

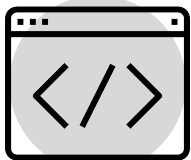
Profissão: Cientista de Dados



BOAS PRÁTICAS



Regressão Logística I



- **Conheça a regressão logística**
- **Bivariada**
- **Aplique regressão logística**
- **Use mais de uma variável**
- **Qualidade do modelo**



Conheça a regressão logística

- **Compreenda a natureza da sua variável resposta:**

A regressão logística é adequada para variáveis binárias, ou seja, variáveis que podem assumir apenas dois valores. Portanto, antes de escolher o modelo, é importante entender a natureza da sua variável resposta.

- **Use a regressão logística para analisar a ocorrência ou não de um evento:**

A regressão logística é útil para analisar a ocorrência ou não de um evento, como no exemplo da aula, onde foi usada para analisar os óbitos de pacientes com COVID-19.



Bivariada

- Sempre comece com uma análise descritiva dos dados:**

Antes de mergulhar em modelos complexos, é importante entender os dados que você está trabalhando. Uma análise descritiva pode revelar padrões e tendências que podem ser úteis mais tarde.

- Entenda a diferença entre probabilidade e chance:**

Embora esses termos sejam frequentemente usados de forma intercambiável, eles têm significados distintos em estatística. A probabilidade é a razão de eventos favoráveis para o total de eventos possíveis, enquanto a chance é a razão de eventos favoráveis para eventos não favoráveis.



Bivariada

● Use a razão de chances para comparar grupos:

A razão de chances é uma maneira útil de comparar a chance de um evento ocorrer em diferentes grupos. Ela pode ser usada para quantificar a força de uma associação entre uma variável explicativa e uma variável de resposta.

● Aprenda a interpretar a razão de chances:

A razão de chances pode ser um conceito difícil de entender, mas é uma parte crucial da regressão logística. Uma razão de chances maior que 1 indica que o evento é mais provável no primeiro grupo, enquanto uma razão de chances menor que 1 indica que o evento é mais provável no segundo grupo.



Aplique Regressão Logística



- Ao lidar com variáveis categóricas, como sexo ou classe de viagem, pode ser útil criar variáveis dummy. Isso permite que o modelo capture o efeito de cada categoria individualmente.
- A razão de chances é uma métrica útil para comparar a probabilidade de um evento ocorrer em um grupo em relação à probabilidade de ocorrer em outro grupo. A interpretação da razão de chances é uma das principais vantagens da regressão logística.
- Ao interpretar os parâmetros do modelo, lembre-se de que um parâmetro negativo para uma variável dummy indica que a categoria representada por essa variável tem menos probabilidade de experimentar o evento de interesse do que a categoria de referência.
- Sempre verifique a relevância e a importância das variáveis utilizadas no modelo. Variáveis irrelevantes podem distorcer os resultados e a interpretação do modelo.

Use mais de uma variável

- Sempre verifique a significância das variáveis em seu modelo. Se uma variável não é estatisticamente significativa na presença de outras, considere removê-la para simplificar o modelo.
- Ao adicionar uma nova variável ao modelo, observe como isso afeta a significância das outras variáveis. Se a adição de uma nova variável torna outra insignificante, você pode precisar escolher entre elas.
- Ao interpretar os resultados do modelo, considere o contexto. Por exemplo, neste caso, a chance de sobrevivência aumenta à medida que o valor do bilhete aumenta. Isso faz sentido no contexto do Titanic, onde passageiros que pagaram mais por seus bilhetes provavelmente tinham acesso a acomodações mais seguras.



Qualidade do modelo

- Avalie a capacidade de discriminação do seu modelo: Seu modelo deve ser capaz de distinguir corretamente entre diferentes resultados. Se não for, você pode precisar ajustar seu modelo ou considerar um modelo diferente.
- Use a curva ROC para avaliar a capacidade de discriminação do seu modelo: A curva ROC é uma ferramenta gráfica útil que pode ajudá-lo a entender como seu modelo está se saindo em termos de acertos e erros para diferentes pontos de corte.
- Use a estatística KS para comparar distribuições de probabilidade: A estatística KS pode ajudá-lo a entender a distância entre as curvas de distribuição acumulada de dois grupos, o que pode ser útil para avaliar a discriminação do seu modelo.
- Considere o índice de Gini ao avaliar a discriminação do modelo: O índice de Gini é uma medida de desigualdade que pode ser útil para entender como seu modelo está discriminando entre diferentes resultados.



Bons estudos!

