# Amostragem aleatória simples sem reposição (parte 2)

Prof. Caio Azevedo

# Estimação da proporção populacional

- População: observações univariadas - $y_1, ..., y_N$ (variáveis não aleatórias), em que $y_i$ é a observação relativa ao indivíduo $i$ (podemos também considerar observações multivariadas).
- Temos que $y_i = 1$ se o indivíduo $i$ possui a característica de interesse e 0 caso contrário.

# Estimação da proporção

- Exemplos: presença de alguma doença, procedência (1 se é oriundo de determinado lugar, 0, caso contrário), inadimplência (1 se inadimplente, 0 caso contrário).
- Os procedimentos definidos anteriormente, em princípio, se mantêm (slides $AAS_s$ parte 1, link). A principal diferença, de forma geral, reside na estrutura da variável de interesse.
- Parâmetro de interesse: $p = \frac{1}{N}\sum_{i=1}^{N} y_i$.
- Lembremos que $y_i = y_i^k$, $\forall k \in \mathbb{R}^+$.

# Estimação da proporção

- Estimador "natural":

$$\begin{aligned} \widehat{p} &= \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i\in \boldsymbol{S}} y_i \\ &= \frac{1}{n}\sum_{i=1}^{N} F_i y_i \end{aligned}$$

- Note que, neste caso, a variância populacional $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - p)^2$, toma a seguinte forma:

$$\begin{aligned} \sigma^2 &= \frac{1}{N}\sum_{i=1}^{N}\left(y_i^2 - 2y_i p + p^2\right) = \frac{1}{N}\left(Np - Np^2\right) \\ &= p(1-p) = pq, q = 1-p \end{aligned}$$

- Consequentemente, $s^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - p\right)^2 = \frac{N}{N-1}\sigma^2 = \frac{N}{N-1}pq$

# Propriedades do estimador

- Note que, essencialmente, $\widehat{p}$ é uma média amostral (de variáveis binárias), semelhante à $\widehat{\mu}$ (aqui).
- Portanto, as propriedades de $\widehat{p}$ são semelhantes as de $\widehat{\mu}$ (lembrando que $f = \frac{n}{N}$), sob $AAS_s$, por exemplo:
  - $\mathcal{E}_{A_2}(\widehat{p}) = \mathcal{E}_{A_2}(\widehat{\mu}) = \frac{1}{n}\sum_{i=1}^{N} y_i \mathcal{E}(F_i) = \frac{1}{n}\sum_{i=1}^{N} y_i \frac{n}{N} = \frac{1}{N}\sum_{i=1}^{N} y_i = p.$
  - $\mathcal{V}_{A_2}(\widehat{p}) = \mathcal{V}_{A_2}(\widehat{\mu}) = (1-f)\frac{s^2}{n} = (1-f)\frac{N}{N-1}\frac{pq}{n} = \frac{N-n}{N-1}\frac{pq}{n}.$
- Estimativa: $\widetilde{p} = \frac{1}{n}\sum_{i \in \boldsymbol{s}} y_i = \frac{1}{n}\sum_{i=1}^{N} f_i y_i$

# Propriedades do estimador

- Vimos também que um estimador não viciado para a variância populacional ($s^2 = \frac{N}{N-1}pq$) é dado por

$$\begin{aligned} \widehat{s^2} &= \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \widehat{\mu})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \widehat{p})^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{N}F_i\,(y_i - \widehat{p})^2 \end{aligned}$$

- Note, no entanto, que neste caso

$$\begin{aligned} \widehat{s^2} &= \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i^2 - 2Y_i\widehat{p} + \widehat{p}^2\right) = \frac{1}{n-1}\left(n\widehat{p} - n\widehat{p}^2\right) \\ &= \frac{n}{n-1}\widehat{p}\widehat{q}, \widehat{q} = 1 - \widehat{p} \end{aligned}$$

# Propriedades do estimador

- Consequentemente, um estimador não viciado para a variância do estimador é dado por:

$$(1-f)\frac{\widehat{s^2}}{n} = (1-f)\frac{n\widehat{pq}}{n(n-1)} = (1-f)\frac{\widehat{pq}}{n-1}$$

- Analogamente ao caso da média, temos que

$$\frac{\widehat{p}-p}{\sqrt{\frac{N-n}{N-1}\frac{pq}{n}}} \xrightarrow[\substack{n\to\infty,\\ N-n\to\infty}]{D} N(0,1); \frac{\widehat{p}-p}{\sqrt{(1-f)\frac{\widehat{pq}}{(n-1)}}} \xrightarrow[\substack{n\to\infty,\\ N-n\to\infty}]{D} N(0,1)$$

# Comparação dos estimadores sob os planos $A_1$ e $A_2$

- O estimador para a proporção sob $AAS_c$ ou $AAS_s$, é o mesmo.
- Temos que $\mathcal{E}_{A_i}(\widehat{p}) = p$, $i = 1, 2$.
- Portanto, o efeito do planejamento (EPA), do estimador sob o plano $A_2$ em relação ao plano $A_1$, é dado por:

$$EPA = \frac{\mathcal{V}_{A_2}(p)}{\mathcal{V}_{A_1}(p)} = \frac{(1-f)\frac{N}{N-1}\frac{pq}{n}}{\frac{pq}{n}} = (1-f)\frac{N}{N-1}$$

- Assim, quando $N \to \infty$, $\mathcal{V}_{A_2}(\widehat{p}) \to \mathcal{V}_{A_1}(\widehat{p})$ (semelhante ao que ocorre com $\widehat{\mu}$).

# Comparação dos estimadores sob os planos $A_1$ e $A_2$

- Consequentemente, temos que o plano $AAS_s$ é melhor do que $AAS_c$, tendendo ambos a serem equivalentes, à medida que o tamanho da população tende a infinito.
- Exercício: Calcular o EPA para $\widehat{\tau}$.
- Defina um estimador para o EPA, da seguinte forma: $\widehat{EPA} = \frac{\widehat{\mathcal{V}}_{A_2}(\widehat{\theta})}{\widehat{\mathcal{V}}_{A_1}(\widehat{\theta})}$, em que $\widehat{\theta}$ é um estimador escolhido, para o parâmetro de interesse $(\theta)$ e $\widehat{\mathcal{V}}_A(\widehat{\theta})$ é o estimador da variância do estimador de $\widehat{\theta}$. Exercício: calcular o $\widehat{EPA}$ para os estimadores $\widehat{\mu}$, $\widehat{\tau}$ e $\widehat{p}$.
- OBS: Se for necesssário, é possível ser mais específico com relação à notação do EPA ($\widehat{EPA}$), ou seja, escrevendo $EPA(\theta)(\widehat{EPA}(\theta))$.

# Intervalo de Confiança

- Erro-padrão do estimador: $EP_{A_2}(\widehat{p}) = \sqrt{\mathcal{V}_{A_2}(\widehat{p})}$.
- Um estimador da variância do estimador: $\widehat{\mathcal{V}}_{A_2}(\widehat{p}) = (1 - f)\,\frac{\widehat{p}\widehat{q}}{n-1}$.
- Um estimador do erro-padrão do estimador: $\widehat{EP}_{A_2}(\widehat{p}) = \sqrt{\widehat{\mathcal{V}}_{A_2}(\widehat{p})}$.

# Intervalo de Confiança

- Assim, dois intervalos de confiança (assintóticos) com coeficiente de confiança de aproximadamente $\gamma$, são dados por

$$IC(\mu,\gamma) \approx \left[\widehat{p} - z_\gamma\sqrt{(1-f)\frac{\widehat{pq}}{n-1}}; \widehat{p} + z_\gamma\sqrt{(1-f)\frac{\widehat{pq}}{n-1}}\right] \quad (1)$$

$$IC(\mu,\gamma) \approx \left[\widehat{p} - z_\gamma\sqrt{\frac{1-f}{4(n-1)}}; \widehat{p} + z_\gamma\sqrt{\frac{1-f}{4(n-1)}}\right] \quad (2)$$

em que $P(Z \leq z_\gamma) = \frac{1+\gamma}{2}$ e $Z \sim N(0,1)$.

- Erro da estimativa: $z_\gamma\sqrt{(1-f)\frac{\widehat{pq}}{n-1}}$ ou $z_\gamma\sqrt{\frac{1-f}{4(n-1)}}$.
- O comprimento do intervalo (2) sempre será maior (ou igual) ao comprimento do intervalo (1).

# Testes de Hipótese

- Hipóteses usuais ($p_0$ conhecido, $q_0 = 1 - p_0$)
  1. $H_0 : p = p_0$ vs $H_1 : p < p_0$.
  2. $H_0 : p = p_0$ vs $H_0 : p > p_0$.
  3. $H_0 : p = p_0$ vs $H_0 : p \neq p_0$.
- Estatística do teste $Z_t = \frac{\widehat{p} - p_0}{\sqrt{[(N-n)/(N-1)]p_0 q_0 / n}}$.
- Sob $H_0$, vimos que $Z_t \approx N(0,1)$, para n e N-n suficientemente grandes.
- Defina $z_t = \frac{\widetilde{p} - p_0}{\sqrt{[(N-n)/(N-1)]p_0 q_0 / n}}$ o valor calculado da estatística do teste e $z_c$ o(s) valor(es) crítico(s).
- Defina ainda $Z \sim N(0,1)$. Os procedimentos são análogos ao caso da média, com as devidas adaptações.

# Determinação do tamanho amostral: erro da estimativa

Analogamente ao caso da média populacional, temos que

$$\delta = z_\gamma \sqrt{(1-f)\frac{pq}{n}} \rightarrow n = \frac{1}{\frac{\delta^2}{z_\gamma^2 pq} + \frac{1}{N}} \tag{3}$$

Podemos usar estimativas de $p$ obtidas em pesquisas anteriores, sob uma amostra piloto ou, considerar o pior caso, em termos da variabilidade dos dados. Neste último caso, temos que:

$$n = \frac{1}{\frac{4\delta^2}{z_\gamma^2} + \frac{1}{N}} \tag{4}$$

Isto vale para qualquer um dos dois critérios: erro da estimativa e precisão. Note que o tamanho da amostra fornecido por (4) será maior ou igual àquele fornecido por (3).

# Estudos de simulação

- Distribuição assintótica do estimador para a proporção. Tamanho da população $N = 100.000$.
- Vários cenários, variando em função do valor verdadeiro da proporção populacional $p$.
- $p =$ $(0,01; 0,05; 0,10; 0,25; 0,35; 0,50; 0,65; 0,75; 0,9; 0,95; 0,99)^T$.
- A distribuição, (em princípio), da variável de interesse é Bernoulli($p$).

# Estudos de simulação

- Quatro tamanhos amostrais $(30, 50, 100, 1000)$, em termos percentuais, com relação ao tamanho da população (0,03%,0,05%,0,1%,1%).
- Estudar a distribuição amostral (empírica) com base em $R = 1.000$ réplicas (amostras selecionadas da população de interesse).

# p=0,01

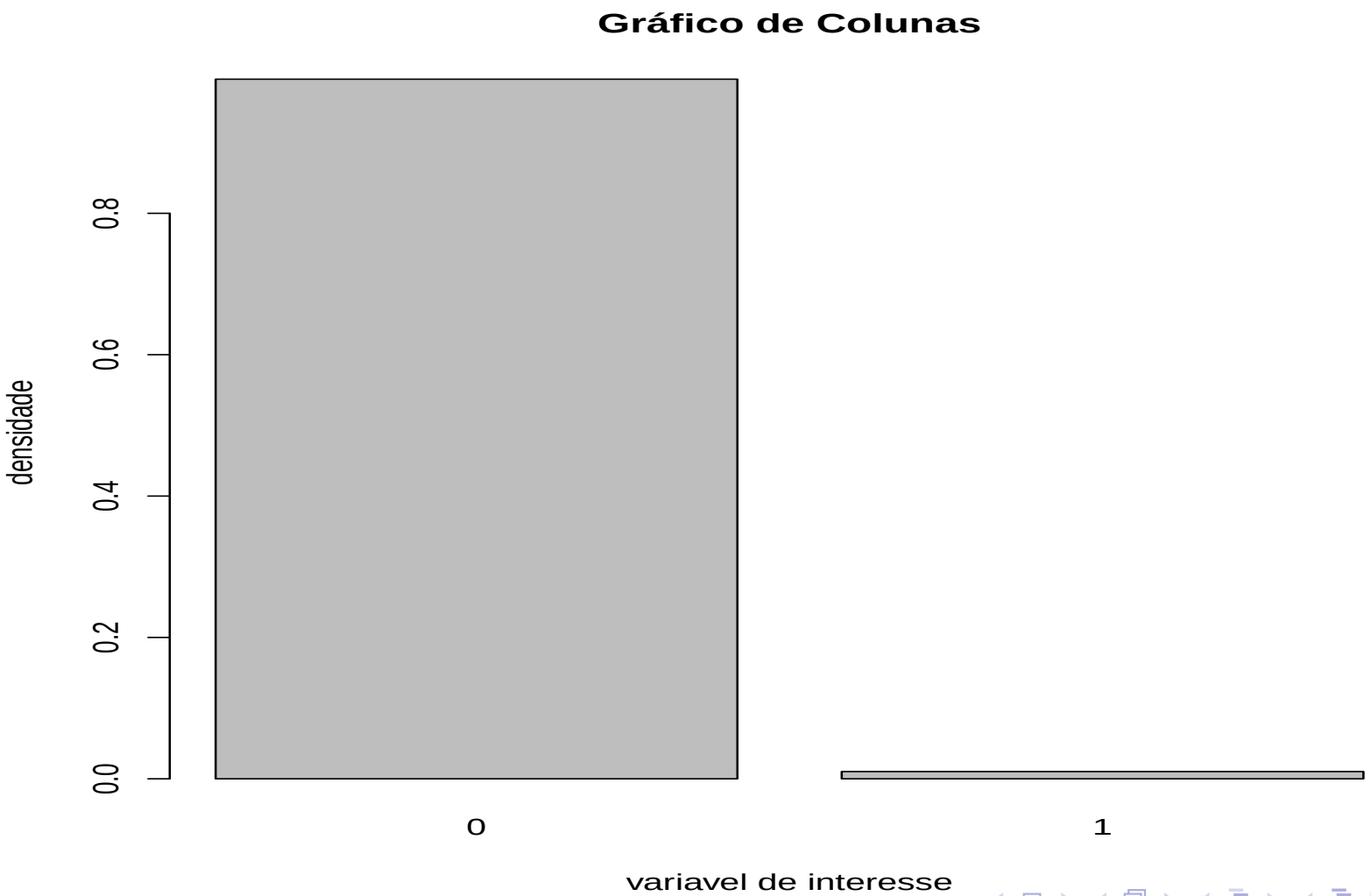
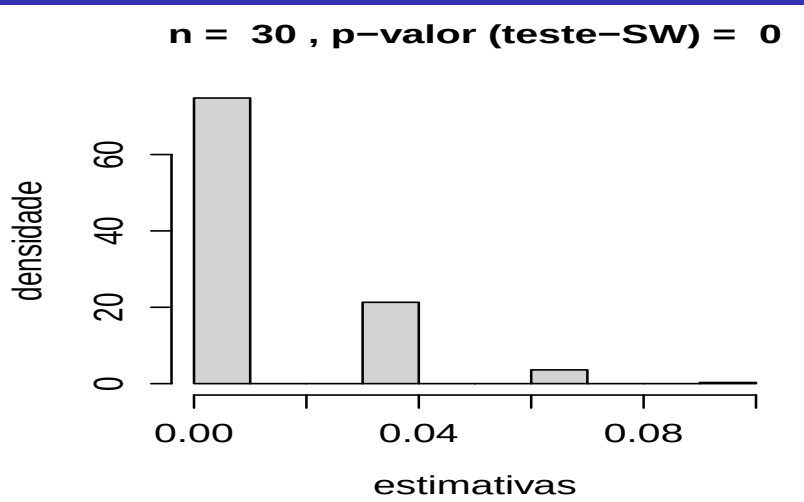
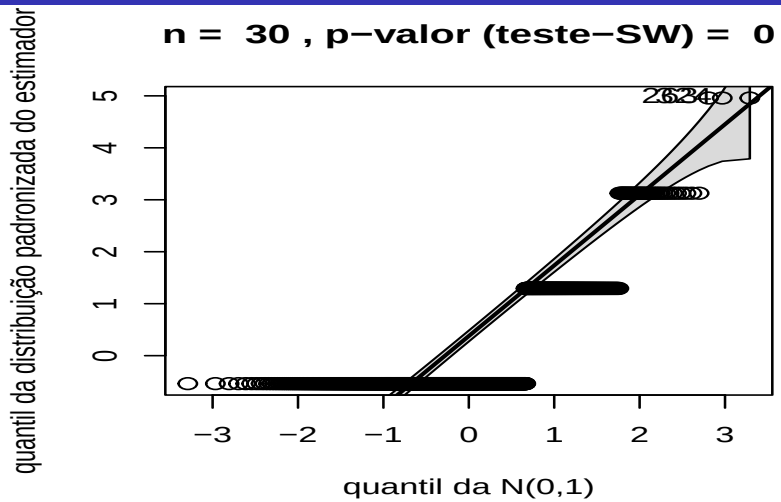
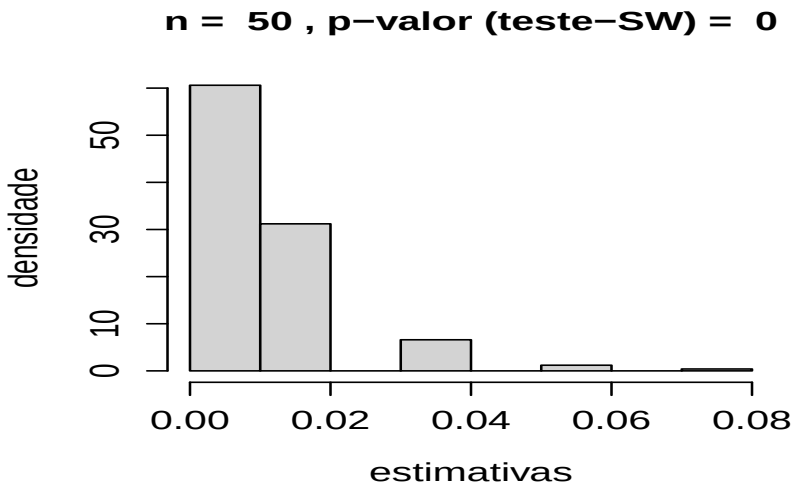


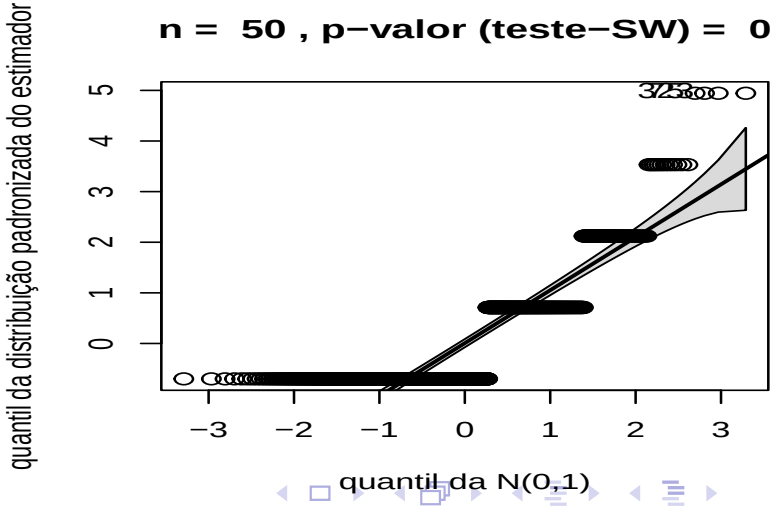
Gráfico de Colunas
densidade
0.0
0.2
0.4
0.6
0.8
0
1
variavel de interesse

# p=0,01

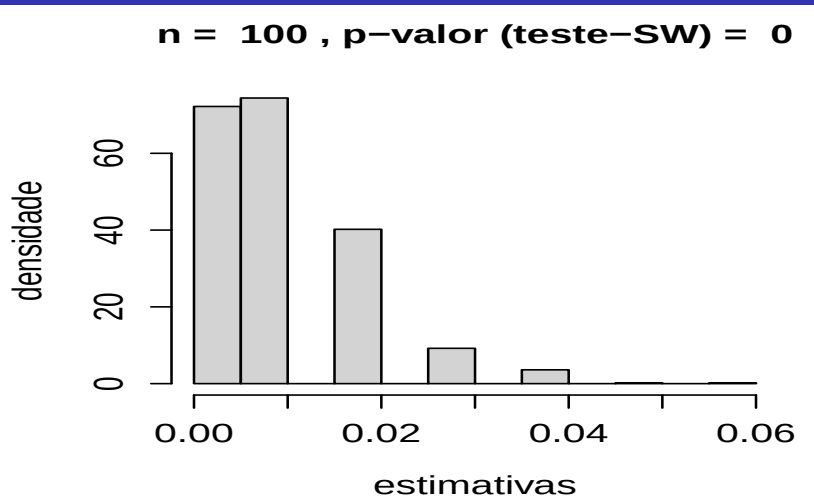

n = 30 , p−valor (teste−SW) = 0
densidade
0
20
40
60
0.00
0.04
0.08
estimativas

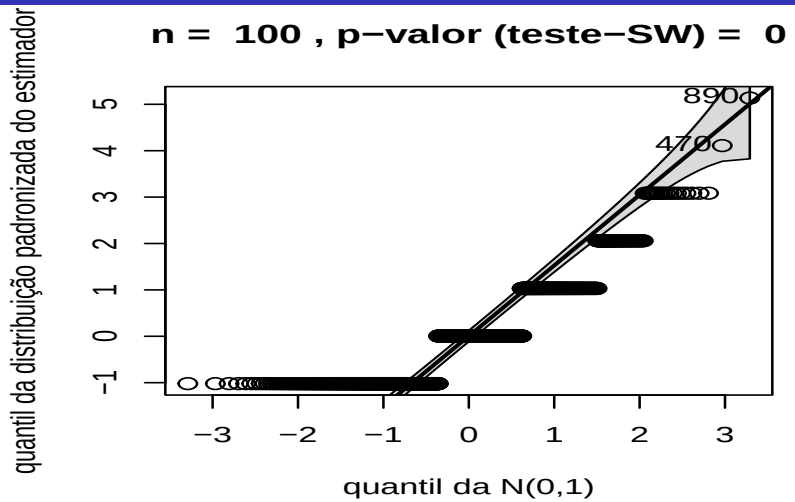

n = 30 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
0
1
2
3
4
5
−3
−2
−1
0
1
2
3
quantil da N(0,1)

