Crítica à Avaliação Intervalar na NBR 14.653-02

Autores:

- Luiz Fernando Palin Droubi
- Carlos Augusto Zilli
- Norberto Hochheim
- Willian Zonatto

Inferência Estatística

Inferência Estatística

É o conjunto de técnicas e procedimentos que permitem dar ao pesquisador um **grau de confiabilidade**, de confiança, nas afirmações que faz para a população, baseadas nos resultados das amostras.

Média simples

Dada uma variável aleatória X com média $\mu(X)$ e variância $\mathrm{Var}(X) = \sigma^2$ desconhecidas, estimar $\hat{\mu}(X)$.

É sabido que $\hat{\mu}(X) = \mathbb{E}(X) = \bar{x} = \frac{1}{n}\sum x_i$.

Qual a precisão de $\hat{\mu}(x)$?

Variância

Pode-se mostrar que a variância da média amostral $\sigma_{\bar{x}}^2$ é igual à variância populacional σ dividida pelo tamanho da amostra n, ou seja:

$$\mathrm{Var}(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

Se a variância populacional σ é desconhecida, esta deve ser estimada. Pode-se mostrar que um estimador não-viesado da variância populacional é:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\therefore \sigma_{\bar{x}}^2 = \frac{s^2}{n} \quad ou \quad \sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$\sigma_{\bar{x}}$ é conhecido como o erro-padrão do estimador $\bar{x}$.

Intervalo de Confiança

Erro-padrão é uma medida de erro de um estimador. O intervalo de confiança é construído em torno de algumas vezes o erro-padrão deste estimador.

$$IC_{\hat{\mu}} = \hat{\mu} \pm Z_{\alpha} s.e.(\hat{\mu})$$

Dependendo do nível de confiança desejado, a amplitude do IC pode ser maior ou menor.

Exemplos

$$IC_{95} \approx \hat{\mu} \pm 1,96 s.e.(\hat{\mu})$$

$$IC_{90} \approx \hat{\mu} \pm 1,64 s.e.(\hat{\mu})$$

$$IC_{80} \approx \hat{\mu} \pm 1,28 s.e.(\hat{\mu})$$

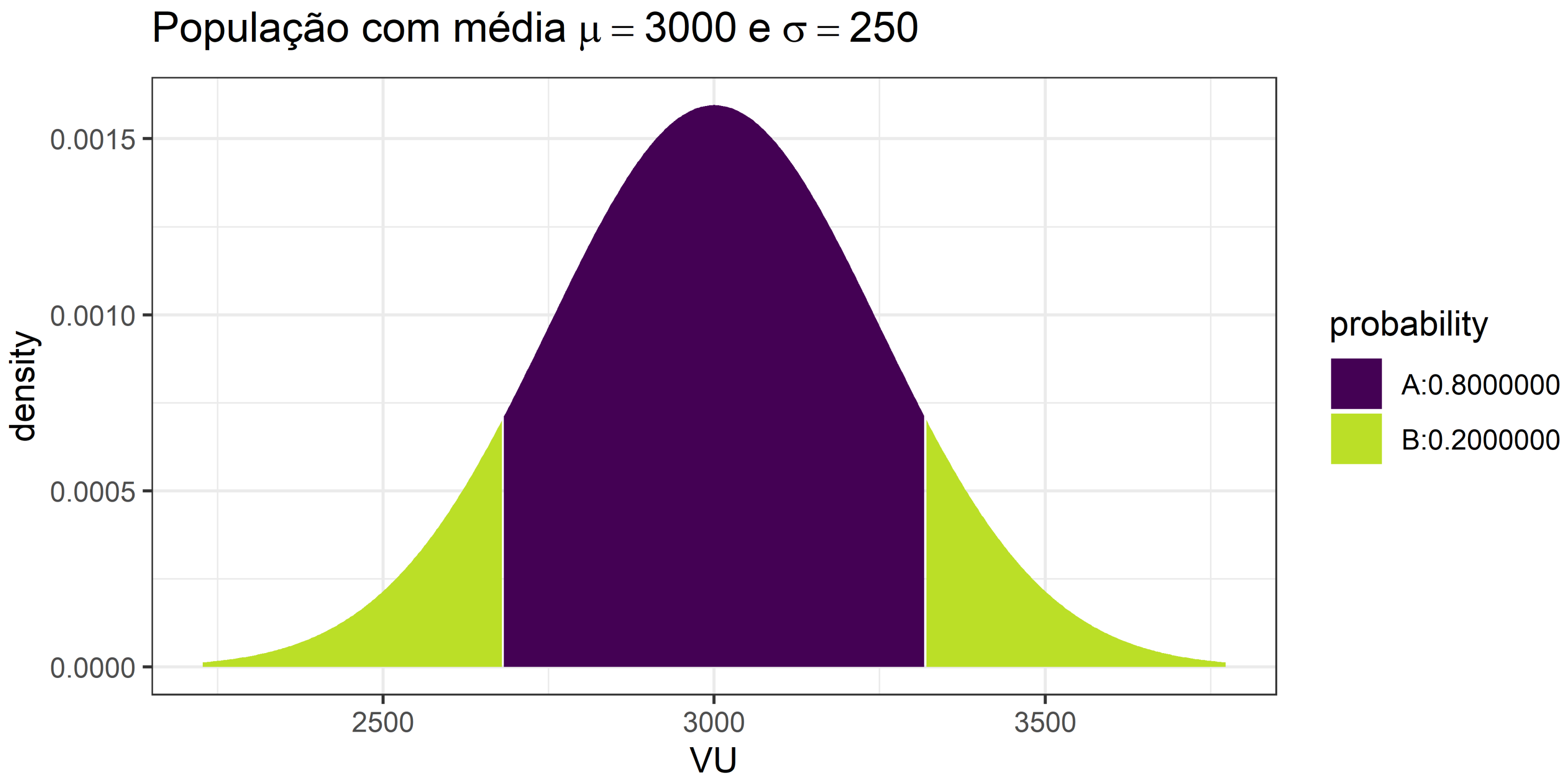
Atenção

Se o número de dados da amostra é pequeno, utiliza-se a distribuição t de Student.

