The University of Kansas

Department of Economics

Midterm **Mar/03/2020**
Econ 526 - Introduction to Econometrics **Instructor: Caio Vigo Pereira**

Name:

SECTION A - MULTIPLE CHOICE

$\boxed{3\%}$ 1. 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 $Corr(X, Y)$

[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:

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

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

$\boxed{3\%}$ 2. What is the expected value of the estimator W? (i.e., what is $E(W)$?)
A. $\frac{1}{2}\mu$
B. μ
C. 2μ
D. $\frac{3}{2}\mu$

$\boxed{3\%}$ 3. 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

[This statement refers to the dataset presented in Table I below]
Table I shows a random sample with 60 observations (data points) from a population. Thus, your observations are $\{(x_i, y_i) : i = 1, 2, \ldots, n\}$, where $n = 60$. 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	65	10	-66.35	-30	4,402.32	900.00	1,990.50	83.88	-18.88	2,253.70	356.34
2	127	40	-4.35	0	18.92	0.00	0.00	131.35	-4.35	0.00	18.92
3	66	0	-65.35	-40	4,270.62	1,600.00	2,614.00	68.05	-2.05	4,006.58	4.21
4	174	80	42.65	40	1,819.02	1,600.00	1,706.00	194.65	-20.65	4,006.58	426.32
5	131	40	-0.35	0	0.12	0.00	0.00	131.35	-0.35	0.00	0.12
6	113	35	-18.35	-5	336.72	25.00	91.75	123.44	-10.44	62.60	108.95
7	91	10	-40.35	-30	1,628.12	900.00	1,210.50	83.88	7.12	2,253.70	50.74
8	182	70	50.65	30	2,565.42	900.00	1,519.50	178.82	3.18	2,253.70	10.09
9	75	10	-56.35	-30	3,175.32	900.00	1,690.50	83.88	-8.88	2,253.70	78.80
10	227	80	95.65	40	9,148.92	1,600.00	3,826.00	194.65	32.35	4,006.58	1,046.68
11	165	70	33.65	30	1,132.32	900.00	1,009.50	178.82	-13.82	2,253.70	191.08
12	122	30	-9.35	-10	87.42	100.00	93.50	115.53	6.47	250.41	41.92
13	142	50	10.65	10	113.42	100.00	106.50	147.17	-5.17	250.41	26.77
14	69	10	-62.35	-30	3,887.52	900.00	1,870.50	83.88	-14.88	2,253.70	221.32
15	118	10	-13.35	-30	178.22	900.00	400.50	83.88	34.12	2,253.70	1,164.39
16	89	20	-42.35	-20	1,793.52	400.00	847.00	99.70	-10.70	1,001.64	114.52
17	82	10	-49.35	-30	2,435.42	900.00	1,480.50	83.88	-1.88	2,253.70	3.52
18	160	70	28.65	30	820.82	900.00	859.50	178.82	-18.82	2,253.70	354.31
19	113	30	-18.35	-10	336.72	100.00	183.50	115.53	-2.53	250.41	6.38
20	159	50	27.65	10	764.52	100.00	276.50	147.17	11.83	250.41	139.85
21	82	10	-49.35	-30	2,435.42	900.00	1,480.50	83.88	-1.88	2,253.70	3.52
22	150	50	18.65	10	347.82	100.00	186.50	147.17	2.83	250.41	7.98
23	161	70	29.65	30	879.12	900.00	889.50	178.82	-17.82	2,253.70	317.66
24	85	15	-46.35	-25	2,148.32	625.00	1,158.75	91.79	-6.79	1,565.07	46.09
