Lista 2a

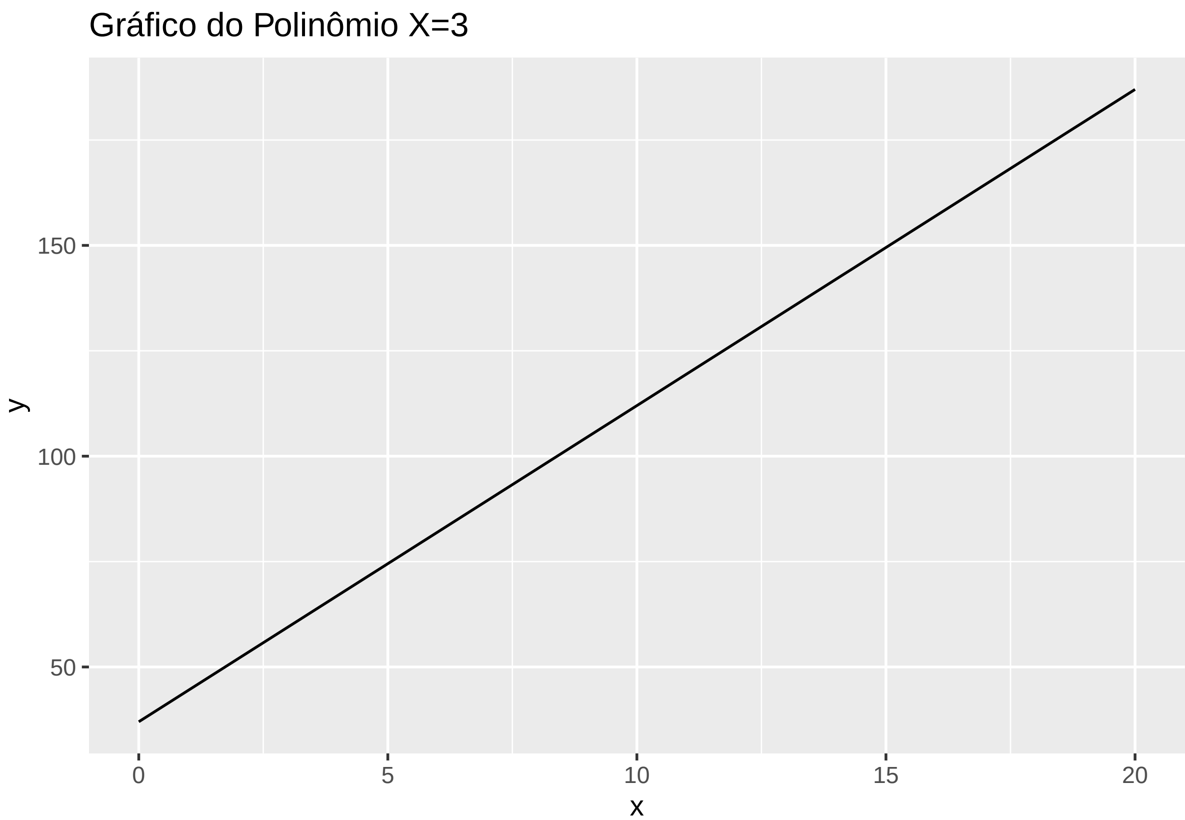
Análise de Regressao Linear

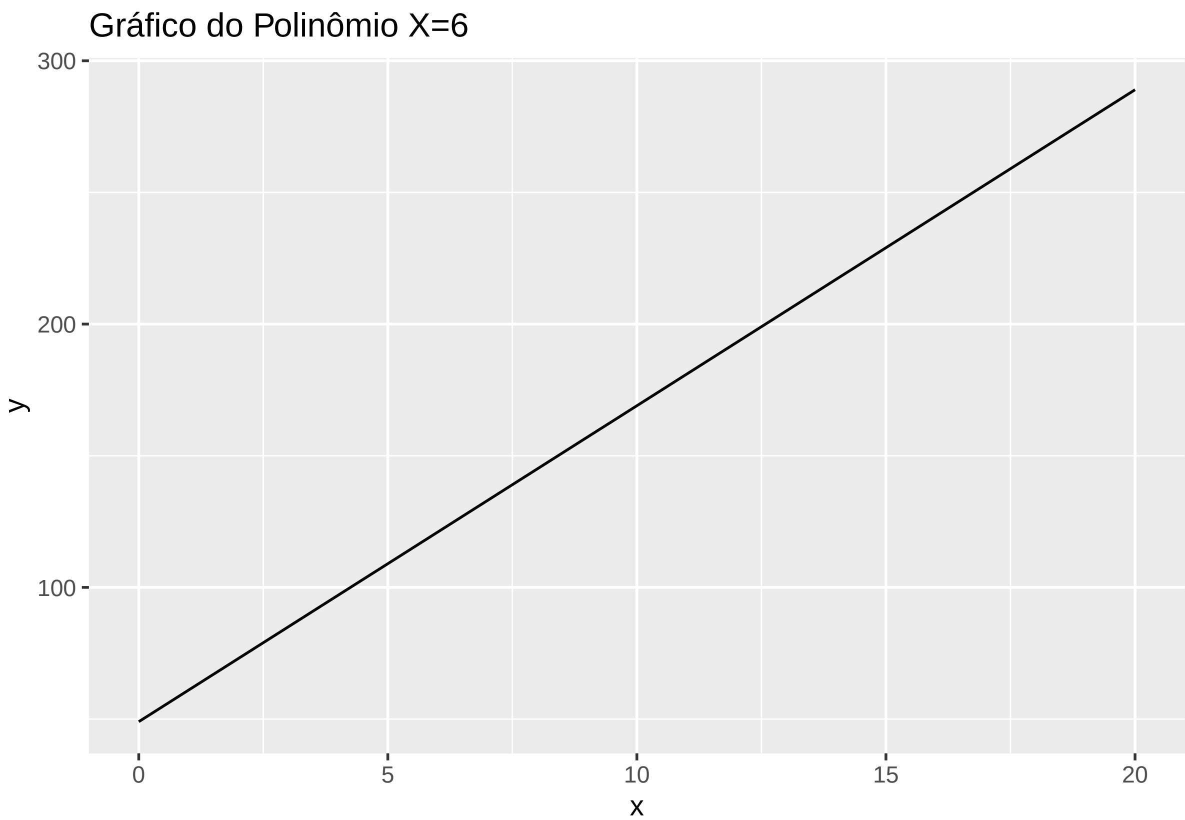
Davi Wentrick

July 10, 2023

1 Considere a função resposta: $E(Y) = 25 + 3X_1 + 4X_2 + 1,5X_1X_2$

1.1 Faça o gráfico de $E(Y) \times X_1$ quando $X_2 = 3$ e $X_2 = 6$.





1.2 Os efeitos de X_1 e X_2 são aditivos? Como você identificou isto no gráfico obtido no item a.

O modelo é aditivo podemos perceber que a escala ela dobra quando dobramos o X indicando que o valores sao maiores

2 Estabeleça a matriz X e os vetores Y β e para os seguintes modelos (assuma que $i = 1, 2, 3, 4$).

2.1 $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i1} X_{i2} + \epsilon_i$

2.2 $\sqrt{Y_i} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$

3 Por que não é siginificativo atribuir um sínal ao coeficiente de correlação múltipla, embora façamos isso para o coeficiente de correlação linear simples?

