

Análise estatística por aleatorização, *bootstrap* e Monte Carlo

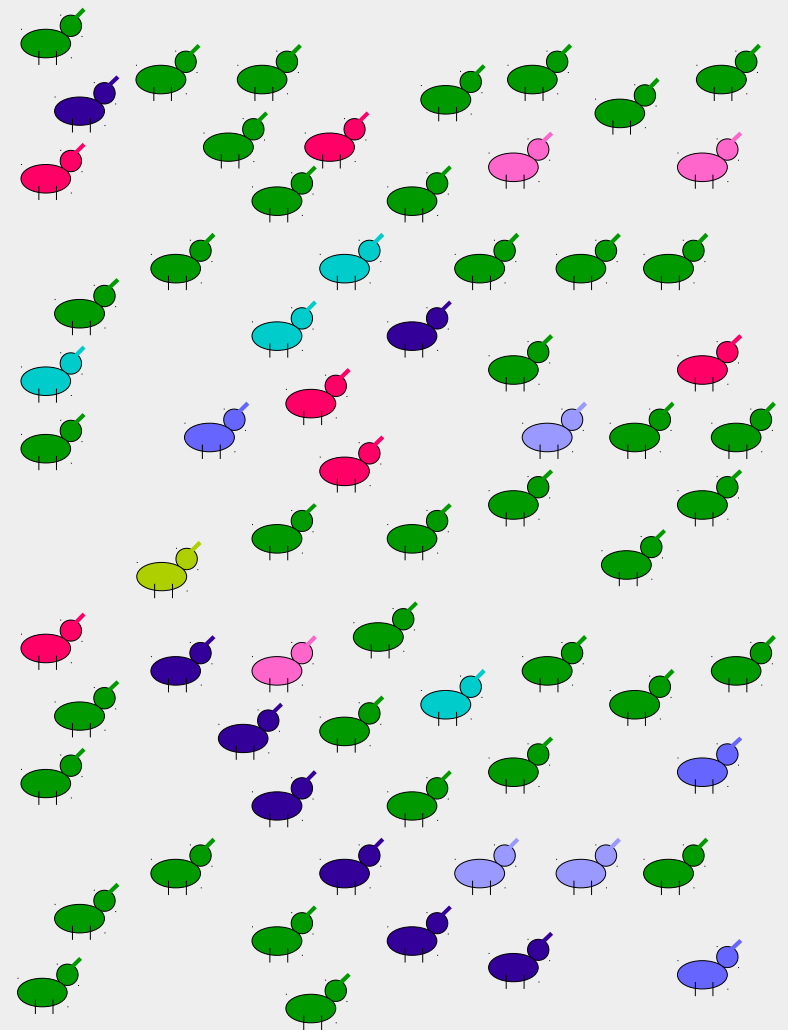
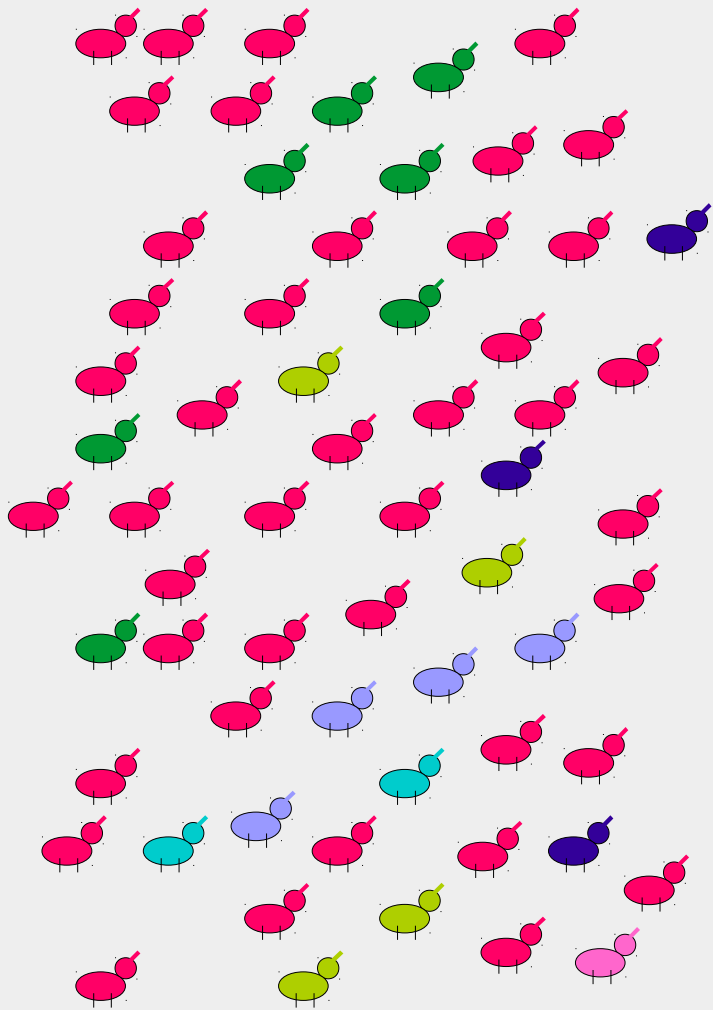
Pavel Dodonov
pdodonov@gmail.com

Laboratório de Ecologia Aplicada à Conservação (LEAC)
Universidade Estadual de Santa Cruz (UESC)
Ilhéus - BA

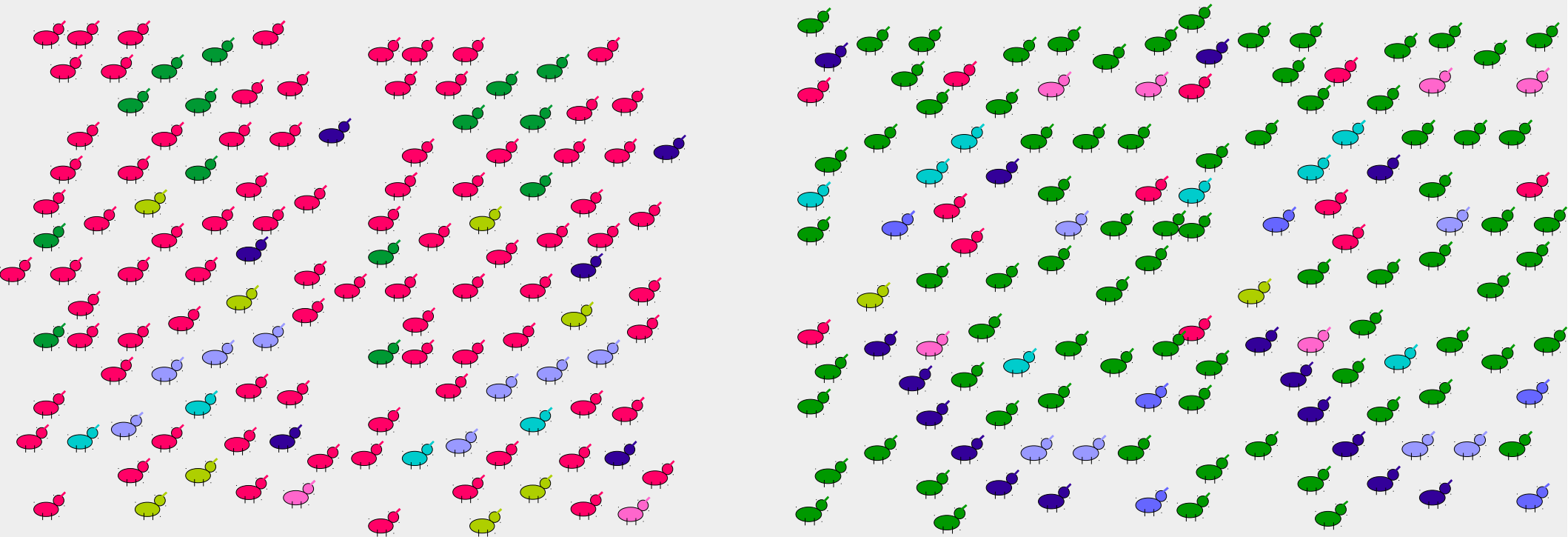
Aula teórica 2

Testes por permutação:
Comparando médias
Comparando variâncias
ANOVA
Regressão linear

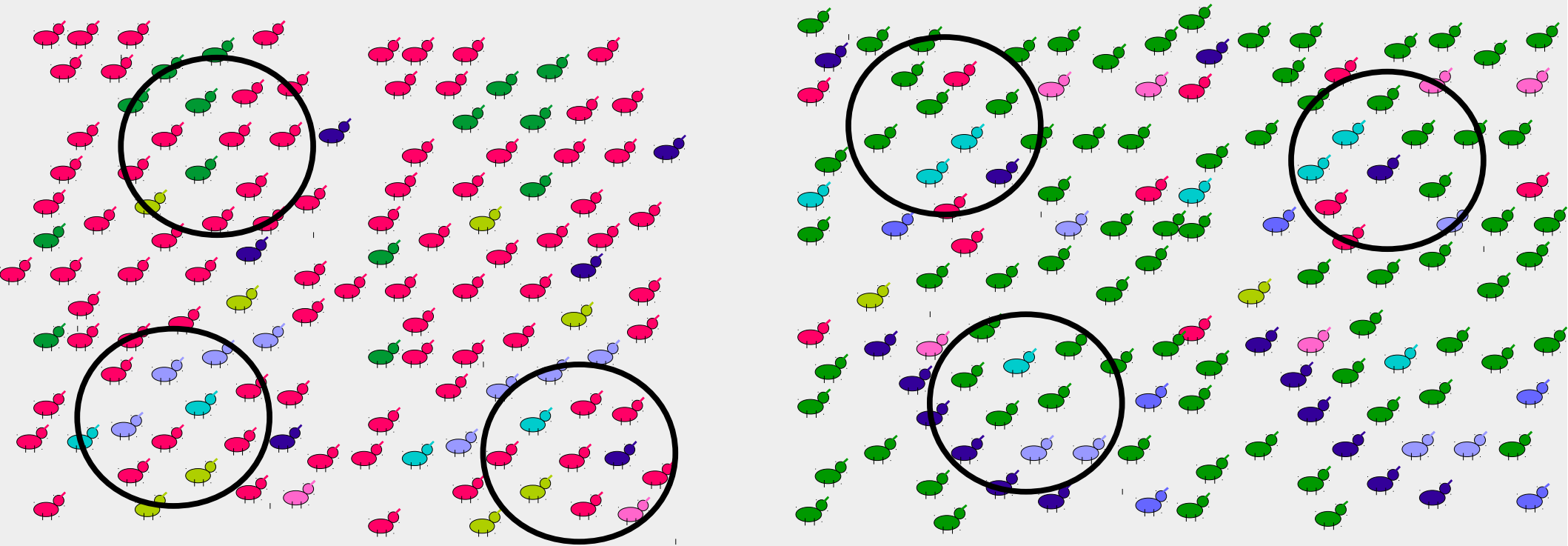
Duas amostras independentes



Hipótese: duas populações

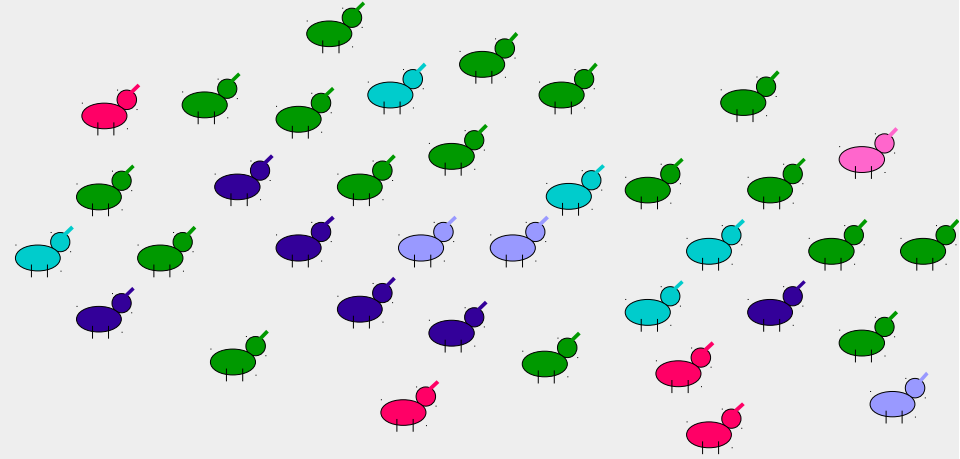
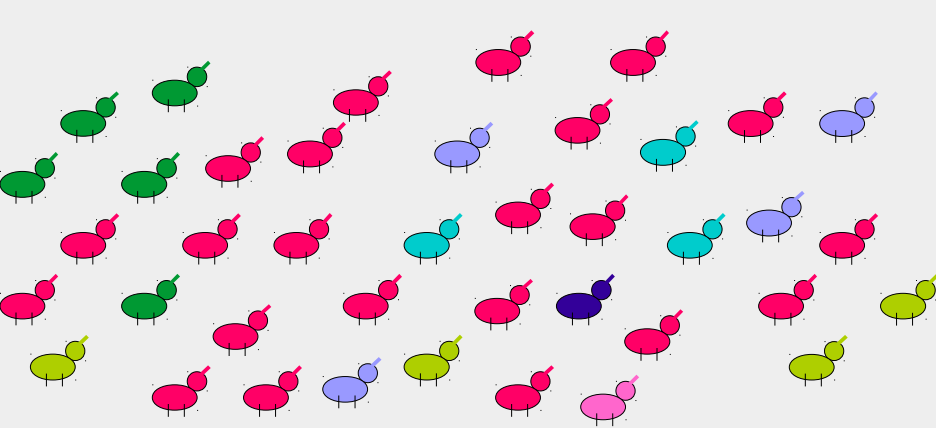


Hipótese: duas populações

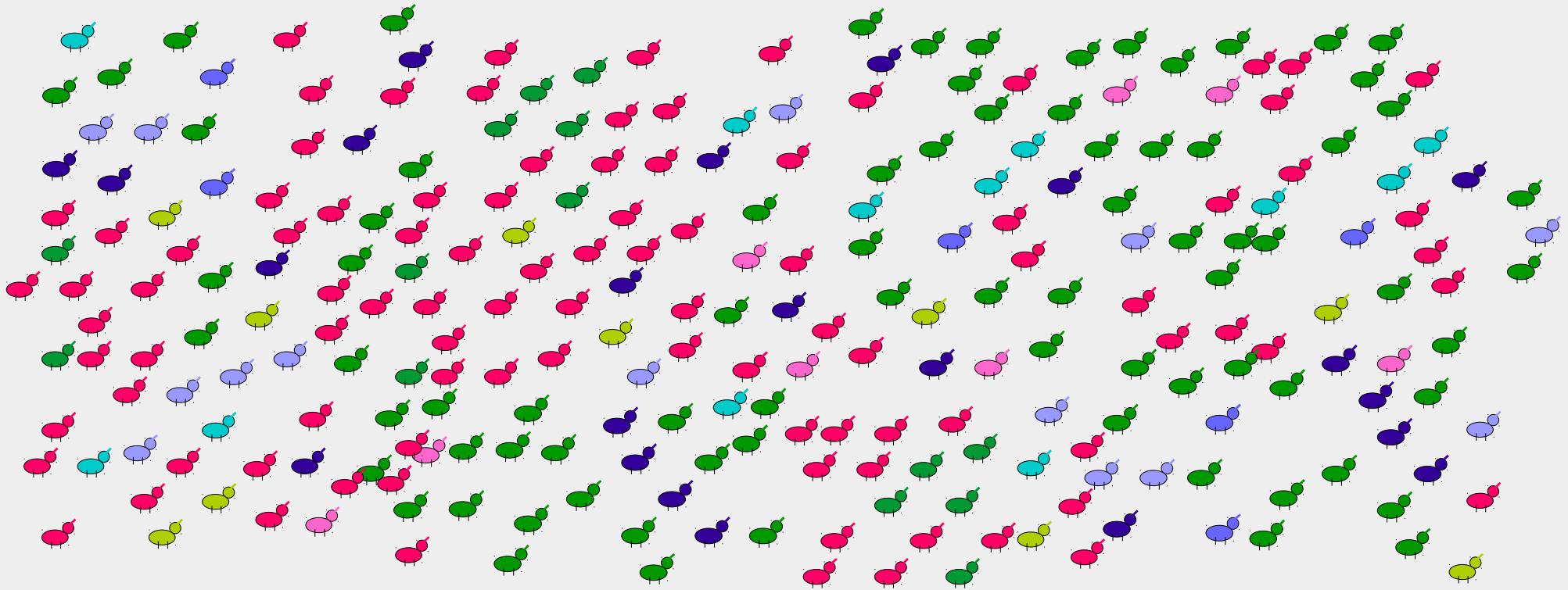


Amostragem

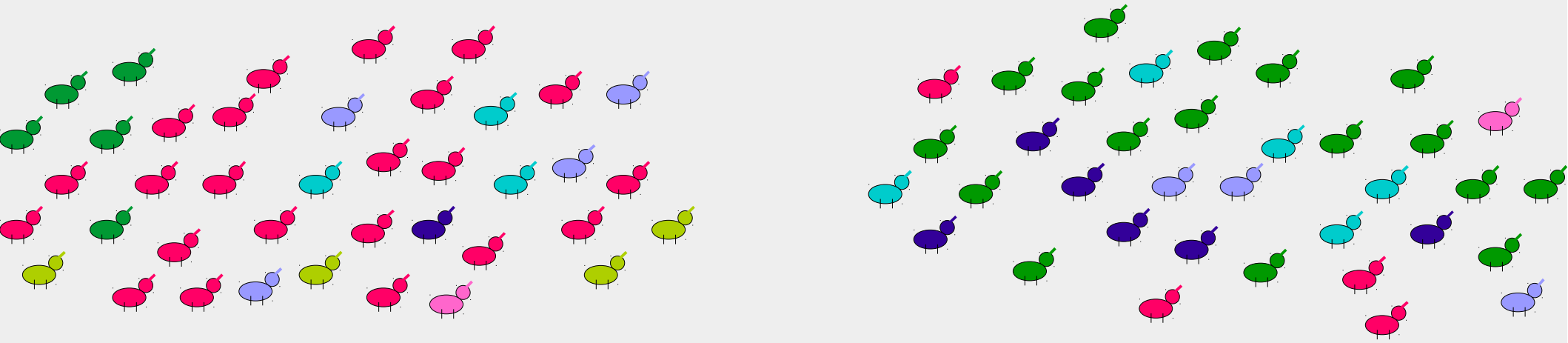
Duas amostras independentes



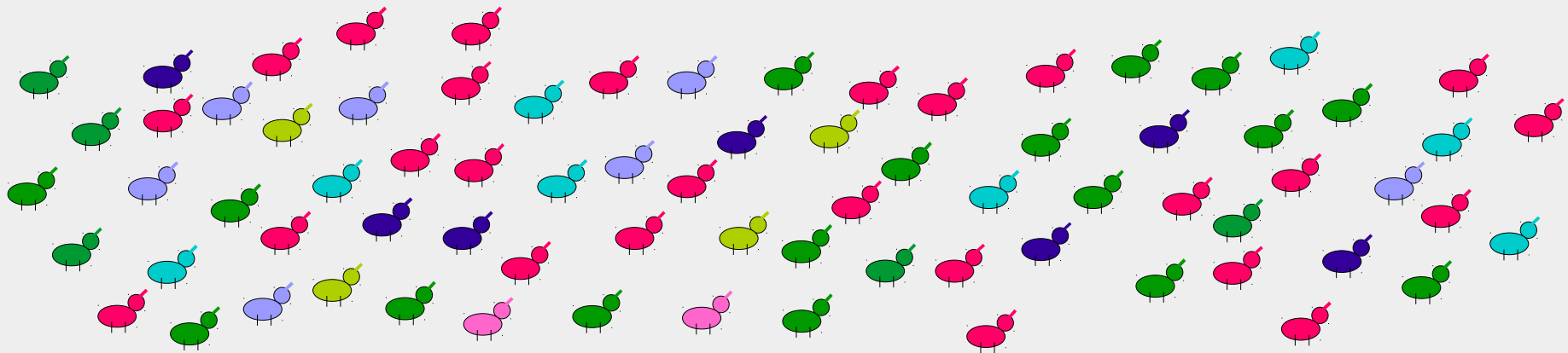
H_0 : Uma única população



Problema: Só temos as amostras!