25	173	65	41.65	25	1,734.72	625.00	1,041.25	$\boxed{E}$	not provided	not provided	not provided
26	210	90	78.65	50	6,185.82	2,500.00	3,932.50	210.47	-0.47	6,260.28	0.22
27	225	80	93.65	40	8,770.32	1,600.00	3,746.00	194.65	30.35	4,006.58	921.27
28	123	10	-8.35	-30	69.72	900.00	250.50	83.88	39.12	2,253.70	1,530.62
29	62	10	-69.35	-30	4,809.42	900.00	2,080.50	83.88	-21.88	2,253.70	478.60
30	63	0	-68.35	-40	4,671.72	1,600.00	2,734.00	68.05	-5.05	4,006.58	25.53
31	83	10	-48.35	-30	2,337.72	900.00	1,450.50	83.88	-0.88	2,253.70	0.77
32	125	40	-6.35	0	40.32	0.00	0.00	131.35	-6.35	0.00	40.32
33	121	20	-10.35	-20	107.12	400.00	207.00	99.70	21.30	1,001.64	453.64
34	174	70	42.65	30	1,819.02	900.00	1,279.50	178.82	-4.82	2,253.70	23.26
35	162	45	30.65	5	939.42	25.00	153.25	not provided	$\boxed{D}$	not provided	not provided
36	210	90	78.65	50	6,185.82	2,500.00	3,932.50	210.47	-0.47	6,260.28	0.22
37	153	70	21.65	30	468.72	900.00	649.50	178.82	-25.82	2,253.70	666.84
38	199	90	67.65	50	4,576.52	2,500.00	3,382.50	210.47	-11.47	6,260.28	131.61
39	114	20	-17.35	-20	301.02	400.00	347.00	99.70	14.30	1,001.64	204.45
40	101	30	-30.35	-10	921.12	100.00	303.50	115.53	-14.53	250.41	210.99
41	82	10	-49.35	-30	2,435.42	900.00	1,480.50	83.88	-1.88	2,253.70	3.52
42	135	40	3.65	0	13.32	0.00	0.00	131.35	3.65	0.00	13.32
43	71	10	-60.35	-30	3,642.12	900.00	1,810.50	83.88	-12.88	2,253.70	165.81
44	123	40	-8.35	0	69.72	0.00	0.00	131.35	-8.35	0.00	69.72
45	196	80	64.65	40	4,179.62	1,600.00	2,586.00	194.65	1.35	4,006.58	1.83
46	90	10	-41.35	-30	1,709.82	900.00	1,240.50	83.88	6.12	2,253.70	37.49
47	180	50	48.65	10	2,366.82	100.00	486.50	147.17	32.83	250.41	1,077.52
48	183	50	51.65	10	2,667.72	100.00	516.50	147.17	35.83	250.41	1,283.47
49	62	10	-69.35	-30	4,809.42	900.00	2,080.50	83.88	-21.88	2,253.70	478.60
50	108	30	-23.35	-10	545.22	100.00	233.50	115.53	-7.53	250.41	56.63
51	158	60	26.65	20	710.22	400.00	533.00	163.00	-5.00	1,001.64	24.99
52	110	30	-21.35	-10	455.82	100.00	213.50	115.53	-5.53	250.41	30.53
53	181	60	49.65	20	2,465.12	400.00	993.00	163.00	18.00	1,001.64	324.04
54	84	10	-47.35	-30	2,242.02	900.00	1,420.50	83.88	0.12	2,253.70	0.02
55	200	70	68.65	30	4,712.82	900.00	2,059.50	178.82	21.18	2,253.70	448.46
56	90	10	-41.35	-30	1,709.82	900.00	1,240.50	83.88	6.12	2,253.70	37.49
57	183	90	51.65	50	2,667.72	2,500.00	2,582.50	210.47	-27.47	6,260.28	754.71
58	124	40	-7.35	0	54.02	0.00	0.00	131.35	-7.35	0.00	54.02
59	159	50	27.65	10	764.52	100.00	276.50	147.17	11.83	250.41	139.85
60	116	40	-15.35	0	235.62	0.00	0.00	131.35	-15.35	0.00	235.62
Sum	**7,881**	**2,400**	$\boxed{A}$	$\boxed{B}$	**127,122**	**44,700**	**70,735**	**7,881**	$\boxed{C}$	not provided	**15,188**