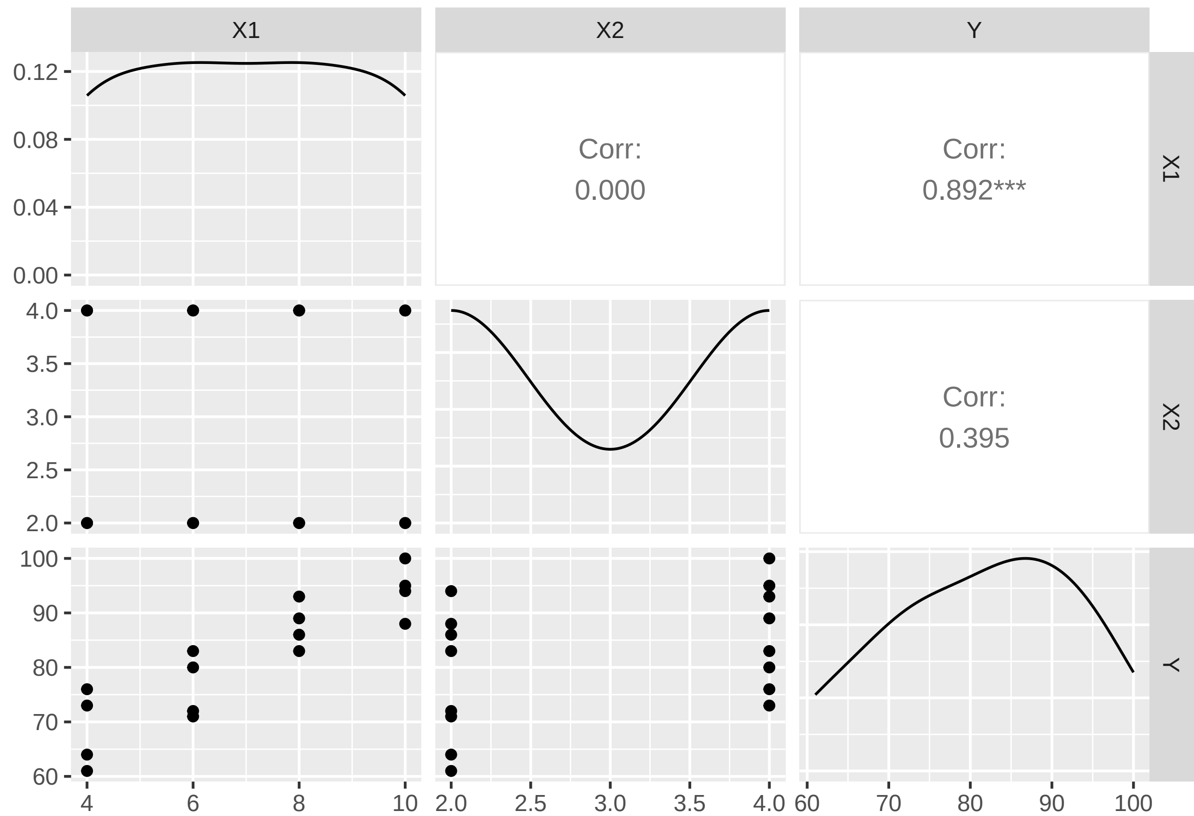
4 Preferência de marca : Vamos estudar a relacao entre gosto da marca (Y),Teor de Umidade (X_1) e Doçura (X_2).

Table 1: Dados do estudo de preferência de marca

	X1	X2	Y
1	4.00	2.00	64.00
2	4.00	4.00	73.00
3	4.00	2.00	61.00
4	4.00	4.00	76.00
5	6.00	2.00	72.00
6	6.00	4.00	80.00
7	6.00	2.00	71.00
8	6.00	4.00	83.00
9	8.00	2.00	83.00
10	8.00	4.00	89.00
11	8.00	2.00	86.00
12	8.00	4.00	93.00
13	10.00	2.00	88.00
14	10.00	4.00	95.00
15	10.00	2.00	94.00
16	10.00	4.00	100.00

4.1 Obtenha o grafico de dispersao e a matriz de correlacao

Vamos utilizar o pacote `GGally` e a funcao `ggpairs()` para criar uma matriz que contem os gráficos de dispersao, densidade e correlacao entre as variaveis do banco



Podemos notar pelo grafico de dispersao que a variavel X_2 é categorica com 2 valores 4 e 2.

4.2 Ajuste o modelo aos dados. Escreva a funcao de regressao e interprete β_1

Ao ajustar o modelo no R temos o seguinte resultado

```
##
## Call:
## lm(formula = Y ~ X1 + X2, data = dados)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  46.4000     2.3129  20.061 3.66e-11 ***
## X1            4.4250     0.3011  14.695 1.78e-09 ***
## X24           8.7500     1.3466   6.498 2.01e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.693 on 13 degrees of freedom
## Multiple R-squared:  0.9521, Adjusted R-squared:  0.9447
## F-statistic: 129.1 on 2 and 13 DF,  p-value: 2.658e-09
```

