

**Universidade de São Paulo
Instituto de Matemática e Estatística**

Centro de Estatística Aplicada

Relatório de Análise Estatística

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

**“Concentrações de proteína C-reativa ultrasensível materna e no cordão umbilical
e tamanho do recém-nascido”**

Evandro Santalha Ferreira
Rick Armstrong de Santanna Cesar
Airlane Pereira Alencar

São Paulo, junho de 2025

CENTRO DE ESTATÍSTICA APLICADA - CEA – USP

TÍTULO: Relatório de Análise Estatística sobre ofic Projeto: “Concentrações de proteína C-reativa ultrasensível materna e no cordão umbilical e tamanho do recém-nascido”.

PESQUISADORA: Aldaisa Pereira Lopes

ORIENTADORA: Patrícia Helen de Carvalho Rondó

INSTITUIÇÃO: Faculdade de Saúde Pública – Universidade de São Paulo

FINALIDADE DO PROJETO: Mestrado

RESPONSÁVEIS PELA ANÁLISE:

Evandro Santalha Ferreira

Rick Armstrong de Santanna Cesar

Airlane Pereira Alencar

REFERÊNCIA DESTE TRABALHO: ALENCAR, A.P; FERREIRA, E.S; CESAR, R.A.S (2025). **Relatório de análise estatística sobre o projeto: “Concentrações de proteína C-reativa ultrasensível materna e no cordão umbilical e tamanho do recém-nascido”**. São Paulo, IME-USP. (RAE-CEA-25P05)

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS:

ALTMAN, D.G. (1991). **Practical Statistics for Medical Research**. London: Chapman & Hall.

CHAWLA, N. et al. (2002). SMOTE: Synthetic Minority Over-sampling Technique. **Journal Of Artificial Intelligence Research**, 16, 321-357.

NEVES, F. S. et al. (2020). Tri-ponderal mass index is useful for screening children and adolescents with insulin resistance. **Rev. paul. pediatr.**, 38, 1-8.

PAULA, G. A. (2024). **Modelos de regressão: com apoio computacional**. 4. ed. São Paulo: Blucher.

RONDÓ, P. H. C. et al. (2003). Maternal psychological stress and distress as predictors of low birth weight, prematurity, and intrauterine growth retardation. **European Journal of Clinical Nutrition**, 57, 266–272.

TERÁN, J. M. et al. (2017). New birthweight charts according to parity and type of delivery **for the Spanish population**. **Gaceta sanitaria**, 31, 116-122.

VILLACORTA, H. et al. (2007). C-Reactive Protein: an Inflammatory Marker with Prognostic Value in Patients with Decompensated Heart Failure. **Arq Bras Cardiol.**, 88, 585-589.

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

Google Docs

Microsoft Excel for Windows

R (versão 4.5.0)

RStudio

Python (versão 3.9)

Visual Studio Code

TÉCNICAS ESTATÍSTICAS UTILIZADAS

Análise Descritiva Unidimensional (03:010)

Análise Descritiva Multidimensional (03:020)

Regressão Logística (07:090)

ÁREA DE APLICAÇÃO

Medicina Epidemiologia (14:040)

RESUMO

Bebês que nascem pequenos têm risco elevado para diversas enfermidades durante toda a vida, portanto estudar fatores de risco para bebês pequenos é de grande interesse para a saúde pública. Em estudos passados, foi estabelecida uma possível relação entre a concentração da proteína C-Reativa ultrasensível no sangue materno durante o terceiro trimestre com o risco do bebê nascer pequeno. O estudo visa testar esta relação novamente, conjuntamente com a concentração de proteína C-Reativa no cordão umbilical do bebê.

A partir do ajuste de modelos lineares generalizados, não foi possível estabelecer uma relação estatisticamente significativa entre ambas a concentração da proteína C-Reativa ultrasensível no sangue materno e a concentração de proteína C-Reativa ultrasensível no cordão umbilical do bebê com o índice ponderal e o peso para a idade gestacional.

Sumário

1. Introdução	8
2. Objetivo	8
3. Descrição do estudo	9
4. Descrição das variáveis	9
5. Análise descritiva	10
6. Análise inferencial	13
7. Conclusões	17
APÊNDICE A	18
APÊNDICE B	38
APÊNDICE C	63

1. Introdução

O sistema imunológico da mulher desempenha um papel fundamental durante a gestação, alterando-se diversas vezes ao longo da mesma. Mulheres com complicações durante a gestação apresentam atividade imunológica alterada em relação às mulheres com gestação normal.

Um dos indicadores da alteração do sistema imunológico é a proteína C-Reativa (PCR-us), cujo aumento da concentração no sangue está associada com inflamação. Alterações na concentração da PCR-us durante a gestação estão associadas com diversas complicações, como pré-eclâmpsia, comprometimento cognitivo grave, restrição de crescimento intrauterino (RCIU), prematuridade, baixo peso ao nascer (BPN) e tamanho pequeno para a idade gestacional (PIG).

Crianças PIG apresentam maior risco para complicações no período neonatal, distúrbios durante a puberdade, e uma maior mortalidade na vida adulta. Portanto a prevenção de nascimentos PIG é de suma importância para a saúde pública.

Outro indicador importante de tamanho no recém-nascido é o índice ponderal, definido como peso sobre comprimento ao cubo.

2. Objetivo

O objetivo do estudo é avaliar a relação das concentrações da PCR-us no sangue materno e no cordão umbilical do recém-nascido com o tamanho do recém-nascido e outras variáveis do recém-nascido.

3. Descrição do estudo

As gestantes foram selecionadas de 37 Unidades Básicas de Saúde no município de Araraquara e arredores.

As participantes selecionadas responderam a questionários padronizados previamente testados em gestantes (Rondó et al, 2003), que consistiram em avaliação de fatores demográficos e socioeconômicos, antropométricos, de hábitos de vida, obstétricos e morbidade no terceiro período de gestação. Também foram determinadas as concentrações sanguíneas PCR-us no mesmo período da gestação e no sangue do cordão umbilical.

Foram excluídas deste estudo mulheres com gestações múltiplas, síndromes hipertensivas da gestação, diabetes mellitus, doenças infecciosas pelo grupo TORCH (toxoplasmose, rubéola, citomegalovírus e herpes), sífilis, varicela e crianças com doenças congênitas e infecciosas.

Foram retiradas 15 das 1108 observações coletadas por dados faltantes/inconsistentes.

4. Descrição das variáveis

As variáveis que caracterizam a amostra são:

- **Idade materna** – anos
- **Cor da pele** – (branca, preta, amarela, indígena, parda)
- **Estado civil** – (solteira, solteira com acompanhante, casada, separada/viúva)
- **Renda per capita** – da família (reais)
- **Número de pessoas no domicílio**
- **Anos de estudo**
- **Fumante na gestação** – (sim, não)
- **Ingestão de bebida alcoólica na gestação** – (sim, não)
- **Número de partos anteriores**
- **Estatura materna** – centímetros

- **Circunferência do pescoço materna** – centímetros
- **Circunferência do braço materna** – centímetros
- **Peso materno** – quilogramas
- **Índice de massa corporal (IMC) materno** – kg/m^2
- **PCR-us materna** – mg/dL
- **PCR-us no cordão umbilical** – mg/dL
- **PCR-us maternal** – (alto, baixo). Se o PCR-us materno for menor ou igual à 5, ele é considerado baixo; e alto, caso contrário.
- **PCR-us no cordão umbilical do bebê** – (alto, baixo). Se o PCR-us no cordão umbilical for menor ou igual à 3, ele é considerado baixo; e alto, caso contrário.
- **Sexo do recém-nascido** – (masculino, feminino)
- **Idade gestacional** – dias
- **Peso ao nascer** – gramas
- **Comprimento ao nascer** – centímetros
- **Perímetro encefálico ao nascer** – centímetros
- **Prematuridade** – (normal, prematuro)
- **Peso para idade gestacional** – (PIG, AIG, GIG)
- **Índice ponderal** - g/cm^3

5. Análise descritiva

A Tabela A.1 apresenta as estatísticas descritivas das variáveis contínuas do estudo, incluindo média, mínimo, máximo e desvio-padrão (DP). Por exemplo, a idade materna variou de 13,76 a 44,69 anos (média=27,14; DP=6,46), o IMC materno oscilou entre 15,96 e 52,45 kg/m^2 (média=29,25; DP=5,46), e a PCR-us materna variou de 0,10 a 95,70 mg/dL (média=7,78; DP=8,01). Esses valores mostram ampla variabilidade nas condições maternas e níveis inflamatórios, fornecendo uma base para as análises bivariadas subsequentes. Vemos uma inconsistência no valor de renda per capita, com

algumas observações tendo valores extremamente baixos e discrepantes (como 0), porém, visto que esta variável não é de suma importância para o estudo, estas observações não serão retiradas. O tamanho da amostra é o mesmo para todas as variáveis, pois foram mantidas somente as gestantes que tinham resposta para todas as variáveis de interesse.

Na Tabela A.2, destacam-se pares com coeficientes de correlação de Pearson absolutas acima de 0,3. Observamos, por exemplo, circunferência do pescoço e peso materno ($r=0,88$), IMC e peso materno ($r=0,87$) e circunferência do braço e circunferência do pescoço materno ($r=0,76$). Vemos que as variáveis morfológicas da mãe são altamente correlacionadas entre si, tanto quanto as do bebê. Conseguimos também observar uma relação entre a renda per capita e os anos de estudo, e renda per capita com número de pessoas no domicílio.

A Tabela A.3 mostra que entre os 31 recém-nascidos prematuros, 19,4% foram classificados como PIG, enquanto esse percentual cai a 11% nos casos a termo. Apesar da grande diferença, a mesma não é significativa (teste exato de Fisher, valor- $p = 0,131$). No *box plot* da PCR-us materna por prematuridade (Figura B.1), observa-se mediana ligeiramente superior e IQR mais amplo nas gestantes de prematuros, indicando maior variabilidade inflamatória neste grupo. Na Figura B.1 foram incluídos todos os bebês, não somente os PIG.

O *box plot* da idade materna por peso para idade gestacional (Figura B.2) revela medianas muito semelhantes para as categorias AIG e GIG, enquanto PIG tende a ter idade levemente menor, mas com ampla sobreposição dos IQRs, sugerindo homogeneidade de faixa etária independente do desfecho de crescimento. Vemos na Figura B.3 que os anos de estudo não parecem ter grande influência no tamanho do bebê. Na Figura B.4, o IMC materno apresenta mediana mais baixa na categoria PIG e deslocamento do IQR para baixo, contrastando com GIG, cujo IQR está deslocado para

valores mais altos, refletindo maior prevalência de sobrepeso/obesidade. Vemos padrões similares nas Figuras B.5 a B.8, com graus diferentes de intensidade.

Vemos na Figura B.9 que o número de pessoas no domicílio é maior para bebês GIG, porém a Figura B.10, sobre a renda per capita, não aparenta mostrar diferenças entre as categorias de peso para idade gestacional. A Figura B.11 mostra que o número mediano de partos anteriores é muito próximo entre as categorias AIG e GIG, mas a categoria PIG tem mediana igual a 0, indicando que mulheres na primeira gestação possam ter um risco maior de bebês pequenos para idade gestacional.

Vemos no *box plot* da PCR-us materna por peso para idade gestacional (Figura B.12) que as medianas e intervalos interquartis são parecidos em todas as categorias, isto é, a PCR-us materna não aparenta ter influência no risco do bebê ser PIG.

As Tabelas A.4 a A.9 apresentam as distribuições de frequências conjuntas entre a variável peso para idade gestacional e algumas variáveis qualitativas. Por meio da aplicação de testes de hipóteses, não foi possível observar associação entre a variável peso para idade gestacional e as variáveis fumante na gestação (teste exato de Fisher, valor-p = 0,106), ingestão de bebida alcoólica durante a gestação (teste qui-quadrado, valor-p = 0,631), cor da pele (teste exato de Fisher, valor-p = 0,117), escolaridade (teste exato de Fisher, valor-p = 0,107), estado civil (teste exato de Fisher, valor-p = 0,726) e sexo do recém-nascido (teste qui-quadrado, valor-p = 0,097).

Vemos no *box plot* da PCR-us no cordão umbilical por peso para idade gestacional (Figura B.13) que, em todas as categorias, o intervalo interquartil é uma barra centrada perto de 0. Isto se deve ao fato de a vasta maioria das observações ter valor pequeno nesta variável (vide Figura B.14). De fato, 886 observações, isto é, 80,96% da amostra, têm valor da PCR-us no cordão umbilical igual a 0,1. Para facilitar a visualização de um possível efeito desta variável no peso para idade gestacional, criamos a variável PCR-us no cordão umbilical do bebê, após discussão com as

pesquisadoras sobre valores limítrofes adequados, decidindo-se por 3 (Villacorta, Humberto et al., 2007). Embora a Tabela A.10 mostre uma possível associação entre as variáveis peso para idade gestacional e PCR-us no cordão umbilical do bebê, o teste de exato de Fisher não foi significativo (valor-p = 0.080). De maneira similar, criamos a variável PCR-us maternal, a pedido das pesquisadoras. Vemos na Tabela A.11 que não há evidência de associação entre as variáveis peso para idade gestacional e PCR-us maternal (valor-p do teste qui-quadrado igual a 0,413).

Nas Figuras B.16 a B.31, temos uma repetição da análise acima para a variável índice ponderal. No geral, não vemos muita correlação entre as outras variáveis e o índice ponderal. Vemos uma correlação baixa entre as variáveis morfológicas da mulher relacionadas com o peso e o índice ponderal, com este aumentando à medida que as variáveis relacionadas ao peso aumentam (Figuras B.18 a B.21). Vemos também que as variáveis relacionadas ao PCR-us não aparentam ter relação com o índice ponderal (Figuras B.26, B.31 e B.32).

6. Análise inferencial

A metodologia apresentada no Apêndice C foi utilizada para ajustar modelos de regressão para as variáveis respostas índice ponderal e peso para idade gestacional (sem e com o procedimento SMOTE) em função de várias variáveis explicativas.

6.1 Índice ponderal

A Tabela A.11 apresenta os resultados do ajuste do modelo para o índice ponderal. Vemos que o sinal da estimativa do parâmetro atrelado à circunferência do pescoço materna e ao peso materno são diferentes, o que não era esperado, já que estas variáveis são correlacionadas positivamente (Tabela A.2). A Tabela A.12 nos mostra um fator generalizado de inflação da variância (GVIF) relativamente alto para o peso materno e para a circunferência do pescoço materna. Ajustando modelos onde não incluímos as duas variáveis mencionadas, simultaneamente, observamos que

(Tabelas A.13 e A.14) as estimativas dos parâmetros associados a estas variáveis têm o mesmo sinal, indicando que a multicolinearidade estava afetando as estimativas dos efeitos de ambas as variáveis. Vemos nas Figuras B.33 a B.36 que os modelos estão relativamente bem ajustados.

Pela Tabela A.13 vemos que apenas os efeitos de cor da pele indígena e peso materno são significantes a 5%. A estimativa do parâmetro associado a cor da pele indígena tem valor -0.193, isto é, fixados os valores das demais variáveis explicativas, uma mulher com cor de pele indígena tem valor esperado do índice ponderal igual a 0,824 vezes o valor esperado de uma mulher com cor de pele branca. Devemos considerar esse efeito com cautela, visto que temos apenas 6 mulheres indígenas na amostra. A estimativa do parâmetro associado ao peso materno é igual a 0,0005; isto é, fixados os valores das demais variáveis explicativas, o valor esperado do índice ponderal aumenta 0,05% a cada aumento de um quilo no peso materno.

Pela Tabela A.14, vemos que apenas os efeitos de cor da pele indígena e primigesta são significantes a 5%. Fixados os valores das demais variáveis explicativas, uma mulher com cor de pele indígena tem valor esperado do índice ponderal igual a 0,822 vezes o valor esperado de uma mulher com cor de pele branca e uma mulher primigesta tem valor esperado do índice ponderal igual a 0,985 vezes o valor esperado de uma mulher que não é primigesta.

6.2 Peso para a idade gestacional dicotomizada – sem SMOTE

Para a variável resposta peso para a idade gestacional dicotomizada (sem SMOTE), inicialmente foi ajustado um modelo em que a variável cor foi selecionada pelo método GAIC. Entretanto, os erros padrões das estimativas dos parâmetros associados a esta variável eram enormes e, portanto, essa variável foi retirada do modelo. A Tabela A.15 apresenta os resultados do ajuste do modelo logístico sem a variável cor. Vemos que os sinais das estimativas dos parâmetros atrelados à

circunferência do pescoço materna e à circunferência do braço materna são diferentes do associado ao peso materno, o que não era esperado, já que estas variáveis são correlacionadas positivamente (Tabela A.2). A Tabela A.16 nos mostra um fator generalizado de inflação da variância (GVIF) relativamente alto para o peso materno, para a circunferência do pescoço materna e para a circunferência do braço materna. Ajustando modelos onde incluímos apenas uma das três variáveis de cada vez, observamos que (Tabelas A.17 a A.19) as estimativas dos parâmetros associados a estas variáveis têm o mesmo sinal, indicando que a multicolinearidade estava afetando as estimativas dos efeitos das variáveis. Vemos nas Figuras B.37 a B.42 que as suposições sobre os resíduos dos três modelos não são violadas.

Pela Tabela A.17, vemos que as variáveis peso materno e primigesta têm efeitos significativos a 5%. Fixados os valores das demais variáveis explicativas

- a chance de um bebê nascer PIG diminui 3,8% a cada aumento de um quilo no peso materno
- a chance de um bebê nascer PIG aumenta 58,4% quando a mãe é primigesta.

Pela Tabela A.18, vemos que apenas a variável primigesta tem efeito significativo a 5%. Fixados os valores das demais variáveis explicativas, a chance de um bebê nascer PIG aumenta 67,9% quando a mãe é primigesta.

Pela Tabela A.19, vemos que as variáveis circunferência do pescoço materna e primigesta têm efeitos significativos a 5%. Fixados os valores das demais variáveis explicativas

- a chance de um bebê nascer PIG diminui 6,8% a cada aumento de um centímetro na circunferência do pescoço materna
- a chance de um bebê nascer PIG aumenta 62,1% quando a mãe é primigesta.

6.3 Peso para a idade gestacional dicotimizada – com SMOTE

Ainda para a variável resposta peso para a idade gestacional dicotimizada, vamos repetir a análise, agora aplicando o procedimento SMOTE. Ajustamos os mesmos modelos da Seção 6.2, cujos resultados se encontram nas Tabelas A.20 a A.24. Como observamos o mesmo padrão de multicolinearidade, ajustamos três modelos incluindo em cada um deles apenas uma das variáveis, peso materno, circunferência do braço materna e circunferência do pescoço materna. As Figuras B.43 a B.48 mostram que os resíduos não violam as suposições e os modelos parecem bem ajustados.