Solução Monte Carlo: simula a população a partir das amostras



População simulada

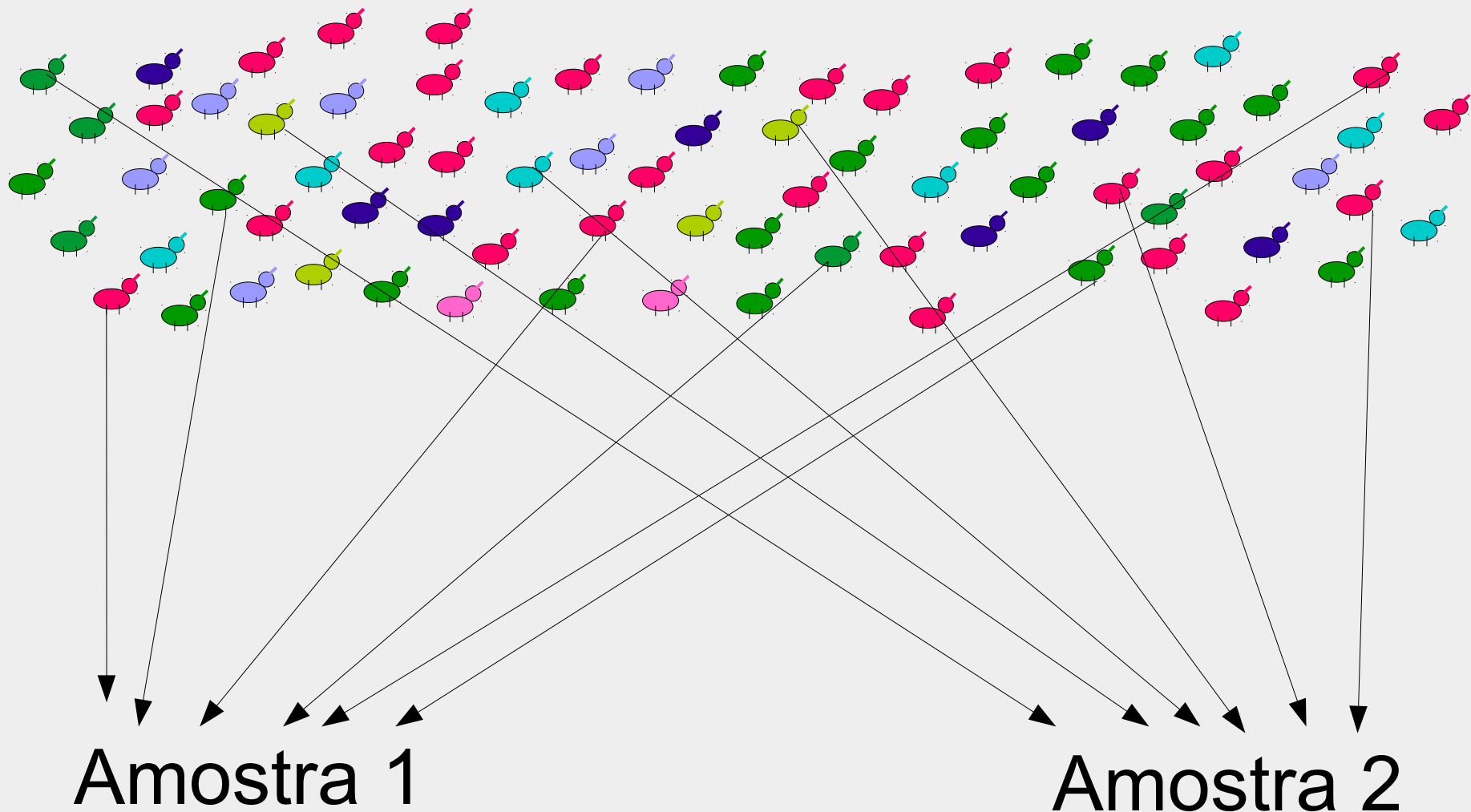
Hipótese nula

**Todos os
indivíduos fazem
parte da mesma
população**

Hipótese nula

**Cada indivíduo
tem a mesma
probabilidade de
entrar na amostra
1 ou na amostra 2**

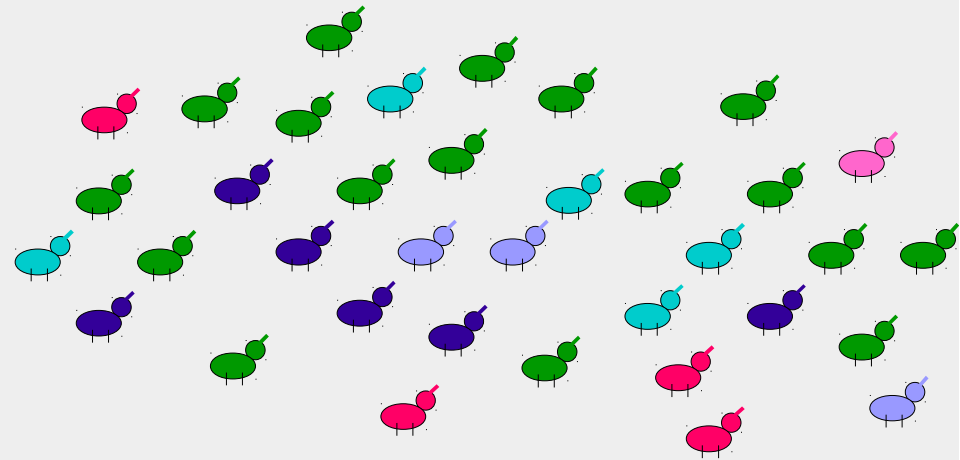
Solução Monte Carlo: simula a população a partir das amostras



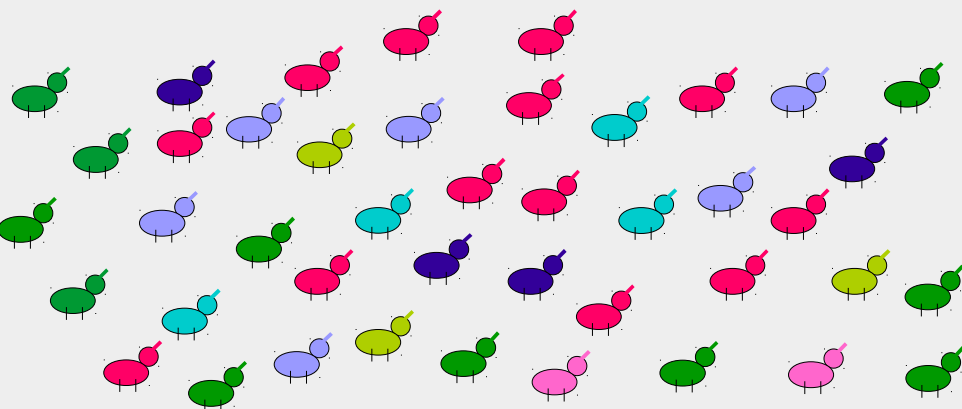
Observado VS Esperado na H_0



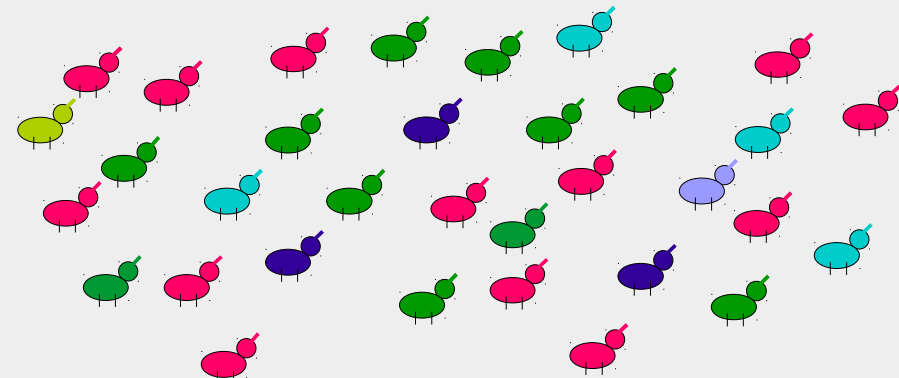
Amostra 1 - Observado



Amostra 2 - Observado



Amostra 1 - H_0

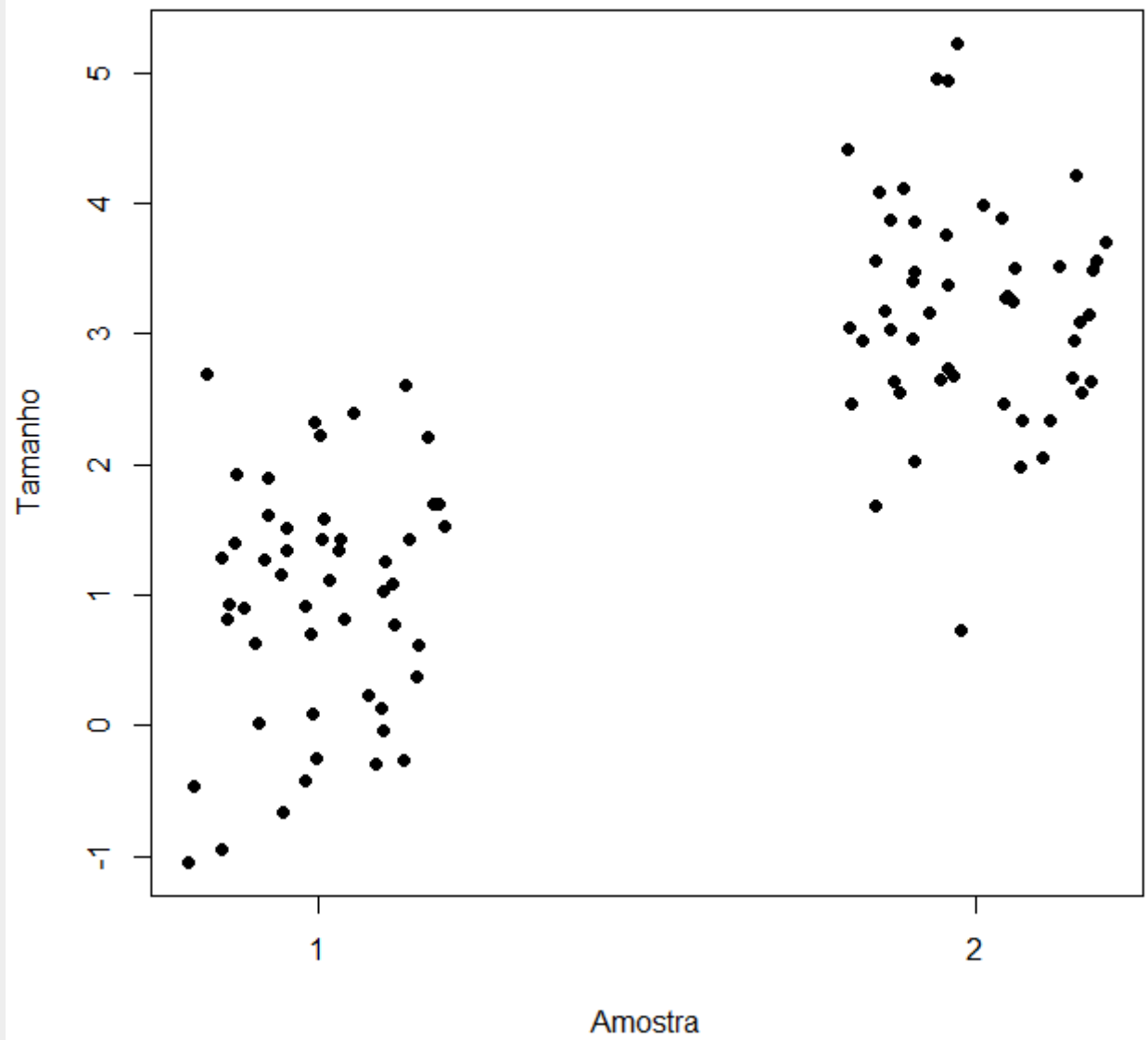


Amostra 2 - H_0

Abordagem Monte Carlo

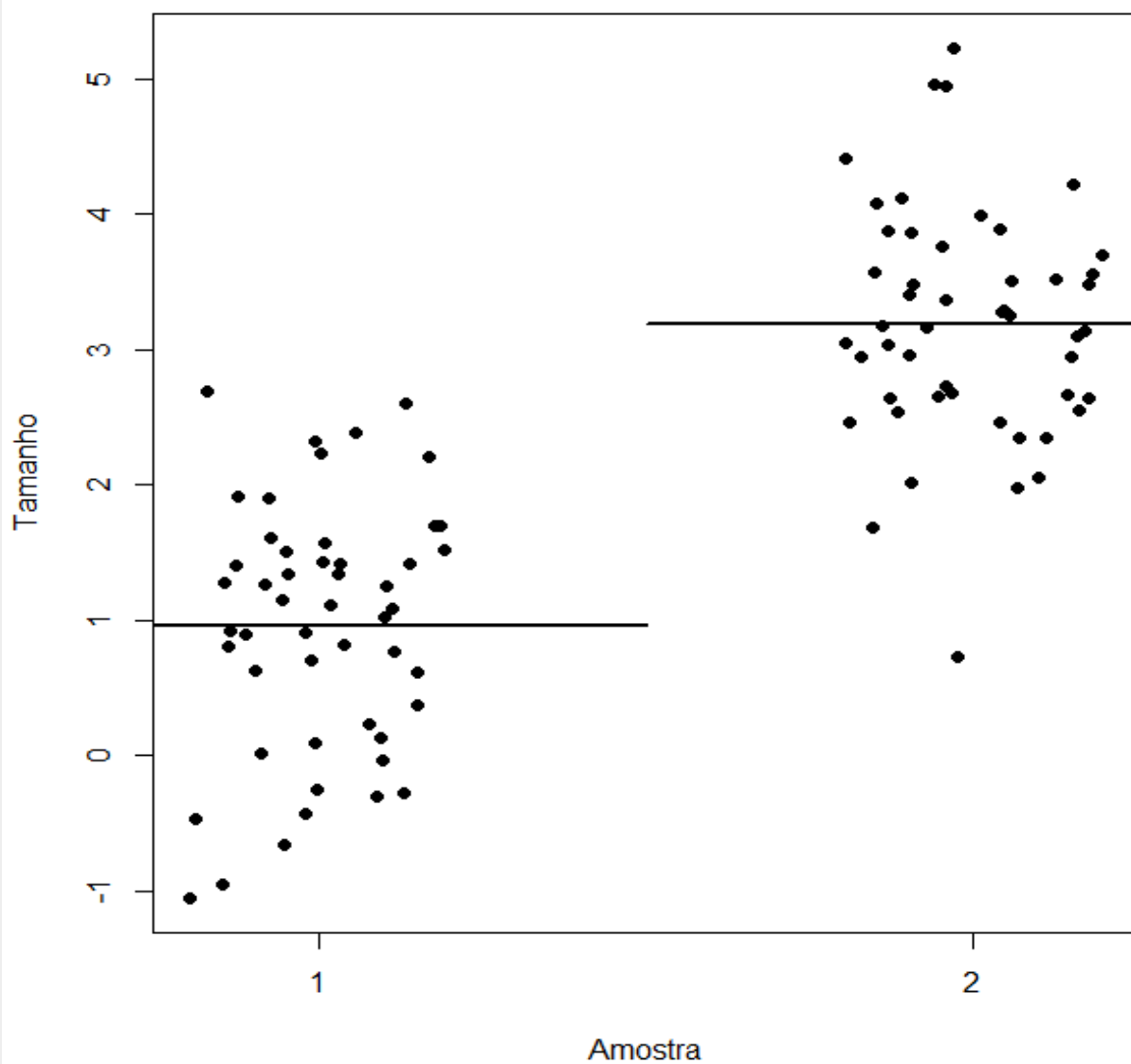
Realizar **muitas**
(e.g. 5000)
simulações da H_0
e comparar com o
valor **observado**

