Universidade de São Paulo / Faculdade de Filosofia, Letras e Ciências Humanas

Departamento de Ciência Política

FLP-0468 & FLS-6183

2º semestre / 2016

Prova Parcial

Exercício 1. (10 pontos) Which of the following statements are accurate about the population regression model?

(a) u_i is the stochastic component of Yi.

(b) $\hat{\alpha} + \hat{\beta} X_i$ is the systematic component of Yi.

(c) Both (a) and (b) are correct.

(d) Neither (a) nor (b) are correct.

Answer: (c)

Exercício 2. (10 pontos) True or False. In a situation where we estimate a regression that includes X and Y but improperly exclude Z, the estimated effect of X on Y will be equal to the true effect of X on Y, so long as the sample was chosen randomly. Please justify your answer.

Answer: False. This is a case of omitted variable bias. The estimated effect of X on Y will be equal to the true effect of X on Y plus the product of the effect of Z on Y and the bivariate association between X and Z. There are two exceptions: 1) If Cov (X,Y) =0; or 2) $\beta_z = 0$.

Exercício 3. (10 pontos) True or False. "High multicollinearity simply means that there is not enough information in the data to estimate the model parameters accurately and the standard errors rightfully reflect this." Please justify your answer.

Answer: True.

As Gujarti notes, "multicollinearity violates no regression assumptions. Unbiased, consistent estimates will occur, and their standard errors will be correctly estimated. The only effect of multicollinearity is to make it hard to get coefficient estimates with small standard error (326)."

He goes on to elaborate "*multicollinearity is essentially a sample (regression) phenomenon* in the sense that, even if the *X* variables are not linearly related in the population, they may be so related in the particular sample at hand: When we postulate the theoretical or population regression function (PRF), we believe that all the *X* variables included in the model have a separate or independent influence on the dependent variable *Y*. But it may happen that in any given sample that is used to test the PRF some or all of the *X* variables are so highly collinear that we cannot isolate their individual influence on *Y*. So to speak, our sample lets us down, although the theory says that all the *X*'s are important. In short, our sample may not be "rich" enough to accommodate all *X* variables in the analysis."

Students should note that this also is correct in the case of interaction models as is argued by Brambor, Clark and Golder. In their article, ""Understanding Interaction Models: Improving Empirical Analyses," Brambor, Clark and Golder state "Even if there really is high multicollinearity and this leads to large standard errors on the model parameters, it is important to remember that these standard errors are never in any sense ''too'' large—they are always the ''correct'' standard errors. High multicollinearity simply means that there is not enough information in the data to estimate the model parameters accurately and the standard errors rightfully reflect this (70)."

Exercício 4. (40 pontos) No artigo "*Benchmarking across Borders: Electoral Accountability and the Necessity of Comparison*", Kayser e Peress (2012) reportam, na Tabela 1, os seguintes resultados de regressões baseadas em uma amostra de 22 países da OCDE e 213 eleições realizadas entre 1948 e 2008:

TABLE 1. Aggregate-level Results for Benchmarking in the Economic Vote

		Decomposition Method		
	(1)	(2) Median	(3) PC	(4) Trade
Independent Variables				
Constant	34.095	37.446	35.069	32.764
	(1.699)***	(3.944)***	(1.997)***	(4.372)***
Growth	0.604			
	(0.267)*			
Local growth		0.818	1.261	0.820
		(0.331)*	(0.352)***	(0.404)*
Global growth		0.098	−0.044	0.904
		(0.485)	(0.364)	(0.614)
Unemployment	−0.248			
	(0.206)			
Local unemployment		−0.210	−0.335	0.027
		(0.210)	(0.229)	(0.211)
Global unemployment		−0.590	−0.130	−0.277
		(0.523)	(0.264)	(0.541)
***p*-Value from Wald Test**				
Local growth = global growth		0.233	0.006**	0.905
Local unem. = global unem.		0.471	0.422	0.576
Number of elections	213	213	213	146
Number of countries	22	22	22	22
R^2	0.029	0.035	0.061	0.034

Notes: Heteroskedasticity robust standard errors in parentheses. All results are restricted to OECD countries. We obtained nearly identical results when standard errors were clustered by country.
*5.0% significance level; **1.0% significance level; ***0.1% significance level.

a) (10 pontos) Os resultados reportados na Coluna 1 referem-se ao seguinte modelo:

$$vote = \alpha + \beta_1\ growth + \beta_2\ unemployment + \varepsilon \qquad \text{(Modelo 1)}$$

Onde:

- *vote* = % de votos recebidos pelo partido incumbente;
- *growth* = crescimento do PIB real (em %);
- *unemployment* = taxa de desemprego (em %)

Por favor discuta o que os resultados reportados na **Coluna 1** da Tabela 1 nos dizem a respeito da hipótese do voto econômico que está sendo testada pelos autores.

