

# Monitoria – 08/10/2019

---

MONITOR: LEANDRO ANAZAWA

PPGE – ECONOMIA APLICADA

# Pontos a serem abordados

---

- Alguns pontos importantes da primeira parte da matéria (30 min).
- Apresentação sobre IDEB (30 min).
- Conversa aberta sobre o desafio 1.

# Links

---

- Seguem alguns links importantes para a implementação computacional:
- UCLA – Stata e R: <https://stats.idre.ucla.edu/other/dae/>
- Introduction to Econometrics with R: <https://www.econometrics-with-r.org/>

# Probit e Logit

---

# Funções de transformação

---

- Probit e logit são apenas funções transformadoras que limitam a função  $F(Y)$  para valores entre 0 e 1.

$$F(Y) = \Phi^{-1}(Y) \quad \textit{probit}$$

$$F(Y) = \log \left[ \frac{Y}{(1 - Y)} \right] \quad \textit{logit}$$

## Resultados por MV

---

- Estimaco dos parmetros por mxima verossimilhana.
- Essas estimativas obtidas no tem interpretao direta de probabilidade.
  
- No probit: a variao de X em 1 unidade leva a uma mudana de  $\beta$  unidades no z-valor de Y.
- No logit: a variao de 1 unidade de X (varivel independente) influencia em  $\beta$  unidades o log das chances (odds) de acontecer Y (varivel dependente).

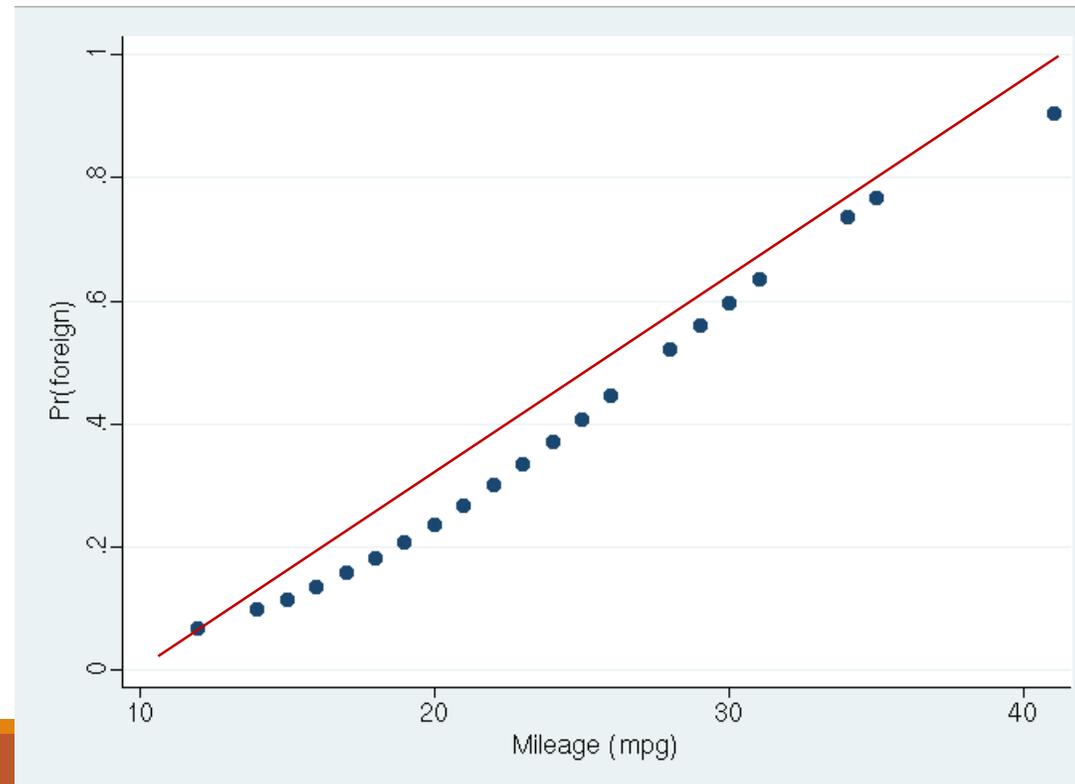
# Efeitos marginais

---

- Para obter o impacto em termos de probabilidade, calculamos os efeitos marginais.
- Importante! O logit e o probit são funções transformadoras limitadas entre 0 e 1. Entretanto, isso não significa que os seus efeitos marginais também estão limitados nesse intervalo!

