

O CÁLCULO AMOSTRAL

é um fantasma para você?

Se a sua resposta for SIM... vem cá que vamos tentar te ajudar a superar esse filme de terror!



O Halloween ainda não chegou, mas esse é o pesadelo de muitos pesquisadores por aí, não é mesmo?

E para auxiliar você, a CEUA/UFRJ está lançando...



O novo Manual de Orientações para Cálculo de Tamanho Amostral e Poder Estatístico

O cálculo de tamanho amostral e poder estatístico é essencial para a autorização de experimentos em animais pela Comissão de Ética no Uso de Animais - CEUA/UFRJ, pois a ausência desses cálculos pode resultar em experimentos ineficazes e expor animais a procedimentos desnecessários.

4 parâmetros básicos são necessários para calcular o tamanho de amostra recomendado para um experimento:

1

O poder estatístico desejado

2

O limiar de significância utilizado

3

O tamanho do efeito que o estudo planeja detectar.

4

A variabilidade esperada da amostra (i.e., desvio-padrão ou variância dos grupos), na mesma unidade em que a diferença foi expressa.

QUE PARÂMETROS SÃO ESSES?

Poder estatístico

Mede a probabilidade de um teste estatístico detectar um efeito na amostra estudada como significativo (i.e., com um valor de p abaixo do limiar de significância), quando este efeito de fato existe na população.

Limiar de significância

Corresponde à taxa de resultados falso-positivos de um determinado desenho experimental caso o efeito a ser detectado não exista (i.e., seja exatamente 0), um caso em que se costuma dizer que “a hipótese nula é verdadeira”.

Tamanho de Efeito

Este efeito se refere ao desfecho primário de seu experimento – ou seja, à medida de maior interesse no estudo.

Como determinar o tamanho de efeito a ser detectado?

Um experimento é projetado para estimar o tamanho de efeito populacional, ou seja, o efeito "real" de um tratamento em uma população teórica. No entanto, o tamanho de efeito encontrado em uma amostra será apenas uma estimativa imperfeita devido ao erro amostral. É necessário ter uma ideia do tamanho de efeito desejado para detectar no experimento, semelhante a escolher uma ferramenta adequada para uma tarefa específica. Para isso, pode-se consultar estudos anteriores ou realizar um estudo piloto, levando em consideração a relevância biológica do efeito mínimo de interesse.

A Comissão de Ética no Uso de Animais (CEUA) recomenda basear o cálculo do tamanho amostral na percepção do tamanho de efeito relevante, em vez de uma estimativa do efeito esperado.

Como determinar a variabilidade esperada?

Ao estipular o tamanho de efeito a ser detectado, seja como uma diferença absoluta ou relativa, é essencial acompanhar essa estimativa com uma medida de variabilidade, como o desvio-padrão esperado em cada grupo. Isso permite que a diferença seja expressa de forma padronizada, necessária para o cálculo de poder e para testes estatísticos. Para obter o desvio-padrão esperado, recomenda-se consultar resultados prévios, preferencialmente do mesmo modelo usado no experimento, seja no próprio laboratório ou na literatura. Se não houver estimativas disponíveis, pode-se realizar um experimento piloto, considerando que a precisão da estimativa dependerá do tamanho da amostra. É crucial que tanto a diferença entre médias quanto o desvio-padrão estejam na mesma unidade de medida.

Como realizar o cálculo?

Ferramentas e Softwares Recomendados:

Estatística
Bauru



<http://estatistica.bauru.usp.br/calculoamostral/calculos.php>

Statulator



<https://www.statulator.com/>

UCSF



<http://www.stat.ubc.ca/~rollin/stats/ssize/>

GPower



<http://www.gpower.hhu.de/>

Recomendamos ainda as seguintes referências disponíveis online:

<https://doi.org/10.1525/collabra.33267>

<https://doi.org/10.1136/bmj.306.6886.1181>

<https://doi.org/10.3352%2Fjeehp.2021.18.17>

**A CEUA possui profissionais que podem ajudá-lo.
Para esclarecer dúvidas, entre em contato:**

ceua@ccsdecania.ufrj.br

Esperamos ter ajudado a transformar o Fantasma do Cálculo Amostral em um Anjinho de Paz!

Conte com a CAMBE!

Entre em contato conosco para mais informações ou suporte técnico.

Estamos aqui para ajudar você a alcançar resultados precisos e confiáveis em seus experimentos.



Cálculo de tamanho amostral e poder estatístico: orientações

Seguindo a tendência de diversos órgãos reguladores ao redor do mundo, a CEUA-CCS exige a realização de cálculos de tamanho amostral para a autorização de experimentos em animais.

A razão é simples: a precisão de um experimento para estimar os efeitos ou parâmetros de interesse é dependente do tamanho amostral. Caso testes estatísticos sejam usados para testar uma hipótese, as probabilidades de resultados falsos-negativos e falsos-positivos também são dependentes do tamanho da amostra. Desta forma, um projeto sem um cálculo de tamanho amostral expõe animais a experimentos inapropriados para responder à pergunta a que ele se propõe.

De forma geral, cálculos de tamanho amostral podem ser realizados com base no *poder estatístico desejado* (quando o experimento será analisado através de um teste estatístico que tem como resultado um valor p) ou na *precisão de uma estimativa* (quando o experimento visa estimar um tamanho de efeito com um determinado intervalo de confiança). O primeiro caso é mais comum em experimentos em animais, e será detalhado inicialmente.

1. Cálculo de tamanho amostral baseado em poder estatístico

1.1. Parâmetros Necessários para o Cálculo

Três parâmetros básicos são necessários para calcular o tamanho de amostra recomendado para um experimento:

1. O *poder estatístico desejado*
2. O *limiar de significância utilizado*
3. O *tamanho do efeito* que o estudo planeja detectar.

Caso o tamanho de efeito seja expresso como uma medida padronizada (e.g. d de Cohen ou r), estes parâmetros são suficientes. Caso ele seja expresso como uma diferença absoluta (e.g. na unidade em que a medida está sendo realizada) ou relativa (e.g. em porcentagem do controle), é necessário informar também:

4. A *variabilidade* esperada da amostra (i.e., desvio-padrão ou variância dos grupos), na mesma unidade em que a diferença foi expressa.

1.2. O que são estes parâmetros (e como defini-los)

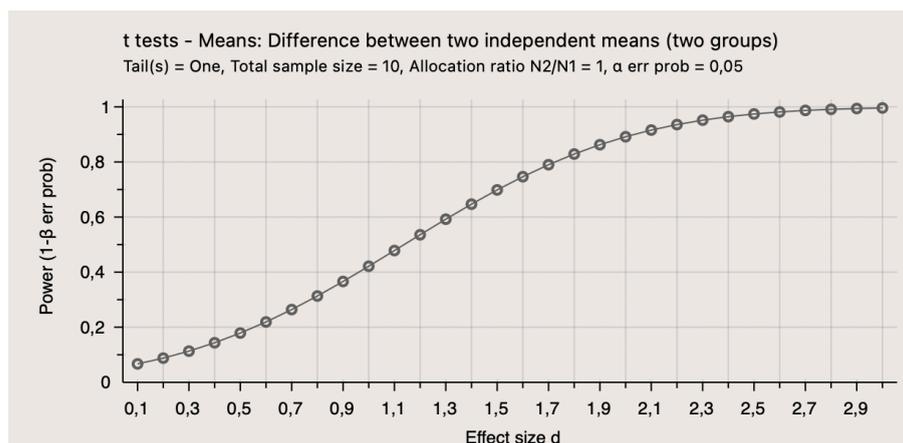
Poder estatístico

O poder estatístico mede a probabilidade de um teste estatístico detectar um efeito na amostra estudada como significativo (i.e., com um valor de p abaixo do limiar de significância), quando este efeito de fato existe na população.

Naturalmente, esta probabilidade depende do *tamanho real do efeito* – ou seja, o quão grande é o efeito na população em estudo. Efeitos muito pequenos (como os de uma manipulação comportamental ou um tratamento farmacológico

em dose muito baixa) serão difíceis de detectar mesmo com tamanhos de amostra grandes, enquanto efeitos muito grandes (e.g. o de um tratamento tóxico em dose alta) podem ser detectados mesmo em experimentos com amostras pequenas.