Pela Tabela A.22, vemos que as variáveis idade materna, peso materno e primigesta têm efeitos significativos a 5%. Fixados os valores das demais variáveis explicativas

- a chance de um bebê nascer PIG aumenta 3,6% a cada aumento de um ano na idade da mãe
- a chance de um bebê nascer PIG diminui 3,9% a cada aumento de um quilo no peso materno
- a chance de um bebê nascer PIG aumenta 184% quando a mãe é primigesta.

Pela Tabela A.23, vemos que apenas as variáveis idade materna e primigesta têm efeito significativo a 5%. Fixados os valores das demais variáveis explicativas

- a chance de um bebê nascer PIG aumenta 3,1% a cada aumento de um ano na idade da mãe
- a chance de um bebê nascer PIG aumenta 199% quando a mãe é primigesta.

Pela Tabela A.24, vemos que as variáveis idade materna, circunferência do pescoço materna e primigesta têm efeitos significativos a 5%. Fixados os valores das demais variáveis explicativas

- a chance de um bebê nascer PIG aumenta 3,3% a cada aumento de um ano na idade da mãe
- a chance de um bebê nascer PIG diminui 8,2% a cada aumento de um centímetro na circunferência do pescoço materna
- a chance de um bebê nascer PIG aumenta 192% quando a mãe é primigesta.

A Tabela A.25 mostra que os modelos logísticos sem SMOTE apresentam alta especificidade (até 100%), mas sensibilidade muito baixa (1,6%), indicando viés para a classe majoritária. Com a aplicação do SMOTE, há um ganho expressivo de sensibilidade — especialmente no modelo com peso materno (de 1,6% para 20,7%) — com reduções moderadas de especificidade. Esse comportamento é esperado em contextos com forte desequilíbrio de classes, como no caso do peso para a idade gestacional, evidenciando que o SMOTE melhora a capacidade de identificar corretamente os casos da classe minoritária.

7. Conclusão

Concluimos que as variáveis PCR-us materna/no cordão umbilical não têm efeito significativo nem sobre o índice ponderal nem sobre o peso para a idade gestacional.

A média da variável resposta índice ponderal é maior quando o peso materno aumenta e é menor quando a mãe tem cor da pele indígena e é primigesta.

Além disso, vemos que quanto maior o peso materno e a circunferência do pescoço materna, menor a chance do bebê ser PIG; mulheres primigestas têm uma chance maior de ter um bebê PIG. A idade materna se mostrou significativa apenas no modelo com SMOTE, sendo que quanto maior é a idade materna, maior a chance do bebê ser PIG.

APÊNDICE A

Tabelas

Tabela A.1 Medidas resumo das variáveis quantitativas

Variável	Média	Mínimo	Máximo	Desvio padrão
Idade materna	27,14	13,76	44,69	6,46
Renda per capita	825,37	0,00	5800,00	582,99
Número de pessoas no domicílio	3,34	1,00	9,00	1,33
Número de partos anteriores	1,06	0,00	11,00	1,29
Estatura materna	161,85	142,50	181,50	6,13
Circunferência do braço materna	33,21	23,00	44,50	2,32
Circunferência do pescoço materna	30,03	16,10	48,75	4,34
Peso materno	76,80	39,20	143,10	15,44
Índice de massa corporal materno	29,25	15,96	52,45	5,46
PCR-us materna	7,78	0,10	95,70	8,01
PCR-us no cordão umbilical	0,59	0,10	63,80	3,31
Idade gestacional	275,81	225,00	308,00	10,31
Peso ao nascer	3219,46	1435,00	4890,00	466,55
Comprimento ao nascer	48,95	39,00	55,00	2,23
Perímetro encefálico ao nascer	33,83	27,00	39,00	1,57
Anos de estudo	11,01	3,00	24,00	2,41
Índice ponderal	2,73	1,68	4,66	0,31

Tabela A.2 Pares de variáveis com valor absoluto do coeficiente de correlação de Pearson maior que 0,30

Variável 1	Variável 2	Correlação
Circunferência do pescoço materna	Peso materno	0,88
Índice de massa corporal materno	Peso materno	0,87
Circunferência do pescoço materna	Índice de massa corporal materno	0,85
Circunferência do braço materna	Peso materno	0,76
Circunferência do braço materna	Circunferência do pescoço materna	0,76
Comprimento ao nascer	Peso materno	0,70
Circunferência do braço materna	Índice de massa corporal materno	0,69
Perímetro encefálico ao nascer	Peso ao nascer	0,68
Idade gestacional	Peso ao nascer	0,50
Comprimento ao nascer	Perímetro encefálico ao nascer	0,49
Comprimento ao nascer	Idade gestacional	0,47
Número de pessoas no domicílio	Renda per capita	-0,44
Idade materna	Número de partos anteriores	0,44
Idade gestacional	Perímetro encefálico ao nascer	0,38
Número de partos anteriores	Número de pessoas no domicílio	0,31
Índice ponderal	Peso ao nascer	0,48
Índice ponderal	Perímetro encefálico ao nascer	0,32
Renda per capita	Anos de estudo	0,46

Tabela A.3 Distribuição de frequências conjunta entre Prematuridade e Peso para idade gestacional.

Peso para idade gestacional	Prematuridade		Total
	normal	premature	
PIG	117 (11%)	6 (19,4%)	123 (11,3%)
AIG	863 (81,3%)	21 (67,7%)	884 (80,9%)
GIG	82 (7,7%)	4 (12,9%)	86 (7,9%)
Total	1.062 (100,0%)	31 (100,0%)	1.093 (100,0%)

Tabela A.4 Distribuição de frequências conjunta entre Fumante na gestação e Peso para idade gestacional.

Peso para idade gestacional	Fumante na gestação		Total
	não	sim	
PIG	111 (10,86%)	12 (16,22%)	123 (11,3%)
AIG	824 (80,91%)	60 (81,08%)	884 (80,9%)
GIG	84 (8,22%)	2 (2,70%)	86 (7,9%)
Total	1.019 (100,0%)	74 (100,0%)	1.093 (100,0%)

Tabela A.5 Distribuição de frequências conjunta entre Ingestão de bebida alcoólica na gestação e Peso para idade gestacional.

Peso para idade gestacional	Ingestão de bebida alcoólica na gestação		Total
	não	sim	
PIG	96 (10,81%)	27 (13,04%)	123 (11,3%)
AIG	721 (81,47%)	163 (78,74%)	884 (80,9%)
GIG	69 (7,77%)	17 (8,21%)	86 (7,9%)
Total	886 (100%)	207 (100%)	1.093 (100%)

Tabela A.6 Distribuição de frequências conjunta entre Cor da pele e Peso para idade gestacional.

Peso para idade gestacional	Cor da pele					Total
	branca	preta	amarela	indígena	parda	
PIG	53 (10.8%)	18 (12.6%)	0 (0%)	2 (100%)	50 (11.1%)	123 (11,3%)
AIG	406 (82.6%)	115 (80.4%)	6 (100%)	0 (0%)	357 (79.3%)	884 (80,9%)
GIG	33 (6.7%)	10 (7%)	0 (0%)	0 (0%)	43 (9.6%)	86 (7,9%)
Total	492 (100%)	143 (100%)	6 (100%)	2 (100%)	450 (100%)	1.093 (100%)

Tabela A.7 Distribuição de frequências conjunta entre Estado civil e Peso para idade gestacional.

Peso para idade gestacional	Estado civil				Total
	casada	solteira	solteira com acompanhante	separada/ viúva	
PIG	41 (11.3%)	63 (10.5%)	16 (13.9%)	3 (17.6%)	123 (11,3%)
AIG	289 (79.8%)	494 (82.3%)	88 (77.4%)	13 (76.5%)	884 (80,9%)
GIG	32 (8.8%)	43 (7.2%)	10 (8.7%)	1 (5.9%)	86 (7,9%)
Total	362 (100%)	600 (100%)	114 (100%)	17 (100%)	1.093 (100%)

Tabela A.8 Distribuição de frequências conjunta entre Sexo do recém nascido e Peso para idade gestacional.

Peso para idade gestacional	Sexo do recém nascido		Total
	feminino	masculino	
PIG	62 (11.4%)	61 (11.1%)	123 (11,3%)
AIG	446 (82.5%)	438 (79.3%)	884 (80,9%)
GIG	33 (6.1%)	53 (9.6%)	86 (7,9%)
Total	541 (100%)	552 (100%)	1.093 (100%)

Tabela A.9 Distribuição de frequências conjunta entre PCR-us no cordão umbilical do bebê e Peso para idade gestacional.

Peso para idade gestacional	PCR-us no cordão umbilical do bebê		Total
	baixo	alto	
PIG	121 (11.4%)	2 (5.7%)	123 (11,3%)
AIG	857 (81%)	27 (77.7%)	884 (80,9%)
GIG	80 (7.6%)	6 (17.1%)	86 (7,9%)
Total	1.058 (100%)	35 (100%)	1.093 (100%)

Tabela A.10 Distribuição de frequências conjunta entre PCR-us maternal e Peso para idade gestacional.

Peso para idade gestacional	PCR-us maternal		Total
	baixo	alto	
PIG	59 (11.7%)	64 (10.9%)	123 (11,3%)
AIG	413 (81.6%)	472 (80.3%)	884 (80,9%)
GIG	24 (6.7%)	52 (8.8%)	86 (7,9%)
Total	496 (100%)	597 (100%)	1.093 (100%)

Tabela A.11 Estimativas dos parâmetros do modelo com intervalos de 95% de confiança para o Índice ponderal

Variável	Estimativa	e^{β}	Intervalo de confiança (95%) para e^{β}		Valor-p
			Limite inferior	Limite superior	
Intercepto	0,983	N/A	N/A	N/A	<0,001 *
Renda per capita	10^{-5}	1,00001	0,999998	1,00002	0,107
Fumante na gestação	0,020	1,020	0,993	1,046	0,139
Circunferência do pescoço materna	-0,004	0,995	0,992	0,9992	0,017 *
Peso materno	0,001	1,001	1,0005	1,002	0,001 *
Cor da pele – preta	-0,007	0,992	0,972	1,013	0,463
Cor da pele – amarela	0,050	1,051	0,967	1,153	0,227
Cor da pele – indígena	-0,188	0,827	0,712	0,962	0,014 *
Cor da pele – parda	0,002	1,002	0,988	1,016	0,786
Primigesta	-0,014	0,986	0,973	0,9996	0,040 *
log PCR-us materna	0,001	1,001	0,993	1,010	0,712
log PCR-us no cordão umbilical	-0,003	0,996	0,990	1,003	0,299

Tabela A.12 Fator de inflação de variância generalizada (GVIF) para as variáveis da Tabela A.11

Variável	GVIF
Renda per capita	1,11
Fumante na gestação	1,02
Circunferência do pescoço materna	4,51
Peso materno	4,45
Cor da pele	1,13
Primigesta	1,10
log PCR-us materna	1,20
log PCR-us no cordão umbilical	1,03

Tabela A.13 Estimativas dos parâmetros do modelo com intervalos de 95% de confiança para o Índice ponderal, considerando apenas o Peso materno

Variável	Estimativa	e^{β}	Intervalo de confiança (95%) para e^{β}		Valor-p
			Limite inferior	Limite superior	
Intercepto	0,941	N/A	N/A	N/A	<0,001 *
Renda per capita	$9 \cdot 10^{-4}$	1,0001	0,999998	1,00002	0,105
Fumante na gestação	0,018	1,018	0,992	1,045	0,171
Peso materno	$5 \cdot 10^{-4}$	1,0005	1,00008	1,0009	0,013 *
Cor da pele – preta	-0,007	0,993	0,972	1,013	0,476
Cor da pele – amarela	0,055	1,056	0,968	1,155	0,213
Cor da pele – indígena	-0,193	0,824	0,709	0,958	0,011 *
Cor da pele – parda	0,001	1,002	0,988	1,016	0,788
Primigesta - sim	-0,013	0,987	0,973	1,0005	0,054
log PCR-us materna	$2 \cdot 10^{-4}$	1,0003	0,992	1,008	0,838
log PCR-us no cordão umbilical	-0,004	0,995	0,989	1,002	0,258

Tabela A.14 Estimativas dos parâmetros do modelo com intervalos de 95% de confiança para o Índice ponderal, considerando apenas a Circunferência do pescoço materna

Variável	Estimativa	e^{β}	Intervalo de confiança (95%) para e^{β}		Valor-p
			Limite inferior	Limite superior	
Intercepto	0,958	N/A	N/A	N/A	<0,001 *
Renda per capita	10^{-5}	1,00001	0,9999991	1,00002	0,070
Fumante na gestação	0,019	1,019	0,993	1,045	0,151
Circunferência do pescoço materna	$6 \cdot 10^{-4}$	1,0006	0,9990	1,002	0,422
Cor da pele – preta	-0,006	0,993	0,973	1,014	0,531
Cor da pele – amarela	0,056	1,058	0,969	1,156	0,207
Cor da pele – indígena	-0,195	0,822	0,707	0,956	0,011 *
Cor da pele – parda	0,002	1,002	0,989	1,015	0,737
Primigesta - sim	-0,014	0,985	0,972	0,9993	0,040 *
log PCR-us materna	0,002	1,002	0,994	1,009	0,544
log PCR-us no cordão umbilical	-0,004	0,995	0,988	1,002	0,213

Tabela A.15 Estimativas dos parâmetros do modelo logístico com intervalos de 95% de confiança para as razões de chances, para o Peso para a idade gestacional - sem SMOTE

Variável	Estimativa	e^{β}	Intervalo de confiança (95%) para e^{β}		P-valor
			Limite inferior	Limite superior	
Intercepto	-4,445	N/A	N/A	N/A	0,111
Circunferência do braço materna	0,135	1,145	0,992	1,321	0,063
Fumante na gestação	0,626	1,871	0,948	3,693	0,070
Circunferência do pescoço materna	0,110	1,116	1,009	1,235	0.032 *
Peso materno	-0,082	0,921	0,892	0,950	<0,001 *
Primigesta	0,503	1,653	1,115	2,451	0,012 *
log PCR-us materna	0,031	1,032	0,814	1,308	0,793
log PCR-us no cordão umbilical do bebê	-0,181	0,834	0,644	1,080	0,169

Tabela A.16 Fator de inflação de variância generalizada (GVIF) para as variáveis da Tabela A.15

Variável	GVIF
Circunferência do braço materna	2,601
Fumante na gestação	1,033
Circunferência do pescoço materna	4,255
Peso materno	4,241
Primigesta	1,045
log PCR-us materna	1,145
log PCR-us no cordão umbilical do bebê	1,018

Tabela A.17 Estimativas dos parâmetros do modelo logístico com intervalos de 95% de confiança para as razões de chances, para o Peso para a idade gestacional, considerando apenas o Peso materno- sem SMOTE

Variável	Estimativa	e^{β}	Intervalo de confiança (95%) para e^{β}		P-valor
			Limite inferior	Limite superior	
Intercepto	-0,064	N/A	N/A	N/A	0,918
Fumante na gestação	0,648	1,911	0,972	3,759	0,060
Peso materno	-0,037	0,962	0,947	0,977	<0,001 *
Primigesta	0,460	1,584	1,073	2,339	0,020 *
log PCR-us materna	0,100	1,105	0,877	1,393	0,396
log PCR-us no cordão umbilical do bebê	-0,179	0,835	0,646	1,080	0,171

Tabela A.18 Estimativas dos parâmetros do modelo logístico com intervalos de 95% de confiança para as razões de chances, para o Peso para a idade gestacional, considerando apenas a Circunferência do braço materna - sem SMOTE

Variável	Estimativa	e^{β}	Intervalo de confiança (95%) para e^{β}		P-valor
			Limite inferior	Limite superior	
Intercepto	0,218	N/A	N/A	N/A	0,881
Fumante na gestação	0,617	1,854	0,955	3,600	0,068
Circunferência do braço materna	-0,087	0,916	0,838	1,000	0,052
Primigesta	0,518	1,679	1,142	2,468	0,008 *
log PCR-us materna	-0,025	0,974	0,774	1,226	0,827
log PCR-us no cordão umbilical do bebê	-0,174	0,839	0,652	1,081	0,176

Tabela A.19 Estimativas dos parâmetros do modelo logístico com intervalos de 95% de confiança para as razões de chances, para o Peso para a idade gestacional, considerando apenas a Circunferência do pescoço materna - sem SMOTE

Variável	Estimativa	e^{β}	Intervalo de confiança (95%) para e^{β}		P-valor
			Limite inferior	Limite superior	
Intercepto	-0,657	N/A	N/A	N/A	0,392
Fumante na gestação	0,641	1,899	0,975	3,701	0,059
Circunferência do pescoço materna	-0,070	0,932	0,886	0,979	0,005 *
Primigesta	0,483	1,621	1,101	2,388	0,014 *
log PCR-us materna	0,027	1,027	0,812	1,299	0,819
log PCR-us no cordão umbilical do bebê	-0,170	0,843	0,654	1,086	0,188

Tabela A.20 Estimativas dos parâmetros do modelo logístico com intervalos de 95% de confiança para as razões de chances, para o Peso para a idade gestacional - com SMOTE

Variável	Estimativa	e^{β}	Intervalo de confiança (95%) para e^{β}		P-valor
			Limite inferior	Limite superior	
Intercepto	-4,767	N/A	N/A	N/A	0,002 *
Idade materna	0,038	1,039	1,009	1,070	0,010 *
Circunferência do braço materna	0,190	1,210	1,066	1,373	0,003 *
Peso materno	-0,091	0,912	0,886	0,939	<0,001 *
Renda per capita)	$-2 \cdot 10^{-4}$	0,9997	0,9993	1,00006	0,108
Circunferência do pescoço materna	0,097	1,102	1,009	1,204	0,030 *
Primigesta	1,107	3,027	2,072	4,423	<0,001 *
log PCR-us materna	0,115	1,122	0,912	1,381	0,274
log PCR-us no cordão umbilical do bebê	-0,068	0,933	0,758	1,150	0,518

Tabela A.21 Fator de inflação de variância generalizada (GVIF) para as variáveis da Tabela A.20

Variável	GVIF
Circunferência do braço materna	2,934
Circunferência do pescoço materna	4,662
Peso materno	4,860
Primigesta	1,387
log PCR-us materna	1,185
log PCR-us no cordão umbilical do bebê	1,021
Renda per capita	1,149
Idade materna	1,387

Tabela A.22 Estimativas dos parâmetros do modelo logístico com intervalos de 95% de confiança para as razões de chances, para o Peso para a idade gestacional, considerando apenas o Peso materno- com SMOTE

Variável	Estimativa	e^{β}	Intervalo de confiança (95%) para e^{β}		P-valor
			Limite inferior	Limite superior	
Intercepto	0,516	N/A	N/A	N/A	0,410
Idade materna	0,039	1,040	1,010	1,071	0,008 *
Peso materno	-0,042	0,958	0,946	0,971	<0,001 *
Renda per capita)	$-3 \cdot 10^{-4}$	0,9996	0,9993	1,00002	0,072
Primigesta - sim	1,099	3,003	2,068	4,360	<0,001 *
log PCR-us materna	0,180	1,197	0,978	1,465	0,081
log PCR-us no cordão umbilical do bebê	-0,050	0,950	0,775	1,166	0,627

Tabela A.23 Estimativas dos parâmetros do modelo logístico com intervalos de 95% de confiança para as razões de chances, para o Peso para a idade gestacional, considerando apenas a Circunferência do braço materna - com SMOTE

Variável	Estimativa	e^{β}	Intervalo de confiança (95%) para e^{β}		P-valor
			Limite inferior	Limite superior	
Intercepto	1,058	N/A	N/A	N/A	0,406
Idade materna	0,030	1,031	1,002	1,061	0,035 *
Circunferência do braço materna	-0,095	0,909	0,843	0,980	0,013 *
Renda per capita	$-3 \cdot 10^{-4}$	0,9996	0,9992	0,99998	0,042 *
Primigesta	1,094	2,986	2,075	4,297	<0,001 *
log PCR-us materna	0,032	1,032	0,850	1,253	0,746
log PCR-us no cordão umbilical do bebê	-0,054	0,946	0,775	1,154	0,587

Tabela A.24 Estimativas dos parâmetros do modelo logístico com intervalos de 95% de confiança para as razões de chances, para o Peso para a idade gestacional, considerando apenas a Circunferência do pescoço materna - com SMOTE

Variável	Estimativa	e^{β}	Intervalo de confiança (95%) para e^{β}		P-valor
			Limite inferior	Limite superior	
Intercepto	0,253	N/A	N/A	N/A	0,726
Idade materna	0,033	1,033	1,004	1,064	0,023 *
Circunferência do pescoço materna	-0,085	0,918	0,884	0,957	<0,001 *
Renda per capita	$-3 \cdot 10^{-4}$	0,999	0,9993	1,000001	0,051
Primigesta	1,070	2,917	2,022	4,209	<0,001 *
log PCR-us materna	0,108	1,114	0,911	1,361	0,290
log PCR-us no cordão umbilical do bebê	-0,048	0,952	0,779	1,163	0,635

Tabela A.25 Sensitividade e especificidade para os modelos finais com e sem SMOTE.