Table 2: Matriz de Variancia-Covariancia dos Betas

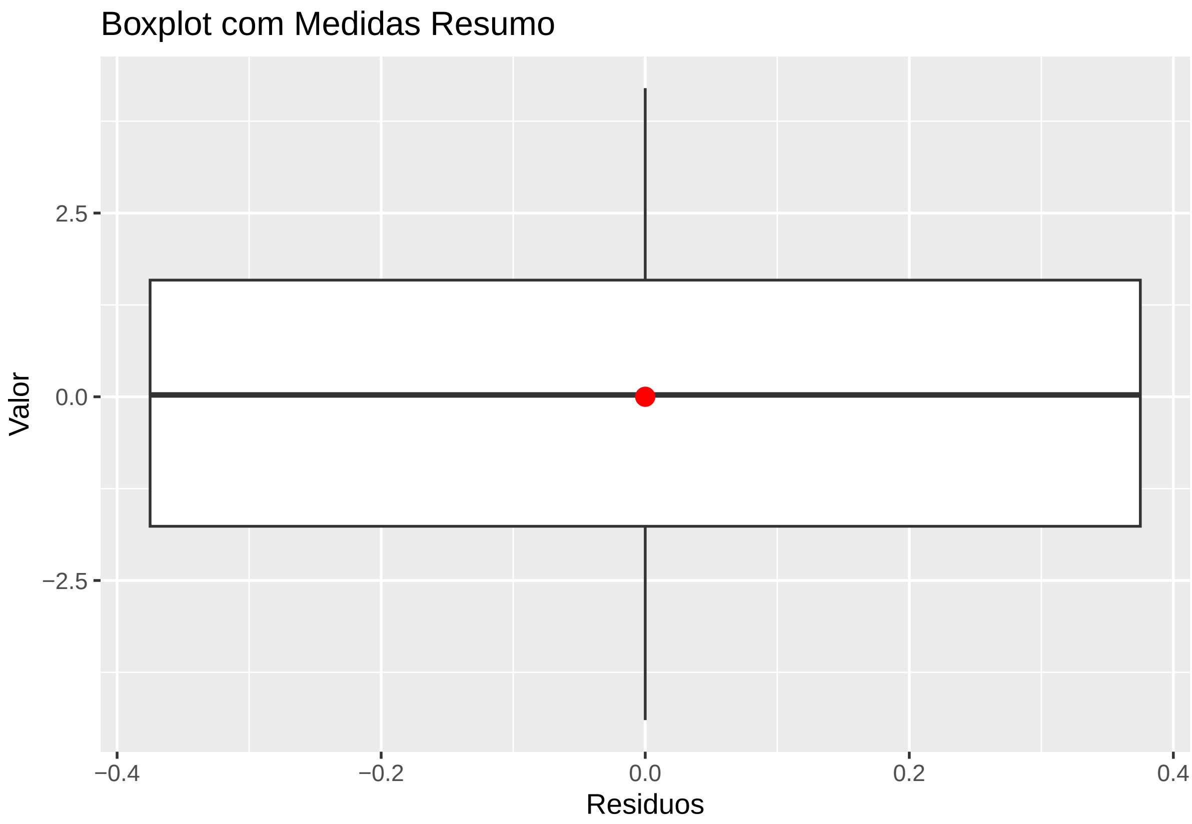
	(Intercept)	X1	X24
(Intercept)	5.35	-0.63	-0.91
X1	-0.63	0.09	0.00
X24	-0.91	0.00	1.81

