**Universidade de São Paulo**

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

**Departamento de Ciência Política**

**Métodos Quantitativos e Técnicas em Ciência Política**

1º semestre / 2021

**Lab #9. Inferência e Significância Estatística**

**Parte I – Retomando semana passada**

No lab da semana passada, se pedia na parte II que a turma respondesse às questões a respeito da construção do intervalo de confiança. O exercício todo será reproduzido aqui, mas vocês NÃO precisarão responde-lo; vocês terão apenas que discutir a questão a seguir:

Como seriam os diagramas dos exercícios 5 e 8? Qual seria a diferença entre eles?

REPRODUÇÃO DO EXERCÍCIO

“Nesta atividade, vamos nos concentrar em tentar compreender os conceitos da teoria das probabilidades, que enfatiza o entendimento de por que o intervalo de confiança é uma quantidade aleatória, sujeita à variabilidade de amostragens[[1]](#footnote-1).

A tabela abaixo é um resumo que irá ajudá-lo a compreender os exemplos desenvolvidos neste laboratório.

|  |
| --- |
| **Tabela 1. Resumos dos Métodos de Estimação para a Média e a Proporção para Populações Infinitas** |
| **Parâmetro** | **Ponto Estimado** | **Erro padrão Estimado** | **Intervalo de Confiança**  |
| Média | $$\overbar{y}=\frac{1}{n}\sum\_{i=1}^{n}y\_{i}$$ | $ $ $s\_{\overbar{y }}=\frac{s\_{y}}{\sqrt{n}}$, onde $s\_{y}=\sqrt{\frac{\sum\_{i=1}^{n}(y\_{i}-\overbar{y})^{2}}{n-1}}$ |  |
| Proporção | $$\hat{π}$$ | $s\_{\hat{π}}=\sqrt{\frac{\hat{π}(1-\hat{π})}{n}}$  | $$\hat{π}\pm zs\_{\hat{π}}$$ |

**Exercício 1. Intervalos de confiança para uma variável contínua**

Neste Laboratório vamos utilizar o banco de dados construído a partir de um questionário sobre o tempo gasto para chegar até a USP.

Selecione 5 repostas de forma aleatória. Registre, agora, na tabela abaixo o ID obtido em cada sorteio e verifique, para cada um deles, o tempo de deslocamento até a universidade. Assumimos nesta atividade que os dados coletados nesta turma compõem uma amostra aleatória dos alunos da USP.

|  |  |  |
| --- | --- | --- |
| Caso selecionado  | ID | Valor |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |

**Exercício 2**. Use os valores que foram obtidos a partir de sua amostra para calcular a média para esta amostra de n = 5.

(Dica: Utilize as fórmulas fornecidas na Tabela 1.).

**Exercício 3**. Por favor, use os resultados apresentados na Tabela 2 para calcular o desvio padrão para a sua amostra.

**Exercício 4.** Use a informação que você calculou acima para calcular o intervalo de confiança de 68% para o valor médio da avaliação usando a distribuição normal (mesmo a amostra sendo pequena, por favor use a distribuição normal).

**Exercício 5.** Agora vamos combinar nossos resultados para criar um gráfico das médias amostrais e intervalos de confiança de 68% obtidos por cada grupo. Por favor, registre a linha de intervalo de confiança e média amostral relatado por cada grupo na figura abaixo.

Média Populacional

**Exercício 6**. Qual % dos intervalos contêm a média da população (valor real)? É isso que você esperava? Por quê?

**Exercício 7**. Vamos agora fazer os mesmos cálculos utilizando a distribuição *t*. Calcule o intervalo de confiança de 68%, utilizando uma distribuição *t* com 4 graus de liberdade para a sua amostra.

**Exercício 8**. Agora vamos combinar nossos resultados para criar um gráfico das médias amostrais e intervalos de confiança de 68% obtidos para cada grupo, utilizando a distribuição *t*. Registre a linha de intervalo de confiança e média amostral relatado por cada grupo na figura abaixo.

Média Populacional

**Exercício 9**. Qual % dos intervalos contêm a média da população (valor real)? É isso que você esperava? Por quê?**”**

**\*\*\*\*\*\*\*\***

**Parte II – Inferência e significância**

O debate a respeito da interpretação dos resultados de uma inferência tem sido bastante intenso atualmente. A comparação entre os valores estimados para um parâmetro populacional qualquer e as conclusões a que os pesquisadores chegam tem sido o alvo central da discussão. Afinal, o que aprendemos ao fazer uma inferência e comparar com um valor teste de uma hipótese nula? Este laboratório tem o objetivo de proporcionar uma discussão em torno deste tema.

O primeiro exercício é considerarmos uma situação hipotética de comparação entre médias. Assuma que duas versões de um medicamento foram testadas em um experimento. Os resultados obtidos mostram o efeito médio com o respectivo desvio padrão e tamanho da amostra:

|  |  |  |
| --- | --- | --- |
|  | Grupo 1 | Grupo 2 |
| Efeito médio | 25 | 10 |
| Desvio padrão | 100 | 100 |
| N | 100 | 100 |

1. Assumindo que $H\_{0}: μ=0$ e que $H\_{1}: μ\ne 0$, qual a conclusão a que se chega a respeito dos valores médios obtidos para cada um dos dois grupos?

Mas isto está correto?

Vamos fazer uma comparação de médias. Neste caso, criamos uma nova variável, que chamaremos de $d\_{i-j}$, que será a diferença de médias entre o grupo *i* e o grupo *j*. Essa variável, então, é igual a $d\_{i-j}=d\_{i}-d\_{j}$. Seu erro padrão é:

$$EP\left(\overbar{Y\_{1}}-\overbar{Y\_{2}}\right)=\sqrt{\frac{\left(n\_{1}-1\right)s\_{1}^{2}+\left(n\_{2}-1\right)s\_{2}^{2}}{n\_{1}+n\_{2}-2}}\*\sqrt{\left(\frac{1}{n\_{1}}+\frac{1}{n\_{2}}\right)}$$

Em que $Y\_{i}$ é a média obtida para o grupo *i*, $n\_{i}$ é o tamanho do grupo *i* e $s\_{i}^{2}$é a variância do grupo *i*.

O teste t aplicado seria, então:

$$t=\frac{\overbar{Y\_{1}}-\overbar{Y\_{2}}}{EP\left(\overbar{Y\_{1}}-\overbar{Y\_{2}}\right)}$$

Pode-se, assim, realizar o seguinte teste de hipóteses: $H\_{0}: d\_{i-j}=0$ contra $H\_{1}: d\_{i-j}\ne 0$

1. O que se testa no caso acima?
2. Construa este teste. Qual a sua conclusão? Discuta.

Considere agora que um terceiro grupo pesquisado, cujas informações são: média = 2,5; desvio padrão = 50 e amostra de tamanho 2.500.

1. Este resultado é estatisticamente significante? Discuta.
2. Este resultado é diferente de algum dos outros dois grupos? Discuta.

Para efeitos do segundo problema enfrentado pelo processo tradicional de inferência, tome a comparação entre os resultados dos 1º e 3º grupos acima. **Considere ainda que o *N* do Grupo 1 agora seja igual a 2500 também.**

1. De acordo com os resultados obtidos em termos de significância estatística, qual é a conclusão a que devemos chegar ao tomarmos estes dois casos?
2. Já do ponto de vista substantivo, qual medicamento deveria ser escolhido como preferido?
1. O exercício 1 é baseado no capítulo 8 de Gelman e Nolan (2002). Para uma discussão, veja as páginas 126-130. [↑](#footnote-ref-1)