Métodos Quantitativos
Aula 05

Metodologia de Comparações Experimentais

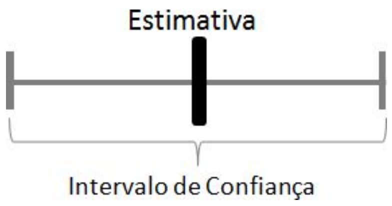
Roberto Massi de Oliveira
Alex Borges Vieira

Revisão: Amostra e População

Característica	Parâmetro (refere-se à população)	Estatística (refere-se à amostra)
Número de elementos	N	n
Média	$\bar{x}$	μ
Desvio Padrão	S	σ
Variância	S^2	σ^2

Intervalo de Confiança

- Inferência estatística:
  - Estimar um parâmetro da população através de uma estatística da amostra
- Objetivo:
  - Estabelecer generalizações seguras com base em resultados amostrais
- Intervalo de confiança para a média:
  - Usado juntamente com a média amostral para estimar a média populacional



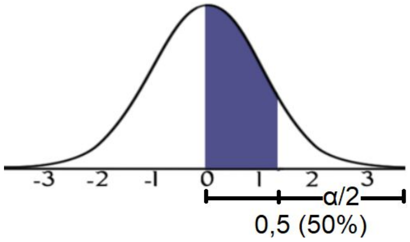
Intervalo de Confiança

- Duas fórmulas para intervalo de confiança (IC):
    - Acima de 30 elementos/amostras de qualquer distribuição:
        - Usar distribuição-z (Normal)
    - Pequenas amostras de populações **normalmente distribuídas**:
        - Usar distribuição-t (Student)
- Níveis de confiança mais usados para cálculos de IC: **90%, 95% e 99%**

Intervalo de Confiança

- Acima de 30 elementos/amostras de qualquer distribuição:
  - Usar distribuição-z (Normal)
  - Média estimada da população ($\hat{y}$), calculada a partir das médias amostrais, é dada por:

$$\hat{y} = \mu \pm Z_{0,5-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$



- **α** é o nível de significância. Nível de confiança = 100*(1 - α)%
  - Ex.: uma média estimada com 99% de confiança possui **α** = 0,01
- $Z_{0,5-\alpha/2}$ é um valor tabelado:
  - https://drive.google.com/open?id=1_MjCcbHR6IYp20uO1EC_-8gGqXjV7Eus

Intervalo de Confiança

$$\hat{y} = \mu \pm Z_{0,5-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

- Ex. (distribuição-Z): Dada uma amostra com 35 elementos {10, 16, 47, 48, 74, 30, 81, 42, 57, 67, 7, 13, 56, 44, 54, 17, 60, 32, 45, 28, 33, 60, 36, 59, 73, 46, 10, 40, 35, 65, 34, 25, 18, 48, 63}. Qual o valor estimado para a média da população com um nível de confiança de 90%?
    - μ = 42,1
    - σ = 20,1
    - n = 35
    - Nível de confiança = 90%, α = 0,1
    - $\hat{y} = 42{,}1 \pm Z_{0,45}\left(\frac{20{,}1}{\sqrt{35}}\right)$
    - $\hat{y} = 42{,}1 \pm 5.6$ = (36,5 ; 47,7)

$Z_{0,45}$??

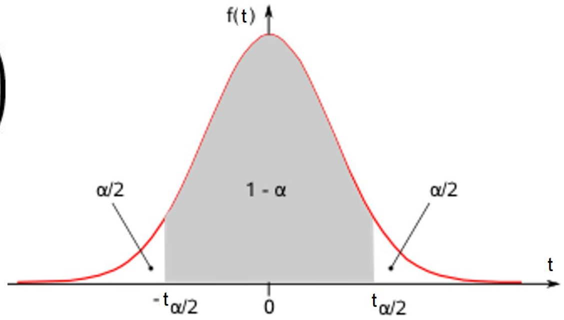
$Z_{0,45} \sim 1{,}645$

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505

Intervalo de Confiança

- Pequenas amostras de populações **normalmente distribuídas**:
  - Usar distribuição-t (Student)
  - Média estimada da população ($\hat{y}$), calculada a partir das médias amostrais, é dada por:

$$\hat{y} = \mu \pm t_{[1-\alpha;\, DF]}\left(\frac{\sigma}{\sqrt{n}}\right)$$



- DF (GL ou v) é o grau de liberdade: *DF = n - 1*
- $t_{[1-\alpha;\, DF]}$ é um valor tabelado:
  - https://drive.google.com/open?id=1ZTUXt7AOdecvKaUV8W2EQZ5TxkAdFtM7

