

Comparação de métodos de imputação de dados ausentes sob diferentes mecanismos

Aluna: Fernanda Barros

Orientadora: Thaís Paiva

Introdução

- ▷ Dados ausentes em pesquisas;
- ▷ Inferência de informações;
- ▷ Métodos existentes: exclusão de observações com valores ausentes, análise de caso completo (ACC) e imputação múltipla (IM).

Banco de dados

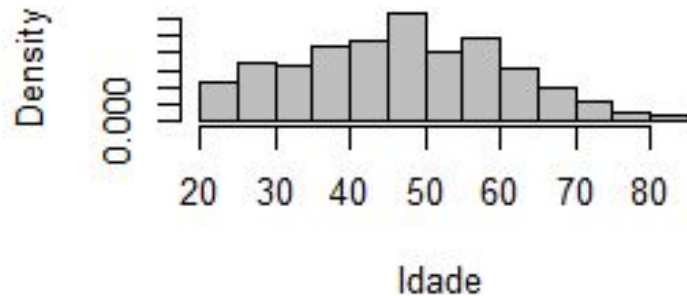
- ▷ Amostra de 500 domicílios retirada da pesquisa de finanças de consumidores (*Survey of Consumer Finances*), realizada em 2004 nos Estados Unidos da América (EUA).
- ▷ Variáveis: renda, gênero, idade, estado civil, anos de escolaridade e etnia.
- ▷ Objetivo: realizar inferência para a renda dado às demais variáveis, sob a presença de dados faltantes utilizando imputação múltipla.

Imputação Múltipla (IM)

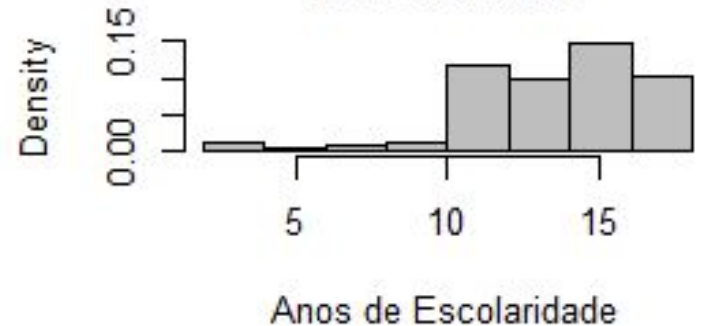
- ▷ Perda completamente aleatória: o motivo pelo qual os dados estão ausentes não está relacionado às variáveis do estudo.
- ▷ Perda aleatória: a razão para um valor estar ausente está relacionada às outras variáveis observadas, mas não está relacionada à variável em que há valores ausentes.
- ▷ Perda não aleatória: o motivo pelo qual os dados estão ausentes está diretamente relacionado aos valores não observados da variável de interesse.