Exemplo numérico



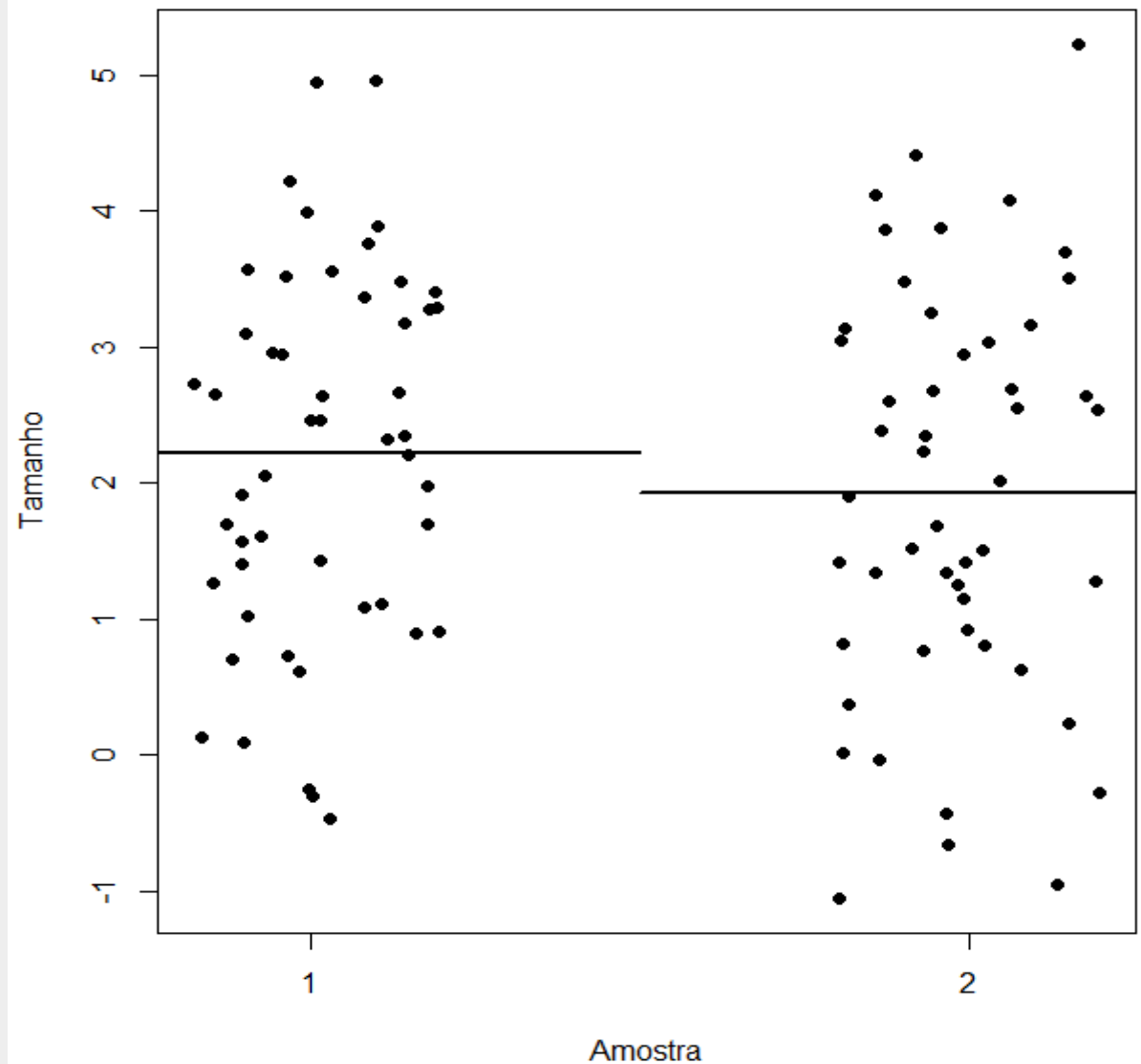
Exemplo numérico

Diferença nas
médias



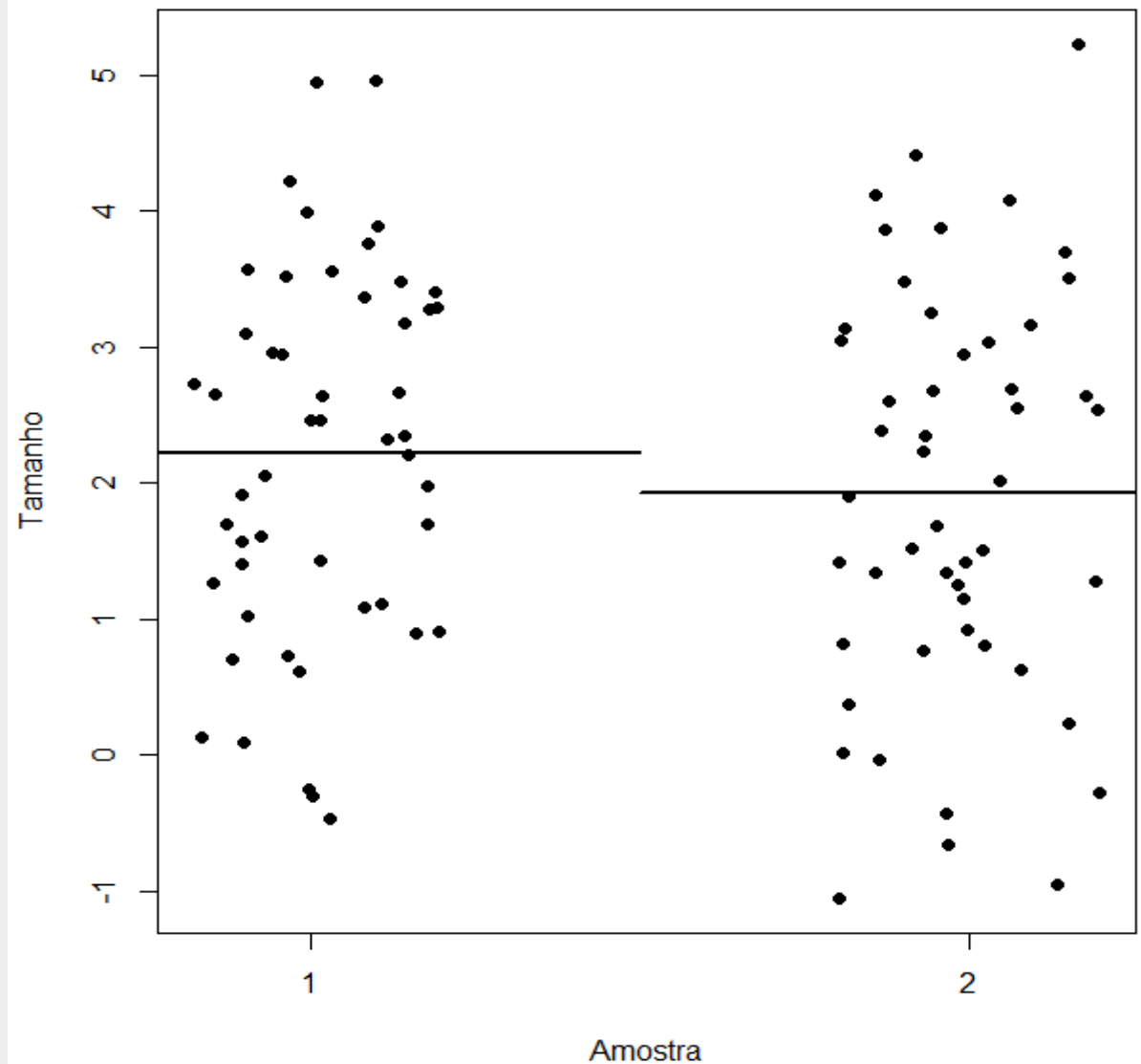
Exemplo numérico

Simulação –
aleatorizando
pontos entre
amostras



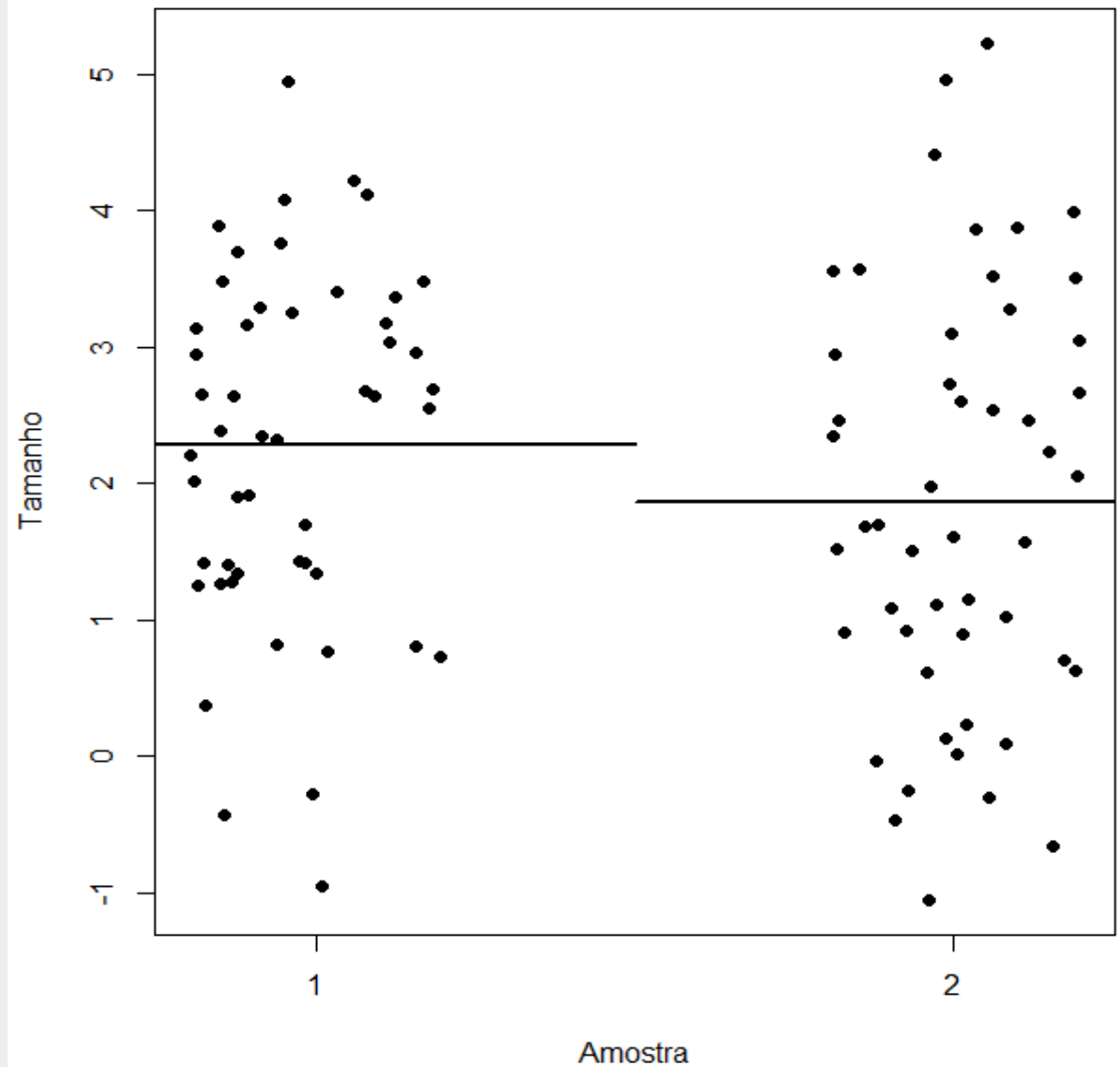
Exemplo numérico

Simulação –
aleatorizando
pontos entre
amostras



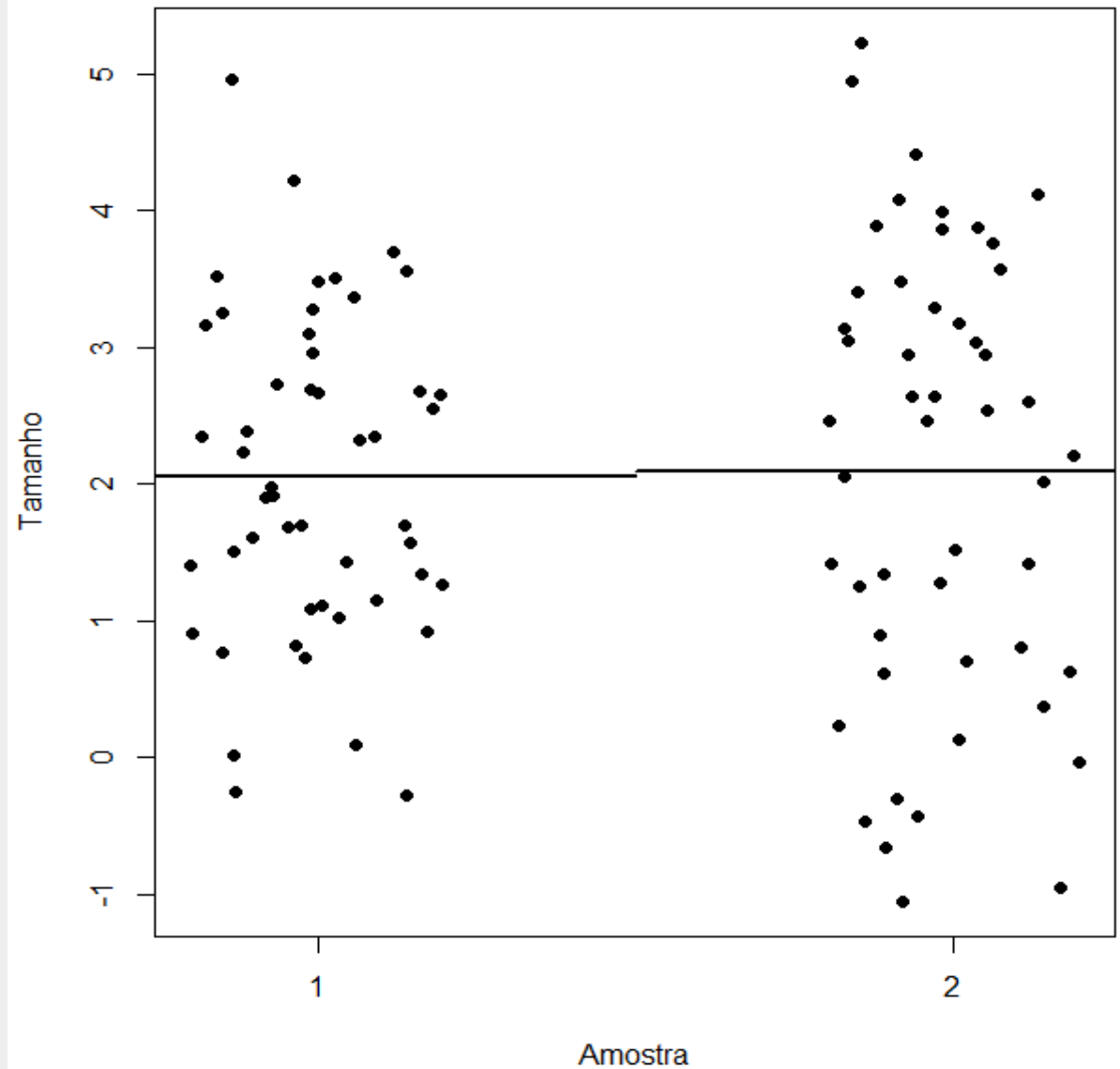
Exemplo numérico

Simulação –
aleatorizando
pontos entre
amostras



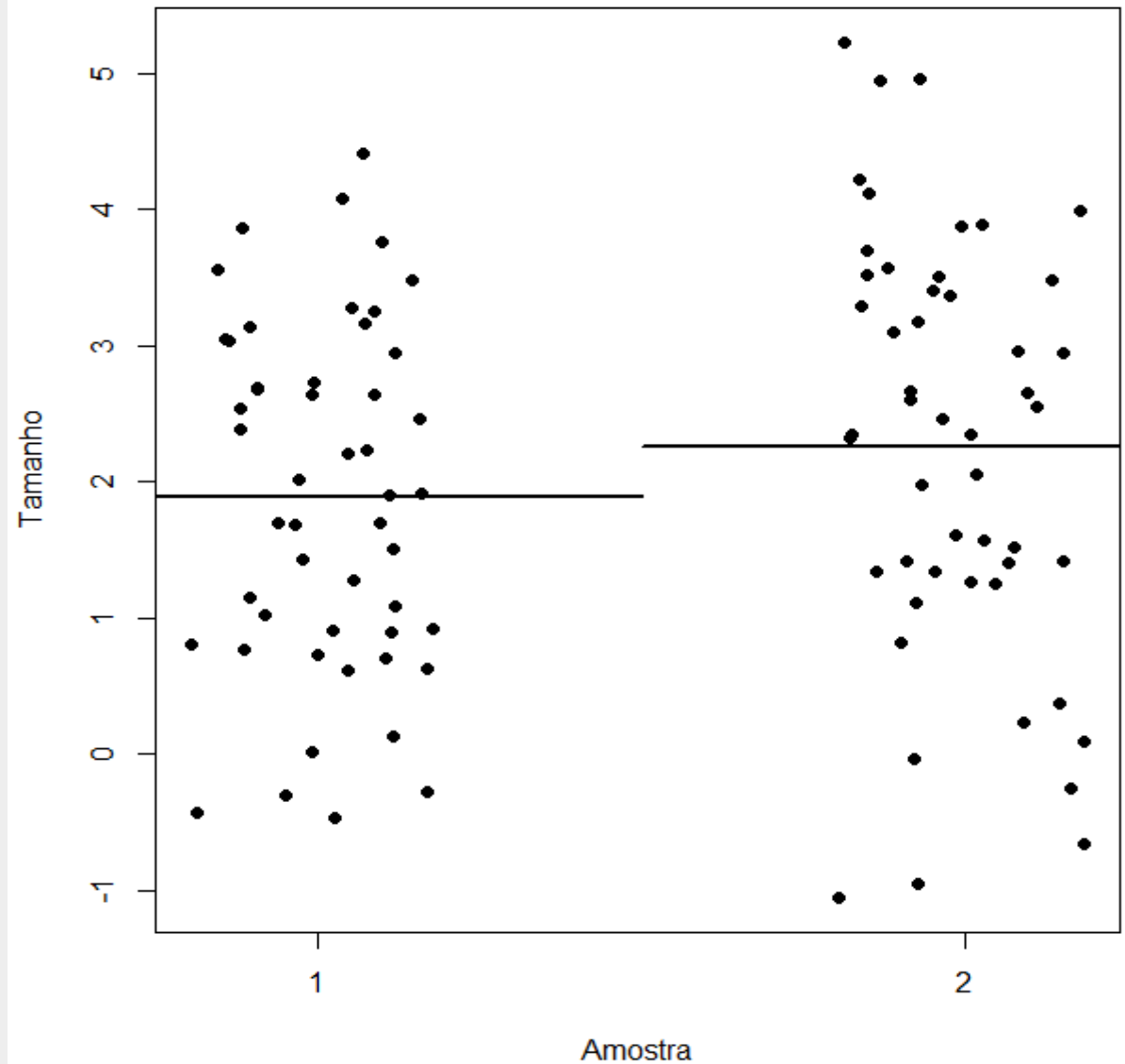
