

Inferência em Dados Categóricos

Bioestatística em R

André M Ribeiro-dos-Santos

10 de 04, 2017

- Reconhecer variáveis categóricas.
- Identificar quando aplicar teste binomial.
- Identificar quando aplicar um teste de qui-quadrado e/ou Fisher.
- Compreender qual a pergunta estatística resolvida por cada teste.
- Realizar os testes.
- Visualizar e Ilustrar os resultados.

Testes de Proporção

Num estudo de Fibrose Cística na população de Belém. Você começou a reunir amostras de casos da doença no HUJBB. Na primeira coleta, você obteve 10 amostras das quais 7 (70%) apresentam um quadro de hipoproteinemia. No entanto, esperava-se que apenas 45% dos casos apresentem hipoproteinemia. A elevada incidência pode indicar um fator genético atuando na população.

A frequência de hipoproteinemia diferencia-se da esperada?

$$H_o : \hat{p} = p$$

$$H_a = \hat{p} \neq p$$

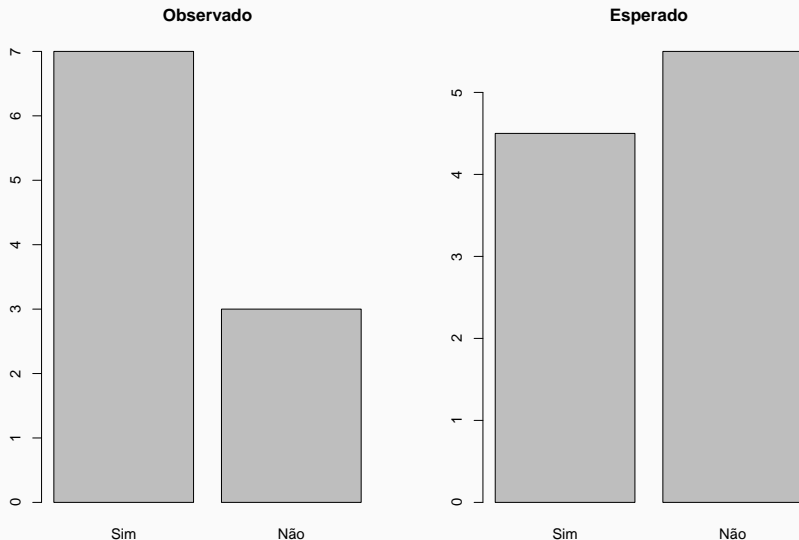
Quais as características do experimento?

- Qual Tamanho da amostra?
- Qual a variável em questão?
- A variável é numérica ou categórica?
- Quantos casos foram observados e quantos seriam esperados?

Quais as características do experimento?

- Qual Tamanho da amostra? **10**
- Qual a variável em questão? **presença de hipoproteinemia**
- A variável é numérica ou categórica? **categórica, com duas respostas**
- Quantos casos foram observados e quantos seriam esperados? **Foram observados 7, mas esperava-se 4.5**

```
> par(mfrow = c(1,2))  
> barplot(c("Sim" = 7, "Não" = 3), main = "Observado")  
> barplot(c("Sim" = 4.5, "Não" = 5.5), main = "Esperado")
```



Distribuição Binomial $B(n, p)$

A distribuição de **variáveis categóricas com duas alternativas (ou grupos)** são descritas por uma **distribuição binomial** (ou de bernoulli). A qual descreve a probabilidade *discreta* de observa-se x “sucessos” em n independente experimentos com probabilidade p .

$$P(x; n, p) = \binom{n}{x} p^x (1 - p)^{n-x}$$

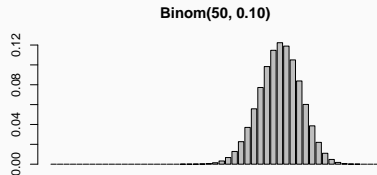
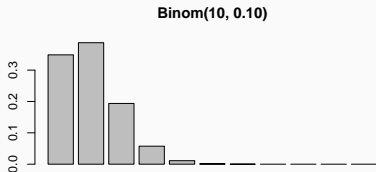
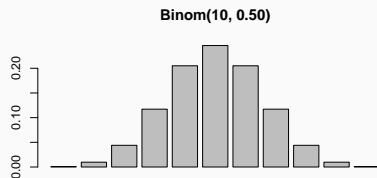
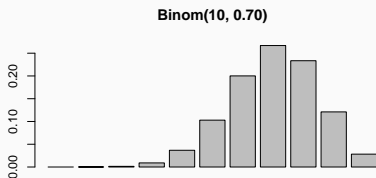
```
> dbinom(x, size, prob, log = FALSE)
> pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)
> qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)
> rbinom(n, size, prob)
```



```

> par(mfrow = c(2,2))
> barplot(dbinom(0:10, size = 10, p = 0.7), main = "Binom(10, 0.70)")
> barplot(dbinom(0:10, size = 10, p = 0.5), main = "Binom(10, 0.50)")
> barplot(dbinom(0:10, size = 10, p = 0.1), main = "Binom(10, 0.10)")
> barplot(dbinom(0:50, size = 50, p = 0.7), main = "Binom(50, 0.10)")

