

A large, light gray, 3D-style logo of the letter 'R' is centered in the background. The 'R' has a thick, rounded top and a vertical stem that curves into a horizontal bar at the bottom right. The text is overlaid on this logo.

BIE5782

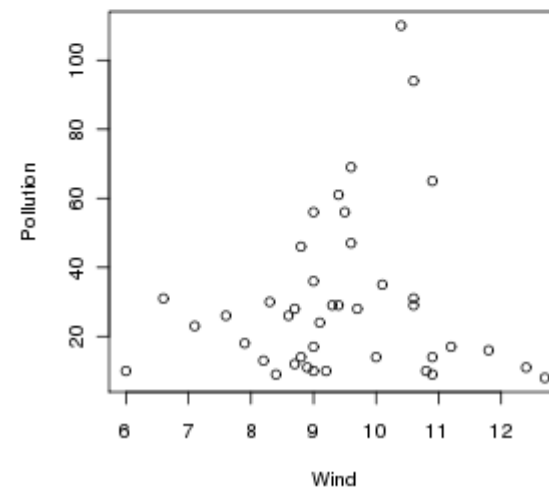
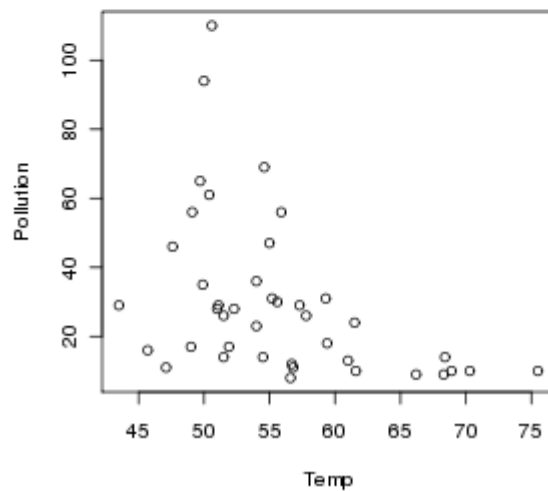
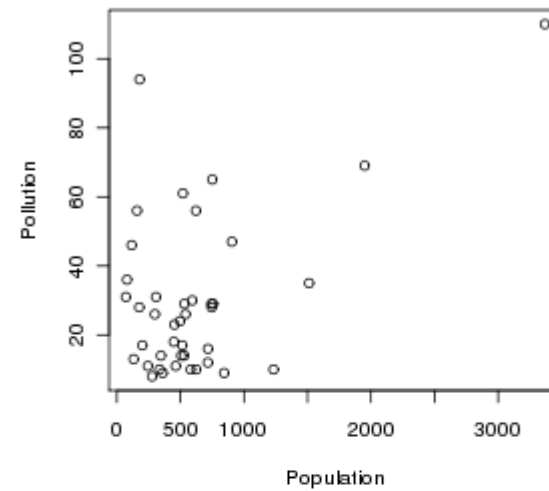
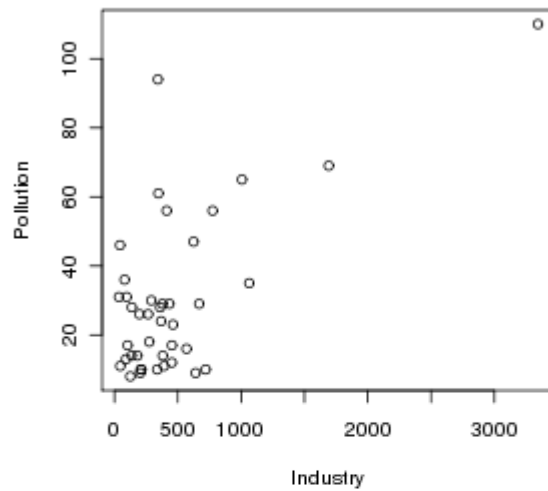
Unidade 7:

