



PUC Minas
Virtual

educação sem distância

Comparação de Agrupamento da Mortalidade Infantil entre indígenas e não-indígenas - 2019

Nahari Terena

Pós-graduação Lato Sensu em
Ciência de Dados e Big Data

Abril, 2023.

Problema
Proposto

```
graph TD; A[Problema Proposto] --> B[Coleta de dados]; A --> C[Processamento de dados]; A --> D[Análise de Componentes Principais]; A --> E[Análise e Exploração de dados]; A --> F[Modelagem]; A --> G[Resultados];
```

The diagram is a flowchart illustrating the stages of a data science project. At the top, a light green box labeled 'Problema Proposto' is connected by a horizontal orange line to six yellow boxes below it. The yellow boxes are labeled 'Coleta de dados', 'Processamento de dados', 'Análise de Componentes Principais', 'Análise e Exploração de dados', 'Modelagem', and 'Resultados'. The background features abstract green and yellow geometric shapes on the right side.

Coleta de
dados

Processamento
de dados

Análise de
Componentes
Principais

Análise e
Exploração de
dados

Modelagem

Resultados

Por quê?

- ODS 3: Assegurar uma vida saudável e promover o bem-estar para todos, em todas as idades

Quem?

- Dados oficiais do Ministério da Saúde divulgados pela Coordenação-Geral de Informações e Análises Epidemiológicas (CGIAE)
 - Sistema de Informação sobre Mortalidade (SIM)
 - Sistema de Informação sobre Nascidos Vivos (Sinasc)

O que?

- Existem padrões nos óbitos infantis?
- Há distinção significativa entre indígenas e não-indígenas?
- Quais as diferenças entre os padrões em cada grupo?

Onde?

- Óbitos de crianças de até 1 ano de vida registrados no Brasil.

Quando?

- Óbitos infantis registrados no período de em 2019.

Coleta de dados

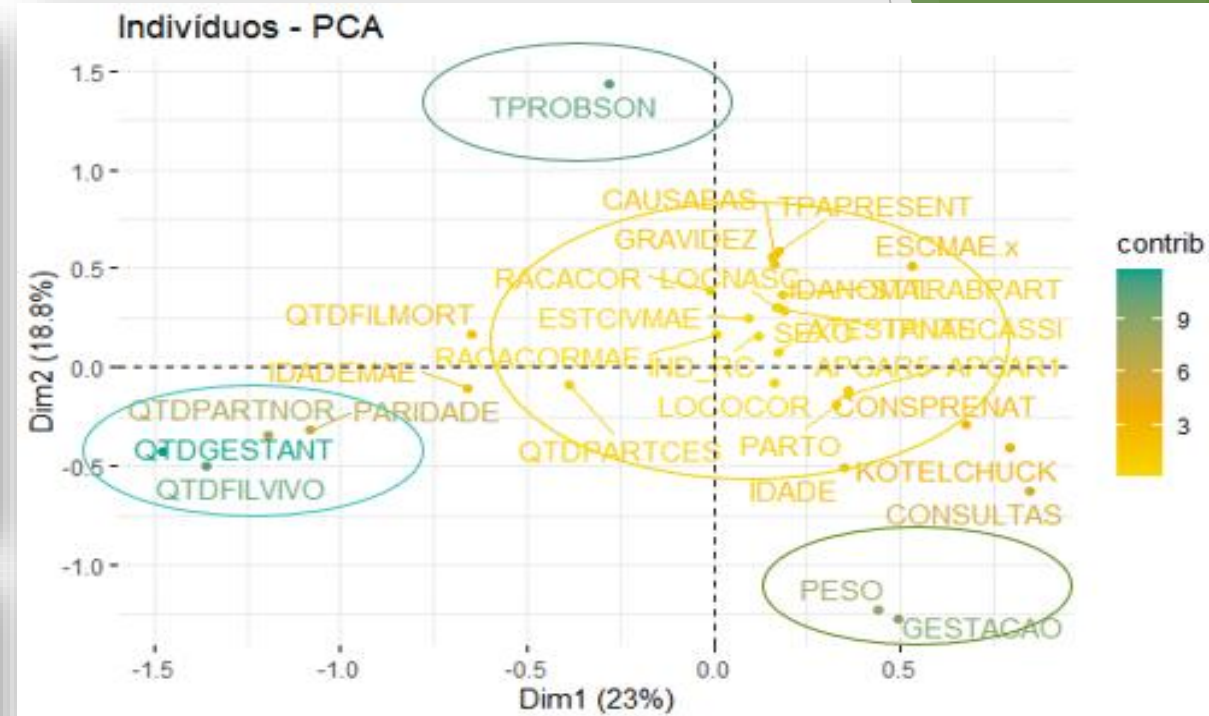
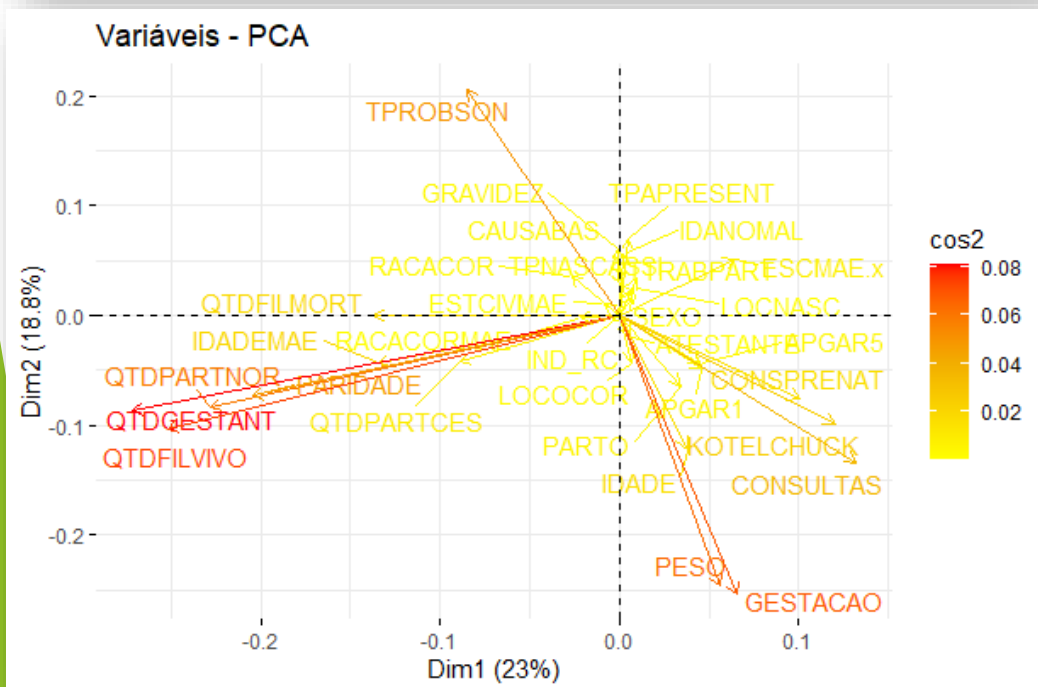
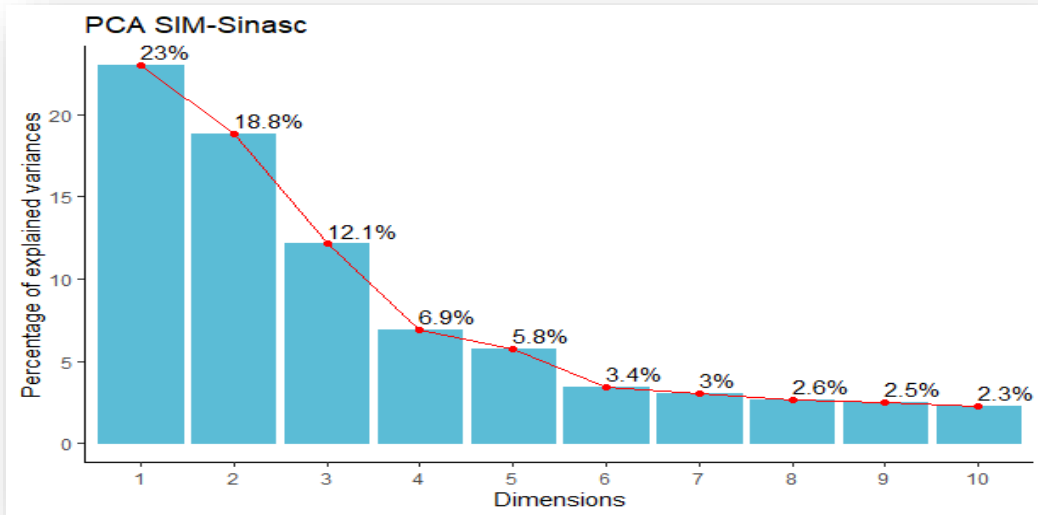
- União das bases do SIM de crianças de até 1 ano de idade e o Sinasc a fim de obter os recém-nascidos para o ano de 2019 que vieram a óbito no mesmo período.
- Para realização, foi necessário agendamento na sala segura do Ministério da Saúde e depois do *linkage* entre as bases, anonimização e tratamentos de segurança.

