



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz



ESCOLA NACIONAL DE SAÚDE PÚBLICA
SERGIO AROUCA
ENSP

“Métodos e interpretação de modelos estatísticos de análise de medidas repetidas: uma aplicação a ensaio clínico”

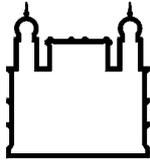
por

Gabriela de Queiroz

Dissertação apresentada com vistas à obtenção do título de Mestre em Ciências, na área de Epidemiologia em Saúde Pública.

Orientadora: Prof.^a Dr.^a Marília Sá Carvalho

Rio de Janeiro, maio de 2012.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz



ESCOLA NACIONAL DE SAÚDE PÚBLICA
SERGIO AROUCA
ENSP

Esta dissertação, intitulada

“Métodos e interpretação de modelos estatísticos de análise de medidas repetidas: uma aplicação a ensaio clínico”

apresentada por

Gabriela de Queiroz

foi avaliada pela Banca Examinadora composta pelos seguintes membros:

Prof.^a Dr.^a Mariane Branco Alves

Prof.^a Dr.^a Katia Vergetti Bloch

Prof.^a Dr.^a Marília Sá Carvalho – Orientadora

Catálogo na fonte
Instituto de Comunicação e Informação Científica e Tecnológica
Biblioteca de Saúde Pública

Q3 Queiroz, Gabriela de
Métodos e interpretação de modelos estatísticos de análise de
medidas repetidas: uma aplicação a ensaio clínico. / Gabriela de
Queiroz. -- 2012.

132 f. : il. ; tab. : graf.

Orientador: Carvalho, Marília Sá

Dissertação (Mestrado) – Escola Nacional de Saúde Pública
Sergio Arouca, Rio de Janeiro, 2012

1. Obesidade-terapia. 2. Sobrepeso. 3. Prevalência. 4. Estudos
Longitudinais. 5. Modelos Estatísticos. 6. Ensaio Clínico como
Assunto. 7. Brasil. I. Título.

CDD - 22.ed. – 616.3980981

*Para as três grandes mulheres da minha vida:
minha mãe, Dadi e Tati*

AGRADECIMENTOS

À minha orientadora Marilia, que me acolheu como orientanda e que contribuiu imensamente para que este trabalho pudesse ser realizado e concluído.

À Mariane Branco Alves, por me transmitir o que existe de mais valioso: o conhecimento.

À Katia Vergetti Bloch, por ceder os dados utilizados neste trabalho e por estar sempre a disposição todas as vezes que precisei de ajuda.

It takes no compromise to give people their rights.

It takes no money to respect the individual.

It takes no political deal to give people freedom.

It takes no survey to remove repression.

Harvey Milk

Progress comes from being willing to walk a mile in someone else's shoes.

We need to ask ourselves, "How would it feel if it were a crime to love the person I love?"

How would it feel to be discriminated against for something about myself that I cannot change?"

This challenge applies to all of us as we reflect upon deeply held beliefs, as we work to embrace tolerance and respect for the dignity of all persons, and as we engage humbly with those with whom we disagree in the hope of creating greater understanding.

Hillary Clinton

RESUMO

Introdução: Estudos longitudinais são aqueles nos quais as variáveis são medidas repetidamente, ou seja, em diferentes momentos para o mesmo indivíduo. Em estudos longitudinais as observações de um indivíduo geralmente não são independentes entre si, e por isso é necessário o uso de técnicas estatísticas que considerem o fato de que medidas repetidas do mesmo indivíduo são correlacionadas. Além disso, em estudos de coorte, esses dados geralmente são desbalanceados (medidas obtidas em momentos diferentes nos pacientes e/ou número diferente de medidas repetidas por paciente) ou tem observações faltantes.

Objetivo: Apresentar técnicas de análise exploratória de dados aplicadas a medidas repetidas e ajustar modelos utilizando a metodologia de modelos de efeitos mistos.

Dados: Os dados referem-se a um ensaio clínico randomizado do uso de orlistat em pacientes hipertensos em coorte acompanhada no Hospital Universitário Clementino Fraga Filho no ano de 2000.

Métodos: Foram aplicados métodos gráficos de análise exploratória específicos para dados longitudinais e modelos de efeitos mistos, tendo como variáveis resposta o peso e a proporção de peso perdido.

Resultados: Os resultados encontrados não foram diferentes do apresentado no estudo original onde observou-se uma redução de peso em ambos os grupos. Porém, no presente estudo, verificou-se uma perda de peso significativa entre os pacientes que utilizaram o medicamento ainda que esta perda não seja expressiva. É importante ressaltar que os dados deste estudo apresentavam problemas comuns - desbalanceamento e perdas de seguimento - o que são complicantes adicionais para a análise de dados longitudinais.

Discussão: A incorporação de modelos de efeitos mistos faz jus à qualidade e custo de estudos longitudinais. Além disso, a dependência entre as medidas repetidas, o uso de dados desbalanceados e as observações faltantes, tornam necessário o uso de um tipo de modelo capaz de lidar com essas características.

ABSTRACT

Background: Longitudinal studies are those in which variables are measured repeatedly, i.e. at different times for the same subject. In longitudinal studies, a subject's observations are not independent, making it necessary to use statistical techniques that take into consideration the fact that repeated observations for a given subject are correlated. Furthermore, cohort studies usually yield unbalanced data (measurements obtained at different times for different subjects and/or a different number of observations per subject) or missing data.

Aim: Present techniques for exploratory data analysis applied to repeated measures and adjust models using the mixed effects modeling methodology.

Dataset: The data comes from a randomized clinical trial of the use of orlistat on hypertense patients in a cohort performed at Hospital Universitario Clementino Fraga Filho in 2000.

Methods: Several graphical methods applied to longitudinal data were used. Mixed effects models were adjusted for the response variables weight and percentage of weight loss.

Results: The results were not different from those presented in the original study where there was a reduction in weight for both groups. However, in the present study, there was a significant weight loss in patients who have used the drug even though this loss is not significant. Importantly, our data showed common problems - imbalance and loss of follow up - what are additional complicating the analysis of longitudinal data.

Discussion: The incorporation of mixed effects models is justified by the quality and cost of longitudinal studies. Moreover, the dependency between repeated measurements, the use of unbalanced data and the existence missing observations make it necessary to use a modeling technique capable of dealing with these characteristics.

Sumário

1	Introdução.....	13
2	O Problema em Análise.....	15
2.1	Obesidade no Brasil.....	16
2.2	Tratamento da Obesidade.....	17
2.3	A Utilização de Medicamentos.....	17
2.3.1	Orlistat	18
3	Objetivos.....	19
3.1	Objetivo Geral.....	19
3.2	Objetivos Específicos.....	19
4	Revisão dos Métodos	20
4.1	Análise Exploratória.....	20
4.1.1	Análise Exploratória de Dados Longitudinais	21
4.2	Modelagem Estatística.....	22
4.2.1	Modelagem de Dados Longitudinais.....	22
4.2.1.1	Modelos de Transição.....	23
4.2.1.2	Modelos Marginais	24
4.2.1.3	Modelos Mistos	25
4.3	Escolha do Modelo.....	29
4.4	Diagnóstico do Modelo.....	29
5	Materiais e Métodos.....	31
5.1	População e Desenho de Estudo.....	31
5.1.1	Participantes.....	31
5.1.2	Desenho de Estudo.....	31
5.1.3	Tamanho da Amostra.....	32
5.1.4	Prescrição da Medicação.....	33
6	Aplicação.....	34
6.1	Análise Exploratória.....	34
6.1.1	Descrição Geral.....	34
6.1.2	Descrição Individual.....	41

6.2	Modelagem.....	59
6.2.1	Definição do Modelo.....	60
6.2.2	Ajuste e Escolha do Modelo.....	60
6.2.2.1	Variável Resposta Peso.....	60
6.2.2.2	Variável Resposta Proporção de Peso Perdido.....	62
6.2.2.3	Variável Resposta Peso e Interação Tratamento-Tempo.....	63
6.2.2.4	Saída do R	64
6.2.3	Resultados dos Modelos.....	67
	Modelos cuja variável resposta é o peso.....	67
	Modelos cuja variável resposta é a proporção de peso perdido.....	69
	Modelos cuja variável resposta é o peso e interação tratamento-tempo.....	72
6.2.4	Examinando os Modelos Ajustados.....	74
6.2.4.1	Exemplos de diferentes representações gráficas.....	74
6.2.4.2	Examinando e comparando os “melhores” modelos.....	78
7	Discussão.....	85
8	Comentários Finais.....	89
9	Referências Bibliográficas.....	90
10	Apêndices	98
10.1	Apêndice A1: Script utilizado para a análise exploratória e modelagem.....	99
10.2	Apêndice A2: Gráficos dos modelos cuja variável resposta é o peso	116
10.3	Apêndice A3: Gráficos dos modelos cuja variável resposta é a proporção de peso perdido.....	125
10.4	Apêndice A4: Gráficos dos modelos cuja variável resposta é o peso e interação tratamento-tempo	130

Lista de Tabelas

Tabela 1: Número de pessoas que participaram da análise por sexo e grupo.....	34
Tabela 2: Características dos participantes do estudo quanto a certos fatores de risco cardiovascular..	35
Tabela 3: Número de pacientes que foram a cada uma das consultas.....	38
Tabela 4: Número de medidas do peso.....	39
Tabela 5: Estimativas das partes fixa e aleatória dos modelos com efeito aleatório no intercepto.....	67
Tabela 6: Estimativas das partes fixa e aleatória dos modelos com interações e efeito aleatório no intercepto.....	68
Tabela 7: Estimativas das partes fixa e aleatória dos modelos com efeitos aleatórios no intercepto e na inclinação.....	69
Tabela 8: Estimativas das partes fixa e aleatória dos modelos com efeito aleatório no intercepto.....	70
Tabela 9: Estimativas das partes fixa e aleatória dos modelos com interações e efeito aleatório no intercepto.....	70
Tabela 10: Estimativas das partes fixa e aleatória dos modelos com efeitos aleatórios no intercepto e na inclinação.....	71
Tabela 11: Estimativas das partes fixa e aleatória dos modelos com interações e efeito aleatório no intercepto.....	72
Tabela 12: Estimativas das partes fixa e aleatória dos modelos com interações e efeitos aleatórios no intercepto e na inclinação.....	73

Lista de Figuras

Figura 1: Evolução da variável resposta Y com diferentes interceptos.	26
Figura 2: Evolução da variável resposta Y com diferentes inclinações.....	27
Figura 3: Evolução da variável resposta Y com diferentes inclinações e interceptos.....	28
Figura 4: Fluxo do Ensaio Clínico.....	32
Figura 5: Box-plot do peso (kg) e idade (anos) no início do estudo.....	36
Figura 6: Box-plot das medidas laboratoriais.....	37
Figura 7: Histogramas do número de dias entre o início do estudo e a primeira consulta,entre a primeira e a segunda consulta e entre a segunda e terceira consulta.....	39
Figura 8: Box-plot do peso no início do estudo e nas consultas seguintes nos dois grupos.....	40
Figura 9: Box-plot com os intervalos de confiança para a mediana do peso no início do estudo e nas consultas seguintes.....	41
Figura 10: Trajetórias individuais de cada paciente do grupo tratamento (em vermelho) e do grupo controle (em azul).....	42
Figura 11: Trajetórias individuais de cada paciente dos grupos tratamento e controle.....	43
Figura 12: Trajetórias individuais de cada paciente por grupo sobrepostas pelo box-plot.....	44
Figura 13: Perfil médio do peso de cada grupo ao longo das consultas.....	45
Figura 14: Perfis individuais do peso por grupo.....	46
Figura 15: Evolução do peso de cada paciente do grupo tratamento cujo peso inicial era menor que 70kg (a) e entre 70 e 80kg (b).....	47
Figura 16: Evolução do peso de cada paciente do grupo tratamento cujo peso inicial era entre 80 e 90kg (c) e entre 90 e 100kg (d).....	48
Figura 17: Evolução do peso de cada paciente do grupo tratamento cujo peso inicial era entre 100 e 110kg (e) e maior que 110kg (f).	48
Figura 18: Evolução do peso de cada paciente do grupo controle cujo peso inicial era menor que 70kg (a), entre 70 e 80kg (b), 80 e 90kg (c) e entre 90 e 100kg (d).....	49
Figura 19: Evolução do peso de cada paciente do grupo controle cujo peso inicial era entre 100 e 110kg (e) e maior que 110kg (f).	50
Figura 20: Trajetórias individuais do peso.	51
Figura 21: Trajetórias individuais do peso.....	52

Figura 22: Gráfico interativo do peso – início do estudo.....	53
Figura 23: Gráfico interativo do peso – primeira consulta.....	54
Figura 24: Gráfico interativo do peso – segunda consulta.....	54
Figura 25: Gráfico interativo do peso – terceira consulta.....	55
Figura 26: Proporção de peso perdido por grupo.....	56
Figura 27: Box-plot da proporção de peso perdido por grupo.....	57
Figura 28: Trajetórias individuais da proporção de peso perdido por grupo.....	58
Figura 29: Gráficos dos valores ajustados versus resíduos.....	74
Figura 30: Gráficos de probabilidade normal para os resíduos.....	75
Figura 31: Gráficos de probabilidade normal do efeito aleatório.....	77
Figura 32: Valores ajustados versus resíduos dos modelos 10 (a),1 (b) e 4 (c).	79
Figura 33: Gráficos de probabilidade normal (Q-Q plot) para os resíduos dos modelos 10 (a),1 (b) e 4 (c).....	80
Figura 34: Gráficos de probabilidade normal (Q-Q plot) para os efeitos aleatórios (intercepto e sexo, respectivamente) dos modelos 10 (a),1 (b) e 4 (c).....	81
Figura 35: Gráficos de probabilidade normal dos efeitos aleatórios dos modelos 10 (a),1 (b) e 4 (c)....	83
Figura 36: Diagnósticos para o modelo 1.....	116
Figura 37: Diagnósticos para o modelo 2.....	117
Figura 38: Diagnósticos para o modelo 3.....	118
Figura 39: Diagnósticos para o modelo 4.....	119
Figura 40: Diagnósticos para o modelo 5.....	120
Figura 41: Diagnósticos para o modelo 6.....	121
Figura 42: Diagnósticos para o modelo 7.....	122
Figura 43: Diagnósticos para o modelo 8.....	123
Figura 44: Diagnósticos para o modelo 9.....	124
Figura 45: Diagnósticos para o modelo 2.....	125
Figura 46: Diagnósticos para o modelo 3.....	126
Figura 47: Diagnósticos para o modelo 4.....	127
Figura 48: Diagnósticos para o modelo 5.....	128
Figura 49: Diagnósticos para o modelo 6.....	129
Figura 50: Diagnósticos para o modelo 1.....	130

Figura 51: Diagnósticos para o modelo 2.....	131
Figura 52: Diagnósticos para o modelo 3.....	132

1 Introdução

Estudo longitudinal é aquele que acompanha o eventos de saúde de cada indivíduo incluído no estudo ao longo do tempo. Nele, segue-se a mesma pessoa ao longo de um período que pode ser meses, anos, ou até décadas. Os dados podem ser coletados prospectivamente, ou seja, seguindo os indivíduos ao longo do tempo, ou retrospectivamente, através da extração de diversas medidas de cada pessoa pelo seu histórico.

Estudos longitudinais apresentam vantagens em relação a outros desenhos, em especial quando o que se deseja é estimar fatores associados à incidência de eventos em saúde. Este tipo de estudo permite ao investigador separar efeitos da idade (por exemplo, mudanças ao longo do tempo no indivíduo) dos efeitos da coorte (por exemplo, diferenças entre os sujeitos na medida inicial) além de se poder avaliar mais precisamente as mudanças que ocorrem no grupo e/ou no indivíduo ao longo do período de observação. Ele permite, portanto, a análise dinâmica ao longo do tempo (Sanson & Smart, 2011).

Porém, a análise dos dados provenientes desse tipo de desenho apresenta dificuldades. Entre essas, o fato de existir uma correlação entre as medidas repetidas ocasionada por medirmos a mesma variável no mesmo indivíduo em tempos diferentes. Como consequência, essas observações não são independentes. Do ponto de vista estatístico, observações dependentes devem ser analisadas levando-se em consideração a correlação existente para obter inferências válidas sobre os coeficientes dos modelos de regressão. Para isto é necessário métodos estatísticos específicos (Frees, 2004).

Diversos estudos longitudinais, ainda hoje, não utilizam técnicas estatísticas adequadas para este tipo de dado. Em alguns estudos considera-se apenas as medidas iniciais e finais de cada variável, excluindo os dados entre estes dois momentos. Em outros, supõe-se que as medidas repetidas em um indivíduo são independentes, o que pode introduzir vieses nas estimativas dos parâmetros dos modelos, por exemplo.

Este trabalho se propõe a apresentar e discutir técnicas de análise exploratória e de modelagem adequadas para tratar dados longitudinais. Para isto, reanalisaremos os dados apresentados em Bloch *et al.* (2003), resultantes de um ensaio clínico do uso do orlistat, um medicamento para emagrecer, realizado na unidade de medicina interna do Hospital Universitário Clementino Fraga Filho (HUCFF). Os participantes do estudo eram homens e mulheres não grávidas

com idade entre 18 e 75 anos com IMC maior do que 25kg/m^2 e hipertensão arterial ($\text{PA} \geq 140 \times 90$ mmHg) diagnosticada há pelo menos três meses.

No capítulo a seguir será apresentada uma breve revisão bibliográfica sobre o problema analisado.

2 O Problema em Análise

Obesidade é uma doença crônica com crescente prevalência no mundo (Seidell, 1998) e que contribui para o desenvolvimento de diabetes e doenças cardiovasculares. Muitas dessas doenças requerem um tratamento de longo prazo e são associadas ao crescente risco de morbidade e mortalidade (Pi-Sunyer, 1993; Bray, 1996). Podemos citar alguns riscos associados ao sobrepeso e à obesidade, entre eles: hipertensão arterial, dislipidemias, diabetes mellitus tipo 2, doenças cardiovasculares e cerebrovasculares, doenças biliares, osteoartrite, apneia do sono e certos tipos de câncer, tais como do endométrio, mama, próstata e cólon (Hernán-Daza, 2002; Sosa, 2004).

As medidas antropométricas representadas pelo Índice de Massa Corporal (IMC), Razão Cintura-Quadril (RCQ) e Circunferência Abdominal (CA) representam uma maneira eficiente de se conhecer o volume e a distribuição de gordura no corpo, e são bastante utilizadas na prática clínica. É considerado com sobrepeso o indivíduo que apresenta IMC maior ou igual a 25kg/m^2 e menor que 30kg/m^2 e obeso o que apresenta IMC maior ou igual a 30kg/m^2 . Além do volume corpóreo total, a circunferência da cintura é um bom marcador da deposição central de tecido adiposo, associando-se fortemente ao risco de doenças crônicas não-transmissíveis como doenças cardiovasculares e diabetes, e sendo, portanto, um importante indicador adicional do estado nutricional de adultos.

A obesidade dobra o risco de diabetes e eleva em 40% o risco das doenças biliares (Francischi *et al.*, 2000; Hernán-Daza, 2002; Ferrari, 2008). Em relação ao risco de câncer, acredita-se que mulheres com IMC maior que 35kg/m^2 tenham 3,2 vezes mais possibilidade de falecer por câncer cervical, e as com IMC maior que 40kg/m^2 possuam 6,2 vezes maior risco de morrer por câncer de útero, além de risco 4,7 vezes maior de mortalidade por câncer renal e 2,7 vezes maior de óbito por câncer de pâncreas. Em homens com IMC maior que 35kg/m^2 os riscos de mortalidade por cânceres hepático, pancreático e gastroesofágico aumentam 4,5, 2,6 e 1,9 vezes, respectivamente (Chen-Ku & Sandí-Bogantes, 2004).

Dieta, inatividade física, fatores genéticos, o meio em que vive, e condições de saúde contribuem para o sobrepeso e a obesidade. Os potenciais benefícios na saúde ao se reduzir a prevalência de sobrepeso - em particular a obesidade - são de significativa importância na saúde pública.

2.1 Obesidade no Brasil

No Brasil, em uma pesquisa realizada pelo IBGE (2002-2003), a prevalência de sobrepeso observada foi de 41% para os homens e 39,2% para as mulheres. Mais recentemente, dados nacionais apresentaram uma prevalência de excesso de peso de 43,1% e de obesidade de 16,1% entre o sexo feminino (PNDS, 2006).

Na população feminina entre 15 e 49 anos, a prevalência de sobrepeso e de obesidade foi de 43% e 26% respectivamente. Neste grupo, o risco de exposição à obesidade se mostrou elevado em todos os estratos, sendo ligeiramente maior nas regiões do Centro-sul do que nas regiões Norte e Nordeste, mostrando tendência de diminuição com o aumento da escolaridade da mulher (PNDS, 2006). Prevalências elevadas de sobrepeso foram encontradas em todas as macrorregiões brasileiras, variando de 41,0% na região Nordeste a 45,2% na região Centro-Oeste e a prevalência de obesidade variando de 12,7% na região Norte a 19,7% na região Sul. A frequência do excesso de peso assim como da obesidade aumentou com a idade da mulher e com o número de filhos tidos, e diminuiu com o aumento da escolaridade da mulher, mas não com o aumento do poder aquisitivo familiar. A situação de residência em área urbana ou rural não modificou substancialmente a prevalência do excesso de peso nem a prevalência de obesidade (PNDS, 2006).

O valor da circunferência da cintura é uma condição indicativa de risco elevado para o desenvolvimento de doenças cardiovasculares em mulheres quando maior ou igual a 80 cm. Neste estudo foram encontrados valores de circunferência maior ou igual a 80 cm em 55,5% das mulheres (PNDS, 2006). A idade e o número de filhos tidos aumentaram a frequência dessa condição: cerca de 3/4 das mulheres entre 45 e 49 anos de idade ou com quatro ou mais filhos apresentaram circunferência da cintura maior ou igual a 80 cm (PNDS, 2006). O aumento do nível de escolaridade também esteve associado à diminuição da frequência de cinturas maior ou igual a 80 cm. Valores da circunferência da cintura maior ou igual a 88 cm, condição indicativa de risco muito elevado à saúde, foram encontrados em 32,1% das mulheres. A idade e o número de filhos também aumentaram a frequência de circunferência de cinturas maior ou igual a 88 cm, condição presente em cerca de metade das mulheres entre 45 e 49 anos de idade ou com quatro ou mais filhos. A frequência de cinturas maior ou igual a 88 cm também diminuiu com o aumento da escolaridade. Não foi encontrado uma associação entre aumento do poder aquisitivo familiar com as frequências de circunferência de cinturas maior ou igual a 80 cm ou maior ou igual a 88 cm (PNDS, 2006).

2.2 Tratamento da Obesidade

O tratamento da obesidade tem como principal ponto a mudança no estilo de vida. Por isso, deve estar claro ao paciente que a perda de peso não é apenas uma medida estética e sim representa a diminuição da morbidade e mortalidade associadas à obesidade. É importante ressaltar que ao reduzir 5% a 10% de seu peso, o indivíduo estará diminuindo consideravelmente a pressão arterial, glicemia, valores séricos de lipídios, o número de apneias e hiponeias durante o sono e o diabetes (Finer, 1998; Nonino-Borges *et al.*, 2006). No entanto, dieta, mudanças no comportamento e atividades físicas geralmente não são suficientes para causar perda e manutenção do peso. Portanto, tratamentos farmacológicos devem ser avaliados, pois podem aumentar a motivação do paciente e a aderência ao tratamento. Por essa razão, são considerados uma abordagem de extrema importância.

2.3 A Utilização de Medicamentos

O uso de remédios para emagrecer deve estar associado a mudanças no estilo de vida do indivíduo (melhoria da dieta, prática regular de atividade física, etc), uma vez que ao interromper o uso deste tipo de medicamento, o indivíduo apresenta ganho de peso. Por isso mesmo, medicamentos para emagrecer não devem ser utilizados apenas por estética (Fortes *et al.*, 2006). Desse modo, a utilização de medicamentos para emagrecer está indicada para pacientes de ambos os sexos, com IMC igual ou maior a 30kg/m² e que não obtiveram resultados na perda de peso com mudanças no estilo de vida. Os remédios antiobesidade podem ser indicados também para pacientes acima do peso (IMC entre 27 e 30 kg/m²), e que possuam pelo menos uma comorbidade, ou seja, pacientes que apresentem doenças crônicas, tais como hipertensão arterial, diabetes mellitus tipo 2, dislipidemia, artrite, câncer de próstata, útero, entre outras (Bouchard, 2000; Sosa, 2004).

Entre as drogas para a perda de peso, sibutramina e seus metabólitos afetam o sistema nervoso central, inibindo a reabsorção de serotonina e noradrenalina, aumentando a sensação de saciedade e termogênese (Connoley *et al.*, 1995; Halford *et al.*, 1995; Connoley *et al.*, 1999). Orlistat, por outro lado, é um inibidor das lipases gastrointestinais, o que diminui a absorção de gorduras em aproximadamente 30%, reduzindo assim a absorção de calorias.

2.3.1 Orlistat

No grupo de drogas que afetam a absorção de nutrientes encontram-se os inibidores de lipases, que atuam no nível gastrointestinal, diminuindo a absorção de gordura. O orlistat (nome comercial Xenical®) inibe as lipases que agem no tubo gastrointestinal, diminuindo a absorção de gordura no nível intestinal. Porém, esta inibição não é total, sendo de aproximadamente 1/3 da gordura ingerida, o que leva a uma perda de peso gradual. Além disso, ele bloqueia a disponibilidade de vitaminas solúveis em gordura (vitaminas A, D, E e K). Portanto, pacientes que fazem o uso deste medicamento devem tomar suplemento vitamínico (NICE, 2001). Esta droga é a única forma atual de medicamento prescrita para obesidade que atua no sistema gastrointestinal e que reduz a absorção de gordura no intestino. A gordura não absorvida é eliminada através do movimento intestinal (Hollywood *et al.*, 2011).

Orlistat, no entanto, tem alguns efeitos colaterais desagradáveis, resultado do impacto da não absorção de gordura, que incluem diarreia, urgência para ir ao banheiro, cólicas abdominais e incontinência fecal. Esses efeitos colaterais são mais intensos após uma dieta rica em gordura, pois a droga faz com que a gordura consumida seja removida do corpo.

Padwal *et al.* (2003) reportaram que pacientes tomando orlistat por um ano perderam 2,7kg a mais do que os que não tomaram. O estudo também indicou que o orlistat reduz os níveis de colesterol e pressão arterial e melhora o controle glicêmico quando comparado ao placebo. Avenell *et al.* (2004) fizeram uma revisão sistemática de ensaios envolvendo combinação de dietas, terapias farmacológicas, exercícios e terapia de comportamento e concluíram que adicionar orlistat a uma dieta melhora a perda de peso em 3,26kg em até 24 meses.

Da mesma forma, Phelan & Wadden (2002) concluíram em sua revisão que a adição de orlistat à modificação do estilo de vida melhora tanto a perda de peso quanto a sua manutenção. Além disso, em uma recente metanálise, Rucker *et al.* (2007) concluíram que os pacientes em tratamentos com droga ativa tiveram mais chances de chegar a 5% e 10% de perda de peso, e que as perdas de peso para três medicamentos-chave foram os seguintes: sibutramina - 4,2kg; rimonabant - 4,7kg; e orlistat - 2,9kg. Esta pesquisa, portanto, indica que o orlistat pode melhorar a perda de peso se usado juntamente com as intervenções comportamentais e de estilo de vida.

3 Objetivos

3.1 Objetivo Geral

Utilizar métodos de análise - exploratória e modelagem - de medidas repetidas visando discutir adequabilidade e facilidade de uso. Como exemplo motivador utilizaremos o estudo do uso de orlistat em pacientes hipertensos em coorte acompanhada no Hospital Universitário Clementino Fraga Filho.

3.2 Objetivos Específicos

- a. Aplicar e discutir técnicas de análise exploratória de dados apropriadas para medidas repetidas.
- b. Ajustar e interpretar modelos para medidas repetidas considerando diferentes medidas relacionadas ao peso corporal.

4 Revisão dos Métodos

4.1 Análise Exploratória

John W. Tukey (1977) foi o primeiro estatístico a fornecer uma descrição detalhada do que é análise exploratória de dados (AED). Ele a definiu como “um trabalho de detetive - um trabalho numérico de detetive – ou trabalho de contagem de detetive - ou trabalho gráfico de detetive”. Análise exploratória de dados é, sobretudo, uma filosofia de análise de dados, onde o pesquisador examina os dados a fim de descobrir o que eles podem lhe dizer sobre os fenômenos estudados. Tukey contrasta isso com a análise de dados de confirmação (ADC), uma área de análise de dados que está principalmente preocupada com testes estatísticos de hipóteses, intervalos de confiança, estimação, etc. AED e ADC não devem ser usados separadamente um do outro, mas sim de forma complementar. O analista explora os dados procurando padrões e estruturas que levem a hipóteses e modelos (Martinez *et al.*, 2010).

A análise exploratória é composta por procedimentos numéricos e gráficos utilizados para resumir um conjunto de dados de forma clara e compreensível e que ajuda a simplificar grandes quantidades de dados de uma maneira sensata (Behrens, 1997). Os procedimentos numéricos são por exemplo medidas estatísticas como a média e o desvio padrão, considerados mais simples do que a abordagem gráfica. Representações gráficas para resumir dados, no entanto, têm inúmeras vantagens em relação às medidas de resumo numéricas, já que gráficos podem descrever simultaneamente uma grande quantidade de valores e seus padrões (Behrens, 1997).

Podemos resumir os objetivos da Análise Exploratória de Dados como:

- Descrição dos dados: brutos e indicadores
- Relação entre a variável resposta e as covariáveis (e entre essas)
- Visualização de padrões longitudinais
- Visualização de padrões transversais
- Identificação de indivíduos e observações incomuns

Em função de suas características, é importante destacar que a análise exploratória deve ser feita antes de qualquer construção de modelos ou teste estatísticos.

4.1.1 Análise Exploratória de Dados Longitudinais

Alguns tipos de dados, quando apresentados na forma de tabela e de algumas formas gráficas, são difíceis de se visualizar. Métodos gráficos tradicionais tais como ramo-e-folha, histogramas, ou box-plots não são suficientes para representar dados longitudinais devido à própria característica deste tipo de dado (por exemplo, padrões não lineares, heterocedasticidade, dependência e variabilidade entre os indivíduos). Isso se torna ainda mais complexo à medida que o número de variáveis aumenta.

No entanto, a visualização de dados é importante e útil na construção de modelos: ela nos ajuda a identificar parâmetros significativos em potencial, especialmente quando não há base teórica ou conhecimento para orientar a formulação do modelo; melhora a compreensão das estatísticas resumo e testes significativos; estimula novas hipóteses a serem testadas, e detecta possíveis padrões (Ker, 2010).

Como alternativa aos gráficos tradicionais acima mencionados, podemos usar técnicas gráficas simples, como o gráfico de barras, gráfico de dispersão e gráfico de perfis, ou gráficos mais complexos, como o radial plot e gráficos dinâmicos. Estes formatos são particularmente adequados para ilustrar relações e tendências de forma concisa.

Técnicas gráficas aplicadas a dados longitudinais devem levar em conta a importância da trajetória dos dados ao longo do tempo, e portanto possibilitar:

1. Descrever o dado ao longo do tempo em cada indivíduo
2. Identificar o momento onde importantes mudanças ocorrem
3. Descrever a forma do crescimento tanto no âmbito do grupo como individual (ex.: linear e não linear)
4. Identificar padrões de comportamento ao longo do tempo
5. Identificar indivíduos incomuns (fora do padrão) na sua trajetória
6. Sugerir modelos estatísticos adequados

4.2 Modelagem Estatística

Segundo Kaplan (2009), “modelagem” é um processo de fazer perguntas e “estatística” refere-se, em parte, aos dados. Segundo ele, a palavra “modelagem” enfatiza que os seus objetivos, as suas crenças e seu conhecimento influenciam a análise dos dados. O papel principal do método científico é a formação de hipóteses que podem ser testadas e, talvez, refutadas por experimento ou observação. Da mesma forma, em modelagem estatística examinam-se os dados para ver se eles são consistentes com a hipótese que sustenta a compreensão do sistema em estudo.

Um modelo estatístico pode ser entendido como a relação entre duas variáveis: uma variável dependente Y , também conhecida como variável desfecho ou resposta, e uma ou mais variáveis independentes X_1, \dots, X_k , também conhecidas como variáveis preditoras, variáveis explicativas, variáveis regressoras ou covariáveis.

Modelos estatísticos podem ser descritivos ou preditivos. Nos modelos descritivos - caso deste trabalho - tem-se por objetivo estimar medidas de associação entre uma variável desfecho e os fatores que se relacionam com ela. Já os modelos preditivos, são usados para fazer previsões sobre fenômenos do mundo real. Eles são construídos a partir de dados coletados representativos dos processos a serem analisados. Quanto mais preciso o modelo, mais confiança temos nas suas previsões. Ou seja, quanto melhor um modelo descrever o processo a partir dos dados coletados, mais apuradas serão as previsões por ele geradas (Field, 2009).

4.2.1 Modelagem de Dados Longitudinais

O objetivo principal do estudo longitudinal é caracterizar as alterações nos valores repetidos da variável resposta e determinar as variáveis explicativas mais associadas com qualquer alteração. Estudos longitudinais apresentam algumas complexidades estatísticas já que a suposição habitual de que todas as observações são independentes geralmente não se sustenta (Finucane *et al.*, 2007).

Como as múltiplas observações de uma variável sobre uma pessoa em particular são provavelmente correlacionadas positivamente (isto é, os erros podem refletir uma tendência sistemática dentro de cada indivíduo), algumas opções para lidar com isso são utilizadas. Muito

frequentemente analisa-se apenas os dados de base da coorte. No entanto, este método utiliza apenas parte dos dados disponíveis, e é altamente ineficiente e desaconselhável. Tal abordagem ineficiente vai levar a estimativas menos precisas, e não permitirá uma avaliação da exposição da variação do tempo e dos seus resultados (Finucane *et al.*, 2007).

Em geral, a escolha do modelo de análise deve depender não apenas da natureza da variável e do tipo de desenho, mas também do objetivo do estudo. Segundo Diggle *et al.* (1994) existem três distintas classes de modelos para análise de dados longitudinais: modelo marginal, modelo de transição e modelo misto ou de efeitos aleatórios. Abaixo descreveremos brevemente os modelos marginais e os de transição e depois, mais detalhadamente, os modelos mistos – foco deste trabalho.

4.2.1.1 Modelos de Transição

Um modelo transição pode ser especificado incluindo os valores de respostas anteriores como covariáveis “adicionais” e, por isso, é semelhante a um modelo clássico de regressão para dados transversais (Wu, 2010). Por exemplo, se a suposição da cadeia de Markov de primeira ordem é válida para um processo longitudinal, então um modelo de transição pode ser escrito como:

$$E(y_{ij}) = g(y_{i,j-1}, \mathbf{x}_i, \beta), \quad i = 1, 2, \dots, \quad (1)$$

onde g é uma função de ligação conhecida, i é o indivíduo e j a observação. Efeitos aleatórios podem também ser introduzidos em um modelo de transição para incorporar variações entre indivíduos, e tal modelo de efeitos mistos de transição pode ser útil para modelar dados longitudinais agrupados (Cook *et al.*, 2004). Maiores detalhes deste tipo de modelo podem ser encontrados em Diggle *et al.* (2002).

4.2.1.2 Modelos Marginais

Os modelos marginais são apropriados para a modelagem de dados longitudinais quando inferências sobre a média populacional é o principal objetivo do estudo. Por exemplo, quando na compreensão da diferença entre a média do grupo tratamento e do grupo controle não é preciso levar em consideração a diferença entre indivíduos (Diggle *et al.*, 1994; Androozzi, 2002). Tomando como exemplo o banco de dados utilizado neste estudo, os modelos marginais podem responder questões do tipo:

- A perda média de peso é maior no grupo tratamento que no grupo controle?
- Como o efeito do medicamento na perda de peso se comporta ao longo do tempo na população?

Além de modelar os efeitos das covariáveis na esperança marginal (resposta média das observações partilhando das mesmas covariáveis), os modelos marginais também permitem a especificação de um modelo de associação entre observações de cada sujeito. Nesta abordagem o analista está interessado em modelar a esperança marginal como função das variáveis preditoras (Diggle *et al.*, 1994; Androozzi, 2002; Horton & Lipsitz, 1999).

Uma limitação desta modelagem, no entanto, é que não são incorporados fatores individuais que afetam a resposta, isto é, não se observa as trajetórias individuais de cada paciente levando em conta as suas diferenças, por exemplo, as diferenças biológicas. Quando se introduz um efeito latente para cada indivíduo pode-se medir quanto o intercepto de um indivíduo difere do restante.

Outra limitação deste tipo de modelo é que as formas de estruturar a matriz de variância e covariância somente podem ser utilizadas em dados longitudinais balanceados, ou seja, o número de medidas e o tempo entre estas medidas deve ser o mesmo para todos os indivíduos.

4.2.1.3 Modelos Mistos

Um modelo linear que apresenta somente fatores de efeitos fixos - além do erro experimental, que é sempre aleatório - é denominado modelo fixo. Os modelos que apresentam tanto fatores de efeitos fixos como aleatórios, além do erro experimental e da constante μ , são chamados de modelos mistos. Os modelos mistos são também conhecidos como modelos de efeitos aleatórios ou modelos hierárquicos e podem ser considerados como uma abordagem flexível onde a variabilidade entre os indivíduos reflete uma heterogeneidade natural devido a fatores não mensurados (Diggle *et al.*, 1994).

É razoável acreditar que a resposta de cada indivíduo em um estudo longitudinal tem vários componentes: um efeito fixo, que é uma função das covariáveis; um efeito aleatório, que expressa a variação entre os indivíduos, e um erro, que é devido à medição ou ao não registro de variáveis (Faraway, 2006).

Efeitos aleatórios são efeitos específicos estimados, para cada indivíduo, como provenientes de uma distribuição comum (geralmente uma distribuição normal). Assim, ao contrário dos modelos lineares – que não incluem um efeito associado ao nível individual, nos modelos de efeito misto, inclinações e/ou interceptos podem ter diferentes valores para cada indivíduo (Schielzeth & Forstmeier, 2008).

Os modelos de efeitos mistos para dados longitudinais permitem que os coeficientes da regressão (coeficiente angular e o intercepto) variem entre os indivíduos. Esses modelos têm dois componentes: um intra individual (uma mudança longitudinal intra individual é descrita pelo modelo de regressão com um intercepto e inclinação populacional) e outro entre indivíduos (variação no intercepto e inclinação individual). Neste modelo, as medidas dos indivíduos não precisam ser igualmente espaçadas e balanceadas e as análises podem ser conduzidas com os dados de indivíduos que foram perdidos de seguimento (pessoas que abandonaram o estudo, por exemplo) ou que apresentam ausência de informação em algum momento do estudo (Fausto *et al.*, 2008). Portanto, na presença de dados faltantes, os modelos de efeitos mistos oferecem mais uma vantagem de utilização frente aos outros tipos de modelos para dados longitudinais (Wu, 2010).

Para um melhor entendimento deste tipo de modelo, apresentaremos a seguir a interpretação do modelo com apenas intercepto aleatório, do modelo com somente inclinação aleatória e do

modelo com intercepto e inclinação aleatórios, respectivamente. Como neste modelo é possível trabalhar com dados desbalanceados o índice do tempo vai variar de 1 até m_i .

