

Rodada 2 - LabEst

Table of contents

1	Importação e preparação dos dados	2
2	Dicionário das variáveis	2
3	Análise Univariada	4
3.1	Variáveis Categóricas	4
3.2	Variáveis Numéricas	11
4	Variáveis resposta	12
4.1	Teste de Normalidade	12
5	Boxplot das Escalas PEDSQL	13
6	Associação das variáveis resposta (pedsqlglobal36)	14
6.1	Por gênero	14
6.2	Por Faixa Etária Materna	15
6.3	Por Idade Gestacional	16
6.4	Por Tipo de Parto	17
7	Boxplots das Variáveis Categóricas vs PEDSQL Global	17
7.1	Boxplots para Variáveis com Associação Significativa ($p < 0.05$)	17
7.1.1	Via de Parto ($p = 0.015$)	18
7.1.2	Horário Fixo para Telas ($p = 0.031$)	19
7.1.3	Limite de Tempo de Tela ($p = 0.003$)	20
7.2	Variáveis Demográficas e Perinatais	21
7.3	Variáveis de Desenvolvimento (Bayley)	22
7.4	Outras Variáveis de Interesse	23
8	Boxplot Completo - Todas as Variáveis Categóricas	25

9	Análise de Colinearidade e Correlação entre Variáveis de Tela	26
9.1	Matriz de Correlação - Variáveis de Tela	26
9.1.1	Interpretação da Matriz de Correlação	27
9.2	Análise VIF - Variáveis de Tela	27
9.2.1	Interpretação da Análise VIF	27
10	Modelagem de Regressão Multivariada	28
10.0.1	Estratégia de Modelagem	28
10.0.2	Fase 1: Preparação e Construção do Modelo Completo	29
10.0.3	Modelo 2: Modelo com Seleção Stepwise (Exploratório)	33
10.0.4	Modelo 3: Modelo com Seleção Stepwise (A partir das Variáveis Pré-Selecionadas)	35
10.0.5	Modelo 4: Modelo Hierárquico Sequencial	37
10.0.6	Fase Final: Comparação e Seleção do Modelo	42
11	Análise de Resíduos	43
11.1	Diagnóstico do Modelo Stepwise Refinado	44
11.1.1	Teste de Normalidade dos Resíduos (Shapiro-Wilk)	45
11.2	Diagnóstico do Modelo Hierárquico Final	46
11.2.1	Teste de Normalidade dos Resíduos (Shapiro-Wilk)	47
12	Modelagem Avançada: Regressão Beta	47
12.1	Por que um novo modelo?	47
12.2	Implementação da Regressão Beta	47
12.2.1	Passo 1: Preparação da Variável Resposta e Seleção de Preditoras	47
12.2.2	Fase 2: Construção e Comparação dos Modelos	48
12.2.3	Fase 3: Seleção, Diagnóstico e Interpretação	52
13	Conclusão Final	58

1 Importação e preparação dos dados

2 Dicionário das variáveis

Table 1: Dicionário resumido das variáveis

Variavel	Explicacao	Categorias
idademaecat	Categorização da idade da mãe da criança	Adulta: 1 / Adolescente: 0
pesocat	Categorização peso de nascimento da criança	≥ 2.500 g: 1 / < 2.500 g: 0

Variavel	Explicacao	Categorias
igcat	Categorização da idade gestacional de nascimento da criança	≥ 37 semanas: 1 / < 37 semanas: 0
partocat	Categorização da via de parto	Vaginal ou fórceps: 1 / Cesariana: 0
lme6	Criança em aleitamento materno exclusivo aos 6 meses de idade	Sim: 1 / Não: 0
corcat24	Categorização da etnia	Branco: 1 / Não branco: 0
estadocivilmae24	Estado civil da mãe da criança	Casada ou em união estável: 1 / Solteira ou divorciada: 0
escolamaecat36	Categorização da escolaridade da mãe da criança	Superior/Pós: 1, Médio: 2, Fundamental: 3
abepcat36	Categorização social pelo Critério Brasil - ABEP	A, B, C, DE
genero	Gênero	Feminino/Masculino
tanu	A criança realizou a Triagem Auditiva Neonatal Universal (TANU)?	Sim/Não
resultadotanu	O resultado da TANU foi normal?	Sim ou resultado desconhecido
resultqtai36	Resultado do QTAI aos 36 meses	Positivo: Risco / Negativo: Sem Risco
creche36	Criança frequentando creche	Sim/Não
ludicascat36	Categorização da quantidade de atividades lúdicas	Adequada: 4+ / Inadequada: 1-3
tempotelacat36	Categorização do tempo de tela por dia	Usa $< 1h$: 1 / Usa $\geq 1h$: 0
horacertacat36	Categorização do horário para usar telas	Sempre: 1 / Às vezes ou Nunca: 0
telacasa36	Maneira que a criança usa telas em casa	Acompanhado (adultos), Acompanhado (criança), Sozinho
telalimitetempoCat36	Cuidadores limitam o tempo de uso de telas?	Sim: 1 / Às vezes ou Não: 0
telainteraçãocat36	Cuidadores interagem com a criança durante o uso de telas?	Sim: 1 / Às vezes ou Não: 0
telalimiteconteuCat36	Cuidadores limitam os conteúdos que a criança assiste em telas?	Sim: 1 / Às vezes ou Não: 0
pedsqlfisico36	Peds-QL - Escore Físico	Contínuo (quanto maior, melhor)
pedsqlpsico36	Peds-QL - Escore Psicossocial (Escolar + Emocional + Social)	Contínuo (quanto maior, melhor)
pedsqlglobal36	Peds-QL - Escore Global	Contínuo (quanto maior, melhor)

Variavel	Explicacao	Categorias
classcogbal36	Bayley - Domínio Cognitivo Escore Balanceado - Classificação	Normal: 1 / Atraso: 0
classcogcomp36	Bayley - Domínio Cognitivo Escore Composto - Classificação	Normal: 1 / Atraso: 0
claslrbal36	Bayley - Domínio Linguagem Receptiva - Classificação	Normal: 1 / Atraso: 0
classlebal36	Bayley - Domínio Linguagem Expressiva - Classificação	Normal: 1 / Atraso: 0
classlingcomp36	Bayley - Domínio Linguagem Composto (Receptiva e Expressiva) - Classificação	Normal: 1 / Atraso: 0
classmfbal36	Bayley - Domínio Motor Fino - Classificação	Normal: 1 / Atraso: 0
classmgbal36	Bayley - Domínio Motor Grosso - Classificação	Normal: 1 / Atraso: 0
classmotorcomp36	Bayley - Domínio Motor Composto (Fino e Grosso) - Classificação	Normal: 1 / Atraso: 0

3 Análise Univariada

3.1 Variáveis Categóricas

A tabela a seguir é o primeiro passo da nossa investigação. Ela tem dois objetivos principais: primeiro, descrever as características das crianças e famílias do estudo (ex: quantos são meninos, quantas mães são adolescentes, etc.); segundo, fazer uma triagem inicial para identificar quais dessas características parecem estar associadas à qualidade de vida das crianças (medida pela escala PEDSQL Global).

Por exemplo, para a “Via de Parto”: - Crianças que nasceram de Cesariana tiveram uma pontuação média de qualidade de vida de 93.87. - Crianças que nasceram de Vaginal/Fórceps tiveram uma pontuação média de 90.22. Apenas olhando para esses números, parece haver uma diferença. Mas essa diferença é “real” ou poderia ter acontecido por puro acaso? É aí que entram os testes estatísticos.

Para cada variável, a tabela mostra: - **N (%)**: A contagem e o percentual de crianças em cada categoria. - **Média (DP)** e **Mediana (IQR)**: A média e a mediana do escore de qualidade de vida para cada grupo. Como a nossa variável de qualidade de vida não segue uma distribuição normal (conforme verificado no Teste de Normalidade), a **mediana** é a medida mais confiável para comparação. - **p-valor**: O resultado de um teste estatístico que avalia se a diferença observada entre os grupos é real ou se poderia ter acontecido por acaso. O **p (Mediana)** (Teste de Wilcoxon/Kruskal-Wallis) é o mais importante para nós.

Decidimos destacar em verde todas as variáveis com um **p-valor menor que 0.25** em qualquer um dos testes. Isso não significa que a associação é definitiva, mas sim que a variável é uma **candidata forte** para ser incluída na próxima etapa da análise: o modelo de regressão multivariada.

Table 2: Análise descritiva das variáveis categóricas com PEDSQL Global

Variável	Categoria	N (%)	Média (DP)	Mediana (IQR)	P-valor (Média)	p-valor (Mediana)
Idade da Mãe	Adolescente	3 (2.2%)	87.11 (± 12.2)	90.14 (81.91–93.82)	56.1%	43.8%
Idade da Mãe	Adulta	131 (97.8%)	91.98 (± 7.75)	93.99 (88.94–96.88)		
Peso ao Nascer	< 2500g	13 (9.8%)	88.31 (± 13.28)	90.14 (87.02–97.5)	30.7%	27.7%
Peso ao Nascer	2500g	119 (90.2%)	92.29 (± 7.03)	94.23 (88.94–96.88)		
Idade Gestacional	< 37 sem.	5 (3.6%)	90.23 (± 7.82)	89.9 (85.94–95.31)	64.5%	48.3%
Idade Gestacional	37 sem.	132 (96.4%)	91.99 (± 7.82)	94.11 (88.94–96.88)		
Via de Parto	Cesariana	60 (44.4%)	93.87 (± 5.54)	95.75 (90.47–98.08)	0.4%	1.5%
Via de Parto	Vaginal/Fórceps	75 (55.6%)	90.22 (± 8.96)	92.81 (86.84–96.2)		
Aleitamento Exclusivo (6m)	Não	84 (62.7%)	91.56 (± 6.96)	92.9 (88.16–96.32)	57.2%	14.8%
Aleitamento Exclusivo (6m)	Sim	50 (37.3%)	92.42 (± 9.27)	94.77 (90.94–97.93)		

Variável	Categoria	N (%)	Média (DP)	Mediana (IQR)	p- valor (Média)	p-valor (Mediana)
Etnia	Não Branco	73 (56.6%)	92.51 (±6.83)	94.38 (90.14–96.88)	68.9%	77.9%
Etnia	Branco	56 (43.4%)	92.01 (±7.21)	93.99 (88.22–97.64)		
Estado Civil da Mãe	Solteira/Div.	52 (40.9%)	91.47 (±6.35)	92.49 (88.22–96.15)	27.6%	9.0%
Estado Civil da Mãe	Casada/União	75 (59.1%)	92.8 (±7.22)	94.38 (90.26–97.3)		
Categorização da escolaridade da mãe da criança	1	53 (38.7%)	92.99 (±5.48)	94.59 (90.5–96.25)	N/A	N/A
Categorização da escolaridade da mãe da criança	2	79 (57.7%)	91.29 (±9.04)	93.12 (88.22–97.5)		
Categorização da escolaridade da mãe da criança	3	5 (3.6%)	90.7 (±7.74)	92.67 (84.38–95.19)		
Categorização social pelo Critério Brasil - ABEP	A	14 (10.2%)	95.52 (±4.6)	96.05 (95.01–98.8)	N/A	N/A
Categorização social pelo Critério Brasil - ABEP	B	30 (21.9%)	92.12 (±5.98)	93.44 (88.52–96.45)		
Categorização social pelo Critério Brasil - ABEP	C	75 (54.7%)	91.49 (±8.88)	93.44 (88.94–97)		
Categorização social pelo Critério Brasil - ABEP	DE	18 (13.1%)	90.64 (±7.29)	90.38 (87.38–96.69)		
Gênero	Feminino	57 (41.6%)	90.69 (±9.03)	92.81 (86.56–96.51)	13.6%	18.6%

Variável	Categoria	N (%)	Média (DP)	Mediana (IQR)	p-valor (Média)	p-valor (Mediana)
Gênero	Masculino	80 (58.4%)	92.81 (±6.71)	94.48 (90.32–96.94)		
Resultado QTAI (Risco para perda auditiva)	NEGATIVA	83 (60.6%)	92.03 (±7.26)	93.44 (88.58–96.69)	85.8%	90.7%
Resultado QTAI (Risco para perda auditiva)	POSITIVA	54 (39.4%)	91.77 (±8.63)	94.38 (88.94–96.88)		
Criança frequentando creche	Não	43 (31.4%)	89.72 (±10.32)	93.12 (84.06–96.25)	6.2%	17.6%
Criança frequentando creche	Sim	94 (68.6%)	92.94 (±6.13)	94.59 (89.24–97.06)		
Categorização da quantidade de atividades lúdicas	adequado	102 (74.5%)	92.18 (±7.03)	94.3 (88.94–96.88)	58.1%	91.5%
Categorização da quantidade de atividades lúdicas	inadequado	35 (25.5%)	91.19 (±9.78)	93.44 (89.22–96.88)		
Categorização do tempo de tela por dia	0	108 (78.8%)	91.33 (±8.2)	93.28 (88.16–96.51)	3.6%	9.1%
Categorização do tempo de tela por dia	1	29 (21.2%)	94.14 (±5.65)	95.19 (90.94–99.04)		
Categorização do horário para usar telas	0	85 (65.4%)	90.79 (±8.79)	93.03 (86.56–96.51)	0.7%	3.1%
Categorização do horário para usar telas	1	45 (34.6%)	94.24 (±5.45)	95.19 (91.83–98.08)		
Maneira que a criança usa telas em casa	Acompanhado pelos cuidadores (adultos)	100 (76.9%)	92.22 (±7.04)	94.38 (88.84–96.88)	N/A	N/A

Variável	Categoria	N (%)	Média (DP)	Mediana (IQR)	p- valor (Média)	p-valor (Mediana)
Maneira que a criança usa telas em casa	Acompanhado por outra criança	14 (10.8%)	89.63 (±13.83)	94.18 (87.94–96.88)		
Maneira que a criança usa telas em casa	Sozinho	16 (12.3%)	93 (±6.57)	93.83 (89.84–97.84)		
Cuidadores limitam o tempo de uso de telas?	0	45 (34.4%)	89.31 (±8.37)	92.5 (84.13–95.94)	0.7%	0.3%
Cuidadores limitam o tempo de uso de telas?	1	86 (65.6%)	93.42 (±7.34)	95.19 (90.2–98.08)		
Cuidadores interagem com a criança durante o uso de telas?	0	49 (37.7%)	92.05 (±6.55)	93.03 (88.22–96.88)	92.1%	57.2%
Cuidadores interagem com a criança durante o uso de telas?	1	81 (62.3%)	91.92 (±8.71)	94.95 (89.42–96.88)		
Cuidadores limitam os conteúdos que a criança assiste em telas?	0	9 (6.8%)	90.6 (±6.15)	89.18 (86.06–92.67)	50.0%	23.6%
Cuidadores limitam os conteúdos que a criança assiste em telas?	1	123 (93.2%)	92.12 (±8.01)	94.38 (89.18–96.88)		
Bayley: Cognitivo (Comp.)	Atraso	5 (3.6%)	84.66 (±9.83)	80.94 (78.44–88.46)	16.2%	7.2%
Bayley: Cognitivo (Comp.)	Normal	132 (96.4%)	92.2 (±7.62)	94.3 (89.12–96.88)		
Bayley: Ling. Receptiva	Atraso	9 (6.7%)	92.72 (±8.1)	96.88 (88.46–97.5)	89.5%	51.0%
Bayley: Ling. Receptiva	Normal	126 (93.3%)	92.35 (±6.67)	93.87 (89–96.78)		

Variável	Categoria	N (%)	Média (DP)	Mediana (IQR)	p-valor (Média)	p-valor (Mediana)
Bayley: Ling. Expressiva	Atraso	24 (17.5%)	92.08 (±7.11)	94.38 (88.82–96.94)	91.2%	93.9%
Bayley: Ling. Expressiva	Normal	113 (82.5%)	91.9 (±7.97)	93.75 (88.94–96.88)		
Bayley: Linguagem (Comp.)	Atraso	19 (13.9%)	90.22 (±12.67)	95.19 (87.44–98.27)	51.3%	78.4%
Bayley: Linguagem (Comp.)	Normal	118 (86.1%)	92.2 (±6.74)	93.75 (89–96.51)		
Bayley: Motor Fino	Atraso	7 (5.1%)	89.38 (±6.1)	88.46 (85.52–94.06)	30.1%	19.3%
Bayley: Motor Fino	Normal	130 (94.9%)	92.07 (±7.88)	94.3 (89–96.88)		
Bayley: Motor Grosso	Atraso	10 (7.4%)	89.31 (±8.11)	87.44 (82.56–96.88)	28.9%	23.5%
Bayley: Motor Grosso	Normal	126 (92.6%)	92.27 (±7.66)	94.3 (89.54–96.88)		
Bayley: Motor (Comp.)	Atraso	12 (8.8%)	89.51 (±8.3)	90.64 (83.7–96.39)	30.7%	29.8%
Bayley: Motor (Comp.)	Normal	125 (91.2%)	92.16 (±7.75)	94.23 (89.18–96.88)		

3.1.0.1 Como Interpretar a Tabela Acima

A tabela acima é um resumo completo que nos ajuda a entender o perfil das crianças no estudo e a procurar por pistas sobre o que pode influenciar sua qualidade de vida (medida pela escala PEDSQL Global).

