

Associações entre Variáveis

Objetivo: descrever simultaneamente a variabilidade de duas ou mais variáveis de forma que cada conjunto de variáveis é observado para uma mesma unidade observational (pessoas, animal, planta, peça, etc).

Vamos iniciar com um par de variáveis (x, y) , sendo $(x_i, y_i), i=1, \dots, n$ pares de observações de duas variáveis:

| x | y |
|--------------|--------------|
| Qualitativa | Qualitativa |
| " | Quantitativa |
| Quantitativa | Qualitativa |

Para cada combinação de pares de variáveis serão um tipo de análise a ser realizada.

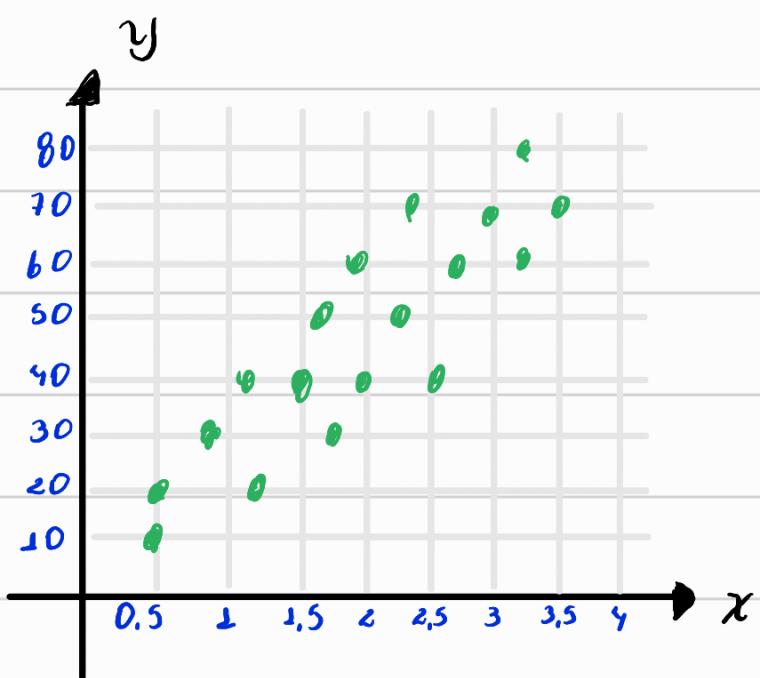
Para estudarmos a associação entre x e y

precisamos de uma medida de associação, que deve avaliar se essa associação é forte ou fraca, positiva ou negativa.

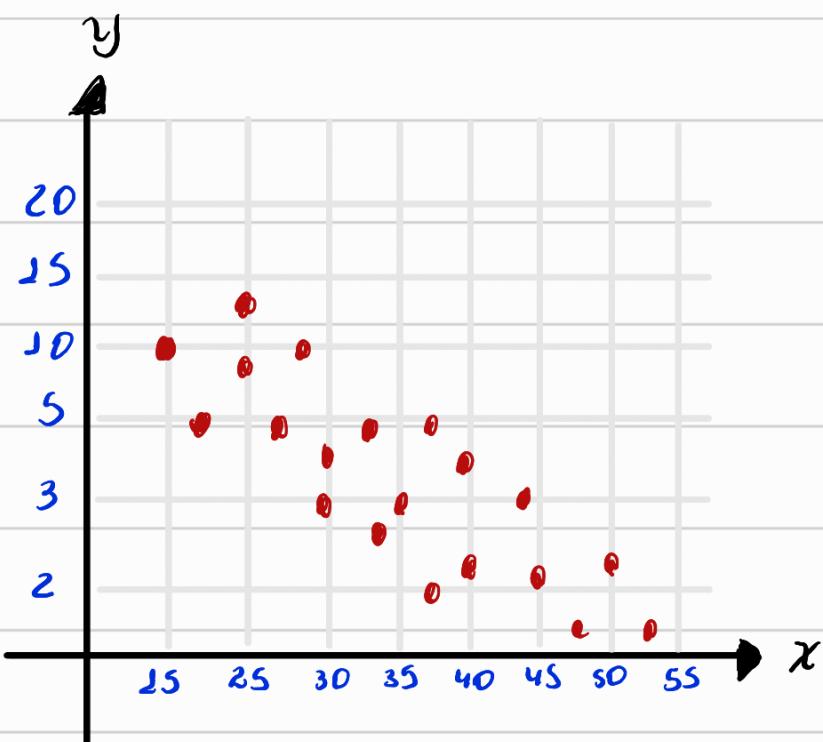
Outra possibilidade é através de representação gráfica, sendo que esse associação pode ser:

