# ISCTE-IUL

**UC: Modelação Estocástica**  
 **Docente: Catarina Marques**  
 **Ano Letivo:** 2023/2024  
  
**Problem Set 1**

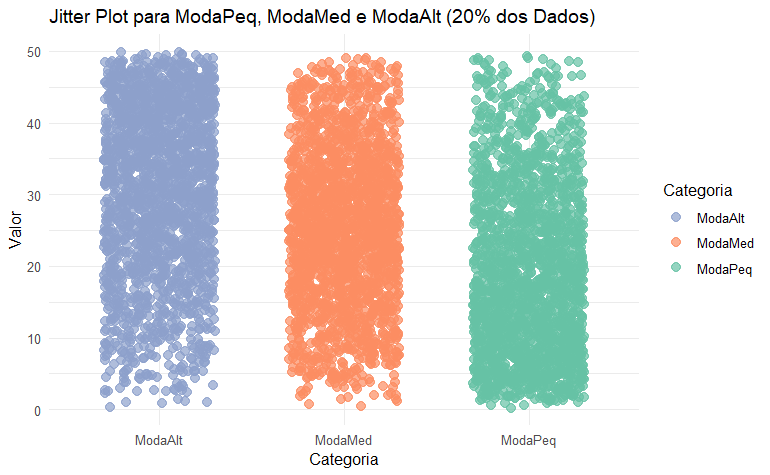


**08 de Dezembro de 2023**

Trabalho realizado por:

* Allan Kardec, nº 103380, CDC1
* Diogo Freitas, nº 104841, CDC1

# Exercício 1)

A chart with different colored squares

Description automatically generatedA diagram of a triangular shape

Description automatically generated with medium confidenceA graph with a red line with Ryugyong Hotel in the background

Description automatically generatedPara a realização deste exercício, foi criada uma função distribuição de densidade triangular de raiz, como também um algoritmo de [aceitação rejeição](#_Pseudocódigo_do_Algoritmo); estes encontram-se explicados e comentados no ficheiro rmd. Para testar este algoritmo, foram realizadas 3 distribuições diferentes (num intervalo de 0 a 50), com uma amostra de 1000 números, diferenciando apenas a moda em cada uma destas (5, 25 e 45). Foram criados 3 histogramas diferentes para uma visualização mais fácil. À direita, podemos visualizar o histograma com a moda centrada e, se analisado com atenção, é percetível o formato de um triângulo formado pelas diferentes barras. Esta distribuição triangular possui o valor da moda como 25 e, analisando com atenção, é possível reparar que o “pico” do triangulo (ponto mais alto), se encontra no valor 25 (correspondente à 8ª barra). Logo abaixo, podemos então visualizar todos os candidatos criados pelo método de aceitação-rejeição, estando a vermelho os rejeitados e a verde os aceites. Em anexo, encontram-se os restantes [histogramas](#_Histograma_com_a), sendo possível visualizar os diferentes triângulos formados, como também os pontos aceites e rejeitados. Por último, podemos visualizar 3 *boxplots* diferentes, cada um representando uma das 3 diferentes distribuições. Como já era esperado, se o valor da moda for muito parecido/próximo com um dos valores dos extremos, maior será a concentração dos pontos dessa distribuição no extremo respetivo. É importante ter em conta que, todas as distribuições têm pontos próximos do máximo e do mínimo, existindo alguma dispersividade devido ao elevado número da amostra. Por último, podemos visualizar um *jitterplot* que dá a mesma informação que os *boxplots*, mas fornece uma melhor visualização da dispersão dos pontos.

# Exercício 2)

A graph with a bar chart and a bar chart

Description automatically generated with medium confidenceA graph with colored squares and lines

Description automatically generated with medium confidenceA graph with different colored squares

Description automatically generatedNeste exercício serão comparados 2 estimadores diferentes, cujo nomes se encontram em [anexo](#_Estimadores_–_Skewness). Para realizar esta comparação, serão usadas 3 amostras diferentes, cada uma com 100 subamostras, mas com 3 tamanhos diferenciados: 20, 100, 1000. Tendo em conta que se trata de uma *t.student*, foi pensado num valor para os graus de liberdade situado entre 3 e 30, pois a *skewness* não tem um valor definido para distribuições que têm menos de três graus de liberdade e se o grau de liberdade for superior a 30, existe uma grande aproximação de uma normal. Tendo isto em conta, foi escolhido 5 como valor para os graus de liberdade. Os 2 primeiros *Boxplots* aqui apresentados demonstram que, em ambos os estimadores, quanto maior o número de subamostras criados, mais concentrados são os valores médios de zero. É importante relembrar que a *skewness* da *t.student* é, teoricamente, zero devido à sua simetria[[1]](#footnote-2), logo, o estimador que apresentar os valores mais concentrados/próximos de zero, é melhor. De seguida, podemos visualizar uma comparação das 3 subamostras de cada um dos estimadores e, analisando com atenção, o segundo estimador (s2), apresenta valores mais próximos de zero (em todas as amostras) dando assim a entender que o segundo estimador é melhor que o primeiro. Para avaliar os estimadores, podemos visualizar a [tabela](#_Tabela_–_Avaliação) situada nos anexos que demonstra que o erro padrão e o MSE do segundo estimador são melhores, ou seja, este possui uma menor variabilidade, como também apresenta valores mais próximos da *skewness* real, que é zero. Em [anexo](#_Escolha_dos_graus) podemos visualizar o quanto os graus de liberdade afetam esta conclusão.

# Anexos

## Pseudocódigo do Algoritmo Aceitação-rejeição

**Função *AceitaçãoRejeição*(n, g(x), f(x), C):**

amostras = []

enquanto tamanho de amostras < n:

gerar candidato x' de acordo com g(x)

gerar número aleatório u entre 0 e 1

calcular R = f(x') / (C \* g(x'))

se u <= R:

adicionar x' a amostras

retornar amostras

**Legenda**

n := número de amostras que se deseja criar

g(x) := função de densidade de probabilidade da distribuição dos candidatos

f(x) := função de densidade de probabilidade da distribuição alvo

C := constante tal que

## Histograma com a moda baixa

A diagram of a distribution overlay with red and green dots

Description automatically generatedA graph of a graph

Description automatically generated

## A graph with a red line Description automatically generatedA green and red dotted chart Description automatically generatedHistograma com a moda alta

## Estimadores – Skewness

Quantile-based measures

Groeneveld and Meeden's coefficient



## Tabela – Avaliação dos Estimadores

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | S1\_20 | S1\_100 | S1\_1000 | S2\_20 | S2\_100 | S2\_1000 |
| Erro Padrão | 0.063 | 0.014 | 0.0014 | 0.0484 | 0.01 | 0.001 |
| Erro quadrático médio (MSE) | 0.079 | 0.0195 | 0.002 | 0.0466 | 0.0092 | 0.0009 |

## A graph with a red line Description automatically generatedA graph with lines and numbers Description automatically generatedEscolha dos graus de liberdade

Um valor de grau de liberdade inferior a 3 faz com que o primeiro estimador apresente resultados melhores que o segundo e, a partir de 3 para a frente, o segundo passa a ser melhor que o primeiro. É importante relembrar que um valor de grau de liberdade inferior a 3 não faz sentido.

1. Quanto maior o número de graus de liberdade, mais a distribuição se aproxima de uma distribuição normal padrão, que é perfeitamente simétrica [↑](#footnote-ref-2)