Variável	Modelo	Especificidade	Sensitividade
Peso materno	Sem SMOTE	99,9%	1,6%
	Com SMOTE	90,5%	20,7%
Circunferência do braço materna	Sem SMOTE	100%	1,6%
	Com SMOTE	97,6%	3,3%
Circunferência do pescoço materna	Sem SMOTE	100%	1,6%
	Com SMOTE	95,6%	9,3%

APÊNDICE B

Figuras

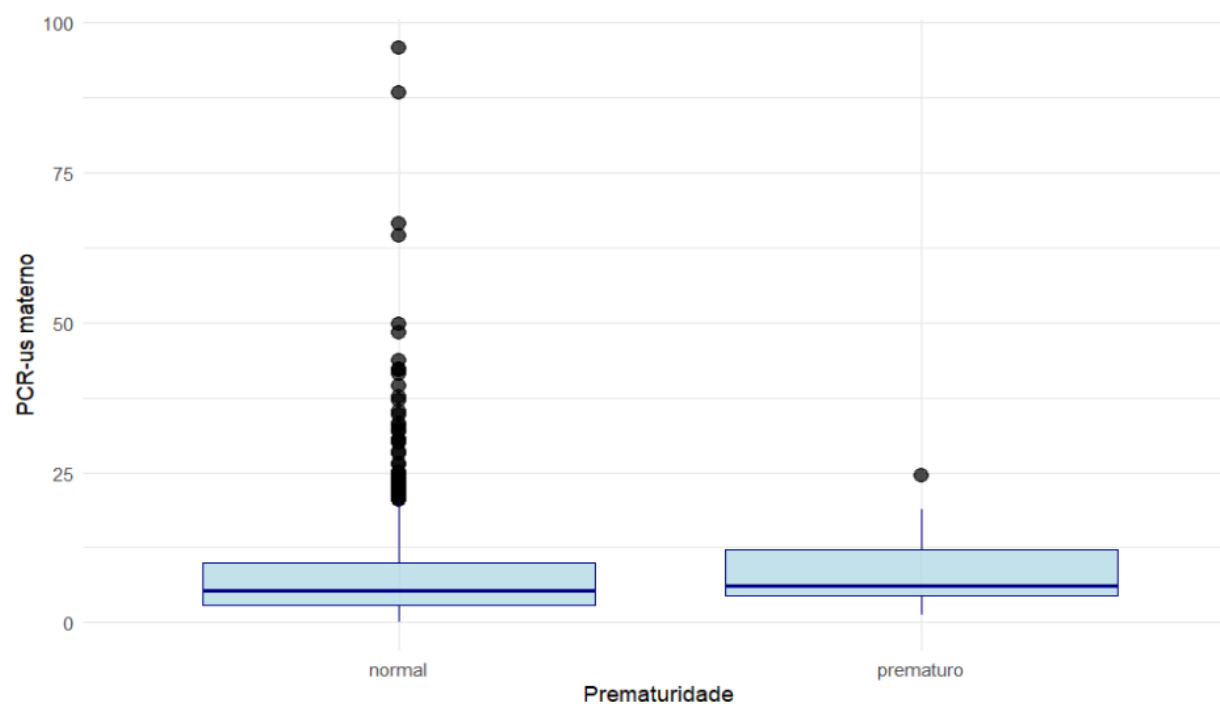


Figura B.1 *Box plot* do PCR-us materno por Prematuridade

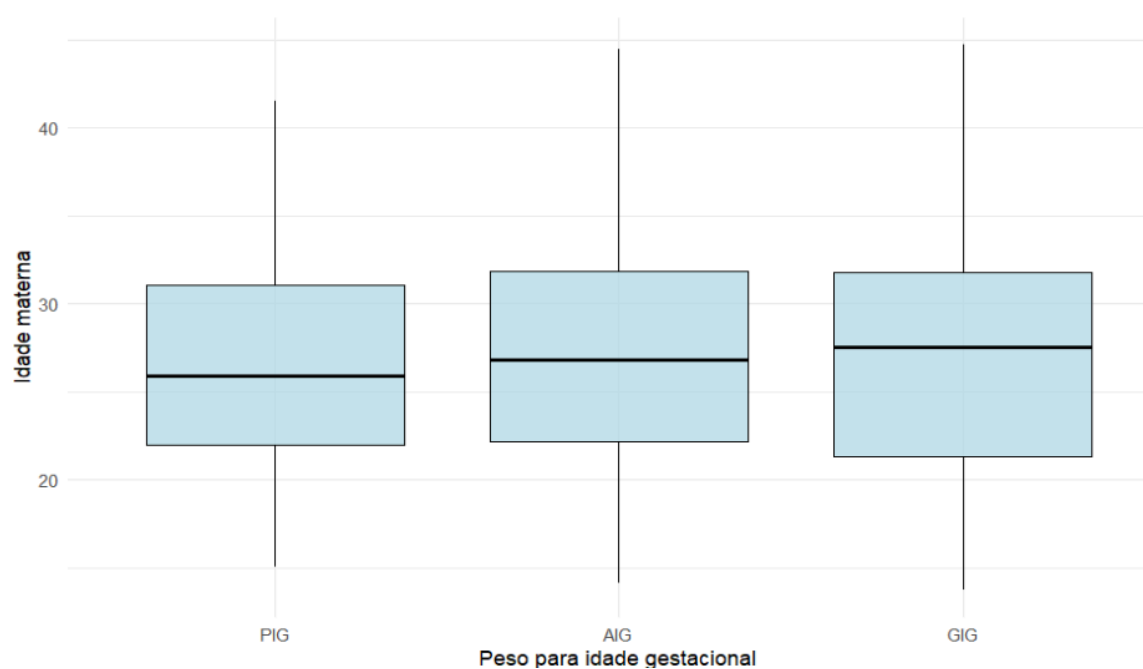


Figura B.2 *Box plot* da Idade materna por Peso para idade gestacional

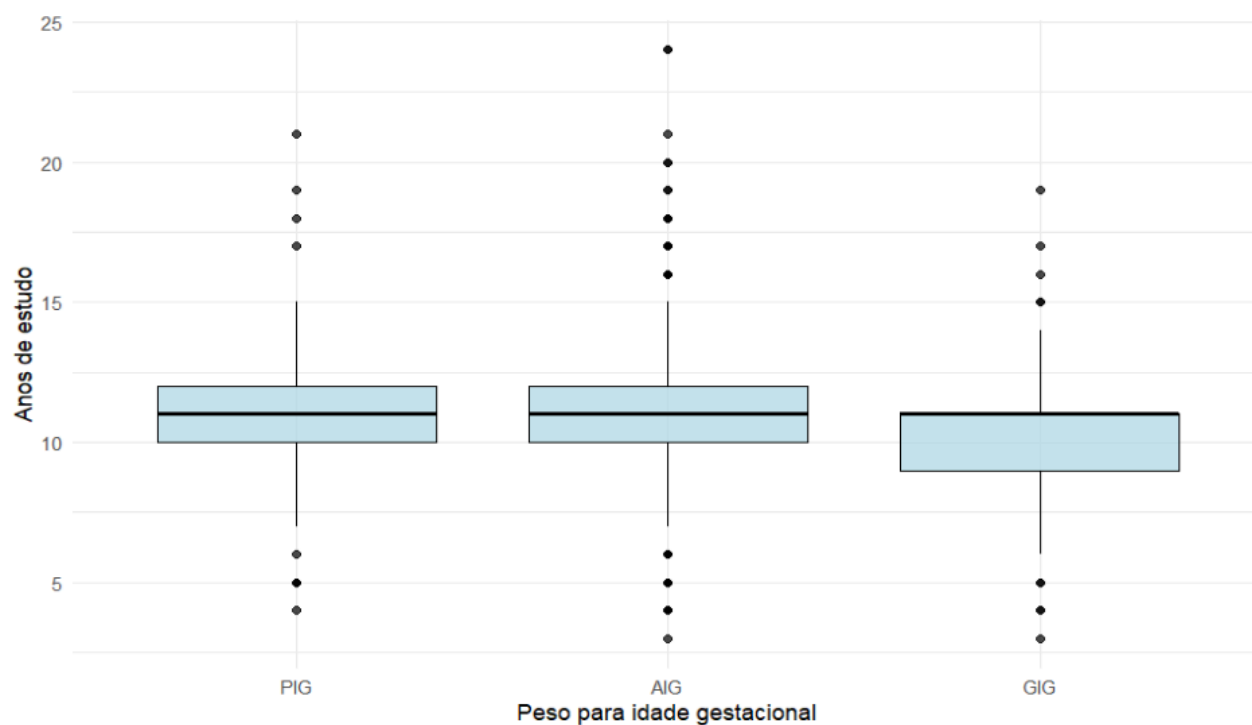


Figura B.3 *Box plot* de Anos de estudo por Peso para idade gestacional

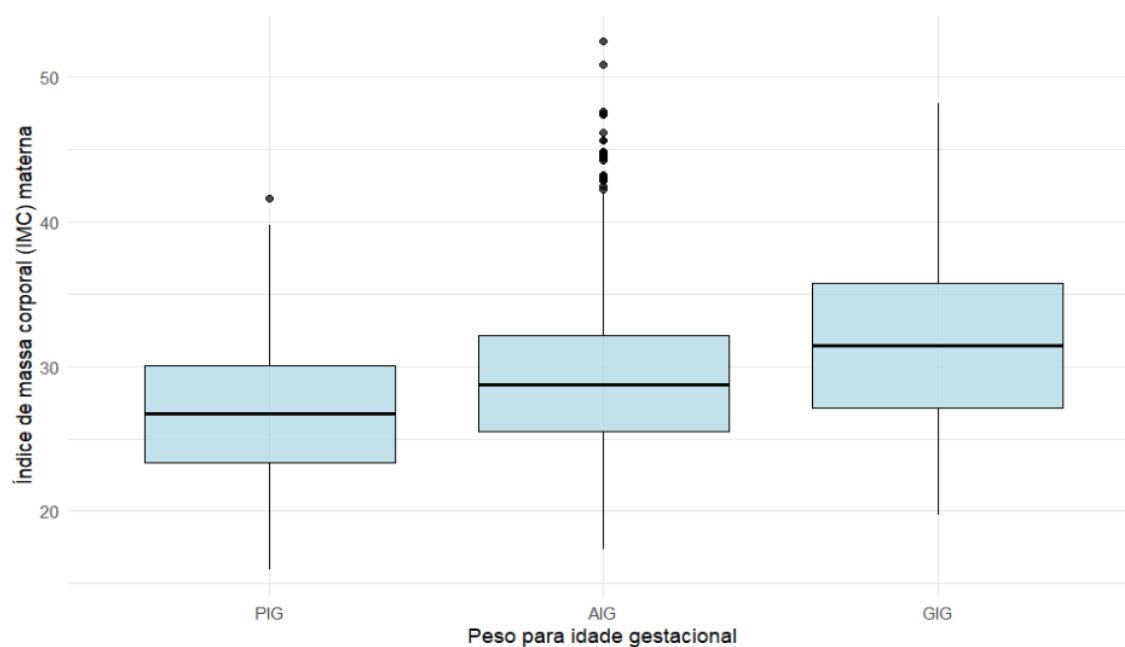


Figura B.4 *Box plot* do Índice de massa corporal (IMC) materna por Peso para idade gestacional

Box plot showing the distribution of maternal weight gain (Circunferência do pescoço materna) for three groups: FIG, AIG, and GIG. The y-axis ranges from 20 to 50. The x-axis is labeled 'Peso para idade gestacional'. The FIG group has a median around 29, the AIG group has a median around 30, and the GIG group has a median around 32. The AIG group shows several outliers above 40.

Figura B.6 *Box plot* da Circunferência do pescoço materna por Peso para idade gestacional

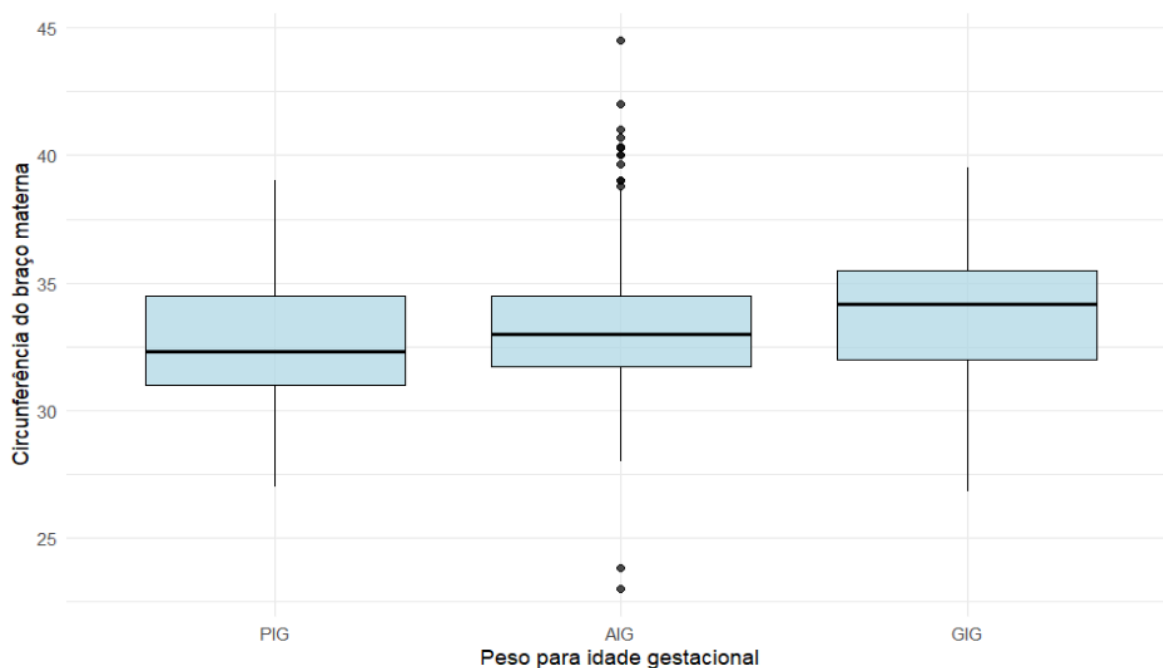


Figura B.7 *Box plot* da Circunferência do braço materna por Peso para idade gestacional

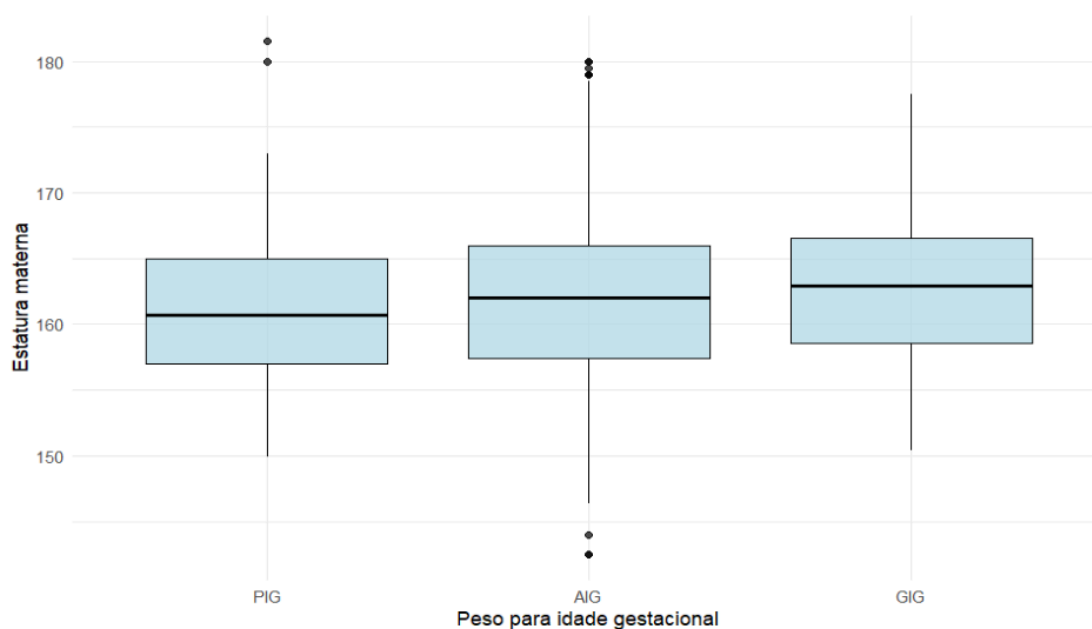


Figura B.8 *Box plot* da Estatura materna por Peso para idade gestacional

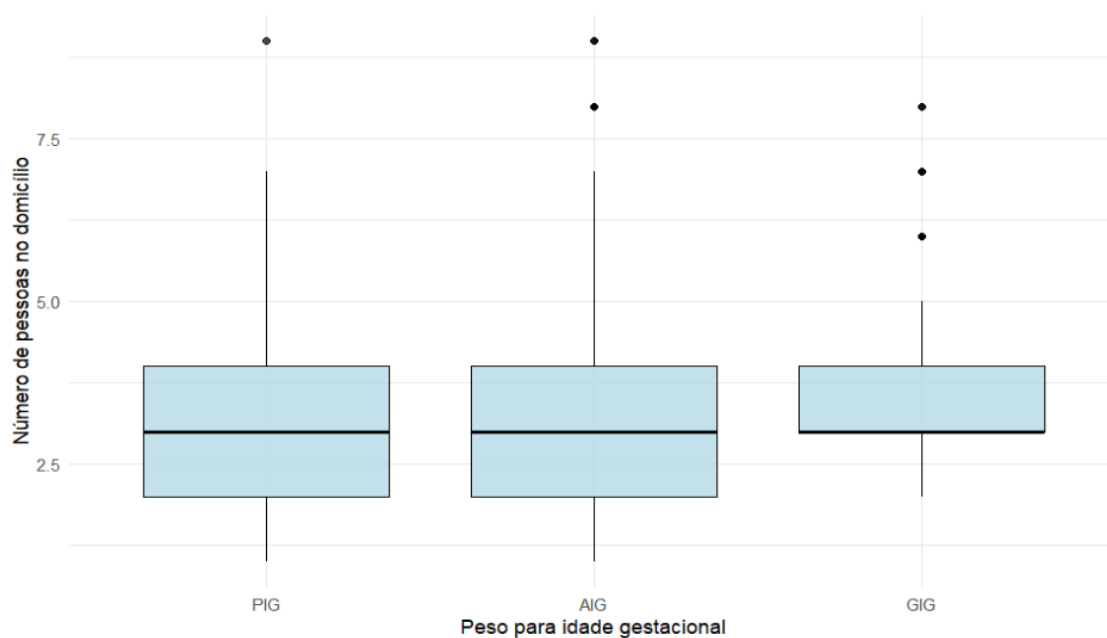


Figura B.9 *Box plot* do Número de pessoas no domicílio por Peso para idade gestacional