- **Coluna N (%):** Mostra quantas crianças estão em cada categoria e qual a porcentagem

que isso representa do total. Por exemplo, podemos ver a proporção de crianças que frequentam ou não a creche.

- **Coluna ‘Média (DP)’:**** Apresenta a média da pontuação de qualidade de vida (PED-SQL Global) para cada grupo, junto com o Desvio Padrão (DP), que indica a variabilidade dos scores em torno da média.
- **Coluna ‘Mediana (IQR)’:**** Apresenta a mediana da pontuação de qualidade de vida, com a Amplitude Interquartil (IQR), que mostra a variação dos 50% centrais dos dados. Como a distribuição da nossa variável resposta não é “normal” (simétrica), a mediana é uma medida de tendência central mais robusta que a média.
- **Colunas p-valor:** São resultados de testes estatísticos que comparam os grupos.
 - O p (Média) (Teste T/ANOVA) compara as médias e assume que os dados são normalmente distribuídos.
 - O p (Mediana) (Teste de Wilcoxon/Kruskal-Wallis) compara as medianas e não assume normalidade, sendo mais adequado para nosso caso.
 - De forma simplificada, o p-valor nos ajuda a responder: “A diferença que vemos entre os grupos é real ou provavelmente aconteceu por acaso?”.

O que significam as cores?

Para não perdermos nenhuma associação potencialmente importante, decidimos (com base na reunião do dia 30/05) destacar em **verde** todas as variáveis com um **p-valor menor que 0.25** em qualquer um dos testes (média ou mediana). Isso não significa que a associação é definitivamente comprovada, mas sim que ela é um **candidato forte** para ser incluído em nossa próxima etapa de análise: a construção de um modelo estatístico multivariado. As variáveis com p-valor maior que 0.25 (em cinza) são consideradas menos prováveis de terem uma associação forte nesta análise inicial.

3.1.0.2 Por que alguns p-valores são “N/A”?

Você notou corretamente que variáveis com mais de duas categorias, como “Escolaridade da Mãe”, “Classe Social” e “Como Usa Telas”, não apresentam um p-valor, aparecendo como “N/A”.

Isso acontece porque os testes estatísticos usados para comparar três ou mais grupos (como **ANOVA** para médias ou **Kruskal-Wallis** para medianas) exigem um número mínimo de participantes em *cada* categoria para produzir um resultado válido. Em nosso conjunto de dados, algumas categorias dentro dessas variáveis têm um número muito pequeno de observações (por exemplo, apenas 5 crianças no grupo “Fundamental” de escolaridade da mãe). Quando o teste tenta rodar com um grupo tão pequeno, ele não consegue fazer uma comparação estatisticamente confiável e falha.

Isso é um problema?

Não para o nosso objetivo final. A ausência de um p-valor aqui é uma limitação da *análise univariada*, mas não nos impede de investigar a importância dessas variáveis no modelo de regressão. Como parte da nossa estratégia de modelagem, **nós incluímos proativamente todas as variáveis com p-valor “N/A” como candidatas**, garantindo que elas tenham a oportunidade de serem avaliadas no contexto multivariado, onde o modelo consegue lidar melhor com essas situações.

3.2 Variáveis Numéricas

Table 3: Sumário estatístico das variáveis numéricas/resposta

pedsqlfisico36	pedsqlpsico36	pedsqlglobal36
Min. : 59.38	Min. : 25.00	Min. : 48.44
1st Qu.: 87.50	1st Qu.: 86.54	1st Qu.: 88.94
Median : 93.75	Median : 92.31	Median : 93.99
Mean : 93.50	Mean : 90.36	Mean : 91.93
3rd Qu.:100.00	3rd Qu.: 98.08	3rd Qu.: 96.88
Max. :100.00	Max. :100.00	Max. :100.00

4 Variáveis resposta

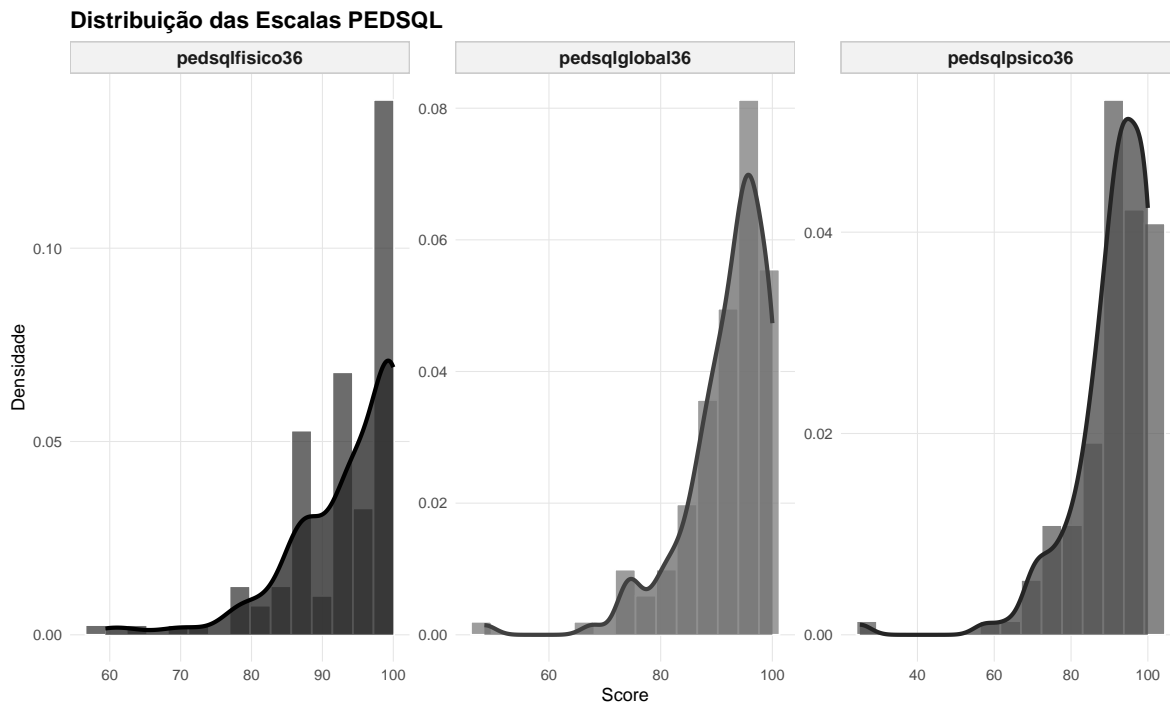


Figure 1: Distribuição das Escalas PEDSQL

4.1 Teste de Normalidade

Table 4: Teste de Normalidade de Shapiro-Wilk para as Escalas PEDSQL

	Escala	N	Media	DP	W_Shapiro	p_valor	Interpretacao
W...1	PEDSQL Físico	137	93.50	7.91	0.7920	<0.001	Não Normal ($p < 0.05$)
W...2	PEDSQL Global	137	91.93	7.80	0.8350	<0.001	Não Normal ($p < 0.05$)
W...3	PEDSQL Psicossocial	137	90.36	10.40	0.7887	<0.001	Não Normal ($p < 0.05$)

