



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA

**ANÁLISE DE DADOS ANTROPOMÉTRICOS FACIAIS: UM ESTUDO
NÃO CASUALISADO**

BRUNNO KALYXTON SOUSA RAMOS

PEDRO LUIZ PINTO DE LIMA

Brasília

2014

BRUNNO KALYXTON SOUSA RAMOS

PEDRO LUIZ PINTO DE LIMA

ANÁLISE DE DADOS ANTROPOMÉTRICOS FACIAIS: UM ESTUDO NÃO CASUALISADO

Relatório apresentado à disciplina Estágio Supervisionado II do curso de graduação em Estatística, Instituto de Ciências Exatas, Universidade de Brasília, como parte de requisitos necessários para a obtenção do grau de Bacharel em Estatística.

Orientador: Prof. Dr. Lúcio José Vivaldi

Brasília

2014

DEDICATÓRIAS

Dedico este trabalho aos meus pais Eunice Torquato, que com arte de professores me ensinaram com muito amor e carinho desde ler e a escrever até agir com honestidade e gratidão em todos os momentos da minha vida, e Sinval Garcia, exemplo de dignidade. Às minhas irmãs Paty e Cláudia pilares de incentivo e carinho em todos os momentos. Aos meus sobrinhos Suzana, Guilherme e Larissa que com seus corações infantis sempre me fizeram ver o lado bom de qualquer momento adverso. À Carolina Andrade que do início ao fim deste projeto me brindou com sua inteligência, carinho e amor. A cada amizade verdadeira que já passou pela minha caminhada em particular ao Pedro que esteve sempre ao meu lado durante todo o curso.

Brunno Kalyton Sousa Ramos

Dedico este trabalho ao meu pai e à minha mãe, Geraldo e Clélia, por tudo que são e representam para mim, apoio e valores passados, ao meu irmão e irmã, Geraldo e Bel, pelo amor e carinho, ao meus tio Arnaldo e tia Bia, pela força e apoio, e a todos demais familiares. Dedico também a meus verdadeiros amigos e amigas. À Flora pelo amor e compreensão e ao Brunno pela grande parceria e amizade durante toda a graduação.

Pedro Luiz Pinto de Lima

AGRADECIMENTOS

Agradecemos, primeiramente, a Deus que, por meio de uma feliz combinação de eventos aleatórios, uniu as mentes de discípulos e mestre, tornando assim este trabalho possível.

Aos nossos pais e companheiros, que desde o primeiro passo na caminhada acadêmica nos assistiram e apoiaram com olhos de confiança, orgulho e esperança de que fizéssemos, com muito suor e esforço, a diferença, estivéssemos onde estiver, e assim o fizemos.

Aos nossos amigos, de pouca ou longa data, os que permaneceram e os que já partiram, pois sabemos que mesmo nas caminhadas mais árduas e sérias um sorriso e uma distração são essenciais para se recuperar o foco e as energias.

Ao nosso grande orientador, Prof. Dr. Lúcio José Vivaldi, por sua orientação, aulas, conselhos, dedicação e inestimável auxílio durante todo o trabalho. Somos privilegiados e nos sentimos honrados em ter a oportunidade de trabalhar tão próximos e com sua orientação.

Aos demais professores, funcionários e colegas do Departamento de Estatística da Universidade de Brasília, pela contribuição e apoio.

E por fim, porém não menos importante, a todos que acreditaram e principalmente aos que duvidaram deste trabalho, pois sabemos que a motivação dos verdadeiros vencedores é forjada nas adversidades.

RESUMO

O trabalho em questão foi originado de estudos de um policial federal. O objetivo deste policial com a pesquisa era desenvolver e averiguar ferramentas eficientes na identificação de pessoas por meio de fotografias utilizando a ciência da Antropometria.

Este trabalho teve como foco principal mostrar uma estratégia de como se conduzir a análise de um experimento não casualizado com dados repetidos no espaço, aplicado em um estudo de diferenciação de métodos de marcações antropométricas.

Tendo em vista o número pequeno de estudos com a problemática de análise em experimentos não casualizados, este trabalho expôs várias linhas de raciocínio e alternativas para conduzir a análise deste tipo de experimento por meio das técnicas de análise exploratória, delineamento de experimentos, modelos mistos, análises univariadas e multivariadas.

Iniciou-se com a análise exploratória observando a característica dos dados por meio de gráficos, tabelas de médias e variâncias. Logo após procurou-se realizar a modelagem do experimento utilizando a metodologia de parcelas divididas e, no fim deste tópico, a modelagem em modelos mistos. Também foi feita a análise univariada e multivariada, devido ao tipo de dados que estavam sendo analisados, e em caráter confirmatório a análise de normalidade das variáveis.

A análise exploratória inicial indicou pequenas diferenças entre os métodos de marcação estudados, quando analisados mais profundamente em modelos mistos notou-se que existia diferença estatisticamente significativa. As análises univariadas e multivariadas também confirmaram a diferença significativa entre os métodos.

LISTA DE FIGURAS

Figura 1: Marcações Antropométricas na face humana.....	13
Figura 2: Diferença entre métodos na coordenada Y para o indivíduo 1	19
Figura 3: Diferença entre métodos na coordenada Y para os indivíduos 2 ao 5	20
Figura 4: Diferença entre métodos na coordenada Y para os indivíduos 6 ao 9	21
Figura 5: Diferença entre métodos na coordenada Y para os indivíduos 10 ao 13.....	22
Figura 6: Diferença entre métodos na coordenada Y para os indivíduos 14 ao 17.....	23
Figura 7: Diferença entre métodos na coordenada Y para o indivíduo 18.....	24
Figura 8: Gráficos de Andrews para os indivíduos 1 ao 6	25
Figura 9: Gráficos de Andrews para os indivíduos 7 ao 14.....	26
Figura 10: Gráficos de Andrews para os indivíduos 15 ao 18.....	27
Figura 11: Distâncias entre métodos nos locais 1 ao 4.....	29
Figura 12: Distâncias entre métodos nos locais 5 ao 8.....	30
Figura 13: Distâncias entre métodos nos locais 9 ao 12	30
Figura 14: Distâncias entre métodos nos locais 13 ao 16	31
Figura 15: Distâncias entre métodos nos locais 17 ao 20	31
Figura 16: Distâncias entre métodos nos locais 21 ao 24	32
Figura 17: Locais onde a as distâncias médias foi muito grande	46
Figura 18: Análise de Normalidade do ponto Endocanthion Esquerdo	58
Figura 19: Análise de Normalidade do ponto Glabela.....	58
Figura 20: Análise de Normalidade do ponto Iridion Medial Esquerdo.....	58
Figura 21: Análise de Normalidade do Ponto Labial Inferior	59
Figura 22: Análise de Normalidade do Ponto Labial Superior	59
Figura 23: Análise de Normalidade do Ponto Nasion.....	59
Figura 24: Análise de Normalidade do Ponto Stomion.....	60
Figura 25: Análise de Normalidade do Ponto Subnasal.....	60
Figura 26: Análise de Normalidade do Ponto Zygion Esquerdo	60
Figura 27: Análise de Normalidade do ponto Iridion Medial Esquerdo.....	61
Figura 28: Análise de Normalidade para o ponto Labial Inferior.....	61
Figura 29: Análise de normalidade para o Ponto Nasion.....	62
Figura 30: Análise de Normalidade para o ponto Stomion	62
Figura 31: Análise de Normalidade para o ponto Subnasal	62
Figura 32: Análise de Normalidade para o ponto Endocanthion Direito.....	63
Figura 33: Análise de Normalidade para o ponto Endocanthion Esquerdo	63
Figura 34: Análise de Normalidade para o ponto Iridion Medial Esquerdo.....	64
Figura 35: Análise de Normalidade para o ponto Labial Inferior.....	64
Figura 36: Análise de Normalidade para o ponto Stomion	64
Figura 37: Análise de Normalidade para o ponto Subnasal	65
Figura 38: Locais onde a normalidade foi prejudicada por um outlier para distância	66

LISTA DE TABELAS

Tabela 1: Médias de X e Y por examinadores.....	15
Tabela 2: Médias de X e Y por indivíduos.....	16
Tabela 3: Variâncias de X e Y por examinadores.....	17
Tabela 4: Variâncias de X e Y por indivíduos.....	18
Tabela 5: Média das Distâncias para cada local.....	28
Tabela 6: Anova para cada local de Y e X.....	33
Tabela 7: Diferenças significativas pelo teste F para todos os pontos.....	34
Tabela 8: Variância de cada local.....	35
Tabela 9: ANOVA para distancia entre métodos.....	35
Tabela 10: Resumo da ANOVA exploratória das distancias entre métodos.....	36
Tabela 11: ANOVA de Delineamento em parcelas divididas para a coordenada Y.....	40
Tabela 12: ANOVA de Delineamento em parcelas divididas para a coordenada X.....	40
Tabela 13: ANOVA com aproximação HF para a coordenada Y.....	41
Tabela 14: ANOVA com aproximação HF para a coordenada X.....	41
Tabela 15: ANOVA Delineamento em parcelas divididas para a distancia entre métodos.....	42
Tabela 16: ANOVA com aproximação HF para a distancia entre métodos.....	42
Tabela 17: Matriz de Variâncias e Covariâncias para locais.....	43
Tabela 18: Teste Type 3 para efeitos fixos.....	44
Tabela 19: Distância média para cada examinador em cada local.....	45
Tabela 20: ANOVA Multivariada.....	49
Tabela 21: Teste para os efeitos dos Examinadores em Y.....	52
Tabela 22: Testes para os efeitos dos Métodos em Y.....	52
Tabela 23: Testes para a interação Examinador*Método em Y.....	52
Tabela 24: Testes para os efeitos dos Locais em Y.....	53
Tabela 25: Testes para a interação Local*Examinador em Y.....	53
Tabela 26: Teste para a interação Local*Métodos em Y.....	53
Tabela 27: Testes para a interação Local*Examinador*Método em Y.....	53
Tabela 28: Testes para o efeito dos Examinadores em X.....	53
Tabela 29: Testes para o efeito de Métodos em X.....	54
Tabela 30: Testes para a interação Examinador*Método em X.....	54
Tabela 31: Testes para os efeitos dos Locais em X.....	54
Tabela 32: Testes para a interação Local*Examinadores em X.....	54
Tabela 33: Testes para a interação Local*Métodos em X.....	54
Tabela 34: Testes para a interação Local*Examinador*Métodos em X.....	55
Tabela 35: Testes para o efeito dos Examinadores na distancia.....	55
Tabela 36: Testes para o efeito dos Locais na distancia.....	55
Tabela 37: Testes para a interação Local*Examinadores na distancia.....	55

Sumário

1.	INTRODUÇÃO	8
2.	NATUREZA DOS DADOS.....	10
3.	ANÁLISE DESCRITIVA DOS DADOS.....	14
3.1.	CARACTERÍSTICAS DOS DADOS	14
3.2.	MÉDIAS E VARIÂNCIAS	15
3.3.	GRÁFICOS	19
3.4.	DISTÂNCIAS	27
3.5.	ANOVA EXPLORATÓRIA.....	33
3.6.	CONCLUSÃO BASEADA NA ANÁLISE DESCRITIVA.....	37
4.	ANÁLISE DE EXPERIMENTOS	38
4.1.	EXPERIMENTOS CASUALIZADOS	38
4.2.	MODELOS	39
4.3.	CONCLUSÃO BASEADA NOS MODELOS.....	47
5.	ANÁLISE MULTIVARIADA.....	48
5.1.	INTRODUÇÃO	48
5.2.	INFERÊNCIA EM ANÁLISE MULTIVARIADA	50
6.	TESTE PARA NORMALIDADE.....	57
6.1.	Normalidade em Y.....	57
6.2.	Normalidade em X.....	61
6.3.	Normalidade na distância entre pontos.....	63
7.	Conclusão	70
8.	Referências Bibliográficas	73
9.	Apêndice.....	75

1. INTRODUÇÃO

Este trabalho tem como objetivo apresentar uma análise dos dados de um estudo comparativo entre dois métodos de identificações antropométricas faciais, os quais são usados para a identificação de suspeitos pela Polícia Federal através do retrato falado. Trata-se de um estudo observacional estruturado com dados repetidos no espaço. Os dados analisados foram gerados por cinco examinadores (peritos em desenho de identificação facial) cada um dos quais marcou vinte e quatro pontos antropométricos faciais, que caracterizam o rosto humano, a partir de dezoito fotos de faces de dezoito indivíduos distintos, utilizando as duas técnicas de marcação em questão (tradicional e alternativa). Como são dados pouco conhecidos vamos descrever resumidamente o processo de origem.

A antropometria, ciência que estuda as medidas de tamanho, peso e proporções do corpo humano, fornece dados objetivos na avaliação da morfologia craniofacial, mediante uma série de medidas da cabeça e da face. Além disto, oferece inúmeras vantagens sobre outros métodos de avaliação da morfologia do complexo craniofacial por utilizar técnicas simples, não invasivas, sem risco para o sujeito e com baixo custo, além de fornecer dados que podem ser comparados, uma vez que foi estabelecido o padrão da normalidade de medidas faciais para a população brasileira. Diferenças nas proporções faciais são demonstradas em estudos com populações compostas de diversas raças e etnias. Na Antropometria direta, a medida é obtida diretamente do sujeito por meio de paquímetros ou fita métrica. Na Antropometria indireta, as medidas são coletadas mediante fotografia, cefalometria do perfil do tecido mole e imagens computadorizadas da superfície craniofacial. Todos esses métodos têm suas vantagens e desvantagens e devem envolver três elementos básicos do exame: localização das marcas craniofaciais, execução das medidas, avaliação dos achados e comparação com os dados da normalidade.

No estudo em questão será utilizada a antropometria indireta. Para a primeira etapa, o pesquisador selecionou cinco examinadores com experiência em cefalometria e pediu para que marcassem os seguintes 24 pontos faciais: Glabella, Nasion, Subnasal, Labial Superior, Stomion, Labial Inferior, Labiamental, Gnathion,

Zygion Direito, Zygion Esquerdo, Gonio Direito, Gonio Esquerdo, Ectoanthion Direito, Ectoanthion Esquerdo, Iridion Medial Direito, Iridion Medial Esquerdo, Iridion Lateral Direito, Iridion Lateral Esquerdo, Alar Direito, Alar Esquerdo, Chelion Direito e Chelion Esquerdo, no capítulo 2, este tópico será revisto.

Uma reflexão posterior qualifica o estudo como um experimento não casualizado, com dados repetidos no espaço e neste caso, várias etapas são necessárias para a análise dos dados. Num primeiro momento foram aplicadas as técnicas elementares como a análise de gráficos e estatísticas exploratórias, as quais se resumem aqui em médias, variâncias e desvios padrões de cada ponto antropométrico analisado, utilizando assim o suporte gráfico para melhor visualizar a configuração dos dados. Como orientação para analisar os dados, seguimos em boa parte, o livro de Hoaglin, Mosteller e Tukey(1991), que contém procedimentos recomendados para analisar dados como estes.

Embora sem um delineamento fixado, no capítulo 4, alguns modelos foram utilizados para comparar os métodos em três perspectivas diferentes (eixo horizontal da foto, eixo vertical da foto e distância entre as marcações do mesmo ponto pelas duas metodologias). Em um outro nível inferencial, a análise univariada irá expor medidas de variância e covariância. Num outro capítulo, utilizamos também modelos de análise de variância multivariada para de uma forma global cobrir todas as inferências. Como não houve casualização, foram aplicadas técnicas para detectar não normalidade dos dados.

2. NATUREZA DOS DADOS

A antropometria, do grego, *anthropos* que significa homem e *metrikos* que significa justa proporção, é a “ciência que estuda e avalia as medidas de tamanho, massa e proporções do corpo humano.” (Martins, 2003). Na antropometria são aferidas medidas como massa, estatura, perímetros, dobras cutanêas, diâmetros e distâncias de pontos específicos nos ossos e órgãos.

Há indícios de que aferições antropométricas eram feitas desde a antiguidade grega e egípcia. Atualmente, seu uso é fundamental no acompanhamento do desenvolvimento infanto-juvenil pela puericultura (ciência médica que se dedica ao estudo dos cuidados com o ser humano em desenvolvimento, mais especificamente com o acompanhamento do desenvolvimento infantil), pela ortopedia (ciência médica que se dedica ao tratamento das doenças e deformidades dos ossos, músculos, ligamentos e articulações), pela traumatologia (ciência médica que lida com os possíveis traumas do aparelho músculo-esquelético) e pelos órgãos de segurança pública.

Formado pelas estruturas ósseas, musculares e tecidos moles, cada rosto apresenta seu próprio conjunto de características que formam o perfil do indivíduo. Sexo, raça, genética, ambiente e acidentes sua formação e configuração.

O objeto de pesquisa em questão foi originado de estudos de um policial federal. O objetivo deste policial com esta pesquisa é desenvolver e averiguar ferramentas eficientes na identificação de pessoas por meio de fotografias utilizando a ciência da Antropometria.

As medidas podem ser aferidas de maneira direta ou indireta. Na antropometria direta, a medida é obtida diretamente do sujeito por meio de paquímetros e fita métrica. Na Antropometria indireta, as medidas são coletadas mediante fotografia, cefalometria do perfil do tecido mole e imagens computadorizadas da superfície craniofacial. Todos esses métodos têm suas vantagens e desvantagens e devem envolver três elementos básicos do exame: localização das marcas craniofaciais, execução das medidas, avaliação dos achados e comparação com os dados da normalidade.

Neste estudo serão utilizadas fotos de rostos de dezoito indivíduos diferentes e a antropometria para aferir vinte e quatro pontos que caracterizam o rosto dos indivíduos. O estudo como um todo envolve várias etapas, tais como a análise de estatísticas exploratórias, análise de modelos de delineamento e a análise univariada e multivariada.

Será utilizada a antropometria indireta para a primeira etapa, o pesquisador selecionou cinco examinadores com experiência em cefalometria e pediu para que marcassem os 24 pontos faciais a partir de imagens de 18 faces escolhidas anteriormente. As medidas para cada examinador em cada imagem será observada de três formas distintas, são elas: a distância, em pixels, da borda esquerda da imagem até o ponto selecionado pelo examinador (análise em X); a distância, em pixels, da borda inferior da imagem até o ponto selecionado (análise em Y); e a distância das coordenadas do ponto selecionado até o ponto médio entre os examinadores (análise espacial).

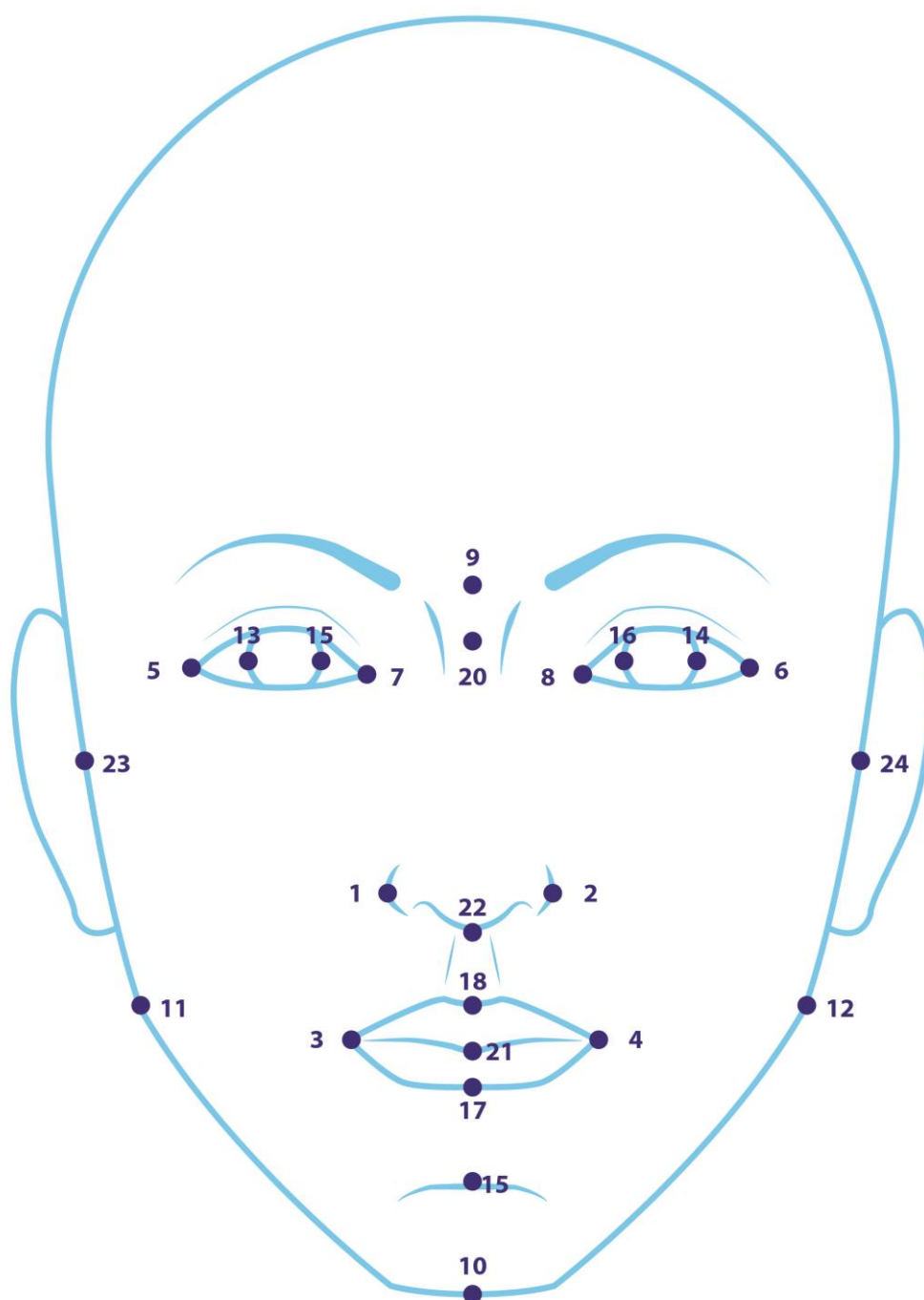
Os pontos escolhidos foram determinados para a realização do retrato falado, de maneira que seja a mais precisa possível. Os pontos utilizados neste estudo serão:

1. ALAR DIREITO: ponto mais externo ao canto da narina direita.
2. ALAR ESQUERDO: ponto mais externo ao canto da narina esquerda.
3. CHELION DIREITO: ponto de encontro dos lábios superior e inferior, no canto da boca, ao lado direito.
4. CHELION ESQUERDO: ponto de encontro dos lábios superior e inferior, no canto da boca, ao lado esquerdo.
5. ECTOANTHION DIREITO: ponto de encontro das palpebras, próximo à tempora, no olho direito.
6. ECTOANTHION ESQUERDO: ponto de encontro das palpebras, próximo à tempora, no olho esquerdo.
7. ENDOCANTHION DIREITO: ponto de encontro das palpebras, próximo ao nariz, no olho direito.
8. ENDOCANTHION ESQUERDO: ponto de encontro das palpebras, próximo ao nariz, no olho esquerdo.
9. GLABELA: ponto medial mais proeminente na fronte, entre as sobrancelhas.

10. GNATHION: ponto no bordo anterior da mandíbula que mais se projeta para baixo, ponta do queixo)
11. GONIO DIREITO: ponto de vértice do ângulo do maxilar inferior à direita.
12. GONIO ESQUERDO: ponto de vértice do ângulo do maxilar inferior à esquerda.
13. IRIDION LATERAL DIREITO: ponto mais lateral da borda da íris do olho direito.
14. IRIDION LATERAL ESQUERDO: ponto mais lateral da borda da íris do olho esquerdo.
15. IRIDION MEDIAL DIREITO: ponto mais medial na borda da íris do olho direito.
16. IRIDION MEDIAL ESQUERDO: ponto mais medial na borda da íris do olho esquerdo.
17. LABIAL INFERIOR: ponto médio da borda inferior do lábio inferior.
18. LABIAL SUPERIOR: ponto médio da borda superior do lábio superior.
19. LABIOMENTAL: ponto de dobra acima do queixo, que inicia o queixo.
20. NASION: ponto da sutura naso-frontal, imediatamente inferior à glabella.
21. STOMION: ponto da dobra ao centro entre os lábios.
22. SUBNASAL: ponto da interseção inferior da columella com o rosto, logo abaixo do nariz.
23. ZYGION DIREITO: ponto mais lateral da maçã do rosto ao lado direito.
24. ZYGION ESQUERDO: ponto mais lateral da maçã do rosto ao lado esquerdo.

As Numerações utilizadas acima ao lado de cada nomenclatura dos pontos serão utilizadas em alguns momentos deste trabalho afim de facilitar a apresentação de tabelas e gráficos.

Figura 1: Marcações Antropométricas na face humana



3. ANÁLISE DESCRITIVA DOS DADOS

3.1. CARACTERÍSTICAS DOS DADOS

Antes de se iniciar uma análise descritiva de um conjunto de dados, deve-se definir como eles são; embora sua forma, objetivos do estudo, etc., já tenham sido comentados, nesta seção vamos ampliar esta caracterização.

Um conjunto de dados pode surgir de pelo menos de três origens: por amostragem, por um estudo observacional ou por um experimento. Dados provenientes de um estudo por amostragem são obtidos de numa população sem a interferência do pesquisador, mas de acordo com um modelo probabilístico que define quais elementos da população pertencerão à amostra; assim a estrutura de variâncias e covariâncias é conhecida e pode ser usada para a análise dos dados, entretanto, não fica definido qual ou quais modelos são os adequados. Num estudo observacional os dados são obtidos de uma população mas sem um modelo probabilístico que indique quais elementos vão fazer parte do estudo; nesta categoria estão os estudos feitos com voluntários, típicos na área de saúde; pesquisas florestais conduzidas em parte de uma floresta ou reserva. Neste caso não há um conhecimento da estrutura de variâncias e covariâncias nem definição de modelos. Dados que vem de um experimento são aqueles que foram gerados por um sistema construído pelo pesquisador, no qual foi aplicado um estímulo ou uma interferência, chamados de tratamentos.

Um outro qualificativo utilizado com frequência em análise de dados são dados estruturados. Um conjunto de dados tem uma estrutura quando ele pode ser dividido em classes: gênero, cidade, idade, etc. É evidente que um experimento gera sempre um conjunto de dados estruturados, mas tanto em amostragem com em estudos observacionais os dados também podem ser estruturados. A existência de uma estrutura possibilita uma análise mais detalhada, mediante modelos.

Diante do que foi dito, os dados do estudo em questão foram gerados por um experimento e portanto são estruturados, entretanto, não houve casualização. Este ponto será desenvolvido no próximo capítulo. O objetivo principal do estudo do

pesquisador foi testar a diferença entre dois métodos mas nosso interesse aqui vai também em direção a como analisar um experimento desta magnitude que não foi casualizado, com duas variáveis resposta e com dados repetidos. Para um conhecimento geral vamos conduzir primeiramente uma análise descritiva, seguindo os desenvolvimento recomendados por Hoaglin, Mosteller e Tukey(1991), autores do livro “*Fundamentals of Exploratory Analysis Of Variance*”. A proposta dos autores é, mediante tabelas de médias, gráficos e ANOVA de um conjunto de dados estruturados, desenvolver uma conclusão preliminar.

3.2. MÉDIAS E VARIÂNCIAS

Em estatística, uma análise exploratória da dados estruturados consiste na apresentação de medidas básicas, tais quais média, variância e desvio padrão. O volume de dados é grande e parte das tabelas estão no apêndice, entretanto, algumas mais importantes serão apresentadas adiante. Para o cálculo das medidas descritivas os valores das coordenadas X e Y foram divididos por 10 a fim de facilitar a visualização dos valores obtidos.

Seguem as tabelas das médias avaliadas primeiramente para cada um dos examinadores de acordo com a metodologia e em seguida para cada um dos indivíduos também de acordo com cada metodologia:

Tabela 1: Médias de X e Y por examinadores

EXA	X			Y		
	MET		MEDIA	MET		MEDIA
	1	2		1	2	
1	84.681	84.943	84.812	62.090	62.092	62.091
2	84.767	84.990	84.879	62.122	62.125	62.124
3	84.951	84.949	84.950	62.141	62.063	62.102
4	84.573	84.934	84.753	62.027	62.095	62.061
5	84.998	84.909	84.953	62.113	62.075	62.094
MEDIA	84.794	84.945	84.870	62.099	62.090	62.094

Tabela 2: Médias de X e Y por indivíduos

IND	X			Y		
	MET		MEDIA	MET		MEDIA
	1	2		1	2	
1	83.1	83.2	83.1	63.6	63.6	63.6
2	84.6	85.1	84.8	61.3	61.3	61.3
3	81.7	81.8	81.8	62.8	62.9	62.9
4	84.4	84.5	84.4	61	61	61
5	79.7	79.8	79.8	59.1	59	59.1
6	84.5	84.9	84.7	61	61	61
7	87.1	87	87	59.5	59.5	59.5
8	87.2	87.6	87.4	62.6	62.5	62.6
9	86.3	86.6	86.4	59.9	59.9	59.9
10	88.3	88.2	88.2	63.2	63.3	63.2
11	85.1	85.4	85.2	62.3	62.4	62.4
12	83.4	83.2	83.3	66.9	66.9	66.9
13	82.6	82.7	82.6	58.7	58.7	58.7
14	81.7	81.8	81.7	67	67	67
15	86.2	86.4	86.3	63.8	63.8	63.8
16	88.2	88.4	88.3	61.1	61.1	61.1
17	85.4	85.8	85.6	57.5	57.4	57.5
18	86.7	86.9	86.8	66.4	66.5	66.4
MEDIA	84.8	84.9	84.9	62.1	62.1	62.1

A partir da observação das médias gerais observa-se na Tabela 1 que as diferenças entre examinadores é muito pequena, variando em torno de 84.8 para X e em torno de 62.1 para Y. A variação entre métodos aparenta ser menor ainda. As médias gerais baseadas em 24 locais dos indivíduos apresentam uma maior variação mas que é esperada.

Entre locais, as médias tanto para examinadores quanto para métodos, apresentam maior variabilidade mas isto também é esperado de local para local. Também não há indicação de interação de locais com examinadores e com métodos, o que também é esperado .

Outra medida exploratória importante em um estudo exploratório é a variância, que consiste em uma medida de dispersão indicando quão longe os valores observados se encontram de um suposto valor esperado.

A seguir encontram-se as tabelas de variância por examinador e por indivíduos comparando as duas metodologias:

Tabela 3: Variâncias de X e Y por examinadores

EXA	X			Y		
	MET		TOTAL	MET		TOTAL
	1	2		1	2	
	Var	Var	Var	Var	Var	Var
1	220.7	226	223.1	184.3	185	184.4
2	219.7	223.7	221.5	184.3	184.9	184.4
3	229.4	224.9	226.9	176	182.9	179.3
4	229.9	225.8	227.6	179.6	185.4	182.3
5	215.6	224.9	220	185.7	184.6	184.9
TOTAL	222.6	224.6	223.6	181.6	184.2	182.9

Tabela 4: Variâncias de X e Y por indivíduos

IND	X			Y		
	MET		TOTAL	MET		TOTAL
	1	2		1	2	
	Var	Var	Var	Var	Var	Var
1	249	257.2	252.1	180.9	181.9	180.6
2	251	246	247.5	197.9	204	200.1
3	230.1	223.8	226	174.4	178	175.5
4	236.3	241.5	237.9	167.2	168	166.9
5	267.1	280.1	272.5	199.5	198.4	198.1
6	242.1	234.8	237.5	175.2	182.9	178.3
7	215	219.9	216.5	174.2	173.2	172.9
8	225.6	228	225.9	172.4	175.7	173.3
9	178.5	178.7	177.9	164.9	169.6	166.5
10	199.2	197.3	197.4	147.6	148.1	147.2
11	169.8	168.3	168.4	143.9	148.4	145.5
12	241.3	248.6	243.9	187.5	184.3	185.1
13	185.4	191.2	187.5	167.4	170.2	168.1
14	232.1	238.8	234.5	185.2	186.4	185
15	217.1	224	219.6	182.6	182.1	181.6
16	195.5	190.1	192	182.3	186.7	183.7
17	221.1	216.7	218	206.6	212.7	208.8
18	185.3	188.9	186.3	154.9	157.4	155.5
TOTAL	222.6	224.6	223.6	181.6	184.2	182.9

Assim como na análise das médias, a variância baseada em todos os dados também apresenta, em sua maioria, valores muito próximos entre examinadores e entre os indivíduos, o que indica que há pouca variabilidade nos resultados do método tradicional e do método alternativo.

3.3. GRÁFICOS

Para uma melhor visualização das informações apresentadas e analisadas acima seguem alguns gráficos para a coordenada Y. Como a diferença entre métodos é a mais importante, os gráficos abaixo representam a diferença entre eles em cada individuo nos diferentes locais; a cor vermelha indica o método 1 e a cor azul o método 2

Figura 2: Diferença entre métodos na coordenada Y para o indivíduo 1

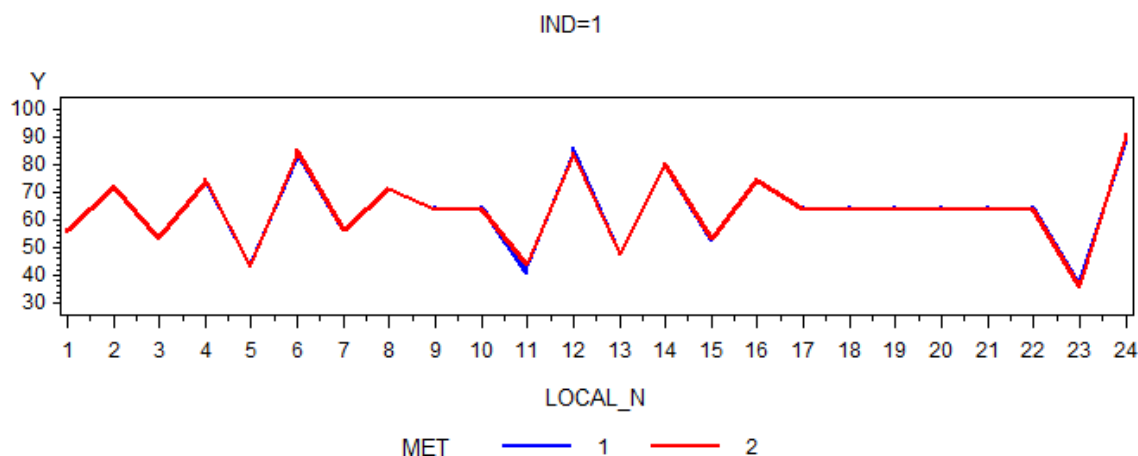


Figura 3: Diferença entre métodos na coordenada Y para os indivíduos 2 ao 5

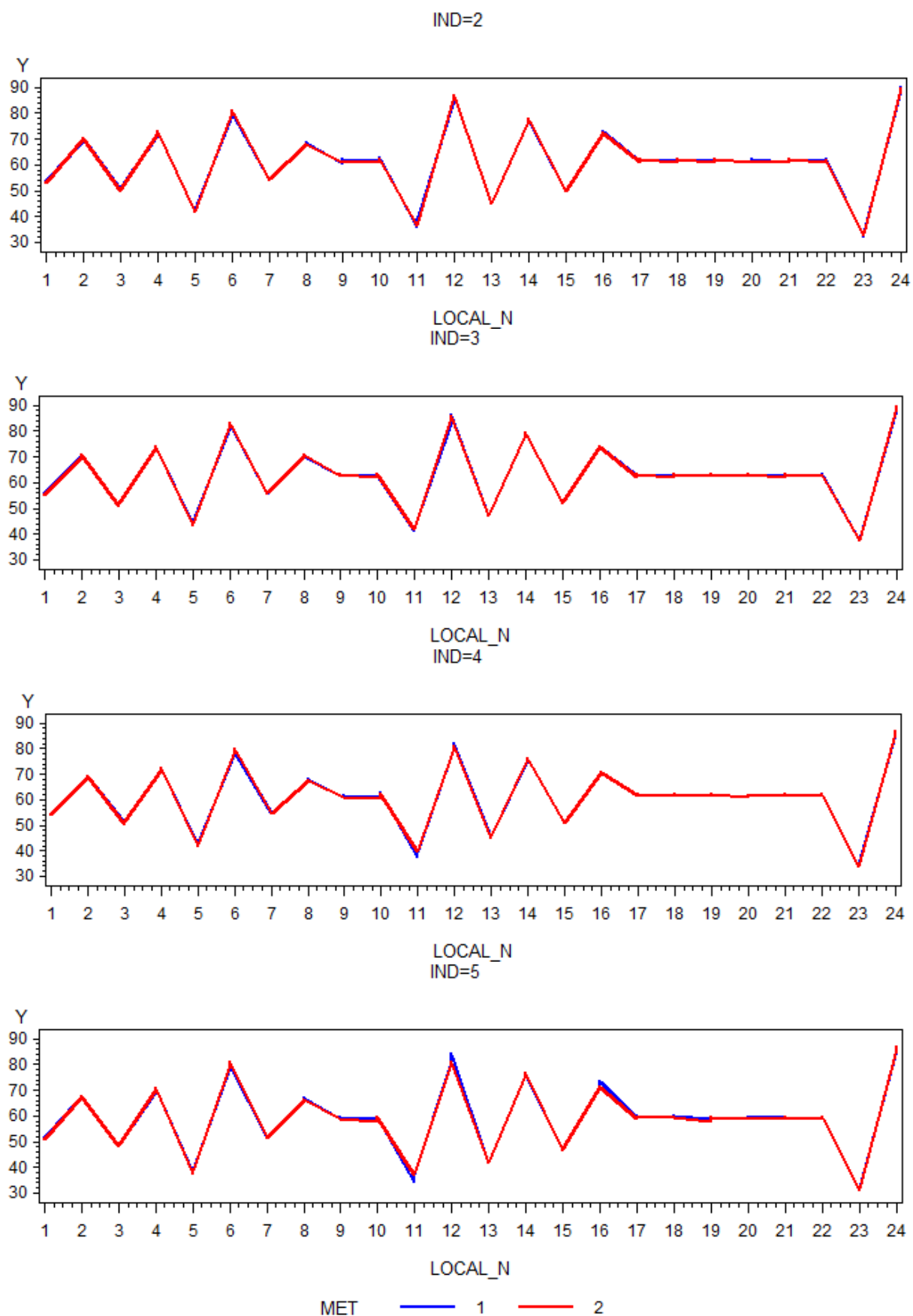


Figura 4: Diferença entre métodos na coordenada Y para os indivíduos 6 ao 9

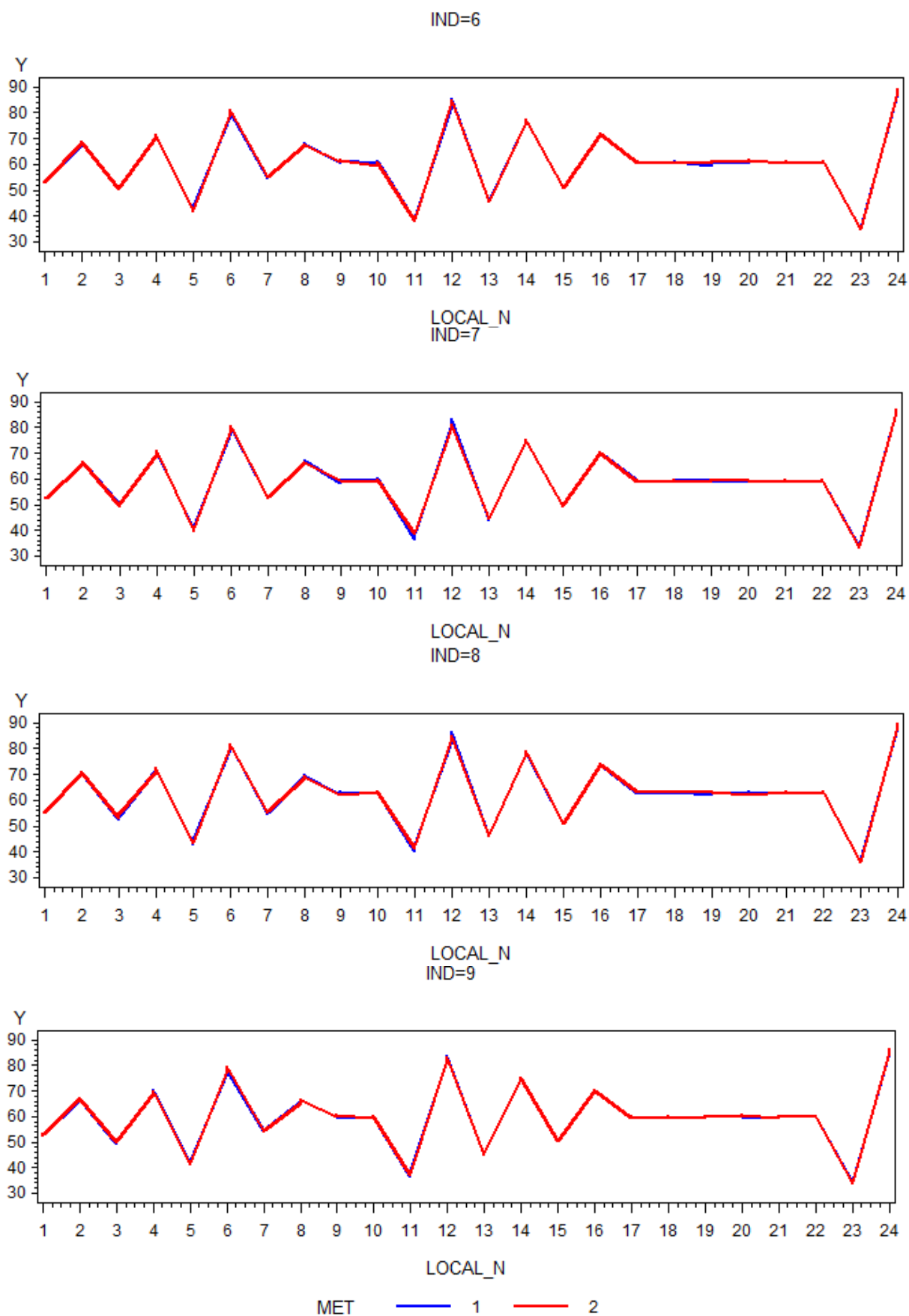


Figura 5: Diferença entre métodos na coordenada Y para os indivíduos 10 ao 13

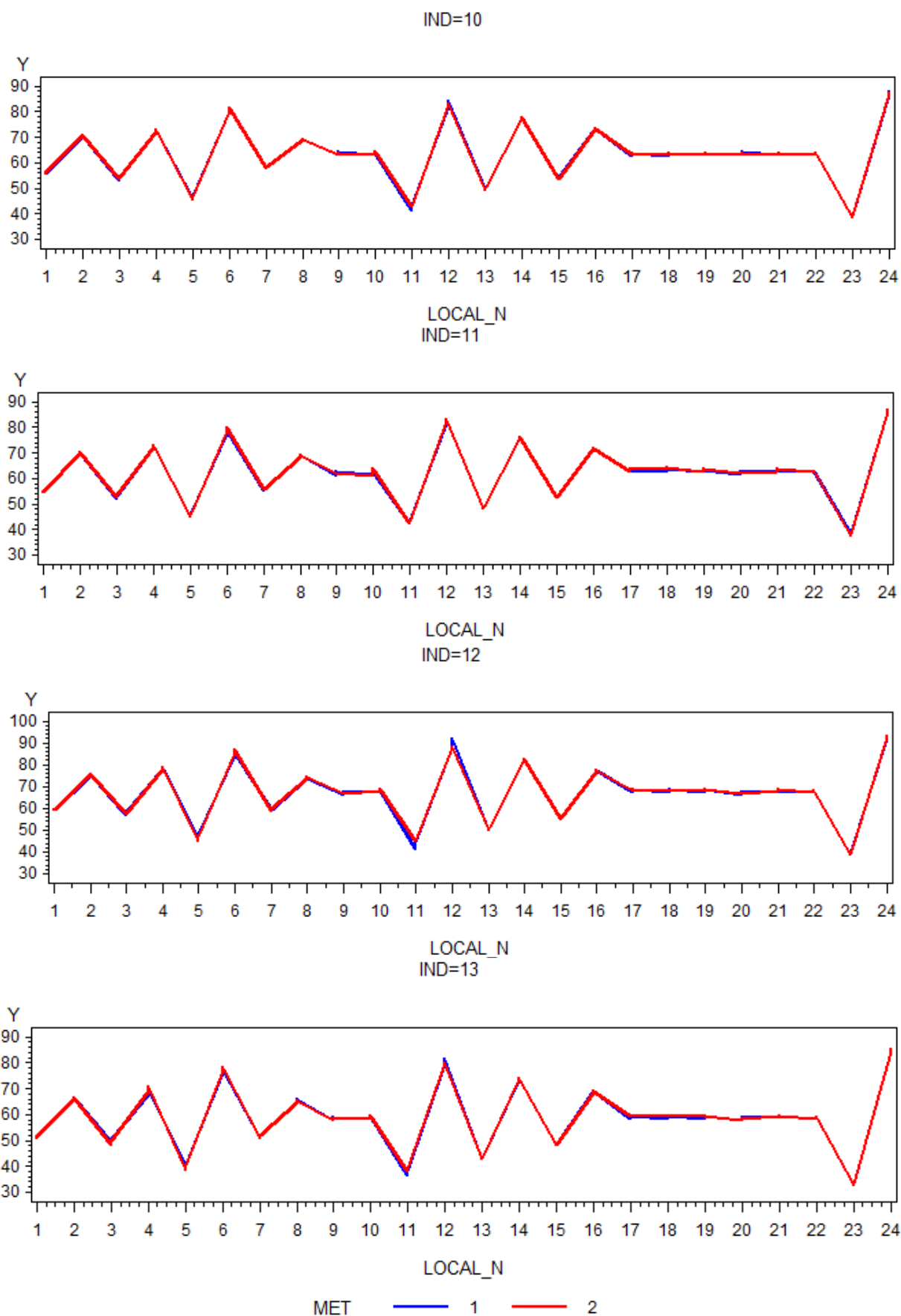


Figura 6: Diferença entre métodos na coordenada Y para os indivíduos 14 ao 17

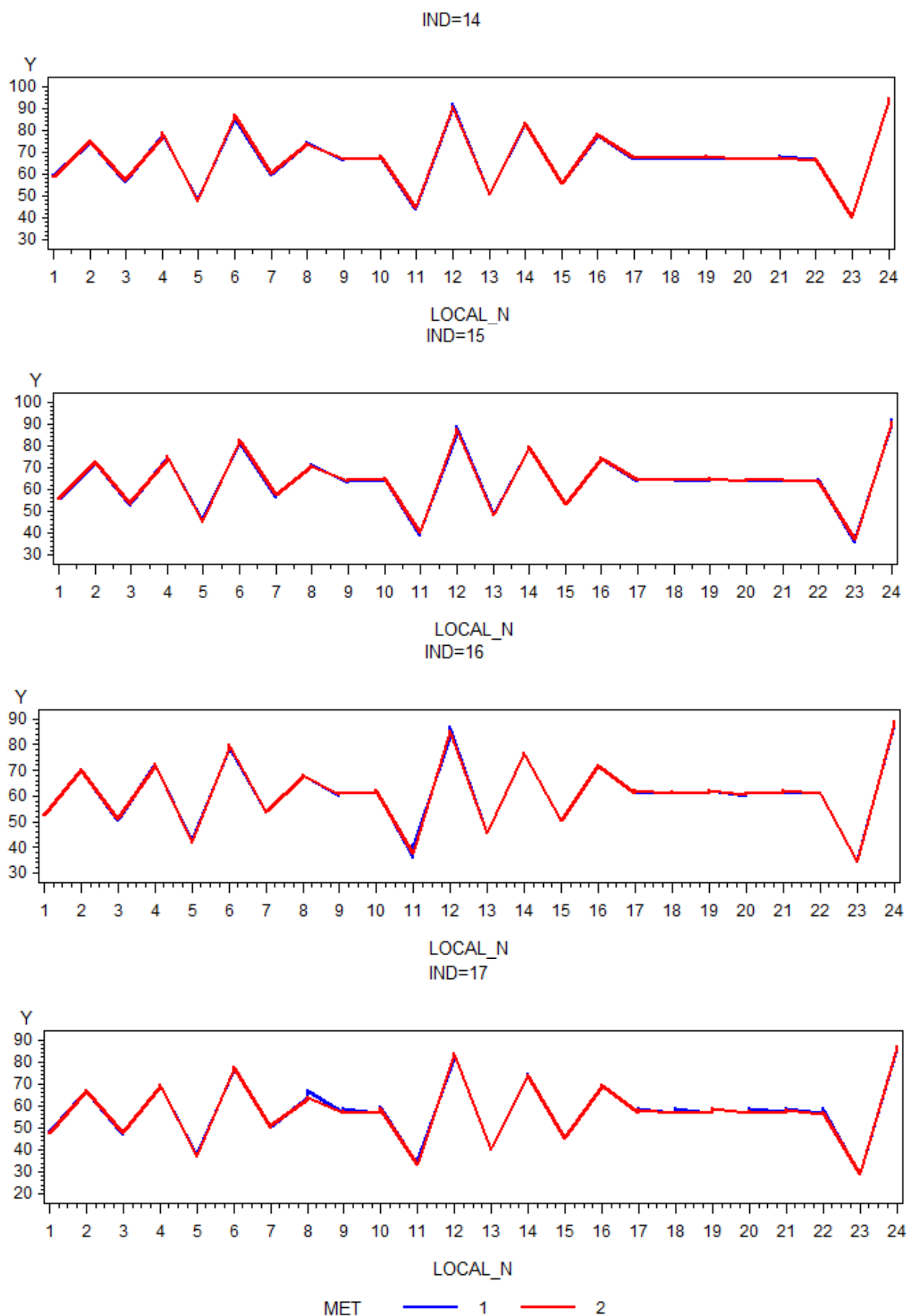
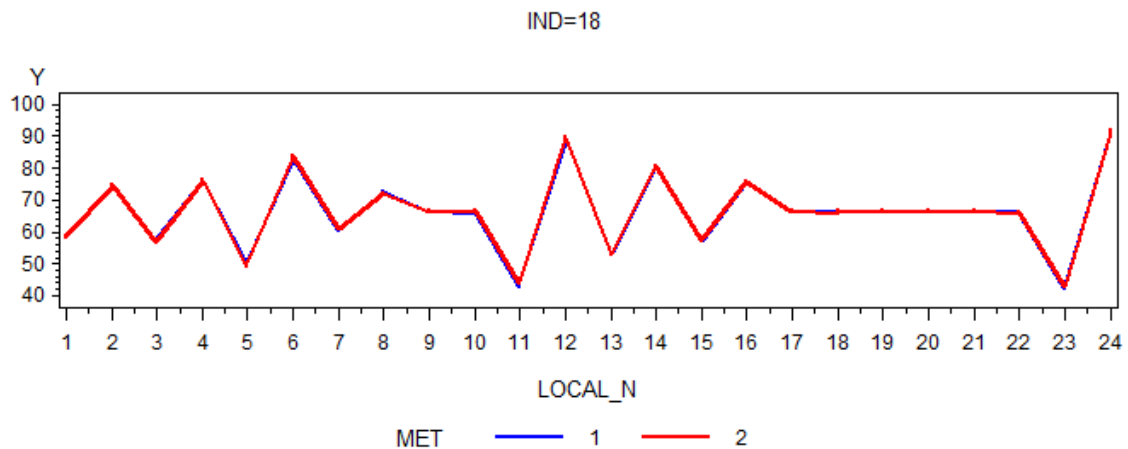


Figura 7: Diferença entre métodos na coordenada Y para o indivíduo 18



Notamos que nos gráficos de perfil de cada um dos 18 indivíduos que as linhas representativas das duas metodologias diferem muito pouco uma da outra, em cada um dos pontos estudados, até mesmo se sobrepondo em alguns momentos. Desta forma, os gráficos confirmam visualmente o que já havia sido apresentado nas tabelas 1, 2, 3 e 4, ou seja, uma diferença muito pequena entre os métodos em cada local. Embora mostramos aqui somente alguns gráficos, o que foi exposto é válido também para a coordenada X .

Outro tipo de gráfico para análise exploratória foi o proposto por Andrews (1972), próprio para dados multivariados. Se X_1, X_2, \dots, X_p são as variáveis observadas num elemento do conjunto, define-se a função

$$f(t) = X_1 / \sqrt{2} + X_2 \sin(t) + X_3 \cos(t) + X_4 \sin(2t) + X_5 \cos(2t) + \dots$$

e constrói-se um gráfico $(f(t), t)$, com $-\pi < t < \pi$, para cada elemento dos dados. A ideia é representar cada elemento que está num espaço multidimensional, por uma curva num espaço de duas dimensões.