**REGRESSÃO LINEAR
(MULTIPLA)**

ROTEIRO

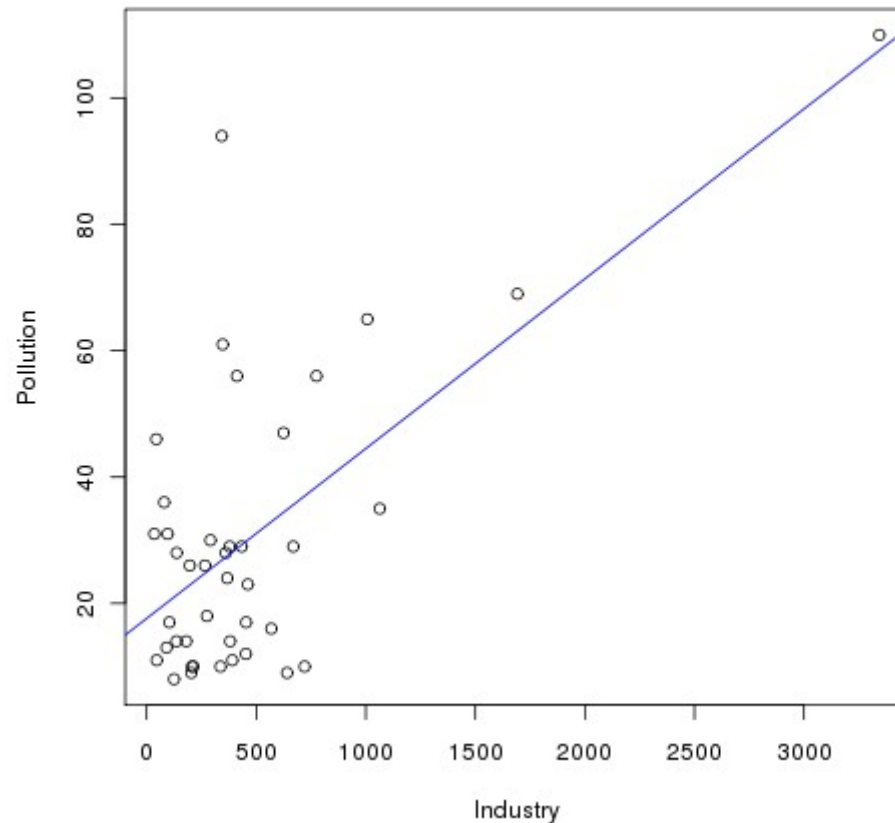
1. Motivação
2. Incluindo mais preditoras: ciclo de ajuste e avaliação
3. Modelos sem interação (comparação do intercepto)
4. Interação (comparação da inclinação)
5. Colinearidade
6. Matriz de modelos lineares e interpretação dos coeficientes

Regressão Linear Múltipla



lm()

Ajusta Modelo Linear Gaussiano



```
> pol.m1 <- lm(Pollution~Industry, data=poluicao)
> plot(Pollution~Industry, data=poluicao)
> abline(pol.m1, col="blue")
```

anova.lm()

Avalia o Modelo*

Analysis of Variance Table

Response: Pollution

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Industry	1	9173.0	9173.0	27.808	5.269e-06	***
Residuals	39	12864.9	329.9			

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

*** Uma avaliação comparativa!!!**



lm ➡ update ➡ anova

O Ciclo de Ajuste e Avaliação

```
> pol.m1 <- lm(Pollution~Industry, data=poluicao)
> pol.m2 <- update(pol.m1, .~. + Temp)
```

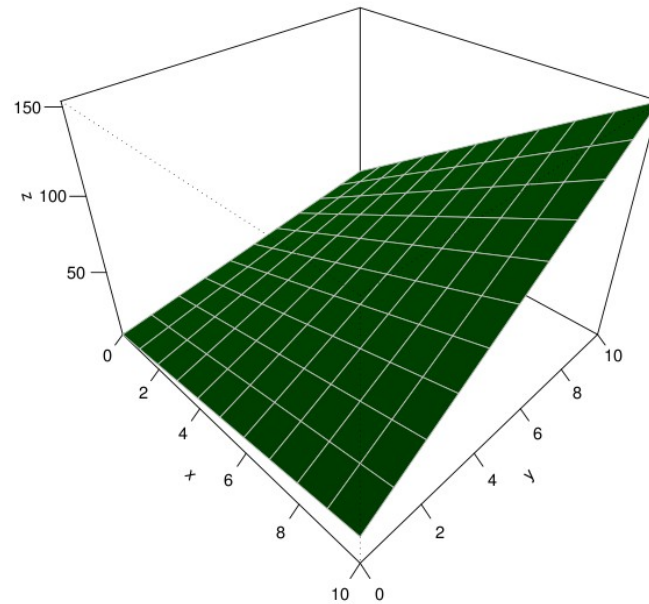
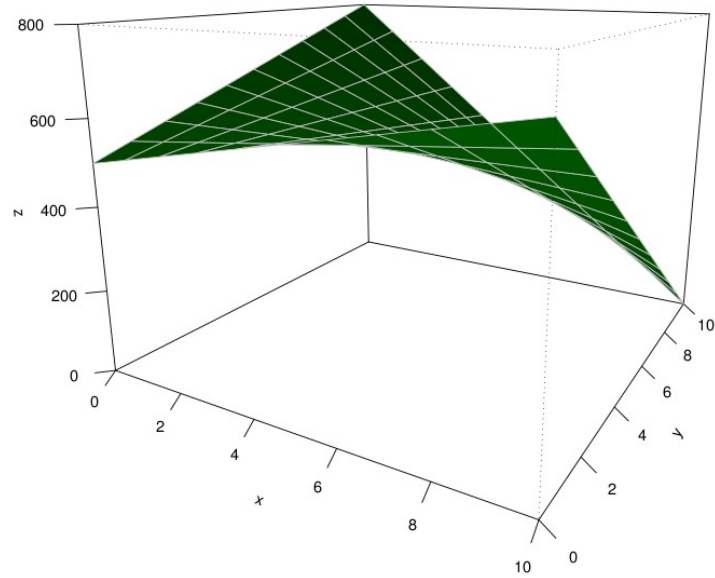
```
> anova(pol.m1, pol.m2)
Analysis of Variance Table
```

