

Gráfico de órbita (orbitplot): Uma nova ferramenta para a apresentação multivariada de contrastes estatísticos emparelhados

Fernando Ferreira-Santos^{1,2}, Eva C. Martins³, João Marques-Teixeira¹, & Michelle de Haan²

1. Laboratório de Neuropsicofisiologia, Universidade do Porto (Portugal)

2. Developmental Cognitive Neuroscience Unit, UCL Institute of Child Health (Reino Unido)

3. Departamento de Ciências Sociais e do Comportamento/UNIDEP-CINEICC, Instituto Superior da Maia (Portugal)

Resumo: Estado de Arte: Existem poucos tipos de gráficos capazes de resumir dados multivariados. O gráfico de radar é um exemplo, permitindo apresentar simultaneamente dados de um número arbitrário de variáveis. Cada variável é representada por um eixo unidimensional com uma origem comum, sendo cada eixo disposto de forma equiangular.

Novas perspectivas: As principais limitações do gráfico de radar consistem na dificuldade em representar barras de erro (e.g., intervalos de confiança) e valores negativos (cf. eixos unidimensionais). A representação precisa de contrastes estatísticos e tamanhos de efeito depende criticamente destas possibilidades. O gráfico de órbita consiste numa adaptação do gráfico de radar que permite incorporar barras de erro e valores negativos (utilizando eixos bidimensionais). Implicações teóricas e práticas: O gráfico de órbita constitui uma nova ferramenta para a visualização de dados multivariados implementada na função "orbitplot" para o software estatístico R.

Palavras-chave: gráficos multivariados; tamanho do efeito; intervalos de confiança; estatística.

INTRODUÇÃO

A apresentação gráfica de dados ou resultados científicos é um fator que pode afetar a eficácia da comunicação científica em múltiplos contextos, como a submissão de artigos para publicação, a candidatura a financiamento para investigação, ou apresentações e posters em encontros científicos (Frankel & DePace, 2012). Gráficos simples e intuitivos são normalmente os mais eficazes a transmitir a informação pretendida, mas ao mesmo tempo a simplicidade não deve sobrepor-se à necessidade de transmitir todos os dados necessários à correta interpretação do gráfico.

Uma forma usual de apresentação gráfica de resultados é a ilustração de medidas de tendência central (tipicamente a média) acompanhadas de medidas de dispersão (usualmente, desvio padrão, erro padrão da média ou intervalos de confiança) utilizando, por exemplo, gráficos de barras com marcadores de erro. Esta forma de apresentação gráfica é muito útil pois transmite toda a informação necessária para que o leitor interprete o efeito encontrado. No entanto, na Psicologia e domínios afins, o uso de desenhos de investigação e análises estatísticas multivariadas é uma prática comum e os resultados destas análises são muitas vezes difíceis de resumir graficamente. Em desenhos multivariados com vários grupos e/ou múltiplas condições experimentais a apresentação de todas as médias para todas as combinações do plano fatorial torna-se pouco prática. Uma forma de contornar este problema pode ser representar não os dados descritivos, mas os tamanhos de efeito dos contrastes estatísticos analisados. A apresentação de tamanhos de efeito retém as vantagens da apresentação de estatísticas descritivas permitindo uma correta interpretação do resultado, mas simplifica a informação a apresentar ao mostrar um valor por contraste entre cada par de variáveis.

Os gráficos de radar (também denominados gráficos estrela, polígonos irregulares, ou gráficos aranha) permitem a apresentação de dados multivariados com um número arbitrário de variáveis (Friendly, 1991). Cada eixo representa uma variável e é desenhado como uma linha unidimensional começando na origem do gráfico e prolongando-se para o exterior. Diferentes variáveis serão desenhadas como eixos adicionais separados de forma equiangular. Esta estrutura permite aos gráficos de radar

acomodar um número arbitrário de variáveis simplesmente ajustando o ângulo entre cada par de variáveis (o valor deste ângulo para cada variável, ao contrário dos gráficos polares ou radiais, não é informativo uma vez que depende apenas do número de variáveis que são apresentadas simultaneamente no mesmo gráfico radar). Depois de estabelecida a estrutura do gráfico com todos os eixos, o valor de cada variável corresponde a um ponto no eixo respectivo e são traçadas linhas entre pontos adjacentes. Isto leva à construção de um polígono irregular onde os vértices são os valores das variáveis (Figura 1).

