ALEX DA SILVA TEMOTEO

ANÁLISE CONJUNTA DE FATORES: DISTRIBUIÇÃO AMOSTRAL DA IMPORTÂNCIA RELATIVA POR SIMULAÇÃO DE DADOS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-graduação em Estatística Aplicada e Biometria, para obtenção do título de "*Magister Scientiae*".

VIÇOSA
MINAS GERAIS - BRASIL
2008

ALEX DA SILVA TEMOTEO

ANÁLISE CONJUNTA DE FATORES: DISTRIBUIÇÃO AMOSTRAL DA IMPORTÂNCIA RELATIVA POR SIMULAÇÃO DE DADOS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-graduação em Estatística Aplicada e Biometria, para obtenção do título de "*Magister Scientiae*".

APROVADA: 17 de novembro de 2008

Luiz Alexandre Peternelli
Co-Orientador

José Ivo Ribeiro Júnior

Sebastião Martins Filho

Valéria Paula Rodrigues Minim

Carlos Henrique Osório Silva
Orientador

À minha mãe Inez, à minha esposa Cássia e minha filha Isabelly.

Agradecimentos

- Agradeço primeiramente à Deus, porque a fé nos dá força na hora que mais precisamos.
- Agradeço também às pessoas que direta ou indiretamente, participaram dessa conquista: minha esposa Cássia, pelo apoio, paciência e credibilidade em mim depositada; minha filha Isabelly, que é a razão de todas as minhas conquistas; minha mãe Inez, que com tanta dificuldade me apoiou e acreditou em mim em todos os momentos;
- ao Professor Carlos Henrique, que além de meu orientador, me trata como um verdadeiro amigo, e quero que saiba que a recíproca também é verdadeira;
- agradecimento especial à Consuelo Arellano (North Carolina State University) pela grande colaboração e dicas de programação para simulação dos dados no SAS;
- à Cleuza Salomé, incentivadora e uma das responsáveis por minha vinda para Viçosa;
- aos professores do departamento, que sempre se mostraram dispostos a ajudar;
- aos membros da banca, pelo voto de confiança;
- ao Fernando Bastos pela força com o Tex;
- Thiago, que sempre dava um jeito de dar uma navegada na internet (até que comprou seu notebook); Danilo, meu companheiro de arroz com pequi e Maria Nilsa com a família, que veio para Viçosa e ganhou um estimadorzinho.
- Às diretoras Rita Queiroz e Maria Amélia da Escola Estadual Dr Mariano da Rocha, pelo incentivo e disposição na elaboração de documentos e esforços necessários para se fazer o pedido de licença.
- à todos da UNIPAC, pela confiança em mim depositada. Aos meus afilhados do sexto período da matemática 2008, turma que tive o prazer de, além de amigo, ser professor, coordenador e paraninfo.

Sumário

Lista de Figuras

Lista de Tabelas

Resumo

TEMOTEO, Alex da Silva M.Sc., Universidade Federal de Viçosa, novembro de 2008. **Análise conjunta de fatores: distribuição amostral da Importância Relativa por simulação de dados.** Orientador: Carlos Henrique Osório Silva. Coorientadores: Luiz Alexandre Peternelli, Fabyano Fonseca e Silva, Adair José Regazzi.

Conjoint analysis ou análise conjunta de fatores (ANCF) é uma análise de regressão que utiliza um modelo com variáveis explicativas indicadoras ou dummy, para se estudar a preferência de consumidores por tratamentos que podem ser serviços ou produtos, e que são definidos pela combinação de níveis de diversos atributos ou fatores. Com esta técnica estima-se a Importância Relativa (IR) de cada fator que compõe os tratamentos avaliados. Tais estudos são importantes por permitir decidir, com base nas estimativas das IR de cada fator, quais devem ser observados com maior atenção na definição do tratamento. No presente trabalho foi realizado um estudo por simulação para se investigar a robustez da distribuição amostral do estimador da IR de um fator, à variações na distribuição do erro aleatório do modelo de regressão empregado na ANCF. Foram gerados erros aleatórios com a distribuição normal e também três outras distribuições alternativas obtidas por uma transformação de locação e escala da beta: uma distribuição assimétrica à direita, outra assimétrica à esquerda e uma com forma de U. Para cada distribuição, utilizou-se desvio-padrão $\sigma = 2,8$ e $\sigma = 0,5$, portanto para oito condições foram simulados 100 conjuntos de dados referentes a avaliações (notas de aceitação) de 108 consumidores para cada um dos 36 tratamentos formados pela combinação de 4 fatores (A, B, C e D) num esquema fatorial

completo $3^2 \times 2^2$. Definiu-se com base em um modelo de regressão para ANCF, valores de referência para as IR's iguais a 44,25%, 25,66%, 26,55% e 3,54%, respctivamente para os fatores A, B, C e D. Na avaliação dos resultados com base em intervalos de confiança percentil e pela aproximação normal, ambos a 95%, verificou-se intervalos mais estreitos pela aproximação normal. Conforme esperado, verificou-se intervalos de confiança para as IR's mais amplos quando $\sigma = 2,8$.

Observou-se que todos os intervalos de confiança incluíram os valores das IR's tomados como referência, exceto para os seguintes casos: (i) intervalo de confiança pela aproximação normal para a simulação de erros com distribuição normal e $\sigma = 2,8$, para os fatores A e B; (ii) com intervalo pela aproximação normal e $\sigma = 0,5$, (iia) para os fatores A e C com distribuição normal, em forma de U e assimétrica à esquerda; (iib) para o fator B com distribuição em forma de U; e (iic) para o fator D com distribuição normal e em forma de U . Entretanto, neste casos de não inclusão do valor IR de referência nos intervalos, observou-se que o valor estava próximo ao limite do IC, tanto à esquerda quanto à direita. As estimativas de IR obtidas no estudo por simulação também foram avaliadas pelo Erro Médio Relativo (EMR) com relação aos respectivos valores de referência. Exceto para o fator D na simulação com erros normais e $\sigma = 2,8$, na qual se obteve $EMR = 7,91\%$, em todas as demais situações simuladas obteve-se $EMR < 5\%$. Adicionalmente, o teste de Kolmogorov-Smirnov indicou normalidade ($p > 0,05$) das distribuições amostrais em todos os casos.

Concluiu-se que o estimador da IR pode ser considerado como robusto à não normalidade da distribuição do erro aleatório do modelo de regressão utilizado na ANCF. Adicionalmente, pode-se considerar que a distribuição amostral da IR seja normal e que portanto métodos inferenciais que requerem normalidade podem ser aplicados às estimativas de IR's obtidas na ANCF.

Abstract

TEMOTEO, Alex da Silva M.Sc., Universidade Federal de Viçosa, november of 2008. **Conjoint analysis: sampling distribution of the Relative Importance by data simulation** Adviser: Carlos Henrique Osório Silva. Co-advisors: Luiz Alexandre Peternelli, Fabyano Fonseca e Silva and Adair José Regazzi.

Conjoint analysis is a regression analysis that uses a model with dummy or indicator explanatory variables to study consumer preference for treatments that can be products or services, and are defined by combining levels of each attribute or factor. It allows the estimation of the Relative Importance (RI) of each factor that makes up the treatments. Such studies are important to help decide, based on RI estimates obtained from the CA, to which factors should be given more attention when developing the products and/or services. In this research work we conducted a simulation study in order to investigate the robustness of the RI sampling distribution to departures from normality for the distribution of the random error term (ϵ) of the CA model. We simulated four alternative distributions for ϵ and generated data (acceptance notes) that allowed estimation of RI for hypothetical factors A, B, C and D considered in our study. In addition to the normal distribution, we used a location and scale transformation of the beta density to generate three alternative distributions: right skewed, left skewed, and also an U-shape distribution. Each one of these four distributions was tested with two standard error values ($\sigma = 2.8$ and 0.5) which resulted in eight alternative scenarios. Our simulation study considered factors A and B with 3 levels and factors C and D with two levels, hence 36 treatments in a full factorial design. We set

reference RI values of 44.25%, 25.66%, 26.55% and 3.54%, respectively for factors A, B, C and D, and simulated data such that each treatment was evaluated by 108 consumers. This data set with 3888 observations was simulated 100 times for each scenario and analyzed by CA which resulted in 100 RI estimates for each factor at every scenario.

Results were investigated by 95% confidence intervals (CI) using the usual normal approximation and also percentile intervals, histograms of RI values sampling distribution to check normality, and also we calculated relative mean errors of estimation (RME) with respect to the reference RI values. It was observed that the confidence intervals included the values of RI's taken as reference in all scenarios, with the exception of: (i) factors A and B, with the normal CI using normal distribuition and $\sigma = 2.8$; (ii) with normal CI and $\sigma = 0.5$, (iia) for factors A and C with normal distribution, U shaped and left skewed; (iib) for factor B with U shaped model and (iic) for factor D with normal distribution and U shaped. In all these cases were the CI missed the RI reference value, we observed close miss left and miss right results. We observed RME $< 5\%$ in all scenarios except for normal distribution and factor D only, for which RME $= 7.91\%$.

We concluded that the sampling distribution of the estimator of the RI of a factor is relatively robust to departures from the normal distribution. In fact, results showed that its sampling distribution must be close to the normal, regardless of the distribution of the random error term of the CA model.

Capítulo 1

Introdução

Conjoint Analysis ou Análise Conjunta de Fatores (ANCF) é uma análise de regressão que tem sido utilizada em estudos da preferência do consumidor, tanto para avaliar produtos quanto para avaliar serviços (tratamentos).

Os tratamentos estudados são formados pela combinação de diversos níveis de fatores ou atributos. Por exemplo, a aceitação e escolha de embalagens de óleo de soja, pode ser entre embalagem PET ou Lata, pelas informações do rótulo como marca (conhecida ou não), presença ou ausência de informação nutricional, informação sobre soja transgência ou não e preço alto ou baixo. Outro exemplo é a embalagem de couve minimamente processada, que pode ser escolhida pela cor (verde ou amarelo), informação (com ou sem informação), tipo de produção (orgânico, sem produtos químicos ou sem informação), visibilidade do produto (pouca ou muita visibilidade) e preço (baixo ou alto). Ainda nas ciências agrárias podemos considerar a preferência por frutos de maçã, que pode ser investigada quanto aos fatores cor (vermelho ou verde), brix (baixo ou alto), tamanho do fruto (pequeno, médio ou grande) e acidez (baixo ou alto).

Os três exemplos citados foram propositalmente escolhidos da área de tecnologia de alimentos, a qual foi responsável pela motivação do tópico da presente dissertação.

Outras áreas do conhecimento tais como em ciências humanas, indústria hoteleira e de prestação de serviços (bancos, lojas e seguradoras), indústria automobilística, indústria de calçados, etc. também empregam a ANCF em estudos que visam definir perfis de produtos e serviços.

1.1 Objetivos da Pesquisa

A importância relativa (IR%) de um fator ou atributo é uma das principais informações resultantes da aplicação da ANCF. Por exemplo, se quatro fatores são investigados tem-se $\sum_{i=1}^{4} IR_i = 100\%$ e ainda que IR_i representa o impacto que o i-ésimo fator causa na aceitação do consumidor pelo tratamento. Portanto, se um fator possui $IR = 80\%$ ele é mais importante do que um fator com $IR = 10\%$. Mais importante no sentido de que deve ser dado mais atenção a ele na elaboração do tratamento (definição do produto e/ou serviço). Entretando, se a ANCF estima IR's iguais a 20% e 22% para dois fatores, o que pode ser concluído quanto à significância estatística da diferença observada? Existem pesquisadores que analisam os resultados aplicando-se testes de média como Tukey e Duncan, onde uma das pressuposições é a normalidade dos dados. Adicionalmente, como a variável resposta é uma nota de aceitação, pode-se questionar a validade de se empregar um modelo de regressão que requer normalidade dos erros aleatórios para validar testes de significância. Portanto, outra questão de interesse é a robustez da distribuição amostral do estimador da IR de um atributo à variações na distribuição do erro aleatório do modelo de regressão empregado na ANCF.

O entendimento destas questões permitirá, em trabalhos futuros, que se construa testes ou intervalos de confiança para se inferir com base nas estimativas de IR obtidas na ANCF, e, consequentemente uma melhor compreensão dos resultados obtidos em estudos da aceitação que utlizam a ANCF.

No presente trabalho apresentaremos um estudo realizado com dados obtidos por simulação, visando-se investigar a robustez da distribuição amostral do estimador da IR. O objetivo geral deste trabalho é estudar o efeito da não normalidade da distribuição dos erros aleatórios do modelo de regressão empregado na ANCF. Como objetivos específicos temos:

- comparar a amplitude e também a inclusão dos valores das IR de referência nos intervalos de confiança percentis e também nos intevalos obtidos pela aproximação normal.

- verificar se a distribuição amostral do estimador da IR, pode ser considerada como normal, mesmo com a não normalidade da distribuição dos erros aleatórios.

- verificar a acurácia preditiva do estimador da IR, por intermédio das estimativas dos erros relativos médios.

Capítulo 2

Revisão de Literatura

2.1 Análise conjunta de fatores

O termo conjoint analysis é da língua inglesa e em português a tradução como análise conjunta de fatores (ANCF) foi proposta por CARNEIRO, SILVA E MINIM (2006). Outros termos propostos são análise de preferência e análise combinada de fatores. ANCF é uma técnica que pode ser aplicada em qualquer área de tomada de decisões.

Em Ciências Agrárias ou afins tais como na Ciência e Tecnologia de Alimentos, a ANCF é uma alternativa que vêm sendo utilizada em estudos da intenção de compra dos consumidores por rótulos e embalagens de alimentos.

É uma análise de regressão utilizada para decompor a preferência de indivíduos por tratamentos a ele apresentado em partes devidas a cada um dos fatores (ou atributos) que o compõem. Os denominados tratamentos podem ser objetos, serviços, produtos, etc, mas estes devem ser formados pela combinação dos mesmos fatores, porém em diferentes versões ou níveis. É também uma metodologia fundamentada numa análise de decomposição, na qual os entrevistados (ou consumidores) reagem a um produto informando sua aceitação global sobre

ele e, à partir desta, calcula-se o valor das contribuições que cada nível de cada fator tem sobre ela, decompondo-a (Green e Srinivasan, 1987). Desta forma, em vez do consumidor avaliar cada fator (ou cada nível) separadamente, ele avalia conjuntamente os fatores, por meio de combinações dos seus níveis, as quais formam os produtos em avaliação (MOSKOWITZ et al, 2004). Assim, assume-se que o tratamento avaliado pode ser decomposto em seus componentes, podendo ser estimada a importância que cada um deles tem sobre a decisão do consumidor (DELLA LUCIA, 2008).

Como uma regra geral, (SIQUEIRA, 2000) aconselha-se que nenhum dos fatores possua número de níveis muito maior do que dos outros, pois isso faria com que esse fator chamasse mais a atenção em relação aos outros, o que implicaria em uma maior importância relativa para esse fator.

Se a variável resposta for métrica ou ordinal, o método de estimação geralmente utilizado é o Método dos Mínimos Quadrados (MMQ). Quando a variável dependente é binária, deve-se fazer uso do Método da Máxima Verossimilhança (MMV) (SIQUEIRA, 2000). Os resultados da ANCF são avaliados quanto à contribuição de cada nível de cada fator e quanto às importâncias relativas dos fatores (SAS, 1993), conforme explicaremos resumidamente a seguir, adaptado de CARNEIRO, SILVA E MINIM (2006). Considere um experimento com n fatores cada um com m_i níveis, sendo estes fatores variáveis qualitativas ou quantitativas. Cada tratamento é obtido pela combinação dos níveis dos fatores, portanto, os tratamentos constituem um fatorial. Todos os tratamentos (fatorial completo) ou uma parte cientificamente estabelecida (fatorial fracionado) são submetidos à avaliação pelos consumidores, em geral quanto à aceitação ou intenção de compra. Os tratamentos

são apresentados aos consumidores e estes podem, por exemplo, atribuir notas, em uma escala previamente estabelecida. Neste caso as notas são as variáveis respostas ou os dados submetidos à análise. Um modelo estatístico comumente utilizado em ANCF é,

$$Y_{jk} = \tau_j + \epsilon_{jk},$$

com

$$\tau_j = \beta_0 + \sum_{i=1}^{m_1} \beta_{1i} X_{1i}^j + \cdots + \sum_{i=1}^{m_n} \beta_{ni} X_{ni}^j$$

onde Y_{jk} é a variável resposta, (nota de aceitação ou intenção de compra) dada ao j-ésimo tratamento pelo k-ésimo consumidor, para j=1, 2, ..., N tratamentos avaliados por cada um dos k consumidores participantes da pesquisa. Para s= 1, 2, ..., n fatores cada um com m_i níveis, tem-se $X_{si}^j = 1$ quando o i-ésimo nível do s-ésimo fator está presente no j-ésimo tratamento e $X_{si}^j = 0$ caso contrário. β_{si} é o coeficiente de preferência (CP) associado ao i-ésimo nível do s-ésimo fator (denominados *part-worths* em *conjoint analysis*). O termo ϵ_{jk} é o erro aleatório não observável associado à observação Y_{jk}. O erro aleatório pode ser definido como a diferença entre a nota dada pelo consumidor k e a nota média de todos os consumidores para um mesmo tratamento, para um número muito grande de consumidores que tende ao infinito. O modelo acima pode ser apresentado compactamente na notação matricial como: $Y = X\beta + \epsilon$, em que Y é o vetor de observações correspondente aos tratamentos avaliados, X é a matriz de variáveis indicadoras da presença ou ausência dos níveis dos fatores e β é o vetor de parâmetros a serem estimados (part-worths ou CP´s). O modelo acima é denominado de efeitos principais, pois não inclui interações entre os fatores.

