

UNIVERSIDADE DE BRASÍLIA
DEPARTAMENTO DE ESTATÍSTICA

Lei dos Julgamentos Categóricos de Thurstone em Avaliações Discentes

Patrícia Aguiar Moreira

Brasília - DF

2013

PATRÍCIA AGUIAR MOREIRA - 09/0043863

Lei dos Julgamentos Categóricos de Thurstone em Avaliações Discentes

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

Orientador: *Prof.^o George F. von Borries*

Brasília - DF

2013

Monografia intitulada “**Lei dos Julgamentos Categóricos de Thurstone em Avaliações Discentes**”, defendida por *Patrícia Aguiar Moreira* e que será apresentada no primeiro semestre de 2013, na Universidade de Brasília - Distrito Federal, pela banca examinadora assim constituída:

George Freitas von Borries (Orientador)

PhD. em Estatística, Kansas State University, K.S.U., Estados Unidos, 2008

Departamento de Estatística - UnB

Maria Teresa Leão Costa

Mestre em Estatística, Universidade de Brasília, UnB, Brasil, 1985

Departamento de Estatística - UnB

Eduardo Freitas da Silva

Dr. em Ciências da Saúde, Universidade de Brasília, UnB, Brasil, 1998

Departamento de Estatística - UnB

Brasília - DF

2013

Aos meus pais

Mauro Carlos e Anna Christina.

Ao meu irmão

Mauro Filho.

E aos meus avós

Américo e Joice.

Agradecimentos

Aos meus pais, irmão e a toda minha família que, com muito carinho e apoio, não mediram esforços para que eu chegasse até esta etapa de minha vida.

Ao meu orientador, professor George pela paciência e dedicação que tornaram possível a conclusão desta monografia.

A todos os professores do curso, que foram tão importantes na minha vida acadêmica.

Aos amigos e colegas, pelo companheirismo e apoio constantes.

Ao SAS[®] *Institute* Brasil por possibilitar a utilização desse *software* através de parceria acadêmica com o Departamento de Estatística da UnB.

“Live as if you were to die tomorrow.
Learn as if you were to live forever.”
(Mahatma Gandhi)

Tradução: “Aprenda como se você fosse viver para
sempre. Viva como se você fosse morrer amanhã.”
(Mahatma Gandhi)

Resumo

O presente trabalho tem como objetivo geral, apresentar um visão diferente para andos dados coletados em avaliações discentes. A finalidade é de fornecer informações consistentes e úteis aos professores, de forma que eles possam utilizá-las como ferramenta de auxílio para melhoria da atividade de ensino. O método a ser apresentado é conhecido por Lei dos Julgamentos Categóricos de Thurstone, geralmente utilizado em áreas ligadas à psicologia para análise de julgamentos feitos numa escala não contínua. A teoria apresentada é fortemente baseada no artigo de Souza (2001) [13] e as macros SAS foram desenvolvidas originalmente por Souza e Vales (2002)¹ [14], sendo ajustadas para a versão 9.2 do SAS e adequação ao problema aqui estudado.

Palavras-chave: Julgamentos Categóricos, Thurstone, Avaliações Discentes.

¹ Os autores permitiram o uso destes materiais.

Lista de Tabelas

3.1	Frequência dos Estímulos nas Categorias	12
3.2	Frequências Relativas do Estímulo i na Categoria j	15
3.3	Frequências Relativas Acumuladas do Estímulo i até a Categoria j	16
3.4	Probitos das Frequências Relativas Acumuladas do Estímulo i até a Categoria j	16
3.5	Derivada dos Probitos das Frequências Relativas Acumuladas do Estímulo i até a Categoria j	16
3.6	Contrastes entre os Valores de Escala μ_i	22
5.1	Frequência de Classificação dos Estímulos em cada Categoria da Turma A	29
5.2	Testes de Hipóteses $\beta's = 0$ para a Turma A	30
5.3	Estimação dos Parâmetros para a Turma A	30
5.4	Teste Qui-Quadrado para Bondade do Ajuste para a Turma A	31
5.5	Parâmetros Ordenados pela Intensidade para a Turma A	31
5.6	Frequência de Classificação dos Estímulos em cada Categoria da Turma B	33
5.7	Testes de Hipóteses $\beta's = 0$ para a Turma B	34
5.8	Estimação dos Parâmetros para a Turma B	34
5.9	Teste Qui-Quadrado para Bondade do Ajuste para a Turma B	35
5.10	Parâmetros Ordenados pela Intensidade para a Turma B	35
5.11	Frequência de Classificação dos Estímulos em cada Categoria da Turma C	37
5.12	Testes de Hipóteses $\beta's = 0$ para a Turma C	38
5.13	Estimação dos Parâmetros para a Turma C	38
5.14	Teste Qui-Quadrado para Bondade do Ajuste para a Turma C	39
5.15	Parâmetros Ordenados pela Intensidade para a Turma C	39
5.16	Frequência de Classificação dos Estímulos em cada Categoria da Turma D	41

5.17	Testes de Hipóteses $\beta's = 0$ para a Turma D	42
5.18	Estimação dos Parâmetros para a Turma D	42
5.19	Teste Qui-Quadrado para Bondade do Ajuste para a Turma D	42
5.20	Parâmetros Ordenados pela Intensidade para a Turma D	43
5.21	Frequência de Classificação dos Estímulos em cada Categoria da Turma E	44
5.22	Testes de Hipóteses $\beta's = 0$ para a Turma E	45
5.23	Estimação dos Parâmetros para a Turma E	45
5.24	Teste Qui-Quadrado para Bondade do Ajuste para a Turma E	45
5.25	Parâmetros Ordenados pela Intensidade para a Turma E	46
5.26	Frequência de Classificação dos Estímulos em cada Categoria da Turma F	48
5.27	Testes de Hipóteses $\beta's = 0$ para a Turma F	49
5.28	Estimação dos Parâmetros para a Turma F	49
5.29	Teste Qui-Quadrado para Bondade do Ajuste para a Turma F	50
5.30	Parâmetros Ordenados pela Intensidade para a Turma F	50
5.31	Posição dos Itens por Turma	52

Sumário

Resumo	vi
Lista de Tabelas	vii
1 Introdução	1
2 As Avaliações Discentes	3
2.1 Histórico	3
2.2 Fatores que podem influenciar as avaliações discentes	4
2.3 A Inflação de Escores (“Grade Inflation”)	5
2.4 Análises feitas a partir das avaliações discentes	6
3 A Lei dos Julgamentos Categóricos de Thurstone	8
3.1 O “Continuum Psicológico”	9
3.2 Obtenção do Modelo	10
3.3 Os modelos B, C e D	12
3.4 A estimação dos parâmetros	13
3.4.1 Modelo D	18
4 Coleta dos Dados	23
5 Resultados	25
5.1 Turma A	29
5.1.1 Regressão Logística	29
5.1.2 Lei dos Julgamentos Categóricos de Thurstone	31
5.2 Turma B	33
5.2.1 Regressão Logística	33
5.2.2 Lei dos Julgamentos Categóricos de Thurstone	35
5.3 Turma C	37
5.3.1 Regressão Logística	37
5.3.2 Lei dos Julgamentos Categóricos de Thurstone	38

5.4	Turma D	41
5.4.1	Regressão Logística	41
5.4.2	Lei dos Julgamentos Categóricos de Thurstone	42
5.5	Turma E	44
5.5.1	Regressão Logística	44
5.5.2	Lei dos Julgamentos Categóricos de Thurstone	45
5.6	Turma F	48
5.6.1	Regressão Logística	48
5.6.2	Lei dos Julgamentos Categóricos de Thurstone	50
5.7	Comparação entre as Duas Técnicas	52
6	Conclusão	54
	Referências Bibliográficas	56

1 Introdução

O presente trabalho tem como objetivo geral, apresentar um visão diferente para andos dados coletados em avaliações discentes. A finalidade é de fornecer informações consistentes e úteis aos professores, de forma que eles possam utilizá-las como ferramenta de auxílio para melhoria da atividade de ensino. O método a ser apresentado é conhecido por Lei dos Julgamentos Categóricos de Thurstone, geralmente utilizado em áreas ligadas à psicologia para análise de julgamentos feitos numa escala não contínua. A teoria apresentada é fortemente baseada no artigo de Souza (2001) [13] e as macros SAS foram desenvolvidas originalmente por Souza e Vales (2002)¹ [14], sendo ajustadas para a versão 9.2 do SAS e adequação ao problema aqui estudado.

Para o alcance do objetivo proposto, dividiu-se a dissertação em cinco capítulos. O capítulo que se segue é um resumo do que foi estudado sobre as avaliações discentes para o desenvolvimento deste trabalho. Esse capítulo traz referências sobre estudos já realizados a fim de mostrar alguns problemas relacionados com a utilização dessas avaliações para mensurar o desempenho dos professores.

O segundo capítulo expõe o método a ser aplicado, a Lei dos Julgamentos Categóricos de Thurstone. Ele aborda o tipo de dado que é analisado por esse método, a teoria tomada como base para o desenvolvimento do método, a obtenção do modelo a ser aplicado, os métodos de estimação que podem ser utilizados e os resultados que serão obtidos.

No terceiro capítulo é descrito como se deu a coleta dos dados reais. Em seguida, no quarto capítulo, são apresentados os principais resultados obtidos nas seis turmas que fizeram parte da coleta. Primeiro, tem-se a interpretação do modelo tradicional de regressão logística, e depois, a interpretação da Lei dos Julgamentos categóricos.

¹ Os autores permitiram o uso destes materiais.

Por fim, o quinto capítulo contém a conclusão do trabalho. Este capítulo sintetiza os principais pontos levantados no texto e destaca os resultados considerados mais relevantes.

2 As Avaliações Discentes

Com o intuito de conhecer melhor o universo das avaliações discentes, foi feita uma pesquisa bibliográfica. Os resultados relevantes são apontados neste capítulo e são fundamentais para compreender as vantagens de utilizar a Lei dos Julgamentos Categóricos de Thurstone. Sendo que o foco principal é a utilização desse método para gerar insumos consistentes aos professores para que, a partir desses insumos, possam ter mais subsídios para compreender a visão dos estudantes sobre seu desempenho em atividades de ensino e poder utilizar esta informação para melhoria da qualidade desta atividade.

2.1 Histórico

Ao final das disciplinas cursadas na UnB (e em diversos outros centros de ensino superior), os alunos avaliam diversos itens referentes ao conteúdo da disciplina, ao desempenho do professor, ao seu próprio rendimento enquanto aluno, ao suporte disponível para execução da disciplina, entre outros. Os questionários são computados e uma das ações resultantes é o envio de um resumo, aos professores, contendo informações sobre as avaliações feitas por seus respectivos alunos.

Segundo Loveland (2007) [6], avaliações de ensino feitas por estudantes são ferramentas importantes para o corpo docente. Estas avaliações visam ajudar a aumentar a performance dos docentes e são usadas como base para avaliação da efetividade do ensino em decisões administrativas, por exemplo, promoção (Sheehan e DuPrey, 1999) [11]. Devido a importância que essas avaliações têm para as universidades, vários pesquisadores se dedicaram a estudar os aspectos relacionados a esse instrumento. H. H. Remmers iniciou as primeiras pesquisas nessa área por volta de 1920 e também foi nessa época que algumas universidades como Harvard e University of Texas começaram a introduzir o programa de avaliação discente (Marsh, 1987) [7]. Desde então, essa ferramenta foi ga-

nhando cada vez mais espaço no universo acadêmico. Muitas escolas e universidades têm adotado o uso de avaliações feitas pelos estudantes como uma das medidas da efetividade do ensino (d'Apollonia e Abrami, 1997) [2].

Com tamanho destaque no ambiente acadêmico, é preciso avaliar se esse instrumento é realmente válido para mensurar o que é de interesse das universidades. Entretanto, a avaliação do desempenho em atividades docentes não é uma atividade fácil. Acredita-se que as avaliações discentes possam ser úteis para estimar a performance dos docentes. Esse instrumento será melhor, à medida que satisfaça duas condições. A primeira, que diz respeito ao nível de associação existente entre as avaliações discente e o desempenho do professor. A outra, se refere aos fatores externos que podem influenciar as avaliações. O ideal seria que apenas o desempenho do professor influenciasse as avaliações, porém, estudos indicam que fatores externos à performance do docente também influenciam as avaliações discentes. Alguns destes fatores são destacados a seguir.

2.2 Fatores que podem influenciar as avaliações discentes

Em diversos trabalhos, o foco dos pesquisadores foi estudar a validade das avaliações discentes como instrumento para medir o desempenho do professor. Alguns estudiosos acreditam que o instrumento é válido, outros discordam. Apesar da controvérsia, as instituições de ensino superior continuam se empenhando em obter as avaliações feitas por seus alunos. Isso porque é melhor ter informação das avaliações discentes do que não possuir informação alguma e essas avaliações são pouco onerosas às instituições.

Alguns fatores que parecem influenciar as avaliações discentes são o tipo de disciplina, gênero do professor ou do aluno, tamanho da classe, notas esperadas, experiência ou tipo do professor (substituto, adjunto, titular) e semestre do curso. Este fatores foram abordados em muitos estudos, e tiveram sua influência confirmada.

Preocupados com o viés causado pela influência de fatores externos ao desempenho do professor, alguns estudiosos Seldlmeier (2006) [10], Millea e Grimes (2002) [9], Marsh (1987) [7] alertam para o uso das avaliações discentes na tomada de decisões administrativas resultarem num fenômeno denominado por Inflação de Escores (“Grade Inflation”),

que está associado às notas que os alunos esperam obter no momento em que preenchem as avaliações.

2.3 A Inflação de Escores (“Grade Inflation”)

Alguns autores apontam evidências de que as avaliações discentes possam ser influenciadas pela menção que os alunos supõem que irão obter na disciplina que estão avaliando e, conseqüentemente, cursando. Estudos sugerem que os alunos deem maiores notas aos professores que de certa maneira tendem a aprovar mais alunos, como uma forma de retribuição. Por outro lado, um maior índice de reprovação da classe pode resultar em uma má avaliação do professor, como uma forma dos alunos protestarem contra o rigor de tais docentes (Millea e Grimes, 2002) [9]. Esse fenômeno é conhecido, em inglês, por Grade inflation e denominaremos aqui por Inflação de Escores (Johnson, 2003) [4].