Esta dependência pode ser observada no exemplo abaixo. O experimento em questão (com um n de 10/grupo) tem um poder estatístico de quase 100% para detectar um efeito com d de Cohen = 3 (ou seja, uma diferença de 3 desvios-padrão entre dois grupos), mas de apenas 20% para detectar um tamanho de efeito de 0,5.



O poder a ser utilizado no cálculo de tamanho amostral depende da sensibilidade que você pretende ter para encontrar um determinado tamanho de efeito. Caso você opte por um poder de 90%, por exemplo, você está assumindo um risco de 10% do experimento apresentar um resultado negativo (acima do limiar de significância escolhido), caso o efeito real tenha o tamanho que foi assumido.

De forma geral, recomendamos que experimentos tenham poder estatístico de pelo menos 80% para detectar um tamanho de efeito julgado relevante pelos pesquisadores, já que números abaixo disso geram taxas de resultados falso-negativos muito altas. Dito isso, há diversas situações em que um poder mais alto pode ser desejável.

Limiar de significância

O limiar de significância corresponde à taxa de resultados falso-positivos de um determinado desenho experimental *caso o efeito a ser detectado não exista* (i.e., seja exatamente 0), um caso em que se costuma dizer que “a hipótese nula é verdadeira”.

A pesquisa biomédica costuma usar o valor de 5% como limiar de significância (o que significa admitir que 1 em cada 20 experimentos gerará um resultado falso-positivo, caso não exista um efeito real). Dito isso, este não é um critério particularmente rigoroso, e existem diversas situações em que um limiar mais estrito pode ser desejável.

Exemplos de circunstâncias em que um limiar de significância mais estrito deve ser considerado incluem (a) experimentos em cujas conclusões se deseja ter

um grau de convicção mais alto do que o usual, (b) experimentos em que muitas variáveis são medidas, levando a um potencial maior de falsos-positivos e (c) experimentos com hipóteses improváveis, em que um grau de evidência maior será necessário para aceitá-las.

Para um mesmo desenho experimental e tamanho amostral, quanto mais estrito for o limiar de significância, menor será o poder estatístico, já que um valor de p mais baixo será necessário para considerar um efeito como significativo.

Tamanho de Efeito

Para definir este parâmetro, é importante ter claro *qual o efeito que você pretende identificar*, e como ele será medido.

De forma geral, este efeito se refere ao desfecho primário de seu experimento – ou seja, à medida de maior interesse no estudo. Caso você tenha múltiplos desfechos, cabe ao pesquisador decidir entre (a) realizar o cálculo amostral para cada um deles e utilizar um tamanho de amostra capaz de garantir um poder adequado para todos os desfechos, ou (b) usar o cálculo para o desfecho mais importante, assumindo o risco de que haja um poder menor para outros desfechos.

Diferentes tipos de experimento podem apresentar medidas de efeito distintas. Em experimentos comparando médias de uma variável quantitativa entre dois grupos (o caso mais comum em experimentação animal), a medida de efeito costuma ser uma *diferença entre médias*. Ela pode ser expressa:

- Em unidades absolutas (e.g. 5 kg).
- Em unidades relativas (e.g. 20% do controle).
- Em unidades padronizadas (e.g. 0,5 desvios-padrão, ou $d=0,5$).

O tamanho de efeito relevante para o cálculo amostral é o tamanho de efeito padronizado (e.g. d de Cohen). Dito isso, unidades absolutas ou relativas são uma forma mais intuitiva de pensar numa diferença. Caso você queira basear seu raciocínio nelas (o que geralmente é recomendável), será necessário adicionar como parâmetro *a variabilidade da amostra* (e.g. o desvio-padrão esperado entre os grupos), expresso na mesma unidade da diferença.

1.3. Como determinar o tamanho de efeito a ser detectado

De forma geral, um experimento é planejado para estimar um tamanho de efeito populacional (e.g. o efeito “real” de um tratamento em uma população teórica do animal em estudo), que é o efeito que nos interessa para o cálculo. O tamanho de efeito encontrado em uma amostra de animais será uma estimativa imperfeita deste efeito real, por conta do erro amostral.

Naturalmente, você não tem como saber *qual é* o tamanho de efeito real – afinal, você está fazendo seu experimento justamente para estimá-lo! Dito isso, você tem que ter uma ideia do tamanho de efeito que *gostaria de detectar* para pensar no experimento que será necessário.

Fazendo uma analogia, seu experimento é uma ferramenta para detectar um efeito, e é preciso saber se a ferramenta tem precisão suficiente. A situação é análoga a escolher outras ferramentas – uma balança de farmácia terá precisão adequada para pesar uma pessoa, mas não um camundongo. E ainda que você não saiba o peso exato do que quer pesar de antemão, você precisa de uma ideia da ordem de grandeza a ser medida para selecionar uma balança adequada.

Uma forma de estimar esta ordem de grandeza para seu tamanho de efeito é buscar estudos prévios que tenham realizado medidas em circunstâncias semelhantes. Se você está buscando um efeito anti-hipertensivo de um novo fármaco, por exemplo, pode se basear no efeito de fármacos existentes no mesmo modelo – seja em dados publicados, seja em seu próprio laboratório.

Outra forma de pensar neste problema é pensar qual o efeito mínimo de interesse – ou seja, no tamanho de efeito que você *não gostaria de deixar de detectar, caso ele exista*. Caso você julgue que reduções na pressão arterial abaixo de 10% são biologicamente irrelevantes, por exemplo, você deveria utilizar um valor de pelo menos 10% em seu cálculo.

Caso não haja nenhum dado prévio para permitir a estimativa do tamanho de efeito, uma última opção é realizar um estudo piloto. Dito isso, deve-se levar em conta que estudos pilotos com amostras pequenas possuem uma margem grande de erro em suas estimativas. De forma geral, a CEUA recomenda que os pesquisadores baseiem o cálculo de tamanho amostral em sua percepção de qual o tamanho de efeito *relevante* (i.e. aquele que não deveria deixar de ser detectado), mais do que em uma estimativa do tamanho de efeito *esperado* (que pode ser difícil de prever).

1.4. Como determinar a variabilidade esperada

Como colocado acima, caso você estipule o tamanho de efeito a ser detectado como uma diferença absoluta (e.g. uma redução de 5 g) ou relativa (e.g. uma redução de 10% no peso corporal), ele deve ser acompanhado de uma estimativa de variabilidade – ou seja, o desvio-padrão esperado em cada grupo. Isso é necessário para que a diferença possa ser expressa de forma *standardizada* (i.e. em número de desvios-padrão), que é o que o cálculo de poder e o teste estatístico levarão em conta.

Para isso, novamente é necessário algum conhecimento prévio sobre o comportamento da variável que você está estudando. A estratégia mais comum para estimar o desvio-padrão esperado é olhar para resultados prévios utilizando o mesmo modelo, idealmente em seu próprio laboratório. De forma geral, uma estimativa de variabilidade no grupo controle é suficiente para este fim. É claro que a intervenção em estudo (e.g. um tratamento farmacológico) pode afetar a variabilidade, mas isso é difícil de prever de antemão.

Caso você nunca tenha trabalhado com o modelo em uso, é possível utilizar resultados da literatura com o mesmo modelo, que geralmente podem fornecer alguma referência. No caso de você não possuir nenhuma estimativa de variabilidade – e.g. um modelo experimental ou instrumento nunca utilizado antes –, uma última alternativa é fazer um experimento piloto para estimar o

desvio-padrão, com a ressalva de que a margem de erro dessa estimativa dependerá do tamanho da amostra.

Uma ressalva importante é que, caso você use a diferença entre médias e o desvio-padrão em seu cálculo (ver abaixo), ambos devem ser expressos *na mesma unidade*. Caso o efeito a ser detectado seja uma média de 5 gramas, o desvio-padrão deve ser expresso em gramas; caso ele seja uma diferença de 10% em relação ao controle, o desvio-padrão também deve ser expresso em % da média do controle.