Intervalo de Confiança

- Ex. (distribuição-t): Dada uma amostra com 10 elementos normalmente distribuídos {148, 166, 170, 191, 187, 114, 168, 180, 177, 204}. Qual o valor estimado para a média da população com um nível de confiança de 99%?
  - μ = 170,5
  - σ = 25,1
  - n = 10, *DF = n - 1 = 9*
  - Nível de confiança = 99%, α = 0,01
  - $\hat{y} = 170{,}5 \pm t_{[0,99;\,9]}\left(\frac{25{,}1}{\sqrt{10}}\right)$
  - $\hat{y} = 170{,}5 \pm 25{,}8$ = (144,7 ; 196.3)

$t_{[0,99;\,9]}??$ $\qquad$ $t_{[0,99;\,9]} = 3{,}2498$

	80%	90%	95%	98%	99%	99.9%	2-Tail Confidence Level
df	0.20	0.10	0.05	0.02	0.01	0.001	2-Tail Alpha
1	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192	
2	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991	
3	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240	
4	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103	
5	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688	
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588	
7	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079	
8	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413	
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809	

Intervalo de Confiança

- Em Python:

```python
from scipy import stats
import math

interval = 0.95
z = stats.norm.ppf((interval + 1)/2)
lower_bound_z = u - z * stdev / math.sqrt(n)
upper_bound_z = u + z * stdev / math.sqrt(n)

t = stats.t.ppf((interval + 1)/2, n)
lower_bound_t = u - t * stdev / math.sqrt(n)
upper_bound_t = u + t * stdev / math.sqrt(n)
```

Intervalo de Confiança

$$\hat{y} = \mu \pm t_{[1-\alpha; DF]} \left(\frac{\sigma}{\sqrt{n}}\right)$$

- Quanto maior o intervalo de confiança, menor a **precisão** da estimativa
- O tamanho do intervalo de confiança é influenciado pelo nível de confiança
  - Maior o nível de confiança, menor a precisão (maior o intervalo de confiança)

Student's t Distribution Table

	90%	95%	97.5%	99%	99.5%	99.95%	1-Tail Confidence Level
	80%	90%	95%	98%	99%	99.9%	2-Tail Confidence Level
	0.100	0.050	0.025	0.010	0.005	0.0005	1-Tail Alpha
df	0.20	0.10	0.05	0.02	0.01	0.001	2-Tail Alpha
1	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192	

- Melhorar precisão sem mexer no nível do confiança, aumentar **n**

Tamanho da Amostra

- É possível calcular o tamanho ideal de uma amostra, dependendo do intervalo de confiança desejado

$$n=\left(\frac{Z_{0,5-\alpha/2}\,\sigma}{e\,\mu}\right)^2$$

- e é o erro da estimativa:
  - margem de erro admitida
  - diferença máxima entre a MÉDIA AMOSTRAL e a verdadeira MÉDIA POPULACIONAL.

Tamanho da Amostra

- Ex.: Suponha que o tempo médio para gravar um arquivo é 7,94 seg com desvio padrão de 2,14. Aproximadamente, quantas medidas serão requeridas se nós desejamos um IC de 90% e que a média esteja dentro de um intervalo de 3.5%.

$$n = \left(\frac{z_{0,5-\alpha/2}\sigma}{e\mu}\right)^2 = \left(\frac{1.645(2.14)}{0.035(7.94)}\right)^2 = 160.46$$

$$n = 161$$

Tamanho da Amostra

- Se soubermos o erro em porcentagem ($\pm r$ %), temos:

$$n = \left(\frac{100 z \sigma}{r \mu}\right)^2$$

- Para uma proporção, temos:

$$n = z^2 \frac{p(1-p)}{r^2}$$

Proporções

- Se n_1 de n experimentos dão um certo resultado, então pode-se dizer que a proporção das amostras é dada por:

$$p = \frac{n_1}{n}$$

- Intervalo de confiança da proporção, se *np > 10*, é dado por:

$$IC \rightarrow p \pm z_{0,5-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

Proporções

$$IC \rightarrow p \pm z_{0,5-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

- Ex.: Um experimento foi repetido em dois sistemas 40 vezes. O sistema A foi superior ao B em 26 repetições. Podemos dizer que, com confiança de 99%, o sistema A é superior ao sistema B? E com uma confiança de 90%?

$$\underbrace{p}_{p_A} = \frac{26}{40} = 0.65 \qquad \sqrt{\frac{p(1-p)}{n}} = 0.075$$

$$p_B = 1 - p_A$$

0,5

$$p_A < p_B \qquad\qquad p_A > p_B$$

$$p_A = p_B$$

Proporções

$$IC \rightarrow p \pm z_{0,5-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

- Ex.: Um experimento foi repetido em dois sistemas 40 vezes. O sistema A foi superior ao B em 26 repetições. Podemos dizer que, com confiança de 99%, o sistema A é superior ao sistema B? E com uma confiança de 90%?

$$p = \frac{26}{40} = 0.65 \qquad \sqrt{\frac{p(1-p)}{n}} = 0.075$$

$$\text{IC de } 99\% = 0.65 \pm (2.576)0.075 = (0.46, 0.84)$$

IC inclui <= 0.5, logo, A não é superior a B para o nível de confiança de 99%

$$\text{IC de } 90\% = 0.65 \pm (1.645)0.075 = (0.53, 0.77)$$

IC > 0.5, logo, A é superior a B para o nível de confiança de 90%

Significância Estatística

“Statisticians are largely concerned with whether observations on data are significant. Computational analysis will readily find a host of patterns and correlations in any interesting data set. But does a particular correlation reflect a real phenomena, as opposed to just chance? In other words, when is an observation really significant?”