Sentido da Associação: Positivo ou Negativo
Intensidade da Associação: Forte, moderada ou fraca

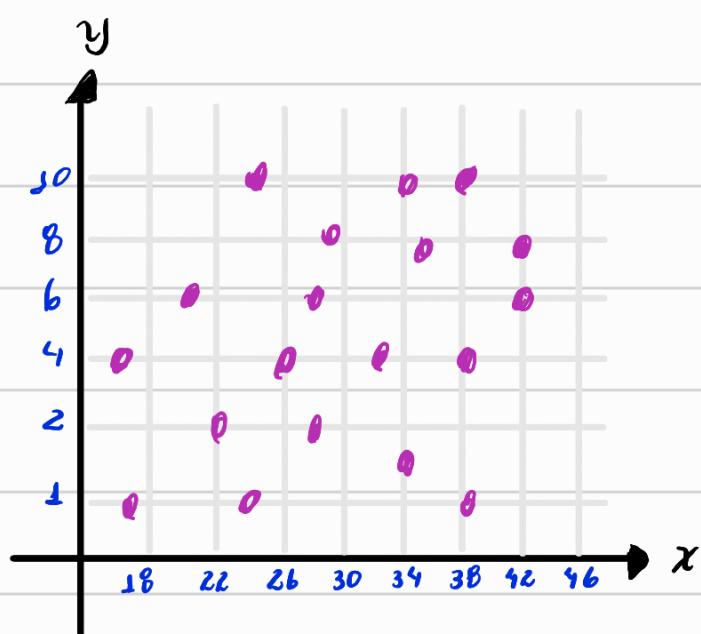
Entender o sentido e sua intensidade é necessário para poder realizar previsões nos dados.



x = Altura
 y = largura de folha



$x = \text{Jelade}$
 $y = \text{nº de acidentes}$



$x = \text{nº de sopa}$
 $y = \text{Nota final}$

Tipos de relações entre variáveis

a) Causal Unilateral: y depende de x ou x depende de y .

Exemplo: Preço do renda de um produto (y) depende do local da renda (x).

$$x \rightarrow y$$

Neste exemplo, o preço do produto depende do local da renda e também pode depender de outros fatores mas esse local, não depende do preço do produto, comprovando isso é ação unilateral (Preço x local).

b) Causal bilateral: y depende de x e x depende de y

Exemplo: Peso (x) círcunferência abdominal (y) de uma pessoa

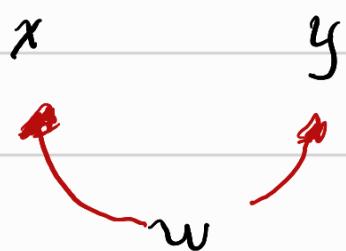
$$x \leftrightarrow y$$

Neste caso, se o peso cresce, a círcunferência também cresce, se a círcunferência aumenta,

to, o peso também aumenta, então, uma depende de outra, nos levando a uma relação causal bilateral.