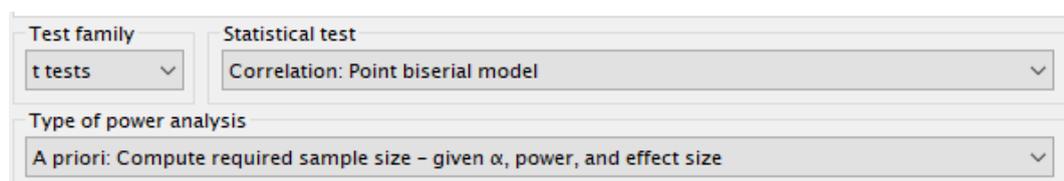
2. Como realizar o cálculo

A realização de cálculos de poder estatístico para a maior parte dos experimentos a serem analisados por estatística paramétrica é simples e pode ser realizada em softwares gratuitos (como o G*Power, disponível em <http://www.gpower.hhu.de/>), calculadoras online (como em <http://estatistica.bauru.usp.br/calculoamostral/calculos.php>, <https://www.statulator.com/> ou <http://www.stat.ubc.ca/~rollin/stats/ssize/>) ou pacotes em linguagem R (como *pwrss* e *superpower*).

Ainda que estes métodos se baseiem na mesma matemática, diferentes testes estatísticos requerem parâmetros distintos para o cálculo. Além disso, cada um dos softwares ou calculadoras acima pode requerer que estes parâmetros (por exemplo, a definição do tamanho da diferença entre os grupos) sejam informados de maneira ligeiramente diferente.

Para ilustrar o processo com um exemplo simples, vamos mostrar como os parâmetros devem ser usados para calcular o tamanho da amostra em alguns casos, utilizando o G*Power 3.1. Mais informações sobre o uso do software para outros tipos de teste podem ser encontradas neste guia aberto em <https://osf.io/zqphw> ou no manual de instruções do software em <https://www.psychologie.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisches-Naturwissenschaftliche-Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf>.

Para iniciar o cálculo, a “família de teste” (*test family*) e o “teste estatístico” (*statistical test*) devem ser selecionados no programa. A “família” envolve tipos gerais de teste (e.g. testes t, ANOVAs, testes de qui-quadrado), enquanto o “teste estatístico” envolve o caso estatístico do teste a ser utilizado.



The image shows a screenshot of the G*Power software interface. It features three dropdown menus for configuration:

- Test family:** Set to "t tests".
- Statistical test:** Set to "Correlation: Point biserial model".
- Type of power analysis:** Set to "A priori: Compute required sample size - given α , power, and effect size".

Além do teste, o programa solicita o tipo de cálculo a ser realizado (“*Type of power analysis*”). No G*Power, você pode optar pelas seguintes opções:

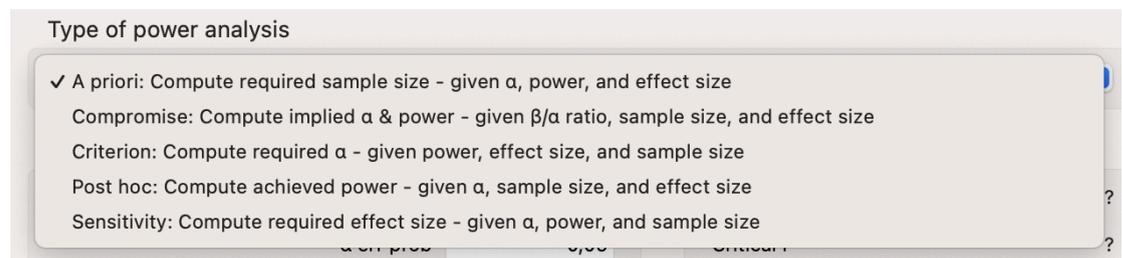
(a) calcular um tamanho de amostra para obter um determinado poder estatístico para detectar um determinado tamanho de efeito (“A priori”)

(b) o poder estatístico de um determinado tamanho de amostra para detectar um determinado tamanho de efeito (“Post-hoc”)

(c) o tamanho de efeito para o qual se tem determinado poder estatístico com um determinado tamanho de amostra (“Sensitivity”).

(d) o valor de alfa a ser utilizado para se obter um determinado poder, dado um tamanho de efeito a ser detectado e o tamanho de amostra (“Criterion”).

(e) os valores de alfa e beta (ou de limiar de significância e poder estatístico) a serem utilizados, dado um tamanho de efeito a ser detectado e o tamanho de amostra (“Compromise”).



Para fins de cálculo de tamanho de amostra durante o planejamento experimental (que é a etapa em que projetos são submetidos à CEUA), o caso (a) é de longe a situação mais comum, então recomendamos que a opção “A priori” seja usada.

Abaixo, discutiremos alguns dos casos mais comuns de cálculos e testes estatísticos utilizados em pesquisa com animais:

2.1. Comparação de uma variável quantitativa entre dois grupos independentes

Um teste t para amostras independentes compara as médias de uma variável quantitativa entre dois grupos. Para este caso, selecione “t test” em “family” e “Means: Difference between two independent means (two groups)” em “statistical test”.

Test family		Statistical test	
t tests		Means: Difference between two independent means (two groups)	
Type of power analysis			
A priori: Compute required sample size - given α , power, and effect size			
Input Parameters		Output Parameters	
Determine =>	Tail(s)	One	Noncentrality parameter δ
	Effect size d	0.5	Critical t
	α err prob	0.05	Df
	Power ($1-\beta$ err prob)	0.95	Sample size group 1
	Allocation ratio $N2/N1$	1	Sample size group 2
		Total sample size	
		Actual power	

Os parâmetros de análise devem ser definidos nos campos acima:

“*Type of power analysis*” – Para fins de cálculo de tamanho de amostra, normalmente a opção “A priori” será usada.

“*Tail(s)*” – Define se o teste t será unicaudal (“One”) – ou seja, testará uma hipótese unidirecional – ou bicaudal (“Two”) – ou seja, testará a presença de um efeito em ambos os sentidos. Via de regra, consideramos que a opção de um teste bicaudal costuma ser a mais adequada, a não ser que os pesquisadores julguem que não há interesse em detectar um efeito no sentido oposto ao esperado.

“*Effect size d* ” – É o tamanho de efeito para o qual se está realizando o cálculo de tamanho amostral, expresso em d de Cohen. Caso você prefira inserir o tamanho da diferença em uma escala absoluta ou relativa (o que costuma ser mais intuitivo), clique em “*Determine*” para abrir uma nova janela que permite essa opção (ver descrição abaixo).

“ *α err prob*” – É o valor de α , que corresponde ao limiar de significância a ser utilizado (tipicamente 0,05, mas podendo ser mais estrito caso os pesquisadores desejem ser mais rigorosos).

“*Power ($1-\beta$ err prob)*” – É o poder estatístico desejado. Um valor comumente utilizado é 80%, mas existem diversos casos em que um poder maior pode ser desejável.

“*Allocation ratio $N2/N1$* ” – É a razão entre os tamanhos amostrais de ambos os grupos. Idealmente, o melhor poder estatístico de um estudo será obtido com grupos experimentais do mesmo tamanho (neste caso, use 1 neste campo). Se por algum motivo isso não for viável, você pode alterar o parâmetro inserindo a razão entre o n dos dois grupos. Nesse caso, o resultado apresentará tamanhos amostrais diferentes para cada grupo, respeitando essa razão.

Caso você *não* queira preencher o tamanho de efeito em Cohen's *d* e sim como uma diferença absoluta ou relativa (o que geralmente consideramos a opção mais recomendável), clique em *Determine* para abrir a janela abaixo

The image shows a software interface for calculating effect size *d*. It features two radio button options: *n1 != n2* and *n1 = n2*. The *n1 = n2* option is selected. Under the *n1 != n2* option, there are input fields for Mean group 1 (0), Mean group 2 (1), and SD σ within each group (0.5). Under the *n1 = n2* option, there are input fields for Mean group 1 (0), Mean group 2 (1), SD σ group 1 (0.5), and SD σ group 2 (0.5). At the bottom, there is a 'Calculate' button, an 'Effect size *d*' field with a question mark, and a 'Calculate and transfer to main window' button.

Nesta janela, você deve especificar o tamanho da diferença que deseja detectar, inserindo as médias esperadas para os dois grupos caso esse valor de diferença exista (e.g. *Mean group 1* e *Mean group 2*).

