



UNIVERSIDADE DE BRASÍLIA
DEPARTAMENTO DE ESTATÍSTICA

**Modelos Estatísticos para Avaliação da Qualidade
Culinária de Arroz: Textura e Propriedades
Viscoamilográficas**

Érica dos Santos Rios - 10/0010385

Brasília - DF

2015

ÉRICA DOS SANTOS RIOS - 10/0010385

Modelos Estatísticos para Avaliação da Qualidade Culinária de Arroz: Textura e Propriedades Viscoamilográficas

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
2015**

Agradecimentos

Este trabalho foi financiado pela bolsa PIBIC fornecida pela Embrapa Arroz e Feijão - CNPAF (Centro Nacional de Pesquisa em Arroz e Feijão), durante o período de 01/10/2014 a 31/07/2015.

Em primeiro lugar, agradeço a Deus por ter me dado a força e determinação necessárias para chegar até aqui. Sempre presente, Ele me ajudou a enfrentar os momentos de desânimo e a olhar para frente.

Em segundo lugar, agradeço a minha família por todo apoio financeiro e emocional concedidos a mim nessa jornada. Mesmo sem compreenderem a minha decisão de curso, eles foram meu alicerce e fonte de inspiração durante esses anos da graduação.

Agradeço também a Dr. Priscila Zaczuk Bassinello pela ajuda e pelos dados fornecidos, que possibilitaram a realização deste trabalho.

Por fim, agradeço ao Professor Dr. George Freitas von Borries, pelos diversos conhecimentos transmitidos a mim de forma tão sábia. Seu exemplo sempre me instiga a voar mais alto intelectualmente.

Resumo

A qualidade do arroz pode ser analisada através da avaliação de textura desse grão. O método mais indicado para se fazer essa análise é a avaliação sensorial, que é demorada e de alto custo, visto que envolve treinamento, aptidão e disponibilidade de pessoas. Por essa razão, este trabalho averiguou a possível substituição da avaliação sensorial verificando a relação entre medidas sensoriais de textura, medidas instrumentais de textura e medidas de viscosidade. Modelos capazes de prever a avaliação sensorial foram criados aplicando as técnicas estatísticas Análise de Componentes Principais e Regressão Logística Politômica. O grau de eficiência da previsão desses modelos foi obtido através da estimativa da taxa do erro de classificação por validação cruzada, e utilizando o recurso gráfico curva ROC. Os resultados mostraram que as medidas instrumentais de textura possuem relação consistente com as medidas de textura sensorial. De forma análoga, as medidas de viscosidade parecem permitir a previsão dos resultados obtidos por meio da avaliação sensorial de textura.

Palavras-chave: Textura de Arroz, Avaliação Sensorial, Regressão Logística Politômica, Componentes Principais, Medidas Instrumentais de Textura, Medidas de Viscosidade.

Sumário

1	Introdução	1
2	Metodologia	3
2.1	Avaliação da textura de arroz	5
2.2	Componentes Principais	12
2.3	Regressão Logística	15
2.4	Classificação e Discriminação	23
3	Resultados	29
3.1	Análise descritiva dos dados para o ano de 2013	29
3.2	Análise de Componentes Principais para o ano de 2013	35
3.3	Resultados da Regressão Logística	38
4	Conclusão	61
5	Considerações finais	63
	Referências Bibliográficas	65
	Apêndice	69
5.1	Tabelas e Figuras da Regressão Logística para o ano de 2013	69
5.2	Tabelas e Figuras da Regressão Logística para o ano de 2014	81
5.3	Tabelas e Figuras da Regressão Logística para o ano de 2013 unido ao de 2014	92
5.4	Tabelas e Figuras da previsão para o ano de 2014 utilizando os modelos de Regressão Logística obtidos por meio das observações de 2013	113

1 Introdução

O preço e a qualidade do arroz cozido que chega a casa dos consumidores é resultado de diversas características intrínsecas desse grão. Algumas dessas características são determinadas visualmente, como o tamanho, a coloração e a translucidez. Já outras podem ser percebidas sensorialmente, como é o caso do aroma e da textura [5, 17] .

A textura é a característica mais importante entre aquelas que podem ser percebidas sensorialmente. De acordo com a norma ISO 1990, textura é “o conjunto de propriedades mecânicas, geométricas e de superfície de um produto, detectáveis pelos receptores mecânicos e tácteis e, eventualmente pelos receptores visuais e auditivos”[10]. Para o caso específico do arroz, a textura é consequência da estrutura interna do grão e é determinada através do tato.

Em geral, a avaliação sensorial de textura do arroz é demorada e de alto custo, visto que envolve treinamento, aptidão e disponibilidade de pessoas. Assim, este trabalho procurou alternativas para a avaliação de textura. Utilizando medidas instrumentais de textura e medidas laboratoriais, denominadas medidas de perfil viscoamilográfico, buscou-se minimizar o tempo e o custo gastos nessa avaliação.

