

Aula 9 - Testes Estatísticos

Na nossa nona aula, começamos com um momento de discussão, esclarecimento de dúvidas e revisão de conceitos vistos nas aulas passadas!

Testes para k amostras emparelhadas

Teste Q de Cochran

É utilizado para dados binários, testando se as proporções são iguais ou se pelo menos uma delas é diferente

$$\begin{cases} H_0: p_1 = p_2 = \dots = p_k \\ H_1: \text{pelo menos um } \neq \end{cases}$$

O teste em R, utilizando a base dados escola.xlsx

```
cochran.qtest(Resposta ~ Pratica | Aluno, data = escola)
```

Cochran's Q test

```
data: Resposta by Pratica, block = Aluno
Q = 16.9535, df = 3, p-value = 0.0007225
alternative hypothesis: true difference in probabilities is not equal to 0
sample estimates:
proba in group      <NA>      <NA>      <NA>
0.1428571      0.5714286      0.7857143      0.8571429
```

Teste de Friedman

É utilizado quando temos dados ordinais, testando se as medianas são iguais ou se pelo menos uma é diferente.

Consiste em uma alternativa a ANOVA para dados emparelhados quando não atende o pressuposto!!

$$\begin{cases} H_0: \text{mediana}_1 = \text{mediana}_2 = \dots = \text{mediana}_k \\ H_1: \text{pelo menos um é } \neq \end{cases}$$

$$\begin{cases} H_0: \mu_1 = \mu_2 = \dots = \mu_k \\ H_1: \text{pelo menos um é } \neq \end{cases}$$

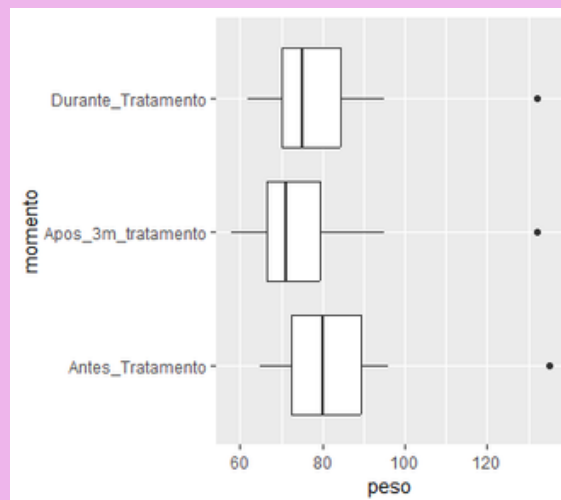
O teste em R, utilizando a base dados tratamentos2.xlsx

```
friedman.test(peso ~ momento | ID, data = tratamentos)
```

Friedman rank sum test

data: peso and momento and ID

Friedman chi-squared = 27.527, df = 2, p-value = 1.053e-06



Testes para k amostras independentes

Teste Qui-quadrado

É utilizado para dados nominais ou ordinais (sendo mais interessante para os nominais), para entender se há diferenças significativas entre frequências e, assim, verificar se há associação ou não.

$\begin{cases} H_0: \text{não há diferença significativa entre as frequências observadas e esperadas (não há associação)} \\ H_1: \text{há diferença significativa entre as frequências observadas e esperadas (há associação)} \end{cases}$

O teste em R, utilizando a base dados produtividade.xlsx

```
chisq.test(produtividade$Produtividade, produtividade$Turno)
```

Pearson's Chi-squared test

```
data: produtividade$Produtividade and produtividade$Turno
X-squared = 13.143, df = 6, p-value = 0.04082
```

Teste Kruskal Wallis

É utilizado quando temos dados ordinais, testando se as medianas são iguais ou se pelo menos uma delas é diferente. Consiste em uma alternativa a ANOVA para dados independentes quando não atende o pressuposto!!

$$\begin{cases} H_0: \text{mediana}_1 = \text{mediana}_2 = \dots = \text{mediana}_k \\ H_1: \text{pelo menos um é } \neq \end{cases}$$

$$\begin{cases} H_0: \mu_1 = \mu_2 = \dots = \mu_k \\ H_1: \text{pelo menos um é } \neq \end{cases}$$

O teste em R, utilizando a base dados produtividade.xlsx

```
kruskal.test(peso ~ tratamento, data = Peso_Planta)
```

```
Kruskal-Wallis rank sum test

data:  peso by tratamento
Kruskal-Wallis chi-squared = 2.1053, df = 2, p-value = 0.349
```

Medidas de Associação

Teste Qui-quadrado

Ele traz um "sentimento" de associação. Então, quanto menor o p-valor, mais eu rejeito H0 e assim, apresenta uma maior associação.

No entanto, existe crítica!

$$\begin{cases} H_0: \text{não há diferença significativa entre as frequências observadas e esperadas (não há associação)} \\ H_1: \text{há diferença significativa entre as frequências observadas e esperadas (há associação)} \end{cases}$$

Coeficiente Phi

Utiliza a estatística do teste Qui-quadrado e penaliza pelo número de indivíduos da tabela!

$$\text{Phi} = \sqrt{\frac{\chi^2}{n}}$$

Para que phi varie apenas entre 0 e 1, é necessário que a tabela de contingência seja 2x2.

Quanto maior o valor de phi maior a associação entre duas variáveis.

Coeficiente de contingência

Quando C máximo, significa que existe associação perfeita entre as variáveis, porém, esse limite nunca assume o valor 1.

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

O coeficiente de contingência tem como limite inferior o valor 0, indicando que não existe relação entre as variáveis, porém o limite superior varia em função do número de categorias, ou seja:

$$0 \leq C \leq \sqrt{\frac{q-1}{q}} \quad q = \min(I, J)$$

Coeficiente V de Cramer

O valor 0 indica que as variáveis não tem nenhum tipo de associação e o valor 1 revela que elas são perfeitamente associadas.

$$V = \sqrt{\frac{\chi^2}{n(q-1)}} \quad q = \min(I, J)$$

Quando a tabela de contingência é 2x2 o V acaba sendo phi. Este é uma alternativa ao coeficiente de phi, ao coeficiente de contingência e seu valor está sempre limitado no intervalo [0,1], independente da quantidade de categorias nas linhas e colunas.

Coeficiente de contingência

O valor 0 indica que não existe uma associação entre as variáveis X e Y, o valor 1 indica uma associação positiva perfeita e o valor -1 associação negativa perfeita.

$$r_{sp} = 1 - \frac{6 \cdot \sum_{k=1}^n d_k^2}{n(n^2-1)}$$

É uma medida de associação para variáveis ordinais, assim, é baseada na ordenação, onde criam-se rankings.



Lembre-se da tarefa de casa!!