

TESTES DE CORRELAÇÃO

Os coeficientes de correlação são métodos estatísticos para se **medir as relações entre variáveis** e o que elas representam.

O que a correlação procura entender é como uma variável se comporta em um cenário onde outra está variando, visando identificar se existe alguma relação entre a variabilidade de ambas. Embora não implique em causalidade, o coeficiente de correlação exprime em números essa relação, ou seja, quantifica a relação entre as variáveis.

Nesse trabalho iremos acompanhar um teste de correlação que mede a relação da idade dos usuários com o salário afim de medir a correlação entre essas variáveis por meio do banco de dados usuários utilizando o software R.

Principais testes de correlação:

Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson (r), também chamado de correlação linear ou **r de Pearson**, é um grau de relação **entre duas variáveis quantitativas** e exprime o grau de correlação através de valores situados entre -1 e 1.

Coeficiente de Correlação de Spearman

Denominado pela letra grega *rho* (ρ), o coeficiente de correlação de postos de Spearman é uma medida de correlação não paramétrica também avaliado no intervalo entre -1 e 1.

Ao contrário do coeficiente de Pearson, o coeficiente de Spearman **não exige a suposição de que a relação entre as variáveis seja linear, nem requer que as mesmas sejam quantitativas** – pode inclusive ser utilizado para verificar relação entre variáveis medidas no nível ordinal.

Coeficiente de Correlação de Kendall

Expresso pela letra grega *tau* (τ), o coeficiente de **correlação de Kendall é uma medida de associação para variáveis ordinais**. Uma vantagem de *tau* sobre o coeficiente de Spearman é a possibilidade de ser generalizado para um coeficiente de correlação parcial.

Correlação de Pearson

Sejam duas variáveis X e Y , ambas quantitativas, preferencialmente contínuas.

Correlação de Spearman

Utilizado quando não existe normalidade e/ou não existe relação linear, Ideal quando temos variáveis medidas apenas em uma escala ordinal.

Correlação de Kendall

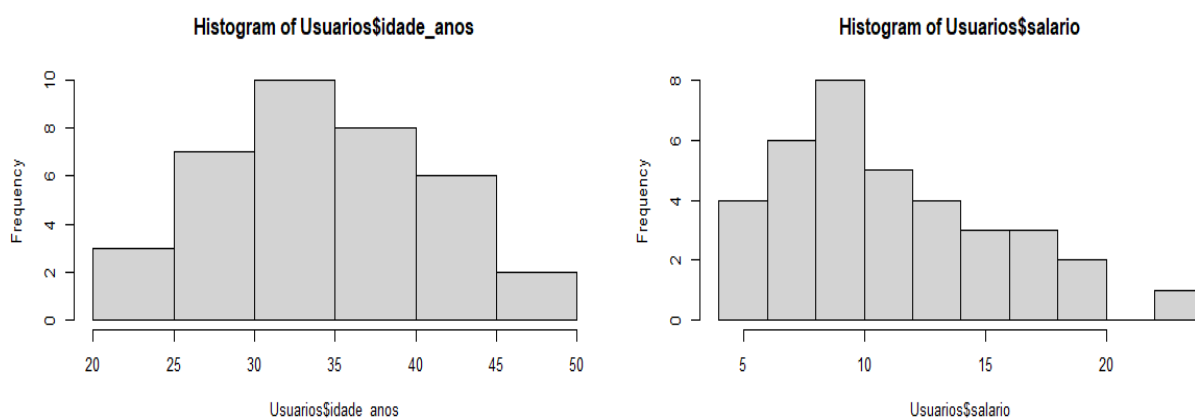
O coeficiente de correlação Tau de Kendall serve para verificar se existe correlação entre duas variáveis ordinais. É um método adequado quando amostras têm tamanhos reduzidos, pois o método é mais preciso

