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

Regressão Linear

- 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 variável: regressão linear <u>simples</u>
      - Outras variáveis: regressão linear <u>múltipla</u>
  - De relevante importância em
    - Economia, administração, contabilidade

Regressão Linear

- Regressão linear
  - Dados são emparelhados
  - Cada observação tem duas ou mais variáveis
  - Exemplos
    - Amostra de pessoas
      - Nome, consumo, renda, escolaridade, idade, altura, peso
    - Amostra de alunos
      - Nome, desempenho, hor_est_sem, idade, altura, peso
    - Amostra de empresas
      - Empresa, RSC, endividamento, rentabilidade, tangibilidade, valor da empresa

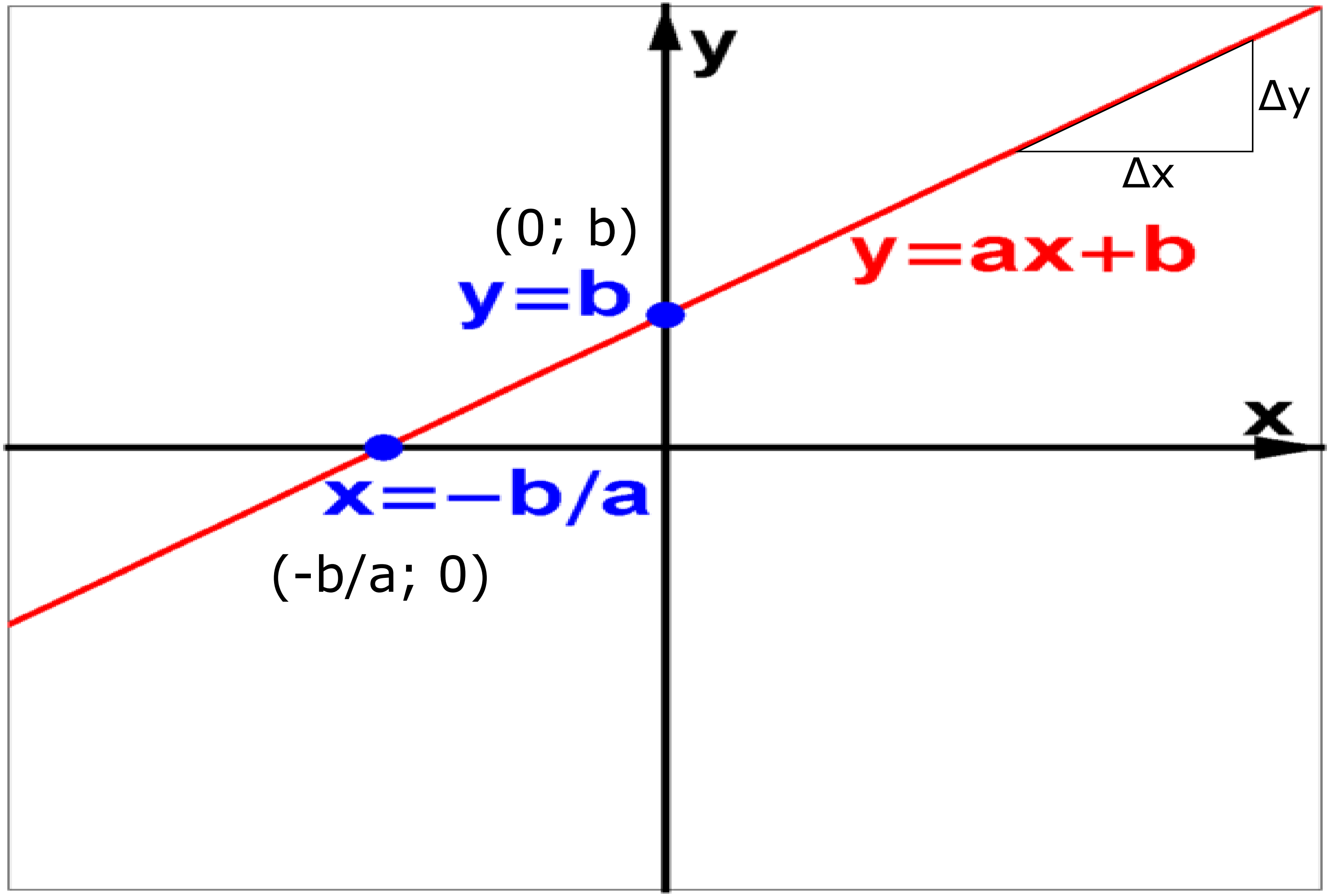
Regressão Linear

- Usos da equação de Regressão linear
  - Calcular/Estimar valores futuros/não observados
  - Calcular valores para evitar experimentos caros e/ou destrutivos
    - Em função de dados históricos obtém-se a equação
  - Estimação de efeitos
    - Adoção de políticas e efeitos na economia
    - Uso de certas técnicas e seus efeitos
  - Observação:
    - ***Sempre***, **a lógica da relação deve originar-se de teorias externas ao âmbito da estatística**

Regressão Linear

- A equação de Regressão linear
  - Corresponde a uma equação de uma reta
    - $\boldsymbol{y} = b + a\underline{\boldsymbol{x}} \approx a\underline{\boldsymbol{x}} + b \approx b_0 + b_1.\underline{\boldsymbol{x}}$

  - b = termo independente, constante, intercepto
  - a = coeficiente angular da reta
    - $a = \Delta y / \Delta x$
  - A reta intercepta o eixo ***y*** no ponto (0; b)
  - A reta intercepta o eixo ***x*** no ponto (-b/a; 0)


y
Δy
Δx
(0; b)
y=ax+b
y=b
x
x=−b/a
(-b/a; 0)