Em Johnson (2003) [4] são apresentadas duas teorias propostas por estudiosos para explicar a relação entre notas e avaliações discentes: a teoria da efetividade do professor (“teacher-effectiveness”), e a teoria da leniência de notas (“grade-lenieny”). A primeira supõe que os estudantes aprendem mais em cursos dados por bons professores. Por aprenderem mais, os alunos recebem notas mais altas e isso resulta na correlação positiva entre as notas e as avaliações discentes. Entretanto, o básico da segunda teoria é de que os alunos fornecem boas avaliações aos professores que lhes dão boas notas. De acordo com essa teoria, a correlação positiva entre as notas médias dos alunos e as avaliações discentes representa um viés. Isso é considerado um viés na medida que o efeito das notas nas avaliações discentes é um fator não relacionado ao desempenho do professor ou ao aprendizado do aluno.

Independente de qual teoria seja mais adequada à realidade das avaliações discentes, o fato de que vários docentes percebem uma correlação positiva entre as notas dadas aos alunos e as avaliações discentes tem conseqüências importantes quando também existe a percepção de que essas avaliações são utilizadas para tomada de decisões administrativas, principalmente para promoções e aumentos salariais (Johnson, 2003) [4]. Uma conseqüência conhecida da conjuntura apresentada é a inflação de escores. Visando aumentos salariais ou promoções, os professores são incentivados a obter boas avaliações

dos seus alunos e para alcançar isso, podem assumir a postura de serem menos rigorrosos ao atribuir notas pelo desempenho de seus alunos.

Dentre os problemas causados pela conjuntura apresentada acima, um dos mais preocupantes talvez seja a dificuldade em separar os alunos que merecem destaque dos demais. Johnson (2003) [4] coloca que a inflação de notas concentra todas as notas no topo, o que dificulta diferenciar os melhores dos muito bons, os muito bons dos bons e os bons dos medíocres. Além disso, ele acrescenta que ao aprovar a mediocridade desencoraja-se a excelência.

2.4 Análises feitas a partir das avaliações discentes

Em geral, o instrumento utilizado para coletar as avaliações discentes consiste em um questionário aplicado ao final das disciplinas. Para o desenvolvimento do presente trabalho foram utilizados dados provenientes da Universidade de Brasília. Nesta universidade, os estudantes são levados a responder diversos itens em relação a, por exemplo, estrutura da sala de aula, perfil do professor, auto avaliação de rendimento do aluno, carga horária, objetivos e formas de avaliação claras, entre outros. A cada item o aluno dava uma nota na escala hedônica de 1 a 10, até o ano de 2011. Em 2012, houve algumas alterações e passou-se a utilizar uma escala de 1 até 5, sendo 1 para os itens avaliados como ruins e 5 para os itens avaliados como bons.

Na Universidade de Brasília, os professores recebem um relatório das avaliações discentes referente a cada turma em que deu aula durante o semestre. Neste relatório são apresentadas algumas medidas descritivas. Como exemplo, a média e desvio padrão das avaliações referentes a cada item. No entanto, tendo em vista a natureza dos dados, o uso dessas estatísticas não é o mais adequado.

Uma vez que a escala de medida é ordinal, ou seja, os rótulos 1, 2, 3, 4 e 5 representando categorias que seguem uma ordem, o uso de estatísticas como média e desvio padrão não são adequadas porque exigem a noção de distância, que é básica na definição de uma medida intervalar. Na escala utilizada, se um aluno marca a opção 4 no item A e a opção 2 no item B, não significa que este aluno considera o quesito A duas vezes melhor que o quesito B. Nesse caso, a noção de distância não existe, os números

não representam quantidades, são apenas rótulos. Sendo assim, medidas como média e desvio padrão não são consideradas adequadas para resumir as informações.

Existem análises estatísticas específicas para serem aplicadas a este tipo de dados (veja, por exemplo, Agresti (2002) [1]).

Thurstone foi precursor de uma técnica conhecida atualmente como Lei dos julgamentos Categóricos de Thurstone e que será o foco do deste trabalho. Essa técnica é tipicamente utilizada em dados de avaliações feitas por juízes selecionados aleatoriamente da população que classificam determinados estímulos em classes ordenadas. No caso em estudo, os alunos são os juízes, os critérios avaliados são os estímulos e as classes ordenadas são representadas pelos valores da escala utilizada no questionário de avaliação. Detalhes da Lei dos julgamentos Categóricos serão abordados na metodologia.

3 A Lei dos Julgamentos Categóricos de Thurstone

Nesta seção será apresentado o método que será utilizado para analisar as avaliações discentes. A teoria original da Lei dos Julgamentos Categóricos foi desenvolvida por Thurstone em 1927 [16]. Souza (2001) [13] apresenta métodos mais eficientes para estimação dos parâmetros do modelo. Souza e Vales (2002) [14] elaboraram um programa em SAS que ajusta o modelo descrito em Souza (2001) através da estimação via mínimos quadrados generalizados. Não é objetivo deste trabalho desenvolver novas técnicas, os trabalhos anteriores fornecem uma boa base metodológica para ser aplicada no estudo das avaliações discentes. A presente seção reproduz em mais detalhes o que já foi desenvolvido nos trabalhos citados acima, com o objetivo de explicar como se chega aos resultados gerados pelo SAS.

De forma geral, a Lei dos Julgamentos Categóricos de Thurstone foi desenvolvida para análise de dados provenientes da seguinte situação: um conjunto de juízes, selecionados aleatoriamente de uma população, avaliam um conjunto de estímulos e classificam cada um deles em uma das categorias apresentadas. Thurstone (1927) [16] formulou um modelo de comportamento mental para explicar o processo de classificação de um dado estímulo em uma das categorias possíveis.

Considere um conjunto de r estímulos representado por $S = \{S_1, \dots, S_r\}$, $r \geq 2$. Considere também, um conjunto de $m \geq 2$ categorias exaustivas e mutuamente excluídas representadas por $C = \{C_1, \dots, C_m\}$, nas quais os r estímulos devem ser alocados. Além disso, as categorias devem ser ordenadas de forma crescente segundo uma característica de interesse. Assim, tem-se $C_1 < C_2 < \dots < C_m$, onde C_1 contém estímulos menos intensos que aqueles pertencentes a C_2 em relação a característica de interesse, C_2 contém

estímulos menos intensos que aqueles pertencentes a C_3 e relação à mesma característica de interesse, e assim sucessivamente.

No caso das avaliações discentes, os alunos assumem o papel de juízes e são responsáveis por classificar cada estímulo em uma das categorias disponíveis. Os estímulos passam a ser os diversos itens do questionário de avaliação discente aplicado nas turmas. E, por fim, as categorias correspondem a escala utilizada para medir a opinião dos estudantes a respeito de cada item.

Os dados oriundos desse tipo de levantamento podem ser organizados em uma tabela de contingência com os estímulos nas linhas e as categorias nas colunas. Segundo Souza (2001) [13], dados gerados por esse processo são conhecidos como dados politômicos com medidas em escala ordinal. O modelo idealizado por Thurstone possibilita derivar uma série de equações que relacionam os parâmetros dos estímulos e das categorias com as frequências da tabela de contingência. Essas equações são conhecidas por Lei dos Julgamentos Categóricos.

3.1 O “Continuum Psicológico”

Thurstone (1927) [16] admite que as avaliações psicológicas que levam os juízes a classificar um estímulo S_i em uma das categorias C_j são efetuadas no que ele designou por continuum psicológico. Quando o juiz é exposto a um estímulo S_i , isso dá início a um processo mental chamado de processo de discriminação modal. Esse mecanismo é latente no juiz, ou seja, é um mecanismo mental inconsciente e conhecido apenas através de suas manifestações perceptíveis. No caso, essas manifestações podem ser observadas através da tabela de contingência que armazena os dados.

Durante o processo de discriminação modal, os estímulos S_i são traduzidos por valores de escala μ_i no continuum psicológico. Da mesma forma, as categorias C_j são traduzidas por uma partição da reta real de forma que a categoria C_1 seja representada pelo intervalo $(-\infty, \tau_1]$, a categoria C_2 seja representada por $(\tau_1, \tau_2]$ e assim sucessivamente até a categoria C_m , representada por $(\tau_{m-1}, +\infty)$. Sendo assim, o continuum psicológico pode ser visto como um conjunto de números reais percebidos mentalmente. Segundo a teoria de Thurstone, um juiz classifica um estímulo S_i em uma das j primeiras categorias se, e somente se, $\mu_i \leq \tau_j$, onde τ_j é o limite superior da categoria j .

Na psicologia, existem duas áreas que estudam o comportamento dos juízes. A psicologia comportamental se refere ao estudo das reações de um mesmo juiz. Enquanto a psicologia diferencial se refere ao estudo das reações de juízes diferentes. O caso das avaliações discentes está inserido no ambiente da segunda área, a psicologia diferencial. Como as reações dos juízes podem ser diferentes e um estímulo avaliado duas vezes por um mesmo juiz também pode gerar valores diferentes no continuum psicológico, é possível a existência de variações nos julgamentos dos juízes. Pela aleatoriedade do processo de escalagem dos estímulos e das categorias no continuum psicológico, é preciso utilizar variáveis aleatórias para representá-los de forma adequada. Os valores de escala μ_i , relacionados aos estímulos, serão representados pela variável ξ_i . Enquanto os valores de localização τ_j relacionados às categorias, serão representados pela variável η_j . Em função disso, seguem os conceitos:

$$\begin{aligned} E[\xi_i] &= \mu_i \\ V[\xi_i] &= \sigma_i^2 \\ \xi_i(\mu_i, \sigma_i^2) & \quad (i = 1, 2, \dots, r) \\ E[\eta_j] &= \tau_j \\ V[\eta_j] &= \phi_j^2 \\ \eta_j(\tau_j, \phi_j^2) & \quad (j = 1, 2, \dots, m-1) \end{aligned}$$

Além disso, são feitas suposições de que os valores de escala ξ_i são independentes e que (ξ_i, η_j) formam uma normal bivariada com coeficiente de correlação igual a ρ_{ij} . A partir desses conceitos o próximo passo será desenvolver um modelo estatístico capaz de estimar esses parâmetros. O interesse principal será na estimação das diferenças $\mu_i - \mu_j$. A partir dessas diferenças, será possível ordenar os estímulos.

3.2 Obtenção do Modelo

Seja agora, π_{ij} a probabilidade de classificar o estímulo i em uma das j primeiras categorias, isto é,

$$\pi_{ij} = P \left\{ S_i \in \bigcup_{l=1}^j C_l \right\} \quad i = 1, \dots, r \quad e \quad j = 1, \dots, m-1 \quad (3.1)$$

De acordo com a teoria de Thurstone,

$$\begin{aligned}\pi_{ij} &= P\{\xi_i \leq \eta_j\} \\ &= P\{\xi_i - \eta_j \leq 0\}\end{aligned}\tag{3.2}$$

Aplicando-se a padronização da distribuição Normal, tem-se

$$\begin{aligned}\pi_{ij} &= P\left\{\frac{(\xi_i - \eta_j) - (\mu_i - \tau_j)}{\sqrt{V(\xi_i - \eta_j)}} \leq -\frac{(\mu_i - \tau_j)}{\sqrt{V(\xi_i - \eta_j)}}\right\} \\ &= P\left\{Z \leq -\frac{(\mu_i - \tau_j)}{\sqrt{V(\xi_i - \eta_j)}}\right\}\end{aligned}\tag{3.3}$$

Seja $g(\pi_{ij})$ a função de transformação da probabilidade π_{ij} para o modelo probit, obtem-se a equação

$$g(\pi_{ij}) = -\frac{(\mu_i - \tau_j)}{\sqrt{V(\xi_i - \eta_j)}} \quad i = 1, \dots, r \quad e \quad j = 1, \dots, m - 1\tag{3.4}$$

onde

$$g(\pi_{ij}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\pi_{ij} - \mu_{ij})^2}{2\sigma^2}\right]\tag{3.5}$$

Existem outras duas opções típicas para transformação da probabilidade π_{ij} . O modelo logit, em que

$$g(\pi_{ij}) = \ln\left\{\frac{\pi_{ij}}{(1 - \pi_{ij})}\right\}\tag{3.6}$$

e o modelo complemento log-log, onde

$$g(\pi_{ij}) = \ln\{-\ln(1 - \pi_{ij})\}\tag{3.7}$$

Na equação (3.4), a probabilidade π_{ij} pode ser estimada através da frequência relativa acumulada em j , denotada por $\hat{\pi}_{ij}$.

Seja \hat{p}_{il} a proporção de vezes em que os n juízes selecionados da população classificaram o estímulo S_i na categoria C_l .

Tabela 3.1: Frequência dos Estímulos nas Categorias

Estímulo	C_1	C_2	\cdots	C_j	\cdots	C_m
S_1	f_{11}	f_{12}	\cdots	f_{1j}	\cdots	f_{1m}
S_2	f_{21}	f_{22}	\cdots	f_{2j}	\cdots	f_{2m}
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
S_i	f_{i1}	f_{i2}	\cdots	f_{ij}	\cdots	f_{im}
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
S_r	f_{r1}	f_{r2}	\cdots	f_{rj}	\cdots	f_{rm}

$$\hat{p}_{ij} = \frac{f_{ij}}{n} \quad (3.8)$$

Assim, a probabilidade do juiz classificar o estímulo i em uma das j primeiras pode ser estimada por

$$\hat{\pi}_{ij} = \hat{p}_{i1} + \hat{p}_{i2} + \cdots + \hat{p}_{ij} \quad (3.9)$$

Ao substituir π_{ij} por $\hat{\pi}_{ij}$ e acrescentar o erro experimental aleatório u_{ij} na equação (3.4), obtem-se o modelo estatístico

$$g(\hat{\pi}_{ij}) = -\frac{(\mu_i - \tau_j)}{\sqrt{V(\xi_i - \eta_j)}} + u_{ij} \quad i = 1, \dots, r \quad e \quad j = 1, \dots, m-1 \quad (3.10)$$

com os vetores $u_i^T = (u_{i1}, \dots, u_{im-1})$ independentes e com matrizes de covariância distintas.