No primeiro caso, a forma mais simples de análise de coeficientes aleatórios, considera-se apenas um efeito aleatório e uma única variável independente, a covariável tempo, como:

$$Y_{ij} = \beta_{0i} + \beta_1 t_{ij} + \varepsilon_{ij} \quad i=1, \dots, n \quad e \quad j=1, \dots, m_i, \quad (2)$$

em que Y_{ij} é a resposta observada para o i -ésimo indivíduo na observação j , β_{0i} é o intercepto aleatório, t_{ij} é o tempo, β_1 é o coeficiente de regressão do tempo, e ε_{ij} é o erro para o i -ésimo indivíduo na observação j .

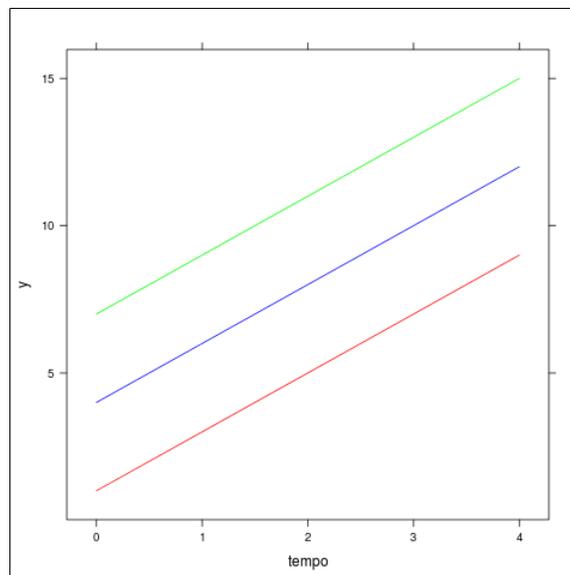


Figura 1: Evolução da variável resposta Y com diferentes interceptos.

Como podemos observar no gráfico acima, cada uma das retas representa um indivíduo. A única diferença entre elas é valor do intercepto - valor da variável resposta Y quando a variável t (tempo) é igual a zero.

No segundo caso, consideramos a situação onde o intercepto não é aleatório, mas a inclinação da curva ao longo do tempo é considerada aleatória. Ou seja, a relação entre a variável resposta e determinada variável independente é diferente para cada um dos indivíduos. A equação é apresentada abaixo:

$$Y_{ij} = \beta_0 + \beta_{1i} t_{ij} + \varepsilon_{ij} \quad i=1, \dots, n \text{ e } j=1, \dots, m_i, \quad (3)$$

em que β_0 é o intercepto fixo, β_{1i} é o coeficiente de regressão aleatório para o tempo, e ε_{ij} é o erro para o i -ésimo indivíduo na observação j .

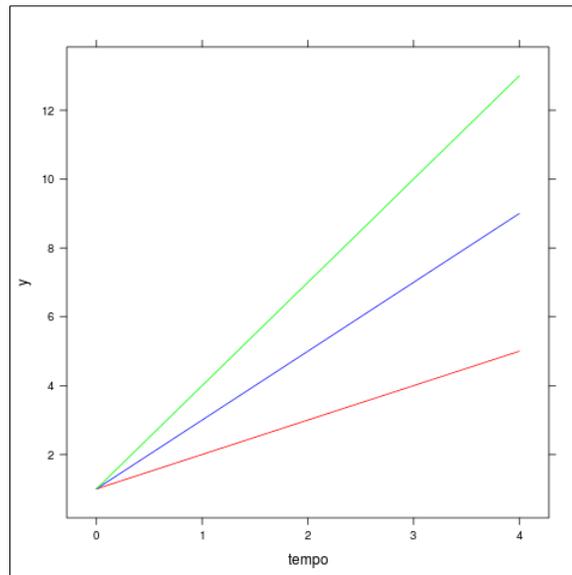


Figura 2: Evolução da variável resposta Y com diferentes inclinações.

Na figura 2 acima, os 3 indivíduos possuem o mesmo intercepto porém diferentes inclinações. A inclinação mede o aumento ou redução da variável resposta Y para cada aumento de uma unidade da variável t (tempo).

No terceiro caso, uma forma mais geral considera a combinação de intercepto e inclinações diferentes ao longo do tempo, ou seja, neste caso tanto o intercepto quanto a inclinação são aleatórios.

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \varepsilon_{ij} \quad i=1, \dots, n \text{ e } j=1, \dots, m_i, \quad (4)$$

em que β_{0i} é o intercepto aleatório, β_{1i} é o coeficiente de regressão aleatório para o tempo, e ε_{ij} é o erro para o i -ésimo indivíduo na observação j .

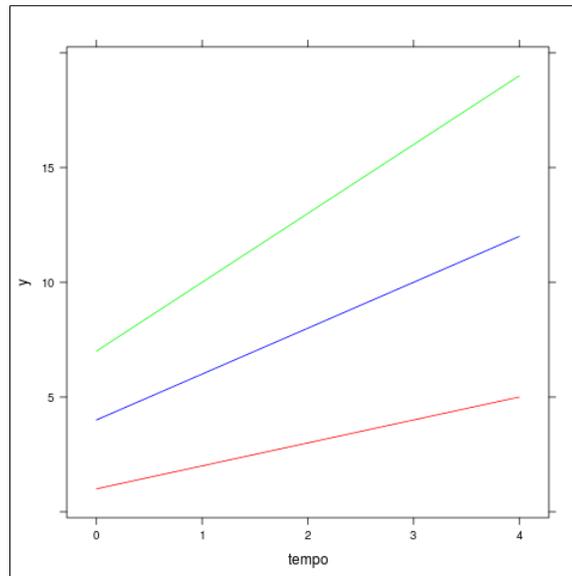


Figura 3: Evolução da variável resposta Y com diferentes inclinações e interceptos.

A forma geral para o modelo linear misto proposto por Laird & Ware (1982) e que expressa o vetor resposta y_i de dimensão n_i para o i -ésimo de M indivíduos é dada por:

$$y_i = X_i \boldsymbol{\beta} + Z_i \mathbf{b}_i + \epsilon_i, \quad i = 1, \dots, M,$$

$$\mathbf{b}_i \sim N(\mathbf{0}, \Sigma) \tag{5}$$

$$\epsilon_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

onde $\boldsymbol{\beta}$ é o vetor de dimensão p de efeitos fixos, \mathbf{b}_i é um vetor de efeitos aleatórios de dimensão q e que assume-se ser normalmente distribuído com média zero e matriz de variância-covariância Σ , e ϵ_i é um vetor de erros (dentro dos indivíduos) de dimensão n_i que também segue uma distribuição normal. Os efeitos aleatórios \mathbf{b}_i e os erros dentro dos indivíduos assumem-se serem independentes para diferentes indivíduos e independentes um do outro para um mesmo indivíduo. X_i (de tamanho $n_i \times p$) é a matriz de desenho observada relacionada com os efeitos fixos, $X_i \boldsymbol{\beta}$ é a componente global ou fixa do modelo. Z_i (de tamanho $n_i \times q$) é matriz de desenho para o indivíduo i ; $Z_i \mathbf{b}_i$ representa os efeitos aleatórios para o indivíduo i .

4.3 Escolha do Modelo

É uma importante tarefa selecionar o “melhor” modelo, ou seja, um modelo que é parcimonioso em termos do número de parâmetros utilizados, e que ao mesmo tempo tem o melhor ajuste (West *et al.*, 2007). Ao selecionar o “melhor” modelo para um determinado conjunto de dados, levamos em conta os objetivos da investigação, o desenho de estudo, o conhecimento prévio sobre importantes preditores e considerações relevantes sobre o assunto em questão (West *et al.*, 2007).

Para a seleção do modelo adequado, os critérios mais comuns são o AIC (Critério de Informação de Akaike) e o BIC (Critério de Informação Bayesiano). O AIC e BIC são definidos como:

$$\begin{aligned} AIC &= -2\log Lik + 2n_{par} \\ BIC &= -2\log Lik + n_{par} \log(N) \end{aligned} \tag{6}$$

onde *Lik* representa a verossimilhança, n_{par} o número de parâmetros do modelo e *N* o número total de observações usadas para ajustar o modelo. Como decisão, adota-se como “melhor” modelo aquele que apresenta o menor valor em qualquer um desses critérios.

4.4 Diagnóstico do Modelo

Após a modelagem, é importante efetuar o diagnóstico do modelo para verificar se os pressupostos básicos de distribuição para os modelos de efeitos mistos são verificados e se o modelo proposto é sensível a observações não usuais. O processo de diagnóstico do modelo envolve técnicas formais e informais. Métodos de diagnóstico para modelos de regressão linear estão bem estabelecidas na literatura. Em contraste, diagnósticos para modelos de efeitos mistos são mais difíceis de executar e interpretar, devido à complexidade do modelo (West *et al.*, 2007).

Segundo Pinheiro & Bates (2000), existem dois tipos básicos de pressupostos da distribuição dos modelos de efeitos mistos:

1. os erros dentro do grupo são independentes e identicamente distribuídos seguindo uma distribuição normal, com média zero e variância σ^2 , além de serem independentes dos efeitos aleatórios.
2. os efeitos aleatórios são normalmente distribuídos, com média zero e matriz de covariância Σ (não dependendo do grupo) e são independentes para diferentes grupos.

As bibliotecas *nlme* (Pinheiro & Bates, 2000) e *lme4* (Bates *et al.*, 2011) do R fornecem vários métodos para se avaliar a validade destas suposições. Os métodos mais usados para essa avaliação baseiam-se em gráficos tais como os de resíduos e os de valores ajustados. A validade das hipóteses de distribuição também pode ser formalmente avaliada através de testes de hipóteses. No entanto, raramente as conclusões de um teste de hipóteses sobre algum pressuposto do modelo contradizem as informações exibidas nos gráficos de diagnóstico (Pinheiro & Bates, 2000).

5 Materiais e Métodos

5.1 População e Desenho de Estudo

5.1.1 Participantes

Os pacientes do estudo aqui analisado eram acompanhados no ambulatório de Programa de Hipertensão Arterial - PROHART, do Serviço de Clínica Médica do Hospital Universitário Clementino Fraga Filho - HUCFF. Homens e mulheres não grávidas com idade entre 18 e 75 anos com IMC maior do que 25kg/m² e hipertensão arterial (PA \geq 140 x 90 mmHg) diagnosticada há pelo menos três meses foram considerados elegíveis para inclusão. Pacientes com osteoporose conhecida, doença hepática ou biliar ou que estavam tomando remédios para perda de peso foram excluídos. Eram considerados hipertensos os que recebiam qualquer tratamento anti-hipertensivo por no mínimo dois meses.

5.1.2 Desenho de Estudo

O ensaio clínico aleatorizado é um tipo de estudo prospectivo, cujo delineamento permite eliminar diversos vieses e fatores de confusão, pois os grupos de tratamento e controle são alocados aleatoriamente, e espera-se que as características sejam distribuídas de maneira semelhante em ambos os grupos (Buehler *et al.*, 2009). O ensaio clínico aleatorizado é também submetido a um controle e gerenciamento mais intensivos (Benson & Hartz, 2000).

O estudo analisado neste trabalho é um ensaio clínico aleatorizado realizado no ano de 2000 e cujos dados foram cedidos pela pesquisadora que coordenou a pesquisa. Os pacientes do grupo tratamento receberam o Orlistat e orientação dietética, com visitas mensais a nutricionistas e prescrição de dieta hipocalórica. Os pacientes do grupo controle receberam apenas orientação dietética com visitas mensais a nutricionistas e prescrição de dieta hipocalórica. O estudo foi

“aberto”, ou seja, tanto os médicos e nutricionistas como os pacientes, sabiam a qual grupo os pacientes pertenciam pois não foi utilizado placebo.

O estudo procedeu conforme a Declaração de Helsinki e foi aprovado pelo comitê de ética local. Todos os participantes assinaram um termo de consentimento.

5.1.3 Tamanho da Amostra

O cálculo para o tamanho da amostra do ensaio clínico foi feito estimando-se que 40% do grupo tratado com Orlistat e 20% do grupo controle teriam uma redução maior que 5% do peso corporal. Para um poder de 80% de significância seriam necessários 91 pacientes. Considerando ainda a possibilidade de uma perda de 20% de participação, foi definido que o tamanho da amostra seria de 110 pacientes em cada grupo. Na Figura 4 abaixo é apresentado o fluxo dos pacientes do ensaio clínico.

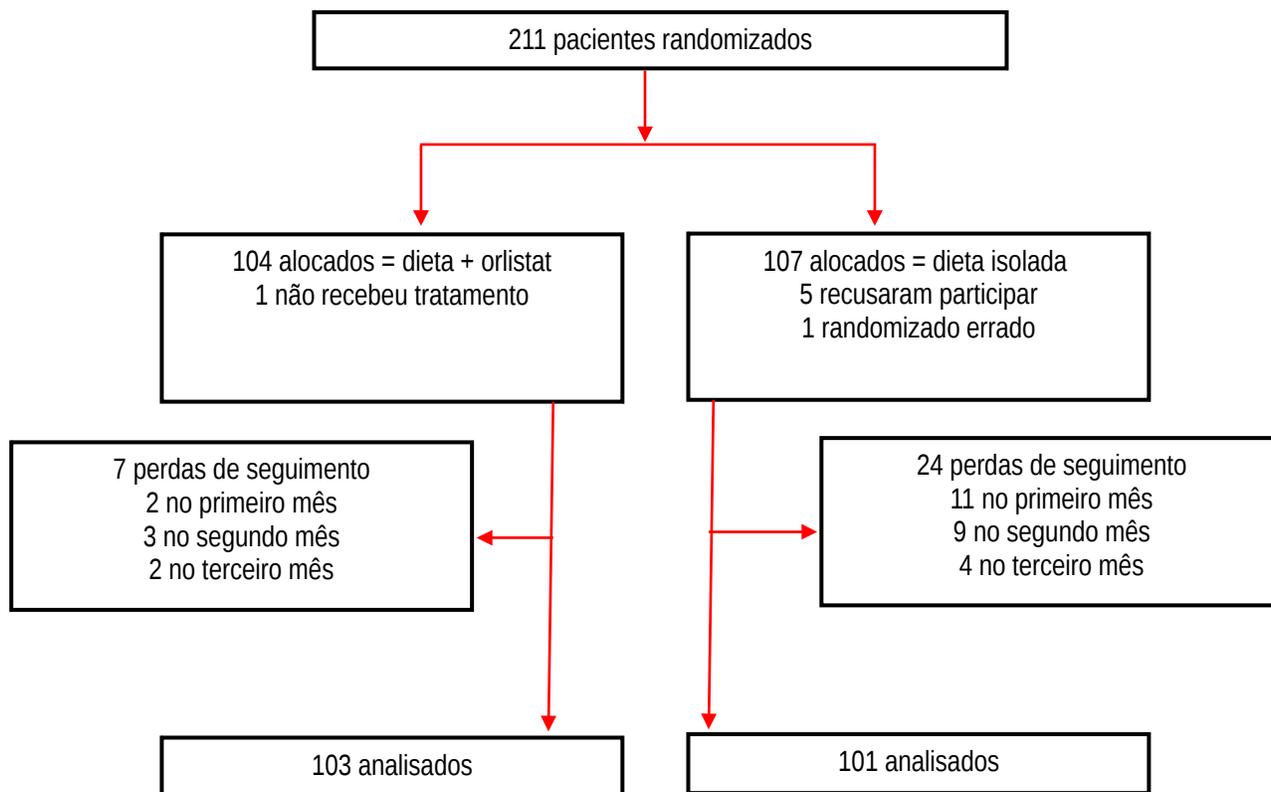


Figura 4: Fluxo do Ensaio Clínico

5.1.4 Prescrição da Medicação

Aos pacientes do grupo tratamento foi prescrito Orlistat 120 mg, três vezes ao dia (durante ou até uma hora após as refeições). Esses pacientes receberam a medicação para quatro semanas, após as quais retornavam à consulta para avaliação e recebiam medicação por mais quatro semanas, e assim subsequentemente.

Em caso de intolerância ao medicamento por efeitos colaterais tais como aumentos das evacuações e incontinência fecal, a medicação foi suspensa; em caso contrário o tratamento foi mantido por três meses, com avaliações mensais.

6 Aplicação

Neste capítulo são apresentados os resultados obtidos a partir do uso da metodologia descrita anteriormente. Este capítulo está estruturado em duas partes que consistem na análise exploratória dos dados e no ajuste do modelo linear misto. Toda a análise apresentada nesse capítulo foi realizada no software livre R (R Development Core Team, 2012) e as bibliotecas *lme4* (Bates *et al.*, 2011), *nlme* (Pinheiro *et al.*, 2011), *ggplot2* (Wickham, 2009) e *car* (Fox & Weisberg, 2011), disponíveis em seus repositórios.

6.1 Análise Exploratória

6.1.1 Descrição Geral

Esta seção apresenta a análise exploratória dos dados do ensaio clínico descrito anteriormente. Como o objetivo é avaliar o efeito do tratamento, todos os resultados exploratórios são apresentados considerando a alocação no ensaio clínico.

Dos 211 pacientes que foram randomizados no início do estudo, foram excluídos um que não recebeu o tratamento, cinco que se recusaram a participar e um que foi randomizado erradamente. Além disso, uma pessoa que não possuía a medida do peso inicial também foi excluída.

O número de pacientes que participaram desta análise foi de 203 sendo que 103 (19 homens e 84 mulheres) foram alocados para o grupo de tratamento e 100 (16 homens e 84 mulheres) para o grupo controle conforme a Tabela 1 abaixo:

Tabela 1: Número de pessoas que participaram da análise por sexo e grupo.

Sexo	Tratamento	Controle	Total
Masculino	19	16	35
Feminino	84	84	168
Total	103	100	203

Na primeira entrevista, os pacientes responderam a um questionário quanto a diversos fatores de risco cardiovascular: tabagismo, sedentarismo, história de doença coronariana precoce na família, história de hipertensão na família, comorbidade, diabetes, dislipidemia e se nos últimos três meses praticou alguma atividade física. Essas características podem ser observadas na Tabela 2.

A maioria dos indivíduos, nos dois grupos, é sedentária e diabética, apresenta alguma comorbidade, tem história de hipertensão na família, tem dislipidemia, não tem história de doença coronariana precoce na família e não fuma.

Tabela 2: Características dos participantes do estudo quanto a certos fatores de risco cardiovascular.

Variável	Tratamento n (%)		Controle n (%)	
	Sim	NA	Sim	NA
Tabagismo	6(0,06)	2(0,02)	12(0,12)	4(0,04)
Sedentarismo	60(0,6)	2(0,02)	67(0,65)	4(0,04)
Histcor*	31(0,31)	10(0,1)	29(0,28)	12(0,12)
Histha**	77(0,77)	7(0,07)	82(0,8)	7(0,07)
Comorbidade	48(0,48)	0(0)	43(0,42)	0(0)
Diabetes	37(0,37)	3(0,03)	38(0,37)	5(0,05)
Dislipidemia	53(0,53)	5(0,05)	66(0,64)	7(0,07)
Ativ. Fis. 3m***	21(0,21)	36(0,36)	41(0,4)	16(0,16)

* - história de doença coronariana precoce na família

** - história de hipertensão na família

*** - atividade física nos últimos 3 meses

NA - não disponível

No início do estudo a média de peso foi de 90 kg no grupo tratamento e 85 kg no grupo controle. Em relação à idade, a média foi de 55 anos no grupo tratamento e 57 anos no grupo controle.

Para apresentação destes resultados, utilizamos o box-plot (“diagrama em caixa”), que mostra os valores centrais (mediana), a dispersão (primeiro e terceiros quartis) e os valores máximos e mínimos. O gráfico box-plot dos pesos e das idades (Figura 5) foi utilizado para ilustrar a semelhança das distribuições destas duas variáveis entre os grupos.

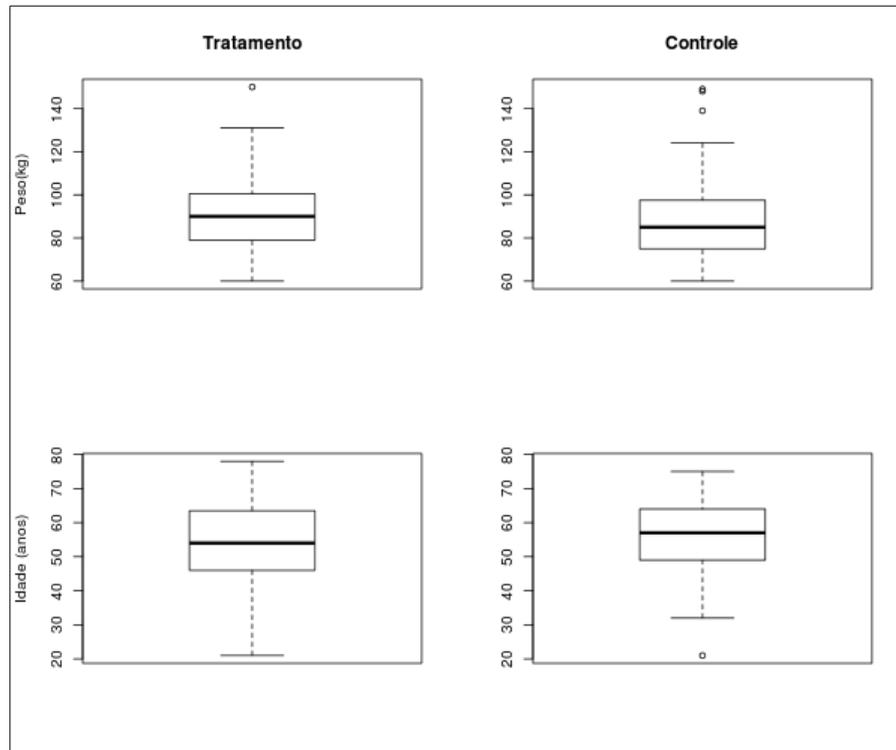
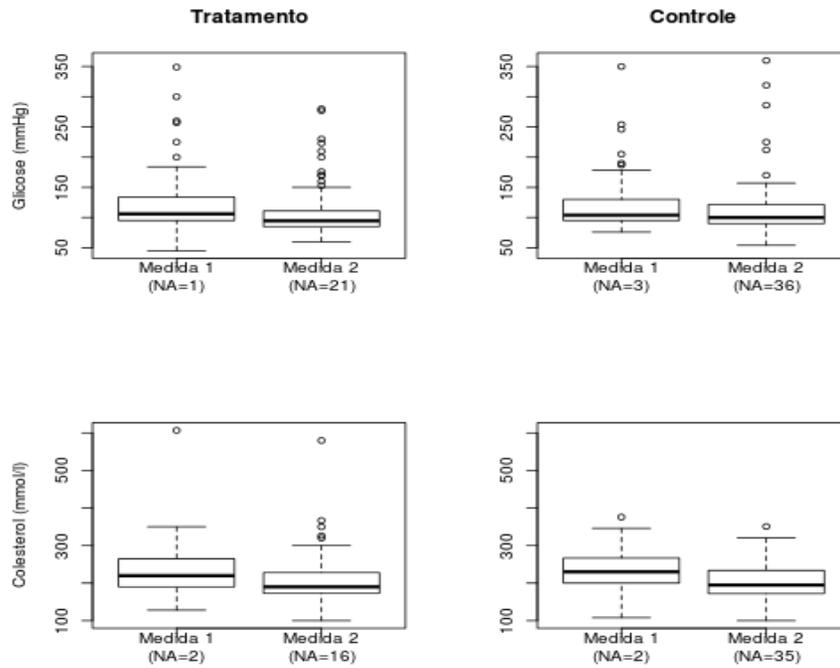


Figura 5: Box-plot do peso (kg) e idade (anos) no início do estudo.

Medidas laboratoriais foram colhidas em dois momentos: no início e no fim do estudo. Estas podem ser visualizadas na Figura 6.

(a) Glicose e Colesterol



(b) HDL e Triglicerídeos

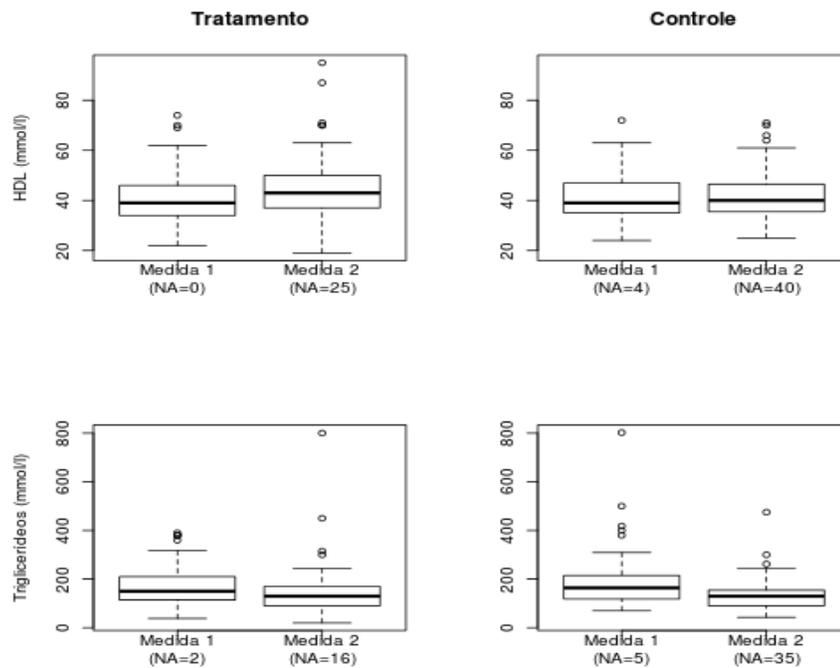


Figura 6: Box-plot das medidas laboratoriais.

A mediana das medidas iniciais de glicose e colesterol é semelhante para os dois grupos. O mesmo acontece com as medidas finais. Nota-se também que a mediana da segunda medida em todos os casos é menor que a da primeira medida.

Observando as variáveis HDL e triglicérides, pode-se perceber que a distribuição nos dois grupos é bastante similar nas duas medidas e que não há aparentemente uma significativa mudança nestes níveis ao fim do estudo.

Em relação a cada consulta (1, 2 e 3 meses após o início do estudo), 96 pacientes do grupo tratamento e 77 do grupo controle foram à terceira consulta, ou seja 93,2% do grupo tratamento e 77% do grupo controle continuaram no estudo até o final (Tabela 3). Mas isso não significa que quem ficou até o final foi em todas as consultas. Algumas pessoas foram, por exemplo, na terceira consulta mas não foram na segunda ou na primeira. Basta notarmos que 96 pessoas do grupo tratamento foram na terceira consulta porém na segunda apenas 94 deste mesmo grupo compareceram. Da mesma maneira acontece com o grupo controle, onde 77 participantes foram à terceira consulta porém 74 tinham ido na segunda. Isso pode ser melhor observado na Tabela 4.

Tabela 3: Número de pacientes que foram a cada uma das consultas.

Consultas	Tratamento	Controle	Total
0 (baseline)	103	100	203
1	100	85	185
2	94	74	168
3	96	77	173

Uma hipótese levantada para explicar o porquê de tantas perdas no grupo controle é o fato do estudo ter sido aberto, o que explica o maior número de perdas nesse grupo pois muitos desses pacientes já tinham história de insucessos com dietoterapia e desistiram de participar.

Tabela 4: Número de medidas do peso.

Medidas	Tratamento	Controle	Total
1	1	8	9
2	4	15	19
3	8	10	18
4	90	67	157

Uma das justificativas para o uso do tipo de modelagem que utilizaremos neste trabalho é o fato dos dados não serem balanceados. Para termos uma ideia melhor do intervalo de cada consulta, construímos o histograma abaixo. Nele podemos verificar que a grande maioria dos pacientes tem as medidas colhidas num mesmo espaço de tempo, 28 dias, porém, em torno de 10% dos pacientes não apresentam regularidade nos intervalos de cada medida.

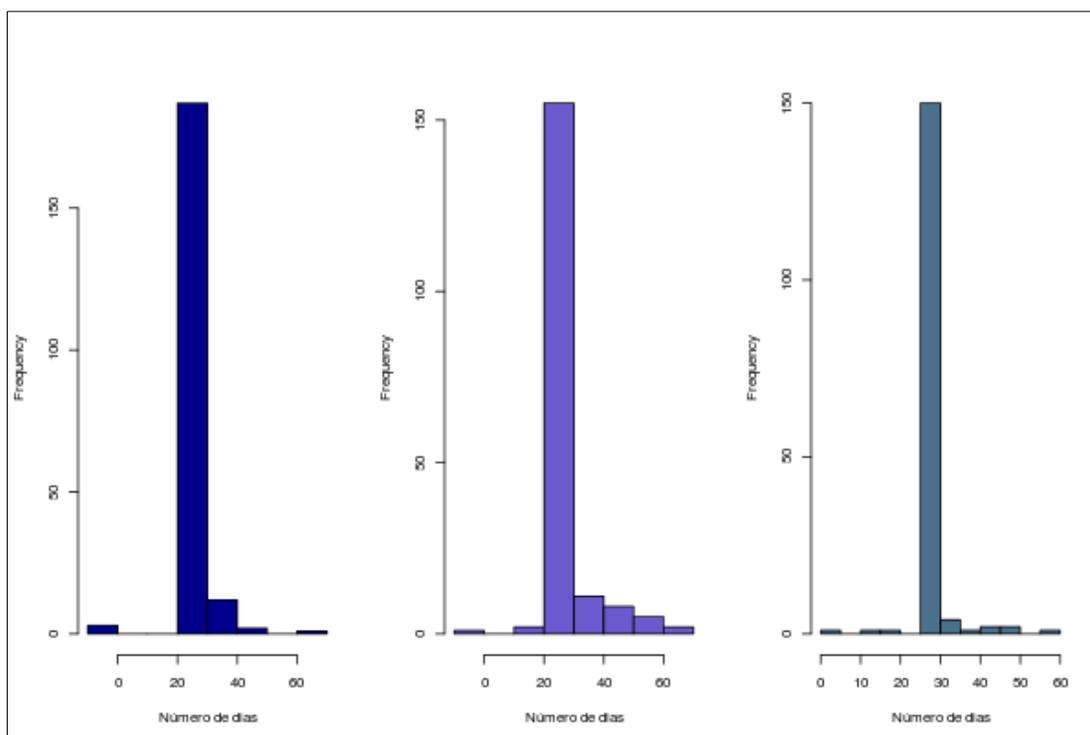


Figura 7: Histogramas do número de dias entre o início do estudo e a primeira consulta, entre a primeira e a segunda consulta e entre a segunda e terceira consulta.

Na figura abaixo podemos observar a evolução do peso ao longo do estudo para ambos os grupos.

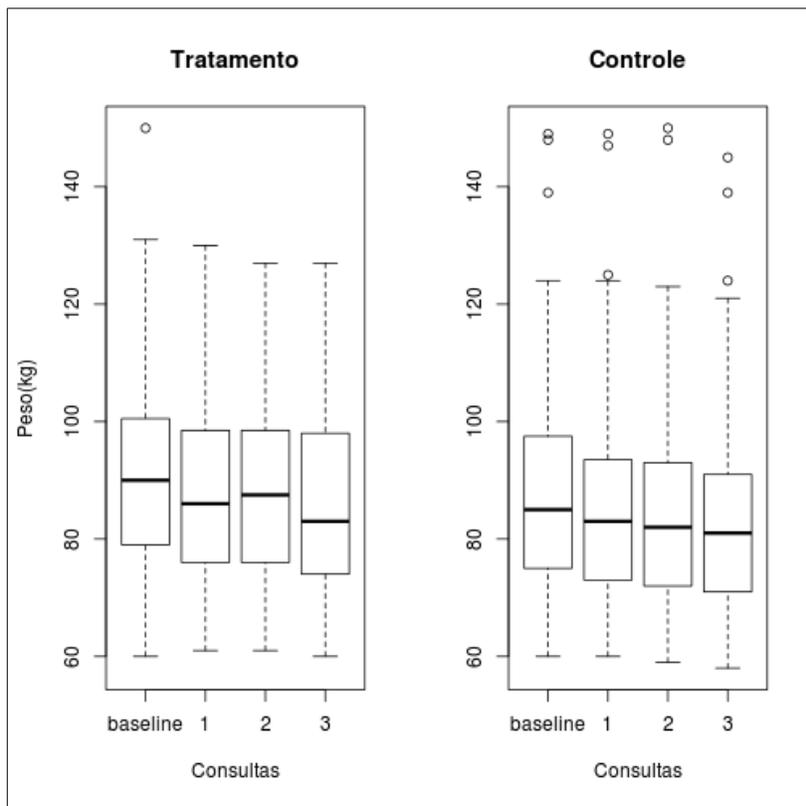


Figura 8: Box-plot do peso no início do estudo e nas consultas seguintes nos dois grupos.

Assim como o peso inicial (baseline), as outras medições do peso, em ambos os grupos, não apresentam grande variabilidade entre os meses subsequentes. Em geral o peso diminui ao longo do tempo nos dois casos, porém observa-se um ligeiro aumento na segunda medida no grupo tratamento. É observado também alguns “outliers” no grupo controle, isto é, alguns pacientes apresentam valores de peso que destoam do peso do seu grupo.

Uma outra maneira interessante de representar essa evolução do peso utilizando o box-plot pode ser visto na Figura 9. Ele mostra os intervalos de confiança para a mediana. Para isto, basta adicionar um entalhe em cada um deles. Quando os entalhes nas caixas relativas aos diferentes grupos não se sobrepõem, temos fortes evidências que as medianas são diferentes (Chambers *et al.*, 1983). É um gráfico útil para as medianas e que nos indica se há ou não diferença nas medianas entre os grupos. Porém, este gráfico não deve substituir o tradicional teste da mediana. Observando o

gráfico abaixo, podemos perceber que os entalhes se sobrepõem, não indicando, portanto, forte evidência de que as medianas dos grupos tratamento e controle são diferentes.

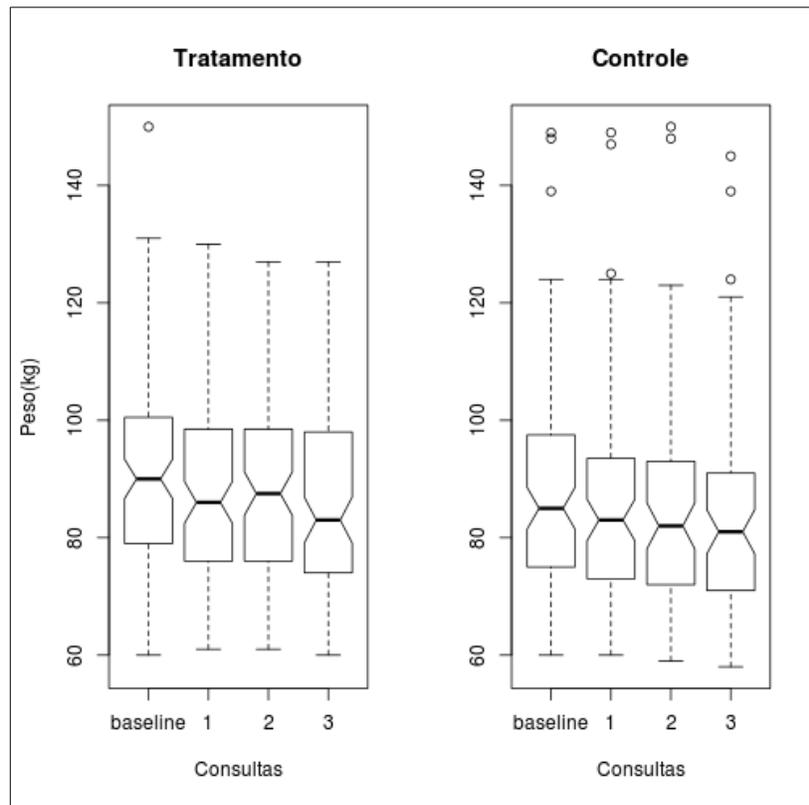


Figura 9: Box-plot com os intervalos de confiança para a mediana do peso no início do estudo e nas consultas seguintes.

6.1.2 Descrição Individual

Uma das representações gráficas mais importantes na análise descritiva longitudinal é o gráfico de perfis individuais e médios (Ker, 2010). Esse gráfico mostra a relação entre a variável resposta no eixo y - no nosso caso o peso - e o tempo no eixo x - no nosso caso as consultas (baseline, consultas 1, 2 e 3). Dentre os objetivos deste tipo de representação gráfica podemos destacar: identificar observações discrepantes, comparar as variabilidades de cada grupo, sugerir possíveis modelos para explicar a variação da trajetória média ao longo do tempo (Ker, 2010).

Nas figuras abaixo (Figura 10, Figura 11, Figura 12 e Figura 13) observamos diferentes e interessantes formas deste tipo de representação gráfica. As trajetórias individuais de cada paciente ao longo tempo, ou seja, o seu peso ao longo das consultas, podem ser observadas na Figura 10. As linhas de cor vermelha são os pacientes do grupo tratamento e as linhas azuis são os do grupo controle.

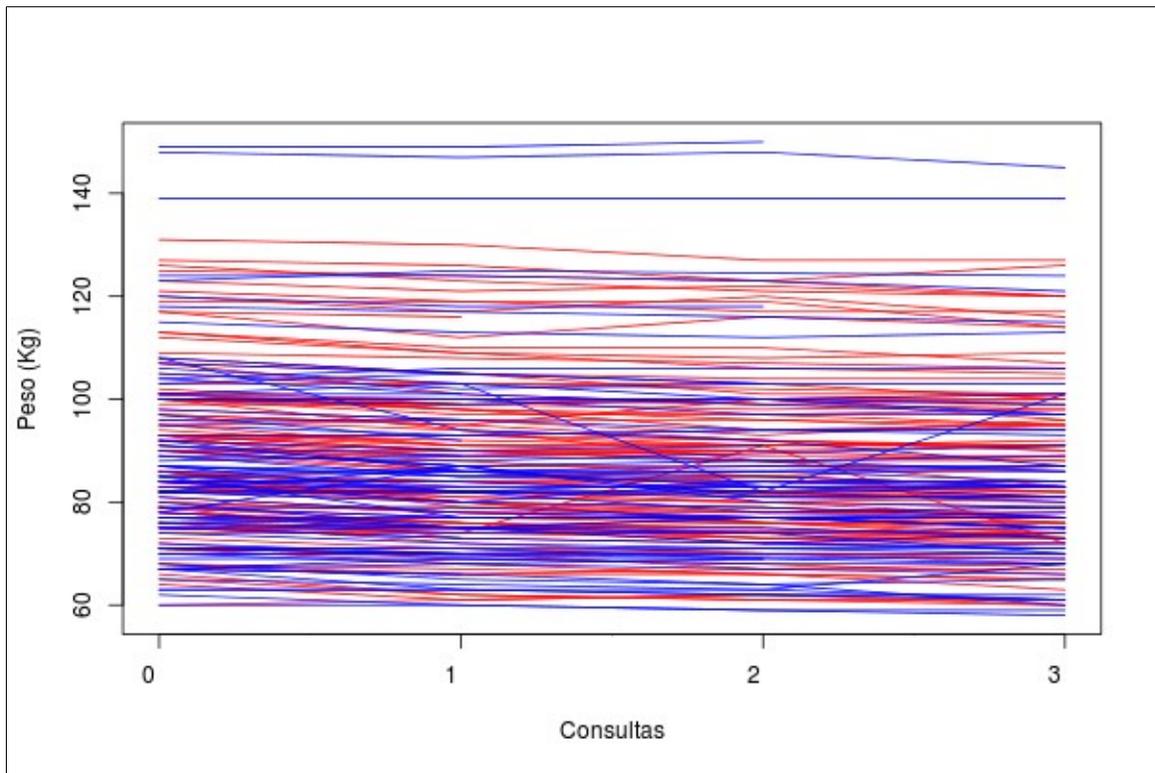


Figura 10: Trajetórias individuais de cada paciente do grupo tratamento (em vermelho) e do grupo controle (em azul).

