



Trabalho de Conclusão de Curso

Tamanho de amostra e poder no Software R

Luana Giongo Pedrotti

27 de janeiro de 2018

Luana Giongo Pedrotti

Tamanho de amostra e poder no Software R

Trabalho de Conclusão apresentado à comissão de Graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

Orientador: Profa. Dra. Stela Maris de Jesus Castro

Porto Alegre
Janeiro de 2018

Luana Giongo Pedrotti

Tamanho de amostra e poder no Software R

Este Trabalho foi julgado adequado para obtenção dos créditos da disciplina Trabalho de Conclusão de Curso em Estatística e aprovado em sua forma final pelo Orientador e pela Banca Examinadora.

Orientador: _____
Profa. Dra. Stela Maris de Jezus Castro, UFRGS
Doutora pela Universidade Federal do Rio Grande do Sul, Porto Alegre, RS

Banca Examinadora:

Profa. Dra. Luciana Neves Nunes, UFRGS
Doutora pela Universidade Federal do Rio Grande do Sul - Porto Alegre, RS

Porto Alegre
Janeiro de 2018

Agradecimentos

Agradeço a todas as pessoas que de uma forma ou outra acompanharam minha trajetória e meu aprendizado na graduação em Estatística e que torceram por mim para concluir essa etapa da minha vida.

Em especial aos meus pais, Neiva e Maurício, agradeço por entender os meus momentos de ausência, por sempre me apoiarem nas minhas escolhas e pelo incentivo incondicional para eu chegar até aqui.

Ao meu namorado e companheiro de todas as horas, Matias, obrigada pelo incentivo, suporte e carinho de todos esses anos juntos.

À professora de matemática do Ensino Médio, Glória, obrigada por me apresentar a Estatística e me encorajar a entrar neste curso.

À professora Stela, por todo aprendizado, motivação e auxílio na orientação desta monografia.

Obrigada por tudo, sem vocês nada disso seria possível!

Resumo

Esta monografia tem como principal objetivo produzir um material prático e acessível a todos os pesquisadores sobre cálculo de tamanho de amostra e poder no Software R a fim de servir como auxílio em monografias, artigos ou trabalhos científicos, e apresentar paralelamente exemplos aplicados em diversas áreas da estatística.

A estrutura deste trabalho se dá em apresentar os pacotes até então conhecidos de tamanho de amostra e poder, seus autores e colaboradores, assim como, o procedimento para o funcionamento de cada função em cada pacote estatístico e sua aplicação em exemplos práticos.

Entre as principais técnicas para as quais são apresentadas funções para tais cálculos estão: estimação de um parâmetro, teste de hipóteses para uma e duas amostras, teste de associação, análise de variância (ANOVA) e correlação, aplicadas nos seguintes pacotes do R: *samplingbook*, *pwr*, *pwr2* e *TrialSize*.

Palavras-Chave: Tamanho de amostra, Poder, Software R.

Abstract

This monograph main target is to provide a practical and accessible material to all researchers on how to obtain sample size and power on the free software R, in order to work as an aid in monographs, articles and scientific works, and also present examples applied in several statistical areas.

The monograph is structured as follows: presentation of the packages that work with sample size and power known until now, their authors and contributors and, mainly, how to proceed to utilize the functions in each package followed by applications in practical examples.

Among the main techniques for which these functions are presented are: one-parameter estimation, one- and two-sample hypothesis tests, association test, analysis of variance (ANOVA) and correlation, applied in the following R packages: *samplingbook*, *pwr*, *pwr2* and *TrialSize*.

Keywords: Sample Size, Power, R Software.

Sumário

1	Introdução	9
2	Desenvolvimento	11
2.1	Estimação de Parâmetros	14
2.1.1	Média	14
2.1.1.1	Exemplo	14
2.1.1.2	Exemplo aplicado no pacote <i>samplingbook</i>	14
2.1.2	Proporção	15
2.1.2.1	Exemplo	16
2.1.2.2	Exemplo aplicado no pacote <i>samplingbook</i>	16
2.2	Teste de Hipóteses para uma amostra	17
2.2.1	Para uma média	17
2.2.1.1	Exemplos	17
2.2.1.2	Exemplo aplicado no pacote <i>TrialSize</i>	18
2.2.1.3	Exemplo aplicado no pacote <i>pwr</i>	18
2.2.2	Para uma proporção	20
2.2.2.1	Exemplos	20
2.2.2.2	Exemplo aplicado no pacote <i>TrialSize</i>	21
2.2.2.3	Exemplo aplicado no pacote <i>pwr</i>	22
2.2.3	Equivalência	24
2.2.3.1	Para uma média	24
2.2.3.1.1	Exemplos	25
2.2.3.1.2	Exemplo aplicado no pacote <i>TrialSize</i>	25
2.2.3.2	Para uma proporção	26
2.2.3.2.1	Exemplos	27
2.2.3.2.2	Exemplo aplicado no pacote <i>TrialSize</i>	27
2.3	Teste de Hipóteses para duas amostras	29
2.3.1	Comparação entre duas médias de grupos independentes	29
2.3.1.1	Exemplo	29
2.3.1.2	Exemplo aplicado no pacote <i>TrialSize</i>	29
2.3.1.3	Exemplo aplicado no pacote <i>pwr</i>	30
2.3.2	Comparação entre duas proporções de grupos independentes	32
2.3.2.1	Exemplos	32
2.3.2.2	Exemplo aplicado no pacote <i>TrialSize</i>	32
2.3.2.3	Exemplo aplicado no pacote <i>pwr</i>	33
2.3.3	Equivalência	36

2.3.3.1	Para duas médias de grupos independentes	36
2.3.3.1.1	Exemplos	37
2.3.3.1.2	Exemplo aplicado no pacote <i>TrialSize</i>	37
2.3.3.2	Para duas proporções de grupos independentes	38
2.3.3.2.1	Exemplos	39
2.3.3.2.2	Exemplo aplicado no pacote <i>TrialSize</i>	39
2.3.4	Comparação entre duas médias de grupos pareados	41
2.3.4.1	Exemplo	41
2.3.4.2	Exemplo aplicado no pacote <i>pwr</i>	41
2.4	Análise de Variância (ANOVA)	43
2.4.1	Exemplo	43
2.4.2	Exemplo aplicado no pacote <i>pwr</i>	43
2.4.3	Exemplo aplicado no pacote <i>pwr2</i>	44
2.5	Teste de Associação para variáveis categóricas	45
2.5.1	Exemplo	46
2.5.2	Exemplo aplicado no pacote <i>pwr</i>	46
2.6	Correlação	47
2.6.1	Exemplo	47
2.6.2	Exemplo aplicado no pacote <i>pwr</i>	48
3	Considerações Finais	50
4	Pacotes do R	52
4.1	<i>samplingbook</i>	52
4.2	<i>pwr</i>	52
4.3	<i>pwr2</i>	52
4.4	<i>TrialSize</i>	53
	Referências Bibliográficas	53

1 Introdução

Muitas vezes, trabalhos científicos e pesquisas em geral são estruturados a partir de um planejamento inicial que deve conter além de um embasamento teórico sobre o assunto previsto, métodos e técnicas estatísticas a serem utilizadas, quando necessário.

Uma das ferramentas estatísticas mais importantes para o planejamento de uma pesquisa é a amostragem. Definir um plano amostral passa por cumprir um conjunto de etapas, desde a forma em que os dados serão coletados, podendo ser probabilística ou não, até a quantidade de unidades elementares - que é o objeto ou entidade portadora das informações que pretende-se coletar (Bolfarine e Bussab, 2005) - que chamamos de tamanho de amostra, cuja notação tradicional é n .

É importante frisar que todos os cálculos para tamanho de amostra apresentados nesta monografia referem-se à amostragem probabilística, mais especificamente do tipo amostragem aleatória simples, no qual toda amostra possível de mesmo tamanho tem a mesma probabilidade de ser selecionada a partir da população.

A partir da definição do tamanho amostral é possível delimitar a viabilidade do projeto, a fim de evitar desperdício de recursos (seja humanos e/ou financeiros) e tempo, assim como, falta de precisão dos resultados encontrados, garantindo assim conclusões confiáveis.

A não utilização de um cálculo de tamanho de amostra adequado pode levar a amostras mal planejadas, podendo resultar em: amostras menores do que o necessário que podem não produzir uma resposta definitiva e permitir que diferenças importantes passem despercebidas ou amostras exageradamente grandes que podem fazer com que diferenças irrelevantes sejam estatisticamente significantes (Tavares, 2008).

Apesar da reconhecida importância de se obter um tamanho de amostra suficiente para um bom resultado no seu trabalho, nem sempre o pesquisador faz uso dessa prática, sendo, muitas vezes, deixada de lado no planejamento da pesquisa, seja devido a problemas de compreensão da teoria ou falta de referências de fácil acesso.

Existem vários softwares gratuitos que possuem implementadas rotinas para o cálculo de tamanho de amostra e poder. Dentre os gratuitos pode-se citar o WinPepi (Abramson et al., 1993) (composto por 7 programas e mais de 120 módulos, além de um manual totalmente referenciado), o R (R Core Team, 2017) (composto por inúmeros pacotes) e, também, uma ferramenta online disponibilizada pelo Laboratório

de Epidemiologia e Estatística - Lee (USP, 1995), que realiza cálculos para tamanho de amostra e dispõe de referências bibliográficas e um glossário com os principais conceitos básicos.

O software R, além de ser livre, tem a vantagem de estar em contínuo desenvolvimento ou atualização, pois seus desenvolvedores e/ou usuários estão sempre agregando mais conteúdos a medida em que vão sendo descobertas novas técnicas na área da estatística. Além disso, é de fácil acesso e conta com uma ampla rede de informações e materiais na *Internet*.

Em vista de tudo isso, o objetivo deste trabalho é congregiar formas de cálculo de tamanho de amostra e poder disponíveis em diversos pacotes do Software R, bem como ser um guia prático de consulta para os pesquisadores, incentivando a prática deste tipo de cálculo no planejamento de uma pesquisa.

Cabe salientar que esta monografia não tem o intuito de ensinar estatística, ou seja, supõe-se que o leitor a quem se destina este material já saiba os objetivos do projeto de pesquisa, as hipóteses de pesquisa e conseqüentemente que técnica estatística deverá aplicar.

Esta monografia será estruturada em três seções. Primeiramente será apresentado um referencial teórico de cada técnica a ser utilizada, revisando conceitos essenciais. Na próxima etapa será apresentado um exemplo que posteriormente será aplicado e discutido no Software R. Vale ressaltar que todos os exemplos são meramente ilustrativos, ou seja, dados fictícios criados apenas com o intuito didático.

Todos os pacotes apresentados nesta monografia estão disponíveis em <https://cran.r-project.org/>, assim como no capítulo 4 são apresentados os criadores, desenvolvedores e editores dos pacotes e o *link* para acesso direto ao pacote citado. Para suporte em relação a instalação do R, recomenda-se o tutorial do próprio software, disponível em: <https://cran.r-project.org/doc/manuals/r-release/R-admin.html#Installing-R-under-Windows>.