SKIENA, S. S. The Data Science Design Manual. Springer, 2017.

Significância Estatística

- Significância estatística serve para mostrar se há ou não genuína diferença entre duas amostras

- Não mede a importância ou a magnitude dessa diferença

- Pequenas diferenças numa amostra podem ter elevada significância em testes estatísticos

- Testes de significância também são conhecidos como testes de hipótese

Teste de Hipótese

- Hipótese Nula (H_0):
  - Afirma que não existe relação entre duas amostras comparadas
  - Assumindo-se H_0 , tenta-se provar, através de testes de significância, que não há diferença entre os alvos de uma comparação estatística
  - Ex.: Hipótese nula: observando-se resultados de desempenho, GPUs NVIDIA e Radeon não possuem diferença significativa entre si
- Hipótese Alternativa (H_1):
  - Se opõe à hipótese nula, afirmando que há significativa diferença entre amostras comparadas
  - Tenta-se provar que há diferença entre os alvos de uma comparação estatística
  - Ex.: Hipótese alternativa: observando-se resultados de desempenho, NVIDIA é bem melhor

Comparando Amostras

- Num projeto de pesquisa, geralmente, procura-se o melhor sistema, o melhor algoritmo:
    - Ex.:
        - Determinar o sistema que apresenta a melhor relação QoS-preço, onde QoS é medido experimentalmente
        - Mostrar que um algoritmo Y executa mais rápido que outros existentes e sejam similares funcionalmente
- Métodos diferentes para observações pareadas e não-pareadas
    - Pareadas se cada observação em uma amostra for comparável com observações em outra
    - Não-pareadas, caso contrário

Comparando Observações Pareadas

1. Tratar o problema como uma amostra de n pares

2. Para cada teste: calcule as diferenças dos resultados

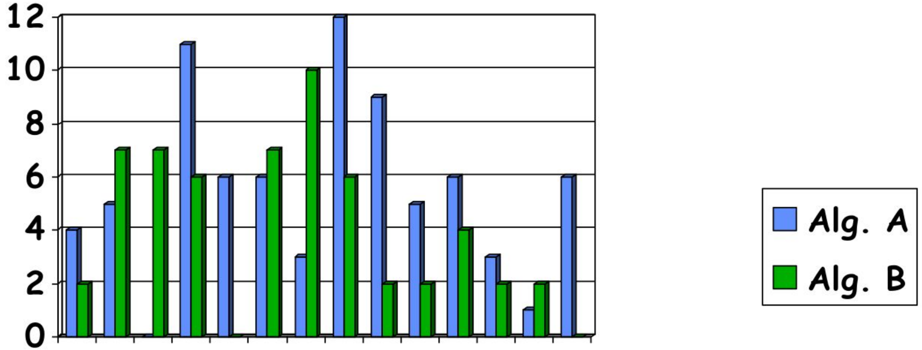
3. Calcule o intervalo de confiança para a média das diferenças ($\hat{y}_{\mu_{(A-B)}}$)

4. Se o intervalo inclui 0 (zero), os objetos de comparação (ex.: sistemas, algoritmos, etc) não são diferentes com a dada confiança

5. Se o intervalo não inclui zero, o sinal da diferença indica qual dos objetos é melhor, baseado nos dados experimentais

Comparando Observações Pareadas

- Ex.: Considere dois métodos de busca A e B que são avaliados em função do nº de documentos relevantes (em um total de 100) que cada um retorna. Num teste com várias consultas, o algoritmo A retorna mais documentos relevantes que o B? Amostra de testes com 14 consultas:



Alg. A	4	5	0	11	6	6	3	12	9	5	6	3	1	6
Alg. B	2	7	7	6	0	7	10	6	2	2	4	2	2	0

Comparando Observações Pareadas

- Ex. (continuação):