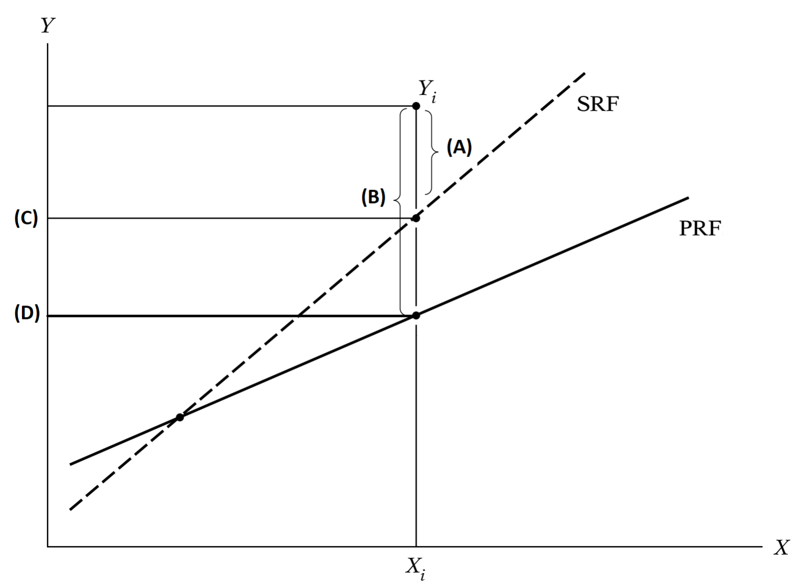
$\boxed{3\%}$ 4. Refer to Table I. What is the sample variance of X equal to? (i.e., what is the S^2 of X?)
A. 757.6
B. 2,154.6
C. 1,198.9
D. 257.4

$\boxed{3\%}$ 5. Refer to Table I again. What is the sample covariance of Y and X equal to? (i.e., what is the $\hat{\sigma}_{Y,X} = \widehat{Cov}(Y, X)$?)
A. 757.6
B. 1,897.2
C. 1,198.9
D. 257.4

$\boxed{3\%}$ 6. Refer to Table I again. What is $\hat{\beta}_1$ equal to?
A. 1.58
B. 0.63
C. 7.37
D. 1.80

$\boxed{3\%}$ 7. Refer to Table I again. What is the predicted value of y at $\overline{x}$ (i.e., what is the value of $\hat{y}_i$ when $x_i = \overline{x}$)?
A. 139.94
B. 93.33
C. 362.85
D. 131.35

Figure I



3% 8. Based on Figure I, choose the correct answer for observation i.

A. **(A)** is the error term, **(B)** is the residual, **(C)** is $\widehat{y}_i$, and **(D)** is $\widehat{\beta}_0$.
B. **(A)** is the error term, **(B)** is $\widehat{\beta}_0$, **(C)** is $\widehat{y}_i$, and **(D)** is the residual.
C. **(A)** is the residual, **(B)** is the error term, **(C)** is $\widehat{y}_i$, and **(D)** is $E(y|X = x_i)$.
D. **(A)** is the residual, **(B)** is the error term, **(C)** is $E(y|X = x_i)$, and **(D)** is $\widehat{\beta}_0$.

SECTION B - TRUE OR FALSE

3% 1. 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

3% 2. We say that we have nonexperimental data whenever we run controlled experiments, which is the **most** common way of obtaining data in economics.

◯ True ◯ False

3% 3. 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

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 $y = \beta_0 + \beta_1 x + u$, x is the unknown (populational) parameter to be estimated using data.

◯ True ◯ False

SECTION C - SHORT ANSWER

1. Assume that an online store has a customer rating system for the products that are sold on their website. After a customer buys any product online, s/he is invited to send his/her review about the product acquired. The customer can choose any integer number from 0 to 5 to give his/her overall evaluation of the product. This is done choosing the number of stars in the online store website. With these customer evaluations, the online store compute the following formula:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

where X_i is the number of stars given by the customer i. Then, the online store shows this customer rating score (i.e., $\overline{X}$) in their website for all products that they sell. This is done showing the number of stars, where the number of stars is equal to $\overline{X}$.

