



# **Universidade Federal do Ceará**

## **Faculdade de Economia**

### **Métodos Quantitativos**

Vicente Lima Crisóstomo

Fortaleza, 2020

# Sumário

- Introdução
- Estatística Descritiva
- Probabilidade
- Distribuições de Probabilidades
- Amostragem e Distribuições Amostrais
- Estimação
- Testes de Significância
- Análise de Variância
- Teste de Significância para Proporções
- Testes Não Paramétricos
- Correlação e Regressão

# Correlação e Regressão

## ■ Correlação e Regressão

### ■ Técnicas relacionadas

### ■ Fazem uma estimação

### ■ Correlação e Regressão

- Estimam relações que possam existir entre variáveis da população
- Analisam dados amostrais buscando
  - Existência e a forma de relação de uma variável com outra

### ■ Técnicas anteriores

- Fazem estimação relativas a um parâmetro populacional

# Correlação e Regressão

- Análise de Correlação
  - Avalia se há relacionamento entre variáveis
  - Encontra/calcula um número que exprime o grau de relacionamento entre duas variáveis
  - Fundamental em trabalho exploratório
  - Permite uma visão geral de relacionamentos entre variáveis
  - Avalia importância de variáveis no contexto
  - Identifica potenciais relações e respectivas “forças”

# Correlação e Regressão

- **Análise de Regressão**
  - Ênfase na natureza do relacionamento
  - Busca uma Equação matemática
    - Capaz de descrever o relacionamento entre variáveis
    - Equação pode ser usada para estimar valores de uma variável com base em valores de outra(s)
  - De relevante importância em
    - Economia, administração, contabilidade

# Correlação e Regressão

## ■ Correlação e Regressão

- Dados são emparelhados
- Cada observação tem duas ou mais variáveis

## ■ Exemplos

- Amostra de pessoas
  - Nome, consumo, renda, escolaridade
  - Nome, idade, altura, peso
- Amostra de alunos
  - Nome, indicador\_desempenho, horas\_estudo\_semanal
- Amostra de empresas
  - Empresa, RSC, endividamento, rentabilidade, tangibilidade

# Análise de Regressão

- Correlação e Regressão
  - Dados são emparelhados
  - Cada observação tem duas ou mais variáveis
  - Exemplos
    - Amostra de pessoas
      - Nome, consumo, renda, escolaridade
      - Nome, idade, altura, peso
    - Amostra de alunos
      - Nome, indicador\_desempenho, horas\_estudo\_semanal,
    - Amostra de empresas
      - Empresa, RSC, endividamento, rentabilidade, tangibilidade

# Análise de Correlação

- Análise de Correlação
  - Estudo investigativo correlacional
  - Correlação    Correlacionamento    Co-relacionamento
  - Grau de associação entre valores de duas variáveis
  - Exemplos
    - Idade e renda?
    - Escolaridade e renda?
    - rentabilidade e qualidade da gestão da empresa?
    - Desempenho acadêmico e horas de estudo?
    - Consumo e renda?
    - Temperatura e dedicação ao trabalho?
    - Estrutura de propriedade e endividamento da empresa?



# Análise de Correlação

## ■ Análise de Correlação

### ■ Exemplos

- Inflação na Alemanha e criminalidade na Colômbia?
- Renda per capita em países desenvolvidos e nível de pobreza em países pobres?
- Satisfação do trabalhador e produtividade?
- Nível salarial e produtividade?
- Inflação e consumo?
- Temperatura e venda de casacos de frio?
- Preço e nível de venda de um bem?
- Empreendedorismo e crescimento econômico do país?
- Investimento empresarial e nível de emprego?

# Análise de Correlação

## ■ Análise de Correlação

- Avalia possíveis associações (ou correlacionamentos) entre variáveis
- O resultado da análise é um indicador
  - Um **Coeficiente de Correlação**
  - Valor que exprime o grau de correlação entre as duas variáveis analisadas
- Correlação entre variáveis com dados contínuos
  - Forma mais usada de análise de correlação
  - O Coeficiente ***r*** de Pearson
    - Expressa o grau de relacionamento entre duas variáveis com dados contínuos

# Análise de Correlação

## ■ Análise de Correlação

### ■ O **Coeficiente de Correlação $r$** de Pearson

- Matemático Karl Pearson
- Requisitos/supostos para validade do  **$r$**  de Pearson
  - As duas variáveis aleatórias e contínuas
  - Distribuição de frequência conjunta
    - das duas variáveis, i.e., pares  $(x, y)$  é normal
    - Distribuição normal bivariada