Através do gráfico acima vê-se que há um ligeiro declínio do peso ao longo das consultas em ambos os grupos. Verificamos três linhas mais acima no gráfico e que são justamente os “outliers” observados anteriormente nos box-plots. Além disso, observa-se que alguns indivíduos tem uma trajetória oscilante, em uma consulta perdem muito e em outra ganham muito peso. É uma situação pouco provável e que possivelmente seja um erro de aferição ou de digitação.

Para se ter uma ideia melhor desta evolução individual por grupo, a Figura 11 mostra um gráfico para cada um deles. Já na Figura 12 é exibida uma outra forma gráfica na qual sobrepõe-se

esta evolução ao gráfico de box-plot. Pode parecer um gráfico confuso, porém ele é bastante útil para observar a trajetória individual de certa variável e informações sobre a sua distribuição.

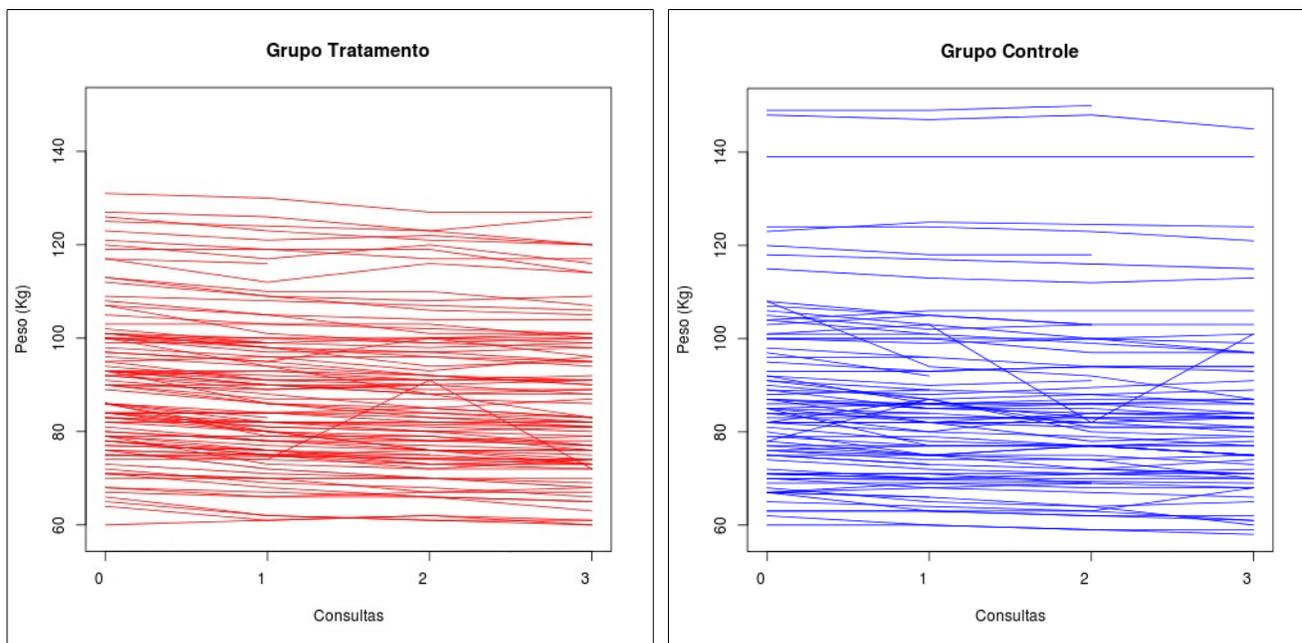


Figura 11: Trajetórias individuais de cada paciente dos grupos tratamento e controle.

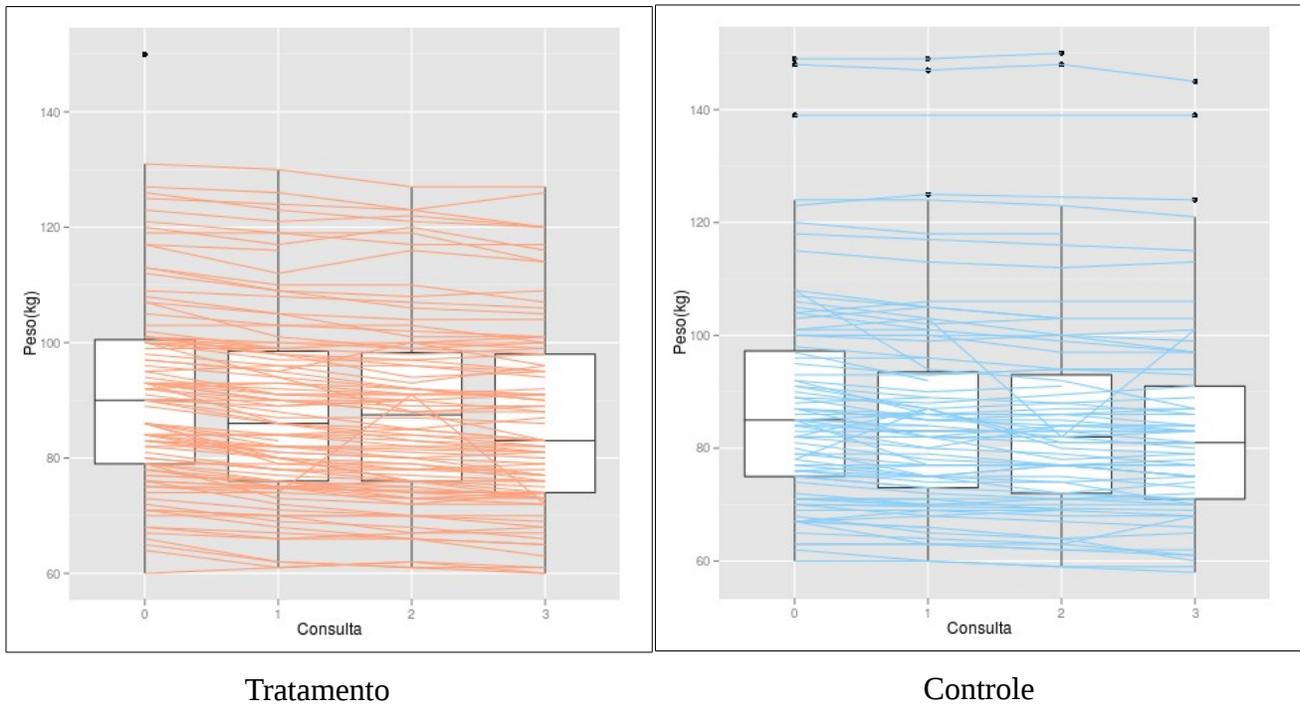


Figura 12: Trajetórias individuais de cada paciente por grupo sobrepostas pelo box-plot.

Pelas figuras acima, observa-se que há uma diminuição do peso ao longo do tempo nos dois grupos, com a maior perda de peso ocorrendo no primeiro mês do estudo. Além disso, observa-se também, que esta perda é praticamente linear.

Para se ter uma ideia geral de cada grupo, os gráficos abaixo (Figura 13) mostram o peso médio de cada grupo e sua evolução ao longo das três consultas.

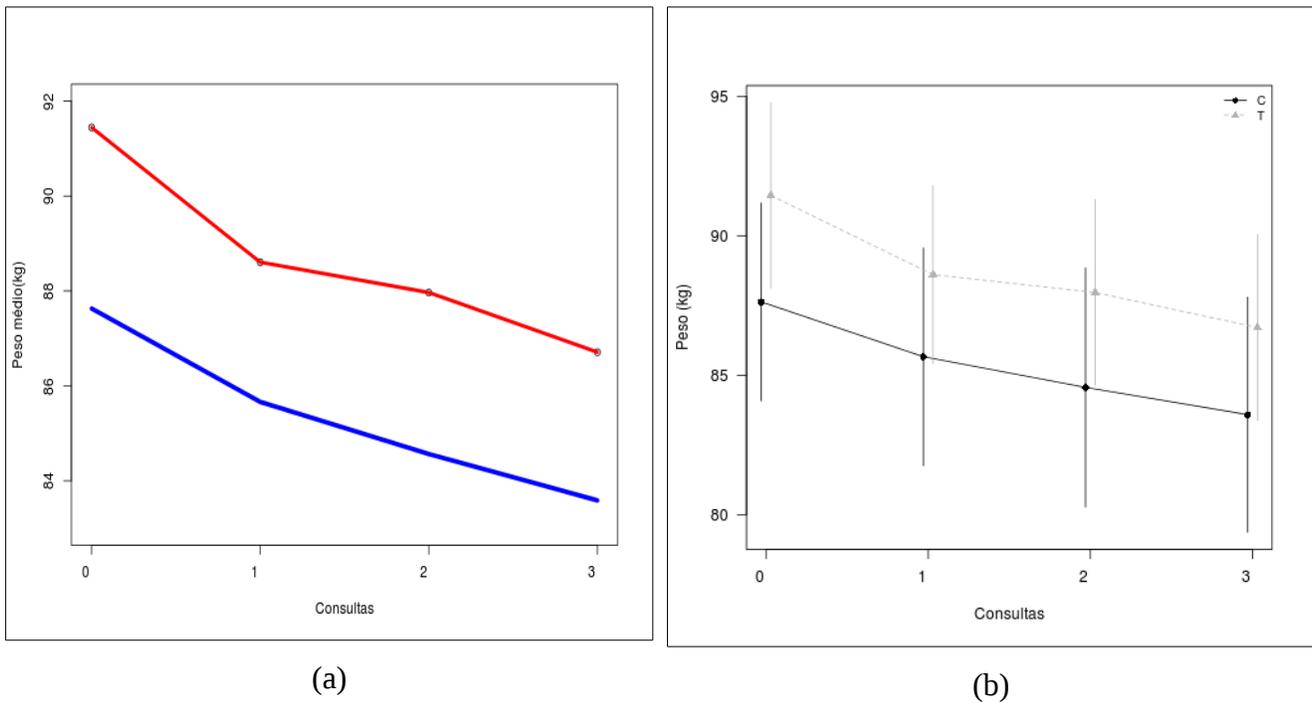


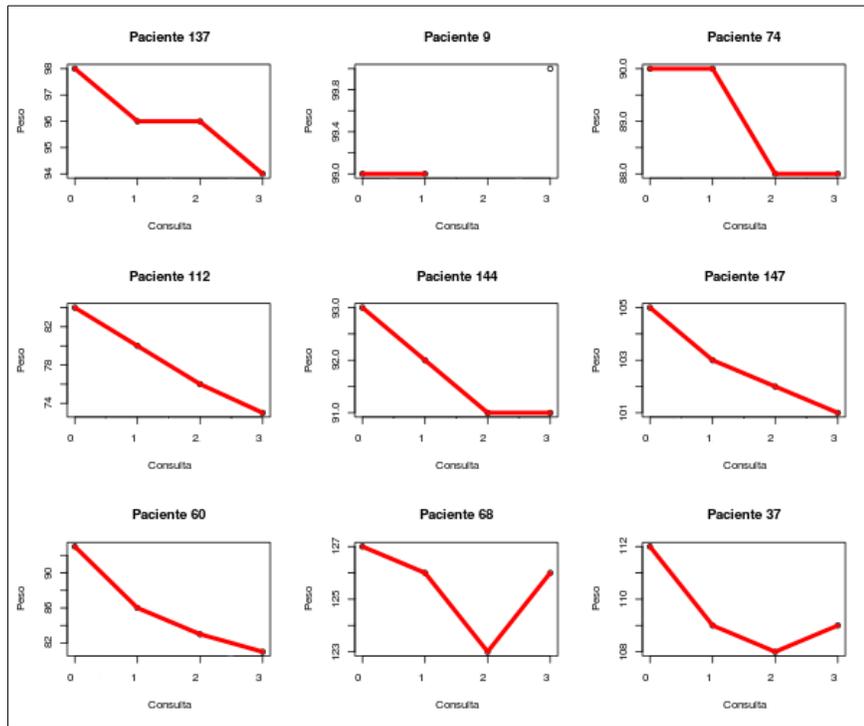
Figura 13: Perfil médio do peso de cada grupo ao longo das consultas.

O primeiro gráfico é uma maneira mais simples de se representar esta evolução. Já no segundo utiliza-se uma forma mais elaborada no qual é adicionada a variação do intervalo do peso médio em cada consulta e em cada um dos grupos. Ou seja, mais informação sem sobrecarregar a figura.

O perfil médio por consulta, definido como a média do peso por consulta por grupo, apresenta um decaimento ao longo do tempo para ambos os grupos como já foi observado anteriormente.

A Figura 14, a seguir, contém perfis individuais de alguns pacientes selecionados aleatoriamente e sugere a existência de uma grande variabilidade entre os indivíduos.

(a) Grupo Tratamento



(b) Grupo Controle

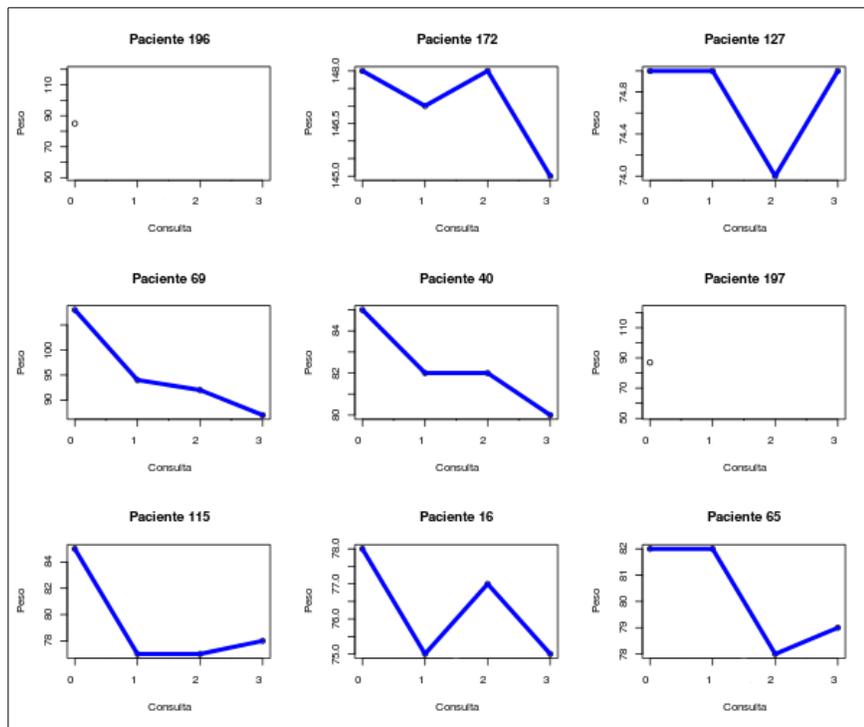
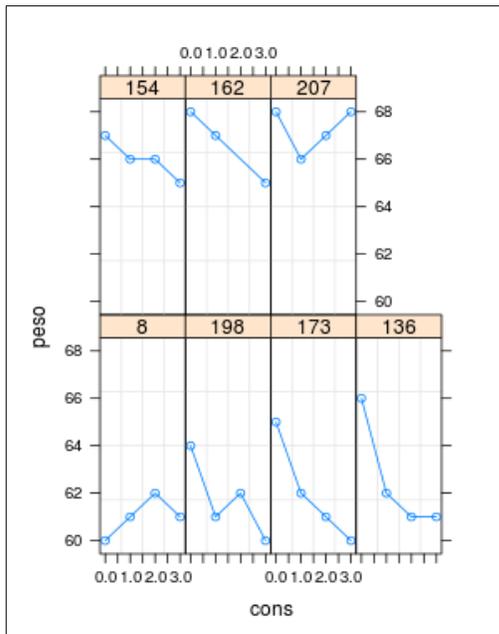
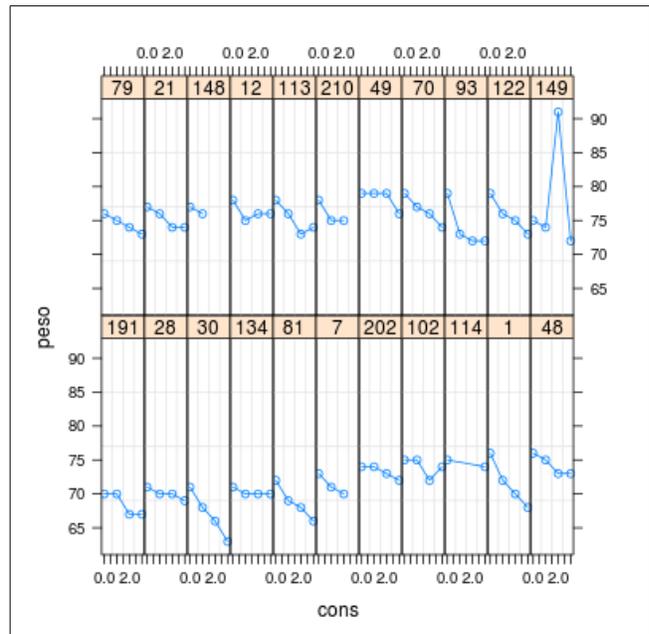


Figura 14: Perfis individuais do peso por grupo.

Como neste ensaio clínico temos uma grande quantidade de pacientes e cujos pesos variaram de 58kg a 150 kg, a maneira encontrada para ilustrar a evolução do peso de cada paciente e de cada grupo ao longo das consultas foi organizando cada um de acordo com o seu peso inicial. O primeiro gráfico à esquerda na figura abaixo, por exemplo, mostra todos os pacientes cujo peso inicial era menor que 70kg e que pertenciam ao grupo tratamento.



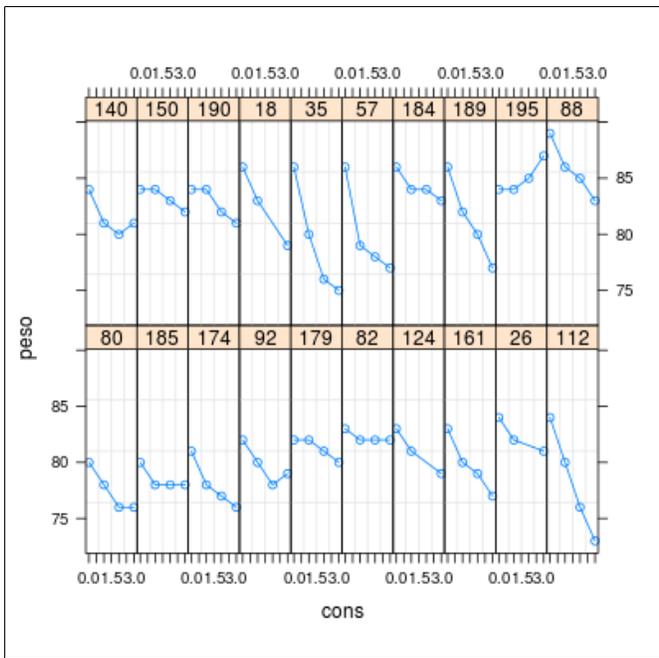
(a)



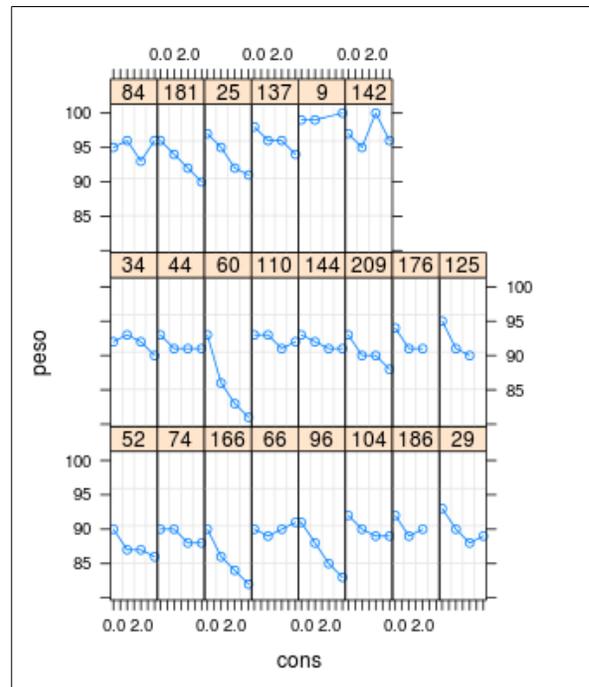
(b)

Figura 15: Evolução do peso de cada paciente do grupo tratamento cujo peso inicial era menor que 70kg (a) e entre 70 e 80kg (b).

Observe, por exemplo, o paciente 149 (figura 15-b). Ele começou o estudo pesando 75kg, depois diminuiu para 74kg, em seguida aumentou para 91kg, e na última consulta diminuiu para 72kg. É pouco provável que este paciente tenha ganho e perdido tanto peso no curto espaço de tempo deste estudo. Este paciente é considerado, portanto, um possível candidato a outlier deste banco de dados. Outros possíveis candidatos a outliers são os pacientes: 194 (figura 17-e), 183 (figura 18-b), 128 (figura 19-e), entre outros.

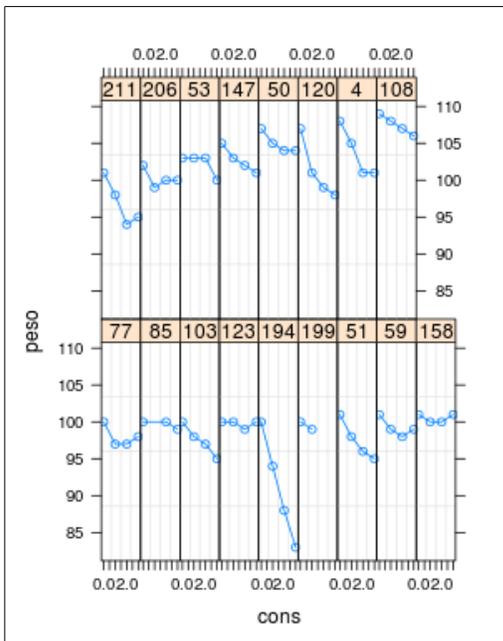


(c)

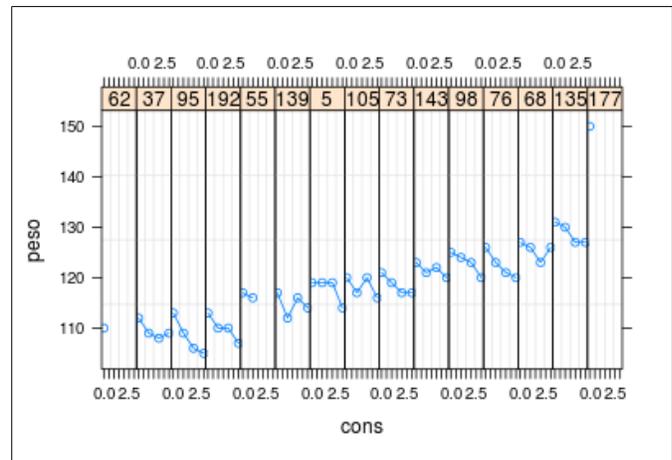


(d)

Figura 16: Evolução do peso de cada paciente do grupo tratamento cujo peso inicial era entre 80 e 90kg (c) e entre 90 e 100kg (d).

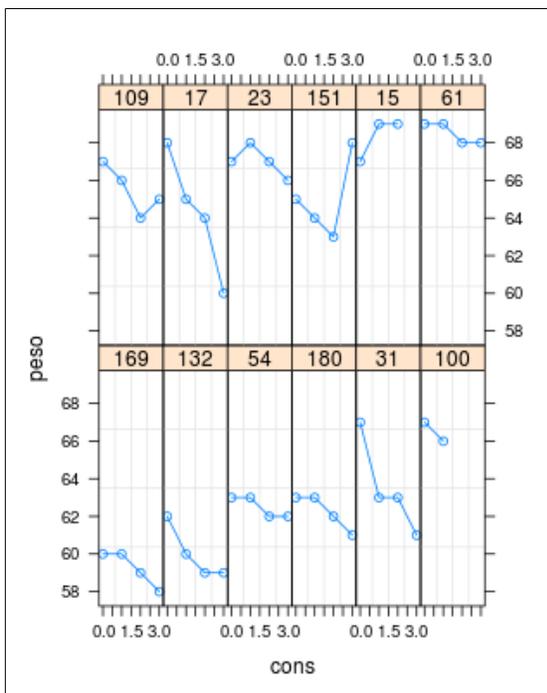


(e)

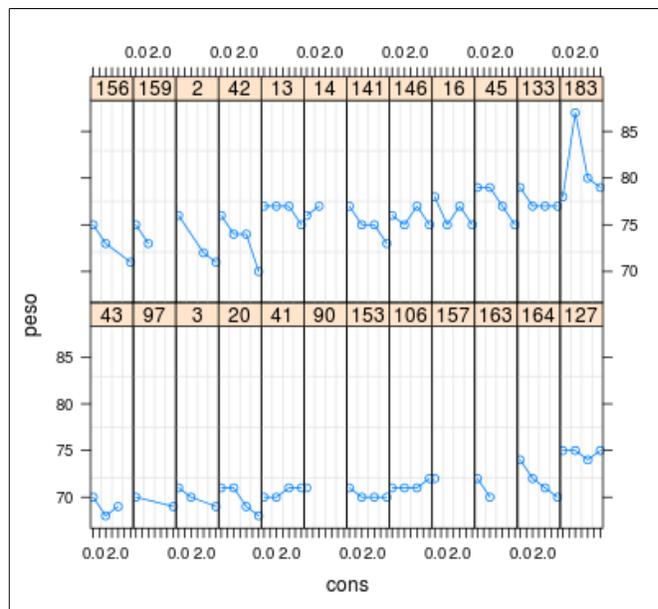


(f)

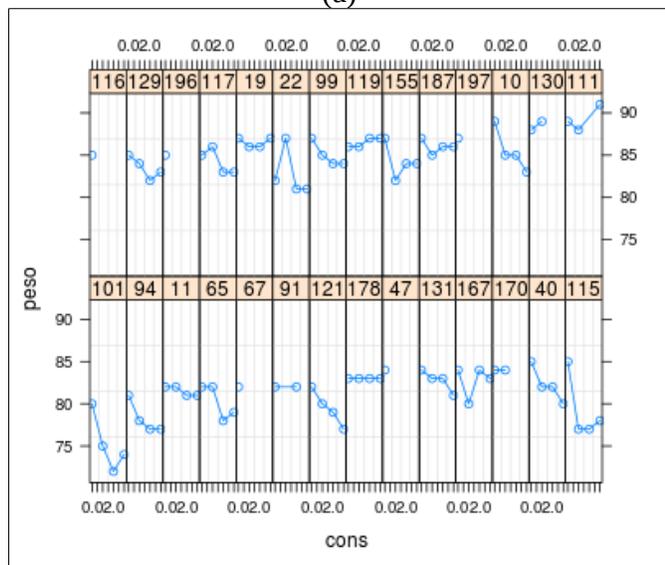
Figura 17: Evolução do peso de cada paciente do grupo tratamento cujo peso inicial era entre 100 e 110kg (e) e maior que 110kg (f).



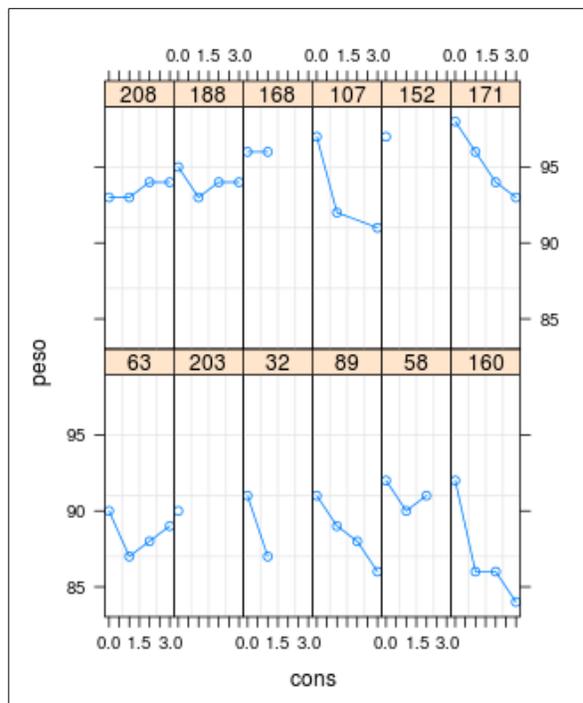
(a)



(b)

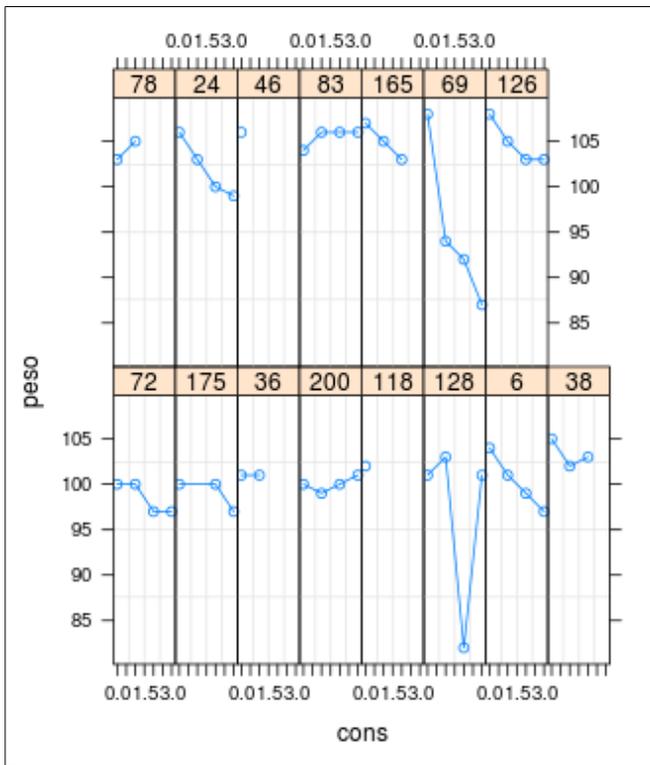


(c)

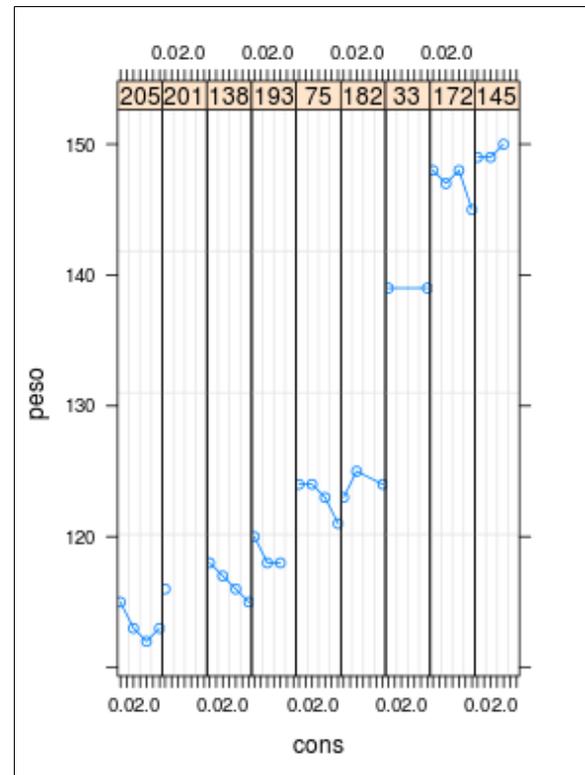


(d)

Figura 18: Evolução do peso de cada paciente do grupo controle cujo peso inicial era menor que 70kg (a), entre 70 e 80kg (b), 80 e 90kg (c) e entre 90 e 100kg (d).



(e)



(f)

Figura 19: Evolução do peso de cada paciente do grupo controle cujo peso inicial era entre 100 e 110kg (e) e maior que 110kg (f).

A relação mais importante de se observar em dados longitudinais com inúmeros indivíduos, é a evolução da variável resposta ao longo do tempo por indivíduo. Esse gráfico, em que os dados para indivíduos diferentes são mostrados em painéis separados, permite a verificação e a comparação da evolução do peso em cada paciente. Nos gráficos apresentados anteriormente temos uma ideia clara da evolução de cada paciente. Porém, podemos ilustrar essa mesma evolução utilizando um gráfico visualmente mais simpático por apresentar cores diferentes para cada paciente.

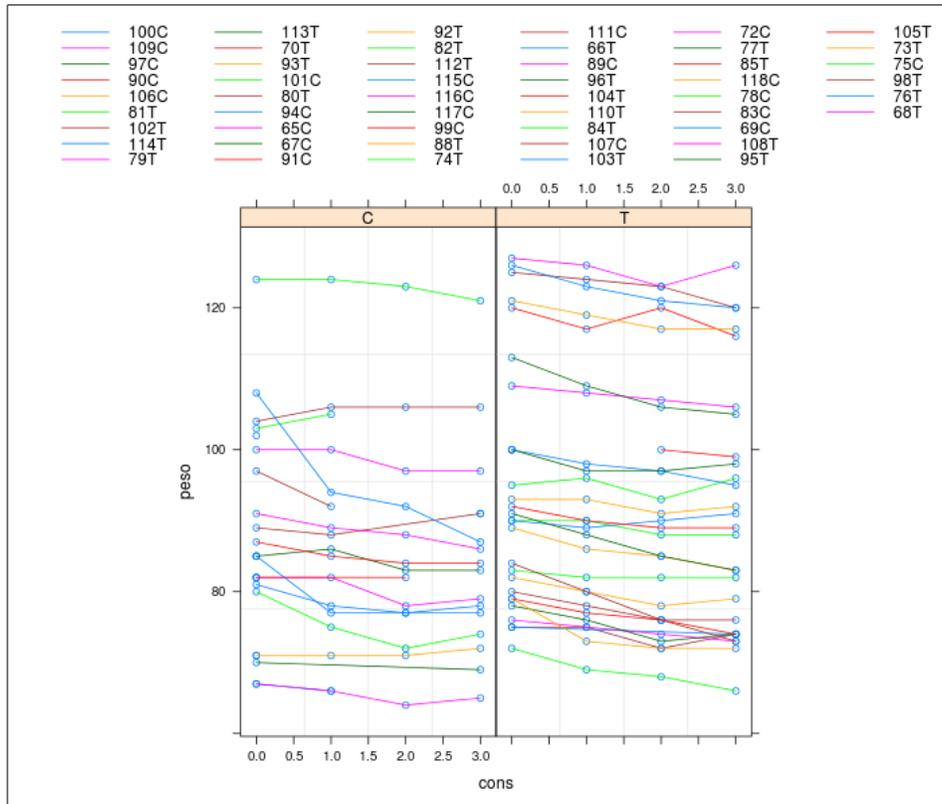
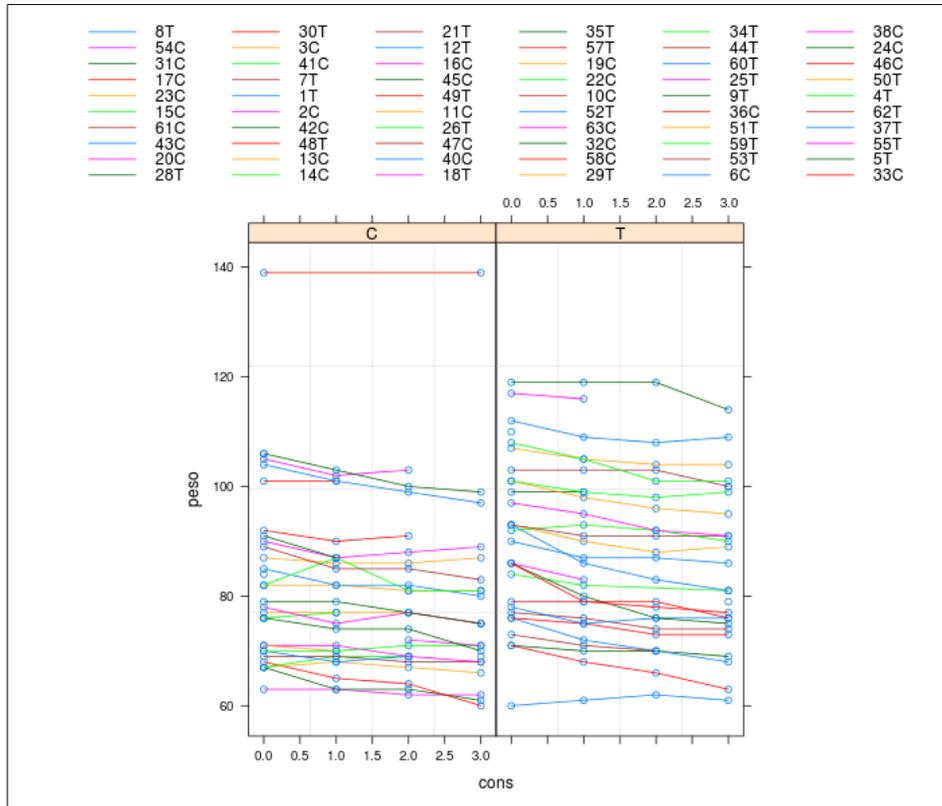


Figura 20: Trajetórias individuais do peso.