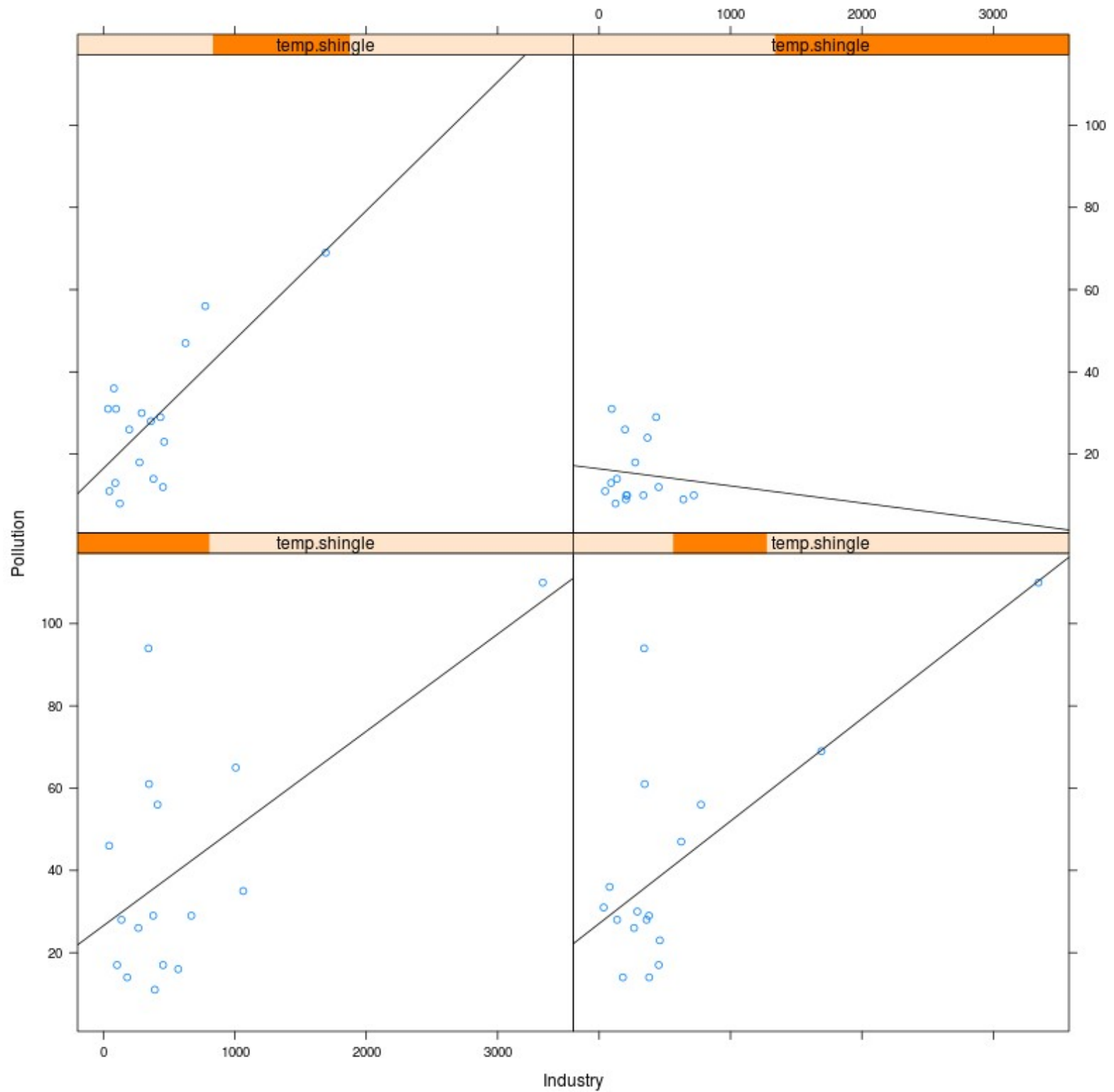
Model 1: Pollution ~ Industry

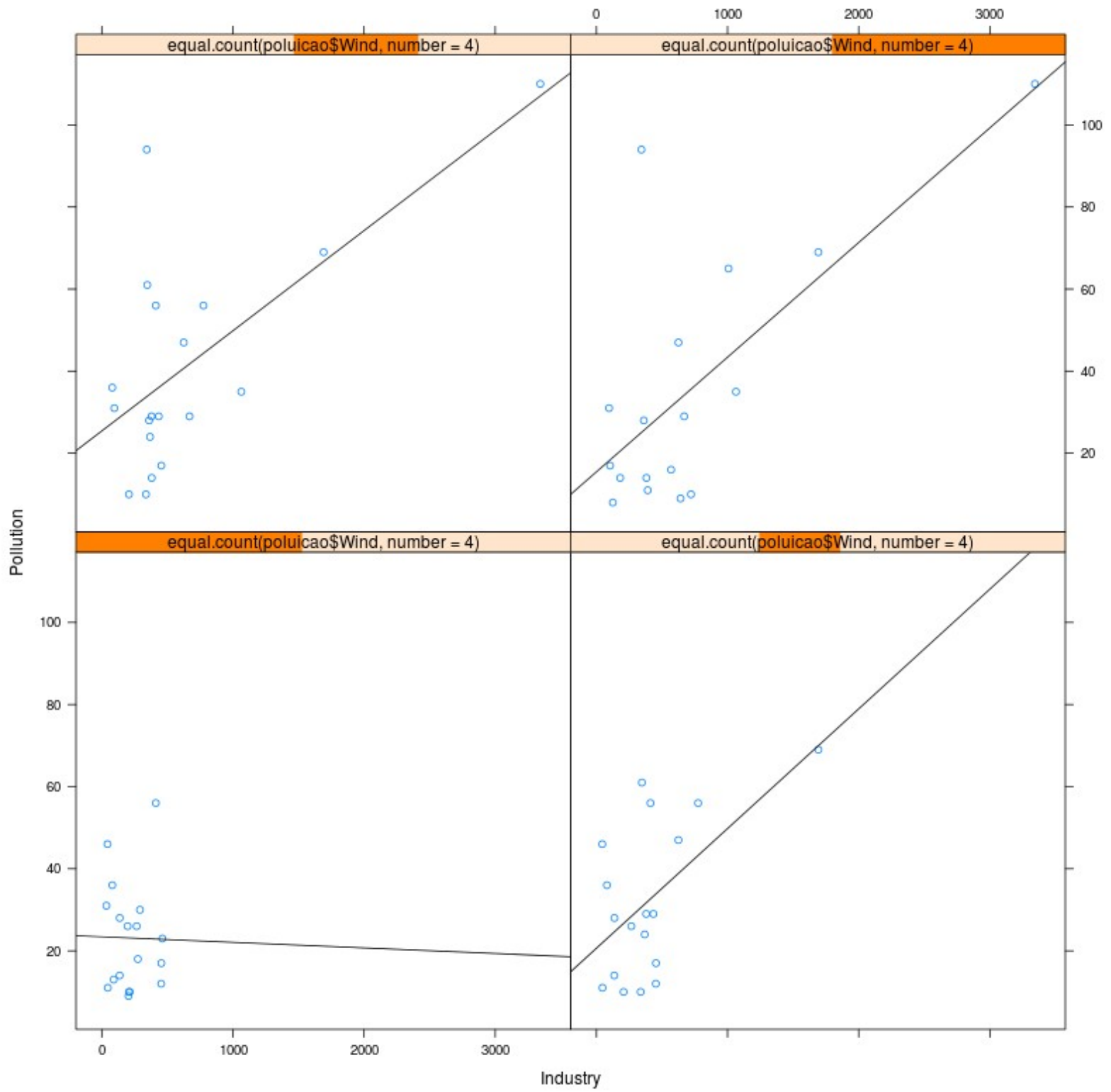
Model 2: Pollution ~ Industry + Temp

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)	
1	39	12864.9					
2	38	10635.6	1	2229.2	7.9649	0.007546	**

