

Análise estatística por aleatorização, *bootstrap* e Monte Carlo

Pavel Dodonov
pdodonov@gmail.com

Considerações finais

Considerações finais

Bibliografia

Significância

Descrevendo algoritmos

Apresentando os p-valores

Limitações dos testes por aleatorização

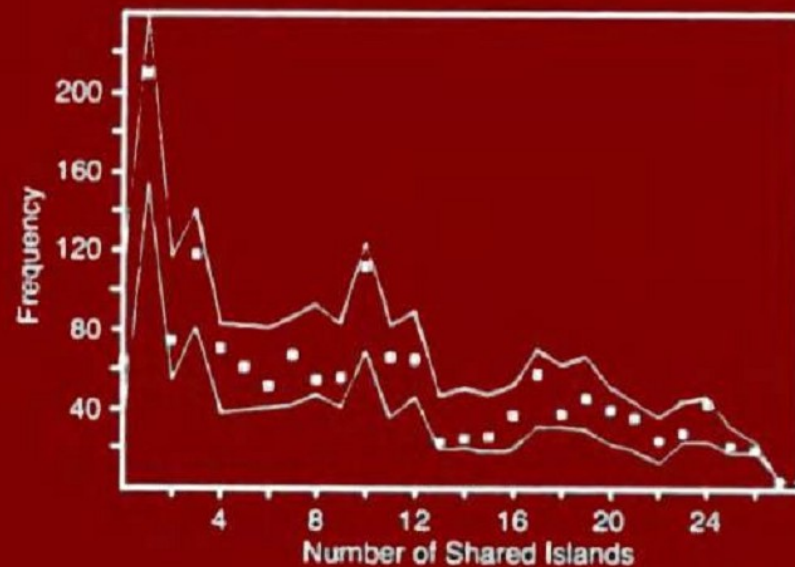
Justificativa para testes por aleatorização

Fim

Bibliografia

Randomization, Bootstrap and Monte Carlo Methods in Biology

Third Edition



Bryan F. J. Manly

Advanced R

Hadley Wickham



CRC Press

Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Use R!

Christian P. Robert
George Casella

Introducing Monte Carlo Methods with R

 Springer



Community Experience Distilled

R Object-oriented Programming

A practical guide to help you learn and understand the programming techniques necessary to exploit the full power of R

Kelly Black

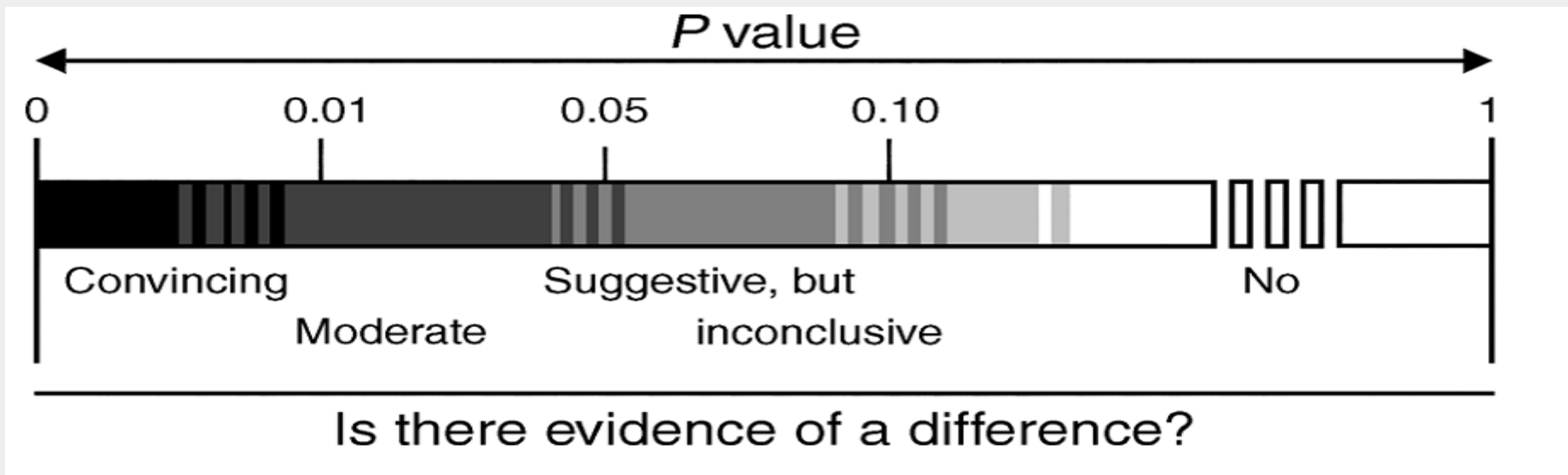
[PACKT] open source*
PUBLISHING community experience distilled

Significância

Significância

What are P values for? The purpose of P values is to convince a skeptic that a pattern in data is real. Or, when you are the skeptic, the purpose of P values is to convince others that a pattern in data could plausibly have arisen by chance alone. When there is a scientific need for skeptical reasoning with noisy data, the logic of P values is inevitable.

Significância



Significância

Ecology, 95(3), 2014, pp. 611–617
© 2014 by the Ecological Society of America

In defense of P values

PAUL A. MURTAUGH¹

Department of Statistics, Oregon State University, Corvallis, Oregon 97331 USA

noting that most of them “concern the misuse and misinterpretation of significance and P values by investigators and not the inherent properties ... of the tests or P values themselves,” and they make suggestions

ing on details of the application at hand. I conclude that P values, confidence intervals, and information-theoretic criteria all have their places in sound statistical practice, and that none of them should be excluded based on dogmatic, a priori considerations.