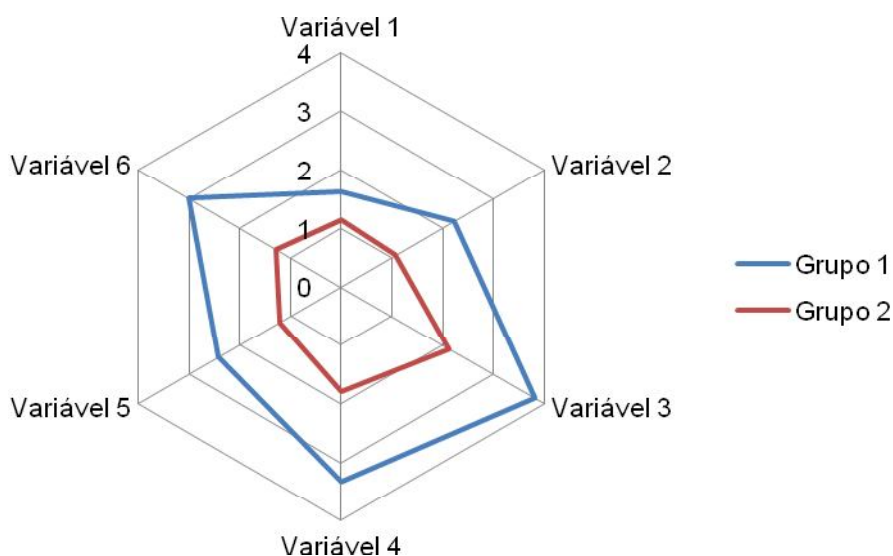


Figura 1. Exemplo de um gráfico de radar sumariando os resultados de dois grupos experimentais para seis variáveis.

Neste sentido, os gráficos de radar podem ser utilizados para a apresentação de dados multivariados, mas apresentam duas grandes limitações: torna difícil representar barras de erro (e.g., intervalos de confiança) e valores negativos (cf. eixos unidimensionais). A representação precisa de contrastes estatísticos e tamanhos de efeito depende criticamente destas possibilidades.

De modo a contornar as limitações apresentadas pelos gráficos de radar, desenvolvemos um novo tipo de gráfico para a apresentação multivariada de contrastes estatísticos a que chamamos gráfico de órbita (orbitplot). A estrutura do gráfico de órbita é semelhante à do gráfico de radar, com a exceção de que as linhas equiangulares representam eixos bipolares para o desenho de contrastes estatísticos ou tamanhos de efeito. Ao permitir valores positivos e negativos, os eixos bipolares permitem representar a direção do efeito estatístico. Na origem do gráfico colocamos uma condição ou grupo central e cada linha irá ligá-la a cada outra condição ou grupo. O valor do tamanho de efeito do contraste entre cada duas condições será desenhado no respetivo eixo que as une. O ponto médio de cada eixo entre duas condições ou grupos representa o efeito nulo (e.g., 0 para o d de Cohen), com valores mais próximos da origem a representar um efeito que favorece a condição central e valores mais extremos que o valor nulo a indicar um efeito que favorece a condição periférica. Cada uma destas três possibilidades é codificada com cor para facilitar a visualização: um efeito nulo é codificado a cinzento, um efeito na direção da condição central é codificado a vermelho, um efeito na direção da condição periférica é codificado a azul. Os contrastes ou tamanhos de efeito são representados pelos círculos preenchidos e a respetiva medida de erro é indicada pela circunferência em torno desse círculo. Esta medida de erro pode ser qualquer medida de variabilidade de interesse, mas a utilização

de intervalos de confiança é de particular utilidade. Utilizando intervalos de confiança de 95% é possível aferir a significância dos efeitos: se um tamanho de efeito é significativamente diferente do valor nulo, então o seu intervalo de confiança não atravessa esse valor nulo e, para facilidade de visualização no gráfico de órbita, o círculo preenchido é substituído por um asterisco. Metaforicamente, podemos dizer que os tamanhos de efeito "orbitam" em torno da condição ou grupo central e quanto mais forte é o efeito na direção dessa condição, mais a condição central "puxa" o efeito para si (por outro lado, tamanhos de efeito com a direção oposta serão "puxados" para a periferia). O gráfico de órbita é exemplificado na Figura 2.

