ME613 - Análise de Regressão

Parte 6

Benilton S Carvalho e Rafael P Maia - 2S2020

Regressão Linear Múltipla

Regressão Linear Múltipla

Imagine que algum pesquisador apresente o seguinte resultado: há relação entre uso de balinhas de menta (X, total por dia) e função pulmonar (Y, FEV).

O que você diria?

Você poderia argumentar, por exemplo, que fumantes consomem mais balinhas de menta e que o fato de ser fumante influencia na função pulmonar, não as balinhas.

O pesquisador então perguntaria: como eu poderia convencer você do efeito das balinhas?

Você poderia dizer que estaria convencido se, por exemplo: não-fumantes consumidores de balinhas de menta apresentam função pulmonar menor do que fumantes não consumidores de balinhas de menta; ou se fumantes consumidores de balinhas de menta apresentam função pulmonar melhor do que os fumantes não consumidores de balinhas de menta.

Regressão Linear Múltipla

Ou seja, para verificar o efeito do consumo de balinhas de menta, você gostaria de manter o efeito do cigarro (fumantes e não fumante) fixo.

A técnica de regressão linear múltipla pode ser usada neste caso: ela avaliará a relação entre um preditor e a resposta, enquanto “controla” pelas demais variáveis no modelo.

Modelo com duas variáveis preditoras

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

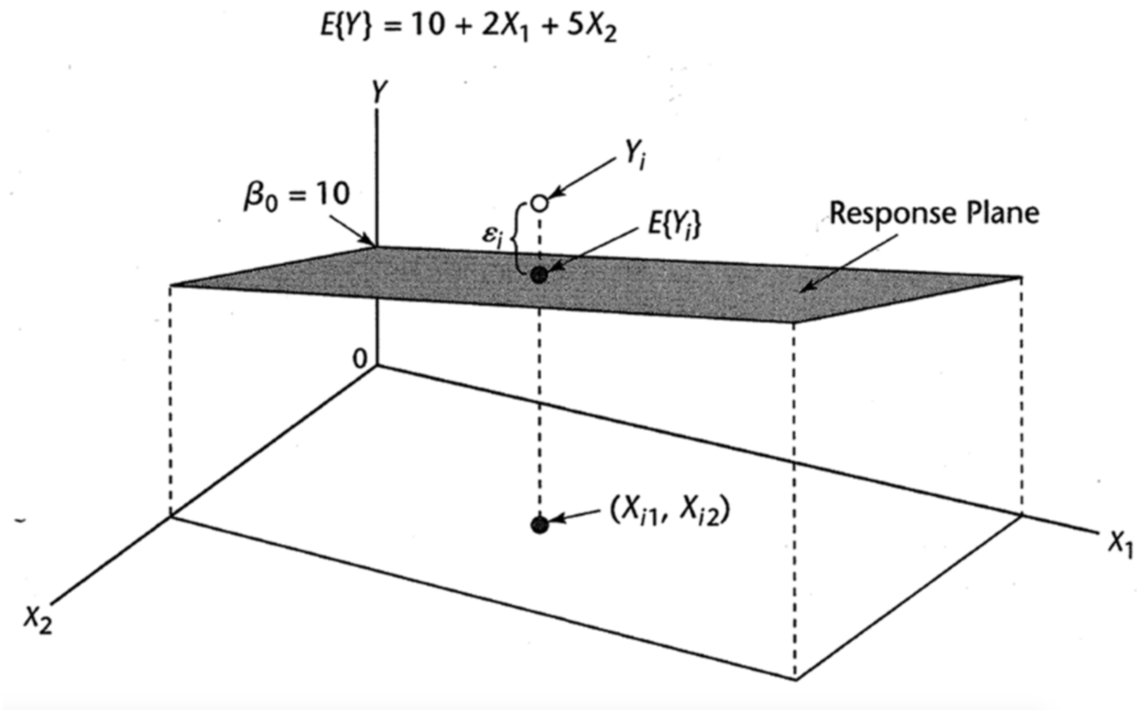
X_{i1} e X_{i2} são valores de duas variáveis preditoras para a observação i.

Assumindo que $E(\varepsilon_i) = 0\,, \forall i$:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Exemplo

Na situação com duas variáveis preditoras, a função de regressão representa um plano:



Exemplo

Interpretação dos coeficientes:

- β_0 (intercepto): valor esperado de Y quando $X_1 = 0$ e $X_2 = 0$.
- β_1: indica a mudança no valor esperado de Y para cada unidade de aumento de X_1, quando X_2 é mantida constante.
- β_2: indica a mudança no valor esperado de Y para cada unidade de aumento de X_2, quando X_1 é mantida constante.

Exemplo, se fixamos $X_2 = 2$:

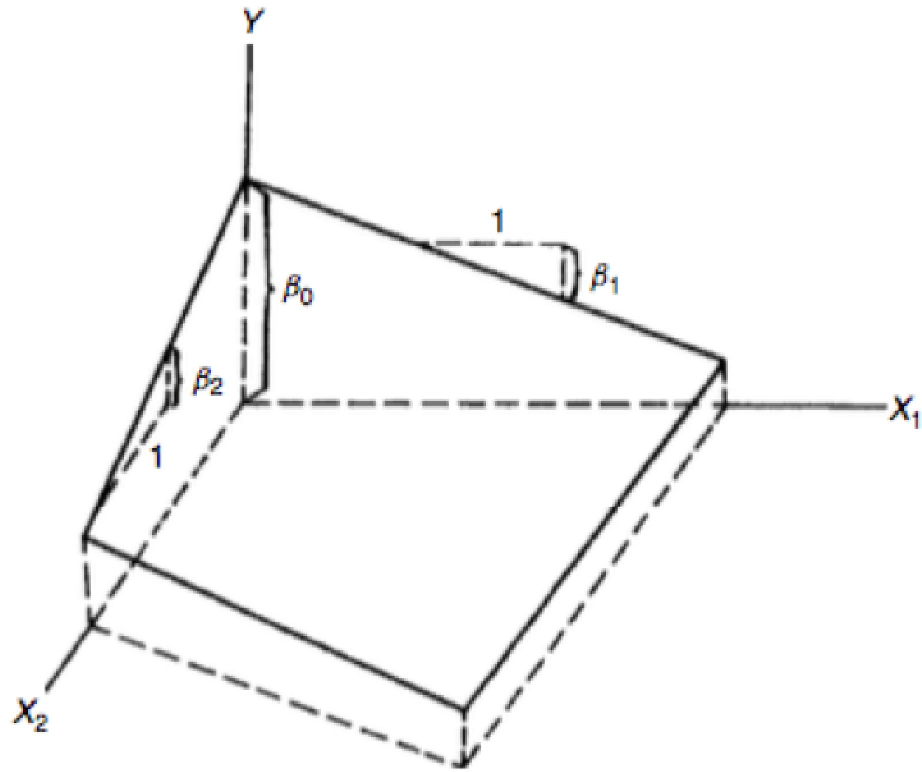
$$E(Y) = 10 + 2X_1 + 5 \times 2 = 20 + 2X_1$$

Exemplo

- Se $\beta_1 = 2$: o valor esperado de Y aumenta 2 unidades a cada aumento de 1 unidade de X_1 e X_2 mantida constante.
- Se $\beta_2 = 5$: o valor esperado de Y aumenta 5 unidades a cada aumento de 1 unidade de X_2 e X_1 mantida constante.

Exemplo

Na situação com duas variáveis preditoras, a função de regressão representa um plano:



Modelo de regressão linear múltipla geral

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

- $\beta_0, \beta_1, \ldots, \beta_{p-1}$ são parâmetros.
- $X_{i1}, \ldots, X_{i,p-1}$ são constantes conhecidas.
- $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$.
- $i = 1, 2, \ldots, n$.

Se $X_{i0} = 1$, podemos escrever:

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i$$

Modelo de regressão linear múltipla geral

Função de regressão (hiperplano):

$$E(Y) = \beta_0 + \beta_1 X_1 + \ldots + \beta_{p-1} X_{p-1}$$

Regressão Linear Múltipla com notação matricial

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i \,, \quad \varepsilon_i \overset{iid}{\sim} \mathbf{N}(0, \sigma^2) \,, \quad i = 1, 2, \ldots, n$$

$$\mathbf{Y}_{n\times 1} = \mathbf{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\varepsilon}_{n\times 1} \,, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

$$\mathbf{Y}_{n\times 1} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$$

$$\mathbf{X}_{n\times p} = \begin{pmatrix} 1 & X_{11} & X_{12} & \ldots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \ldots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & X_{n1} & X_{n2} & \ldots & X_{n,p-1} \end{pmatrix} \quad \boldsymbol{\beta}_{p\times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} \quad \boldsymbol{\varepsilon}_{n\times 1} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Regressão Linear Múltipla com notação matricial

$$E(\boldsymbol{\varepsilon})_{n\times 1} = \mathbf{0}_{n\times 1}$$

$$Var(\boldsymbol{\varepsilon})_{n\times n} = \sigma^2\mathbf{I}_{n\times n}$$

$$E(\mathbf{Y}) = E(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta}$$

$$Var(\mathbf{Y}) = \sigma^2\mathbf{I}$$

Mínimos Quadrados

Queremos encontrar $\hat{\boldsymbol{\beta}}$ que minimiza:

$$
\begin{aligned}
S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{Y}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} \\
&= \mathbf{Y}^T\mathbf{Y} - 2\boldsymbol{\beta}^T\mathbf{X}^T\mathbf{Y} + \boldsymbol{\beta}^T\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}
\end{aligned}
$$

$$
\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T\mathbf{Y} + 2\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}
$$

Equação normal: $\mathbf{X}^T\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y}$

$$
\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}
$$

Mínimos Quadrados

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}) &= Var\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}\right] \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T Var(\mathbf{Y})\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \end{aligned}$$

$\mathbf{H}$ é a matriz de projeção ortogonal no espaço coluna de $\mathbf{X}$.

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \underbrace{\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T}_{\mathbf{H}}\mathbf{Y}$$

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{H}\mathbf{Y} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

Preditores qualitativos

Muitas vezes as variáveis preditoras podem ser do tipo qualitativo:

- Sexo: feminino/masculino
- Tem ensino superior? sim/não
- etc

Exemplo

Modelo de regressão para tempo de permanência no hospital (Y) considerando a idade (X_1) e o sexo (X_2) do paciente.

$$X_2 = \begin{cases} 1 & \text{se feminino} \\ 0 & \text{se masculino} \end{cases}$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

- X_{i1} é a idade do paciente i.
- X_{i2} é o sexo do paciente i.

Se $X_2 = 0$ (paciente masculino): $E(Y) = \beta_0 + \beta_1 X_1$.

Se $X_2 = 1$ (paciente feminino): $E(Y) = (\beta_0 + \beta_2) + \beta_1 X_1$.

Preditores qualitativos

Em geral, representamos uma variável qualitativa com c classes através de $c - 1$ variáveis indicadoras.