Exemplo numérico

Simulação –
aleatorizando
pontos entre
amostras



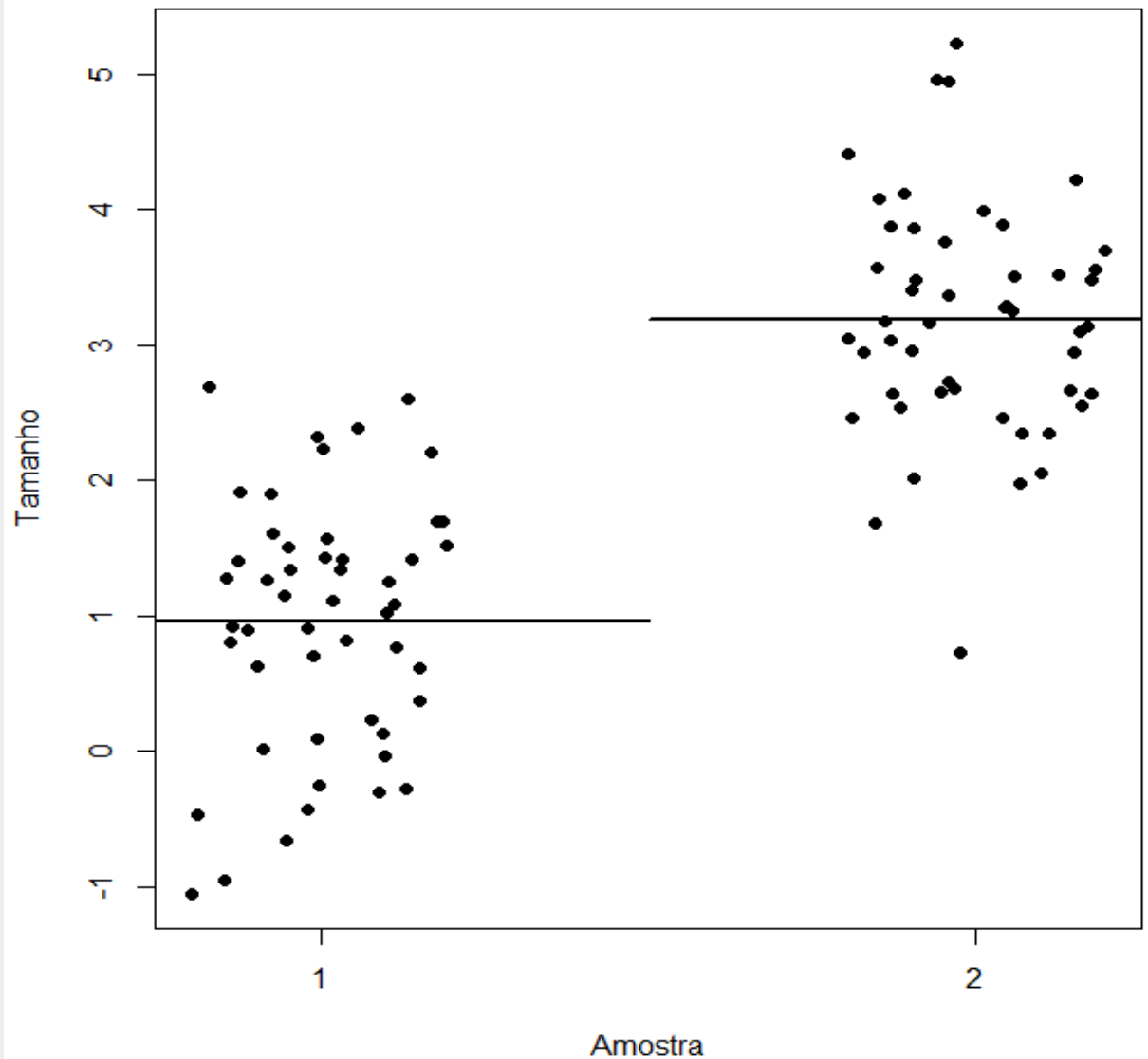
Exemplo numérico

Simulação –
aleatorizando
pontos entre
amostras

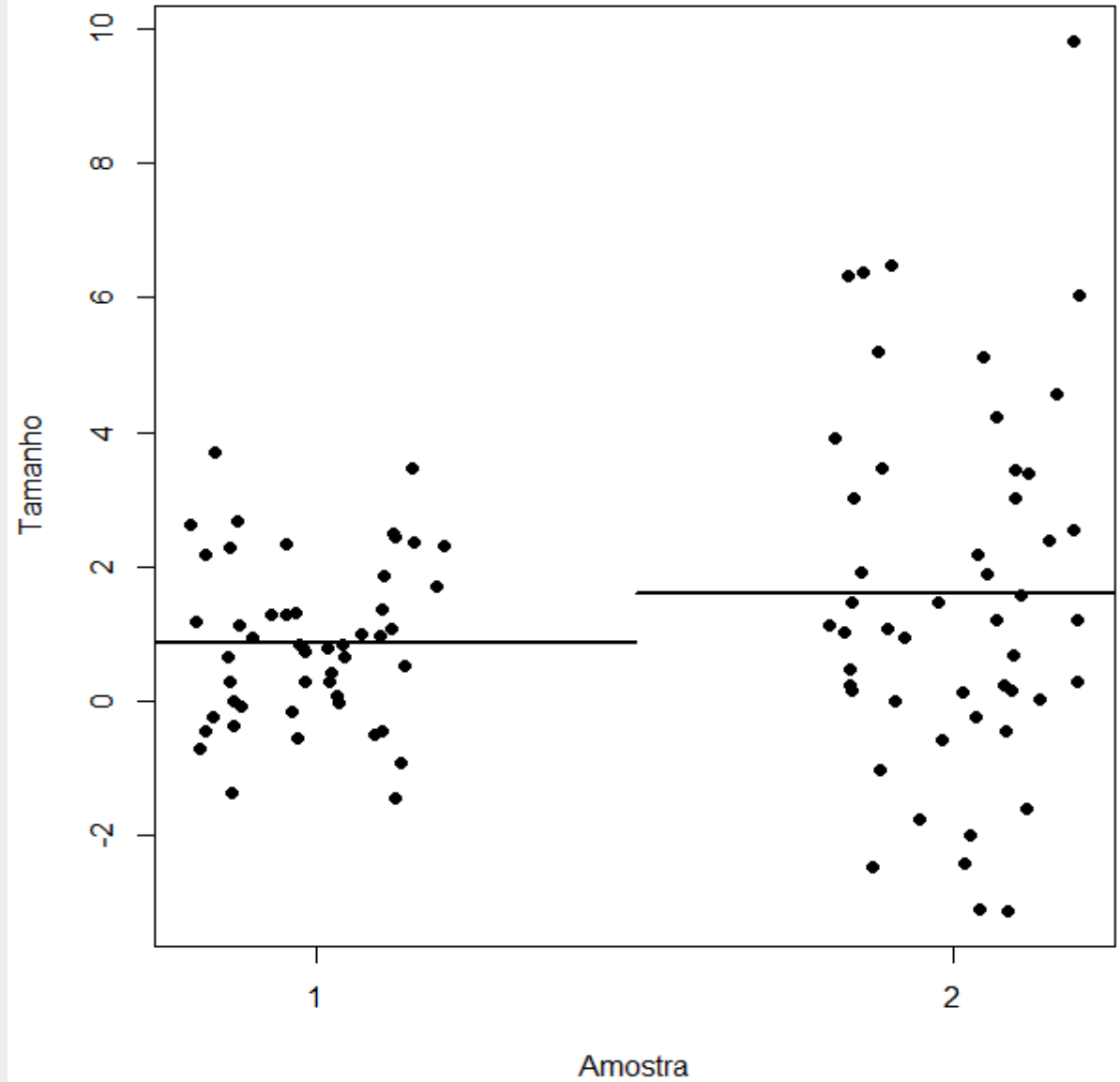


Para testar significância

Comparar a
diferença real
entre as
médias com
as simuladas

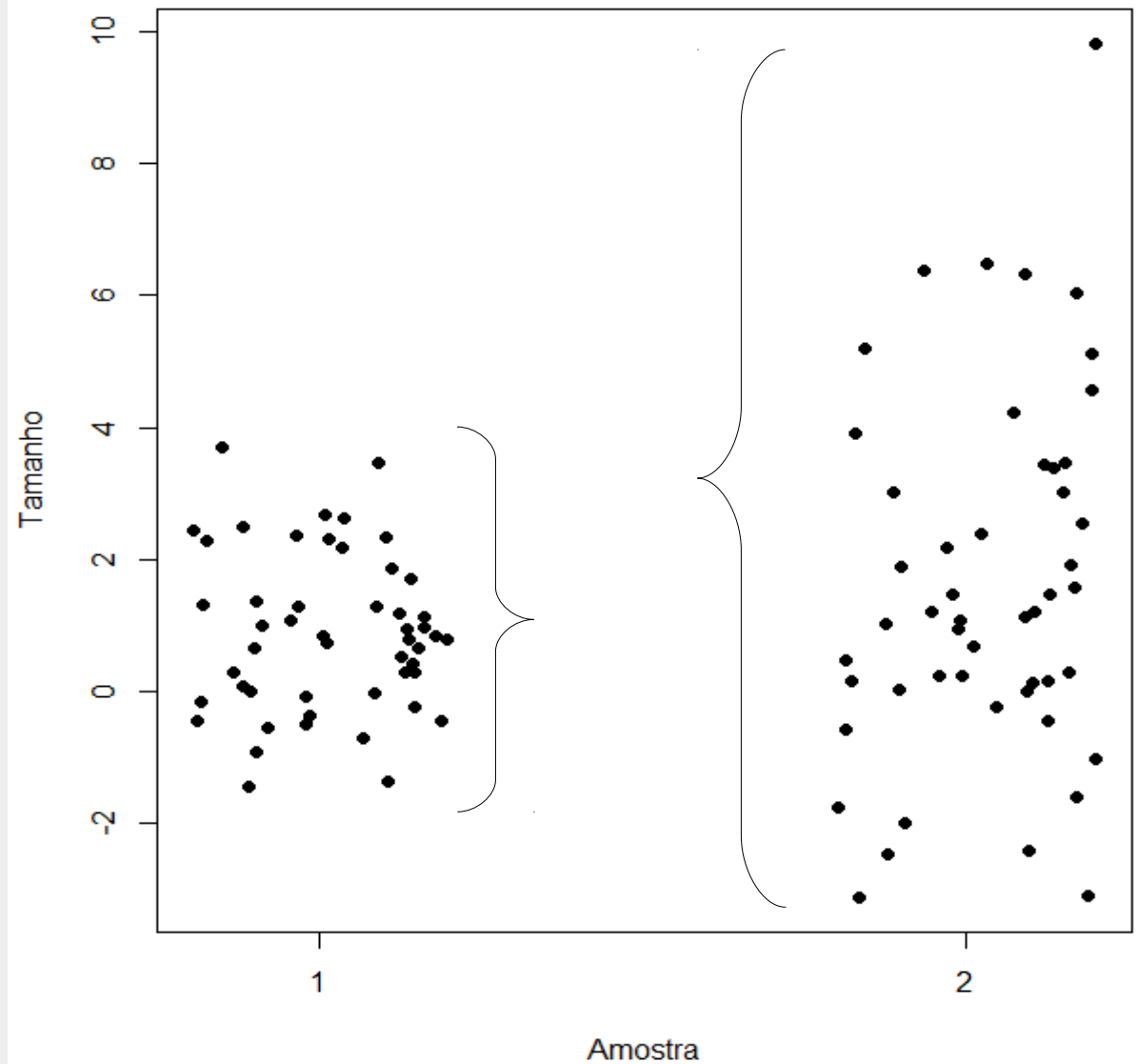


Outro exemplo numérico



Outro exemplo numérico

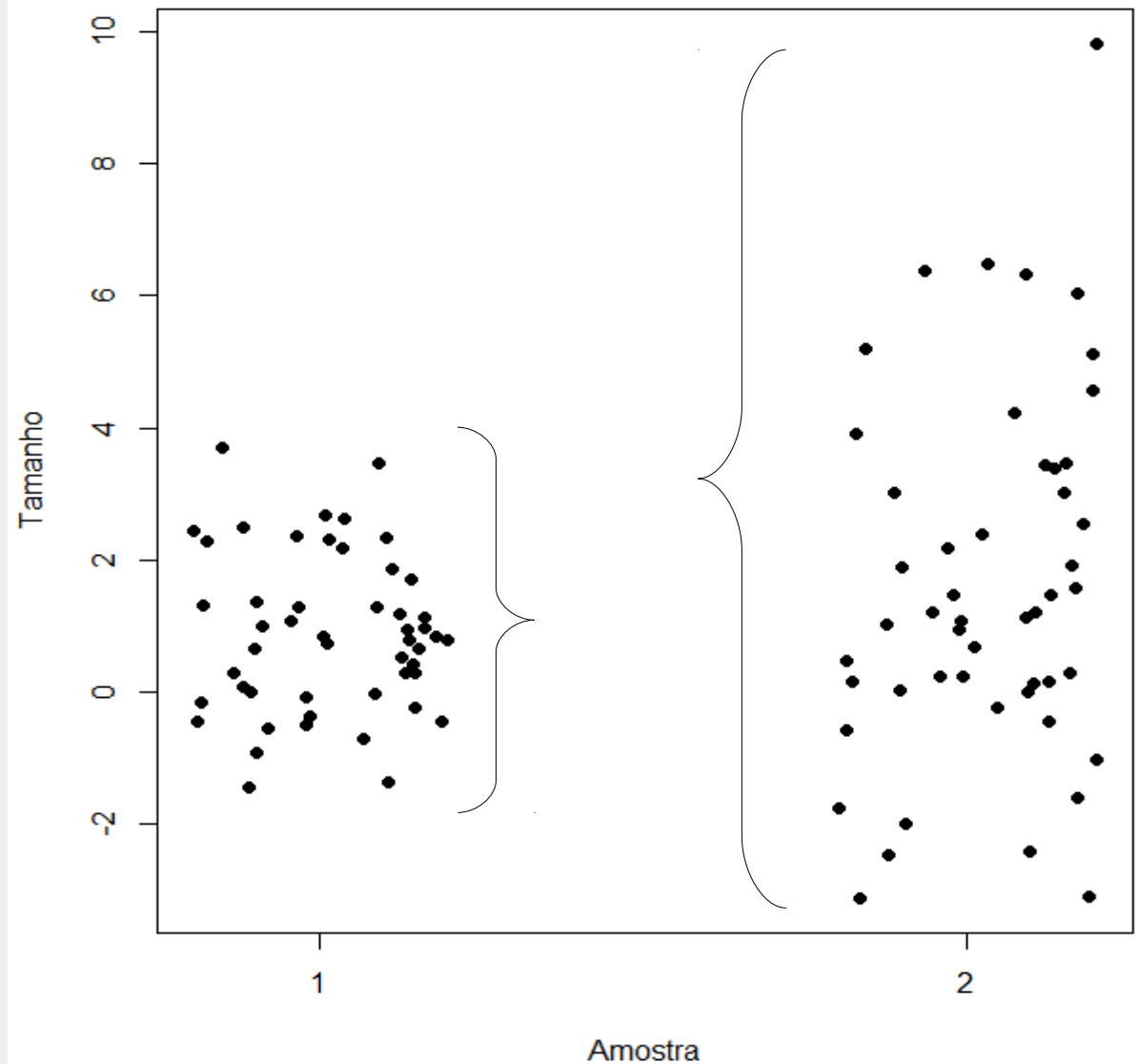
Diferentes