```



Teste Binomial

Uma vez que a variável em questão segue uma *distribuição binomial*, podemos avaliar se frequência observada diferencia-se do esperado usando um **teste binomial**.

```
> ? binom.test
> ## Exact Binomial Test
> ##
> ## Description:
> ##     Performs an exact test of a simple null
> ##     hypothesis about the probability of
> ##     success in a Bernoulli experiment.
> ## Usage:
> ##     binom.test(x, n, p = 0.5,
> ##               alternative = c("two.sided", "less", "greater"),
> ##               conf.level = 0.95)
> ## ...
```

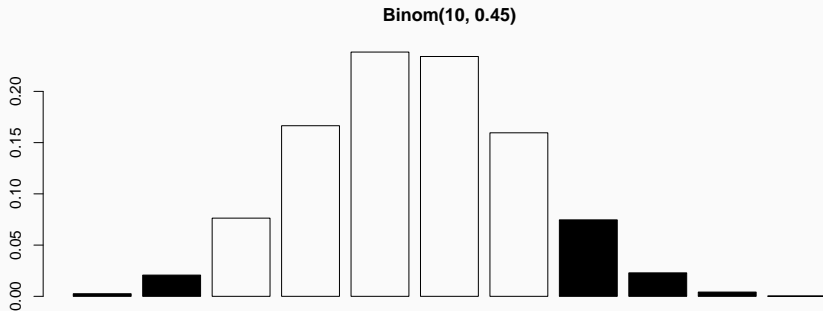
A frequência de hipoproteinemia diferencia-se da esperada?

$$H_o : \hat{p} = p; \quad H_a = \hat{p} \neq p$$

```
> binom.test(7, n = 10, p = 0.45)
```

```
##  
## Exact binomial test  
##  
## data: 7 and 10  
## number of successes = 7, number of trials = 10, p-value = 0.1253  
## alternative hypothesis: true probability of success is not equal to 0.45  
## 95 percent confidence interval:  
## 0.3475471 0.9332605  
## sample estimates:  
## probability of success  
## 0.7
```

```
> marked <- (0:10 < floor(7 - 4.5)) | (0:10 >= 7)
> probs <- dbinom(0:10, size = 10, p = 0.45)
> barplot(probs, col = marked, main = "Binom(10, 0.45)")
```



```
> sum(probs[marked])
```

```
## [1] 0.125252
```

Exercícios - Teste de Proporção

Questão (1)

Após a segunda coleta de amostras, os pesquisadores observaram mais 11 casos de hipoproteinemia entre as 15 amostras coletadas. Após reunir as amostras, qual a frequência de hipoproteinemia observada? Os pesquisadores constataram diferenças significativas da proporção esperada?

```
> x      <- 7 + 11
> n      <- 10 + 15
> x / n
```

```
## [1] 0.72
```

```
> binom.test(x, n, p = 0.45)
```

```
##
## Exact binomial test
##
## data:  x and n
## number of successes = 18, number of trials = 25, p-value =
## 0.008143
## alternative hypothesis: true probability of success is not equal to 0.45
## 95 percent confidence interval:
##  0.5061232 0.8792833
## sample estimates:
## probability of success
##                                0.72
```

Questão (2)

Qual foi o *Odds-Ratio* (ou risco relativo) de hipoproteinemia em relação ao esperado?

$$\text{Odds}(p) = p/(1 - p)$$

$$OR = \frac{\text{Odds}(p_o)}{\text{Odds}(p_e)}$$

$$OR = \frac{p_o/(1 - p_o)}{p_e/(1 - p_e)}$$


```
> p_o <- x / n  
> p_e <- 0.45  
>  
> (p_o / (1 - p_o))
```

```
## [1] 2.571429
```

Questão (3)

Qual o intervalo de confiança da proporção observada? Represente o valor em gráficos.

```
> (cnfint <- qbinom(c(0.025, 0.975), n, p_o) / 25)
```

```
## [1] 0.52 0.88
```

```
> par(mfrow = c(1,2))
```

```
>
```

```
> barplot(c("Esperado" = p_e, "Observado" = p_o), ylim=c(0, 1))
```

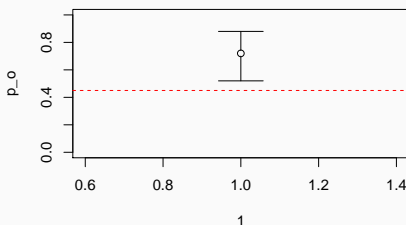
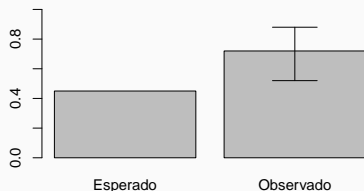
```
> arrows(1.9, cnfint[1], 1.9, cnfint[2], code = 3, angle = 90)
```

```
>
```

```
> plot(1, p_o, ylim=c(0,1))
```

```
> abline(h = p_e, lty = 'dashed', col = 'red')
```

```
> arrows(1, p_o + c(-0.02, 0.02), 1, cnfint, angle=90)
```



