

Trabalho de Estatística Computacional II

André Luiz Becker e Eduardo Pereira Lima

24 de novembro de 2017

Problema:

Testes para independência em tabelas de contingência. Obter as curvas de poder do teste chi-quadrado (`chisq.test()`) e do teste de Fisher (`fisher.test()`) para a hipótese de independência entre variáveis aleatórias categóricas (independência de linhas e colunas). Avaliar a taxa de rejeição da hipótese nula para um experimento fatorial considerando os fatores: tamanho da amostra (n), número de categorias nas variáveis ($p \times q$, e.g. tabelas 2×2 , 3×2 , 3×3 , 3×4 , 4×4 , etc), grau de dependência entre as variáveis.

Etapas:

- Iniciamos simulando dados de uma Normal Bivariada.
- 1000 valores foram simulados
- Foi criado Tabelas de Contingência e com a contagem dos valores a respeito de sua média.
- Realizado os Testes de Chi-Quadrado e o Teste de Fisher, salvo os p-valores.
- Processo foi repetido 1000 vezes para se analisar as curvas de poder do teste.
- Avaliada a Taxa de Rejeição dos Dados.

P-Valores dos Testes Chi Quadrado e Teste de Fisher na Normal Bivariada

#Simulando dados de uma Normal Bivariada

`require(MASS)`

Loading required package: MASS

```
set.seed(1234)
n <- 1000
simul <- function(n){
  Sigma <- matrix(c(10,3,3,1),nrow=2,ncol=2) # Matriz de Variância-Covariância
  mu <- c(0,0) # Médias
  dados <- mvrnorm(n=1000, mu, Sigma)
  names(dados) <- c("x", "y")
  dados
  ## criando categorias
  q1 <- quantile(dados[,1], probs=c(0.5))
  q2 <- quantile(dados[,2], probs=c(0.5))
  sx <- sum(dados[,1]<q1)
  sy <- sum(dados[,2]<q1)

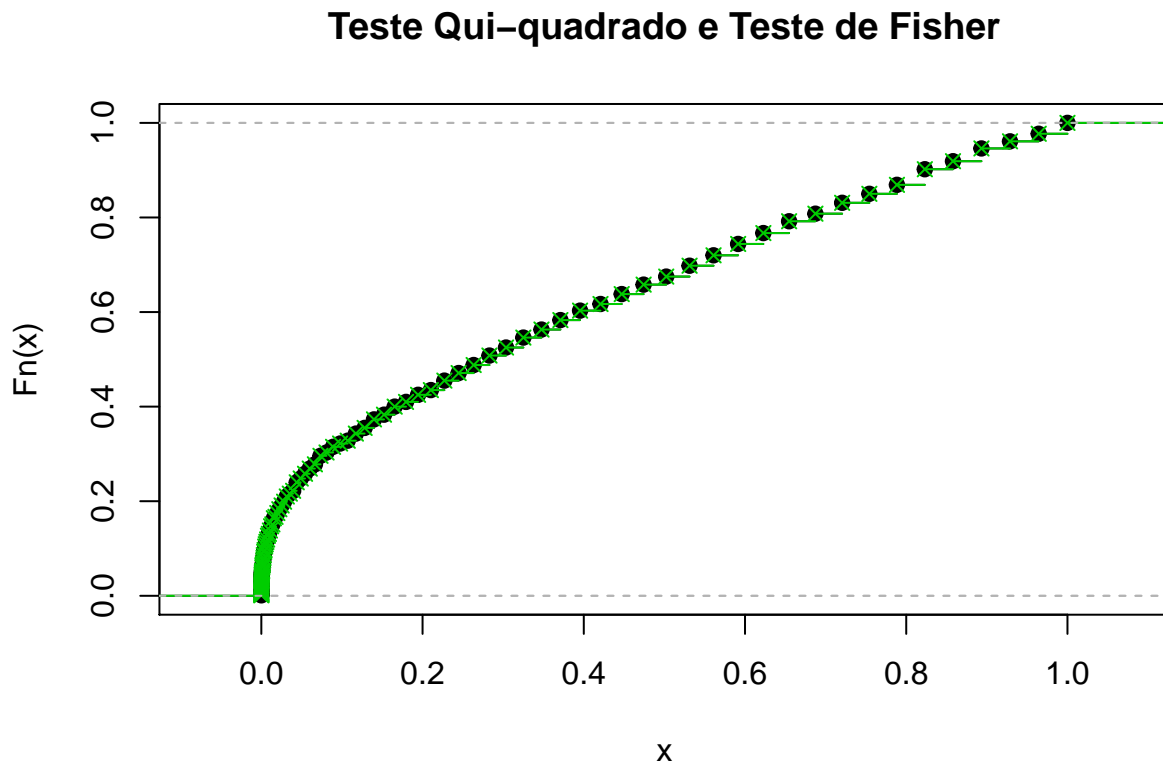
  #Arrumando a tabela de contingência
  tab <- matrix(c(sx, 1000-sx, sy, 1000-sy), nrow=2, ncol=2, byrow=TRUE,
    dimnames = list(c("x", "y"), c(">0", "<0")))
  tab
  ##Teste Qui-quadrado e Teste de Fisher
  cbind(chisq.test(tab)$p.value,fisher.test(tab)$p.value)
}
simul(n)
```

```
##           [,1]      [,2]
## [1,] 0.01761107 0.01758896
```

