

### 3. QUESTÃO 3

#### 3.1. ANÁLISE ESTATÍSTICA DESCRITIVA

##### 3.1.1 GERAL

Na tabela a seguir estão os valores de análise estatística descritiva geral, obtidos através do software R, em conformidade com o script anexado (ANEXO 2).

**Tabela 2:** Estatística descritiva geral dos dados.

Razão Eficiência da Dieta (g dieta/g ganho de peso) dos pintinhos	
Soma	30,85
Mín.	1,540
1º Quartil	1,583
Mediana	1,660
Média	1,714
3º Quartil	1,772
Máx.	2,200
Amplitude	0,66
Variância	0,031
Desvio Padrão	0,176
Coef. Variação (CV) (%)	10,279
Coef. de Assimetria (As)	1,357
Coef. de Curtose (CUR)	4,241

\* Os valores possuem unidade em g dieta/g ganho de peso e o C.V. em porcentagem

Observa-se a partir da tabela 2, que a soma geral das razões de eficiência (em g dieta / g ganho de peso) foi de 30,85 g/g. A máxima dos valores é de 2,200 e a mínima é de 1,540. Com até 1,583 g/g, mostram-se 25% das observações, enquanto com 1,772 g/g, apresentam-se 75% dos valores.

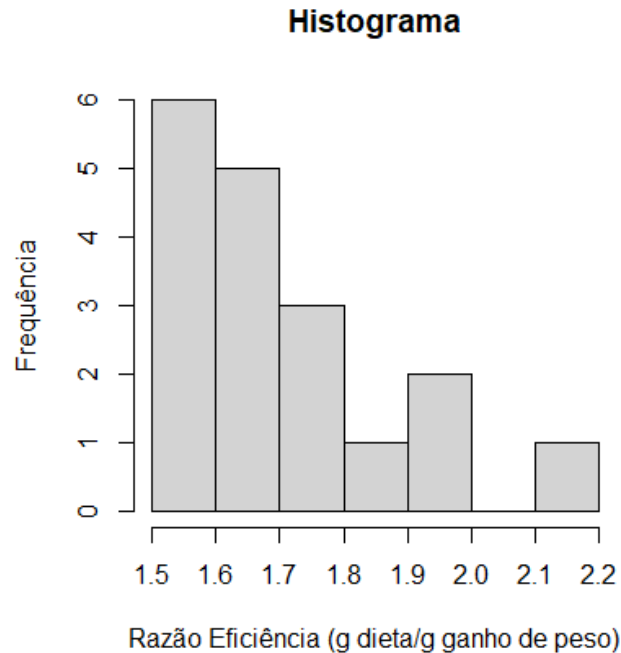
A mediana apresentou valor de 1,66 g/g, significando que 50% das observações ficaram abaixo e acima desses valores. Já a média das razões de eficiência das dietas foi de 1,714 g/g, com desvio padrão de 0,176. O coeficiente de variação (C.V.) apresenta valor de 10,28%, o que significa que os dados possuem uma dispersão média (PIMENTEL GOMES, 2000).

O Coeficiente de assimetria é um valor que quando se tem valor igual a 0 diz-se que a distribuição é simétrica. Quando esse valor é maior que 0, como observou-se um valor de 1,357, tem-se uma a distribuição assimétrica é positiva.

O coeficiente de curtose é um valor que igual a 3 distribuição com caudas neutras (normais – mesocúrtica). Já se esse número for menor que 3, a distribuição é com caudas longas e pesadas (leptocúrtica). Já se o valor calculado for maior que 3,