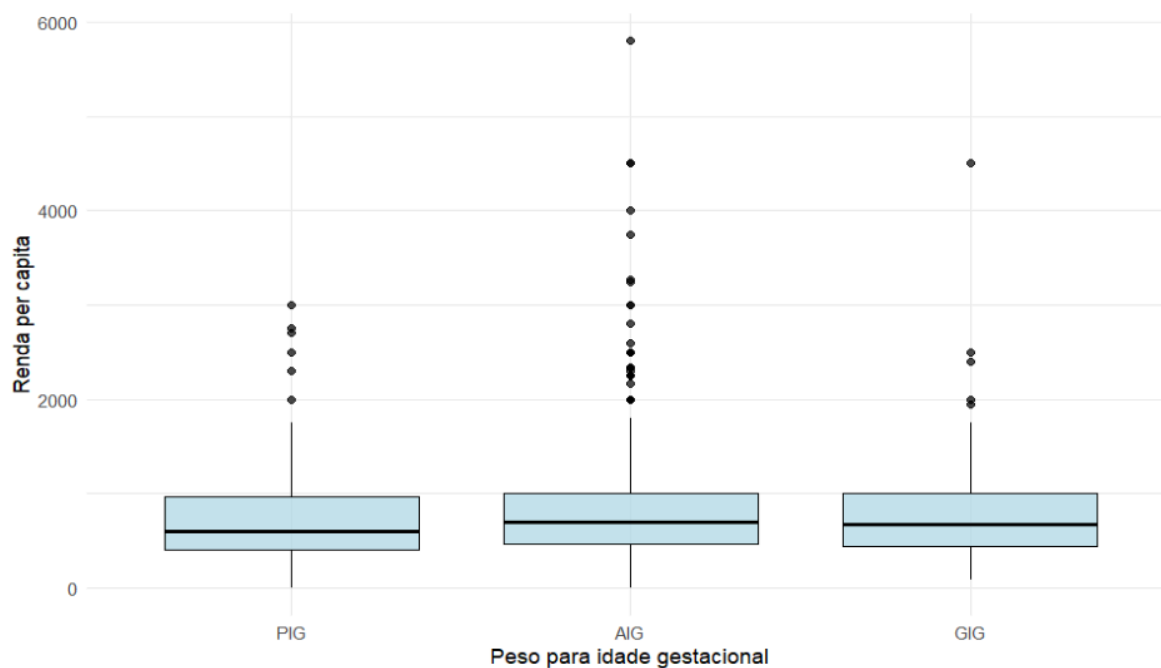


Figura B.10 *Box plot* da Renda per capita por Peso para idade gestacional

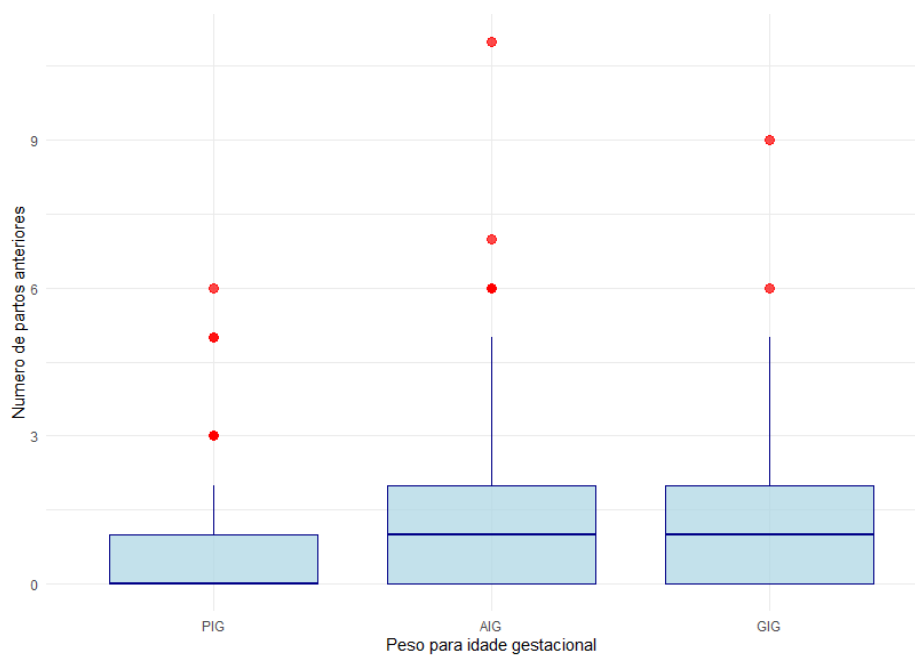


Figura B.11 *Box plot* do Número de partos anteriores por Peso para idade gestacional

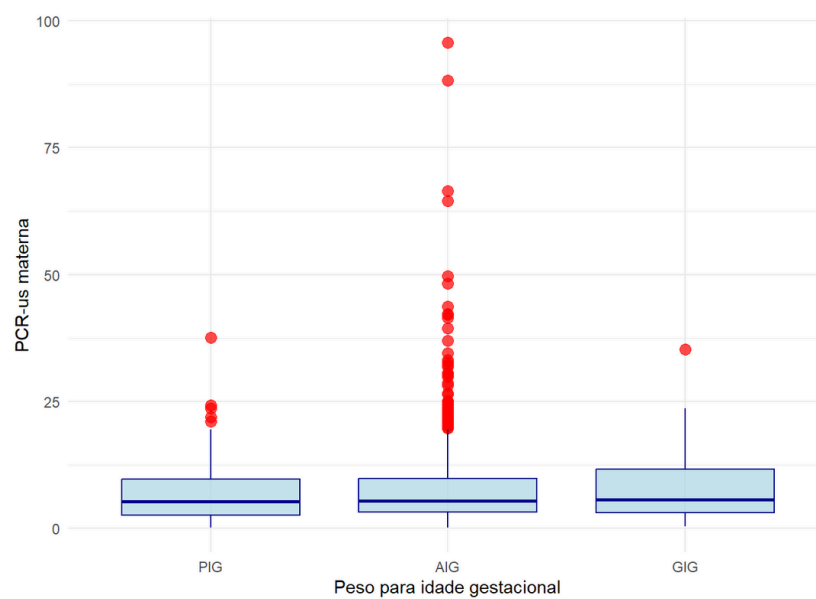


Figura B.12 *Box plot* da PCR-us materna por Peso para idade gestacional

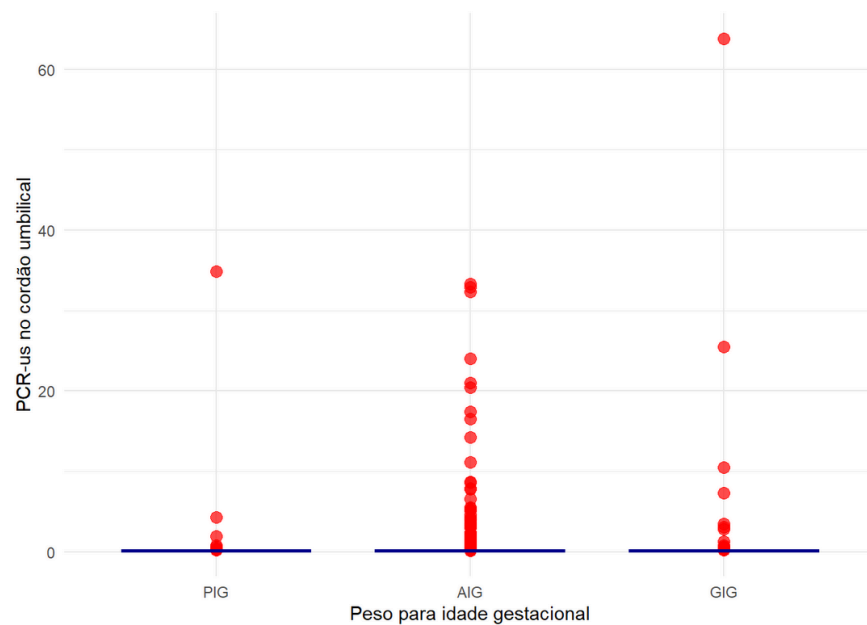


Figura B.13 *Box plot* da PCR-us no cordão umbilical por Peso para idade gestacional

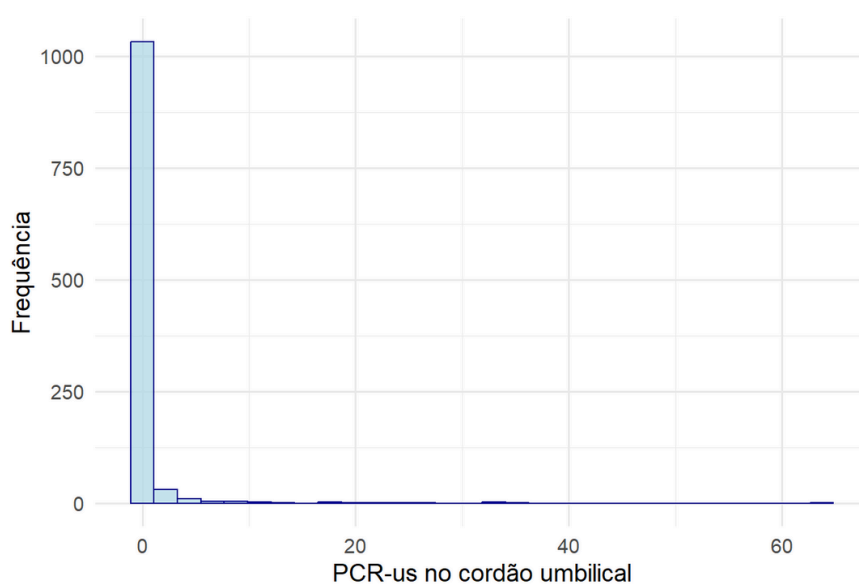


Figura B.14 Histograma da variável PCR-us no cordão umbilical.

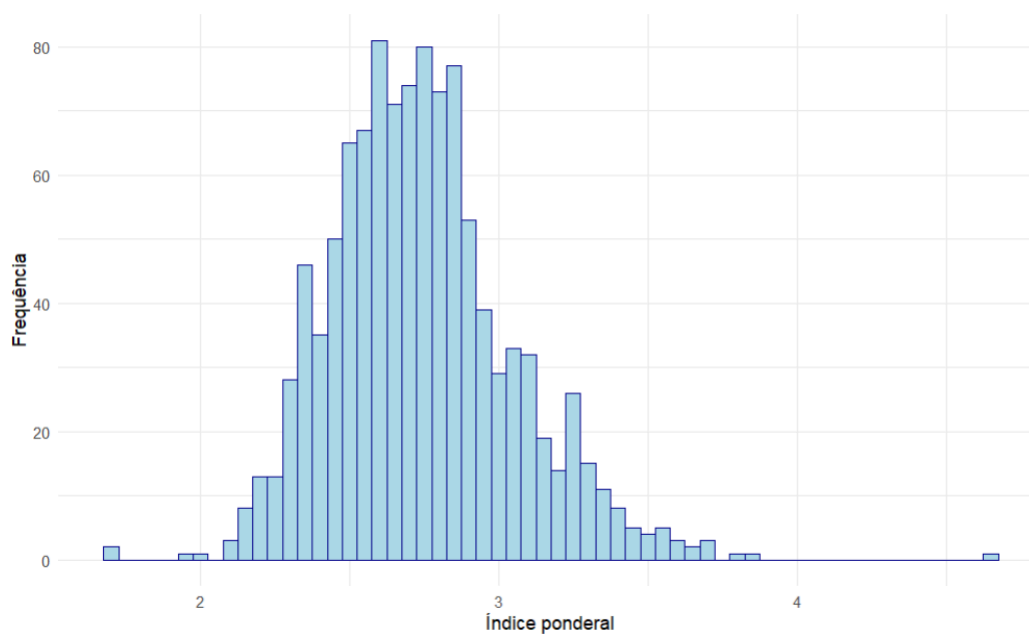


Figura B.15 Histograma da variável Índice Ponderal

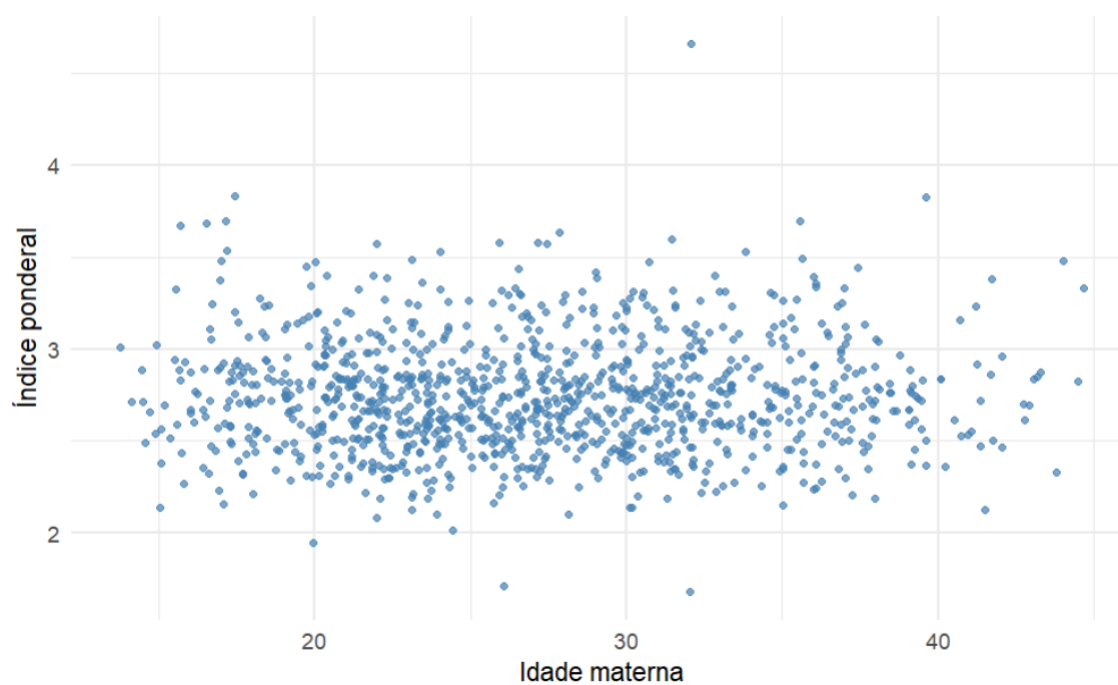


Figura B.16 Gráfico de dispersão da Idade materna pelo Índice ponderal

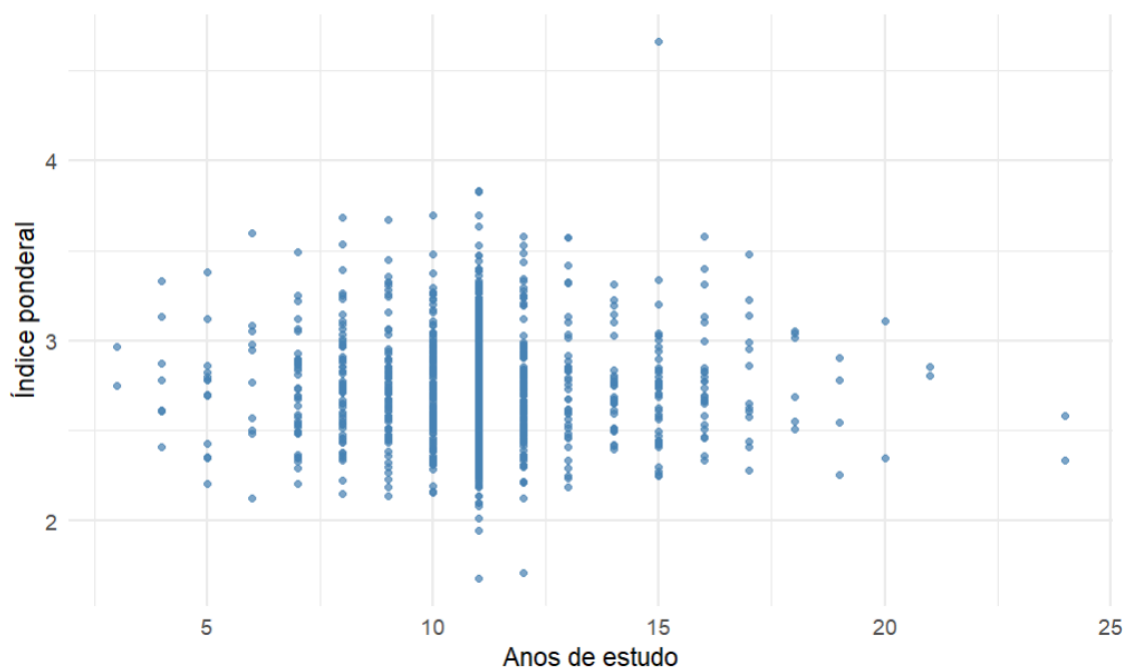


Figura B.17 Gráfico de dispersão dos Anos de estudo pelo Índice ponderal

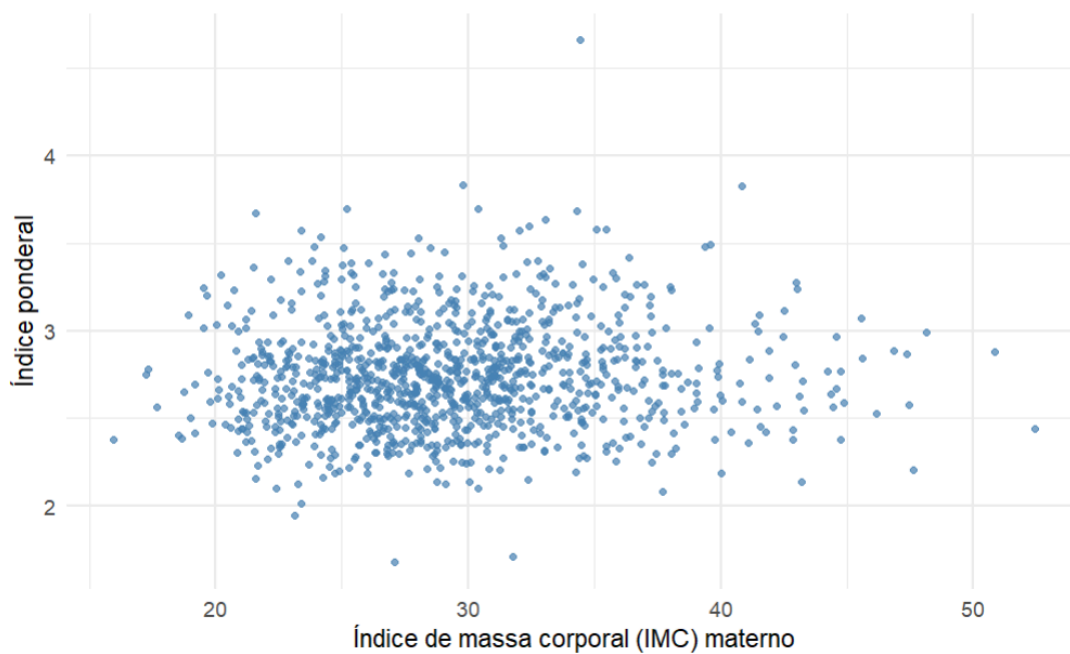


Figura B.18 Gráfico de dispersão do Índice de massa corporal (IMC) materno pelo Índice ponderal

