UFMG
GESTÃO PÚBLICA

DCP098
Fundamentos para Avaliação Quantitativa de Políticas Públicas

Precisão das estimativas
Teste de hipótese

Aulas 11-12
09 e 11 de maio de 2022

Ana Paula Karruz

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

FOCO DE HOJE: aleatoriedade

Aleatoriedade (compromete a **precisão**)

Endogeneidade (compromete a **acurácia**)

$$\hat{\beta}_1$$

(ou qualquer outro coeficiente de inclinação estimado)

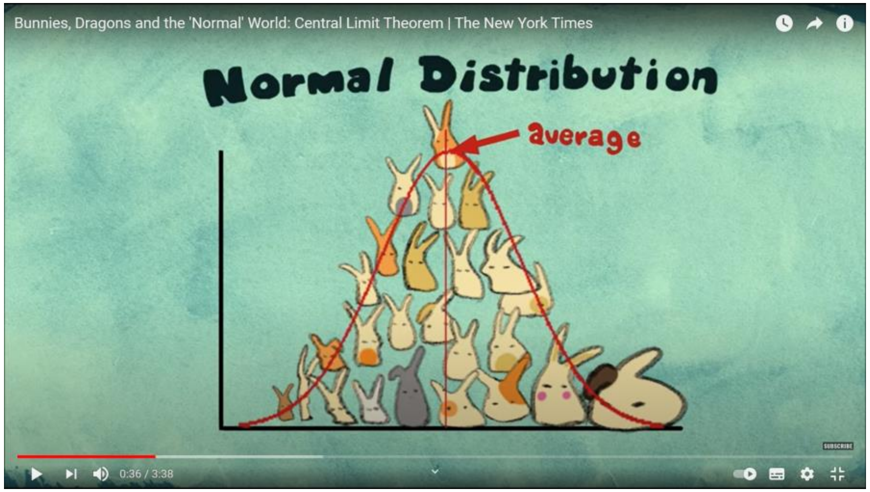
Distribuição de densidade de $\beta_1$hats se aproxima de uma normal

- **Fórmula** do $\beta_1$hat pode ser **reescrita como uma média** (Bailey, 2016: 85, nota 8)

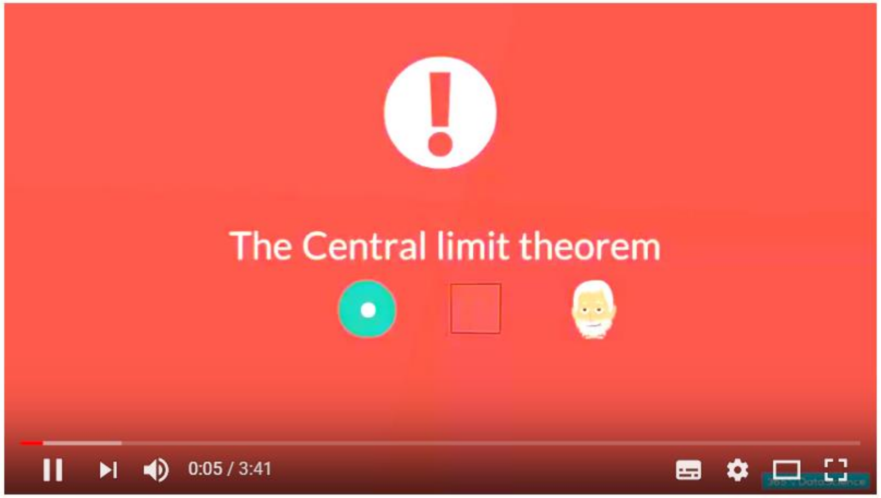
- Como tal, **$\beta_1$hat é uma estatística amostral sujeita ao Teorema Central do Limite** (aka Teorema do Limite Central), segundo o qual:

  - 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/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)

- **Variância = (Desvio padrão)^2**

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

$$var(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2$$

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

- Para calcular a variância de X numa dada amostra, substituímos o N por N-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

- 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 media dos quadrados dos desvios

- 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

- 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)

- **Variância = (Desvio padrão)^2**

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

$$var(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2$$

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

- Para calcular a variância de X numa dada amostra, substituímos o N por N-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

- 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 media dos quadrados dos desvios

O **desvio padrão** mede a dispersão de uma **variável** em uma população ou amostra. O **erro padrão** mede a **precisão de uma estimativa** (i.e., a dispersão da distribuição "teórica" de infinitas estimativas obtidas a partir de amostras de mesmo tamanho)

- 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.

