

Pequenas amostras

Encolhimento



11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



IBAPE NACIONAL
Instituto Brasileiro de Avaliações
e Perícias de Engenharia

Introdução

- É fato amplamente conhecido que o estimador dos MQO é BLUE: *Best Linear Unbiased Estimator*
- Não seria surpreendente afirmar que existem estimadores melhores que o MQO
 - P. ex.: um estimador não-linear (regressão à mediana, regressão não-linear, mínimos quadrados generalizados, etc.)
- Porém, pode parecer estranho dizer que existem estimadores melhores, porque viesados.
- Pode-se pensar no *trade-off* entre viés e variância: um pouco de viés pode ser útil!

Exemplo Motivador

- Imagine que se queira prever o número médio de rebatidas de um grupo de 18 jogadores de beisebol ao final da temporada, baseado nas médias das suas primeiras 45 rebatidas
- O estimador mais natural para a média de rebatidas ao longo da temporada, para um determinado jogador, é a sua própria média de rebatidas.
 - P. Ex.: Se Clemente obteve sucesso em 18 das 45 primeiras rebatidas, o natural seria estimar que, ao final do ano, Clemente teria um aproveitamento de 0,400 ou 40,0%.
- Em termos estatísticos, este seria o melhor estimador não-viesado para o aproveitamento médio de rebatidas de cada jogador

Exemplo Motivador

- Intuição:
 - Um rebatedor que acerte sempre no terceiro arremesso, teria um aproveitamento de 0,333.
 - Um rebatedor que acerte sempre no segundo arremesso, teria um aproveitamento de 0,500.
 - Babe Ruth tinha um aproveitamento de 0,342, em 2.873 rebatidas.
 - Porém, Clemente apresentou um aproveitamento de 0,400 nas primeiras 45 rebatidas da temporada
 - Outros dois jogadores também apresentaram aproveitamento mais altos do que Babe Ruth nas primeiras 45 rebatidas

Exemplo Motivador

- Caso o problema fosse o de estimar o aproveitamento nas rebatidas de um único jogador, o estimador “ordinário” (não-viesado) seria o mais adequado;
- No entanto, se a nossa intenção for estimar o sucesso nas rebatidas de um grupo de 3 ou mais jogadores, pode-se demonstrar que o melhor estimador, *i.e.* aquele que resulta num menor erro quadrático total (para o conjunto de 18 jogadores), não é este estimador “ordinário”, mas o estimador de Efron-Morris, que **encolhe** as médias em direção à “Grande Média” $\bar{y}$, ou média do grupo.

$$z = \bar{y} + c(y - \bar{y})$$

- Por exemplo, para Clemente:

$$z = 0,265 + 0,212(0,400 - 0,265) = 0,294$$

Estimadores do tipo James-Stein

- O estimador de James-Stein encolhe um conjunto de k médias (desde que $k \geq 3$) em direção à zero:

$$\theta_i^{JS} = \left(1 - \frac{c}{S^2}\right) X_i$$
$$0 < c < 2(k - 2)$$

- O estimador de Efron-Morris, encolhe um conjunto de k médias (desde que $k \geq 4$) em direção à “grande-média” $\bar{X}$.

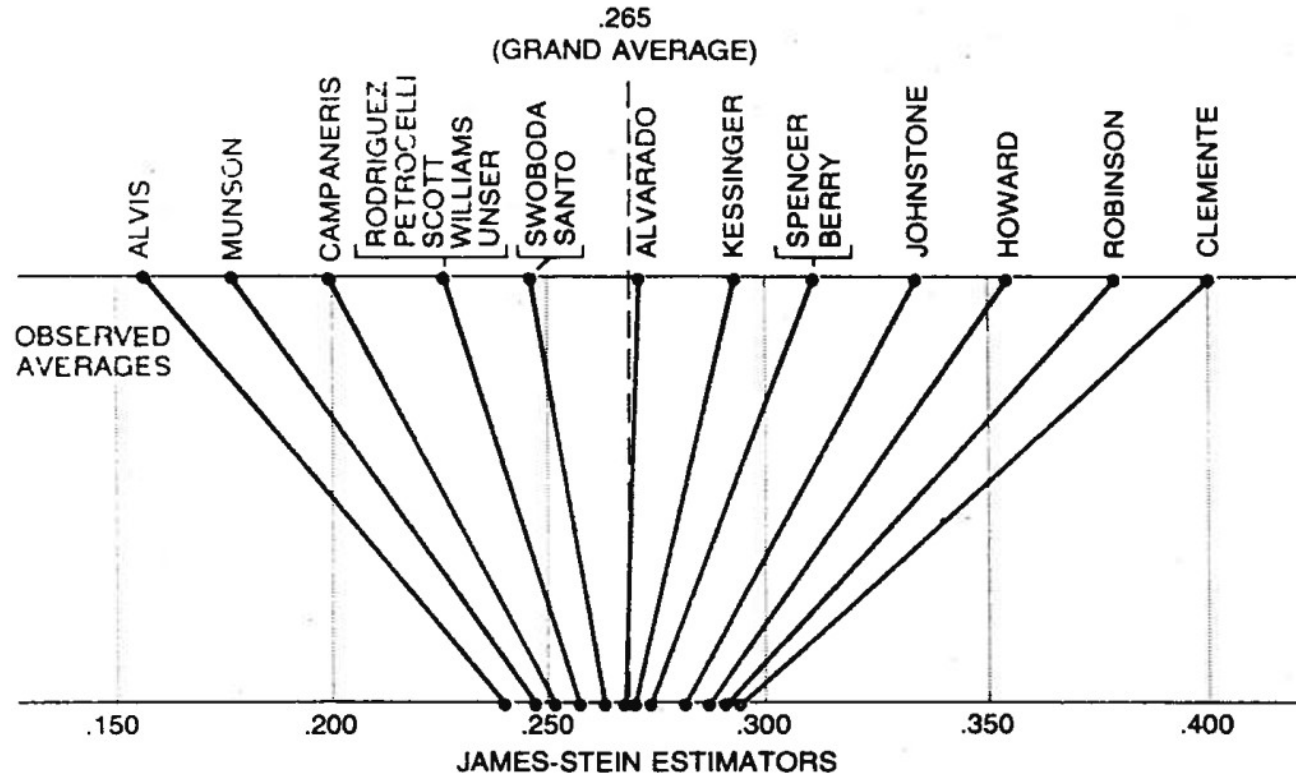
$$\theta_i^{EM} = \bar{X} + \left(1 - \frac{c}{S'^2}\right) (X_i - \bar{X})$$
$$0 < c < 2(k - 3)$$

11^A15

SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

Exemplo Motivador



Exemplo Motivador

- Para este caso em tela, as médias “verdadeiras” podem ser (e foram) observadas ao final da temporada.
- Do grupo de 18 jogadores:
 - 16 vezes o estimador de James-Stein teve erro médio quadrático menor do que o estimador “ordinário”
 - 2 vezes o estimador ordinário teve erro médio quadrático menor do que o estimador de James-Stein
- O erro médio quadrático total diminui de 0,077 para 0,022 (3,5x menor).