2 Desenvolvimento

Como já foi dito anteriormente é fundamental um bom planejamento do estudo para podermos obter resultados confiáveis. Além disso, há algumas atividades imprescindíveis que devem ser realizadas nos levantamentos por amostragem. Iniciando com a formulação do problema e de maneira conjunta, a definição dos objetivos e as hipóteses da pesquisa (mesmo que de forma preliminar), seguindo com a definição da população de estudo e as possíveis variáveis e/ou características que serão observadas e analisadas.

Um aspecto fundamental de um estudo por amostragem é a seleção da amostra. Métodos de amostragem probabilística devem ser utilizados, caso se deseje obter resultados fidedignos (Silva, 1998). Outros critérios de elegibilidade, como por exemplo, critérios de inclusão e/ou exclusão, também devem ser considerados para selecionar os sujeitos que farão parte do estudo. É necessário estabelecê-los e cumpri-los a fim de não interferir na qualidade dos dados, assim como nas futuras inferências.

A ideia principal de fazer inferências é que, a partir dos dados amostrais, é possível tirar conclusões e fazer generalizações sobre uma população, da qual a amostra foi retirada. As duas principais áreas de inferência são: inferência clássica e inferência bayesiana. Nesta monografia trataremos apenas da primeira, a inferência clássica.

As inferências estatísticas a serem feitas a partir do resultado dos dados amostrais se dividem em dois grandes tópicos: estimação de parâmetros e teste de hipóteses.

Parâmetros são valores desconhecidos relativos à população de interesse que resumem a informação relativa a uma variável e, considerando que o seu verdadeiro valor apenas será conhecido ao ser realizado um censo (isto é, estudar todos os elementos da população), geralmente é de grande interesse estimá-los a partir de uma amostra. Este procedimento se denomina, como o nome já diz, estimação de parâmetros.

Há duas formas de estimar um parâmetro: por ponto, o qual fornece apenas um único valor da estimativa; ou por intervalo, o qual fornece um intervalo denominado intervalo de confiança, usualmente chamado de IC.

O intervalo de confiança (IC) pode ser entendido como um intervalo de valores plausíveis para o parâmetro, obtidos com determinado nível ou grau de confiança, usualmente denotado por $100 \times (1 - \alpha)\%$, sendo mais populares os níveis 95% ou 99% de confiança. De forma geral, quando se tem 95% de confiança significa que dos intervalos de confiança construídos a partir das amostras de mesmo tamanho,

95% deles contêm o verdadeiro valor do parâmetro.

Os parâmetros de interesse a serem estimados dependem do tipo de variável que está sendo estudada, uma vez que são recomendados procedimentos estatísticos diferentes em cada situação. A principal divisão ocorre entre variáveis quantitativas e qualitativas. O principal parâmetro de interesse das variáveis quantitativas, em geral, é a média e, das variáveis qualitativas (dicotômicas ou categóricas), é a proporção de ocorrência de uma de suas categorias de resposta.

Ao fazer inferências se está sujeito a cometer erros amostrais. O erro amostral é inerente ao processo de amostragem e se refere à variabilidade que pode ocorrer de amostra para amostra. Não se pode evitar a ocorrência deste erro, porém pode-se limitar seu valor por meio de um dimensionamento adequado da amostra. Este limite é o que se denomina erro máximo aceitável, tratado nesta monografia como erro máximo ou também margem de erro.

A outra forma de inferência estatística é o teste de hipóteses. Neste procedimento, o pesquisador não está interessado em conhecer a magnitude de um parâmetro, mas sim avaliar uma hipótese sobre a população, ou seja, a partir dos dados amostrais, verificar a veracidade de uma hipótese feita sobre a(s) população(ões) envolvida(s). Por meio desta técnica é possível rejeitar ou não uma hipótese sobre um parâmetro de interesse, associando à conclusão uma probabilidade máxima de erro ([Callegari-Jacques, 2003](#)). Hipóteses estatísticas são suposições ou afirmações sobre o parâmetro que o pesquisador deseja testar.

Há duas hipóteses estatísticas:

- Hipótese nula ou de nulidade (H_0 – leia-se h zero): estabelece a ausência de diferença entre os parâmetros de diferentes grupos ou entre o parâmetro de um grupo e uma medida qualquer.
- Hipótese alternativa (H_1 – leia-se h um): é a hipótese complementar à hipótese nula. Em muitas situações de pesquisa, é a que o pesquisador quer ver confirmada.

Dentre as maneiras de apresentar a hipótese alternativa as mais usuais são, diferente (\neq), maior ($>$) ou menor ($<$) que algum valor. A natureza do estudo vai definir como deve ser formulada e qual opção escolher.

Por exemplo, θ é o parâmetro a ser testado:

$$H_0 : \theta = \theta_0 \tag{2.1}$$

$$\left\{ \begin{array}{l} H_1 : \theta \neq \theta_0, \\ H_1 : \theta > \theta_0, \\ H_1 : \theta < \theta_0. \end{array} \right. \tag{2.2}$$

$$\tag{2.3}$$

$$\tag{2.4}$$

Em [2.2](#), temos que o teste é bilateral ou bicaudal. Em [2.3](#), temos que o teste é unilateral ou unicaudal à direita (ou superior) e em [2.4](#), unilateral ou unicaudal à esquerda (ou inferior).

Ao realizar um teste de hipóteses podemos tomar uma de duas decisões estatísticas: rejeitar H_0 ou aceitar (não rejeitar) H_0 . Devido a isso, estamos sujeitos a dois tipos de erros:

- O “erro do tipo I” que é definido como “rejeitar a hipótese de nulidade (H_0) quando a mesma for verdadeira”, e
- O “erro do tipo II” que é definido como “não rejeitar hipótese de nulidade (H_0) quando esta for falsa”.

Outra definição importante é o nível de significância (ou região de significância), conhecido como alfa (α), determinado como a probabilidade máxima de cometer o erro do tipo I. Já beta (β) é a probabilidade máxima de cometer o erro do tipo II.

Por via de regra, usualmente é usado $1-\beta$ (complementar do erro tipo II), conhecido como o poder estatístico de um teste (que por simplificação será chamado apenas de *poder*), correspondente à probabilidade de rejeitar a hipótese de nulidade (H_0) quando a mesma é falsa. Em outras palavras, é a probabilidade de detectar um efeito (uma diferença) quando ele (ela) realmente existir. Sendo assim, este deve ser suficiente para o objetivo desejado de detectar diferenças.

Essa diferença é também chamada de magnitude do efeito ou tamanho de efeito. Normalmente é estabelecida a partir de um dos seguintes métodos: uma estimativa do pesquisador; um estudo piloto, no qual o pesquisador conduz um estudo pequeno e determina o valor dessa diferença; ou então a partir da literatura, através de um estudo semelhante. Também é possível calcular essa diferença de diversas maneiras para diferentes tipos de estudos, como pode ser visto em [Cohen \(1988\)](#).

Segundo [Cohen \(1988\)](#) um nível mínimo aceitável para o poder é de 0,80, ou seja, 80%. Porém, vale ressaltar que quando o objetivo de um estudo é provar equivalência (neste caso, o pesquisador deseja aceitar H_0), o poder deve ser mais alto, assim havendo uma menor probabilidade de errar (menor probabilidade de erro tipo II).

Em geral, fixa-se um valor do poder, frequentemente 80%, 85%, 90%, 95% ou até mesmo 99% ([Callegari-Jacques, 2003](#)). Um poder alto é sempre desejável, porém aumentar o poder mantendo um nível de significância baixo acarreta num aumento acentuado no tamanho da amostra, sendo ideal encontrar um equilíbrio entre os dois tipos de erro, para não tornar o estudo inviável.

Uma forma de avaliar a significância de um teste é analisar pelo nível descritivo, mais conhecido como *p-valor*, que é uma maneira mais pragmática e ágil de interpretar os testes, já que não é necessário comparar com valores tabelados. Entretanto compare-se o *p-valor* com um α fixado (por exemplo, H_0 é rejeitada se *p-valor* for menor ou igual que α).

Vale ressaltar que, nessa monografia, só serão vistos testes paramétricos. Para casos em que se faz necessária a utilização de procedimentos não-paramétricos, aconselha-se a leitura de [Castro e Souza \(2012\)](#), onde são apresentadas formas de cálculo de tamanho de amostra e poder para alguns dos principais testes não-paramétricos.

Outro aspecto a considerar é que na área da Saúde, são realizados testes de equivalência farmacêutica, os quais comprovam que os medicamentos genéricos possuem os mesmos fármacos, na mesma dosagem e na mesma forma farmacêutica encontradas

em medicamentos de referência, além de cumprirem com as mesmas especificações físicas e físico-químicas delimitadas pelo controle de qualidade.

Também são realizados testes (ou ensaios) de bioequivalência que segundo a Agência Nacional de Vigilância Sanitária (ANVISA), consiste na demonstração de equivalência farmacêutica entre produtos apresentados quando estudados sob um mesmo delineamento experimental.

Em algumas situações há a necessidade de se trabalhar com uma transformação das medidas farmacocinéticas, sendo a logarítmica a mais usada, porém esse tópico não será abordado nessa monografia. Para mais detalhamento sobre esses testes e afins recomenda-se a leitura de [Guewehr \(2004\)](#).

2.1 Estimação de Parâmetros

2.1.1 Média

Um estudo pode ter como objetivo estimar o valor médio (μ) de uma variável quantitativa referente à população de interesse. Ao calcular o tamanho da amostra para este objetivo, estaremos supondo que a variável que contém a resposta de interesse segue uma distribuição normal, com uma determinada média, representada por μ e uma determinada variância, representada por σ^2 ([Zar, 1999](#)).

Para os casos onde a variância é supostamente conhecida, utiliza-se o valor tabelado de uma distribuição Normal Padrão no cálculo do tamanho da amostra. Caso contrário, é necessário estimar a variância por meio da variância amostral, o que geralmente é mais comum, sendo assim necessário usar o valor tabelado da distribuição *t de Student*.

2.1.1.1 Exemplo

Um pesquisador deseja estimar a pressão arterial sistólica média de enfermeiras no hospital de sua região. Um estudo realizado em adultos no Rio Grande do Sul mostrou que a pressão arterial sistólica média é de 128mmHg, com desvio padrão de 24mmHg. Consideremos um erro máximo de 3mmHg e um nível de confiança de 95%.

Obs.: Pensemos agora que o tamanho da população seja finito. No caso do exemplo, suponhamos que se totaliza 120, o número total de mulheres enfermeiras que trabalham no hospital.

2.1.1.2 Exemplo aplicado no pacote *samplingbook*

No pacote *samplingbook* existe a função *sample.size.mean* para calcular o tamanho da amostra para estimação de uma média considerando ou não uma correção para

população finita.

Para a função `sample.size.mean` são necessários os seguintes argumentos:

- **e**, número positivo que especifica a precisão, que é metade da amplitude do intervalo de confiança (erro máximo).
- **S**, desvio padrão da variável de interesse na população.
- **N**, tamanho da população. *Default* é `N=Inf`, que significa que os cálculos se dão sem correção para população finita.
- **level**, nível de confiança. *Default* é `level=0.95`.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do exemplo:

```
> n=sample.size.mean(e=3, S=24, N=Inf, level=0.95)
> n

sample.size.mean object: Sample size for mean estimate
without finite population correction: N=Inf, precision e=3 and standard deviation S=24

sample size needed: 246

> n=sample.size.mean(e=3, S=24, N=120, level=0.95)
> n

sample.size.mean object: Sample size for mean estimate
with finite population correction: N=120, precision e=3 and standard deviation S=24

sample size needed: 81
```

Para o exemplo acima, o valor que retorna para o tamanho estimado da amostra, quando `N=Inf` é **n=246 enfermeiras**, porém quando especificamos `N=120`, o valor que retorna é **n=81 enfermeiras**.