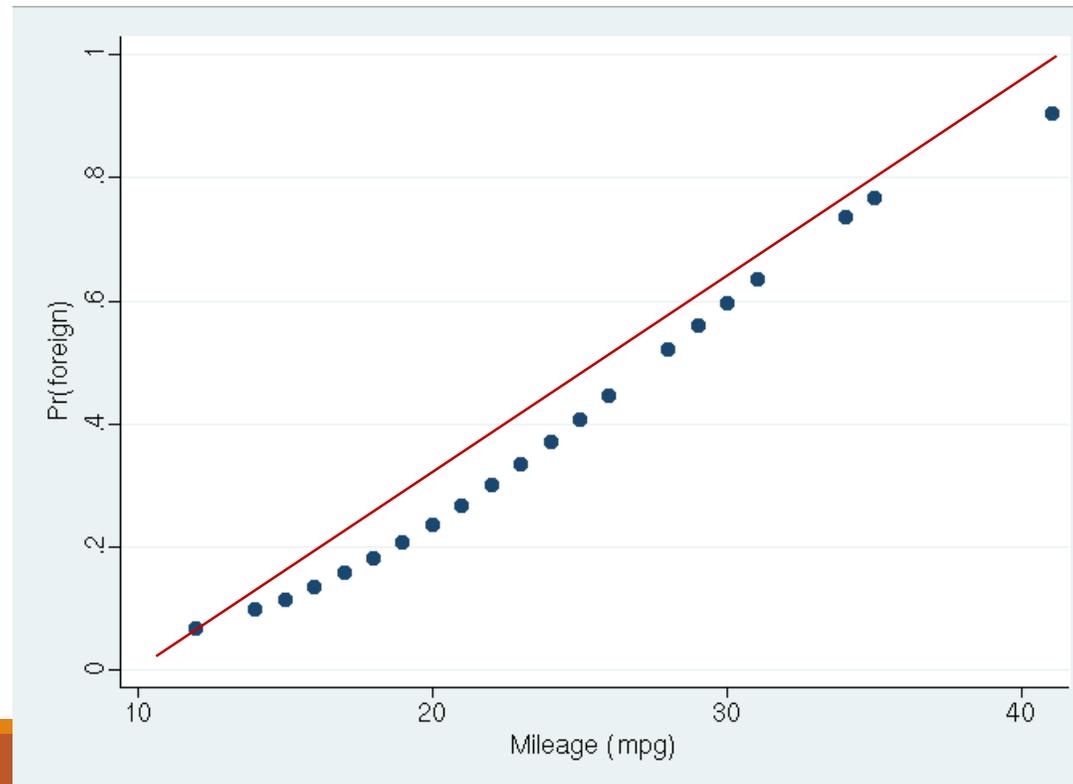
# Efeitos marginais

- No gráfico a seguir temos um modelo probit de regressão de “foreign” (Y) contra “mpg” (X). Os efeitos marginais (derivadas ao longo da curva de pontos) são semelhantes à inclinação da reta de 45 graus (linha vermelha).



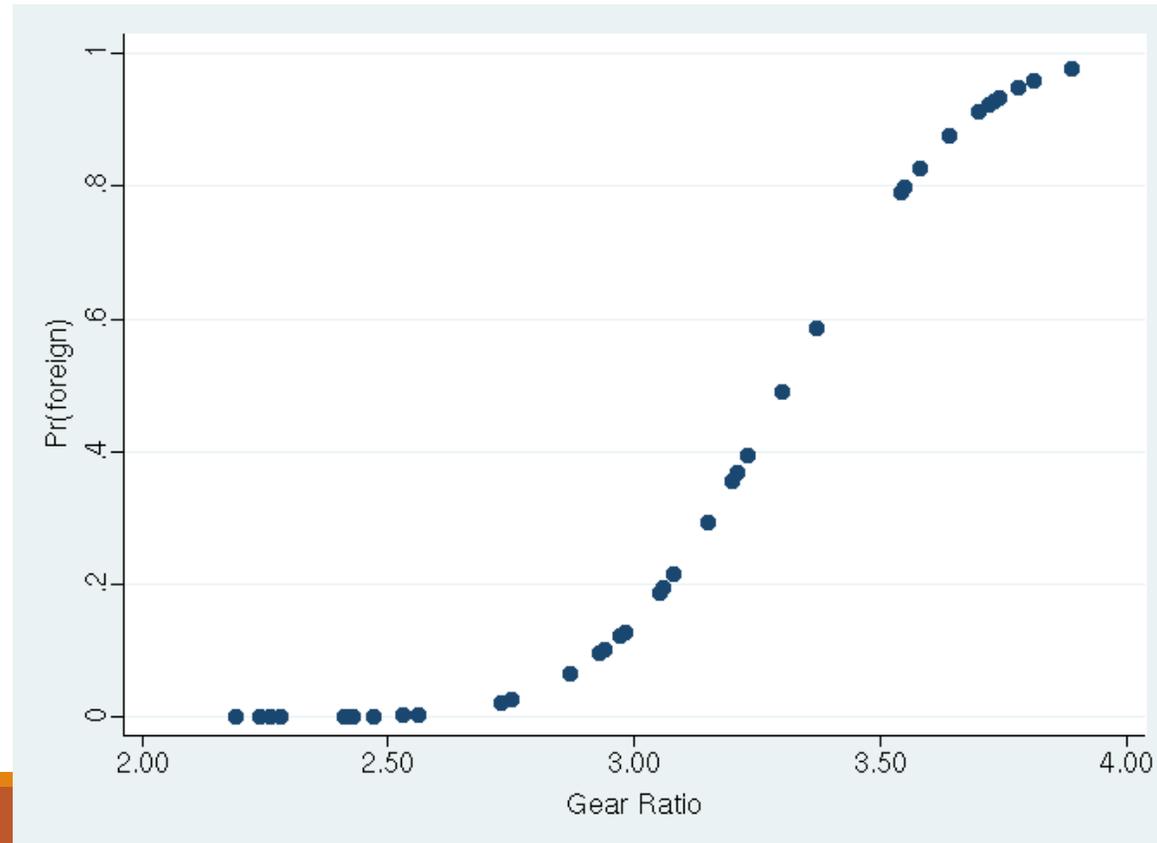
# Efeitos marginais

- Nesse caso, os efeitos marginais provavelmente estarão, aproximadamente, dentro do intervalo de 0 a 1.