No caso em questão um individuo do estudo é caracterizado por 24 variáveis (locais), para a dimensão Y e outras 24 para a dimensão X e a função para a dimensão Y fica da forma

$$f(t) = Y_1 / \sqrt{2} + Y_2 \sin(t) + Y_3 \cos(t) + Y_4 \sin(2t) + Y_5 \cos(2t) + \dots \\ Y_{22} \sin(11t) + Y_{23} \cos(11t) + Y_{24} \sin(12t)$$

Como o estudo visa discriminar diferenças entre métodos, podemos obter a média dos 24 locais, para cada indivíduo em cada método, baseado em 5 examinadores e então obter a função de Andrews para cada indivíduo e método. Usando o programa SAS obtemos os seguintes gráficos para a coordenada Y:

Figura 8: Gráficos de Andrews para os indivíduos 1 ao 6

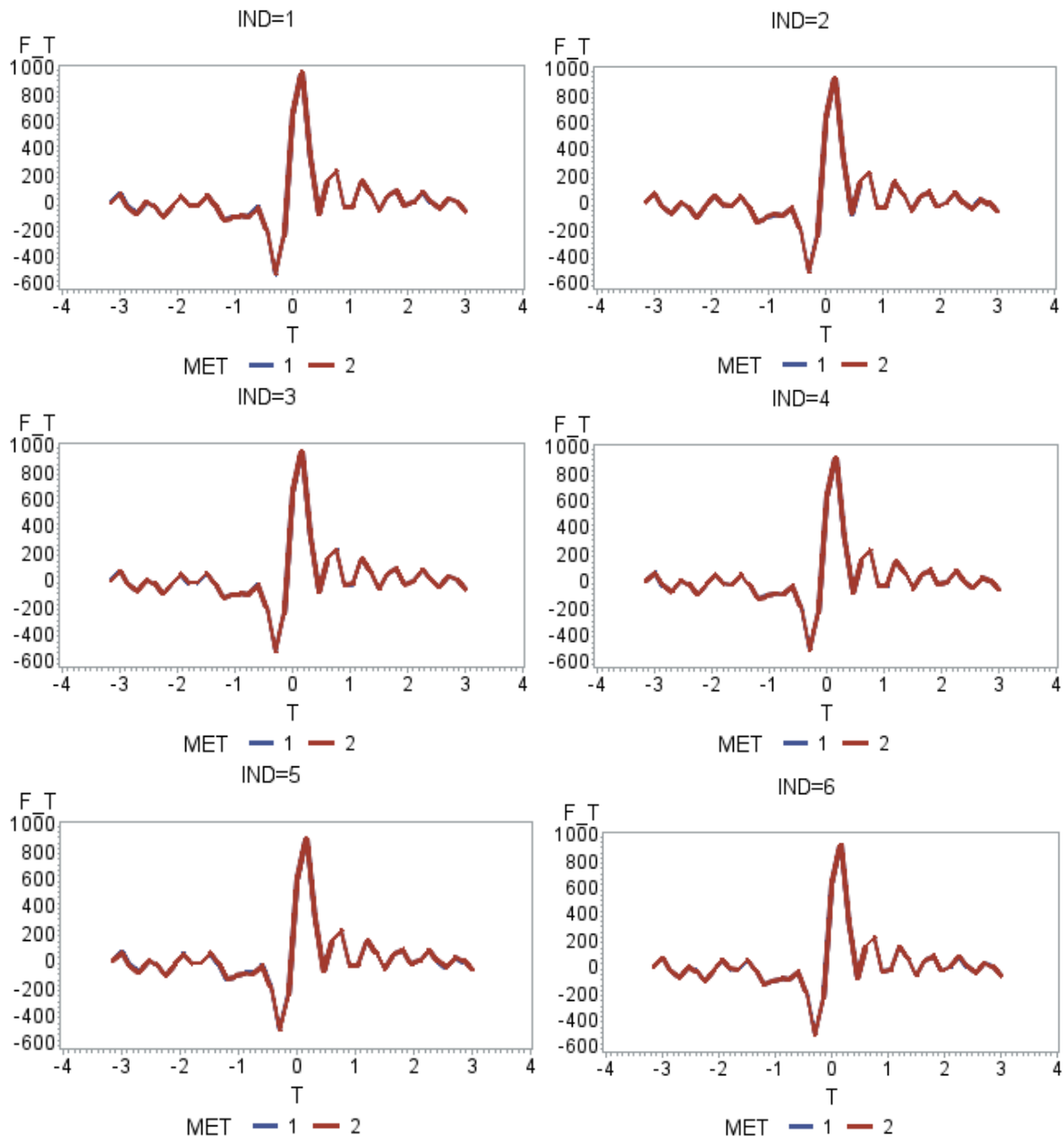


Figura 9: Graficos de Andrews para os indivíduos 7 ao 14

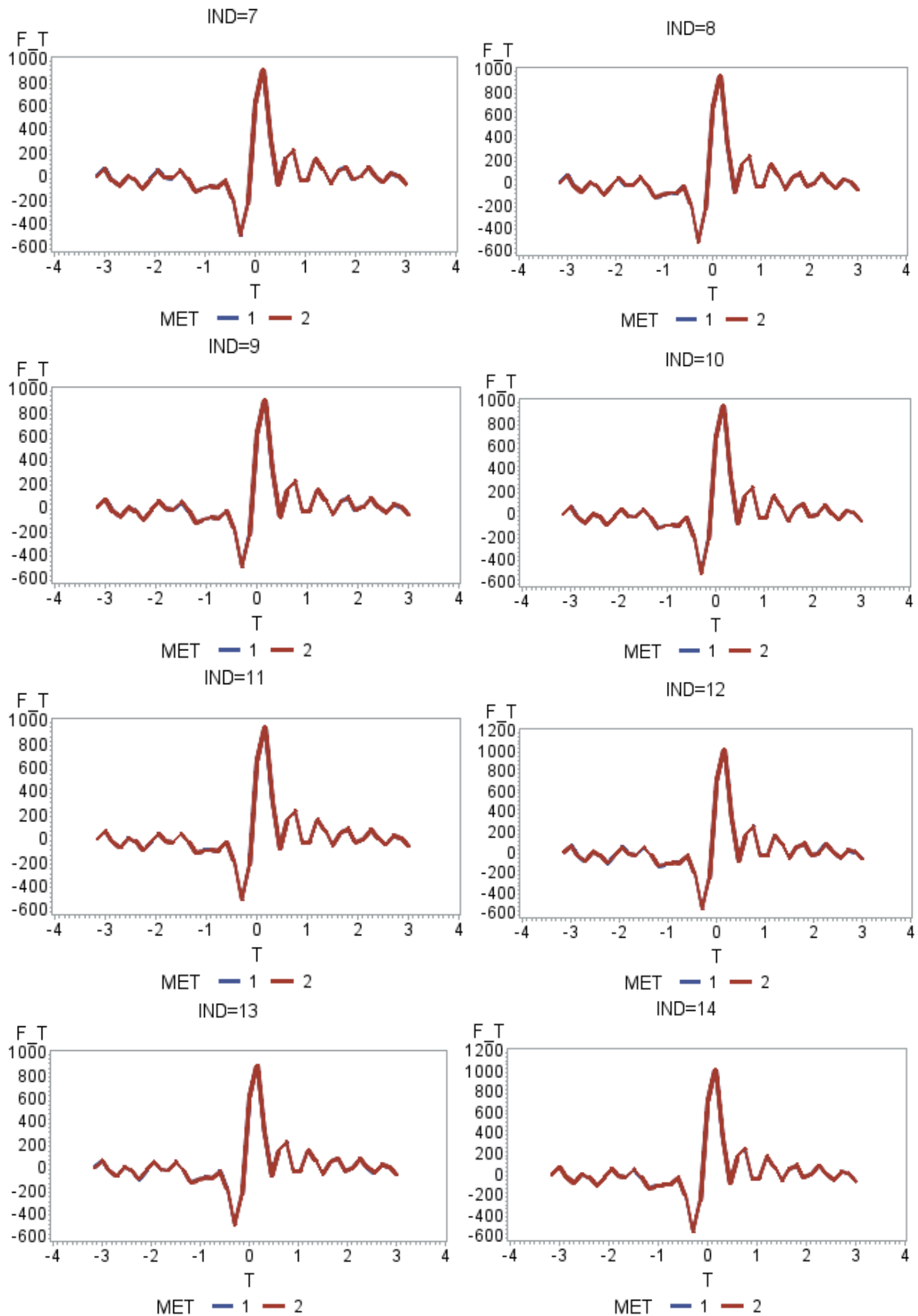
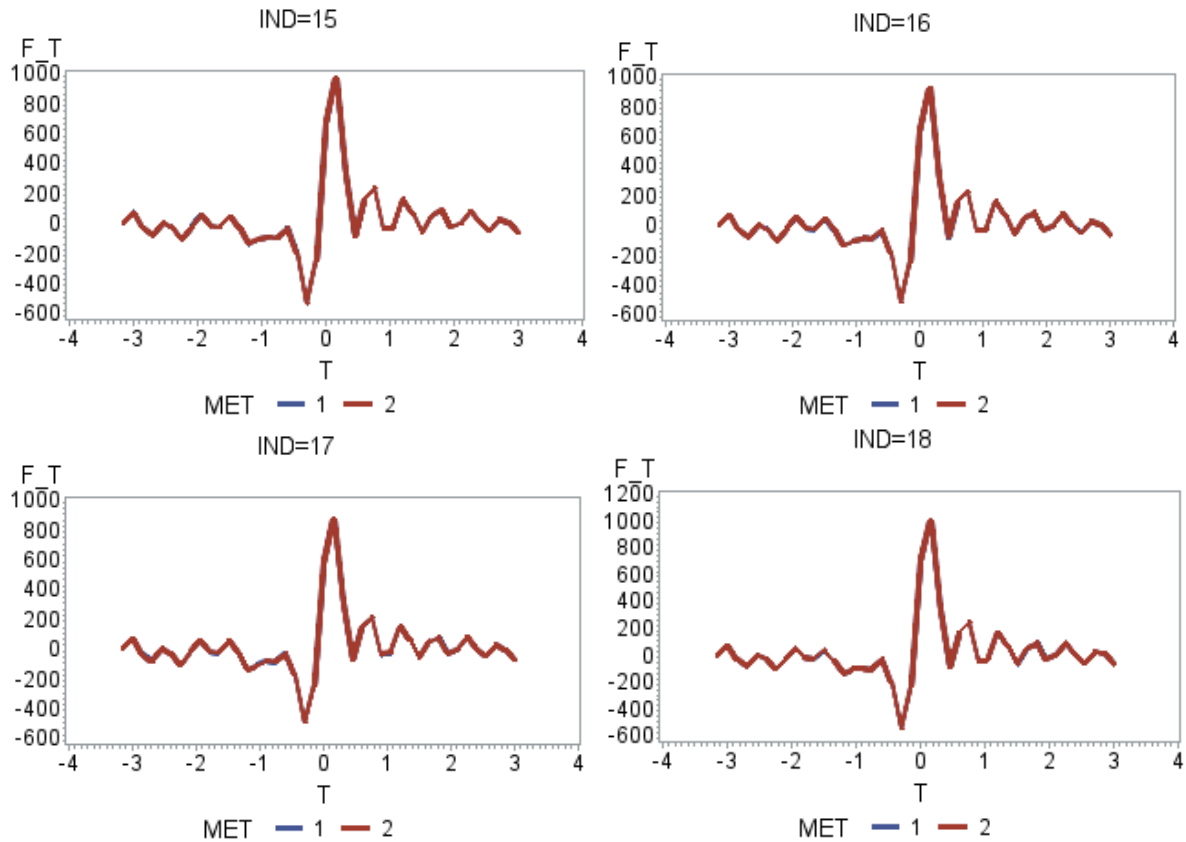


Figura 10: Gráficos de Andrews para os indivíduos 15 ao 18



Com as médias e gráficos dos dados vistos ate aqui, podemos concluir preliminarmente que neste estudo a diferença entre examinadores e entre métodos é muito pequena e que não há indicação de interação entre métodos e examinadores.

3.4. DISTÂNCIAS

Como o experimento tem duas variáveis resposta em cada local da face e são coordenadas num plano, vamos estabelecer um outro argumento que use simultaneamente ambas coordenadas. Seja P_1 o ponto obtido pelo examinador 1 com o método 1, isto é, $P_1 = (Y_1, X_1)$ e seja P_2 o ponto obtido pelo examinador 1 com o método 2, isto é, $P_2 = (Y_2, X_2)$. Estes dois pontos estão no plano e se a distância entre eles for próxima de zero é porque a diferença entre os dois métodos é pequena. A distância entre eles é dada por:

$$d = \sqrt{(Y_1 - Y_2)^2 + (X_1 - X_2)^2}$$

Assim um estudo da distância entre métodos pode auxiliar na análise exploratória dos dados deste experimento.

Mediante uma macro do SAS, obtemos todas as distâncias entre os dois métodos e o quadro de médias das distâncias para cada local é o seguinte:

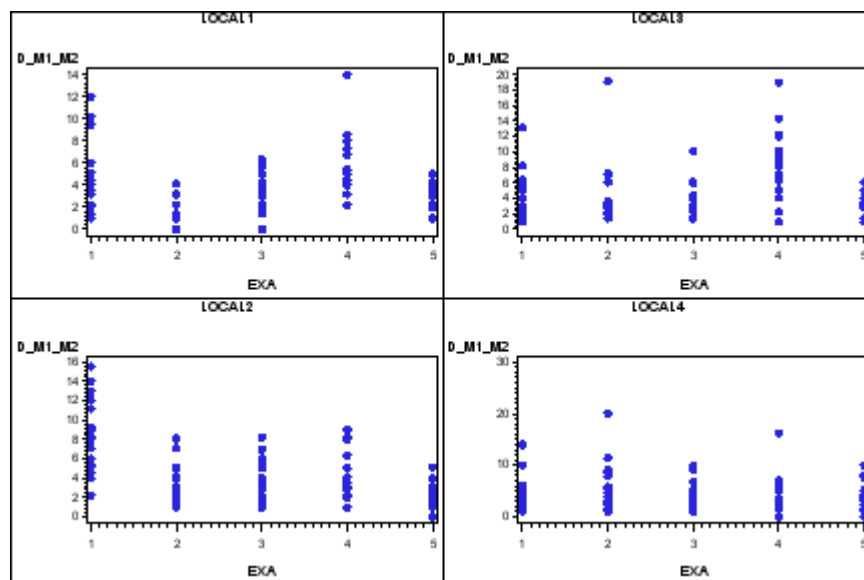
Tabela 5: Média das Distâncias para cada local

LOCAIS	DISTANCIA ENTRE METODOS					
	EXAMINADORES					MEDIA
	1	2	3	4	5	
1	5.06	1.99	2.98	5.4	2.99	3.68
2	8.26	3.43	4.04	5.25	2.27	4.65
3	4.37	4.34	3.3	7.96	2.79	4.55
4	4.3	5.69	4.06	4.42	3.66	4.43
5	7.35	2.88	12.47	9.94	12.04	8.94
6	9.31	3.36	12.12	13.16	10.82	9.75
7	3.88	1.59	2.33	2.78	2.31	2.58
8	4.72	1.62	3.59	4.54	3.24	3.54
9	9.58	11.96	30.18	27.72	15.25	18.94
10	7.01	4.94	4.11	5.84	4.34	5.25
11	29.15	17.08	24.82	22.88	39.68	26.72
12	20.25	18.1	24.18	28.18	34.86	25.12
13	1.91	1.81	2.52	2.74	2.79	2.35
14	2.4	1.34	2.27	2.93	2.29	2.25
15	1.56	1.67	1.66	1.75	1.27	1.58
16	1.67	1.26	2.11	4.27	1.58	2.18
17	3.47	5.67	5.59	4.25	3.2	4.44
18	4.73	6.55	4.07	11.53	2.55	5.89
19	9.7	12.67	8.16	10.21	8.81	9.91
20	10.98	17.46	5.99	19.42	18.05	14.38
21	2.24	5.28	5.68	3.23	2.21	3.73
22	3.59	4.29	2.38	2.95	3.21	3.28
23	26.23	21.47	13.94	27.32	21.55	22.1
24	29.85	15.22	16.36	21.81	20.98	20.85
MEDIA	8.82	7.15	8.29	10.44	9.28	8.79

Com as distâncias entre métodos podemos observar que em alguns locais, como o local 15, a diferença entre métodos é bem pequena ($d=1.58$) e em outros a distância é grande, como a do local 11 ($d=26.72$). Isto evidencia que a diferença entre os dois métodos varia de local para local. Em alguns poderá não haver diferença e em outros, grande diferença, ou seja, interação métodos x locais.

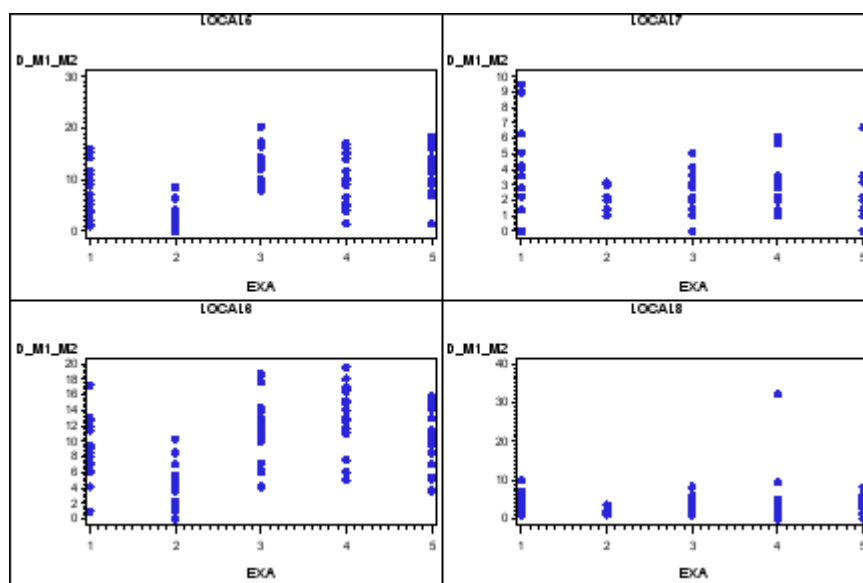
Para ilustrar a variabilidade das distâncias entre os examinadores, seguem alguns gráficos das distâncias obtidas com o SAS GRAPH:

Figura 11: Distâncias entre métodos nos locais 1 ao 4



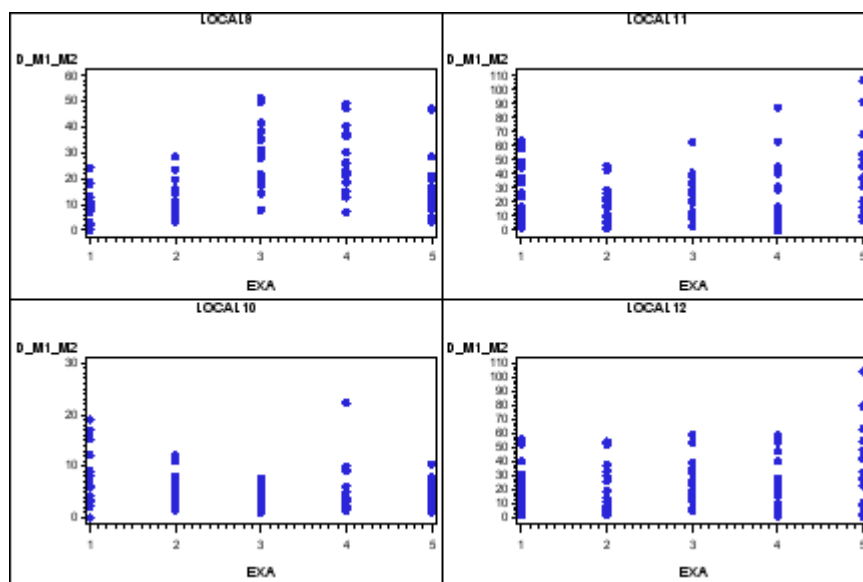
Avaliando os quatro primeiros locais, nota-se um comportamento destacado dos métodos nos examinadores 1 e 4, pois estes apresentaram maiores diferenças no tratamento dos 18 indivíduos. O examinador 2 também se destaca na observação do local 2.

Figura 12: Distâncias entre métodos nos locais 5 ao 8



Nos locais 5 e 6, os examinadores apresentaram comportamento semelhante quando aplicado os diferentes métodos, já no local 7 o examinador 1 apresentou grande variação das distâncias para cada indivíduo e no local 8, o examinador 4 apresentou apenas uma marcação bem distante das demais assinaladas.

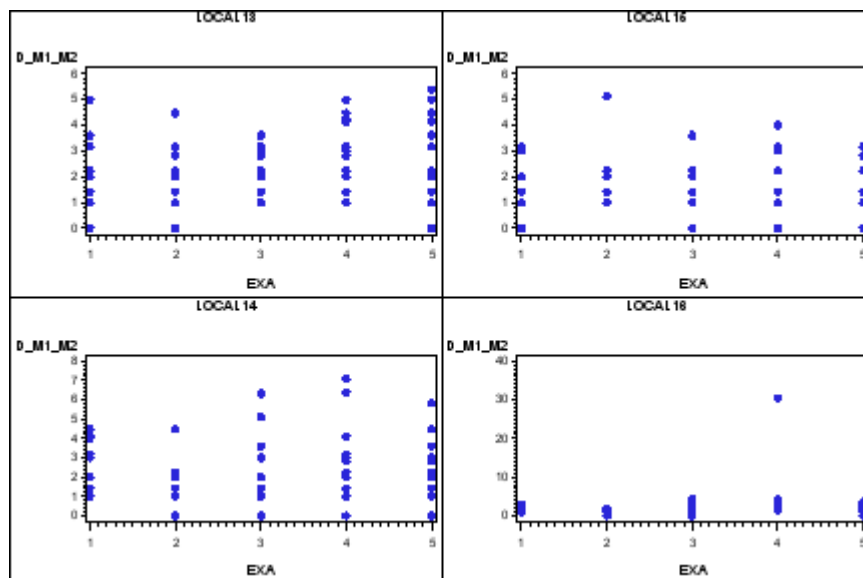
Figura 13: Distâncias entre métodos nos locais 9 ao 12



Da Figura 11 tem-se o comportamento das distâncias entre os dois métodos nos locais 9, 10, 11 e 12. Os locais 9, 11 e 12 apresentam comportamento bastante similares na dispersão dos pontos, já o local 10 os examinadores 1 e 4

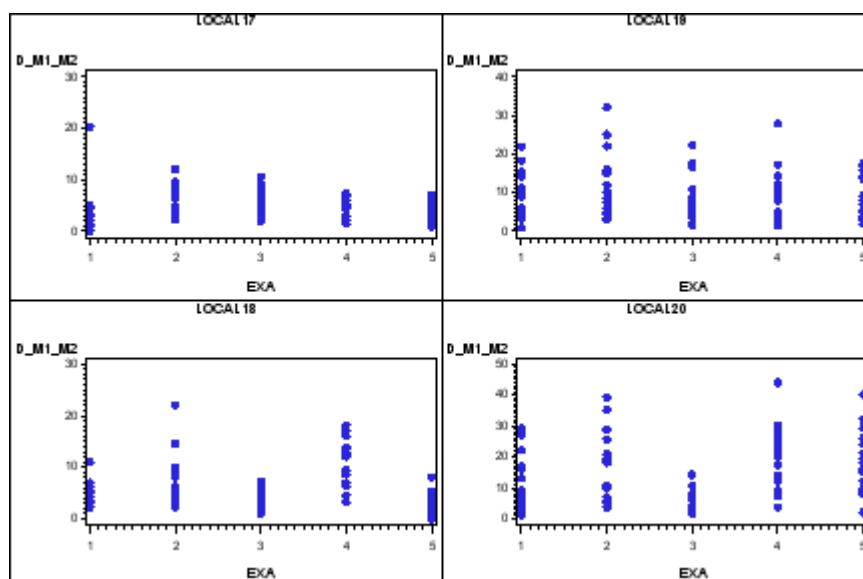
apresentaram maior dispersão das distâncias com destaque para o examinador 4 que apresentou apenas uma marcação bem distante das demais assinaladas.

Figura 14: Distâncias entre métodos nos locais 13 ao 16



Os locais 13, 14 e 15 também apresentaram dispersão dos pontos semelhantes entre cada um dos examinadores, porém no local 16 observa-se um outlier na marcação do examinador 4.

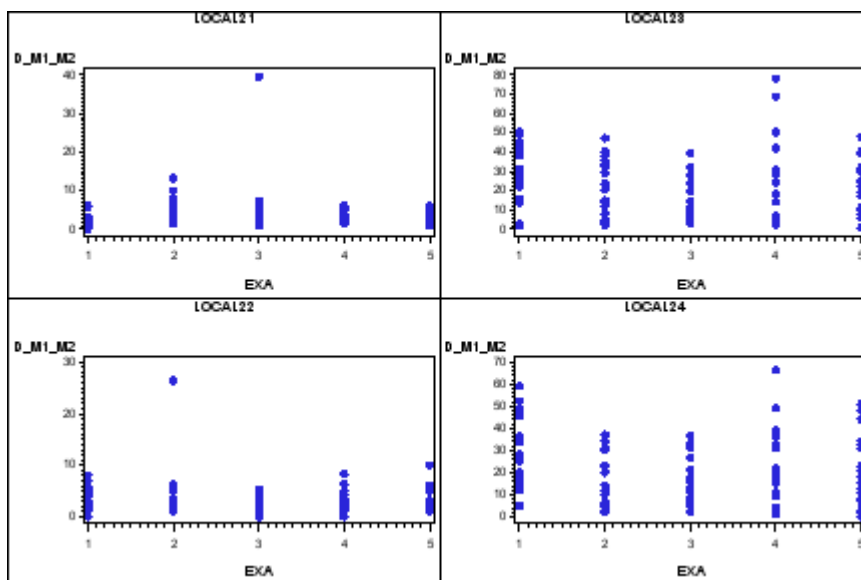
Figura 15: Distâncias entre métodos nos locais 17 ao 20



Entre os locais 17 ao 20 muita pouca diferença pode ser notada quando avaliada a dispersão dos pontos marcados por cada examinador. Merece destaque um ponto discrepante no local 17 pelo examinador 1 e maior concentração dos

pontos marcados pelo examinador 3 no local 20, quando comparado com os demais examinadores.

Figura 16: Distâncias entre métodos nos locais 21 ao 24



Dentre os locais 21 e 24 também há pouca diferença entre as dispersões de pontos marcados pelos 5 examinadores. Dentre estes locais merecem destaque os outliers observados no local 21 pelo examinador 3 e no local 22 pelo examinador 2.

Com base nos gráficos de todos os 24 pontos das distâncias pelas duas metodologias observados a partir da marcação de 5 examinadores, pode ser questionada a questão da diferença entre as metodologias, onde primeiramente não havia sido notado indicação de contestação. Uma vez que foi observado na maioria dos pontos grandes diferenças de marcação, chegando a alguns pontos até a 30 pixels, de acordo com a metodologia, e também a média geral das distâncias entre os dois métodos é de 8.79, entretanto esta conclusão só será definida após as análises de modelos e testes multivariados.

3.5. ANOVA EXPLORATÓRIA

Seguindo as recomendações de Hoaglin, Mosteller e Tukey(1991), de que a ANOVA pode ser aplicada em qualquer conjunto de dados estruturados, faremos uma análise de variância para as coordenadas Y e X, para cada local. Uma análise de variancia simples para cada local de Y e X, pode ser feita da seguinte forma:

Tabela 6: Anova para cada local de Y e X

FV	GL
INDIVIDUOS	17
EXAMINADORES	4
METODOS	1
INTERAÇÃO	4
ERRO	153
TOTAL	179

São 24 análises, um resumo sobre o que foi significativo esta na seguinte tabela 7. Nesta tabela, 1 (um) indica que houve diferença significativa pelo teste F e 0 (zero) indica que não houve. Observa-se que poucos foram os casos em que não houve significância, apesar de que os resultados descritivos indicassem diferenças bem pequenas. Na maior parte dos locais houve diferença entre ao examinadores e entre os dois métodos. Também a interação foi significativa na maior parte dos locais. Estes resultados foram mais presentes ainda na coordenada X.

Tabela 7: Diferenças significativas pelo teste F para todos os pontos

LOCAL	FONTES DE VARIAÇÃO					
	EXA		EXA*MET		MET	
	SIGY	SIGX	SIGY	SIGX	SIGY	SIGX
ALAR DIREITO	1	1	1	1	1	1
ALAR ESQUERDO	1	1	1	1	1	1
CHELION DIREITO	1	1	1	1	0	1
CHELION ESQUERDO	1	1	1	1	0	1
ECTOANTHION DIREITO	1	1	1	1	1	1
ECTOANTHION ESQUERDO	1	1	1	0	1	1
ENDOCANTHION DIREITO	1	1	1	0	1	1
ENDOCANTHION ESQUERDO	1	1	0	1	1	0
GLABELA	1	1	1	1	0	1
GNATHION	0	1	0	1	0	1
GONIO DIREITO	1	1	1	1	1	1
GONIO ESQUERDO	1	1	1	1	0	0
IRIDION LATERAL DIREITO	1	1	1	0	1	1
IRIDION LATERAL ESQUERDO	1	1	0	1	1	1
IRIDION MEDIAL DIREITO	1	1	1	1	1	1
IRIDION MEDIAL ESQUERDO	0	1	0	0	1	0
LABIAL INFERIOR	0	1	0	1	0	1
LABIAL SUPERIOR	0	1	0	1	0	1
LABIOMENTAL	0	1	0	1	0	1
NASION	1	1	1	1	1	1
STOMION	0	0	0	0	0	0
SUBNASAL	0	1	0	1	1	1
ZYGION DIREITO	1	1	1	1	1	1
ZYGION ESQUERDO	1	0	1	1	1	1

Ainda relacionada com ANOVA descritiva por local e coordenada, a tabela seguinte contém as variâncias de cada local (QM).

Tabela 8: Variância de cada local

Obs	LOCAL	_TYPE_	QM_Y	QM_X
1	ALAR DIREITO	ERROR	0.8883	3.997
2	ALAR ESQUERDO	ERROR	0.7518	5.010
3	CHELION DIREITO	ERROR	98.074	1.125
4	CHELION ESQUERDO	ERROR	91.605	1.275
5	ECTOANTHION DIREITO	ERROR	103.963	2.484
6	ECTOANTHION ESQUERDO	ERROR	99.476	3.179
7	ENDOCANTHION DIREITO	ERROR	24.302	1.045
8	ENDOCANTHION ESQUERDO	ERROR	86.334	1.450
9	GLABELA	ERROR	90.381	60.807
10	GNATHION	ERROR	175.455	1.374
11	GONIO DIREITO	ERROR	457.920	276.170
12	GONIO ESQUERDO	ERROR	371.778	216.639
13	IRIDION LATERAL DIREITO	ERROR	0.6371	1.376
14	IRIDION LATERAL ESQUERDO	ERROR	0.8896	1.720
15	IRIDION MEDIAL DIREITO	ERROR	0.4025	0.971
16	IRIDION MEDIAL ESQUERDO	ERROR	36.613	3.934
17	LABIAL INFERIOR	ERROR	82.455	4.757
18	LABIAL SUPERIOR	ERROR	78.148	6.145
19	LABIOMENTAL	ERROR	127.537	45.765
20	NASION	ERROR	79.904	52.349
21	STOMION	ERROR	73.733	9.696
22	SUBNASAL	ERROR	60.062	2.912
23	ZYGION DIREITO	ERROR	41.832	181.539
24	ZYGION ESQUERDO	ERROR	38.617	145.017

Este é um resultado importante: as variâncias são muito diferentes, principalmente na coordenada X e isto deve ser considerado na análise estatística envolvendo todos os dados.

A ANOVA exploratória para distância entre métodos tem o seguinte formato:

Tabela 9: ANOVA para distancia entre métodos

FV	GL
IND	17
EXA	4
ERRO	68
TOTAL	89

Um resumo da ANOVA exploratória das distâncias entre métodos esta na tabela 10, onde MEDIA é a média geral da distância em cada local, F é a estatística F para testar diferenças entre os examinadores, PROB é o P valor de F e VARIANCIA é o quadrado médio do ERRO da ANOVA.

Tabela 10: Resumo da ANOVA exploratória das distancias entre métodos

LOCAL	MEDIA	F	PROB	VARIANCIA
1	3,68	7,77	0,001	5,03
2	4,65	18,04	0,001	5,22
3	4,55	8,63	0,001	8,54
4	4,43	1,71	0,158	6,14
5	8,94	22,94	0,001	12,23
6	9,75	23,49	0,001	11,36
7	2,58	5,05	0,002	2,52
8	3,54	2,3	0,068	12,08
9	18,94	24,22	0,001	65,65
10	5,25	1,69	0,163	15,09
11	26,72	3,46	0,013	370,7
12	25,12	2,27	0,071	353,12
13	2,35	4,16	0,006	0,94
14	2,25	2,79	0,034	2,11
15	1,58	0,65	0,63	0,94
16	2,18	2,72	0,038	9,66
17	4,44	3,15	0,021	7,63
18	5,89	20,4	0,001	10,61
19	9,91	1,59	0,189	34,15
20	14,38	11,86	0,001	49,42
21	3,73	3	0,025	16,49
22	3,28	1,08	0,375	8,52
23	22,1	2,36	0,063	213,18
24	20,85	3,18	0,02	189,69

Dos 24 locais, em 17 houve diferença significativa entre os examinadores, indicando interação entre métodos e locais e que há grande variabilidade entre locais, concordando com a ANOVA exploratória das coordenadas Y e X.

3.6. CONCLUSÃO BASEADA NA ANÁLISE DESCRITIVA

Nesta primeira discussão sobre os resultados, devemos lembrar que o objetivo do experimento foi comparar dois métodos de marcações dos pontos antropométricos, usando 5 examinadores. Pelo que se sabe, um bom resultado seria se não houvesse diferenças entre métodos, entre examinadores e não houvesse interação.

Por meio das médias dos fatores, as diferenças entre examinadores e entre métodos não são grandes e a interação é reduzida. Com os diferentes tipos de gráficos podemos também observar que estas diferenças são pequenas. Entretanto, por meio da ANOVA exploratória das coordenadas Y e X e das distâncias entre métodos, há evidências que as diferenças entre examinadores, entre métodos e a interação são significativas e interagem com locais.

Na ANOVA exploratória leva-se em conta apenas a estrutura do conjunto de dados e não há preocupação com casualização e distribuição dos dados e isto terá que ser considerado mais tarde. Por outro lado, uma análise conjunta com todos os locais poderá evidenciar melhor algumas das afirmações advindas da análise descritiva.

4. ANÁLISE DE EXPERIMENTOS

4.1. EXPERIMENTOS CASUALIZADOS

Experimentos vem sendo feitos há muito tempo para avaliar e comparar tratamentos. Muito dos cuidados na realização dos experimentos já eram levados em conta antes e durante a realização do experimento, mas foi Fisher quem introduziu a casualização, uma ação que retira do pesquisador a decisão sobre quais parcelas vão receber um determinado tratamento. Fisher aplicou um modelo probabilístico para indicar quais parcelas iriam receber um tratamento. Embora tenha escrito sobre o assunto a partir de 1922, foi em seu livro, *The Design of Experiments*, em 1935, que a casualização foi realmente colocada em evidência. Posteriormente, vários autores estudaram a casualização proposta por Fisher e exploraram as propriedades que dela podiam vir. Kempthorne (1952) foi um destes cientistas, mas outros como Nelder (1965) também contribuíram, em muito, para a teoria da casualização em experimentos.

Uma consequência da casualização é que ela define o modelo estatístico do experimento, isto é, a forma como a casualização foi realizada é que define o modelo, seja blocos ao acaso, parcelas divididas, etc. Também define como analisar o experimento pela ANOVA, técnica também introduzida por Fisher. Embora Fisher (1935) tenha comentado, Kempthorne e seus seguidores concluíram que é a casualização que valida as estimativas e os testes de significância na ANOVA, e não a normalidade e a independência dos erros. Também concluíram que é a casualização que define a estrutura de variância e covariância dos dados experimentais.

Diante deste conhecimento é fato que não conhecemos o modelo do experimento que está sendo analisado e nem a correspondente estrutura de variâncias e covariância desde que não houve casualização no experimento básico envolvendo indivíduos, examinadores e métodos. Mesmo nestas condições a ANOVA Fisheriana pode ser usada, mas com alguns cuidados. Em primeiro lugar vamos trabalhar com os possíveis modelos para este experimento.

4.2. MODELOS

Como já foi caracterizado, este experimento é um fatorial 5 x 2, com 18 repetições, isto é, 5 examinadores, dois métodos e 18 indivíduos (blocos). Além disto, as variáveis resposta, Y e X, foram medidas em 24 locais. Numa situação ideal este seria o caso de um delineamento em blocos ao acaso com medidas repetidas no espaço. Qual seria então, o modelo dos possíveis modelos para analisar este experimento? O modelo mais simples seria o de blocos ao acaso com parcelas divididas, com o fatorial 5 x 2 nas parcelas e os locais nas sub parcelas. Para a coordenada Y é o seguinte:

$$Y_{ijkl} = \mu + E_i + M_j + (EM)_{ij} + I_k + \varepsilon_{ijk} + L_l + (EL)_{il} + (ML)_{jl} + (EML)_{ijl} + \phi_{ijkl}$$

com

E_i = efeito do examinador i.

M_j = efeito do método j.

I_k = efeito do individuo k.

ε_{ijk} = erro experimental.

L_l = efeito do local l.

$(EM)_{ij}$, $(EL)_{il}$, $(ML)_{jl}$ e $(EML)_{ijl}$ = efeitos das respectivas interações.

ϕ_{ijkl} = erro gerado pelos locais.

O modelo seria o mesmo para a coordenada X.

O segundo modelo (para Y e X) é o mesmo mas usando a teoria de Huynh e Feldt (1970), bem explicada por exemplo por Milliken e Johnson (2009), que propôs um ajustamento dos graus de liberdade para as fontes de variação que envolvem locais, desde que eles foram sequenciais e não casualizados. Este ajustamento foi proposto anteriormente por Greenhouse e Geisser (1959), mas o primeiro é mais conhecido. Basicamente, Huynh e Feldt (1970) estabeleceram as condições em que experimentos com dados repetidos no tempo e ou espaço poderiam ser analisados

como se fossem sub parcelas e tivessem sido casualizados. O ajustamento dos graus de liberdade é feito pela estimativa do parâmetro θ de Box que tem os seguintes limites :

$$\frac{1}{t} < \theta < 1$$

Com t sendo o número de locais. Se θ está próximo de 1 então dizemos que existe a condição HF caso contrário, não. Neste último caso, o grau de liberdade de cada fonte de variação é multiplicado por θ se a análise segue em frente. Para estes dois modelos a análise de variância seria a seguinte :

Tabela 11: ANOVA de Delineamento em parcelas divididas para a coordenada Y

FV	GL	SQ	QM	F	Pr>f
EXA	4	1,761287	0.440322	2.90	0.0239
MET	1	0,079225	0.079225	0.52	0.4713
EXA*MET	4	2,5275	0.631875	4.16	0.0032
IND	17	31526	1854.451531	-	-
Erro(a)	153	23,2405	0.1519	0.18	1.0000
LOCAL	23	754810,4	32817.8426	39559.1	<.0001
EXA*LOCAL	92	90,734	0.9862	1.19	0.1077
MET*LOCAL	23	78,6194	3.4182	4.12	<.0001
EXA*MET*LOCAL	92	70,2136	0.7632	0.92	0.6929
Erro(b)	3910	3243,702	0.8296	-	-
Total	4319	789846,9	-	-	-

Tabela 12: ANOVA de Delineamento em parcelas divididas para a coordenada X

FV	GL	SQ	QM	F	Pr>F
EXA	4	26,26589	6.566473	8.15	<.0001
MET	1	24,55567	24.555669	30.47	<.0001
EXA*MET	4	30,8127	7.703175	9.56	<.0001
IND	17	23680	1392.917066	-	-
Erro(a)	153	123,2837	0.8058	0.42	1.0000
LOCAL	23	932928,1	40562.0891	20911.9	<.0001
EXA*LOCAL	92	412,1548	4.4799	2.31	<.0001
MET*LOCAL	23	498,5837	21.6776	11.18	<.0001
EXA*MET*LOCAL	92	389,9426	4.2385	2.19	<.0001
Erro(b)	3910	7584,08	1.9397	-	-
Total	4319	965697,3	-	-	-

Tabela 13: ANOVA com aproximação HF para a coordenada Y

FV	GL	SQ	QM	F	Pr>F	G-G	H-F-L
LOCAL	23	754810	32817	39559	<.0001	<.0001	<.0001
LOCAL*EXA	92	90,734	0,9862	1.19	0.1077	0.2884	0.2874
LOCAL*MET	23	78,6194	3,4182	4.12	<.0001	0.0069	0.0066
LOCAL*EXA*MET	92	70,2136	0,7632	0.92	0.6929	0.5255	0.5268
Erro(b)	3910	3243,702	0,8296	-	-	-	-

Greenhouse-Geisser Epsilon 0.1286

Huynh-Feldt-Lecoutre Epsilon 0.1311

Tabela 14: ANOVA com aproximação HF para a coordenada X

FV	GL	SQ	QM	F	Pr>F	G-G	H-F-L
LOCAL	23	932928	40562	20911	<.0001	<.0001	<.0001
LOCAL*EXA	92	412,1548	4,4799	2.31	<.0001	0.0074	0.0070
LOCAL*MET	23	498,5837	21,6776	11.18	<.0001	<.0001	<.0001
LOCAL*EXA*MET	92	389,9426	4,2385	2.19	<.0001	0.0118	0.0112
Erro(b)	3910	7584,08	1,9397	-	-	-	-

Greenhouse-Geisser Epsilon 0.1286

Huynh-Feldt-Lecoutre Epsilon 0.1311

Os dois modelos apresentam os mesmos resultados para as coordenadas Y e X em relação a alguns efeitos. A diferença entre examinadores e interação entre métodos e examinadores é significativa em ambas coordenadas. A diferença entre métodos é significativa apenas para a coordenada X. Com relação às interações com locais, todas são significativas para a coordenada X, enquanto que para a coordenada Y, apenas a interação com métodos foi significativa. Também, os valores de θ de Box foram 0.1311 para HF e 0.1266 para GG e para este caso θ menor valor é 0.042. Como as estimativas de θ esta longe de 1 não ocorreu a

condição HF e houve ajustamento dos graus de liberdade das fontes de variação que contem LOCAL e do Erro(b).

O terceiro modelo seria o mesmo que o primeiro mas como um modelo misto onde a estrutura de variância e covariância dos locais não é conhecida. O PROC MIXED do SAS tem muitas destas estruturas que podem ser testadas em um modelo misto como este, entretanto, em todos os casos tentados, a função de verossimilhança não converge e por isso não há solução. Neste estudo já foi visto que as variâncias são bem diferentes em relação aos locais, condição que inviabiliza o primeiro modelo. Por outro lado todos os modelos dependem de normalidade, ponto que será discutido mais adiante.

Considerando as distâncias entre métodos envolvendo as duas coordenadas, teríamos também os três modelos citados anteriormente: parcelas divididas, graus de liberdade ajustado e misto. O modelo básico seria o seguinte:

$$D_{ijk} = \mu + E_i + I_j + \varepsilon_{ij} + L_k + (EL)_{ik} + \phi_{ijk},$$

Onde a fonte de variação métodos foi absorvida pela distância. A análise de variância para os dois primeiros modelos é a seguinte:

Tabela 15:ANOVA Delineamento em parcelas divididas para a distancia entre métodos

FV	GL	SQ	QM	F	Pr>F
IND	17	4052,918287	238.406958	2.50	0.0040
EXA	4	2541,346603	635.336651	6.66	0.0001
Erro(a)	68	6489,5101	95.4340	1.43	0.0138
LOCAL	23	132230,5742	5749.1554	85.89	<.0001
EXA*LOCAL	92	23806,3936	258.7651	3.87	<.0001
Erro(b)	1955	130856,0537	66.9340	-	-
Total	2159	299976,7964	-	-	-

Tabela 16: ANOVA com aproximação HF para a distancia entre métodos

FV	DF	SQ	QM	F	Pr>F	G.G	H.F.L
LOCAL	23	132230,5742	5749,1554	85.89	<.0001	<.0001	<.0001
LOCAL*EXA	92	23806,3936	258,7651	3.87	<.0001	<.0001	<.0001
Error(LOCAL)	1955	130856,0537	66,934	-	-	-	-

Greenhouse-Geisser Epsilon 0.1700

Huynh-Feldt-Lecoutre Epsilon 0.1792

Para este modelo o ajustamento da matriz de variâncias e covariâncias no modelo misto foi feito via PROC MIXED do SAS. A matriz que se adaptou melhor pelo critério de Scharzt foi a denominada matriz não estruturada (UN), uma simétrica heterogênea, onde as variâncias e covariâncias são diferentes.

Os resultados foram os seguintes:

Tabela 17: Matriz de Variâncias e Covariâncias para locais

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12
1	5.62	1.78	1.13	1.18	-1.67	0.43	0.84	0.79	3.29	0.31	-3.62	-2.81
2	1.78	6.52	-0.5	0.34	0.03	1.58	0.54	-0.21	3.66	1.16	11.33	6.06
3	1.13	-0.5	11.2	5.14	-1.1	-1.2	0.56	0.75	1.97	-1.13	-4.13	0.05
4	1.18	0.34	5.14	11.33	-0.49	-0.17	0.9	1.43	1.66	-2.27	-10.46	-9.38
5	-1.67	0.03	-1.1	-0.49	14.58	7.68	0.18	1.88	0.62	-2.5	7.07	18.72
6	0.43	1.58	-1.2	-0.17	7.68	13.7	0.92	0.39	8.94	0.19	1.61	9.43
7	0.84	0.54	0.56	0.9	0.18	0.92	2.36	0.18	2.14	-0.03	-2.98	-3.35
8	0.79	-0.21	0.75	1.43	1.88	0.39	0.18	12.58	-3.27	0.6	14.2	12.88
9	3.29	3.66	1.97	1.66	0.62	8.94	2.14	-3.27	95.18	2.41	22.9	23.63
10	0.31	1.16	-1.13	-2.27	-2.5	0.19	-0.03	0.6	2.41	15.97	6.66	5.74
11	-3.62	11.33	-4.13	-10.46	7.07	1.61	-2.98	14.2	22.9	6.66	425.1	305.97
12	-2.81	6.06	0.05	-9.38	18.72	9.43	-3.35	12.88	23.63	5.74	305.97	386.9
13	0.48	-0.01	0.59	0.3	-0.57	-0.5	-0.06	-0.23	-1.26	-0.53	-5.63	-4.96
14	0.54	-0.18	1	0.16	-0.25	-0.2	-0.06	-0.23	-4.62	0.14	-11.72	-5.39
15	0.52	0.5	0.31	-0.08	-0.97	0.19	0.08	-0.87	3.05	0.46	-2.77	-2.9
16	1.3	1.61	-1.06	0.12	-2.54	-1.74	-0.53	-1.82	6.91	-0.82	-12.67	-5.9
17	-0.38	0.42	1.27	0.86	-1.67	-1.27	-0.64	-0.21	4.82	4.26	0.65	1.43
18	-0.57	-1.03	-0.19	0.31	-1.81	-0.34	0.09	2.48	-1.92	4.05	5.43	-4.41
19	-2.51	0.1	0.18	-3.52	-1.91	-1.22	-0.69	-0.05	-5.32	3.07	2.92	6.58
20	0.35	1.63	0.36	6.46	-1.06	3.11	-1.59	-0.3	5.66	-6.35	10.56	9.99
21	0.11	-1.26	-0.04	-0.83	-0.67	-1.61	0.41	1.57	5.34	5.08	-6.24	1.74
22	-0.43	0.96	-1.19	1.09	0.45	1.86	0.64	2.09	0.6	2.45	2.05	-0.92
23	3.69	9.21	-4.44	-6.6	4.16	8.89	0.61	3.06	14.1	-9.49	-2.52	-36.5
24	3.37	9.06	2.17	-6.13	-2.89	3.88	1.87	-9.93	32.79	-0.84	-32.16	-21.9

Row	Col13	Col14	Col15	Col16	Col17	Col18	Col19	Col20	Col21	Col22	Col23	Col24
1	0.48	0.54	0.52	1.3	-0.38	-0.57	-2.51	0.35	0.11	-0.43	3.69	3.37
2	-0.01	-0.18	0.5	1.61	0.42	-1.03	0.1	1.63	-1.26	0.96	9.21	9.06
3	0.59	1	0.31	-1.06	1.27	-0.19	0.18	0.36	-0.04	-1.19	-4.44	2.17
4	0.3	0.16	-0.08	0.12	0.86	0.31	-3.52	6.46	-0.83	1.09	-6.6	-6.13
5	-0.57	-0.25	-0.97	-2.54	-1.67	-1.81	-1.91	-1.06	-0.67	0.45	4.16	-2.89
6	-0.5	-0.2	0.19	-1.74	-1.27	-0.34	-1.22	3.11	-1.61	1.86	8.89	3.88
7	-0.06	-0.06	0.08	-0.53	-0.64	0.09	-0.69	-1.59	0.41	0.64	0.61	1.87
8	-0.23	-0.23	-0.87	-1.82	-0.21	2.48	-0.05	-0.3	1.57	2.09	3.06	-9.93
9	-1.26	-4.62	3.05	6.91	4.82	-1.92	-5.32	5.66	5.34	0.6	14.1	32.79
10	-0.53	0.14	0.46	-0.82	4.26	4.05	3.07	-6.35	5.08	2.45	-9.49	-0.84
11	-5.63	-11.72	-2.77	-12.67	0.65	5.43	2.92	10.56	-6.24	2.05	-2.52	-32.16
12	-4.96	-5.39	-2.9	-5.9	1.43	-4.41	6.58	9.99	1.74	-0.92	-36.5	-21.9
13	1.32	0.49	0.11	-0.21	0.08	0.3	-0.94	-1	-0.37	-0.57	-1.14	2
14	0.49	2.28	0.01	0.1	-0.16	0.2	-0.98	-1.87	-0.45	-0.37	1.41	2.67
15	0.11	0.01	1.32	-0.01	1.09	0.29	0.14	-0.38	0.46	-0.54	-3.58	0.68
16	-0.21	0.1	-0.01	9.22	-0.36	-2.44	-1.97	-0.39	0.88	-0.73	0.99	2.93
17	0.08	-0.16	1.09	-0.36	8.8	2.55	4.76	-2.61	3.76	0.54	-9.32	-3.21
18	0.3	0.2	0.29	-2.44	2.55	11.99	1.5	-4.15	2.07	4.93	-7.08	-5.94
19	-0.94	-0.98	0.14	-1.97	4.76	1.5	40.97	-0.58	1.11	-0.29	3.03	-5.52
20	-1	-1.87	-0.38	-0.39	-2.61	-4.15	-0.58	76.4	-2.22	0.59	-12.01	-23.32
21	-0.37	-0.45	0.46	0.88	3.76	2.07	1.11	-2.22	18.79	0.66	-9.45	-3.03
22	-0.57	-0.37	-0.54	-0.73	0.54	4.93	-0.29	0.59	0.66	9.62	2.21	1.3
23	-1.14	1.41	-3.58	0.99	-9.32	-7.08	3.03	-12.01	-9.45	2.21	249.68	130.46
24	2	2.67	0.68	2.93	-3.21	-5.94	-5.52	-23.32	-3.03	1.3	130.46	216.68

Tabela 18: Teste Type 3 para efeitos fixos

FV	NumGL	DenGL	F	Pr>F
IND	17	68	9,43	<.0001
EXA	4	68	5,85	0.0004
LOCAL	23	68	72,91	<.0001
EXA*LOCAL	92	68	7,14	<.0001

Os resultados nos três modelos foram equivalentes no que tange aos teste de hipóteses: diferenças significativas entre examinadores, locais e interação. O que pode ser notado mediante o ajustamento de uma matriz de variâncias e covariâncias, é que a variabilidade entre os locais difere grandemente. Por exemplo, para o local 7 é 2.36 e para o local 11 é 425.10, quase duzentas vezes maior. Particularmente aqui, como a distância absorveu as duas coordenadas e os dois métodos, vamos fazer um comentário mais detalhado.

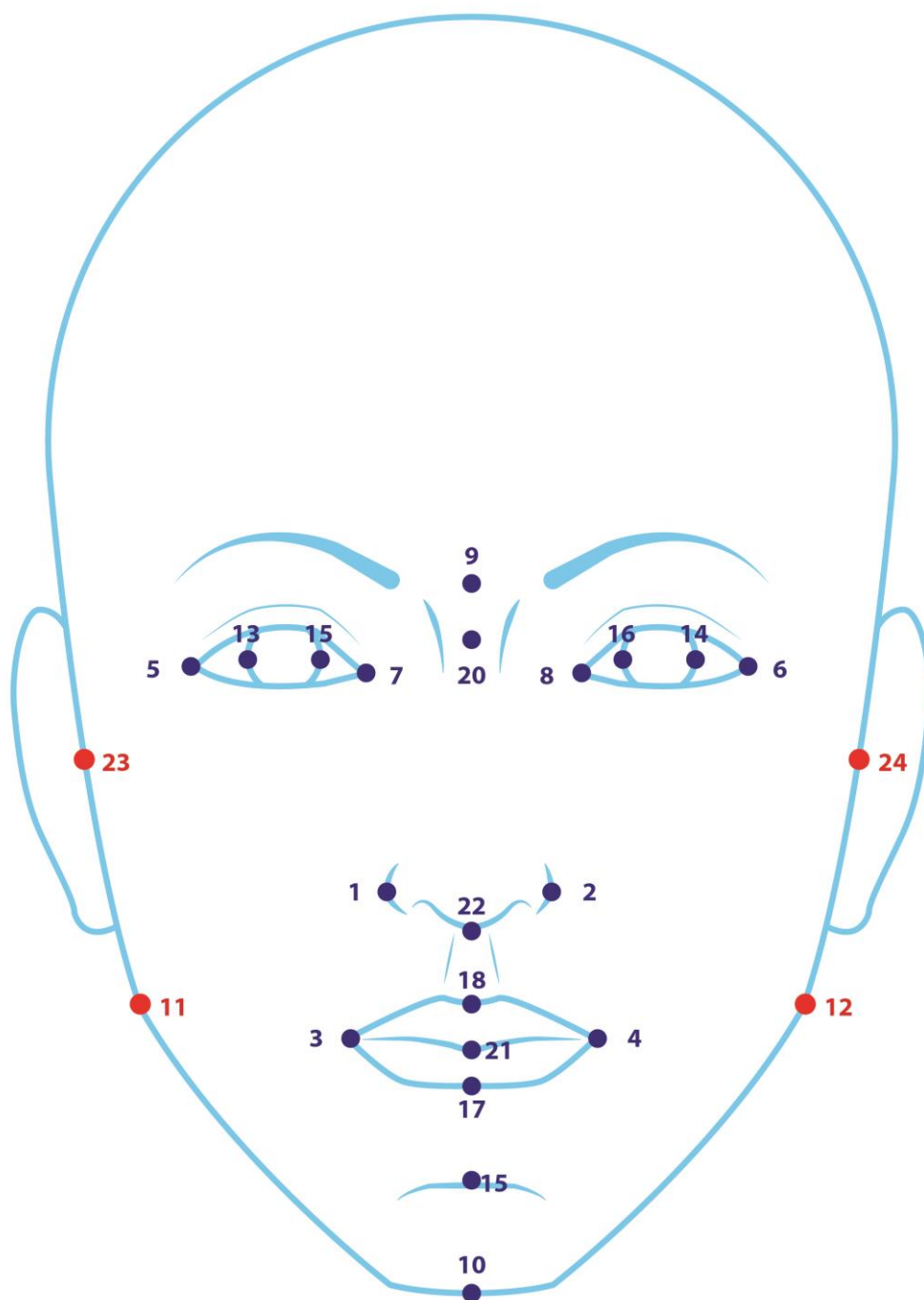
Primeiramente, vamos apresentar as médias das distâncias para a interação Examinadores x Locais. Estas médias estão na tabela 19. Quanto mais próxima de zero menor a diferença entre os métodos. O teste T da hipótese $H_0 : \text{dist} = 0$, $H_a : \text{dist} > 0$, feito pelo PROC MIXED para cada distância média indicou que se ela for maior ou igual a 3.20 então a hipótese nula é rejeitada e portanto houve diferença entre os métodos. Assim para os locais 13, 14 e 15, não houve diferença entre os métodos mas para os locais 11, 12, 23 e 24, as diferenças entre os métodos foram muito grandes. Este resultado reforça a existência das interações Examinadores x Métodos, Examinadores x Locais e Métodos x Locais.

Tabela 19: Distância média para cada examinador em cada local

distância média												
examinador	local											
	1	2	3	4	5	6	7	8	9	10	11	12
1	5.06	8.26	4.67	4.3	7.35	9.31	3.88	4.72	9.58	7.01	29.15	20.25
2	1.99	3.43	4.34	5.69	2.88	3.36	1.59	1.62	11.96	4.94	17.08	18.1
3	2.98	4.04	3.3	4.06	12.47	12.12	2.33	3.59	30.18	4.11	24.82	24.18
4	5.4	5.25	7.96	4.42	9.94	13.16	2.78	4.54	27.72	5.84	22.88	28.18
5	2.99	2.27	2.79	3.66	2.04	10.82	2.31	3.24	15.25	4.34	39.68	34.86

distância média												
examinador	local											
	13	14	15	16	17	18	19	20	21	22	23	24
1	1.91	2.4	1.56	1.67	3.47	4.73	9.7	10.98	2.24	3.59	26.23	29.85
2	1.81	1.34	1.67	1.26	5.67	6.55	12.67	17.46	5.28	4.29	21.47	15.22
3	2.52	2.27	1.66	2.11	5.59	4.07	8.16	5.99	5.68	2.38	13.94	16.36
4	2.74	2.93	1.75	4.27	4.25	11.53	10.21	19.42	3.23	2.95	27.32	21.81
5	2.79	2.29	1.27	1.58	3.2	2.55	8.81	18.05	2.21	3.21	21.55	20.98

Figura 17: Locais onde a as distâncias médias foi muito grande



4.3. CONCLUSÃO BASEADA NOS MODELOS

Pela análise dos dados deste experimento não casualizado com alguns possíveis modelos podemos fazer os seguintes comentários. Há diferença entre os métodos e interação entre os métodos e os examinadores, métodos e locais e examinadores e locais. Também a variação na coordenada Y é menor do que na coordenada X e grande variação entre o locais. Apenas para alguns locais a diferença entre os métodos é bem pequena e não significativa. O impacto destas conclusões parciais sobre os objetivos do pesquisador será explorado mais adiante.