Por exemplo, se temos uma variável qualitativa do estado de incapacidade do paciente com as seguintes classes: incapaz, parcialmente incapaz, não incapaz. Utilizamos as seguinte vairáveis indicadoras:

$$X_3 = \begin{cases} 1 & \text{se não incapaz} \\ 0 & \text{caso contrário} \end{cases}$$

$$X_4 = \begin{cases} 1 & \text{se parcialmente incapaz} \\ 0 & \text{caso contrário} \end{cases}$$

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i$$

Exemplo: conjunto de dados `swiss`

```
require(datasets); data(swiss); ?swiss
```

A data frame with 47 observations on 6 variables, each of which is in percent, i.e., in [0, 100].

[,1] Fertility Ig, ‘common standardized fertility measure’

[,2] Agriculture % of males involved in agriculture as occupation

[,3] Examination % draftees receiving highest mark on army examination

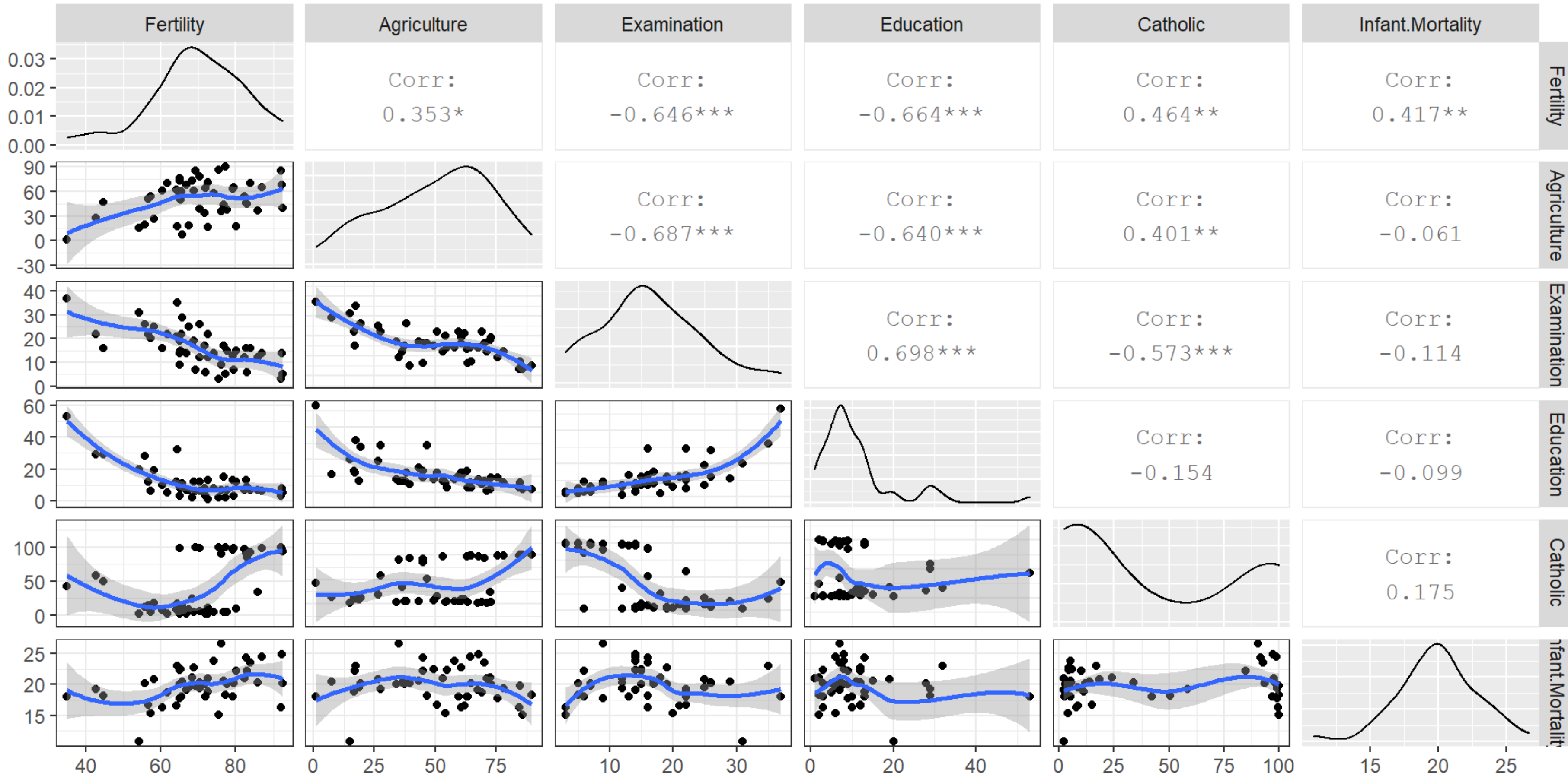
[,4] Education % education beyond primary school for draftees.

[,5] Catholic % ‘catholic’ (as opposed to ‘protestant’).

[,6] Infant.Mortality live births who live less than 1 year.

All variables but ‘Fertility’ give proportions of the population.

