Segunda Avaliação

Estatística não paramétrica

Paulo Ricardo Seganfredo Campana

8 de março de 2024

Questão 1. A uma amostra aleatória de 100 adultos foi inquirido sobre sua opinião a respeito do aborto devido a pobreza, antes e depois de um debate realizado sobre o assunto. Os resultados estão na tabela seguinte. Use estes resultados e $\alpha = 0{,}05$ para testar se o debate mudou a opinião da população sobre o aborto devido a pobreza.

Vou usar o teste de McNemar para testar se as probabilidades de mudança de opinião de ambos os lados são as mesmas ou não

$$\begin{cases} \mathcal{H}_0 : P(\text{contra antes, a favor depois}) = P(\text{a favor antes, contra depois}) \\ \mathcal{H}_1 : P(\text{contra antes, a favor depois}) \neq P(\text{a favor antes, contra depois}) \end{cases}$$

$$\text{Estatística do teste:} \qquad \frac{(|b-c|-1)^2}{b+c} = \frac{(|30-8|-1)^2}{30+8} = \frac{21^2}{38} = 11.61$$

```
# Quantil da distribuição
qchisq(0.95, 1)
## [1] 3.841
```

A estatística do teste observada é um valor mais extremo doque o quantil de 95% da distribuição em $\mathcal{H}_0$, portanto rejeitamos a hipótese nula então o debate mudou a opinião dos adultos.

Questão 2. Numa classe de 24 alunos, comparou-se o rendimento de estudantes provenientes de escolas particulares e escolas públicas. Os resultados seguem abaixo: Existe diferença entre os rendimentos, a depender do tipo de escola? Use $\alpha = 0{,}05$.

Para verificar a independência das variáveis de rendimento e tipo de escola, irei usar o teste exato de fisher devido ao pequeno tamanho da amostra. O teste é bilateral:

$$\begin{cases} \mathcal{H}_0 : P(\text{acima da média, particular}) = P(\text{acima da média, pública}) \\ \mathcal{H}_1 : P(\text{acima da média, particular}) \neq P(\text{acima da média, pública}) \end{cases}$$

```
p <- function(a, b, c, d) {
    choose(a + b, a) * choose(c + d, c) / choose(a + b + c + d, a + c)
}

2 * (p(5, 7, 10, 2) + p(4, 8, 11, 1) + p(3, 9, 12, 0))
## [1] 0.08938
```

O p-valor do teste é superior ao nível de significancia de 5%, portanto não rejeitamos a hipótese de que o rendimento dos alunos de escolas públicas e particulares são as mesmas.

Questão 3. Um grupo de 8 pacientes se submete a um tratamento por meio de estímulos. Na tabela seguinte estão as medidas de pressão sanguínea (em mm/Hg) antes e depois do tratamento. O pesquisador deseja saber se os estímulos aumentam a pressão sanguínea. Use $\alpha = 0{,}05$.

```
q3 <- data.frame(
    antes = c(118, 120, 128, 124, 130, 136, 128, 140),
    depois = c(127, 128, 136, 131, 135, 138, 125, 136)
)
```

Usarei um teste unilateral de Wilcoxon pois os dados são quantitativos, para testar se existe diferença nas medianas das amostras.

$$\begin{cases} \mathcal{H}_0 : \tilde{\mu}_D \leq 0 \\ \mathcal{H}_1 : \tilde{\mu}_D > 0 \end{cases}$$

```
q3$diff <- q3$depois - q3$antes #  9   8   8   7   5   2  -3  -4
q3$sign <- sign(q3$diff)        #  +   +   +   +   +   +   -   -
q3$rank <- rank(abs(q3$diff))   #  8 6.5 6.5   5   4   1   2   3
sum(q3$rank[q3$sign == -1])
## [1] 5
```

O quantil tabelado para $n = 8$ e $\alpha = 0.05$ é 6, a estatística do teste observada é $T^- = 5$, menor que o quantil, rejeitamos a hipótese nula e assim os estimulos aumentam a pressão sanguínea.

