1. (10 points) You are hired by the U.S. department of commerce to analyze the effect of changes in U.S. gross domestic product (*gdp*) on changes in the volume of U.S. imports (*im*) over the period 1940 to 2015.

(a) Present economic model of U.S. GDP and volume of U.S. imports.(**2 points**)

(b) Turn your economic model in (a) into an econometric model. (**2 points**)

(c) Explain the zero conditional mean assumption using your model in (b). (**2 points**)

(d) Explain the homoskedasticity assumption using your model in (b). (**2 points**)

(e) Explain what dataset (cross-sectional or time series data) is needed for estimating your model in (b). (**2 points**)

2. (16 points) As an economist, you collect data on births to women in Alaska. Two variables of interest are the dependent variable, infant birth weight in ounces (*bwght*), and an explanatory variable, average number of cigarettes the mother smoked per day during pregnancy (*cigs*). The following simple regression is estimated using data on *n*=14,000 births:

$$\widehat{bwght} = 119.77 - 0.514cigs$$

(a) Interpret the intercept **<u>clearly</u>** (**3 points**).

(b) Interpret the coefficient on *cigs* **<u>clearly</u>** (**3 points**).

(c) To predict a birth weight of 125 ounces, what would *cigs* have to be? Does it make sense? **<u>Explain clearly</u>** (**4 points**)

(d) What is the predicted birth weight when *cigs*=20 (one pack per day)? (**3 points**)

(e) **<u>Agree</u>** or **<u>disagree</u>** with the following statement, and **<u>explain clearly</u>**: "The sum, and therefore the sample average of the OLS errors, is zero." (**3 points**)

3. (12 points) You study whether there is a trade-off between the time spent sleeping per week and the time spent in paid work. Two variables of interest are the dependent variable, minutes spent sleeping at night per week (*sleep*), and an explanatory variable, total minutes worked during the week (*totwrk*). The following simple regression is estimated using data on n=706 people:

$$\widehat{sleep} = 3{,}586.4 - 0.148 totwrk$$
$$R^2 = 0.021$$

(a) Interpret the coefficient on *totwrk* **clearly** (**3 points**).

(b) If *totwrk* increases by 3 hours, by how much is *sleep* estimated to fall? (**3 points**).

(c) Would you say *totwrk* explain much of the variation in *sleep*? Why or why not? **Explain clearly** (**3 points**)

(d) **Agree** or **disagree** with the following statement, and **explain clearly**: "The homoskedasticity assumption plays a key role in providing an unbiased estimator of the ceteris paribus relationship between *totwrk* and *sleep*?" (**3 points**)

4. (12 points) Let *salary* be annual CEO salary in thousands of dollars, and let *sales* be annual firm sales, measured in millions of dollars. Using data, we obtain the following:

$$\log(\widehat{salary}) = 4.822 + 1.237 \log(sales)$$

$$n = 209,\ R^2 = 0.217$$

(a) Interpret the coefficient on log(*sales*) **clearly**. Is the sign of this estimate what you expect it to be? **Explain** (**4 points**).

(b) **Explain** the meaning $R^2 = 0.217$ **clearly** (**3 points**).

(c) What assumptions do we need to provide an unbiased estimator of the ceteris paribus elasticity of *log(salary)* with respect to *log(sales)*? List **all** necessary assumptions (**3 points**).

(d) Decide whether the following statement is **true** (**T**) or **false** (**F**): "The sample covariance between the independent variables and the OLS residuals is zero. Consequently, the sample covariance between the OLS fitted values and the OLS residuals is zero." (**2 points**)

5. (10 points) Using the 526 observations on workers in Alaska, we include *exper* (years of labor market experience) in an equation explaining log(*wage*):

$$\log(\hat{wage}) = 0.43 + 0.0125 exper$$

$$n = 526, \qquad R^2 = 0.182$$

(a) Interpret the effect of experience on wage **clearly** **(3 points)**.

(b) Explain the meaning of R^2=0.182 **clearly** (**3 points**).