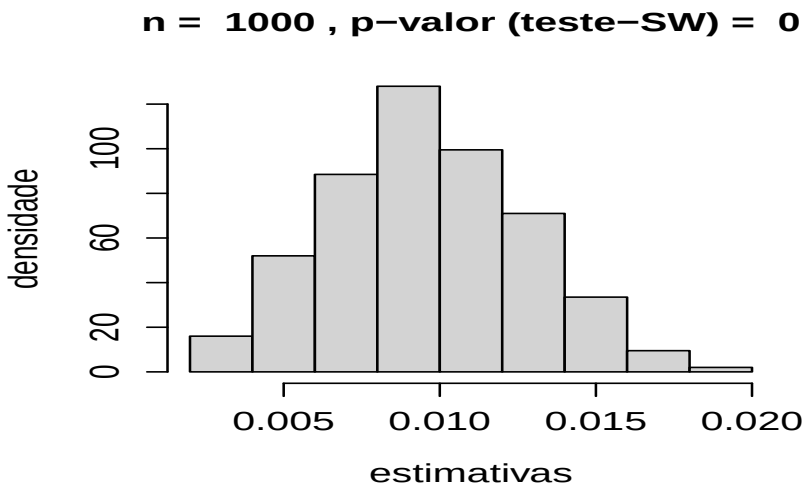

n = 50 , p−valor (teste−SW) = 0
densidade
0
10
30
50
0.00
0.02
0.04
0.06
0.08
estimativas

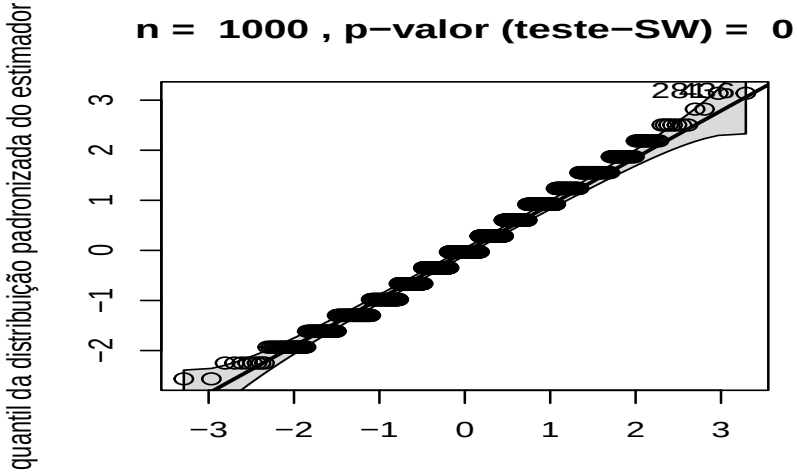

n = 50 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
0
1
2
3
4
5
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# p=0,01


n = 100 , p−valor (teste−SW) = 0
densidade
0
20
40
60
0.00
0.02
0.04
0.06
estimativas


n = 100 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−1
0
1
2
3
4
5
890
470
−3
−2
−1
0
1
2
3
quantil da N(0,1)


n = 1000 , p−valor (teste−SW) = 0
densidade
0
20
60
100
0.005
0.010
0.015
0.020
estimativas


n = 1000 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−2
−1
0
1
2
3
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# p=0,05

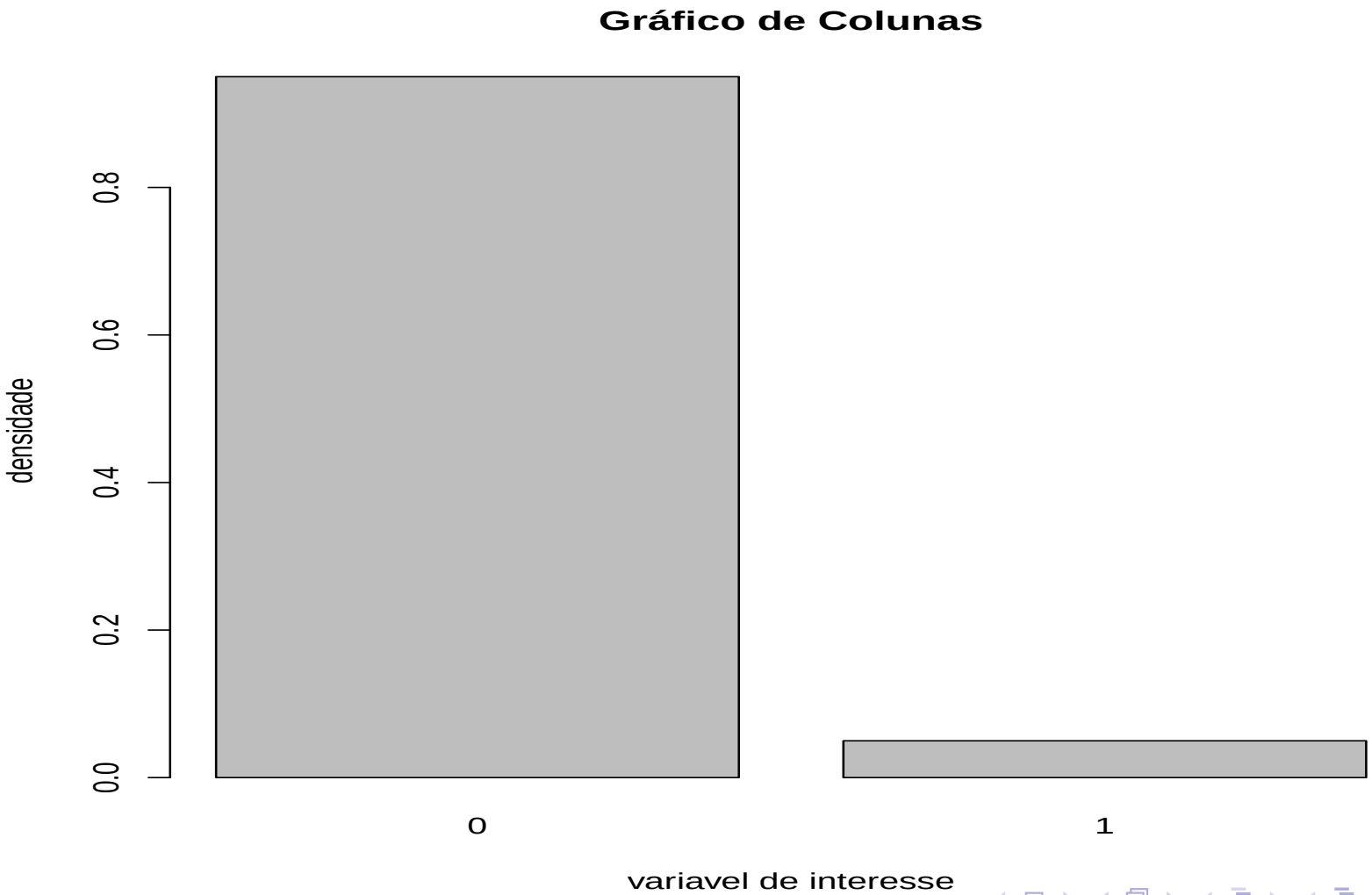
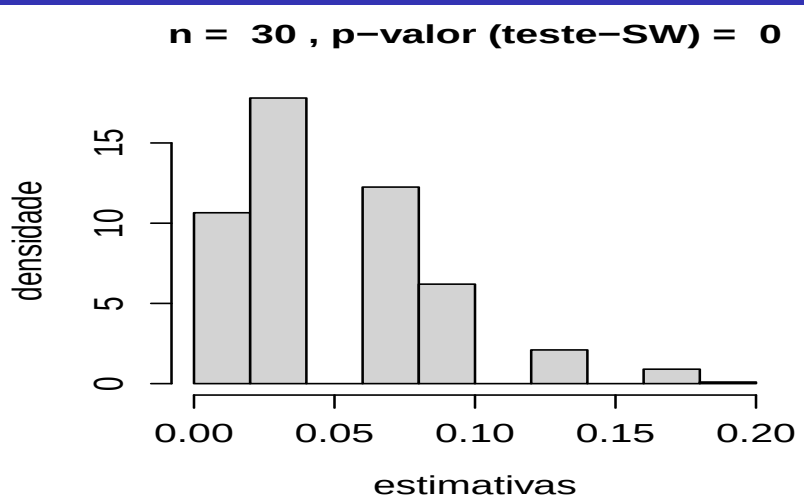
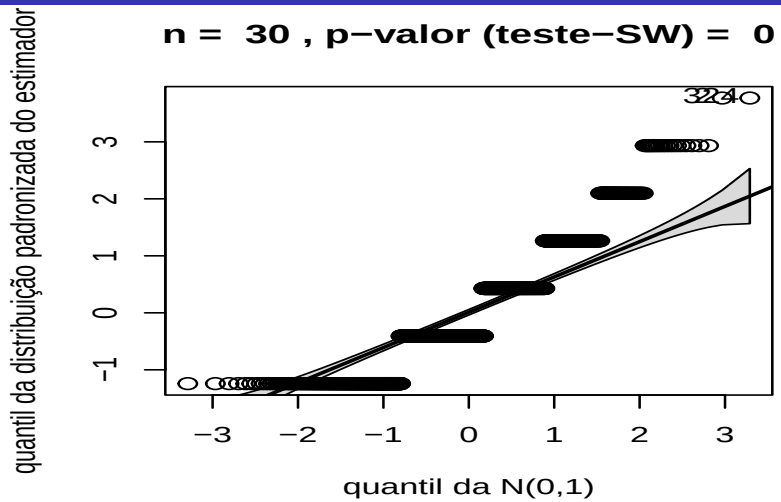
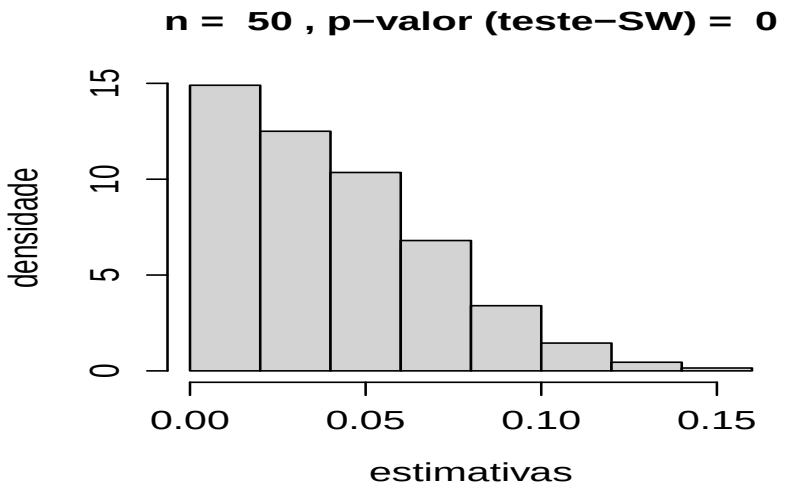


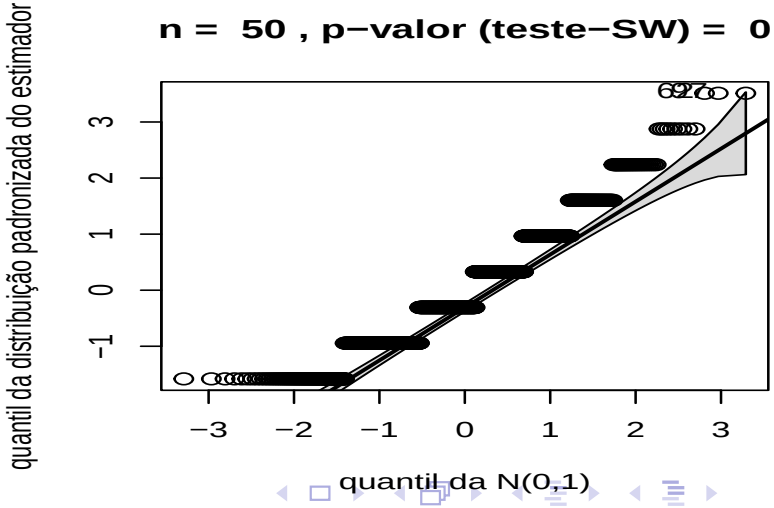
Gráfico de Colunas
densidade
0.0
0.2
0.4
0.6
0.8
0
1
variavel de interesse

# p=0,05

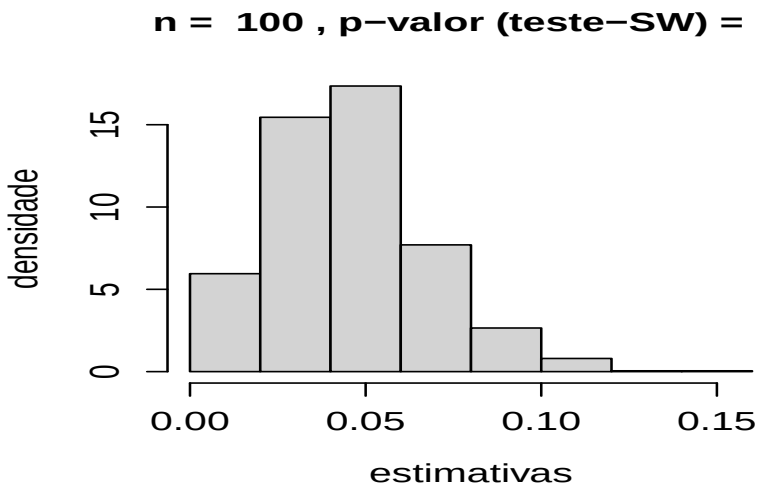

n = 30 , p−valor (teste−SW) = 0
densidade
0
5
10
15
0.00
0.05
0.10
0.15
0.20
estimativas

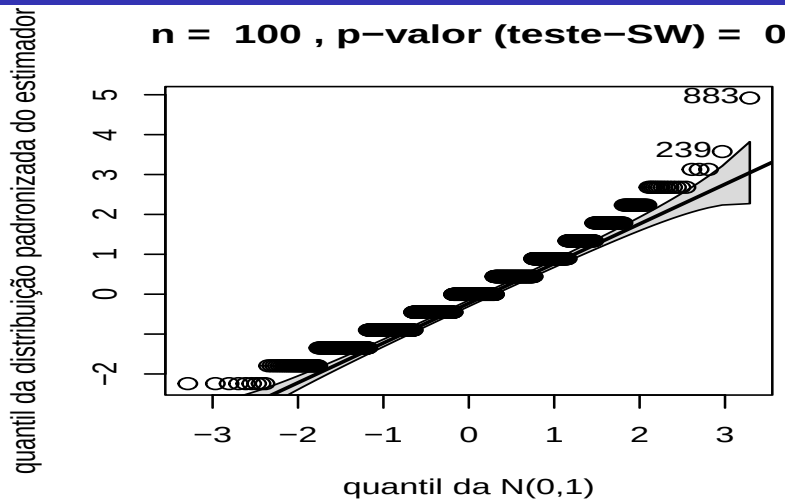

n = 30 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−1
0
1
2
3
−3
−2
−1
0
1
2
3
quantil da N(0,1)

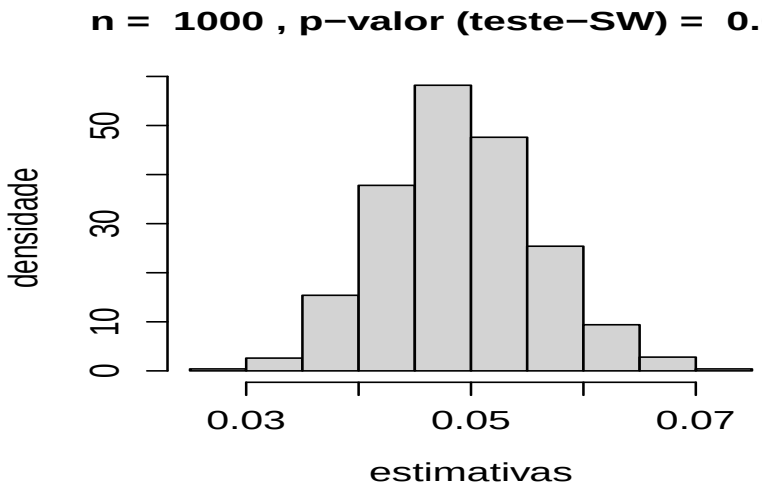

n = 50 , p−valor (teste−SW) = 0
densidade
0
5
10
15
0.00
0.05
0.10
0.15
estimativas

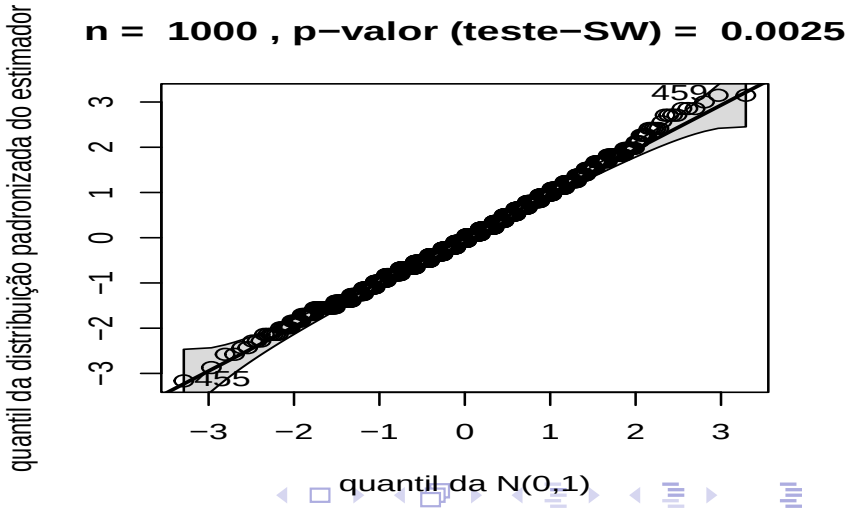

n = 50 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−1
0
1
2
3
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# p=0,05


n = 100 , p−valor (teste−SW) = 0
densidade
0
5
10
15
0.00
0.05
0.10
0.15
estimativas


n = 100 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
883
239
−2
0
1
2
3
4
5
−3
−2
−1
0
1
2
3
quantil da N(0,1)


n = 1000 , p−valor (teste−SW) = 0.0025
densidade
0
10
30
50
0.03
0.05
0.07
estimativas


n = 1000 , p−valor (teste−SW) = 0.0025
quantil da distribuição padronizada do estimador
4596
455
−3
−2
−1
0
1
2
3
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# p=0,10

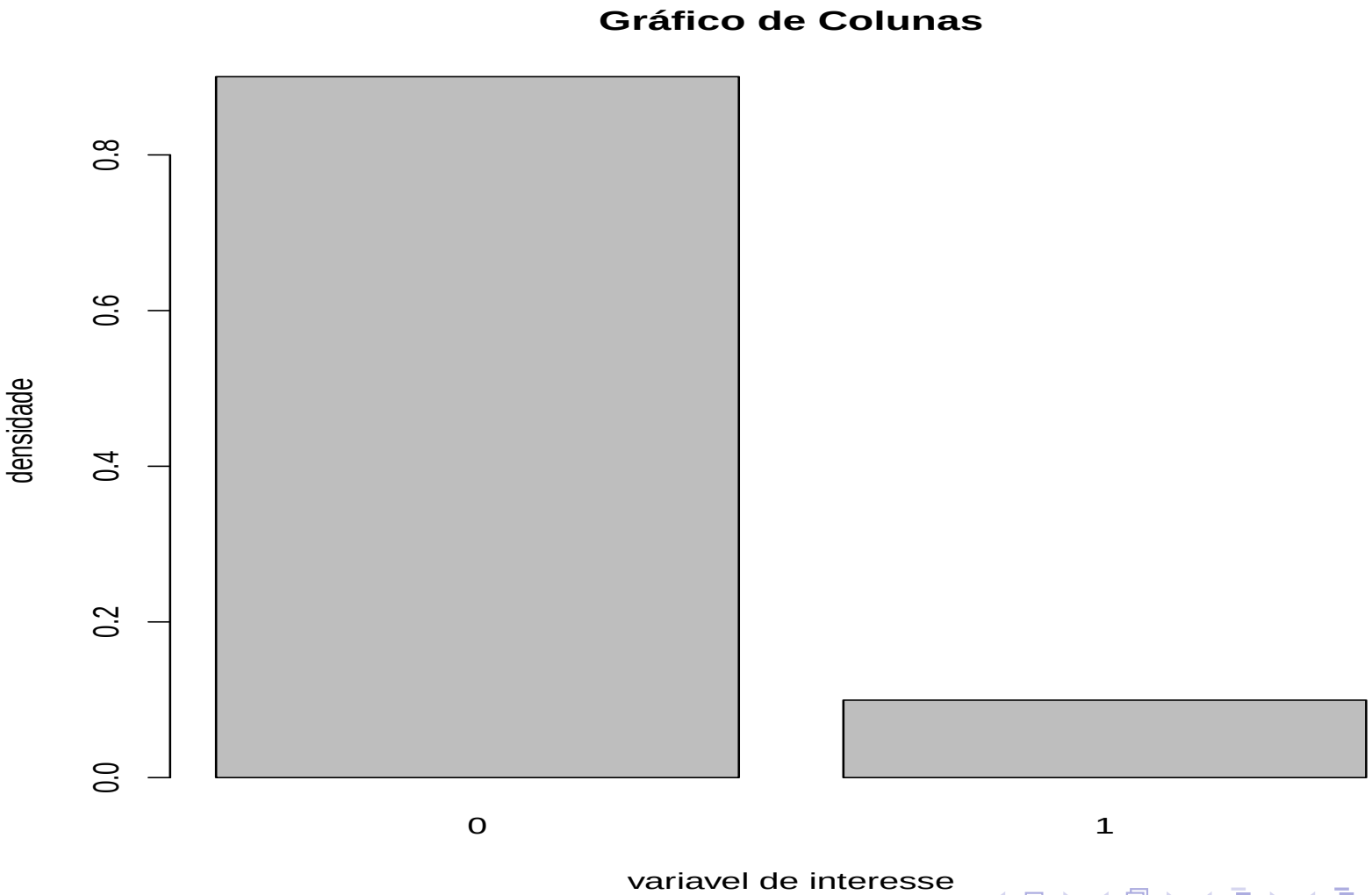
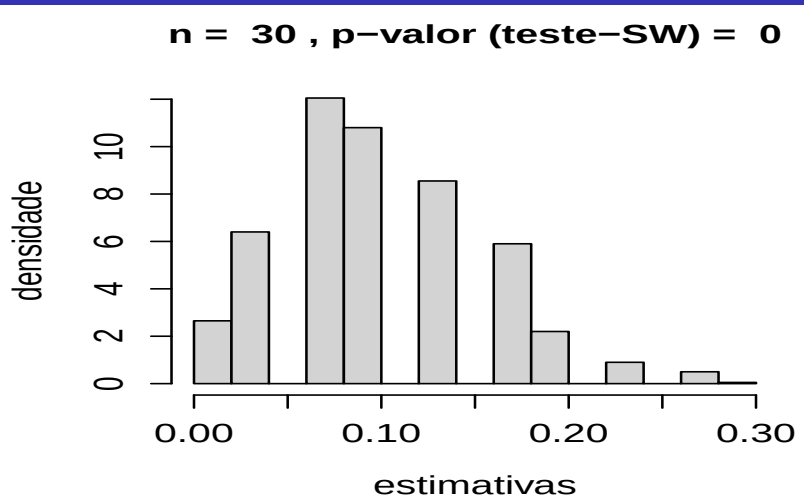
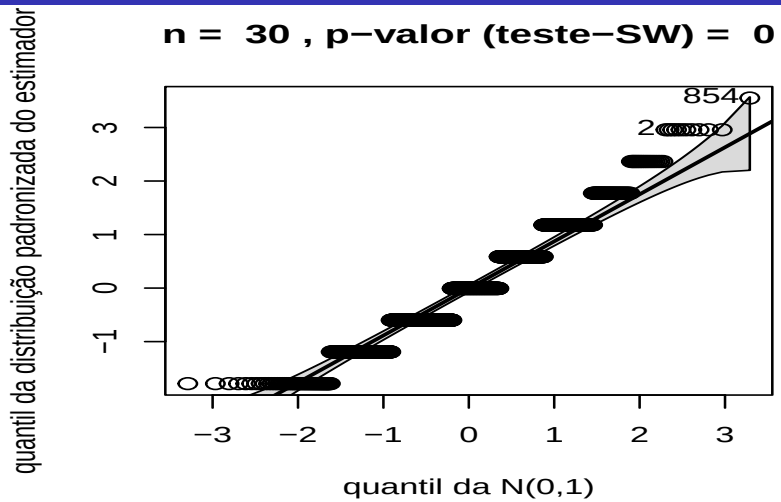
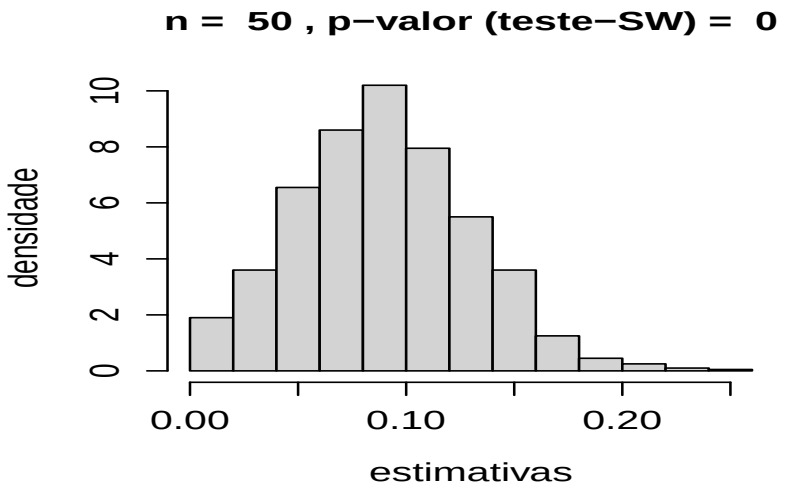