Significância

Ecology, 95(3), 2014, pp. 627–630
© 2014 by the Ecological Society of America

P values are only an index to evidence: 20th- vs. 21st-century statistical science

K. P. BURNHAM¹ AND D. R. ANDERSON

We were surprised to see a paper defending *P* values and significance testing at this time in history. We

[*available online*]).² Oakes (1986) summed it up over 25 years ago, “It is extraordinarily difficult to find a statistician who argues explicitly in favor of the retention of significance tests ...”

Significância

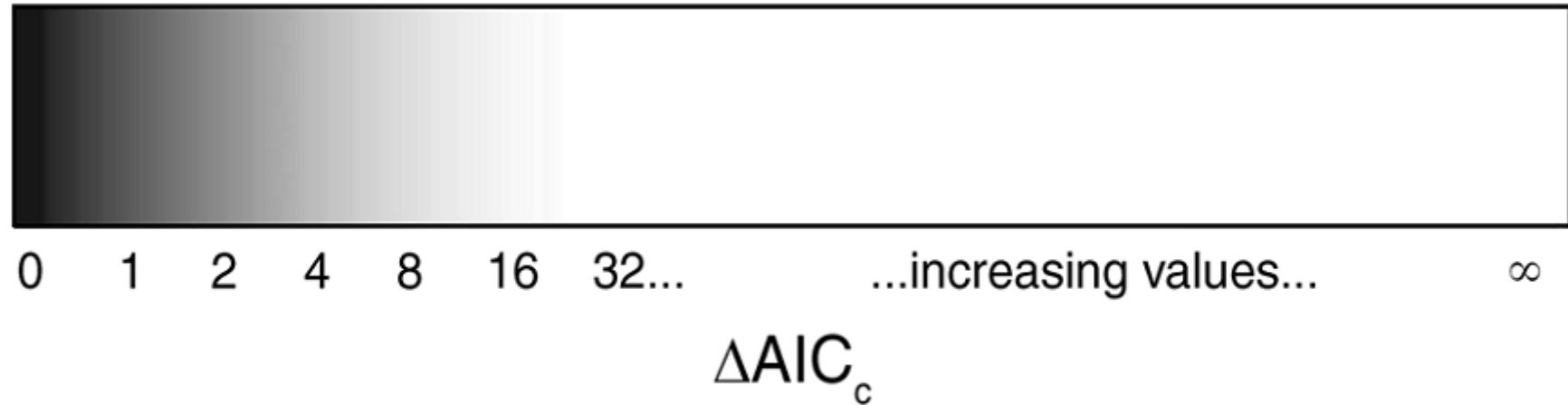


FIG. 3. Interpretation of ΔAIC , from Burnham et al. (2011). “Plausible hypotheses are identified by a narrow region in the continuum where $\Delta <$ perhaps four to seven (black and dark gray). The evidence in the light grey area is inconclusive and value judgments for hypotheses in this region are equivocal. Implausible models are shown in white, $\Delta >$ about 14.” (The

