



Universidade de Brasília - UnB  
Instituto de Ciências Exatas - IE  
Departamento de Estatística - EST

# **Principais medidas de magnitude do efeito utilizadas na comparação de dois grupos**

Marcos Douglas Rodrigues de Sousa

Orientador: Prof. Dr. Eduardo Yoshio Nakano

Brasília

2018



Marcos Douglas Rodrigues de Sousa

**Principais medidas de magnitude do efeito utilizadas na  
comparação de dois grupos**

Monografia apresentada para obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Eduardo Yoshio Nakano

Brasília

2018

Marcos Douglas Rodrigues de Sousa

Principais medidas de magnitude do efeito utilizadas na comparação de dois grupos/ Marcos Douglas Rodrigues de Sousa. – Brasília, 2018-  
109 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. Eduardo Yoshio Nakano

Relatório Final – Universidade de Brasília

Instituto de Ciências Exatas

Departamento de Estatística

Trabalho de Conclusão de Curso de Graduação, 2018.

1. TCC 1. 2. TCC 2. 3. Pesquisa. 4.  $\LaTeX$ . 5. Motivação. 6. Dedicção.

Marcos Douglas Rodrigues de Sousa

## **Principais medidas de magnitude do efeito utilizadas na comparação de dois grupos**

Monografia apresentada para obtenção do título de Bacharel em Estatística.

Trabalho aprovado. Brasília, 10 de julho de 2018:

---

**Prof. Dr. Eduardo Yoshio Nakano**  
Orientador

---

**Prof. Dr. André Luiz Fernandes  
Cançado**  
Membro da Banca

---

**Prof. Dr. Helton Saulo Bezerra dos  
Santos**  
Membro da Banca

Brasília  
2018



# Agradecimentos

Agradeço a Deus, por ter me concedido força e fé para prosseguir no curso e assim realizar este trabalho.

Agradeço aos meus pais, Vanderleia e Valter por terem sido meu suporte durante o período do curso, sem eles não seria possível. Agradeço à minha irmã Milene, por ser uma grande companheira e por sempre me trazer ânimo.

Agradeço ao orientador deste trabalho, o Professor Eduardo Nakano que sempre atendeu as minhas dúvidas e tornou possível a realização deste trabalho. Agradeço aos demais professores do departamento de Estatística pelo aprendizado ao longo desses anos.

Agradeço a todos os amigos e colegas de curso pelo companherismo durante esses anos. E por fim, agradeço ao pessoal da secretaria do departamento, que sempre estiveram disponíveis para nos atender.





*“A persistência é o caminho  
do êxito.”  
(Charles Chaplin)*



# Resumo

As medidas de magnitude do efeito (*Effect Size*) trazem informações sobre a significância prática dos resultados obtidos em estudos que realizam a comparação de dois grupos de tratamentos. Neste tipo de estudo, as pesquisas científicas reportam a significância estatística dos resultados obtidos nos testes de hipóteses através de seu nível descritivo (*valor-p*). Portanto, o *Effect Size* é uma forma para avaliar a dimensão ou o tamanho do efeito entre os dois grupos e permite que os pesquisadores se direcionem no estudo descritivo da magnitude do efeito entre os grupos. O objetivo deste trabalho foi realizar um estudo das principais medidas de magnitude do efeito adotadas na comparação de dois grupos, verificando qual medida é a mais adequada para determinada situação e a interpretação do resultado obtido no cálculo dessas medidas. Os procedimentos utilizados neste trabalho foram ilustrados por meio de dados simulados.

**Palavras-chave:** *Effect Size*, magnitude do efeito, significância prática, testes de hipóteses, comparação de grupos, *valor-p*.



# Abstract

The Effect Size measures information on the practical significance of the results obtained in studies comparing two groups of treatments. In this type of study, scientific researches report the statistical significance of the results obtained in the hypothesis tests through its descriptive level ( $p$ -value). Therefore, the Effect Size is a form to evaluate the size of the effect between two groups and allows researchers to address the descriptive study of effect size between groups. The objective of this project was to perform a study of the main measures of effect size adopted in the comparison of two groups, verifying which measure is the most adequate for a given situation and the interpretation of the result obtained in the calculation of these measures. The procedures used in this work were illustrated of simulated data.

**Keywords:** Effect Size, practical significance, hypothesis tests, comparison of two groups,  $p$ -value.



# Lista de ilustrações

- Figura 1 – Gráfico de curvas com valores de  $p$  do teste Qui-Quadrado segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ ) 61
- Figura 2 – Gráficos de curvas com os valores de  $p$  do teste de McNemar segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ ) 64
- Figura 3 – Gráfico de curvas com os valores de  $p$  do teste de Mann-Whitney segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ ) 67
- Figura 4 – Gráfico de curvas com os valores de  $p$  do teste de Wilcoxon segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ ) 70
- Figura 5 – Gráfico de curvas com os valores de  $p$  do teste  $t$  não-pareado segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ ) 73
- Figura 6 – Gráfico de curvas com os valores de  $p$  do teste  $t$  não-pareado segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ ) 76
- Figura 7 – Gráfico de curvas com os valores de  $p$  do teste  $t$  pareado segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ ) 78
- Figura 8 – Principais medidas de magnitude do efeito segundo escala de mensuração 83





# Lista de tabelas

Tabela 1 – Tabela de contingência $2 \times 2$ . . . . .	21
Tabela 2 – Tabela de contingência $2 \times 2$ . . . . .	23
Tabela 3 – Tabela de contingência $2 \times 2$ . . . . .	25
Tabela 4 – Valores Críticos ( $W_c$ ) do teste de Wilcoxon . . . . .	28
Tabela 5 – Valores Críticos ( $d$ ) do teste de Mann-Whitney . . . . .	32
Tabela 6 – Níveis de Magnitude do Efeito ( $ES$ ) do Coeficiente $\phi$ . . . . .	40
Tabela 7 – Níveis de Magnitude do Efeito ( $ES$ ) do $W$ de Cohen . . . . .	41
Tabela 8 – Níveis de Magnitude do Efeito ( $ES$ ) do $V$ de Cramér . . . . .	42
Tabela 9 – Tabela de contingência $2 \times 2$ . . . . .	42
Tabela 10 – Tabela de contingência $2 \times 2$ . . . . .	43
Tabela 11 – Níveis de Magnitude do Efeito ( $ES$ ) da $OR$ e $RR$ . . . . .	43
Tabela 12 – Níveis de Magnitude do Efeito ( $ES$ ) do $Q$ de Yule . . . . .	45
Tabela 13 – Níveis de Magnitude do Efeito ( $ES$ ) do $\delta$ de Cliff . . . . .	47
Tabela 14 – Níveis de Magnitude de Efeito ( $ES$ ) do $A$ de Vargha-Delaney . . . . .	48
Tabela 15 – Níveis de Magnitude do Efeito ( $ES$ ) do Coeficiente $r$ . . . . .	49
Tabela 16 – Níveis de Magnitude do Efeito ( $ES$ ) do $\tau$ de Kendall . . . . .	50
Tabela 17 – Níveis de Magnitude do Efeito ( $ES$ ) do $d$ de Cohen . . . . .	53
Tabela 18 – Níveis de Magnitude do Efeito ( $ES$ ) do $g$ de Hedges . . . . .	54
Tabela 19 – Níveis de Magnitude do Efeito ( $ES$ ) do $\Delta$ de Glass . . . . .	56
Tabela 20 – <i>Effect Size</i> muda ao duplicar o tamanho da amostra? . . . . .	57
Tabela 21 – Valores $p$ do teste Qui-Quadrado segundo o tamanho da amostra $n$ e Coeficiente $\phi$ . . . . .	60
Tabela 22 – Tamanhos de amostra $n$ segundo valores $p$ do teste Qui-Quadrado e Coeficiente $\phi$ . . . . .	62
Tabela 23 – Valores $p$ do teste de McNemar segundo o tamanho da amostra $n$ e Coeficiente $\phi$ . . . . .	63
Tabela 24 – Tamanhos de amostra $n$ segundo valores $p$ do teste de McNemar e Coeficiente $\phi$ . . . . .	64
Tabela 25 – Valores $p$ do teste de Mann-Whitney segundo o tamanho da amostra $n$ e $\delta$ de Cliff . . . . .	66
Tabela 26 – Tamanhos de amostra $n$ segundo valores $p$ do teste de Mann-Whitney e $\delta$ de Cliff . . . . .	68
Tabela 27 – Valores $p$ do teste de Wilcoxon segundo o tamanho da amostra $n$ e Coeficiente $r$ . . . . .	69
Tabela 28 – Tamanhos de amostra $n$ segundo valores $p$ do teste de Wilcoxon e Coeficiente $r$ . . . . .	71

Tabela 29 – Valores $p$ do teste $t$ não-pareado segundo o tamanho da amostra $n$ e $d$ de Cohen . . . . .	72
Tabela 30 – Tamanhos de amostra $n$ segundo valores $p$ do teste $t$ não-pareado e $d$ de Cohen . . . . .	73
Tabela 31 – Valores $p$ do teste $t$ não-pareado segundo o tamanho da amostra $n$ e $g$ de Hedges . . . . .	75
Tabela 32 – Tamanhos de amostra $n$ segundo valores $p$ do teste $t$ não-pareado e $g$ de Hedges . . . . .	76
Tabela 33 – Valores $p$ do teste $t$ pareado segundo o tamanho da amostra $n$ e $d$ de Cohen . . . . .	77
Tabela 34 – Tamanhos de amostra $n$ segundo valores $p$ do teste $t$ pareado e $d$ de Cohen . . . . .	79

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>19</b>
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>21</b>
2.1	Teste de McNemar	21
2.2	Teste Qui-Quadrado (Homogeneidade)	23
2.3	Teste Exato de Fisher	25
2.4	Teste de Wilcoxon	27
2.5	Teste de Mann - Whitney	29
2.6	Correlação $\tau$ de Kendall	33
2.6.1	Teste da correlação $\tau$ de Kendall	34
2.7	Teste $t$ Pareado	35
2.8	Teste $t$ não-pareado	36
<b>3</b>	<b>MEDIDAS DE MAGNITUDE DO EFEITO</b>	<b>39</b>
<b>3.1</b>	<b>Escala Nominal</b>	<b>39</b>
3.1.1	Coeficiente $\phi$	39
3.1.2	$W$ de Cohen	40
3.1.3	$V$ de Cramér	41
3.1.4	Razão de chances ( <i>Odds Ratio</i> )/Risco relativo	42
3.1.5	$Q$ de Yule	44
<b>3.2</b>	<b>Escala Ordinal</b>	<b>45</b>
3.2.1	Delta ( $\delta$ ) de Cliff	45
3.2.2	$A$ de Vargha-Delaney	47
3.2.3	Coeficiente $r$	49
3.2.4	$\tau$ de Kendall	50
<b>3.3</b>	<b>Escala Intervalar</b>	<b>51</b>
3.3.1	$d$ de Cohen	51
3.3.2	$g$ de Hedges	54
3.3.3	Delta ( $\Delta$ ) de Glass	55
<b>3.4</b>	<b>Influência do tamanho da amostra no Effect Size</b>	<b>57</b>
<b>4</b>	<b>SIMULAÇÕES</b>	<b>59</b>
<b>4.1</b>	<b>Escala Nominal</b>	<b>60</b>
4.1.1	Simulação da medida associada ao teste Qui-quadrado ( $\phi$ )	60
4.1.2	Simulação da medida associada ao teste de McNemar ( $\phi$ )	63
<b>4.2</b>	<b>Escala Ordinal</b>	<b>66</b>

4.2.1	Simulação da medida associada ao teste de Mann-Whitney ( $\delta$ ) . . . . .	66
4.2.2	Simulação da medida associada ao teste de Wilcoxon ( $r$ ) . . . . .	69
<b>4.3</b>	<b>Escala Intervalar</b> . . . . .	<b>72</b>
4.3.1	Simulação das medidas associadas ao teste $t$ não-pareado ( $d$ e $g$ ) . . . . .	72
4.3.1.1	$d$ de Cohen . . . . .	72
4.3.1.2	$g$ de Hedges . . . . .	75
4.3.2	Simulação da medida associada ao teste $t$ pareado ( $d$ ) . . . . .	77
<b>5</b>	<b>CONCLUSÃO</b> . . . . .	<b>81</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>85</b>
	<b>Anexos</b> . . . . .	<b>87</b>

# 1 Introdução

Em um estudo cujo o interesse é realizar a comparação de dois grupos específicos, denotados por  $X$  e  $Y$ , pode-se utilizar a metodologia estatística dos testes de hipóteses para a comparação de duas populações. Para utilizar esta metodologia, deve-se atentar aos pressupostos exigidos por cada teste, como por exemplo o nível de mensuração da variável estudada. A verificação desses pressupostos auxiliará na escolha do teste mais adequado para o caso em análise.

As pesquisas científicas costumam atribuir a significância estatística dos resultados obtidos em um teste de hipóteses através de seu nível descritivo, também conhecido como *valor-p* [Conboy, 2000], sendo que o *valor-p* é a probabilidade de ocorrer valores da estatística mais extremos do que o observado supondo ser verdadeira, a hipótese nula,  $H_0$  [Bussab & Morettin, 2013] e o *valor-p* é sempre uma função do número de observações na amostra  $n$ , seja qual for a importância prática da diferença ou relação observada.

Como comentado por [Conboy, 2003], uma grande amostra pode gerar um *valor-p* reduzido, exagerando a importância aparente das diferenças (diferenças grandes). Em contrapartida, diferenças com significado prático podem dar origem a valores  $p$  grandes, se a amostra for considerada pequena. Sendo assim, quanto maior for o tamanho da amostra, maior será o poder do teste estatístico (poder do teste rejeitar a hipótese de igualdade de grupos), trazendo evidências de que uma diferença estatisticamente significativa sempre poderá ser alcançada, mesmo quando a diferença observada não for relevante na prática.

Uma proposta, cada vez mais aceita por pesquisadores de diversas áreas, para resolver esse problema do *valor-p* é combinar os resultados dos testes de significância estatísticos com uma medida, denominada Magnitude do Efeito (*Effect Size* -  $ES$ ). O uso dessas medidas representa uma tentativa de substituir o conceito de significância estatística com noções mais úteis de significância prática específica [Kirk, 1996], ou seja, as estimativas das medidas de magnitude são úteis para determinar a importância prática ou teórica de um efeito e permitem que os pesquisadores se direcionem no estudo descritivo do  $ES$  entre os grupos. Estudos com diferentes tamanhos de amostra, mas com a mesma base descritiva, como por exemplo distribuições, médias, desvios-padrões e intervalos de confiança, diferem em seus valores de significância estatística, mas não em suas estimativas de  $ES$  e quanto maior for o  $ES$ , maior é o poder do teste e menor a probabilidade de um erro do tipo II [Cohen, 1992].

No presente trabalho, em um dos capítulos foi apresentada uma revisão bibliográfica dos principais testes estatísticos que realizam a comparação de dois grupos e em um outro capítulo foi apresentada uma revisão das principais medidas de magnitude do efeito. De

certa forma, existem diversos *ES* [Huberty, 2002], no entanto, neste contexto, o presente trabalho visou realizar um estudo das principais medidas utilizadas na comparação de dois grupos, atentando-se ao nível de mensuração da variável em estudo.

Os resultados deste projeto foram ilustrados por meio de dados simulados e tais simulações foram feitas para alguns dos *ES*. Portanto, foram geradas algumas amostras para assegurar (aproximadamente) o mesmo nível de magnitude do efeito, sendo uma amostra inicial de  $n_0 = 5$  (para as medidas com escala de mensuração ordinal e intervalar) e uma amostra inicial  $n_0 = 10$  (para as medidas com escala de mensuração nominal). Tais amostras foram repetidas diversas vezes e o intuito da simulação foi verificar o comportamento de cada *ES* à medida que aumentava-se o tamanho da amostra, apresentando os valores  $p$  do teste estatístico associado a cada *ES*. Os cálculos e análises foram realizados através do Software R.

## 2 Revisão bibliográfica

O objetivo deste capítulo é apresentar os principais testes estatísticos que realizam comparações de duas populações.

Para o nível de mensuração nominal, os principais testes utilizados são o teste de Qui-Quadrado (homogeneidade), teste de McNemar e o teste exato de Fisher. Considerando o nível de mensuração ordinal, utiliza-se os testes não-paramétricos de Mann-Whitney (teste específico para duas populações independentes), o de Wilcoxon (teste específico para duas populações dependentes) e o teste da correlação  $\tau$  de Kendall; e ainda para o nível de mensuração intervalar, utiliza-se os testes  $t$  pareado e o  $t$  não-pareado.

### 2.1 Teste de McNemar

O teste de McNemar consiste em detectar mudanças em planejamentos do tipo “antes e depois”, em que cada unidade amostral é seu próprio controle. O teste possui o pressuposto de que as amostras devem ser pareadas (relacionadas).

Considerando duas amostras  $X_1, \dots, X_n$  e  $Y_1, \dots, Y_n$ , do tipo “antes e depois”, com valores pareados, pode-se construir uma tabela de contingência  $2 \times 2$  para observar os desfechos.

Tabela 1 – Tabela de contingência  $2 \times 2$

		Depois	
		(+)	(-)
Antes	(+)	A	B
	(-)	C	D

Note que  $(B + C)$  representa o número total de indivíduos que acusaram alguma modificação.

As hipóteses de interesse são:

$$\begin{cases} H_0 : p_B = p_C \\ H_a : p_B \neq p_C \end{cases}$$

Onde,

$p_B$  indica a probabilidade de uma mudança do status “(+)” para o status “(-)”;

$p_C$  indica a probabilidade de uma mudança do status “(-)” para o status “(+)”.

Sob a hipótese nula,  $(B + C)/2$  representa a frequência esperada, tanto para a casela  $B$  como para a casela  $C$ . Portanto, o teste de McNemar consiste em fazer um teste Qui-Quadrado de aderência com os valores de  $B$  e  $C$ . Sendo assim, a estatística de teste é dada por

$$Q_{McNemar}^2 = \frac{(B - C)^2}{B + C} \approx \chi_{(1)}^2. \quad (2.1)$$

Quando  $(B + C) < 25$ , a aproximação pela distribuição Qui-Quadrado pode ficar comprometida. Para estes casos, recomenda-se utilizar a correção de continuidade de Yates. Sendo assim, a estatística do teste de McNemar fica da forma

$$Q_{McNemar-corrigido}^2 = \frac{(|B - C| - 1)^2}{B + C} \approx \chi_{(1)}^2. \quad (2.2)$$

O nível descritivo do teste (*valor-p*) é dado por

$$valor-p = P(\chi_{(1)}^2 > Q_{McNemar}^2), \quad (2.3)$$

e a hipótese nula será rejeitada se  $valor-p < \alpha$ , sendo que  $\alpha$  é o nível de significância do teste.

**Nota:** Se a frequência esperada,  $(B + C)/2$ , for menor do que 5, deve-se utilizar o teste Binomial ao invés do teste de McNemar.

### Código para realização do teste de McNemar no Software R :

`mcnemar.test(x)` (com correção de continuidade)

`mcnemar.test(x,correct=F)` (sem correção de continuidade)



## 2.2 Teste Qui-Quadrado (Homogeneidade)

O teste Qui-Quadrado é utilizado para dados em tabelas de contingência. Sendo assim, o tamanho das amostras devem ser suficientemente grandes para que se possa realizar a aproximação para a distribuição Qui-Quadrado.

O teste de homogeneidade baseado na estatística de Qui-Quadrado, verifica se determinados grupos possuem características semelhantes, segundo uma variável de interesse, tais grupos são representados por amostras extraídas de populações específicas.

Segundo [Conover, 1999], supondo que existam duas amostras aleatórias extraídas de duas populações específicas, pressupõe-se que tais amostras devem ser mutuamente independentes e considerando ainda que existam duas categorias, as observações de cada amostra podem ser classificadas em cada uma das categorias. Na tabela de contingência apresentada na Tabela 2, observa-se como pode ser feita a distribuição de frequência para ambas as amostras em cada uma das categorias.

Tabela 2 – Tabela de contingência  $2 \times 2$

	<b>Categoria 1</b>	<b>Categoria 2</b>	<b>Total</b>
<b>População 1</b>	$O_{11}$	$O_{12}$	$n_1$
<b>População 2</b>	$O_{21}$	$O_{22}$	$n_2$
<b>Total</b>	$C_1$	$C_2$	$N$

$O_{ij}$ : o número de observações que se enquadra em uma das duas categorias para cada amostra,  $i, j = 1, 2$ ;

$n_i$ : o número de observações na  $i$ -ésima linha;

$C_j$ : o número de observações na  $j$ -ésima coluna;

$N$ : representa o número total de observações, isto é,  $N = n_1 + n_2$ .

A homogeneidade das duas populações é testada por meio das seguintes hipóteses:

$$\begin{cases} H_0 : p_1 = p_2 \\ H_a : p_1 \neq p_2 \end{cases}$$

Em que  $p_i$  é a probabilidade de um indivíduo da população  $i$  pertencer à Categoria 1. Consequentemente,  $(1 - p_i)$  é a probabilidade de um indivíduo da população  $i$  pertencer à Categoria 2,  $i = 1, 2$ .

