

	<b>Instituto Federal de Goiás</b>
	<b>Curso:</b> Ciência da Computação
	<b>Disciplina:</b> Probabilidade e Estatística
	<b>Semestre:</b> 2024/1
	<b>Professor:</b> Éder Brito

## Trabalho Final - Testes de Hipóteses

### Instruções:

- O objetivo dessa atividade que vocês revisem o conteúdo de Testes de Hipóteses e, simultaneamente, organizem e entendam os códigos necessários para realizá-los de acordo com o software que preferirem (R, python ou outros).
- A atividade pode ser feita em grupos de até 4 pessoas.
  - Lembrem-se que a ideia é conhecer/entender os códigos e compartilhá-los entre vocês. Eu sugiro que façam em grupo, compartilhando o que cada um fizer e discutindo os resultados. Além de ser um trabalho que vale nota, é também uma forma de estudar e praticar o conteúdo.
- O que eu quero que vocês entreguem?
  - Um arquivo pdf com os resultados dos testes **organizados**.
- O que precisa ser feito em cada questão?
  - Em um teste de hipóteses não podem faltar AS HIPÓTESES! (nula e alternativa).
  - O  $p$ -valor obtido e a conclusão a partir dele. Obviamente, quero a conclusão do problema (e não apenas dizer sobre rejeitar ou não a hipótese nula).
  - Algumas questões carecem de mais de um teste ou outras análises. Apresente o que for necessário para analisar o que é pedido na questão.
  - Claro que nos resultados dos testes vocês verão outras informações interessantes. Fiquem a vontade para decidir inseri-las ou não no trabalho.
- O que vou avaliar?
  - Vou avaliar a entrega e verificar se os resultados estão coerentes e objetivamente discutidos em cada questão.
  - A ideia, como disse acima, é que vocês revisem o conteúdo e pesquisem ou adaptem os códigos para realizar os testes.. Por isso optei por questões simples, que certamente darão pouquíssimo trabalho.
  - Por isso espero que sejam bem feitas!

### Questão 1. Teste de Normalidade

A distribuição abaixo apresenta os tempos, em segundos, que 75 computadores aleatoriamente escolhidos por um cientista da computação gastaram para executar determinada tarefa.

89, 93, 96, 98, 98, 101, 102, 102, 104, 105, 106, 108, 109, 109, 110,  
110, 110, 111, 111, 112, 112, 113, 113, 114, 114, 114, 114, 114, 114,  
115, 115, 115, 116, 116, 116, 116, 116, 116, 116, 117, 117, 117, 118, 118,  
118, 119, 119, 119, 119, 119, 119, 119, 120, 120, 120, 121, 121, 121, 122,  
122, 123, 123, 124, 125, 125, 126, 127, 127, 127, 128, 128, 129, 131, 133.

- Discuta (ao nível de 5%) a normalidade desses dados.
- Faça um histograma e um boxplot dos dados. O resultado sobre a normalidade se alteraria caso os outliers fossem desconsiderados?

### Questão 2. Teste para média em amostra única

Cinquenta dados de magnitude foram selecionados aleatoriamente de 10.594 terremotos registrados em um ano, em uma localidade no sul da Califórnia. Estes valores estão exibidos logo abaixo. Os dados são do *Southern California Earthquake Data Center (Centro de Dados de Terremotos da Califórnia do Sul)*. Deseja-se verificar se a população de terremotos tem uma magnitude média maior do que 1,00. Use o nível de significância de 0,05.

0.70, 2.20, 1.64, 1.01, 1.62, 1.28, 0.92, 1.00, 1.49, 1.42,  
0.74, 1.98, 1.32, 1.26, 1.83, 0.83, 1.00, 2.24, 0.84, 1.35,  
0.64, 0.64, 2.95, 0.00, 0.99, 1.34, 0.79, 2.50, 1.42, 0.93,  
0.39, 1.22, 0.90, 0.65, 1.56, 0.54, 0.79, 1.79, 1.00, 0.40,  
0.70, 0.20, 1.76, 1.46, 0.40, 1.25, 1.44, 1.25, 1.25, 1.39.

