

**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
DE SÃO PAULO**

DAVI FERNANDES SIQUEIRA CAMPOS - SP3082148

LUIZA ROSITO REIS DOS SANTOS - SP 3099814

SARAH MENDES SANTIAGO - SP3083047

**Análise estatística Sistema de Informação
sobre Nascidos Vivos Sinasc(2010)**

São Paulo

2024

Sumário

Introdução.....	3
Coleta de Dados.....	4
Tratamento de dados.....	4
Estatística Descritiva.....	4
Análise de Probabilidade.....	8
Inferência Estatística.....	9
Teste de normalidade.....	9
Histogramas.....	11
Teste de Shapiro.....	12
Teste Homogeneidade de médias.....	12
Teste T.....	13
Discussão.....	14
Conclusão.....	15

Introdução

O Sistema de Informações sobre Nascidos Vivos (Sinasc) foi oficialmente implantado em 1990 com a missão de coletar e fornecer dados abrangentes sobre nascimentos em todo o território nacional, atendendo a todas as esferas do Sistema de Saúde. A implantação do SINASC foi gradual em todas as unidades da Federação e, desde 1994, muitos municípios brasileiros já apresentaram um número de registros superior ao publicado pelo IBGE, com base nos dados dos Cartórios de Registro Civil. O monitoramento dos nascimentos é essencial para compreender a situação de saúde de uma população e para avaliar políticas e ações de vigilância e atenção à saúde materno-infantil. O Sinasc, ao coletar informações detalhadas sobre nascimentos, permite a identificação de padrões e tendências, facilitando a tomada de decisões baseadas em evidências no âmbito da saúde pública.

Os dados registrados no Sinasc incluem informações socioeconômicas, local de residência e ocorrência dos nascimentos, anomalias congênitas, além de detalhes sobre parto e pré-natal. O sistema também registra uma ampla gama de dados, incluindo características sociodemográficas das mães, condições do parto, peso e condição de saúde dos recém-nascidos. Esses dados são fundamentais para a formulação de políticas públicas eficazes, a alocação de recursos e a implementação de intervenções que visam melhorar a saúde materno-infantil no Brasil.

Este estudo tem como objetivo analisar os dados do Sinasc disponíveis na plataforma DATASUS no ano de 2010, utilizando técnicas de estatística descritiva, probabilidade e inferência estatística. Através desta análise, pretendemos compreender melhor as condições de nascimento e saúde dos recém-nascidos no Brasil, explorar padrões e tendências nos dados e avaliar a eficácia das políticas de saúde materno-infantil implementadas nas últimas décadas. Com isso, esperamos fornecer insights valiosos que possam contribuir para a melhoria contínua dos serviços de saúde no país.

Coleta de Dados

Os dados analisados neste estudo são os pertencentes ao Sistema de Informação sobre Nascidos Vivos – Sinasc do ano de 2010.

Disponível

em:

