

A large, light gray, 3D-style logo consisting of the letters 'P' and 'R' intertwined. The 'P' is on the left and the 'R' is on the right, with the 'R' partially overlapping the 'P'. The letters have a slight shadow and a beveled edge, giving them a three-dimensional appearance.

**BIE5782**

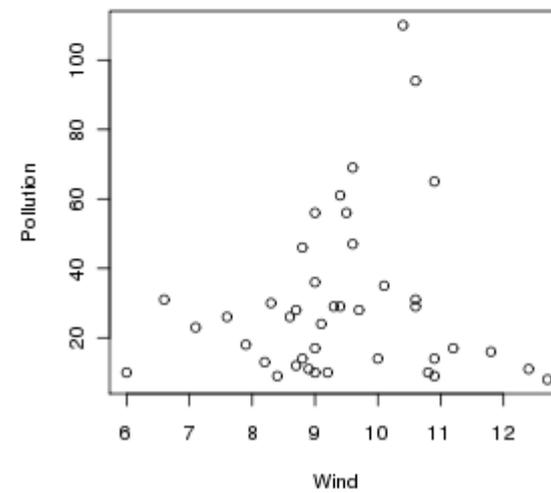
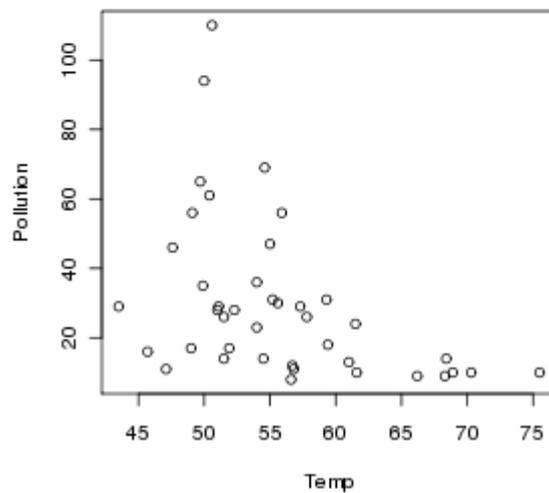
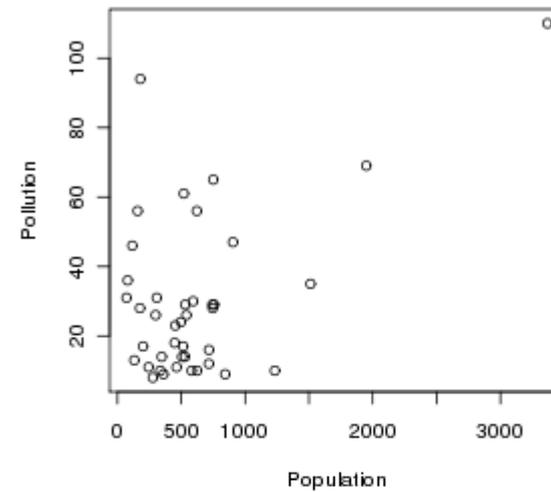
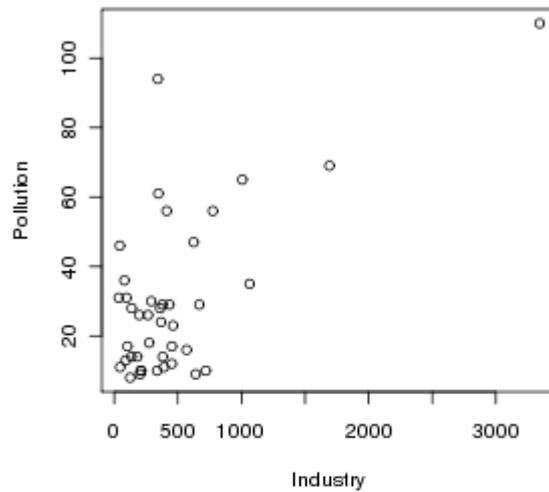
Aula 8:

**MODELOS LINEARES MÚLTIPLOS**

# Tópicos

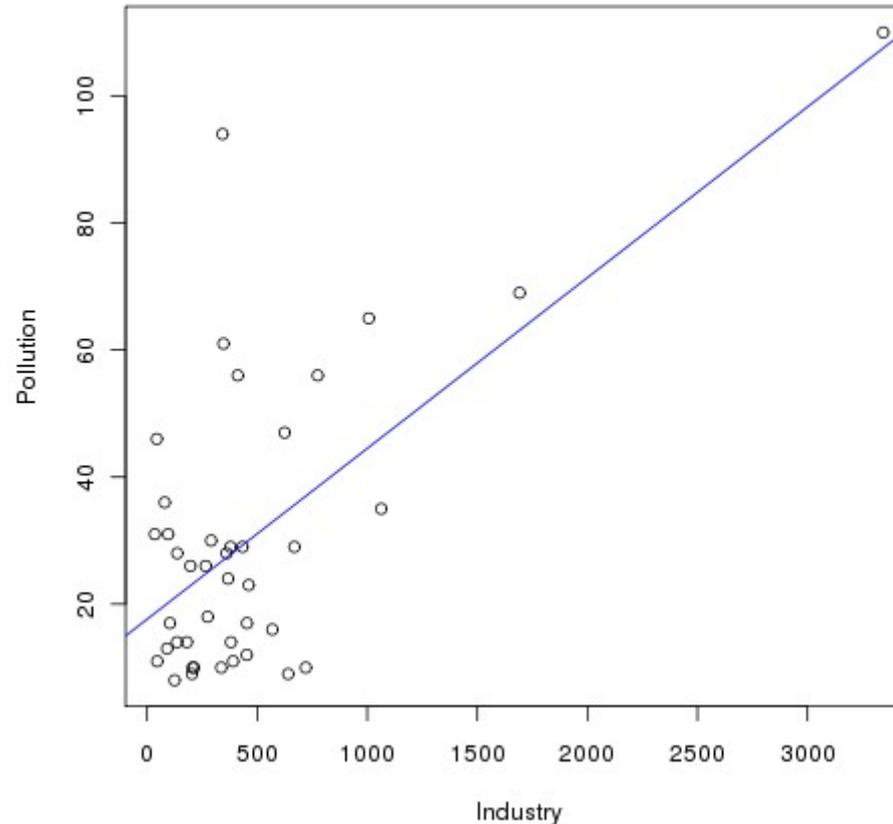
- Efeito de mais de uma preditora:
  - Como incluir no modelo
  - Como interpretar
  - Tipos de efeitos: aditivos e interações
  - Como avaliar sua importância para o modelo
  - Problemas: colinearidade
  - Notação matricial de modelos lineares