Exemplo: conjunto de dados `swiss`



Exemplo: conjunto de dados `swiss`

```
modelo <- lm(Fertility ~ . , data = swiss)
summary(modelo)$coefficients
```

```
##                    Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept)      66.9151817 10.70603759  6.250229 1.906051e-07
## Agriculture      -0.1721140  0.07030392 -2.448142 1.872715e-02
## Examination      -0.2580082  0.25387820 -1.016268 3.154617e-01
## Education        -0.8709401  0.18302860 -4.758492 2.430605e-05
## Catholic          0.1041153  0.03525785  2.952969 5.190079e-03
## Infant.Mortality  1.0770481  0.38171965  2.821568 7.335715e-03
```

Exemplo: conjunto de dados `swiss`

- `Agriculture`: expressa em porcentagem (0 - 100)
- Estimativa é -0.172114.
- Segundo o modelo, espera-se um decréscimo de 0.17 na fertilidade para cada 1% de aumento de pessoas do sexo masculino envolvidas na agricultura, mantendo as demais variáveis fixas.
- O teste-t para $H_0 : \beta_{Agri} = 0$ versus $H_a : \beta_{Agri} \neq 0$ é significante.
- A título de curiosidade, a estimativa do efeito de agricultura, sem ajustar pelas demais variáveis é:

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 60.3043752 4.25125562 14.185074 3.216304e-18
## Agriculture  0.1942017 0.07671176  2.531577 1.491720e-02
```

(Paradoxo de Simpson)

Simulação

Ao considerarmos outras variáveis no modelo, o sinal do efeito de uma dada variável pode inverter. Vamos simular um caso para exemplificar.

- Simulamos 100 v.a. com relação linear: Y, X_1 e X_2.
- X_1 tem relação linear com X_2.
- X_1 tem um efeito ajustado negativo sobre Y.
- X_2 tem um efeito ajustado positivo sobre Y.

Simulação

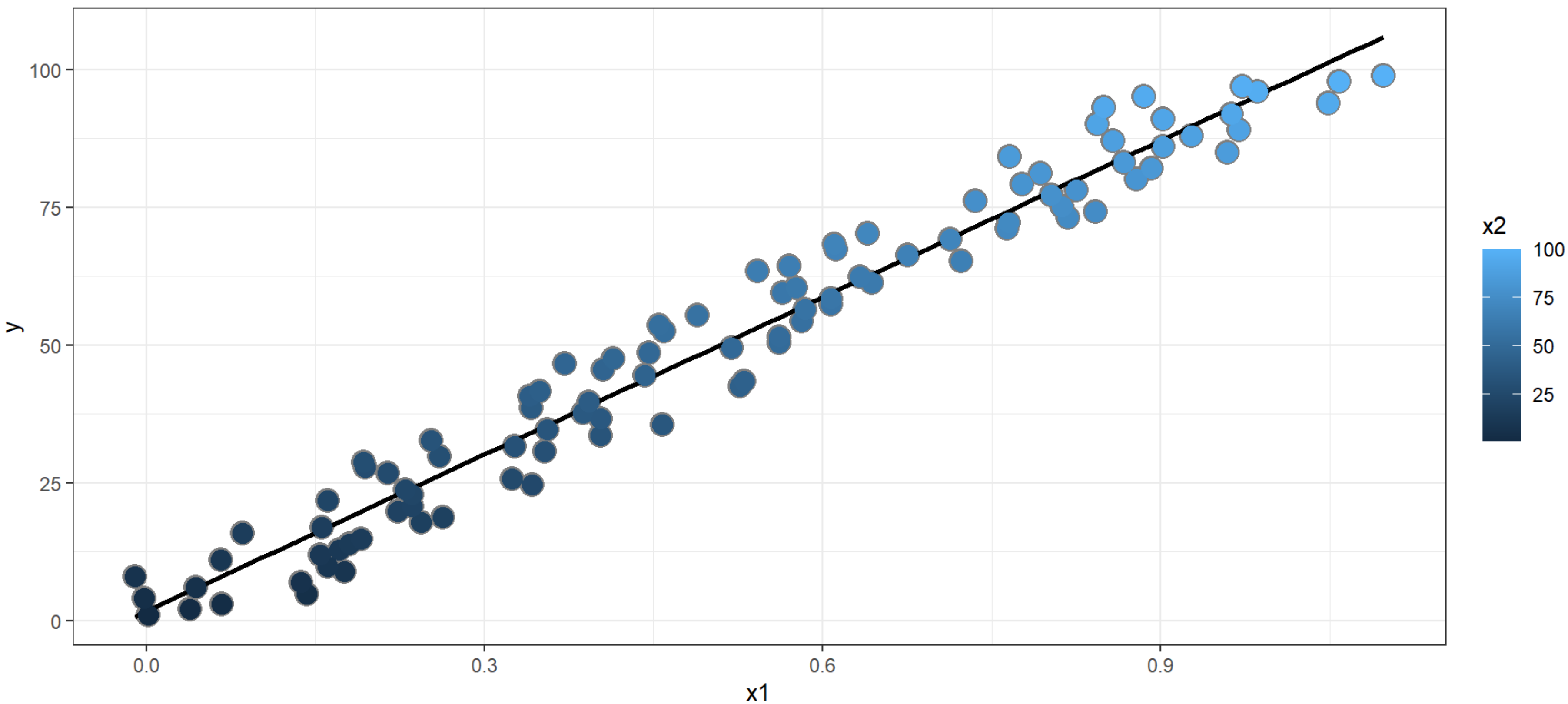
```r
n <- 100
x2 <- 1 : n
x1 <- .01 * x2 + runif(n, -.1, .1)
y = -x1 + x2 + rnorm(n, sd = .01)
summary(lm(y ~ x1))$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  1.657251   1.139998  1.453732 1.492156e-01
## x1          95.030050   1.939222 49.004216 9.766532e-71
```