É claro que todas as análises anteriores são análises univariadas desde que somente uma variável resposta foi colocada no modelo. Entretanto, os locais podem também serem considerados como respostas multivariadas do experimento e assim a análise de variância multivariada pode ser outro caminho. Por outro lado, as análises feitas neste capítulo admitiram normalidade dos dados e em alguns modelos, até variâncias homogêneas e isto ainda não foi estudado. No próximo capítulo trataremos destes dois aspectos.

5. ANÁLISE MULTIVARIADA

5.1. INTRODUÇÃO

Um experimento com dados repetidos no espaço, como o que esta sendo analisado, pode ser estudado sob o seguinte ponto de vista. O experimento diretor é composto dos dezoito indivíduos, que fazem o papel do blocos, os cinco examinadores e os dois métodos que são os tratamentos em forma de um fatorial 5 x 2. Pode-se considerar as observações em cada local como o de uma variável reposta do experimento e assim tem-se 24 variáveis resposta para a coordenada Y, outras 24 para X e 24 distâncias, ou seja, com esta perspectiva o modelo de análise passa a ser multivariado e neste caso, análise de variância multivariada.

O primeiro passo é definir o modelo para a ANOVA multivariada. O modelo básico que temos usado é o de blocos ao acaso univariado para Y

$$Y_{ijk} = \mu + E_i + M_j + (EM)_{ij} + I_k + \varepsilon_{ijk}$$

Sendo:

E_i o efeito do examinador i.

M_j o efeito do método j.

$(EM)_{ij}$ o efeito da interação examinadores e métodos.

I_k o efeito do individuo k.

O modelo univariado para o local o seguinte:

$$Y_{ijkl} = \mu_l + E_{il} + M_{jl} + (EM)_{ijl} + I_{kl} + \varepsilon_{ijkl}$$

e o modelo multivariado é construido pela junção dos 24 modelos univariados em um. Usando uma forma matricial, vamos definir o vetor \mathbf{Y}_l com o vetor com as observações do local l, \mathbf{X} como a matriz de incidência dos tratamentos, média e

blocos, β_1 o vetor que contem os efeitos dos tratamentos, efeitos dos blocos e a média para o local l e ε_1 o vetor dos erros do local l. Com esta notação o modelo multivariado é assim representado:

$$[Y_1 Y_2, \dots, Y_{24}] = [X\beta_1 X\beta_2, \dots, X\beta_{24}] + [\varepsilon_1 \varepsilon_2, \dots, \varepsilon_{24}] \text{ ou } Y = X\beta + \varepsilon$$

Sendo Y uma matriz 180 x 24, X uma matriz 180 x 24, β uma matriz 24 x 24 e ε uma matriz 180 x 24. Também as matrizes X e β podem ser particionadas para representar separadamente a média, os efeitos de tratamentos e os efeitos de blocos:

$$Y = 1\mu + X_t t + X_b b + \varepsilon$$

ou mais ainda para separar examinadores (t_1), métodos (t_2) e interação (t_3) e suas respectivas matrizes,

$$Y = 1\mu + X_1 t_1 + X_2 t_2 + X_3 t_3 + X_b b + \varepsilon$$

Todos os componentes deste modelo são matrizes com suas respectivas dimensões.

A teoria da análise multivariada não vai ser desenvolvida aqui e pode ser encontrada em muitos textos como o de Mardia *et al.* (1908) entretanto, vamos descrever de forma sucinta como funciona a ANOVA multivariada.

A tabela da ANOVA univariada é formada por soma de quadrados enquanto que a multivariada é formada por matrizes cujos elementos são somas de quadrados e somas de produtos. É então:

Tabela 20: ANOVA Multivariada

FV	GL	MSQP
INDIVÍDUOS	17	S_1
EXAMINADORES	4	S_2
MÉTODOS	1	S_3
INTERAÇÃO	4	S_4
ERRO	153	S_5
TOTAL	179	T

Sendo $MSQP$ as respectivas matrizes de soma de quadrados e produtos para cada fonte de variação. Na diagonal de cada matriz esta a soma de quadrados para a respectiva fonte de variação para cada local e fora da diagonal a soma de produtos cruzados entre um local e outro. As matrizes S representam as fontes de variação que vão ser submetidas aos testes de hipóteses e, assim como na análise univariada, a matriz T é a soma de todas as outras.

5.2. INFERÊNCIA EM ANÁLISE MULTIVARIADA

Na ANOVA univariada de experimentos, as observações são assumidas serem independentes e com distribuição normal univariada, e na ANOVA multivariada as linha da matriz Y é que são independentes, com distribuição normal multivariada. Para representar melhor esta parte é comum também nos textos definir a matriz Y da seguinte forma:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}$$

sendo y_i o vetor linha da observação multivariada i , $i=1,2,\dots,n$. No experimento aqui analisado, cada linha tem 24 observações (24 locais) e $n=180$ observações multivariadas. Assim dizemos que y_i é iid $\approx N(\mu_i, \Sigma)$. A matriz Σ tem dimensão 24×24 com variâncias e covariâncias diferentes e, na linguagem de modelos mistos é chamada de matriz de variâncias e covariâncias não estruturada. Esta é a suposição para a ANOVA multivariada.

A inferência primeira é em relação a β do modelo $Y = X\beta + \varepsilon$, a estimativa de β é formada pelas estimativas individuais de quadrados mínimos de cada componente de β , no caso 24 betas e podem ser usados nas inferências individuais para cada local. Isto é

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}$$

A segunda inferência diz respeito aos testes de hipóteses que na ANOVA univariada é escrita da forma $H_0: \mathbf{C}\beta = \mathbf{0}$, $H_a: \mathbf{C}\beta \neq \mathbf{0}$ e o teste F é o aplicado para o teste. Na análise multivariada, hipóteses deste tipo também podem ser testadas mas a forma geral dos testes é a seguinte :

$$H_0: \mathbf{C}\beta\mathbf{L} = \mathbf{0}, H_a: \mathbf{C}\beta\mathbf{L} \neq \mathbf{0}$$

A matriz \mathbf{C} define os testes entre os parâmetros de cada local separadamente e a matriz \mathbf{L} define os testes entre locais que é o principal ganho da análise multivariada. Como exemplo, suponha que queiramos testar a diferença entre os dois métodos para cada local ou seja o método 1 é igual ao método 2 em todos os locais; então vamos usar $H_0: \mathbf{C}\beta = \mathbf{0}$, $H_a: \mathbf{C}\beta \neq \mathbf{0}$, para uma especificada matriz \mathbf{C} . Por outro lado, se a hipótese for que a diferença dos dois métodos é uma no local 1 e outra no local 2 (interação métodos locais) então a hipótese é da forma $H_0: \mathbf{C}\beta\mathbf{L} = \mathbf{0}$, $H_a: \mathbf{C}\beta\mathbf{L} \neq \mathbf{0}$, para uma certa matriz \mathbf{L} . Por sua vez assim como na análise univariada existem duas estatísticas básicas para um teste multivariado.

Suponha que o teste seja da forma $H_0: \mathbf{C}\beta\mathbf{L} = \mathbf{0}$, $H_a: \mathbf{C}\beta\mathbf{L} \neq \mathbf{0}$. Então as duas estatísticas são:

$$\mathbf{H} = \mathbf{L}'(\mathbf{C}\hat{\beta})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\hat{\beta})\mathbf{L}$$

$$\mathbf{E} = \mathbf{L}'(\mathbf{Y}'\mathbf{Y} - \hat{\beta}'(\mathbf{X}'\mathbf{X})\hat{\beta})\mathbf{L}$$

semelhantes às aquelas da análise univariada.

Os testes multivariados mais comuns são os seguintes :

$$\text{Lambda de Wilks} = \frac{\det(\mathbf{E})}{\det(\mathbf{H} + \mathbf{E})}$$

$$\text{Traço de Pillai} = \text{traço}(\mathbf{H}(\mathbf{H} + \mathbf{E})^{-1})$$

Traço de Hotelling-Lawley = $\text{traço}(\mathbf{E}^{-1}\mathbf{H})$

Teste de Roy = maior autovalor de $(\mathbf{E}^{-1}\mathbf{H})$

Para o experimento que esta sendo analisado, para cada fonte de variação existem as matrizes **C**, **L**, **H** e **E** usadas nos 4 testes. Alguns deste testes são aproximados e outros exatos, dependendo dos parâmetros do teste; o SAS aplica o exato quando existe e quando não, o aproximado. Vamos aplicá-los ao experimento deste trabalho mediante o PROC GLM do SAS. Os resultados foram os seguintes:

Para a coordenada Y:

Tabela 21: Teste para os efeitos dos Examinadores em Y

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilk's Lambda	0,026902	8,04	96	517,51	<.0001
Pillai's Trace	2,285929	7,39	96	532	<.0001
Hotelling-Lawley Trace	6,489152	8,69	96	422,08	<.0001
Roy's Greatest Root	3,018715	16,73	24	133	<.0001

Tabela 22: Testes para os efeitos dos Métodos em Y

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilk's Lambda	0,180869	24,53	24	130	<.0001
Pillai's Trace	0,819132	24,53	24	130	<.0001
Hotelling-Lawley Trace	4,528879	24,53	24	130	<.0001
Roy's Greatest Root	4,528879	24,53	24	130	<.0001

Tabela 23: Testes para a interação Examinador*Método em Y

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilk's Lambda	0,0805462	4,79	96	517,51	<.0001
Pillai's Trace	1,82279935	4,64	96	532	<.0001
Hotelling-Lawley Trace	3,67493404	4,92	96	422,08	<.0001
Roy's Greatest Root	1,51523272	8,4	24	133	<.0001

Tabela 24: Testes para os efeitos dos Locais em Y

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilks' Lambda	0	1222081	23	131	<.0001
Pillai's Trace	1	1222081	23	131	<.0001
Hotelling-Lawley Trace	214563,8718	1222081	23	131	<.0001
Roy's Greatest Root	214563,8718	1222081	23	131	<.0001

Tabela 25: Testes para a interação Local*Examinador em Y

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilks' Lambda	0,03098031	7,96	92	521,01	<.0001
Pillai's Trace	2,23372251	7,37	92	536	<.0001
Hotelling-Lawley Trace	6,0550133	8,53	92	422,76	<.0001
Roy's Greatest Root	2,68737859	15,66	23	134	<.0001

Tabela 26: Teste para a interação Local*Métodos em Y

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilks' Lambda	0,18163155	25,66	23	131	<.0001
Pillai's Trace	0,81836845	25,66	23	131	<.0001
Hotelling-Lawley Trace	4,50565152	25,66	23	131	<.0001
Roy's Greatest Root	4,50565152	25,66	23	131	<.0001

Tabela 27: Testes para a interação Local*Examinador*Método em Y

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilks' Lambda	0,0966671	4,56	92	521,01	<.0001
Pillai's Trace	1,71740215	4,38	92	536	<.0001
Hotelling-Lawley Trace	3,35340012	4,72	92	422,76	<.0001
Roy's Greatest Root	1,50356047	8,76	23	134	<.0001

Para a coordenada X:

Tabela 28: Testes para o efeito dos Examinadores em X

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilks' Lambda	0,01102377	11,43	96	517,51	<.0001
Pillai's Trace	2,67059516	11,13	96	532	<.0001
Hotelling-Lawley Trace	8,70005102	11,66	96	422,08	<.0001
Roy's Greatest Root	3,49206376	19,35	24	133	<.0001

Tabela 29: Testes para o efeito de Métodos em X

Statistic	Value	F	NumDF	DenDF	Pr>F
Wilks' Lambda	0,13793662	33,85	24	130	<.0001
Pillai's Trace	0,86206338	33,85	24	130	<.0001
Hotelling-Lawley Trace	6,24970662	33,85	24	130	<.0001
Roy's Greatest Root	6,24970662	33,85	24	130	<.0001

Tabela 30: Testes para a interação Examinador*Método em X

Statistic	Value	F	NumDF	DenDF	Pr>F
Wilks' Lambda	0,02919966	7,76	96	517,51	<.0001
Pillai's Trace	2,28843046	7,41	96	532	<.0001
Hotelling-Lawley Trace	6,059604	8,12	96	422,08	<.0001
Roy's Greatest Root	2,71774439	15,06	24	133	<.0001

Tabela 31: Testes para os efeitos dos Locais em X

Statistic	Value	F	NumDF	DenDF	Pr.F
Wilks' Lambda	0	1279778	23	131	<.0001
Pillai's Trace	1	1279778	23	131	<.0001
Hotelling-Lawley Trace	224693,9	1279778	23	131	<.0001
Roy's Greatest Root	224693,9	1279778	23	131	<.0001

Tabela 32: Testes para a interação Local*Examinadores em X

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilks' Lambda	0,01204286	11,63	92	521,01	<.0001
Pillai's Trace	2,63407668	11,24	92	536	<.0001
Hotelling-Lawley Trace	8,48240889	11,95	92	422,76	<.0001
Roy's Greatest Root	3,47826977	20,26	23	134	<.0001

Tabela 33: Testes para a interação Local*Métodos em X

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilks' Lambda	0,14054833	34,83	23	131	<.0001
Pillai's Trace	0,85945167	34,83	23	131	<.0001
Hotelling-Lawley Trace	6,11499038	34,83	23	131	<.0001
Roy's Greatest Root	6,11499038	34,83	23	131	<.0001

Tabela 34: Testes para a interação Local*Examinador*Métodos em X

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilks' Lambda	0,03552809	7,5	92	521,01	<.0001
Pillai's Trace	2,17708244	6,96	92	536	<.0001
Hotelling-Lawley Trace	5,70193101	8,03	92	422,76	<.0001
Roy's Greatest Root	2,68314543	15,63	23	134	<.0001

Para a distância entre métodos:

Tabela 35: Testes para o efeito dos Examinadores na distancia

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilks' Lambda	0,00224524	6,9	96	180,79	<.0001
Pillai's Trace	3,08949163	6,79	96	192	<.0001
Hotelling-Lawley Trace	15,21337115	6,92	96	135,44	<.0001
Roy's Greatest Root	6,00815944	12,02	24	48	<.0001

Tabela 36: Testes para o efeito dos Locais na distancia

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilks' Lambda	0,0300799	64,49	23	46	<.0001
Pillai's Trace	0,9699201	64,49	23	46	<.0001
Hotelling-Lawley Trace	32,24478663	64,49	23	46	<.0001
Roy's Greatest Root	32,24478663	64,49	23	46	<.0001

Tabela 37: Testes para a interação Local*Examinadores na distancia

Statistic	Value	F Value	NumDF	DenDF	Pr>F
Wilks' Lambda	0,00287567	6,79	92	184,57	<.0001
Pillai's Trace	3,02618138	6,62	92	196	<.0001
Hotelling-Lawley Trace	14,22507307	6,9	92	137,98	<.0001
Roy's Greatest Root	5,97287928	12,72	23	49	<.0001

Os resultados das análise de variância multivariada feitos aqui confirmam os resultados das análises univariadas: todas as fontes de variação importantes do experimento foram significativas. O ponto positivo sobre os resultados da análise univariada foi o de testar todas as hipóteses com uma matriz de variâncias e covariâncias não estruturada, considerando o fato de que elas diferem muito de um local para o outro. Além disto, devemos levar em conta que os testes multivariados são mais conservadores do que os testes univariados.

Com estes resultados, qual a conclusão sobre o experimento? Discutiremos este aspecto na parte final do trabalho.

6. TESTE PARA NORMALIDADE

A análise multivariada assume que a distribuição é normal, y_i é iid $\approx N(\mu_i, \Sigma)$, isto é, cada conjunto dos 24 locais tem distribuição normal com variâncias e covariâncias iguais. Até aqui não discutimos este aspecto nem para a análise univariada nem para a multivariada. Como o experimento não foi casualizado deve-se pesquisar investigar tanto uma como a outra.

Para o caso da análise de variância multivariada existem alguns estudos dispersos na literatura mas em todos começa com a análise de cada variável separadamente.

Embora distribuição marginal normal não implique em distribuição multivariada normal, como ressalta Gnanadesikan (1977), frequentemente a não normalidade multivariada é causada por uma anomalia em uma das marginais. Assim é necessário primeiro o estudo de cada uma e para isto existem testes para normalidade como o teste de Shaphiro-Wilk, gráficos como o Q-Q, histogramas, etc. No caso univariado, usando o PROC MIXED do SAS com a opção INFLUENCE, foram construídos os histogramas e os gráficos Q-Q para os erros estimados pelo modelo univariado de cada local.

Para analisar a normalidade destes pontos foram considerados gráficos Q-Q, histogramas e gráficos de resíduos, todos estes gráficos se encontram no apêndice deste trabalho.

6.1. Normalidade em Y

A seguir serão discutidos os 24 pontos para a variável Y (eixo y) quanto à normalidade.

Dos 24 pontos analisados, 9 não apresentaram indicação de normalidade: ENDOCANTHION ESQUERDO, GLABELA, IRIDION MEDIAL ESQUERDO*, LABIAL INFERIOR*, LABIAL SUPERIOR, NASION, STOMION*, SUBNASAL*, ZYGION ESQUERDO.

Figura 18: Análise de Normalidade do ponto Endocanthion Esquerdo

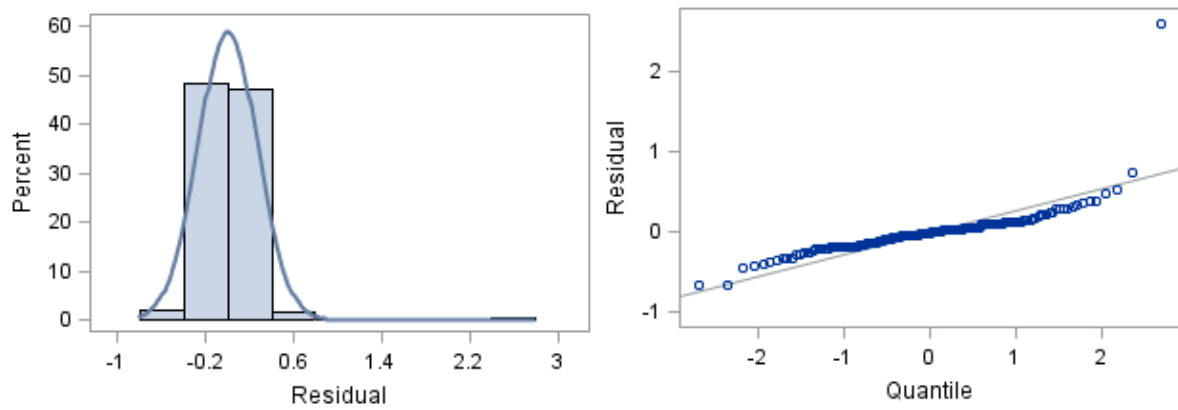


Figura 19: Análise de Normalidade do ponto Glabella

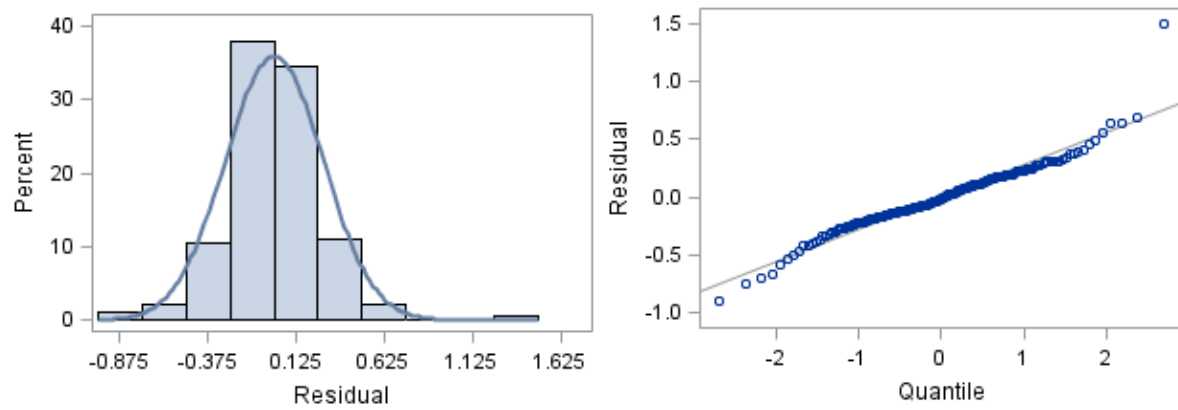


Figura 20: Análise de Normalidade do ponto Iridion Medial Esquerdo

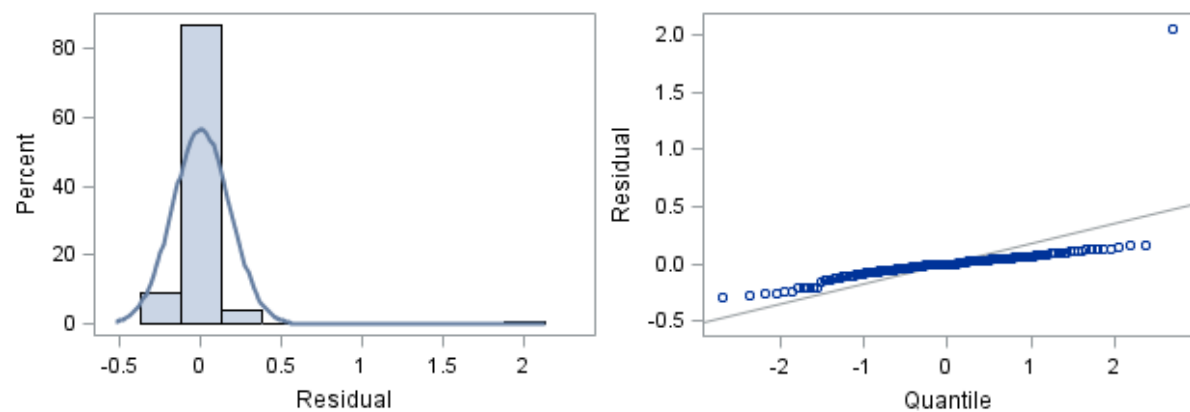


Figura 21: Análise de Normalidade do Ponto Labial Inferior

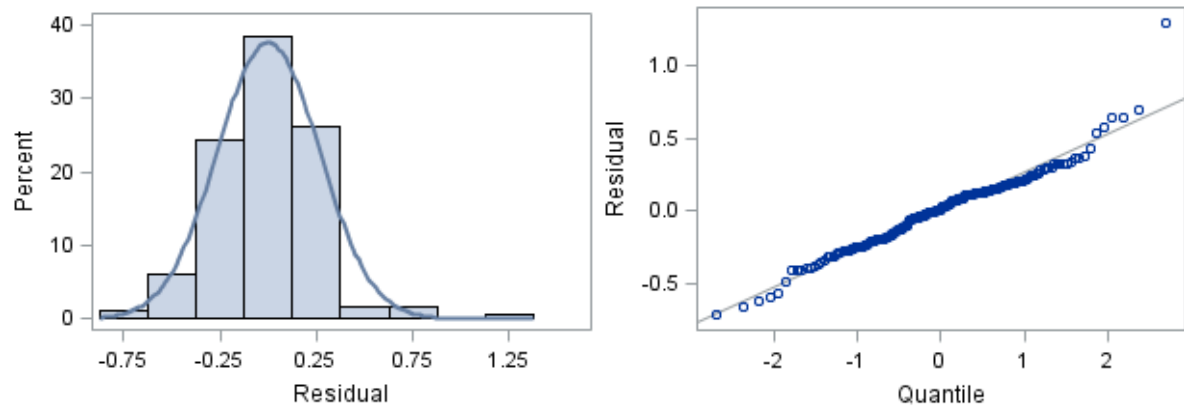


Figura 22: Análise de Normalidade do Ponto Labial Superior

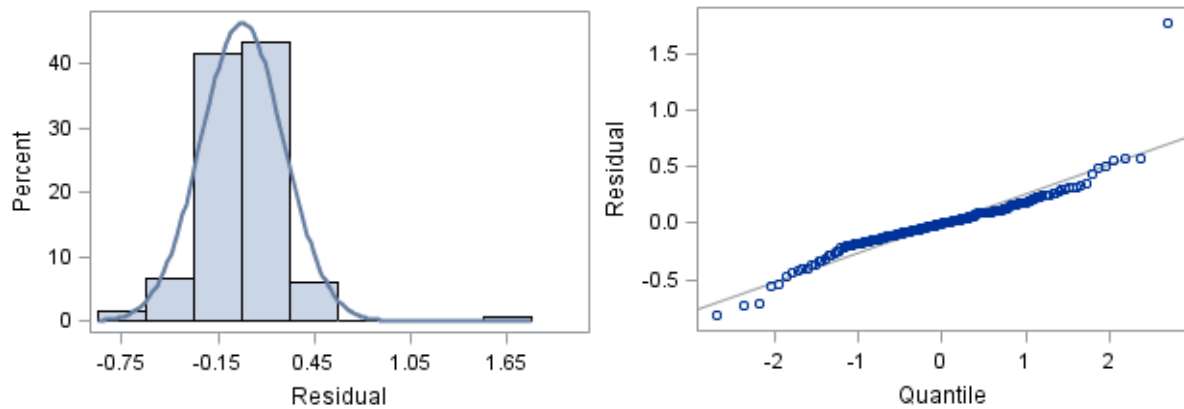


Figura 23: Análise de Normalidade do Ponto Nasion

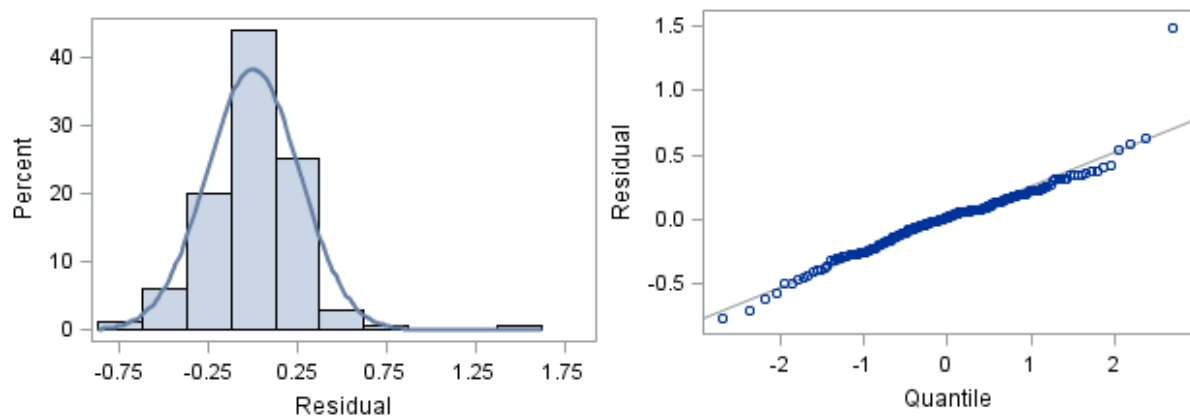


Figura 24: Análise de Normalidade do Ponto Stomion

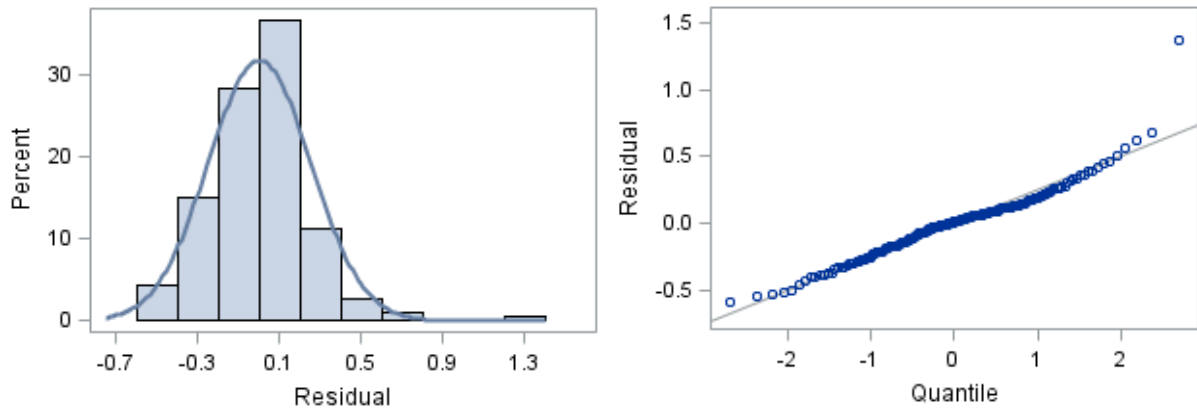


Figura 25: Análise de Normalidade do Ponto Subnasal

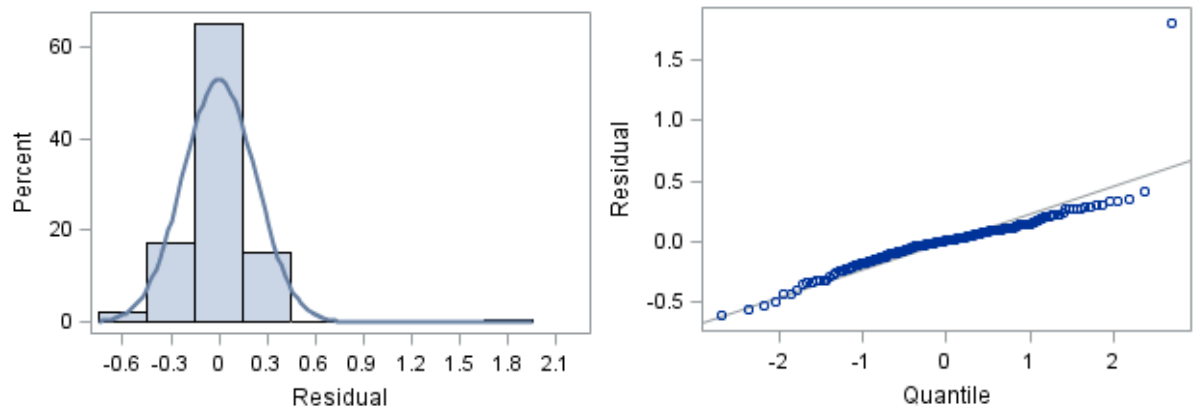
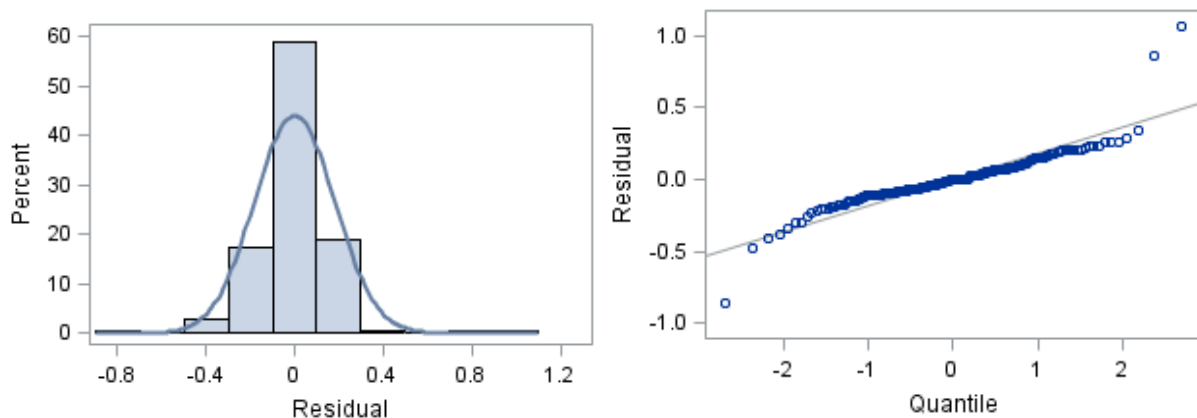


Figura 26: Análise de Normalidade do Ponto Zygion Esquerdo



Os pontos faciais que estão destacados com um asterisco (*) representando locais em que apenas um ponto muito discrepante (outlier) influenciou de maneira decisiva para a não normalidade neste local. Este outlier pode ter sido causado por um erro na calibragem do programa ou um erro pessoal causado pelo próprio examinador.

6.2. Normalidade em X

A seguir serão discutidos os 24 pontos para a variável X (eixo x) quanto à normalidade.

Dos 24 pontos analisados, 6 não apresentaram indicação de normalidade: IRIDION MEDIAL ESQUERDO*, LABIAL INFERIOR*, NASION, STOMION*, SUBNASAL*.

Figura 27: Análise de Normalidade do ponto Iridion Medial Esquerdo

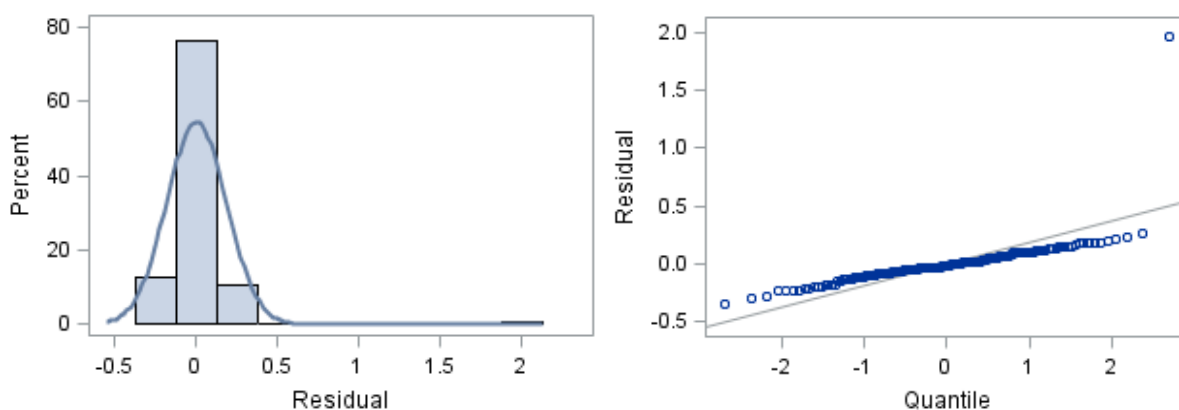


Figura 28: Análise de Normalidade para o ponto Labial Inferior

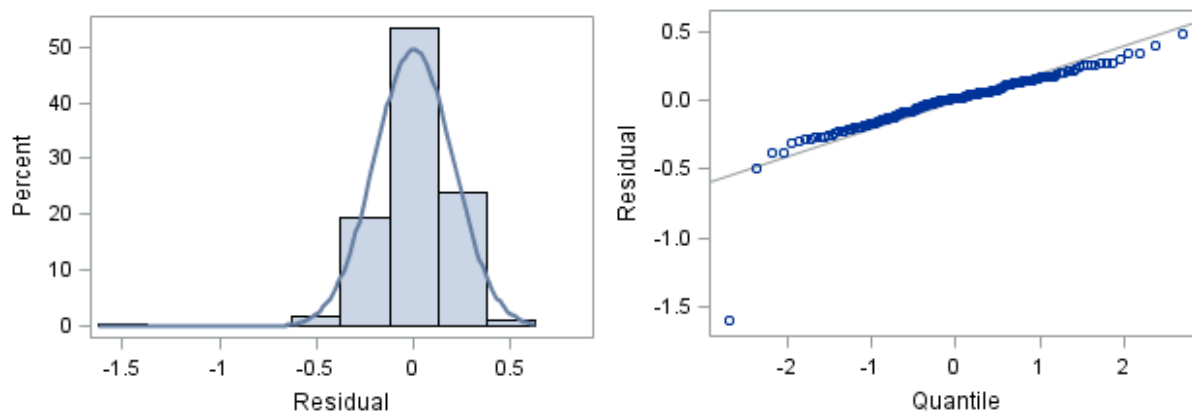


Figura 29: Análise de normalidade para o Ponto Nasion

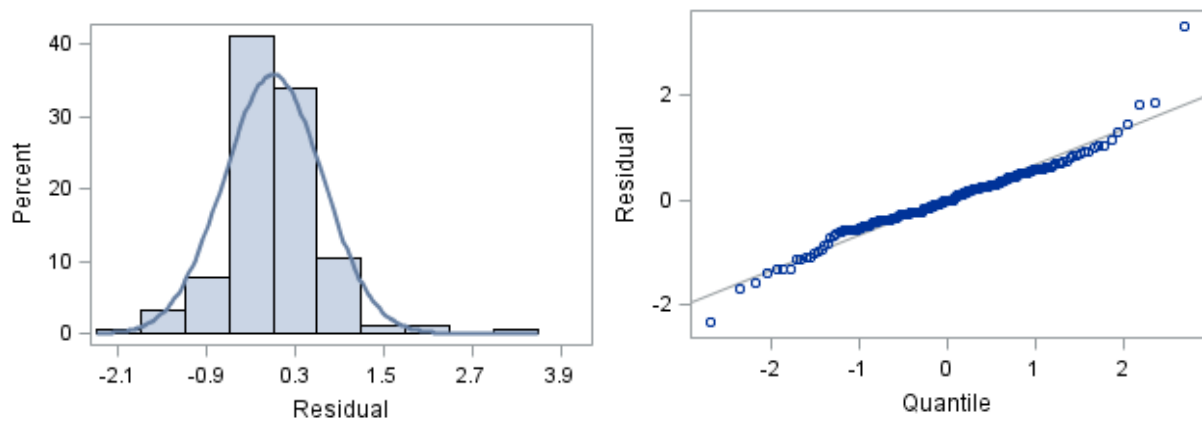


Figura 30: Análise de Normalidade para o ponto Stomion

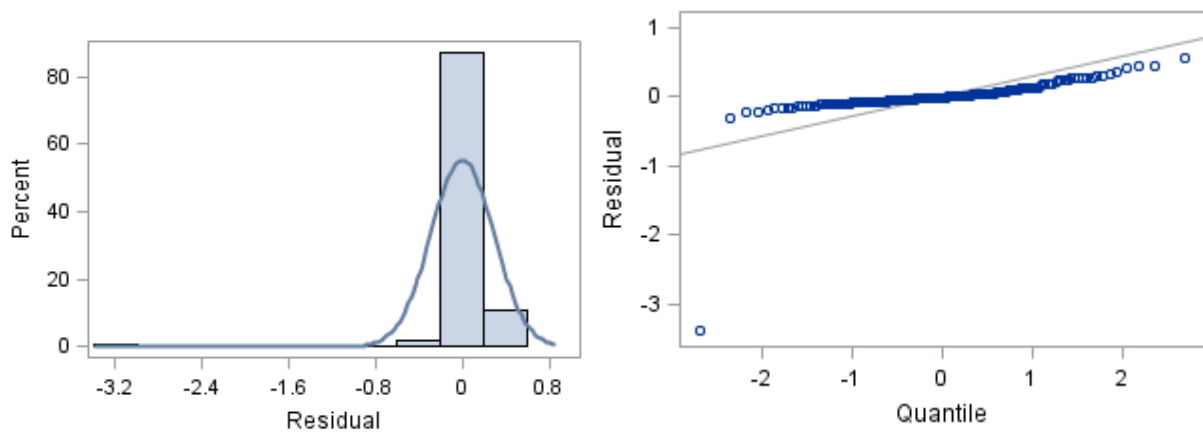
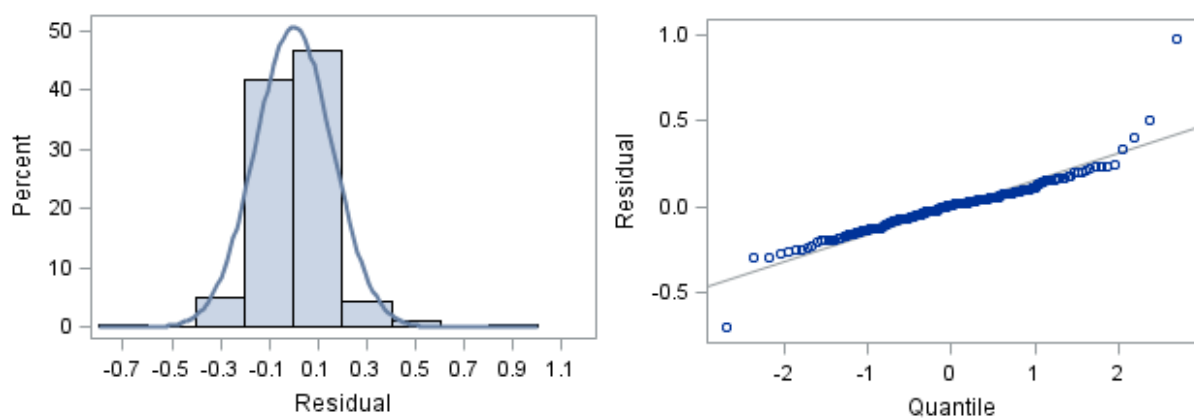


Figura 31: Análise de Normalidade para o ponto Subnasal



Os pontos faciais que estão destacados com um asterisco (*) representando locais em que apenas um ponto muito discrepante (outlier) influenciou de maneira decisiva para a não normalidade neste local. Este outlier pode ter sido causado por

um erro na calibragem do programa ou um erro pessoal causado pelo próprio examinador.

6.3. Normalidade na distância entre pontos

Como dito anteriormente, a distância entre pontos é obtida a partir da diferença entre as marcações de um ponto pelas duas metodologias estudadas. A seguir serão apresentadas as análises referentes à normalidade para esta variável criada.

Dos 24 pontos analisados, 6 não apresentaram indicação de normalidade: ENDOCANTHION DIREITO, ENDOCANTHION ESQUERDO, IRIDION MEDIAL ESQUERDO*, LABIAL INFERIOR*, STOMION*, SUBNASAL*.

Figura 32: Análise de Normalidade para o ponto Endocanthion Direito

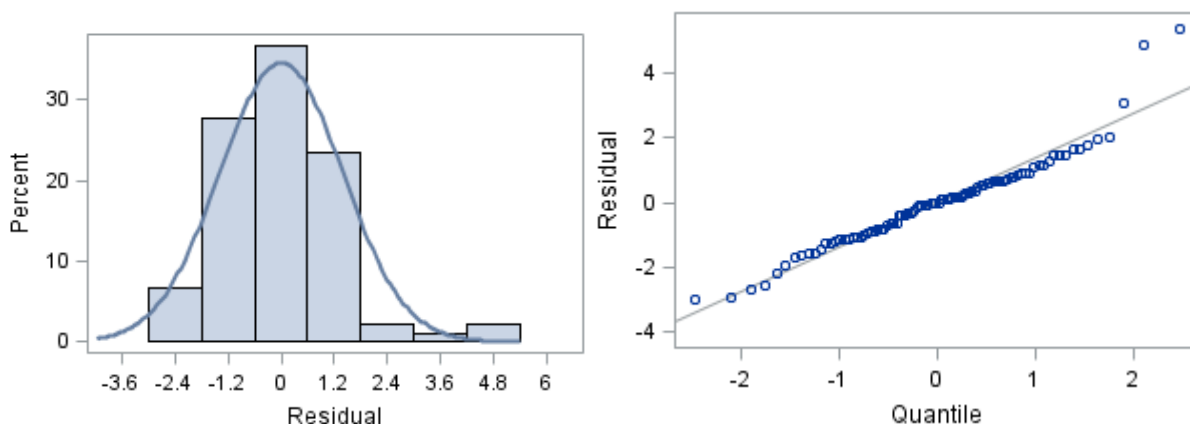


Figura 33: Análise de Normalidade para o ponto Endocanthion Esquerdo

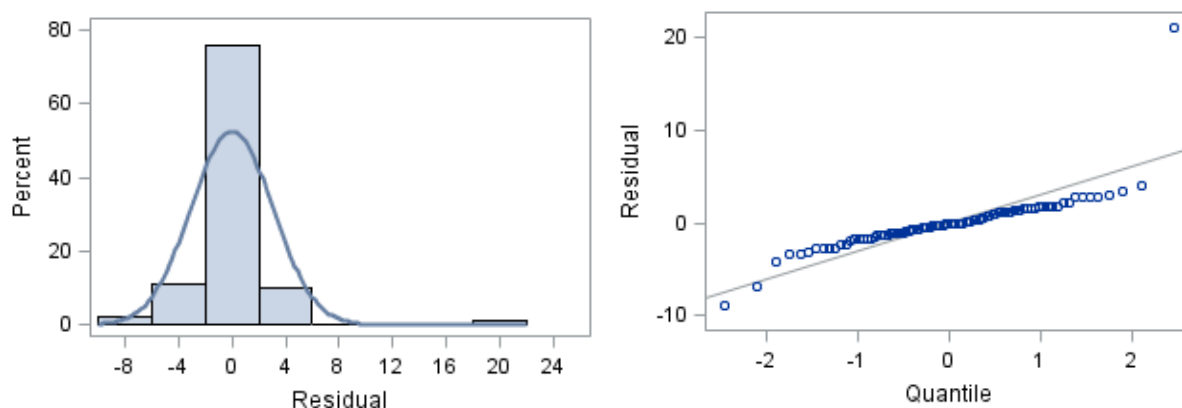


Figura 34: Análise de Normalidade para o ponto Iridion Medial Esquerdo

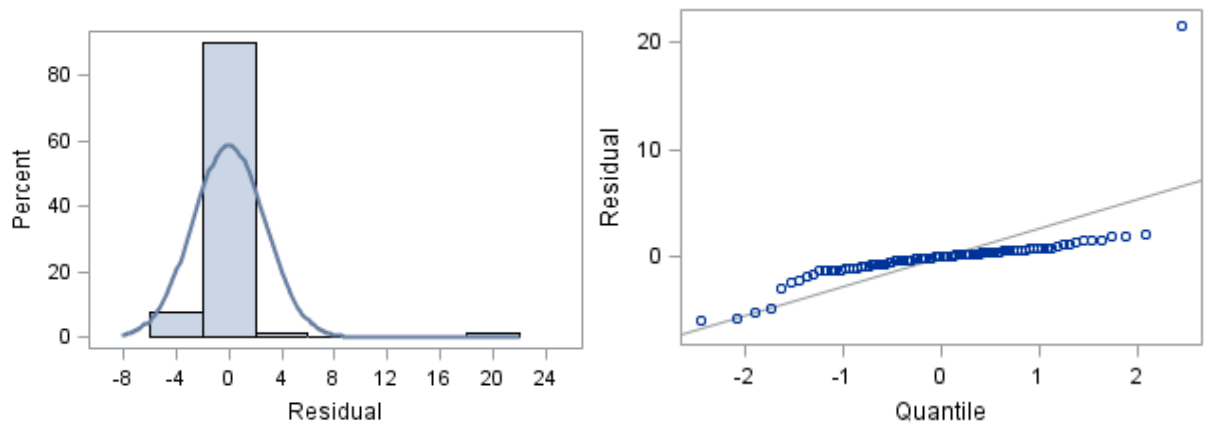


Figura 35: Análise de Normalidade para o ponto Labial Inferior

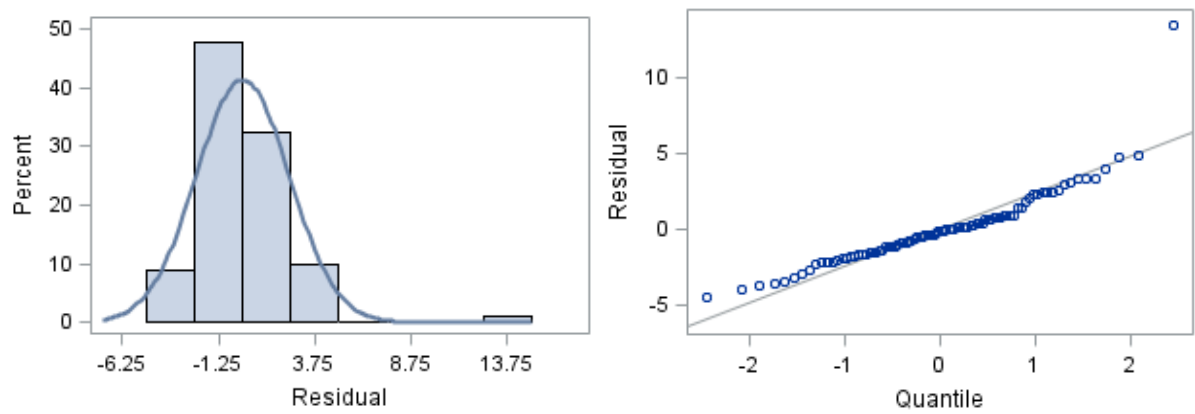


Figura 36: Análise de Normalidade para o ponto Stomion

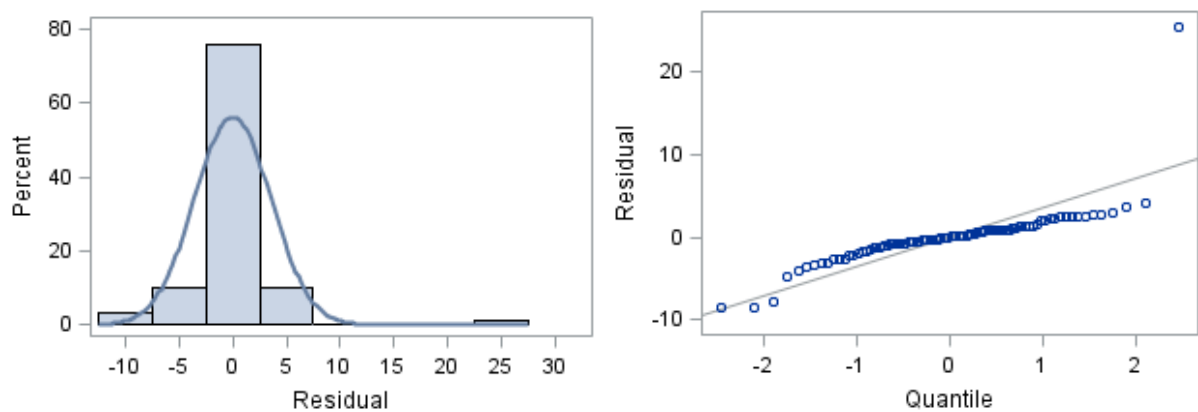
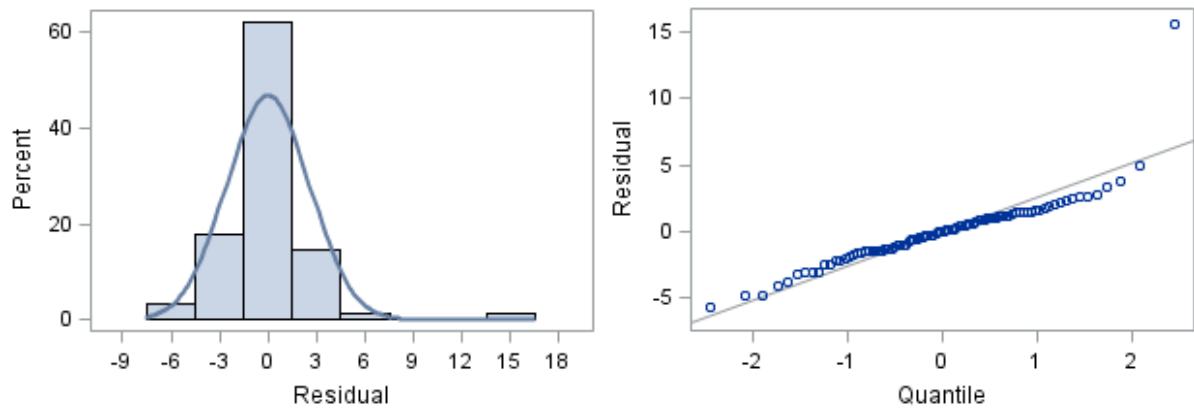
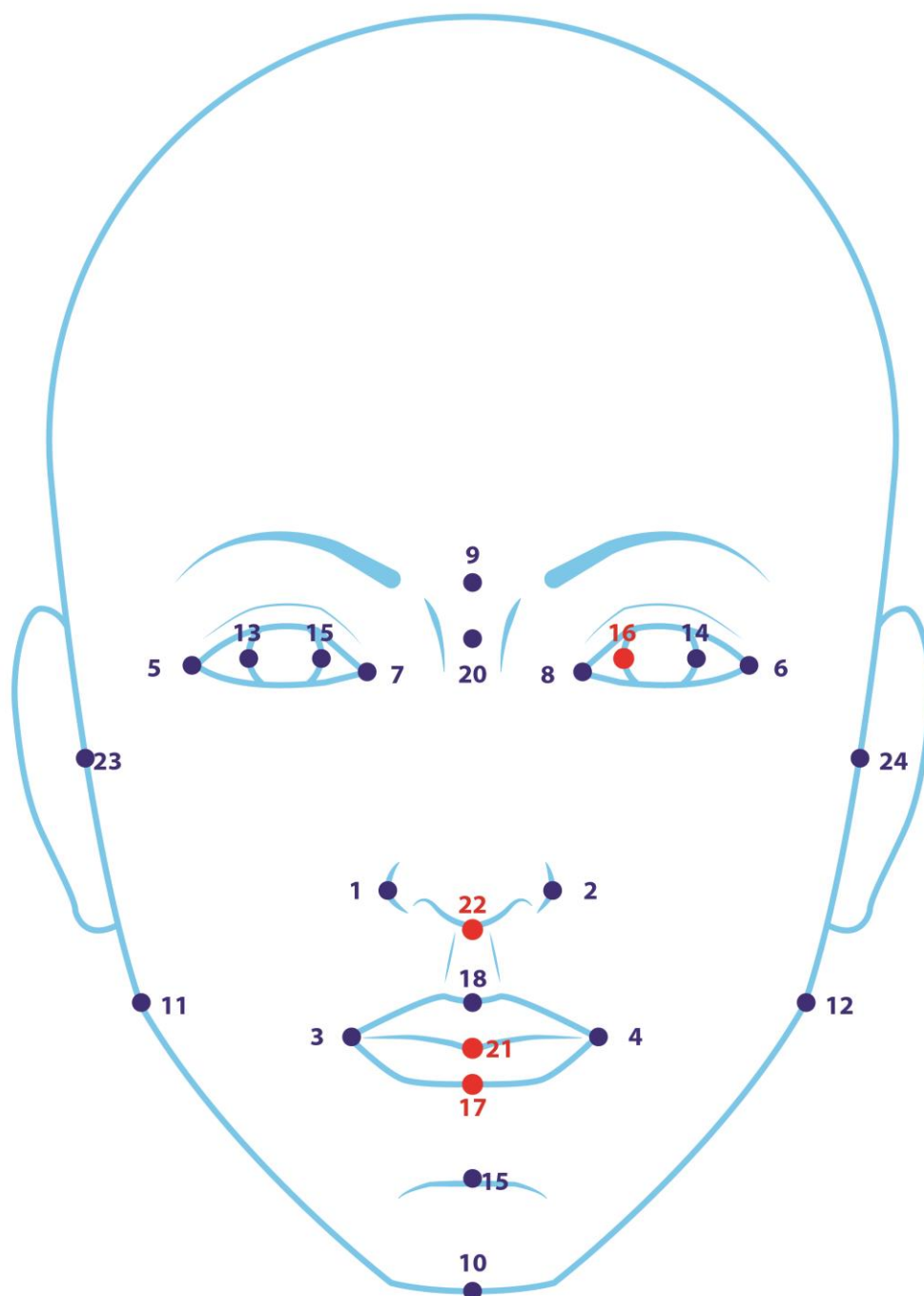


Figura 37: Análise de Normalidade para o ponto Subnasal



Os pontos faciais que estão destacados com um asterisco (*) representando locais em que apenas um ponto muito discrepante (outlier) influenciou de maneira decisiva para a não normalidade neste local. Este outlier pode ter sido causado por um erro na calibragem do programa ou um erro pessoal causado pelo próprio examinador.