5 Boxplot das Escalas PEDSQL

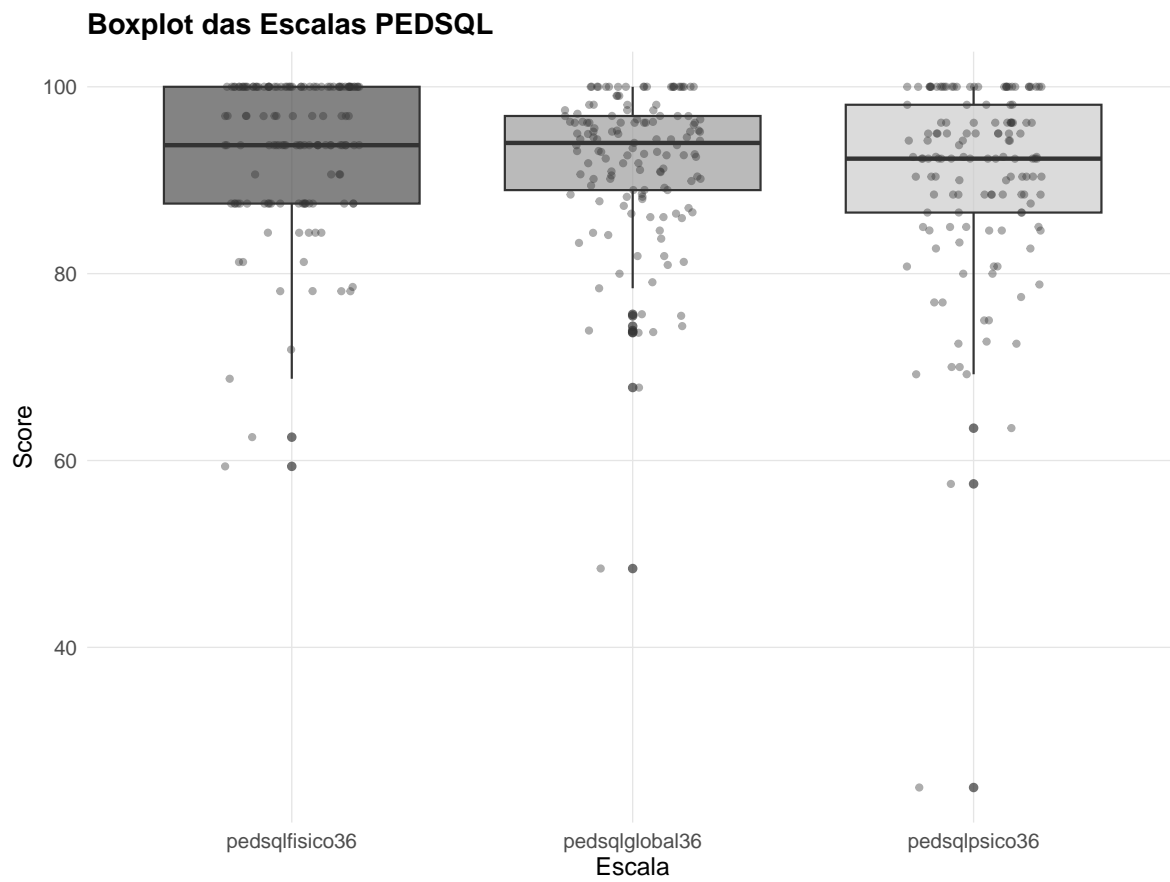


Figure 2: Boxplot das Escalas PEDSQL

6 Associação das variáveis resposta (pedsqlglobal36)

6.1 Por gênero

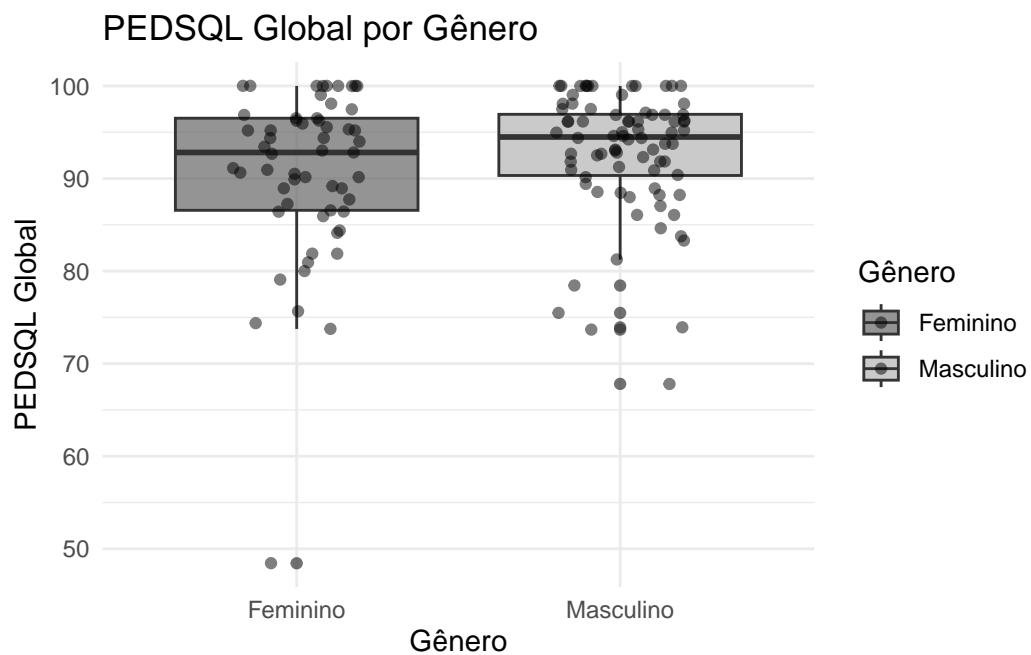


Figure 3: PEDSQL Global por Gênero

6.2 Por Faixa Etária Materna

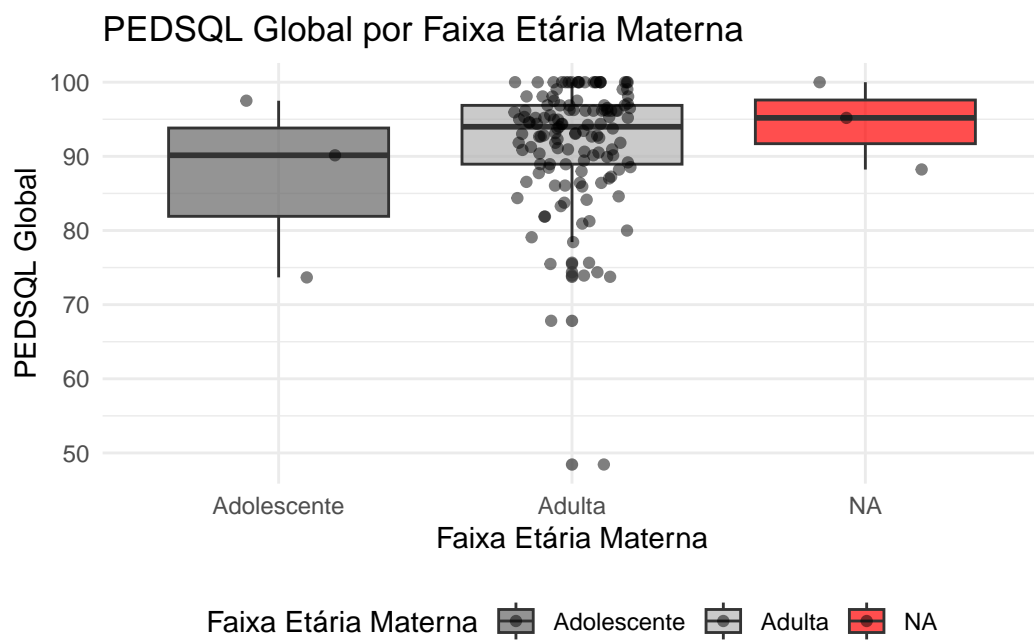


Figure 4: PEDSQL Global por Faixa Etária Materna

6.3 Por Idade Gestacional

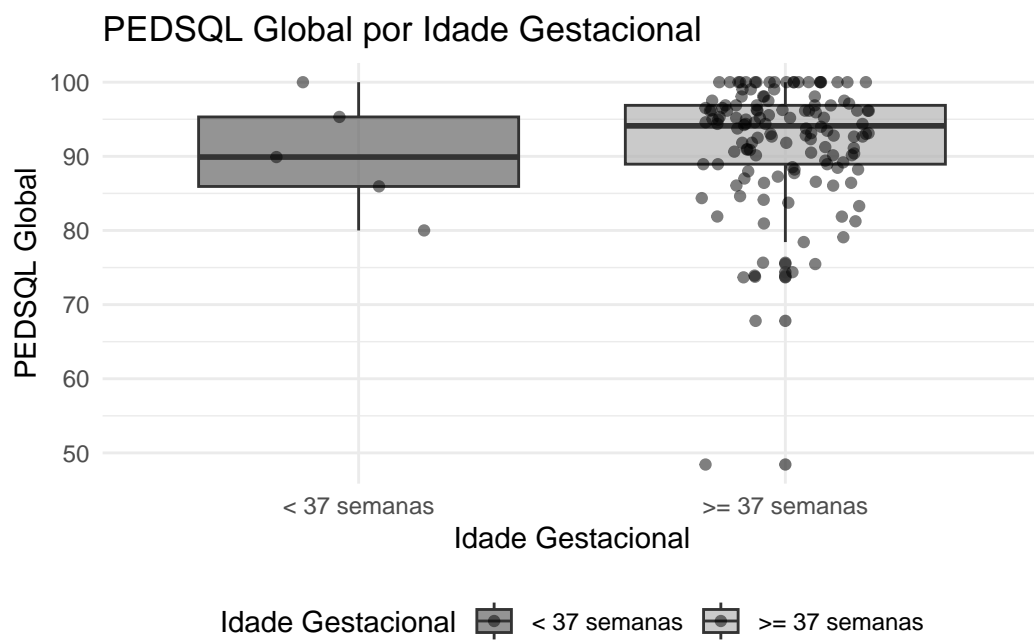


Figure 5: PEDSQL Global por Idade Gestacional

6.4 Por Tipo de Parto

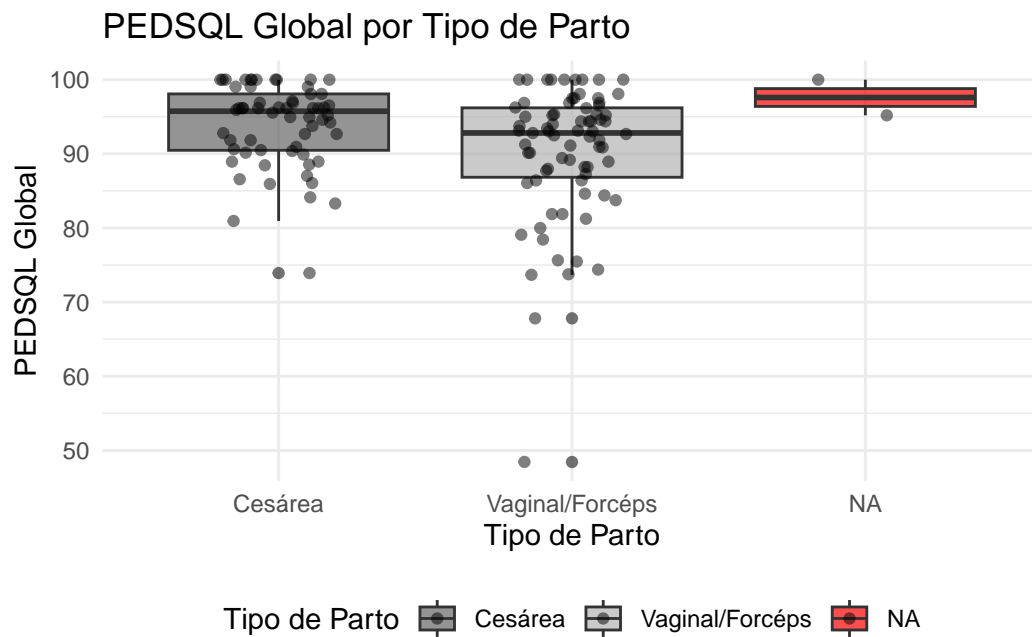


Figure 6: PEDSQL Global por Tipo de Parto

7 Boxplots das Variáveis Categóricas vs PEDSQL Global

7.1 Boxplots para Variáveis com Associação Significativa ($p < 0.05$)

Como a variável resposta (PEDSQL Global) não apresentou distribuição normal, os gráficos a seguir mostram as variáveis que tiveram uma associação estatisticamente significativa ($p < 0.05$) com a qualidade de vida, com base no teste de comparação de medianas (Teste de Wilcoxon/Kruskal-Wallis), que é mais robusto para estes casos.