```r
summary(lm(y ~ x1 + x2))$coef
```

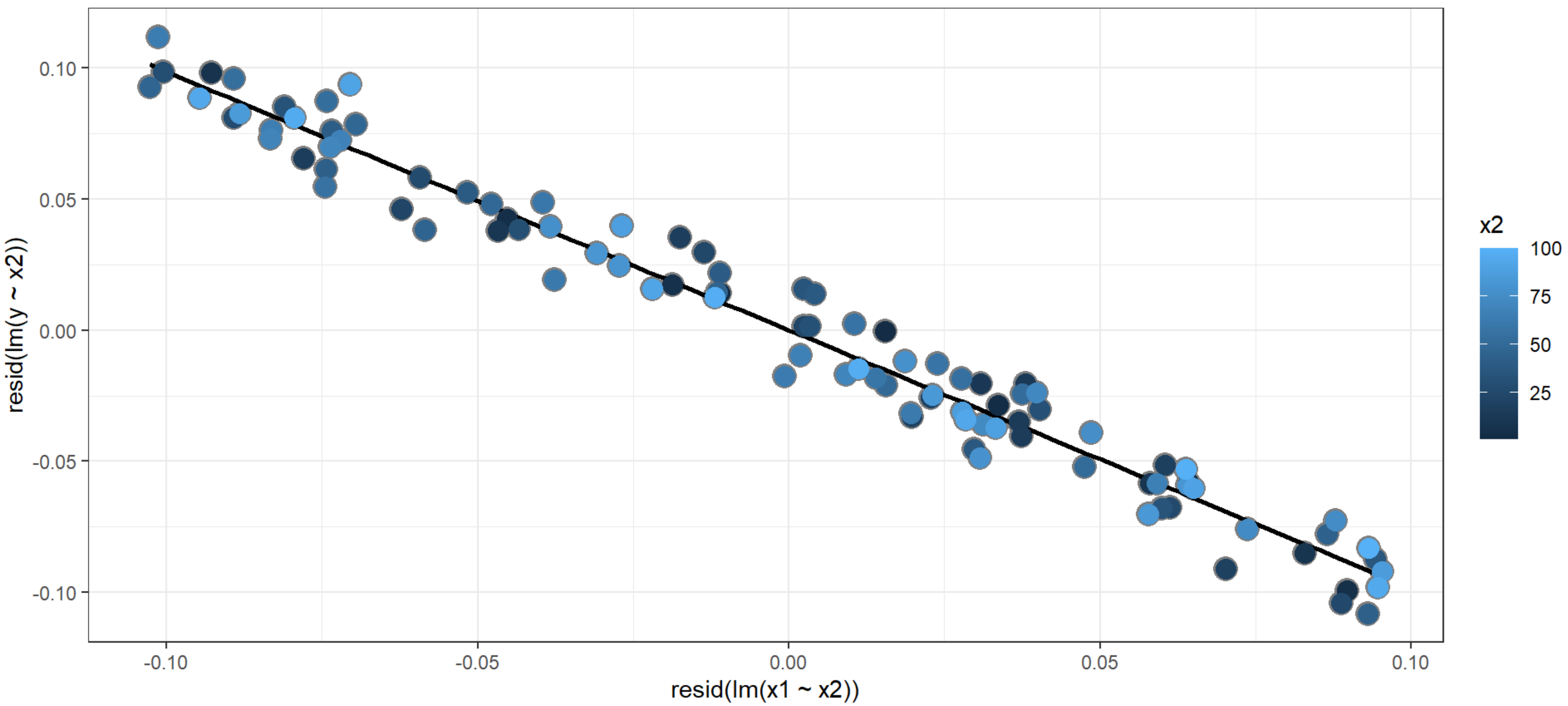
```
##                  Estimate   Std. Error      t value      Pr(>|t|)
## (Intercept) -0.0003312952 0.0020561458   -0.1611244  8.723306e-01
## x1          -0.9844021684 0.0176502642  -55.7726591  1.670030e-75
## x2           0.9998552934 0.0001802353 5547.4984934 1.225662e-268
```

Simulação



Y e X_1 têm relação positiva (não ajustada). Note que X_2 também aumenta com Y.

Simulação



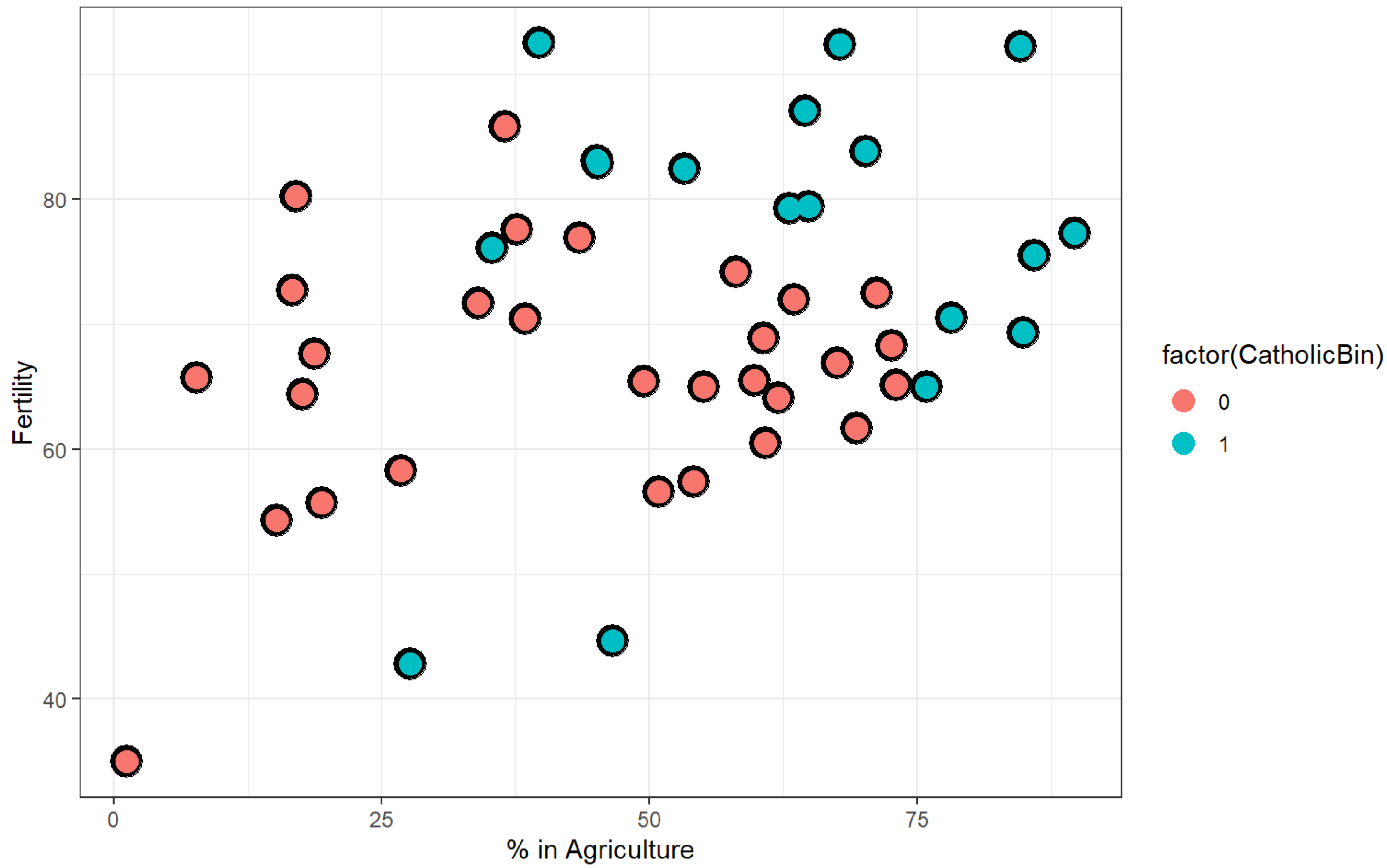
Ajustando X_1 e Y através do resíduo da regressão de cada uma em X_2 temos a relação correta entre X_1 e Y.

Exemplo: conjunto de dados `swiss`

Vamos considerar a seguinte variável qualitativa:

```
library(dplyr);
swiss = mutate(swiss, CatholicBin = 1 * (Catholic > 50))
```

Exemplo: conjunto de dados `swiss`



Exemplo: conjunto de dados `swiss`

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

- Y_i: `Fertility`
- X_{i1}: `Agriculture`
- X_{i2}: `CatholicBin`

Exemplo: conjunto de dados `swiss`

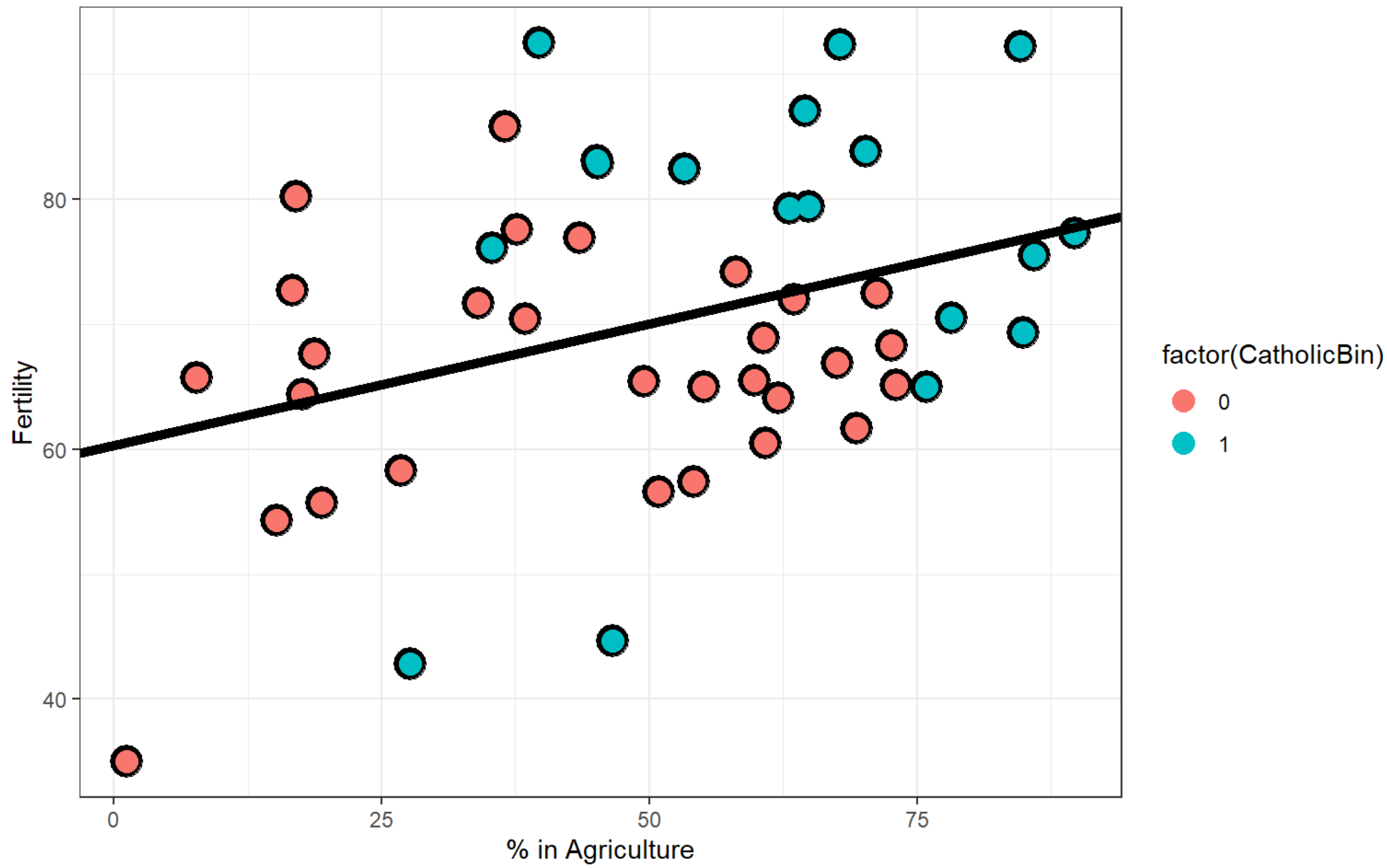
Sem considerar X_{i2}:

```
summary(lm(Fertility ~ Agriculture, data = swiss))$coef
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 60.3043752 4.25125562 14.185074 3.216304e-18
## Agriculture  0.1942017 0.07671176  2.531577 1.491720e-02
```

Este modelo assume que ajustamos apenas uma reta.

Exemplo: conjunto de dados `swiss`



Exemplo: conjunto de dados `swiss`

No modelo:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$$

Temos que, se $X_{i2} = 0$:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$$

e se $X_{i2} = 1$:

$$Y_i = (\beta_0 + \beta_2) + \beta_1 X_{i1} + \varepsilon_i$$

Ou seja, temos duas retas paralelas ajustadas (uma para cada categoria de `CatholicBin`).

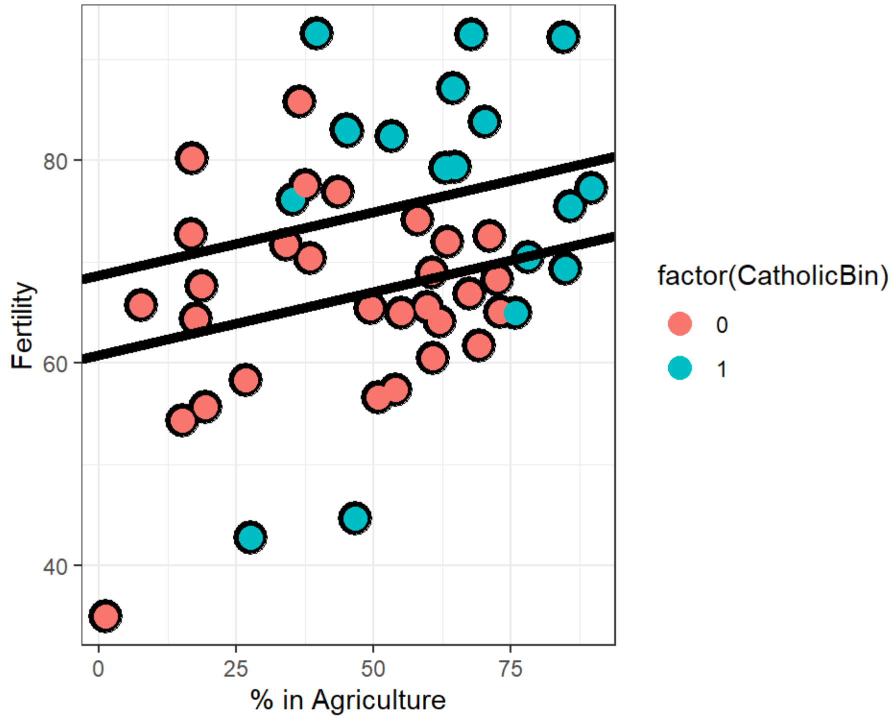
Exemplo: conjunto de dados `swiss`

```
summary(lm(Fertility ~ Agriculture + factor(CatholicBin), data = swiss))$coef
```

```
##                       Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)         60.8322366  4.1058630 14.815944 1.032493e-18
## Agriculture          0.1241776  0.0810977  1.531210 1.328763e-01
## factor(CatholicBin)1  7.8843292  3.7483622  2.103406 4.118221e-02
```

Segundo o modelo, 7.88 é a mudança esperada no intercepto da relação linear entre agricultura e fertilidade quando comparamos não-católicos a católicos.

Exemplo: conjunto de dados `swiss`



Exemplo: conjunto de dados `swiss`

Podemos também considerar um modelo que permite diferentes interceptos e diferentes coeficientes angulares (retas não paralelas). Isto é obtido considerando termo de **interação**.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

Agora, quando $X_{i2} = 0$:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i$$

e quando $X_{i2} = 1$:

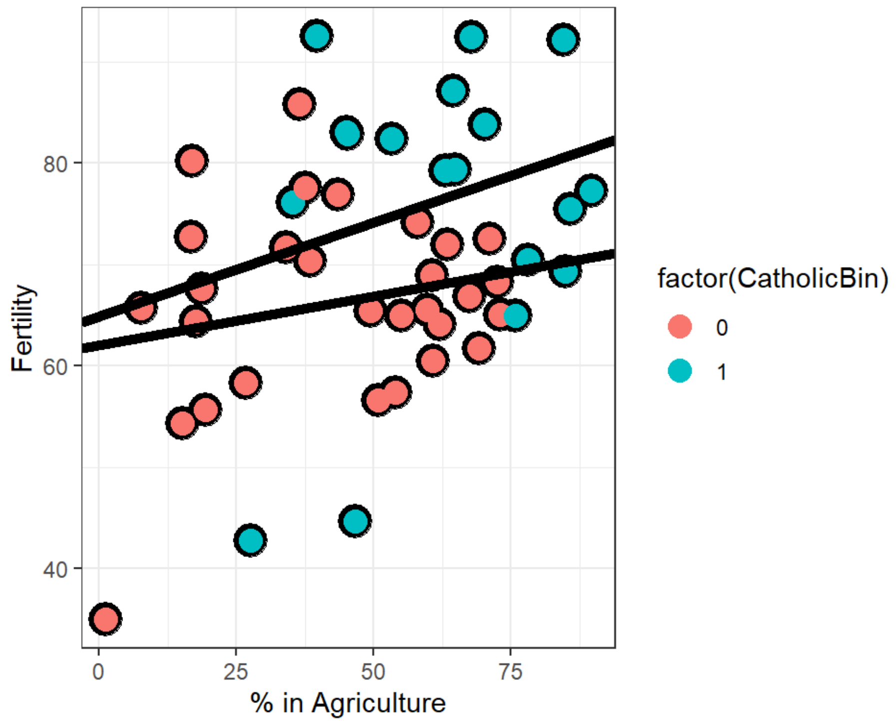
$$Y_i = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_{i1} + \varepsilon_i$$

Exemplo: conjunto de dados `swiss`