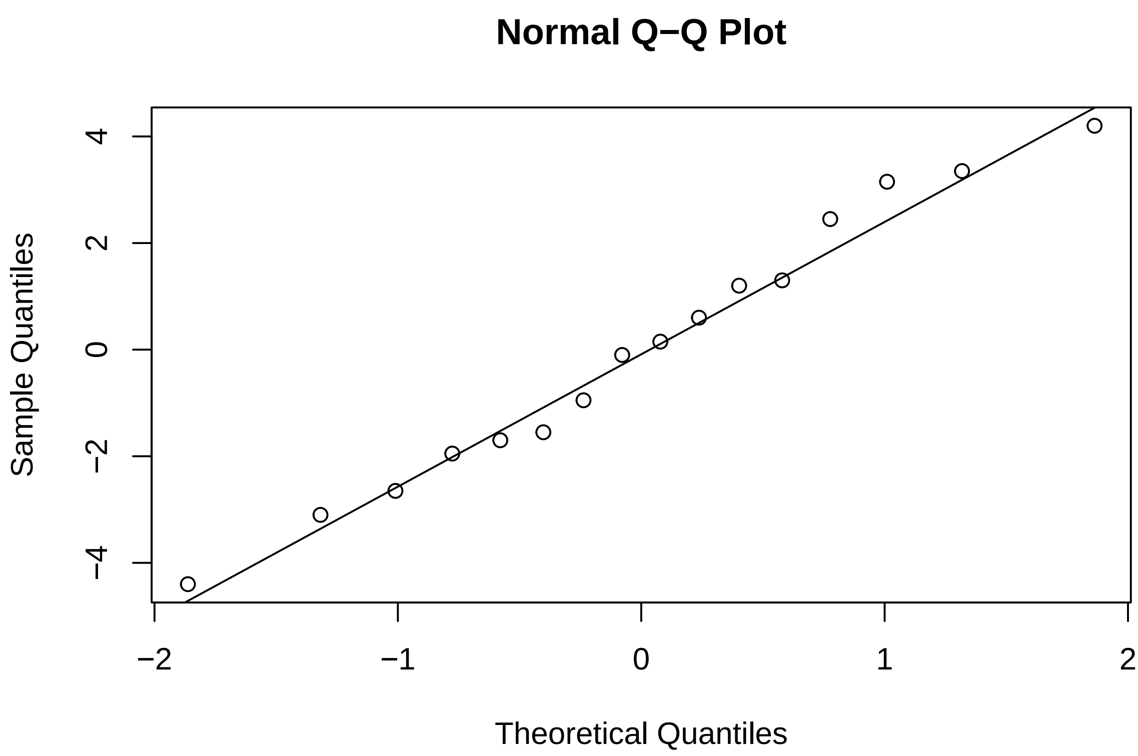
E obtemos a seguinte equacao de regressao:

$$Y = 46,4 + 4.425X_1 + 8.75X_2$$

Nosso β_1 é o aumento que teremos no Gosto da marca (Y) a cada unidade fixado um X_2

4.3 Obtenha os residuo, monte um boxplot e interprete.





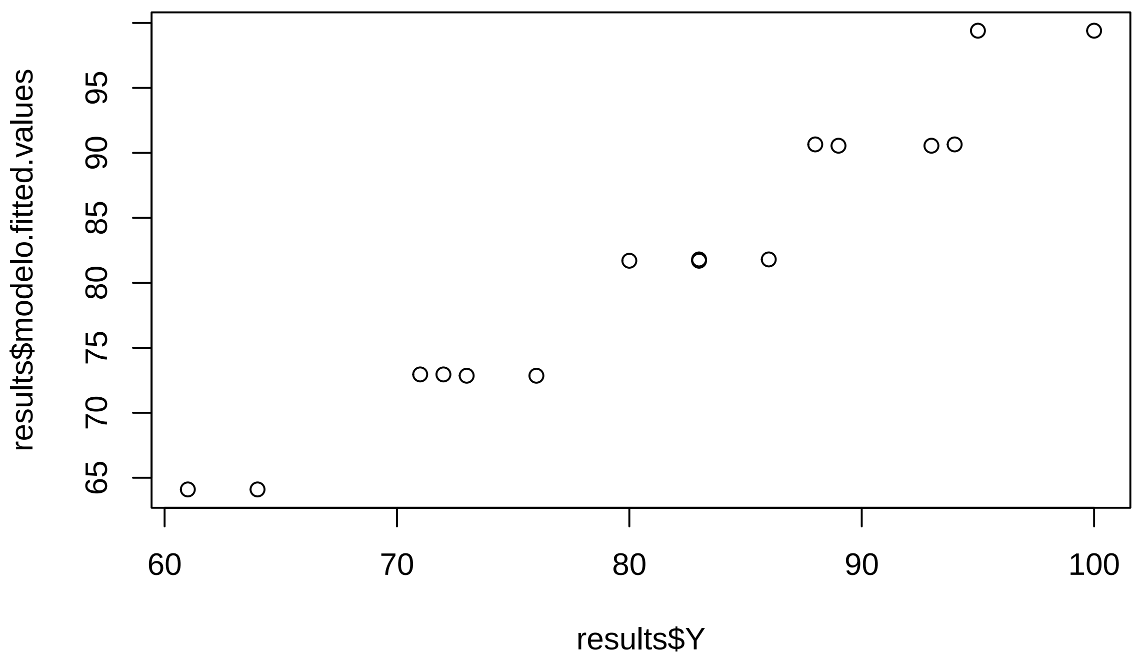
O boxplot parece indicar que os residuos seguem normalidade e podemos confirmar isso pelo gráfico QQplot feito usando os valores residuais. Logo podemos realizar um teste de Shapiro para verificar essa hipotese:

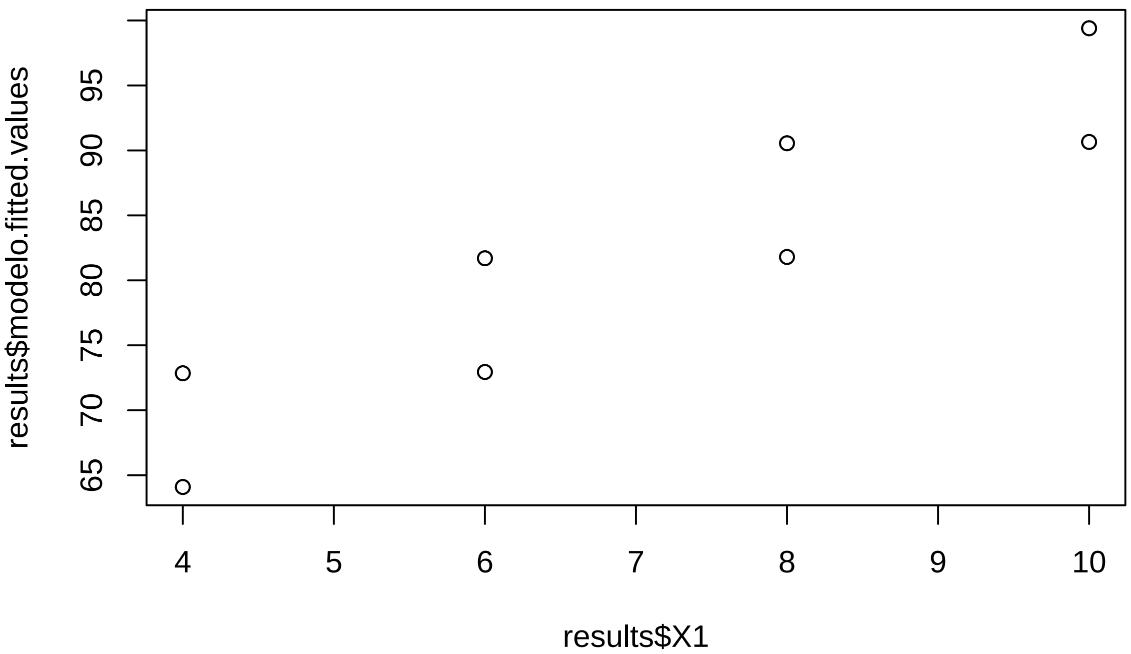
```
##
##  Shapiro-Wilk normality test
##
## data:  residuos
## W = 0.97585, p-value = 0.9222
```

