

# Laboratório 07 - Estatística Aplicada - Manhã

**Aluno: Marcus Vinícius Leite Costa**  
**Matrícula: 116110728**

1) O código utilizado foi o seguinte:

```
##### QUESTAO 01 #####

# hipotese h0: Testar se a moeda favorece cara

# hipotese h1: Testar se a moeda nao favorece cara

prob_cara = 2/5

prob_zero_caras = dbinom(0, 4, prob_cara)

prob_uma_cara = dbinom(1, 4, prob_cara)

prob_duas_caras = dbinom(2, 4, prob_cara)

prob_tres_caras = dbinom(3, 4, prob_cara)

prob_quatro_caras = dbinom(4, 4, prob_cara)

valores_observados = c(72,204,228,101,20)
valores_esperados = c(625*prob_zero_caras, 625*prob_uma_cara,
625*prob_duas_caras, 625*prob_tres_caras, 625*prob_quatro_caras)

qui_quadrado = sum( ((valores_observados - valores_esperados) **2) /
valores_esperados)

pchisq(qui_quadrado, df = 4, lower.tail = FALSE)

##### Como P = 0.4637673 > 0.05, entao, nao rejeitamos h0, logo,
existem evidencias que a moeda favorece cara #####
```

2)

a) Código utilizado na letra A:

```
##### QUESTAO 02 #####
```

```
dados = dados_cia = read.csv('DadosCiaMB.csv', sep = ";")
```

```
attach(dados)
```

```
# Letra A, criando a tabela de distribuicao conjunta para as variaveis  
grau de instrucao e regio de procedencia
```

```
table(Instrucao, procedencia)
```

Tabela obtida na letra A:

	\procedencia		
Instrucao	capital	interior	outra
f	4	3	5
m	5	7	6
s	2	2	2

b)

```
# Letra B
```

```
linha_um <- c(tabela[1,1:3])
```

```
linha_dois <- c(tabela[2,1:3])
```

```
linha_tres <- c(tabela[3,1:3])
```

```
col_um <-c(tabela[1:3,1])
```

```
col_dois <-c(tabela[1:3,2])
```

```
col_tres <-c(tabela[1:3, 3])
```

```
sum_linhas = c( sum(linha_um), sum(linha_dois), sum(linha_tres) )
```

```
sum_colunas = c( sum(col_um), sum(col_dois), sum(col_tres) )
```

```
quiQ2 = 0
```

```
calculaEsperado <- function (i, j){  
  return ((sum_linhas[i] * sum_colunas[j]) / sum(sum_linhas))  
}
```

```
(sum_linhas[1] * sum_colunas[1]) / sum(sum_linhas)
```

```
calculaEsperado(1,1)
```

```
sum(sum_linhas)
```

```
for(i in 1:3) {
```

```

    for(j in 1:3) {
        quiQ2 = (quiQ2 + ((tabela[i,j] - calculaEsperado(i, j))**2) /
calculaEsperado(i,j))
    }
}

pchisq(quiQ2, df = 4, lower.tail = TRUE)

# H0: Existe dependencia entre as variaveis
# H1: Nao existe dependencia entre as variaveis

# Como P = 0.04400023 < 0.05, entao, rejeitamos H0, logo, existem
evidencias que nao ha dependencia entre as variaveis

```

c) Código utilizado (com a biblioteca gmodels) para montar a tabela de distribuição conjunta das variáveis grau de instrução e região de procedência, e também, obter o valor de qui-quadrado

```

##### LETRA C #####
library(gmodels)

CrossTable(Instrucao,procedencia,chisq=TRUE)

```

O resultado obtido foi o seguinte:

Total observations in Table: 36

Instrucao	procedencia capital	interior	outra	Row Total
f	4	3	5	12
	0.030	0.250	0.103	0.333
	0.333	0.250	0.417	
	0.364	0.250	0.385	
	0.111	0.083	0.139	
m	5	7	6	18
	0.045	0.167	0.038	0.500
	0.278	0.389	0.333	
	0.455	0.583	0.462	
	0.139	0.194	0.167	
s	2	2	2	6
	0.015	0.000	0.013	0.167
	0.333	0.333	0.333	
	0.182	0.167	0.154	
	0.056	0.056	0.056	
Column Total	11	12	13	36
	0.306	0.333	0.361	

Statistics for All Table Factors

Pearson's Chi-squared test

-----  
Chi^2 = 0.6614219      d.f. = 4      p = 0.9559998

d) Utilizando-se do pacote gmodels, foi-se calculado o qui-quadrado para as variáveis estado civil e grau de instrução da seguinte maneira:

```
##### LETRA D #####
```

```
CrossTable(Instrucao,Civil,chisq=TRUE)
```