Sistemas de Informação	ID	Variáveis para Pareamento	ID	Variáveis para Revisão Manual
SIM	a	Nome	1	Idade da mãe
	b	Sobrenome	2	Sexo
	c	Data de nascimento	3	Município de Ocorrência
	d	Município de Residência	4	Nome da mãe
Sinasc	a	Nome	1	Idade da mãe
	b	Sobrenome	2	Sexo
	c	Data de nascimento	3	Município de Ocorrência
	d	Município de Residência	4	Nome da mãe

Processamento de Dados

- ▶ A base liberada pela CGIAE/MS apresentou 35.293 registros com 59 variáveis.
- ▶ Para informações repetidas, considerou-se prioritariamente as informações do SIM.
- ▶ Números de DN inválida foram retirados.
- ▶ A idade foi separada em três grupos: Neonatal precoce (até 6 dias de vida); Neonatal tardio (até 27 dias de vida); pós-neonatal (até 1 ano de idade).
- ▶ A cor/raça foi alterada para dois grupos: indígenas e não-indígenas.
- ▶ A base para a aplicação de componentes principais registrou 26.738 com 36 variáveis.
- ▶ Para aplicar o PCA, retirou-se as colunas com mais de 10% de dados nulos: Fonte de investigação, TPPOS, mês de início do pré-natal. Após a retirada dessas colunas, foram excluídos os registros com alguma informação faltante.

Análise de Componentes Principais



- ▶ A base registrava 19.506 com 32 variáveis.
- ▶ Quantidade de gestações, quantidade de filho vivo e morto, peso ao nascer, TPRobson são as variáveis que mais contribuem para as duas dimensões.
- ▶ Possibilidade de 4 grupos.

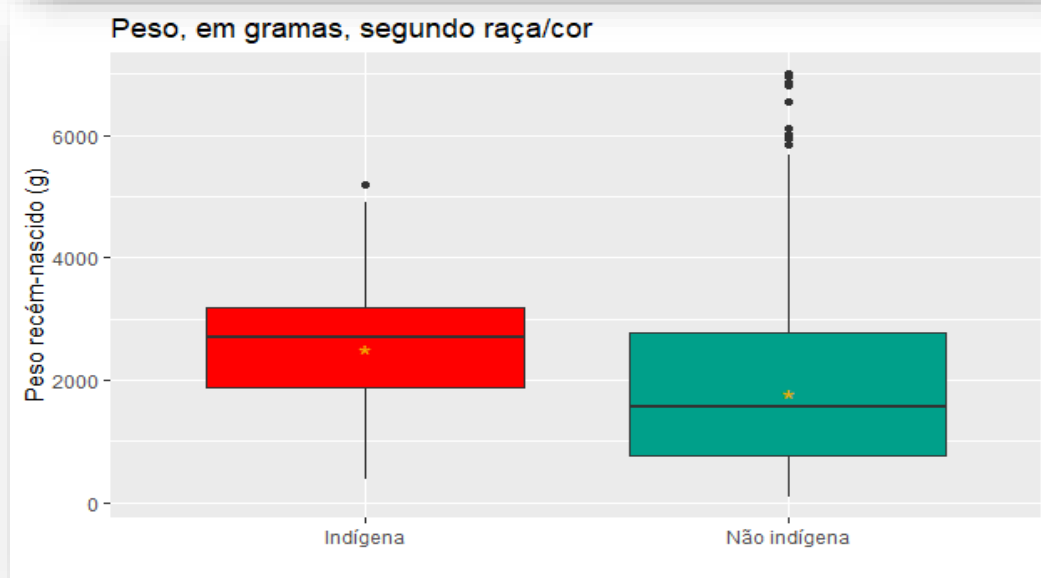
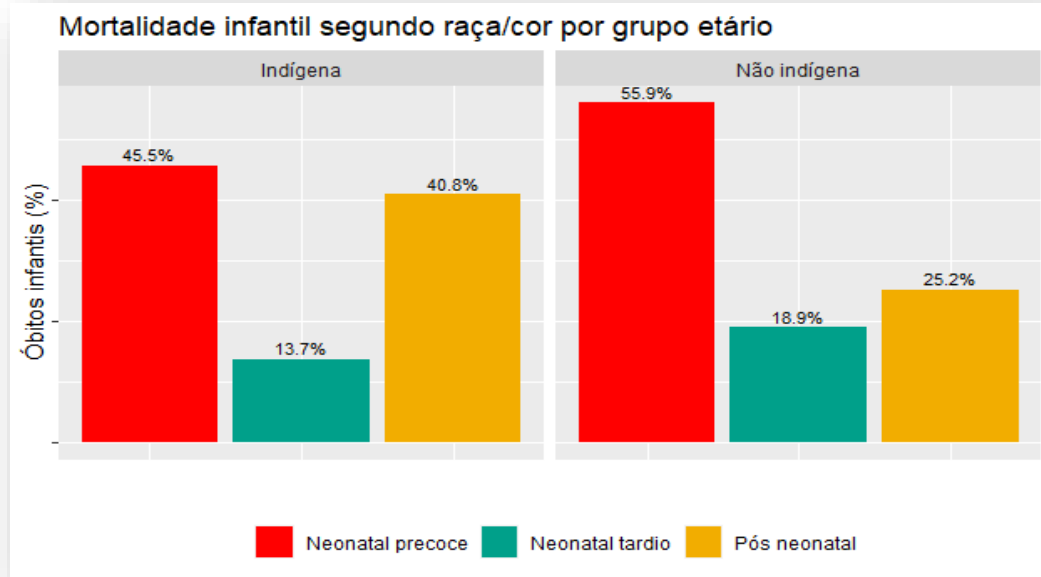
Análise e Exploração de Dados

