Nome: Lucas Silva Neris

Estudo de Caso papelão ondulado - Projetos industriais

**1.Introdução**

**1.1. Livro-texto usado como base**

O livro utilizado como base para a realização das análises foi o livro 'Econometria Básica - Gujarati e Porter'.

---

**1.2. Premissas importantes ao se trabalhar com regressões**

Antes de começar a análise dos dados, precisamos relembrar alguns pontos ao tratarmos de regressões. Entre elas:

Não pode haver autocorrelação entre uma variável, ou seja, o valor de uma variável t não pode depender de sua antecessora.

Os erros quadráticos devem ser aleatórios.

Os regressores não podem estar correlacionados(não pode haver alta multicolinearidade)

**2; Busca por um melhor regressor para explicar a variável “consumo de papelão”**

**2.1. Identificando correlações entre os dados apresentados.**

Primeiramente vamos ver se há uma correlação entre os candidatos à variáveis explicativas. Para isso faremos uma plotagem dos dados a fim de ver se há correlação entre os dados(*não pode haver uma forte multicolinearidade numa regressão múltipla, ou seja, a correlação entre as variáveis explicativas deve ser baixa).*

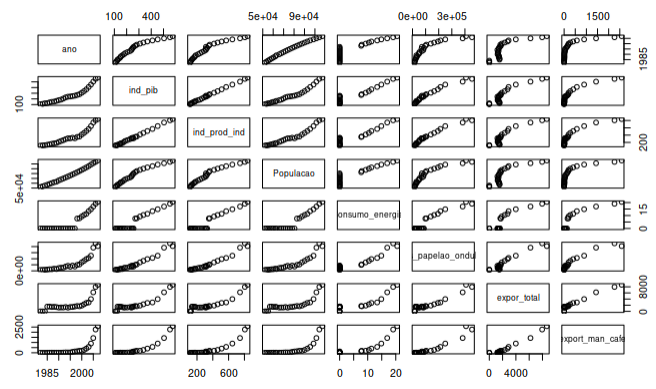
***2.1.1. Código utilizado e os resultados obtidos***

*“*

*dados <- read.csv('dados\_problema.csv')*

*plot(dados)*

*“*



**2.1.2. Conclusões obtidas a partir da plotagem**

Como podemos ver, existe grande correlação entre as variáveis: *ind\_pib, ind\_prod\_ind e ind\_papelao\_ondulado*. Outros pontos que também devem ser levados em consideração, são:

* Os dados sobre o consumo de energia industrial estão com uma amostragem baixa
* Consumo de energia industrial e o índice de produção industrial são variáveis correlacionadas, uma vez que uma maior produção industrial requer um maior consumo de energia.
* Exportação total está relacionado com o índice do PIB também. Assim, também torna-se uma variável altamente correlacionada.
* Portanto, para o cálculo do consumo de papelão ondulado, vamos utilizar a variável ind\_pib e a variável tempo.

**2.2. Escolhendo o melhor tipo de regressão para prever a demanda de papelão**

**2.2.1. Arrumando os dados das variáveis “ano” e “consumo de papelão”.**

Foi necessário a mudança da unidade de medida, da variável papelão, de Toneladas para “1000 toneladas”. Isso tornou-se necessário pois, caso contrário, a plotagem a fim de visualizar como irá comportar os dados de papelão, ficaria prejudicada.

**2.2.2. Código utilizado para arrumar os valores**

#vamos representar os dados de consumo de papelão como 10³toneladas

dados$cons\_papelao\_ondulado = (dados$cons\_papelao\_ondulado/1000)

#Vamos arrumar também a variável ano

dados$ano = (dados$ano - dados$ano[1])

**2.2.3. testes para encontrar a melhor regressão linear**

Vamos agora fazer as regressões, uma simples(explicação do consumo de papelão ondulado com base no índice do PIB) e uma múltipla(explicação do consumo de papelão com base no índice do PIB e do tempo).

