The University of Kansas

Department of Economics

Midterm **Jun/27/2019**
Econ 526 - Introduction to Econometrics **Instructor: Caio Vigo Pereira**

Name:

REGRESSION (A)

Consider a model relating the annual number of crimes on college campuses to the number of police officers and student enrollment. The econometric model is:

$$log(crime) = \beta_0 + \beta_1 log(enroll) + u$$

where *crime* is total campus crimes, *police* is the number of employed officers and *enroll* is the total enrollment.
The *R* output is:

```
===============================================
                        Dependent variable:
                    ---------------------------
                            log(crime)
-----------------------------------------------
log(enroll)                  1.270***
                             (0.110)

Constant                    -6.631***
                             (1.034)

-----------------------------------------------
Observations                    97
R2                            0.585
Adjusted R2                   0.580
Residual Std. Error      0.895 (df = 95)
F Statistic          133.792*** (df = 1; 95)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

SECTION A - MULTIPLE CHOICE

3% 1. Based on the Regression (A) above, what is the effect on the dependent variable if *enroll* increases 10%?
   A. $\widehat{crime}$ will increase 12.70%
   B. $\widehat{log(crime)}$ will increase 1.270%
   C. $\widehat{log(crime)}$ will increase 12.70%
   D. $\widehat{crime}$ will increase 1.270%

3% 2. Among the measures of central tendency of a distribution we have:
A. $Med(X)$ and $sd(X)$
B. $E(X)$ and $sd(X)$
C. $E(X)$ and $Mode(X)$
D. $Mode(X)$ and $Var(X)$

3% 3. Let X and Y be two discrete random variables. Knowing that the conditional expectation of X given Y is given by:

$$\sum_{j=1}^{m} x_j f_{X|Y}(x_j|y)$$

What is the term $f_{X|Y}(x_j|y)$ used in this conditional expectation?

A. the conditional probability of X given Y
B. the joint distribution of X given Y
C. the joint distribution of Y given X
D. the probability density function of X

3% 4. Consider any simple linear regression model, such as: $y = \beta_0 + \beta_1 x + u$. What is the Explained Sum of Squares (ESS) equal to?
A. $\sum_{i=1}^{n}(y_i - \bar{y})^2$

B. $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$

C. $\sum_{i=1}^{n} \hat{u}_i^2$

D. $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

[This statement refers to the following four questions]
Let X_1, X_2, and X_3 be i.i.d. random variables from a population with mean μ and variance σ^2. Consider the following estimators for the mean μ:

$$W = \sum_{i=1}^{3} \frac{1}{i^2} X_i$$

$$H = \sum_{i=1}^{3} \frac{1}{i} X_i$$

$$V = \sum_{i=1}^{3} \frac{1}{3} X_i$$

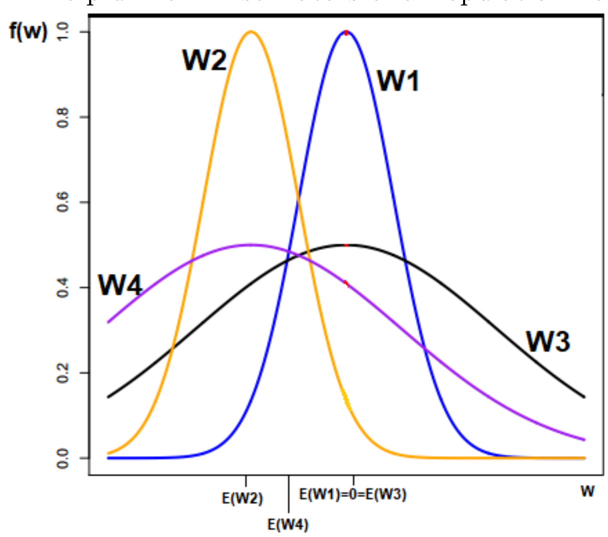
3% 5. What is the $E(W)$ of the estimator?
A. $\frac{11}{6}\mu$
B. μ
C. 3μ
D. $\frac{49}{36}\mu$

3% 6. What is the $E(H)$ of the estimator?
A. $\frac{11}{6}\mu$
B. μ
C. 3μ
D. $\frac{1}{3}\mu$

3% 7. What can you tell about the bias of the estimators W and H?
A. W and H are both **unbiased** estimators for the mean μ
B. W is a **biased** and H is an **unbiased** estimator for the mean μ
C. W is an **unbiased** and H is a **biased** estimator for the mean μ
D. W and H are both **biased** estimators for the mean μ

3% 8. What is the variance of V?