Informações relativas à mãe

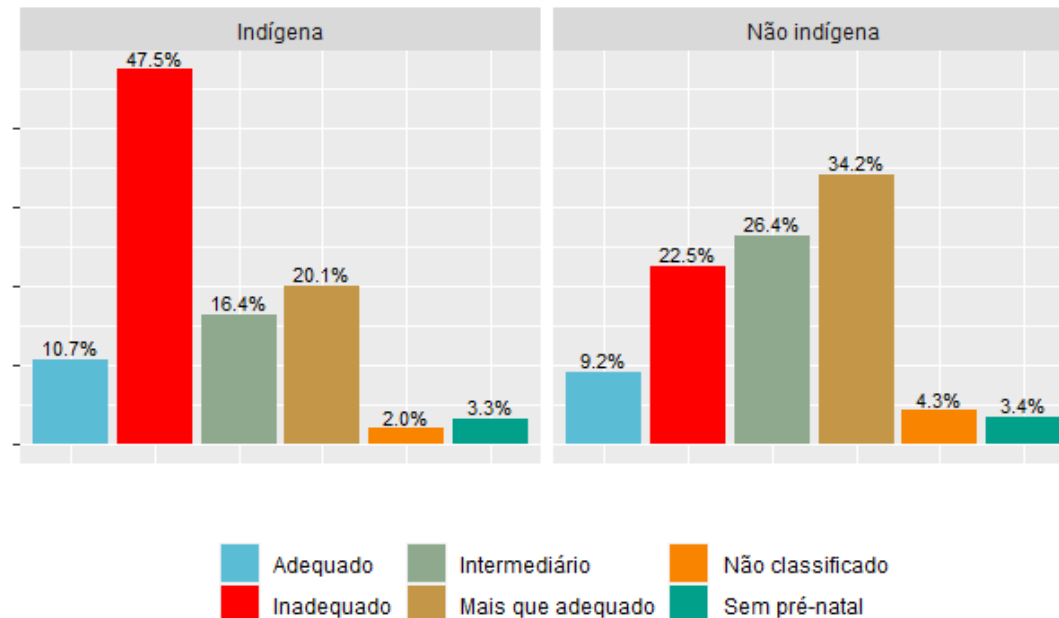
Variável	N	(%)
Raça/cor		
Parda	12.972	57,80%
Branca	6.876	30,64%
Preta	1.712	7,63%
Indígena	316	1,41%
Amarela	80	0,36%
Não informado	486	2,17%
Estado Civil		
Solteiro	11.285	50,29%
Casado	6.074	27,07%
União estável	4.531	20,19%
Separado judicialmente/divorciado	314	1,40%
Viúvo	46	0,20%
Não informado	192	0,86%
Escolaridade		
Superior Incompleto	12.334	54,96%
Médio	4.209	18,76%
Superior Completo	3.543	15,79%
Fundamental II	803	3,58%
Fundamental I	520	2,32%
Não informado	1.033	4,60%