Análise descritiva

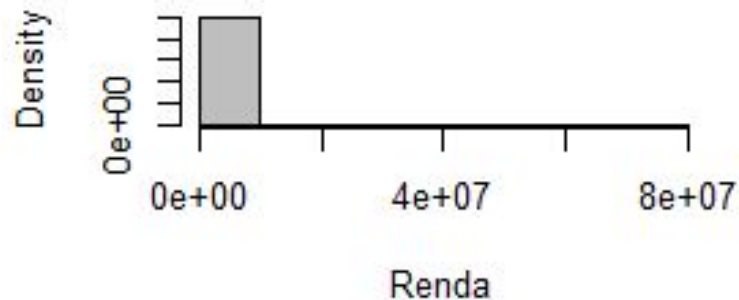
Histograma da Idade



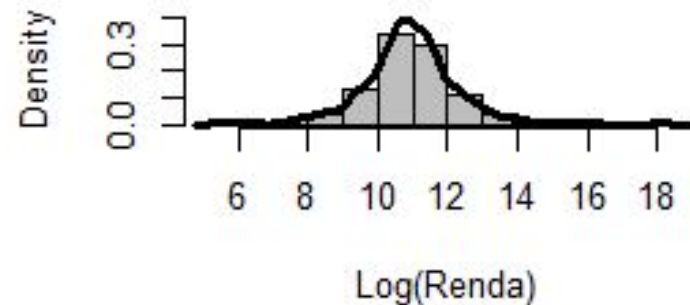
Histograma dos Anos de Escolaridade



Histograma da Renda

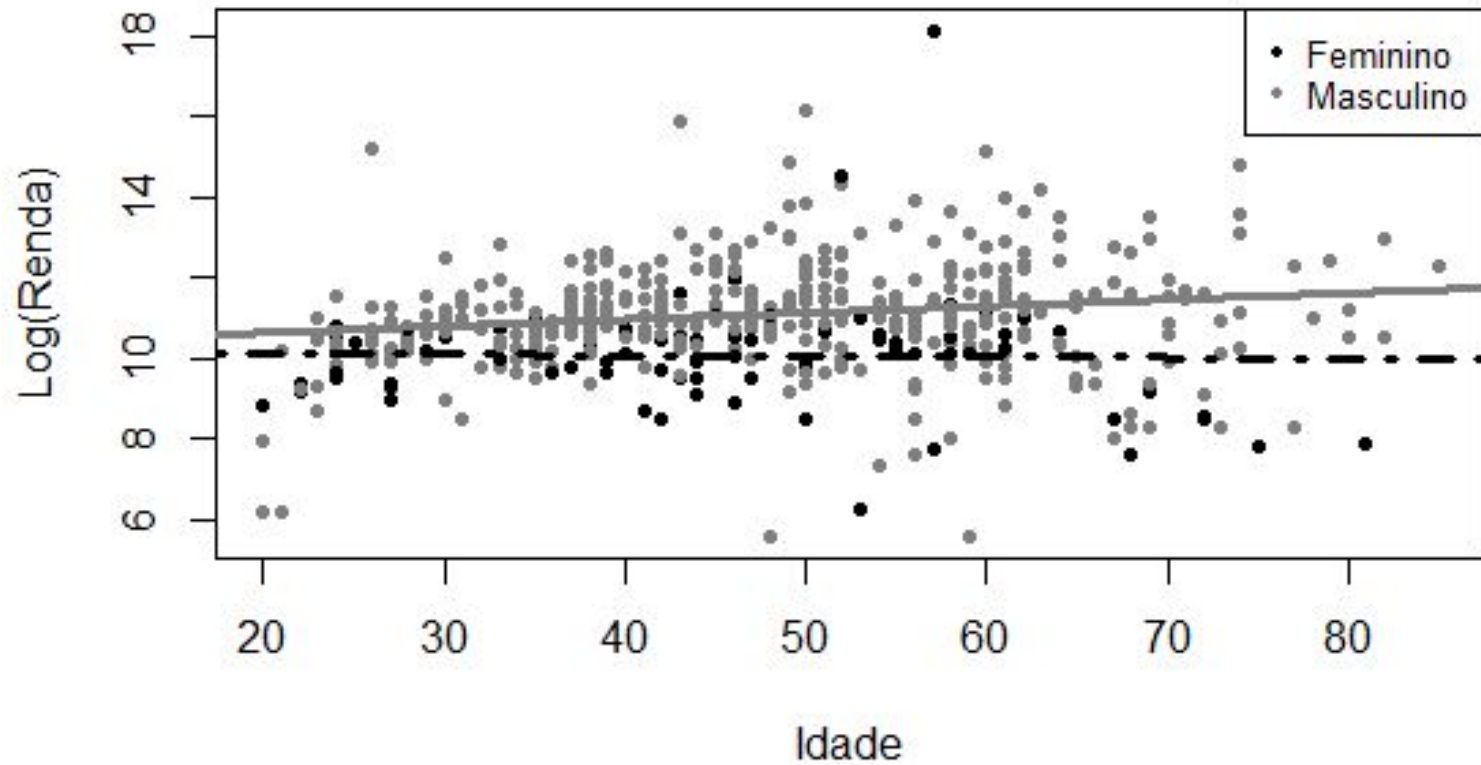


Histograma da Log(Renda)