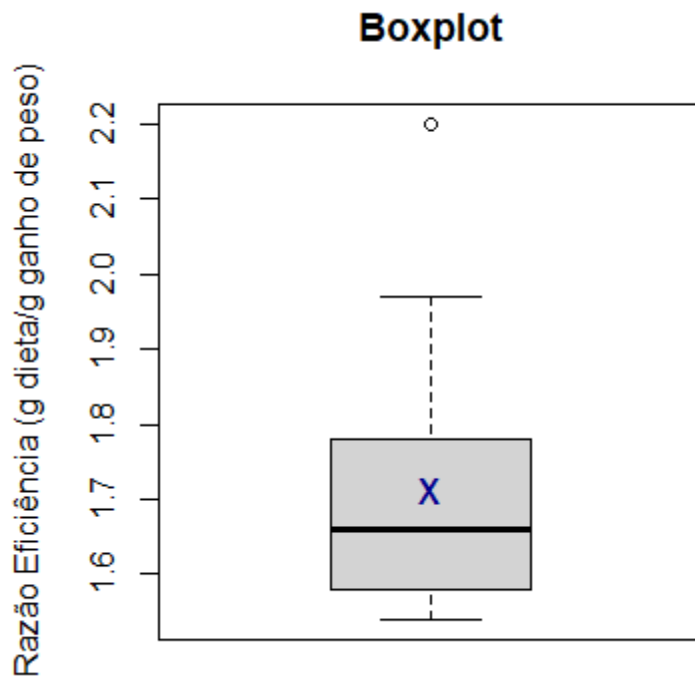
tem-se distribuição com caudas curtas ou leves (platicúrtica). No caso da presente análise, o valor da curtose foi de 4,24 e, com isso, permite dizer que a curva é de cauda curta, sendo classificada platicúrtica.

**Figura 1:** Histograma das Razões de Eficiência das Dietas.



No histograma (figura 1) é possível, através de análise gráfica, analisar o comportamento da variável resposta, em relação a sua distribuição. Nota-se que a distribuição não é simétrica, tendo uma concentração maior de valores menores, sendo que a maior frequência fora entre 1,5 e 1,6 g/g. Já na Figura 2, se apresenta o gráfico box-plot, que indica a área de concentração dos valores, em relação aos quartis, bem como a média geral. Novamente, nota-se uma dispersão assimétrica positiva, visto que a dispersão superior é maior. Ainda, pode-se destacar também a existência de um possível outlayer acima da cauda superior.

**Figura 2:** Histograma das Razões de Eficiência das Dietas.



### 3.1.2. POR TRATAMENTO

Os valores para análise descritiva discriminada por tratamento se encontram na Tabela 3.

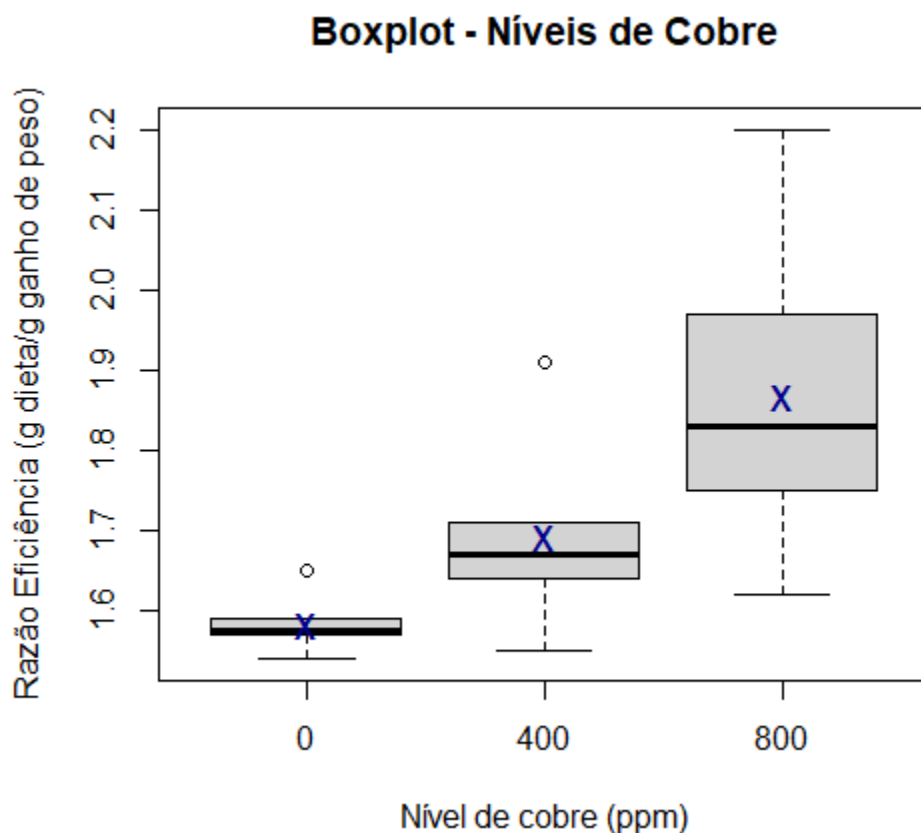
**Tabela 3:** Estatística descritiva dos dados da razão da eficiência da dieta (g dieta / g ganho de peso), por tratamento.

