3. Construindo superfícies de resposta

Assim como fazemos diagramas de dispersão bidimensional para verificar a relação entre duas variáveis, um gráfico tridimensional pode ser feito com o exemplo usado aqui. Para tal, utilizaremos a função `scatterplot3d()` do pacote homônimo, que, embora não forneca soluções gráficas muito sofisticada, é uma implementação simples e de fácil uso.

```
> library(scatterplot3d)
> s3d <- scatterplot3d(x = DS, y = US, z = RP, angle = 45)
```

Feito isso, e tendo-se ajustado o modelo (`m2`), é possível adicionar ainda um plano 3D representando o ajuste da regressão.

```
> m2 <- lm(RP ~ DS + US, data = solo)
> s3d$plane3d(m2, col = "blue")
```

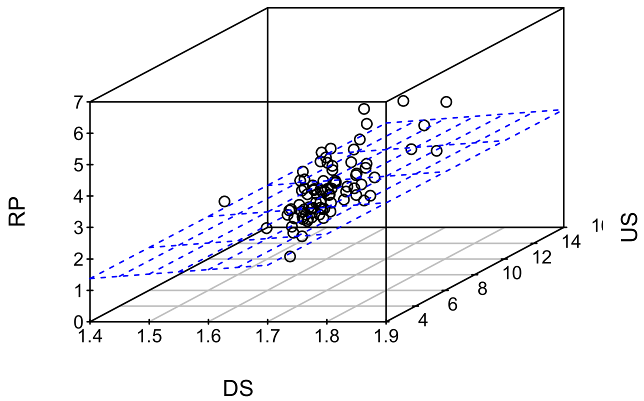

FIGURA 3. Diagrama de dispersão em 3D e plano de regressão ajustado para resistência à penetração em função da densidade e umidade do solo.

Nota: É importante salientar que a função `plane3d()` serve, unicamente, para construir **planos** de regressão. Quaisquer outros modelos, contendo não apenas os termos lineares, devem ser representados graficamente com outras funções mais gerais, tal como a função `persp()`.

4. Regressão não linear

A relação entre uma variável explicativa ou explanatória X e uma variável resposta Y pode ser do tipo linear ou não linear. Mas atenção, NÃO é esta forma de associação que define se um modelo de regressão é linear ou não. Isto é, na verdade, definido pela forma como os parâmetros do modelo se relacionam. Assim, dizemos que a equação $y = b_0 + b_1x + b_2x^2 + \epsilon$ constitui um modelo de regressão *linear nos parâmetros*. Matematicamente, dizemos que um modelo é não linear se pelo menos uma das derivadas parciais em relação aos parâmetros for uma função de pelo menos um parâmetro. Por exemplo, o modelo $y = b_0 + x^{b_1} + \epsilon$ é não linear, pois:

$$\frac{\partial y}{\partial b_1} = f(b_1) = x^{b_1} \log(x)$$

Um método de estimação amplamente utilizado é o *método dos mínimos quadrados não lineares*, cujo princípio é o mesmo do método de mínimos quadrados para modelos lineares. Contudo, modelos não lineares raramente possuem solução analítica para os estimadores. Assim sendo, a estimação é feita por um processo iterativo, de acordo com algum algoritmo de busca, como Gauss-Newton, Newton-Raphson, Gauss-Marquardt, entre outros. Por isso, o uso de processos iterativos requer valores (ou *chutes*) iniciais para os parâmetros.

Exemplo. Visando determinar o tamanho ideal de parcelas experimentais, Lessman & Atkins (1963) propuseram o seguinte modelo para o coeficiente de variação experimental:

$$y = b_0 x^{-b_1} + \epsilon$$

em que y é o coeficiente de variação experimental; x é o tamanho da parcela; b_0 e b_1 são parâmetros.