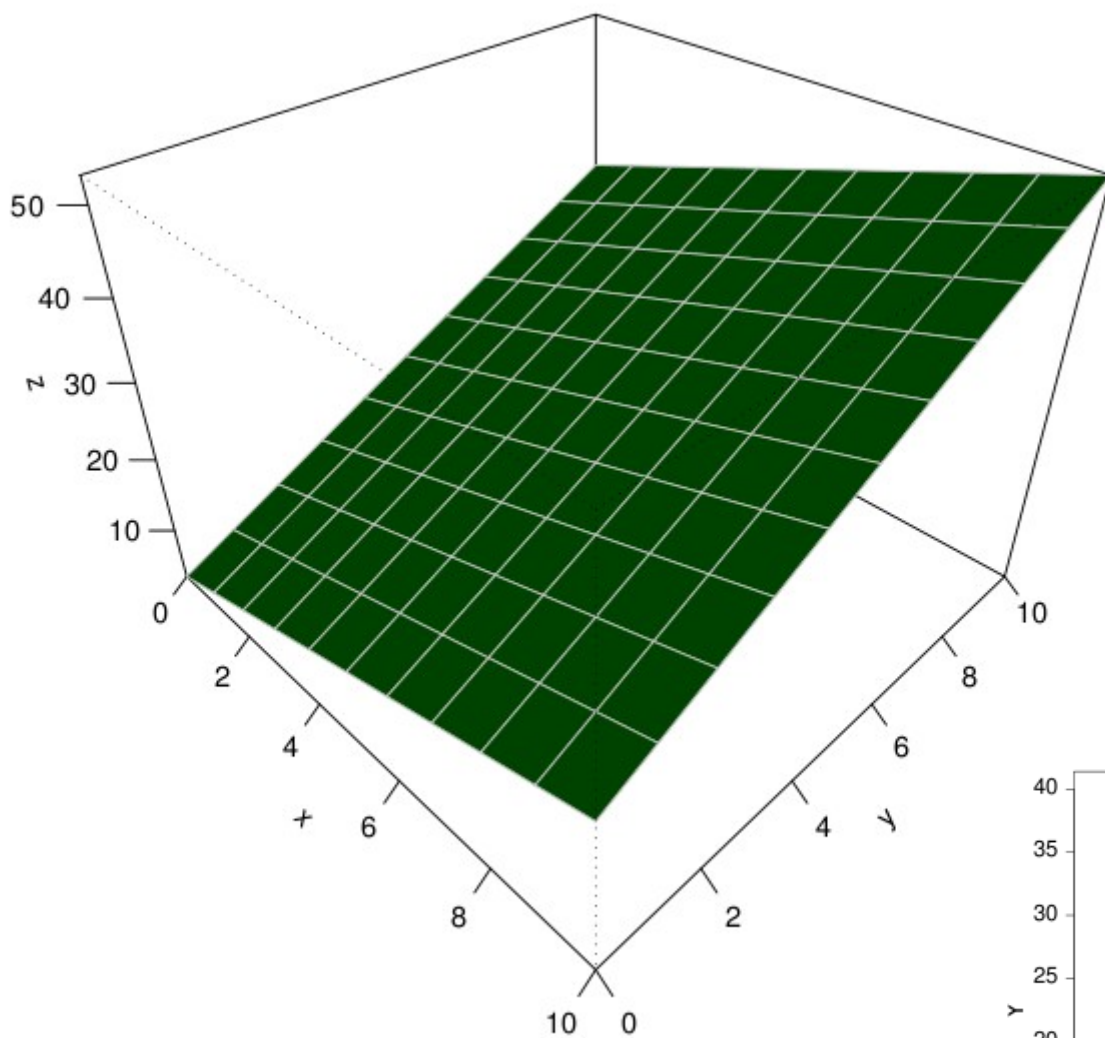
Interações



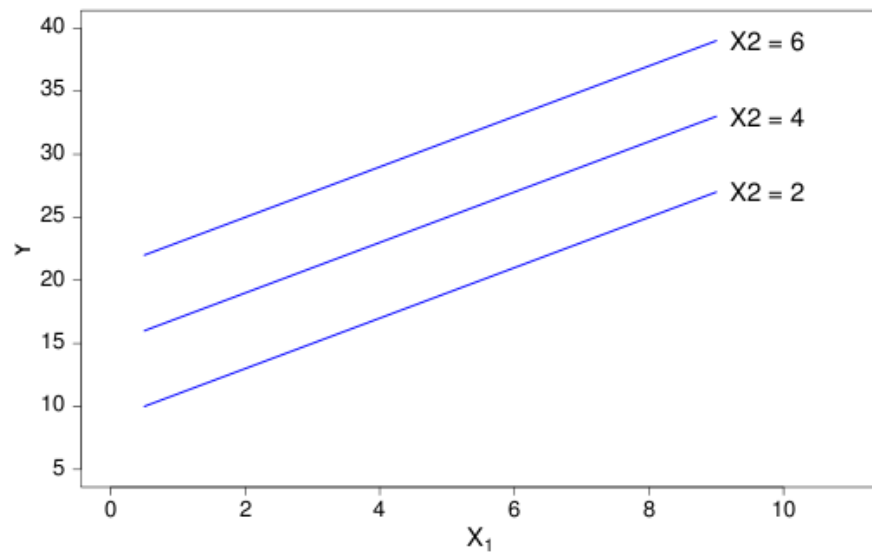




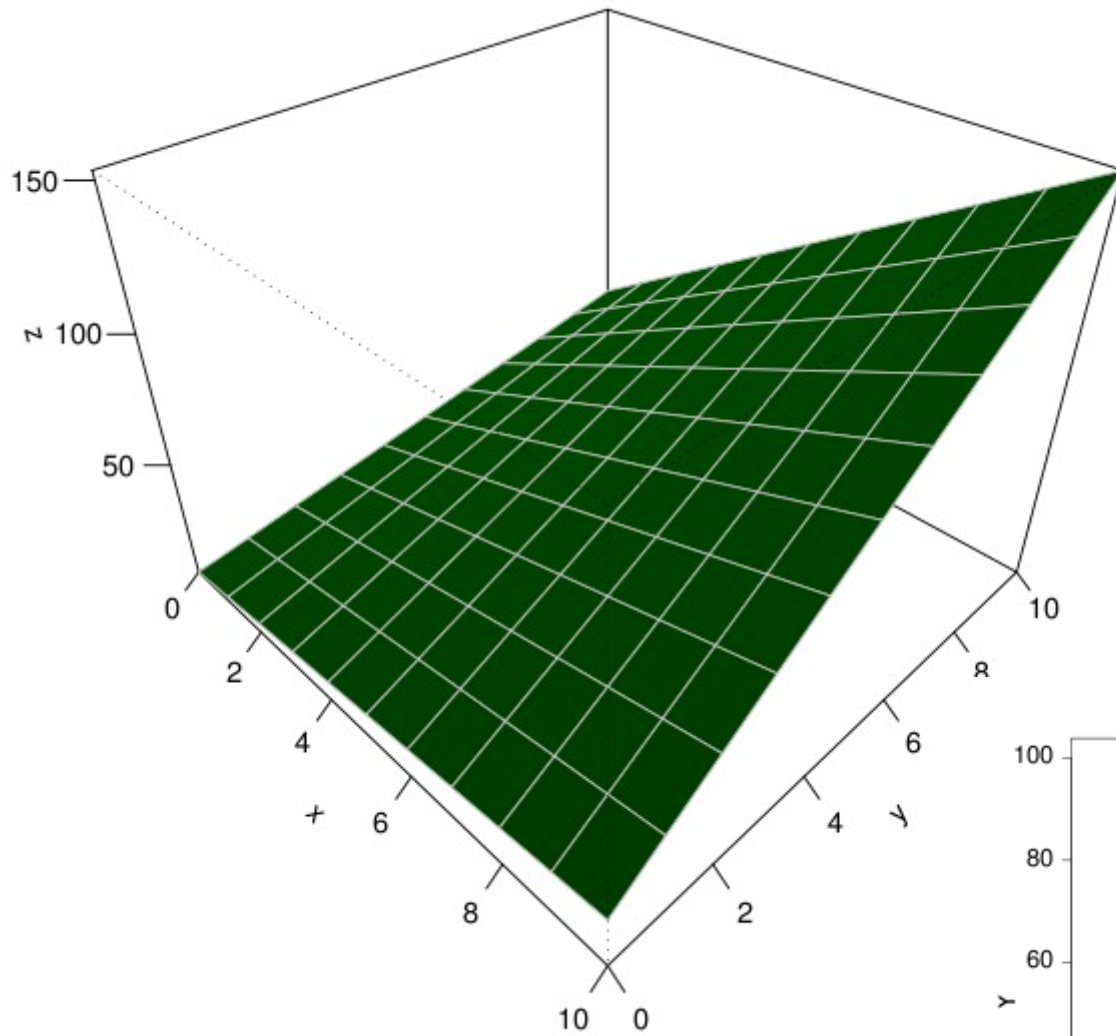
