

Introdução aos modelos de regressão múltipla e análise de dados: parte 1

Prof. Caio Azevedo

Ajuste de modelos de regressão linear simples normais homocedásticos no R

- Função *lm*.
- Comando geral $lm(y \sim x_1)$, y : variável resposta, x_1 : variável explicativa.
- Modelo sem intercepto $lm(y \sim -1 + x_1)$, y : variável resposta, x_1 : variável explicativa.

Exemplo 1: sem considerar as etiologias cardíacas

$$Y_i = \beta_0 + \beta_1 x_i + \xi_i$$

Parâmetro	Estimativa	EP	IC(95%)	Estat. t	p-valor
β_0	6,563	0,356	[5,859 ; 7,268]	18,434	<0,0001
β_1	0,085	0,006	[0,072 ; 0,100]	12,516	<0,0001

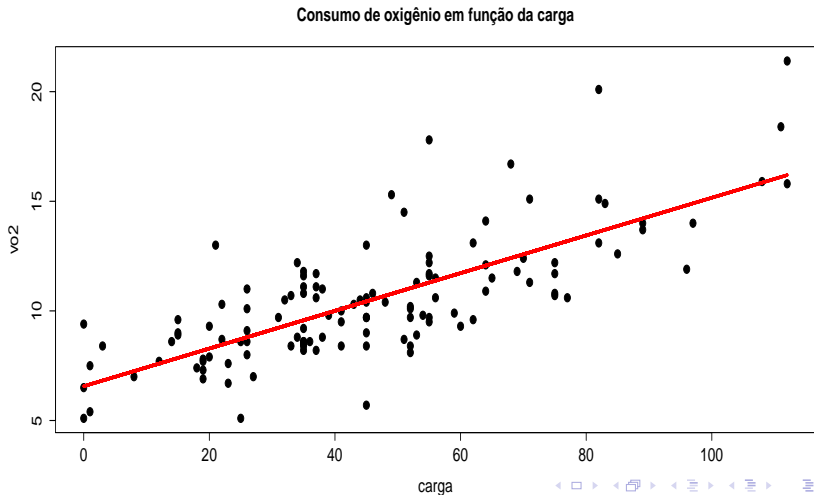
- Os dois parâmetros são diferentes de 0.
- A carga influencia positivamente o consumo de oxigênio.

Exemplo 1: sem considerar as etiologias cardíacas

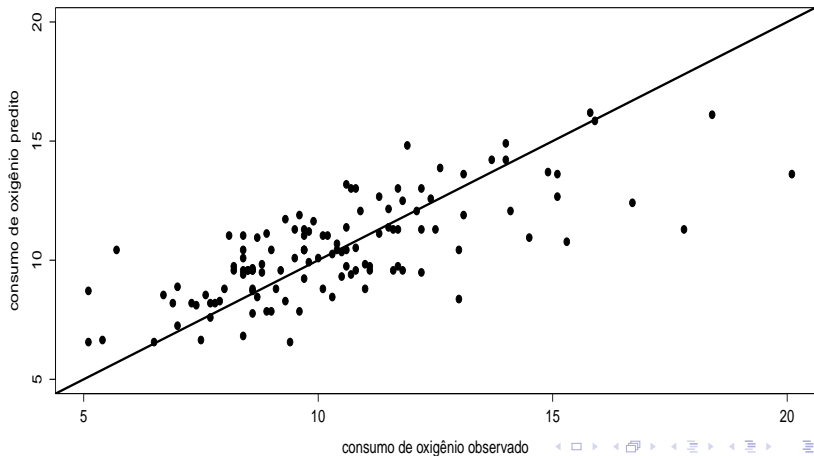
$$Y_i = \beta_0 + \beta_1 x_i + \xi_i \text{ (cont.)}$$

- O consumo de oxigênio para pacientes submetidos à carga 0 tende a se apresentar entre 5,859 e 7,268 ml/(kg.min).
- Por outro lado, o aumento esperado no consumo para o aumento em uma unidade da carga tende a se apresentar entre 0,072 e 0,100 ml/(kg.min).
- **A etapa de verificação de qualidade de ajuste do modelo, que deve preceder a sua utilização para fins inferenciais, será discutida posteriormente. Isso vale para todos os exemplos que veremos.**

Dispersão entre carga e consumo e reta ajustada



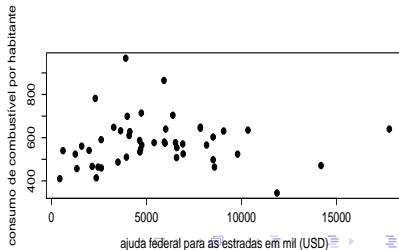
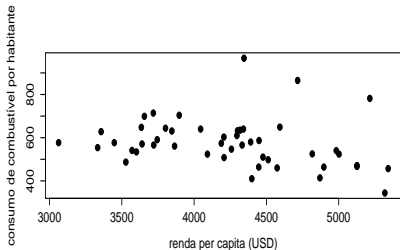
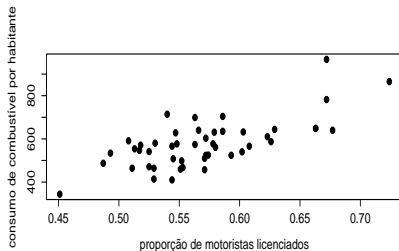
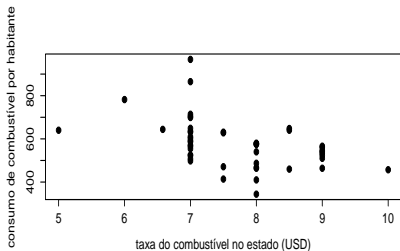
Consumos de oxigênio observado e predito modelo