Figura 38: Locais onde a normalidade foi prejudicada por um outlier para distância



Como foi dito anteriormente, mesmo que em todos os 24 locais os erros estimados seguissem a distribuição normal, seria necessário investigar se a distribuição multivariada é normal. Entretanto, segundo Vonesh e Chinchilli (1997), para grandes amostras os estimadores de interesse, tem distribuição aproximadamente multivariada normal pelo teorema central do limite multivariado e, além disto, sob condições não muito restritivas os quatro testes multivariados usados aqui, são relativamente robustos à não normalidade. É claro também que este resultado vale para a análise univariada. No experimento em questão, o número de observações 180, pode ser considerado grande e assim os testes serem válidos. Todavia, vamos aplicar um teste para normal multivariada.

Um teste conhecido e mais simples para testar normal multivariada é o teste baseado nas medidas de assimetria e curtose multivariada de Mardia (1980). Sob o ponto de vista teórico $\beta_{1,p} = E(((\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}))^3)$ e

$\beta_{2,p} = E(((\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}))^2)$ são respectivamente as medidas de assimetria e curtose multivariada de Mardia. Se a distribuição é normal então $\beta_{1,p} = 0$ e $\beta_{2,p} = p(p+2)$. Com uma amostra de n observações multivariadas, as estimativas dos betas são as seguintes:

$$\hat{\beta}_{1,p} = \frac{1}{n^2} \sum_i \sum_j g_{ij}^3$$

$$\hat{\beta}_{2,p} = \frac{1}{n} \sum_i g_{ii}^2,$$

$$g_{ij} = (\mathbf{y}_i - \mathbf{1}\bar{y}_i)' \mathbf{S}_n^{-1} (\mathbf{y}_i - \mathbf{1}\bar{y}_i)$$

$$\mathbf{S}_n = \frac{1}{n} \mathbf{Y}' (\mathbf{I}_n - \frac{1}{n} \mathbf{1}\mathbf{1}') \mathbf{Y} \text{ (matriz dos erros)}$$

Segundo Mardia(1980), para grandes amostras, $\kappa_1 = \frac{n\hat{\beta}_{1,p}}{6}$ tem distribuição aproximada de χ^2 com graus de liberdade = $p(p+1)(p+2)/6$ e a hipótese nula de normalidade multivariada é rejeitada para valores grandes de κ_1 .

$\kappa_2 = (\hat{\beta}_{2,p} - p(p+2))/\sqrt{(8p(p+2)/n)}$ tem distribuição aproximada $N(0,1)$ e hipótese nula de normalidade multivariada é rejeitada para valores críticos de κ_2 .

O teste para normalidade multivariada de assimetria e de curtosis de Mardia foi adaptado por Vonesh e Vernon(1997) para condições mais gerais da ANOVA e funciona da seguinte forma. Se o modelo multivariado é dado por

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

então é conhecido que os erros na ANOVA são gerados pela seguinte expressão :

$$\hat{\boldsymbol{\varepsilon}} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}$$

Entretanto, a parte dos dados que realmente geram os erros esta no espaço do erro, que tem dimensão $(n-g)$ igual ao grau de liberdade do erro da ANOVA. Segundo Vonesh e Vernon(1997), o teste de normalidade multivariada deve ser feito com os erros gerados deste espaço que não são contaminados pelos efeitos fixos (semelhante ao argumento da verossimilhança restrita ou residual). Assim sendo constrói-se uma matriz \mathbf{H} , de dimensão $(n-g) \times n$, de posto $(n-g)$ tal que $\mathbf{H}\mathbf{X} = \mathbf{0}$ e $\mathbf{H}\mathbf{H}' = \mathbf{I}$. Os erros agora são obtidos pela expressão

$$\hat{\boldsymbol{\varepsilon}}_V = \mathbf{H}\mathbf{Y}$$

No experimento que esta sendo analisado, $n=180$, $g=27$ e $p=24$ e a matriz dos erros agora tem dimensão 153×24 . Resta agora testar a normalidade multivariada dos 24 vetores usando os testes de Mardia. Por um programa SAS com o IML o teste foi aplicado para a coordenada Y e os resultados foram os seguintes:

$$\kappa_1 = \frac{n\hat{\beta}_{1,p}}{6} = 2530.7934 \text{ com } p\text{-valor } 0.8312475$$

$$\kappa_2 = (\hat{\beta}_{2,p} - p(p+2)) / \sqrt{(8p(p+2)/n)} = -1.734112 \text{ e } p_valor \text{ } 0.0828983$$

E assim a hipótese de normalidade multivariada para Y é aceita.

Para a coordenada X os resultados foram os seguintes:

$$\kappa_1 = \frac{n\hat{\beta}_{1,p}}{6} = 2620.0844 \text{ com } p_valor \text{ } 0.3870384$$

$$\kappa_2 = (\hat{\beta}_{2,p} - p(p+2)) / \sqrt{(8p(p+2)/n)} = 0.006589$$

e também aceitamos a hipótese nula.

Para a distância a $n=90$, $g=22$ e $p=24$ e a matriz dos erros agora tem dimensão 68×24 e os resultados foram :

$$\kappa_1 = \frac{n\hat{\beta}_{1,p}}{6} = 2536.7737 \text{ com } p_valor \text{ } 0.8091412$$

$$\kappa_2 = (\hat{\beta}_{2,p} - p(p+2)) / \sqrt{(8p(p+2)/n)} = -1.755153 \text{ e } p_valor \text{ } 0.0792331$$

Aqui também aceitamos a normalidade multivariada.

7. Conclusão

A análise de experimentos casualizados segue um protocolo bem definido principalmente pela ANOVA de Fisher, acompanhado de comparações de médias e outras técnicas pertinentes à análise de variância. Entretanto, quando o experimento não foi casualizado a análise deve seguir um caminho mais longo para que seus resultados possam ser confiáveis. A casualização define a estrutura de variâncias e covariâncias dos dados e o modelo de análise para a ANOVA e se não houve casualização, tudo isto deve passar por um estudo cuidadoso.

Nesta monografia o experimento que foi analisado não foi casualizado, haviam 2 variáveis resposta e ainda os dados foram repetidos no espaço, no caso 24 locais da face de uma pessoa. O objetivo do trabalho foi pesquisar quais recursos deve-se usar na análise de um experimento não casualizado. Foi visto que as técnicas conhecidas de análise exploratória devem ser aplicadas, principalmente o uso de tabelas e gráficos. Um texto importante para atingir os objetivos do trabalho foi o livro de Hoaglin, Mosteller e Tukey (1991) intitulado “*Fundamentals of Exploratory Analysis of Variance*”; os autores são profundos conhecedores de análise de dados e procuraram neste trabalho unir ANOVA e técnicas de análise exploratória de dados estruturados.

Em cada fase do processo de análise houve um espaço para as conclusões dali advindas, procurando traduzir o que os resultados da análise dos gráficos, das médias e das variâncias significavam para quem realizou o experimento. A escolha de um modelo semelhante ao de blocos ao acaso para representar o conjunto de dados veio pela forma como os dados foram gerados: um indivíduo gerou todas as parcelas e portanto podem ser consideradas homogêneas, o que caracteriza um bloco. Por sua vez a combinação examinadores e métodos definem um fatorial 5 x 2.

A análise univariada permitiu um conhecimento particular de cada local em relação aos efeitos dos fatores e interação, e também do nível de variação de local para local. Embora pequena, as diferenças entre os examinadores foi significativa para todos os locais, com exceção do STOMION, assim como as diferenças entre os dois métodos e a interação. Ficou claro que havia heterogeneidade das variâncias

em relação aos locais, o que dificultaria uma análise conjunta dos locais (Tabelas 3 e 4).

Como as variáveis resposta são as coordenadas de pontos, foi criada uma nova variável que uniu as duas coordenadas e que representou a distância entre os dois pontos no plano. Se, para algum examinador, a distância é zero então não houve diferença entre os dois métodos, porque o primeiro ponto foi obtido com um método 1 e o segundo com o método 2. Esta nova variável também foi submetida às mesmas técnicas que as duas coordenadas, mas a interpretação dos resultados é mais concentrada. Por exemplo, uma ordenação das distâncias mostra onde as distâncias não diferem de zero, ou seja, onde os métodos foram equivalentes (Tabela 5). Pela distância entre métodos foi verificada a interação entre métodos e examinadores (Tabela 7). Também para distância entre métodos foi verificada heterogeneidade das variâncias entre os locais foi observada.

Seguindo o texto citado, as técnicas de análise de dados repetidos no espaço foram aplicadas ao experimento para as 3 variáveis e pelos resultados, também seriam confirmados os obtidos pela análise univariada. Aqui houve a possibilidade de se verificar a interação entre os fatores do experimento e os locais em todos os casos. Um ponto que impediu o uso de um modelo misto com modelagem da matriz de variâncias e covariâncias, foi que pelo PROC MIXED do SAS (2003) até para casos simples, o processo de iteração não convergiu na análise das coordenadas Y e X, mas convergiu para a distância em alguns casos.

Desde que não foi possível uma modelagem adequada da matriz de variâncias e covariâncias mediante um modelo misto, foi aplicada a análise multivariada (MANOVA). Neste caso, a matriz é a mais geral de todas, chamada pelo SAS de UN (Unistructured – Não-Estruturada), que possui as variâncias e covariâncias todas diferentes. Os resultados com os testes multivariados confirmaram os obtidos pelos modelos univariados.

Testes de hipóteses multivariados (e também os univariados) assumem que a distribuição dos dados é normal, isto é, cada um dos 180 vetores de 24 observações (locais) tem uma distribuição normal com a mesma matriz de variâncias. Um estudo de não normalidade foi feito em cada um dos modelos univariados. As distribuições foram aproximadamente normais para a maioria dos locais e as exceções, em 4

locais, foram devido a presença de uma medida discrepante em cada local. Um estudo dos erros após a análise sem as medidas discrepantes indicou uma distribuição aproximadamente normal para estes locais e as inferências não mudaram.

Pelos gráficos que foram feitos, pelas análises individuais das coordenadas e das distâncias e pela análise multivariada das coordenadas e distâncias entre métodos, chegamos a conclusões convergentes: Há diferença entre os examinadores, entre os métodos e evidência forte de interação entre examinadores e métodos. Com relação aos locais, também concluímos que há forte interação entre locais e examinadores e entre locais e métodos. Por outro lado, como a amostra é grande podemos usar o teorema central do limite para a dados multivariados e univariados referido por Vonesh e Chinchilli (1997) e considerar válidas as inferências das análises.

Outro ponto importante relacionado com a conclusão é por em discussão o que o pesquisador esperava. Era esperado que houvesse diferença entre os métodos e as interações? Sim, era esperado, pois havia a expectativa de se encontrar diferença entre as metodologias, em específico que a metodologia 2 (alternativa) teria uma variância menor, podendo assim ser uma boa opção de uso, em especial quando utilizado em trabalhos descontínuos que é quando há a possibilidade de troca do examinador ao longo do processo. Aqui vamos parafrasear Miller (1981): a significância em um estudo não deve ser somente estatística, também tem que ser significativo na parte social, econômica, fisiológica...

8. Referências Bibliográficas

- ANDREWS, D. F. **Plots of high-dimensional data**. Biometrics 28. p125-136. 1972.
- CHATFIELD, C.; COLLINS, A. J.; **Introduction to multivariate analysis**. ed. Chapman & Hall/crc, primeira edição 1980, reimpressão 2000.
- GNANADESIKAN, R. **Methods for Statistical Data Analysis of Multivariate Observations**. Wiley, 1977.
- HOAGLIN, D. C. ; MOSTELLER, F. ; TUKEY, J. W. **Fundamentals of Exploratory Analysis of Variance**. Wiley, 1991.
- JOE, H. **Multivariate models and dependence concepts**. ed. Chapman & Hall/crc. primeira edição 1997, reimpressão 2001.
- KUTNER, M. H. **Applied linear statistical models**.- Rev. ed. de: Applied linear regression models, quarta edição, 2004.
- LITTELL, Ramon C.; MILLIKEN, George A.; STROUP, Walter W.; WOLFINGER, Russell D.; SCHABENBERGER, Oliver. **SAS® for Mixed Models**. Second Edition. Cary, NC: SAS Institute Inc. 2006.
- MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate Analysis**. Academic Press. 1979.
- MILLIKEN, George A.; JOHNSON, Dallas E. **Analysis of messy data**. 2 v. : ill. 1989.
- MONTARROYOS, Ellen C. L.; COSTA, Kelem R. L.; FORTES, Renata C. **Antropometria e sua importância na avaliação do estado nutricional de crianças escolares**.
<http://www.escs.edu.br/pesquisa/revista/2013Vol24_1_3_Antropometriaimportancia.pdf> 2014.
- MONTGOMERY, Douglas G.; Arizona State University. **Design and Analysis of Experiment**. Fifth edition, John Wiley & Sons, INC. 2001.
- PEREIRA, Cimar A. **Manual de antropometria**. Pesquisa nacional de saúde. IBGE, rio de janeiro. 2013.

PREEDY, Victor R. **Handbook of antropometry**, physical mesures in human form in health and disease, volume 1, parts 1-6. Ed Springer. King`s college London. Londres 2012.

RAMIRES, Rossana R., **Correlação entre cefalometria e antropometria para determinação do tipo facial**. PUC-SP, São Paulo. 2008.

SILVA, Giovani L. **Modelos Lineares Generalizados - da teoria à pratica**. -M. Antonia Amaral Turkman DEIO/FC e CEAUL, Universidade de Lisboa E DM/IST e CMA, Universidade Técnica de Lisboa - Lisboa, 2000.

TINSLEY, H.; BROWN, S. **Handbook of Applied Multivariate Statistics and Mathematical Modeling**. Academic Press. 2000.

VIVALDI, Lucio J. **Análise de experimentos com dados repetidos ao longo do tempo ou espaço**. EMBRAPA DOCUMENTOS n.8 p1-52 dezembro 1999.

VONESH, E.F.; CHINCHILLI, V.M. **Linear and Nonlinear Models for the Analysis of Repeated Measurements**. CRC Press. 1997.

Laboratório de Avaliação Nutricional de Populações – LANPOP/HNT/FSP/USP. **Manual de técnicas antropométricas**. <<http://hygeia.fsp.usp.br/lanpop/manual.pdf>> 2014.

9. Apêndice

```
LIBNAME PROJETO2 "H:\KINGSTON\PEDRO_BRUNO"; RUN;

PROC PRINT DATA=PROJETO2.ANTROP; RUN;

/* DISTANCIA ENTRE PONTOS */

TITLE DISTANCIA ENTRE PONTOS;

PROC PRINT DATA=PROJETO2.MULTIYX; RUN;

DATA DECCA;

INPUT IND EXA D_M1_M2 LOCAL;

DATALINES;

0 0 0 0

PROC PRINT DATA=DECCA; RUN;

%MACRO DISTANCIA;

%DO J1 =1 %TO 24 ;

DATA TESTED; SET PROJETO2.MULTIYX;

KEEP IND EXA MET LY&J1 LX&J1;

/*PROC PRINT DATA=TESTED; RUN;*/

DATA MET_1; SET TESTED; IF MET=1;

LYM1L&J1=LY&J1;

LXM1L&J1=LX&J1;

DROP LY&J1 LX&J1 MET;

/*PROC PRINT DATA=MET_1; RUN;*/

DATA MET_2; SET TESTED; IF MET=2;

LYM2L&J1=LY&J1;

LXM2L&J1=LX&J1;

DROP LY&J1 LX&J1 MET;

/*PROC PRINT DATA=MET_2; RUN;*/

PROC SORT DATA=MET_1; BY IND EXA; RUN;

PROC SORT DATA=MET_2; BY IND EXA; RUN;

DATA MET_12; MERGE MET_1 MET_2; BY IND EXA;

D1=(LYM1L&J1-LYM2L&J1)**2;

D2=(LXM1L&J1-LXM2L&J1)**2;

D_M1_M2=10*SQRT(D1+D2);

LOCAL=&J1;

KEEP LOCAL IND EXA D_M1_M2;

/*PROC PRINT DATA=MET12; RUN;*/

DATA DECCA; SET DECCA MET_12; RUN;

/*PROC PRINT DATA=DECCA; RUN;*/

%END;

%MEND DISTANCIA;
```

```

RUN;
%DISTANCIA
PROC PRINT DATA=DECCA; RUN;
DATA DISAT1; SET DECCA; IF IND=0 THEN DELETE;
OPTIONS LS=90 PS=80;
PROC PRINT DATA=DISAT1;
FORMAT D_M1_M2 5.2;
RUN;

/* ANALISE DAS COMO PARCELAS DIVIDIDAS */
TITLE ANALISE DAS DISTANCIAS COMO PARCELAS DIVIDIDAS;
PROC GLM DATA=DISAT1;
CLASS IND EXA LOCAL;
MODEL D_M1_M2 = IND EXA IND*EXA LOCAL EXA*LOCAL;
RANDOM IND*EXA/TEST ;
RUN;
QUIT;

/* ***** */
TITLE LEITURA DA DISTANCIA MULTIVARADA;
DATA M_DISAT1; SET DISAT1;
PROC PRINT DATA=M_DISAT1; RUN;
PROC SORT DATA=M_DISAT1; BY IND EXA LOCAL;
PROC PRINT DATA=M_DISAT1; RUN;

/* LEITURA MULTIVARIADA DAS DISTANCIAS */
TITLE LEITURA MULTIVARIADA DOS DADOS;
PROC TRANSPOSE DATA=M_DISAT1 OUT=M_DISAT2 PREFIX=D;
VAR D_M1_M2;
BY IND EXA ;
RUN;
OPTIONS LS=160;
PROC PRINT DATA=M_DISAT2 ; RUN;
OPTIONS LS=90 PS=64 center;
TITLE1 ANALISE COMO PARCELAS DIVIDIDAS;
TITLE2 COM GRAUS DE LIBERDADE AJUSTADO;
PROC GLM DATA=M_DISAT2;
CLASS IND EXA ;
MODEL D1-D24 = EXA/NOUNI;
REPEATED LOCAL 24/PRINTE;
RUN;
QUIT;
TITLE1 ANALISE DAS DISTANCIAS COMO PARCELAS DIVIDIDAS;
TITLE2 MIXED COM VARIANCIAS DIFERENTES(CONVERGIU PARA DISTANCIAS);

```

```

PROC MIXED DATA=DISAT1;
CLASS IND EXA LOCAL;
MODEL D_M1_M2 = IND EXA LOCAL EXA*LOCAL;
REPEATED LOCAL /TYPE=CS SUBJECT=IND*EXA;
/*LSMEANS EXA*LOCAL;*/
ODS OUTPUT LSMEANS=MED_INT;
RUN;
QUIT;
TITLE1 ANALISE DAS DISTANCIAS COMO PARCELAS DIVIDIDAS;
TITLE2 MIXED COM VARIANCIAS DIFERENTES(CONVERGIU PARA DISTANCIAS);
PROC MIXED DATA=DISAT1;
CLASS IND EXA LOCAL;
MODEL D_M1_M2 = IND EXA LOCAL EXA*LOCAL;
REPEATED LOCAL /TYPE=un SUBJECT=IND*EXA;
/*LSMEANS EXA*LOCAL;*/
ODS OUTPUT LSMEANS=MED_INT;
RUN;
QUIT;
TITLE1 ANALISE DAS DISTANCIAS COMO PARCELAS DIVIDIDAS;
TITLE2 MIXED COM VARIANCIAS DIFERENTES(CONVERGIU PARA DISTANCIAS);
PROC MIXED DATA=DISAT1;
CLASS IND EXA LOCAL;
MODEL D_M1_M2 = IND EXA LOCAL EXA*LOCAL;
REPEATED LOCAL /TYPE=UN SUBJECT=IND*EXA R;
/*LSMEANS EXA*LOCAL;*/
ODS OUTPUT LSMEANS=MED_INT;
ODS OUTPUT R=MATRIZR;
RUN;
QUIT;
TITLE;
OPTIONS LS=90 PS=64 CENTER;
PROC PRINT NOOBS DATA=MATRIZR;
VAR ROW COL1-COL24;
FORMAT COL1-COL24 6.2;
RUN;
PROC PRINT DATA=MED_INT;RUN;
/* ONDE OS METODOS SÃO MAIS DIFERENTES*/
TITLE ONDE OS METODOS SÃO MAIS DIFERENTES;
PROC SORT DATA=MED_INT OUT=ORDEM_1; BY ESTIMATE; RUN;
PROC PRINT DATA=ORDEM_1;RUN;
PROC SORT DATA=MED_INT OUT=ORDEM_2; BY EXA ESTIMATE; RUN;

```

```

PROC PRINT DATA=ORDEM_2;RUN;
PROC SORT DATA=MED_INT OUT=ORDEM_3; BY LOCAL ESTIMATE; RUN;
PROC PRINT DATA=ORDEM_3;RUN;
PROC SORT DATA=MED_INT OUT=ORDEM_4; BY LOCAL EXA RUN;
PROC PRINT DATA=ORDEM_4;RUN;
PROC TABULATE DATA=MED_INT FORMAT=5.2;
CLASS EXA LOCAL;
VAR Estimate;
TABLE LOCAL , ESTIMATE*EXA*MEAN/RTS=9;
LABEL ESTIMATE="DISTANCIA MEDIA" EXA="EXAMINADORES";
KEYLABEL MEAN=" " ;
RUN;
PROC TABULATE DATA=MED_INT FORMAT=5.2;
CLASS EXA LOCAL;
VAR Estimate;
TABLE EXA , ESTIMATE*LOCAL*MEAN/RTS=9;
LABEL ESTIMATE="DISTANCIA MEDIA" ;
KEYLABEL MEAN=" " ;
RUN;
/* UNIVARIADA */
OPTIONS LS=90 PS=60;
LIBNAME PROJETO2 "H:\KINGSTON\PEDRO_BRUNO"; RUN;
OPTIONS NODATE NONUMBER;
TITLE LEITURA UNIVARIADA;
PROC PRINT DATA=PROJETO2.ANTROP; RUN;
/* MUDANÇA DE ESCALA*/
TITLE DIVIDIR Y E X POR 10;
DATA TESTE; SET PROJETO2.ANTROP;
Y=Y/10;
X=X/10;
PROC SORT DATA=TESTE; BY IND EXA MET;
PROC PRINT DATA=TESTE;
VAR LOCAL IND EXA MET Y X;RUN;
/* MEDIAS E VARIANCIAS*/
TITLE TABELA DE INTERAÇÃO;
OPTIONS LS=94 PS=70;
PROC TABULATE DATA=TESTE FORMAT=4.1;
CLASS IND EXA MET LOCAL;
VAR Y X;
/*TABLE EXA ALL,(X Y)*(MET ALL)*MEAN/RTS=12;
TABLE IND ALL , (X Y)*(MET ALL)*MEAN;*/

```

```

TABLE LOCAL , (Y X)*(EXA ALL)*MEAN/RTS=27;
KEYLABEL MEAN=" " ALL="MEDIA" ;
RUN;

/* MEDIAS E VARIANCIAS*/
TITLE TABELA DE INTERAÇÃO;
PROC TABULATE DATA=TESTE FORMAT=6.1;
CLASS IND EXA MET;
VAR Y X;
TABLE EXA ALL ,(X Y)*(MET ALL)*VAR/RTS=8;
TABLE IND ALL , (X Y)*(MET ALL)*VAR/RTS=8;
KEYLABEL MEAN=" " ALL="TOTAL" ;
RUN;

/* MEDIA E ERRO PADRÃO*/
TITLE TABELA DE INTERAÇÃO;
PROC TABULATE DATA=TESTE FORMAT=6.1;
CLASS IND EXA MET;
VAR Y X;
TABLE EXA ALL ,(X Y)*(MET ALL)*STD/RTS=8;
TABLE IND ALL , (X Y)*(MET ALL)*STD/RTS=8;
KEYLABEL MEAN=" " ALL="TOTAL" ;
RUN;

/*GRAFICOS*/
/* LOCAL NUMERICO*/
PROC SORT DATA=PROJETO2.ANTROP ; BY LOCAL;
PROC PRINT DATA=PROJETO2.ANTROP; RUN;
OPTIONS NODATE NONUMBER;
TITLE LEITURA UNIVARIADA DOS DADOS;
DATA TESTE; SET PROJETO2.ANTROP;
C + 1;
LOCAL_N=CEIL(C/180);
DROP C;
y=y/10;
X=X/10;
PROC PRINT DATA=TESTE;
VAR LOCAL LOCAL_N IND MET EXA Y X ;
RUN;

/* MEDIAS */
TITLE TABELA DE INTERAÇÃO COM LOCAIS;
OPTIONS LS=90 PS=70;
PROC TABULATE DATA=TESTE FORMAT=5.1;
CLASS IND EXA MET LOCAL_N;

```



```

VAR Y X;

/*TABLE EXA ALL,(X Y)*(MET ALL)*MEAN/RTS=12;
TABLE IND ALL , (X Y)*(MET ALL)*MEAN;*/

TABLE LOCAL_N , (Y X)*(EXA ALL)*MEAN/RTS=12;
TABLE LOCAL_N , (Y X)*(MET ALL)*MEAN/RTS=12;
KEYLABEL MEAN=" " ALL="MEDIA" ;

RUN;

/* MEDIAS PARA OS GRAFICOS*/

DATA TESTE1; SET TESTE;

PROC SORT DATA=TESTE1 ; BY IND MET LOCAL_N;

PROC MEANS DATA=TESTE1 MEAN NOPRINT ; BY IND MET LOCAL_N;

VAR Y X;

OUTPUT OUT=MEDIAS1 MEAN=YM XM;

RUN;

PROC PRINT DATA=MEDIAS1; RUN;

OPTIONS RESET=ALL;

TITLE H=1.3 'GRAFICOS DOS PERFIS DE CADA INDIVIDUO ';

proc gplot data=TESTE1;; BY IND;

symbol1 I=JOINT c=BLUE W=2;
symbol2 I=JOINT c=RED W=2;

plot Y*LOCAL_N=MET ;

run;

QUIT;

PROC SORT DATA=TESTE1; BY MET; RUN;

OPTIONS RESET=ALL;

TITLE H=1.3 'GRAFICOS DOS PERFIS DE CADA METODO ';

proc gplot data=TESTE1;; BY MET;

symbol1 I=JOINT c=BLUE W=2;
symbol2 I=JOINT c=RED W=2;

plot Y*LOCAL_N=IND ;

run;

QUIT;

/******//
/* GRAFICO DE ANDREWS*/
/* LEITURA MULTIVARIADA*/

OPTIONS LS=160;

TITLE LEITURA MULTIVARIADA DOS DADOS PARA Y;

PROC PRINT DATA=PROJETO2.MULTIY; RUN;

/* ARQUIVO PROJETO2.MULTIY SENDO ENVIADO*/

TITLE LEITURA MULTIVARIADA DOS DADOS PARA X;

PROC PRINT DATA=PROJETO2.MULTIX; RUN;

```

/* ARQUIVO PROJETO2.MULTIX SENDO ENVIADO*/

TITLE LEITURA MULTIVARIADA DE Y E X;

PROC PRINT DATA=PROJETO2.MULTIYX; RUN;

/* ARQUIVO PROJETO2.MULTIYX SENDO ENVIADO*/

/* FUNÇÃO DE ANDREWS PARA Y*/

TITLE MEDIAS PARA CADA IND E MET PARA Y;

PROC SORT DATA=PROJETO2.MULTIY ; BY IND MET;RUN;

DATA ERRADOY; SET PROJETO2.MULTIY; RUN;

PROC SORT DATA=ERRADOY ; BY IND MET;RUN; PROC PRINT DATA=ERRADOY; RUN;

PROC MEANS DATA=ERRADOY MEAN NOPRINT; BY IND MET;

VAR LY1-LY24;

OUTPUT OUT=MEDIAS1Y MEAN=LY1M LY2M LY3M LY4M LY5M LY6M

LY7M LY8M LY9M LY10M LY11M LY12M

LY13M LY14M LY15M LY16M LY17M LY18M

LY19M LY20M LY21M LY22M LY23M LY24M;

RUN;

PROC PRINT DATA=MEDIAS1Y;

RUN;

DATA TESTE; SET MEDIAS1Y;

C+1;

PROC PRINT DATA=TESTE; RUN;

TITLE VALORES DA FUNÇÃO DE ANDREWS PARA Y;

DATA ANDREWS2; SET TESTE;

DO T = -3.14 TO 3.14 BY 0.15;

F_T = LY1M/SQRT(2) + LY2M*SIN(T) + LY3M*COS(T)+

LY4M*SIN(2*T) + LY5M*COS(2*T)+

LY6M*SIN(3*T) + LY7M*COS(3*T)+

LY8M*SIN(4*T) + LY9M*COS(4*T)+

LY10M*SIN(5*T) + LY11M*COS(5*T)+

LY12M*SIN(6*T) + LY13M*COS(6*T)+

LY14M*SIN(7*T) + LY15M*COS(7*T)+

LY16M*SIN(8*T) + LY17M*COS(8*T)+

LY18M*SIN(9*T) + LY19M*COS(9*T)+

LY20M*SIN(10*T) + LY21M*COS(10*T)+

LY22M*SIN(11*T) + LY23M*COS(11*T)+

LY24M*SIN(12*T) ;

OUTPUT;

END;

RUN;

PROC PRINT DATA=ANDREWS2;

VAR IND MET C T F_T;

```

RUN;
TITLE GRAFICO DE ANDREWS PARA CADA MET - Y;
PROC SORT DATA=ANDREWS2; BY MET;
OPTIONS RESET=ALL;
PROC GPLOT DATA=ANDREWS2; BY MET;
SYMBOL I=J ;
PLOT F_T*T=IND;
RUN;
QUIT;
TITLE GRAFICO DE ANDREWS PARA CADA IND - Y;
PROC SORT DATA=ANDREWS2; BY IND;
OPTIONS RESET=ALL;
PROC GPLOT DATA=ANDREWS2; BY IND;
SYMBOL I=J W=2;
PLOT F_T*T=MET;
RUN;
QUIT;
/* GRAFICOS */
LIBNAME PROJETO2 "H:\KINGSTON\PEDRO_BRUNO"; RUN;
PROC PRINT DATA=PROJETO2.ANTROP; RUN;
OPTIONS LS=90 PS=64;
DATA TESTE; SET PROJETO2.ANTROP;
C + 1;
LOCAL_N=CEIL(C/180);
Y=Y/10;
X=X/10;
DROP C;
PROC SORT DATA=TESTE; BY IND EXA LOCAL_N MET;
PROC PRINT DATA=TESTE;
VAR LOCAL_N IND EXA MET Y X ;RUN;
proc univariate data=teste;
var y x;
run;
proc sort data=teste; by x;
proc print data=teste; run;
TITLE GRAFICO GERAL;
OPTIONS RESET=ALL;
PROC SORT DATA=TESTE; BY MET; RUN;
PROC GPLOT DATA=TESTE;
PLOT Y*X=MET;
RUN;

```

```

QUIT;
DATA TESTEL1; SET TESTE;
IF LOCAL_N=1;
PROC PRINT DATA=TESTEL1; RUN;
TITLE GRAFICO LOCAL1;
OPTIONS RESET=ALL;
PROC SORT DATA=TESTEL1; BY MET; RUN;
PROC GLOT DATA=TESTEL1;
PLOT Y*X=MET;
RUN;
QUIT;
options ls=90 ps=64;
PROC SORT DATA=TESTE; BY IND MET; RUN;
PROC MEANS DATA=TESTE MEAN NOPRINT; BY IND MET;
VAR Y X;
OUTPUT OUT=MEDIASYX MEAN=YM XM;
RUN;
PROC PRINT DATA=MEDIASYX; RUN;
TITLE GRAFICO COM MEDIAS DE IND;
OPTIONS RESET=ALL;
PROC GLOT DATA=MEDIASYX;
PLOT YM*XM=MET;
RUN;
QUIT;
/* DISTANCIA ENTRE PONTOS */
TITLE DISTANCIA ENTRE PONTOS;
PROC PRINT DATA=PROJETO2.MULTIYX; RUN;
DATA TESTED; SET PROJETO2.MULTIYX;
DATA TESTED1; SET TESTED;
KEEP IND EXA MET LY1 LX1 ;
PROC PRINT DATA=TESTED1; RUN;
DATA MET1; SET TESTED1; IF MET=1;
LY1M1=LY1;
LX1M1=LX1;
DROP LY1 LX1 MET;
PROC PRINT DATA=MET1; RUN;
DATA MET2; SET TESTED1; IF MET=2;
LY1M2=LY1;
LX1M2=LX1;
DROP LY1 LX1 MET;
PROC PRINT DATA=MET2; RUN;

```

```

PROC SORT DATA=MET1; BY IND EXA; RUN;
PROC SORT DATA=MET2; BY IND EXA; RUN;
DATA MET12; MERGE MET1 MET2; BY IND EXA;
D_M1_M2=10*SQRT((LY1M1-LY1M2)**2 + (LX1M1-LX1M2)**2);
PROC PRINT DATA=MET12; RUN;
OPTIONS RESET=ALL HSIZE=3.5 VSIZE=3.0;
PROC GPLOT DATA=MET12;
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
QUIT;
OPTIONS RESET=ALL;
PROC ANOVA DATA=MET12;
CLASS IND EXA;
MODEL D_M1_M2 = IND EXA;
MEANS EXA;
RUN;
QUIT;
DATA DECCA;
INPUT IND EXA D_M1_M2 LOCAL;
DATALINES;
0 0 0 0
PROC PRINT DATA=DECCA; RUN;
%MACRO DISTANCIA;
%DO J1 =1 %TO 24 ;
DATA TESTED; SET PROJETO2.MULTIYX;
KEEP IND EXA MET LY&J1 LX&J1;
/*PROC PRINT DATA=TESTED; RUN;*/
DATA MET_1; SET TESTED; IF MET=1;
LYM1L&J1=LY&J1;
LXM1L&J1=LX&J1;
DROP LY&J1 LX&J1 MET;
/*PROC PRINT DATA=MET_1; RUN*/
DATA MET_2; SET TESTED; IF MET=2;
LYM2L&J1=LY&J1;
LXM2L&J1=LX&J1;
DROP LY&J1 LX&J1 MET;
/*PROC PRINT DATA=MET_2; RUN;*/
PROC SORT DATA=MET_1; BY IND EXA; RUN;
PROC SORT DATA=MET_2; BY IND EXA; RUN;
DATA MET_12; MERGE MET_1 MET_2; BY IND EXA;

```

```

D1=(LYM1L&J1-LYM2L&J1)**2;
D2=(LXM1L&J1-LXM2L&J1)**2;
D_M1_M2=10*SQRT(D1+D2);
LOCAL=&J1;
KEEP LOCAL IND EXA D_M1_M2;
/*PROC PRINT DATA=MET12; RUN;*/
DATA DECCA; SET DECCA MET_12; RUN;
/*PROC PRINT DATA=DECCA; RUN;*/
%END;
%MEND DISTANCIA;
RUN;
%DISTANCIA
PROC PRINT DATA=DECCA; RUN;
DATA DISAT1; SET DECCA; IF IND=0 THEN DELETE;
OPTIONS LS=90 PS=80;
PROC PRINT DATA=DISAT1;
FORMAT D_M1_M2 5.2;
RUN;
TITLE GRAFICOS DAS DISTANCIAS - LOCAIS 1 A 4;
DATA G4L1; SET DISAT1; IF LOCAL=1 ; PROC PRINT ; RUN;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L1 GOUT= WORK.DISAT1;
TITLE H=1 'LOCAL1';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
QUIT;
DATA G4L2; SET DISAT1; IF LOCAL=2 ; PROC PRINT ; RUN;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L2 GOUT= WORK.DISAT1;
TITLE H=1 'LOCAL2';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
QUIT;
DATA G4L3; SET DISAT1; IF LOCAL=3 ; PROC PRINT ; RUN;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0 ;
PROC GPLOT DATA=G4L3 GOUT= WORK.DISAT1;
TITLE H=1 'LOCAL3';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;

```

```

RUN;
QUIT;
DATA G4L4; SET DISAT1; IF LOCAL=4 ; PROC PRINT ; RUN;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L4 GOUT= WORK.DISAT1;
TITLE H=1 'LOCAL4';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
QUIT;
ODS HTML;
ODS GRAPHICS ON;
TITLE QUATRO GRAFICOS EM UMA PAGINA;
PROC GREPLAY IGOUT=WORK.DISAT1 TC=SASHELP.TEMPLT TEMPLATE=L2R2 NOFS;
TREPLAY 1:1 2:2 3:3 4:4 ;
RUN;
ODS GRAPHICS OFF;
ODS HTML CLOSE;
QUIT;
DATA DISAT2; SET DECCA; IF IND=0 THEN DELETE; RUN;
TITLE GRAFICOS DAS DISTANCIAS - LOCAIS 5 A 8;
DATA G4L5; SET DISAT2; IF LOCAL=5 ;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L5 GOUT= WORK.DISAT2;
TITLE H=1 'LOCAL5';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
QUIT;
DATA G4L6; SET DISAT2; IF LOCAL=6 ;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L6 GOUT= WORK.DISAT2;
TITLE H=1 'LOCAL6';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
QUIT;
DATA G4L7; SET DISAT2; IF LOCAL=7 ;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L7 GOUT= WORK.DISAT2;
TITLE H=1 'LOCAL7';

```

```

SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
QUIT;
DATA G4L8; SET DISAT2; IF LOCAL=8 ;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L8 GOUT= WORK.DISAT2;
TITLE H=1 'LOCAL8';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
QUIT;
ODS HTML;
ODS GRAPHICS ON;
TITLE QUATRO GRAFICOS EM UMA PAGINA;
PROC GREPLAY IGOUT=WORK.DISAT2 TC=SASHELP.TEMPLT TEMPLATE=L2R2 NOFS;
TREPLAY 1:1 2:2 3:3 4:4 ;
RUN;
ODS GRAPHICS OFF;
ODS HTML CLOSE;
QUIT;
DATA DISAT3; SET DECCA; IF IND=0 THEN DELETE; RUN;
TITLE GRAFICOS DAS DISTANCIAS - LOCAIS 9 A 12;
DATA G4L9; SET DISAT3; IF LOCAL=9 ;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L9 GOUT= WORK.DISAT3;
TITLE H=1 'LOCAL9';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
DATA G4L10; SET DISAT3; IF LOCAL=10 ;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L10 GOUT= WORK.DISAT3;
TITLE H=1 'LOCAL10';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
DATA G4L11; SET DISAT3; IF LOCAL=11 ;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L11 GOUT= WORK.DISAT3;
TITLE H=1 'LOCAL11';

```



```

SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
DATA G4L12; SET DISAT3; IF LOCAL=12 ;
OPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L12 GOUT= WORK.DISAT3;
TITLE H=1 'LOCAL12';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
ODS HTML;
ODS GRAPHICS ON;
TITLE QUATRO GRAFICOS EM UMA PAGINA;
PROC GREPLAY IGOUT=WORK.DISAT3 TC=SASHELP.TEMPLT TEMPLATE=L2R2 NOFS;
TREPLAY 1:1 2:2 3:3 4:4 ;
RUN;
ODS GRAPHICS OFF;
ODS HTML CLOSE;
QUIT;
DATA DISAT4; SET DECCA; IF IND=0 THEN DELETE; RUN;
TITLE GRAFICOS DAS DISTANCIAS - LOCAIS 13 A 16;
DATA G4L13; SET DISAT4; IF LOCAL=13 ;
OPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L13 GOUT= WORK.DISAT4;
TITLE H=1 'LOCAL13';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
DATA G4L14; SET DISAT4; IF LOCAL=14 ;
OPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L14 GOUT= WORK.DISAT4;
TITLE H=1 'LOCAL14';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
DATA G4L15; SET DISAT4; IF LOCAL=15 ;
OPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L15 GOUT= WORK.DISAT4;
TITLE H=1 'LOCAL15';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;

```

```

RUN;
DATA G4L16; SET DISAT4; IF LOCAL=16 ;
OPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L16 GOUT= WORK.DISAT4;
TITLE H=1 'LOCAL16';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
ODS HTML;
ODS GRAPHICS ON;
TITLE QUATRO GRAFICOS EM UMA PAGINA;
PROC GREPLAY IGOUT=WORK.DISAT4 TC=SASHELP.TEMPLT TEMPLATE=L2R2 NOFS;
TREPLAY 1:1 2:2 3:3 4:4 ;
RUN;
ODS GRAPHICS OFF;
ODS HTML CLOSE;
QUIT;
DATA DISAT5; SET DECCA; IF IND=0 THEN DELETE; RUN;
TITLE GRAFICOS DAS DISTANCIAS - LOCAIS 17 A 20;
DATA G4L17; SET DISAT5; IF LOCAL=17 ; PROC PRINT ; RUN;
OPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L17 GOUT= WORK.DISAT5;
TITLE H=1 'LOCAL17';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
DATA G4L18; SET DISAT5; IF LOCAL=18 ;
OPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L18 GOUT= WORK.DISAT5;
TITLE H=1 'LOCAL18';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
DATA G4L19; SET DISAT5; IF LOCAL=19 ;
OPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L19 GOUT= WORK.DISAT5;
TITLE H=1 'LOCAL19';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
DATA G4L20; SET DISAT5; IF LOCAL=20 ;

```

```

GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L20 GOUT= WORK.DISAT5;
TITLE H=1 'LOCAL20';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
ODS HTML;
ODS GRAPHICS ON;
TITLE QUATRO GRAFICOS EM UMA PAGINA;
PROC GREPLAY IGOUT=WORK.DISAT5 TC=SASHELP.TEMPLT TEMPLATE=L2R2 NOFS;
TREPLAY 1:1 2:2 3:3 4:4 ;
RUN;
ODS GRAPHICS OFF;
ODS HTML CLOSE;
QUIT;
DATA DISAT6; SET DECCA; IF IND=0 THEN DELETE; RUN;
TITLE GRAFICOS DAS DISTANCIAS - LOCAIS 21 A 24;
DATA G4L21; SET DISAT6; IF LOCAL=21 ; proc print ; run;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L21 GOUT= WORK.DISAT6;
TITLE H=1 'LOCAL21';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
TITLE GRAFICOS DAS DISTANCIAS - LOCAIS 21 A 24;
DATA G4L22; SET DISAT6; IF LOCAL=22 ;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L22 GOUT= WORK.DISAT6;
TITLE H=1 'LOCAL22';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
TITLE GRAFICOS DAS DISTANCIAS - LOCAIS 21 A 24;
DATA G4L23; SET DISAT6; IF LOCAL=23 ;
GOPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L23 GOUT= WORK.DISAT6;
TITLE H=1 'LOCAL23';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
TITLE GRAFICOS DAS DISTANCIAS - LOCAIS 21 A 24;

```

```

DATA G4L24; SET DISAT6; IF LOCAL=24 ;
OPTIONS RESET=ALL HSIZE=4.5 VSIZE=3.0;
PROC GPLOT DATA=G4L24 GOUT= WORK.DISAT6;
TITLE H=1 'LOCAL24';
SYMBOL VALUE=DOT;
PLOT D_M1_M2*EXA/HAXIS=1 TO 5 ;
RUN;
ODS HTML;
ODS GRAPHICS ON;
TITLE QUATRO GRAFICOS EM UMA PAGINA;
PROC GREPLAY IGOUT=WORK.DISAT6 TC=SASHELP.TEMPLT TEMPLATE=L2R2 NOFS;
TREPLAY 1:1 2:2 3:3 4:4 ;
RUN;
ODS GRAPHICS OFF;
ODS HTML CLOSE;
QUIT;
OPTIONS LS=90 PS=80;
TITLE TABELA DE INTERAÇÃO;
PROC TABULATE DATA=DISAT1 FORMAT=6.2;
CLASS EXA local;
VAR D_M1_M2;
TABLE LOCAL all, D_M1_M2*(EXA all)*MEAN/RTS=12;
KEYLABEL MEAN=" " ALL="MEDIA" ;
RUN;
PROC SORT DATA=DISAT1; BY LOCAL; RUN;
TITLE ANALISE DE TODOS OS LOCAIS - GLM;
PROC GLM DATA=DISAT1 OUTSTAT=RESD NOPRINT ;BY LOCAL;
CLASS IND EXA ;
MODEL D_M1_M2= IND EXA ;
RUN;
QUIT;
PROC PRINT DATA=RESD; RUN;
DATA RESD1; SET RESD;
QM=SS/DF;
IF _TYPE_="SS1" THEN DELETE;
IF _SOURCE_="IND" THEN DELETE;
PROC PRINT DATA=RESD1; RUN;
TITLE ANOVA DESCRITIVA PARA EXAMINADORES;
DATA ERROD1; SET RESD1;
IF _TYPE_="ERROR";
KEEP LOCAL QM;

```

```

PROC PRINT DATA=ERROD1; RUN;
DATA ERROD2; SET RESD1;
IF _SOURCE_="EXA";
KEEP LOCAL F PROB;
PROC PRINT DATA=ERROD2; RUN;
PROC SORT DATA=DISAT1 ; BY LOCAL;
PROC MEANS DATA=DISAT1 MEAN NOPRINT ; BY LOCAL;
VAR D_M1_M2;
OUTPUT OUT=MEDIASDL MEAN=MEDIA;
RUN;
PROC PRINT DATA=MEDIASDL ; RUN;
TITLE ANOVA DESCRITIVA PARA EXAMINADORES EM CADA LOCAL;
OPTIONS NODATE NONUMBER;
DATA ANOVAD1; MERGE MEDIASDL ERROD1 ERROD2; BY LOCAL;
VARIANCIA=QM;
DROP _TYPE_ _FREQ_ QM;
PROC PRINT DATA=ANOVAD1;
VAR MEDIA F PROB VARIANCIA;
RUN;
TITLE ANALISE DE VARIANCIA - GLM;
PROC GLM DATA=TESTE;
CLASS IND EXA MET;
MODEL Y = EXA MET EXA*MET IND;
RUN;
QUIT;
TITLE ANALISE DE UM LOCAL - MIXED;
PROC MIXED DATA=TESTE;
CLASS IND EXA MET;
MODEL Y = EXA MET EXA*MET ;
RANDOM IND;
RUN;
QUIT;
DATA TESTE1; SET PROJETO2.ANTROP;
IF LOCAL="GLABELA" OR LOCAL="SUBNASAL";
PROC PRINT DATA=TESTE1; RUN;
TITLE ANALISE - GLM SPLIT PARA LOCAL;
PROC GLM DATA=TESTE;
CLASS IND EXA MET LOCAL;
MODEL Y = EXA MET EXA*MET IND IND*EXA*MET
LOCAL LOCAL*EXA LOCAL*MET LOCAL*EXA*MET;
RANDOM IND IND*EXA*MET/ TEST;

```

```

RUN;
QUIT;
TITLE ANALISE DE TODOS OS LOCAIS - MIXED;
PROC MIXED DATA=PROJETO2.ANTROP;
CLASS IND EXA MET LOCAL;
MODEL Y = EXA MET EXA*MET
      LOCAL LOCAL*EXA LOCAL*MET LOCAL*EXA*MET;
RANDOM IND IND*EXA*MET;
RUN;
QUIT;
TITLE TABELA DE INTERAÇÃO;
PROC TABULATE DATA=PROJETO2.ANTROP FORMAT=6.1;
CLASS IND EXA MET LOCAL;
VAR Y X;
TABLE EXA ALL, Y*(MET ALL)*MEAN/RTS=12;
KEYLABEL MEAN=" " ALL="MEDIA" ;
RUN;
PROC SORT DATA=DATA=PROJETO2.ANTROP; BY LOCAL;
TITLE ANALISE DE TODOS OS LOCAIS - GLM;
PROC GLM DATA=PROJETO2.ANTROP OUTSTAT=RESYX;BY LOCAL;
CLASS IND EXA MET ;
MODEL Y X= IND EXA MET EXA*MET;
RUN;
QUIT;
PROC PRINT DATA=RESYX; RUN;
DATA RESY; SET RESYX; IF _NAME_="Y";
DROP _NAME_ DF X;
IF _SOURCE_="IND" THEN DELETE;
IF _TYPE_="SS1" THEN DELETE;
DROP _TYPE_ Y SS;
FONTE_Y=_SOURCE_;
F_Y=F;
PROB_Y=PROB;
DROP _SOURCE_ F PROB;
PROC PRINT DATA=RESY; RUN;
DATA RESX; SET RESYX; IF _NAME_="X";
DROP _NAME_ DF Y;
IF _SOURCE_="IND" THEN DELETE;
IF _TYPE_="SS1" THEN DELETE;
DROP _TYPE_ X SS;
FONTE_X=_SOURCE_;

```

```

F_X=F;
PROB_X=PROB;
DROP _SOURCE_ F PROB;
PROC PRINT DATA=RESX; RUN;
PROC SORT DATA=RESY; BY LOCAL;
PROC SORT DATA=RESX; BY LOCAL;
DATA RESYX_J; MERGE RESY RESX ; BY LOCAL;
IF FONTE_Y="ERROR" THEN DELETE;
IF FONTE_X="ERROR" THEN DELETE;
PROC PRINT DATA=RESYX_J; RUN;
PROC PRINT DATA=RESX; RUN;
/* TESTE DE MARDIA PARA NORMALIDADE MULTIVARIADA */
LIBNAME PROJETO2 "H:\KINGSTON\PEDRO_BRUNO"; RUN;
PROC SORT DATA=PROJETO2.ANTROP; BY LOCAL;
PROC PRINT DATA=PROJETO2.ANTROP; RUN;
DATA PROJETO2.MULTIY; SET TESTE6; RUN;
OPTIONS LS=140;
PROC PRINT DATA=PROJETO2.MULTIY; RUN;
DATA ANOVA1; SET PROJETO2.MULTIY; RUN;
/* ***** */
/* PARA Y */
TITLE MATRIZ H DE VONESH E CHINCHILLI E ERROS;
OPTIONS LS=100 PS=100;
PROC IML ;
USE ANOVA1;
READ ALL VAR{IND} INTO B1;
READ ALL VAR{EXA} INTO B2;
READ ALL VAR{MET} INTO B3;
READ ALL VAR{LY1 LY2 LY3 LY4 LY5 LY6 LY7 LY8 LY9 LY10
LY11 LY12 LY13 LY14 LY15 LY16 LY17 LY18
LY19 LY20 LY21 LY22 LY23 LY24} INTO Y;
UM=J(180,1,1);
X1=DESIGN(B1);
X2=DESIGN(B2);
X3=DESIGN(B3);
X4=J(180,10,0);
K=0;
DO I = 1 TO 5;
DO J = 1 TO 2;
K=K+1;
X4[,K]=X2[,I]#X3[,J];

```

```

END;
END;
A=UM||X1||X2||X3||X4;
AA=A`*A;
AA_=GINV(AA);
PT=A*GINV(A`*A)*A`;
S=I(180)-PT;
D=EIGVAL(S);
V=EIGVEC(S);
DI=DIAG(D);
P=V*DI;
H=P[,1:153];
H=H`;
P_P_=H*H`;
ERROV=H*Y;
CREATE DADOSV FROM ERROV;
append from ERROV;
close DADOSV;
QUIT;
PROC PRINT DATA=DADOSV; RUN;
TITLE ESTIMATIVA DA MATRIZ DE VARIANCIAS E COVARIANCIAS;
proc anova DATA=ANOVA1 outstat=ttv;
    class IND EXA MET;
    model LY1-LY24=IND EXA MET EXA*MET/nouni;
    MANOVA/PRINTE;
    ODS OUTPUT ErrorSSCP=ERRO;
RUN;
QUIT;
PROC PRINT DATA=ERRO; RUN;
TITLE TESTE DE MARDIA;
proc iml;
USE ERRO;
READ ALL VAR{LY1 LY2 LY3 LY4 LY5 LY6 LY7 LY8 LY9 LY10
    LY11 LY12 LY13 LY14 LY15 LY16 LY17 LY18
    LY19 LY20 LY21 LY22 LY23 LY24} INTO SY; PRINT SY;
DY=EIGVAL(SY); PRINT DY;
use DADOSV;
read all into y;
n = nrow(y) ; PRINT N;
p = ncol(y) ; PRINT P;
dfchi = p*(p+1)*(p+2)/6 ; PRINT DFCHI;

```



```

S=(1/153)*SY;
s_inv = inv(s) ; PRINT S_INV;
q = i(n) - (1/n)*j(n,n,1);
g_matrix = q*y*s_inv*y`*q;
beta1hat = ( sum(g_matrix#g_matrix#g_matrix) )/(n*n);
beta2hat =trace( g_matrix#g_matrix )/n ;
kappa1 = n*beta1hat/6 ;
kappa2 = (beta2hat - p*(p+2) ) /sqrt(8*p*(p+2)/n) ;
pvalskew = 1 - probchi(kappa1,dfchi) ;
pvalkurt = 2*( 1 - probnorm(abs(kappa2)) );
print s ;
print s_inv ;
print 'TESTS: ';
print 'Based on skewness: ' beta1hat kappa1 pvalskew ;
print 'Based on kurtosis: ' beta2hat kappa2 pvalkurt;
QUIT;
/*****
/* PARA X */
PROC PRINT DATA=PROJETO2.MULTIX; RUN;
DATA ANOVA2; SET PROJETO2.MULTIX; RUN;
OPTIONS LS=100 PS=100;
PROC IML ;
USE ANOVA2;
READ ALL VAR{IND} INTO B1;
READ ALL VAR{EXA} INTO B2;
READ ALL VAR{MET} INTO B3;
READ ALL VAR{LX1 LX2 LX3 LX4 LX5 LX6 LX7 LX8 LX9 LX10
LX11 LX12 LX13 LX14 LX15 LX15 LX16 LX18
LX19 LX20 LX21 LX22 LX23 LX4} INTO Y;
UM=J(180,1,1);
X1=DESIGN(B1);
X2=DESIGN(B2);
X3=DESIGN(B3);
X4=J(180,10,0);
K=0;
DO I = 1 TO 5;
DO J = 1 TO 2;
K=K+1;
X4[,K]=X2[,I]#X3[,J];
END;
END;