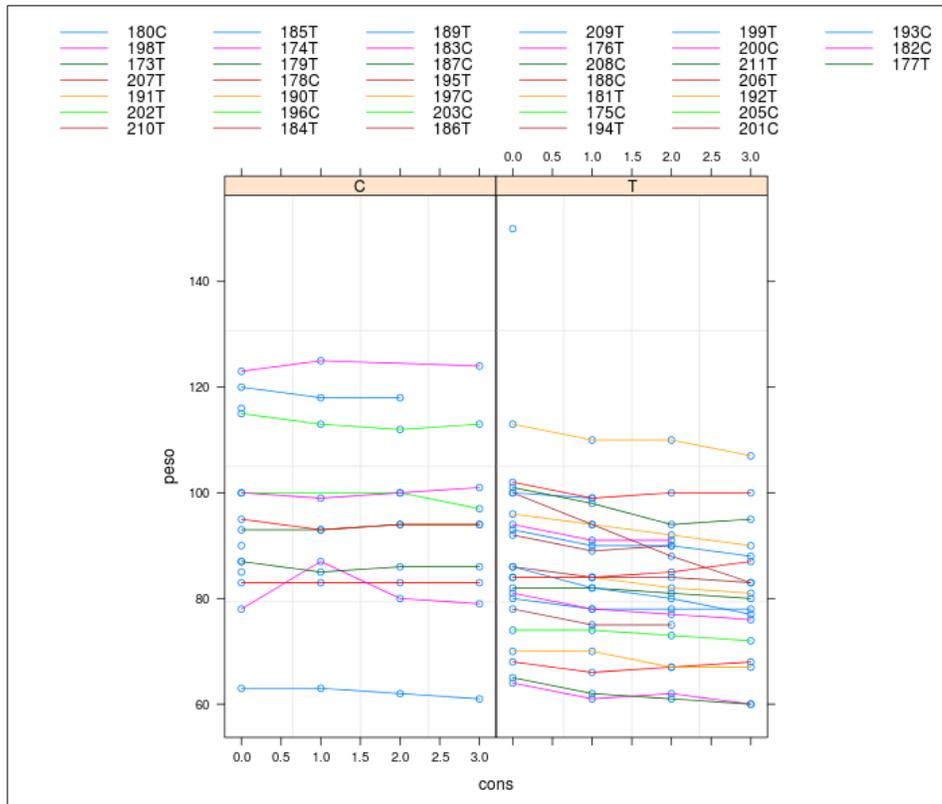
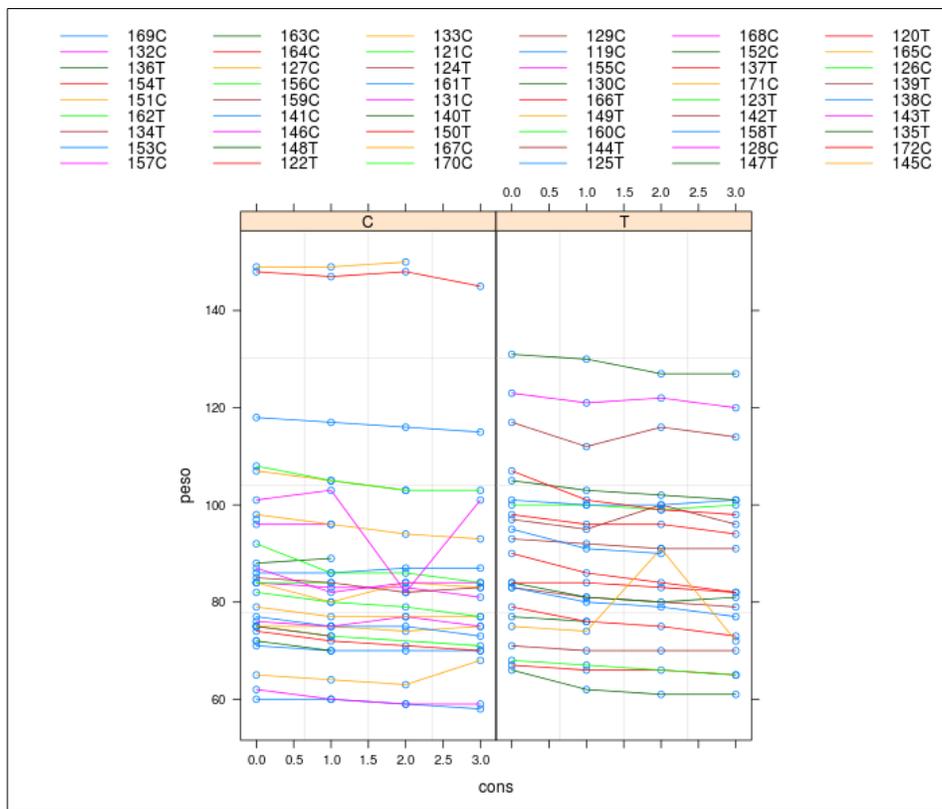


Figura 21: Trajetórias individuais do peso.

Além dos gráficos mostrados acima, uma outra forma de se visualizar dados, e que vem sendo crescentemente utilizada ao longo dos últimos anos, é a visualização de dados de uma forma dinâmica. Visualiza-se as imagens gráficas em movimento, e assim, é possível observar mais facilmente padrões nos dados.

Este gráfico pode ser acessado em: <http://goo.gl/3U5dD>. Nele pode-se escolher quais variáveis usar em cada eixo, assim como a cor e o tamanho dos círculos. O indicado, porém, é escolher as variáveis idade para o eixo y, peso para o eixo x, grupo para a cor (“color”), e o tamanho dos círculos (“size”) como sendo do mesmo tamanho (“same size”). Para visualizar este gráfico em movimento, basta apertar o botão “play” que fica no canto inferior à esquerda. Assim, veremos a evolução do peso ao longo das consultas. Observe que a consulta inicial está representada pelo número 2000, a consulta seguinte por 2001, a segunda consulta por 2002 e a terceira por 2003. Este é um formato padrão para este tipo de gráfico.

Abaixo, estão representados os pesos dos pacientes em cada uma das consultas onde os círculos de cor verde são os pacientes do grupo tratamento e os de cor azul os do grupo controle.

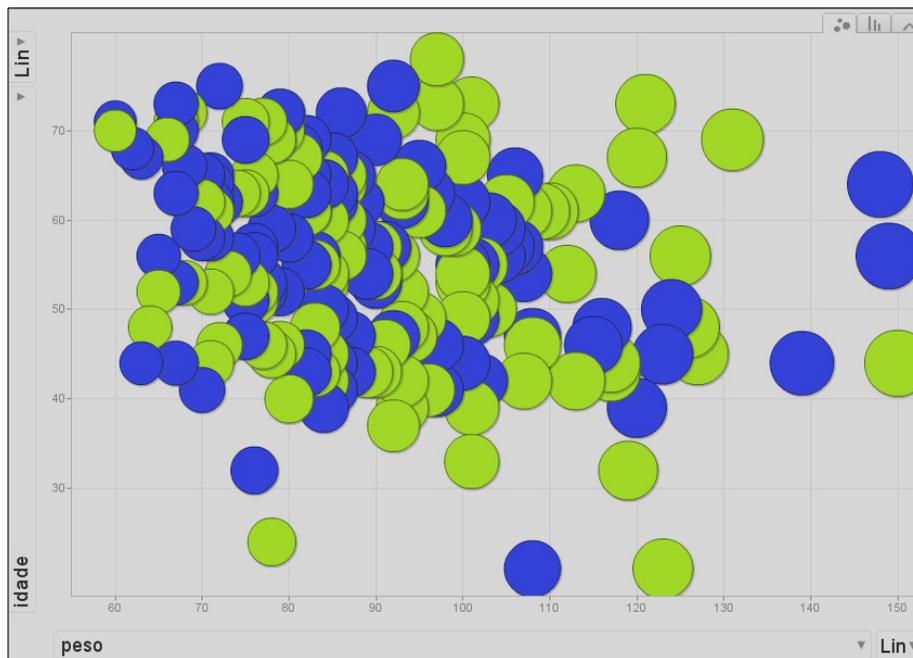


Figura 22: Gráfico interativo do peso – início do estudo.

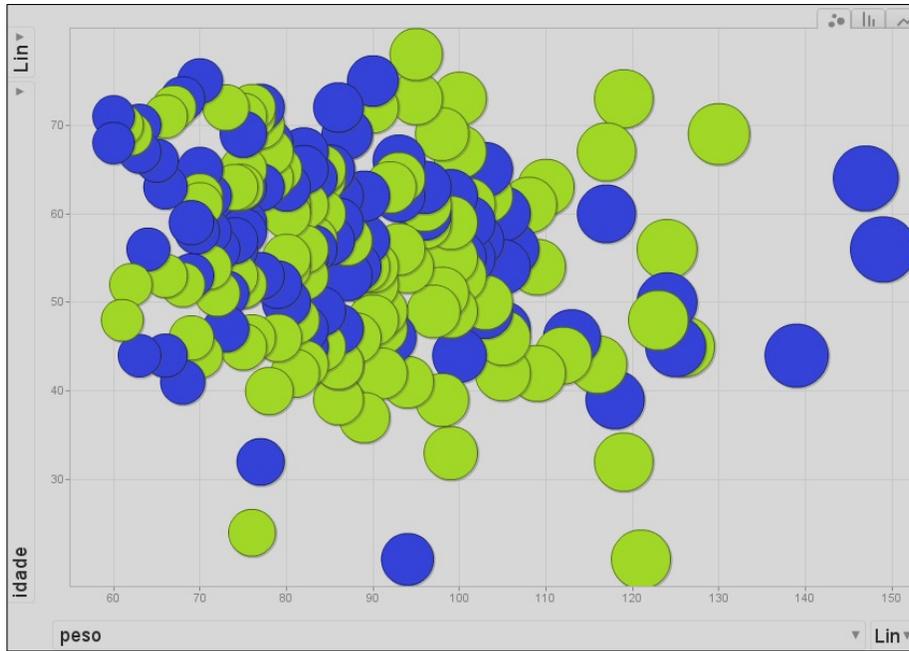


Figura 23: Gráfico interativo do peso – primeira consulta.

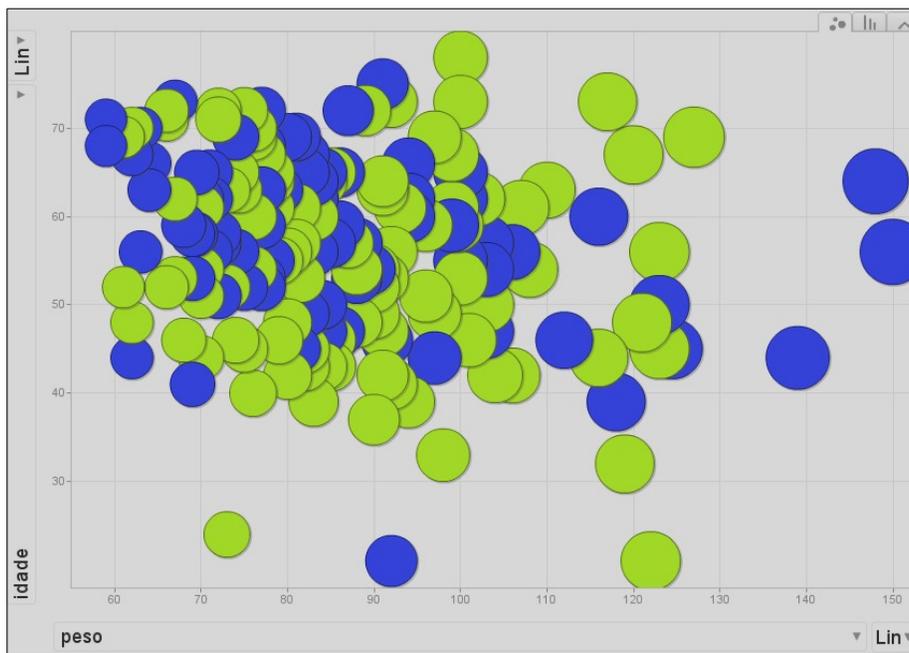


Figura 24: Gráfico interativo do peso – segunda consulta.

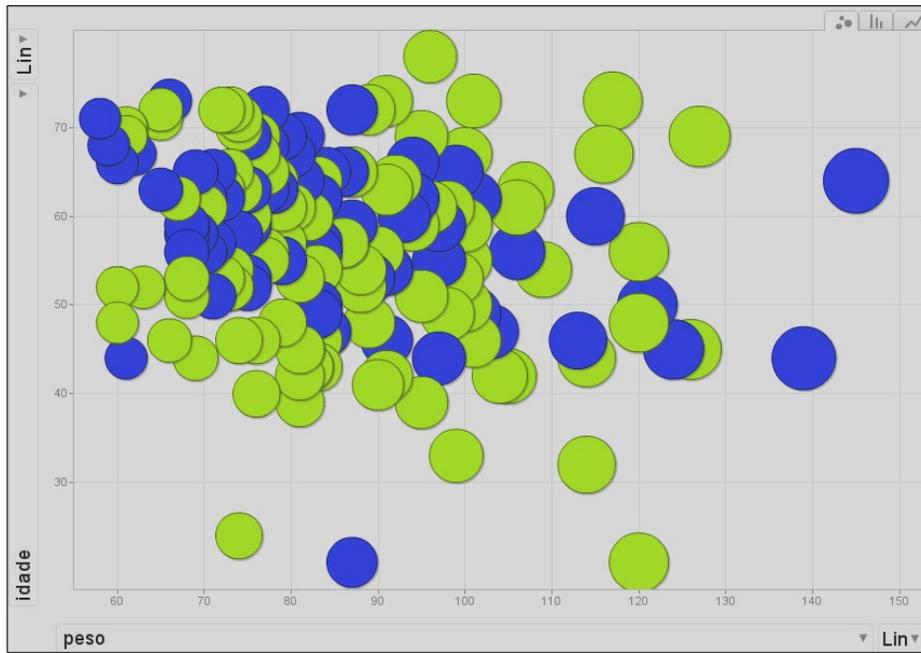


Figura 25: Gráfico interativo do peso – terceira consulta.

É interessante notar que parece não existir um padrão de comportamento entre os indivíduos, ou seja, cada paciente tem seu próprio comportamento no que diz respeito à evolução do peso. Algumas pessoas perdem e ganham peso ao longo do período de estudo, outras só perdem, outras só ganham. Além disso, cada paciente tem uma velocidade de perda ou de ganho bastante particular. Esse comportamento não padronizado sugere que cada participante deste ensaio clínico apresenta uma evolução de peso particular e que deve ser influenciada por suas próprias condições, o que é um dos fatores motivadores da utilização do modelo linear misto.

Tendo em vista que alguns estudos além de trabalharem com o peso, trabalham também com a diferença de peso (proporcional ou não) ou de IMC (Berkowitz *et al.*, 2006; Chanoine *et al.*, 2005; Sheehan *et al.*, 2003; Berkey *et al.*, 2004; Alley *et al.*, 2010), incluiremos na análise deste trabalho a variável proporção de peso perdido. Abaixo mostraremos como ela foi criada:

$$\text{prop. de peso perdido}[c] = \frac{\text{peso}[c] - \text{peso inicial}}{\text{peso inicial}}$$

$c = 1, 2, 3$
 em que c é o número da consulta.

(7)

É importante destacar que esta variável é sempre calculada em relação ao peso inicial. Portanto, se o seu valor for negativo significa que houve perda de peso em relação ao peso inicial e se for positivo houve ganho de peso em relação ao peso inicial. A variável proporção de peso perdido foi criada após a retirada de alguns valores considerados “outliers” conforme comentado na página 43. Além disso, devido a grande quantidade de valores iguais ou próximos a zero - que poderia dificultar o processo de modelagem -, adicionamos ao peso inicial de cada paciente 0,1kg com o intuito de diminuir os excessos de zero.

Para visualizarmos de que maneira essa variável se comporta nos dois grupos de interesse, mostraremos abaixo alguns gráficos.

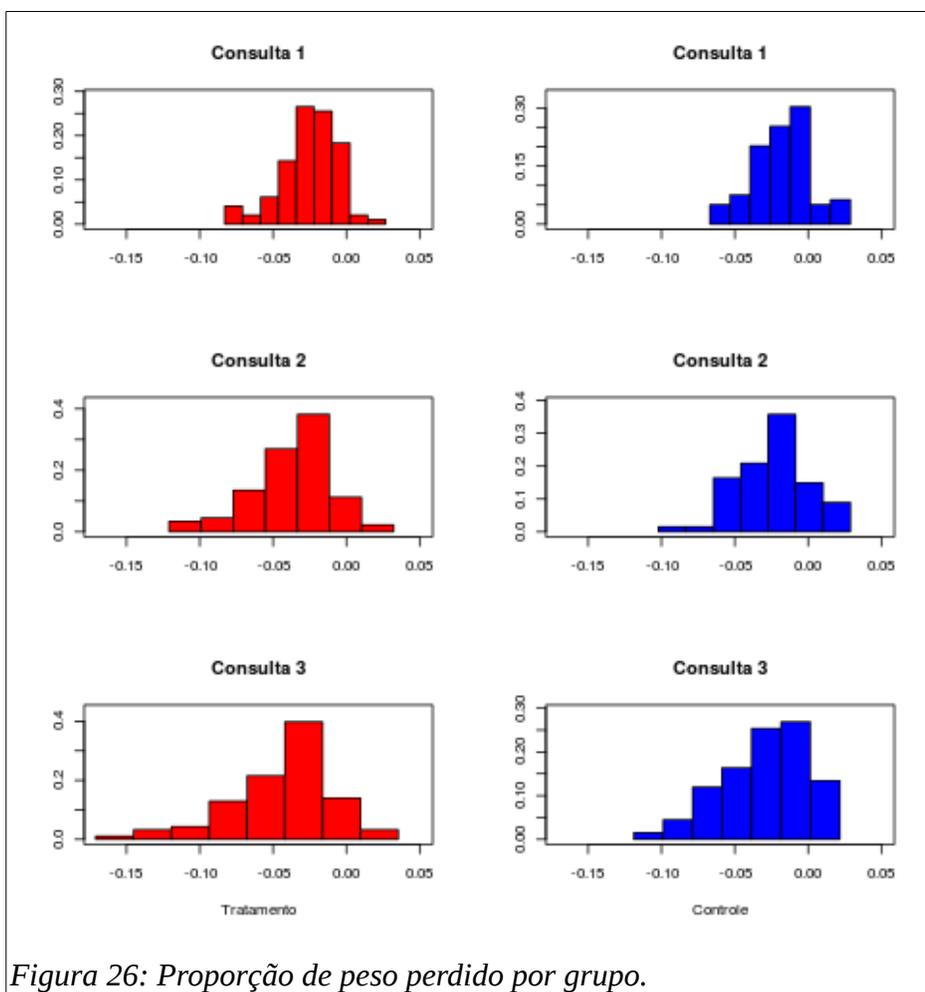


Figura 26: Proporção de peso perdido por grupo.

Na Figura 26 acima estão representadas as frequências relativas da variável proporção de peso perdido em cada grupo. Observa-se que a distribuição dessa proporção é bastante parecida em

todas as consultas e em cada grupo, e que mais da metade das pessoas perdem em média até 5% do seu peso inicial .

Uma outra maneira de se observar a distribuição desta variável é utilizar o gráfico de box-plot conforme a figura 27. Nele notamos a presença de alguns “outliers” e que a proporção de peso diminui ao longo das consultas nos dois grupos, ou seja, os pacientes de ambos os grupos perdem peso ao longo das consultas.

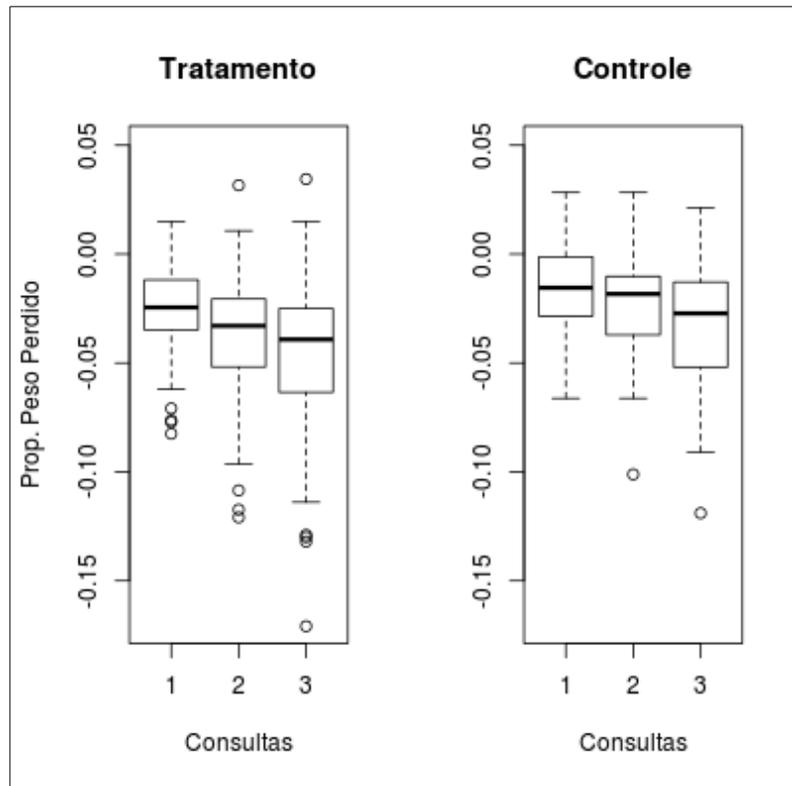


Figura 27: Box-plot da proporção de peso perdido por grupo.

Pela Figura 28 fica mais claro de se observar o comportamento de cada indivíduo dentro de cada grupo em relação a proporção de peso perdido. A trajetória é similar em ambos os grupos, porém, parece que uma quantidade maior de pacientes do grupo tratamento tem uma maior perda de peso ao compararmos com o grupo controle. Além disso, observa-se que a trajetória desta variável é praticamente linear nos dois grupos.

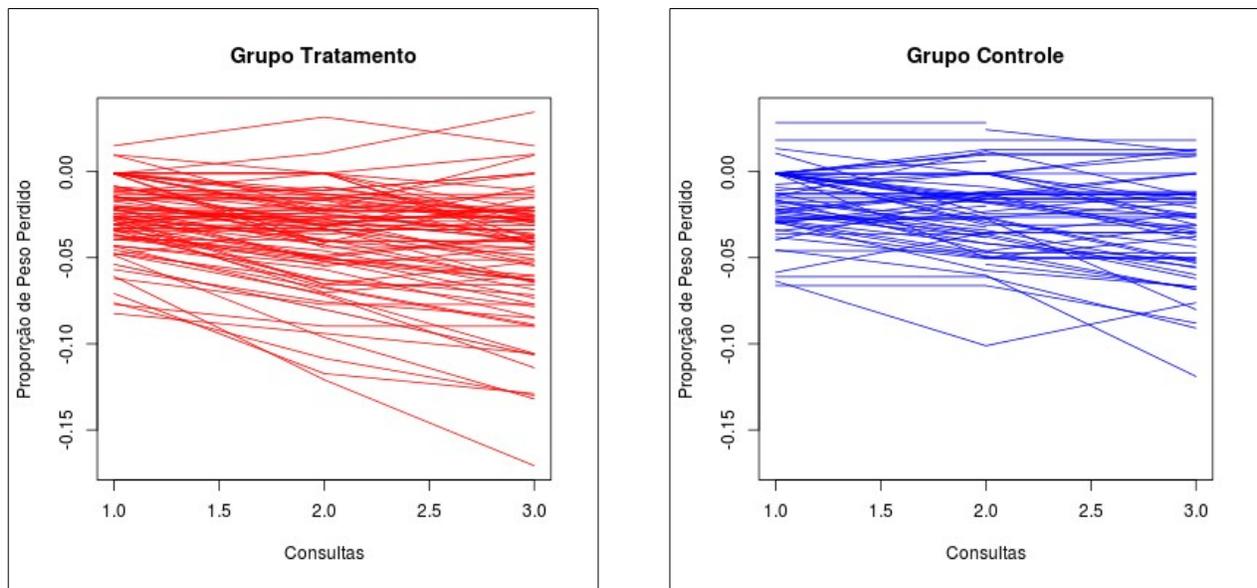


Figura 28: Trajetórias individuais da proporção de peso perdido por grupo.

Após visualizarmos e comentarmos as características das variáveis de interesse deste trabalho, apresentaremos na seção seguinte a parte de modelagem dos dados.

6.2 Modelagem

Esta seção apresenta a aplicação de modelos lineares mistos no banco de dados do ensaio clínico realizado no Hospital Universitário da UFRJ. Vários modelos de efeitos aleatórios foram sugeridos e aplicados a estes dados. As variáveis utilizadas foram: grupo (controle e tratamento), sexo (masculino e feminino), idade, peso inicial e número da consulta. Vale destacar que a idade foi transformada em idade centrada na média e o peso inicial em peso inicial centrado na média. A idade varia entre 21 e 78 anos e o peso inicial entre 60 e 150 kg, porém o intercepto representa o peso do paciente (eixo y), por exemplo, na idade 0 ou no peso inicial 0 kg (eixo x). A extrapolação para o zero resultará em uma alta correlação entre as estimativas das inclinações e suas correspondentes estimativas do intercepto. Para remover essa correlação é necessário centralizar os dados (Endres & Tofghi, 2007; Pinheiro & Bates, 2000).

Nesta parte da modelagem, estamos supondo normalidade para os resíduos do modelo ainda que isto deva ser melhor verificado. Além disso, os modelos que serão apresentados a seguir não tratarão formalmente do fato de existir uma dependência entre as observações. Esperamos que, ao usarmos certas variáveis preditoras que possuam alguma correlação, elas deem conta de qualquer autocorrelação existente de tal forma que não sobre autocorrelação residual.

Consideraremos um efeito linear do início do estudo para a primeira consulta, da primeira pra segunda, e da segunda pra terceira. Apesar da teoria da perda de peso afirmar que esta perda é uma exponencial negativa, no período inicial de perda ela é praticamente linear. Como a duração do estudo utilizado neste trabalho é de apenas três meses e conforme observamos nos gráficos na análise exploratória, é razoável supormos linearidade.

As medidas laboratoriais não entraram no processo de modelagem devido a grande quantidade de dados faltantes.

6.2.1 Definição do Modelo

O interesse deste trabalho é ajustar modelos de efeitos mistos considerando as variáveis peso e proporção de peso perdido como respostas, incorporando nos modelos a variabilidade existente de cada indivíduo. Tomando como referência o modelo com efeito aleatório no intercepto, optou-se por ajustar modelos diferentes em relação à variável resposta, ao número de variáveis incluídas no modelo e na especificação dos efeitos aleatórios.

6.2.2 Ajuste e Escolha do Modelo

A parte da modelagem foi dividida em três blocos de acordo com a variável resposta: o primeiro em que se utiliza o peso; o segundo em que se usa a proporção de peso perdido; o terceiro que é igual ao primeiro com a inclusão da interação do grupo com o tempo.

Os modelos foram ajustados por meio do método de máxima verossimilhança restrita (REML) (Verbeke & Molenberghs, 2000) utilizando a função *lmer* do pacote *lme4* (Bates *et al.*, 2011) do R. A seguir, apresentaremos cada um dos modelos que iremos ajustar.

6.2.2.1 Variável Resposta Peso

Este primeiro bloco foi dividido em três partes. Para a primeira parte, construímos quatro modelos contendo as variáveis grupo, idade e sexo e o intercepto aleatório. O efeito aleatório no intercepto representa a heterogeneidade natural entre os indivíduos decorrente de fatores (genéticos ou ambientais) não medidos (Diggle *et al.*, 1994; Molenberghs & Verbeke, 2001). O primeiro modelo apresenta apenas o grupo e o efeito aleatório no intercepto. Ao segundo modelo adicionamos as variáveis idade e sexo. No terceiro testou-se a interação entre sexo e grupo e no quarto a interação entre a idade e o grupo. Esta interação significa que o efeito do sexo na resposta média depende do grupo e vice-versa, assim como o efeito da idade na resposta média depende do grupo e vice-versa.

Na segunda parte, levantamos a questão se o peso inicial deveria entrar como uma covariável.

Por corresponder ao peso do paciente antes do início do tratamento, o peso inicial pode influenciar o quanto o paciente perde ou ganha bem como a taxa desta perda ou deste ganho de peso. As etapas seguidas foram as mesmas da primeira parte.

Para a última parte, utilizamos como modelo de base o sexto modelo - o segundo da parte 2- por ser um modelo que contém variáveis importantes e por apresentar um dos menores valores de AIC. Nesta parte, incluímos primeiramente um efeito aleatório na idade e depois na variável sexo. Adicionar um efeito aleatório nestas variáveis é justificado pela hipótese de se ter uma variabilidade do desfecho dependendo da idade ou do sexo do paciente.

Abaixo resumiremos cada um dos modelos deste bloco:

Parte I

1. $\text{peso} = \text{grupo} + (1|\text{individuo})$
2. $\text{peso} = \text{grupo} + \text{idade} + \text{sexo} + (1|\text{individuo})$
3. $\text{peso} = \text{idade} + \text{sexo} * \text{grupo} + (1|\text{individuo})$
4. $\text{peso} = \text{sexo} + \text{idade} * \text{grupo} + (1|\text{individuo})$

(8)

Parte II

5. $\text{peso} = \text{pesoinicial} + \text{grupo} + (1|\text{individuo})$
6. $\text{peso} = \text{pesoinicial} + \text{grupo} + \text{idade} + \text{sexo} + (1|\text{individuo})$
7. $\text{peso} = \text{pesoinicial} + \text{idade} + \text{sexo} * \text{grupo} + (1|\text{individuo})$
8. $\text{peso} = \text{pesoinicial} + \text{sexo} + \text{idadecentrada} * \text{grupo} + (1|\text{individuo})$

(9)

Parte III

9. $\text{peso} = \text{pesoinicial} + \text{grupo} + \text{idade} + \text{sexo} + (\text{idade}|\text{individuo})$
10. $\text{peso} = \text{pesoinicial} + \text{grupo} + \text{idade} + \text{sexo} + (\text{sexo}|\text{individuo})$

(10)

6.2.2.2 Variável Resposta Proporção de Peso Perdido

Para o ajuste dos modelos cuja variável resposta é a proporção de peso perdido, a estratégia de modelagem adotada foi: primeiramente utilizou-se as variáveis grupo, idade e sexo e intercepto aleatório; em seguida adicionou-se interações; e, por último, acrescentou-se efeito aleatório na idade e no sexo.

Abaixo resumiremos cada uma dessas partes e os respectivos modelos:

Parte I

1. $prop. de peso perdido = peso_{inicial} + grupo + (1|individuo)$ (11)
2. $prop. de peso perdido = peso_{inicial} + grupo + idade + sexo + (1|individuo)$

Parte II

3. $prop. de peso perdido = peso_{inicial} + idade + sexo * grupo + (1|individuo)$ (12)
4. $prop. de peso perdido = peso_{inicial} + sexo + idade * grupo + (1|individuo)$

Parte III

5. $prop. de peso perdido = peso_{inicial} + grupo + idade + sexo + (idade|individuo)$ (13)
6. $prop. de peso perdido = peso_{inicial} + grupo + idade + sexo + (sexo|individuo)$

6.2.2.3 Variável Resposta Peso e Interação Tratamento-Tempo

Ao ajustarmos os modelos acima uma pergunta que levantamos foi: qual seria o efeito do tratamento ao longo do tempo? Para respondê-la, ajustamos quatro modelos que incluem a interação entre o tratamento e o tempo (consulta) e os organizamos em duas partes. A primeira contém dois modelos com esta interação e efeito aleatório no intercepto. A segunda contém dois modelos que possuem além do efeito aleatório no intercepto, um efeito aleatório na idade (modelo 3) e no sexo (modelo 4).

Parte I

$$\begin{aligned} 1. \text{ peso} &= \text{pesoinicial} + \text{grupo:cons} + (1|\text{individuo}) \\ 2. \text{ peso} &= \text{pesoinicial} + \text{sexo} + \text{idade} + \text{grupo:cons} + (1|\text{individuo}) \end{aligned} \tag{14}$$

Parte II

$$\begin{aligned} 3. \text{ peso} &= \text{pesoinicial} + \text{sexo} + \text{idade} + \text{grupo:cons} + (\text{idade}|\text{individuo}) \\ 4. \text{ peso} &= \text{pesoinicial} + \text{sexo} + \text{idade} + \text{grupo:cons} + (\text{sexo}|\text{individuo}) \end{aligned} \tag{15}$$

Antes de analisar cada um dos modelos descritos acima, mostraremos como é a saída (“output”) de um modelo ajustado, utilizando a metodologia de modelos de efeitos mistos, no R. Explicaremos como esta saída está estruturada e como se interpreta os valores que nela se encontram.

6.2.2.4 Saída do R

Linear mixed model fit by REML

Formula: peso ~ peso inicial + grupo + idade + sexo + (1|indivíduo)

Data: bancolimpo

AIC	BIC	logLik	deviance	REMLdev
2086	2116	-1036	2056	2072

Random effects:

Groups	Name	Variance	Std.Dev.
indivíduo	(Intercept)	3.6301	1.9053
	Residual	1.9827	1.4081

Number of obs: 493, groups: indivíduo, 187

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	86.908179	0.420445	206.71
peso inicial	1.003247	0.009445	106.22
tratamento	-1.276101	0.313936	-4.06
idade	0.022522	0.015242	1.48
sexo2	1.008663	0.419419	2.40

A saída do modelo está dividida em 4 seções principais: a primeira que descreve o modelo que foi ajustado; a segunda que apresenta algumas estatísticas que caracterizam o ajuste do modelo; a terceira que contém um resumo das propriedades dos efeitos aleatórios; e a quarta que fornece um resumo das estimativas dos parâmetros dos efeitos fixos.

No modelo que ajustamos existem duas fontes de variabilidade: a variabilidade de grupo para grupo, e a variabilidade residual ou a variabilidade por observação também chamada de variabilidade dentro do grupo. O nome “residual” é usado em modelagem estatística para designar a parte da variabilidade que não é explicada ou modelada com os outros termos. É a variação nos dados observados e que “sobraram” mesmo depois do ajuste do modelo (Bates, 2010).

Para este exemplo, pode-se observar na primeira seção da saída do modelo ajustado, que a variável resposta é o peso, as covariáveis são o peso inicial, o grupo, a idade e o sexo, e o efeito

aleatório é no indivíduo. Na segunda seção, são apresentadas as estatísticas que caracterizam o ajuste do modelo e que são utilizadas para comparar os diferentes modelos: o critério de informação de Akaike (AIC), o critério Bayesiano de informação (BIC), o logaritmo da verossimilhança ($\log\text{Lik}$), a deviance (menos duas vezes o $\log\text{Lik}$) e a deviance do REML (menos duas vezes o logaritmo da função de verossimilhança restrita).

Na terceira parte do “output” do R, relativa aos efeitos aleatórios, a linha “indivíduo”, nome da variável que indica cada participante da pesquisa, mostra que o efeito aleatório adicionado é modelado como variável aleatória cujo valor da variância estimada é de 3.6301, e do desvio-padrão correspondente é 1.9053. A linha rotulada de “Residual” fornece a estimativa da variância dos resíduos e seu correspondente desvio-padrão, cujos respectivos valores são 1.9827 e 1.4081. A última linha da tabela de efeitos aleatórios indica o número de observações para o qual o modelo foi ajustado e o número de níveis dos “fatores de agrupamento”. Neste caso, temos um único termo de efeito aleatório, e o fator de agrupamento para esse termo é “indivíduo”. Há um total de 187 efeitos aleatórios, um para cada nível de “indivíduo”.

A parte final da saída do R apresenta as estimativas e erros-padrão dos parâmetros de efeitos fixos e os respectivos valores da estatística de teste (t de Student). A estimativa dos efeitos fixos é interpretada da mesma maneira que na regressão linear tradicional. O intercepto é interpretado como a média da variável desfecho, neste exemplo o peso, quando todas as variáveis preditoras têm valores iguais a zero. Por exemplo, aumentar uma unidade no preditor idade – um ano a mais de idade – corresponde a um acréscimo de 0.022522 no peso, ou seja, 22 gramas. Enquanto que aumentar uma unidade no peso inicial corresponde a um aumento de 1.003247 no peso das medidas subsequentes. Além disso, as variáveis categóricas grupo (tratamento) e sexo têm coeficientes iguais a -1.276101 e 1.008663, respectivamente, o que significa que, em média, o peso do grupo tratamento é 1,28 vezes menor do que o peso no grupo controle e que, em média, o peso do sexo feminino (sexo2) é 1,01 vezes maior do que o peso do sexo masculino.

O grupo controle e o sexo masculino (sexo1) foram automaticamente codificados como categoria de referência pela função *lmer* porque, como a maioria das funções do R, a categoria com o menor valor numérico (ou primeira letra do alfabeto) é codificado como categoria de referência.

É importante destacar que para um grande número de graus de liberdade (>100) a distribuição t se aproxima da distribuição normal e, portanto, uma forma simples de avaliar se as variáveis são

estatisticamente significativas ao nível de 5%, é verificar o valor absoluto da estatística t. Se este valor exceder 2, a variável pode ser considerada estatisticamente significativa (Baayen, 2008).

6.2.3 Resultados dos Modelos

O objetivo desta seção é analisar cada um dos modelos, interpretar os valores encontrados e levantar hipóteses que justifiquem possíveis mudanças nas estimativas de um ou mais coeficientes.

Os resultados dos modelos foram organizados de acordo com a variável resposta. Nas tabelas encontram-se as estimativas da parte fixa e seus respectivos erros-padrão, as variâncias e desvios-padrão da parte aleatória, e o valor do AIC. Para dar maior destaque às variáveis que foram significativas, isto é, cujo valor da estatística t foi maior, em módulo, do que 2, as suas respectivas estimativas estão em negrito. Os valores da estatística t não serão mostrados nas tabelas a seguir porém, para calculá-los, basta dividir o valor da estimativa pelo seu respectivo erro padrão. Para toda a análise utilizou-se $\alpha=5\%$.

- **Modelos cuja variável resposta é o peso**

Os resultados desta modelagem encontram-se nas três tabelas a seguir.

Tabela 5: Estimativas das partes fixa e aleatória dos modelos com efeito aleatório no intercepto.

	mod1		mod2		mod5		mod6	
Parte fixa	estim.	e.p.	estim.	e.p.	estim.	e.p.	estim.	e.p.
(Intercepto)	85,450	1,834	92,224	3,255	87,793	0,234	86,908	0,420
tratamento	2,206	2,508	1,271	2,453	-1,342	0,316	-1,276	0,314
idade	-	-	-0,315	0,114	-	-	0,023	0,015
sexo feminino	-	-	-7,388	3,243	-	-	1,009	0,419
peso inicial	-	-	-	-	0,996	0,009	1,003	0,009
Parte aleatória	variância	d.p.	variância	d.p.	variância	d.p.	variância	d.p.
individuo (Intercpt.)	295,051	17,177	278,063	16,675	3,766	1,941	3,630	1,905
residuo	2,056	1,434	2,056	1,434	1,983	1,408	1,983	1,408
AIC	2890		2879		2084		2086	

No primeiro modelo, observa-se que a variável tratamento não está significativamente associada com o peso. Do ponto de vista do estudo original, praticamente bastaria esse modelo, uma vez que se considera que sendo um desenho com aleatorização dos indivíduos as variáveis de

confusão estariam controladas. Ao introduzir as variáveis idade e sexo no modelo, o tratamento continua não significativo, porém, idade e sexo são significativas, o peso diminuindo com a idade e com o sexo feminino.

A inclusão do peso inicial (modelo 6) altera completamente os resultados. Em primeiro lugar é importante notar que nos dois primeiros modelos temos uma enorme variância do efeito aleatório e que, ao inserirmos o peso inicial, ela se torna bastante pequena. Isso quer dizer que, se nos modelos iniciais boa parte da variabilidade era explicada pelos indivíduos, ao introduzir o peso inicial, ele “explica” as diferenças entre os indivíduos. Além disso, verificou-se que houve modificação no sinal e no valor da estimativa para as variáveis grupo e sexo quando comparadas ao modelo 2. E o tratamento passa a estar negativa e significativamente associado ao peso! Essa mudança somente pode ser interpretada a partir da compreensão do modelo de efeitos aleatórios. Nos modelos 2 e 3 a ausência de uma variável tão importante como o peso inicial, quando se considera a resposta os pesos subsequentes, faz com que praticamente toda a variabilidade seja atribuída às características do indivíduo, modeladas por meio de efeito aleatório. Ou seja, qualquer explicação das variáveis está condicionada a um fator não mensurado. Quando se inclui esse fator, o comportamento das demais variáveis muda.

Conforme podemos observar na Tabela 6, a possibilidade de existência de interações entre o sexo e o grupo e entre a idade e o grupo não foi confirmada. Nenhuma das interações inseridas nos modelos foi significativa.