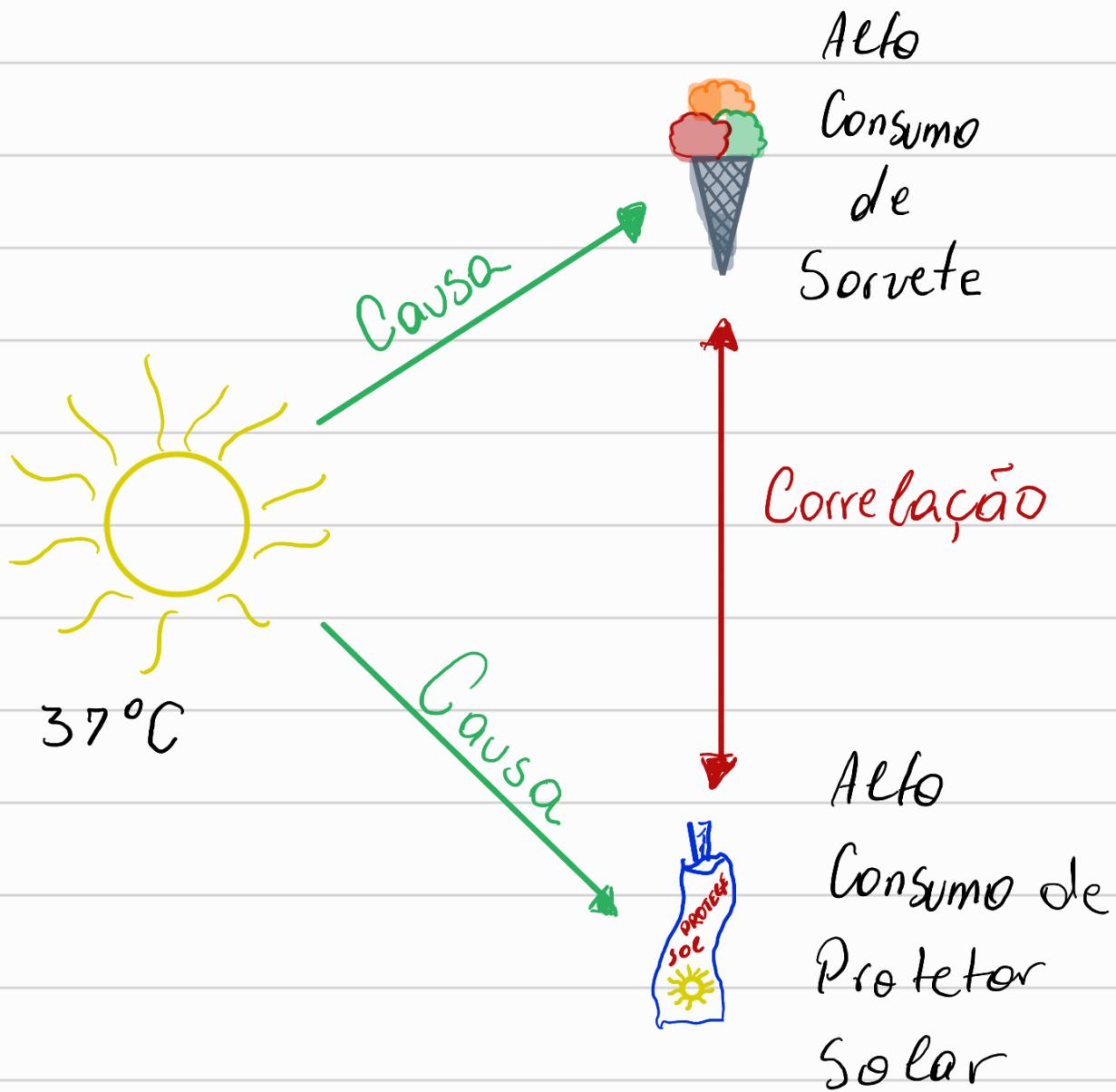
c) Dependência Indireta: Considerando uma condição que gera discussão quando usamos estatística devido ao fato de existir uma correlação mas não existe causa.

Exemplo: Vendas de sorvete na praia (x), casos de afogamento (y) e temperatura (w)



w possui uma relação causal unilateral com x , e w também tem uma relação causal unilateral com y ou seja, aumentar w causa aumento tanto em x quanto em y .

