













# Classificação Supervisionada

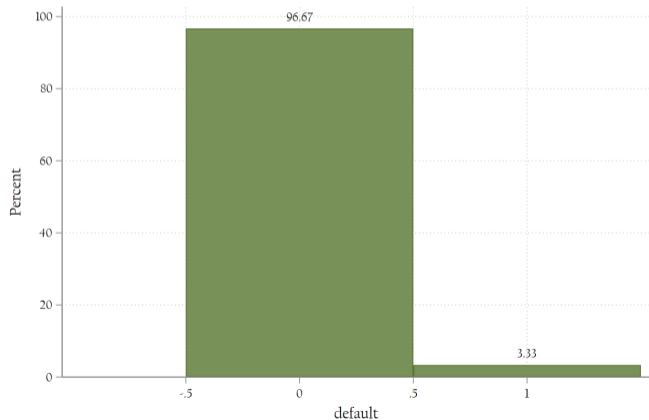
- Lembrando das 03 etapas necessárias:
  - 1 Classificação dos documentos (Definição e Categorização);
  - 2 Treinamento do modelo (Aprendizado do Modelo);
  - 3 Validação (Inferência).



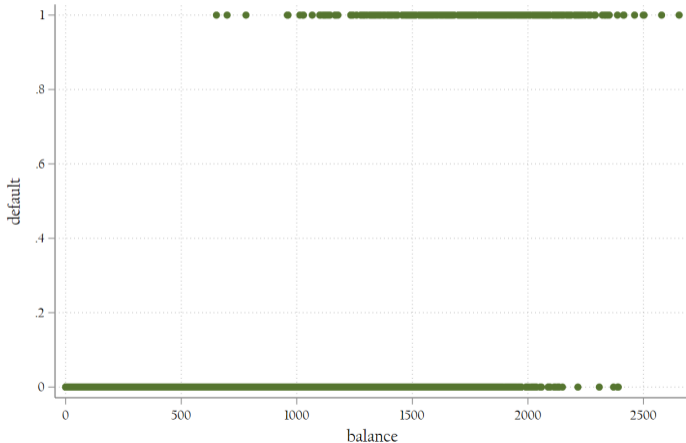


# Caso 1: Variável binária *dummy* com distribuição assimétrica

Default =  $f(\text{Student}, \text{Balance}, \text{Income})$



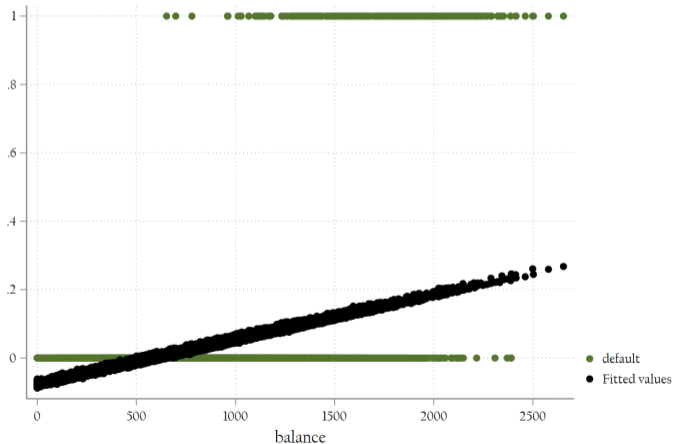
# Caso 1: Variável binária *dummy* com distribuição desbalanceada



# Caso 1: Variável binária *dummy* com distribuição desbalanceada

- População N=10.000
- Population Regression Function using OLS:  
 $Default = -.092 + 0.0001 Balance + 0.000000467 Income$

# Caso 1: Variável binária *dummy* com distribuição desbalanceada



# Training and Test Data Sets

- Training data set (80% da amostra)
- Test data set (20% da amostra)

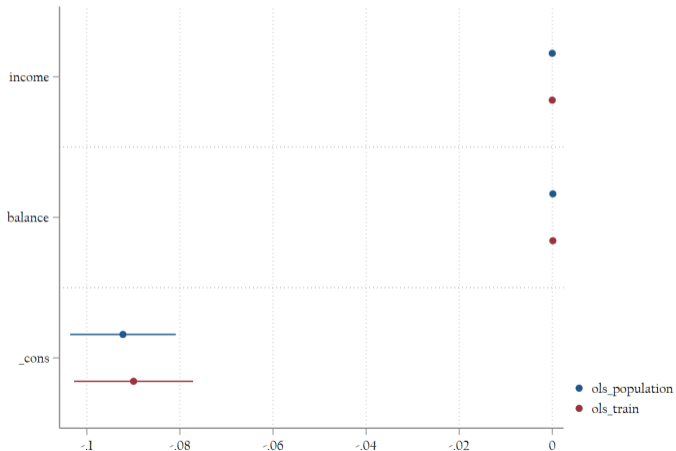
# Caso 1: Variável binária *dummy* com distribuição desbalanceada



# Caso 1: Variável binária *dummy* com distribuição desbalanceada

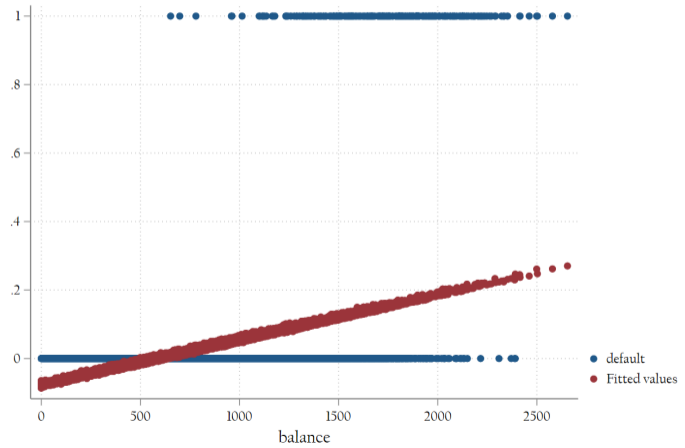
- Training Sample N=8.000
- Training Data Set (Random Sample of 80% of cases) Regression Function using OLS:  $Default = -.090 + .0001 Balance + 0.000000388 Income$

# Training and Population Data Sets





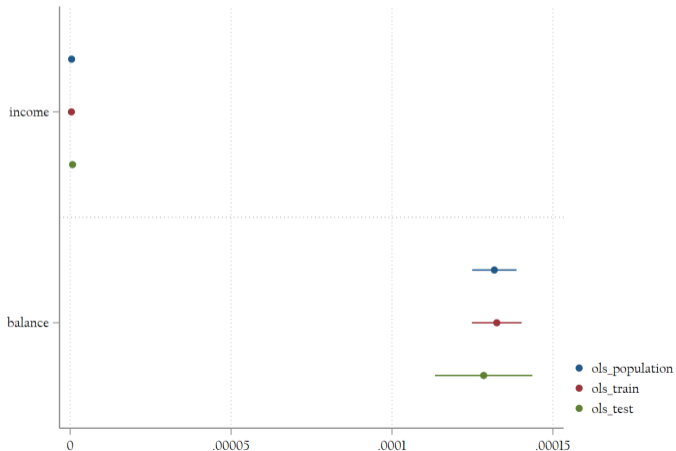
# Training Data Set



# Caso 1: Variável binária *dummy* com distribuição desbalanceada

- Test Sample N=2.000
- Test Data Set (Random Sample of 20% of cases)

# Population, Training and Test Data Sets







# Laboratório 8