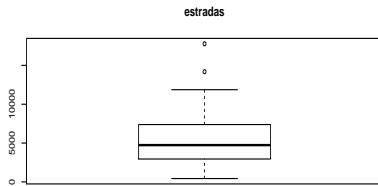
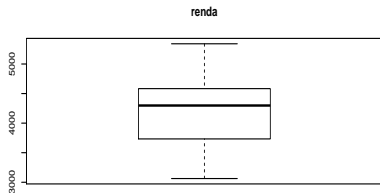
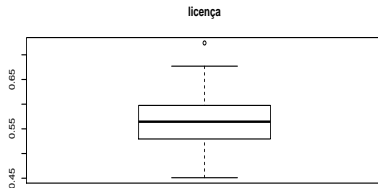
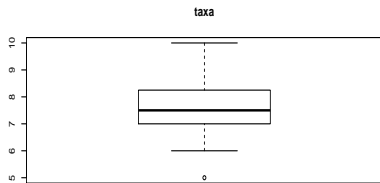
Exemplo 4: consumo de combustível

- Dos 48 estados (contíguos) dos Estados Unidos (em um certo ano) mediu-se:
 - Taxa do combustível no estado em USD - taxa.
 - Proporção de motoristas licenciados - licença.
 - Renda per capita em USD - renda.
 - Ajuda federal para as estradas em mil USD - estradas.
 - Consumo de combustível por habitante - consumo.
- Objetivo: tentar explicar o consumo de combustível em função das outras variáveis. Fonte de consulta: Paula (2013). Fonte original (Gray, 1989).

Diagrama de dispersão entre consumo e as var. explicativas



Box plot das variáveis explicativas



Medidas resumo das variáveis

Variável	Média	DP	CV(%)	Min.	Mediana	Max.
consumo	576,77	111,89	19,40	344,00	568,50	968,00
taxa	7,67	0,95	12,40	5,00	7,50	10,00
licença	0,57	0,06	9,73	0,45	0,56	0,72
renda	4241,83	573,62	13,52	3063,00	4298,00	5342,00
estradas	5565,42	3491,51	62,74	431,00	4735,50	17782,00

Correlações entre as variáveis

	consumo	taxa	licença	renda	estradas
consumo	.	-0,45	0,70	-0,24	0,02
taxa	.	.	-0,29	0,01	-0,52
licença	.	.	.	0,16	-0,06
renda	0,05
estradas

Exemplo 4: consumo de combustível

Modelo 1

$$\text{consumo}_i = \beta_0 + \beta_1 \text{taxa}_i + \beta_2 \text{licença}_i + \beta_3 \text{renda}_i + \beta_4 \text{estradas}_i + \xi_i, i = 1, \dots, 48$$

≡

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \xi_i, i = 1, \dots, 48$$

- $\xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- β_0 : consumo esperado para estados com valor zero para todas as covariáveis simultaneamente.
- β_j : incremento (positivo ou negativo) no consumo esperado para o aumento em uma unidade da variável $j, j = 1, 2, 3, 4$, mantendo-se as outras fixas.

Exemplo 4: consumo de combustível

Modelo 2

$$Y_i = \beta_0 + \beta_1(x_{1i} - \bar{x}_1) + \beta_2(x_{2i} - \bar{x}_2) + \beta_3(x_{3i} - \bar{x}_3) + \beta_4(x_{4i} - \bar{x}_4) + \xi_i, \quad (1)$$

$$i = 1, \dots, 48, \bar{x}_j = \frac{1}{48} \sum_{i=1}^{48} x_{ij}, j = 1, 2, 3, 4$$

- $\xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.
- β_0 : consumo esperado para estados com valor de cada covariável igual à sua respectiva média.
- β_j : incremento (positivo ou negativo) no consumo esperado para o aumento em uma unidade da variável $j, j = 1, 2, 3, 4$, mantendo-se as outras fixas.

Inferência

- Já vimos como ajustar o modelo (estimar os parâmetros), através da metodologia de MQO.
- Veremos agora como construir intervalos de confiança e testar hipóteses.
- Dos resultados apresentados nos slides “Introdução aos modelos de regressão normais lineares”, temos que

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 \psi_j), \quad \frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-p)}^2 \text{ e } \hat{\beta}_j \perp \frac{(n-p)\hat{\sigma}^2}{\sigma^2}.$$

- Logo, $\frac{\hat{\beta}_j - \beta_j}{\sqrt{\hat{\sigma}^2 \psi_j}} \sim t_{(n-p)}$, portanto (considerando-se $P(X \leq t_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$, $X \sim t_{(n-p)}$), temos que
- $$IC(\beta_j, \gamma) = \left[\hat{\beta}_j - t_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}^2 \psi_j}; \hat{\beta}_j + t_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}^2 \psi_j} \right]$$

Testes de hipóteses

- Suponha que queremos testar $H_0 : \beta_j = \beta_{j0}$ vs $H_1 : \beta_j \neq \beta_{j0}$, para algum j , em que β_{j0} é um valor fixado.
- Estatística do teste $T_t = \frac{\hat{\beta}_j - \beta_{j0}}{\sqrt{\hat{\sigma}^2 \psi_j}}$, em que $\hat{\beta}_j$ é o estimador de MQO de β_j e $\hat{\sigma}^2 = \frac{1}{n-p} (\mathbf{Y} - \mathbf{X}\hat{\beta})' (\mathbf{Y} - \mathbf{X}\hat{\beta})$.