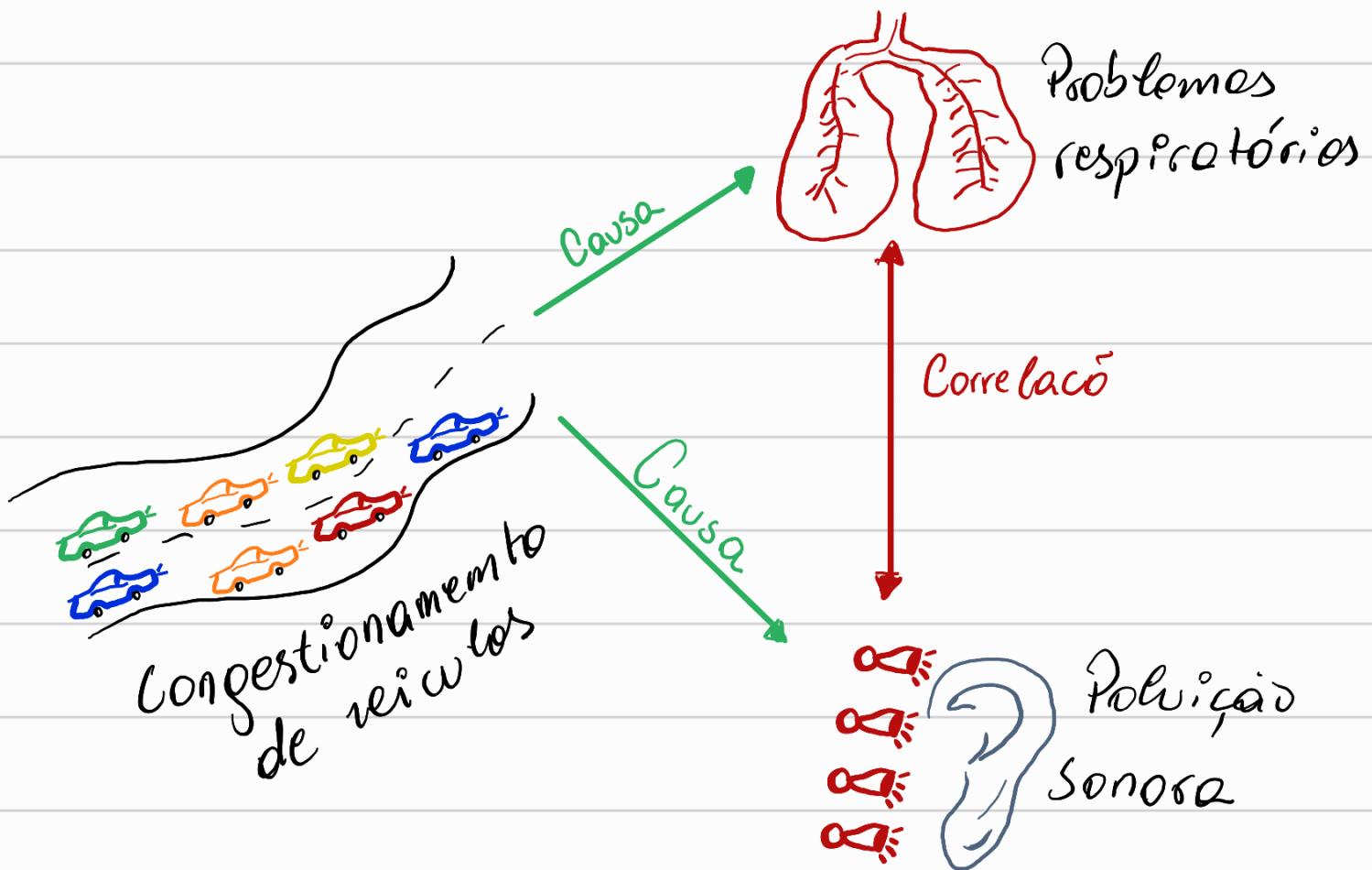
Correlação não é Causalidade



Neste exemplo, podemos observar que as altas temperaturas (variável independente numérica contínua) podem causar o aumento dos vendas (variável dependente, numérica e contínua) de sorvete e protetor solar.

Então, podemos observar que existe uma

correlações entre o consumo de sorvetes e o consumo de protetor solar mas, um não é a causa do outro (Associação não causal)



Temos também uma correlação positiva entre os problemas respiratórios e a poluição sonora, ambos causados pelo alto fluxo de veículos em uma rodovia porém, esse relogio não é de causalidade e sim, uma correlação.

Obs: ver spurious Correlations (curiosos)

Associação entre Variáveis

Já falamos sobre isso no aula passada mas, para seguiremos com a aula de hoje, vamos iniciar descrevendo essas associações:

i) Qualitativa x Qualitativa: tabelos de contingência, teste qui-quadrado, gráficos de mosaico.

ii) Quantitativa x Quantitativa: gráfico de dispersão, covariância e correlação

iii) Quantitativa x Qualitativa: boxplot, teste para comparação de médias

Associação entre variáveis Quantitativas

Segom $x \in x_1, \dots, x_k$ e $y \in y_1, \dots, y_m$,
 $1 \leq k \leq n$ e $1 \leq m \leq n$

Seja f_{ij} a frequência absoluta do par (x_i, y_j) ,
 $i = 1, \dots, k$, $j = 1, \dots, m$.

As tabelas de distribuição condicional também
são tabelas de frequências relativas, mas qual
usar?

