

MAE 399 - Análise de Dados e Simulação  
Prof. Fábio Machado  
Lista 4 - 09/05/2022

1. Recentemente introduzimos o tema *Simulação* e a técnica conhecida como *método da transformada inversa*. Aprendemos, por exemplo, que o comando do R, `rexp(n, λ)` gera  $n$  cópias independentes de uma variável aleatória  $X$ , com distribuição exponencial de taxa  $\lambda$  (no jargão dos Probabilistas:  $X \sim \mathcal{E}(\lambda)$ , lembrando que  $\mathbb{E}(X) = \frac{1}{\lambda}$ ) a partir de  $n$  uniformes independentes  $(U_1, U_2, \dots, U_n)$  e da transformação  $X_i = -\frac{1}{\lambda} \ln(U_i)$ . Faça uma ilustração para isto:
  - a) Gere 10.000 valores para exponencial de parâmetro  $\lambda$  (de sua escolha) com o comando `rexp(n, λ)`;
  - b) Gere 10.000 valores independentes para exponencial de parâmetro  $\lambda$  (mesmo do item acima) através do *método da transformada inversa*;
  - c) Faça e coloque lado a lado, os histogramas dos conjuntos de dados gerados nos itens anteriores.
2. No gabarito apresentado pelo monitor, Gustavo, a variável distância entre estação de retirada e estação de devolução, foi considerada (pag 4 e 5). Lembre que na análise do Gustavo a variável tempo de uso da bicicleta foi aparada (ou "trimmed") da seguinte maneira: as linhas com valores negativos ou valores superiores a 60 minutos na variável tempo de uso da bicicleta, foram retiradas. Vamos aprender a gerar valores com distribuição análoga aquela encontrada no gabarito do Gustavo (importante: considere o Citi Bike de 02/2021).
  - a) Considerando que a distância entre as estações é sempre superior a 350m e que a velocidade média de uma bicicleta, em "modo passeio", é superior a 16 Km/h, retire também as linhas cujos valores da variável tempo de uso são inferiores ao tempo de percorrer 350 m a uma velocidade de 16 km/h.
  - b) Retire também as linhas cujo valor para a variável tempo de uso é superior a 60 min.
  - c) Refaça o histograma apresentado no gabarito feito pelo Gustavo.
  - d) Com os dados resultantes estime  $p$ , a proporção de viagens cujas locais de retirada e devolução das bicicletas são os mesmos.
  - e) Com os dados resultantes, estime  $\bar{X}_0$  a média das distâncias percorridas, sem considerar os valores 0's.
  - f) Simule  $X_1, X_2, \dots, X_n$ , com  $n = 10,000$ , valores para a variável distância entre estação de retirada e estação de devolução, através da seguinte ideia:
    - Gere  $U_1, U_2, \dots, U_n$ , uniformes em  $[0,1]$ ;
    - Se  $U_i \leq p$ , então  $X_i = 0$ ;
    - Caso contrário,  $X_i \sim \mathcal{E}(\lambda)$ , onde  $\lambda = \frac{1}{\bar{X}_0}$ ;
    - Faça e coloque lado a lado, o histograma original (item c) e o histograma de dados gerados por esta ideia;
    - Responda: você acha que a Exponencial um bom modelo para simular os dados originais (item c)?