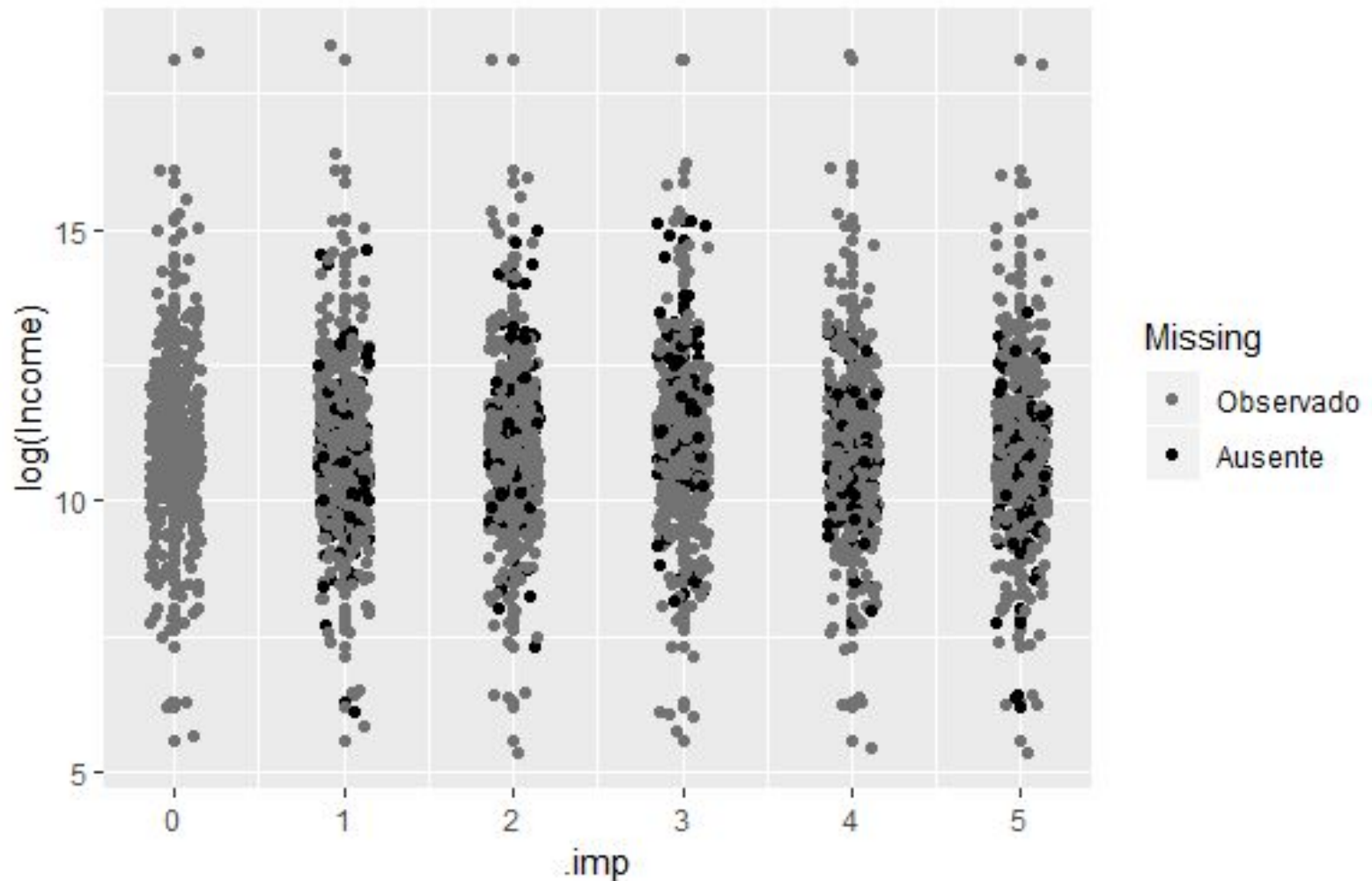


Análise descritiva

Gráfico da Idade e a Log(Renda)