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):

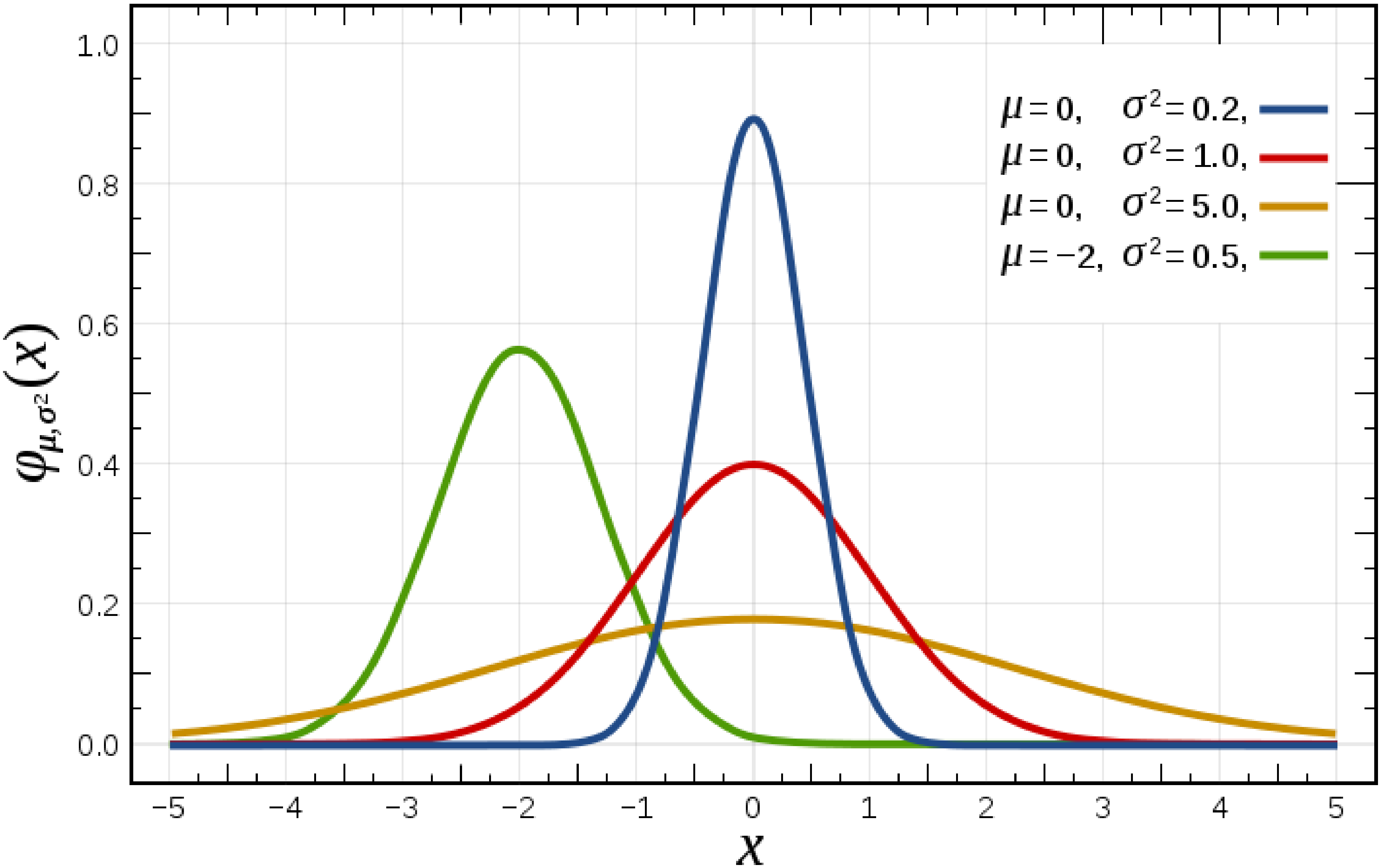
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}$$