2.1.2 Proporção

No caso de estudos em que o objetivo seja alcançado através de variáveis categóricas, o parâmetro de interesse é a proporção de ocorrência das categorias de resposta destas variáveis. No caso de se estimar uma proporção, o motivo principal de se calcular o tamanho da amostra é garantir uma determinada precisão na estimativa que será obtida ([Agranonik e Hirakata, 2011](#)).

Para se calcular o tamanho amostral nesta técnica, além do nível de confiança, do erro máximo aceitável (e quando conhecido, o tamanho da população), é necessária uma suposta proporção de ocorrência da categoria de interesse da variável estudada que pode ser encontrada via estudo piloto ou numa busca prévia na literatura. Esta suposta proporção vai definir a variabilidade do estimador do parâmetro proporção que se pretende estimar.

2.1.2.1 Exemplo

Uma engenheira de qualidade deseja conhecer a proporção de peças defeituosas de sua fábrica de peças automotivas na produção de uma semana. Supondo um nível de confiança de 95% e margem de erro de 0,05. Como a engenheira baseou-se em estudos passados da fábrica utilizou-se uma proporção de 0,1 de peças defeituosas, ou seja, a cada 100 peças produzidas, 10 são defeituosas.

Obs.: Pensemos agora que a engenheira diga que são produzidas 1500 peças por semana.

2.1.2.2 Exemplo aplicado no pacote *samplingbook*

No pacote *samplingbook* existe a função *sample.size.prop* para calcular o tamanho da amostra para estimação de uma proporção considerando ou não uma correção para população finita.

Para a função *sample.size.prop* são necessários os seguintes argumentos:

- **e**, número positivo que especifica a precisão, que é metade da amplitude do intervalo de confiança (erro máximo).
- **P¹**, proporção esperada para os eventos, é um valor entre 0 e 1. *Default* é P=0.5.
- **N**, tamanho da população. *Default* é N=Inf, que significa que os cálculos se dão sem correção para população finita.
- **level**, nível de confiança. *Default* é level=0.95.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do exemplo:

```
> n=sample.size.prop(e=0.05, P=0.1, N=Inf, level=0.95)
> n

sample.size.prop object: Sample size for proportion estimate
without finite population correction: N=Inf, precision e=0.05 and expected proportion P=0.1

sample size needed: 139

> n=sample.size.prop(e=0.05, P=0.1, N=1500, level=0.95)
> n

sample.size.prop object: Sample size for proportion estimate
with finite population correction: N=1500, precision e=0.05 and expected proportion P=0.1

sample size needed: 127
```

¹Quando P=0,5 assume-se o máximo de variabilidade, assim obtém-se o maior tamanho da amostra possível, mantendo os outros parâmetros fixos.

Para o exemplo acima, o valor que retorna para o tamanho estimado da amostra, quando $N = \text{Inf}$ é **$n = 139$ peças**, porém quando especificamos $N = 1500$, o valor que retorna é **$n = 127$ peças**.

2.2 Teste de Hipóteses para uma amostra

2.2.1 Para uma média

Essa técnica é utilizada quando se deseja comparar a média obtida de uma amostra em relação um valor já estabelecido, podendo ser um valor de referência ou um valor histórico (μ_0).

As hipóteses de interesse neste contexto, para um teste bilateral, são:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Já para um teste unilateral, são:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0 \text{ ou } H_1 : \mu < \mu_0$$

Ao utilizar essa técnica supõe-se que a variável resposta, a qual se quer testar, siga uma distribuição normal com desvio padrão conhecido.

2.2.1.1 Exemplos

Exemplo 1: Uma fábrica de café gostaria de verificar se a máquina nova está moendo uma quantidade significativamente diferente do valor médio por saco especificado pelo fornecedor que é de 50,5kg com desvio padrão de 1kg. Considera-se um nível de significância de 5% e poder 80% para detectar uma diferença mínima de 0,4 kg.

As hipóteses nula e alternativa são:

$$H_0 : \mu = 50,5\text{Kg}$$

$$H_1 : \mu \neq 50,5\text{Kg}$$

Exemplo 2: Supondo mesmo dados do exemplo 1, porém deseja-se testar se uma média é maior que 50,5 kg.

As hipóteses nula e alternativa são:

$$H_0 : \mu = 50,5\text{Kg}$$

$$H_1 : \mu > 50,5\text{Kg}$$

2.2.1.2 Exemplo aplicado no pacote *TrialSize*

No pacote *TrialSize* existe a função *OneSampleMean.Equality* para calcular o tamanho de amostra para esta situação.

Para a função *OneSampleMean.Equality* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **sigma**, desvio padrão da variável de interesse X.
- **margin**, matematicamente, $margin = \bar{x} - \mu_0$, ou seja, é a diferença entre \bar{x} (média amostral) e um valor de referência μ_0 . Porém na prática, como ainda não se tem a amostra, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 1**:

```
> n=OneSampleMean.Equality(alpha=0.05, beta=0.2, sigma=1, margin=0.4)
> n
[1] 49.0555
```

O valor que retorna para o tamanho estimado da amostra é $n=49,06$, ou seja, arredondando para cima, **$n=50$ sacos de café**.

2.2.1.3 Exemplo aplicado no pacote *pwr*

No pacote *pwr* existe a função *pwr.t.test* para calcular o tamanho da amostra e poder para testar as mesmas hipóteses.

Obs.: Cabe ressaltar que para as funções deste pacote é necessário passar um dos argumentos como *NULL*, ou seja, nulo. Quando se deseja calcular o tamanho amostral, passa-se $n = NULL$, caso estejamos interessados no poder, passa-se $power = NULL$.

Para a função *pwr.t.test* são necessários os seguintes argumentos:

- **n**, número de observações (por amostra).
- **d**, magnitude do efeito (tamanho de efeito).
- **sig.level**, nível de significância (Probabilidade do erro tipo I).
- **power**, poder (1-Probabilidade do erro tipo II).
- **type**, definição do teste *t de Student*: para uma amostra (“one.sample”), duas amostras independentes (“two.sample”) ou duas amostras pareadas (“paired”).

- **alternative**, especifica do tipo de teste de acordo com a hipótese alternativa, com as opções: “two.sided” (bilateral), “greater” (unilateral superior) ou “less” (unilateral inferior). *Default* é `alternative=“two.sided”`.

Tamanho de efeito:

Sabendo que este pacote é baseado nos métodos de [Cohen \(1988\)](#), a magnitude do efeito (tamanho de efeito) d é calculada da seguinte maneira:

$$d = \frac{\mu_1 - \mu_0}{\sigma}, \quad (2.5)$$

onde μ_0 é a média de referência, μ_1 é a média de referência somada com a diferença mínima a ser detectada e σ é o desvio padrão. Portanto para o **exemplo 1**, tem-se:

$$d = \frac{50.9 - 50.5}{1} = 0.4. \quad (2.6)$$

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 1**:

```
> n=pwr.t.test(n=NULL, d=0.4, sig.level=0.05, power=0.8, type="one.sample", alternative="two.sided")
> n

one-sample t test power calculation

      n = 51.00945
      d = 0.4
sig.level = 0.05
  power = 0.8
alternative = two.sided
```

O valor que retorna para o tamanho estimado da amostra é **n=51 sacos de café**.

Poder:

Suponhamos que gostaríamos de calcular o poder, tendo um tamanho de amostra igual a $n=74$. A função deve ser utilizada da seguinte forma para o **exemplo 1**:

```
> poder=pwr.t.test(n=74, d=0.4, sig.level=0.05, power=NULL, type="one.sample", alternative="two.sided")
> poder

one-sample t test power calculation

      n = 74
      d = 0.4
sig.level = 0.05
  power = 0.924407
alternative = two.sided
```

O valor que retorna para o poder é de **0,92**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,92.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 2**:

```
> n=pwr.t.test(n=NULL, d=0.4, sig.level=0.05, power=0.8, type="one.sample", alternative="greater")
> n

One-sample t test power calculation

      n = 40.02908
      d = 0.4
sig.level = 0.05
  power = 0.8
alternative = greater
```

O valor que retorna para o tamanho estimado da amostra é $n=40,03$, ou seja, arredondando para cima, **$n=41$ sacos de café**.

Poder:

Suponhamos que gostaríamos de calcular o poder, tendo um tamanho de amostra igual a $n=74$. A função deve ser utilizada da seguinte forma para o **exemplo 2**:

```
> poder=pwr.t.test(n=74, d=0.4, sig.level=0.05, power=NULL, type="one.sample", alternative="greater")
> poder

one-sample t test power calculation

      n = 74
      d = 0.4
sig.level = 0.05
  power = 0.9611276
alternative = greater
```

O valor que retorna para o poder é de **0,96**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,96.

2.2.2 Para uma proporção

Essa técnica é utilizada quando se deseja comparar a proporção obtida de uma amostra contra um valor já estabelecido, podendo ser, por exemplo, um valor de referência, um valor histórico ou a prevalência de uma doença (P_0).

As hipóteses de interesse neste contexto, para um teste bilateral, são:

$$H_0 : P = P_0$$

$$H_1 : P \neq P_0$$

Já para um teste unilateral, são:

$$H_0 : P = P_0$$

$$H_1 : P > P_0 \text{ ou } H_1 : P < P_0$$

2.2.2.1 Exemplos

Exemplo 1: Um hospital está interessado na proporção de melhora dos pacientes devido a um tratamento com novo medicamento que tem menor custo. Sabe-se que a diferença mínima detectável desejada é de 0,075. Considerando 5% de significância

e 80% de poder, quantos pacientes o hospital precisa para atestar que a taxa de melhora é diferente a do antigo medicamento que é de 0,7?

As hipóteses nula e alternativa são:

$$H_0 : P = 0,7$$

$$H_1 : P \neq 0,7$$

Exemplo 2: Supondo os mesmos dados do exemplo 1, porém deseja-se testar se a proporção é maior que 0,7.

As hipóteses nula e alternativa são:

$$H_0 : P = 0,7$$

$$H_1 : P > 0,7$$

2.2.2.2 Exemplo aplicado no pacote *TrialSize*

No pacote *TrialSize* existe a função *OneSampleProportion.Equality* para calcular o tamanho de amostra para esta situação.

Para a função *OneSampleProportion.Equality* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **p**, a proporção verdadeira da categoria de interesse da variável estudada na população.
- **delta**, matematicamente, $delta = P - P_0$, ou seja, é a diferença entre P (verdadeira proporção) e um valor de referência P_0 . Porém na prática, como ainda não se tem a amostra, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 1**:

```
> n=OneSampleProportion.Equality(alpha=0.05, beta=0.2, p=0.7,delta=0.075)
> n
[1] 293.0248
```

O valor que retorna para o tamanho estimado da amostra é $n=293,02$, ou seja, arredondando para cima, **n=294 pacientes**.