Alguns estatísticos aconselham:

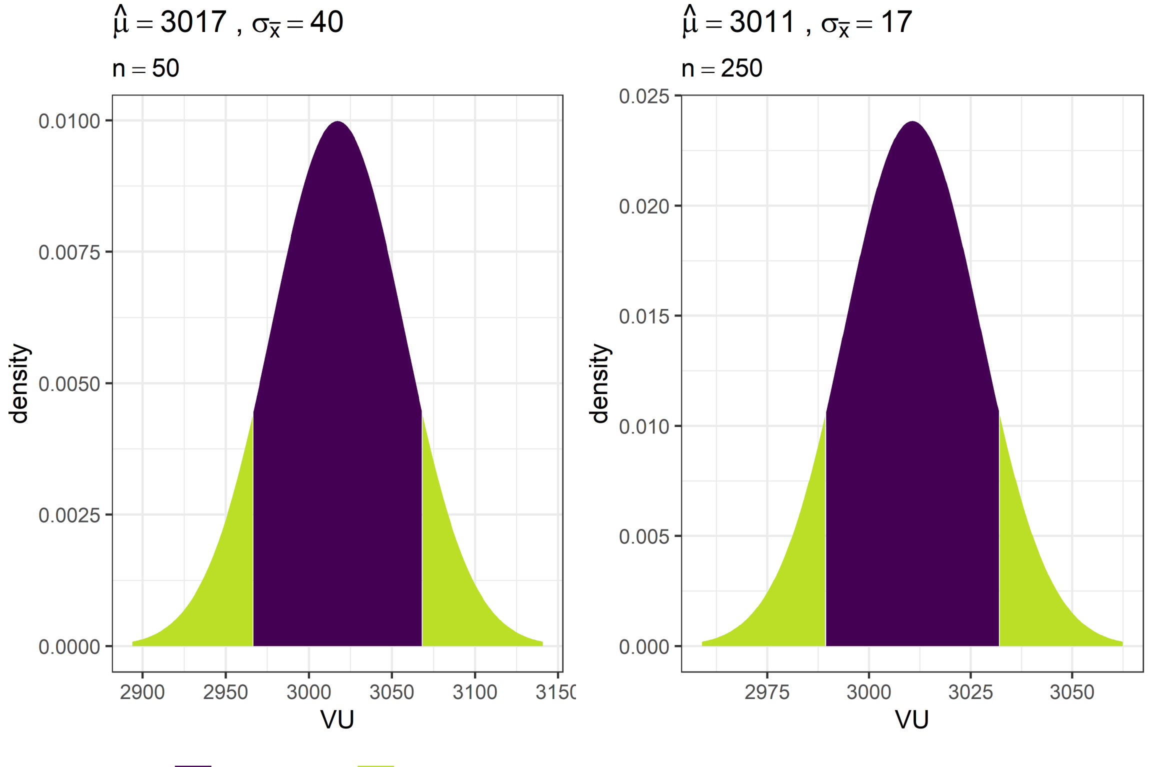
- Se $n \leq 30$, utiliza-se a distribuição t de Student
- Se $n > 30$, pode-se assumir a normalidade.