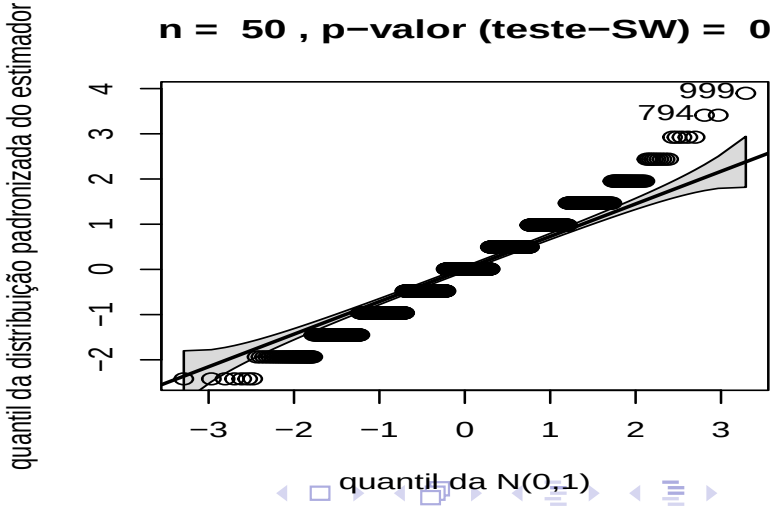
Gráfico de Colunas
densidade
0.0
0.2
0.4
0.6
0.8
0
1
variavel de interesse

# p=0,10

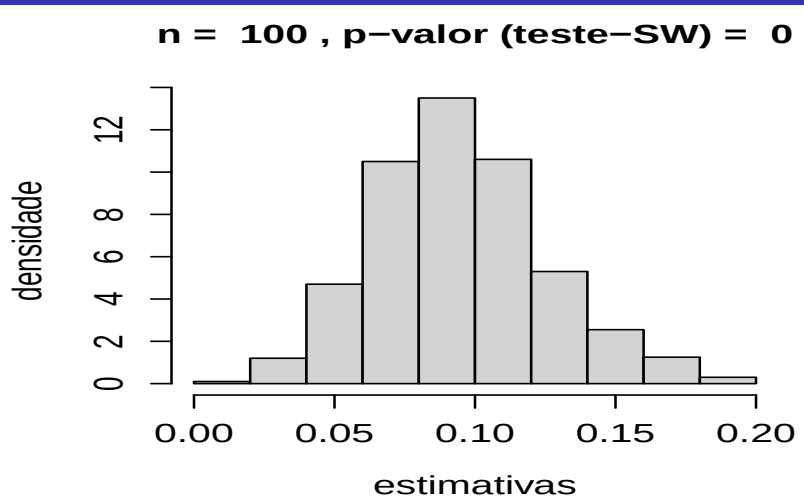

n = 30 , p−valor (teste−SW) = 0
densidade
0 2 4 6 8 10
0.00 0.10 0.20 0.30
estimativas

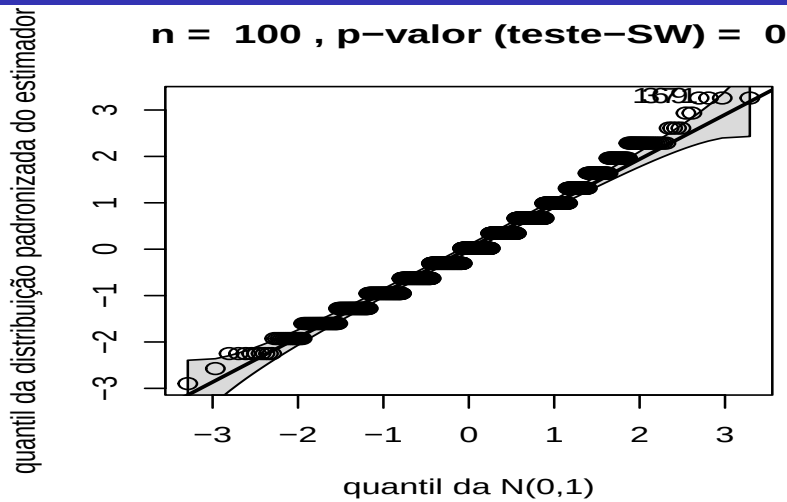

n = 30 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−1 0 1 2 3
854
2
−3 −2 −1 0 1 2 3
quantil da N(0,1)

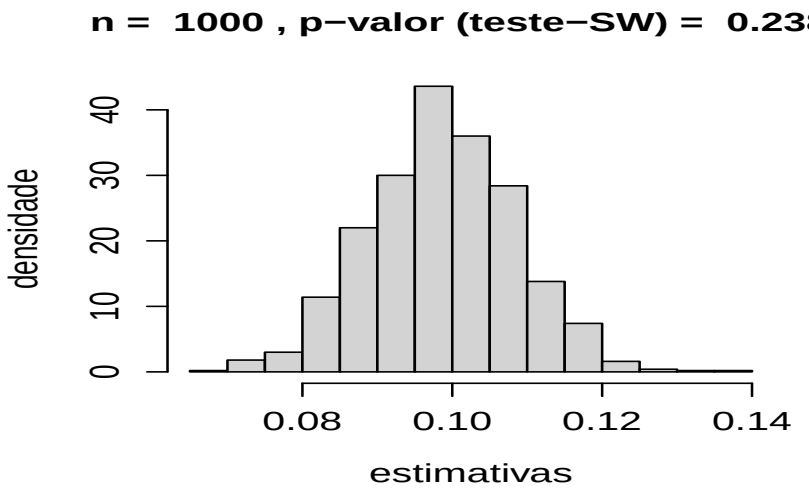

n = 50 , p−valor (teste−SW) = 0
densidade
0 2 4 6 8 10
0.00 0.10 0.20
estimativas

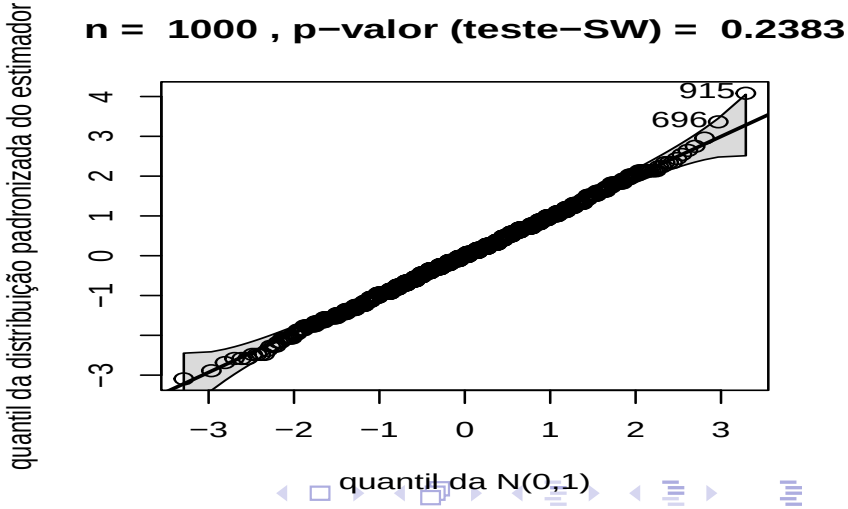

n = 50 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−2 −1 0 1 2 3 4
999
794
−3 −2 −1 0 1 2 3
quantil da N(0,1)

# p=0,10


n = 100 , p−valor (teste−SW) = 0
densidade
0 2 4 6 8 12
0.00 0.05 0.10 0.15 0.20
estimativas


n = 100 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3 −2 −1 0 1 2 3
−3 −2 −1 0 1 2 3
quantil da N(0,1)


n = 1000 , p−valor (teste−SW) = 0.2383
densidade
0 10 20 30 40
0.08 0.10 0.12 0.14
estimativas


n = 1000 , p−valor (teste−SW) = 0.2383
quantil da distribuição padronizada do estimador
−3 −1 0 1 2 3 4
915
696
−3 −2 −1 0 1 2 3
quantil da N(0,1)

# p=0,25

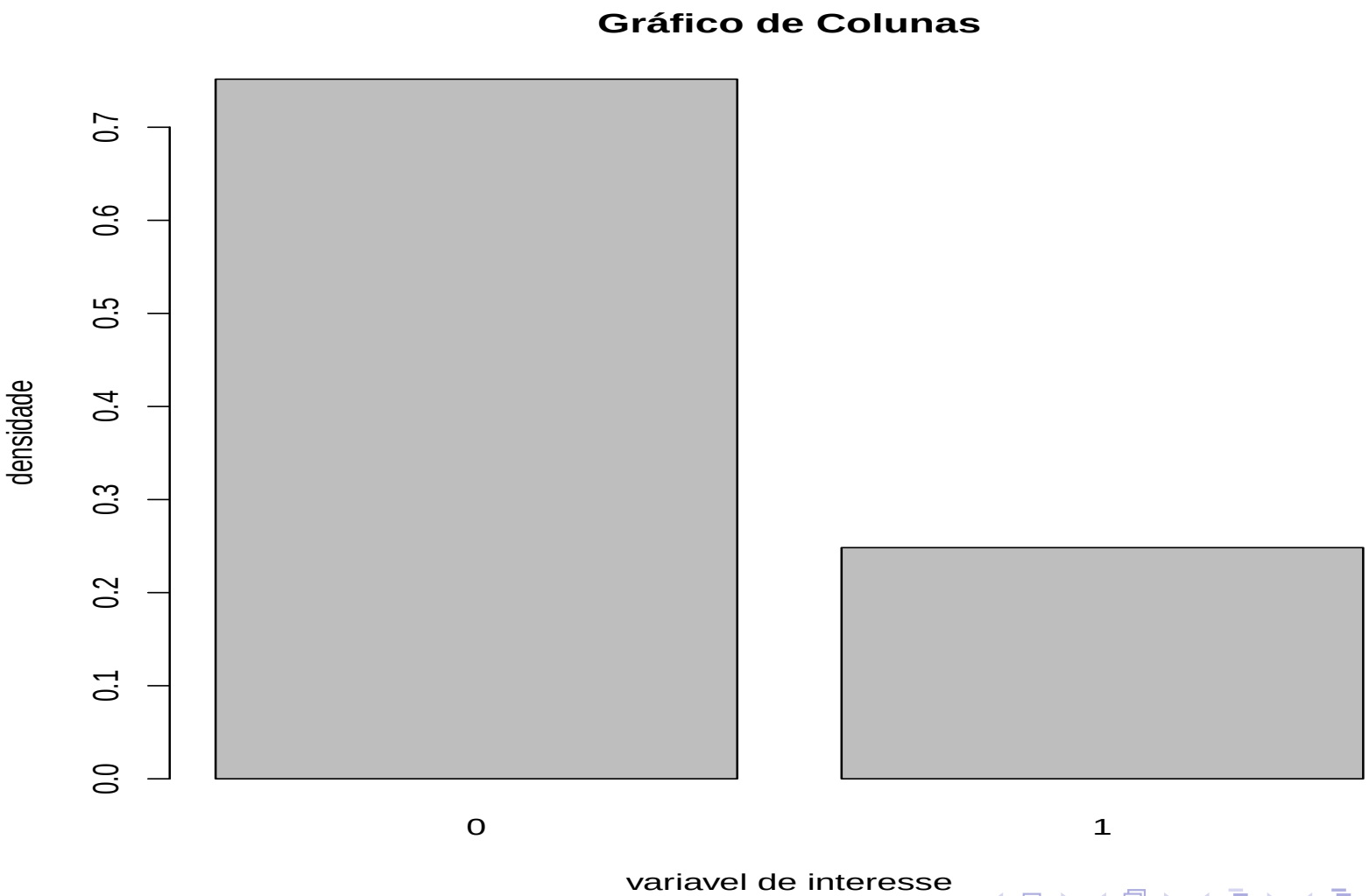
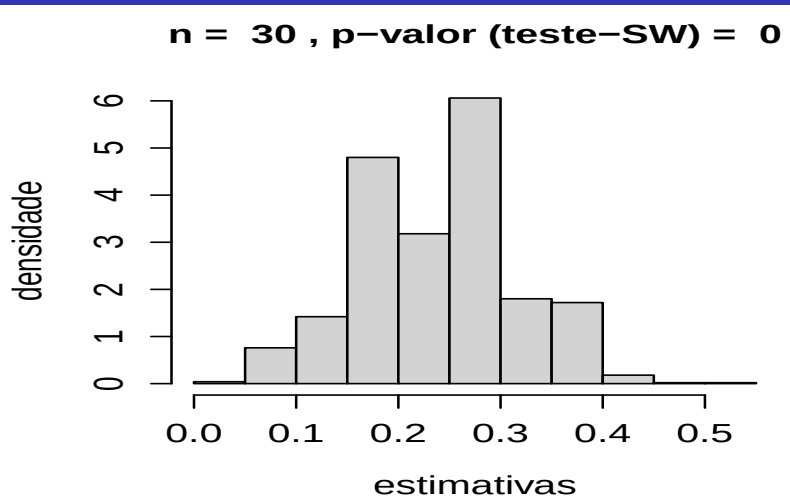
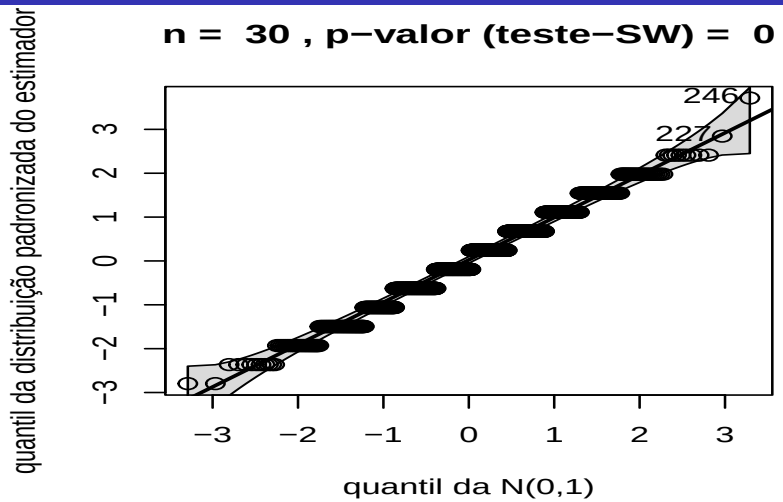
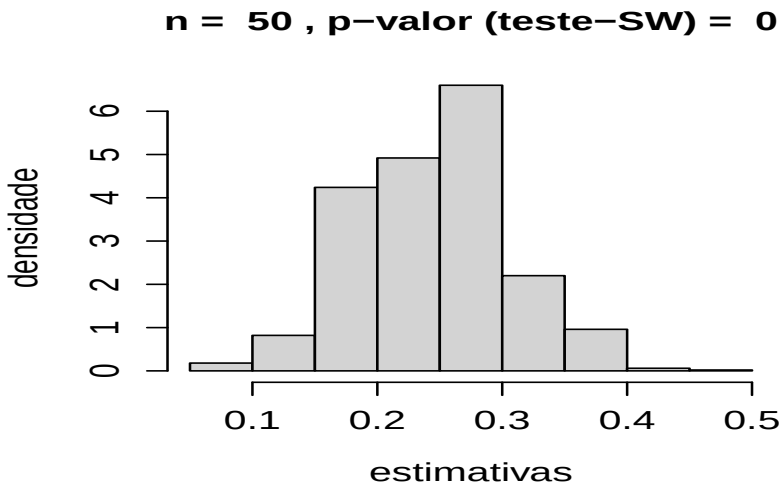


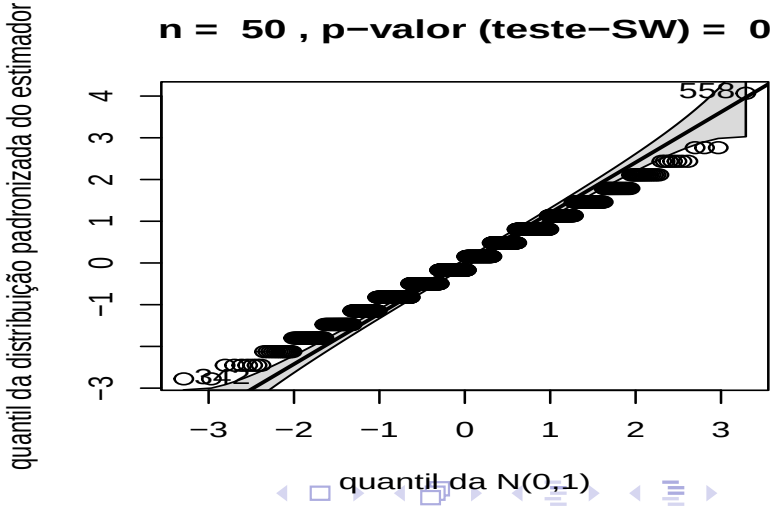
Gráfico de Colunas
densidade
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1
variavel de interesse

# p=0,25

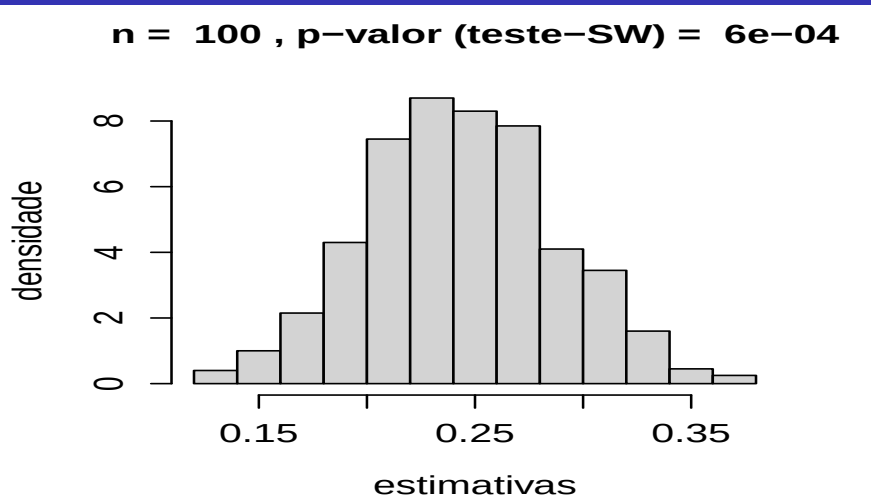

n = 30 , p−valor (teste−SW) = 0
densidade
0 1 2 3 4 5 6
0.0 0.1 0.2 0.3 0.4 0.5
estimativas

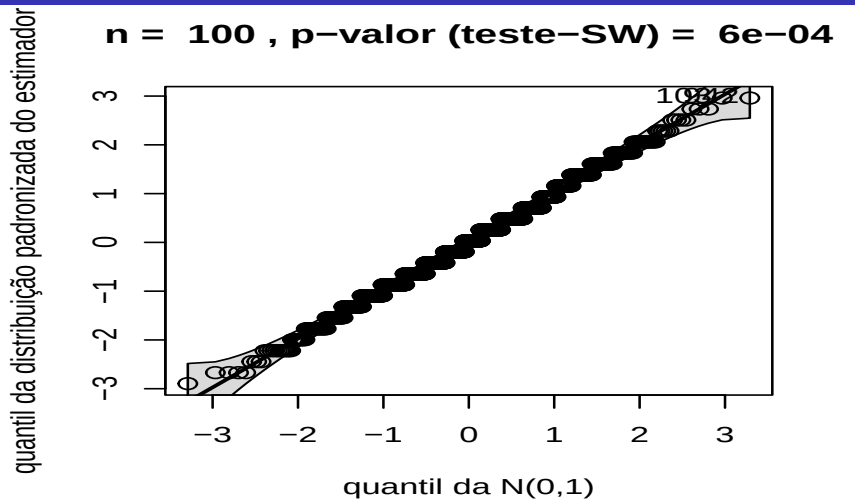

n = 30 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3 −2 −1 0 1 2 3
2460
2270
−3 −2 −1 0 1 2 3
quantil da N(0,1)