Para melhor auxiliar na elaboração de sua resposta, o resultado do *Stata* para o Modelo 1 foi reproduzido abaixo:

```
. reg votelead gr_an unem_an, robust

Linear regression                                      Number of obs =      213
                                                       F(  2,    210) =     3.39
                                                       Prob > F      =   0.0356
                                                       R-squared     =   0.0289
                                                       Root MSE      =   10.486
```

votelead	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
gr_an	.6043563	.2672755	2.26	0.025	.0774695	1.131243
unem_an	-.2484676	.2055033	-1.21	0.228	-.6535814	.1566463
_cons	34.0951	1.698902	20.07	0.000	30.74601	37.44418

In explaining the findings from this regression, students should be careful to explain what insights we can draw regarding the expected incumbent party vote when growth and unemployment are both zero (the meaning of the intercept, its prediction and statistical significance), as well as the interpretation of the null hypothesis test for each variable (growth and unemployment).

In their comments in the article on this regression, Kayser and Peress state that "Substantively, a 1% increase in growth leads to a 0.604% increase in the leader party's vote share, and 1% increase in unemployment corresponds to a 0.248% decrease in the leader party's vote share (668)." However, unemployment is not statistically significant at the 1,5 or 10% level.

b) (10 pontos) Os resultados reportados na Coluna 2 da Tabela 1 podem ser resumidos por:

$$vote = \alpha + \beta_1 \, local\ growth + \beta_2 \, global\ growth + \\ + \beta_3 \, local\ unemployment + \beta_4 \, global\ unemployment + \varepsilon \qquad \text{(Modelo 2)}$$

Onde:

- *vote* = % de votos recebidos pelo partido incumbente para o país *p* e para a eleição *e*;
- *local growth* = (mediana do) crescimento do PIB real do país *p* e eleição *e* (em %);
- *global growth* = (mediana do) crescimento do PIB real dos 22 países na eleição *e* (em %);
- *local unemployment* = (mediana da) taxa de desemprego para o país *p* e eleição *e* (em %);
- *global unemployment* = (mediana da) taxa de desemprego para os 22 países na eleição *e* (em %)

Por favor discuta o que os resultados reportados na **Coluna 2** da Tabela 1 nos dizem a respeito da hipótese do voto econômico que está sendo testada pelos autores.

Para melhor auxiliar na elaboração de sua resposta, o resultado do *Stata* para o <u>Modelo 2</u> foi reproduzido a seguir:

```
. reg votelead gr_loc_med_an gr_glob_med_an unem_loc_med_an unem_glob_med_an, robust

Linear regression                                      Number of obs =     213
                                                       F(  4,   208) =    1.95
                                                       Prob > F      =  0.1028
                                                       R-squared     =  0.0351
                                                       Root MSE      =  10.503

------------------------------------------------------------------------------
                 |             Robust
        votelead |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
   gr_loc_med_an |   .8184615   .3314926     2.47   0.014      .1649454    1.471978
  gr_glob_med_an |   .0977744   .4850076     0.20   0.840     -.8583864    1.053935
 unem_loc_med_an |  -.2095095   .2095363    -1.00   0.319     -.6225966    .2035777
unem_glob_med_an |   -.590107   .5234412    -1.13   0.261     -1.622037    .4418231
           _cons |   37.44574   3.943787     9.49   0.000      29.67082    45.22066
------------------------------------------------------------------------------
```

In explaining the findings from this regression, students should be careful to explain what insights we can draw regarding the expected incumbent party vote when median local and median global economic growth and median local unemployment and median global unemployment are zero. In addition to examining the intercept (its prediction and statistical significance), students should discuss the interpretation of the null hypothesis test for each variable (local median growth, global median growth, local median unemployment and global median unemployment).

As Kayser and Peress explain, "Local growth has a positive and statistically significant effect, whereas global growth has a statistically insignificant effect. This strongly suggests that voters respond to their country's deviation from various measures of average international performance, but not to the international benchmark itself."

With respect to local unemployment or global employment, voters do not seem to benchmark on either type of unemployment.

The F-Test, which tests the null hypothesis that all coefficient estimates are jointly equal to zero), can not be rejected at a confidence level of 90%. Students should have noted this result with concern.

An answer that received full credit also elaborated on discussing how the findings correspond to economic voting theory and the substantive (versus statistical) significance of the estimates obtained from Model 1.

c) (10 pontos) Sobre a Tabela 1 de Kayser e Peress (2012:669), um colega nota que

> "*A diferença do tamanho dos efeitos entre as coluna (1) e coluna (2) chama atenção. De acordo com o Modelo (1), um aumento de 1% no crescimento está associado a um aumento de 0,604% na porcentagem de votos do partido incumbente. De acordo com o Modelo (2), um aumento de 1% no crescimento local está associado a um aumento de 0,81% na porcentagem de votos do líder do partido. O tamanho do efeito estimado, então, aumenta quando se compara os dois modelos e o modelo 2 apresenta efeitos maiores.*"

Você está inseguro a respeito da afirmação de seu colega. Que tipo de análise pode ser útil para melhor julgar se os resultados reportados são *estatisticamente* significativos? Por favor, forneça uma resposta detalhada com exemplos de procedimentos e exemplos de como você julgaria se os resultados são ou não *estatisticamente* significativos.