Adicionalmente, podem ser incluídas interações entre dois ou mais fatores, entretanto, a inclusão de interações implica na necessidade de estimação de novos parâmetros β_{si} e conseqüentemente na necessidade de um maior número de observações, isto é, na avaliação de um maior número de tratamentos pelos consumidores, o que em termos práticos pode inviabilizar o estudo. Portanto, a ANCF é uma técnica que requer um cuidadoso planejamento experimental.

O vetor β pode ser estimado pelo método dos mínimos quadrados ordinários o que requer a solução do sistema de equações normais $X'X\widehat{\beta} = X'Y$.

Em ANCF, assim como em muitas aplicações em estatística experimental, a matriz X não é de posto coluna completo, e portanto, artifícios se fazem necessários se for desejado uma solução única $\widehat{\beta}$ para o vetor β. Para facilitar a interpretação das estimativas $\widehat{\beta}_{si}$ impõe-se as restrições $\sum_{i=1}^{m_i} \beta_{si}=0$, para todo fator s. Estas restrições completam o posto da matriz X , de modo que o sistema de equações normais passa a ter solução única

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

.

As restrições também permitem interpretações importantes para as estimativas dos $\widehat{\beta}_{si}$. Pode-se concluir que $\widehat{\beta}_{si} < 0$ significam efeito desfavorável, ou seja, que diminuem a nota de preferência pelo produto, enquanto $\widehat{\beta}_{si} > 0$ significam efeito favorável na preferência do consumidor. Com os valores $\widehat{\beta}_{si}$, podemos estimar a Importância de um fator s, que é dado pelo máximo valor de $\widehat{\beta}_{si}$ subtranindo do mínimo valor de $\widehat{\beta}_{si}$, ou seja, $I_s = max(\widehat{\beta}_s) - min(\widehat{\beta}_s)$. Estima-se a Importância Relativa (IR) de cada fator, pela importância dele dividido pela soma das importâncias de todos os fatores,

$$IR_s\% = \frac{I_s}{\sum_{s=1}^{n} I_s} \times 100\% \tag{2.1}$$

A importância relativa pode ser interpretada como o "impacto", ou o efeito que o fator tem sobre a aceitação do produto pelo consumidor. A soma de todas as IR's resulta em 100%.

DELLA LUCIA (2008) aplicou a ANCF na avaliação da intenção de compra e da escolha do iogurte light sabor morango. Ela estimou a IR de três fatores, todos com dois níveis cada um e obteve os seguintes resultados: (i) Informação sobre o conteúdo de açúcar, $\widehat{IR} = 60{,}2\%$, (ii) Informação sobre o conteúdo de gordura, $\widehat{IR} = 10{,}6\%$ e (iii) Informação sobre o conteúdo de proteína, $\widehat{IR} = 29{,}2\%$

CARNEIRO (2007) aplicou ANCF para avaliar os fatores da embalagem e do rótulo de cachaça na aceitação dos consumidores. Foram avaliados cinco fatores, todos eles com dois níveis cada um: (i) Marca, (ii) Ilustração do rótulo, (iii) Tempo de envelhecimento, (iv) Tipo de madeira e (v) Tipo de embalagem. As estimativas das IR's foram apresentadas para diversos grupos.

CASTRO (2006) aplicou a ANCF na Indústria Hoteleira para avaliar os pacotes de serviços preferidos pelos clientes de um hotel na cidade de São Paulo. Foram avaliados cinco fatores: (i) Equipamentos (com três níveis) que resultou em uma $\widehat{IR} = 27{,}60\%$, (ii) Apartamentos (com três níveis) com $\widehat{IR} = 32{,}39\%$, (iii) Serviços de quarto (com três níveis) e $\widehat{IR} = 17{,}15\%$, (iv) Café da Manhã (dois níveis) e $\widehat{IR} = 12{,}77\%$ e (v) Conforto do Banheiro (dois níveis) com $\widehat{IR} = 10{,}19\%$.

ABADIO (2003) utilizou a técnica para avaliar o efeito de diferentes fatores de in-

formação da embalagem de suco de abacaxi na intenção de compra do consumidor. Foram avaliados cinco níveis: (i) Marca (três níveis) que apresentou $\widehat{IR} = 24\%$, (ii) Definição do Produto (dois níveis) com $\widehat{IR} = 1{,}4\%$, (iii) Informação sobre tecnologia (três níveis) com $\widehat{IR} = 10{,}6\%$, (iv) Tipo de produção (dois níveis) com $\widehat{IR} = 9{,}5\%$ e (v) Preço (dois níveis) com $\widehat{IR} = 25{,}9\%$. Neste estudo também foi aplicado o teste t para inferir quanto à diferença significativa entre os coeficientes de preferência.

CARNEIRO (2002) aplicou ANCF para avaliar os fatores que causam impacto na embalagem de óleo de soja na intenção de compra dos consumidores. Os resultados foram apresentados para quatro grupos distintos obtidos por análise de agrupamento dos consumidores com base nos CP's. Apresentaremos para ilustar os resultados do grupo 1. Foram avaliados quatro fatores, todos com dois níveis cada um: (i) Marca que resultou na $\widehat{IR}$ = 1,9%, (ii) Preço com $\widehat{IR} = 6.5\%$, (iii) Informação nutricional, com $\widehat{IR} = 2{,}6\%$e (iv) Informação sobre o tipo de soja com $\widehat{IR} = 89\%$. Neste estudo, foi aplicado o Teste t para verificar diferença significativa entre os coeficientes de preferência.

KOHLI (1988) propôs dois testes de significância para os fatores avaliados na ANCF. Um pode ser aplicado quando as preferências dos consumidores pelos fatores podem ser ordenados a priori e o outro pode ser usado quando esta ordenação não é permitida.

2.2 Intervalos de confiança

Não encontramos relatos na literatura à respeito da construção de intervalos de confiança para a IR de um fator. Acreditamos que inferir à respeito das IR's dos fatores investigados numa ANCF possa ser realizado com base em intervalos de confiança (IC). Entretanto, a construção de IC para a IR requer que se conheça a sua distribuição amostral ou pelo

menos que se possa estimar o seu desvio-padrão de modo que se possa construir um IC pela aproximação normal, e neste caso, a validade do intervalo precisa ser investigada. O método delta (CASELLA e BERGER, 2002) pode ser uma alternativa para se obter uma estimativa do desvio-padrão de $\widehat{IR}$. Um intervalo de confiança é uma afirmação probabilística,

$$P[LI \leq \mu \leq LS] = 1 - \alpha,$$

sendo $1 - \alpha$ o nível de confiança do intervalo. Portanto, IC e testes de hipótese são procedimentos equivalentes, pois se o valor do parâmetro sob H_0, pertencer ao intervalo recém construido, H_0 não deve ser rejeitada, caso contrário, rejeita-se a hipótese nula. É conveniente salientar que, se a hipótese nula é do tipo $H_0 : \mu = \mu_0$, o intervalo deve ser do tipo $[LI; LS]$; se $H_0 : \mu \geq \mu_0$, então, o intervalo é $[LI; \infty)$; e se $H_0 : \mu \leq \mu_0$ o intervalo é $(-\infty; LS]$.

Segundo FERREIRA (2005), a estimação pontual não fornece idéia de margem de erro que é cometida ao se estimar um determinado parâmetro. A estimação por intervalo procura corrigir essa lacuna a partir da criação de um intervalo associado a uma probabilidade de se conter o verdadeiro valor do parâmetro. Essa probabilidade é dada por $1 - \alpha$ de modo que α se refere à probabilidade de o real valor do parâmetro não pertencer ao intervalo gerado. Por outro ponto de vista, o coeficiente de confiança pode ser interpretado como a proporção média de intervalos de confiança, gerados à partir de um grande número de amostras, que incluirá o parâmetro, ou seja, que conterá o valor real do parâmetro entre seus limites. O intervalo de confiança pela aproximação da distribuição normal é motivado pelo Teorema Central do Limite (ver por exemplo MAGALHÃES(2006)), o qual garante que para grandes amostras ($n \to \infty$) tem-se $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ e portanto pode-se definir α tal que:

$$P\left(-Z_{\alpha/2} \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq Z_{\alpha/2}\right) \approx 1-\alpha$$

o que nos fornece

$$P\left(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \approx 1-\alpha$$

onde os limites inferiores e superiores do IC são dados por:

$$LI = \overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

e

$$LS = \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

onde μ e σ são a média e o desvio padrão da população, respectivamente.

TRINDADE e ROTONDARO (2004) aplicaram a ANCF e propuseram a modificação da escala por postos fornecida pelo entrevistado em uma nova escala, ainda discreta, que mantenha a posição relativa entre os produtos originalmente ordenados. O objetivo desta mudança é estimar um modelo em que a reconstrução da ordem original seja mais eficiente. Utilizaram intervalos de confiança a 95% para a média ou proporção para verificar a validade das respostas. Foram gerados 1000 ordenamentos aleatórios válidos dos produtos em cada cenário. No cenário com quatro fatores e três níveis, apenas 25,4% das respostas foram consideradas válidas. No cenário com 6 fatores e 2 níveis, por sua vez, a proporção de respostas válidas foi de 89,2%.

FONSECA e SALAY (2005) utilizaram intervalos de confiança a 95% para analisar os riscos alimentares de acordo com a localização demografica e socioeconômica de 158

consumidores da região de campinas.

SPERS, SAES e SOUZA (2004) aplicaram a ANCF para caracterizar o consumidor de café, tentando identificar os principais fatores que levam ao consumo desse produto. Foram entrevistados consumidores de São Paulo e Belo Horizonte, que degustaram três tipos de café. Os fatores mais importantes para a amostra estudada foram: (i) preço com $\widehat{IR} = 30\%$, (ii) tipo de café com $\widehat{IR} = 20\%$, (iii) marca, também com $\widehat{IR} = 20\%$, (iv) tipo de preparo com $\widehat{IR} = 15\%$ e (v) embalagem com $\widehat{IR} = 10\%$. Nesse estudo, cosntruiram-se intervalos de confiança a 95% para analisar a média entre as notas dos cafés analisados.

Capítulo 3

Material e Métodos

Este estudo foi desenvolvido utilizando-se os recursos e instalações do Departamento de Informática, Setor de Estatística, da Universidade Federal de Viçosa (UFV). Foi utilizado principalmente o software SAS, licenciado para a UFV.

3.1 Análise conjunta de fatores, um estudo por simulação de dados

3.1.1 Fatores e níveis

Neste estudo por simulação de dados, considerou-se um produto hipotético formado pela combinação num esquema fatorial completo, dos níveis de quatro fatores conforme apresentado na Tabela 3.1.

3.1.2 Modelo utilizado na análise

Optamos por não considerar a interação entre os fatores, já que além de não acrescentar informações adicionais ao estudo, implicaria na necessidade de estimação de um maior número de parâmetros e portanto na simulação de mais dados. Portanto, considerou-se o

Tabela 3.1: Fatores e respectivos níveis considerados no estudo por simulação de dados para a análise conjunta de fatores

Fatores	Níveis
A	pouco, médio e muito
B	fraco, médio e forte
C	sim e não
D	baixo e alto

seguinte modelo no estudo por simulação de dados,

$$\begin{aligned} Y_{jk} &= \tau_j + \epsilon_{jk} \\ &= \beta_0 + \sum_{i=1}^{3} \beta_{1i} X_{1i}^{j} + \sum_{i=1}^{3} \beta_{2i} X_{2i}^{j} + \sum_{i=1}^{2} \beta_{3i} X_{3i}^{j} + \sum_{i=1}^{2} \beta_{4i} X_{4i}^{j} + \epsilon_{jk} \end{aligned} \tag{3.1}$$

onde, Y_{jk} é a nota atribuída pelo k-ésimo consumidor ao j-ésimo tratamento, cujo efeito é τ_j, para $k = 1, 2, 3, \ldots, 108$ e $j = 1, 2, 3, \ldots, 36$; β_{si} é a *part-worth* ou o coeficiente de preferência associado ao i-ésimo nível do s-ésimo fator, $X_{si}^{j} = 0$ ou 1, é a variável indicadora da presença do i-ésimo nível do s-ésimo fator no j-ésimo tratamento (referimos o leitor à Seção 2.1).

O termo ϵ_{jk} mereceu especial atenção em nosso estudo. ϵ_{jk} é o erro aleatório do modelo, que no presente estudo foi gerado considerando-se quatro modelos de distribuição de probabilidades alternativos: distribuição normal, distribuição assimétrica à direita (AD), distribuição assimétrica à esquerda (AE), e, distribuição em forma de U (FU). Os modelos AD, AE e FU utilizados para gerarem valores aleatórios ϵ_{jk} foram obtidos por meio de uma transformação de locação e escala da distribuição beta, conforme explicaremos adiante no texto em 3.1.4. O objetivo principal deste estudo era investigar a robustez da distribuição

amostral do estimador da Importância Relativa de um fator frente à distribuições alternativas para o erro aleatório,

3.1.3 Importâncias Relativas (IR%) dos fatores

Com base no modelo 3.1 define-se a Importância (I_s) do s-ésimo fator como,

$$I_s = max(\beta_{si}) - min(\beta_{si}) \qquad s = 1, 2, 3, 4$$

e consequentemente a Importância Relativa ($IR_s\%$) do s-ésimo fator como,

$$IR_s\% = \frac{I_s}{\sum_{s=1}^{4} I_s} \times 100\% \tag{3.2}$$

Com base em pesquisas bibliográficas, tomou-se $\beta_0 = 4$ e também os valores β_{si} apresentados na Tabela 3.2. Estes valores IR% serão denominados valores de referência ou paramétricos do estudo por simulação.

Tabela 3.2: Valores de referência dos coeficientes de preferência (β_{si}) dos níveis e das respectivas importâncias relativas (IR%) dos fatores utilizados no estudo por simulação

Fator A			Fator B			Fator C		Fator D	
β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	β_{31}	β_{32}	β_{41}	β_{42}
-1,20	-0,10	1,30	-0,50	-0,45	0,95	-0,75	0,75	-0,10	0,10
IR= 44, 25%			IR= 25, 66%			IR= 26, 55%		IR= 3, 54%	

Portanto, o fatorial completo, 3^2 x 2^2, com os fatores apresentados na Tabela 3.1 resulta em 36 tratamentos conforme especificados na Tabela 3.3, juntamente com a contribuição da parte determinística do modelo 3.1, para o valor da nota Y_{jk}.

Tabela 3.3: Tratamentos resultantes da combinação no esquema fatorial completo, dos níveis dos fatores considerados no estudo por simulação de dados, com os respectivos valores dos efeitos fixos τ_j [/1]

Tratamentos	Fatores e níveis				τ_j
	A	B	C	D	
1	pouca	fraco	sim	baixo	1,45
2	pouca	fraco	sim	alto	1,65
3	pouca	fraco	não	baixo	2,95
4	pouca	fraco	não	alto	3,15
5	pouca	médio	sim	baixo	1,50
6	pouca	médio	sim	alto	1,70
7	pouca	médio	não	baixo	3,00
8	pouca	médio	não	alto	3,20
9	pouca	forte	sim	baixo	2,90
10	pouca	forte	sim	alto	3,10
11	pouca	forte	não	baixo	4,40
12	pouca	forte	não	alto	4,60
13	medio	fraco	sim	baixo	2,55
14	medio	fraco	sim	alto	2,75
15	medio	fraco	não	baixo	4,05
16	medio	fraco	não	alto	4,25
17	medio	médio	sim	baixo	2,60
18	medio	médio	sim	alto	2,80
19	medio	médio	não	baixo	4,10
20	medio	médio	não	alto	4,30
21	medio	forte	sim	baixo	4,00
22	medio	forte	sim	alto	4,20
23	medio	forte	não	baixo	5,50
24	medio	forte	não	alto	5,70
25	muito	fraco	sim	baixo	3,95
26	muito	fraco	sim	alto	4,15
27	muito	fraco	não	baixo	5,45
28	muito	fraco	não	alto	5,65
29	muito	médio	sim	baixo	4,00
30	muito	médio	sim	alto	4,20
31	muito	médio	não	baixo	5,50
32	muito	médio	não	alto	5,70
33	muito	forte	sim	baixo	5,40
34	muito	forte	sim	alto	5,60
35	muito	forte	não	baixo	6,90
36	muito	forte	não	alto	7,10

[/1]– Efeitos τ_j conforme o modelo (3.1)

3.1.4 Simulação de distribuições alternativas para o erro aleatório do modelo da ANCF

O problema inicial foi gerar valores de notas Y_{jk} atribuídas pelos consumidores para a ANCF conforme 3.1 para os tratamentos definidos conforme Tabelas 3.1 e 3.3. Para tanto, inicialmente gerou-se valores aleatórios ϵ_{jk} da distribuição contínua $g_\epsilon(\varepsilon)$ obtida por meio de uma transformação de locação e escala da distribuição beta.