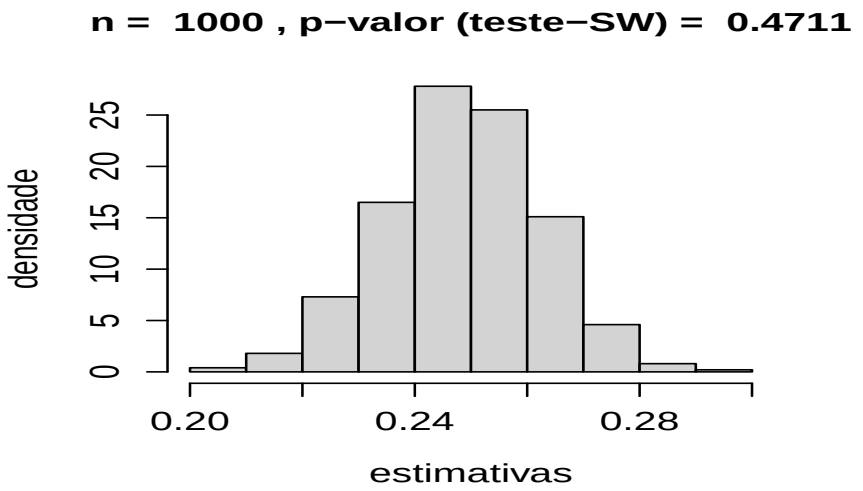

n = 50 , p−valor (teste−SW) = 0
densidade
0 1 2 3 4 5 6
0.1 0.2 0.3 0.4 0.5
estimativas

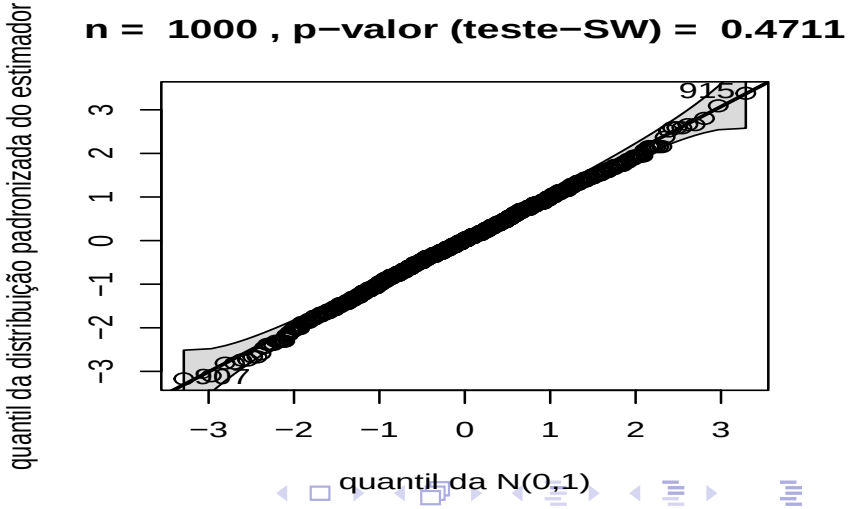

n = 50 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3 −1 0 1 2 3 4
5580
3412
−3 −2 −1 0 1 2 3
quantil da N(0,1)

# p=0,25


n = 100 , p−valor (teste−SW) = 6e−04
densidade
0
2
4
6
8
0.15
0.25
0.35
estimativas


n = 100 , p−valor (teste−SW) = 6e−04
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
3
quantil da N(0,1)


n = 1000 , p−valor (teste−SW) = 0.4711
densidade
0
5
10
15
20
25
0.20
0.24
0.28
estimativas


n = 1000 , p−valor (teste−SW) = 0.4711
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# p=0,35

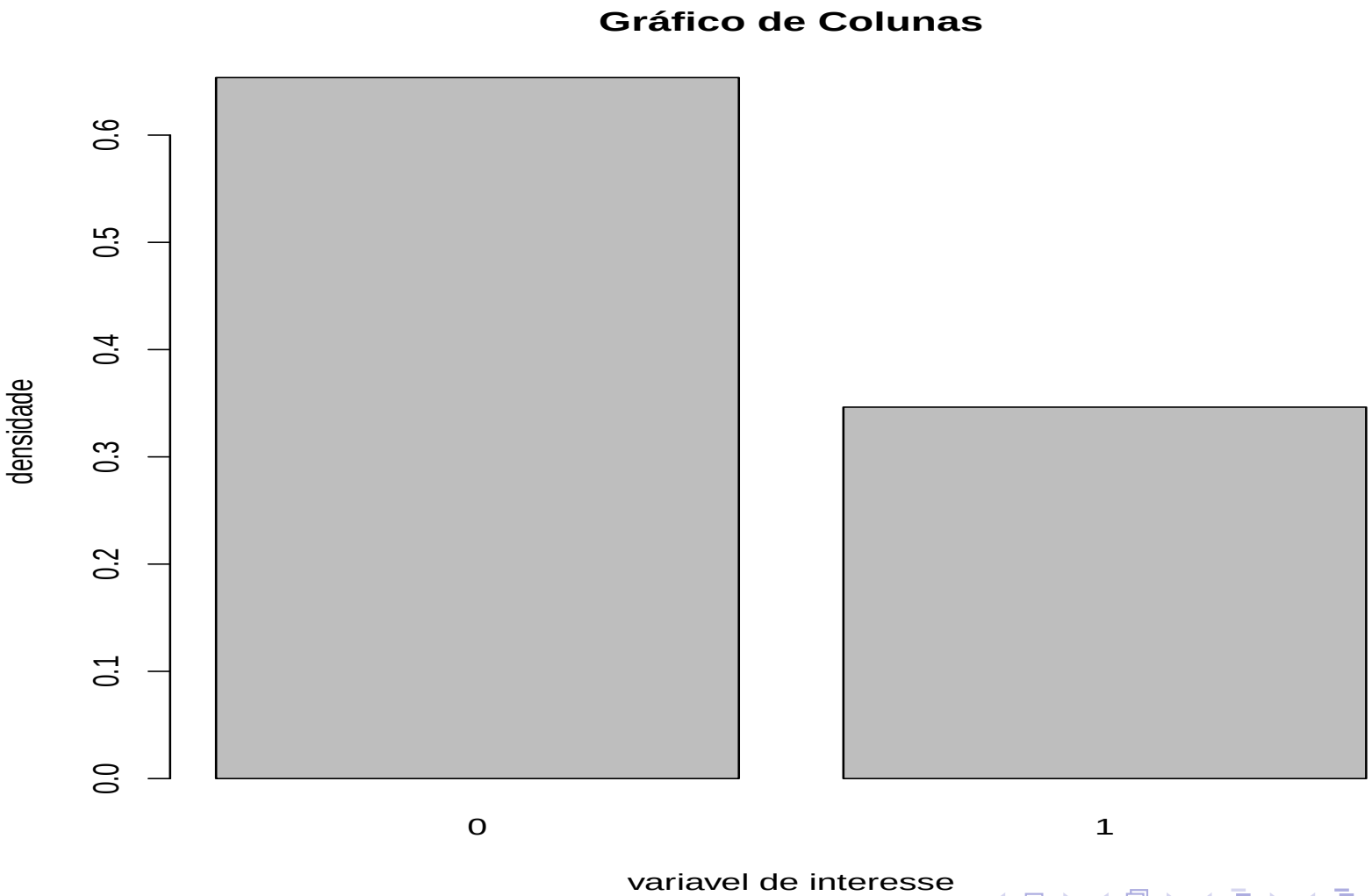
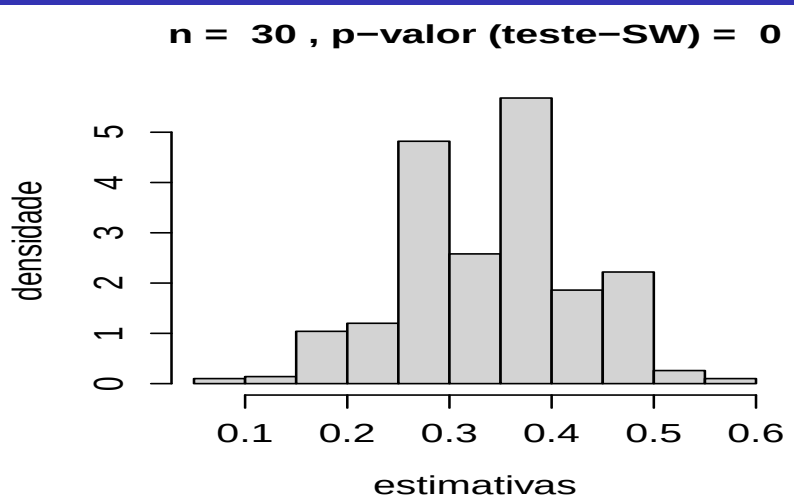
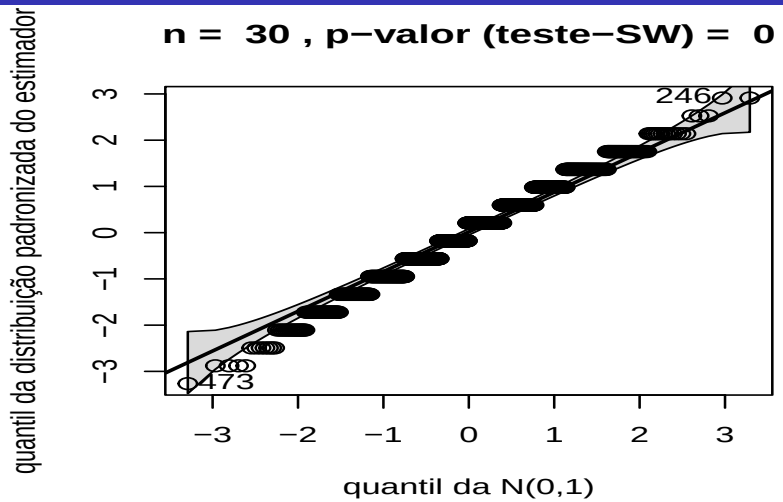
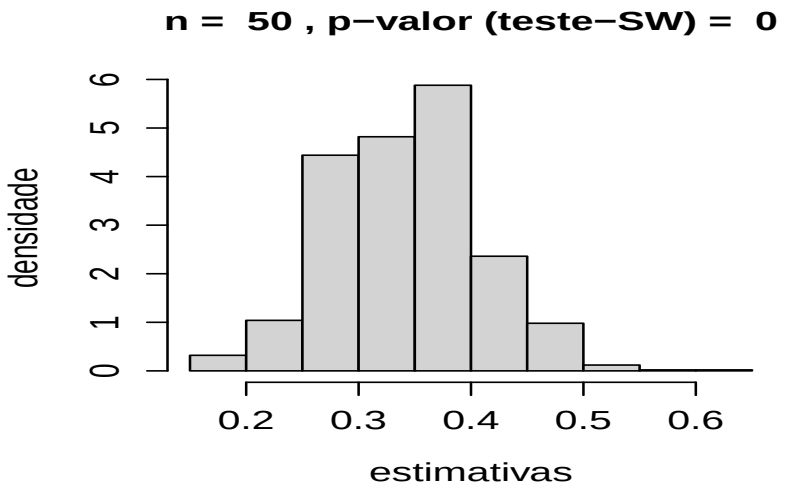


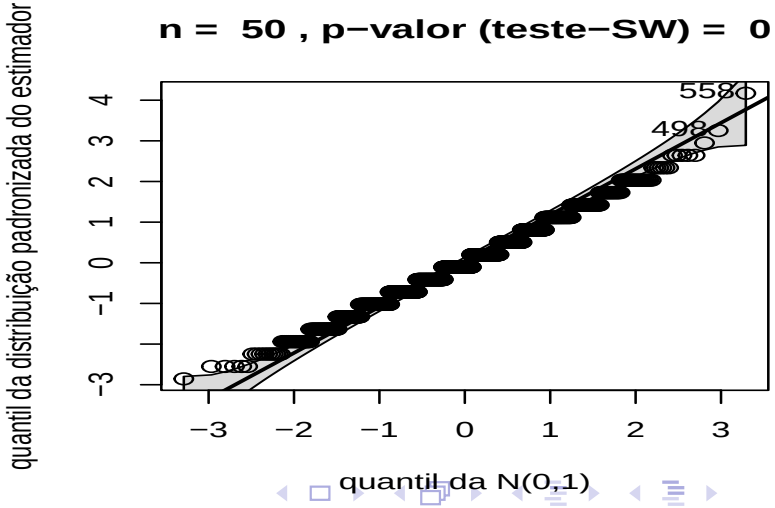
Gráfico de Colunas
densidade
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
1
variavel de interesse

# p=0,35

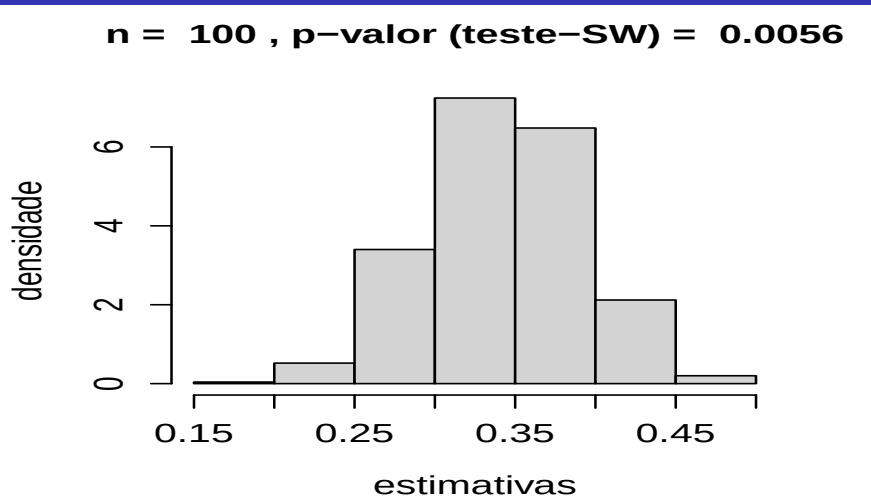

n = 30 , p−valor (teste−SW) = 0
densidade
0 1 2 3 4 5
0.1 0.2 0.3 0.4 0.5 0.6
estimativas

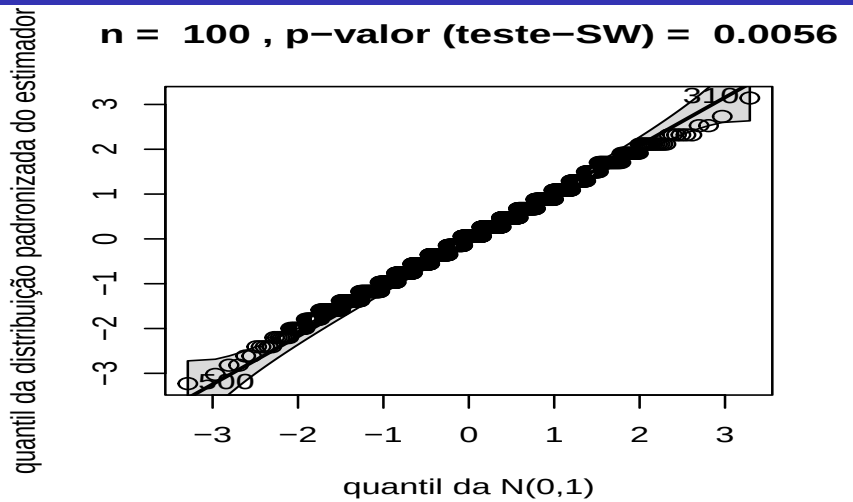

n = 30 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3 −2 −1 0 1 2 3
246
473
−3 −2 −1 0 1 2 3
quantil da N(0,1)

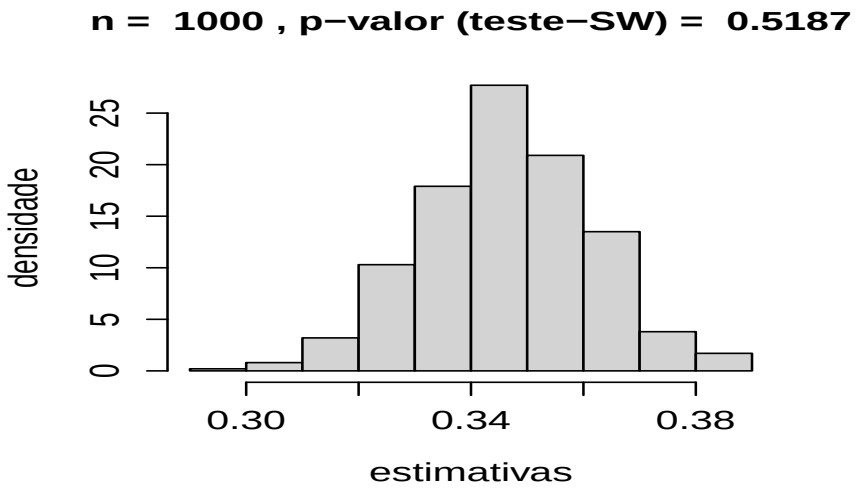

n = 50 , p−valor (teste−SW) = 0
densidade
0 1 2 3 4 5 6
0.2 0.3 0.4 0.5 0.6
estimativas

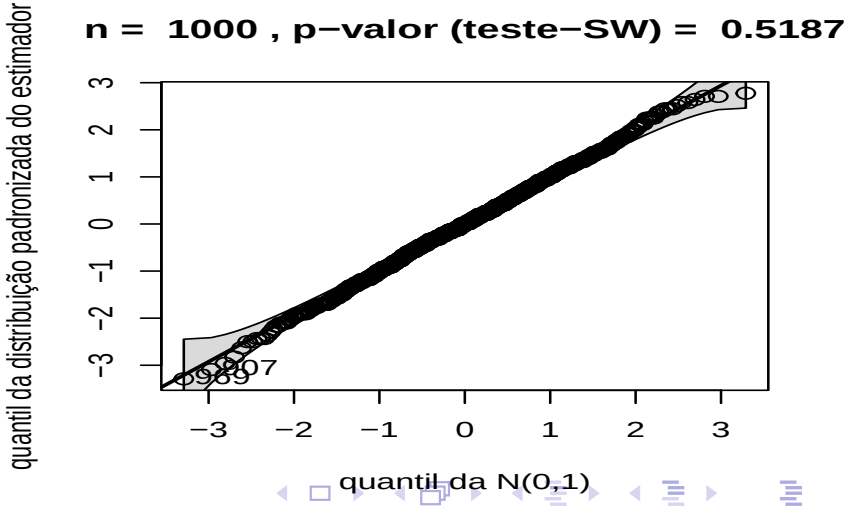

n = 50 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3 −1 0 1 2 3 4
558
498
−3 −2 −1 0 1 2 3
quantil da N(0,1)

# p=0,35


n = 100 , p−valor (teste−SW) = 0.0056
densidade
0
2
4
6
0.15
0.25
0.35
0.45
estimativas


n = 100 , p−valor (teste−SW) = 0.0056
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
3
quantil da N(0,1)


n = 1000 , p−valor (teste−SW) = 0.5187
densidade
0
5
10
15
20
25
0.30
0.34
0.38
estimativas


n = 1000 , p−valor (teste−SW) = 0.5187
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# p=0,50

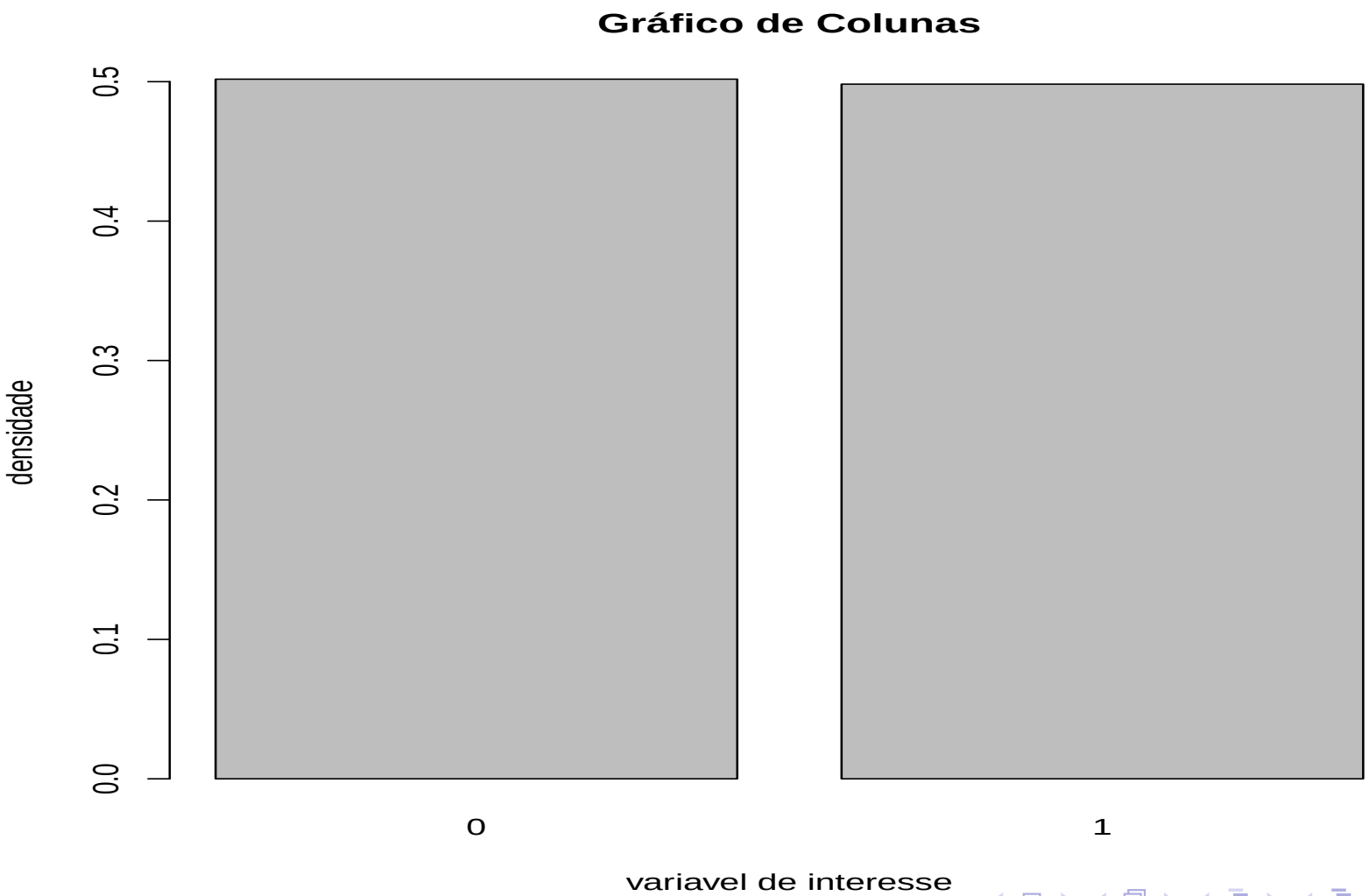
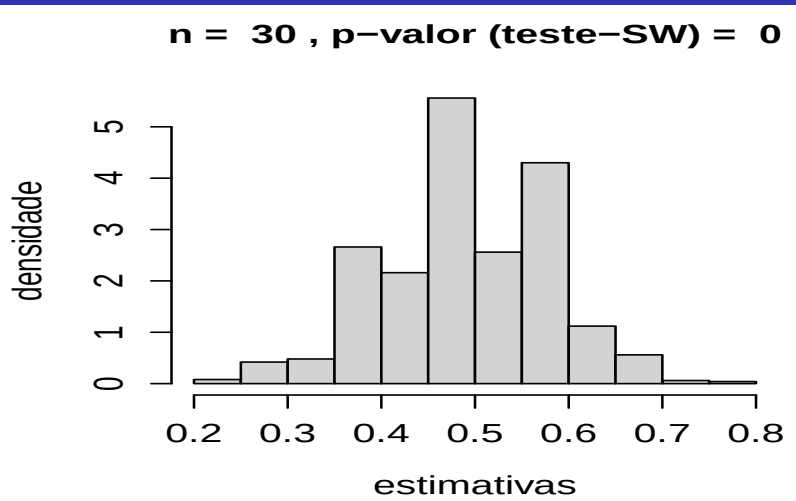
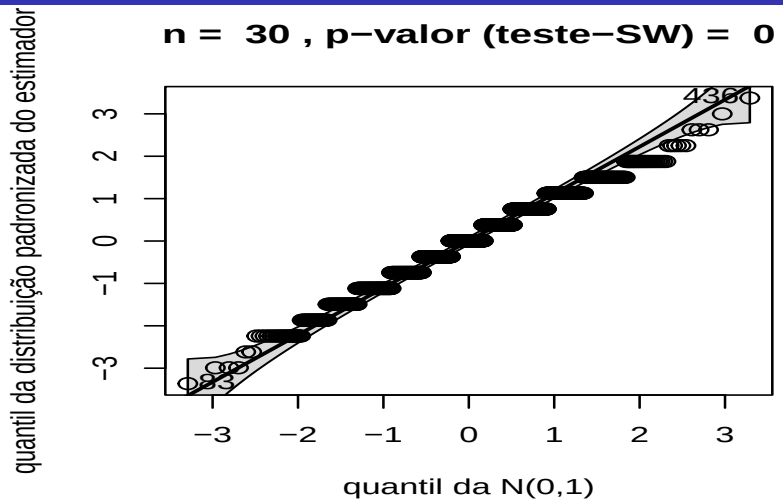
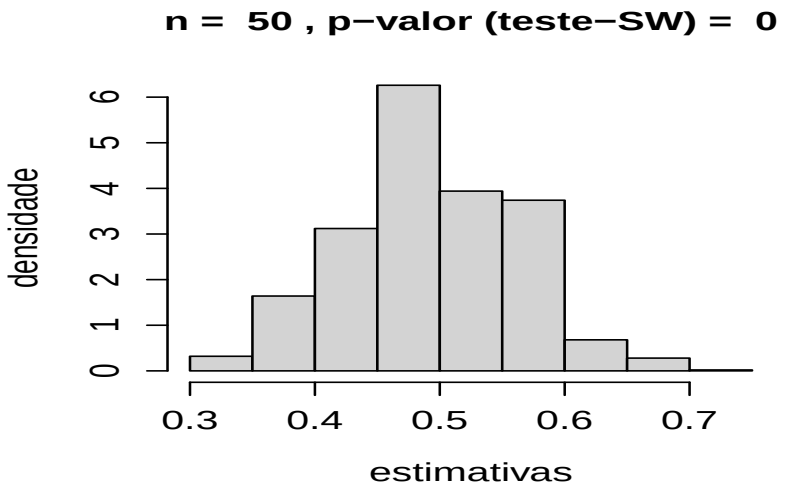


