Um Exemplo Prático com Quarto: A Distribuição de Poisson

Prof. Dr. Sadraque E. F. Lucena

Este é um arquivo usado como exemplo na disciplina ESTAT0109 - Mineração de Dados em Estatística, com o intuito de explorar a aplicação do Quarto, uma ferramenta de publicação técnica de código aberto projetada para auxiliar os cientistas de dados na compartilhamento de suas análises. Como exemplo, realizaremos uma breve exploração da distribuição de Poisson.

## A distribuição de Poisson

A distribuição de Poisson é uma distribuição de probabilidade discreta que descreve o número de eventos que ocorrem em um intervalo fixo de tempo ou espaço. A distribuição de Poisson é frequentemente utilizada em problemas do mundo real, como prever o número de chamadas recebidas em um *call center* em uma hora, o número de defeitos em um lote de produtos, ou a contagem de eventos raros em geral.

## A Fórmula Matemática

De acordo com MAGALHÃES (2006), seja uma variável aleatória com distribuição de Poisson (MAGALHÃES, 2006) com parâmetro , denotada como . A função de probabilidade de é então dada por (MORETTIN; BUSSAB, 2017)

onde:

* é a função exponencial;
* é o número de ocorrências de um evento ().
* (lambda) é um número real positivo que representa a média de ocorrências no intervalo especificado.

## Exemplo Prático de Aplicação

Vamos imaginar algumas situações onde a Poisson se aplica. A [Tabela 1](#tbl-letters), criada com a sintaxe simples do Markdown, nos dá alguns exemplos:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tabela 1: Exemplos de aplicação da distribuição de Poisson.   | Fenômeno Observado | Intervalo | Valor de Lambda () | | --- | --- | --- | | Carros passando por um cruzamento | 1 minuto | 10 | | Erros de digitação em uma página de um livro | Por página | 2 | | Gols marcados em uma partida de futebol | Por partida | 2.5 | |

## Simulando e Visualizando Dados da Poisson em R

Agora, a mágica do Quarto: vamos usar um bloco (*chunk*) de código R para simular dados de uma distribuição de Poisson e visualizá-los. Vamos simular 1000 observações de uma Poisson com . Usaremos o pacote ggplot2 para criar um histograma elegante.

# 1. Carregar o pacote ggplot2 para visualização  
library(ggplot2)  
  
# 2. Definir o nosso parâmetro lambda  
lambda\_valor <- 5  
  
# 3. Gerar 1000 números aleatórios de uma distribuição Poisson  
dados\_poisson <- rpois(n = 1000, lambda = lambda\_valor)  
  
# 4. Criar o histograma  
ggplot(data = data.frame(dados = dados\_poisson), aes(x = dados)) +  
 geom\_histogram(binwidth = 1, fill = "skyblue", color = "black", alpha = 0.8) +  
 labs(  
 title = "Distribuição de 1000 Observações de uma Poisson",  
 subtitle = paste("Lambda (taxa média) =", lambda\_valor),  
 x = "Número de Eventos (x)",  
 y = "Frequência"  
 ) +  
 theme\_minimal()

|  |
| --- |
| Figura 1: Histograma de 1000 valores simulados de uma distribuição de Poisson com lambda = 5. |

Como podemos ver na [Figura 1](#fig-gerar-e-plotar-poisson), a maior parte dos nossos dados se concentra em torno do valor de lambda, que é 5.

## Tabela de Estatísticas Descritivas

Para complementar nossa análise, podemos criar uma tabela de resumo profissional dos dados simulados. Usaremos o pacote gtsummary, que é fantástico para criar tabelas prontas para publicação com um código muito simples.

library(knitr)  
resumo\_df <- data.frame(  
 "Estatística" = c("Tamanho da Amostra (n)",  
 "Valor Mínimo",  
 "Valor Máximo",  
 "Média",  
 "Variância"),  
 Valor = c(  
 length(dados\_poisson),  
 min(dados\_poisson),  
 max(dados\_poisson),  
 mean(dados\_poisson),  
 var(dados\_poisson)  
 )  
)  
kable(resumo\_df,  
 caption = "Estatísticas Descritivas dos Dados Simulados",  
 digits = 2)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tabela 2: Estatísticas Descritivas dos Dados Simulados   | Estatística | Valor | | --- | --- | | Tamanho da Amostra (n) | 1000.00 | | Valor Mínimo | 0.00 | | Valor Máximo | 13.00 | | Média | 4.88 | | Variância | 5.13 | |

A [Tabela 2](#tbl-resumo) acima confirma o que esperávamos: a média e a variância dos nossos dados simulados são muito próximas de .

# Referências

MAGALHÃES, M. N. **Probabilidade e variáveis aleatórias**. *[S. l.]*: Edusp, 2006.

MORETTIN, P. A.; BUSSAB, W. O. **Estatı́stica básica**. 9. ed. São Paulo: Saraiva, 2017.