# 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!!!



O Ciclo de Ajuste e Avaliação

# 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
```

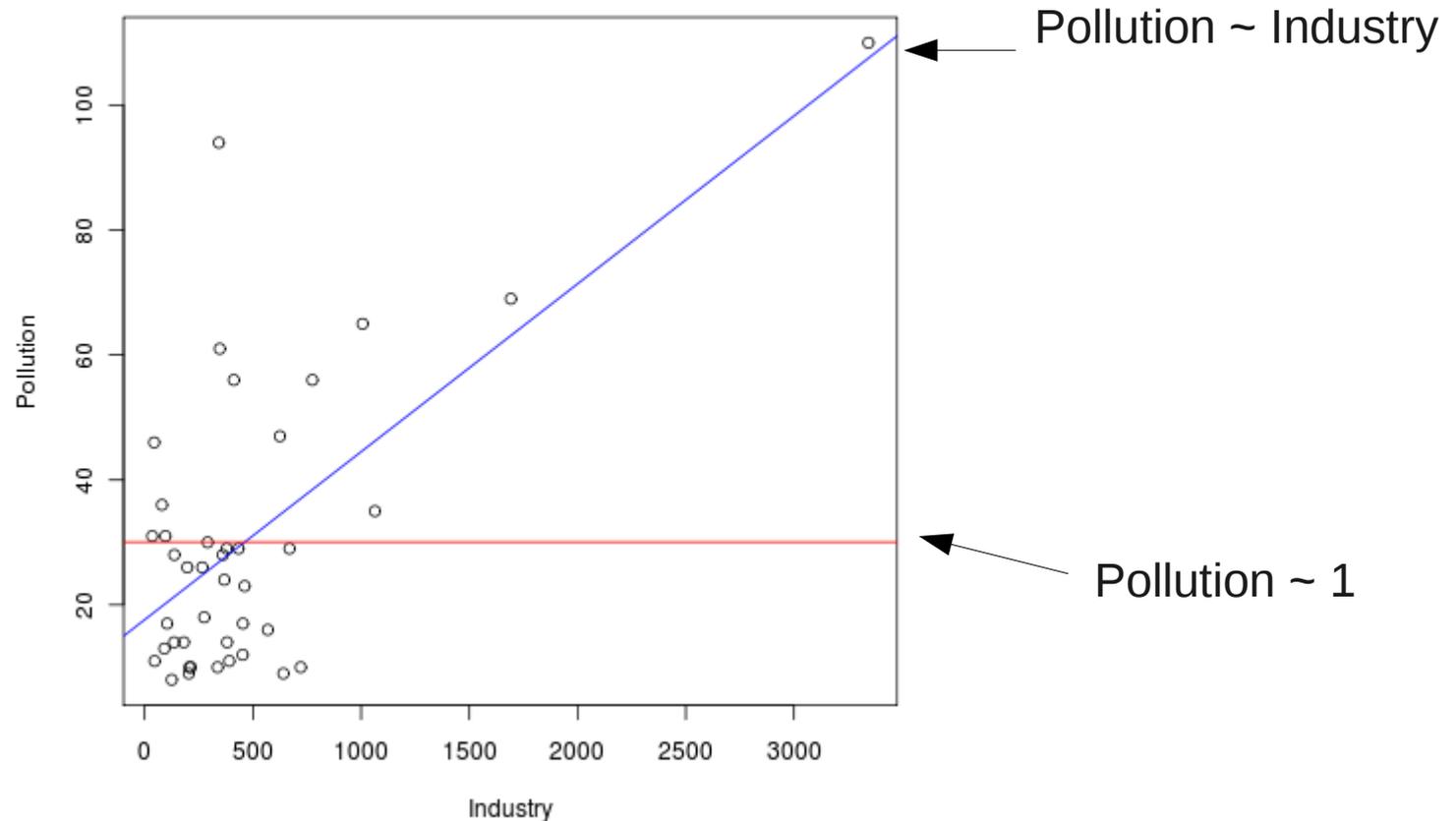
Model 1: Pollution ~ Industry

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	**

# anova()

Uma função de comparação de modelos



- > `pol.m0 <- lm(Pollution~1, data=poluicao)`
- > `anova(pol.m0, pol.m1)`
- > `anova(pol.m1) # o mesmo que os comandos anteriores`

# summary X anova

Atenção para a diferença dos testes!

