

REGRESSÃO LINEAR MÚLTIPLA.

Um experimento para avaliar quais dos elementos químicos: $X_1=N$, $X_2=P$, $X_3=K$, $X_4=CA$, $X_5=MG$, $X_6=S$ influenciam na granulometria do solo, obteve-se os dados apresentados a seguir, onde $Y = \%$ terra retida na peneira 18.

Fazer a análise de regressão linear múltipla, fazendo a seleção do modelo pelo método Backward.

X1	X2	X3	X4	X5	X6	Y
2,92	0,53	1,40	0,06	0,12	0,21	17,54
2,75	0,52	1,45	0,08	0,13	0,22	5,08
2,78	0,51	1,52	0,09	0,13	0,20	6,90
2,73	0,52	1,53	0,09	0,13	0,21	19,34
2,45	0,47	1,55	0,14	0,14	0,20	4,57
2,89	0,48	1,49	0,12	0,14	0,20	7,28
2,98	0,53	1,52	0,07	0,12	0,22	9,40
2,78	0,51	1,42	0,07	0,13	0,20	15,66
2,32	0,48	1,50	0,10	0,13	0,20	10,76
2,87	0,49	1,51	0,08	0,13	0,21	13,94
2,51	0,48	1,51	0,11	0,14	0,19	6,28
2,58	0,50	1,42	0,08	0,13	0,22	13,28
2,78	0,48	1,46	0,08	0,12	0,21	14,86
2,80	0,51	1,51	0,08	0,13	0,21	15,68
3,01	0,54	1,49	0,10	0,14	0,22	6,80
3,05	0,53	1,53	0,07	0,12	0,21	18,68
3,04	0,52	1,46	0,09	0,13	0,20	5,47
3,13	0,51	1,49	0,09	0,13	0,20	8,76
3,10	0,50	1,42	0,07	0,12	0,20	9,68
3,18	0,50	1,45	0,07	0,12	0,21	11,47
3,37	0,46	1,47	0,12	0,13	0,20	11,41
3,30	0,50	1,48	0,11	0,12	0,22	9,38
3,35	0,51	1,42	0,09	0,13	0,20	5,10
3,45	0,47	1,49	0,12	0,13	0,20	15,01
3,46	0,51	1,45	0,07	0,12	0,16	27,93
3,48	0,49	1,45	0,09	0,14	0,19	5,85
3,59	0,48	1,53	0,13	0,14	0,18	10,38
4,58	0,51	1,44	0,07	0,13	0,21	12,57
3,51	0,49	1,47	0,09	0,14	0,21	3,43
3,56	0,52	1,47	0,10	0,14	0,21	6,13
3,74	0,51	1,44	0,08	0,13	0,20	7,46
3,71	0,45	1,50	0,08	0,13	0,21	7,18
3,95	0,47	1,47	0,12	0,13	0,20	10,59
4,00	0,52	1,46	0,09	0,13	0,21	13,96
3,90	0,51	1,51	0,11	0,13	0,20	15,47
3,96	0,45	1,41	0,12	0,13	0,20	6,48

```
#####
```

```
### ENTRADA DE DADOS
```

```
### DEFINA AS VARIÁVEIS COMO Y, X1, X2, ....
```

```
rlm <- read.table("C://GER//A7_RLM.TXT", h=T)
```

```
rlm
```

```
X1 <- rlm[,1]
```

```
X2 <- rlm[,2]
```

```
X3 <- rlm[,3]
```

```
X4 <- rlm[,4]
```

```
X5 <- rlm[,5]
```

```
X6 <- rlm[,6]
```

```
Y <- rlm[,7]
```

```
X1; X2; X3; X4; X5; X6; Y
```

```
### DEFINA O MODELO ###
```

```
mod <- lm(Y~X1+X2+X3+X4+X5+X6)
```

```
mod
```

```
### PLOTE A MATRIZ DE DISPERSÃO #####
```

```
install.packages("car")
```

```
require(car)
```

```
pairs(rlm)
```

```
plot(rlm)
```

```
#####
```

```
### CARREGUE A ROTINA PARA ANÁLISE DE DIAGNÓSTICO ###
```

```
### FAÇA OS DIAGNÓSTICOS, ELIMINE OBSERVAÇÕES, SE FOR O CASO.
```

```
##### ROTINA DIAGNÓSTICOS - REG. LIN. MULTIPLA #####
```

```
N=length(Y) # Tamanho da amostra
```

```
P=length(mod$coeffient) # numero de parametros
```

```
N; P
```

```
# Selecionando os vetores para analise de disgnostico (rs - Rstudent) e (h - influencia)
```

```
rs <- rstudent(mod)
```

```
yp <- predict.lm(mod)
```

```
h <- lm.influence(mod)$hat
```

```
lh <- 2*P/N
```

```
rs; yp; h; lh
```

```
### Gráficos de diagnósticos
```

```
par(mfrow=c(1,3))
```

```
hist(rs, main="histograma")
```

```
boxplot(rs, main="boxplot")
```

```
qqnorm(rs, main="normalidade"); qqline(rs)
```

```
### Gráfico para Análise do Residuo
par(mfrow=c(1,1))
plot(y,rs,main="Análise do Resíduo", xlab="Valores preditos", ylab="RStudent")
abline(h=0)
```

```
### Grafico para detectar dados influentes e outliers
minrs=min(rs,-3)
maxrs=max(rs,3)
ymin=minrs-.1
ymax=maxrs+.1
maxh=max(h,4/N)
minh=min(h)
xmin=minh-.1
xmax=maxh+.1
par(mfrow=c(1,1))
plot(c(xmin,xmax),c(ymin,ymax), type="n", xlab="h - leverage", ylab="RStudent")
abline(h=-3, col="red"); abline(h=3,col="red"); abline(v=lh, col="blue"); points(h,rs)
```

```
sh <- sort(h);sh
rlm <- rlm[-c(25,32),]
rlm
```

```
#####
```

```
### CARREGUE A ROTINA PARA SELEÇÃO DE MODELOS
# MÉTODO Cp ou MÉTODO MAXIMO R2 ou
FORWARD/BACKWARD/STEPWISE.
# FAÇA A SELEÇÃO DO MODELO
```

```
##### Selecao de Modelo - método Cp (Menor Cp - melhor o modelo) ###
# COMPLETE O cbind COM O TOTAL DE VARIÁVEIS ESPLANATÓRIAS
install.packages("leaps")
require(leaps)
selcp <- leaps(x=cbind(X1,X2,X3,X4,X5,X6), y=Y,method=c("Cp"))
nparcp <- selcp$size
Cp <- selcp$Cp
cbind(nparcp,Cp,selcp$which)
```

```
##### Selecao de Modelo - método R2 Ajustado (Maior R2 - melhor o modelo)
###
# COMPLETE O cbind COM O TOTAL DE VARIÁVEIS ESPLANATÓRIAS
install.packages("leaps")
require(leaps)
selr2 <- leaps(x=cbind(X1,X2,X3,X4,X5,X6), y=Y,method=c("adjr2"))
nparr2 <- selr2$size
r2.aj <- selr2$adjr2
```

```
cbind(nparr2,r2.aj,selr2$which)
```

```
##### Selecao de Modelo - método backward | forward | stepwise###  
# a seleção é feita pelo valor de AIC  
step(mod, direction="backward") # ou forward ou both (stepwise).  
#####
```

```
#####  
#### COMPLETE A ANÁLISE COM O MODELO SELECIONADO  
mods <- lm(Y~X3+X5)  
anvar <- anova(mods)  
anvar  
summary(mods)  
#####
```