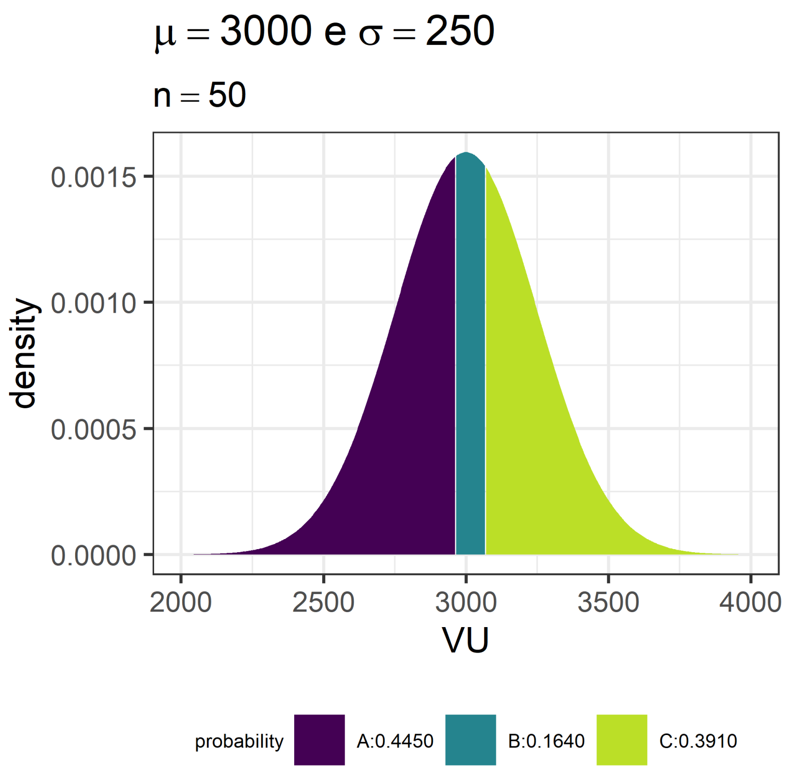
Exemplo

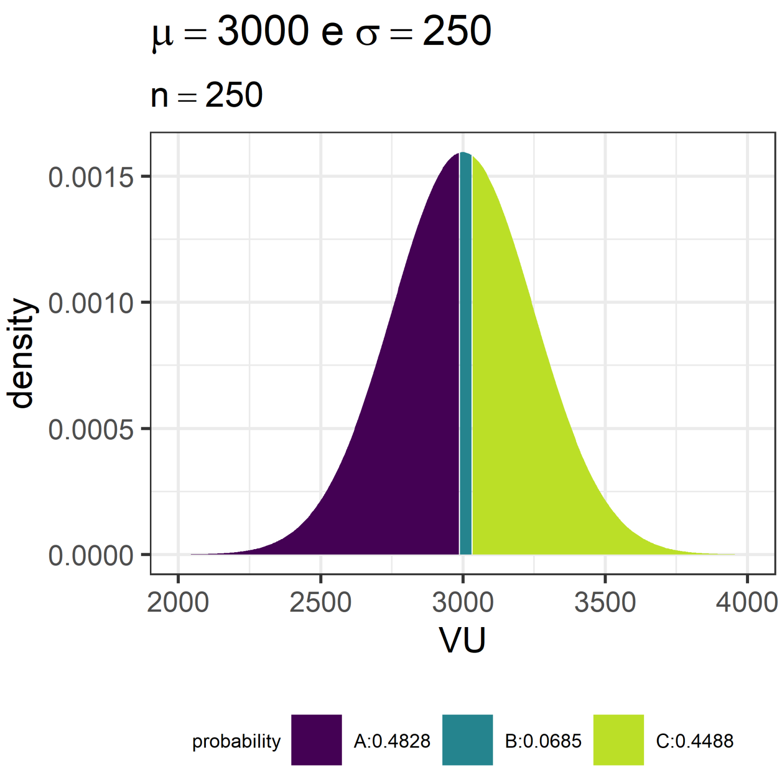


População com média $\mu = 3000$ e $\sigma = 250$
0.0015
0.0010
0.0005
0.0000
density
2500
3000
3500
VU
probability
A:0.8000000
B:0.2000000