Para grandes amostras, a distribuição assintótica de  $Q^2$  é dada pela aproximação da distribuição Qui-Quadrado, com 1 grau de liberdade. Sendo assim a estatística de teste

é calculada como

$$Q^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \approx \chi_{(1)}^2, \quad (2.4)$$

em que a frequência esperada  $E_{ij}$  é denotada da forma

$$E_{ij} = \frac{(\text{total da coluna } j) \times (\text{total da linha } i)}{\text{total geral}} = \frac{n_i C_j}{N}, i, j = 1, 2. \quad (2.5)$$

Em casos de tabelas de contingência no formato  $2 \times 2$  com frequência esperada menor que 5 ( $E_{ij} < 5$ ), a aproximação para a distribuição Qui-Quadrado é considerada ruim. Sendo assim para estes casos, pode-se utilizar a correção de continuidade de Yates, dada por

$$Q_{Yates}^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(|O_{ij} - E_{ij}| - 0,5)^2}{E_{ij}} \approx \chi_{(1)}^2. \quad (2.6)$$

O nível descritivo do teste (*valor-p*) é dado por

$$\text{valor-p} = P(\chi_{(1)}^2 > Q^2), \quad (2.7)$$

e a hipótese nula será rejeitada se  $\text{valor-p} < \alpha$ , sendo que  $\alpha$  é o nível de significância do teste.

A razão de chances (*Odds Ratio*) é considerada como uma medida de associação para tabelas  $2 \times 2$ , representada pela Tabela 2. Sendo assim, pode-se representar  $p_1$  como a probabilidade de sucesso da Categoria 1, na linha 1 (Grupo 1) e  $p_2$  como a probabilidade de sucesso da Categoria 1, na linha 2 (Grupo 2).

A razão de chances também é considerada como a razão das *odds* de dois grupos (linhas), sendo definida da seguinte forma:

$$\theta = \frac{\text{odds}_1}{\text{odds}_2} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}. \quad (2.8)$$

Sendo assim, se ( $p_1 = p_2$ ), então  $\theta = 1$ , caracterizando que os grupos são homogêneos. Quando ( $1 < \theta < +\infty$ ), a *odds* de sucesso na Categoria 1 é maior na linha 1 do que na linha 2, e portanto  $p_1 > p_2$ . Da mesma forma a *odds* de sucesso na Categoria 1

será menor na linha 1 do que na linha 2, quando ( $0 < \theta < 1$ ), e portanto  $p_1 < p_2$ . Na interpretação da razão de chances, dois valores para  $\theta$  representarão a mesma associação, porém em direções opostas, quando um valor é o inverso do outro.

### Código para realização do teste Qui-Quadrado no Software R :

`chisq.test(x,correct=T)` (com correção de continuidade)

`chisq.test(x,correct=F)` (sem correção de continuidade)

## 2.3 Teste Exato de Fisher

O Teste exato de Fisher é utilizado em situações em que existem tabelas de contingência no formato  $2 \times 2$ , com amostras de tamanho pequeno e valor esperado  $E_{ij} < 5$ . Portanto, o teste consiste em determinar a probabilidade exata de ocorrência de uma frequência observada, ou de valores mais extremos.

Na tabela  $2 \times 2$ , cada observação é classificada em uma determinada célula. Para a realização do teste, supõe-se que as amostras devem ser aleatórias e independentes [Conover, 1999].

Tabela 3 – Tabela de contingência  $2 \times 2$

	<b>Categoria 1</b>	<b>Categoria 2</b>	<b>Total</b>
<b>População 1</b>	$x$	$r - x$	$r$
<b>População 2</b>	$c - x$	$N - c - r + x$	$N - r$
<b>Total</b>	$c$	$N - c$	$N$

$c$ : número total de observações na Categoria 1;

$r$ : número total de observações na População 1;

$x$ : Frequência observada na População 1 e Categoria 1;

$N$ : Número total de observações.

No teste exato de Fisher, os totais das linhas ( $r$  e  $N - r$ ), e totais das colunas ( $c$  e  $N - c$ ) são considerados fixos, sendo não aleatórios.

Considerando  $p_1$  a probabilidade de uma observação da População 1 ser classificada na Categoria 1 e  $p_2$ , a probabilidade de uma observação da População 2 ser classificada na Categoria 1, tem-se que as hipóteses de interesse serão:

$$\begin{cases} H_0 : p_1 = p_2 \\ H_a : p_1 \neq p_2 \end{cases}$$

A estatística de teste denotada por  $T$ , representa o número de observações na célula da primeira linha e primeira coluna. Portanto, a distribuição exata de  $T$ , considerando  $H_0$  verdadeira, é dada pela distribuição hipergeométrica, isto é

$$P(T = x) = \frac{\binom{r}{x} \binom{N-r}{c-x}}{\binom{N}{c}}; x = 0, 1, \dots, \min(r, c) \quad (2.9)$$

Para amostras grandes, pode-se utilizar a aproximação para a distribuição normal padrão, dada pela seguinte expressão [Conover, 1999]:

$$T_{obs} = \frac{x - \frac{rc}{N}}{\sqrt{\frac{rc(N-r)(N-c)}{N^2(N-1)}}} \approx N(0, 1) \quad (2.10)$$

Para a decisão do teste, calcula-se o *valor-p* utilizando a Equação (2.9), sendo as probabilidades referentes às frequências observadas e das demais situações extremas. Para o caso de teste bilateral, o nível descritivo (*valor-p*) do teste é definido como

$$valor-p = 2 \times \min\{P(T \leq t_{obs}), P(T \geq t_{obs})\} \quad (2.11)$$

Assim, considerando um nível significância  $\alpha$ , a hipótese  $H_0$  é rejeitada se *valor-p*  $\leq \alpha$ .

### Código para realização do teste Exato de Fisher no Software R :

```
fisher.test(x)
```

## 2.4 Teste de Wilcoxon

Com o pressuposto de que os dados devem ter no mínimo escala ordinal, o teste de Wilcoxon pode ser utilizado para a comparação de dois grupos ( $X$  e  $Y$ ) pareados, para os quais pode haver medidas da mesma unidade amostral em dois tempos ou em duas situações diferentes. Sendo assim, o teste se baseia em verificar se há diferença entre os efeitos de dois tipos de tratamentos denotados pelos grupos  $X$  e  $Y$  [Conover, 1999].

Ao observar duas amostras  $X_1, \dots, X_n$  e  $Y_1, \dots, Y_n$  de valores pareados, calcula-se a diferença dos valores entre os grupos para as  $n$  observações, e em seguida atribui-se o posto do valor absoluto para cada diferença. Se existir diferença sistemática entre as duas condições, o posto mais alto pertencerá à uma condição e o mais baixo dos postos pertencerá à outra condição [Borba, 2015]. Se ocorrer duas ou mais diferenças com o mesmo valor resultante, atribui-se o mesmo posto para todos os empates e se a diferença entre as observações dos dois grupos for nula, o posto referente a essa diferença é excluído da análise.

Para testar se o efeito dos dois grupos provenientes de duas populações  $X$  e  $Y$  são iguais ou diferentes, utiliza-se as hipóteses

$$\begin{cases} H_0 : \text{Os grupos } X \text{ e } Y \text{ são iguais} \\ H_a : \text{Os grupos } X \text{ e } Y \text{ são diferentes} \end{cases}$$

Como a estatística do teste de Wilcoxon se baseia na utilização dos postos, ela será denotada para cada caso, da seguinte forma:

1º Caso – Quando não ocorrer empates entre os postos, a estatística de teste é o menor valor entre  $S_p$  e  $S_n$ , sendo denotada da seguinte forma:

$$W = \min\{S_p, S_n\} \quad (2.12)$$

Aqui,  $S_p$  é a soma dos postos das diferenças positivas e  $S_n$  é a soma dos postos das diferenças negativas. Quanto menor o valor de  $W_{Wilcoxon}$ , menor é a probabilidade dos postos das diferenças terem ocorrido ao acaso. Assim, a hipótese de não existir diferenças ( $H_0$ ) é rejeitada se  $W < W_c$ , onde  $W_c$  é o valor crítico encontrado na tabela do teste de Wilcoxon [Borba, 2015], apresentada na Tabela 4.

Tabela 4 – Valores Críticos ( $W_c$ ) do teste de Wilcoxon

	$w_{0.005}$	$w_{0.01}$	$w_{0.025}$	$w_{0.05}$	$w_{0.1}$	$w_{0.2}$	$w_{0.3}$	$w_{0.4}$	$w_{0.5}$	$\frac{n(n+1)}{2}$
<b>n=4</b>	0	0	0	0	1	3	3	4	5	10
<b>5</b>	0	0	0	1	3	4	5	6	7.5	15
<b>6</b>	0	0	1	3	4	6	8	9	10.5	21
<b>7</b>	0	1	3	4	6	9	11	12	14	28
<b>8</b>	1	2	4	6	9	12	14	16	18	36
<b>9</b>	2	4	6	9	11	15	18	20	22.5	45
<b>10</b>	4	6	9	11	15	19	22	25	27.5	55
<b>11</b>	6	8	11	14	18	23	27	30	33	66
<b>12</b>	8	10	14	18	22	28	32	36	39	78
<b>13</b>	10	13	18	22	27	33	38	42	45.5	91
<b>14</b>	13	16	22	26	32	39	44	48	52.5	105
<b>15</b>	16	20	26	31	37	45	51	55	60	120
<b>16</b>	20	24	30	36	43	51	58	63	68	136
<b>17</b>	24	28	35	42	49	58	65	71	76.5	153
<b>18</b>	28	33	41	48	56	66	73	80	85.5	171
<b>19</b>	33	38	47	54	63	74	82	89	95	190
<b>20</b>	38	44	53	61	70	83	91	98	105	210
<b>21</b>	44	50	59	68	78	91	100	108	115.5	231
<b>22</b>	49	56	67	76	87	100	110	119	126.5	253
<b>23</b>	55	63	74	84	95	110	120	130	138	276
<b>24</b>	62	70	82	92	105	120	131	141	150	300
<b>25</b>	69	77	90	101	114	131	143	153	162.5	325

Fonte: Conover (1999)

2º Caso - Se ocorrer empates e se a amostra for suficiente grande, deve-se utilizar a aproximação para a distribuição normal. Sendo assim, a estatística de teste é dada da forma

$$Z_{Wilcoxon} = \frac{\min(S_p, S_n) - \frac{m(m+1)}{4}}{\sqrt{\frac{m(m+1)(2m+1)}{24} - \frac{\sum_{j=1}^l (t_j^3 - t_j)}{48}}} \approx N(0, 1), \quad (2.13)$$

em que

$S_p$  : Soma dos postos das diferenças positivas

$S_n$  : Soma dos postos das diferenças negativas

$m$  : Número de diferenças diferentes de zero

$l$  : Número de empates (empates dos postos)

$t_j$  : Número de elementos no  $j$ -ésimo empate (dos postos).

O nível descritivo do teste (*valor-p*) é dado por

$$\text{valor-p} = 2 \times P(Z > |Z_{Wilcoxon}|) \quad (2.14)$$

Considerando um nível de significância  $\alpha$ , a hipótese nula é rejeitada se  $\text{valor-p} < \alpha$ .

**Código para realização do teste de Wilcoxon no Software R :**

`wilcox.test(x,y,paired=T)` (com correção de continuidade)

`wilcox.test(x,y,paired=T,correct=F)`

## 2.5 Teste de Mann - Whitney

O teste de Mann - Whitney se baseia na comparação de dois grupos ( $X$  e  $Y$ ) independentes, com a suposição de que os dados devem possuir no mínimo escala ordinal. De forma geral, o objetivo do teste é verificar se duas amostras independentes são provenientes de uma mesma população. Para a utilização do teste, observa-se uma amostra aleatória  $(x_1, \dots, x_n)$ , procedente de uma população e outra amostra aleatória  $(y_1, \dots, y_m)$ , procedente de uma outra população [Conover, 1999].

Com as observações dos dois grupos, verifica-se os postos dos valores sem discriminar os grupos, ou seja, ordena-se as observações e atribui-se os postos para ambos os grupos. Se a distribuição dos postos são bem próximas em ambas as amostras, existe a possibilidade de que as amostras são provenientes de uma mesma população.

Para verificar se ambas as amostras são procedentes da mesma população, as hipóteses de interesse serão:

$$\begin{cases} H_0 : \text{Os grupos } X \text{ e } Y \text{ são iguais} \\ H_a : \text{Os grupos } X \text{ e } Y \text{ são diferentes} \end{cases}$$

Para a população  $Y$ , a estatística  $U$  será dada da forma

$$U_y = R_y - \frac{m(m+1)}{2}, \quad (2.15)$$

onde

$R_y$  representa a soma dos postos do grupo  $Y$ ,

$m$  representa o número de observações no grupo  $Y$ .

Para o grupo  $X$  deve ser feito o mesmo cálculo sendo representado por  $U_x$ . Quando houver poucos empates ou nenhum empate, a estatística de teste será o mínimo entre  $U_x$  e  $U_y$ , sendo denotada da forma

$$U_{obs} = \min(U_x, U_y). \quad (2.16)$$

Portanto, ao se obter o valor crítico  $d$  na tabela do teste de Mann-Whitney (Tabela 5), rejeita-se  $H_0$  a um nível de significância específico se  $U_{obs} < d$ .

Para o caso de grandes amostras, utiliza-se uma aproximação para a distribuição normal padrão e portanto a estatística de teste sem correção de continuidade é dada por

$$U = \frac{|U_x - \frac{nm}{2}|}{\sqrt{\frac{nm}{(n+m)(n+m-1)} \left( \frac{(n+m)^3 - n - m}{12} - \sum_{i=1}^l \frac{t_i^3 - t_i}{12} \right)}} \approx N(0, 1), \quad (2.17)$$

e a estatística com correção de continuidade é dada por

$$U = \frac{|U_x - \frac{nm}{2}| - 0.5}{\sqrt{\frac{nm}{(n+m)(n+m-1)} \left( \frac{(n+m)^3 - n - m}{12} - \sum_{i=1}^l \frac{t_i^3 - t_i}{12} \right)}} \approx N(0, 1), \quad (2.18)$$

em que

$n$  : número de observações na amostra aleatória referente à população X.

$m$  : número de observações na amostra aleatória referente à população Y.

$l$  é o número de empates dos postos.

$t_j$  : é o número de elementos no  $j$ -ésimo empate (do posto).

O nível descritivo do teste (*valor-p*) é dado por

$$\text{valor-p} = 2 \times P(Z > |U|) \quad (2.19)$$

Considerando um nível de significância  $\alpha$ , a hipótese nula é rejeitada se *valor-p*  $< \alpha$ .

**Código para realização do teste de Mann-Whitney no Software R :**



```
wilcox.test(x,y, correct=FALSE)
```

Tabela 5 – Valores Críticos ( $d$ ) do teste de Mann-Whitney

$n$	$p$	$m=2$	3	4	5	6	7	8	9	10
2	0,001	3	3	3	3	3	3	3	3	3
	0,005	3	3	3	3	3	3	3	3	3
	0,01	3	3	3	3	3	3	3	3	3
	0,025	3	3	3	3	3	3	4	4	4
	0,05	3	3	3	4	4	4	5	5	5
	0,1	3	4	4	5	5	5	6	6	7
3	0,001	6	6	6	6	6	6	6	6	6
	0,005	6	6	6	6	6	6	6	7	7
	0,01	6	6	6	6	6	7	7	8	8
	0,025	6	6	6	7	8	8	9	9	10
	0,05	6	7	7	8	9	9	10	11	11
	0,1	7	8	8	9	10	11	12	12	13
4	0,001	10	10	10	10	10	10	10	10	11
	0,005	10	10	10	10	11	11	12	12	13
	0,01	10	10	10	11	12	12	13	14	14
	0,025	10	10	11	12	13	14	15	15	16
	0,05	10	11	12	13	14	15	16	17	18
	0,1	11	12	14	15	16	17	18	20	21
5	0,001	15	11	12	13	14	15	16	17	17
	0,005	15	15	15	16	17	17	18	19	20
	0,01	15	15	16	17	18	19	20	21	22
	0,025	15	16	17	18	19	21	22	23	24
	0,05	16	17	18	20	21	22	24	25	27
	0,1	17	18	20	21	23	24	26	28	29
6	0,001	21	21	21	21	21	21	23	24	25
	0,005	21	21	22	23	24	25	26	27	28
	0,01	21	21	23	24	25	26	28	29	30
	0,025	21	23	24	25	27	28	30	32	33
	0,05	22	24	25	27	29	30	32	34	36
	0,1	23	25	27	29	31	33	35	37	39
7	0,001	28	28	28	28	29	30	31	32	34
	0,005	28	28	29	30	32	33	35	36	38
	0,01	28	29	30	32	33	35	36	38	40
	0,025	28	30	32	34	35	37	39	41	43
	0,05	29	31	33	35	37	40	42	44	46
	0,1	30	33	35	37	40	42	45	47	50
8	0,001	36	36	36	37	38	39	41	42	43
	0,005	36	36	38	39	41	43	44	46	48
	0,01	36	37	39	41	43	44	46	48	50
	0,025	37	39	41	43	45	47	50	52	54
	0,05	38	40	42	45	47	50	52	55	57
	0,1	39	42	44	47	50	53	56	59	61
9	0,001	45	45	45	47	48	49	51	53	54
	0,005	45	46	47	49	51	53	55	57	59
	0,01	45	47	49	51	53	55	57	60	62
	0,025	46	48	50	53	56	58	61	63	66
	0,05	47	50	52	55	58	61	64	67	70
	0,1	48	51	55	58	61	64	68	71	74
10	0,001	55	55	56	57	59	61	62	64	66
	0,005	55	56	58	60	62	65	67	69	72
	0,01	55	57	59	62	64	67	69	72	75
	0,025	56	59	61	64	67	70	73	76	79
	0,05	57	60	63	67	70	73	76	80	83
	0,1	59	62	66	69	73	77	80	84	88

Fonte: Conover (1999)

## 2.6 Correlação $\tau$ de Kendall

A correlação  $\tau$  de Kendall é utilizada para dados ordinais. Suponha que existam dois grupos de variáveis aleatórias conjuntas  $X$  e  $Y$ , denotados por  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ . Para o cálculo da correlação, deve-se obter os postos de cada variável. Após obter os postos de cada uma das variáveis, ordena-se os postos segundo uma das variáveis ( $X$  ou  $Y$ ). Com a ordenação dos postos da variável específica, define-se quantos pares de postos da outra variável também estão em sua ordenação natural.

A partir destes resultados, verifica-se número de pares com a ordenação correta e os pares com a ordenação errada.

Quando não houver empates, a correlação de Kendall é definida por

$$\tau_{Kendall} = \frac{N_C - N_E}{\frac{1}{2}(n(n-1))}, \quad (2.20)$$

onde

$N_C$  representa o número de pares com a ordenação correta;

$N_E$  representa o número de pares com a ordenação errada;

$n$  tamanho da amostra.

Note que:

- se todos os pares estiverem na ordenação correta, tem-se que  $\tau_{Kendall} = 1$ ;
- se todos os pares estiverem na ordenação errada, tem-se que  $\tau_{Kendall} = -1$ ;
- se metade dos valores estiverem na ordenação correta e metade errada, então não há nenhuma associação entre  $X$  e  $Y$  e  $\tau_{Kendall} = 0$ ;

Quando houver empates, a correlação será denotada da forma

$$\tau_{Kendall} = \frac{N_C - N_E}{\frac{1}{2}\sqrt{n(n-1) - \sum_{i=1}^I t_{x;i}(t_{x;i} - 1)}\sqrt{n(n-1) - \sum_{j=1}^J t_{y;j}(t_{y;j} - 1)}}, \quad (2.21)$$

em que

$t_{x;i}$  é o número de elementos do  $i$ -ésimo empate da variável  $X$ ,  $i = 1, 2, \dots, I$ ;

$t_{y;j}$  é o número de elementos do  $j$ -ésimo empate da variável  $Y$ ,  $j = 1, 2, \dots, J$ .

### 2.6.1 Teste da correlação $\tau$ de Kendall

As hipóteses de interesse são:

$$\begin{cases} H_0 : \text{Não existe associação linear entre } X \text{ e } Y \\ H_a : \text{Existe associação linear entre } X \text{ e } Y \end{cases}$$

A estatística de teste é denotada da forma

$$z_{Kendall} = \frac{N_C - N_E}{\sqrt{\frac{v_0 - v_x - v_y}{18} + \frac{v_1}{2n(n-1)} + \frac{v_2}{9n(n-1)(n-2)}}} \approx N(0, 1), \quad (2.22)$$

onde

$$v_0 = n(n-1)(2n+5)$$

$$v_x = \sum_{i=1}^I t_{x;i}(t_{x;i}-1)(2t_{x;i}+5)$$

$$v_y = \sum_{j=1}^J t_{y;j}(t_{y;j}-1)(2t_{y;j}+5)$$

$$v_1 = (\sum_{i=1}^I t_{x;i}(t_{x;i}-1))(\sum_{j=1}^J t_{y;j}(t_{y;j}-1))$$

$$v_2 = (\sum_{i=1}^I t_{x;i}(t_{x;i}-1)(t_{x;i}-2))(\sum_{j=1}^J t_{y;j}(t_{y;j}-1)(t_{y;j}-2))$$

**Nota:** Por sugestão de Kendall, se o tamanho da amostra  $n \geq 8$  já existe uma boa aproximação para a distribuição normal.

Para o critério de decisão, pode-se utilizar o *valor-p*, que é calculado da forma

$$\text{valor-p} = 2 \times P[Z > |z_{Kendall}|]. \quad (2.23)$$

Considerando um nível de significância  $\alpha$ , rejeita-se a hipótese nula  $H_0$  se *valor-p* <  $\alpha$ .

**Código para realização do teste da correlação de Kendall no Software R :**

`cor.test(x,y,method="kendall")`

## 2.7 Teste $t$ Pareado

O teste  $t$  pareado consiste em verificar se as médias de duas populações (População 1 e 2) dependentes são iguais.