Tabela 6: Estimativas das partes fixa e aleatória dos modelos com interações e efeito aleatório no intercepto.

	mod3		mod4		mod7		mod8	
Parte fixa	estim.	e.p.	estim.	e.p.	estim.	e.p.	estim.	e.p.
(Intercepto)	94,152	4,470	92,411	3,263	87,590	0,583	86,934	0,426
tratamento	-2,155	5,963	1,090	2,461	-2,437	0,758	-1,292	0,317
idade	-0,314	0,114	-0,451	0,185	0,022	0,015	0,014	0,027
sexo feminino	-9,684	4,879	-7,321	3,245	0,200	0,636	1,000	0,421
sexo fem:tratam	4,120	6,533	-	-	1,393	0,828	-	-
idade:tratamento	-	-	0,219	0,235	-	-	0,013	0,032
peso inicial	-	-	-	-	1,003	0,009	1,003	0,010
Parte aleatória	variância	d.p.	variância	d.p.	variância	d.p.	variância	d.p.
indivíduo (Intercpt.)	278,975	16,703	278,251	16,681	3,585	1,893	3,651	1,911
resíduo	2,056	1,434	2,056	1,434	1,983	1,408	1,983	1,408
AIC	2875		2882		2084		2093	

Na tabela abaixo, incluímos efeitos aleatórios na estimativa do efeito da idade (modelo 9) e do sexo (modelo 10). A inclusão deste tipo de efeito na variável idade, por exemplo, representa a heterogeneidade observada, entre e dentro dos indivíduos, com a modificação da idade.

Tabela 7: Estimativas das partes fixa e aleatória dos modelos com efeitos aleatórios no intercepto e na inclinação.

	mod9		mod10	
Parte fixa	estim.	e.p.	estim.	e.p.
(Intercepto)	86,909	0,420	86,854	0,592
tratamento	-1,276	0,314	-1,123	0,287
idade	0,023	0,015	0,018	0,014
sexo feminino	1,007	0,419	0,967	0,589
peso inicial	1,003	0,009	1,000	0,009
Parte aleatória	variância	d.p.	variância	d.p.
indivíduo (Intercpt.)	3,631	1,905	9,206	3,034
idade	0,000	0,000	-	-
sexo feminino	-	-	3,647	1,910
resíduo	1,983	1,408	1,988	1,410
AIC	2090		2071	

No modelo 9 da tabela acima (Tabela 7), observa-se que a variância do efeito aleatório na idade é praticamente igual a zero. Isso não quer dizer que não exista nenhuma variação dentro dos grupos. A estimativa da variância ser zero simplesmente indica que o nível de variabilidade “dentro do grupo” não é suficiente para justificar a incorporação de efeitos aleatórios na idade neste modelo (Bates, 2010). No modelo 10, as estimativas da parte fixa pouco se alterou porém, observa-se que a variância do efeito do aleatório no intercepto aumentou quando comparado ao modelo 9.

- **Modelos cuja variável resposta é a proporção de peso perdido**

A elaboração deste modelo teve como principal objetivo verificar quais são os fatores que influenciam de modo significativo a perda de peso. Utilizando a proporção de peso perdido, calculada conforme a equação 6, como variável resposta, os coeficientes estimados representam a mudança de peso esperada a cada consulta, controlada pelas covariáveis peso inicial, além de grupo, idade e sexo. Além disso, possivelmente essa resposta é a de maior importância clínica.

Tabela 8: Estimativas das partes fixa e aleatória dos modelos com efeito aleatório no intercepto.

	mod1		mod2	
Parte fixa	estim.	e.p.	estim.	e.p.
(Intercepto)	-0,021	0,003	-0,030	0,005
tratamento	-0,015	0,004	-0,014	0,004
idade	-	-	0,000	0,000
sexo feminino	-	-	0,011	0,005
peso inicial	0,000	0,000	0,000	0,000
Parte aleatória	variância	d.p.	variância	d.p.
individuo (Intercpt.)	0,000	0,022	0,000	0,022
residuo	0,000	0,016	0,000	0,016
AIC	-2305		-2284	

Ajustando um modelo com apenas a variável grupo (modelo 1), observa-se que ser do grupo tratamento é associado significativamente com a proporção de peso perdido. O sinal da associação é negativa o que indica que o tratamento diminui a diferença de peso em relação ao peso inicial em cada consulta. Ou seja, controlado pelo peso inicial o tratamento diminui o peso do paciente.

Na Tabela 9, são apresentados os modelos com intercepto aleatório e interações entre sexo e grupo (modelo 3), e entre idade e grupo (modelo 4). Observa-se que tanto no modelo 3 quanto no modelo 4 as interações não foram significativas.

Tabela 9: Estimativas das partes fixa e aleatória dos modelos com interações e efeito aleatório no intercepto.

	mod3		mod4	
Parte fixa	estim.	e.p.	estim.	e.p.
(Intercepto)	-0,024	0,007	-0,030	0,005
tratamento	-0,025	0,009	-0,014	0,004
idade	0,000	0,000	0,000	0,000
sexo feminino	0,003	0,007	0,010	0,005
sexo fem:tratam	0,013	0,009	-	-
idade:tratamento	-	-	0,000	0,000
peso inicial	0,000	0,000	0,000	0,000
Parte aleatória	variância	d.p.	variância	d.p.
individuo (Intercpt.)	0,000	0,022	0,000	0,022
residuo	0,000	0,016	0,000	0,016
AIC	-2276		-2268	

Na Tabela 10 observa-se que, no modelo 5 todas as variáveis foram significativas exceto a idade enquanto, no modelo 6, somente a variável tratamento foi significativamente associada com a proporção de peso perdido.

Tabela 10: Estimativas das partes fixa e aleatória dos modelos com efeitos aleatórios no intercepto e na inclinação.

	mod5		mod6	
Parte fixa	estim.	e.p.	estim.	e.p.
(Intercepto)	-0,030	0,005	-0,031	0,006
tratamento	-0,014	0,004	-0,013	0,003
idade	0,000	0,000	0,000	0,000
sexo feminino	0,011	0,005	0,010	0,006
peso inicial	0,000	0,000	0,000	0,000
Parte aleatória	variancia	d.p.	variancia	d.p.
individuo (Intercpt.)	0,000	0,022	0,001	0,033
idade	0,000	0,000	-	-
sexo feminino	-	-	0,000	0,020
residuo	0,000	0,016	0,000	0,016
AIC	-2281		-2293	

- **Modelos cuja variável resposta é o peso e interação tratamento-tempo**

Para uma última análise, investigamos a interação do tratamento com o tempo - que neste caso é o número da consulta. Esses resultados estão nas tabelas abaixo.

Tabela 11: Estimativas das partes fixa e aleatória dos modelos com interações e efeito aleatório no intercepto.

	mod1		mod2	
Parte fixa	estim.	e.p.	estim.	e.p.
(Intercepto)	88.587	0.200	87.736	0.400
peso inicial	0.995	0.009	1.003	0.009
sexo feminino	-	-	1.010	0.418
idade	-	-	0.026	0.015
controle:consulta	-0.505	0.088	-0.512	0.088
tratamento:consulta	-1.006	0.077	-1.001	0.077
Parte aleatória	variância	d.p.	variância	d.p.
indivíduo (Intercpt.)	4.002	2.000	3.852	1.963
resíduo	1.309	1.144	1.308	1.144
AIC	1962		1964	

Neste modelo, utilizamos o efeito do tempo considerando que o paciente é do grupo controle e o efeito do tempo considerando que ele é do grupo tratamento. Observa-se que os dois modelos são bem parecidos. As estimativas, a variância da parte aleatória e o AIC têm valores próximos. O termo de interação grupo e consulta foi estatisticamente significativo, isto é, o efeito do uso (ou não) do medicamento na perda de peso varia com o tempo.

Olhando as estimativas destas variáveis, observa-se que as duas têm sinais negativos e estão estatisticamente associadas com o peso. Tanto no modelo 1 quanto no 2, os dois grupos perdem peso a cada consulta.

Ao modelo 2, incluiu-se efeitos aleatórios na idade (modelo 3) e no sexo (modelo 4). A Tabela 12, abaixo, resume as informações destes dois modelos.

Tabela 12: Estimativas das partes fixa e aleatória dos modelos com interações e efeitos aleatórios no intercepto e na inclinação.

	mod3		mod4	
Parte fixa	estim.	e.p.	estim.	e.p.
(Intercepto)	87,736	0,400	87,769	0,586
peso inicial	1,003	0,009	1,000	0,009
sexo feminino	1,009	0,418	0,973	0,592
idade	0,026	0,015	0,021	0,014
controle:consulta	-0,512	0,088	-0,532	0,086
tratamento:consulta	-1,001	0,077	-0,989	0,076
Parte aleatória	variância	d.p.	variância	d.p.
individuo (Intercpt.)	3,852	1,963	9,571	3,094
idade	0,000	0,000	-	-
sexo feminino	-	-	6,140	2,478
residuo	1,308	1,144	1,309	1,144
AIC	1968		1947	

O valor das estimativas das covariáveis em ambos os modelos pouco se alterou. No modelo 3, a variância do efeito aleatório na idade foi praticamente zero.

Porém, a variância do efeito aleatório no intercepto no modelo 4 teve seu valor aumentado três vezes quando comparado ao modelo 3, ou seja, a heterogeneidade entre os pacientes aumentou. Isso pode ser atribuído ao fato de termos incluído um efeito na variável sexo.

No processo de modelagem, como visto acima, a escolha de quais variáveis resposta e covariáveis usar depende do objetivo do pesquisador. Neste trabalho, usou-se como variáveis resposta o peso e a proporção de peso perdido. Utilizar o peso como variável resposta é verificar quais variáveis estão associadas a ele. Ao se usar a proporção de peso perdido, significa modelar a perda de peso dos pacientes. Para entender melhor o resultado do orlistat, utilizar um modelo cuja variável resposta seja a proporção de peso perdido é mais plausível dado que estamos interessados em observar o efeito do tratamento na perda do peso.

A seguir, examinaremos os modelos ajustados neste trabalho analisando e comentando os gráficos de diagnósticos.

6.2.4 Examinando os Modelos Ajustados

Após o ajuste dos modelos, é importante realizar o diagnóstico do modelo para verificar se os pressupostos básicos de distribuição para os modelos de efeitos mistos são satisfeitos (West *et al.*, 2007).

Primeiro, compararemos diferentes representações gráficas para um mesmo tipo de gráfico utilizando como exemplo os gráficos gerados no modelo 6 do primeiro bloco.

6.2.4.1 Exemplos de diferentes representações gráficas

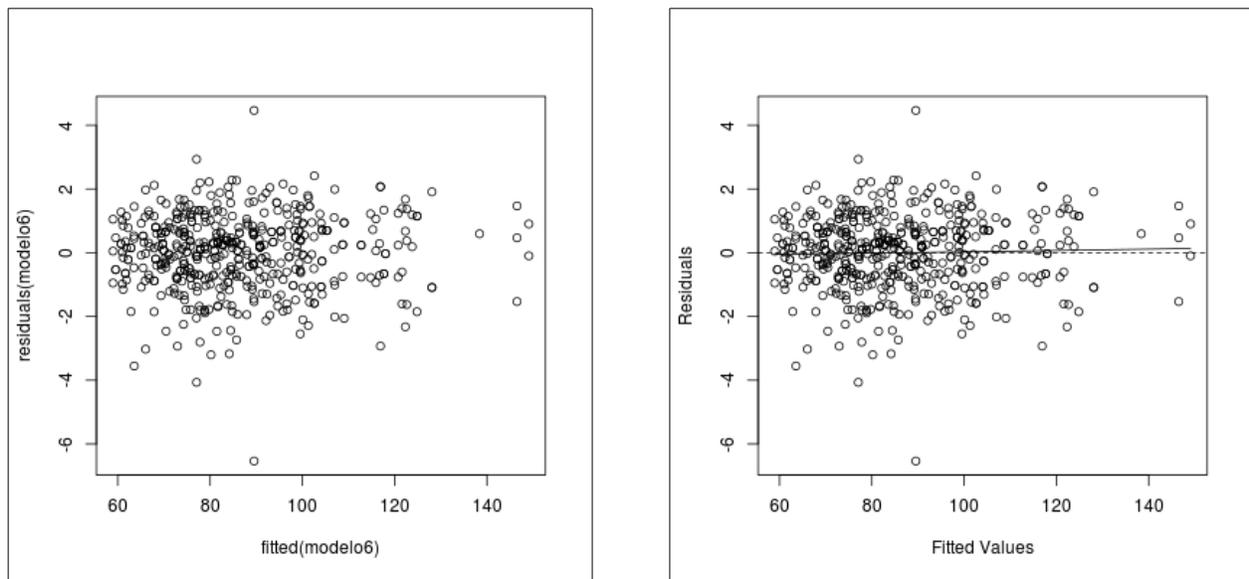


Figura 29: Gráficos dos valores ajustados versus resíduos.

Nestes dois primeiros gráficos, os valores ajustados estão no eixo x e os resíduos no eixo y. Os resíduos representam o quanto “sobrou” após o modelo ser ajustado, isto é, a diferença entre o valor ajustado e o valor observado. A finalidade desses gráficos é observar o comportamento dos resíduos. Este gráfico é útil para verificar se os erros são centrados no zero (isto é, $E[\epsilon]=0$) e se têm variância constante em todos os grupos $Var[\epsilon_{ij}]=\sigma^2$.

A diferença entre os gráficos da Figura 29 é que no segundo temos uma linha tracejada e outra contínua. O objetivo dessas linhas é nos guiar, já que os valores dos resíduos devem estar em torno de zero (por isso a linha tracejada) e não apresentar nenhum padrão (por isso a linha contínua).

Os resíduos, nos gráficos da Figura 29, aparentemente não possuem padrão particular ou indicação de tendência. Se não há padrões e, acima de tudo, se não existe uma tendência, então o modelo está respeitando a suposição em relação à homogeneidade (variância constante) e à normalidade.

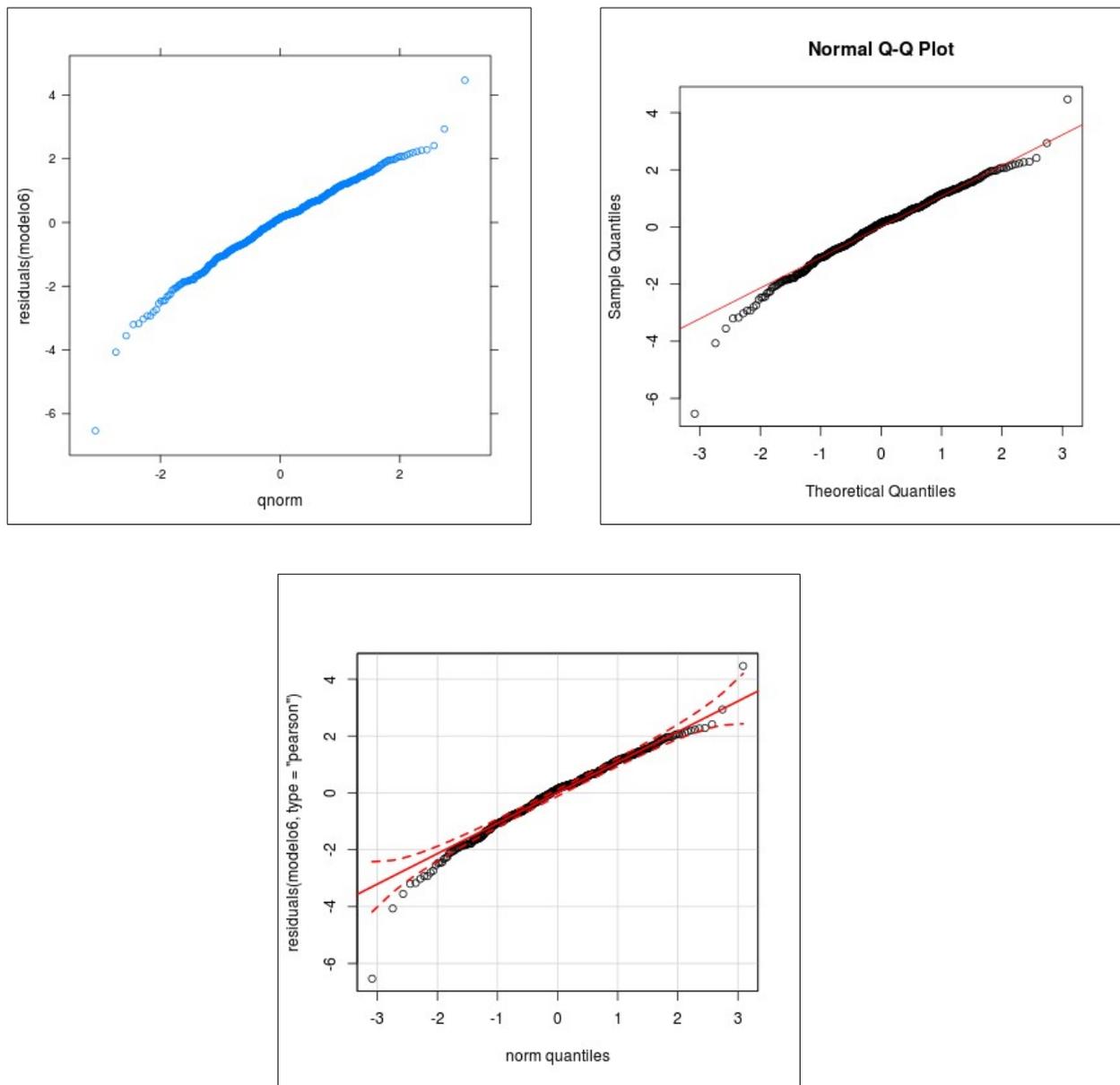


Figura 30: Gráficos de probabilidade normal para os resíduos.

Os gráficos anteriores (Figura 30) são também conhecidos como gráficos de quantis dos resíduos *versus* os quantis da distribuição normal ou Q-Q plot. Eles são utilizados para verificar a suposição de normalidade dos resíduos. Quanto mais próximos os pontos estão de uma reta, mais próximos os resíduos estão da distribuição normal. Os segundos e terceiros gráficos fornecem esta reta sendo que neste último é fornecido também o intervalo de 95% de confiança. Além dos pontos estarem nesta reta, eles devem cair dentro do envelope (Atkinson, 1985).

Na Figura 31, os gráficos apresentados são os Q-Q plots do efeito aleatório. Gráficos como o segundo desta figura, com linhas verticais, são por vezes chamados de “caterpillar plots” devido à sua aparência. Esta linha vertical, que se estende $\pm 1,96$ desvios padrão em cada direção, indica a precisão da distribuição condicional dos efeitos aleatórios. É necessário observar, portanto, se esses intervalos criados se sobrepõem ou não com a linha que passa por zero. Se todos ou grande parte se sobrepuserem, então é provável não ser necessário adicionar este efeito aleatório ao modelo.

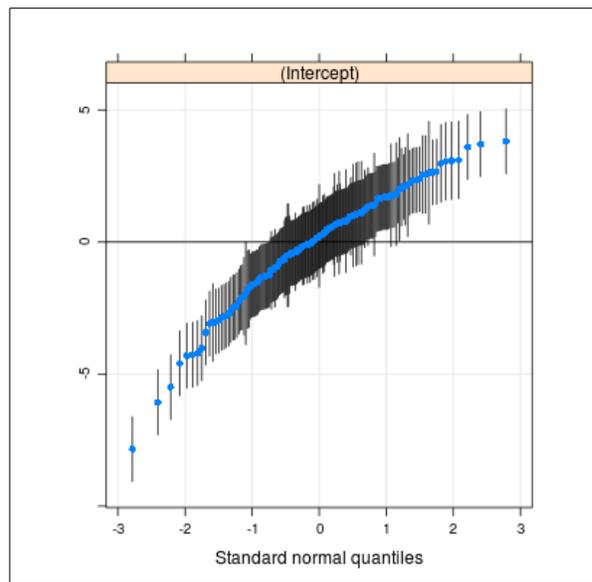
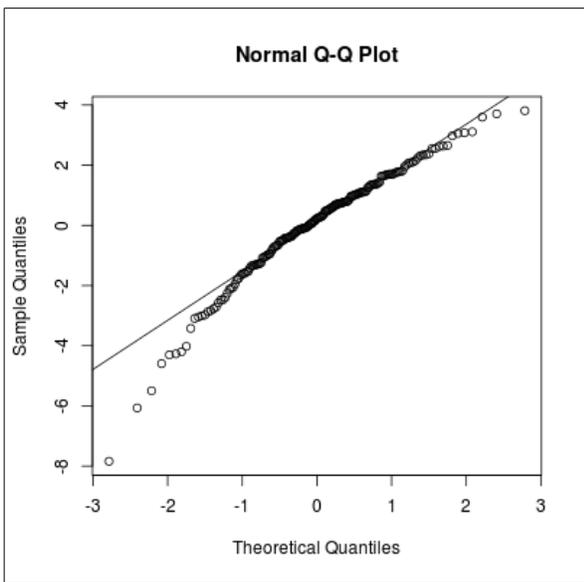
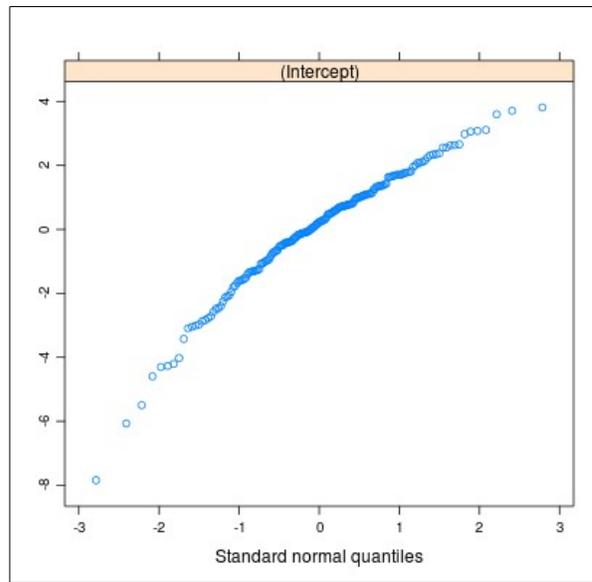
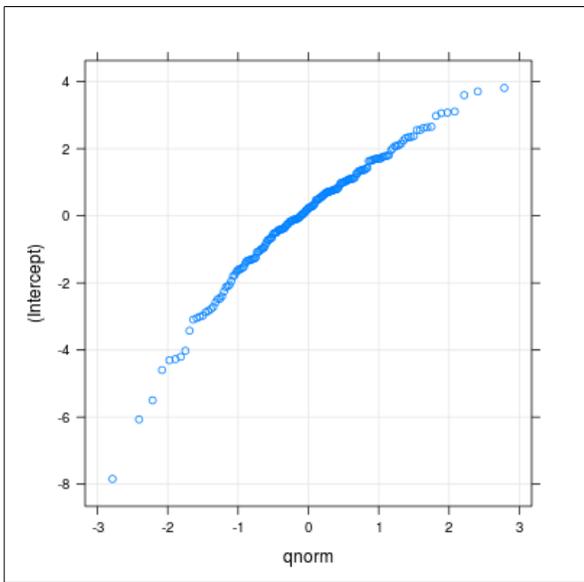


Figura 31: Gráficos de probabilidade normal do efeito aleatório.

6.2.4.2 Examinando e comparando os “melhores” modelos

A seguir, verificaremos os pressupostos de um modelo linear misto em cada um dos “melhores” modelos ajustados de cada bloco. Para o bloco cuja variável resposta é o peso, o modelo que teve menor AIC foi o modelo 10, para o bloco da proporção de peso perdido foi o modelo 1 e para o bloco da resposta peso ao longo do tempo, o modelo 4 foi o que apresentou menor AIC. Os gráficos dos outros modelos podem ser encontrados no Apêndice deste trabalho. A seguir, exibimos a forma de cada um desses modelos:

$$\begin{aligned} 10. \text{ peso} &= \text{pesoinicial} + \text{grupo} + \text{idade} + \text{sexo} + (\text{sexo}|\text{individuo}) \\ 1. \text{ proporção de peso} &= \text{pesoinicial} + \text{grupo} + (1|\text{individuo}) \\ 4. \text{ peso} &= \text{pesoinicial} + \text{sexo} + \text{idade} + \text{grupo} : \text{cons} + (\text{sexo}|\text{individuo}) \end{aligned} \tag{16}$$

Inicialmente falaremos sobre os gráficos relacionados ao resíduo e, em seguida, sobre os gráficos relacionados aos efeitos aleatórios. Diversos gráficos foram gerados para cada modelo, porém, apresentaremos, a seguir, apenas dois gráficos para os resíduos e dois para os efeitos aleatórios. O restante pode ser encontrado no Apêndice.

O primeiro gráfico dos resíduos a ser analisado é o dos valores ajustados *versus* os valores dos resíduos (Figura 32). Nos gráficos (a), (b) e (c), observa-se que os valores se distribuem aleatoriamente em torno de zero. Percebe-se, ainda, a presença de alguns candidatos a “outliers”, cujos resíduos estão afastados mais de 2 desvios padrão do zero. Esses três gráficos não apresentam nenhum padrão. No entanto, o gráfico (b) apresenta uma linha de pontos na diagonal, no canto superior à direita, que pode ser devido à grande quantidade de zeros ou valores próximos a zero nesta variável.

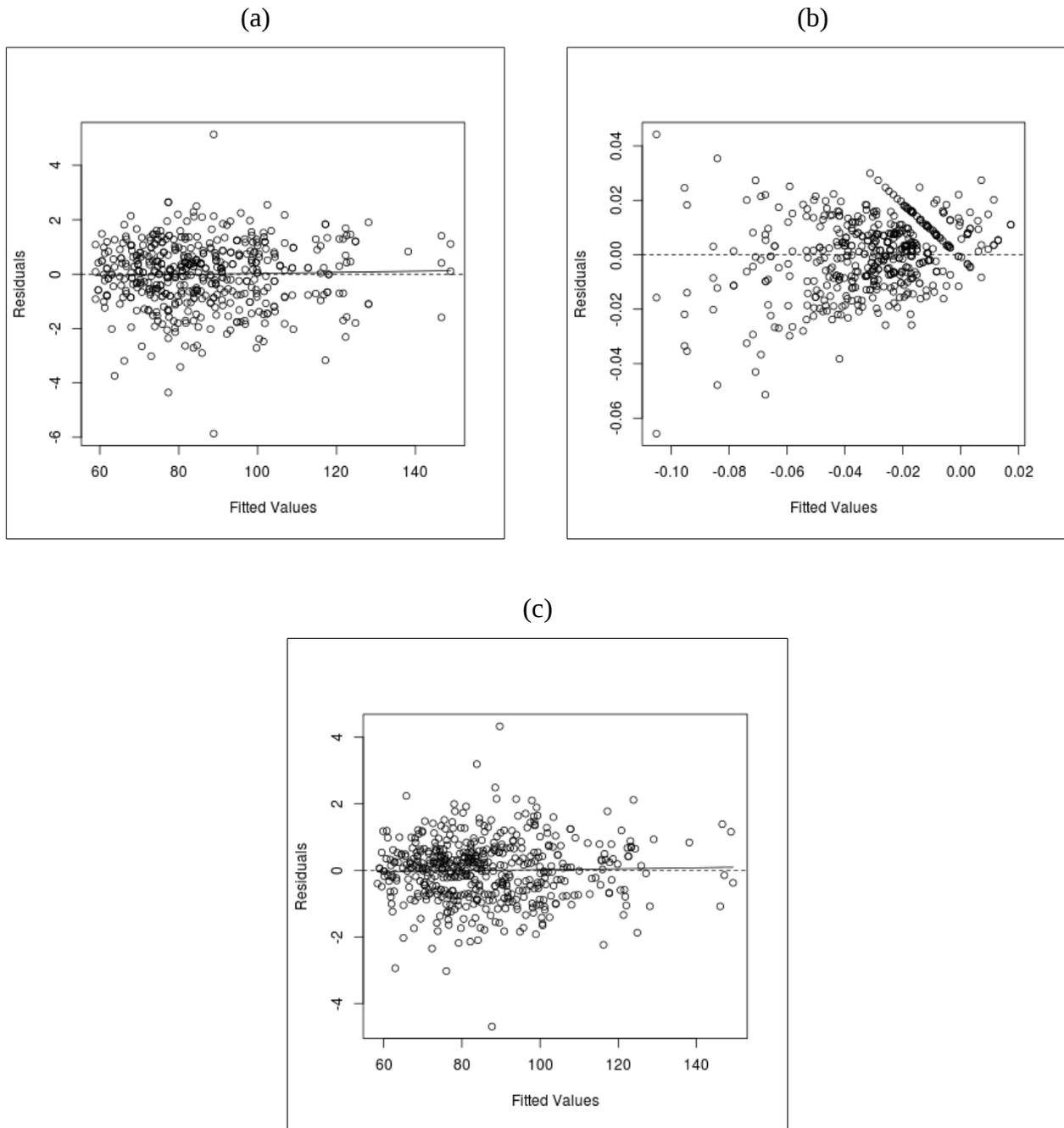


Figura 32: Valores ajustados versus resíduos dos modelos 10 (a), 1 (b) e 4 (c).

A suposição para os erros, dentro do grupo, pode ser avaliada com o gráfico de probabilidade normal dos resíduos (Figura 33). Observando os três gráficos, a suposição de normalidade parece plausível. No entanto, a presença de alguns “outliers” indica uma investigação mais profunda.

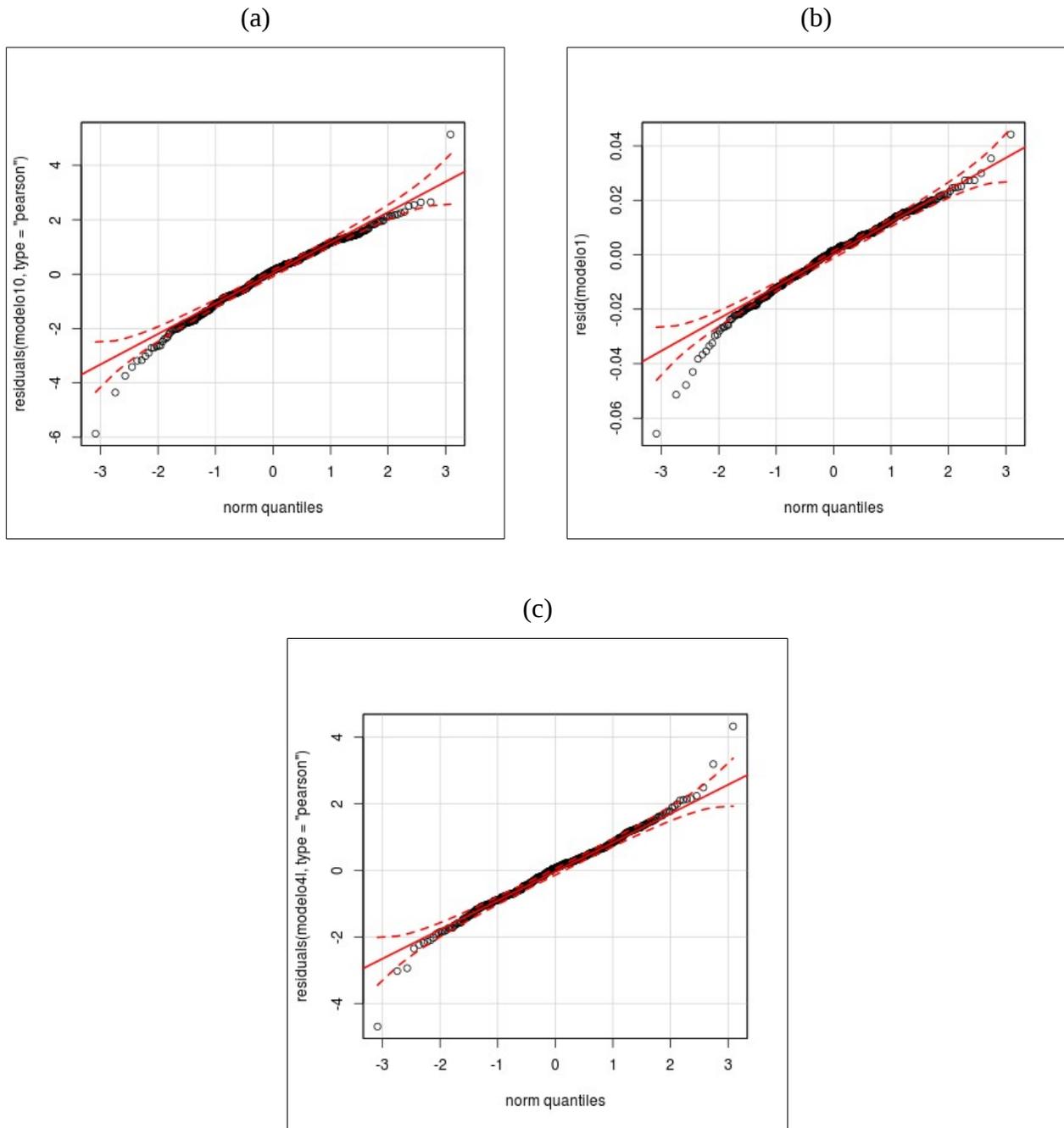


Figura 33: Gráficos de probabilidade normal (Q-Q plot) para os resíduos dos modelos 10 (a), 1 (b) e 4 (c).

O primeiro pressuposto em relação aos efeitos aleatórios que vamos analisar é o que diz respeito à normalidade. Para essa avaliação utilizaremos o Q-Q plot (Figura 34).

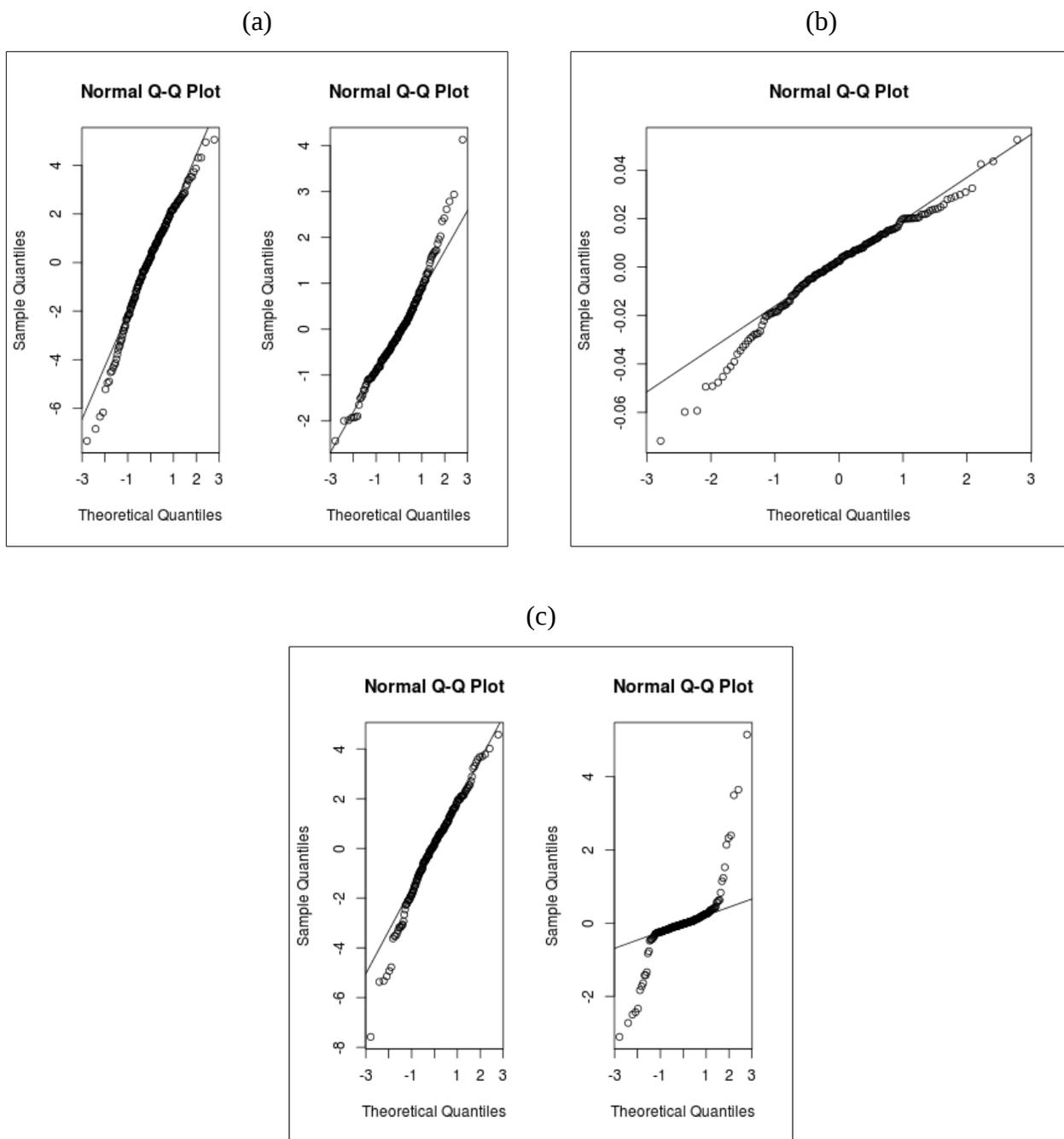


Figura 34: Gráficos de probabilidade normal (Q-Q plot) para os efeitos aleatórios (intercepto e sexo, respectivamente) dos modelos 10 (a), 1 (b) e 4 (c).

Dos três modelos acima, a suposição de normalidade parece razoável apenas para os efeitos aleatórios dos modelos dos gráficos (a) e (b), e para o primeiro efeito aleatório (intercepto) do modelo do gráfico (c). Nos outros gráficos, a assimetria na distribuição dos efeitos aleatórios é bastante evidente, assim como a presença de “outliers”.

Na Figura 35 encontram-se os “caterpillar plots” de cada modelo. Acrescentando ao que foi comentado sobre a figura acima, observa-se que grande parte dos intervalos do segundo gráfico de (c) sobrepõe a linha contínua que passa por zero. Isso é um indício de que não seria necessário adicionar este efeito aleatório aos seus respectivo modelo.

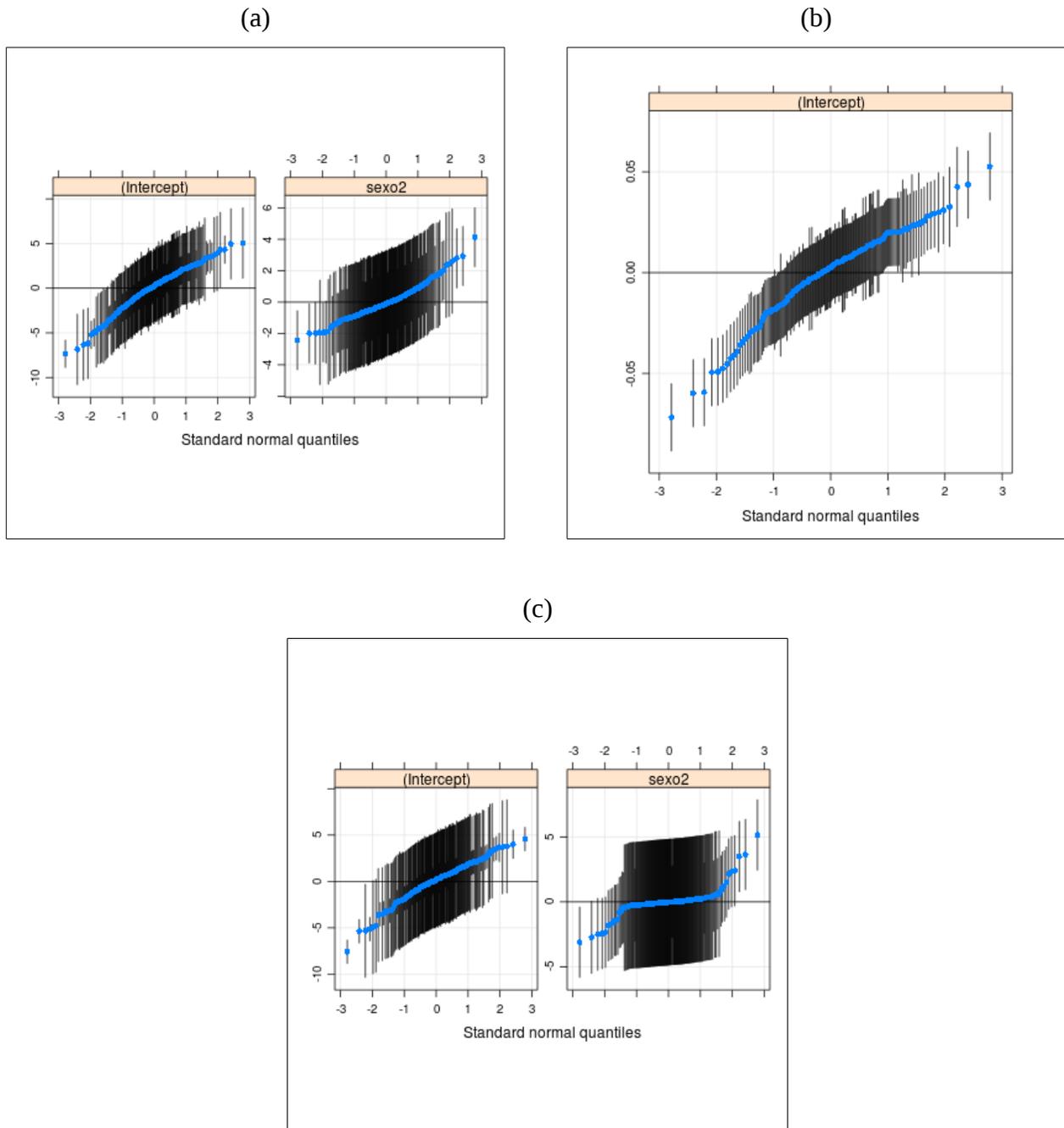


Figura 35: Gráficos de probabilidade normal dos efeitos aleatórios dos modelos 10 (a), 1 (b) e 4 (c).