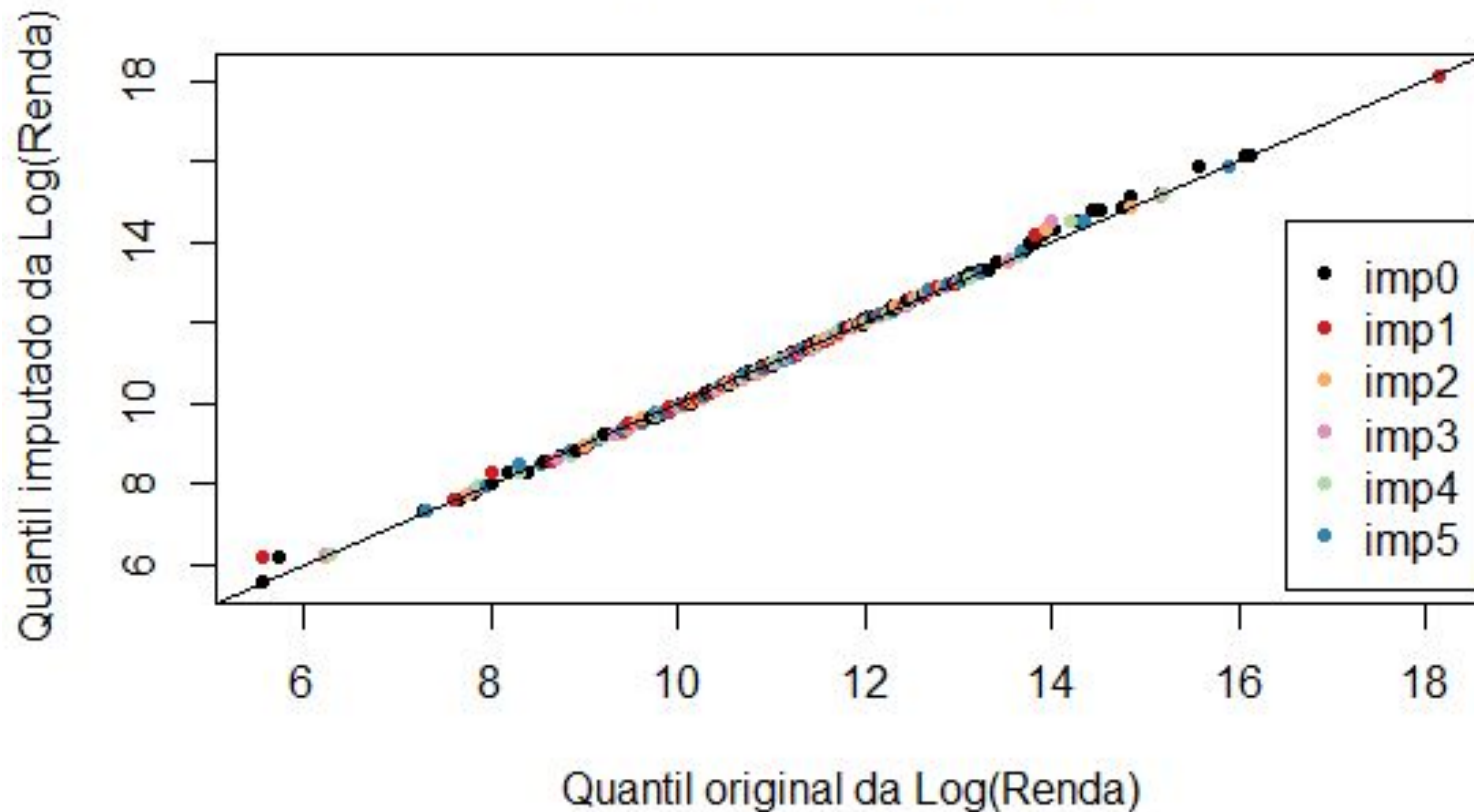


Perda Completamente Aleatória

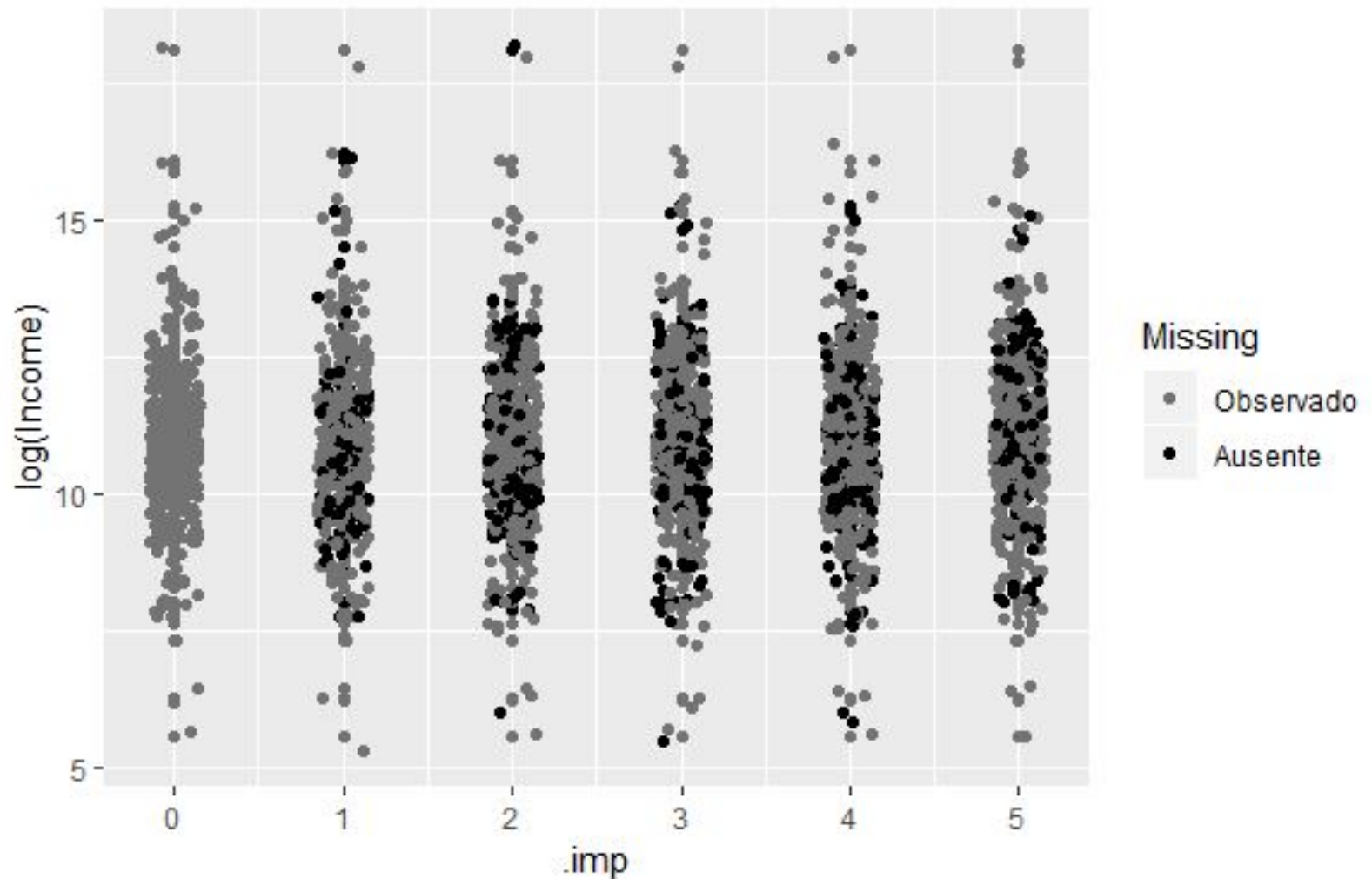