# Análise de Correlação

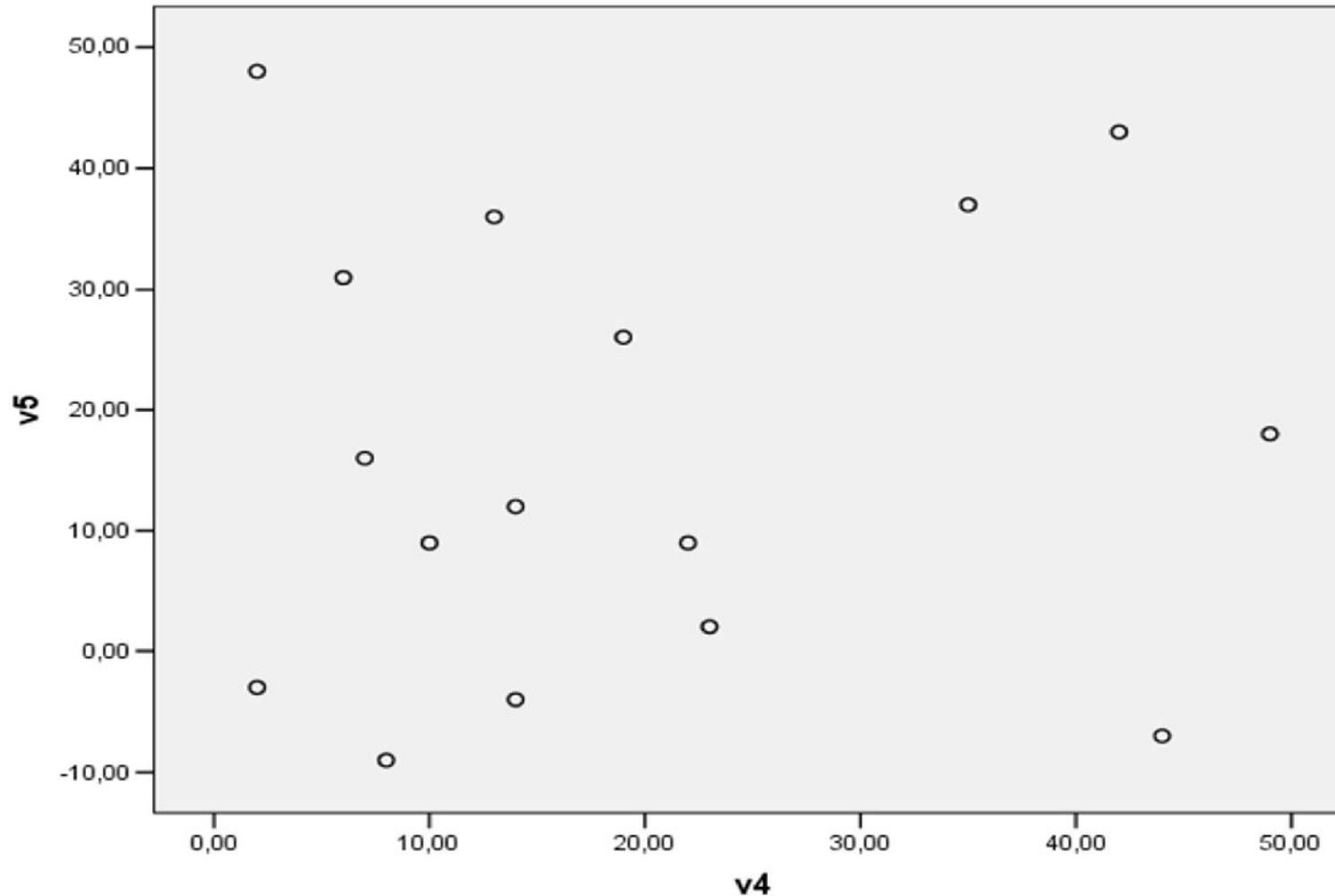
## ■ Análise de Correlação

### ■ ***Coeficiente de Correlação $r$*** de Pearson

- Propriedades do  ***$r$***  de Pearson
  - Magnitude
    - [0; 1]
    - Indica o grau de correlacionamento entre as variáveis
    - Quão próximo de uma reta estão os pontos (x, y)
  - Sinal (+ ou -)
    - Equivale ao coeficiente angular de uma reta imaginária

