**Universidade de São Paulo**

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

**Departamento de Ciência Política**

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

**FLS5028 - Métodos Quantitativos e Técnicas em Ciência Política**

1º semestre / 2019

Prof. Glauco Peres da Silva

**Laboratório 6 - Probabilidades**

No laboratório de hoje, tentaremos entender os conceitos da teoria da probabilidade.

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 |  |  |
| Variância | Var (Y)  Var (Y) = E(Y2)-[E(Y)]2  Var () |  |
| Número de observações | N | N |

**Exercício 1. 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.

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. Mas como calcular essa probabilidade?

Primeiramente, vamos focar no caso de um único dado de seis lados.

**Questão 1.** Qual a probabilidade de se tirar um determinado número, digamos 3, em um único lançamento de um dado? Como você chegou a esta resposta?

É 1/6. Pois podemos calcular a probabilidade de um evento pela razão entre o número de eventos favoráveis e o número de eventos possíveis.

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 do lançamento do dado. 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 ), todos igualmente suscetíveis (tendo cada um a probabilidade de 1/6).

Note que esta probabilidade é teórica. Ou seja, não foi preciso lançar o dado para que fosse atribuída a chance de que determinado resultado ocorra. Aqui temos, assim, que , em que *L* é o número de lados do dado lançado. Essa probabilidade é calculada pela frequência de um resultado diante do espaço amostral. Outras formas de calcular a probabilidade determinam os resultados a partir de perspectivas diferentes sobre a ocorrência dos fenômenos.

A propósito, na aula passada, trabalhamos com um histograma de uma variável aleatória que media o número efetivo de partidos em nosso banco de dados.

**Questão 2.** Qual a relação que podemos encontrar entre a abordagem frequentista de probabilidade e um histograma? Discuta.

O histograma representa graficamente a frequência de valores de uma determinada variável. Quando abordamos a probabilidade pela frequência de ocorrências, podemos utilizar o Teorema do Limite Central, em que diz que a distribuição da média de uma amostra aleatória, se for grande o suficiente, é aproximadamente igual a uma distribuição normal. Assim, se temos um histograma descrevendo a frequência das médias de uma amostra suficientemente grande, podemos verificar que segue um formato de sino.

Retomando a avaliação de uma variável aleatória, para qualquer abordagem de probabilidade utilizada, o valor esperado de Y para o lançamento de um dado de 6 lados pode ser representado por:

Note que o valor esperado de Y, *E(Y)*, é a média ponderada de cada valor possível de Y por sua probabilidade de ocorrência. Ou por sua frequência de ocorrência. Se todas as ocorrências possíveis possuem a mesma probabilidade de ocorrência, temos que o valor esperado é igual a média aritmética.

Do mesmo modo, podemos calcular a variância de Y:



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

**Questão 2.** O Excel tem uma função “Aleatório Entre” que iremos utilizar para simular jogos de dados. Utilizaremos este comando para simular a obtenção de amostras aleatórias de uma população.

Por favor, use a função "Aleatório Entre" para simular uma única jogada de um dado de seis lados. 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?

|  |  |
| --- | --- |
| **Jogada** | **Valor obtido** |
| 1 | 6 |
| 2 | 4 |
| 3 | 6 |

**Questão 3.** Agora, assumindo que jogamos o dado três vezes, vamos utilizar os valores obtidos para calcular a média e o desvio padrão para esta distribuição.

Média: x = (6 + 4 + 6) / 3 = 5.3

Variância: s² = (6 – 5.3)² + (4 – 5.3)² + (6 – 5.3)² / (3-1) = 1.3

Desvio padrão: s = 1.15

|  |  |  |
| --- | --- | --- |
| Agora repita o exercício, mas com 600 jogadas de um único dado de 6 lados. Complete a tabela abaixo com as frequências obtidas. Observe que, em alguns casos, a frequência relativa poderá ser maior do que 1/6 e, em outros, poderá ser menor.  **Tabela 2. Uma amostra de 600 jogadas de um dado** | | |
| **Resultado** | **Frequência** | **Frequência relativa** |
| 1 | 107 | 107/600 = 0.1783 |
| 2 | 88 | 88/600 = 0.147 |
| 3 | 100 | 100/600 = 0.167 |
| 4 | 107 | 107/600 =0.1783 |
| 5 | 103 | 103/600 = 0.1717 |
| 6 | 95 | 95/600 = 0.1583 |

**Questão 4.** Por favor, use os resultados apresentados na Tabela 2 para calcular a média e a variância da amostra. Não se esqueça de demonstrar cálculos e raciocínios.

Valor esperado é a média ponderada. Pra cada um dos resultados possíveis vezes a frequência.

**Questão 5.** O que você pode falar sobre a média amostral e o valor esperado da média? Justifique.

A média amostral e o valor esperado estão bem próximos com a amostra de 600 jogadas.

**Questão 6.** Com base na Lei dos Grandes Números, conforme o número de jogadas aumente (n) o que irá acontecer?

Caso o número de jogadas aumente a média amostral tende a ficar mais próxima do valor esperado.

**Questão 7.** Teste sua hipótese simulando 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 | 3.5 | 1.64 |
| 60 | 3.67 | 1.51 |
| 600 | 3.56 | 1.63 |
| 6000 | 3.5 | 1.72 |

Note que a partir deste momento, podemos comparar o resultado obtido através do lançamento dos dados com a frequência teórica.

**Questão 8.** Determine com base em 6000 lançamentos, qual seria probabilidade de você lançar uma vez o dado e tirar 1. E qual seria a probabilidade de tirar 5? São diferentes da resposta teórica? Por que?

**P(1) =** 1009/6000 = 0.168

**P(2) =** 1032/6000 = 0.172

São diferentes da resposta teórica, pois calculamos a probabilidade a partir das ocorrências, no qual não obtivemos a mesma frequência de ocorrências para todos os resultados possíveis.

**Questão 9.** O que aconteceria com as estimativas da média e da variância se jogássemos o dado 600 vezes e, em seguida, repetíssemos o mesmo procedimento?

Os valores, provavelmente, iriam mudar, pois como a amostra é aleatória, diversos valores podem ser incluídos, porém é esperado que ele esteja próximo da resposta teórica.

**Questão 10.** Com base nos resultados obtidos na Questão 3, calcule o erro padrão da média de uma distribuição amostral.