7.1.1 Via de Parto ($p = 0.015$)

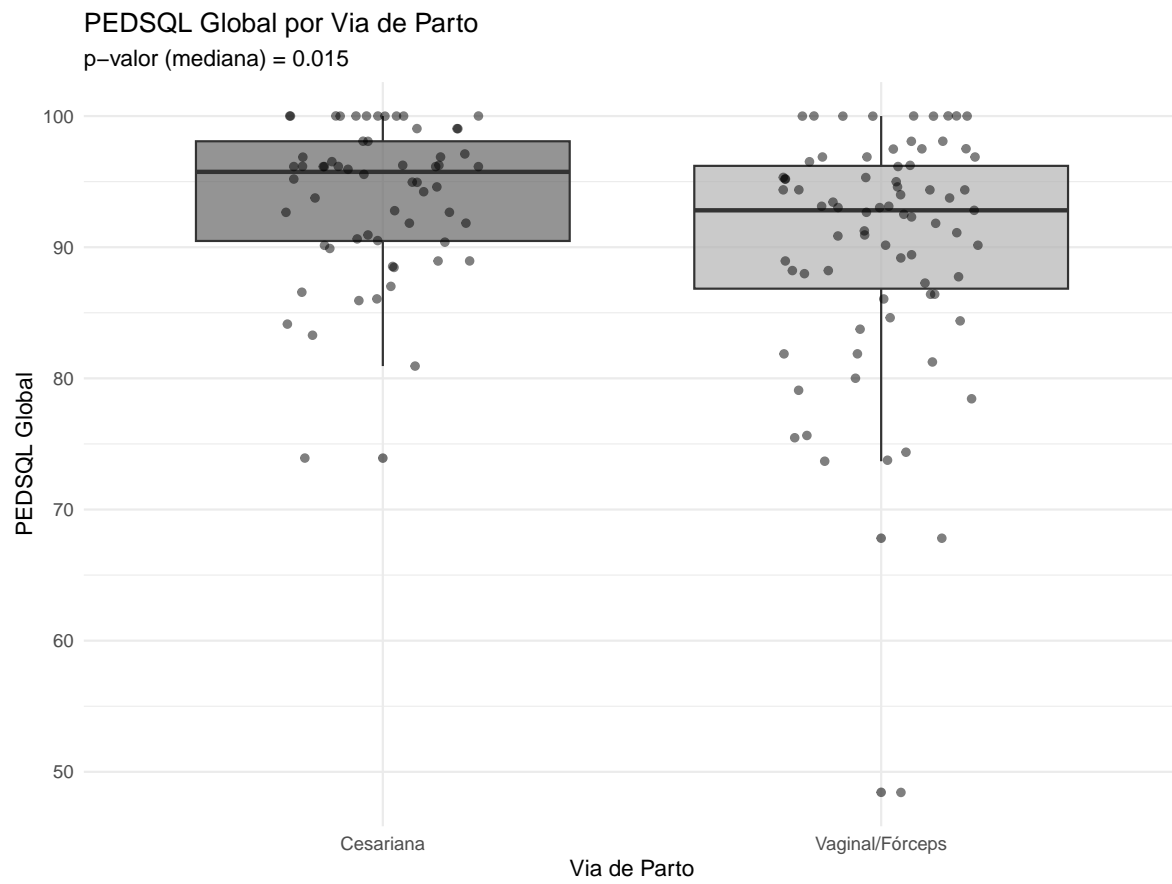


Figure 7: PEDSQL Global por Via de Parto

7.1.2 Horário Fixo para Telas (p = 0.031)

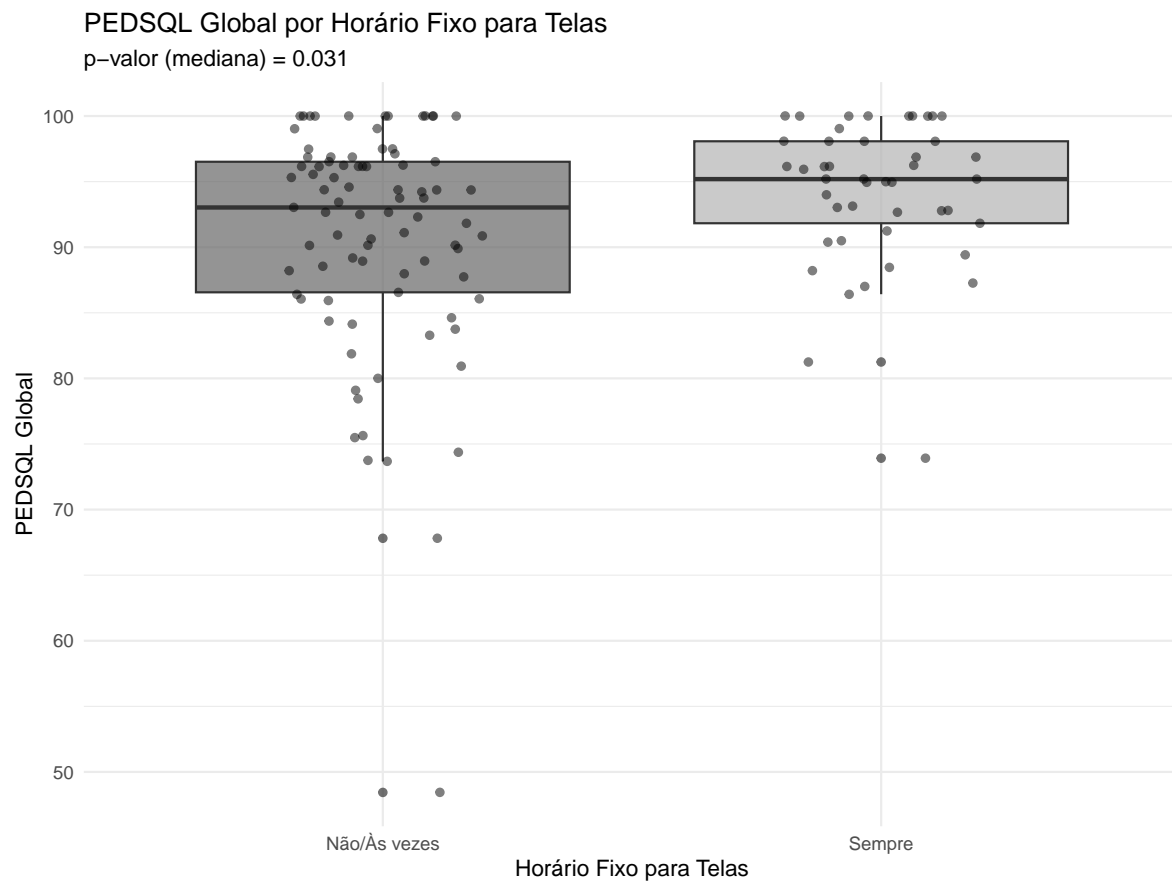


Figure 8: PEDSQL Global por Horário Fixo para Telas

7.1.3 Limite de Tempo de Tela ($p = 0.003$)

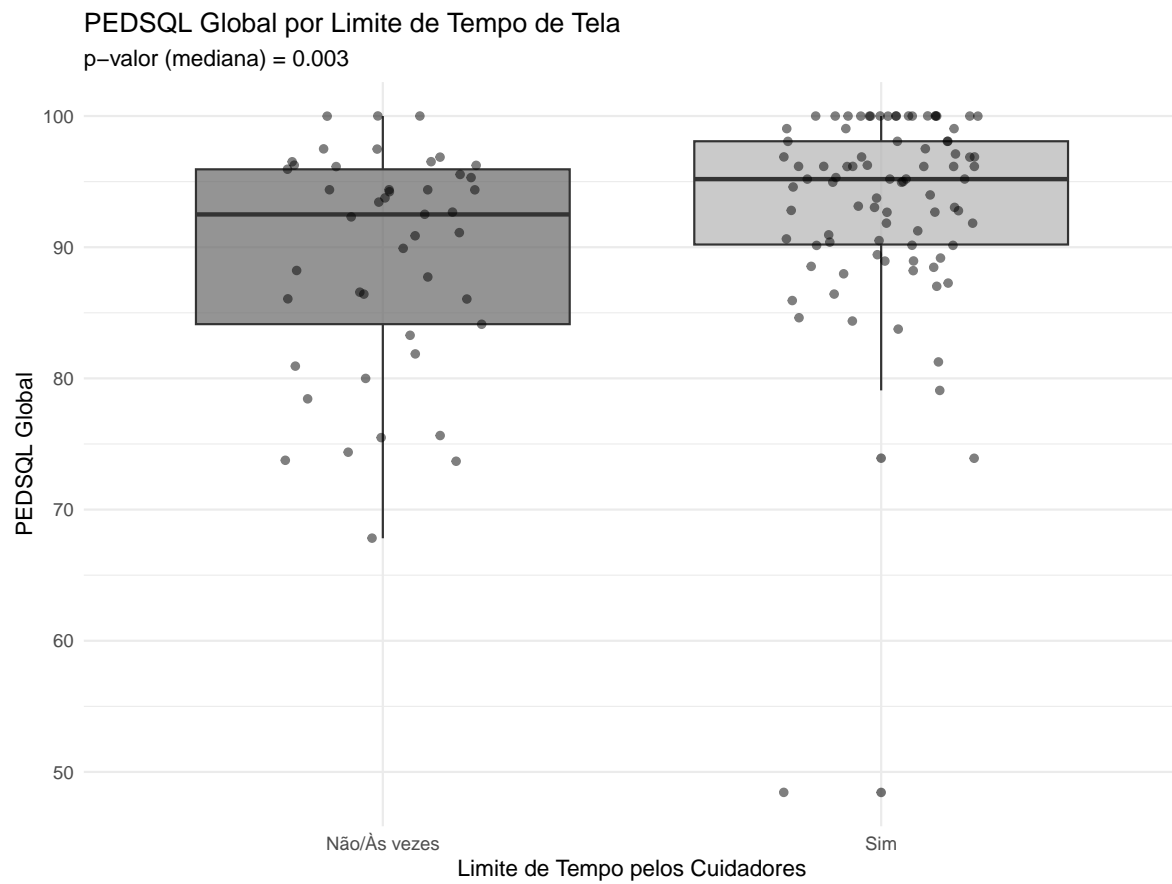


Figure 9: PEDSQL Global por Limite de Tempo de Tela

7.2 Variáveis Demográficas e Perinatais

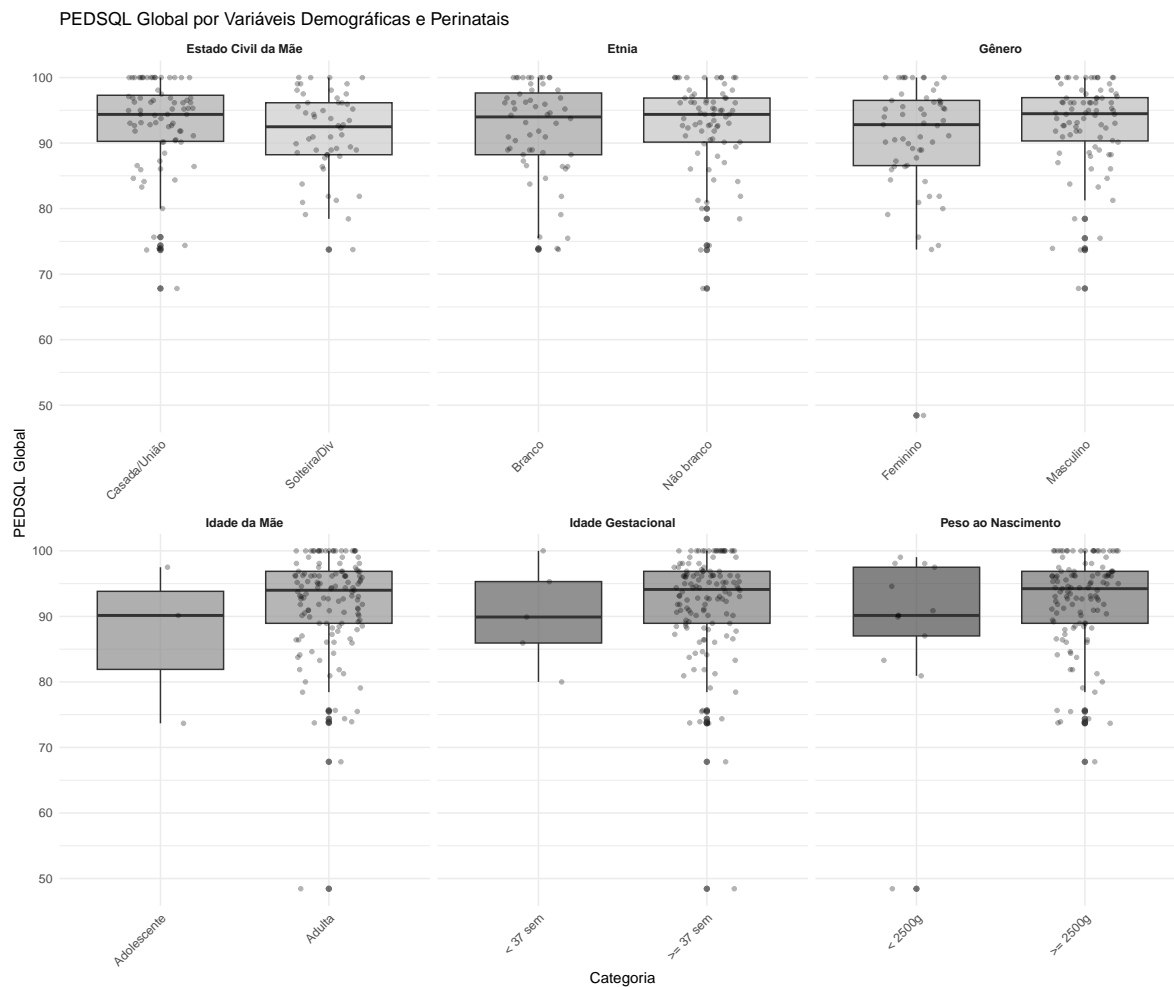


Figure 10: PEDSQL Global por Variáveis Demográficas e Perinatais

7.3 Variáveis de Desenvolvimento (Bayley)

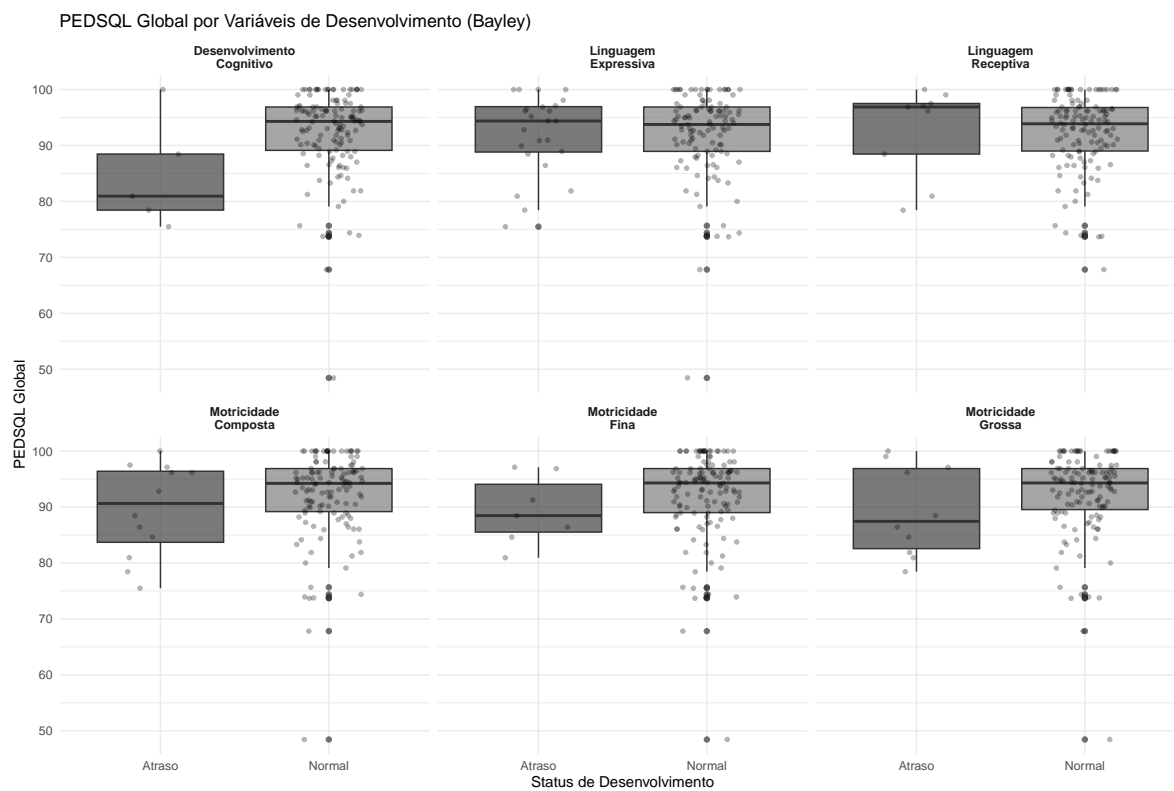


Figure 11: PEDSQL Global por Variáveis de Desenvolvimento

7.4 Outras Variáveis de Interesse

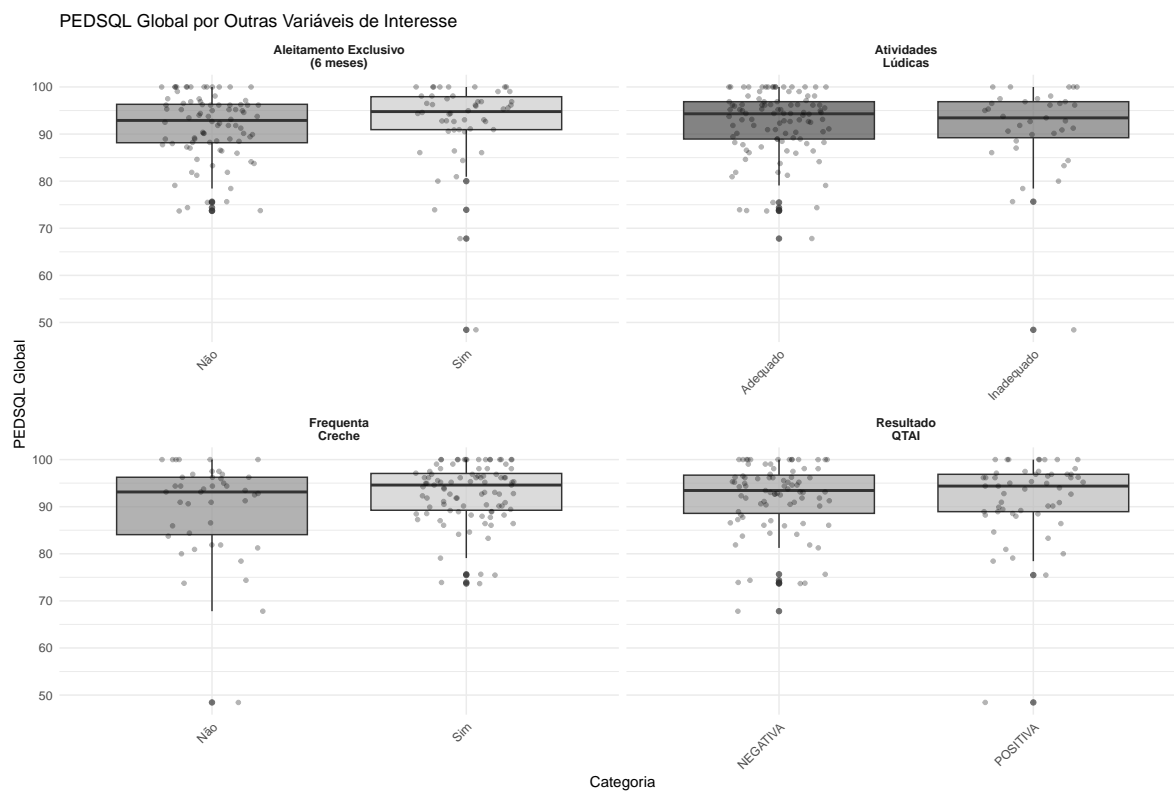


Figure 12: PEDSQL Global por Outras Variáveis

8 Boxplot Completo - Todas as Variáveis Categóricas

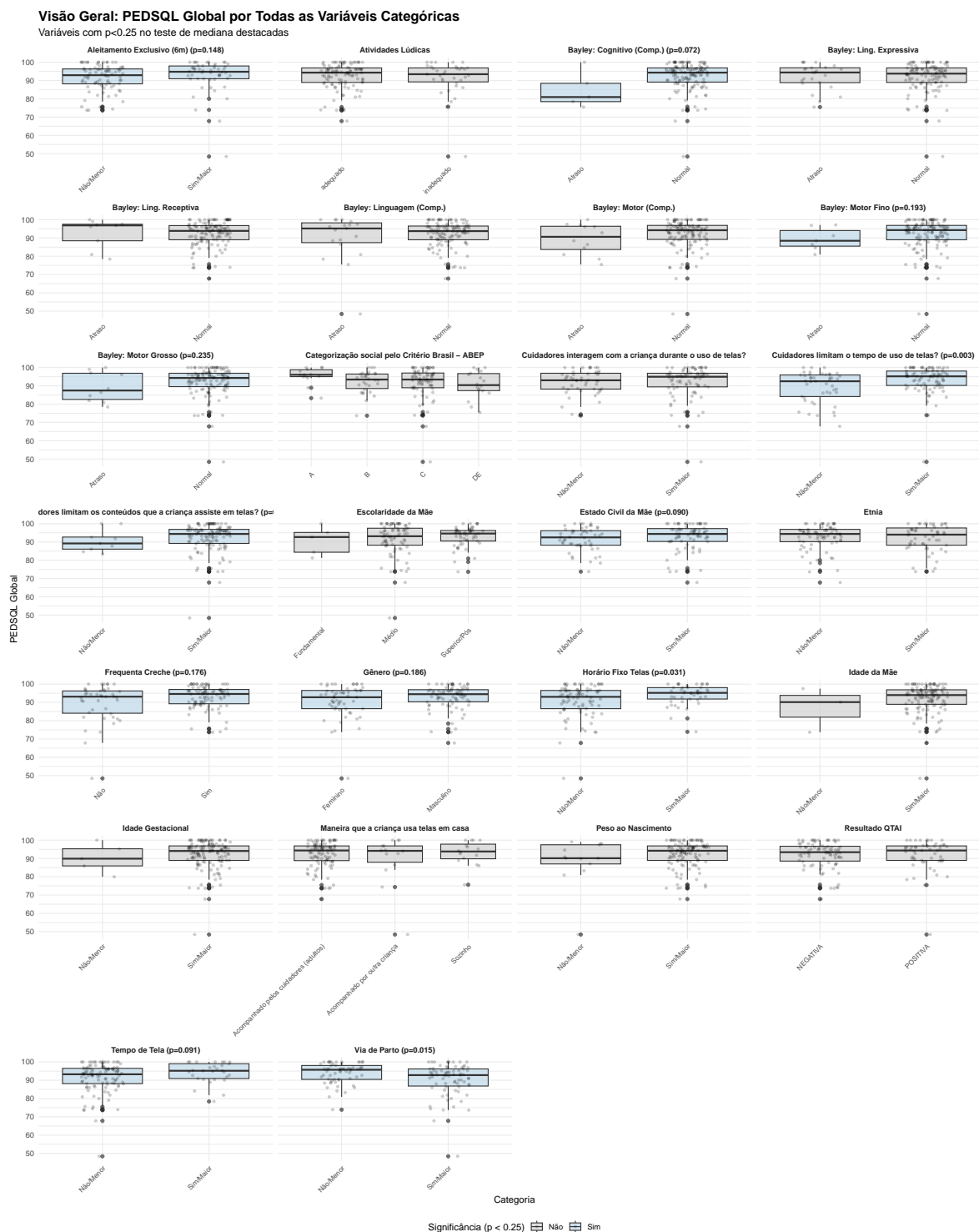


Figure 13: PEDSQL Global por Todas as Variáveis Categóricas

9 Análise de Colinearidade e Correlação entre Variáveis de Tela

Para as variáveis relacionadas ao uso de telas, é fundamental verificar a existência de **multicolinearidade**. Como essas variáveis medem aspectos relacionados dos hábitos de uso de tela (tempo, horário, supervisão, etc.), existe uma possibilidade de que estejam fortemente correlacionadas entre si.

Por que isso é importante para variáveis de tela?

- **Sobreposição conceitual:** Variáveis como “limitação de tempo” e “limitação de conteúdo” podem medir aspectos similares do controle parental sobre o uso de telas.
- **Interpretação clara:** Queremos entender qual aspecto específico do uso de telas tem maior associação com a qualidade de vida.

Para investigar isso, realizamos duas análises focadas nas variáveis de tela:

1. **Matriz de Correlação:** Visualiza a força da relação linear entre as variáveis de tela.
2. **Fator de Inflação da Variância (VIF):** Identifica se alguma variável de tela é redundante com as outras.

9.1 Matriz de Correlação - Variáveis de Tela

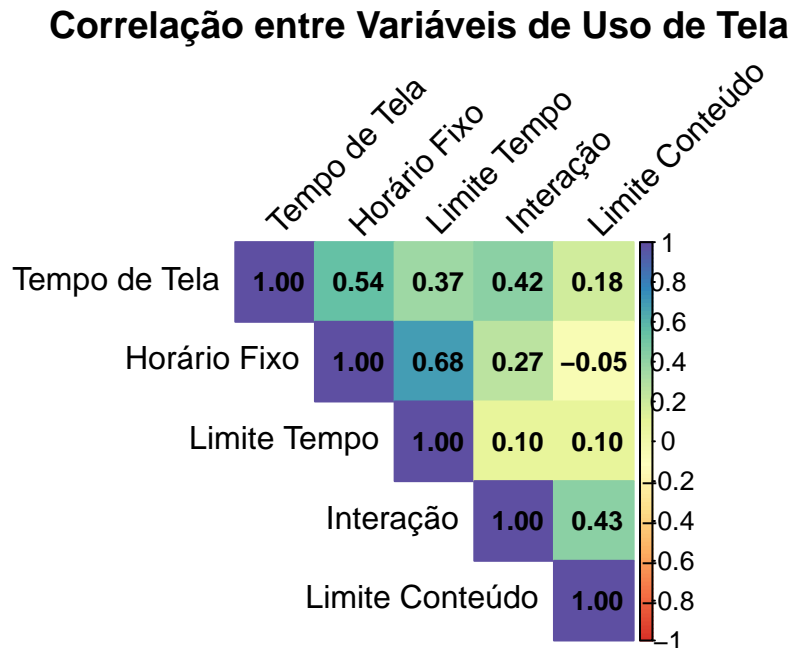


Figure 14: Matriz de correlação entre as variáveis de uso de tela

9.1.1 Interpretação da Matriz de Correlação

A matriz de correlação acima mostra a força da associação linear entre as diferentes variáveis relacionadas ao uso de telas. Os valores variam de -1 a +1:

- **Valores próximos de +1:** Indicam uma correlação positiva forte (quando uma variável aumenta, a outra também tende a aumentar).
- **Valores próximos de -1:** Indicam uma correlação negativa forte (quando uma variável aumenta, a outra tende a diminuir).
- **Valores próximos de 0:** Indicam ausência de correlação linear.

O que procuramos: Correlações muito altas ($|r| > 0.7$) podem indicar redundância entre variáveis, sugerindo que elas medem aspectos muito similares do comportamento de uso de telas.

9.2 Análise VIF - Variáveis de Tela

O Fator de Inflação da Variância (VIF) quantifica o quanto a variância de um coeficiente de regressão aumenta devido à colinearidade com outras variáveis.

Table 5: Análise VIF para Variáveis de Uso de Tela

Variável	VIF	Interpretação	Status
Tempo de Tela	1.63	Baixa colinearidade	Aceitável
Horário Fixo	2.62	Colinearidade moderada	Aceitável
Limite Tempo	2.05	Baixa colinearidade	Aceitável
Interação	1.55	Baixa colinearidade	Aceitável
Limite Conteúdo	1.38	Baixa colinearidade	Aceitável

9.2.1 Interpretação da Análise VIF

O VIF nos ajuda a identificar se alguma variável de tela é redundante com as outras:

- **VIF < 2.5:** Baixa colinearidade - a variável contribui com informação única.
- **VIF 2.5-5:** Colinearidade moderada - ainda aceitável, mas deve ser monitorada.
- **VIF 5-10:** Colinearidade alta - pode causar problemas na interpretação do modelo.
- **VIF > 10:** Colinearidade muito alta - a variável é altamente redundante e deve ser considerada para remoção.