*Com somente uma interação os testes parecem ser bem próximos e sem diferença significativa.

Curva de Poder do Teste

```
Pv <- matrix(replicate(1000, simul(n)), ncol=2, byrow=TRUE)
pa <- Pv[,1]
pb <- Pv[,2]
## -----
plot(ecdf(pa), main= "Teste Qui-quadrado e Teste de Fisher")
plot(ecdf(pb), col= 3, pch=4 ,add=TRUE)
```



Taxa de Rejeição

```
C <- (sum(pa<0.05))/(sum(pa))
F <- (sum(pb<0.05))/(sum(pb))
T <- cbind(C,F)
T
```

```
##           C           F
## [1,] 0.696028 0.6960374
```

Tabelas 3x3

P-Valores Associados

```
set.seed(4321)
simul1 <- function(n){
  Sigma <- matrix(c(8,1,8,3,3,5,8,2,10),nrow=3,ncol=3, byrow=TRUE)
  mu <- c(0,0,0) # Médias
  dados1 <- mvrnorm(n=1000, mu, Sigma)
  names(dados1) <- c("x", "y", "z")
  cor(dados1)
  head(dados1)
  ## criando categorias
  sx <- sum(dados1[,1]>0.66)
  sy <- sum(dados1[,2]>0.66)
  sz <- sum(dados1[,3]>0.66)
  dx <- sum(dados1[,1]<0.33)
  dy <- sum(dados1[,2]<0.33)
  dz <- sum(dados1[,3]<0.33)
  #Arrumando a tabela de contingência
  tab <- matrix(c(sx, 1000-sx-dx, dx, sy, 1000-sy-dy, dy, sz, 1000-sz-dz, dz),
                nrow=3, ncol=3, byrow=TRUE, dimnames = list(c("x", "y", "z"),
                                                            c(">0.6", "Entre 0.3 e 0.6", "<0.3")))

  tab
  ##Teste Qui-quadrado e Teste de Fisher
  a <- cbind(chisq.test(tab[,1:2])$p.value, fisher.test(tab[,1:2])$p.value)
  b <- cbind(chisq.test(tab[,2:3])$p.value, fisher.test(tab[,2:3])$p.value)
  cbind(a,b)
}
simul1(n)
```