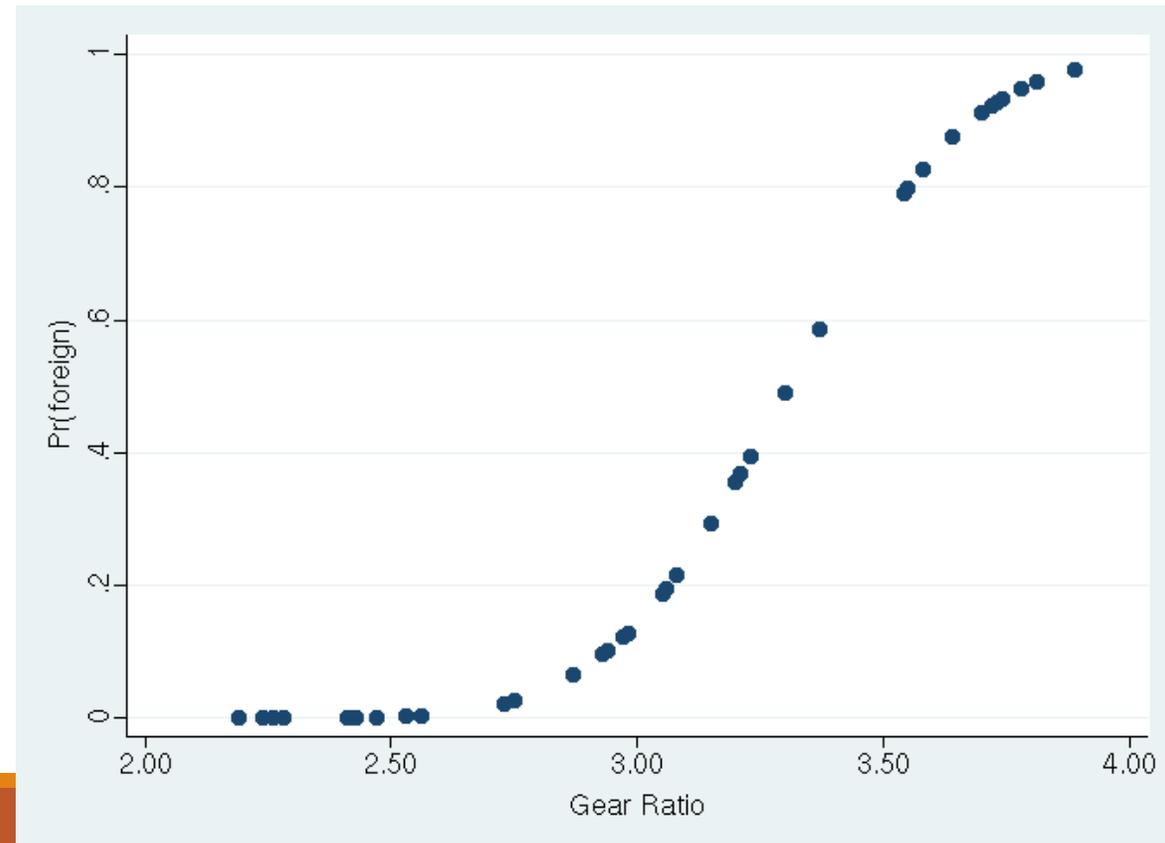
# Efeitos marginais

- E nesse exemplo, será que os efeitos marginais estão limitados entre 0 e 1?



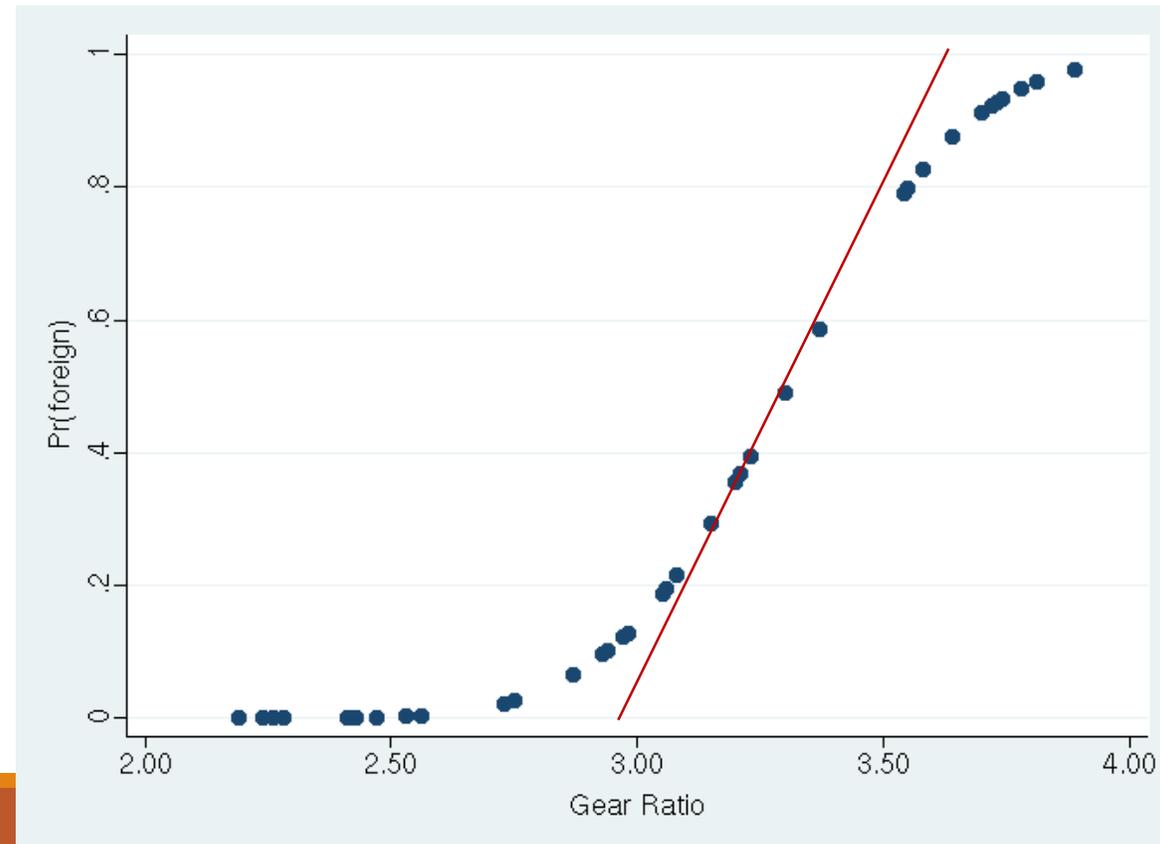
# Efeitos marginais

- A resposta é não!



# Efeitos marginais

- No ponto de “gear ratio” igual a 3,3, temos que o efeito marginal é igual a 1,38!



# Multinomial ordenado e não ordenado

---

## *Logit multinomial*

---

- Qual a diferença das versões multinomiais (não ordenada e ordenada) do logit?
- Em termos de intuição, é só que os resultados são apresentados sempre em relação à categoria base.
  - Mas os efeitos marginais sempre têm a mesma interpretação do probit/logit binários!

# Heckit

---

# Equação de interesse

---

- Suponha a equação de salários (**EQUAÇÃO DE INTERESSE**):

$$W_i = \mathbf{X}_i \boldsymbol{\beta} + \varepsilon_i \quad , \quad \begin{array}{l} \varepsilon \sim N(0, \sigma_\varepsilon^2) \\ \mathbf{X} \text{ independente de } \varepsilon \end{array}$$

- $W_i$  só é observado para quem trabalha. Então, precisamos levar em consideração a decisão de entrar no mercado de trabalho dos indivíduos (auto seleção)!