ESTATÍSTICA DESCRITIVA

```
> summary(Usuarios)
      ...1      N      estado_civil      grau_instrucao
Min.   : 1.00   Min.   : 1.00   Length:36   Length:36
1st Qu.: 9.75   1st Qu.: 9.75   Class :character   Class :character
Median :18.50   Median :18.50   Mode  :character   Mode  :character
Mean   :18.50   Mean   :18.50
3rd Qu.:27.25   3rd Qu.:27.25
Max.   :36.00   Max.   :36.00

      n_filhos      salario      idade_anos      idade_meses
Min.   :0.00   Min.   : 4.000   Min.   :20.00   Min.   : 0.000
1st Qu.:1.00   1st Qu.: 7.553   1st Qu.:30.00   1st Qu.: 3.750
Median :2.00   Median :10.165   Median :34.50   Median : 6.000
Mean   :1.65   Mean   :11.122   Mean   :34.58   Mean   : 5.611
3rd Qu.:2.00   3rd Qu.:14.060   3rd Qu.:40.00   3rd Qu.: 8.000
Max.   :5.00   Max.   :23.300   Max.   :48.00   Max.   :11.000
NA's   :16
reg_procedencia
Length:36
Class :character
Mode  :character
```

DISTRIBUIÇÃO DOS DADOS

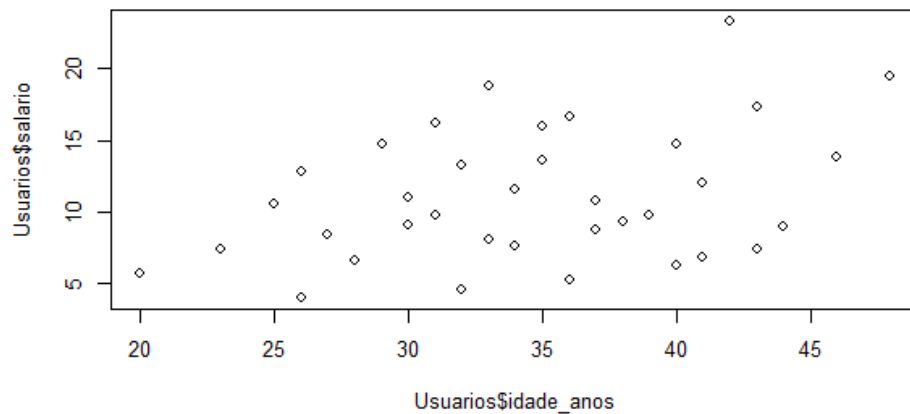


Podemos visualizar todas as estatísticas descritivas do banco de dados temos na variável idade uma média de 34 anos e na variável salário uma média de R\$ 11.122,00 assim como podemos verificar a sua distribuição através dos gráficos de histograma que mostra uma maior dispersão dos dados na variável salário e uma distribuição simétrica na variável idade.

Porém para esse trabalho em específico desejamos verificar apenas o grau de correlação das variáveis idade e salário. A seguir veremos a aplicação do teste e suas interpretações.

GRÁFICOS DE DISPERSÃO

IDADE X SÁLARIO



Podemos verificar através do gráfico acima que os dados estão dispostos de forma meio aleatório porém com uma leve tendência a subir conforme a idade aumenta , porém não é possível medir essa correlação somente com a leitura do gráfico vamos aplicar um teste de correlação para verificar se existe ou não uma correlação assim como sua força.

TESTE DE CORRELAÇÃO

Coeficiente de Correlação de Pearson

Sejam duas variáveis X e Y, ambas quantitativas, preferencialmente contínuas. A existência de relação linear entre essas variáveis pode ser detectada com auxílio do Diagrama de Dispersão, mas, também, com auxílio do Coeficiente de Correlação Linear de Pearson.

RESULTADO DO TESTE:

IDADE X SÁLARIO

```
> cor.test(x = Usuarios$idade_anos,  
+ y = Usuarios$salario,  
+ method = "pearson")  
  
Pearson's product-moment correlation  
  
data: Usuarios$idade_anos and Usuarios$salario  
t = 2.2742, df = 34, p-value = 0.02938  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 0.03954753 0.61810957  
sample estimates:  
      cor  
0.3633622
```

Através das análises do resultado do teste temos que seu ***p valor é = 0,02938*** um valor menor que **0,05** o que podemos aceitar a hipótese alternativa de que a correlação não é igual a 0, ou seja, existe sim uma correlação entre as variáveis. Esse grau de correlação se é forte ou fraco será dimensionado pelo valor da cor.

Temos um valor de **correlação de 0,3633622** o que nos indica que há uma relação linear moderada entre a idade e o salário.

ANA PAULA DE SOUZA VANDERLEY

ESPECIALISTA EM ESTATÍSTICA

2022

