**Universidade de São Paulo**

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

**Departamento de Ciência Política**

FLS 5028 – Métodos Quantitativos e Técnicas de Pesquisa em Ciência Política

FLP 0406 – Métodos e Técnicas de Pesquisa em Ciência Política

1º Semestre de 2024

Prof. Dr. Glauco Peres da Silva

Laboratório 7 - Inferência

No laboratório de hoje, tentaremos entender os conceitos da teoria da probabilidade discutidos. A ideia é a de integrar o uso de probabilidades com os conteúdos anteriormente tratados no curso.

A tabela abaixo é um resumo que ajudará a entender os exemplos discutidos neste laboratório.

| **Tabela 1. A média e a variância da População e da Amostra** |
| --- |
|  | **População** | **Amostra** |
| Média | $μ=E(Y)$$\overline{Y}=\frac{1}{n}\sum\_{i=1}^{n}Y\_{i}$ | $\overline{y}=\frac{1}{n}\sum\_{i=1}^{n}y\_{i} $  |
| Variância  | Var (Y)$=σ^{2}$Var (Y) = E(Y2)-[E(Y)]2Var ($\overline{Y}$) $=\frac{σ^{2}}{n}$ | $Var\_{y}=\frac{\sum\_{i=1}^{n}(y\_{i}-\overline{y})^{2}}{n-1}$$s\_{y}=\sqrt{\frac{\sum\_{i=1}^{n}(y\_{i}-\overline{y})^{2}}{n-1}}$$s\_{\overline{y }}=\frac{s\_{y}}{\sqrt{n}}$ |
| Número de observações | N | N |

1. Por que há uma fórmula para o cálculo da média e da variância para a população e outra para a amostra?
2. Quais são as diferenças entre elas?

**Simulações com dado**

O valor esperado de uma variável aleatória é uma média ponderada de todos os valores possíveis que esta variável aleatória pode assumir.

1. Assumindo um lançamento de dado convencional de 6 faces, construa uma variável aleatória que compute os resultados observados de cada lançamento. Qual o espaço amostral dessa variável? Qual o tipo de variável criada? E qual a probabilidade de cada evento?
2. Como você fez para determinar a probabilidade de cada um dos eventos no item anterior? Qual o nome damos a este tipo de abordagem à probabilidade das coisas?

No caso de uma variável aleatória discreta, os pesos utilizados no cálculo desta média correspondem às probabilidades de cada um dos possíveis eventos ou resultados.

Primeiramente, vamos focar no caso de um único dado de seis lados. Em geral, a probabilidade de rolar qualquer um dos valores dos lados é exatamente 1/L para uma única jogada de um dado de L lados. Este é um exemplo de uma distribuição uniforme discreta.

No caso de uma única jogada de um dado de 6 lados, existem seis diferentes resultados que podem advir da jogada. Y representa o resultado de uma jogada de um dado de seis lados. Os valores possíveis de Y são 1, 2, 3, 4, 5, 6 (que geralmente são chamados de $y\_{i}$), todos igualmente suscetíveis (tendo cada um a probabilidade de 1/6). O valor esperado de Y é:

$E\left(Y\right)=1\* \frac{1}{6}+ 2\*\frac{1}{6}+ 3\*\frac{1}{6}+ 4\*\frac{1}{6}+5\*\frac{1}{6}+ 6\*\frac{1}{6}= \frac{21}{6} =3,5$

Do mesmo modo, podemos calcular a variância de Y:

$=\frac{1}{6}\left[\left(1-3,5\right)^{2}+\left(2-3,5\right)^{2}+\left(3-3,5\right)^{2}+\left(4-3,5\right)^{2}+\left(5-3,5\right)^{2}+\left(6-3,5\right)^{2}\right]$

$=\frac{1}{6}\left[\left(-2,5\right)^{2}+\left(-1,5\right)^{2}+\left(-0,5\right)^{2}+\left(0,5\right)^{2}+\left(1,5\right)^{2}+\left(2,5\right)^{2}\right]=\frac{17,5}{6}≈2,9$

O desvio padrão do valor esperado de Y, então, seria:

$s\_{y}=\sqrt{\frac{17,5}{6}}=1,7$

1. Vamossimular os resultados deste lançamento. Você pode usar o software que quiser. O Excel tem uma função “Aleatório Entre” que podemos utilizar para simular jogos de dados. No *R*, podemos utilizar o comando *sample[[1]](#footnote-0).* Utilize comandos como estes para simular a obtenção de amostras aleatórias de uma população.

Podemos pensar em cada jogada como um meio de obtenção de uma amostra aleatória. Repita este procedimento de amostragem três vezes. Qual é o valor obtido em cada uma das três jogadas?

**Tabela 1. Diferentes resultados para uma jogada de dado**

| **Jogada** | **Valor obtido** |
| --- | --- |
| 1 |  |
| 2 |  |
| 3 |  |

1. Agora, assumindo que jogamos o dado três vezes, vamos utilizar os valores obtidos para calcular a média e o desvio padrão.

| Agora repita o exercício, mas com 600 jogadas de um único dado de 6 lados. Complete a tabela abaixo com as frequências obtidas. **Tabela 2. Uma amostra de 600 jogadas de um dado** |
| --- |
| **Resultado** | **Frequência** | **Frequência relativa** |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |
| 6 |  |  |

1. Compare e comente os resultados da tabela acima com a frequência esperada. Também use os resultados apresentados na Tabela 2 para calcular a média e a variância da amostra.
2. O que você pode falar sobre a média amostral e o valor esperado da média? Justifique.
3. Compare os resultados das simulações com 6, 60, 600 e 6000 jogadas de dado. Usando a informação, calcule a média da amostra e o desvio padrão da amostra. O que acontece com a média da amostra e com o desvio padrão à medida que o número de jogadas aumenta?

| **Número de jogadas** | **Média amostral** | **Desvio padrão da amostra** |
| --- | --- | --- |
| 6 |  |  |
| 60 |  |  |
| 600 |  |  |
| 6000 |  |  |

1. Podemos comparar os resultados obtidos com a simulação com os valores esperados para a média e para o desvio padrão. Como faria essa comparação?
1. Você pode criar uma variável **x1** utilizando o comando sample da seguinte forma: **x1 <- sample(min : max, n, replace=T)** em que *min* é o valor mínimo do intervalo; *max* é o valor máximo do intervalo e *n* é o número de observações geradas e o *replace* indica se a amostra é feita com ou sem reposição. No caso, o *T* significa *true*, indicando que há reposição na amostragem. Se você digitar o comando **x1,** você verá os valores selecionados. [↑](#footnote-ref-0)