**2.2.3.1. Testando a regressão linear “consumo papelão(x) = A + índice do pib\*B + ano\*C” e “consumo papelão(x) = índice do pib\*A + B” com intercepto**

Código utilizado

*fit<- lm(dados$cons\_papelao\_ondulado ~ dados$ind\_pib + dados$ano)*

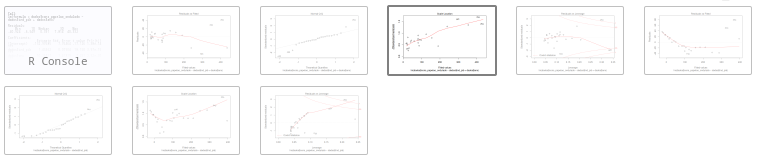
*fit2 <- lm(dados$cons\_papelao\_ondulado ~dados$ind\_pib)*

*summary(fit)*

*summary(fit2)*

*plot(fit)*

*plot(fit2)*



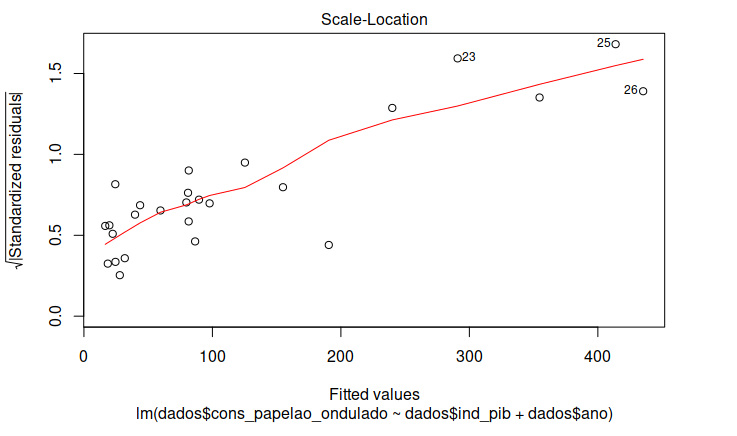


Imagem do resultado obtido a partir das duas regressões. Note que existem outras imagens, como por exemplo o relatório com os valores das regressões.

Note que em ambos os casos o intercepto é negativo, o que não faz sentido. Assim, vamos testar as regressões lineares sem intercepto.

**2.2.3.2. Testando a regressão linear “consumo papelão(x) = índice do pib\*A + ano\*B” e “consumo papelão(x) = índice do PIB\*A” sem intercepto.**

Código utilizado para realizar as regressões

fit3 <- lm(dados$cons\_papelao\_ondulado ~dados$ind\_pib + dados$ano -1)

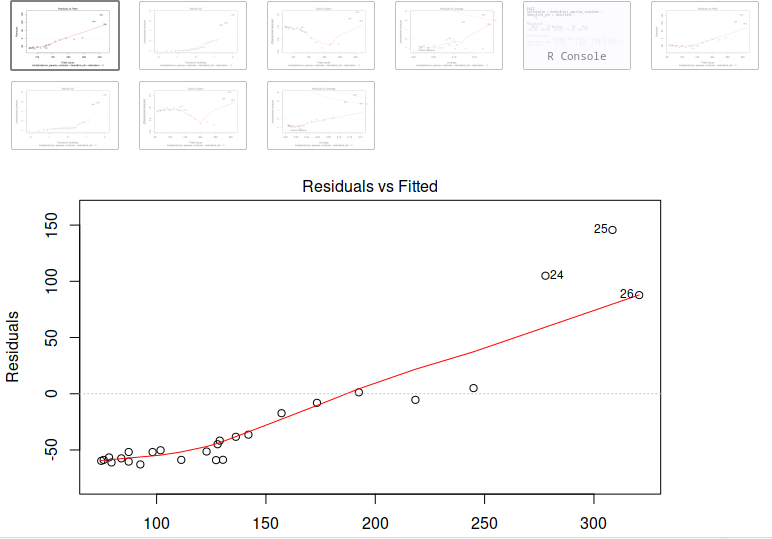
plot(fit3)

summary(fit3)

fit4 <- lm(dados$cons\_papelao\_ondulado ~dados$ind\_pib -1)

summary(fit4)

plot(fit4)



Note que os resíduos não são aleatórios*(*fere o princípio da hetorocedasticidade), *uma vez que é possível realizar uma regressão entre os mesmos.*

*Assim sendo, vamos testar uma regressão polinomial.*

**2.2.3.3. Testando a regressão polinomial linear “consumo papelão(x) = índice do PIB²\*A” sem intercepto.**

Vamos testar a regressão polinomial também, uma vez que plotando o “índice do PIB” x consumo de papelão, o formato é simelhante ao de uma parábola.