- Use um teste  $t$  para essa análise.
- Lembre-se que uma premissa desse teste é que os dados sejam normalmente distribuídos.

### Questão 3. Teste para proporção em amostra única

Quando Mendel realizou seus famosos experimentos sobre genética, uma amostra de descendentes consistia em 428 ervilhas verdes e 152 ervilhas amarelas. Teste a um nível de significância de 5%, a afirmativa de Mendel de que, sob as mesmas circunstâncias, 25% das ervilhas descendentes seriam amarelas.

#### Questão 4. Teste para comparar médias de duas amostras pareadas

O gerente de uma empresa de táxi está tentando decidir se o uso de novo tipo de pneus aumenta a economia de combustível de sua frota. Doze carros foram equipados com os novos pneus e conduzidos por um curso de teste pré-estabelecido. Sem mudar de motorista, os mesmos carros foram então equipados com pneus dos mesmos tipos anteriores e conduzidos novamente durante o curso de teste. O consumo de gasolina, em quilômetros por litro, foi registrado como segue:

Carro	Km/L pneus tradicionais	Km/L pneus novos
1	4,2	4,1
2	4,7	4,9
3	6,6	6,2
4	7,0	6,9
5	6,7	6,8
6	4,5	4,4
7	5,7	5,7
8	6,0	5,8
9	7,4	6,9
10	4,9	4,7
11	6,1	6,0
12	5,2	5,9

Assumindo que as populações sejam normalmente distribuídas, pode-se concluir que carros equipados com os novos pneus fornecem maior economia de combustível que os antigos? (Use 5% de significância).

#### Questão 5. Teste para comparar médias de duas amostras independentes

Em um estudo realizado numa certa cidade, os níveis de ácido ascórbico plasmático de mulheres grávidas foram comparados para fumantes versus não-fumantes. Trinta e duas mulheres nos últimos três meses de gravidez, livres de grandes distúrbios de saúde e com idades entre 15 e 32 anos, foram selecionadas para o estudo. Antes da coleta de 20 ml de sangue, as participantes foram orientadas a evitar o café da manhã, abandonar seus suplementos vitamínicos e evitar alimentos ricos em ácido ascórbico.

A partir das amostras de sangue, foram determinados os seguintes valores de ácido ascórbico plasmático, em miligramas por 100 mililitros:

Não Fumantes		Fumantes
0,97	1,16	0,48
0,72	0,86	0,71
1	0,85	0,98
0,81	0,58	0,68
0,62	0,57	1,18
1,32	0,64	1,36
1,24	0,98	0,78
0,99	1,09	1,64
0,9	0,92	
0,74	0,78	
0,88	1,24	

Existem provas suficientes para concluir que há uma diferença entre os níveis de ácido ascórbico no plasma de fumantes e não fumantes?

**Obs.:** Não se esqueça que existe diferença na estatística de teste caso as variâncias das populações das duas amostras sejam iguais ou não. Portanto, não deixe de verificar (testar) a igualdade ou não das variâncias populacionais antes de testar qualquer coisa sobre suas médias.

### Questão 6. Teste para comparar médias de duas amostras independentes

Em um experimento de adubação de pastagens foram comparadas dois tipos de adubação química e orgânica. Seja  $X_q$  a altura das plantas adubadas com adubo químico e  $X_o$  a altura das adubadas com adubo orgânico, de modo que ambas tenham distribuição aproximadamente normal. Observaram-se 12 parcelas de  $X_q$  e 10 parcelas de  $X_o$ . Os resultados são apresentados a seguir:

$X_q$	67,1	56,4	63,5	60	62,5	59,6	70	67,9	69,9	74,1	65,9	58
$X_o$	82,7	71,8	70	71,8	78	69,6	75,8	62,8	72,4	74,9		

Verificar se, ao nível de 5%, existe diferença significativa entre as alturas das plantas em relação ao tipo de adubo utilizado na pastagem. Caso exista, verifique qual adubo é mais indicado para obter maiores alturas dessa planta.