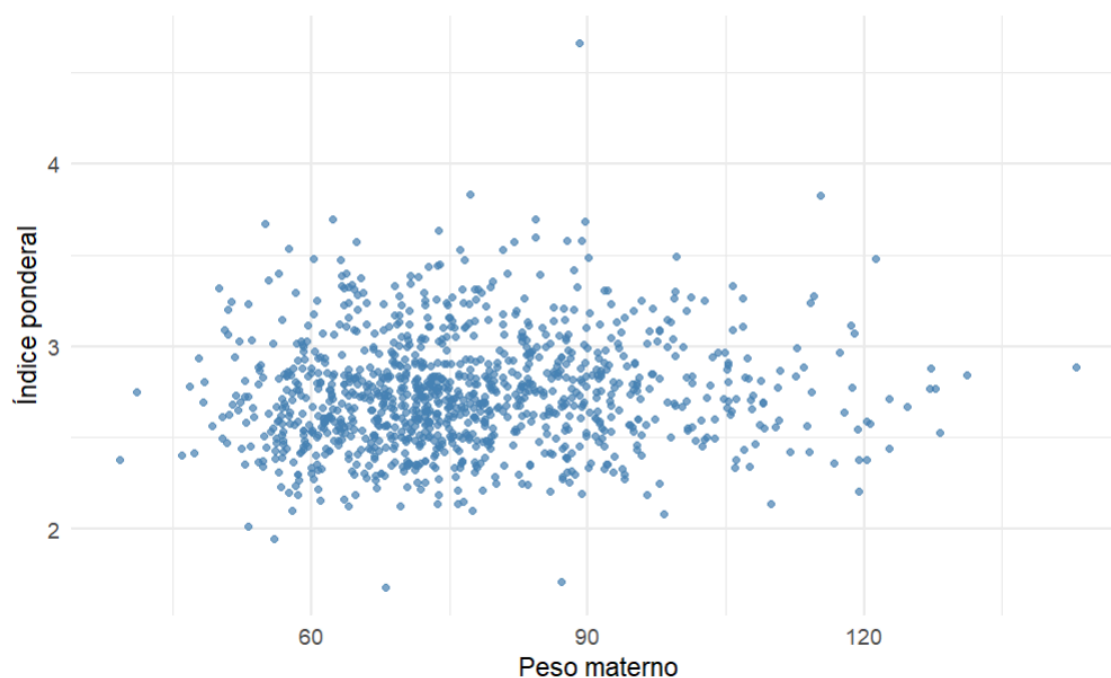


Figura B.19 Gráfico de dispersão do Peso materno pelo Índice ponderal

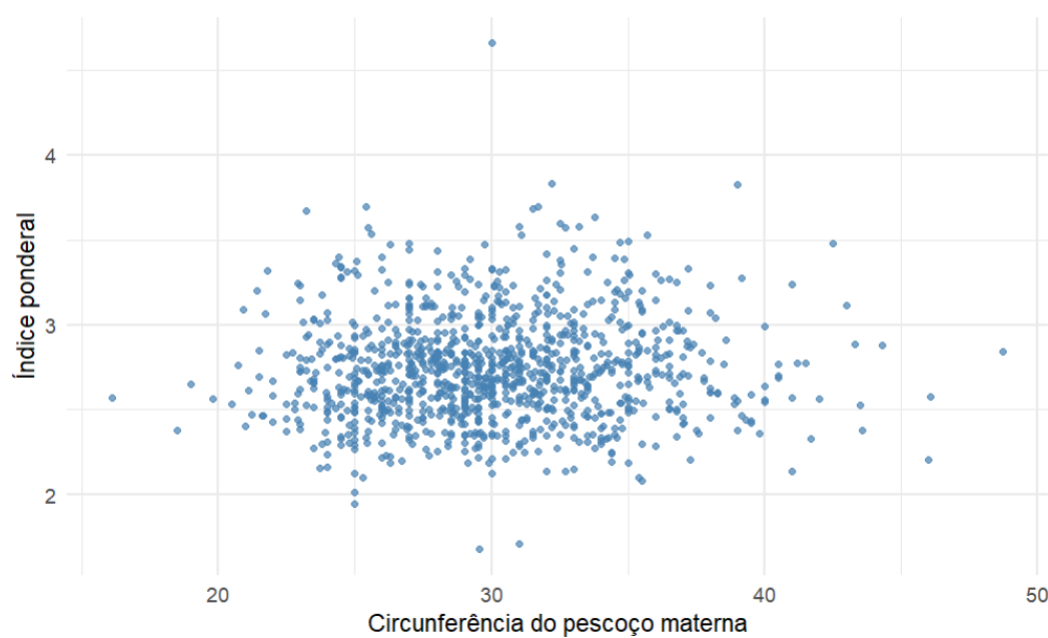


Figura B.20 Gráfico de dispersão da Circunferência do pescoço materna pelo Índice ponderal

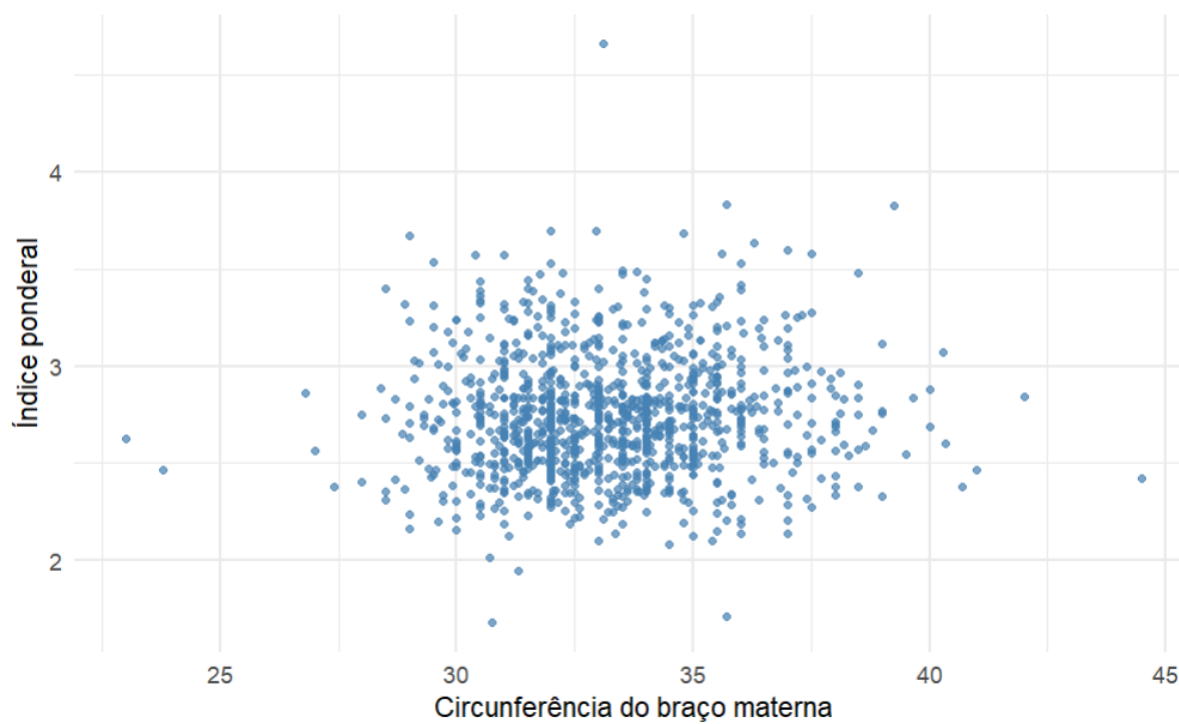


Figura B.21 Gráfico de dispersão da Circunferência do braço materna pelo Índice ponderal

