The University of Kansas

Department of Economics

Final **May/13/2019**

Econ 526 - Introduction to Econometrics **Instructor: Caio Vigo Pereira**

Name:

SECTION A - MULTIPLE CHOICE

4% 1. For the past 3 months you verified that **every time** the price of stock A raised, the price of stock B raised, and **every time** the price of stock A dropped, the price of stock B dropped. Then, based on your data, what is the Corr(A, B)? based on Quiz 1, A-3

A. -1
B. 1
C. 0
D. 0.5

4% 2. Knowing that the estimator of the variance of the error term u given the explanatory variables $x_1, x_2, \ldots, x_k$, i.e., the estimator of $Var(u|x_1, x_2, \ldots, x_k)$ is given by:

$$\hat{\sigma}^2 = \frac{SSR}{df}$$

What is the *Residual Standard Error*: $\hat{\sigma}$? based on Quiz 6, C-1

A. $\sqrt{\dfrac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-k-1}}$
B. $\sqrt{\dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}}$
C. $\sqrt{\dfrac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n-k-1}}$
D. $\sqrt{\dfrac{\sum_{i=1}^{n}(\bar{y} - \hat{y}_i)^2}{n-k}}$

4% 3. Assume that the **Classical Linear Model (CLM)** assumptions hold. What is the distribution of $\dfrac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}$?

based on Quiz 7, A-6

A. t_{df}, where $df = n - k - 1$
B. $F_{(q,n-k-1)}$
C. $N(0, k^2)$
D. None of the above

4% 4. Consider a multiple linear regression model such as: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$. It is known that under the Gauss-Markov assumptions, the OLS estimators are BLUE. What "B" refers to?

A. That the OLS estimators have the smallest variance among the unbiased estimators
B. That $E(\hat{\beta}_j^{OLS}) = \beta_j$ for any $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$
C. That the OLS estimators have the smallest variance among all possible estimators
D. That the OLS estimators are consistent

4% 5. The ________ is used to compare across models that have different numbers of explanatory variables but where one is **not** a special case of the other (i.e., **nonnested models**).

A. R^2
B. *t test*
C. *Adjusted* R^2
D. *F test*

2.5% 6. EXTRA POINTS Among the statements below, which one is **NOT** under the *Classical Linear Model* assumptions?
A. the error term u is normally distributed
B. the error term u is independent of the explanatory variables
C. the error term u has mean 0
D. the variance of the error term u is a function of the explanatory variables

2.5% 7. EXTRA POINTS Which of the following can cause the usual OLS t statistics to be invalid (that is, not to have t distributions under the null hypothesis)?
A. Heteroskedasticity
B. Multicollinearity
C. Homoskedasticity
D. Exogenous variables

SECTION B - TRUE OR FALSE

3% 1. Let X and Y be two independent random variables. Then $Cov(X, Y) = 0$. based on Quiz 1, B-1
◯ True ◯ False

3% 2. We say that an estimator is unbiased if it has the smallest variance among all other estimators. based on Quiz 2, B-4
◯ True ◯ False

3% 3. Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables with mean μ, and variance σ^2. Consider the following estimator: $W = Y_1$. Then, W is a **biased** estimator of μ. based on Quiz 2, B-6
◯ True ◯ False

3% 4. The following regression model: $log(y) = \beta_0 + \beta_1 log(x_1) + u$ is also known as constant elasticity model. based on Quiz 4, B-4
◯ True ◯ False

3% 5. Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables with mean μ. The *Law of Large Numbers (LLN)* states that $\bar{Y}$ is an unbiased and efficient estimator of μ. based on Midterm, C-1
◯ True ◯ False

3% 6. Consider the following models:

$$\text{Model 1:} \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

$$\text{Model 2:} \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

Then, $SSR_{model1} \geq SSR_{model2}$. based on Quiz 5, B-3

◯ True ◯ False

3% 7. Exogenous explanatory variables is not a necessary assumption in order to the OLS estimator to be unbiased, however the assumption $E(u|x_1, \ldots, x_k) = 0$ is necessary.

◯ True ◯ False

3% 8. Large absolute t statistics are associated with **large** p-values.

◯ True ◯ False

3% 9. Multicollinearity violates the Gauss-Markov assumptions, and therefore the OLS estimators are not BLUE.