2.2.2.3 Exemplo aplicado no pacote *pwr*

No pacote *pwr* existe a função *pwr.p.test* para calcular o tamanho da amostra e poder para esta situação. Porém, para proceder com a utilização desta função faz-se necessário o cálculo do tamanho de efeito. Para tal, deve-se usar a função *ES.h*, encontrada neste mesmo pacote, que calcula a magnitude do efeito (tamanho de efeito) entre duas proporções.

Sabendo que a função *ES.h* é baseada nos métodos de [Cohen \(1988\)](#), a magnitude do efeito (tamanho de efeito) h é calculada da seguinte maneira:

$$h = \phi_1 - \phi_2, \text{ para hipótese unilateral} \quad (2.7)$$

ou

$$h = |\phi_1 - \phi_2|, \text{ para hipótese bilateral} \quad (2.8)$$

onde $\phi_1 = 2 \times \arcsin \sqrt{p_1}$, $\phi_2 = 2 \times \arcsin \sqrt{p_2}$ e p_1 e p_2 são a primeira e a segunda proporção, respectivamente.

Para a função *ES.h* são necessários os seguintes argumentos:

- **p1**, primeira proporção.
- **p2**, segunda proporção.

Tamanho de efeito:

Essa função é utilizada da seguinte forma no caso do **exemplo 1**:

```
> h=ES.h(0.775,0.7)
> h
[1] 0.1708474
```

A primeira proporção é a soma da proporção de referência (0,7) com a diferença mínima que se deseja detectar (0,075), resultando em 0,775. A segunda proporção é o próprio valor de referência (0,7).

Então, o valor que retorna para o tamanho de efeito é **h=0.1708**.

A partir de então usaremos o resultado de h no argumento do tamanho de efeito na função *pwr.p.test* tanto para o exemplo 1 como para o exemplo 2. E por isso escreve-se **h=h** no código do R.

Para a função *pwr.p.test* são necessários os seguintes argumentos:

- **h**, magnitude do efeito (tamanho de efeito).
- **n**, número de observações.
- **sig.level**, nível de significância (Probabilidade do erro tipo I).
- **power**, poder do teste (1-Probabilidade do erro tipo II).

- **alternative**, especifica o tipo de teste de acordo com a hipótese alternativa, com as opções: “two.sided” (bilateral), “greater” (unilateral superior) ou “less” (unilateral inferior). *Default* é `alternative=“two.sided”`.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 1**:

```
> n=pwr.p.test(h=h, n=NULL, sig.level=0.05, power=0.8, alternative="two.sided")
> n

proportion power calculation for binomial distribution (arcsine transformation)

      h = 0.1708474
      n = 268.8994
sig.level = 0.05
  power = 0.8
alternative = two.sided
```

O valor que retorna para o tamanho estimado da amostra é $n=268,90$, ou seja, arredondando para cima, **$n=269$ pacientes**.

Poder:

Suponhamos que gostaríamos de calcular o poder, dado o tamanho de amostra igual a $n=200$, temos que a função é utilizada da seguinte forma para o **exemplo 1**:

```
> poder=pwr.p.test(h=h, n=200, sig.level=0.05, power=NULL, alternative="two.sided")
> poder

proportion power calculation for binomial distribution (arcsine transformation)

      h = 0.1708474
      n = 200
sig.level = 0.05
  power = 0.6758768
alternative = two.sided
```

O valor que retorna para o poder é de **0,676**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,676.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 2**:

```
> n=pwr.p.test(h=h, n=NULL, sig.level=0.05, power=0.8, alternative="greater")
> n

proportion power calculation for binomial distribution (arcsine transformation)

      h = 0.1708474
      n = 211.8124
sig.level = 0.05
  power = 0.8
alternative = greater
```

O valor que retorna para o tamanho estimado da amostra é $n=211,81$, ou seja, arredondando para cima, **$n=212$ pacientes**.

Poder:

Suponhamos que gostaríamos de calcular o poder, dado o tamanho de amostra igual a $n=200$, temos que a função é utilizada da seguinte forma para o **exemplo 2**:

```
> poder=pwr.p.test(h=h, n=200, sig.level=0.05, power=NULL, alternative="greater")
> poder

proportion power calculation for binomial distribution (arcsine transformation)

      h = 0.1708474
      n = 200
sig.level = 0.05
  power = 0.7797335
alternative = greater
```

O valor que retorna para o poder é de **0,78**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,78.

2.2.3 Equivalência

É um tipo de análise para determinar se é possível afirmar que a diferença entre a média ou a proporção estimada da população e do valor-alvo está dentro de uma margem denominada intervalo de equivalência.

Ao utilizar um teste de equivalência para uma amostra, deve especificar um intervalo de valores que esteja “perto o suficiente” para ser considerado equivalente ao alvo. Este intervalo de equivalência, também chamado de zona de equivalência, é baseado em algum conhecimento prévio, por exemplo, da literatura, e deve ser determinado antes de realizar o teste. (Minitab, 2017)

2.2.3.1 Para uma média

Essa técnica é utilizada quando se deseja avaliar se a média de uma variável na população está suficientemente próxima de um valor-alvo (ou um valor de referência) a ponto de ser considerada equivalente.

As hipóteses nula e alternativa para testar a equivalência são:

$$H_0 : |\bar{x} - \mu_0| \geq \text{delta}$$

$$H_1 : |\bar{x} - \mu_0| < \text{delta}$$

As hipóteses nula e alternativa para testar a superioridade e não-inferioridade são:

$$H_0 : \bar{x} - \mu_0 \leq \text{delta}$$

$$H_1 : \bar{x} - \mu_0 > \text{delta}$$

2.2.3.1.1 Exemplos -

Exemplo 1: Um analista do setor de controle de qualidade quer determinar se a força média necessária para abrir vedações de sacos de salgadinhos é equivalente ao valor de referência de 4,2 Newtons, sendo que se quer detectar uma diferença mínima de 0,1 Newtons com desvio padrão de 0,3 Newtons. Considerando que se deseja um poder de 80% e um nível de significância de 5%, qual o tamanho da amostra para testar a equivalência sendo que o intervalo de equivalência é $(-0,2 ; +0,2)$ Newtons?

As hipóteses nula e alternativa são:

$$H_0 : |\bar{x} - 4,2| \geq 0,2$$

$$H_1 : |\bar{x} - 4,2| < 0,2$$

Exemplo 2: Um laboratório deseja atestar que seu novo medicamento para certa condição não é inferior ao que já está estabelecido no mercado quanto ao tempo de entrada na corrente sanguínea. O medicamento referência entra na corrente sanguínea em 180 segundos em média com desvio de 30 segundos. Considerando um nível de significância de 5%, um poder de 80% e que se deseja detectar uma diferença mínima de 10 segundos, qual o tamanho de amostra necessário para testar a não inferioridade com uma margem de 10%, isto é, 18 segundos?

As hipóteses nula e alternativa são:

$$H_0 : \bar{x} - 180 \leq 18$$

$$H_1 : \bar{x} - 180 > 18$$

2.2.3.1.2 Exemplo aplicado no pacote *TrialSize* -

No pacote *TrialSize* existem duas funções para calcular o tamanho de amostra para esta situação, que são *OneSampleMean.Equivalence*, para equivalência e *OneSampleMean.NIS*, para não-inferioridade e superioridade.

Para a função *OneSampleMean.Equivalence* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **sigma**, desvio padrão da variável de interesse X.
- **margin**, matematicamente, $margin = \bar{x} - \mu_0$, ou seja, é a diferença entre \bar{x} (média amostral) e um valor de referência μ_0 . Porém na prática, como ainda não se tem a amostra, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).
- **delta**, margem de não-inferioridade ou superioridade.

Obs.: O argumento *margin*, neste caso, é um valor em módulo, ou seja, está programado para desconsiderar o sinal do valor.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 1**:

```
> n=OneSampleMean.Equivalence(alpha=0.05, beta=0.2, sigma=0.3, margin=0.1, delta=0.2)
> n
[1] 55.64302
```

O valor que retorna para o tamanho estimado da amostra é $n=55,64$, ou seja, arredondando para cima, **$n=56$ sacos de salgadinho**.

Para a função *OneSampleMean.NIS* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **sigma**, desvio padrão da variável de interesse X.
- **delta**, margem de não-inferioridade ou superioridade.
- **margin**, matematicamente, $margem = \bar{x} - \mu_0$, ou seja, é a diferença entre \bar{x} (média amostral) e um valor de referência μ_0 . Porém na prática, como ainda não se tem a amostra, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).

Obs.: O argumento *margin*, neste caso, pode ser um valor positivo ou negativo, sendo considerado o sinal do valor. Claramente, se executarmos o exemplo 1 (com um valor positivo) o resultado será igual ao apresentado anteriormente, ou seja, 56 sacos de salgadinho.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 2**:

```
> n=OneSampleMean.NIS(alpha=0.05, beta=0.2, sigma=30, delta=18, margin=-10)
> n
[1] 7.097323
```

O valor que retorna para o tamanho estimado da amostra é $n=7,1$, ou seja, arredondando para cima, **$n=8$ pacientes**.

Obs.: Salientando que para o exemplo, nessa função, utilizou-se o limite de não-inferioridade, mas é possível usar também o limite de superioridade.

2.2.3.2 Para uma proporção

Essa técnica é utilizada quando se deseja avaliar se a proporção de uma variável na população está suficientemente próxima de um valor-alvo (ou um valor de referência) a ponto de ser considerada equivalente.

As hipóteses nula e alternativa para testar a equivalência são:

$$H_0 : |p - p_0| \geq \text{delta}$$

$$H_1 : |p - p_0| < \text{delta}$$

As hipóteses nula e alternativa para testar a superioridade e não-inferioridade são:

$$H_0 : p - p_0 \leq \text{delta}$$

$$H_1 : p - p_0 > \text{delta}$$

2.2.3.2.1 Exemplos -

Exemplo 1: Um médico está interessado em analisar a mudança da densidade óssea pós-tratamento com um novo medicamento para osteoporose em termos de taxa de resposta. Assume-se que a taxa esperada de resposta do medicamento é de 0,6 e que uma diferença mínima clinicamente importante é 0,05. Considerando que se deseja um poder de 80% e um nível de significância de 5% , qual o tamanho da amostra para testar a equivalência sendo que o intervalo de equivalência é (-0,2 ; +0,2)?

As hipóteses nula e alternativa são:

$$H_0 : |p - 0,6| \geq 0,05$$

$$H_1 : |p - 0,6| < 0,05$$

Exemplo 2: Supondo os mesmos dados do exemplo 1, porém deseja-se testar a não-inferioridade do novo medicamento considerando um limite de não-inferioridade de 0,1.

As hipóteses nula e alternativa são:

$$H_0 : p - 0,6 \leq 0,05$$

$$H_1 : p - 0,6 > 0,05$$

2.2.3.2.2 Exemplo aplicado no pacote *TrialSize* -

No pacote *TrialSize* existem duas funções para calcular o tamanho de amostra para esta situação, que são *OneSampleProportion.Equivalence* e *OneSampleProportion.NIS*.

Para a função *OneSampleProportion.Equivalence* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **p**, a proporção verdadeira da categoria de interesse na variável estudada na população.

- **delta**, matematicamente, $delta = P - P_0$, ou seja, é a diferença entre P (verdadeira proporção) e um valor de referência P_0 . Porém na prática, como ainda não se tem a amostra, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).
- **margin**, margem de não-inferioridade ou superioridade.