Questão (4)

Tendo em vista que a hipoproteïnemia ocorre geralmente associada a anemia, os pesquisadores decidiram avaliar se anemia também estava alterada na amostra. Utilizando a tabela abaixo, avalie se houve mudança em relação a proporção de anêmicos esperados (35%). Qual a conclusão?

```
> cftr <- read.table('cftr-ex.tsv', header = T)
```

```
> table(cftr$anemia)
```

```
##
```

```
## Não Sim
```

```
## 28 22
```

```
> binom.test(table(cftr$anemia), p = 1 - 0.35)
```

```
##
```

```
## Exact binomial test
```

```
##
```

```
## data: table(cftr$anemia)
```

```
## number of successes = 28, number of trials = 50, p-value = 0.1849
```

```
## alternative hypothesis: true probability of success is not equal to 0.65
```

```
## 95 percent confidence interval:
```

```
## 0.4125441 0.7000928
```

```
## sample estimates:
```

```
## probability of success
```

```
## 0.56
```

5. Dada uma variável x corresponde ao número de sucessos em uma série de 50 experimentos de Bernoulli cuja probabilidade de sucesso é 30%. Calcule:
- $P(x = 15)$, $P(x \geq 15)$, e $P(x < 15)$.
 - $P(15 < x \leq 35)$, ilustre num gráfico da distribuição a área selecionada.
 - Quantis $x_{0.025}$ e $x_{0.975}$.
6. Qual a frequência do sexo (**sexo**) e tipo de resposta ao tratamento (**response**) na amostra da tabela (**cft**)?
7. A amostra apresenta uma proporção similar de homens e mulheres?

• $P(x = 15)$, $P(x \geq 15)$, e $P(x < 15)$?

```
> dbinom(15, 50, 0.3) ## ou sum(dbinom(15:50, 50, 0.3))
```

```
## [1] 0.1223469
```

```
> pbinom(14, 50, 0.3) ## ou sum(dbinom(0:14, 50, 0.3))
```

```
## [1] 0.4468316
```

```
> pbinom(14, 50, 0.3, lower.tail = F)
```

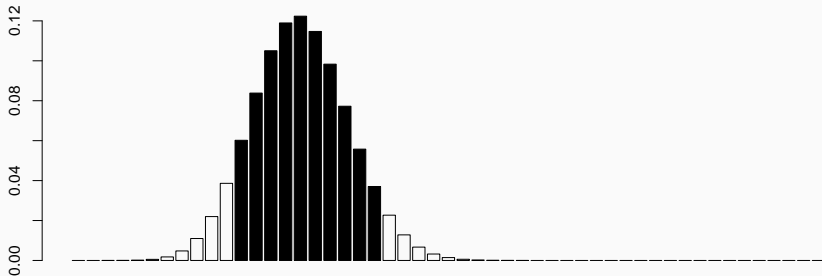
```
## [1] 0.5531684
```

- $P(15 < x \leq 35)$, ilustre num gráfico da distribuição a área selecionada.

```
> ## sum(dbinom(9:25, size = 50, p = 0.3))  
> pbinom(25, 50, 0.3) - pbinom(10, 50, 0.3)
```

```
## [1] 0.9202162
```

```
> barplot(dbinom(0:50, 50, 0.3), col = (0:50 > 10) & (0:50 <= 20))
```



- Quantis $x_{0.025}$ e $x_{0.975}$.

```
> qbinom(c(0.025, 0.975), 50, 0.3)
```

```
## [1] 9 22
```

Qual a frequência do sexo (**sexo**) e tipo de resposta ao tratamento (**response**) na amostra da tabela (**cftr**)?

```
> prop.table(table(cftr$sexo))
```

```
##  
##    F    M  
## 0.6 0.4
```

```
> prop.table(table(cftr$response))
```

```
##  
##    I    II   III  
## 0.28 0.50 0.22
```

A amostra apresenta uma proporção similar de homens e mulheres?

```
> binom.test(table(cftr$sexo))
```

```
##  
## Exact binomial test  
##  
## data: table(cftr$sexo)  
## number of successes = 30, number of trials = 50, p-value = 0.2026  
## alternative hypothesis: true probability of success is not equal to 0.5  
## 95 percent confidence interval:  
## 0.4517940 0.7359216  
## sample estimates:  
## probability of success  
## 0.6
```

Testes de Independência

Uma questão de independência

Ao pesquisar os sintomas de *Fibrose Cística*, os pesquisadores constataram que *anemia* geralmente acompanha os pacientes com *hipoproteinemia*. Sendo a ocorrência de anemia muito mais comum entre pacientes com *hipoproteinemia*. Na amostra estudada, eles observaram a seguinte tabela de confusão (ou contingência):

Hipoproteinemia	Anemia	
	Não	Sim
Não	2	6
Sim	11	7

A anemia é distribuída independentemente da hipoproteinemia?