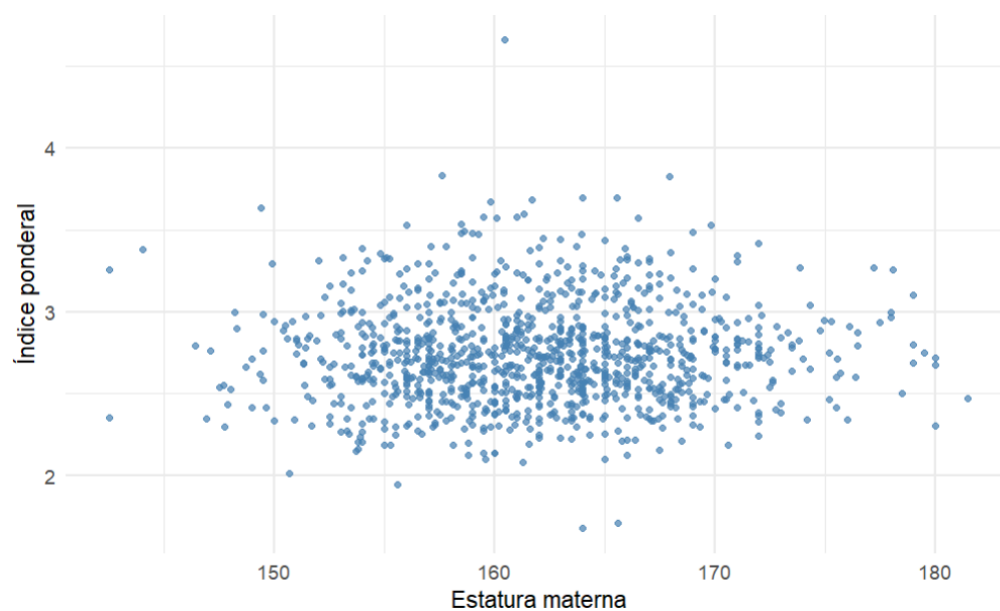


Figura B.22 Gráfico de dispersão da Estatura materna pelo Índice ponderal

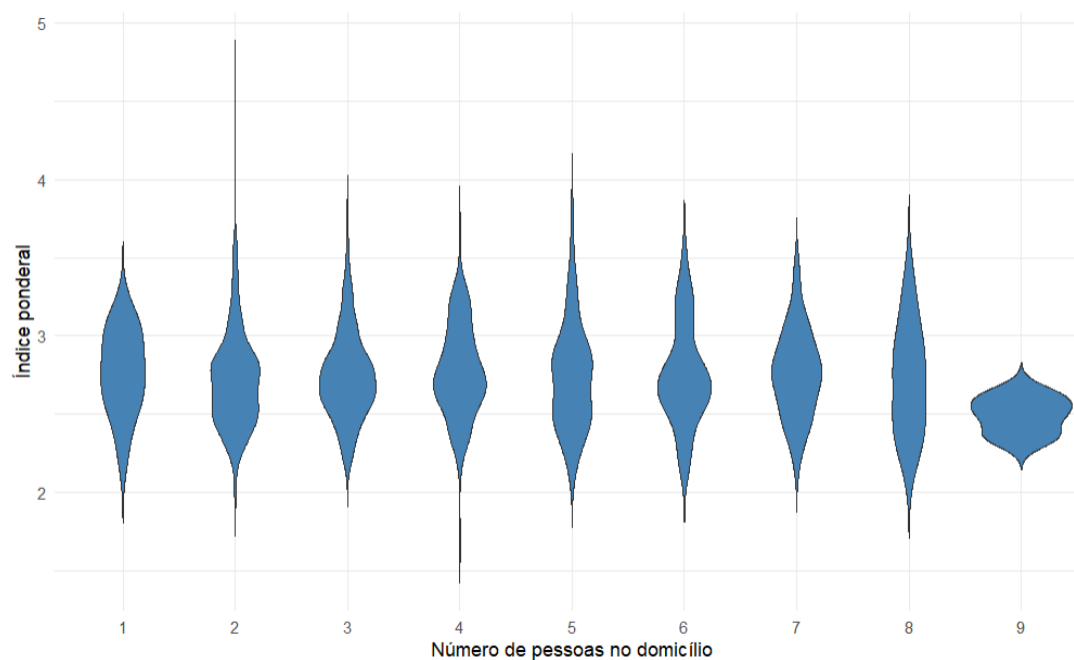


Figura B.23 *Violin plot* do Número de pessoas no domicílio pelo Índice ponderal

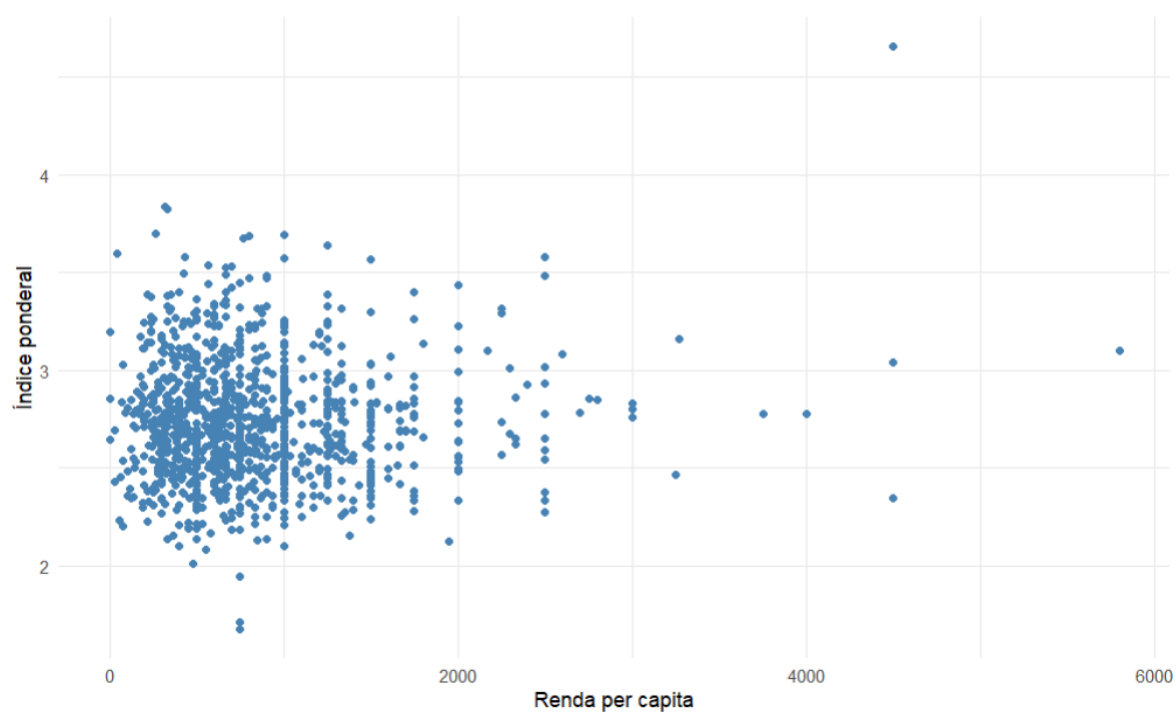


Figura B.24 Gráfico de dispersão da Renda per capita pelo Índice ponderal

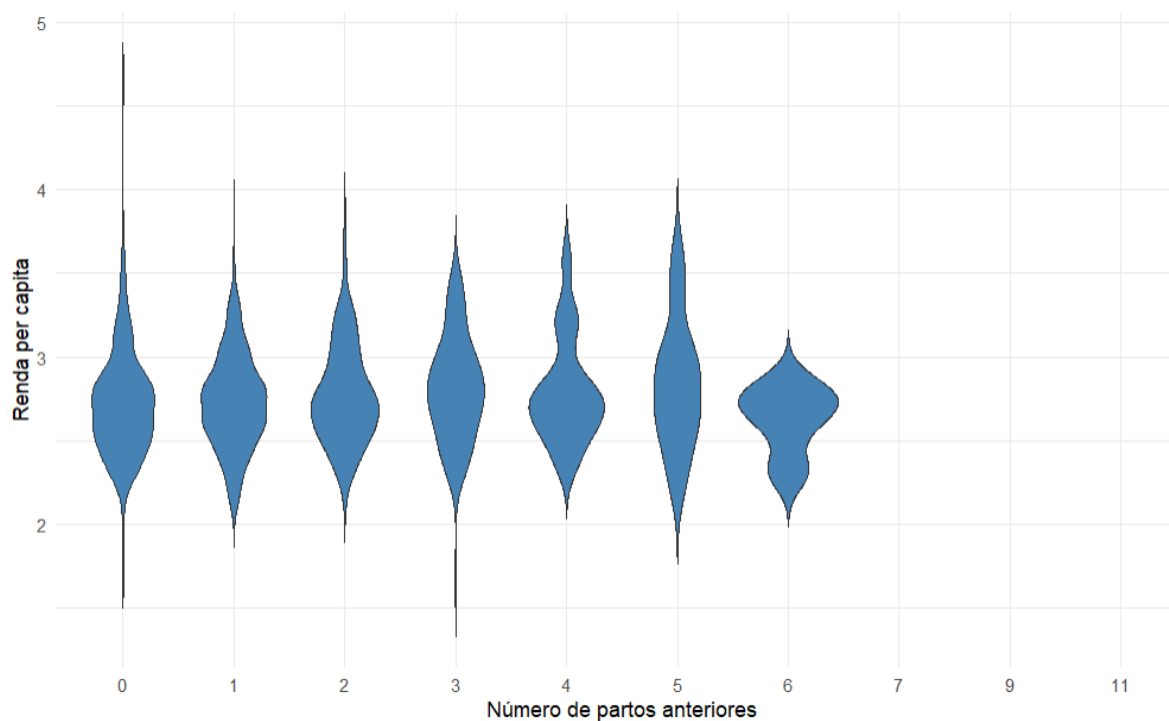


Figura B.25 *Violin plot* do Número de partos anteriores pelo Índice ponderal

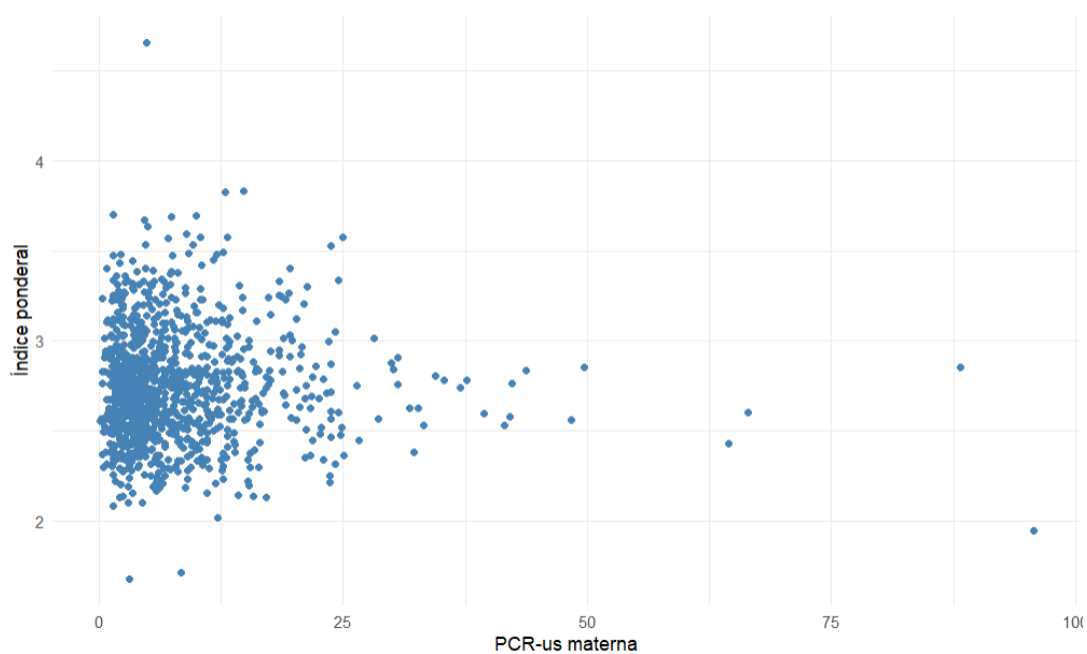


Figura B.26 Gráfico de dispersão da PCR-us materna pelo Índice ponderal

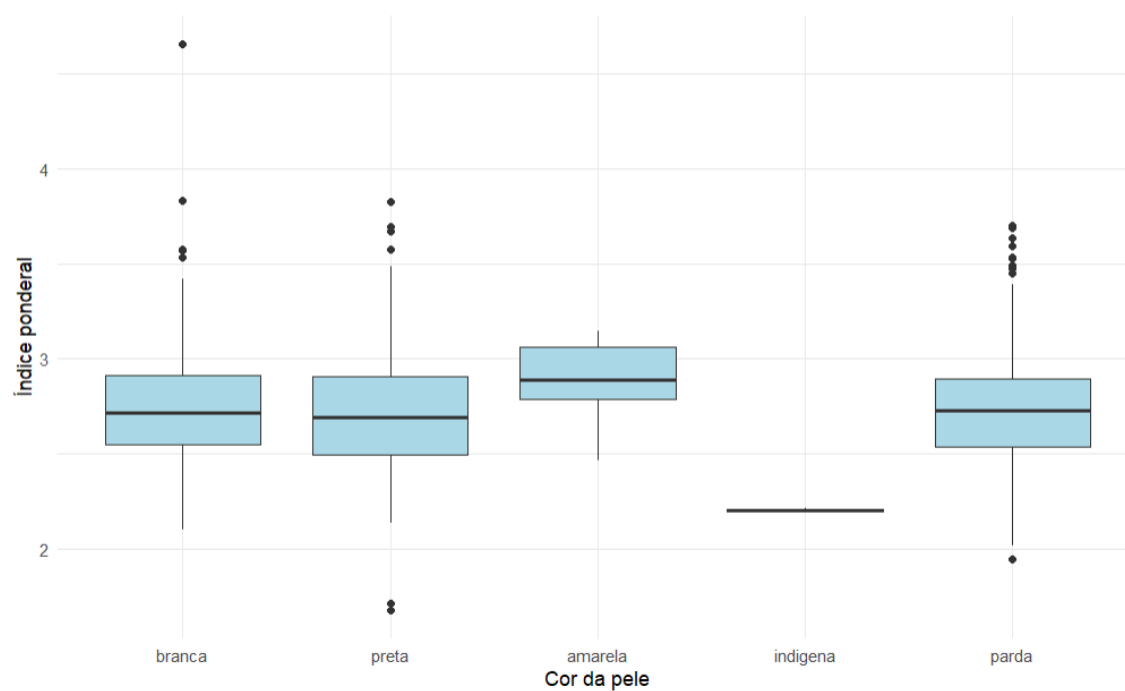


Figura B.27 *Box plot* do Índice ponderal por Cor da pele

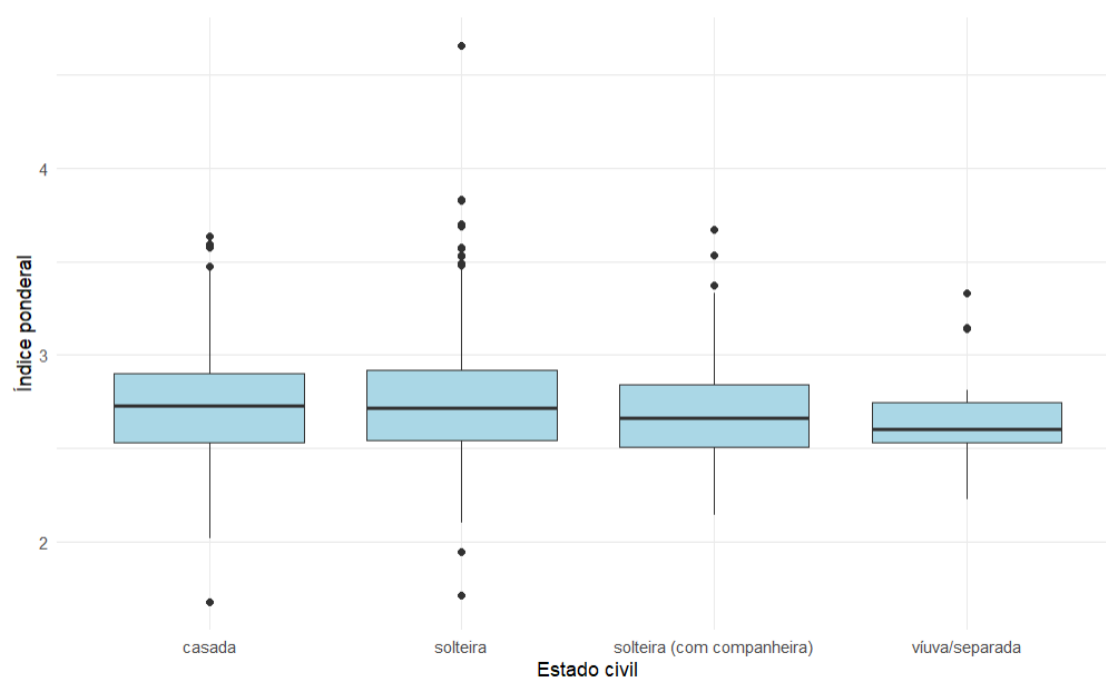


Figura B.28 *Box plot* do Índice ponderal por Estado civil

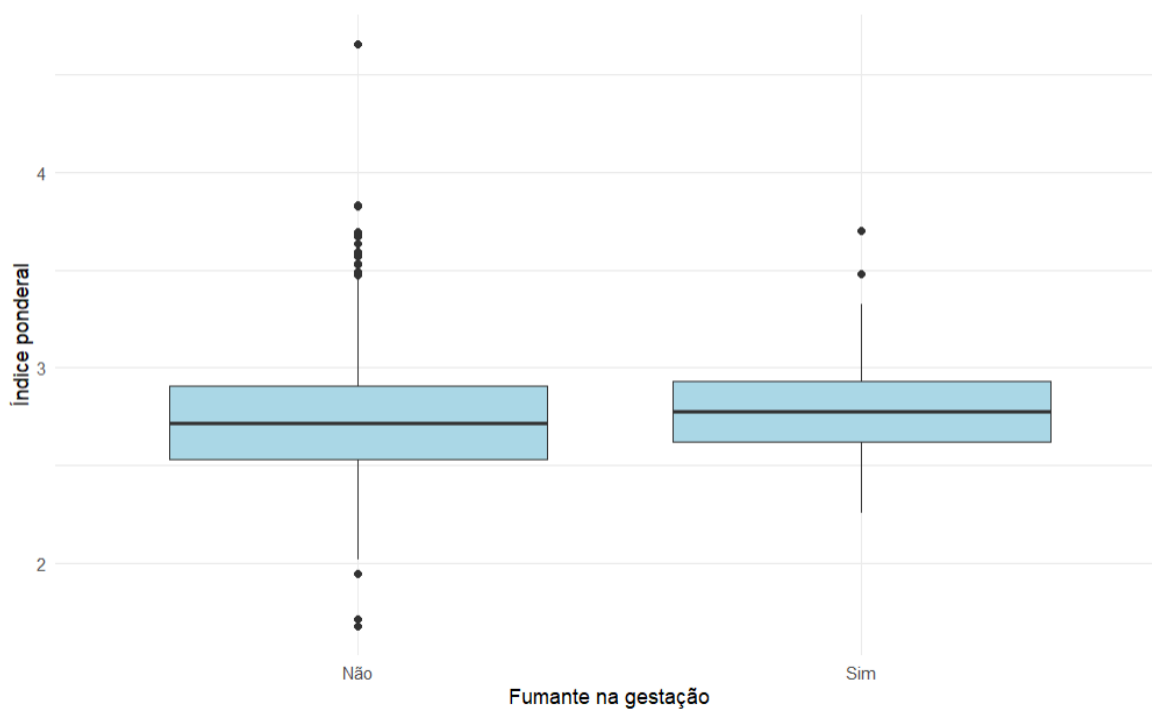


Figura B.29 *Box plot* de Índice ponderal por Fumante na gestação

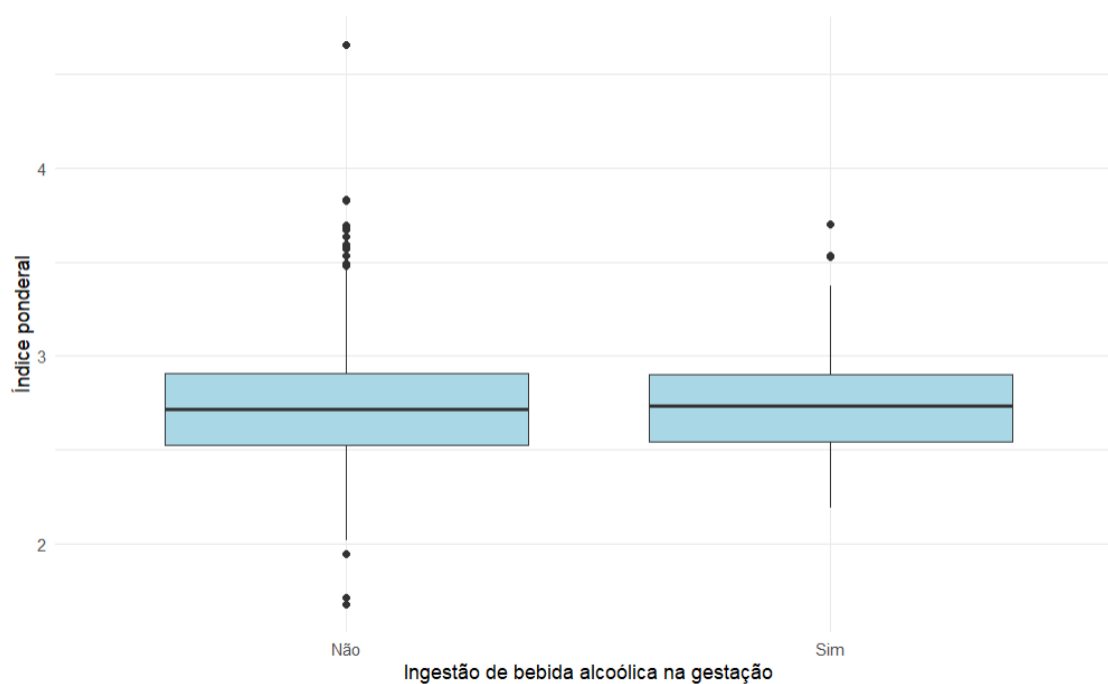


Figura B.30 *Box plot* de Índice ponderal por Ingestão de bebida alcoólica na gestação