No estudo por simulação, foram consideradas três formas alternativas para a distribuição $g_\epsilon(\varepsilon)$, além da distribuição normal como uma quarta alternativa. Detalhes da metodologia são apresentados à seguir.

Transformação de locação e escala da distribuição beta

Seja X uma variável aleatória tal que $X \sim beta(\alpha_1, \alpha_2)$, com parâmetros $\alpha_1, \alpha_2 > 0$. A função densidade de probabilidade (fdp) de X é dada por:

$$f_X(x) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}, \qquad 0 \leq x \leq 1,$$

com média e variância dados por,

$$E(X) = \frac{\alpha_1}{\alpha_1 + \alpha_2} \quad \text{e} \quad V(X) = \frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2(\alpha_1 + \alpha_2 + 1)}.$$

Conforme Casella & Berger (1990), citado por SILVA(2000), para $\theta > 0$, $T(X) = \epsilon = 2\theta(X - 0,5)$, define uma transformação de locação e escala da distribuição beta, com $-\theta \leq \epsilon \leq \theta$ tal que o Jacobiano da transformação é:

$$J = \frac{\partial T^{-1}(\epsilon)}{\partial \epsilon} = \frac{\partial \left(\frac{\epsilon}{2\theta} + 0,5\right)}{\partial \epsilon} = \frac{1}{\frac{\partial T(X)}{\partial X}} = \frac{1}{2\theta}$$

Portanto, como $g_\epsilon(\varepsilon) = g_X\left(T^{-1}(\varepsilon)\right)\ |J|$, resulta em:

$$g_\epsilon(\varepsilon) = \frac{1}{2\theta} f_X\left(\frac{\varepsilon+\theta}{2\theta}\right) = \frac{1}{2\theta}\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\left(\frac{\varepsilon+\theta}{2\theta}\right)^{\alpha_1-1}\left(\frac{\theta-\varepsilon}{2\theta}\right)^{\alpha_2-1}$$

A média e a variância de ϵ são dadas por:

$$E(\epsilon) = \frac{\theta(\alpha_1-\alpha_2)}{\alpha_1+\alpha_2} \quad \text{e} \quad \sigma^2 = V(\epsilon) = 4\theta^2 V(X) = \frac{4\theta^2\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+1)}$$

Formas das distribuições utilizadas no estudo

A forma da distribuição $g_\epsilon(\varepsilon)$ que se deseja gerar é determinada pelos valores dos parâmetros α_1 e α_2. Optou-se por simular 3 formas alternativas, conforme abaixo:

$$\begin{aligned}
\alpha_1 < 1 \text{ e } \alpha_2 < 1 \ &\rightarrow \quad \text{forma de U - FU} \\
1 < \alpha_2 < \alpha_1 \ &\rightarrow \quad \text{assimétrica à esquerda - AE, e} \\
1 < \alpha_1 < \alpha_2 \ &\rightarrow \quad \text{assimétrica à direita - AD}
\end{aligned}$$

Entretanto, para que os resultados pudessem ser adequadamente comparados em termos das distribuições amostrais das IR´s dos fatores, os valores θ e $\sigma^2 = V(\epsilon)$ foram escolhidos de modo que houvesse homocedasticidade entre as distribuições FU, AE e AD utilizadas no estudo.

Para que isso ocorresse, utilizou-se $\sigma = 2,8$ para o caso da normal; $\theta = 4$, $\alpha_1 = 0,5$ e

$\alpha_2 = 0,5$ para o caso da forma de U; $\theta = 7$, $\alpha_1 = 2$ e $\alpha_2 = 3$ para o caso da forma assimétrica à direita e $\theta = 7$, $\alpha_1 = 3$ e $\alpha_2 = 2$ para o caso da forma assimétrica à esquerda.

Posteriormente, afim de diminuir a variação nas notas y encontradas, utilizou-se também $\sigma = 0,5$ para o caso normal; $\theta = 0,7$; $\alpha_1 = 0,5$ e $\alpha_2 = 0,5$ para o caso da forma de U, $\theta = 1,25$; $\alpha_1 = 2$ e $\alpha_2 = 3$ para o caso da forma assimétrica à direita e $\theta = 1,25$; $\alpha_1 = 3$ e $\alpha_2 = 2$ para o caso da forma assimétrica à esquerda.

Verificou-se a distribuição da importância relativa de cada fator para cada uma das diferentes distribuições de erros e com cada um dos valores de desvio-padrão. Para essa verificação, foram feitas 100 simulações, considerando 108 consumidores para cada uma das distribuições simuladas. Portanto, para cada distribuição e valor σ temos um conjunto com 36 tratamentos x 108 consumidores x 100 simulações = 388800 dados.

3.1.5 Re-escalonamento das variáveis resposta Y obtidas na simulação

Optou-se por simular valores Y_{jk} iguais a $1, 2, 3, \ldots, 9$, conforme em muitos estudos na área de Tecnologia de Alimentos, nos quais estes são os valores de notas de aceitação fornecidas pelos consumidores participantes de um estudo. Entretanto, a simulação de dados proposta em 3.1.4, com base na transformação de locação e escala da distribuição beta, irá gerar valores aleatórios $Y^*_{jk} = \tau_j + \epsilon_{jk}$.

Portanto, os valores Y_{jk} iguais a $1, 2, 3, \ldots, 9$ foram obtidos mediante uma correção de escala conforme LAW e KELTON (2000) citado por PETERNELLI e MELLO (2007),

$$Y_{jk} = \text{INT}\left\{\left[\frac{Y^* - Y^*_{(1)}}{Y^*_{(n)} - Y^*_{(1)}}\right] \times \text{d} + 1\right\} \qquad \text{d} = Y_{jk}(n) - Y_{jk}(1) = 9 - 1 = 8 \tag{3.3}$$

onde Y_{jk} é a nota gerada por simulação após a correção, um número inteiro $\in \{1, 2, 3, \ldots, 9\}$,

INT ( ) é a função que retorna a parte inteira, Y^* é a nota gerada na simulação antes da corrreção, $Y^*_{(1)}$ é o menor valor e $Y^*_{(n)}$ é o maior valor das notas geradas pela simulação antes da corrreção.

3.2 Apresentação dos resultados da simulação

3.2.1 Normalidade

Para averiguar a normalidade da distribuição amostral do estimador da IR, optou-se por apresentar apenas o teste de Kolmogorov-Smirnov juntamente com a sobreposição da curva normal nos histogramas das IR's para cada fator.

3.2.2 Intervalos de confiança

Intervalos de confiança foram utilizados para verificar se a distribuição amostral da IR é robusta em relação à distribuição dos erros. Verificou-se a sobreposição dos intervalos para os fatores B e C cujos valores de IR de referência foram próximos. Verificou-se também se os intervalos de confiança continham os valores de referência das IR's.

Intervalo de confiança percentil

Apresentaremos intevalos de confiança percentis com $\alpha = 5\%$, os quais são indicados por [LI, LS] = $[\widehat{IR}_{2,5\%}; \widehat{IR}_{97,5\%}]$, onde LI e LS são obtidos, colocando-se os dados em ordem crescente, e excluíndo-se da amostra os $2,5\%$ menores e os $2,5\%$ maiores valores $\widehat{IR}_i$, i=1, 2, 3, $\cdots$, 100, assim, apresentamos os quantis $2,5\%$ e $97,5\%$ como limites do intervalo.

Intervalo de confiança pela aproximação da distribuição normal

Também apresentaremos intervalos de confiança pela aproximação da normal, com os valores de σ^2 substituídos por s^2, que é a variância resultante dos 100 valores estimados para IR_i; assim, tem-se que: $s^2 = \sum_{i=1}^{100} \frac{\widehat{IR}_i - \overline{\widehat{IR}}}{99}$ onde:

$\widehat{IR}_i$ é o i-ésimo valor estimado para IR_i dos fatores A, B, C e D,

$\overline{\widehat{IR}} = \frac{\sum_{i=1}^{100} \widehat{IR}_i}{100}$ é a média amostral.

3.2.3 Erro Médio Relativo

Definiu-se o Erro Médio Relativo (EMR), como

$$EMR = \left| \frac{\overline{\widehat{IR}} - IR}{IR} \right| 100\%$$

onde $\overline{\widehat{IR}}$ é o valor médio das 100 IR's geradas para os fatores A, B C e D em cada distribuição alternativa para o erro aleatório, e IR é o valor de referência utilizado na simulação dos dados.

Capítulo 4

Resultados e Discussão

À seguir são apresentados os intervalos de confiança, histogramas da distribuição amostral das Importâncias Relativas dos fatores, bem como também são apresentadas figuras para ilustrar as formas das distribuições dos erros aleatórios e das respectivas variáveis resposta Y e notas de aceitação. Na apresentação dos resultados utilizamos os símbolos ϵ e σ; e também os termos variável resposta Y e notas de aceitação, para designar o seguinte:

(i) ϵ são os erros aleatórios do modelo da ANCF, simulados por quatro modelos de distribuição, a normal (N) e outros três com uma transformação de locação e escala da distribuição beta: em forma de U (FU), assimétrica à esquerda (AE) e assimétrica à direita (AD).

(ii) σ é o desvio-padrão da distribuição utilizada para simular valores ϵ.

(iii) **variável resposta Y** são os valores gerados pelo modelo da ANCF. São os valores das CP's (Tabela 3.2) somados aos valores ϵ.

(iv) **notas de aceitação (Y) na forma decimal**, são os valores da variável resposta Y re-escalonadas para o intervalo 1 a 9, conforme fórmula 3.3, sem a aplicação do

comando INT. Ou seja, valores com casas decimais.

(v) **notas de aceitação** são as variáveis resposta Y re-escalonadas conforme fórmula 3.3 de modo que sejam valores inteiros de 1 a 9.

São apresentados os resultados tanto dos valores ϵ quanto para as notas de aceitação (Y) para os tratamentos 1 e 36.

4.1 Valores simulados

4.1.1 Erros Aleatórios

As figuras 2.1 e 2.2 apresentam os 10800 valores gerados por simulação para cada uma das quatro formas alternativas de distribuições dos erros aleatórios, com respectivos valores σ, somente para os tratamentos 1 e 36. Na tabela 4.1 são apresentados os respectivos valores da média, desvio padrão, mínimos e máximos para cada gráfico. Verificou-se que com $\sigma = 2,8$ obteve-se erros em um intervalo muito amplo, o que motivou a utilização do $\sigma = 0,5$, para diminuir tal amplitude. Observou-se que os erros foram gerados com as formas desejadas para todos os tratamentos (resultados não apresentados).

4.1.2 Variável resposta Y

Após adicionar os erros ao modelo estatístico 3.1 e com valores betas da tabela 3.2, obtemos variáveis resposta Y, cujas distribuições são apresentadas nas figuras 2.3 e 2.4 para $\sigma = 2,8$ e $\sigma = 0,5$ respectivamente.

Observou-se que os valores Y possuiam a mesma forma de distribuição dos erros para todos os tratamentos. Aprsentamos na tabela 4.2 um resumo descritivo, para todas as dis-

tribuições simuladas e valores σ, sobre a variável resposta Y gerados para os tratamentos 1 e 36.

4.1.3 Notas de aceitação

Nas figuras 2.5 e 2.6 apresentamos para os tratamentos 1 e 36 respectivamente, os valores da variável resposta Y re-escalonados para o intervalo 1 a 9 conforme fórmula proposta por LAW e KELTON (2000), citado por PETERNELLI e MELLO (2007) sem o comando INT, ou seja, notas com casas decimais.

As notas de aceitação apresentadas nas Figuras 2.7 e 2.8, respectivamente para $\sigma = 2,8$ e $\sigma = 0,5$, são valores inteiros de 1 a 9, utilizadas na ANCF.

4.2 Distribuições amostrais das importâncias relativas geradas por simulação

À seguir são apresentados os intervalos de confiança, histogramas e resultados do teste de normalidade (Kolmogorov Smirnov), da distribuição amostral das Importâncias Relativas dos fatores A, B, C e D.

As importâncias relativas para os fatores A, B, C e D obtidas na simulação, possuem distribuição amostral conforme figuras 2.9, 2.10, 2.11 e 2.12 respectivamente. Podemos observar uma boa sobreposição da curva normal sobre os histogramas das distribuição das IR's para todos os fatores, o que nos dá evidências de normalidade. Adicionalmente, as tabelas 4.3 e 4.4 apresentam o teste de normalidade de Kolmogorov-Smirnov para as distribuições

amostrais das IR's geradas por simulação. Adotando-se $\alpha = 5\%$ conclui-se que em todos os casos não se rejeita a hipótese H_0 de que as IR's sigam a distribuição normal.

4.2.1 Intervalos de Confiança

Com a finalidade de verificar a robustez das IR's perante a não-normalidade dos erros aleatórios, construiu-se intervalos de confiança percentis. Esses intervalos foram obtidos tomando-se o intervalo de [5%; 95%] dos quantis dos dados em ordem crescente. As tabelas 4.5 e 4.6 nos mostram esses intervalos de confiança para cada fator, e todas as formas de distribuições alternativas para os erros com $\sigma = 2,8$ e $\sigma = 0,5$ respectivamente.

Observou-se que em todos os casos, o IC incluiu os valores de referência das IR's. Observou-se também que, no caso de $\sigma = 2,8$, como a amplitude dos intervalos de confiança para as IR's foi grande, existe sobreposição de intervalos para as IR´s dos fatores B e C, cujos valores de referência foram $26,55\%$ e $25,66\%$. Ou seja, inferência com base nestes IC´s concluiria erroneamente que B e C possuem IR´s estatisticamente iguais.

Considerando-se que houve indícios de normalidade pelas tabelas 4.3 e 4.4, construiu-se também intervalos de confiança pela distribuição normal, apresentado nas tabelas 4.7 e 4.8. Vale ressaltar que o desvio-padrão utilizado na construção dos intervalos pela aproximação normal foi o obtido com base nos valores IR obtidos nas simulações. Este é um problema para a construção do IC normal na prática, como obter um valor para o desvio-padrão? Uma solução seria obter uma aproximação pelo método delta (Casella e Berger, 2002).

Observou-se que todos os intervalos de confiança incluíram os valores das IR's tomados como referência, exceto para os seguintes casos: (i) intervalo de confiança pela aproximação normal para a simulação de erros com distribuição normal e $\sigma = 2,8$, para os fatores A

e B; (ii) com intervalo pela aproximação normal e $\sigma = 0,5$, (iia) para os fatores A e C com distribuição normal, em forma de U e assimétrica à esquerda; (iib) para o fator B com distribuição em forma de U; e (iic) para o fator D com distribuição normal e em forma de U . Entretanto, neste casos de não inclusão do valor IR de referência nos intervalos, observou-se que o valor estava próximo ao limite do IC, tanto à esquerda quanto à direita. Novamente existe sobreposição de intervalos entre os fatores B e C quando $\sigma = 2,8$, que possuem IR's próximas. Portanto, assim como no IC percentil, novamente não podemos concluir que a IR de $25,66\%$ é estatisticamente diferente da IR de $26,55\%$ quando $\sigma = 2,8$.

4.2.2 Erro Médio Relativo

Para avaliar a acurácia do estimador da IR, definiu-se o Erro Médio Relativo (EMR) dado por

$$EMR = \left| \frac{\overline{\widehat{IR}} - IR}{IR} \right| 100\%.$$

As tabelas 4.9 e 4.10 apresentam os EMR, respectivamente para $\sigma = 2,8$ e $\sigma = 0,5$, de todos os fatores para todas as distribuições alternativas para o erro aleatório.

Um resumo dos resultados é apresentado nas tabelas 4.9 e 4.10. Para $\sigma = 2,8$, com excessão dos EMR do fator D para as formas alternativas dos erros aleatórios normal e em forma de U, iguais a $7,65\%$ e $4,95\%$, respectivamente, todos os demais valores foram inferiores a 2%. Já para $\sigma = 0,5$, Com excessão dos EMR do fator D para as formas alternativas dos erros aleatórios normal e em forma de U, iguais a $4,62\%$ e $2,14\%$, respectivamente, todos os demais valores foram inferiores a $1,5\%$. Estes resultados reforçam as evidências de que os estimadores da IR são robustos à não-normalidade dos erros. Existem evidências ainda de que quanto menor a IR, mais sujeito à variabilidade ela está, pois qualquer alteração pode representar uma variação muito grande em porcentagem (como no caso do fator D para as formas alternativas dos erros aleatórios normal e em forma de U).