3.3 Os modelos B, C e D

O termo $\sqrt{V(\xi_i - \eta_j)}$, da equação (3.10), pode ser expandido em $\sqrt{\sigma_i^2 + \phi_j^2 - 2\rho_{ij}\sigma_i\phi_j}$. Assim, a equação (3.10) passa a ser

$$E(g(\pi_{ij})) = -\frac{(\mu_i - \tau_j)}{\sqrt{\sigma_i^2 + \phi_j^2 - 2\rho_{ij}\sigma_i\phi_j}} \quad (3.11)$$

Segundo Torgerson (1958) [17] a equação acima é a forma completa da Lei dos Julgamentos Categóricos. No entanto, não há solução para forma completa. Não importa o numero de estímulos ou categorias, sempre haverá mais incógnitas que equações. Para r estímulos classificados em m categorias, tem-se $2r$ parâmetros que representam os estímulos, $2(m-1)$ parâmetros que representam os limites das categorias e $r(m-1)$

parâmetros que representam as correlações entre estímulos e limites das categorias. Somando os parâmetros, tem-se $2(r+m-1) + r(m-1)$ incógnitas. E por outro lado, apenas $r(m-1)$ equações, uma para cada uma das proporções independentes e observáveis.

Dessa forma, torna-se necessário simplificar algumas suposições do modelo. Várias restrições podem ser utilizadas para tornar a equação geral resolvível. Neste trabalho serão apresentadas três delas, todas baseadas em simplificações do termo $\sqrt{\sigma_i^2 + \phi_j^2 - 2\rho_{ij}\sigma_i\phi_j}$. Elas resultam nos modelos B, C e D de Torgerson (1958) [17], e que também podem ser vistos em Souza (2001) [13]. O modelo B supõe que a variância varia apenas com o estímulo, e tem-se o modelo estatístico

$$g(\hat{\pi}_{ij}) = -\frac{\mu_i - \tau_j}{\delta_i} + u_{ij} \quad (3.12)$$

No modelo C, a variância varia apenas com as categorias, e tem-se o modelo

$$g(\hat{\pi}_{ij}) = -\frac{\mu_i - \tau_j}{\delta_j} + u_{ij} \quad (3.13)$$

O modelo D assume que a variância é constante e tem-se

$$g(\hat{\pi}_{ij}) = -(\mu_i - \tau_j) + u_{ij} \quad (3.14)$$

Os dois primeiros modelos são não-lineares, e o último é linear. Esses modelos não são identificáveis e precisam de restrições paramétricas adicionais que serão vistas em detalhe adiante.

3.4 A estimação dos parâmetros

Independente do modelo escolhido, B, C ou D, a estimação dos parâmetros segue uma estrutura semelhante. Para a análise dos dados coletados os três modelos foram aplicados. Apesar do modelo D ser menos geral que os modelos B e C, os três modelos apresentaram resultados muito próximos. Dessa forma, optou-se por aplicar apenas o modelo D, por este ser mais simples. Comparações entre os três modelos podem ser vistas em Souza (2000) [12].

Agora, seguem os procedimentos gerais para estimação via mínimos quadrados generalizados.

Seja o vetor resposta $G(\hat{\boldsymbol{\pi}}) = (G_1^T(\hat{\boldsymbol{\pi}}_1), \dots, G_r^T(\hat{\boldsymbol{\pi}}_r))^T$ com $\hat{\boldsymbol{\pi}} = (\hat{\boldsymbol{\pi}}_1, \dots, \hat{\boldsymbol{\pi}}_r)^T$, onde

$$\mathbf{G}_i(\hat{\boldsymbol{\pi}}_i) = \begin{bmatrix} g(\hat{\boldsymbol{\pi}}_{i1}) \\ g(\hat{\boldsymbol{\pi}}_{i2}) \\ \vdots \\ g(\hat{\boldsymbol{\pi}}_{im-1}) \end{bmatrix} \quad e \quad \boldsymbol{\pi}_i = \begin{bmatrix} \pi_{i1} \\ \pi_{i2} \\ \vdots \\ \pi_{im-1} \end{bmatrix} \quad (3.15)$$

Segue da expansão de Taylor de primeira ordem da função $\mathbf{G}_i(\cdot)$ em torno do verdadeiro parâmetro $\boldsymbol{\pi}_i$, as seguintes equações:

$$\begin{aligned} g(\hat{\boldsymbol{\pi}}_{i1}) &= g(\boldsymbol{\pi}_{i1}) + g'(\boldsymbol{\pi}_{i1})(\hat{p}_{i1} - p_{i1}) \\ g(\hat{\boldsymbol{\pi}}_{i2}) &= g(\boldsymbol{\pi}_{i2}) + g'(\boldsymbol{\pi}_{i2})(\hat{p}_{i1} - p_{i1}) + g'(\boldsymbol{\pi}_{i2})(\hat{p}_{i2} - p_{i2}) \\ &\vdots \\ g(\hat{\boldsymbol{\pi}}_{im-1}) &= g(\boldsymbol{\pi}_{im-1}) + g'(\boldsymbol{\pi}_{im-1})(\hat{p}_{i1} - p_{i1}) + \dots + g'(\boldsymbol{\pi}_{im-1})(\hat{p}_{im-1} - p_{im-1}) \end{aligned} \quad (3.16)$$

Esse procedimento pode ser visto em mais detalhes no Apêndice A (6). Traduzindo em forma de matriz, tem-se

$$\mathbf{G}_i \begin{pmatrix} \hat{\boldsymbol{\pi}}_{i1} \\ \hat{\boldsymbol{\pi}}_{i2} \\ \vdots \\ \hat{\boldsymbol{\pi}}_{im-1} \end{pmatrix} = \mathbf{G}_i \begin{pmatrix} \boldsymbol{\pi}_{i1} \\ \boldsymbol{\pi}_{i2} \\ \vdots \\ \boldsymbol{\pi}_{im-1} \end{pmatrix} + \begin{pmatrix} g'(\boldsymbol{\pi}_{i1}) & 0 & \dots & 0 \\ g'(\boldsymbol{\pi}_{i2}) & g'(\boldsymbol{\pi}_{i2}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g'(\boldsymbol{\pi}_{im-1}) & g'(\boldsymbol{\pi}_{im-1}) & \dots & g'(\boldsymbol{\pi}_{im-1}) \end{pmatrix} \begin{pmatrix} \hat{p}_{i1} - p_{i1} \\ \hat{p}_{i2} - p_{i2} \\ \vdots \\ \hat{p}_{im-1} - p_{im-1} \end{pmatrix} \quad (3.17)$$

De acordo com a transformação escolhida, a função $g'(\boldsymbol{\pi}_{ij})$ pode assumir uma das formas a seguir:

- Probito:

$$g'(\boldsymbol{\pi}_{ij}) = \sqrt{2\pi} \exp \left\{ \frac{g^2(\boldsymbol{\pi}_{ij})}{2} \right\}$$

- Logística:

$$g'(\boldsymbol{\pi}_{ij}) = \frac{1}{\boldsymbol{\pi}_{ij}(1 - \boldsymbol{\pi}_{ij})}$$

- Log-log:

$$g'(\boldsymbol{\pi}_{ij}) = \frac{1}{\boldsymbol{\pi}_{ij} \ln(1 - \boldsymbol{\pi}_{ij})}$$

Represente por \mathbf{H}_i a matriz triangular em (3.17) com os elementos $g'(\hat{\pi}_{ij})$. E seja, ainda, \mathbf{V}_i a matriz de covariância do vetor de frequências relativas $\hat{p}_i^T = (\hat{p}_{i1}, \dots, \hat{p}_{im-1})$ associado ao i -ésimo estímulo. Resulta que o modelo de regressão em (3.10) tem vetor residual com média zero e variância

$$\mathbf{V} = \text{diag}(\mathbf{H}_1 \mathbf{V}_1 \mathbf{H}_1^T, \dots, \mathbf{H}_r \mathbf{V}_r \mathbf{H}_r^T) \quad (3.18)$$

que pode ser estimada por

$$\mathbf{V} = \text{diag}(\hat{\mathbf{H}}_1 \hat{\mathbf{V}}_1 \hat{\mathbf{H}}_1^T, \dots, \hat{\mathbf{H}}_r \hat{\mathbf{V}}_r \hat{\mathbf{H}}_r^T) \quad (3.19)$$

usando as quantidades \hat{p}_{ij} e $\hat{\pi}_{ij}$ no lugar de p_{ij} e π_{ij} , respectivamente.

A estrutura de \mathbf{V} é a mesma para qualquer que seja o modelo utilizado.

Exemplo

Para ilustrar a obtenção da matriz \mathbf{V} , segue um exemplo com dados elaborados artificialmente. Considere uma pesquisa realizada com 100 estudantes selecionados aleatoriamente, que classificaram 4 estímulos, S_1 , S_2 , S_3 e S_4 , em uma das 4 categorias C_1 , C_2 , C_3 ou C_4 , ordenadas segundo uma característica de interesse qualquer. A categoria C_1 está associada aos estímulos menos intensos, enquanto a categoria C_4 está associada aos estímulos mais intensos. As frequências relativas dos estímulos em cada categoria pode ser vista na seguinte tabela de contingência

Tabela 3.2: Frequências Relativas do Estímulo i na Categoria j

Estímulo	C_1	C_2	C_3	C_4
S_1	0,05	0,10	0,15	0,70
S_2	0,10	0,15	0,25	0,50
S_3	0,65	0,20	0,10	0,05
S_4	0,40	0,30	0,20	0,10

A seguir obteve-se as frequências relativas e as acumulou em relação às categorias.

Tabela 3.3: Frequências Relativas Acumuladas do Estímulo i até a Categoria j

Estímulo	C_1	C_2	C_3
S_1	0,05	0,15	0,30
S_2	0,10	0,25	0,50
S_3	0,65	0,85	0,95
S_4	0,40	0,70	0,90

Sob a transformação probito, tem-se

Tabela 3.4: Probitos das Frequências Relativas Acumuladas do Estímulo i até a Categoria j

Estímulo	C_1	C_2	C_3
S_1	-1,6448536	-1,0364334	-0,5244005
S_2	-1,2815516	-0,6744898	0,0000000
S_3	0,3853205	1,0364334	1,6448536
S_4	-0,2533471	0,5244005	1,2815516

E por último, a derivada da função de transformação

Tabela 3.5: Derivada dos Probitos das Frequências Relativas Acumuladas do Estímulo i até a Categoria j

Estímulo	C_1	C_2	C_3
S_1	9,695969	4,288923	2,876104
S_2	5,698060	3,146865	2,506628
S_3	2,699791	4,288923	9,695969
S_4	2,588377	2,876104	5,698060

Sabe-se que $\mathbf{V} = \text{diag}(\mathbf{H}_1 \mathbf{V}_1 \mathbf{H}_1^T, \dots, \mathbf{H}_r \mathbf{V}_r \mathbf{H}_r^T)$. Portanto, é preciso calcular \mathbf{V}_1 e \mathbf{H}_1 . Tem-se que \mathbf{V}_1 é a matriz de covariância do vetor de frequências relativas \hat{p}_1 associado ao estímulo S_1 , dada por

$$\begin{aligned}
\mathbf{V}_1 &= Cov(\hat{\boldsymbol{\rho}}_1) = Cov \begin{pmatrix} \hat{\rho}_{11} \\ \hat{\rho}_{12} \\ \hat{\rho}_{13} \end{pmatrix} = Cov \begin{pmatrix} 0,05 \\ 0,15 \\ 0,30 \end{pmatrix} \\
&= \left(\begin{pmatrix} 0,05 & 0 & 0 \\ 0 & 0,15 & 0 \\ 0 & 0 & 0,30 \end{pmatrix} - \begin{pmatrix} 0,05 \\ 0,15 \\ 0,30 \end{pmatrix} x \begin{pmatrix} 0,05 & 0,15 & 0,30 \end{pmatrix} \right) x \frac{1}{100} \\
&= \begin{pmatrix} 0,000475 & -0,000075 & -0,00015 \\ -0,000075 & 0,001275 & -0,00045 \\ -0,000150 & -0,000450 & 0,00210 \end{pmatrix}
\end{aligned}$$

A estrutura da matriz \mathbf{H}_1 está definida na equação (3.17), é a matriz triangular com os elementos $g'(\hat{\boldsymbol{\rho}}_{ij})$. Seus elementos encontram-se na tabela (3.5).

$$\mathbf{H}_1 = \begin{pmatrix} 9,695969 & 0 & 0 \\ 4,288923 & 4,288923 & 0 \\ 2,876104 & 2,876104 & 2,876104 \end{pmatrix}$$

Assim,

$$\mathbf{H}_1 \mathbf{V}_1 \mathbf{H}_1^T = \begin{pmatrix} 0,044655615 & 0,016634106 & 0,006971653 \\ 0,016634106 & 0,02943177 & 0,01233539 \\ 0,006971653 & 0,01233539 & 0,020679931 \end{pmatrix}$$

Por fim, a matriz V é contruída a partir das quatro matrizes $\mathbf{H}_i \mathbf{V}_i \mathbf{H}_i^T$

$$\mathbf{V} = \begin{pmatrix} 0,045 & 0,017 & 0,007 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,017 & 0,029 & 0,012 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0,007 & 0,012 & 0,021 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,013 & 0,003 & -0,005 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,003 & 0,008 & -0,008 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0,005 & -0,008 & -0,014 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0,027 & -0,107 & -0,404 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0,107 & -0,377 & -1,445 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0,404 & -1,445 & -5,173 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0,007 & -0,029 & -0,111 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0,029 & -0,084 & -0,328 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0,111 & -0,328 & -1,136 \end{pmatrix}$$

Visto isso, segue a estimação dos parâmetros para o modelo D. Os modelos B e C não foram utilizados neste trabalho e, por isso, não serão apresentados em detalhe.

3.4.1 Modelo D

O modelo D é baseado na suposição de que os termos σ_i , ϕ_j e ρ_{ij} são constantes para todos os valores de i e j . Assim, a equação na forma completa (3.11) se reduz a

$$E(g(\pi_{ij})) = \tau_j - \mu_i \quad (3.20)$$

Em forma matricial, tem-se

$$E(\mathbf{G}) = \mathbf{X}\boldsymbol{\beta} \quad (3.21)$$

onde

$$\begin{pmatrix} E(g(\hat{\pi}_{11})) \\ E(g(\hat{\pi}_{12})) \\ \vdots \\ E(g(\hat{\pi}_{1(m-1)})) \\ \vdots \\ E(g(\hat{\pi}_{r1})) \\ E(g(\hat{\pi}_{r2})) \\ \vdots \\ E(g(\hat{\pi}_{r(m-1)})) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & -1 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & -1 \end{pmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_{m-1} \\ \mu_1 \\ \mu_2 \\ \vdots \\ \mu_r \end{pmatrix}$$

Este modelo assemelha-se ao modelo de Análise de Variância, ANOVA. A matriz \mathbf{X} pode ser vista como uma matriz de delineamento. Da mesma forma como ocorre na ANOVA, o modelo não é identificável porque essa matriz não possui posto completo. Como o interesse é nas comparações entre os μ_i 's, não há problema ao impor a restrição $\Sigma\mu_i = 0$ ou igualar um dos μ_i 's a zero. Sob qualquer uma dessas restrições adicionais, o modelo torna-se estimável. Desse modo, apesar de não ser possível estimar os valores μ_i 's, todos os contrastes entre os μ_i 's são estimáveis.