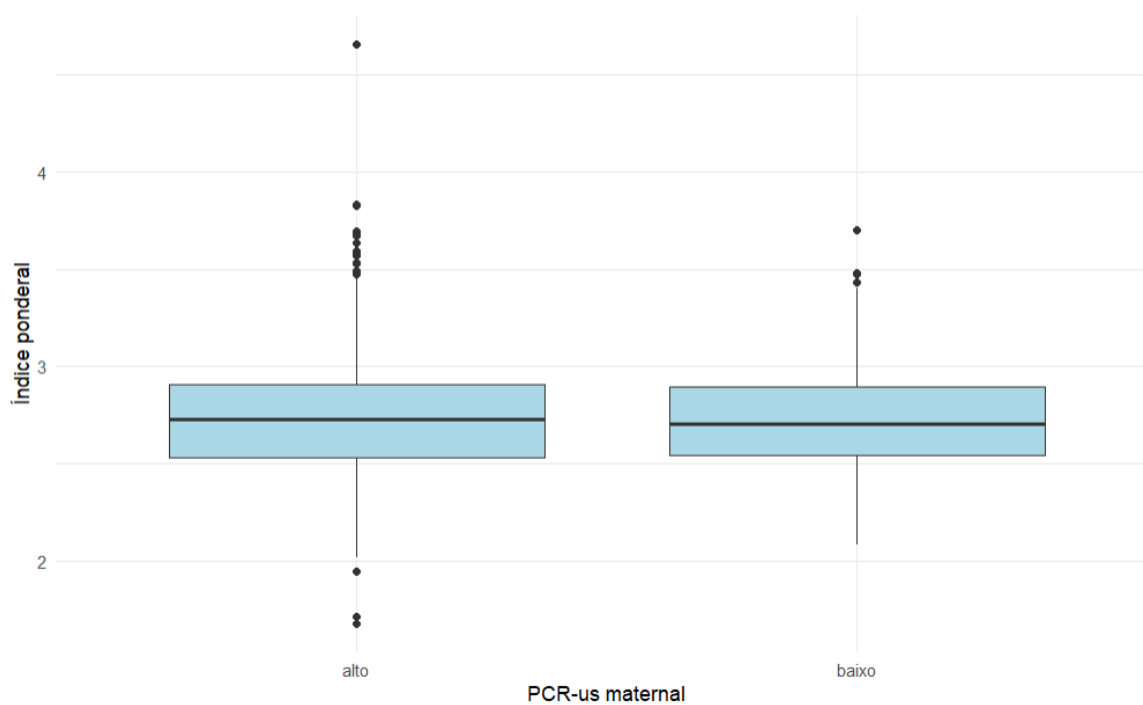


Figura B.31 *Box plot* de Índice ponderal por PCR-us maternal

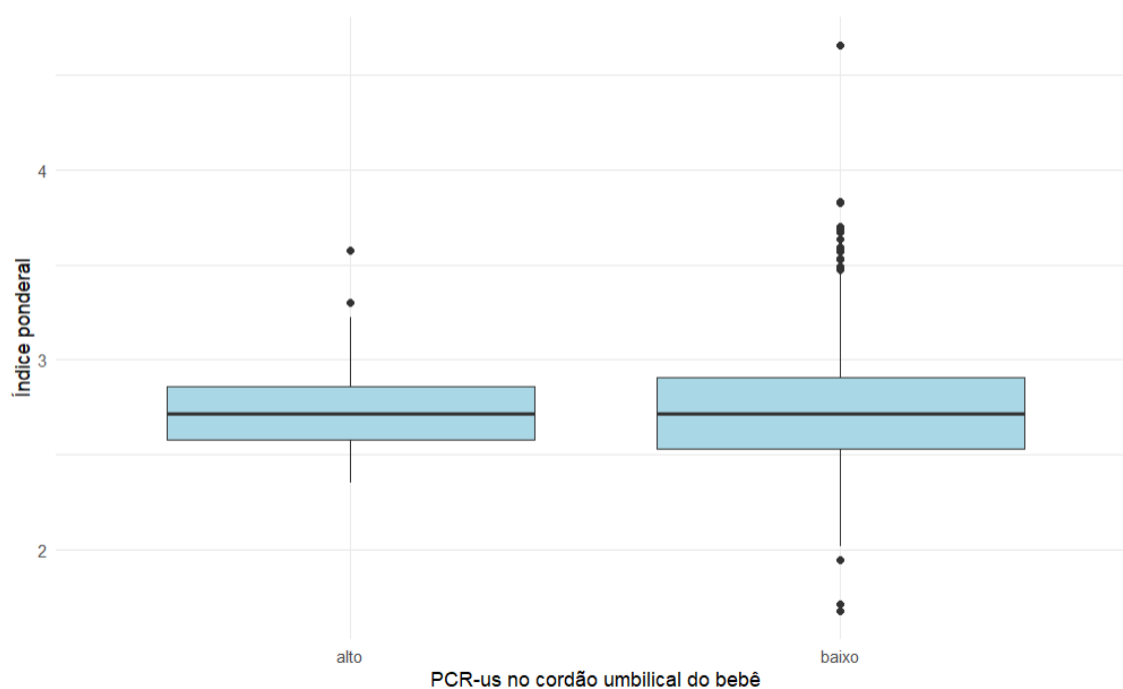


Figura B.32 *Box plot* de Índice ponderal por PCR-us no cordão umbilical do bebê

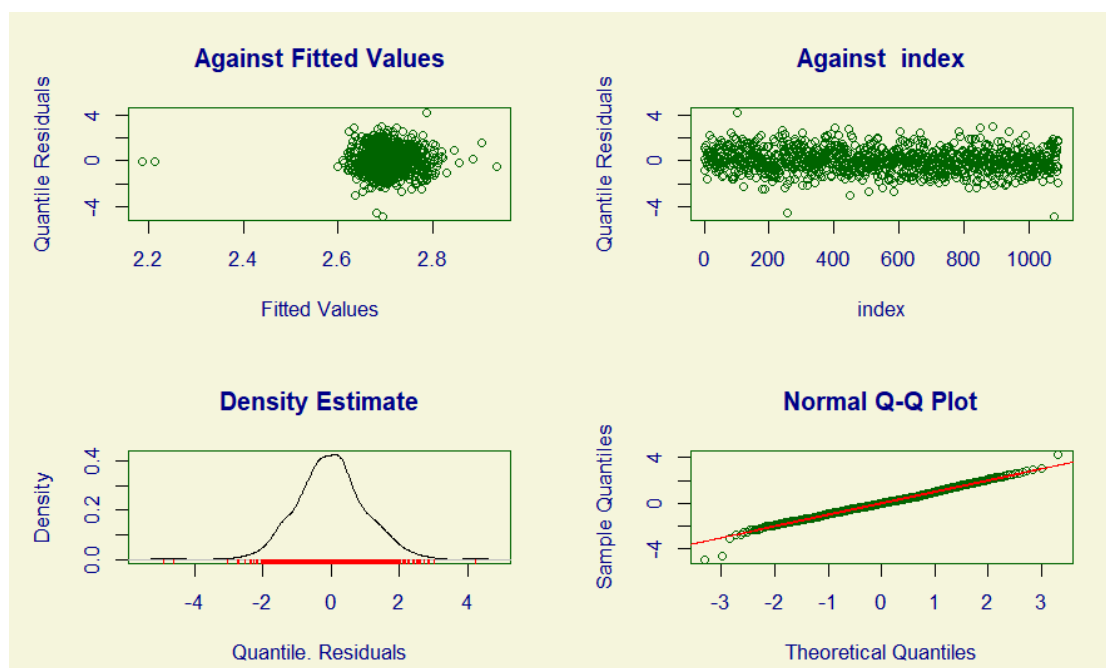


Figura B.33 Gráficos de resíduos do modelo para o Índice ponderal, considerando o Peso materno

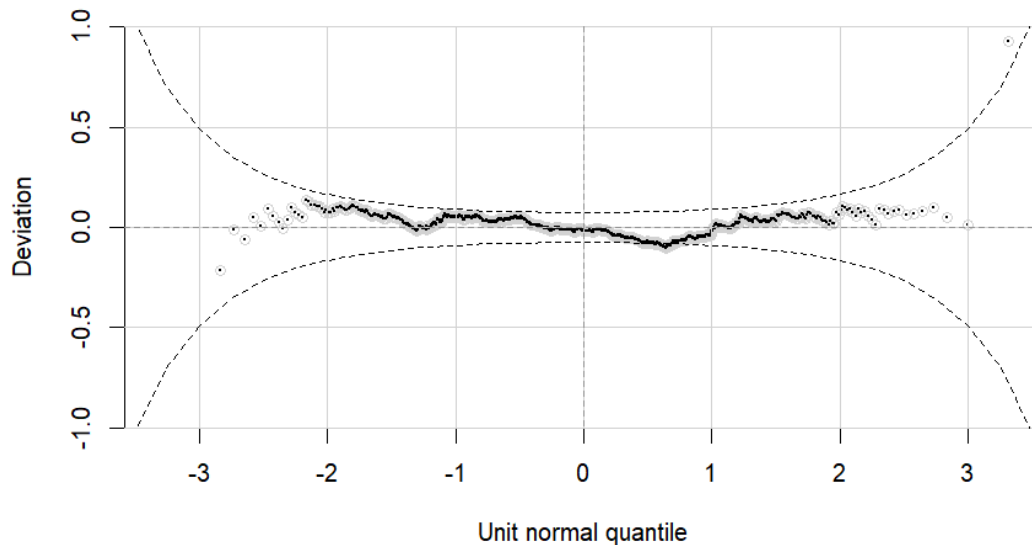


Figura B.34 *Worm plot* do modelo para o Índice ponderal, considerando o Peso materno

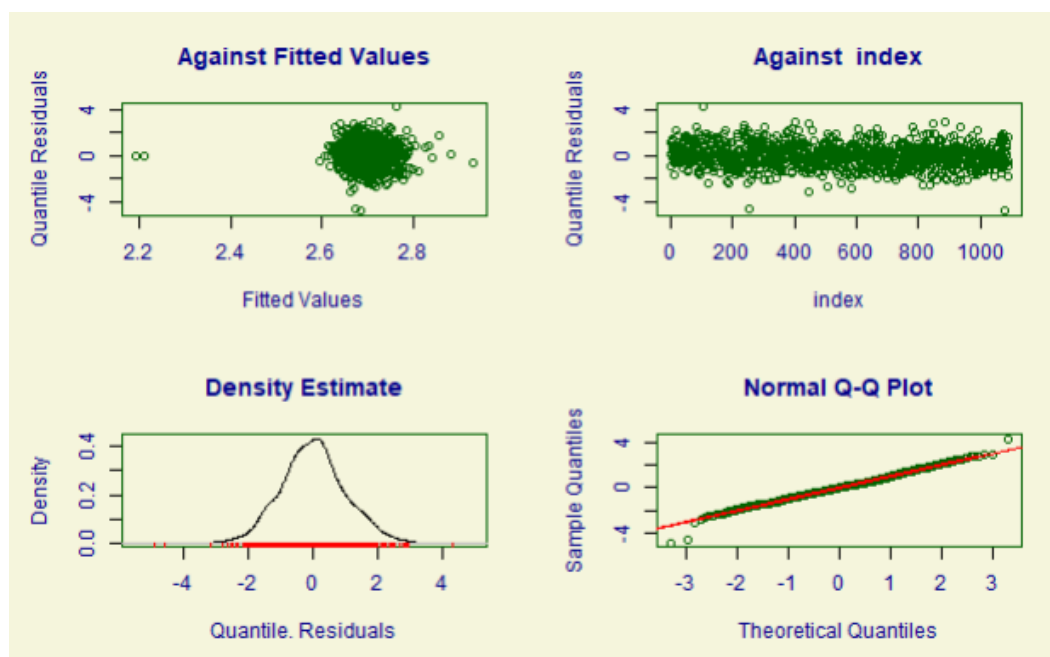


Figura B.35 Gráficos de resíduos do modelo para o Índice ponderal, considerando a Circunferência do pescoço materna

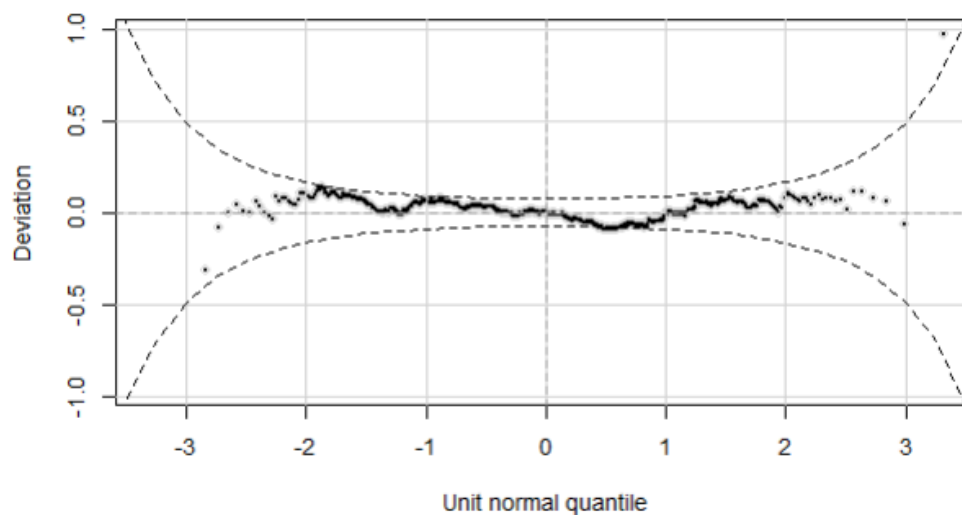


Figura B.36 *Worm plot* do modelo para o Índice ponderal, considerando a Circunferência do pescoço materna

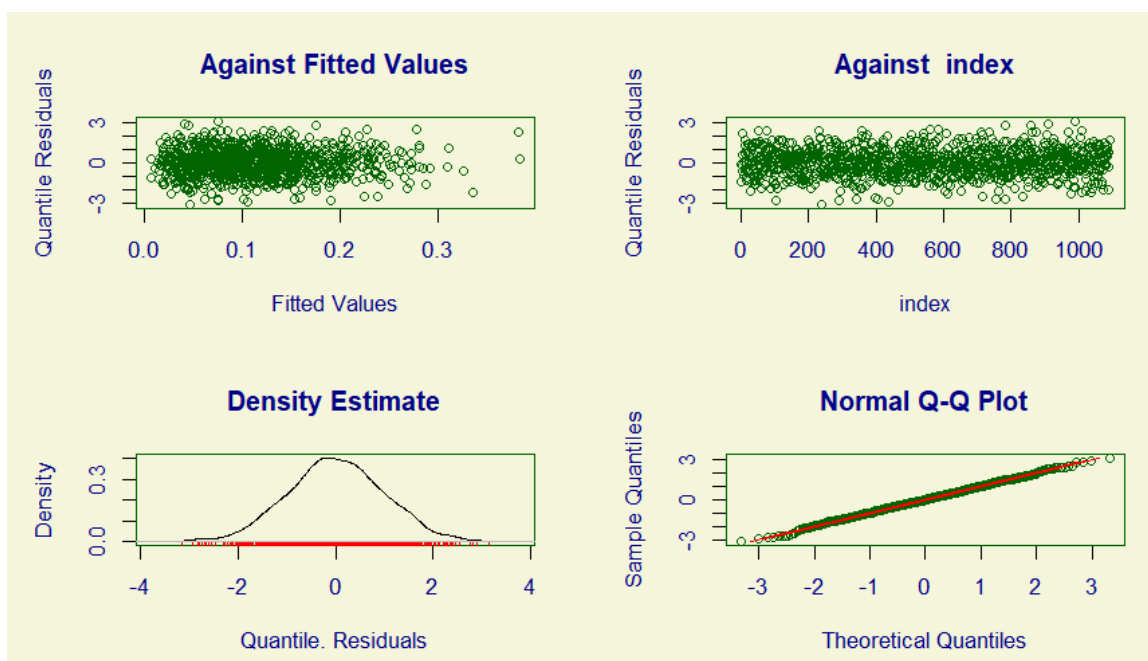


Figura B.37 Gráficos de resíduos do modelo logístico para o Peso para a idade gestacional, considerando o Peso materno – sem SMOTE

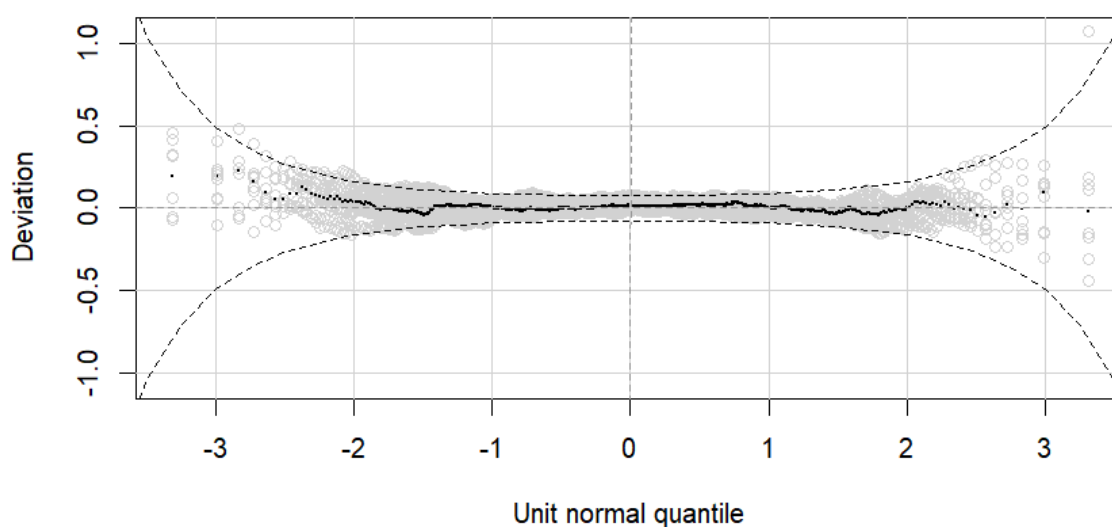


Figura B.38 *Worm plot* do modelo logístico para o Peso para a idade gestacional, considerando o Peso materno – sem SMOTE

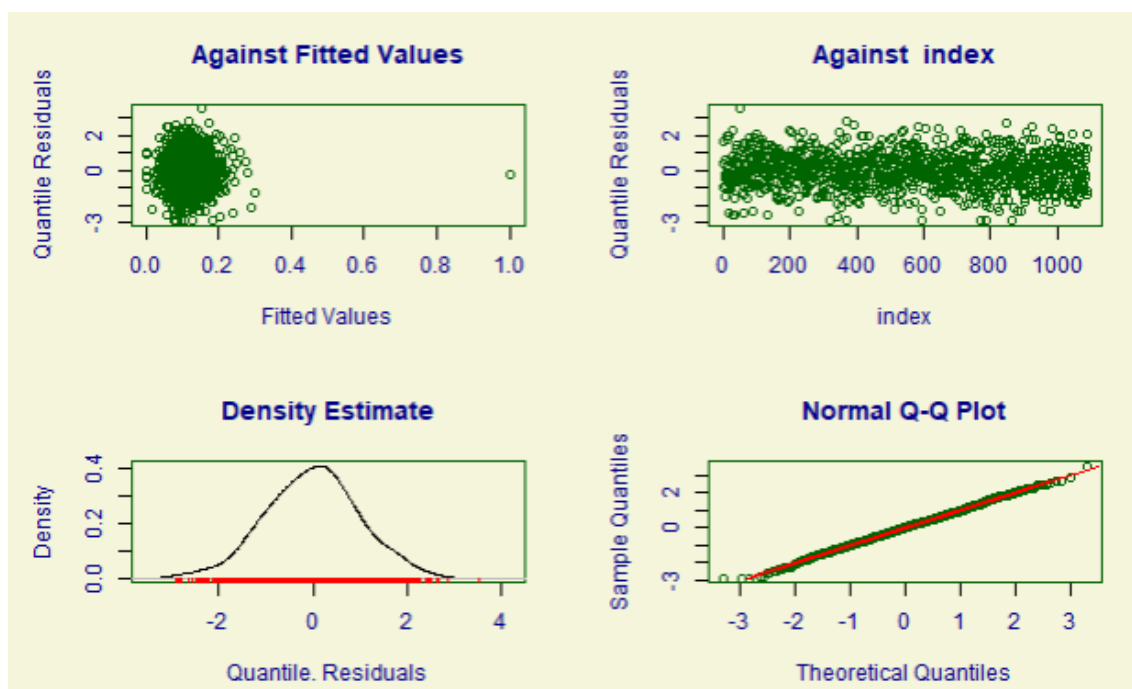


Figura B.39 Gráficos de resíduos do modelo logístico para o Peso para a idade gestacional, considerando a Circunferência do braço materna – sem SMOTE

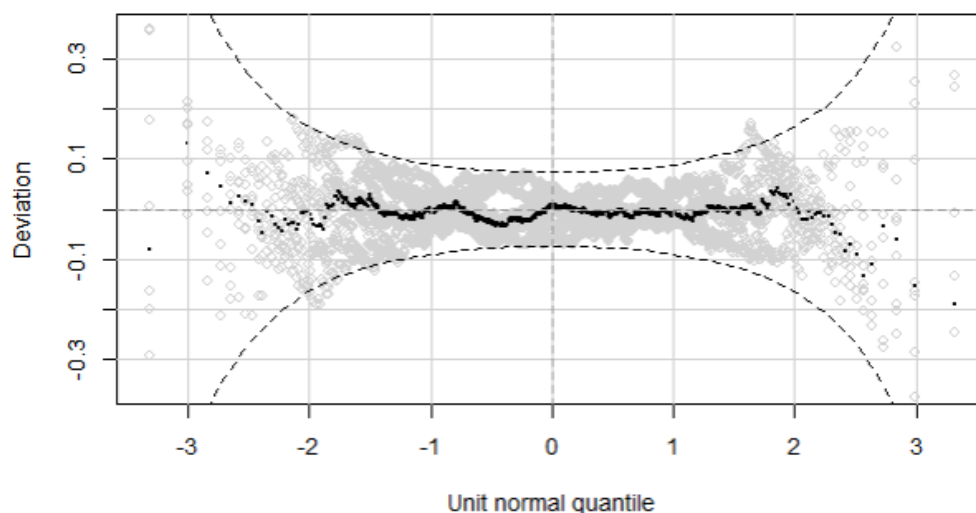


Figura B.40 *Worm plot* do modelo logístico para o Peso para a idade gestacional, considerando a Circunferência do braço materna – sem SMOTE