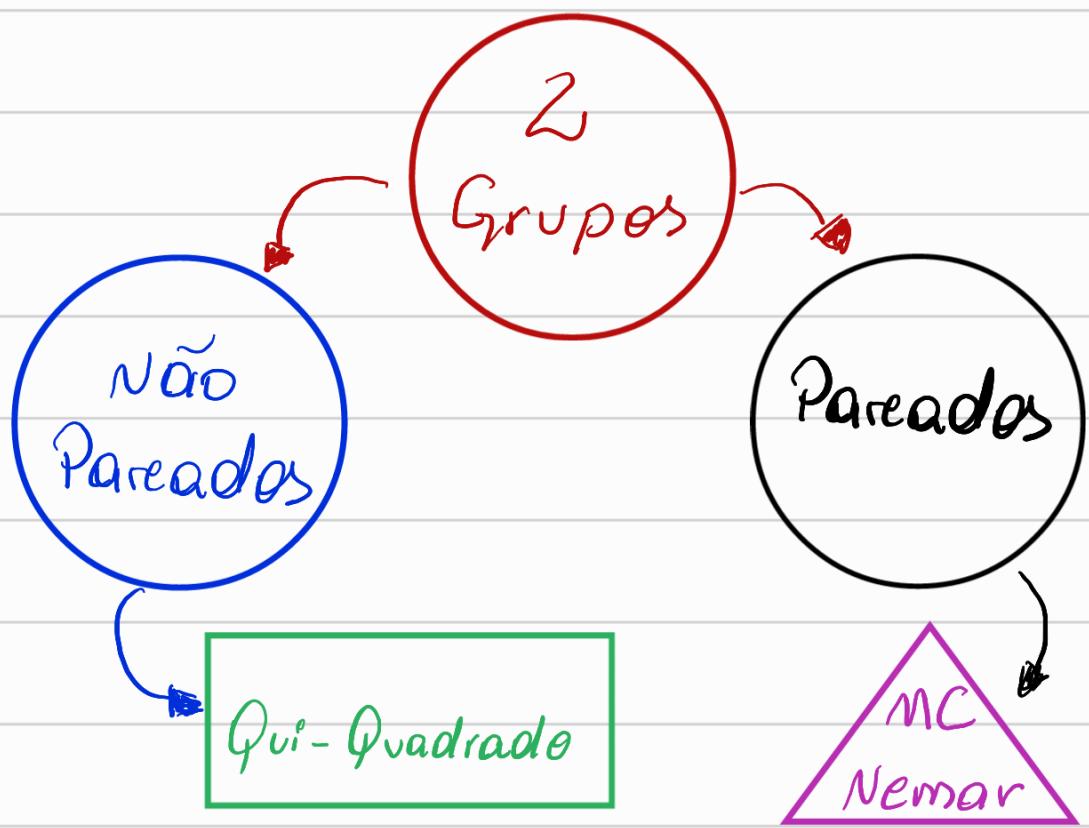
i) Relação causal bilateral ($x \leftrightarrow y$):

Tabela de frequência relativa (dividido por
todos os elementos por n)

ii) Relação causal unilateral ($x \rightarrow y$):

Distribuição condicional

Testes para variáveis Qualitativas
Nominais para 2 grupos



Não Pareados:

25 Grupo de homens > var. independente
 25 n mulheres > (agrupamento)

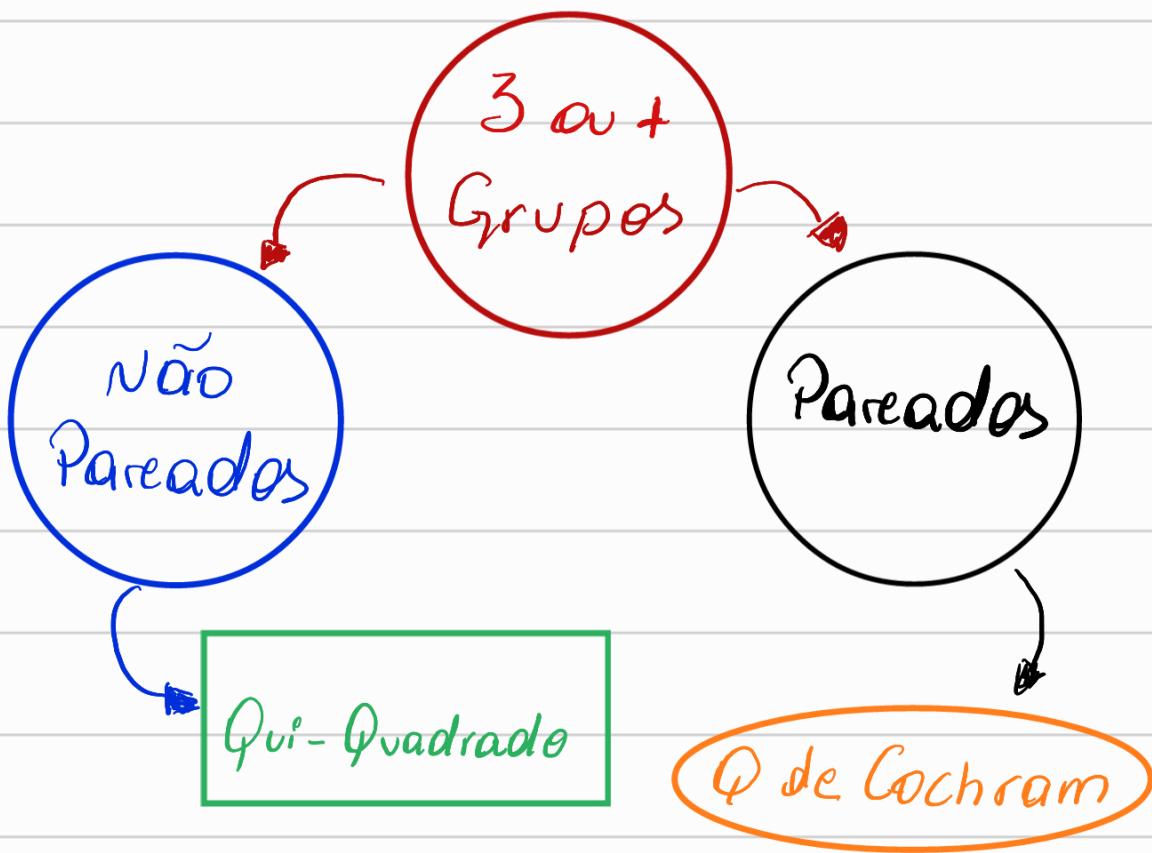
Esses grupos então são compostos de outros de um nível dependente que é o classificar o indivíduo com estandardize a normal, outros são teste qui-quadrado apontando os percentuais para cada grupo.