Significância

**Monte Carlo: a
significância pode
mudar entre
repetições da
mesma análise.**

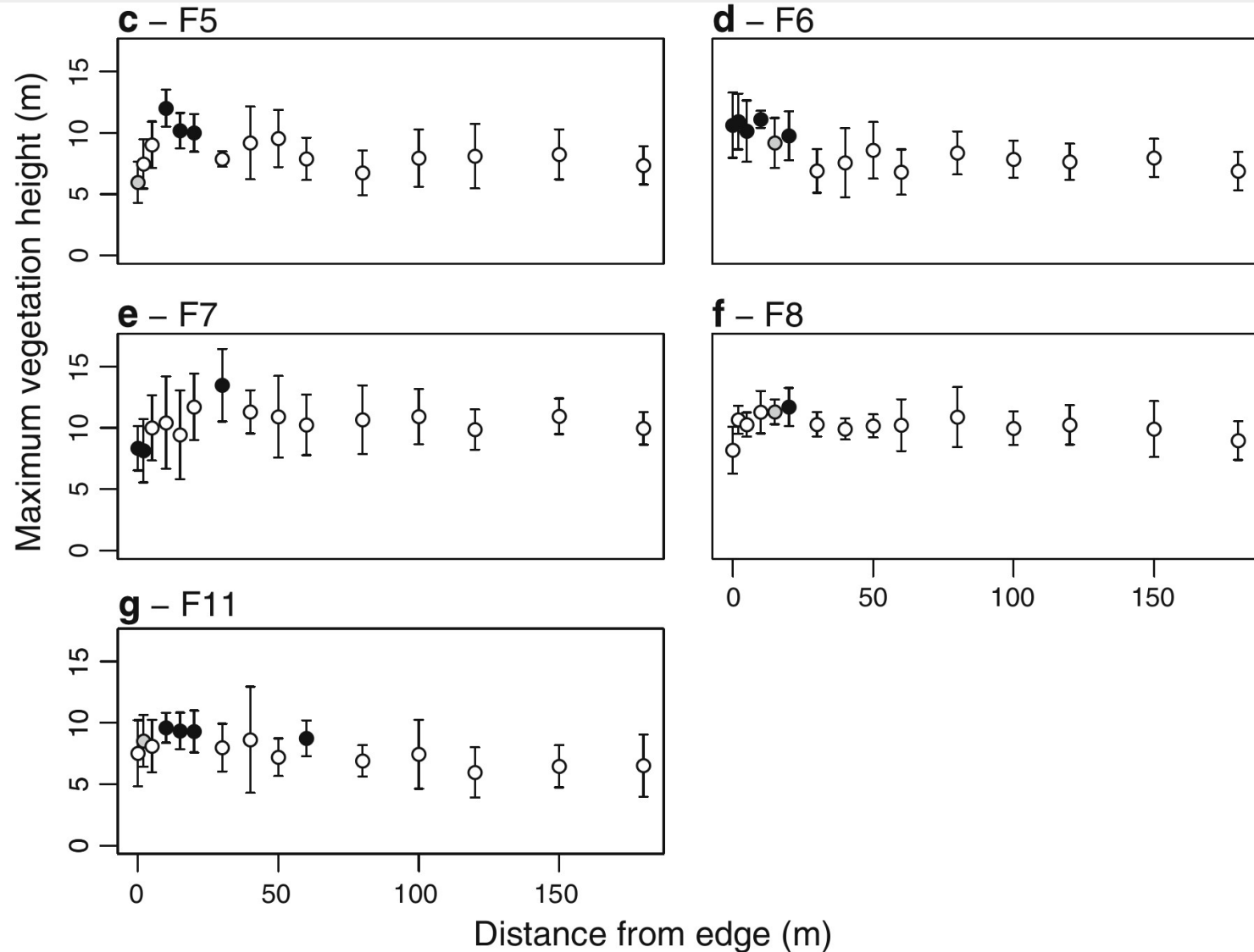
Significância

Monte Carlo: a significância pode mudar entre repetições da mesma análise.

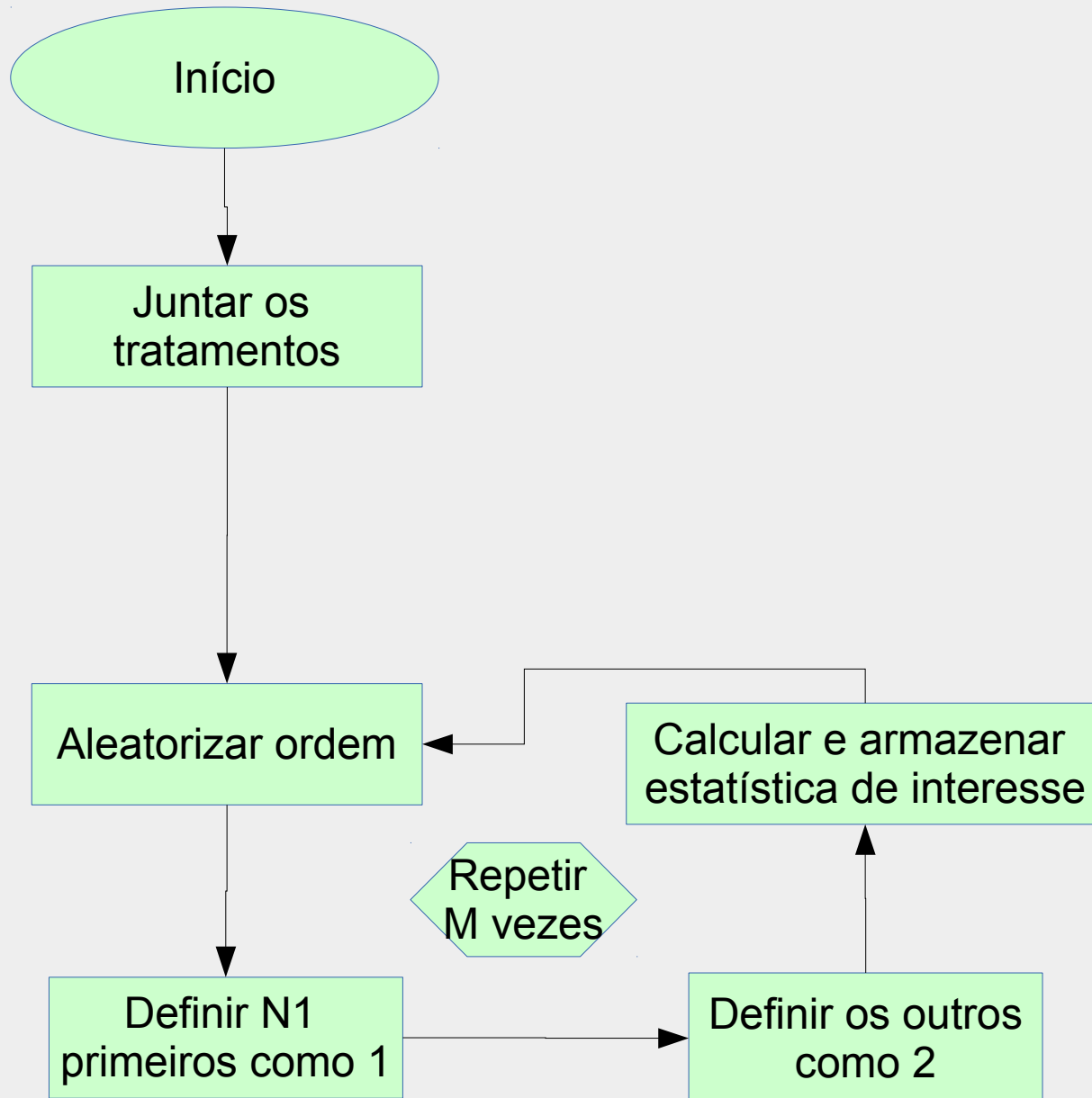
Não é recomendável usar um limiar (e.g. $p < 0.05$) para decidir se há um efeito significativo ou não.

Possível excessão

Vários testes sendo feitos; observamos o padrão geral encontrado (não cada teste individualmente)



Descrevendo algoritmos



Descrevendo algoritmos

This analysis proceeds as follows:

- 1. calculate the magnitude of edge influence using the values at a distance from the edge and the reference values;*
- 2. create a dataset containing both the edge values and the reference values;*
- 3. randomly assign five of these values as edge and the remaining as reference;*
- 4. recalculate magnitude of edge influence for the randomized dataset and repeat steps 2-4.*

The magnitude of edge influence values from 9 999 randomizations plus the observed value were then used to calculate the significance of the difference between edge and interior values for each distance. We considered distances with a p-value of 0.05 or lower as significantly different from reference conditions.