Perda Completamente Aleatória

QQ-plot das imputações

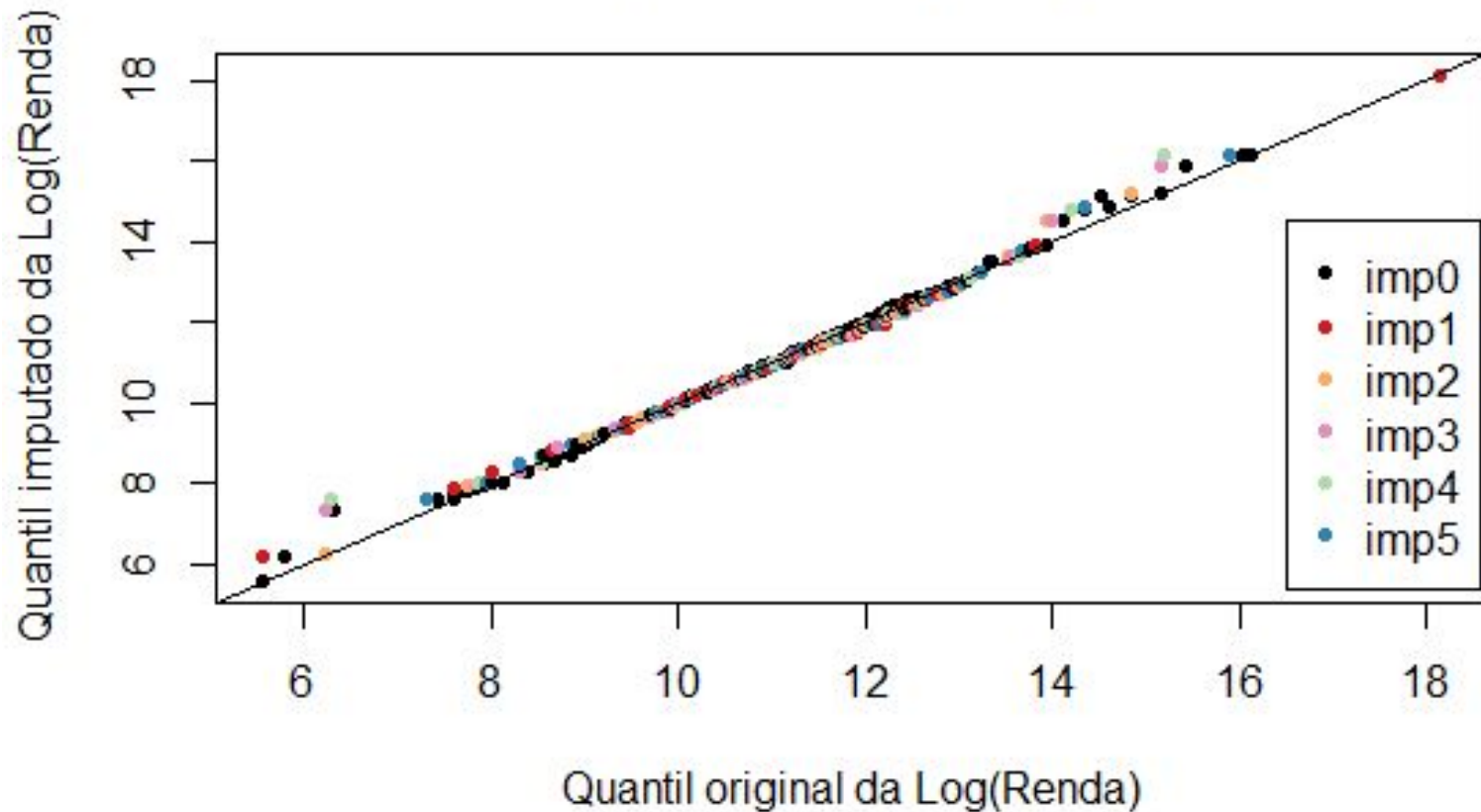