Exemplo Motivador

- A classe de estimadores do tipo James-Stein é uma classe de estimadores *viesados* do número de rebatidas;
- Viés não necessariamente é um defeito para um estimador;
- Viés também pode ser uma qualidade desejada para um estimador;
- P. Ex.: imaginem que um grande banco de dados armazene o número de rebatidas e sucessos de cada jogador da liga de beisebol ao longo de toda a história.
 - Babe Ruth teve média de sucessos de 0,342 (ou 34,2% de sucesso)
 - Imagine agora que um novato acaba de ter sucesso nas suas 2 primeiras rebatidas. O seu aproveitamento, portanto, é de 100%. O estimador viesado, no entanto, proporciona uma estimativa bem mais realista, apesar de viesada:

$$z = 0,265 + 0,212(1,000 - 0,265) = 0,421$$

Exemplo Motivador

- No entanto, ainda, este estimador pode ser melhorado.
- Pode-se pensar que o estimador deveria apresentar encolhimento maior para o jogador que rebateu apenas 2 vezes do que o encolhimento para os outros jogadores, que rebateram 45 vezes
- Em termos estatísticos, deve-se pensar num estimador que tenha um encolhimento proporcional à variância de cada média
- O próximo exemplo irá explorar essa propriedade

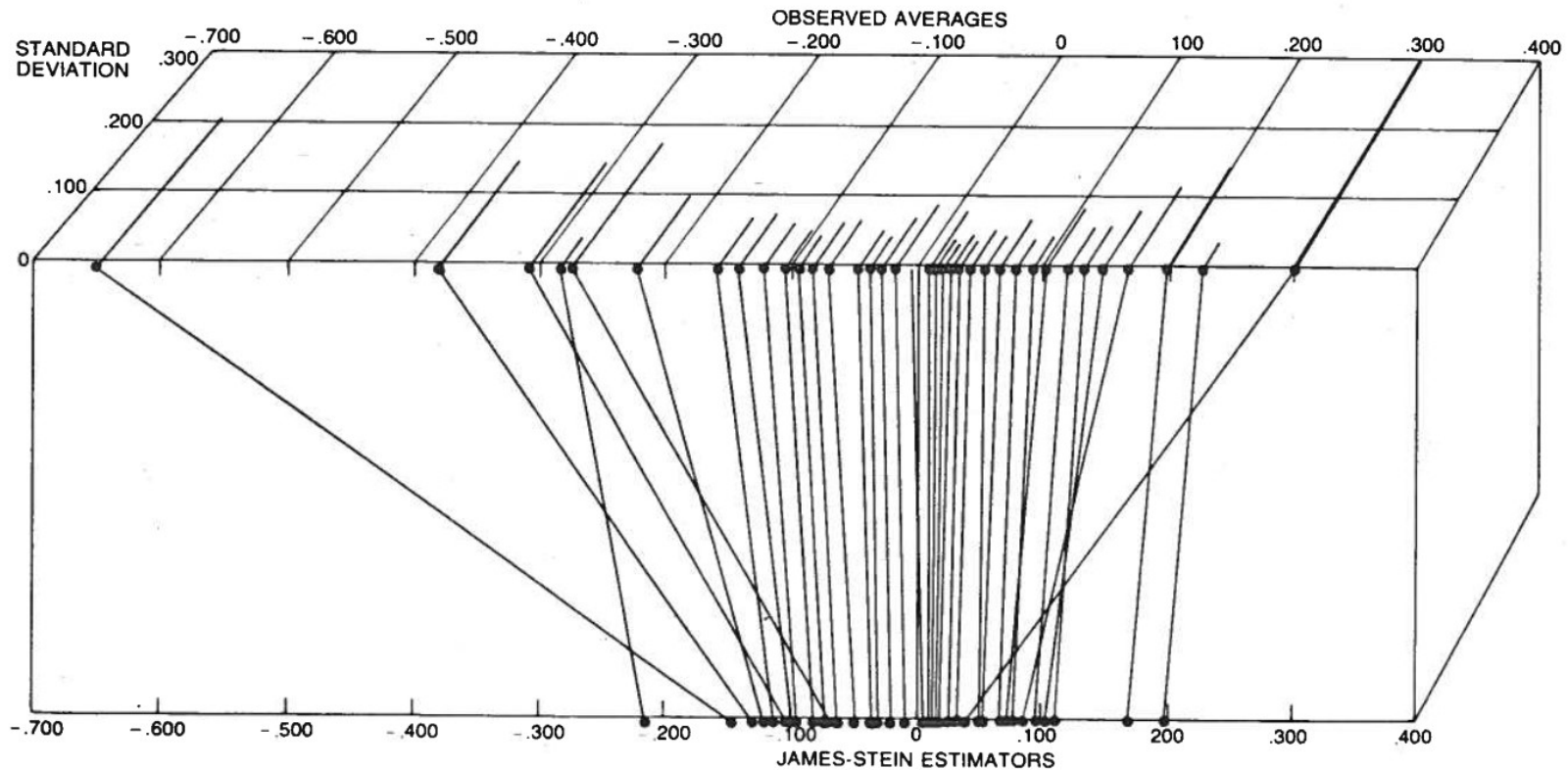
11^A15

SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



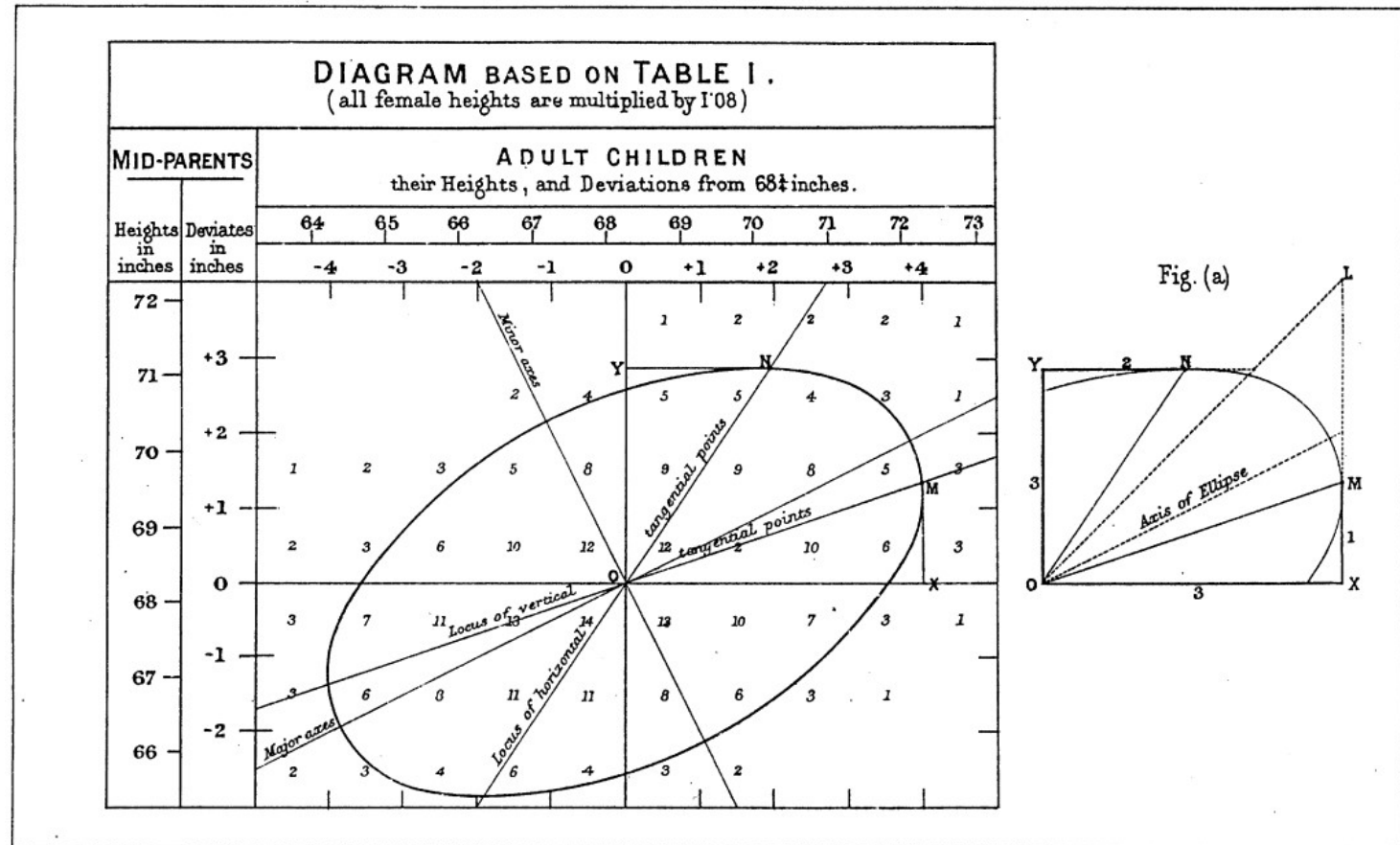
Exemplo Motivador 2



11^A15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

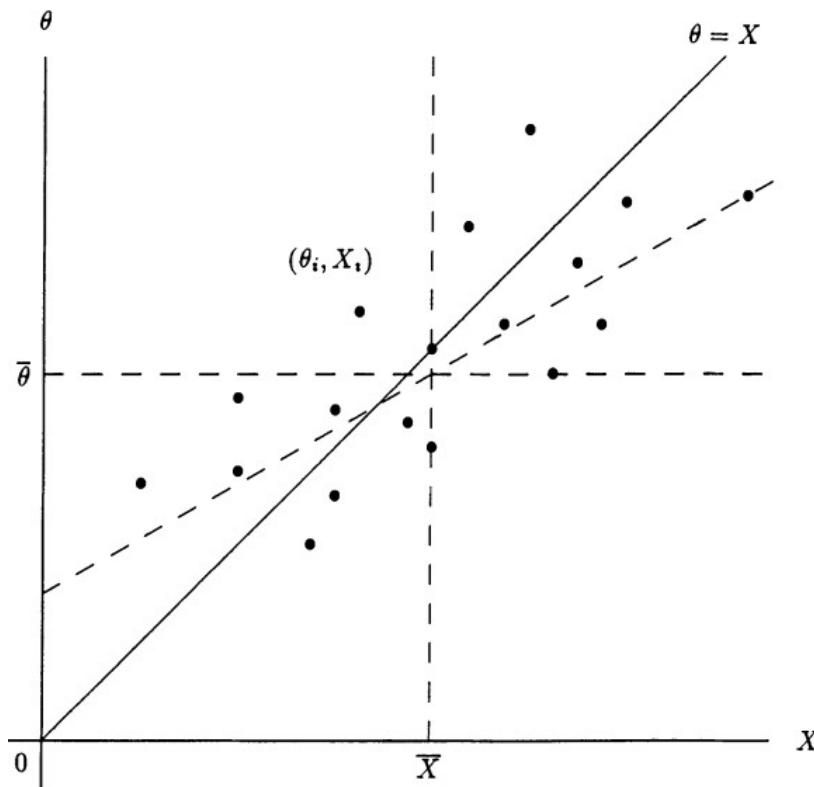
Explicação intuitiva: uma perspectiva Galtoniana