Análise e Exploração de Dados

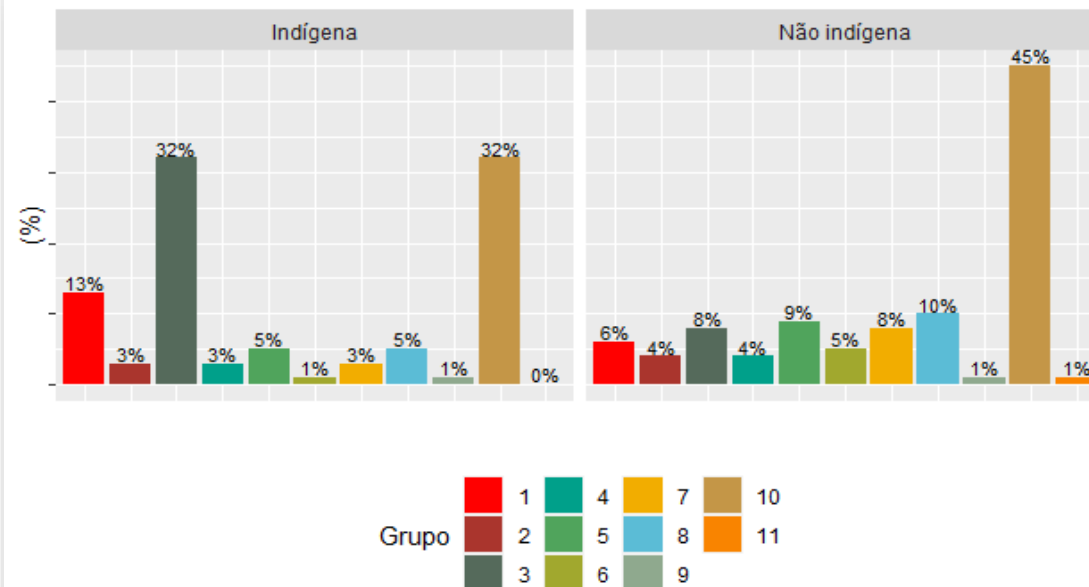
► Informações infantis



Assistência de pré-natal (APNCU) por raça/cor



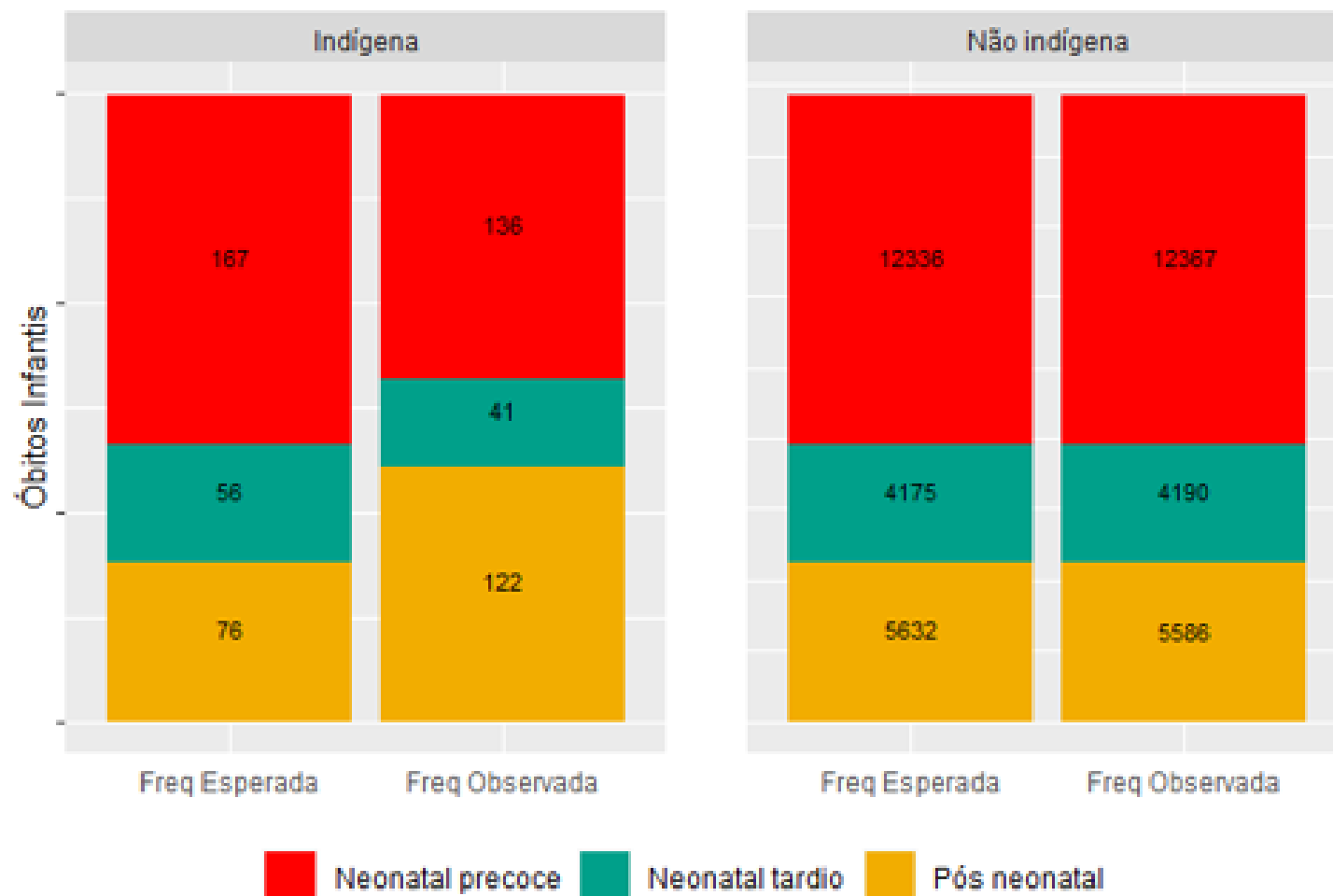
Grupo de Robson por raça/cor



Teste de Hipóteses

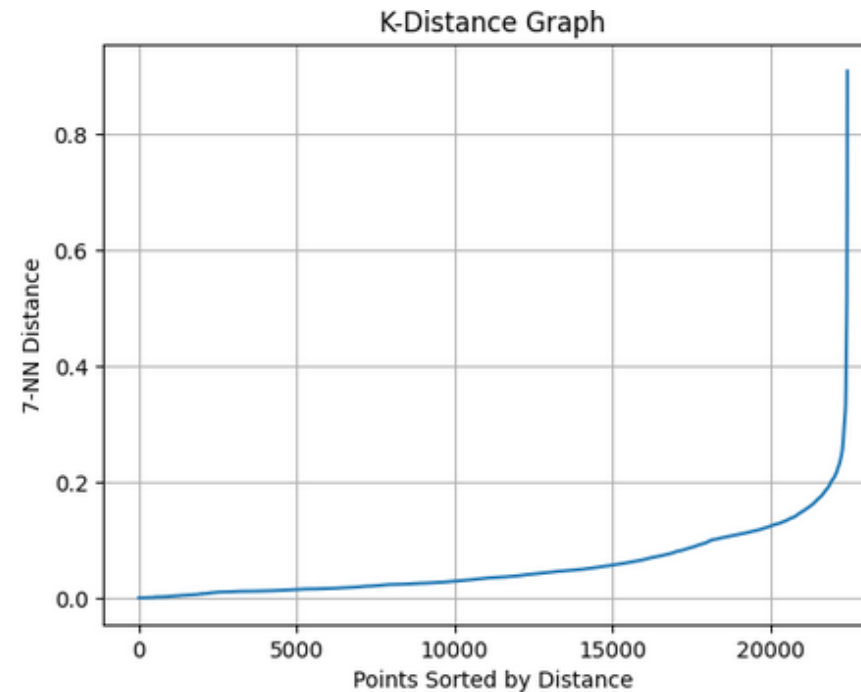
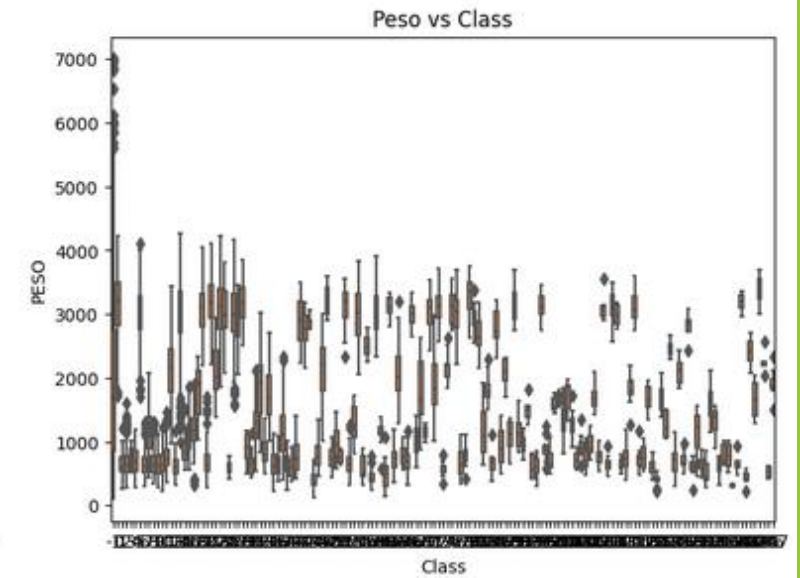
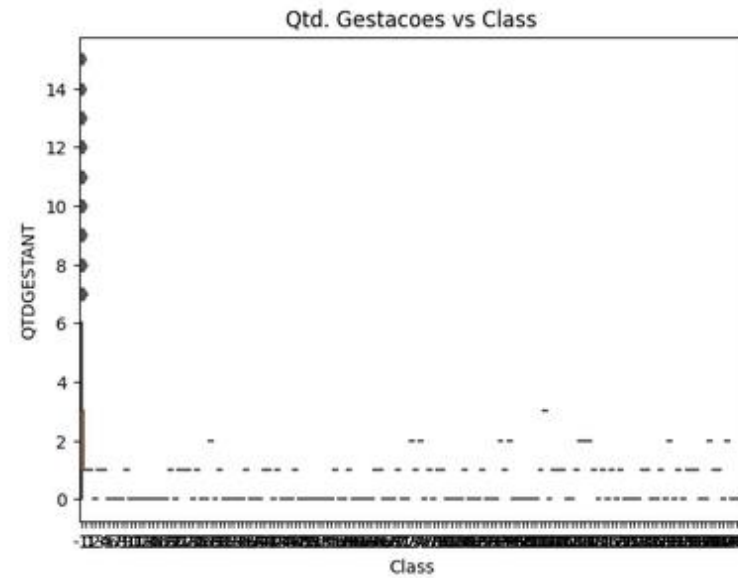
- ▶ **Hipótese nula:** Há independência entre cor/raça e grupo etário
- ▶ **Hipótese alternativa:** Não há independência entre cor/raça e grupo etário
- ▶ O χ^2 é de 38,07.
- ▶ O p-valor é menor que 0,01.
- ▶ Comparando as frequências esperadas e observadas de acordo com a cor/raça, o grupo indígena tem diferença significativa entre os valores.

Mortalidade infantil observada e esperada por cor/raça



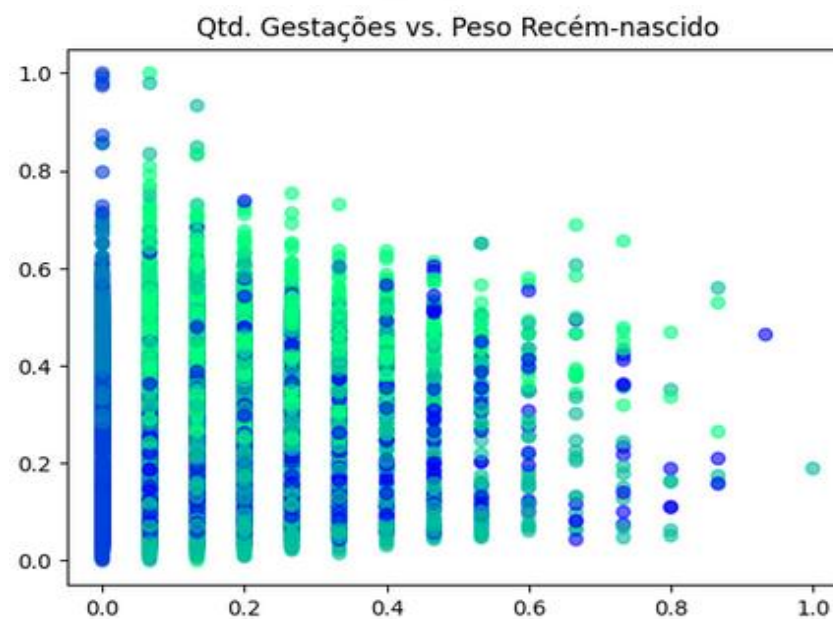
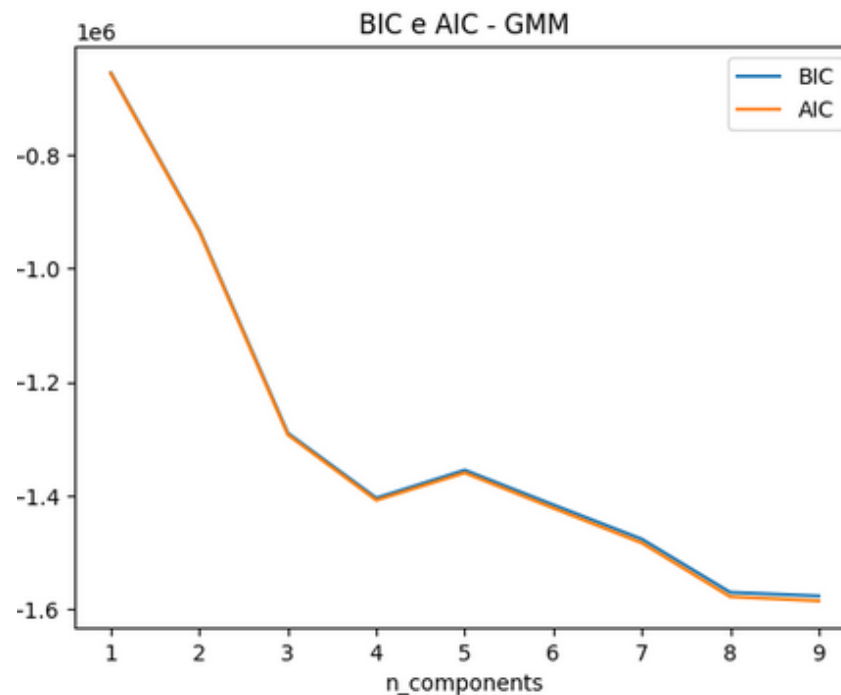
Modelagem DBSCAN