De acordo com as figuras acima adicionado ao que foi comentado anteriormente, observamos que o modelo que teve o melhor comportamento em relação aos resíduos e aos efeitos aleatórios foi

o modelo composto pela variável resposta peso, pelas covariáveis peso inicial, grupo, idade e sexo e pelo efeito aleatório na variável sexo (modelo 10).

7 Discussão

Estudos longitudinais têm sido amplamente utilizados na área de saúde. No entanto, as análises estatísticas que acompanham esses estudos muitas vezes se limitam a testes estatísticos básicos, como testes de médias ou variâncias. Esses testes, embora úteis em um primeiro plano, limitam a análise a um nível superficial, pois, em alguns casos, não utilizam todos os dados disponíveis, enquanto em outros não levam em conta a estrutura dos dados (correlação e dependência das observações). Não se pode deixar de ressaltar que estudos longitudinais possuem um alto custo para serem feitos seja qual for o seu período de abrangência e, portanto, a não utilização de uma metodologia adequada para o tratamento dos dados significa subutilizar tanto a informação disponível como o investimento realizado, já que é nestes tipos de estudos que se encontram informações valiosas sobre o comportamento dos indivíduos ao longo do tempo.

Os métodos apropriados para analisar estes tipos de dados vêm sendo aperfeiçoados ao longo dos anos. No entanto, já existem pacotes disponíveis desde 1997, no pacote estatístico SAS (Verbeke & Molenberghs, 1997), e 2000, no pacote estatístico R por meio das bibliotecas *nlme* e *lme4* desenvolvidas, respectivamente, por Pinheiro & Bates (2000) e por Bates *et al.* (2011).

Apesar disso, diversos estudos epidemiológicos recentes sobre perda, mudança ou manutenção de peso apresentam a limitação acima citada e não utilizam os métodos adequados. Halpern *et al.* (2010) avaliaram a eficácia e tolerabilidade da associação de sibutramina e orlistat no tratamento da obesidade em um ensaio aberto de 6 meses. Quatrocentos e quarenta e seis pacientes obesos ou com sobrepeso que procuraram tratamento para obesidade na clínica privada foram avaliados a cada 2 semanas em um período de 3 e 6 meses. Para comparar os dados, utilizaram o teste t de Student pareado. Este tipo de teste é utilizado quando o que se deseja é verificar se a diferença observada entre dois momentos (efeito do tratamento por exemplo) é significativa. Portanto, este tipo de teste leva somente em consideração a medida inicial e final do estudo, ignorando as medidas entre estas. Halpern *et al.* (2010) concluíram neste estudo que a associação de orlistat e sibutramina é bastante eficiente e parece promover uma maior taxa de perda de peso do que a relatada em estudos clínicos realizados com cada droga separadamente.

Makoundou *et al.* (2011) avaliaram a eficácia de um programa específico de longo prazo para a manutenção da perda de peso, usando uma nova abordagem do orlistat, em 50 pacientes obesos que

foram seguidos durante 4 anos. Para isso, foi usado o teste não paramétrico dos postos sinalizados de Wilcoxon para comparar valores da linha de base e após o fim do estudo. Além disso, foi utilizada análise simples de regressão linear para avaliar relações entre mudanças no peso e para analisar o efeito da evolução do peso relativo à sensibilidade à insulina. Teste Qui-quadrado foi usado para comparar frequências percentuais do tratamento com orlistat e oficinas motivacionais entre grupos que ganharam e não ganharam peso. Os resultados encontrados neste estudo foram de que 73% dos indivíduos mantiveram mais do que 10% da sua perda de peso. Indivíduos com uma grande perda de peso estão em alto risco para a recuperação do peso.

Harrison *et al.* (2009) realizaram um estudo para determinar se o orlistat combinado com uma dieta de restrição calórica em pessoas acima do peso com esteatose hepática não alcoólica (NASH) resulta em perda de peso e melhora na histopatologia do fígado. Estas pessoas foram acompanhadas por 36 semanas. Testes não paramétricos (Mann-Whitney) foram usados para variáveis ordinais e nominais. Esse teste é alternativo ao teste t de Student para comparar as médias de duas amostras independentes. O único pressuposto exigido para a aplicação do teste Mann-Whitney é que as duas amostras sejam independentes e aleatórias, e que as variáveis em análise sejam numéricas ou ordinais (os pressupostos para a aplicabilidade do teste t-Student são mais exigentes: as populações de onde as amostras provêm têm distribuição normal; as amostras são independentes e aleatórias; as populações têm uma variância comum). Neste estudo, o grupo que utilizou orlistat perdeu em média 8,3% do peso enquanto que comparado à perda de 6,0% no grupo que fez uso da dieta mais vitamina E .

Apesar de termos ainda muitos estudos que não utilizam o método adequado para analisar dados longitudinais, tem-se observado que alguns estudos mais recentes preocupam-se em adotar modelos de efeitos mistos para analisar seus dados. Dentre eles podemos citar alguns.

Yancy *et al.* (2010), utilizaram modelos lineares mistos para testar hipóteses de diferenças de tratamento (orlistat mais uma dieta com baixo teor de gordura vs dieta com baixo teor de carboidrato) para as variáveis respostas contínuas (peso, pressão arterial) em um estudo de 48 semanas com 146 pacientes com sobrepeso ou obesos. Concluíram que pacientes que seguiram apenas a dieta tiveram melhorias semelhantes aos pacientes que combinaram dieta e orlistat para os parâmetros de peso, colesterol total e glicemia.

LeCheminant *et al.* (2005), compararam o uso de substitutos de refeições ou de orlistat durante a manutenção do peso após a perda de peso usando uma dieta de baixa caloria em noventa e dois adultos com sobrepeso ou obesos. Para determinar as diferenças de grupo durante a perda de

peso e durante a manutenção do peso, os dados foram analisados utilizando um modelo de efeitos mistos. Neste estudo eles concluíram que tratamentos posteriores à perda de peso a partir de uma dieta de baixa caloria, os substitutos de refeição e o orlistat foram ambos eficazes na manutenção de peso ao longo de um período de 52 semanas. Os substitutos de refeição podem ser uma estratégia alternativa viável aos medicamentos para manutenção do peso.

Apesar de encontrarmos alguns artigos que fazem uso de modelos de efeitos mistos em sua análise, essa quantidade ainda é pequena. Podemos citar dois motivos que justificam o pouco uso desse método: a dificuldade da própria metodologia e a escassez de material disponível. Para o entendimento, é necessário um investimento no estudo desta metodologia e, por isso, muitas vezes, o pesquisador opta por utilizar técnicas mais difundidas e mais simples. Além disso, um dos poucos softwares que disponibilizam funções para a análise de modelos lineares mistos, o *R*, ainda é pouco utilizado na área da saúde. A escassa documentação da biblioteca *lme4* - necessária para o uso dessas funções - torna a aplicação destes modelos ainda mais difícil.

Na análise de evolução da perda de peso, é imprescindível considerar a dependência entre as medidas do peso repetidas para o mesmo paciente. Os modelos mistos são úteis na modelagem deste tipo de desenho longitudinal com medidas repetidas, uma vez que não têm restrições caso os dados sejam desbalanceados (medidas obtidas em momentos diferentes nos pacientes e/ou número diferente de medidas repetidas por paciente) ou tenham observações faltantes.

O uso de modelos de efeitos mistos neste trabalho, não encontrou resultados diferentes do apresentado por Bloch *et al.* (2003). Assim como no estudo original, observou-se uma redução de peso em ambos os grupos. Porém, conforme observado nos resultados dos modelos, verifica-se que em diversos modelos os coeficientes associados ao grupo tratamento foram negativamente significativos, isto é, houve uma perda de peso significativa entre os pacientes que utilizaram o medicamento ainda que esta perda não seja expressiva. O paciente perde peso em resposta ao tratamento medicamentoso e provavelmente diminui a rigorosidade na dieta subsequentemente por achar que basta apenas a medicação. Os indivíduos do grupo controle apresentam uma perda progressiva, embora menor, em resposta à dietoterapia. Essa informação é de grande utilidade no acompanhamento de pacientes com objetivo de perda ponderal e que pode ser apresentada como uma vantagem do método proposto.

Neste trabalho, além de se observar a redução do peso como principal desfecho, outras abordagens seriam possíveis. Poderíamos analisar, por exemplo, a questão do dado faltante. Além

disso, uma pergunta interessante seria: que características afetam a participação dos indivíduos no estudo? Utilizando um modelo binomial se poderia buscar entender as características dos indivíduos que abandonaram o estudo, e orientar inclusive a imputação de dados, caso desejável. Como a variável resposta é binomial, o modelo adequado para tal análise seria o modelo linear generalizado misto.

É importante destacar que, em um processo de modelagem estatística, devemos dedicar uma grande parcela de tempo na análise exploratória dos dados. Observar como cada variável se comporta e qual a natureza da variável resposta que se deseja modelar, é essencial para que o ajuste dos modelos se dê de maneira correta. Ao utilizarmos, neste trabalho, a proporção de peso perdido como variável resposta, notamos a presença de uma grande quantidade de zeros, presente em aproximadamente 10% das medidas. Como isto poderia dificultar o processo de modelagem, adicionamos ao peso inicial de cada paciente 0,1kg com o intuito de diminuir os excessos de zero. Porém, conforme visto nos gráficos de resíduos, essa não tenha sido a melhor escolha. Uma possibilidade para um trabalho futuro seria, por exemplo, utilizar a distribuição beta inflacionada como proposta por Ospina & Ferrari (2011)

Além disso, neste trabalho, não tratamos formalmente da autocorrelação residual – que é inerente a dados longitudinais, nem analisamos se os modelos ajustados deram conta disso. Portanto propostas que tratem formalmente desta autocorrelação e que a avaliem, podem ser possíveis extensões deste trabalho.

O ganho maior deste trabalho foi em relação à utilização de formas gráficas alternativas de visualização de dados longitudinais, bem como o entendimento dos modelos de efeitos mistos. O ganho só não foi maior porque o banco de dados que possuíamos pertenceu a um estudo com um período curto de observação e, portanto, com poucas medidas. Uma das grandes vantagens deste método é exatamente quando se tem uma considerável variabilidade ao longo do tempo de uma ou mais variáveis - o que não aconteceu neste estudo.

8 Comentários Finais

A análise longitudinal constitui um ferramental que deve ser amplamente utilizado para a total e completa exploração de dados gerados por estudos de coorte, sejam eles balanceados ou não e/ou com dados faltantes. Utilizar somente duas medidas - geralmente início e fim - é descartar um número valioso de informações além de gerar uma interpretação precária dos resultados.

Apesar da biblioteca *lme4* do software *R* ser um excelente pacote para análise de modelos lineares mistos, a pouca quantidade de material disponível englobando teoria e aplicação torna a sua utilização pouco difundida. O rascunho da documentação oficial desta biblioteca foi lançado somente em 2010 (Bates, 2010), porém, como muito ainda se discute sobre os modelos de efeitos mistos, este livro ainda não foi lançado. O que se consegue encontrar são alguns capítulos de livros que apresentam parte da teoria e alguns exemplos, dentre os quais podemos destacar: Everitt & Hothorn, 2009; Zuur *et al.*, 2009; Baayen, 2008.

Através de uma perspectiva de aplicação, o presente trabalho ilustrou como abordagens alternativas de representação gráfica e de modelagem longitudinal podem ser usadas em conjunto para fornecer informações e conclusões complementares à utilização tradicional de uma simples análise de medidas resumidas, histogramas, testes-t, testes Qui-quadrado, ANOVA entre outros.

Em função dos resultados obtidos, é possível afirmar que este trabalho conseguiu atingir um objetivo básico que é a utilização de ferramentas gráficas e de modelos de efeitos mistos para facilitar um uso mais consistente de investigação e análise de dados provenientes de estudos longitudinais.

9 Referências Bibliográficas

Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974; 19: 716-23.

Alley DE, Metter EJ, Griswold ME, Harris TB, Simonsick EM, Longo DL et al. Changes in weight at the end of life: characterizing weight loss by time to death in a cohort study of older men. *Am J Epidemiol* 2010; 172(5):558-565.

Andreozzi, VL. Modelagem de Dados Longitudinais Aplicada a um Estudo Aleatorizado do Efeito da Vitamina A na Diarréia Infantil [Tese de Doutorado]. Rio de Janeiro: COPPE, Universidade do Rio de Janeiro; 2002.

Atkinson AC. *Plots, transformations and regression : an introduction to graphical methods of diagnostic regression*. Oxford, Oxford Science Publications, 1985.

Avenell A, Brown TJ, McGee MA, Campbell MK, Grant AM, Broom J, Jung RT. What interventions should we add to weight reducing diets in adults with obesity? A systematic review of randomized controlled trials of adding drug therapy, exercise, behaviour therapy or combinations of these interventions. *Journal of Human Nutrition and Dietetics* 2004; 17(Pt 4):293-316.

Baayen RH. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press, 2008.

Bates DM. *lme4: Mixed-Effects Modeling with R*. New York: Springer; Prepublication version at: <http://lme4.r-forge.r-project.org/book/>; 2010.

Bates D, Maechler M, Bolker B. *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42. <http://CRAN.R-project.org/package=lme4> ; 2011.

Behrens JT. Principles and procedures of exploratory data analysis. *Psychological Methods* 1997; 2:131-160.

Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000; 342:1878-86.

Berkey CS, Rockett HR, Field AE, Gillman MW, Colditz GA. Sugar-added beverages and adolescent weight change. *Obes Res* 2004; 12:778-88.

Berkowitz RI, Fujioka K, Daniels SR, Hopping AG, Owen S, Perry AC, et al. Effects of sibutramine treatment in obese adolescents: a randomized trial. *Ann Intern Med* 2006;145:81-90.

Bloch KV, Salles GF, Muxfeldt ES, Rocha, NA. Orlistat in hypertensive overweight/obese patients: results of a randomized clinical trial. *J Hypertension*. 2003; 21: 2159-65.

Bouchard C. *Physical activity and obesity*. Champaign: Human kinetics; 2000.

Bray GA. Health hazards of obesity. *Endocrinol Metab Clin North Am* 1996; 25:907-19.

Buehler AM, Cavalcanti AB, Suzumura EA, Carballo MT, Berwanger O. Como avaliar criticamente um ensaio clínico de alocação aleatória em terapia intensiva. *Rev Bras Ter Intensiva* 2009; 21:219-25.

Chambers JM, Cleveland WS, Kleiner B, Tukey PA. *Graphical Methods for Data Analysis*. Belmont (CA): Wadsworth, 1983.

Chanoine JP, Hampl S, Jensen C, Boldrin M, Hauptman J. Effect of orlistat on weight and body composition in obese adolescents: a randomized controlled trial. *JAMA* 2005; 293:2873-83.

Chen-Ku CH, Sandí-Bogantes MG. Actualización en obesidad. *Acta Med Costar* 2004; 46 Suppl 1:7-14.

Connoley IP, Heal DJ, Stock MJ. A study in rats of the effects of sibutramine on food intake and thermogenesis. *Br J Pharmacol* 1995; 114 Suppl:388.

Connoley IP, Liu Y-L, Frost I, Reckless IP, Heal DJ, Stock MJ. Thermogenic effects of sibutramine and its metabolites. *Br J Pharmacol* 1999; 126(6):1487-95.

Cook RJ, Yi GY, Lee KA, Gladman DD. A conditional Markov model for clustered progressive multistate processes under incomplete observation. *Biometrics* 2004; 60: 436-43.

Diggle, PJ, Liang, K-Y, Zeger, SL. *Analysis of Longitudinal Data*. Oxford: Clarendon Press; 1994.

Enders, CK, Tofighi D. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods* 2007; 12:121-38.

Everitt BS, Hothorn T: *A Handbook of Statistical Analysis Using R*, 2nd edition London: CRC Press; 2009.

Faraway JJ. *Extending the Linear Model with R*. Boca Raton, FL: Chapman & Hall; 2006.

Fausto, MA, Carneiro M, Antunes CMF, Pinto JA, Colosimo EA. O modelo de regressão linear misto para dados longitudinais: uma aplicação na análise de dados antropométricos desbalanceados. *Cad. Saúde Pública* 2008; 24(3):513-24.

Ferrari CKB. Metabolic syndrome and obesity: epidemiology and prevention by physical activity and exercise. *J Exerc Sci Fitness* 2008; 6(1):87-96.

Field AP. *Discovering statistics using SPSS: and sex and drugs and rock 'n' roll*. 3rd edition. London: Sage Publications; 2009.

Finer N. Clinical assessment, investigation and principles of management: realistic weight goals. *Clinical Obesity*, Kopelman PG & Stock MJ Eds., Blackwell Science, London, 1998; 350-76.

Finucane MM, Samet JH, Horton NJ. Translational methods in biostatistics: linear mixed effect regression models of alcohol consumption and HIV disease progression over time. *Epidemiol Perspect Innov.* 2007; 4:1-14.

Fortes RC, Guimarães NG, Haack A, Torres AAL, Carvalho KMB. Orlistat e sibutramina: bons coadjuvantes para perda e manutenção de peso? *Rev Bras Nutr Clin* 2006; 21(3):244-51.

Fox J, Weisberg S. *An {R} Companion to Applied Regression, Second Edition.* Thousand Oaks CA: Sage. URL: <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion> ; 2011.

Francischi RP, Pereira LO, Freitas CS, Klopfer M, Santos RC, Vieira P et al. Obesidade: atualização sobre sua etiologia, morbidade e tratamento. *Rev Nutr* 2000; 13:17-28.

Frees EW. *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences.* Cambridge University Press; 2004.

Gelman A, Hill J. *Data analysis using regression and multilevel/hierarchical models.* Cambridge: Cambridge University Press; 2007.

Halford JCG, Heal DJ, Blundell JE. Effects in the rat of sibutramine on food intake and the behavioural satiety sequence. *British Journal of Pharmacology* 1995; 114 Suppl:387.

Halpern A, Pepe RB, Monegaglia AP, Beyruti M. Efficacy and Tolerability of the Association of Sibutramine and Orlistat for Six Months in Overweight and Obese Patients . *Journal of Obesity*, 2010; 1-5.

Harrison SA, Fecht W, Brunt EM, Neuschwander-Tetri BA. Orlistat for overweight subjects with nonalcoholic steatohepatitis: A randomized, prospective trial. *Hepatology* 2009; 49(1):80-6.

Hernán-Daza CH. La obesidad: un desorden metabólico de alto riesgo para la salud. *Colomb Med* 2002; 33(1):72-80.

Hollywood A, Ogden J. Taking Orlistat: Predicting Weight Loss over 6 Months. *Journal of Obesity* 2011; 1-7.

Horton NJ, Lipsitz SR. Review of software to fit generalized estimating equation regression models. *Am Statistician* 1999; 53:160-69.

Instituto Brasileiro de Geografia e Estatística (IBGE). Pesquisa de Orçamentos Familiares (POF) 2002-2003. Ministério do Planejamento, Orçamento e Gestão, 2006.

Kaplan D. *Statistical modeling: A fresh approach*. CreateSpace; 2009.

Ker, H. *Visual-Graphical Methods for Exploring Longitudinal Data*. World Academy of Science, Engineering and Technology 2010; 62:445-452.

Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; 38:963-74.

LeCheminant JD, Jacobsen DJ, Hall MA, Donnelly JE. A comparison of meal replacements and medication in weight maintenance after weight loss. *J Am Coll of Nutr* 2005; 24(5):347-53.

Makoundou V, Pataky Z, Bobbioni-Harsch E, Gachoud JP, Habicht F, Golay A. Multi-factorial approach associated with a new 'on/off' Orlistat® use in a weight loss maintenance programme: 4 years follow-up. *Obes Facts* 2011; 4(3):191-6.

Martinez, WL, Martinez, AR, Solka J. *Exploratory data analysis with MATLAB, 2nd Edition*. Computer Science and Data Analysis Series. Boca Raton, FL: CRC Press; 2010.

Ministério da Saúde. Pesquisa Nacional de Demografia e Saúde da criança e da mulher - PNDS 2006: Dimensões do processo reprodutivo e da saúde da criança. Brasília; 2009. Available from: http://bvsms.saude.gov.br/bvs/publicacoes/pnds_crianca_mulher.pdf

Molenberghs G, Verbeke G. A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling* 2001; 1:235-69.

National Institute for Clinical Excellence (NICE). Guidance on the use of orlistat for the treatment of obesity in adults. *Technology Appraisal Guidance No.22*; 2001.

Nonino-Borges CB, Borges RM, Santos JE. Tratamento clínico da obesidade. *Medicina, Ribeirão Preto* 2006; 39(2):246-52.

Ospina R, Ferreira SLP. A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis* 2012; 56(6):1609-23.

Padwal R, Li SK, Lau DC. Long-term pharmacotherapy for overweight and obesity: a systematic review and meta-analysis of randomized controlled trials. *International Journal of Obesity* 2003; 27(Pt 12):1437-1446.

Pinheiro JC, Bates D. *Mixed-effects models in S and S-PLUS*. New York: Springer; 2000.

Pinheiro J, Bates D, DebRoy S, Sarkar D, the R Development Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-102; 2011.

Pi-Sunyer FX. Medical hazards of obesity. *Ann Intern Med* 1993; 119(7 Pt 2):655-60.

Phelan S, Wadden TA. Combining behavioral and pharmacological treatments for obesity. *Obesity research* 2002; 10(Pt 6):560-74.

R Development Core Team (2012). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

Rucker D, Padwal R, Curioni C, Lau DC. Long term pharmacotherapy for obesity and overweight: updated meta-analysis. *BMJ* 2007; 335(7631): 1194-9.

Sanson A, Smart D. Longitudinal research on mental health and behavioural problems of children and youth: New findings. *Australian Journal of Psychology* 2011; 63:1-5.

Schieleth H, Forstmeier W. Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology* 2008; 20:416-20.

Seidell JC. Dietary fat and obesity: an epidemiologic perspective. *American Journal of Clinical Nutrition* 1998; 67 Suppl 3: 546-50.

Seltman, HJ. *Experimental design and analysis*. Carnegie Mellon University; 2010. <http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf> (acessado em 29/março/2012).

Sheehan TJ, DuBrava S, DeChello LM, Fang Z. Rates of weight change for black and white Americans over a twenty year period. *Int J Obes Relat Metab Disord* 2003; 27:498-504.

Sosa RAB. ¿Cuál es el papel actual del tratamiento farmacológico de la obesidad en personas adultas? *Rev Endocrinol Nutr* 2004;12 Suppl 3:130-5.

Tukey JW. *Exploratory Data Analysis*. Reading: Addison-Wesley; 1977.

Verbeke G, Molenberghs G. *Linear Mixed Models In Praticce: A SAS Oriented Approach*. New York: Springer; 1997.

Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*. New York: Springer-Verlag; 2000. (Springer Series in Statistics).

West BT, Welch KB, Galecki AT. *Linear Mixed Models: A Practical Guide using Statistical Software*. Boca Raton, FL: Chapman & Hall/CRC Press; 2007.

Wickham H. ggplot2: elegant graphics for data analysis. Springer New York; 2009.

Wu L. Mixed Effects Models for Complex Data. Boca Raton, FL: Chapman & Hall/CRC; 2010.

Yancy WS Jr., Westman EC, McDuffie JR, Grambow SC, Jeffreys AS, Bolton J, et al. A randomized trial of a low-carbohydrate diet vs orlistat plus a low-fat diet for weight loss. Arch Intern Med 2010; 170 (2):136-45.

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM: Mixed effects models and extensions in ecology with R. New York: Springer; 2009.

10 Apêndices

10.1 Apêndice A1: Script utilizado para a análise exploratória e modelagem

```
## Análise Exploratória

# Abrindo o banco de dados
bancodadosconsulta <- read.csv("bancodadosconsulta.csv")
# Para ver a estrutura dos dados
str(bancodadosconsulta)
# Transformando as variáveis em fatores categóricos
bancodadosconsulta$sexo <- as.factor(bancodadosconsulta$sexo)
bancodadosconsulta$diabet <- as.factor(bancodadosconsulta$diabet)
bancodadosconsulta$sedent <- as.factor(bancodadosconsulta$sedent)
bancodadosconsulta$tab <- as.factor(bancodadosconsulta$tab)
bancodadosconsulta$histha <- as.factor(bancodadosconsulta$histha)
bancodadosconsulta$dislip <- as.factor(bancodadosconsulta$dislip)
bancodadosconsulta$histcor <- as.factor(bancodadosconsulta$histcor)
bancodadosconsulta$atif_s3m <- as.factor(bancodadosconsulta$atif_s3m)

# Abrindo o banco de dados
bancolongo <- read.csv("bancolongo.csv")
# Para ver a estrutura dos dados
str(bancolongo)
# Transformando para categorica
bancolongo$sexo <- as.factor(bancolongo$sexo)

# Criando Tabela 1
# "sexo" 1 = masc 2 = femin
table(bancodadosconsulta$sexo, bancodadosconsulta$grupo, useNA = "always")
table(bancodadosconsulta$sexo)
table(bancodadosconsulta$grupo)
```

```

## Criando Tabela 2
# "tab" 1 = sim 2 = não 3 = ign
table(bancodadosconsulta$tab, bancodadosconsulta$grupo, useNA = "always")
# "sedent" 1 = não 2 = sim 3 = ign
table(bancodadosconsulta$sedent, bancodadosconsulta$grupo, useNA = "always")
# "histcor" 1 = sim 2 = não 3 = ign
table(bancodadosconsulta$histcor, bancodadosconsulta$grupo, useNA = "always")
# "histha" 1 = sim 2 = não 3 = ign
table(bancodadosconsulta$histha, bancodadosconsulta$grupo, useNA = "always")
# "comorb" 1 = sim 2 = não
table(bancodadosconsulta$comorb, bancodadosconsulta$grupo, useNA = "always")
# "diabet" 1 = sim 2 = não 3 = ign
table(bancodadosconsulta$diabet, bancodadosconsulta$grupo, useNA = "always")
# "dislip" 1 = sim 2 = não 3 = ign
table(bancodadosconsulta$dislip, bancodadosconsulta$grupo, useNA = "always")
# "atif_s3m" 1 = sim 2 = não
table(bancodadosconsulta$atif_s3m, bancodadosconsulta$grupo, useNA = "always")

```

```

## Criando Figura 5
par(mfrow = c(2, 2))
ylim <- range(bancodadosconsulta[, "peso"], na.rm = TRUE)
tratamento <- subset(bancodadosconsulta, grupo == "T")[, "peso"]
boxplot(tratamento, main = "Tratamento", ylab = "Peso(kg)", ylim = ylim)
controle <- subset(bancodadosconsulta, grupo == "C")[, "peso"]
boxplot(controle, main = "Controle", ylim = ylim)
ylim <- range(bancodadosconsulta[, "idade"], na.rm = TRUE)
tratamento <- subset(bancodadosconsulta, grupo == "T")[, "idade"]
boxplot(tratamento, ylab = "Idade (anos)", ylim = ylim)
controle <- subset(bancodadosconsulta, grupo == "C")[, "idade"]
boxplot(controle, ylim = ylim)

```

```

## Criando Figura 6
# Carregando o pacote "ares" --> ele possui a funcao count.na que usarei abaixo
library(ares)
# Glicose, Colesterol
par(mfrow = c(2, 2))

```

```

#Glicose
ylim <- range(bancodadosconsulta[, c("glicose", "glicose0")], na.rm = TRUE)
tratamiento <- subset(bancodadosconsulta, grupo == "T")[, c("glicose", "glicose0")]
boxplot(tratamiento, main = "Tratamiento", ylab = "Glicose (mmHg)",
        names = c(paste(" Medida 1 \n (NA = ", count.na(tratamiento$glicose)[2],
") ", sep = "")),paste(" Medida 2 \n (NA = ", count.na(tratamiento$glicose0)[2],
") ", sep = "")),ylim = ylim)
controle <- subset(bancodadosconsulta, grupo == "C")[, c("glicose", "glicose0")]
boxplot(controle, main = "Controle", names = c(paste(" Medida 1 \n (NA = ",
count.na(controle$glicose)[2], ") ", sep = ""), paste(" Medida 2 \n (NA = ",
count.na(controle$glicose0)[2], ") ", sep = ""))), ylim = ylim)

#Colesterol
ylim <- range(bancodadosconsulta[, c("coltot", "coltot01")], na.rm = TRUE)
tratamiento <- subset(bancodadosconsulta, grupo == "T")[, c("coltot", "coltot01")]
boxplot(tratamiento, ylab = "Colesterol (mmol/l)", names = c(paste(" Medida 1 \n (NA
= ", count.na(tratamiento$coltot)[2], ") ", sep = ""), paste(" Medida 2 \n (NA = ",
count.na(tratamiento$coltot01)[2], ") ", sep = ""))), ylim = ylim)
controle <- subset(bancodadosconsulta, grupo == "C")[, c("coltot", "coltot01")]
boxplot(controle, names = c(paste(" Medida 1 \n (NA = ", count.na(controle$coltot)
[2], ") ", sep = ""), paste(" Medida 2 \n (NA = ", count.na(controle$coltot01)[2],
") ", sep = ""))),ylim = ylim)

# HDL, Triglicérido
par(mfrow = c(2, 2))
#HDL
ylim <- range(bancodadosconsulta[, c("hdl", "hdl01")], na.rm = TRUE)
tratamiento <- subset(bancodadosconsulta, grupo == "T")[, c("hdl", "hdl01")]
boxplot(tratamiento, main = "Tratamiento", ylab = "HDL (mmol/l)", names = c(paste("
Medida 1 \n (NA = ", count.na(tratamiento$hdl)[2], ") ", sep = ""), paste(" Medida 2
\n (NA = ", count.na(tratamiento$hdl01)[2], ") ", sep = ""))), ylim = ylim)
controle <- subset(bancodadosconsulta, grupo == "C")[, c("hdl", "hdl01")]
boxplot(controle, main = "Controle", names = c(paste(" Medida 1 \n (NA = ",
count.na(controle$hdl)[2], ") ", sep = ""), paste(" Medida 2 \n (NA = ",
count.na(controle$hdl01)[2], ") ", sep = ""))), ylim = ylim)

```

```

#Triglicerídeo
ylim <- range(bancodadosconsulta[, c("trig", "trig01")], na.rm = TRUE)
tratamento <- subset(bancodadosconsulta, grupo == "T")[, c("trig", "trig01")]
boxplot(tratamento, ylab = "Triglicerídeos (mmol/l)", names = c(paste(" Medida 1 \n
(NA = ", count.na(tratamento$trig)[2], ") ", sep = ""), paste(" Medida 2 \n (NA =
", count.na(tratamento$trig01)[2], ") ", sep = "")), ylim = ylim)
controle <- subset(bancodadosconsulta, grupo == "C")[, c("trig", "trig01")]
boxplot(controle, names = c(paste(" Medida 1 \n (NA = ", count.na(controle$trig)
[2], ") ", sep = ""), paste(" Medida 2 \n (NA = ", count.na(controle$trig01)[2], ")
", sep = "")), ylim = ylim)

```

```

## Criando Tabela 3

```

```

tab1 <- table(bancolongo$cons, bancolongo$grupo)
row.sums <- apply(tab1, 1, sum)
cbind(tab1, row.sums)

```

```

## Criando Tabela 4

```

```

# Criando a variavel consulta total
consultatotal <- aggregate(bancolongo$numcons,
by = as.data.frame(bancolongo$ficha), FUN = sum, na.rm = TRUE)
# Dando nomes as colunas
names(consultatotal) <- c("ficha", "consultatotal")
# Juntado esta nova variavel ao banco de dados
bancodadosconsulta <- merge(bancodadosconsulta, consultatotal, by = "ficha")
# Criando a tabela
tab2 <- table(bancodadosconsulta$consultatotal, bancodadosconsulta$grupo)
row.sums <- apply(tab2, 1, sum)
cbind(tab2, Total = row.sums)

```

```

## Criando Figura 8
# Peso ao longo do tempo
layout(matrix(1:2, nrow = 1))
ylim <- range(bancodadosconsulta[, c("peso", "peso01", "peso02", "peso03")],
na.rm = TRUE)
tratamento <- subset(bancodadosconsulta, grupo == "T")[, c("peso", "peso01",
"peso02", "peso03")]
boxplot(tratamento, main = "Tratamento", ylab = "Peso(kg)", xlab = "Consultas",
names = c("baseline", 1, 2, 3), ylim = ylim)
controle <- subset(bancodadosconsulta, grupo == "C")[, c("peso", "peso01",
"peso02", "peso03")]
boxplot(controle, main = "Controle", xlab = "Consultas", names = c("baseline", 1,
2, 3), ylim = ylim)

```

```

## Criando Figura 9
# Peso ao longo do tempo with NOTCH
layout(matrix(1:2, nrow = 1))
ylim <- range(bancodadosconsulta[, c("peso", "peso01", "peso02", "peso03")],
na.rm = TRUE)
tratamento <- subset(bancodadosconsulta, grupo == "T")[, c("peso", "peso01",
"peso02", "peso03")]
boxplot(tratamento, main = "Tratamento", ylab = "Peso(kg)", xlab = "Consultas",
names = c("baseline", 1, 2, 3), ylim = ylim, notch = TRUE)
controle <- subset(bancodadosconsulta, grupo == "C")[, c("peso", "peso01",
"peso02", "peso03")]
boxplot(controle, main = "Controle", xlab = "Consultas", names = c("baseline", 1,
2, 3), ylim = ylim, notch = TRUE)

```

```

## Criando Figura 10
idunicoC <- unique(bancolongo$ficha[bancolongo$grupo == "C"])
idunicoT <- unique(bancolongo$ficha[bancolongo$grupo == "T"])
bancolongoC <- bancolongo[bancolongo$grupo == "C", ]
bancolongoT <- bancolongo[bancolongo$grupo == "T", ]
ylim <- range(bancolongo$peso, na.rm = TRUE)
xlim <- range(bancolongo$cons)
condicaoC <- bancolongoC$ficha %in% idunicoC[1]

```

```

condicaoT <- bancolongoT$ficha %in% idunicoT[1]
plot(bancolongoT$cons[condicaoT], bancolongoT$peso[condicaoT], type = "l",
ylim = ylim, xlim = xlim, xlab = "Consultas", ylab = "Peso (Kg)", col = "gray")
for (i in 1:length(idunicoT)){
  condicaoT <- bancolongoT$ficha %in% idunicoT[i]
  lines(bancolongoT$cons[condicaoT], bancolongoT$peso[condicaoT], col = "red")
}
for (i in 1:length(idunicoC)){
  condicaoC <- bancolongoC$ficha %in% idunicoC[i]
  lines(bancolongoC$cons[condicaoC], bancolongoC$peso[condicaoC], col = "blue")
}

## Criando Figura 11
plot(bancolongoT$cons[condicaoT], bancolongoT$peso[condicaoT], type = "l", ylim =
ylim, xlim = xlim, xlab = "Consultas", ylab = "Peso (Kg)", main = "Grupo
Tratamento", col = "gray")
for (i in 1:length(idunicoT)){
  condicaoT<-bancolongoT$ficha %in% idunicoT[i]
  lines(bancolongoT$cons[condicaoT], bancolongoT$peso[condicaoT], col = "red")
}
plot(bancolongoC$cons[condicaoC], bancolongoC$peso[condicaoC], type = "l", ylim =
ylim, xlim = xlim, xlab = "Consultas", ylab = "Peso (Kg)", main = "Grupo Controle",
col = "gray")
for (i in 1:length(idunicoC)){
  condicaoC<-bancolongoC$ficha %in% idunicoC[i]
  lines(bancolongoC$cons[condicaoC], bancolongoC$peso[condicaoC], col = "blue")
}

```

```

## Criando Figura 12
library(ggplot2)
graphT <- ggplot(bancolongoT, aes(factor(cons), peso), na.rm = T) + geom_boxplot()
+ xlab("Consulta") + ylab("Peso(kg)")
graphT + geom_line(aes(group = ficha), colour = "lightsalmon")
graphC <- ggplot(bancolongoC, aes(factor(cons), peso), na.rm = T) + geom_boxplot()
+ xlab("Consulta") + ylab("Peso(kg)")
graphC + geom_line(aes(group = ficha), colour = "lightskyblue")

```

```
## Criando Figura 13
```

```
# Figura 13-a
```

```

ylim <- range(83:92, na.rm = TRUE)
y <- cbind(mean((bancodadosconsulta$peso)[bancodadosconsulta$grupo == "T"]),
mean((bancodadosconsulta$peso01)[bancodadosconsulta$grupo == "T"], na.rm = TRUE),
mean((bancodadosconsulta$peso02)[bancodadosconsulta$grupo == "T"], na.rm = TRUE),
mean((bancodadosconsulta$peso03)[bancodadosconsulta$grupo == "T"], na.rm = TRUE))

```

```
x <- c(0, 1, 2, 3)
```

```
plot(x, y, ylim = ylim, xlab = "Consultas", ylab = "Peso médio(kg)")
```

```
lines(x, y, lwd = 4, col = "red")
```

```

z <- cbind(mean((bancodadosconsulta$peso)[bancodadosconsulta$grupo == "C"]),
mean((bancodadosconsulta$peso01)[bancodadosconsulta$grupo == "C"], na.rm = TRUE),
mean((bancodadosconsulta$peso02)[bancodadosconsulta$grupo == "C"], na.rm = TRUE),
mean((bancodadosconsulta$peso03)[bancodadosconsulta$grupo == "C"], na.rm = TRUE))

```

```
lines(x, z, lwd = 5, col = "blue")
```

```
# Figura 13-b
```

```
source("http://www.biostat.jhsph.edu/~pmurakam/xtgraph.R")
```

```

xtgraph(y = bancolongo$peso, id = bancolongo$ficha, time = bancolongo$cons, group =
bancolongo$grupo, las = 1, ylab = "Peso (kg)", xlab = "Consultas")

