PROGRAMA DE
PÓS-GRADUAÇÃO EM
CIÊNCIA POLÍTICA | UFMG

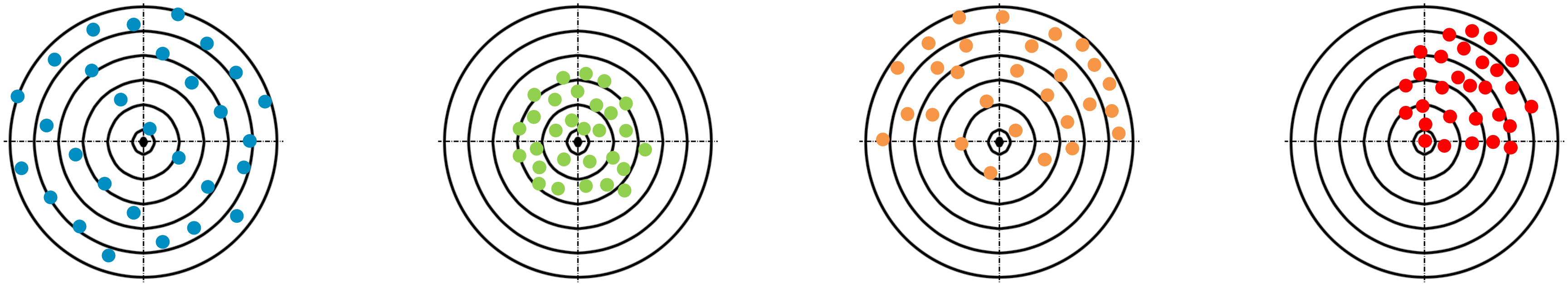
DCP919B
Metodologia II

Fundamentos para teste de hipótese

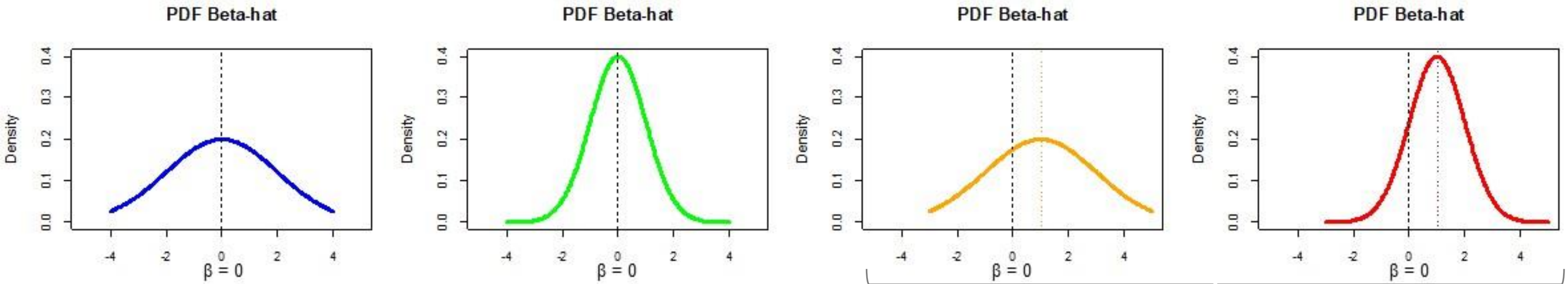
Aula 3
21 de setembro de 2022

Ana Paula Karruz

Acurácia, precisão e a distribuição teórica de βhats



Acurácia (ausência de viés)	✓	✓		
Precisão (baixa dispersão)		✓		✓



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

Dois desafios à análise estatística inferencial: aleatoriedade e endogeneidade

Fontes de incerteza quanto ao efeito estimado de X sobre Y

FOCO DE HOJE: aleatoriedade

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

Aleatoriedade (compromete a **precisão**)

Variáveis omitidas correlacionadas com X: existência dessas variáveis implica espuriedade

Endogeneidade (compromete a **acurácia**)

$$\hat{\beta}_j$$

(qualquer 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].

Bailey (2016: 13)

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_j herda aleatoriedade (amostral e/ou modelada)

Portanto, $\beta_j$hat é uma variável aleatória. ***What?***

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

Variáveis aleatórias quantitativas dividem-se em dois tipos:

- **Discretas:** têm um número contável de valores entre quaisquer dois valores
  - E.g.: número de conversões de tiro livre no basquete; número de consultas ao médico em um determinado ano; avaliação de um serviço, medida em uma escala de 0 a 5, em intervalos de 0,5

Note: uma variável aleatória discreta não necessariamente pertence aos inteiros

- **Contínuas:** têm um número infinito de valores entre dois valores quaisquer
  - E.g.: estatura; massa; PIB; hora

A estatura pode ser representada em um quantidade não enumerável de casas decimais; em "zoom", 1,80m pode assumir infinitos valores: 1,801m, 1,8001m, 1,80007m, ...

$\beta_j$hat é uma variável aleatória contínua

Para entender $\beta_j$hats como aleatórios, convém pensar em sua distribuição

- Distribuição de $\beta_j$hats: compreende **todos os possíveis valores que $\beta_j$hat pode assumir**, e a **probabilidade** relativa desses valores

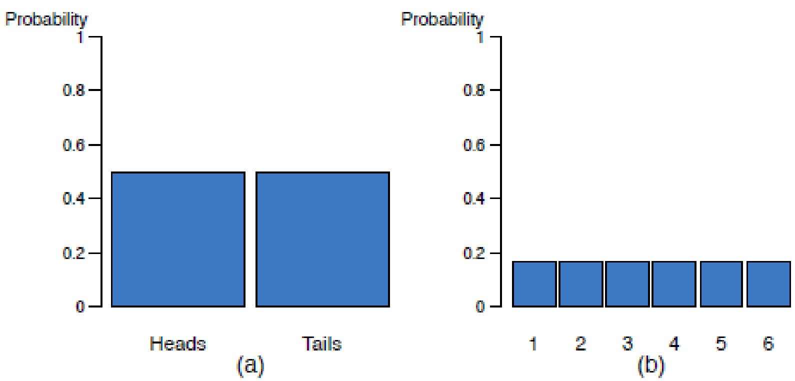
- Essa distribuição, chamada **distribuição amostral de $\beta_j$hat**, é uma distribuição **"teórica"**: não nos é visível, pois observamos apenas o $\beta_j$hat gerado por nossa (tipicamente única) amostra