Testes de hipóteses

- Sob H_0 , $T_t \sim t_{(n-p)}$. Assim, rejeita-se H_0 se $|t_t| \geq t_c$, em que $t_t = \frac{\tilde{\beta}_j - \beta_{j0}}{\sqrt{\tilde{\sigma}^2 \psi_j}}$ e $P(X \geq t_c | H_0) = \alpha/2$, $X \sim t_{(n-p)}$, $\tilde{\beta}_j$ é o j -ésimo elemento de $\tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}\mathbf{y}$ e $\tilde{\sigma}^2 = \frac{1}{n-p} (\mathbf{y} - \mathbf{X}\tilde{\beta})' (\mathbf{y} - \mathbf{X}\tilde{\beta})$
- De modo equivalente, rejeita-se H_0 se p-valor $\leq \alpha$, em que p-valor $= 2P(X \geq |t_t| | H_0)$, $X \sim t_{(n-p)}$

ANOVA para modelos de regressão

- Suponha o seguinte modelo:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_{p-1} x_{(p-1)i} + \xi_i, \xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$$

- Logo $Y_i \stackrel{ind.}{\sim} N(\beta_0 + \sum_{j=1}^{p-1} \beta_j x_{ji}, \sigma^2)$.
- O modelo acima define uma média (condicional aos valores de $x_{ji}, j = 1, \dots, p - 1; i = 1, \dots, n$) para cada observação Y_i .
- Defina $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_{p-1} x_{(p-1)i}$ (valor predito pelo modelo).
- O resíduo é definido por $R_i = \hat{\xi}_i = Y_i - \hat{Y}_i$.

- Nosso objetivo é considerar um modelo que explique adequadamente a variabilidade dos dados, ou seja, um modelo para o qual os resíduos sejam “pequenos”.
- Pode-se provar que, a soma de quadrados total $SQT = \sum_{i=1}^n (Y_i - \bar{Y})^2$, pode ser decomposta como:

$$SQT = \underbrace{\sum_{i=1}^n (\bar{Y} - \hat{Y}_i)^2}_{SQM} + \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{SQR}$$

- Assim, quanto maior for o valor de SQM em relação à SQR, maior será a contribuição da parte sistemática para explicar a variabilidade dos dados. Portanto, mais “provável” que exista (pelo menos um) $\beta_j \neq 0, j = 1, \dots, p - 1$.

- SQM: soma de quadrados do modelo ; SQR: soma de quadrados dos resíduos.
- Portanto, como estatística de teste, podemos comparar, de alguma forma, as somas de quadrados (do modelo e dos resíduos).
- Pergunta: como construir uma estatística de teste adequada ?
Adequada: que serve para testar as hipóteses de interesse, que possua distribuição conhecida (sob H_0 e sob H_1) e que tenha um “razoável” poder (assumindo-se ser possível fixar o nível de significância α).

- Lembrando da forma matricial do modelo: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$, pode-se demonstrar que:
 - $\text{SQT} = \mathbf{Y}' (\mathbf{I} - n^{-1}\mathbf{J}) \mathbf{Y}$, em que $\mathbf{J} = \mathbf{1}\mathbf{1}'$.
 - $\text{SQM} = \mathbf{Y}' (\mathbf{H} - n^{-1}\mathbf{J}) \mathbf{Y}$, em que $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ (chamada de matriz de projeção ou matriz “hat”).
 - $\text{SQR} = \mathbf{Y}' (\mathbf{I} - \mathbf{H}) \mathbf{Y}$.
- Pode-se provar que as matrizes $\mathbf{A} = \mathbf{H} - n^{-1}\mathbf{J}$ e $\mathbf{B} = \mathbf{I} - \mathbf{H}$ são ortogonais, ou seja, $\mathbf{AB} = \mathbf{0}$ (provando-se que $n^{-1}\mathbf{HJ} = n^{-1}\mathbf{J}$).
- Dizemos que $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$ e $\mathbf{R} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y} = (\mathbf{I} - \mathbf{H})$ projetam \mathbf{Y} em dois subespaços ortogonais, pois $\mathbf{H}(\mathbf{I} - \mathbf{H}) = \mathbf{0}$.
- Um caminho para a construção de estatísticas do teste é o estudo das propriedades de formas quadráticas aleatórias (normais).

Distribuição de formas quadráticas normais

- Seja $\mathbf{Z} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Dizemos, então, que $Y = \mathbf{Z}'\mathbf{A}\mathbf{Z}$ é uma forma quadrática (escalar) normal. Em geral, \mathbf{A} é uma matriz não aleatória, simétrica e real. Em todos os resultados apresentados no curso, consideraremos que a matriz \mathbf{A} é não aleatória e real (eventualmente a suposição de simetria pode ser relaxada).
- Pode-se provar, por exemplo, que:

$$Y = (\mathbf{Z} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \boldsymbol{\mu}) \sim \chi_{(p)}^2 \quad (2)$$

- Vamos demonstrar a validade do resultado dado pela equação (2).
- Por simplicidade, admita que $\boldsymbol{\mu} = \mathbf{0}$.

- Assim, temos que :

$$\begin{aligned}
 M_Y(t) &= \mathcal{E}(e^{Yt}) = \int_{\mathcal{R}^p} e^{t'y} f_Z(\mathbf{z}) d\mathbf{z} \\
 &= \int_{\mathcal{R}} \int_{\mathcal{R}} \dots \int_{\mathcal{R}} e^{t\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}} |\boldsymbol{\Sigma}|^{-1/2} (2\pi)^{-p/2} \exp\left\{-\frac{1}{2}\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}\right\} d\mathbf{z} \\
 &= \int_{\mathcal{R}} \int_{\mathcal{R}} \dots \int_{\mathcal{R}} |\boldsymbol{\Sigma}|^{-1/2} (2\pi)^{-p/2} \exp\left\{-\frac{1}{2}\mathbf{z}'\boldsymbol{\Sigma}^{-1}\mathbf{z}(1-2t)\right\} d\mathbf{z}
 \end{aligned}$$