Alg. A	4	5	0	11	6	6	3	12	9	5	6	3	1	6
Alg. B	2	7	7	6	0	7	10	6	2	2	4	2	2	0

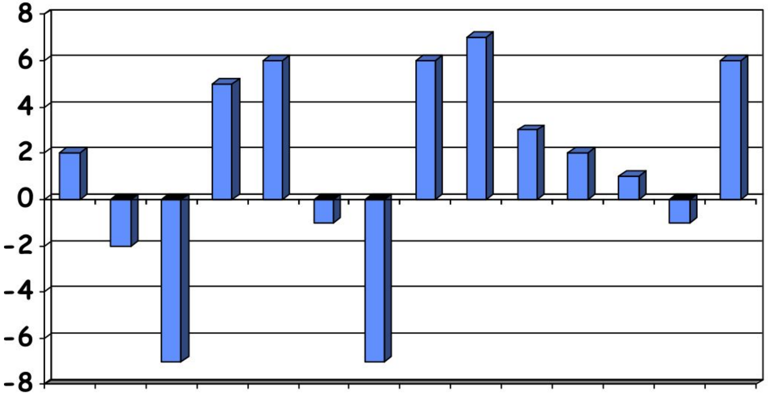
$\hat{y}_{\mu_{(A-B)}} = 0$

$A < B$ $\quad$ $A > B$

$A = B$

Comparando Observações Pareadas

- Ex. (continuação):



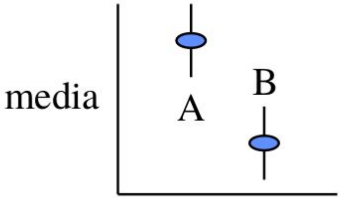
  - Diferenças entre algoritmos (A - B): 2, -2, -7, 5, 6, -1, -7, 6, 7, 3, 2, 1, -1, 6
  - Média 1.4, intervalo de 90% (-0.75, 3.6)
  - Intervalo de confiança inclui 0, portanto, não há diferença significativa entre A e B
  - Intervalo de 70% é (0.10, 2.76), A tem desempenho melhor que B

Comparando Observações Não-Pareadas

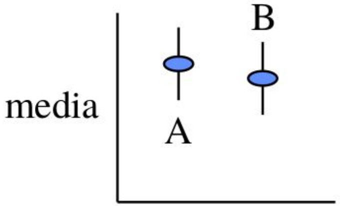
1. Calcule a média das amostras

2. Calcule o intervalo de confiança (IC)

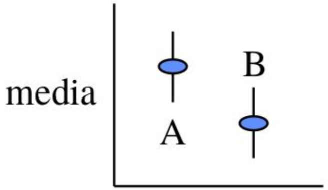
    a. Se não houver sobreposição:
        - amostras são diferentes



    b. Se houver sobreposição e cada (IC) contém a outra média:
        - Amostras não são diferentes neste nível

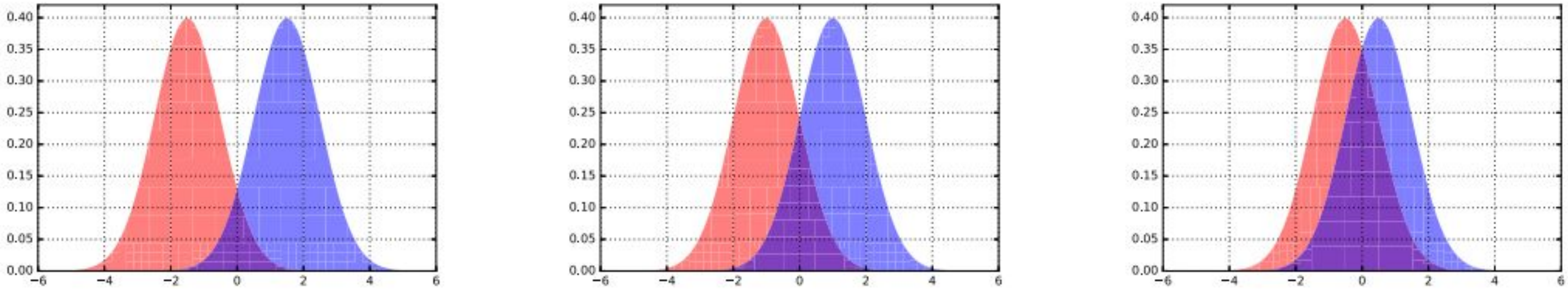


    c. Se houver sobreposição e uma média não está no IC da outra:
        - Fazer Teste-T



Teste-T

- Também conhecido como teste **T de Student**
- Verifica se diferença entre médias de 2 amostras/populações é significativa



- Ex.: São feitos 20 testes de Q.I. em homens e mulheres. As médias dos resultados diferem. Essa diferença é **significativa**?