Duas amostragens



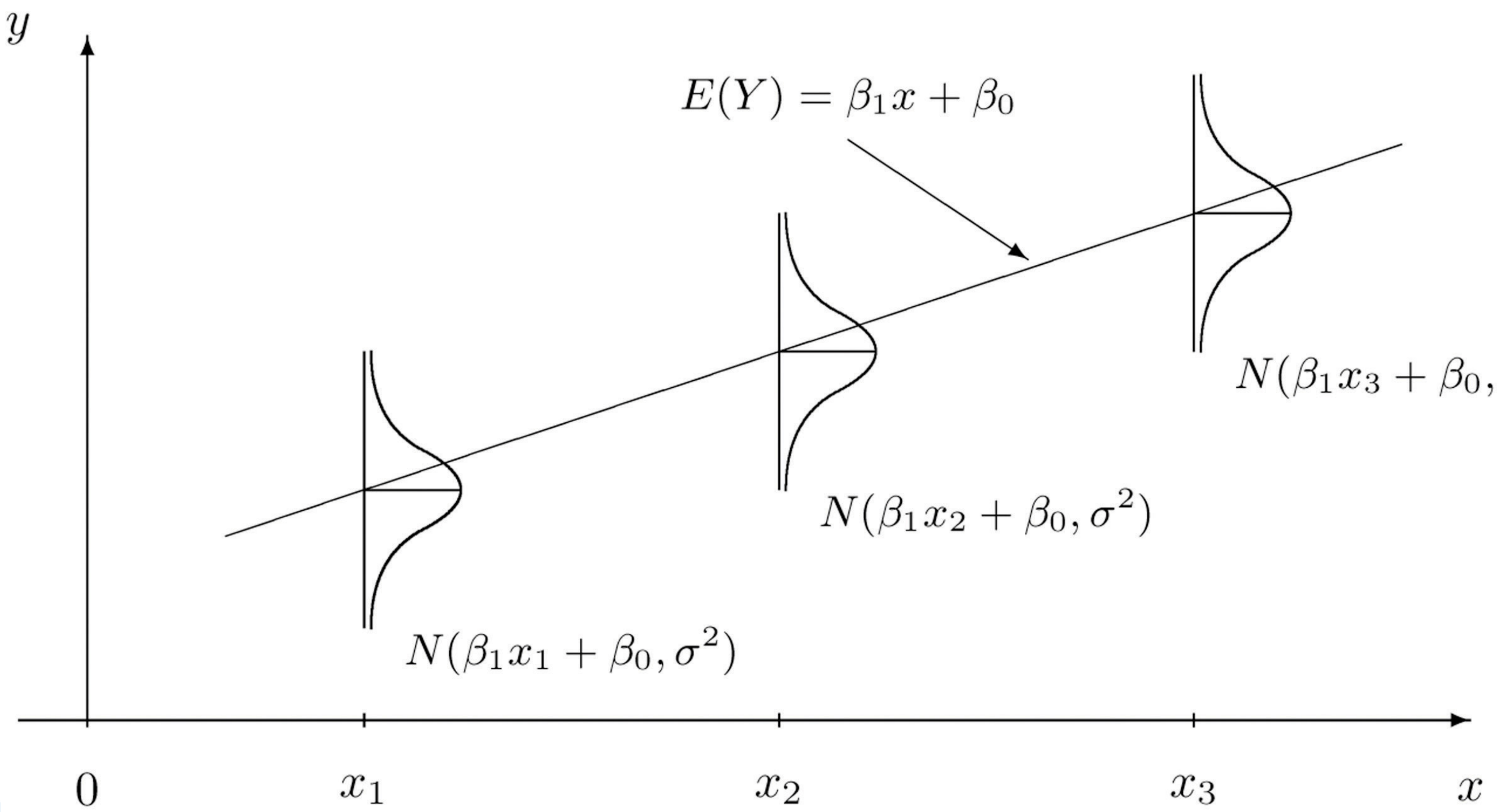

$\mu = 3000$ e $\sigma = 250$
$n = 50$
density
0.0015
0.0010
0.0005
0.0000
2000
2500
3000
3500
4000
VU
probability
A:0.4450
B:0.1640
C:0.3910


$\mu = 3000$ e $\sigma = 250$
$n = 250$
density
0.0015
0.0010
0.0005
0.0000
2000
2500
3000
3500
4000
VU
probability
A:0.4828
B:0.0685
C:0.4488

Resumindo

População	n	$\hat{\mu}$	$\sigma_{\bar{x}}$	**IC**	**IP**	CA_{inf}	p_{CAinf}
P1	50	3.017,28	39,94	[2.965,41, 3.069,16]	[2.646,85, 3.387,72]	2.564,69	5.94\%
P1	250	3.010,69	16,73	[2.989,19, 3.032,19]	[2.670,08, 3.351,30]	2.559,08	4.48\%

Regressão Linear


y
$E(Y) = \beta_1 x + \beta_0$
$N(\beta_1 x_3 + \beta_0, \sigma^2)$
$N(\beta_1 x_2 + \beta_0, \sigma^2)$
$N(\beta_1 x_1 + \beta_0, \sigma^2)$
0
x_1
x_2
x_3
x

Média Condicional

A equação de regressão linear nada mais é do que a estimativa da função média condicional de $Y|X$.

$$\mathbb{E}(Y|X) = X\beta + \epsilon$$

Como $\mathbb{E}(\epsilon) = 0$

$$\hat{\mu}(Y|X) = X\beta$$
$$\epsilon = Y - \hat{\mu}(Y|X)$$
$$Y = X\beta + \epsilon$$

Inferência clássica na regressão linear

Satisfeitas as hipóteses da inferências estatística clássica, *i.e.*, homoscedasticidade, normalidade, etc, temos:

Estimativa Central

$$\hat{y}_0 = \hat{\mu}(x_0) = x_0^T \hat{\beta}$$

Erro-padrão

$$se(\hat{y}_0) = se(x_0^T \hat{\beta}) = \sqrt{\mathrm{Var}(x_0^T \hat{\beta})} = \sqrt{\hat{\sigma}^2 x_0^T (X^T X)^{-1} x_0}$$

Intervalo de Confiança

$$IC = \hat{y}_0 \pm t_{0{,}90, n-k-1} \hat{\sigma} \sqrt{x_0^T (X^T X)^{-1} x_0}$$

Intervalo de Predição

Na predição de valores, estamos interessado na variância da população como um todo e não apenas na variância da média condicional, ou seja, queremos saber:

$$\text{Var}(y) = \text{Var}(x_0\hat{\beta} + \epsilon)$$

Portanto, para o cálculo do intervalo de predição, adiciona-se à incerteza do estimador, a variância da população:

$$IP = \hat{y}_0 \pm t_{0,90,n-k-1}\sqrt{\text{Var}(x_0^T\hat{\beta} + \epsilon)}$$