variâncias



Outro exemplo numérico

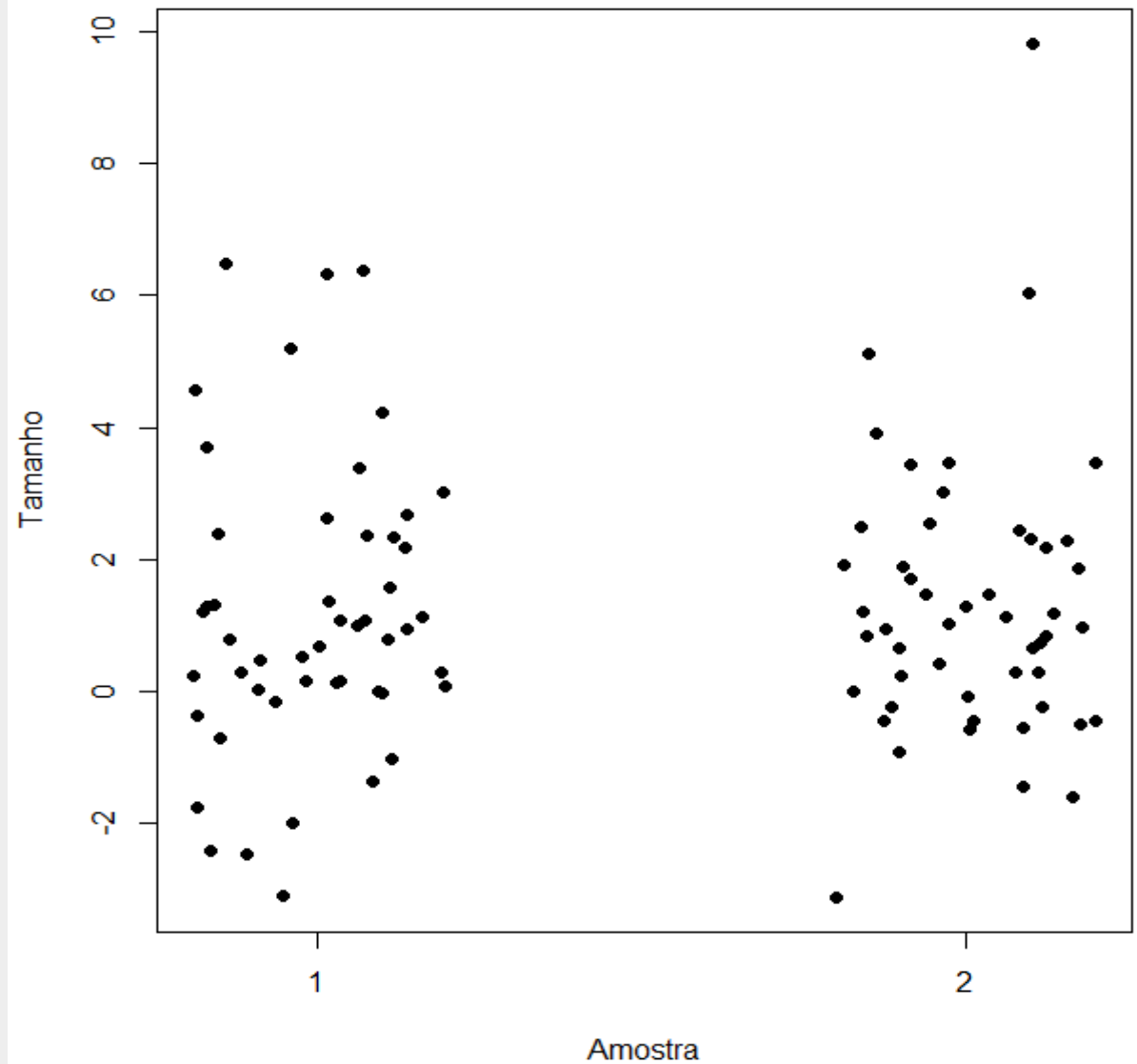
Razão de
variâncias

Variância maior
dividida pela
variância menor



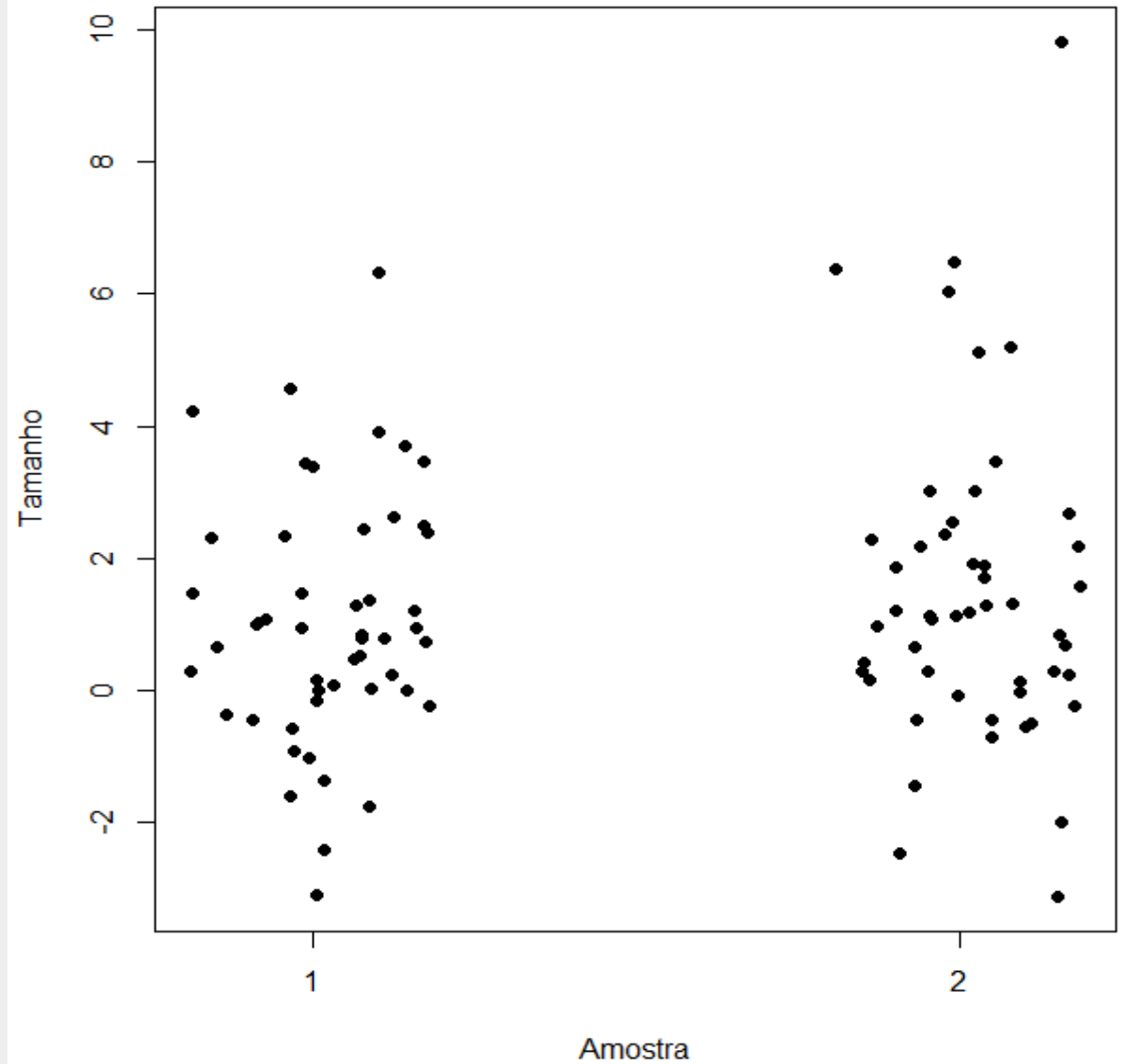
Outro exemplo numérico

Simulando...



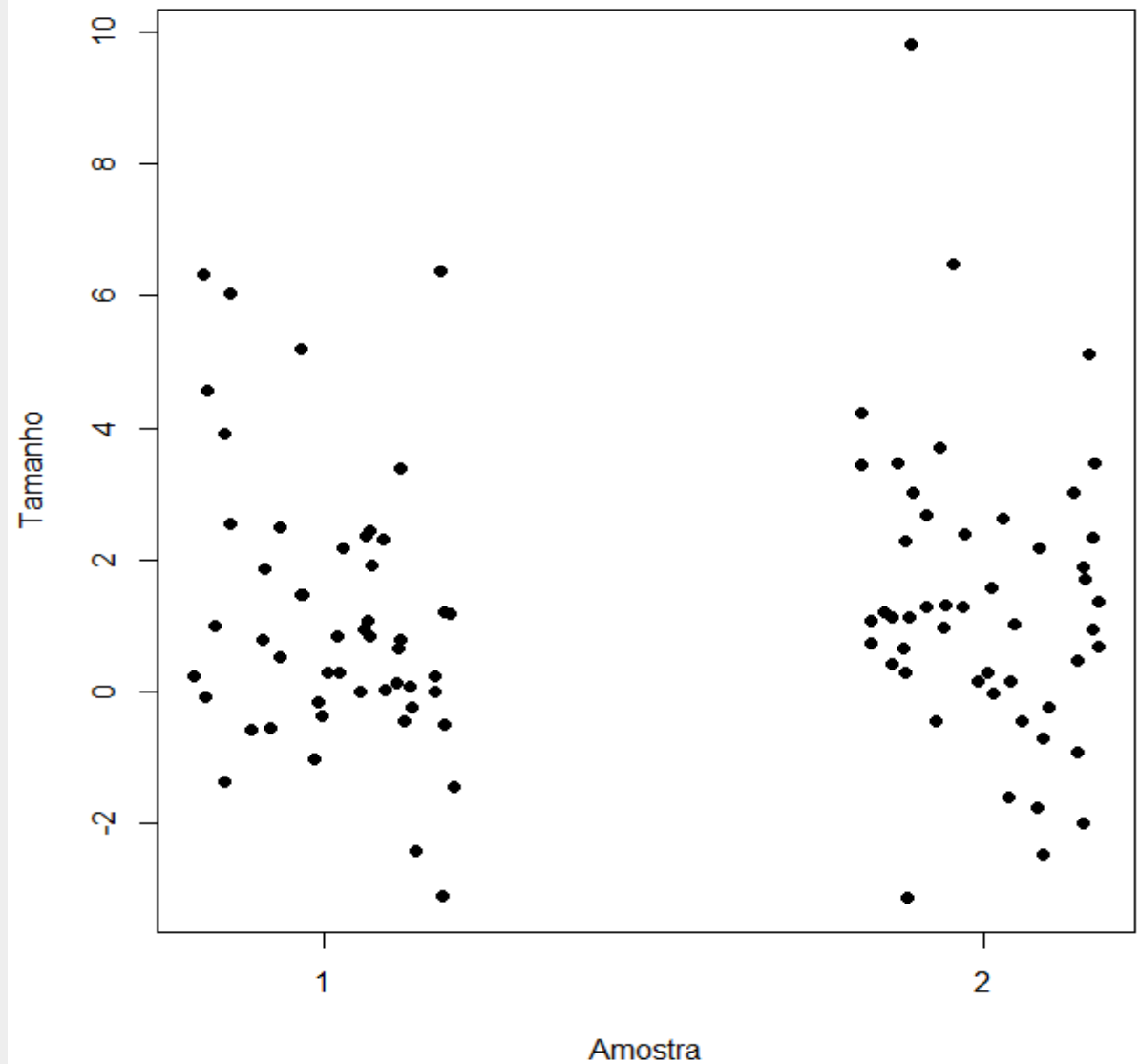
Outro exemplo numérico

Simulando...



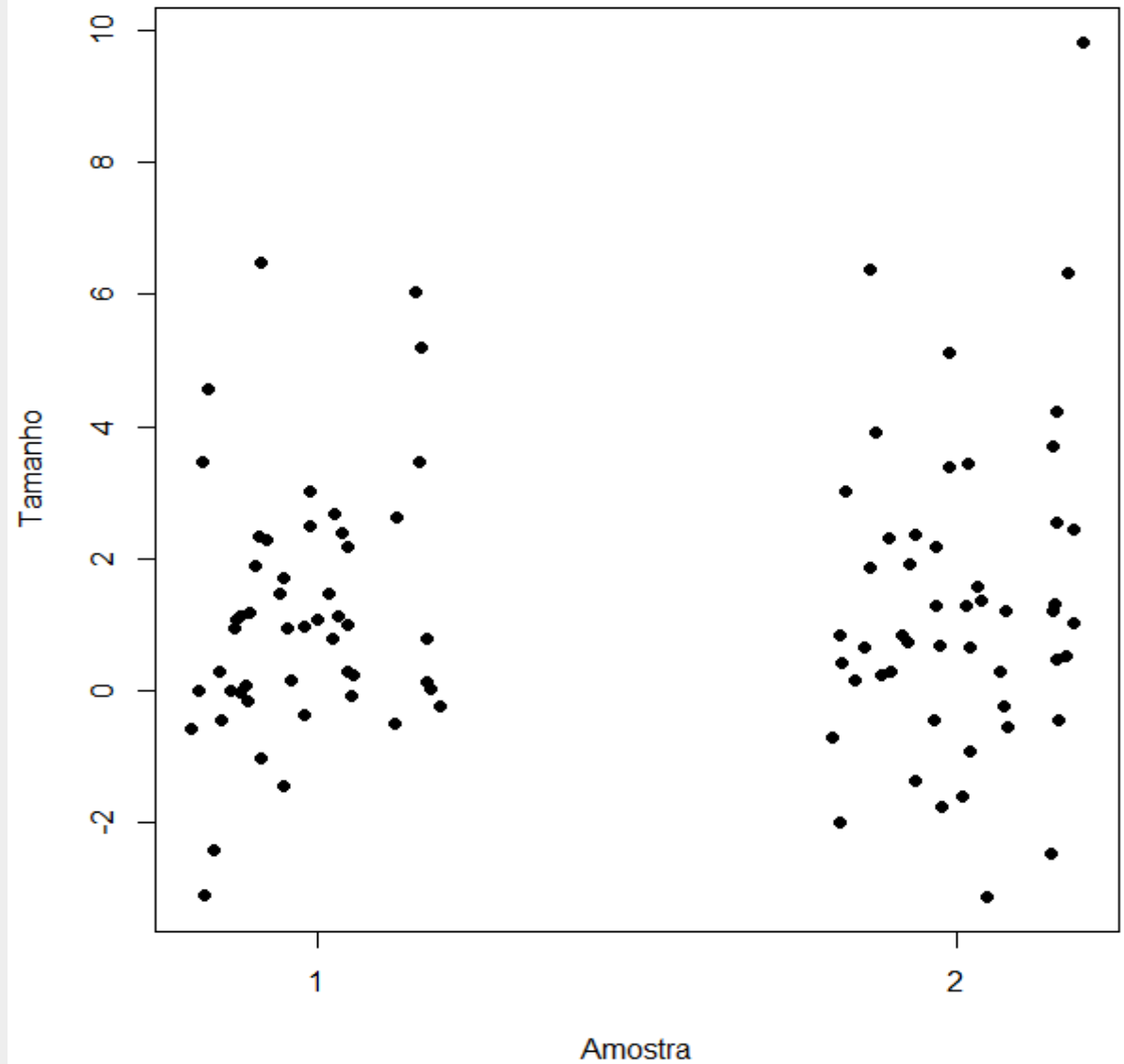
Outro exemplo numérico

Simulando...



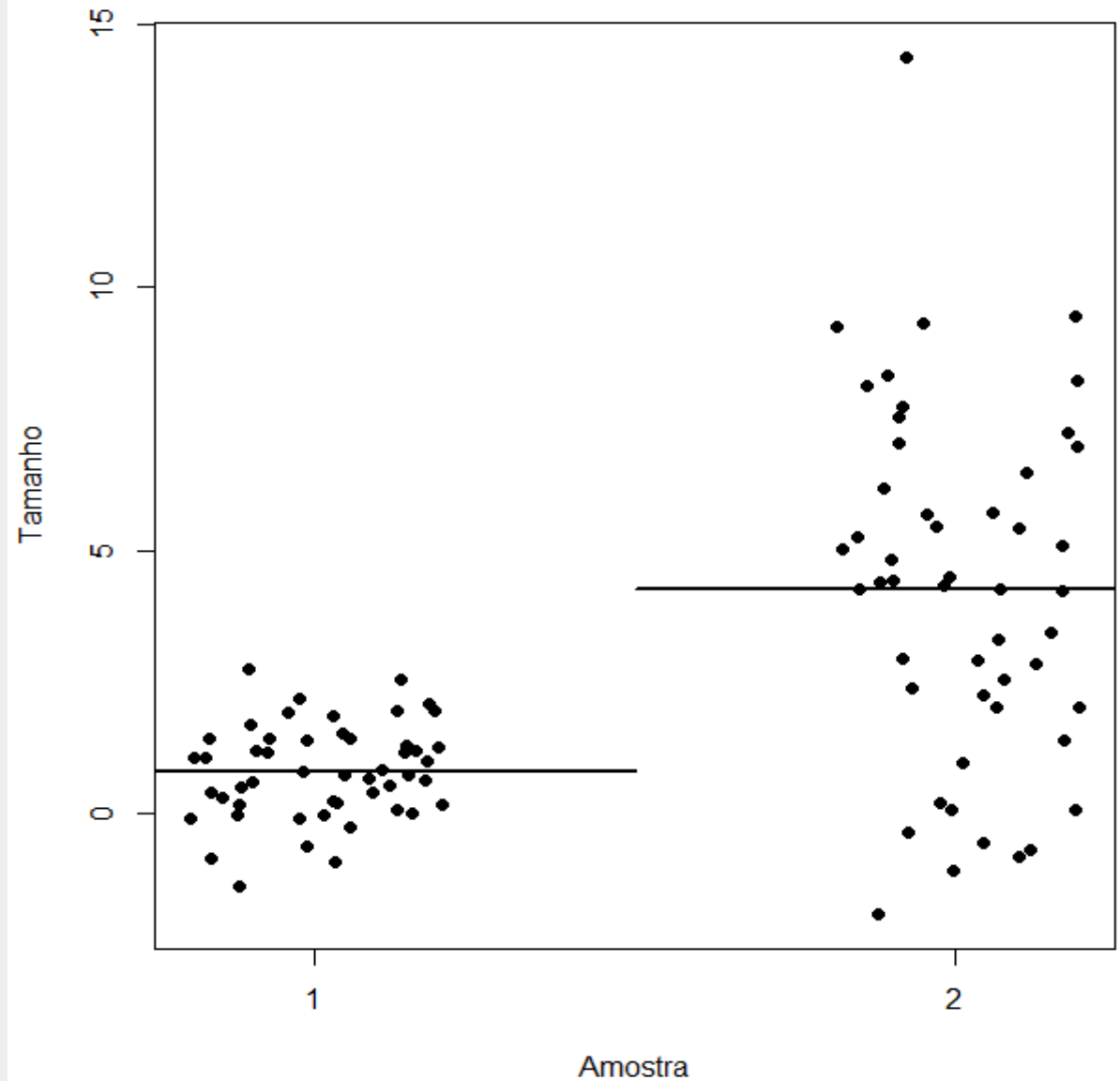
Outro exemplo numérico

Simulando...