Para entender $\beta_j$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

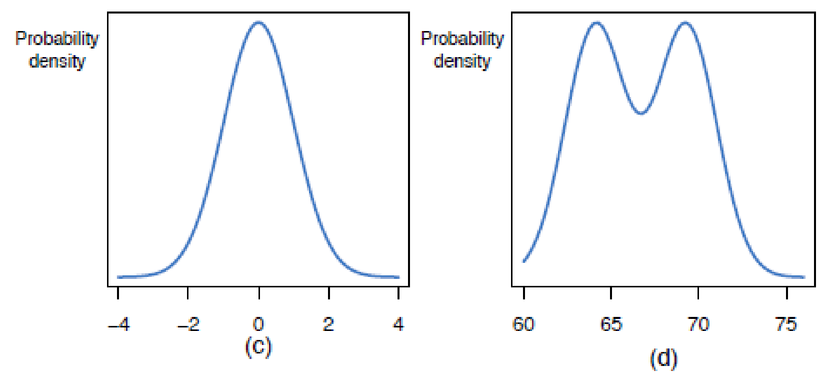


FIGURE 3.4: Four Distributions

Adult heights (inches)

Fonte: Adaptado de Bailey (2016: 82).

$\beta_j$hat é uma variável aleatória contínua, para a qual identificamos a probabilidade de <u>intervalores</u> 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**

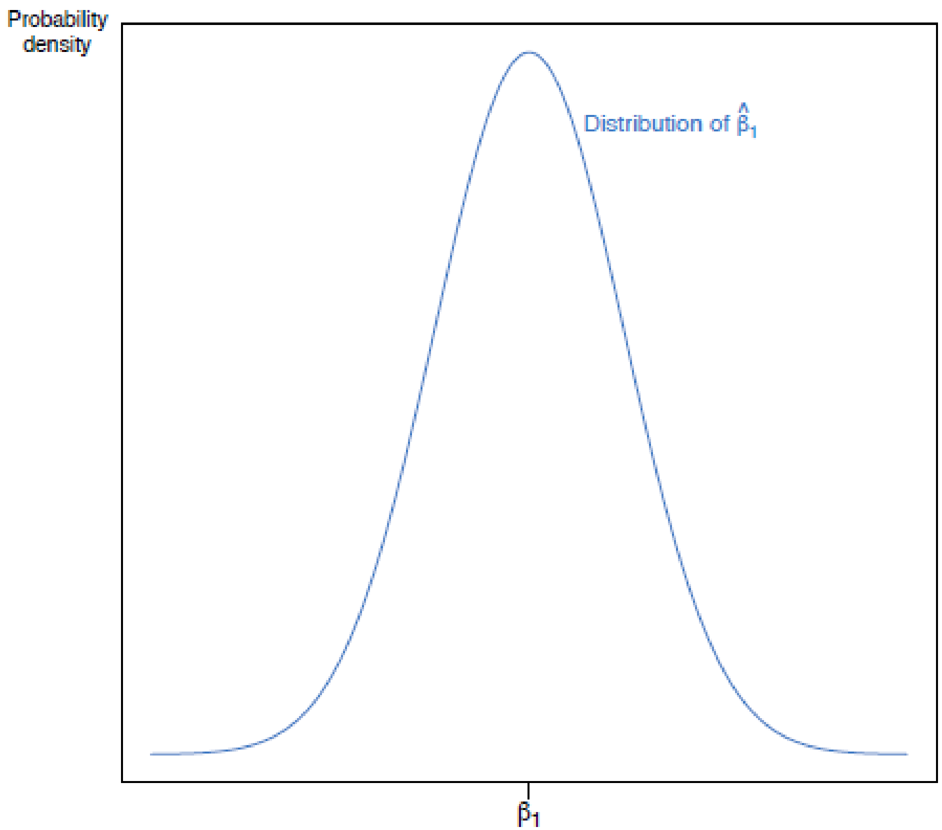


FIGURE 3.5: Distribution of $\hat{\beta}_1$

Fonte: Bailey (2016: 88).

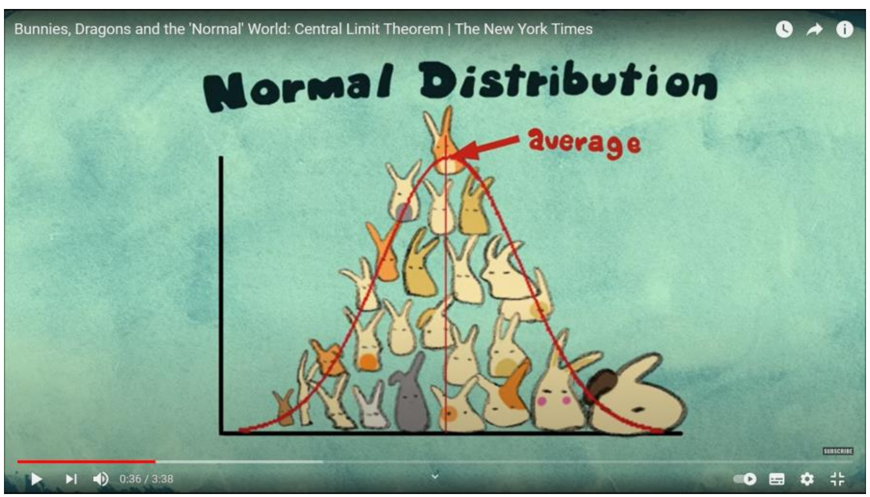
Distribuição de densidade de $\beta_j$hats se aproxima de uma normal

- **Fórmula** do $\beta_j$hat pode ser **reescrita como uma média** (Bailey, 2016: 85, nota 8)

