

## O software R e a pesquisa com Alimentos e Nutrição

### Eric Batista Ferreira

Docente do Departamento de Estatística (DEst) e do  
Programa de Pós-graduação em Estatística Aplicada e Biometria (PPGEAB)  
Diretor Adjunto do Centro de Educação Aberta e a Distância (CEAD)  
Coordenador Adjunto da Universidade Aberta do Brasil (UAB)  
Universidade Federal de Alfenas (Unifal-MG)  
✉ [eric.ferreira@unifal-mg.edu.br](mailto:eric.ferreira@unifal-mg.edu.br)

Data de mais de 5000 anos antes de Cristo os rudimentos do ábaco, primeira máquina de fazer contas que se tem notícia. Há muito o ser humano vem criando tecnologias que o auxiliem na missão de calcular das mais simples às mais complexas tarefas. Após as máquinas mecânicas vieram as eletrônicas e, associado ao desenvolvimento da indústria e da pesquisa básica em Física e Engenharia de materiais, a criação dos computadores e assim, o século XX – principalmente sua segunda metade – nos mostrou como podemos fazer coisas incríveis com os recursos tecnológicos.

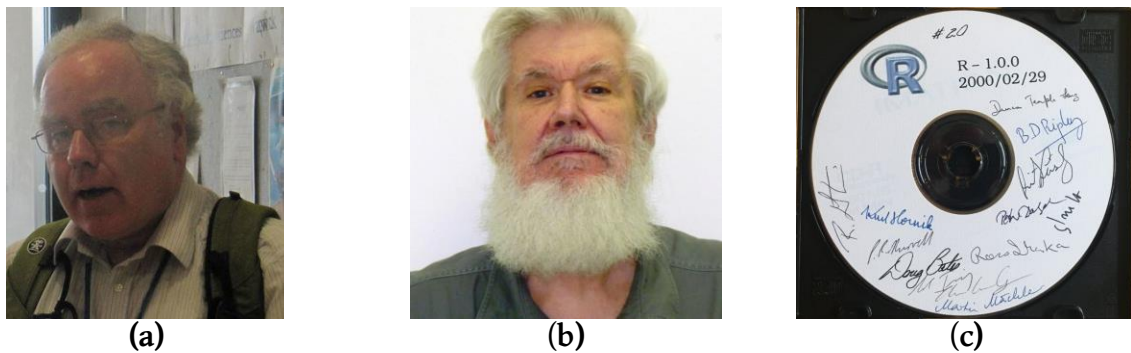
Essas cadeias se retroalimentam! Assim como a pesquisa básica permite o desenvolvimento de tecnologias, o uso de recursos tecnológicos impulsiona a ciência para velocidades nunca antes imaginadas. E isso não seria diferente com a Estatística Aplicada. O poderio da Estatística pura vem sendo cada vez mais velozmente implementado em linguagens computacionais, e isso coloca à disposição do pesquisador aplicado um arsenal atualmente invejável, se comparado ao disponível no final do século passado.

A linguagem de programação S foi desenvolvida por John Chambers, Rick Becker e Allan Wilks nos laboratórios Bell, na década de 1970. Em seguida, a empresa Macsoft lançou o software estatístico S+, baseado na linguagem S. Esse software era proprietário e vendido a preços altos para a época. O S+ se tornou popular entre estatísticos, mas com o desenvolvimento do sistema operacional Linux e a resistência/demora do lançamento do S+ para essa plataforma, surge o software R (Figura 1c), que aproveita a oportunidade para se impulsionar para o mundo!

O software R surge com a criação da linguagem de programação R. Criada pelos pesquisadores Brian Ripley e Bill Venables (Figura 1a e 1b), a linguagem R tem a aparência da

linguagem S e a semântica da linguagem Scheme. Esses pesquisadores formaram um time de especialistas, o R Core Team, que matem até hoje o software atualizado, de código aberto e gratuito. A primeira versão do software R foi lançada propositalmente aos 29 dias do mês de fevereiro do ano 2000, ano bissexto.

