



Predizendo uma variável por outra - Regressão Linear - Parte 3

# Aula	45
<input checked="" type="checkbox"/> Preparada	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> Revisada	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> Lecionada	<input checked="" type="checkbox"/>

▼ A Análise de Regressão Linear exige cuidados!

▼ Cuidado com as extrapolações!

Imagine estar modelando a fermentação de bactérias em assados a partir da temperatura de cozimento. Você mediu (e fez gráficos!) com temperaturas variando entre 50 e 150 graus Celsius. A associação é positiva, a relação é linear e forte, o R^2 ficou acima de 90%, então fica claro que quanto maior a temperatura, mais efetiva é a fermentação bacteriana dos assados.

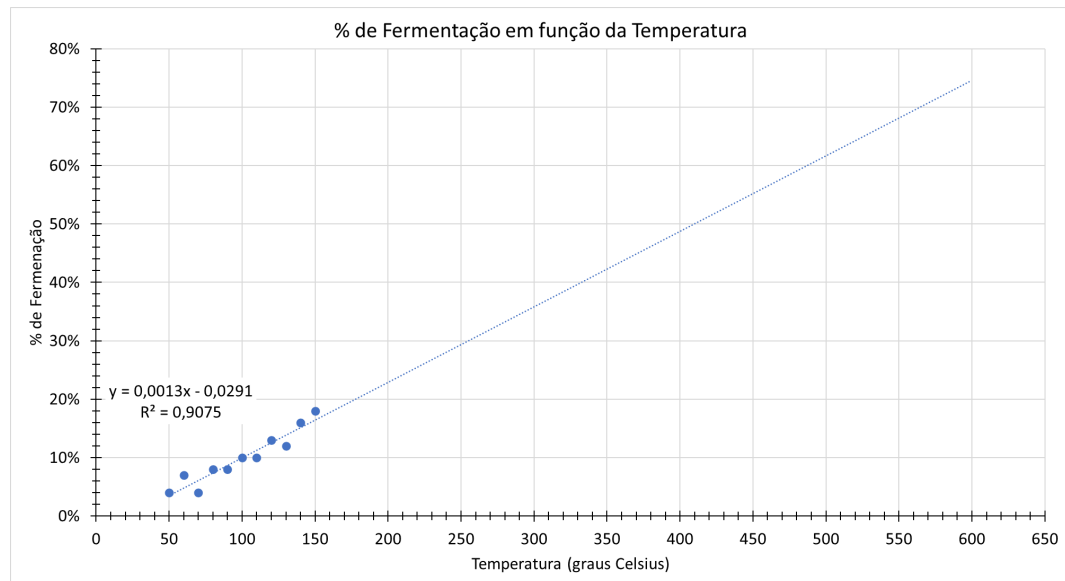
Alguém olha os seus gráficos e passa a sonhar com um forno industrial que atinge 600 graus Celsius, achando que vai ter um nível altíssimo de fermentação em assados.

A ideia faz sentido, mas é possível assegurar isso?

Não! Seus dados vão até 150 graus Celsius. Ainda que extrapolar “um pouco” seja aceitável, extrapolar neste nível é assumir que o padrão encontrado se mantenha

mesmo estando num patamar totalmente diferente de temperaturas.

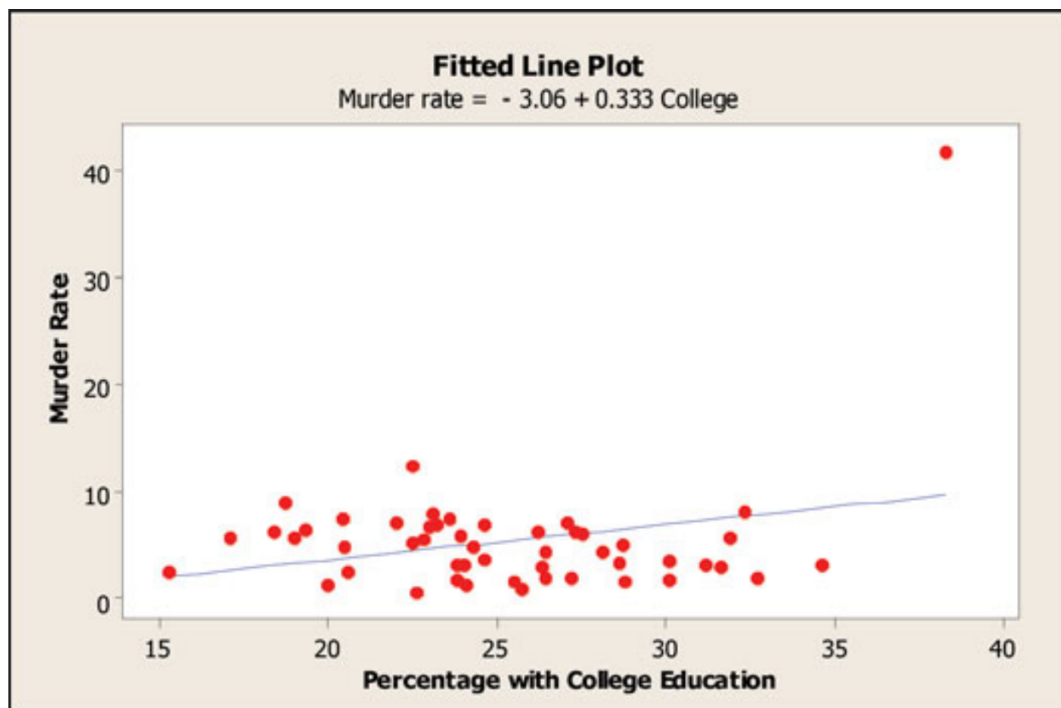
▼ Graficamente...