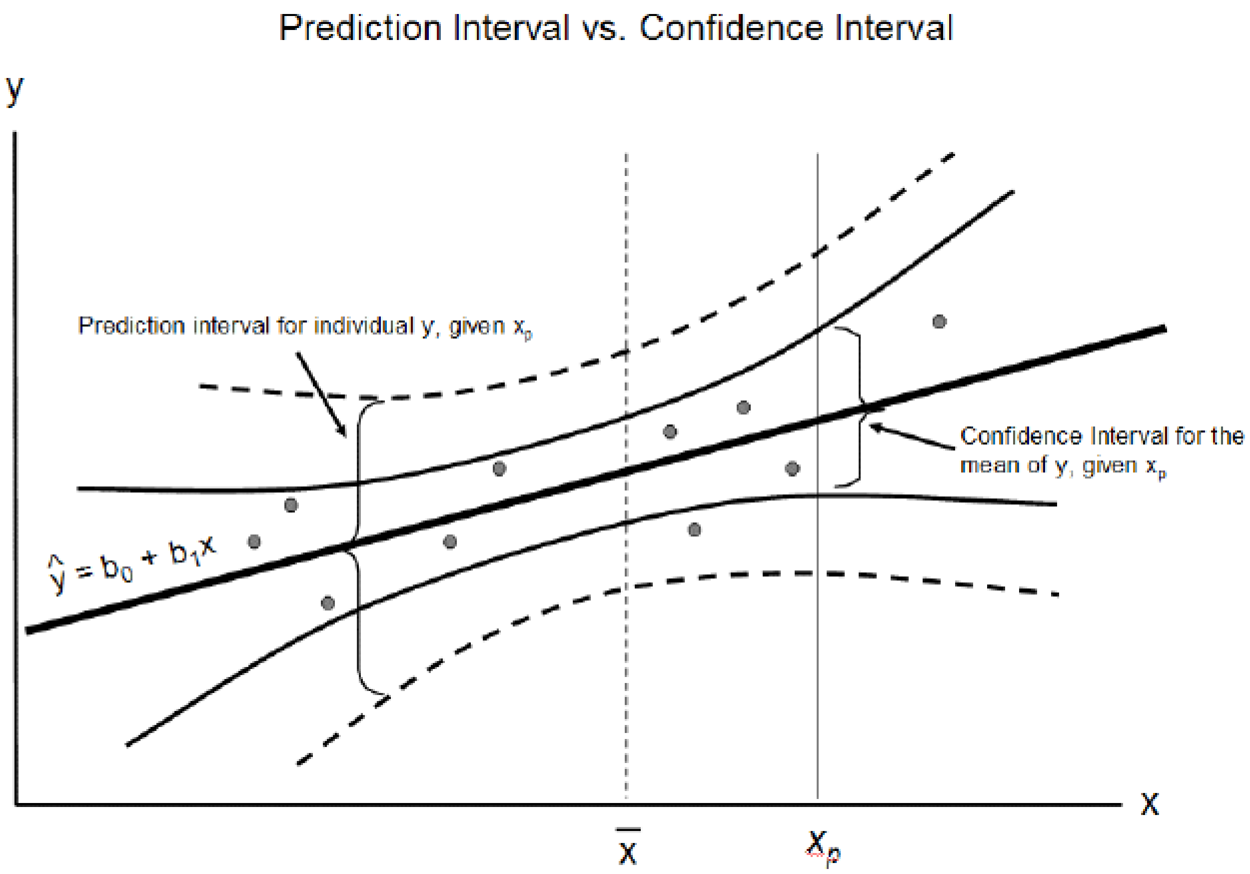
Se os erros forem independentes, então pode-se concluir que $\text{Var}(x_o^T\hat{\beta} + \epsilon) = \text{Var}(x_o^T\hat{\beta}) + \text{Var}(\epsilon)$.

Então:

$$IP = \hat{y}_0 \pm t_{0,90,n-k-1}\sqrt{\text{Var}(x_0^T\hat{\beta}) + \text{Var}(\epsilon)}$$

$$IP = \hat{y}_0 \pm t_{0,90,n-k-1}\sqrt{\hat{\sigma}^2 x_0^T(X^TX)^{-1}x_0 + \hat{\sigma}^2}$$

$$IP = \hat{y}_0 \pm t_{0,90,n-k-1}\hat{\sigma}\sqrt{x_0^T(X^TX)^{-1}x_0 + 1}$$


Prediction Interval vs. Confidence Interval
y
Prediction interval for individual y, given x_p
Confidence Interval for the mean of y, given x_p
$\hat{y} = b_0 + b_1 x$
x
$\bar{x}$
x_p

