

Cálculo de tamanho amostral e poder estatístico: orientações

Seguindo a tendência de diversos órgãos reguladores ao redor do mundo, o CEUA-CCS exige a realização de cálculos de poder estatístico e tamanho amostral para a autorização de experimentos em animais. A razão é simples: experimentos com baixo poder estatístico tem uma grande chance de gerarem tanto resultados falsos-negativos (pela própria definição do conceito de poder estatístico, que é a capacidade de um teste estatístico detectar um determinado tamanho de efeito se ele existir) como falsos-positivos (já que, em se utilizando um limiar de significância fixo, um resultado significativo terá menos chance de ser verdadeiro se o poder estatístico for baixo). Desta forma, a ausência de um cálculo de tamanho amostral e poder estatístico pode expor animais a experimentos inapropriados para responder uma determinada pergunta.

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 inúmeras calculadoras online (como em <http://powerandsamplesize.com> e <http://www.stat.ubc.ca/~rollin/stats/ssize/>) ou softwares gratuitos (como o GPower, disponível em <http://www.gpower.hhu.de/>).

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

1. O tamanho mínimo de diferença entre grupos que se pretende detectar com o estudo (o qual deve ser estabelecido pelo pesquisador).
2. O poder estatístico desejado (recomendamos que o mesmo seja de pelo menos 80%).
3. Para variáveis quantitativas, a variância ou desvio padrão esperado para os grupos experimentais.
4. O limiar de significância a ser utilizado no teste estatístico.

Diferentes testes paramétricos ou calculadoras podem requerer que as informações (e.g. a definição de tamanho de diferença) sejam descritos de maneira discretamente diferente. Para ilustrar o processo com um exemplo simples, vamos mostrar como os parâmetros devem ser utilizados para cálculos de tamanho amostral com diferentes tipos de variáveis e desenhos experimentais.

(i) Comparação de uma variável quantitativa entre dois grupos

Para demonstrar o cálculo de tamanho amostral de um teste t bicaudal para amostras independentes, um dos casos mais comuns em pesquisa em modelos animais, utilizaremos o site <http://powerandsamplesize.com>:

1. Entre no site
2. Clique em “go straight to the calculators” e em “Compare 2 means – 2 samples – 2 sided equality” (isso levará você para este link: <http://powerandsamplesize.com/Calculators/Compare-2-Means/2-Sample-Equality>).

3. No primeiro campo de cima para baixo, você pode optar em calcular (a) um tamanho amostral para obter um determinado poder estatístico ou (b) o poder estatístico de um determinado tamanho amostral. Para fins de planejamento experimental, o caso (a) será mais comum, então vamos analisá-lo a seguir.

4. Na linha abaixo, você deve inserir (a) o poder estatístico desejado (em forma de decimal – e.g. “0.8” significa 80%) e (b) a chance do erro tipo 1, que equivale ao limiar de significância estatística desejado (e.g. 5% para um valor p de 0.05). O campo “sample size” não deve ser preenchido, pois é exatamente este número que vamos calcular em nosso exemplo.

Na CEUA-CCS, recomendamos que o poder estatístico utilizado para o cálculo seja de pelo menos 80%, mas também é possível calcular o tamanho de amostra necessário para obter poderes maiores (e.g. 90% ou 95%), dependendo da intenção dos pesquisadores.

5. Nas próximas linhas (“Group ‘A’ mean” e “Group ‘B’ mean”) você deve especificar o tamanho de diferença que pretende detectar, ao inserir as médias esperadas para os dois grupos no caso desta diferença existir. Por exemplo, se você está desenhando 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, por exemplo), pode inserir 40 no grupo A (controle) e 20 no grupo B (tratado). Note que as unidades nesse caso não importam (e.g. você poderia inserir “100” e “50” se quisesse considerar o valor do controle como 100%), mas lembre que *o desvio padrão deverá ser inserido na mesma unidade que as médias.*

É óbvio que, previamente ao experimento, o experimentador não tem como prever o tamanho da diferença que de fato encontrará. No entanto, o tamanho de diferença inserido não se refere ao resultado do experimento, e sim do *tamanho de diferença que o experimento terá poder para detectar.* Desta forma, a diferença inserida deve ser aquela que você consideraria *importante não deixar de detectar em um teste estatístico*, o que geralmente corresponde àquela que você consideraria biologicamente relevante. Lembre que quanto menor a diferença que você quiser detectar com confiança, *maior* será o tamanho amostral.

7. Na linha seguinte, você deverá inserir o desvio padrão. No caso do teste t, o mesmo usa um desvio padrão calculado a partir dos desvios dos dois grupos (calculado por $s_{X_1X_2} = \sqrt{(s_{X_1}^2 + s_{X_2}^2)}$). Novamente, o desvio padrão do experimento em si não tem como ser conhecido antes do mesmo. Desta forma, você deve utilizar a *melhor estimativa possível para este* – seja a partir de experimentos anteriores, dados da literatura ou um experimento piloto.

Se você não dispuser de um desvio padrão esperado para o grupo tratado, utilizar a estimativa de desvio padrão dos controles é uma alternativa razoável, a não ser que seu tratamento aumente de forma significativa a variância da variável a ser medida. Novamente, lembre que o desvio padrão deve estar na mesma unidade que as médias – i.e. se você espera uma média de 40 mg/dl, um desvio padrão de 10 mg/dl e quer detectar uma redução de 50%, você pode inserir tudo em unidades brutas (médias de 40 e 20, desvio padrão de 10) ou

normalizadas (médias de 100 e 50 , desvio padrão de 25) sem que isso mude o resultado do teste.