11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

Explicação intuitiva: uma perspectiva Galtoniana

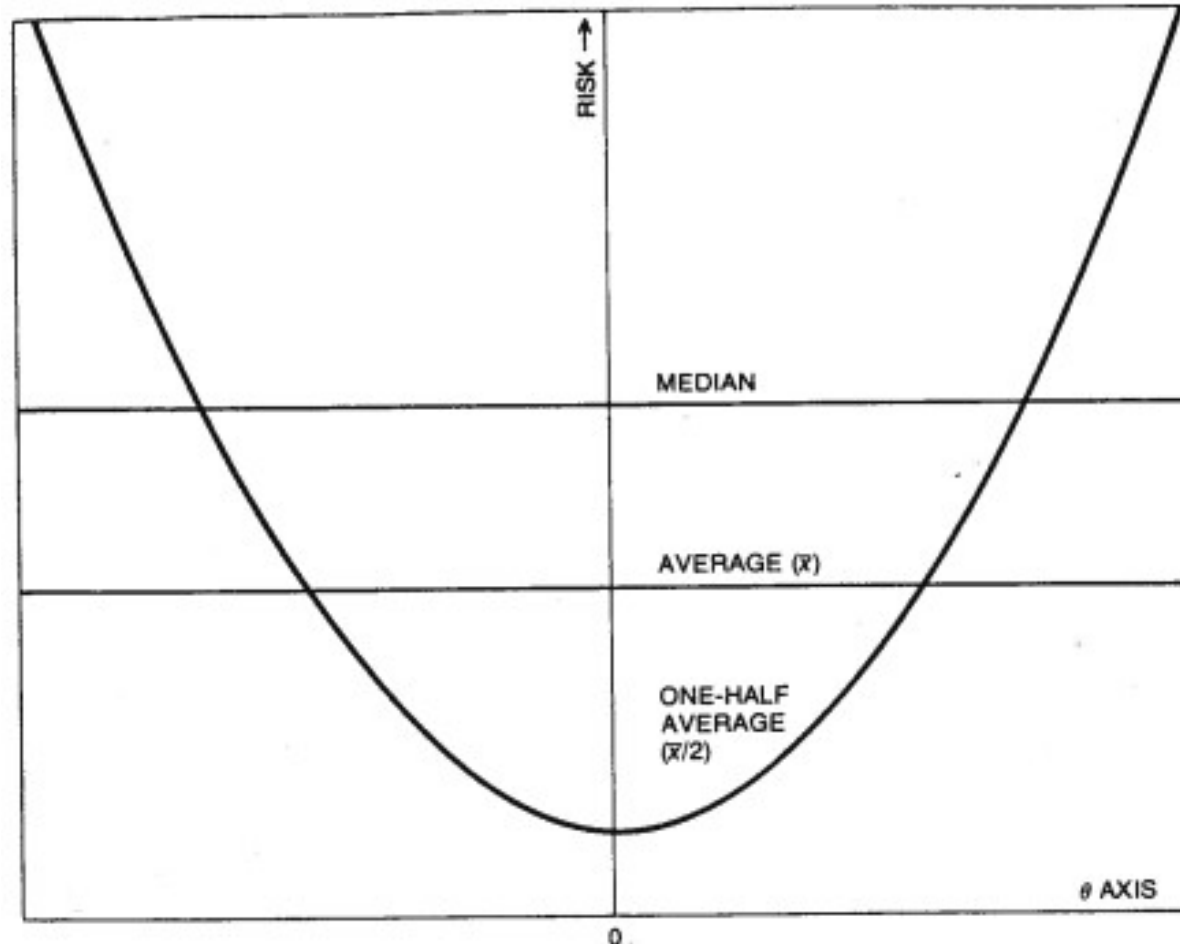


$$E(\theta|X) \neq E(X|\theta)$$

11^A15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

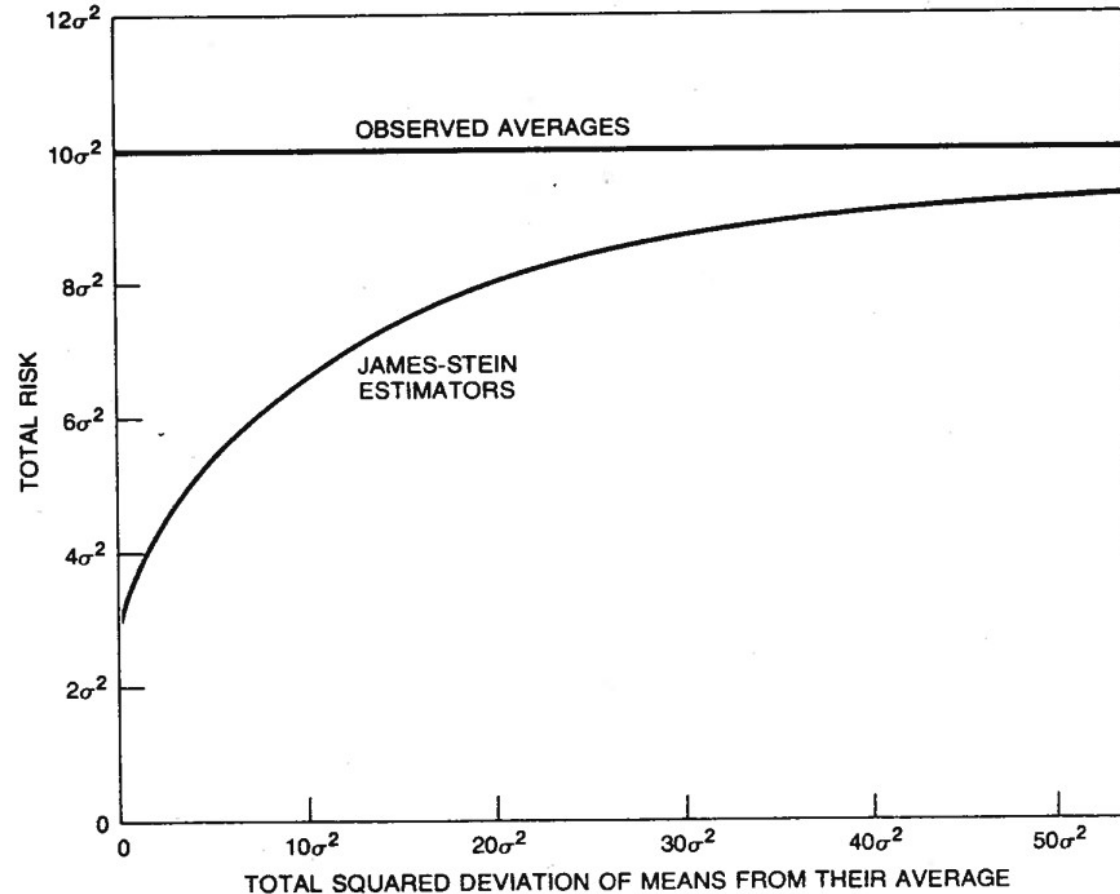
Admissibilidade de um estimador



11^A15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

Admissibilidade de um estimador



Aplicações na Engenharia de Avaliações

- Na Avaliação em massa de imóveis, os avaliadores precisam avaliar o valor de um lote paradigma em diversas localidades ao mesmo tempo. Neste caso, a estimação em conjunto, como vimos, reduz o erro médio quadrático total;
- Na imensa maioria das vezes não são encontrados muitos elementos amostrais em uma ou mais localidades;
- Nestas localidades a amostragem reduzida pode induzir a falsas conclusões (como no exemplo do novato do beisebol)
- Os estimadores do tipo James-Stein podem ser úteis aos avaliadores nestes casos.