Suponha que existam duas amostras  $X_1, \dots, X_n$  e  $Y_1, \dots, Y_n$  em que as observações são dependentes (pareadas). Ao definir a variável aleatória  $D = X - Y$ , obtém-se uma amostra  $D_1, \dots, D_n$  que é o resultado das diferenças entre os valores de cada par [Bussab & Morettin, 2013]. Considerando que a variável aleatória  $D$  tem distribuição normal  $N(\mu_D, \sigma_D^2)$ ,  $\sigma_D^2$  desconhecido, pode-se inferir que :

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n (X_i - Y_i) = \bar{X} - \bar{Y} \sim N\left(\mu_D, \frac{\sigma_D^2}{n}\right) \quad (2.24)$$

Como  $\mu_D = E(D) = E(X - Y) = E(X) - E(Y) = \mu_1 - \mu_2$ , qualquer informação sobre  $\mu_1 - \mu_2$  corresponde a  $\mu_D$ . Portanto, para testar se as médias dos grupos são iguais, as hipóteses de interesse serão:

$$\begin{cases} H_0 : \mu_D = 0 \\ H_a : \mu_D \neq 0 \end{cases}$$

Considerando que

$$S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2, \quad (2.25)$$

a estatística de teste é dada por

$$T_{obs} = \frac{\sqrt{n}(\bar{D} - \mu_D)}{S_D}, \quad (2.26)$$

e terá distribuição  $t$  de Student, com  $(n - 1)$  graus de liberdade.

Para o critério de decisão, pode-se utilizar o *valor-p*, neste caso de teste bilateral, o *valor-p* é calculado da forma

$$valor-p = 2 \times P(t_{(n-1)} > |T_{obs}|), \quad (2.27)$$

tendo por base na distribuição  $t$  de Student, com  $(n - 1)$  graus de liberdade. Considerando um nível de significância  $\alpha$ , rejeita-se  $H_0$ , se  $valor-p < \alpha$ .

**Código para realização do teste  $t$  pareado no Software R :**

t.test(x,y,paired=TRUE)

## 2.8 Teste $t$ não-pareado

O teste  $t$  não-pareado consiste na comparação de duas médias de duas populações específicas, averiguando se ambas as médias são iguais ou diferentes.

Suponha que existam duas amostras independentes  $x_1, \dots, x_{n_1}$  e  $y_1, \dots, y_{n_2}$ , extraídas de duas populações  $X \sim N(\mu_1, \sigma_1^2)$  e  $Y \sim N(\mu_2, \sigma_2^2)$ , respectivamente. Para testar se existe diferença entre as médias das duas populações, não importando a direção, as hipóteses de interesse serão da forma [Bussab & Morettin, 2013]

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_a : \mu_1 \neq \mu_2 \end{cases}$$

Os estimadores da média e da variância para cada amostra são calculados da forma

$$\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i \quad (2.28)$$

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2 \quad (2.29)$$

$$\bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i \quad (2.30)$$

$$S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 \quad (2.31)$$

Este teste se adapta a dois casos específicos. Sendo assim, para cada um dos casos existe uma estatística de teste para que se possa verificar se existe diferença entre as médias das populações.

Antes, deve-se realizar um teste de igualdade de variâncias para verificar se as variâncias das duas populações são iguais ou não; Para realizar este teste supõe-se independência entre as duas populações e as hipóteses de interesse serão:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_a : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

Visto que  $S_1^2$  e  $S_2^2$  são as variâncias amostrais das respectivas populações, a estatística para o teste de igualdade de variâncias é dada por

$$W_{obs} = \frac{S_1^2}{S_2^2}, \quad (2.32)$$

que segue uma distribuição de Fisher-Snedecor com  $(n_1 - 1)$  e  $(n_2 - 1)$  graus de liberdade.

O nível descritivo (*valor-p*) do teste de igualdade de variâncias é dado por

$$valor-p = 2 \times \min\{P(F_{(n_1-1, n_2-1)} > W_{obs}), P(F_{(n_1-1, n_2-1)} < W_{obs})\}. \quad (2.33)$$

Considerando um nível de significância  $\alpha$ , a hipótese  $H_0$  (de igualdade de variâncias) é rejeitada se  $valor-p < \alpha$ .

1º Caso - Variâncias populacionais iguais e desconhecidas:

Suponha que o teste de igualdade de variâncias não rejeitou a hipótese de igualdade de variâncias. A estatística de teste das hipóteses

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_a : \mu_1 \neq \mu_2 \end{cases}$$

é dada por:

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{1/n_1 + 1/n_2}} \sim t_{(n_1+n_2-2)}, \quad (2.34)$$

em que

$$S_p^2 = \frac{\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2}{n_1 + n_2 - 2}. \quad (2.35)$$

O nível descritivo do teste (*valor-p*) é dado por

$$valor-p = 2 \times P[T_{(n_1+n_2-2)} > |T_{obs}|]. \quad (2.36)$$

Considerando um nível de significância  $\alpha$ , a hipótese nula é rejeitada se  $valor-p < \alpha$ .

**Código para realização do teste  $t$  não-pareado considerando variâncias iguais no Software R :**

```
t.test(x,y,var.equal=TRUE)
```

2º Caso - Variâncias populacionais desiguais e desconhecidas:

Neste caso, se ao realizar o teste de igualdade de variâncias, e a hipótese nula for rejeitada, concluí-se que as variâncias populacionais são diferentes, sendo assim, A estatística de teste das hipóteses

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_a : \mu_1 \neq \mu_2 \end{cases}$$

é dada por

$$T_{obs} = \frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim t_v, \quad (2.37)$$

com

$$v = \frac{(A + B)^2}{A^2/(n_1 - 1) + B^2/(n_2 - 1)}, \quad (2.38)$$

na qual

$$A = s_1^2/n_1 \quad (2.39)$$

e

$$B = s_2^2/n_2. \quad (2.40)$$

O nível descritivo do teste (*valor-p*) é dado por

$$valor-p = 2 \times P(t_{(v)} > |T_{obs}|). \quad (2.41)$$

Considerando um nível de significância  $\alpha$ , a hipótese nula é rejeitada se  $valor-p < \alpha$ .

**Código para realização do teste  $t$  não-pareado considerando variâncias desiguais no Software R :**

```
t.test(x,y,var.equal=FALSE)
```



## 3 Medidas de magnitude do efeito

---

Embora os testes de hipóteses estatísticos possuam grande importância para o estudo da comparação de dois grupos, eles costumam fornecer informações limitadas, tais como: rejeita-se a hipótese nula ou falha-se em rejeitá-la [Conboy, 2003].

O *Effect Size* - *ES* pode ser interpretado como o tamanho, dimensão ou magnitude do efeito e pode ser definido como o grau em que o fenômeno está presente na população [Cohen, 1988]. Quanto maior for o *ES*, maior será a manifestação do fenômeno na população [Loureiro & Gameiro, 2011]. Sendo assim, o *ES* pode informar sobre a direção das diferenças entre os grupos e o grau em que diferem [Cardoso, Mello & Freitas, 2013].

Um resumo das medidas de magnitude do efeito é apresentado no Apêndice A deste trabalho.

### 3.1 Escala Nominal

Esta seção apresenta uma revisão das principais medidas de magnitude do efeito para variáveis categóricas nominais.

#### 3.1.1 Coeficiente $\phi$

O coeficiente Phi ( $\phi$ ) é uma medida indicada para tabelas de contingência  $2 \times 2$ , sendo um caso específico da medida mais generalizada  $V$  de Cramér, podendo ser associada ao teste Qui-Quadrado quando os dois grupos são não-pareados ou ao teste de McNemar quando os dois grupos são pareados.

No caso de dois grupos não-pareados, após obter o valor do estatística de Qui-Quadrado  $Q^2$  (Expressão 2.4), o valor do coeficiente é denotado da forma [Cohen, 1988]

$$\phi = \sqrt{\frac{Q^2}{n}}, \quad (3.1)$$

onde

$n$  representa o número de observações na amostra;

$Q^2$  é a estatística do teste Qui-Quadrado apresentado pela Expressão (2.4).

Quando os dois grupos são pareados, o coeficiente  $\phi$  é baseado na estatística do teste de McNemar (Expressão 2.1) e é definido por

$$\phi = \sqrt{\frac{Q_{McNemar}^2}{n}} = \sqrt{\frac{(B-C)^2}{B+C}}. \quad (3.2)$$

O coeficiente  $\phi$  pode variar de 0 a 1. Quanto maior é o valor de  $\phi$ , maior é a diferença entre os dois grupos. A Tabela 6 traz os valores para os níveis de magnitude do efeito ( $ES$ ) sugeridos por [Cohen, 1992] para o coeficiente  $\phi$ .

Tabela 6 – Níveis de Magnitude do Efeito ( $ES$ ) do Coeficiente  $\phi$

<i>Effect Size</i>	$\phi$
<b>Fraco</b>	0,10
<b>Médio</b>	0,30
<b>Forte</b>	0,50

### Código para cálculo do coeficiente $\phi$ no Software R :

```
data = matrix(c(25, 5, 15, 15), ncol=2, byrow=T)
library(psych)
phi(data, digits = 4)
library(DescTools)
Phi(data)
```

### 3.1.2 $W$ de Cohen

Cohen introduziu uma estatística denominada  $W$ , sendo que essa estatística fornece um resultado do  $ES$  e é associada ao teste Qui-Quadrado. A sua fórmula é semelhante à do Qui-Quadrado, no entanto utiliza proporções ao invés de frequências [Cohen, 1988].

Sua expressão é denotada da forma

$$w = \sqrt{\sum_{i=1}^m \frac{(P_{1i} - P_{0i})^2}{P_{0i}}}, \quad (3.3)$$

em que

$m$  representa o número de caselas na tabela de contingência.

$P_{0i}$  representa a proporção na  $i$ -ésima casela postulada pela hipótese nula, ou seja, as proporções esperadas que são especificadas pela hipótese nula.

$P_{1i}$  representa a proporção na  $i$ -ésima casela postulada pela hipótese alternativa, ou seja, as proporções observadas nos dados.

A Tabela 7 a seguir traz os valores para os níveis de magnitude de efeito ( $ES$ ) sugeridos na literatura [Cohen, 1988].

Tabela 7 – Níveis de Magnitude do Efeito ( $ES$ ) do  $W$  de Cohen

<i>Effect Size</i>	<i>w de Cohen</i>
<b>Fraco</b>	0,10
<b>Médio</b>	0,30
<b>Forte</b>	0,50

### Código para cálculo do $W$ de Cohen no Software R :

```
library (rcompanion)
data = matrix(c(25, 5, 15, 15), ncol=2, byrow=T)
cohenW(data)
```

### 3.1.3 $V$ de Cramér

O  $V$  de Cramér é uma medida associada ao teste Qui-Quadrado. Se o número de colunas e linhas na tabela de contingência for igual a dois, então  $V = \phi$ . A indicação para o uso de  $V$  como um  $ES$  ocorre quando o número de linhas ou colunas na tabela de contingência for maior que dois.

A sua expressão é dada da forma [Cohen, 1988]

$$V = \sqrt{\frac{Q^2}{n(k-1)}}, \quad (3.4)$$

onde

$n$  representa o número de observações na amostra;

$k$  representa o menor valor entre o número de linhas e colunas;

$Q^2$  é a estatística do teste Qui-Quadrado em (2.4).

A Tabela 8 traz os valores para os níveis de magnitude do efeito (*ES*) sugeridos por [Cohen, 1988].

Tabela 8 – Níveis de Magnitude do Efeito (*ES*) do *V* de Cramér

<i>Effect Size</i>	<b>Cramér's <i>V</i></b> <b>k=2*</b>	<b>Cramér's <i>V</i></b> <b>k=3*</b>	<b>Cramér's <i>V</i></b> <b>k=4*</b>
<b>Fraco</b>	0,10	0,07	0,06
<b>Médio</b>	0,30	0,21	0,17
<b>Forte</b>	0,50	0,35	0,29

\*k é o número mínimo de categorias em linhas ou colunas

### Código para cálculo do *V* de Cramer no Software R :

```
data = matrix(c(25, 5, 15, 15), ncol=2, byrow=T)
library(rcompanion)
cramerV(data)
```

#### 3.1.4 Razão de chances (*Odds Ratio*)/Risco relativo

A razão de chances (*Odds Ratio*) fornece um resultado mais concreto do *ES* do que um mero valor de Qui-Quadrado ou do *valor-p* que lhe está associado [Conboy, 2003]. Sendo assim, a *Odds Ratio* é projetada para contrastar dois grupos em uma variável dependente binária (dicotômica), sendo a razão entre a chance de ocorrência do evento em um grupo e a chance de ocorrência do evento em outro grupo, podendo ser calculada através dos resultados obtidos em uma tabela de contingência  $2 \times 2$ , como na Tabela 9.

Tabela 9 – Tabela de contingência  $2 \times 2$

	Sim	Não
Grupo A	$a_{11}$	$a_{12}$
Grupo B	$a_{21}$	$a_{22}$
<b>Total</b>	$a_{11} + a_{21}$	$a_{12} + a_{22}$

Considerando os totais das colunas fixos, o cálculo da razão de chances é feito da forma

$$OR = \frac{a_{11}/a_{21}}{a_{12}/a_{22}}. \quad (3.5)$$

Diferente da razão de chances, o risco relativo é uma razão entre o risco de ocorrência do evento em um grupo e o risco de ocorrência do evento em outro grupo.

O risco relativo estima a magnitude entre o ‘desfecho’, que é a variável de interesse no estudo, e a exposição ao ‘fator de risco’, que é uma variável que supõe-se ser associada ao desfecho, indicando quantas vezes a ocorrência do desfecho é maior do que aquela entre os não-expostos [Wagner & Callegari-Jacques, 1998].

Tabela 10 – Tabela de contingência  $2 \times 2$ 

	Sim	Não	<b>Total</b>
Grupo A	$a_{11}$	$a_{12}$	$a_{11} + a_{12}$
Grupo B	$a_{21}$	$a_{22}$	$a_{21} + a_{22}$

Na tabela acima, pode-se considerar o ‘Grupo A’ como o grupo exposto ao fator de risco e o ‘Grupo B’ como o grupo não-exposto ao fator de risco. Considerando os totais das linhas fixos, o cálculo do risco relativo é feito da forma

$$RR = \frac{a_{11}/(a_{11} + a_{12})}{a_{21}/(a_{21} + a_{22})}. \quad (3.6)$$

Na tabela abaixo encontram-se os valores para interpretação dos níveis de magnitude do efeito (*ES*) da razão de chances e do risco relativo [Sullivan & Feinn, 2012].

Tabela 11 – Níveis de Magnitude do Efeito (*ES*) da *OR* e *RR*

<i>Effect Size</i>	<i>OR</i>	<i>RR</i>
<b>Fraco</b>	1,50	2,00
<b>Médio</b>	2,00	3,00
<b>Forte</b>	3,00	4,00

**Código para cálculo da Razão de Chances/ Risco Relativo no Software R :**

```
library (DescTools)

x11<-30
x12<-20
x21<-25
x22<-40
x<-matrix (0 ,2 ,2)
x [1 ,1]<-x11
x [1 ,2]<-x12
x [2 ,1]<-x21
x [2 ,2]<-x22
```

```
OddsRatio(x)
```

```
x11<-30
x12<-20
x21<-25
x22<-40
x<-matrix(0,2,2)
x[1,1]<-x11
x[1,2]<-x12
x[2,1]<-x21
x[2,2]<-x22
RelRisk(x)
```

### 3.1.5 $Q$ de Yule

O  $Q$  de Yule é uma transformação algébrica direta da razão de chances *Odds Ratio*, sendo que a sua variação é de -1 a 1. Tal medida possui utilidade para o estudo do *ES* [Bakeman & Quera, 2011].

Apesar de possuir uma interpretação não tão descritiva igual à *Odds Ratio*, o  $Q$  de Yule é menos vulnerável a uma contagem de células iguais a 0 e pode ser uma alternativa para a *Odds Ratio* como *ES*. Se o resultado da estatística for  $Q = 0$  indica-se que não há efeito entre as duas variáveis.

Considerando a Tabela 9, o cálculo do  $Q$  de Yule é denotado da seguinte forma:

$$Q = \frac{(a_{11} \times a_{22}) - (a_{12} \times a_{21})}{(a_{11} \times a_{22}) + (a_{12} \times a_{21})}, \quad (3.7)$$

e pode ser denotado também da seguinte forma:

$$Q = \frac{(OR - 1)}{(OR + 1)}, \quad (3.8)$$

em que

$OR$  representa a Razão de Chances, expressão obtida em (3.5).

A Tabela 12 traz os valores sugeridos para os níveis de magnitude do efeito (*ES*), sendo valores equivalentes aos níveis de magnitude do efeito do Coeficiente  $Q$  de Yule [Bakeman & Quera, 2011].

Tabela 12 – Níveis de Magnitude do Efeito ( $ES$ ) do  $Q$  de Yule

<i>Effect Size</i>	$ Q $
<b>Fraco</b>	0,11
<b>Médio</b>	0,33
<b>Forte</b>	0,50

### Código para cálculo do $Q$ de Yule no Software R :

```
library(psych)

x11<-30
x12<-20
x21<-25
x22<-40
x<-matrix(0,2,2)
x[1,1]<-x11
x[1,2]<-x12
x[2,1]<-x21
x[2,2]<-x22
YuleQ(x)
```

## 3.2 Escala Ordinal

Esta seção apresenta uma revisão das principais medidas de magnitude do efeito para variáveis ordinais.

### 3.2.1 Delta ( $\delta$ ) de Cliff

O Delta ( $\delta$ ) de Cliff é uma medida não paramétrica que permite quantificar a magnitude da diferença entre dois grupos de observação que não atendem aos pressupostos de normalidade [Cardoso, Mello & Freitas, 2013]. Portanto, o  $\delta$  de Cliff é uma medida utilizada para grupos independentes, que possuem escala de mensuração ordinal e consequentemente é associado ao teste de Mann-Whitney.

Tal medida estima a probabilidade de uma observação aleatoriamente selecionada de um grupo ser maior do que uma observação aleatoriamente selecionada de outro grupo, menos a probabilidade reversa [Cliff, 1996]. Supondo dois grupos de tratamentos  $x$  e  $y$  procedentes de duas populações  $X$  e  $Y$ , o cálculo da estimativa do  $\delta$  de Cliff é dado da forma [Cardoso, Mello & Freitas, 2013]

$$\delta = \frac{\sum_{i=1}^n \sum_{j=1}^m (\mathbb{I}_{(x_i > y_j)}(x_i, y_j) - \mathbb{I}_{(x_i < y_j)}(x_i, y_j))}{mn}, \quad (3.9)$$

em que

$$\mathbb{I}_{(x_i > y_j)}(x_i, y_j) = \begin{cases} 1, & \text{se } x_i > y_j \\ 0, & \text{Caso contrário} \end{cases}$$

e

$$\mathbb{I}_{(x_i < y_j)}(x_i, y_j) = \begin{cases} 1, & \text{se } x_i < y_j \\ 0, & \text{Caso contrário,} \end{cases}$$

onde

$x_i$  : representa a  $i$ -ésima observação do grupo  $x$ ;

$y_j$  : representa a  $j$ -ésima observação do grupo  $y$ ;

$n$  : representa o número de observações do grupo  $x$ ;

$m$  : representa o número de observações do grupo  $y$ .

Em (3.9), cada observação de um grupo é comparada a cada observação no outro grupo, sendo assim, verifica-se quantas vezes as observações de um grupo são maiores ou inferiores ao número de observações do outro grupo. O valor de  $\delta$  varia entre -1 e 1. Se o seu resultado for igual a 1, então todos os valores de um grupo são superiores aos valores do outro grupo; o inverso ocorre quando o resultado for igual a -1.

Como o  $\delta$  de Cliff é linearmente relacionado à estatística  $U$  de Mann-Whitney, dada a estatística de Mann-Whitney, a expressão é denotada da forma [Cliff, 1996]

$$\delta = \frac{2U}{mn} - 1, \quad (3.10)$$

onde

$U$  : representa o resultado da estatística de Mann-Whitney (2.16);

$n$  : representa o número de observações do grupo  $x$ ;

$m$  : representa o número de observações do grupo  $y$ .

A magnitude é avaliada utilizando os limiares fornecidos em [Romano, 2006], e apresentados na Tabela 13.



Tabela 13 – Níveis de Magnitude do Efeito ( $ES$ ) do  $\delta$  de Cliff

<i>Effect Size</i>	$ \delta $
<b>Insignificante</b>	$<0,147$
<b>Fraco</b>	$<0,33$
<b>Médio</b>	$<0,474$
<b>Forte</b>	$>0,474$

**Código para cálculo do  $\delta$  de Cliff no Software R :**

```
library( effsize )

grupo1 <- c(10,10,20,20,20,30,30,30,40,50)
grupo2 <- c(10,20,30,40,40,50)
resultado = cliff.delta( grupo1 , grupo2 , return.dm=T)
print( resultado )
print( resultado$dm)
```

**3.2.2 A de Vargha-Delaney**

O  $A$  de Vargha-Delaney é uma transformação linear simples do  $\delta$  de Cliff, sendo também associada ao teste de Mann-Whitney e utilizada para grupos independentes com escala de mensuração ordinal. Essa estatística varia de 0 a 1, diferente do  $\delta$  de Cliff que varia de -1 a 1 [Vargha & Delaney, 2000].