```

```

A=UM||X1||X2||X3||X4;
AA=A`*A;
AA_=GINV(AA);
PT=A*GINV(A`*A)*A`;
S=I(180)-PT;
D=EIGVAL(S);
V=EIGVEC(S);
DI=DIAG(D);
P=V*DI;
H=P[,1:153];
H=H`;
P_P_=H*H`;
ERROV=H*Y;
CREATE DADOSVX FROM ERROV;
append from ERROV;
close DADOSVX;
QUIT;
PROC PRINT DATA=DADOSVX; RUN;
proc anova DATA=ANOVA2 outstat=ttv;
    class IND EXA MET;
    MODEL LX1 LX2 LX3 LX4 LX5 LX6 LX7 LX8 LX9 LX10
    LX11 LX12 LX13 LX14 LX15 LX16 LX17 LX18
    LX19 LX20 LX21 LX22 LX23 LX24 = IND EXA MET EXA*MET;
MANOVA/PRINTE;
    ODS OUTPUT ErrorSSCP=ERROX;
RUN;
QUIT;
PROC PRINT DATA=ERROX; RUN;
proc iml;
USE ERROX;
READ ALL VAR{LX1 LX2 LX3 LX4 LX5 LX6 LX7 LX8 LX9 LX10
    LX11 LX12 LX13 LX14 LX15 LX16 LX17 LX18 LX19 LX20
    LX21 LX22 LX23 LX24} INTO SX; ;
S=(1/153)*SX;
use DADOSVX;
read all into y;
n = nrow(y) ; PRINT N;
p = ncol(y) ; PRINT P;
dfchi = p*(p+1)*(p+2)/6 ; PRINT DFCHI;
q = i(n) - (1/n)*j(n,n,1); PRINT Q;
s_inv = inv(s) ;

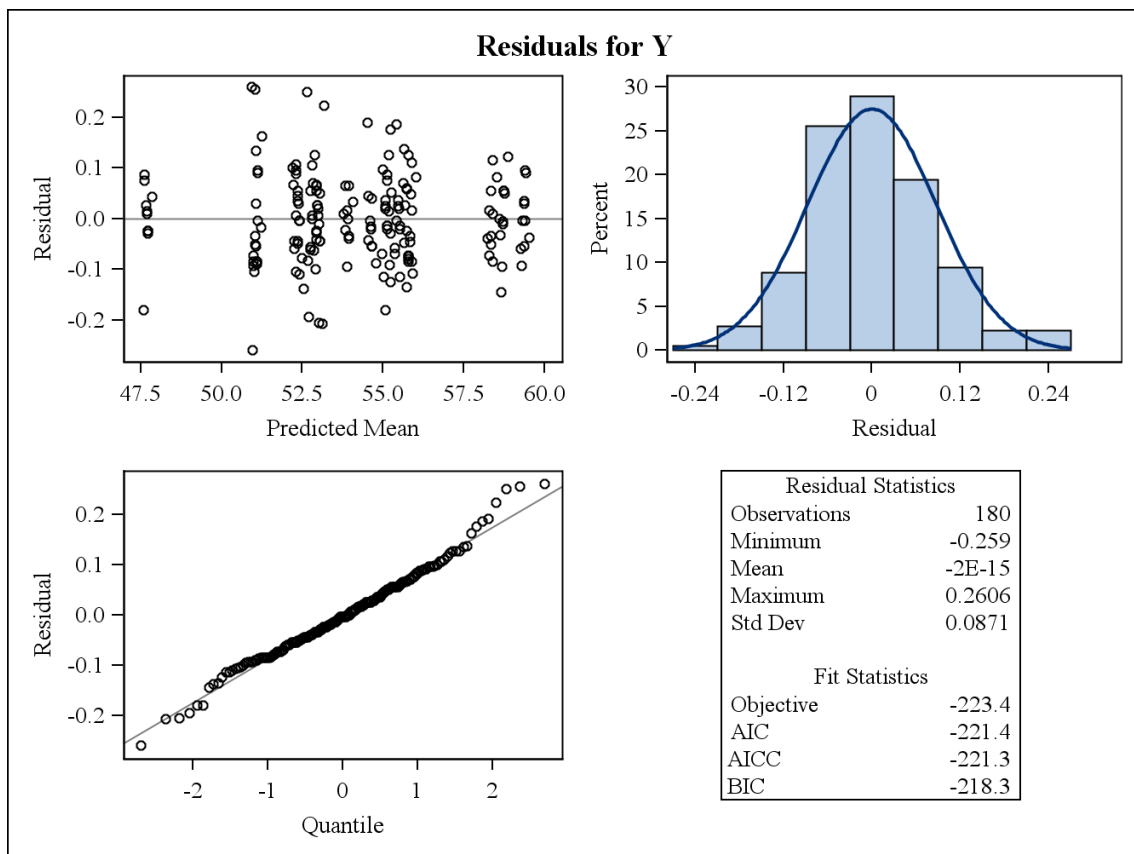
```

```

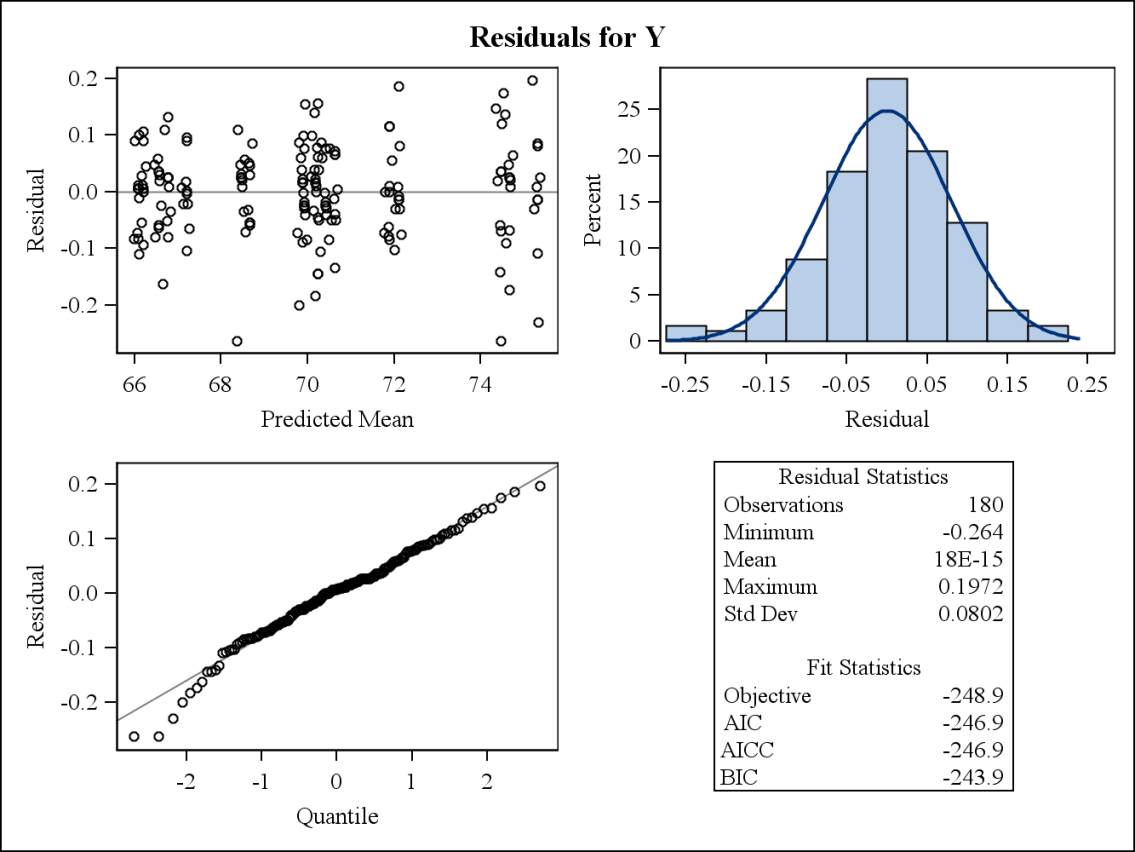
g_matrix = q*y*s_inv*y`*q;
beta1hat = ( sum(g_matrix#g_matrix#g_matrix) )/(n*n);
beta2hat =trace( g_matrix#g_matrix )/n ;
kappa1 = n*beta1hat/6 ;
kappa2 = (beta2hat - p*(p+2) ) /sqrt(8*p*(p+2)/n) ;
pvalskew = 1 - probchi(kappa1,dfchi) ;
pvalkurt = 2*( 1 - probnorm(abs(kappa2)) );
print s ;
print s_inv ;
print 'TESTS: ';
print 'Based on skewness: ' beta1hat kappa1 pvalskew ;
print 'Based on kurtosis: ' beta2hat kappa2 pvalkurt;
QUIT;