(c) Decide whether the following statement is **true** (**T**) or **false** (**F**): "The errors are computed from the data, whereas the residuals are never observable." (**2 points**)

(d) Decide whether the following statement is **true** (**T**) or **false** (**F**): "If Assumption 4 (sample variation in the independent variable) fails, then we will not be able to obtain the OLS estimates." (**2 points**)

6. (30 points) The attached table contains information on chief executive officers for U.S. corporations. The variable *salary* is annual compensation, in millions of dollars, and *roe* is return on equity for the CEO's firm for the previous three years (Return on equity is defined in terms of net income as a percentage of common equity) (**all answers should be rounded to three digits after the decimal**)

   (a) Estimate the relationship between CEO salary and return on equity using OLS; that is obtain the intercept and slope estimates in the equation (**5 points**)

$$salar\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 roe$$

   (b) Interpret the relationship between CEO salary and return on equity **clearly** **(3 points)**.

   (c) Interpret the intercept **clearly**. (**3 points**)

   (d) What is the predicted value of *salary* when *roe* =30? **Explain** **(2 points).**

(e) Compute the fitted values and residuals for each observation and fill the Table (**10 points**).

CEO	y_i **(salary)**	x_i **(roe)**	$\hat{y}_i$ ($\widehat{salary}$)	$\hat{u}_i$ ($y_i - \hat{y}_i$)	$(y_i - \bar{y})^2$	$\hat{u}_i^2$
1	1.1	14				
2	0.8	11				
3	1.4	24				
4	1.2	6				
5	0.8	14				
6	1.2	20				
7	1.7	16				
8	1.0	15				
Sum						
Mean						

(f) Verify that the residuals (approximately) sum to zero (**3 points**).

(g) How much of the variation in *salary* is explained by *roe*? **Explain** (**4 points**).

Table 1. Data and the calculations to estimate the regression equation for the salary-return on equity problem. You can use this Table to obtain the intercept and slope estimates in the equation (**<u>all answers should be rounded to three digits after the decimal</u>**) .

CEO	$\boldsymbol{y_i}$ **(salary)**	$\boldsymbol{x_i}$ **(roe)**	$(y_i - \bar{y})$	$(x_i - \bar{x})$	$(y_i - \bar{y})(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	1.1	14				
2	0.8	11				
3	1.4	24				
4	1.2	6				
5	0.8	14				
6	1.2	20				
7	1.7	16				
8	1.0	15				
Sum						
Mean						

Answer 1

a. The U.S. government influences process and stability through the use of policy (manipulating tax rates and spending programs) and monetary policy (manipulating the number of money in circulation). This stimulates demand and encourages the process. Cuts in government spending have the opposite effect.

b. within the short term, the process is caused by an increase in aggregate demand (AD). If there's spare capacity within the economy, then a rise in aggregate demand will cause a far better level of real GDP.

c. economic process may be a rise within the assembly of economic goods and services, compared from one period of some time to a special. it's often measured in nominal or real (adjusted for inflation) terms. Traditionally, the mixture process is measured in terms of gross national product (GNP) or gross domestic product (GDP), although alternative metrics are sometimes used. The economic process may be a rise within the production of products and services in an economy.

d. Increases in capital goods, labor force, technology, and human capital can all contribute to the process.

e. economic process is typically measured in terms of the increase within the aggregated market value of additional goods and services produced, using estimates like GDP.

Answer 2

a) The intercept term is the maximum birth weight of a newborn. It corresponds to a number of cigarettes= 0. In other words, if a mother does not smoke even a single cigarette during pregnancy then the predicted birth weight is 119.77 ounces.

b) If the number of cigarettes smoked goes up by 1 unit, then the birth weight of the infant goes down by 0.514 ounces.

c) Predicted birth weight = 125 ounces.

Putting in the equation given,

125= 119.77-0.514 cig

Or, cig= (125-119.77) ÷ (-0.514) = -10.2

No, it does not make sense because cigarettes smoked can't be negative.

d) Predicted birth weight when cig= 20 is = 119.77-0.514×**20= 109.49 ounces**

e) Agree. The sum of the OLS errors and therefore their average is always 0. This comes from the first normal equation which we get when we try to minimize the error sum of squares.