**Figura 1:** Criadores da linguagem e programa estatístico R.



**Legenda:** (a) – Brian Ripley; (b) – Bill Venables; (c) – Primeira versão do R.  
**Fonte:** Google, 2020.

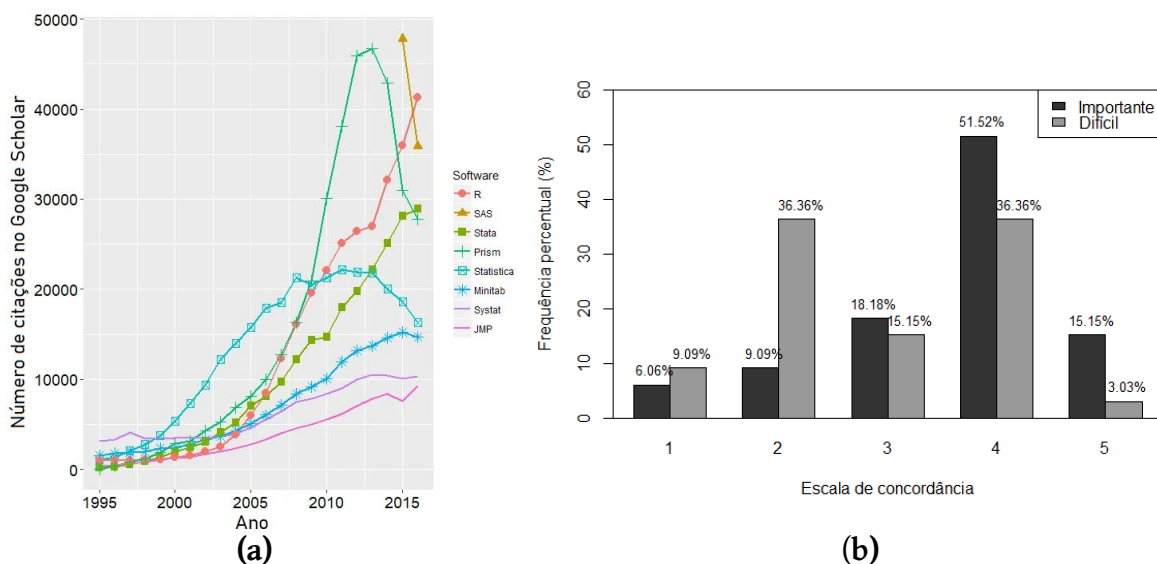
Desde então, o software R (R CORE TEAM, 2020) vem sendo utilizado de forma crescente por pesquisadores das mais variadas áreas do conhecimento. Enquanto sua primeira versão (v1.0.0) contava com apenas alguns pacotes de funções, sua versão v3.6.3 vem automaticamente com algumas dezenas de pacotes básicos, mas estão disponíveis 15.540 pacotes contribuídos, para download. Tais pacotes são especializados tanto a áreas da Estatística, como Econometria, Séries Temporais, Multivariada, Bayesiana, Regressão etc; como a outras áreas do conhecimento como Ecologia, Matemática, Ciências Sociais, Economia, Finanças, dentre outras (CRAN, 2020).

O crescente uso do software R foi bem descrito por Muenchen (2016). O autor mostra, entre outras muitas descobertas que, entre 1995 e 2016, o R teve seu uso continuamente aumentado demonstrando melhor desempenho entre todos os softwares ascendentes. Ele foi seguido pelo Stata, Systat e JMP, enquanto programas como Prism, Statistica e Minitab apresentaram queda no número de citações (Figura 2).

Particularmente, na pesquisa em Alimentos e Nutrição, o software R desempenha importante papel na análise, visualização e predição de dados. Por exemplo, Stefos e Kouliani (2016), juntamente com seus alunos universitários equatorianos, estudaram a relação entre qualidade dos alimentos e performance escolar de estudantes gregos de 10 a 12

anos. Eles utilizaram o software R e ferramentas de Estatística Descritiva e Classificação Hierárquica. As análises mostraram que estudantes que consomem frequentemente alimentos de baixa qualidade tendem a ter pior desempenho escolar. Ao final, os universitários responderam a questionários sobre a experiência de utilizar o R para fazer tais análises estatísticas, e o resultado foi altamente positivo. As Figura 2a e 2b apresentam resultados adaptados desses autores e mostram que boa parte dos estudantes (>65%) creem que o R será importante e sua vida profissional e mais de 50% dos graduandos não acharam o R um software difícil de se aprender.