3.4.1.1 Estimabilidade

Supondo um caso hipotético com $m = 3$ categorias e $r = 2$ estímulos. Este modelo pode ser representado em forma matricial por

$$\begin{pmatrix} E(g(\hat{\pi}_{11})) \\ E(g(\hat{\pi}_{12})) \\ E(g(\hat{\pi}_{21})) \\ E(g(\hat{\pi}_{22})) \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \\ \mu_1 \\ \mu_2 \end{pmatrix}$$

A partir dos dados coletados, que no caso são fictícios, é possível construir as equações:

$$\tau_1 - \mu_1 = 4$$

$$\tau_2 - \mu_1 = 6$$

$$\tau_1 - \mu_2 = 8$$

$$\tau_2 - \mu_2 = 10$$

Como o posto da matriz X é igual a 3, ou seja, apenas 3 linhas são linearmente independentes, o modelo é não identificável. Para que a estimação dos parâmetros seja viável, é necessário impor alguma restrição. Duas soluções são possíveis.

Set to zero:

$$\begin{aligned}\mu_1 = 0 &\Rightarrow \tau_1 = 4 \\ \mu_2 &= -4 \\ \tau_2 &= 6\end{aligned}$$

$$\begin{aligned}\mu_2 = 0 &\Rightarrow \tau_2 = 10 \\ \mu_1 &= 4 \\ \tau_1 &= 8\end{aligned}$$

Sum to zero:

$$\begin{aligned}\mu_1 + \mu_2 = 0 &\Rightarrow \tau_2 = 8 \\ \mu_1 &= 2 \\ \tau_2 &= 6 \\ \mu_2 &= -2\end{aligned}$$

Nas três soluções apresentadas acima os valores dos parâmetros são diferentes, e portanto, não estimáveis. Mas note que $\mu_1 - \mu_2 = 4$ em todas as situações. Por isso, se diz que apesar dos parâmetros não serem estimáveis, os contrastes são.

Continuando a estimação para o modelo D, se \mathbf{X} representa a matriz de delineamento reduzida por alguma restrição paramétrica de modo que fique não-singular, o estimador de mínimos quadrados generalizados do vetor de coeficientes $\boldsymbol{\beta} = (\tau_1, \dots, \tau_{m-1}, \mu_1, \dots, \mu_r)$ é dado por

$$\hat{\beta} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{G}(\hat{\pi}) \quad (3.22)$$

com matriz de covariância

$$\text{Cov}(\hat{\beta}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \quad (3.23)$$

Exemplo

Para ilustrar a estimação dos parâmetros do modelo D, será retomado o exemplo anterior usado para estimar a matriz de covariância dos resíduos \mathbf{V} . No exemplo, a equação (3.21) fica da seguinte forma:

$$\begin{pmatrix} E(g(\hat{\pi}_{11})) \\ E(g(\hat{\pi}_{12})) \\ E(g(\hat{\pi}_{13})) \\ E(g(\hat{\pi}_{21})) \\ E(g(\hat{\pi}_{22})) \\ E(g(\hat{\pi}_{23})) \\ E(g(\hat{\pi}_{31})) \\ E(g(\hat{\pi}_{32})) \\ E(g(\hat{\pi}_{33})) \\ E(g(\hat{\pi}_{41})) \\ E(g(\hat{\pi}_{42})) \\ E(g(\hat{\pi}_{43})) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix}$$

A estimação via mínimos quadrados generalizados é feita através da fórmula $\hat{\beta} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{G}(\hat{\pi})$. Sob a restrição $\Sigma \mu_i = 0$, chega-se ao resultado:

$$\hat{\beta} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} = \begin{pmatrix} -0,726171 \\ -0,047999 \\ 0,5802597 \\ 1,0592232 \\ 0,5874257 \\ -1,100485 \\ -0,546164 \end{pmatrix}$$

A tabela a seguir apresenta os contrastes entre os estímulos e suas significâncias.

Tabela 3.6: Contrastes entre os Valores de Escala μ_i

Contraste	Valor	Desvio	Razão t	P-valor
$\mu_1 - \mu_2$	0,47180	0,17298	2,72739	0,00638
$\mu_1 - \mu_3$	2,15971	0,19212	11,24123	0,00000
$\mu_1 - \mu_4$	1,60539	0,17155	9,35794	0,00000
$\mu_2 - \mu_3$	1,68791	0,19212	8,78553	0,00000
$\mu_2 - \mu_4$	1,13359	0,17155	6,60779	0,00000
$\mu_3 - \mu_4$	-0,55432	0,19340	-2,86614	0,00416

Todos os contrastes foram considerados significativos. Isto significa que os quatro estímulos obtiveram avaliações diferentes dos juízes.

4 Coleta dos Dados

Devido à dificuldade em obter dados através da Universidade de Brasília, optou-se por ir a campo e coletar os dados diretamente nas turmas. Em geral, os questionários de avaliação discente são aplicados ao final das disciplinas. Dessa forma, os alunos têm tempo suficiente para analisar e avaliar seus professores. Para que a nota final não influencie nas avaliações, costuma-se aplicar os questionários antes da divulgação das notas finais. A coleta dos dados reais foi realizada em seis turmas da Universidade de Brasília no segundo semestre letivo de 2012, que devido à greve ocorrida neste mesmo ano, se estendeu até o dia 8 de março de 2013. Os questionários foram aplicados entre os dias 18 e 22 de fevereiro de 2013, levando em consideração os cuidados citados anteriormente.

O questionário de avaliação discente aplicado na UnB até 2011 utilizava uma escala de Likert de 0 a 10, 0 para insatisfatório e 10 para satisfatório. Em 2012 foram feitas algumas alterações, a escala passou a ser de 1 a 5 e o questionário passou a ser respondido via internet no ambiente utilizado pelos alunos para matrícula em disciplinas. Os dados reais utilizados no trabalho foram coletados nas próprias turmas com os itens do questionário antigo, mas com a escala atual de 1 a 5.

As turmas visitadas foram selecionadas por conveniência. O critério de seleção utilizado buscou captar cenários diferentes, de forma que fosse possível analisar os resultados em diversas situações. Como exemplo, disciplinas ofertadas no início do curso e disciplinas ofertadas no final, classes com muitos alunos e classes com poucos alunos, e diferentes tipos de professores (adjunto, titular, ou outros). O questionário foi aplicado em seis turmas, cada uma com um professor diferente. As turmas foram visitadas somente com a autorização dos seus respectivos professores e a identidade deles será mantida em sigilo, bem como a identidade dos alunos que participaram deste estudo.

No momento da aplicação dos questionários em cada turma, solicitei aos respectivos professores que aguardassem fora de sala para que os alunos se sentissem mais à

vontade para preencher os itens. Antes de distribuir os questionários aos estudantes, foi explicada a finalidade da pesquisa e a importância deste estudo para que as avaliações discentes possam contribuir cada vez mais para o aperfeiçoamento do ensino na Universidade. Também foi destacado que os respondentes não seriam identificados e que seus respectivos professores não teriam acesso às avaliações realizadas. Todos esses cuidados foram tomados para que as respostas obtidas fossem o mais fiel possível à realidade, de forma que os estudantes não tivessem receio de dar a sua opinião e ser penalizado por isso de alguma forma. O tempo médio de aplicação dos questionários foi de 25 minutos, desde as instruções de preenchimento até receber o último questionário.

Após a coleta, os questionários foram processados e armazenados em bases de dados da forma necessária para aplicar a rotina desenvolvida por Souza e Vales (2002) [14].

5 Resultados

O principal objetivo do estudo foi fazer uma análise das avaliações discente com foco no bloco de desempenho do professor, de modo que fossem geradas informações relevantes para fornecer aos professores meios de saber em quais itens ele tem se destacado e quais itens podem ser melhorados. Para o alcance deste objetivo, primeiro foi aplicado aos dados uma regressão logística. Esta é a técnica tradicional para análise de variáveis respostas com múltiplas categorias e pode ser vista no quinto capítulo de McCullagh and Nelder (1989) [8]. Em seguida, foi realizada a análise através da Lei dos Julgamentos Categóricos de Thurstone, como foi descrita anteriormente. E por fim, compara-se os resultados obtidos nas duas análises.

O procedimento foi aplicado em cada uma das seis turmas visitadas. Primeiro, é realizada uma análise descritiva. Uma tabela de dupla entrada com os itens nas linhas e as categorias nas colunas é apresentada para que seja observado como os alunos avaliaram os itens de 10 a 30, referentes ao desempenho do professor. Nota-se que alguns alunos não preencheram o questionário corretamente, deixando algum item em branco. Nestes casos, o item em branco foi desconsiderado da análise e a probabilidade do juiz classificar um estímulo S_i em uma das j primeiras categorias foi estimada apenas com as respostas válidas.

Em sequência, seguem os resultados da regressão logística. Para esta análise foi utilizado o modelo que pode ser escrito da forma

$$E(\text{logit}(\pi_{ij})) = \theta_j - \beta_i$$

onde θ_j representam os efeitos de cada categoria e β_i os efeitos dos estímulos. Apenas contrastes sobre os valores de β são estimáveis. Para construção do modelo adotou-se

como referência o estímulo de maior intensidade. Ou seja, o parâmetro β referente ao estímulo de maior intensidade foi igualado a zero.

Os resultados do modelo logístico são apresentados em duas tabelas. A primeira mostra o resultado do teste que verifica se os β 's são iguais a zero. A segunda contém os valores estimados destes parâmetros, seus respectivos desvios e o teste para verificar se a inclusão do parâmetro no modelo é significativa. Nesta tabela, quanto maiores foram os valores de β , mais este item difere do item tomado como referência. Se o estímulo tomado como referência é o que obteve a melhor das avaliações, os itens que mais diferem deste são aqueles que receberam piores notas dos alunos. Sendo assim, no modelo de regressão logística os parâmetros que obtêm p-valores significativos são os que foram piores avaliados pelos alunos.

Em seguida, para apresentação dos resultados da Lei dos Julgamentos Categóricos de Thurstone também foram apresentadas duas tabelas por turma. A primeira tabela contém o resultado do teste de ajustamento do modelo. Pode-se enxergar esta situação de forma análoga a uma regressão. Após o ajuste do modelo, faz-se um teste para analisar a qualidade deste ajuste. Na regressão, a partir de valores de uma variável explicativa X , e valores observados de Y para cada X , é possível estimar os parâmetros de forma que os valores preditos sejam o mais próximo possível dos valores observados. As diferenças entre estes dois valores são os resíduos, que são utilizados na estatística do teste de ajustamento do modelo. No caso dos julgamentos categóricos, X é substituída por uma combinação de estímulos e categorias, enquanto Y representa a frequência com que os juízes classificaram determinado estímulo em uma das k primeiras categorias. Dessa forma, o objetivo é estimar os parâmetros de forma que os resíduos, valores observados menos valores preditos, sejam os menores possíveis. Para verificar o ajuste do modelo, aplicou-se o seguinte teste de hipóteses:

H_0 : O modelo ficou bem ajustado aos dados.

H_a : O modelo não ficou bem ajustado aos dados.

A estatística utilizada para este teste tem como base a soma dos quadrados dos resíduos, que pode ser dada por:

$$SSE = \sum_{i=1}^r \sum_{j=1}^{m-1} \hat{r}_{ij}^2 = \sum_{i=1}^r \sum_{j=1}^{m-1} (\pi_{ij} - \hat{\pi}_{ij})^2 \quad (5.1)$$

onde r = número de estímulos;

m = número de categorias;

π_{ij} = valor observado, probabilidade de classificar o estímulo i em uma das j primeiras categorias;

$\hat{\pi}_{ij}$ = valor predito pelo modelo.

Como o modelo ajustado pela Lei dos Julgamentos Categóricos é baseado na suposição de que os valores de escala μ_i e os valores de localização τ_j seguem uma distribuição Normal, seus resíduos também seguem esta mesma distribuição. E dessa forma, a soma dos quadrados dos resíduos dividida por seu número de graus de liberdade segue uma distribuição Qui-quadrado. O número de graus de liberdade é equivalente ao número de valores observados, $r.(m-1)$, subtraído do número de parâmetros estimados, $(r-(m-1))$, ou seja:

$$gl = r.(m-1) - (r-(m-1))$$

Dessa forma, se os valores preditos pelo modelo ficarem próximos dos valores observados, a soma dos quadrados dos resíduos será pequena. Conseqüentemente, o p-valor, que é igual a probabilidade de se obter um valor menor ou igual ao do SSE obtido sob a hipótese de que o modelo ficou bem ajustado, será alto. Ou seja, ao observar um p-valor alto, não haverá evidências de que o modelo não ficou bem ajustado aos dados. Por outro lado, um p-valor baixo indica que o modelo não teve um bom ajuste e deve-se ter cuidado ao interpretar os resultados nestes casos.

A segunda tabela contém os valores estimados dos estímulos referentes ao desempenho do professor, que são os itens de 10 a 30 do questionário de avaliação discente aplicado. Nesta tabela também são apresentados os desvios e intensidades dos respectivos estímulos. Os estímulos estarão ordenados pela intensidade e através dos grupos é possível identificar entre quais estímulos existe diferença significativa.

Os grupos englobam estímulos que não são significativamente diferentes, ou seja, que não receberam avaliações suficientemente distintas dos juízes. A construção dos grupos é feita a partir dos contrastes entre os estímulos. No total, são 210 contrastes para cada turma, o que é equivalente a uma combinação de 21 estímulos selecionados

dois a dois. Para testar estes 210 contrastes a um nível de significância de 5% é preciso realizar um teste de comparações múltiplas. O método utilizado foi o de Bonferroni. Neste caso, para obter uma conclusão geral a um nível de significância de 5%, cada contraste é avaliado isoladamente a um nível de significância específico calculado por

$$\alpha = \frac{0,05}{total \cdot de \cdot contrastes} = \frac{0,05}{210} = 0,000238$$

Visto isso, os tópicos a seguir apresentarão os resultados dos questionários aplicados nas seis turmas, que serão chamadas de Turmas A, B, C, D, E e F.