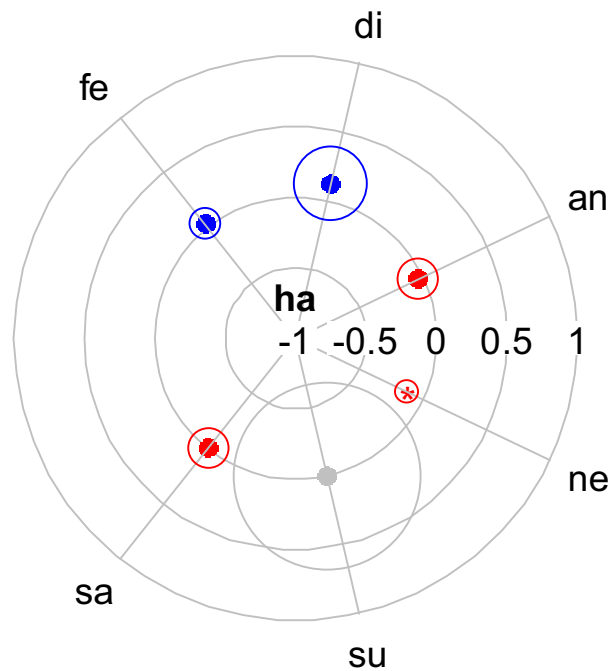


Figura 2. Exemplo de gráfico de órbita ilustrando tamanhos de efeito d de Cohen e respetivos intervalos de confiança a 95% para seis contrastes entre a condição central ($ha=happiness$) e as condições periféricas ($fe=fear$, $di=disgust$, $an=anger$, $ne=neutral$, $su=surprise$, $sa=sadness$). Valores coloridos a vermelho vão no sentido da condição central, a azul vão no sentido da respetiva condição periférica e a cinzento são iguais a zero. O contraste entre $happiness$ e $neutral$ é o único resultado estatisticamente significativo e, como tal, está assinalado com um asterisco.

METODOLOGIA

Para a produção de gráficos de órbita foi desenvolvida uma função computacional intitulada "orbitplot()" para o software de análise estatística R (R Core Team, 2012), dependente do pacote de funções gráficas "plotrix" (Lemon, 2006).

A função "orbitplot()" aceita como argumentos de entrada (inputs) três matrizes de dados com n -por- n elementos cada, contendo respetivamente: (1) tamanhos de efeito ou contrastes estatísticos, (2) limites inferiores das barras de erro, (3) limites superiores das barras de erro. Todos estes valores podem tomar quaisquer grandezas estatísticas de interesse para o utilizador. Os valores de tamanho de efeito ou contraste estatístico podem ser, por exemplo, diferenças médias estandardizadas d de Cohen ou valores t . Por sua vez, as barras de erro podem também representar desvios padrão, erros padrão ou intervalos de confiança. O número de elementos n corresponde ao número de grupos ou condições em comparação. Por exemplo, no caso de haver três grupos em comparação com um grupo

de referência, a matriz terá $3 \times 3 = 9$ elementos (n.b., os elementos da diagonal não são interpretados). O resultado da função consiste em n gráficos de órbita (cada um com uma condição ou grupo no centro), com $n - 1$ eixos.

RESULTADOS

O resultado do presente trabalho consiste na própria função "orbitplot()" descrita acima. A utilização da função para fins de investigação não comerciais é gratuita. Para obter a função para estes fins é favor contactar o primeiro autor.

CONCLUSÕES

O gráfico de órbita foi desenvolvido como uma solução para os problemas identificados na introdução, permitindo a apresentação idónea de dados multivariados com medidas de erro.

AGRADECIMENTOS

FFS foi apoiado por uma bolsa da Fundação para a Ciência e Tecnologia (SFRH/BD/64071/2009). O presente trabalho foi realizado no âmbito do projeto PTDC/PSI-PCO/114953/2009 financiado pela Fundação para a Ciência e Tecnologia.

CONTACTO PARA CORRESPONDÊNCIA

Fernando Ferreira-Santos, Laboratório de Neuropsicofisiologia, Faculdade de Psicologia e de Ciências da Educação da Universidade do Porto, Rua Alfredo Allen 4200-135 Porto, frsantos@fpce.up.pt

REFERÊNCIAS BIBLIOGRÁFICAS

- Frankel, F. C., & DePace, A. H. (2012). *Visual strategies: A practical guide to graphics for scientists and engineers*. New Haven, CT: Yale University Press.
- Friendly, M. (1991). Statistical graphics for multivariate data. *Proceedings of the Sixteenth Annual SAS Users Group International Conference* (pp. 1157–1162). Cary, NC: SAS Institute Inc. Retrieved from http://www.sascommunity.org/wiki/SAS_Global_Forum_Proceedings/SUGI_91
- Lemon, J. (2006). Plotrix: a package in the red light district of R. *R-News*, 6(4), 8–12. Retrieved from http://cran.r-project.org/doc/Rnews/Rnews_2006-4.pdf
- R Core Team. (2012). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org/>