Modelo sem Interação



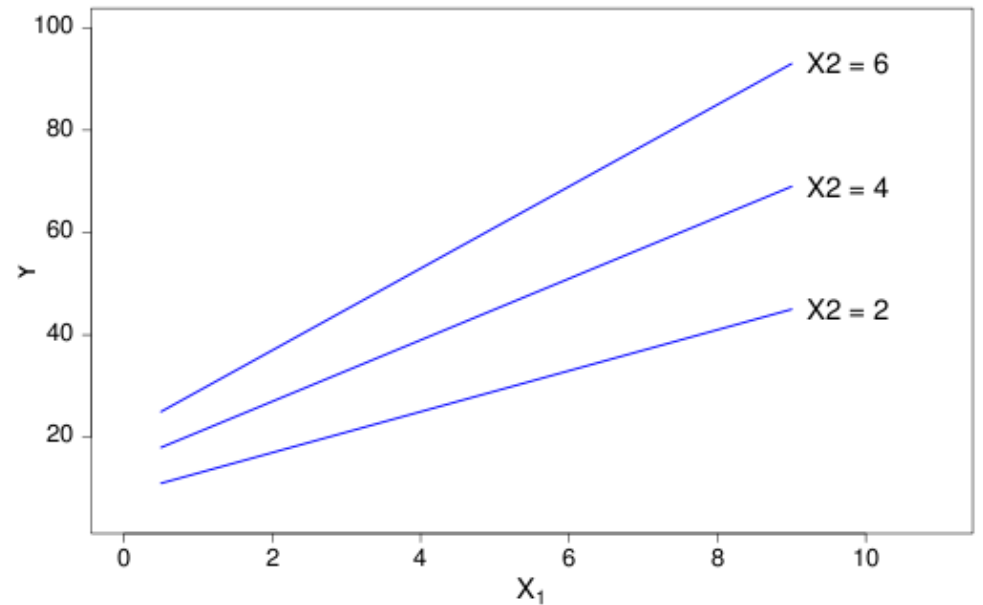
$$y \sim a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_i \cdot x_i$$