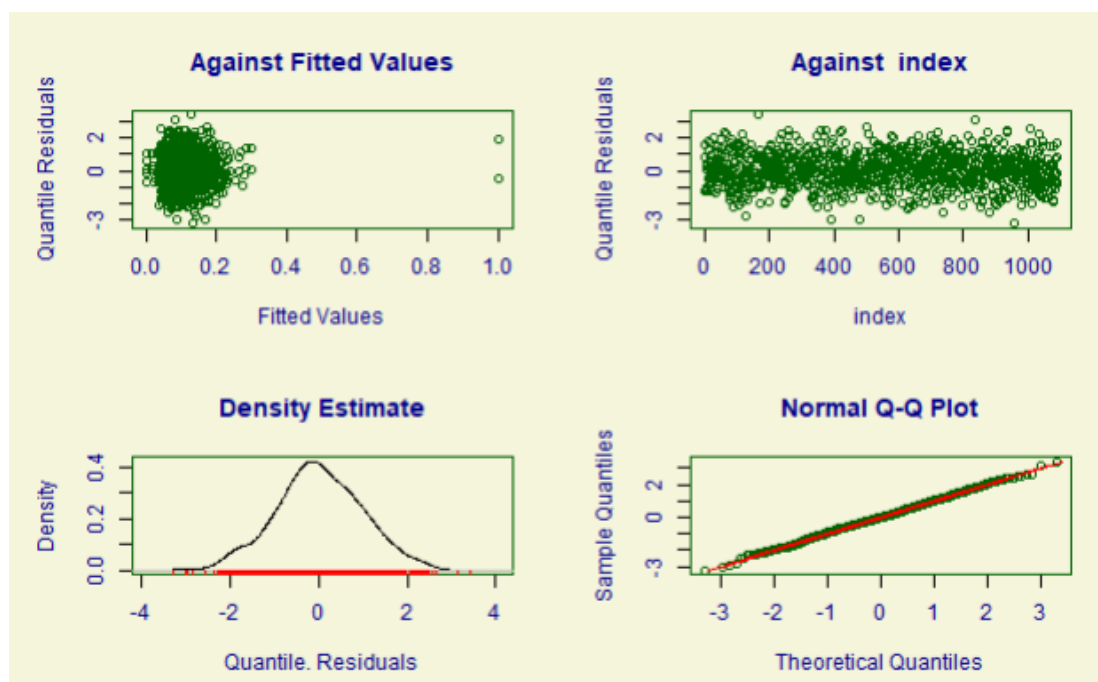


Figura B.41 Gráficos de resíduos do modelo logístico para o Peso para a idade gestacional, considerando a Circunferência do pescoço materna – sem SMOTE

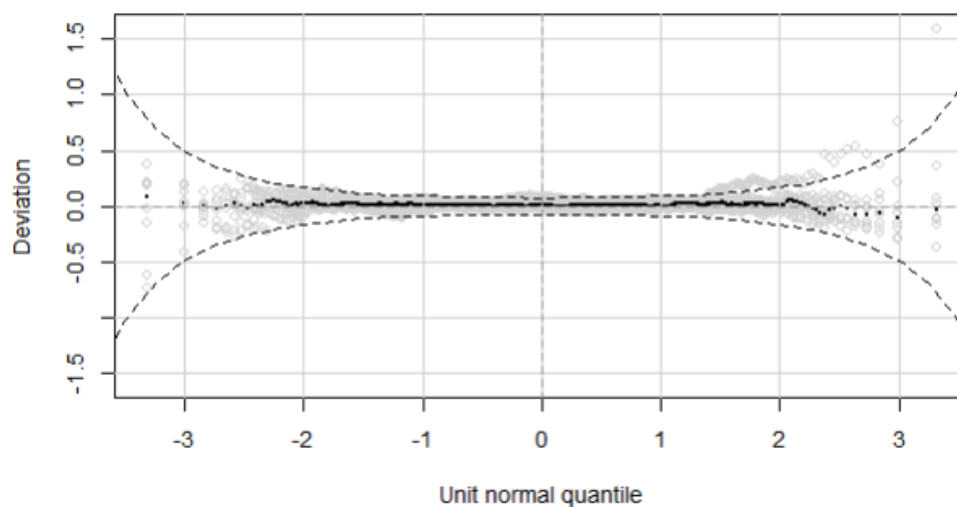


Figura B.42 *Worm plot* do modelo logístico para o Peso para a idade gestacional, considerando a Circunferência do pescoço materna – sem SMOTE

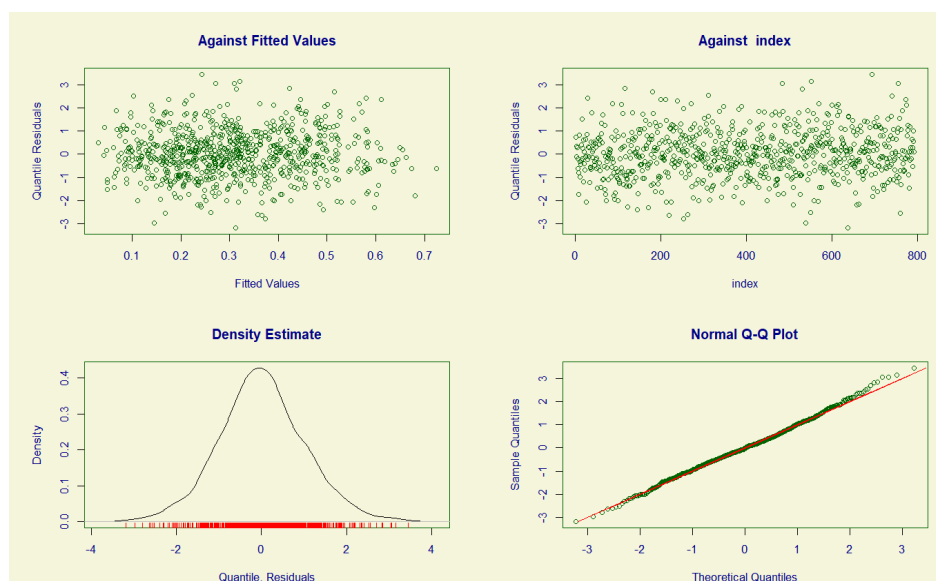


Figura B.43 Gráficos de resíduos do modelo logístico para o Peso para a idade gestacional, considerando o Peso materno – com SMOTE

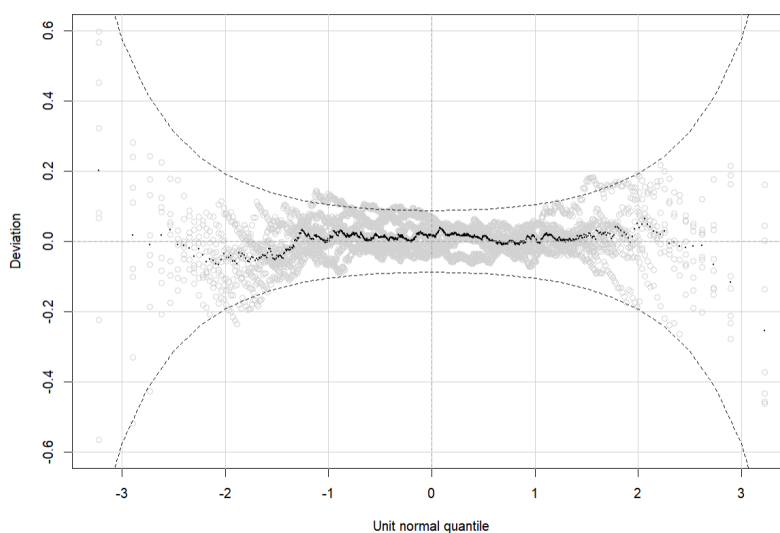


Figura B.44 *Worm plot* do modelo logístico para o Peso para a idade gestacional, considerando o Peso materno – com SMOTE

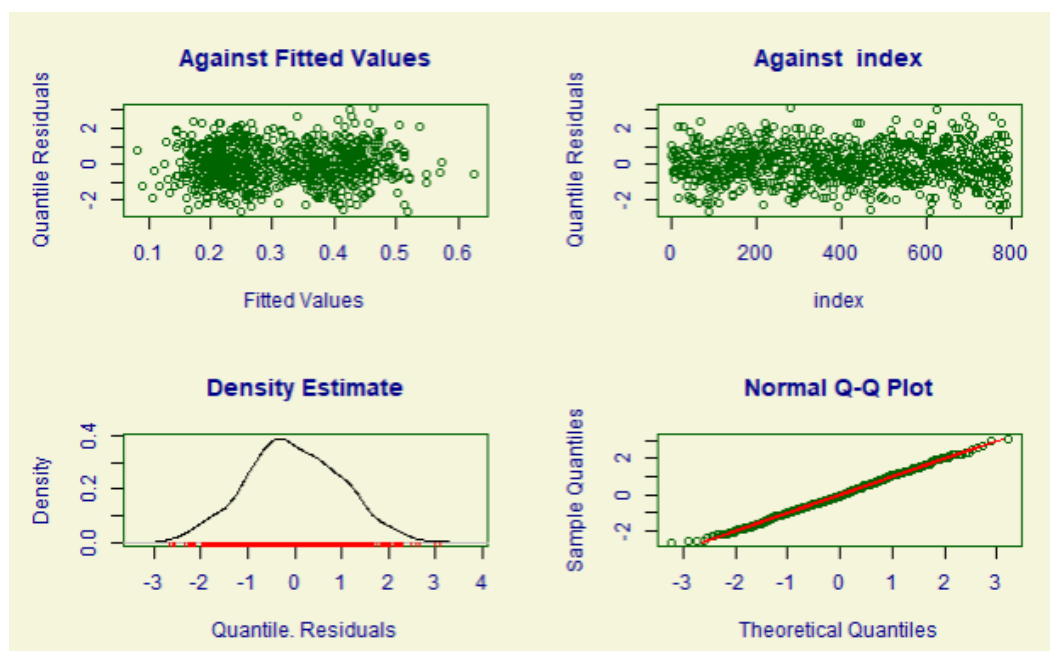


Figura B.45 Gráficos de resíduos do modelo logístico para o Peso para a idade gestacional, considerando a Circunferência do braço materna – com SMOTE

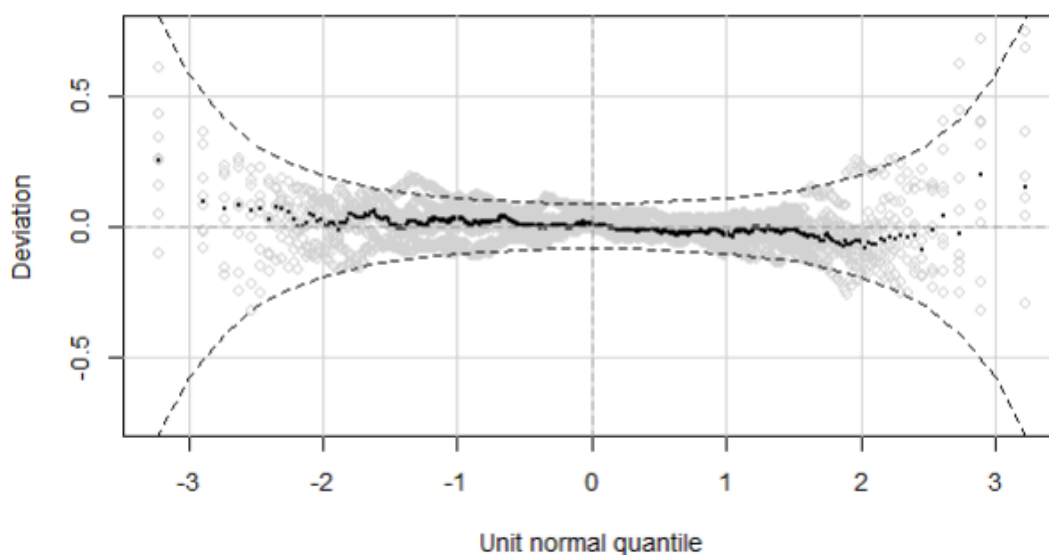


Figura B.46 *Worm plot* do modelo logístico para o Peso para a idade gestacional, considerando a Circunferência do braço materna – com SMOTE

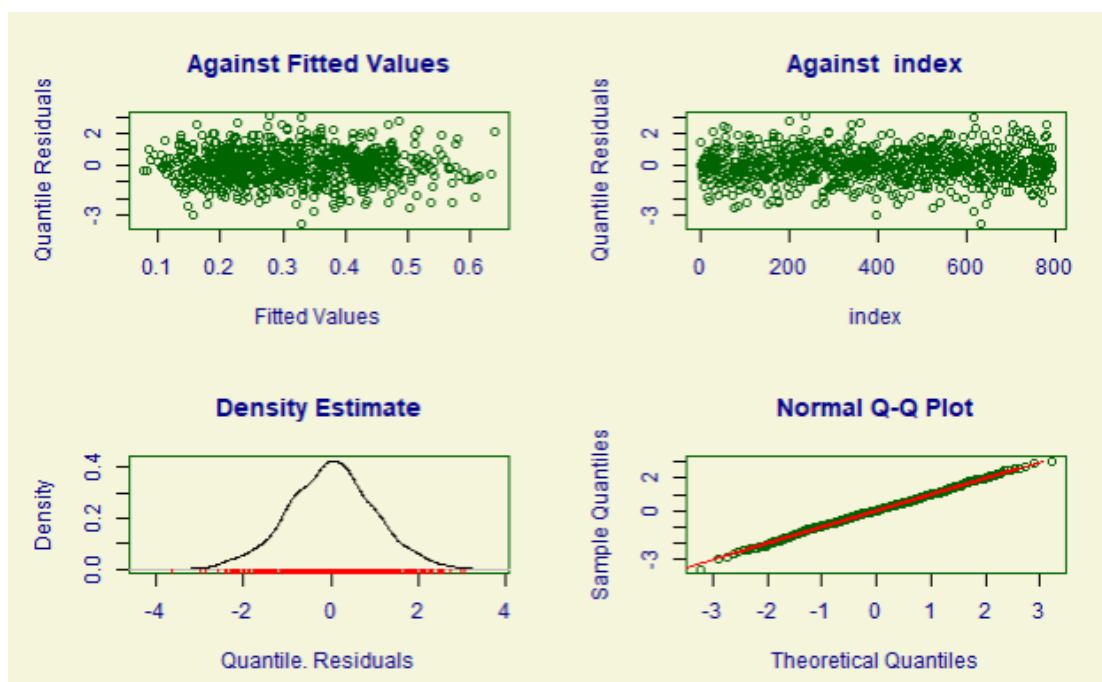


Figura B.47 Gráficos de resíduos do modelo logístico para o Peso para a idade gestacional, considerando a Circunferência do pescoço materna – com SMOTE

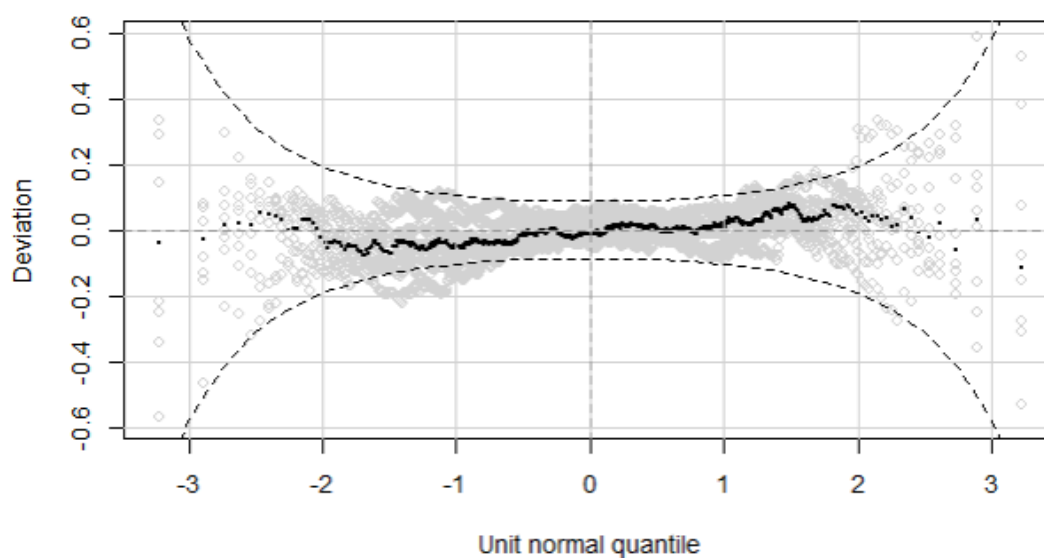


Figura B.48 Worm plot do modelo logístico para o Peso para a idade gestacional, considerando a Circunferência do pescoço materna – com SMOTE

APÊNDICE C

Metodologia

Optamos por usar modelos lineares generalizados (Paula, 2024) para a modelagem das variáveis respostas índice ponderal e peso por idade gestacional. Nos nossos modelos, não entrarão como variáveis explicativas as variáveis morfológicas do bebê, com exceção do PCR-us no cordão umbilical. Inicialmente, todas as outras variáveis entrarão no modelo, sendo um modelo final selecionado usando o procedimento *stepwise* com o critério generalizado de Akaike (GAIC). Tal critério maximiza a verossimilhança, mas penaliza o modelo que tem um número maior de parâmetros.

Decidimos utilizar as seguintes transformações nas variáveis explicativas:

- Número de partos anteriores: nos modelos, entrará como fator com dois níveis, primigesta (mulher tendo sua primeira gestação) e não-primigesta. Isto foi decidido após conversas com a pesquisadora sobre o fenômeno de mulheres primigestas terem um risco maior de bebês pequenos (Teran et al., 2017).
- PCR-us materna: visto que esta é uma variável que tem magnitude pequena, ou “explode”, será considerado no modelo o log desta variável (Altman, 1991).
- PCR-us no cordão umbilical: no modelo será considerado o log desta variável, pelo mesmo motivo da variável PCR-us materna.

Após o exame do histograma do índice ponderal (Figura B.15), optamos por inicialmente ajustar um modelo com distribuição gama para esse índice. Após termos resultados insatisfatórios usando a gama, optamos pela distribuição gama generalizada, com ligação log para os parâmetros μ e σ , e identidade para ν .

Optamos por dicotomizar a variável Peso por idade gestacional em níveis PIG e não-PIG, visto que o foco da pesquisadora são os bebês pequenos para idade gestacional, e ajustar um modelo de regressão logística. Como a proporção de bebês PIG é pequena (evento de interesse é raro), será utilizado o procedimento SMOTE, Synthetic Minority Oversampling TEchnique (Chawla et al., 2002).

O SMOTE é um procedimento de reamostragem que gera exemplos sintéticos da classe minoritária a partir dos próprios dados dessa classe, e pode ser combinado com undersampling da classe majoritária para equilibrar os dados.

Para cada ponto minoritário, são identificados os seus k-vizinhos mais próximos também na classe minoritária. Em seguida, seleciona-se aleatoriamente um desses vizinhos e gera-se um novo exemplo sintético pela interpolação.

Visto que nossos dados incluem variáveis categóricas, temos que usar uma modificação do SMOTE chamada SMOTE-NC (Chawla et al., 2002). A modificação é que, ao calcular a distância no algoritmo KNN, para cada variável categórica de níveis diferentes entre observações, é adicionada a mediana do desvio padrão das variáveis contínuas na distância. Na interpolação, o nível do fator é o majoritário nos k-vizinhos mais próximos.

Na modelagem, fizemos uma superamostragem de 200% na classe minoritária, e uma subamostragem de 50% na classe majoritária. Isto se deve ao fato do SMOTE enviesar a amostra e usar apenas a superamostragem para balancear os dados levaria a um viés muito grande.