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

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

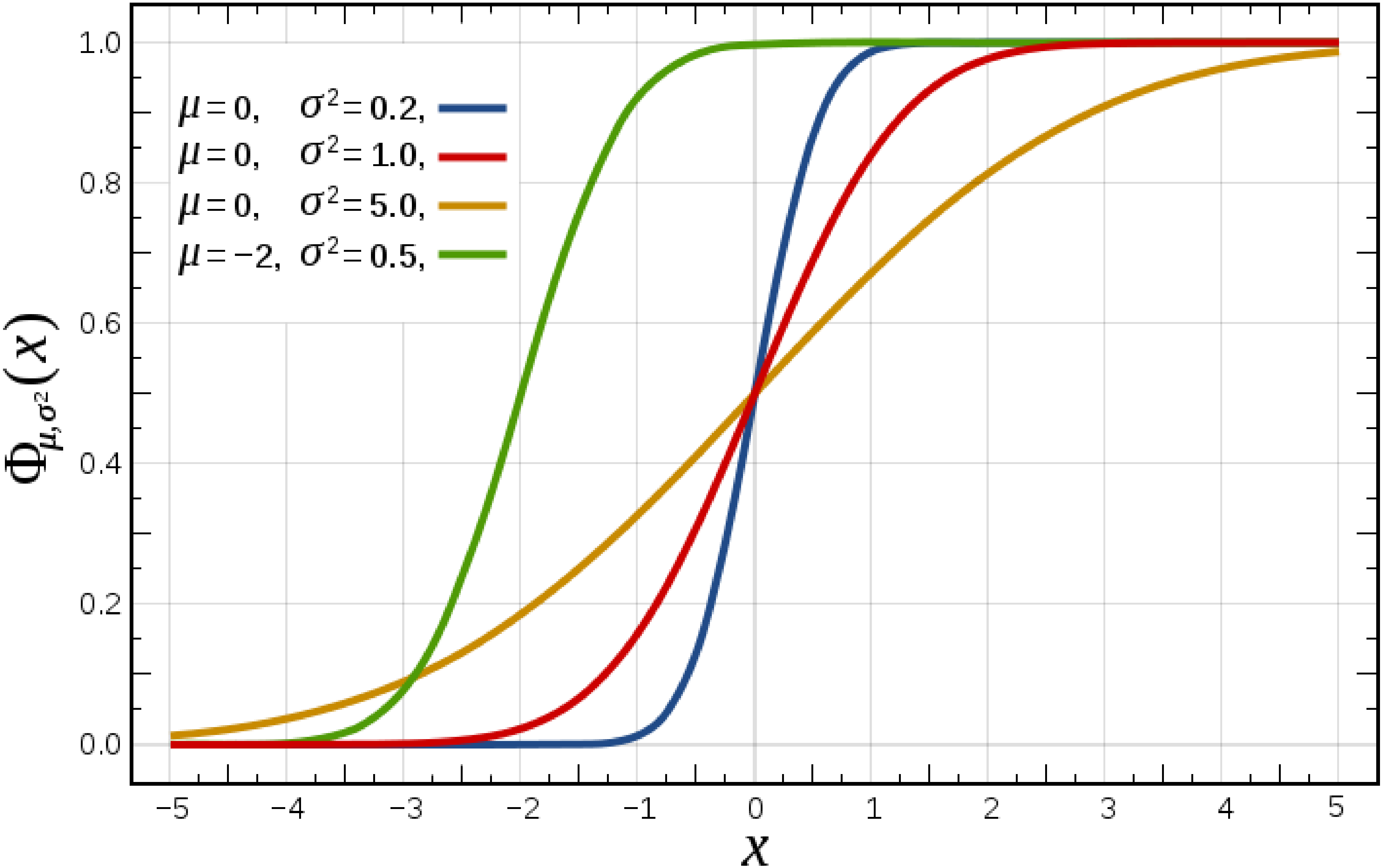
Fonte: Bailey (2016:774-783). Nota: Qualquer livro de estatística básica terá uma seção sobre distribuição normal.