EPS	Samples	Silhouette Score
0.20	7	-0.45
0.20	8	-0.43
0.30	7	-0.40
0.30	8	-0.39
0.40	7	-0.33
0.40	8	-0.34



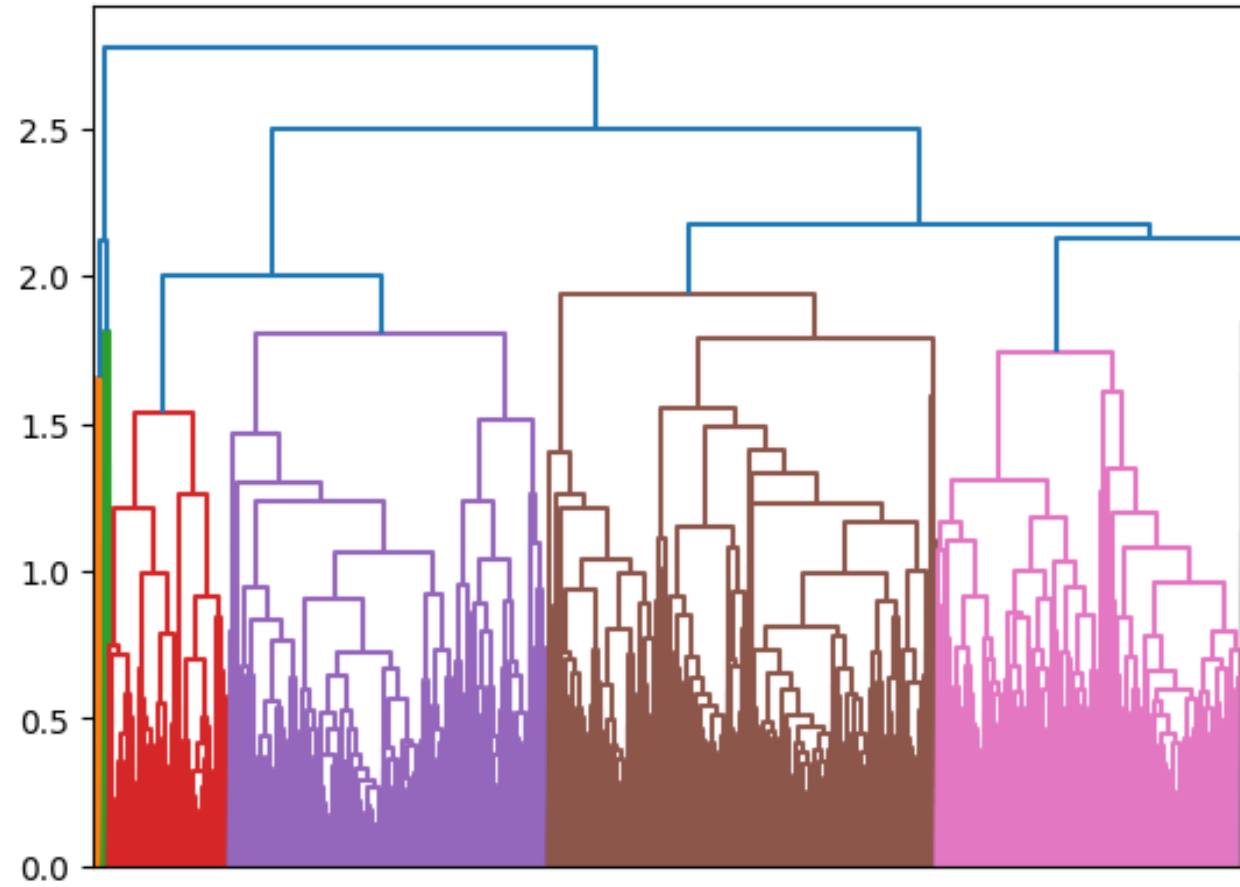
Modelagem Gaussian Mixture Model

Número Componentes	GMM Score
2	4.53
3	4.53
4	1.87
5	1.86
6	7.04
7	2.65
8	6.87



Modelagem Hierárquico

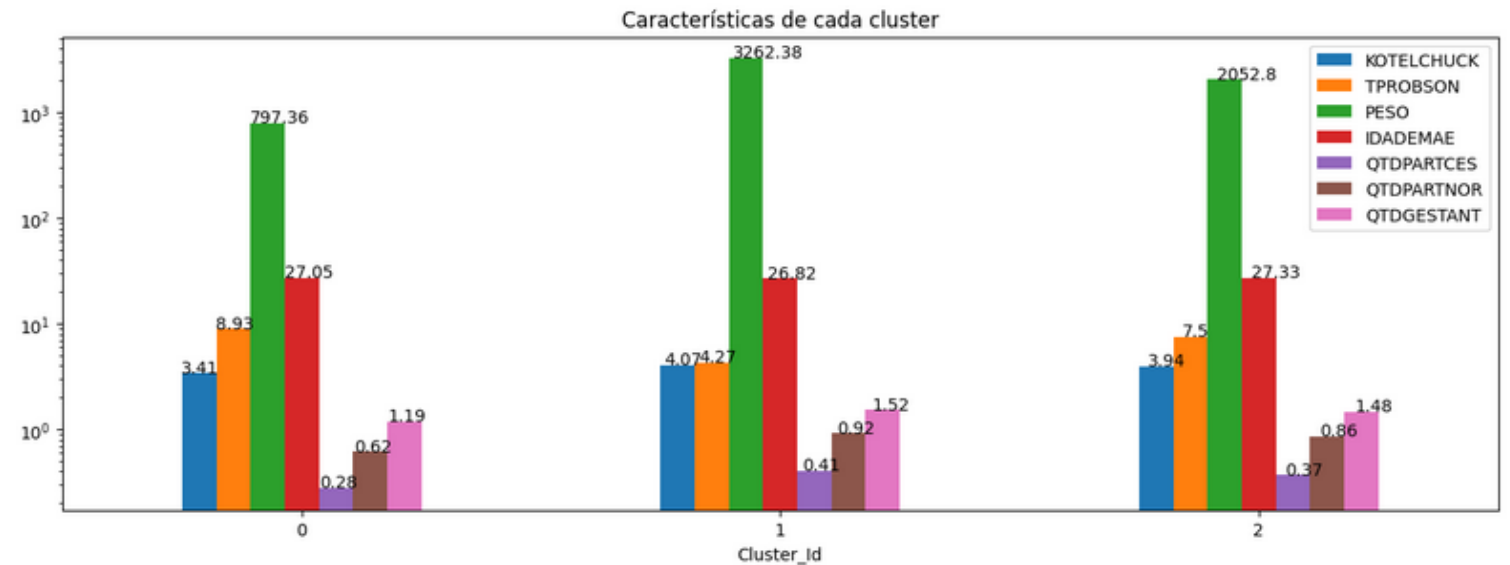
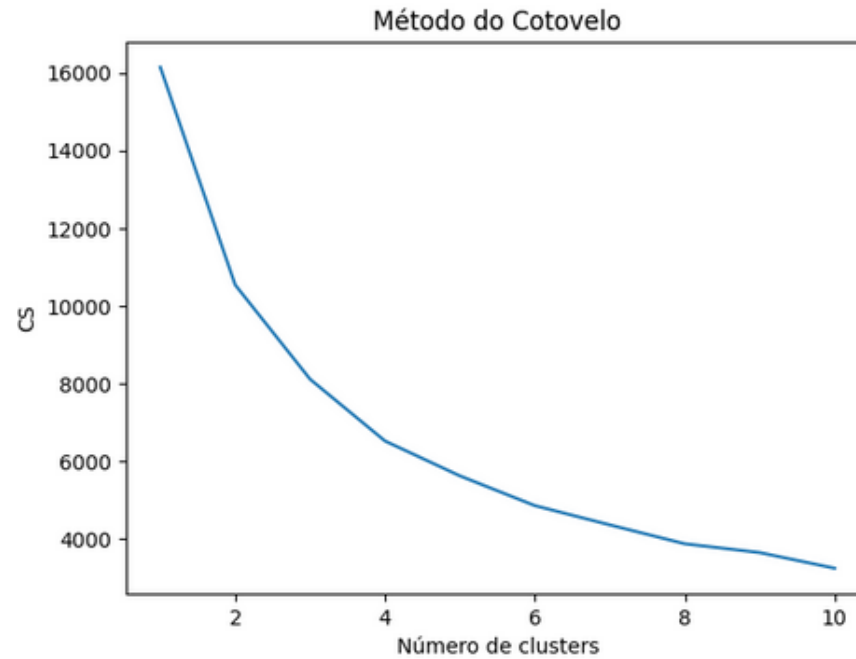
Número de Clusters	Silhouette Score
2	0.37
3	0.36
4	0.33
5	0.32
6	0.32
7	0.36