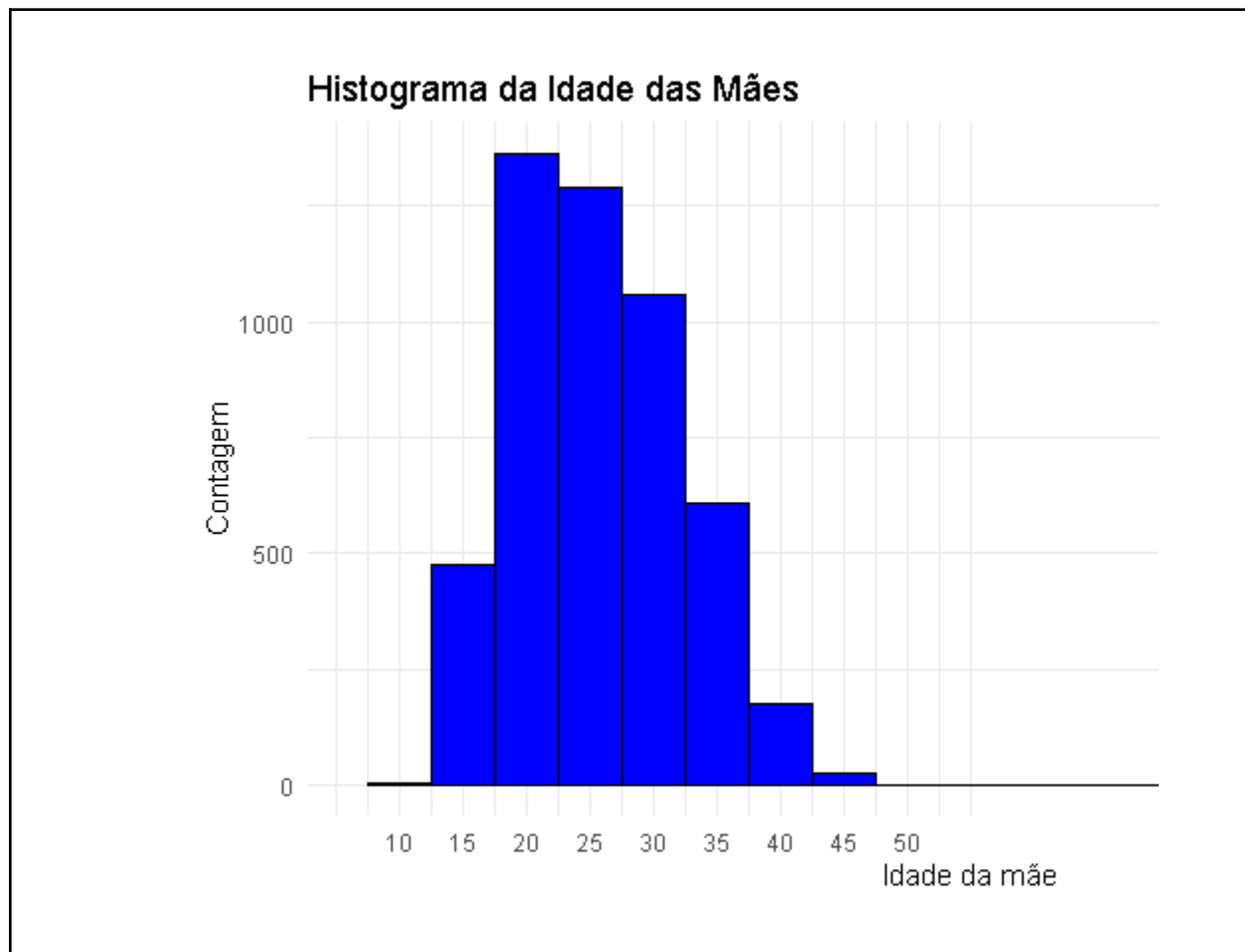
<https://opendatasus.saude.gov.br/dataset/sistema-de-informacao-sobre-nascidos-vivos-sinasc>

Tratamento de dados

Foi feito um tratamento inicial, visto que a base de dados possuía um número de registros massivos, onde foi retirado uma amostra de 5000 registros aleatórios.