Questão 4. Considere as seguintes amostras aleatórias: Testar a hipótese de que as correspondentes populações têm a mesma mediana. Use um nível de 5% de significância.

```r
q4 <- data.frame(
    valor = c(
        18, 35, 53, 48, 60, 40,
        73, 20, 70, 68, 82, 42, 30, 62, 88, 32
    ),
    variavel = c(rep("X", 6), rep("Y", 10))
)
```

Já que o tamanho das duas amostras são diferentes, usarei o teste bilateral de Wilcoxon-Mann-Whitney

Como X é a amostra de menor tamanho, a estatítica do teste é a soma dos postos para X da amostra combinada.

$$\begin{cases} \mathcal{H}_0 : \tilde{\mu}_X = \tilde{\mu}_Y \\ \mathcal{H}_1 : \tilde{\mu}_X \neq \tilde{\mu}_Y \end{cases}$$

```r
q4$rank <- rank(q4$valor)
sum(q4$rank[q4$variavel == "X"])
## [1] 39
```

Os valores críticos tabelados do teste para $\alpha = 0.05$, $n_1 = 6$ e $n_2 = 10$ são 32 e 70, a estatística do teste observada é 39 que pertence a este intervalo, assim não rejeitamos a hipótese de que as duas amostras tenham mesma mediana.

Questão 5. Um pesquisador deseja saber se duas regiões de uma mesma imagem apresentam a mesma distribuiçao de valores (desconhecida). Para testar esta hipótese, amostrou-se 15 pontos independentes de cada região. Os valores observados são apresentados na tabela abaixo. O que se conclui a partir destes valores? Use um nível de 5% de significância.

```r
q5 <- data.frame(
    região1 = c(81, 78, 61, 89, 69, 58, 64, 84, 89, 83, 88, 56, 87, 95, 75),
    região2 = c(56, 55, 76, 54, 83, 97, 85, 66, 78, 80, 61, 69, 71, 55, 91)
)
```

Para testar se as duas amostras tem mesma distribuição usarei o teste de Smirnov

$$\begin{cases} \mathcal{H}_0 : \hat{F}_x(\omega) = \hat{G}_x(\omega), \quad \forall \omega \\ \mathcal{H}_1 : \exists \omega \quad \hat{F}_x(\omega) \neq \hat{G}_x(\omega) \end{cases}$$

```r
ks.test(q5$região1, q5$região2)
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  q5$região1 and q5$região2
## D = 0.27, p-value = 0.6
## alternative hypothesis: two-sided
```

O p-valor do teste de Kolmogorov-smirnov é bastante alto, não rejeitamos a hipótese de que as duas amostras possuem a mesma distribuição.

Questão 6. A tabela seguinte representa o consumo total de carne (em kg), por pessoas adultas, em diferentes meses do ano. Testar se o consumo é o mesmo nos 4 meses considerados. Use um nível de 5% de significância.

```r
q6 <- data.frame(
    fev = c(4.6, 4.9, 5.0, 4.8, 4.7),
    mai = c(4.7, 4.4, 4.3, 4.4, 4.1),
    ago = c(4.8, 4.7, 4.6, 4.5, 4.6),
    nov = c(4.9, 5.2, 5.4, 5.1, 5.6)
)
```

Como o tamanho de amostra dos 4 grupos são iguais, usarei o teste de Friedman

$$\begin{cases} \mathcal{H}_0 : \tilde{\mu}_{\text{fev}} = \tilde{\mu}_{\text{mai}} = \tilde{\mu}_{\text{ago}} = \tilde{\mu}_{\text{nov}} \\ \mathcal{H}_1 : \tilde{\mu}_i \neq \tilde{\mu}_j, \quad \text{para pelo menos 2 meses} \end{cases}$$

```r
friedman.test(as.matrix(q6))
##
##  Friedman rank sum test
##
## data:  as.matrix(q6)
## Friedman chi-squared = 12, df = 3, p-value = 0.007
```

O p-valor do teste é inferior ao nível de significância de 5%, portanto rejeitamos a hipótese nula de que o consumo de carne é igual em todos os meses.

Questão 7. Obtiveram-se amostras aleatórias e independentes de três marcas diferentes I, II e III de certo tipo de aparelho eletrônico, e registrou-se o tempo, em horas, de funcionamento contínuo e sem reparos. Os dados estão na tabela seguinte.

```r
q7 <- data.frame(
    I   = c(48, 67, 5 , 53, 36),
    II  = c(60, 2 , 55, 33, 49),
    III = c(31, 98, 71, 42, 59)
)
```

De maneira similar, os testes de Friedman e Kruskal-Wallis podem ser usados para testar se há diferenças entre grupos.

```r
friedman.test(as.matrix(q7))$p.value
## [1] 0.5488
```

```r
kruskal.test(q7)$p.value
## [1] 0.5434
```

Ambos os testes tem p-valores parecidos e altos, oque indica que não há diferenças estatísticamente significativas na mediana entre os grupos.