



DCP098

Fundamentos para Avaliação Quantitativa de Políticas Públicas

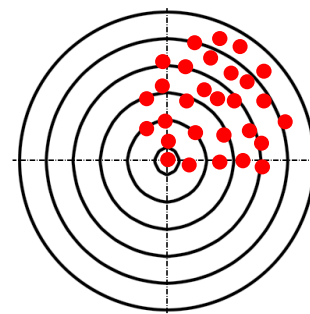
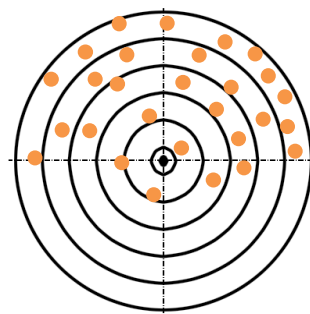
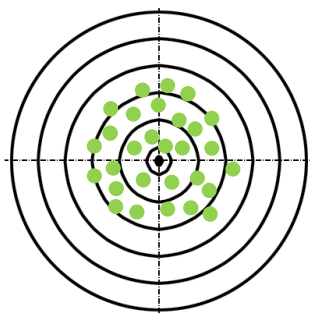
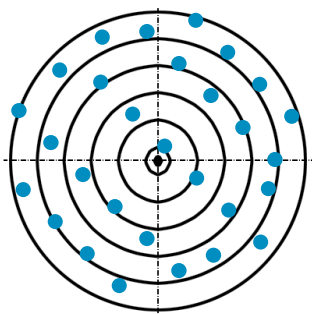
Precisão das estimativas

Teorema Central do Limite

Aula 10
04 de maio de 2022

Ana Paula Karruz

Acurácia, precisão e a distribuição teórica de β hats



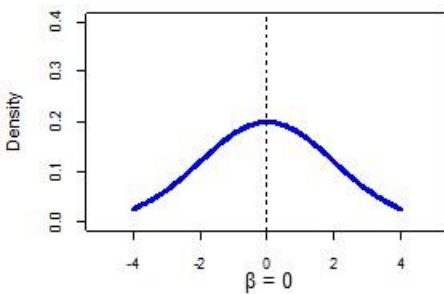
Acurácia
(ausência
de viés)



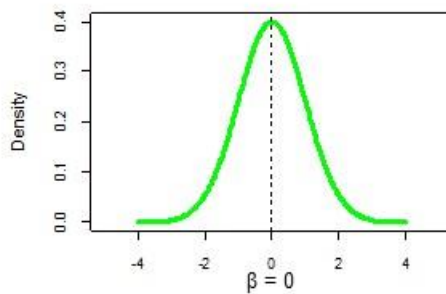
Precisão
(baixa
dispersão)



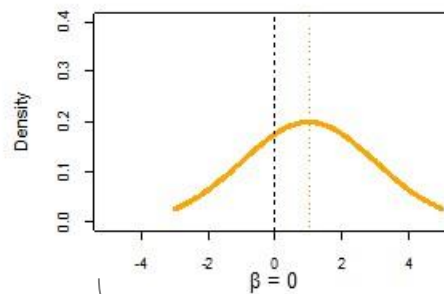
PDF Beta-hat



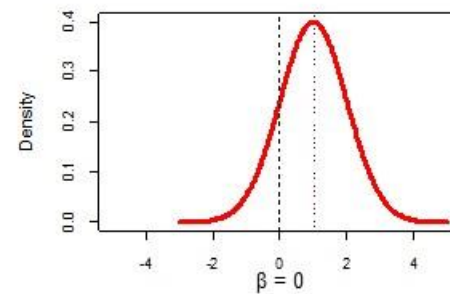
PDF Beta-hat



PDF Beta-hat



PDF Beta-hat



Estes cenários mostram viés positivo (centro da distribuição de β hats está à direita do verdadeiro β). Viés negativo é igualmente prejudicial à acurácia do modelo

Dois desafios à análise estatística: aleatoriedade e endogeneidade

Fontes de incerteza quanto ao efeito estimado de X sobre Y

Sampling randomness: amostras de diferentes tamanhos geram coeficientes estimados diferentes; amostras diferentes de um mesmo tamanho também geram coeficientes estimados diferentes; na estatística frequentista, coeficiente populacional é fixo)

Modeled randomness: aleatoriedade e complexidade na formação de Y redundam em variáveis omitidas; nota: aqui não estamos falando de variáveis omitidas correlacionadas com X

Variáveis omitidas correlacionadas com X: existência dessas variáveis implica espuriedade

Aleatoriedade
(compromete a
precisão)

Endogeneidade
(compromete a
acurácia)

FOCO DE HOJE:
aleatoriedade

$\hat{\beta}_1$

(ou qualquer outro
coeficiente de
inclinação estimado)

Sampling
randomness

Sempre que observamos uma relação nos dados, precisamos ter em mente que alguma coincidência poderia explicá-la. Talvez tenhamos escolhido algumas pessoas incomuns para nosso conjunto de dados. Ou talvez tenhamos escolhido pessoas perfeitamente representativas, mas elas tinham medidas incomuns no dia em que as medimos.