The two variables are not directly comparable. Model 1 measures the effect of growth whereas Model 2 measures the effect of local median economic growth. But, if we buy the argument of Kayser and Peres (2012) that these two coefficients can be compared, we still should be very concerned about the statement the claim that the size of the effect increases and is greater in Model 2. To assess if the difference is statistically significant, Gelman and Stern (2007) would argue that we need to carry out a test to assess if the differences are statistically significant by calculating: z = (B1 - B2) / √(seB1^2 + seB2^2). In terms of magnitude, we could also compare both results by examining the standardized beta coefficients.

d) (10 pontos) Que tipo de análise pode ser útil para melhor julgar se os resultados reportados são *substancialmente* significativos? Por favor, forneça uma resposta detalhada com exemplos de procedimentos e exemplos de como você julgaria se os resultados são ou não *substancialmente* significativos.

If we want to show that we have identified a variable that has substantive importance, we should show that changes in local economic growth are decisive in explaining election outcomes. Based on the results obtained, we could estimate the results and assess if they suggest that changes in local economic conditions are critical in shaping election outcomes in the sample of 22 countries. It would be helpful to calculate the predicted incumbent vote share based on the most likely values for local growth, for example.

We will show in class this week that Clarify is a type of method to aid in assessing substantive significance. Clarify helps us to present substantively meaningful measures of effect magnitude.

Exercise 5. (30 pontos) The regression output and the table below reports the results of the regression results based on survey data from the American National Election Study of 2004.

The dependent variable is an additive index of support for the social welfare state ranging from 0 (least supportive) to 1 (most supportive). The explanatory variables are an indicator variable for gender (1 if Female, 0 if other), an indicator of partisanship (1 if Republican, 0 if Democrat) and the interaction of these two variables.

The model can be summarized as:

$$\text{WelfareSupport} = \alpha + \beta\text{Female} + \delta\text{Republican} + \lambda\text{Female} \times \text{Republican} + \varepsilon$$

```
. regress socwel female Republican femRep

      Source |       SS           df       MS      Number of obs   =     1077
-------------+----------------------------------   F(  3,   1073)  =   103.65
       Model |  9.35811969         3  3.11937323   Prob > F        =   0.0000
    Residual |  32.2907398      1073  .030093886   R-squared       =   0.2247
-------------+----------------------------------   Adj R-squared   =   0.2225
       Total |  41.6488595      1076  .038707118   Root MSE        =   .17348

------------------------------------------------------------------------------
      socwel |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -.0031436    .014435    -0.22   0.828    -.0314676    .0251805
  Republican |  -.2204654   .0155474   -14.18   0.000    -.2509722   -.1899586
      femRep |   .0836563   .0213704     3.91   0.000     .0417238    .1255888
       _cons |   .7451486   .0109716    67.92   0.000     .7236204    .7666768
------------------------------------------------------------------------------
```

a) Please explain what the authors are seeking to test in the model and the interpretation of the λ coefficient.

Answer:. The authors are testing if being a woman increases support for the welfare state (beta), if being a Republican increases support for the welfare state (delta); and if the gender gap for social welfare support is contingent on partisanship (lambda).

Given the interaction, we must assess these hypothesis taking into account the interaction.

Thus, the effect of gender on welfare support:

$$\frac{\partial \text{WelfareSupport}}{\partial \text{Female}} = \beta + \lambda\text{Republican}$$

In turn, the effect of Republican on welfare support:

$$\frac{\partial \text{WelfareSupport}}{\partial \text{Republican}} = \delta + \lambda\text{Female}$$

b) Based on the regression model, we have also calculated the predicted support for social welfare under different conditions. The results based on two estimations in Stata are summarized below. Based on the results, what can we conclude?