Como os estimador JS podem ser úteis aos avaliadores

- Imaginem uma pesquisa de lotes em diversos bairros de uma cidade;
- Agora imaginem que, após a amostragem, os dados são homogeneizados quanto às suas características físicas, ou seja, seus VU são transformados em VU_{hom} para um lote paradigma em cada uma das localidades;
- Uma vez que esses valores homogeneizados são computados para cada elemento amostral, calculam-se $\overline{VU}_{hom}$ para cada localidade;
- Então calcula-se a grande média dos $\overline{VU}_{hom}$ e, por fim, faz-se o encolhimento destes em direção a esta grande média.
- [Planilha](#)

11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



ESTUDO DE CASO

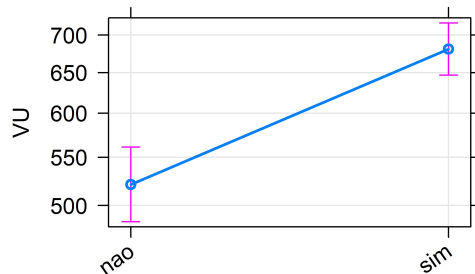
Confecção de Planta Genérica de Valores em Campo Grande/MS

11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

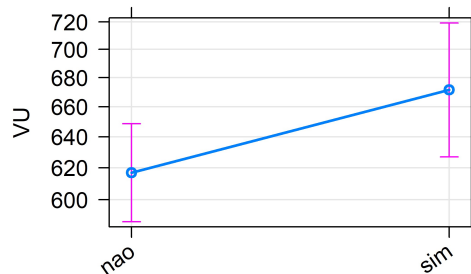
REALIZAÇÃO



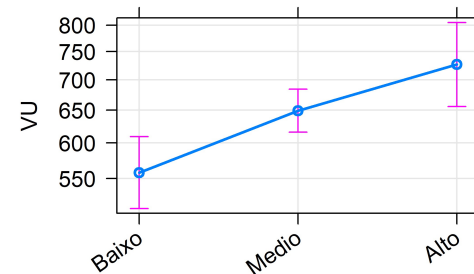
Planta Genérica de Valores



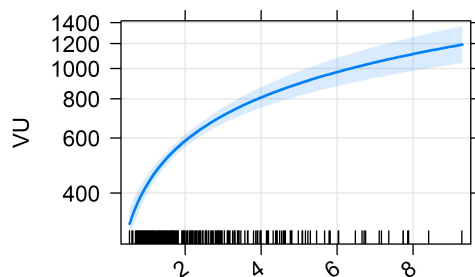
PAV



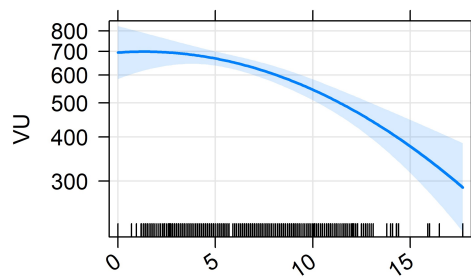
ESQUINA



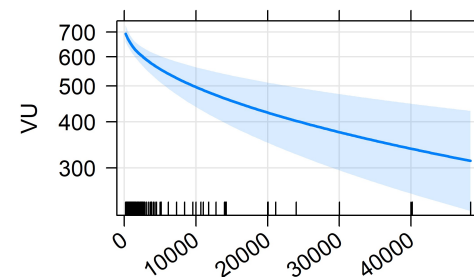
PADRAO



REDA



DISTCENTRO

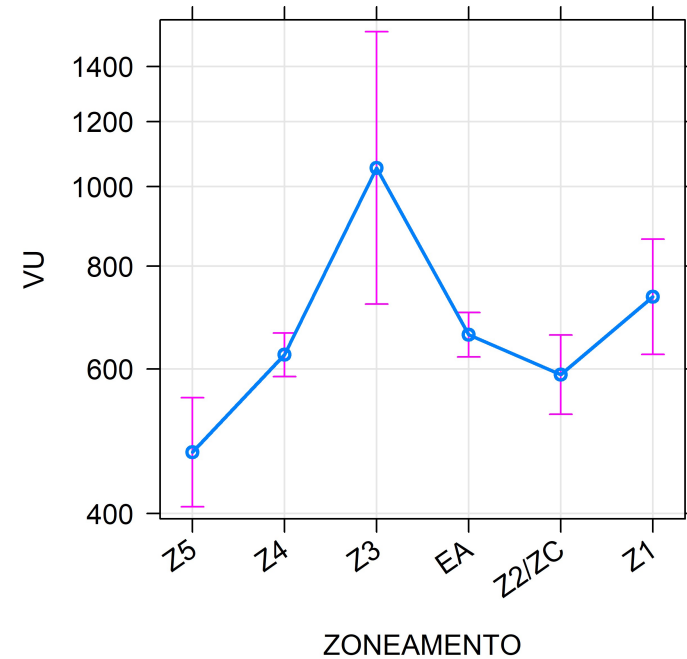
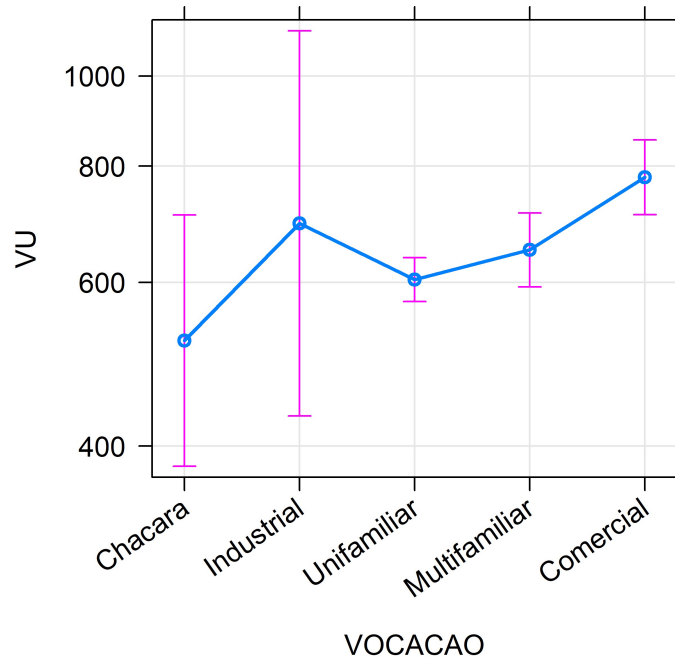