Teste-T

- Conclusões **preliminares** de significância:
  - Grande diferença entre duas médias
    - Ex.: No brasil, peso médio de homens é de 73 kg e de mulheres é de 63 kg
  - Pequeno desvio padrão
    - Ex.: Homens e mulheres têm, em média, 10 dedos. Os valores individuais são próximos da média ({10,10,9,10,10,10…}). Se fossem muito dispersos, dificultariam conclusões
  - Grande número de amostras
    - Ex.: Pode-se afirmar que, em média, homens possuem menos dedos que mulheres, por trabalharem mais com ferramentas que podem decepá-los. Para validar essa afirmação sobre esse raro evento, seria necessário um grande número de amostras

Teste-T

- Intervalo de confiança:

$$(\mu_1 - \mu_2) \pm t_{[1-\alpha;v]}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- μ_i , σ_i e n_i são a média, o desvio padrão e o tamanho da amostra **i**
- Grau de liberdade (**df** ou **v**):

$$\nu = \frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{\sigma_1^4}{n_1^2(n_1-1)} + \frac{\sigma_2^4}{n_2^2(n_2-1)}} - 2$$

- Tabela T de student: https://drive.google.com/open?id=1ZTUXt7AOdecvKaUV8W2EQZ5TxkAdFtM7
- Se intervalo de confiança inclui 0, não há significância

Teste-T

- Ex.: O tempo de processamento necessário para executar uma tarefa foi medido em dois sistemas. Esses sistemas são significativamente diferentes num nível de confiança (NC) de 90%?

Sistema A: {5.36, 16.57, 0.62, 1.41, 0.64, 7.26}.
Sistema B: {19.12, 3.52, 3.38, 2.50, 3.60, 1.74}

$\mu_A = 5{,}31 \quad \sigma_A^2 = 31{,}60$

$\mu_B = 5{,}64 \quad \sigma_B^2 = 36{,}76$

$v = 7{,}94 \sim 8$

$\hat{y} = (5{,}31 - 5{,}64) +/- 1{,}8595(3{,}38)$

$\hat{y} = (-6{,}61\ ;\ 5{,}94)$

$t_{[0,90;8]}$??

$t_{[0,90;8]} = 1{,}8595$

	80%	90%	95%	2-Tail Confidence Level
df	0.20	0.10	0.05	2-Tail Alpha
1	3.0777	6.3138	12.7062	
2	1.8856	2.9200	4.3027	
3	1.6377	2.3534	3.1824	
4	1.5332	2.1318	2.7764	
5	1.4759	2.0150	2.5706	
6	1.4398	1.9432	2.4469	
7	1.4149	1.8946	2.3646	
8	1.3968	1.8595	2.3060	

$\boxed{H_0}$

Intervalo de confiança inclui **0**, logo a diferença entre A e B não é significante para NC = 90%

Teste-T

- Em Python, o cálculo de t é dado por:

```python
import numpy as np
from scipy import stats
```

```python
n = 10
a = np.random.randn(n)+2
b = np.random.randn(n)
```

```python
t,p = stats.ttest_ind(a,b)
print("t:",t,"\np:", p)
```

```
t: 2.043227621046106
p: 0.0559516345812018
```

- Observem que há cada valor de **t** há um valor de **p** (p-value) associado
- Quanto menor o valor de **p**, maior a significância estatística
  - Ex.: Se p = 0,05, isso significa que há apenas 5% de chance da diferença ser ao acaso (H_0)

Teste Qui-Quadrado

- Verificar se a frequência com que um determinado acontecimento observado em uma amostra se desvia significativamente ou não da frequência com que ele é esperado
- Teste não paramétrico: não depende de estatísticas amostrais (e.g., média, variância)
- Condições:
  - Grupos independentes
  - Itens aleatoriamente selecionados (sem viés)
  - As observações devem ser frequências ou contagens
  - Frequência observada maior que 5 (entre 5 e 10, também não é muito bom sem correção)

Teste Qui-Quadrado

- H_0 : frequências observadas = frequências esperadas (casualidade)
- H_1 : frequências observadas ≠ frequências esperadas (significativa)
- O valor de χ^2 ao nível de significância α é denominado χ^2 crítico ou tabelado
- Graus de Liberdade (GL, DF ou v):
  - GL (1 grupo) = r - 1, onde k é o número de linhas da tabela
  - GL (2 grupos) = (r-1)(c-1), onde r é o número de linhas e c é o número de colunas da tabela
  - Obs.: A quantidade de grupos é definida pela quantidade de colunas com valores observados

Teste Qui-Quadrado

- Regras de decisão:
  - É necessário obter duas estatísticas :
    - X^2 calculado: obtido diretamente dos dados das amostras.
    - X^2 tabelado: depende do número de graus de liberdade e do α adotado
  - Se X^2 calculado ≥ X^2 tabelado: Rejeita-se H_0
  - Se X^2 calculado < X^2 tabelado: Aceita-se H_0
  - Quando o valor α não for pré-definido, usa-se 5% por padrão