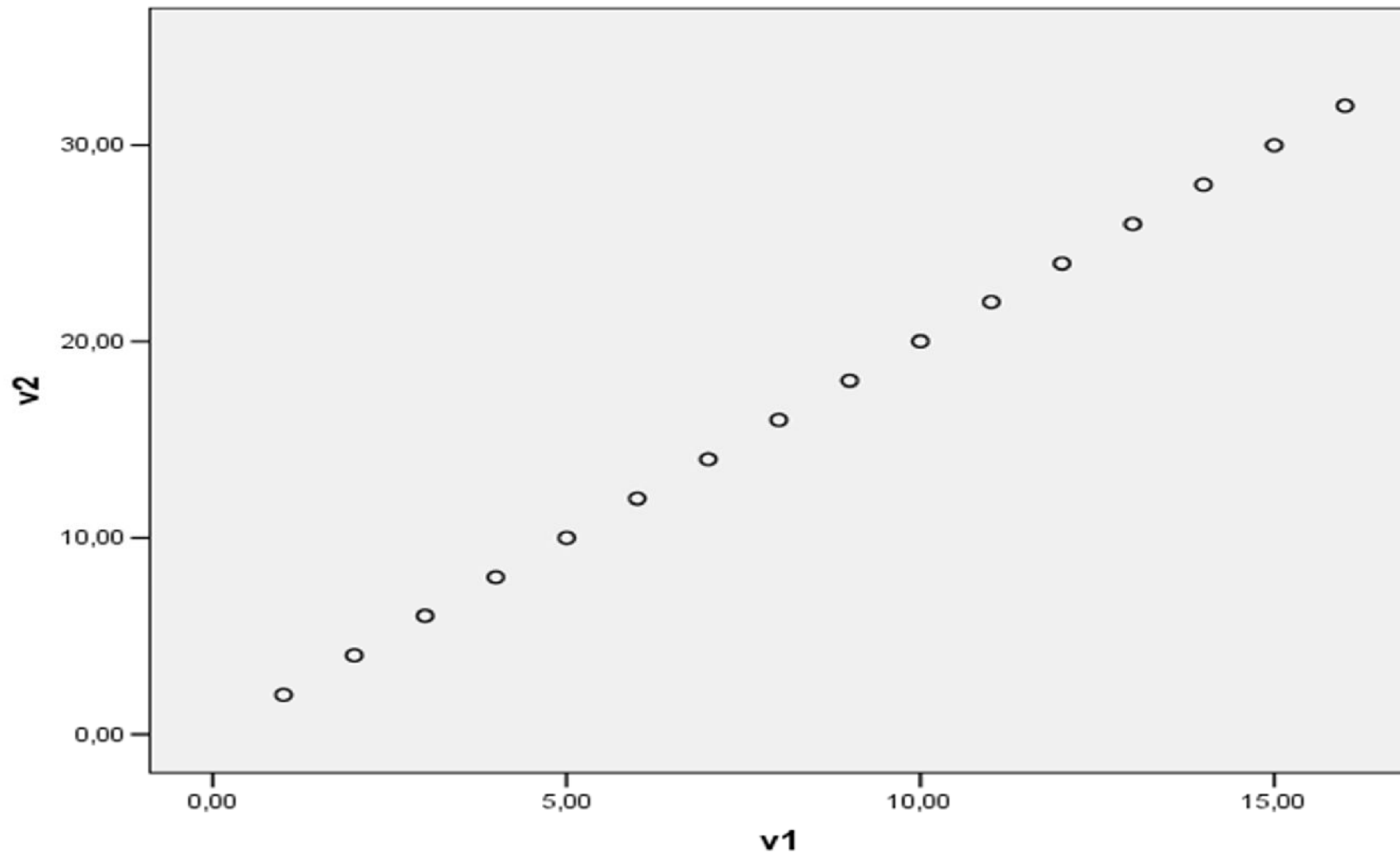
## ■ Fraca Correlação: $r = 0,067$

- Quase ausência de relacionamento entre variáveis



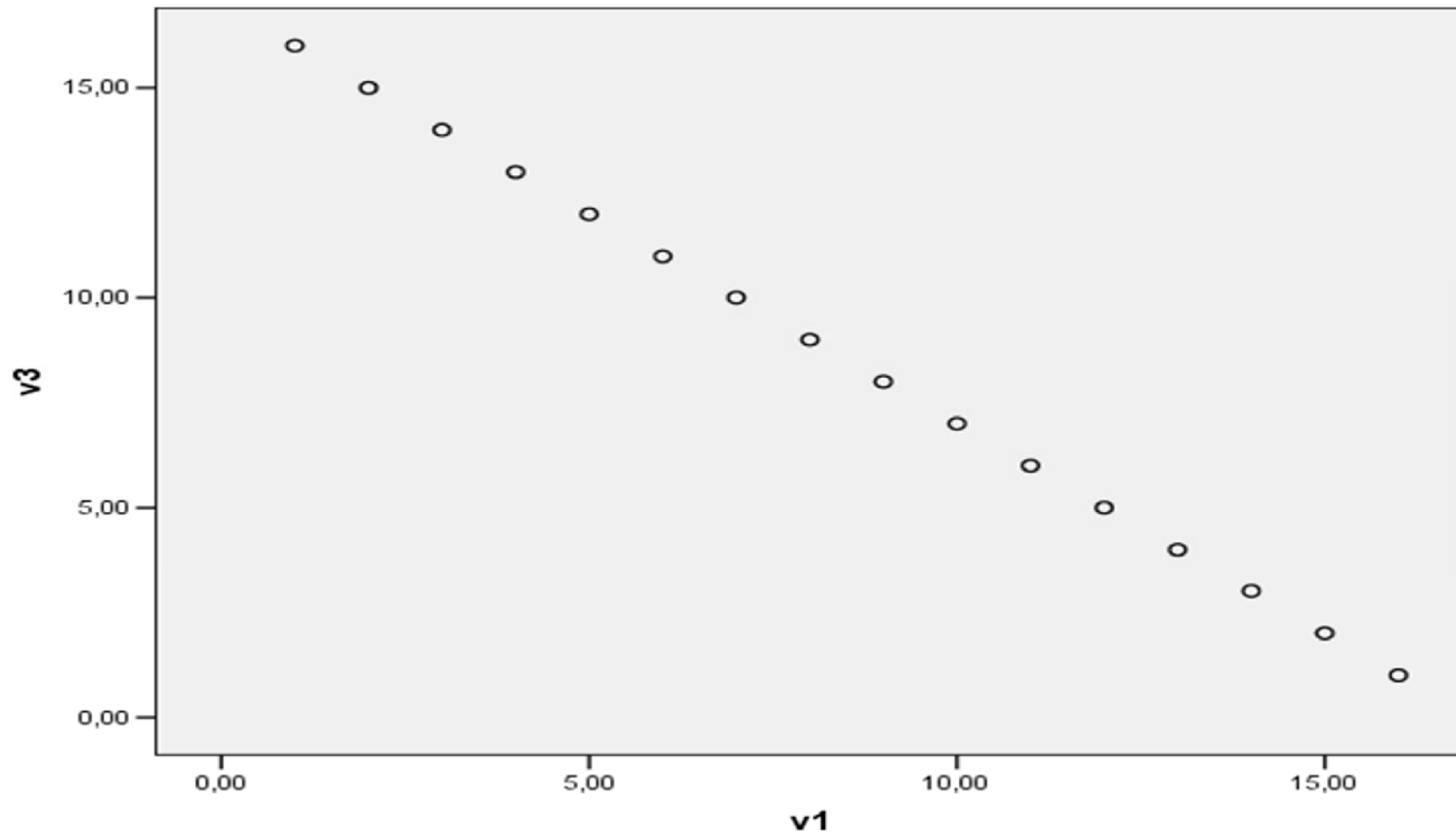
## ■ Forte Correlação Positiva: $r = +1,0$

- Relacionamento positivo, perfeito



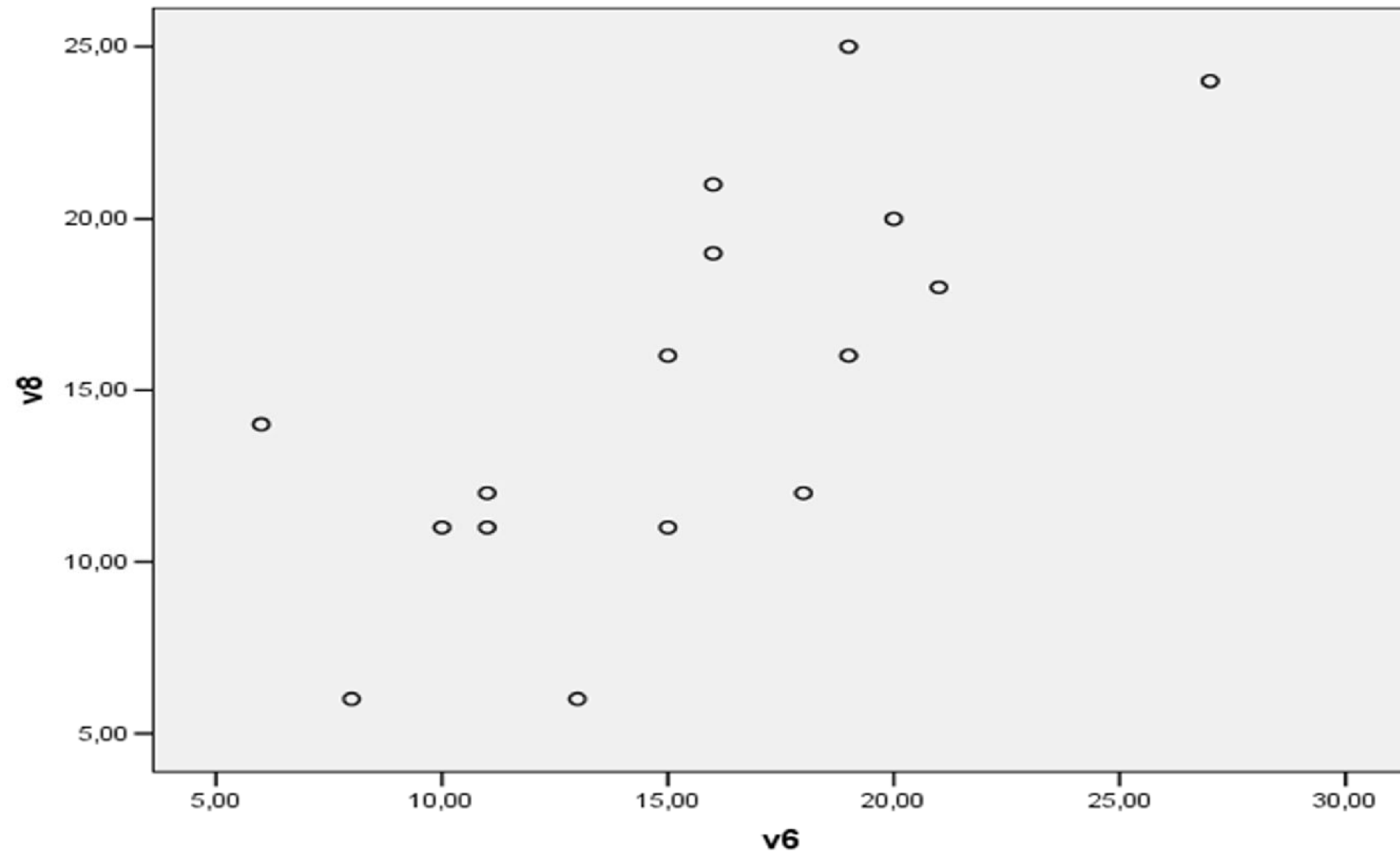
## ■ Forte Correlação Negativa: $r = -1,0$

