1. A variable is standardized in the sample:
2. By multiplying by its mean.
3. By subtracting off its mean and multiplying by its standard deviation.
4. By subtracting off its mean and dividing by its standard deviation.
5. By multiplying by its standard deviation.
6. Suppose that you are interested in estimating the average impact a job training program has on wages. However, you recognize that there are some observed factors that influence wage, participation on the training program, or both. You may still get the unbiased estimate for the program effectiveness by:
7. Excluding those observed factors from your model and running a simple linear regression
8. Including only the factors that predict wage but not participation as controls and running a multiple linear regression
9. Including only the factors that predict participation but not wage as controls and running a multiple linear regression
10. None of the above
11. Which of the following statements is correct?
12. Correlation can be greater than 1
13. R-squared can be less than 0
14. *t* statistics can be less than 0
15. R-squared can be greater than 1
16. The significance level of a test is:
17. The probability of rejecting the null hypothesis when it is false.
18. One minus the probability of rejecting the null hypothesis when it is false.
19. The probability of rejecting the null hypothesis when it is true.
20. One minus the probability of rejecting the null hypothesis when it is true.
21. Suppose you estimate a regression model and obtain and p-value for testing

against What is the *p*-value for testing against ?

1. .086
2. .172
3. .043
4. None of the above
5. The normality assumption implies that:
6. The population error *u* is dependent on the explanatory variables and is normally distributed with mean equal to one and variance .
7. The population error *u* is independent of the explanatory variables and is normally distributed with mean equal to one and variance .
8. The population error *u* is dependent on the explanatory variables and is normally distributed with mean zero and variance .
9. The population error *u* is independent of the explanatory variables and is normally distributed with mean zero and variance .
10. *Female* is a binary variable taking on the value one for females and the value zero for males. State the null hypothesis H0 that there is no discrimination against women, and the alternative hypothesis H1 that there is discrimination against women.
11. is an estimator (not necessarily the OLS estimator) of the population coefficient in the usual linear regression model. is an unbiased estimator of if:
12. The expectation of is equal to
13. converges to as the sample grows
14. has minimum variance
15. is normally distributed
16. If the total sum of squares (SST) in a regression equation is 81, and the residual sum of squares (SSR) is 25, what is the explained sum of squares (SSE)?
17. 64
18. 56
19. 32
20. 18
21. Which of the following statements is true?
22. If the calculated value of *F* statistic is higher than the critical value, we reject the alternative hypothesis in favor of the null hypothesis.
23. The *F* statistic is always nonnegative as SSRr is never smaller than SSRur.
24. Degrees of freedom of a restricted model is always less than the degrees of freedom of an unrestricted model.
25. The *F* statistic is more flexible than the *t* statistic to test a hypothesis with a single restriction.
26. In the following equation, *gdp* refers to gross domestic product, and FDI refers to foreign direct investment.

Which of the following statements is then true?

1. If bank credit increases by 1, *gdp* increases by 0.527%, the level of FDI remaining constant.
2. If bank credit increases by 1%, *gdp* increases by 0.527%, the level of FDI remaining constant.
3. If bank credit increases by 1%, *gdp* increases by 52.7%, the level of FDI remaining constant.
4. If bank credit increases by 1%, *gdp* increases by log(0.527)%, the level of FDI remaining constant.
5. Which of the following correctly identifies a limitation of logarithmic transformation of variables?
6. Taking log of variables make OLS estimates more sensitive to extreme values in comparison to variables taken in level.
7. Logarithmic transformations cannot be used if a variable takes on zero or negative values.
8. Logarithmic transformations of variables are likely to lead to heteroskedasticity.
9. Taking log of a variable often expands its range which can cause inefficient estimates.
10. Which of the following correctly defines *F* statistic if SSRr represents sum of squared residuals from the restricted model of hypothesis testing, SSRur represents sum of squared residuals of the unrestricted model, and *q* is the number of restrictions placed?
11. Which of the following statements is true?
12. When the standard error of an estimate increases, the confidence interval for the estimate narrows down.
13. The lower bound of the confidence interval for a regression coefficient is given by .
14. The lower bound of the confidence interval for a regression coefficient is given by .
15. None of the above.
16. Which of the following correctly identifies a reason why some authors prefer to report the standard errors rather than the *t* statistic?
17. Having standard errors makes it easier to compute confidence intervals.
18. Standard errors are always positive.
19. The *F* statistic can be reported just by looking at the standard errors.
20. Standard errors can be used directly to test multiple linear regressions.

Answer:

1 – 5. C D C C C

6 – 10. D D A B B

11 – 15. B B B D A

1. Suppose that *u* is independent of the explanatory variables, and it takes on the values −2, −1, 0, 1, and 2 with equal probability of 1/5. Does this violate the Gauss-Markov assumptions? Does this violate the CLM assumptions?
2. Consider relating individual performance on a standardized test, score, to a variety of other variables. School factors include average class size, per-student expenditures, average teacher compensation, and total school enrollment. Other variables specific to the student are family income, mother’s education, father’s education, and number of siblings. The model is

State the null hypothesis that student-specific variables have no effect on standardized test performance once school-related factors have been controlled for. What are *k* and *q* for this example? Write down the restricted version of the model.

1. In the birth weight equation

suppose that *faminc* is measured in dollars rather than in thousands of dollars. Thus, define the variable . How will the OLS statistics (slope, intercept, standard error, *t* statistic, etc.) change when *fincdol* is substituted for *faminc*? For the purpose of presenting the regression results, do you think it is better to measure income in dollars or in thousands of dollars?

1. Recall that in the example of the effects of attendance on final exam performance, the model to explain the standardized outcome on a final exam (*stndfnl*) in terms of percentage of classes attended, prior college grade point average, and ACT score is

.

If we add the term to the equation above, what is the partial effect of *atndrte* on *stndfnl*?

1. Suppose that the model

satisfies the first four Gauss-Markov assumptions, where *score* is score on a final exam, *skipped* is number of classes skipped, and *priGPA* is GPA prior to the current semester. If is from the simple regression of *score* on *skipped*, what is the direction of the bias in ?

Answer:

1. Under these assumptions, the Gauss-Markov assumptions are satisfied: *u* is independent of the explanatory variables, so , and . Further, it is easily seen that . Therefore, MLR.4 and MLR.5 hold. The classical linear model assumptions are not satisfied because *u* is not normally distributed (which is a violation of MLR.6).
2. The restricted version of the model is
3. Because , the coefficient on *fincdol* will be the coefficient on *faminc* divided by 1,000. The intercept remains unchanged. The standard error also drops by a factor of 1,000, so the *t* statistic does not change. For readability, it is better to measure family income in thousands of dollars.
4. The new model would be

.

Therefore, the partial effect of *atndrte* on *stndfnl* is . This is what we multiply by Δ*atndrte* to obtain the ceteris paribus change in *stndfnl*.

1. Assuming that *score* depends positively on *priGPA* (β2>0) and *skipped* and *priGPA* are negatively correlated (δ1<0), it follows that β2δ1<0. Because β1 is thought to be negative (or at least nonpositive), a simple regression is likely to overestimate the importance of skipping classes.