Descrevendo algoritmos

*Finally, we used a permutation t-test (Manly, 2007) to compare the mean number of *M. albicans* emergent seedlings between burned and unburned samples and between the depth levels (litter layer, 0-4 cm deep and 5-8 cm deep) for each sampling date, in Past 2.15 (Hammer et al., 2001). We also used bootstrap-T (Manly, 2007) to estimate confidence intervals for the mean number of seedlings, in R 2.14.2 (R Development Core Team, 2012; script in Supplementary Material S1)*

Descrevendo algoritmos

For the GLM randomizations, we (1) extracted the deviance of the observed relation between the response variable and habitat cover; (2) randomized the relation between the two variables; (3) performed a GLM on the randomized data set; and (4) extracted the deviance of this GLM. Significance was calculated as the proportion of deviance values in the randomized data set that were equal to or smaller than the observed deviance. (...) We used a total of 4999 randomizations and included the observed deviation in the comparison because it is one of the possible combinations under the null model (Manly 2007).

Relatando a significância

Improving the reporting of *P*-values generated by randomization methods

Graeme D. Ruxton^{1*} and Markus Neuhäuser²

- Existem alguns problemas em como os resultados de testes por permutação são relatados...

Improving the reporting of *P*-values generated by randomization methods

Graeme D. Ruxton^{1*} and Markus Neuhäuser²

- Valor real da estatística de teste
 - Pode ser incluído ou não
 - x_j : valor de cada permutação
 - X : valor real
 - I : 1 se a desigualdade for verdadeira, 0 se for falsa

$$P = \frac{1 + \sum_{j=1}^B I(x_j \geq X)}{B + 1}$$

$$P = \frac{\sum_{j=1}^B I(x_j \geq X)}{B}$$

Improving the reporting of *P*-values generated by randomization methods

Graeme D. Ruxton^{1*} and Markus Neuhäuser²

Maior garantia que o nível de significância seja o relatado

$$P = \frac{1 + \sum_{j=1}^B I(x_j \geq X)}{B + 1}$$

Os dados observados são uma das possibilidades se H_0 for verdadeira

Improving the reporting of *P*-values generated by randomization methods

Graeme D. Ruxton^{1*} and Markus Neuhäuser²

- Número de permutações

1000 para $p=0.05$,
5000 para $p=0.01$

(Manly 2007, p. 97)

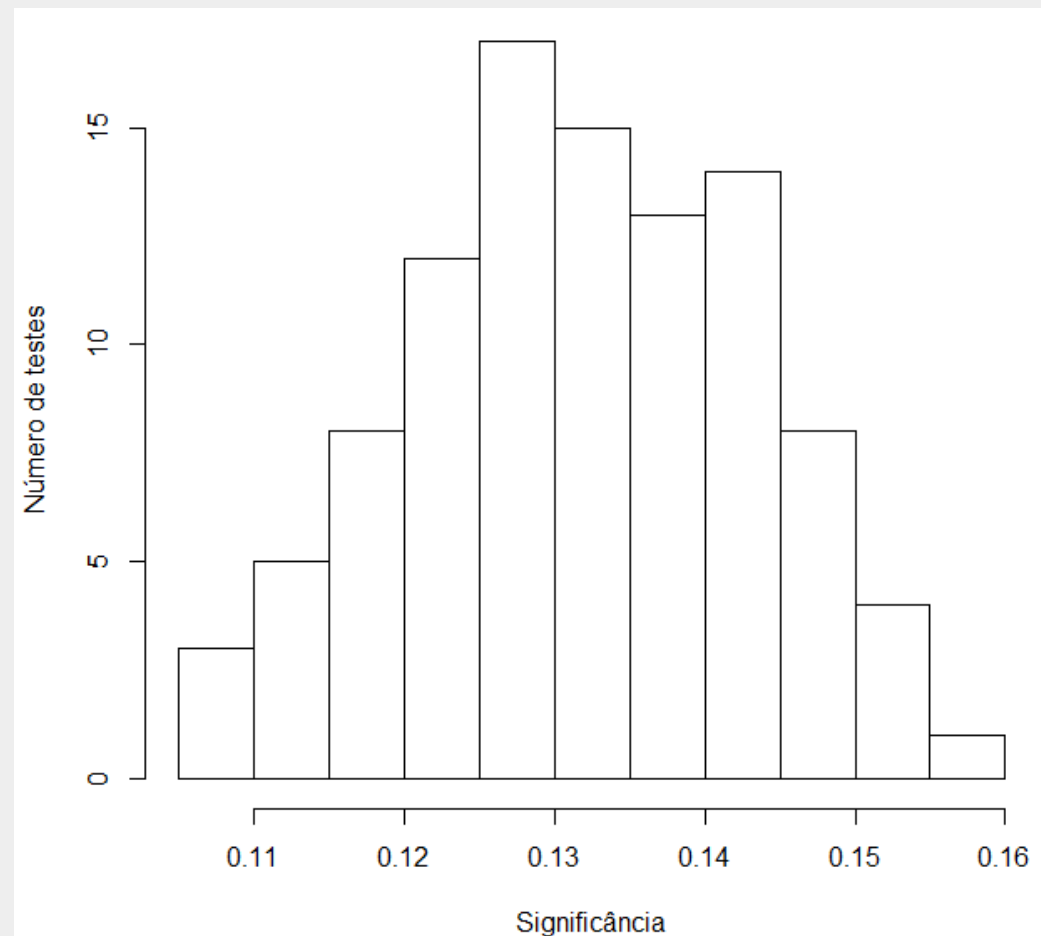
Significância real = 0.05		
	Limite	
N	Inferior	Superior
200	0.010	0.090
500	0.025	0.075
1000	0.032	0.068
5000	0.042	0.058

