

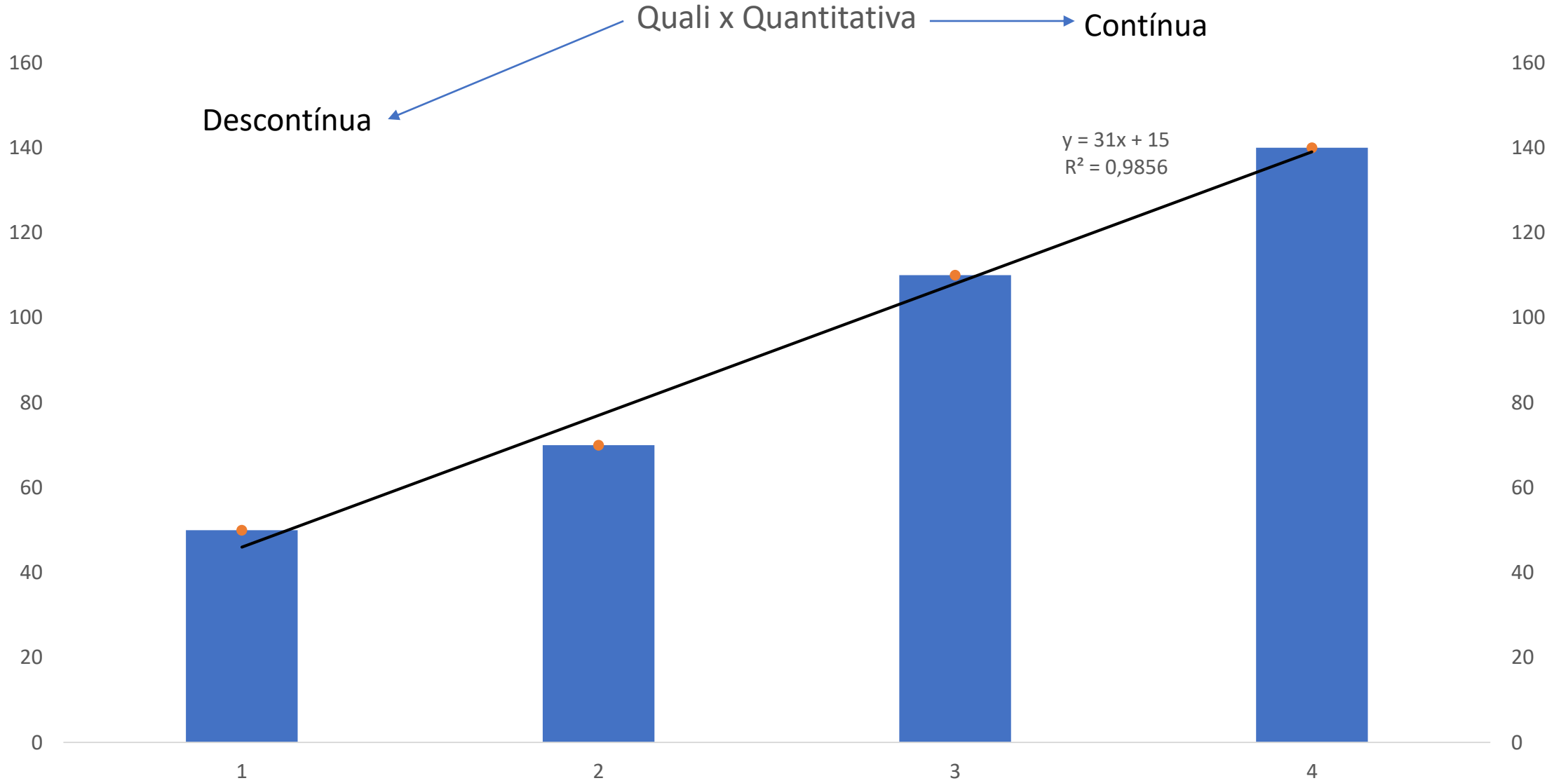
Regressão Linear Simples

Mario Andrade Lira Junior

**Direitos autorais reservados segundo licença
Creative Commons 3.0: com Atribuição; Não
Comercial; Compartilha Igual**

Termos importantes

- Tipos de variáveis
 - Qualitativa x Quantitativa
 - Dependente x Independente
- Regressão
- Correlação
- Relação causa-efeito
- Extrapolação x Interpolação
- Valor ajustado - estimativa da população
- Resíduo- diferença entre ajustado e real (variação do acaso)



