

**#013, 1.1 Regressão Linear Múltipla**

Uma regressão linear múltipla trabalha com modelos mais robustos, uma vez que incorpora duas ou mais variáveis independentes para predizer o comportamento de uma variável dependente.

No mercado financeiro, é utilizada para explicar a relação entre duas variáveis, testar teorias ou fazer previsões. Essa relação precisa ser comprovada posteriormente, por meio de um teste de hipótese podemos dizer se existe alguma relação, para determinado intervalo de confiança.

De um modelo pode ser estimado usando um número de observações, dada a equação:

* é a variável dependente,
* um número de variáveis independentes
* o intercepto é medido quando todas as variáveis são iguais a 0.
* Os coeficientes parciais medidos pela taxa de variação na variável dependente () ao variarmos 1 unidade em cada variável independente () e manter todas as demais variáveis constantes. Pode ser dado em um modelo discretizado [e.g. ] ou como equação diferencial se tomarmos grandezas infinitesimais, quantidades muito pequenas que variam muito pouquinho [e.g. ].

Pode parecer abstrato, mas um exemplo clássico é a elasticidade de um produto onde a quantidade demandada depende do preço (), renda disponível () e preço de um bem substituto ()

Para validar o modelo precisamos garantir 5 condições:

1. **Linearidade:** existe relação linear entre a variável X e Y. (i.e. variam de forma proporcional)
2. **Homoscedasticidade:** a variância dos erros é constante em todas as observações (é comum em modelos mal ajustado o erro aumentar(diminuir) à medida que variamos a variável independente.
3. **Independência dos Erros:** o erro de uma observação não está correlacionado ao erro de outra observação (são independentes).
4. **Normalidade:** os erros devem ser normalmente distribuídos. (impossibilitaria o uso de testes estatísticos)
5. **Ausência de Multicolinearidade:** as variáveis independentes não devem ser aleatórias, ou existir correlação linear entre elas. (dificulta determinar o efeito de cada variável na amostra)

Gráficos de dispersão podem ajudar na avaliação de linearidade, tornam simples identificar se é necessário algum ajuste para modelos lin-log ou log-log. Plotar os resíduos, pode ajudar a identificar a homoscedasticidade ou independência de erros.

Ref. Quantitative Methods. [s. L.]: Cfa Institute, 2025.

**1.2 Avaliação e Gráficos de verificação.**

Uma matriz de dispersões nos ajuda a avaliar pares de variáveis e suas relações. É esperado uma reta o mais próximo de zero nos gráficos de dispersão. Uma inclinação muito positiva(negativa) pode indicar uma relação linear entre as variáveis.

Uma relação linear pode ser causada pela ocorrência de valores extremos (*outliers*). Podemos verificar, de forma mais geral, em uma nuvem de dispersão entre os valores prescritos e os resíduos (erro da regressão). Ou o resíduo vs. cada variável independente para avaliar mais especificamente os outliers de cada parâmetro.

Um gráfico quantil-quantil (Q-Q plot), é utilizado para comparar uma distribuição normal teórica vs. os resíduos da regressão normalizados. Assim, é possível avaliar a normalidade: se os erros são normalmente distribuídos. A homoscedasticidade é verificada também, observando os pontos que extrapolam a distribuição teórica, geralmente utilizamos 2 desvios como ponto de corte, ou seja, um intervalo de confiança de 95%. De forma geral, em uma distribuição normal 68% das variáveis prescritas estão dentro de 1 desvio padrão, 95% para 2 desvios e 99.7% para 3 desvios.

BULLET POINTS

* Regressão linear mútipla é utilizada para relacionar duas ou mais variáveis independentes (x, parametros) para descrever a variação de uma variável dependente (Y).
* Grafico Quantil-Quantil (Q-Q plot) é utilizado para comparar a distribuição dos resíduos normalizados com uma distribuição normal teórica.
* Coeficientes Parciais de Regressão medem a taxa de variação de cada parâmetro mantendo todos os demais constantes. O cálculo de uma derivada parcial, podemos desconsiderar o intercepto, assim como os demais parametros. (derivada de constante é nula)
* Uma matriz de dispersão mostra a relação entre diferentes parametros nos triangulos, tal como a distribuição de cada parametro na diagonal principal. Possibilita avaliar multicolinearidade.