- Considerando a mesma transformação anteriormente utilizada ($\mathbf{w} = \boldsymbol{\Psi}\mathbf{z}$), em que $\boldsymbol{\Sigma} = \boldsymbol{\Psi}\boldsymbol{\Psi}'$, temos que:

$$\begin{aligned}
 M_Y(t) &= \prod_{i=1}^p \int_{\mathcal{R}} \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}w_i^2(1-2t)\right\}}_{N(0, (1-2t)^{-1})} dw_i \\
 &= (1-2t)^{-p/2}
 \end{aligned}$$

- A qual corresponde à fgm de uma distribuição $\chi^2(p)$

Distribuição t de Student não-central

- Defina $T = \frac{Z+\mu}{\sqrt{V/\nu}}$, em que $Z \perp V$, $Z \sim N(0, 1)$ e $V \sim \chi_\nu^2$. Dizemos que T tem distribuição t de student não central, com ν graus de liberdade e parâmetro de não centralidade μ .
- Uma forma de se apresentar a densidade é:

$$f_T(t) = \frac{\nu^\nu/2 e^{-\frac{\nu\mu^2}{2(t^2+\nu)}}}{\sqrt{\pi}\Gamma(\nu/2)2^{\frac{\nu-1}{2}}(t^2 + \nu)^{(\nu+1)/2}} \times \left(\int_0^\infty y^\nu \exp \left\{ -\frac{1}{2} \left(y - \frac{\mu t}{\sqrt{t^2 + \nu}} \right)^2 \right\} dy \right) \mathbb{1}_{(-\infty, \infty)}(t)$$

Distribuição qui-quadrado não central

- Sejam $X_1, X_2, \dots, X_n \stackrel{ind}{\sim} N(\mu_i, \sigma_i^2)$.
- Defina $Y = \sum_{i=1}^n \left(\frac{X_i}{\sigma_i}\right)^2$. Dizemos então que Y tem distribuição qui-quadrado não central com n graus de liberdade e parâmetro de não-centralidade $\delta = \sum_{i=1}^n \left(\frac{\mu_i}{\sigma_i}\right)^2$
- Notação $Y \sim \chi_{(n,\delta)}^2$, cuja fdp é dada por

$$f_Y(y) = \sum_{i=0}^{\infty} \frac{e^{-\delta/2} (\delta/2)^i}{i!} f_{W_{n+2i}}(y) \mathbb{1}_{(0,\infty)}(y),$$

em que $W_{n+2i} \sim \chi_{(n+2i)}^2$

- Se $\delta = 0$, então $Y \sim \chi_{(n)}^2$.

Distribuição F não central

- Seja V uma outra v.a., independente de Y , $V \sim \chi^2_{(m)}$.
- Defina $F = \frac{Y/n}{V/m}$. Então, F tem distribuição F não central com graus de liberdade, n e m e parâmetro de não centralidade δ .
- Notação $F \sim \chi^2_{(n,m,\delta)}$, cuja fdp é dada por

$$f_F(f) = \sum_{i=0}^{\infty} \frac{e^{-\delta/2}(\delta/2)^i}{\beta(m/2, n/2 + i)i!} \left(\frac{n}{m}\right)^{n/2+i} \left(\frac{m}{m+nf}\right)^{(n+m)/2+i} \\ \times f^{n/2-1+i} \mathbb{1}_{(0,\infty)}(f)$$

em que $\beta(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$

Principais teoremas de formas quadráticas normais

- Seja $\mathbf{Y} \sim N_p(\mathbf{0}, \mathbf{I})$ e \mathbf{A} uma matriz simétrica e não-aleatória. Então

$$\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi^2_{[r(\mathbf{A})]}$$

se, e somente se \mathbf{A} for idempotente, em que $r(\mathbf{A}) = \text{rank}(\mathbf{A})$.

- Seja $\mathbf{Y} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$ e \mathbf{A} uma matriz simétrica e não-aleatória. Então

$$\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi^2_{[r(\mathbf{A})]}$$

se, e somente se $\mathbf{A}\mathbf{\Sigma}$ for idempotente.

Cont.

- Seja $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \mathbf{I})$ e \mathbf{A} uma matriz simétrica e não-aleatória. Então

$$\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi^2_{[r(\mathbf{A}), \delta = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}]}$$

se, e somente se \mathbf{A} for idempotente.

- Seja $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e \mathbf{A} uma matriz simétrica e não-aleatória. Então

$$\mathbf{Y}'\mathbf{A}\mathbf{Y} \sim \chi^2_{[r(\mathbf{A}), \delta = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}]}$$

se, e somente se $\mathbf{A}\boldsymbol{\Sigma}$ for idempotente.

- Naturalmente, se $\delta = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} = 0$, as distribuições passam a ser qui-quadrados centrais.

Cont.

- Seja $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e \mathbf{A} e \mathbf{B} matrizes não-aleatórias. Então $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ e $\mathbf{B}\mathbf{Y}$ são independentes se, e somente se $\mathbf{B}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{0}$.
- Seja $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e \mathbf{A} e \mathbf{B} matrizes não-aleatórias. Então $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ e $\mathbf{Y}'\mathbf{B}\mathbf{Y}$ são independentes se, e somente se $\mathbf{B}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{0}$.
- Demonstrações: pesquisar a respeito.

Voltando ao problema

- Lembremos que:

- $SQM = \mathbf{Y}' (\mathbf{H} - n^{-1}\mathbf{J}) \mathbf{Y}$, em que $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

- $SQR = \mathbf{Y}' (\mathbf{I} - \mathbf{H}) \mathbf{Y}$.

- $\mathbf{A} = \mathbf{H} - n^{-1}\mathbf{J}$ e $\mathbf{B} = \mathbf{I} - \mathbf{H}$

- $\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$.