Pareados:

Grupo de meninos > var. independente
n " meninos

Para a métrica, podemos então avaliar esse grupo contendo meninos e meninas no 1º ano da faculdade e depois, avaliar esses mesmos indivíduos no último ano anterior do teste de NC Nemar.

Testes para variáveis Qualitativas
Nominais para 3 Grupos



Não pareados:

Grupo de meninos

n " meninos

n " funcionários

grupos de pessoas diferentes

var. independente

Pareados:

Grupo de meninos

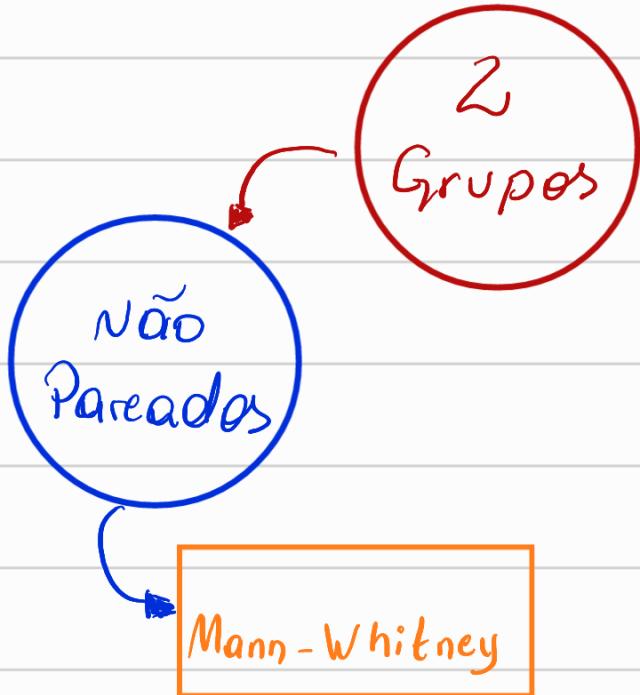
n " meninos

n " funcionários

mesma unidade amostral

Esses grupos foram avaliados no início, 3 meses depois, 6 meses depois, 1 ano depois

Qualitativos Ordinais ou Quantitativos
sem distribuição Normal



Não pareados:

