**Universidade de São Paulo**

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

**Departamento de Ciência Política**

**FLS5208 - Métodos Quantitativos e Técnicas em Ciência Política**

**FLP0406 – Métodos e Técnicas de Pesquisa em Ciência Política**

**1º semestre / 2019**

Prof. Glauco Peres da Silva

**Laboratório nº 5: Estatística Descritiva & Amostragem**

08 de abril (not) e 10 de abril (vesp)

No laboratório de hoje, o objetivo é avaliar a importância do processo de amostragem para os resultados empíricos que obtemos. Os trabalhos devem ser feitos, numa primeira etapa, individualmente, e depois em grupos de 3 ou 4 pessoas.

No estudo sobre eleições, teoricamente, espera-se que o montante de recursos gastos em uma campanha ajude o candidato a se eleger em um contexto no qual os partidos não são atalhos informacionais fortes para os eleitores no momento de sua decisão (p.ex., Freille, 2015). Eleitores se guiariam pela campanha, considerando o que os candidatos veiculam durante este período. Neste sentido, é esperado encontrarmos uma relação positiva entre ambos eventos: quanto maior o valor do financiamento da campanha, maior o número de votos recebido pelos candidatos.

Há um banco de dados no moodle para esta atividade. Ele é composto de 5 variáveis: ID - código de identificação do candidato a deputado federal em um dado estado; Num\_votos – o número de votos total recebido pelo candidato em uma eleição; Financiamento – valor total gasto em campanha naquela campanha; Eleito – *dummy* indicando 1 se o candidato se elegeu; 0 em caso contrário; e Região – variável categórica indicando duas regiões do distrito eleitoral, A e B.

Neste laboratório, o primeiro passo é utilizar uma amostra desta base de dados. Para isto, tome seu número USP e some os algarismos (Ex.: 1154668 => 1 + 1 + 5 + 4 + 6 + 6 + 8 = 31). Este será o número de observações que você selecionará da base de dados que chamaremos de Selecione as primeiras observações deste banco.

1. Calcule estatísticas descritivas das variáveis nesta amostra.
2. Calcule o coeficiente de correlação entre as variáveis (pode usar o Excel em sua fórmula CORREL). O valor está de acordo com o esperado teoricamente?

Não, na verdade, apresentou um pouco de correlação negativa/inversamente proporcional.

COR = -0.03690861

Refaça o processo de seleção, mas agora, escolha deste banco a mesma quantidade de observações anteriores. Chamaremos essa amostram agora de .

1. Calcule as mesmas estatísticas descritivas do item a). São iguais às anteriores? Quais seriam razões para uma eventual diferença?

Não, os valores diferem. As amostras foram recolhidas através de processos diferentes.

1. Calcule novamente o índice de correlação. Agora o valor está de acordo com o esperado teoricamente? É possível chegarmos a alguma conclusão sobre a teoria com esta observação?

A correlação para a amostra n1 foi de 0,716. Bem diferente da amostra anterior. Com esse resultado, a hipótese da teoria tende a ser válida, indicando que há uma correlação diretamente proporcional.

Como vocês sabem, o processo de amostragem é bastante caro à avaliação que somos capazes de fazer utilizando dados empíricos. É um processo que pode gerar bastante viés na análise e contaminar o resultado obtido. A amostra não pode velar tendências desconhecidas do pesquisador, pois isto conduzirá a um resultado que não é válido.

1. O primeiro processo de seleção gerou algum viés? E o segundo? Por que?

Como a planilha foi organizada em ordem crescente de votos, a primeira amostra está enviesada com apenas os menores. Já a segunda foi utilizada a aleatorização para amostragem, removendo o viés da ordem da planilha.

Há uma dimensão importante que é a aleatoriedade na seleção da amostra. É ela que evitará o viés amostral.

1. Por que a aleatoriedade é importante?

Ela é importante, pois cada observação tem a mesma chance de estar na amostra a ser recolhida. Assim, não há viés na escolha da amostra.

1. Forme um grupo de 3 ou 4 indivíduos. Dentro destes grupos, discutam quais procedimentos foram utilizados por vocês para a seleção dos casos para a formação da amostra . Não é preciso comparar resultados; devem comparar apenas os procedimentos.

Ao selecionar seus dados, surgiram (ou ao menos deveriam ter surgido) questionamentos a respeito de como se escolhe elementos de uma população. Há ao menos duas questões fundamentais: a *ordem* em que os dados são observados e se os dados foram coletados com ou sem *reposição*.

1. Qual a importância destas questões ao se selecionar uma amostra? Discuta em grupo a validade de seus procedimentos diante deles.

A ordem não tem importância. Mas ao coletar dado com reposição há o risco de perder representatividade da amostra.

Ordem de sorteio quase nunca importa na nossa disciplina. Na maior parte das vezes é quando relacionado em que a ordem do sorteio importa para o resultado final, por exemplo, letras. Assim, a ordem importa no número de sorteios possíveis. Por exemplo: sorteio de 3 letras para placa, se a ordem importa então são placas diferentes: AAB, ABA e BAA.

* Sem Reposição:
  + Garante I.I.D. (Variáveis aleatórias independentes e identicamente distribuídas)
  + Vantagens
    - Informações
  + Desvantagens
    - Chances diferentes de sorteio
* Com Reposição
  + Vantagem
    - Mesmas chances
  + Desvantagens
    - Informações/Representatividade

Com reposição só em casos específicos. Na maioria das vezes iremos lhe dar como as populações forem infinitas, e assim usamos na maioria das vezes usamos sem reposição. Fator de correção de populações finitas: recurso para Com Reposição.

Porém o conceito de variável independente é muito caro, pois, por exemplo, o ato de candidatos gastarem mais com campanha influenciam outros candidatos a fazerem o mesmo, assim, não é exatamente independente.

1. Selecionem, então, em conjunto, 35 observações desta amostra respeitando o que debateram no item anterior. Calcule as estatísticas descritivas agora e o índice de correlação. Calcule as mesmas estatísticas para o banco de dados completo, inclusive a correlação. Compare os resultados.