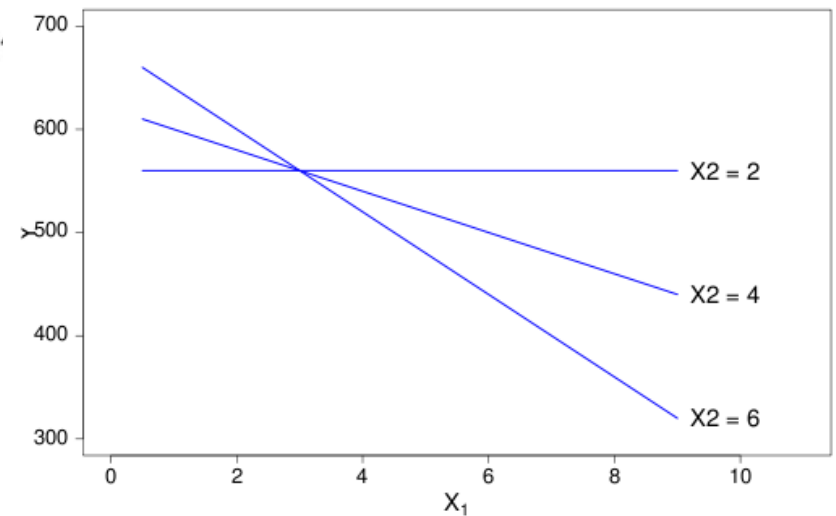
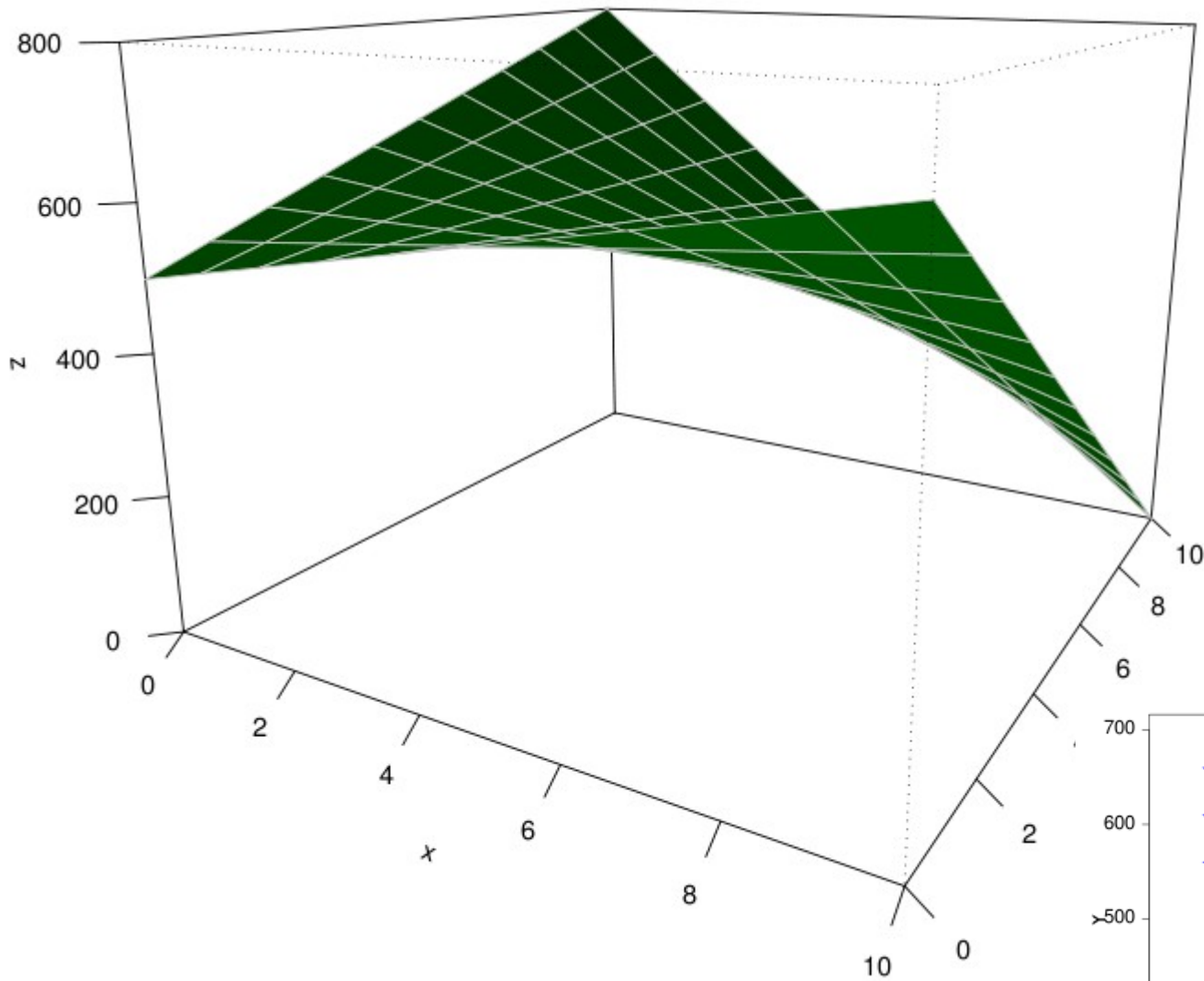
Modelo com Interação