Teste Qui-Quadrado

- Ex. (1 grupo): Foi desenvolvido um algoritmo para gerar números aleatórios inteiros no intervalo 0-9. Ao executar o algoritmo e gerar 1000 valores, deseja-se verificar se há ou não aleatoriedade nos resultados. Para isso, a tabela a seguir contém as frequências observadas:

	o (observado)	e (esperado)	d^2/e (d = o - e)
0	94	100	0,36
1	93	100	0,49
2	112	100	1,44
3	101	100	0,01
4	101	100	0,01
5	104	100	0,16
6	95	100	0,25
7	100	100	0
8	99	100	0,01
9	101	100	0,01

Frequência esperada (e) = 1/10

$$H_0 = aleatório\ (não\ há\ significância)$$

$$H_1 = não\ aleatório\ (há\ significância)$$

$$X^2(calculado) = \sum \frac{(o-e)^2}{e} = \sum \frac{d^2}{e} = 2{,}74$$

Teste Qui-Quadrado

- Ex. (1 grupo - continuação):
  - GL = r - 1 = 10 - 1 = 9
  - **α** = 5% por padrão

$$X^2(tabelado) = 16{,}92$$

X² calculado (2,74) < X² tabelado (16,92)

Comprova-se H_0

$$H_0 = aleatório\ (não\ há\ significância)$$

Graus de Liberdade

p ▶	99,5%	99%	97,5%	95%	90%	10%	5%
1	0,000	0,000	0,001	0,004	0,016	2,706	3,841
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488
5	0,412	0,554	0,831	1,145	1,610	9,236	11,070
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067
8	1,344	1,646	2,180	2,733	3,490	13,362	15,507
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919

- Tabela Qui-Quadrado:
  - https://drive.google.com/open?id=117bXzIpnwWRL7JKY73knYw8mdt-T71tZ

Teste Qui-Quadrado

- Correção de Yates (1 grupo):

$$\chi^2 = \frac{(|o_1 - e_1| - \boxed{0{,}5})^2}{e_1} + \frac{(|o_2 - e_2| - \boxed{0{,}5})^2}{e_2}$$

- Regras de uso:
  - Se X^2 calculado > X^2 tabelado
    - N < 40 (N, nesse caso, é o tamanho total da amostra)
    - Ocorrência de pelo menos uma frequência esperada menor que 5 (e < 5)

Teste Qui-Quadrado

- Teste de independência:
  - Pode-se verificar se frequências comparadas são, de fato, independentes
- Ex. (3 grupos): Um inspetor de qualidade toma uma amostra de 220 artigos num centro de distribuição. Se sabe que cada produto pode vir de uma de três fábricas e pode ou não estar defeituoso. O inspetor avalia todos os produtos e obtém os seguintes resultados:
  - GL = (r-1)(c-1) = (2-1)(3-1) = 2
  - α padrão = 5%
  - X² tabelado = 5,991

Tabela de resultados observados

	F_1	F_2	F_3	
D	8	15	11	34
ND	62	67	57	186
	70	82	68	220

$H_0 =$ A proporção de produtos defeituosos é a mesma para todas as fábricas (são independentes)

Teste Qui-Quadrado

Tabela de resultados observados

	F_1	F_2	F_3	
D	8	15	11	34
ND	62	67	57	186
	70	82	68	220

$$E_{11} = \frac{70 \times 34}{220} = 10.810 \qquad E_{21} = \frac{70 \times 186}{220} = 59.180$$

$$E_{12} = \frac{82 \times 34}{220} = 12.673 \qquad E_{22} = \frac{82 \times 186}{220} = 69.327$$

$$E_{13} = \frac{68 \times 34}{220} = 10.509 \qquad E_{23} = \frac{68 \times 186}{220} = 57.490$$

Tabela de resultados esperados

	F_1	F_2	F_3	
D	10.81	12.67	10.51	34
ND	59.18	69.33	57.49	186
	70	82	68	220

$$\chi^2 = \sum \frac{d^2}{e} : \frac{(8 - 10.81)^2}{10.81} + \dots + \frac{(57 - 57.49)^2}{57.49} = 1.398$$

- X² calculado (1,40) < X² tabelado (5,99): comprova-se H_0 (há independência)

Teste Qui-Quadrado

- Se GL = 1, não há necessidade de usar valores esperados (e)
- Ex. (tabela 2x2):

Sexo	Maligna	Benigna	Total
M	15 (a)	35 (b)	50 (n1)
F	6 (c)	24 (d)	30 (n2)
Total	21 (n3)	59 (n4)	80 (N)

$$\chi^2 = \frac{(ad - bc)^2 N}{n1 \;\; n2 \;\; n3 \;\; n4}$$

X^2 calculado = 0,968

X^2 tabelado = 3,841

Teste Qui-Quadrado

- Ainda para o caso de GL = 1:

Correção de Yates / Continuidade

$$\chi^2 = \frac{(|ad - bc| - \boxed{0{,}5.N})^2\, N}{n1\ \ n2\ \ n3\ \ n4}$$

REGRAS DE APLICAÇÃO:

1) o valor de X^2 obtido é **maior** que o crítico e o valor de N é menor que 40 ou
2) o valor de X^2 obtido é **maior** que o crítico e há pelo menos uma classe com frequência esperada menor que 5.