Regressão Linear

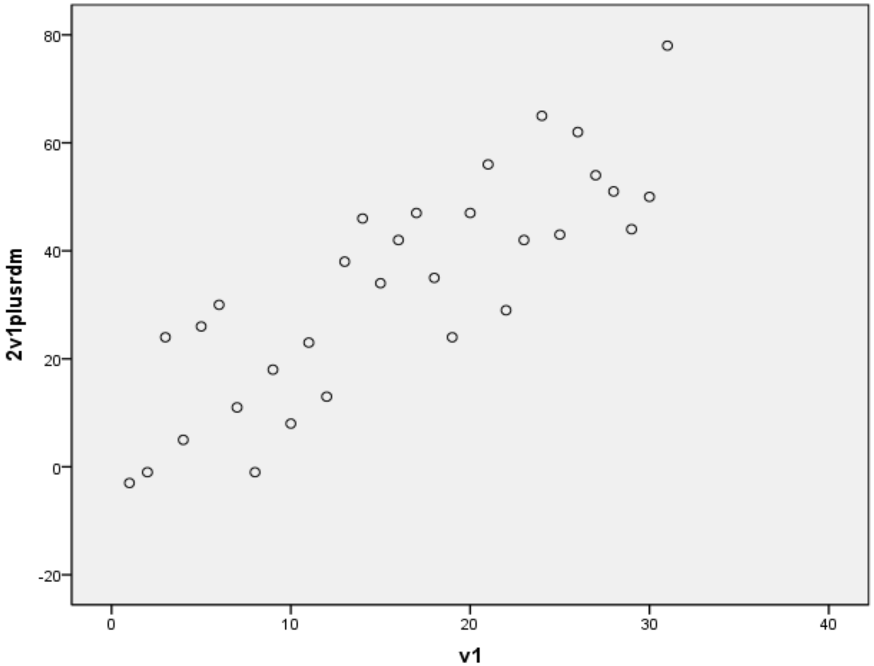
- A equação de Regressão linear
  - Corresponde a uma equação de uma reta
    - $\boldsymbol{y} = b + a\underline{\boldsymbol{x}} \approx a\underline{\boldsymbol{x}} + b \approx b_0 + b_1.\underline{\boldsymbol{x}}$
  - a = coeficiente angular da reta
    - $a = \Delta y / \Delta x$
    - ***O coeficiente angular indica quantas unidades <u>y</u> varia a cada variação de uma unidade de <u>x</u>***

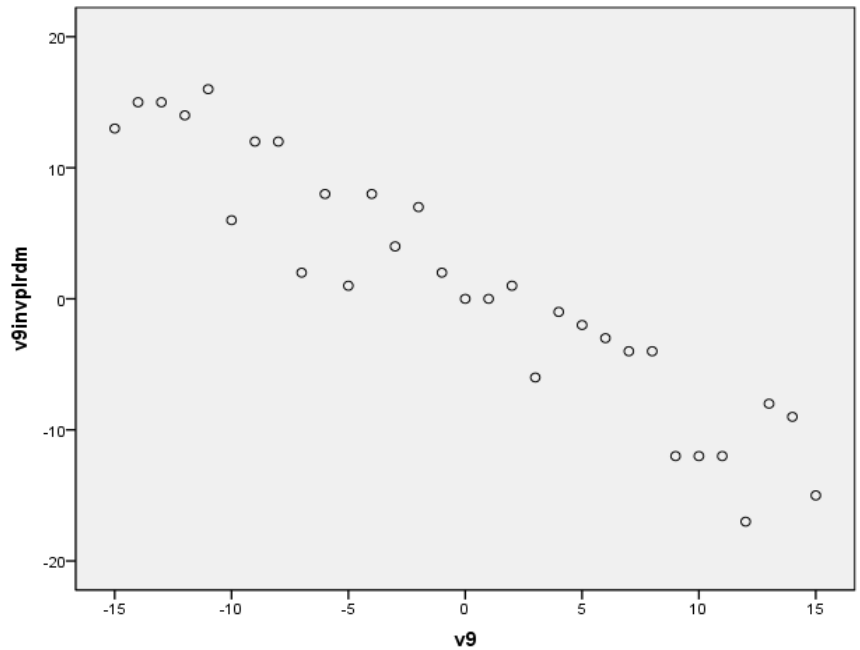
Regressão Linear

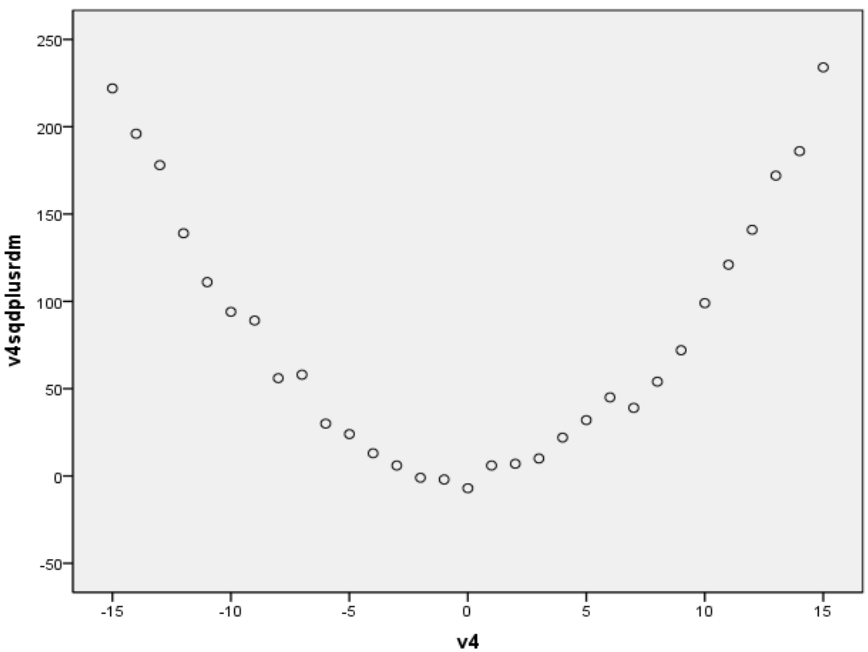
- A equação de Regressão linear
  - Corresponde a uma equação de uma reta
    - $\boldsymbol{y} = b + a\underline{\boldsymbol{x}} \approx a\underline{\boldsymbol{x}} + b \approx b_0 + b_1.\underline{\boldsymbol{x}}$
  - Como equação de uma regressão linear
    - ***y*** = variável a ser **predita**, **estimada**, **calculada**, "**explicada**" = variável dependente
    - ***x*** = variável **preditora**, **explicativa**, **independente**