$$y \sim a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_1 x_2 \dots$$



Modelo com Interação



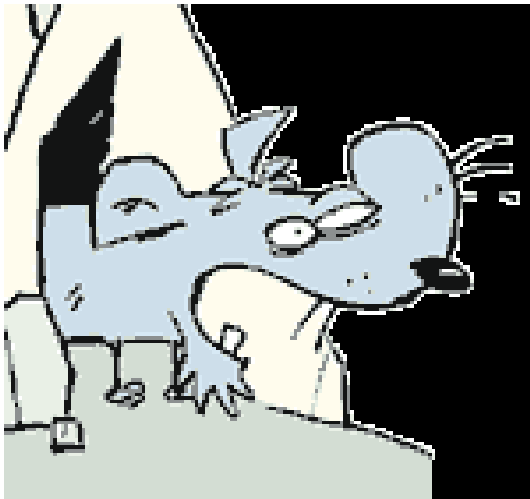
$$y \sim a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_1 x_2 \dots$$

Notação de Modelos no R

Símbolos utilizados nas Fórmulas Estatísticas para definir diferentes Delineamentos Experimentais

Expressão	Significado
$Y \sim X$	Modele Y como função estatística de X
$A + B$	inclui ambos os fatores A e B
$A - B$	inclui todos os efeitos em A , exceto os que estão em B
$A * B$	$A + B + A:B$
A / B	$A + B \%in\% (A)$ modelos hierárquicos
$A:B$	efeito da interação entre os fatores A e B
$B \%in\% A$	efeitos de B dentro dos níveis de A
A^m	todos os termos de A cruzados até à ordem m

INTERAÇÕES E COLINEARIDADE



O Fantasma da Colinearidade

1. Grandes mudanças nas estimativas dos coeficientes quando variáveis são incluídas ou excluídas do modelo.
2. Testes não significativos para variáveis consideradas importantes.
3. Estimativas dos coeficientes de regressão com sinal errado.
4. Alta correlação (duas-a-duas) entre as variáveis preditoras.

Efeitos da Colinearidade

- 3. Não é realista avaliar o efeito de uma variável preditora mantendo as demais constantes.**
- 4. A soma de quadrados extra de uma variável preditora depende das demais variáveis presentes no modelo.**
- 5. Testes t simultâneos são problemáticos;**
 - eles podem aceitar o conjunto de hipóteses $H_0 : \beta_k = 0$, quando o teste F pela SQ-extra rejeitaria o mesmo conjunto de H_0 's.

Efeitos da Colinearidade

1. Nenhum efeito sobre a possibilidade de se obter um bom ajuste global.

2. A variância amostral dos coeficientes de regressão é grande:

- Nenhum parâmetro é estatisticamente significativo (não se rejeita $H_0 : \beta_k = 0$);
- Mas o conjunto de variáveis preditoras é estatisticamente significativo (teste F do modelo é significativo).

A MATRIZ

(do modelo)

(Intercept)	Industry	Temp
1	1	368 61.5
2	1	291 55.6
3	1	775 55.9
4	1	137 51.0
5	1	136 68.4
6	1	44 47.6
7	1	641 66.2
8	1	1064 49.9
9	1	197 57.8
10	1	347 50.4
11	1	434 57.3



model.matrix()

A Matriz do Modelo

```
> names(pol.m2)
[1] "coefficients" "residuals" "effects" "rank"
[5] "fitted.values" "assign" "qr" "df.residual"
[9] "xlevels" "call" "terms" "model"
```



```
> formula(pol.m2)
Pollution ~ Industry + Temp
```

```
> model.matrix(pol.m2)
  (Intercept) Industry Temp
1           1         368 61.5
2           1         291 55.6
3           1         775 55.9
4           1         137 51.0
5           1         136 68.4
...

```

Cálculo Matricial dos Esperados

$$y = a.X$$

$$\begin{bmatrix} 1 & 368 & 61,5 \\ 1 & 991 & 77,5 \\ 1 & 775 & 51,1 \end{bmatrix} \times \begin{bmatrix} 77,4 \\ 0,02 \\ -1,05 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 \times 77,4 + 0,02 \times 368 - 1,05 \times 61,5 \\ 1 \times 77,4 + 0,02 \times 991 - 1,05 \times 77,5 \\ 1 \times 77,4 + 0,02 \times 775 - 1,05 \times 51,2 \end{bmatrix}$$

MATRIZ DO MODELO E VARIÁVEIS CATEGÓRICAS



FIM DA UNIDADE 7

Resto da tarde:

Plantão de Tutoriais e Exercícios