Regressão x Correlação

- Diferença

- Regressão - equação ligando duas ou mais variáveis
- Correlação – medida do grau de ligação entre duas variáveis

- Usos

- Regressão – interpolar valores na faixa usada no experimento
- Correlação – indicar variáveis com ligação no padrão de comportamento
 - Semelhante
 - Oposto

Coeficientes

- Correlação (r)
 - Mede o grau de relação entre variáveis
 - Faixa de valores
 - 0 a 1
 - Positivo ou negativo
 - Interpretação
 - Quanto mais próximo de 1/-1, maior a ligação entre as variáveis
- Determinação (r^2)
 - Quanto a regressão explica dos dados
 - Faixa de valores
 - 0 a 1
 - Interpretação
 - Quanto maior, mais a regressão explica

Causas de variação

- Médias - variável independente
 - Assumimos um efeito linear neste caso
 - Existem muitos outros modelos de regressão em que a variável independente tem outros tipos de efeito
- Acaso

Requisitos

- Variável independente medida sem erro (variação do acaso)
 - Ou seja, só tem variação do acaso para Y
- Y pode ser explicado por reta em função de X
 - Para este modelo de regressão – outros modelos admitem outros “formatos”
- Para cada X os Y's são
 - Independentes, ou seja, variação do acaso independente
 - Resíduos com média zero
- Homocedase