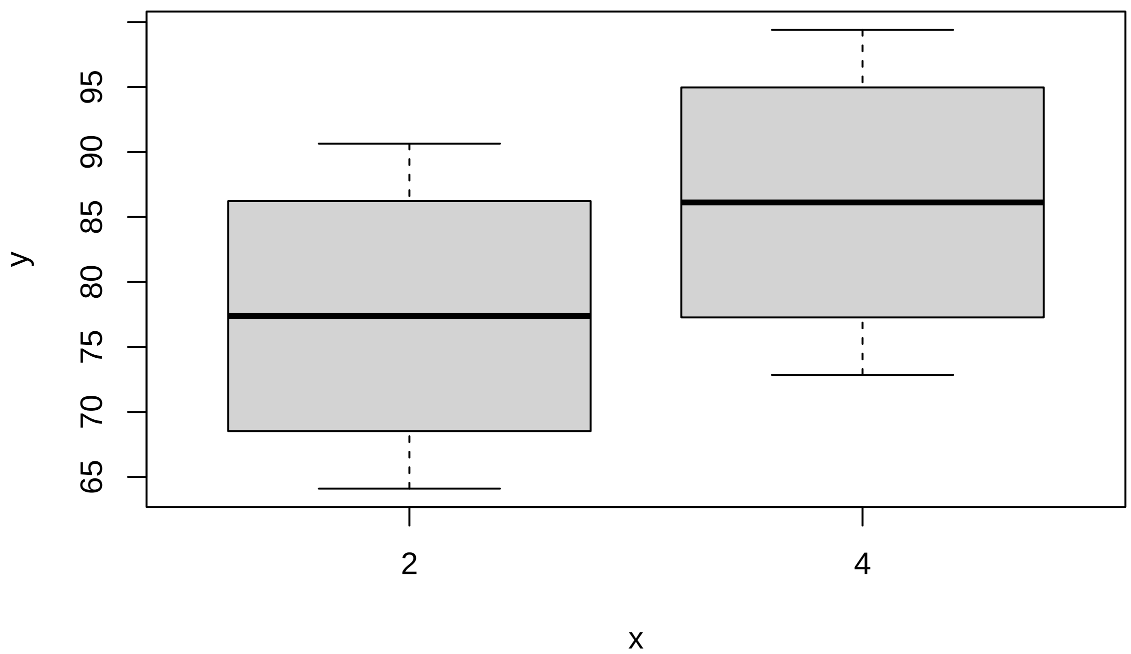
Agora com o p-valor de 0.922 podemos assumir que os residuos seguem normalidade

4.4 Faca o plot dos residuos contra $\hat{Y}$,X_1,X_2 e X_1X_2 em graficos separados. E faca um grafico de probabilidade Normal e interprete os resultados

O grafico de normalidade ja foi feito anteriormente




results$modelo.fitted.values
65
70
75
80
85
90
95
4
5
6
7
8
9
10
results$X1


y
65
70
75
80
85
90
95
2
4
x

4.5 Realize o teste de Breusch-Pagan para a consistencia da variancia do erro com $\alpha = 0.01$.

```
## [1] 0.9999942
```

4.6 Realize um teste de falta de ajuste com $\alpha = 0.01$

```
## [1] 0.007077954
```

4.7 Assumindo que a regressao tem erro normal e independente:

4.7.1 Teste se tem um relacao de regressao com $\alpha = 0.01$. O que seu teste diz sobre β_1 e β_2

4.7.2 Qual é o p-valor do teste?

4.7.3 Estime β_1 e β_2 conjuntamente pelo metodo de bonferroni com uam confianca de 99%. interprete os resultados

```
## Intervalo de confiança de Bonferroni para beta 1: [ 4.19745 , 4.65255 ]

## Intervalo de confiança de Bonferroni para beta 1: [ 4.198992 , 13.30101 ]
```

4.7.4 Calcule o coeficiente de determinacao multiplo R^2. Como ele é interpretado?

```
## [1] 0.952059
```

4.7.5 Calcule o coeficiente de determinacao simples entre Y_i e $\hat{Y}_i$. Ele é igual ao coeficiente de determinacao multiplo?

4.7.6 Obtenha o intervalo de estimacao para $E(Y_h)$ quando $X_{h1} = 5$ e $X_{h2} = 4$ use um intervalo com 99% de confianca para sua estimativa

```
## O valor estimado Y quando X1 e X2 = 5 4 é: 103.525

## Intervalo de confiança para a média de Y: [ 34.96545 , 172.0846 ]
```

4.7.7 Obtenha um intervalo de predicao para uma nova observacao $Y_{h(nova)}$ quando $X_{h1} = 5$ e $X_{h2} = 4$ use um intervalo com 99% de confianca para sua estimativa

```
## O valor estimado Y quando X1 e X2 = 5 4 é: 103.525

## Intervalo de confiança de Bonferroni para a média de Y: [ 34.96545 , 172.0846 ]
```