Parâmetros estatísticos	Razão Eficiência Dieta (g dieta/g ganho de peso) para os Níveis de Cobre (ppm)		
	Tratamentos		
	0	400	800
Soma	9,50	10,15	11,20
Mín.	1,540	1,550	1,620
1º Quartil	1,570	1,647	1,758
Mediana	1,575	1,670	1,830
Média	1,583	1,692	1,867
3º Quartil	1,587	1,700	1,948
Máx.	1,650	1,910	2,200
Amplitude	0,11	0,36	0,58
Variância	0,0013	0,0143	0,0408
Desvio Padrão	0,0367	0,1197	0,2020
Coef. Variação (%)	2,318	7,078	10,819
Coef. Assimetria	0,933	0,946	0,556

Em relação as médias, o maior valor ocorreu no tratamento de 800ppm (1,867 g/g), seguido pelo 400ppm (1,670 g/g) e por fim o tratamento com 0ppm (1,583 g/g). Apesar disso, os valores de desvio padrão seguiram também a sequência, isto é, a maior média apresentou o valor mais alto de desvio padrão (0,2020), seguido por 0,1197 em 400 ppm e 0,0367 para o tratamento 0ppm. Os valores máximos e mínimos também seguiram esse padrão, sendo o máximo 2,200 do tratamento 800ppm e o mínimo 1,540 referente a 0ppm.

Os três tratamentos tiveram coeficientes de variação baixos, um indício de homogeneidade dos dados, isto é, pouca dispersão.

**Figura 5.** Gráfico Boxplot separado por tratamento.



Na Figura 5, encontram-se os gráficos boxplot de cada nível do tratamento, onde se nota visualmente os apontamentos feitos acima. A dispersão maior dos dados nos níveis maiores de cobre. Também, destaca-se a presença de possíveis pontos discrepantes nos tratamentos de 0 e 400 ppm.

### 3.2. ANÁLISE DAS SUPOSIÇÕES DO MODELO

Tendo em vista ser um modelo de efeito fixo, têm-se como modelo estatístico o seguinte:

$$Y_{ij} = m + t_i + e_{ij}$$

onde:

$i$  = tratamentos;

$j$  = repetições;

$Y_{ij}$ : Razão da eficiência da dieta de pintinhos (g dieta / g ganho de peso) a  $i$ -ésima dose e  $j$ -ésima repetição;

$m$ : Média geral;

$t_i$ : Efeito da  $i$ -ésima dose;

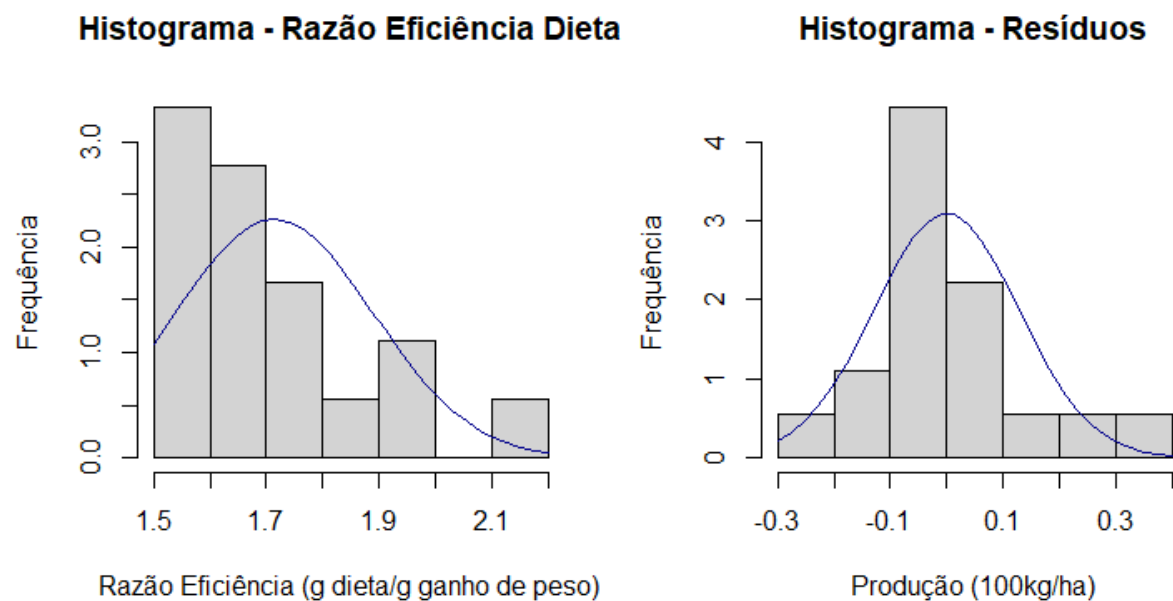
$e_{ij}$ : Erro experimental associado a  $Y_{ij}$ .