```
> summary(pol.m2)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	77.429499	21.464395	3.607	0.000888
Industry	0.024354	0.004782	5.093	9.94e-06
Temp	-1.051604	0.372617	-2.822	0.007546

```
> anova(pol.m2)
```

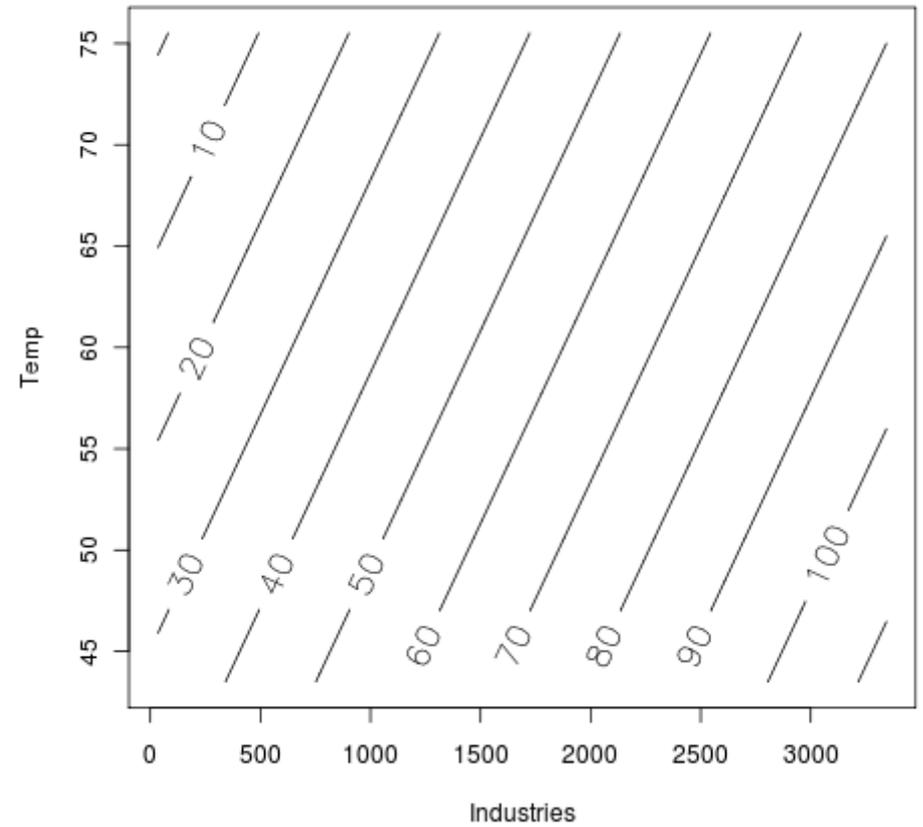
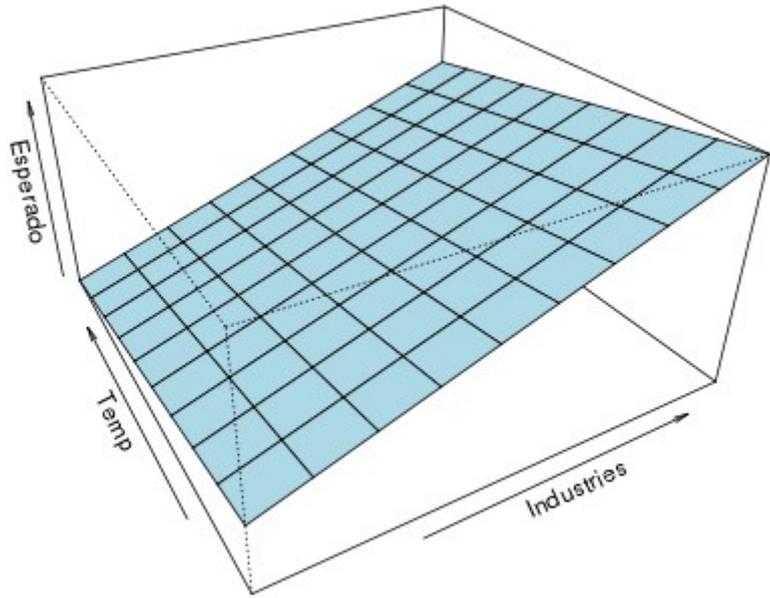
Response: Pollution

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Industry	1	9173.0	9173.0	32.7743	1.358e-06
Temp	1	2229.2	2229.2	7.9649	0.007546
Residuals	38	10635.6	279.9		

