**Universidade de São Paulo**

**Faculdade de Filosofia, Letras e Ciências Humanas**

**Departamento de Ciência Política**

1º semestre / 2022

Prova Parcial

**Instruções**

Esta prova representa a avaliação parcial desta disciplina. Ela deve ser feita individualmente; porém, é esperado que vocês conversem e discutam a respeito destes exercícios. Não será tolerada, sob hipótese alguma, qualquer forma de plágio.

A primeira parte deve ser respondida obrigatoriamente pelos alunos da graduação e pelos alunos da pós-graduação. A segunda parte deve ser feita apenas pelos alunos da pós-graduação.

A entrega será feita via *moodle* até as 23:59h do dia 23/10/2022 através do link adequado criado para tal. Não serão recebidas as respostas por nenhum outro meio sob nenhuma justificativa.

Cada discente deverá enviar um arquivo com o próprio nome e número USP tanto no título do arquivo, como na primeira página do arquivo com as respostas. O modelo a ser seguido para o nome do arquivo é *Glauco Peres\_NUSP 1104679*. Aqueles que por ventura não tenham o número USP, utilizem apenas o próprio nome. Serão aceitos arquivos em formato .doc, .docx ou .pdf.

Sejam detalhistas, claros e específicos em suas respostas, mostrando as etapas dos raciocínios envolvidos. Não apresentem apenas resultados, mas utilizem comentários e discussões envolvidas. Não apresentem saídas diretas dos softwares, se for o caso, mas utilizem tabelas ou gráficos que ajudem na compreensão dos resultados quando for conveniente. Não há uma única maneira de responder cada exercício e, assim, espera-se que vocês desenvolvam as respostas de maneira ampla.

Bom trabalho!

**1ª Parte- Graduação e Pós-graduação**

Considere a seguinte tabela com os resultados de quatro modelos distintos de regressão. A variável dependente é uma variável Y não especificada aqui.

Tabela 1 – Resultados da Regressão – Variável Dependente: Y



Note que a tabela está preenchida com letras em alguns lugares ao invés de algarismos. Para converter estas letras em números, é preciso tomar o seu número USP e proceder da forma a seguir (aqueles que não possuem número USP podem usar o RG):

Considere o seguinte número USP: 12345678

13485660

Associamos para cada letra um dos numerais em ordem: a = 1; b = 2; c = 3; e assim sucessivamente até h = 8. Nos termos da tabela em que constam as letras, você deve substituir pelo algarismo correspondente. Como exemplo, para a expressão da primeira linha do modelo 2 [**a**, (**b**+1)2], o termo deve ser substituído por: 1,(2+1)2 = 1,32.

1. Assim, antes de iniciar, você deve reescrever a tabela acima com todos os algarismos completados. Note que é preciso utilizar 8 algarismos em 6 valores diferentes.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | (1) | (2) | (3) | (4) |
| X1 | 1,30 | 1,42 | 3,61 | 1,44 |
|  | (0,02)\*\*\* | (0,01)\*\*\* | 1,30 | (0,02)\*\*\* |
| X2 |  | -4,12 | 6,60 | -3,85 |
|  |  | (1,90)\* | (6,7) | (1,02)\*\*\* |
| X3 |  |  | -0,15 |  |
|  |  |  | (0,20) |  |
| X4 |  |  |  | 2,45 |
|  |  |  |  | (0,25)\*\*\* |
| Constante | 2,10 | 1,54 | -3,29 | 1,29 |
|  | (4,85) | (0,56)\*\* | (2,81)\* | (0,01)\*\*\* |
| N | 150 | 150 | 150 | 150 |
| R2 | 31% | 35% | 87% | 41% |

A partir do que é apresentado nesta tabela, responda às questões a seguir:

1. Escreva a equação que representa o único modelo bivariado estimado. Interprete-a;
2. O que significam os asteriscos ao lado do termo em parênteses? Apresente o teste de hipóteses que justifica esta indicação. Interprete o resultado obtido para o coeficiente estimado de X1 para o primeiro modelo;
3. Explique também o que significam os termos apresentados em parênteses em toda a tabela. Explique em que situação se deve apresentar os valores calculados desta forma;
4. Corrija o nível de significância do parâmetro estimado para o termo da constante do primeiro modelo, mostrando como chegou ao resultado;
5. Escreva a equação para o 2º modelo apresentado. Interprete cada um dos parâmetros estimados;
6. Como é possível comparar este resultado com o obtido no primeiro modelo? Apresente um teste que permita dizer se o novo parâmetro se alterou em relação ao primeiro modelo. Discuta;
7. O que significa o termo R2? Interprete o valor encontrado para o segundo modelo;

No 3º modelo indicado na tabela, há dois problemas. O primeiro é uma violação clássica de uma das hipóteses do modelo de MQO; o segundo parece ser um erro de digitação a respeito da significância de um dos parâmetros estimados.

1. Explique qual é a violação mencionada e quais são as evidências que suportam a sua interpretação. Quais são os impactos desta violação na interpretação dos resultados?
2. Como podemos identificar este erro de digitação? Discuta.
3. Quais efeitos produzem a introdução de uma nova variável explicativa em um modelo de regressão? Discuta utilizando os resultados obtidos no 4º modelo em comparação com os demais.

**2ª parte – Pós-graduação**

Para responder esta parte, deve-se tomar o banco de dados que está junto com este arquivo de enunciado para a prova (*Base\_prova\_parte2.csv*).

A variável dependente é *votos* e as demais colunas são todas variáveis explicativas.

Deve-se selecionar uma amostra aleatória de tamanho igual a 500 para esse exercício. Isto deve ser feito utilizando dois comandos no R: set.seed() e o sample\_n().

Para o primeiro comando, deve-se novamente utilizar o número USP como o seed. Para um número USP como o do exemplo anterior, deve-se digitar no início do código: *set.seed(12345678)* e em seguida *sample\_n(500)*.

De posse deste novo banco de dados, deve-se buscar encontrar um modelo de bom ajuste para a previsão dos votos. O livro de código das variáveis está resumido a seguir:

|  |  |
| --- | --- |
| Variável | Significado |
| Votos | Votação total recebida por um candidato a deputado federal na eleição no ano t |
| financ | Valor (em milhares) declarado pelo candidato da arrecadação da campanha |
| idade | Idade do candidato (em anos) no dia 31/07 do ano t |
| Fpart | Proporção dos recursos de campanha do candidato recebida do partido |
| Votos\_t1 | Votação total recebida por um candidato a deputado federal na eleição no ano t-1. |

Em sua resposta, você deve apresentar o modelo de MQO que melhor se adéqua aos dados selecionados. Se for necessário, considere a situação de apresentar mais de um modelo. Apresente também gráficos e estatísticas descritivas das variáveis que contribuam na sua decisão sobre qual o melhor modelo. Teste para a validade das hipóteses do modelo de MQO e mostre os resultados obtidos. Apresente ao final as primeiras 25 linhas do seu banco de dados também.