Considerando uma amostra aleatória  $(x_1, \dots, x_n)$ , procedente de uma população  $X$  e outra amostra aleatória  $(y_1, \dots, y_m)$ , procedente de uma outra população  $Y$ , a estimativa do  $A$  de Vargha-Delaney é dada da forma [Vargha & Delaney, 2000]

$$A = \frac{\sum_{i=1}^n \sum_{j=1}^m (\mathbb{I}_{(x_i > y_j)}(x_i, y_j) + 0.5 \times \mathbb{I}_{(x_i = y_j)}(x_i, y_j))}{mn}, \quad (3.11)$$

em que

$$\mathbb{I}_{(x_i > y_j)}(x_i, y_j) = \begin{cases} 1, & \text{se } x_i > y_j \\ 0, & \text{Caso contrário} \end{cases}$$

e

$$\mathbb{I}_{(x_i = y_j)}(x_i, y_j) = \begin{cases} 1, & \text{se } x_i = y_j \\ 0, & \text{Caso contrário,} \end{cases}$$

onde

$x_i$  : representa a  $i$ -ésima observação do grupo  $x$ ;

$y_j$  : representa a  $j$ -ésima observação do grupo  $y$ ;

$n$  : representa o número de observações do grupo  $x$ ;

$m$  : representa o número de observações do grupo  $y$ .

Baseado nos postos obtidos no teste de Mann-Whitney, o cálculo da estimativa do  $A$  de Vargha-Delaney em (3.11) pode ser feito da forma

$$A = \frac{\frac{R_x}{n} - \frac{(n+1)}{2}}{m}, \quad (3.12)$$

onde

$R_x$  representa a soma dos postos do grupo  $x$ ;

$n$  número de observações no grupo  $x$ ;

$m$  número de observações no grupo  $y$ .

Como o  $A$  de Vargha-Delaney é uma transformação linear do  $\delta$  de Cliff, o seu valor pode ser obtido da forma

$$A = \frac{\delta + 1}{2}, \quad (3.13)$$

em que

$\delta$  representa o resultado do  $ES$   $\delta$  de Cliff (3.9).

Os níveis de magnitude do efeito ( $ES$ ) estão representados na Tabela 14 [Vargha & Delaney, 2000].

Tabela 14 – Níveis de Magnitude de Efeito ( $ES$ ) do  $A$  de Vargha-Delaney

<i>Effect Size</i>	<i>A</i>
<b>Fraco</b>	0,56
<b>Médio</b>	0,64
<b>Forte</b>	0,71

**Código para cálculo do  $A$  de Vargha and Delaney no Software R :**

```
library ( effsize )
```

```
grupoa = rnorm (100 , mean=10)
```

```
grupob = rnorm (100 , mean=12)
```

$d = (c(\text{grupos}a, \text{grupos}b))$   
 $VD.A(\text{grupos}a, \text{grupos}b)$

### 3.2.3 Coeficiente $r$

O coeficiente  $r$  é um *ES* utilizado para grupos de tratamentos pareados, com nível de mensuração ordinal e, conseqüentemente, ele é associado ao teste de Wilcoxon.

Pressupondo  $n$  pares de valores  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , o coeficiente  $r$  que obtém o grau da diferença entre as duas variáveis  $X$  e  $Y$  é denotado da forma [Pallant, 2007]

$$r = \frac{Z}{\sqrt{n_x + n_y}}, \quad (3.14)$$

onde

$Z$  é o resultado da estatística do teste de Wilcoxon, dada pela expressão em (2.13);

$n_x$  representa o número de observações no grupo 1;

$n_y$  representa o número de observações no grupo 2.

A Tabela 15 traz os valores para os níveis de magnitude do efeito (*ES*) sugeridos pela literatura [Pallant, 2007].

Tabela 15 – Níveis de Magnitude do Efeito (*ES*) do Coeficiente  $r$

<i>Effect Size</i>	$r$
<b>Fraco</b>	0,10
<b>Médio</b>	0,30
<b>Forte</b>	0,50

#### Código para cálculo do coeficiente $r$ no Software R :

```
x0<-runif(5,min = 10, max = 20)
diff0<-runif(5,min = -12, max= 30)
y0<-x0+diff0

x<-x0
y<-y0
p<-wilcox.test(x0,y0,paired = T,exact=F,correct=F)$p.value

z<-qnorm(1-(p/2))
z1<-z/sqrt(10)
```

### 3.2.4 $\tau$ de Kendall

O  $\tau$  de Kendall é um *ES* utilizado quando as duas variáveis possuem nível de escala ordinal. É similarmente uma medida não paramétrica que varia entre -1 e 1 [Espírito-Santo & Daniel, 2017].

Sua fórmula é a mesma da correlação de Kendall dada pela expressão (2.20)

$$\tau_{Kendall} = \frac{N_C - N_E}{\frac{1}{2}(n(n-1))}, \quad (3.15)$$

em que

$N_C$  representa o número de pares com a ordenação correta;

$N_E$  representa o número de pares com a ordenação errada;

$n$  tamanho da amostra.

Os níveis de magnitude do efeito (*ES*) estão representados na Tabela 16 [Espírito-Santo & Daniel, 2017].

Tabela 16 – Níveis de Magnitude do Efeito (*ES*) do  $\tau$  de Kendall

<i>Effect Size</i>	$ \tau $
<b>Fraco</b>	0,10
<b>Médio</b>	0,30
<b>Forte</b>	0,50

#### Código para cálculo do $\tau$ de Kendall no Software R :

```
x<-c(11,15,14,17,19,13,15,21,15,12,16,24)
y<-c(3,15,8,13,21,7,0,14,11,4,8,15)
kendall<-cor.test(x,y,method="kendall")
```

### 3.3 Escala Intervalar

Esta seção apresenta uma revisão das principais medidas de magnitude do efeito para variáveis com nível de mensuração intervalar.

#### 3.3.1 $d$ de Cohen

Dentre os  $ES$ , uma medida bastante utilizada é o  $d$  de Cohen. A sua utilização pode abranger grupos não-pareados ou pareados com nível de mensuração intervalar, portanto tal medida é associada ao teste  $t$  não-pareado quando as duas amostras são não-pareadas e é associada ao teste  $t$  pareado quando as duas amostras são pareadas.

Considerando que os dois grupos sejam independentes, a expressão para o cálculo do  $d$  de Cohen é dada da forma [Lakens, 2013]

$$d = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}}, \quad (3.16)$$

onde

$\bar{X}$  representa a média do grupo 1;

$\bar{Y}$  representa a média do grupo 2;

$S_1^2$  representa a variância amostral do grupo 1;

$S_2^2$  representa a variância amostral do grupo 2;

$n$  representa o número de observações no grupo 1;

$m$  representa o número de observações no grupo 2.

A fórmula do  $d$  de Cohen (3.16) apresenta uma diferença padronizada da média dos dois grupos no numerador e a expressão no denominador representa a estimativa do desvio padrão das amostras combinado. A medida é influenciada pelo tamanho da amostra, ou seja, o seu resultado sofrerá uma alteração resultando um valor com uma diferença não muito grande à medida que o tamanho da amostra aumenta.

Para resolver essa questão da mudança no valor do  $d$  de Cohen, propusemos utilizar os estimadores de máxima verossimilhança para  $\sigma_1^2$  e  $\sigma_2^2$  no denominador da medida, ao invés dos estimadores não viciados  $S_1^2$  e  $S_2^2$ .

Os estimadores de máxima verossimilhança de  $\sigma_1^2$  e  $\sigma_2^2$  são dados, respectivamente,

por

$$\hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (3.17)$$

e

$$\hat{\sigma}_2^2 = \frac{1}{m} \sum_{j=1}^m (Y_j - \bar{Y})^2 \quad (3.18)$$

em que

$X_i$  representa a  $i$ -ésima observação no grupo 1;

$Y_j$  representa a  $j$ -ésima observação no grupo 2;

$\bar{X}$  representa a média do grupo 1;

$\bar{Y}$  representa a média do grupo 2;

$n$  representa o número de observações no grupo 1;

$m$  representa o número de observações no grupo 2.

Considerando que as duas amostras sejam pareadas, a estimativa do  $ES$  pode ser calculada diretamente do valor da estatística  $T$  obtida ao realizar o teste  $t$  pareado, sendo calculada da forma [Rosenthal, 1991]

$$d = \frac{T}{\sqrt{n}}, \quad (3.19)$$

onde

$T$  é o resultado da estatística  $T$ , resultado obtido no teste  $t$  pareado (2.26);

$n$  número de observações nas duas amostras pareadas, sendo amostras pareadas  $n = m$ .

Assim como o  $d$  de Cohen associado ao teste  $t$  não-pareado, o  $d$  de Cohen associado ao teste  $t$  pareado também é influenciado pelo tamanho da amostra, sendo que o seu resultado apresentará uma diferença não muito grande à medida que o tamanho da amostra aumenta.

Como no caso anterior, uma proposta para poder manter este valor invariante à medida que aumenta o tamanho da amostra é utilizar o estimador de máxima verossimilhança  $\sigma_D^2$  ao invés do estimador  $S_D^2$ .

O estimador de máxima verossimilhança de  $\sigma_D^2$  é dado por

$$\hat{\sigma}_D^2 = \frac{1}{n} \sum_{i=1}^n (D_i - \bar{D})^2. \quad (3.20)$$

Cohen definiu os níveis de magnitude do efeito (*ES*). Tais níveis estão especificados na Tabela 17 [Cohen, 1988].

Tabela 17 – Níveis de Magnitude do Efeito (*ES*) do *d* de Cohen

<i>Effect Size</i>	$ d $
<b>Fraco</b>	0,20
<b>Médio</b>	0,50
<b>Forte</b>	0,80

Sendo assim, pelo critério de Cohen(1988), um efeito  $d = 0,2$  é considerado *pequeno*,  $d = 0,5$  é considerado *médio* e  $d = 0,8$  é considerado *grande*.

### Código para cálculo do *d* de Cohen no Software R :

```
library(effsize)

grupoa = rnorm(100, mean=10)
grupob = rnorm(100, mean=12)
d = (c(grupoa, grupob))
cohen.d(grupoa, grupob)

x0<-rnorm(5, 2, 2.838)
diff0<-rnorm(5, 1, 2.838)
y0<-x0+diff0
n<-5
nx<-length(x0)
ny<-length(y0)

t<-t.test(x0,y0, paired = T)$statistic
cohen<-t/sqrt(n)
```

### 3.3.2 $g$ de Hedges

Em casos que grupos apresentam tamanhos diferentes de pequena dimensão, o  $g$  de Hedges é o  $ES$  mais indicado, sendo uma medida também associada ao teste  $t$  não-pareado [Hedges & Olkin, 1986].

O  $g$  de Hedges é semelhante à medida  $d$  de Cohen. No entanto há uma alteração em sua estimativa para corrigir o viés de pequenas amostras [Cardoso, Mello & Freitas, 2013]. Quando o número de observações na amostra for menor que 20, o  $g$  de Hedges mostra um resultado diferente do  $d$  de Cohen, já se os tamanhos de amostra forem maior que 20, os resultados de ambas as estatísticas são aproximadamente equivalentes.

Em sua fórmula há um ajustamento para pequenas amostras, sendo sua estimativa denotada da forma [Cardoso, Mello & Freitas, 2013]:

$$\hat{g} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}}} \left( 1 - \frac{3}{4(n-m) - 9} \right), \quad (3.21)$$

onde

$\bar{X}$  representa a média do grupo 1;

$\bar{Y}$  representa a média do grupo 2;

$S_1^2$  representa a variância amostral do grupo 1;

$S_2^2$  representa a variância amostral do grupo 2;

$n$  representa o número de observações no grupo 1;

$m$  representa o número de observações no grupo 2.

Em (3.21), mesmo com o ajuste para o cálculo do  $ES$ , o  $g$  de Hedges, assim como o  $d$  de Cohen muda o seu valor à medida que aumenta o tamanho da amostra. Para manter este resultado invariante pode-se substituir os estimadores  $S_1^2$  e  $S_2^2$  pelos estimadores de máxima verossimilhança  $\hat{\sigma}_1^2$  (3.17) e  $\hat{\sigma}_2^2$  (3.18).

O resultado do valor obtido pela medida  $g$  de Hedges em (3.21) é interpretado de forma semelhante ao valor obtido pela medida de  $d$  de Cohen.

Tabela 18 – Níveis de Magnitude do Efeito ( $ES$ ) do  $g$  de Hedges

<i>Effect Size</i>	$ g $
<b>Fraco</b>	0,20
<b>Médio</b>	0,50
<b>Forte</b>	0,80



Sendo assim, pelo critério de Cohen(1988), um efeito  $g = 0,2$  é considerado *pequeno*,  $g = 0,5$  é considerado *médio* e  $g = 0,8$  é considerado *grande*.

### Código para cálculo do $g$ de Hedges no Software R :

```
library(effsize)
grupoa = rnorm(100,mean=10)
grupob = rnorm(100,mean=12)
d = (c(grupoa , grupob))
f = rep(c("grupoa" , "grupob"), each=100)

cohen.d(d,f,hedges.correction=TRUE)
```

### 3.3.3 Delta ( $\Delta$ ) de Glass

O Delta ( $\Delta$ ) de Glass é um  $ES$  associado ao teste  $t$  não-pareado. Sendo utilizado quando os grupos de tratamentos são independentes e quando os desvios padrão dos dois grupos são diferentes, havendo um comprometimento da homogeneidade de variâncias [Ellis, 2010].

O cálculo do  $\Delta$  de Glass consiste no valor amostral da diferença da média padronizada através do desvio-padrão do grupo de controle [Lakens, 2013], supondo que o desvio padrão do grupo de controle não é afetado pelos efeitos do grupo de tratamento, refletindo mais de perto o desvio padrão populacional [Ellis, 2010]. Seu cálculo é feito por meio da expressão

$$\Delta = \frac{\bar{X} - \bar{Y}}{S_{controle}}, \quad (3.22)$$

em que

$\bar{X}$  representa a média do grupo 1;

$\bar{Y}$  representa a média do grupo 2;

$S_{controle}$  representa o desvio padrão amostral do grupo de controle.

O resultado do valor obtido pela medida  $\Delta$  de Glass em (3.22) é interpretado de forma semelhante ao valor obtido pela medida de  $d$  de Cohen:

Tabela 19 – Níveis de Magnitude do Efeito ( $ES$ ) do  $\Delta$  de Glass

<i>Effect Size</i>	$ \Delta $
<b>Fraco</b>	0,20
<b>Médio</b>	0,50
<b>Forte</b>	0,80

Sendo assim, pelo critério de [Cohen, 1988], um efeito  $\Delta = 0,2$  é considerado *pequeno*,  $\Delta = 0,5$  é considerado *médio* e  $\Delta = 0,8$  é considerado *grande*.

**Código para cálculo do  $\Delta$  de Glass no Software R :**

```
tratamento = rnorm(50, mean = 35)
controle = rnorm(50, mean = 25)
desvio_trat = sqrt(var(tratamento))
desvio_cont = sqrt(var(controle))
media_trat = mean(tratamento)
media_cont = mean(controle)
delta = ((media_trat - media_cont) / desvio_cont)
```

### 3.4 Influência do tamanho da amostra no *Effect Size*

Visto que os *ES* devem medir a significância prática (e não a significância estatística), é natural que eles não sejam influenciadas pelo tamanho da amostra. No entanto, não são todos os *ES* que apresentam essa propriedade. A Tabela 20 apresenta uma lista com as medidas que satisfazem ou não essa propriedade.

Tabela 20 – *Effect Size* muda ao duplicar o tamanho da amostra?

Escala de mensuração	<i>Effect Size</i>	Resultado altera ao dobrar tamanho da amostra	Resultado não altera ao dobrar tamanho da amostra
<b>Nominal</b>	Coeficiente $\phi$ - associado ao teste de McNemar		×
	Coeficiente $\phi$ - associado ao teste de Qui-Quadrado		×
	$W$ de Cohen		×
	$V$ de Cramér		×
	<i>Odds Ratio</i>		×
	Risco Relativo		×
	$Q$ de Yule		×
<b>Ordinal</b>	$\delta$ de Cliff		×
	$A$ de Vargha-Delaney		×
	Coeficiente $r$	×	
	$\tau$ de Kendall		×
<b>Intervalar</b>	$d$ de Cohen - associado ao teste $t$ pareado	×	
	$d$ de Cohen - associado ao teste $t$ não-pareado	×	
	$g$ de Hedges	×	
	$\Delta$ de Glass	×	

Observando a Tabela 20, percebe-se todas as medidas com escala de mensuração nominal e boa parte das medidas com escala de mensuração ordinal satisfazem a propriedade de não serem influenciadas pelo tamanho amostral. Nota-se que o *ES* permanece invariante quando dobra-se o tamanho da amostra nessas medidas.

No entanto, percebe-se que todas as medidas com escala de mensuração intervalar, incluindo o  $d$  de Cohen que é o principal *ES*, alteram-se quando dobra-se o tamanho da amostra. Essa alteração, embora não muito grande quando os tamanhos de amostra são grandes, acontece devido ao estimador da variância amostral que é utilizado no cálculo

dessas medidas. Como comentado anteriormente, fizemos uma proposta para manter estes valores invariantes, sendo utilizar os estimadores de máxima verossimilhança para  $\sigma_1^2$  e  $\sigma_2^2$  no denominador da medida, ao invés dos estimadores não viciados  $S_1^2$  e  $S_2^2$ .

Dentre as medidas com escala de mensuração ordinal, o coeficiente  $r$  também mostrou que é influenciado pelo tamanho da amostra, sendo que apresenta alteração em seu resultado quando há duplicação no tamanho da amostra.

## 4 Simulações

Esta seção apresenta os resultados da ilustração de dados simulados para algumas das medidas estudadas neste projeto.

Nas simulações, para garantir (aproximadamente) o mesmo nível de magnitude do efeito ( $ES$ ), uma amostra inicial ( $n_0 = 5$  ou  $n_0 = 10$ ) foi considerada e a mesma foi repetida diversas vezes, sendo que foram geradas diferentes amostras para cada nível de magnitude; este procedimento foi realizado com o auxílio do software R. Com as repetições, pôde-se observar os valores  $p$  do teste estatístico associado a cada  $ES$ , segundo os níveis de magnitude do efeito ( $ES$ ) que são sugeridos na literatura e tamanhos de amostra  $n$ .

Também é apresentada uma proposta de cálculo do tamanho de amostra necessário para se obter um determinado *valor-p* de acordo com o nível de magnitude do efeito ( $ES$ ) definido. Para obter estes tamanhos de amostra  $n$  segundo o nível de magnitude do efeito e *valor-p* utilizou-se o método de interpolação linear. As variáveis utilizadas na interpolação linear foram o tamanho de amostra  $n$  e o *valor-p*.

A interpolante linear é a linha entre dois pontos, considerando os pontos  $(x_0, y_0)$  e  $(x_1, y_1)$ . Para um valor de  $x$  no intervalo  $(x_0, x_1)$ , o valor de  $y$  é a linha entre os pontos. Pode-se definir a seguinte relação:

$$\frac{(y - y_0)}{(x - x_0)} = \frac{(y_1 - y_0)}{(x_1 - x_0)} \quad (4.1)$$

Isolando o valor de  $y$  em (4.1) tem-se

$$y = y_0 + (y_1 - y_0) \frac{x - x_0}{x_1 - x_0}, \quad (4.2)$$

que é a interpolação linear do valor de  $y$  referente a  $x$ , onde

$y$  representa o tamanho da amostra  $n$ ;

$x$  representa o *valor-p*.

## 4.1 Escala Nominal

Para essa escala de mensuração, a ilustração de dados simulados foi realizada para dados não-pareados (utilizando o coeficiente  $\phi$  - associado ao teste Qui-Quadrado) e para dados pareados (utilizando o coeficiente  $\phi$  - associado ao teste de McNemar).

Os níveis de magnitude do efeito ( $ES$ ) utilizados nas tabelas abaixo representam valores próximos aos níveis definidos pela literatura, como por exemplo o nível médio sugerido pela literatura para o coeficiente  $\phi$  associado ao teste Qui-Quadrado é igual a 0,3 e o nível de magnitude do efeito ( $ES$ ) utilizado na simulação foi 0,3273, assim como o coeficiente  $\phi$  associado ao teste de McNemar que o nível de magnitude do efeito ( $ES$ ) utilizado na simulação foi igual a 0,3162.

### 4.1.1 Simulação da medida associada ao teste Qui-quadrado ( $\phi$ )

A Tabela 21 e Figura 1 apresentam os valores  $p$  do teste Qui-Quadrado segundo o tamanho da amostra e  $ES$ .