Uma questão prática de interesse na ANCF seria estabelecer qual o menor valor de IR estimado para se considerar o fator como relevante, ou seja, para ele não ser descartado de futuros estudos.

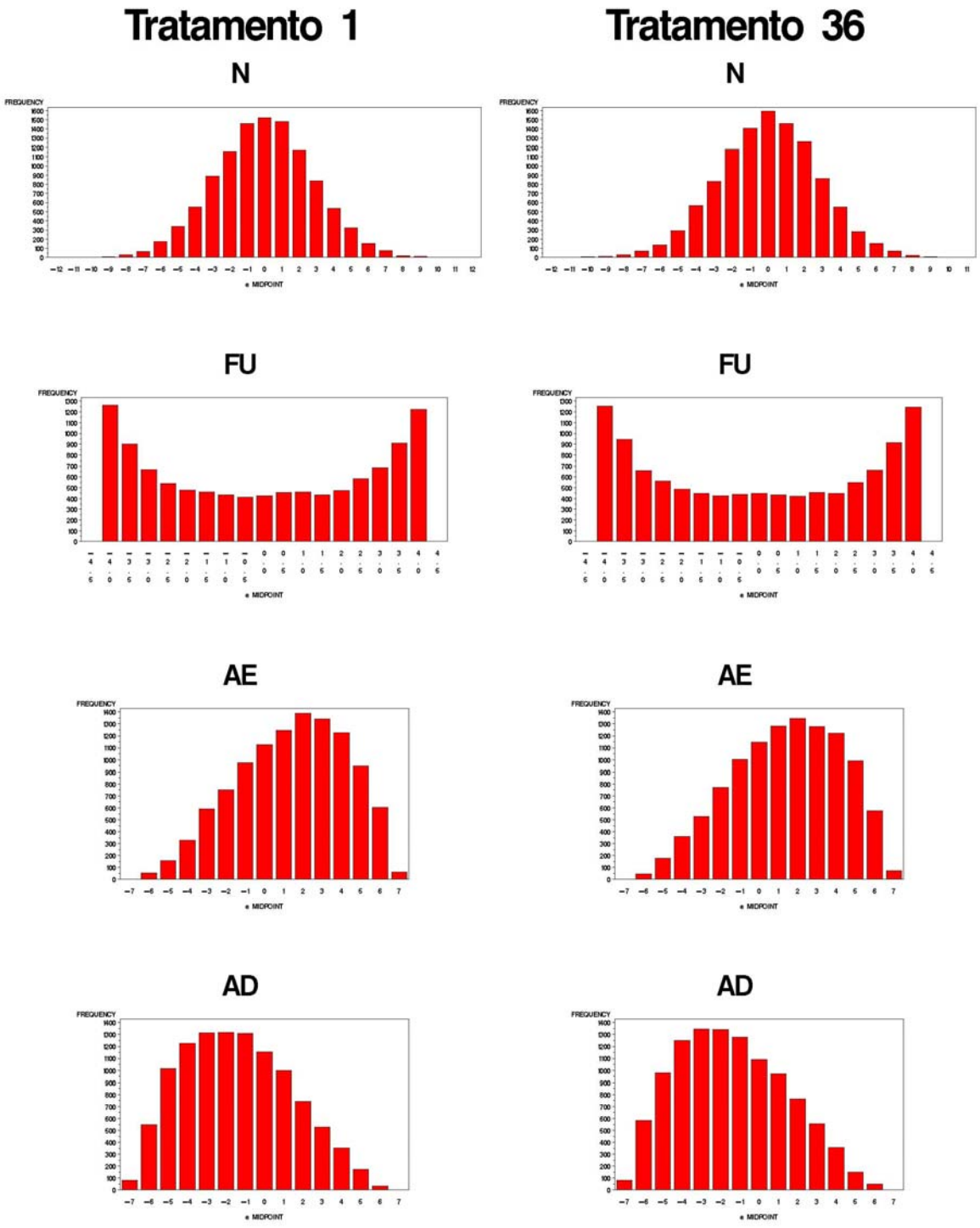

Figura 2.1: Distribuições dos erros aleatórios com $\sigma = 2,8$ para os tratamentos 1 e 36. Modelos normal (N), em forma de U (FU), assimétrico à esquerda (AE) e à direita (AD)

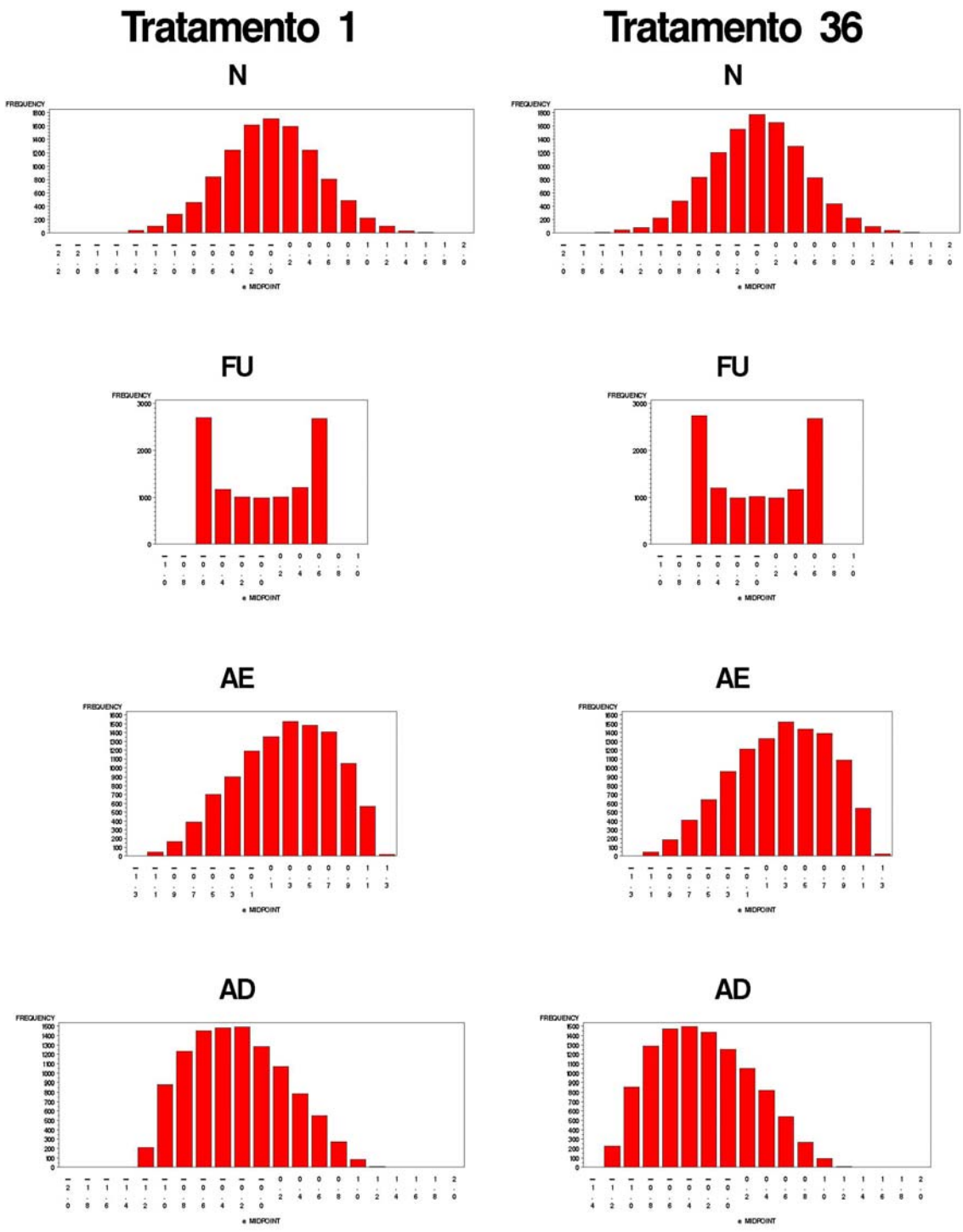

Figura 2.2: Distribuições dos erros aleatórios com $\sigma = 0,5$ para os tratamentos 1 e 36. Modelos normal (N), em forma de U (FU), assimétrico à esquerda (AE) e à direita (AD)

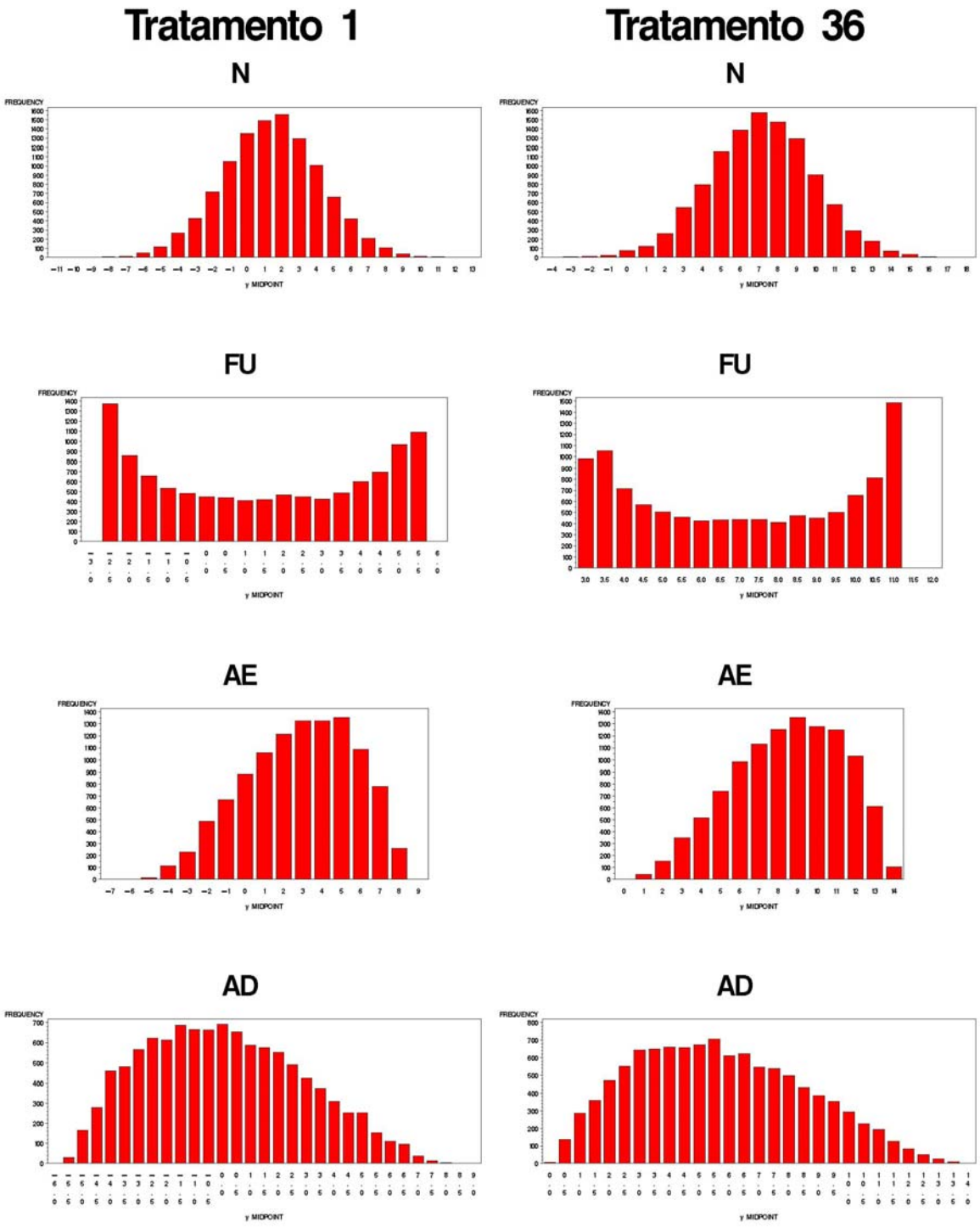

Figura 2.3: Formas das distribuições das variáveis resposta Y para os 10800 valores gerados para o tratamento 1 e 36 com $\sigma = 2,8$

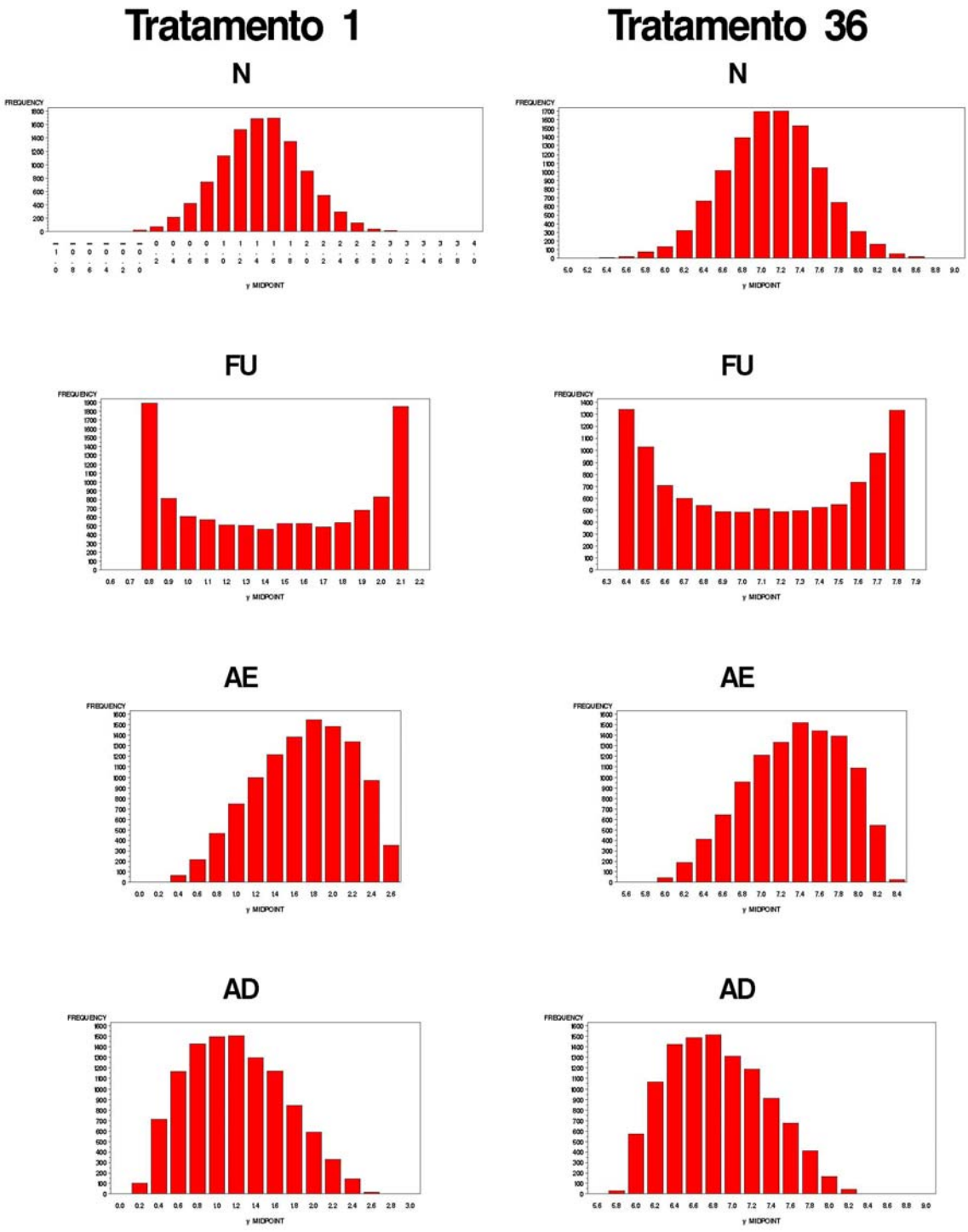

Figura 2.4: Formas das distribuições das variáveis resposta Y para os 10800 valores gerados para o tratamento 1 e 36 com $\sigma = 0,5$