Implicações para o modelo: Se encontrarmos variáveis de tela com VIF alto, isso sugere que elas medem aspectos muito similares do comportamento de uso de telas. Nesse caso, podemos optar por manter apenas a variável mais relevante clinicamente ou criar um índice composto.

10 Modelagem de Regressão Multivariada

Nesta seção, nosso objetivo é construir um modelo estatístico para entender quais fatores estão mais fortemente associados à qualidade de vida das crianças, medida pelo escore `pedsqglobal36`. Como vimos na análise univariada, diversas variáveis parecem ter alguma associação, mas um modelo multivariado nos ajuda a entender a contribuição de cada uma quando consideramos todas ao mesmo tempo.

10.0.1 Estratégia de Modelagem

Para encontrar o modelo mais robusto e interpretável, adotaremos a seguinte estratégia:

1. **Seleção de Variáveis Candidatas:** Começaremos com as variáveis que mostraram alguma associação potencial na análise univariada. Adotamos o critério de incluir todas as variáveis com **p-valor** < 0.25 em qualquer um dos testes (comparação de médias ou medianas). Essa abordagem mais liberal na fase inicial evita a exclusão prematura de variáveis que podem ser importantes no contexto multivariado. Também incluiremos variáveis cujo p-valor não pôde ser calculado (N/A) por limitações amostrais.
2. **Construção e Comparação de Modelos:** Construiremos e compararemos três modelos diferentes:
 - **Modelo 1 (Completo):** Um modelo inicial com todas as variáveis candidatas.
 - **Modelo 2 (Hierárquico):** Um modelo em blocos, adicionando grupos de variáveis em uma ordem teórica (ex: fatores perinatais, depois contexto social, e por último, desenvolvimento).
 - **Modelo 3 (Stepwise):** Um modelo otimizado por um algoritmo (AIC) para fins de exploração e comparação.
3. **Diagnóstico e Interpretação:** Todos os modelos serão avaliados quanto à sua adequação (ex: análise de resíduos) e a multicolinearidade será verificada (VIF). O modelo final será escolhido com base no ajuste estatístico, na simplicidade e, crucialmente, na sua capacidade de responder às perguntas da pesquisa de forma clara.

10.0.2 Fase 1: Preparação e Construção do Modelo Completo

10.0.2.1 Seleção de Variáveis Candidatas

Primeiro, vamos identificar programaticamente as variáveis candidatas com base no critério definido: variáveis com **p-valor (Mediana) < 0.25** na análise univariada e aquelas cujo p-valor não pôde ser calculado (N/A).

```
# Inverter o mapa de variáveis para facilitar a busca pelo nome
map_vars_invertido <- setNames(names(vars_info), unlist(vars_info))

# Filtrar a tabela de resultados para obter variáveis candidatas
# Critério: p-valor da mediana < 0.25 ou p-valor não aplicável (N/A)
vars_candidatas_nomes <- tabela_completa %>%
  filter(p_valor_mediana < 0.25 | is.na(p_valor_mediana)) %>%
  distinct(Variável) %>%
  pull(Variável)

# Mapear os nomes de volta para os nomes das variáveis no dataset
vars_candidatas <- map_vars_invertido[vars_candidatas_nomes] %>%
  unname()

# Mostrar as variáveis selecionadas
cat("Variáveis candidatas selecionadas para o modelo (p-mediana < 0.25 ou N/A):\n")
```

Variáveis candidatas selecionadas para o modelo (p-mediana < 0.25 ou N/A):

```
cat(paste("-", vars_candidatas, "\n", collapse = "\n"), sep = "\n")
```

```
- ` partocat `
- ` lme6 `
- ` estadocivilmae24 `
- ` escolamaecat36 `
- ` abepcat36 `
- ` genero `
- ` creche36 `
- ` tempotelacat36 `
- ` horacertacat36 `
- ` telacasa36 `
- ` telalimitetempocat36 `
- ` telalimiteconteudocat36 `
- ` classcogcomp36 `
```

- ` classmfbal36 `
- ` classmgbal36 `

10.0.2.2 Modelo 1: Modelo de Regressão Completo

Agora, ajustamos o primeiro modelo de regressão linear incluindo todas as variáveis candidatas que selecionamos. Este modelo nos dará uma visão inicial de quais variáveis se destacam quando analisadas em conjunto.

```
# Criar a fórmula do modelo dinamicamente
formula_completa <- as.formula(
  paste("pedsqlglobal36 ~", paste(vars_candidatas, collapse = " + "))
)
cat("Fórmula do modelo:\n")
```

Fórmula do modelo:

```
print(formula_completa)
```

```
pedsqlglobal36 ~ partocat + lme6 + estadocivilmae24 + escolamaecat36 +
  abepcat36 + genero + creche36 + tempotelacat36 + horacertacat36 +
  telacasa36 + telalimitetempocat36 + telalimiteconteudocat36 +
  classcogcomp36 + classmfbal36 + classmgbal36
```

```
cat("\n")
```

```
# Ajustar o modelo linear
modelo_completo <- lm(formula_completa, data = data, na.action = na.exclude)

# Criar a tabela de regressão para ser usada tanto na tabela quanto no gráfico
tbl_modelo_completo <- tbl_regression(
  modelo_completo,
  exponentiate = FALSE,
  pvalue_fun = ~style_pvalue(.x, digits = 3),
  intercept = TRUE
)

# Apresentar o modelo em formato de tabela
tbl_modelo_completo %>%
  bold_p(t = 0.05) %>%
```

```

add_glance_table(
  include = c(adj.r.squared, AIC),
  label = list(adj.r.squared ~ "R2 Ajustado", AIC ~ "AIC")
) %>%
as_gt() %>%
tab_header(
  title = "Modelo Completo: Fatores Associados à Qualidade de Vida",
  subtitle = "Variável dependente: PEDSQL Global (pedsqlglobal36)"
) %>%
tab_source_note(md("**Nota:** Coeficientes representam a mudança no escore PEDSQL Global. V

```

10.0.2.3 Visualização do Modelo: Forest Plot

Para facilitar a interpretação dos resultados do modelo, podemos usar um **Forest Plot**. Este gráfico mostra o efeito de cada variável de forma visual, sendo uma alternativa poderosa à tabela de regressão.

Como interpretar o Forest Plot:

- **Eixo Vertical:** Lista todas as variáveis (e suas categorias) incluídas no modelo.
- **Eixo Horizontal:** Representa a escala dos coeficientes. O valor zero, marcado por uma linha pontilhada vertical, é a nossa referência de “nenhum efeito”.
- **Pontos:** Cada ponto representa a estimativa do coeficiente para aquela variável — ou seja, o quanto o escore de qualidade de vida muda, em média, para aquela categoria, mantendo as outras variáveis constantes.
- **Linhas Horizontais:** As linhas que passam por cada ponto são os **intervalos de confiança de 95%**. Elas nos dão uma ideia da incerteza em torno da nossa estimativa.

Regra prática: Se a linha horizontal de uma variável **cruza a linha do zero**, o efeito daquela variável não é considerado estatisticamente significativo (equivalente a um p-valor > 0.05). Se a linha está inteiramente à direita ou à esquerda do zero, o efeito é significativo.

Table 6: Modelo de Regressão Linear Completo para o Escore PEDSQL Global

Modelo Completo: Fatores Associados à Qualidade de Vida

Variável dependente: PEDSQL Global (pedsqlglobal36)

Characteristic	Beta	95% CI	p-value
(Intercept)	82	71, 94	<0.001
partocat	-2.0	-4.7, 0.69	0.145
lme6	2.3	-0.33, 5.0	0.085
estadocivilmae24	0.10	-2.7, 2.9	0.945
escolamaecat36	-0.37	-2.9, 2.2	0.777
abepcat36			
A	—	—	
B	-2.5	-7.4, 2.4	0.323
C	-2.9	-7.3, 1.5	0.194
DE	-2.5	-8.3, 3.3	0.396
genero			
Feminino	—	—	
Masculino	1.1	-1.5, 3.8	0.401
creche36			
Não	—	—	
Sim	1.9	-1.0, 4.8	0.206
tempotelacat36	1.4	-2.2, 5.0	0.446
horacertacat36	0.84	-2.1, 3.8	0.577
telacasa36			
Acompanhado pelos cuidadores (adultos)	—	—	
Acompanhado por outra criança	3.3	-1.0, 7.7	0.133
Sozinho	1.7	-2.2, 5.5	0.395
telalimitetempocat36	3.3	0.17, 6.4	0.039
telalimiteconteudocat36	3.8	-2.1, 9.6	0.206
classcogcomp36	1.9	-6.1, 9.9	0.638
classmfbal36	1.6	-5.5, 8.7	0.652
classmgbal36	1.0	-5.2, 7.3	0.742
R ² Ajustado	0.088		
AIC	797		

Abbreviation: CI = Confidence Interval

Nota: Coeficientes representam a mudança no escore PEDSQL Global. Valores em negrito indicam significância estatística ($p < 0.05$).

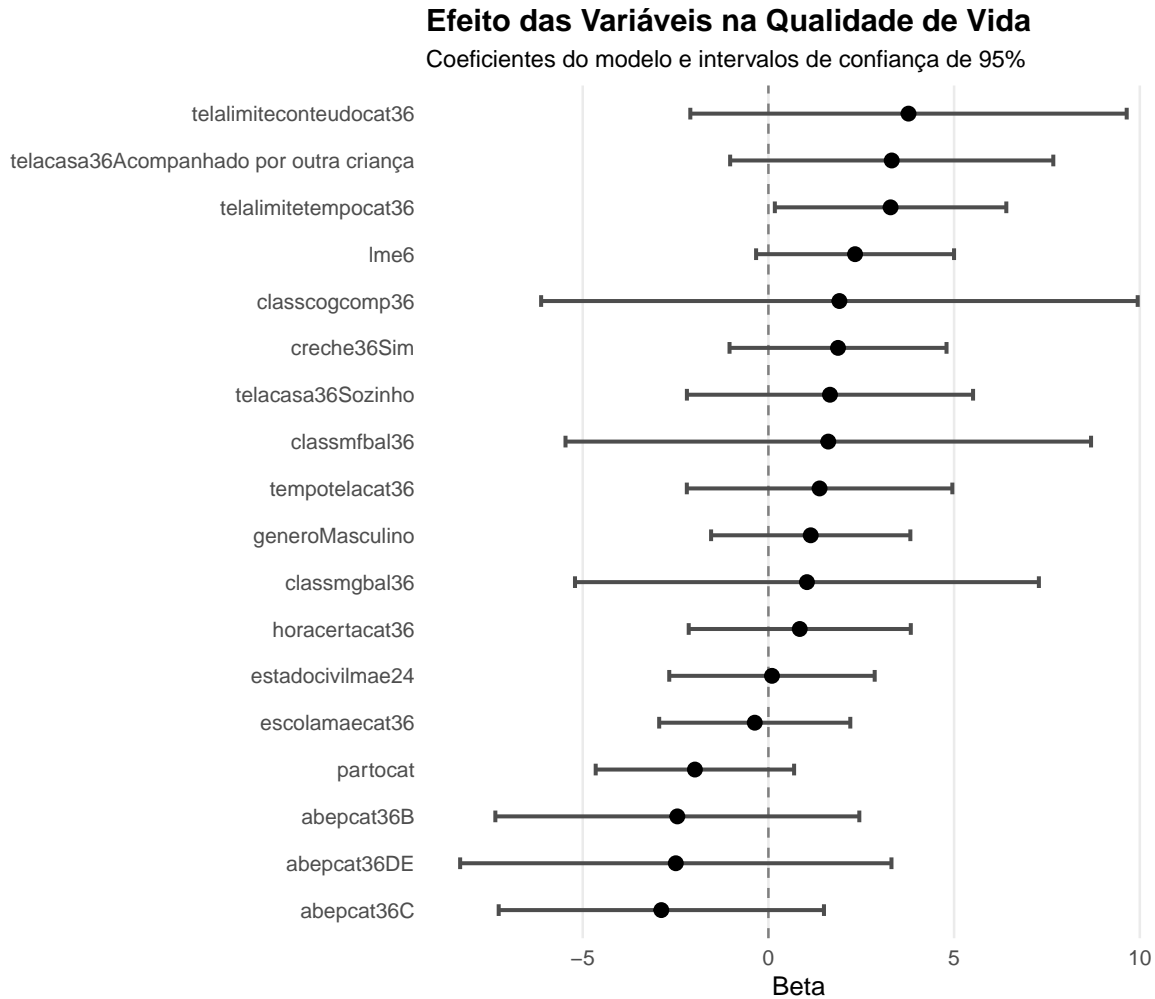


Figure 15: Forest Plot dos coeficientes do Modelo Completo

10.0.3 Modelo 2: Modelo com Seleção Stepwise (Exploratório)

Diferente da abordagem anterior, onde partimos de variáveis pré-selecionadas, agora vamos adotar uma estratégia puramente exploratória. O método **Stepwise** construirá um modelo do zero, tendo a liberdade de testar **todas as variáveis disponíveis no conjunto de dados original**.

O método funciona passo a passo, adicionando ou removendo variáveis com base no critério AIC (*Akaike Information Criterion*), que busca o melhor equilíbrio entre o ajuste do modelo aos dados e sua simplicidade (número de variáveis). O resultado é um modelo otimizado

Table 7: Modelo de Regressão com Seleção Stepwise (AIC) a partir de todas as variáveis

Modelo Stepwise: Fatores Associados à Qualidade de Vida

Variáveis selecionadas automaticamente (AIC) a partir de todo o dataset

Characteristic	Beta	95% CI	p-value
(Intercept)	86	79, 93	<0.001
partocat	-2.4	-4.9, 0.19	0.070
telalimitetempocat36	4.0	1.3, 6.7	0.004
classcogcomp36	7.5	-1.3, 16	0.095
claslrbal36	-7.2	-14, -0.50	0.035
classmgbal36	5.0	-0.38, 10	0.068
R ² Ajustado	0.146		
AIC	747		

Abbreviation: CI = Confidence Interval

estatisticamente, que pode revelar associações inesperadas, mas que deve ser interpretado com cautela, pois não parte de hipóteses teóricas.

Fórmula do modelo final selecionado pelo Stepwise:

```
pedsqlglobal36 ~ partocat + telalimitetempocat36 + classcogcomp36 +
  claslrbal36 + classmgbal36
```

10.0.3.1 Visualização do Modelo Stepwise: Forest Plot

Assim como fizemos para o modelo completo, vamos visualizar os resultados do modelo stepwise para facilitar a interpretação.

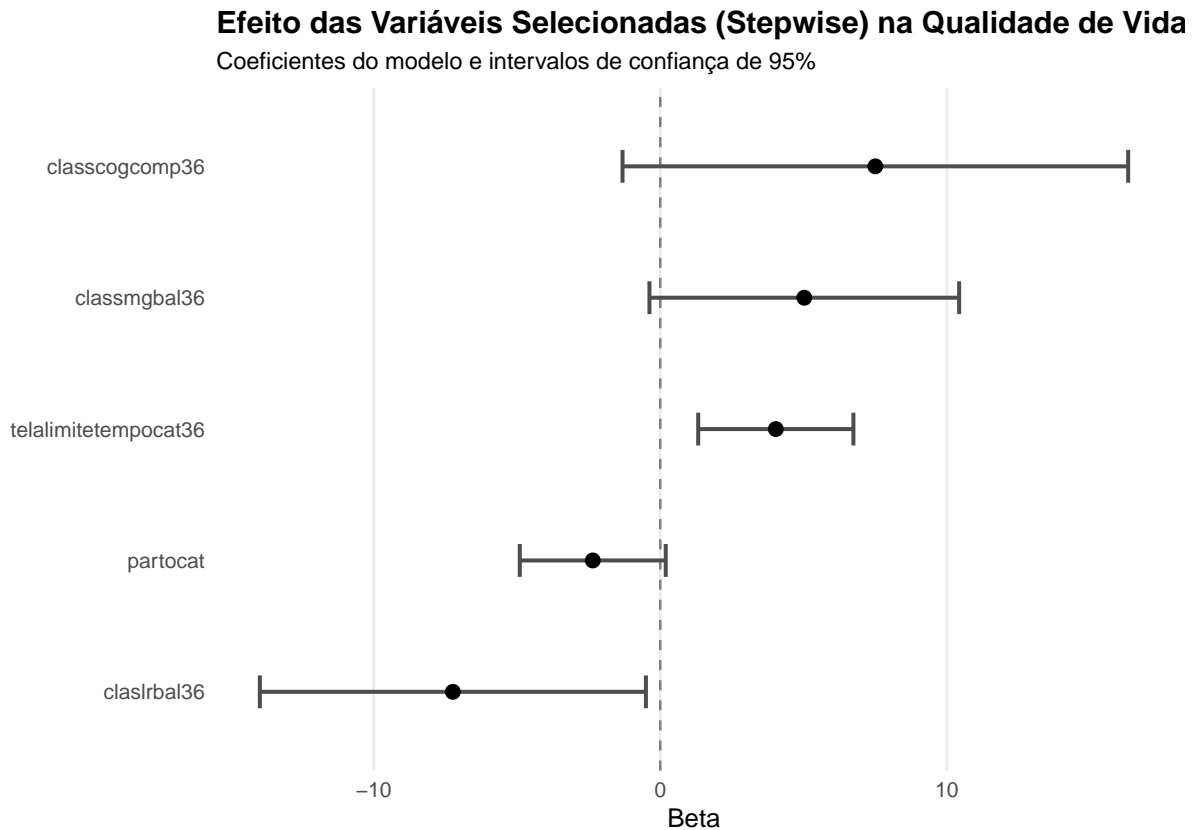


Figure 16: Forest Plot dos coeficientes do Modelo Stepwise

Com os resultados dos modelos completo e stepwise em mãos, o próximo passo da nossa estratégia será construir o modelo hierárquico. Isso nos permitirá comparar as três abordagens e escolher a mais adequada.