**Figura 2:** Evolução temporal de softwares estatísticos e satisfação de estudantes ao utilizar o R para analisar dados nutricionais.



**Legenda:** (b) 1: discordo fortemente; 2: discordo; 3: não concordo nem discordo; 4: Concordo; 5: concordo fortemente.

**Fonte:** (a) Adaptado de Muenchen (2016) e (b) de Stefos e Koulianidi (2016).

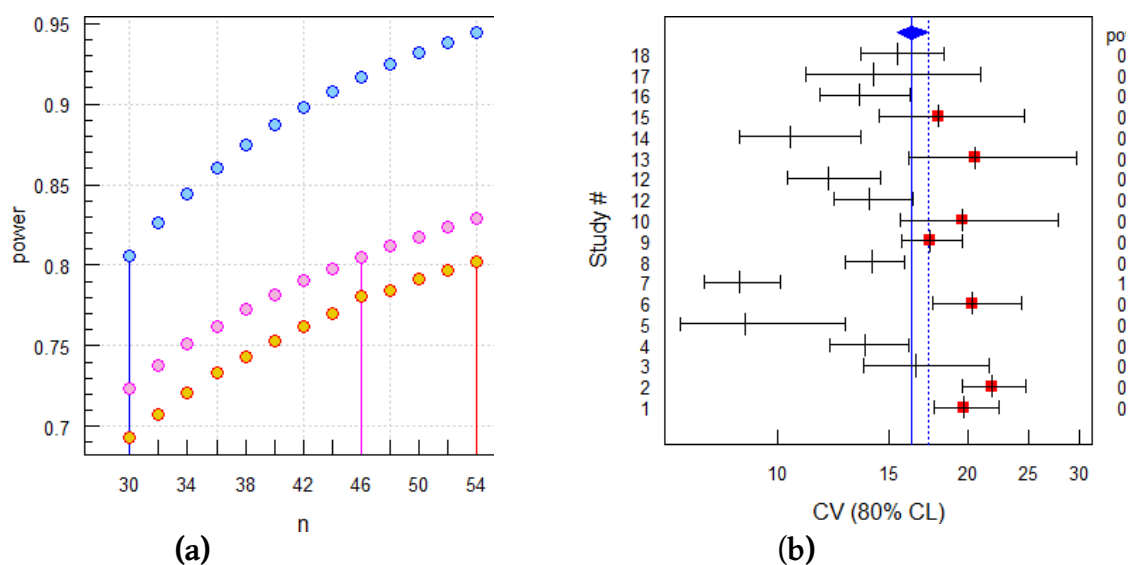
Particularmente, na pesquisa em Alimentos e Nutrição, o software R desempenha importante papel na análise, visualização e predição de dados. Por exemplo, Stefos e Koulianidi (2016), juntamente com seus alunos universitários equatorianos, estudaram a relação entre qualidade dos alimentos e performance escolar de estudantes gregos de 10 a 12 anos. Eles utilizaram o software R e ferramentas de Estatística Descritiva e Classificação Hierárquica. As análises mostraram que estudantes que consomem frequentemente alimentos de baixa qualidade tendem a ter pior desempenho escolar. Ao final, os universitários responderam a questionários sobre a experiência de utilizar o R para fazer tais

análises estatísticas, e o resultado foi altamente positivo. As Figura 2a e 2b apresentam resultados adaptados desses autores e mostram que boa parte dos estudantes (>65%) creem que o R será importante e sua vida profissional e mais de 50% dos graduandos não acharam o R um software difícil de se aprender.

O R possui um grande número de pacotes de funções destinados ao delineamento de ensaios clínicos, monitoramento e análise (CRAN, 2020). Como exemplos, podem ser citados o pacote TrialSize (ZHANG, 2013), que permite calcular o tamanho de amostras para ensaios clínicos; PowerTOST (LABES, SCHÜTZ e LANG, 2020), para calcular o poder e tamanho amostral em estudos de bioequivalência; survival (THERNEAU, 2020), para estudos de análise de sobrevivência; multcomp (HOTHORN, BRETZ e WESTFALL, 2008), que permite testes e intervalos de confiança simultâneos para diversas classes de modelos lineares, generalizados, mistos, paramétricos e não paramétricos; e muitos outros.