Obs.: O argumento *margin*, neste caso, é um valor em módulo, ou seja, está programado para desconsiderar o sinal do valor.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 1**:

```
> n=OneSampleProportion.Equivalence(alpha=0.05, beta=0.2, p=0.6, delta=0.05, margin=0.2)
> n
[1] 91.34771
```

O valor que retorna para o tamanho estimado da amostra é $n=91,35$, ou seja, arredondando para cima, **$n=92$ pacientes**.

Para a função *OneSampleProportion.NIS* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **p**, a proporção verdadeira da categoria de interesse na variável estudada na população.
- **delta**, matematicamente, $delta = P - P_0$, ou seja, é a diferença entre P (verdadeira proporção) e um valor de referência P_0 . Porém na prática, como ainda não se tem a amostra, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).
- **margin**, margem de não-inferioridade ou superioridade.

Obs.: O argumento *margin*, neste caso, pode ser um valor positivo ou negativo, sendo considerado o sinal do valor. Claramente, se executarmos o exemplo 1 (com um valor positivo) o resultado será igual ao apresentado anteriormente, ou seja, 92 pacientes.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 2**:

```
> n=OneSampleProportion.NIS(alpha=0.05, beta=0.2, p=0.6, delta=0.05, margin=-0.1)
> n
[1] 65.94728
```

O valor que retorna para o tamanho estimado da amostra é $n=65,95$, ou seja, arredondando para cima, **$n=66$ pacientes**.

Obs.: Salientando que para o exemplo, nessa função, utilizou-se o limite de não-inferioridade, mas é possível usar também o limite de superioridade.

2.3 Teste de Hipóteses para duas amostras

2.3.1 Comparação entre duas médias de grupos independentes

Nesta técnica deseja-se calcular o tamanho amostral de um estudo cujo objetivo é comparar se a média de dois grupos independentes diferem ou não em relação à variável interesse, com um certo nível de confiança e poder.

As hipóteses de interesse neste contexto são:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2 \text{ ou } \mu_1 > \mu_2 \text{ ou } \mu_1 < \mu_2$$

Ao utilizar essa técnica supõe-se que as variâncias (e conseqüentemente os desvios padrões também) são iguais e que o número de observações é o mesmo nas duas amostras com exceção da função *pwr.t2n.test* do pacote *pwr* onde é calculado o poder para duas amostras de tamanhos diferentes. Já na função *TwoSampleMean.Equality* do pacote *TrialSize* é possível o cálculo para ambos os casos, delimitando isso no argumento *k*.

2.3.1.1 Exemplo

Uma fábrica de erva mate está interessada em detectar se as duas máquinas de empacotar, a nova, de última geração, e a antiga, estão empacotando a mesma quantidade de erva mate, em Kg. Com base em informações do fabricante, sabe-se que o desvio é de 0,4Kg em cada máquina. O gerente afirma que deseja um nível de significância igual a 5% e poder 0,80 para detectar uma diferença mínima de 0,3Kg.

2.3.1.2 Exemplo aplicado no pacote *TrialSize*

No pacote *TrialSize* existe a função *TwoSampleMean.Equality* para calcular o tamanho de amostra para esta situação.

Para a função *TwoSampleMean.Equality* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **sigma**, desvio padrão combinado dos dois grupos.
- **k**, $k = \frac{n_1}{n_2}$, razão entre o tamanho das duas amostras.

- **margin**, matematicamente, $margin = \mu_2 - \mu_1$, ou seja, é a diferença entre μ_2 (média da variável de interesse na amostra 2) e μ_1 (média da variável de interesse na amostra 1). Porém na prática, como ainda não se tem a média de uma variável, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do exemplo:

```
> n=TwoSampleMean.Equality(alpha=0.05, beta=0.2, sigma=0.4, k=1, margin=0.3)
> n
[1] 27.90713
```

O valor que retorna para o tamanho estimado da amostra é $n=27,91$, ou seja, arredondando para cima, **$n=28$ pacotes de café por máquina.**

2.3.1.3 Exemplo aplicado no pacote *pwr*

No pacote *pwr* existem duas funções para calcular o tamanho de amostra e poder para esta situação, que são *pwr.t.test* e *pwr.2tn.test*.

Para a função *pwr.t.test* são necessários os seguintes argumentos:

- **n**, número de observações (por amostra).
- **d**, magnitude do efeito (tamanho de efeito).
- **sig.level**, nível de significância (Probabilidade do erro tipo I).
- **power**, poder (1-Probabilidade do erro tipo II).
- **type**, definição do teste *t de Student*: para uma amostra (“one.sample”), duas amostras independentes (“two.sample”) ou duas amostras pareadas (“paired”).
- **alternative**, especifica o tipo de teste de acordo com a hipótese alternativa, com as opções: “two.sided” (bilateral), “greater” (unilateral superior) ou “less” (unilateral inferior). *Default* é *alternative* = “two.sided”.

Tamanho de efeito:

Sabendo que este pacote é baseado nos métodos de [Cohen \(1988\)](#), a magnitude do efeito (tamanho de efeito) d é calculada da seguinte maneira:

$$d = \frac{\mu_2 - \mu_1}{\sigma} = \frac{0,3}{0,4} = 0,75, \quad (2.9)$$

onde $\mu_2 - \mu_1$, neste caso, é a diferença mínima a ser detectada em função de não termos as informações das médias.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do exemplo:

```
> n=pwr.t.test(n=NULL, d=0.75, sig.level=0.05, power=0.8, type="two.sample", alternative="two.sided")
> n

Two-sample t test power calculation

      n = 28.89959
      d = 0.75
sig.level = 0.05
power = 0.8
alternative = two.sided

NOTE: n is number in *each* group
```

O valor que retorna para o tamanho estimado da amostra é $n=28,90$, ou seja, arredondando para cima, **$n=29$ pacotes de café por máquina.**

Poder:

Suponhamos que gostaríamos de calcular o poder, dado o tamanho de amostra igual a $n=23$, temos que a função é utilizada da seguinte forma:

```
> poder=pwr.t.test(n=23, d=0.75, sig.level=0.05, power=NULL, type="two.sample", alternative="two.sided")
> poder

Two-sample t test power calculation

      n = 23
      d = 0.75
sig.level = 0.05
power = 0.7011076
alternative = two.sided

NOTE: n is number in *each* group
```

O valor que retorna para o poder é de **0,70**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,70.

Para a função *pwr.t2n.test* são necessários os seguintes argumentos:

- **n1**, número de observações na amostra 1.
- **n2**, número de observações na amostra 2.
- **d**, magnitude do efeito (tamanho de efeito).
- **sig.level**, nível de significância (Probabilidade do erro tipo I).
- **power**, poder (1-Probabilidade do erro tipo II).
- **alternative**, especifica o tipo de teste de acordo com a hipótese alternativa, com as opções: “two.sided” (bilateral), “greater” (unilateral superior) ou “less” (unilateral inferior). *Default* é `alternative="two.sided"`.

Poder:

Suponhamos que gostaríamos de calcular o poder dado que temos duas amostras de tamanho $n1=32$ e $n2=48$, temos que a função é utilizada da seguinte forma:

```
> poder=pwr.t2n.test(n1=32, n2=48, d=0.75, sig.level=0.05, power=NULL, alternative="two.sided")
> poder

t test power calculation

      n1 = 32
      n2 = 48
       d = 0.75
sig.level = 0.05
  power = 0.9007169
alternative = two.sided
```

O valor que retorna para o poder é de **0,90**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,90.

2.3.2 Comparação entre duas proporções de grupos independentes

Nesta técnica deseja-se calcular o tamanho amostral de um estudo cujo objetivo é comparar se as proporções em dois grupos distintos diferem ou não em relação à variável de interesse, com um certo nível de confiança e poder. Para isso, assume-se que o número de observações é o mesmo nas duas amostras.

Por exemplo, saber se a prevalência de uma determinada doença ou condição (por exemplo, óbito) é diferente ou não entre os pacientes que possuem um fator de risco ou condição ([Agranonik e Hirakata, 2011](#)).

As hipóteses de interesse neste contexto são:

$$H_0 : P_1 = P_2$$

$$H_1 : P_1 \neq P_2$$

2.3.2.1 Exemplos

Exemplo 1: Uma neurologista deseja comparar a proporção de idosos com a doença de Alzheimer, entre mulheres e homens, de uma cidade. As prevalências encontradas em um estudo recente foram de 0,45 e 0,15, respectivamente. Assume-se que o nível de significância adotado é de 5% e o poder do teste de 80%. Calcule o tamanho amostral para ambos grupos adotando uma diferença mínima de 0,3.

Exemplo 2: Considere os dados do exemplo 1, mas que já foi coletada uma amostra de 43 idosos do sexo masculino, qual o tamanho da amostra que deverá ser coletada do sexo feminino?

2.3.2.2 Exemplo aplicado no pacote *TrialSize*

No pacote *TrialSize* existe a função *TwoSampleProportion.Equality* para calcular o tamanho de amostra para esta situação.

Para a função *TwoSampleProportion.Equality* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **p1**, proporção para amostra 1.
- **p2**, proporção para amostra 2.
- **k**, $k = \frac{n_1}{n_2}$, razão entre o tamanho das duas amostras.
- **delta**, matematicamente, $delta = p_1 - p_2$, ou seja, é a diferença entre p_1 (proporção do desfecho de interesse na amostra 1) e p_2 (proporção do desfecho de interesse na amostra 2). Porém na prática, como ainda não se tem a amostra, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 1**:

```
> n=TwoSampleProportion.Equality(alpha=0.05, beta=0.2, p1=0.45, p2=0.15, k=1, delta=0.3)
> n
[1] 32.70367
```

O valor que retorna para o tamanho estimado da amostra é $n=32,70$, ou seja, arredondando para cima, **$n=33$ idosos por sexo**.

Obs.: o argumento delta, neste caso, é um valor em módulo, ou seja, está programado para desconsiderar o sinal do valor. Para ambos valores, positivos ou negativos, resultam o mesmo valor de tamanho de amostra.

2.3.2.3 Exemplo aplicado no pacote *pwr*

No pacote *pwr* existem duas funções para calcular o tamanho de amostra e poder para esta situação, que são *pwr.2p.test* e *pwr.2p2n.test*. Porém, para proceder com a utilização desta função faz-se necessário o cálculo do tamanho de efeito. Para tal, deve-se usar a função *ES.h*, encontrada neste mesmo pacote, que calcula a magnitude do efeito (tamanho de efeito) h entre duas proporções.

O cálculo que tal função faz está descrito em (2.7) e (2.8).

Para a função *ES.h* são necessários os seguintes argumentos:

- **p1**, primeira proporção.
- **p2**, segunda proporção.

Tamanho de efeito:

Essa função é utilizada da seguinte forma no caso dos exemplos:

```
> h=ES.h(0.45, 0.15)
> h
[1] 0.6752301
```

O valor que retorna para o tamanho de efeito é **$h=0.6752$** .

A partir de então usaremos o resultado de h no argumento do tamanho de efeito nas funções *pwr.2p.test* e *pwr.2p2n.test* tanto para o exemplo 1 como para o exemplo 2. E por isso escreve-se **$h=h$** no código do R.

