

# EXPERIMENTO DOSAGEM DE $P_2O_5$ - DBC

Elderson Silva, Évelyn Muniz, João Borré

Os dados referem-se a produções de milho (Y), em kg/parcela, de um experimento casualizado em blocos de adubação de milho com diferentes doses (X) de  $P_2O_5$ .

```
dados <- data.frame(
  Bloco = c("I", "II", "III", "I", "II", "III", "I",
            "II", "III", "I", "II", "III", "I", "II", "III"),
  Tratamento = c(1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5),
  Dose = c(0, 0, 0, 25, 25, 25, 50, 50, 50, 75, 75, 75, 100, 100, 100),
  Producao = c(85.0, 86.0, 84.0, 94.5,
               96.0, 95.8, 99.5, 98.0, 104.0, 93.0, 96.0,
               90.5, 83.0, 80.0, 78.5)
)

attach(dados)

dados
```

	Bloco	Tratamento	Dose	Producao
1	I	1	0	85.0
2	II	1	0	86.0
3	III	1	0	84.0
4	I	2	25	94.5
5	II	2	25	96.0
6	III	2	25	95.8
7	I	3	50	99.5
8	II	3	50	98.0
9	III	3	50	104.0
10	I	4	75	93.0
11	II	4	75	96.0
12	III	4	75	90.5
13	I	5	100	83.0
14	II	5	100	80.0
15	III	5	100	78.5

1:

a) Existe diferença significativa entre as doses de  $P_2O_5$ ?  $\alpha = 5\%$

O experimento proposto é do tipo delineamento em blocos casualizados (DBC), com tratamentos quantitativos, dessa forma, para verificar se existe diferença entre os níveis de dosagem, deve-se proceder com a ANOVA.

Inicialmente, foi proposto o modelo aov, para estudar a variabilidade entre as doses:

```
modelo <- aov(Producao ~ as.factor(Dose)
              + as.factor(Bloco))
# factor transforma valores categóricos em níveis distintos.
```

Posteriormente, a análise de variância foi executada:

```
anova(modelo)
```

Analysis of Variance Table

Response: Producao

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(Dose)	4	782.45	195.613	32.9999	5.087e-05 ***
as.factor(Bloco)	2	1.07	0.536	0.0904	0.9145
Residuals	8	47.42	5.928		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

As hipóteses a serem testadas por meio da ANOVA:

$H_0$  : As médias das produções de milho são iguais para todas as doses.

$H_1$  : Pelo menos uma média é diferente.

Ao nível de significância de 5%, com 4 graus de liberdade, a hipótese nula foi rejeitada, uma vez que o  $p$ -value foi significativo ( $5,087^{-5}$ ), concluindo que existe pelo menos uma média de produção de milho com valor estatisticamente diferente.

### Pressupostos:

Para validar a inferência obtida via ANOVA, é necessária a validação de pressupostos estatísticos de normalidade, homocedasticidade e independência de resíduos. Dessa forma:

### Normalidade:

Para normalidade de resíduos foi aplicado o teste *Shapiro. Wilk*, sob as hipóteses:

$H_0$  : Os resíduos seguem distribuição normal.

$H_1$  : Os resíduos não seguem distribuição normal.

```
shapiro.test(residuals(modelo))
```

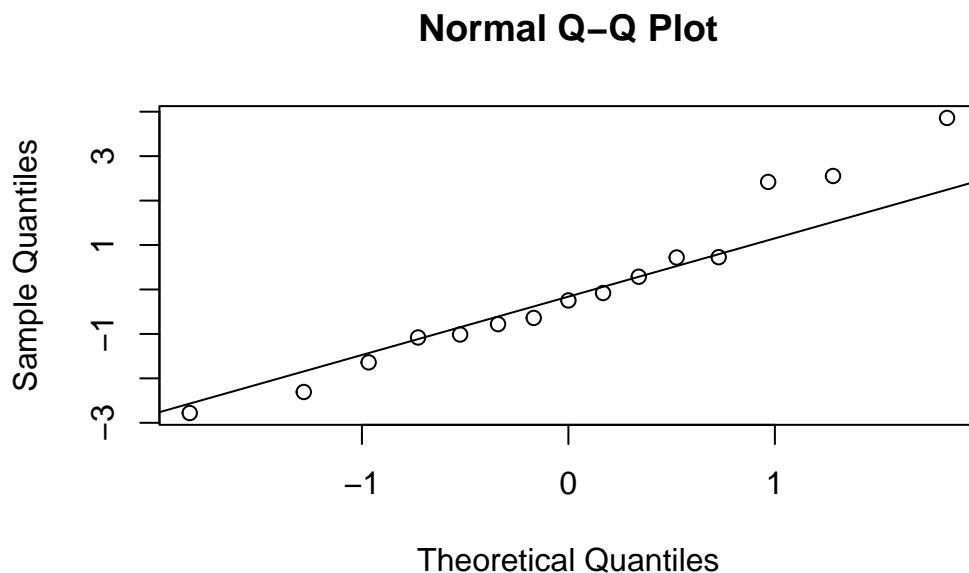
Shapiro-Wilk normality test

```
data: residuals(modelo)
W = 0.95543, p-value = 0.6136
```