X^2 calculado = 0,968 < X^2 tabelado = 3,841

NÃO É NECESSÁRIO

Teste Qui-Quadrado

- Python:

```
import numpy as np
from scipy import stats

observed = np.array([20., 20., 0., 0.])
expected = np.array([.25, .25, .25, .25]) * np.sum(observed)
x2 = stats.chisquare(observed, expected)
```

x2 =

Índice	Tipo	Tamanho	Valor
0	float64	1	40.0
1	float64	1	1.065509033425585e-08

- A saída da função é composta por 2 parâmetros (tabela da parte direita da figura):
  - Índice 0 (40.0, no exemplo): X^2 calculado
  - Índice 1 (1.065e-08, no exemplo): p-value (quanto menor o p-value, maior a significância)
    - Se p = 0,01, há apenas 1% de chance de constatação da hipótese H_0

Teste Kolmogorov-Smirnov

- Observa a máxima diferença absoluta entre a distribuição assumida para os dados, no caso a Normal, e a distribuição empírica dos dados

- Pode ser usado para avaliar as hipóteses:
    - H_0 : os dados seguem uma distribuição normal (não há significativa diferença)
    - H_1 : os dados não seguem uma distribuição normal (há significativa diferença)

- Pode ser usado para testar a amostra antes do uso da distribuição-t

- O valor crítico, assim como nos demais testes, é tabelado:
    - https://drive.google.com/open?id=1Ujc39foBoqPJvYCNKKx9MbRpOTTY0CAI

Teste Kolmogorov-Smirnov

x_i	$\hat{F}_n(x_i)$	$\hat{F}_n(x_i^-)$	$z_i = \frac{x_i - \bar{x}}{s}$	$F_0(z_i)$ $(F_0(x_i))$	$\lvert F_0(x_i) - \hat{F}_n(x_i) \rvert$ **(1)**	$\lvert F_0(x_i) - \hat{F}_n(x_i^-) \rvert$ **(2)**
1 representante de cada elemento da amostra	frequência relativa acumulada	0, frequência relativa acumulada de X_0 a X_{n-1}	Z calculado	tabela Z acumulada	tabela Z acumulada (menos) frequência relativa acumulada	tabela Z acumulada (menos) 0, frequência relativa acumulada de X_0 a X_{n-1}

$$D_n = \max((1),(2))$$

Se $D_n\ calculado > D_n\ crítico$(tabelado), rejeita-se H_0 com $(1-\alpha)100\%$ de confiança. Caso contrário, comprova-se H_0.

Obs: Tabela Z <u>Acumulada</u>

https://drive.google.com/open?id=1pepnGOaIs0C1jTJAZ-gvtWUurZUz5216

Teste Kolmogorov-Smirnov

- Ex.: O RU deve servir, por contrato, uma média de 290 g de carne por refeição. No entanto, alguns alunos queixaram-se da quantidade de carne servida. O cozinheiro chefe lhes disse que a quantidade de carne servida por refeição tinha aproximadamente **distribuição normal** com **média 290 g** e **desvio padrão de 56 g**. Duvidando disso, alguns alunos recolheram as suas refeições ao longo de vários dias, resultando em uma amostra de 10 refeições. Os dados obtidos são os seguintes:

  198 254 262 272 275 278 285 287 287 292

  Ao nível de significância de 5%, há evidência para rejeitar a hipótese de que o cozinheiro seguia as regras?

Teste Kolmogorov-Smirnov

- Ex. (continuação):

$H_0: X \sim N(290, 56^2)$ $\quad$ $H_1: X \not\sim N(290, 56^2)$.

Vamos calcular o F(x) do primeiro elemento para aprender a usar a tabela Z acumulada:

198 254 262 272 275 278 285 287 287 292

$$F_0(198) = P(X \leq 198) = P(Z \leq \frac{198 - 290}{56}) =$$

$$= P(Z \leq -1.64) = 1 - 0.9495 = 0.0505$$

z	0,00	0,01	0,02	0,03	0,04
0,0	0,5000	0,5040	0,5080	0,5120	0,5160
0,1	0,5398	0,5438	0,5478	0,5517	0,5557
0,2	0,5793	0,5832	0,5871	0,5910	0,5948
0,3	0,6179	0,6217	0,6255	0,6293	0,6331
0,4	0,6554	0,6591	0,6628	0,6664	0,6700
0,5	0,6915	0,6950	0,6985	0,7019	0,7054
0,6	0,7257	0,7291	0,7324	0,7357	0,7389
0,7	0,7580	0,7611	0,7642	0,7673	0,7704
0,8	0,7881	0,7910	0,7939	0,7967	0,7995
0,9	0,8159	0,8186	0,8212	0,8238	0,8264
1,0	0,8413	0,8438	0,8461	0,8485	0,8508
1,1	0,8643	0,8665	0,8686	0,8708	0,8729
1,2	0,8849	0,8869	0,8888	0,8907	0,8925
1,3	0,9032	0,9049	0,9066	0,9082	0,9099
1,4	0,9192	0,9207	0,9222	0,9236	0,9251
1,5	0,9332	0,9345	0,9357	0,9370	0,9382
1,6	0,9452	0,9463	0,9474	0,9484	0,9495