Perda Aleatória - Variável Gênero

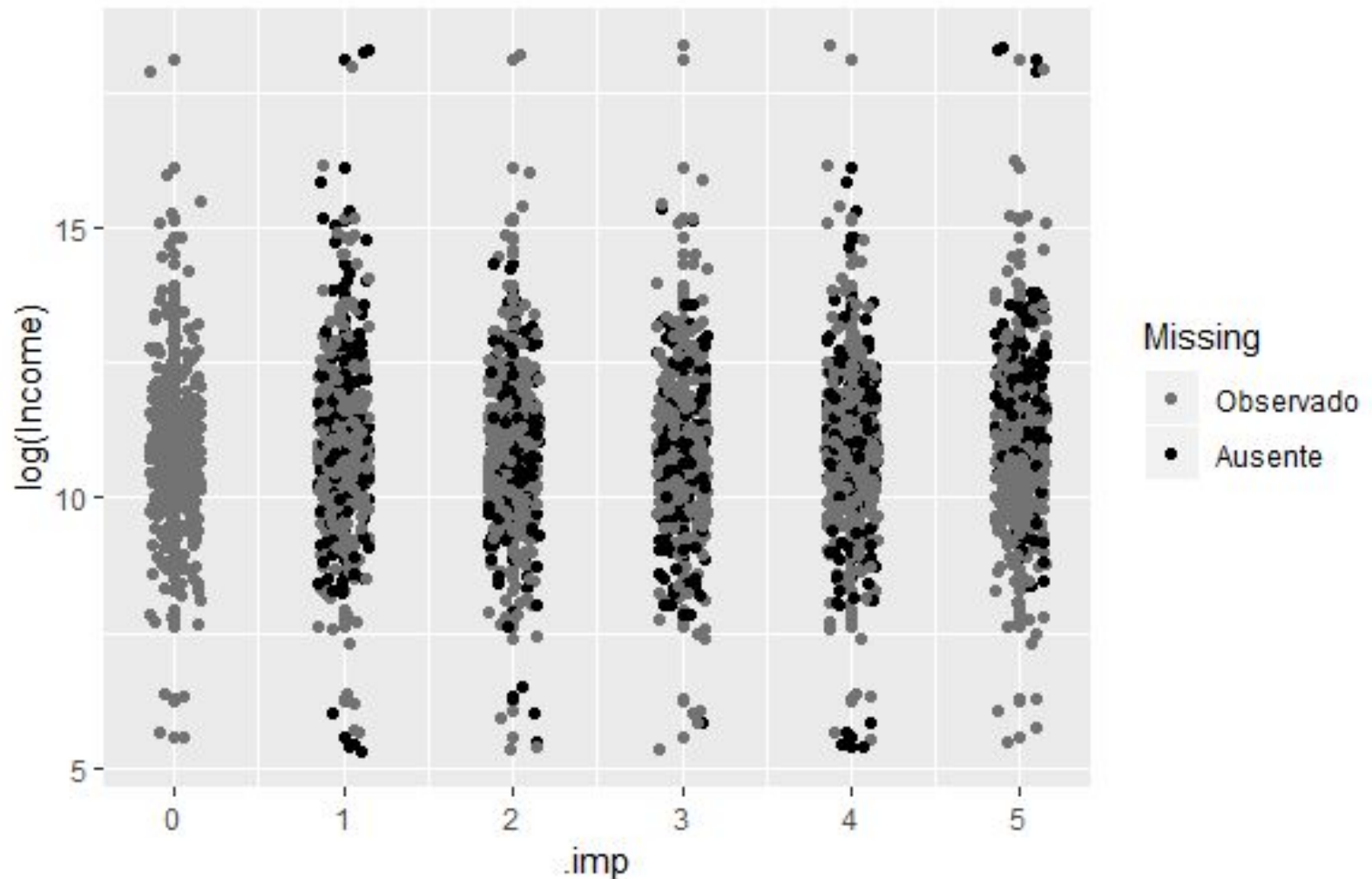


Perda Aleatória - Variável Gênero