**Obs.:** Não se esqueça que existe diferença na estatística de teste caso as variâncias das populações das duas amostras sejam iguais ou não. Portanto, não deixe de verificar (testar) a igualdade ou não das variâncias populacionais antes de testar qualquer coisa sobre suas médias.

### Questão 7. Teste Qui-quadrado para aderência

O tempo residual do feito de um agrotóxico está sendo analisado. Estudos anteriores com produtos similares, indicam que o modelo Exponencial com média de 3 dias poderia ser adequado. Qual a conclusão, ao nível de 5%, se a análise em laboratório de uma amostra de 300 aplicações do agrotóxico forneceu os seguintes tempos residuais:

Faixa de Tempo	[0,1)	[1,2)	[2,3)	[3,4)	[4,5)	[5,6)	[6,∞)
Frequência	89	60	43	40	25	22	21

- Lembre-se que para esse teste basta comparar o vetor de frequências da tabela com as probabilidades da distribuição em teste para os mesmos intervalos definidos na tabela.

### Questão 8. Teste Qui-quadrado para independência de duas variáveis qualitativas

Estudando-se os índices de natalidade em uma certa cidade, levantou-se a questão se a renda familiar poderia ter alguma relação com o número de filhos da família. Uma amostra de 257 famílias foi observada e os dados coletados estão apresentados na tabela abaixo

Renda - Filhos	0	1	2	> 2	Total
menos de 2000	15	27	50	43	135
2000 a 5000	25	37	12	8	82
5000 ou mais	8	13	9	10	40
Total	48	77	71	61	257

Verifique se existe dependência entre o número de filhos e a renda familiar, a um nível de 5%.

### Questão 9. Teste Qui-quadrado para homogeneidade de populações

A reação ao tratamento por quimioterapia foi estudada em quatro grupos de pacientes com câncer. Retirou-se uma amostra de pacientes de cada grupo e classificou-se a reação em três categorias: pouca, média e alta. Teste, ao nível de 5%, se todos os tipos de câncer reagem da mesma maneira.

Câncer - Reação	Pouca	Média	Alta	Total
Tipo I	51	33	16	100
Tipo II	58	29	13	100
Tipo III	48	42	30	120
Tipo IV	26	38	16	80

Se, por curiosidade, você retirasse um desses grupos (o que tiver comportamento proporcionalmente diferente dos demais) e refizer o teste? Poderia ter um outro resultado?

### Questão 10. ANOVA

No conjunto de dados abaixo estão listadas as quantidades de nicotina (mg por cigarro) encontradas em amostras de três tipos de cigarros: cigarros tamanho *king*, cigarros de 100mm mentolados e nos cigarros de 100mm sem mentol. Os cigarros tamanho *king* são sem filtro, sem mentol e são não light. Os cigarros de 100mm mentolados são com filtro e não light. Os cigarros de 100 mm sem mentol têm filtro e são não light.

Use o nível de significância de 0,05 para testar a afirmativa de que as três categorias de cigarros resultam na mesma quantidade média de nicotina.

Uma vez que apenas os cigarros tamanho king são sem filtro, os filtros parecem fazer alguma diferença?

KING	MENTH	NonMENTH
1.1	1.1	0.4
1.7	0.8	1.0
1.7	1.0	1.2
1.1	0.9	0.8
1.1	0.8	0.8
1.4	0.8	1.0
1.1	0.8	1.1
1.4	0.8	1.1
1.0	0.9	1.1
1.2	0.8	0.8
1.1	0.8	0.8
1.1	1.2	0.8
1.1	0.8	0.8
1.1	0.8	1.0
1.1	1.3	0.2
1.8	0.7	1.1
1.6	1.4	1.0
1.1	0.2	0.8
1.2	0.8	1.0
1.5	1.0	0.9
1.3	0.8	1.1
1.1	0.8	1.1
1.3	1.2	0.6
1.1	0.6	1.3
1.1	0.7	1.1

### Questão 11. ANOVA

Qual sua cor de M&M's preferida? Você já se perguntou se existe diferença nos pesos dos M&M's de acordo com a cor de cada um?