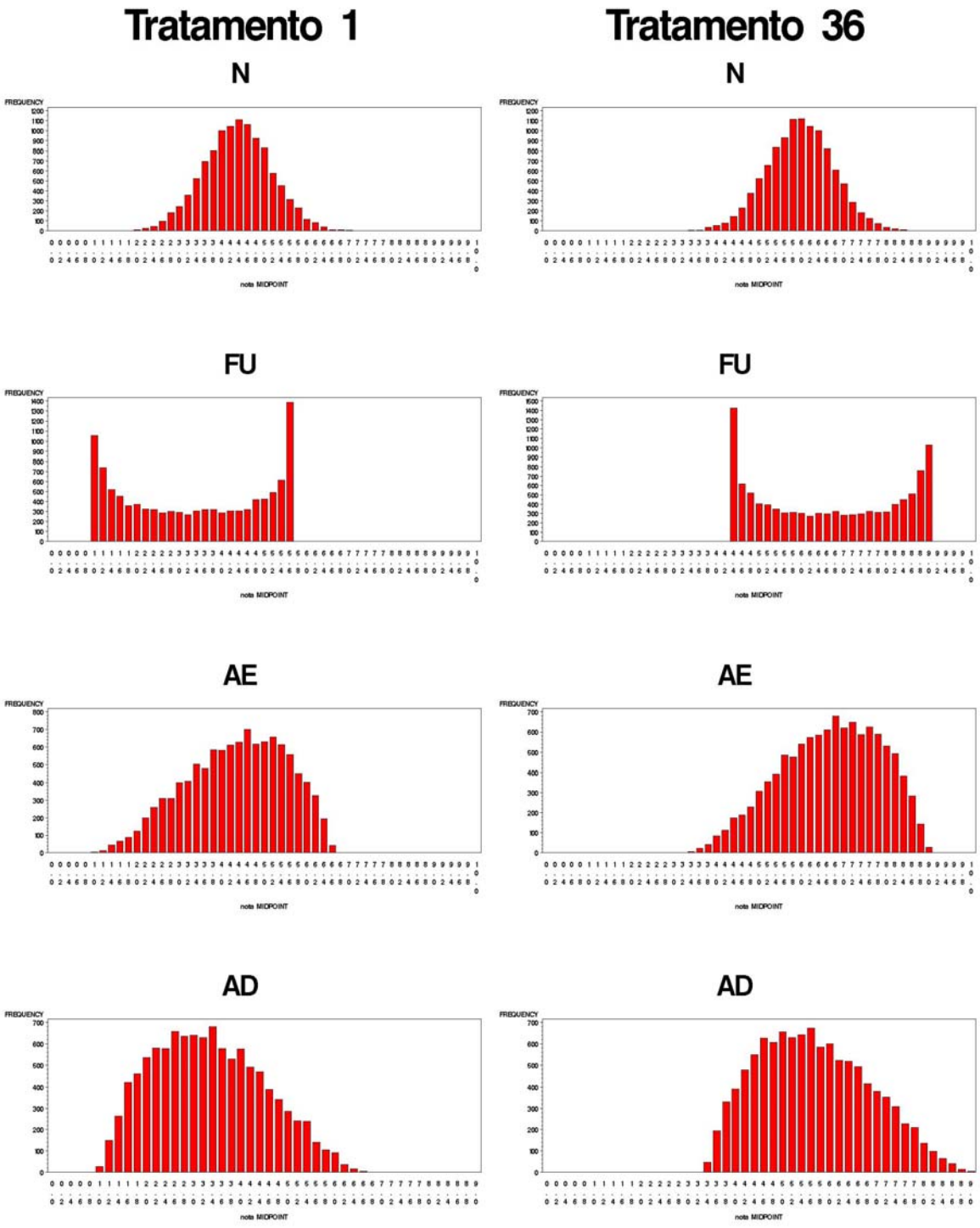

Figura 2.5: Formas das distribuições dos valores das notas de aceitação (Y) na forma decimal para os 10800 valores gerados para o tratamento 1 e 36 com $\sigma = 2,8$

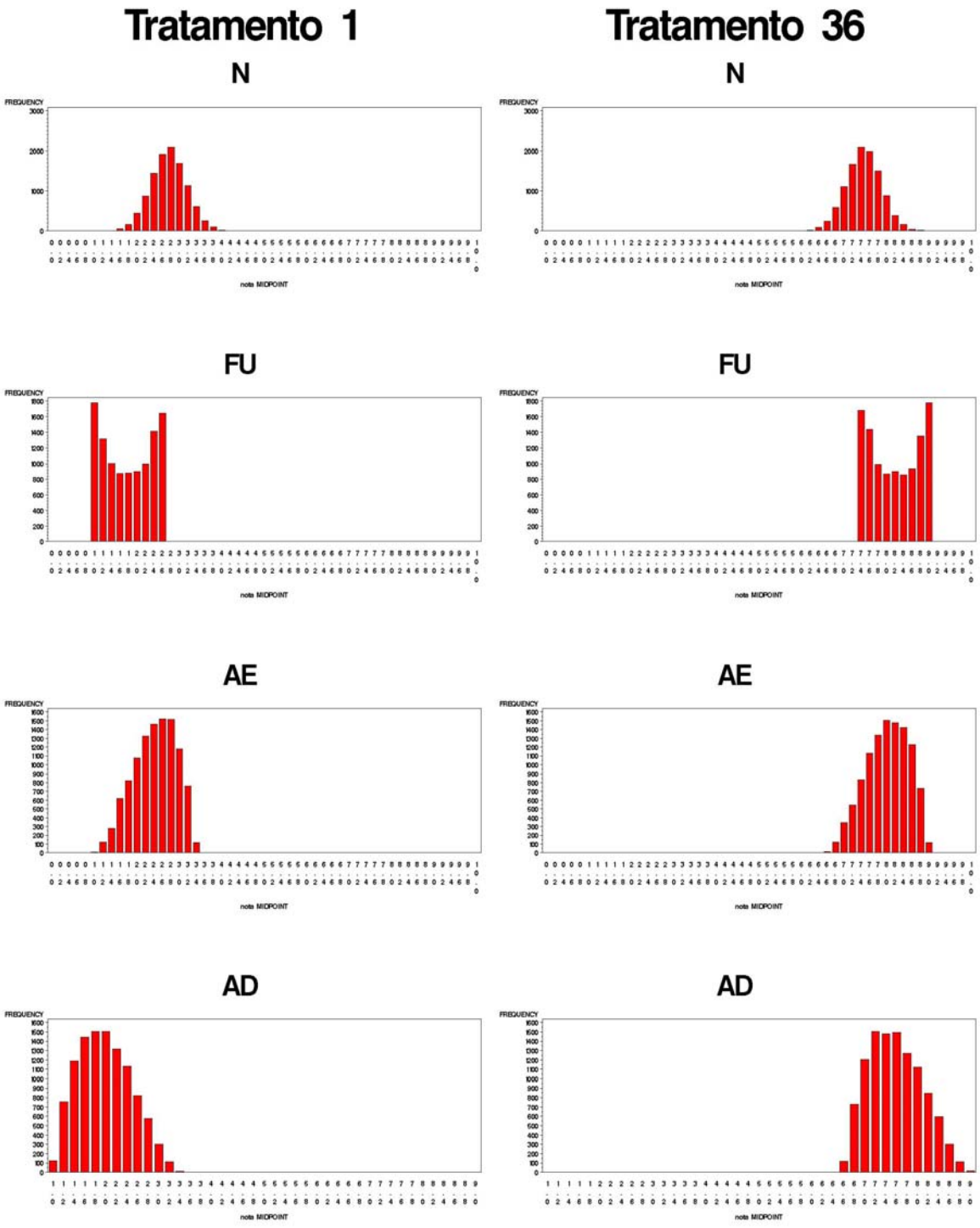

Figura 2.6: Formas das distribuições dos valores das notas de aceitação (Y) na forma decimal para os 10800 valores gerados para o tratamento 1 e 36 com $\sigma = 0,5$

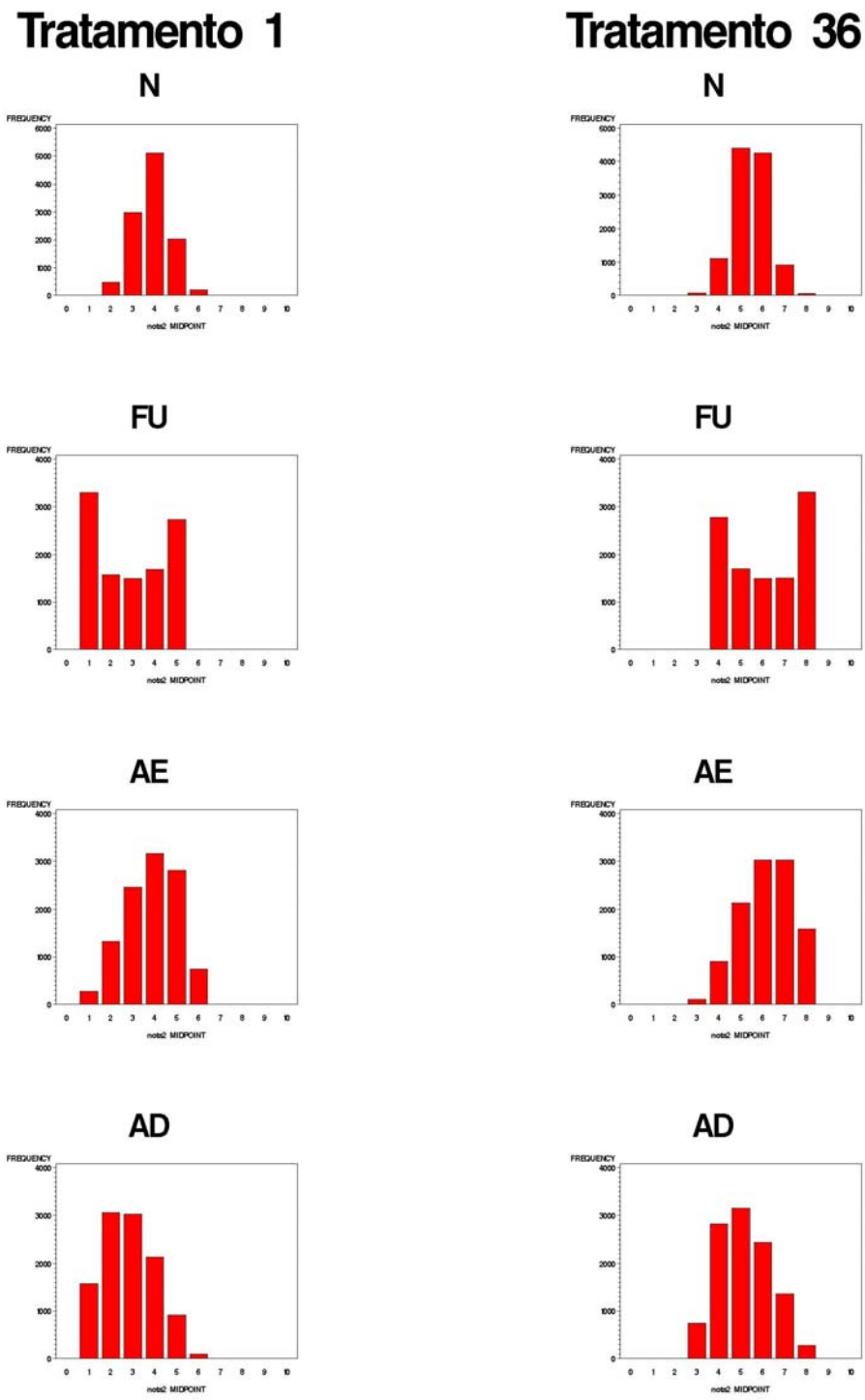

Figura 2.7: Formas das distribuições dos valores das notas de aceitação (Y) para os 10800 valores gerados para o tratamento 1 e 36 com $\sigma = 2,8$

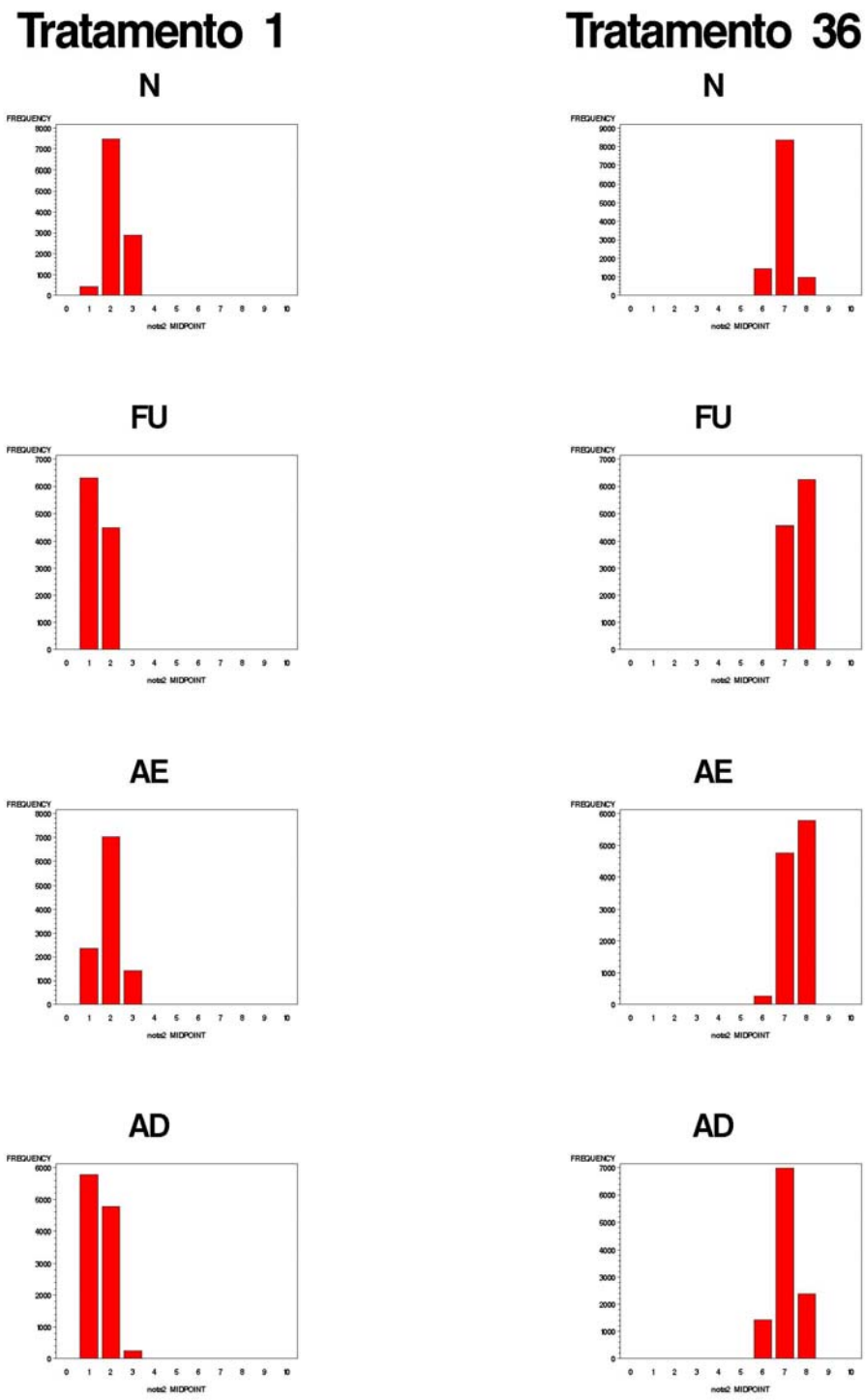

Figura 2.8: Formas das distribuições dos valores das notas de aceitação (Y) para os 10800 valores gerados para o tratamento 1 e 36 com $\sigma = 0,5$