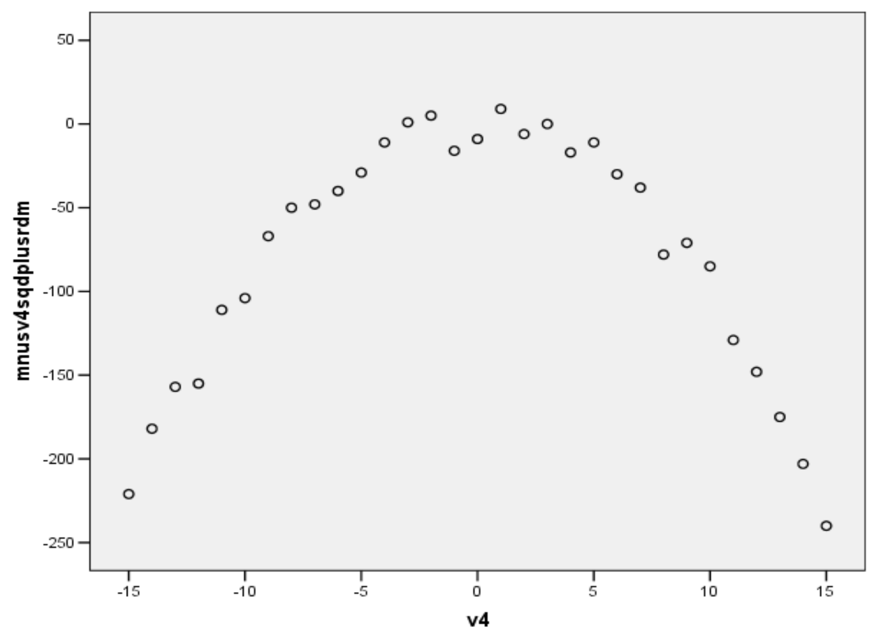
Regressão Linear

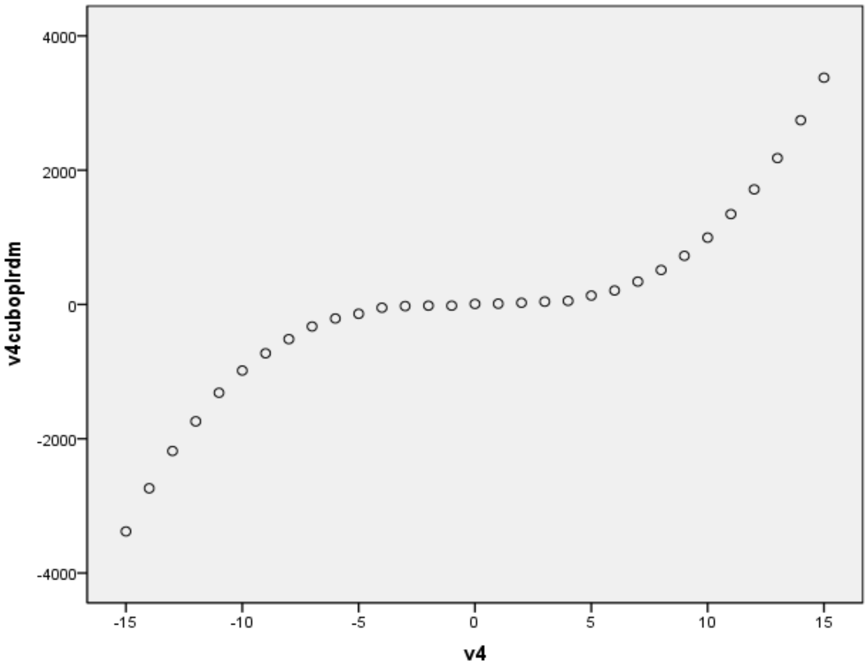
- Nem toda relação é aproximada por uma equação linear
- Ao examinar uma relação entre duas variáveis
  - Pode haver:
    - Inexistência de relação
    - Relação linear
    - Relação não linear
      - Quadrática
      - Cúbica
      - ...

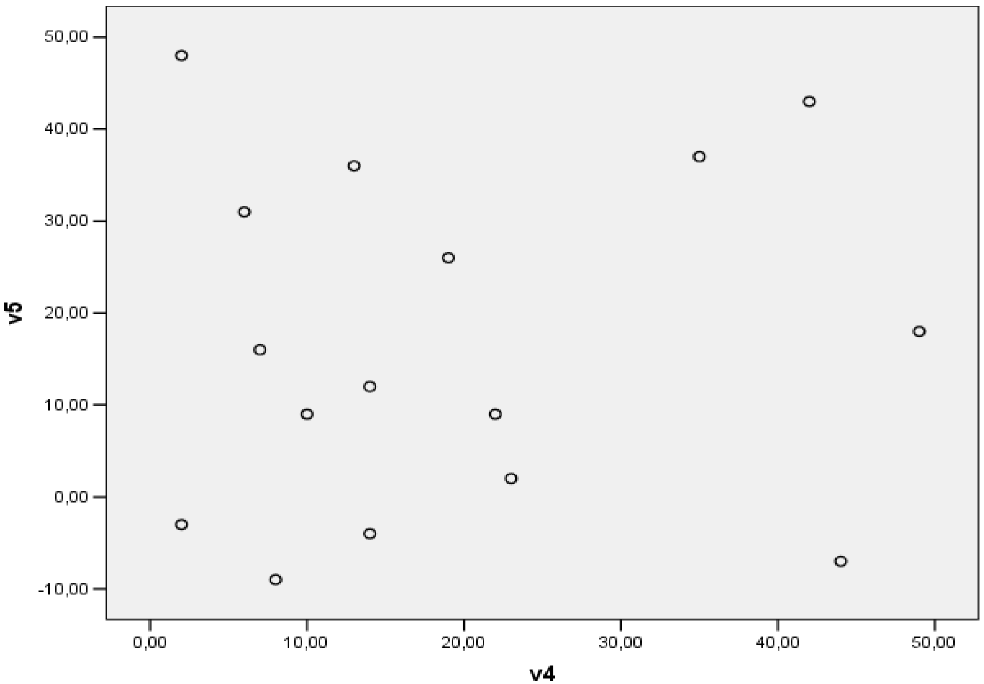

2v1plusrdm
v1


v9invplrdm
v9


v4sqdplusrdm
v4


mnusv4sqdplusrdm
v4


4000
2000
0
-2000
-4000
v4cuboplrdm
-15
-10
-5
0
5
10
15
v4


50,00
40,00
30,00
20,00
10,00
0,00
-10,00
v5
0,00
10,00
20,00
30,00
40,00
50,00
v4

Regressão Linear

- Determinação da equação linear
  - Valores de **y** podem ser preditos a partir de **x**?
    - Equação amostral (observada)