Código utilizado:

fit5 <- lm(dados$cons\_papelao\_ondulado ~ I(dados$ind\_pib^2)\*x -1)

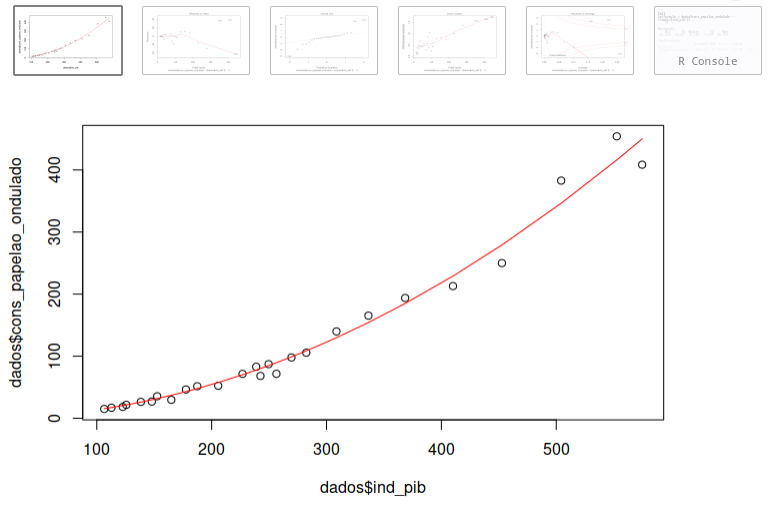
plot(dados$ind\_pib,dados$cons\_papelao\_ondulado)

points(dados$ind\_pib, predict(fit5),type = 'l', col='red')

plot(fit5)

summary(fit5)

Imagem obtida:



Plotagem dos dados (“índice PIB” x “consumo de papelão”). A linha vermelha indica o comportamento dos dados da nossa regressão.

Como podemos perceber, o nosso R² possui um valor altíssimo. Assim sendo, o índice do pib explica muito bem o consumo de papelão, o que já era esperado uma vez que o papelão é utilizado tanto no transporte quanto na armazenagem de produtos

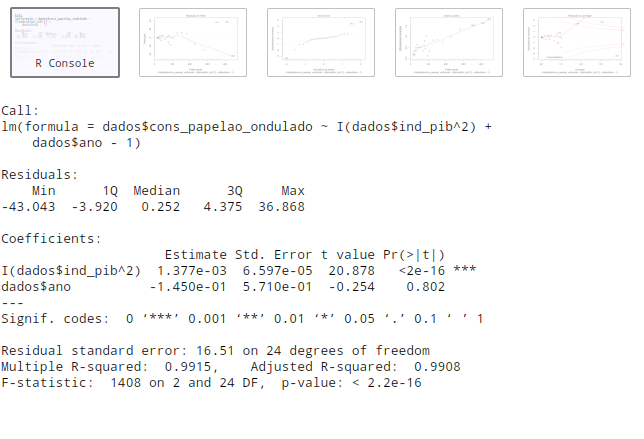
**2.2.3.4. Testando a regressão polinomial linear “consumo papelão(x) = A\*índice PIB² + B\*Índice do pib” sem intercepto.**

Código utilizado:

fit6 <- lm(dados$cons\_papelao\_ondulado ~I(dados$ind\_pib^2) + dados$ano -1)

summary(fit6)

plot(fit6)



Logo, selecionaremos o fit5 pois possuí um F-teste alto e seus regressões são estatisticamente significativos.

Logo, **nosso B1 será 1.362e-03** (Obtido no item 2.2.3.3.)

**3 Projeção do PIB para os próximos 5 anos**

Vamos agora calcular o PIB para os próximos 5 anos. Para isso vamos utilizar a forma quantitativa, vamos prever a evolução do pib ao passar dos anos.

Vamos tentar dois modelos de regressão, o linear e o polinomial. A partir da plotagem do índice do pib, os valores aparentam seguir a forma de um polinômio do segundo grau.