Por exemplo, se você está projetando um experimento capaz de detectar uma redução de 50% em um determinado parâmetro em relação ao grupo controle (cuja média esperada é de 40 mg/dL), você pode inserir 40 no grupo 1 (controle) e 20 no grupo 2 (tratado). Você também pode inserir os valores como diferença relativa (e.g. 100 (%) para o controle e 50 (%) para o grupo tratado), mas lembre-se de que neste caso o desvio padrão deve ser inserido na mesma unidade.

Na(s) próxima(s) linha(s) (*SD σ group 1* e *SD σ group 2*), você deve inserir o desvio-padrão para os dois grupos. Você pode usar a mesma estimativa para ambos os grupos (uma situação comum), mas caso tenha uma razão para prever que a variação num dos grupos seja maior (e.g. um tratamento ao qual a resposta seja heterogênea), pode inserir valores diferentes para cada grupo se ambos tiverem o mesmo tamanho amostral (opção *n1 = n2*). Para casos em que o *n* é diferente (opção *n1 \neq n2*), o software não recomenda esse procedimento e solicita um único valor.

Novamente, lembre-se de que o desvio padrão deve estar na mesma unidade da média. Com isso, se você espera um desvio-padrão de 10 mg/dL no caso acima (em que os valores do grupo controle e tratado são 40 mg/dL e 20 mg/dL), pode inserir todos os valores nesta unidade. Mas no caso de ter inserido os valores das médias em % (i.e. 100 e 50), o desvio-padrão deve ser inserido na mesma unidade (no caso, 25% do valor da média).

Após inserir os valores, clique em "*calculate and transfer to main window*" para que o campo "*Effect size *d**" seja calculado automaticamente a partir dos valores inseridos. Conforme mostra a figura abaixo, o tamanho de efeito em *d* (número

de desvios-padrão) no exemplo acima (uma diferença de 20 mg/dL para um desvio-padrão de 10 mg/dL) é de 2.

The screenshot shows a software interface for power analysis. It is divided into several sections:

- Type of power analysis:** A dropdown menu set to "A priori: Compute required sample size - given α , power, and effect size".
- Input parameters:** Includes a "Determine" button, "Tail(s)" set to "Two", "Effect size d" set to 2, " α err prob" set to 0,05, "Power (1- β err prob)" set to 0,8, and "Allocation ratio N2/N1" set to 1.
- Output parameters:** A list of parameters with question marks next to them: Noncentrality parameter δ , Critical t, Df, Sample size group 1, Sample size group 2, Total sample size, and Actual power.
- Group Settings:** Two radio buttons for "n1 \neq n2" and "n1 = n2". The "n1 = n2" option is selected. Below it, input fields for "Mean group 1" (40), "Mean group 2" (20), "SD σ group 1" (10), and "SD σ group 2" (10) are visible.
- Buttons:** "Calculate", "Calculate and transfer to main window", and "Close effect size drawer".

Após isso, clique em “*calculate*” na janela principal.

Na imagem abaixo, vemos que nesse exemplo o tamanho da amostra para obter um poder de 80% seria de 6 animais para cada grupo (ou 12 no total).

This screenshot shows the same software interface after the calculation has been performed. The results are now displayed in the output parameters section:

- Test family:** "t tests".
- Statistical test:** "Means: Difference between two independent means (two groups)".
- Type of power analysis:** "A priori: Compute required sample size - given α , power, and effect size".
- Input parameters:** Same as the previous screenshot.
- Output parameters:** Now filled with numerical values: Noncentrality parameter δ (3,4641016), Critical t (2,2281389), Df (10), Sample size group 1 (6), Sample size group 2 (6), Total sample size (12), and Actual power (0,8764178).

Se em um exemplo semelhante (veja a imagem abaixo), quiséssemos detectar um tamanho de efeito menor (por exemplo, uma diferença de 10 mg/dL, ou 1 desvio-padrão, com os demais parâmetros iguais, vemos que o tamanho de amostra necessário para obter um poder de 80% seria maior (17 animais para cada grupo).

Type of power analysis

A priori: Compute required sample size - given α , power, and effect size

Input parameters

Tail(s) Two

Determine

Effect size d 1

α err prob 0,05

Power (1- β err prob) 0,8

Allocation ratio N2/N1 1

Output parameters

Noncentrality parameter δ 2,9154759

Critical t 2,0369333

Df 32

Sample size group 1 17

Sample size group 2 17

Total sample size 34

Actual power 0,8070367

$n1 \neq n2$

Mean group 1 40

Mean group 2 30

SD σ within each group 10

$n1 = n2$

Mean group 1 40

Mean group 2 30

SD σ group 1 10

SD σ group 2 10

Calculate Effect ?

Calculate and transfer to main window

Close effect size drawer

Se em um terceiro exemplo (ver imagem abaixo), tivermos uma diferença entre médias semelhante ao caso original, mas um desvio-padrão maior, vemos que o mesmo efeito acontece, já que o parâmetro de interesse (o tamanho de efeito em d) representa a razão entre a diferença e o desvio-padrão.

Type of power analysis

A priori: Compute required sample size - given α , power, and effect size

Input parameters

Tail(s) Two

Determine

Effect size d 1

α err prob 0,05

Power (1- β err prob) 0,8

Allocation ratio N2/N1 1

Output parameters

Noncentrality parameter δ 2,9154759

Critical t 2,0369333

Df 32

Sample size group 1 17

Sample size group 2 17

Total sample size 34

Actual power 0,8070367

$n1 \neq n2$

Mean group 1 40

Mean group 2 30

SD σ within each group 10

$n1 = n2$

Mean group 1 40

Mean group 2 20

SD σ group 1 20

SD σ group 2 20

Calculate Effect 1

Calculate and transfer to main window

Close effect size drawer

2.2. Comparação de uma variável quantitativa entre dois grupos pareados

Caso o teste estatístico a ser utilizado seja pareado (e.g. analisando as diferenças individuais entre pares, como no caso de comparações entre dois pontos do tempo num mesmo animal, ou de animais/amostras tratado(s) que possuem controles específicos), selecione “Means: Difference between two dependent means (matched pairs)” em “statistical test”.

O procedimento neste caso é semelhante ao do teste t para amostras independentes, com os mesmos parâmetros sendo inserido (ver item 2.1). Mas para calcular um tamanho de efeito no caso pareado, você precisará adicionalmente informar a *correlação esperada* entre os indivíduos (e.g. o grau de correlação entre as medidas de indivíduos de um mesmo par).

Test family: t tests
Statistical test: Means: Difference between two dependent means (matched pairs)

Type of power analysis: A priori: Compute required sample size - given α , power, and effect size

Input parameters:
Tail(s): Two
Effect size dz: 1
 α err prob: 0,05
Power (1- β err prob): 0,8

Output parameters:
Noncentrality parameter δ : 3,1622777
Critical t: 2,2621572
Df: 9
Total sample size: 10
Actual power: 0,8030969

From differences: Mean of difference: 0, SD of difference: 1

From group parameters: Mean group 1: 40, Mean group 2: 20, SD group 1: 20, SD group 2: 20, Correlation between groups: 0,5

Calculate: Effect size dz: 1

Note que nesse caso o valor de tamanho amostral informado pelo programa (“total sample size”) se refere ao *número de pares* (ou seja, o número de animais/amostras individuais será o dobro deste número).

Cabe notar também que, quanto maior a correlação entre os pares (ou seja, quanto mais efetivo for o pareamento), maior será o tamanho de efeito estandardizado, e menor o tamanho amostral necessário para atingir o poder estatístico. Caso a correlação esperada aumente de 0,5 para 0,8, por exemplo, o número de pares necessário para detectar o tamanho de efeito acima com poder de 80% cai para 6.

Test family: t tests
Statistical test: Means: Difference between two dependent means (matched pairs)

Type of power analysis: A priori: Compute required sample size - given α , power, and effect size

Input parameters:
Tail(s): Two
Effect size dz: 1,581139
 α err prob: 0,05
Power (1- β err prob): 0,8

Output parameters:
Noncentrality parameter δ : 3,8729838
Critical t: 2,5705818
Df: 5
Total sample size: 6
Actual power: 0,8677077

From differences: Mean of difference: 0, SD of difference: 1

From group parameters: Mean group 1: 40, Mean group 2: 20, SD group 1: 20, SD group 2: 20, Correlation between groups: 0,8

Calculate: Effect size dz: 1,581139

Outra opção para calcular o poder estatístico de um teste pareado é informar o *desvio-padrão das diferenças entre pares* ao invés do desvio-padrão de cada grupo (opção “From differences” na janela de tamanho de efeito), mas isso geralmente representa uma alternativa menos intuitiva e mais difícil de estabelecer *a priori*.