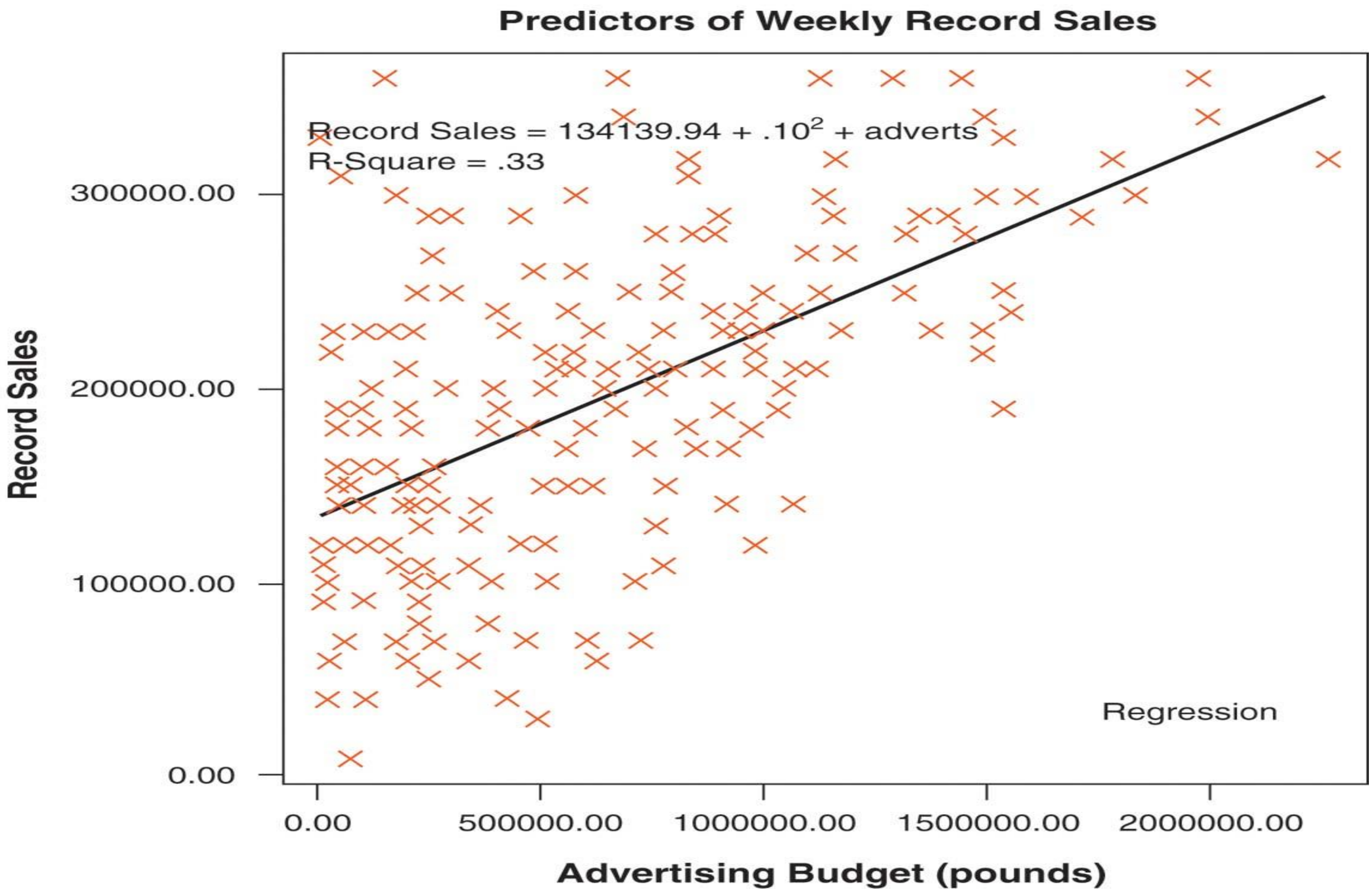
$$y_c = b_0 + b_1 x + erro$$

- A partir de dados amostrais pode-se determinar uma equação que bem represente a relação entre duas variáveis da população?
  - Equação que representa função de regressão populacional (não observada, estimada)

$$y_c = \beta_0 + \beta_1 x + erro$$

Regressão Linear

- Método para encontrar, ou ajustar, uma linha reta ao conjunto de pontos (pares ordenados) determinados pelas duas variáveis
  - Mínimos Quadrados Ordinários (MQO)
    - *Ordinary Least Squares* (OLS)
  - Outros


Predictors of Weekly Record Sales
Record Sales = 134139.94 + .10² + adverts
R-Square = .33
300000.00
200000.00
100000.00
0.00
Record Sales
0.00
500000.00
1000000.00
1500000.00
2000000.00
Advertising Budget (pounds)
Regression

Regressão Linear

- Mínimos Quadrados Ordinários (MQO)
  - Método mais simples e mais usado
  - Busca *minimizar os desvios* entre cada observação (par ordenado) e a reta da equação de regressão
  - Busca ***minimizar o erro***
  - Busca a *reta que apresente menores desvios* em relação aos vários pontos
  - Observações
    - *Soma dos desvios verticais dos pontos em relação à reta "estimada" é zero*
    - ***Soma dos quadrados dos desvios dos pontos em relação à reta "estimada" é mínima***

Regressão Linear

- Mínimos Quadrados Ordinários (MQO)
  - Método mais comum para ajustar uma linha ao conjunto de pontos observados
  - Duas características da reta resultante:
    - Soma dos desvios verticais (valor observado – valor calculado pela equação linear) é zero
    - Reta é a mais "próxima" de todos os pontos
      - Soma dos quadrados das distâncias é mínima
  - Equação da reta de regressão

$$y_c = b_0 + b_1 x + erro$$

Regressão Linear

- Mínimos Quadrados Ordinários (MQO)
  - Valor minimizado
    - Diferença entre y_i e y_c.
      - y_i = um valor observado de y
      - y_c = um valor calculado de y usando a equação de mínimos quadrados com valor observado de x emparelhado com valor y.

$$\sum (y_i - y_c)^2$$

$$y_c = b_0 + b_1 x + erro$$

Regressão Linear

- Mínimos Quadrados Ordinários (MQO)
  - Valores de b_0 e b_1 que minimizam a soma dos quadrados dos desvios são as soluções do seguinte sistema de equações:
    - n = número de pares de observações

$$\sum y = nb_0 + b_1\left(\sum x\right)$$

$$\sum xy = b_0\left(\sum x\right) + b_1\left(\sum x^2\right)$$

Regressão Linear

- Mínimos Quadrados Ordinários (MQO)
  - Valores de b_0 e b_1

$$b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$b_0 = \frac{\sum y - b_1 \sum x}{n}$$

- Exemplo:
  - Quilometragem de um carro pode afetar seu preço de venda? Pode-se propor que sim como hipótese? Considerando que sim. Qual o grau de influência da **quilometragem** sobre o preço de venda?
    - x = quilômetros rodados (em milhares)
    - y = preço do veículo

Regressão Linear

- Outros fatores além da quilometragem tambem podem afetar o preço
- “isolando” outros possíveis fatores
- Tentar encontrar uma equação que “indique” o efeito da quilometragem sobre o preço

Regressão Linear

- Dispondo de uma amostra (conjunto de observações de carros vendidos) pode-se calcular b_0 e b_1

$$b_1 = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$b_0 = \frac{\sum y - b_1 \sum x}{n}$$

- Exemplo: quilometragem e preço do carro
  - $y = 2.933,6 - 38,56\,x$

Obs.	x	y	x.y	x2	y2		
1	40	1.000	40.000	1.600	1.000.000		
2	30	1.500	45.000	900	2.250.000	n =	14
3	30	1.200	36.000	900	1.440.000		
4	25	1.800	45.000	625	3.240.000	b1 =	-38,56
5	50	800	40.000	2.500	640.000		
6	60	1.000	60.000	3.600	1.000.000		
7	65	500	32.500	4.225	250.000	b0 =	2.933,6
8	10	3.000	30.000	100	9.000.000		
9	15	2.500	37.500	225	6.250.000		
10	20	2.000	40.000	400	4.000.000		
11	55	800	44.000	3.025	640.000		
12	40	1.500	60.000	1.600	2.250.000		
13	35	2.000	70.000	1.225	4.000.000		
14	30	2.000	60.000	900	4.000.000		
soma	505	21.600	640.000	21.825	39.960.000		