```

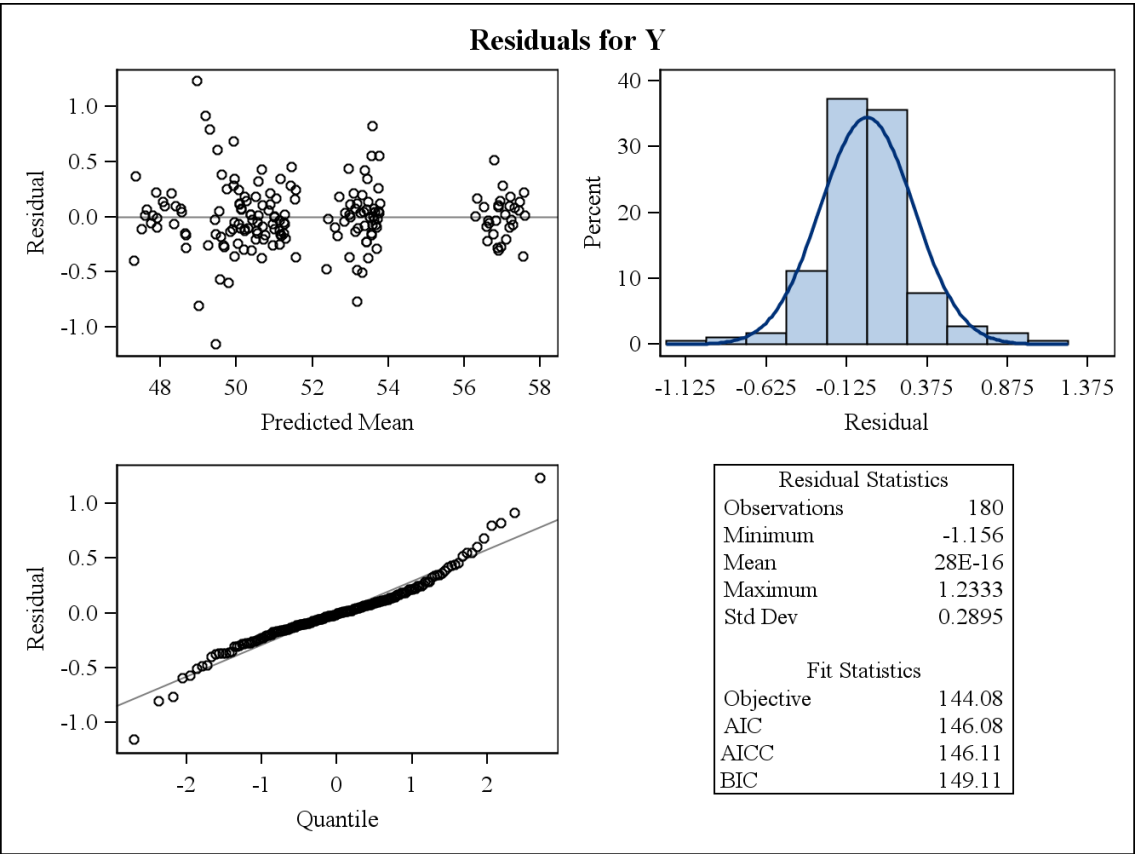
LOCAL=ALAR DIREITO



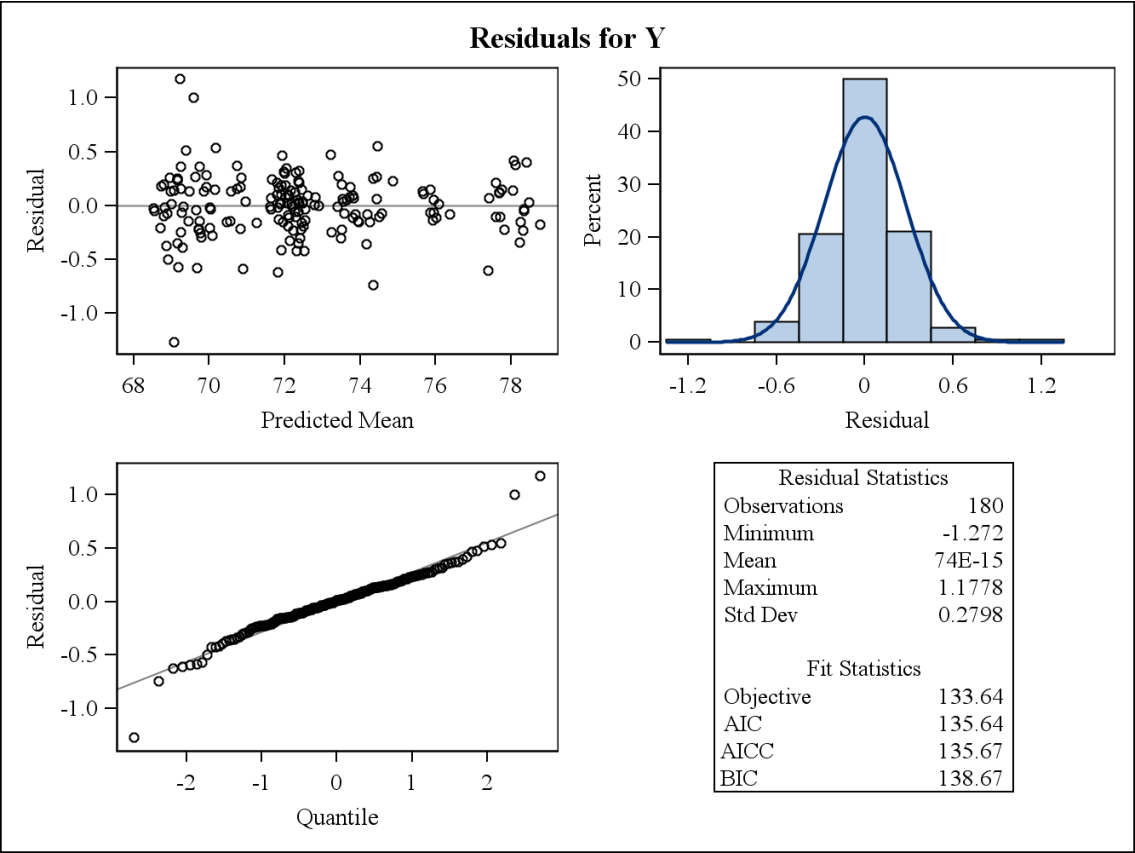
LOCAL=ALAR ESQUERDO



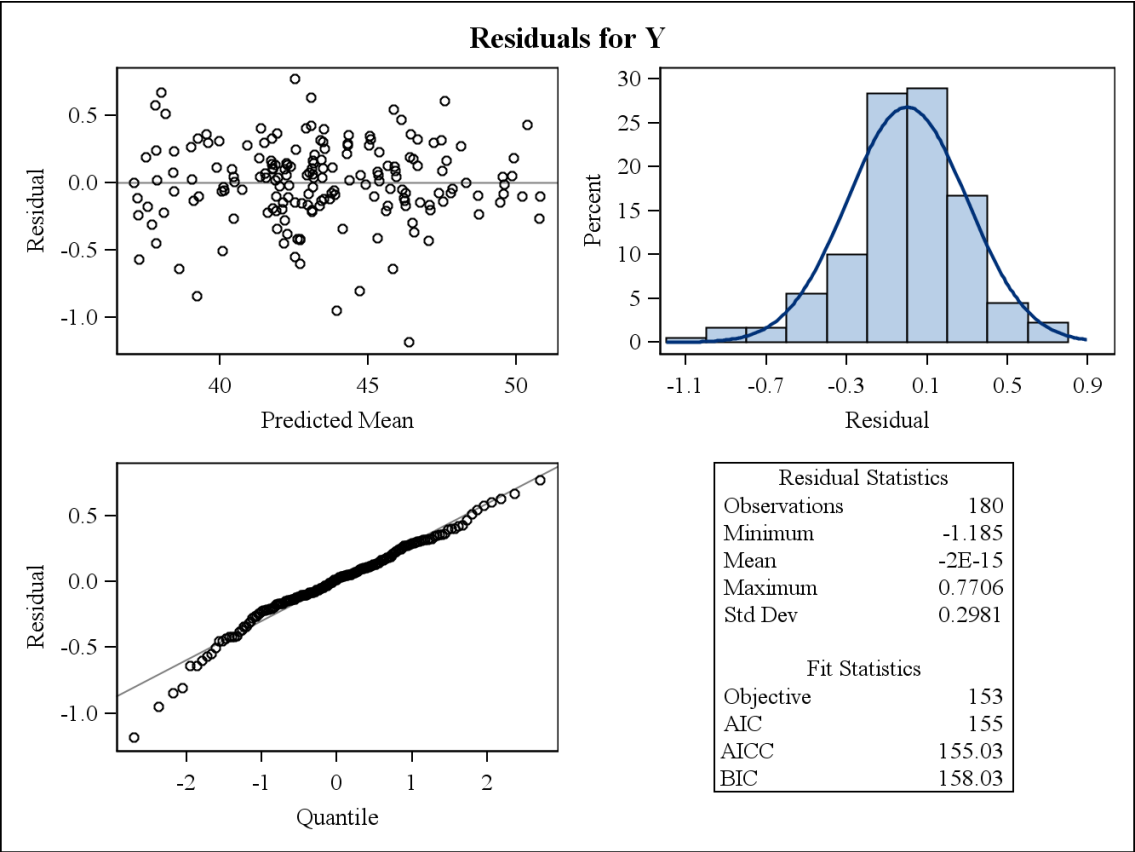
LOCAL=CHELION DIREITO



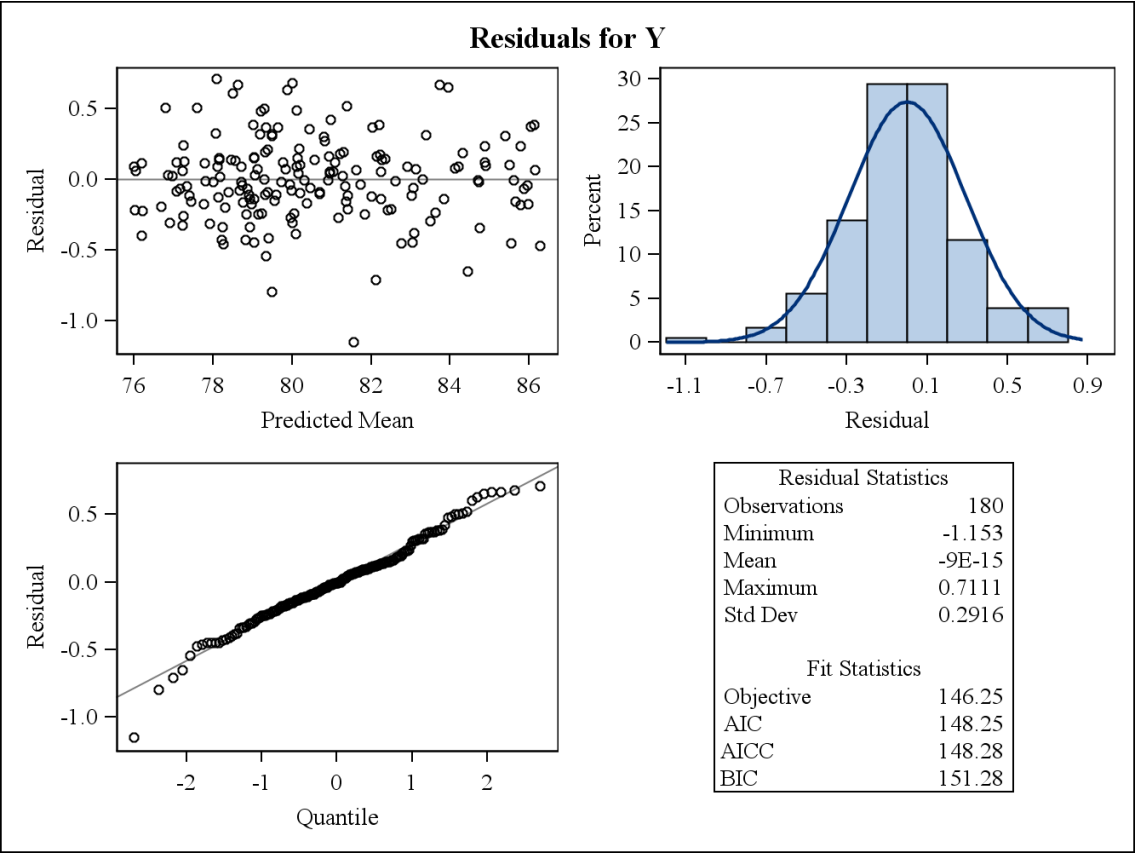
LOCAL=CHELION ESQUERDO



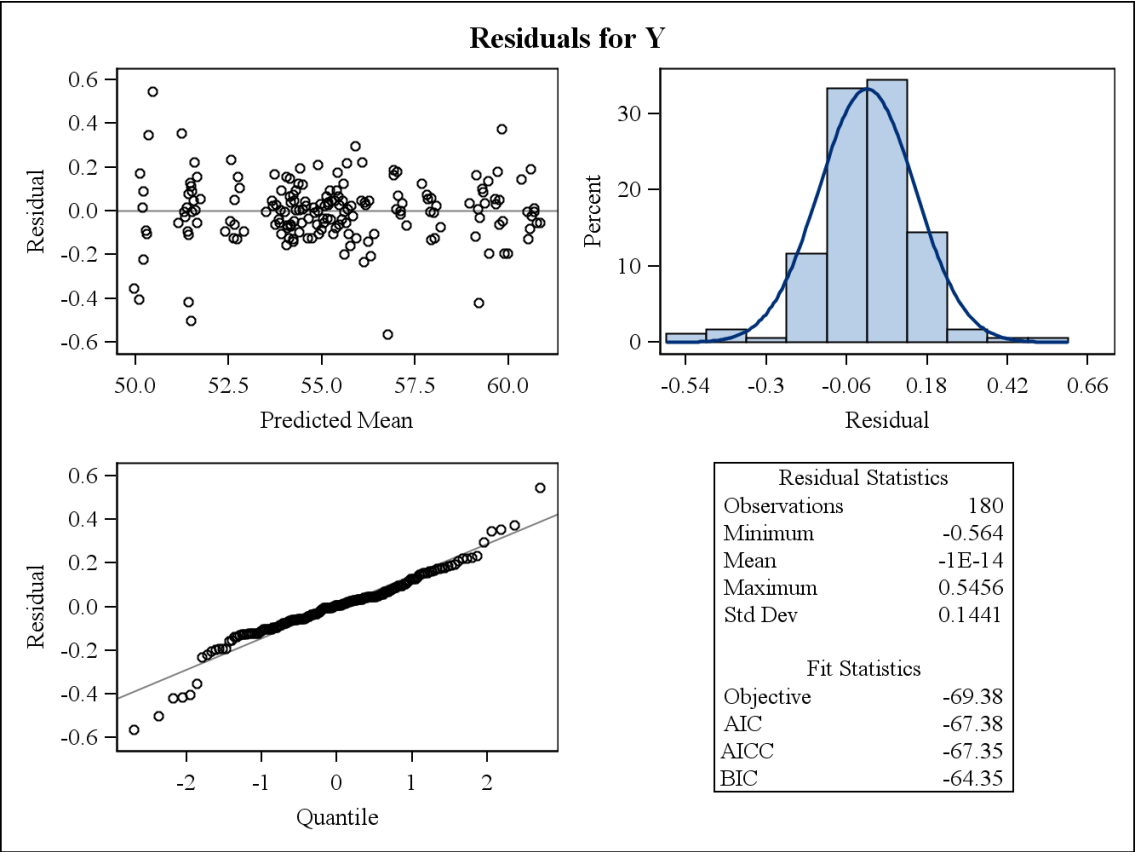
LOCAL=ECTOANTHION DIREITO



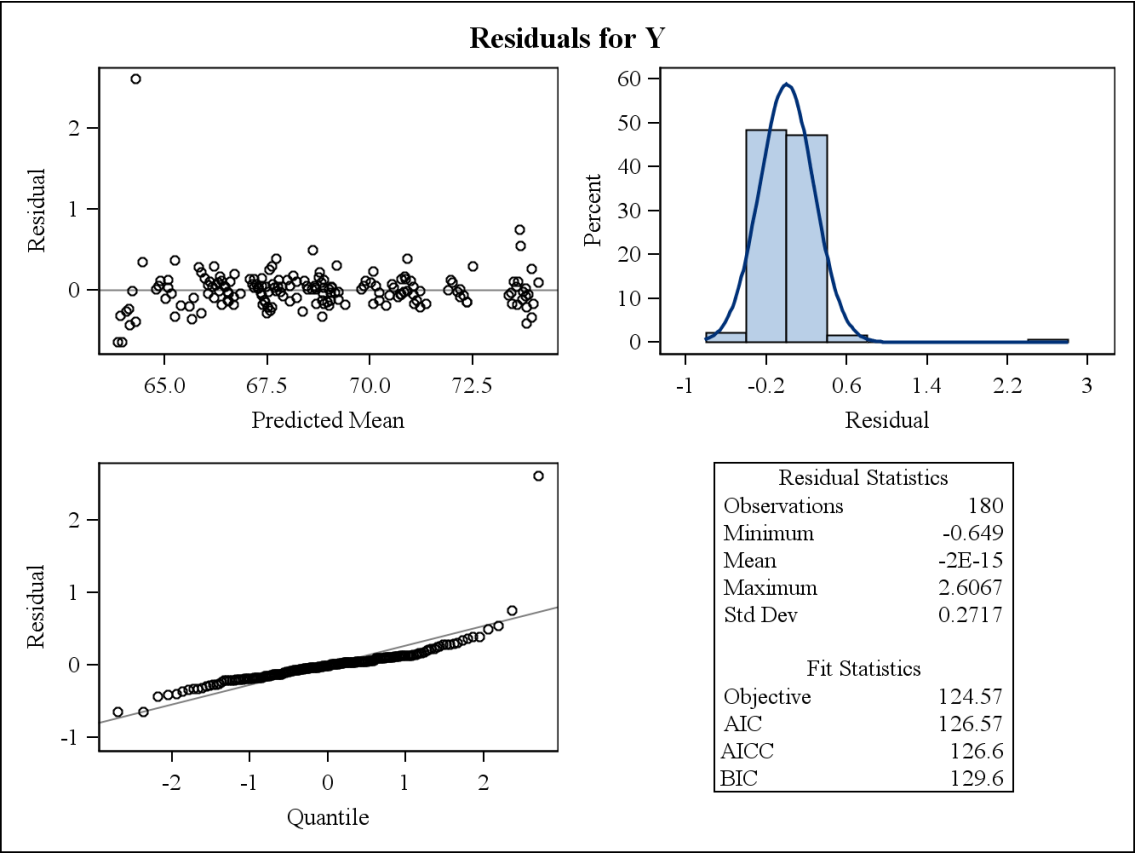
LOCAL=ECTOANTHION ESQUERDO



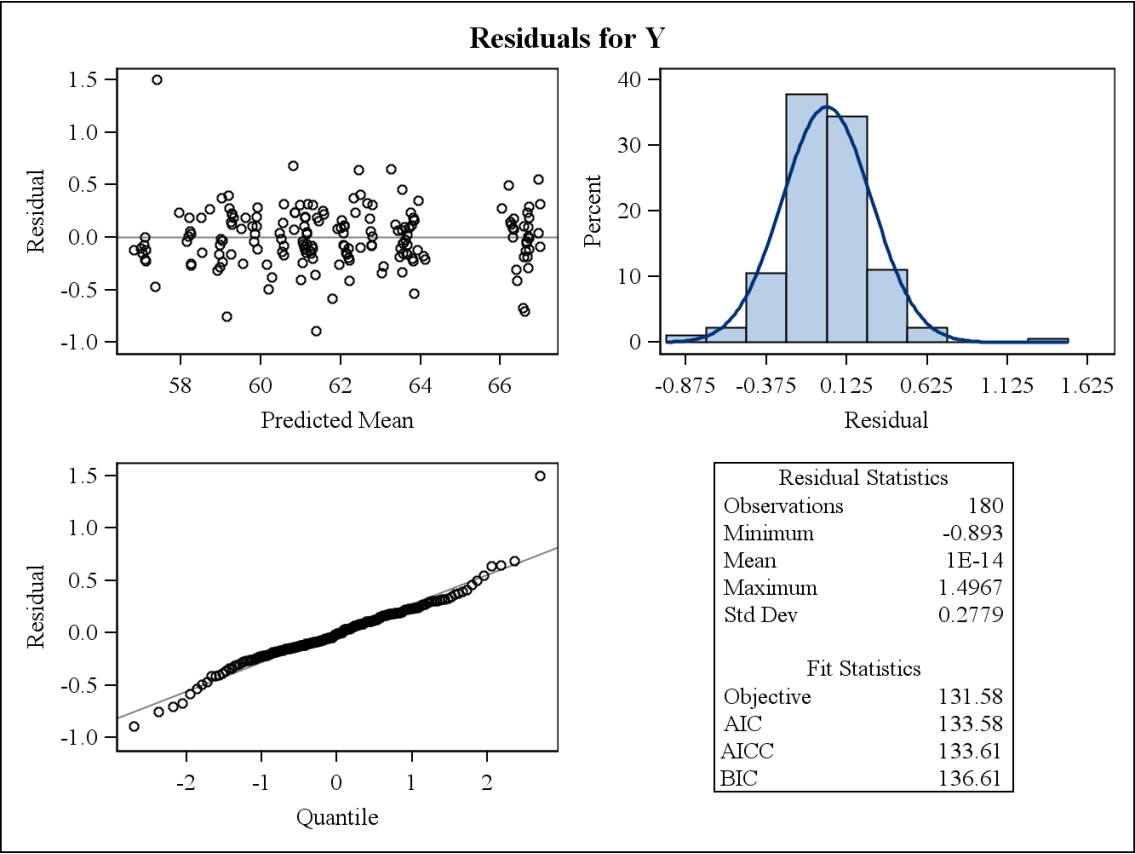
LOCAL=ENDOCANTHION DIREITO



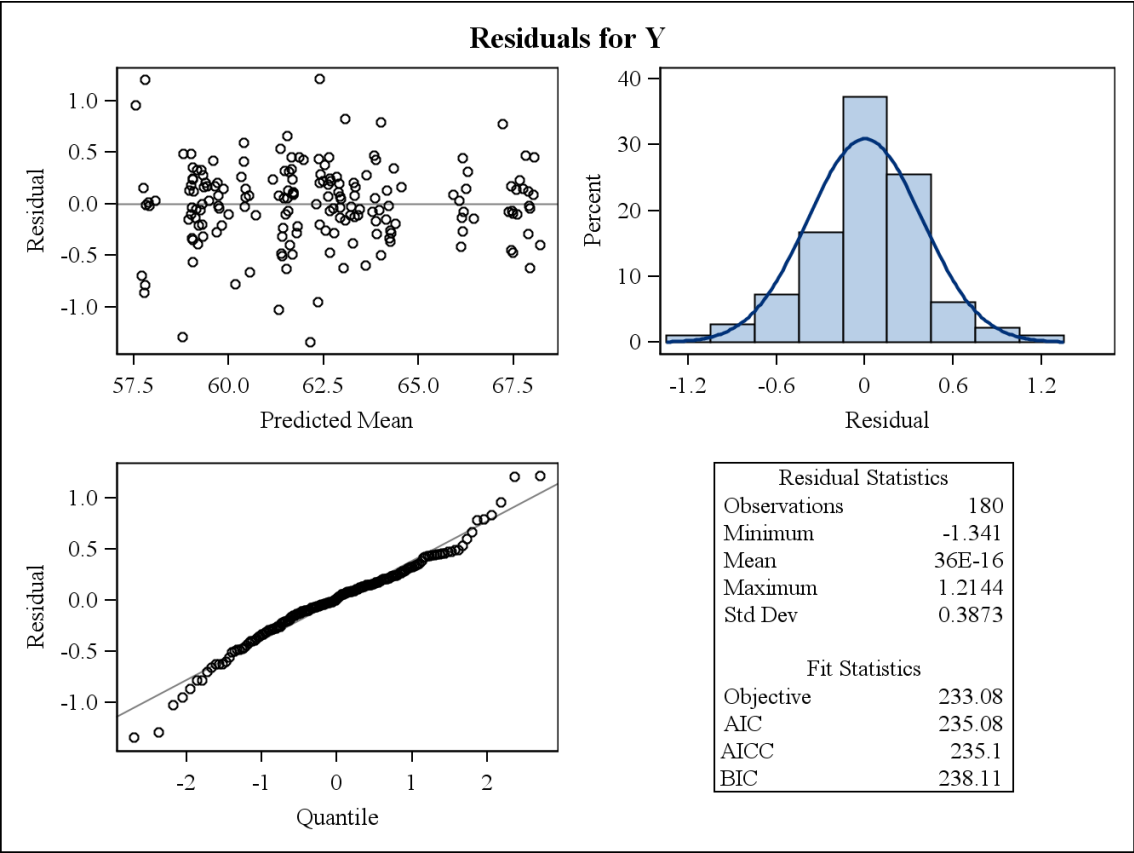
LOCAL=ENDOCANTHION ESQUERDO



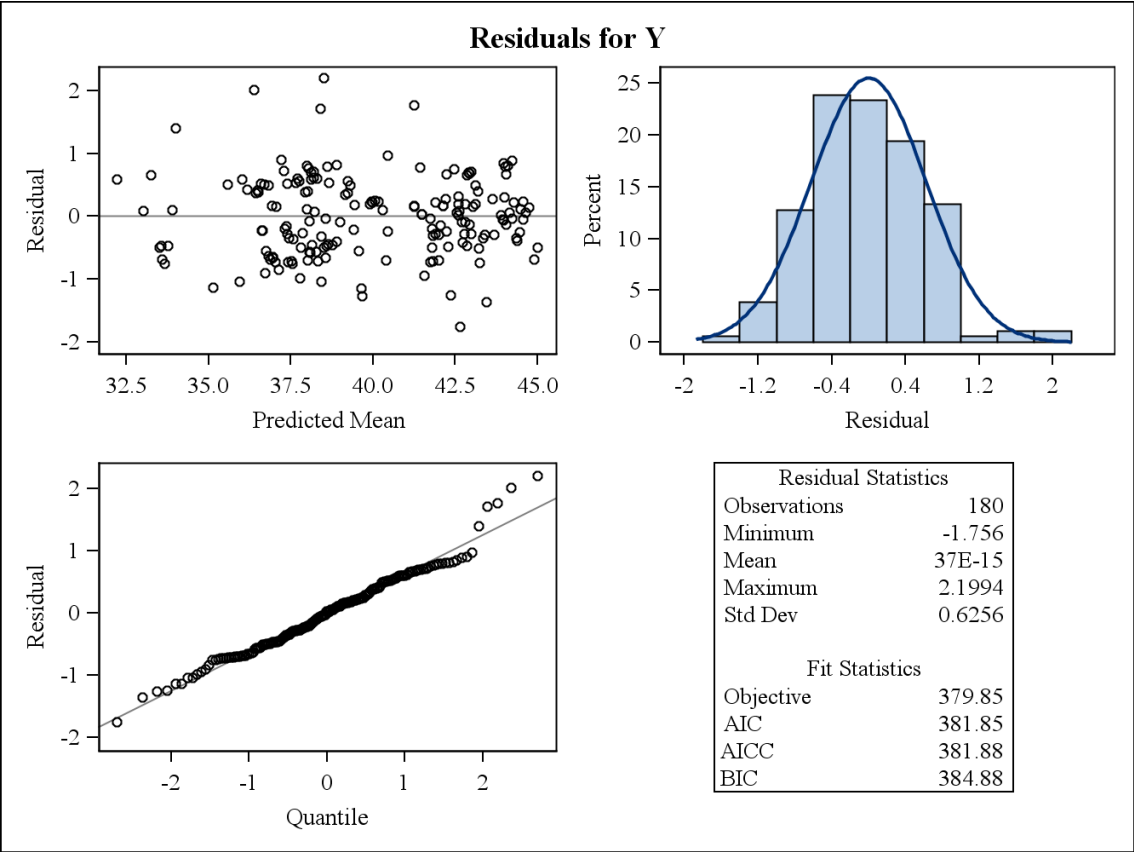
LOCAL=GLABELA



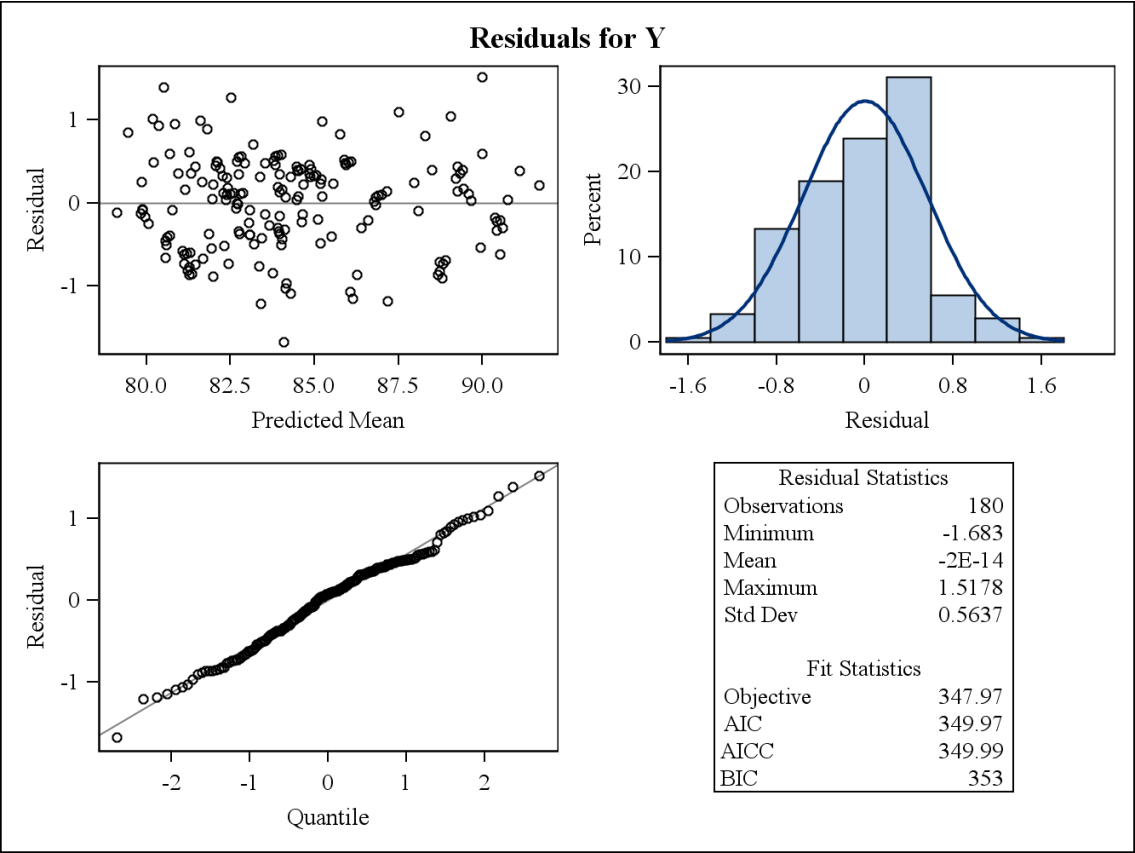
LOCAL=GNATHION



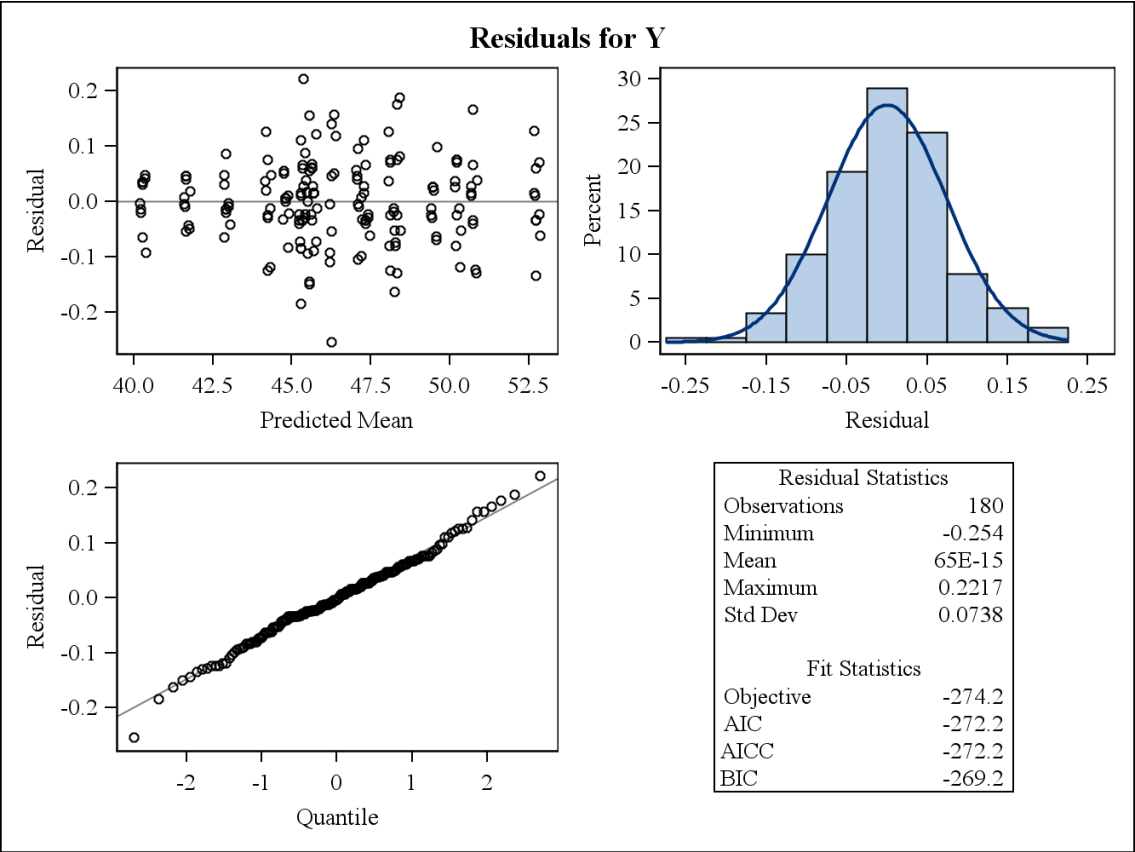
LOCAL= GONIO DIREITO



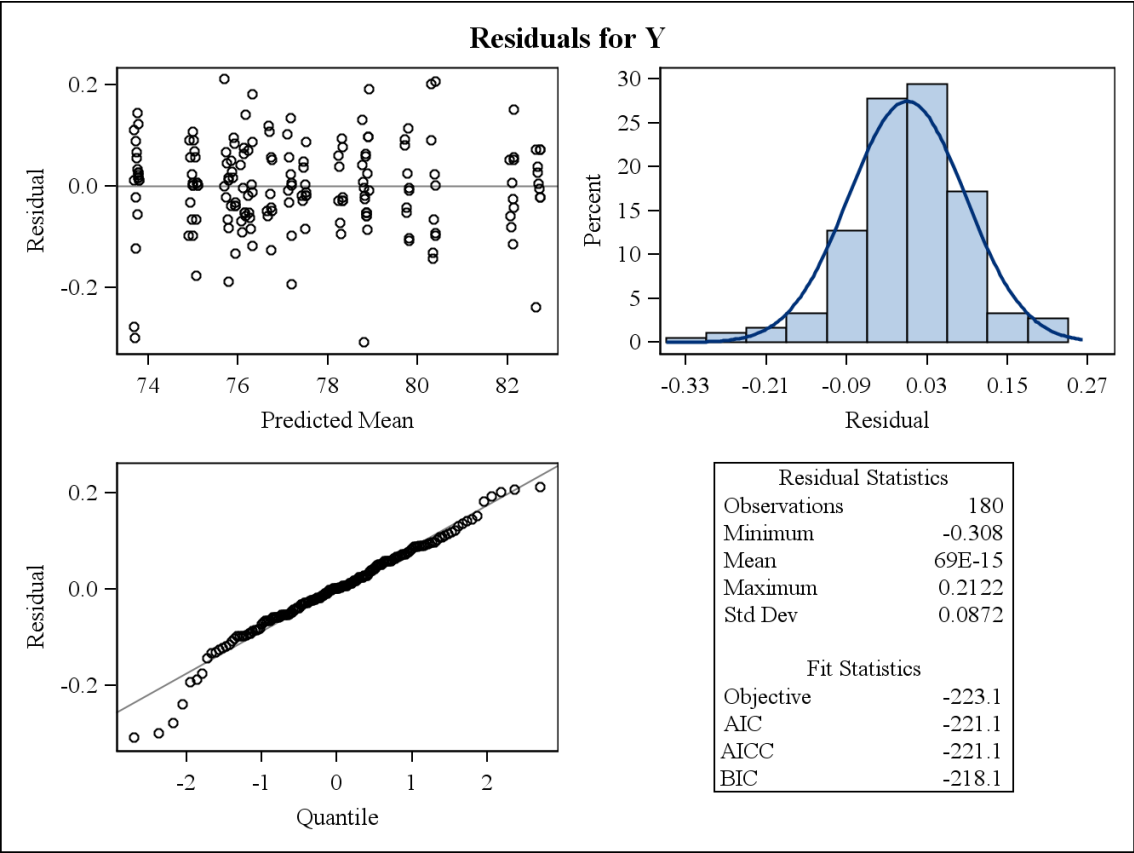
LOCAL= GONIO ESQUERDO



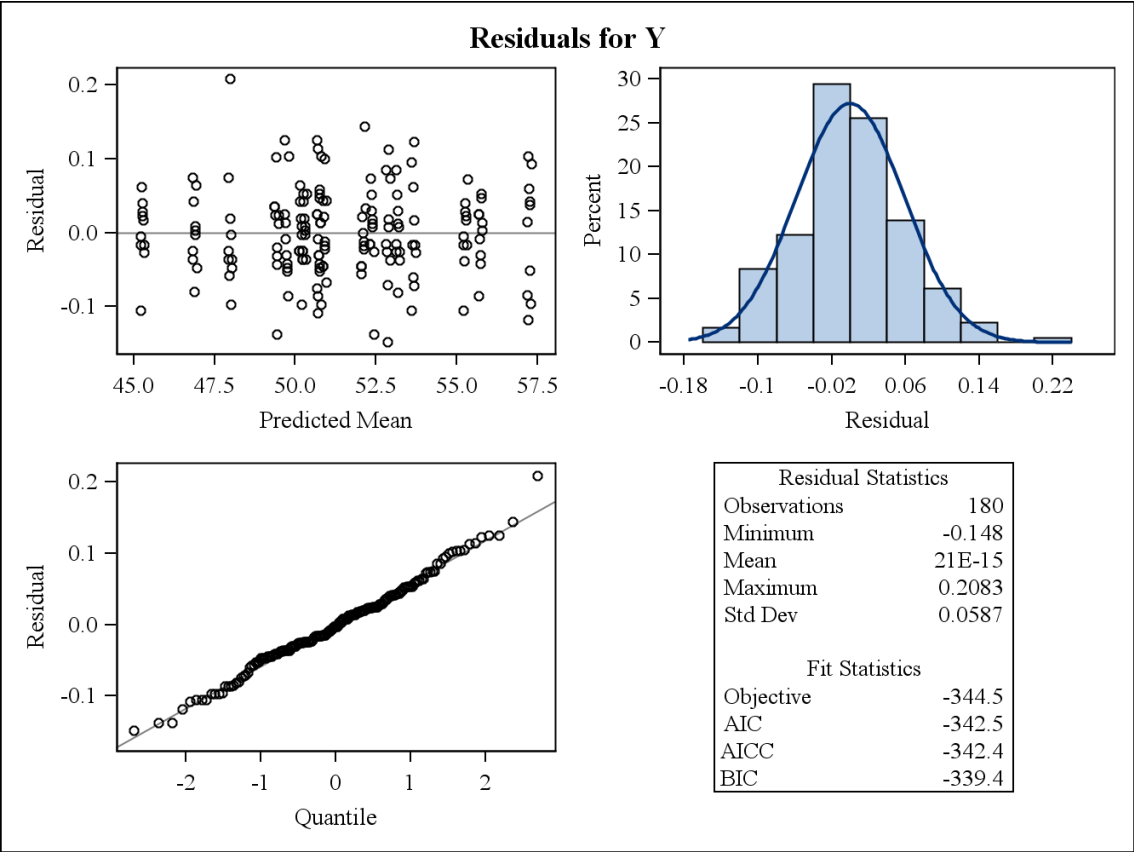
LOCAL=IRIDION LATERAL DIREITO



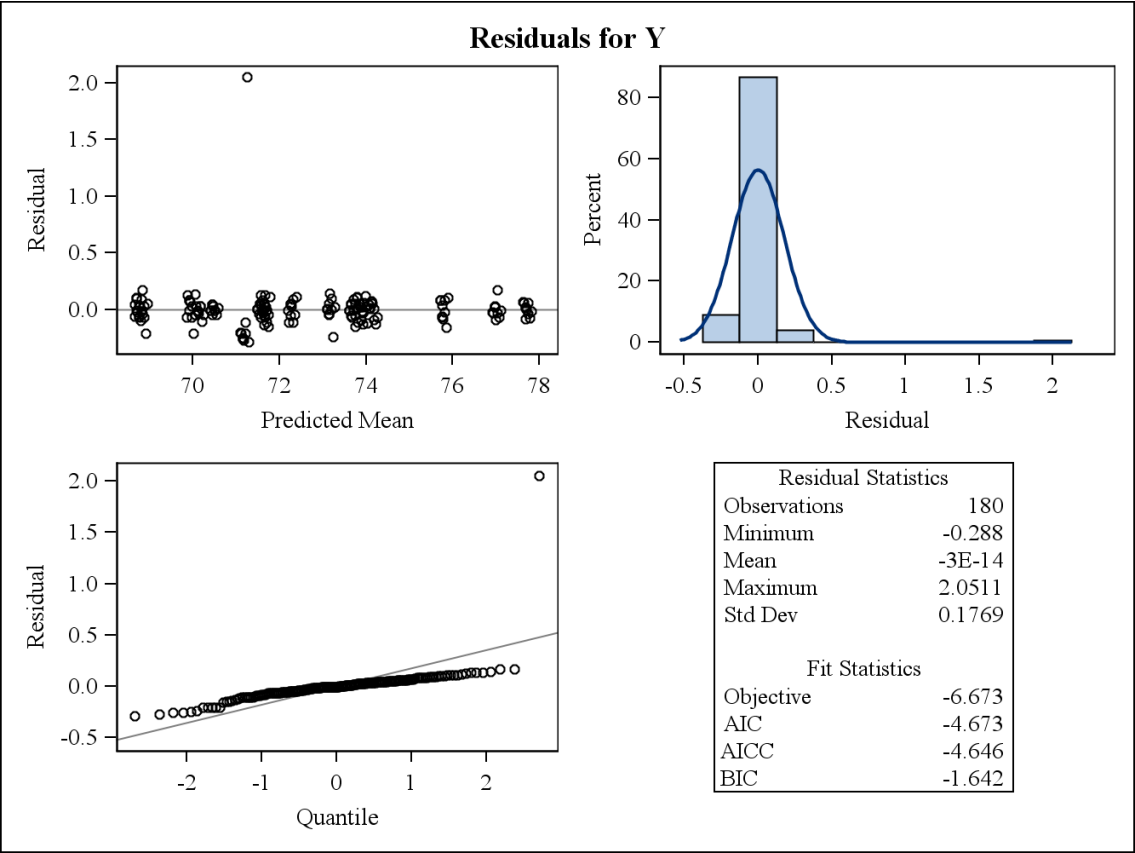
LOCAL=IRIDION LATERAL ESQUERDO



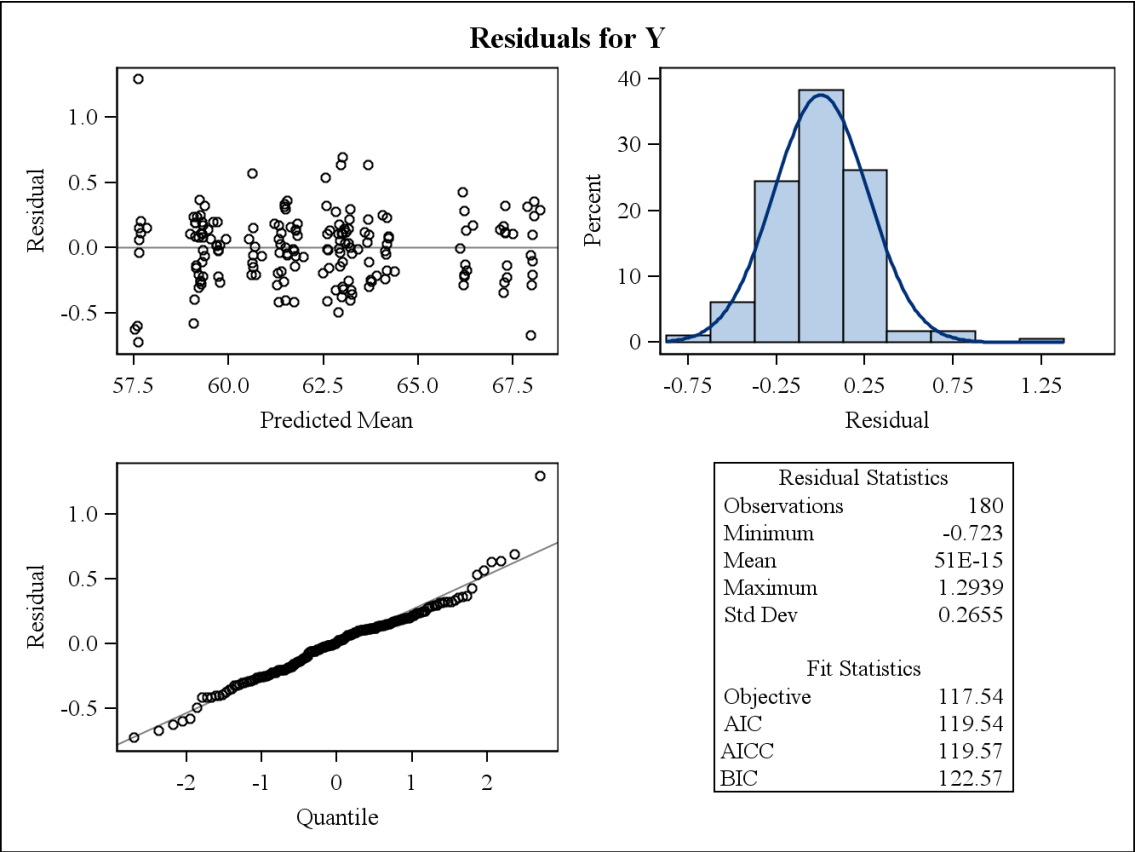
LOCAL=IRIDION MEDIAL DIREITO



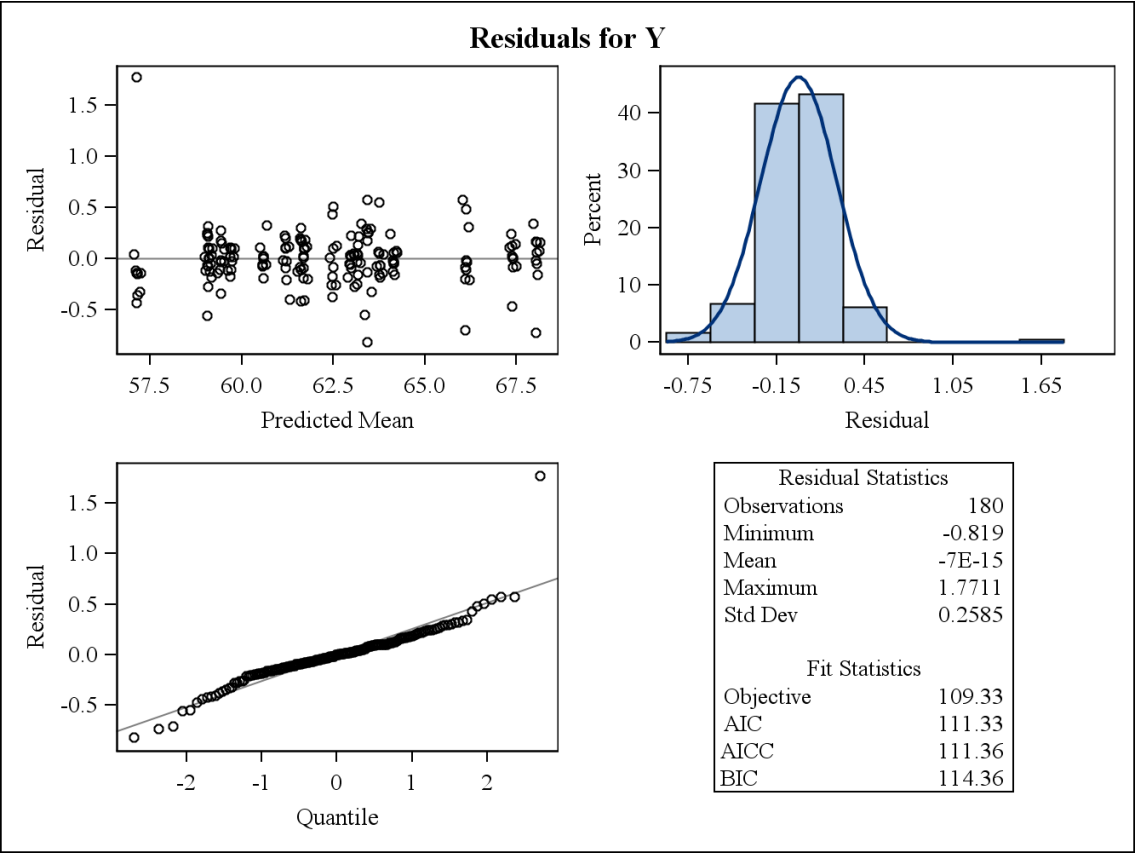
LOCAL=IRIDION MEDIAL ESQUERDO



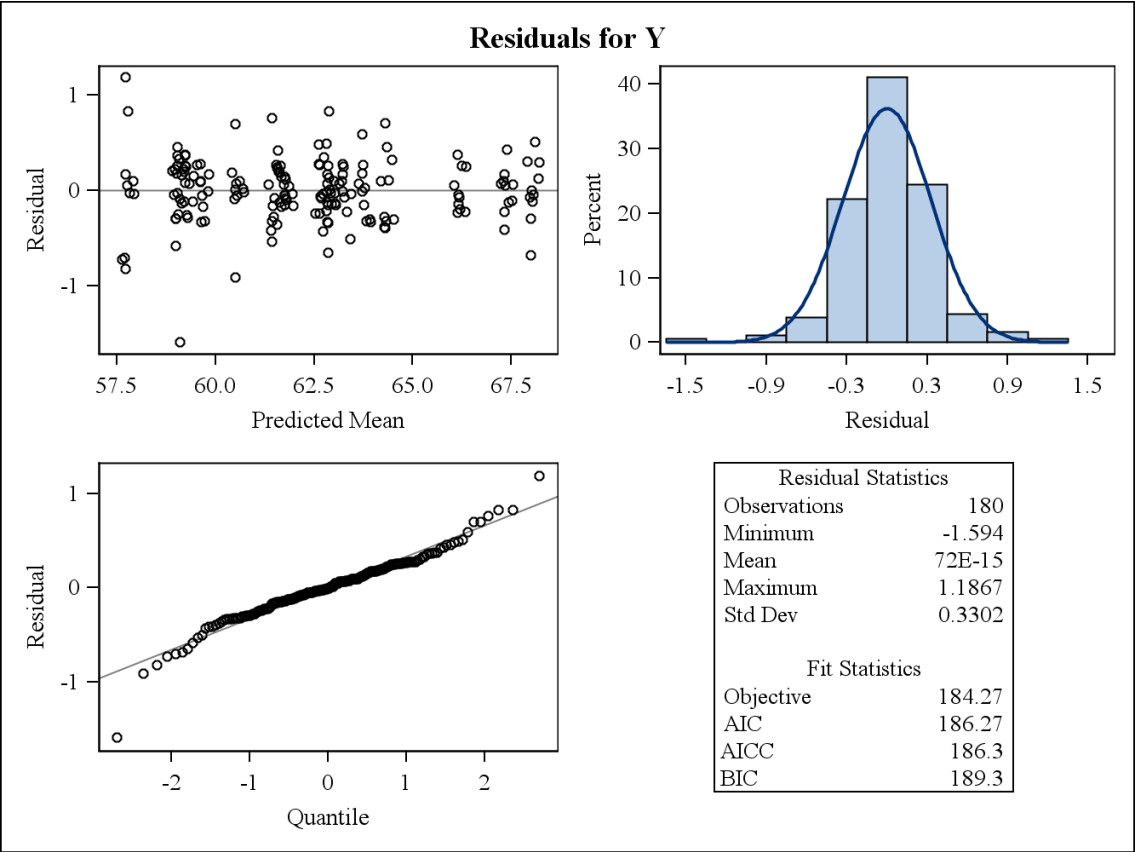
LOCAL=LABIAL INFERIOR



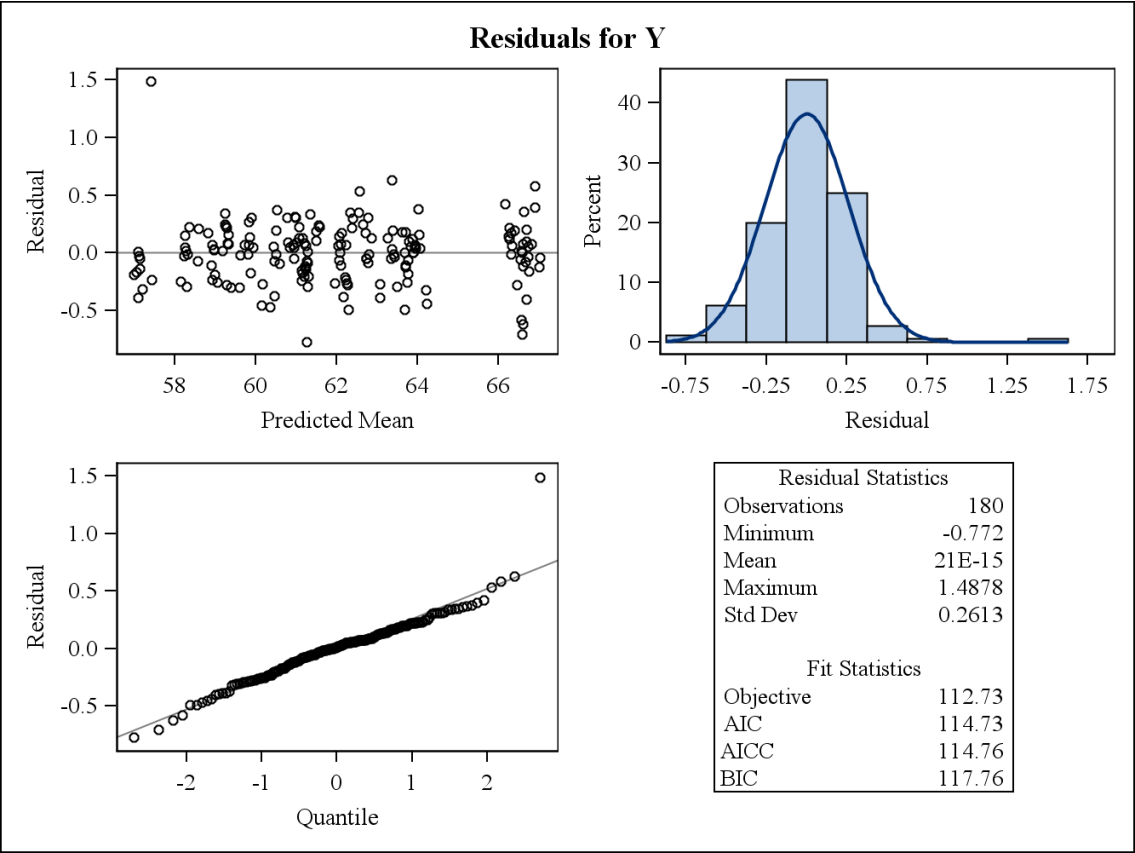
LOCAL=LABIAL SUPERIOR



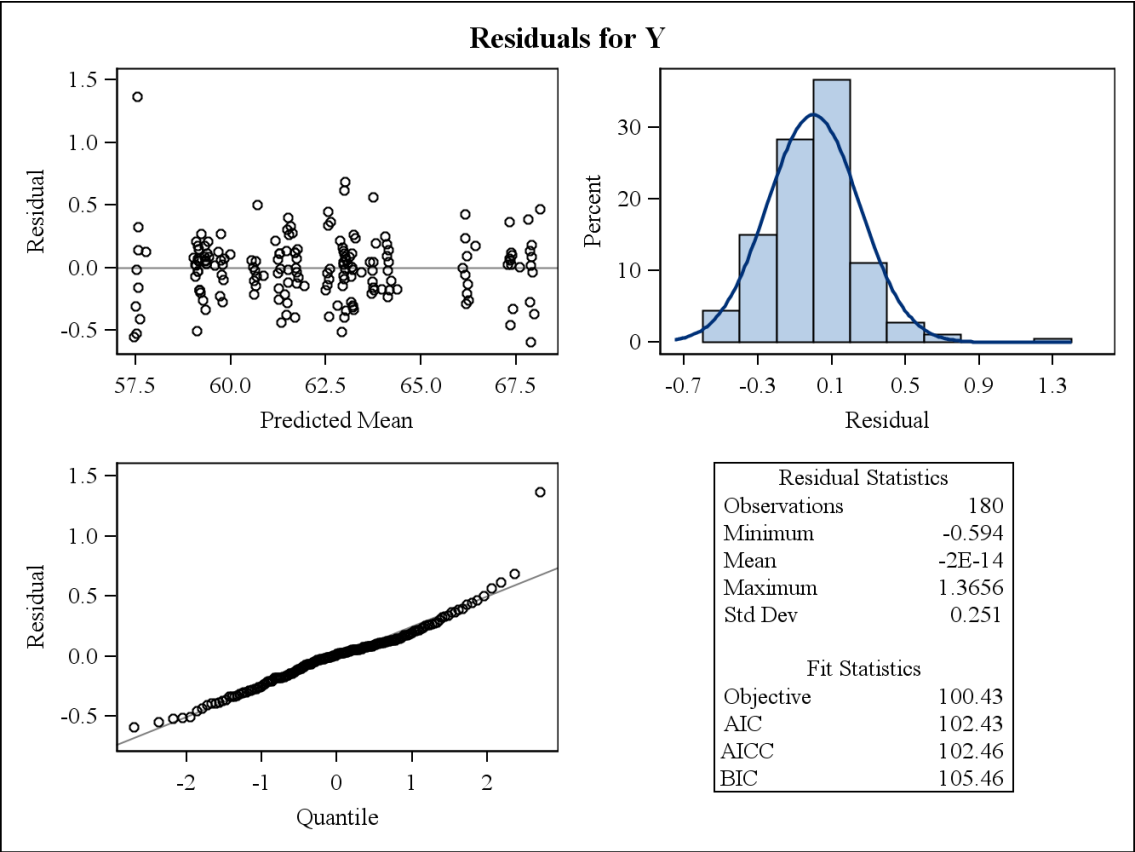
LOCAL=LABIOMENTAL



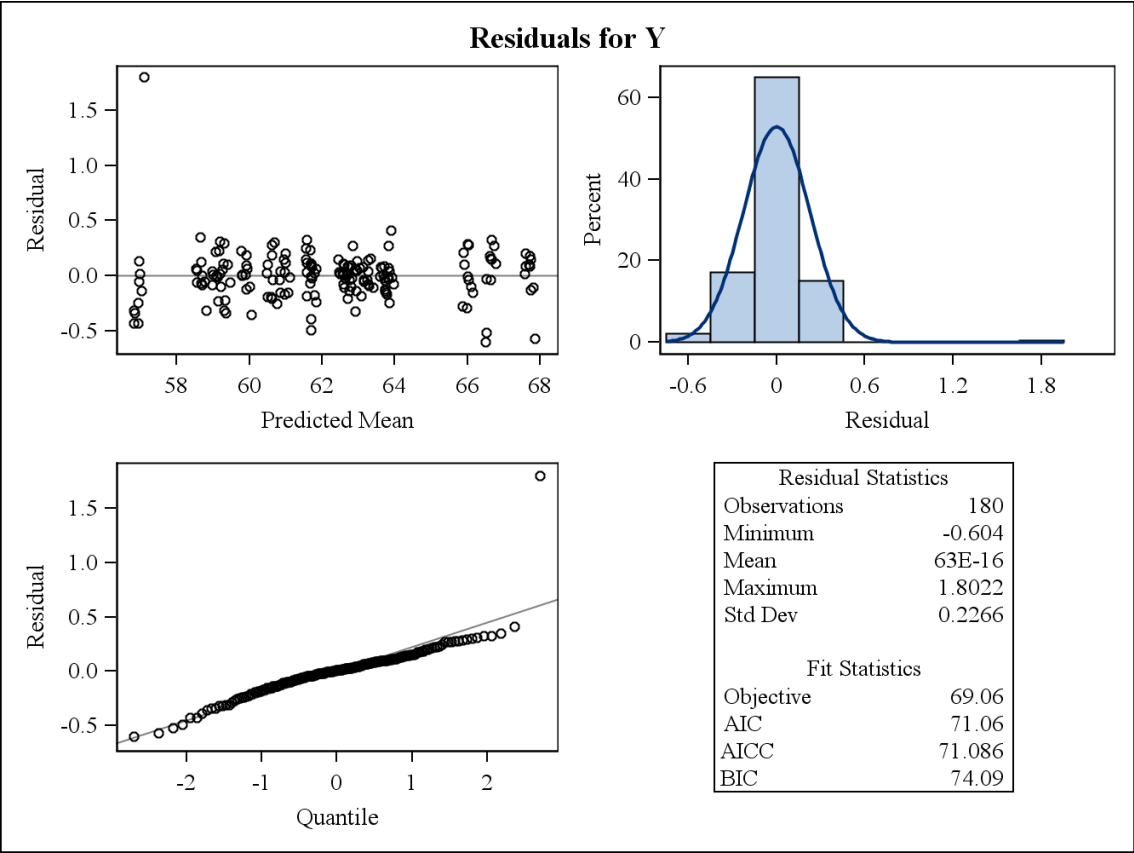
LOCAL=NASION



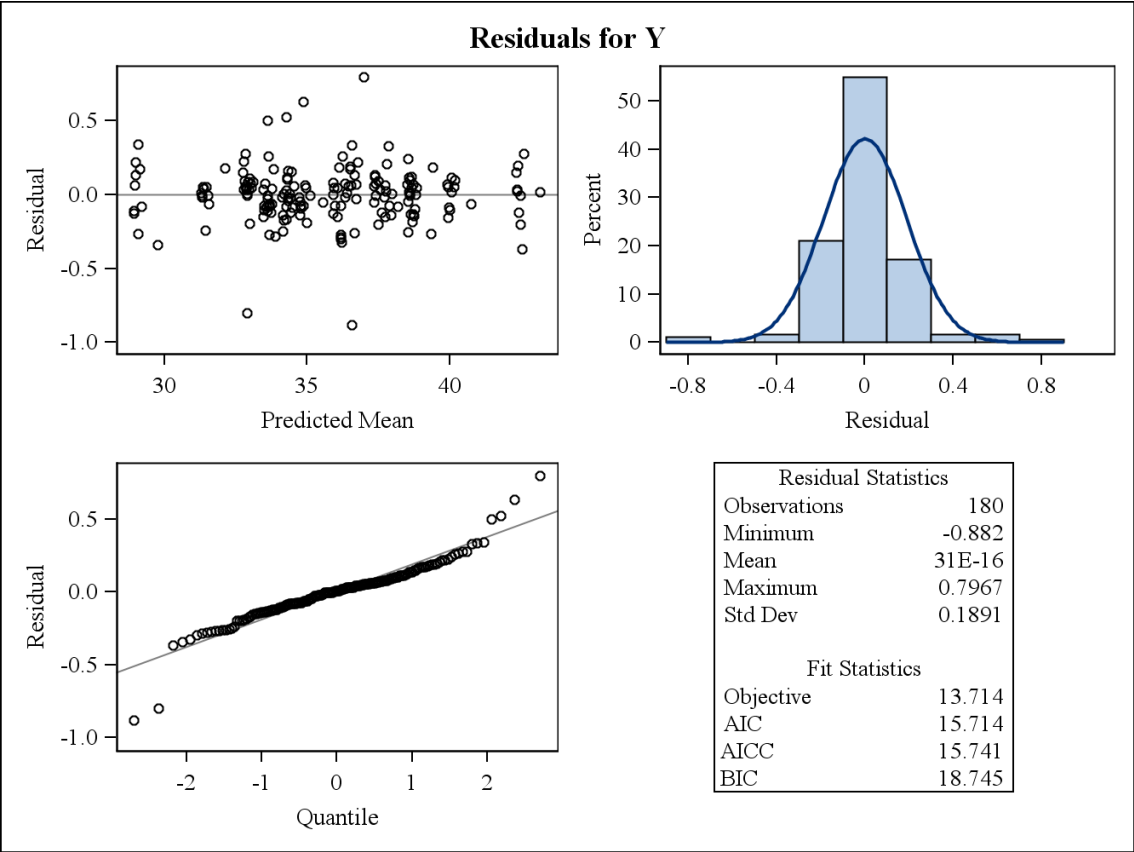
LOCAL=STOMION



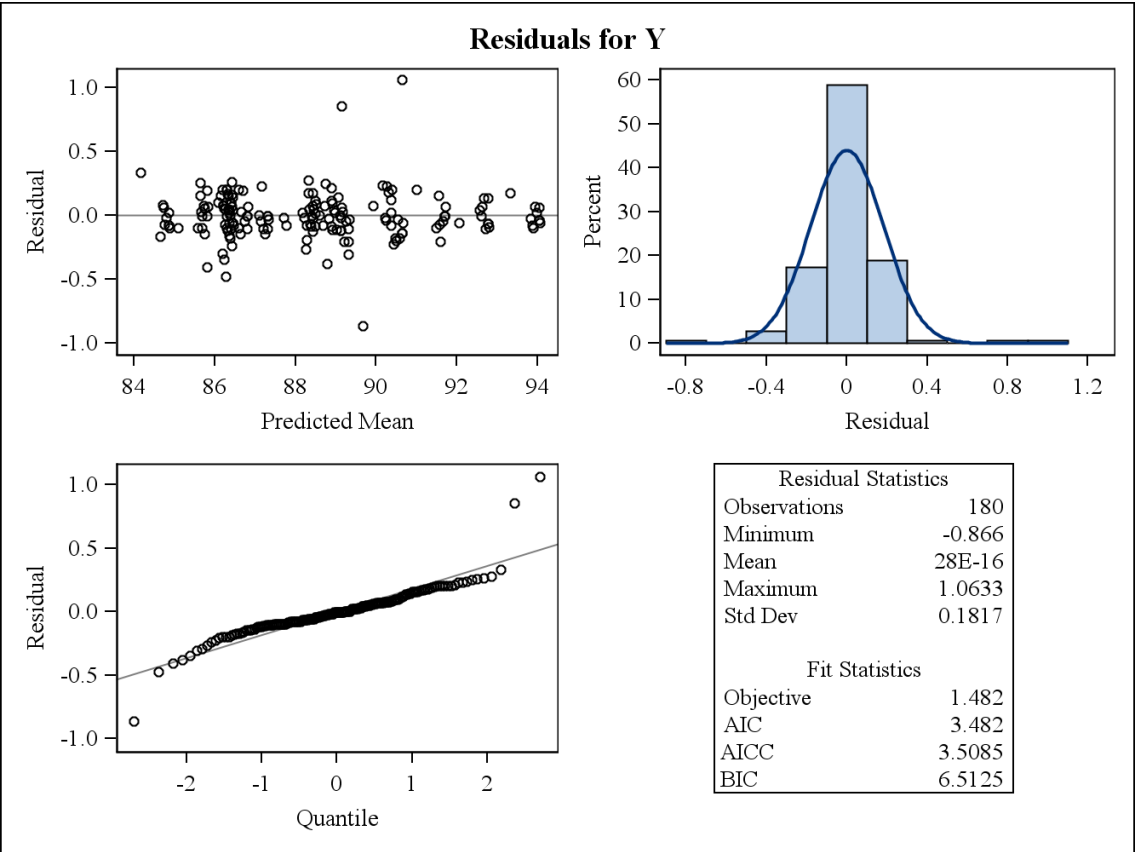
LOCAL=SUBNASAL