Quais as características do experimento?

- Qual Tamanho da amostra?
- Quais as variáveis em questão?
- As variáveis são numérica ou categórica?
- Descreva o problema estatístico em questão.

Quais as características do experimento?

- Qual Tamanho da amostra? 25
- Quais as variáveis em questão? **presença de hipoproteinemia e anemia**
- As variáveis são numérica ou categórica? **ambas categóricas, com duas respostas**
- Descreva o problema estatístico em questão. **deseja-se avaliar se o conhecimento da presença de anemia afeta a chance do individuo apresentar hipoproteinemia**

Uma vez que desejamos avaliar a independência entre duas variáveis categóricas com duas possíveis respostas, podemos utilizar o **teste de Fisher**.

$$H_0 : P(B|A) = P(B); \quad H_a : P(B|A) \neq P(B)$$


```
> ? fisher.test
> ## Fisher's Exact Test for Count Data
> ## Description:
> ##      Performs Fisher's exact test for testing the null of
> ##      independence of rows and columns in a contingency
> ##      table with fixed marginals.
> ## Usage:
> ##      fisher.test(x, y = NULL, workspace = 200000,
> ##                  hybrid = FALSE, control = list(),
> ##                  or = 1, alternative = "two.sided",
> ##                  conf.int = TRUE, conf.level = 0.95,
> ##                  simulate.p.value = FALSE, B = 2000)
> ##...
```

```

> table <- matrix(c(2, 11, 6, 7), nrow=2,
+                 dimnames = list("hipoproteinemia"=c("Não", "Sim"),
+                 "anemia" =c("Não", "Sim")))
> addmargins(table)

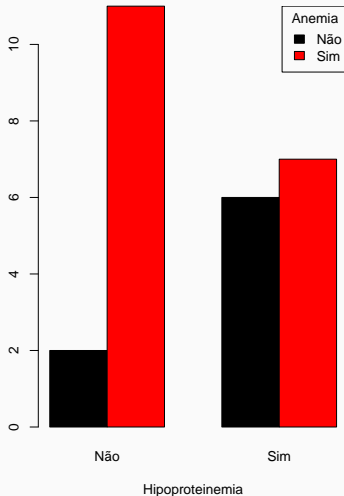
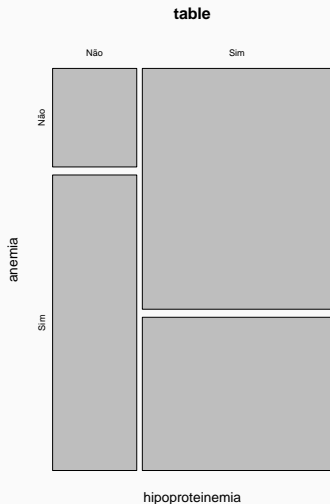
```

```

##              anemia
## hipoproteinemia Não Sim Sum
##              Não   2   6   8
##              Sim  11   7  18
##              Sum  13  13  26

```

```
> par(mfrow=c(1, 2))
> mosaicplot(table)
> barplot(table, beside=TRUE, col=1:2, xlab = "Hipoproteinemia")
> legend("topright", rownames(table), fill=1:2, title="Anemia")
```



```
> fisher.test(table)
```

```
##
```

```
## Fisher's Exact Test for Count Data
```

```
##
```

```
## data:  table
```

```
## p-value = 0.2016
```

```
## alternative hypothesis: true odds ratio is not equal to 1
```

```
## 95 percent confidence interval:
```

```
## 0.0175156 1.7390386
```

```
## sample estimates:
```

```
## odds ratio
```

```
## 0.2257303
```

Os pesquisadores também acreditam que a presença de *hipoproteinemia* está associado ao *risco* de resposta adversa ao tratamento. O risco é dividido em três classes (I, II, e III). Ao avaliar os pacientes estudados, obtiveram a seguinte tabla de confusão:

Resposta	Hipoproteinemia	
	Não	Sim
I	0	6
II	3	10
III	3	3

O risco é independente da hipoproteinemia?

Neste caso, uma das variáveis em questão possui três possíveis respostas, não sendo mais adequado utilizar o **teste de Fisher**, no lugar podemos utilizar o **teste de Qui-quadrada** para avaliar a independência.

$$H_0 : P(B|A) = P(B); \quad H_a : P(B|A) \neq P(B)$$

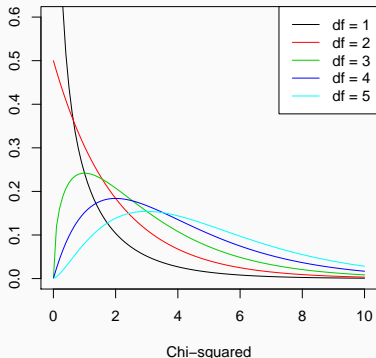
O teste usa a distribuição qui-quadrada ou χ^2 para avaliar o teste.

