

[10:10, 17/11/2017] Aluna: Bom dia, professor! Gostaria de tirar uma dúvida sobre o trabalho de interpretação de regresso linear simples- SAS.

[10:11, 17/11/2017] Professor: Diga lá

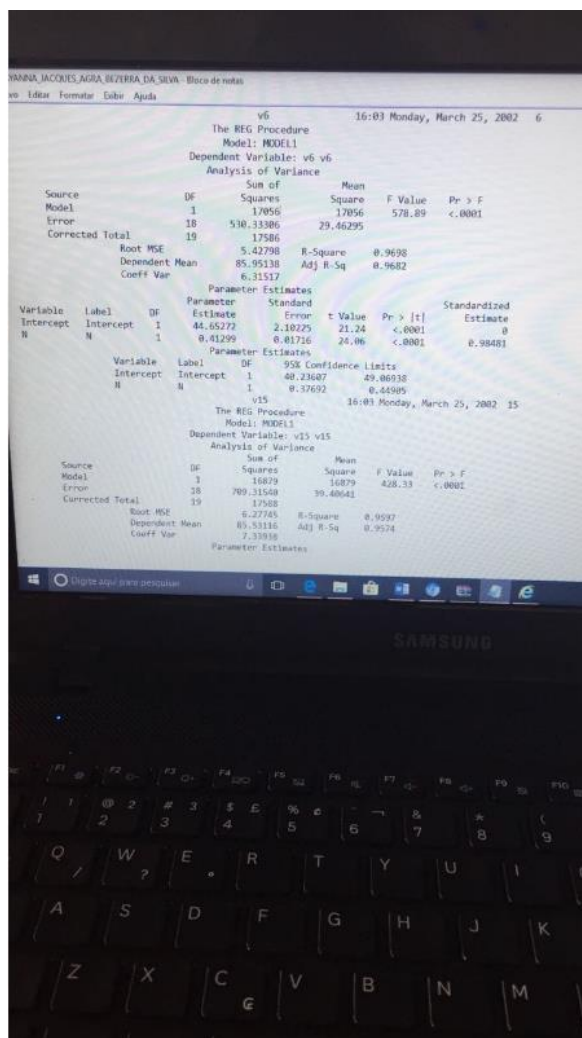
[10:13, 17/11/2017] Aluna: Para saber qual a dose que obteve melhor resposta a produtividade eu preciso identificar o valor de a na equação. Quanto o maior o a, maior a produtividade. Porém, não estou conseguindo identificar o ar no modelo do SAS.

[10:14, 17/11/2017] Aluna: Qual seria o meu valor de a nesse modelo?

[10:14, 17/11/2017] Professor: Pense um pouco em quem é a e b

[10:26, 17/11/2017] Aluna: Intercept = b

N = a



Levando em consideração que a e b são conhecidos e meu x é a dose, seriam esses os respectivos valores para as variáveis?

Variable	Label	DF	Parameter Estimate	Standard Error	t
Intercept	Intercept	1	44.65272	2.10225	
N	N	1	0.41299	0.01716	

[10:29, 17/11/2017] Aluna: Ou

Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	44.65272	2.10225	21.24	<.0001
N	1	0.41299	0.01716	24.06	<.0001

[10:37, 17/11/2017] Professor: Eu mando uma resposta mais detalhada logo que chegar na Rural, para poder usar um teclado mais decente?

[10:37, 17/11/2017] Professor: Ok?

[10:38, 17/11/2017] Aluna: Sem problemas, professor! Obrigada! 😊

Esta foi a conversa, mas agora vamos por partes para entender melhor o conjunto da obra.

O primeiro ponto é que me parece que está vendo uma certa confusão sobre a pergunta.

As perguntas que fiz neste trabalho foram:

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

E nenhuma destas corresponde bem à pergunta como foi expressa pela aluna. Vejam que ela está pensando na pergunta “Qual a dose que obteve melhor resposta a produtividade?”, ou seja uma dose específica (x kg N/ha) que levariam a uma maior produtividade da cana (y t/ha).

Como estamos trabalhando com uma regressão linear, ou seja, na forma de uma reta, este valor seria obrigatoriamente o maior valor possível no intervalo em que o experimento foi conduzido, já que a equação descreve uma reta (que esperamos neste caso ser crescente com a dose), e portanto não tem uma inflexão a partir da qual haveria redução na produtividade.

A pergunta real, eu acho, é à primeira pergunta, ou seja, em qual das regressões (teoricamente uma regressão para cada “variedade”) a variedade de cana tem a maior resposta ao N (ou seja, para qual variedade um acréscimo de 1 kg N/ha) levaria a um maior aumento na produtividade.

Considerando a boa e velha equação

$$y = a + bx$$

podemos observar que o a é completamente independente do valor de x. Assim, a não pode ser um indicador de resposta ao N, já que não tem ligação. Neste caso, temos que quanto maior (para $b > 0$)/menor ($b < 0$) o valor de b, maior a variação no valor de y.

Por sinal, isto admite que a pergunta que a aluna quer responder é a primeira (ou a segunda que é o inverso da primeira) pergunta do trabalho. Isto é reforçado pelo ponto que ela fala em intercept=b, N=a às 10:26, em que indica que está pensando na equação

$$y = ax + b$$

Em que a lógica é invertida em relação ao que costumamos usar em estatística. Este ponto fica bem claro pela discussão correta de quem é quem, fora a inversão que já comentei.

Agora partindo para um ponto mais pragmático, depois desta discussão mais geral, a dúvida que ela realmente parece estar é qual o conjunto de valores que ela deve examinar. Aí o ponto é bem mais simples

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	44.65272	2.10225	21.24	<.0001
N	N	1	0.41299	0.01716	24.06	<.0001

Variable	Label	DF	95% Confidence Limits
Intercept	Intercept	1	40.23607 - 49.06938
N	N	1	0.37692 - 0.44905

Olhando as duas telas que ela mandou (e que eu marquei em vermelho) fica claro que embora as duas falem em Parameter Estimates (ou Estimativas dos parâmetros em português), a da direita apresenta uma faixa de valores, enquanto na esquerda (que é a primeira das duas) apresenta um valor único (para cada parâmetro, lógico) que é o centro da estimativa apresentada na direita.

A da esquerda é realmente a estimativa encontrada para aquela amostra em particular, ou seja a que foi calculada pelo SAS como a que levaria ao melhor ajuste da equação. Na direita, temos os limites superiores e inferiores de uma faixa de valores em que esperamos encontrar os reais parâmetros da população (aquele que não temos, e nunca teremos já que a população real é desconhecida). Ou seja, tem menos de 5% de chance (neste caso) de que um a ou b real da população não esteja dentro do limite apresentado nesta análise.