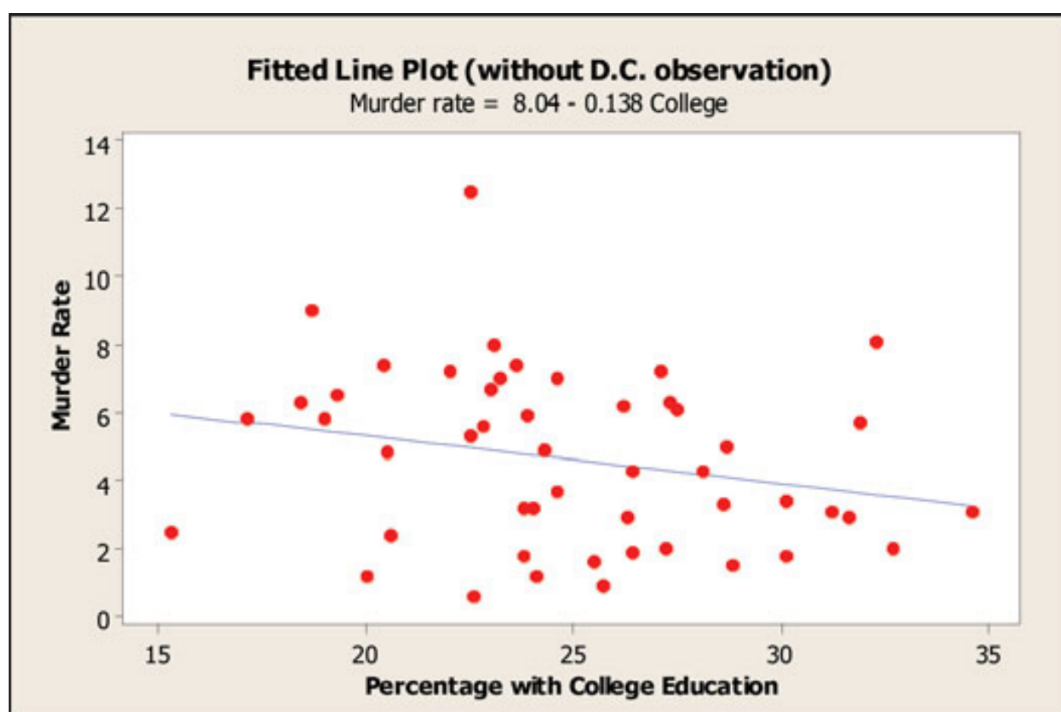
▼ Cuidado com outliers influentes!

Estudos da quantidade anual de assassinatos por grupo de 100.000 pessoas na população (aqui chamado de Murder Rate) em função do percentual da população com nível superior (aqui chamado de Percentage with College Education). Cada observação diz respeito a um estado norte-americano ou ao Distrito de Columbia (o equivalente norte-americano do nosso Distrito Federal).

▼ É de se esperar que quanto maior o nível de educação, menores os níveis de crimes violentos...



- A manutenção do Distrito de Columbia (o ponto afastado do padrão) na análise de regressão tem o efeito de puxar a inclinação da reta de regressão para cima. Então a linha de regressão sugere, no primeiro caso, que os crimes violentos aumentam com a maior escolaridade. Fato é que todo o restante dos dados sugere que os crimes violentos diminuem com o aumento da escolaridade.



Outliers comprometem o ajuste do modelo.

Mais um motivo para não nos esquecermos de fazer gráficos, no mínimo para inspecionar os dados e identificar os outliers!

▼ Cuidado com a diferença entre correlação e causalidade!

Em análise de regressão, suponha que à medida em que x cresce, y também tenda a crescer (ou a diminuir). Será que podemos concluir que existe uma conexão causal, em que mudanças em x causam mudanças em y ?

O conceito de causalidade é central para a ciência. Nós estamos familiarizados com ele, pelo menos num sentido informal. Sabemos, por exemplo, que a exposição a um vírus pode causar o contrair uma doença. Mas a mera percepção da associação entre duas variáveis não é suficiente para implicar numa conexão causal. Pode haver alguma explicação alternativa para a associação observada.

Sempre que duas variáveis estejam associadas, outras variáveis podem ter influenciado esta associação. Por isso dizemos mais uma vez: **Correlação não implica em causalidade!**

Sempre que observamos a correlação entre as variáveis x e y , talvez existam mais variáveis correlacionadas tanto com x quanto com y que sejam responsáveis pela associação deles. Estas variáveis são chamadas de variáveis ocultas.

▼ Um exemplo para “afogar” quaisquer dúvidas sobre este ponto:

- Em um determinado litoral se percebeu, em observações diárias, a existência de uma correlação elevada entre mortes por afogamento no mar e o quantidade de sorvete vendido na praia.

- Daria para dizer que a maior quantidade de sorvete vendido na praia causa mais afogamentos? De maneira nenhuma!
- Acontece que a quantidade de sorvete vendido na praia revela os dias mais quentes, em que a praia acaba sendo bem mais frequentada. Naturalmente, nestes dias existem mais afogamentos...
- A temperatura é a variável oculta que explica a elevada correlação entre a quantidade de sorvete vendido na praia e a quantidade de afogamentos no mar.

De maneira mais genérica, podemos dizer que **associação não implica em causalidade!**

Por isso, cuidado com as primeiras impressões quando for interpretar uma associação!