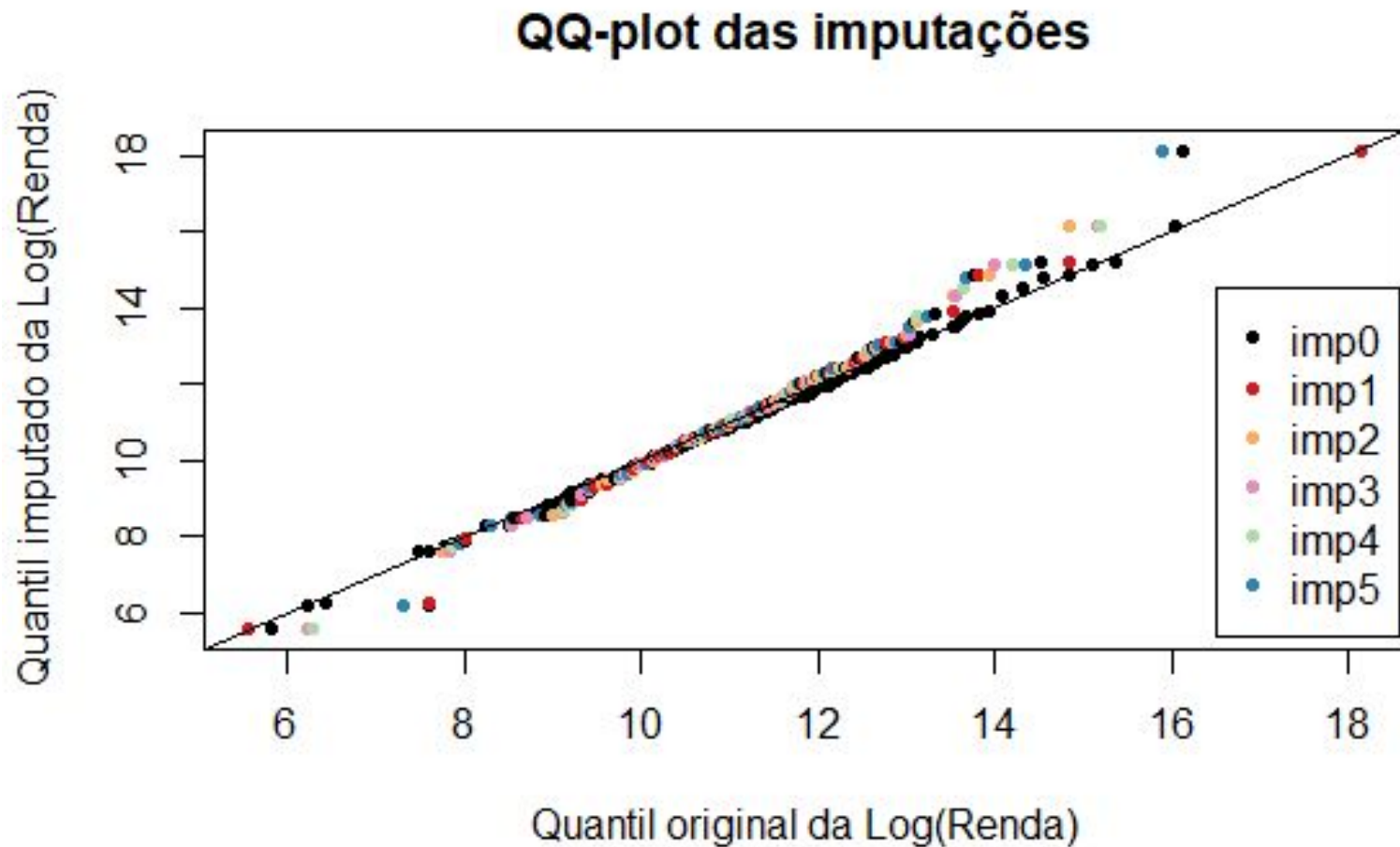
QQ-plot das imputações



Perda Aleatória - Variável Tipo de Ensino



Perda Aleatória - Variável Tipo de Ensino



Perda Não Aleatória

Perda Não Aleatória

Obrigada!