Significância real = 0.01		
	Limite	
N	Inferior	Superior
600	0.000	0.020
1000	0.002	0.018
5000	0.006	0.014
10000	0.007	0.013

Improving the reporting of *P*-values generated by randomization methods

Graeme D. Ruxton^{1*} and Markus Neuhäuser²

- Incerteza no valor do p
 - A distribuição de valores possíveis é apenas aproximada
 - Varia entre testes



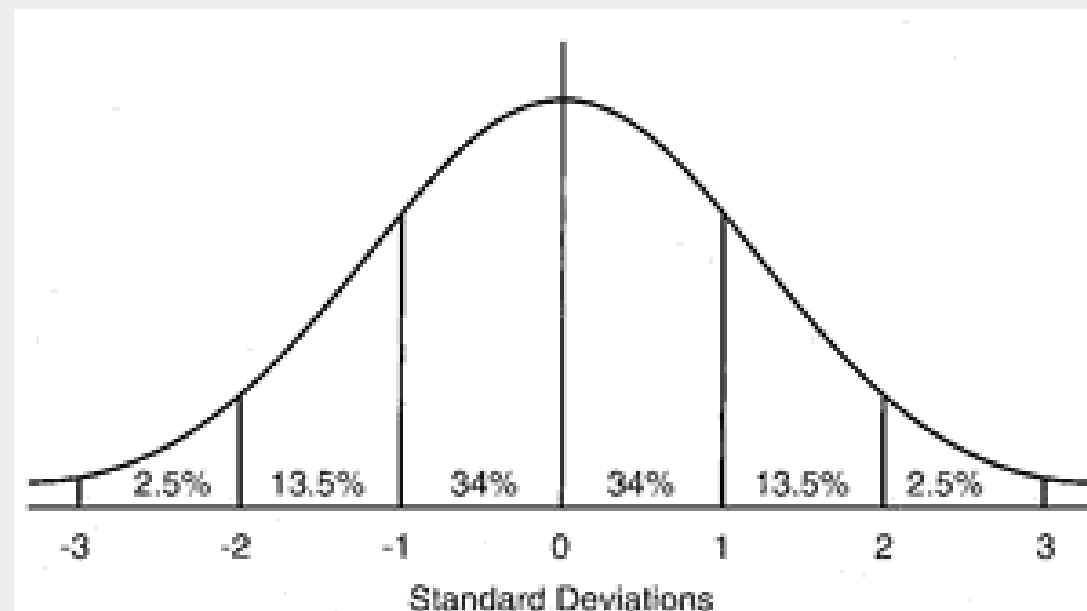
Improving the reporting of *P*-values generated by randomization methods

Graeme D. Ruxton^{1*} and Markus Neuhauser²

- Incerteza no valor do *p*

- A distribuição de valores possíveis é apenas aproximada

$$P \pm z_{1-0.5\alpha} \sqrt{\frac{P(1-P)}{N_{rand} + 1}}$$



Improving the reporting of *P*-values generated by randomization methods

Graeme D. Ruxton^{1*} and Markus Neuhauser²

- Incerteza no valor do *p*
 - A distribuição de valores possíveis é apenas aproximada

$$P \pm 1.96 \sqrt{\frac{P(1-P)}{N_{rand} + 1}}$$

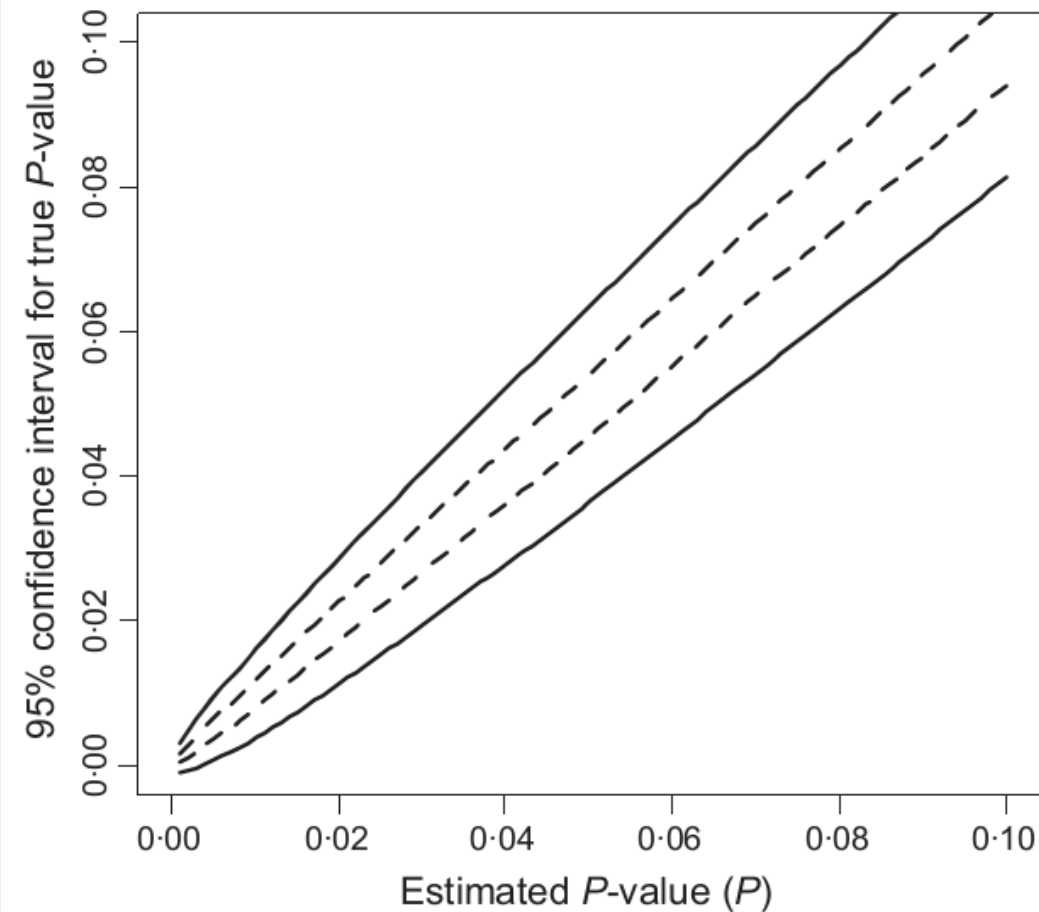
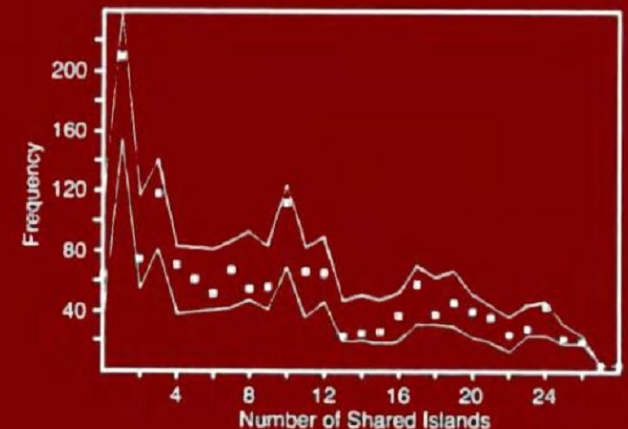