LOCAL=ZYGION DIREITO

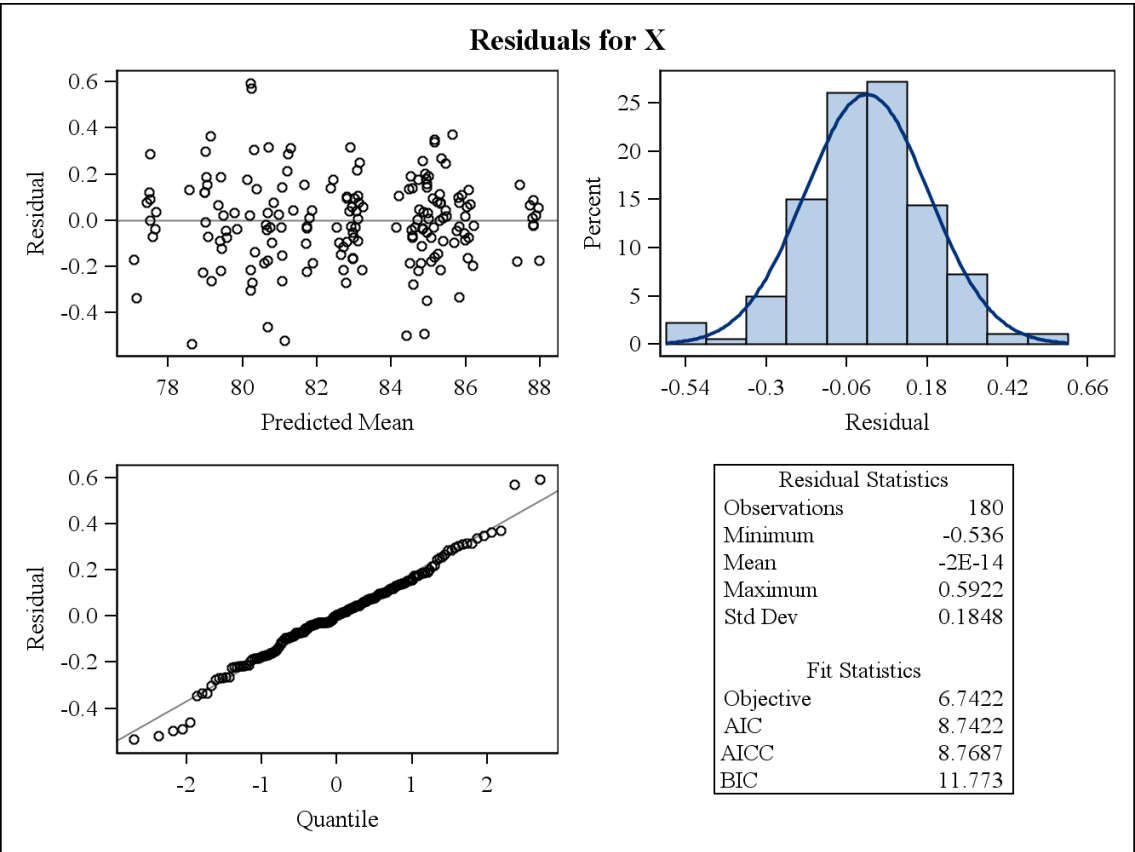


LOCAL=ZYGION ESQUERDO

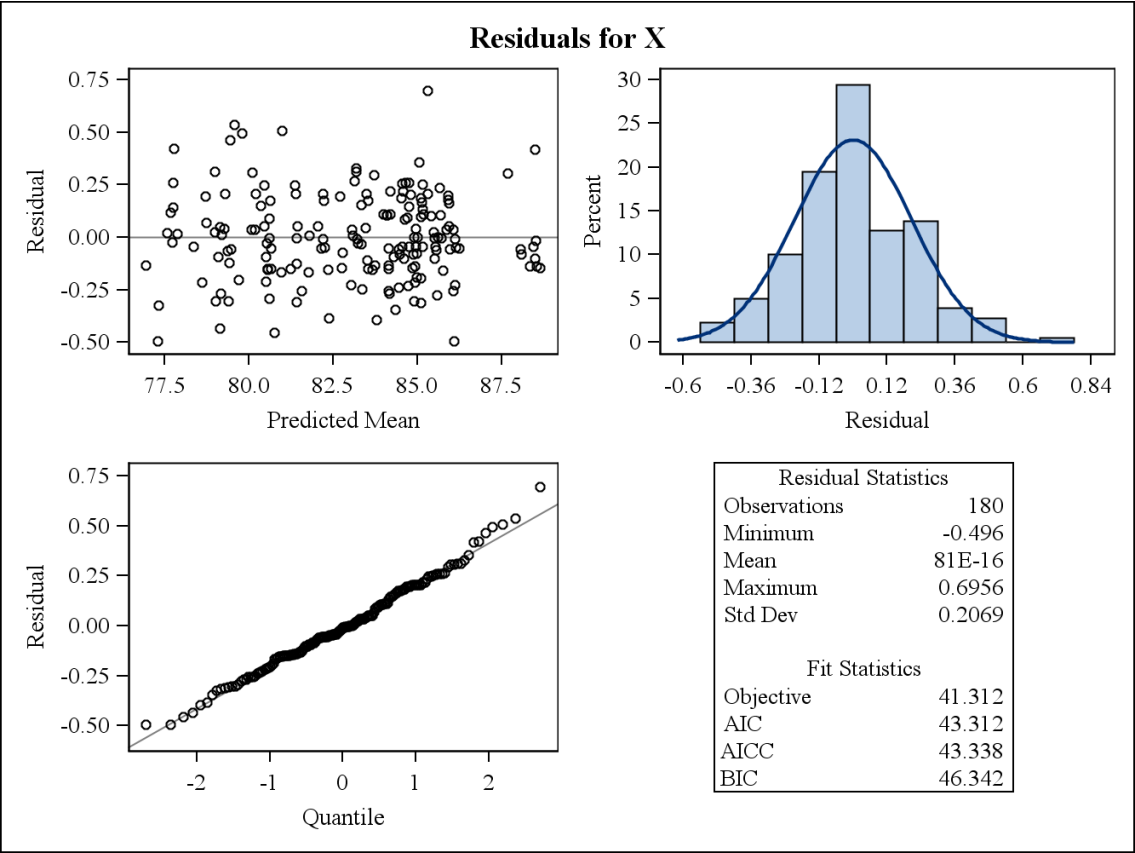


ANÁLISE EM X

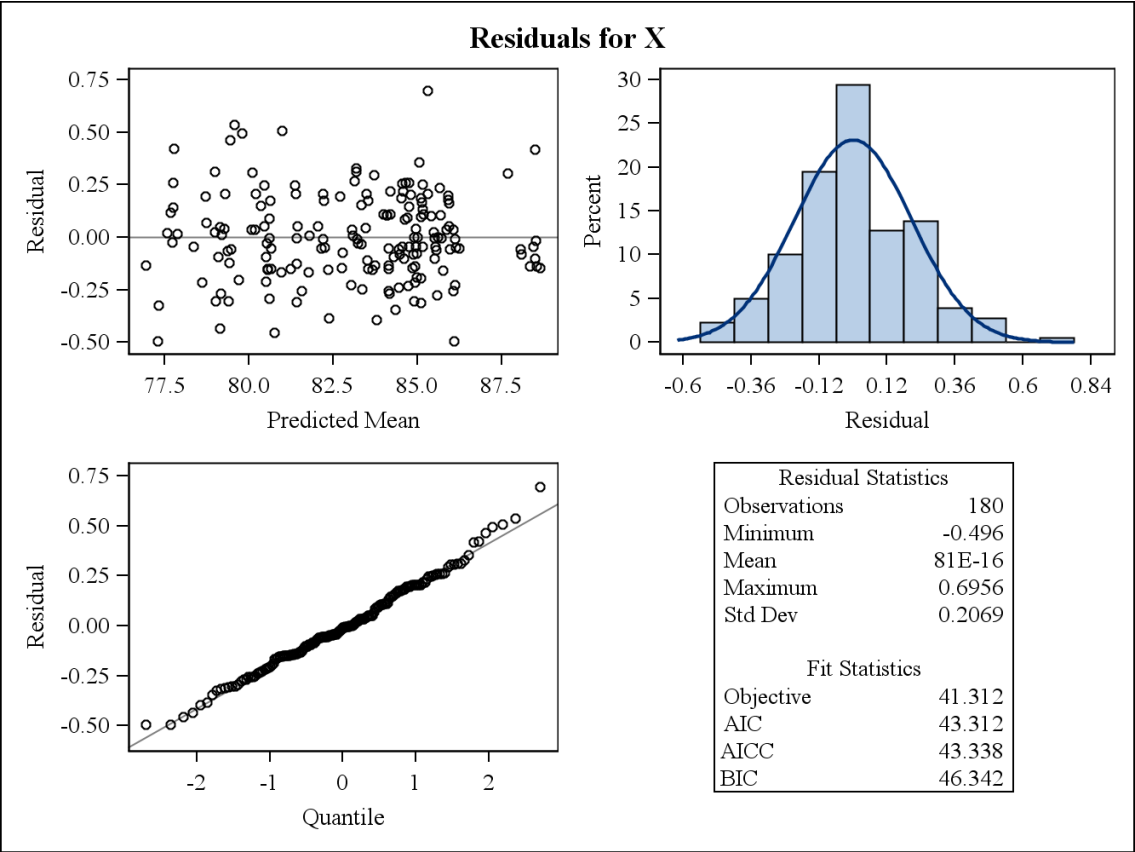
LOCAL=ALAR DIREITO



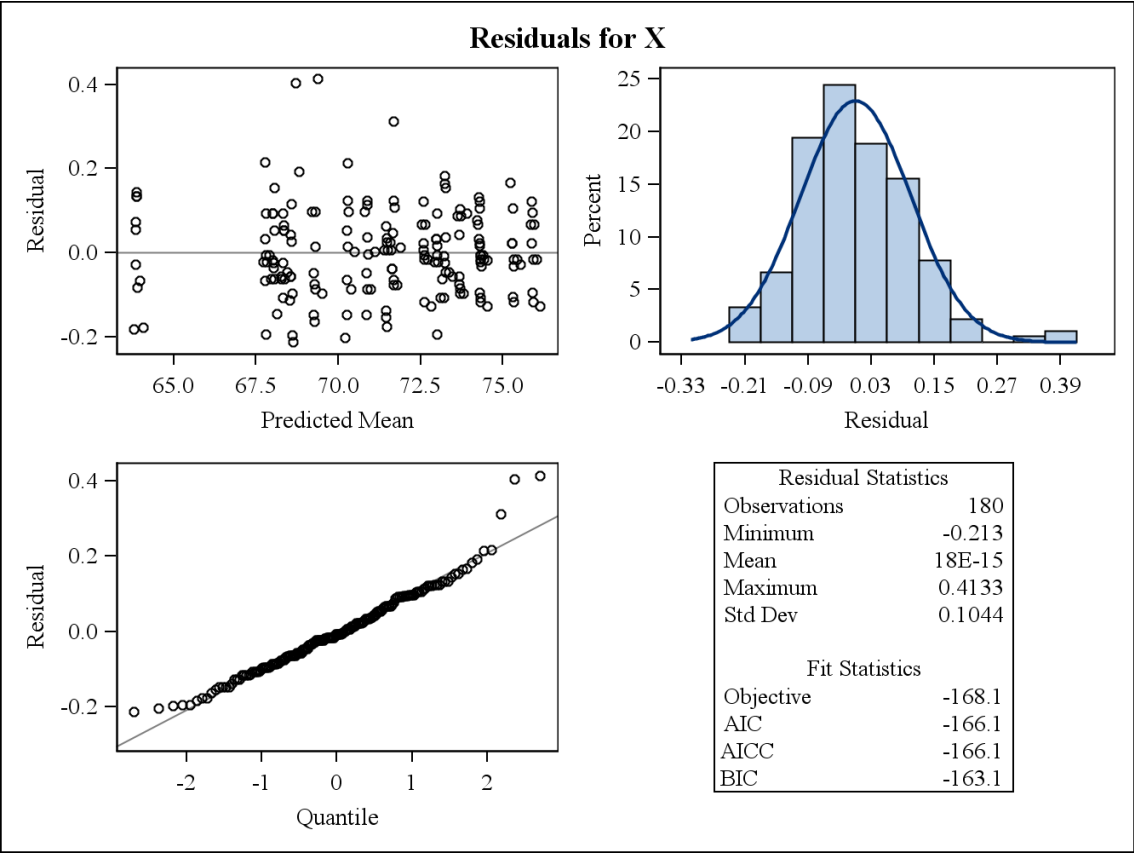
LOCAL= ALAR ESQUERDO



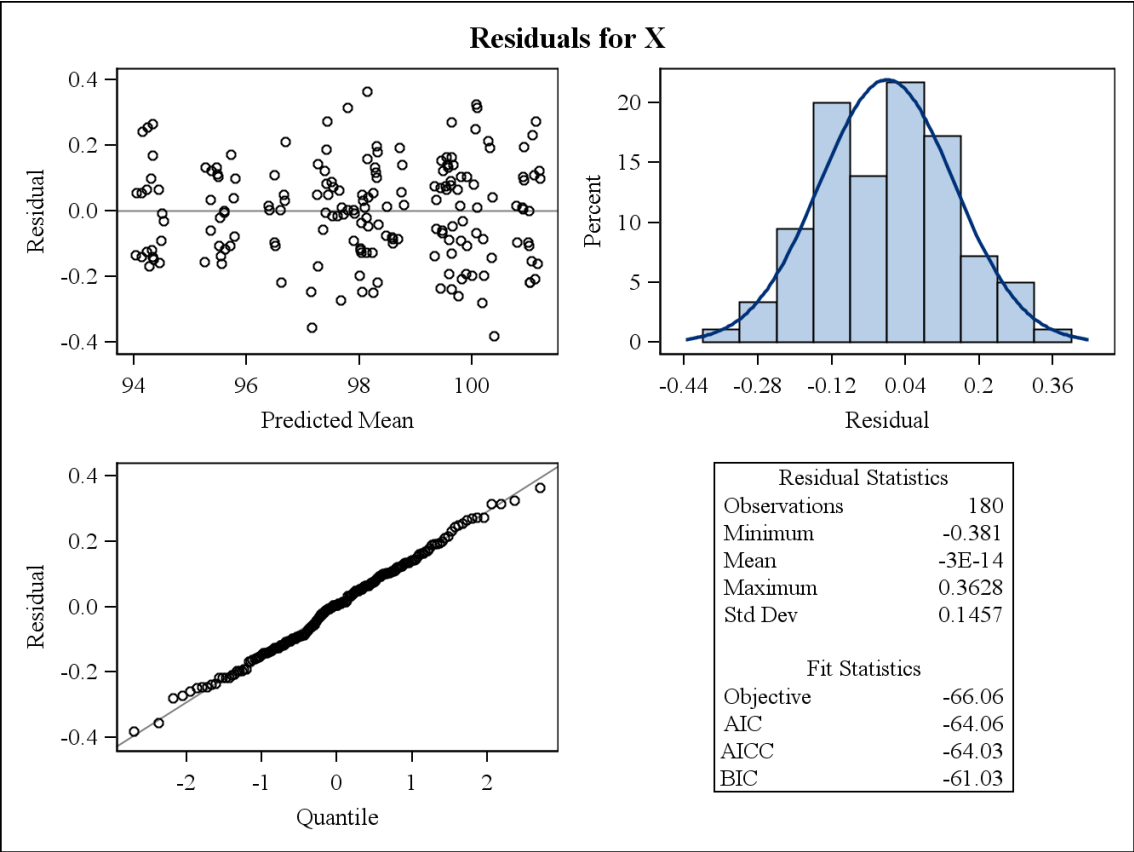
LOCAL=CHELION DIREITO



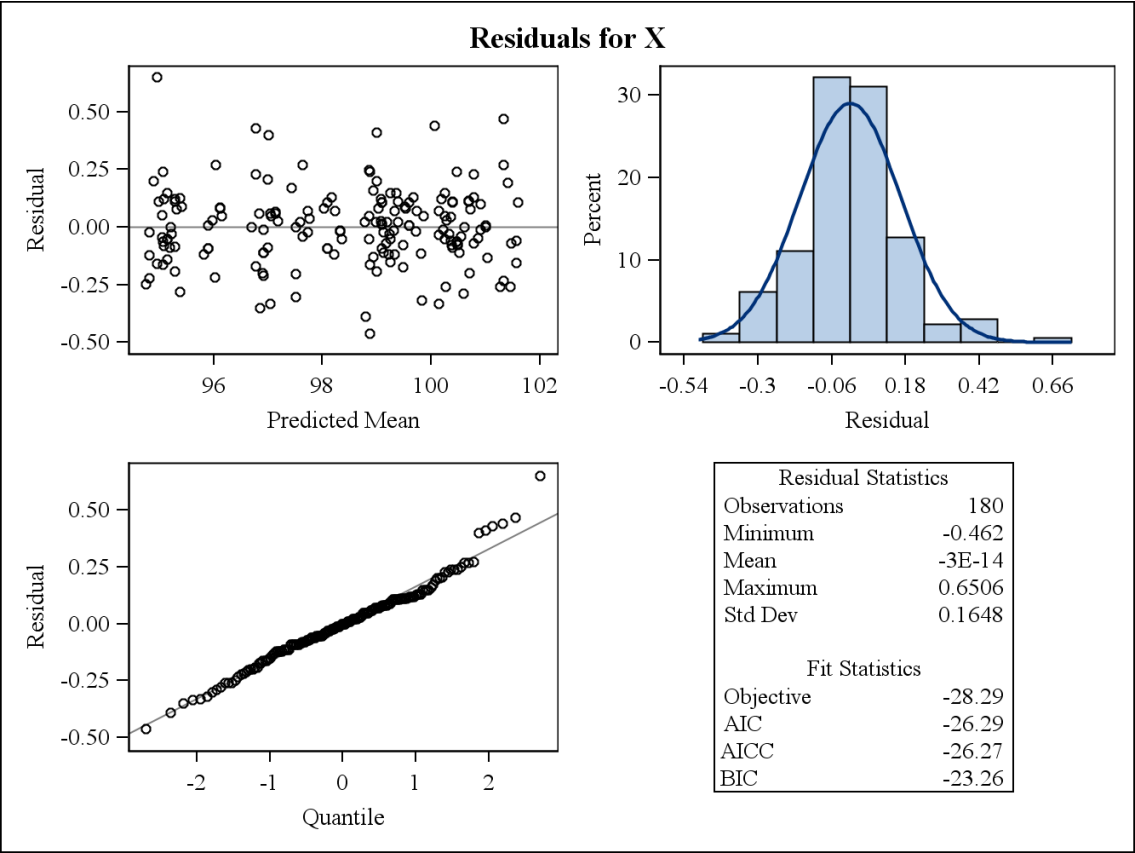
LOCAL= CHELION ESQUERDO



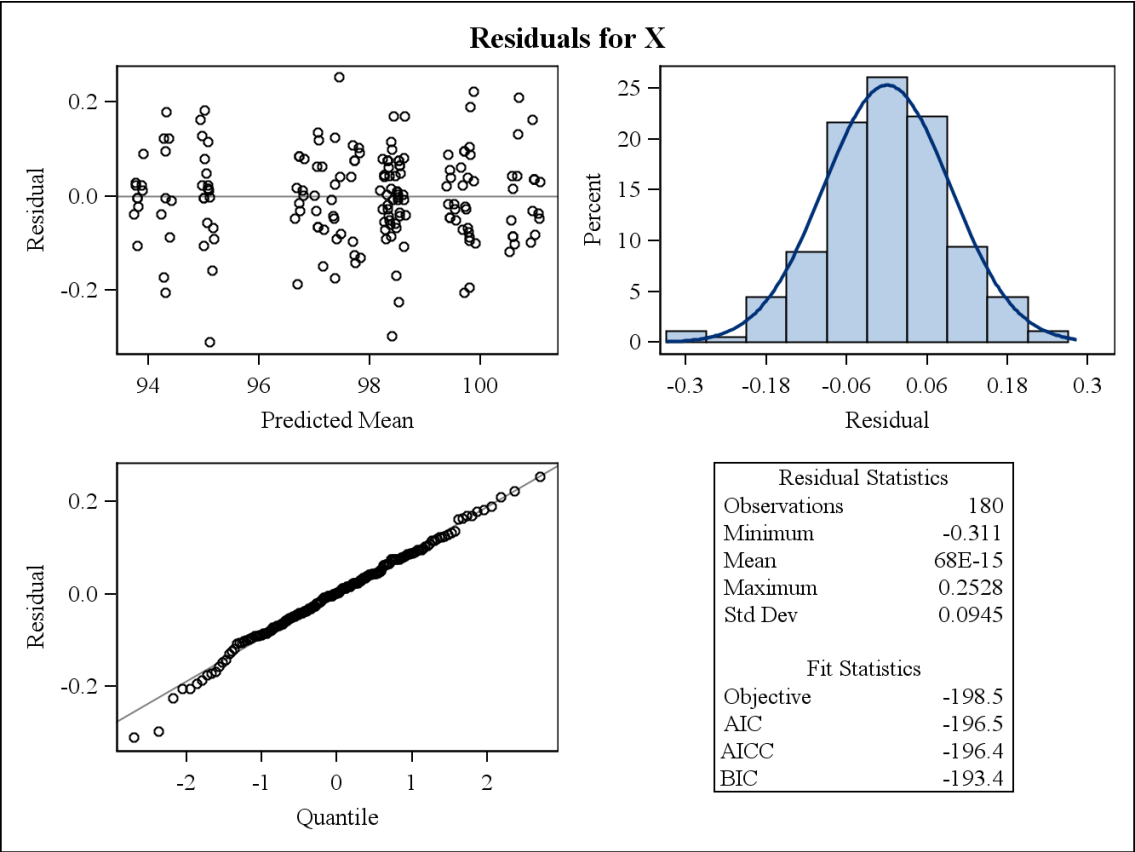
LOCAL=ECTOANTHION DIREITO



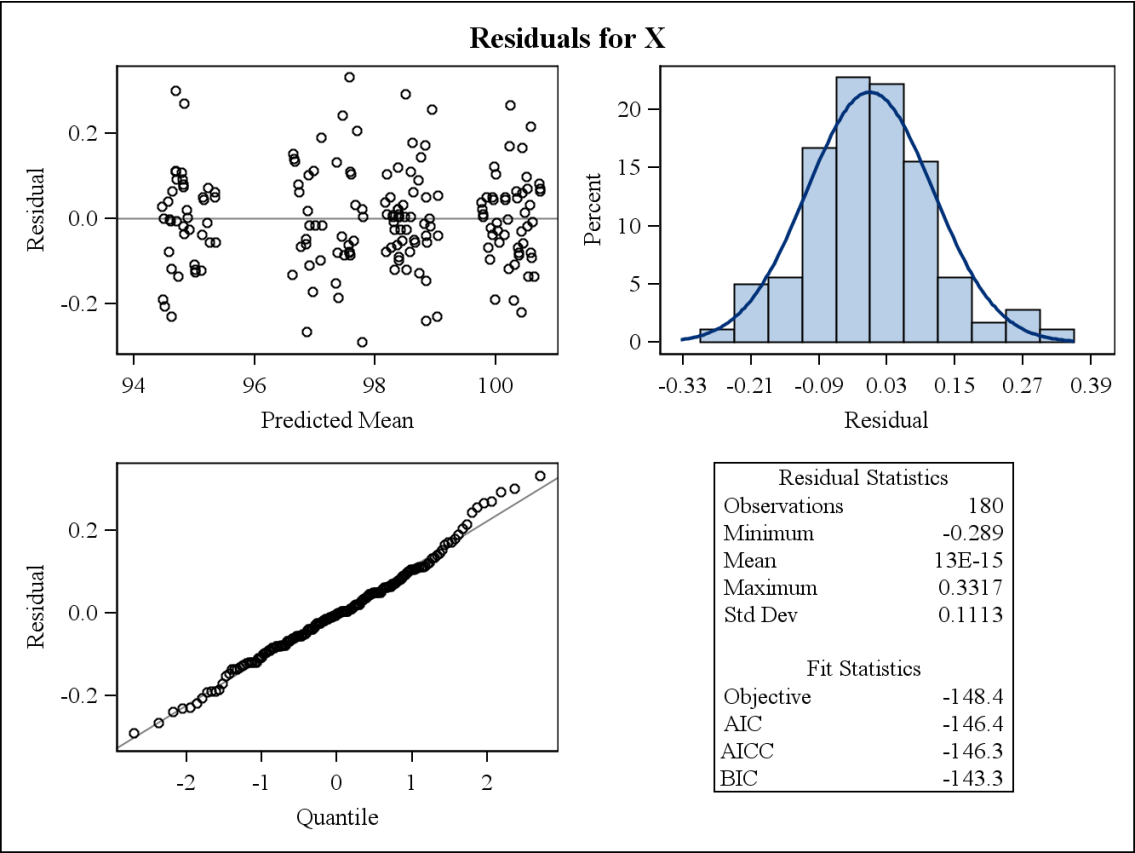
LOCAL=ECTOANTHION ESQUERDO



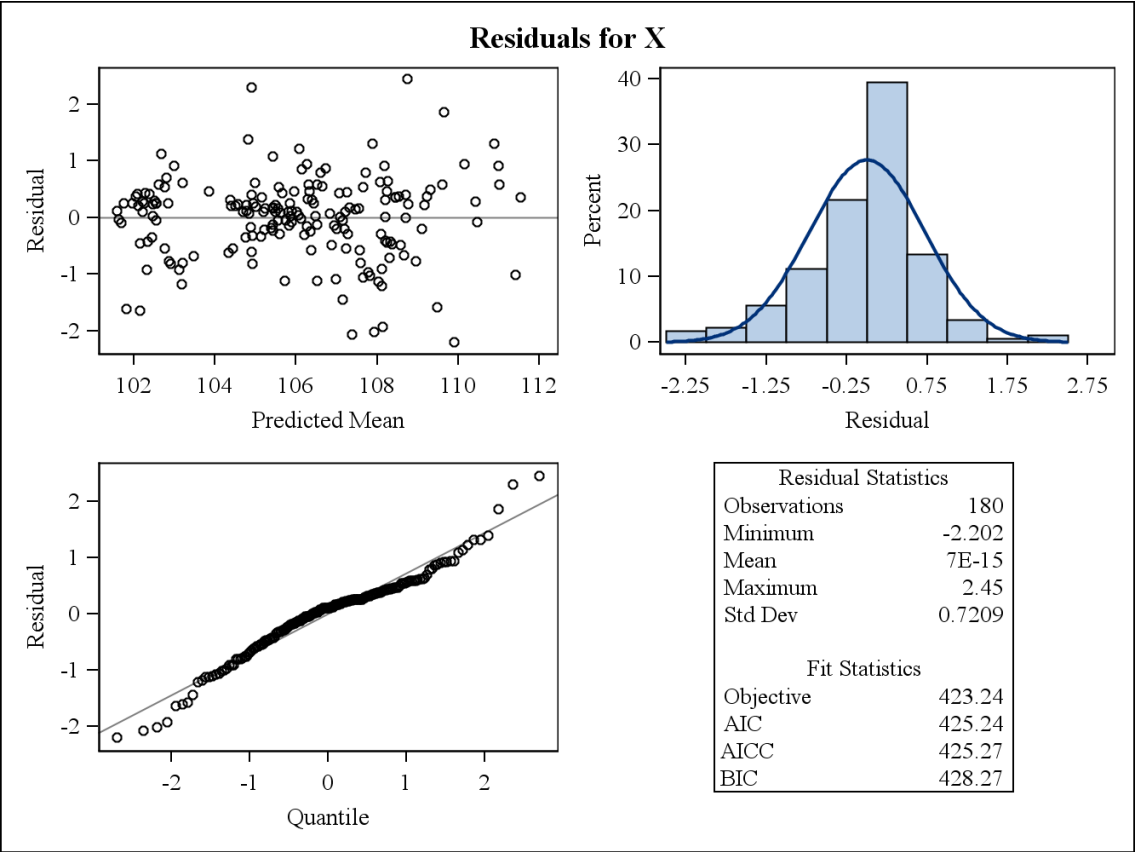
LOCAL=ENDOCANTHION DIREITO



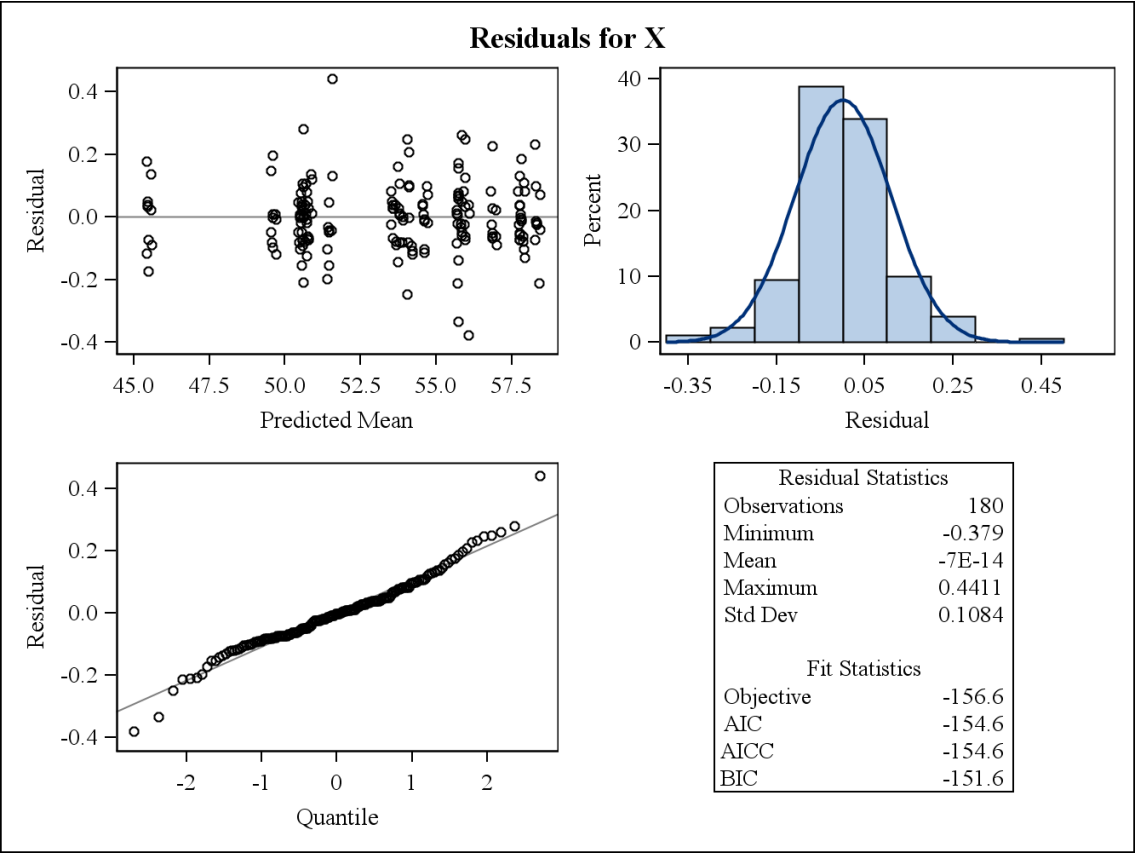
LOCAL= ENDOCANTHION ESQUERDO



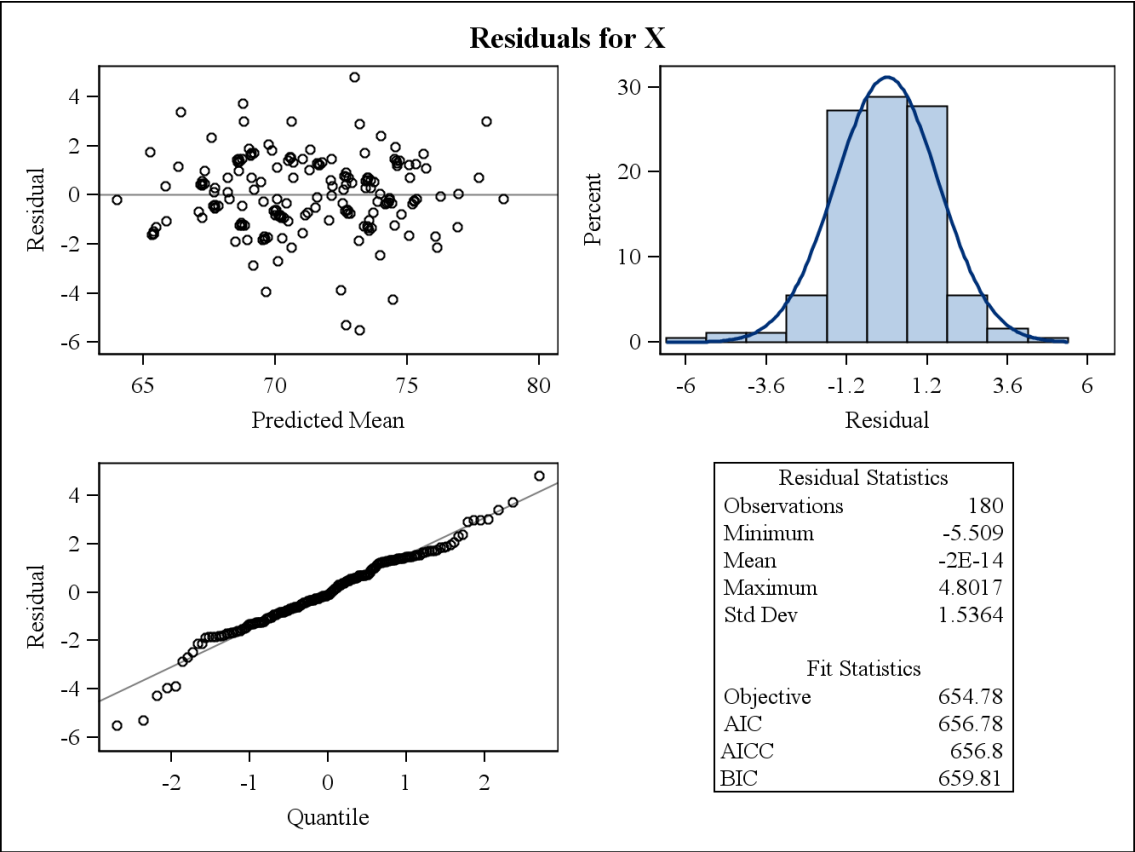
LOCAL=GLABELA



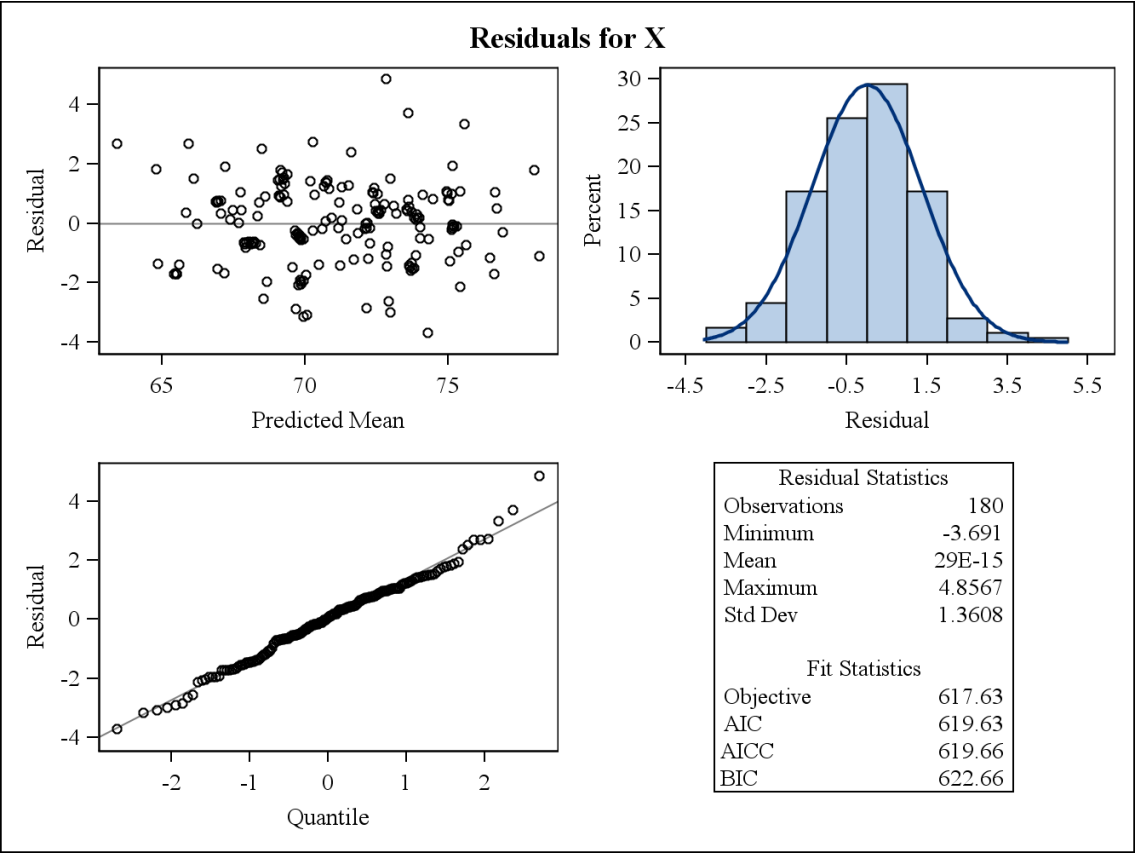
LOCAL=GNATHION



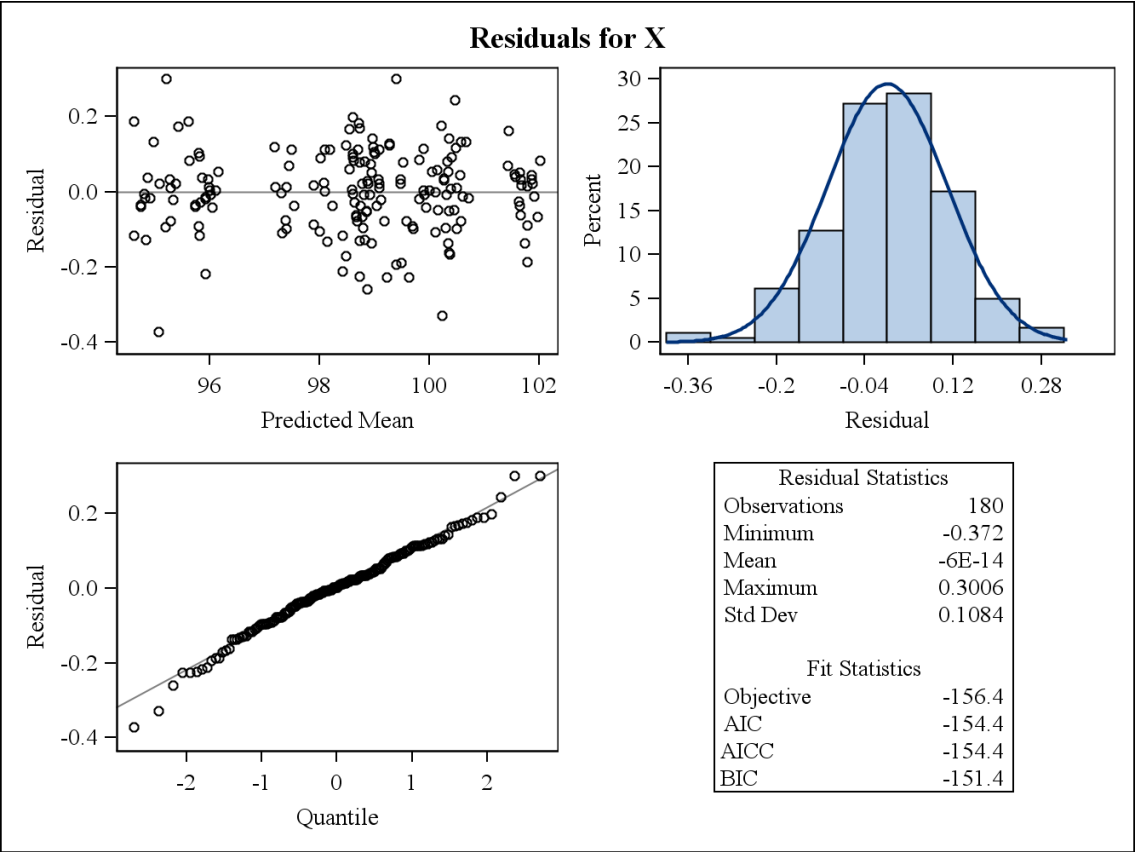
LOCAL=GONIO DIREITO



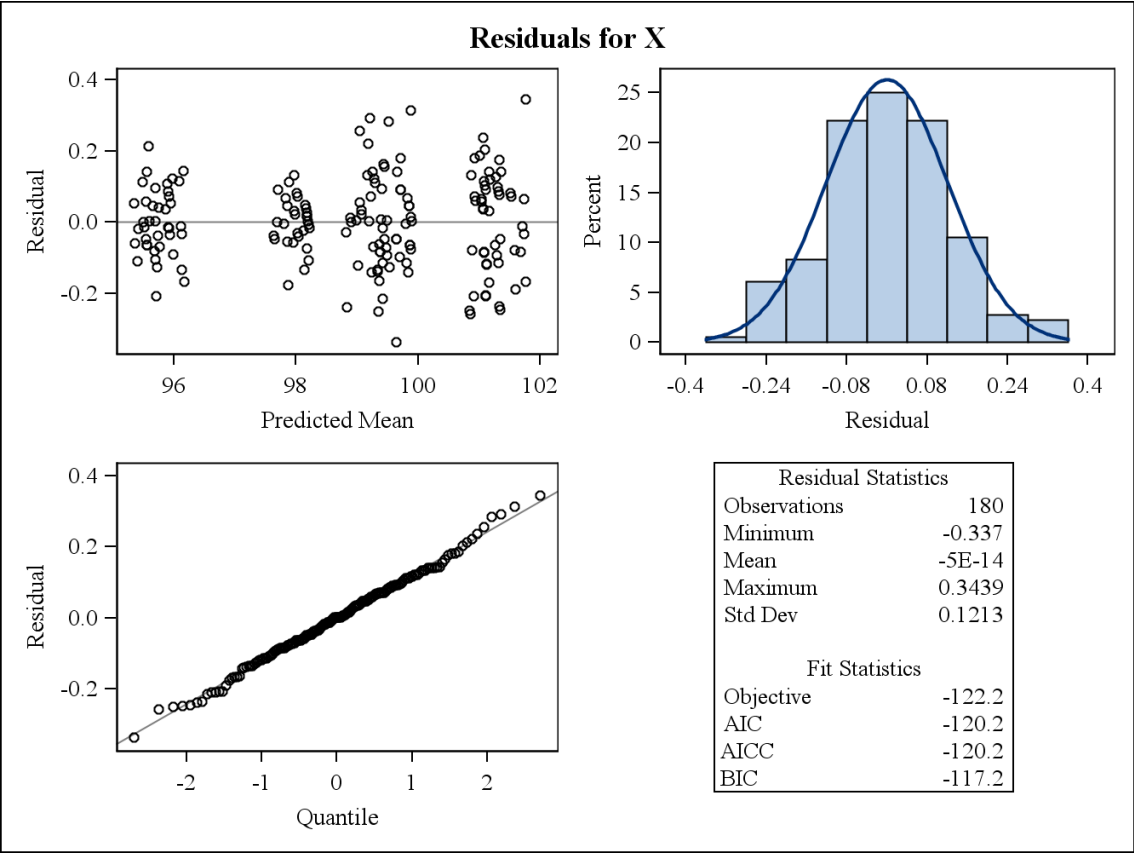
LOCAL=GONIO ESQUERDO



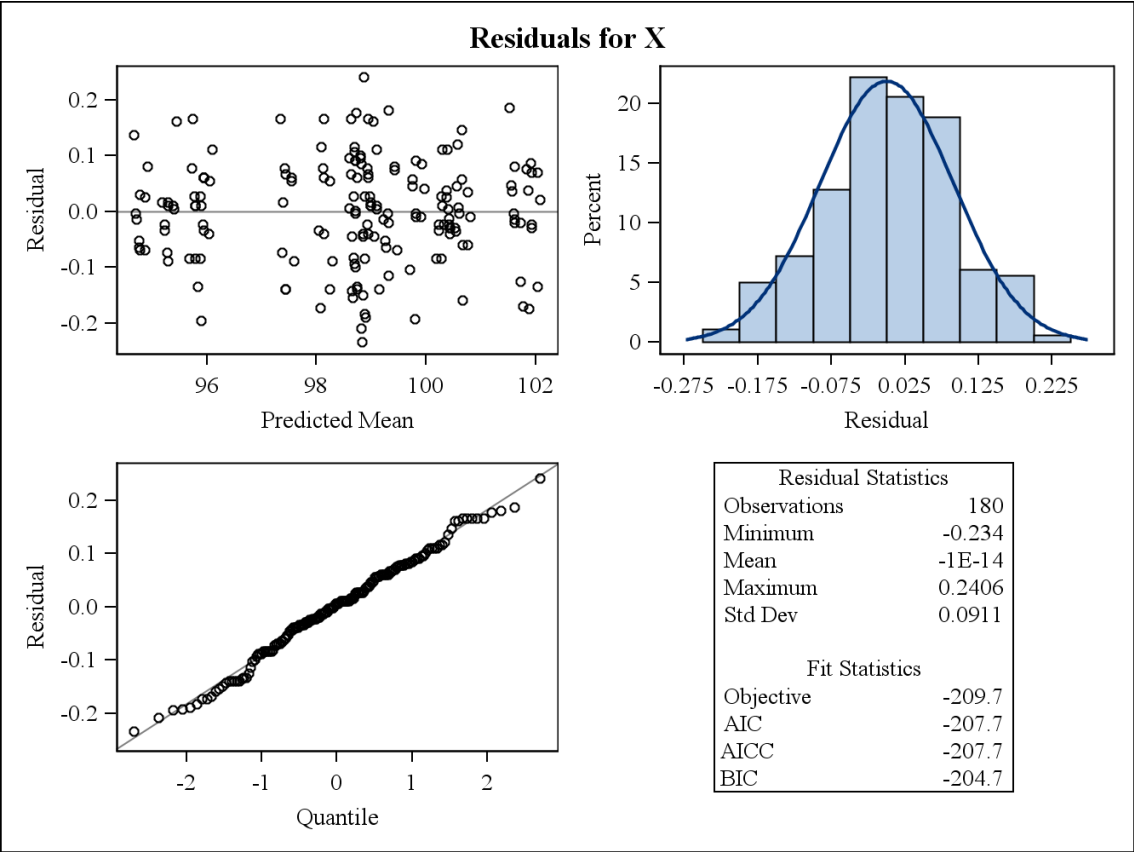
LOCAL=IRIDION LATERAL DIREITO



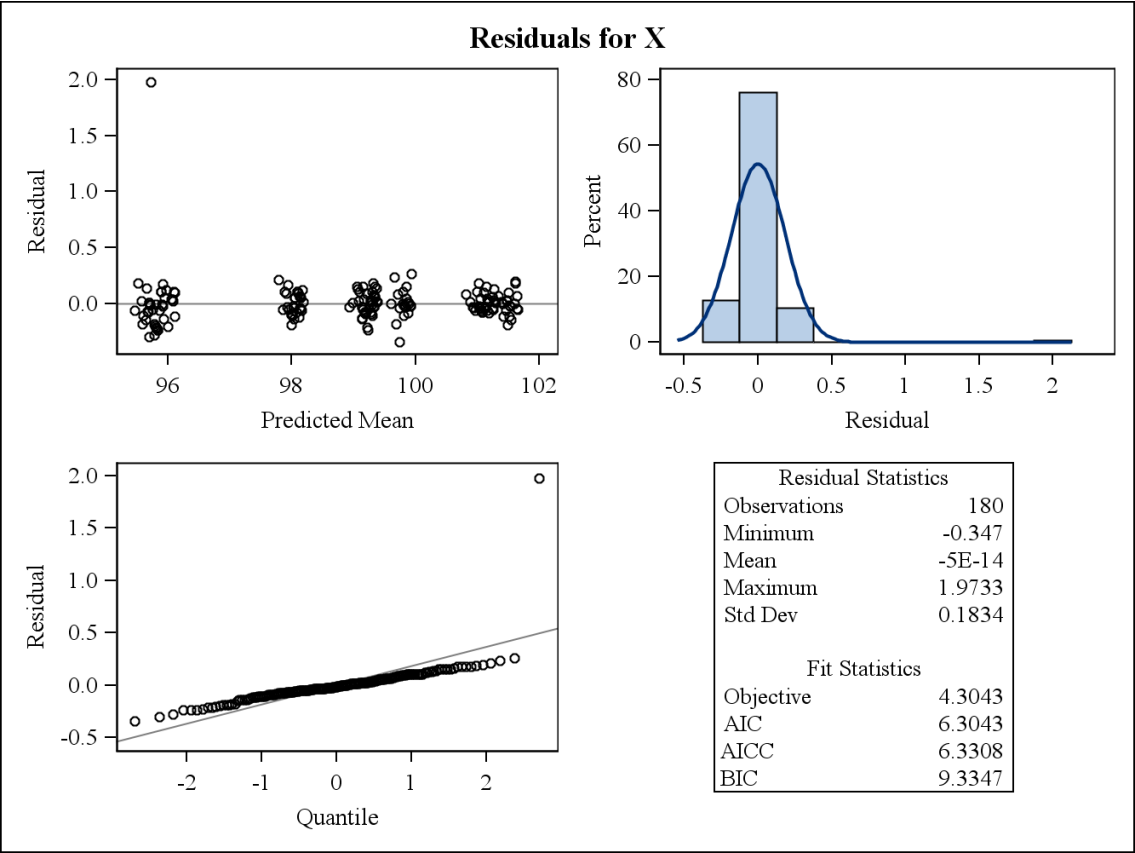
LOCAL=IRIDION LATERAL ESQUERDO



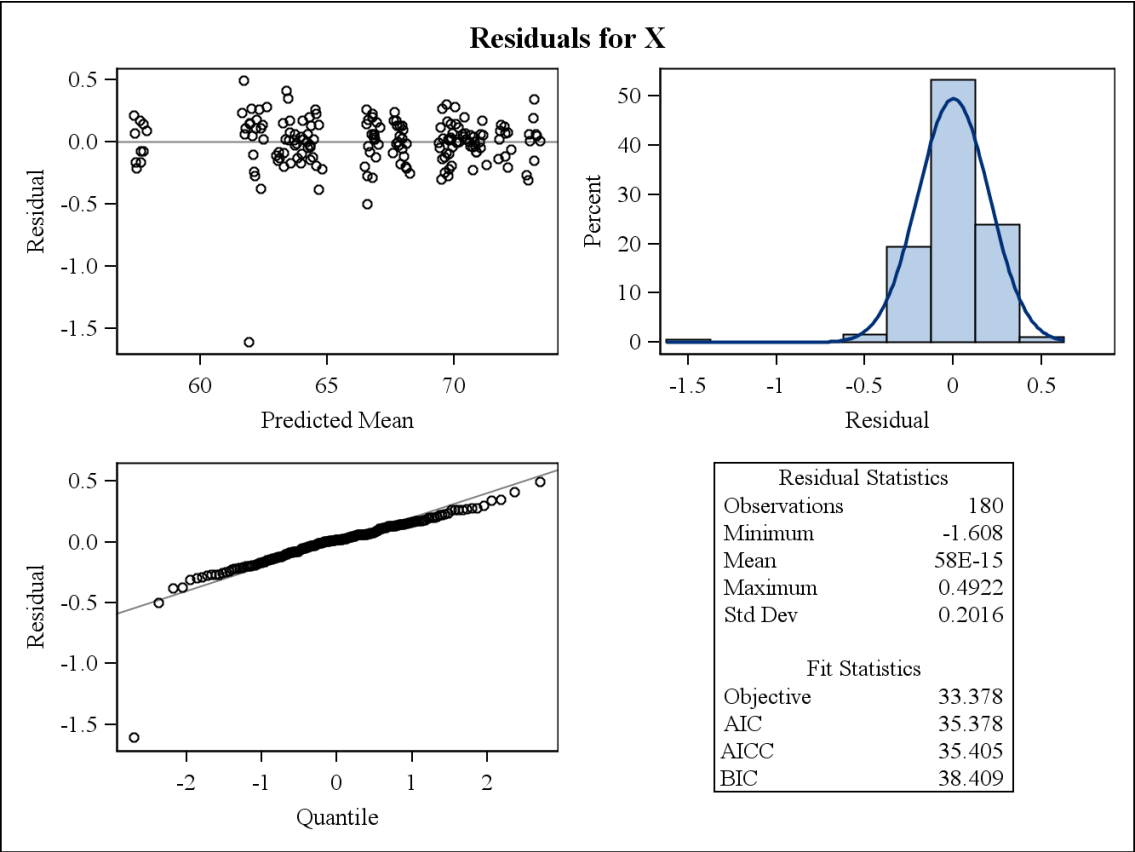
LOCAL=IRIDION MEDIAL DIREITO



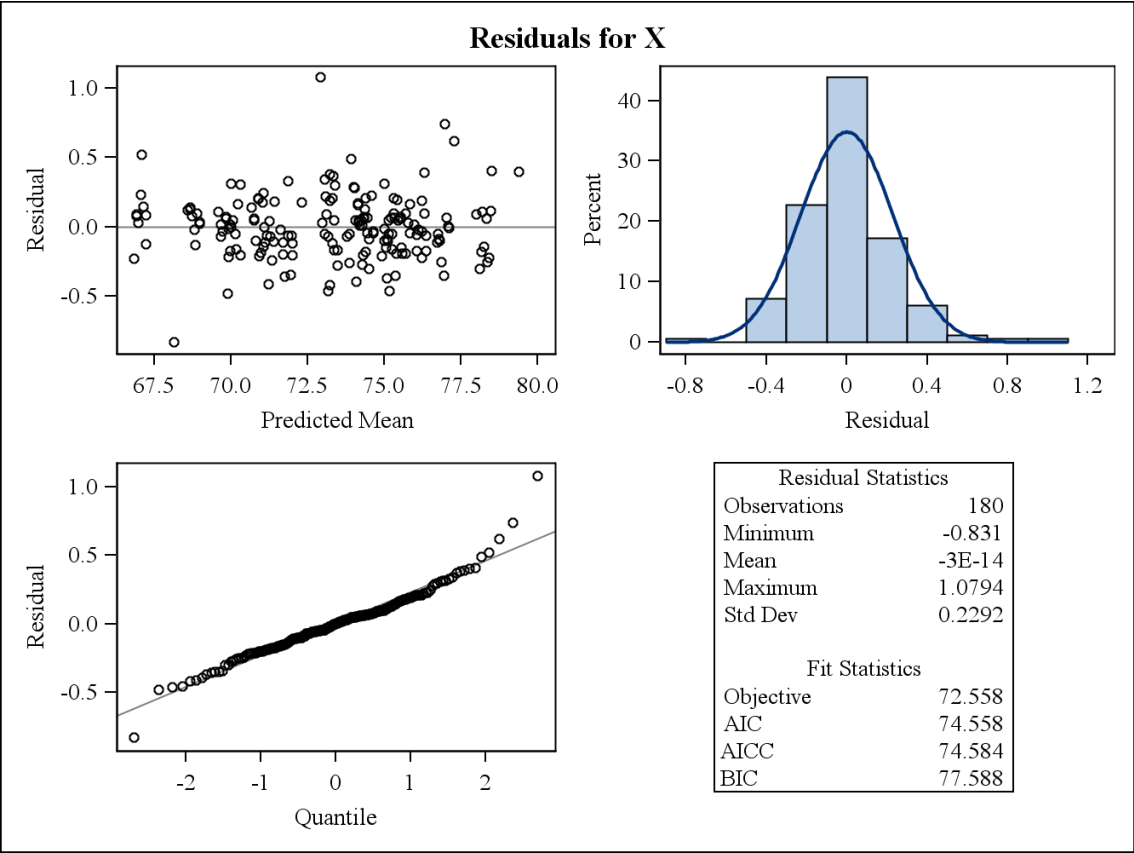
LOCAL=IRIDION MEDIAL ESQUERDO



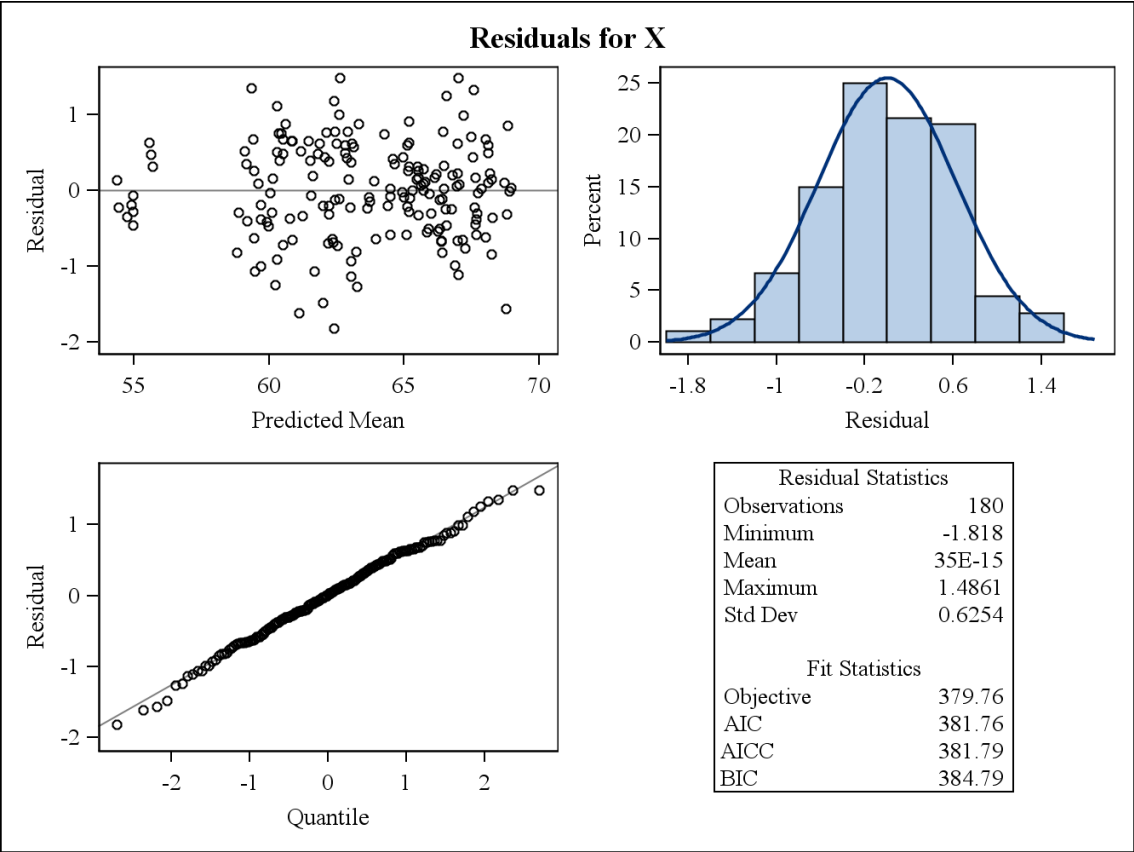
LOCAL=LABIAL INFERIOR



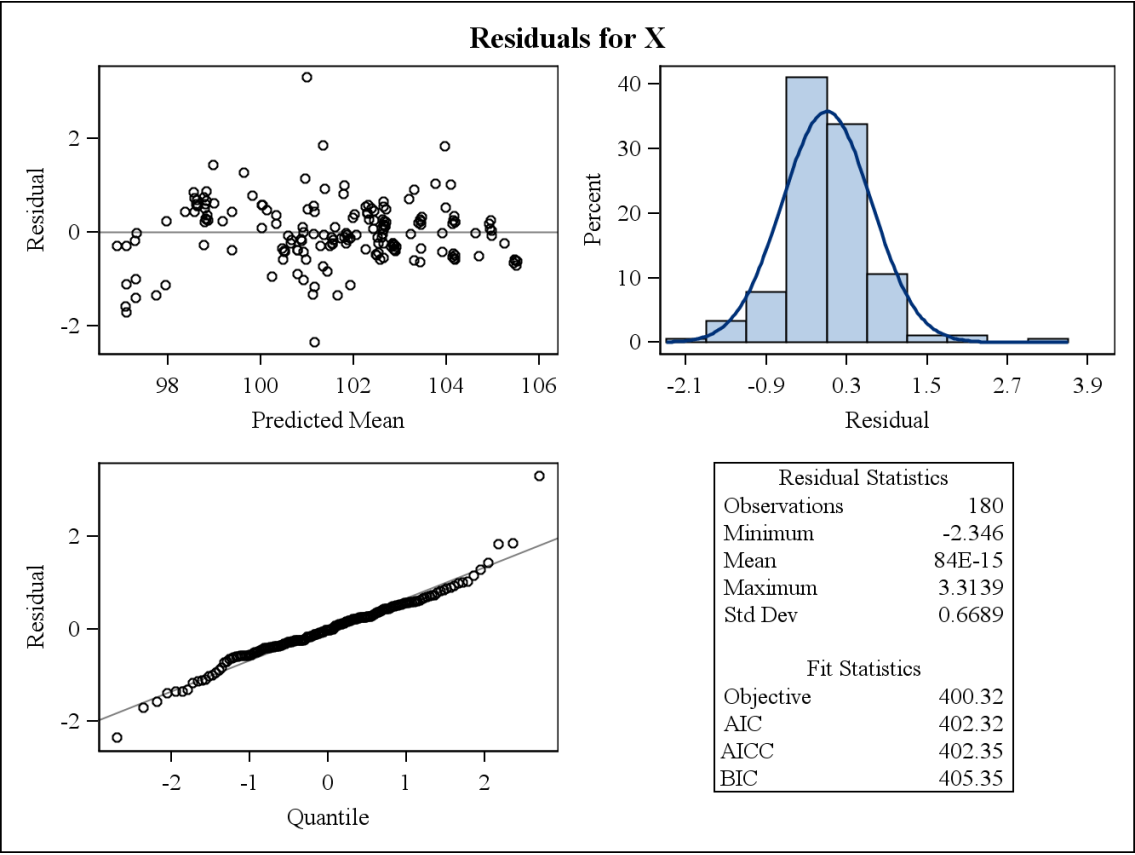
LOCAL=LABIAL SUPERIOR



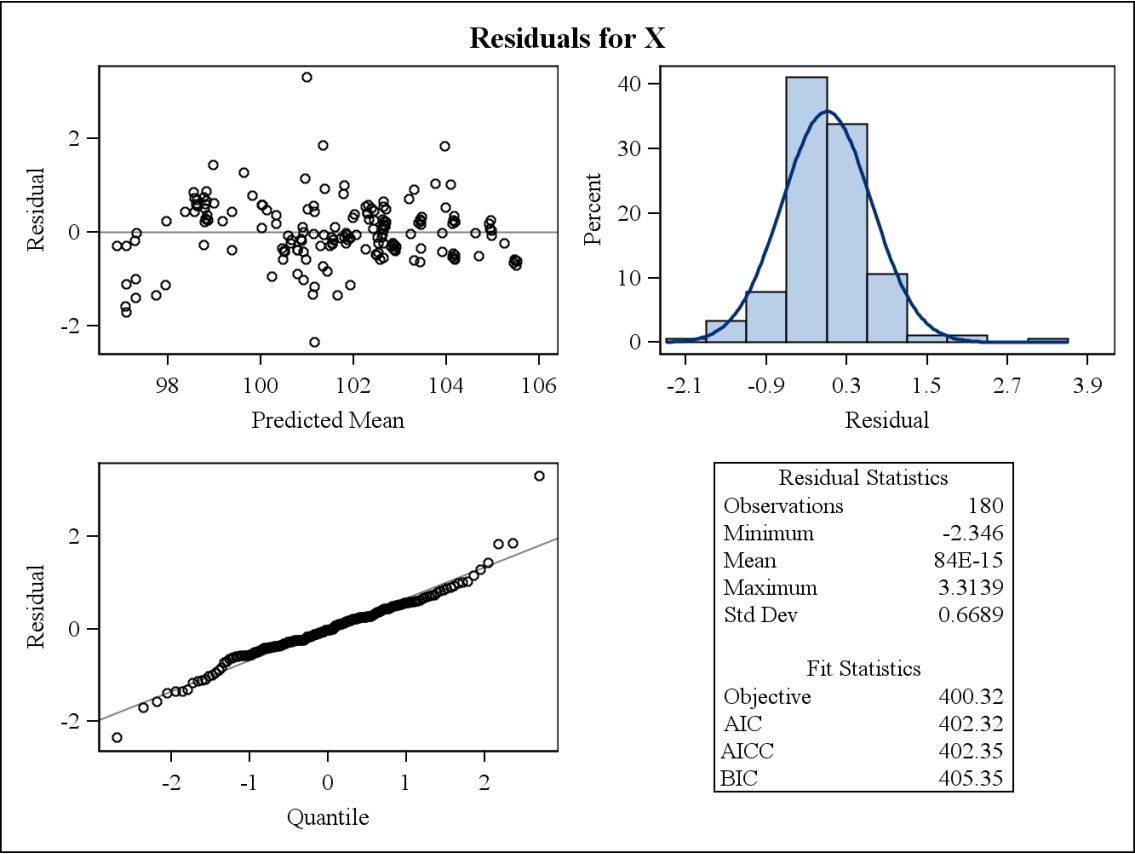
LOCAL=LABIOMENTAL



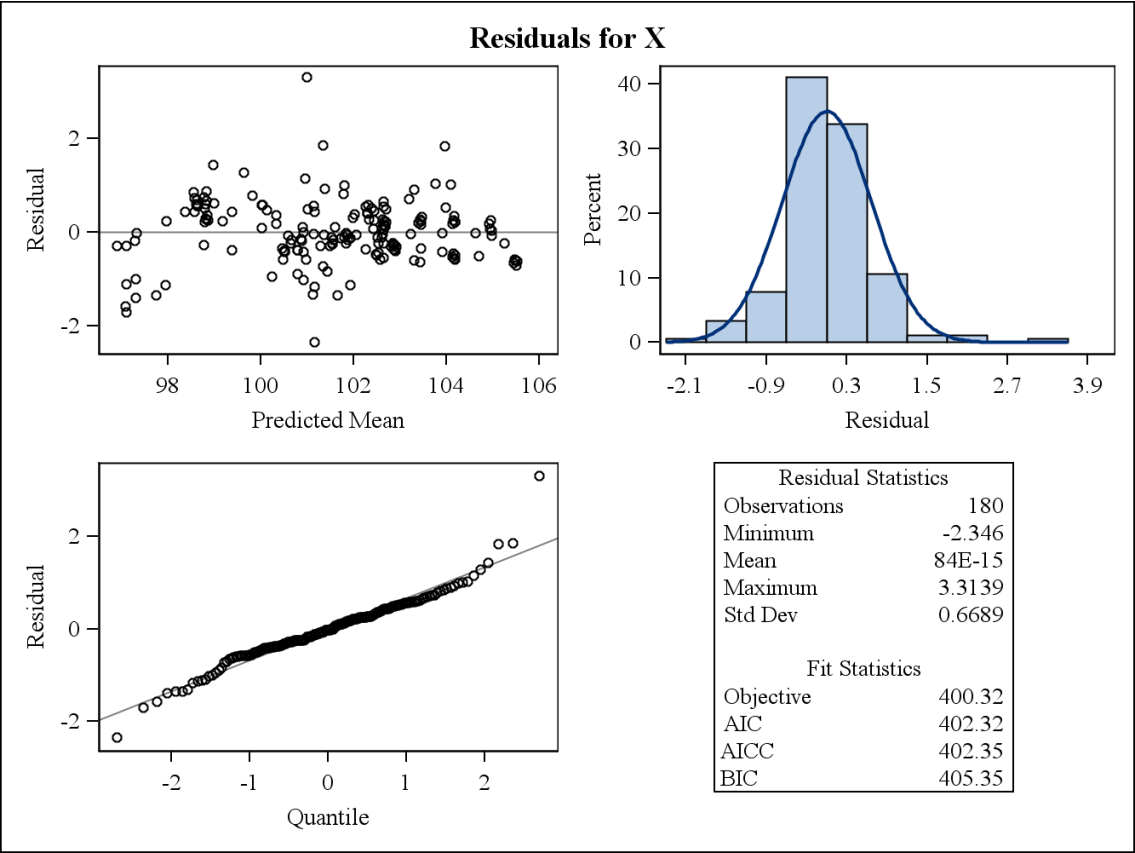