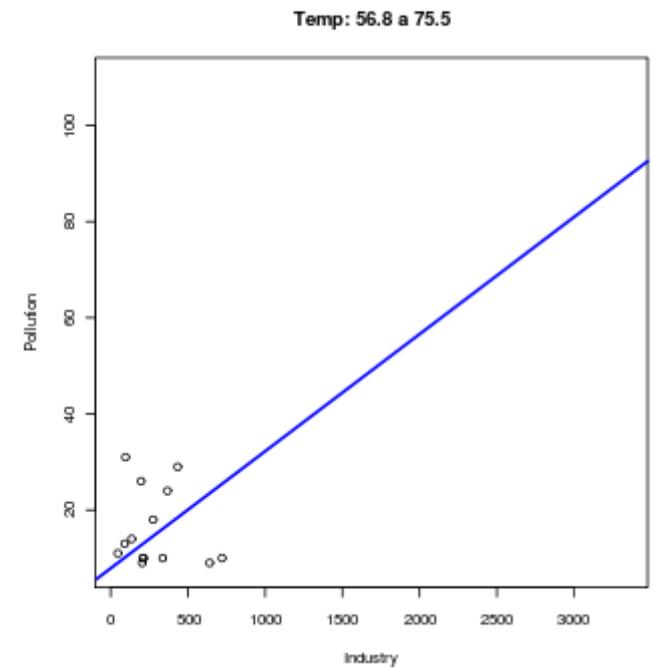
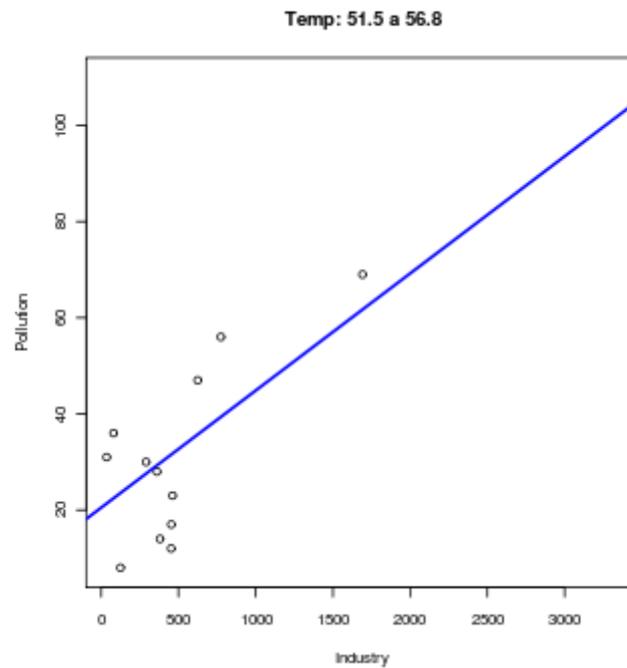
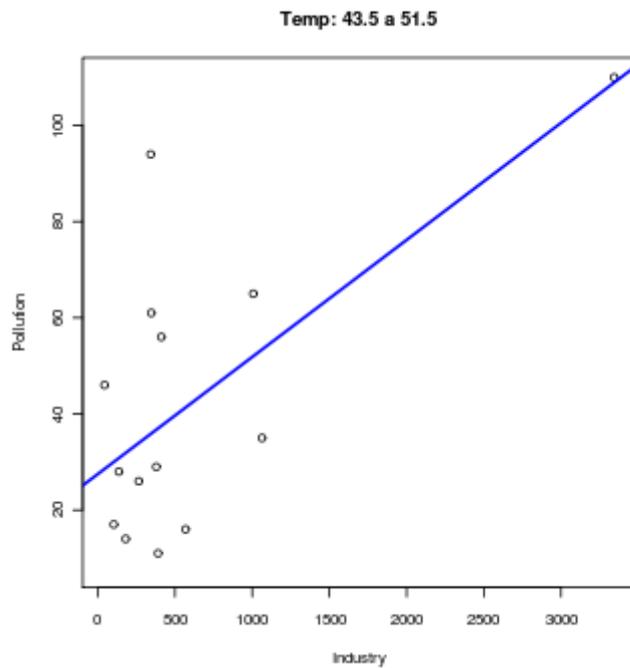
# TESTES SEQUENCIAIS E MARGINAIS



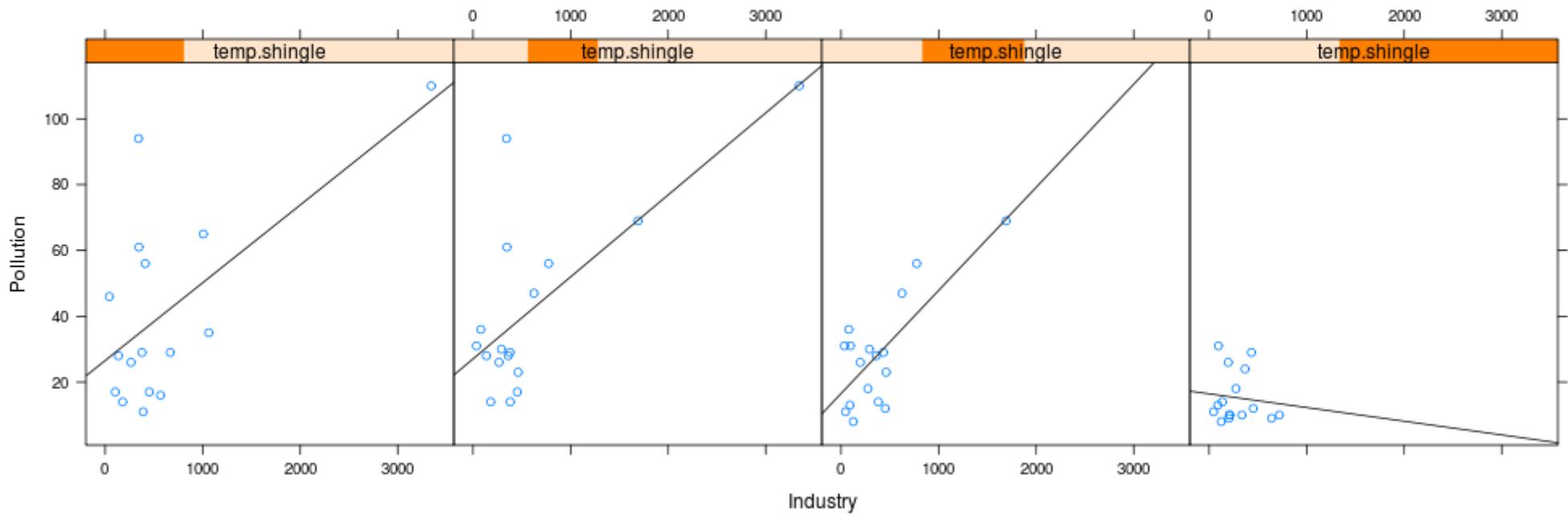
# Duas preditoras: Superfície de Resposta