Obs.: Se $z < 0$, então $\Phi(z) = P(-\infty < Z \leq z) = 1 - \Phi(-z)$.

Teste Kolmogorov-Smirnov

- Ex. (continuação):

x_i	$\hat{F}_{10}(x_i)$	$\hat{F}_{10}(x_i^-)$	$z_i = \frac{x_i - 290}{56}$	$F_0(z_i)$	$\lvert F_0(x_i) - \hat{F}_{10}(x_i) \rvert$	$\lvert F_0(x_i) - \hat{F}_{10}(x_i^-) \rvert$
198	0,1	0	-1,64	0,0505	*0,0495*	*0,0505*
254	0,2	0,1	-0,64	0,2611	*0,0611*	*0,1611*
262	0,3	0,2	-0,5	0,3085	*0,0085*	*0,1085*
272	0,4	0,3	-0,32	0,3745	*0,0255*	*0,0745*
275	0,5	0,4	-0,27	0,3936	*0,1064*	*0,0064*
278	0,6	0,5	-0,21	0,4168	*0,1832*	*0,0832*
285	0,7	0,6	-0,09	0,4641	*0,2359*	*0,1359*
287	0,9	0,7	-0,05	0,4801	*0,4199*	*0,2199*
292	1	0,9	0,04	0,516	***0,484***	*0,384*

$$D_{10} = \max(0{,}484; 0{,}384) = 0{,}484$$

Falta calcular o D crítico (tabelado)

Teste Kolmogorov-Smirnov

- Ex. (continuação): Encontrando D crítico (tabelado) para um nível de confiança de 5% (do enunciado), na tabela Kolmogorov-Smirnov:

			α
n	0.20	0.10	0.05
1	0.900	0.95	0.975
2	0.684	0.776	0.842
3	0.565	0.636	0.708
4	0.493	0.565	0.624
5	0.447	0.509	0.563
6	0.410	0.468	0.519
7	0.381	0.436	0.483
8	0.358	0.410	0.454
9	0.339	0.387	0.430
10	0.323	0.369	0.409

H_0: $X \sim N(290, 56^2)$ H_1: $X \not\sim N(290, 56^2)$.

D calculado (0,484) > D crítico (0,409)

Portanto, rejeita-se a hipótese nula. Ou seja, a regra do peso da carne não está sendo devidamente seguida pelo RU.

Comprova-se hipótese alternativa. A amostra não segue a distribuição normal especificada.

Teste Kolmogorov-Smirnov

- Python:

```python
import numpy as np
from scipy import stats


d = stats.kstest(stats.t.rvs(100,size=100),'norm')
```

d =

Índice	Tipo	Tamanho	Valor
0	float64	1	0.07201892916547126
1	float64	1	0.6763006286247913

- A saída da função é composta por 2 parâmetros (tabela da parte direita da figura):
    - Índice 0 (0.0072, no exemplo): d calculado
    - Índice 1 (0.6763, no exemplo): p-value (quanto menor o p-value, maior a significância)
        - Se p = 0.67, há 67,63% de chance de constatação da hipótese H_0

AAG04 Tarefa em Dupla

- Desenvolva em Python um script que execute os seguintes passos:
    1. Escolha um contexto (tema) e colete ou gere pequenas amostras comparáveis, com <u>n < 30</u>
    2. Execute Kolmogorov-Smirnov para verificar se as amostras seguem distribuição Normal (p-value de pelo menos 0.8)
    3. Repita de 1 e 2 até encontrar pelo menos 2 amostras Normais
    4. Verifique se as médias das amostras possuem diferença significativa (Teste-T)
    5. Calcule o intervalo de confiança para as médias para níveis de confiança de 90%, 95% e 99%
    6. **Apresentar conclusões sobre os resultados obtidos nas etapas 2, 4 e 5**

- Regras:
    1. Funções prontas de bibliotecas python DEVEM ser usadas ao máximo possível
    2. Código e resultados devem ser explicados em Markdown com comandos LaTex
    3. Os formatos de entrega devem ser .pdf e .ipynb (código fonte+markdowns)
    4. Os dados devem ser entregues em anexo