O *p-value* obtido foi de 0,6136 indicando que não há evidências estatísticas contra  $H_0$  ao nível de significância de 5%, ou seja, não há evidências para rejeitar a hipótese de que os resíduos apresentam normalidade.

Para corroborar com o teste, foi aplicado o **gráfico Q-Q**:

```
qqnorm(residuals(modelo))
qqline(residuals(modelo))
```



O gráfico obtido indica a normalidade dos resíduos, uma vez que os pontos estão próximos da reta, caso houvesse uma dispersão ou distanciamento da reta, seria um indicativo de não-normalidade.

**Homocedasticidade:**

Foi aplicado o teste de *Bartlett* sob as hipóteses:

$H_0$  : Os resíduos seguem distribuição normal.

$H_1$  : Os resíduos não seguem distribuição normal.

```
bartlett.test(Producao ~ Tratamento, data = dados)
```

Bartlett test of homogeneity of variances

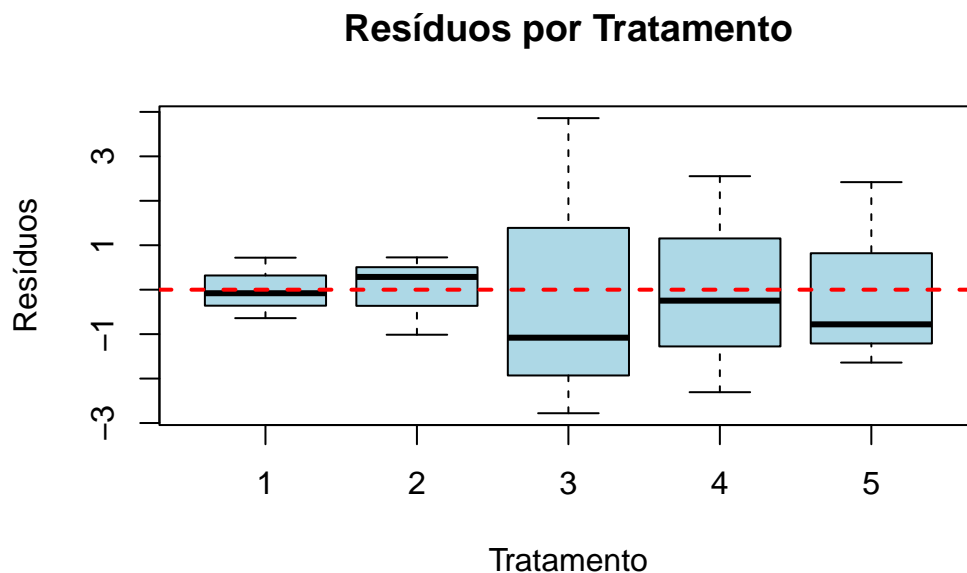
data: Producao by Tratamento

Bartlett's K-squared = 3.9054, df = 4, p-value = 0.419

O *p-value* obtido foi de 0,419 indicando que não há evidências estatísticas contra a hipótese nula, ao nível de significância de 5%, ou seja, não há evidências para rejeitar a hipótese de que os resíduos são homocedásticos.