Função densidade de probabilidade para distribuições normais



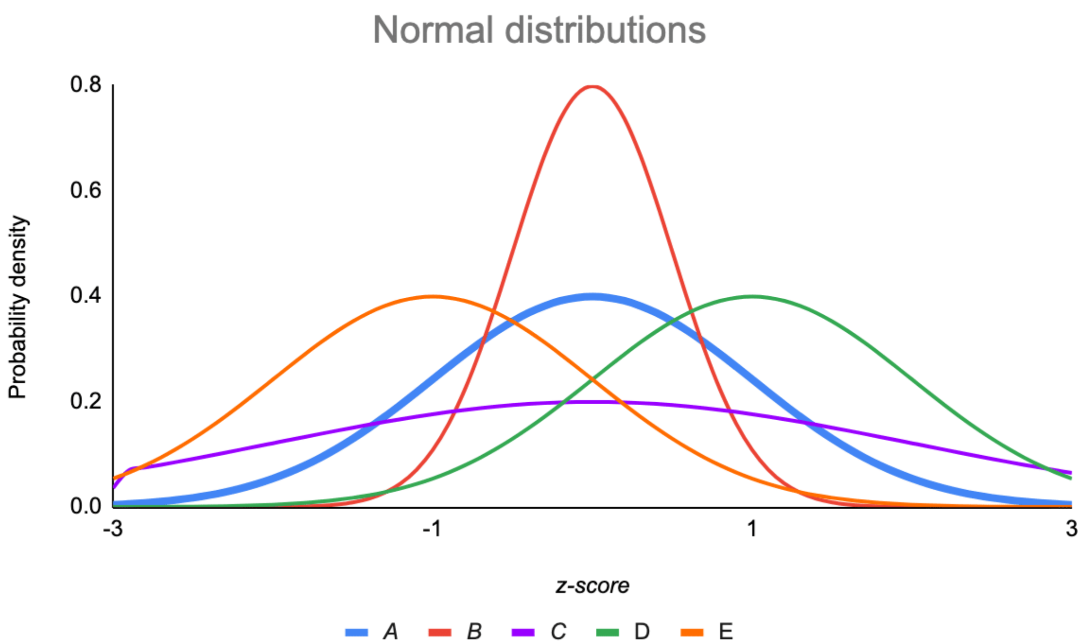
Fonte: https://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg

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:

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

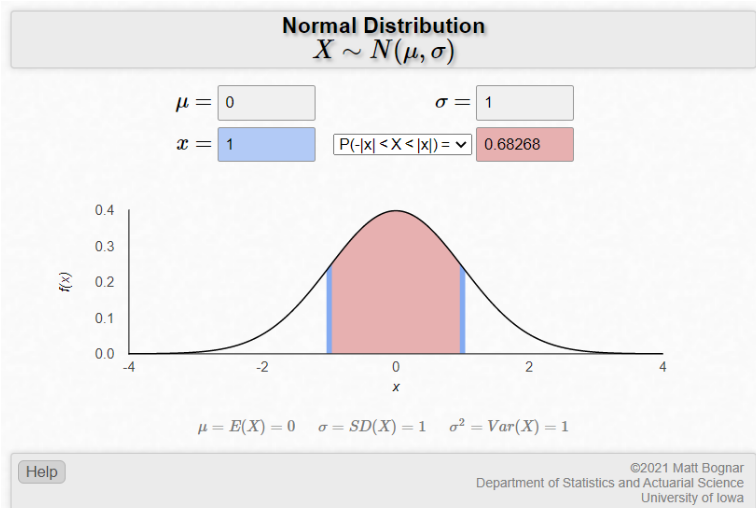
em que:
x = qualquer valor da distribuição normal original (aquela que se quer padronizar)
μ = média da distribuição normal original
σ = desvio padrão da distribuição normal original