Tabela 21 – Valores  $p$  do teste Qui-Quadrado segundo o tamanho da amostra  $n$  e Coeficiente  $\phi$

$n$	Coeficiente $\phi$			
	<b>Fraco (0,1021)</b>	<b>Médio (0,3273)</b>	<b>Forte (0,5)</b>	<b>Muito Forte (0,7638)</b>
<b>10</b>	1,00000	0,86303	0,42920	0,12051
<b>20</b>	1,00000	0,39315	0,09353	0,00502
<b>30</b>	0,92577	0,19538	0,02248	0,00023
<b>40</b>	0,80873	0,10123	0,00566	0,00001
<b>50</b>	0,71822	0,05376	0,00146	0
<b>60</b>	0,64468	0,02901	0,00038	0
<b>70</b>	0,58304	0,01584	0,00010	0
<b>80</b>	0,53027	0,00872	0,00003	0
<b>90</b>	0,48437	0,00484	0,00001	0
<b>100</b>	0,44399	0,0027	0	0
<b>150</b>	0,29757	0,00015	0	0
<b>200</b>	0,20661	0,00001	0	0
<b>250</b>	0,1464	0	0	0
<b>300</b>	0,10513	0	0	0
<b>400</b>	0,04346	0	0	0
<b>500</b>	0,03015	0	0	0
<b>600</b>	0,01658	0	0	0
<b>700</b>	0,00922	0	0	0
<b>800</b>	0,00517	0	0	0
<b>900</b>	0,00291	0	0	0
<b>1000</b>	0,00165	0	0	0

O nível descritivo (*valor-p*) do teste diminui à medida que o tamanho da amostra aumenta. Este é um resultado esperado, pois quanto maior o tamanho da amostra, maior é o poder do teste. Com os resultados apresentados na Tabela 21 e observando a Figura 1, percebe-se que as curvas com os valores  $p$  reduzem mais rapidamente com um  $ES$  ( $\phi = 0,5$ ) à medida que o tamanho da amostra aumenta, diferente de um  $ES$  fraco ( $\phi = 0,1021$ ) para o qual os valores  $p$  demoram a reduzir à medida que o tamanho da amostra aumenta.

Considerando um tamanho de amostra  $n = 60$  (30 para cada grupo) e coeficiente  $\phi = 0,3273$  ( $ES$  médio), o nível descritivo, ou seja, o *valor-p* do teste Qui-Quadrado será igual a 0,02901, sendo assim, com este resultado pode-se concluir que haveria rejeição da hipótese nula do teste Qui-Quadrado (hipótese da igualdade de grupos), com um nível de significância de 5%.

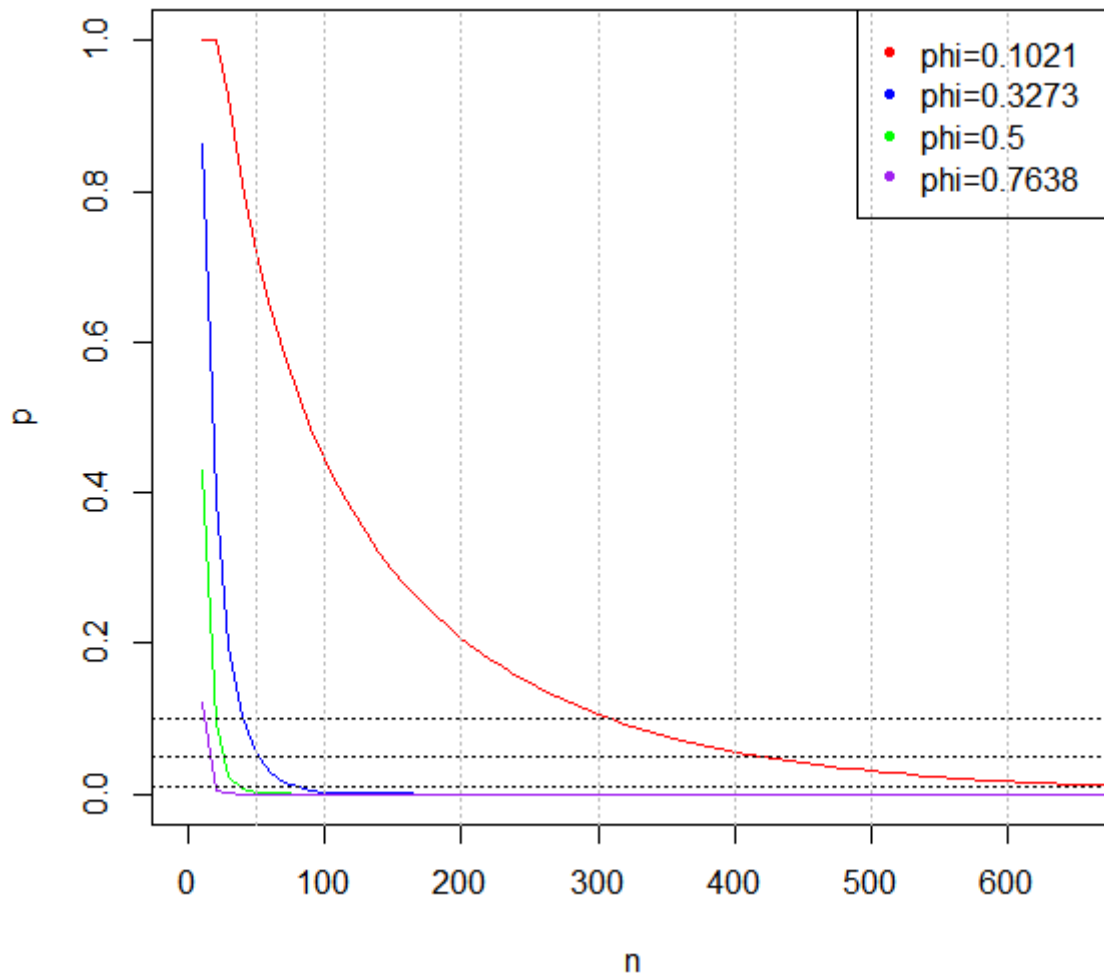


Figura 1 – Gráfico de curvas com valores de  $p$  do teste Qui-Quadrado segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ )

A Tabela 22 apresenta os tamanhos de amostra necessários para a rejeição da hipótese nula  $H_0$  para diversos níveis de significância e  $ES$ .

Tabela 22 – Tamanhos de amostra  $n$  segundo valores  $p$  do teste Qui-Quadrado e Coeficiente  $\phi$

$\phi$	<i>valor-p</i>			
	<b>p = 0,2</b>	<b>p = 0,1</b>	<b>p = 0,05</b>	<b>p = 0,01</b>
<b>Fraco (0,1021)</b>	206	309	390	690
<b>Médio (0,3273)</b>	30	41	52	79
<b>Forte (0,5)</b>	17	20	27	38

Para rejeitar a hipótese nula do teste Qui-Quadrado (hipótese da igualdade de dois grupos) com um nível de significância maior que 5 %, considerando um  $ES$  fraco ( $\phi = 0,1021$ ) seria necessário um tamanho de amostra  $n = 390$  (195 para cada grupo).

Por outro lado, se considerarmos um  $ES$  médio ( $\phi = 0,3273$ ), o tamanho de amostra  $n$  necessário para a rejeição da hipótese nula do teste Qui-Quadrado, considerando um nível de significância maior que 5%, seria de apenas  $n = 52$  (26 para cada grupo).



4.1.2 Simulação da medida associada ao teste de McNemar ( $\phi$ )

A Tabela 23 e Figura 2 apresentam os valores  $p$  do teste de McNemar segundo o tamanho da amostra e  $ES$ .

Tabela 23 – Valores  $p$  do teste de McNemar segundo o tamanho da amostra  $n$  e Coeficiente  $\phi$

$n$	Coeficiente $\phi$			
	<b>Fraco (0,1054)</b>	<b>Médio (0,3162)</b>	<b>Forte (0,5164)</b>	<b>Muito Forte (0,7379)</b>
<b>10</b>	1,00000	0,61708	0,22067	0,04550
<b>20</b>	0,81366	0,28884	0,04331	0,00218
<b>30</b>	0,70031	0,14891	0,00952	0,00012
<b>40</b>	0,61708	0,08012	0,0022	0,00001
<b>50</b>	0,55098	0,04417	0,00052	0
<b>60</b>	0,49624	0,02474	0,00013	0
<b>70</b>	0,44969	0,01402	0,00003	0
<b>80</b>	0,40940	0,00801	0,00001	0
<b>90</b>	0,37406	0,00461	0	0
<b>100</b>	0,34278	0,00266	0	0
<b>150</b>	0,22823	0,00018	0	0
<b>200</b>	0,15672	0,00001	0	0
<b>250</b>	0,10960	0	0	0
<b>300</b>	0,07758	0	0	0
<b>400</b>	0,03983	0	0	0
<b>500</b>	0,02089	0	0	0
<b>600</b>	0,01112	0	0	0
<b>700</b>	0,00598	0	0	0
<b>800</b>	0,00324	0	0	0
<b>900</b>	0,00177	0	0	0
<b>1000</b>	0,00097	0	0	0

Como esperado, o nível descritivo (*valor-p*) do teste diminui à medida que o tamanho da amostra aumenta, pois quanto maior o tamanho da amostra, maior é o poder do teste. Com os resultados apresentados na Tabela 23 e observando a Figura 2, percebe-se que as curvas com os valores  $p$  reduzem mais rapidamente com um  $ES$  forte ( $\phi = 0,5164$ ) à medida que o tamanho da amostra aumenta, diferente de um  $ES$  fraco ( $\phi = 0,1054$ ) que os valores  $p$  demoram a reduzir à medida que o tamanho da amostra aumenta.

Considerando um tamanho de amostra  $n = 20$  e coeficiente  $\phi = 0,5164$  ( $ES$  forte), o nível descritivo, ou seja, *valor-p* do teste de McNemar será igual a 0,04331. Sendo assim, com este resultado pode-se concluir que haveria rejeição da hipótese nula do teste de McNemar, considerando um nível de significância de 5%.

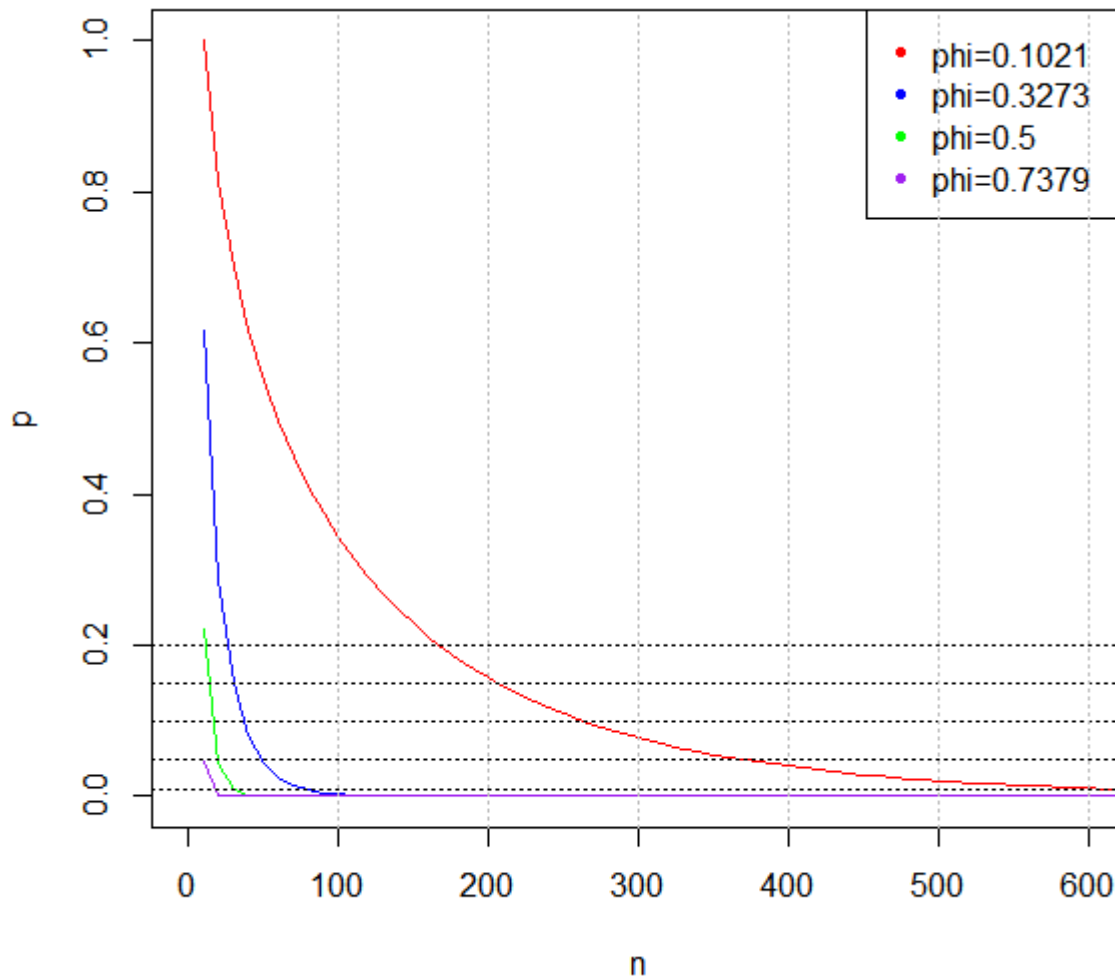


Figura 2 – Gráficos de curvas com os valores de  $p$  do teste de McNemar segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ )

A Tabela 24 apresenta os tamanhos de amostra necessários para a rejeição da hipótese nula  $H_0$  para diversos níveis de significância e  $ES$ .

Tabela 24 – Tamanhos de amostra  $n$  segundo valores  $p$  do teste de McNemar e Coeficiente  $\phi$

$\phi$	<i>valor-p</i>			
	p = 0,2	p = 0,1	p = 0,05	p = 0,01
<b>Fraco</b> <b>(0,1054)</b>	170	265	374	622
<b>Médio</b> <b>(0,3162)</b>	27	38	49	77
<b>Forte</b> <b>(0,5164)</b>	12	17	20	30

Para rejeitar a hipótese nula do teste de McNemar (hipótese da igualdade de dois grupos) com um nível de significância maior que 5 %, considerando um  $ES$  fraco

( $\phi = 0,1054$ ) seria necessário um tamanho de amostra  $n = 374$  (187 para cada grupo).

Por outro lado, se considerarmos um *ES* forte ( $\phi = 0,5164$ ), o tamanho de amostra  $n$  necessário para a rejeição da hipótese nula do teste de McNemar, considerando um nível de significância maior que 5%, seria de apenas  $n = 20$  (10 para cada grupo).

## 4.2 Escala Ordinal

Para essa escala de mensuração, a ilustração de dados simulados foi realizada para dados não-pareados (utilizando o  $\delta$  de Cliff) e para dados pareados (utilizando o coeficiente  $r$ ).

Os níveis de magnitude do efeito ( $ES$ ) utilizados nas tabelas representam valores próximos aos níveis definidos pela literatura, como por exemplo, o nível médio sugerido pela literatura para o  $\delta$  de Cliff é igual a 0,33 e o  $ES$  utilizado na simulação foi 0,36. Já para o coeficiente  $r$ , o nível médio sugerido pela literatura é igual a 0,3 e o  $ES$  utilizado na simulação é igual a 0,384.

### 4.2.1 Simulação da medida associada ao teste de Mann-Whitney ( $\delta$ )

A Tabela 25 e Figura 3 apresentam os valores  $p$  do teste de Mann-Whitney segundo o tamanho da amostra e  $ES$ .

Tabela 25 – Valores  $p$  do teste de Mann-Whitney segundo o tamanho da amostra  $n$  e  $\delta$  de Cliff

$n$	$\delta$ de Cliff			
	<b>Fraco (0,2)</b>	<b>Médio (0,36)</b>	<b>Forte (0,52)</b>	<b>Muito Forte (0,76)</b>
<b>5</b>	0,69048	0,42063	0,22222	0,05556
<b>10</b>	0,44798	0,172	0,04851	0,00393
<b>15</b>	0,34853	0,09153	0,0148	0,00037
<b>20</b>	0,27699	0,05037	0,00471	0,00004
<b>25</b>	0,22302	0,02828	0,00153	0
<b>30</b>	0,18118	0,01609	0,00051	0
<b>35</b>	0,14818	0,00925	0,00017	0
<b>40</b>	0,12181	0,00535	0,00006	0
<b>50</b>	0,08326	0,00182	0,00001	0
<b>100</b>	0,01406	0,00001	0	0
<b>150</b>	0,00261	0	0	0
<b>200</b>	0,00051	0	0	0
<b>300</b>	0,00002	0	0	0

O nível descritivo ( $valor-p$ ) do teste diminui à medida que o tamanho da amostra aumenta, este é um resultado esperado, pois quanto maior o tamanho da amostra, maior é o poder do teste. Com os resultados apresentados na Tabela 25 e observando a Figura 3, percebe-se que as curvas com os valores  $p$  reduzem mais rapidamente com um  $ES$  forte ( $\delta = 0,52$ ) à medida que o tamanho da amostra aumenta, diferente de um  $ES$  fraco ( $\delta = 0,20$ ) para o qual os valores  $p$  demoram a reduzir à medida que o tamanho da amostra aumenta.

Considerando tamanhos de amostras  $n$  e  $m$  iguais a 10 e  $\delta = 0,52$  ( $ES$  forte), o nível descritivo ( $valor-p$ ) do teste de Mann-Whitney será igual a 0,04851, sendo assim, com este resultado pode-se concluir que haveria rejeição da hipótese nula do teste de Mann-Whitney, com um nível de significância de 5%.

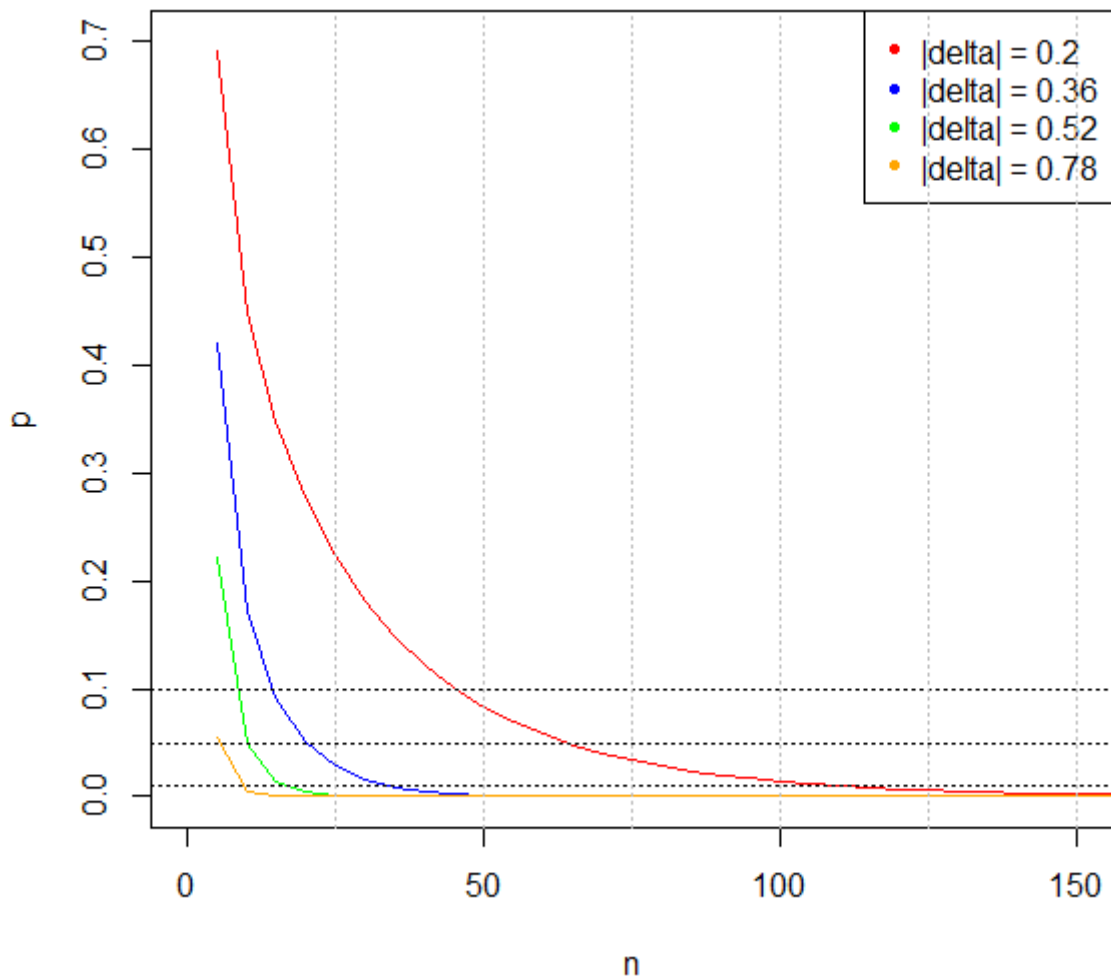


Figura 3 – Gráfico de curvas com os valores de  $p$  do teste de Mann-Whitney segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ )

A Tabela 26 apresenta os tamanhos de amostra necessários para a rejeição da hipótese nula  $H_0$  para diversos níveis de significância e  $ES$ .

Tabela 26 – Tamanhos de amostra  $n$  segundo valores  $p$  do teste de Mann-Whitney e  $\delta$  de Cliff

$\delta$	<i>valor-p</i>			
	<b>p = 0,2</b>	<b>p = 0,1</b>	<b>p = 0,05</b>	<b>p = 0,01</b>
<b>Fraco (0,2)</b>	28	46	75	118
<b>Médio (0,36)</b>	10	15	21	35
<b>Forte (0,52)</b>	6	9	10	18
<b>Muito Forte (0,76)</b>	<5	<5	6	10

Para rejeitar a hipótese nula do teste de Mann-Whitney (hipótese da igualdade de dois grupos) com um nível de significância maior que 5 %, considerando um *ES* forte ( $\delta = 0,52$ ) seriam necessários tamanhos de amostras para  $n$  e  $m$  iguais a 10, sendo que  $n$  representa o número de observações na amostra  $x$  e  $m$  representa o número de observações na amostra  $y$ .

Por outro lado, se considerarmos um *ES* fraco ( $\delta = 0,2$ ), os tamanhos de amostra necessários para a rejeição da hipótese nula do teste de Mann-Whitney, considerando um nível de significância maior que 5%, seriam  $n$  e  $m$  iguais a 75.