Exemplo: quilometragem e preço do carro

- $y = 2.933,6 - 38,56\ x$

obs	x_Km	y_preco observado	y_preco calculado	(yi-yc)^2
1	40	1,000.00	1,391.60	153350.56
2	30	1,500.00	1,777.20	76839.84
3	30	1,200.00	1,777.20	333159.84
4	25	1,800.00	1,970.00	28900
5	50	800.00	1,006.00	42436
6	60	1,000.00	620.40	144096.16
7	65	500.00	427.60	5241.76
8	10	3,000.00	2,548.40	203942.56
9	15	2,500.00	2,355.60	20851.36
10	20	2,000.00	2,162.80	26503.84
11	55	800.00	813.20	174.24
12	40	1,500.00	1,391.60	11750.56
13	35	2,000.00	1,584.40	172723.36
14	30	2,000.00	1,777.20	49639.84
				1269609.9

- Mínimo

- $y = 2.933,6 - 38,56\ x$
- A equação indica que a cada 1.000 quilômetros o carro perde $38,56 de valor

Observ.	x	Y observado	Y calculado
8	10	3.000	2.548
	12		*2.471*
9	15	2.500	2.355
	18		*2.240*
10	20	2.000	2.162
4	25	1.800	1.970
2	30	1.500	1.777
3	*32*		*1.700*
14	*33*		*1.661*
13	35	2.000	1.584
1	*38*		*1.469*
12	40	1.500	1.391
	45		*1.199*
5	50	800	1.006
	53		*890*
11	55	800	813
6	60	1.000	620
	62		*543*
7	65	500	428

Regressão Linear

- Sobre a equação que aproxima a função de regressão populacional

$$y_c = \beta_0 + \beta_1 x + erro$$

  - O resultado é uma estimativa
  - Equação de regressão é uma estimação da real relação
  - Trata-se de uma relação média
  - O valor estimado pela equação pode não ser exato
  - A relação linear pode não manter-se fora do escopo da amostra

Regressão Linear

- Inferência em análise de regressão

$$y_c = \beta_0 + \beta_1 x + erro$$

  - b_0 e b_1 são estimativas pontuais dos parâmetros populacionais β_0 e β_1
  - *erro* representa a dispersão na população
  - Ausência de *erro* significaria todos os pontos sobre uma linha reta => relacionamento perfeito
  - Relacionamento perfeito “quase” impossível
  - Há outras variáveis **x** que influenciam o valor de **y**
  - Cada variável **x** terá seu grau de influência em **y**

Regressão Linear

- Inferência em análise de regressão

$$y_c = \beta_0 + \beta_1 x + erro$$

  - Cada variável **x** terá seu grau de influência em **y**
  - Haveria distintos conjuntos de valores da variável **x** capazes de explicar **y**?
    - Dois fatores minimizam esta possibilidade
      - Idoneidade da amostra
      - Elevado número de observações
  - Haveria outros conceitos (variáveis) capazes de explicar **y**?
      - Idade, Conservação, Cidade, Número de proprietários, Categoria dos proprietários ...

Regressão Linear

- Inferência em análise de regressão
  - Análise de regressão supõe que
    - Para cada **x** há uma distribuição de potenciais **y** com distribuição normal
    - Dado **x**, os vários valores de **y** correspondentes têm distribuição normal
    - A média de cada distribuição equivale ao valor médio de **y** na população

Regressão Linear

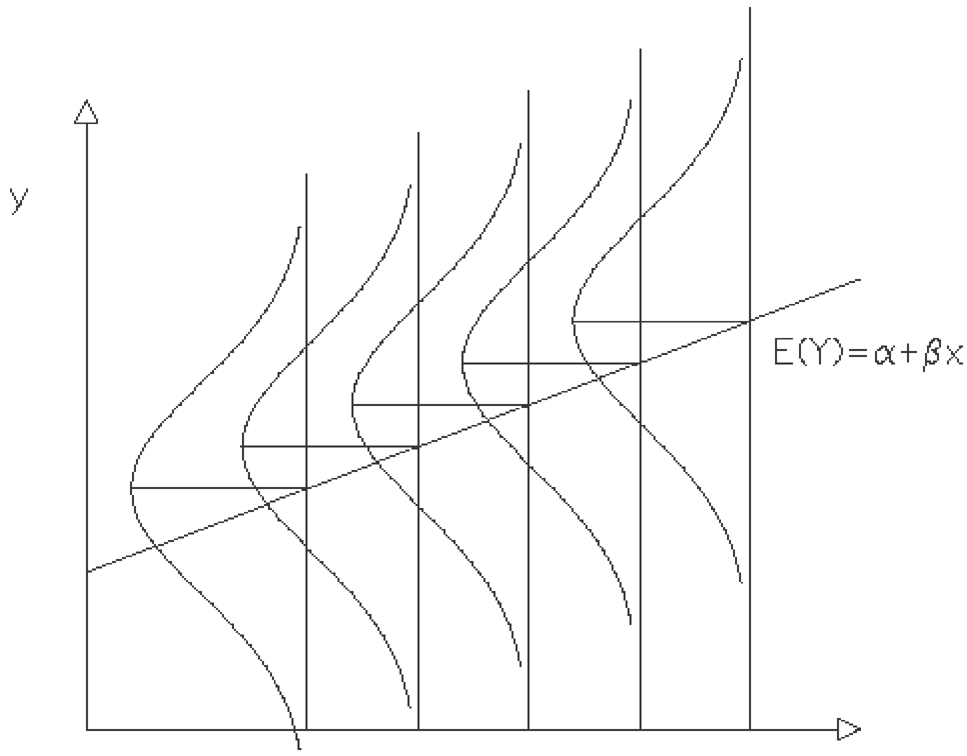
- Inferência em análise de regressão
  - Análise de regressão supõe que
    - Para cada **x** há uma distribuição de potenciais **y** com distribuição normal
    - Dado **x**, os vários valores de **y** correspondentes têm distribuição normal
      - Distribuição condicional de **y** dado **x**
    - A média de cada distribuição equivale ao valor médio de **y** na população
    - Distribuições condicionais de **y**, para cada valor de **x**, têm mesmo desvio padrão

Regressão Linear

- Hipóteses de Análise de Regressão
  - Variável dependente (**y**) é aleatória
  - Para cada **x** há uma distribuição condicional de **y** que é uma distribuição normal
  - As distribuições normais de **y** têm igual desvio padrão (homoscedasticidade = homogeneidade ou uniformidade de variância)