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

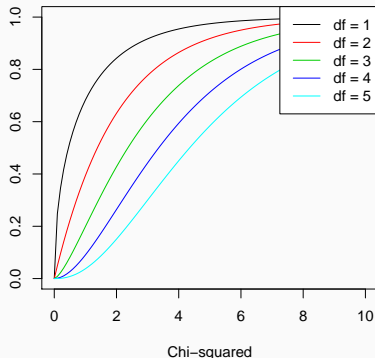
A distribuição Qui-quadrada ou $\chi^2(k)$ com k graus de liberdade representa a distribuição da soma dos quadrados de k independentes curvas normais.

$$df = (|linhas| - 1) * (|colunas| - 1)$$

Chi-squared distribution



Chi-squared cumulative distribution



```
> ? chisq.test
> ## Pearson's Chi-squared Test for Count Data
> ## Description:x
> ##      'chisq.test' performs chi-squared contingency table
> ##      tests and goodness-of-fit tests.
> ## Usage:
> ##      chisq.test(x, y = NULL, correct = TRUE,
> ##                p = rep(1/length(x), length(x)),
> ##                rescale.p = FALSE,
> ##                simulate.p.value = FALSE, B = 2000)
```



```
> obs <- matrix(c(0, 3, 3, 6, 10, 3), nrow = 3,  
+               dimnames = list("Resposta" = c("I", "II", "III"),  
+                               "Hipoproteinemia" = c("Não", "Sim")))  
> exp <- rowSums(obs) %*% t(colSums(obs)) / sum(obs)  
> (chi <- sum((obs-exp)^2/exp))
```

```
## [1] 4.124494
```

```
> df <- (nrow(obs) - 1) * (ncol(obs) - 1)  
> 1 - pchisq(chi, df = df)
```

```
## [1] 0.1271679
```

```
> chisq.test(obs)
```

```
##
```

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

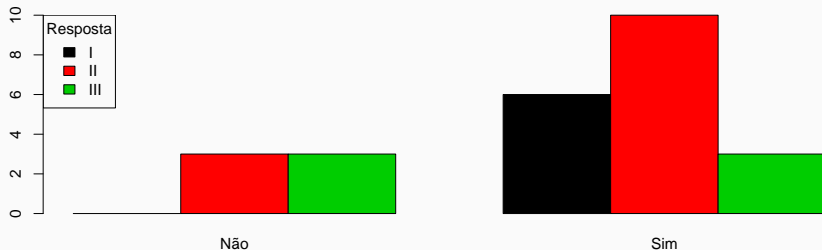
```
##
```

```
## data: obs
```

```
## X-squared = 4.1245, df = 2, p-value = 0.1272
```

```
> barplot(obs, beside = T, col = 1:3)
```

```
> legend("topleft", rownames(obs), fill = 1:3, title = "Resposta")
```



Quando usar Fisher ou Qui-quadrado?

- Mais indicado **Teste de Fisher**:
 - Tabela de contingência 2x2.
 - Preferido para pequenas amostragens.
- Mais indicado **Qui-quadrado**:
 - alguma das variáveis possui 3 ou mais categorias.
 - o tamanho amostral maior do que 1000.

Exercícios - Teste de Independência

Quesstão (1)

Nos pacientes de Fibrose Cística coletadas, podemos investigar se existe associação entre anemia (**anemia**) e hipoproteinemia (**hipoptr**). Usando a tabela **cftr-ex.tsv** faça:

- A tabela de confusão (ou contingência) entre anemia e hipoproteinemia.
- Represente graficamente esta distribuição.
- Existe evidência de associação entre as variáveis? Qual a sua conclusão.

```
> ## Carregar a tabela  
> cf <- read.table('cftr-ex.tsv', header = T)
```

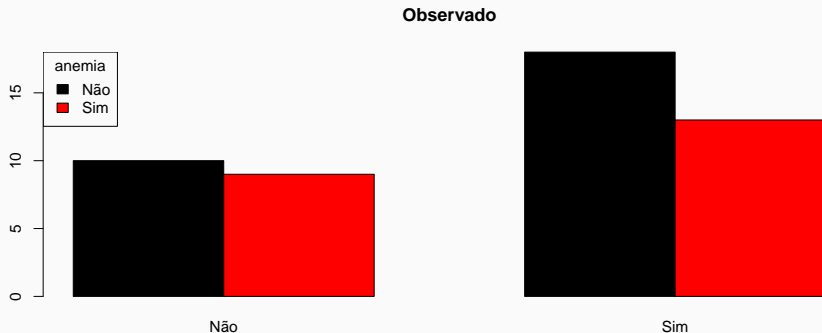
- A tabela de confusão (ou contingência) entre anemia e hipoproteinemia.

```
> ctg <- table(cf$anemia, cf$hipoptr)
> ## ctg <- xtabs(~anemia+hipoptr, data = cf)
> addmargins(ctg)
```

```
##
##      Não Sim Sum
## Não   10  18  28
## Sim    9  13  22
## Sum   19  31  50
```