5.1 Turma A

Primeiro, faremos uma breve análise descritiva dos dados coletados nesta turma. Nesta turma 15 alunos responderam ao questionário de avaliação discente. Segue abaixo a tabela de frequências dos estímulos por categorias. Vale ressaltar que os 15 alunos responderam às 21 questões referentes ao desempenho do professor.

Tabela 5.1: Frequência de Classificação dos Estímulos em cada Categoria da Turma A

Estímulos	Categorias				
	1	2	3	4	5
(10) Domínio do conteúdo	0	1	0	4	10
(11) Qualidade na transmissão	0	3	1	9	2
(12) Entusiasmo na transmissão	0	1	9	2	3
(13) Incentivo à participação	0	1	8	3	3
(14) Incentivo ao aprofundamento	0	1	8	4	1
(15) Uso de estratégias	0	7	4	2	0
(16) Ritmo de apresentação	0	1	8	5	1
(17) Adequação das atividades	0	1	5	6	3
(18) Profundidade do conteúdo	0	1	2	10	2
(19) Integração com conteúdos conexos	0	1	5	7	1
(20) Integração teoria/pesquisa/prática	0	3	6	4	2
(21) Qualidade dos exemplos	0	0	6	9	0
(22) Coerência	0	1	1	8	5
(23) Discussão das avaliações	0	2	5	6	2
(24) Uso dos recursos de ensino	0	1	1	5	8
(25) Disponibilidade para tirar dúvidas	0	2	2	6	5
(26) Cordialidade	0	0	0	7	8
(27) Respeito às idéias dos alunos	0	0	1	5	9
(28) Diponibilidade extra classe	0	2	2	6	4
(29) Cumprimento dos horários	0	5	5	2	3
(30) Assiduidade	0	1	3	7	4

Aparentemente os itens 10, 26 e 27 foram melhor avaliados pelos alunos. Estes itens correspondem a domínio do conteúdo ministrado, cordialidade na relação com os alunos, e respeito às ideias manifestadas pelos alunos acerca dos temas abordados, respectivamente.

5.1.1 Regressão Logística

Para estes dados o modelo de regressão logística apresentou os seguintes resultados:

Tabela 5.2: Testes de Hipóteses $\beta's = 0$ para a Turma A

Teste	Qui-quadrado	GL	P-valor
Razão de Verossimilhança	997,029	20	< 0,0001
Score	828,018	20	< 0,0001
Wald	890,631	20	< 0,0001

Nos três testes a conclusão foi a mesma. Ou seja, a variável resposta, que no caso é a categoria, é significativamente influenciada pelos estímulos. A próxima tabela contém informações sobre os parâmetros estimados.

Tabela 5.3: Estimação dos Parâmetros para a Turma A

Item	Valor estimado	Wald Qui-quadrado	P-valor
(15) Uso de estratégias	48,806	413,010	<,0001
(29) Cumprimento dos horários	33,928	213,758	<,0001
(14) Incentivo ao aprofundamento	31,868	189,586	<,0001
(20) Integração teoria/pesquisa/prática	30,224	171,145	<,0001
(16) Ritmo de apresentação	28,859	156,452	<,0001
(12) Entusiasmo na transmissão	28,464	152,301	<,0001
(19) Integração com conteúdos conexos	27,169	139,050	0,0002
(13) Incentivo à participação	26,833	135,696	0,0002
(23) Discussão das avaliações	25,722	124,873	0,0004
(21) Qualidade dos exemplos	24,666	114,969	0,0007
(17) Adequação das atividades	21,821	90,222	0,0027
(11) Qualidade na transmissão	21,579	88,250	0,0030
(28) Diponibilidade extra classe	20,355	78,605	0,0051
(18) Profundidade do conteúdo	18,614	65,823	0,0103
(30) Assiduidade	16,985	54,873	0,0192
(25) Disponibilidade para tirar dúvidas	15,779	47,393	0,0295
(22) Coerência	12,266	28,626	0,0907
(24) Uso dos recursos de ensino	0,610	0,692	0,4054
(26) Cordialidade	0,371	0,250	0,6169
(27) Respeito às idéias dos alunos	0,240	0,104	0,7476

Para elaboração do modelo o estímulo tomado como referência foi o item 10, por ter apresentado menor intensidade através da análise pela Lei dos Julgamentos Categóricos. Quando o p-valor da tabela é pequeno o suficiente, isto indica que o respectivo parâmetro é significativo para o modelo. Na turma A, a maioria dos parâmetros foi significativa, com exceção dos itens 22, 24, 26 e 27. A seguir será possível observar que este resultado é congruente com a análise por meio da Lei dos Julgamentos Categóricos de Thurstone.

5.1.2 Lei dos Julgamentos Categóricos de Thurstone

Agora serão apresentados os resultados da análise através da Lei dos Julgamentos Categóricos, na qual obteve-se o seguinte teste:

Tabela 5.4: Teste Qui-Quadrado para Bondade do Ajuste para a Turma A

SSE	GL	P-valor
120,0246	60	< 0,0001

A soma dos resíduos ao quadrado foi igual a 120,0246 e sob uma distribuição Qui-quadrado com 60 graus de liberdade, este valor sugere que o modelo não parece adequado ao dados. Ou seja, os valores preditos ficaram distantes dos valores observados. Portanto, deve-se ter cuidado com os resultados obtidos para esta turma. A seguir observa-se a tabela com os valores estimados dos parâmetros referentes aos estímulos de desempenho do professor, bem como seus respectivos desvios e intensidades. Ao lado, tem-se os grupos que permitem verificar entre quais estímulos há diferença significativa.

Tabela 5.5: Parâmetros Ordenados pela Intensidade para a Turma A

Item	Valor estimado	Desvio	Intensidade	Grupos
(10) Domínio do conteúdo	0,995	0,313	0,100	A
(27) Respeito às idéias dos alunos	0,869	0,300	0,088	A
(26) Cordialidade	0,680	0,317	0,073	A B
(24) Uso dos recursos de ensino	0,674	0,290	0,073	A
(22) Coerência	0,349	0,282	0,052	A B
(25) Disponibilidade para tirar dúvidas	0,207	0,276	0,045	A B
(30) Assiduidade	0,150	0,274	0,043	A B
(18) Profundidade do conteúdo	0,044	0,284	0,038	A B
(28) Diponibilidade extra classe	0,000	0,269	0,037	A B
(17) Adequação das atividades	-0,058	0,270	0,035	A B
(11) Qualidade na transmissão	-0,104	0,283	0,033	A B
(23) Discussão das avaliações	-0,210	0,270	0,030	A B
(13) Incentivo à participação	-0,234	0,277	0,029	A B
(21) Qualidade dos exemplos	-0,247	0,318	0,029	A B
(12) Entusiasmo na transmissão	-0,285	0,286	0,028	A B
(19) Integração com conteúdos conexos	-0,310	0,272	0,027	A B
(20) Integração teoria/pesquisa/prática	-0,344	0,271	0,026	A B
(29) Cumprimento dos horários	-0,401	0,273	0,024	A B
(16) Ritmo de apresentação	-0,411	0,280	0,024	A B
(14) Incentivo ao aprofundamento	-0,475	0,274	0,023	A B
(15) Uso de estratégias	-0,885	0,281	0,015	B

Os estímulos que mais se destacaram foram os itens 10 e 27, com intensidade de 0,100 e 0,088, respectivamente. Estes itens refletem o domínio do conteúdo ministrado e o respeito às idéias manifestadas pelos alunos acerca dos temas abordados. Por outro lado, o estímulo 15, referente ao uso de estratégias para motivar os alunos em relação ao conteúdo foi o item pior avaliado. Os alunos desta classe acreditam que o professor domina o conteúdo, no entanto, a forma como este conteúdo é transmitida é inferior em relação aos outros quesitos. Sob o nível de significância de 0,000238 do Teste de Bonferroni, apenas três dos 210 contrastes foram significativos. Os contrastes significativos são do item 15 com os itens 10, 27 e 24.

Ao comparar as duas análises pode-se notar que os parâmetros que não foram significativos no modelo de regressão logística, 22, 24, 26 e 27, são os parâmetros que obtiveram intensidades mais próximas do item 10, o qual foi tomado como referência na regressão logística. Isso se deve ao fato dos itens 22, 24, 26 e 27 receberem avaliações semelhantes ao item 10, e por isso, a inclusão desses parâmetros não é significativa em relação ao modelo nulo que reflete a valiação do item tomado como referência.

5.2 Turma B

Na turma B, 12 alunos responderam ao questionário de avaliação discente, sendo que todos marcaram respostas válidas nos itens correspondentes ao desempenho do professor. Segue abaixo a tabela de frequências dos estímulos por categorias.

Tabela 5.6: Frequência de Classificação dos Estímulos em cada Categoria da Turma B

Estímulos	Categorias				
	1	2	3	4	5
(10) Domínio do conteúdo	0	0	0	4	8
(11) Qualidade na transmissão	0	0	0	3	9
(12) Entusiasmo na transmissão	0	0	0	0	12
(13) Incentivo à participação	0	0	1	1	10
(14) Incentivo ao aprofundamento	0	1	2	3	6
(15) Uso de estratégias	0	0	2	6	4
(16) Ritmo de apresentação	0	0	1	6	5
(17) Adequação das atividades	0	0	0	4	8
(18) Profundidade do conteúdo	0	0	3	5	4
(19) Integração com conteúdos conexos	0	0	5	5	2
(20) Integração teoria/pesquisa/prática	0	0	1	5	6
(21) Qualidade dos exemplos	0	0	0	5	7
(22) Coerência	0	0	0	4	8
(23) Discussão das avaliações	0	0	3	4	5
(24) Uso dos recursos de ensino	0	0	1	2	9
(25) Disponibilidade para tirar dúvidas	0	0	0	4	8
(26) Cordialidade	0	0	0	1	11
(27) Respeito às idéias dos alunos	0	0	0	2	10
(28) Disponibilidade extra classe	0	0	0	6	6
(29) Cumprimento dos horários	0	0	0	1	11
(30) Assiduidade	0	0	0	2	10

À princípio, os itens 12, 26 e 29 se destacaram obtendo as melhores avaliações dos alunos. Estes itens referem-se a entusiasmo demonstrado na transmissão do conteúdo, cordialidade na relação com os alunos e cumprimento dos horários de início, respectivamente. Também nota-se que todos os itens obtiveram avaliações iguais ou maiores que a categoria 3. Aparentemente este professor está sendo bem avaliado em todos os itens.

5.2.1 Regressão Logística

Para estes dados o modelo de regressão logística apresentou os seguintes resultados:

Tabela 5.7: Testes de Hipóteses $\beta's = 0$ para a Turma B

Teste	Qui-quadrado	GL	P-valor
Razão de Verossimilhança	661,007	20	< 0,0001
Score	574,565	20	< 0,0001
Wald	460,082	20	0,0008

Nos três testes a conclusão foi de que os $\beta's$ são diferentes de zero, ou seja, os julgamentos são significativamente influenciados pelos estímulos. A próxima tabela contém informações sobre os parâmetros estimados.

Tabela 5.8: Estimação dos Parâmetros para a Turma B

Item	Valor estimado	Wald Qui-quadrado	P-valor
(19) Integração com conteúdos conexos	160,785	0,002	0,9642
(18) Profundidade do conteúdo	151,447	0,002	0,9663
(15) Uso de estratégias	149,354	0,002	0,9667
(23) Discussão das avaliações	148,766	0,002	0,9669
(14) Incentivo ao aprofundamento	146,271	0,002	0,9674
(16) Ritmo de apresentação	145,013	0,002	0,9677
(20) Integração teoria/pesquisa/prática	142,319	0,002	0,9683
(28) Disponibilidade extra classe	140,895	0,002	0,9686
(21) Qualidade dos exemplos	138,061	0,002	0,9693
(10) Domínio do conteúdo	134,864	0,001	0,9700
(17) Adequação das atividades	134,864	0,001	0,9700
(22) Coerência	134,864	0,001	0,9700
(25) Disponibilidade para tirar dúvidas	134,864	0,001	0,9700
(24) Uso dos recursos de ensino	132,179	0,001	0,9706
(11) Qualidade na transmissão	131,069	0,001	0,9708
(13) Incentivo à participação	127,166	0,001	0,9717
(27) Respeito às idéias dos alunos	126,147	0,001	0,9719
(30) Assiduidade	126,147	0,001	0,9719
(26) Cordialidade	118,397	0,001	0,9736
(29) Cumprimento dos horários	118,397	0,001	0,9736

O item 12 foi tomado como referência para a elaboração do modelo. Nesta turma, nenhum dos parâmetros foi considerado significativo. Isto significa que os itens foram avaliados de forma semelhante ao item tomado como referência, item 12. Este resultado também foi obtido na análise via Lei dos Julgamentos Categóricos de Thurstone, que será mostrada a seguir.

5.2.2 Lei dos Julgamentos Categóricos de Thurstone

Agora serão apresentados os resultados da análise através da Lei dos Julgamentos Categóricos, na qual obteve-se o seguinte resultado:

Tabela 5.9: Teste Qui-Quadrado para Bondade do Ajuste para a Turma B

SSE	GL	P-valor
15,2790	60	1

O valor de 15,279 na soma dos quadrados dos resíduos não é grande o suficiente para rejeitar a hipótese nula de que o modelo não é adequado. Portanto, pode-se concluir que a inclusão dos valores de escala μ_i foi significativa para o ajustamento do modelo.

A seguir observa-se a tabela com os valores estimados para os estímulos de desempenho do professor, bem como seus respectivos desvios e intensidades. Ao lado, tem-se os grupos que permitem verificar entre quais estímulos há diferença significativa.