Grupos de meninos → var. independente
n n meninos



1 - 2 - 3 - 4

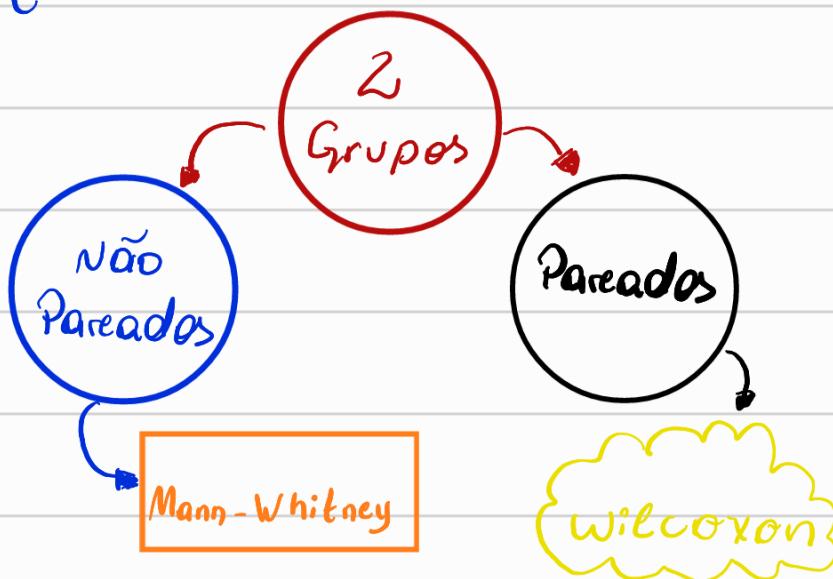
categorias
concretas



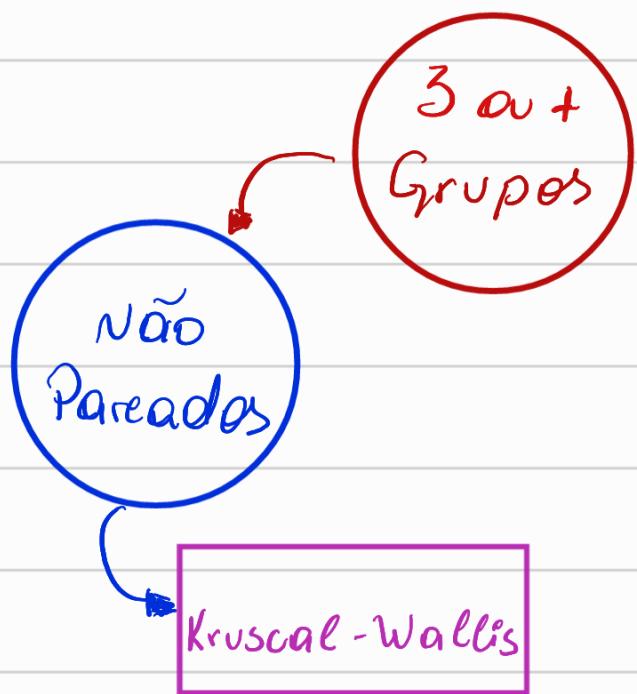
1 - 2 - 3 - 4

Pareados:

Mesmas unidades medidas mais de uma vez



Testes para variáveis Qualitativas Ordinais para 3 Grupos



Regras: Te dou uma resposta parcial ou limitada (grupos diferentes ou não) porém quero saber em quais grupos acontece essa diferença.

Não pareados:

Grupo de meninos

n " meninos

n " funcionários

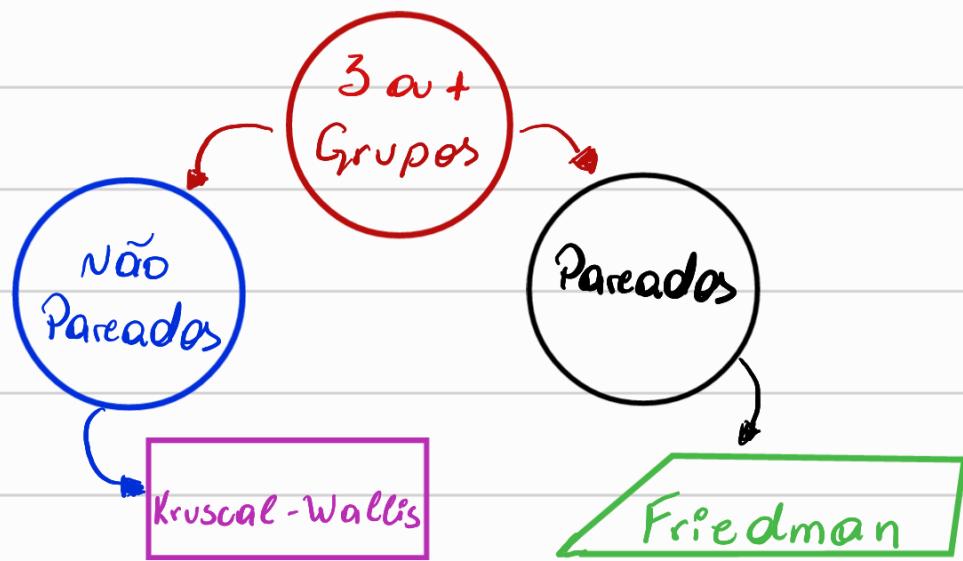
grupos de pessoas diferentes

var. independente

Havendo diferença estatisticamente significativa, torna-se necessário a utilização de testes de comparação múltiple ou pós ROC

