Q1. The regression output reveals a high but many of the explanatory variables are statistically insignificant at the 5% significance level. What problem in the regression could give rise to this symptomatic result?

|  |
| --- |
| When R Square is high, and many explanatory variables are statistically insignificant at given level of significance, this is the problem of multicollinearity. It reduces the precision of the estimated coefficients, which weakens the statistical power of the regression model and not be able to trust the p-values to identify explanatory variables that are statistically significant. Additionally, the reason for high R square is that irrelevant variables added in the regression model, thereby, R square of the model increases. |

Q2. Correct this false statement: If an estimator is unbiased, it gives a correct estimate of the population parameter of interest for any sample size.

|  |
| --- |
| If an estimator is unbiased, it gives a correct estimate of the population parameter of interest, but only on average. Technically, because is an unbiased estimator of where is an unknown parameter. |

Q3. Suppose the estimated regression model is Provide an interpretation of the slope coefficient

|  |
| --- |
| If 1% increase in x, then it gives rise to an increase in y by 0.5%. |

Q4. Holding other things constant, what is the effect of a larger sample size on the variance of the OLS estimator? Explain.

|  |
| --- |
| The effect of a larger sample size (n large) on the variance of the OLS estimator tend to zero (decreases). In formula of variance of OLS estimator, “n” shown in indicator which implies that if “n” tend to infinity then variance of the OLS estimator tend to zero.  Formula:  =  n (Large)  variance of OLS estimator (Decrease) |

Q5. A researcher runs a regression that yields the following result:

Write down the long-run relationship between and

|  |
| --- |
|  |

Q6. State two violations of the CLRM assumptions that will give rise to inefficient OLS estimators.

|  |
| --- |
| **1.** When the conditional expected value of the error term but a non-zero constant c, i.e., E(. This case violates when E( , implies that wrongly estimated intercept while the other beta coefficients are not affected.  **2.** Suppose least squares is used to fit a line relating y and x, yi = b1+b2xi+ei. Assume that in our data all the x’s are identical, so that at least some of the x’s are different from their sample mean x bar. When the least square estimators “inefficient” of biased estimators. Thus, E (, a constant. Hence, these are the violations of the CLRM assumptions. |

Use this information to answer Q7 to Q10.

A real estate agent collects a dataset comprising 88 houses and their characteristics. The house features are:

* assess = the assessed value of the house ($’000)
* bdrms = the number of bedrooms
* colonial = a dummy which equals 1 if the house has a colonial style
* price = the price of the house ($’000)
* sqrft = the size of the house in square feet
* lotsize = the size of the lot (the area where the house resides) in square feet

The table below shows the correlation between the variables of interest.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | LOTSIZE | SQRFT | PRICE | ASSESS | BDRMS | COLONIAL |
| LOTSIZE | 1.000000 | 0.183842 | 0.347124 | 0.328146 | 0.136326 | 0.014019 |
| SQRFT | 0.183842 | 1.000000 | 0.787907 | 0.865634 | 0.531474 | 0.065421 |
| PRICE | 0.347124 | 0.787907 | 1.000000 | 0.905279 | 0.508084 | 0.137946 |
| ASSESS | 0.328146 | 0.865634 | **0.905279** | 1.000000 | 0.482474 | 0.082936 |
| BDRMS | 0.136326 | 0.531474 | 0.508084 | 0.482474 | 1.000000 | 0.304575 |
| COLONIAL | 0.014019 | 0.065421 | 0.137946 | 0.082936 | 0.304575 | 1.000000 |

Q7. The agent wanted to study the relationship between house prices and their assessed values. So he plotted the two series together. GRAPH 1 shows the plot for “assess” and “price” across the 88 houses. GRAPH 2 shows the scatter plot of the two series, with each series represented on one of the axes.

|  |  |
| --- | --- |
| GRAPH 1 | GRAPH 2 |

What can you infer about the relationship between the two series?

|  |
| --- |
| Graph 1: The timer series plot shows that these data show cyclic movements because the cycles do not repeat at regular intervals and do not have the same shape. As the assess values increases then the house of prices increases. Around, 40th and 72th number of observations of houses shows higher prices as compared to other houses.  Graph 2: The scatter plot indicates the relationship between house prices and their assessed values which shows that the plot seems to scatter in an upward line or direction, there is a positive relation or strong (positive) correlation i.e., 0.90. |

Q8. The agent believes that the house’s assessed value has predictive power over the settlement price of the house. He ran a regression of ‘PRICE’ on ‘ASSESS’, and the results are shown below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: PRICE | | |  |  |
| Method: Least Squares | | |  |  |
| Sample: 1 88 | |  |  |  |
| Included observations: 88 | | |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|  |  |  |  |  |
|  |  |  |  |  |
| ASSESS | 0.933507 | 0.014175 | 65.85659 | 0.0000 |
|  |  |  |  |  |
|  |  |  |  |  |
| R-squared | 0.817871 | Mean dependent var | | 293.5460 |
| Adjusted R-squared | 0.817871 | S.D. dependent var | | 102.7134 |
| S.E. of regression | 43.83456 | Akaike info criterion | | 10.41002 |
| Sum squared residual | 167167.8 | Schwarz criterion | | 10.43817 |
| Log likelihood | -457.0409 | Hannan-Quinn criterion. | | 10.42136 |
| Durbin-Watson stat | 1.943617 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

What can be inferred from this regression result about the average price of houses based on their assessed values?

|  |
| --- |
| The dependent variable in the model is price and independent variables is assess. The slope coefficient of the assess is positive 0.93, so the interpretation is that for each increase of 1$ in assessed value of the house, then the value of price is estimated to increase by $0.93.  The R square of the model is 0.81 or 81% of the variation in the dependent variable (price) can be explained by the variation in the independent variable (assess). Thus, the model is good-fit and appropriately explains observed data. |

Q9. The agent obtained the descriptive statistic of the resulting residuals from the above regression in Q8. Based on the descriptive statistics, it can be seen that one of the assumptions of the classical linear regression model is violated.



State the assumption that is violated. Explain your answer.

|  |
| --- |
| The detection of autocorrelation in the regression model. So, that the assumption is violated. The Durbin-Watson test value is 1.94 which indicates that the positive autocorrelation because the value is less than 2. |

Q10. How would you change the regression model to ensure that the assumption in Q9 is not violated?

|  |
| --- |
| An AR1 model can be included to reduce the effects of this autocorrelation. The autoregressive process of order p or AR(p) is defined by the equation  Where  , is the vector of model coefficients and p non-negative integer.  The AR model establishes that a realization at time t is a linear combination of the p previous realization plus some noise term.  For p = 0, Xt = ωt and there is no autoregression term.  Now, the modified linear regression model is  Where and |p| < 1, |