Answer 3

a) The coefficient of totwrk shows as per the given question shows that when the entire minutes worked during the week increases by 1 unit, the minute spent sleeping in the dark per week will decrease [thanks to negative relationship] by 0.148 units.

b) If totwrk increases by 3hrs then the autumn in sleep is estimated by multiply change with the coefficient of totwrk and its = 3*0.148 = 0.444hrs

c) we've given coefficient of determination R2 = 0.021 which shows that only 2.1% variation within the variable sleep is demonstrated by the explanatory variable totwrk.

d) Under the heteroskedasticity assumption population variance remains constant for the given values of the explanatory variable. And when this assumption violates the estimators remain unbiased but they're not efficient estimators. The heteroskedasticity assumption affects the efficiency of the estimators but not the unbiasedness.

Hence, we will afflict the given statement.

Answer 4

Let, the true regression model be-

$$log(salary) = \beta_0 + \beta_1 log(sales) + u$$

Here, u is the disturbance term which includes all variables other than sales which affect the salary and they are unobserved.

Now, after estimating this model, you will get-

$$log(salary) = 4.822 + 1.237 log(sales)$$

Here, n is total number of observation which is 209 and $R^2 = 0.217$.

a) The term log (sales) implies the total volume of sales and the coefficient is 1.237 which implies if there is the one unit increase in log of sales then the salary of CEO has been increased by 1.237 units ceteris paribus and this should be positive because if the log of sales increases then log of salary will also increase because if sales increases it leads to increase in total revenue as well as total profit and which leads to increase in salary.

You can understand this by differentiation.

$$\frac{dlog(salary)}{dlog(sales)} = 1.237 > 0$$

It actually measures the elasticity of salary with respect to sales and which tells that one unit increase in sales implies 1.237 units increase in salary, ceteris paribus.

b) First, I would like to discuss about R^2. It is actually the measure of goodness of fit. It measures how the model is fitted best. The formula of measuring R^2 is-

$$R^2 = \frac{ESS}{TSS}$$

ESS is explained sum of square.

TSS is total sum of square.

Now, $R^2 = \frac{ESS}{TSS} = 0.217$

Thus, R^2 measures how much the dependent variable that is log of salary is explained by independent variable that is log of sales and if the value is one then the independent variable can totally explain the dependent variable and the model is good fitted but if the value is zero then the independent variable cannot explain the dependent variable and the model is not good fitted and in this case the value is 0.217 thus, you can explain the 21.7% of the log of salary by using the log of sales and rest 78.3% is explained by the other variables which are not included in the model. Thus, your model is not good fitted but it is fitted moderately.

c) An estimator is called unbiased estimator when the expected value of the sample estimator is exactly equal with population estimator.

Mathematically, you can write-

$$E(\hat{\beta}_1) = \beta_1$$

Where, $\hat{\beta}_1$ is the estimated value and it is 1.237. If the expected value of this is exactly equal with the population estimator that is β_1, then you can say that your model is unbiased.

To get the unbiased estimator, you will have to infer some assumptions.

They are as follows-

I) $E(u) = 0$. That is the disturbance term has zero mean.

Ii) $Cov(u, log(sales)) = 0$. There are no correlation between explanatory variable that is log of sales and disturbance term that is u.

Iii) $Cov(log(salary), log(sales)) \neq 0$. The log of sales and log of salary are correlated with each other.

These are the necessary assumptions that you have to infer to derive the unbiased estimator.

d) The sample covariance between independent variables and OLS residuals are zero.

Thus, mathematically, you can write-

$Cov(x, e) = 0$

Where x is the dependent variable and e is the OLS residual.

Now, prove it-

Let, the regression model be-

$y = \beta_2 x + u$

Now, the fitted model is-

$\hat{y} = \hat{\beta}_2 x$

Now, $e = \hat{u} = y - \hat{y} = y - \hat{\beta}_2 x$

Now, $Cov(x, e) = Cov(x, y - \hat{\beta}_2 x)$

Or, $Cov(x, e) = Cov(x, y) - \hat{\beta}_2 V(x) = Cov(x, y) - \frac{Cov(x, y)}{V(x)} * V(x) = 0$ [Proved]

Now, you have to verify that whether, $Cov(\hat{\beta}_2, e) = 0$ or not.