- Defina $W = SQM/\sigma^2 = \mathbf{Y}' \left(\frac{\mathbf{H} - n^{-1}\mathbf{J}}{\sigma^2} \right) \mathbf{Y}$ e

$$V = SQR/\sigma^2 = \mathbf{Y}' \left(\frac{\mathbf{I} - \mathbf{H}}{\sigma^2} \right) \mathbf{Y}.$$

Voltando ao problema

- Temos que $\frac{\mathbf{A}}{\sigma^2}(\sigma^2\mathbf{I})\frac{\mathbf{B}}{\sigma^2} = \frac{1}{\sigma^2}\mathbf{AB} = \mathbf{0}$ (são ortogonais). Então $W \perp V$.
- Por outro lado, $\frac{\mathbf{A}}{\sigma^2}(\sigma^2\mathbf{I}) = \mathbf{A}$ e $\frac{\mathbf{B}}{\sigma^2}(\sigma^2\mathbf{I}) = \mathbf{B}$. Como \mathbf{A} e \mathbf{B} são matrizes idempotentes, logo \mathbf{AI} e \mathbf{BI} , também o são.
- Pode-se provar ainda que $\mathcal{E}(V) = n - p$ e que, sob H_0 , $\mathcal{E}(W) = p - 1$ (os parâmetros de não centralidade são iguais à 0). Basta usar a fórmula da esperança de formas quadráticas dos slides “A distribuição normal multivariada”.

- Assim, $W \sim \chi^2_{(p-1)}$ (sob H_0) e $V \sim \chi^2_{(n-p)}$, em que $W \perp V$.
- Logo, $F_t = \frac{W/(p-1)}{V/(n-p)} \sim F_{[(p-1),(n-p)]}$, sob H_0 .
- Assim, ao se optar por usar formas quadráticas para se testar hipóteses, é necessário verificar:
 - A distribuição de \mathbf{Y} .
 - Algumas propriedades das matrizes núcleo (com a matriz de covariâncias de \mathbf{Y}).
 - As esperanças das formas quadráticas (sob H_0 e H_1).
- Portanto, rejeita-se H_0 se $f_t \geq f_c$ ou, analogamente, se $p - \text{valor} = P(X \geq f_t | H_0) \leq \alpha$, em que f_t é o valor calculado da estatística F_t e $P(X \geq f_c | H_0) = \alpha, X \sim F_{(p-1, n-p)}$.

Tabela de ANOVA (matricial)

- Para testar $H_0 : \beta_1 = \beta_2 = \dots = \beta_{(p-1)} = 0$ vs $H_1 : \text{Há pelo menos uma diferença.}$

FV	SQ	GL	QM	Estatística F	pvalor
Modelo	$SQM = \mathbf{Y}'\mathbf{A}\mathbf{Y}$	p-1	$QMM = \frac{SQM}{p-1}$	$F_t = \frac{QMM}{QMR}$	$P(X \geq f_t H_0)$
Resíduo	$SQR = \mathbf{Y}'\mathbf{B}\mathbf{Y}$	n-p	$QMR = \frac{SQR}{n-p}$		
Total	SQT	n-1			

FV: fonte de variação, SQ: soma de quadrados, GL: graus de liberdade, QM: quadrado médio.

Comentários

- Quando não há intercepto no modelo ou quando há mais de um, eventualmente, o procedimento descrito anteriormente pode não ser adequado para se testar as hipóteses de interesse.
- Alternativas: o procedimento pode ser adaptado ou pode-se usar outros tipo de testes como os “do tipo” $\mathbf{C}\beta$ (o qual veremos mais adiante).
- Os testes do tipo $\mathbf{C}\beta$ são úteis (como veremos) para testar outras hipóteses de interesse como, por exemplo, descobrir quais componentes do vetor β são diferentes de 0.

Ajuste de modelos de regressão linear múltipla normais homocedásticos no R

- Função *lm*.
- Comando geral $lm(y \sim x_1 + x_2 + \dots + x_{(p-1)})$, y : variável resposta, $x_1, x_2, \dots, x_{(p-1)}$: variáveis explicativas.
- Modelo sem intercepto $lm(y \sim -1 + x_1 + x_2 + \dots + x_{(p-1)})$, y : variável resposta, $x_1, x_2, \dots, x_{(p-1)}$: variáveis explicativas.

Cont.

- Suponha que o objeto “result” contenha a saída da aplicação da função *lm*

Quantidade	Comando
$\tilde{\beta}$ (vetor com as estimativas dos parâmetros de regressão)	<code>result\$coef</code>
$\tilde{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$ (estimativa da matriz de variâncias e covariâncias dos estimadores de β)	<code>vcov(result)</code>
$\tilde{\sigma}$ (estimativa do desvio-padrão residual)	<code>(summary(result))\$sigma</code>
\mathbf{X} (matriz de planejamento)	<code>model.matrix(result)</code>
$\tilde{\mathbf{y}} = \mathbf{X}\tilde{\beta}$ (valor predito modelo)	<code>fitted.values(result)</code>

Exemplo 4: consumo de combustível

- Modelo com as covariáveis centradas, equação (1).
- Tabela ANOVA

FV	GL	SQ	QM	Estatística F	p-valor
Modelo	4	399316,51	99829,13	22,71	<0,0001
Resíduo	43	189049,97	4396,51	-	-

- Pelo menos um dos coeficientes de regressão $(\beta_1, \beta_2, \beta_3, \beta_4)'$ é diferente de 0. Também se diz que “existe regressão”.

