

Integrantes da Dupla

Marcos Cavalcanti – 1920533

Carlos Santos – 1910425

PARTE I

`n <- 47` → Número de observações da amostra

`p <- 6` → Número de parâmetros do modelo

`y <- swiss$Fertility`

`agr <- swiss$Agriculture`

`exm <- swiss$Examination`

`edc <- swiss$Education`

`cth <- swiss$Catholic`

`imt <- swiss$Infant.Mortality`

`summary(lm(y ~ agr + exm + edc + cth + imt))`

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	66.91518	10.70604	6.250	1.91e-07	***
agr	-0.17211	0.07030	-2.448	0.01873	*
exm	-0.25801	0.25388	-1.016	0.31546	
edc	-0.87094	0.18303	-4.758	2.43e-05	***
cth	0.10412	0.03526	2.953	0.00519	**
imt	1.07705	0.38172	2.822	0.00734	**

`t_icpt <- abs(6.250)`

`t_exm <- abs(-1.016)`

`t_cth <- abs(2.953)`

`t_agr <- abs(-2.448)`

`t_edc <- abs(-4.758)`

`t_imt <- abs(2.822)`

`alfa <- 0.005` → Escolhemos um alfa bastante conservador, de 0,5%.

`t <- abs(qt(alfa/2,n-p))` → Agora vamos compará-lo com cada um dos 6 parâmetros do modelo.

`t < t_icpt`

`t < t_exm`

`t < t_cth`

`[1] TRUE`

`[1] FALSE`

`[1] FALSE`

`t < t_agr`

`t < t_edc`

`t < t_imt`

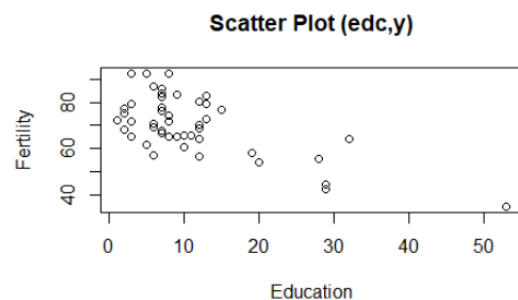
`[1] FALSE`

`[1] TRUE`

`[1] FALSE`

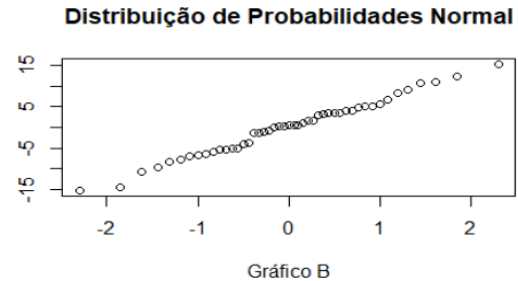
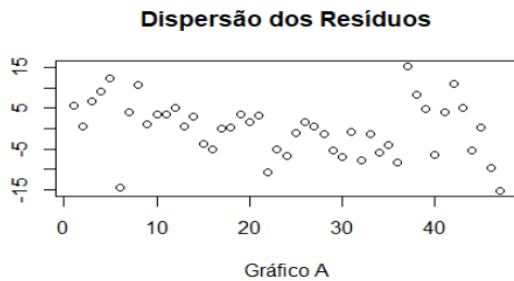
→ Interessante notar que quando optamos por um nível de significância bem restritivo, como o de 0,5%, apenas uma das variáveis regressoras se mostrou estatisticamente significativa. E, admiravelmente, essa variável foi o nível educacional da província.

`plot(edc,y,xlab = "Education",ylab = "Fertility", main = "Scatter Plot (edc,y)")`



➔ Agora vamos analisar a normalidade dos resíduos com um (a) gráfico de dispersão dos resíduos, (b) gráfico de probabilidade normal e (c) o nível de curtose da distribuição.

```
uhat <- residuals(lm(y ~ agr + exm + edc + cth + imt))
```



```
curtose <- function(distribuicao) {mean(distribuicao^4)/(mean(distribuicao^2)^2)}
```

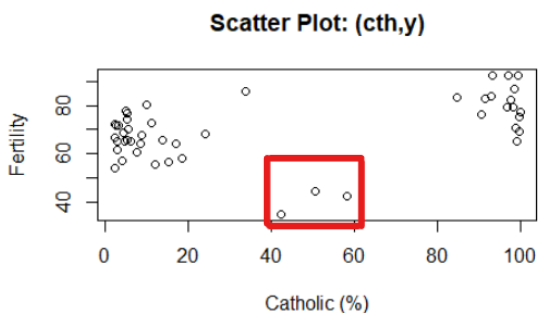
```
(c) curtose(uhat)
```

```
[1] 2.763938
```

➔ Os resíduos parecem seguir uma característica normal, porém não *muito* forte. A medida da curtose ficou dentro de um intervalo considerável para uma v.a normal, [2.6,3.2], mas não tão próximo do resultado idealmente esperado que é o número 3.

PARTE II

```
plot(cth,y,xlab = "Catholic (%)",ylab = "Fertility",main = "Scatter Plot: (cth,y)")
```



➔ “min”: população católica minoritária e “maj”: população católica majoritária

```
fator <- c("min")
```

```
for(i in 2:n) { If(cth[i] < 50.00) { fator[i] <- c("min") } else { fator[i] <- c("maj") } }
```

```
fator <- factor(fator) ; anova(lm(y~fator))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
fator	1	1164.7	1164.68	8.7159	0.004998 **
Residuals	45	6013.3	133.63		

➔ Sim, é estatisticamente significativo! Podemos notar um valor-p abaixo de 0,05, o que nos fornece uma confiança de no mínimo 95%.