If there is the correlation between x and e then the estimator will be biased and there is the endogeneity problem and result will be inefficient and that's why you have to assume $Cov(x, e) = 0$. If the estimated coefficient is correlated with the residual then the same problem will appear and the regression is inefficient. Thus, if you assume that there is no correlation between x and e then you has to assume that there is no correlation between the estimated coefficient and e. Thus, the statement is true. This is same for multiple linear regression models.

Answer 5

(a) The given equation may be a log-level model. One term is in logs and therefore the other isn't . Here a rise in experience says, 1-year increase wage by 100*(coefficient of experience).In this case 100*0.0125 = 1.25% .The wage increases by 1.25% when experiences increase by 1 unit.

(b) The R-squared measures the goodness of fit of the model. It explains what proportion of the variation within the variable is explained by the experimental variable . Here it equals 0.182. this suggests 18.2 % of the variation within the variable is explained by the experimental variable .

(c) False, because residuals are calculated because the difference between the observed value and estimated value. So it's observable. Since truth value of the population can't be observed errors are computed from the info .

(d) True. The independent variables are independent of every other. Otherwise, problems of multicollinearity arise.

Answer 6

Table 1

CEO	y_i (Salary)	x_i (roe)	$y_i-\bar{y}$	$x_i-\bar{x}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})^2$	b=(xi-x̄) (yi-ȳ)/(xi-x̄)2	a = ȳ - bx̄
1	1.1	14	-0.050	-1	0.050	0.002	1	0.050	0.400
2	0.8	11	-0.350	-4	1.400	0.123	16	0.088	-0.163
3	1.4	24	0.250	9	2.250	0.063	81	0.028	0.733
4	1.2	6	0.050	-9	-0.450	0.003	81	-0.006	1.233
5	0.8	14	-0.350	-1	0.350	0.123	1	0.350	-4.100

6	1.2	20	0.050	5	0.250	0.003	25	0.010	1.000
7	1.7	16	0.550	1	0.550	0.303	1	0.550	-7.100
8	1	15	-0.150	0	0.000	0.023	0	-	-
SUM	9.2	120	0.0000	0	4.400	0.640	206	1.070	-7.996
Mean	1.15	15	0.0000	0	0.550	0.080	25.750	0.134	-0.999

ŷ =

ax+ b

Table-2

CEO	y_i (Salary)	x_i (roe)	$\hat{y}$= ax+ b	$\hat{u} = y_i - \hat{y}$	$(y_i-\bar{y})^2$	$(\hat{u})^2$
1	1.1	14	1.100	0	0.002	0.000
2	0.8	11	0.800	0	0.123	0.000
3	1.4	24	1.400	0	0.063	0.000
4	1.2	6	1.200	0	0.003	0.000
5	0.8	14	0.800	0	0.123	0.000
6	1.2	20	1.200	0	0.003	0.000
7	1.7	16	1.700	0	0.303	0.000
8	1	15	-	-	0.023	-
SUM	9.2	120	8.200	0	0.640	0.000

Mean	1.15	15	1.025	0	0.080	0.000

A. Intercept = 0.830 and Slope Estimate = 0.021

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.383203158							
R Square	0.14684466							
Adjusted R	0.004652104							

Square								
Standard Error	0.301667427							
Observations	8							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	0.093980583	0.093980583	1.032717	0.348734571			
Residual	6	0.546019417	0.091003236					
Total	7	0.64						
	Coefficients	Standard	t Stat	P-value	Lower 95%	Upper 95%	Lower	Upper

		Error					95.0%	95.0%
Intercept	0.830	0.332824493	2.492640019	0.046994	0.015219454	1.644003847	0.015219454	1.644003847
xi	0.021	0.021018166	1.016226809	0.348735	- 0.030070375	0.072788822	- 0.030070375	0.072788822

B.As the return on equity (ROE) increase by 1 unit the salary of CEO increases by 0.021

C. when the ROE is 0, the salary of CEO is 0.830

D. **predicted value of salary =** 0.830 + 30*0.021 = 1.46

E. **Refer attached Table 1 and Table 2**

F. Refer to **Table 2**

G.

R Square	0.14684466

14.68% variations in salary are explained by *roe*.