Para a função *pwr.2p.test* são necessários os seguintes argumentos:

- **h**, magnitude do efeito (tamanho de efeito).
- **n**, número de observações (por amostra).
- **sig.level**, nível de significância (Probabilidade do erro tipo I).
- **power**, poder do teste (1-Probabilidade do erro tipo II).
- **alternative**, especifica o tipo de teste de acordo com a hipótese alternativa, com as opções: “two.sided” (bilateral), “greater” (unilateral superior) ou “less” (unilateral inferior). *Default* é `alternative=“two.sided”`.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 1**:

```
> n=pwr.2p.test(h=h, n=NULL, sig.level=0.05, power=0.8, alternative="two.sided")
> n

      Difference of proportion power calculation for binomial distribution (arcsine transformation)
      h = 0.6752301
      n = 34.42969
      sig.level = 0.05
      power = 0.8
      alternative = two.sided

NOTE: same sample sizes
```

O valor que retorna para o tamanho estimado da amostra é $n=34,43$, ou seja, arredondando para cima, **$n=35$ idosos por sexo**.

Poder:

Suponhamos que gostaríamos de calcular o poder, dado o tamanho de amostra igual a $n=20$, temos que a função é utilizada da seguinte forma:

```
> poder=pwr.2p.test(h=h, n=20, sig.level=0.05, power=NULL, alternative="two.sided")
> poder

Difference of proportion power calculation for binomial distribution (arcsine transformation)

      h = 0.6752301
      n = 20
sig.level = 0.05
  power = 0.5695995
alternative = two.sided

NOTE: same sample sizes
```

O valor que retorna para o poder é de **0,57**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,57.

Para a função *pwr.2p2n.test* são necessários os seguintes argumentos:

- **h**, magnitude do efeito (tamanho de efeito).
- **n1**, número de observações na amostra 1.
- **n2**, número de observações na amostra 2.
- **sig.level**, nível de significância (Probabilidade do erro tipo I).
- **power**, poder do teste (1-Probabilidade do erro tipo II).
- **alternative**, especifica o tipo de teste de acordo com a hipótese alternativa, com as opções: “two.sided” (bilateral), “greater” (unilateral superior) ou “less” (unilateral inferior). *Default* é alternative=“two.sided”.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 2**:

```
> n=pwr.2p2n.test(h=h, n1=43, n2=NULL, sig.level=0.05, power=0.8, alternative="two.sided")
> n

difference of proportion power calculation for binomial distribution (arcsine transformation)

      h = 0.6752301
      n1 = 43
      n2 = 28.70795
sig.level = 0.05
  power = 0.8
alternative = two.sided

NOTE: different sample sizes
```

O valor que retorna para o tamanho estimado da amostra é $n=28,71$, ou seja, arredondando para cima, **$n=29$ idosos**.

Como curiosidade, a Tabela 2.1 apresenta alguns cenários possíveis fixando o número de observações na amostra 1 (n_1) em 25, 31, 43 e 58, e os respectivos números de observações na amostra 2 (n_2) mantendo o poder desejado.

Tabela 2.1: Diferentes combinações fixando n1

n1	n2
25	56
31	39
43	29
58	25

Poder:

Suponhamos que gostaríamos de calcular o poder dado que temos duas amostras de tamanho $n_1=26$ e $n_2=35$, respectivamente, temos que a função é utilizada da seguinte forma:

```
> poder=pwr.2p2n.test(h=h, n1=26, n2=35, sig.level=0.05, power=NULL, alternative="two.sided")
> poder

difference of proportion power calculation for binomial distribution (arcsine transformation)

      h = 0.6752301
     n1 = 26
     n2 = 35
sig.level = 0.05
  power = 0.7415216
alternative = two.sided

NOTE: different sample sizes
```

O valor que retorna para o poder é de **0,74**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,74.

2.3.3 Equivalência

É um tipo de análise para determinar se é possível afirmar que a diferença entre duas médias ou proporções estimadas está dentro de uma margem denominada intervalo de equivalência.

Ao utilizar um teste de equivalência para duas amostras independentes, assim como no teste para uma amostra, deve-se especificar um intervalo de valores que esteja “perto o suficiente” para considerar as duas médias ou duas proporções equivalentes.

2.3.3.1 Para duas médias de grupos independentes

Essa técnica é utilizada quando se deseja testar se duas médias de amostras independentes estão próximas o suficiente para serem consideradas equivalentes.

As hipóteses nula e alternativa para testar a equivalência são:

$$H_0 : |\mu_2 - \mu_1| \geq \text{delta}$$

$$H_1 : |\mu_2 - \mu_1| < \text{delta}$$

As hipóteses nula e alternativa para testar a superioridade e não-inferioridade são:

$$H_0 : \mu_2 - \mu_1 \leq \text{delta}$$

$$H_1 : \mu_2 - \mu_1 > \text{delta}$$

2.3.3.1.1 Exemplos -

Exemplo 1: Um analista do setor de qualidade quer determinar se a quantidade média de ingrediente ativo em um tipo genérico de analgésico está dentro de 1mg da quantidade média de uma marca conceituada de analgésico. A partir das amostras anteriores, o analista estima que o desvio padrão da população é 0,41. Considera-se um poder de 80% e o nível de significância de 5%, e que se deseja detectar uma diferença mínima de 0,7mg, qual o tamanho da amostra para testar a equivalência?

Exemplo 2: Supondo que o analista deseja testar a não-inferioridade do medicamento genérico considerando um limite de não-inferioridade de 0,05. Mantendo o desvio padrão de 0,41, o nível de significância de 5% e o poder de 80%, porém querendo detectar uma diferença mínima de 0,2mg, qual o tamanho da amostra para testar a não-inferioridade?

2.3.3.1.2 Exemplo aplicado no pacote *TrialSize* -

No pacote *TrialSize* existem duas funções para calcular o tamanho de amostra para esta situação, que são *TwoSampleMean.Equivalence* e *TwoSampleMean.NIS*.

Para a função *TwoSampleMean.Equivalence* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **sigma**, desvio padrão combinado dos dois grupos.
- **k**, $k = \frac{n_1}{n_2}$, razão entre o tamanho das duas amostras.
- **delta**, margem de não-inferioridade ou superioridade.
- **margin**, matematicamente, $\text{delta} = \mu_2 - \mu_1$, ou seja, é a diferença entre μ_2 (média da variável de interesse na amostra 2) e μ_1 (média da variável de interesse na amostra 1). Porém na prática, como ainda não se tem a média de uma variável, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).

Obs.: o argumento *margin*, neste caso, é um valor em módulo, ou seja, está programado para desconsiderar o sinal do valor.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 1**:

```
> n=TwoSampleMean.Equivalence(alpha=0.05, beta=0.2, sigma=0.41, k=1, delta=1, margin=0.7)
> n
[1] 31.99073
```

O valor que retorna para o tamanho estimado da amostra é $n=31,99$, ou seja, arredondando para cima, **$n=32$ unidades do remédio em cada grupo de analgésicos.**

Para a função *TwoSampleMean.NIS* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **sigma**, desvio padrão combinado dos dois grupos.
- **k**, $k = \frac{n_1}{n_2}$, razão entre o tamanho das duas amostras.
- **delta**, margem de não-inferioridade ou superioridade.
- **margin**, matematicamente, $delta = \mu_2 - \mu_1$, ou seja, é a diferença entre μ_2 (média da variável de interesse na amostra 2) e μ_1 (média da variável de interesse na amostra 1). Porém na prática, como ainda não se tem a média de uma variável, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 2**:

```
> n=TwoSampleMean.NIS(alpha=0.05, beta=0.2, sigma=0.41, k=1, delta=-0.05, margin=0.2)
> n
[1] 33.25721
```

O valor que retorna para o tamanho estimado da amostra é $n=33,26$, ou seja, arredondando para cima, **$n=34$ unidades do remédio em cada grupo de analgésicos.**

Obs.: Deixa-se claro que para o exemplo, nessa função, usamos o limite de não-inferioridade, mas é possível usar também o limite de superioridade.

2.3.3.2 Para duas proporções de grupos independentes

Essa técnica é utilizada quando se deseja testar se duas proporções de amostras independentes estão próximas o suficiente para serem consideradas equivalentes.

As hipóteses nula e alternativa para testar a equivalência são:

$$H_0 : |p_1 - p_2| \geq margin$$

$$H_1 : |p_1 - p_2| < margin$$

As hipóteses nula e alternativa para testar a superioridade e não-inferioridade são:

$$H_0 : p_1 - p_2 \leq \textit{margin}$$

$$H_1 : p_1 - p_2 > \textit{margin}$$

2.3.3.2.1 Exemplos -

Exemplo 1: Um dermatologista deseja testar a equivalência entre a taxa de cura (desfecho de interesse) de dois antibióticos para tratamento de infecções de pele de pacientes do hospital onde trabalha. Supondo que a taxa de cura do medicamento 1 é 0,75 e do medicamento 2 é 0,80 e que o limite de equivalência de é 20%. Considerando um poder de 80%, o nível de significância de 5%, qual o tamanho da amostra para testar a equivalência?

As hipóteses nula e alternativa são:

$$H_0 : |0,75 - 0,80| \geq 0,2$$

$$H_1 : |0,75 - 0,80| < 0,2$$

Exemplo 2: Supondo os mesmos dados do exemplo 1, porém deseja-se testar a superioridade do medicamento 1 considerando um limite de superioridade de 0,05, e que a taxa de cura do medicamento 1 é 0,65 e do medicamento 2 é 0,85, qual o tamanho da amostra para testar a superioridade?

As hipóteses nula e alternativa são:

$$H_0 : 0,65 - 0,85 \leq 0,05$$

$$H_1 : 0,65 - 0,85 > 0,05$$

2.3.3.2.2 Exemplo aplicado no pacote *TrialSize* -

No pacote *TrialSize* existem duas funções para calcular o tamanho de amostra para esta situação, que são *TwoSampleProportion.Equivalence* e *TwoSampleProportion.NIS*.

Para a função *TwoSampleProportion.Equivalence* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **p1**, proporção para amostra 1.
- **p2**, proporção para amostra 2.
- **k**, $k = \frac{n_1}{n_2}$, razão entre o tamanho das duas amostras.
- **delta**, matematicamente, $\textit{delta} = p_1 - p_2$, ou seja, é a diferença entre p_1 (proporção do desfecho de interesse da amostra 1) e p_2 (proporção do desfecho de interesse na amostra 2). Porém na prática, como ainda não se tem a amostra, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).

- **margin**, margem de não-inferioridade ou superioridade.

Obs.: o argumento *margin*, neste caso, é um valor em módulo, ou seja, está programado para desconsiderar o sinal do valor.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 1**:

```
> n=TwoSampleProportion.Equivalence(alpha=0.05, beta=0.2, p1=0.75, p2=0.80, k=1, delta=0.05, margin=0.2)
> n
[1] 95.48616
```

O valor que retorna para o tamanho estimado da amostra é $n=95,48$, ou seja, arredondando para cima, **n=96 pacientes em cada grupo**.