“Leitura” de uma regressão linear simples

- $Y = a + bx$

- Y – variável dependente

- a – y para $x = 0$

- b – quanto y varia para cada x

- Quanto maior em absoluto, mais Y responde a X

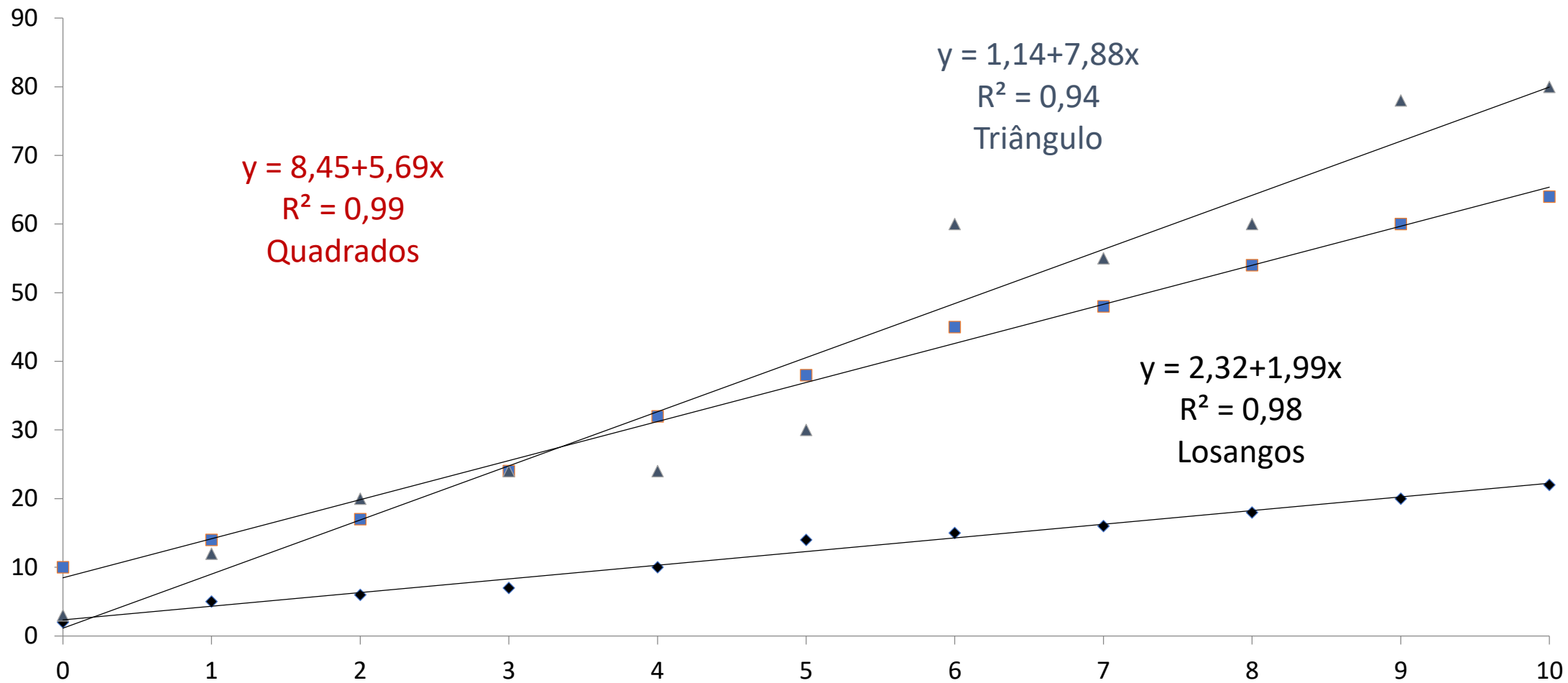


Figura 1 – Representação esquemática de três regressões lineares simples aleatórias, para interpretação dos coeficientes

Equações para cálculo da regressão

$$r = \frac{\sum xy - \left(\frac{(\sum x)(\sum y)}{N} \right)}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{N} \right] \left[\sum y^2 - \frac{(\sum y)^2}{N} \right]}}$$

$$b = \frac{\sum xy - \frac{(\sum x)(\sum y)}{N}}{\left[\sum x^2 - \frac{(\sum x)^2}{N} \right]}$$

$$a = \bar{y} - b\bar{x}$$

Um experimento foi conduzido em Saturno em 2100 avaliando a resposta de cana de açúcar à adubação nitrogenada, com os tratamentos e resultados abaixo. Calcule a regressão entre as duas variáveis

N (kg.ha ⁻¹)	Produtividade (t.ha ⁻¹)
0	64,64
0	75,06
0	88,14
0	105,71
50	88,88
50	109,85
50	130,66
50	152,69
100	113,12
100	146,47
100	170,06
100	211,42
150	153,52
150	177,59
150	207,39
150	246,65
200	185,84
200	216,04
200	248,87
200	293,63

Cálculo

x	y	x ²	y ²	xy
0	64,64	0	4178,3296	0
0	75,06	0	5634,0036	0
0	88,14	0	7768,6596	0
0	105,71	0	11174,6041	0
50	88,88	2500	7899,6544	4444
50	109,85	2500	12067,0225	5492,5
50	130,66	2500	17072,0356	6533
50	152,69	2500	23314,2361	7634,5
100	113,12	10000	12796,1344	11312
100	146,47	10000	21453,4609	14647
100	170,06	10000	28920,4036	17006
100	211,42	10000	44698,4164	21142
150	153,52	22500	23568,3904	23028
150	177,59	22500	31538,2081	26638,5
150	207,39	22500	43010,6121	31108,5
150	246,65	22500	60836,2225	36997,5
200	185,84	40000	34536,5056	37168
200	216,04	40000	46673,2816	43208
200	248,87	40000	61936,2769	49774
200	293,63	40000	86218,5769	58726
2000	3186,2	300000	585295,035	394859,5

$$b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\left[\sum x^2 - \frac{(\sum x)^2}{n} \right]}$$

$$b = \frac{(0 \times 64,64 + 0 \times 75,06 + \dots + 200 \times 293,63) - \frac{(0+0+\dots+200) \times (64,64 + 75,06 + \dots + 293,63)}{20}}{\left[(0^2 + 0^2 + \dots + 20^2) - \frac{(0+0+\dots+20)^2}{20} \right]}$$

$$b = \frac{394859,5 - \frac{2000 \times 3186,23}{20}}{\left[300000 - \frac{2000^2}{20} \right]} \cong 0,76$$

$$a = \bar{y} - b\bar{x} = \frac{\sum y}{n} - b \frac{\sum x}{n} = \frac{3186,23}{20} - 0,762365 \frac{2000}{20} \cong 83,08$$