Regressão Linear

- Homoscedasticidade



11

x	y	y-calc	DN y
10	3.000	2.548	2.551
10		2.548	2.547
10		2.548	2.544
10		2.548	2.543
10		2.548	2.547
10		2.548	2.541
10		2.548	2.545
10		2.548	2.558
10		2.548	2.556
30	2.500	1.777	1.772
30		1.777	1.784
30		1.777	1.769
30		1.777	1.781
30		1.777	1.774
30		1.777	1.789
30		1.777	1.779
30		1.777	1.783
30		1.777	1.780
30		1.777	1.771
65	500	428	436
65		428	436
65		428	417
65		428	436
65		428	429
65		428	433
65		428	431
65		428	439
65		428	439
65		428	432
65		428	430

Regressão Linear

- Erro Padrão da Estimativa
  - Qual a precisão das estimativas de regressão?
  - Dispersão populacional pode ser estimada com base na dispersão amostral (em relação à reta de regressão)?
    - Quanto menor o erro (dispersão em relação à reta de regressão)
      - Maior a precisão das estimativas
      - Maior a capacidade explicativa do modelo

Regressão Linear

- Erro Padrão da Estimativa
  - Dispersão na população afeta precisão da estimação
  - Maior dispersão => menor precisão da estimação
  - Estimação do *erro* (dispersão) populacional com base na dispersão amostral
    - **Desvio padrão** em relação à reta de regressão
      - (n – 2): dois graus de liberdade ao calcular-se dois valores (b_0 e b_1) na equação de regressão

$$s_e = \sqrt{\frac{\sum (y_i - y_c)^2}{n-2}}$$

Regressão Linear

- Erro Padrão da Estimativa
  - Baseado na hipótese de igualdade de desvio padrão entre as várias distribuições condicionais de **y** para cada **x**
    - Hipótese de dispersão uniforme
      - ***homoscedasticidade***
- Fórmula alternativa para erro padrão da estimativa que não requer y_c

$$s_e = \sqrt{\frac{\sum y^2 - b_0 \sum y - b_1 \sum xy}{n-2}}$$

Regressão Linear

- **Inferência sobre o coeficiente angular da reta**
  - Testar se parâmetro *coeficiente angular da reta* (b_1) é, ou não, nulo
  - Um coeficiente angular nulo ($b_1 = 0$) significa que **x** não tem influência sobre **y**
    - Ausência de relacionamento entre **x** e **y**
  - Hipótese a ser verificada (para o parâmetro populacional)

H_0: $\beta_1 = 0$

significa que x *<u>não tem</u>* efeito sobre y

H_1: $\beta_1 \neq 0$ ($\neq$, > , <)

significa que x *<u>tem</u>* efeito sobre y

Regressão Linear

- Inferência sobre o coeficiente angular da reta (b_1)
  - Hipótese a ser verificada
    - **H_0: $\beta_1 = 0$**
    - **H_1: $\beta_1 \neq 0$**
  - Estatística de teste
    - Diferença entre coeficiente amostral (**b_1**) e **0** (zero)
    - dividido pelo desvio padrão da distribuição amostral do coeficiente angular

$$t = \frac{valor\ amostral - valor\ referência}{desvio\ padrao\ da\ distribuicao\ amostral\ do\ coeficiente\ angular}$$

Regressão Linear

- Inferência sobre o coeficiente angular da reta (b_1)
  - Desvio padrão da distribuição amostral do coeficiente angular

$$s_b = s_e \cdot \sqrt{\frac{1}{\sum x^2 - \left[\frac{(\sum x)^2}{n}\right]}}$$

  - Estatística de teste

$$t_{teste} = \frac{b_1 - 0}{s_b}$$

Regressão Linear

- Inferência sobre o coeficiente angular da reta (b_1)
  - Hipótese a ser verificada

    H_0: $\beta_1 = 0$

    coeficiente angular é nulo. x não tem efeito sobre y.

    H_1: $\beta_1 \neq 0$

    coeficiente angular é distinto de zero. x tem efeito sobre y. cada incremento de x ocasiona um efeito de b_1 unidades em y.
  - Teste de significância de b_1
    - Se t_{teste} supera $t_{crítico}$ (zona de rejeição de H0)
      - Rejeita-se H_0 e aceita-se H_1
        - Significa que o coeficiente angular é significativamente distinto de zero

- Exemplo: quilometragem e preço do carro
  - *Teste de significância de b_1*
  - $y = 2.933{,}6 - 38{,}56\,x$

Obs.	x	y	xy	x2	y2		
1	40	1.000	40.000	1.600	1.000.000		
2	30	1.500	45.000	900	2.250.000	n = 14	14
3	30	1.200	36.000	900	1.440.000		
4	25	1.800	45.000	625	3.240.000	b1 =	-38,56
5	50	800	40.000	2.500	640.000		
6	60	1.000	60.000	3.600	1.000.000		
7	65	500	32.500	4.225	250.000	b0 =	2.933,6
8	10	3.000	30.000	100	9.000.000		
9	15	2.500	37.500	225	6.250.000	s_e =	324,55
10	20	2.000	40.000	400	4.000.000	s_b =	5,4
11	55	800	44.000	3.025	640.000		
12	40	1.500	60.000	1.600	2.250.000		
13	35	2.000	70.000	1.225	4.000.000	t_{teste}=	***-7,12078***
14	30	2.000	60.000	900	4.000.000		
soma	505	21.600	640.000	21.825	39.960.000		

Probabilidades da Distribuição 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

Regressão Linear

- Inferência sobre o coeficiente angular da reta (b_1)
  - Hipótese a ser verificada

    H_0: $\beta_1 = 0$

    H_1: $\beta_1 \neq 0$
  - Teste de significância de b_1
    - Se t_{teste} supera $t_{crítico}$ em valor absoluto
      - Rejeita-se H_0 e aceita-se H_1
      - No exemplo, t_{teste} (-7,12078) supera $t_{crítico}$ de modo que o coeficiente angular (-38,56) é significativamente distinto de zero (no caso é inferior a zero)
      - Há uma correlação significativamente negativa entre x e y
      - De fato, a quilometragem do veículo afeta negativamente seu valor de venda
      - A cada 1.000Km o preço do veículo cai $ 38,56.