2.3. Comparação de uma variável quantitativa entre mais de dois grupos

No caso da comparação de uma variável quantitativa entre mais de dois grupos com tratamentos distintos, o procedimento mais comum em pesquisa com

animais é utilizar uma análise de variância (ANOVA) seguida de uma comparação pós-hoc entre grupos individuais, em que o limiar de significância de cada comparação é ajustado conforme o número de comparações através de métodos como os de Bonferroni, Sidak ou Tukey.

Calcular o tamanho amostral para uma ANOVA de uma via envolve expressar o tamanho de efeito através de um valor de f (a razão entre o desvio-padrão entre as médias e o desvio-padrão entre as unidades) ou de η^2 (a porcentagem da variância da variável dependente que é explicada pelo grupo), o que não é muito intuitivo para a maior parte dos pesquisadores. Dado isso, o G*Power permite que você calcule um valor de f a partir das médias esperadas nos vários grupos, facilitando essa estimativa.

Para uma ANOVA de uma via, escolha “F tests” em “Test Family” e “ANOVA: fixed effects, omnibus, one-way” em “Statistical Test”. Após informar o poder desejado, limiar de significância e número de grupos, clique em “Determine” para gerar um valor de f .

Group	Mean	Size
1	40	10
2	40	10
3	40	10
4	20	10

Na caixa da direita, você deve informar o desvio-padrão esperado e a média de cada grupo. O tamanho de cada grupo importa apenas para determinar a proporção entre os diferentes grupos (já que o número exato é justamente o que você está tentando calcular).

Após calcular o tamanho de efeito e transferi-lo para a janela principal, clique em “calculate” para calcular o tamanho da amostra, que nesse caso será informado como o *número total de indivíduos*. No caso de grupos de tamanho semelhante (que costuma ser a melhor opção para maximizar o poder estatístico), ele deve ser dividido pelo número de grupos para chegar ao número de animais por grupo (por exemplo, 96 animais em 4 grupos representam 24 animais por grupo).

Dito isso tudo, o caso mais frequente em uma ANOVA de uma via é que as comparações individuais entre grupos através do teste pós-hoc interessem mais a você do que o resultado da ANOVA em si. Como esses testes costumam ter um poder estatístico menor do que o da ANOVA (e portanto necessitar de um

tamanho de amostra maior), costuma ser mais relevante calcular o tamanho de amostra com base no teste pós-hoc – que normalmente é uma versão modificada do teste t , com alguma forma de ajuste para múltiplas comparações. Para este fim, você deverá retornar ao caso de uma comparação entre dois grupos (conforme descrito na seção 2.1).

Uma heurística simples neste caso é realizar o cálculo para uma comparação entre dois grupos (vide a seção 2.1), mas corrigindo o limiar de significância através da correção da Bonferroni – o que equivale a dividir o valor de alfa informado ao programa pelo *número de comparações entre grupos a serem realizadas*. Note que este número é diferente do número de *grupos* do experimento: um experimento com 4 grupos, por exemplo, requer 6 comparações para que todos eles sejam comparados entre si. Neste caso, para atingir um α de 0,05 para o conjunto das comparações (o dito *family-wise error rate*), o limiar de significância das comparações individuais deve ser atualizado para 0,008.

Outros testes pós-hoc tipicamente usados (e.g. Šidák, Tukey, Dunnett) costumam fornecer poder estatístico maior do que o teste de Bonferroni, o que significa que o cálculo acima deve fornecer poder suficiente para qualquer um deles.

Independente do teste usado, cabe ressaltar que, quanto mais comparações entre grupos são realizadas, mais estrito se torna o limiar de significância, e portanto maior o tamanho amostral necessário por grupo para que o poder estatístico seja mantido. Desta forma, limitar o número de comparações a serem realizadas àquelas que são de fato informativas é sempre recomendado, já que permite um poder estatístico melhor com um número menor de animais).

2.4. Comparação de uma variável quantitativa entre mais de dois grupos em um desenho fatorial

Inúmeros desenhos experimentais utilizam *mais de uma variável independente* na definição dos grupos. Um experimento, por exemplo, pode estar interessado em comparar o efeito de um tratamento e seu veículo em animais machos e fêmeas. Neste caso, temos 4 grupos experimentais (machos tratados, machos controle, fêmeas tratadas e fêmeas controle), mas o desenho experimental pode ser melhor descrito como um desenho fatorial 2x2 (i.e. duas variáveis independentes – tratamento e sexo –, cada qual com 2 categorias).

Casos como este são comumente analisados em pesquisa de laboratório por ANOVAs de múltiplas vias, que avaliam separadamente o efeito de cada uma das variáveis independentes (“vias”), bem como de sua interação. No exemplo acima, por exemplo, uma ANOVA de 2 vias testaria o efeito do tratamento (analisando machos e fêmeas em conjunto) e o efeito do sexo (analisando animais tratados e controles em conjunto) sobre a variável de interesse, além da interação entre tratamento e sexo (e.g. até que ponto o efeito do tratamento é o mesmo em ambos os sexos).

Desenhos fatoriais são interessantes para analisar mais de uma variável maximizando o poder estatístico de um experimento. No exemplo acima, por exemplo, o efeito do tratamento é medido no conjunto de ambos os sexos em uma ANOVA de duas vias, proporcionando um poder maior do que comparando os grupos individuais entre si.

O G*Power oferece diversas opções para cálculo de poder em desenhos fatoriais – dito isso, todas elas requerem que você especifique o tamanho de efeito como f ou η^2 (a porcentagem da variância explicada pela variável independente de interesse). Caso isso seja intuitivo para você, escolha “F tests” e procure seu caso: uma ANOVA de múltiplas vias sem pareamento, por exemplo (como no caso acima), seria analisada em “ANOVA: Fixed effects, special, main effects and interactions”.

Neste caso, você deve informar o tamanho de efeito, o limiar de significância, o poder estatístico, o número de grupos e os graus de liberdade do denominador (*numerator df*), que corresponde ao número de categorias da variável de interesse – 1 (no caso de um desenho 2 x 2, por exemplo, 2-1 = 1 para ambas as variáveis). No exemplo acima, assim, isso seria preenchido da seguinte forma para um η^2 de 0.3:

The screenshot shows the G*Power software interface. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Fixed effects, special, main effects and interactions'. The 'Type of power analysis' is 'A priori: Compute required sample size - given alpha, power, and effect size'. Under 'Input parameters', 'Effect size f' is 0,6546537, 'alpha err prob' is 0,05, 'Power (1-beta err prob)' is 0,8, 'Numerator df' is 1, and 'Number of groups' is 4. Under 'Output parameters', 'Noncentrality parameter lambda' is 9,0000008, 'Critical F' is 4,4513218, 'Denominator df' is 17, 'Total sample size' is 21, and 'Actual power' is 0,8070465. On the right, the 'Direct' method is selected, showing 'Partial eta squared' as 0,3. The 'Calculate' button is active, and the 'Effect size f' is displayed as 0,6546537.

Caso tamanhos de efeito expressos em f ou η^2 não sejam intuitivos para você – o que costuma ser o caso para a maior parte dos pesquisadores – você tem algumas opções para contornar o problema. Uma delas é simular resultados a partir de médias para cada grupo e desvios-padrão em outro programa e verificar que tipo de η^2 isso gera.

Caso isso ainda pareça complicado, a CEUA considera que considerar cada uma das comparações da ANOVA como uma comparação simples entre dois grupos é uma simplificação razoável. No exemplo acima, por exemplo, caso a variável em estudo mais importante seja o tratamento, pode-se utilizar o tamanho de diferença a ser detectado e o desvio-padrão levando em conta machos e fêmeas

em conjunto, e realizar o cálculo de tamanho amostral para um teste *t* de amostras independentes entre dois grupos como uma aproximação do número necessário para a detecção deste efeito na ANOVA. Nesse caso, é importante ter em mente que o número de animais calculado por grupo corresponderá ao *total de animais tratados* e ao *total de animais controle* – logo, ele deverá ser dividido por dois para chegar ao número de machos e fêmeas necessário em cada um dos grupos.