Estatística Descritiva

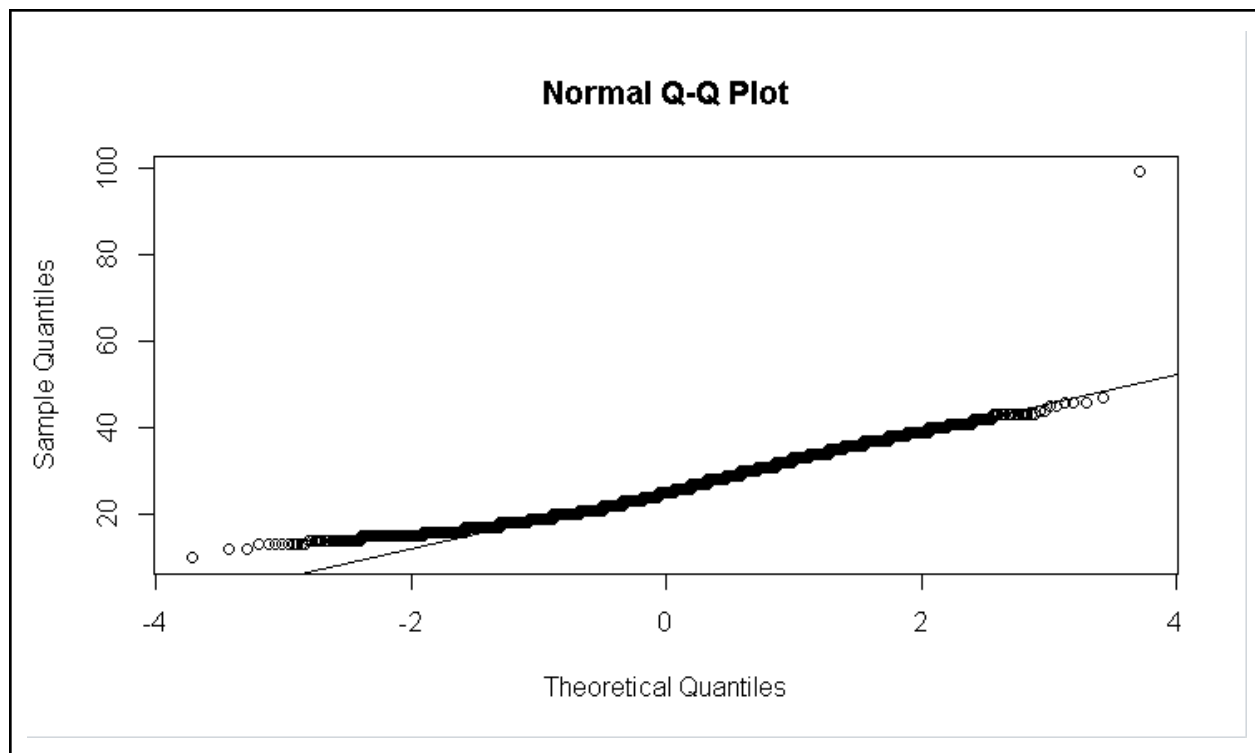
Iniciaremos a análise, verificando a distribuição das idades das mães em todo o território nacional registradas em 2010.



Histograma distribuição idade das mães SINASC 2010

Com um histograma rápido, podemos ver uma possível tendência de distribuição normal, onde levando em conta fatores como o período mais fértil e idade de mais segurança de gestação, faz-se coerente o pico se mostrar mais elevado entre 20 a 30 anos.

Entretanto, para assegurar a hipótese de ser uma distribuição normal, vamos realizar os outros testes estatísticos possíveis. Como analisar a distribuição das idades comparando-a com uma distribuição normal teórica.



