

## Trabalho No.1

14 de agosto de 2023

Redigir de maneira individual e entregar na área correspondente no sistema **Microsoft Teams** um relatório eletrônico até o dia **04 de setembro de 2023**.

### Capítulo 1- Dados Categóricos: introdução

- 1- Rous et al. (2005) coletaram amostras de urina de 750 gestantes assintomáticas em Rotterdam, Holanda, para estimar a prevalência de clamídia (chlamydia) na população correspondente. Das 750 amostras, 48 testaram positivo para a doença. Com essas informações, responda:
  - a) Descreva as suposições que seriam necessárias para usar uma distribuição binomial para a amostra.
  - b) Presumindo que as suposições sejam satisfeitas, encontre um intervalo de confiança para estimar a prevalência de clamídia. Use o procedimento de intervalo de confiança mais apropriado para este problema e interprete os resultados.
- 2- O Aptima Combo 2 Assay da Gen-Probe é usado para detectar simultaneamente clamídia e gonorreia em indivíduos. Cada linha do conjunto de dados contém informações sobre a precisão do teste para combinações específicas de doenças (clamídia, gonorreia), sexo (masculino, feminino), tipo de amostra (swab, urina) e status dos sintomas (sintomáticos, assintomáticos).

A primeira linha de dados, juntamente com alguns cálculos adicionais, são fornecidos abaixo:

```
aptima <- read.table( file = "http://leg.ufpr.br/~lucambio/CE073/202225/Aptima_combo.csv",
                    header = TRUE , sep = ",")
c.table <- array ( data = c( aptima [1 ,6] , aptima [1 ,7],
                          aptima [1 ,9] , aptima [1 ,8] ) , dim = c(2 ,2) ,
                  dimnames = list ( True = c("+", "-") , Assay = c("+", "-")))
c.table
##      Assay
## True  +   -
##   + 190   7
##   -  15 464
Se.hat <- c.table [1 ,1]/ sum(c.table [1 ,]) # Sensitivity
Sp.hat <- c.table [2 ,2]/ sum(c.table [2 ,]) # Specificity
data.frame (Se.hat , Sp.hat)
##      Se.hat  Sp.hat
## 1 0.964467 0.9686848
```

Listados na primeira linha do conjunto de dados, havia 676 homens sintomáticos que forneceram amostras de swab para testes de clamídia. Destes 676 indivíduos, 190 testaram positivo e foram considerados verdadeiramente positivos. Observe que os resultados de uma combinação de outros procedimentos de teste foram usados para fazer o diagnóstico verdadeiro. Da mesma forma, 464 indivíduos foram verdadeiramente negativos e testaram negativo. Houve 22 erros com 15 testes positivos para indivíduos verdadeiramente negativos e 7 testes negativos para indivíduos verdadeiramente positivos.

O objetivo deste problema é calcular intervalos de confiança para medidas de acurácia de sensibilidade  $S_e$  e especificidade  $S_p$ . A sensibilidade é definida como a proporção de indivíduos verdadeiramente positivos com teste positivo e a especificidade é definida como a proporção de indivíduos verdadeiramente negativos com teste negativo. Idealmente, é desejável ter  $S_e$  e  $S_p$  o mais próximo possível de 1. As estimativas pontuais,  $\hat{S}_e$  e  $\hat{S}_p$ , para essas medidas são fornecidas na saída. Usando este plano de fundo, complete o seguinte:

- a) Calcule os intervalos de Clopper-Pearson para  $S_e$  e  $S_p$  entre os homens sintomáticos que forneceram amostras de swab para teste de clamídia. Interprete os intervalos.
- b) Calcule os intervalos de Clopper-Pearson para as outras combinações de doença, gênero, espécime e sintoma e exiba-os de maneira organizada, por exemplo, eles podem ser colocados em um quadro de dados com a rotulagem apropriada.

- 3- Aseffa et al. (1998) examinam a prevalência de HIV entre mulheres que visitam clínicas de saúde no noroeste da Etiópia. Juntamente com o teste de HIV, informações adicionais foram coletadas sobre cada mulher, como o uso de preservativos. A tabela abaixo fornece uma tabela de contingência resumindo os dados.

TABELA: Dados sobre HIV e uso de preservativos na Etiópia. A fonte de dados é Aseffa et al. (1998).

		HIV		
		Positivo	Negativo	Total
Preservativo	Nunca	135	434	569
	Sempre	15	9	24
Total		150	443	539

- a) Calcule os intervalos de confiança de Wald e Agresti-Caffo para a diferença nas probabilidades de ser HIV positivo com base no uso de preservativo. Interprete os intervalos.
- b) Realize um teste score, teste qui-quadrado de Pearson e LRT para testar a igualdade das probabilidades de sucesso.
- c) Estime a razão de chances e calcule o intervalo de confiança correspondente para ela. Interprete a estimativa e o intervalo.
- d) Geralmente, pensa-se que o uso do preservativo ajuda a prevenir a transmissão do HIV. Os resultados aqui concordam com isso? Se não, quais fatores podem ter levado a esses resultados? Observe que Aseffa et al. (1998) fornecem a razão de chances estimada e um intervalo de Wald correspondente, mas falham em interpretá-los.

- 4- Semelhante ao intervalo de Wilson para  $\pi$ , um intervalo para  $\pi_1 - \pi_2$  pode ser encontrado usando uma estatística de teste de um teste de hipótese para os parâmetros. Considere o teste  $H_0 : \pi_1 - \pi_2 = d$  vs.  $H_1 : \pi_1 - \pi_2 \neq d$  onde  $d$  denota o valor hipotético da diferença  $\pi_1 - \pi_2$ . O intervalo é o conjunto de todos os valores possíveis de  $d$  tais que

$$\frac{|\hat{\pi}_1 - \hat{\pi}_2 - d|}{\sqrt{\hat{\pi}_1^{(0)}(1 - \hat{\pi}_1^{(0)})/n_1 + \hat{\pi}_2^{(0)}(1 - \hat{\pi}_2^{(0)})/n_2}} < Z_{1-\alpha/2} \quad (1)$$

seja satisfeito.

Observe que  $\hat{\pi}_1^{(0)}$  e  $\hat{\pi}_2^{(0)}$  denotam os MLEs de  $\pi_1$  e  $\pi_2$  sob a restrição de que  $\pi_1 - \pi_2 = d$ . Esse intervalo é chamado de intervalo score para a diferença de duas probabilidades.

Infelizmente, não há solução de forma fechada e procedimentos numéricos iterativos devem ser usados para encontrar os limites do intervalo de confiança. A função **diffscoreci()** do pacote **PropCIs** executa esses cálculos. Calcule este intervalo score para os dados de Larry Bird. Compare o intervalo com os intervalos Wald e Agresti-Caffo para ele. Considere  $d = 0$ .

- 5- Mostre que  $OR = 1$  quando  $\pi_1 = \pi_2$ .