### 4.2.2 Simulação da medida associada ao teste de Wilcoxon ( $r$ )

A Tabela 27 e Figura 4 apresentam os valores  $p$  do teste de Wilcoxon segundo o tamanho da amostra e  $ES$ .

Tabela 27 – Valores  $p$  do teste de Wilcoxon segundo o tamanho da amostra  $n$  e Coeficiente  $r$

$n$	Coeficiente $r$		
	<b>Fraco (0,128)</b>	<b>Médio (0,384)</b>	<b>Forte (0,554)</b>
<b>5</b>	0,68583	0,22492	0,07962
<b>10</b>	0,50624	0,07352	0,01223
<b>15</b>	0,39233	0,02615	0,00207
<b>20</b>	0,31135	0,00967	0,00037
<b>25</b>	0,25066	0,00366	0,00007
<b>30</b>	0,20378	0,00141	0,00001
<b>35</b>	0,16686	0,00055	0
<b>40</b>	0,13738	0,00022	0
<b>50</b>	0,09428	0,00003	0
<b>100</b>	0,01633	0	0
<b>150</b>	0,00312	0	0
<b>200</b>	0,00062	0	0
<b>300</b>	0,00003	0	0

Como esperado, o nível descritivo (*valor-p*) do teste diminui à medida que o tamanho da amostra aumenta, pois quanto maior o tamanho da amostra, maior é o poder do teste. Com os resultados apresentados na Tabela 27 e observando a Figura 4, percebe-se que as curvas com os valores  $p$  reduzem mais rapidamente com um  $ES$  forte ( $r = 0,554$ ) à medida que o tamanho da amostra aumenta, diferente de um  $ES$  fraco ( $r = 0,128$ ) para o qual os valores  $p$  demoram a reduzir à medida que o tamanho da amostra aumenta.

Considerando tamanhos de amostras  $n$  e  $m$  iguais a 15 e coeficiente  $r = 0,384$  ( $ES$  médio), o nível descritivo (*valor-p*) do teste de Wilcoxon será igual a 0,02615, sendo assim, com este resultado pode-se concluir que haveria rejeição da hipótese nula do teste de Wilcoxon, com um nível de significância de 5%.

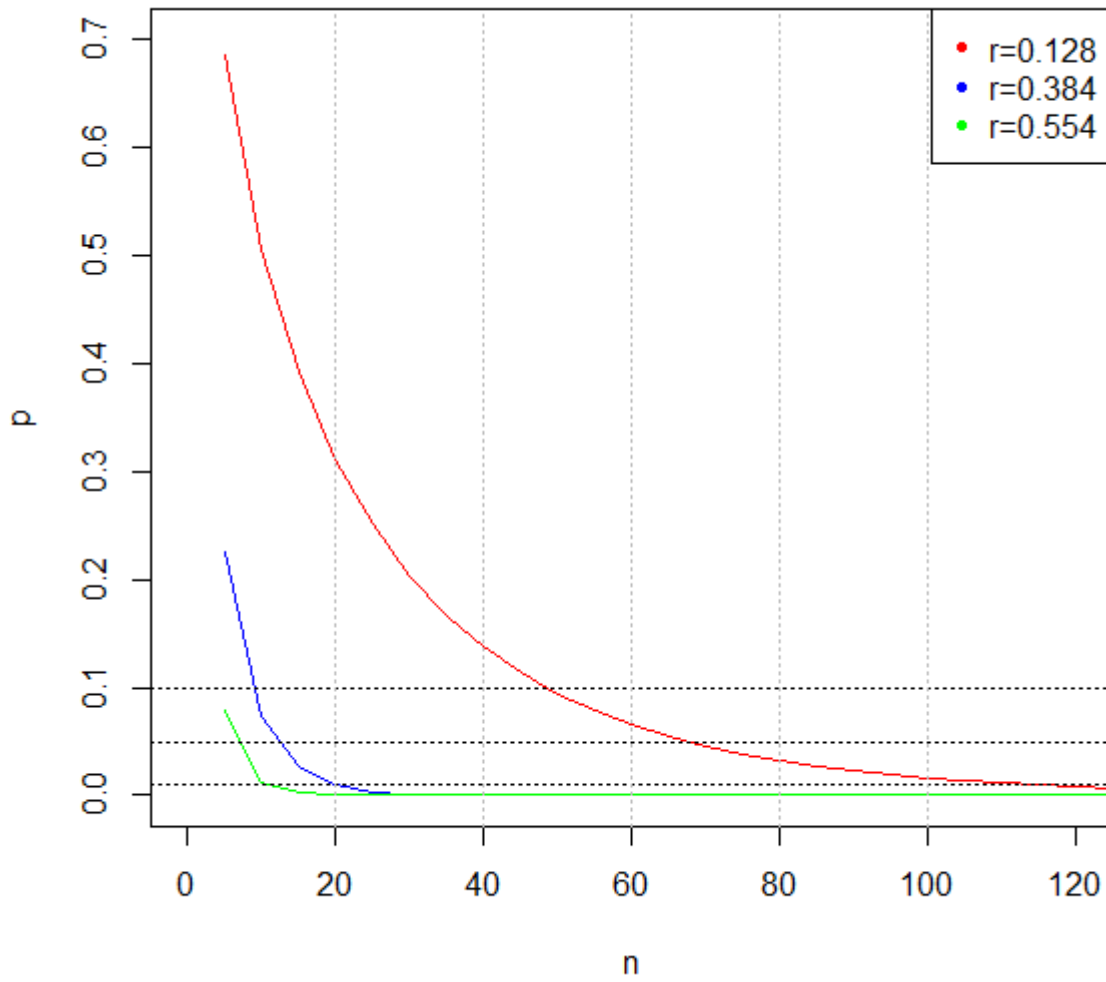


Figura 4 – Gráfico de curvas com os valores de  $p$  do teste de Wilcoxon segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ )



A Tabela 28 apresenta os tamanhos de amostra necessários para a rejeição da hipótese nula  $H_0$  para diversos níveis de significância e  $ES$ .

Tabela 28 – Tamanhos de amostra  $n$  segundo valores  $p$  do teste de Wilcoxon e Coeficiente  $r$

$r$	<i>valor-p</i>			
	<b>p = 0,2</b>	<b>p = 0,1</b>	<b>p = 0,05</b>	<b>p = 0,01</b>
<b>Fraco (0,128)</b>	31	49	79	124
<b>Médio (0,384)</b>	6	10	13	20
<b>Forte (0,554)</b>	<5	<5	8	12

Como comentado anteriormente, o coeficiente  $r$  sofre uma pequena alteração quando dobra-se o tamanho da amostra, sendo que a diferença não é muito grande. Portanto os tamanhos de amostra obtidos na tabela acima são aproximados.

Para rejeitar a hipótese nula do teste de Wilcoxon (hipótese da igualdade de dois grupos) com um nível de significância maior que 5 %, considerando um  $ES$  fraco ( $r = 0,128$ ) seriam necessários tamanhos de amostras para  $n$  e  $m$  iguais a 79, sendo que  $n$  representa o número de observações na amostra  $x$  e  $m$  representa o número de observações na amostra  $y$ .

Por outro lado, se considerarmos um  $ES$  médio ( $r = 0,384$ ), os tamanhos de amostra necessários para a rejeição da hipótese nula do teste de Wilcoxon, considerando um nível de significância maior que 5%, seriam  $n$  e  $m$  iguais a 13.

## 4.3 Escala Intervalar

Para essa escala de mensuração, a ilustração de dados simulados foi realizada para dados não-pareados (utilizando o  $d$  de Cohen - associado ao teste  $t$  não-pareado e o  $g$  de Hedges) e para dados pareados (utilizando o  $d$  de Cohen - associado ao teste  $t$  pareado).

### 4.3.1 Simulação das medidas associadas ao teste $t$ não-pareado ( $d$ e $g$ )

#### 4.3.1.1 $d$ de Cohen

A Tabela 29 e Figura 5 apresentam os valores  $p$  do teste  $t$  não-pareado segundo o tamanho da amostra e  $ES$ .

Tabela 29 – Valores  $p$  do teste  $t$  não-pareado segundo o tamanho da amostra  $n$  e  $d$  de Cohen

$n$	$d$ de Cohen				
	<b>Fraco (0,2)</b>	<b>Médio (0,5)</b>	<b>Forte (0,8)</b>	<b>Muito Forte (1,0)</b>	<b>Muito Forte (2,0)</b>
<b>5</b>	0,76195	0,45764	0,25025	0,16211	0,01802
<b>10</b>	0,64241	0,25499	0,07813	0,03223	0,0003
<b>15</b>	0,56007	0,15314	0,02706	0,00721	0,00001
<b>20</b>	0,49595	0,09511	0,0098	0,00169	0
<b>25</b>	0,44335	0,06023	0,00364	0,00041	0
<b>30</b>	0,39891	0,03865	0,00138	0,0001	0
<b>40</b>	0,32727	0,01633	0,0002	0,00001	0
<b>50</b>	0,27178	0,00706	0,00003	0	0
<b>100</b>	0,11768	0,00013	0	0	0
<b>150</b>	0,0548	0	0	0	0
<b>200</b>	0,02642	0	0	0	0
<b>300</b>	0,00649	0	0	0	0

Como esperado, o nível descritivo ( $valor-p$ ) do teste diminui à medida que o tamanho da amostra aumenta, pois quanto maior o tamanho da amostra, maior é o poder do teste. Com os resultados apresentados na Tabela 29 e observando a Figura 5, percebe-se que as curvas com os valores  $p$  reduzem mais rapidamente com um  $ES$  muito forte ( $d = 1$ ) à medida que o tamanho da amostra aumenta, diferente de um  $ES$  fraco ( $d = 0,2$ ) para o qual os valores  $p$  demoram a reduzir à medida que o tamanho da amostra aumenta.

Considerando tamanhos de amostras  $n$  e  $m$  iguais a 200 e  $d = 0,2$  ( $ES$  fraco), o nível descritivo ( $valor-p$ ) do teste  $t$  não-pareado será igual a 0,02642. Sendo assim, com este resultado pode-se concluir que haveria rejeição da hipótese nula do teste  $t$  não-pareado, com um nível de significância de 5%.

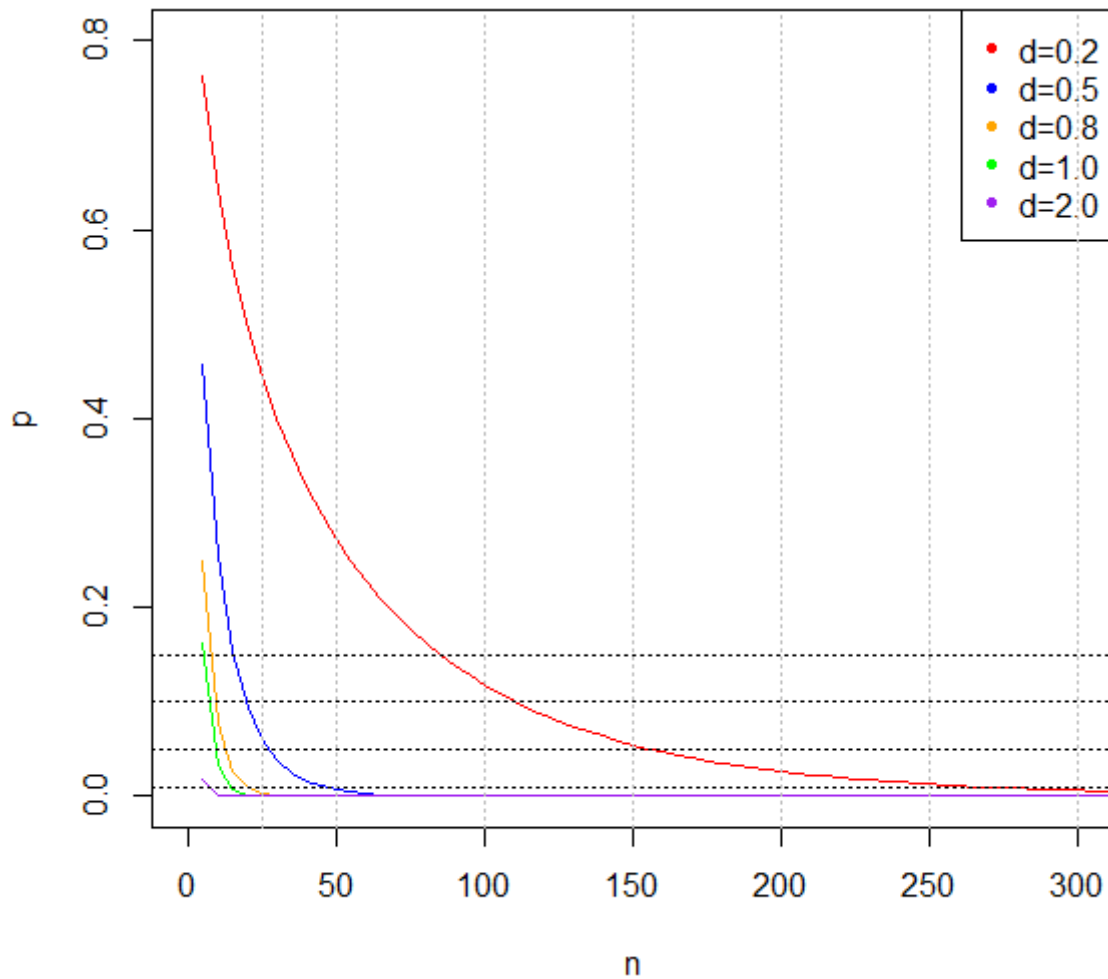


Figura 5 – Gráfico de curvas com os valores de  $p$  do teste  $t$  não-pareado segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ )

A Tabela 30 apresenta os tamanhos de amostra necessários para a rejeição da hipótese nula  $H_0$  para diversos níveis de significância e  $ES$ .

Tabela 30 – Tamanhos de amostra  $n$  segundo valores  $p$  do teste  $t$  não-pareado e  $d$  de Cohen

$d$	<i>valor-p</i>			
	$p = 0,2$	$p = 0,1$	$p = 0,05$	$p = 0,01$
<b>Fraco</b> <b>(0,2)</b>	74	115	159	283
<b>Médio</b> <b>(0,5)</b>	13	20	28	47
<b>Forte</b> <b>(0,8)</b>	7	10	13	20
<b>Muito Forte</b> <b>(1,0)</b>	<5	8	10	15

O  $d$  de Cohen associado ao teste  $t$  não-pareado sofre uma pequena alteração quando

dobra-se o tamanho da amostra, sendo que a diferença não é muito grande. Portanto os tamanhos de amostra obtidos na tabela acima também são aproximados. Os estimadores de máxima verossimilhança não foram utilizados no cálculo do  $d$  de Cohen para realizar a simulação.

Para rejeitar a hipótese nula do teste  $t$  não-pareado (hipótese da igualdade de dois grupos) com um nível de significância maior que 5%, considerando um  $ES$  médio ( $d = 0,5$ ) seriam necessários tamanhos de amostras para  $n$  e  $m$  iguais a 28, sendo que  $n$  representa o número de observações na amostra  $x$  e  $m$  representa o número de observações na amostra  $y$ .

Por outro lado, se considerarmos um  $ES$  forte ( $d = 0,8$ ), os tamanhos de amostra necessários para a rejeição da hipótese nula do teste  $t$  pareado, considerando um nível de significância maior que 5%, seriam  $n$  e  $m$  iguais a 13.

4.3.1.2  $g$  de Hedges

A Tabela 31 e Figura 6 apresentam os valores  $p$  do teste  $t$  não-pareado segundo o tamanho da amostra e  $ES$ .

Tabela 31 – Valores  $p$  do teste  $t$  não-pareado segundo o tamanho da amostra  $n$  e  $g$  de Hedges

$n$	$g$ de Hedges				
	<b>Fraco (0,2)</b>	<b>Médio (0,5)</b>	<b>Forte (0,8)</b>	<b>Muito Forte (1,0)</b>	<b>Muito Forte (2,0)</b>
<b>5</b>	0,73756	0,41324	0,20821	0,12776	0,01164
<b>10</b>	0,60751	0,20986	0,05384	0,01966	0,00011
<b>15</b>	0,51918	0,11558	0,01554	0,00342	0
<b>20</b>	0,45135	0,06598	0,0047	0,00063	0
<b>25</b>	0,39647	0,03846	0,00146	0,00012	0
<b>30</b>	0,35075	0,02274	0,00046	0,00002	0
<b>40</b>	0,27853	0,00816	0,00005	0	0
<b>50</b>	0,22412	0,003	0,00001	0	0
<b>100</b>	0,0835	0,00002	0	0	0
<b>150</b>	0,03364	0	0	0	0
<b>200</b>	0,01406	0	0	0	0
<b>300</b>	0,0026	0	0	0	0

O nível descritivo (*valor-p*) do teste diminui à medida que o tamanho da amostra aumenta. Este é um resultado esperado, pois quanto maior o tamanho da amostra, maior é o poder do teste. Com os resultados apresentados na Tabela 31 e observando a Figura 6, percebe-se que as curvas com os valores  $p$  reduzem mais rapidamente com um  $ES$  muito forte ( $g = 1$ ) à medida que o tamanho da amostra aumenta, diferente de um  $ES$  fraco ( $g = 0, 20$ ) para o qual os valores  $p$  demoram a reduzir à medida que o tamanho da amostra aumenta.

Considerando tamanhos de amostras  $n$  e  $m$  iguais a 25 e  $g = 0, 5$  ( $ES$  médio), o nível descritivo (*valor-p*) do teste  $t$  não-pareado será igual a 0,03846. Sendo assim, com este resultado pode-se concluir que haveria rejeição da hipótese nula do teste  $t$  não-pareado, com um nível de significância de 5%.

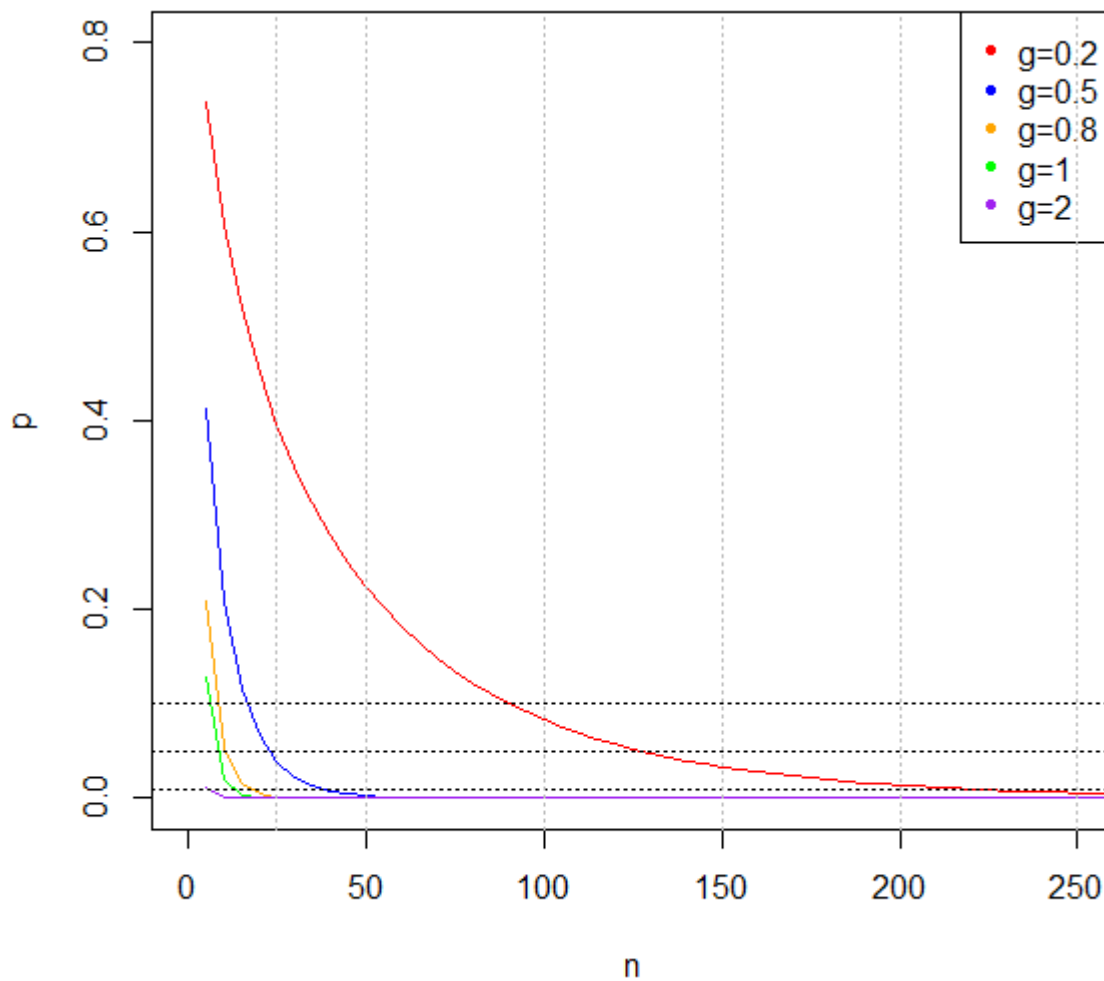


Figura 6 – Gráfico de curvas com os valores de  $p$  do teste  $t$  não-pareado segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ )

A Tabela 32 apresenta os tamanhos de amostra necessários para a rejeição da hipótese nula  $H_0$  para diversos níveis de significância e  $ES$ .