Para a função *TwoSampleProportion.NIS* são necessários os seguintes argumentos:

- **alpha**, nível de significância.
- **beta**, poder = 1-beta.
- **p1**, proporção para amostra 1.
- **p2**, proporção para amostra 2.
- **k**, $k = \frac{n_1}{n_2}$, razão entre o tamanho das duas amostras.
- **delta**, matematicamente, $delta = p_1 - p_2$, ou seja, é a diferença entre p_1 (proporção do desfecho de interesse da amostra 1) e p_2 (proporção do desfecho de interesse na amostra 2). Porém na prática, como ainda não se tem a amostra, fixamos tal diferença mínima que se deseja detectar (com o poder desejado).
- **margin**, margem de não-inferioridade ou superioridade.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do **exemplo 2**:

```
> n=TwoSampleProportion.NIS(alpha=0.05, beta=0.2, p1=0.65, p2=0.85, k=1, delta=0.2, margin=0.05)
> n
[1] 97.54701
```

O valor que retorna para o tamanho estimado da amostra é $n=97,55$, ou seja, arredondando para cima, **n=98 pacientes em cada grupo**.

Obs.: Salientando que para o exemplo, nessa função, utilizou-se o limite de superioridade, mas é possível usar também o limite de não-inferioridade.

2.3.4 Comparação entre duas médias de grupos pareados

Nesta técnica deseja-se calcular o tamanho amostral de um estudo cujo objetivo é comparar se as médias de dois grupos dependentes, ou seja, dois grupos relacionados ou comumente dito pareados, diferem ou não em relação à resposta de interesse, com um certo nível de significância e poder.

Neste caso, um exemplo comum de dois grupos pareados é medir uma característica do indivíduo antes e depois do mesmo ser submetido a um tratamento.

As hipóteses de interesse neste contexto são:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2 \text{ ou } \mu_1 > \mu_2 \text{ ou } \mu_1 < \mu_2$$

Ao utilizar essa técnica supõe-se que as variâncias (e conseqüentemente os desvios padrões também) são iguais e que o número de observações é o mesmo nas duas amostras.

2.3.4.1 Exemplo

Um pesquisador da área da Nutrição gostaria de verificar o efeito da utilização de um novo suplemento de ação rápida na média de peso levantado por atletas de mesmo nível. O peso, em kg, levantado será medido da mesma forma antes e após a dieta de 2 dias com o suplemento. O pesquisador deseja detectar uma magnitude de efeito “média”, portanto, baseado em [Cohen \(1988\)](#)² ele utiliza $d = 0,5$. Quantos atletas são necessários para detectar tal magnitude do efeito, com nível de significância de 5% e poder 80%?

2.3.4.2 Exemplo aplicado no pacote *pwr*

No pacote *pwr* existe a funções *pwr.t.test* para calcular o tamanho de amostra e poder para esta situação.

Para a função *pwr.t.test* são necessários os seguintes argumentos:

- **n**, número de observações (por amostra).
- **d**, magnitude do efeito (tamanho de efeito).
- **sig.level**, nível de significância (Probabilidade do erro tipo I).
- **power**, poder (1-Probabilidade do erro tipo II).
- **type**, definição do teste *t de Student*: para uma amostra (“one.sample”), duas amostras independentes (“two.sample”) ou duas amostras pareadas (“paired”).

²O autor considera, para esta técnica, $d = 0,2$, $d = 0,5$ e $d = 0,8$ como tamanhos de efeito pequeno, médio e grande, respectivamente.

- **alternative**, especifica do tipo de teste de acordo com a hipótese alternativa, com as opções: “two.sided” (bilateral), “greater” (unilateral superior) ou “less” (unilateral inferior). *Default* é `alternative=“two.sided”`.

Tamanho de efeito:

Sabendo que este pacote é baseado nos métodos de [Cohen \(1988\)](#), a magnitude do efeito (tamanho de efeito) d é calculada da seguinte maneira para amostras pareadas:

$$d = \frac{|\mu_2 - \mu_1|}{\sigma}, \quad (2.10)$$

onde μ_2 é o peso médio levantado depois da dieta, μ_1 é o peso médio levantado antes da dieta e σ é o desvio padrão.

Obs.: Para este exemplo não é calculado pois ainda não se dispõe das médias, por isso utiliza-se um tamanho de efeito de acordo com [Cohen \(1988\)](#).

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do exemplo:

```
> n=pwr.t.test(n=NULL, d=0.5, sig.level=0.05, power=0.8, type="paired", alternative="two.sided")
> n

Paired t test power calculation

      n = 33.36713
      d = 0.5
sig.level = 0.05
  power = 0.8
alternative = two.sided

NOTE: n is number of *pairs*
```

O valor que retorna para o tamanho estimado da amostra é $n=33,37$, ou seja, arredondando para cima, **$n=34$ atletas**.

Poder:

Suponhamos que gostaríamos de calcular o poder, dado o tamanho de amostra igual a $n=20$, temos que a função é utilizada da seguinte forma:

```
> n=pwr.t.test(n=20, d=0.5, sig.level=0.05, power=NULL, type="paired", alternative="two.sided")
> n

Paired t test power calculation

      n = 20
      d = 0.5
sig.level = 0.05
  power = 0.5645044
alternative = two.sided

NOTE: n is number of *pairs*
```

O valor que retorna para o poder é de **0,56**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,56.

2.4 Análise de Variância (ANOVA)

A Análise de Variância, mais conhecida como ANOVA, é utilizada para comparar as médias de três ou mais grupos independentes com o objetivo de saber se os grupos diferem ou não em relação à resposta média de interesse (μ).

Para essa técnica é necessário atender algumas suposições, a saber:

- As amostras são aleatórias e independentes.
- Variâncias homogêneas, ou seja, variâncias semelhantes nas diferentes amostras.
- Variável de interesse normalmente distribuída.

Na prática, esses pressupostos não precisam ser todos rigorosamente satisfeitos. Os resultados são empiricamente verdadeiros sempre que as variáveis de interesse nas populações são aproximadamente normais (isso é, não muito assimétricas) e têm variâncias próximas (Milone, 2009).

As hipóteses de interesse neste contexto são:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

$$H_1 : \mu_i \neq \mu_j, \text{ onde } i = 1, \dots, k, j = 1, \dots, k \text{ e } i \neq j$$

2.4.1 Exemplo

Em um colégio deseja-se avaliar a proficiência dos alunos de cada ano do ensino médio (1º, 2º e 3º anos). A coordenadora deseja detectar uma magnitude de efeito “grande”, portanto, baseado em Cohen (1988)³ ela utiliza $f = 0,4$. Quantos alunos deve-se submeter a um teste dado que se deseja um nível de significância de 5% e poder de 80%.

As hipóteses nula e alternativa são:

$$H_0 : \mu_{1^\circ \text{ ano}} = \mu_{2^\circ \text{ ano}} = \mu_{3^\circ \text{ ano}}$$

$$H_1 : \mu_i \neq \mu_j, \text{ onde } i, j = 1^\circ \text{ ano}, 2^\circ \text{ ano ou } 3^\circ \text{ ano e } i \neq j$$

2.4.2 Exemplo aplicado no pacote *pwr*

No pacote *pwr* existe a função *pwr.anova.test* para calcular o tamanho da amostra e poder para esta situação.

Para a função *pwr.anova.test* são necessários os seguintes argumentos:

- **k**, número de grupos.
- **n**, número de observações (por grupo).

³O autor considera, para esta técnica, $f = 0,1$, $f = 0,25$ e $f = 0,4$ como tamanhos de efeito pequeno, médio e grande, respectivamente.

- **f**, magnitude do efeito (tamanho de efeito).
- **sig.level**, nível de significância (Probabilidade do erro tipo I).
- **power**, poder (1-Probabilidade do erro tipo II).

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do exemplo:

```
> n=pwr.anova.test(k=3, n=NULL, f=0.4, sig.level=0.05, power=0.8)
> n

Balanced one-way analysis of variance power calculation

      k = 3
      n = 21.10364
      f = 0.4
sig.level = 0.05
power = 0.8

NOTE: n is number in each group
```

O valor que retorna para o tamanho estimado da amostra é $n=21,10$, ou seja, arredondando para cima, **$n=22$ alunos para cada ano do ensino médio.**

Poder:

Suponhamos que gostaríamos de calcular o poder, dado o tamanho de amostra igual a $n=15$, temos que a função é utilizada da seguinte forma:

```
> poder=pwr.anova.test(k=3, n=15, f=0.4, sig.level=0.05, power=NULL)
> poder

Balanced one-way analysis of variance power calculation

      k = 3
      n = 15
      f = 0.4
sig.level = 0.05
power = 0.6350412

NOTE: n is number in each group
```

O valor que retorna para o poder é **0,635**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,635.

2.4.3 Exemplo aplicado no pacote *pwr2*

No pacote *pwr2* existe a função *pwr.1way* para calcular o poder para esta situação.

Para a função *pwr.1way* são necessários os seguintes argumentos:

- **k**, número de grupos.
- **n**, número de observações (por grupo).
- **f**, magnitude do efeito (tamanho de efeito).

- **sig.level**, nível de significância (Probabilidade do erro tipo I).
- **delta**, menor diferença entre os k grupos.
- **sigma**, desvio padrão da variável de interesse X.

Poder:

Suponhamos os dados do exemplo acima e que gostaríamos de calcular o poder, dado o tamanho de amostra $n=14$ em cada ano do ensino médio, e como para esta função são necessários mais dois argumentos, suponhamos então um desvio padrão de 1 e a menor diferença entre grupos de 1. Temos então que a função é utilizada da seguinte forma:

```
> n=pwr.1way(k=3, n=14, alpha=0.05, f=0.4, delta=1.0, sigma=1.0)
> n

Balanced one-way analysis of variance power calculation

      k = 3
      n = 14
effect.size = 0.4
sig.level = 0.05
power = 0.6005168

NOTE: n is number in each group, total sample = 42 power = 0.600516815800322
```

O valor que retorna para o poder é **0,60**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,60.

2.5 Teste de Associação para variáveis categóricas

Essa técnica é utilizada quando se deseja descobrir se existe associação entre duas variáveis categóricas, geralmente agrupadas em tabelas de contingência.

As hipóteses de interesse neste contexto são:

H_0 : Não existe associação entre as variáveis

H_1 : Existe associação entre as variáveis

A estatística de teste segue distribuição qui-quadrado (χ^2), que tem como parâmetros graus de liberdade (para mais detalhes sobre a distribuição, recomenda-se [Callegari-Jacques \(2003\)](#)).

Os graus de liberdade para este cálculo são dados por:

$$\text{graus de liberdade} = (L - 1) \times (C - 1) ,$$

onde L é o número de categorias da primeira variável e C é o número de categorias da segunda variável.

2.5.1 Exemplo

Uma concessionária está interessada em testar a associação entre o sexo dos clientes e a preferência por marcas de carro. As possíveis categorias para sexo são feminino e masculino e para as marcas são Fiat, Ford, Volkswagen e Chevrolet. A gerente deseja detectar uma magnitude de efeito “média”, portanto, baseado em [Cohen \(1988\)](#)⁴ ela utiliza $w = 0,3$. Considerando um nível de significância igual a 5% e poder 80%, quantos clientes são necessários para tal?

As hipóteses nula e alternativa são:

H_0 : Não existe associação entre o sexo dos clientes e a preferência por marcas de carro

H_1 : Existe associação entre o sexo dos clientes e a preferência por marcas de carro

2.5.2 Exemplo aplicado no pacote *pwr*

No pacote *pwr* existe a função *pwr.chisq.test* para calcular o tamanho da amostra e poder para esta situação.