8. A última linha (sampling ratio) contempla a possibilidade de experimentos com grupos de tamanho diferente. Idealmente, o maior poder estatístico de um estudo será obtido com grupos experimentais do mesmo tamanho (nesse caso, use 1 neste campo). Entretanto, se por alguma razão isso não for factível, você pode alterar o parâmetro inserindo o valor de kappa (definido como a razão entre o tamanho amostral dos dois grupos – i.e. n do grupo A/n do grupo B). Neste caso, o resultado dado pelo teste no campo sample size representará o tamanho amostral necessário para o grupo B, e você deverá calcular o tamanho amostral do grupo A a partir desta razão.

9. Depois de preencher todos os itens, clique em “Calculate” e veja o número que aparece no campo “Sample Size”.

Tomando um exemplo simples, se você tem um experimento planejado para (a) detectar uma redução de 50% em um parâmetro cuja média nos controles é 40 mg/dl, (b) com um desvio padrão de 10 mg/dl, (c) com um poder estatístico de 80%, (d) um limiar de significância de $p=0.05$, (e) com grupos de tamanho igual, o tamanho de amostra para obter o poder estatístico desejado será de 4 animais por grupo.

Note que para detectar uma diferença menor (e.g. uma redução de 25%) com o mesmo poder estatístico, o tamanho amostral será maior (nesse caso, 16 animais por grupo). Caso o poder estatístico desejado seja maior (e.g. 90%), o tamanho amostral também aumenta (no caso anterior, de 4 para 6).

(ii) Comparação de uma variável quantitativa entre mais de dois grupos

No caso da comparação de uma variável quantitativa entre dois grupos, o procedimento mais comum 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 análise de variância envolve expressar o tamanho de efeito através de um valor de F, o que não é muito instintivo para a maior parte dos pesquisadores. Dito isso, se você quiser fazê-lo, o G*Power permite que você estime um valor de F a partir das médias esperadas nos vários grupos (para uma ANOVA de uma via, escolha “F tests”, “ANOVA: fixed effects, omnibus, one-way” e clique em “Determine” para gerar um valor de F a partir de um resultado esperad).

Dito isso, se as comparações individuais entre grupos através do pós-hoc interessam mais a você do que o resultado da ANOVA em si, uma forma prática de estimar o tamanho amostral necessário para estas (que será geralmente maior do que aquele necessário para a detecção de um efeito geral através da ANOVA) é utilizar a calculadora disponível no [powerandsamplesize.com](http://powerandsamplesize.com/Calculators/Compare-k-Means/1-Way-ANOVA-Pairwise) em “Compare K-means, 1-Way ANOVA Pairwise, 2-sided equality”.
<http://powerandsamplesize.com/Calculators/Compare-k-Means/1-Way-ANOVA-Pairwise>

Neste link, você simplesmente informará a diferença que pretende detectar através das médias, o poder estatístico desejado e o limiar significância, como no caso do teste t explicado no item (i). Além disso, porém, você preencherá no item “number of pairwise comparisons” o *número de comparações entre grupos* que pretende realizar (e.g. três comparações no caso de três grupos comparados entre si em todas as combinações possíveis, ou duas no caso de dois grupos tratados, cada um dos quais comparado com um controle). A partir do número de comparações, o tamanho amostral será ajustado para que o poder estatístico seja mantido para cada uma das comparações pós-hoc. Note que quanto mais comparações forem realizadas, maior será o tamanho amostral necessário, já que o limiar de significância utilizado em cada uma delas será mais estrito.

(iii) Comparação 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), entre em “Compare 2 proportions”; “2 sample, 2-sided equality”: <http://powerandsamplesize.com/Calculators/Compare-2-Proportions/2-Sample-Equality>

O procedimento é similar ao utilizado para variáveis qualitativas em relação à informação do poder estatístico, do limiar de significância utilizado e do *sampling ratio* (e.g. equilíbrio do tamanho amostral entre os dois grupos). No entanto, o desvio padrão não se aplica, já que o mesmo não existe para variáveis categóricas. Ao invés disso, você deve simplesmente 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. Note também que os 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 necessitar 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.

(iv) Outros cálculos

Cálculos de outros testes paramétricos, como testes t de 1 amostra e testes de não-inferioridade, também são relativamente simples e podem ser realizados no mesmo site. Já o G*Power proporciona cálculos para uma gama muito maior de testes, tais como testes t pareados, ANOVAs de medidas repetidas, correlações, regressões lineares, etc. Mais detalhes sobre cada um destes casos podem ser encontrados na documentação do software http://www.gpower.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf

ou, em uma versão mais simplificada, em <http://www.ats.ucla.edu/stat/gpower/>

(v) Cálculos para métodos estatísticos não-paramétricos

Cálculos de tamanho amostral para métodos não-paramétricos (como os testes de Mann-Whitney e Kruskal-Wallis) são mais complexos e geralmente dependem de uma estimativa prévia de distribuição, o que pode ser mais difícil de realizar. Nestes casos, uma regra simples é calcular o tamanho amostral para uma comparação paramétrica e adicionar 15%, o que representa uma estimativa conservadora do tamanho amostral necessário.

Para mais informações, qualquer livro de bioestatística possui uma seção de cálculo de tamanho amostral (para uma introdução a essa e outras questões estatísticas, em termos de simplicidade, recomendamos o livro *Intuitive Biostatistics*, de Harvey Motulsky). Recomendamos ainda as seguintes referências disponíveis online:

<http://www.graphpad.com/guides/prism/6/statistics/>

<http://emj.bmj.com/content/20/5/453.full.pdf+html>

<http://www.nature.com/nrn/journal/v14/n5/abs/nrn3475.html>

Atenciosamente,

CEUA- CCS