Tabela 32 – Tamanhos de amostra  $n$  segundo valores  $p$  do teste  $t$  não-pareado e  $g$  de Hedges

$g$	<i>valor-p</i>			
	$p = 0,2$	$p = 0,1$	$p = 0,05$	$p = 0,01$
<b>Fraco</b> <b>(0,2)</b>	59	95	134	236
<b>Médio</b> <b>(0,5)</b>	11	17	23	39
<b>Forte</b> <b>(0,8)</b>	6	9	11	18
<b>Muito Forte</b> <b>(1,0)</b>	<5	7	9	13

O  $g$  de Hedges também sofre uma pequena alteração quando dobra-se o tamanho da amostra, sendo que a diferença não é muito grande. Portanto os tamanhos de amostra obtidos na tabela acima também são aproximados. Os estimadores de máxima verossimilhança não foram utilizados no cálculo do  $g$  de Hedges para realizar a simulação.

Para rejeitar a hipótese nula do teste  $t$  não-pareado (hipótese da igualdade de dois grupos) com um nível de significância maior que 5 %, considerando um  $ES$  fraco ( $g = 0,2$ ), seriam necessários tamanhos de amostras para  $n$  e  $m$  iguais a 134, sendo que  $n$  representa o número de observações na amostra  $x$  e  $m$  representa o número de observações na amostra  $y$ .

Por outro lado, se considerarmos um  $ES$  forte ( $g = 0,8$ ), os tamanhos de amostra necessários para a rejeição da hipótese nula do teste  $t$  pareado, considerando um nível de significância maior que 5%, seriam  $n$  e  $m$  iguais a 11.

#### 4.3.2 Simulação da medida associada ao teste $t$ pareado ( $d$ )

A Tabela 33 e Figura 7 apresentam os valores  $p$  do teste  $t$  pareado segundo o tamanho da amostra e  $ES$ .

Tabela 33 – Valores  $p$  do teste  $t$  pareado segundo o tamanho da amostra  $n$  e  $d$  de Cohen

$n$	$d$ de Cohen				
	<b>Fraco (0,2)</b>	<b>Médio (0,5)</b>	<b>Forte (0,80)</b>	<b>Muito Forte (1,0)</b>	<b>Muito Forte (2,0)</b>
<b>5</b>	0,67786	0,32591	0,14811	0,08899	0,01105
<b>10</b>	0,51916	0,12765	0,02506	0,00846	0,00009
<b>15</b>	0,41682	0,05505	0,00479	0,00092	0
<b>20</b>	0,34195	0,02475	0,00096	0,00011	0
<b>25</b>	0,28417	0,0114	0,0002	0,00001	0
<b>30</b>	0,23825	0,00533	0,00004	0	0
<b>40</b>	0,17047	0,0012	0	0	0
<b>50</b>	0,12395	0,00028	0	0	0
<b>100</b>	0,02835	0	0	0	0
<b>150</b>	0,00711	0	0	0	0
<b>200</b>	0,00186	0	0	0	0
<b>300</b>	0,00014	0	0	0	0

O nível descritivo ( $valor-p$ ) do teste diminui à medida que o tamanho da amostra aumenta. Este é um resultado esperado, pois quanto maior o tamanho da amostra, maior é o poder do teste. Com os resultados apresentados na Tabela 33 e observando a Figura 7, percebe-se que as curvas com os valores  $p$  reduzem mais rapidamente com um  $ES$  muito forte ( $d = 1$ ) à medida que o tamanho da amostra aumenta, diferente de um nível de

magnitude do efeito fraco ( $d = 0,20$ ) para o qual os valores  $p$  demoram a reduzir à medida que o tamanho da amostra aumenta.

Considerando tamanhos de amostras  $n$  e  $m$  iguais a 10 e  $d = 0,8$  ( $ES$  forte), o nível descritivo ( $valor-p$ ) do teste  $t$  pareado será igual a 0,02506. Sendo assim, com este resultado pode-se concluir que haveria rejeição da hipótese nula do teste  $t$  pareado, com um nível de significância de 5%.

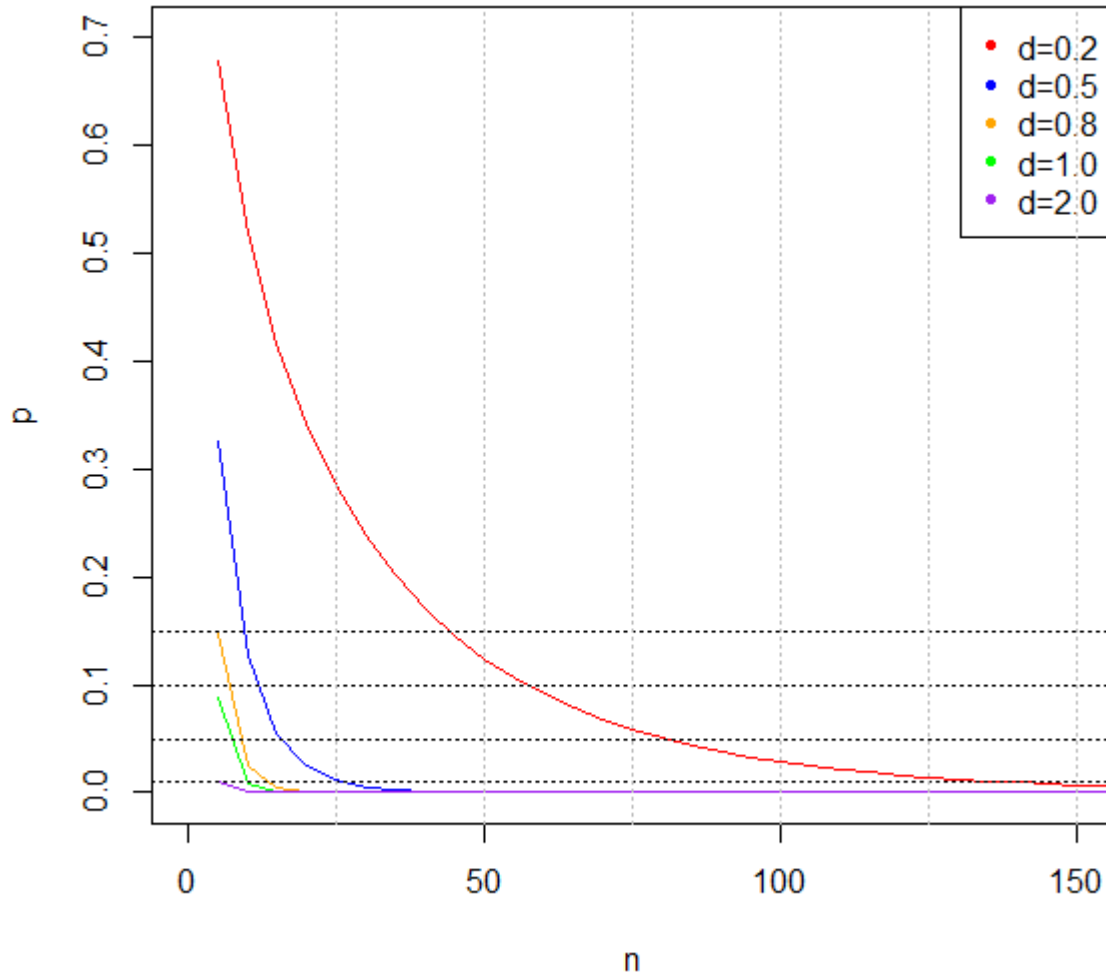


Figura 7 – Gráfico de curvas com os valores de  $p$  do teste  $t$  pareado segundo o tamanho de amostra  $n$  para diversos níveis de magnitude do efeito ( $ES$ )



A Tabela 34 apresenta os tamanhos de amostra necessários para a rejeição da hipótese nula  $H_0$  para diversos níveis de significância e  $ES$ .

Tabela 34 – Tamanhos de amostra  $n$  segundo valores  $p$  do teste  $t$  pareado e  $d$  de Cohen

$d$	<i>valor-p</i>			
	<b>p = 0,2</b>	<b>p = 0,1</b>	<b>p = 0,05</b>	<b>p = 0,01</b>
<b>Fraco (0,2)</b>	26	63	89	144
<b>Médio (0,5)</b>	9	12	16	27
<b>Forte (0,8)</b>	<5	7	9	14
<b>Muito Forte (1,0)</b>	<5	<5	8	10

Como comentado anteriormente, o  $d$  de Cohen associado ao teste  $t$  pareado sofre uma pequena alteração quando dobra-se o tamanho da amostra, sendo que a diferença não é muito grande. Portanto os tamanhos de amostra obtidos na tabela acima são aproximados. Os estimadores de máxima verossimilhança não foram utilizados no cálculo do  $d$  de Cohen para realizar a simulação.

Para rejeitar a hipótese nula do teste  $t$  pareado (hipótese da igualdade de dois grupos) com um nível de significância maior que 5 %, considerando um  $ES$  fraco ( $d = 0,2$ ) seriam necessários tamanhos de amostras para  $n$  e  $m$  iguais a 89, sendo que  $n$  representa o número de observações na amostra  $x$  e  $m$  representa o número de observações na amostra  $y$ .

Por outro lado, se considerarmos um  $ES$  médio ( $d = 0,5$ ), os tamanhos de amostra necessários para a rejeição da hipótese nula do teste  $t$  pareado, considerando um nível de significância maior que 5%, seriam  $n$  e  $m$  iguais a 16.



## 5 Conclusão

---

O presente trabalho fez uma breve apresentação das principais medidas de magnitude do efeito (*Effect Size*) utilizadas na comparação de dois grupos. O *ES* nada mais é do que uma maneira de quantificar a magnitude, ou o tamanho da diferença entre dois grupos, sendo uma proposta para complementar os resultados dos testes de significância estatística, promovendo uma abordagem mais científica.

Neste estudo, os *ES* foram classificadas segundo o nível de mensuração da variável. Para a escala de mensuração nominal, o coeficiente  $\phi$  se caracteriza como a medida mais utilizada para quantificar a magnitude da diferença entre dois grupos. O  $W$  de Cohen e o  $V$  de Cramér também são medidas utilizadas para esta escala de mensuração. Já a *Odds Ratio* e o Risco relativo são *ES* amplamente utilizados em estudos epidemiológicos e o  $Q$  de Yule é uma alternativa para a *Odds Ratio*.

Na escala de mensuração ordinal, o  $\delta$  de Cliff e o  $A$  de Vargha-Delaney são medidas utilizadas para grupos independentes, sendo que o  $A$  é uma transformação linear do  $\delta$  de Cliff. Já o Coeficiente  $r$  e o  $\tau$  de Kendall são medidas utilizadas para grupos pareados. As medidas mais utilizadas para grupos independentes e associadas ao teste  $t$  não-pareado ( $d$  de Cohen,  $g$  de Hedges e o  $\Delta$  de Glass) são medidas com escala de mensuração intervalar (normal), sendo que o  $d$  de Cohen é considerado o principal *ES* e ainda pode ser utilizado para grupos pareados.

Com o intuito de mostrar o desempenho dos *ES*, na seção 4 deste trabalho estão contidos os resultados da ilustração de dados simulados para algumas das medidas estudadas neste projeto. Com os resultados obtidos, pôde-se verificar que quando há um *ES* considerado ‘forte’ entre os grupos, haverá uma redução mais rápida dos valores  $p$  do teste estatístico à medida que se aumenta o tamanho da amostra, sendo que a rejeição da hipótese de igualdade entre dois grupos será mais evidente, diferente dos casos com *ES* ‘fraco’ em que serão necessários grandes tamanhos de amostra para poder gerar pequenos valores  $p$ . Este resultado mostra que quanto maior o *ES*, maior é o poder do teste.

Com os níveis de magnitude do efeito (*ES*) definidos, foi apresentada uma proposta de cálculo do tamanho de amostra necessário para a obtenção de determinados valores  $p$ . Esses tamanhos de amostra foram obtidos para algumas das medidas estudadas neste trabalho.

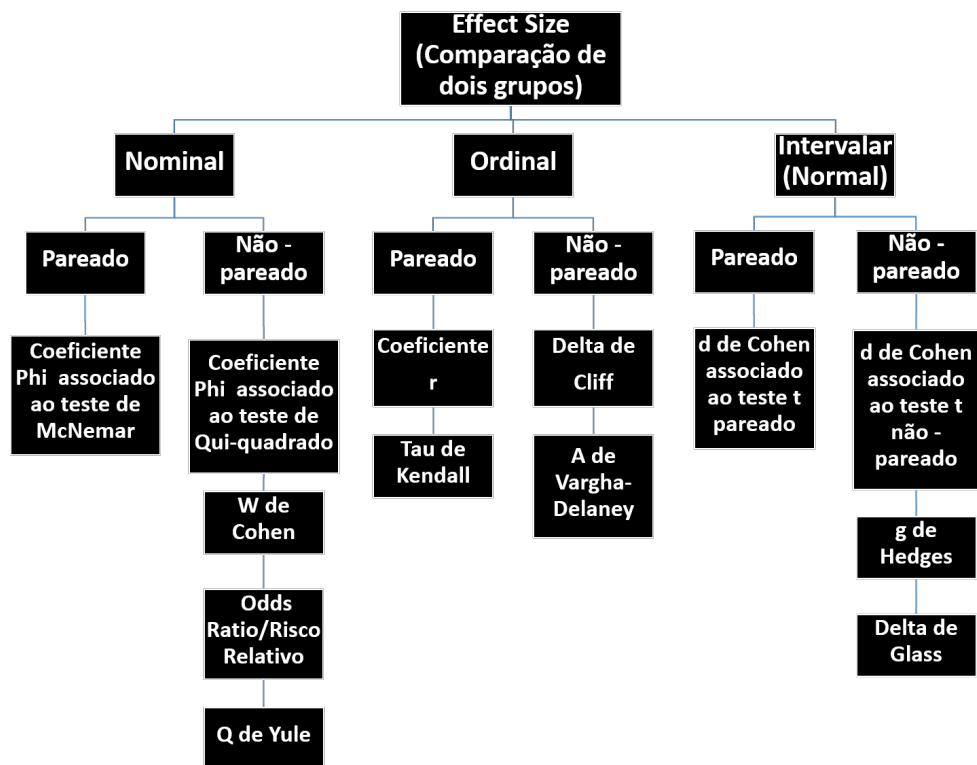
Os estudos com diferentes tamanhos de amostra diferem em seus valores de significância estatística. No entanto, os *ES* possuem a característica de não sofrer alteração.

Com o propósito de verificar essa propriedade, pôde-se concluir que boa parte dos *ES* satisfizeram essa propriedade, com exceção do Coeficiente  $r$  e das medidas com escala de mensuração intervalar (normal), sendo o  $d$  de Cohen,  $g$  de Hedges e  $\Delta$  de Glass. Esse fato ocorre com as medidas com escala de mensuração intervalar, devido ao fato dessas medidas utilizarem a variância amostral dos grupos em seu cálculo.

# Apêndice A - Principais medidas de magnitude do efeito

A Figura 8 traz um resumo das medidas estudadas neste trabalho segundo a escala de mensuração da variável.

Figura 8 – Principais medidas de magnitude do efeito segundo escala de mensuração





# Referências

- [1] CONOVER, W.J. (1999). *Practical nonparametric statistics* 3.ed. New York: J Wiley.
- [2] CONBOY, J.E. (2003). *Algumas medidas típicas univariadas da magnitude do efeito. Análise Psicológica*. v.2, n.11, p.155.
- [3] WAGNER, M.B. and CALLEGARI-JACQUES S.M. (1998). *Medidas de associação em estudos epidemiológicos: risco relativo e odds ratio. Jornal de Pediatria* p.247-251.
- [4] SULLIVAN, GM. and FEINN, R. (2012). *Using Effect Size-or Why the P Value Is Not Enough. Journal of Graduate Medical Education* p.279-282.
- [5] BAKEMAN, R. and QUERA, V. (2011). *Sequential Analysis and Observational Methods for the Behavioral Sciences* New York, NY, US: Cambridge University Press.
- [6] CARDOSO, T.S.G., MELLO C.B. and FREITAS P.M. (2013). *Use of Effectiveness Quantitative Measures to Neuropsychological Rehabilitation Psicologia em Pesquisa* p.121-131.
- [7] COHEN, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* 2.ed. Hillsdale, NJ: Erlbaum.
- [8] ROMANO, J., KROMREY J. D., CORAGGIO, J. Coraggio and SKOWRONEK, J. (2006). *Appropriate statistics for ordinal level data: Should we really be using t-test and cohen's d for evaluating group differences on the NSSE and other surveys?* Annual meeting of the Florida Association of Institutional Research.
- [9] VARGHA, A. and DELANEY, H. D. (2000). *A critique and improvement of the CL common language effect size statistics of McGraw and Wong. Journal of Educational and Behavioral Statistics* p.101-132.
- [10] PALLANT, J. (2007). *SPSS Survival Manual* 3.ed. Open University Press, New York, p.224-225.
- [11] ESPÍRITO-SANTO, H. and DANIEL F. (2017). *Calcular e apresentar tamanhos do efeito em trabalhos científicos (2): Guia para reportar a força das relações Revista Portuguesa de Investigação Comportamental e Social* p.53-64. 3.ed. Open University Press, New York, p.224-225.
- [12] LAKENS, D. (2013). *Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs.*

- [13] ROSENTHAL, R. (1991). *Meta-analytic procedures for social research* Newbury Park, CA: SAGE Publications, Incorporated.
- [14] HEDGES, L.V and OLKIN, I. (1986). *Statistical methods for meta-analysis*. Academic Press.
- [15] BUSSAB, W. e MORETTIN, P. (2013). *Estatística Básica* 8.ed. Ed. Saraiva, São Paulo.
- [16] BORBA, M.C.V (2015). *Testes Estatísticos de Concordância Trabalho de Conclusão de Curso (Bacharelado em Estatística) – Universidade de Brasília - UnB*, Distrito Federal. [Orientador: Prof. Dr. Eduardo Yoshio Nakano].
- [17] ELLIS, P.D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge University Press.
- [18] HUBERTY, C. (2002). *A history of effect size indices*. **Educational and Psychological Measurement** 62 (2), p. 227-240.
- [19] COHEN, J. (1992). *Statistical Power Analysis Current Directions in Psychological Science*, p. 98-101.
- [20] KIRK, R.E. (1996). *Practical significance: A concept whose time has come*. **Educational and Psychological Measurement**, 56, p. 746-759.
- [21] LOUREIRO, L.M.J. and GAMEIRO, M.G.H. (2011). *Interpretação crítica dos resultados estatísticos: para lá da significância estatística*. Referência. Série 3, nº 3, p. 151-162.
- [22] CLIFF, N. (1996). *Ordinal Methods for Behavioral Data Analysis* Lawrence Erlbaum Associates.
- [23] CONBOY, J.E. (2000). *Estimando a dimensão do N de uma amostra*. **INUAF STUDIA**. v.1, n.2, p. 227-274.
- [24] HAYS, W.L. (1973). *Statistics for the social sciences*. 2.ed. New York: Holt, Rinehart and Winston, 954p.