- Relacionamento negativo, perfeito



## ■ Correlação Positiva: $r = +0,714$

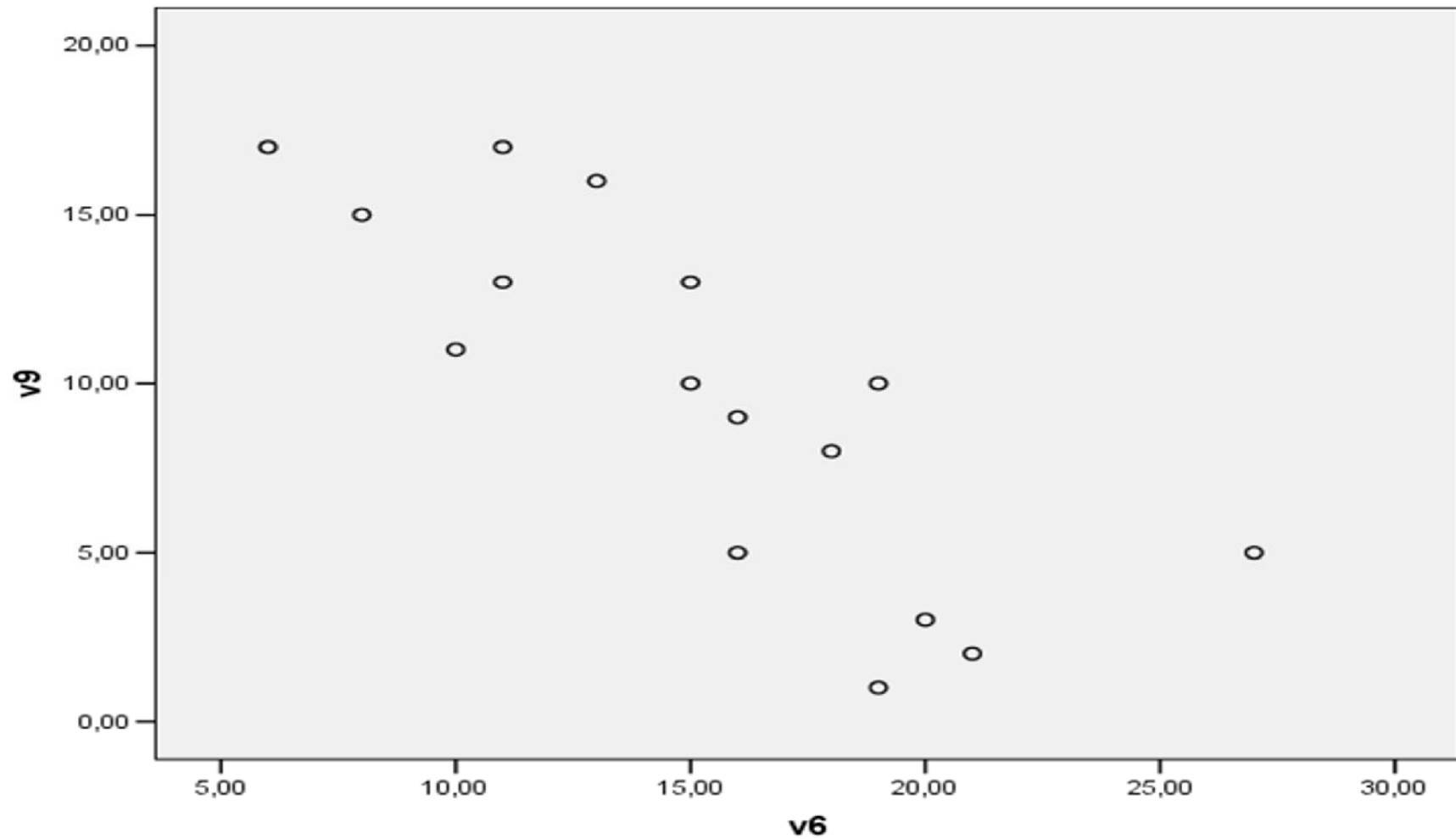
- Relacionamento positivo, moderado





## ■ Correlação Negativa: $r = -0,798$

- Relacionamento negativo, moderado



# Análise de Correlação

## ■ **Coeficiente de Correlação $r$ de Pearson**

- Verifica se posição, ou situação, relativa
  - das observações de um grupo estão relacionadas com
  - as posições do outro grupo
- Cálculo/medição da posição relativa em um grupo
  - Função da média e desvio padrão (DP)
  - **$z = (\text{valor} - \text{média}) / DP$** 
    - Padroniza cada um dos valores de cada variável
    - Assim, tornam-se comparáveis os grupos/variáveis
    - Valores padronizados (z)
      - Usados para calcular valor que meça uma situação combinada
      - Usados para calcular valor que meça posição relativa em ambos os grupos

# Análise de Correlação

## ■ **Coeficiente de Correlação $r$ de Pearson**

### ■ **$z = (\text{valor} - \text{média}) / DP$**

- Valores padronizados ( $z$ ) de cada variável
  - Usados para calcular valor que meça uma situação combinada
  - Valores efetivos acima da média geram mais altos  $z$  que são positivos
  - Valores efetivos abaixo da média geram mais baixos  $z$  que são negativos

## ■ Produto dos dois escores padronizados

- $z_x \times z_y$

$$+ \times + = +$$

$$- \times - = +$$

$$- \times + = -$$

$$+ \times - = -$$

# Análise de Correlação

## ■ **Coeficiente de Correlação $r$ de Pearson**

### ■ Produto dos dois escores padronizados

- $z_x \times z_y$
- Se variáveis estão correlacionadas positivamente
  - Escores de  $x$  ( $z_x$ ) estão emparelhados com escores de  $y$  ( $z_y$ ) progressivamente
  - Mais baixo  $z_x$  está emparelhado com mais baixo  $z_y$
  - Mais alto  $z_x$  está emparelhado com mais alto  $z_y$
  - Tendência
    - $z_x \times z_y > 0$

# Análise de Correlação

## ■ **Coeficiente de Correlação $r$ de Pearson**

### ■ Produto dos dois escores padronizados

- $z_x \times z_y$
- Se variáveis estão correlacionadas negativamente
  - Escores de  $x$  ( $z_x$ ) estão emparelhados com escores de  $y$  ( $z_y$ ) inversamente
  - Mais baixo  $z_x$  está emparelhado com mais alto  $z_y$
  - Mais alto  $z_x$  está emparelhado com mais baixo  $z_y$
  - Tendência
    - $z_x \times z_y < 0$

# Análise de Correlação

## ■ **Coeficiente de Correlação $r$ de Pearson**

- O Coeficiente Correlação  $r$  de Pearson
- É a média dos produtos dos dos escores padronizados das duas variáveis ( $z_x \times z_y$ )
  - Calcula-se a soma dos produtos ( $z_x \times z_y$ )
  - Divide-se pelo número de produtos
  - Observe-se
    - Valores emparelhados geram produto ( $z_x \times z_y$ ) positivo
    - Isto ocasiona
      - maior valor da soma dos produtos e assim
      - maior valor da média, que é o coeficiente de correlação