Para a função *pwr.chisq.test* são necessários os seguintes argumentos:

- **w**, magnitude do efeito (tamanho de efeito).
- **N**, número total de observações.
- **df**, graus de liberdade.
- **sig.level**, nível de significância (Probabilidade do erro tipo I).
- **power**, poder (1-Probabilidade do erro tipo II).

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do exemplo:

```
> n=pwr.chisq.test(w=0.3, N=NULL, df=(2-1)*(4-1), sig.level=0.05, power=0.8)
> n

Chi squared power calculation

      w = 0.3
      N = 121.1396
      df = 3
sig.level = 0.05
power = 0.8

NOTE: N is the number of observations
```

O valor que retorna para o tamanho estimado da amostra é $n=121,14$, ou seja, arredondando para cima, **$n=122$ clientes**.

⁴O autor considera, para esta técnica, $w = 0,1$, $w = 0,3$ e $w = 0,5$ como tamanhos de efeito pequeno, médio e grande, respectivamente.

Poder:

Suponhamos que gostaríamos de calcular o poder, dado o tamanho de amostra igual a $n=100$, temos que a função é utilizada da seguinte forma:

```
> poder=pwr.chisq.test(w=0.3, N=100, df=(2-1)*(4-1), sig.level=0.05, power=NULL)
> poder
```

```
Chi squared power calculation
```

```
      w = 0.3
      N = 100
      df = 3
sig.level = 0.05
power = 0.7112536
```

```
NOTE: N is the number of observations
```

O valor que retorna para o poder é **0,71**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,71.

2.6 Correlação

Essa técnica é utilizada para avaliar se existe relação linear entre duas variáveis quantitativas. É possível avaliar a correlação de duas maneiras: via gráfico de dispersão ou via coeficiente de correlação, cuja fórmula foi proposta por Karl Pearson em 1896, e por isso denominado coeficiente de correlação de Pearson.

O coeficiente de correlação, representado por ρ (lê-se rho) pode variar de -1 a +1, sendo que valores negativos indicam uma correlação inversa (negativa), ou seja, quando uma variável cresce a outra diminui; valores positivos indicam uma correlação direta (positiva), ou seja, as duas variáveis variam para o mesmo sentido. Quando igual a zero, podemos concluir que não há correlação linear.

As hipóteses de interesse neste contexto são:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

2.6.1 Exemplo

O engenheiro de uma indústria de alumínio deseja verificar se existe correlação entre o conteúdo de hidrogênio e a porosidade das fundições de liga de alumínio de certa peça. Espera-se uma correlação de 0,5, isto é, que existe uma correlação positiva entre as variáveis, ou seja, à medida que aumenta o hidrogênio, a porosidade também aumenta. Considerando um poder de 80% e um nível de significância de 5%, quantas peças serão necessárias para compor a amostra?

As hipóteses nula e alternativa são:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

2.6.2 Exemplo aplicado no pacote *pwr*

No pacote *pwr* existe a função *pwr.r.test* para calcular o tamanho da amostra e poder para esta situação. Por questões computacionais, essa função utiliza a transformação z' dada pela seguinte fórmula:

$$z' = \arctan r + \frac{r}{2 \times (n + 1)} .$$

Para a função *pwr.r.test* são necessários os seguintes argumentos:

- **n**, número de observações.
- **r**⁵, coeficiente de correlação linear.
- **sig.level**, nível de significância (Probabilidade do erro tipo I).
- **power**, poder (1-Probabilidade do erro tipo II).
- **alternative**, especifica o tipo de teste de acordo com a hipótese alternativa, com as opções: “two.sided” (bilateral), “greater” (unilateral superior) ou “less” (unilateral inferior). *Default* é *alternative*=“two.sided”.

Tamanho da amostra:

Essa função é utilizada da seguinte forma no caso do exemplo:

```
> n=pwr.r.test(n=NULL, r=0.5, sig.level=0.05, power=0.8,alternative="two.sided")
> n

approximate correlation power calculation (arctangh transformation)

      n = 28.24841
      r = 0.5
sig.level = 0.05
  power = 0.8
alternative = two.sided
```

O valor que retorna para o tamanho estimado da amostra é $n=28,25$, ou seja, arredondando para cima, **$n=29$ peças**.

Poder:

Suponhamos que gostaríamos de calcular o poder, dado o tamanho de amostra igual a $n=20$, temos que a função é utilizada da seguinte forma:

⁵Note que quanto mais próximo de zero for r , maior o tamanho da amostra


```
> poder=pwr.r.test(n=20, r=0.5, sig.level=0.05, power=NULL, alternative="two.sided")
> poder

approximate correlation power calculation (arctangh transformation)

      n = 20
      r = 0.5
sig.level = 0.05
  power = 0.6378746
alternative = two.sided
```

O valor que retorna para o poder é de **0,64**, portanto, a probabilidade de detectar o tamanho de efeito estabelecido, se ele realmente existir, é de 0,64.

3 Considerações Finais

Existem diversos *softwares* que abordam cálculo de tamanho de amostra e poder. O objetivo principal desta monografia foi a criação de um material sobre cálculos de tamanho de amostra e poder, disponibilizados pelo software R, acessível a qualquer indivíduo que tenha interesse no uso desta metodologia, buscando incentivar estes cálculos que, muitas vezes, são deixados de lado.

Ao longo deste trabalho foram analisados quatro pacotes com diversas funções para desenvolvimento de algumas técnicas mais recorrentes na estatística. Para cada técnica, foram apresentados conceitos básicos e construídos exemplos práticos posteriormente aplicados nos pacotes possíveis. Algumas dificuldades foram encontradas na obtenção de maiores detalhes sobre como funcionam algumas funções, sendo por vezes necessária a consulta a referências externas, não bastando apenas a documentação dos pacotes.

A contextualização e aplicação em diversas áreas e a utilização de exemplos práticos visou tornar este material acessível. De maneira geral, espera-se que esta monografia seja útil para auxiliar os pesquisadores a superar a barreira do cálculo de tamanho de amostra e poder e tornar tal prática corriqueira entre pesquisas.

Ressalta-se novamente, por fim, a importância do dimensionamento correto da amostra nas pesquisas visando a obtenção de resultados confiáveis e diminuição de mau uso dos recursos, seja em questões financeiras ou até mesmo éticas, como por exemplo mais pacientes do que necessário a algum tipo de tratamento, seres vivos a testes invasivos, etc.

Visa-se em trabalhos futuros a expansão dessa monografia compreendendo outros tipos de amostragens, tais como amostragem estratificada e por conglomerados, bem como outras técnicas estatísticas mais específicas. Alguns pacotes do software R não contemplados nesta monografia ficam como sugestão também para complementação e ampliação do trabalho. São estes:

- **HMISC**, pacote de miscelâneas, conta com algumas funções de tamanho de amostra e poder, tais como: `bpower` (Poder e tamanho de amostra para o teste Binomial para duas amostras), `cpower` (Poder para o teste Cox/log-rank para duas amostras), `popower` (Poder e tamanho de amostra para Resposta Ordinal) e `samplesize.bin` (Tamanho de amostra para o teste Binomial para duas amostras).

- ***PowerSurvEpi***, conta com funções para calcular o poder e tamanho de amostra para testar o efeito principal ou de interação em modelos de análise de sobrevivência, tais como: `powerCT` (Cálculo de poder na Análise de Sobrevivência em ensaios clínicos), `powerEpi` (Cálculo de poder para Regressão de Cox com Riscos Proporcionais com duas covariáveis) e `ssize.stratify` (Tamanho de amostra para Análise de Sobrevivência com preditor binário e função de Sobrevivência Exponencial).
- ***Longpower***, conta com algumas funções para cálculo de tamanho de amostra e poder para dados longitudinais, tais como: `lmpower` (Tamanho de amostra para modelos lineares mistos para taxa de mudança) e `power.mmrn` (Tamanho de amostra para modelos lineares mistos com medidas repetidas).

4 Pacotes do R

4.1 *samplingbook*

Foi desenvolvido por Juliane Manitz, contribuição de Mark Hempelmann, Goeran Kauermann, Helmut Kuechenhoff, Shuai Shao, Cornelia Oberhauser, Nina Westerheide, Manuel Wiesenfarth no ano de 2017.

Este pacote é baseado nos procedimentos de cálculo de tamanho amostral do livro 'Stichproben - Methoden und praktische Umsetzung mit R' de Goeran Kauermann e Helmut Kuechenhoff (2010).

Link para acesso: <https://cran.r-project.org/web/packages/samplingbook/samplingbook.pdf>

4.2 *pwr*

Foi desenvolvido por Stephane Champely, contribuição de Claus Ekstrom, Peter Dalgaard, Jeffrey Gill, Stephan Weibelzahl, Aditya Anandkumar, Clay Ford, Robert Volcic, Helios De Rosario no ano de 2017.

Este pacote conta com funções que calculam tamanho de amostra e poder seguindo as linhas de [Cohen \(1988\)](#).

Link para acesso: <https://cran.r-project.org/web/packages/pwr/pwr.pdf>

4.3 *pwr2*

Foi desenvolvido por Pengcheng Lu, Junhao Liu e Devin Koestler no ano de 2017.

Este pacote conta com funções para cálculo de tamanho de amostra diferenciados e poder para modelos ANOVA balanceados.

Link para acesso: <https://cran.r-project.org/web/packages/pwr2/pwr2.pdf>

4.4 *TrialSize*

Foi desenvolvido por Ed Zhang, Vicky Qian Wu, Shein-Chung Chow e Harry G. Zhang no ano de 2013.

Este pacote conta com mais de 80 funções entre tamanho de amostra e diversos temas, porém todos relacionados à área da Pesquisa Clínica.

Usa como referência o livro [Chow et al. \(2003\)](#).

Link para acesso: <https://cran.r-project.org/web/packages/TrialSize/TrialSize.pdf>

Referências Bibliográficas

- Abramson, J., Peritz, E., e Gahlinger, P. (1993). Winpepi: Windows programs for epidemiologists. <http://www.brixtonhealth.com/pepi4windows.html>.
- Agranonik, M. e Hirakata, V. (2011). Cálculo de tamanho de amostra: proporções. *Revista HCPA*, páginas 382–388.
- Bolfarine, H. e Bussab, W. O. (2005). *Elementos de Amostragem*. ABE.
- Callegari-Jacques, S. (2003). *Bioestatística Princípios e Aplicações*, volume 1. Art-med.
- Castro e Souza, P. (2012). Tamanho de amostra e poder para três testes não-paramétricos. *LUME UFRGS*.
- Chow, S.-C., Shao, J., e Wang, H. (2003). *Sample Size Calculations in Clinical Research*, volume 1. Marcel Dekker.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, volume 2. Lawrence Erlbaum Associates.
- Guewehr, K. (2004). Ensaio de bioequivalência: do planejamento à análise. *LUME UFRGS*.
- Milone, G. (2009). *Estatística Geral e Aplicada*, volume 1. Centage Learning.
- Minitab (2017). *Suporte ao Minitab 18*. Minitab.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Silva, N. (1998). *Amostragem probabilística: um curso introdutório*, volume 18. Edusp.
- Tavares, A. P. (2008). Qual deve ser o tamanho da minha amostra? *Einstein: Educação Continuada em Saúde*, páginas 153–154.
- USP (1995). Laboratório de epidemiologia e estatística - lee. <http://www.lee.dante.br>.
- Zar, J. (1999). *Biostatistical analysis*, volume 1. Pearson Education India.