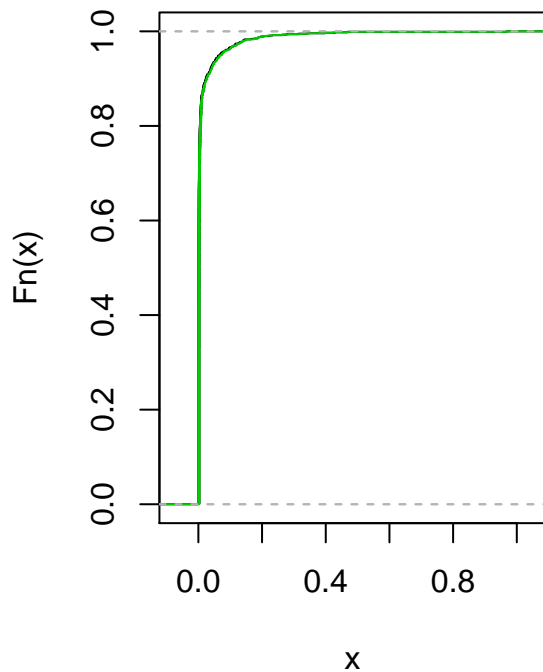
```
##           [,1]           [,2]           [,3]           [,4]
## [1,] 1.596908e-06 2.647957e-06 0.003778046 0.004065158
```

Curva de Poder do Teste (caso 3x3)

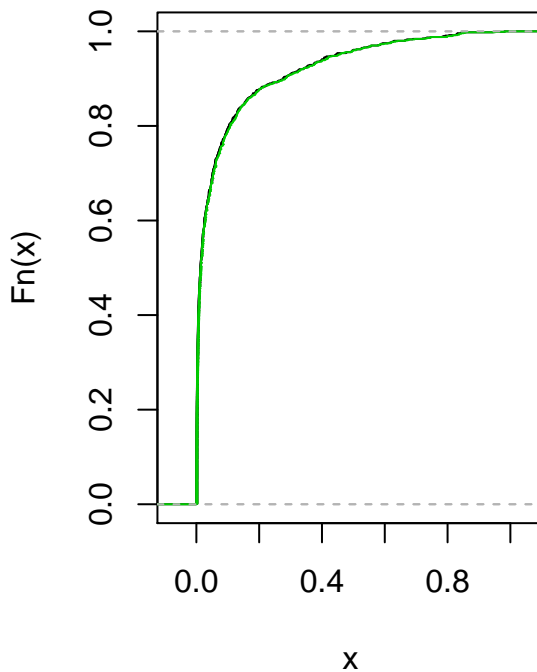
```
Pv1 <- matrix(replicate(1000, simul1()), ncol=4, byrow=TRUE)
pa <- Pv1[,1]
pc <- Pv1[,3]
pb <- Pv1[,2]
pd <- Pv1[,4]
## -----
par(mfrow=c(1,2))
plot(ecdf(pa), main= "Testes Correlacionados")
plot(ecdf(pb), col= 3, pch=4 ,add=TRUE)

plot(ecdf(pc), main= "Testes Não-correlacionado")
plot(ecdf(pd), col= 3, pch=4 ,add=TRUE)
```

Testes Correlacionados



Testes Não-correlacionado



Taxa de Rejeição

```
C <- (sum(pa<0.05))/(sum(pa<0.05)+sum(pa>0.05))
F <- (sum(pb<0.05))/(sum(pb<0.05)+sum(pa>0.05))
T <- cbind(C,F)
T
```

```
##           C           F
## [1,] 0.936 0.9358717
```

Caso 4x4

P-Valores Associados

```
set.seed(123)
simul2 <- function(){
  Sigma <- matrix(c(10,5,0.8,0.7,2,1,0.8,3,0.5,0.8,1,0.5,0.7,0.8,2,10),
                  nrow=4,ncol=4, byrow=TRUE)
  mu <- c(0.5,0.5,0.5,0.5) # Médias
  dados2 <- mvrnorm(n=50, mu, Sigma)
  names(dados2) <- c("x", "y", "z", "w")
  cor(dados2)
  ## criando categorias
```

```

sx <- sum(dados2[,1]>1)
sy <- sum(dados2[,2]>1)
sz <- sum(dados2[,3]>1)
sw <- sum(dados2[,4]>1)
Dx <- sum(dados2[,1]<0.5)
Dy <- sum(dados2[,2]<0.5)
Dz <- sum(dados2[,3]<0.5)
Dw <- sum(dados2[,4]<0.5)
dx <- sum(dados2[,1]<0)
dy <- sum(dados2[,2]<0)
dz <- sum(dados2[,3]<0)
dw <- sum(dados2[,4]<0)
#Arrumando a tabela de contingência
tab <- matrix(c(sx, 50-sx-Dx, Dx-dx, dx, sy, 50-sy-Dy, Dy-dy, dy,
               sz, 50-sz-Dz, Dz-dz, dz, sw, 50-sw-Dw, Dw-dw, dw),
              nrow=4, ncol=4, byrow=TRUE, dimnames = list(c("x", "y", "z", "w"),
                                                         c(">1", "Entre 0.5 e 1", "Entre 0 e 0.5", "<0.5")))
tab
##Teste Qui-quadrado e Teste de Fisher
a <- cbind(chisq.test(tab[,1:2])$p.value, fisher.test(tab[,1:2])$p.value)
b <- cbind(chisq.test(tab[,2:3])$p.value, fisher.test(tab[,2:3])$p.value)
c <- cbind(chisq.test(tab[,3:4])$p.value, fisher.test(tab[,3:4])$p.value)
return(cbind(a,b,c))
}
simul2()