## Equação de seleção

---

- Suponha uma equação de variável latente que é a “decisão de entrar no mercado de trabalho”

$$E_i^* = \mathbf{Z}_i \boldsymbol{\gamma} + u_i \quad , \quad \begin{array}{l} u_i \sim N(0, \sigma_u^2) \\ Z \text{ independente de } u \end{array}$$

# Equação de seleção

---

- Suponha uma equação de variável latente que é a “decisão de entrar no mercado de trabalho”

$$E_i^* = \mathbf{Z}_i \boldsymbol{\gamma} + u_i \quad , \quad \begin{array}{l} u_i \sim N(0, \sigma_u^2) \\ Z \text{ independente de } u \end{array}$$

- Temos que  $\mathbf{Z}$  é um vetor de características individuais e produtivas.
- Assim, podemos representar essa variável latente pela seguinte função indicadora (**EQUAÇÃO DE SELEÇÃO**):

$$E_i = \mathbf{1}(\mathbf{Z}_i \boldsymbol{\gamma} + u_i > 0)$$

# Equação de seleção

- Suponha uma equação de variável latente que é a “decisão de entrar no mercado de trabalho”

$$E_i^* = \mathbf{Z}_i \boldsymbol{\gamma} + u_i \quad , \quad \begin{array}{l} u_i \sim N(0, \sigma_u^2) \\ Z \text{ independente de } u \end{array}$$

- Temos que  $\mathbf{Z}$  é um vetor de características individuais e produtivas.
- Assim, podemos representar essa variável latente pela seguinte função indicadora (**EQUAÇÃO DE SELEÇÃO**):

$$E_i = \mathbf{1}(\mathbf{Z}_i \boldsymbol{\gamma} + u_i > 0)$$

Função  
indicadora



## Juntando as duas equações

---

$$W_i = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i \quad , \quad \varepsilon \sim N(0, \sigma_\varepsilon^2)$$

- Tomando o valor esperado dessa equação, temos:

$$\begin{aligned} E(W_i | \mathbf{X}_i, E_i = 1) &= E(W_i | \mathbf{X}_i, \mathbf{Z}_i, u_i) = \mathbf{X}_i\boldsymbol{\beta} + E(\varepsilon_i | \mathbf{X}_i, \mathbf{Z}_i, u_i) \\ E(W_i | \mathbf{X}_i, \mathbf{Z}_i, u_i) &= \mathbf{X}_i\boldsymbol{\beta} + E(\varepsilon_i | \mathbf{Z}_i, u_i) \\ E(W_i | \mathbf{X}_i, \mathbf{Z}_i, u_i) &= \mathbf{X}_i\boldsymbol{\beta} + E(\varepsilon_i | u_i > \mathbf{Z}_i\boldsymbol{\gamma}) \end{aligned}$$

- A correção proposta por Heckman é a seguinte:

$$E(\varepsilon_i | u_i > \mathbf{Z}_i\boldsymbol{\gamma}) = \rho_{\varepsilon u} \sigma_\varepsilon \lambda_i(-\mathbf{Z}_i\boldsymbol{\gamma})$$

- $\rho_{\varepsilon u}$ : correlação entre os termos de erro.
- $\lambda_i(-\mathbf{Z}_i\boldsymbol{\gamma})$ : razão inversa de Mills.

## Na prática

- 
- Como fazer essa correção na prática? [VER DO-FILE “AULA 05 – DO FILE.DO” A PARTIR DA LINHA DE COMANDO 38].

- **1º passo)** Estimar o seguinte modelo de regressão probit (**EQUAÇÃO DE SELEÇÃO**):

$$Prob(E = 1) = \Phi(\mathbf{Z}_i\boldsymbol{\gamma})$$

- E obter a estimativa da razão inversa de Mills,  $\hat{\lambda}_i$ .

# Na prática

---

- Como fazer essa correção na prática?

- **2º passo)** Estimar por MQO a **EQUAÇÃO DE INTERESSE**, incluindo como variável independente a estimativa da razão inversa de Mills:

$$W_i = \mathbf{X}_i\boldsymbol{\beta} + \beta_\lambda \hat{\lambda}_i + \varepsilon_i$$

- Pronto! Agora você tem a equação de interesse, controlando pela possível auto seleção presente nos dados.
- $\beta_\lambda$  mede a covariância entre os dois resíduos ( $\varepsilon_i$  e  $u_i$ ) (Söderbom, 2011, pág. 12).
- Se  $\hat{\beta}_\lambda$  for estatisticamente significativo podemos dizer que existe um processo de auto seleção.

## Na prática

---

- É importante não utilizar todas as variáveis independentes da equação de seleção na equação de interesse!
- Isso porque a sua inclusão leva a estimações imprecisas pelo modelo de Heckman (ver Wooldridge, 2010, pág 565).
- Essa imprecisão decorre da multicolinearidade gerada pela inclusão da razão inversa de Mills na equação.

## Na prática

---

- Cuidado! O primeiro estágio de Heckman foi construído a partir de um modelo *probit*. Se você precisar utilizar outro tipo de modelo (ex: multinomial *probit/logit*), acesse: <https://core.ac.uk/download/pdf/43021578.pdf>

# IDEB – Índice de Desenvolvimento da Educação Básica

---

# História

---

- Avaliação do sistema educacional
  - Início com o Sistema de Avaliação da Educação Básica (SAEB) em 1990.
  - O SAEB consiste de testes de Leitura e de Matemática dos alunos de 5º ano, 9º ano do ensino fundamental e 3º ano do ensino médio.
  - Mas é em caráter amostral (representatividade ao nível de estados e país).
  - Aplicado a cada dois anos para alunos da rede pública e privada de educação.
- 1º problema enfrentado:
  - Como comparar os resultados do SAEB de diferentes anos e de diferentes séries?
  - Solução: utilizar a Teoria de Resposta ao Item (TRI) a partir do SAEB de 1995.
  - Resultados comparáveis entre anos e acumulativo entre séries.

# História

---

- Outros problemas enfrentados:
  - O SAEB gera resultados que variam de 0 a 400+ pontos.
  - Mas esses resultados não são intuitivos.
  - Além disso, existe o problema de elevado grau de repetência escolar no Brasil.
- No ano de 2007, existe um movimento no Ministério da Educação (MEC) para implementar o Plano de Desenvolvimento da Educação (PDE).
- O PDE focou nas questões de qualidade educacional, monitoramento e responsabilidade.