10.0.4 Modelo 3: Modelo com Seleção Stepwise (A partir das Variáveis Pré-Seleccionadas)

Finalmente, vamos criar um terceiro modelo para comparação. Este modelo também usará a seleção **Stepwise**, mas em vez de começar com todas as variáveis do dataset, ele partirá do nosso `modelo_completo` original — aquele que contém apenas as variáveis que pré-selecionamos com base na análise univariada ($p\text{-mediana} < 0.25$ ou N/A).

O objetivo aqui é refinar nossa seleção inicial, permitindo que o algoritmo remova variáveis que, embora promissoras na análise univariada, podem não ser significativas quando consideradas em conjunto com as outras.

Table 8: Modelo de Regressão com Seleção Stepwise (a partir das variáveis pré-selecionadas)

Modelo Stepwise Refinado: Fatores Associados à Qualidade de Vida

Variáveis selecionadas (AIC) a partir do grupo pré-selecionado

Characteristic	Beta	95% CI	p-value
(Intercept)	87	81, 92	<0.001
partocat	-2.5	-5.1, -0.01	0.049
telalimitetempocat36	4.3	1.6, 6.9	0.002
classmfbal36	4.7	-0.84, 10	0.096
R ² Ajustado	0.129		
AIC	747		

Abbreviation: CI = Confidence Interval

Fórmula do modelo final selecionado pelo Stepwise (Refinado):

`pedsqlglobal36 ~ partocat + telalimitetempocat36 + classmfbal36`

10.0.4.1 Visualização do Modelo Stepwise Refinado: Forest Plot

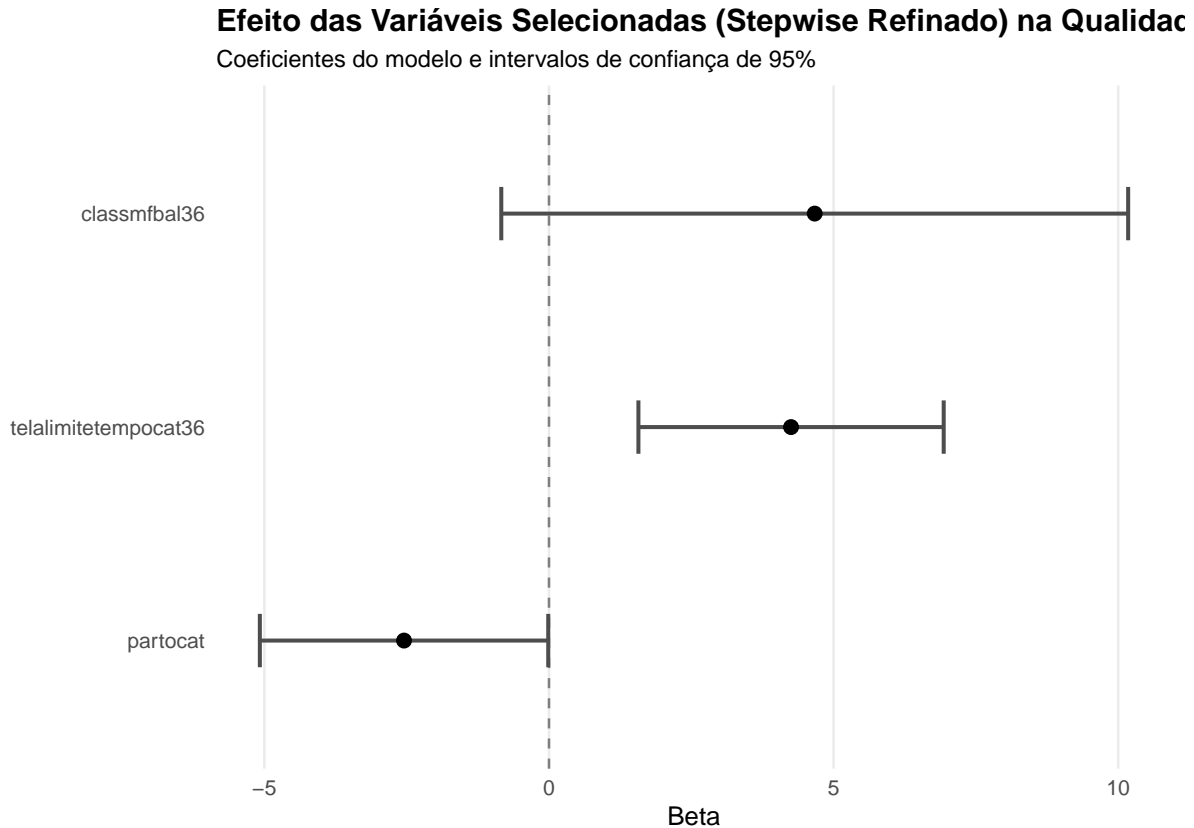


Figure 17: Forest Plot dos coeficientes do Modelo Stepwise Refinado

10.0.5 Modelo 4: Modelo Hierárquico Sequencial

Conforme a estratégia definida com a orientadora, construiremos o modelo final de forma sequencial e teórica. A ideia é adicionar blocos de variáveis em uma ordem lógica. Uma variável que se mostrar significativa ($p < 0.10$) em qualquer etapa será acumulada e mantida em todos os passos subsequentes.

Os blocos são definidos da seguinte forma: - **Bloco 1 (Saúde Perinatal):**

- ``partocat`` (Via de Parto)
- ``lme6`` (Aleitamento Exclusivo aos 6m)

- **Bloco 2 (Contexto Social):**

- `estadocivilmae24` (Estado Civil da Mãe)
- `escolamaecat36` (Escolaridade da Mãe)

Table 9: Modelo Hierárquico - Passo 1: Saúde Perinatal

Passo 1: Gênero + Saúde Perinatal

Characteristic	Beta	95% CI	p-value
(Intercept)	92	89, 95	<0.001
genero			
Feminino	—	—	
Masculino	1.8	-0.94, 4.5	0.2
partocat	-3.5	-6.2, -0.79	0.012
lme6	1.2	-1.5, 4.0	0.4
R ² Ajustado	0.050		
AIC	919		

Abbreviation: CI = Confidence Interval

- abepcat36 (Classe Social ABEP)
- creche36 (Frequenta Creche)

- **Bloco 3 (Hábitos de Tela):**

- tempotelacat36 (Tempo de Tela)
- horacertacat36 (Horário Fixo para Telas)
- telacasa36 (Como Usa Telas)
- telalimitetempocat36 (Limite de Tempo de Tela)
- telalimiteconteudocat36 (Limite de Conteúdo da Tela)

- **Bloco 4 (Desenvolvimento):**

- classcogcomp36 (Bayley: Cognitivo Composto)
- classmfbal36 (Bayley: Motor Fino)
- classmgbal36 (Bayley: Motor Grosso)

A variável **genero** será incluída como controle em todas as etapas.

10.0.5.1 Passo 1: Gênero + Bloco de Saúde Perinatal

Começamos com o modelo mais básico, incluindo apenas o gênero e os fatores perinatais.

No primeiro passo, a variável **partocat** (Via de Parto) mostrou-se significativa ($p=0.012$). Conforme nossa regra, ela será mantida em todas as etapas subsequentes do modelo.

Table 10: Modelo Hierárquico - Passo 2: Contexto Social

Passo 2: Gênero + Sig. (Passo 1) + Contexto Social

Characteristic	Beta	95% CI	p-value
(Intercept)	94	88, 100	<0.001
genero			
Feminino	—	—	
Masculino	1.2	-1.2, 3.7	0.3
partocat	-2.3	-4.9, 0.20	0.070
estadocivilmae24	1.3	-1.3, 3.8	0.3
creche36			
Não	—	—	
Sim	1.6	-1.1, 4.2	0.2
escolamaecat36	-0.31	-2.8, 2.1	0.8
abepcat36			
A	—	—	
B	-2.8	-7.3, 1.8	0.2
C	-2.8	-6.9, 1.3	0.2
DE	-2.8	-8.2, 2.6	0.3
R ² Ajustado	0.030		
AIC	844		

Abbreviation: CI = Confidence Interval

10.0.5.2 Passo 2: Adição do Bloco de Contexto Social

Agora, adicionamos as variáveis do contexto social, mantendo a variável significativa da etapa anterior.

Após adicionar o bloco social, **partocat** manteve sua significância ($p=0.070$). Nenhuma nova variável se mostrou significativa nesta etapa. Seguimos para o próximo bloco, mantendo **partocat** no modelo.

10.0.5.3 Passo 3: Adição do Bloco de Hábitos de Tela

Continuamos o processo, agora incluindo as variáveis sobre hábitos de tela.

Table 11: Modelo Hierárquico - Passo 3: Hábitos de Tela

Passo 3: Gênero + Sig. (Passo 2) + Hábitos de Tela

Characteristic	Beta	95% CI	p-value
(Intercept)	88	81, 95	< 0.001
genero			
Feminino	—	—	
Masculino	1.3	-1.6, 4.2	0.4
partocat	-2.5	-5.4, 0.43	0.094
tempotelacat36	2.4	-1.4, 6.1	0.2
horacertacat36	1.4	-1.9, 4.7	0.4
telacasa36			
Acompanhado pelos cuidadores (adultos)	—	—	
Acompanhado por outra criança	-0.56	-5.2, 4.1	0.8
Sozinho	0.93	-3.3, 5.2	0.7
telalimitetempocat36	2.5	-0.80, 5.9	0.13
telalimiteconteudocat36	1.6	-4.6, 7.8	0.6
R ² Ajustado	0.066		
AIC	884		

Abbreviation: CI = Confidence Interval

Table 12: Modelo Hierárquico Final

Modelo Hierárquico Final
Resultado após a adição sequencial de todos os blocos

Characteristic	Beta	95% CI	p-value
(Intercept)	86	77, 94	<0.001
genero			
Feminino	—	—	
Masculino	1.5	-1.2, 4.3	0.3
partocat	-2.7	-5.5, 0.03	0.053
telalimitetempocat36	3.0	-0.02, 5.9	0.051
classcogcomp36	2.2	-6.5, 11	0.6
classmfbal36	0.97	-6.3, 8.2	0.8
classmgbal36	2.0	-4.5, 8.5	0.5
R ² Ajustado	0.065		
AIC	891		

Abbreviation: CI = Confidence Interval

Nesta etapa, **partocat** manteve-se como uma variável de tendência ($p=0.094$). Nenhuma nova variável atingiu o limiar de significância de $p < 0.10$. No entanto, por interesse teórico e para explorar seu potencial impacto, a variável **telalimitetempocat36** também será mantida e levada para a etapa final, juntamente com **partocat**.

10.0.5.4 Passo 4: Adição do Bloco de Desenvolvimento (Modelo Final)

Finalmente, adicionamos as variáveis de desenvolvimento, mantendo **partocat** e **telalimitetempocat36** no modelo, para chegarmos ao nosso modelo hierárquico final.

No modelo final, após a inclusão de todos os blocos, as variáveis **partocat** ($p=0.053$) e **telalimitetempocat36** ($p=0.051$) mantiveram-se como as únicas significativas no limiar de $p < 0.10$. Com base nisso, construiremos nosso modelo final refinado apenas com estas variáveis.

10.0.5.5 Modelo Hierárquico Final (Refinado)

O modelo abaixo contém apenas as variáveis (**genero**, **partocat** e **telalimitetempocat36**) que se mostraram robustas ao final do processo de seleção hierárquica.

Table 13: Modelo Hierárquico Final (Refinado)

Modelo Hierárquico Final (Refinado)

Characteristic	Beta	95% CI	p-value
(Intercept)	90	87, 94	<0.001
genero			
Feminino	—	—	
Masculino	1.2	-1.5, 4.0	0.4
partocat	-2.7	-5.5, 0.00	0.050
telalimitetempocat36	3.5	0.65, 6.4	0.017
R ² Ajustado	0.082		
AIC	896		

Abbreviation: CI = Confidence Interval

10.0.6 Fase Final: Comparação e Seleção do Modelo

Chegamos à etapa decisiva: comparar os modelos que construímos e selecionar o que melhor responde à nossa pergunta de pesquisa.

Construímos os seguintes modelos: 1. **Modelo Completo:** Incluiu todas as variáveis com p -mediana < 0.25 na análise univariada. 2. **Modelo Stepwise (Exploratório):** Construído automaticamente a partir de todas as variáveis do dataset. 3. **Modelo Stepwise (Refinado):** Construído automaticamente, mas partindo apenas das variáveis pré-selecionadas no Modelo 1. 4. **Modelo Hierárquico Sequencial:** O nosso modelo mais teórico, construído passo a passo por blocos, acumulando variáveis com $p < 0.10$. 5. **Modelo Hierárquico (Refinado):** Construído apenas com as variáveis que se mostraram significativas no final do processo de seleção hierárquica.

A tabela abaixo resume o desempenho de cada um.

- **R² Ajustado:** Indica a proporção da variabilidade da qualidade de vida que o modelo consegue explicar, ajustado pelo número de variáveis. Quanto maior, melhor.
- **AIC/BIC:** São critérios de informação que penalizam a complexidade do modelo. Quanto menores, melhor.

10.0.6.1 Discussão e Escolha do Modelo Final

Ao analisar a tabela de comparação, observamos que o **Modelo Stepwise (Exploratório)** e o **Stepwise (Refinado)** apresentam os melhores indicadores estatísticos, com o maior R²

Table 14: Comparação das Métricas de Ajuste entre os Modelos de Regressão

Comparativo de Desempenho dos Modelos Finais
Métricas para avaliar o ajuste e a complexidade

Modelo	R ² Ajustado	AIC	BIC	N
Completo (Pré-selecionado)	0.088	797.171	852.414	117
Stepwise (Exploratório)	0.146	746.828	765.857	112
Stepwise (Refinado)	0.129	747.091	760.683	112
Hierárquico (Refinado)	0.082	895.732	910.031	129

Ajustado e os menores valores de AIC e BIC. Isso sugere que, do ponto de vista puramente preditivo, esses modelos são superiores.

No entanto, o **Modelo Hierárquico Sequencial**, embora com um poder de explicação um pouco menor (R² ajustado de 0.057 vs 0.129 do Stepwise Refinado), tem a grande vantagem de ter sido construído com base em uma lógica teórica clara, conforme sugerido pela orientadora. Ele mostra como as associações mudam à medida que o modelo é ajustado por novos fatores, o que enriquece a discussão.

Recomendação: Para o objetivo final do trabalho, que é entender os fatores associados à qualidade de vida de uma forma interpretável, tanto o **Modelo Stepwise (Refinado)** quanto o **Modelo Hierárquico** são fortes candidatos. A escolha entre eles depende do peso que se quer dar à otimização estatística versus à fundamentação teórica.

A seguir, faremos a análise de resíduos de ambos para garantir que seus pressupostos são válidos.

11 Análise de Resíduos

Para garantir a validade de nossos modelos de regressão, é essencial verificar se os pressupostos do modelo linear foram atendidos. Faremos isso analisando os resíduos dos nossos dois melhores candidatos: o `modelo_stepwise_refinado` (melhor estatisticamente) e o `modelo_hierarquico_refinado` (melhor teoricamente).