De forma visual, foi realizado o *boxplot* dos dados, afim de verificar visualmente o comportamento dos resíduos:

```
boxplot(residuals(modelo) ~ dados$Tratamento,  
        col = "lightblue",  
        main = "Resíduos por Tratamento",  
        xlab = "Tratamento", ylab = "Resíduos")  
abline(h = 0, col = "red", lty = 2, lwd = 2)
```

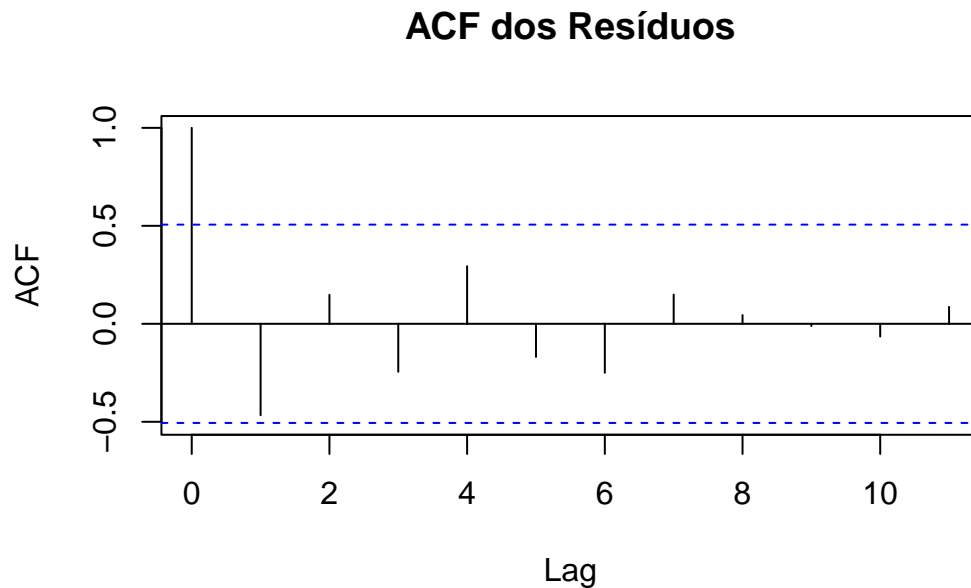


É possível identificar que os tratamentos/dosagens 3, 4 e 5 apresentam uma maior variabilidade em relação aos tratamentos/dosagens 1 e 2, devido ao achatamento das caixas, porém, sem presença de *outliers*.

### Independência de resíduos:

Para verificar se os resíduos estão autocorrelacionados, foi criado o gráfico *acf*:

```
acf(residuals(modelo),  
    main = "ACF dos Resíduos")
```



Por meio do gráfico, é possível identificar que os resíduos são independentes, pois nenhuma linha ultrapassou a linha pontilhada azul. Entretanto, para análise ser ainda mais robusta, também foi utilizado o teste de *Durbin-Watson*:

```
library(lmtest)
```

Carregando pacotes exigidos: zoo

Anexando pacote: 'zoo'

Os seguintes objetos são mascarados por 'package:base':

```
as.Date, as.Date.numeric
```

```
dwtest(modelo)
```

Durbin-Watson test

data: modelo

DW = 2.8764, p-value = 0.7882

alternative hypothesis: true autocorrelation is greater than 0

As hipóteses:

$H_0$  : Os resíduos são independentes.

$H_1$  : Os resíduos não são independentes.

Ao nível de significância de 5% e com  $p\text{-value} = 0,7882$ , não há evidências para rejeitar a hipótese nula, sendo assim, conclui-se com 95% de confiança que os resíduos são independentes.

Com todos os pressupostos atendidos, foi proposto um modelo que representasse a relação entre a dosagem e a produção de milho:

b) Equação de regressão polinomial mais adequada:

Após concluir que existe diferença entre as dosagens, faz sentido ajustar um modelo de regressão que seja capaz de modelar como os valores de dosagens influenciam os diferentes níveis de produção:

**Modelo polinomial de primeiro grau:**

```
Lm1 <- lm(Producao ~ (Dose) + (Bloco)) #Linear Model 1
Lm2 <- lm(Producao ~ as.factor(Dose)
          + as.factor(Bloco)) #Linear Model 2
anova(Lm1,Lm2)
```

Analysis of Variance Table

Model 1: Producao ~ (Dose) + (Bloco)

Model 2: Producao ~ as.factor(Dose) + as.factor(Bloco)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	11	791.79				
2	8	47.42	3	744.37	41.858	3.089e-05 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

As hipóteses testadas:

$H_0$  : O modelo é adequado.

$H_1$  : O modelo não é adequado.

Com  $p\text{-value} = 3,089^{-5}$ , ao nível de significância de 5%, os resíduos de regressão foram significativos, indicando que existe um modelo de grau superior mais adequado.

