1. Jae Park, CFA, is a manager of a hedge fund that bases its security selection on advanced quantitative analysis. For several open job positions with the fund, Park is looking to hire people with scientific and research backgrounds. Using multiple regression, she would like to evaluate the relationship between the expected salary of the candidates based on their years of experience (EXP), number of published research papers (PRP), and amount of grant funding received in their career (GF). The results of that regression are shown below, along with sample critical values. Park wishes to test the results at a 5% significance level (α = 0.05).

**Exhibit 1 Selected Regression Output and ANOVA Data**

| | Coefficient | Standard Error | *t*-Statistic | | |
|---|---|---|---|---|---|
| Intercept | 94.222 | 11.785 | 7.995 | | |
| EXP | 5.080 | 1.116 | 4.550 | | |
| PRP | −0.820 | 1.873 | −0.438 | | |
| GF | 0.212 | 0.136 | 1.552 | | |
| **ANOVA Data** | ***df*** | **Sum of Squares (SS)** | **Mean SS** | ***F*** | **Significance *F*** |
| Regression (*k*) | 3 | 30,430.34 | 10,143.40 | 18.643 | 0 |
| Residual (*n* − *k* − 1) | 22 | 11,969.66 | 544.08 | | |
| Total | 25 | 42,400.00 | | | |
| Observations | 26 | | | | |
| $R^2$ | 0.718 | | | | |
| Standard error | 23.325 | | | | |

**Exhibit 2 Sample Values from *t*-Distribution Table**

| Significance - Two-tailed | 0.100 | 0.050 |
|---|---|---|
| Significance - One-tailed | 0.050 | 0.025 |
| ***df*** | | |
| 21 | 1.7207 | 2.0796 |
| 22 | 1.7171 | 2.0739 |
| 23 | 1.7139 | 2.0687 |
| 24 | 1.7109 | 2.0639 |
| 25 | 1.7081 | 2.0595 |

Park also notices that each candidate attended one of five universities. She is considering how to add a variable for university attended to the regression model and believes dummy variables are the best way to capture this.

Finally, Park suspects that her regression in its current form may violate regression assumptions. Her concern is that her model might have an artificially large $R^2$ and *t*-statistics that are understated.

**Question 1 of 6**

**Based on the data in Exhibit 1, the regression is *most likely* a good predictor of projected salary since:**

a. it has a high $R^2$.
b. the $F$-statistic has a low $p$-value.
c. most coefficients are statistically significant.

---

2. **Question 2 of 6**

**According to Exhibit 1, the *most appropriate* interpretation of the coefficients is that a higher expected salary will result from:**

a. more published papers and less grant funding.
b. more published papers and more grant funding.
c. fewer published papers and more grant funding.

---

3. **Question 3 of 6**

**Based on Exhibits 1 and 2, which slope coefficient is *most likely* to be statistically significant?**

a. GF
b. EXP
c. PRP

---

4. **Question 4 of 6**

**Based on Exhibits 1 and 2, the 95% confidence interval for the EXP coefficient has a lower bound *closest* to:**

a. −4.36
b. 2.77
c. 3.16

---

**5. Question 5 of 6**

**To account for the candidates' universities, the number of dummy variables that Park should add to the regression is *closest* to:**

a. 1
b. 4
c. 5

---

**6. Question 6 of 6**

**If Park's suspicions about her model are correct, then the model would *most likely* show signs of*:***

a. multicollinearity.
b. serial correlation.
c. heteroskedasticity.

---

7. Lana Marek, CFA, a pricing analyst at a European auto manufacturer, is analyzing competitor pricing data. She is working with Micah Hould, CFA, a quantitative analyst at the same company, to create multiple regression models that help identify factors affecting vehicle pricing.

As part of her analysis, she has chosen three independent variables to regress against vehicle price: engine power output (*HP*), vehicle seating capacity (*SC*), and size of the car's interior (*CI*). Marek assumes that all three variables are positively related to price. To test this, she has compiled data for 30 different vehicle models from last year and developed a regression equation and statistics at a 5% level of significance.