Fig. 1. The 95% confidence intervals calculated for the true *P*-value as a function of the estimated *P*-value *P* for *B* = 999 replicates (solid lines) and *B* = 9999 replicates (broken lines).

Limitações das análises por aleatorização

Randomization, Bootstrap and Monte Carlo Methods in Biology

Third Edition



Bryan F. J. Manly

Limitações das análises por aleatorização

Aleatorizações

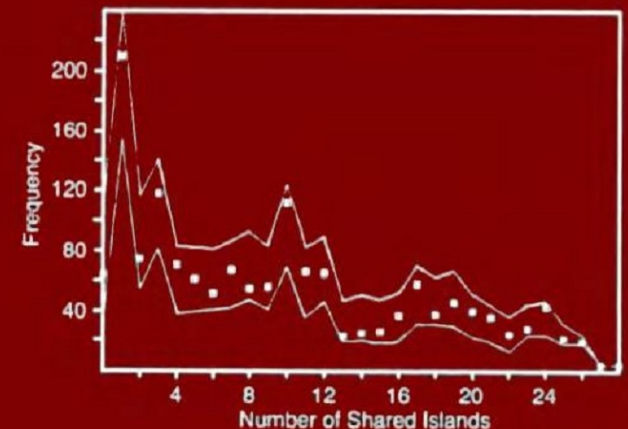
Comparações de dois ou mais grupos

Hipóteses de que a ordem das observações em um dos grupos é aleatória

Jusitifativas para um teste Monte Carlo

Randomization, Bootstrap and Monte Carlo Methods in Biology

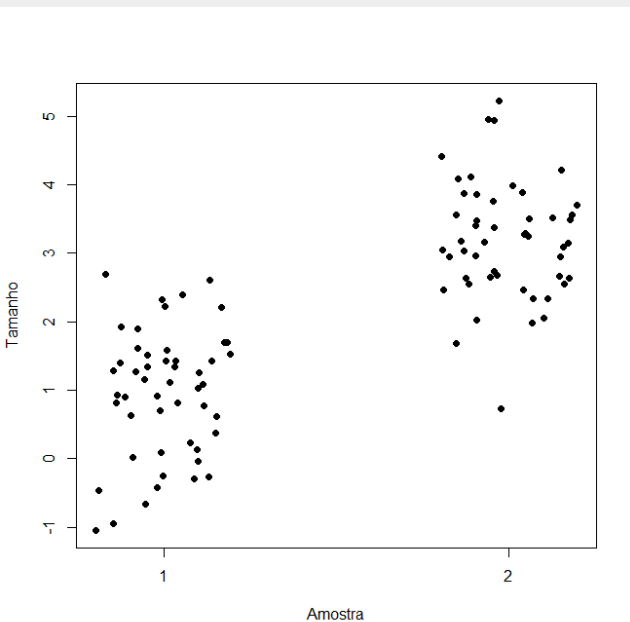
Third Edition



Bryan F. J. Manly

Comparação de duas amostras

“Esse teste pode ser geralmente justificado (...) se o desenho experimental envolve alocação aleatória de itens aos dois grupos antes deles serem tratados diferentemente ou se os dois grupos são amostras aleatórias independentes de duas populações.”



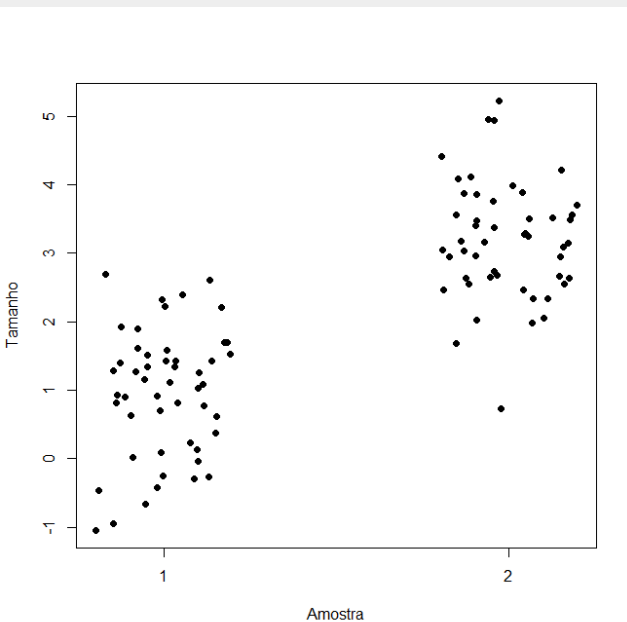
Comparação de duas amostras

Alocação aleatória

Na H_0 o experimento escolhe uma das alocações possíveis da distribuição aleatória.

