

## UNIVERSIDADE DE COIMBRA FACULDADE DE CIÊNCIAS E TECNOLOGIA

## Departamento de Engenharia Informática

Pólo II - Pinhal de Marrocos, 3030 Coimbra - Portugal

## Exame de Metodologias Experimentais em Informática

20 de julho d	e 2020						Tempo má	x1mo: 2h00m
Nome:							Nº:	
D 1 ~	4) 200	<b>a</b> ) 150	<b>31</b> ) 150	<b>3</b> ) <b>3</b> 0%	4) 100	<b>E</b> ) 200		

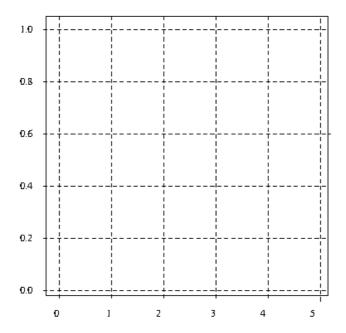
Ponderações: 1) 20% 2a) 15% 2b) 15% 3) 20% 4) 10% 5) 20%

- 1) Um banco desenvolveu um novo sistema de base de dados com o objetivo de melhorar o desempenho das transações bancárias dos seus clientes através da Internet. O tempo de resposta para cada tipo de transação depende de vários fatores que estão essencialmente relacionados com o tipo de *queries* efetuadas pelo clientes e com a carga no sistema. De modo a avaliar a resposta do novo sistema, o Banco pediu ao Departamento de Sistemas de Informação que efetuasse uma experiência para testar o desempenho deste e comparar com o desempenho do sistema atual. Para a experiência foram selecionadas várias contas bancárias representativas e com perfis de transações típicas para assegurar que as medições do tempo de resposta em ambos os sistemas fossem comparáveis. Descreva como organizaria esta experiência. A sua resposta deve ser a mais completa possível, com ênfase na definição dos passos do desenho experimental. Indique as variáveis dependentes e independentes, os níveis para cada variável independente, as hipóteses mais relevantes e as técnicas de teste de hipóteses a usar.
- 2) Assuma que desenvolveu um novo algoritmo de ordenação denominado QueSort e que decidiu efetuar uma experiência para comparar o seu desempenho com o do Quicksort em termos de tempo de execução. Para efetuar a experiência, executou ambos os algoritmos de ordenação em cinco sequências de valores gerados aleatoriamente e guardou os tempos de execução obtidos por cada algoritmo. A seguinte tabela apresenta os resultados obtidos em segundos, ordenados por ordem não-decrescente.

QueSort	1	2	2	2	4
QuickSort	2	3	3	3	5

a) Qual é o intervalo de confiança a 90% para a diferença entre as médias do tempo de execução de ambos algoritmos? Explique o significado do intervalo obtido.

b) Aplique o teste de Kolmogorov-Smirnov para duas amostras para testar se ambas as amostras seguem a mesma distribuição para 10% de significância. Desenha as funções de distribuição empírica na seguinte figura.



3) Complete a seguinte tabela ANOVA para 2 fatores (Compilador e Computador). Na coluna "Decisão" deve apresentar a sua decisão relativamente à hipótese nula (Rejeitar H<sub>0</sub> / Não rejeitar H<sub>0</sub>). Justique a sua decisão e apresente todos os cálculos necessários. Assuma um nível de significância de 5%, que o tamanho da amostra para cada combinação de níveis de ambos os factores é 3 e que os pressupostos do teste estatístico são válidos.

	SS	df	MS	F value	Decisão
Compilador	872.44	2			
Computador	2005.6	1			
Interaction	53.78				
Within	240				
Total	3171.82				

4) Em que condições poderá ser preferível efetuar um teste paramétrico em vez de um teste nãoparamétrico? Justique a sua resposta. 5) Pretende avaliar o desempenho do novo sistema de base de dados que está a propor no seu trabalho de mestrado relativamente ao número de queries. Testou o seu sistema com várias queries e recolheu os seguintes tempos de resposta na seguinte tabela

Nº. queries	Tempo
10	10
20	22
30	43
40	89
50	165

Ao aplicar uma regressão linear do tipo y = ax + b a estes dados, obteve o seguinte modelo linear: a = 3.7 and b=-47.3. Calcule o coeficiente de determinação ( $\mathbb{R}^2$ ), discuta a qualidade deste modelo linear e discuta o resultado no contexto do problema apresentado. Se aplicável, discuta possíveis transformações.