**Modelo polinomial de segundo grau:**

```
Lm3 <- lm(Producao ~ (Dose)
          + I(Dose^2) + Bloco) #Modelo 3

Lm4 <- lm(Producao ~ as.factor(Dose) +
          as.factor(Dose^2) +
          as.factor(Bloco)) #Modelo 4
```

As hipóteses testadas:

$$H_0 : \text{O modelo é adequado.}$$

$$H_1 : \text{O modelo não é adequado.}$$

```
anova(Lm3,Lm4)
```

Analysis of Variance Table

```
Model 1: Producao ~ (Dose) + I(Dose^2) + Bloco
Model 2: Producao ~ as.factor(Dose) + as.factor(Dose^2) + as.factor(Bloco)
  Res.Df    RSS Df Sum of Sq    F Pr(>F)
1      10 55.942
2       8 47.421  2    8.5208 0.7187 0.5163
```

Com  $p\text{-value} = 0,5163$ , ao nível de significância de 5%, os resíduos de regressão não foram significativos, indicando que o modelo polinomial de grau 2 é adequado, e que, modelos com graus superiores não são estatisticamente significativos quanto sua capacidade descrever os dados, sendo assim, é cabível escolher o modelo mais parcimonioso.

c) Gráfico associado ao modelo ajustado com valores observados e a curva ajustada:

```
anova(Lm4)
```

Analysis of Variance Table

```
Response: Producao
      Df Sum Sq Mean Sq F value    Pr(>F)
as.factor(Dose)  4 782.45  195.613 32.9999 5.087e-05 ***
as.factor(Bloco)  2   1.07   0.536  0.0904  0.9145
Residuals       8  47.42   5.928
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(ExpDes.pt)
reg.poly(Producao, Dose, glres=8,
         SQres=47.42, gltrat=4, SQtrat=782.45)
```

## Ajuste de modelos polinomiais de regressao

---

### Modelo Linear

```
=====
      Estimativa Erro.padrao   tc   valor.p
-----
b0  93.1733      1.0888      85.5738    0
b1  -0.0451      0.0178     -2.5347 0.0350
-----
```

### R2 do modelo linear

```
-----
0.048669
-----
```

### Analise de variancia do modelo linear

```
=====
              GL    SQ      QM    Fc   valor.p
-----
Efeito linear      1 38.0813 38.0813  6.42   0.035
Desvios de Regressao 3 744.3687 248.1229 41.86 3e-05
Residuos           8  47.4200   5.9275
-----
-----
```

### Modelo quadratico

```
=====
      Estimativa Erro.padrao   tc   valor.p
-----
b0  84.8019      1.3229      64.1038    0
b1   0.6246      0.0627      9.9653 0.00001
b2  -0.0067      0.0006     -11.1419    0
-----
```

### R2 do modelo quadratico

```
-----
0.989111
-----
```



# Analise de variancia do modelo quadratico

	GL	SQ	QM	Fc	valor.p
Efeito linear	1	38.0813	38.0813	6.42	0.035
Efeito quadratico	1	735.8486	735.8486	124.14	0
Desvios de Regressao	2	8.5201	4.2600	0.72	0.51636
Residuos	8	47.4200	5.9275		

## Modelo cubico

	Estimativa	Erro.padrao	tc	valor.p
b0	84.7986	1.3956	60.7628	0
b1	0.6256	0.1420	4.4054	0.0023
b2	-0.0067	0.0036	-1.8644	0.0993
b3	0.000000	0.00002	0.0075	0.9942

## R2 do modelo cubico

0.989111

# Analise de variancia do modelo cubico

	GL	SQ	QM	Fc	valor.p
Efeito linear	1	38.0813	38.0813	6.42	0.035
Efeito quadratico	1	735.8486	735.8486	124.14	0
Efeito cubico	1	0.0003	0.0003	0	0.9942
Desvios de Regressao	1	8.5198	8.5198	1.44	0.26488
Residuos	8	47.4200	5.9275		

## \$`Quadro de medias`

	Niveis	Medias Observadas
1	0	85.00000
2	100	80.50000
3	25	95.43333
4	50	100.50000

5            75                    93.16667

```
$`Coeficientes reta`
```

```
      [,1]
```

```
[1,] 93.17333333
```

```
[2,] -0.04506667
```

```
$`R2 reta`
```

```
[1] 0.04866935
```

```
$`Coeficientes parabola`
```

```
      [,1]
```

```
[1,] 84.801904762
```

```
[2,] 0.624647619
```

```
[3,] -0.006697143
```

```
$`R2 parabola`
```

```
[1] 0.989111
```

```
$`Coeficientes cubica`
```

```
      [,1]
```

```
[1,] 8.479857e+01
```

```
[2,] 6.256032e-01
```

```
[3,] -6.723810e-03
```