Sendo assim, as suposições associadas ao modelo são: linearidade e aditividade, independência dos erros, normalidade dos dados e homocedasticidade das variâncias.

### 3.2.1 NORMALIDADE

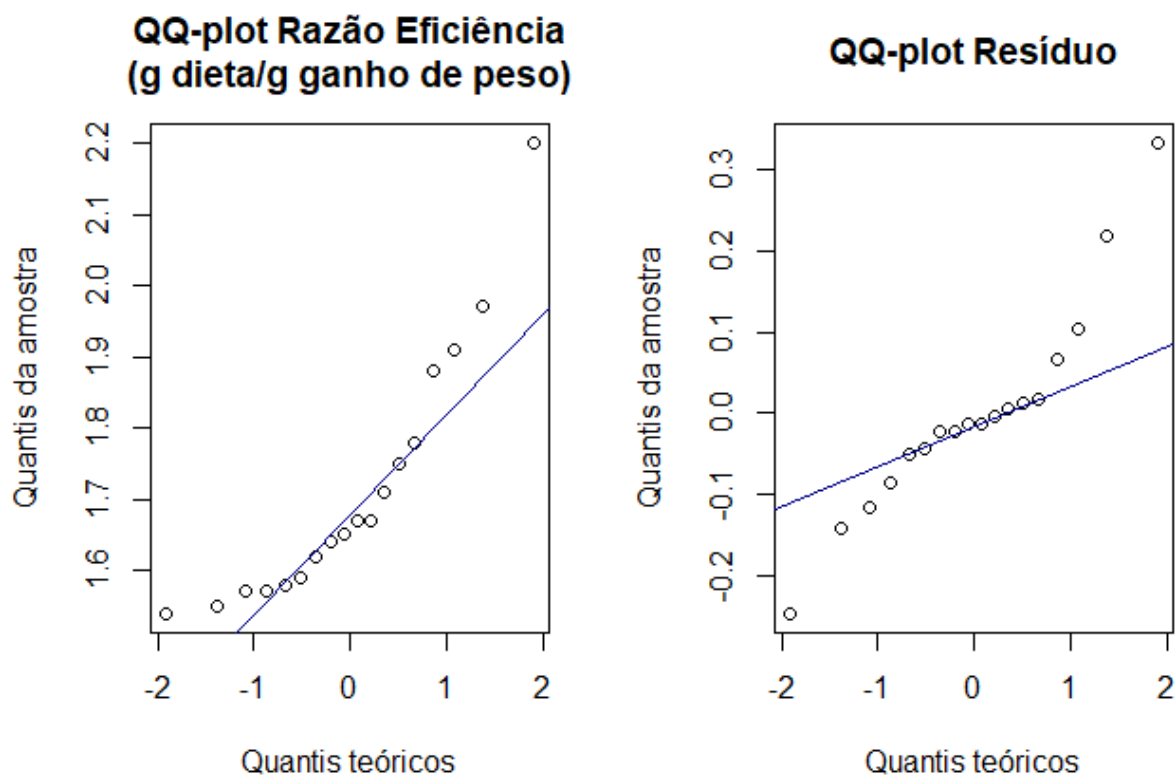
Na Figura 6, observam-se os histogramas da distribuição dos dados e dos resíduos. Busca-se nesses gráficos avaliar a similaridade da curva com o formato de sino da distribuição normal. Com a curva, nota-se, visualmente, que os valores da variável resposta se encontram um pouco assimétrico, com concentração à esquerda e cauda mais alongada à direita. Porém, realizando a análise dos resíduos, nota-se uma distribuição mais simétrica dos valores, o que pode inferir subjetivamente a normalidade dos dados.