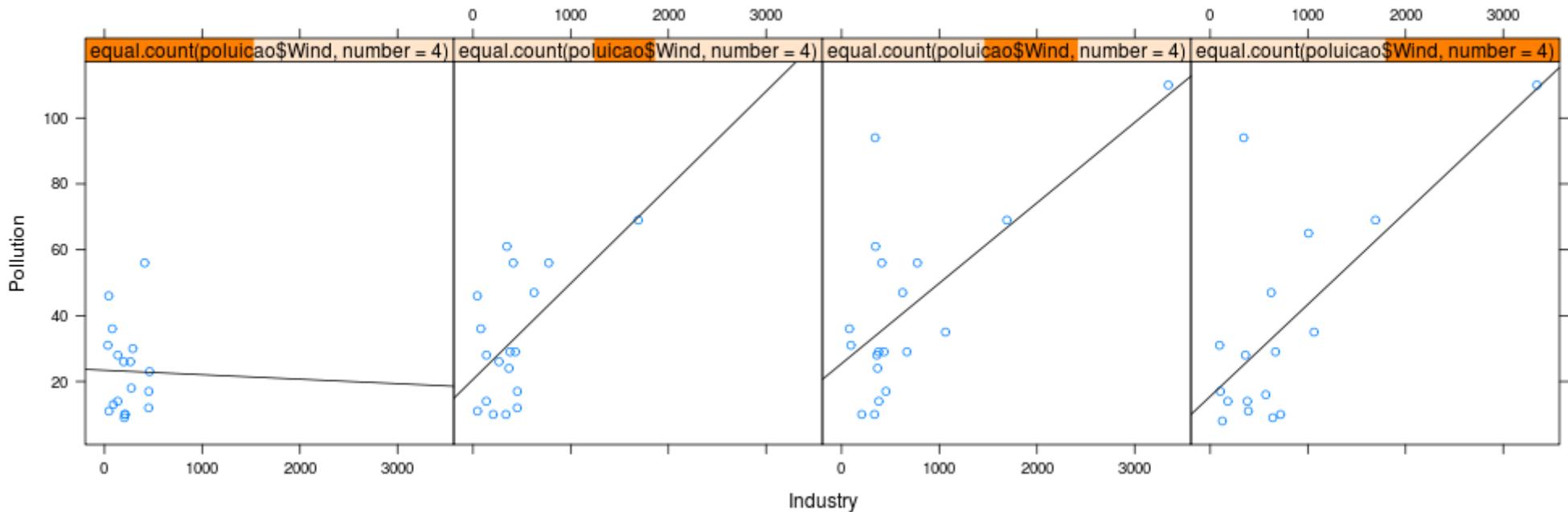
# Duas preditoras: Como Inspeccionar?



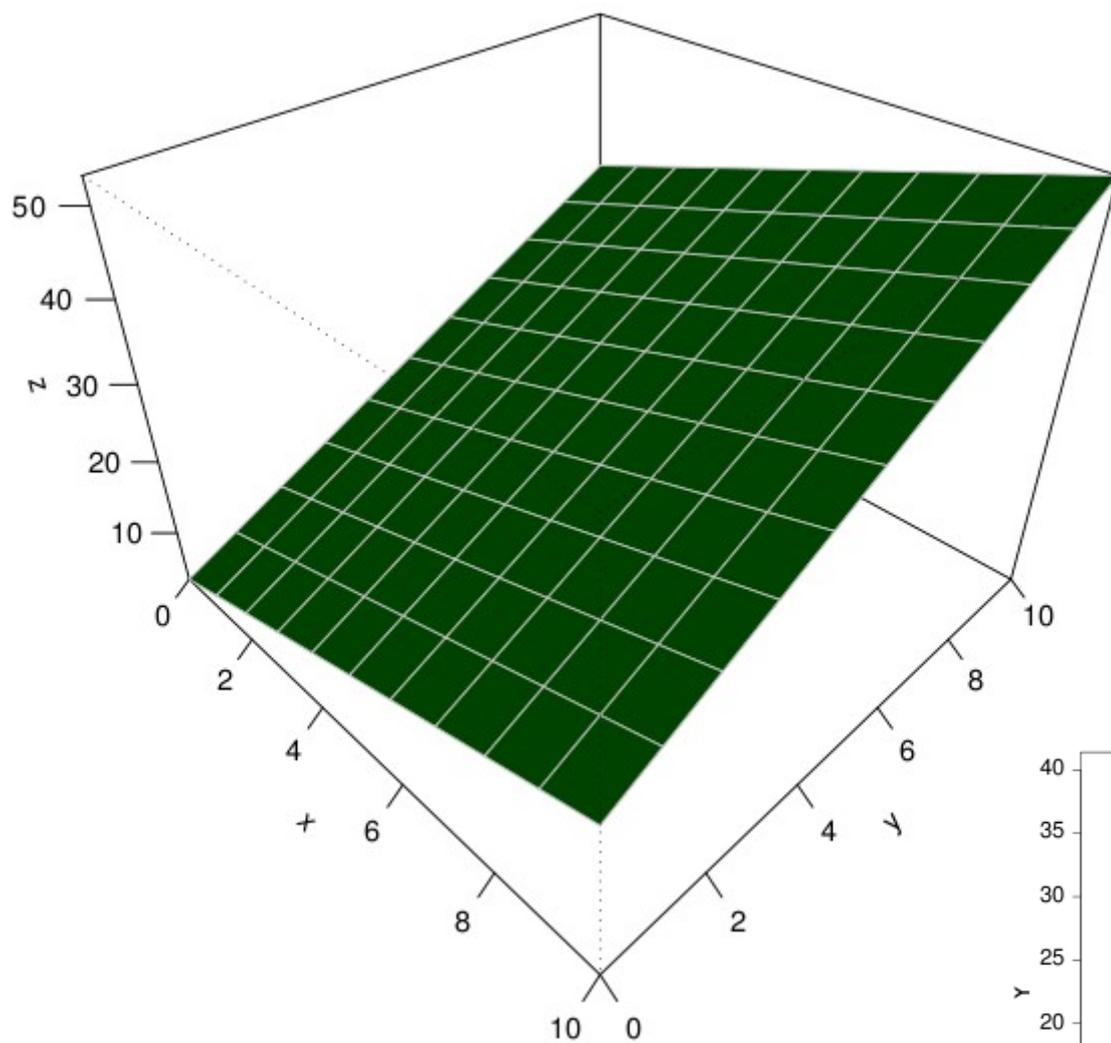
# Duas preditoras: Interação Temperatura x Indústrias?