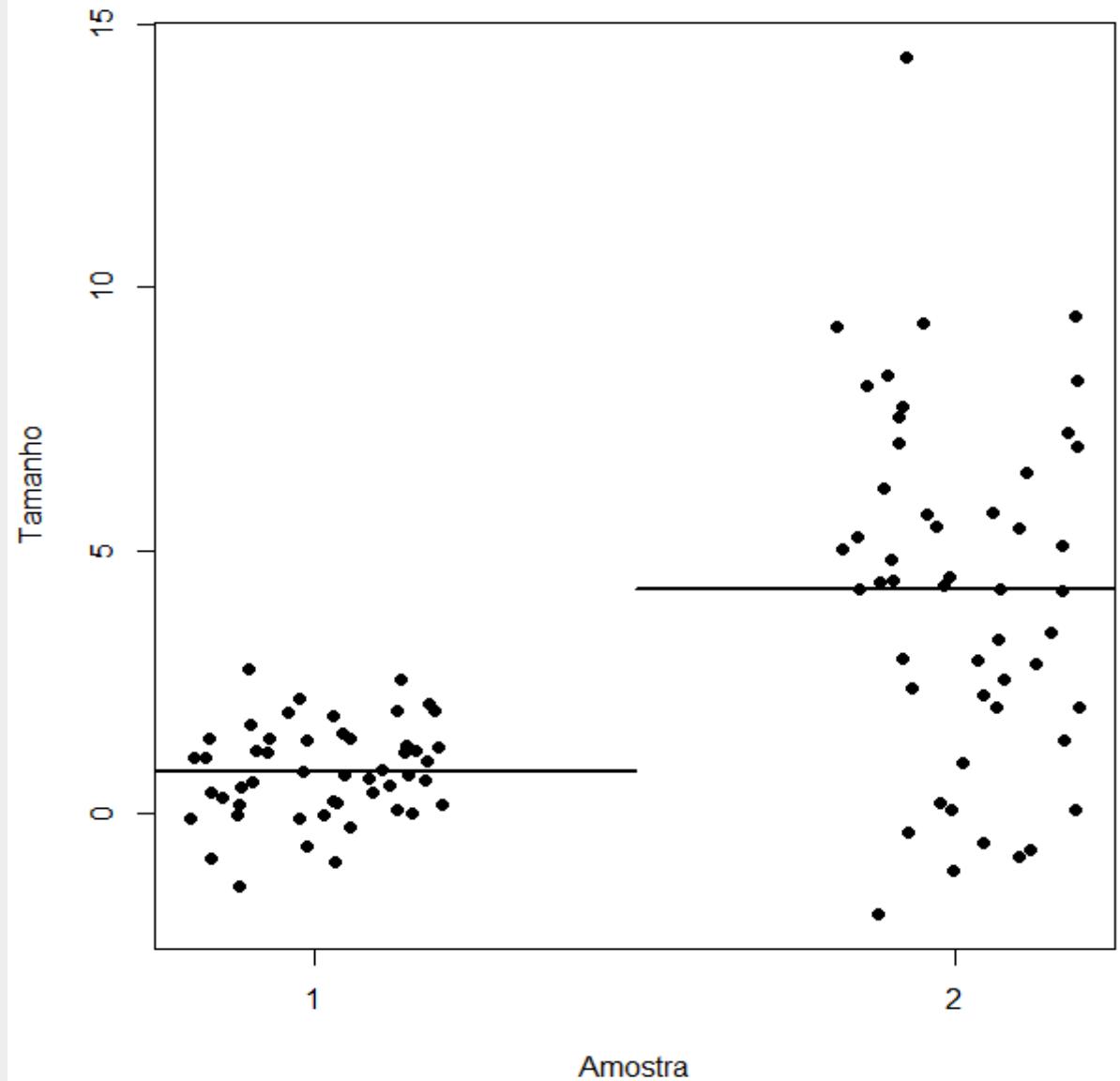
Mais um exemplo numérico

Diferenças nas
médias e nas
variâncias



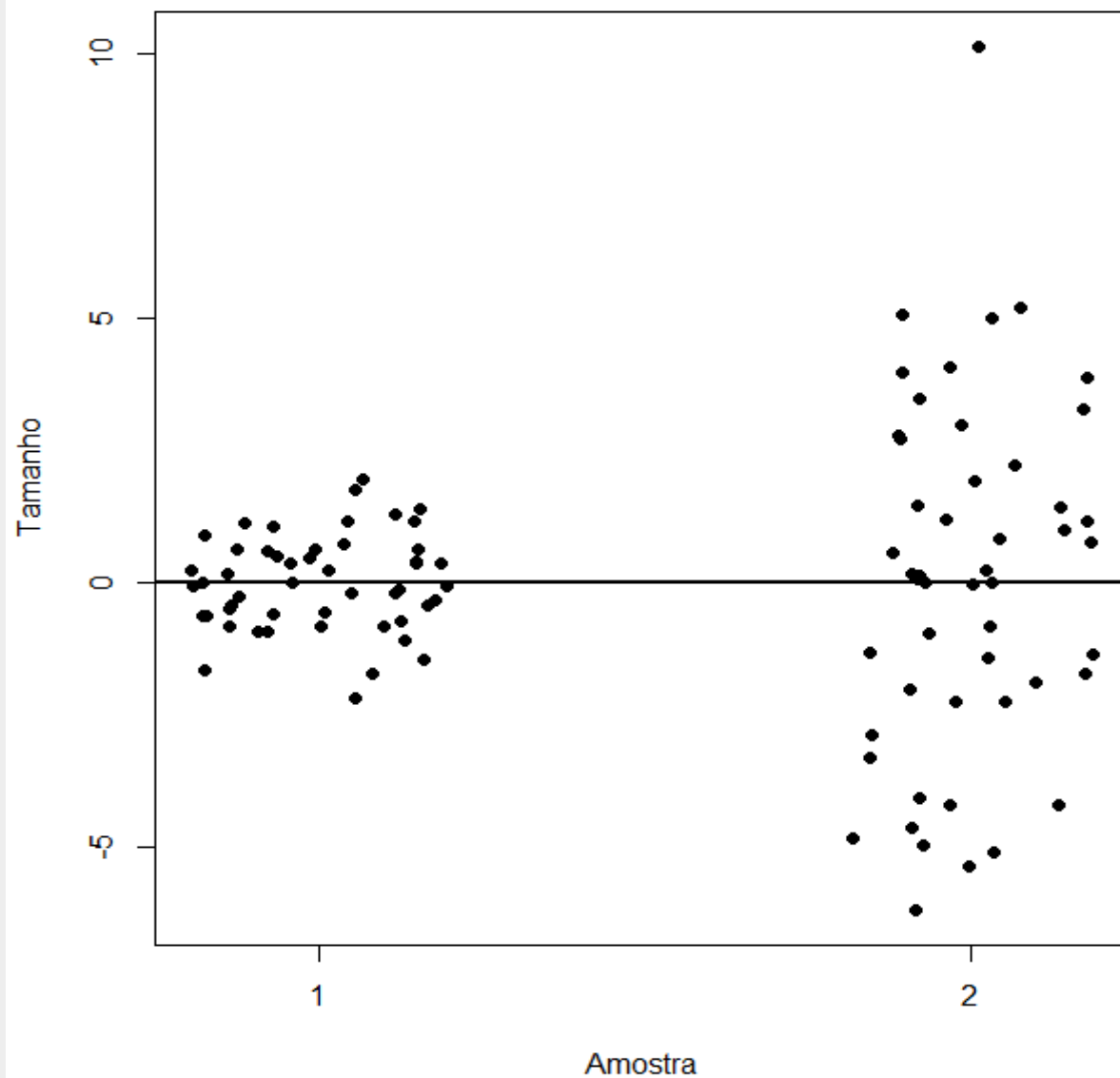
Mais um exemplo numérico

Mais
complicado,
mas não
impossível
de analisar



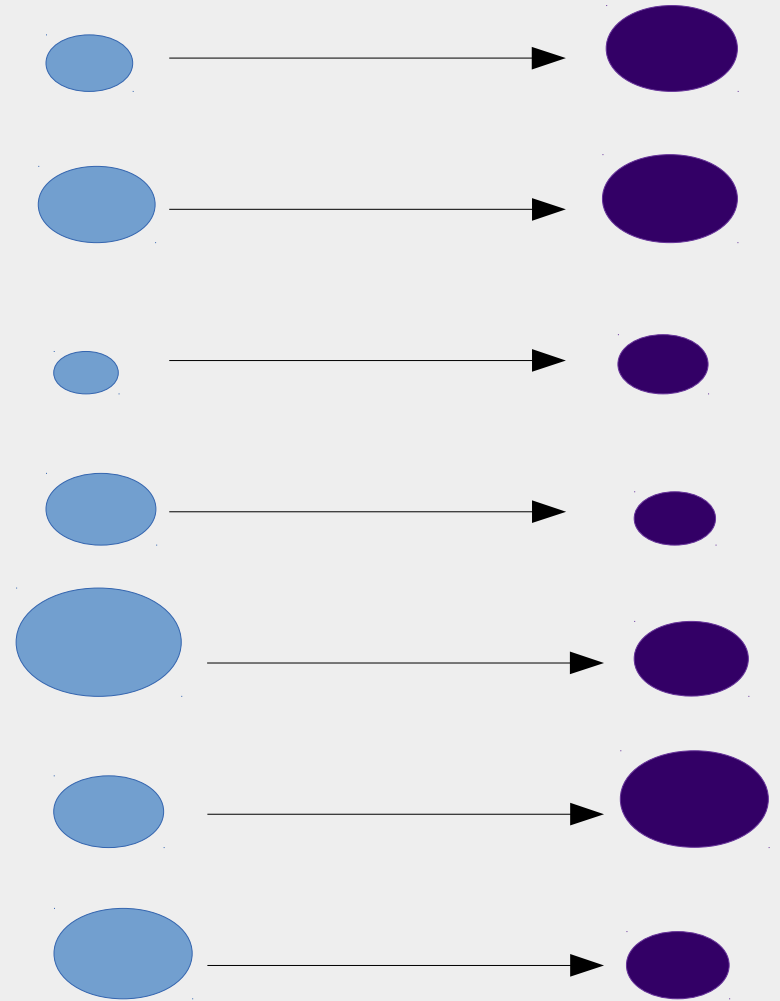
Mais um exemplo numérico

Para
comparar
variância:
usar
resíduos



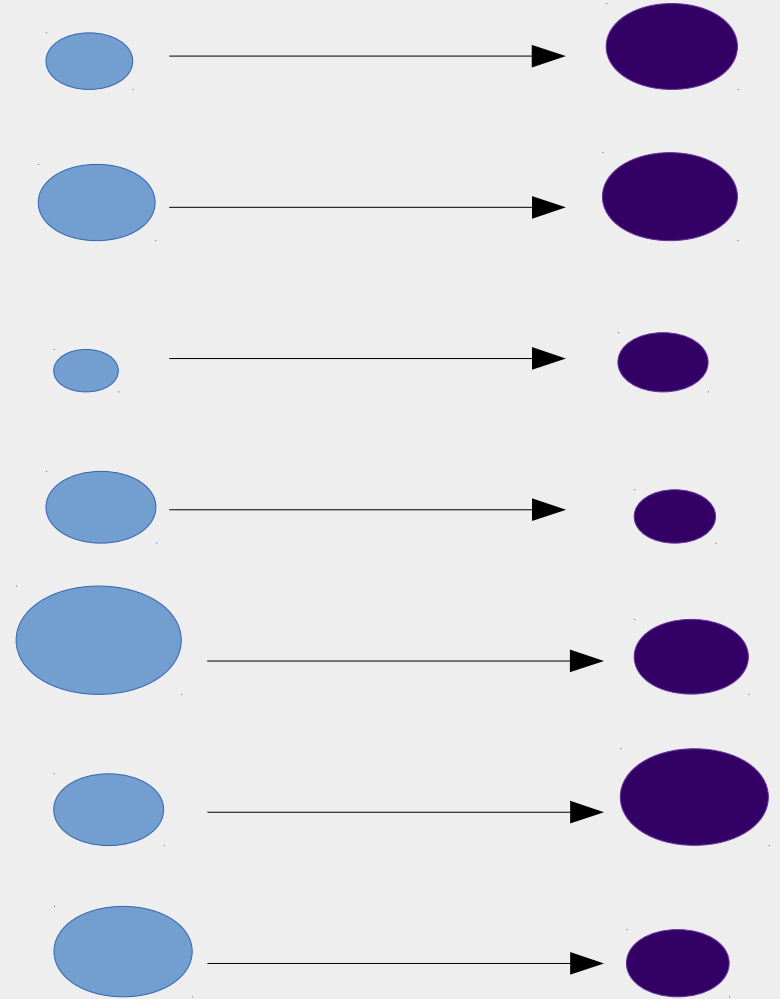
Amostras pareadas

- Valor médio difere?
 - H_0 : não existe efeito do tratamento



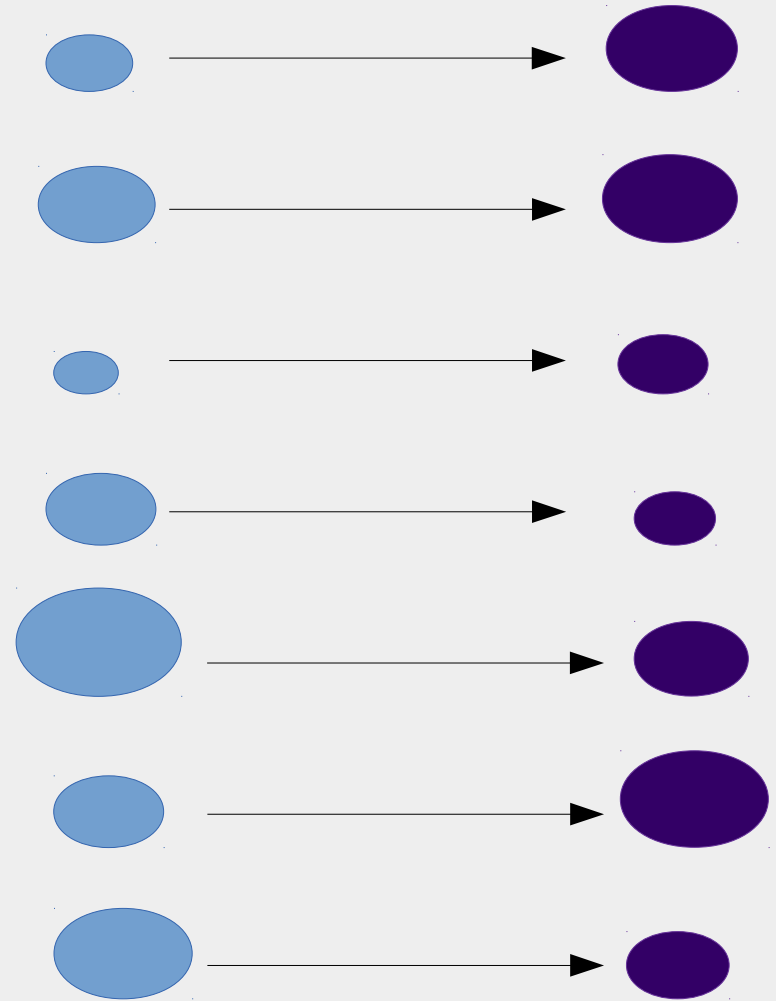
Amostras pareadas

- Valor médio difere?
 - H_0 : não existe efeito do tratamento
 - Diferenças positivas e negativas são igualmente prováveis
 - Trabalhamos com a soma (absoluta) das diferenças



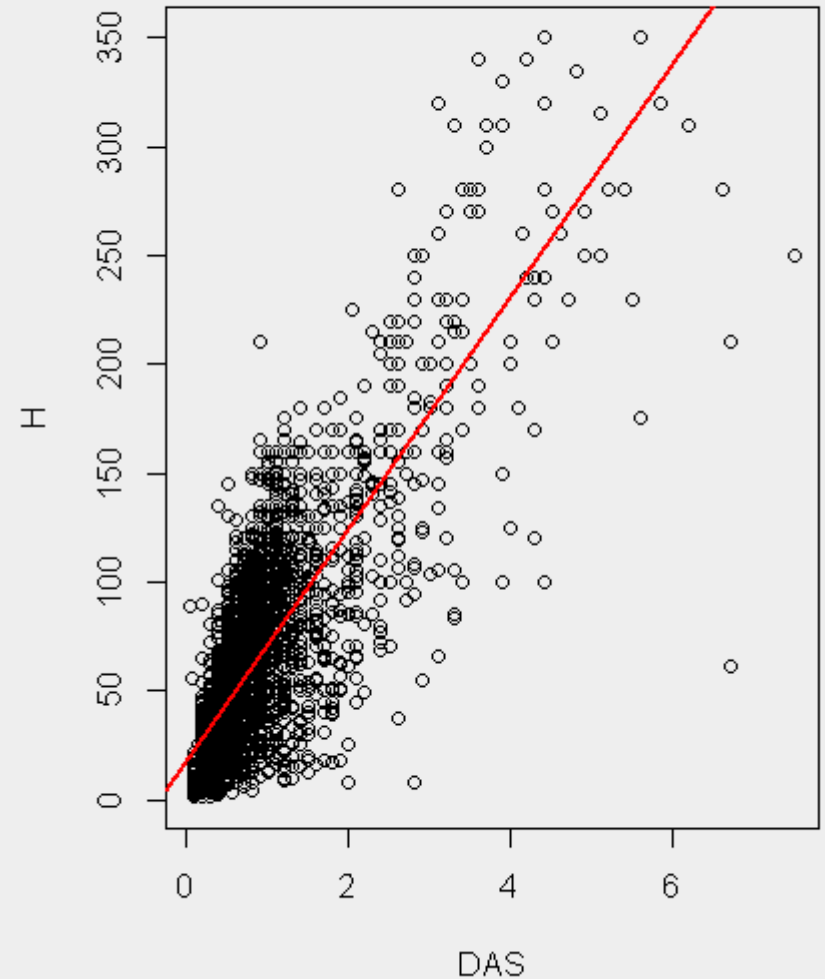
Amostras pareadas

- Valor médio difere?
 - H_0 : não existe efeito do tratamento
 - Diferenças positivas e negativas são igualmente prováveis
 - Trabalhamos com a soma (absoluta) das diferenças
 - Alocar cada amostra aleatoriamente a um tratamento