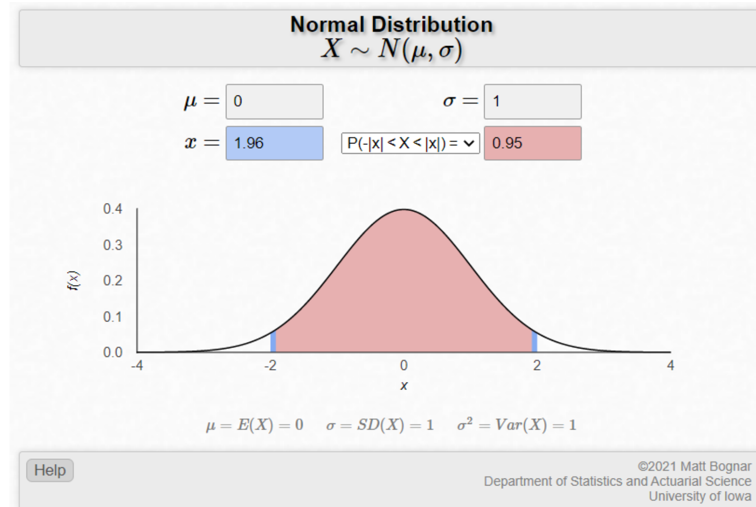
- Um **z-score** (valor de Z para um dado valor de X) positivo significa que x é maior que a média de X
  - z-score > 0 ↔ x > μ
  - z-score < 0 ↔ x < μ
  - z-score = 0 ↔ x = μ

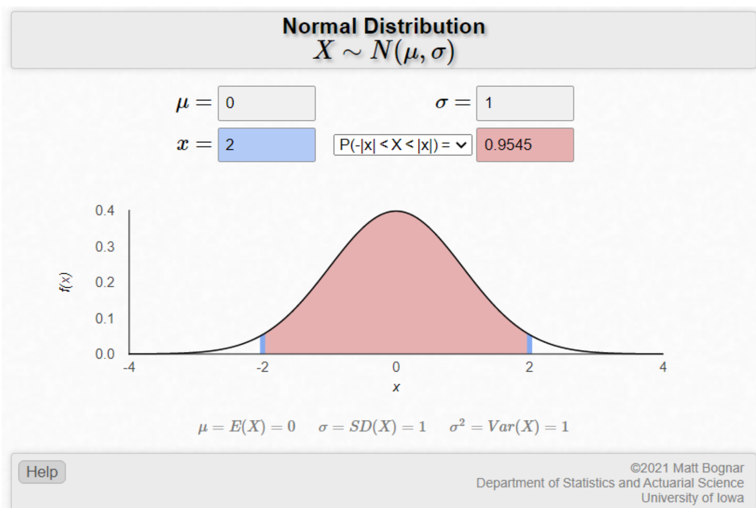
- Entre outras vantagens, **converter** uma distribuição normal **em uma distribuição normal padrão permite**:
  - Comparar pontuações em diferentes distribuições com diferentes médias e desvios padrão
  - Encontre a probabilidade de observações em uma distribuição caírem acima ou abaixo de um determinado valor

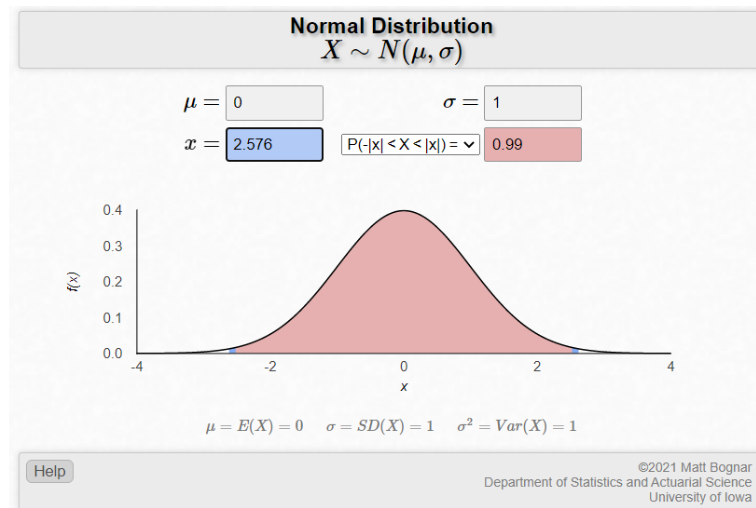
Probabilidade em intervalos de uma distribuição normal padrão