**3.1. Testando a regressão “índice do pib = A\*ano + B\*Ano²), com intercepto**

Código Utilizado:

#plotanto o Ano x Índice PIB

plot(dados$ano,dados$ind\_pib)

#note que há uma tendência polinomial. Assim, vamos tentar regressões polinomiais com intercepto e sem intercepto

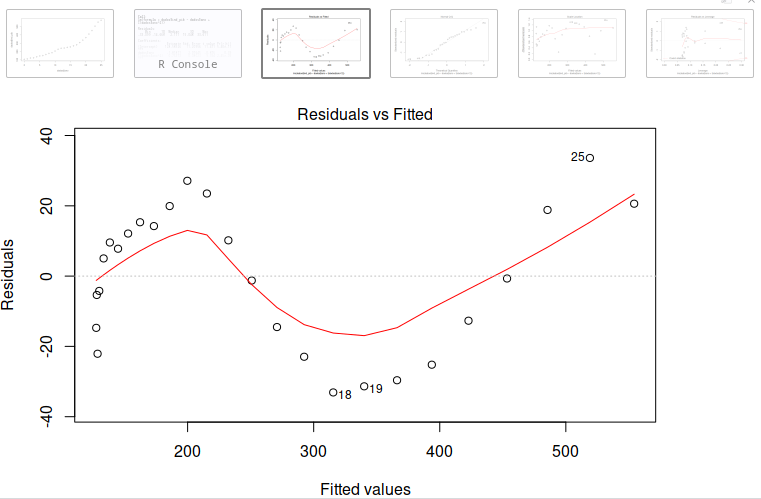
#Com intercepto

fit8 <- lm(dados$ind\_pib ~ dados$ano + I(dados$ano^2))

summary(fit8)

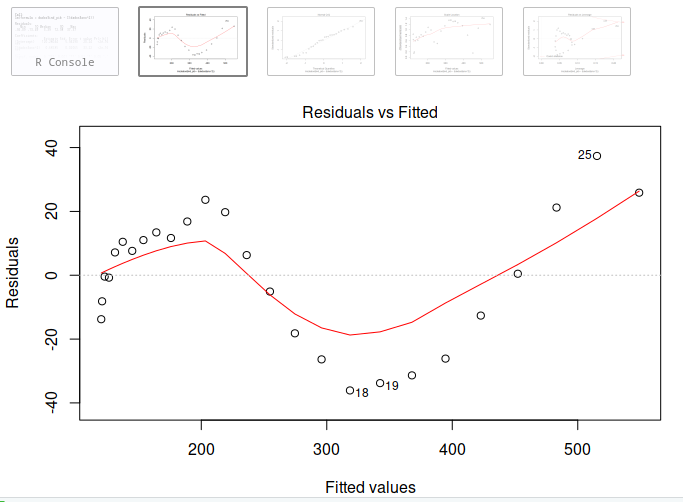
plot(fit8)

Resultado obtido



Note que os resíduos não são aleatórios.

**3.2. Testando a regressão “índice do PIB = A + B\*Ano²”, com intercepto**



Note que os resíduos não são aleatórios.

**3.3. Testando a regressão “índice do PIB = B\*Ano²”, sem intercepto**

Código utilizado:

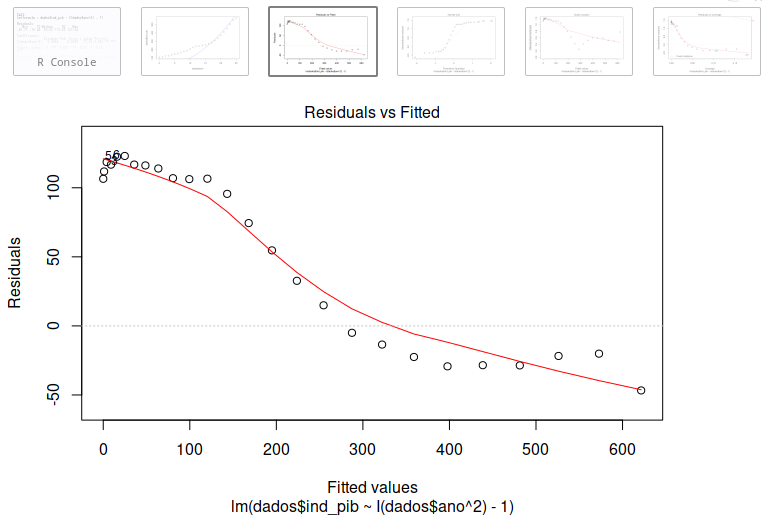
fit10 <- lm(dados$ind\_pib ~ I(dados$ano^2) -1)

summary(fit10)

plot(dados$ano,dados$ind\_pib)

points(dados$ano, predict(fit10),type = 'l', col = 'blue')

plot(fit10)



Note que nenhuma dessas regressões funcionou de modo aceitável. Isso deve-se ao fato de ter ocorrido uma mudança nos dados, como pode-se perceber a partir do ponto 13.