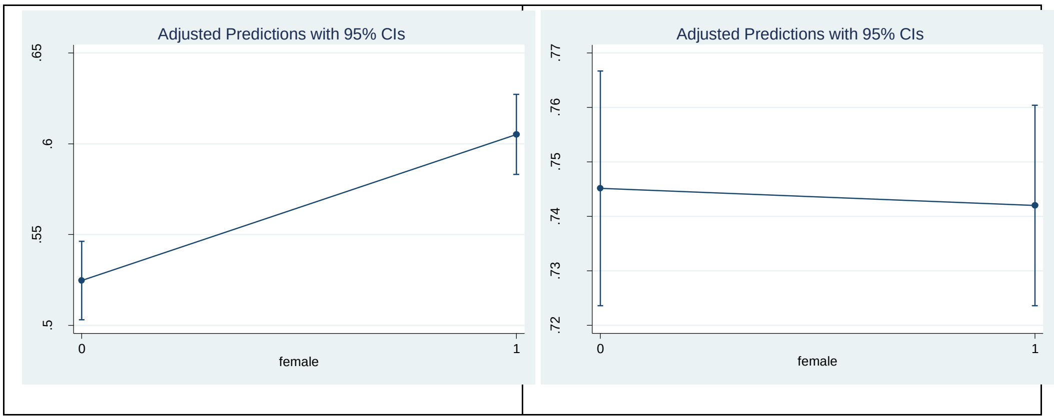
Answer: We are testing if the gender gap for social welfare support is contingent on partisanship. Based on the results, there is little difference in social welfare support of female and male Democrats, but there is a gender gap between female and male Republicans. Thus, the gender gap for social welfare support is contingent on partisanship. We can see these results by comparing the confidence intervals in the table below.

The Predicted Effect of Gender and Partisanship on Welfare Support
(95% Confidence Intervals in Parenthesis)

	Democrat	Republican
Female	0.74 (0.72, 0.76)	0.60 (0.58, 0.62)
Male	0.74 (0.72, 0.76)	0.52 (0.50, 0.54)

```
. margins,  at( female =(0 1) Republican =1)

Adjusted predictions                              Number of obs   =       1077
Model VCE     : OLS

Expression    : Linear prediction, predict()

1._at         : female          =           0
                Republican      =           1

2._at         : female          =           1
                Republican      =           1

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .5246832   .0110157    47.63   0.000     .5030684     .546298
          2  |   .6051959   .0112685    53.71   0.000     .5830852    .6273066
------------------------------------------------------------------------------

. margins,  at( female =(0 1) Republican =0)

Adjusted predictions                              Number of obs   =       1077
Model VCE     : OLS

Expression    : Linear prediction, predict()

1._at         : female          =           0
                Republican      =           0

2._at         : female          =           1
                Republican      =           0

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         _at |
          1  |   .7451486   .0109716    67.92   0.000     .7236204    .7666768
          2  |    .742005   .0093805    79.10   0.000     .7235988    .7604112
------------------------------------------------------------------------------
```

c) The same results are summarized below with two graphs. Are your conclusions in (b) confirmed?



Answer: Yes. In the graph on the left, there is a difference in the effect of being a female on welfare support among Republicans. In the graph on the right, there is not a difference in the effect of being a female on welfare support among Democrats.

Exercício 6. (30 pontos) Below please find three different models and the partial effects derivatives that show how changes in each explanatory variable influence changes in the dependent variable. Please explain the difference between the following three models in terms of which interactions are being tested and concentrate your discussion only on X (Hint: draw Venn diagrams if helpful):

Model 1:

$$y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_6 ZW + \varepsilon$$

$$\frac{\partial y}{\partial x} = \beta_1 + \beta_4 Z$$

$$\frac{\partial y}{\partial z} = \beta_2 + \beta_4 X + \beta_6 W$$

$$\frac{\partial y}{\partial w} = \beta_3 + \beta_6 Z$$

Model 2:

$$y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_5 XW + \beta_6 ZW + \varepsilon$$

$$\frac{\partial y}{\partial x} = \beta_1 + \beta_4 Z + \beta_5 W$$

$$\frac{\partial y}{\partial z} = \beta_2 + \beta_4 X + \beta_6 W$$

$$\frac{\partial y}{\partial w} = \beta_3 + \beta_5 X + \beta_6 Z$$

Model 3:

$$y = \alpha + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_5 XW + \beta_6 ZW + \beta_7 XZW + \varepsilon$$

$$\frac{\partial y}{\partial x} = \beta_1 + \beta_4 Z + \beta_5 W + \beta_7 ZW$$

$$\frac{\partial y}{\partial z} = \beta_2 + \beta_4 X + \beta_6 W + \beta_7 XW$$

$$\frac{\partial y}{\partial w} = \beta_3 + \beta_5 X + \beta_6 Z + \beta_7 XZ$$

Answer: The difference between models 1 and 2 is that in 2 we are testing the hypothesis that the effect of X on Y depends on W and Z (separately) whereas in 1 we hypothesize that the effect of X on Y depends only on the value of Z. In model 3, we test the hypothesis that the effect of X on Y depends on the value of combination of the values of Z and W. Model 3 is the only model in which there is a three-way interaction.

It is important to note that there is nothing wrong with the specification in model 2 in which the XWZ term is not included. The constitutive term is only necessary if theory posits that such three way interaction is necessary.