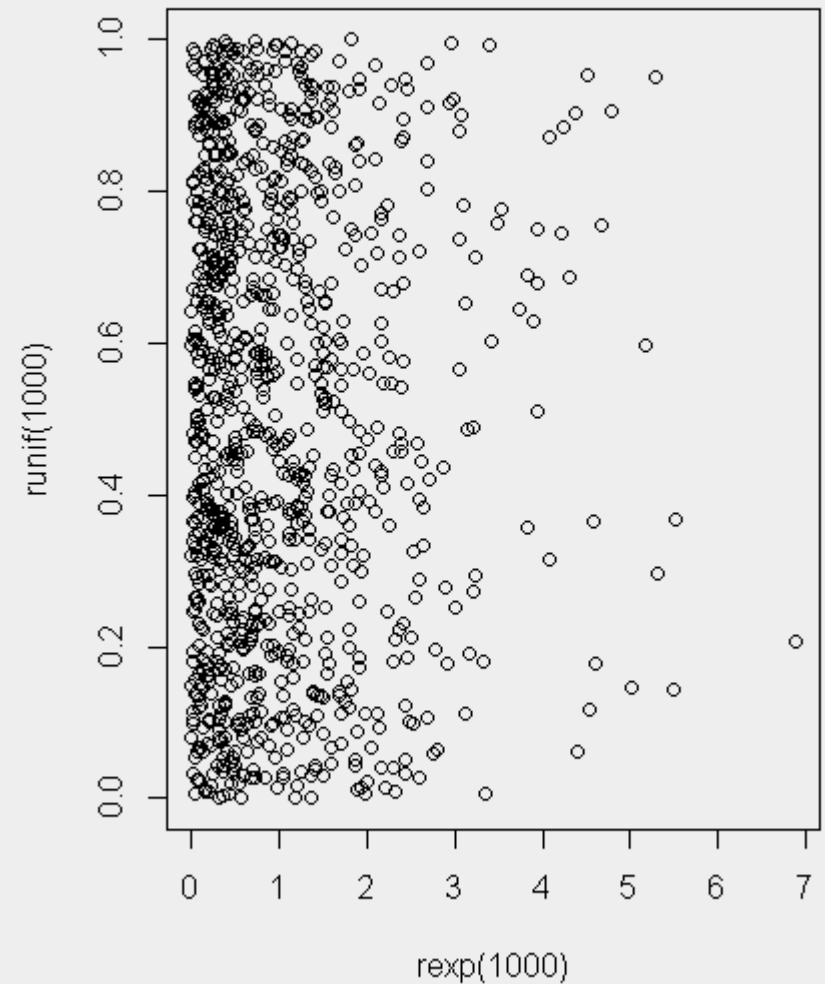
Regressão linear

- Relação entre duas variáveis
 - Coeficiente de correlação e inclinação da regressão são estatísticas equivalentes para permutações



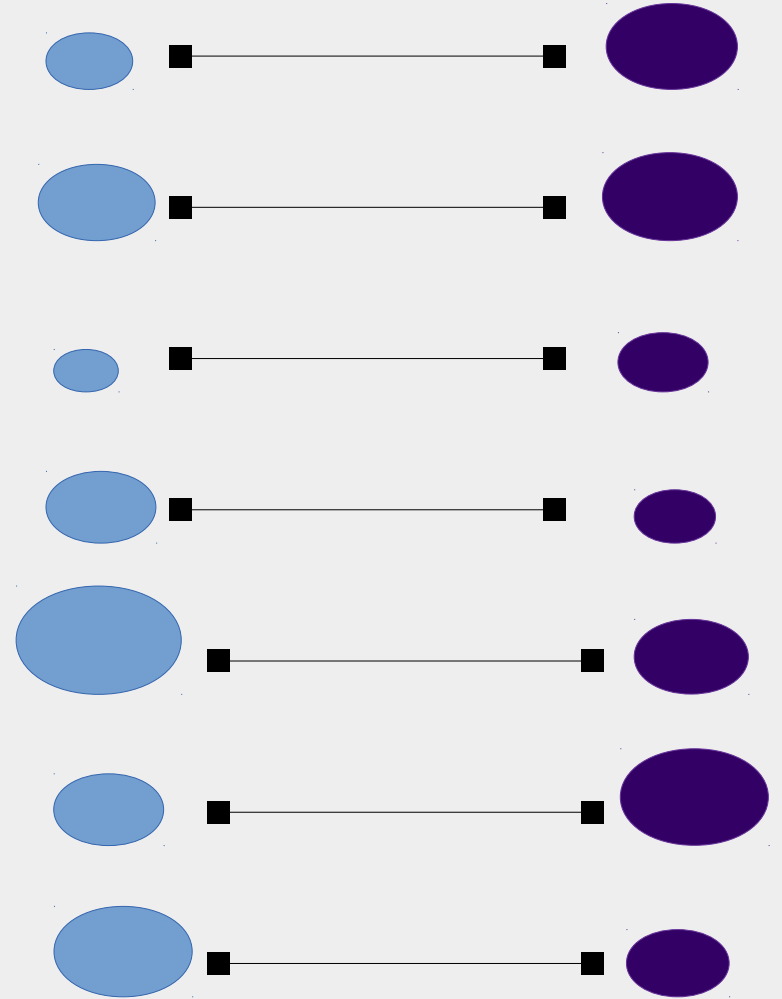
Regressão linear

- Hipótese nula
 - Não existe relação entre as duas variáveis



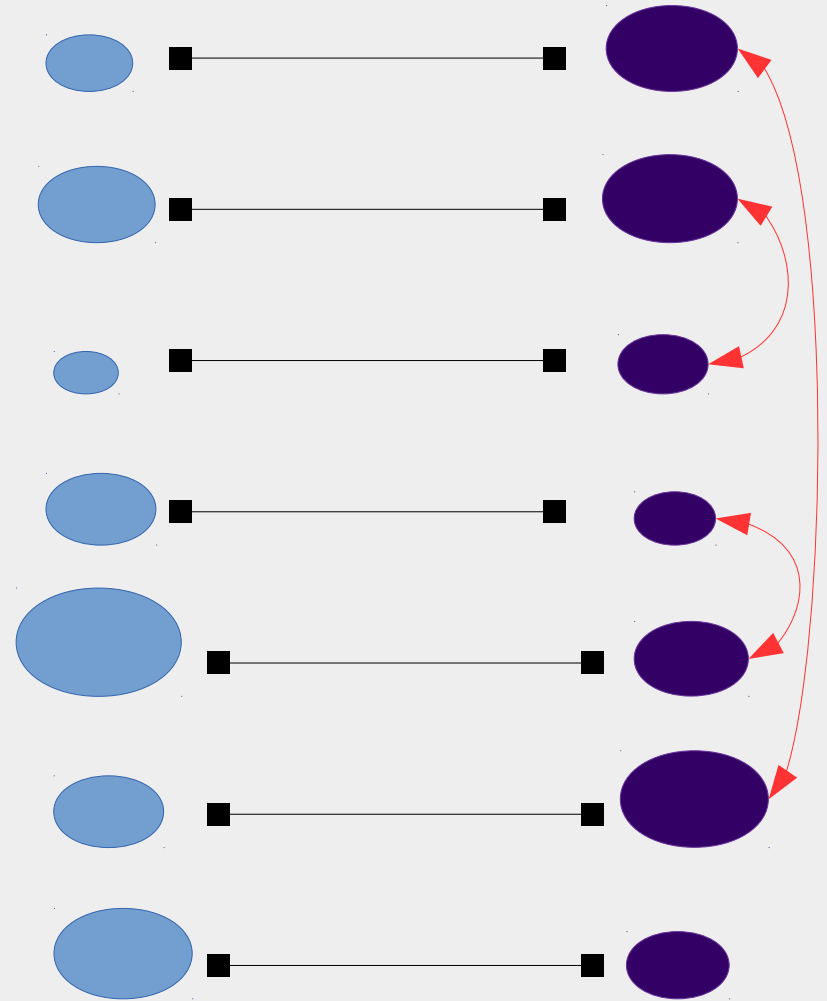
Regressão linear

- Como aleatorizar?



Regressão linear

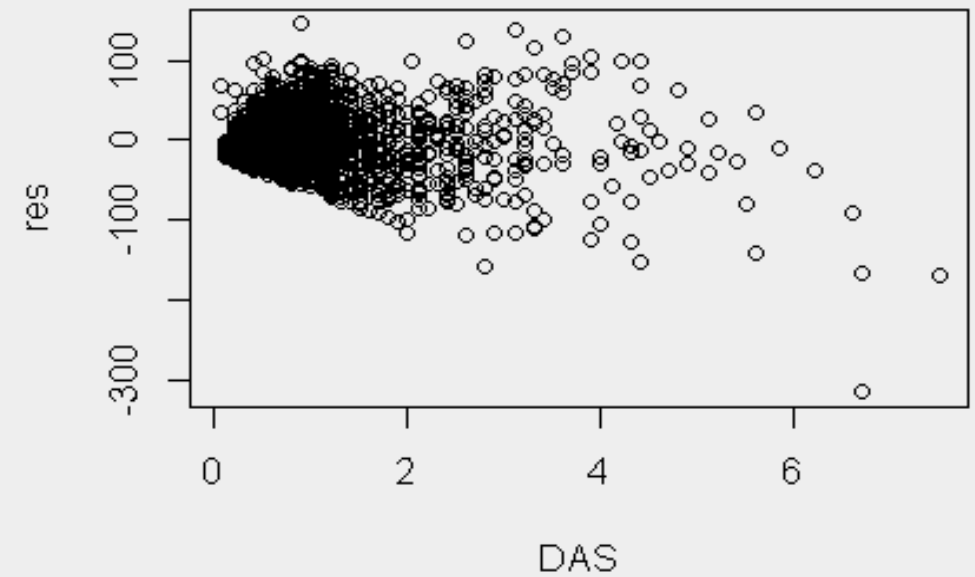
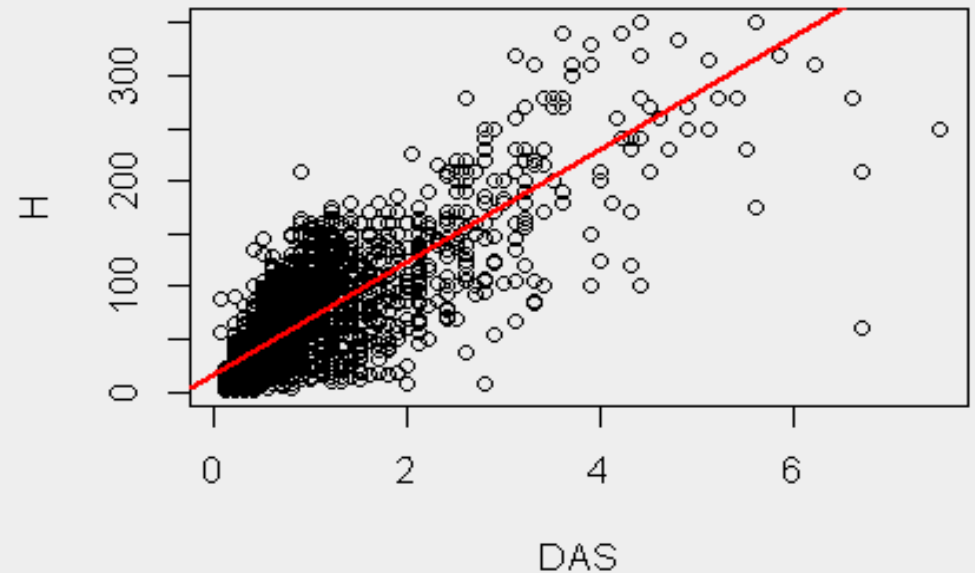
- Aleatorizando valores
 - Associamos aleatoriamente valores de X e de Y



Regressão linear

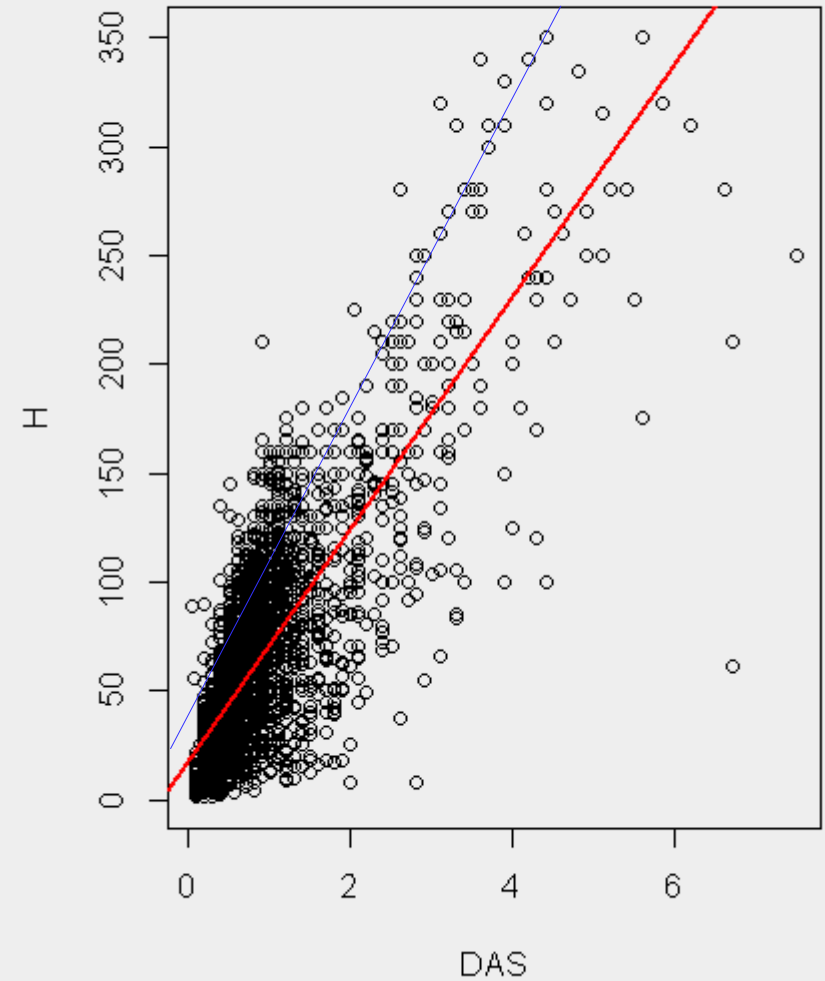
- Aleatorizando resíduos
 - Resíduos e valores de X não são correlacionados
 - Aproximaria a distribuição de resíduos na hipótese nula

$$t = \frac{b}{SE(b)}$$



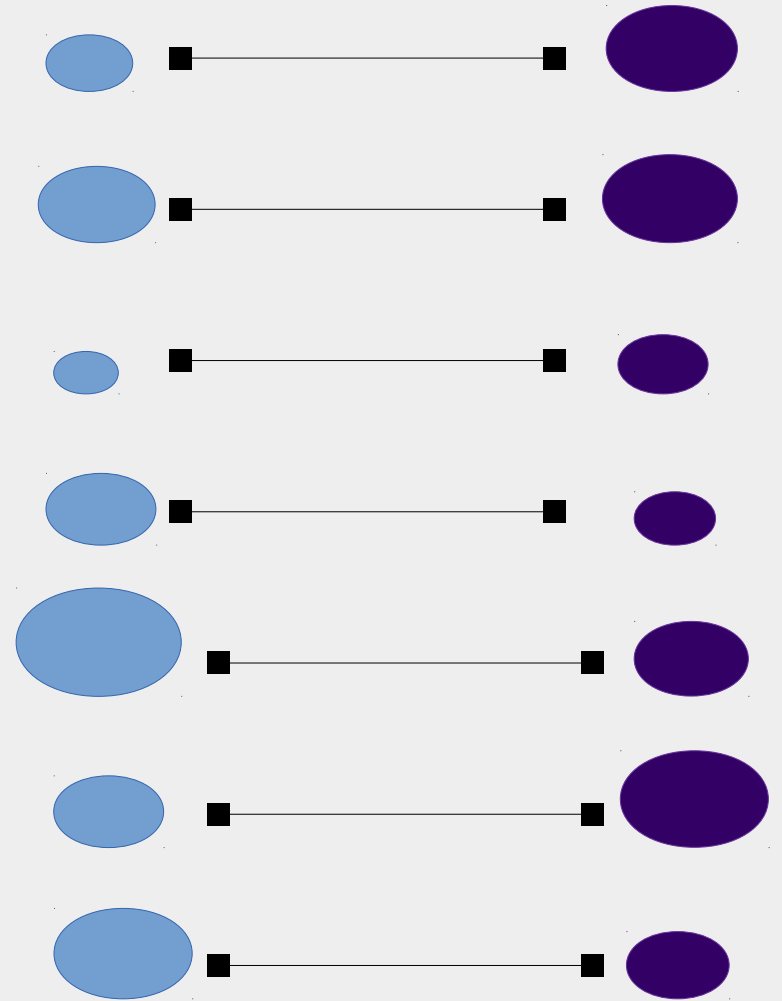
Duas regressões lineares

- Relação entre duas variáveis em dois grupos



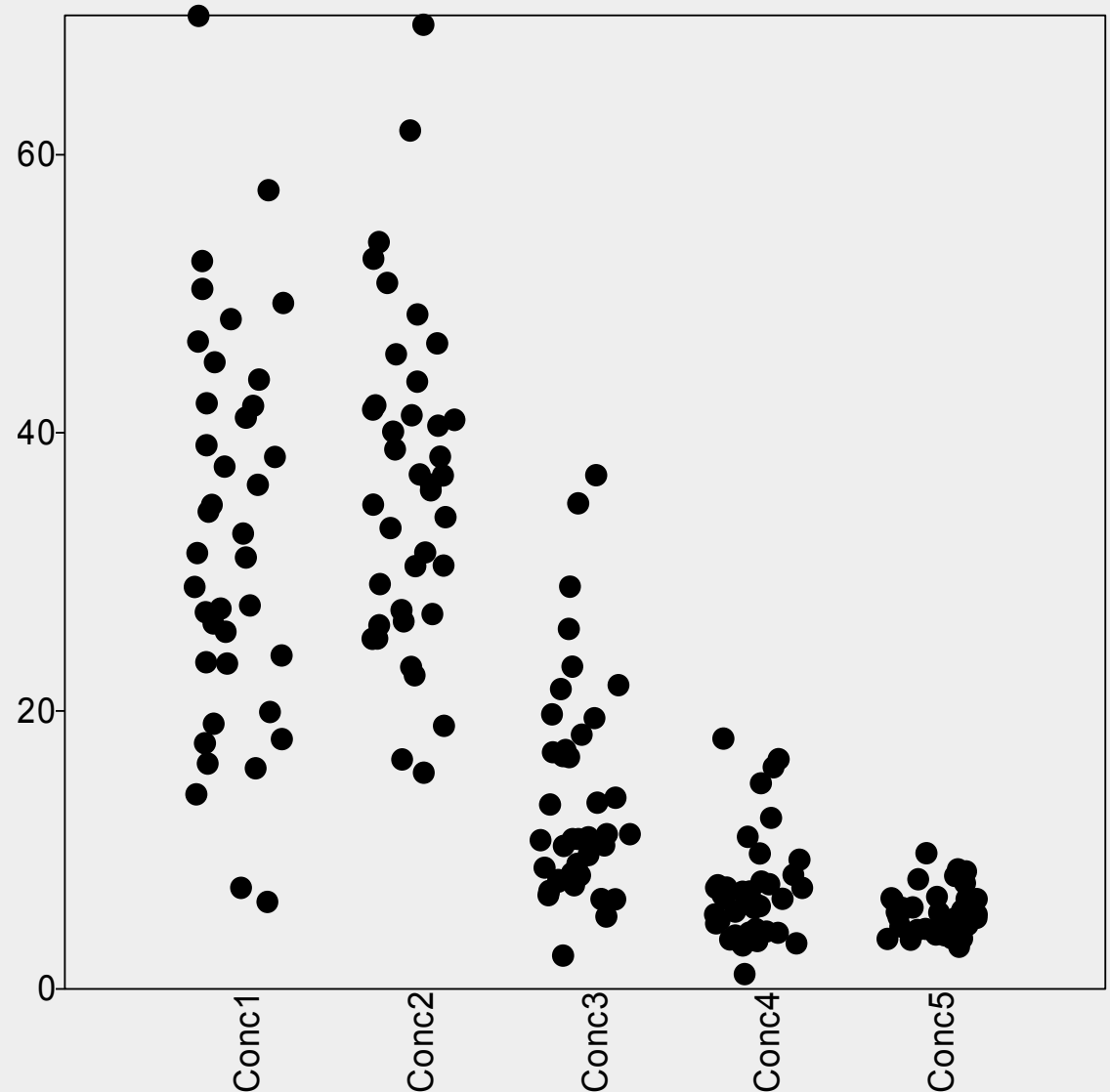
Duas regressões

- Como aleatorizar?