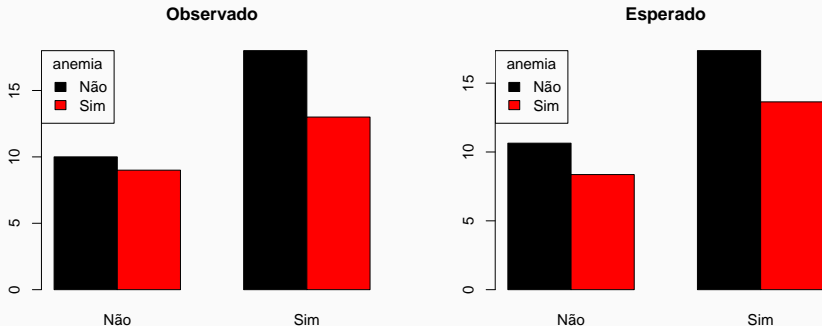
- Represente graficamente esta distribuição.

```
> barplot(ctg, main = "Observado", beside = T, col = 1:2)  
> legend("topleft", rownames(ctg), fill = 1:2, title="anemia")
```



- Represente graficamente esta distribuição.

```
> exp <- rowSums(ctg) %*% t(colSums(ctg)) / sum(ctg)
>
> par(mfrow = c(1, 2))
> barplot(ctg, main = "Observado", beside = T, col = 1:2)
> legend("topleft", rownames(ctg), fill = 1:2, title="anemia")
>
> barplot(exp, main = "Esperado", beside = T, col = 1:2)
> legend("topleft", rownames(ctg), fill = 1:2, title="anemia")
```



- Existe evidência de associação entre as variáveis? Qual a sua conclusão.

```
> fisher.test(ctg)
```

```
##  
## Fisher's Exact Test for Count Data  
##  
## data:  ctg  
## p-value = 0.7742  
## alternative hypothesis: true odds ratio is not equal to 1  
## 95 percent confidence interval:  
##  0.2194035 2.9592002  
## sample estimates:  
## odds ratio  
##  0.8060332
```

Questão (2)

Da mesma forma, os pesquisadores desejam investigar a associação entre o tipo de resposta ao tratamento (**response**).

- Gere a tabela de contingência entre o tipo de resposta ao tratamento e hipoproteinemia.
- Represente graficamente a distribuição.
- As variáveis são distribuídas independentemente? Qual a sua conclusão.

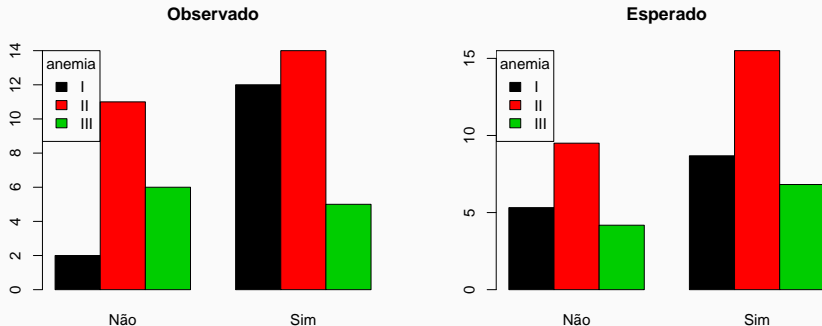
- Gere a tabela de contingência entre o tipo de resposta ao tratamento e hipoproteinemia.

```
> ctg <- table(cf$response, cf$hipoptr)
> addmargins(ctg)
```

```
##
##      Não Sim Sum
##  I      2  12  14
##  II     11  14  25
##  III     6   5  11
##  Sum    19  31  50
```

- Represente graficamente a distribuição.

```
> exp <- rowSums(ctg) %*% t(colSums(ctg)) / sum(ctg)
>
> par(mfrow = c(1, 2))
> barplot(ctg, main = "Observado", beside = T, col = 1:3)
> legend("topleft", rownames(ctg), fill = 1:3, title="anemia")
>
> barplot(exp, main = "Esperado", beside = T, col = 1:3)
> legend("topleft", rownames(ctg), fill = 1:3, title="anemia")
```



- As variáveis são distribuídas independentemente? Qual a sua conclusão.

```
> chisq.test(ctg)
```

```
##
```

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

```
##
```

```
## data:  ctg
```

```
## X-squared = 5.0019, df = 2, p-value = 0.08201
```

Questão (3)

Para investigar o excesso de hipoproteinemia na amostra, os pesquisadores levantaram duas mutações de interesse (mut_a e mut_b). O primeiro passo para toda pesquisa de genética é investigar se as mutações estão em equilíbrio de Hardy-Weinberg. Para cada mutação calcule a frequência alélica e genotípica (esperada e observada) e avalie se a mutação está em equilíbrio.