Tanto em experimentos em Nutrição quanto em Alimentos, é muito comum o uso da Análise de Variância (ANOVA). Para essa finalidade, um dos pacotes mais utilizados no Brasil é o ExpDes (FERREIRA, CAVALCANTI e NOGUEIRA, 2014) e sua versão em Português, ExpDes.pt (FERREIRA, CAVALCANTI e NOGUEIRA, 2011). Esses pacotes são capazes de analisar dados de experimentos balanceados simples e complexos em uma única rodada, por meio de modelos de efeitos fixos, conduzidos em delineamento inteiramente casualizado (DIC), em blocos casualizados (DBC) ou em Quadrado Latino (DQL), em esquema fatorial ou em parcelas subdivididas, com ou sem tratamentos testemunha.

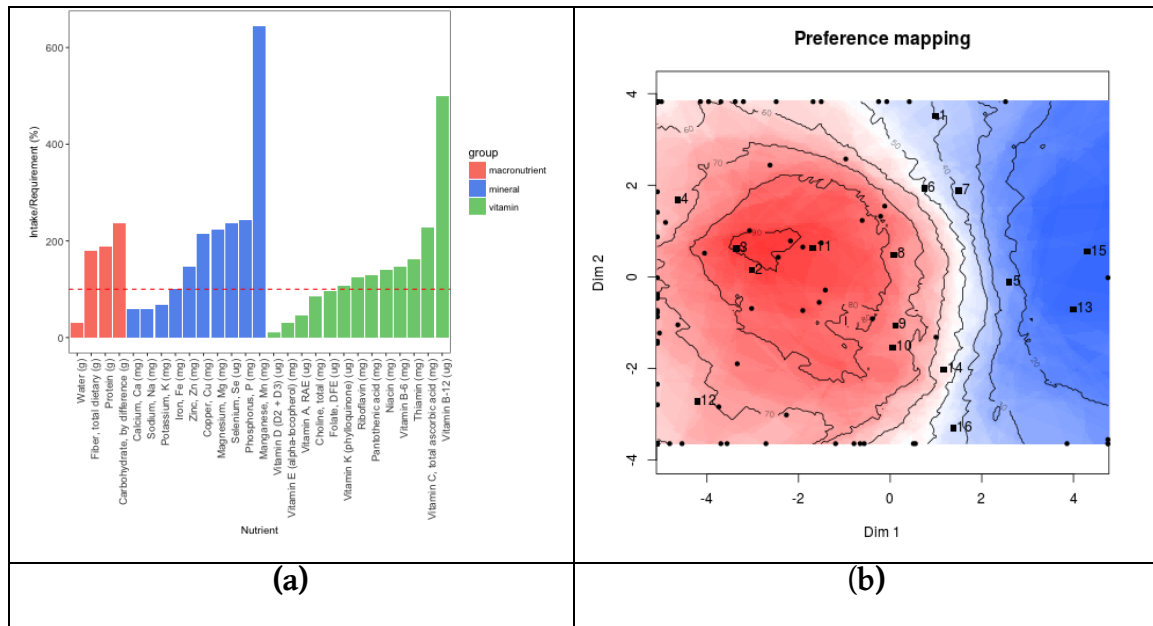
**Figura 3:** Cálculo de poder em ensaios clínicos (a) e intervalos de confiança simultâneos (b), ambos no pacote PowerTOST.



**Fonte:** Adaptado de Schütz (2019).

Para as pesquisas em Nutrição, o R dispõe também do pacote NutrienTracker (RODRIGUEZ-MARTINEZ et al., 2018a) que usa dados de composição de alimentos de vários bancos de dados de referência, como USDA (Estados Unidos), CIQUAL (França) e BEDCA (Espanha), e então calcula os níveis de ingestão de macronutrientes e micronutrientes (Figura 4a) e os compara com as doses dietéticas recomendadas.

**Figura 4** – Gráfico de ingestão de nutrientes (a) e mapa de preferência.



**Fonte:** (a) Adaptado de Rodriguez-Martinez et al. (2018b) e (b) de Wingfeet (2014).

Para análise sensorial de alimentos, o destacamos o pacote SensMineR (HUSSON, LE e CADORET, 2017) que reúne métodos muito clássicos, saídas gráficas fáceis de interpretar, bem como sínteses de resultados provenientes de vários modelos uni e multivariados. O SensMineR trata a caracterização de produtos, avaliação do desempenho de painel sensorial, links entre dados sensoriais e instrumentais, preferências do consumidor (Figura 4b), avaliação do *napping* e delineamentos de otimização.

