C is useful in explaining EPS.**

The test is a two-tailed test because the analyst is testing whether either variable B or C is significantly different from 0, which means it does help explain EPS. Using the Student's *t*-Distribution table, the critical statistic for 26 ($n - k - 1$ = 30 – 3 – 1) degrees of freedom with 0.025 in each tail (0.05 significance level divided by 2 because it is a two-tailed test) is ±2.056. The *t*-statistic for independent variable B is –3.118 and for independent variable C, it is 2.054. Since variable B's *t*-statistic falls outside the *t*-critical range of ±2.056, the analyst can conclude at a 0.05 significance level that it is useful in explaining EPS. But, because variable C's *t*-statistic does not fall outside the *t*-critical range of ±2.056, the analyst should conclude at a 0.05 significance level that it is not useful in explaining EPS.