Comparando três grupos ou mais

- ANOVA
 - Testes t repetidos: problema das comparações múltiplas
 - Compara a variação que existe entre grupos e dentro de grupos



Análise de Variância

- Variação entre grupos
 - Efeito dos tratamentos
 - Quão longe a média de cada grupo está da média global

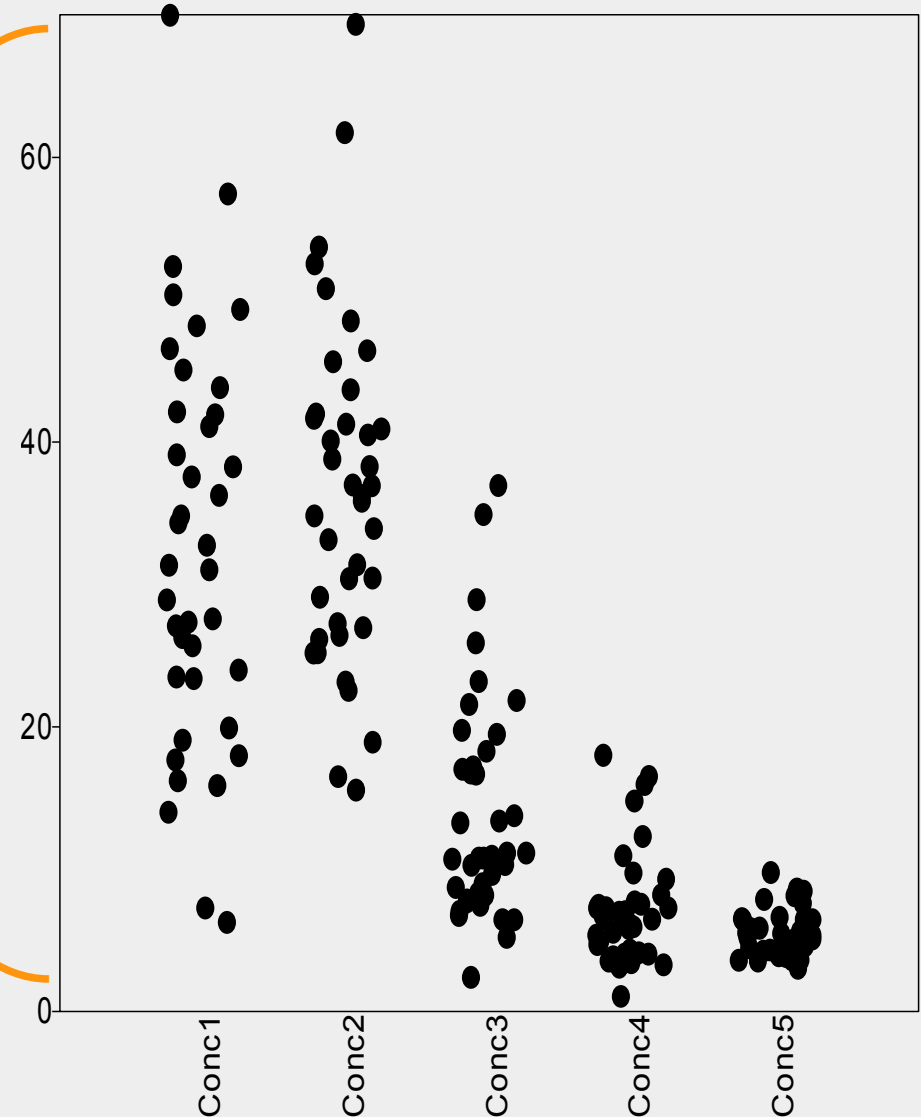
$$SS_{entre} = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X})^2$$

$$DF_{entre} = k - 1$$

$$MS_{entre} = \frac{SS_{entre}}{DF_{entre}}$$

k grupos

Variância



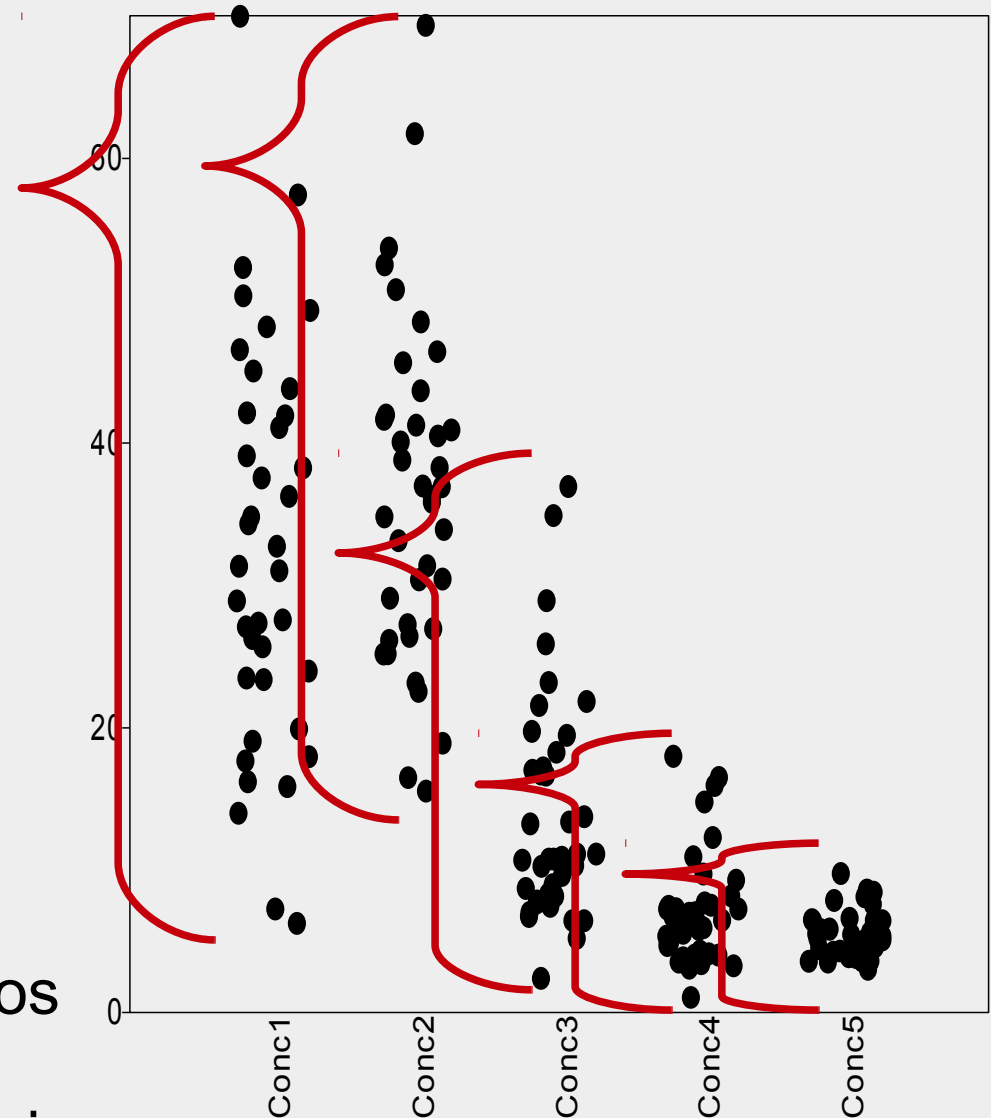
Análise de Variância

- Variação intra grupos
 - Variação residual
 - O que o tratamento não explica
 - Distância entre cada ponto e a média do grupo

$$SS_{intra} = \sum_{i=1}^k \left[\sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 \right]$$
$$DF_{intra} = N - k$$
$$MS_{intra} = \frac{SS_{intra}}{DF_{intra}}$$

k grupos

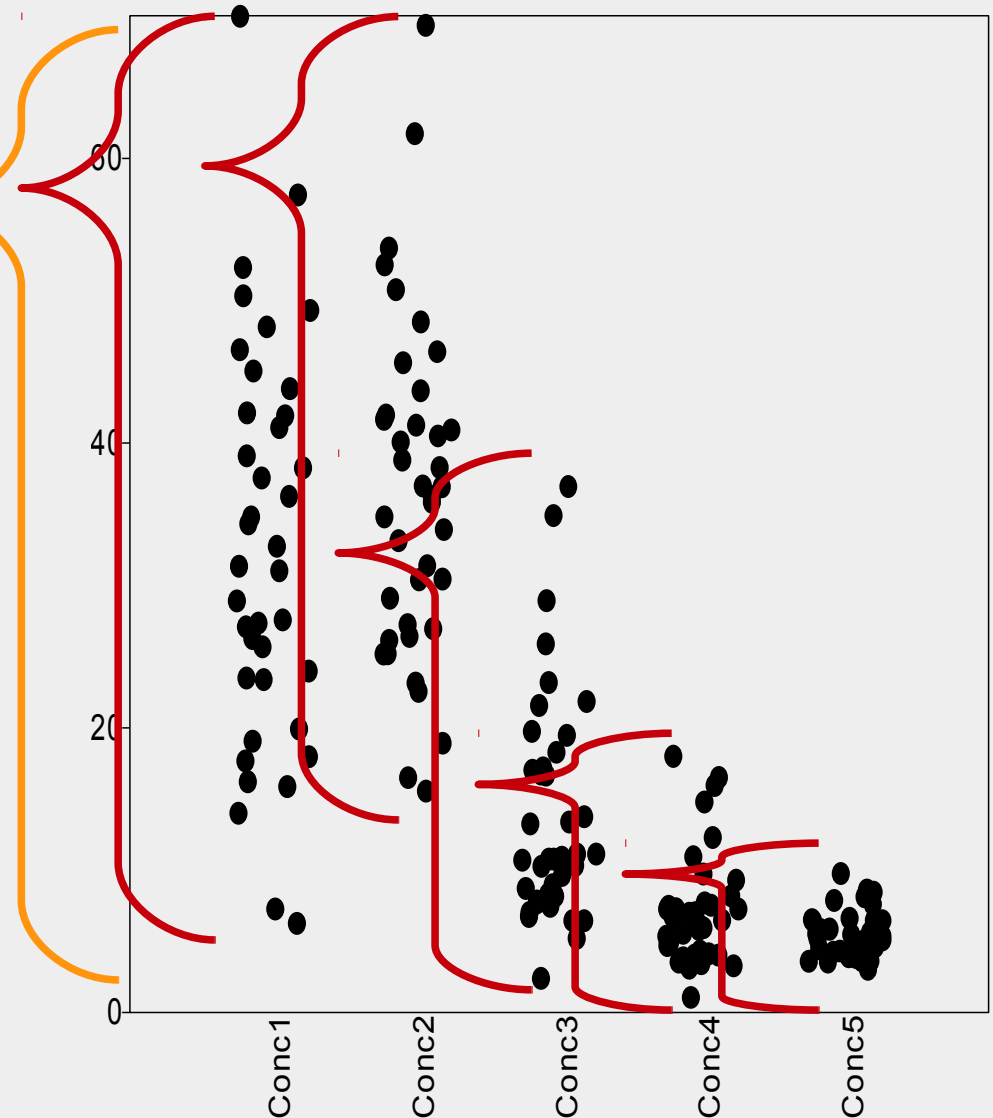
Variância



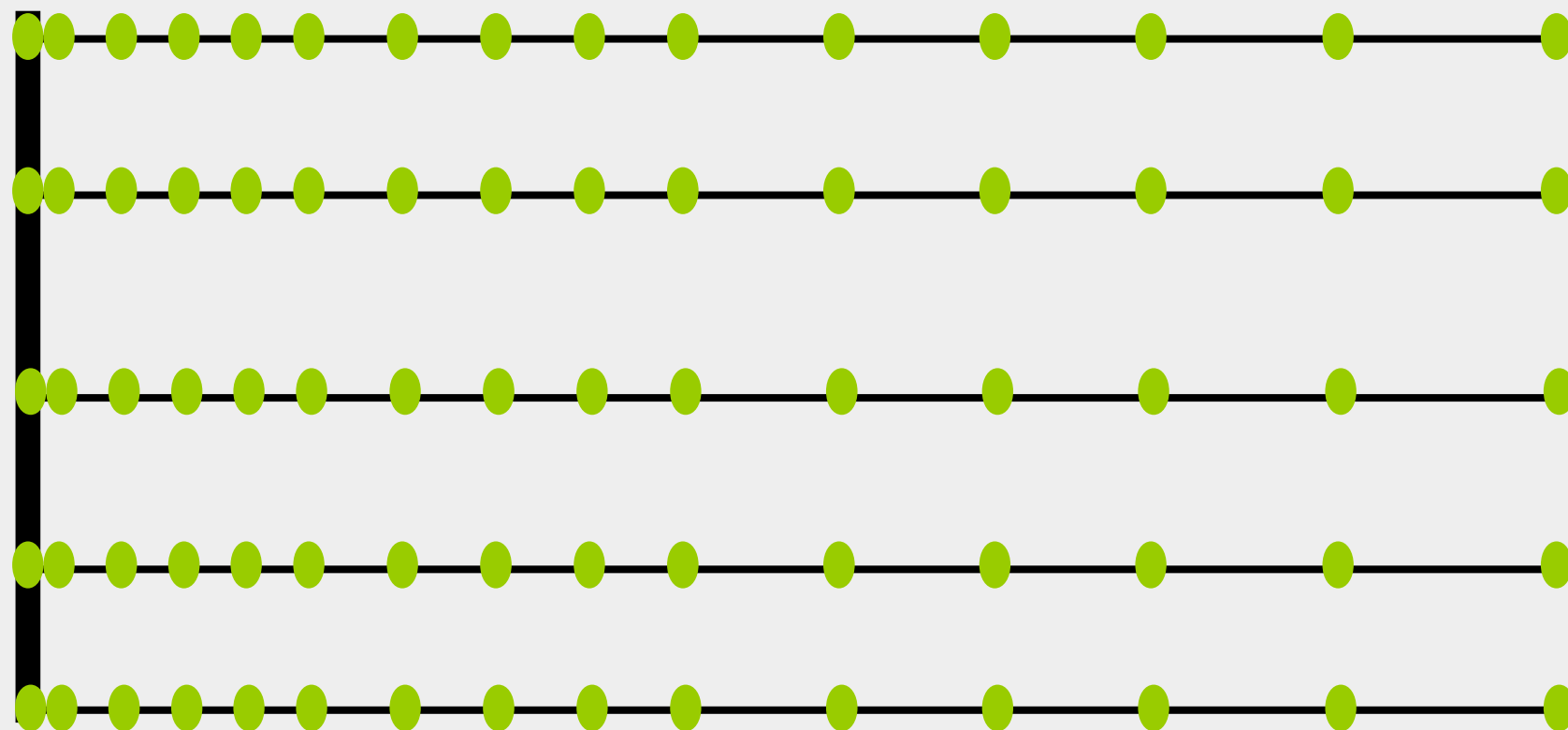
Análise de Variância

- Estatística F
 - Variação entre grupos pela variação residual

$$F = \frac{MS_{entre}}{MS_{intra}}$$



Exemplo – efeitos de borda



Borda do
fragmento

Interior do
fragmento

Exemplo – efeitos de borda

esa

ECOSPHERE

Quantifying distance of edge influence:
a comparison of methods and a new randomization method

K. A. HARPER^{1,†} AND S. E. MACDONALD²

The role of edge contrast and forest structure in edge influence: vegetation and microclimate at edges in the Brazilian cerrado

**Pavel Dodonov · Karen A. Harper ·
Dalva M. Silva-Matos**