Comparação distribuição das idades das mães 2010 X normalidade teórica

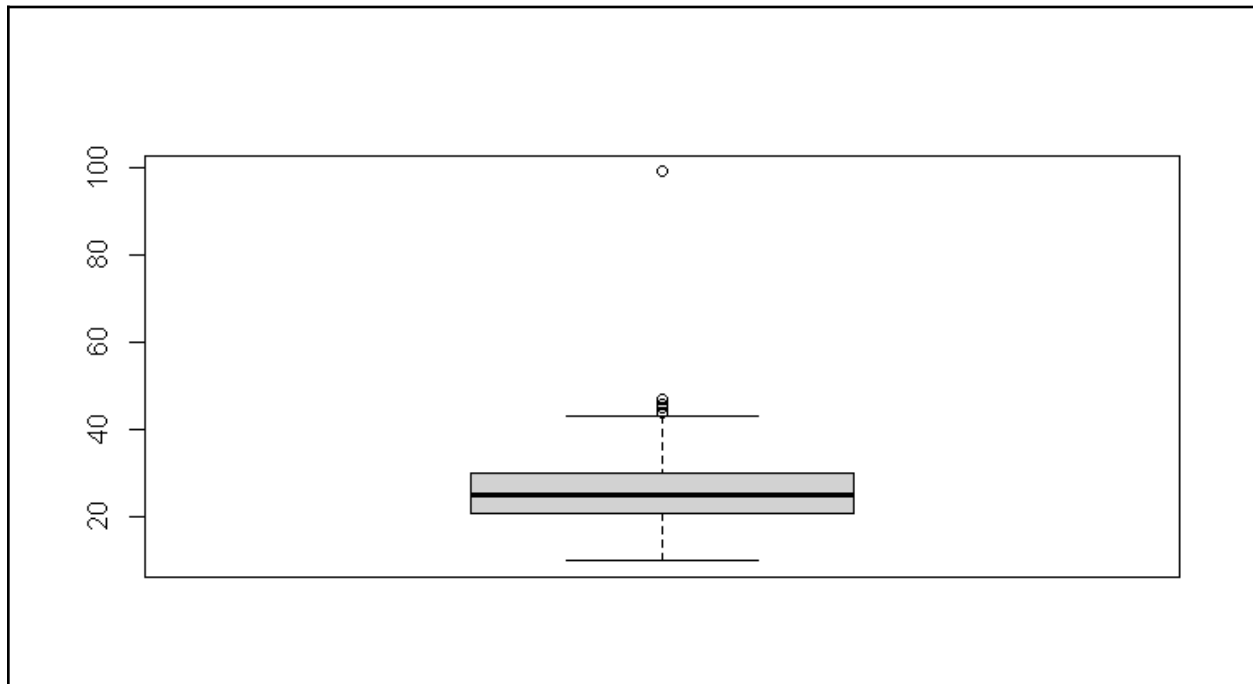
Traçando uma linha normal teórica contra a distribuição de idade das mães na amostra, podemos manter a ideia de uma distribuição normal.

Prosseguindo com um teste final de normalidade, vamos aplicar o **teste de Shapiro**:

```
shapiro-wilk normality test  
data:  amostra$IDADEMAE  
W = 0.96608, p-value < 2.2e-16
```

Teste shapiro - idade das mães SINASC 2010

Ao contrário do que imagina-se inicialmente, o nosso p-value, se mostra a princípio se tratar de uma hipótese alternativa (<0.05), ou seja, não normal, o que está indo contra as análises iniciais feitas. Um fator que pode ter influenciado este resultado, seria outliers na amostra, para isso vamos analisar através do **boxplot**.



Boxplot idade das mães registradas SINASC 2010

Podemos verificar não só vários outliers, como um outlier extremo, logo, podemos dizer que a idade das mães registradas no Brasil no ano de 2010 possui muito provavelmente uma distribuição normal, onde mesmo com um teste Shapiro inconclusivo, outras aferições nos mostraram o contrário, sendo essa diferença potencialmente causada por outliers.

Análise de Probabilidade

Investigaremos a probabilidade de um recém-nascido ser de raça/cor **branca**, dado que o parto foi **cesariana** através da probabilidade condicional.

Usando a formula:

$$P(A / B) = \frac{P(A \cap B)}{P(B)}$$

Temos:

Branco e Cesário = 0.30

Cesário = 0.52

$$P(A|B) = 0.3042932 / 0.5258639 = 0.5786539$$

$0.5786539 \times 100 = 57.87\%$ de probabilidade do parto Bebê ser de etnia **branca** sendo que o parto foi cesariana.

Inferência Estatística

Realizaremos uma inferência estatística para averiguar se existe diferença estatisticamente significativa entre o **peso do nascimento do bebê** dependendo da **forma do parto (Vaginal ou Cesariana)**.