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n} \right] \times \left[\sum y^2 - \frac{(\sum y)^2}{n} \right]}} = \frac{394859,5 - \frac{2000 \times 3186,23}{20}}{\sqrt{\left[300000 - \frac{2000^2}{20} \right] \times \left[585295 - \frac{(3186,23)^2}{20} \right]}} = 0,86$$

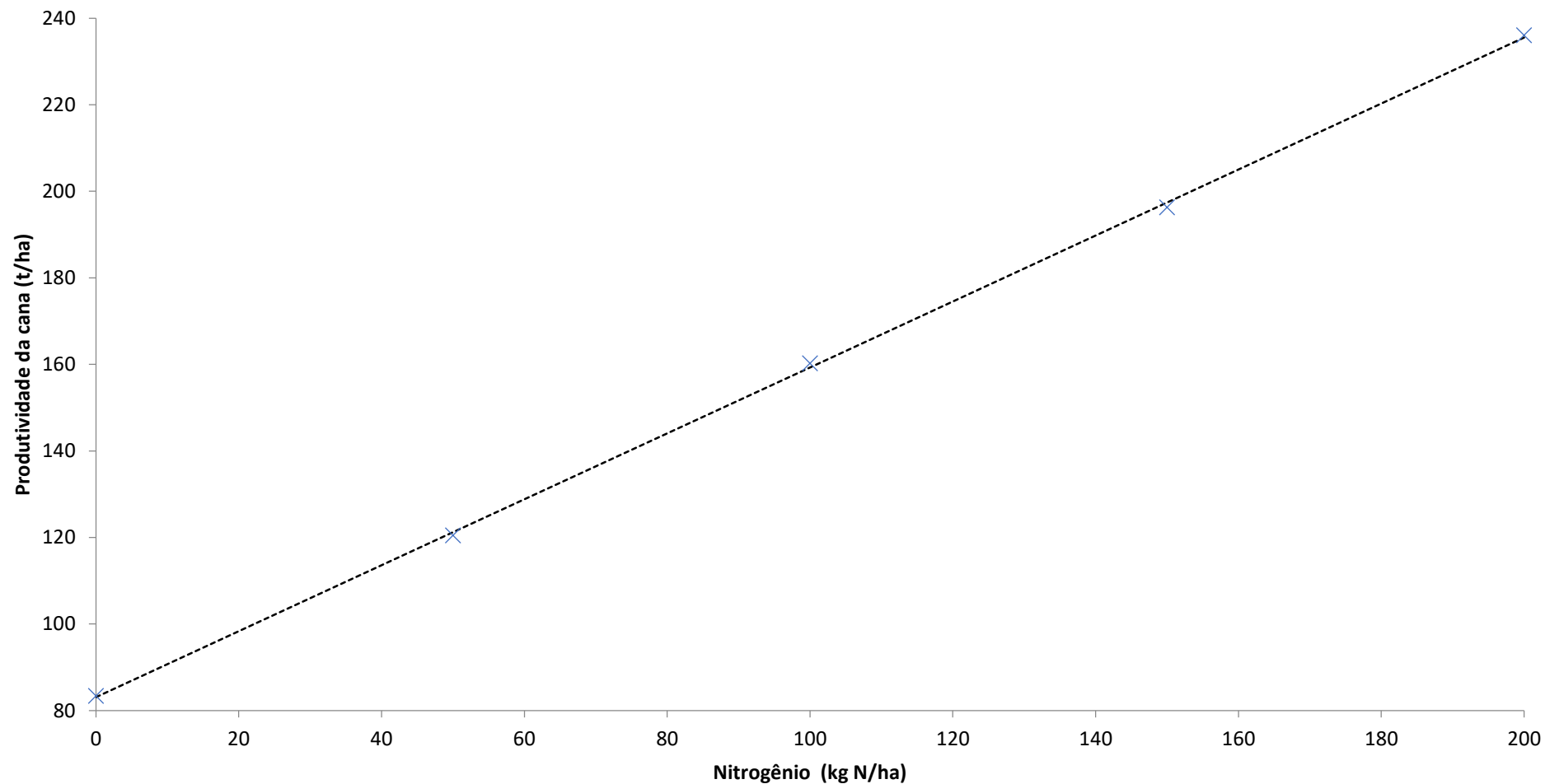


Figura 2 – Produtividade agrícola da cana de açúcar em Saturno, 2100, sob fertilização nitrogenada. A linha representa a estimativa de regressão e os x, as médias reais.

Interpretação de regressão linear simples - SAS

Model: MODEL1 - Dependent Variable: _800125888

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	58120	58120	53,45	<,0001
Error	18	19572	1087,32861		
Corrected Total	19	77692			

Root MSE		32,97467	R-Square	0,7481
Dependent Mean		159,31150	Adj R-Sq	0,7341
Coeff Var		20,69823		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr> t	Standardized Estimate
Intercept	1	83,07500	12,77103	6,50	<,0001	0
N	1	0,76237	0,10428	7,31	<,0001	0,86492

Parameter Estimates

Variable	DF	95% Confidence Limits	
Intercept	1	56,24405	109,90595
N	1	0,54329	0,98144

Trabalho de Análise de Dados

- Valerá um ponto da primeira VA
- Individual
 - Dados individuais especificados na planilha “Distribuição” do arquivo anexado
- Deverá ser entregue até o dia 28/11/2017
 - Às 23:59
- O trabalho constará dos seguintes arquivos
 - Arquivo de análise estatística. Saída do computador e entrada dos comandos, caso feito no computador, **OU** PDF com análise manual
 - Apresentação dos resultados em formas de figuras e ou tabelas (arquivo Word com figuras e tabelas incluídas)
 - Resposta às perguntas feitas, com sua justificativa (arquivo Word)
- Dados fictícios
 - No arquivo zip, com seus nomes