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 $\beta_1$hats se aproxima de uma normal

- Para **amostras grandes** (digamos, 100+), **$\beta_1$hats são normalmente distribuídos**
  - Se os erros do modelo (ε) forem normalmente distribuídos, então $\beta_1$hats serão normalmente distribuídos independentemente do tamanho da amostra
- Portanto, enquanto normal, a **distribuição amostral (“teórica”) de $\beta_1$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 $\beta_1$hat

- **$\beta_1$hats são aleatórios**; conforme a sorte, podemos obter uma estimativa distante de β_1

- Além da **centralidade**, interessa considerar a **dispersão** da distribuição "teórica" de $\beta_1$hats, para avaliarmos quão provável é que estimemos um coeficiente próximo do parâmetro populacional

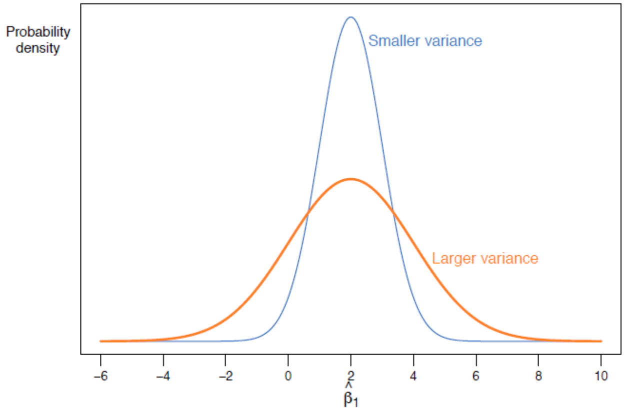


FIGURE 3.6: Two Distributions with Different Variances of $\hat{\beta}_1$

Fonte: Bailey (2016: 94).

– Se a condição de **exogeneidade** estiver atendida, a **média dos $\beta_1$hats é β_1**

– O que podemos dizer sobre a **variância de $\beta_1$hat**?

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

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

Erro padrão de $\beta_1$hat, o **se($\beta_1$hat)** = Raiz quadrada da var($\beta_1$hat)

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

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

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}{N - j} = \frac{\sum_{i=1}^{N}\hat{\epsilon}_i^2}{N - j}$$

$\hat{\sigma}^2$

Variância da regressão

- Mede quão bem o modelo explica a variação de Y
- Seu cálculo baseia-se nos resíduos
- j = número de parâmetros estimados (incluindo o intercepto)
- É também uma estimativa da variância de ε
- **Intuição:** média do quadrado da distância entre valores observados e previstos de Y

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 amostral de X

- 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

$$var(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})^2$$

Em amostras pequenas, divida o somatório por 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.

Massa em libras (pounds)

Ilustração:

Comando para estimar regressão em R

```
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
```