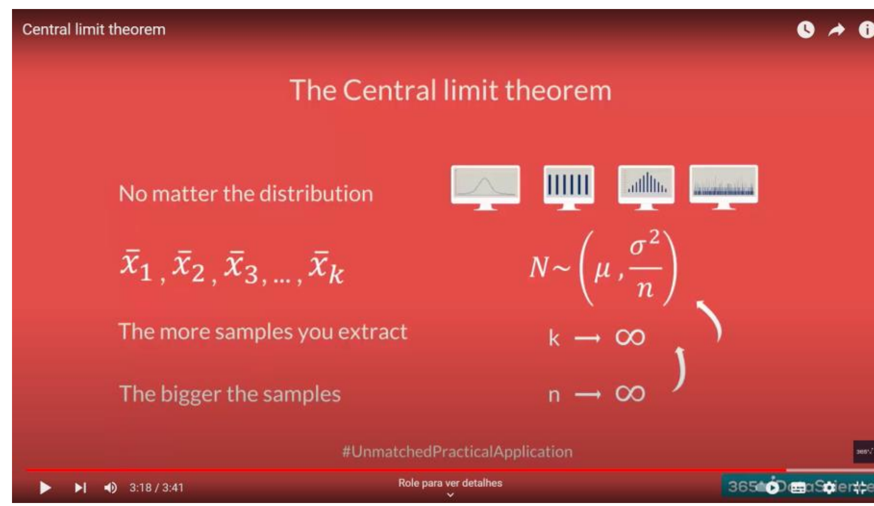
- Como tal, **$\beta_j$hat é uma estatística amostral sujeita ao Teorema Central do Limite** (aka Teorema do Limite Central), segundo o qual:
  - As médias amostrais de qualquer variável aleatória se distribuem de maneira próxima a uma distribuição normal centrada na média populacional (μ)

$$\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots, \bar{x}_k \sim N(\mu, {\sigma^2}/{n})$$

  - Quanto maior o número de amostras (k), e quanto maior o tamanho de cada amostra (n), 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/YAlJCEDH2uY

[Pausa para um recordatório de Estatística]

Variância e desvio padrão

- A variância e o desvio padrão são medidas de dispersão: informam quão dispersos se encontram os valores de uma variável; quanto maior a variância, mais espalhados se encontram as observações (o mesmo vale para o desvio padrão)

- A **variância populacional** de uma variável (e.g., X) é dada por:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

Fórmula assume que conhecemos a média populacional de X (i.e., μ)

- Para calcular a **variância amostral** de X (i.e., variância numa dada amostra), substituímos o N (tamanho da população) por n – 1 (tamanho da amostra – 1); o "– 1" é um ajuste pela perda de graus de liberdade incorrida no cálculo da média amostral; em amostras grandes, não importa se usamos n ou n – 1 no denominador, pois o ajuste não causará diferença material no valor da variância amostral

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

- É útil **"desconstruir"** o que a formula da variância faz:
  - Para cada observação, obtenha o desvio em relação à média
  - Eleve cada desvio ao quadrado
  - Tire a média dos quadrados dos desvios

- **Variância = (Desvio padrão)^2**

- Ao calcularmos a **raiz quadrada da variância (para obter o desvio padrão)** trazemos o resultado final **de volta à escala original da variável**

Fonte: Bailey (2016:768). Nota: Qualquer livro de estatística básica terá uma seção sobre variância.

Desvio padrão e erro padrão

Ambos são medidas de variação

Desvio padrão

- Representa a **variabilidade (dispersão) de uma variável em uma amostra** (note: a amostra pode conter todas as unidades da população ou não)
  - E.g.: podemos obter o desvio padrão de X, em que X = peso de coelhos selvagens
- Corresponde à **raiz quadrada da variância da variável**
- **Desvio padrão populacional (*σ*):** requer que conheçamos *μ*, a média populacional de X

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

- **Desvio padrão amostral (*s*):** utilizamos a média amostral de X, i.e., Xbar

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

Erro padrão

- Representa a **variabilidade (dispersão) de uma estimativa entre amostras** (i.e., mede a dispersão da distribuição "teórica" de infinitas estimativas obtidas a partir de amostras de mesmo tamanho)
  - E.g.: podemos obter o erro padrão de Xbar, em que Xbar = média amostral do peso de coelhos selvagens em infinitas amostras com 50 coelhos cada
- Corresponde à **raiz quadrada da variância da estimativa**
- **Erro padrão da média (populacional):** requer que conheçamos *σ*, o desvio padrão populacional de X

$$EP = \frac{\sigma}{\sqrt{n}}$$

- **Erro padrão da média (estimado):** utilizamos *s*, o desvio padrão amostral de X

$$EP_{estimado} = \frac{s}{\sqrt{n}}$$

Na análise de regressão, frequentemente estamos interessados em estimar o erro padrão de βhats, a partir de uma fórmula diferente desta e que será detalhada adiante

Distribuição normal

- Trata-se uma distribuição de probabilidade de variável aleatória contínua (digamos, X) baseada em dois parâmetros: *μ* (média populacional de X) e *σ* (desvio padrão populacional de X)

$$X \sim N(\mu, \sigma)$$

- Os valores da distribuição normal são dados pela fórmula abaixo; com ela, podemos "desenhar" uma distribuição, desde que conheçamos *μ* e *σ*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- A distribuição normal é simétrica, centrada em sua média (*μ*) e apresenta forma de sino; o espalhamento da distribuição normal é definido por (*σ*)

- O pico (valor mais alto, localizado em x = *μ*) é dado pela seguinte fórmula (obtida pela simples substituição de x = *μ* na fórmula acima):