Assim, os dados anteriores não serão úteis para prever as variações futuras. Faremos então uma regressão polinomial excluindo as 13 primeiras variáveis.

**3.3. Refazendo as regressões excluindo os 13 primeiros dados.**

**3.3.1. Testando a regressão “índice do pib = A\*ano + B\*Ano²), com intercepto**

Código utilizado:

ind\_pib = dados$ind\_pib[13:26]

ind\_ano = dados$ano[13:26]

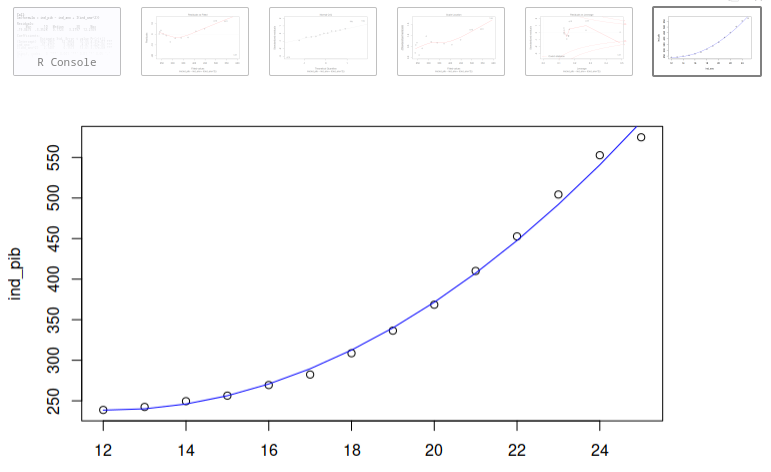
fit1\_ <- lm(ind\_pib ~ ind\_ano + I(ind\_ano^2))

summary(fit1\_)

plot(fit1\_)

plot(ind\_ano,ind\_pib)

points(ind\_ano,predict(fit1\_), type = 'l', col='blue')



Como podemos ver, a previsão do PIB pode ser dado por: Índice PIB = B1\*ano + B2\*ano²

Com o código abaixo podemos saber a previsão para os próximos 5 anos.

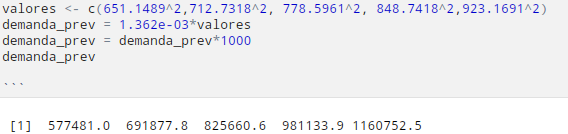


Os valores são:



**4 Projetando o consumo de papelão para os próximos 5 anos**

Vamos agora utilizar esses valores para estipular o consumo de papelão nos próximos 5 anos. O código, assim como o resultado obtido, são:



Logo,a previsão para os próximos 5 anos, serão: (577481.0, 691877.8, 825660.6, 981133.9, 1160752.5)toneladas

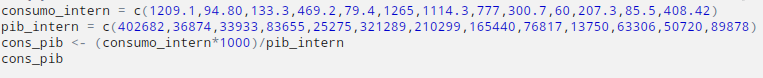
**5 Previsão da oferta para papeis ondulados**

Faz sentido pensar que no futuro a oferta de papelão tenderá a alinhar-se com a demanda do mesmo.

A fim de fazer essa comparação, vamos analisar países desenvolvidos e ver se há um múltiplo semelhante entre o consumo de papelão

e o PIB.

Código utilizado e o resultado obtido:





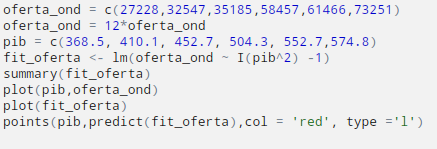
Note que há uma grande variação entre os países.