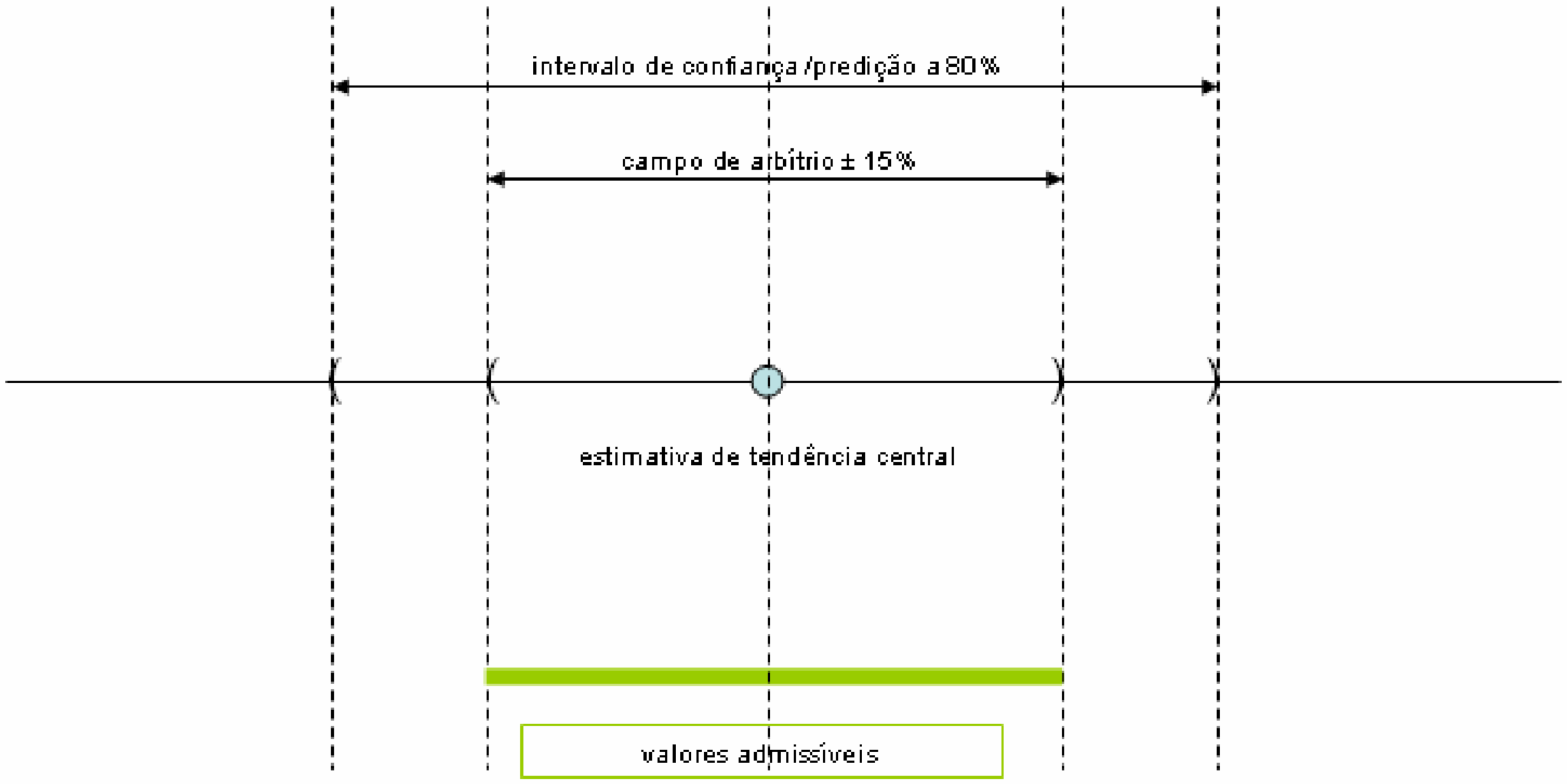
NBR 14.653-02

Estimativa de tendência central

Quando for adotada a estimativa de tendência central para a avaliação, o intervalo de valores admissíveis deve estar limitado **simultaneamente**:

A) ao intervalo de predição ou ao intervalo de confiança de 80 % para a estimativa de tendência central;

B) ao campo de arbítrio.

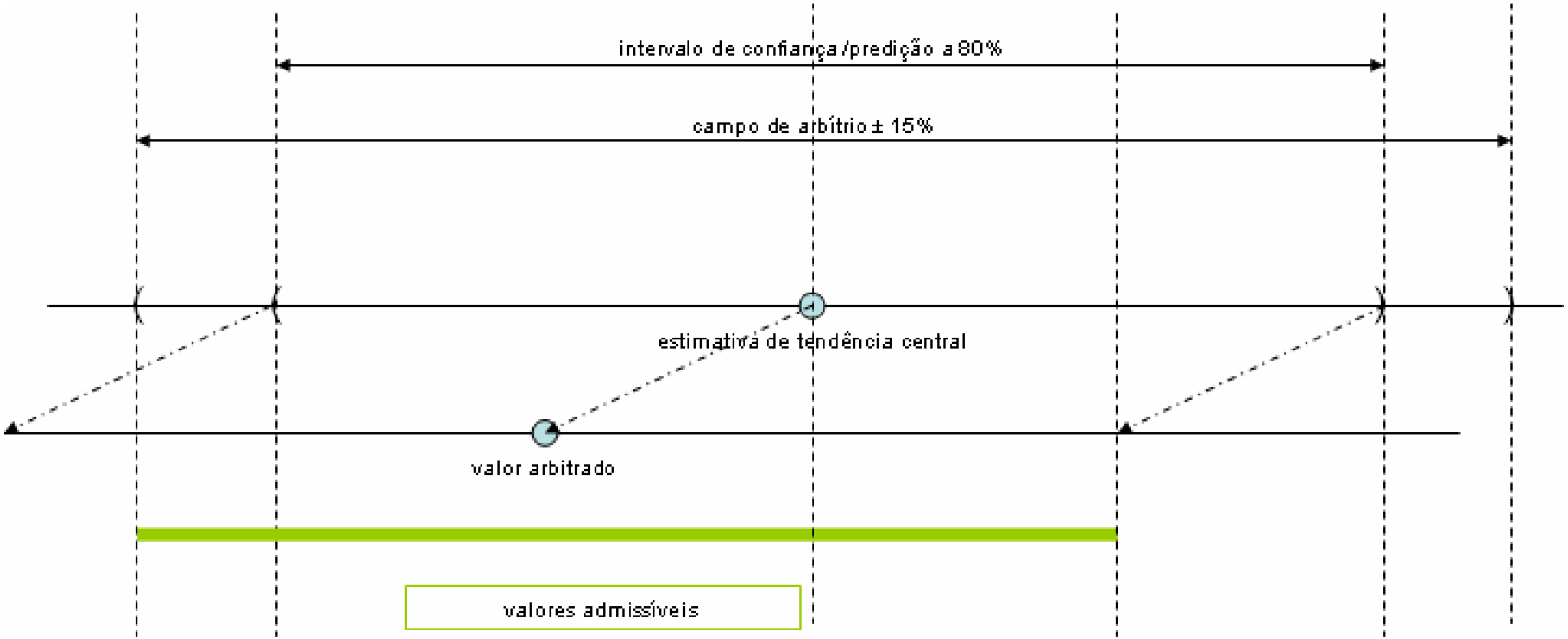

intervalo de confiança /predição a 80%
campo de arbítrio ± 15%
estimativa de tendência central
valores admissíveis