```

```

## Criando Figura 14
# Tirando uma sample de pacientes de tamanho 9 para construir gráficos individuais
sample(bancodadosconsulta$ficha[bancodadosconsulta$grupo == "T"], 9,
replace = FALSE, prob = NULL)
# 137  9  74 112 144 147  60  68  37
sample(bancodadosconsulta$ficha[bancodadosconsulta$grupo == "C"], 9,
replace = FALSE, prob = NULL)
# 196 172 127  69  40 197 115  16  65

# Construindo os gráficos individuais
# Tratamento
par(mfrow = c(3, 3))
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 137],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 137],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 137],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 137])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 137")
lines(x, y, lwd = 4, col = "red")
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 9],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 9],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 9],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 9])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 9")
lines(x, y, lwd = 4, col = "red")
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 74],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 74],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 74],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 74])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 74")
lines(x, y, lwd = 4, col = "red")
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 112],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 112],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 112],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 112])
x <- c(0, 1, 2, 3)

```

```

plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 112")
lines(x, y, lwd = 4, col = "red")
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 144],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 144],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 144],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 144])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 144")
lines(x, y, lwd = 4, col = "red")
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 147],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 147],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 147],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 147])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 147")
lines(x, y, lwd = 4, col = "red")
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 60],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 60],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 60],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 60])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 60")
lines(x, y, lwd = 4, col = "red")
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 68],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 68],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 68],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 68])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 68")
lines(x, y, lwd = 4, col = "red")
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 37],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 37],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 37],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 37])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 37")
lines(x, y, lwd = 4, col = "red")

```

```

# Controle
par(mfrow = c(3, 3))
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 196],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 196],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 196],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 196])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 196")
lines(x, y, lwd = 4, col = "blue")

y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 172],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 172],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 172],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 172])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 172")
lines(x, y, lwd = 4, col = "blue")

y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 127],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 127],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 127],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 127])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 127")
lines(x, y, lwd = 4, col = "blue")

y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 69],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 69],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 69],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 69])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 69")
lines(x, y, lwd = 4, col = "blue")

y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 40],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 40],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 40],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 40])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 40")
lines(x, y, lwd = 4, col = "blue")

```

```

y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 197],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 197],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 197],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 197])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 197")
lines(x, y, lwd = 4, col = "blue")
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 115],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 115],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 115],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 115])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 115")
lines(x, y, lwd = 4, col = "blue")
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 16],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 16],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 16],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 16])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 16")
lines(x, y, lwd = 4, col = "blue")
y <- cbind((bancodadosconsulta$peso)[bancodadosconsulta$ficha == 65],
           (bancodadosconsulta$peso01)[bancodadosconsulta$ficha == 65],
           (bancodadosconsulta$peso02)[bancodadosconsulta$ficha == 65],
           (bancodadosconsulta$peso03)[bancodadosconsulta$ficha == 65])
x <- c(0, 1, 2, 3)
plot(x, y, xlab = "Consulta", ylab = "Peso", main = "Paciente 65")
lines(x, y, lwd = 4, col = "blue")

## Criando Figura 15
# Figura 15-a
library(nlme)
cons0T <- bancolongoT[bancolongoT$cons == 0, c("ficha", "peso", "cons")]
bancoT70 <- bancolongoT[bancolongoT$ficha%in%cons0T[cons0T$peso < 70, ][, 1], ]
gdT70 <- groupedData( peso ~ cons | ficha, data = bancoT70)
plot(gdT70)

```

```

# Figura 15-b
bancoT7080 <- bancolongoT[bancolongoT$ficha%in%cons0T[cons0T$peso >= 70 &
cons0T$peso < 80, ][, 1], ]
gdT7080 <- groupedData( peso ~ cons | ficha, data = bancoT7080)
plot(gdT7080)

## Criando Figura 16
# Figura 16-c
bancoT8090 <- bancolongoT[bancolongoT$ficha%in%cons0T[cons0T$peso >= 80 &
cons0T$peso < 90, ][, 1], ]
gdT8090 <- groupedData( peso ~ cons | ficha, data = bancoT8090)
plot(gdT8090)
# Figura 16-d
bancoT90100 <- bancolongoT[bancolongoT$ficha%in%cons0T[cons0T$peso >= 90 &
cons0T$peso < 100, ][, 1], ]
gdT90100 <- groupedData( peso ~ cons | ficha, data = bancoT90100)
plot(gdT90100)

## Criando Figura 17
# Figura 17-e
bancoT100110 <- bancolongoT[bancolongoT$ficha%in%cons0T[cons0T$peso >= 100 &
cons0T$peso < 110, ][, 1], ]
gdT100110 <- groupedData( peso ~ cons | ficha, data = bancoT100110)
plot(gdT100110)
# Figura 17-f
bancoT110 <- bancolongoT[bancolongoT$ficha%in%cons0T[cons0T$peso >= 110, ][, 1], ]
gdT110 <- groupedData( peso ~ cons | ficha, data = bancoT110)
plot(gdT110)

## Criando Figura 18
# Figura 18-a
cons0C <- bancolongoC[bancolongoC$cons == 0, c("ficha", "peso", "cons")]
bancoC70 <- bancolongoC[bancolongoC$ficha%in%cons0C[cons0C$peso < 70, ][, 1], ]
gdC70 <- groupedData( peso ~ cons | ficha, data = bancoC70)
plot(gdC70)

```

```

# Figura 18-b
bancoC7080 <- bancolongoC[bancolongoC$ficha%in%cons0C[cons0C$peso >= 70 &
cons0C$peso < 80, ][, 1], ]
gdC7080 <- groupedData( peso ~ cons | ficha, data = bancoC7080)
plot(gdC7080)

# Figura 18-c
bancoC8090 <- bancolongoC[bancolongoC$ficha%in%cons0C[cons0C$peso >= 80 &
cons0C$peso < 90, ][, 1], ]
gdC8090 <- groupedData( peso ~ cons | ficha, data = bancoC8090)
plot(gdC8090)

# Figura 18-d
bancoC90100 <- bancolongoC[bancolongoC$ficha%in%cons0C[cons0C$peso >= 90 &
cons0C$peso < 100, ][, 1], ]
gdC90100 <- groupedData( peso ~ cons | ficha, data = bancoC90100)
plot(gdC90100)

## Criando Figura 19
# Figura 19-e
bancoC100110 <- bancolongoC[bancolongoC$ficha%in%cons0C[cons0C$peso >= 100 &
cons0C$peso < 110, ][, 1], ]
gdC100110 <- groupedData( peso ~ cons | ficha, data = bancoC100110)
plot(gdC100110)

# Figura 19-f
bancoC110 <- bancolongoC[bancolongoC$ficha%in%cons0C[cons0C$peso >= 110, ][, 1], ]
gdC110 <- groupedData( peso ~ cons | ficha, data = bancoC110)
plot(gdC110)

## Criando Figuras 20 e 21
gdbancolongofichagrupo1 <- groupedData(peso ~ cons | fichagrupo,
data = bancolongofichagrupo)
plot(gdbancolongofichagrupo1[1:221, ], outer = ~ grupo, aspect = 2)
plot(gdbancolongofichagrupo1[222:403, ], outer = ~ grupo, aspect = 2)
plot(gdbancolongofichagrupo1[404:598, ], outer = ~ grupo, aspect = 2)
plot(gdbancolongofichagrupo1[599:729, ], outer = ~ grupo, aspect = 2)

```

```
#####
##### Modelagem

# Abrindo o banco de dados
bancolimpo <- read.csv("bancolongofichagrupoLIMPO.csv")
bancolimpo$sexo <- as.factor(bancolimpo$sexo)

# Idade Centrada na média
bancolimpo$idadecentrada <- bancolimpo$idade - mean(bancolimpo$idade)
# Peso Inicial Centrado na média
bancolimpo$pesoinicialcentrado <- bancolimpo$pesoinicial -
mean(bancolimpo$pesoinicial,na.rm = TRUE)

### Modelos Bloco I - resposta peso
library(lme4)
# Parte I
modelo1 <- lmer(peso ~ grupo + (1|fichagrupo), data = bancolimpo)
modelo2 <- lmer(peso ~ grupo + idadecentrada + sexo + (1|fichagrupo),
data = bancolimpo)
modelo3 <- lmer(peso ~ idadecentrada + sexo*grupo + (1|fichagrupo),
data = bancolimpo)
modelo4 <- lmer(peso ~ sexo + idadecentrada*grupo + (1|fichagrupo),
data = bancolimpo)
# Parte II
modelo5 <- lmer(peso ~ pesoinicialcentrado + grupo + (1|fichagrupo),
data = bancolimpo)
modelo6 <- lmer(peso ~ pesoinicialcentrado + grupo + idadecentrada + sexo +
(1|fichagrupo), data = bancolimpo)
modelo7 <- lmer(peso ~ pesoinicialcentrado + idadecentrada + sexo*grupo +
(1|fichagrupo), data = bancolimpo)
modelo8 <- lmer(peso ~ pesoinicialcentrado + sexo + idadecentrada*grupo +
(1|fichagrupo), data = bancolimpo)
# Parte III
modelo9 <- lmer(peso ~ pesoinicialcentrado + grupo + idadecentrada + sexo +
(idadecentrada|fichagrupo), data = bancolimpo)
modelo10 <- lmer(peso ~ pesoinicialcentrado + grupo + idadecentrada + sexo + (sexo|
fichagrupo), data = bancolimpo)
```

```

### Modelos Bloco II - resposta proporcao de peso perdido
# Parte I
modelo1 <- lmer(propessonovo ~ pesoinicialnovocentrado + grupo + (1|fichagrupo),
data = bancolimpo)
modelo2 <- lmer(propessonovo ~ pesoinicialnovocentrado + grupo + idadecentrada +
sexo + (1|fichagrupo), data = bancolimpo)
# Parte II
modelo3 <- lmer(propessonovo ~ pesoinicialnovocentrado + idadecentrada + sexo*grupo
+ (1|fichagrupo), data = bancolimpo)
modelo4 <- lmer(propessonovo ~ pesoinicialnovocentrado + sexo + idadecentrada*grupo
+ (1|fichagrupo), data = bancolimpo)
# Parte III
modelo5 <- lmer(propessonovo ~ pesoinicialnovocentrado + grupo + idadecentrada +
sexo + (idadecentrada|fichagrupo), data = bancolimpo)
modelo6 <- lmer(propessonovo ~ pesoinicialnovocentrado + grupo + idadecentrada +
sexo + (sexo|fichagrupo), data = bancolimpo)

```

```

### Modelos Bloco III - resposta peso e interaçao tratamento-tempo
# Parte I
modelo11 <- lmer(peso ~ pesoinicialcentrado + grupo:cons + (1|fichagrupo),
data = bancolimpo)
modelo21 <- lmer(peso ~ pesoinicialcentrado + sexo + idadecentrada + grupo:cons +
(1|fichagrupo), data = bancolimpo)
# Parte II
modelo31 <- lmer(peso ~ pesoinicialcentrado + sexo + idadecentrada + grupo:cons +
(idadecentrada|fichagrupo), data = bancolimpo)
modelo41 <- lmer(peso ~ pesoinicialcentrado + sexo + idadecentrada + grupo:cons +
(sexo|fichagrupo), data = bancolimpo)

```

```

# Melhores modelos
modelo10 <- lmer(peso ~ pesoinicialcentrado + grupo + idadecentrada + sexo + (sexo|
fichagrupo), data = bancolimpo)
modelo1 <- lmer(propessonovo ~ pesoinicialnovocentrado + grupo + (1|fichagrupo),
data = bancolimpo)
modelo41 <- lmer(peso ~ pesoinicialcentrado + sexo + idadecentrada + grupo:cons +
(sexo|fichagrupo), data = bancolimpo)
# Criando Figura 32
# Figura 32-a
plot(fitted(modelo10), residuals(modelo10), xlab = "Fitted Values",
ylab = "Residuals")
abline(h = 0, lty = 2)
# Figura 32-b
plot(fitted(modelo1), residuals(modelo1), xlab = "Fitted Values",
ylab = "Residuals")
abline(h = 0, lty = 2)
# Figura 32-c
plot(fitted(modelo41), residuals(modelo41), xlab = "Fitted Values",
ylab = "Residuals")
abline(h = 0, lty = 2)

# Criando Figura 33
library(car)
# Figura 33-a
qqPlot(resid(modelo10))
# Figura 33-b
qqPlot(resid(modelo1))
# Figura 33-c
qqPlot(resid(modelo41))

## Criando Figura 34
# Figura 34-a
par(mfrow = c(1, 2))
rr1 <- ranef(modelo10)
qqnorm(rr1$fichagrupo[[1]])
qqline(rr1$fichagrupo[[1]])
qqnorm(rr1$fichagrupo[[2]])

```

```

qqline(rr1$fichagrupo[[2]])
# Figura 34-b
rr1 <- ranef(modelo1)
qqnorm(rr1$fichagrupo[[1]])
qqline(rr1$fichagrupo[[1]])
# Figura 34-c
par(mfrow = c(1, 2))
rr1 <- ranef(modelo41)
qqnorm(rr1$fichagrupo[[1]])
qqline(rr1$fichagrupo[[1]])
qqnorm(rr1$fichagrupo[[2]])
qqline(rr1$fichagrupo[[2]])

## Criando Figura 35
# Figura 35-a
qqmath(ranef(modelo10, post = TRUE))$fichagrupo
# Figura 35-b
qqmath(ranef(modelo1, post = TRUE))$fichagrupo
# Figura 35-c
qqmath(ranef(modelo41, post = TRUE))$fichagrupo

```

10.2 Apêndice A2: Gráficos dos modelos cuja variável resposta é o peso

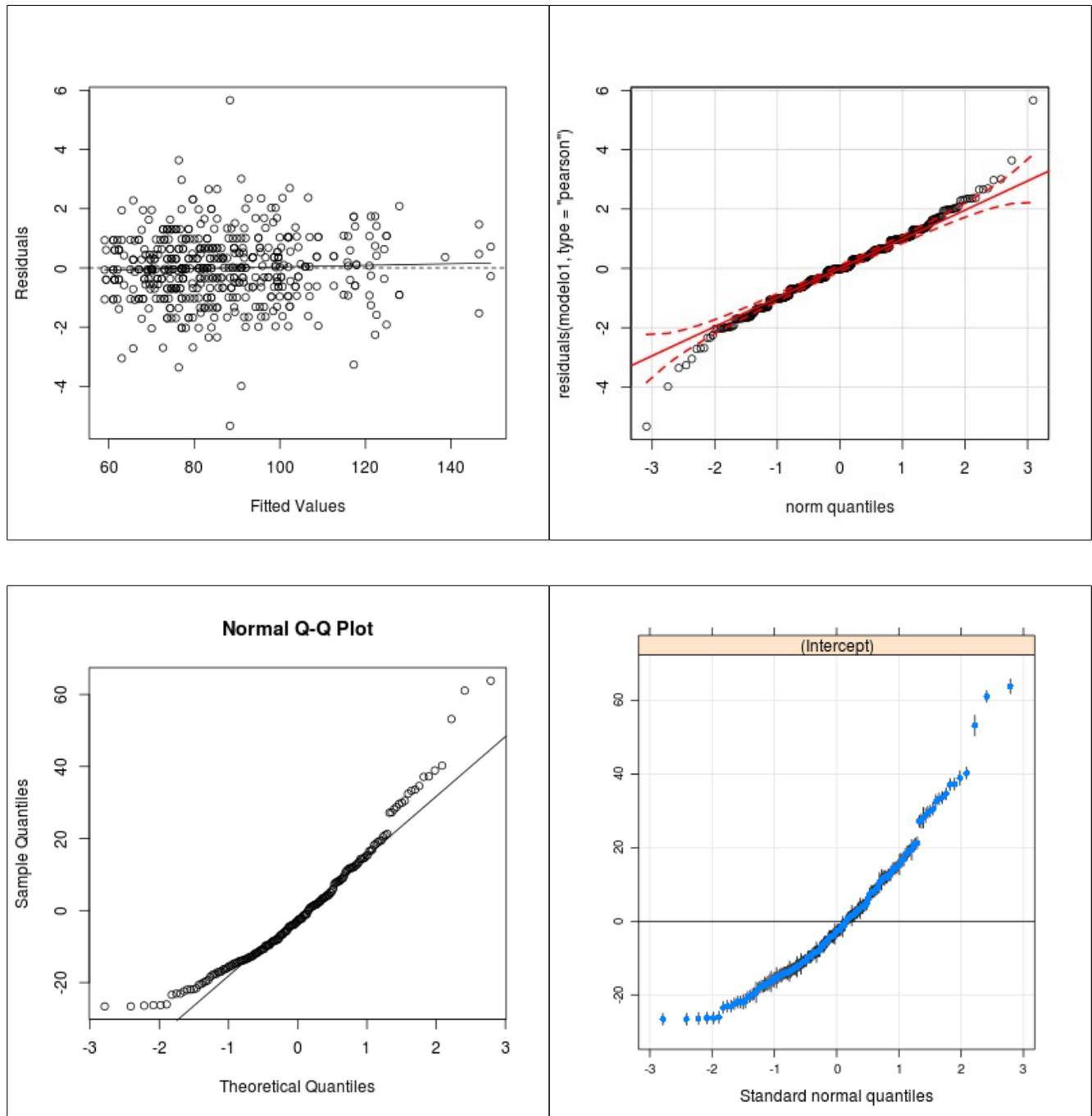


Figura 36: Diagnósticos para o modelo 1.

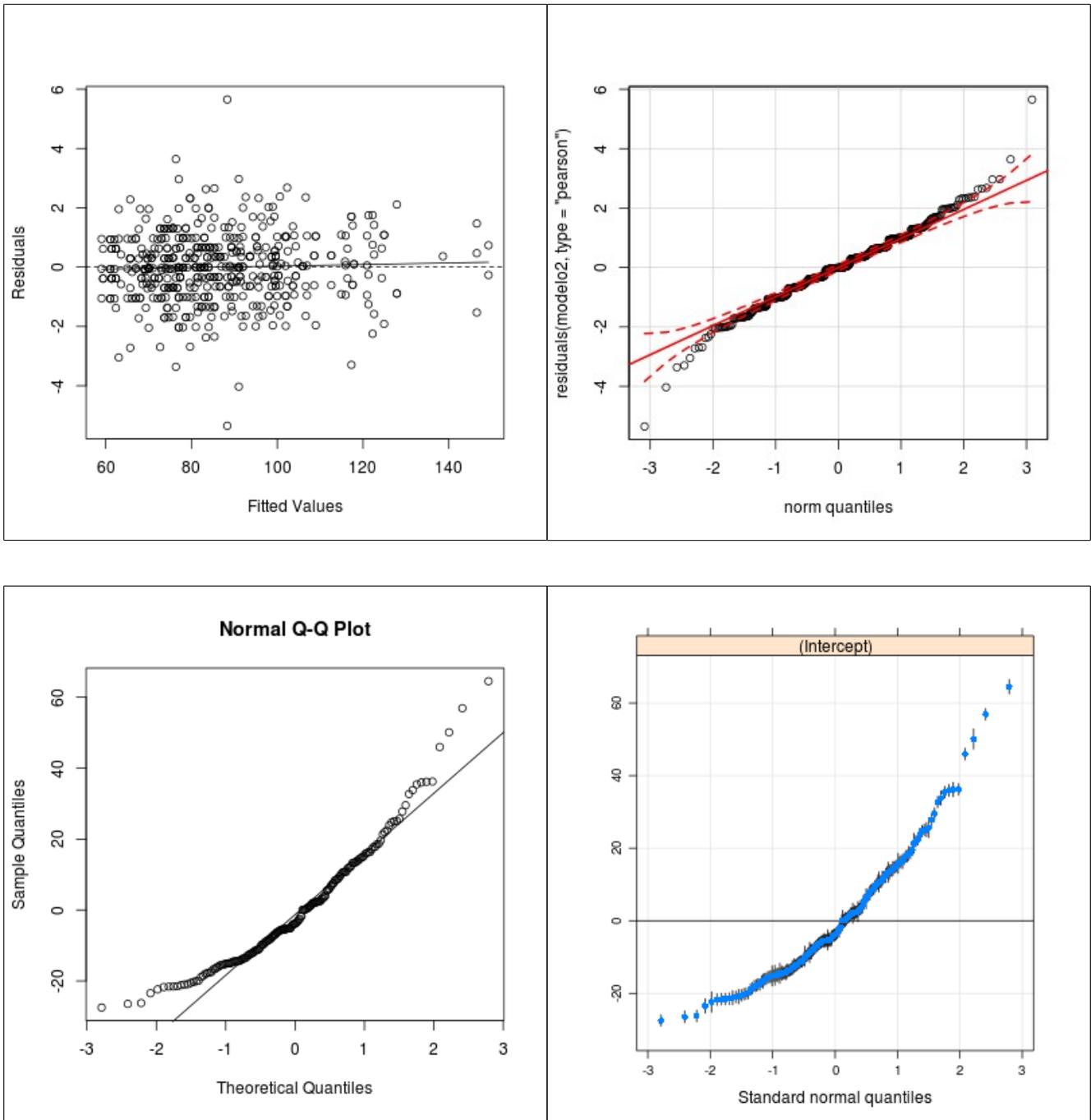


Figura 37: Diagnósticos para o modelo 2.

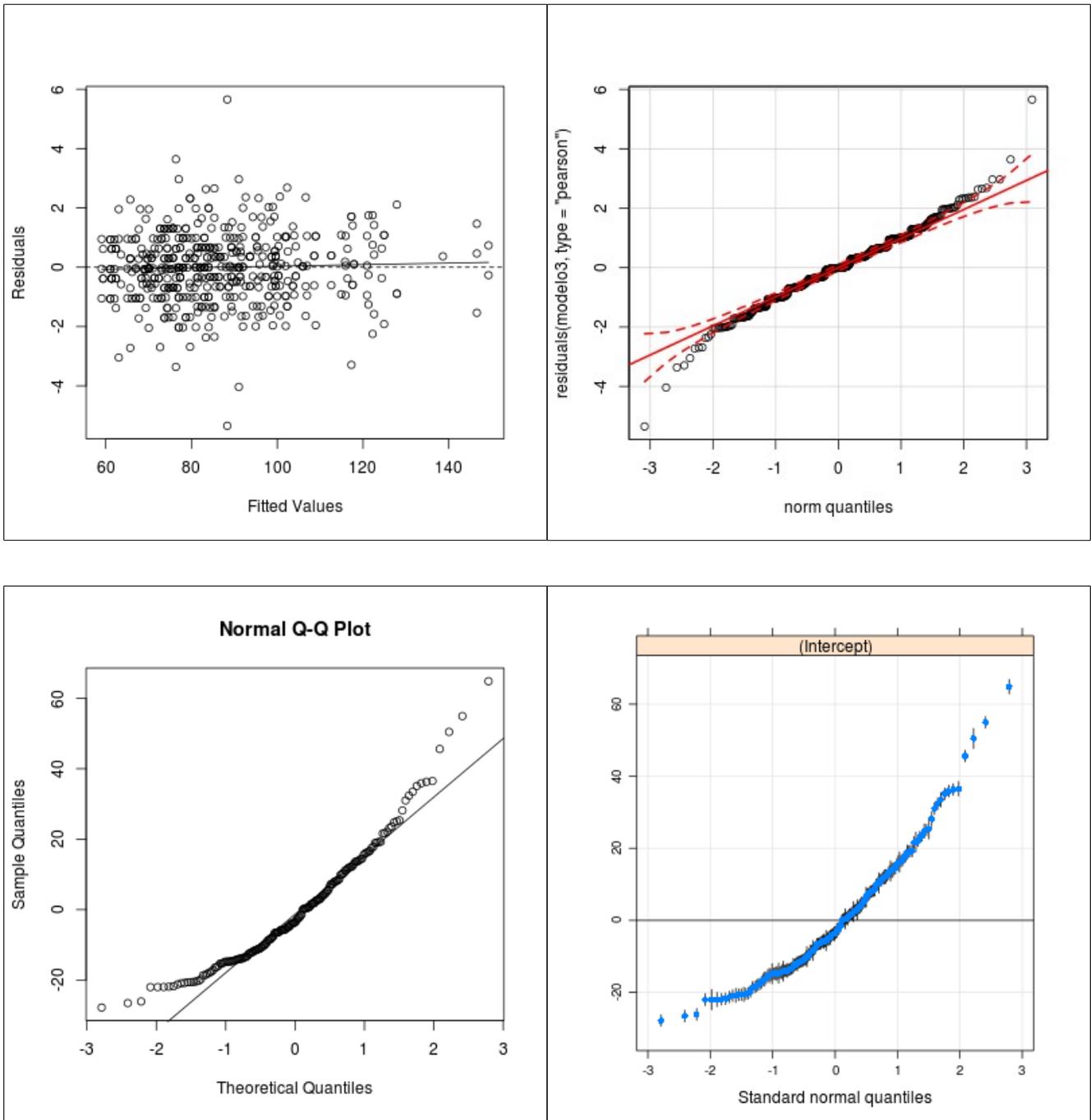


Figura 38: Diagnósticos para o modelo 3.

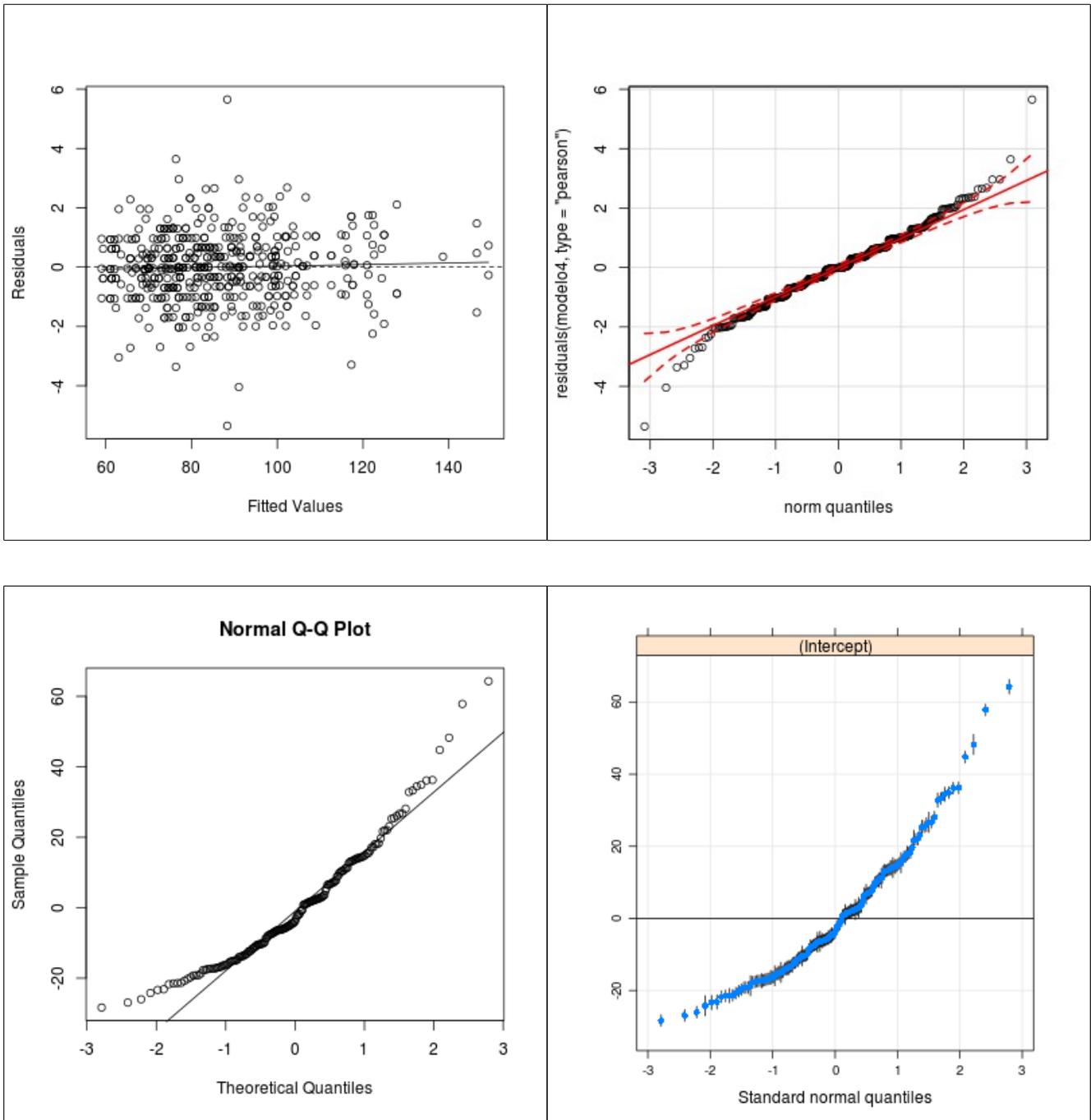


Figura 39: Diagnósticos para o modelo 4.

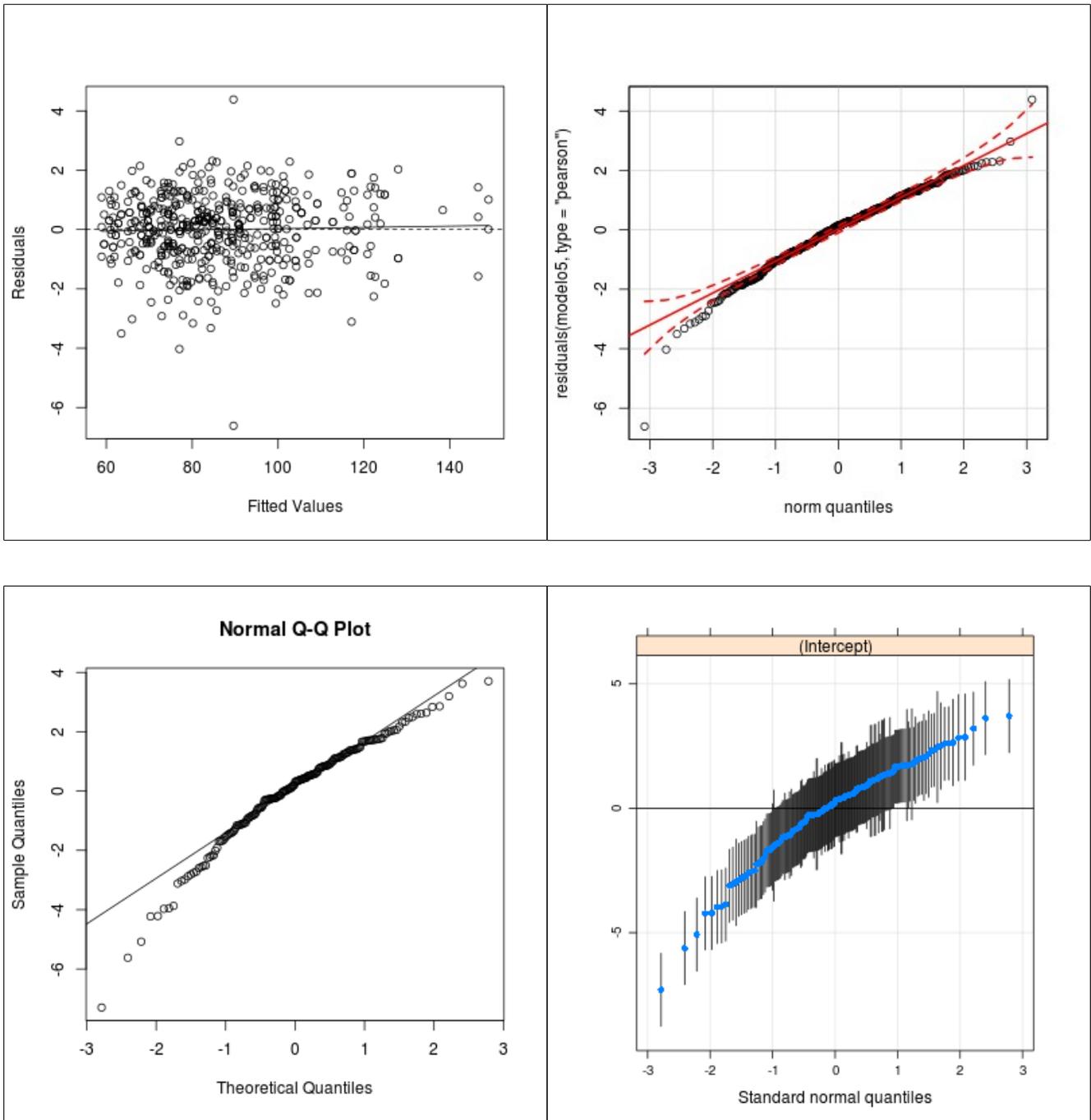


Figura 40: Diagnósticos para o modelo 5.

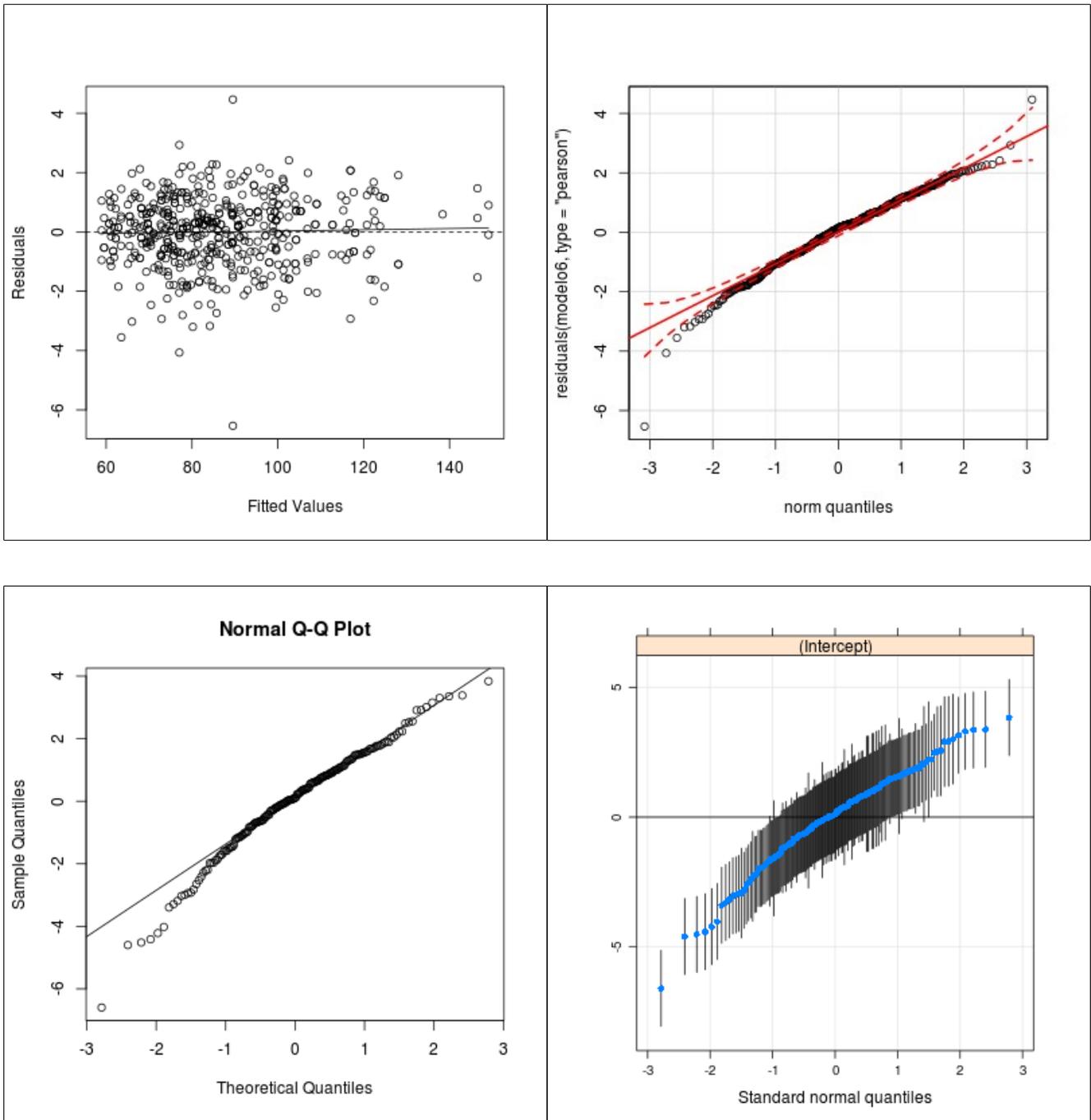


Figura 41: Diagnósticos para o modelo 6.

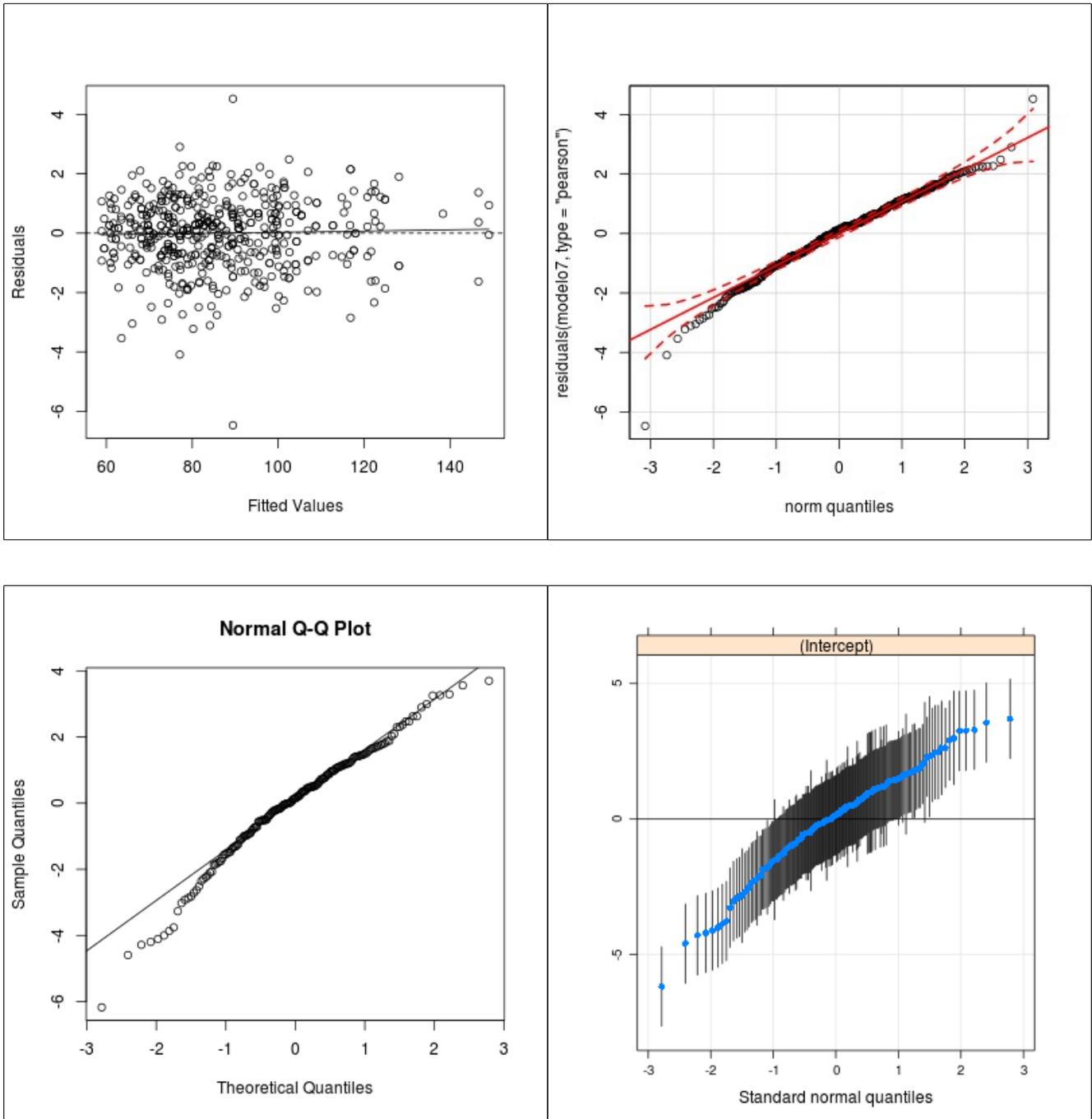


Figura 42: Diagnósticos para o modelo 7.