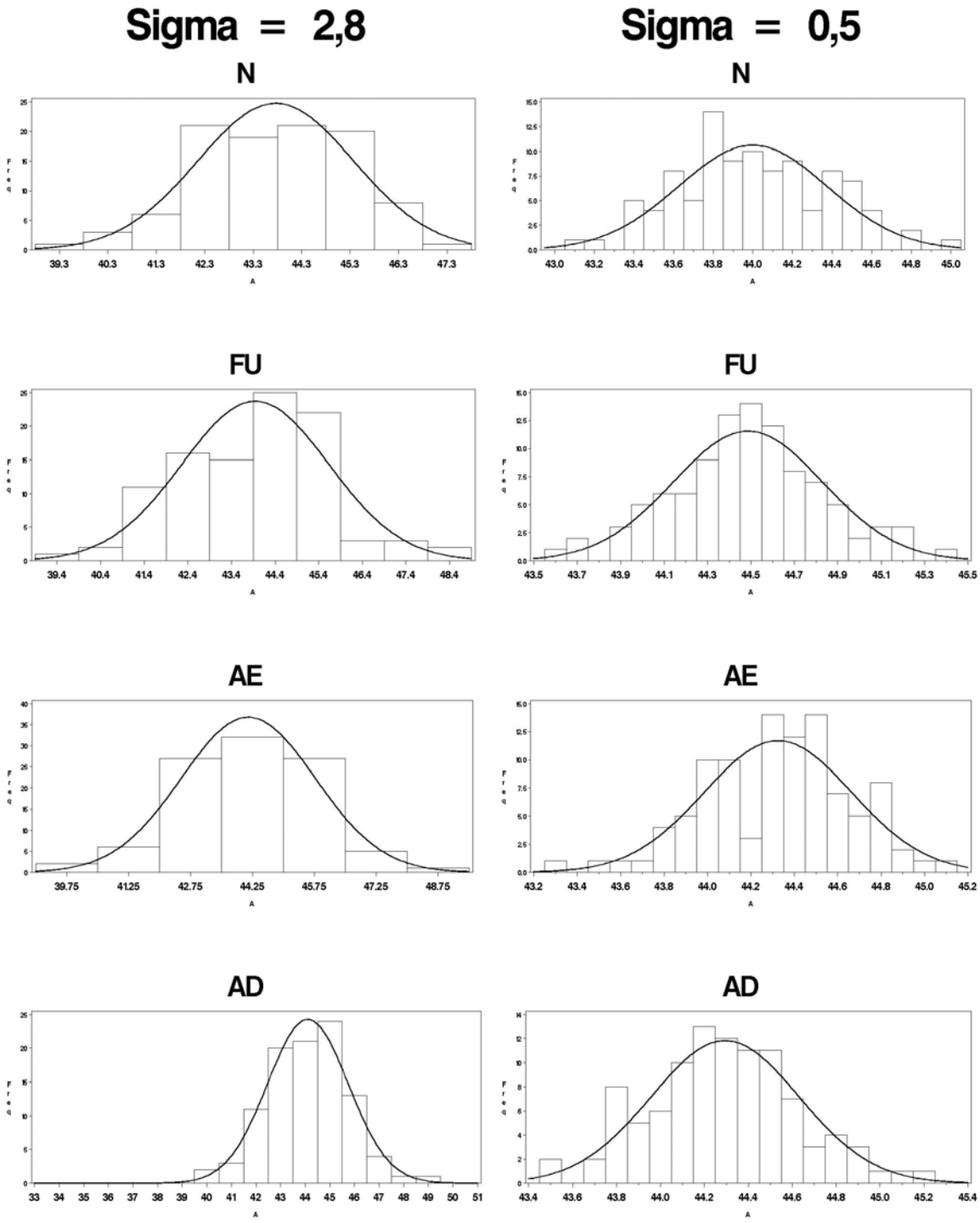

Figura 2.9: Importância relativa do fator A com a presença de erros na forma da distribuição normal (N), em forma de U (FU), assimétrica à esquerda (AE) e assimétrica à direita (AD)

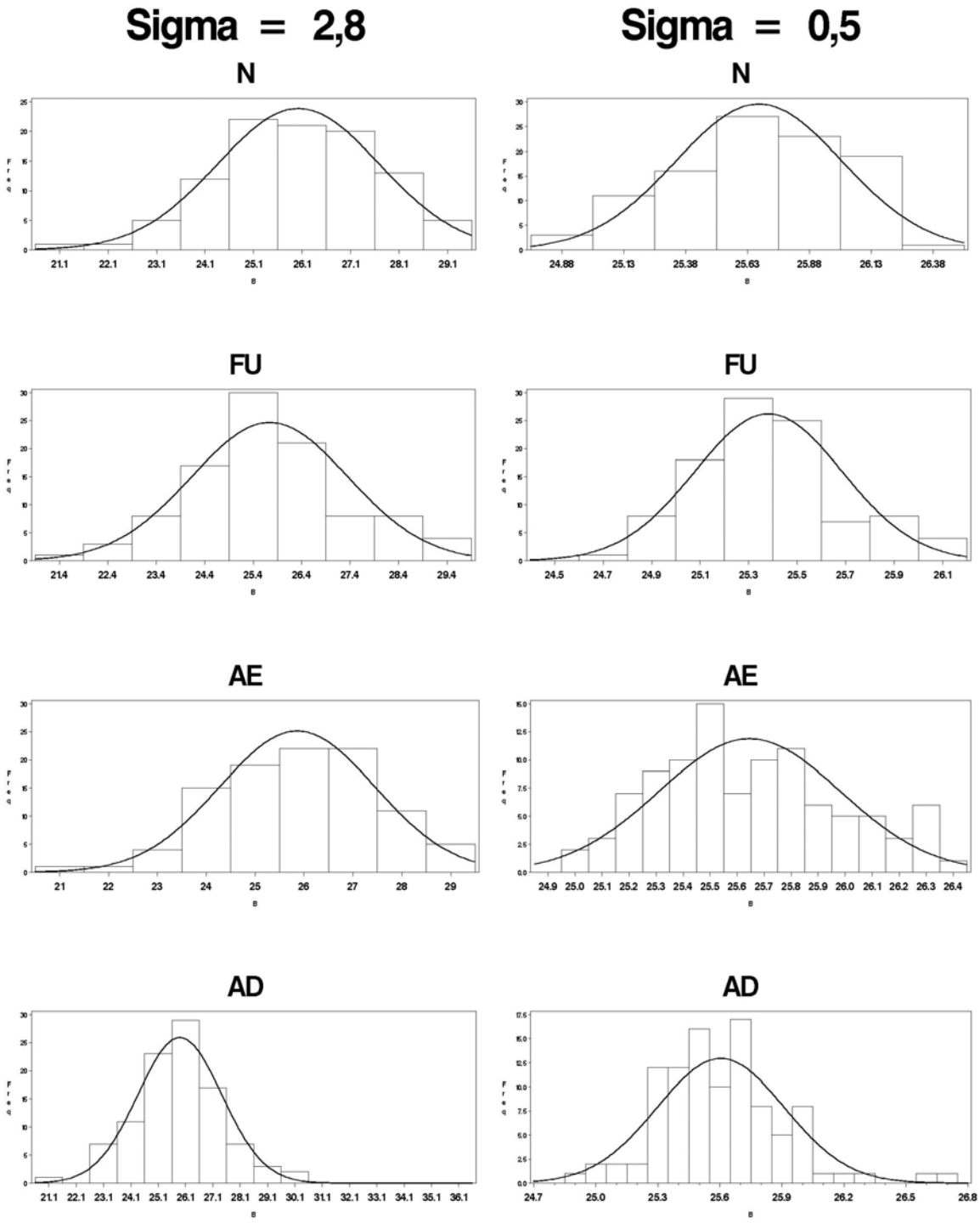

Figura 2.10: Importância relativa do fator B com a presença de erros na forma da distribuição normal (N), em forma de U (FU), assimétrica à esquerda (AE) e assimétrica à direita (AD)

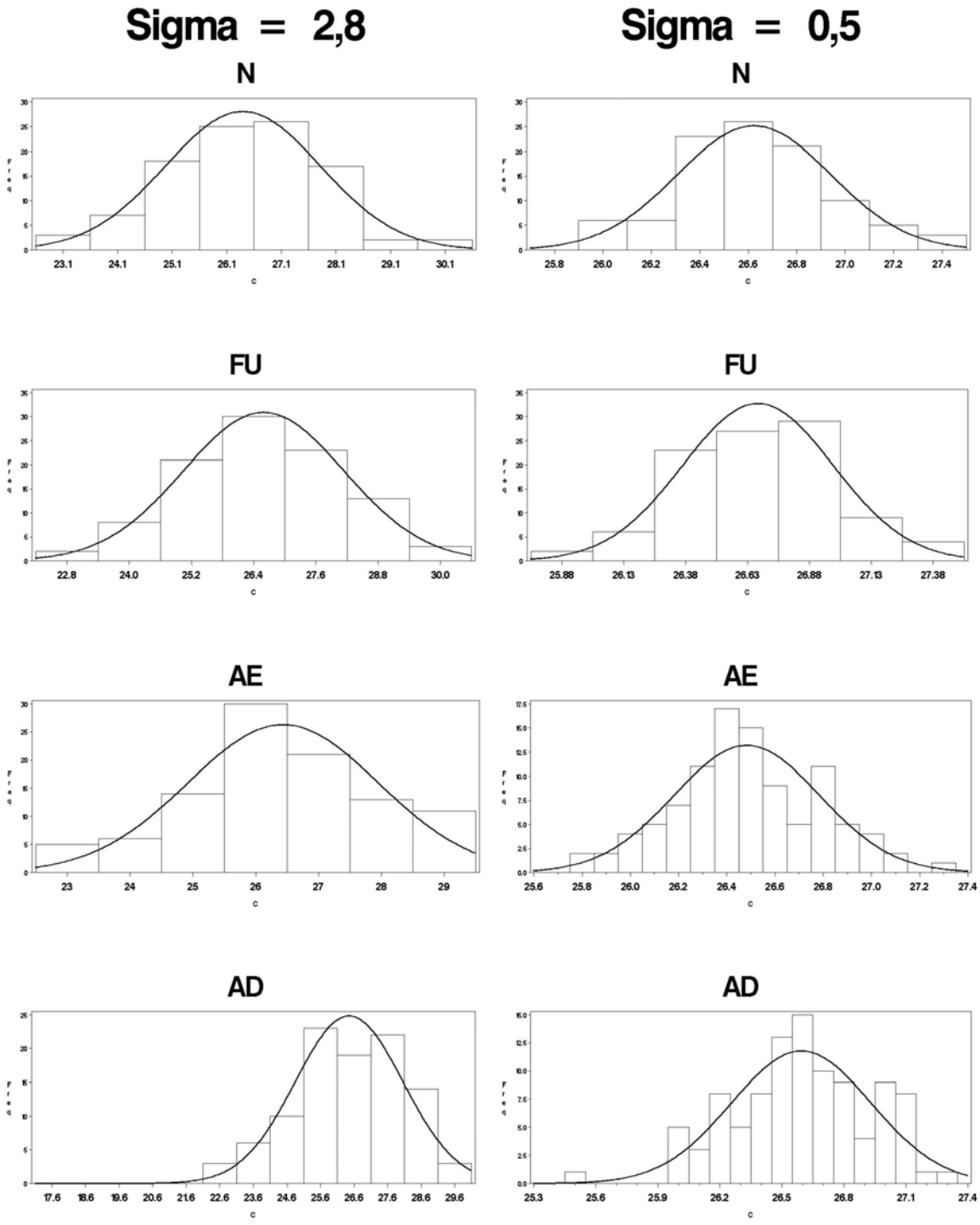

Figura 2.11: Importância relativa do fator C com a presença de erros na forma da distribuição normal (N), em forma de U (FU), assimétrica à esquerda (AE) e assimétrica à direita (AD)

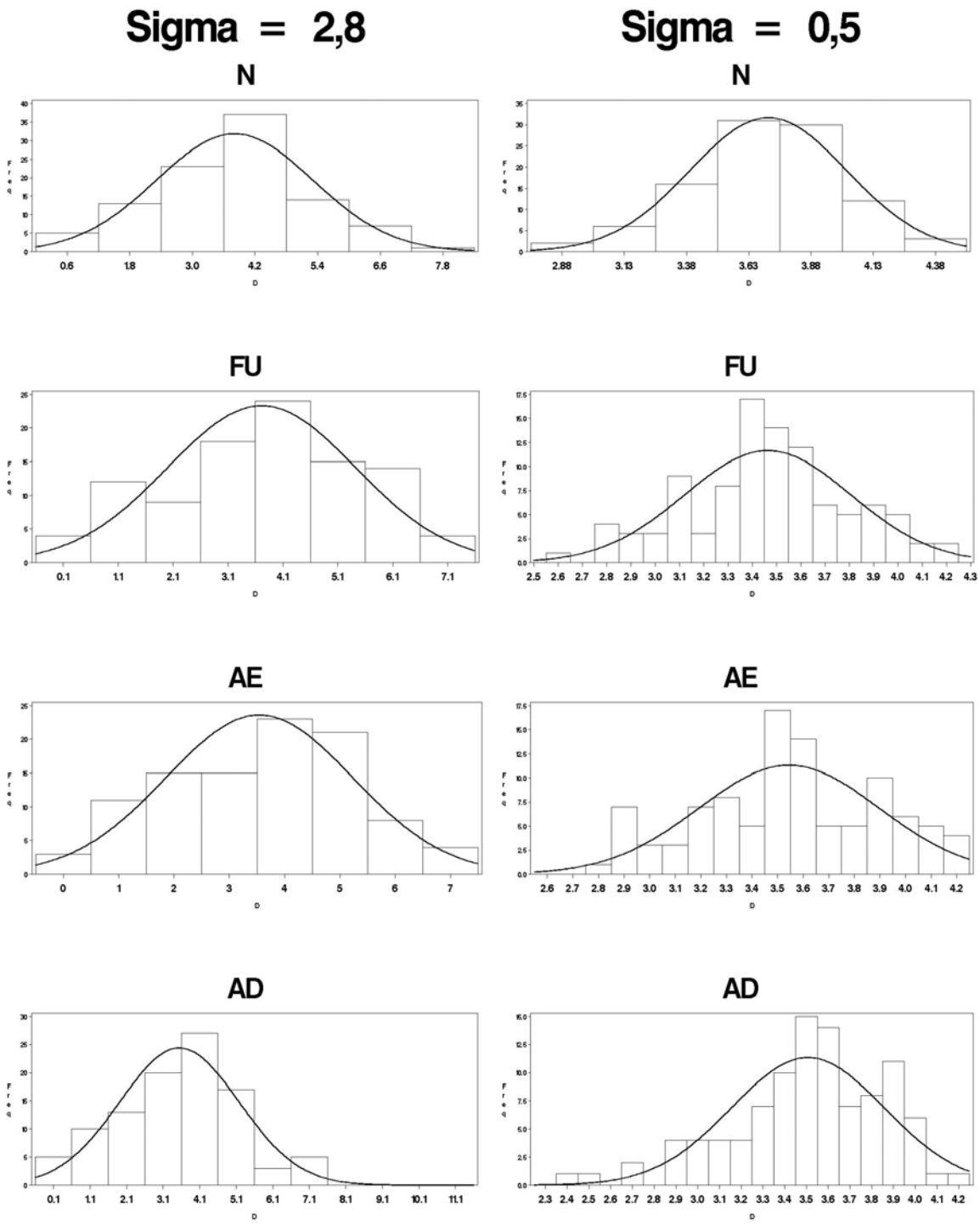

Figura 2.12: Importância relativa do fator D com a presença de erros na forma da distribuição normal (N), em forma de U (FU), assimétrica à esquerda (AE) e assimétrica à direita (AD)

Tabela 4.1: Informações sobre os erros aleatórios gerados para cada tipo de distribuição para os tratamentos 1 e 36

Distribuição	Trat	σ	Média	Desvio Padrão	Mínimo	Máximo
N	1	2,8	-0,03	2,80	-11,76	11,15
		0,5	-0,01	0,50	-2,10	1,99
	36	2,8	0,01	2,77	-11,03	10,38
		0,5	0,00	0,49	-1,97	1,85
FU	1	2,8	0,00	2,84	-4,00	4,00
		0,5	0,00	0,50	-0,70	0,70
	36	2,8	-0,02	2,85	-4,00	4,00
		0,5	-0,01	0,50	-0,70	0,70
AE	1	2,8	1,40	2,81	-6,63	6,95
		0,5	0,25	0,50	-1,18	1,24
	36	2,8	1,37	2,82	-6,61	6,94
		0,5	0,24	0,50	-1,18	1,24
AD	1	2,8	-1,40	2,81	-6,95	6,61
		0,5	-0,25	0,50	-1,24	1,18
	36	2,8	-1,42	2,82	-6,95	6,58
		0,5	-0,25	0,50	-1,24	1,17

Tabela 4.2: Informações sobre as variáveis resposta y geradas para cada tipo de distribuição para os tratamentos 1 e 36

Distribuição	Trat	σ	Média	Desvio Padrão	Mínimo	Máximo
N	1	2,8	1,42	2,80	-10,31	12,60
		0,5	1,44	0,50	-0,65	3,44
	36	2,8	7,11	2,77	-3,93	17,48
		0,5	7,10	0,49	5,13	8,95
FU	1	2,8	1,45	2,84	-2,55	5,45
		0,5	1,45	0,50	0,75	2,15
	36	2,8	7,07	2,85	3,10	11,10
		0,5	7,10	0,50	6,40	7,80
AE	1	2,8	2,85	2,81	-5,18	8,40
		0,5	1,70	0,50	0,27	2,69
	36	2,8	8,47	2,83	0,49	14,04
		0,5	7,34	0,50	5,92	8,34
AD	1	2,8	0,05	2,81	-5,50	8,06
		0,5	1,20	0,50	0,21	2,63
	36	2,8	5,68	2,82	0,15	13,68
		0,5	6,84	0,50	5,86	8,27

Tabela 4.3: Teste de normalidade(Kolmogorov-Smirnov), estatística do teste e p-valor para as IR's de cada fator para cada forma de distribuição dos erros com $\sigma = 2,8$