A. $\frac{1}{9}\sigma^2$
B. σ^2
C. $3\sigma^2$
D. $\frac{1}{3}\sigma^2$

Figure 1: The p.d.f. of 4 Estimators of a Population Parameter



3% 9. Figure 1 shows the p.d.f. of 4 estimators of the population parameter θ. Knowing that $\theta = 0$, which estimator(s) for the parameter θ is(are) **biased**?
A. $W1$
B. $W1$ and $W3$
C. $W2$ and $W4$
D. $W1$ and $W2$

3% 10. Refer to Figure 1 again. Knowing that $\theta = 0$, which estimator(s) is(are) relatively efficient in comparison to another one?
A. $W1$ and $W3$ are efficient relative to $W4$
B. $W1$ is efficient relative to $W3$
C. $W3$ is efficient relative to $W4$
D. $W3$ is efficient relative to $W1$

SECTION B - TRUE OR FALSE

2% 1. Let X and Y be two random variables. Then $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$.
◯ True ◯ False

2% 2. Let c be a constant. Then $\text{Var}(c) = c^2$.
◯ True ◯ False

2% 3. Knowing that KU has the following units/campuses: Lawrence, Edwards Campus, the medical school in Kansas City (besides educational and research sites in Garden City, Hays, Leavenworth, Parsons, Topeka, Salina and Wichita). You are interested to know on average how many hours per week KU students spend doing homework. You go to Lawrence campus and randomly survey students walking to classes on Jayhawk boulevard during one day. Then, this is a random sample representing the entire KU students population.
◯ True ◯ False

20% 4. The *Law of Large Number* (LLN) is related with the concept of convergence in probability, while *The Central Limit Theorem* (CLT) is related with convergence in distribution.
◯ True ◯ False

2% 5. You have a cross-sectional dataset with an independent variable X and a dependent variable Y. You find a positive correlation between X and Y. Then you can conclude that X causes Y.
◯ True ◯ False

2% 6. In a cross-sectional dataset the order of the observations is arbitrary, while in a time series dataset the order is important because it is likely that we have correlated observations.
◯ True ◯ False

2% 7. Consider the following simple linear regression model: $y = \beta_0 + \beta_1 x + u$. The essential assumption to derive the estimators of β_0 and β_1 through the Method of Moments is $E(u|X) = 0$.
◯ True ◯ False

2% 8. Consider the following simple linear regression model: $y = \beta_0 + \beta_1 x + u$. When we derive the estimators for β_0 and β_1 we get 2 First Order Conditions.
◯ True ◯ False

3% 9. [**This question refers to Regression (A) on the first page of your exam**]
This model is also known as constant percentage model.
◯ True ◯ False

3% 10. [**This question refers to Regression (A) on the first page of your exam**]
Based on this model, β_1 represents the elasticity of *crime* with respect to *enroll*.
◯ True ◯ False

SECTION C - SHORT ANSWER

1. **This question refers to Regression (B) below**

Consider a random sample with the Grade Point Average (GPA) and standardized test scores (ACT), along with the performance in an introductory economics course, for students at a large public university. The variable to be explained is *score*, which is the final score in the course measured as a percentage. The variable *hsgpa* is the high school GPA, *actmth* is the ACT in math and *colgpa* is the college GPA of the student prior to take the economics course.

REGRESSION (B)