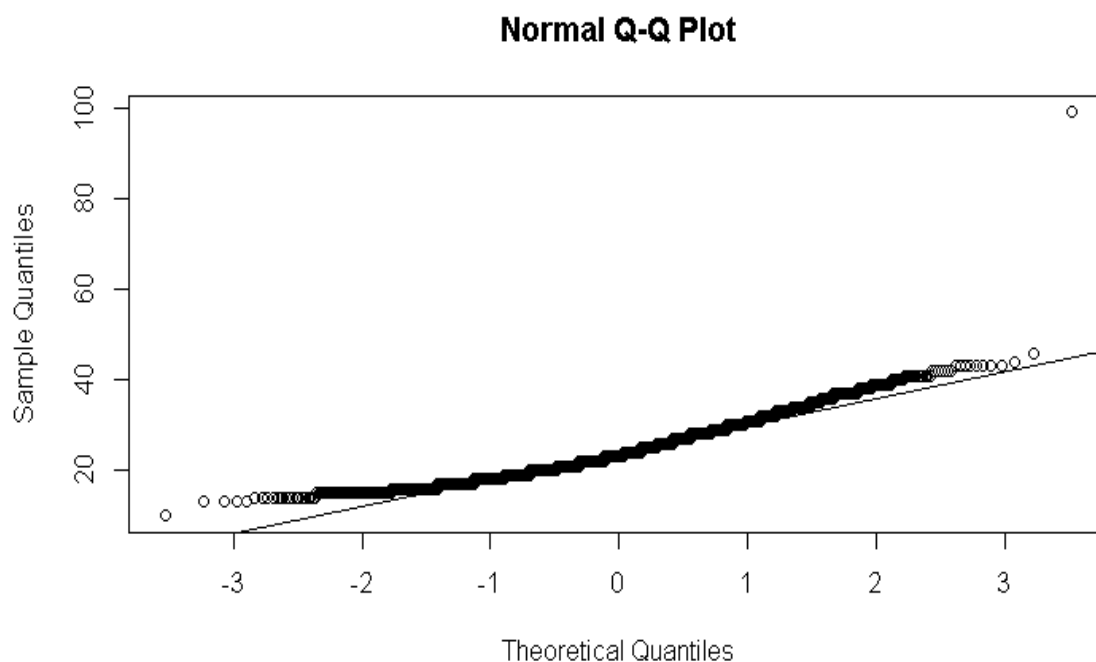
Como pede a inferência estatística de comparação as médias, temos que passar anteriormente pelos testes de Normalidade e Homogeneidade de médias.

Teste de normalidade

Utilizaremos : **Comparação com reta teórica, histograma e teste de Shapiro**

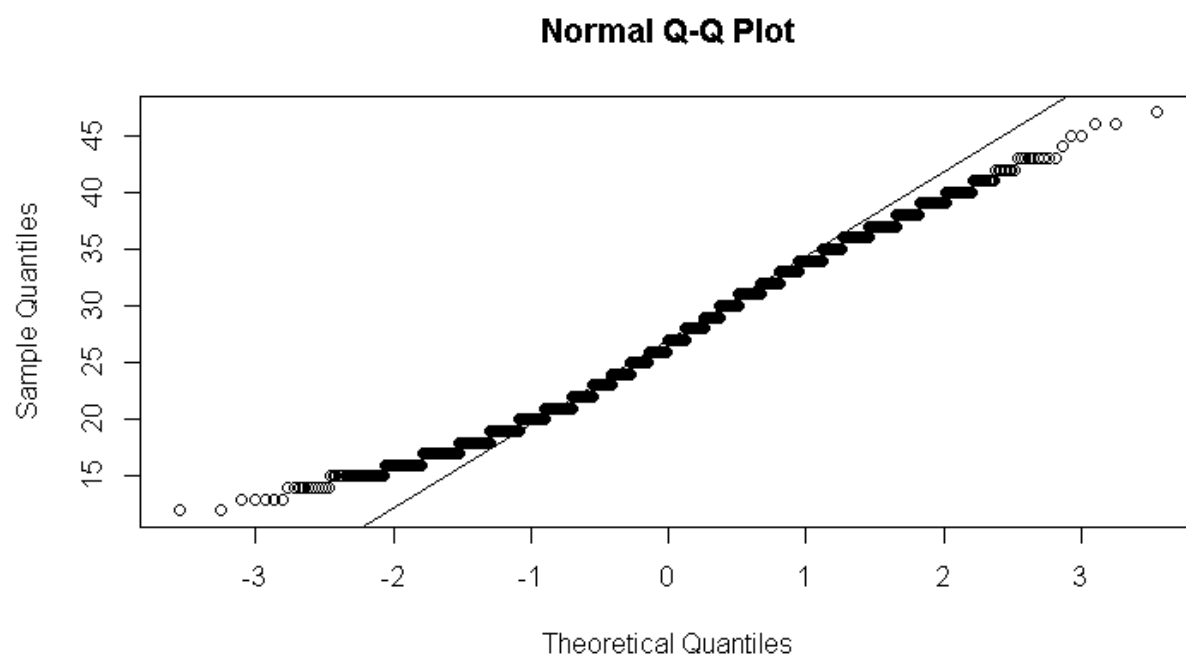
Comparação com reta normal teórica para distribuição de peso