Tabela 5.10: Parâmetros Ordenados pela Intensidade para a Turma B

Item	Valor estimado	Desvio	Intensidade	Grupos
(12) Entusiasmo na transmissão	2,251	1,822	0,300	A
(26) Cordialidade	0,911	0,503	0,078	A
(29) Cumprimento dos horários	0,911	0,503	0,078	A
(30) Assiduidade	0,507	0,423	0,052	A
(27) Respeito às idéias dos alunos	0,507	0,426	0,052	A
(13) Incentivo à participação	0,367	0,419	0,045	A
(11) Qualidade na transmissão	0,221	0,392	0,039	A
(24) Uso dos recursos de ensino	0,139	0,379	0,036	A
(10) Domínio do conteúdo	-0,016	0,373	0,031	A
(17) Adequação das atividades	-0,016	0,373	0,031	A
(25) Disponibilidade para tirar dúvidas	-0,016	0,373	0,031	A
(22) Coerência	-0,016	0,373	0,031	A
(21) Qualidade dos exemplos	-0,231	0,366	0,025	A
(20) Integração teoria/pesquisa/prática	-0,369	0,341	0,021	A
(28) Disponibilidade extra classe	-0,435	0,363	0,020	A
(16) Ritmo de apresentação	-0,514	0,337	0,018	A
(14) Incentivo ao aprofundamento	-0,589	0,334	0,017	A
(23) Discussão das avaliações	-0,72	0,334	0,015	A
(15) Uso de estratégias	-0,736	0,334	0,015	A
(18) Profundidade do conteúdo	-0,843	0,328	0,013	A
(19) Integração com conteúdos conexos	-1,308	0,338	0,008	A

A tabela acima reforça o comentário anterior de que não há muita diferença entre os estímulos. Isto porque observa-se apenas um grande grupo que contém todos os estímulos em estudo. Este professor foi bem avaliado em todos os itens, obtendo notas maiores ou iguais a três. Não há um critério que tenha se destacado dos demais tanto positiva como negativamente.

Comparando esta análise com o modelo de regressão logística nota-se que as duas chegaram ao mesmo resultado.

5.3 Turma C

Nesta turma, 24 alunos responderam ao questionário de avaliação discente. É importante observar que dois alunos deixaram um dos itens em branco. Dessa forma, nas questões 19 e 24 foram computadas apenas 23 respostas válidas. Segue abaixo a tabela de frequências dos estímulos por categorias.

Tabela 5.11: Frequência de Classificação dos Estímulos em cada Categoria da Turma C

Estímulos	Categorias				
	1	2	3	4	5
(10) Domínio do conteúdo	0	1	3	2	18
(11) Qualidade na transmissão	3	5	4	10	2
(12) Entusiasmo na transmissão	2	1	2	7	12
(13) Incentivo à participação	2	1	9	9	3
(14) Incentivo ao aprofundamento	2	4	8	9	1
(15) Uso de estratégias	4	7	10	3	0
(16) Ritmo de apresentação	12	7	2	2	1
(17) Adequação das atividades	2	5	6	8	3
(18) Profundidade do conteúdo	3	3	3	10	5
(19) Integração com conteúdos conexos	1	1	3	11	7
(20) Integração teoria/pesquisa/prática	4	2	2	7	9
(21) Qualidade dos exemplos	1	0	8	8	7
(22) Coerência	2	1	3	10	8
(23) Discussão das avaliações	3	4	6	5	6
(24) Uso dos recursos de ensino	3	4	3	7	6
(25) Disponibilidade para tirar dúvidas	3	3	8	5	5
(26) Cordialidade	1	2	2	9	10
(27) Respeito às idéias dos alunos	1	1	4	6	12
(28) Diponibilidade extra classe	2	6	3	7	6
(29) Cumprimento dos horários	0	0	2	8	14
(30) Assiduidade	0	0	0	1	23

À princípio, os itens 12, 26 e 29 se destacaram positivamente. Estes itens referem-se a entusiasmo demonstrado na transmissão do conteúdo, cordialidade na relação com os alunos e cumprimento dos horários de início, respectivamente.

5.3.1 Regressão Logística

Para estes dados o modelo de regressão logística apresentou os seguintes resultados:

Nos três testes a conclusão foi a mesma. Ou seja, a variável resposta, que no caso é a categoria, é significativamente influenciada pelos estímulos. Como a hipótese de que

Tabela 5.12: Testes de Hipóteses $\beta's = 0$ para a Turma C

Teste	Qui-quadrado	GL	P-valor
Razão de Verossimilhança	1787,714	20	< 0,0001
Score	1438,835	20	< 0,0001
Wald	1306,977	20	< 0,0001

os β 's são iguais a zero foi rejeitada, pode-se concluir que os estímulos possuem efeitos consideravelmente diferentes.

Tabela 5.13: Estimação dos Parâmetros para a Turma C

Item	Valor estimado	Wald Qui-quadrado	P-valor
(16) Ritmo de apresentação	68,468	382,374	<,0001
(15) Uso de estratégias	57,657	276,773	<,0001
(14) Incentivo ao aprofundamento	49,725	206,979	<,0001
(11) Qualidade na transmissão	49,299	203,485	<,0001
(17) Adequação das atividades	48,515	197,118	<,0001
(25) Disponibilidade para tirar dúvidas	47,863	191,907	<,0001
(23) Discussão das avaliações	46,969	184,853	<,0001
(13) Incentivo à participação	45,865	176,317	<,0001
(28) Diponibilidade extra classe	45,521	173,699	<,0001
(24) Uso dos recursos de ensino	44,835	167,709	<,0001
(18) Profundidade do conteúdo	43,958	162,023	<,0001
(21) Qualidade dos exemplos	39,773	132,643	0,0003
(20) Integração teoria/pesquisa/prática	39,671	131,958	0,0003
(22) Coerência	37,563	118,227	0,0006
(19) Integração com conteúdos conexos	37,241	115,607	0,0007
(26) Cordialidade	34,633	100,275	0,0015
(27) Respeito às idéias dos alunos	32,731	89,333	0,0028
(12) Entusiasmo na transmissão	32,410	87,541	0,0031
(29) Cumprimento dos horários	27,151	60,530	0,0139
(10) Domínio do conteúdo	21,759	37,661	0,0523

Nesta turma, o estímulo tomado como referência para elaboração do modelo foi o item 30. Na tabela acima, é possível notar que apenas o item 10 não foi significativo.

5.3.2 Lei dos Julgamentos Categóricos de Thurstone

A seguir, serão apresentados os resultados da análise através da Lei dos Julgamentos Categóricos, na qual obteve-se o seguinte resultado:

Tabela 5.14: Teste Qui-Quadrado para Bondade do Ajuste para a Turma C

SSE	GL	P-valor
115,1875	60	<0,0001

A soma dos resíduos ao quadrado foi igual a 115,1875. A um nível de significância de 5% a conclusão do teste é de que o modelo parece não estar adequado aos dados. Dessa forma, é preciso ter cuidado ao analisar os resultados desta turma.

A seguir observa-se a tabela com os valores estimados para os estímulos de desempenho do professor. Ao lado, tem-se os grupos que permitem verificar entre quais estímulos há diferença significativa.

Tabela 5.15: Parâmetros Ordenados pela Intensidade para a Turma C

Item	Valor estimado	Desvio	Intensidade	Grupos
(30) Assiduidade	2,162	0,439	0,271	A
(10) Domínio do conteúdo	0,942	0,259	0,08	A B
(29) Cumprimento dos horários	0,716	0,237	0,064	A B C
(27) Respeito às idéias dos alunos	0,392	0,222	0,046	A B C D
(12) Entusiasmo na transmissão	0,388	0,225	0,046	A B C D
(26) Cordialidade	0,288	0,219	0,041	B C D E
(19) Integração com conteúdos conexos	0,145	0,221	0,036	B C D E
(22) Coerência	0,119	0,216	0,035	B C D E
(20) Integração teoria/pesquisa prática	0,006	0,217	0,031	B C D E
(21) Qualidade dos exemplos	-0,005	0,216	0,031	B C D E
(18) Profundidade do conteúdo	-0,204	0,213	0,025	B C D E F
(24) Uso dos recursos de ensino	-0,214	0,217	0,025	B C D E F
(28) Disponibilidade extra classe	-0,237	0,212	0,024	B C D E F
(23) Discussão das avaliações	-0,300	0,211	0,023	B C D E F
(13) Incentivo à participação	-0,327	0,212	0,022	C D E F
(25) Disponibilidade para tirar dúvidas	-0,362	0,211	0,021	C D E F
(17) Adequação das atividades	-0,419	0,209	0,02	C D E F
(11) Qualidade na transmissão	-0,460	0,211	0,019	C D E F
(14) Incentivo ao aprofundamento	-0,496	0,213	0,019	D E F
(15) Uso de estratégias	-0,850	0,215	0,013	E F
(16) Ritmo de apresentação	-1,281	0,226	0,008	F

Na tabela acima o item 30 foi o melhor avaliado pelos alunos obtendo a maior das intensidades. No mesmo grupo estão os itens 10, 29, 27 e 12. Por outro lado, os itens 15 e 16 obtiveram as duas menores intensidades. Eles referem-se ao uso de estratégias para

motivar os alunos e relação ao conteúdo e ritmo de apresentação do conteúdo. A ordem em que os itens ficaram ordenados sugere que, na visão dos alunos, este professor possui domínio do conteúdo, mas a forma como este conteúdo é apresentado não está boa em comparação com os outros itens.

Na análise de regressão logística, que tomou como referência o item 30, apenas o item 10 não foi considerado significativo. Na tabela anterior pode-se notar que este é o item com intensidade mais próxima do item 30. Dessa forma, as duas análises obtiveram resultados congruentes.

5.4 Turma D

Na turma D, 15 alunos responderam ao questionário de avaliação discente. Todos os alunos preencheram as questões de forma correta. Segue abaixo a tabela de frequências dos estímulos por categorias.

Tabela 5.16: Frequência de Classificação dos Estímulos em cada Categoria da Turma D

Estímulos	Categorias				
	1	2	3	4	5
(10) Domínio do conteúdo	0	0	0	1	14
(11) Qualidade na transmissão	0	3	1	4	7
(12) Entusiasmo na transmissão	0	0	2	2	11
(13) Incentivo à participação	0	2	2	1	10
(14) Incentivo ao aprofundamento	0	1	3	4	7
(15) Uso de estratégias	0	2	4	4	5
(16) Ritmo de apresentação	0	1	4	2	8
(17) Adequação das atividades	0	0	3	4	8
(18) Profundidade do conteúdo	0	0	1	5	9
(19) Integração com conteúdos conexos	0	0	1	4	10
(20) Integração teoria/pesquisa/prática	0	1	3	3	8
(21) Qualidade dos exemplos	0	1	1	6	7
(22) Coerência	0	0	1	3	11
(23) Discussão das avaliações	1	0	5	2	7
(24) Uso dos recursos de ensino	0	2	4	3	6
(25) Disponibilidade para tirar dúvidas	0	0	1	4	10
(26) Cordialidade	0	0	1	1	13
(27) Respeito às idéias dos alunos	0	0	0	2	13
(28) Disponibilidade extra classe	0	0	0	4	11
(29) Cumprimento dos horários	0	0	1	3	11
(30) Assiduidade	0	0	0	2	13

Aparentemente os itens 10, 27 e 30 se destacaram positivamente nas avaliações dos alunos.

5.4.1 Regressão Logística

Para estes dados o modelo de regressão logística apresentou os seguintes resultados:

Os testes apresentaram a mesma conclusão. Existe diferenças relevantes entre os estímulos em estudo, pois a hipótese nula de que os β 's são iguais a zero foi rejeitada. A próxima tabela contém informações sobre os parâmetros estimados.

Tabela 5.17: Testes de Hipóteses $\beta's = 0$ para a Turma D

Teste	Qui-quadrado	GL	P-valor
Razão de Verossimilhança	485,297	20	0,0004
Score	444,516	20	0,0013
Wald	402,739	20	0,0046

Tabela 5.18: Estimação dos Parâmetros para a Turma D

Item	Valor estimado	Wald Qui-quadrado	P-valor
(15) Uso de estratégias	34,235	89,143	0,0028
(24) Uso dos recursos de ensino	32,803	81,819	0,0042
(23) Discussão das avaliações	30,643	71,280	0,0076
(11) Qualidade na transmissão	29,390	65,455	0,0105
(14) Incentivo ao aprofundamento	28,121	59,774	0,0145
(16) Ritmo de apresentação	27,425	56,752	0,0172
(20) Integração teoria/pesquisa/prática	26,304	52,038	0,0225
(21) Qualidade dos exemplos	26,102	51,208	0,0236
(17) Adequação das atividades	24,787	45,948	0,0321
(13) Incentivo à participação	22,349	36,895	0,0548
(18) Profundidade do conteúdo	21,015	32,340	0,0721
(19) Integração com conteúdos conexos	18,727	25,202	0,1124
(25) Disponibilidade para tirar dúvidas	18,727	25,202	0,1124
(12) Entusiasmo na transmissão	16,821	19,937	0,1579
(22) Coerência	16,017	17,902	0,1809
(29) Cumprimento dos horários	16,017	17,902	0,1809
(28) Diponibilidade extra classe	15,267	16,109	0,2044
(26) Cordialidade	0,810	0,396	0,529
(27) Respeito às idéias dos alunos	0,7404	0,3249	0,5687
(30) Assiduidade	0,740	0,325	0,5687

Na turma D o item 10 foi tomado como referência. Ao observar os p-valores, nota-se que aproximadamente 50% dos itens é significativamente diferente do estímulo de referência.

5.4.2 Lei dos Julgamentos Categóricos de Thurstone

Ao aplicar a Lei dos Julgamentos Categóricos aos dados, obteve-se o seguinte ajustamento do modelo:

Tabela 5.19: Teste Qui-Quadrado para Bondade do Ajuste para a Turma D

SSE	GL	P-valor
27,7663	60	0,9999

A soma dos resíduos ao quadrado igual a 27,7663, isso revela que os valores preditos pelo modelo estão próximos dos valores observados. Portanto, o modelo pode ser considerado adequado aos dados.

A seguir observa-se a tabela com os valores estimados dos parâmetros referentes aos estímulos de desempenho do professor, bem como seus respectivos desvios e intensidades. Ao lado, tem-se os grupos que permitem verificar entre quais estímulos há diferença significativa.

Tabela 5.20: Parâmetros Ordenados pela Intensidade para a Turma D

Item	Valor estimado	Desvio	Intensidade	Grupos
(10) Domínio do conteúdo	1,081	0,476	0,117	A
(27) Respeito às idéias dos alunos	0,696	0,392	0,079	A
(30) Assiduidade	0,696	0,392	0,079	A
(26) Cordialidade	0,651	0,386	0,076	A
(29) Cumprimento dos horários	0,263	0,326	0,051	A
(22) Coerência	0,263	0,326	0,051	A
(28) Diponibilidade extra classe	0,213	0,337	0,049	A
(12) Entusiasmo na transmissão	0,180	0,326	0,047	A
(25) Disponibilidade para tirar dúvidas	0,112	0,312	0,044	A
(19) Integração com conteúdos conexos	0,112	0,312	0,044	A
(18) Profundidade do conteúdo	-0,022	0,301	0,038	A
(13) Incentivo à participação	-0,174	0,311	0,033	A
(17) Adequação das atividades	-0,265	0,294	0,030	A
(20) Integração teoria/pesquisa/prática	-0,309	0,289	0,029	A
(21) Qualidade dos exemplos	-0,319	0,286	0,028	A
(16) Ritmo de apresentação	-0,361	0,291	0,027	A
(14) Incentivo ao aprofundamento	-0,403	0,283	0,026	A
(23) Discussão das avaliações	-0,510	0,287	0,023	A
(11) Qualidade na transmissão	-0,536	0,287	0,023	A
(24) Uso dos recursos de ensino	-0,645	0,279	0,020	A
(15) Uso de estratégias	-0,724	0,277	0,019	A

A tabela acima reforça a conclusão do teste de ajustamento do modelo para a Lei dos Julgamentos Categóricos. Isto porque observa-se apenas um grande grupo que contém todos os estímulos em estudo. Isto pode indicar que este professor fora avaliado de forma parecida em todos os itens.