**Figura 6.** Gráficos Histograma da variável resposta e resíduos com curvas normais.



Os gráficos QQ-plot sugerem a linearidade dos dados e a dispersão dos mesmos em torno da reta. Visualmente, os valores do resíduo se mostram ligeiramente menos ajustados à reta, em comparação com os da razão da eficiência da dieta. Porém ambos apresentam um comportamento semelhante de dispersão.

**Figura 7.** Gráficos QQ-plot da variável resposta e resíduos.



Contudo, para uma análise mais objetiva, realizou-se o teste não paramétrico para distribuição normal de Shapiro-Wilk. Sendo as hipóteses testadas:

**H0:** a produção segue a distribuição normal

**H1:** a produtividade não segue a distribuição normal

**Tabela 7:** Teste de normalidade de Shapiro-Wilk para os dados da dieta

	<b>W<sub>calculado</sub></b>	<b>p-valor</b>
<b>Razão Eficiência Dieta</b>	0,848	0,008
<b>Resíduos</b>	0,917	0,115

Dada a tabela 7, temos que o p-valor obtido para a Razão de Eficiência da Dieta  $0,008 < 0,05$ , rejeita-se assim H0 ao nível de 5% de significância, não podendo afirmar estatisticamente que os dados seguem distribuição normal. Em contrapartida, para os dados de resíduos, o p-valor  $0,115 > 0,05$ , aceitando-se H0, podendo dizer que os resíduos seguem em distribuição normal. Assim, considera-se que os dados atendem o pressuposto de Normalidade.

### 3.2.2 HOMOGENEIDADE DAS VARIÂNCIAS

A fim de verificar a homocedasticidade dos dados, realizou-se o teste de Levene com as seguintes hipóteses:

H0: existe homogeneidade entre as doses de adubação fosfatadas e o tipo de aplicação.

H1: Não existe homogeneidade entre as doses de adubação fosfatadas e o tipo de aplicação.

**Tabela 8:** Teste de Levene para homogeneidade das variâncias a 5% de significância.

	Graus de liberdade	Fvalor	Pr(>F)
Tratamento	2	2,952	0,083 .

. Significante a 0,1

De acordo com os dados dispostos na Tabela 8, a significância do Pr(>F) valor não foi maior que 5%, o que não permite concluir com esse grau de significância a Homocedasticidade dos dados, levando a rejeição de H0. Porém, de acordo com o resultado da tabela, esse valor seria estatisticamente significativo a 10%. Tendo isso em vista, e a normalidade dos resíduos, proceder-se-á a análise dos dados, visando a correlação linear dos mesmos, uma vez que para isso a principal análise é de normalidade dos resíduos.

### 3.3. ANÁLISE DE VARIÂNCIA

Sendo as hipóteses de teste para a análise ANOVA:

H0: os tratamentos são iguais a 0, ou seja, nenhuma dose de cobre exerce influência sobre a variável resposta.

H1: Pelo menos um tratamento é diferente de 0, ou seja, pelo menos uma dose de cobre exerce influência sobre a variável resposta.

**Tabela 7.** Análise de variância (ANOVA).

Fonte de variação	GL	SQ	QM	F	p-valor
Tratamento	2	0,2453	0,1226	6,515	0,009**
Resíduos	15	0,2824	0,0188	-	-
Total	17	0,5276	-	-	-

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

CV = 8,01 %



Com o p-valor encontrado  $0,009 < 0,05$ , rejeita-se  $H_0$  e admite-se, estatisticamente, com 5% de significância, que pelo menos uma dose de cobre exerce influência na razão da eficiência da dieta (g/g) dos pintinhos.