Parto Vaginal:



Teste reta teórica normal X distribuição peso parto vaginal

Parto Cesário:



Teste reta teórica normal X distribuição peso parto cesário

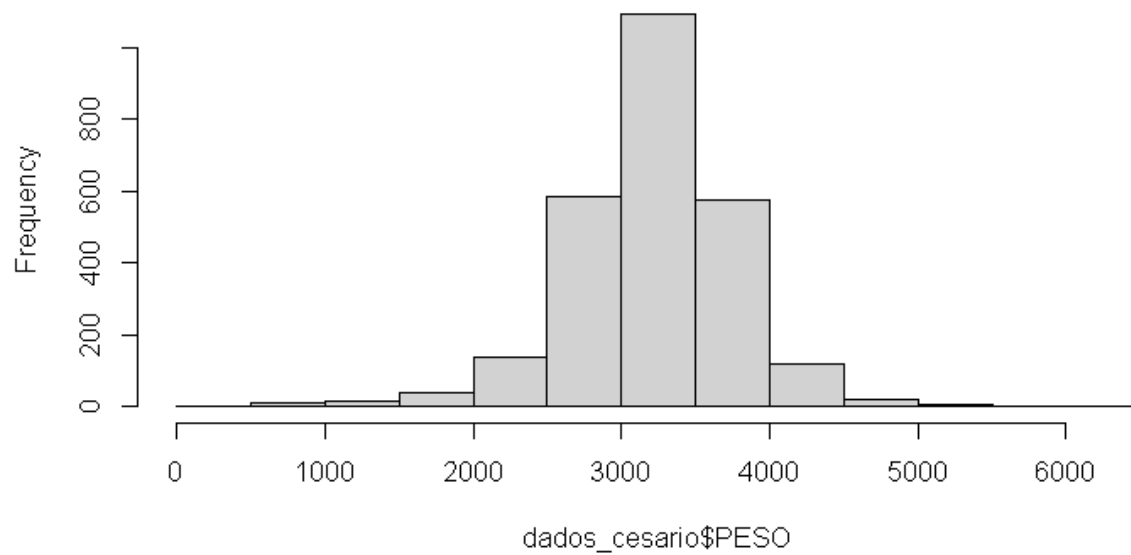
Histogramas:

Histogram of dados_vaginal\$PESO



Histograma distribuição peso bebê parto vaginal

Histogram of dados_cesario\$PESO



Histograma distribuição peso bebê parto cesário

Teste de Shapiro

Parto Vaginal

```
shapiro-wilk normality test  
  
data: dados_vaginal$PESO  
W = 0.93686, p-value < 2.2e-16
```

Parto Cesário:

```
shapiro-wilk normality test  
  
data: dados_cesario$PESO  
W = 0.96356, p-value < 2.2e-16
```

Como foi possível ver, tanto a reta como o histograma evidencia uma tendência normal, embora o teste de Shapiro não nos assegure, o que pode ter sido influenciado pelo grande tamanho da amostra, visto que este teste é sensível, e, outliers. Desta forma, seguiremos com a análise.

Teste Homogeneidade de médias

```
data: dados_vaginal$PESO and dados_cesario$PESO  
F = 1.0033, num df = 2379, denom df = 2606, p-value = 0.9338  
alternative hypothesis: true ratio of variances is not equal to 1  
95 percent confidence interval:  
 0.9275213 1.0854476  
sample estimates:  
ratio of variances  
 1.003311
```

Nosso p-values nos mostram um valor > 0.05 , explicitando uma heterogeneidade de médias das amostras, nos permitindo prosseguir.