# Análise de Correlação

- Resumo do cálculo do Coeficiente Correlação  $r$  de Pearson
- Correlação entre variáveis  $x$  e  $y$ 
  - Padronizar todos os valores (observações)

$$z_x = \frac{(x_i - \bar{x})}{s_x}$$

$$z_y = \frac{(y_i - \bar{y})}{s_y}$$

- Calcular somatório dos produtos dos valores padronizados emparelhados e o Coeficiente Correlação  $r$  de Pearson

$$r = \frac{\sum_{i=1}^n z_x z_y}{n - 1}$$

obsv	X	Y	xi - med_x	zx=(xi-med_x)/dp	yi - med_y	zy=(yi-med_y)/dp	zx X zy	x . y	x2	y2
1	80	1	-9	-1,8	-1,5	-1,5	2,7	80	6400	1
2	82	1	-7	-1,4	-1,5	-1,5	2,1	82	6724	1
3	84	2,1	-5	-1	-0,4	-0,4	0,4	176,4	7056	4,41
4	85	1,4	-4	-0,8	-1,1	-1,1	0,88	119	7225	1,96
5	87	2,1	-2	-0,4	-0,4	-0,4	0,16	182,7	7569	4,41
6	88	1,7	-1	-0,2	-0,8	-0,8	0,16	149,6	7744	2,89
7	88	2	-1	-0,2	-0,5	-0,5	0,1	176	7744	4
8	89	3,5	0	0	1	1	0	311,5	7921	12,25
9	90	3,1	1	0,2	0,6	0,6	0,12	279	8100	9,61
10	91	2,4	2	0,4	-0,1	-0,1	-0,04	218,4	8281	5,76
11	91	2,7	2	0,4	0,2	0,2	0,08	245,7	8281	7,29
12	92	3	3	0,6	0,5	0,5	0,3	276	8464	9
13	94	3,9	5	1	1,4	1,4	1,4	366,6	8836	15,21
14	96	3,6	7	1,4	1,1	1,1	1,54	345,6	9216	12,96
15	98	4	9	1,8	1,5	1,5	2,7	392	9604	16
media	89	2,5						12,6		
dp	5	1						<b>r =</b>	<b>0,9</b>	



# Análise de Correlação

- Resumo do cálculo do Coeficiente Correlação  $r$  de Pearson
- Fórmula alternativa que dispensa padronização

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \cdot \sqrt{n(\sum y^2) - (\sum y)^2}}$$

# Análise de Correlação

- Teste de Significância do Coeficiente de Correlação  $r$  de Pearson
  - $r$  : coeficiente de correlação amostral
  - $\rho$  : coeficiente de correlação populacional
- Hipóteses
  - $H_0: \rho = 0$ 
    - Não há correlação entre as variáveis
  - $H_1: \rho \neq 0$ 
    - Há correlação entre as variáveis

# Análise de Correlação

## ■ Teste de Significância do Coeficiente Correlação $r$ de Pearson

### ■ Hipóteses

- $H_0: \rho = 0$
- $H_1: \rho \neq 0$

### ■ Estatística de teste:

$$t = \frac{r - 0}{\sqrt{\frac{(1 - r^2)}{(n - 2)}}}$$

### ■ Graus de Liberdade: $(n - 2)$

### ■ Teste bilateral de $t$

# Análise de Correlação

## ■ Exemplo

- $n = 16$ ;  $r = 0,067$ ;  $GL = 14$  (variáveis  $x$  e  $y$ );  
 $t = 0,2506$
- $n = 16$ ;  $r = +0,999999999$ ;  $GL = 14$ ;  $t = 26.457,51$
- $n = 16$ ;  $r = -0,999999999$ ;  $GL = 14$ ;  $t = -26.457,51$
- $n = 16$ ;  $r = 0,714$ ;  $GL = 14$ ;  $t = 3,82056292^{***}$
- $n = 16$ ;  $r = 0,798$ ;  $GL = 14$ ;  $t = -4,961748894^{***}$