Para verificar se essa diferença existe ou não, uma amostra de M&M's foi selecionada, os doces foram categorizados de acordo com suas cores e seus pesos foram observados. Os resultados estão listados na tabela abaixo.

Red	Orange	Yellow	Brown	Blue	Green
0.751	0.735	0.883	0.696	0.881	0.925
0.841	0.895	0.769	0.876	0.863	0.914
0.856	0.865	0.859	0.855	0.775	0.881
0.799	0.864	0.784	0.806	0.854	0.865
0.966	0.852	0.824	0.840	0.810	0.865
0.859	0.866	0.858	0.868	0.858	1.015
0.857	0.859	0.848	0.859	0.818	0.876
0.942	0.838	0.851	0.982	0.868	0.809
0.873	0.863			0.803	0.865
0.809	0.888			0.932	0.848
0.890	0.925			0.842	0.940
0.878	0.793			0.832	0.833
0.905	0.977			0.807	0.845
	0.850			0.841	0.852
	0.830			0.932	0.778
	0.856			0.833	0.814
	0.842			0.881	0.791
	0.778			0.818	0.810
	0.786			0.864	0.881
	0.853			0.825	
	0.864			0.855	
	0.873			0.942	
	0.880			0.825	
	0.882			0.869	
	0.931			0.912	
				0.887	
				0.886	

Discuta, ao nível de 5%, se existe diferença no peso médio de M&M's de acordo com as cores.

### Questão 12. Regressão múltipla

Os dados abaixo trazem informações diversas e algumas medidas de indivíduos. Esses dados foram utilizados para um estudo de relação entre altura das pessoas e o tamanho dos pés e sapatos, aplicados à ciência forense (*Estimation of Stature from Foot and Shoe Length: Applications in Forensic Science*).

Estabeleça um modelo de regressão múltipla para a variável altura em função de todas as outras covariáveis. Feito isso, verifique quais covariáveis são significativas para o modelo e o aprimore, se for o caso. Tente obter o melhor modelo possível, que seja capaz de explicar razoavelmente bem a variação de alturas entre indivíduos.

Sexo	Idade	Comprimento do pé	Comprimento do Sapato	Numero do Sapato	Altura
M	67	27.8	31.3	11.0	180.3
M	47	25.7	29.7	9.0	175.3
M	41	26.7	31.3	11.0	184.8
M	42	25.9	31.8	10.0	177.8
M	48	26.4	31.4	10.0	182.3
M	34	29.2	31.9	13.0	185.4
M	26	26.8	31.8	11.0	180.3
M	29	28.1	31.0	10.5	175.3
M	60	25.4	29.7	9.5	177.8
M	48	27.9	31.4	11.0	185.4
M	30	27.5	31.4	11.0	190.5
M	43	28.8	31.6	12.0	195.0
M	54	26.7	31.8	10.0	175.3
M	31	26.7	32.4	10.5	180.3
M	42	25.1	27.6	9.0	172.7
M	21	28.7	31.8	12.5	182.9
M	59	29.2	31.3	11.0	189.2
M	58	27.9	31.3	11.5	185.4
M	42	28.6	34.5	14.0	193.7
F	47	23.2	24.8	7.0	165.1
F	19	24.3	28.6	9.0	166.4
F	20	26.0	25.4	10.0	177.8
F	27	23.8	26.7	8.0	167.6
F	19	25.1	26.7	9.0	168.3
F	21	25.4	27.9	8.5	165.7
F	32	21.9	27.9	8.0	165.1
F	19	26.2	28.9	11.0	165.1
F	27	23.8	27.9	8.0	165.1
F	18	22.2	25.9	9.5	152.4
F	26	24.6	25.4	8.5	162.6
F	36	24.6	28.1	9.0	179.1
F	28	23.7	27.6	9.0	175.9
F	29	25.6	26.5	8.5	166.4
F	58	24.1	26.5	7.0	167.6
F	30	23.8	28.4	9.0	162.6
F	23	23.3	26.5	8.0	167.6
F	26	23.5	26.0	8.0	165.1
F	47	25.1	27.0	10.0	172.7
F	36	24.1	25.1	7.5	157.5
F	19	23.8	27.9	10.0	167.6