Atividade Calculo Computacional

Prof: Eduardo

Aluno: Gabriel Boos Duarte

**Atividade - Modelagem Linear**

1. Plotando gráficos das variáveis da minha análise, para ver qual gráfico desejo fazer um modelo que possa o representar.

Texto

Descrição gerada automaticamente

Interface gráfica do usuário, Diagrama

Descrição gerada automaticamente

1. Após plotar os gráficos, pode-se ver a dispersão dos mesmos, por onde todos são cabíveis a uma modelagem linear para explicá-los, entretanto foi escolhido as variáveis Open e High para fazer uma análise detalhada.



Gráfico, Gráfico de dispersão

Descrição gerada automaticamente

Vendo o gráfico de dispersão das variáveis independentes, o gráfico se aproxima de um modelo linear, por onde em grande parte das vezes, quanto maior o Open, maior o Close, e vice-versa.

1. Vendo a dispersão do Open e High, foi feito um modelo linear em cima dessas variáveis e um summary, para ver os dados descritivos obtidos do mesmo.



Gráfico, Gráfico de dispersão

Descrição gerada automaticamente



Texto, Carta

Descrição gerada automaticamente

Análise estatística:

Ambos os coeficientes relacionados possuem p-values muito pequenos, sugerindo que há uma relação estatisticamente significativa entre base.r.wege.full$High e base.r.wege.full$Open. O p-value ideal para contrapor a hipótese nula é p < -0.05, e como o coeficiente A, possui **Pr(>|t|) < 2e-16,** isso comprova o modelo.

R-squared: 0.9885, demonstrando um modelo linear com poucos erros em relação a equação linear, pois mostra que **97.7%** da variação em Open pode ser explicada pelos preditores do modelo.

A distribuição dos resíduos (erros) parece razoavelmente simétrica, com um residual padrão médio (Residual standard error) de **1.377**.

Em resumo, com base nos resultados, parece que o modelo é estatisticamente significativo e explica uma grande parte da variabilidade em base.r.wege.full$Open. A variável base.r.wege.full$High é estatisticamente significativa e está positivamente associada a base.r.wege.full$Open.

1. Outra medida importante na verificação da modelagem, é plotar esses 4 gráficos Residuals vs Fitted, Normal quantil-quanti, Scale-location e Distâncias de Cook.



Diagrama

Descrição gerada automaticamente

Por meio desses gráficos podemos observar o seguinte. No primeiro gráfico de Residuals vs Fitted temos os resíduos em função dos valores

estimados, e podemos utilizar este gráfico para observar a independência e a homocedasticidade, se os resíduos se distribuem de maneira razoavelmente aleatória e com mesma amplitude em torno do zero. A partir desse gráfico, vê-se que a linha do modelo se aproxima muito de uma amplitude perto do 0, o que significa uma validação da **homoscedasticidade**, e segue a mesma lógica para o gráfico 3 (Scale-location).

A homoscedasticidade é uma das suposições importantes na análise de regressão linear. Se essa suposição não for atendida, os intervalos de confiança e os testes de significância podem ser viesados, levando a conclusões incorretas sobre a relação entre as variáveis.

No gráfico 2 (Q-Q Residuals) , podemos avaliar a normalidade dos resíduos. A linha diagonal pontilhada representa a distribuição normal teórica, e os pontos a distribuição dos resíduos observados. Neste gráfico podemos ver que os resíduos estão perto da linha definida, por isso tem-se um bom modelo, pois espera-se que não exista grande fuga dos pontos em relação à reta teórica.

No gráfico 4(Distâncias de Cook) é uma medida de influência quando pode indicar a presença de outliers que possuem valor maior do que 1, no gráfico é possível ver alguns outliers.

1. Buscar adicionar outra variável que possa melhorar o modelo, dando o mais significância. Para fazer isso pode-se usar o pairs() para analisar a nova variável e ver sua dispersão. Adicionei a variável Low.



Gráfico, Gráfico de dispersão

Descrição gerada automaticamente

• Open e High têm uma forte correlação linear positiva

• Open e Low, ou High e Low têm uma correlação linear positiva modesta, o que pode nos ajudar na amostra

1. Fazer novo modelo adicionando variável explicativa (Low)





Gráfico, Gráfico de dispersão

Descrição gerada automaticamente

1. Buscando dados descritivos que comprovem a significância do meu novo modelo ao adicionar a variável explicativa



Uma imagem contendo Carta

Descrição gerada automaticamente

•Ao fazer o novo summary, observa-se o novo R-squared, que passou de 0.9885 do modelo antigo, para **0.9938** no modelo novo, explicando **98.7%** da variação em Open com base nas predição do modelo.

•A distribuição dos resíduos (RSE) houve uma significativa diminuição, passando de 1.377 para **1.01,** demonstrando uma melhora de correlação.



Gráfico, Histograma

Descrição gerada automaticamente

•O histograma do modelo novo demonstra-se parecido com um histograma de uma curva normal, que é o ideal para histograma de resíduos.

**Conclusão:**

Por meio de toda a análise feita, foi possível ver quais são os fatores que podem contribuir a entender a modelagem, e se ela é significativa ou não. Além disso foi visto ferramentas que podem melhorar a modelagem, como a adição de variáveis explicativas, aumentando os atributos de uma boa correlação entre a linha preditiva e a dispersão dos dados.