5.5 Turma E

Na turma E, 26 alunos responderam ao questionário de avaliação discente. Algumas respostas não foram válidas. Nos itens 15, 26 e 28 houveram uma falha, enquanto no item 23 foram 2 respostas inválidas. Segue abaixo a tabela de frequências dos estímulos por categorias.

Tabela 5.21: Frequência de Classificação dos Estímulos em cada Categoria da Turma E

Estímulos	Categorias				
	1	2	3	4	5
(10) Domínio do conteúdo	0	0	0	3	23
(11) Qualidade na transmissão	1	0	1	5	19
(12) Entusiasmo na transmissão	0	0	2	5	19
(13) Incentivo à participação	1	0	4	13	8
(14) Incentivo ao aprofundamento	0	1	10	11	4
(15) Uso de estratégias	3	1	8	7	6
(16) Ritmo de apresentação	0	0	6	7	13
(17) Adequação das atividades	0	0	2	11	13
(18) Profundidade do conteúdo	0	0	1	5	20
(19) Integração com conteúdos conexos	0	2	8	10	6
(20) Integração teoria/pesquisa/prática	0	1	5	12	8
(21) Qualidade dos exemplos	0	1	3	10	12
(22) Coerência	1	1	2	5	17
(23) Discussão das avaliações	0	1	9	7	7
(24) Uso dos recursos de ensino	0	0	4	7	15
(25) Disponibilidade para tirar dúvidas	0	0	3	2	21
(26) Cordialidade	0	0	2	4	19
(27) Respeito às idéias dos alunos	0	1	2	4	19
(28) Diponibilidade extra classe	1	1	3	9	11
(29) Cumprimento dos horários	0	0	1	4	21
(30) Assiduidade	0	0	1	3	22

À princípio, o item 10, referente a domínio do conteúdo ministrado, se destaca obtendo as melhores avaliações dos alunos.

5.5.1 Regressão Logística

Para estes dados o modelo de regressão logística apresentou os seguintes resultados:

Nos três testes a conclusão foi a mesma. O que significa que a inclusão dos parâmetros no modelo é significativa, pois a hipótese nula de que os β 's são iguais a zero foi rejeitada. A próxima tabela contém informações sobre os parâmetros estimados.

Tabela 5.22: Testes de Hipóteses $\beta's = 0$ para a Turma E

Teste	Qui-quadrado	GL	P-valor
Razão de Verossimilhança	1291,768	20	< 0,0001
Score	1204,917	20	< 0,0001
Wald	1112,819	20	< 0,0001

Tabela 5.23: Estimação dos Parâmetros para a Turma E

Item	Valor estimado	Wald Qui-quadrado	P-valor
(15) Uso de estratégias	35,776	243,435	<,0001
(14) Incentivo ao aprofundamento	33,924	221,795	<,0001
(19) Integração com conteúdos conexos	32,125	199,426	<,0001
(23) Discussão das avaliações	31,181	184,282	<,0001
(20) Integração teoria/pesquisa/prática	26,906	140,224	0,0002
(13) Incentivo à participação	26,240	133,321	0,0003
(28) Diponibilidade extra classe	23,325	103,836	0,0013
(16) Ritmo de apresentação	21,584	89,376	0,0028
(21) Qualidade dos exemplos	21,584	89,375	0,0028
(17) Adequação das atividades	19,056	68,873	0,0087
(24) Uso dos recursos de ensino	17,841	59,914	0,0144
(22) Coerência	15,431	43,946	0,0361
(27) Respeito às idéias dos alunos	11,381	22,738	0,1316
(11) Qualidade na transmissão	10,880	20,618	0,151
(12) Entusiasmo na transmissão	10,733	20,016	0,1571
(26) Cordialidade	0,935	14,629	0,2265
(18) Profundidade do conteúdo	0,836	11,624	0,281
(25) Disponibilidade para tirar dúvidas	0,714	0,827	0,3633
(29) Cumprimento dos horários	0,620	0,608	0,4355
(30) Assiduidade	0,363	0,194	0,6595

Para elaboração do modelo o estímulo tomado como referência foi o item 10. Novamente, cerca de 50% dos itens foram considerados significativos.

5.5.2 Lei dos Julgamentos Categóricos de Thurstone

Ao aplicar a Lei dos Julgamentos Categóricos aos dados, obteve-se o seguinte ajustamento do modelo:

Tabela 5.24: Teste Qui-Quadrado para Bondade do Ajuste para a Turma E

SSE	GL	P-valor
35,9681	60	0,9941

A soma dos resíduos ao quadrado foi igual a 35,9681. Este valor não é suficiente grande para dizer que o modelo não é adequado. O valores preditos ficaram próximos dos valores observados.

A seguir observa-se a tabela com os valores estimados para os estímulos de desempenho do professor, bem como seus respectivos desvios e intensidades. Ao lado, tem-se os grupos que permitem verificar entre quais estímulos há diferença significativa.

Tabela 5.25: Parâmetros Ordenados pela Intensidade para a Turma E

Item	Valor estimado	Desvio	Intensidade	Grupos
(10) Domínio do conteúdo	0,995	0,311	0,104	A
(30) Assiduidade	0,797	0,282	0,085	A B
(29) Cumprimento dos horários	0,669	0,267	0,075	A B
(18) Profundidade do conteúdo	0,557	0,256	0,067	A B
(25) Disponibilidade para tirar dúvidas	0,528	0,268	0,065	A B C
(26) Cordialidade	0,465	0,258	0,061	A B C D
(12) Entusiasmo na transmissão	0,399	0,246	0,057	A B C D
(11) Qualidade na transmissão	0,357	0,245	0,055	A B C D
(27) Respeito às idéias dos alunos	0,307	0,244	0,052	A B C D E
(22) Coerência	0,073	0,231	0,041	A B C D E F
(24) Uso dos recursos de ensino	-0,021	0,226	0,037	A B C D E F
(17) Adequação das atividades	-0,064	0,223	0,036	A B C D E F
(21) Qualidade dos exemplos	-0,215	0,216	0,031	A B C D E F
(16) Ritmo de apresentação	-0,252	0,221	0,029	A B C D E F
(28) Diponibilidade extra classe	-0,319	0,217	0,028	A B C D E F
(13) Incentivo à participação	-0,471	0,211	0,024	B C D E F
(20) Integração teoria/pesquisa prática	-0,504	0,210	0,023	B C D E F
(23) Discussão das avaliações	-0,714	0,219	0,018	C D E F
(19) Integração com conteúdos conexos	-0,780	0,207	0,017	D E F
(14) Incentivo ao aprofundamento	-0,886	0,210	0,015	E F
(15) Uso de estratégias	-0,920	0,211	0,015	F

Os itens que mais se destacaram positivamente foram os itens 10 e 30, com intensidade de 0,104 e 0,085, respectivamente. Por outro lado, os itens 15 e 14, foram os critérios que receberam as piores avaliações dos alunos. Neste caso, apesar deste professor ter domínio do conteúdo ministrado, os alunos deram notas baixas para uso de estratégias e incentivo ao aprofundamento. Evidenciando que este item pode ser melhorado.

5.6 Turma F

Nesta turma, 29 alunos responderam ao questionário de avaliação discente. Algumas respostas não foram válidas. No item 20 houve uma falha, enquanto no item 22 foram 2 respostas inválidas. Segue abaixo a tabela de frequências dos estímulos por categorias.

Tabela 5.26: Frequência de Classificação dos Estímulos em cada Categoria da Turma F

Estímulos	Categorias				
	1	2	3	4	5
(10) Domínio do conteúdo	0	0	1	4	24
(11) Qualidade na transmissão	0	0	0	10	19
(12) Entusiasmo na transmissão	0	1	5	7	16
(13) Incentivo à participação	0	0	4	6	19
(14) Incentivo ao aprofundamento	0	2	7	9	11
(15) Uso de estratégias	0	3	8	12	6
(16) Ritmo de apresentação	2	2	9	7	9
(17) Adequação das atividades	0	0	5	9	15
(18) Profundidade do conteúdo	0	1	3	11	14
(19) Integração com conteúdos conexos	0	3	6	5	15
(20) Integração teoria/pesquisa/prática	1	0	9	8	9
(21) Qualidade dos exemplos	1	0	3	12	13
(22) Coerência	0	0	5	10	13
(23) Discussão das avaliações	1	6	7	6	9
(24) Uso dos recursos de ensino	0	0	5	9	15
(25) Disponibilidade para tirar dúvidas	0	0	1	3	25
(26) Cordialidade	0	0	1	3	25
(27) Respeito às idéias dos alunos	0	0	1	6	22
(28) Diponibilidade extra classe	0	0	2	6	21
(29) Cumprimento dos horários	0	3	4	7	15
(30) Assiduidade	0	0	3	3	23

Aparentemente, os itens 25 e 26 se destacaram obtendo as melhores avaliações dos alunos.

5.6.1 Regressão Logística

Para estes dados o modelo de regressão logística apresentou os seguintes resultados:

Tabela 5.27: Testes de Hipóteses $\beta's = 0$ para a Turma F

Teste	Qui-quadrado	GL	P-valor
Razão de Verossimilhança	1131,030	20	< 0,0001
Score	1031,729	20	< 0,0001
Wald	988,274	20	< 0,0001

Nos três testes a conclusão foi a mesma. A hipótese nula de que os β 's são iguais a zero foi rejeitada. E portanto, pode-se concluir que os itens influenciam as avaliações realizadas pelo alunos.

Tabela 5.28: Estimação dos Parâmetros para a Turma F

Item	Valor estimado	Wald Qui-quadrado	P-valor
(23) Discussão das avaliações	31,541	243,642	<,0001
(16) Ritmo de apresentação	29,223	209,847	<,0001
(15) Uso de estratégias	29,052	207,438	<,0001
(20) Integração teoria/pesquisa/prática	25,729	159,638	<,0001
(14) Incentivo ao aprofundamento	23,756	138,728	0,0002
(19) Integração com conteúdos conexos	20,372	101,257	0,0015
(29) Cumprimento dos horários	19,070	88,283	0,0030
(21) Qualidade dos exemplos	18,843	86,100	0,0033
(22) Coerência	18,779	84,605	0,0036
(18) Profundidade do conteúdo	17,873	77,090	0,0055
(17) Adequação das atividades	17,216	71,252	0,0076
(24) Uso dos recursos de ensino	17,216	71,252	0,0076
(12) Entusiasmo na transmissão	16,941	68,875	0,0087
(13) Incentivo à participação	12,178	34,053	0,065
(11) Qualidade na transmissão	10,401	24,238	0,1195
(28) Diponibilidade extra classe	0,846	15,508	0,213
(27) Respeito às idéias dos alunos	0,643	0,859	0,354
(30) Assiduidade	0,543	0,599	0,4391
(10) Domínio do conteúdo	0,251	0,118	0,7313
(26) Cordialidade	0,000	0,000	1,0000

Para elaboração do modelo o estímulo tomado como referência foi o item 25. Cerca de 50% dos parâmetros foram considerados significativos, ou seja, foram avaliados de forma diferente do estímulo de referência, item 25. De forma geral, os itens significativos estão relacionados à maneira com que o professor transmite o conteúdo e às estratégias utilizadas.

5.6.2 Lei dos Julgamentos Categóricos de Thurstone

Ao aplicar a Lei dos Julgamentos Categóricos aos dados, obteve-se o seguinte resultado:

Tabela 5.29: Teste Qui-Quadrado para Bondade do Ajuste para a Turma F

SSE	GL	P-valor
30,0553	60	0,9996

A soma dos resíduos ao quadrado foi igual a 30,0553. A um nível de significância de 5% a conclusão do teste é de que o modelo ficou adequado aos dados. Ou seja, os valores preditos ficaram próximos dos valores observados.

A seguir observa-se a tabela com os valores estimados para os estímulos de desempenho do professor, bem como seus respectivos desvios e intensidades.

Tabela 5.30: Parâmetros Ordenados pela Intensidade para a Turma F

Item	Valor estimado	Desvio	Intensidade	Grupos
(25) Disponibilidade para tirar dúvidas	0,896	0,277	0,095	A
(26) Cordialidade	0,896	0,277	0,095	A
(10) Domínio do conteúdo	0,772	0,260	0,084	A B
(27) Respeito às idéias dos alunos	0,564	0,239	0,068	A B C
(30) Assiduidade	0,551	0,250	0,067	A B C
(28) Diponibilidade extra classe	0,433	0,233	0,060	A B C D
(11) Qualidade na transmissão	0,214	0,235	0,048	A B C D E
(13) Incentivo à participação	0,190	0,223	0,047	A B C D E
(12) Entusiasmo na transmissão	-0,071	0,207	0,036	A B C D E
(17) Adequação das atividades	-0,095	0,212	0,035	A B C D E
(24) Uso dos recursos de ensino	-0,095	0,212	0,035	A B C D E
(18) Profundidade do conteúdo	-0,115	0,208	0,034	A B C D E
(21) Qualidade dos exemplos	-0,169	0,203	0,033	A B C D E
(22) Coerência	-0,186	0,209	0,032	A B C D E
(29) Cumprimento dos horários	-0,239	0,207	0,030	A B C D E
(19) Integração com conteúdos conexos	-0,306	0,210	0,028	A B C D E
(14) Incentivo ao aprofundamento	-0,452	0,200	0,024	B C D E
(20) Integração teoria/pesquisa prática	-0,543	0,206	0,022	C D E
(16) Ritmo de apresentação	-0,696	0,197	0,019	D E
(15) Uso de estratégias	-0,755	0,196	0,018	E
(23) Discussão das avaliações	-0,789	0,198	0,017	E

Os estímulos que se destacaram positivamente foram os itens 25 e 26, seguidos pelo item 10. É o mesmo resultado obtido através da regressão logística. Por outro lado, os itens 23, 15 e 16 apresentaram as menores intensidades. Os alunos desta classe acreditam que o professor domina o conteúdo, no entanto, a forma como este conteúdo é transmitida é inferior em relação aos outros quesitos.

