The University of Kansas

Department of Economics

Midterm **Oct/10/2019**
Econ 526 - Introduction to Econometrics **Instructor: Caio Vigo Pereira**

Name:

SECTION A - MULTIPLE CHOICE

[This statement refers to the dataset presented in Table I below]
Table I shows a random sample with 40 observations (data points) from a population. Thus, your observations are $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$, where $n = 40$. Consider a simple linear regression model given by $y_i = \beta_0 + \beta_1 x_i + u_i$.

TABLE I

column (1) Obs. #	column (2) y_i	column (3) x_i	column (4) $(y_i - \bar{y})$	column (5) $(x_i - \bar{x})$	column (6) $(y_i - \bar{y})^2$	column (7) $(x_i - \bar{x})^2$	column (8) $(x_i - \bar{x})(y_i - \bar{y})$	column (9) $\hat{y}_i$	column (10) $(y_i - \hat{y}_i)$	column (11) $(\hat{y}_i - \bar{y})^2$	column (12) $(y_i - \hat{y}_i)^2$
1	175	80	36.625	35.5	1341.391	1260.25	1300.188	188.39	-13.39	2501.19	179.21
2	114	30	-24.375	-14.5	594.1406	210.25	353.4375	117.95	-3.95	417.28	15.58
3	127	40	-11.375	-4.5	129.3906	20.25	51.1875	132.04	-5.04	40.19	25.36
4	148	60	9.625	15.5	92.64063	240.25	149.1875	160.21	-12.21	476.82	149.11
5	117	30	-21.375	-14.5	456.8906	210.25	309.9375	117.95	-0.95	417.28	0.90
6	198	90	59.625	45.5	3555.141	2070.25	2712.938	202.47	-4.47	4108.79	20.02
7	181	70	42.625	25.5	1816.891	650.25	1086.938	174.30	6.70	1290.54	44.90
8	91	10	-47.375	-34.5	2244.391	1190.25	1634.438	89.77	1.23	2362.27	1.51
9	78	10	-60.375	-34.5	3645.141	1190.25	2082.938	89.77	-11.77	2362.27	138.58
10	146	30	7.625	-14.5	58.14063	210.25	-110.563	not provided	D	not provided	not provided
11	153	60	14.625	15.5	213.8906	240.25	226.6875	E	not provided	not provided	not provided
12	181	70	42.625	25.5	1816.891	650.25	1086.938	174.30	6.70	1290.54	44.90
13	99	20	-39.375	-24.5	1550.391	600.25	964.6875	103.86	-4.86	1191.31	23.62
14	178	80	39.625	35.5	1570.141	1260.25	1406.688	188.39	-10.39	2501.19	107.89
15	112	10	-26.375	-34.5	695.6406	1190.25	909.9375	89.77	22.23	2362.27	494.09
16	182	80	43.625	35.5	1903.141	1260.25	1548.688	188.39	-6.39	2501.19	40.79
17	84	10	-54.375	-34.5	2956.641	1190.25	1875.938	89.77	-5.77	2362.27	33.31
18	119	40	-19.375	-4.5	375.3906	20.25	87.1875	132.04	-13.04	40.19	169.92
19	129	40	-9.375	-4.5	87.89063	20.25	42.1875	132.04	-3.04	40.19	9.21
20	172	60	33.625	15.5	1130.641	240.25	521.1875	160.21	11.79	476.82	138.98
21	84	10	-54.375	-34.5	2956.641	1190.25	1875.938	89.77	-5.77	2362.27	33.31
22	105	20	-33.375	-24.5	1113.891	600.25	817.6875	103.86	1.14	1191.31	1.30
23	135	50	-3.375	5.5	11.39063	30.25	-18.5625	146.12	-11.12	60.04	123.73
24	125	40	-13.375	-4.5	178.8906	20.25	60.1875	132.04	-7.04	40.19	49.50
25	136	40	-2.375	-4.5	5.640625	20.25	10.6875	132.04	3.96	40.19	15.72
26	210	90	71.625	45.5	5130.141	2070.25	3258.938	202.47	7.53	4108.79	56.63
27	129	20	-9.375	-24.5	87.89063	600.25	229.6875	103.86	25.14	1191.31	632.03
28	177	50	38.625	5.5	1491.891	30.25	212.4375	146.12	30.88	60.04	953.37
29	68	10	-70.375	-34.5	4952.641	1190.25	2427.938	89.77	-21.77	2362.27	474.01
30	200	90	61.625	45.5	3797.641	2070.25	2803.938	202.47	-2.47	4108.79	6.12
31	205	90	66.625	45.5	4438.891	2070.25	3031.438	202.47	2.53	4108.79	6.38
32	157	60	18.625	15.5	346.8906	240.25	288.6875	160.21	-3.21	476.82	10.31
33	193	70	54.625	25.5	2983.891	650.25	1392.938	174.30	18.70	1290.54	349.72
34	100	20	-38.375	-24.5	1472.641	600.25	940.1875	103.86	-3.86	1191.31	14.90
35	91	0	-47.375	-44.5	2244.391	1980.25	2108.188	75.68	15.32	3930.17	234.58
36	165	60	26.625	15.5	708.8906	240.25	412.6875	160.21	4.79	476.82	22.93
37	98	30	-40.375	-14.5	1630.141	210.25	585.4375	117.95	-19.95	417.28	397.91
38	156	60	17.625	15.5	310.6406	240.25	273.1875	160.21	-4.21	476.82	17.73
39	142	40	3.625	-4.5	13.14063	20.25	-16.3125	132.04	9.96	40.19	99.29
40	75	10	-63.375	-34.5	4016.391	1190.25	2186.438	89.77	-14.77	2362.27	218.21
Sum	**5,535**	**1,780**	A	B	**64,127**	**29,190**	**41,123**	**5,535**	C	not provided	**6,195**