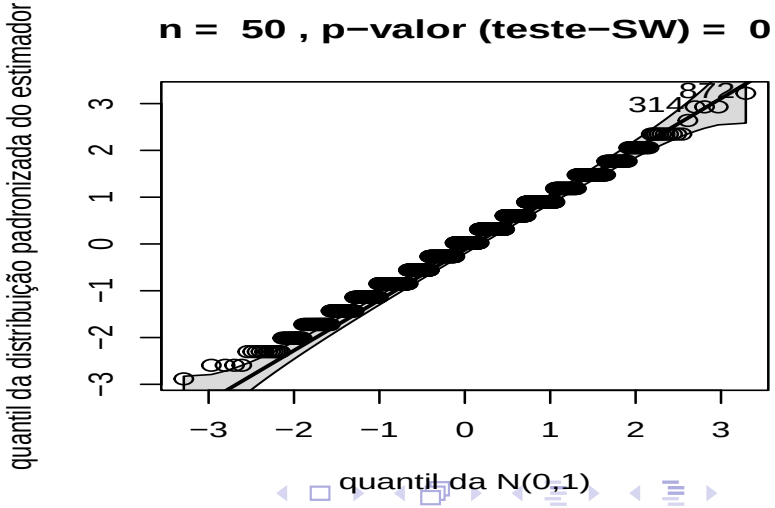
Gráfico de Colunas
densidade
0.0
0.1
0.2
0.3
0.4
0.5
0
1
variavel de interesse

# p=0,50

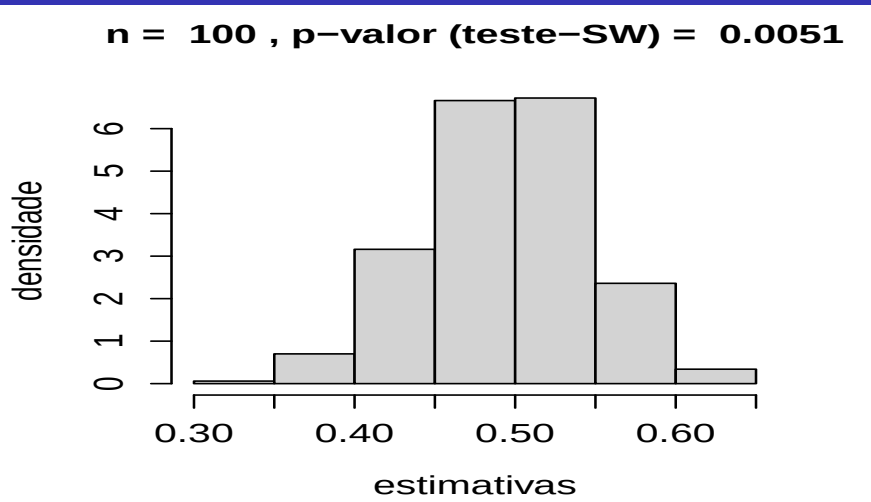

n = 30 , p−valor (teste−SW) = 0
densidade
0 1 2 3 4 5
0.2 0.3 0.4 0.5 0.6 0.7 0.8
estimativas

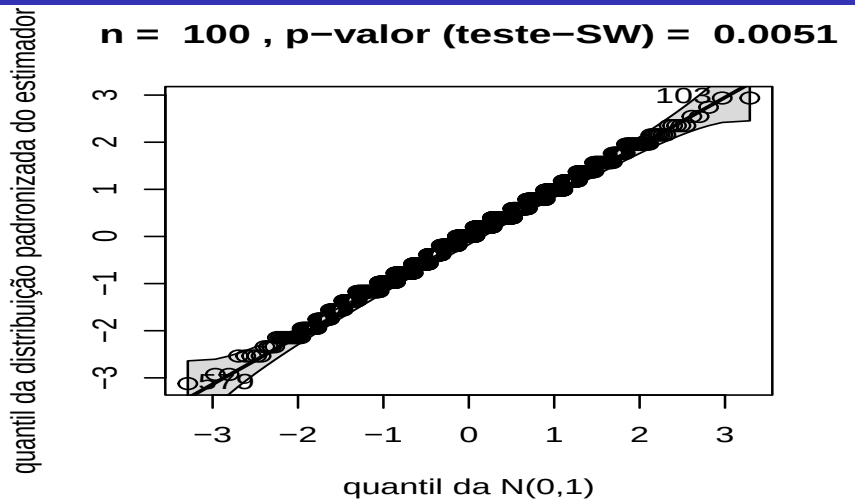

n = 30 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3 −1 0 1 2 3
−3 −2 −1 0 1 2 3
quantil da N(0,1)

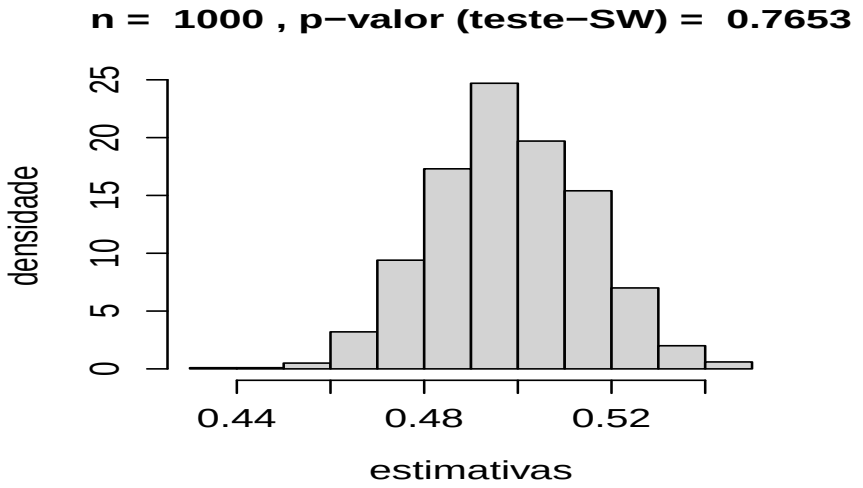

n = 50 , p−valor (teste−SW) = 0
densidade
0 1 2 3 4 5 6
0.3 0.4 0.5 0.6 0.7
estimativas

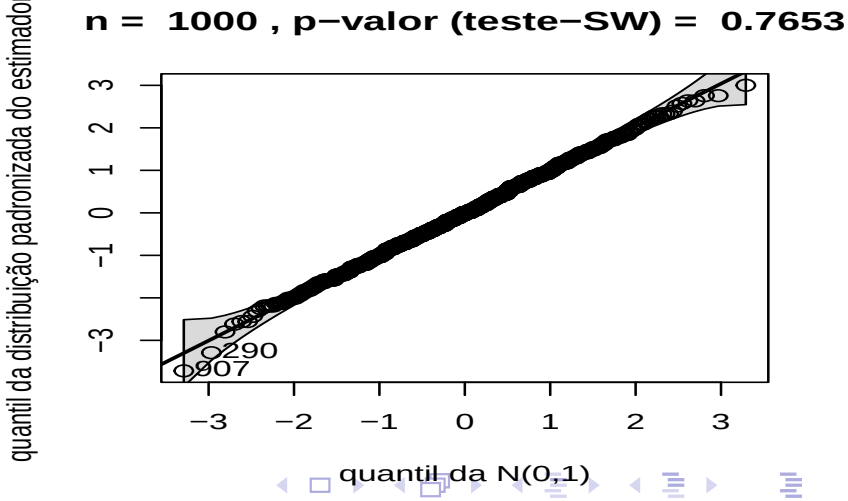

n = 50 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3 −2 −1 0 1 2 3
−3 −2 −1 0 1 2 3
quantil da N(0,1)

# p=0,50


n = 100 , p−valor (teste−SW) = 0.0051
densidade
0 1 2 3 4 5 6
0.30 0.40 0.50 0.60
estimativas


n = 100 , p−valor (teste−SW) = 0.0051
quantil da distribuição padronizada do estimador
−3 −2 −1 0 1 2 3
103
579
−3 −2 −1 0 1 2 3
quantil da N(0,1)


n = 1000 , p−valor (teste−SW) = 0.7653
densidade
0 5 10 15 20 25
0.44 0.48 0.52
estimativas


n = 1000 , p−valor (teste−SW) = 0.7653
quantil da distribuição padronizada do estimador
−3 −1 0 1 2 3
290
907
−3 −2 −1 0 1 2 3
quantil da N(0,1)

# p=0,65

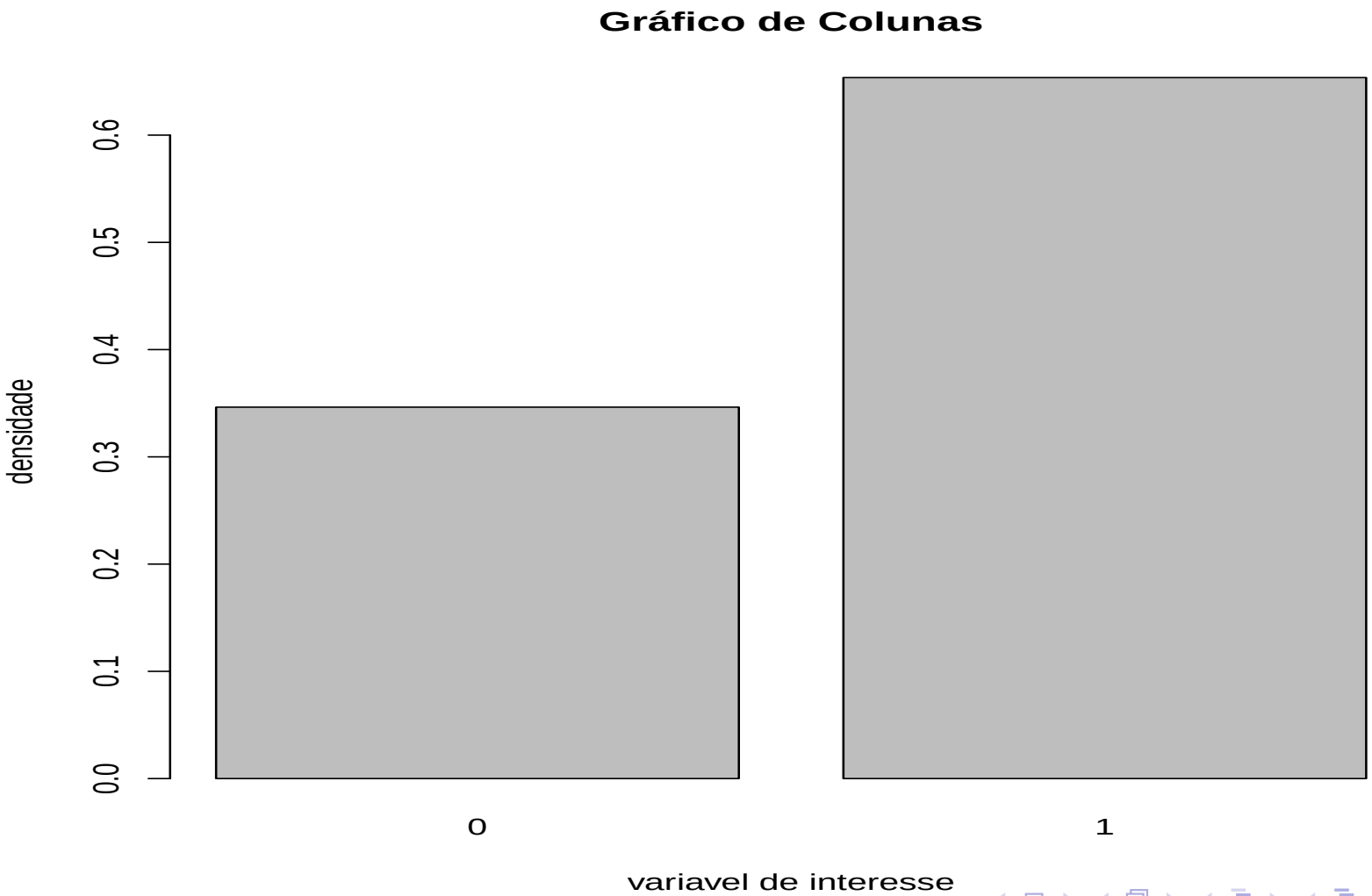
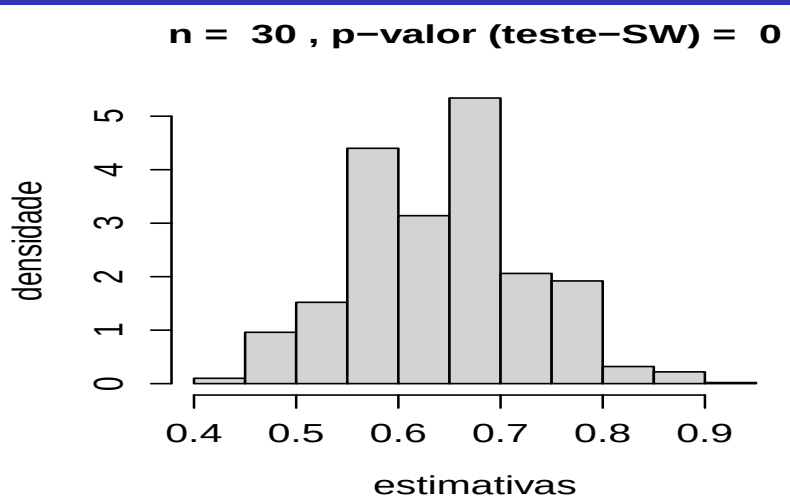
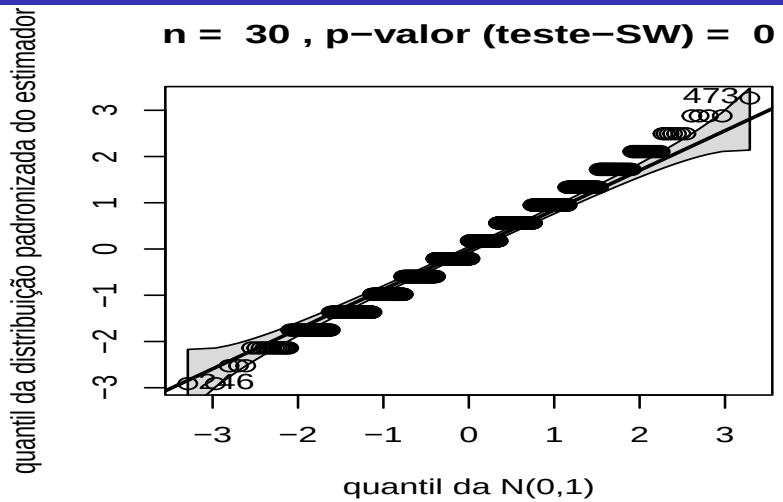
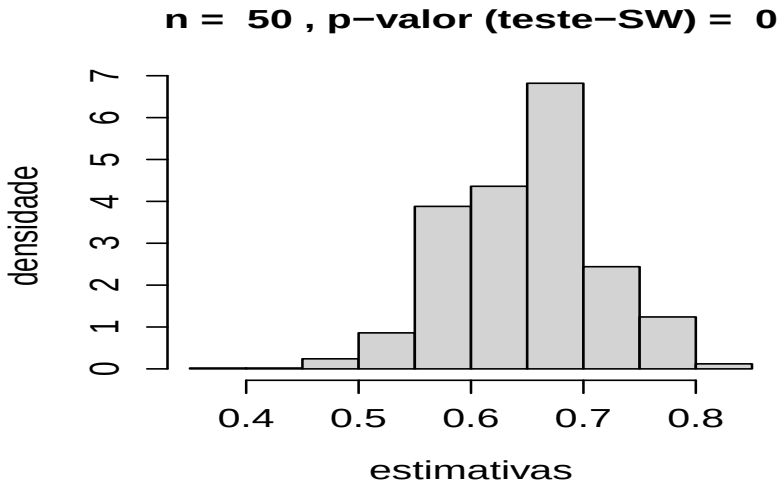


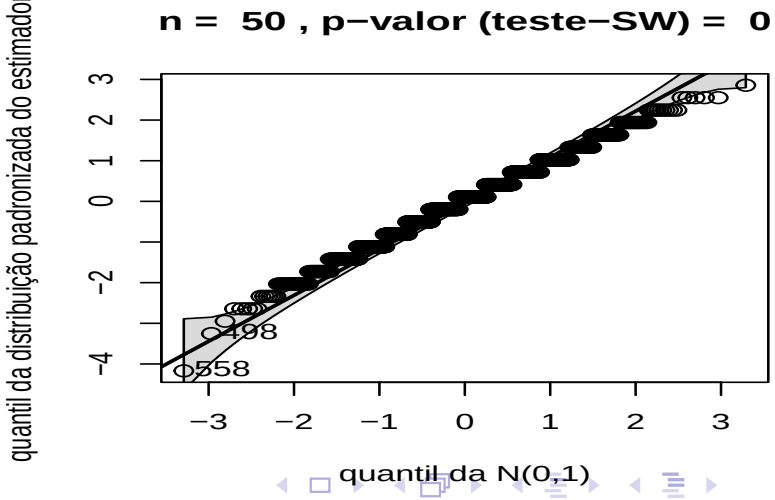
Gráfico de Colunas
densidade
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
1
variavel de interesse

# p=0,65

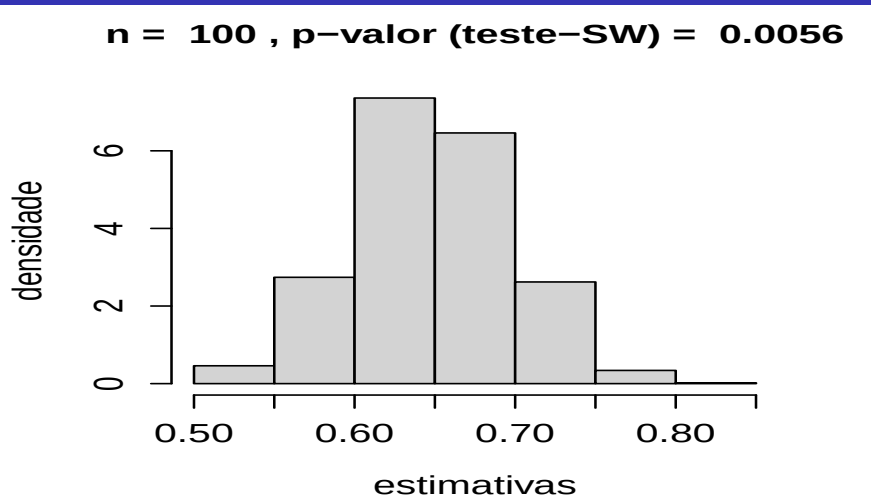

n = 30 , p−valor (teste−SW) = 0
densidade
0 1 2 3 4 5
0.4 0.5 0.6 0.7 0.8 0.9
estimativas

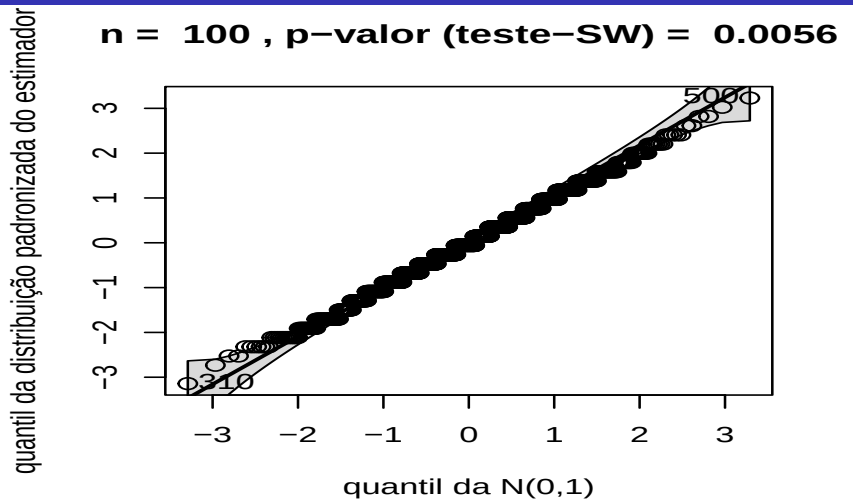

n = 30 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3 −2 −1 0 1 2 3
473
246
−3 −2 −1 0 1 2 3
quantil da N(0,1)