Fator	Formas	Estatística	p-valor (Pr > D)
A	N	D = 0,05584820	$> 0,150$
	FU	D = 0,05644860	$> 0,150$
	AE	D = 0,05340434	$> 0,150$
	AD	D = 0,08393014	$0,083$
B	N	D = 0,05540334	$> 0,150$
	FU	D = 0,06573097	$> 0,250$
	AE	D = 0,04158116	$> 0,150$
	AD	D = 0,04583226	$> 0,150$
C	N	D = 0,04945570	$> 0,150$
	FU	D = 0,06287521	$> 0,150$
	AE	D = 0,04840710	$> 0,150$
	AD	D = 0,08106178	$0,103$
D	N	D = 0,06403950	$> 0,150$
	FU	D = 0,07339083	$> 0,250$
	AE	D = 0,08608175	$0,068$
	AD	D = 0,05916682	$> 0,150$

Tabela 4.4: Teste de normalidade(Kolmogorov-Smirnov), estatística do teste e p-valor para as IR's de cada fator para cada forma de distribuição dos erros com $\sigma = 0,5$

Fator	Formas	Estatística	p-valor (Pr > D)
A	N	D = 0,05227260	$> 0,150$
	FU	D = 0,05607470	$> 0,150$
	AE	D = 0,05799330	$> 0,150$
	AD	D = 0,0597502	$> 0,150$
B	N	D = 0,07182885	$> 0,150$
	FU	D = 0,08838745	$0,053$
	AE	D = 0,08181350	$0,097$
	AD	D = 0,0604332	$> 0,150$
C	N	D = 0,06797566	$> 0,150$
	FU	D = 0,05673800	$> 0,150$
	AE	D = 0,0632306	$> 0,150$
	AD	D = 0,0494557	$> 0,150$
D	N	D = 0,07083845	$> 0,150$
	FU	D = 0,0603286	$> 0,150$
	AE	D = 0,0616599	$> 0,150$
	AD	D = 0,0818474	$0,096$

Tabela 4.5: Limites inferior e superior dos intervalos de confiança percentis a 95% para as Importâncias Relativas, com os respectivos valores de referência (%), para os fatores A, B, C e D em cada distribuição alternativa para o erro com $\sigma = 2,8$

Fator	Distribuição	Referência(%)	Limite Inferior	Limite Superior
A	N	44,25	42,63	45,01
	FU		42,67	45,05
	AE		43,07	45,28
	AD		43,03	45,19
B	N	25,66	24,88	27,15
	FU		24,80	26,69
	AE		24,94	26,94
	AD		25,02	26,78
C	N	26,55	25,42	27,31
	FU		25,55	27,81
	AE		25,50	27,47
	AD		25,45	27,62
D	N	3,54	2,72	4,60
	FU		2,65	4,75
	AE		2,32	4,73
	AD		2,44	4,60

Tabela 4.6: Limites inferior e superior dos intervalos de confiança percentis a 95% para as Importâncias Relativas, com os respectivos valores de referência (%), para os fatores A, B, C e D em cada distribuição alternativa para o erro com $\sigma = 0,5$

Fator	Distribuição	Referência(%)	Limite Inferior	Limite Superior
A	N	44,25	43,77	44,25
	FU		43,31	44,69
	AE		44,07	44,53
	AD		44,09	44,49
B	N	25,66	25,45	25,95
	FU		25,20	25,52
	AE		25,40	25,87
	AD		25,39	25,80
C	N	26,55	26,42	26,78
	FU		26,43	26,88
	AE		26,30	26,69
	AD		26,38	26,83
D	N	3,54	3,52	3,89
	FU		3,27	3,66
	AE		3,31	3,80
	AD		3,34	3,78

Tabela 4.7: Limites inferior e superior dos intervalos de confiança pela aproximação normal com 95% de confiança ($z = 1,96$) para as Importâncias Relativas, com os respectivos valores de referência (IR%), para os fatores A, B, C e D em cada distribuição alternativa para o erro com $\sigma = 2,8$

Fator	Distribuição	IR%	Média	Desvio Padrão	Limite Inferior	Limite Superior
A	N	44,25	43,77	1,61	43,45	**44,08**$^{/D}$
	FU		43,95	1,68	43,62	44,28
	AE		44,15	1,63	43,83	44,47
	AD		44,11	1,64	43,78	44,43
B	N	25,66	26,03	1,67	**25,70**$^{/E}$	26,35
	FU		25,74	1,62	25,43	26,06
	AE		25,86	1,59	25,55	26,18
	AD		25,89	1,54	25,59	26,20
C	N	26,55	26,40	1,42	26,12	26,68
	FU		26,59	1,55	26,29	26,90
	AE		26,44	1,52	26,14	26,74
	AD		26,45	1,61	26,14	26,77
D	N	3,54	3,81	1,50	3,52	4,11
	FU		3,72	1,71	3,38	4,05
	AE		3,55	1,69	3,21	3,88
	AD		3,55	1,64	3,23	3,87

$^{/D}$ e $^{/E}$ - valor de referência à direita e à esquerda do IC respectivamete

Tabela 4.8: Limites inferior e superior dos intervalos de confiança pela aproximação normal com 95% de confiança ($z = 1,96$) para as Importâncias Relativas, com os respectivos valores de referência (IR%), para os fatores A, B, C e D em cada distribuição alternativa para o erro com $\sigma = 0,5$

Fator	Distribuição	IR%	Média	Desvio Padrão	Limite Inferior	Limite Superior
A	N	44,25	44,00	0,38	43,93	**44,07**$^{/D}$
	FU		44,49	0,35	**44,42**$^{/E}$	44,55
	AE		44,32	0,34	**44,26**$^{/E}$	44,39
	AD		44,29	0,34	44,23	44,36
B	N	25,66	25,67	0,34	25,61	25,74
	FU		25,38	0,30	**25,32**$^{/E}$	25,44
	AE		25,65	0,34	25,58	25,71
	AD		25,60	0,31	25,54	25,66
C	N	26,55	26,62	0,32	**26,56**$^{/E}$	26,69
	FU		26,67	0,30	**26,61**$^{/E}$	26,73
	AE		26,48	0,30	26,42	**26,54**$^{/D}$
	AD		26,59	0,34	26,53	26,66
D	N	3,54	3,70	0,32	**3,64**$^{/E}$	3,77
	FU		3,46	0,34	3,40	**3,53**$^{/D}$
	AE		3,55	0,35	3,48	3,61
	AD		3,51	0,35	3,44	3,58

$^{/D}$ e $^{/E}$ - valor de referência à direita e à esquerda do IC respectivamete

Tabela 4.9: Valores médios das IR estimadas ($\overline{\widehat{IR}\%}$), valores de referência (IR%) e Erro Médio Relativo (EMR), para cada fator com a respectiva distribuição do erro aleatório, para $\sigma = 2,8$

Fator	Distribuição	$\overline{\widehat{IR}\%}$	IR%	EMR%
A	N	43,77	44,25	1,09
	FU	43,95		0,68
	AE	44,15		0,23
	AD	44,11		0,32
B	N	26,03	25,66	1,43
	FU	25,74		0,32
	AE	25,86		0,80
	AD	25,89		0,92
C	N	26,40	26,55	0,58
	FU	26,59		0,17
	AE	26,44		0,42
	AD	26,45		0,37
D	N	3,81	3,54	7,65
	FU	3,72		4,95
	AE	3,55		0,17
	AD	3,55		0,17

Tabela 4.10: Valores médios das IR estimadas ($\overline{\widehat{IR}\%}$), valores de referência (IR%) e Erro Médio Relativo (EMR), para cada fator com a respectiva distribuição do erro aleatório, para $\sigma = 0,5$

Fator	Distribuição	$\overline{\widehat{IR}\%}$	IR%	EMR%
A	N	44,00	44,25	0,56
	FU	44,49		0,53
	AE	44,32		0,17
	AD	44,29		0,10
B	N	25,67	25,66	0,05
	FU	25,38		1,08
	AE	25,65		0,05
	AD	25,60		0,22
C	N	26,62	26,55	0,28
	FU	26,67		0,44
	AE	26,48		0,25
	AD	26,59		0,17
D	N	3,70	3,54	4,62
	FU	3,46		2,14
	AE	3,55		0,15
	AD	3,51		0,93

Capítulo 5

Conclusões

Os intervalos de confiança pela aproximação normal foram mais estreitos do que os percentis e os resultados no geral foram satisfatórios com a inclusão dos valores IR de referência na maioria dos IC ou com a proximidade nos casos da não inclusão. Assim, concluímos então que existe evidências de que o estimador da IR é robusto à variações na forma da distribuição do erro aleatório, visto que em todos os casos de distribuição dos erros aleatórios gerados pela simulação, a impotância relativa teve distribuição que pode ser considerada normal pelos testes de Kolmogov-Smirnov, independente do tipo de distribuição e do valor do desvio-pardrão do erro que foi utilizado na simulação.

Somente alguns IC normais não incluíram o valor paramétrico ou de referência, e mesmo assim estes estavam na adjacência do intervalo. Portanto, esta parece ser uma boa alternativa prática em estudos com a ANCF. Sugerimos então investigar se o método delta pode ser utilizado para se aproximar o desvio-padrão da IR visando-se a construção de IC normais para se comparar os valores de IR's.

Verificou-se que para os fatores B e C, houve sobreposição dos intervalos de confiança tanto para o IC percentil quanto para o IC pela aproximação da normal nos casos onde o

desvio-padrão utilizado foi $\sigma = 2,8$. Isso foi devido à grande amplitude dos erros somados aos valores 'tau' fixados na simulação, que gerou valores de y também com amplitude grande e consequentemente geraram IR's também com grande amplitude. Assim, acabou ocorrendo a sobreposição dos intervalos de confiança para os fatores B e C que foram propositalmente escolhidos próximos. Portanto, nesse contexto, não se pode concluir que na simulação apresentada, os fatores B e C podem ser considerados com importâncias relativas distintas, ou se um fator é mais importante do que o outro, ou ainda se um fator causa maior impacto do que o outro na avaliação do produto (ou serviço) pelo consumidor. Tal fato não ocorreu quando diminuímos o desvio-padrão para $\sigma = 0,5$. Nesse caso, como a amplitude dos erros eram pequenas, esses dados geraram valores de y com amplitude pequena e consquentemente IR's com amplitude também pequenas. Assim, não ocorreu sobreposição de intervalos. Nesse caso específico, podemos considerar sim que as IR's dos fatores B e C são estatísticamente distintas. Nesse contexto, nos aproximamos mais da realidade, já que as notas não se alteram tanto dentro de um mesmo tratamento.

Sugerimos também que se tente obter a fdp da IR pelo método do jacobiano da transformação visando-se a construção de intervalos de confiança exatos, ou ainda via inferência Bayesiana, assim, seria possível obter intervalos de credibilidade.

Referências Bibliográficas

[1] ABADIO, F. D. B. **Efeito de diferentes fatores de informação da embalagem de suco de abacaxi (ananas comosus l. merr) no comportamento do consumidor** 2003. 59 p. Tese (Mestrado em Ciência e Tecnologia de Alimentos) - Universidade Federal Rural do Rio de Janeiro, Seropédica - RJ.

[2] ANDRADE, D. F.; OGLIARI, P. J. **Estatística para as ciências agrárias e biológicas: com noções de experimentação** Florianópolis: Editora da UFSC, 2007. 432 p.

[3] ARTES, R. **Análise de preferência "conjoint analysis".** 1991. 189 p. Tese (Mestrado em Estatística Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo - SP.

[4] CARNEIRO, J. D. S. **Estudo dos fatores da embalagem e do rótulo de cachaça no comportamento dos consumidores.** 2007. 109 p. Tese (Doutorado em Ciência e Tecnologia de Alimentos) - Universidade Federal de Viçosa - Viçosa - MG.

[5] CARNEIRO, J. D. S. **Impacto da embalagem de óleo de soja na intenção de compra do consumidor, via conjoint analysis.** 2002. 80 p. Tese (Mestrado em Ciência e Tecnologia de Alimentos) - Universidade Federal de Viçosa, Viçosa - MG.

[6] CASELLA, G.; BERGER, R. L. **Statistical inference.** California: Editora Duxbury. 2ed, 2002.660 p.

[7] CASTRO, L. R. K. **Valor percebido como ferramenta para tomada de decisão: uma aplicação na indústria hoteleira utilizando a análise conjunta.** 2006. 187 p. Tese (Mestrado em Engenharia de Produção) - Escola de Engenharia de São Carlos, Universidade de São Paulo, São Paulo - SP.

[8] DANTAS, M. I. S. **Impacto da embalagem de couve (Brassica oleraceal cv. acephala) minimamente processada na intenção de compra do consumidor.** 2001. 77p. Tese (Mestrado em Ciência e Tecnologia de Alimentos) - Universidade Federal de Viçosa, Viçosa - MG.

[9] DELLA LUCIA, S. M. **Métodos estatísticos para a avaliação da influência de características não sensoriais no comportamento do consumidor.** 2008. 135 p. Tese (Doutorado em Ciência e Tecnologia de Alimentos) - Universidade Federal de Viçosa, Viçosa - MG.

[10] FERREIRA, D. F. **Estatística básica** Lavras: Editora UFLA, 2005. 664 p.

[11] FONSECA, M. C. P.; SALAY, E. **Opinião de consumidores dos município de Campinas (SP) sobre riscos à saúde provenientes de alimentos**. Disponível em www.unicamp.br\nepa\arquivo_san\Opinio_de_Consumidores _e_riscos_alimentares.pdf, acesso em 15 ago. 2008.

[12] KOHLI, R. Assessing Attribute Significance in Conjoint Analysis: Nonparametric Tests and Empirical Validation **Journal of Marketing Research**, v. 25, no. 2 p. 123-133, 1988.

[13] MAGALHÃES, M. N. **Probabilidade e variáveis aleatórias** São Paulo: Editora da Universidade de São Paulo, 2ed, 2006. 428 p.

[14] MINIM, V. P. R. **Análise sensorial: estudos com consumidores.** Viçosa: Editora UFV, 2006. 225p.

[15] SAS Institute Inc., SAS Technical Report R-109, **Conjoint Analysis Examples**, Cary, NC: SAS Institute Inc., 1993. 85 p.

[16] SILVA, C. H. O. **A proposed framework for establishing optimal genetic designs for estimating narrow-sense heritability.** 2000. 198 p. Dissertation (Doctor of Philosophy in Statistics) North Carolina State University - North Carolina.

[17] SIQUEIRA, J. O. **Mensuração da estrutura de preferência do consumidor: uma aplicação da conjoint analysis em marketing.** 2000. 250 p. Dissertação (Mestrado em Administração) - Universidade de São Paulo, São Paulo - SP.

[18] SPERS, E. E.; SAES, M. S. M.; SOUZA, M. C. M. **Análise das preferências do consumidor brasileiro de café: um estudo exploratório dos mercados de São Paulo e Belo Horizonte**. Disponível em www.rausp.usp.br\download.asp?file=V390153.pdf, acesso em 15 ago. 2008.

[19] TRINDADE, A. L. G; ROTONDARO, R. G. **Modificação da escala de classificação por postos utilizada em Análise Conjunta para aprimorar o modelo obtido por regressão com variáveis dummy.** Revista Produção Online, v. 4, p. 2678-2685, 2004.