Valor arbitrado

Quando for adotado o valor arbitrado, o intervalo de valores admissíveis deve estar limitado **simultaneamente**:

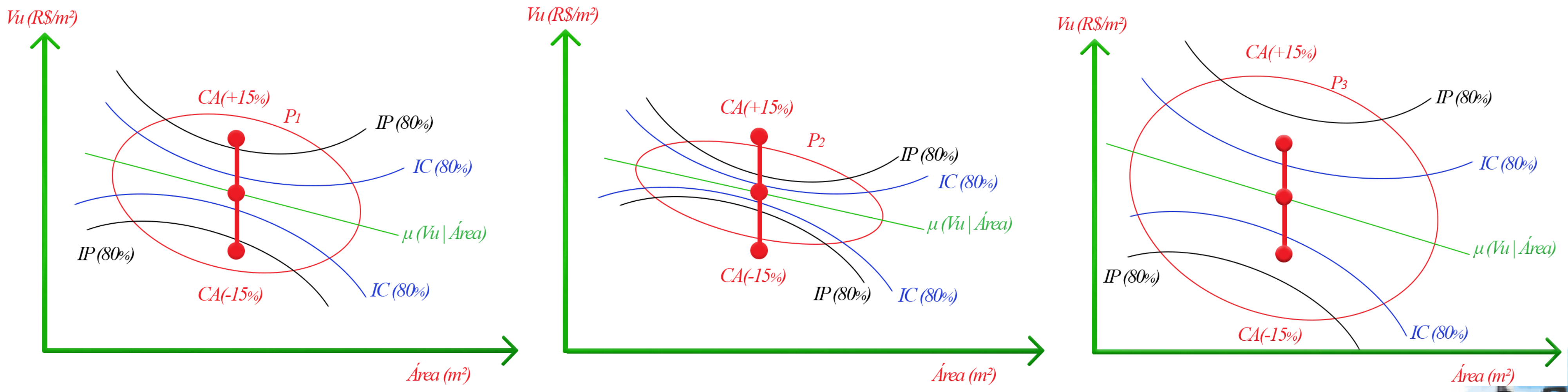
A) ao intervalo em torno do valor arbitrado com amplitude igual à do intervalo de predição ou ao intervalo de confiança de 80% para a estimativa de tendência central;

B) ao campo de arbítrio em torno da estimativa de tendência central.


intervalo de confiança /predição a 80%
campo de arbítrio ± 15%
estimativa de tendência central
valor arbitrado
valores admissíveis

NBR 14.653-02

Aplicação em diferentes cenários



Estudos de casos

Dados de mercado

Coeficientes e estatísticas do modelo adotado.

	Estimate	CI (lower)	CI (upper)	Std. Error	t value	Pr(>|t|)	
(Intercept)	13.564	13.264	13.864	0.230	58.847	<0.001	***
area_total	0.001	0.001	0.002	0.000	5.113	<0.001	***
quartos	0.164	0.118	0.210	0.035	4.626	<0.001	***
suites	0.061	0.017	0.105	0.034	1.810	0.078	.
garagens	0.209	0.165	0.252	0.033	6.247	<0.001	***
log(dist_b_mar)	-0.141	-0.176	-0.105	0.027	-5.174	<0.001	***
rec(padrao)	-0.563	-0.700	-0.426	0.105	-5.360	<0.001	***

Estimativa	Intervalo de valores admissíveis	Campo de Arbítrio
R$961.660,64	[R$924.768,13; R$1.000.024,94]	[R$817.411,55; R$1.105.909,74]
R$1.100.000,00	[R$1.060.973,75; R$1.105.909,74	[R$817.411,55; R$1.105.909,74]

Probabilidades associadas

Média

Valor	t	p
Inferior	3,27	0.108\%
Arbitrado	4,48	0.00298\%
Superior	4,65	0.0017\%

População

Valor	Valor	t	p
Inferior (CA)	817.411,55	-1,17	12.5\%
Arbitrado	1.100.000,00	0,96	83\%
Superior (CA)	1.105.909,74	1,00	83.9\%

Regressão Quantílica

Quantílica (72,5%)	NBR
1.096.396,16	1.100.000,00
1.022.899,48	1.060.973,75
1.175.173,68	1.105.909,74