Amostras aleatórias

Se as duas populações são iguais (H_0), todas as alocações de observações a amostras são igualmente prováveis.



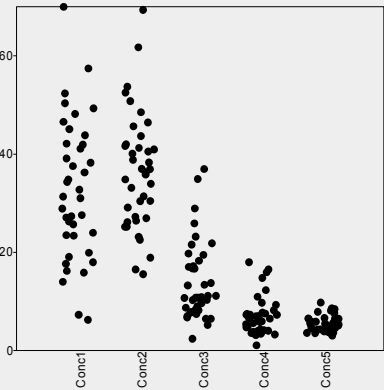
ANOVA

Alocação aleatória

Unidades amostras são alocadas aleatoriamente a grupos antes de receber diferentes tratamentos

H_0 : não há efeito do tratamento

As observações são feitas em uma ordem aleatória



Amostras aleatórias

Os grupos são amostras aleatórias de distribuições de itens possíveis

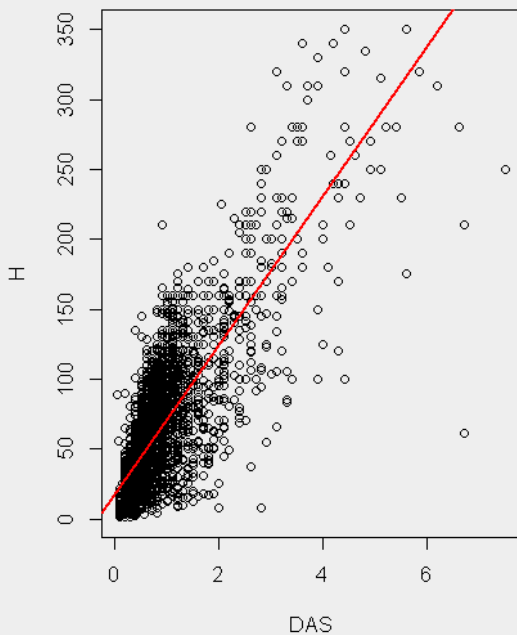
H_0 : as distribuições são as mesmas

Atribuições dos grupos são arbitrários

Todas as alocações de itens aos grupos são igualmente prováveis

Análise de regressão

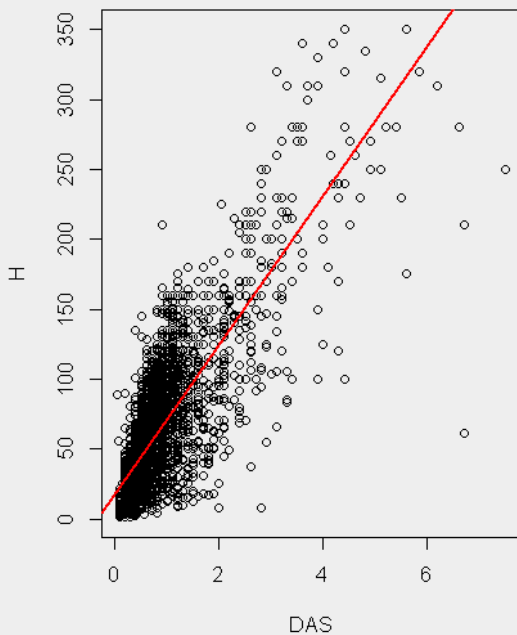
Os dados são uma amostra aleatória de uma população bivariada; H_0 de que X e Y são independentes e portanto todas as combinações possíveis de X e Y podem ocorrer com a mesma probabilidade.



Amostragem aleatória

Análise de regressão

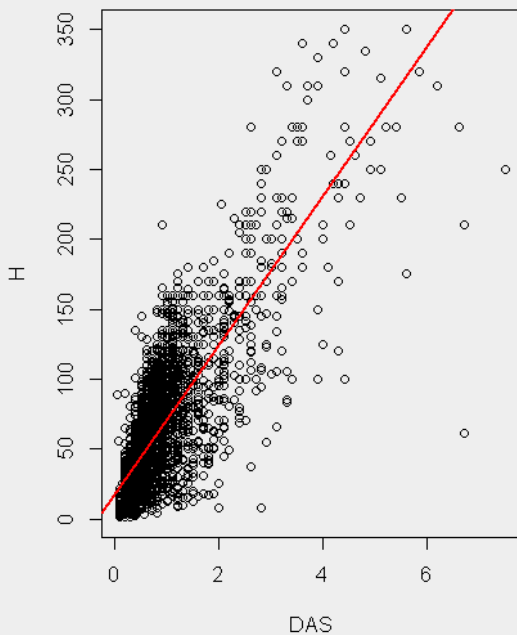
Os dados são obtidos de um experimento em que valores de X são atribuídos aleatoriamente a n unidades amostrais e uma resposta experimental Y é obtida. H_0 de que Y não tem relação com X .



Atribuição aleatória dos valores de X

Análise de regressão

Dados observacionais; H_0 de que o mecanismo que gera os dados torna os valores de X e Y não relacionados, de modo que todos os pareamentos possíveis são igualmente prováveis.