Note que esta abordagem é válida caso o que importe seja estimar o efeito geral do tratamento (i.e. em ambos os sexos), e não para testar o efeito do tratamento em cada sexo, ou a interação entre tratamento e sexo. Caso seja importante para os autores testar a consistência do efeito dentro de cada sexo (i.e. entre machos e entre fêmeas separadamente), seria mais apropriado calcular o poder estatístico dos testes pós-hoc entre os grupos em questão (vide o item 2.3). Dito isso, é importante lembrar que estes testes terão menos poder estatístico do que a ANOVA em si para testar a mesma pergunta, e requererão um tamanho de amostra maior. Por conta disso, a CEUA entende que, em casos em que a variável de interesse possui apenas duas categorias, usar o resultado da própria ANOVA para chegar a conclusões costuma ser mais apropriado.

2.5. Comparação das proporções de uma variável categórica entre dois grupos

Para a comparação de uma variável categórica binomial entre dois grupos (e.g. comparação do percentual de animais vivos em cada grupo após um tratamento), escolha “Exact” em “Test Family” e “Proportions: inequality, two independent groups (Fisher’s exact test)” em “Statistical test”, o que calcula o tamanho amostral necessário para obter o poder estatístico desejado em um teste exato de Fisher.

Test family		Statistical test	
Exact		Proportions: Inequality, two independent groups (Fisher's exact test)	

Type of power analysis

A priori: Compute required sample size - given α , power, and effect size

Input parameters		Output parameters		
Determine	Tail(s)	Two	Sample size group 1	64
	Proportion p1	0,5	Sample size group 2	64
	Proportion p2	0,75	Total sample size	128
	α err prob	0,05	Actual power	0,8017145
	Power (1- β err prob)	0,8	Actual α	0,0306288
	Allocation ratio N2/N1	1		

O procedimento necessário é similar ao utilizado para variáveis quantitativas em relação à inserção do poder estatístico desejado, do limiar de significância e do *allocation ratio* (e.g. equilíbrio do tamanho amostral entre os dois grupos). No

entanto, o desvio padrão não se aplica, já que este parâmetro não existe para variáveis categóricas. Ao invés disso, você deve informar o tamanho da diferença através das proporções esperadas em cada um dos grupos (e.g. 0.5 no grupo controle e 0.25 no grupo tratado, para um experimento planejado para detectar uma redução de mortalidade de 50% para 25% em um modelo de determinada doença, por exemplo).

Note que os tamanhos amostrais necessários para experimentos que medem variáveis categóricas tendem a ser maiores do que aqueles que medem variáveis quantitativas. Estes tamanhos de amostra serão maiores quanto mais afastadas de uma distribuição equilibrada (i.e. 50%/50%) forem as variáveis em questão. Detectar uma redução de mortalidade de 2% para 1%, por exemplo, vai requerer um tamanho amostral muito maior do que detectar uma redução de mortalidade de 50% para 25%, mesmo que em termos relativos a diminuição seja a mesma, conforme o exemplo abaixo.

Test family		Statistical test	
Exact		Proportions: Inequality, two independent groups (Fisher's exact test)	
Type of power analysis			
A priori: Compute required sample size - given alpha, power, and effect size			
Input parameters		Output parameters	
Determine		Sample size group 1	2456
Tail(s)	Two	Sample size group 2	2456
Proportion p1	0,02	Total sample size	4912
Proportion p2	0,01	Actual power	0,8000334
alpha err prob	0,05	Actual alpha	0,0356121
Power (1-beta err prob)	0,8		
Allocation ratio N2/N1	1		

2.6. Comparação de variáveis quantitativas por testes não paramétricos

Testes não-paramétricos para variáveis quantitativas entre dois grupos (como o teste de Mann-Whitney) se baseiam em uma lógica diferente dos testes paramétricos. Eles não utilizam os parâmetros que usamos até aqui, como médias e desvios-padrão, e sim a *ordem* entre os valores individuais e sua distribuição entre os grupos.

O G*Power não realiza cálculos para este tipo de teste, mas eles podem ser realizados a partir da *probabilidade de que um indivíduo qualquer em um grupo apresente um valor superior a um indivíduo em outro grupo* (“ $P(X>Y)$ ”) em <https://homepage.univie.ac.at/robin.ristl/samplesize.php?test=wilcoxon>. Esta é uma medida de tamanho de efeito na lógica não paramétrica, e indica o tamanho de diferença que se pretende detectar com o teste. Uma probabilidade de 50% indica grupos com distribuições semelhantes e representa a “hipótese nula” do teste, enquanto probabilidades muito altas ou muito baixas indicam grupos com diferenças grandes entre suas distribuições. Com isso, o tamanho de amostra

necessário é maior quanto mais esta probabilidade se aproxima de 0,5 (ou seja, quanto menor o tamanho de efeito), conforme os exemplos abaixo.

Input and calculation

P(X>Y)

Alpha two-sided

Power

The required sample size per group is 9.

Input and calculation

P(X>Y)

Alpha two-sided

Power

The required sample size per group is 33.

Input and calculation

P(X>Y)

Alpha two-sided

Power

The required sample size per group is 33.

Input and calculation

P(X>Y)

Alpha two-sided

Power

The required sample size per group is 9.

Note que colocar 0,1 e 0,9 (ou 0,3 e 0,7) gera os mesmos resultados, já que se um indivíduo tratado tem 10% de chance de ter um valor mais alto do que um controle, um indivíduo controle tem 90% de chance de ter um valor mais alto do que um tratado e vice-versa.

Os demais parâmetros a serem informados são o limiar de significância (α) e o poder desejado, como nos casos já vistos acima.

O site ainda permite que você calcule $P(X>Y)$ a partir de médias e desvios-padrão assumindo distribuições normais. Dito isso, cabe lembrar que testes não-paramétricos costumam ser usados precisamente em casos nos quais não se espera uma distribuição normal das amostras ou médias amostrais. Se este de fato for o caso, o procedimento pode levar a estimativas pouco confiáveis.

3. Cálculos de tamanho amostral baseados em precisão:

Caso o objetivo de seu projeto não seja comparar grupos e sim *estimar o valor de um parâmetro* (por exemplo, a prevalência de indivíduos portadores de uma determinada bactéria, ou a concentração inibitória média de um antibiótico), a lógica do cálculo de tamanho amostral é diferente. Neste caso, ele deixa de ser baseado no poder estatístico (já que não há um teste estatístico a ser realizado) e passa a depender da *precisão desejada para a estimativa*.

Mais comumente, a incerteza na estimativa de um parâmetro é dada por um *intervalo de confiança* – a faixa em que você esperaria encontrar o parâmetro real X% das vezes. Um intervalo de confiança de 95% (o mais comumente utilizado), por exemplo, representa a margem de erro que incluiria o parâmetro a ser estimado em 95% dos experimentos, caso a amostra de fato seja representativa da população a ser estudada.

Com isso, a precisão desejada envolve decidir *qual o tamanho do intervalo de confiança* que se deseja obter. Novamente, intervalos de confiança de 95% costumam ser os mais utilizados.

3.1. Estimando a proporção de uma variável categórica

Caso você esteja estimando a *proporção* de uma categoria de uma variável categórica (e.g. a % de animais infectados por um parasita num biotério), acesse <https://statulator.com/SampleSize/ss1P.html> e informe os seguintes parâmetros

a) o *nível de confiança* de seu intervalo (usualmente 95%)

b) a *proporção esperada*. É evidente que você não tem como saber a proporção que será encontrada, mas proporções intermediárias (e.g. 50%) levarão a intervalos de confiança distintos de proporções extremas (e.g. 1%), então é importante estimar a faixa em que você pretende trabalhar.

c) a *precisão desejada* – ou seja, o quão estreito ou largo você pretende que seja o seu intervalo de confiança. Note que a precisão pode ser informada tanto em valores *absolutos* (e.g. X pontos percentuais, como em uma pesquisa eleitoral) como *relativos* (e.g. X% de variação para mais ou para menos em relação à proporção encontrada).

Input Values

Specify input values and click Calculate. Hover over the ? sign to obtain help.

Level of Confidence ?