◯ True ◯ False

3% 10. Consider the following multiple linear regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

Assume that the 95% confidence interval for β_1 is $[-2.254, -1.723]$.
Therefore, $\hat{\beta}_1$ is statistically different from 0 at 5% significance level.

◯ True ◯ False

SECTION C - SHORT ANSWER

Consider a data set containing a random sample with salary information and career statistics for 353 players in the Major League Baseball (MLB). The dataset consists of the following variables (variable's name and description):

salary	1993 season salary measured in dollars
teamsal	team payroll measured in dollars
years	years in major leagues
games	career games played
atbats	career at bats
runs	career runs scored
hits	career hits
doubles	career doubles
triples	career triples
hruns	career home runs
hispan	=1 if hispanic
yrsallst	years as all-star
pcinc	city per capita income

1. (This question refers to **Regression (A)** below) Consider the following regression (*R* output) [**Notice that the significance level "stars" - *, **, *** - were suppressed in this output**]:

Regression (A)

	Dependent variable:
	salary
games	862.1545
	(149.1432)
pcinc	0.8880
	(20.4365)
teamsal	0.0208
	(0.0068)
yrsallst	233,250.3000
	(38,526.0700)
Constant	-79,709.5200
	(406,023.3000)
Observations	353
R2	0.4246
Adjusted R2	0.4180
Residual Std. Error	1,073,697.0000 (df = 348)
F Statistic	64.1907 (df = 4; 348)

4% (a) State the null hypothesis that the team payroll has no *ceteris paribus* effect on a baseball player salary. State the alternative hypothesis that there is an effect? [Two lines answer]

4% (b) Test the hypothesis stated above at the 1% significance level. Find the critical value. [Two lines answer]

4% (c) Do you reject the null hypothesis? Explain the statistical significance of your test at 1% significance level. [Two lines answer]

2% (d) Would you include *teamsal* in a final model explaining *salary* for players in the MLB? Why? Explain. [One line answer]

4% (e) Find the 99% confidence interval for β_{games}. [One line answer]

4% (f) Is the variable *games* statistically significant at 1% significance level? [One line answer]

2. (This question refers to **Regression (B)** below) Consider the following (additional) regression:

Regression (B)

	Dependent variable:
	salary
games	-944.9137*
	(538.2324)
pcinc	0.2409
	(19.7416)
teamsal	0.0203***
	(0.0065)
yrsallst	106,998.2000**
	(47,154.2200)
hruns	5,494.6750***
	(1,237.3210)
hits	1,498.2190***
	(561.1967)
Constant	54,356.1000
	(394,524.3000)
Observations	353
R2	0.4676
Adjusted R2	0.4584
Residual Std. Error	1,035,719.0000 (df = 346)
F Statistic	50.6544*** (df = 6; 346)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
Call:
lm(formula = salary ~ games + pcinc + teamsal + yrsallst + hruns +
    hits)

Residuals:
     Min       1Q   Median       3Q      Max
-4067915  -575667  -275468   354541  3618185

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.436e+04  3.945e+05   0.138  0.89050
games       -9.449e+02  5.382e+02  -1.756  0.08004 .
pcinc        2.409e-01  1.974e+01   0.012  0.99027
teamsal      2.031e-02  6.535e-03   3.109  0.00204 **
yrsallst     1.070e+05  4.715e+04   2.269  0.02388 *
hruns        5.495e+03  1.237e+03   4.441 1.21e-05 ***
hits         1.498e+03  5.612e+02   2.670  0.00795 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1036000 on 346 degrees of freedom
Multiple R-squared:  0.4676, Adjusted R-squared:  0.4584
F-statistic: 50.65 on 6 and 346 DF,  p-value: < 2.2e-16
```

2% (a) Which independent variables are statistically significant at 1% significance level. List their names. [*Hint:* no computation required.] [One line answer]

4% (b) Using the data from both regressions, state the null and alternative hypothesis that *hruns* and *hits* are **jointly** significant. Write down the unrestricted and the restricted model. [Four lines answer]

4% (c) Test the hypothesis stated above at the 1% significance level. Find the critical value. Test the same hypothesis again at the 5% significance level. Find the critical value. [Four lines answer]

4% (d) Do you reject the null hypothesis? Explain the statistical significance of your test at 1% significance level. [*Hint:* Don't forget to use a specific word when explaining the statistical significance.] [Four lines answer]

4% (e) Using **Regression (B)**, state the null and alternative hypothesis of the F statistic for overall significance of a regression. Do you reject the null hypothesis? Explain the statistical significance of your test at 1% significance level. [Three lines answer]

3. (This question refers to **Regression (C)** below).

Regression (C)