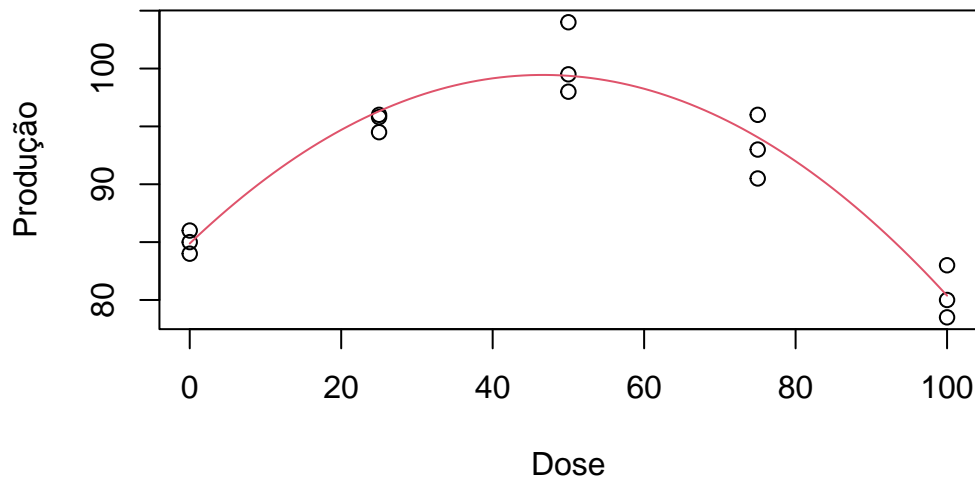
```
[4,] 1.777777e-07
```

```
$`R2 cubica`
```

```
[1] 0.9891114
```

```
x <- seq(0, 100, 1)
plot(Dose, Producao, xlab="Dose",
     ylab="Produção",
     main="Curva ajustada e valores observados")
curve(84.881905 + 0.624648 * x - 0.006697 * x^2,
      0, 100, col=2, add=TRUE)
```

### Curva ajustada e valores observados



O gráfico obtido é uma parábola com concavidade negativa, o que indica que existe um valor máximo de produção, associado a uma dosagem de  $P_2O_5$ . Visualmente, pode-se dizer que a dosagem que oferece o valor máximo de produção está entre 40 e 60 doses, porém, para identificar esse valor, é necessário calcular a derivada do modelo ajustado.

**Pressupostos para o modelo linear:**

**Normalidade:**

$H_0$  : Os resíduos seguem distribuição normal.

$H_1$  : Os resíduos não seguem distribuição normal.

```
shapiro.test(residuals(Lm3))
```

Shapiro-Wilk normality test

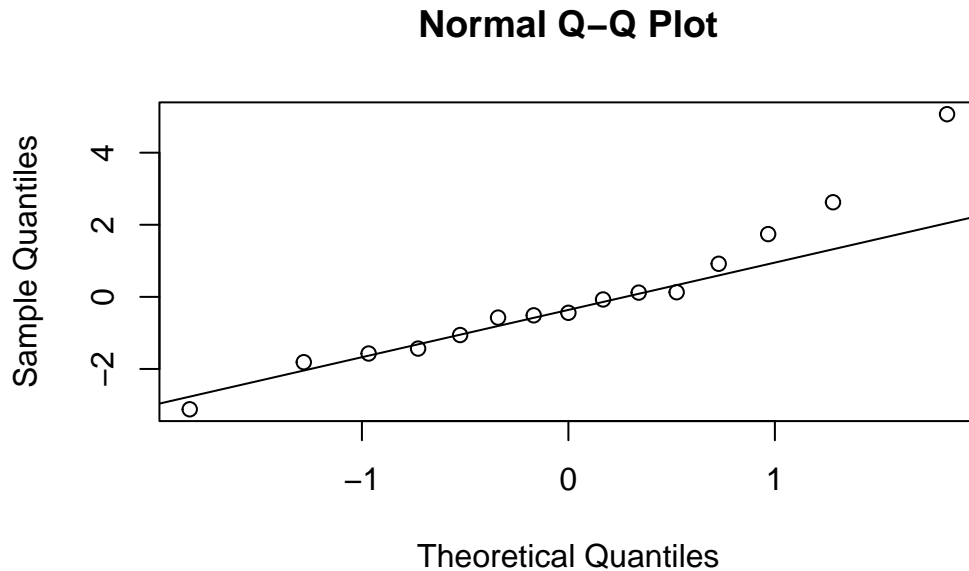
```
data: residuals(Lm3)
```

```
W = 0.92478, p-value = 0.2277
```