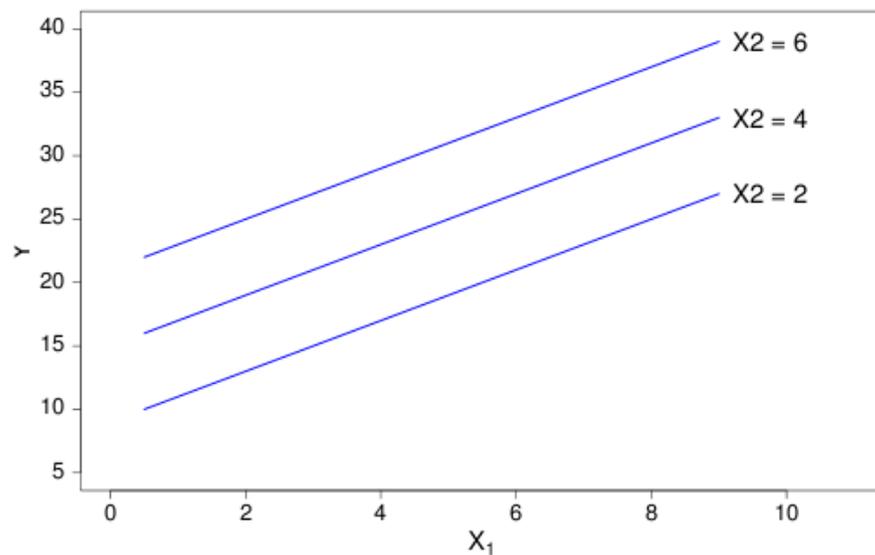
# Duas preditoras: Interação Vento x Indústrias?