# História

---

- A Prova Brasil e o Índice de Desenvolvimento da Educação Básica (IDEB) surgiram nesse movimento de 2007.
- A Prova Brasil é uma extensão do SAEB com o objetivo de aplicar os testes de Leitura e Matemática para todos os alunos de escolas públicas.
- O IDEB tinha como objetivo monitorar e permitir uma fácil comparação entre escolas.
  - O IDEB varia de 0 a 10, sendo que 10 indica a melhor avaliação educacional da escola

# Cálculo

---

- O que é o IDEB?
- O IDEB é uma medida sintética da qualidade de ensino nas escolas brasileiras.
- Esse indicador padronizou os resultados do SAEB para uma escala de 0 a 10;
- Além disso, ponderou esses resultados pelas taxas de aprovação da escola.
  - O objetivo dessa ponderação era penalizar as escolas que tentavam aumentar os seus resultados através da reprovação escolar.

# Cálculo

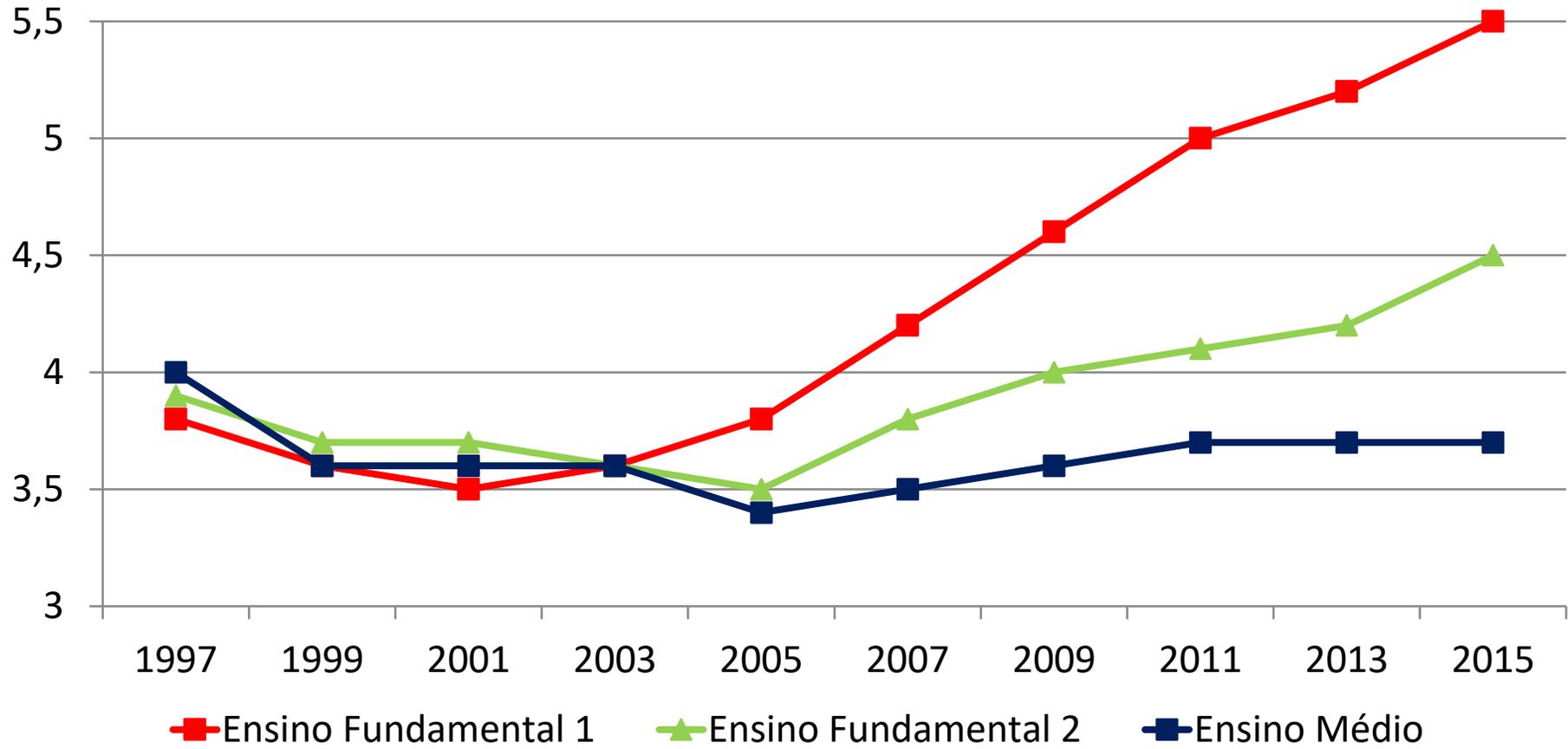
---

$$IDEB_{jt} = N_{jt} * P_{jt}$$

- $N_{jt}$ : média da proficiência em Leitura e Matemática (padronizada para um intervalo de 0 a 10) dos alunos da unidade j e da edição da Prova Brasil (ou SAEB) no ano t.
- $P_{jt}$ : indicador de rendimento (na escala de 0 a 1) baseado na taxa de aprovação da etapa de ensino dos alunos da unidade j no ano t.
- Verificar cálculos detalhados no texto oficial do IDEB.

# Evolução do IDEB

Fonte:  
Retirado da  
apresentação  
“20 anos de  
SAEB/Prova  
Brasil: Uma  
Avaliação de  
Seus  
Resultados” de  
Reynaldo  
Fernandes –  
USP.



# Evolução

---

- A questão da composição dos alunos é um tema investigado desde o final da década de 1980. O efeito composição no contexto de avaliações educacionais ganha destaque no episódio de 1999 envolvendo o Ministério da Educação e a explicação da queda das notas do SAEB entre 1995 e 1999 (Fernandes e Natezon, 2003).
- As notas do SAEB (componente N do IDEB) caíram entre 1995 e 2001.
- Duas hipóteses:
  - 1) A expansão de matrículas piora a qualidade de ensino (as matrículas em escolas aumentaram mais rápido do que os recursos);
  - 2) Efeito composição: o perfil dos alunos que fazem a prova muda.