## REFERÊNCIAS BIBLIOGRÁFICAS

- CRAN Task Views. **R Project**. Disponível em: <<https://brieger.esalq.usp.br/CRAN/>>. Acesso em: 15 de abr. de 2020.
- FERREIRA, E.B.; CAVALCANTI, P.P.; NOGUEIRA, D.A. Experimental Designs: um pacote R para análise de experimentos. **Revista da Estatística da Universidade Federal de Ouro Preto**, v. 1, p. 1-9, 2011.
- FERREIRA, E.B.; CAVALCANTI, P.P.; NOGUEIRA, D.A. ExpDes: an R package for ANOVA and experimental designs. **Applied Mathematics**, v.5, 2014. p.2952-2958. Disponível em: <<http://dx.doi.org/10.4236/am.2014.519280>>.
- HOTHORN, T.; BRETZ, F.; WESTFALL, P. Simultaneous Inference in General Parametric Models. **Biometrical Journal**. v.50, n.3, p.346-363. 2008.
- HUSSON, F.; LE, S.; CADORET, M. **SensomineR: Sensory Data Analysis**. R package version 1.23. 2017. Disponível em: <<https://CRAN.R-project.org/package=SensomineR>>.
- LABES, D.; SCHÜTZ, H.; LANG, B. **PowerTOST: Power and Sample Size for (Bio)Equivalence Studies**. R package version 1.4-9. 2019. Disponível em: <<https://CRAN.R-project.org/package=PowerTOST>>. Acesso em: 16 de abr. de 2020.
- MUENCHEN, R.A. **The Popularity of Data Science Software**. R4stats. 2016. Disponível em: <<http://r4stats.com/articles/popularity/>>. Acesso em: 16 de abr. de 2020.
- R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2020. Disponível em: <<https://www.R-project.org/>>.
- RODRIGUEZ-MARTINEZ, A.; AYALA, R.; DEBBABI, Y.; VIDAL, L.S. **NutrienTracker: Food Composition Information and Dietary Assessment**. R package version 0.99.2. 2018a. Disponível em: <<https://CRAN.R-project.org/package=NutrienTracker>>.
- RODRIGUEZ-MARTINEZ, A.; AYALA, R.; DEBBABI, Y.; VIDAL, L.S. **NutrienTracker**. 2018b. Disponível em: <<https://cran.r-project.org/web/packages/NutrienTracker/vignettes/NutrienTracker.html>>. Acesso em: 16 de abr. de 2020.
- SCHÜTZ, H. **Average Bioequivalence**. PowerTOST. 2019. Disponível em: <<https://cran.r-project.org/web/packages/PowerTOST/vignettes/ABE.html>>. Acesso em: 16 de abr. de 2020.
- STEFOS, E.; KOULIANIDI, G. Nutrition Data Analysis Using R: Applications in Higher Education. **Health Sciences Research**. v.3, n.1, 2016, p.10-16. Disponível em: <<https://www.puce.edu.ec/sitios/biblioteca/pdf/EfstathiosStefos2.pdf>>. Acesso em: 16 de abr. de 2020.
- THERNEAU, T. **A Package for Survival Analysis in R**. R package version 3.1-12. 2020. Disponível em: <<https://CRAN.R-project.org/package=survival>>. Acesso em: 16 de abr. de 2020.
- WINGFEET. **Trying a premap**. R-bloggers. 2014. Disponível em: <<https://www.r-bloggers.com/trying-a-prefmap/>>. Acesso em: 16 de abr. de 2020.
- ZHANG, E.; WU, V.Q.; CHOW, S.; ZHANG, H.G. **TrialSize: R functions in Chapter 3, 4, 6, 7, 9, 10, 11, 12, 14, 15**. R package version 1.3. 2013. Disponível em: <<https://CRAN.R-project.org/package=TrialSize>>. Acesso em: 16 de abr. de 2020.



Este trabalho está licenciado com uma Licença [Creative Commons - Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/).