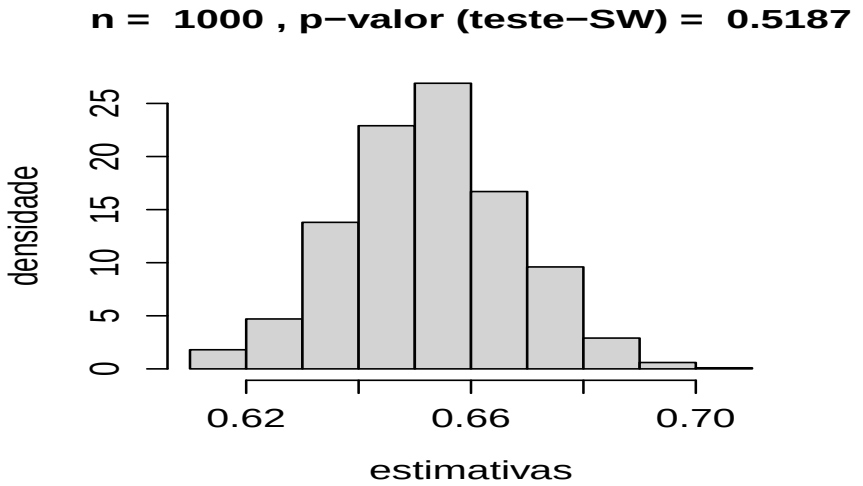

n = 50 , p−valor (teste−SW) = 0
densidade
0 1 2 3 4 5 6 7
0.4 0.5 0.6 0.7 0.8
estimativas

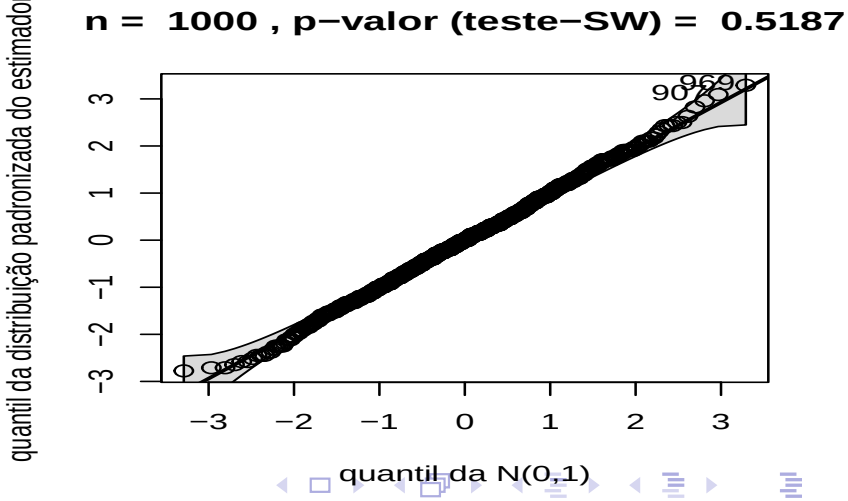

n = 50 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−4 −2 0 1 2 3
498
558
−3 −2 −1 0 1 2 3
quantil da N(0,1)

# p=0,65


n = 100 , p−valor (teste−SW) = 0.0056
densidade
0
2
4
6
0.50
0.60
0.70
0.80
estimativas


n = 100 , p−valor (teste−SW) = 0.0056
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
3
quantil da N(0,1)


n = 1000 , p−valor (teste−SW) = 0.5187
densidade
0
5
10
15
20
25
0.62
0.66
0.70
estimativas


n = 1000 , p−valor (teste−SW) = 0.5187
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# p=0,75

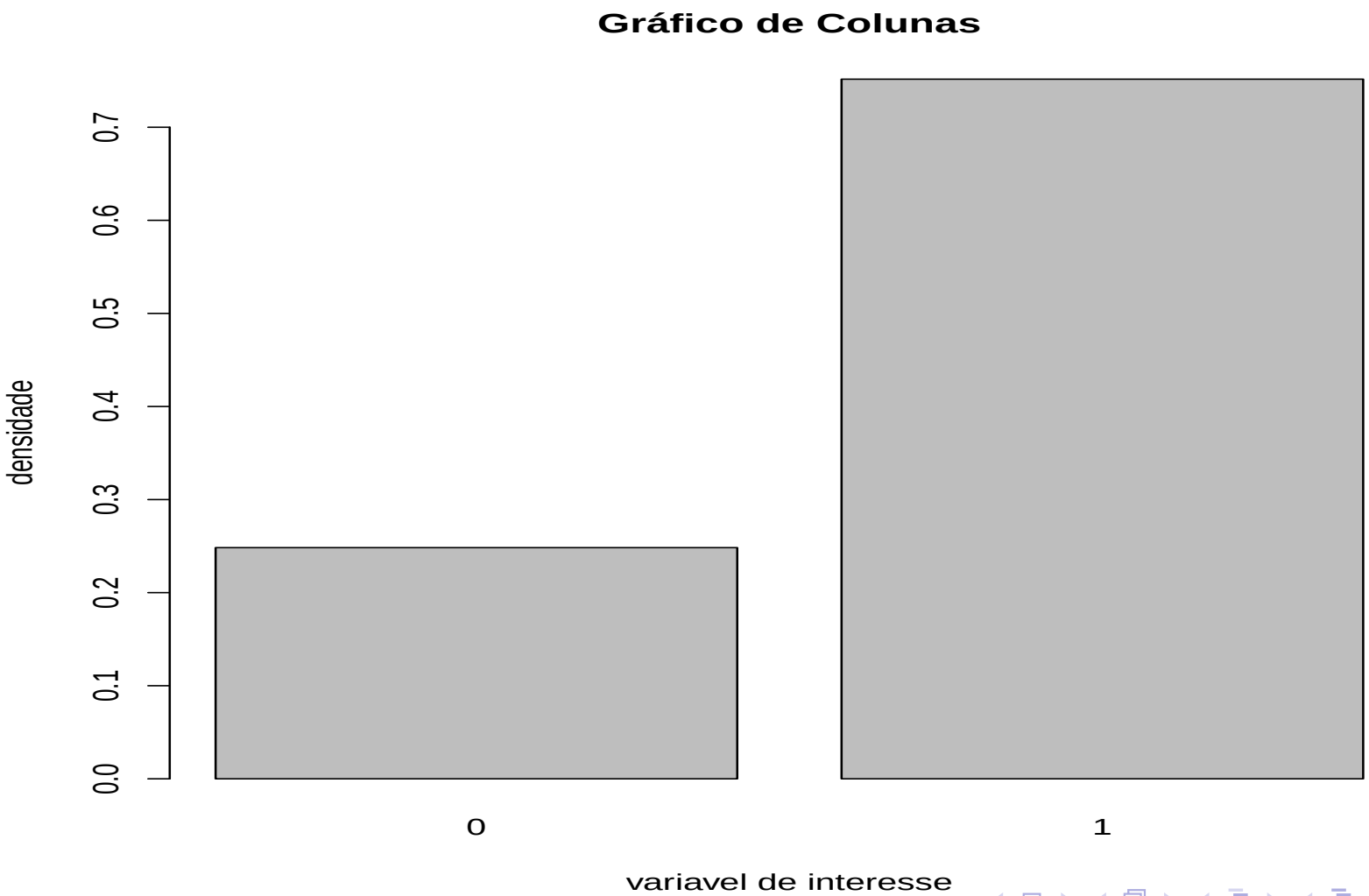
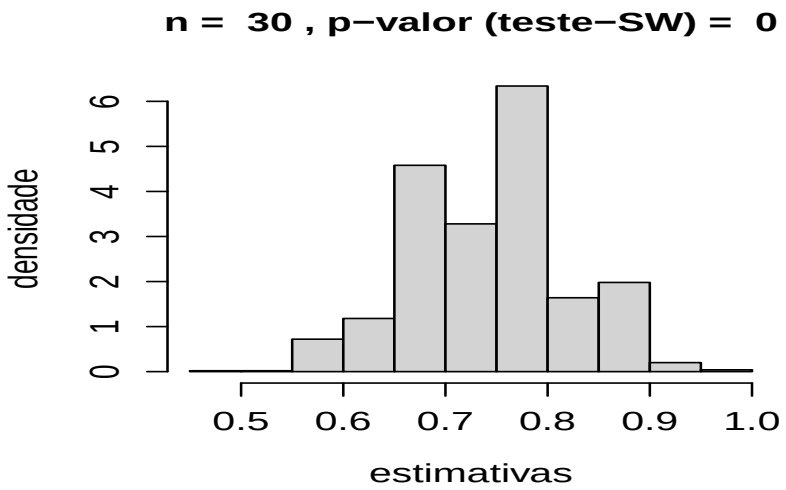
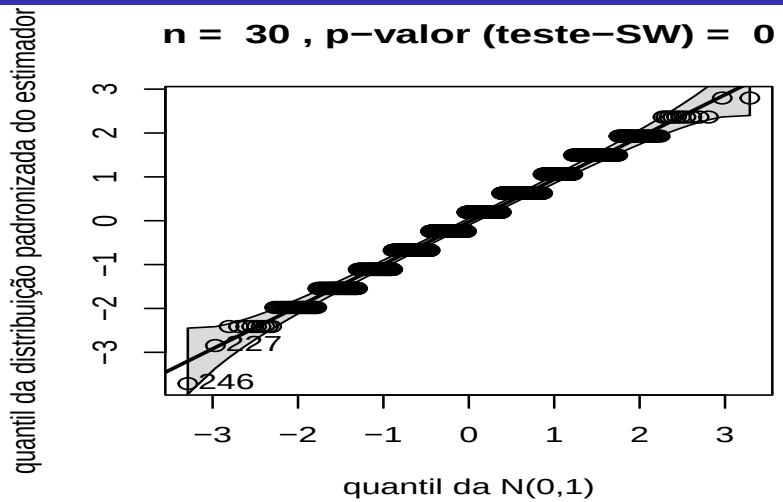
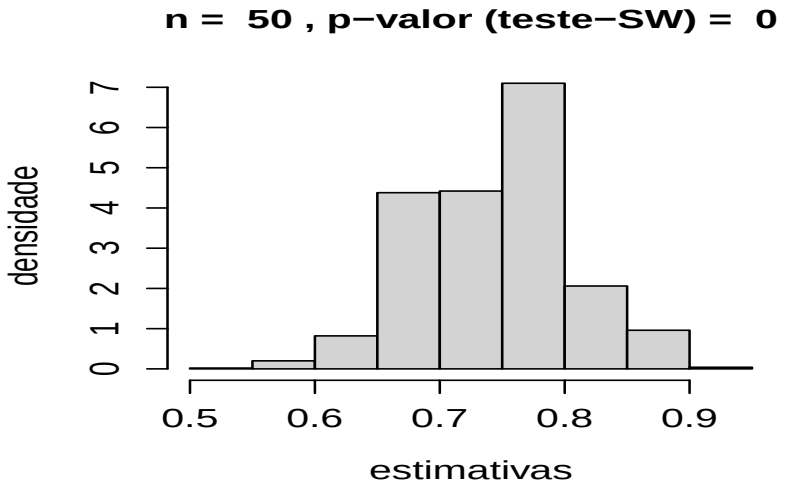


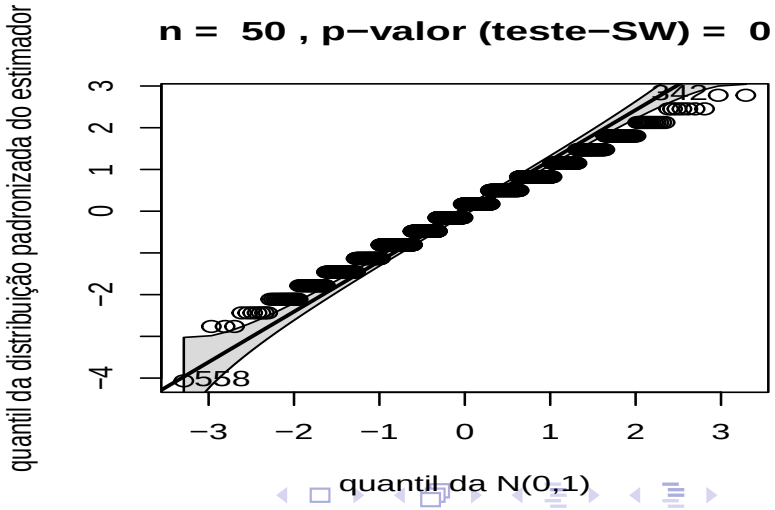
Gráfico de Colunas
densidade
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1
variavel de interesse

# p=0,75

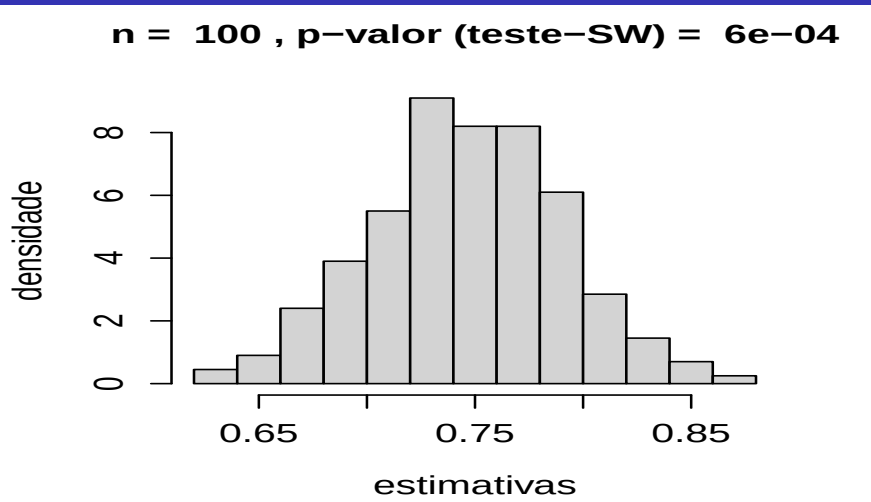

n = 30 , p−valor (teste−SW) = 0
densidade
0 1 2 3 4 5 6
0.5 0.6 0.7 0.8 0.9 1.0
estimativas

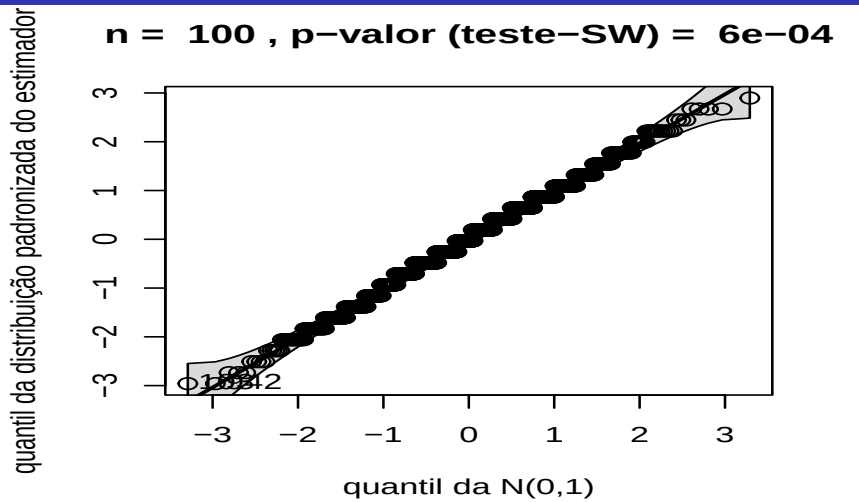

n = 30 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3 −2 −1 0 1 2 3
227
246
−3 −2 −1 0 1 2 3
quantil da N(0,1)

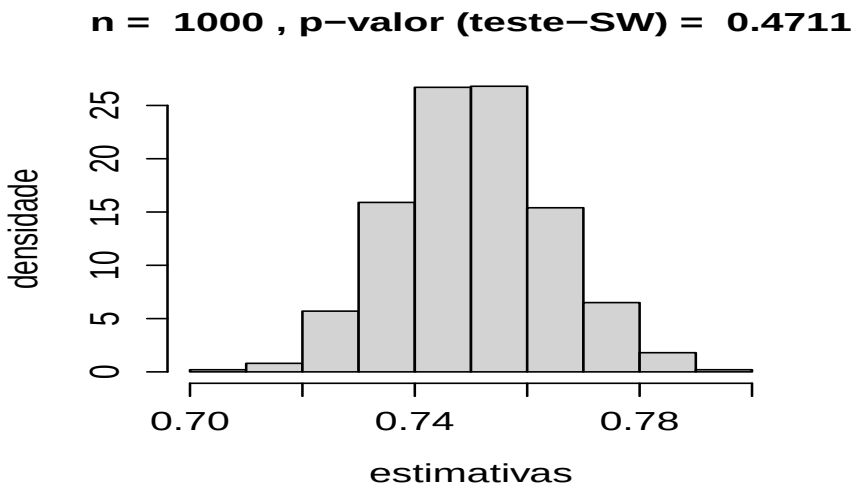

n = 50 , p−valor (teste−SW) = 0
densidade
0 1 2 3 4 5 6 7
0.5 0.6 0.7 0.8 0.9
estimativas

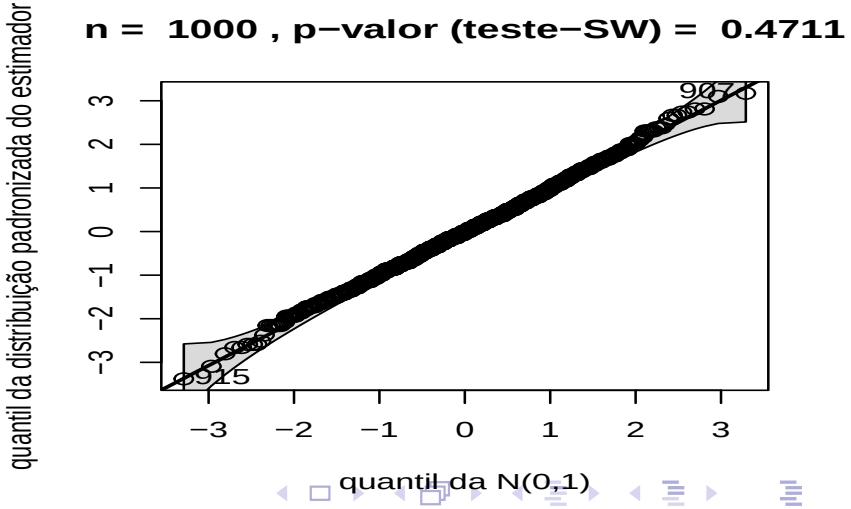

n = 50 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−4 −2 0 1 2 3
342
558
−3 −2 −1 0 1 2 3
quantil da N(0,1)

# p=0,75


n = 100 , p−valor (teste−SW) = 6e−04
densidade
0
2
4
6
8
0.65
0.75
0.85
estimativas


n = 100 , p−valor (teste−SW) = 6e−04
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
3
quantil da N(0,1)


n = 1000 , p−valor (teste−SW) = 0.4711
densidade
0
5
10
15
20
25
0.70
0.74
0.78
estimativas


n = 1000 , p−valor (teste−SW) = 0.4711
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# p=0,90