# Evolução

---

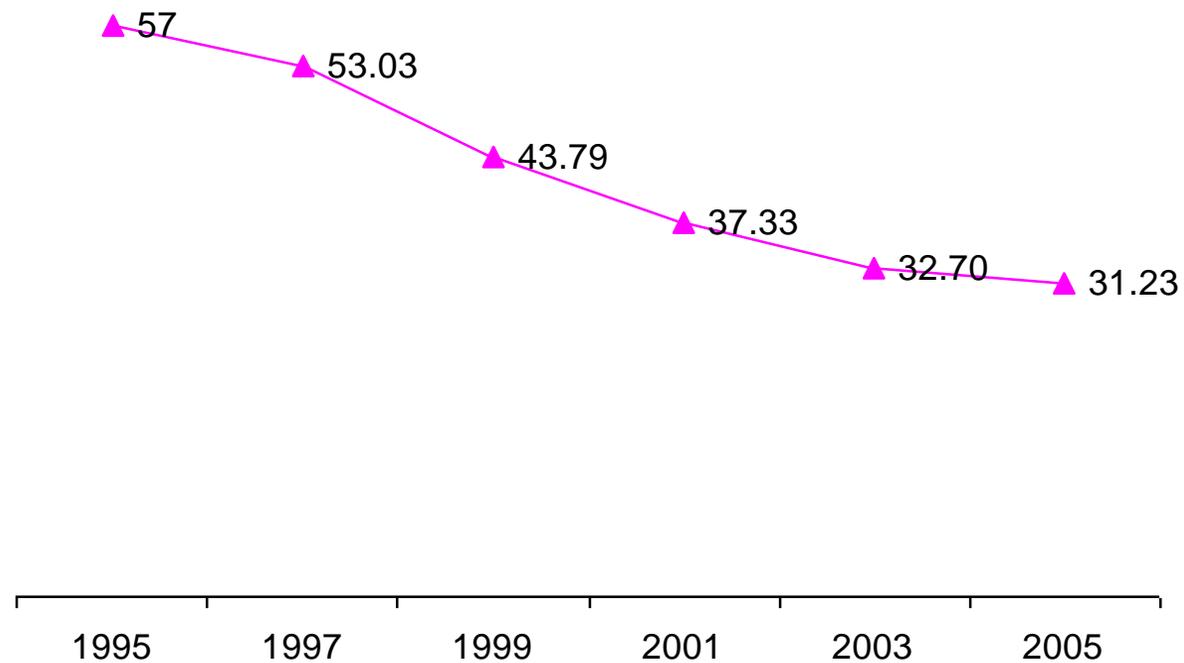
- Existe evidências de que o “Efeito composição” é o predominante.
- Em 1995, 93,9% das crianças de 10 anos de idade já estavam na escola. Em 1999, esse percentual sobe para 97,6%.
- Mas a parcela de crianças atrasada (fora da idade ideal para a série) ou fora da escolas caiu abruptamente a partir de 1995.

# Evolução

---

Fonte: PNAD – IBGE –  
vários anos. Retirado da  
apresentação “20 anos  
de SAEB/Prova Brasil:  
Uma Avaliação de Seus  
Resultados” de  
Reynaldo Fernandes –  
USP.

% de crianças de 10 anos atrasadas ou fora da  
escola



# Evolução

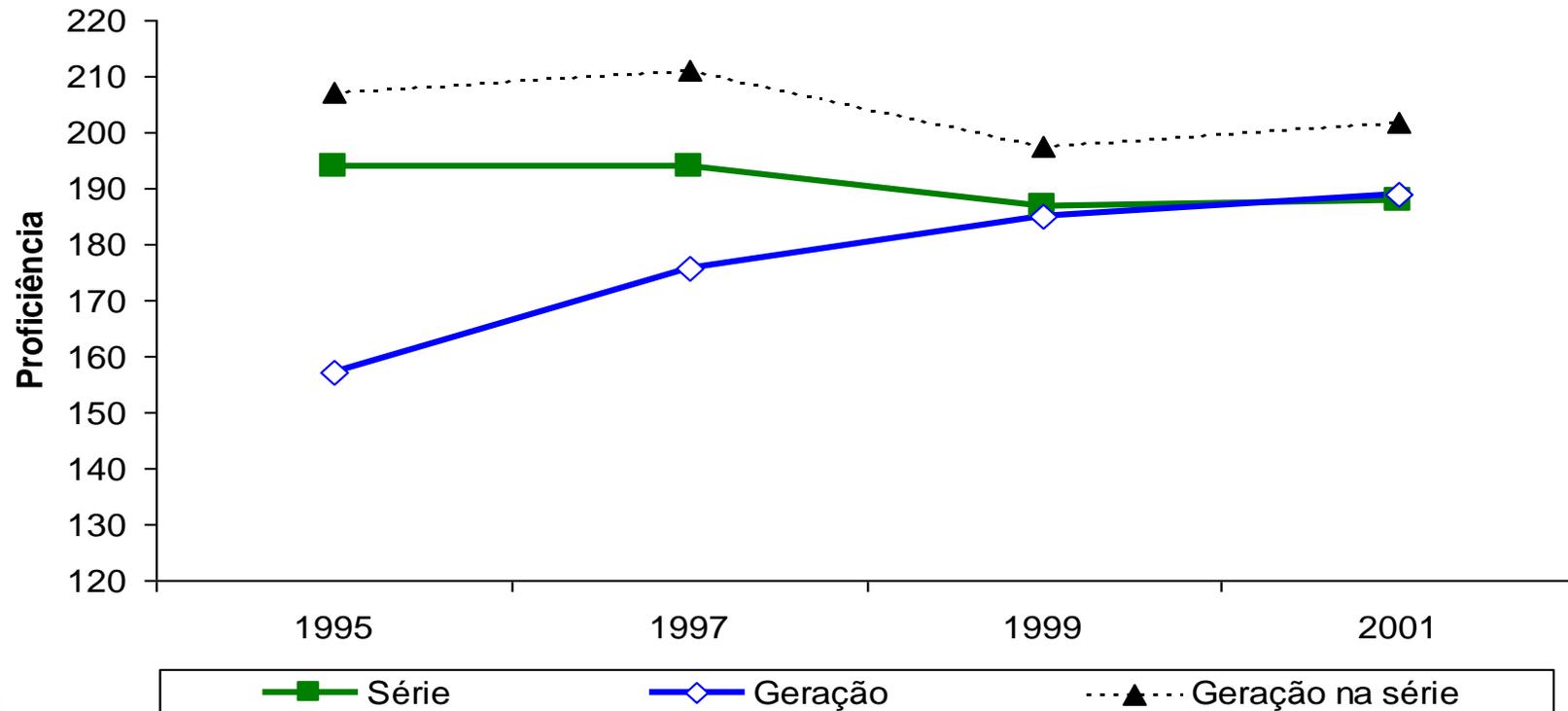
---

- Redução da taxa de reprovação e o efeito onda.
- Efeito onda: redução imediata das notas do SAEB devido ao aumento de alunos atrasados e posterior aumento dessas notas quando o fluxo escolar estiver corrigido.