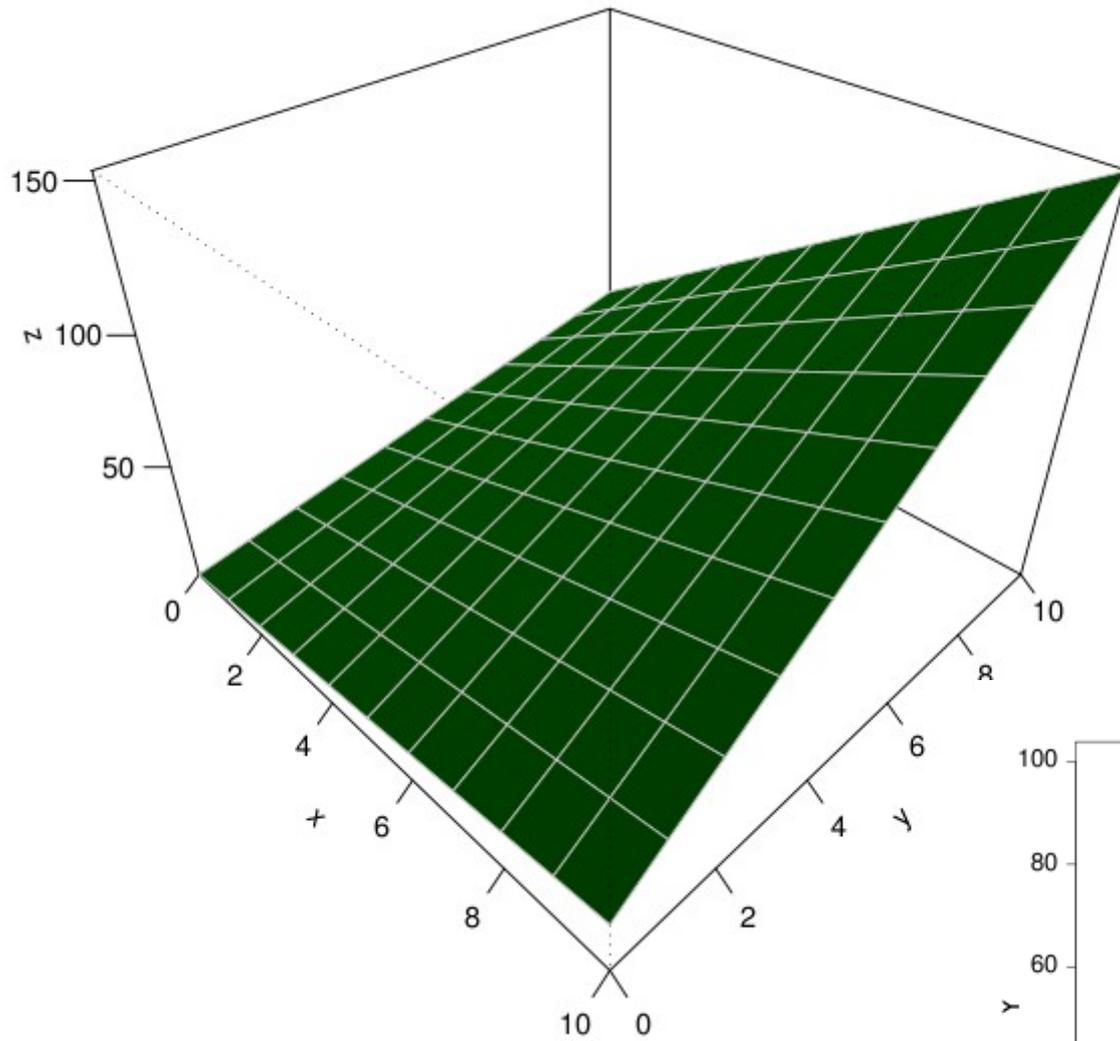
# Modelo sem Interação



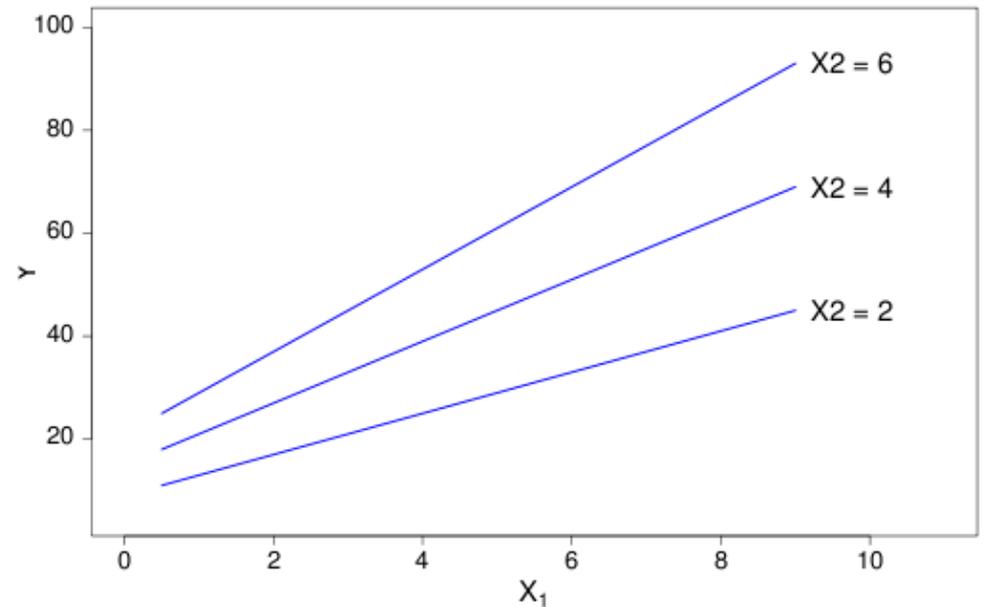
$$y \sim a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_i \cdot x_i$$



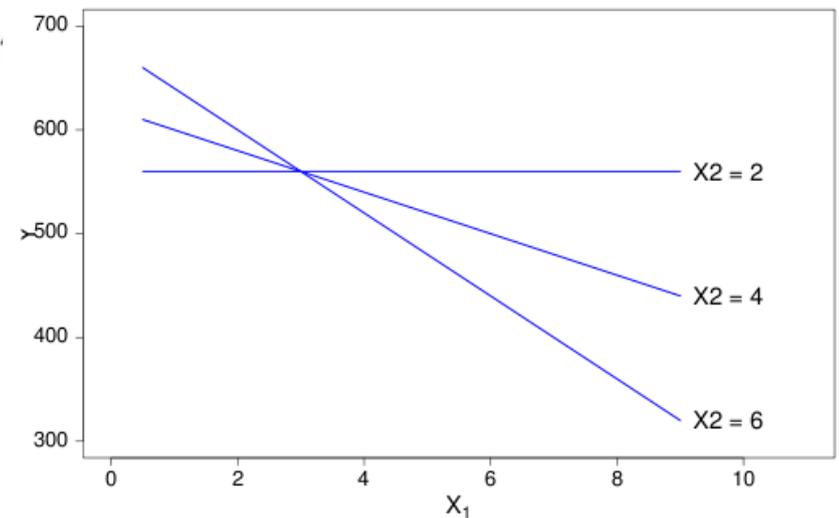
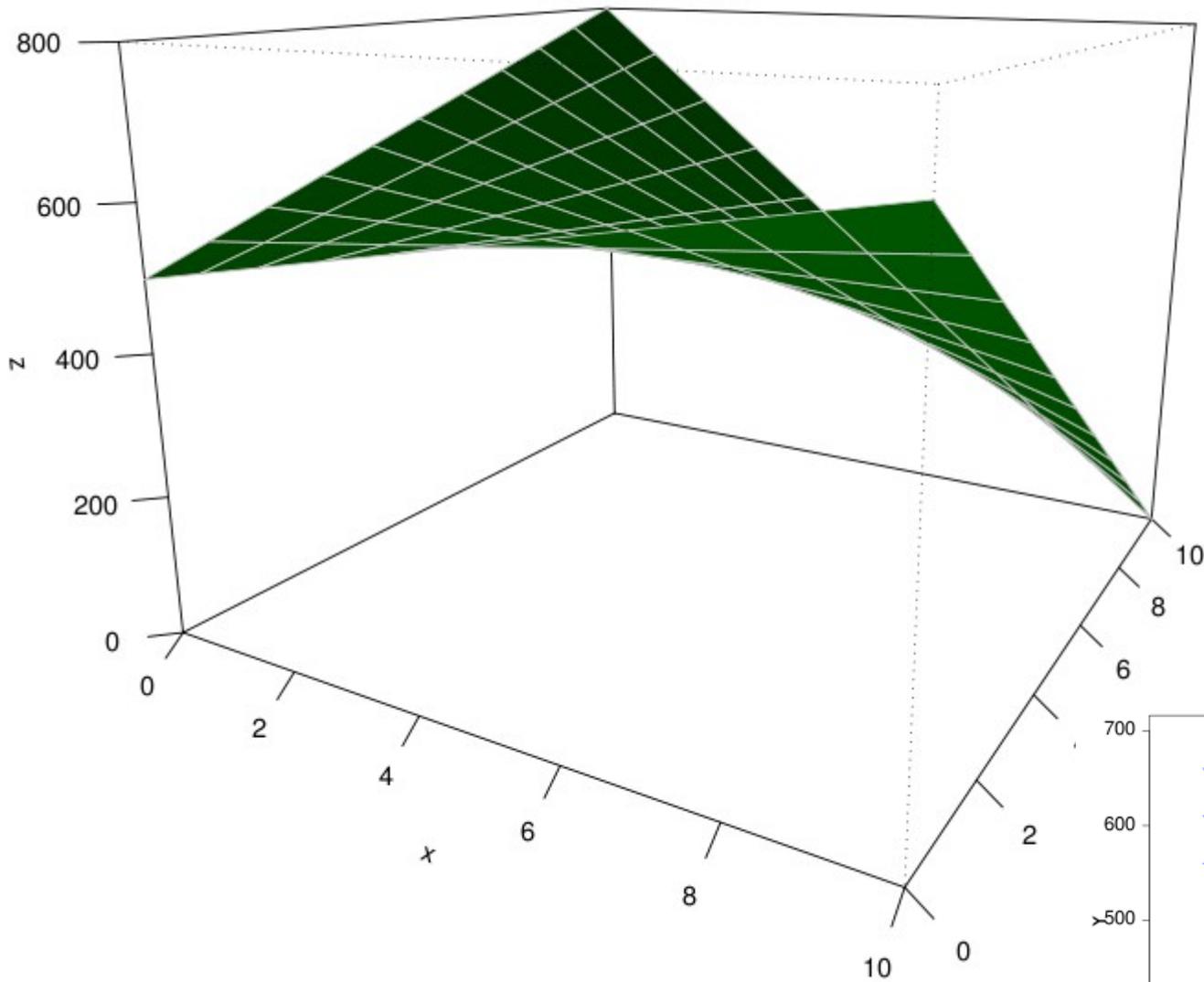
# Modelo com Interação