Modeled
randomness

No exemplo do donut, a possibilidade de tal aleatoriedade deve nos preocupar, pelo menos um pouco. Talvez as pessoas na Figura 1.3 sejam um pouco estranhas. Talvez se tivéssemos mais pessoas, poderíamos ter mais pessoas pesadas que não comem rosquinhas e pessoas magras que as comem. Adicionar essas pessoas à figura mudaria a figura e nossas conclusões. Ou talvez mesmo com o conjunto de pessoas que observamos, podemos ter pegado alguns deles em um dia ruim (ou bom) e, se olharmos para eles outro dia, podemos observar uma relação diferente [entre consumo de donuts e massa].



<https://www.petz.com.br/blog/especies/como-criar-um-furao/>

Pense em alguma população, digamos a população de furões na Flórida. Suponha que queremos saber se os furões velhos dormem mais.

[...]

Mesmo se pudéssemos obter dados sobre cada um deles, nosso modelo teria elementos aleatórios. Os padrões de sono do furão (a variável dependente) estão sujeitos à aleatoriedade que entra no termo de erro. Talvez um furão tenha comido um pouco de aipo a mais que de costume, outro ficou preso em uma gaveta e outro ainda se desentendeu com sua parceira. Ter dados sobre cada furão não mudaria o fato de que fatores não medidos denotados por ϵ afetam o sono do furão.

Em outras palavras, há aleatoriedade inerente no processo de geração de dados, mesmo quando os dados são medidos para uma população inteira. Portanto, mesmo que observemos uma população completa em um determinado momento, de modo que não tenhamos variação amostral, teremos aleatoriedade devido ao processo de geração de dados. Em outras palavras, a perspectiva da aleatoriedade modelada [modeled randomness] destaca o fato de que praticamente todo modelo possui algum componente não medido que explica parte da variação em nossa variável dependente.

Bailey (2016: 79-80)

*Portanto, toda análise estatística legítima levará em conta a aleatoriedade em um **esforço para distinguir resultados que poderiam ocorrer por acaso daqueles que provavelmente não ocorreriam por acaso**. A má notícia é que **nunca escaparemos da possibilidade** de que os resultados que observamos sejam **devidos à aleatoriedade** e não a algum efeito causal. A boa notícia, porém, é que muitas vezes podemos fazer um bom trabalho **caracterizando o quão confiantes estamos** de que os resultados não sejam simplesmente devidos à aleatoriedade.*

Bailey (2016: 13)

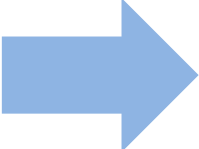
A estimativa de β_1 herda aleatoriedade (amostral e/ou modelada)



Portanto, $\hat{\beta}_1$ é uma variável aleatória. *What?*



Uma variável aleatória é uma variável cujos valores são **resultados numéricos de um fenômeno aleatório**



Variáveis aleatórias podem ser **discretas**, quando assumem um valor enumerável, que podemos contar (e.g., conversões de tiro livre no basquete, resultados de cara ou coroa ou rolagem de dados, número de consultas ao médico em um determinado ano), ou **contínuas** (e.g., estatura, massa, PIB)

A estatura pode ser representada em um número não numerável de casas decimais; em “zoom”, 1,80m pode assumir infinitos valores: 1,801, 1,8001, 1,80007, ...



$\hat{\beta}_1$ é uma variável aleatória contínua

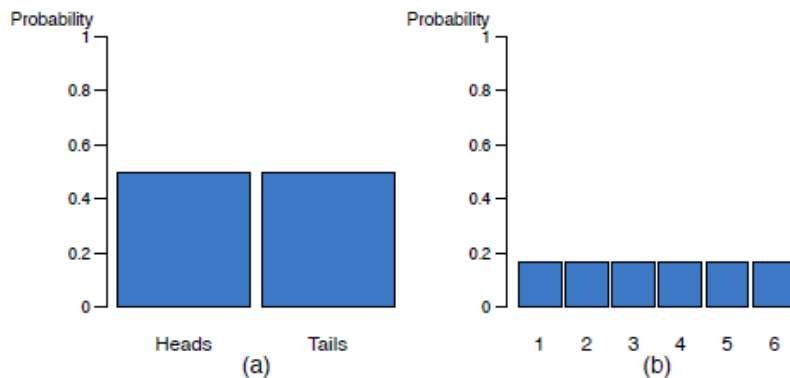
Para entender β_1 hats como aleatórios, convém pensar em sua distribuição