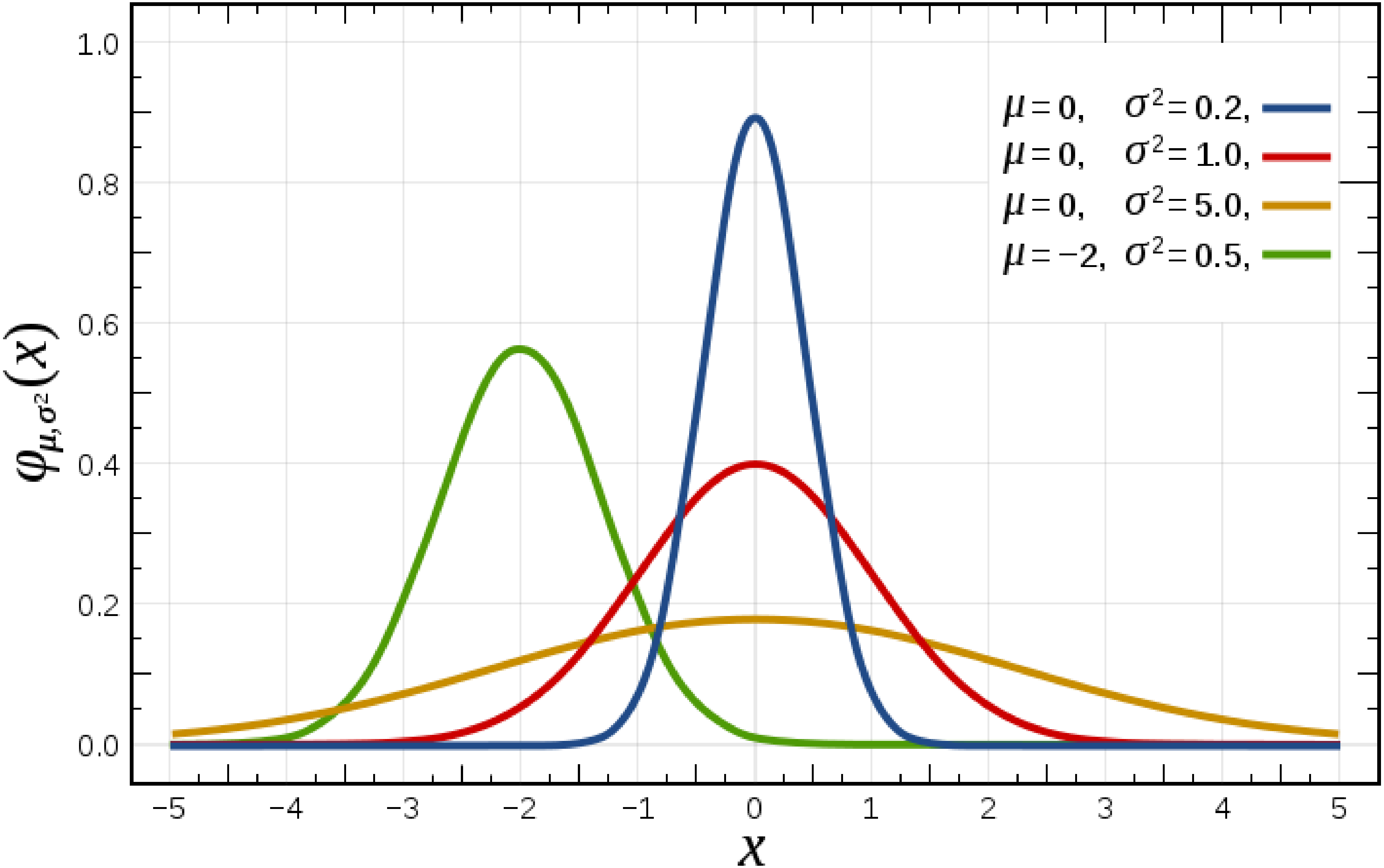
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}$$

- Para $X \sim N(0, 1)$, o valor máximo (pico ou cume) corresponde a aproximadamente 0,40:

$$f(0) = 1/(\sqrt{2\pi}) = 1/(\sqrt{6{,}283185}) = 0{,}398942$$

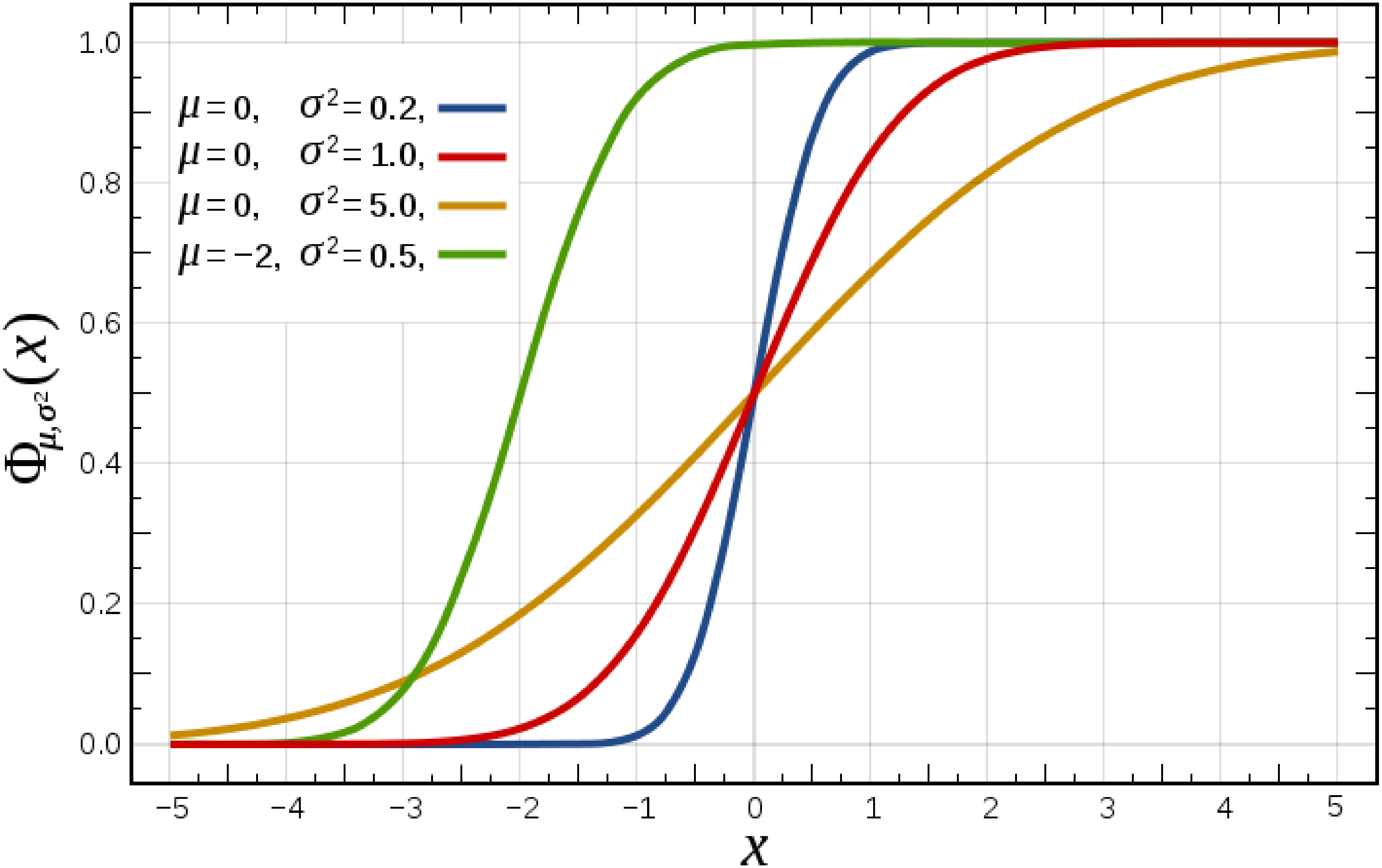
Fonte: Bailey (2016:774-783). Nota: Qualquer livro de estatística básica terá uma seção sobre distribuição normal.