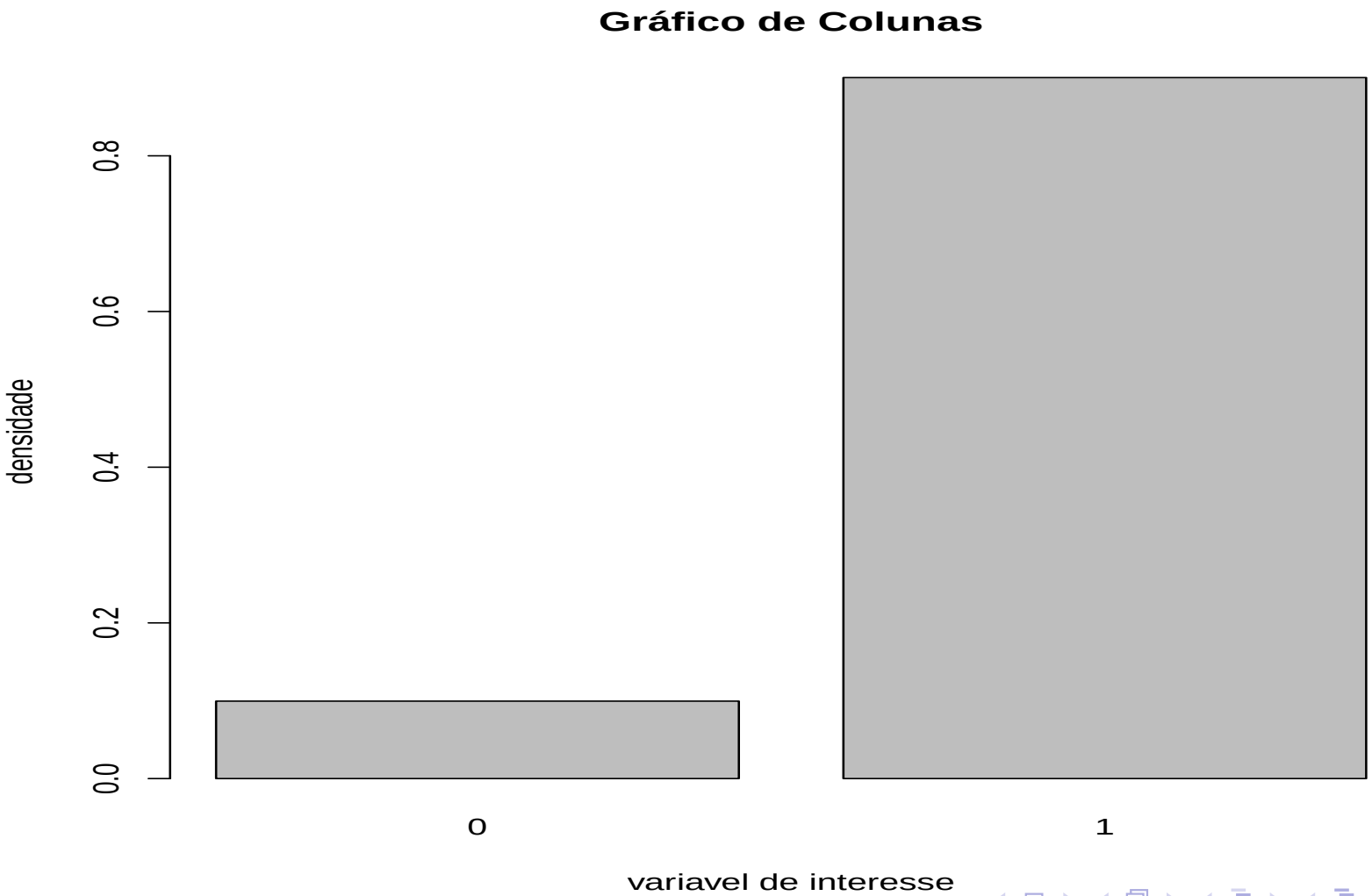
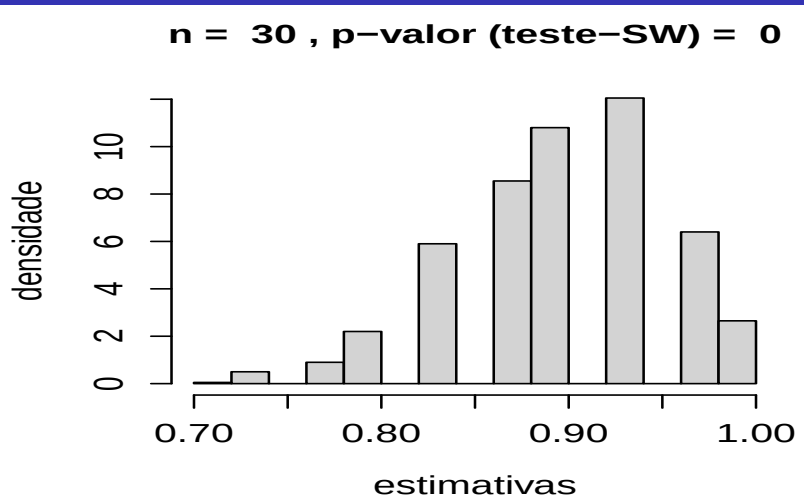
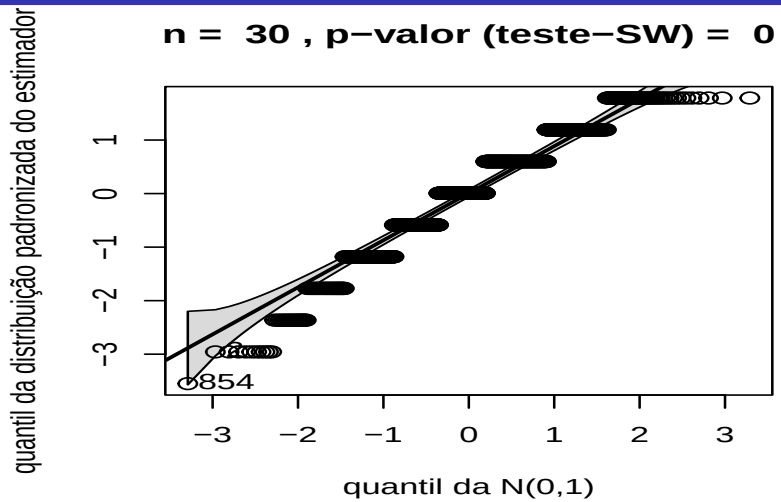
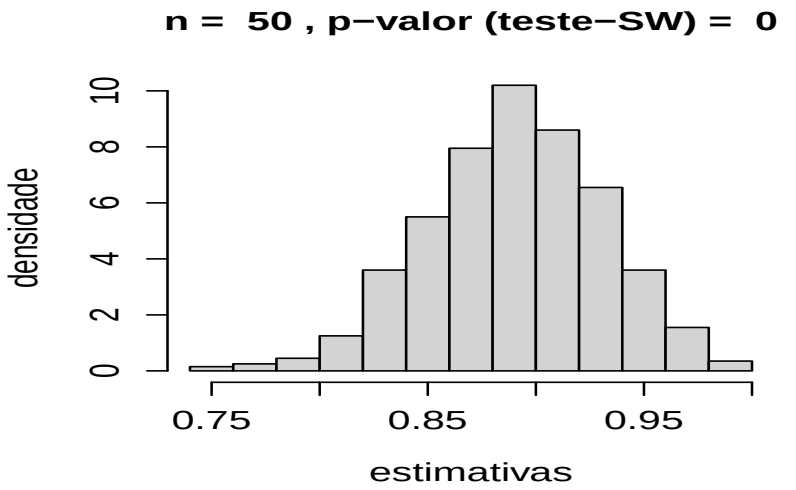


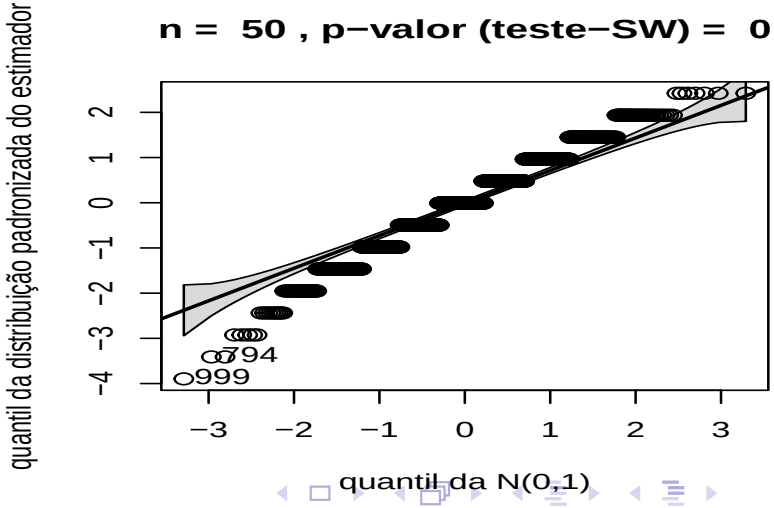
Gráfico de Colunas
densidade
0.0
0.2
0.4
0.6
0.8
0
1
variavel de interesse

# p=0,90

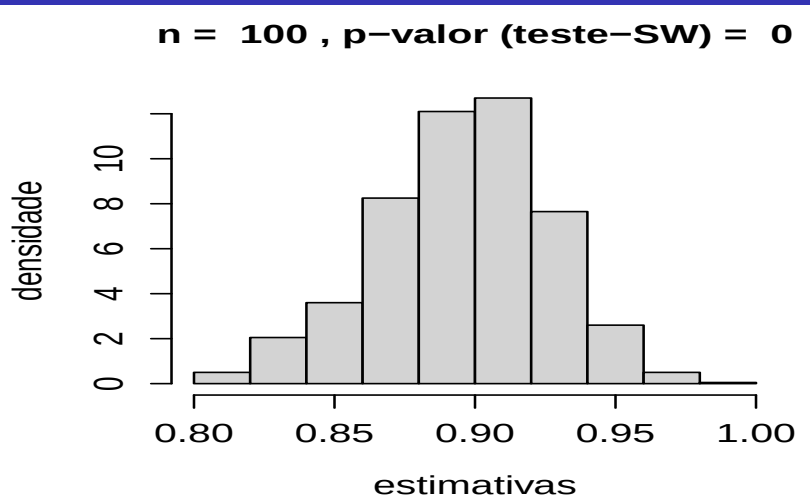

n = 30 , p−valor (teste−SW) = 0
densidade
0 2 4 6 8 10
0.70 0.80 0.90 1.00
estimativas

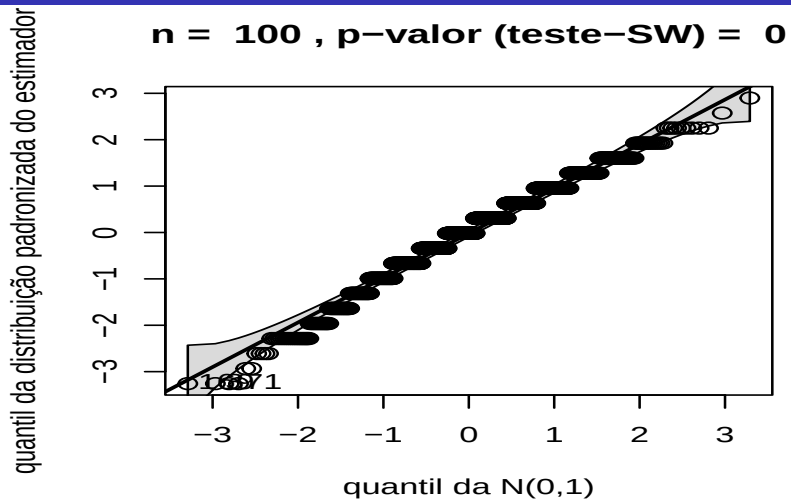

n = 30 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3 −2 −1 0 1
854
−3 −2 −1 0 1 2 3
quantil da N(0,1)

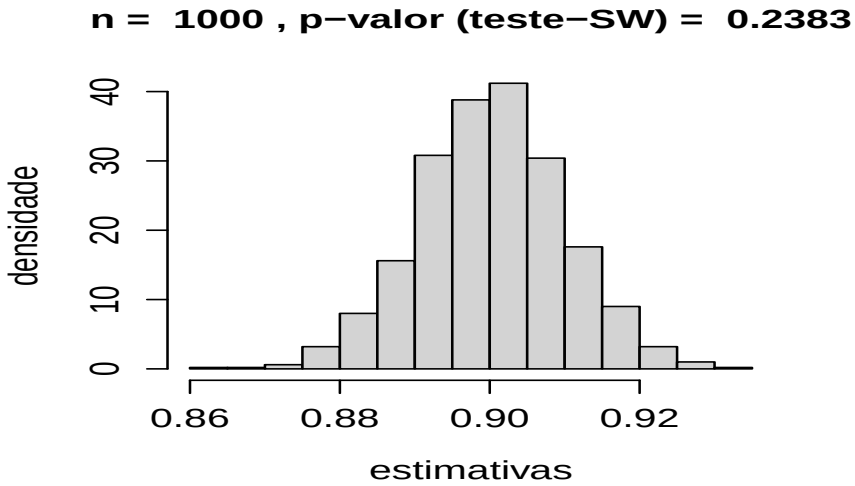

n = 50 , p−valor (teste−SW) = 0
densidade
0 2 4 6 8 10
0.75 0.85 0.95
estimativas

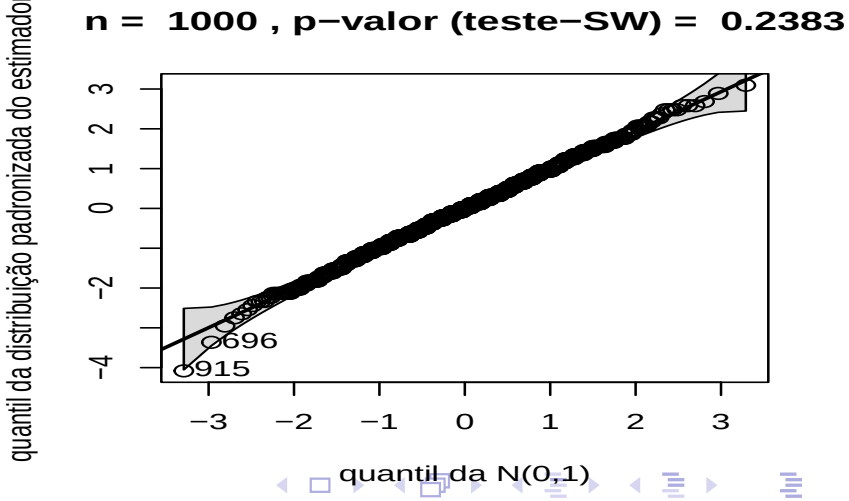

n = 50 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−4 −3 −2 −1 0 1 2
794
999
−3 −2 −1 0 1 2 3
quantil da N(0,1)

# p=0,90


n = 100 , p−valor (teste−SW) = 0
densidade
0 2 4 6 8 10
0.80 0.85 0.90 0.95 1.00
estimativas


n = 100 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3 −2 −1 0 1 2 3
−3 −2 −1 0 1 2 3
quantil da N(0,1)


n = 1000 , p−valor (teste−SW) = 0.2383
densidade
0 10 20 30 40
0.86 0.88 0.90 0.92
estimativas


n = 1000 , p−valor (teste−SW) = 0.2383
quantil da distribuição padronizada do estimador
−4 −2 0 1 2 3
696
915
−3 −2 −1 0 1 2 3
quantil da N(0,1)

# p=0,95

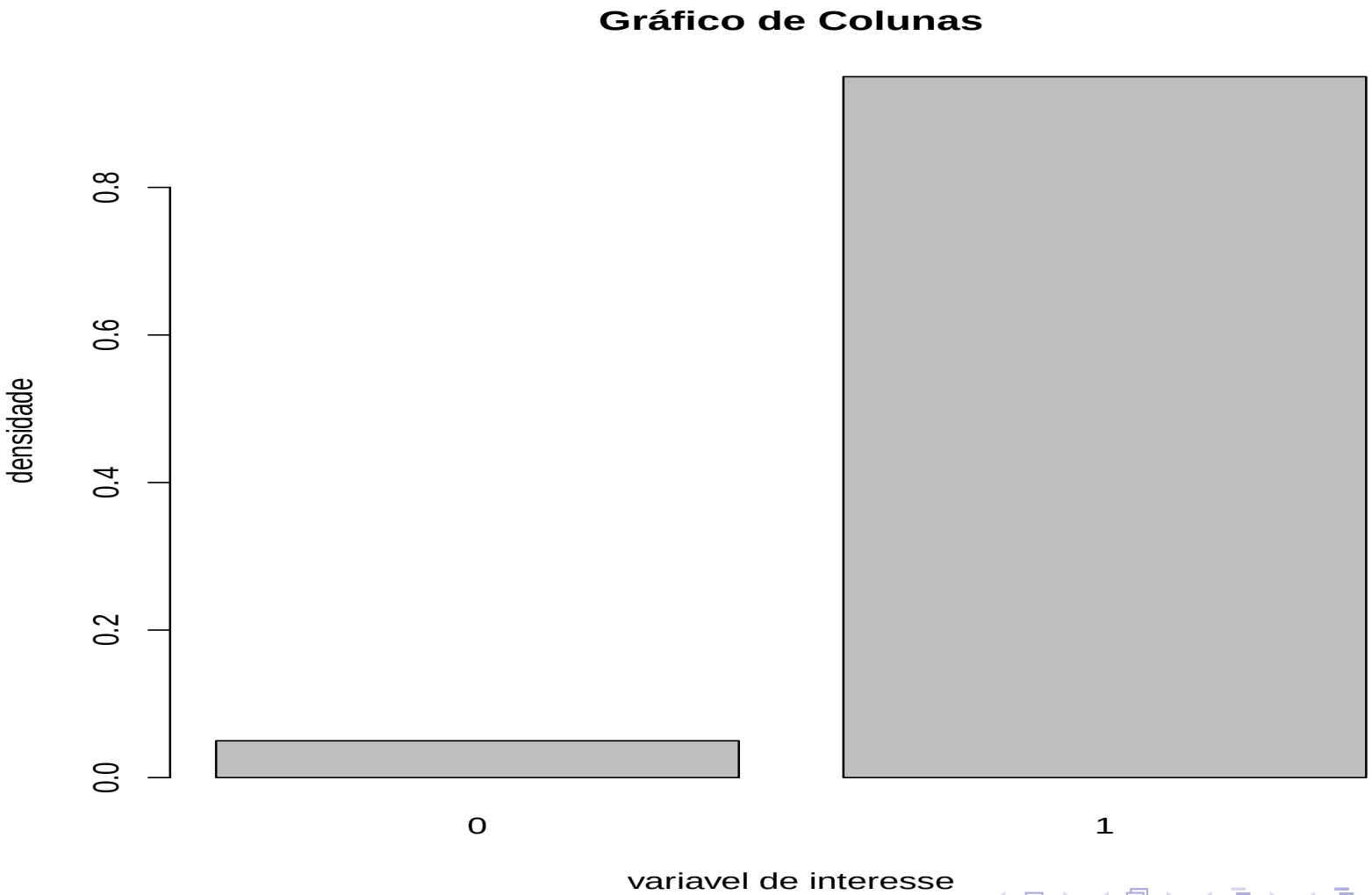
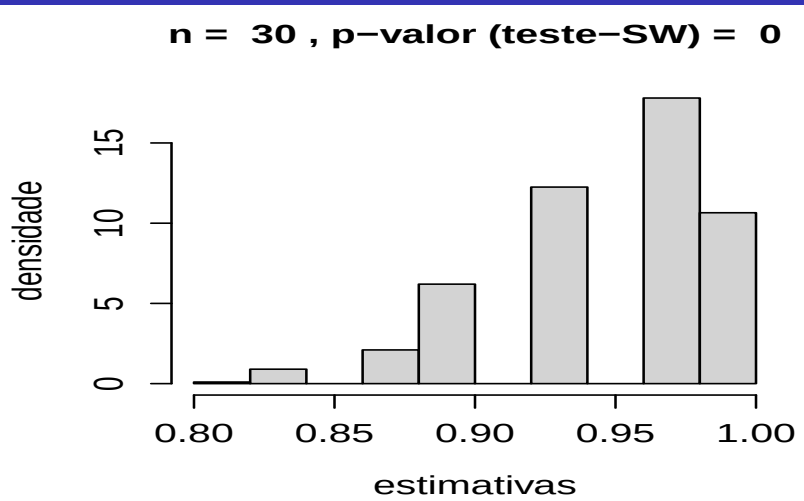
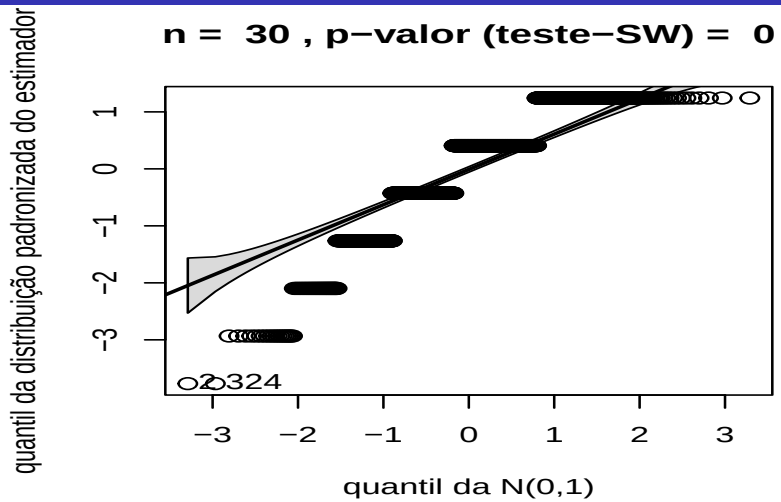
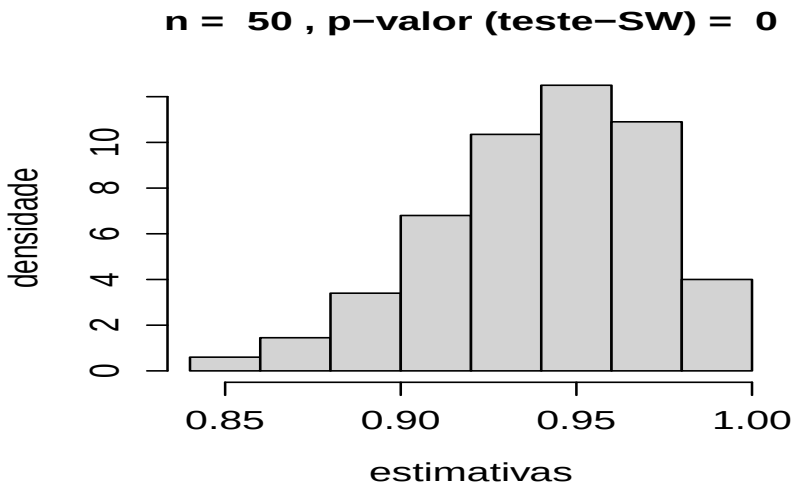


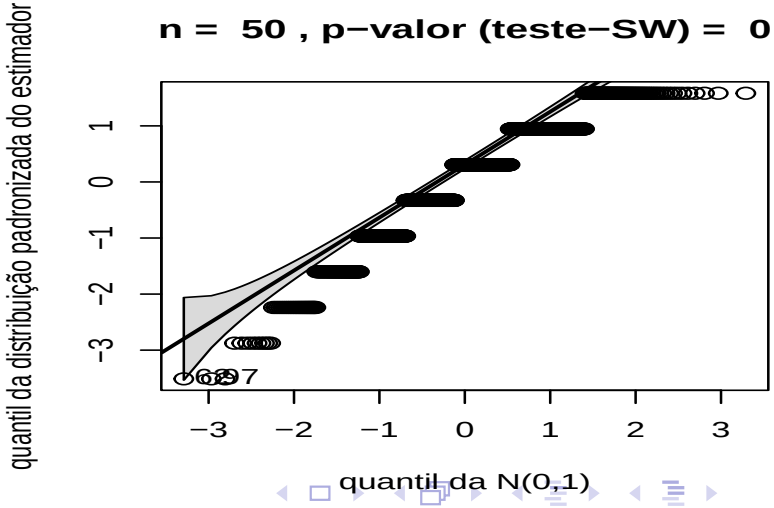
Gráfico de Colunas
densidade
0.0
0.2
0.4
0.6
0.8
0
1
variavel de interesse

# p=0,95

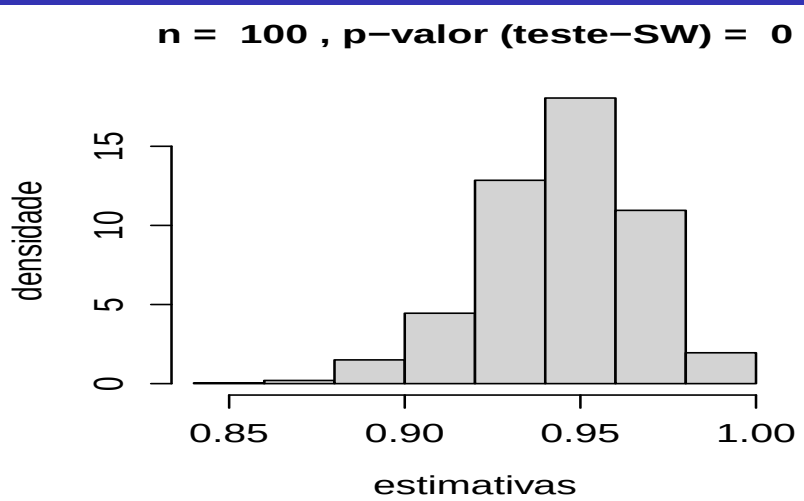

n = 30 , p−valor (teste−SW) = 0
densidade
0
5
10
15
0.80
0.85
0.90
0.95
1.00
estimativas