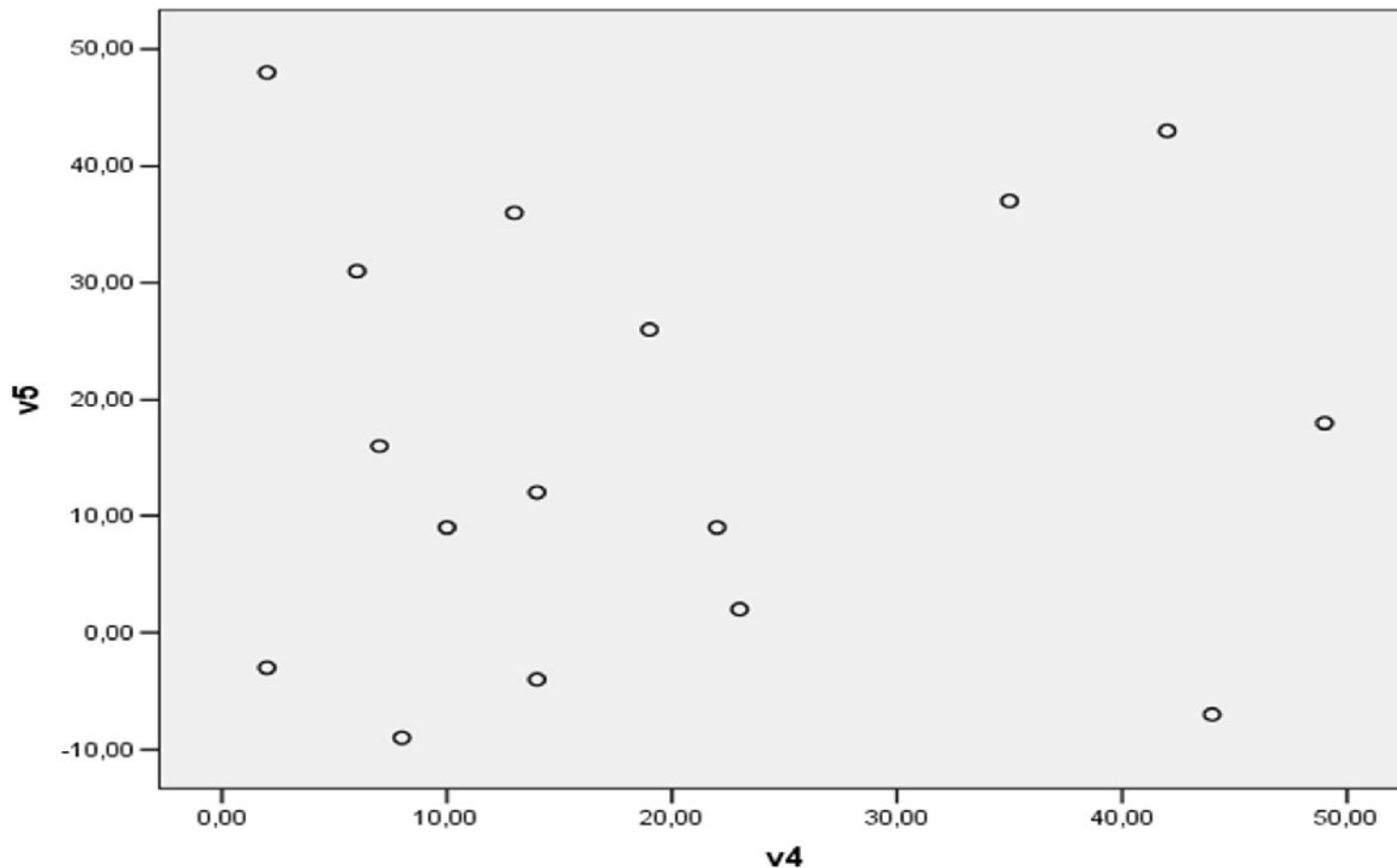
# Pontos Percentuais da Distribuicao t

## Probabilidades na cauda

Uma Cauda		0,100	0,050	0,025	0,010	0,005	0,001	0,0005
Duas Caudas		0,200	0,100	0,050	0,020	0,010	0,002	0,001
D	1	3,078	6,314	12,710	31,820	63,660	318,300	637,000
E	2	1,886	2,920	4,303	6,965	9,925	22,330	31,600
G	3	1,638	2,353	3,182	4,541	5,841	10,210	12,920
R	4	1,533	2,132	2,776	3,747	4,604	7,173	8,610
E	5	1,476	2,015	2,571	3,365	4,032	5,893	6,869
E	6	1,440	1,943	2,447	3,143	3,707	5,208	5,959
S	7	1,415	1,895	2,365	2,998	3,499	4,785	5,408
	8	1,397	1,860	2,306	2,896	3,355	4,501	5,041
O	9	1,383	1,833	2,262	2,821	3,250	4,297	4,781
F	10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
	11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
F	12	1,356	1,782	2,179	2,681	3,055	3,930	4,318
R	13	1,350	1,771	2,160	2,650	3,012	3,852	4,221
E	14	1,345	1,761	2,145	2,624	2,977	3,787	4,140
E	15	1,341	1,753	2,131	2,602	2,947	3,733	4,073
D	16	1,337	1,746	2,120	2,583	2,921	3,686	4,015
O	17	1,333	1,740	2,110	2,567	2,898	3,646	3,965
M	18	1,330	1,734	2,101	2,552	2,878	3,610	3,922

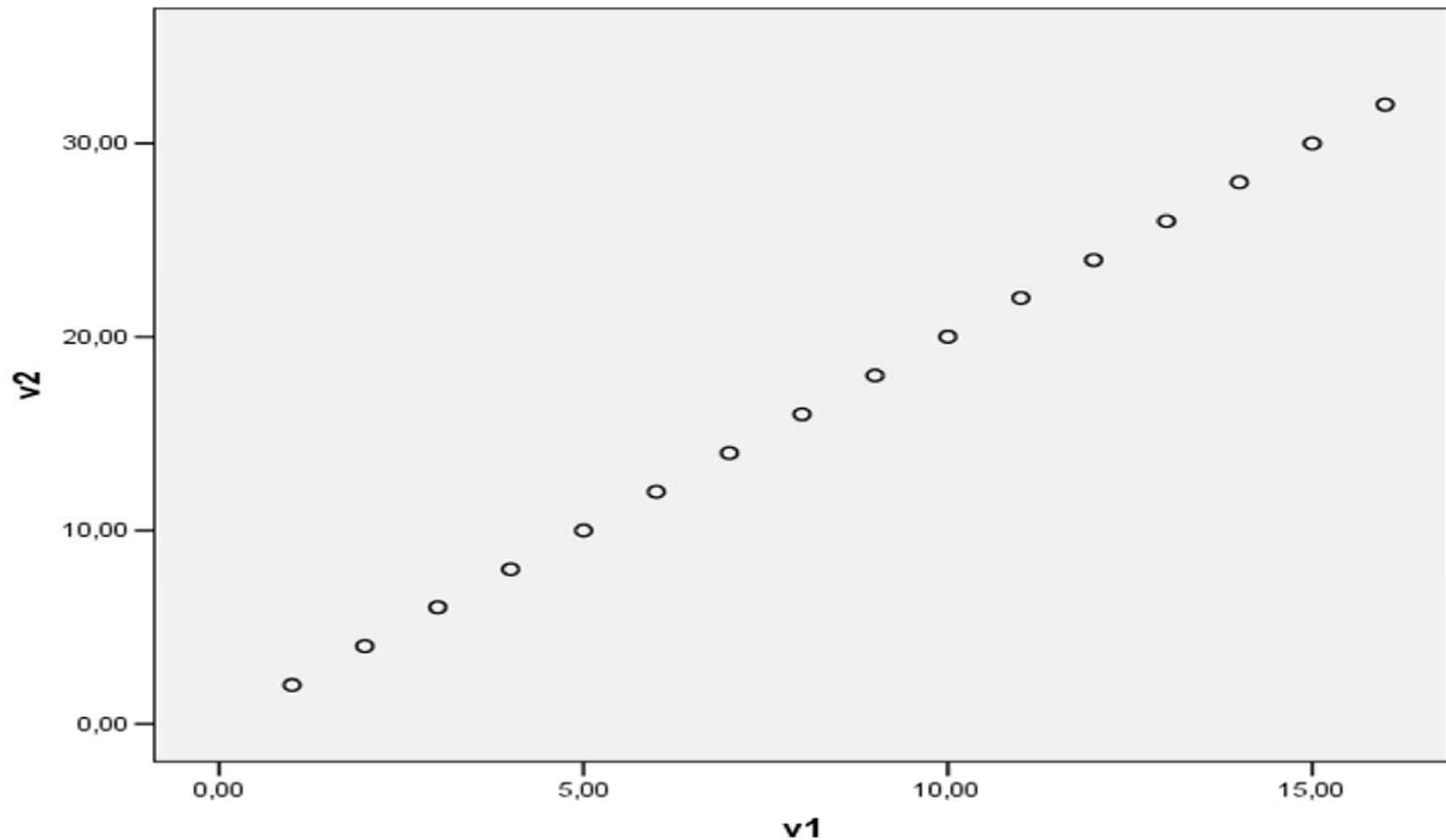
## ■ Fraca Correlação: $r = 0,067$

- $n = 16$ ;  $r = 0,067$ ;  $t = 0,2506 \Rightarrow$  aceitação de  $H_0$  ( $\alpha = 0$ )
  - $t$  menor que  $t$  crítico
- Quase ausência de relacionamento entre variáveis