```

```

##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,] 0.01520519 0.01139531 0.513197 0.5544798 0.0008714971 0.000876426

```

Curva de Poder do Teste (Caso 4x4)

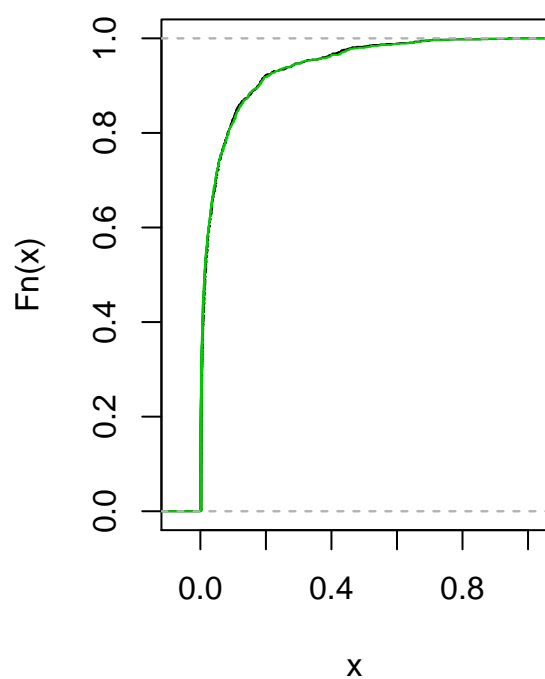
```

Pv2 <- matrix(replicate(1000, simul2()), ncol=6, byrow=TRUE)
pa <- Pv2[,1]
pd <- Pv2[,4]
pb <- Pv2[,2]
pe <- Pv2[,5]
pc <- Pv2[,3]
pf <- Pv2[,6]
## -----
par(mfrow=c(1,2))
plot(ecdf(pa), main= "Testes Correlacionados")
plot(ecdf(pb), col= 3, pch=4 ,add=TRUE)

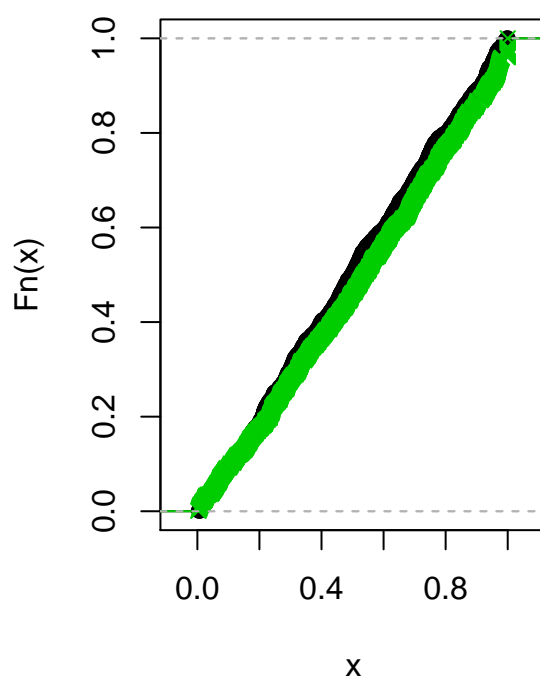
plot(ecdf(pc), main= "Testes Não-Correlacionados")
plot(ecdf(pd), col= 3, pch=4 ,add=TRUE)