Função densidade de probabilidade para distribuições normais



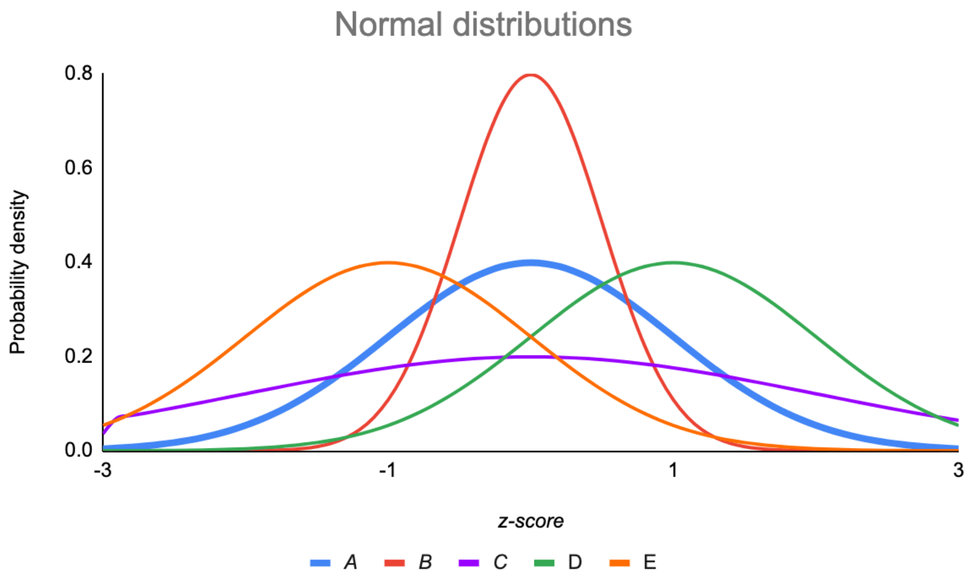
Fonte: https://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg

Função densidade de probabilidade ACUMULADA para distribuições normais



Fonte: https://en.wikipedia.org/wiki/File:Normal_Distribution_CDF.svg

A distribuição $N(0, 1)$ é conhecida como “normal padrão” ou distribuição Z



Curve	Position or shape (relative to standard normal distribution)
A (M = 0, SD = 1)	Standard normal distribution
B (M = 0, SD = 0.5)	Squeezed, because SD < 1
C (M = 0, SD = 2)	Stretched, because SD > 1
D (M = 1, SD = 1)	Shifted right, because M > 0
E (M = −1, SD = 1)	Shifted left, because M < 0

Fonte: https://www.scribbr.com/statistics/standard-normal-distribution/

- Para **“padronizar” uma distribuição normal** (i.e., fazê-la ter média nula e desvio padrão unitário), aplicamos a seguinte fórmula:

$$z = \frac{x - \mu}{\sigma}$$

  em que:
  x = qualquer valor da distribuição normal original (aquela que se quer padronizar)
  μ = média da distribuição normal original
  σ = desvio padrão da distribuição normal original

- Um **z-score** (valor de Z para um dado valor de X) positivo significa que x é maior que a média de X
  - z-score > 0 ↔ x > μ
  - z-score < 0 ↔ x < μ
  - z-score = 0 ↔ x = μ

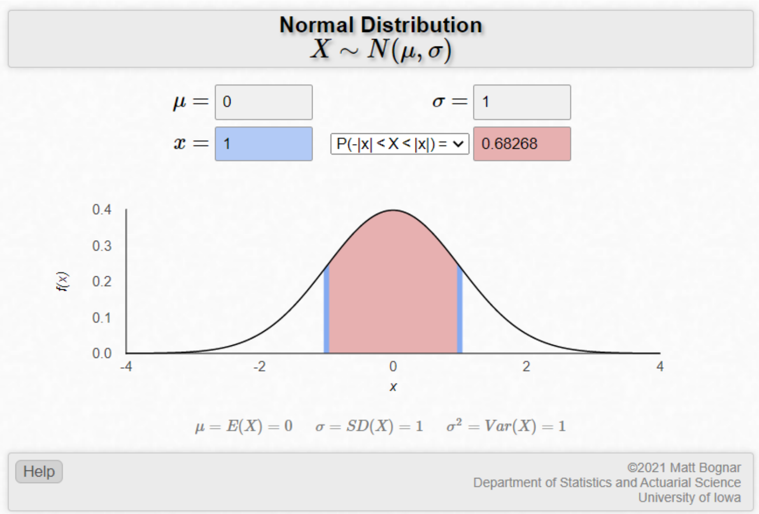
- Entre outras vantagens, **converter** uma distribuição normal **em uma distribuição normal padrão permite**:
  - Comparar valores de diferentes distribuições com diferentes médias e desvios padrão
  - Encontrar rapidamente (numa tabela) a probabilidade de observações em uma distribuição caírem acima ou abaixo de um determinado valor