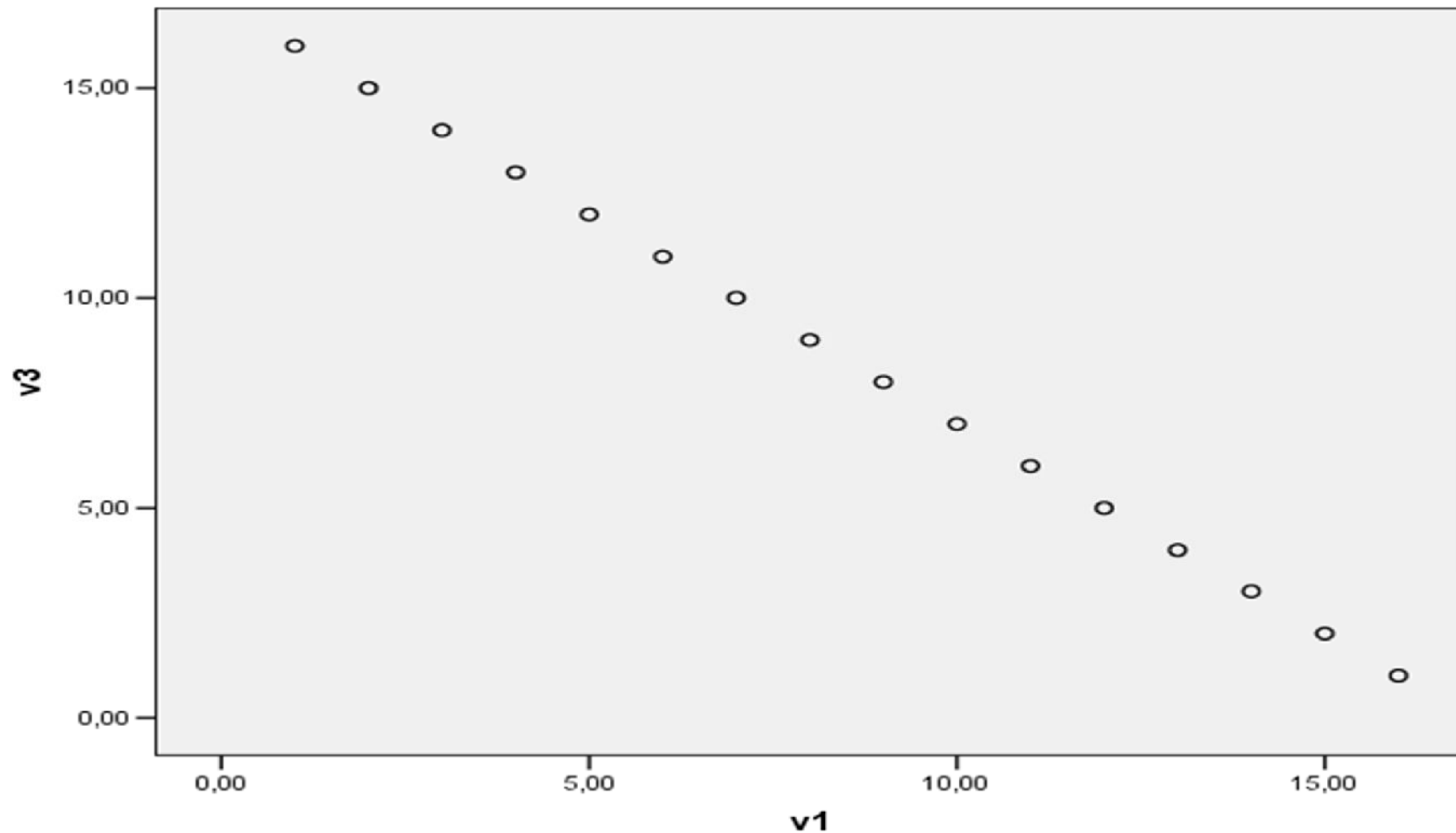
## ■ Forte Correlação Positiva: $r = +1,0$

- $n = 16$ ;  $r = +0,999999999$ ;  $t = 26.457,51 \Rightarrow$  rejeição de  $H_0$  e aceitação de  $H_1$  ( 0)
  - $t$  supera  $t$  crítico



## ■ Forte Correlação Negativa: $r = -1,0$

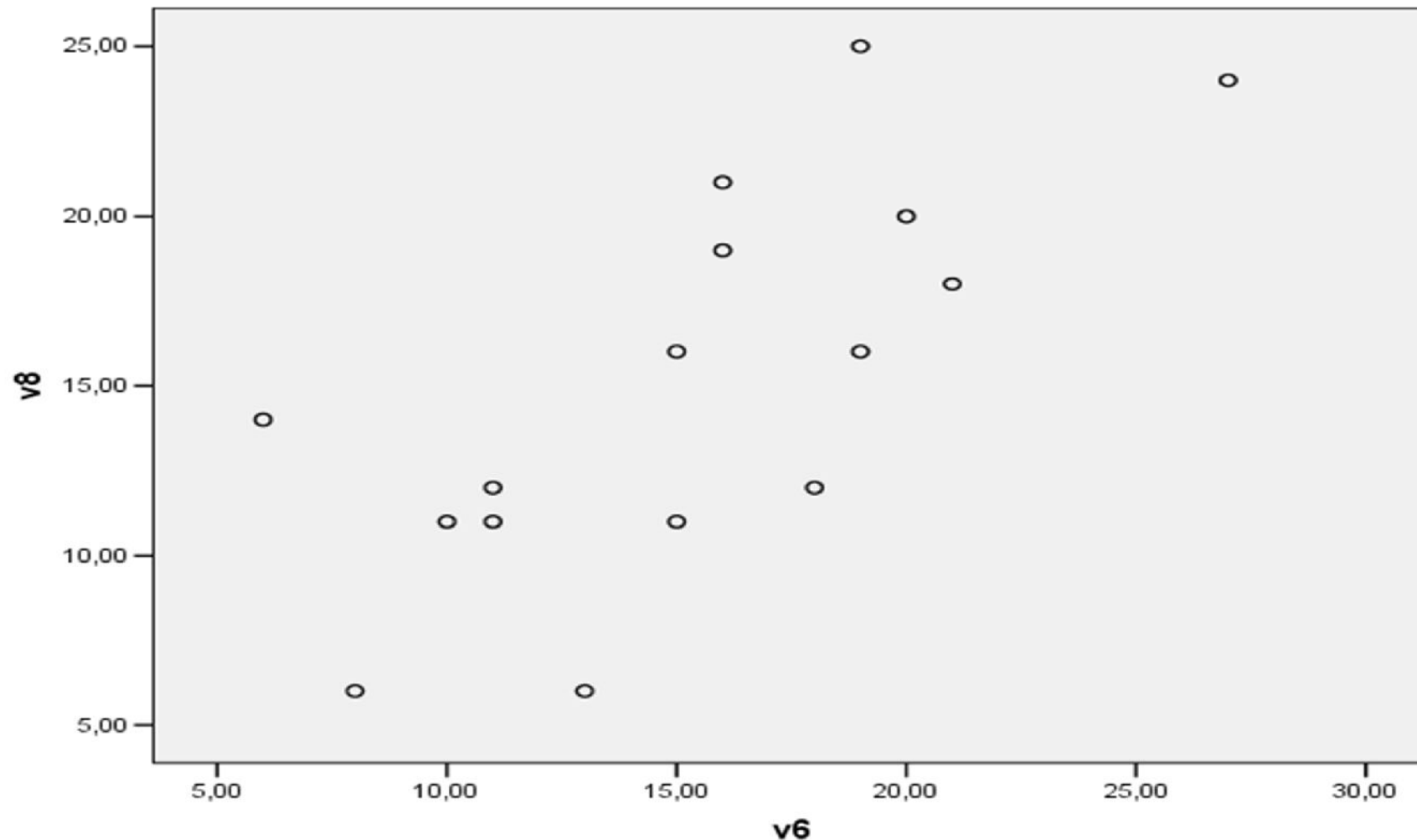
- $n = 16$ ;  $r = -0,999999999$  ;  $t = -26.457,51 \Rightarrow$  rejeição de  $H_0$  e aceitação de  $H_1$  ( 0)
- Relacionamento negativo, perfeito