- Distribuição de β_1 hats: compreende **todos os possíveis valores que β_1 hat pode assumir**, e a **probabilidade** relativa desses valores
- Essa distribuição, chamada **distribuição amostral de β_1 hat**, é uma distribuição “**teórica**”: não nos é visível, pois observamos apenas o β_1 hat gerado por nossa (tipicamente única) amostra

Para entender β_1 hats como aleatórios, convém pensar em sua distribuição

- No caso de **variáveis aleatórias discretas**, podemos identificar a probabilidade de cada resultado possível porque há **uma quantidade enumerável de resultados possíveis**
- No caso de **variáveis aleatórias contínuas**, não podemos identificar a probabilidade de cada resultado possível porque há um **número ilimitado de resultados possíveis**

Discrete random variables with a finite number of possible outcomes



Continuous random variables

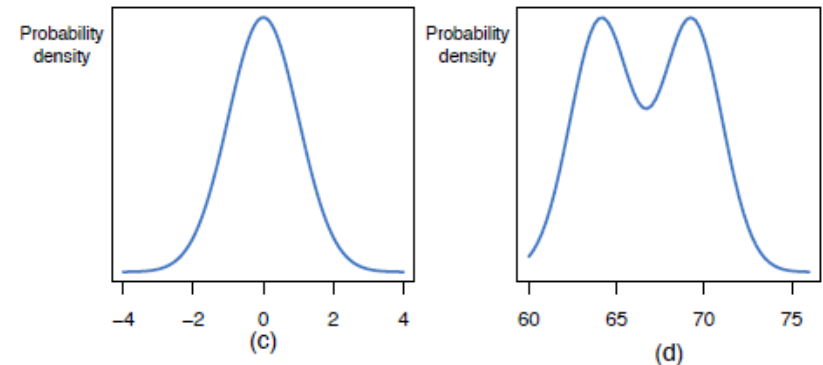


FIGURE 3.4: Four Distributions

Adult heights
(inches)

$\hat{\beta}_1$ é uma variável aleatória contínua, para a qual identificamos a probabilidade de intervalos de valores

- Para variáveis aleatórias contínuas, identificamos a **função densidade de probabilidade (FDP/ PDF – probability density function)**, que descreve a **probabilidade** de uma variável aleatória contínua assumir valores dentro de um intervalo