0.95

Expected Proportion ?

0.5

Precision or Margin of Error ?

Absolute value

Note: You may adjust sample sizes for finite population, clustering and response rate by clicking the 'Adjust' button below.

No exemplo acima, estamos estimando o tamanho de amostra para estimar uma proporção em torno de 50% com um intervalo de confiança de 95% de 2 pontos percentuais (o que normalmente corresponde a situação desejada em uma pesquisa de intenção de voto no segundo turno das eleições, por exemplo). O resultado é que necessitaríamos de 2401 indivíduos.

Results and Live Interpretation Download

Assuming that 50% of the subjects in the population have the factor of interest, the study would require a sample size of:

2401

for estimating the expected proportion with 2% absolute precision and 95% confidence.

In other words, if you select a random sample of 2401 from a population, and determine that 50% of subjects have the factor of interest, you would be 95% confident that between 48% and 52% of subjects in the population have the factor of interest.

Reference: Dhand, N. K., & Khatkar, M. S. (2014). Statulator: An online statistical calculator. Sample Size Calculator for Estimating a Single Proportion. Accessed 31 May 2024 at <http://statulator.com/SampleSize/ss1P.html>

Note: You may adjust the calculated sample size for clustering, response rate or finite population by clicking [here](#) or the 'Adjust' button.

Caso possamos tolerar um intervalo de confiança de 95% de 10 pontos percentuais, esse número cai para 97 indivíduos:

Level of Confidence ?

0.95

Expected Proportion ?

0.5

Precision or Margin of Error ?

Absolute value

Note: You may adjust sample sizes for finite population, clustering and response rate by clicking the 'Adjust' button below.

Results and Live Interpretation Download

Assuming that 50% of the subjects in the population have the factor of interest, the study would require a sample size of:

97

for estimating the expected proportion with 10% absolute precision and 95% confidence.

3.2. Estimando a média de uma variável quantitativa

Para estimar uma *média* com um determinado intervalo de confiança (e.g. a média de peso dos animais de um biotério aos 30 dias), acesse <https://statulator.com/SampleSize/ss1M.html>.

Neste caso, você terá que informar uma *estimativa do desvio-padrão* (na mesma unidade em que a média será estimada), além do nível de confiança e da precisão desejada (conforme o exemplo acima).

No exemplo acima, por exemplo, caso o desvio-padrão do peso seja de 50 gramas, e a precisão desejada para o intervalo de confiança seja ± 20 gramas, necessitaríamos de 28 animais:

Level of Confidence ?

0.95

Expected Standard Deviation ?

50

Precision or Margin of Error ?

20

Note: You may adjust sample size for t-distribution (applied by default), finite population or clustering by clicking the 'Adjust' button below.

▶ Calculate Adjust Reset

Results and Live Interpretation Download

Assuming the expected population standard deviation to be 50, and employing t-distribution to estimate sample size, the study would require a sample size of:

28

to estimate a mean with 95% confidence and a precision of 20.

In other words, if you select a random sample of 28 from a population, and determine the mean to be say y , you would be 95% confident that the mean in the population lies somewhere between $y - 20$ and $y + 20$.

Cabe notar que, caso você queira pensar na comparação de uma variável quantitativa entre dois grupos como um experimento que tenta *estimar uma diferença* de forma quantitativa, o cálculo de poder estatístico pode ser realizado com base na precisão, de forma bastante semelhante ao realizado para uma média, sendo necessária uma estimativa do desvio-padrão, o nível de confiança e o grau de precisão desejados. Tal raciocínio é incomum na pesquisa com animais, mas diversos autores argumentam que ele seria menos dicotômico e mais produtivo (e.g. <https://thenewstatistics.com/> e outros).

4. Recomendações Gerais:

Após ter coberto estes casos particulares (que não esgotam todas as possibilidades de experimentos, mas cobrem a maioria dos que costumamos avaliar na CEUA), gostaríamos de encerrar este manual com algumas recomendações gerais.

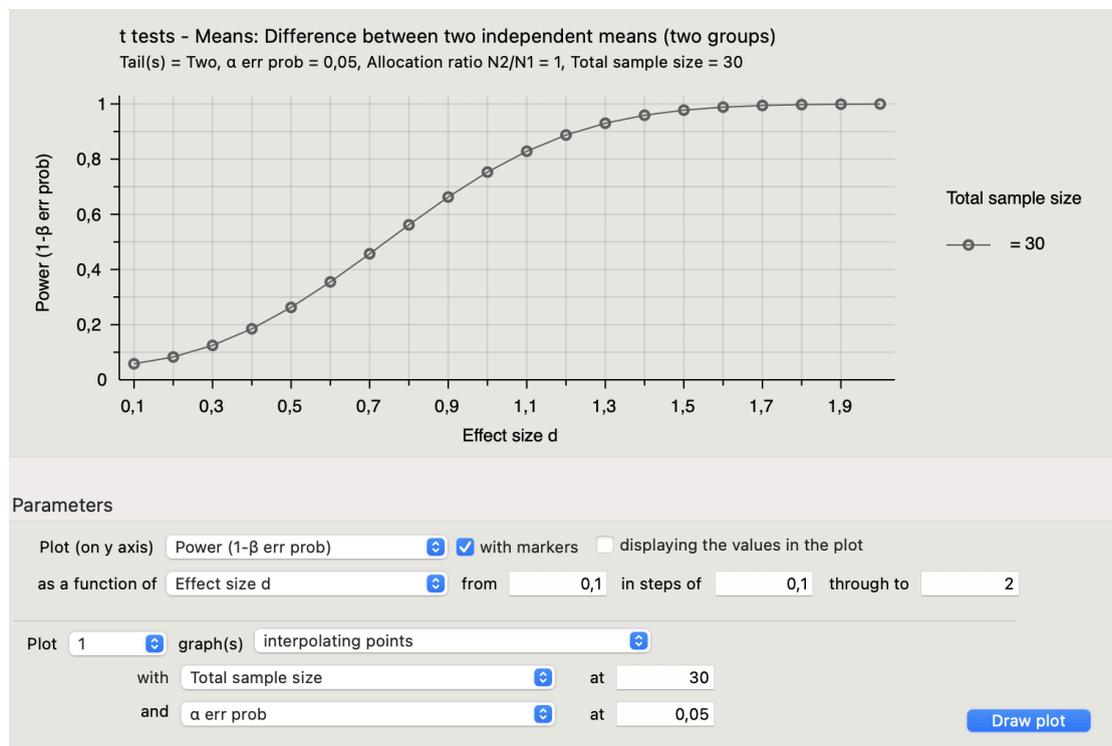
4.1. Não procure uma “resposta certa”.

Um cálculo de tamanho amostral não é uma pergunta que possua uma resposta certa. Pesquisadores diferentes podem estar interessados em tamanhos de efeito distintos, ou optar por graus de rigor distintos em suas escolhas de poder estatístico e limiar de significância. O tamanho amostral necessário para um experimento variará conforme estas decisões, que devem ser feitas pelo pesquisador. Dito isso, é fundamental que você *enxergue as opções que está fazendo*, e as consequências que elas têm sobre a validade de seu experimento, para certificar-se de que os recursos investidos e o sofrimento animal envolvido no mesmo sejam justificados.