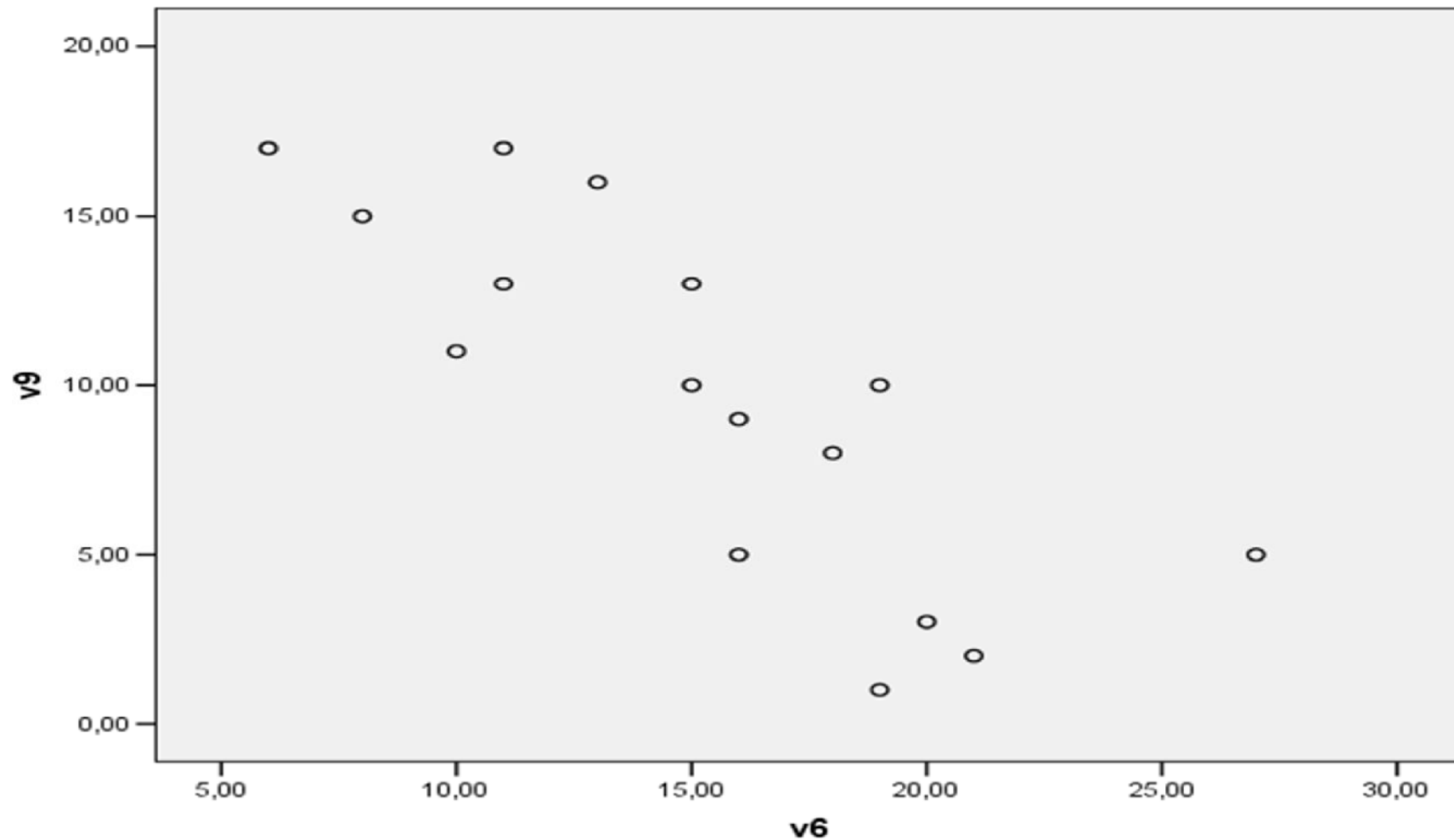
## ■ Correlação Positiva: $r = +0,714$

- $n = 16$ ;  $r = 0,714$ ;  $t = 3,82056292 \Rightarrow$  rejeição de  $H_0$  e aceitação de  $H_1$  (     0)
- Relacionamento positivo, moderado



## ■ Correlação Negativa: $r = -0,798$

- $n = 16$ ;  $r = 0,798$ ;  $t = -4,961748894 \Rightarrow$  rejeição de  $H_0$  e aceitação de  $H_1$  ( 0)
- Relacionamento negativo, moderado



# Análise de Correlação

- Dados por Postos:  $r$  de Spearman
- Dados Nominais: Coeficiente de Contingência

# Análise de Correlação

- Quando os dados forem medidos somente no nível ordinal eles são chamados de não-paramétricos e a correlação de Pearson não é apropriada
- Dados por Postos:  $r$  de Spearman

# Análise de Correlação

- Dados por Postos:  $r$  de Spearman
- Coeficiente de correlação de Spearman
  - estatística não-paramétrica
    - pode ser usado quando os dados violarem suposições paramétricas, tais como dados não-normais
- Teste de Spearman
  - Classifica os dados em primeiro lugar e então aplicando a equação de Pearson aos dados ordenados
  - Categorias que podem ser ordenadas de maneira significativa, os dados são ordinais

# Análise de Correlação

- Dados por Postos:  $r$  de Spearman
- Coeficiente de correlação de Spearman
  - estatística não-paramétrica
    - pode ser usado quando os dados violarem suposições paramétricas, tais como dados não-normais
- Teste de Spearman
  - Classifica os dados em primeiro lugar e então aplicando a equação de Pearson aos dados ordenados
  - Categorias que podem ser ordenadas de maneira significativa, os dados são ordinais