# Anexos

Listing 5.1 – Códigos das simulações do Coeficiente R no Software R

```

set.seed(123456)
x0<-runif(5,min = 10, max = 20)
diff0<-runif(5,min = -12, max= 30)
y0<-x0+diff0
n<-5
x0
diff0
y0

x<-x0
y<-y0
p<-wilcox.test(x0,y0,paired = T,exact=F,correct=F)$p.value

for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-wilcox.test(x,y,paired = T,exact=F,correct=F)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}

z<-qnorm(1-(p/2))

z1<-z/sqrt(10)
z1

plot(n,p,type="l",ylim=c(0,0.7),xlim=c(0,120),col='red')
cbind(n,round(p,5))
cbind(n,round(z1,5))

set.seed(13)

x0<-runif(5,min = 10, max = 20)
diff0<-runif(5,min = -12, max= 30)
y0<-x0+diff0
n<-5
x0
diff0

```

```
y0
```

```
x<-x0
```

```
y<-y0
```

```
p<-wilcox.test(x0,y0,paired = T,exact=F,correct =F)$p.value
```

```
p
```

```
for (i in 2:100){
```

```
  x<-c(x,x0)
```

```
  y<-c(y,y0)
```

```
  p[i]<-wilcox.test(x,y,paired = T,exact=F,correct=F)$p.value
```

```
  n[i]<-5*i
```

```
  cat(length(x),length(y),"\n")
```

```
}
```

```
z<-qnorm(1-(p/2))
```

```
z1<-z/sqrt(10)
```

```
z1
```

```
points(n,p,type="l",col='blue')
```

```
cbind(n,round(p,5))
```

```
cbind(n,round(z1,5))
```

```
set.seed(234)
```

```
x0<-runif(5,min = 10, max = 20)
```

```
diff0<-runif(5,min = -12, max= 30)
```

```
y0<-x0+diff0
```

```
n<-5
```

```
x0
```

```
diff0
```

```
y0
```

```
x<-x0
```

```
y<-y0
```

```
p<-wilcox.test(x0,y0,paired = T,exact=F,correct =F)$p.value
```

```
p
```

```
for (i in 2:100){
```

```
  x<-c(x,x0)
```

```
  y<-c(y,y0)
```

```

p[i]<-wilcox.test(x,y,paired = T,exact=F,correct=F)$p.value
n[i]<-5*i
cat(length(x),length(y),"\n")
}

z<-qnorm(1-(p/2))

z1<-z/sqrt(10)
z1

points(n,p,type="l",col='green')
legend('topright',c("r=0.128","r=0.384","r=0.554"),
col=c("red","blue","green"),pch=rep(20,1))
cbind(n,round(p,5))
abline(h=0.10,lty=3,col=9)
abline(h=0.05,lty=3,col=9)
abline(h=0.01,lty=3,col=9)
abline(v=100,col="gray",lty=3)
abline(v=80,col="gray",lty=3)
abline(v=60,col="gray",lty=3)
abline(v=40,col="gray",lty=3)
abline(v=20,col="gray",lty=3)

```

Listing 5.2 – Códigos das simulações do Coeficiente  $\phi$  - McNemar no Software R

```

A <-0
B <-4
C <-5
D <-1

n<-A+B+C+D
n

x<-matrix(0,2,2)
x[1,1]<-A
x[1,2]<-B
x[2,1]<-C
x[2,2]<-D
x

phi<- ((B-C)^2/(B+C)/n)^0.5
phi

```

```

p<-mcnemar.test(x, correct = T)$p.value
p

for (i in 2:100){
  x[1,1]<-x[1,1]+A
  x[1,2]<-x[1,2]+B
  x[2,1]<-x[2,1]+C
  x[2,2]<-x[2,2]+D
  p[i]<-mcnemar.test(x)$p.value
  n[i]<-10*i
}

x11()
plot(n, p, type="l", ylim=c(0,1), xlim=c(0,650), col='red')
cbind(n, round(p,5))

```

```

A <-1
B <-1
C <-3
D <-5

```

```

n<-A+B+C+D
n

```

```

x<-matrix(0,2,2)
x[1,1]<-A
x[1,2]<-B
x[2,1]<-C
x[2,2]<-D
x

```

```

phi<- ((B-C)^2/(B+C)/n)^0.5
phi

```

```

p<-mcnemar.test(x, correct = T)$p.value
p

```

```

for (i in 2:100){
  x[1,1]<-x[1,1]+A
  x[1,2]<-x[1,2]+B
  x[2,1]<-x[2,1]+C
  x[2,2]<-x[2,2]+D
  p[i]<-mcnemar.test(x)$p.value
  n[i]<-10*i
}

```

```
}

points(n,p,type="l",col='blue')
cbind(n,round(p,5))

A <-0
B <-1
C <-5
D <-4

n<-A+B+C+D
n

x<-matrix(0,2,2)
x[1,1]<-A
x[1,2]<-B
x[2,1]<-C
x[2,2]<-D
x

phi<- ((B-C)^2/(B+C)/n)^0.5
phi

p<-mcnemar.test(x,correct = T)$p.value
p

for (i in 2:100){
  x[1,1]<-x[1,1]+A
  x[1,2]<-x[1,2]+B
  x[2,1]<-x[2,1]+C
  x[2,2]<-x[2,2]+D
  p[i]<-mcnemar.test(x)$p.value
  n[i]<-10*i
}

points(n,p,type="l",col='green')
cbind(n,round(p,5))

A <-1
B <-8
C <-1
```

```

D <-0

n<-A+B+C+D
n

x<-matrix(0,2,2)
x[1,1]<-A
x[1,2]<-B
x[2,1]<-C
x[2,2]<-D
x

phi<- ((B-C)^2/(B+C)/n)^0.5
phi

p<-mcnemar.test(x,correct = T)$p.value
p

for (i in 2:100){
  x[1,1]<-x[1,1]+A
  x[1,2]<-x[1,2]+B
  x[2,1]<-x[2,1]+C
  x[2,2]<-x[2,2]+D
  p[i]<-mcnemar.test(x)$p.value
  n[i]<-10*i
}

points(n,p,type="l",col='purple')
legend('topright',c("phi=0.1021","phi=0.3273","phi=0.5","phi=0.6708"),
,col=c("red","blue","green","purple"),pch=rep(20,1))
cbind(n,round(p,5))
abline(h=0.10,lty=3,col=9)
abline(h=0.05,lty=3,col=9)
abline(h=0.01,lty=3,col=9)
abline(h=0.2,lty=3,col=9)
abline(h=0.15,lty=3,col=9)
abline(v=100,col="gray",lty=3)
abline(v=200,col="gray",lty=3)
abline(v=300,col="gray",lty=3)
abline(v=400,col="gray",lty=3)
abline(v=500,col="gray",lty=3)
abline(v=600,col="gray",lty=3)
abline(v=700,col="gray",lty=3)
abline(v=800,col="gray",lty=3)

```

Listing 5.3 – Códigos das simulações do  $d$  de Cohen - pareado no Software R

```
set.seed(12345)
x0<-rnorm(5,2,2.838)
diff0<-rnorm(5,1,2.838)
y0<-x0+diff0
n<-5
x0
diff0
y0

length(x0)

nx<-length(x0)
ny<-length(y0)

t<-t.test(x0,y0,paired = T)$statistic
t

cohen<-t/sqrt(nx)
cohen

x<-x0
y<-y0
p<-t.test(x,y,paired=T)$p.value

for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-t.test(x,y,paired=T)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}

plot(n,p,type="l",ylim=c(0,0.7),xlim=c(0,150),col='red')
cbind(n,round(p,5))

set.seed(234)
x0<-rnorm(5,2,5.342)
diff0<-rnorm(5,1,5.342)
y0<-x0+diff0
n<-5
```

```
x0
diff0
y0

length(x0)

nx<-length(x0)
ny<-length(y0)

t<-t.test(x0,y0,paired = T)$statistic
t

cohen<-t/sqrt(5)
cohen

x<-x0
y<-y0
p<-t.test(x,y,paired=T)$p.value

for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-t.test(x,y,paired=T)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}
cbind(n,round(p,5))

points(n,p,type="l",col='blue')
cbind(n,round(p,5))

set.seed(567)
x0<-rnorm(5,2,6.375)
diff0<-rnorm(5,1,6.375)
y0<-x0+diff0
n<-5
x0
diff0
y0

length(x0)

nx<-length(x0)
ny<-length(y0)
```



---

```
t<-t.test(x0,y0,paired = T)$statistic
t

cohen<-t/sqrt(5)
cohen

x<-x0
y<-y0
p<-t.test(x,y,paired=T)$p.value

for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-t.test(x,y,paired=T)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}

points(n,p,type="l",col='orange')
cbind(n,round(p,5))

set.seed(234)
x0<-rnorm(5,-4,4)
diff0<-rnorm(5,-1.909,4)
y0<-x0+diff0
n<-5
x0
diff0
y0

length(x0)

nx<-length(x0)
ny<-length(y0)

t<-t.test(x0,y0,paired = T)$statistic
t

cohen<-t/sqrt(5)
cohen

x<-x0
y<-y0
```

```
p<-t.test(x,y,paired=T)$p.value
```

```
for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-t.test(x,y,paired=T)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}
```

```
points(n,p,type="l",col='green')
cbind(n,round(p,5))
```

```
set.seed(234)
x0<-rnorm(5,-7.22805,4)
diff0<-rnorm(5,-7.22805,4)
y0<-x0+diff0
n<-5
x0
diff0
y0
```

```
length(x0)
```

```
nx<-length(x0)
ny<-length(y0)
```

```
t<-t.test(x0,y0,paired = T)$statistic
t
```

```
cohen<-t/sqrt(5)
cohen
```

```
x<-x0
y<-y0
p<-t.test(x,y,paired=T)$p.value
```

```
for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-t.test(x,y,paired=T)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}
```

```

points(n,p,type="l",col='purple')
legend('topright',c("d=0.2","d=0.5","d=0.8","d=1.0","d=2.0"),
col=c("red","blue","orange","green","purple"),pch=rep(20,1))
cbind(n,round(p,5))
abline(h=0.10,lty=3,col=9)
abline(h=0.05,lty=3,col=9)
abline(h=0.01,lty=3,col=9)
abline(h=0.15,lty=3,col=9)
abline(v=25,col="gray",lty=3)
abline(v=50,col="gray",lty=3)
abline(v=75,col="gray",lty=3)
abline(v=100,col="gray",lty=3)
abline(v=125,col="gray",lty=3)
abline(v=150,col="gray",lty=3)

```

Listing 5.4 – Códigos das simulações do  $g$  de Hedges no Software R

```

library(effsize)

set.seed(12345)
n<-5
x0<-rnorm(5,1.031616,2)
y0<-rnorm(5,2.307154,2)

x<-x0
y<-y0
g<-cohen.d(x,y,hedges.correction = TRUE)
g
p<-t.test(x,y,paired=F)$p.value

for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-t.test(x,y,paired=F)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}

plot(n,p,type="l",ylim=c(0,0.8),xlim=c(0,250),col='red')
cbind(n,round(p,7))

```

```

set.seed(12345)
n<-5
x0<-rnorm(5,1.479149,2)
y0<-rnorm(5,2.264623,2)

x<-x0
y<-y0
g<-cohen.d(x,y,hedges.correction = TRUE)
g
p<-t.test(x,y,paired=F)$p.value

for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-t.test(x,y,paired=F)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}

points(n,p,type="l",col='blue')
legend('topright',col=c("red","blue"),pch=rep(20,1))
cbind(n,round(p,5))

```

```

set.seed(12345)
n<-5
x0<-rnorm(5,2.094849,1.5)
y0<-rnorm(5,2.316417,1.5)

x<-x0
y<-y0
g<-cohen.d(x,y,hedges.correction = TRUE)
g
p<-t.test(x,y,paired=F)$p.value

for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-t.test(x,y,paired=F)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}

points(n,p,type="l",col='orange')

```

---

```
legend('topright', col = c("red", "blue", "orange"), pch = rep(20, 1))
cbind(n, round(p, 5))

set.seed(12345)
n<-5
x0<-rnorm(5, 2.00271, 1.5)
y0<-rnorm(5, 1.97925, 1.5)

library(effsize)
x<-x0
y<-y0
g<-cohen.d(x, y, hedges.correction = TRUE)
g
p<-t.test(x, y, paired=F)$p.value

for (i in 2:100){
  x<-c(x, x0)
  y<-c(y, y0)
  p[i]<-t.test(x, y, paired=F)$p.value
  n[i]<-5*i
  cat(length(x), length(y), "\n")
}

points(n, p, type="l", col='green')
legend('topright', col = c("red", "blue", "orange", "green"), pch = rep(20, 1))
cbind(n, round(p, 5))

set.seed(12345)
n<-5
x0<-rnorm(5, 2.80919, 1.2)
y0<-rnorm(5, 1.8103, 1.2)

x<-x0
y<-y0
g<-cohen.d(x, y, hedges.correction = TRUE)
g
p<-t.test(x, y, paired=F)$p.value

for (i in 2:100){
  x<-c(x, x0)
  y<-c(y, y0)
  p[i]<-t.test(x, y, paired=F)$p.value
```

```

n[i]<-5*i
cat(length(x),length(y),"\n")
}

points(n,p,type="l",col='purple')
legend('topright',c("g=0.2","g=0.5","g=0.8","g=1","g=2"),
col=c("red","blue","orange","green","purple"),pch=rep(20,1))
abline(h=0.10, lty=3, col=9)
abline(h=0.05, lty=3, col=9)
abline(h=0.01, lty=3, col=9)
abline(v=25, col="gray", lty=3)
abline(v=50, col="gray", lty=3)
abline(v=100, col="gray", lty=3)
abline(v=150, col="gray", lty=3)
abline(v=200, col="gray", lty=3)
abline(v=250, col="gray", lty=3)

```

Listing 5.5 – Códigos das simulações do  $d$  de Cohen no Software R

```

set.seed(12345)
n<-5
x0<-rnorm(5,1,2)
y0<-rnorm(5,2.307154,2)

x<-x0
y<-y0
d<-cohen.d(x,y)
d
p<-t.test(x,y,paired=F)$p.value

for(i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-t.test(x,y,paired=F)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}

plot(n,p,type="l",ylim=c(0,0.8),xlim=c(0,300),col='red')
cbind(n,round(p,7))

```

---

```
set . seed (12345)
n<-5
x0<-rnorm (5 ,1.4001 ,2)
y0<-rnorm (5 ,2.264623 ,2)

x<-x0
y<-y0
d<-cohen . d(x ,y)
d
p<-t . test (x ,y ,paired=F) $p . value

for (i in 2:100){
  x<-c(x ,x0)
  y<-c(y ,y0)
  p[i]<-t . test (x ,y ,paired=F) $p . value
  n[i]<-5*i
  cat (length(x) ,length(y) ,"\n")
}

points (n ,p ,type="l" ,col='blue ')
legend (' topright ' ,col =c("red" ,"blue") ,pch = rep (20 ,1))
cbind (n ,round (p ,5))

set . seed (12345)
n<-5
x0<-rnorm (5 ,2 ,1.5)
y0<-rnorm (5 ,2.316417 ,1.5)

x<-x0
y<-y0
d<-cohen . d(x ,y)
d
p<-t . test (x ,y ,paired=F) $p . value

for (i in 2:100){
  x<-c(x ,x0)
  y<-c(y ,y0)
  p[i]<-t . test (x ,y ,paired=F) $p . value
  n[i]<-5*i
  cat (length(x) ,length(y) ,"\n")
```

```

}

points(n,p,type="l",col='orange')
legend('topright',col =c("red","blue","orange"), pch = rep(20,1))
cbind(n,round(p,5))

set.seed(12345)
n<-5
x0<-rnorm(5,2,1.5)
y0<-rnorm(5,2.0951,1.5)

library(effsize)
x<-x0
y<-y0
d<-cohen.d(x,y)
d
p<-t.test(x,y,paired=F)$p.value

for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-t.test(x,y,paired=F)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}

points(n,p,type="l",col='green')
legend('topright',col =c("red","blue","orange","green"), pch = rep(20,1))
cbind(n,round(p,5))

set.seed(12345)
n<-5
x0<-rnorm(5,2.80919,1.2)
y0<-rnorm(5,2,1.2)

x<-x0
y<-y0
d<-cohen.d(x,y)
d
p<-t.test(x,y,paired=F)$p.value

for (i in 2:100){

```



```

x<-c(x,x0)
y<-c(y,y0)
p[i]<-t.test(x,y,paired=F)$p.value
n[i]<-5*i
cat(length(x),length(y),"\n")
}

points(n,p,type="l",col='purple')
legend('topright',c("d=0.2","d=0.5","d=0.8","d=1.0","d=2.0"),
col=c("red","blue","orange","green","purple"),pch=rep(20,1))
abline(h=0.10,lty=3,col=9)
abline(h=0.05,lty=3,col=9)
abline(h=0.01,lty=3,col=9)
abline(h=0.15,lty=3,col=9)
abline(v=25,col="gray",lty=3)
abline(v=50,col="gray",lty=3)
abline(v=100,col="gray",lty=3)
abline(v=150,col="gray",lty=3)
abline(v=200,col="gray",lty=3)
abline(v=250,col="gray",lty=3)
abline(v=300,col="gray",lty=3)

```

Listing 5.6 – Códigos das simulações do  $\delta$  de Cliff no Software R

```

library(effsize)
set.seed(12345)
x0<-runif(n=5,min = 20, max = 45)
y0<-runif(n=5,min = 30, max = 50)
n<-5

x<-x0
y<-y0

p<-wilcox.test(x,y,correct=F)$p.value
wilcox.test(x,y,correct=F,exact=T,paired=F)

delta<-cliff.delta(x,y,return.dm=T)$estimate
delta

for(i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)

```

```

p[i] <- wilcox.test(x, y, correct = F)$p.value
n[i] <- 5 * i
cat(length(x), length(y), "\n")
}

plot(n, p, type = "l", ylim = c(0, 0.7), xlim = c(0, 150), col = 2)
cbind(n, round(p, 5))

set.seed(123)
x0 <- runif(n = 5, min = 20, max = 60)
y0 <- runif(n = 5, min = 30, max = 80)
n <- 5

x <- x0
y <- y0

delta <- cliff.delta(x, y, return.dm = T)$estimate
delta

p <- wilcox.test(x, y, correct = F)$p.value
wilcox.test(x, y, correct = F, exact = T, paired = F)

for (i in 2:100){
  x <- c(x, x0)
  y <- c(y, y0)
  p[i] <- wilcox.test(x, y, correct = F)$p.value
  n[i] <- 5 * i
  cat(length(x), length(y), "\n")
}

points(n, p, type = "l", col = 'blue')
legend('topright', col = c("red", "blue"), pch = rep(20, 1))
cbind(n, round(p, 5))

set.seed(567)
y0 <- runif(n = 5, min = 20, max = 40)
x0 <- runif(n = 5, min = 20, max = 40)
n <- 5

x <- x0
y <- y0

```

---

```

p<-wilcox.test(x,y,correct= F)$p.value
wilcox.test(x,y,correct= F,exact = T, paired = F)

delta<-cliff.delta(x,y,return.dm = T)$estimate
delta

for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-wilcox.test(x,y,correct= F)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}

points(n,p,type="l",col='green')
legend('topright',col =c("red","blue","green"),pch = rep(20,1))
cbind(n,round(p,5))

set.seed(189)
y0<-runif(n=5,min = 20, max = 40)
x0<-runif(n=5,min = 20, max = 40)
n<-5

x<-x0
y<-y0

p<-wilcox.test(x,y,correct= F)$p.value
wilcox.test(x,y,correct= F,exact = T, paired = F)

delta<-cliff.delta(x,y,return.dm = T)$estimate
delta

for (i in 2:100){
  x<-c(x,x0)
  y<-c(y,y0)
  p[i]<-wilcox.test(x,y,correct= F)$p.value
  n[i]<-5*i
  cat(length(x),length(y),"\n")
}

points(n,p,type="l",col='orange')
legend('topright',c("|delta| = 0.2","|delta| = 0.36",
"|delta| = 0.52","|delta| = 0.78"),col =c("red","blue","green","orange"),
pch = rep(20,1))
cbind(n,round(p,5))

```

```
abline(h=0.10, lty=3, col=9)
abline(h=0.05, lty=3, col=9)
abline(h=0.01, lty=3, col=9)
abline(v=50, col="gray", lty=3)
abline(v=100, col="gray", lty=3)
abline(v=150, col="gray", lty=3)
abline(v=25, col="gray", lty=3)
abline(v=75, col="gray", lty=3)
abline(v=125, col="gray", lty=3)
```

Listing 5.7 – Códigos das simulações do Coeficiente  $\phi$  no Software R

```
library(DescTools)
library(rcompanion)
library(psych)

x11<-1
x12<-1
x21<-5
x22<-3

n<-x11+x12+x21+x22
n

x<-matrix(0,2,2)
x[1,1]<-x11
x[1,2]<-x12
x[2,1]<-x21
x[2,2]<-x22
x

cohenW(x)

p<-chisq.test(x)$p.value
p

for(i in 2:100){
  x[1,1]<-x[1,1]+x11
  x[1,2]<-x[1,2]+x12
  x[2,1]<-x[2,1]+x21
  x[2,2]<-x[2,2]+x22
  p[i]<-chisq.test(x)$p.value
  n[i]<-10*i
}
```

```
}

x11()
plot(n,p,type="l",ylim=c(0,1),xlim=c(0,650),col='red')
cbind(n,round(p,5))

x11<-2
x12<-0
x21<-5
x22<-3

n<-x11+x12+x21+x22
n

x<-matrix(0,2,2)
x[1,1]<-x11
x[1,2]<-x12
x[2,1]<-x21
x[2,2]<-x22
x

cohenW(x)

p<-chisq.test(x)$p.value
p

for (i in 2:100){
  x[1,1]<-x[1,1]+x11
  x[1,2]<-x[1,2]+x12
  x[2,1]<-x[2,1]+x21
  x[2,2]<-x[2,2]+x22
  p[i]<-chisq.test(x)$p.value
  n[i]<-10*i
}

Phi(x)

points(n,p,type="l",col='blue')
legend('topright',col=c("red","blue"),pch=rep(20,1))
cbind(n,round(p,5))

x11<-2
```

```
x12<-3
x21<-0
x22<-5

n<-x11+x12+x21+x22
n

x<-matrix(0,2,2)
x[1,1]<-x11
x[1,2]<-x12
x[2,1]<-x21
x[2,2]<-x22

p<-chisq.test(x)$p.value
p

for (i in 2:100){
  x[1,1]<-x[1,1]+x11
  x[1,2]<-x[1,2]+x12
  x[2,1]<-x[2,1]+x21
  x[2,2]<-x[2,2]+x22
  p[i]<-chisq.test(x)$p.value
  n[i]<-10*i
}

phi(x, digits = 4)

points(n,p,type="l",col='green')
legend('topright',col=c("red","blue","green"),pch=rep(20,1))
cbind(n,round(p,5))

x11<-2
x12<-0
x21<-1
x22<-7

n<-x11+x12+x21+x22
n

x<-matrix(0,2,2)
x[1,1]<-x11
x[1,2]<-x12
x[2,1]<-x21
x[2,2]<-x22
```

---

```
p<-chisq.test(x)$p.value
p

for (i in 2:100){
  x[1,1]<-x[1,1]+x11
  x[1,2]<-x[1,2]+x12
  x[2,1]<-x[2,1]+x21
  x[2,2]<-x[2,2]+x22
  p[i]<-chisq.test(x)$p.value
  n[i]<-10*i
}

phi(x, digits = 4)

points(n,p,type="l",col='purple')
legend('topright',c("phi=0.1021","phi=0.3273", "phi=0.5", "phi=0.6"),
col =c("red","blue","green","purple"), pch = rep(20,1))
cbind(n,round(p,5))
abline(h=0.10,lty=3,col=9)
abline(h=0.05,lty=3,col=9)
abline(h=0.01,lty=3,col=9)
abline(v=100,col="gray",lty=3)
abline(v=200,col="gray",lty=3)
abline(v=300,col="gray",lty=3)
abline(v=400,col="gray",lty=3)
abline(v=500,col="gray",lty=3)
abline(v=600,col="gray",lty=3)
abline(v=50,col="gray",lty=3)
```