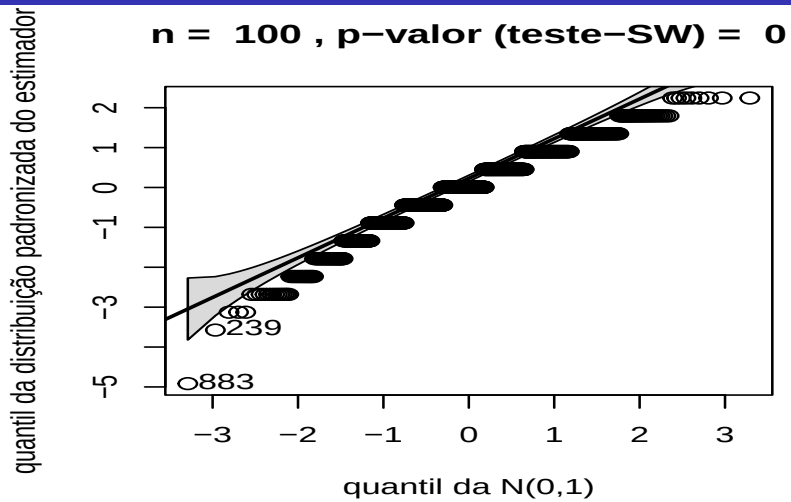

n = 30 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
3
quantil da N(0,1)

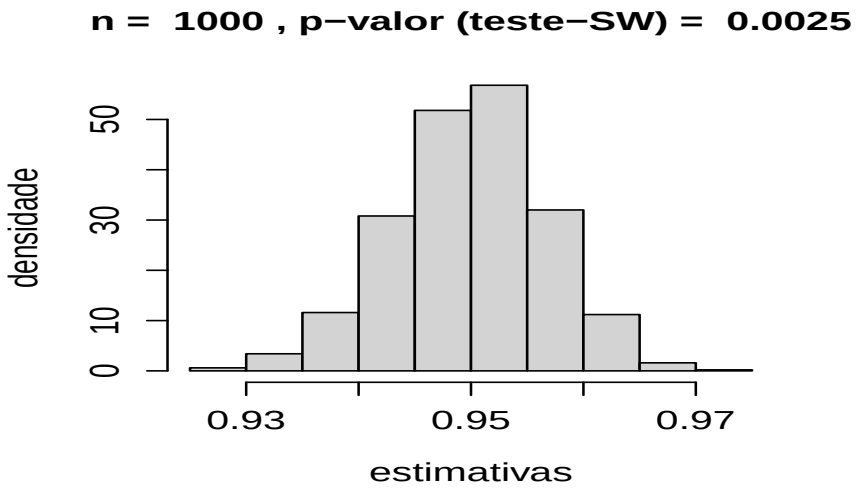

n = 50 , p−valor (teste−SW) = 0
densidade
0
2
4
6
8
10
0.85
0.90
0.95
1.00
estimativas

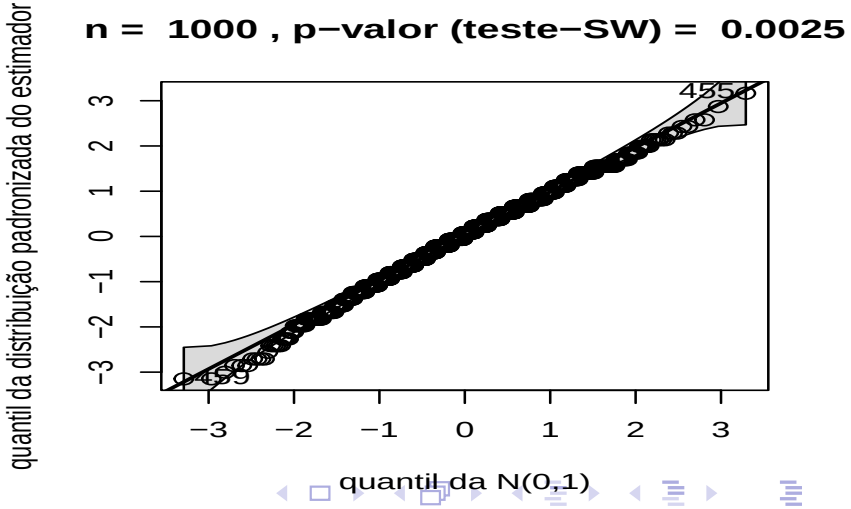

n = 50 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# p=0,95


n = 100 , p−valor (teste−SW) = 0
densidade
0
5
10
15
0.85
0.90
0.95
1.00
estimativas


n = 100 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−5
−3
−1
0
1
2
239
883
−3
−2
−1
0
1
2
3
quantil da N(0,1)


n = 1000 , p−valor (teste−SW) = 0.0025
densidade
0
10
30
50
0.93
0.95
0.97
estimativas


n = 1000 , p−valor (teste−SW) = 0.0025
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
3
4554
459
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# p=0,99

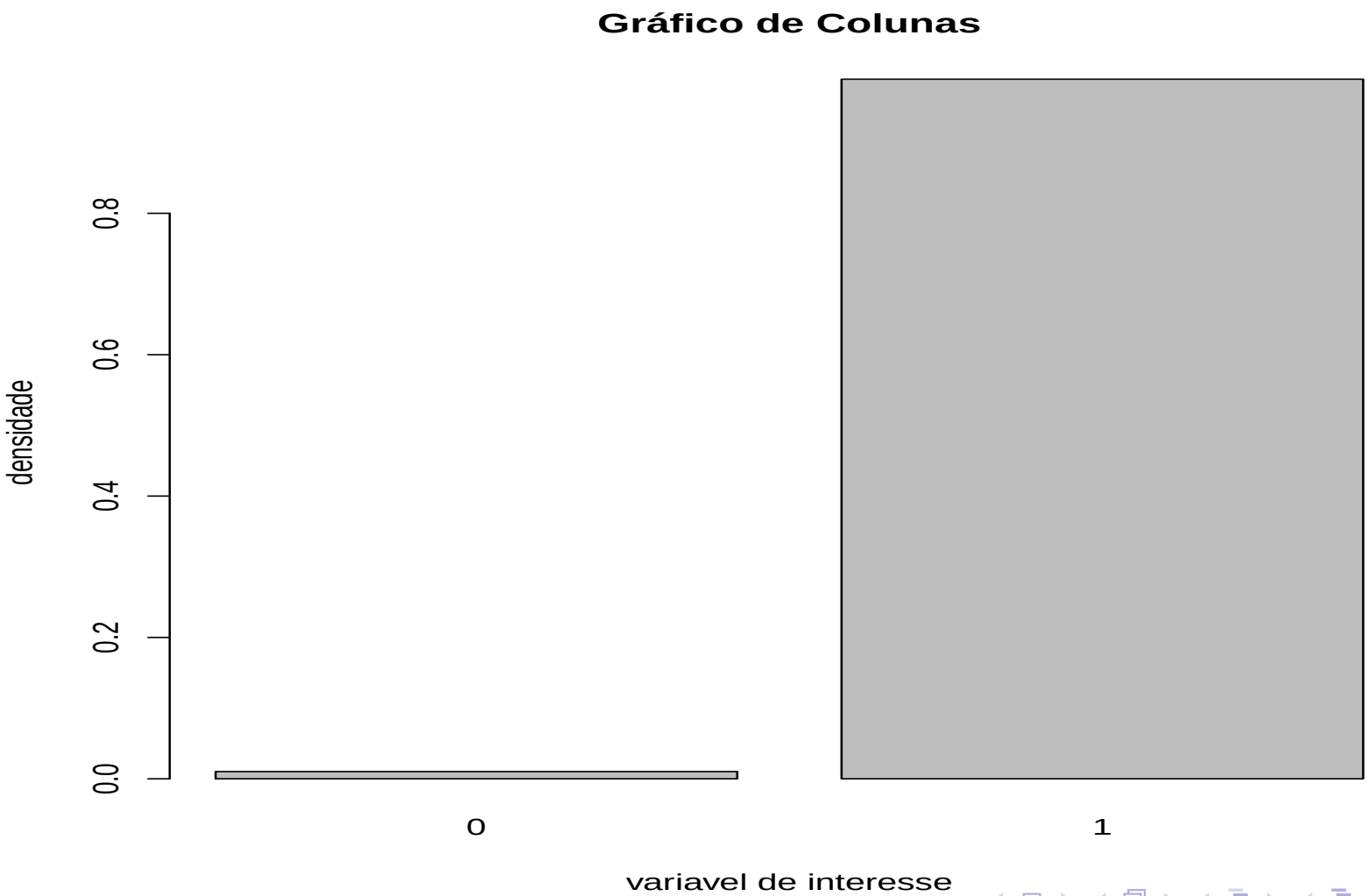
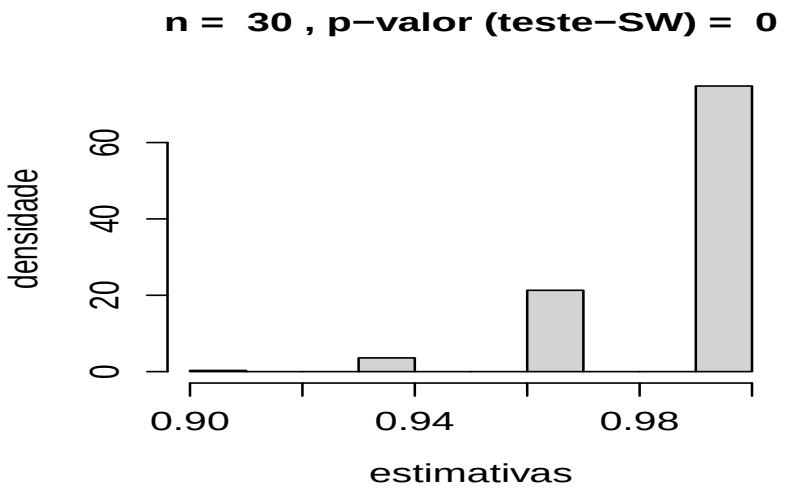
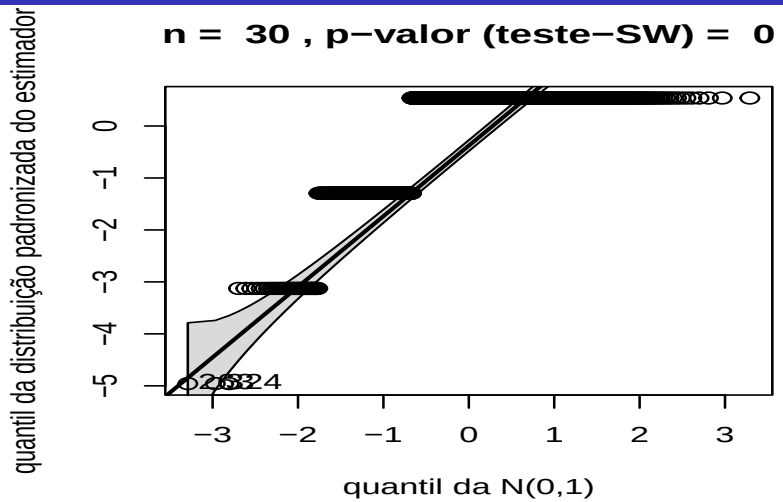
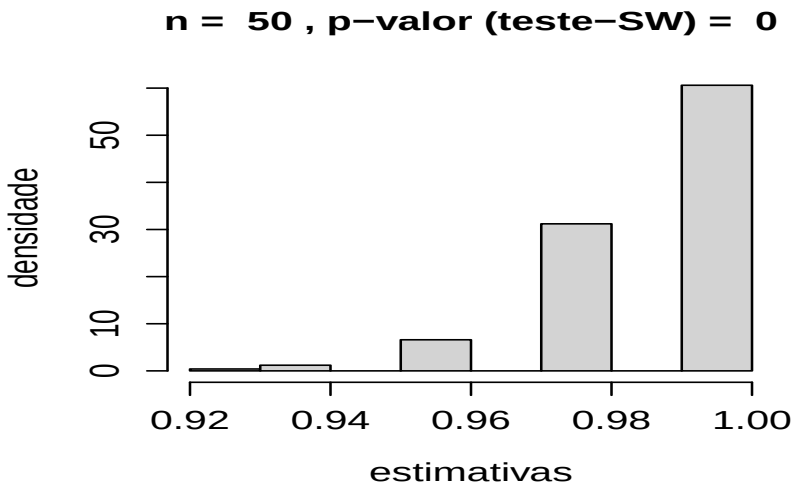


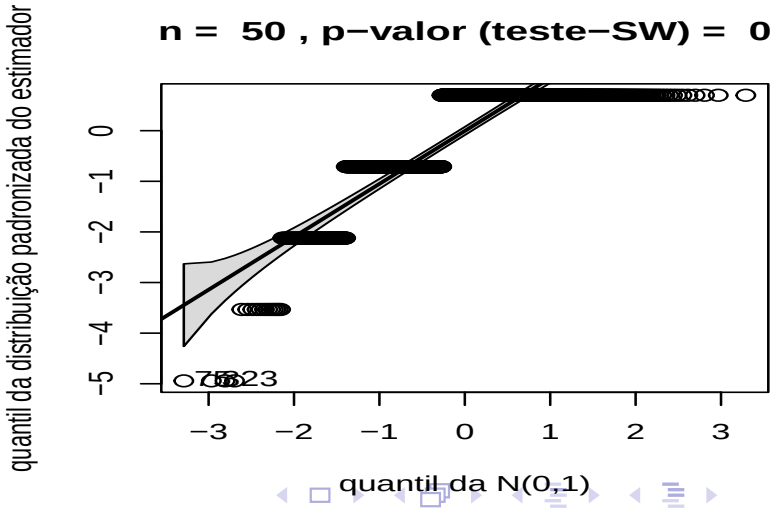
Gráfico de Colunas
densidade
0.0
0.2
0.4
0.6
0.8
0
1
variavel de interesse

# p=0,99

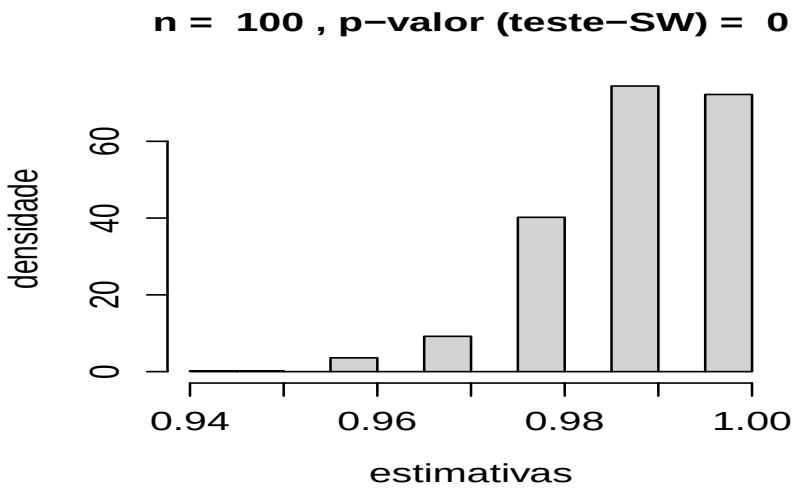

n = 30 , p−valor (teste−SW) = 0
densidade
0
20
40
60
0.90
0.94
0.98
estimativas

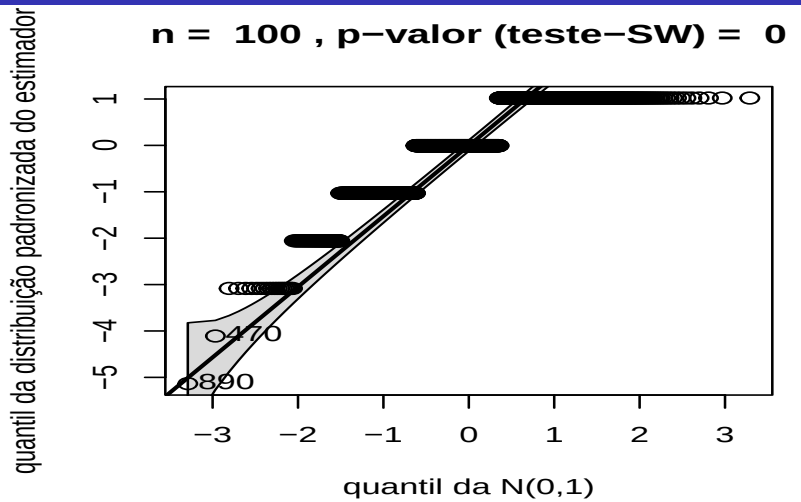

n = 30 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
0
−1
−2
−3
−4
−5
−3
−2
−1
0
1
2
3
quantil da N(0,1)

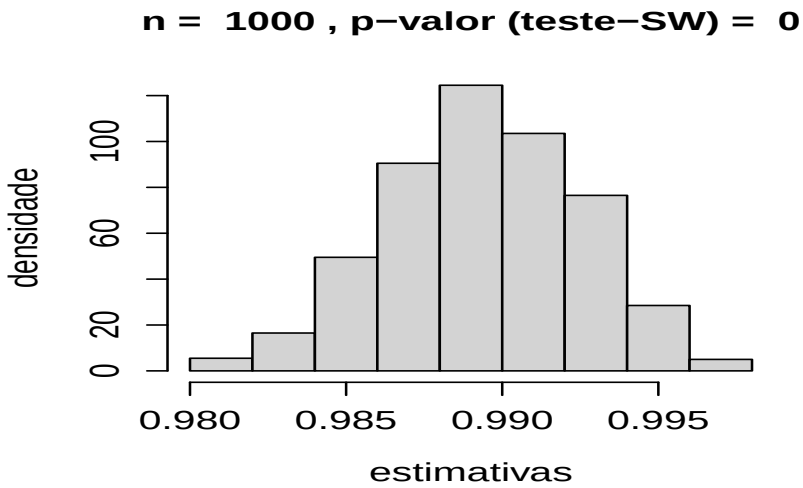

n = 50 , p−valor (teste−SW) = 0
densidade
0
10
30
50
0.92
0.94
0.96
0.98
1.00
estimativas

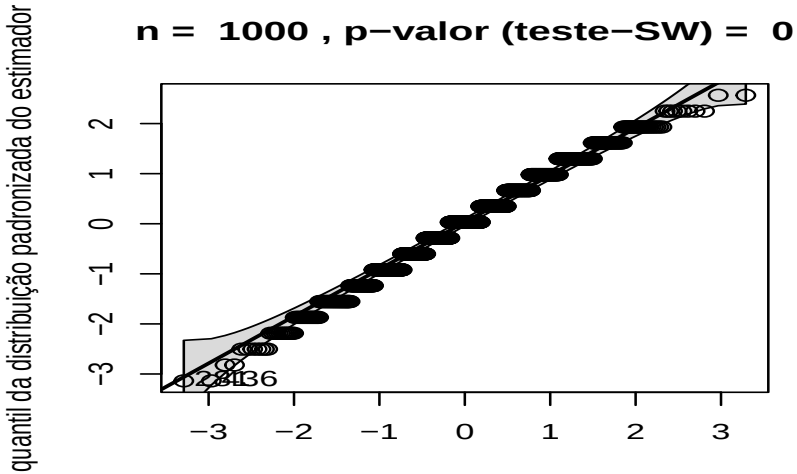

n = 50 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
0
−1
−2
−3
−4
−5
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# p=0,99


n = 100 , p−valor (teste−SW) = 0
densidade
0
20
40
60
0.94
0.96
0.98
1.00
estimativas


n = 100 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−5
−4
−3
−2
−1
0
1
−3
−2
−1
0
1
2
3
quantil da N(0,1)
470
890


n = 1000 , p−valor (teste−SW) = 0
densidade
0
20
60
100
0.980
0.985
0.990
0.995
estimativas


n = 1000 , p−valor (teste−SW) = 0
quantil da distribuição padronizada do estimador
−3
−2
−1
0
1
2
−3
−2
−1
0
1
2
3
quantil da N(0,1)

# Otimalidade dos estimadores

- Vamos nos concentrar na média amostral e na classe de estimadores não viciados que sejam combinações lineares das variáveis aleatórias $(Y_1, ...., Y_n)$.
- Os resultados para os outros parâmetros são análogos.
- A forma geral do estimador em questão é dada por

$$\widehat{\mu}_{sc} = \sum_{i=1}^{n} c_i Y_i$$

- Note que, sob $AAS_s(A_2)$ temos que os $Y_i's$ não são mais independentes (embora ainda sejam "identicamente distribuídas" (pelo menos em relação a alguns momentos)) Exercício 2.9, livro: "Elementos de Amostragem"). Temos que $\mathcal{E}(Y_i) = \mu$, $\mathcal{V}(Y_i) = \frac{N-1}{N} s^2$, $Cov(Y_i, Y_j) = -\frac{s^2}{N}, \forall i \neq j$.

# Otimalidade dos estimadores

- Note que $\mathcal{E}(\widehat{\mu}_{sc}) = \sum_{i=1}^{n} c_i \mathcal{E}(Y_i) = \mu \sum_{i=1}^{n} c_i$.
- Exercício: provar que $\widehat{\mu}_{sc}$ é um estimador não viciado se e somente se

$$\sum_{i=1}^{n} c_i = 1 \tag{5}$$

- Defina $\overline{c} = \frac{1}{n} \sum_{i=1}^{n} c_i$. Pode-se provar (usando (5)) que $\sum_{\substack{i \neq j \\ i,j=1,2,\ldots,N}} c_i c_j = 1 - \sum_{i=1}^{n} c_i^2$).

# Otimalidade dos estimadores

- Assim, temos que:

$$\begin{aligned}
\mathcal{V}(\widehat{\mu}_{sc}) &= \sum_{i=1}^{n} c_i^2 \mathcal{V}(Y_i) + \sum_{\substack{i \neq j \\ i,j=1,2,\ldots,N}} Cov(c_i Y_i, c_j Y_j) \\
&= \frac{N-1}{N} s^2 \sum_{i=1}^{n} c_i^2 - \frac{s^2}{N} \sum_{\substack{i \neq j \\ i,j=1,2,\ldots,N}} c_i c_j \\
&= s^2 \left( \sum_{i=1}^{n} c_i^2 - \frac{1}{N} \right)
\end{aligned}$$

- Devemos então minimizar $g(\boldsymbol{c}) = \sum_{i=1}^{n} c_i^2 - \frac{1}{N} + \lambda \left( \sum_{i=1}^{n} c_i - 1 \right)$, em $c_i = 1, 2, ..., n$, $\boldsymbol{c} = (c_1, ..., c_N)^T$.

# Otimalidade dos estimadores

- Derivando $g(\boldsymbol{c})$ com relação à $\lambda$ e $c_i, i = 1, 2, .., n$, e igualando cada uma das n+1 equações a 0, vem que

$$\sum_{i=1}^{n} c_i - 1 = 0 \tag{6}$$

$$2c_i + \lambda = 0 \tag{7}$$

- De (6) vem que $\sum_{i=1}^{n} c_i = 1(*)$. Assim, utilizando (*) em (7), i=1,2,...,n, tem-se que $\lambda = -\frac{2}{n}(**)$.
- Logo utilizando (**) em (7), tem-se que $c_i = \frac{1}{n}$, $i = 1, 2, ..., n$. O que implica que o estimador ótimo é dado por $\widehat{\mu}_{\boldsymbol{sc}} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.
- Utilizando desenvolvimentos análogos, obtemos resultados semelhantes para a estimação de $\tau$ e $p$.