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



# Modelo com Interação



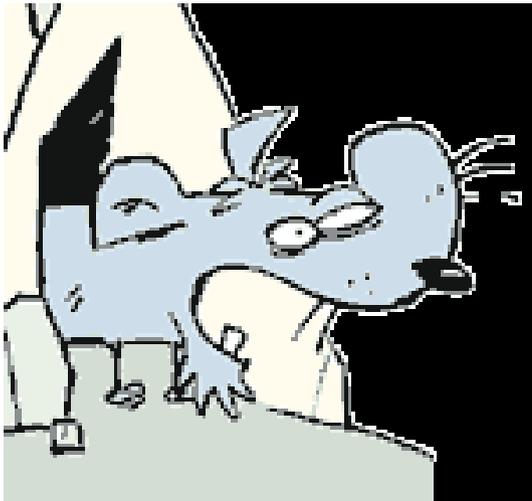
$$y \sim a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_1 x_2 \dots$$

# 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: Bad News

**A variância amostral dos coeficientes de regressão é grande, portanto:**

- Nenhum parâmetro pode ser significativo, mesmo se o modelo é.
- Aumenta a chance de não rejeitar  $H_0 : \beta_k = 0$ ;

# Efeitos da Colinearidade: More Bad News

**A soma de quadrados extra de uma variável preditora depende das demais variáveis presentes no modelo. Portanto:**

- Testes t simultâneos são problemáticos;
- Não é realista avaliar o efeito de uma variável preditora mantendo as demais constantes.

# Efeitos da Colinearidade: Good News

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

# 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
```

```
...
```

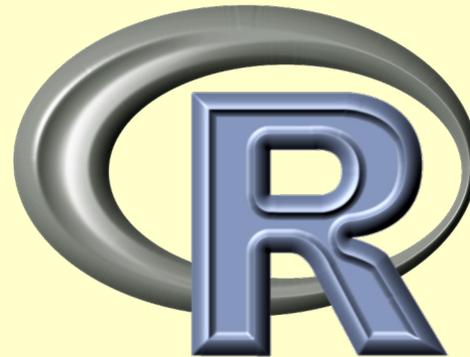
# Cálculo Matricial dos Esperados

$$y = a X$$

$$\begin{bmatrix} 1 & 368 & 61,5 \\ 1 & 291 & 55,6 \\ 1 & 775 & 55,9 \end{bmatrix} \mathbf{X} \begin{bmatrix} 77,4 \\ 0,02 \\ -1,05 \end{bmatrix} =$$

$$= \begin{bmatrix} 1 \times 77,4 + 0,02 \times 368 - 1,05 \times 61,5 \\ 1 \times 77,4 + 0,02 \times 291 - 1,05 \times 55,6 \\ 1 \times 77,4 + 0,02 \times 775 - 1,05 \times 55,9 \end{bmatrix} = \begin{bmatrix} 27,7 \\ 26,1 \\ 37,5 \end{bmatrix}$$

# MATRIZ DO MODELO E VARIÁVEIS CATEGÓRICAS



# FIM DA UNIDADE

# REGRESSÕES LINEARES

Para a próxima aula:

Tutoriais da Unidade

[http://ecologia.ib.usp.br/bie5782/doku.php?id=bie5782:02\\_tutoriais:tut7](http://ecologia.ib.usp.br/bie5782/doku.php?id=bie5782:02_tutoriais:tut7)

Listas de Exercícios:

[http://ecologia.ib.usp.br/bie5782/doku.php?id=bie5782:01\\_curso\\_atual:exercicios7](http://ecologia.ib.usp.br/bie5782/doku.php?id=bie5782:01_curso_atual:exercicios7)

[http://ecologia.ib.usp.br/bie5782/doku.php?id=bie5782:01\\_curso\\_atual:exercicios8](http://ecologia.ib.usp.br/bie5782/doku.php?id=bie5782:01_curso_atual:exercicios8)

# Sugestão de leitura

**John Fox (2002). An R and S-Plus Companion to Applied Regression. Sage Publications, Thousand Oaks, CA, USA.**