# Evolução

- Fernandes e Natenzon (2003):

**Figura 1. Mediana do desempenho em Matemática: Série X Geração**  
Região Sudeste 1995-2001  
4a série do Ensino Fundamental como referência



Fonte:  
Retirado da  
apresentação  
“20 anos de  
SAEB/Prova  
Brasil: Uma  
Avaliação de  
Seus  
Resultados” de  
Reynaldo  
Fernandes –  
USP.

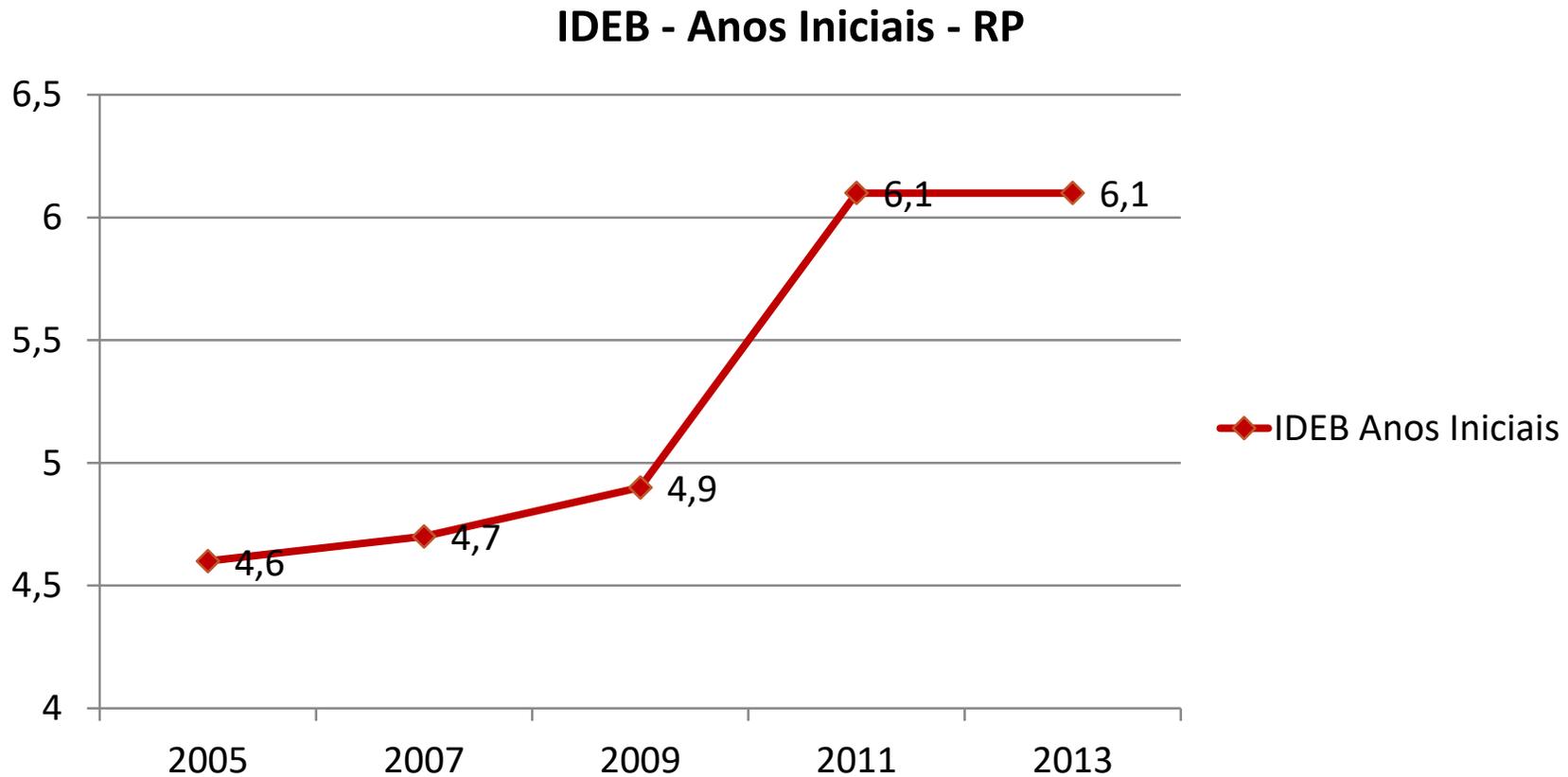
# Evolução

---

- Efeito onda e consequências esperadas.
- A partir de 2007 o IDEB do 5º ano do ensino fundamental aumenta.
  - Supostamente, atingiu-se o “fluxo escolar corrigido” esperado pela tese do “efeito onda”.

# Evolução

Fonte:  
Retirado da  
apresentação  
“20 anos de  
SAEB/Prova  
Brasil: Uma  
Avaliação de  
Seus  
Resultados” de  
Reynaldo  
Fernandes –  
USP.



# Evolução

---

- Efeito onda e consequências esperadas.
- A partir de 2007 o IDEB do 5º ano do ensino fundamental aumenta.
  - Supostamente, atingiu-se o “fluxo escolar corrigido” esperado pela tese do “efeito onda”.