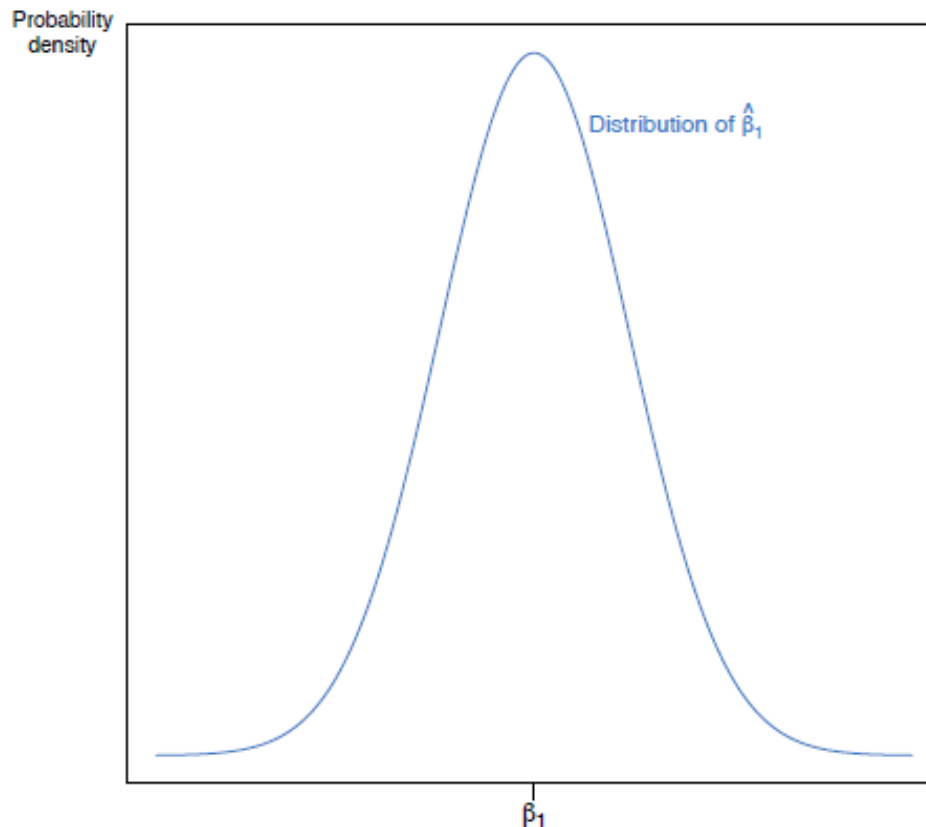
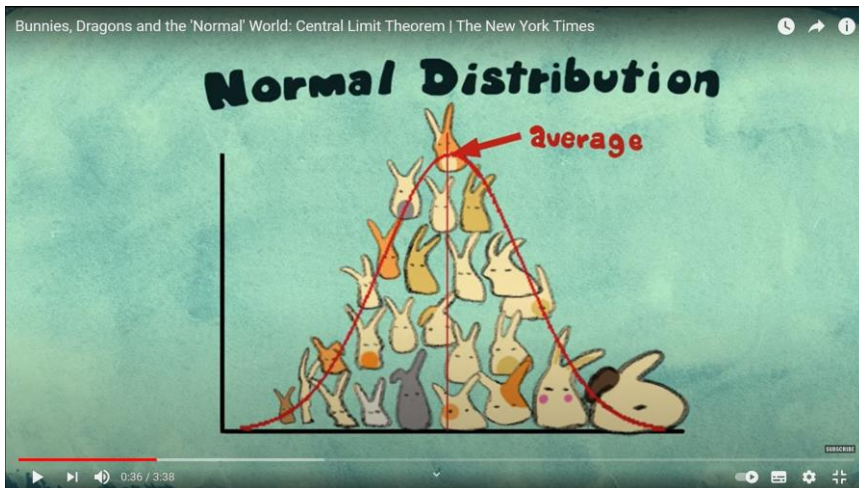


FIGURE 3.5: Distribution of $\hat{\beta}_1$

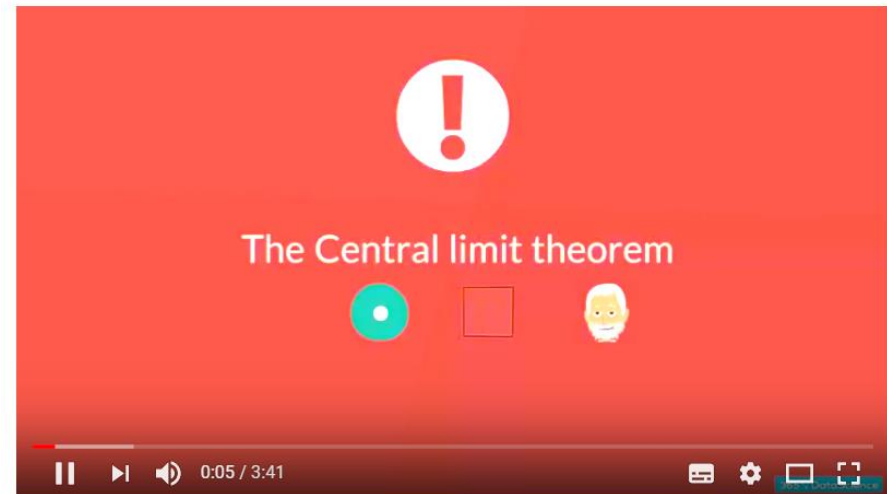
Fonte: Bailey (2016: 88).

Distribuição de densidade de $\beta_1\text{hat}$ se aproxima de uma normal

- **Fórmula** do $\beta_1\text{hat}$ pode ser **reescrita como uma média** (Bailey, 2016: 85, nota 8)
- Como tal, $\beta_1\text{hat}$ é uma **estatística amostral sujeita ao Teorema Central do Limite** (aka Teorema do Limite Central), segundo o qual:
 - A média amostral de qualquer variável aleatória segue uma distribuição normal
 - Quanto maior o tamanho da amostra, a distribuição amostral da média:
 - Será mais parecida com uma distribuição normal
 - Terá centro mais próximo da média populacional



<https://youtu.be/jvoxEYmQHNM>



<https://youtu.be/b5xQmk9veZ4>
<https://youtu.be/YAIJCEDH2uY>



DCP098

Fundamentos para Avaliação Quantitativa de Políticas Públicas

Precisão das estimativas

Teorema Central do Limite

Aula 10
04 de maio de 2022

Ana Paula Karruz