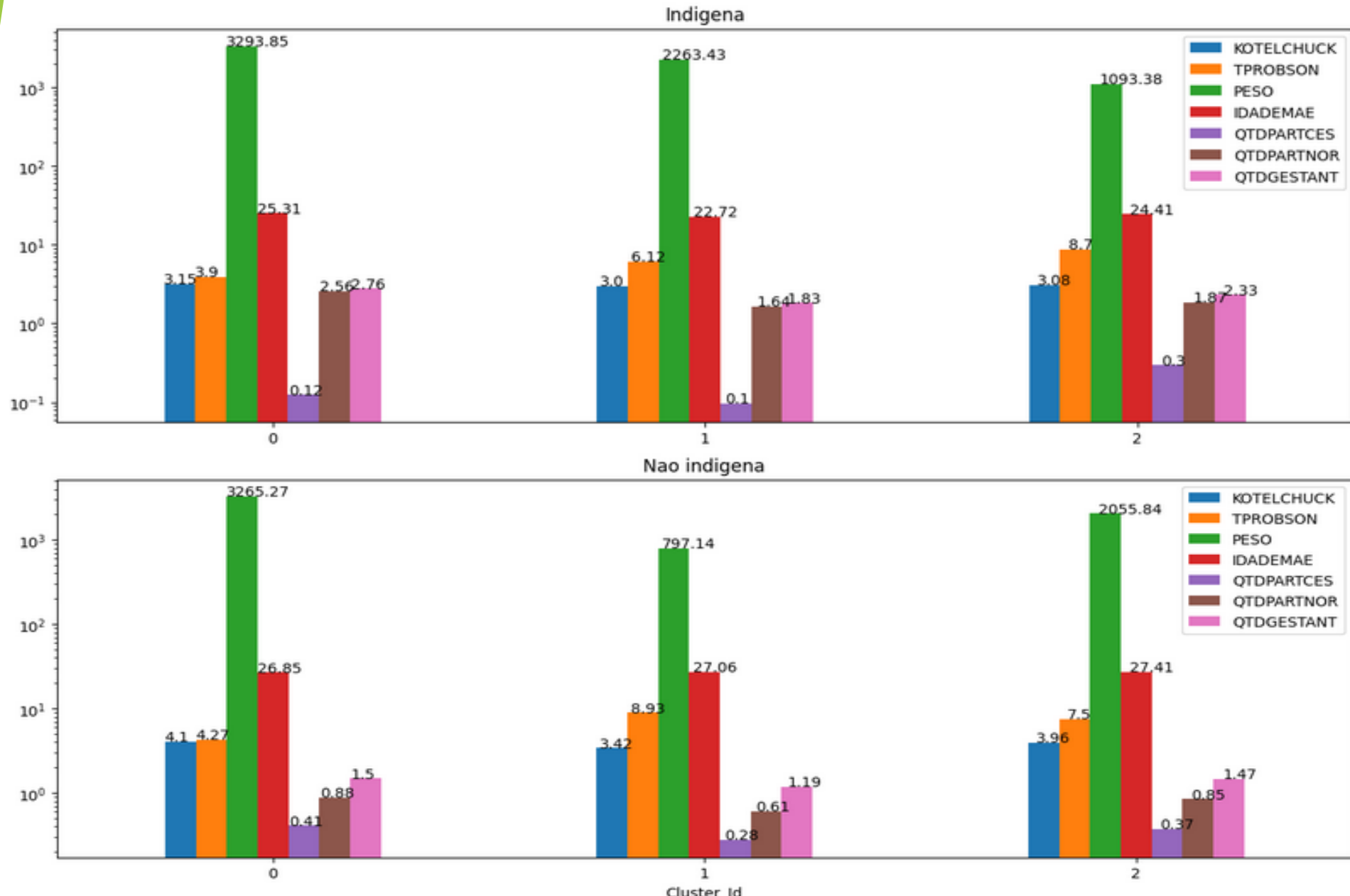
Modelagem K-Means

Número Clusters	Silhouette Score
2	0.68
3	0.62
4	0.58
5	0.56
6	0.54
7	0.53
8	0.53

