

# Respostas: Questões Teóricas

1. **Explique a diferença entre erro padrão, intervalo de confiança e erro de predição.**

**Erro Padrão (EP):** Pense nele como a medida de incerteza de uma *estimativa*. Se nós tirássemos várias amostras da mesma população e calculássemos um coeficiente  $\beta$  (de uma regressão) ou uma média para cada amostra, o erro padrão nos diz o quanto essas estimativas variariam. Ele quantifica a **nossa incerteza sobre o parâmetro** que estimamos.

**Intervalo de Confiança (IC):** É um *intervalo* de valores. Ele usa o erro padrão para construir uma faixa onde temos um certo nível de confiança (ex: 95%) de que o **verdadeiro parâmetro da população** está contido. É uma faixa de incerteza para o *parâmetro* (ex: a média populacional  $\mu$  ou o  $\beta$  verdadeiro).

**Erro de Predição:** Refere-se à incerteza ao prever uma **única observação futura**. Um *intervalo de predição* é sempre mais largo que um intervalo de confiança. A justificativa é que ele inclui duas fontes de incerteza: 1) a incerteza sobre onde está o parâmetro (capturada pelo IC) e 2) a variabilidade aleatória natural dos dados (o erro  $\epsilon$  que o modelo nunca consegue prever).

2. **Em um modelo de regressão linear, o que significa multicolinearidade e como detectá-la?**

**O que é:** A multicolinearidade acontece quando duas ou mais variáveis preditoras (as variáveis  $X$ ) no modelo estão fortemente correlacionadas entre si. Na prática, elas carregam informação redundante.

**Justificativa (Problema):** Isso inflaciona a variância dos coeficientes  $\beta$  estimados. O resultado é que os erros padrão desses coeficientes ficam muito grandes. Isso torna as estimativas individuais instáveis e difíceis de interpretar; o modelo pode ter um  $R^2$  alto (explicando bem os dados), mas nenhum coeficiente  $\beta$  individual é estatisticamente significativo (p valores altos).

**Detecção:** A forma mais comum é calcular o **Fator de Inflação de Variância (VIF)** para cada variável. O VIF mede o quanto a variância de um coeficiente é inflada pela presença de outras preditoras. Uma regra prática comum é que um VIF acima de 5 ou 10 indica um nível problemático de multicolinearidade.

3. **Diferencie overfitting e underfitting; cite uma forma de mitigar cada um.**

**Underfitting (Subajuste):** Ocorre quando o modelo é **simples demais** e não

consegue capturar a tendência principal dos dados. Ele falha em aprender o sinal. O modelo terá um desempenho ruim tanto nos dados de *treino* quanto nos de *teste*. Dizemos que ele tem alto viés (bias).

- **Mitigação:** Aumentar a complexidade. Isso pode ser feito usando um modelo mais potente (ex: trocar regressão linear por uma árvore de decisão ou adicionar termos polinomiais), ou adicionando mais variáveis (features) relevantes.

**Overfitting (Sobreajuste):** Ocorre quando o modelo é **complexo demais**. Ele começa a "decorar" o ruído aleatório presente nos dados de treino, ao invés de aprender o sinal geral. O modelo terá um desempenho excelente no treino, mas muito ruim no teste (ele não consegue generalizar). Dizemos que ele tem alta variância.

- **Mitigação:** Reduzir a complexidade. A técnica mais comum é a **regularização** (como L1 Lasso ou L2 Ridge), que penaliza coeficientes grandes, forçando o modelo a ser mais simples e focar apenas nas variáveis mais importantes. Outra forma é usar mais dados de treino.

#### 4. Explique o conceito de estacionariedade em séries temporais e como testá-la.

**Conceito:** Uma série temporal é considerada (fracamente) estacionária se suas propriedades estatísticas básicas não mudam ao longo do tempo. Especificamente, sua **média é constante** e sua **variância é constante**. Além disso, a autocovariância (a correlação dela com ela mesma no passado) depende apenas da defasagem  $k$ , e não do ponto  $t$  no tempo. Uma série estacionária "flutua" em torno da mesma média, sem tendências ou sazonalidades fortes.

**Teste:** A estacionariedade é fundamental porque a maioria dos modelos (como ARMA) assume isso. Para testar, usamos **testes de raiz unitária**. O mais popular é o **Teste Dickey Fuller Aumentado (ADF)**.

- A hipótese nula ( $H_0$ ) do teste ADF é que a série **possui uma raiz unitária**, o que significa que ela não é estacionária (ex: é um passeio aleatório).
- A hipótese alternativa ( $H_1$ ) é que a série **é estacionária**.
- **Justificativa:** Se o p valor do teste for baixo (geralmente menor que 0.05), nós rejeitamos  $H_0$  e podemos tratar a série como estacionária.

#### 5. Por que o uso de validação cruzada temporal é necessário em séries?

A validação cruzada tradicional (como k folds) embaralha os dados aleatoriamente para criar os conjuntos de treino e teste. Em séries temporais, isso é um erro grave porque **a ordem cronológica dos dados importa**.

**Justificativa:** O valor de amanhã depende do valor de hoje. Se embaralharmos os dados, o modelo pode acabar sendo treinado com dados do "futuro" (ex: ano 2020) para prever o "passado" (ex: ano 2019). Isso é um vazamento de dados (data leakage) que não acontece na vida real. O modelo parecerá ter uma performance ótima nos testes, mas falhará catastroficamente na prática.

Por isso, usamos validação temporal (como *forward chaining* ou janelas deslizantes). O modelo é **sempre treinado no passado para prever o futuro**, respeitando a ordem dos fatos. Por exemplo: treina nos dados [1...100] e testa em [101...110]; depois, treina em [1...110] e testa em [111...120].

## 6. O que é autocorrelação dos resíduos e como pode ser testada?

**Conceito:** É a correlação dos erros (resíduos) do modelo ao longo do tempo. Se existe autocorrelação, significa que o erro no tempo  $t$  está sistematicamente relacionado com o erro no tempo  $t - 1$  (ou  $t - 2$ , etc.).

**Justificativa (Problema):** Isso é um problema sério, especialmente em modelos de regressão OLS (Mínimos Quadrados Ordinários). Uma das premissas chave do OLS é que os erros são independentes. Se eles são autocorrelacionados, significa que o modelo **falhou em capturar toda a estrutura temporal** dos dados; sobrou informação "previsível" nos resíduos.

Embora os coeficientes  $\beta$  possam continuar corretos (não enviesados), os **erros padrão ficam incorretos** (geralmente subestimados). Isso invalida todos os testes de hipótese (p valores) e intervalos de confiança, levando a conclusões erradas sobre a significância das variáveis.

**Teste:** O teste clássico é o de **Durbin Watson (DW)**. Um teste mais geral e consistente (que também testa correlações de ordem superior) é o **Teste de Breusch Godfrey** (um teste LM). Visualmente, podemos inspecionar o gráfico da **Função de Autocorrelação (ACF)** dos resíduos; se houver barras significativas fora do intervalo de confiança, há evidência de autocorrelação.