Os pressupostos são:

1. **Linearidade:** A relação entre os preditores e a variável resposta é linear.
2. **Homocedasticidade:** A variância dos erros (resíduos) é constante.
3. **Normalidade dos Resíduos:** Os erros seguem uma distribuição normal.
4. **Independência dos Erros:** Os erros não são correlacionados entre si.

11.1 Diagnóstico do Modelo Stepwise Refinado

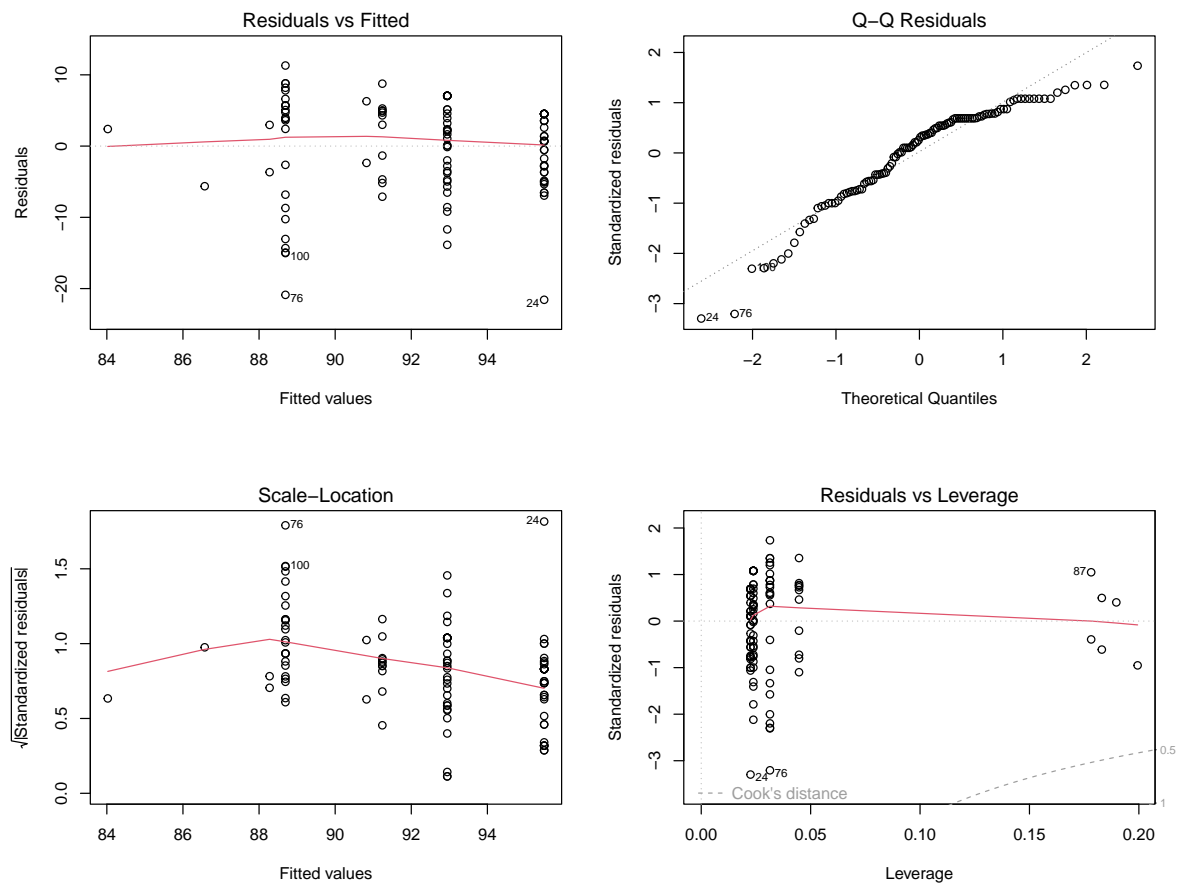


Figure 18: Gráficos de diagnóstico para o Modelo Stepwise Refinado

Interpretação dos Gráficos:

- **Residuals vs Fitted:** A linha vermelha está razoavelmente próxima da linha horizontal em zero, sem um padrão curvo óbvio, o que sugere que o pressuposto de **linearidade** é atendido. A dispersão dos pontos parece relativamente constante, indicando **homocedasticidade**.
- **Normal Q-Q:** Os pontos se desviam consideravelmente da linha diagonal, especialmente nas caudas. Isso indica que os resíduos **não seguem uma distribuição normal**, violando este pressuposto.
- **Scale-Location:** A linha vermelha está quase horizontal, confirmando a **homocedasticidade** (variância constante dos resíduos).
- **Residuals vs Leverage:** Não há pontos no canto superior ou inferior direito, fora da distância de Cook (linhas pontilhadas), o que significa que não há **pontos de alavancagem excessiva ou outliers influentes** que possam distorcer o modelo.

Table 15: Teste de Normalidade de Shapiro-Wilk para o Modelo Stepwise Refinado

W	p-valor	Método
0.9239	0.0000	Shapiro-Wilk normality test

Conclusão: O Modelo Stepwise Refinado atende aos pressupostos de linearidade e homocedasticidade, mas **falha no pressuposto de normalidade dos resíduos**.

11.1.1 Teste de Normalidade dos Resíduos (Shapiro-Wilk)

O teste de Shapiro-Wilk para os resíduos do modelo apresentou um p-valor de <0.001 . Como este valor é menor que 0.05, **rejeitamos a hipótese nula**, o que reforça a conclusão da análise gráfica de que os resíduos **não seguem uma distribuição normal**.

11.2 Diagnóstico do Modelo Hierárquico Final

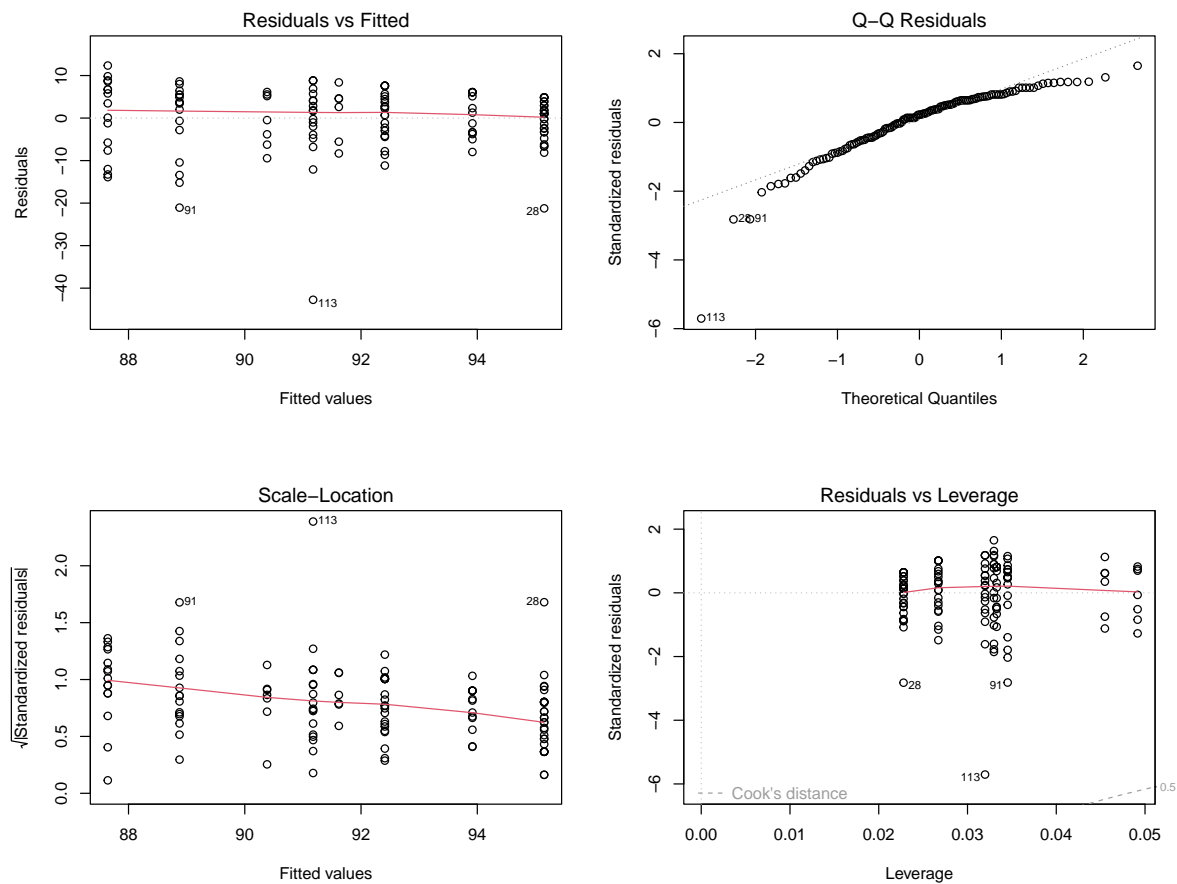


Figure 19: Gráficos de diagnóstico para o Modelo Hierárquico Final

Interpretação dos Gráficos:

- **Residuals vs Fitted:** Similar ao modelo anterior, a linha vermelha está próxima de zero e os pontos estão dispersos sem um padrão claro, apoiando os pressupostos de **linearidade** e **homocedasticidade**.
- **Normal Q-Q:** Os pontos se desviam da linha tracejada, principalmente nas caudas, o que indica que os resíduos **não são normalmente distribuídos**.
- **Scale-Location:** A linha vermelha é quase horizontal, reforçando a conclusão de **homocedasticidade**.
- **Residuals vs Leverage:** Não observamos pontos problemáticos que estejam exercendo uma influência indevida sobre os resultados do modelo.

Conclusão: O Modelo Hierárquico Final também atende aos pressupostos de linearidade e homocedasticidade, mas **viola o de normalidade dos resíduos**.

Table 16: Teste de Normalidade de Shapiro-Wilk para o Modelo Hierárquico (Refinado)

W	p-valor	Método
0.8668	0.0000	Shapiro-Wilk normality test

11.2.1 Teste de Normalidade dos Resíduos (Shapiro-Wilk)

O teste de Shapiro-Wilk para os resíduos deste modelo apresentou um p-valor de <0.001 . Sendo este valor menor que 0.05, há evidências para **rejeitar a hipótese de normalidade dos resíduos**, o que está em concordância com a análise visual do gráfico Q-Q.

12 Modelagem Avançada: Regressão Beta

12.1 Por que um novo modelo?

Nossa análise de resíduos anterior revelou um problema fundamental: os modelos de regressão linear, embora úteis para uma exploração inicial, não são adequados para a nossa variável resposta, a `pedsqglobal36`. A violação do pressuposto de normalidade dos resíduos (confirmada visualmente e pelo teste de Shapiro-Wilk) não é um acaso, mas uma consequência direta da natureza dos dados de qualidade de vida, que são limitados (entre 0 e 100) e geralmente assimétricos.

Para resolver isso, adotaremos uma abordagem mais sofisticada e teoricamente correta: a **Regressão Beta**. Conforme recomendado na literatura especializada (por exemplo, [Hunger et al., 2011](#); [Kharroubi, 2020](#)), a Regressão Beta é um tipo de Modelo Linear Generalizado (GLM) projetado especificamente para variáveis contínuas que representam taxas ou proporções, como o nosso escore.

Em vez de assumir uma distribuição normal, este modelo assume que a variável resposta segue uma **distribuição Beta**, que naturalmente acomoda a assimetria e os limites dos dados.

12.2 Implementação da Regressão Beta

12.2.1 Passo 1: Preparação da Variável Resposta e Seleção de Preditores

A Regressão Beta exige que a variável resposta esteja no intervalo aberto (0, 1). Para isso, aplicamos a transformação $(y * (n - 1) + 0.5) / n$, que é uma abordagem padrão para ajustar escores que podem incluir valores de 0 e 100. Em seguida, reutilizamos as mesmas variáveis candidatas da análise anterior.

Variáveis candidatas para os modelos Beta:

- ` partocat `
- ` lme6 `
- ` estadocivilmae24 `
- ` escolamaecat36 `
- ` abepcat36 `
- ` genero `
- ` creche36 `
- ` tempotelacat36 `
- ` horacertacat36 `
- ` telacasa36 `
- ` telalimitetempocat36 `
- ` telalimiteconteudocat36 `
- ` classcogcomp36 `
- ` classmfbal36 `
- ` classmgbal36 `

12.2.2 Fase 2: Construção e Comparação dos Modelos

Agora, construiremos os três modelos Beta para comparação.

12.2.2.1 Modelo A: Modelo Beta “Completo”

Este modelo inclui todas as variáveis candidatas de uma só vez.

12.2.2.2 Modelo B: Modelo Beta “Stepwise Refinado”

Este modelo usa um algoritmo para refinar o Modelo A, removendo variáveis que não contribuem significativamente.

[1] "100 % of the process"

Table 17: Modelo de Regressão Beta Completo com Todas as Variáveis Candidatas

Modelo Beta Completo
Variável dependente: PedsQL Global (transformado)

Characteristic	exp(Beta)	95% CI	p-value
partocat	0.73	0.54, 1.0	0.046
lme6	1.28	0.95, 1.74	0.11
estadocivilmae24	1.23	0.90, 1.67	0.2
escolamaecat36	1.25	0.93, 1.67	0.13
abepcat36			
A	—	—	
B	0.72	0.40, 1.28	0.3
C	0.79	0.47, 1.35	0.4
DE	0.59	0.30, 1.15	0.12
genero			
Feminino	—	—	
Masculino	1.18	0.88, 1.59	0.3
creche36			
Não	—	—	
Sim	1.20	0.87, 1.66	0.3
tempotelacat36	1.24	0.81, 1.89	0.3
horacertacat36	1.16	0.81, 1.64	0.4
telacasa36			
Acompanhado pelos cuidadores (adultos)	—	—	
Acompanhado por outra criança	1.62	0.99, 2.66	0.054
Sozinho	1.16	0.75, 1.79	0.5
telalimitetempocat36	1.50	1.06, 2.11	0.021
telalimiteconteudocat36	1.58	0.85, 2.93	0.15
classcogcomp36	0.87	0.37, 2.01	0.7
classmfbal36	1.78	0.85, 3.69	0.12
classmgbal36	1.03	0.54, 2.00	>0.9

Abbreviation: CI = Confidence Interval

Table 18: Modelo de Regressão Beta com Seleção Stepwise (via StepBeta)

Modelo Beta Stepwise (Refinado)

Variáveis selecionadas automaticamente por AIC via pacote StepBeta

Characteristic	exp(Beta)	95% CI	p-value
telalimitetempocat36	1.81	1.34, 2.45	<0.001
classmfbal36	1.78	1.03, 3.08	0.040

Abbreviation: CI = Confidence Interval

Table 19: Modelo Beta Hierárquico - Passo 1

Characteristic	exp(Beta)	95% CI	p-value
genero			
Feminino	—	—	
Masculino	1.11	0.83, 1.47	0.5
partocat	0.71	0.53, 0.95	0.022
lme6	1.21	0.90, 1.62	0.2

Abbreviation: CI = Confidence Interval

Table 20: Modelo Beta Hierárquico - Passo 2

Characteristic	exp(Beta)	95% CI	p-value
genero			
Feminino	—	—	
Masculino	1.14	0.86, 1.51	0.4
partocat	0.69	0.52, 0.93	0.016
estadocivilmae24	1.48	1.10, 1.97	0.008
creche36			
Não	—	—	
Sim	1.18	0.88, 1.59	0.3
escolamaecat36	1.25	0.94, 1.65	0.13
abepcat36			
A	—	—	
B	0.64	0.37, 1.11	0.11
C	0.75	0.45, 1.24	0.3
DE	0.59	0.31, 1.11	0.10

Abbreviation: CI = Confidence Interval

Table 21: Modelo Beta Hierárquico - Passo 3

Characteristic	exp(Beta)	95% CI	p-value
genero			
Feminino	—	—	
Masculino	1.11	0.83, 1.50	0.5
partocat	0.75	0.55, 1.01	0.060
estadocivilmae24	1.23	0.91, 1.66	0.2
abepcat36			
A	—	—	
B	0.72	0.40, 1.28	0.3
C	0.82	0.48, 1.40	0.5
DE	0.66	0.34, 1.26	0.2
tempotelacat36	1.25	0.82, 1.90	0.3
horacertacat36	1.10	0.77, 1.57	0.6
telacasa36			
Acompanhado pelos cuidadores (adultos)	—	—	
Acompanhado por outra criança	1.48	0.90, 2.42	0.12
Sozinho	1.01	0.66, 1.56	>0.9
telalimitetempocat36	1.45	1.04, 2.04	0.029
telalimiteconteudocat36	1.39	0.75, 2.56	0.3

Abbreviation: CI = Confidence Interval

12.2.2.3 Modelo C: Modelo Beta “Hierárquico” (Análise Principal)

Esta é a nossa análise principal, seguindo a abordagem teórica em blocos definida anteriormente.