Exemplo 4: ajuste do modelo

Parâmetro	Estimativa	EP	Estat. t	p-valor
β_0	576,771	9,571	60,266	<0,0000
β_1 (taxa)	-34,790	12,970	-2,682	<0,0103
β_2 (licença)	1336,450	192,299	6,950	<0,0000
β_3 (renda)	-0,067	0,017	-3,867	<0,0004
β_4 (estradas)	-0,002	0,0034	-0,716	<0,4780

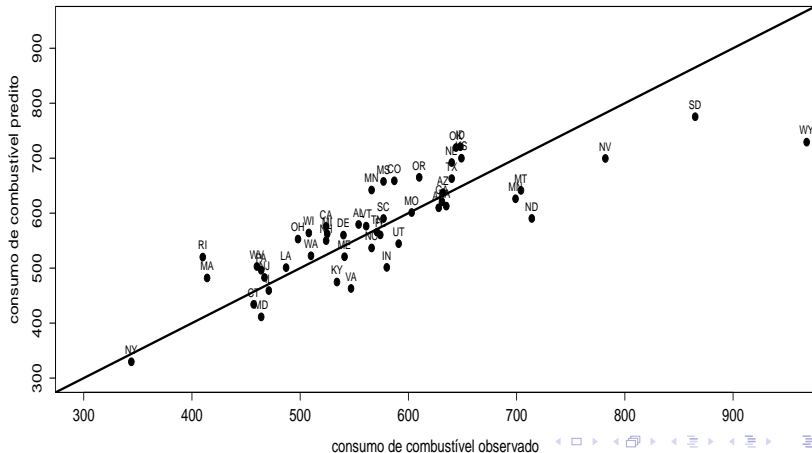
A variável “estradas” não contribui no modelo para explicar a variabilidade do consumo. Vamos ajustar um modelo (reduzido) descartando essa variável.

Exemplo 4: ajuste do modelo reduzido

Par.	Estim.	EP	IC(95%)	Estat. t	p-valor
β_0	576,771	9,517	[557,590 ; 595,952]	60,602	<0,0001
β_1 (taxa)	-29,484	10,584	[-50,814 ; -8,154]	-2,786	0,0079
β_2 (licença)	1374,768	183,670	[1004,607 ; 1744,930]	7,485	<0,0001
β_3 (renda)	-0,068	0,017	[-0,102 ; -0,034]	-3,999	0,0002

Comentários: todas as variáveis explicativas são significativas, sendo que o impacto da taxa e da renda, no consumo, é negativo enquanto que o impacto da licença, no consumo, é positivo. Para medir qual variável explicativa mais impacta no consumo o ideal é padronizá-las.

Consumos de combustível observado e predito modelo



Exemplo 3: efeito do fósforo na produção de milho

- Vamos ajustar dois modelos e compará-los.
- Modelo linear (reta) e modelo quadrático (parábola): ambos fazem parte da família de modelos de regressão linear.

Exemplo 3: medidas resumo por nível de fósforo

Nível de fósforo	Média	DP	CV(%)	Var.	Mínimo	Máximo
0 kg/ha	4,647	2,052	44,156	4,211	2,380	6,770
25 kg/ha	8,255	1,410	17,076	1,987	6,150	9,100
50 kg/ha	8,300	0,951	11,461	0,905	6,920	9,070
75 kg/ha	9,350	0,700	7,482	0,489	8,660	10,240
100 kg/ha	9,640	0,433	4,492	0,188	9,140	10,170

Modelo linear 1: reta

$$Y_i = \beta_0 + \beta_1 x_i + \xi_i, i = 1, 2, \dots, 20$$

- x_i : quantidade de fósforo ministrada a i -ésima parcela.
- β_0 : valor esperado (média) da produção de milho quando a quantidade de fósforo aplicada é igual à 0.
- β_1 : incremento no valor esperado da produção de milho quando a quantidade de fósforo aplicada aumenta em uma unidade.
- $\xi_i \stackrel{iid}{\sim} N(0, \sigma^2)$.

Modelo linear 2: parábola

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \xi_i, i = 1, 2, \dots, 20$$

- x_i : quantidade de fósforo ministrada a i -ésima parcela.
- β_0 : valor esperado (média) da produção de milho quando a quantidade de fósforo aplicada é igual à 0.
- A interpretação isolada dos parâmetros β_1 e β_2 é complicada mas, podemos dizer que $\frac{-\beta_1}{2\beta_2}$ é o máximo (ou mínimo) do valor esperado da produção de milho.
- $\xi_i \stackrel{iid}{\sim} N(0, \sigma^2)$.

ANOVA para o modelo quadrático

■ Tabela ANOVA

FV	GL	SQ	QM	Estatística F	p-valor
Modelo	2	58,16	29,08	17,18	< 0,0001
Resíduo	17	28,77	1,70		

- Pelo menos um dos coeficientes de regressão $(\beta_1, \beta_2)'$ é diferente de 0. Também se diz que “existe regressão”.

Resultados dos ajustes dos modelos

Modelo linear

Parâmetro	Est.	EP	IC(95%)	Estat. t	p-valor
β_0	5,8225	0,5615	[4,6429 ; 7,0021]	10,3701	< 0,0001
β_1	0,0443	0,0092	[0,0251 ; 0,0636]	4,8338	0,0001

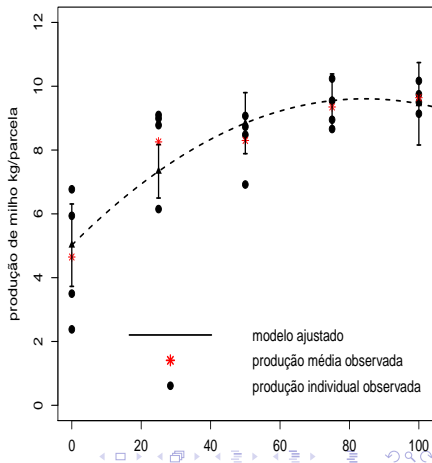
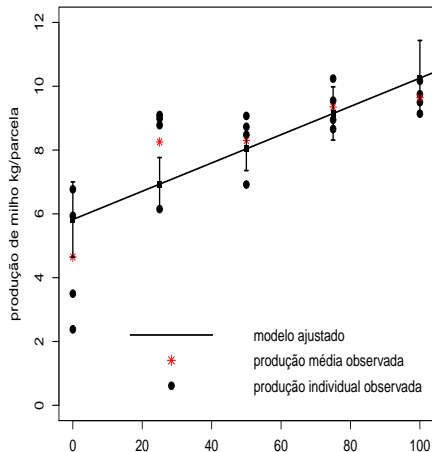
Modelo quadrático