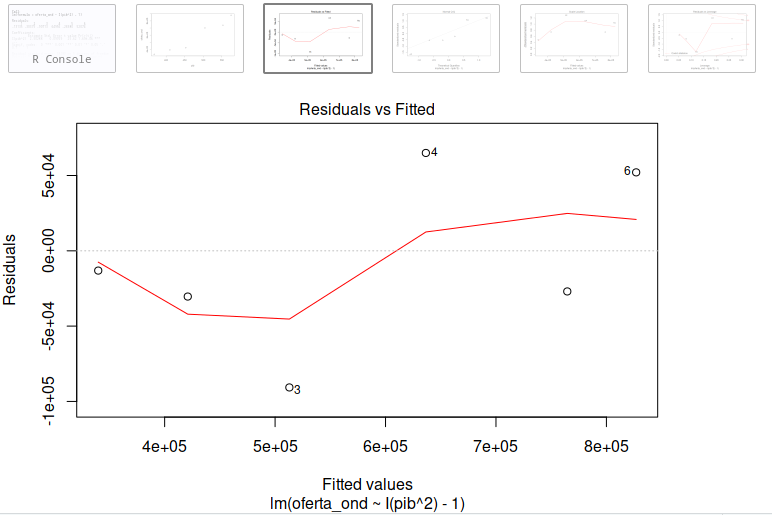
Uma forma de tentarmos prever o quanto será a oferta nos próximos anos é fazendo uma regressão linear, assim como foi feito nos anteriores.

Note que se tratando de uma regressão, apenas 6 valores não garantem uma regressão confiável.

**5.1. Testando regressão “oferta demandada = A\*PIB² ”, sem intercepto**

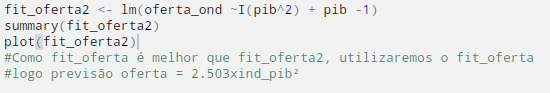
Código utilizado:

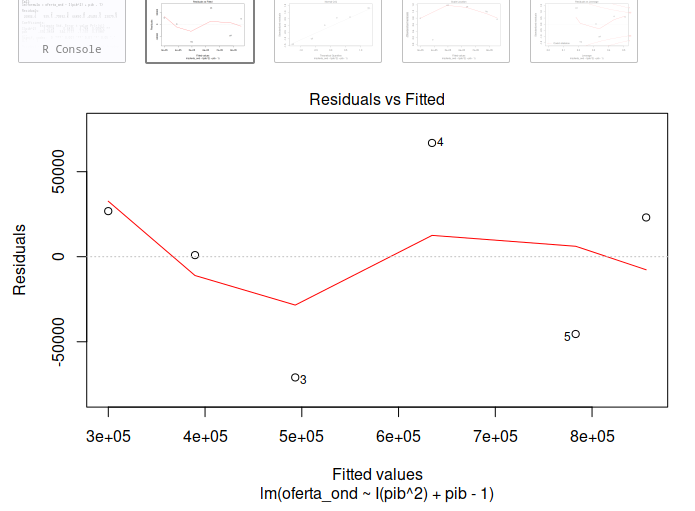




**5.2. Testando regressão “oferta demandada = A\*PIB² + B\*PIB”, sem intercepto**

Código utilizado:





Com base nessas duas regressões, pode-se perceber que a melhor é a do item 5.1., no qual o regressor é dado por: “oferta papelão” = “Previsão PIB²”\*2.503

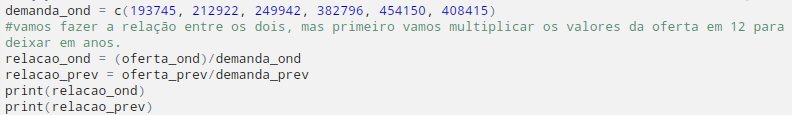
Código utilizado para a previsão da oferta:





**5.3. Fazendo relação entre oferta papelão e “demanda papelão”**

Código utilizado:





**6. Conclusão**

Como podemos ver, há um excesso na oferta tanto atual quanto o futuro. Logo, podemos concluir que a construção de uma nova fábrica não é interessante olhando pela ótica quantitativa. Porém, como o próprio texto diz, existem áreas onde o número de fábricas de papelão é baixo, o que pode ser visto como uma oportunidade.

Outro ponto a ser levado em consideração é o fato de que para uma melhor tomada de decisão seria interessante um maior número de dados, como por exemplo qual a previsão do PIB por UF ou uma série mais longa da oferta.