Uma forma bastante simples de se obter valores iniciais é por meio da *linearização* do modelo não linear, ignorando (obviamente) o erro aleatório. Com o exemplo, podemos fazer:

$$\log(y) = \log(b_0) - b_1 \log(x)$$

que é, agora, um modelo de regressão linear simples.

Considere agora os seguintes dados de $CV\%$ e tamanho de parcela (m^2):

```
> ps <- c(1, 2, 3, 4, 6, 8, 12)
> cv <- c(35.6, 29, 27.1, 25.6, 24.4, 23.3, 21.6)
> lm(log(cv) ~ log(ps))

Call:
lm(formula = log(cv) ~ log(ps))

Coefficients:
(Intercept)      log(ps)
     3.5309      -0.1908
```

Poderíamos então utilizar esses valores iniciais (3.53 e -0.19) para obter as estimativas de mínimos quadrados não lineares, usando a função `nls()`.

```
> nls(cv ~ b0*ps^-b1,
+     start = list(b0 = exp(3.53), b1 = 0.19))
Nonlinear regression model
  model: cv ~ b0 * ps^-b1
   data: parent.frame()
     b0      b1
34.5394  0.1996
 residual sum-of-squares: 3.665

Number of iterations to convergence: 3
Achieved convergence tolerance: 3.061e-06
```

Podemos agora desenhar o ajuste da regressão (Figura 4):

```
> plot(ps, cv)
> curve(34.54*x^-0.1996, add = TRUE, col = "blue")
```

5. *Linear Response Plateau*

Não raramente, em estudos de nutrição de plantas, é esperado que a partir de certo ponto haja uma faixa de estabilização da produção ou crescimento vegetal em função da adição de nutrientes. Nesse caso, parece apropriado ajustar um modelo de regressão do tipo LRP (*linear response plateau*, do inglês) ou platô de resposta linear.

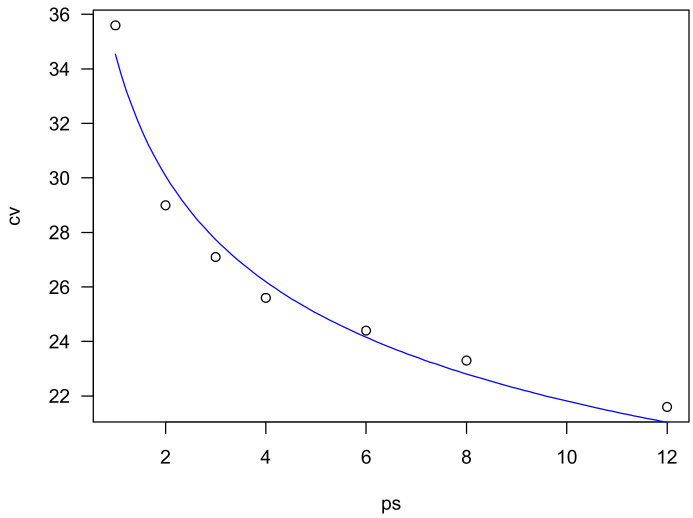

FIGURA 4. Ajuste do modelo não linear para o coeficiente de variação experimental em função do tamanho da parcela.

O modelo LRP nada mais é do que um modelo de regressão segmentado, em duas partes, sendo que em uma delas a resposta é descrita por um modelo de regressão linear simples, e a outra é descrita por um paralelo a x (platô), isto é, onde a resposta é constante. Formalmente, o escrevemos da seguinte forma: $y = h(x) + \epsilon$, em que:

$$h(x) = \begin{array}{ll} b_0 + b_1 x, & x \leq x_0 \\ b_0 + b_1 x_0, & x > x_0 \end{array}$$

sendo o parâmetro x_0 o valor no eixo-x que delimita o platô.

Exemplo. Hartinee et al. (2010) utilizaram o modelo LRP para determinar a exigência de nitrogênio por arroz. Dados de produção em função de doses de um fertilizante nitrogenado são mostrados na tabela a seguir:

TABELA 1. Efeito de N na produção de arroz (Hartinee et al., 2010).