Vide aula 05

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

$n = 13$

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

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

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

$\to 0,3621867$

$0,1311792$

$= 45,81 * 0,3621867$

$= 16,5917727$

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

Não será cobrado

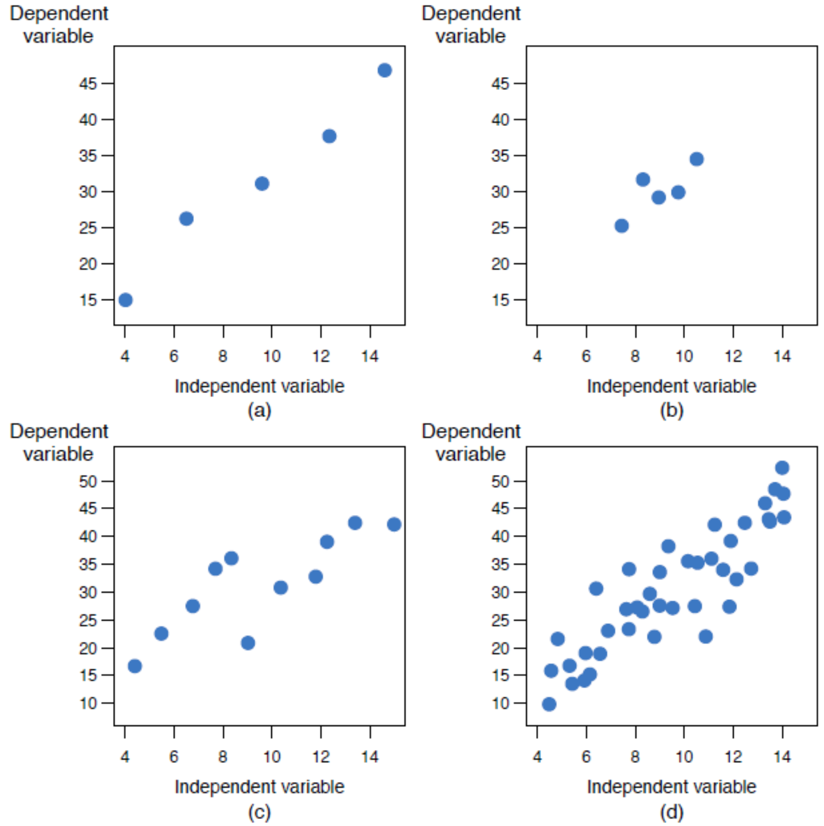


FIGURE 3.7: Four Scatterplots

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_1

Conteúdo não será cobrado

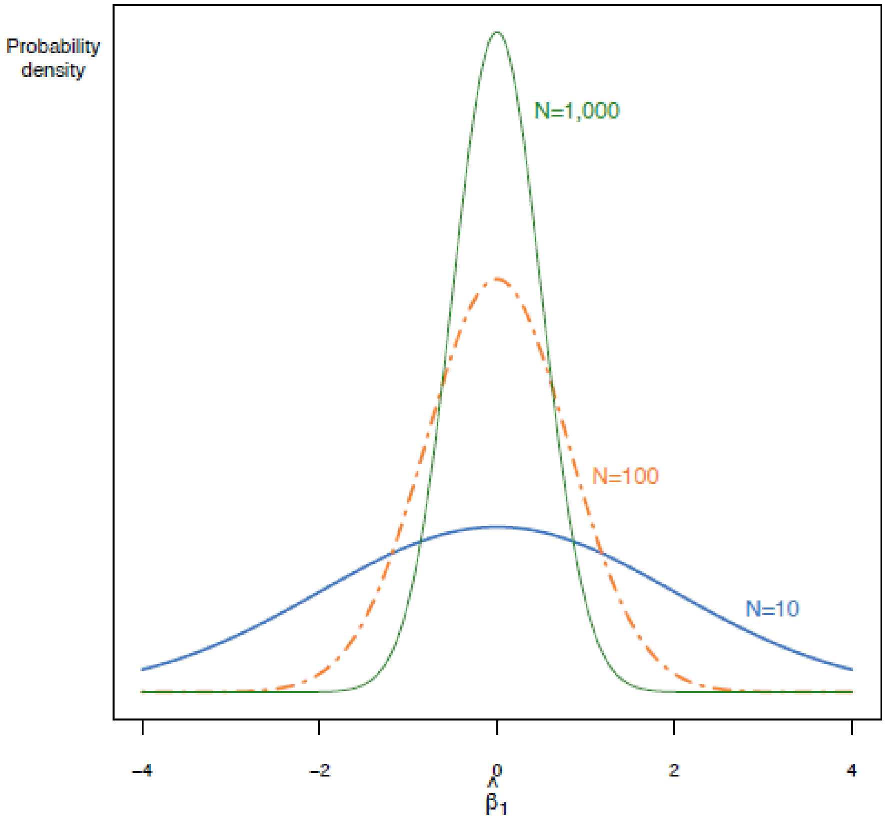


FIGURE 3.8: Distributions of $\hat{\beta}_1$ for different sample sizes

- 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_1

Conteúdo não será cobrado