Descrição do cenário

- Experimento
 - Doses de N (0, 50, 100, 150, 200 kg N/ha)
 - Sete fontes estudadas separadamente
 - Sulfato de amônia
 - Uréia
 - Nitrato de amônio
 - MAP
 - DAP
 - Complexos
 - Esterco bovino
 - Esterco de aves
 - Produção de biomassa seca de capim elefante para produção de energia (t/ha)

Análise de dados

- Regressão linear separada para cada adubo
 - Com todos os valores individuais
 - Com as médias para cada N
- Apresentar os resultados em gráficos separados
 - Todos os pontos
 - Com base nas médias
- Selecionar o adubo com:
 - Melhor resposta
 - Pior resposta
- Comparar as regressões obtidas utilizando todos os valores e as médias

Trabalho de Interpretação de Dados

- Valerá um ponto da primeira VA
- Individual
- Deverá ser entregue até o dia 05/12/2017
- O trabalho constará dos seguintes arquivos
 - Apresentação dos resultados em formas de figuras e ou tabelas (arquivo Word com figuras e tabelas incluídas)
 - Resposta às perguntas feitas, com sua justificativa (arquivo Word)
- Dados fictícios

Descrição do cenário

- Experimento
 - Doses de N (0, 50, 100, 150, 200 kg N/ha)
 - Variedades hipotéticas de cana de açúcar destinadas à produção de energia
 - todas com códigos v algum número

Responder aos itens abaixo

- Selecionar a variedade com:
 - Melhor resposta ao nitrogênio
 - Pior resposta ao nitrogênio
 - Mais produtiva na ausência de adubação
 - Menos produtiva na ausência de adubação
 - Que deverá apresentar resultados mais próximos da previsão
- Estas variedades são significativamente diferentes entre si? Explique sua posição
- Para cada variedade selecionada acima calcular:
 - A produtividade esperada para uma adubação com 175 kg de N
 - A adubação necessária para atingir 150 t/ha