O *p-value* obtido foi de 0,2277 indicando que não há evidências estatísticas contra a hipótese nula, ao nível de significância de 5%, ou seja, não há evidências para rejeitar a hipótese de que os resíduos apresentam normalidade.

Gráfico Q-Q:

```
qqnorm(residuals(Lm3))  
qqline(residuals(Lm3))
```



O gráfico obtido indica a normalidade dos resíduos.

### Homocedasticidade:

Teste de *Breusch-Pagan*:

$H_0$  : Os resíduos são homocedásticos.

$H_1$  : Os resíduos não são homocedásticos.

```
library(lmtest)  
bptest(Lm3)
```

studentized Breusch-Pagan test

data: Lm3

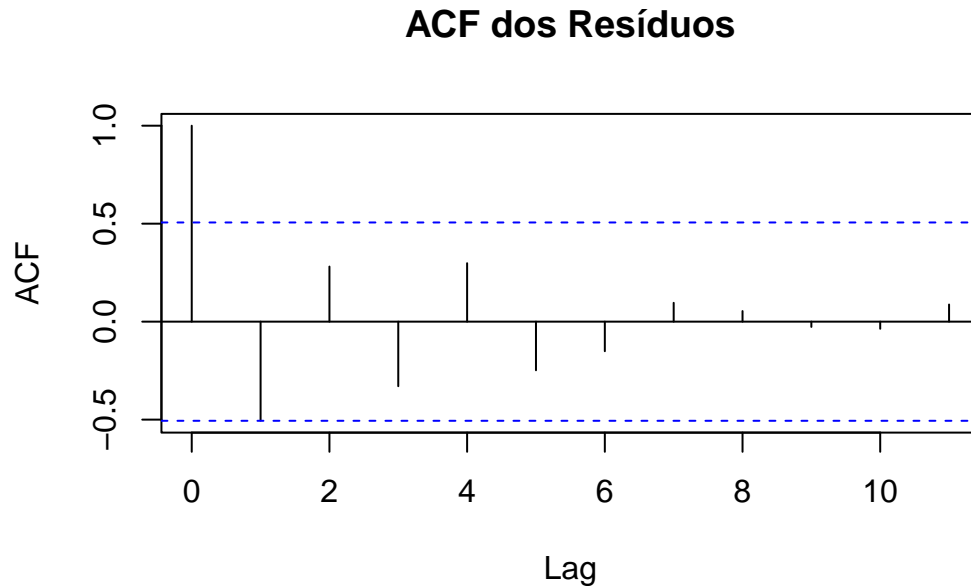
BP = 4.8331, df = 4, p-value = 0.3049

O *p-value* obtido foi de 0,3049 indicando que não há evidências estatísticas contra a hipótese nula ao nível de significância de 5%.

### Independência de resíduos:

Para verificar se os resíduos estão autocorrelacionados, foi criado o gráfico acf:

```
acf(residuals(Lm3),  
    main = "ACF dos Resíduos")
```



Por meio do gráfico *acf*, é possível identificar que os resíduos são independentes, pois nenhuma linha ultrapassou a linha pontilhada azul. Entretanto, para análise ser ainda mais robusta, também foi utilizado o teste de *Durbin-Watson*:

```
library(lmtest)  
dwtest(Lm3)
```

#### Durbin-Watson test

```
data: Lm3  
DW = 2.9718, p-value = 0.9334  
alternative hypothesis: true autocorrelation is greater than 0
```

As hipóteses:

$H_0$  : Os resíduos são independentes.

$H_1$  : Os resíduos não são independentes.

Ao nível de significância de 5% e com  $p\text{-value} = 0,9334$ , não há evidências para rejeitar  $H_0$ .

Com todos os pressupostos atendidos, é possível realizar inferências a partir do modelo proposto, incluindo valores de máxima produção e dosagem mínima que otimize a produção, com  $R\text{-squared} = 0,9891$ , o que implica que o modelo explica 98,91% dos dados.