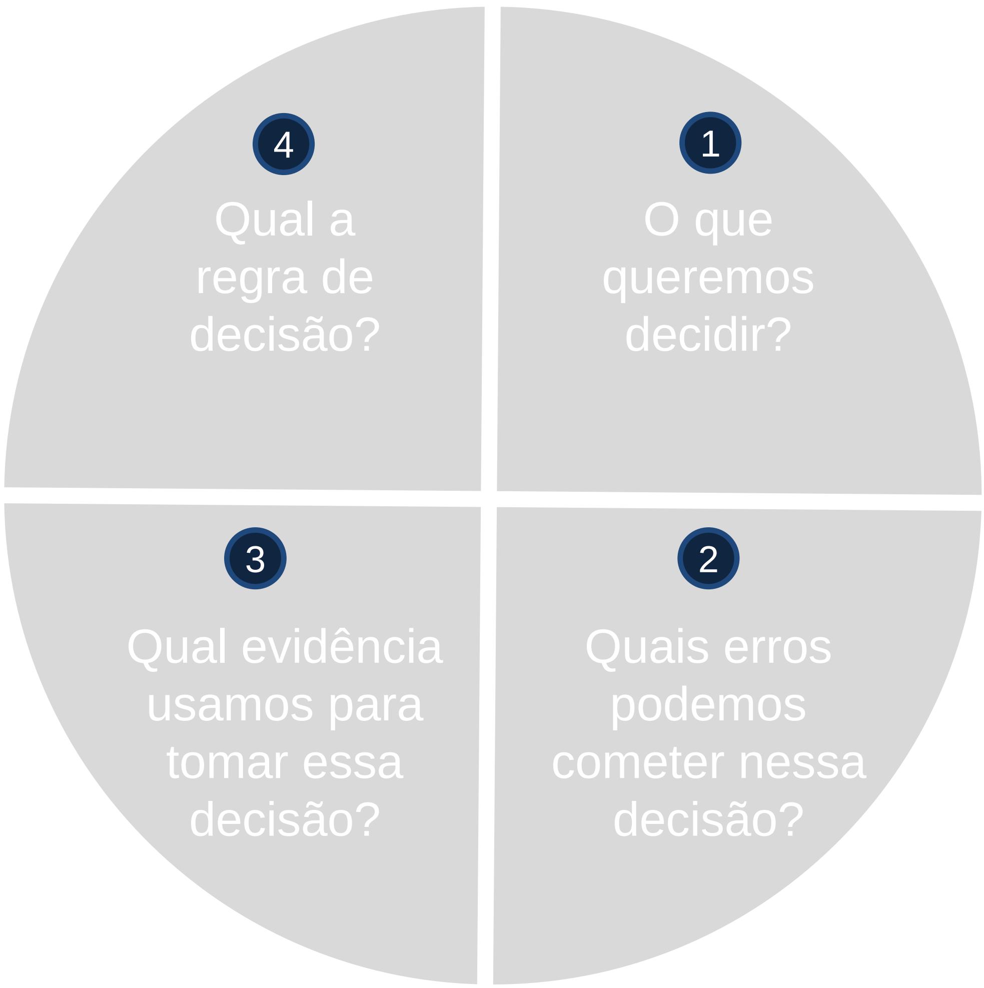
Os livros didáticos às vezes não dão a devida atenção à ***distinção entre ausência de viés [unbiasedness] e consistência****, mas a diferença pode ser importante na prática.* ***Ausência de viés*** *significa que o estimador tem uma distribuição amostral* ***centrada no parâmetro de interesse em uma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

Conclusões sobre efeito de X em Y requerem tomada de decisão sobre o significado da evidência apurada

Teste de hipótese é rotina para essa tomada de decisão



DCP098
Fundamentos para Avaliação Quantitativa de Políticas Públicas

Precisão das estimativas
Teste de hipótese

Aulas 11-12
09 e 11 de maio de 2022

Ana Paula Karruz