5.7 Comparação entre as Duas Técnicas

Para comparar as duas análises ordenou-se os itens. Nas primeiras posições ficam os estímulos melhores avaliados. No modelo de regressão logística os itens foram ordenados segundo o valor estimado do parâmetro, quanto menor este valor, melhor a posição do item. Já no modelo dos Julgamentos Categóricos os itens foram ordenados segundo suas intensidades, quanto maior a intensidade, melhor a posição do item. Na tabela abaixo são apresentadas as posições dos itens obtidas através dos dois modelos. São duas colunas para cada turma. A primeira contém as posições do modelo de regressão logística e a segunda, as posições do modelo de Julgamentos Categóricos.

Tabela 5.31: Posição dos Itens por Turma

Item	A		B		C		D		E		F	
(10) Domínio do conteúdo	1	1	9	12	2	2	1	1	1	1	3	3
(11) Qualidade na transmissão	11	10	7	7	18	18	19	18	8	8	7	7
(12) Entusiasmo na transmissão	15	16	1	1	5	4	8	8	7	7	9	9
(13) Incentivo à participação	13	14	6	6	15	14	12	12	16	16	8	8
(14) Incentivo ao aprofundamento	20	19	17	17	19	19	17	17	20	20	17	17
(15) Uso de estratégias	21	21	19	19	20	20	21	21	21	21	20	19
(16) Ritmo de apresentação	19	17	16	16	21	21	16	16	14	14	19	20
(17) Adequação das atividades	10	11	10	11	17	17	13	13	12	12	10	11
(18) Profundidade do conteúdo	8	8	20	20	11	11	11	11	4	5	12	12
(19) Integração com conteúdos conexos	16	15	21	21	7	7	10	10	19	19	16	16
(20) Integração teoria/pesquisa prática	17	18	14	15	9	9	14	15	17	17	18	18
(21) Qualidade dos exemplos	14	12	13	13	10	10	15	14	13	13	13	14
(22) Coerencia	5	5	12	10	8	8	6	7	10	10	14	13
(23) Discussão das avaliações	12	13	18	18	14	15	18	19	18	18	21	21
(24) Uso dos recursos de ensino	4	4	8	8	12	12	20	20	11	11	11	10
(25) Disponibilidade para tirar dúvidas	6	6	11	9	16	16	9	9	5	4	1	1
(26) Cordialidade	3	3	2	3	6	6	4	4	6	6	2	2
(27) Respeito às idéias dos alunos	2	2	5	5	4	5	2	3	9	9	4	5
(28) Disponibilidade extra classe	9	9	15	14	13	13	7	5	15	15	6	6
(29) Cumprimento dos horários	18	20	3	2	3	3	5	6	3	3	15	15
(30) Assiduidade	7	7	4	4	1	1	3	2	2	2	5	4

A partir da tabela pode-se notar que o resultado das duas técnicas é bem parecido. A maior diferença observada foi de três posições. É importante ressaltar que se um professor ficou em uma posição melhor que outro em determinado item isso não significa que ele

tenha recebido notas maiores do que o outro professor. Por exemplo, um professor pode ter recebido notas 5 em todos os itens exceto o 15, em que ele recebeu notas 4, e neste caso o item 15 ficaria na última posição. Por outro lado, outro professor poderia ter recebido notas 2 em todos os itens exceto o 15, em que recebeu notas 3, e neste caso o item 15 ficaria em primeiro lugar mesmo recebendo notas 3 que são menores do que o primeiro professor recebeu neste mesmo item.

De forma geral, os professores receberam avaliações próximas. Itens como domínio do conteúdo, assiduidade e cordialidade receberam boas avaliações nas seis turmas. Por outro lado, uso de estratégias, incentivo ao aprofundamento e ritmo de apresentação ficaram entre as últimas posições em todas as turmas.

Além disso, alguns itens podem ser destacados por receberem avaliação distinta em uma turma específica, como é o caso do domínio do conteúdo. Este item ficou entre as três primeiras colocações exceto na turma B, em que ficou numa posição abaixo das demais turmas. Em integração com conteúdos conexos e integração teoria/pesquisa/prática a turma C se destaca das demais por ter ficado em uma colocação melhor nestes dois itens. Em relação ao item disponibilidade para tirar dúvidas, a turma F obteve melhor posição enquanto a turma C ficou em pior posição.

6 Conclusão

No presente trabalho, foi apresentada uma nova forma de se analisar os dados provenientes das avaliações discentes, uma análise conhecida como Lei dos Julgamentos Categóricos de Thurstone. Procurou-se uma alternativa que fornecesse aos docentes informações consistentes e úteis sobre a percepção dos alunos com relação ao seu desempenho.

Atualmente, essas avaliações são extremamente importantes para as universidades. São usadas como base para avaliação da efetividade do ensino em decisões administrativas, por exemplo, promoção ou aumento salarial. Apesar da validade das avaliações discentes ser questionada por diversos estudiosos, elas continuam sendo amplamente utilizadas por serem pouco onerosas às instituições de ensino.

No entanto, é preciso ter cuidado ao tirar conclusões a partir dessas avaliações. Tendo em vista que elas podem ser influenciadas por fatores externos ao desempenho do professor, como semestre do curso ou tamanho da classe, não é adequado comparar o resultado de dois professores cujas turmas possuem características diferentes.

Entre os fatores que influenciam as avaliações discentes, a expectativa que o aluno tem da menção que irá obter na disciplina merece destaque. O fato de que vários docentes percebem uma correlação positiva entre as notas dadas aos alunos e as avaliações discentes tem consequências importantes quando também existe a percepção de que essas avaliações são utilizadas para tomada de decisões administrativas, principalmente para promoções e aumentos salariais (Johnson, 2003) [4]. Essa conjuntura tende a incentivar o professor a dar boas notas aos alunos, o que caracteriza um fenômeno chamado de *grade inflation* ou, em português, inflação de notas. Um dos problemas mais preocupantes nesses casos é a dificuldade em separar os alunos que merecem destaque dos demais.

A adequabilidade da técnica estatística ao tipo de dado proveniente das avaliações discentes é outro ponto importante. Medidas descritivas como média e desvio padrão não

são as mais adequadas a este tipo de dado. Isso porque a escala de medida é ordinal e não possui noção de distância, o que é necessário para o cálculo da média. Ou seja, se um aluno marca a opção 4 no item A e a opção 2 no item B, não significa que este aluno considera o quesito A duas vezes melhor que o quesito B, e por isso, não é adequado o uso de certas medidas nesses dados.

Para dados do tipo politômico, que é o caso das avaliações discentes, costuma-se utilizar modelos de regressão logística. E por isso, além de aplicar a Lei dos Julgamentos Categóricos, também foi ajustado um modelo de regressão logística em cada turma. Nas seis turmas a hipótese nula de que os β 's são iguais a zero foi rejeitada, o que significa que existe diferença significativa na forma com que os alunos avaliaram os itens. No entanto, nas seis turmas o teste para verificar a bondade do ajustamento, AIC, sugerem que o modelo não ficou bem ajustado aos dados. Sendo assim, o modelo de regressão logística pode não ser adequado para a análise das avaliações discentes. Em contrapartida, através do teste de bondade do ajuste para o modelo de Julgamentos Categóricos conclui-se que o modelo não ficou bem ajustado apenas nas turmas A e C. Nessas turmas deve-se tomar cuidado ao interpretar os resultados.

As duas técnicas apresentaram resultados bem parecidos com relação à ordenação dos estímulos. Uma vantagem de utilizar a Lei dos Julgamentos Categóricos é que ela gera todos os contrastes entre estímulos, enquanto na regressão logística os contrastes são feitos apenas em relação ao estímulo tomado como referência. Por outro lado, a análise de regressão logística permite explorar outras informações, como a razão de chances. Como o objetivo do presente trabalho era a aplicação da Lei dos Julgamentos Categóricos, não foi realizada uma análise detalhada da regressão logística para verificar qual das duas técnicas traz maiores benefícios. Isso pode ser estudado em trabalhos futuros.

Ao final deste trabalho é possível concluir que a Lei dos Julgamentos Categóricos é uma boa técnica para extrair informações a partir das avaliações discentes. Seus resultados são calculados comparando entre si os critérios de desempenho do professor. Dessa forma, ela permite que o professor saiba em quais itens está se destacando, tanto positivamente quanto negativamente. O que é mais adequado do que comparar determinado professor com outros professores, cujas avaliações podem ser influenciadas por fatores externos ao seu desempenho. Além disso, a Lei dos Julgamentos Categóricos é aplicável a dados que possuem escala ordinal, como é o caso das avaliações discentes.

Referências Bibliográficas

- [1] AGRESTI, A. **Categorical Data Analysis**. 2^a ed., Wiley, New York, 2002.
- [2] d'APOLLONIA, S.; ABRAMI, P. C. **Navigating student ratings of instructions**. American Psychologist, 52, p. 1198-1208, 1997.
- [3] GRIZZLE, J. E.; STAMER, C. F.; KOCK, G. G. **Analysis of Categorical Data by Linear Models**. Biometrics, 25, p. 489-504, 1969.
- [4] Johnson, V. E. **Grade Inflation - A Crisis in College Education**. Springer, 2003.
- [5] KHURI, A. I. **Advanced Calculus with Applications in Statistics**. Second Edition, Wiley, 2003.
- [6] LOVELAND, K. A. **Student Evaluation of Teaching (SET) in Web-based Classes: Preliminary Findings and a Call for Further Research**. Texas AM University, Corpus Christi, 2007.
- [7] MARSH, H. W. **Students' evaluations of university teaching: Research findings, methodological issues, and direction for future research**. International Journal of Educational Research, 11, p. 253-387, 1988.
- [8] McCULLAGH, P.; NELDER, J.A. **Generalized Linear Models**. Chapman and Hall, 2^a ed, 1989.
- [9] MILLEA, M.; GRIMES, P. W. **Grade Expectations and Student Evaluation of Teaching**. College Student Journal, 2002.
- [10] SELDLMEIER, P. **The Role of Scales in Student Ratings**. Learning and Instruction, 16, 2006.
- [11] SHEEHAN, E. P.; DuPREY, T. **Student evaluations of university teaching**. Journal of Instructional Psychology, 26, 188, 1999

- [12] SOUZA, G. S.; von BORRIES, G. F. **Modelo de Julgamentos Categóricos de Thurstone**. Relatório Técnico, Embrapa, 2000.
- [13] SOUZA, G.S. **The Law of Categorical Judgment Revisited**. Brazilian Journal of Statistics, 2001.
- [14] SOUZA, G.S.; VALES, R.C.C. **Análise de Dados Categóricos Ordinais e a Lei dos Julgamentos Categóricos de Thurstone - a Macro Thurst**. Dissertação: Departamento de Estatística da Universidade de Brasília, 2002.
- [15] SOUZA, J. **Métodos de Escalagem Psicossocial**. Thesaurus, Brasília, 1988.
- [16] THURSTONE, L. L. **A Law of Comparative Judgments**. Psychological Review, 34, 1927.
- [17] THURSTONE, L. L. **Theory and Methods of Scaling**. Wiley, New York, 1958.

Expansão de Taylor

Pelo Teorema de Taylor, se uma função $f(x)$ tem derivadas de ordem r , isto é,

$$f^{(r)}(x) = \frac{d^r}{dx^r} f(x)$$

existe então, para qualquer constante a , o polinômio de Taylor de ordem r sobre a dado por

$$T_r(x) = \sum_{i=0}^r \frac{f^{(i)}(a)}{i!} (x-a)^i$$

O Teorema de Taylor garante que o resto da aproximação ($f(x) - T_r(x)$), sempre tende a zero mais rápido que o termo de maior ordem, isto é, se

$$f^{(r)}(a) = \left. \frac{d^r}{dx^r} f(x) \right|_{x=a}$$

existe, então

$$\lim_{x \rightarrow a} \frac{f(x) - T_r(x)}{(x-a)^r} = 0$$

Para $r = 1$ tem-se a expansão de Taylor de primeira ordem. Dessa forma, a função $f(x)$ pode ser aproximada por $T_1(x) = f^{(1)}(a)(x-a)$, que representa a expansão de Taylor de primeira ordem em torno da constante a .

No caso estudado, é preciso obter uma aproximação para a função $g(\hat{\pi}_{ij})$. Através da expansão de Taylor de primeira ordem, esta função pode ser aproximada por

$$g'(\pi_{ij})(\hat{\pi}_{ij} - \pi_{ij})$$

Como exemplo, segue a expansão de $g(\hat{\pi}_{i1})$

$$g'(\boldsymbol{\pi}_{i1})(\hat{\boldsymbol{\pi}}_{i1} - \boldsymbol{\pi}_{i1})$$

onde $\boldsymbol{\pi}_{i1} = p_{i1}$, e portanto a aproximação pode ser escrita da forma

$$g'(\boldsymbol{\pi}_{i1})(\hat{p}_{i1} - p_{i1})$$

Para $g(\hat{\boldsymbol{\pi}}_{i2})$ tem-se

$$g'(\boldsymbol{\pi}_{i2})(\hat{\boldsymbol{\pi}}_{i2} - \boldsymbol{\pi}_{i2})$$

onde $\boldsymbol{\pi}_{i2} = p_{i1} + p_{i2}$, ou seja

$$g'(\boldsymbol{\pi}_{i2})(\hat{p}_{i1} - p_{i1}) + g'(\boldsymbol{\pi}_{i2})(\hat{p}_{i2} - p_{i2})$$

E assim sucessivamente, até $g(\hat{\boldsymbol{\pi}}_{im-1})$, aproximada por

$$g'(\boldsymbol{\pi}_{im-1})(\hat{\boldsymbol{\pi}}_{im-1} - \boldsymbol{\pi}_{im-1})$$

onde $\boldsymbol{\pi}_{im-1} = p_{i1} + p_{i2} + \dots + p_{im-1}$, e portanto pode ser escrita como

$$g'(\boldsymbol{\pi}_{im-1})(\hat{p}_{i1} - p_{i1}) + g'(\boldsymbol{\pi}_{im-1})(\hat{p}_{i2} - p_{i2}) + \dots + g'(\boldsymbol{\pi}_{im-1})(\hat{p}_{im-1} - p_{im-1})$$