```
==============================================
                       Dependent variable:
                   ---------------------------
                           log(salary)
----------------------------------------------
hruns                        0.0025**
                             (0.0010)

hits                         0.0011***
                             (0.0001)

hispan                       -0.0879
                             (0.1244)

Constant                     12.7372***
                             (0.0730)

----------------------------------------------
Observations                   353
R2                            0.4366
Adjusted R2                   0.4318
Residual Std. Error      0.8913 (df = 349)
F Statistic          90.1644*** (df = 3; 349)
==============================================
Note:              *p<0.1; **p<0.05; ***p<0.01
```

3%

(a) What is the estimated average difference in salary between being *hispanic* or not, for players with the same number of hits and home runs? Show your answer using **(i) approximation** and the **(ii) precise estimated average difference**. [*Hint: Notice that the dependent variable is in log*] [Three lines answer]

2%

(b) All other factors being equal, is there any statistical evidence that being a *hispanic* impacts the salary of a MLB player? Consider 1% significance level in your answer. [Two lines answer]

4. **[Gauss-Markov Theorem]**

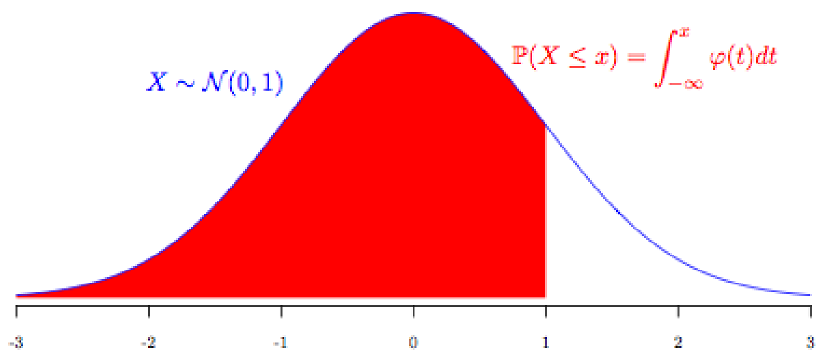
5%

(a) Under which assumptions does the Gauss-Markov theorem holds? State and briefly explain each one of them. [One line answer per assumption]

5%

(b) What does the acronym "BLUE" stands for? [Two lines answer]

Standard Normal Distribution



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Critical Values of the t-distribution

		Significance Level				
1-Tailed:		**.10**	**.05**	**.025**	**.01**	**.005**
2-Tailed:		**.20**	**.10**	**.05**	**.02**	**.01**
	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
Degrees of Freedom	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845
	21	1.323	1.721	2.080	2.518	2.831
	22	1.321	1.717	2.074	2.508	2.819
	23	1.319	1.714	2.069	2.500	2.807
	24	1.318	1.711	2.064	2.492	2.797
	25	1.316	1.708	2.060	2.485	2.787
	26	1.315	1.706	2.056	2.479	2.779
	27	1.314	1.703	2.052	2.473	2.771
	28	1.313	1.701	2.048	2.467	2.763
	29	1.311	1.699	2.045	2.462	2.756
	30	1.310	1.697	2.042	2.457	2.750
	40	1.303	1.684	2.021	2.423	2.704
	60	1.296	1.671	2.000	2.390	2.660
	90	1.291	1.662	1.987	2.368	2.632
	120	1.289	1.658	1.980	2.358	2.617
	∞	1.282	1.645	1.960	2.326	2.576

Source: Wooldridge, Jeffrey M. *Introductory Econometrics*, 2015.

1% Critical Values of the F Distribution

	Numerator Degrees of Freedom									
Denominator Degrees of Freedom	1	2	3	4	5	6	7	8	9	10
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Source: Wooldridge, Jeffrey M. *Introductory Econometrics*, 2015.

5% Critical Values of the F Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom 1	2	3	4	5	6	7	8	9	10
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Source: Wooldridge, Jeffrey M. *Introductory Econometrics*, 2015.