Mutação A

```
> (obs <- table(cf$mut_a))
```

```
##
```

```
## aa Aa AA
```

```
## 29 17 4
```

```
> (p <- (obs[1] * 2 + obs[2]) / (2 * sum(obs)))
```

```
## aa
```

```
## 0.75
```

```
> (exp <- c(p^2, 2 * p * (1-p), (1-p)^2) * sum(obs))
```

```
## aa aa aa
```

```
## 28.125 18.750 3.125
```

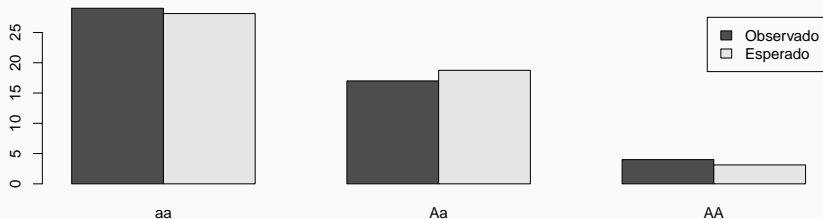
```
> (chi <- sum((obs - exp)^2/exp))
```

```
## [1] 0.4355556
```

```
> pchisq(chi, df=1, lower.tail = FALSE)
```

```
## [1] 0.5092754
```

```
> barplot(rbind(obs, exp), beside = T,  
+         legend = c("Observado", "Esperado"))
```



Por que $df = 1$?

Quando avaliamos o equilíbrio de Hardy-Weinberg, verificamos se os alelos de um cromossomo são independentes do outro. Portanto avaliamos a seguinte tabela de contingência.

paterno\materno		a	A
a	19.0	12.5	
A	12.5	6.0	

Que podemos simplificar para:

	aa	Aa	AA
Observado	19	25	6
Esperado	p^2	$2 * p * (1 - p)$	$(1 - p)^2$

Mutação B

```
> (obs <- table(cf$mut_b))
```

```
##
```

```
## bb Bb BB
```

```
## 41 8 1
```

```
> (p <- (obs[1] * 2 + obs[2]) / (2 * sum(obs)))
```

```
## bb
```

```
## 0.9
```

```
> (exp <- c(p^2, 2 * p * (1-p), (1-p)^2))
```

```
## bb bb bb
```

```
## 0.81 0.18 0.01
```

```
> (chi <- sum((obs - exp)^2/exp))
```

```
## [1] 2431.864
```

```
> pchisq(chi, df=1, lower.tail = FALSE)
```

```
## [1] 0
```

5. Agora na amostra completa, avalie se existe associação da presença de anemia com a hipoproteinemia, tipo de resposta e com os marcadores estudados. Ilustre os resultados obtidos.
6. Constatando que ambas as variantes investigadas estão em equilíbrio de Hardy-Weinberg, os pesquisadores decidiram avaliar a associação delas com o hipoproteinemia. Gere a tabela de contingência entre hipoproteinemia e cada marcador e avalie para cada mutação ilustrando os resultados:
 - Associação com os genótipos
 - Associação com modelo de efeito recessivo e dominante.

Questão (5)

```
> chisq.test(table(cf$anemia, cf$hipoptr))
```

```
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data:  table(cf$anemia, cf$hipoptr)  
## X-squared = 0.0067526, df = 1, p-value = 0.9345
```

```
> chisq.test(table(cf$anemia, cf$response))
```

```
##  
## Pearson's Chi-squared test  
##  
## data:  table(cf$anemia, cf$response)  
## X-squared = 0.88653, df = 2, p-value = 0.6419
```

```
> chisq.test(table(cf$anemia, cf$mut_a))
```

```
##
```

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

```
##
```

```
## data: table(cf$anemia, cf$mut_a)
```

```
## X-squared = 0.65865, df = 2, p-value = 0.7194
```

```
> chisq.test(table(cf$anemia, cf$mut_b))
```

```
##
```

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

```
##
```

```
## data: table(cf$anemia, cf$mut_b)
```

```
## X-squared = 1.0141, df = 2, p-value = 0.6023
```

Questão (6)

```
> (ctgA <- table(cf$hipoptr, cf$mut_a))
```

```
##  
##          aa Aa AA  
## Não 15  4  0  
## Sim 14 13  4
```

```
> chisq.test(ctgA)
```

```
##  
## Pearson's Chi-squared test  
##  
## data:  ctgA  
## X-squared = 6.281, df = 2, p-value = 0.04326
```

```
> (ctgB <- table(cf$hipoptr, cf$mut_b))
```

```
##  
##          bb Bb BB  
## Não  13   5   1  
## Sim  28   3   0
```

```
> chisq.test(ctgB)
```