Davis Bouldin Score: 0.51

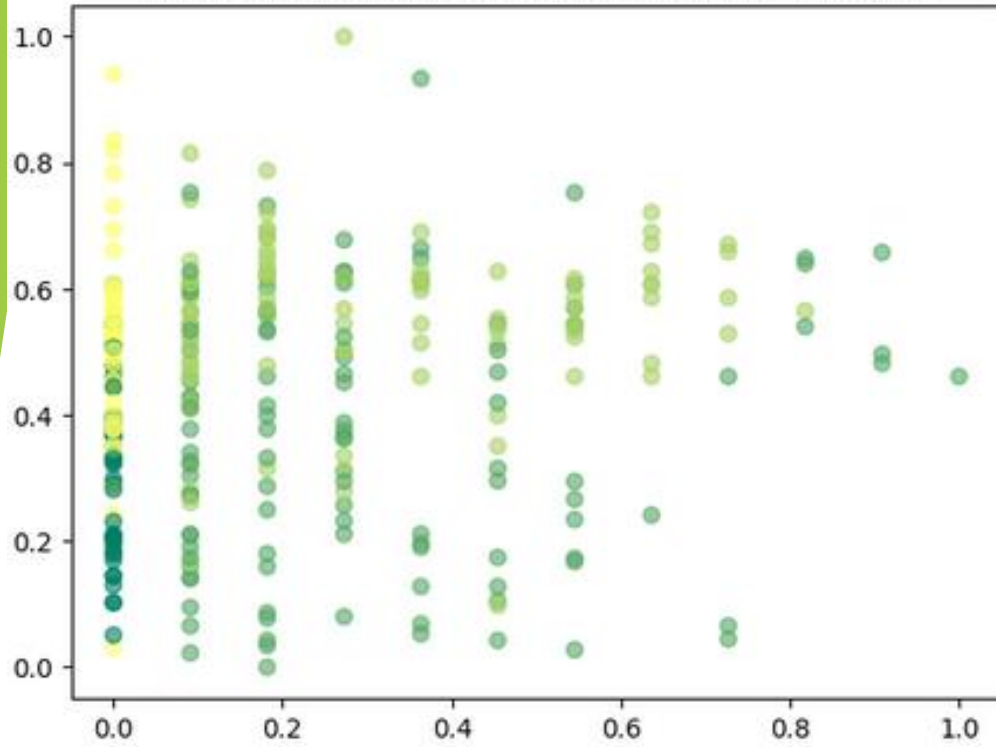


Resultados

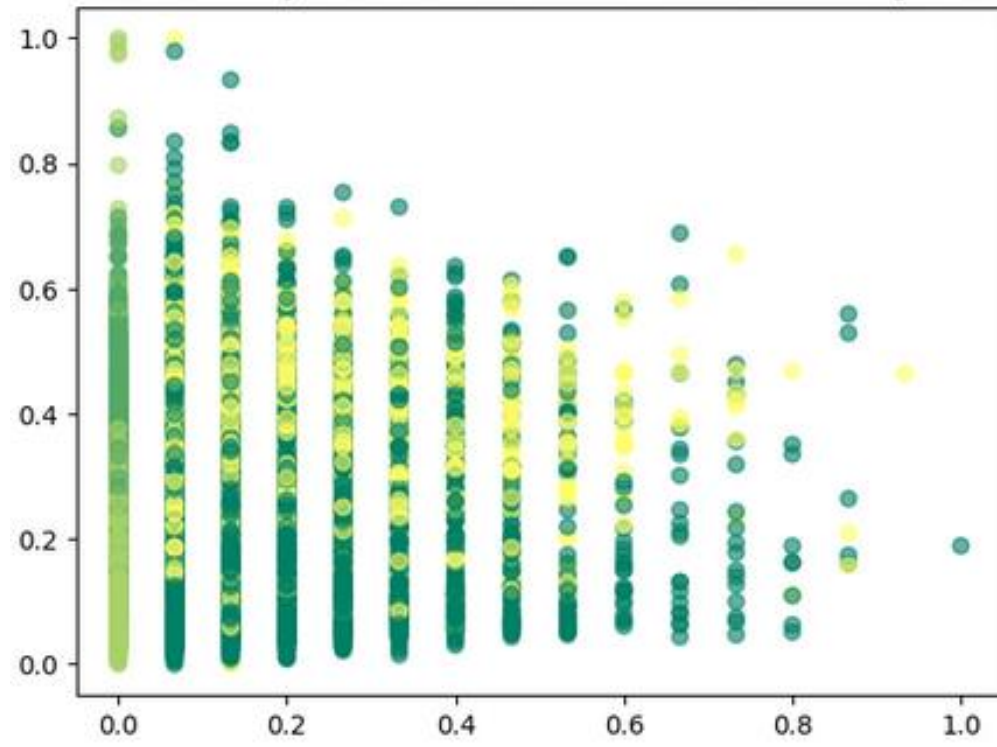


Resultados

Qtd. Gestações vs. Peso Recém-nascido - Indígena

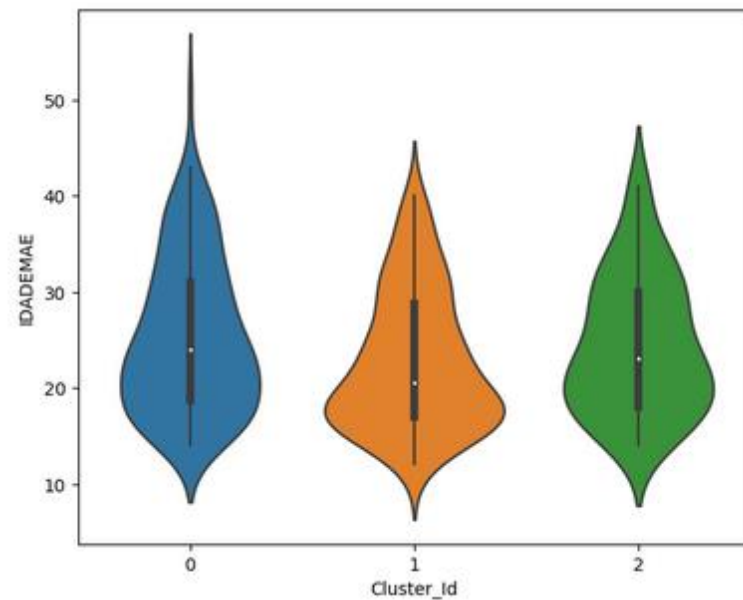
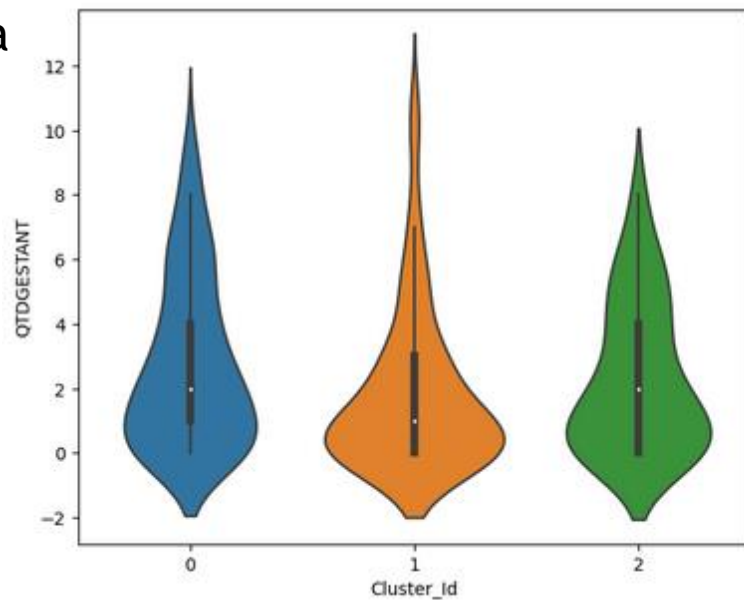


Qtd. Gestações vs. Peso Recém-nascido - Não indígena

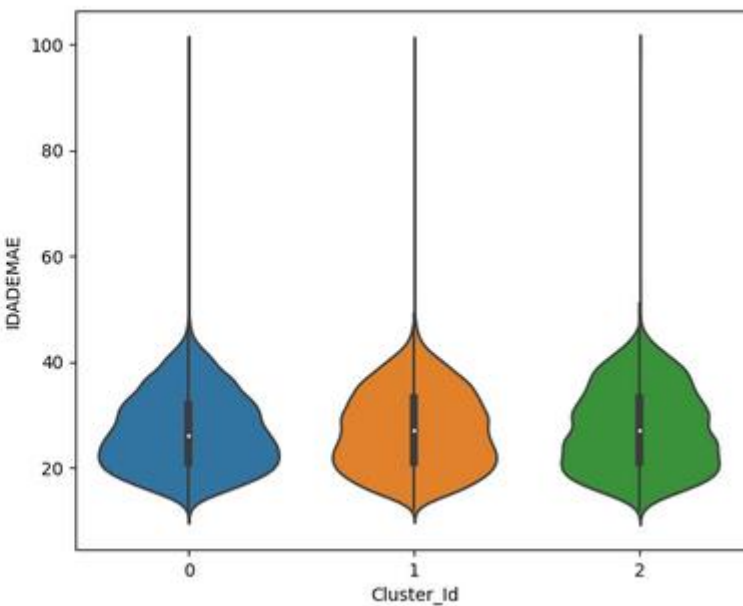
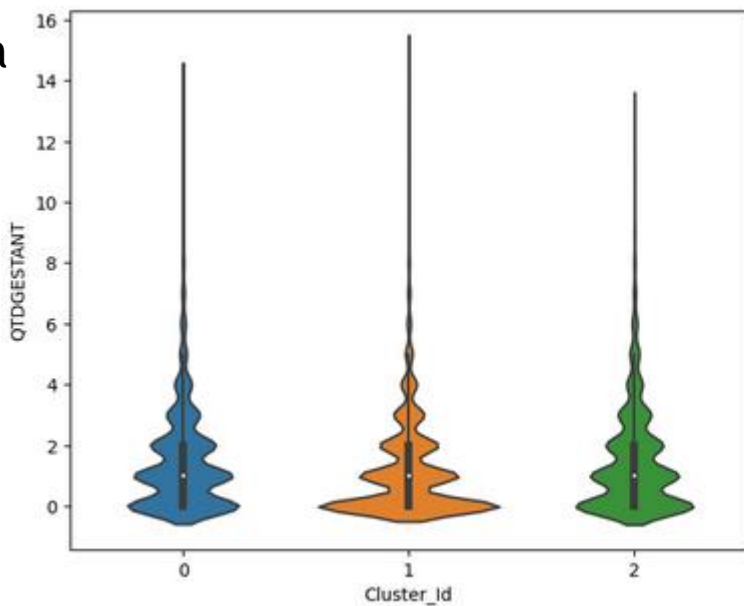