```
===============================================
                        Dependent variable:
                    ---------------------------
                               score
-----------------------------------------------
colgpa                       14.3155***
                              (0.6997)

Constant                     32.3061***
                              (2.0049)

-----------------------------------------------
Observations                    856
R2                             0.3289
Adjusted R2                    0.3281
Residual Std. Error      10.9842 (df = 854)
F Statistic          418.5822*** (df = 1; 854)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

3% (a) Using the variables names, write the simple linear regression model. **[1 line answer]**

3% (b) Using the variables names, write the estimated OLS regression line (also known as SRF or SRL). **[1 line answer]**

3% (c) Using the variables names, write population regression function (PRF). **[1 line answer maximum]**

3% (d) What is the predicted value for whichever is your dependent variable if *colgpa* increases one unit? **[1-3 lines answer]**

3% (e) What is the meaning of the R^2? How is R^2 calculated (formula)? **[2-3 lines answer]**

3% (f) Interpret the R^2 of the regression. **[1-2 lines answer]**

2. **This question refers to Table 1 on next page**

   In this table you have a random sample with 50 data points from a population, i.e., your observations are $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$, where $n = 50$. Considering the following econometric model $y = \beta_0 + \beta_1 x + u$, answer the questions below.

4% (a) What is the OLS estimate of β_1 ? **[1-3 lines answer]**

4% (b) What is the OLS estimate of β_0 ? **[1-3 lines answer]**

4% (c) What is the $\hat{y}$ of observation number 10, i.e., what is $\hat{y}_i$? **[1 line answer]**

4% (d) What is the residual of observation number 10, i.e., what is $\hat{u}_i$? **[1 line answer]**

4% (e) Does the OLS regression line (also known as SRF or SRL) underpredicts or overpredicts y_{10}? Why? **[1 line answer]**

4% (f) Find the value of **A** (located at the bottom - last row - of the table)? Why? **[1-2 lines answer]**

4% (g) Find the value of **B** (located at the 10th row of the table)? **[1 line answer]**

3. Consider the following regression model:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

4% (a) Specify the least squares function that is minimized by OLS. **[1-3 lines answer]**

4% (b) EXTRA POINTS Under which assumptions the OLS estimators for the parameters will be unbiased? State and briefly explain each one of the assumptions. **[4 lines answer]**

1% (c) EXTRA POINTS State precisely the theorem that guarantees that the OLS estimators for the parameters are unbiased? [You may refer to part (b).] **[1-3 lines answer]**

TABLE I

Obs. #	y_i	x_i	$(y_i - \bar{y})$	$(x_i - \bar{x})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	$\hat{y}_i$	$(y_i - \hat{y}_i)$	$(\hat{y}_i - \bar{y})^2$	$(y_i - \hat{y}_i)^2$
1	140	80	45	42	2025	1764	1890	?	?	?	?
2	80	20	-15	-18	225	324	270	?	?	?	?
3	107	50	12	12	144	144	144	?	?	?	?
4	68	30	-27	-8	729	64	216	?	?	?	?
5	52	10	-43	-28	1849	784	1204	?	?	?	?
6	90	40	-5	2	25	4	-10	?	?	?	?
7	60	10	-35	-28	1225	784	980	?	?	?	?
8	101	40	6	2	36	4	12	?	?	?	?
9	45	10	-50	-28	2500	784	1400	?	?	?	?
10	110	30	15	-8	225	64	-120	?	?	?	B
11	50	10	-45	-28	2025	784	1260	?	?	?	?
12	80	30	-15	-8	225	64	120	?	?	?	?
13	150	70	55	32	3025	1024	1760	?	?	?	?
14	50	20	-45	-18	2025	324	810	?	?	?	?
15	77	10	-18	-28	324	784	504	?	?	?	?
16	132	70	37	32	1369	1024	1184	?	?	?	?
17	139	70	44	32	1936	1024	1408	?	?	?	?
18	114	60	19	22	361	484	418	?	?	?	?
19	34	0	-61	-38	3721	1444	2318	?	?	?	?
20	107	40	12	2	144	4	24	?	?	?	?
21	94	40	-1	2	1	4	-2	?	?	?	?
22	100	40	5	2	25	4	10	?	?	?	?
23	40	0	-55	-38	3025	1444	2090	?	?	?	?
24	70	20	-25	-18	625	324	450	?	?	?	?
25	180	90	85	52	7225	2704	4420	?	?	?	?
26	160	80	65	42	4225	1764	2730	?	?	?	?
27	70	0	-25	-38	625	1444	950	?	?	?	?
28	127	40	32	2	1024	4	64	?	?	?	?
29	108	60	13	22	169	484	286	?	?	?	?
30	105	50	10	12	100	144	120	?	?	?	?
31	50	10	-45	-28	2025	784	1260	?	?	?	?
32	137	70	42	32	1764	1024	1344	?	?	?	?
33	140	60	45	22	2025	484	990	?	?	?	?
34	35	0	-60	-38	3600	1444	2280	?	?	?	?
35	56	0	-39	-38	1521	1444	1482	?	?	?	?
36	85	30	-10	-8	100	64	80	?	?	?	?
37	153	90	58	52	3364	2704	3016	?	?	?	?
38	46	10	-49	-28	2401	784	1372	?	?	?	?
39	77	20	-18	-18	324	324	324	?	?	?	?
40	160	90	65	52	4225	2704	3380	?	?	?	?
41	33	20	-62	-18	3844	324	1116	?	?	?	?
42	179	90	84	52	7056	2704	4368	?	?	?	?
43	79	20	-16	-18	256	324	288	?	?	?	?
44	154	70	59	32	3481	1024	1888	?	?	?	?
45	54	10	-41	-28	1681	784	1148	?	?	?	?
46	133	60	38	22	1444	484	836	?	?	?	?
47	96	40	1	2	1	4	2	?	?	?	?
48	127	70	32	32	1024	1024	1024	?	?	?	?
49	65	10	-30	-28	900	784	840	?	?	?	?
50	51	10	-44	-28	1936	784	1232	?	?	?	?
Sum	**4,750**	**1,900**	A	not provided	**84,154**	**40,000**	**55,180**	not provided	not provided	not provided	not provided