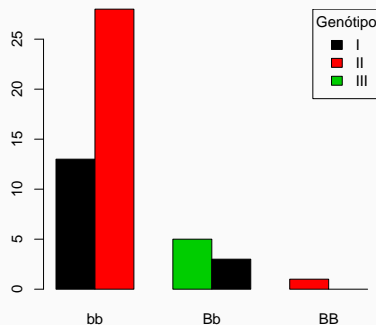
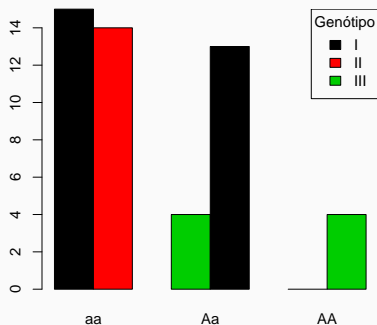
```
##  
## Pearson's Chi-squared test  
##  
## data:  ctgB  
## X-squared = 4.3589, df = 2, p-value = 0.1131
```



```

> par(mfrow = c(1, 2))
>
> barplot(ctgA, beside=T, col=1:3)
> legend("topright", title="Genótipo", rownames(ctg), fill = 1:3)
>
> barplot(ctgB, beside=T, col=1:3)
> legend("topright", title="Genótipo", rownames(ctg), fill = 1:3)

```



```
> fisher.test(table(cf$hipoptr, cf$mut_a != "AA"))
```

```
##
```

```
## Fisher's Exact Test for Count Data
```

```
##
```

```
## data: table(cf$hipoptr, cf$mut_a != "AA")
```

```
## p-value = 0.2839
```

```
## alternative hypothesis: true odds ratio is not equal to 1
```

```
## 95 percent confidence interval:
```

```
## 0.000000 2.416886
```

```
## sample estimates:
```

```
## odds ratio
```

```
## 0
```

```
> fisher.test(table(cf$hipoptr, cf$mut_a == "aa"))
```

```
##
```

```
## Fisher's Exact Test for Count Data
```

```
##
```

```
## data: table(cf$hipoptr, cf$mut_a == "aa")
```

```
## p-value = 0.03738
```

```
## alternative hypothesis: true odds ratio is not equal to 1
```

```
## 95 percent confidence interval:
```

```
## 0.04431456 0.93039597
```

```
## sample estimates:
```

```
## odds ratio
```

```
## 0.2265456
```

```
> fisher.test(table(cf$hipoptr, cf$mut_b != "BB"))
```

```
##
```

```
## Fisher's Exact Test for Count Data
```

```
##
```

```
## data: table(cf$hipoptr, cf$mut_b != "BB")
```

```
## p-value = 0.38
```

```
## alternative hypothesis: true odds ratio is not equal to 1
```

```
## 95 percent confidence interval:
```

```
## 0.04183514 Inf
```

```
## sample estimates:
```

```
## odds ratio
```

```
## Inf
```

```
> fisher.test(table(cf$hipoptr, cf$mut_b == "bb"))
```

```
##
```

```
## Fisher's Exact Test for Count Data
```

```
##
```

```
## data: table(cf$hipoptr, cf$mut_b == "bb")
```

```
## p-value = 0.06705
```

```
## alternative hypothesis: true odds ratio is not equal to 1
```

```
## 95 percent confidence interval:
```

```
## 0.7508666 29.9033941
```

```
## sample estimates:
```

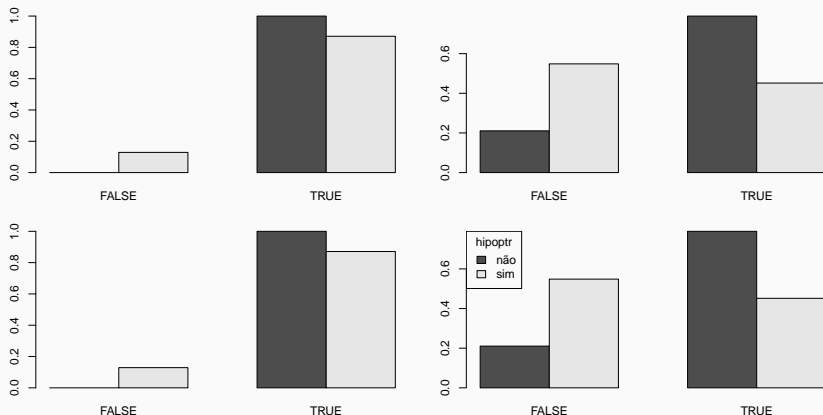
```
## odds ratio
```

```
## 4.171726
```

```

> par(mfrow = c(2,2), mar = rep(2, 4))
> barplot(prop.table(table(cf$hipoptr, cf$mut_a != "AA"), 1), beside = T)
> barplot(prop.table(table(cf$hipoptr, cf$mut_a == "aa"), 1), beside = T)
> barplot(prop.table(table(cf$hipoptr, cf$mut_a != "AA"), 1), beside = T)
> barplot(prop.table(table(cf$hipoptr, cf$mut_a == "aa"), 1), beside = T)
> legend("topleft", c("não", "sim"), title = "hipoptr", fill = grey.colors(2)

```



Ao Final...

O que vimos?

- Aprendemos a reconhecer variáveis categóricas.
- Identificar quando e como aplicar teste binomial.
- Identificar quando e como aplicar teste de independência.
- Saber reconhecer qual teste de independência mais adequado. (**Fisher** ou **Qui-quadrado**).
- Reconhecer a pergunta estatística de cada teste.
- Ilustrar os resultados.

Até a próxima