Apêndice 1

Comandos do SAS (versão 9.1) utilizados para gerar as IR's de referência, sem a presença de erros aleatórios.

```
options nodate pageno=1;


proc format; *to define factor levels; value Af 1 = 'pouca' 2 =
'medio' 3 = 'muito'; value Bf 1 = 'fraco' 2 = 'medio' 3 = 'forte';
value Cf 1 = 'sim'   2 = 'nao'; value Df 1 = 'baixo' 2 = 'alto';
run;


data a; *codede factor levels; format A Af. B Bf. C Cf. D Df. ;
* do cons=1 to 144; retain trt 0;
do A = 1 to 3; do B = 1 to 3; do C = 1 to 2; do D = 1 to 2;
   trt=trt+1;
output; end; end; end; end;* end; run; proc print data = a; run;


data a3; set a;
   IA1 = (A=1);
  IA2 = (A=2);
  IA3 = (A=3);
```

```
  IB1 = (B=1);
  IB2 = (B=2);
  IB3 = (B=3);
  IC1 = (C=1);
  IC2 = (C=2);
  ID1 = (D=1);
  ID2 = (D=2);
b0=4; b11=-1.2; b12=-0.1; b13=1.3; b21=-0.5; b22=-0.45; b23=0.95;
b31=-0.75; b32=0.75; b41=-0.1; b42=0.1; run; proc print data=a3;
run;

data rates; set a3; y =      b0 + IA1*b11 + IA2*b12 + IA3*b13 +
          IB1*b21 + IB2*b22 + IB3*b23 +
          IC1*b31 + IC2*b32 +
          ID1*b41 + ID2*b42 ;
          run;
          proc print data = rates;
          run;
          /* com isso, os valores de y abaixo foram simulados.
          Para facilitar, apenas copiei e colei os valores
          desses y para rodar o proc transreg, e verificar
os valores das ir's que estamos usando como parâmetros.*/
```

```
proc transreg ;

data teste;

input A B C D y;
 datalines;
 1   1   1   1 1.45
 1   1   1   2 1.65
 1   1   2   1 2.95
 1   1   2   2 3.15
 1   2   1   1 1.5
 1   2   1   2 1.7
 1   2   2   1 3
 1   2   2   2 3.2
 1   3   1   1 2.9
 1   3   1   2 3.1
 1   3   2   1 4.4
 1   3   2   2 4.6
 2   1   1   1 2.55
 2   1   1   2 2.75
 2   1   2   1 4.05
 2   1   2   2 4.25
 2   2   1   1 2.6
```

```
 2   2   1   2 2.8
 2   2   2   1 4.1
 2   2   2   2 4.3
 2   3   1   1 4
 2   3   1   2 4.2
 2   3   2   1 5.5
 2   3   2   2 5.7
 3   1   1   1 3.95
 3   1   1   2 4.15
 3   1   2   1 5.45
 3   1   2   2 5.65
 3   2   1   1 4
 3   2   1   2 4.2
 3   2   2   1 5.5
 3   2   2   2 5.7
 3   3   1   1 5.4
 3   3   1   2 5.6
 3   3   2   1 6.9
 3   3   2   2 7.1 ;

   run;
proc print; run; proc gchart data = teste;
   vbar y / TYPE=FREQ//
```

```
   midpoints=(0 to 9 by 1.5);
            run;

proc capability data=teste; var y;

      histogram y / normal(color=yellow w=3)
                        midpoints= 0 to 9 by 1.5
                         vscale    = count;
   run;

proc transreg data=teste utilities  outtest=res.Utils short
dependent=yhat separators='  '; ods select FitStatistics Utilities;
      title 'IR´s da Conjoint Analysis';
title2 'do estudo por simulacao';
 model identity(y) =    class(A B C D / zero=sum);
 output out=saida;
   run;

   proc print data=res.utils;
   run;

   proc print data=saida;
run;
```

```
   proc corr data=saida; var y Ty; run;


   proc transreg data=teste utilities  short;
*ods select FitStatistics Utilities;
      title 'IR´s da Conjoint Analysis';
title2 'do estudo por simulacao';
 model identity(y) =    class(A B C D / zero=sum);
* output out=saida;
   run;
```

Apêndice 2

Comandos do SAS (versão 9.1) utilizados para simular os erros em forma de U com $\sigma = 0,5$ e gerar suas respectivas IR's.

```
options nodate pageno=1;


libname IR2 'C:\Documents and Settings\OK\Desktop\ULTIMA TESE\nova
tentativa nova\3coorientadores\arquivos sas'; /* Generate basic data
structure starts here*/


proc format; *to define factor levels; value Af 1 = 'pouca' 2 =
'medio' 3 = 'muito'; value Bf 1 = 'fraco' 2 = 'medio' 3 = 'forte';
value Cf 1 = 'sim'   2 = 'nao'; value Df 1 = 'baixo' 2 = 'alto';
run;


data a; *codede factor levels; format A Af. B Bf. C Cf. D Df. ;
* do cons=1 to 144; retain trt 0;
do A = 1 to 3; do B = 1 to 3; do C = 1 to 2; do D = 1 to 2;
   trt=trt+1;
output; end; end; end; end;* end; run;
```

```
proc print data=a; run;


data a2(drop=cons);
  set a;
t = 0.7 ; alf = 0.5 ; bet = 0.5 ; drop t alf bet ;


do simul=1 to 100;
    do cons = 1 to 108;
      u = uniform(3895) ;
     r = betainv(u,alf,bet) ;
         e = 2*t*(r-.5);
      id_cons= compress(simul||A||B||C||D||Cons);
      output;
      end;
end; run;


proc sort data=a2; by simul trt id_cons; run; proc print data =
a2(obs=11664); var simul trt id_cons e; run;


proc means data = a2;by trt; var e; output; run;


goptions htitle=6 ftext=swissb ; proc gchart data =
a2(where=(trt=1)); title ; title2 height=4 'FU';
```

```
   vbar e / TYPE=FREQ
            midpoints=(-1 to 1 by .2);
            run;
proc gchart data = a2(where=(trt=36)); title ; title2 height=4 'FU';
   vbar e / TYPE=FREQ
            midpoints=(-1 to 1 by .2);
            run;
proc capability data = a2; var e; title 'FU'; histogram e
/normal(color=yellow w=3)
                      midpoints= -4 to 4 by .2
                       vscale    = count;
       run;


data a3;
 set a2;
  IA1 = (A=1);
  IA2 = (A=2);
  IA3 = (A=3);
  IB1 = (B=1);
  IB2 = (B=2);
  IB3 = (B=3);
  IC1 = (C=1);
  IC2 = (C=2);
```

```
  ID1 = (D=1);
  ID2 = (D=2);
b0=4; b11=-1.2; b12=-0.1; b13=1.3; b21=-0.5; b22=-0.45; b23=0.95;
b31=-0.75; b32=0.75; b41=-0.1; b42=0.1; run;

proc print data=a3; run;

data rates; set a3; y =      b0 + IA1*b11 + IA2*b12 + IA3*b13 +
         IB1*b21 + IB2*b22 + IB3*b23 +
         IC1*b31 + IC2*b32 +
         ID1*b41 + ID2*b42 + e;
  output;
run;

*=====================================================;
proc means data = rates;*by trt; var y; output; run;

proc gchart data = rates(where=(trt=1));
   vbar y / TYPE=FREQ
             midpoints=(0.6 to 2.2 by 0.1);
            title ;
title2 height=4 'FU';
             run;
```

```
proc gchart data = rates(where=(trt=36));
   vbar y / TYPE=FREQ
            midpoints=(6.3 to 7.9 by 0.1);
            title ;
title2 height=4 'FU';
            run;
*=====================================================;

* corrigir para 1 a 9;

data corrigir;set rates;

nota = (((y-0.7500000) / ( 7.8000000-0.7500000)) * 8) +1; nota2 =
int((((y-0.7500000) / ( 7.8000000-0.7500000)) * 8) +1);

if nota2<1 then nota2=1; if nota2>9 then nota2=9; run;

proc print data=corrigir (where=(trt=1)); var nota2; run;

proc means data = corrigir; var nota2; by trt; run;

proc gchart data = corrigir (where=(trt=1));
   vbar nota / TYPE=FREQ
```

```
            midpoints=(0 to 10 by .2);
            title ;
title2 height=4 'FU';
            run;

proc gchart data = corrigir (where=(trt=36));
   vbar nota / TYPE=FREQ
            midpoints=(0 to 10 by .2);
            title ;
title2 height=4 'FU';
            run;

proc gchart data = corrigir (where=(trt=1));
   vbar nota2 / TYPE=FREQ
            midpoints=(0 to 10 by 1);
             title ;
title2 height=4 'FU';
            run;

proc gchart data = corrigir (where=(trt=36));
   vbar nota2 / TYPE=FREQ
            midpoints=(0 to 10 by 1);
             title ;
```

```
title2 height=4 'FU';
             run;

proc print data = corrigir; run;

proc sort data=corrigir;
  by  simul A B C D;
  run;

data tr;
  set corrigir(keep =  simul A B C D id_cons nota2  trt);
  by  simul A B C D;
  run;

proc print data = tr; title 'basic data structure'; run;

 proc transreg data=tr utilities short outtest=Utils separators='  ';
   by  simul;
      ods select FitStatistics Utilities;
      title 'Conjoint Analysis';
      model linear(nota2) =
            class(A B C D / zero=sum);
          run;
```

```
proc print data=Utils; run;


   *---Gather the Importance Values---;
   data Importance;
      set Utils(keep= simul _depvar_ Importance Label);
      by  simul;
      if n(Importance);
      label = substr(label, 1, index(label, '  '));
   run;


   proc print data=Importance; title '  '; run;


   proc transpose out=Importance2(drop=_:);
      by  simul  _depvar_;
      id Label;
   run;


 data IR2.importance2; set importance2; run;
   proc print data=IR2.Importance2;
      title2 'Importance Values';
   run;
```

```
 proc sort data = IR2.importance2;
by  A;
run; proc print; run;

proc sort data = IR2.importance2;
by  B ;
run; proc print; run;

proc sort data = IR2.importance2;
by  C ;
run; proc print; run;

proc sort data = IR2.importance2;
by  D ;
run; proc print; run;
proc means data=Importance2;
   var A B C D;
   title2 'Average Importance';
run;

goptions htitle=6 ftext=swissb ; axis1 value=(height=1.5); axis2
label=('Freq');
 proc capability data=Importance2;
```

```
 var A ;
  title ;
             title2 height=4 'FU';
      histogram A /nolegend haxis = axis1 vaxis = axis2
 normal(color=yellow w=3)
                        midpoints= 43.5 to 45.5 by 0.1
                         vscale    = count;
       run;


proc capability data=Importance2;
   var B;
     title ;
             title2 height=4 'FU';
      histogram B / nolegend haxis = axis1 vaxis = axis2
normal(color=yellow w=3)
                        midpoints= 24.6 to 26 by 0.1
                         vscale    = count;
    run;


   proc capability data=Importance2;
   var C;
     title ;
             title2 height=4 'FU';
```

```
    histogram C /nolegend haxis = axis1 vaxis = axis2
 normal(color=yellow w=3)
                       midpoints= 25.7 to 27.3 by 0.1
                        vscale    = count;
    run;


   proc capability data=Importance2;
   var D;
     title ;
             title2 height=4 'FU';
      histogram D / nolegend haxis = axis1 vaxis = axis2
normal(color=yellow w=3)
                       midpoints= 2.5 to 4.3 by 0.1
                        vscale    = count;
    run;
```

Apêndice 3

Comandos do SAS (versão 9.1) utilizados para simular os erros normais com $\sigma = 0,5$ e gerar suas respectivas IR's.

```
options nodate pageno=1;


libname IR1 'C:\Documents and Settings\OK\Desktop\ULTIMA TESE\ nova
tentativa nova\3coorientadores\arquivos sas'; /* Generate basic data
structure starts here*/


proc format; *to define factor levels; value Af 1 = 'pouca' 2 =
'medio' 3 = 'muito'; value Bf 1 = 'fraco' 2 = 'medio' 3 = 'forte';
value Cf 1 = 'sim'   2 = 'nao'; value Df 1 = 'baixo' 2 = 'alto';
run;


data a; *codede factor levels; format A Af. B Bf. C Cf. D Df. ;
 retain trt 0;
do A = 1 to 3; do B = 1 to 3; do C = 1 to 2; do D = 1 to 2;
   trt=trt+1;
output; end; end; end; end; run;
```

```
proc print data=a; run;


data a2(drop=cons);
  set a;
  do simul=1 to 100;
    do cons = 1 to 108;
      id_cons= compress(simul||A||B||C||D||Cons);
      output;
    end;
  end;
run;


proc print data=a2; run;


data a3;
 set a2;
  IA1 = (A=1);
  IA2 = (A=2);
  IA3 = (A=3);
  IB1 = (B=1);
  IB2 = (B=2);
  IB3 = (B=3);
  IC1 = (C=1);
```

```
  IC2 = (C=2);
  ID1 = (D=1);
  ID2 = (D=2);
b0=4; b11=-1.2; b12=-0.1; b13=1.3; b21=-0.5; b22=-0.45; b23=0.95;
b31=-0.75; b32=0.75; b41=-0.1; b42=0.1; run;

proc print data=a3; run;

data rates; set a3; e = .5 * rannor(3895); tau= b0 + IA1*b11 +
IA2*b12 + IA3*b13 +
         IB1*b21 + IB2*b22 + IB3*b23 +
         IC1*b31 + IC2*b32 +
         ID1*b41 + ID2*b42;
y =      tau + e; output; run;
*================================================;
proc means data = rates; by trt; var tau e; output; run;

goptions htitle=6 ftext=swissb ; proc gchart data =
rates(where=(trt=1));
   vbar e / TYPE=FREQ
            midpoints=(-2.2 to 2 by .2);
            title  'Tratamento 1';
            title2 height=4 'N';
```

```
            run;
proc gchart data = rates(where=(trt=36));
   vbar e / TYPE=FREQ
            midpoints =(-2 to 2 by .2);
            title  'Tratamento 36';
            title2 height=4 'N';
            run;
*=================================================;


*=====================================================;

proc means data = rates;*by trt; var y; output; run;

proc gchart data = rates(where=(trt=1));
   vbar y / TYPE=FREQ
            midpoints=(-1 to 4 by .2);
            title  'Tratamento 1';
            title2 height=4 'N';
            run;
proc gchart data = rates(where=(trt=36));
   vbar y / TYPE=FREQ
            midpoints=(5 to 9 by .2);
            title  'Tratamento 36';
```

```
            title2 height=4 'N';
            run;
*=========================================================;

* corrigir para 1 a 9;

data corrigir;set rates; nota = (((y-(-0.6502759)) /
(8.9532705-(-0.6502759))) * 8) +1; nota2 = int((((y-(-0.6502759)) /
(8.9532705-(-0.6502759))) * 8) +1);

if nota2<1 then nota2=1; if nota2>9 then nota2=9; run; proc print
data=corrigir (where=(trt=1)); var nota2; run;

proc means data = corrigir; var nota2; by trt; run;

proc gchart data=corrigir(where=(trt=1));
   vbar nota / TYPE=FREQ
   midpoints=(0 to 10 by .2);
   title  'Tratamento 1';
            title2 height=4 'N';
            run;

proc gchart data=corrigir(where=(trt=36));
```

```
   vbar nota / TYPE=FREQ
   midpoints=(0 to 10 by .2);
   title  'Tratamento 36';
            title2 height=4 'N';
            run;

proc gchart data=corrigir(where=(trt=1));
   vbar nota2 / TYPE=FREQ
   midpoints=(0 to 10 by 1);
    title  'Tratamento 1';
            title2 height=4 'N';
            run;

proc gchart data=corrigir(where=(trt=36));
   vbar nota2 / TYPE=FREQ
   midpoints=(0 to 10 by 1);
    title  'Tratamento 36';
            title2 height=4 'N';
            run;

proc print data=corrigir; run;

proc capability data=corrigir;
```

```
  var nota2;
    histogram nota2 /normal(color=yellow w=3)
                        midpoints= 0 to 10 by 1
                         vscale    = count;
        run;

        proc means data = corrigir;
        var nota2;
        output;
        run;

proc sort data=corrigir;
  by  simul A B C D;
  run;

data tr;
  set corrigir(keep = simul A B C D id_cons  nota2 trt);
  by  simul A B C D;
  run;

proc print data = tr; title 'basic data structure'; run;

   proc transreg data=tr utilities short outtest=Utils separators='  ';
```

```
   by simul;
      ods select FitStatistics Utilities;
      title 'Conjoint Analysis';
      model linear(nota2) =
            class(A B C D / zero=sum);
          run;

proc print data=Utils; run;

   *---Gather the Importance Values---;
   data Importance;
      set Utils(keep=simul _depvar_ Importance Label);
      by simul;
      if n(Importance);
      label = substr(label, 1, index(label, '  '));
   run;

   proc print data=Importance; title '  '; run;

   proc transpose out=Importance2(drop=_:);
      by simul  _depvar_;
      id Label;
   run;
```

```
data IR1.importance2; set importance2; run;
  proc print data=IR1.Importance2;
     title2 'Importance Values';
  run;

  proc sort data = IR1.importance2;
  by A ;
  run; proc print; run;

  proc sort data = IR1.importance2;
  by B ;
  run; proc print; run;

  proc sort data = IR1.importance2;
  by C ;
  run; proc print; run;

  proc sort data = IR1.importance2;
  by D ;
  run; proc print; run;
  proc means data=IR1.Importance2;
    var A B C D;
```

```
      title2 'Average Importance';
   run;

goptions htitle=6 ftext=swissb ; axis1 value=(height=1.5); axis2
label=('Freq');
 proc capability data=IR1.Importance2;
 var A;
 title 'Sigma = 0,5';
             title2 height=4 'N';
        histogram A /
nolegend haxis = axis1 vaxis = axis2 normal(color=yellow w=3)
                      midpoints= 43 to 45 by 0.1
                       vscale    = count;
       run;

   proc capability data=IR1.Importance2;
   var B;
       title 'Sigma = 0,5';
             title2 height=4 'N';
      histogram B /
nolegend haxis = axis1 vaxis = axis2 normal(color=yellow w=3)
                      midpoints= 24.6 to 26.3 by 0.1
                       vscale    = count;
```

```
        run;


   proc capability data=IR1.Importance2;
   var C;
    title 'Sigma = 0,5';
              title2 height=4 'N';
            histogram C /
nolegend haxis = axis1 vaxis = axis2 normal(color=yellow w=3)
                        midpoints= 25.8 to 27.2 by 0.1
                         vscale    = count;
        run;


   proc capability data=IR1.Importance2;
   var D;
    title 'Sigma = 0,5';
              title2 height=4 'N';
      histogram D /
nolegend haxis = axis1 vaxis = axis2 normal(color=yellow w=3)
                        midpoints= 2.7 to 4.3 by 0.1
                         vscale    = count;
        run;
```