Tarefa 5
Outras medidas de associação

Douglas Rodrigues
Karina Yaginuma

Universidade Federal Fluminense

- Leia o conteúdo e faça os exercícios dos slides.
- Você deve entregar:
  1. Um relatório, em fomato pdf, contendo as respostas dos questionamentos feitos ao longo da tarefa.
  2. Os comandos utilizados.

Resumo

Nas aulas anteriores estudamos algumas medidas de associações entre:

- duas variáveis quantitativas
  - Coeficiente de correlação de Pearson (linear);
  - Correlação de postos de Spearman;
  - Correlação de Kendall.
- Duas variáveis qualitativas
  - Qui-Quadrado;
  - Coeficiente de Contingência (Modificado).

Outras medidas de associação

Nesta tarefa, vamos estudar outros tipos de medidas de associação entre

- variáveis qualitativas ordinais;
- uma variável qualitativa e uma variável quantitativa.

Variáveis qualitativas ordinais

- Já estudamos medidas de associação entre variáveis qualitativas (ordinais e nominais).
- Agora vamos ver como podemos utilizar as correlações de Spearman e de Kendall para variáveis qualitativas ordinais.
- Como as variáveis ordinais possuem uma ordem natural entre as categorias, podemos usar as correlações de Spearman e Kendall que são baseadas na ordenação dos valores.

Cálculo das correlações

- O comando para o cálculo das correlações é idêntico ao utilizado para as variáveis quantitativas.
- Basta alterar as categorias das variáveis por valores numéricos respeitando a ordem das categorias.
- **Exemplo:** Variáveis Grau de Instrução
  - Fundamental $\longleftarrow$ 1
  - Médio $\longleftarrow$ 2
  - Superior $\longleftarrow$ 3

Exercício 1

- Importe os dados do arquivo `Funcionarios.xlsx`. Faça as modificações necessárias para importar os dados corretamente.
- Use a função `factor`, para definir `x` um fator das observações da variável Grau de Instrução.

```
> x <- factor(Funcionarios$`Grau de Instrução`)
> levels(x) # fornece as categorias da variável e
sua ordenação.
```

Exercício 1

- Faça o mesmo procedimento para a variável Classe Social.

```
> y <- factor(Funcionarios$‘Classe Social‘)
> levels(y) # fornece as categorias da variável e
sua ordenação.
```

- O **R** faz a ordenação de maneira automatica. No caso da variável Classe Social, a ordenação das categorias da variável é `"A"`, `"B"`e `"C"`, mas como sabemos que a ordem crescente das categorias é `"C"`, `"B"`e `"A"`. Faça a correção da ordem das categorias, utilizando o argumento `levels`.

```
> y <- factor(Funcionarios$‘Classe Social‘,
levels =c("C", "B", "A"))
```

Exercício 1

- Agora podemos transformar as variáveis `x` e `y` em valores numéricos. Para itsto, podemos utilizar a função `as.numeric`.

```
> x <- as.numeric(x)
> y <- as.numeric(y)
```

- Como tivemos o cuidado de ordenar as categorias corretamente, no momento que transformamos fatores em valores numéricos, a transformação respeita a ordenação do fator.
- Agora podemos calcular a correlação entre as variáveis ordinais Grau de Instrução e Classe Social

```
> cor(x,y,method = "spearman")
> cor(x,y,method = "kendall")
```

Exercício 1

- Agora podemos calcular a correlação entre as variáveis ordinais Grau de Instrução e Classe Social

```
> cor(x,y,method = "spearman")
> cor(x,y,method = "kendall")
```

- O que podemos concluir dos valores obtidos?

Coeficiente de correlação eta (η)

- O coeficiente de correlação eta (η) é um coeficiente de associação não linear.
- Ele mede a intensidade de associação entre variável dependente Y (quantitativa) e independente X (qualitativa).
- Assume valores no intervalo $(0, 1)$.
- η^2 é a percentagem da variabilidade da variável dependente explicada pela variável independente.

Coeficiente de correlação eta (η)

- Para calcular o coeficiente η, assumimos que a variável Y é uma função da variável X, ou seja, $Y = f(X)$.
- O coeficiente η é dado por

$$\eta^2 = \frac{Var(E(f|X))}{Var(f)}.$$

- Ideia dos valores da equação:
  - $Var(E(f|X))$ é a variância da média de Y dado que conhecemos a informação da variável X.
  - $Var(f)$ é a variância calculada sem a informação da variável X.

Cálculo do η^2 no **R**

- Para calcular η^2 no **R**, é preciso instalar o pacote `sjstats`.

```
> install.packages("sjstats")
> library("sjstats")
```

- Use o comando `eta_sq()` para calcular η^2.
- O argumento da função é a relação $Y = f(X)$, que pode ser obtido utilizando o comando `aov()`. Veremos com mais detalhe a utilização do comando no próximo exercício.

Exercício 2

- Vamos utilizar o banco de dados do pacote `sjstats`, chamado `efc`.
- `efc` é um banco de dados do SPSS, contém informações sobre indivíduos idosos de uma instituição de saúde.
- Para acessar o banco de dados, use o comando

```
> library("sjstats")
> data(efc)
```

- Use o comando `str`, para entender a estrutura do banco de dados.

```
> str(efc)
```

Exercício 2

- Suponha que estamos interessados em estudar a relação entre as variáveis `c12hour` e `e16sex`, entre as variáveis `c12hour` e `e42dep`. Sendo
  - `c12hour`: média do número de horas de cuidados por semana;
  - `e16sex`: sexo do indivíduo (1-Masculino e 2-Feminino);
  - `e42dep`: grau de dependência do indivíduo (1-independente, 2-levemente dependente, 3-dependência moderada e 4-dependência severa)
- Crie um data frame com apenas essas três variáveis, nomeie-o `dados`. Altere os nomes das variáveis por `tempo,` `sexo` e `dependencia` (Dica: use os comandos do pacote `dplyr`).

Exercício 2

- Agora precisamos encontrar a função que relaciona as variáveis

  `tempo = função de sexo.`

  Para isto, use o seguinte comando,

```
> ajuste1 <- aov(tempo ~ sexo, data = dados)
```

- Agora é só utilizar o comando `eta_sq`

```
> eta_sq(ajuste1)
  term etasq
1 sexo 0.004
```

Exercício 2

- Podemos concluir que apenas 0.4% da variabilidade da variável tempo é explicada pela variável sexo. Ou seja, temos evidência de que não existe uma relação entre as duas variáveis.
- Repita o mesmo procedimento para calcular o η^2 para as variáveis `tempo` e `dependencia`. O que podemos concluir desta análise?