Teste T

```
data: dados_vaginal$PESO and dados_cesario$PESO
t = -3.0805, df = 4942.4, p-value = 0.002078
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -79.58870 -17.68346
sample estimates:
mean of x mean of y
 3163.497  3212.133
```

Por fim, nosso teste T mostrou um resultado de p-value < 0.05 (0H), o que em outras palavras nos mostra a hipótese não nula, ou seja, existe uma **diferença estatisticamente relevante entre o peso dos recém nascidos de acordo com a forma do parto**.

Discussão

Podemos verificar que a distribuição normal está presente em inúmeros casos ao redor de nós, visto que é muito comum existir uma tendência central nas coisas, como por exemplo o QI de uma população, onde poucos serão muito ruins no teste e poucos serão “gênios”, ficando a maior concentração numa tendência central. Não diferente foi a idade das mães registradas no SINASC em 2010, onde poucos casos foram precoces e poucos tardios.

Quanto a probabilidade da etnia do bebê ser branca, visto que o parto foi cesário, podemos nos questionar se fatores socioeconômicos não estão atrelados a este resultado, visto que a utilização um parto mais profissionalizado esteve quase dois terços concentrado em pessoas brancas. Existe uma possível desigualdade ao acesso de saúde obstétrica a pretos, pardos e indígenas?

Por último, vimos que existe uma diferença estatística no peso dos recém nascidos em comparação ao tipo de parto, que pode incitar mais perguntas a respeito de como isso pode afetar o processo de desenvolvimento dessa criança.

Conclusão

Neste estudo, analisamos os dados do Sistema de Informação sobre Nascidos Vivos (Sinasc) do ano de 2010, disponíveis na plataforma DATASUS. A análise envolveu a extração de uma amostra de 5000 registros aleatórios, devido ao grande volume de dados, e a realização de diversos métodos estatísticos para entender melhor as características e tendências dos nascimentos registrados naquele ano.

Através da estatística descritiva, examinamos a distribuição das idades das mães. A análise preliminar sugeriu uma distribuição normal das idades, com um pico significativo entre os 20 e 30 anos, correspondendo ao período de maior fertilidade e segurança gestacional. No entanto, testes estatísticos subsequentes, incluindo o teste de Shapiro, indicaram que a distribuição não era perfeitamente normal, possivelmente devido à presença de outliers. A inspeção visual via boxplot confirmou a existência de vários outliers, inclusive extremos, que podem ter influenciado os resultados do teste de normalidade. Apesar disso, outras análises corroboram a hipótese de uma distribuição aproximadamente normal para a idade das mães.

Na análise de probabilidade, investigamos a probabilidade condicional de um recém-nascido ser de raça/cor branca, dado que o parto foi cesariana. Os cálculos indicaram uma probabilidade de aproximadamente 57.87%, fornecendo insights valiosos sobre a relação entre a cor/raça do bebê e o tipo de parto.

Em seguida, realizamos uma inferência estatística para verificar se havia uma diferença estatisticamente significativa no peso ao nascimento dos bebês em função do tipo de parto (vaginal ou cesariana). Antes de proceder com a comparação das médias, conduzimos testes de normalidade e homogeneidade. Utilizamos a comparação com a reta teórica normal, histogramas e o teste de Shapiro para confirmar a adequação dos dados para a análise inferencial.

O conjunto de análises realizadas permite uma compreensão aprofundada dos nascimentos no Brasil em 2010, evidenciando a importância de um tratamento cuidadoso dos dados e a aplicação de múltiplos métodos estatísticos para obter resultados confiáveis. Apesar das limitações impostas pelos outliers, as conclusões gerais sobre a distribuição das idades das mães e a relação entre o tipo de parto e características dos recém-nascidos fornecem uma base sólida para futuras investigações e para a formulação de políticas públicas de saúde materno-infantil. Este estudo demonstra a relevância do Sinasc como uma ferramenta essencial para o

monitoramento e avaliação das condições de saúde relacionadas aos nascimentos no Brasil.