```
summary(lm(Fertility ~ Agriculture * factor(CatholicBin), data = swiss))$coef
```

```
##                                    Estimate  Std. Error    t value
## (Intercept)                     62.04993019  4.78915566 12.9563402
## Agriculture                      0.09611572  0.09881204  0.9727127
## factor(CatholicBin)1             2.85770359 10.62644275  0.2689238
## Agriculture:factor(CatholicBin)1  0.08913512  0.17610660  0.5061430
##                                      Pr(>|t|)
## (Intercept)                      1.919379e-16
## Agriculture                      3.361364e-01
## factor(CatholicBin)1             7.892745e-01
## Agriculture:factor(CatholicBin)1 6.153416e-01
```

Exemplo: conjunto de dados `swiss`



Exemplo: conjunto de dados `swiss`

Segundo o modelo ajustado, 2.8577 é a mudança esperada estimada no intercepto da reta de relação entre `Agriculture` e `Fertility` quando comparamos não católicos a católicos.

O termo de interação 0.9891 é a mudança esperada estimada no coeficiente angular.

O intercepto estimado entre os não-católicos é 62.04993 e o intercepto estimado entre os católicos é 62.04993 + 2.85770.

O coeficiente angular da relação entre `Agriculture` e `Fertility` para não-católicos é 0.09612 + 0.08914.

O coeficiente angular da relação entre `Agriculture` e `Fertility` para católicos é 0.09612.

Formas Quadráticas

$$\mathbf{Y}^T\mathbf{A}\mathbf{Y} = \sum_i \sum_j a_{ij} Y_i Y_j \qquad a_{ij} = a_{ji}$$

Exemplos:

$$SQT = \mathbf{Y}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right]\mathbf{Y}$$

$$SQE = \mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$SQReg = \mathbf{Y}^T\left[\mathbf{H} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right]\mathbf{Y}$$

Teorema de Cochran

Seja $X_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ e suponha que

$$\sum_{i=1}^{n} X_i^2 = Q_1 + Q_2 + \ldots + Q_k$$

em que

$$Q_i = \mathbf{X}^T \mathbf{A}_i \mathbf{X}$$

$rank(A_i) = r_i$ e $r_1 + r_2 + \ldots r_k = n$. Então temos que:

- $Q_1, Q_2, \ldots, Q_k$ são independentes
- $Q_i \sim \sigma^2 \chi^2(r_i)$, $i = 1, 2, \ldots, k$.

ANOVA: Regressão Linear Múltipla

Fonte de Variação	gl	SQ	QM
Regressão	$p-1$	$SQReg = \mathbf{Y}^T\left[\mathbf{H} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right]\mathbf{Y}$	$SQReg/(p-1)$
Erro	$n-p$	$SQE = \mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y}$	$SQE/(n-p)$
Total (ajustada)	$n-1$	$\mathbf{Y}^T\left[\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right]\mathbf{Y}$	

Teste F

- H_0: $\beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0$.
- H_1: pelo menos um $\beta_k \neq 0$, $k = 1, 2, \ldots, p-1$.

Estatística do teste:

$$F^* = \frac{SQReg/(p-1)}{SQE/(n-p)} \overset{\text{sob } H_0}{\sim} F_{p-1,n-p}$$

Intervalo de Confiança para β_k

Um intervalo de $100(1-\alpha)\%$ de confiança para β_k é dado por:

$$IC(\beta_k, 1-\alpha) = \left[\hat{\beta}_k - t_{n-p,\alpha/2}\sqrt{\widehat{Var(\hat{\beta}_k)}};\right.$$

$$\left.\hat{\beta}_k + t_{n-p,\alpha/2}\sqrt{\widehat{Var(\hat{\beta}_k)}}\right]$$

Teste de hipótese para β_k

- H_0: $\beta_k = 0$.
- H_1: $\beta_k \neq 0$.

Estatística do teste:

$$t^* = \frac{\hat{\beta}_k}{\sqrt{\widehat{Var(\hat{\beta}_k)}}} \overset{\text{sob } H_0}{\sim} t_{n-p}$$

Agradecimento

- Slides criados por Samara F Kiihl / IMECC / UNICAMP

Leitura

- Applied Linear Statistical Models: Capítulo 6.
- Weisberg - Applied Linear Regression: Capítulos 3, 4 e seção 5.1.
- Faraway - Linear Models with R: Capítulo 5.
- Caffo - Regression Models for Data Science in R: Multivariable regression analysis, Multivariable examples and tricks.