### 3.4. CONSIDERANDO A VARIÁVEL “NÍVEL DE COBRE” QUALITATIVA

Para essa análise, visto que a ANOVA deu resultado significativo, e que o teste de Shapiro-Wilk apontou normalidade dos resíduos, ambos a 5%, procede-se então o teste de comparação de médias.

#### 3.4.1 TESTE DE COMPARAÇÃO DE MÉDIAS

Fazendo uso do teste Tukey ao nível de 5% de significância no software R, aplica-se as seguintes hipóteses para comparar as médias dos tratamentos:

**H<sub>0</sub>:**  $m_u = m_v$  - Todas as médias dos tratamentos são estatisticamente iguais.

**H<sub>1</sub>:**  $m_u \neq m_v$  - Pelo menos uma das médias dos tratamentos é diferente.

**Tabela 8.** Teste Tukey a 5% de significância para comparação de médias.

Tratamentos (Nível de Cobre)	Médias	Teste Tukey (5%)
800	1,867	a
400	1,692	ab
0	1,583	b

a,b: Letras diferentes correspondem a médias estatisticamente diferentes a 5% de significância

O nível 800ppm teve a maior média e diferiu estatisticamente do nível 0ppm, que teve a menor média. O nível de 400ppm não diferiu estatisticamente dos demais, com um grau de 5% de significância.

### 3.5. CONSIDERANDO A VARIÁVEL “NÍVEL DE COBRE” QUANTITATIVA

Fazendo a consideração quantitativa da variável dos tratamentos, é possível então realizar a estatística de regressão dos dados, através do ajuste de modelos polinomiais.

#### 3.5.1 ANÁLISE DE REGRESSÃO LINEAR

Aplicando-se o teste de correlação linear de Pearson (r), chega-se no valor do coeficiente de 0,675, cujo indica uma correlação linear moderada positiva, de forma que com o aumento das doses de cobre, aumenta-se a razão da eficiência da dieta até um valor máximo.

### 3.5.1.1 MODELO LINEAR

A análise de variância para o modelo linear considera as hipóteses:

**H0:**  $\beta_1 = 0$  – O modelo não é adequado.

**H1:**  $\beta_1 \neq 0$  – O modelo é adequado.

**Tabela 9.** Análise de variância (ANOVA).

Fonte de variação	GL	SQ	QM	F	p-valor
Regressão Linear	1	0,2408	0,2408	12,79	0,003**
Falta de ajuste	1	0,0044	0,0044	0,24	0,634
Resíduos	15	0,2824	0,0188	-	-
Total	17	0,5276	-	-	-

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

Como o p-valor da regressão linear  $0,003 < 0,05$ , rejeita-se H0 e com pelo menos 5% de significância, conclui-se que o modelo linear é adequado para explicar a influência do nível de cobre na variável resposta.

**Tabela 10.** Estimativa dos parâmetros do modelo linear.

	ESTIMADO	ERRO	TC	p-valor
$\beta_0$	1,57222	0,05113	30,749	>0,00000
$\beta_1$	0,00035	0,00010	3,577	0,00275

Através da Tabela 10 se obtém os coeficientes do modelo linear, sendo o coeficiente Linear ( $B_0$ )=1,57222, que diz respeito ao valor estimado quando a variável  $x=0$ , e o coeficiente Angular ( $B_1$ )=0,00035, que corresponde ao acréscimo que a variável resposta (razão da eficiência da dieta) apresenta com o aumento de “x” (nível de cobre). Tendo esses valores, determina-se a equação linear do modelo:

$$Y = 1,57222 + 0,00035x$$

O coeficiente de determinação (explicação,  $R^2$ ), cujo indica a porcentagem da variação que é explicada pelo modelo em questão, obtido através do software R foi de  $R^2 = 0,9819$ , indicando que mais de 98% dos valores podem ser representados pelo modelo linear com os coeficientes obtidos acima.