AREA

11^A15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

Planta Genérica de Valores

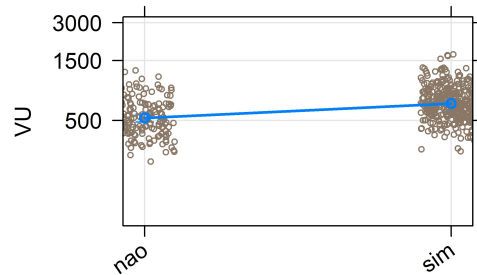


11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

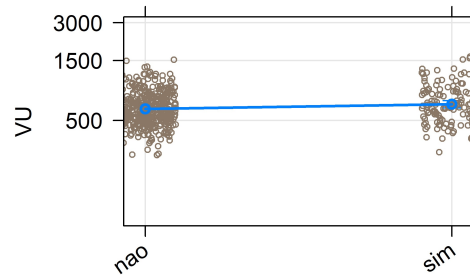
REALIZAÇÃO



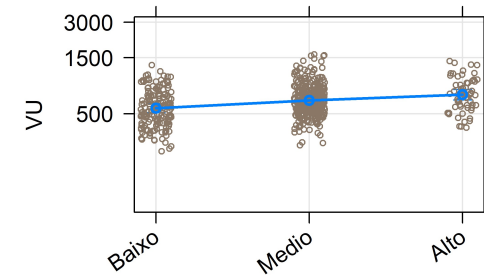
Planta Genérica de Valores



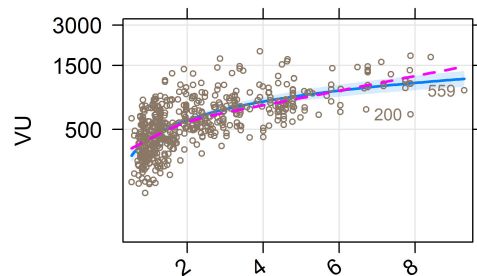
PAV



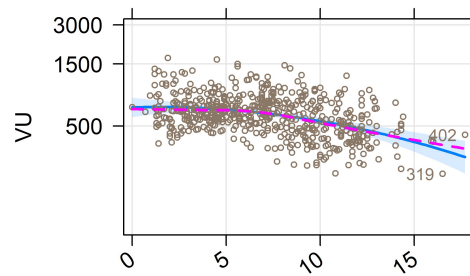
ESQUINA



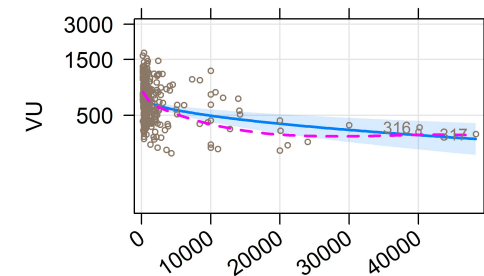
PADRAO



RENDA



DISTCENTRO

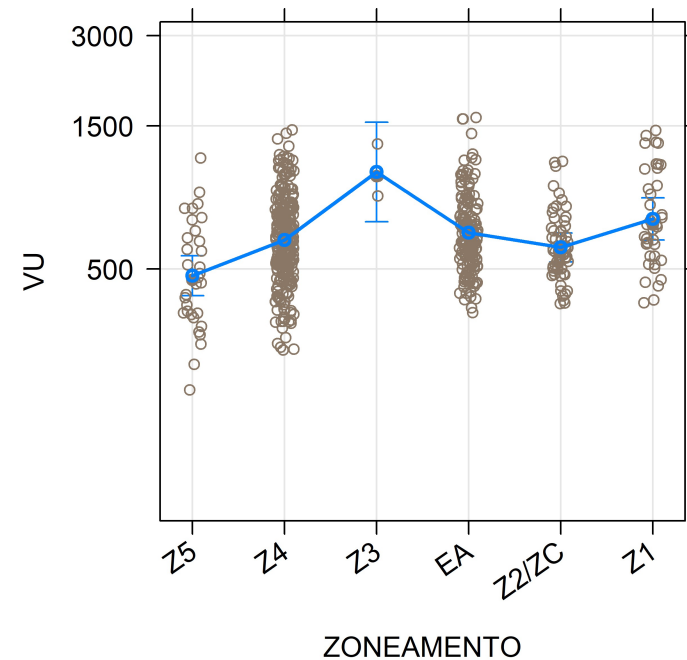
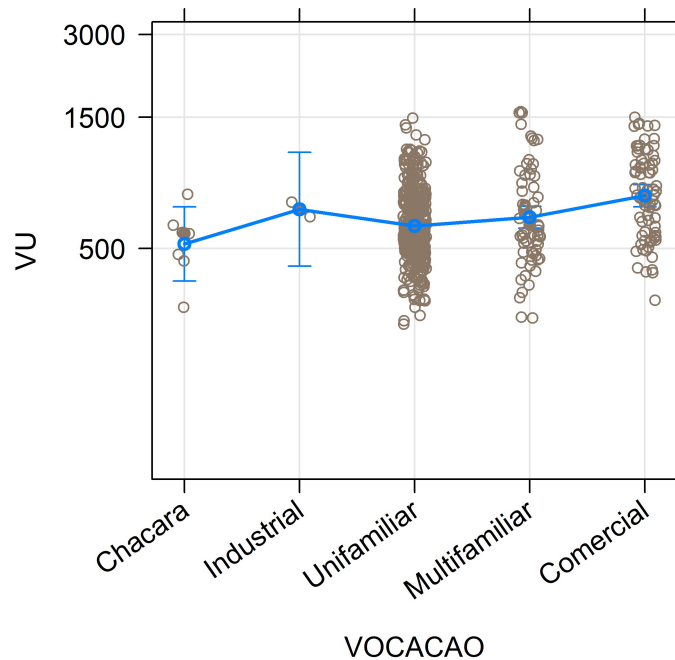