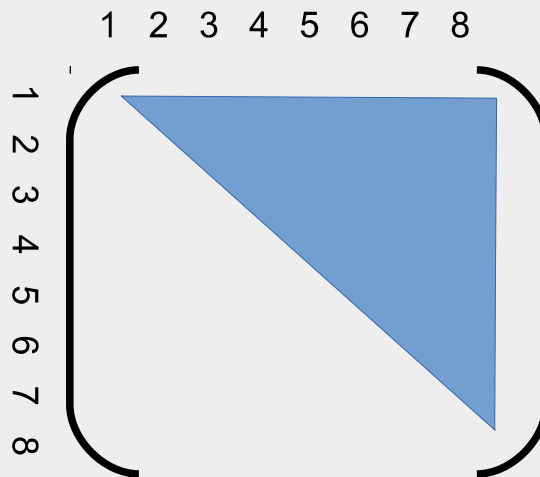
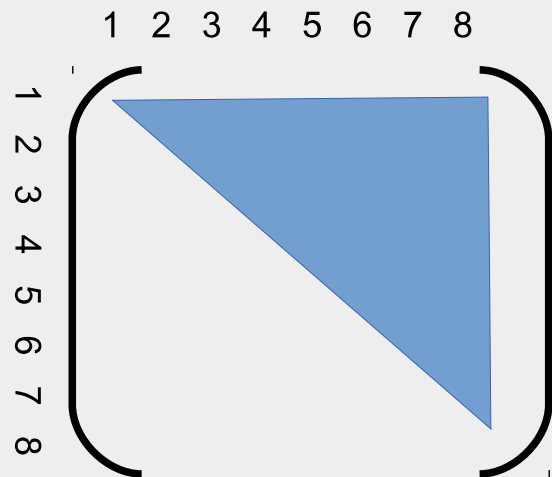


Justificativa mais fraca

Teste de Mantel

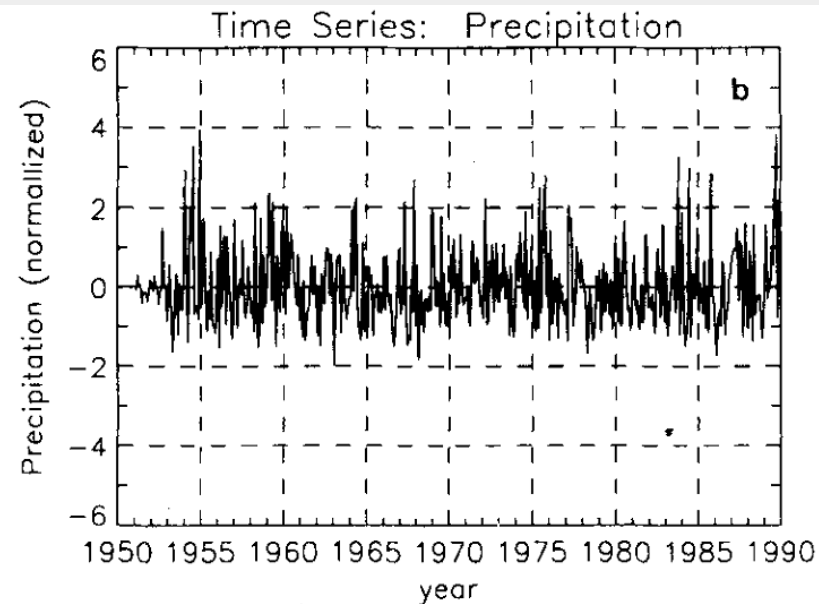
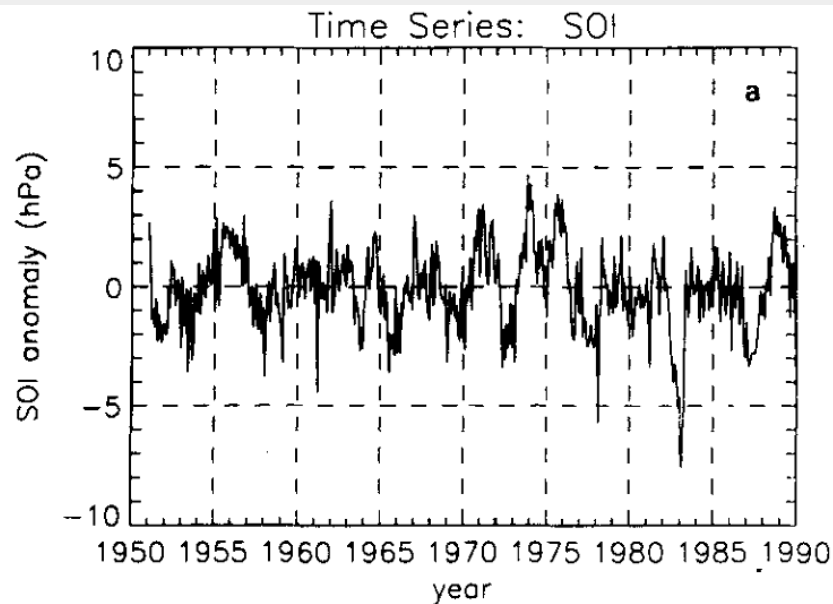
Amostragem aleatória

“Se as distâncias **A** são independentes das distâncias **B** na população, então escolher os mesmos n itens para as distâncias **A** e **B** equivale a tomar uma amostra aleatória de n itens para as distâncias **A** e uma amostra aleatória independente de n itens para as distâncias **B**. O processo amostral garante que todas as permutações dos itens da matriz **A** são igualmente prováveis em relação à ordem na matriz **B**, justificando o teste por aleatorização.”



Séries temporais

“A única justificativa para um teste por aleatorização é na crença de que o mecanismo gerador dos dados pode ser tal que qualquer valor observado pode ocorrer em qualquer posição da série temporal com a mesma probabilidade.”



Se a justificativa não for válida?

Simulações mais complicadas –
aleatorizações
restritas, cadeias
de Markov,
modelos
autoregressivos

Ou outra
abordagem
estatística

E se não der?

E se não der?



“Não se pode tirar leite de pedra, e nem mesmo a análise estatística mais sofisticada pode resgatar um desenho amostral ruim.”

Gotelli & Ellison 2004