### 3.5.1.2 MODELO QUADRÁTICO

A análise de variância para o modelo quadrático considera as hipóteses:

**H0:**  $\beta_2 = 0$  – O modelo não é adequado.

**H1:**  $\beta_2 \neq 0$  – O modelo é adequado.

**Tabela 9.** Análise de variância (ANOVA).

Fonte de variação	GL	SQ	QM	F	p-valor
Regressão Linear	1	0,24083	0,24083	12,79	0,003**
Regressão Quadrática	1	0,00444	0,00444	0,24	0,634
Falta de ajuste	0	0,00000	0	0	1
Resíduos	15	0,28253	0,01882	-	-
Total	17	0,5276	-	-	-

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

Como o p-valor da regressão quadrática  $0,634 > 0,05$ , não se rejeita H0 e com pelo menos 5% de significância, conclui-se que o modelo quadrático não é adequado para explicar a influência do nível de cobre na variável resposta.

### 3.5.1.3 MODELO CÚBICO

Visto que o Modelo Quadrático não foi adequado para explicar a variável resposta mas que não restou Falta de ajuste, o modelo de regressão cúbico não chegou a ser calculado pelo software R.

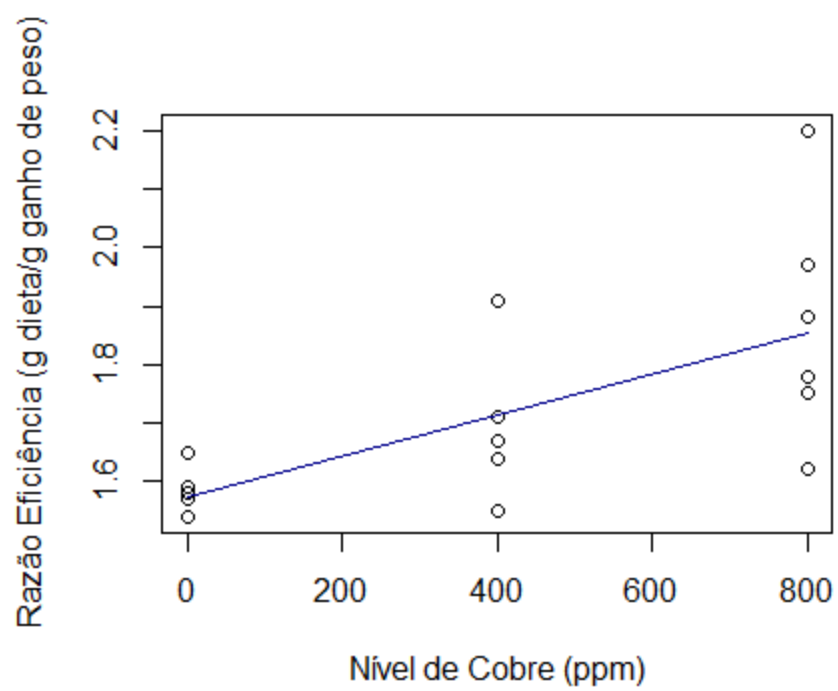
### 3.5.1.4 MODELO ADOTADO

Para dos dados em questão, de acordo com os modelos de regressões polinomiais calculados, o modelo de regressão linear simples se apresentou apropriado para explicar a influência das doses de cobre (em ppm) na razão da eficiência da dieta dos pintinhos, com uma correlação de 98%, sendo este um bom valor (não necessitando de modelos mais complexos).

A equação linear adotada então foi:

$$Y = 1,57222 + 0,00035x$$

**Figura 8.** Gráfico dos dados da variável resposta e preditora com a reta da equação linear ajustada.



## ANEXO 2 – SCRIPT UTILIZADO NO SOFTWARE R PARA QUESTÃO 3

```
dados <- read.table("dados-regressão.txt", header=T)
```

```
dados
```

```
attach(dados)
```

```
##### ANÁLISE DESCRITIVA #####
```

```
### GERAL ###
```

```
summary(razao)
```

```
var(razao)
```

```
sd(razao)
```

```
(sd(razao)/mean(razao))*100 #coeficiente de variação#
```

```
sum(razao)
```

```
max(razao)-min(razao) #amplitude#
```

```
require(moments)
```

```
skewness(razao) #assimetria#
```

```
kurtosis(razao) #curtose#
```

```
# HISTOGRAMA #
```

```
hist(razao, main = "Histograma", xlab = "Razão Eficiência (g dieta/g ganho de peso)",  
ylab = "Frequência", col=c("lightgray"))
```

```
# BOX PLOT #
```

```
média = mean(razao)
boxplot(razao, main = "Boxplot", ylab = "Razão Eficiência (g dieta/g ganho de peso)",
col=c("lightgray"))
points(média, pch='x', cex=1.5, col='darkblue')
```

```
### POR TRATAMENTO ###
```

```
tapply(razao, cobre, summary)
tapply(razao, cobre, var)
tapply(razao, cobre, sd)
(tapply(razao, cobre, sd)/tapply(razao, cobre, mean))*100 #coeficiente de variação#
tapply(razao, cobre, sum)
tapply(razao, cobre, max)-tapply(razao, cobre, min) #amplitude#
```

```
require(moments)
```

```
tapply(razao, cobre, skewness) #assimetria#
tapply(razao, cobre, kurtosis) #curtose#
```

```
# BOX PLOT #
```

```
médias = tapply(razao, cobre, mean)
boxplot(razao~cobre, main = "Boxplot - Níveis de Cobre", xlab = "Nível de cobre (ppm)",
ylab = "Razão Eficiência (g dieta/g ganho de peso)", col=c("lightgray"))
```

```
points(médias, pch='x', cex=1.5, col='darkblue')
```

```
##### ANÁLISE DAS SUPOSIÇÕES DO MODELO #####
```

```
cobreF <- as.factor(dados$cobre)
```

```
anvres <- aov(razao ~ cobreF)
```

```
anova(anvres)
```

```
### Normalidade ###
```

```
# Histograma com curva normal #
```

```
par(mfrow = c(1,2))
```

```
x <- seq(1.5, 2.2, length.out = 30)
```

```
d <- dnorm(x, mean=mean(razao), sd=sd(razao))
```

```
hist(razao, main = "Histograma - Razão Eficiência Dieta", xlab = "Razão Eficiência (g  
dieta/g ganho de peso)",
```

```
ylab = "Frequência", col=c("lightgray"), prob=T)
```

```
lines (x, d, col = "darkblue")
```

```
xres <- seq(-0.3, 0.4, length.out = 30)
```

```
dres <- dnorm(xres, mean=mean(anvres$residuals), sd=sd(anvres$residuals))
```

```
hist(anvres$residuals, main = "Histograma - Resíduos", xlab = "Produção (100kg/ha)",  
ylab = "Frequência",
```

```
col=c("lightgray"), prob=T)
```

```
lines (xres, dres, col = "darkblue")
```

```
# Gráficos QQ-plot #
```

```
par(mfrow = c(1,2))
```

```
qqnorm(razao, main = "QQ-plot Razão Eficiência  
(g dieta/g ganho de peso)",  
ylab="Quantis da amostra", xlab="Quantis teóricos")  
qqline(razao, col = "darkblue")
```

```
qqnorm(anvres$residuals, main = "QQ-plot Resíduo",  
ylab="Quantis da amostra", xlab="Quantis teóricos")  
qqline(anvres$residuals, col = "darkblue")
```

```
# teste de Shapiro-Wilk #
```

```
shapiro.test(razao)  
shapiro.test(anvres$residuals)
```

```
### homogeneidade das variâncias ###
```

```
# teste estatístico #
```

```
require(car)  
leveneTest(razao~cobreF)
```

```
##### ANÁLISE DE VARIÂNCIA #####
```



##### ANÁLISE DE REGRESSÃO #####

```
require(ExpDes)
```

```
## considerando NÍVEL DE COBRE como variável qualitativa ##
```

```
crd(cobre, razao, quali = T, sigF = 0.05, sigT = 0.05)
```

```
## considerando x como variável quantitativa ##
```

```
## R2 = SQReg/SQTrat ##
```

```
## SQ lack of fit(falta de ajuste) = SQTrat - SQReg ##
```

```
crd(cobre, razao, quali = F, sigF = 0.05, sigT = 0.05) ## só está implementada até grau 3 ##
```

```
## LINEAR ##
```

```
## estimativa do modelo
```

```
lin <- lm(razao ~ cobre)
```

```
lin
```

```
## curva do modelo estimado
```

```
plot(cobre, razao, xlab = "Nível de Cobre (ppm)",
```

```
ylab = "Razão Eficiência (g dieta/g ganho de peso)")
```

```
lines(cobre, lin$fi, col="darkblue")
```