Parâmetro	Est.	EP	IC(95%)	Estat. t	p-valor
β_0	5,0182	0,6122	[3,7266 ; 6,3098]	8,1971	0,0000
β_1	0,1087	0,0290	[0,0475 ; 0,1699]	3,7460	0,0016
β_2	-0,0006	0,0003	[-0,0012 ; -0,0001]	-2,3132	0,0335

Intervalos de confiança para as médias obtidas a partir do modelo

- Temos que $\mu_i = \mathcal{E}(Y_i) = \mathbf{X}'_i \boldsymbol{\beta}$ em que \mathbf{X}'_i é a i -ésima linha da matriz \mathbf{X} .
- Em particular $\mu_i = \beta_0 + \beta_1 x_i$, no modelo linear e $\mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$, no modelo quadrático.
- Estimador pontual: $\hat{\mu}_i = \mathbf{X}'_i \hat{\boldsymbol{\beta}} \sim N(\mu_i, \sigma^2 \mathbf{X}'_i (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i)$.
- Pode-se provar que: $\frac{\hat{\mu}_i - \mu_i}{\sqrt{\hat{\sigma}^2 \mathbf{X}'_i (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i}} \sim t_{(n-p)}$.
- Assim, $IC[\mu_i; \gamma] = \left[\hat{\mu}_i - t_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}^2 \mathbf{X}'_i (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i}; \hat{\mu}_i + t_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}^2 \mathbf{X}'_i (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_i} \right]$, em que considerando-se $P(X \leq t_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$, $X \sim t_{(n-p)}$.

Ajuste dos modelos de regressão



Estimativa da quantidade de fósforo que retorna a máxima produtividade

- $\theta = -\frac{\beta_1}{2\beta_2}$.
- Estimador pontual $\hat{\theta} = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$.
- Para fazer inferências sobre θ com base em $\hat{\theta}$ precisamos da distribuição exata ou aproximada deste.
- Sabemos que $(\hat{\beta}_1, \hat{\beta}_2) \sim N_2((\beta_1, \beta_2), \sigma^2 \mathbf{\Omega})$ em que $\mathbf{\Omega} = (\mathbf{X}'\mathbf{X})_{[2:3,2:3]}^{-1}$ (submatriz obtida excluindo-se a primeira linha e a primeira coluna).
- Uma opção: método delta.

Método delta univariado

- Seja $\hat{\theta}$ uma variável aleatória de sorte que, para n suficientemente grande,

$$\hat{\theta} \approx N(\theta, \sigma^2).$$

- Defina $\hat{\tau} = g(\hat{\theta})$. Então, para n suficientemente grande, sob certas condições de regularidade,

$$\hat{\tau} \approx N(g(\theta), \sigma^2 [\psi(\theta)]^2),$$

em que $\psi(\theta) = \frac{d}{d\theta} g(\theta)$.

Método delta univariado

- Voltando ao exemplo, suponha que $\tau = \frac{1}{\beta_1}$. Então

$$\psi(\beta_1) = \frac{d}{d\beta_1} \mathbf{g}(\beta_1) = -\beta_1^{-2}.$$

- Assim, para n suficientemente grande, $\hat{\tau} = \frac{1}{\hat{\beta}_1} \approx N(\tau, \omega\beta_1^{-4})$, em que ω é a variância de $\hat{\beta}_1$.

Método delta multivariado

- Seja $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)'$ de sorte que, para n suficientemente grande

$$\hat{\boldsymbol{\theta}} \approx N_p(\boldsymbol{\theta}, \boldsymbol{\Sigma}).$$

- Defina $\hat{\boldsymbol{\tau}} = (g_1(\hat{\boldsymbol{\theta}}), g_2(\hat{\boldsymbol{\theta}}), \dots, g_r(\hat{\boldsymbol{\theta}}))'$, então, para n suficientemente grande, sob certas condições de regularidade,

$$\hat{\boldsymbol{\tau}} \approx N_r(\boldsymbol{\tau}, \boldsymbol{\Psi}\boldsymbol{\Sigma}\boldsymbol{\Psi}'),$$

em que $\boldsymbol{\tau} = (g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta}), \dots, g_r(\boldsymbol{\theta}))'$.

Método delta multivariado

■ e

$$\Psi = \begin{bmatrix} \frac{\partial}{\partial \theta_1} g_1(\boldsymbol{\theta}) & \frac{\partial}{\partial \theta_2} g_1(\boldsymbol{\theta}) & \frac{\partial}{\partial \theta_3} g_1(\boldsymbol{\theta}) & \cdots & \frac{\partial}{\partial \theta_p} g_1(\boldsymbol{\theta}) \\ \frac{\partial}{\partial \theta_1} g_2(\boldsymbol{\theta}) & \frac{\partial}{\partial \theta_2} g_2(\boldsymbol{\theta}) & \frac{\partial}{\partial \theta_3} g_2(\boldsymbol{\theta}) & \cdots & \frac{\partial}{\partial \theta_p} g_2(\boldsymbol{\theta}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \theta_1} g_r(\boldsymbol{\theta}) & \frac{\partial}{\partial \theta_2} g_r(\boldsymbol{\theta}) & \frac{\partial}{\partial \theta_3} g_r(\boldsymbol{\theta}) & \cdots & \frac{\partial}{\partial \theta_p} g_r(\boldsymbol{\theta}) \end{bmatrix}.$$