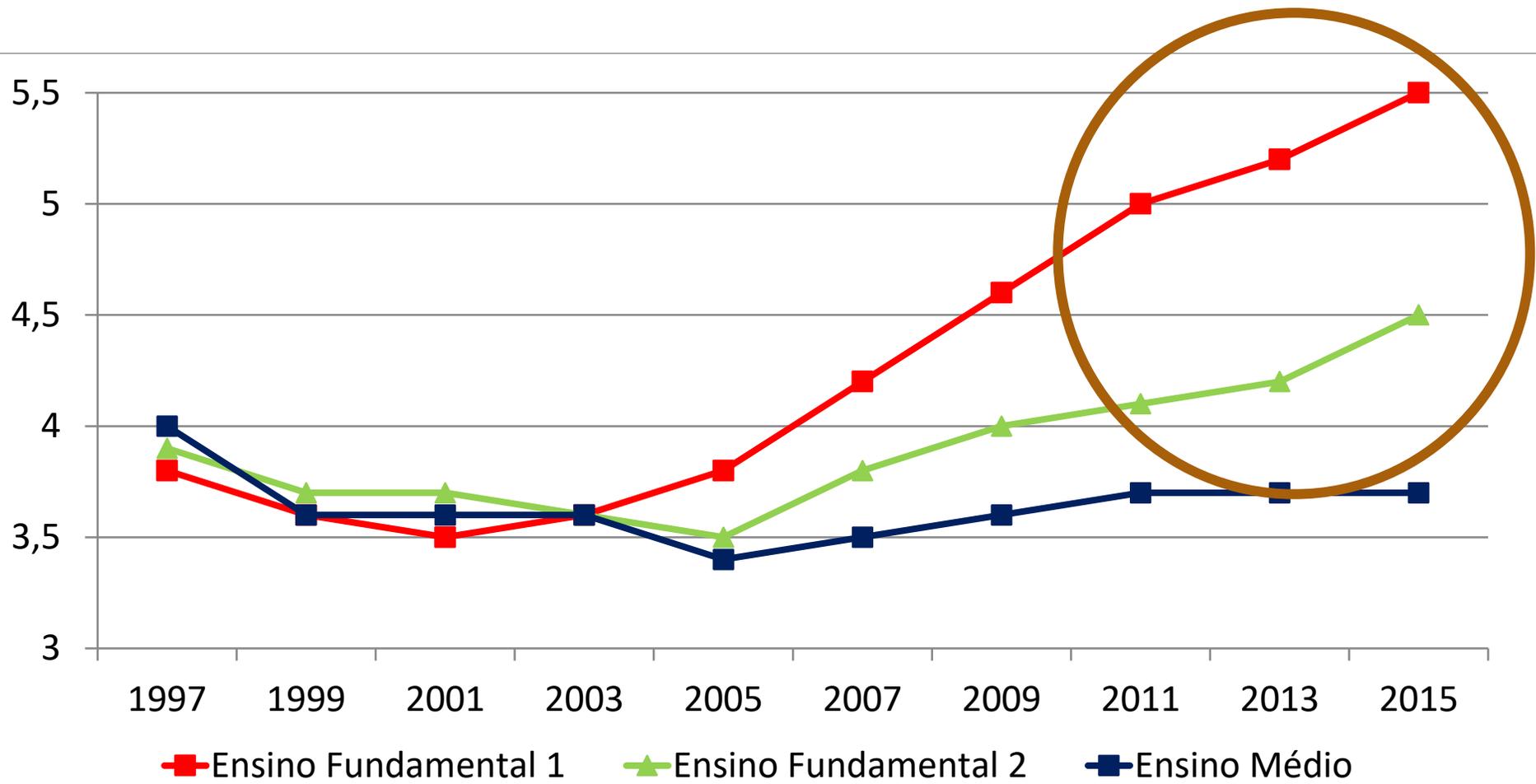
# Evolução

---

- Efeito onda e consequências esperadas.
- A partir de 2007 o IDEB do 5º ano do ensino fundamental aumenta.
  - Supostamente, atingiu-se o “fluxo escolar corrigido” esperado pela tese do “efeito onda”.
- Então, esses alunos do 5º ano em 2007 estariam no 9º ano em 2011.
- Esperava-se então que o IDEB do 9º ano do ensino fundamental chegasse ao patamar do 5º ano a partir de 2011.

# Evolução do IDEB

Fonte:  
Retirado da  
apresentação  
“20 anos de  
SAEB/Prova  
Brasil: Uma  
Avaliação de  
Seus  
Resultados” de  
Reynaldo  
Fernandes –  
USP.



# Evolução

---

- Hipóteses para a inexistência desse catch-up:
  - I) Qualidade de ensino dentro das escolas não melhorou entre 2007 e 2011, o que gerou perdas dos ganhos iniciais de IDEB de 2007.
  - II) Problemas de integração vertical da escala SAEB (comparabilidade cumulativa entre séries escolares).
  - III) Efeito composição.

# Evolução

---

- Hipóteses para a inexistência desse catch-up:
  - I) Qualidade de ensino dentro das escolas não melhorou entre 2007 e 2011, o que gerou perdas dos ganhos iniciais de IDEB de 2007.
  - II) Problemas de integração vertical da escala SAEB (comparabilidade cumulativa entre séries escolares).
  - III) Efeito composição.

**O Desafio 1 se encaixa aqui!**

# Problemas de composição do Desafio 1

---

- Nem todo mundo chega na série correta para a idade;
- Nem todo mundo aparece para fazer a prova;
- Municípios e estados diferem na composição dos alunos que optam pela escola pública.

# Desafio 1

---

- Objetivo: reinventar o IDEB considerando esses problemas de composição.

# Textos sobre o IDEB

---

- [Nota técnica]
- [Oficial] Fernandes (2007). Índice de Desenvolvimento da Educação Básica (IDEB).
- Fernandes e Gremaud (2009). Qualidade da educação: avaliação, indicadores e metas.
- Soares e Xavier (2013). Pressupostos educacionais e estatísticos do IDEB.
- Fernandes e Natenzon (2003). A Evolução Recente do Rendimento Escolar das Crianças Brasileiras: uma reavaliação dos dados do Saeb.

# Bases de dados

---

- Informações sobre o IDEB (por escola, município ou estado) e os seus componentes (N e P) podem ser encontrados no link: <http://portal.inep.gov.br/web/guest/educacao-basica/ideb/resultados>
- Verifiquem a parte “Planilhas do Ideb” desse link

# Bases de dados

---

- As informações sobre características dos alunos (e algumas características das famílias desses alunos) que fizeram a Prova Brasil podem ser obtidas dos microdados da Prova Brasil disponíveis no link: <http://portal.inep.gov.br/microdados>
- Os dados de 2007 a 2011 estão na parte “Prova Brasil”.
- Os dados de 2013 a 2017 estão na parte “Saeb (Aneb/Prova Brasil)”.

# Bases de dados

---

- As informações sobre características das famílias (inclusive das famílias com filhos fora da escola ou fora da idade correta para a série) podem ser encontrada na PNAD no link: [ftp://ftp.ibge.gov.br/Trabalho\\_e\\_Rendimento/Pesquisa\\_Nacional\\_por\\_Amostra\\_de\\_Domicilios\\_anual/microdados/](ftp://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_anual/microdados/)
- Lembre que os dados da PNAD só são representativos ao nível estadual, país ou região metropolitana (ou não metropolitana).