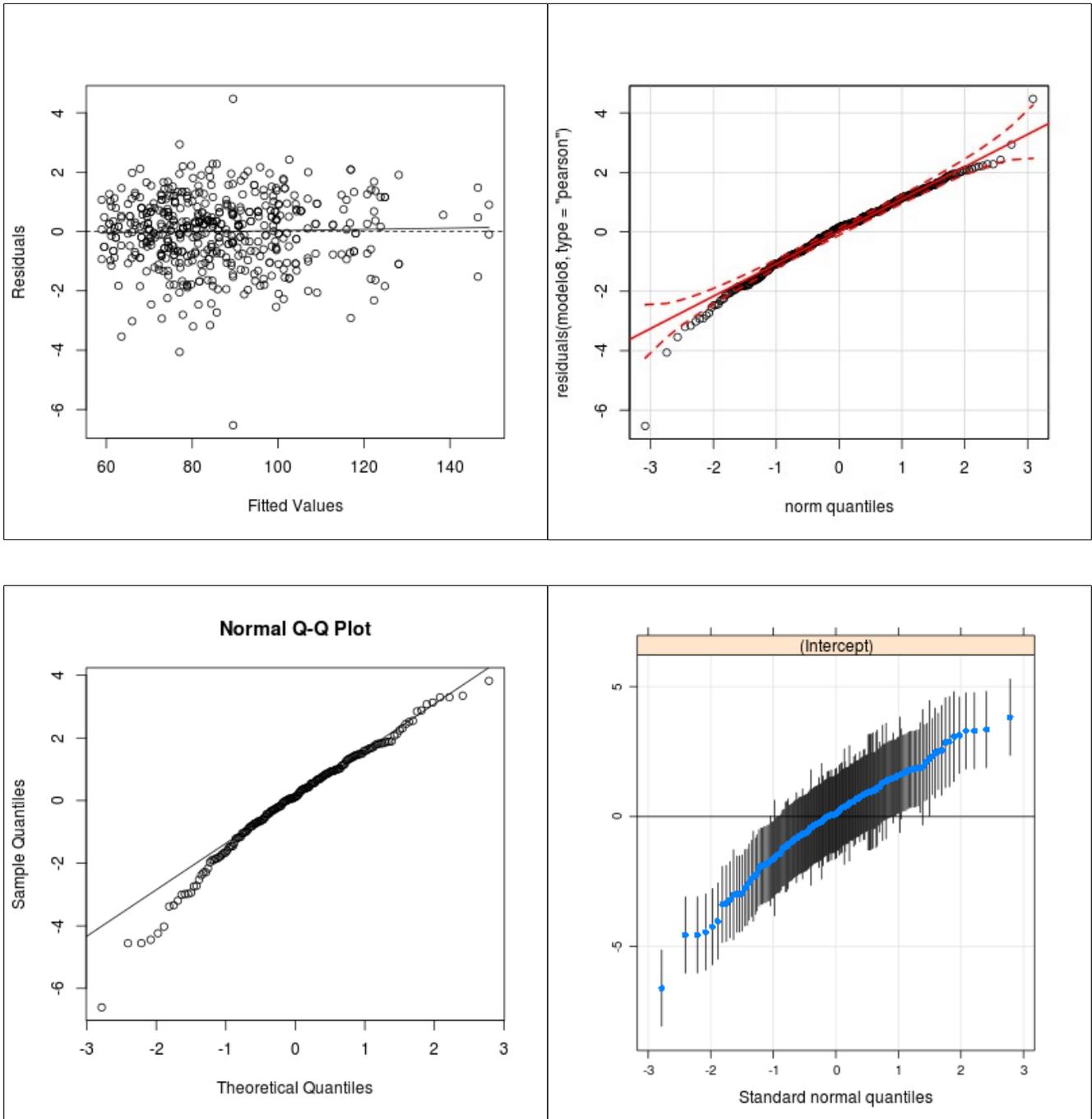


Figura 43: Diagnósticos para o modelo 8.

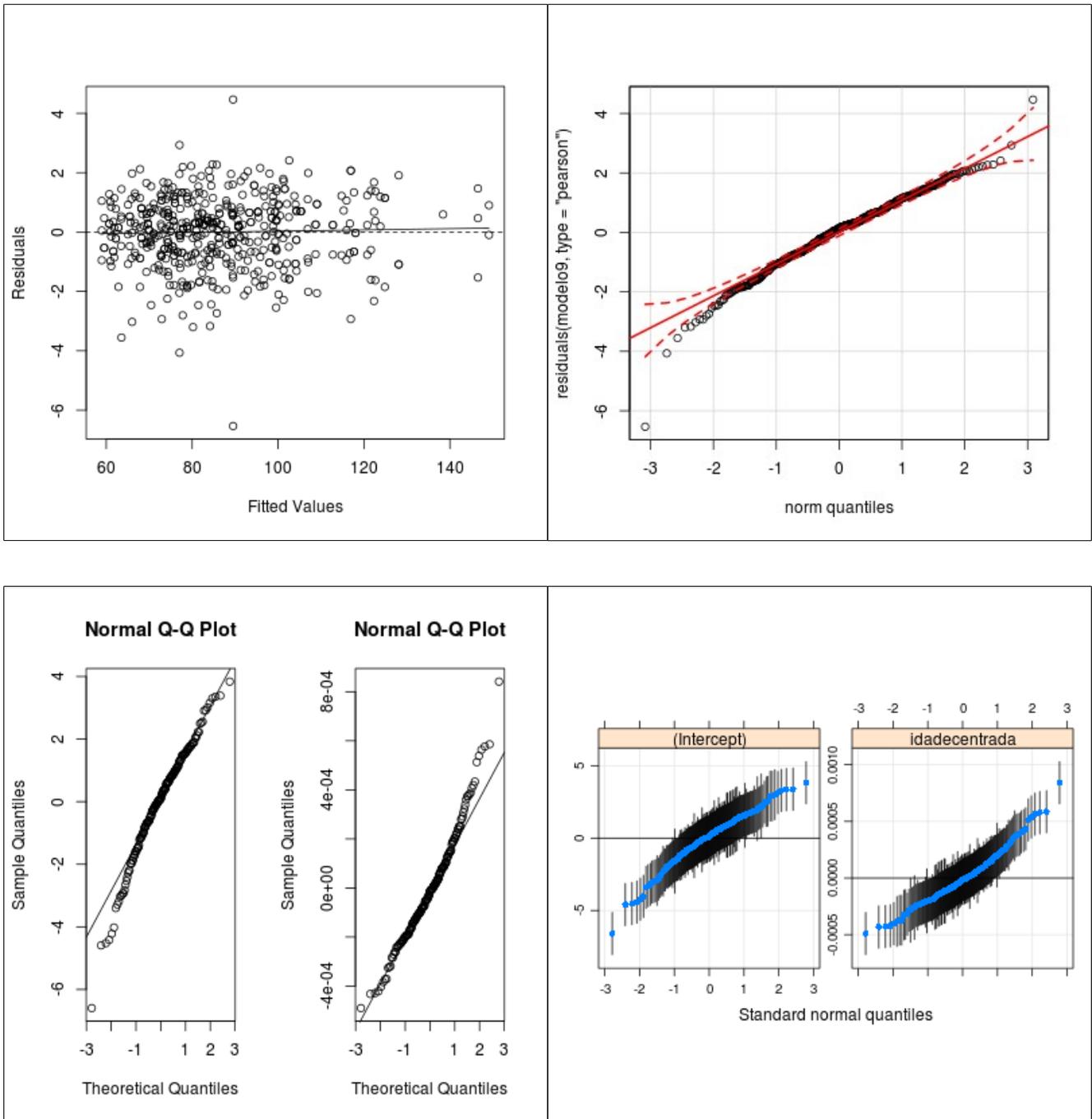


Figura 44: Diagnósticos para o modelo 9.

10.3 Apêndice A3: Gráficos dos modelos cuja variável resposta é a proporção de peso perdido

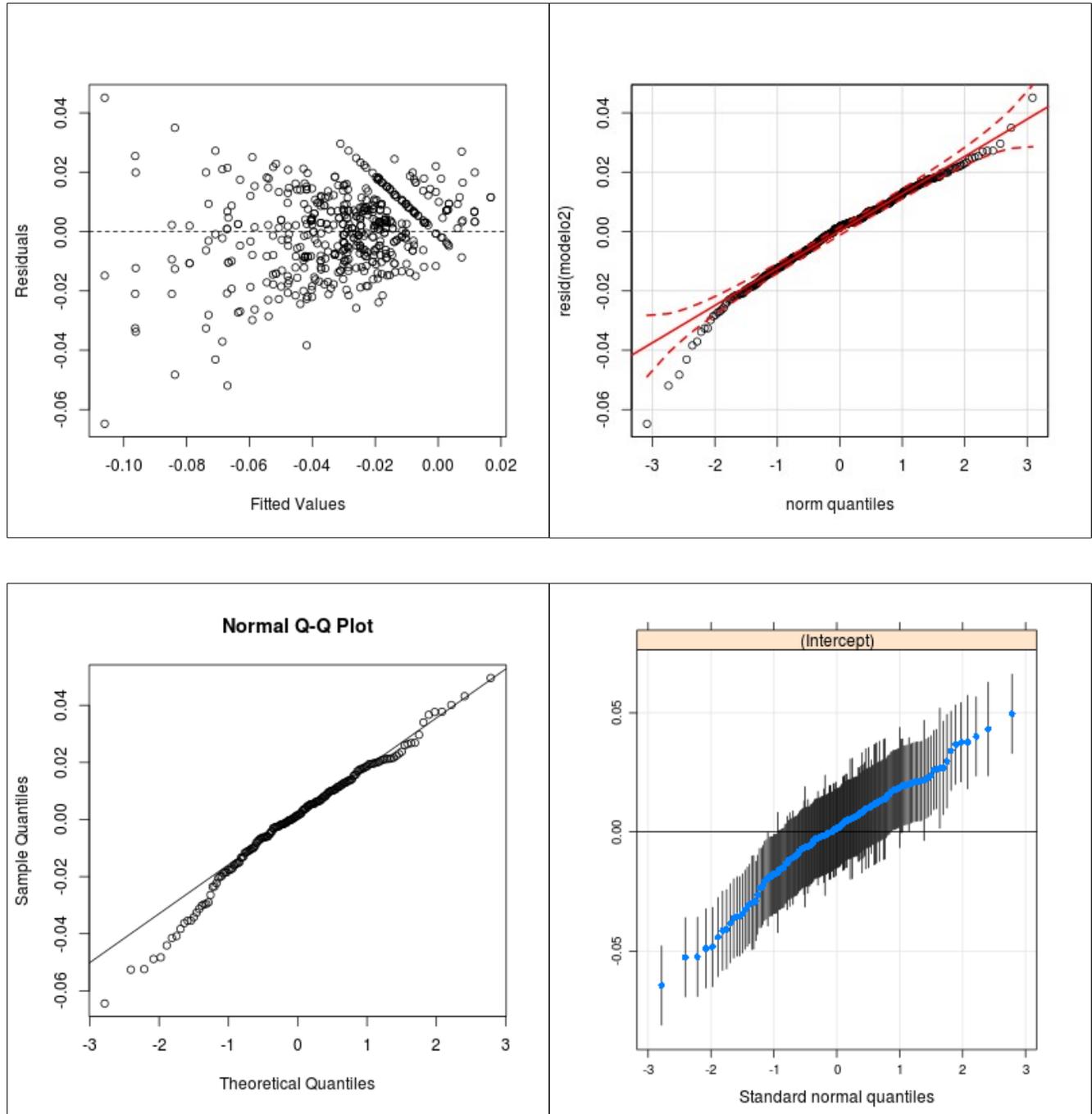


Figura 45: Diagnósticos para o modelo 2.

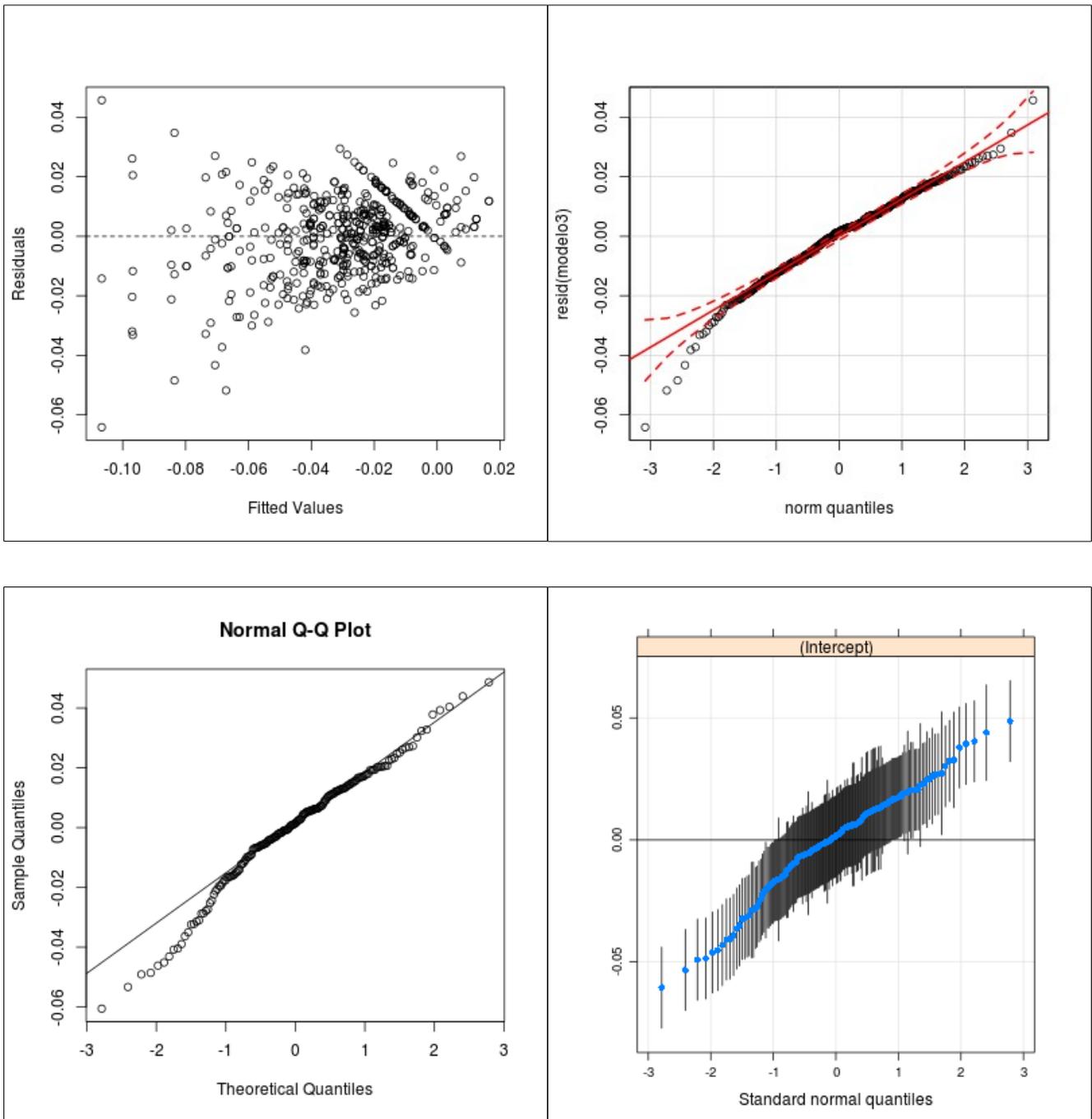


Figura 46: Diagnósticos para o modelo 3.

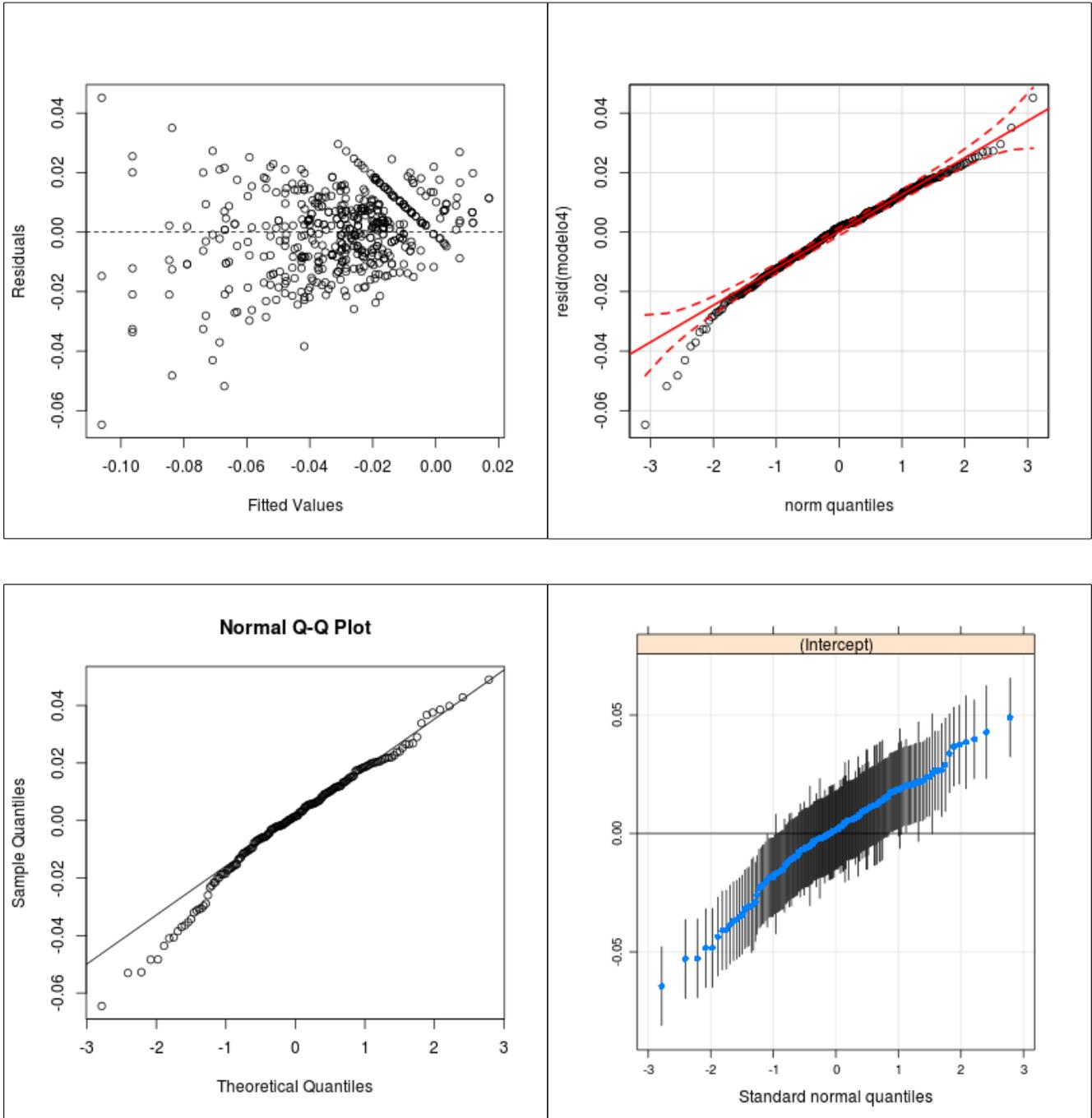


Figura 47: Diagnósticos para o modelo 4.

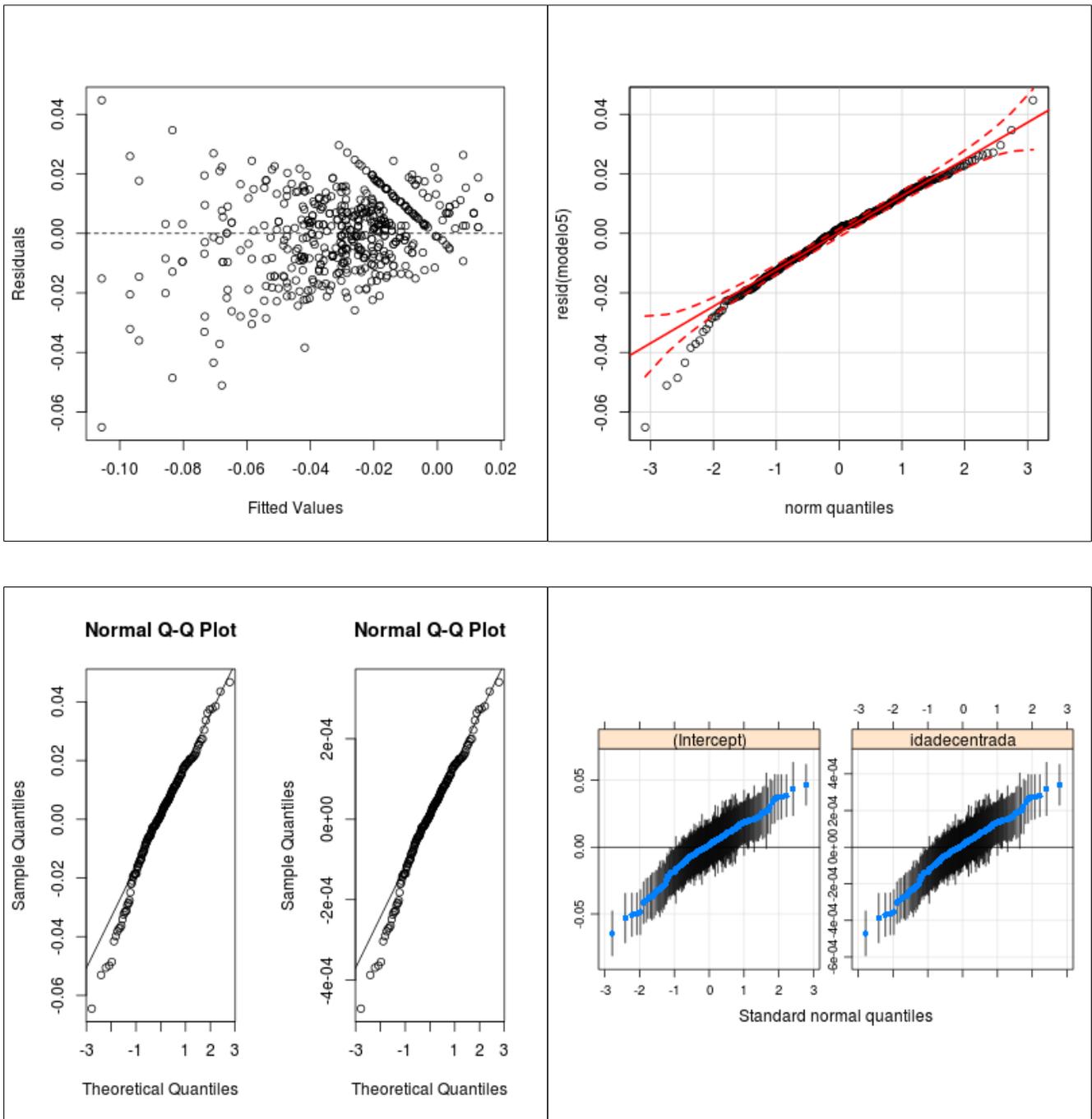


Figura 48: Diagnósticos para o modelo 5.

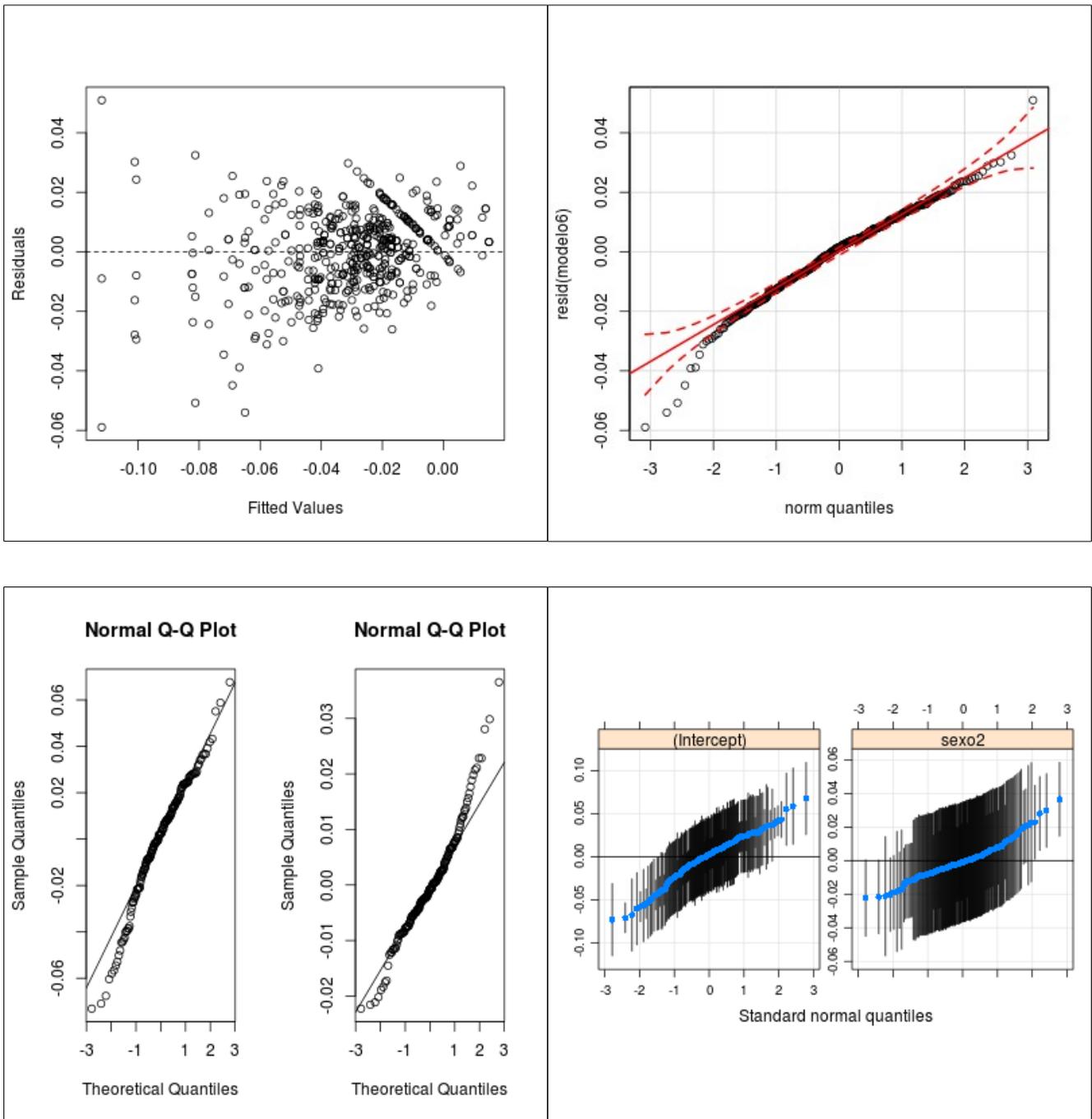


Figura 49: Diagnósticos para o modelo 6.

10.4 Apêndice A4: Gráficos dos modelos cuja variável resposta é o peso e interação tratamento-tempo

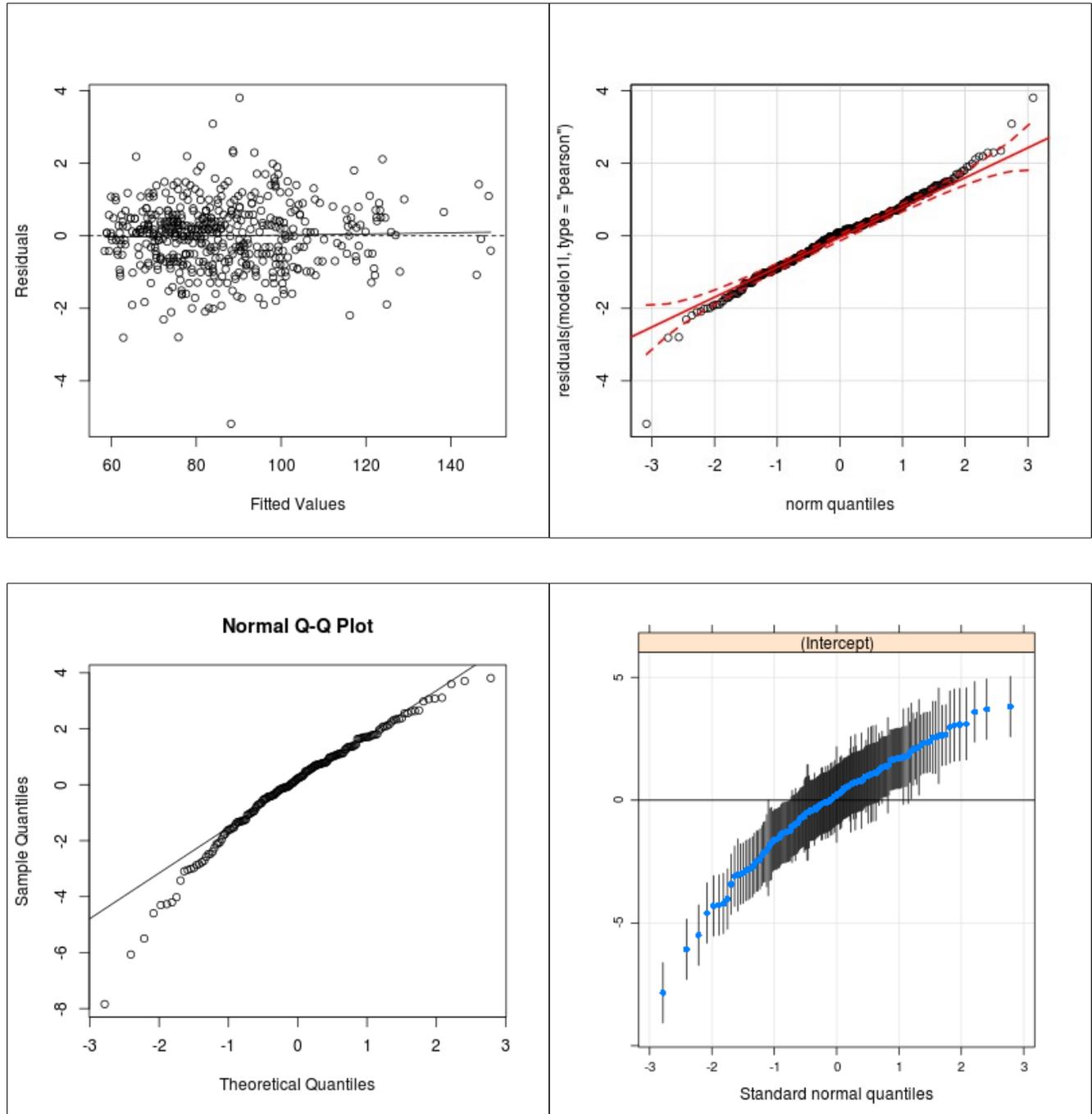


Figura 50: Diagnósticos para o modelo 1.

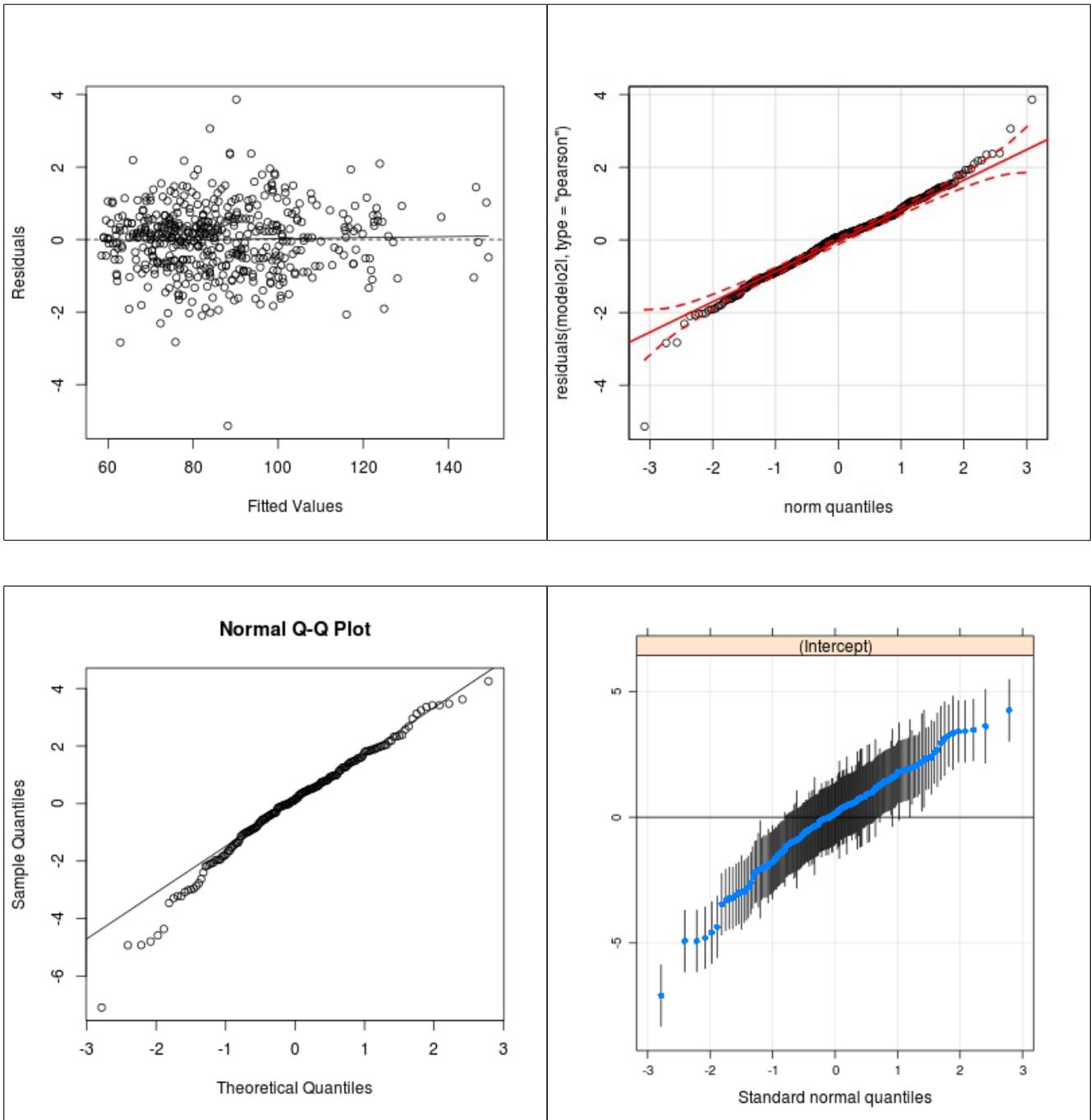


Figura 51: Diagnósticos para o modelo 2.

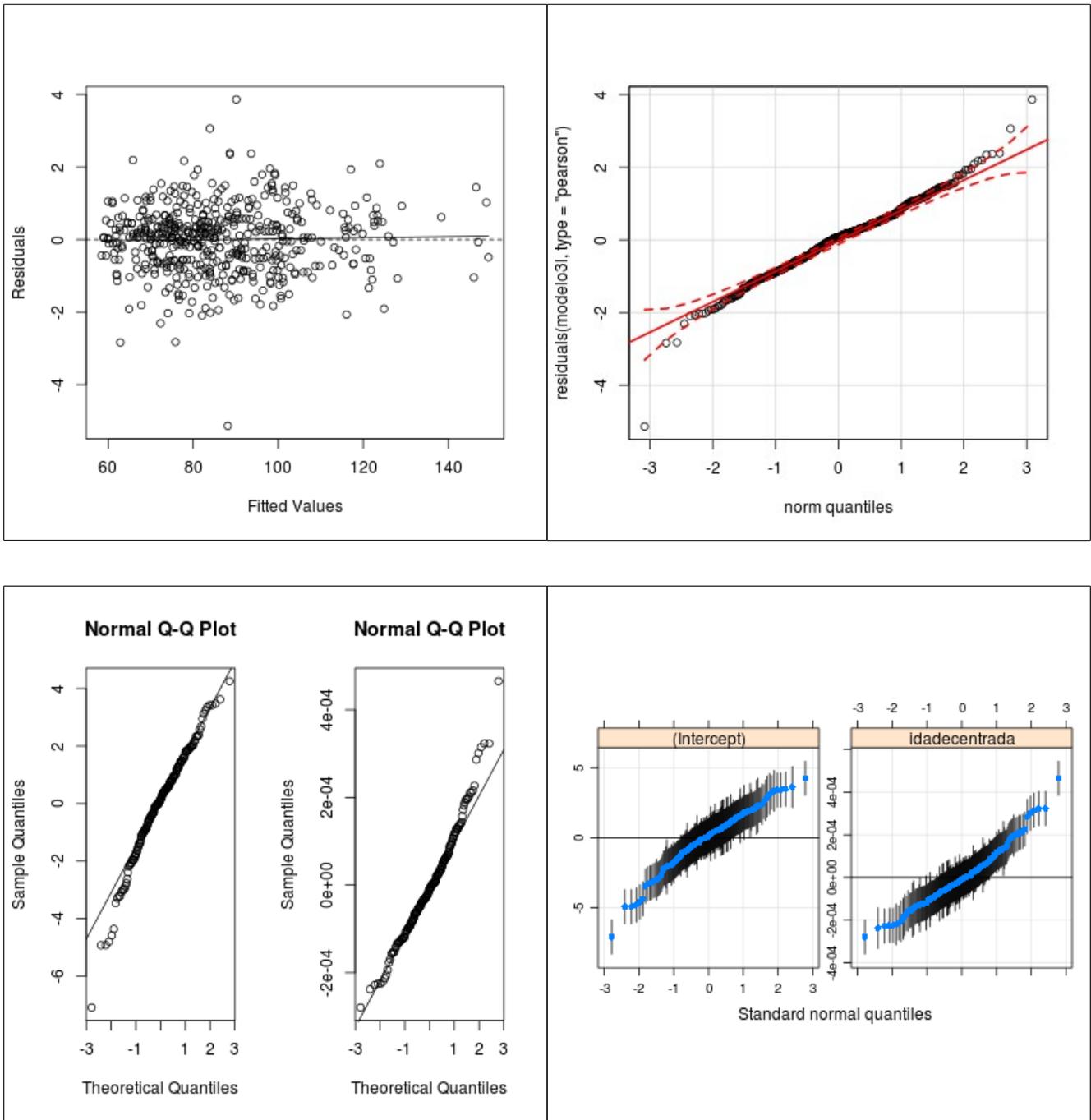


Figura 52: Diagnósticos para o modelo 3.