Marek believes that there is another variable that affects vehicle price: whether the car belongs to a luxury vehicle brand. She adds a dummy variable and assigns a value of 1 for a luxury vehicle and a value of 0 for a nonluxury vehicle. Marek expects that there will be a positive relationship between the brand variable and price. The updated regression statistics and equation are as follows:

**Exhibit 1 Regression Output and ANOVA Data**

| | Coefficient | Standard error | *t*-statistic | *p*-value (two-tailed) | *p*-value (one-tailed) |
|---|---|---|---|---|---|
| *HP* | 0.113 | 0.013 | 8.531 | 0.000 | 0.000 |
| *SC* | −1.596 | 1.270 | −1.257 | 0.220 | 0.110 |
| *CI* | 11.990 | 3.677 | 3.261 | 0.003 | 0.002 |
| *LX* | 1.104 | 2.200 | 0.502 | 0.620 | 0.310 |
| Intercept | −19.197 | 6.262 | −3.066 | 0.005 | 0.003 |
| **ANOVA data** | ***df*** | **Sum of squares (SS)** | **Mean SS** | | |
| Regression ($k$) | 4 | 3164.10 | 791.03 | | |
| Residual ($n - k - 1$) | 25 | 770.20 | 30.81 | | |
| Total ($n - 1$) | 29 | 3934.30 | | | |
| $R^2$ | 0.804 | | | | |
| Adjusted $R^2$ | 0.773 | | | | |
| Standard error | 5.550 | | | | |
| Observations ($n$) | 30 | | | | |

*Price* = Dealership price, expressed in €1,000s
*HP* = Power output of the vehicle's engine, expressed in units of metric horsepower
*SC* = Seating capacity, expressed as number of seats in the vehicle
*CI* = Size of the car's interior, expressed in units of cubic meters
*LX* = Luxury vehicle dummy variable (1 = luxury, 0 = not luxury)

**Marek notices that $R^2$ increased after she added the dummy variable to her analysis, but adjusted $R^2$ (or $\bar{R}^2$ decreased.**

**Finally, Marek is concerned about multicollinearity's effect on her model. She suspects the seating capacity and interior size variables are highly correlated, which may influence the regression results. She shares her concern with Hould, who responds as follows:**

- **Statement 1: "The most reliable way of detecting multicollinearity is to assess the magnitude of correlations between independent variables."**
- **Statement 2: "A key issue with multicollinearity is that *t*-tests on regression coefficients will result in instances of rejecting a true null hypothesis."**
- **Statement 3: "The most straightforward way to correct multicollinearity is to remove one or more correlated variables from the regression."**

**Question 1 of 6**

**Based on Exhibit 1, which is the *most appropriate* alternative hypothesis and conclusion from Marek's assumption on the seating capacity variable (*SC*)?**

a. $H_a$: $SC > 0$; reject the null.
b. $H_a$: $SC > 0$; fail to reject the null.
c. $H_a$: $SC \leq 0$; fail to reject the null.

---

8. **Question 2 of 6**

**If Marek correctly calculated the F-statistic shown in Exhibit 1 and then compared it with a critical F-value of 2.975, the *most appropriate* conclusion is that for this regression model:**

a. there are no statistically significant coefficients.
b. most of the coefficients are statistically significant.
c. there is at least one statistically significant coefficient.

---

9. **Question 3 of 6**

   **Based on the *LX* dummy variable coefficient in Exhibit 1, Marek's *most appropriate* conclusion is that the model projects a luxury vehicle's price to be:**

   a. €1,104 higher, and Marek would reject the null hypothesis.
   b. €1,104 lower, and Marek cannot reject the null hypothesis.
   c. €1,104 higher, and Marek cannot reject the null hypothesis.

---

10. **Question 4 of 6**

    **Based on Exhibit 1, the estimated price (in €) for a vehicle that has a 200 HP engine, seats 5 people, has an interior capacity of 2.5 cubic meters, and is not considered a luxury vehicle is *closest* to:**

    a. 25,400
    b. 26,500
    c. 59,300

---

11. **Question 5 of 6**

    **The decline in $\bar{R}^2$ *most likely* suggests that:**

    a. $R^2$ is the more suitable measure of the regression's goodness of fit.
    b. the regression model does not explain changes in the dependent variable.
    c. it is caused by including an independent variable without a statistically significant relationship to the dependent variable.

---

12. **Question 6 of 6**

    **Which of Hould's statements on multicollinearity is correct?**

    a. Statement 1
    b. Statement 2
    c. Statement 3

---