Regressão Linear

- ***Coeficiente de determinação (r^2)***
  - ***Medida que avalia grau de predição da equação***
  - Grau de ajuste da reta de regressão da amostra aos dados
  - Predição baseada na reta de regressão X predição baseada na média
  - Em que grau as predições da reta de regressão são melhores que as baseadas na média?
  - Menor dispersão levará a melhor predição
  - Dispersão de pontos (y_i) em torno da média de y X dispersão em torno da reta de regressão (y_c)

Regressão Linear

- ***Coeficiente de determinação (r^2)***
  - ***Medida que avalia grau de predição da equação***
  - Dispersão de pontos (y_i) em torno da média de y X dispersão em torno da reta de regressão (y_c)
    - Se a dispersão (erro) em torno da reta de regressão ($y_i - y_c$) é muito menor que aquela (erro) em torno da média ($y_i - y_{med}$)
      - Predições baseadas na reta de regressão são melhores que as da média

Regressão Linear

- Coeficiente de determinação (r^2)

  - **Variação de y_i em torno da média de y**
    - ***Variação TOTAL***
    - Soma de quadrados de desvios entre cada valor observado (y_i) e a média de y

  - **Variação de y em torno da reta de regressão y_c**
    - ***Variação NÃO EXPLICADA***
      - não se sabe a razão da estimação da reta diferir dos valores observados
    - Soma de quadrados de desvios entre cada valor observado (y_i) e cada valor calculado pela equação de regressão (y_c)

Regressão Linear

- Coeficiente de determinação (r^2)
  - Variação de y_i em torno da média de y
    - Variação TOTAL (s_y^2)

$$variacao\ total = \sum (y_i - \bar{y})^2$$

  - Variação de y em torno da reta de regressão y_c
    - Variação NÃO EXPLICADA (s_e^2)

$$variacao\ nao\ explicada = \sum (y_i - y_c)^2$$

Regressão Linear

- Coeficiente de determinação (r^2)
  - Quantidade de desvio explicada pela reta de regressão (variação EXPLICADA)
    - Diferença entre Variação TOTAL e Variação NÃO EXPLICADA

$$variacao\ explicada = variacao\ total - variacao\ nao\ explicada$$

  - **Percentual de variação explicada (r^2)**
    - Razão entre variação explicada e variação total

$$r^2 = \frac{variacao\ explicada}{variacao\ total} = \frac{variacao\ total - variacao\ nao\ explicada}{variacao\ total}$$

Regressão Linear

- Coeficiente de determinação (r^2)
  - Percentual de variação explicada (r^2)

$$r^2 = \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$$

  - Onde
    - s_y^2 é a variância de y em relação à média (variação TOTAL)
    - s_e^2 é a variância de y em relação à reta (variação NÃO EXPLICADA)

Regressão Linear

- s_y^2 é a variância de y em relação à média (total)

$$s_y^2 = \frac{\sum(y_i - \bar{y})^2}{n-2} \qquad s_y^2 = \frac{\left(\sum y^2\right) - \frac{\left(\sum y\right)^2}{n}}{n-2}$$

- s_e^2 é a variância de y em relação à reta (não explicada)

$$s_e^2 = \frac{\sum(y_i - y_c)^2}{n-2} \qquad s_e^2 = \frac{\sum y^2 - b_0 \sum y - b_1 \sum xy}{n-2}$$

- Exemplo: quilometragem e preço do carro
  - $y = 2.933{,}6 - 38{,}56\,x$

Obs.	x	y	xy	x2	y2		
1	40	1.000	40.000	1.600	1.000.000		
2	30	1.500	45.000	900	2.250.000	n	14
3	30	1.200	36.000	900	1.440.000		
4	25	1.800	45.000	625	3.240.000	b_1 =	-38,56
5	50	800	40.000	2.500	640.000		
6	60	1.000	60.000	3.600	1.000.000		
7	65	500	32.500	4.225	250.000	b_0 =	2.933,6
8	10	3.000	30.000	100	9.000.000		
9	15	2.500	37.500	225	6.250.000	t_{teste} =	**-7,12078**
10	20	2.000	40.000	400	4.000.000		
11	55	800	44.000	3.025	640.000	***r^2 =***	***0,81***
12	40	1.500	60.000	1.600	2.250.000		
13	35	2.000	70.000	1.225	4.000.000		
14	30	2.000	60.000	900	4.000.000		
soma	505	21.600	640.000	21.825	39.960.000		

Regressão Linear

- Coeficiente de determinação (r^2)

  - $y = \mathbf{2.933,6} - \mathbf{38,56}\, x$
  - $\boldsymbol{r^2 = 0,81}$

  - Aproximadamente 81% da variação em y é explicada por x

Regressão Linear

- Teste de independência entre as variáveis do modelo
- Análise de Variância para regressão simples
  - Teste dos coeficientes estimados
  - ***“teste da significância global do modelo”***

Regressão Linear

- Análise de Variância para regressão simples

  H0: Há relacionamento entre variáveis

  H1: Não relacionamento entre variáveis, elas são realmente independentes

$$F = \frac{estimativa\ \text{"entre"}\ da\ variancia}{estimativa\ \text{"dentro"}\ da\ variancia}$$

$$F = \frac{\dfrac{(soma\ de\ quadrados\ \text{entre})}{k-1}}{\dfrac{(soma\ de\ quadrados\ \text{dentro})}{n-2}} = \frac{\dfrac{\sum(y_c - \bar{y})^2}{k-1}}{\dfrac{\sum(y_i - y_c)^2}{n-2}}$$

Regressão Linear

- Análise de Variância para regressão simples
  - Teste dos coeficientes estimados
  - ***"teste da significância global do modelo"***
  - F calculado comparado com F crítico
  - Se Fcalculado supera Fcrítico
    - Rejeita-se H0 de igualdade entre coeficientes estimados das variáveis, ou seja, de dependência entre variáveis
    - Aceita-se H1 de não relacionamento entre variáveis, ou seja, de **Independência** das variáveis. O conjunto de variáveis explicativas é bom.

Regressão Linear

- Teste de independência entre os coeficientes

$$F = \frac{\dfrac{(soma\ de\ quadrados\ \text{entre})}{k-1}}{\dfrac{(soma\ de\ quadrados\ \text{dentro})}{n-2}} = \frac{\dfrac{\sum(y_c - \bar{y})^2}{k-1}}{\dfrac{\sum(y_i - y_c)^2}{n-2}}$$

  - F = 5.370.295 / 105.333 = **50,98**
    - df1 = 1; df2 = 12; α = 0,05; $F_{crítico}$ = **4,75**
    - F calculado supera **4,75**   **(50,98 > 4,75)**
      - Rejeita-se H0
      - Aceita-se H1: variáveis são independentes

Regressão Linear Múltipla

- Regressão Linear Múltipla envolve
  - Três ou mais variáveis
  - Duas ou mais variáveis independentes
- Objetivo da regressão linear múltipla
  - Estabelecer uma equação que permita estimar, ou predizer, valores de y a partir de valores de várias variáveis independentes x
  - Mais variáveis independentes melhoram a capacidade de predição em comparação com a regressão simples
  - Técnica de MQO para obtenção de equação
    - Mais cálculos e complexidade

Regressão Linear Múltipla

- Regressão Linear Múltipla
  - Forma da equação de regressão
    - Amostral

$$y_c = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \cdots + b_k x_k + erro$$