AREA

11^A15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

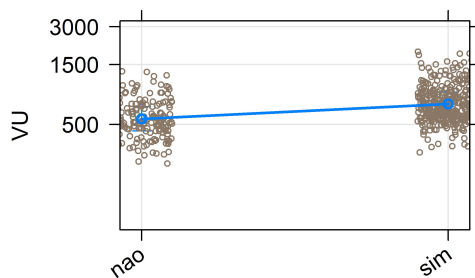
Planta Genérica de Valores



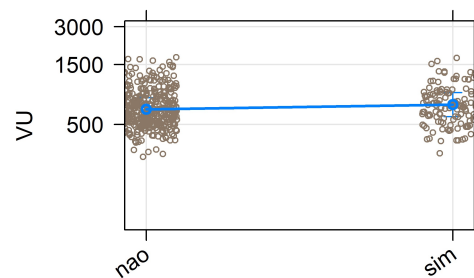
11^A15
SETEMBRO | 2023
SÃO PAULO - SP

Planta Genérica de Valores

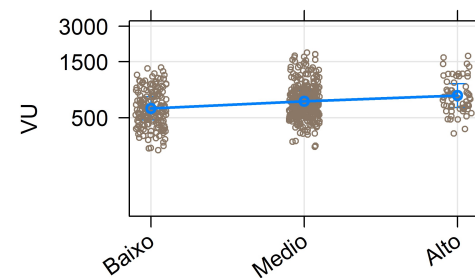
MODELO MISTO:



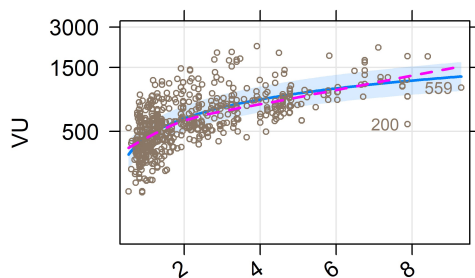
PAV



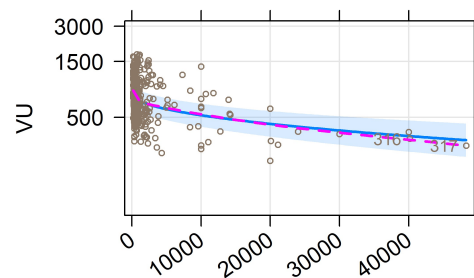
ESQUINA



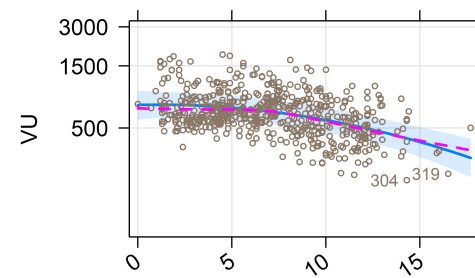
PADRAO



REDA



AREA



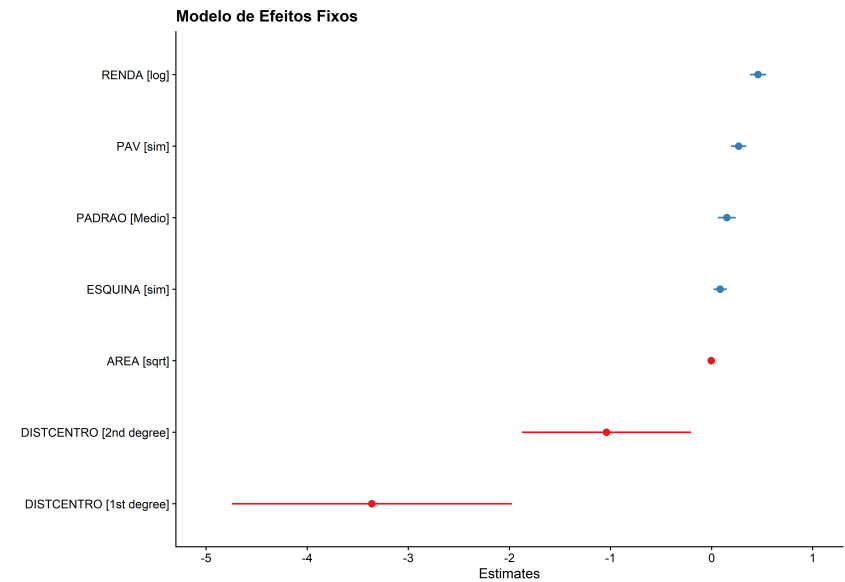
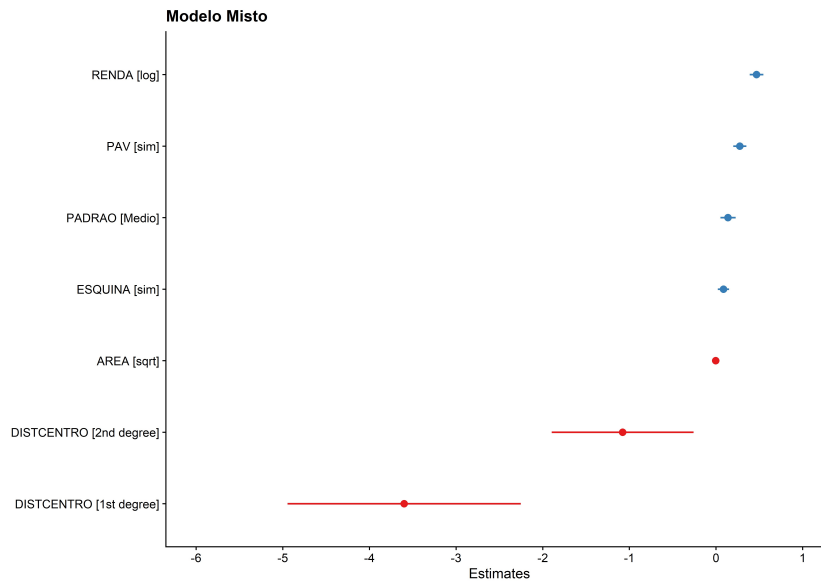
DISTCENTRO

Planta Genérica de Valores

MODELO MISTO vs. MODELO DE EFEITOS FIXOS:

MODELO MISTO

MODELO DE EFEITOS FIXOS

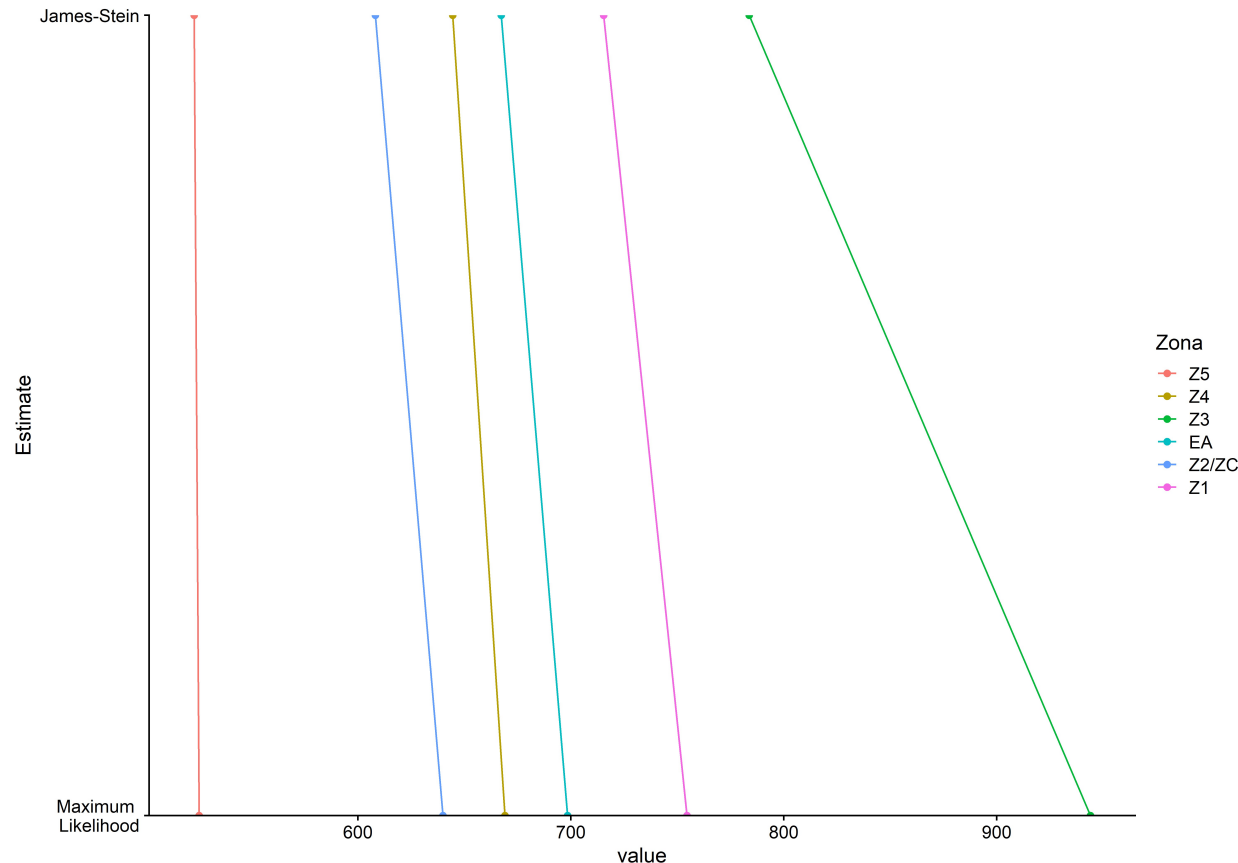


11^A15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

PLANTA GENÉRICA DE VALORES

BENEFÍCIO DO ENCOLHIMENTO DO TERMO ALEATÓRIO:

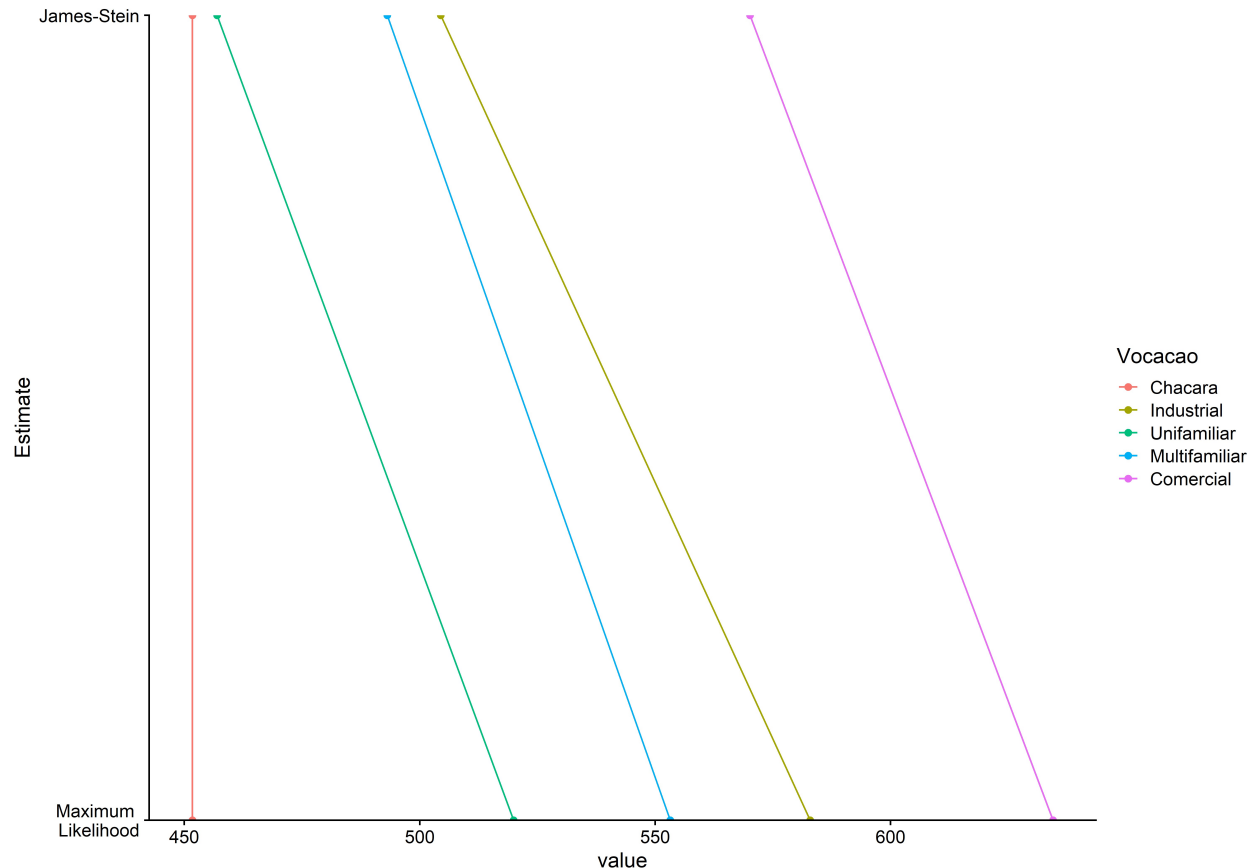


11^A15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO

PLANTA GENÉRICA DE VALORES

BENEFÍCIO DO ENCOLHIMENTO DO TERMO ALEATÓRIO:



11 A 15
SETEMBRO | 2023
SÃO PAULO - SP

REALIZAÇÃO



OBRIGADO!

lfpdroubi@gmail.com

(17)98188-1150