Dados com distribuição Normal

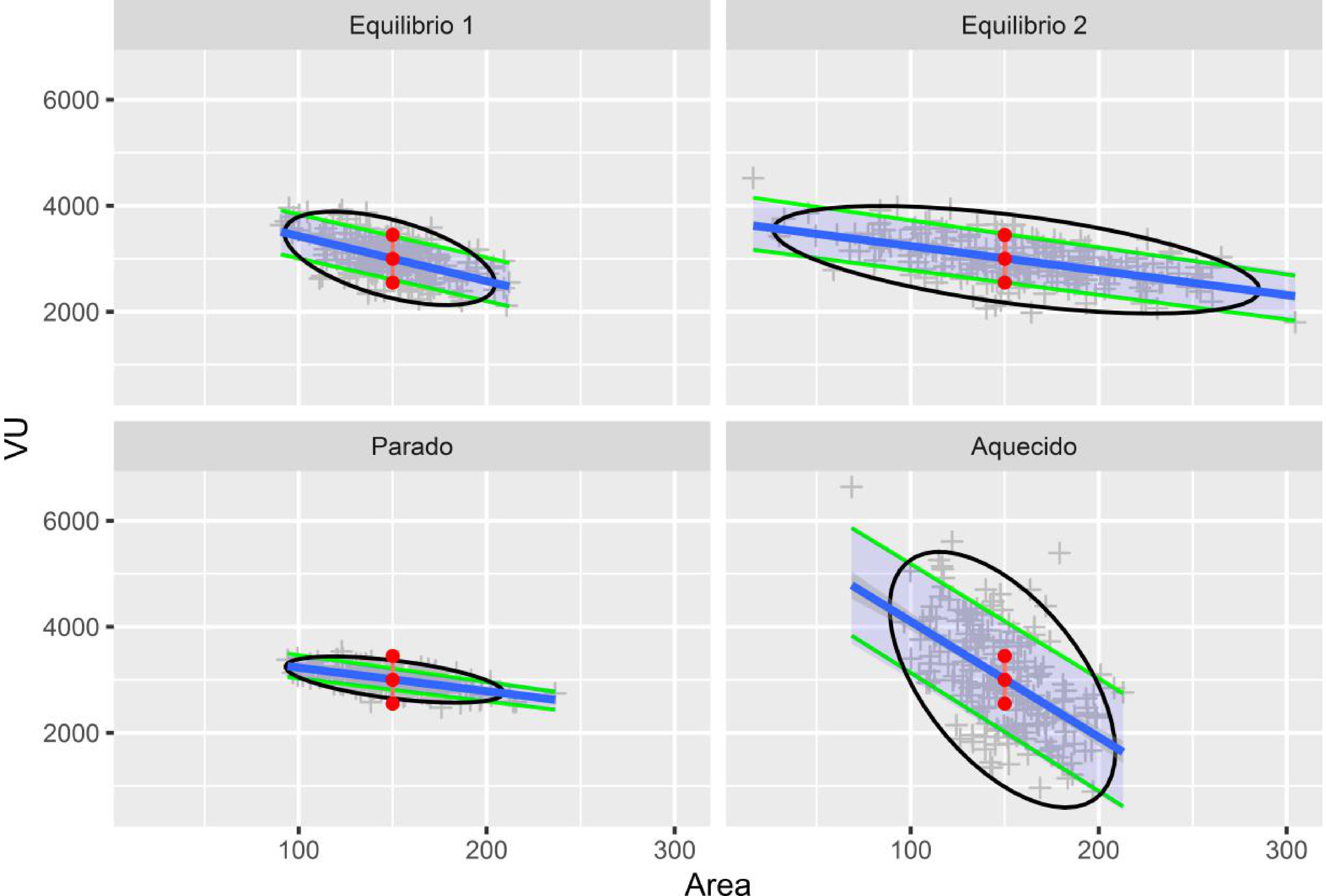
$\hat{\mu}(x_0)$ = R$1.000.000,00. Hipóteses sobre a inferência: $\mathcal{N}$, t_{40}, t_{10}

Amplitude IP	$\sqrt{\sigma^2 + s.e.^2\,(\hat{y}_o)}$	**Semi-amplitude do CA**	Z_{inf}	Z_{sup}	p_{inf}	p_{sup}
30%	117.045,62	150.000	-1,28	1,28	10\%	90\%
40%	156.060,83	150.000	-0,96	0,96	16.8\%	83.2\%
50%	195.076,04	150.000	-0,77	0,77	22.1\%	77.9\%

Amplitude IP	$\sqrt{\sigma^2 + s.e.^2\,(\hat{y}_o)}$	**Semi-amplitude do CA**	t_{inf}	t_{sup}	p_{inf}	p_{sup}
30%	115.112,15	150.000	-1,30	1,30	10\%	90\%
40%	153.482,87	150.000	-0,98	0,98	16.7\%	83.3\%
50%	191.853,58	150.000	-0,78	0,78	21.9\%	78.1\%

Amplitude IP	$\sqrt{\sigma^2 + s.e.^2\,(\hat{y}_o)}$	**Semi-amplitude do CA**	t_{inf}	t_{sup}	p_{inf}	p_{sup}
30%	109.314,81	150.000	-1,37	1,37	10\%	90\%
40%	145.753,09	150.000	-1,03	1,03	16.4\%	83.6\%
50%	182.191,36	150.000	-0,82	0,82	21.5\%	78.5\%

Dados criados randomicamente


Equilibrio 1
Equilibrio 2
Parado
Aquecido
6000
4000
2000
VU
100
200
300
Area

Dados lognormais

	Moda	Mediana	Media
estimativa_central	989.980	999.989	1.005.032
ca_inf	841.483	849.991	854.277
ca_sup	1.138.477	1.149.988	1.155.787
ip_inf	870.567	879.369	883.804
ip_sup	1.125.772	1.137.154	1.142.889

	Moda	Mediana	Media
estimativa_central	46,0	50,0	52,0
ca_inf	4,3	5,3	5,8
ca_sup	90,2	91,8	92,6
ip_inf	8,4	10,0	10,9
ip_sup	88,1	90,0	90,9

Obrigado!