    - Populacional

$$y_c = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + erro$$

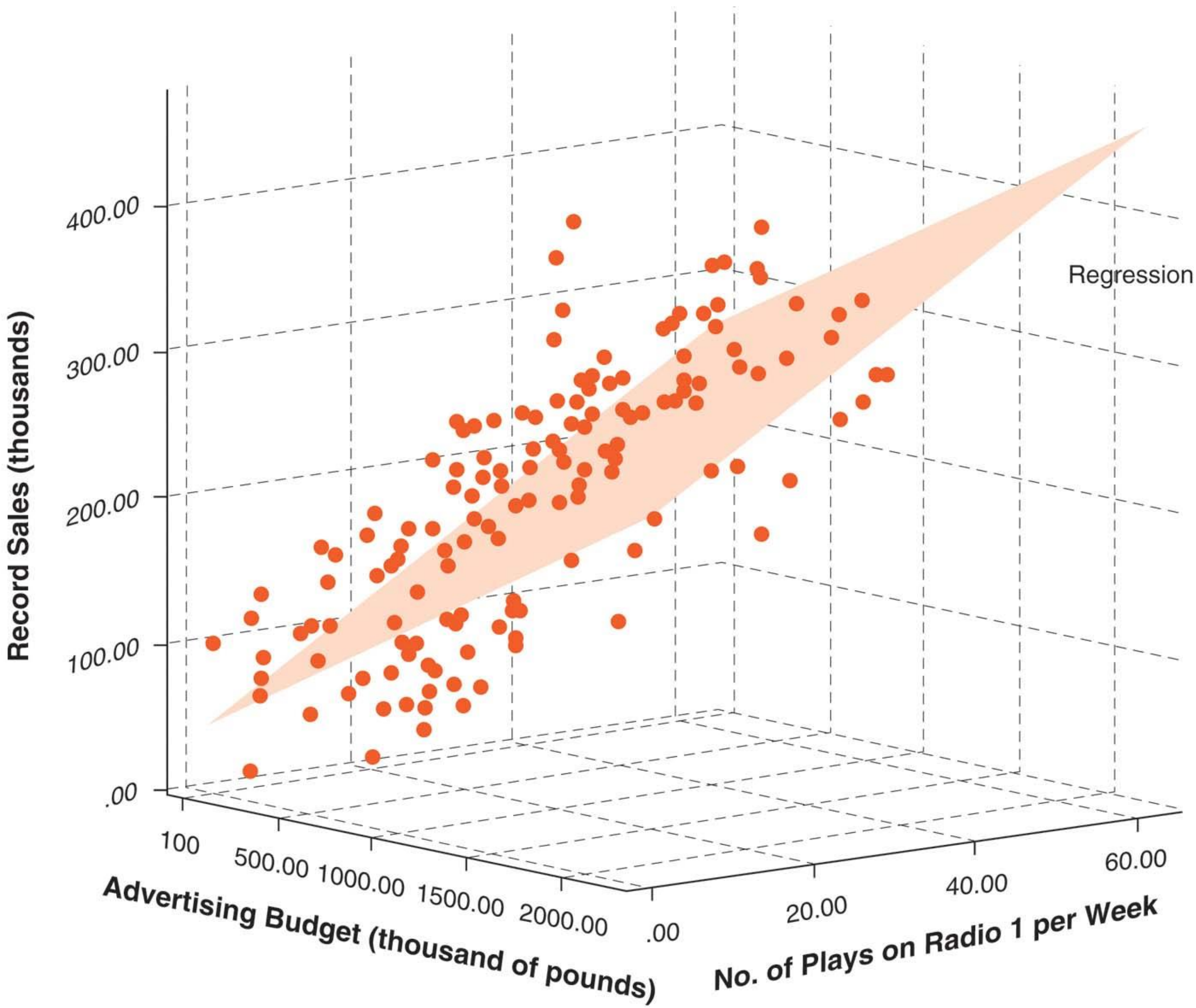
  - β_0 = intercepto
  - k = número de variáveis independentes
  - β_j (j = 1, k) = coeficientes angulares

Regressão Linear Múltipla

- Regressão Linear Múltipla
  - Forma da equação de regressão

$$y_c = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + erro$$

  - Ao invés de linha de regressão tem-se
  - ***<u>“Plano” de regressão</u>*** para três variáveis
  - “Hiperplano” para mais variáveis (k)
  - Quanto menor a dispersão dos pontos em relação ao “plano” de regressão, melhor a precisão das estimações


Record Sales (thousands)
400.00
300.00
200.00
100.00
.00
Regression
100
500.00
1000.00
1500.00
2000.00
Advertising Budget (thousand of pounds)
.00
20.00
40.00
60.00
No. of Plays on Radio 1 per Week

Regressão Linear Múltipla

- Regressão Linear Múltipla
  - Poucos fenômenos podem ser explicados por uma única variável
  - Objetivo é escolher as melhores variáveis explicativas dentre muitas possíveis
  - Ideal
    - ***Mais elevada capacidade explicativa do modelo com o mínimo de variáveis independentes (explicativas)***

Regressão Linear Múltipla

- Regressão Linear Múltipla
  - Escolha de melhores variáveis explicativas
    - Conhecimento do fenômeno estudado é fundamental
    - Revisão da literatura sempre essencial
  - Proposição de hipóteses deve ser independente dos dados

Regressão Linear Múltipla

- Regressão Linear Múltipla
  - Escolha de melhores variáveis explicativas
    - Conhecimento do fenômeno estudado é fundamental
- Exemplos

Var. Dependente	**Variáveis Independentes**
Estrutura de capital	Rentabilidade, tangibilidade, tamanho, estrutura de propriedade...
Volume de vendas do produto	Qualidade, preço, propaganda,...
Salário	Qualificação, inteligência, gênero, dedicação
Investimento	Endividamento, fluxo de caixa, estrutura de propriedade, investimento prévio,...
Safra agrícola	Chuva, tipo de solo, técnica de plantio, técnica de tratamento do solo,...

Regressão Linear Múltipla

- Regressão Linear Múltipla
  - Escolha de melhores variáveis explicativas
    - Conhecimento do fenômeno estudado é fundamental
  - Levantamento de variáveis (conceitos) possíveis
  - Análise de correlação entre variáveis
    - Evitar uso de variáveis independentes muito correlacionadas
  - Estimação de modelos alternativos
    - Avaliação de r^2
      - Indica capacidade explicativa do modelo
    - Avaliação de F
      - Indica grau de independência entre coeficientes

Regressão Linear Múltipla

- Econometria
  - Significa “medida econômica”
  - Aplicação da estatística a dados econômicos para dar suporte a modelos econômicos propostos teoricamente
  - Trata da verificação empírica de leis econômicas
- Contabilometria?
- *metria?