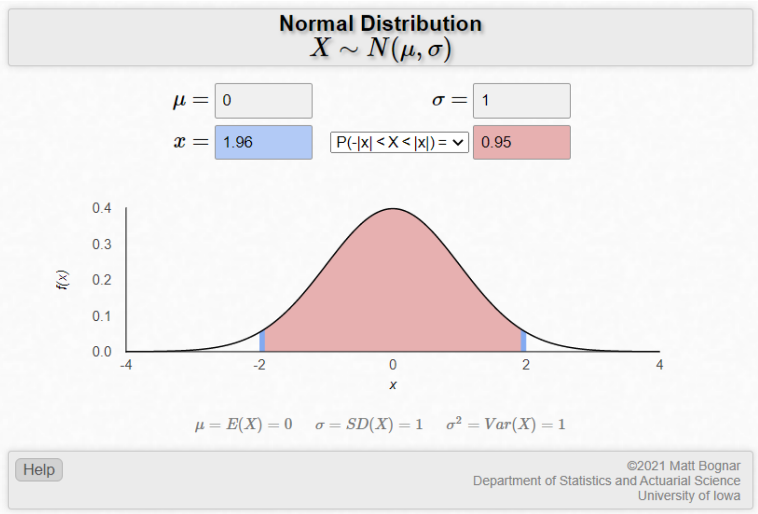
- Quando os parâmetros populacionais μ e σ são desconhecidos, é possível estimar o z-score a partir das estatísticas amostrais $\bar{x}$ (média amostral de X) e s (desvio padrão amostral de X). **Alerta:** s tende a subestimar σ

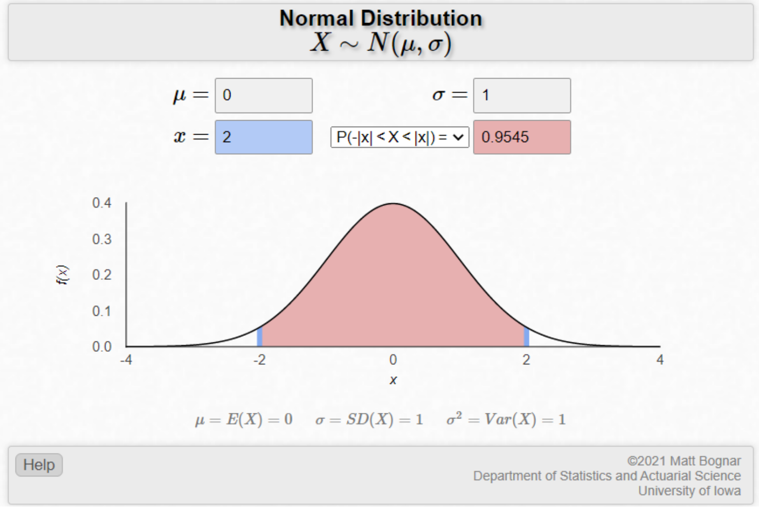
$$z_{estimado} = \frac{x - \bar{x}}{s}$$

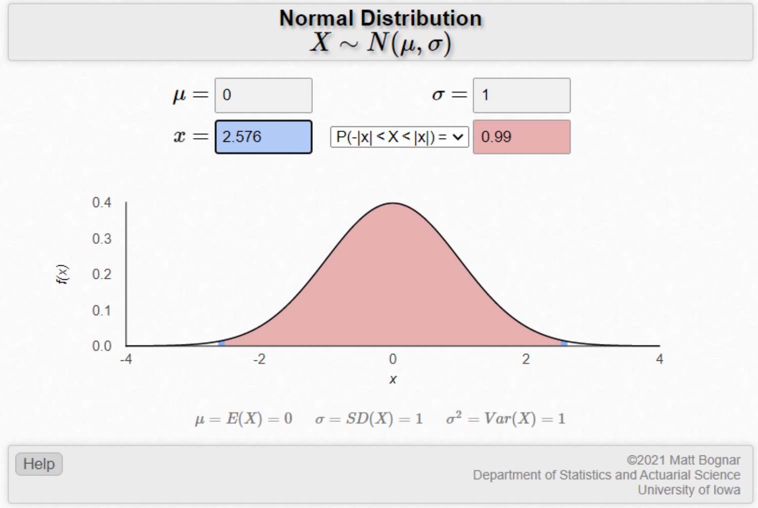
Probabilidade em intervalos de uma distribuição normal padrão

https://homepage.divms.uiowa.edu/~mbognar/applets/normal.html









Outras distribuições de probabilidade disponíveis em:

https://homepage.divms.uiowa.edu/~mbognar/applets

[Fim do recordatório de Estatística]

Distribuição de densidade de $\beta_j$hats se aproxima de uma normal

- Para **amostras grandes** (digamos, 100+), **$\beta_j$hats são normalmente distribuídos**
  - Se os erros do modelo (ε) forem normalmente distribuídos, então $\beta_j$hats serão normalmente distribuídos independentemente do tamanho da amostra
- Portanto, enquanto normal, a **distribuição amostral ("teórica") de $\beta_j$hats**:
  - É simétrica, centrada em sua média (μ), e apresenta forma de sino
  - Tem "espalhamento" definido pela variância (σ^2)

Aleatoriedade, centralidade e dispersão de $\beta_j$hat

- **$\beta_j$hats são aleatórios**; conforme a sorte, podemos obter uma estimativa distante de β_j

- Além da **centralidade**, interessa considerar a **dispersão** da distribuição "teórica" de $\beta_j$hats, para avaliarmos quão provável é que estimemos um coeficiente próximo do parâmetro populacional

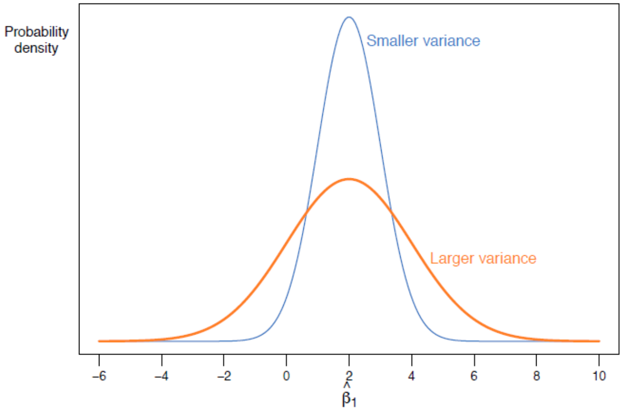


FIGURE 3.6: Two Distributions with Different Variances of $\hat{\beta}_1$

Fonte: Bailey (2016: 94).