```

Testes Correlacionados

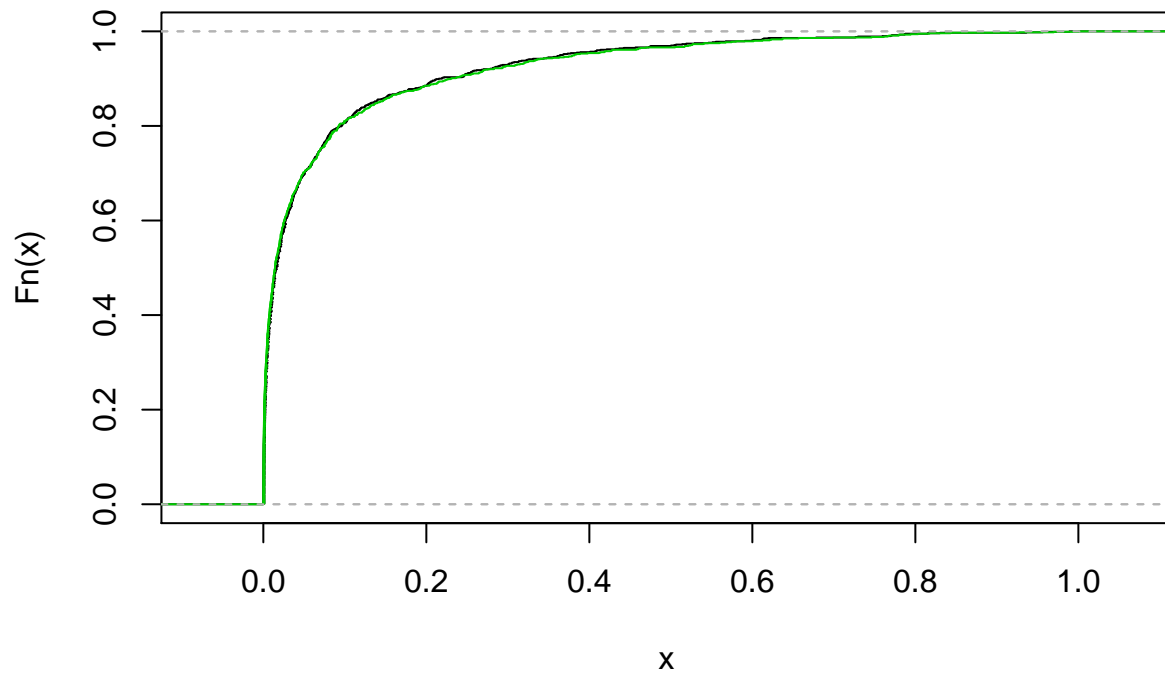


Testes Não-Correlacionados



```
layout(1)
plot(ecdf(pe), main= "Teste Qui-quadrado e Teste de Fisher")
plot(ecdf(pf), col= 3, pch=4 ,add=TRUE)
```

Teste Qui-quadrado e Teste de Fisher



Taxa de Rejeição para Correlacionadas

```
A <- (sum(pa<0.05))/(sum(pa<0.05)+sum(pa>0.05))
B <- (sum(pb<0.05))/(sum(pb<0.05)+sum(pb>0.05))
T <- cbind(A,B)
T
```

```
##           A      B
## [1,] 0.716 0.718
```

Taxa de Rejeição para Não-Correlacionadas

```
E <- (sum(pc<0.05))/(sum(pc<0.05)+sum(pc>0.05))
F <- (sum(pd<0.05))/(sum(pd<0.05)+sum(pd>0.05))
T2 <- cbind(E,F)
T2
```

```
##           E      F
## [1,] NA 0.042
```

Conclusões:

*Nos Testes realizados na Tabela 4x4 o Teste Chi-Quadrado começou a se destacar aproximadamente a partir do ponto 1,5,