Os valores obtidos foram:

```
Pearson's Chi-squared test
```

```
-----  
Chi^2 = 1.9125      d.f. = 2      p = 0.3843314
```

Utilizando as seguintes hipóteses:

H0: As variáveis grau de instrução e estado civil são independentes

H1: As variáveis grau de instrução e estado civil são dependentes

Podemos deduzir que:

Como  $P = 0.3843314 > 0.05$ , então, não rejeitamos  $H_0$ , logo, podemos assumir que as variáveis são independentes.

e) Código utilizado na letra E:

```
##### LETRA E #####

dados$salario <- as.numeric(gsub(",", "\\.", dados$salario))

dados$idade <- as.numeric(gsub(",", "\\.", dados$idade))

cor(dados$salario, dados$idade)

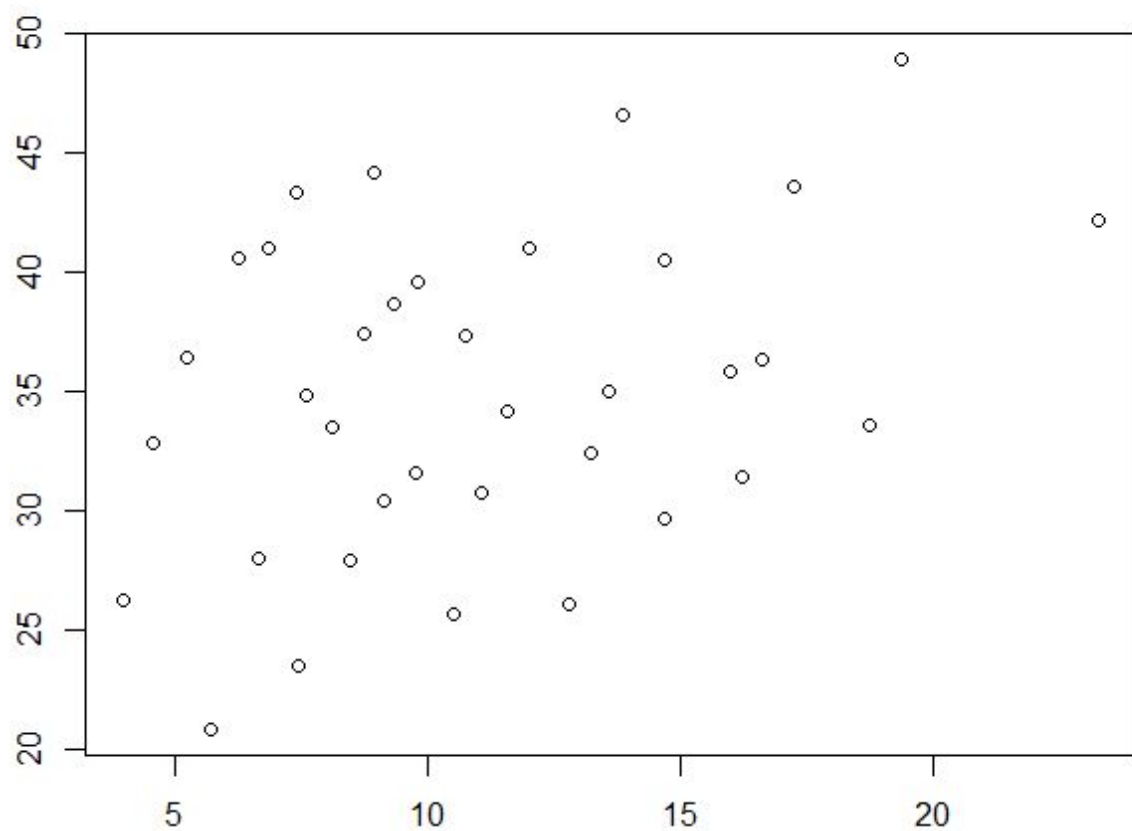
plot(dados$salario, dados$idade)

# H0: Nao ha correlacao linear significativa
# H1: Ha correlacao linear significativa
```

Ao calcular a correlação linear entre as variáveis salário e idade, obtém-se o seguinte valor:

0.365186

Gráfico das variáveis:



Logo, podemos deduzir, a partir do gráfico e o valor de correlação obtido (0.365186), que não existe uma correlação linear significativa entre as variáveis, pois, além de ser observável pelo gráfico, que as variáveis não formam uma reta constante, também é possível deduzir a partir do valor de correlação obtido, que está distante dos extremos (onde existe bastante correlação)  $[-1,1]$