3% 1. Refer to Table I. Knowing A , B and C (located in the bottom of the table), what is $A^2 \cdot B^3 \cdot C$?
A. 2.3
B. 0
C. 2.1
D. 1.9

3% 2. Refer to Table I again. On the bottom of column (6) we have a term equal to 64, 127. In a regression setting, what is the name of this term?
A. Explained Sum of Squares (SSE)
B. Total Sum of Squares (SST)
C. Residual Sum of Squares (SSR)
D. Sum of Errors (SE)

3% 3. Refer to Table I again. What is $\widehat{\beta}_0$ equal to?
A. 75.7
B. 0
C. 0.9
D. 1.4

3% 4. Refer to Table I again. What is $\widehat{\beta}_1$ equal to?
A. 75.7
B. 44.5
C. 0.9
D. 1.4

3% 5. Refer to Table I again. What is the R^2 equal to?
A. 0.90
B. 0.10
C. 0.46
D. 0.66

3% 6. Refer to Table I again. What is the sample variance of X equal to? (i.e., what is the S^2 of X?)
A. 1, 644.3
B. 45.6
C. 748.5
D. 414.9

3% 7. Let X and Y be two random variables. ________, ________, ________ are, respectively, examples of measures of central tendency of X, variability of X and association between X and Y:
A. $Med(X)$, $sd(X)$, and $Var(X)$
B. $E(X)$, $Cov(X, Y)$ and $sd(X)$
C. $E(X)$, $Corr(X, Y)$ and $Cov(X, Y)$
D. $Med(X)$, $Var(X)$ and $Corr(X, Y)$

3% 8. 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

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

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

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

3% 9. What is the expected value of the estimator W? (i.e., what is $E(W)$?)
A. $\frac{11}{6}\mu$
B. μ
C. 3μ
D. $\frac{49}{36}\mu$

3% 10. 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

SECTION B - TRUE OR FALSE

3% 1. We say that an estimator is consistent when the expected value of the estimator is equal to the true parameter.
◯ True ◯ False

3% 2. 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

3% 3. In a random sample with cross-sectional data, the order of observations is important because it is likely that we have correlated observations.
◯ True ◯ False

3% 4. In a simple linear regression model such as $y = \beta_0 + \beta_1 x + u$, 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

3% 5. In a simple linear regression model such as $y = \beta_0 + \beta_1 x + u$, when we derive the estimators for β_0 and β_1 we get 2 First Order Conditions.
◯ True ◯ False

3% 6. In a simple linear regression model such as $y = \beta_0 + \beta_1 x + u$, x is the unknown (populational) parameter to be estimated using data.
◯ True ◯ False

SECTION C - SHORT ANSWER

1. **This question refers to Regression (A) 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 *colgpa* is the college GPA of the student prior to take the economics course.

REGRESSION (A)

```
===============================================
                        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
```

4% (a) Using the variables names, write down the simple linear regression model. **[1 line answer]**

4% (b) Using the variables names, write down the estimated OLS regression line (also known as SRF or SRL). **[1 line answer]**

4% (c) Using the variables names, write down the population regression function (PRF). **[1 line answer]**

4% (d) What is the predicted change in whichever is your dependent variable if *colgpa* increases one unit? **[1-3 lines answer]**

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

Consider a model relating the annual number of crimes on college campuses to the student enrollment. The variable *crime* is the total campus crimes, and *enroll* is the total enrollment.

REGRESSION (B)

	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

4% (a) Using the variables names, write down the estimated OLS regression line (also known as SRF or SRL). **[1 line answer]**

4% (b) How the model estimated in regression (B) is known (name)? **[1-2 lines answer]**

4% (c) Interpret the results of the regression, i.e. explain how a change of either 1 unit or 1% in x (whichever is correct) affect y. **[1-2 lines answer]**

4% (d) How many observations were used in the regression? What is the R^2 of the regression? **[1-2 lines answer]**

4% (e) What is the meaning of the R^2? How is R^2 calculated (formula)? **[2-3 lines answer]**

4% (f) Interpret the R^2 of the regression. **[1-2 lines answer]**

3. **This question refers to Table 1 on the first page of your exam.**

4% (a) Find the residual for observation 10, i.e., find the $\hat{u}_{10}$ given in D ? **[1-2 lines answer]**

4% (b) Find the fitted value for observation 11, i.e., find the $\hat{y}_{11}$ given in E ? **[1-2 lines answer]**

4% (c) For observation 30, does the OLS regression line (also known as SRF or SRL) underpredicts or overpredicts y_{30}? Explain? **[1-2 lines answer]**

5% (d) EXTRA POINTS Specify the least squares function that is minimized by OLS, i.e., write down the objective function of the Least Squares method. Explain in few words what is the goal. **[1-3 lines answer]**