Now consider that for two products having the same features you see the following customer reviews:

(notice that the number to the right of the customer rating score $\overline{X}$ represents n, i.e., the number of customer reviews). Assuming that the customers review X_i, for $i = 1, 2, \ldots, n$ represents a random sample from the population, answer the questions below:

5% (a) If you are a customer comparing both products that have the same features, i.e, you can consider both of them roughly the same, and you want to base your decision to buy either product A or product B using the customer reviews rating (i.e., you are interested to know the true mean). Based on your statistics knowledge, which one would you choose? Give the name of the theorem that you are using in your answer. Disregard any possible price difference between both products. **[3 lines answer]**

5% (b) Explain in your own words this theorem. **[4 lines answer]**

2. **[This question refers to Regression (A).]**

Consider a random sample with the Grade Point Average (GPA), 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:
                   ---------------------------
                           log(score)
---------------------------------------------
colgpa                      0.2165***
                            (0.0113)

Constant                    3.6554***
                            (0.0325)

---------------------------------------------
Observations                   856
R2                           0.2992
Adjusted R2                  0.2984
Residual Std. Error      0.1780 (df = 854)
F Statistic          364.5649*** (df = 1; 854)
=============================================
Note:              *p<0.1; **p<0.05; ***p<0.01
```

$\boxed{5\%}$ (a) Using the variables names, write down the simple linear regression model. Using the variables names, write down the estimated OLS regression line (also known as SRF or SRL). **[2 lines answer]**

$\boxed{5\%}$ (b) How the model estimated in regression (A) is known (name)? **[1 line answer]**

$\boxed{5\%}$ (c) Interpret the results. Explain what is the effect in whichever is your dependent variable if college GPA increases one point? **[1-3 lines answer]**

$\boxed{5\%}$ (d) How many observations were used in the regression? What is the R^2 of the regression? Interpret the R^2 of the regression. **[1-2 lines answer]**

3. **[This question refers to Table 1 on the second page of your exam.]**

$\boxed{5\%}$ (a) Find the residual for observation 35, i.e., find the $\widehat{u}_{35}$ given in $\boxed{\text{D}}$? **[1-2 lines answer]**

$\boxed{5\%}$ (b) Find the fitted (or predicted) value for observation 25, i.e., find the $\widehat{y}_{25}$ given in $\boxed{\text{E}}$? **[1-2 lines answer]**

$\boxed{5\%}$ (c) Find the $\widehat{\beta}_0$ of this regression? **[1-2 lines answer]**

$\boxed{5\%}$ (d) Write down the estimated OLS regression line (also known as SRF or SRL). Interpret this regression. **[2-3 lines answer]**

$\boxed{5\%}$ (e) Find the R^2 of the regression. Interpret the R^2 of the regression. **[1-2 lines answer]**

$\boxed{5\%}$ (f) Find $\boxed{\text{A}}$, $\boxed{\text{B}}$ and $\boxed{\text{C}}$ located in the bottom of the table. Explain. **[2-3 lines answer]**

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

5% 4. EXTRA POINTS

Knowing that we can write $\widehat{\beta}_1 = \widehat{\rho}_{x,y}\dfrac{\widehat{\sigma}_y}{\widehat{\sigma}_x}$, where

- $\widehat{\rho}_{x,y}$ is the sample correlation between X and Y,
- $\widehat{\sigma}_y$ is the sample standard deviation of Y, and
- $\widehat{\sigma}_x$ is the sample standard deviation of X.

Using Table 1, calculate each one of these terms and show that $\widehat{\beta}_1 = \widehat{\rho}_{x,y}\dfrac{\widehat{\sigma}_y}{\widehat{\sigma}_x}$ is exactly the same value of the value found in question **(A-6)**.

Based on the above definition, conclude your answer completing the following sentence:

The sign of $\widehat{\beta}_1$ is always equal to the sign of the sample ________, i.e., if it is positive, then $\widehat{\beta}_1$ is ________, and if it is negative, then $\widehat{\beta}_1$ is ________.