N	0	50	100	150	200
Prod.	7.67	14.46	19.84	20.04	20.59

A relação entre produção e dose de N pode ser melhor entendida quando visualizada graficamente. Isso ainda nos ajudará identificar onde deve estar o platô.

```
> N <- c(0, 50, 100, 150, 200)
> prod <- c(7.67, 14.46, 19.84, 20.04, 20.59)
> plot(prod ~ N)
```

No R, uma forma prática de estimar os parâmetros do modelo LRP é tratando-o como um modelo não linear, via a implementação `nls()`.

```
> lrp <- function (x, b0, b1, x0)
+     ifelse(x <= x0, b0 + b1*x, b0 + b1*x0)
> nls(prod ~ lrp(N, b0, b1, x0),
+     start = list(b0 = 8, b1 = 0.1, x0 = 80))
Nonlinear regression model
  model: prod ~ lrp(N, b0, b1, x0)
```

```
  data: parent.frame()
    b0      b1      x0
 7.6700  0.1358 91.9489
 residual sum-of-squares: 0.3017

Number of iterations to convergence: 2
Achieved convergence tolerance: 1.147e-08
```

Logo, a equação ajustada foi:

$$\hat{y} = \begin{cases} 7.67 + 0.1358x, & x \leq 91.9489 \\ 20.15666, & x > 91.9489 \end{cases}$$

sendo 21.15666 o platô de resposta. A Figura 5 mostra o ajuste do modelo.

```
> curve(lrp(x, b0 = 7.67, b1 = 0.1358, x0 = 91.95), col = "blue", add
    = TRUE)
```

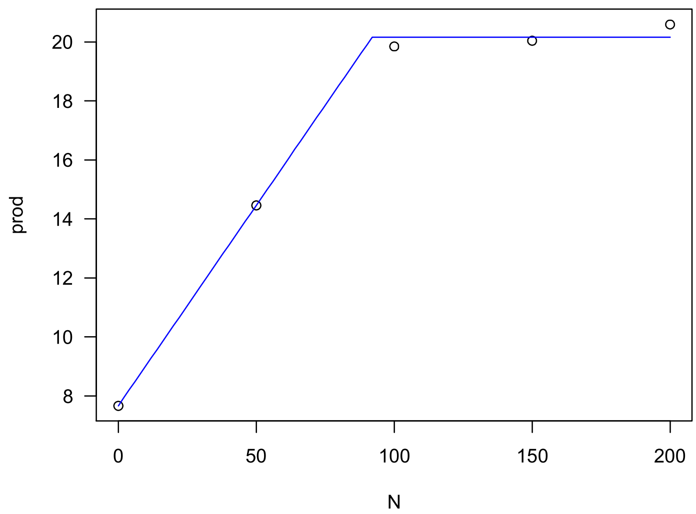

Figura 5. Ajuste do modelo LRP aos dados de produção de arroz em função de doses de fertilizante nitrogenado.

6. Exercícios

(1) Ajuste um modelo de regressão linear múltipla para a resistência à penetração em função da densidade e umidade do solo de cada uma das duas camadas (0-20 e 20-40 cm). Depois construa um diagrama de dispersão 3D, indicando com cores diferentes as observações tomadas em cada camada. Finalmente, adicione a esse gráfico os dois planos de regressão ajustados, separando também pelas cores das observações.

(2) Como você classificaria o seguinte modelo, linear ou não linear?

$$y = b_1\sqrt{x} + b_2 x + \epsilon$$

(3) Busscher (1990) propôs o seguinte modelo para a resistência à penetração:

$$RP = b_0 \theta^{b_1} \rho^{b_2} + \epsilon$$

Em que θ é o conteúdo de água no solo, ρ é o valor da densidade do solo, b_0, b_1 e b_2 são parâmetros. Pede-se: com os dados das duas camadas, ajuste o modelo de Busscher e construa a superfície de resposta.