– Se a condição de **exogeneidade** estiver atendida, a **média dos $\beta_j$hats é β_1**

– O que podemos dizer sobre a **variância de $\beta_j$hat**?

Numa regressão <u>bivariada</u>, a variância de $\beta_1$hat* é dada por:

[Obs.: A variância de $\beta_0$hat é dada por outra fórmula.]

Erro padrão de $\beta_j$hat, o **se($\beta_j$hat)** = Raiz quadrada da var($\beta_j$hat)

$$var(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{n * var(X)}$$

Variância da regressão = (Residual standard error)^2. Obtenha o residual standard error no output da regressão

$\hat{\sigma}^2$

Variância da regressão

- Mede quão bem o modelo explica a variação de Y
- Seu cálculo baseia-se nos resíduos
- Aqui, k = número de variáveis explicativas
- É também uma estimativa da variância de ε
- **Intuição:** média do quadrado da distância entre valores observados e previstos de Y

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n - k - 1}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\hat{\epsilon}_i^2}{n - k - 1}$$

---

n

Tamanho da amostra

- **Intuição:** mais dados implicam menor variância, pois a chance de o acaso nos levar às caudas da distribuição de $\beta_1$hat é menor em amostras maiores (i.e., menor sampling randomness)

$var(X)$

Variância X (amostral)

- Quanto mais X variar, mais precisa será a distribuição de $\beta_1$hat
- **Intuição:** se X varia pouco, não temos muita informação para estimar o efeito da variação de X sobre a variação de Y

$$var(X) = s^2$$

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

* Fórmula de var($\beta_1$hat) é mais complicada quando erros são correlacionados ou heteroscedásticos, mas as intuições sobre variância da regressão, tamanho da amostra e var(X) se aplicam. Voltaremos a esse ponto em aulas futuras.

Extraindo o EP($\beta_1$hat) do output da regressão

```
Call:
lm(formula = dados$Weight..pounds. ~ dados$Donuts.per.week)

Residuals:
    Min       1Q   Median       3Q      Max 
-92.731 -13.508   3.916   36.081   55.716 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)            121.613     16.593   7.329 1.49e-05 ***
dados$Donuts.per.week    9.224      1.959   4.707 0.000643 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 45.81 on 11 degrees of freedom
Multiple R-squared:  0.6683,    Adjusted R-squared:  0.6381 
F-statistic: 22.16 on 1 and 11 DF,  p-value: 0.0006426

######
> sigma(reg.pounds) # to obtain the residual standard error
with more decimal places
[1] 45.81383
> var(dados$Donuts.per.week)
[1] 45.55644
```

Vide aula 02

Calculando o EP($\beta_1$hat) "manualmente"

$$var(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{n * var(X)}$$

$$var(\hat{\beta}_1) = \frac{45{,}81383^2}{13 * 45{,}55644}$$

$$var(\hat{\beta}_1) = \frac{2.098{,}90702}{592{,}23372}$$

$$var(\hat{\beta}_1) = 3{,}54405$$

$$EP(\hat{\beta}_1) = \sqrt{3{,}54405}$$

$$EP(\hat{\beta}_1) = 1.88257$$

R trabalha com mais casas decimais e obtém valor mais alto (1,959)

Numa regressão <u>bivariada</u>, a variância de $\beta_0$hat é dada por:

note: $Var(\hat{\beta}_0) = Var(\bar{Y} - \hat{\beta}_1 \bar{X})$

Erro padrão do Intercepto na regressão SIMPLES

$$EP(\hat{\beta}_0) = \text{residual standard error} * \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{\sum (X-\bar{X})^2}}$$

$$\text{residual standard error} = \sqrt{\frac{\sum (Y-\hat{Y})^2}{DofF}}$$

Exemplo donuts:

residual standard error = 45,81

$n = 13$

$\bar{X}^2 = 29{,}66059$

$\sum (X-\bar{X})^2 = 546{,}67731$

$$EP(\hat{\beta}_0) = 45{,}81 * \underbrace{\sqrt{\underbrace{\underbrace{\frac{1}{13}}_{0{,}0769231} + \underbrace{\frac{29{,}66059}{546{,}67731}}_{0{,}0542561}}_{0{,}1311792}}}_{0{,}3621867}$$