Passo C1: Gênero + Saúde Perinatal

Variáveis significativas ($p \leq 0.10$) para o próximo passo: partocat

Passo C2: Adição do Contexto Social

Variáveis significativas ($p \leq 0.10$) para o próximo passo: partocat, estadocivilmae24, abepcat36

Passo C3: Adição dos Hábitos de Tela

Variáveis significativas ($p \leq 0.10$) para o próximo passo: partocat, estadocivilmae24, abepcat36, telalimitetempocat36

Passo C4: Adição do Bloco de Desenvolvimento

Table 22: Modelo Beta Hierárquico - Passo 4 (Final)

Characteristic	exp(Beta)	95% CI	p-value
genero			
Feminino	—	—	
Masculino	1.02	0.76, 1.35	>0.9
partocat	0.77	0.57, 1.04	0.090
estadocivilmae24	1.23	0.91, 1.66	0.2
abepcat36			
A	—	—	
B	0.67	0.38, 1.18	0.2
C	0.83	0.49, 1.41	0.5
DE	0.70	0.37, 1.31	0.3
telalimitetempocat36	1.65	1.22, 2.23	0.001
classcogcomp36	0.91	0.40, 2.09	0.8
classmfbal36	1.80	0.90, 3.61	0.10
classmgbal36	0.97	0.50, 1.85	>0.9

Abbreviation: CI = Confidence Interval

Variáveis significativas ($p \leq 0.10$) no modelo final: partocat, estadocivilmae24, abepcat36, telalimitetempocat36, classmfbal36

Modelo Hierárquico Final (Refinado) Este é nosso modelo final, contendo apenas as variáveis que se mostraram importantes ao final do processo hierárquico.

12.2.3 Fase 3: Seleção, Diagnóstico e Interpretação

12.2.3.1 Comparação dos Modelos

Comparamos o modelo Stepwise e o Hierárquico para garantir que nossa escolha teórica é estatisticamente sólida.

Ao comparar os modelos, focamos em três métricas principais:

- **Pseudo R²:** Similar ao R² na regressão linear, indica a proporção da “variabilidade” da qualidade de vida que o modelo consegue explicar. Valores maiores são melhores.
- **AIC e BIC:** São critérios de informação que avaliam o ajuste do modelo, mas penalizam a complexidade (número de variáveis). Valores menores são melhores, indicando um bom equilíbrio entre ajuste e simplicidade.

Table 23: Modelo Beta Hierárquico Final (Refinado)

Characteristic	exp(Beta)	95% CI	p-value
genero			
Feminino	—	—	
Masculino	1.01	0.76, 1.35	>0.9
partocat	0.77	0.57, 1.03	0.081
estadocivilmae24	1.23	0.92, 1.65	0.2
abepcat36			
A	—	—	
B	0.67	0.38, 1.18	0.2
C	0.83	0.49, 1.41	0.5
DE	0.70	0.37, 1.31	0.3
telalimitetempocat36	1.63	1.21, 2.19	0.001
classmfbal36	1.75	1.01, 3.02	0.047

Abbreviation: CI = Confidence Interval

Table 24: Comparação de Métricas de Ajuste entre os Modelos Beta

Comparativo de Desempenho dos Modelos Beta

Métricas para avaliar o ajuste e a complexidade

Modelo	Pseudo R ²	AIC	BIC	N
Stepwise (Refinado)	0.112	-367.022	-355.973	117
Hierárquico (Refinado)	0.165	-371.602	-343.727	120

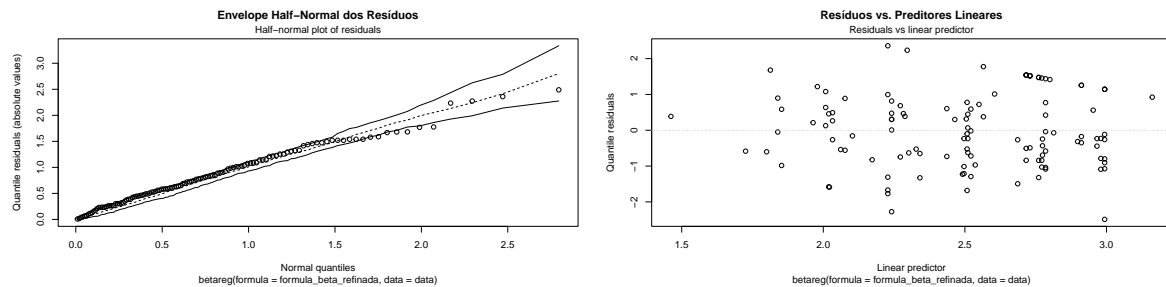
O **Modelo Hierárquico (Refinado)** apresentou um **Pseudo R^2 superior (0.165 vs. 0.112)**, sugerindo que ele explica uma porção maior da variabilidade na qualidade de vida. Embora o modelo Stepwise tenha um AIC/BIC ligeiramente menor, a superioridade do Pseudo R^2 no modelo hierárquico, somada à sua forte fundamentação teórica, nos dá confiança para escolhê-lo como nosso modelo final. Ele não apenas se ajusta bem aos dados, mas o faz de uma maneira que é interpretável e alinhada com as hipóteses da pesquisa.

12.2.3.2 Diagnóstico do Modelo Final (Hierárquico Refinado)

Para verificar se o modelo é adequado avaliamos o comportamento dos **resíduos quantílicos** fornecidos pelo **betareg**. Os principais pontos de atenção são:

1. **Adequação geral (Envelope Half-Normal):** Os pontos devem ficar, em sua maior parte, dentro da banda (envelope) do gráfico half-normal. Pontos fora sugerem má especificação ou outliers.
2. **Homocedasticidade:** No gráfico *Resíduos vs. Preditor Linear* os pontos devem formar uma nuvem aleatória em torno de zero, sem desenho de funil ou curva.
3. **Independência:** Não deve haver padrão sequencial ou clusters sistemáticos nos resíduos nem observações com influência excessiva (investigada via distância de Cook, se necessário).

A seguir, geramos os gráficos para verificar visualmente esses pressupostos.



(a) Gráficos de diagnóstico para o Modelo Beta Hi-erárquico Final

Interpretação do Diagnóstico:

A análise dos gráficos de diagnóstico confirma a boa adequação do nosso modelo final:

1. **Envelope Half-Normal dos Resíduos:** Os pontos dos resíduos deviance seguem aproximadamente a linha diagonal esperada, ficando dentro da banda do envelope. Isso indica que o modelo está **bem ajustado** aos dados. Pontos sistematicamente fora do envelope sugeririam problemas no ajuste.

2. **Resíduos vs. Preditores Lineares:** Os pontos se espalham de forma aleatória em torno da linha horizontal em zero, sem formar um padrão óbvio (como um funil ou uma curva). Isso confirma a **homocedasticidade**, ou seja, a variância dos erros é constante.

Conclusão do Diagnóstico: Como os pressupostos foram atendidos, o modelo hierárquico refinado é considerado estatisticamente válido e robusto, o que nos permite interpretar seus resultados com confiança.

12.2.3.3 Resultados Finais

Por que usamos “Odds Ratios”?

O modelo de regressão beta nos fornece coeficientes (chamados de “betas”) que são um pouco difíceis de interpretar diretamente. Para torná-los mais intuitivos, nós os convertemos em **Odds Ratios (OR)** aplicando uma transformação matemática (exponenciação).

Um Odds Ratio nos diz como a “chance” de uma criança ter uma maior qualidade de vida muda quando comparamos dois grupos.

- **OR = 1:** Não há diferença na chance entre os grupos.
- **OR > 1:** O grupo de interesse tem uma chance *maior* de ter melhor qualidade de vida. Por exemplo, um OR de 1.75 significa que a chance é 75% maior.
- **OR < 1:** O grupo de interesse tem uma chance *menor* de ter melhor qualidade de vida. Por exemplo, um OR de 0.77 significa que a chance é 23% menor.

A tabela a seguir apresenta os coeficientes Beta originais do modelo e sua versão convertida, os Odds Ratios, que usaremos para a interpretação.

Interpretação dos Resultados:

O modelo final destaca os seguintes fatores associados à qualidade de vida infantil aos 36 meses:

1. **Limite de Tempo de Tela (*telalimitetempocat36*, $p < 0.01$):** Esta foi a variável com a associação mais forte. O fato de os cuidadores **limitarem ativamente o tempo de tela** está positivamente associado a uma melhor qualidade de vida. O Odds Ratio de 1.63 significa que as crianças cujos pais limitam o tempo de tela têm **63% mais chances** de apresentar um escore de qualidade de vida mais alto em comparação com aquelas cujos pais não impõem limites, mantendo os outros fatores constantes.
2. **Desenvolvimento Motor Fino (*classmfba136*, $p < 0.05$):** A classificação do desenvolvimento motor fino também se mostrou um fator importante. Crianças com desenvolvimento **normal** nesta área apresentaram maior qualidade de vida. O Odds Ratio de 1.75 indica que crianças com desenvolvimento motor fino adequado têm **75% mais chances** de ter uma maior qualidade de vida em comparação com crianças com atraso motor, controlando pelos outros fatores do modelo.

Table 25: Modelo Final (Beta Hierárquico Refinado): Fatores Associados à Qualidade de Vida

Resultados da Regressão Beta Hierárquica

Variável dependente: PedsQL Global (transformado)

Característica	Beta	Odds Ratio	95% CI	p-value
genero				
Feminino		—	—	
Masculino		1.01	0.76, 1.35	>0.9
partocat	-0.26	0.77	0.57, 1.03	0.081
estadocivilmae24	0.21	1.23	0.92, 1.65	0.2
abepcat36				
A		—	—	
B		0.67	0.38, 1.18	0.2
C		0.83	0.49, 1.41	0.5
DE		0.70	0.37, 1.31	0.3
telalimitetempocat36	0.49	1.63	1.21, 2.19	0.001
classmfbal36	0.56	1.75	1.01, 3.02	0.047

Abbreviation: CI = Confidence Interval

Nota: OR > 1 indica uma associação positiva com maior qualidade de vida; OR < 1 indica uma associação negativa.

3. **Via de Parto (partocat, $p < 0.10$):** A via de parto apareceu como uma **tendência estatística**. O resultado sugere que crianças que nasceram de parto vaginal/fórceps ($\text{partocat}=1$) têm uma chance **23% menor** ($\text{OR} = 0.77$) de apresentar maior qualidade de vida em comparação com as nascidas de cesariana. Embora não seja estatisticamente significativo ao nível de 5%, este achado pode merecer atenção em estudos futuros com amostras maiores.

As demais variáveis, como estado civil da mãe (estadocivilmae24) e classe social (abepcat36), não apresentaram associação estatisticamente significativa no modelo final, sugerindo que seu efeito pode ser menos direto ou explicado por outros fatores incluídos na análise.

Para uma interpretação mais visual, o **Forest Plot** abaixo mostra os Odds Ratios e seus intervalos de confiança de 95%. Se um intervalo de confiança (a linha horizontal) **não cruza a linha vertical em 1**, o efeito da variável é estatisticamente significativo.

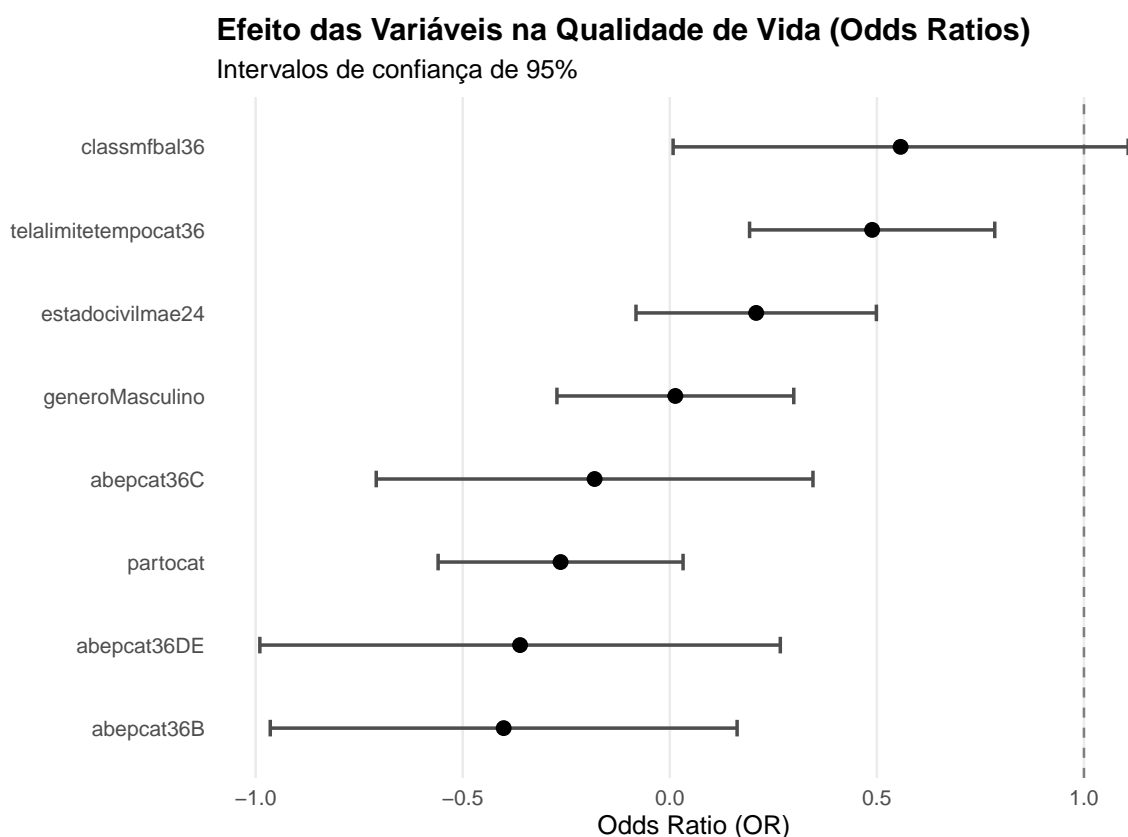


Figure 22: Forest Plot dos Odds Ratios do Modelo Beta Hierárquico Final

No gráfico, podemos ver claramente que os intervalos de confiança para **telalimitetempocat36** e **classmfbal36** estão inteiramente à direita da linha do “1”, confirmando visualmente sua

associação positiva e significativa com a qualidade de vida. As demais variáveis, por outro lado, têm intervalos que cruzam a linha do “1”, indicando que seus efeitos não são estatisticamente distinguíveis de zero no contexto deste modelo.

13 Conclusão Final

Este estudo se propôs a investigar os fatores associados à qualidade de vida de crianças aos 36 meses. Após uma análise descritiva inicial, ficou claro que a natureza da nossa variável resposta (o escore PedsQL Global) exigia uma abordagem estatística mais sofisticada que a regressão linear tradicional. A distribuição assimétrica e limitada do escore nos levou a adotar a **Regressão Beta**, um modelo mais robusto e adequado para este tipo de dado.

A construção de um modelo hierárquico, baseado em blocos teóricos, permitiu uma análise criteriosa e revelou os fatores de maior impacto. Nossos resultados finais indicam de forma consistente que:

- **Hábitos parentais importam:** A imposição de **limites no tempo de uso de telas** emergiu como o preditor mais forte de uma melhor qualidade de vida.
- **O desenvolvimento é chave:** Crianças com **desenvolvimento motor fino adequado** à idade também apresentaram, de forma significativa, uma melhor qualidade de vida.
- **Fatores a observar:** A **via de parto** surgiu como uma tendência estatística, sugerindo uma possível associação que pode ser explorada em futuras pesquisas.

Em suma, os achados desta análise reforçam a importância da intervenção parental ativa nos hábitos digitais das crianças e destacam a conexão intrínseca entre o desenvolvimento motor e o bem-estar geral. Estes resultados fornecem evidências valiosas para profissionais de saúde, pais e educadores sobre os fatores que podem promover uma melhor qualidade de vida na primeira infância.