Resultados

Indígena

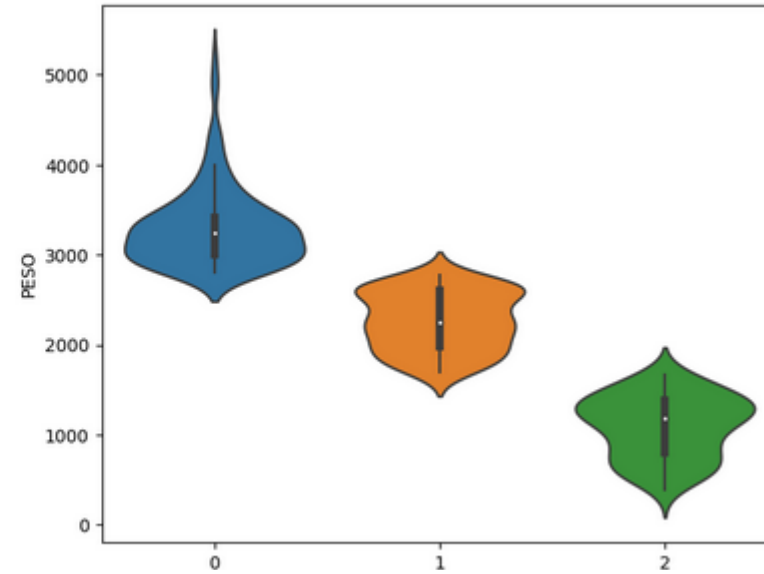
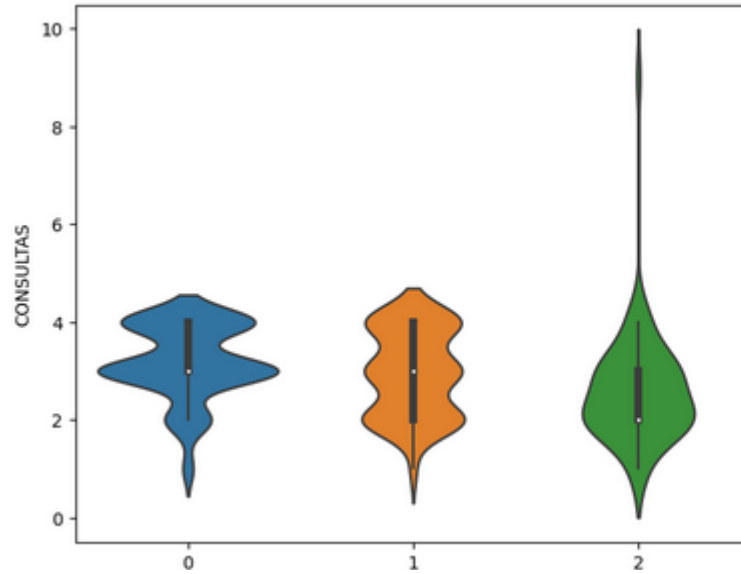


Não-Indígena

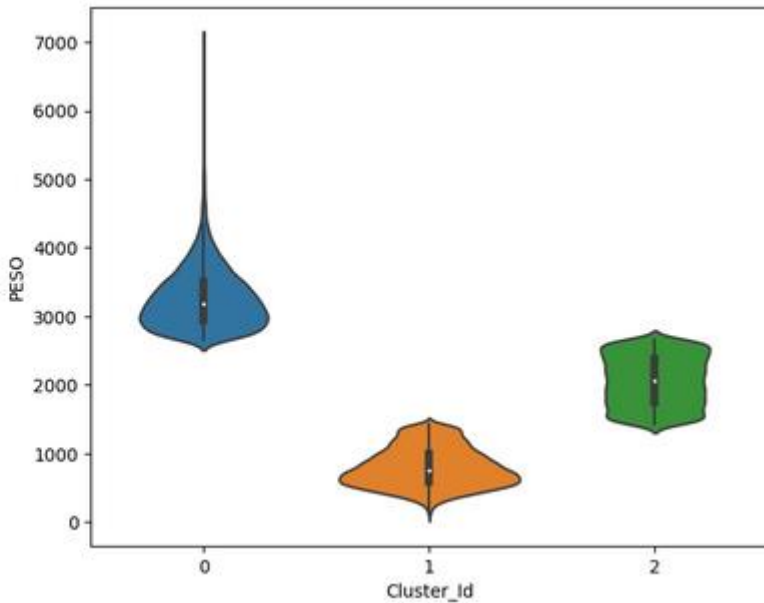
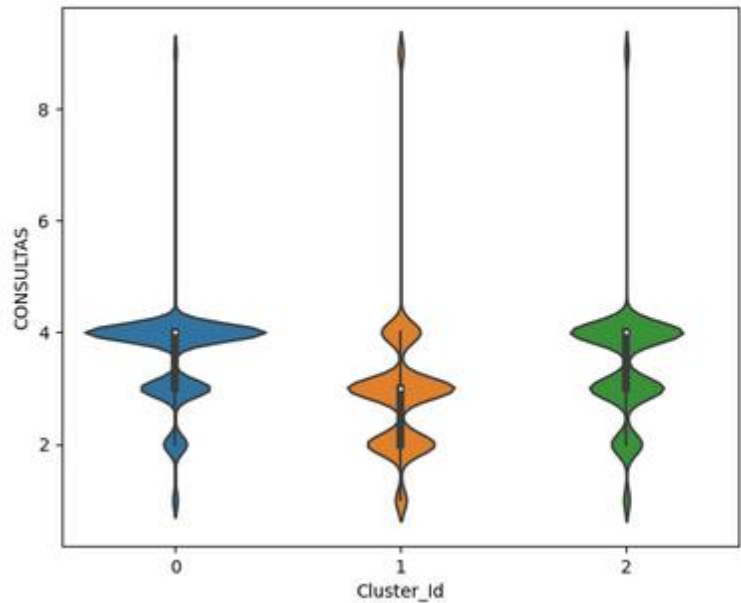


Resultados

Indígena



Não-Indígena



Conclusão

Grupo 0

Recém-nascidos com mais gramas ao nascer, mães mais velhas, outras gestações, pré-natal aquém do esperado e primeiros grupos de Robson.

Crianças com maior peso ao nascer, mães mais novas, maior quantidade de cesáreas, pré-natal adequado e grupos intermediários de Robson.

Grupo 1

Crianças com peso mediano, mães mais novas, menor quantidade de gestações, grupos de Robson intermediários e pré-natal aquém.

Crianças com baixo peso ao nascer, mães de idade mediana, grupos finais de Robson e pré-natal aquém.

Grupo 2

Crianças com baixo peso ao nascer, mães com idade mediana, maior número de cesáreas, grupos finais de Robson e pré-natal aquém.

Crianças com poucos gramas ao nascer, mães mais velhas, pré-natal adequado, grupos intermediários de Robson.

Obrigada

Referências

- Victora CG, Barros FC. Infant mortality due to perinatal causes in Brazil: trends, regional patterns and possible interventions. Sao Paulo Med J. 2001;119:33-42.
- JOHNSON, R. A.; WICHERN, D. W. Applied multivariate statistical analysis. 4th ed. Upper Saddle River, New Jersey: Prentice-Hall, 1999, 815 p. World Health Organization, UNICEF, United Nations Population Fund and The
- World Bank, Trends in Maternal Mortality: 2000 to 2017 WHO, Geneva, 2019.