Pareados

Mesmas unidades medidas mais de uma vez

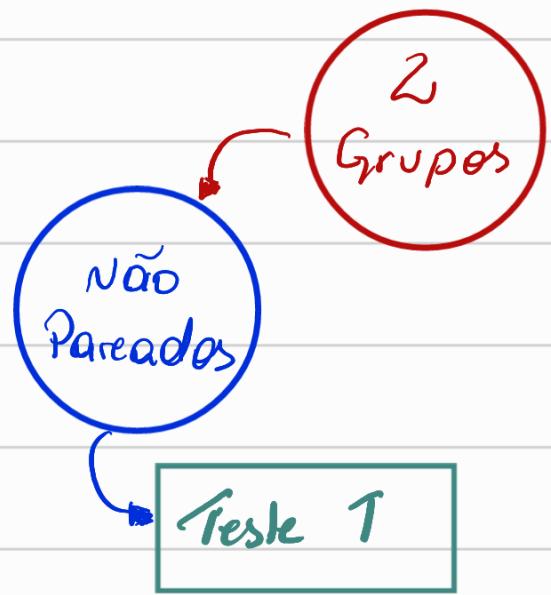


Também se encontrou diferença entre os grupos, desse que rodar um teste de comparação para mensurar essa diferença.

Variáveis quantitativas

Os testes comparativos realizados quando a variável dependente é quantitativa, possuem resultados mais "precisos" se comparado aos testes anteriores.

Quantitativa com distribuição Normal



Não pareados: Teste T de Student

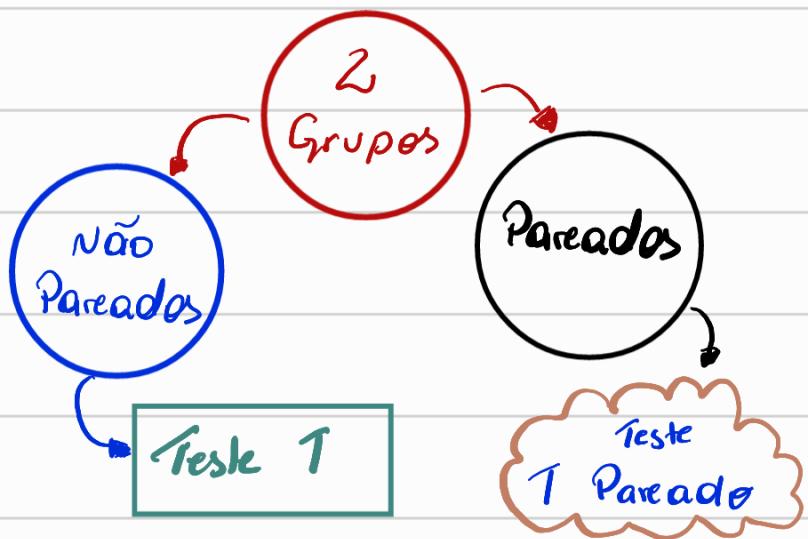


→ 0 até 100

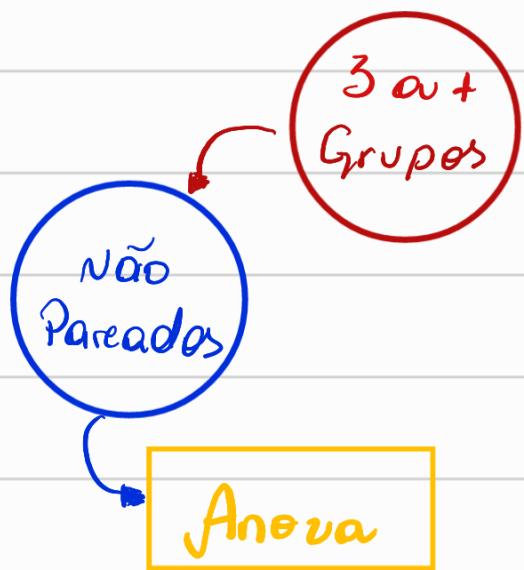


→ 0 até 100

Pareados Mesmo grupo avaliado no inicio e no final da pesquisa



Testes para variações Quantitativas para
3 grupos



Anova: Análise de
variação.
Também quanto
esperar a diferença
precisando comparar
esta diferença.

(Teste mais executado)

Pareados