$= 45{,}81 * 0{,}3621867$

$= 16{,}59177727$

$EP(\hat{\beta}_0)$ no R = 16,593

Não será cobrado

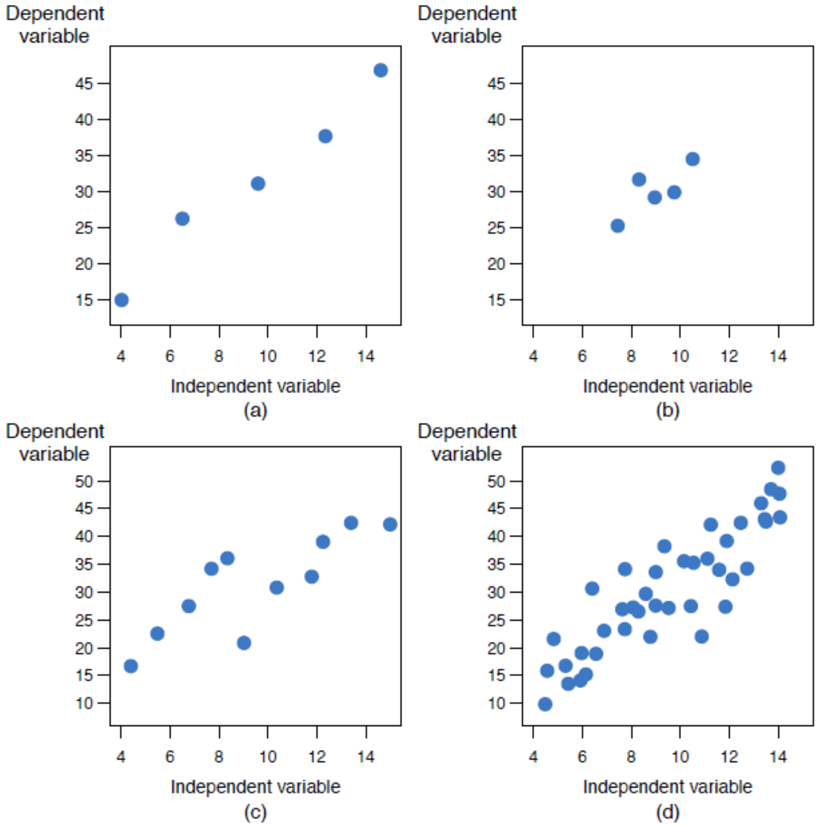


FIGURE 3.7: Four Scatterplots

Discussion Questions

1. Will the variance of $\hat{\beta}_1$ be smaller in panel (a) or panel (b) of Figure 3.7? Why?
2. Will the variance of $\hat{\beta}_1$ be smaller in panel (c) or panel (d) of Figure 3.7? Why?

Fonte: Bailey (2016: 98).

Se a condição de exogeneidade for atendida, MQO produzirá estimador consistente de β_j

Conteúdo não será cobrado

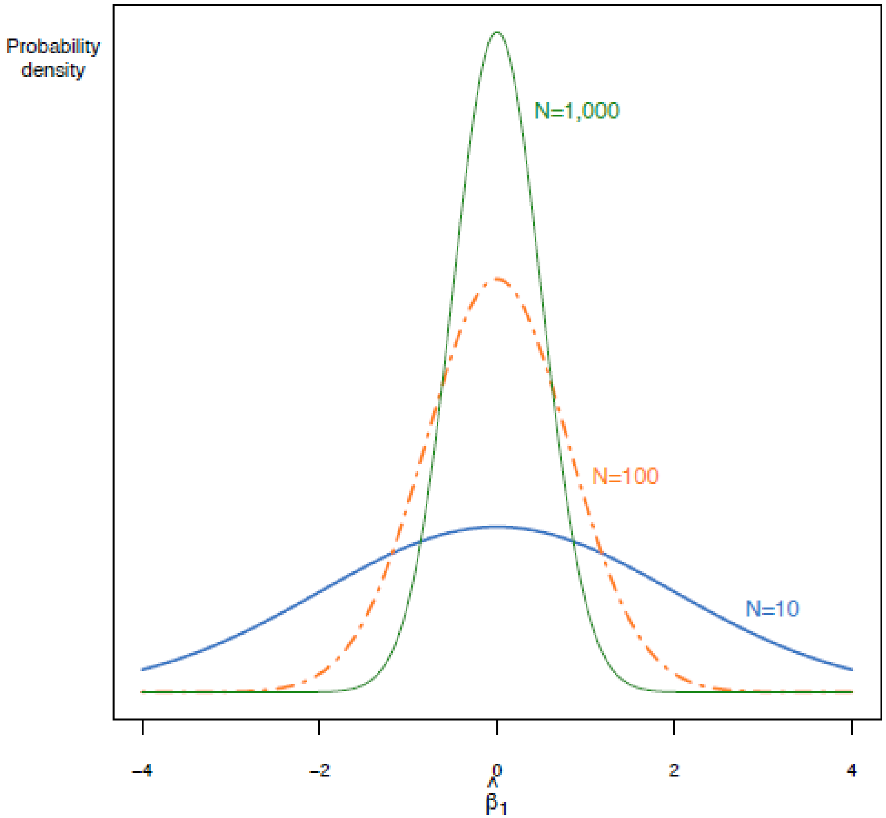


FIGURE 3.8: Distributions of $\hat{\beta}_1$ for different sample sizes

- Um estimador de β_1 é consistente se a respectiva distribuição de $\beta_1$hat se **aproxima** cada vez mais do verdadeiro $\boldsymbol{\beta_1}$ à medida que o tamanho da **amostra aumenta**

- Formalmente, consistência é definida como

$$\text{plim } \hat{\beta}_1 = \beta_1$$

plim = limite de probabilidade
plim de um estimador = valor para o qual a sampling distribution da estimativa gerada pelo estimador converge à medida que o tamanho da amostra aumenta

*As duas melhores coisas que você pode dizer sobre um estimador é que ele é **livre de viés e consistente**. Os estimadores OLS têm essas duas propriedades **quando o erro não está correlacionado com a variável independente**.*

Bailey (2016: 101)

Não será cobrado

Qual a diferença entre acurácia (unbiasedness) e consistência?

Se a condição de exogeneidade for atendida, MQO produzirá estimadores consistentes de β_j

Conteúdo não será cobrado

*Os livros didáticos às vezes não dão a devida atenção à **distinção entre ausência de viés [unbiasedness] e consistência**, mas a diferença pode ser importante na prática. **Ausência de viés** significa que o estimador tem uma distribuição amostral **centrada no parâmetro de interesse em uma amostra de qualquer tamanho**, enquanto **consistência** significa apenas que o estimador **converge para o parâmetro populacional à medida que o tamanho da amostra cresce**.*

Angrist e Krueger (2001: 4-5;
NBER Working Paper 8456;
https://doi.org/10.3386/w8456)

Não será cobrado

PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA POLÍTICA | UFMG

DCP919B
Metodologia II

Fundamentos para teste de hipótese

Aula 3
21 de setembro de 2022

Ana Paula Karruz