Voltando à estimativa da quantidade de fósforo que retorna a máxima produtividade

- Método Delta: sob os resultados obtidos e assumindo válidas as condições de regularidade, temos que $\hat{\theta} \approx N(\theta, \sigma_\theta^2)$, em que

$$\sigma_\theta^2 = \sigma^2 (\mathbf{\Delta})' (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{\Delta})$$

- $\mathbf{\Delta} = \left[0 \quad -\frac{1}{2\beta_2} \quad \frac{\beta_1}{2\beta_2^2} \right]'$.
- Logo $IC(\theta, \gamma) = \left[\hat{\theta} - z_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}_\theta^2}; \hat{\theta} + z_{\frac{1+\gamma}{2}} \sqrt{\hat{\sigma}_\theta^2} \right]$, em que $P(Z \leq z_{\frac{1+\gamma}{2}}) = \frac{1+\gamma}{2}$, $Z \approx N(0, 1)$ e $\hat{\sigma}_\theta^2 = \hat{\sigma}^2 (\hat{\mathbf{\Delta}})' (\mathbf{X}'\mathbf{X})^{-1} (\hat{\mathbf{\Delta}})$ e $\hat{\mathbf{\Delta}} = \left[0 \quad -\frac{1}{2\hat{\beta}_2} \quad \frac{\hat{\beta}_1}{2\hat{\beta}_2^2} \right]'$.
- Em nosso exemplo $\hat{\theta} = 84,45(16,20)$, $IC(\theta, 95\%) = [52,68; 116,20]$.

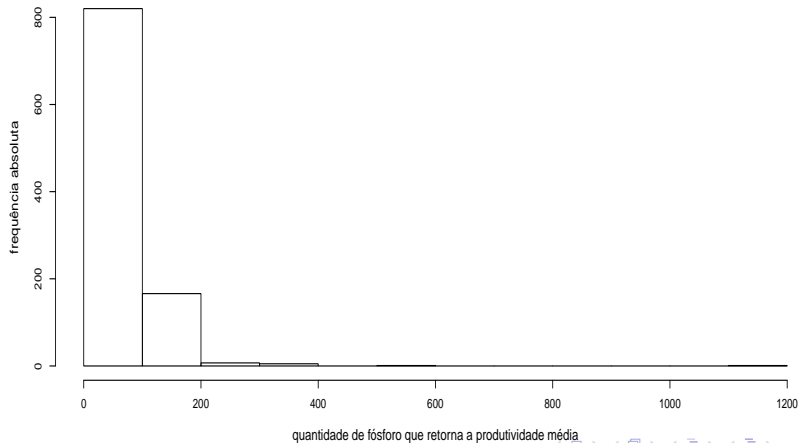
Estimativa da produtividade máxima

- De posse do valor (estimativa pontual) da quantidade de fósforo que retorna a produtividade máxima, podemos estimar a produtividade (média) máxima.
- Ou seja, queremos estimar: $\mu_M = \mu_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$ para $x_i = 84, 45$.
- Utilizando a metodologia apresentada no Slide 46, obtemos $\tilde{\mu}_M = 9,601(0,407)$ e $IC(\mu_M; 95\%) = [8,748; 10,463]$.

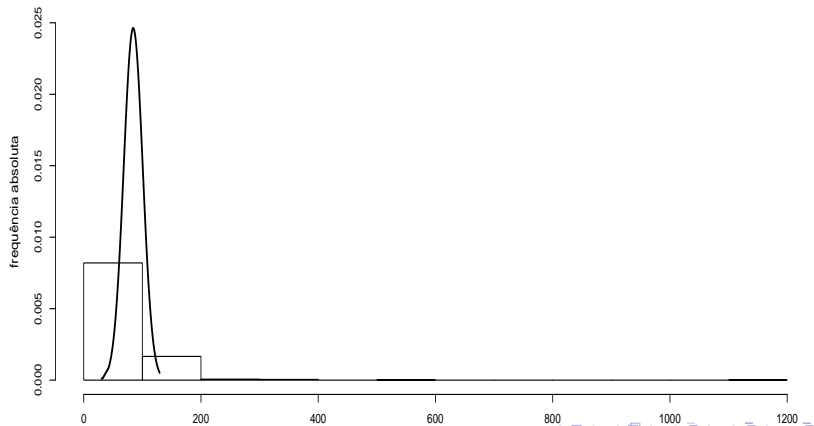
Estimativa da quantidade de fósforo que retorna a máxima produtividade via bootstrap não-paramétrico

- Para $b=1, \dots, B$ (número de amostras) execute os seguintes passos
 - 1 Selecione uma amostra aleatória (com reposição) de tamanho m do conjunto de dados original.
 - 2 Estime os parâmetros β via MQO com base nessa amostra.
 - 3 Obtenha a quantidade de fósforo que retorna a produtividade máxima com as estimativas obtidas no passo 2.
- Ao final ter-se-á uma amostra aleatória de tamanho B da distribuição do estimador da quantidade de fósforo de interesse.

Distribuição empírica do estimador $\hat{\theta}$



Distribuição empírica do estimador $\hat{\theta}$ e curva teórica da distribuição normal obtida pelo método delta



Comentários

- Resultados método Delta:

$$\tilde{\theta} = 84,45(16,20), IC(\theta, 95\%) = [52,68; 116,20].$$

- Resultados bootstrap:

$$\tilde{\theta} = 92,85(46,28), IC(\theta, 95\%) = [68,05; 163,77].$$

- A aproximação (pelo método Delta) da distribuição de $\hat{\theta}$ para normal se mostrou inapropriada.
- Contudo, se nosso interesse for apenas na estimativa pontual da quantidade de fósforo que retorna a produtividade máxima, podemos considerar $\hat{\theta}$ sem problemas.