LOCAL=NASION



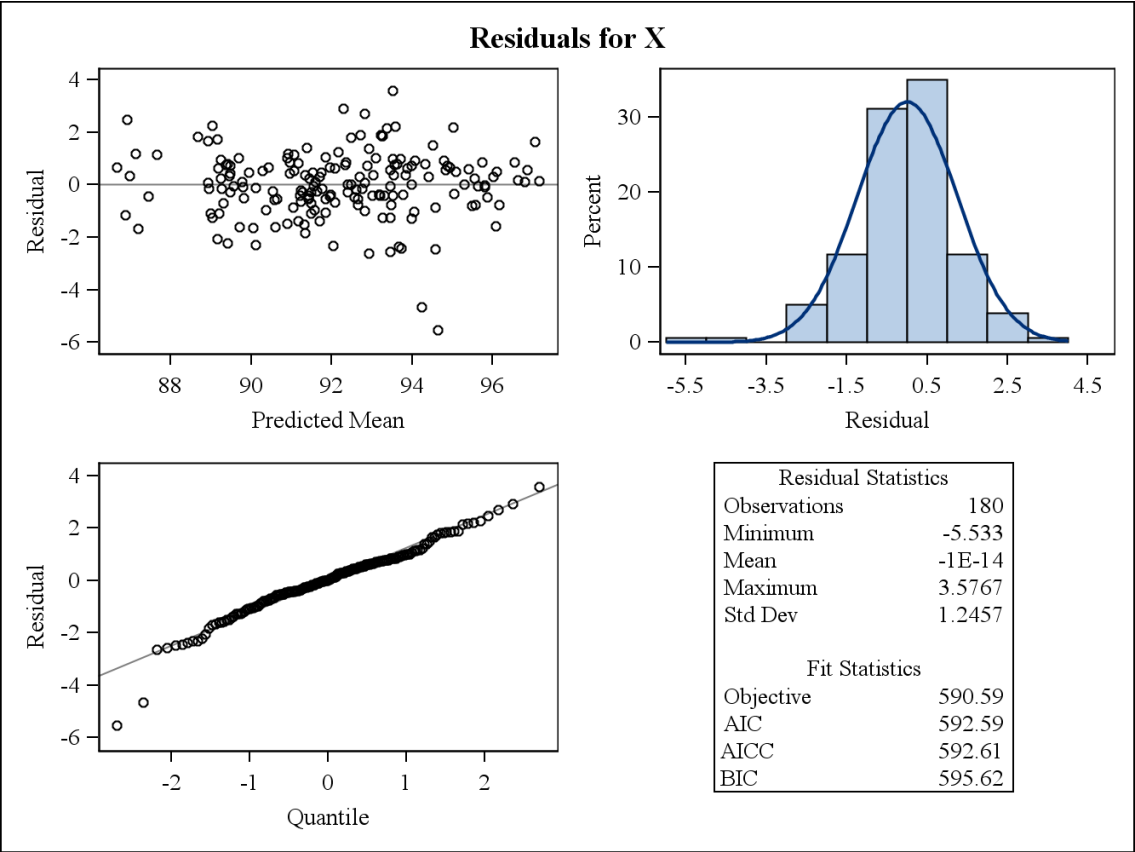
LOCAL=STOMION



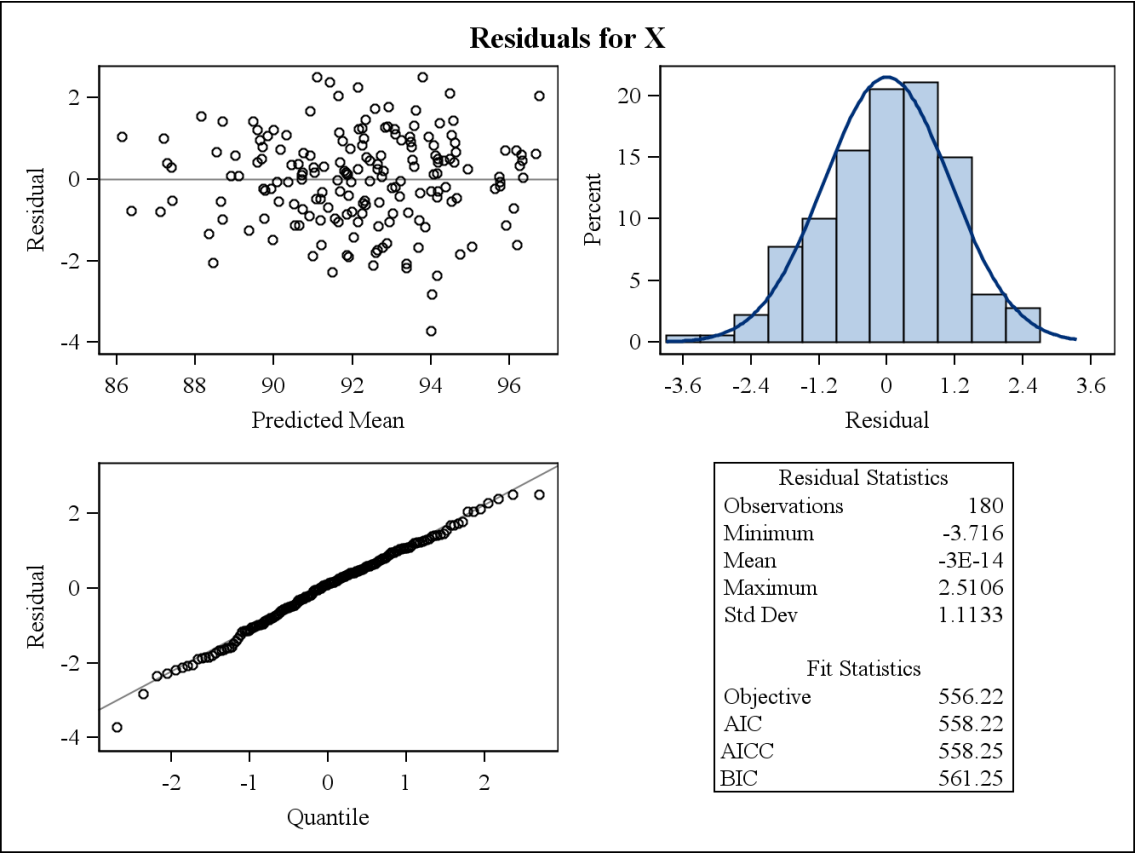
LOCAL=SUBNASAL



LOCAL=ZYGION DIREITO

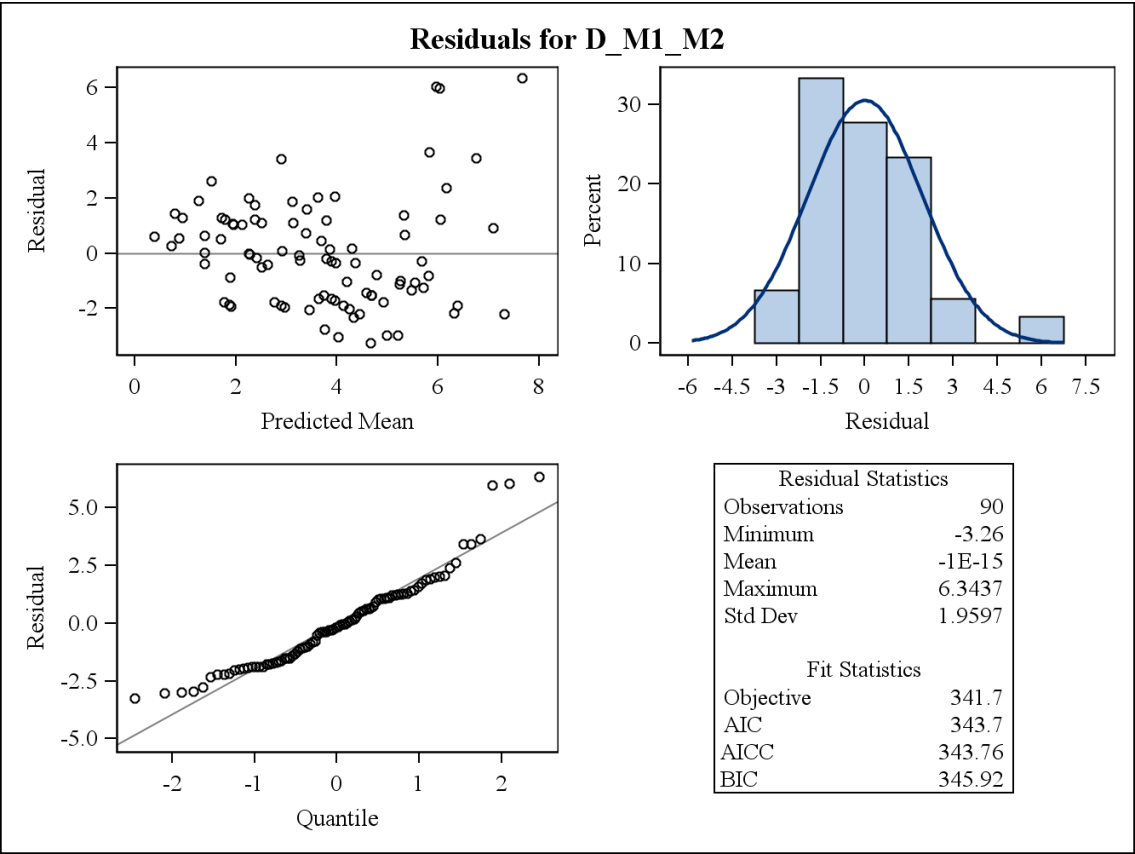


LOCAL=ZYGION ESQUERDO

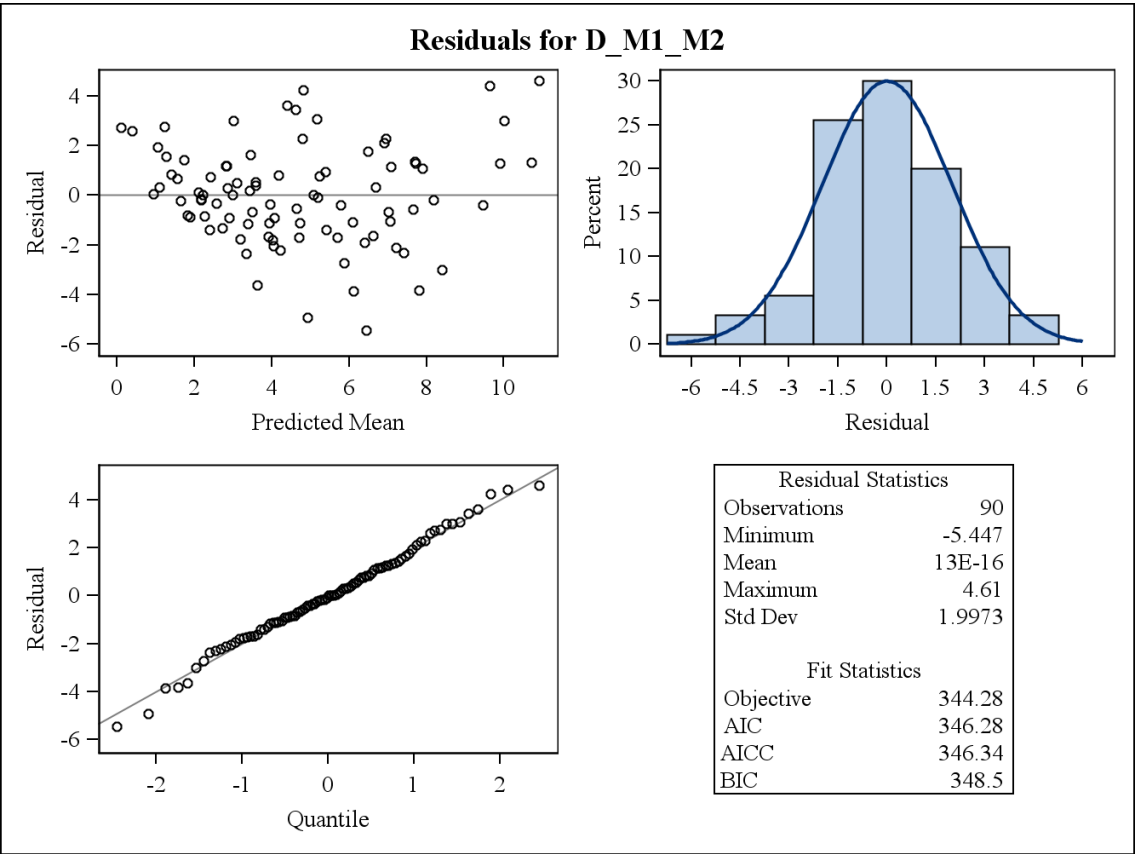


VERIFICAR NORMALIDADE PARA DISTANCIA

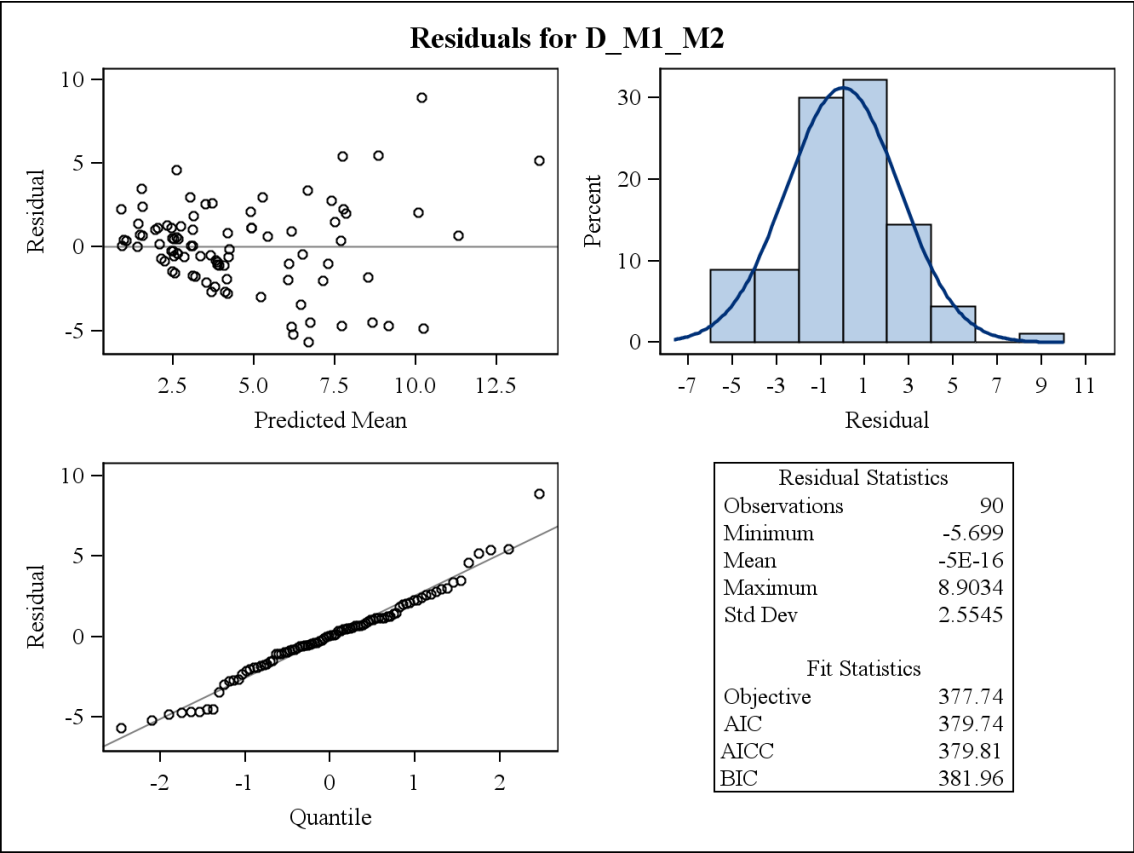
LOCAL=1



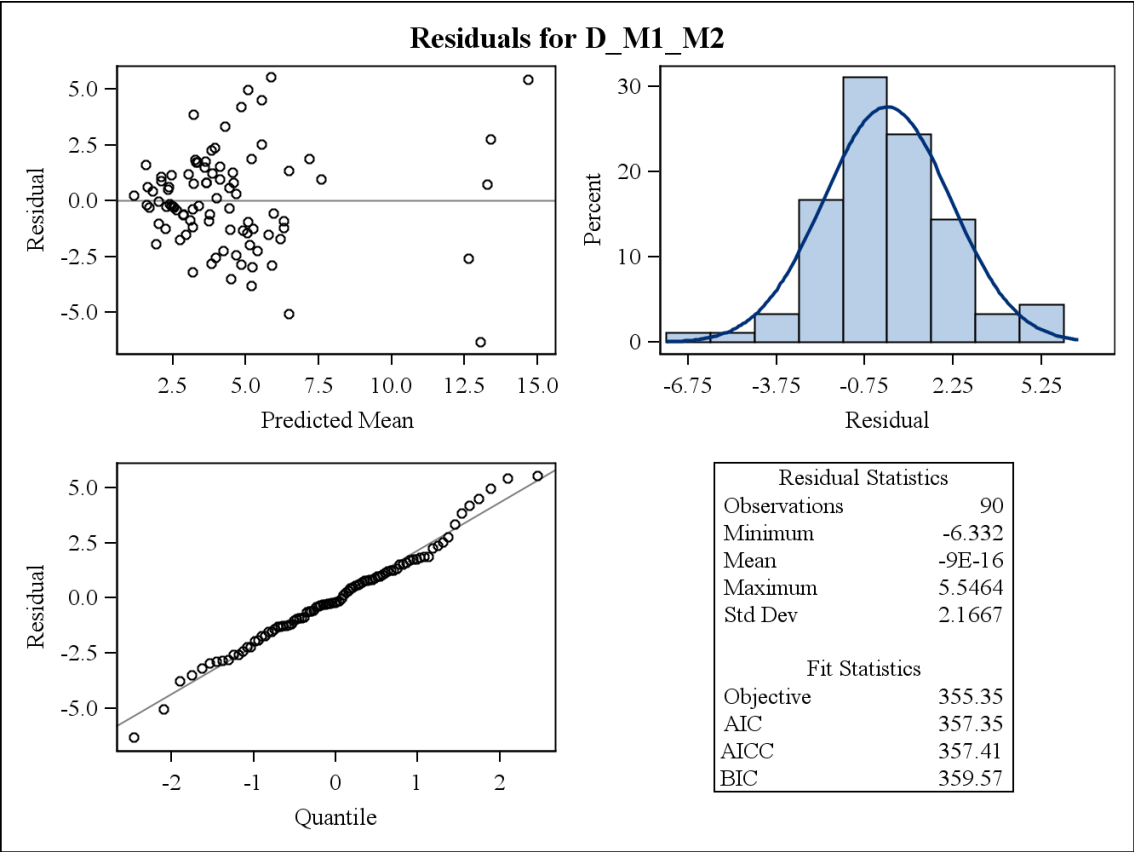
LOCAL=2



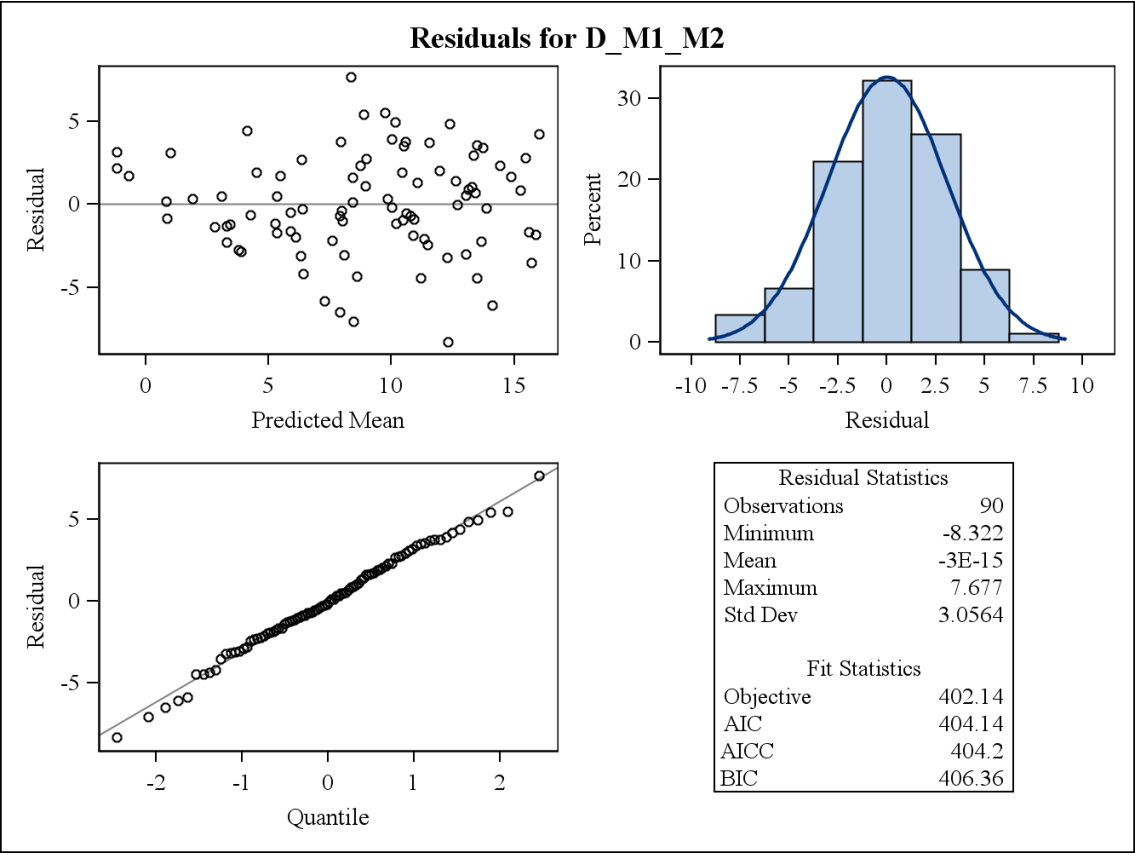
LOCAL=3



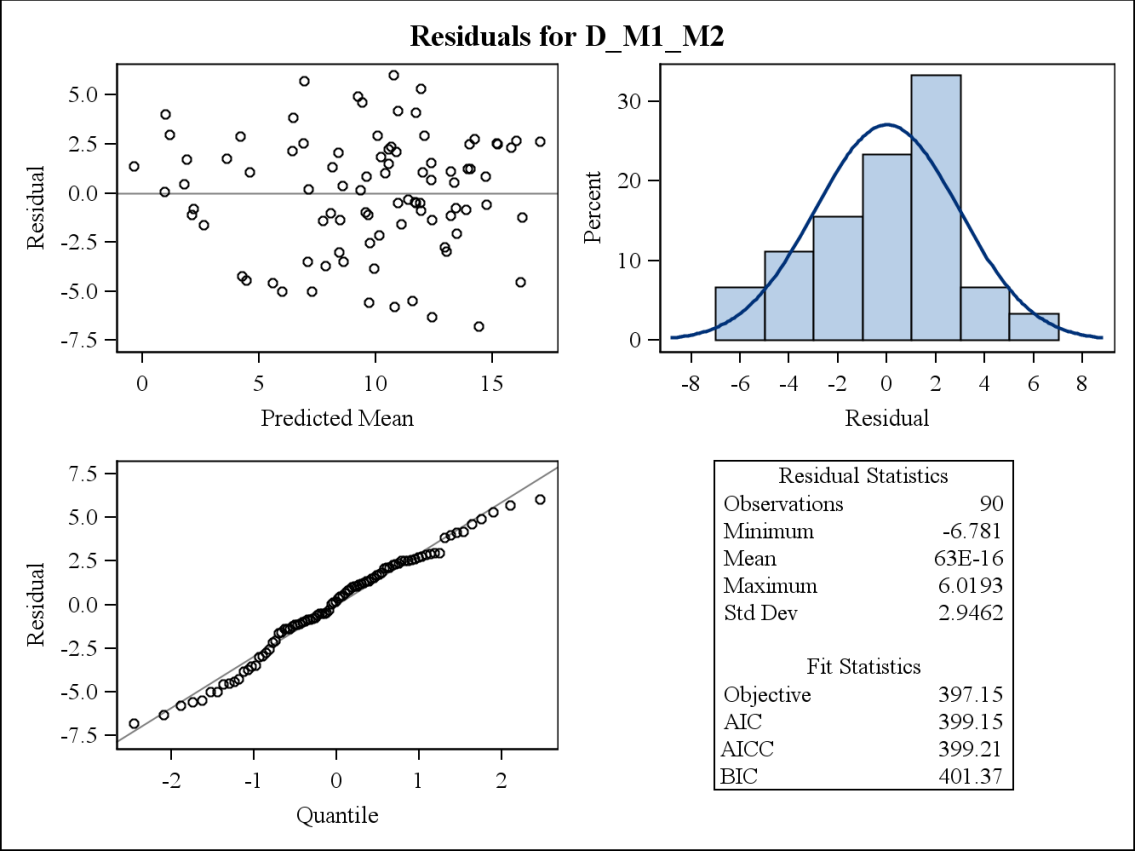
LOCAL=4



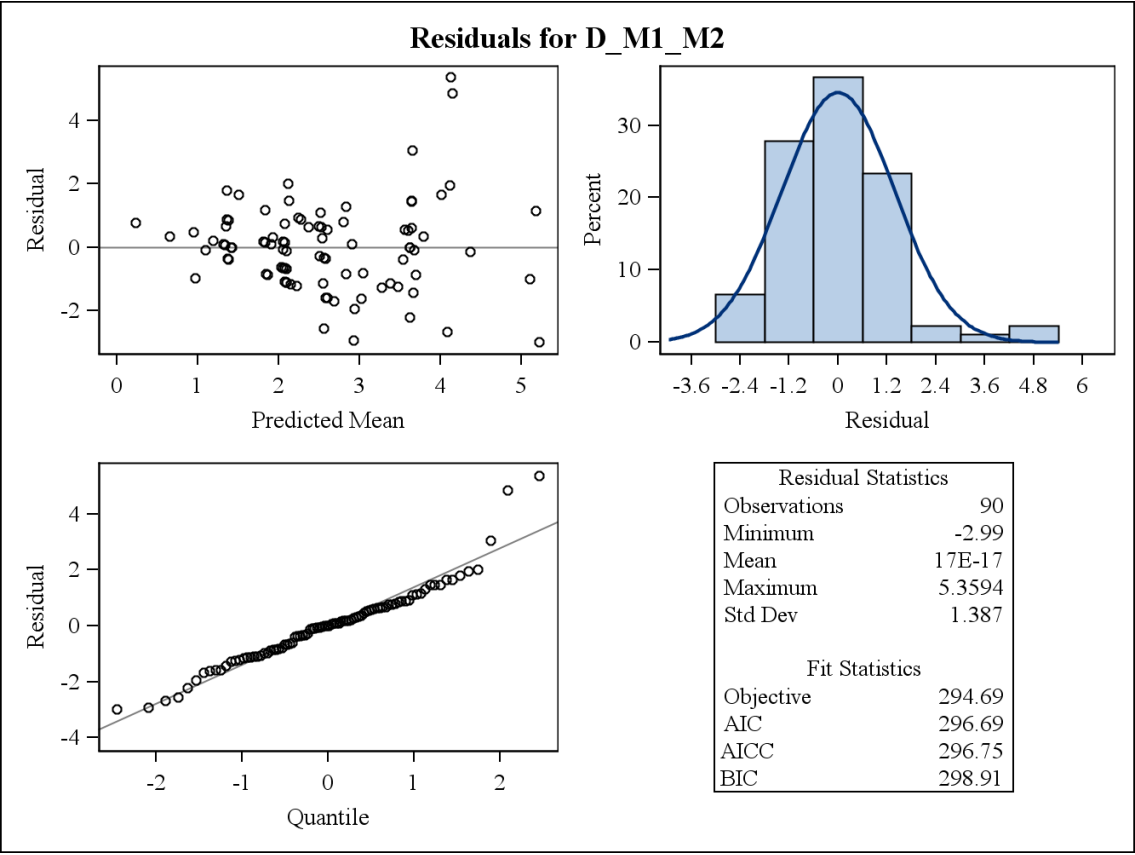
LOCAL=5



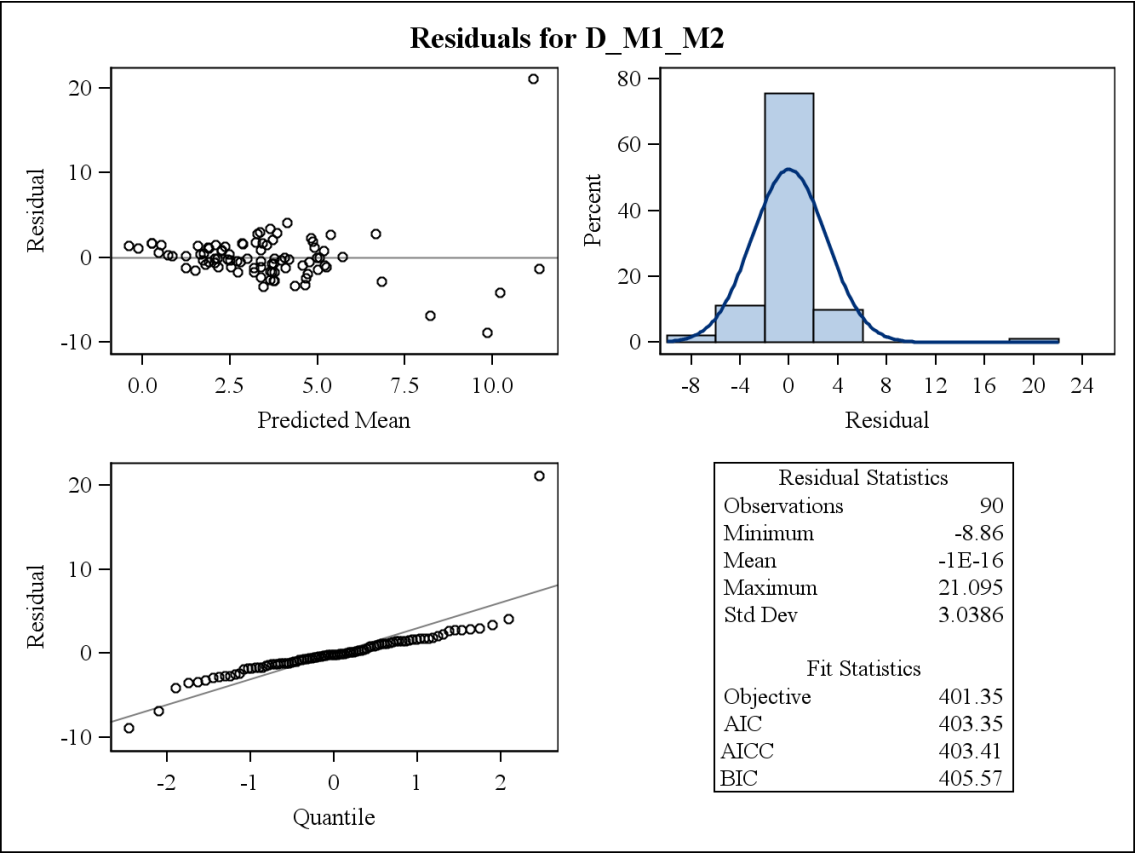
LOCAL=6



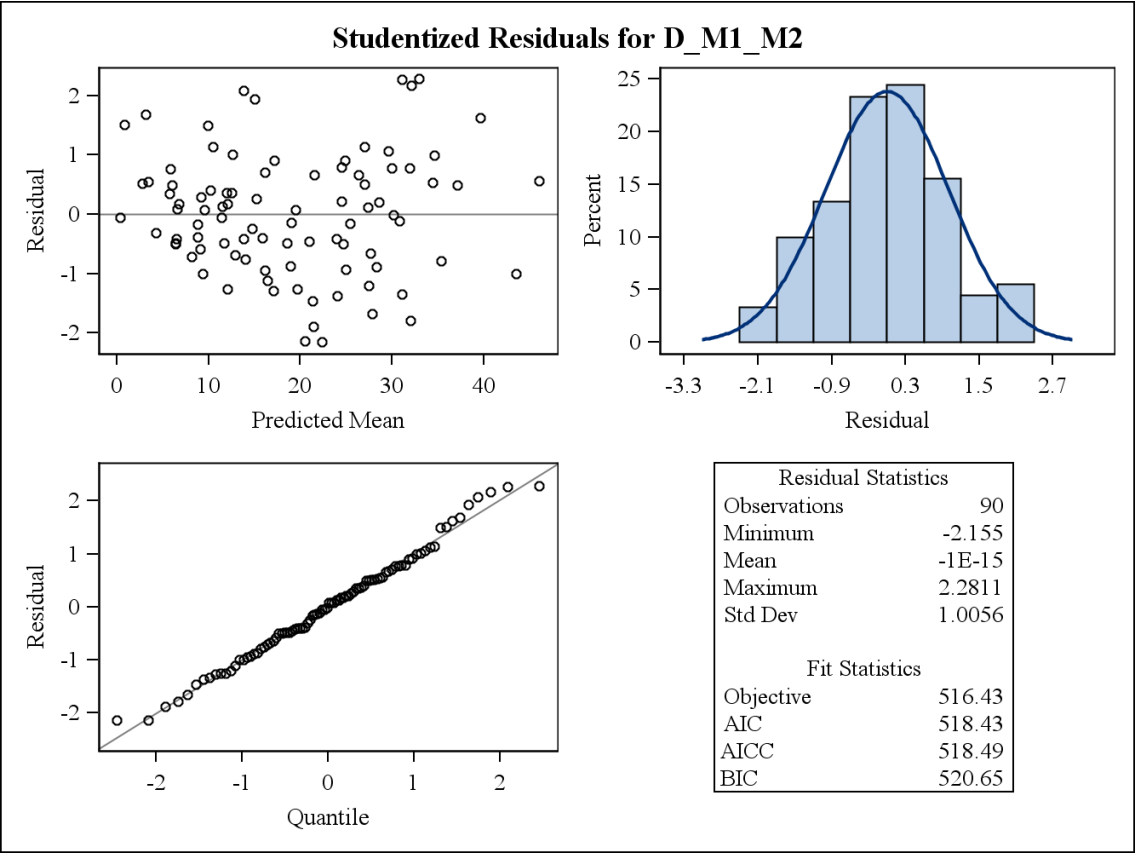
LOCAL=7



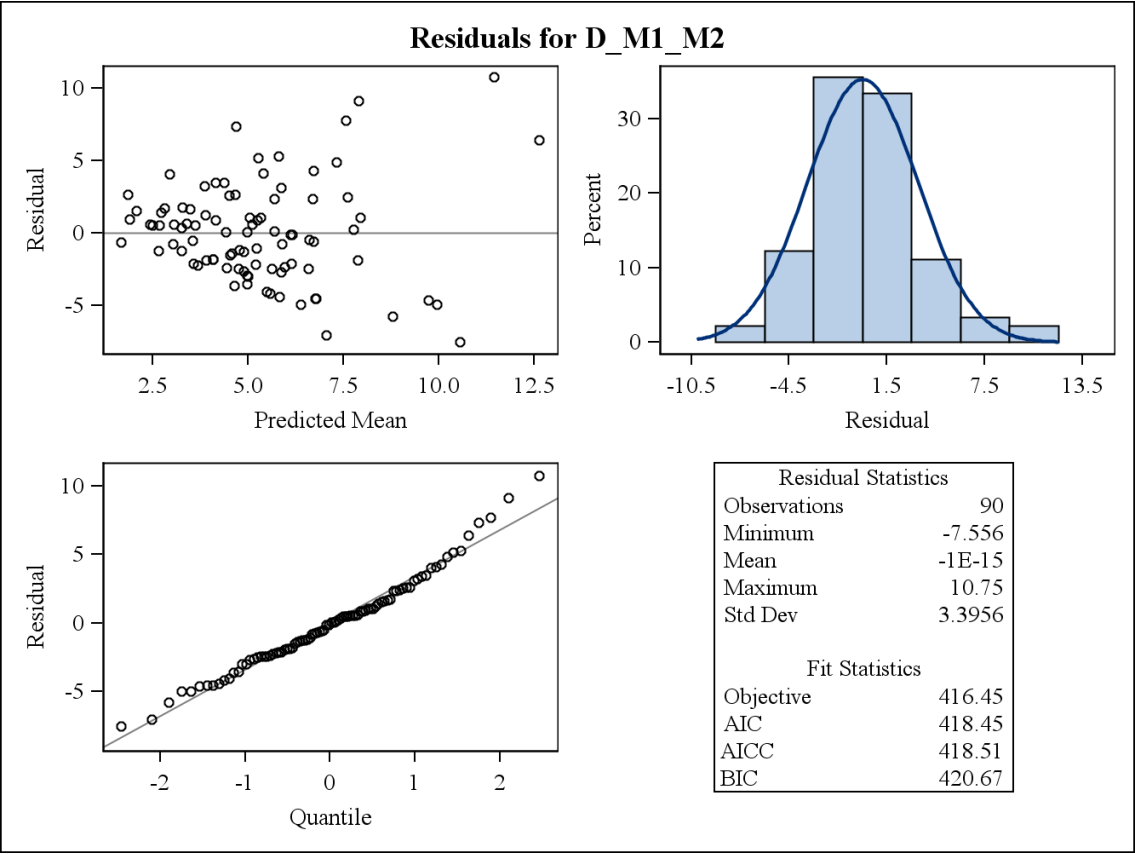
LOCAL=8



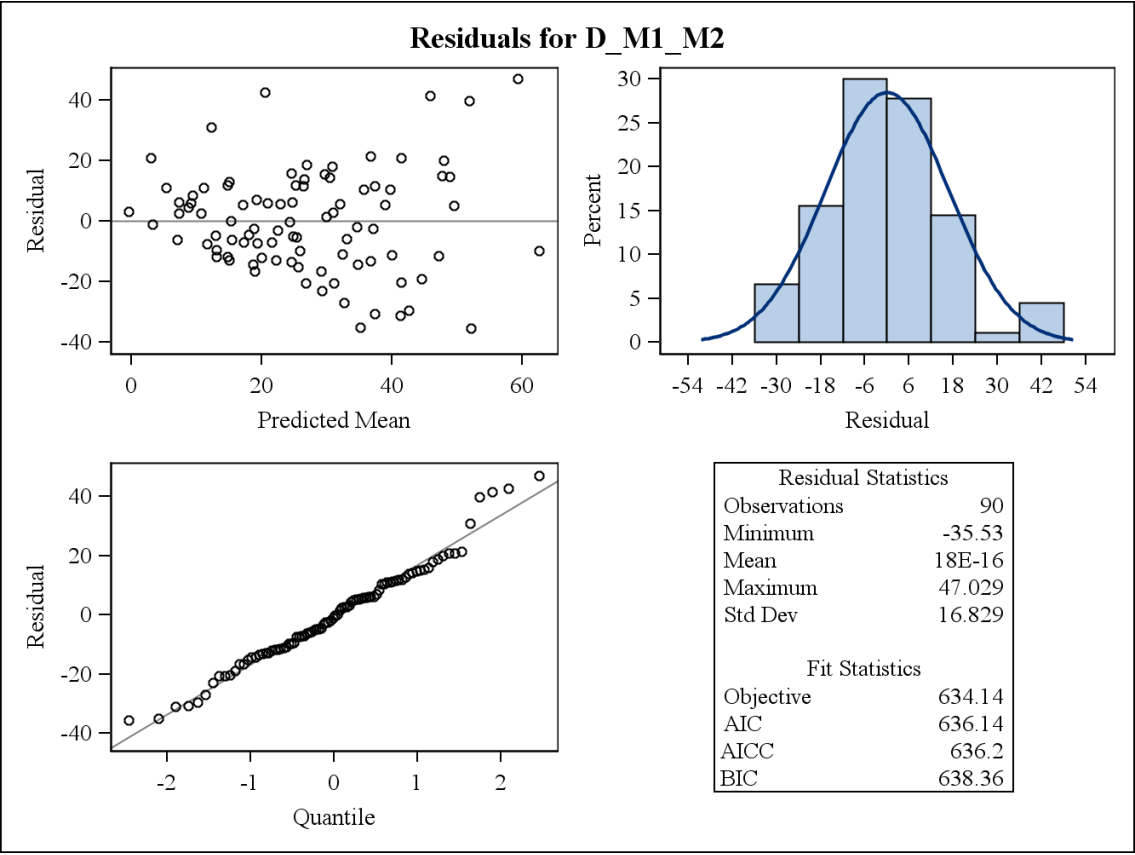
LOCAL=9



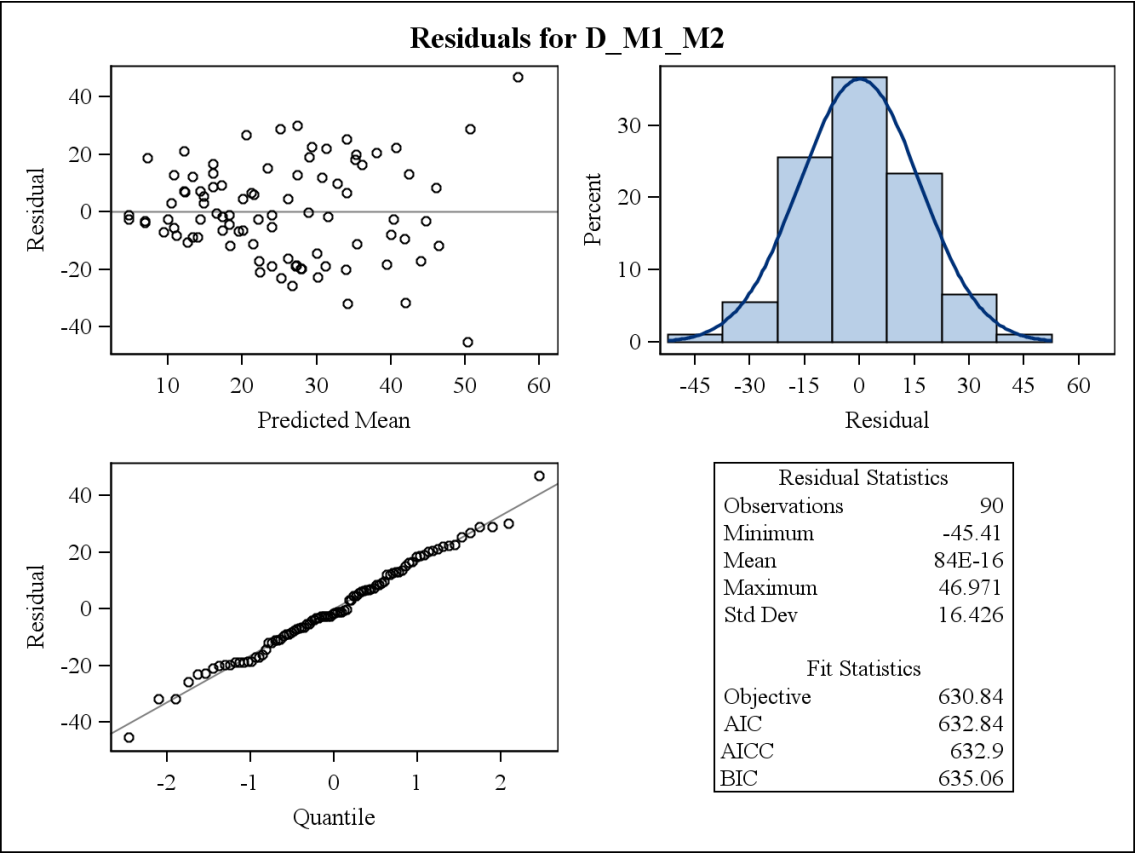
LOCAL=10



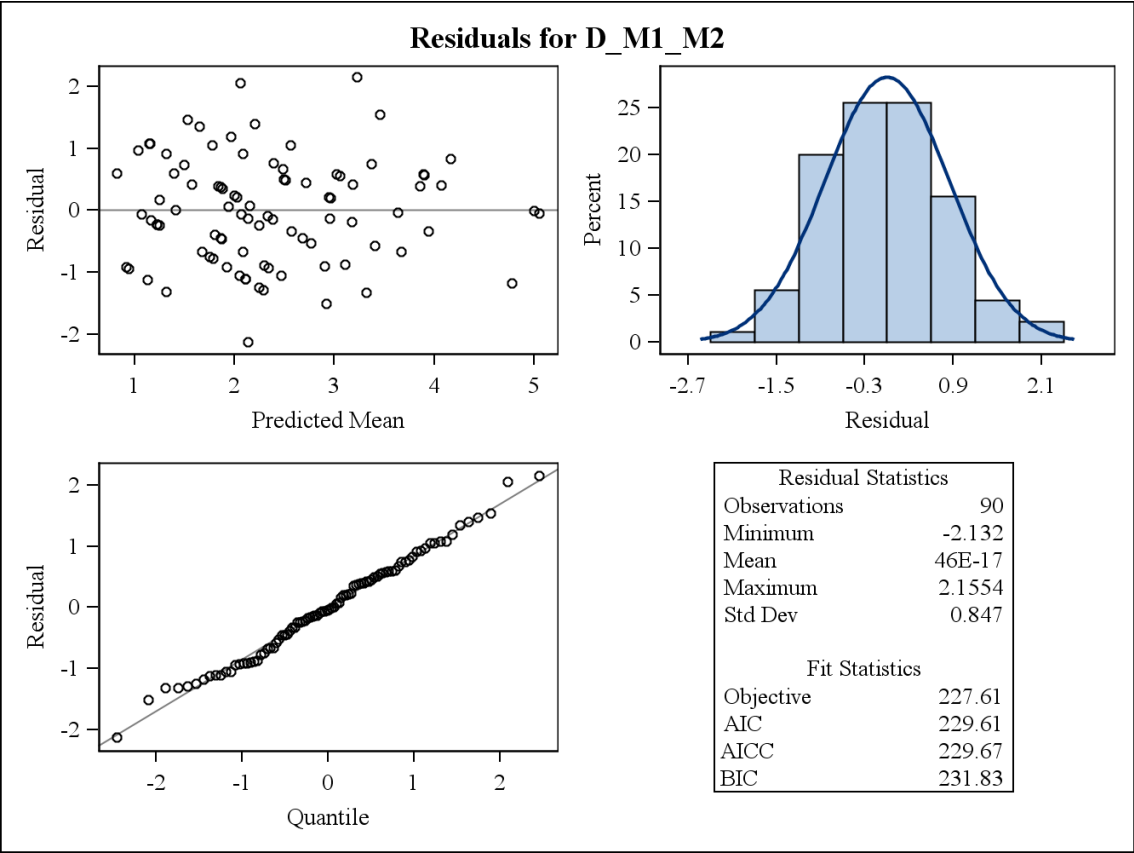
LOCAL=11



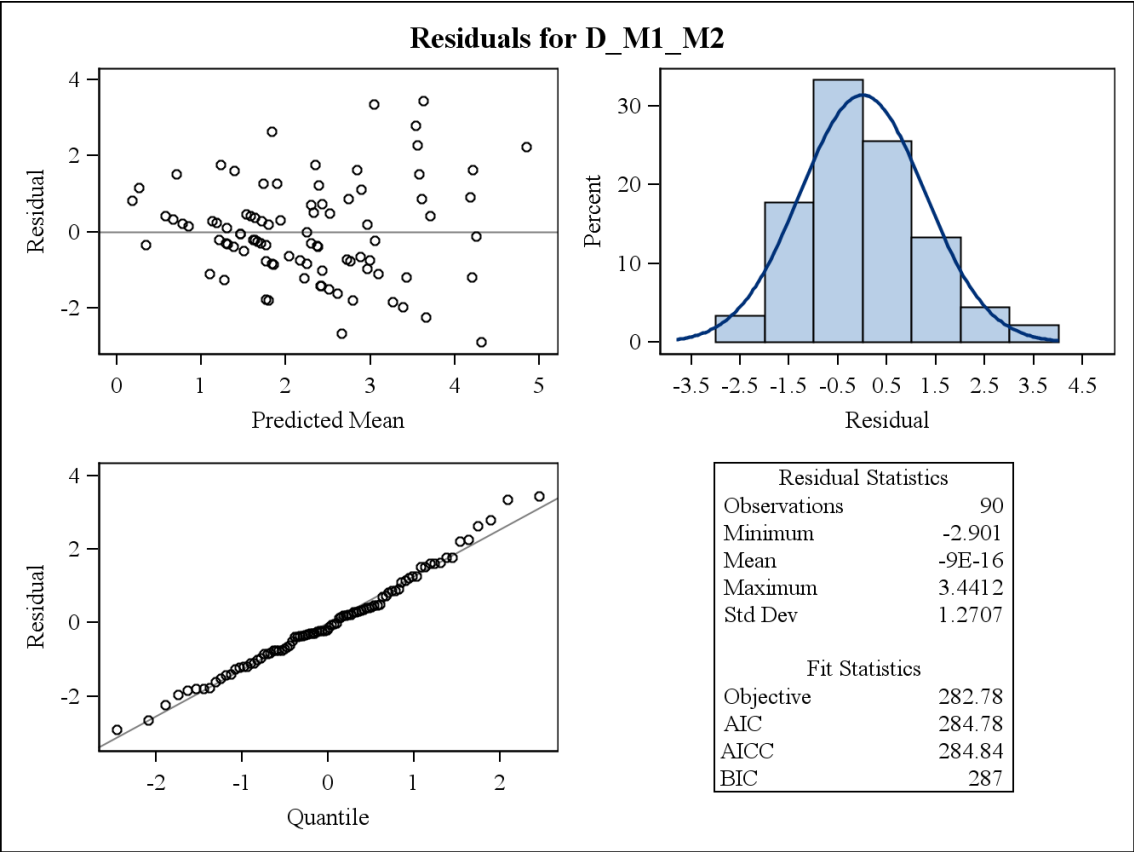
LOCAL=12



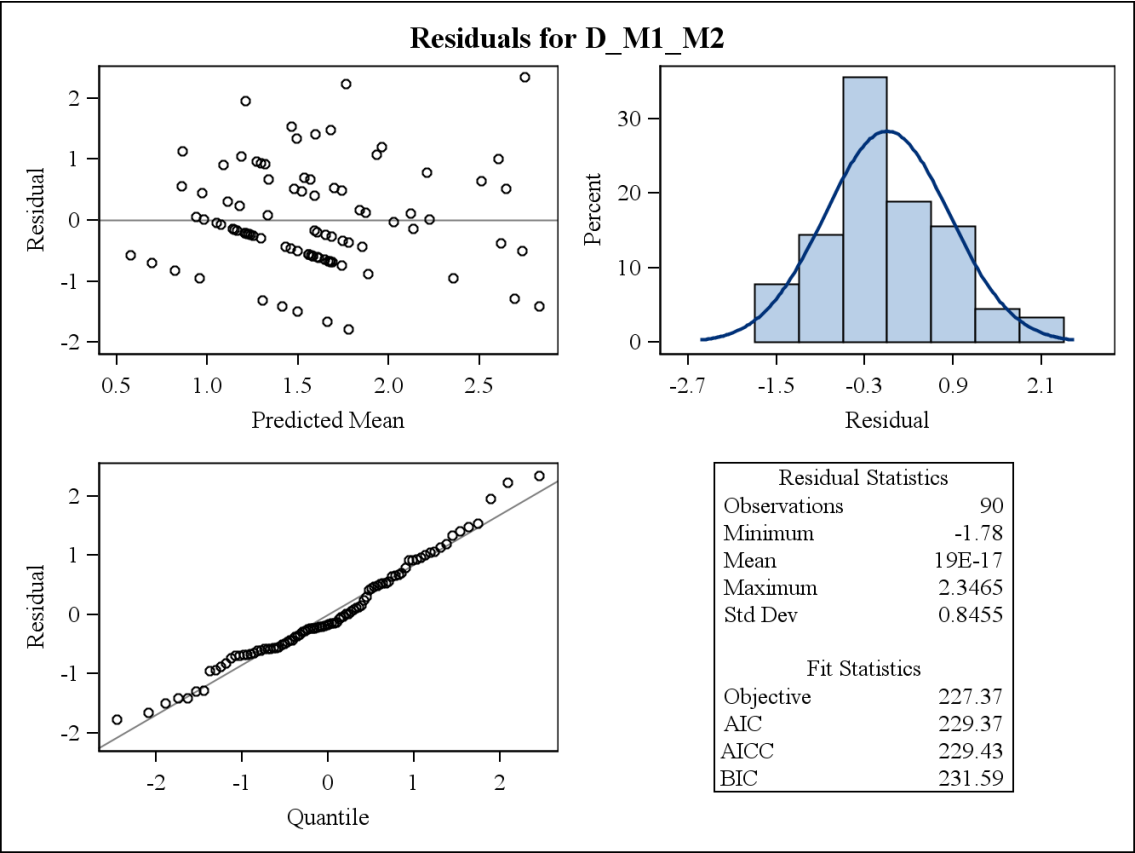
LOCAL=13



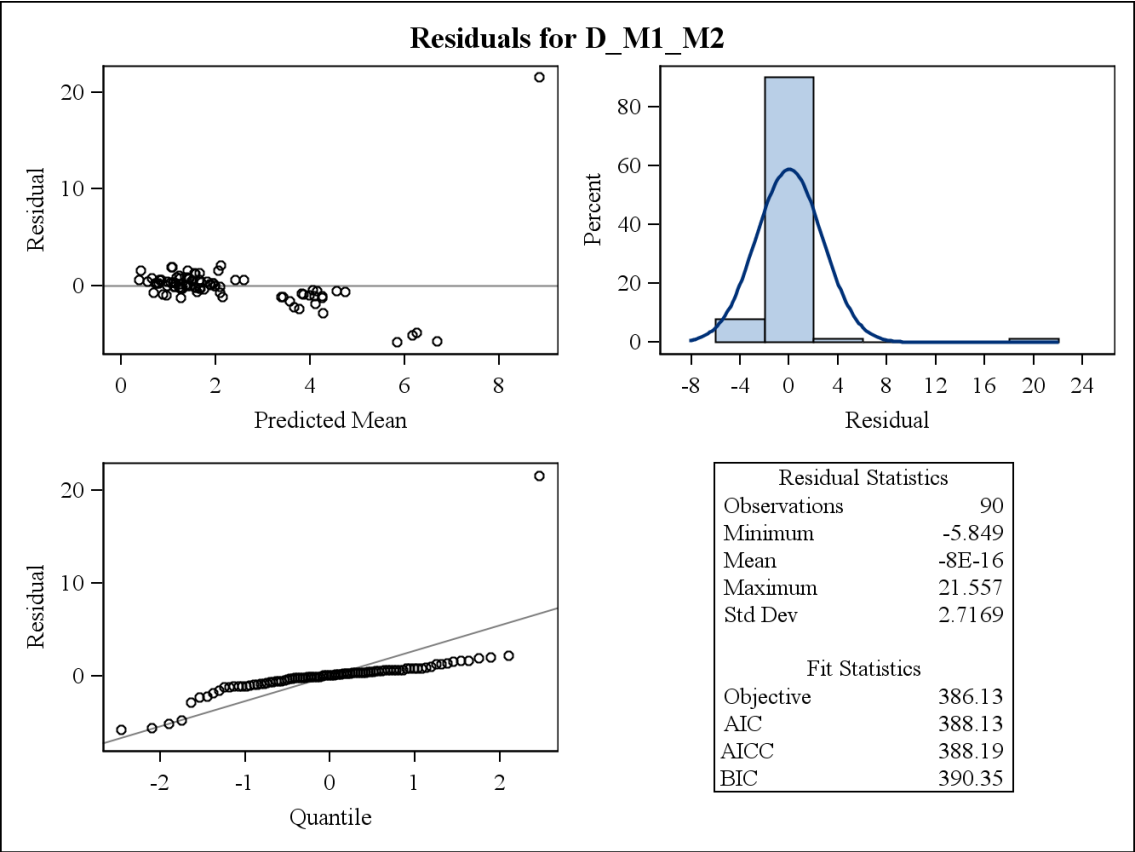
LOCAL=14



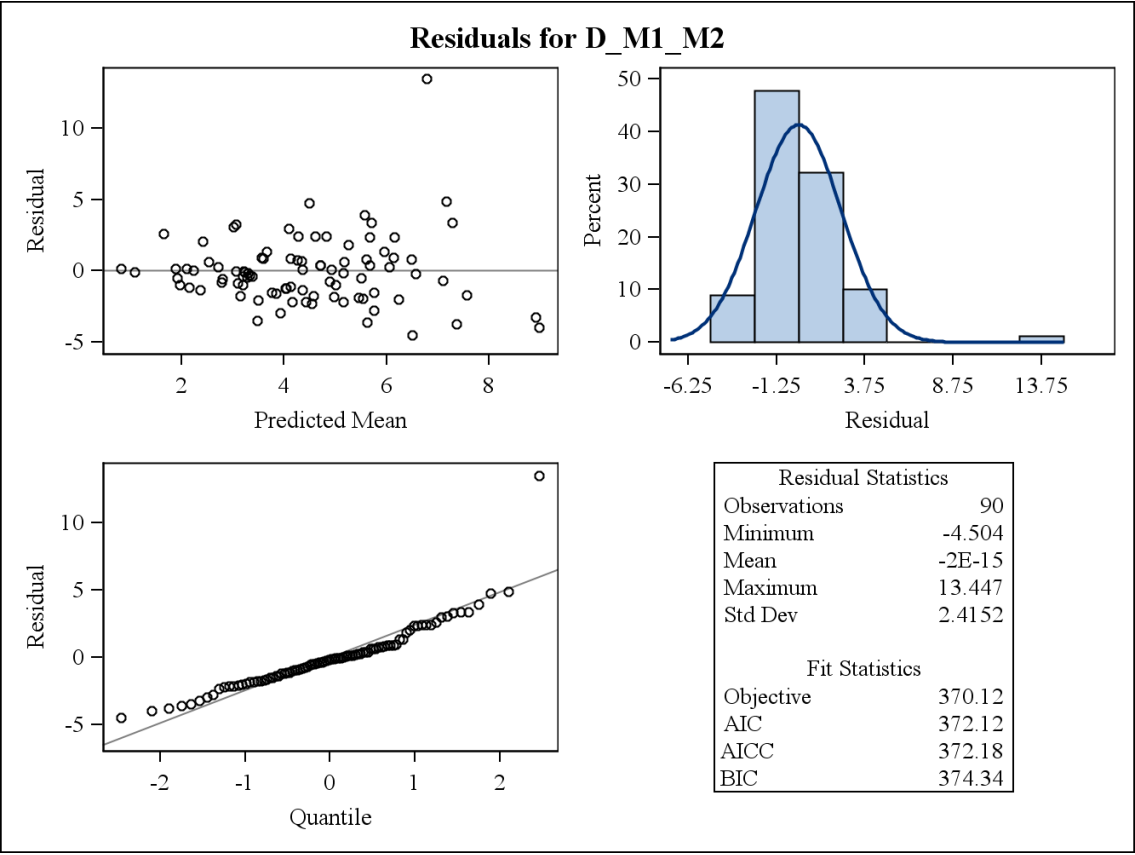
LOCAL=15



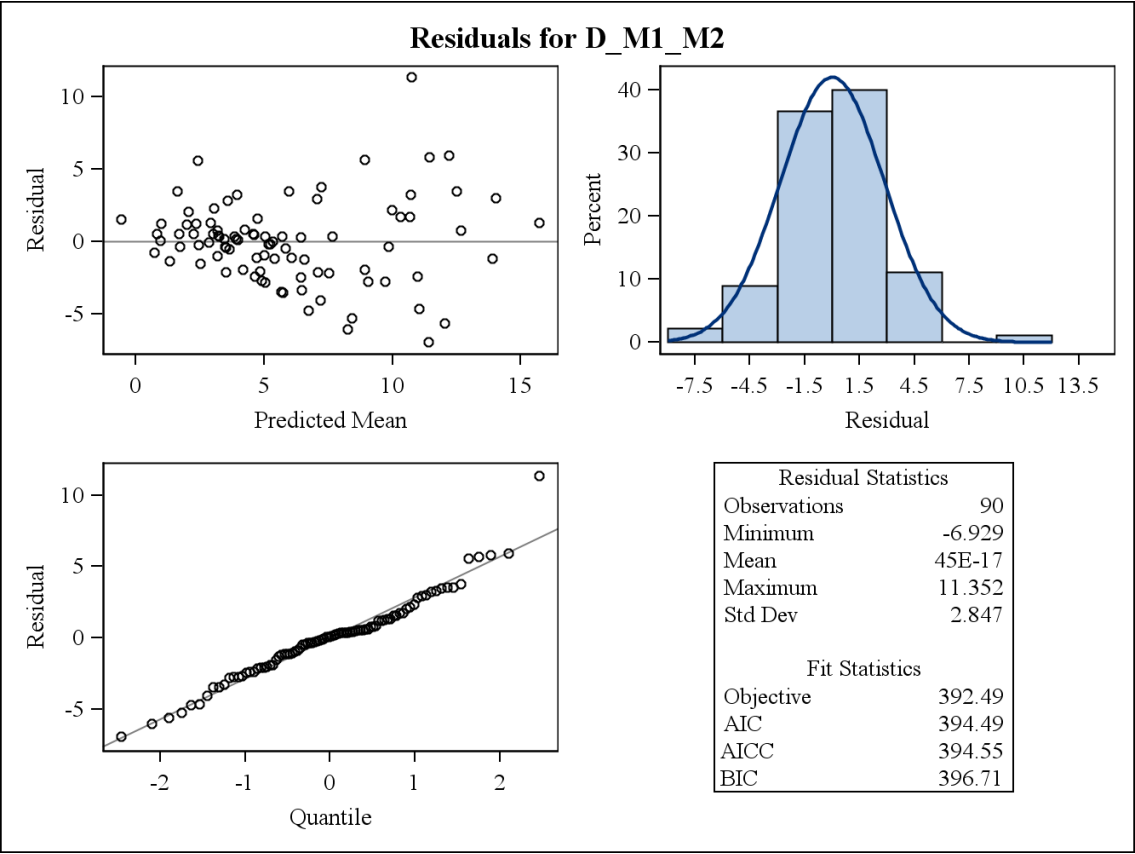
LOCAL=16



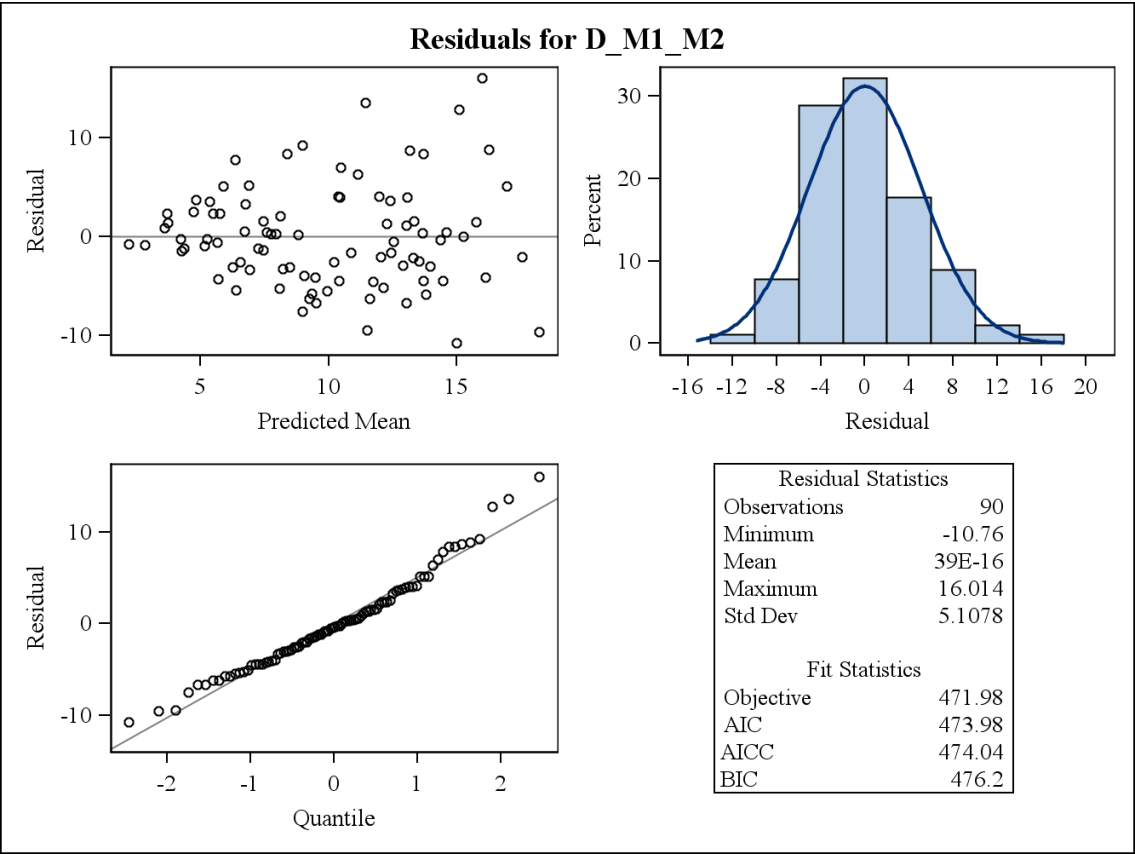
LOCAL=17



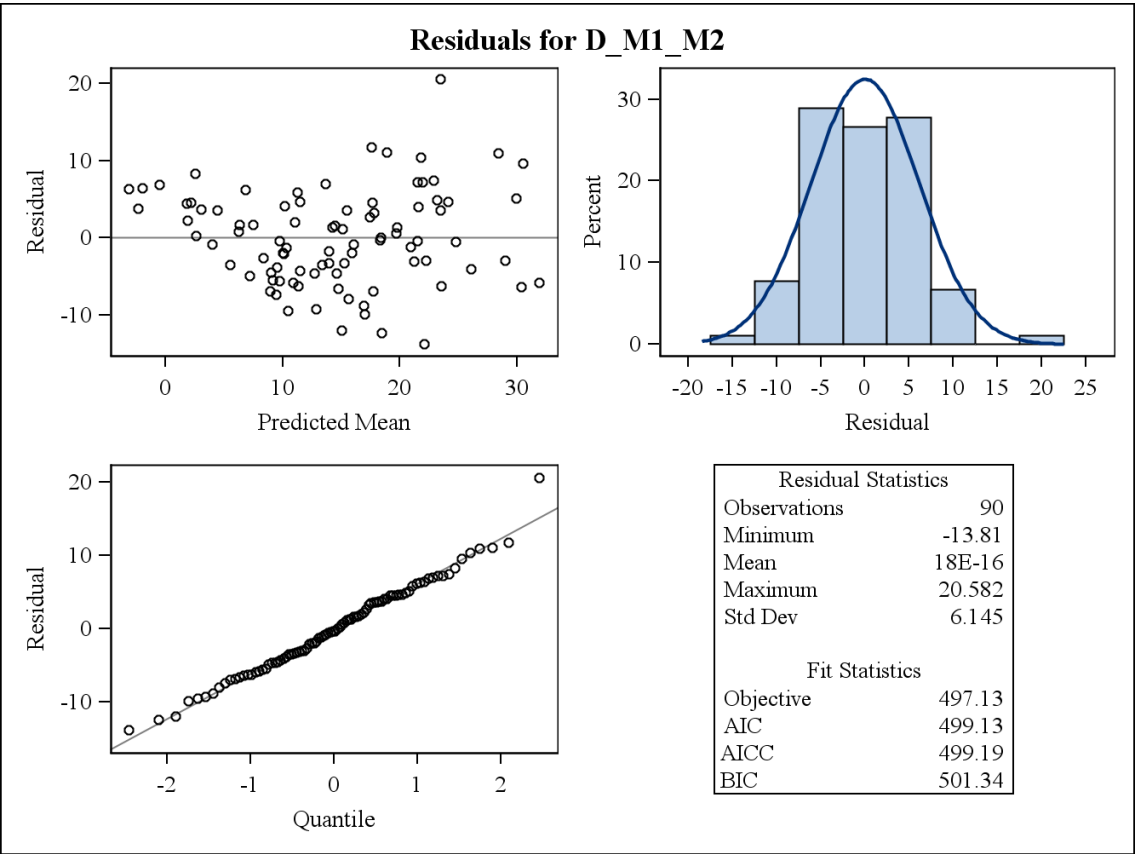
LOCAL=18



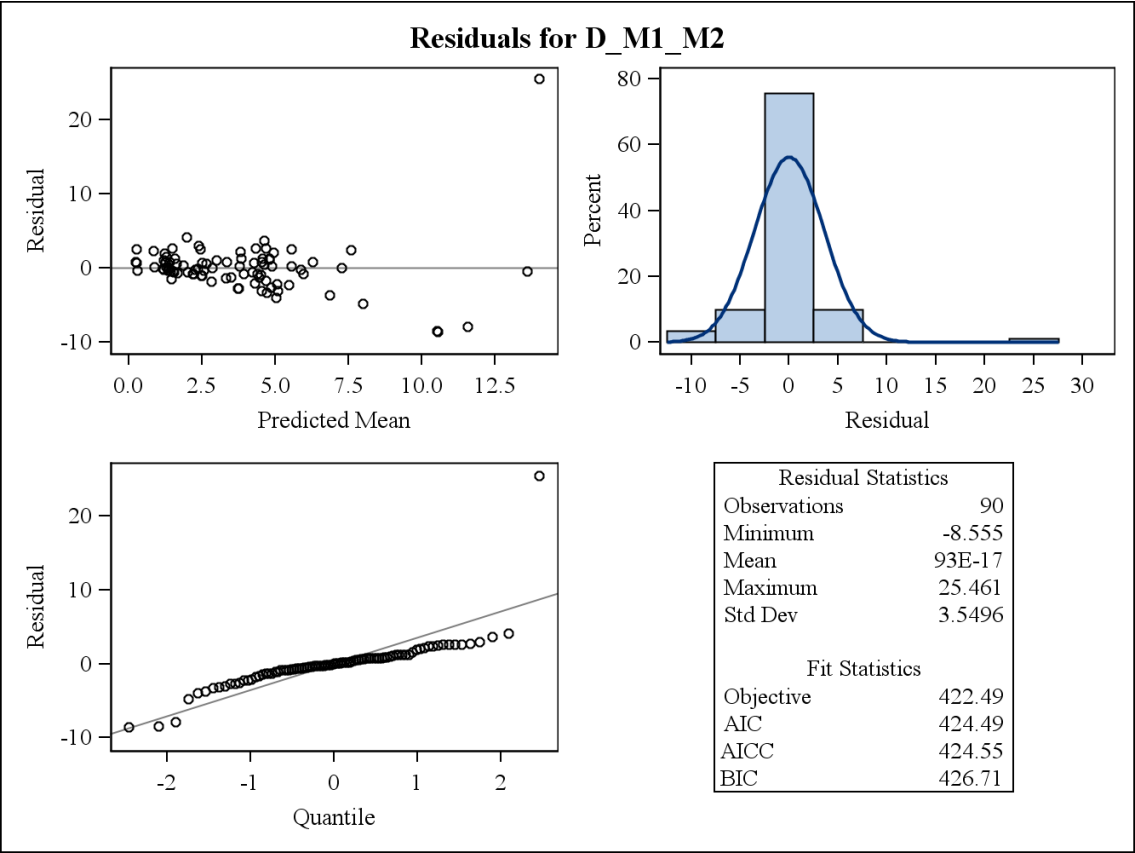
LOCAL=19



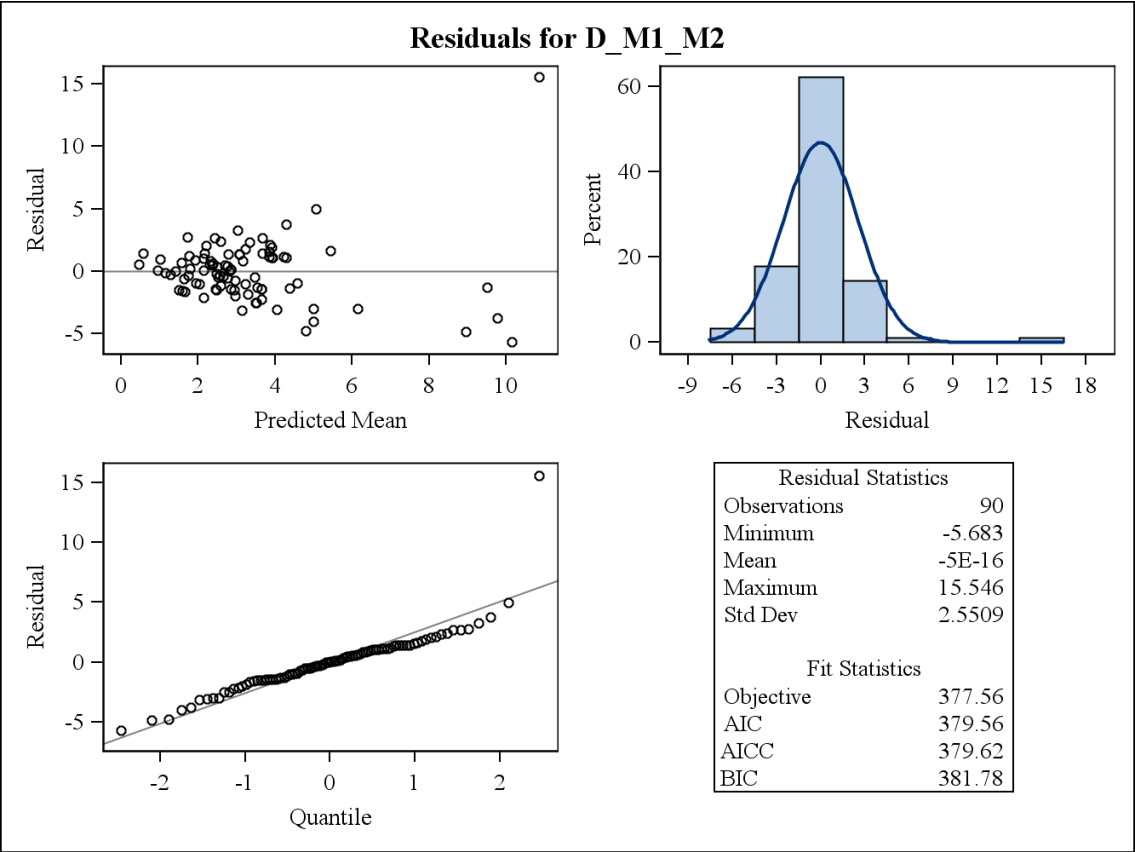
LOCAL=20



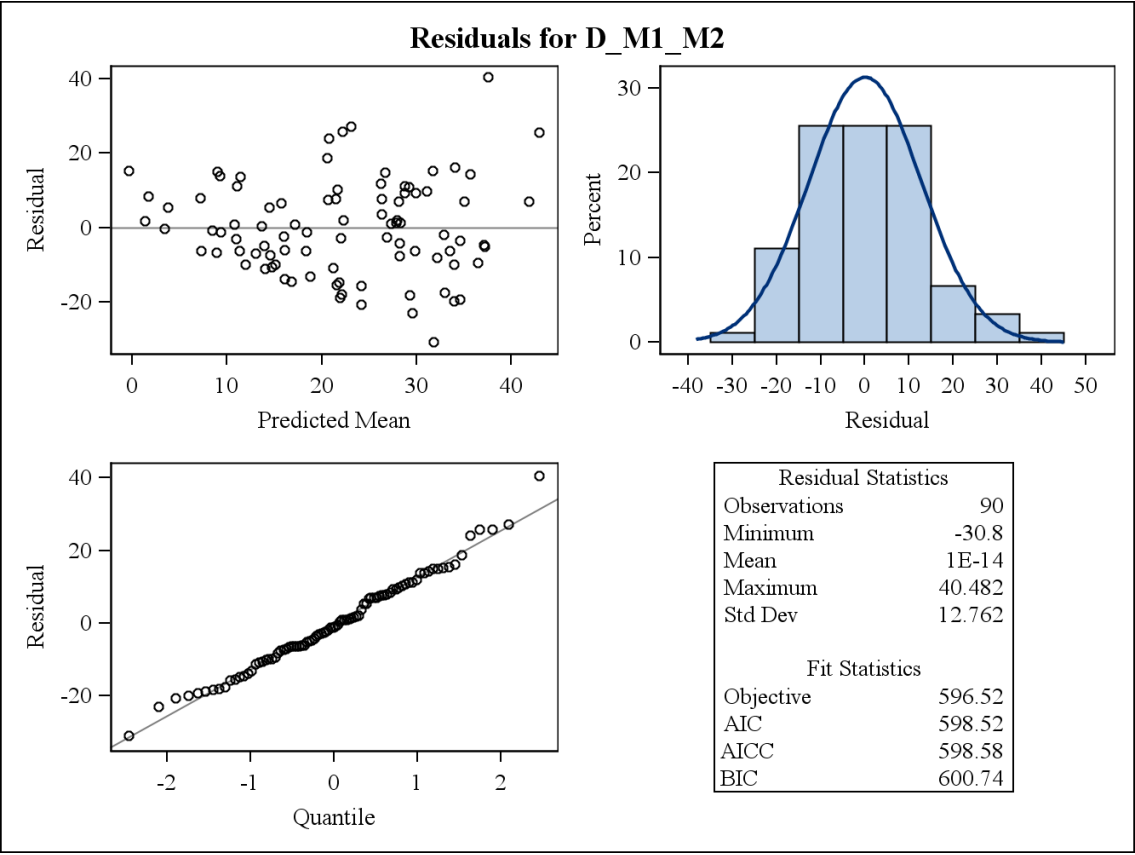
LOCAL=21



LOCAL=22



LOCAL=23



LOCAL=24