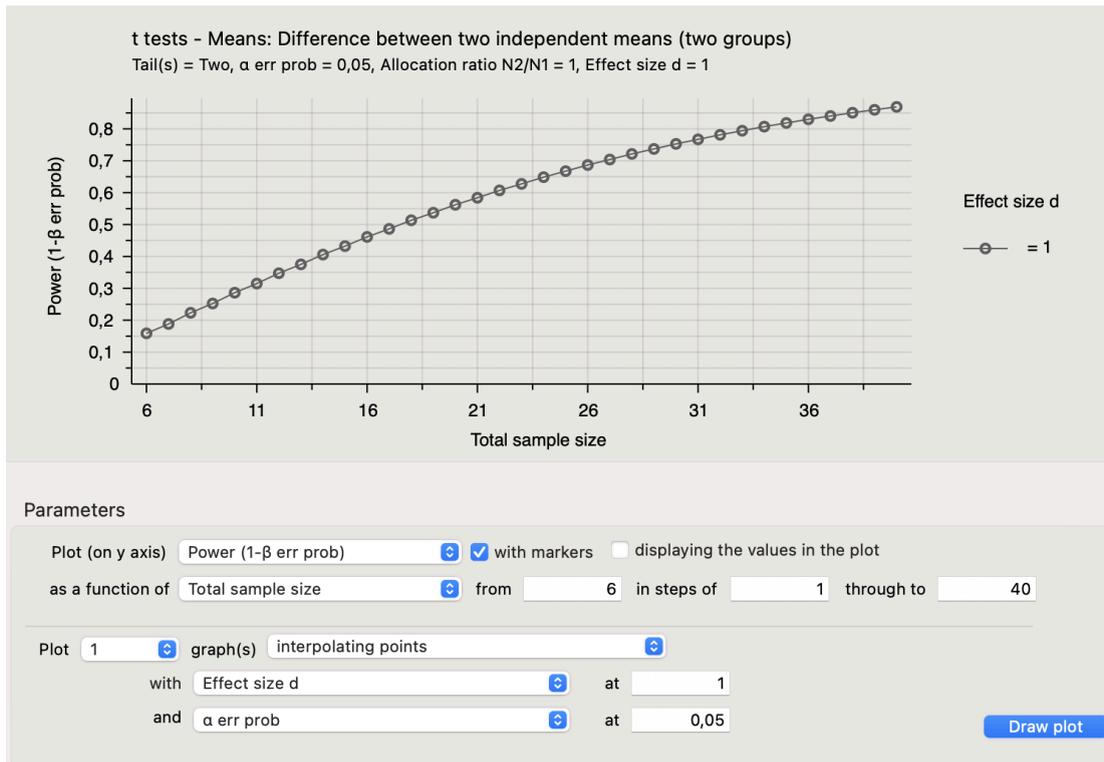
4.2. Explore os parâmetros.

Dado que o cálculo não possui uma resposta certa, sintá-se livre para explorar os parâmetros e verificar o impacto deles sobre o tamanho amostral necessário. Em particular, é bastante útil verificar o poder estatístico de um determinado tamanho de amostra para detectar diferentes tamanhos de efeito. O G*Power permite que você visualize o poder estatístico de diferentes tamanhos de amostra para um tamanho de efeito (ou de um tamanho de amostra para diferentes tamanhos de efeito) através do botão “X-Y plot for a range of values”.

Abaixo, pode-se observar o poder estatístico para diferentes tamanhos de efeito (em d de Cohen) para um experimento com 15 animais por grupo analisado por um teste t bicaudal de amostras independentes, com um limiar de significância de 0,05:



Abaixo, pode-se ver o poder estatístico atingido com diferentes tamanhos de amostra totais (em ambos os grupos) para detectar um efeito com d de Cohen de 1 no mesmo teste:



4.3. Não use parâmetros arbitrários para chegar ao tamanho de amostra que você já desejava utilizar.

Caso um pesquisador tenha uma intenção irredutível de usar um determinado tamanho de amostra, é bastante fácil inserir um conjunto qualquer de parâmetros que leve a este resultado (como demonstrado na seção 4.2). Com isso, uma prática infelizmente comum para convencer a CEUA a aprovar um número de animais específico é inserir parâmetros arbitrários que levem precisamente ao resultado desejado e apresentar este cálculo à comissão, sem pensar na lógica do que foi feito.

Não deveríamos precisar pedir, mas por favor não faça isso. Estudos com tamanhos amostrais insuficientes desperdiçam vidas animais em experimentos inconclusivos ou, pior do que isso, com grande chance de chegar a conclusões erradas. Em contrapartida, estudos com amostras desnecessariamente grandes podem causar sofrimento animal gratuitamente. Existe uma razão concreta e importante para tentar calcular o tamanho amostral que responda sua pergunta de pesquisa com confiança suficiente sem desperdiçar vidas animais à toa. E precisamos que você, como responsável pela pesquisa, se preocupe com isso tanto quanto a CEUA se preocupa.

4.4. Entenda o que você está fazendo.

Use uma forma de cálculo que seja intuitiva para você, e parâmetros cujo sentido biológico você possa compreender. É mais fácil pensar em termos de médias e desvios-padrão do que em tamanhos de efeito estandardizados, por exemplo. Também é mais fácil compreender os parâmetros necessários para uma

comparação entre dois grupos do que aqueles necessários para testes estatísticos mais complexos. Novamente, a intenção do cálculo não é encontrar uma resposta certa, mas entender como diferentes escolhas afetam o poder estatístico do teste e/ou o tamanho amostral necessário. Para que isso funcione, é preciso que você tenha uma compreensão intuitiva tanto dos parâmetros que está inserindo quanto da resposta que está recebendo.

4.5. Evite parâmetros arbitrários.

Algumas áreas da ciência costumam usar definições arbitrárias de tamanhos de efeito “pequenos”, “médios” e “grandes”. De forma geral, consideramos que usar estas definições (originalmente derivadas da pesquisa em psicologia) como base para o cálculo são uma prática ruim. O que constitui um tamanho de efeito “grande” para intervenções comportamentais em humanos, afinal, não corresponde a um efeito “grande” de intervenções tipicamente realizadas em animais de laboratório, que via de regra são mais invasivas e apresentam efeitos maiores.

Da mesma forma, a rigidez em valores arbitrários para o limiar de significância e poder estatístico são contraproducentes. Para um experimento exploratório e uma hipótese com plausibilidade alta, um limiar de significância de 0,05 pode ser suficiente. Porém, em situações em que você precise de um grau de confiança maior nos resultados, esteja testando uma hipótese improvável, ou esteja realizando muitas comparações, este valor costuma fornecer um grau de rigor insuficiente. Caso um experimento seja muito importante para você, considere a possibilidade de ser mais estrito: você não quer basear uma tese de doutorado inteira num experimento com p de 0,04.

4.6. Caso você não possa alcançar o tamanho amostral necessário, pergunte-se se o experimento vale a pena.

A CEUA tem consciência de que recursos de pesquisa são limitados, o que inclui tanto os próprios animais como reagentes, equipamentos e o tempo dos pesquisadores. Com isso, experimentos com grandes números de animais podem ser proibitivos, tanto em termos de recursos utilizados como do grau de sofrimento animal necessário. Você pode e deve levar esses limites em consideração no seu cálculo, caso julgue que chegar a determinados tamanhos de amostra seja inviável.

Caso você seja confrontado com limites no tamanho amostral que pode alcançar, porém, é importante que você se pergunte se o experimento que você é capaz de realizar ainda vale a pena. Eventualmente, aceitar um poder estatístico um pouco menor ou se ater a tamanhos de efeito grandes para usar um tamanho amostral menor pode representar uma solução aceitável. Mas fazer um experimento com 5 animais por grupo quando o ideal seriam 30 quase nunca trará uma resposta apropriada: ou você terá que lidar com taxas de erro muito altas ao diminuir o poder estatístico/elevar o limiar de significância, ou limitará sua sensibilidade apenas a efeitos muito grandes. Nesse caso, a alternativa ética provavelmente é abdicar de realizar o experimento e usar seu tempo, seus recursos, e as vidas dos

animais aos quais você tem acesso para uma pergunta que você consiga responder adequadamente.

4.7. Caso você tenha dúvidas, procure ajuda.

Caso você se sinta inseguro ao realizar um cálculo de tamanho amostral, seja porque ele é particularmente complexo, seja porque você não domina alguns dos fundamentos estatísticos necessários para sua compreensão, não arrisque e consulte ajuda especializada, seja com a CEUA-CCS ou com outros profissionais. O tempo dispendido em procurar este tipo de ajuda provavelmente será muito menor do que aquele que você gastará respondendo pendências da CEUA ou realizando experimentos que darão pouco retorno caso o cálculo não seja adequadamente realizá-lo.

A CEUA possui profissionais que podem ajudá-lo e está disponível para esclarecer dúvidas através do e-mail ceua_ccs@ccsdecania.ufri.br.

Diversas outras referências sobre cálculo de tamanho amostral também podem ser encontradas na literatura científica (e.g. <https://doi.org/10.1525/collabra.33267>, <https://doi.org/10.1136/bmj.306.6886.1181>, <https://doi.org/10.3352%2Fjeehp.2021.18.17>) e também podem ser consultadas para este fim.