Essa problemática é parte do projeto sobre a qualidade dos grãos de arroz, denominado QualiArroz e desenvolvido pela Embrapa Arroz e Feijão (CNPAF), Santo Antônio de Goiás-GO. Portanto, este trabalho é resultado de parceria entre um grupo de pesquisadores da Embrapa CNPAF e o professor George von Borries do EST-UnB. Através da parceria, o trabalho de TCC foi financiado por uma bolsa de PIBIC do CNPq via Embrapa e orientado pelos doutores George von Borries (EST-UnB) e Priscila Zaczuk Bassinello (CNPAF-Embrapa). A parceria resultou num pôster apresentado no congresso XII Conferência Internacional do Arroz para América Latina e Caribe ([22]) e foi selecionado para apresentação oral no 9º Seminário Jovens Talentos da Embrapa Arroz e Feijão [23]. Além disso, um artigo está em fase de preparação para submissão.

Na primeira etapa do trabalho foi realizada uma análise descritiva dos dados. Em seguida, utilizando a técnica de componentes principais, foram exploradas as relações existente nas medidas de viscosidade, buscando dessa forma diminuir a dimensão do estudo. Depois, a regressão logística foi aplicada aos modelos que melhor representaram as relações de interesse. A determinação da textura foi feita utilizando componentes principais e métodos de classificação e discriminação.

A Embrapa CNPAF fornece estudos sobre o cultivo do arroz para diversos produtores do grão no Brasil. Por conseguinte, a motivação para este trabalho foi facilitar a tarefa dessa equipe, por meio do aprimoramento desses estudos que se deu mediante a exploração das relações existentes entre variáveis.

O objetivo desse trabalho foi facilitar e automatizar a análise de textura de arroz de terras altas e terrenos irrigados por meio da relação entre medidas sensoriais de textura, medidas instrumentais de textura e medidas de viscosidade.

Procurando assim, verificar a relação existente entre medidas de textura sensorial e instrumental; minimizar a quantidade de medidas de viscosidade que são necessárias para explicar as medidas de textura sensorial; comparar e identificar os modelos estatísticos que melhor representam a relação entre medidas de textura sensorial e medidas de viscosidade.

2 Metodologia

A base de dados deste trabalho foi fornecida pelos técnicos da Embrapa CNPAF. Os dados foram fornecidos em planilha Excel, lidos e analisados na versão 3.1.0 do software R, utilizando o ambiente de desenvolvimento integrado R-Studio. Os dados contém dezoito variáveis, dentre elas nove são quantitativas e nove são qualitativas.

As medidas de textura sensorial foram apresentadas por duas variáveis qualitativas, em escala ordinal, das quais uma delas é a avaliação sensorial de dureza dos grãos e a outra de pegajosidade. Sete das variáveis quantitativas fazem referência as medidas de perfil viscoamilosgráfico que fornecem informações a respeito da viscosidade e do teor de amilose aparente dos grãos. As medidas de textura instrumental foram apresentadas por duas variáveis quantitativas que são a avaliação instrumental de dureza e de pegajosidade. Sete variáveis qualitativas eram referentes a forma e a área de plantio do arroz, além da identificação do laboratório onde a análise laboratorial foi realizada. Dessas sete, apenas a variável tipo de terreno será utilizada nesse trabalho. Essa é uma variável dicotômica, de modo que, o arroz pode ser de Terrenos Irrigados ou de Terras Altas.

Após a análise descritiva desses dados, a primeira técnica estatística aplicada foi a de Componentes Principais. Essa técnica busca explicar a matriz de variância-covariância de um grupo de variáveis, através de suas combinações lineares. Ela permite uma interpretação mais completa das associações existentes nos dados, pois revela relações entre variáveis que eram inicialmente desconhecidas [8, p. 430]. Neste trabalho, Componentes Principais foi utilizada para a criação de modelos de regressão logística, reduzindo consideravelmente o número de variáveis explicativas nos modelos.

Para a modelagem das relações de interesse foi utilizado a técnica de Regressão Logística Politémica. As variáveis resposta desse estudo são as medidas de textura sensorial que são qualitativas ordinais. Conseqüentemente, o uso de regressão linear nos dados seria inapropriado, uma vez que essa técnica faz a suposição de normalidade dos resíduos, que

só é aceitável quando a variável resposta é quantitativa. As variáveis resposta dureza e pegajosidade possuem sete categorias cada, assumindo-se assim que as contagens para as categorias dessas variáveis tem distribuição multinomial. Por isso, a Regressão Logística que será utilizada é a politômica ao invés da binária, que só incorpora variáveis definidas em duas categorias.

Em seguida, será realizada a Análise de Discriminantes. Essa técnica é utilizada a fim de identificar e classificar um grupo de variáveis em populações previamente definidas [8, p. 575]. Para isso, ela utiliza os dados fornecidos para criar funções de discriminantes. Essas funções determinam a probabilidade que a variável resposta tem de pertencer a uma determinada categoria. Utilizando essas funções é possível reduzir custos gastos na classificação de observações, sejam esses de cunho financeiro, temporal ou de outro tipo.

A técnica de Classificação destaca as probabilidades que a variável resposta tem de pertencer a cada uma de suas categorias, para cada valor específico da(s) variável(is) explicativa(s). De modo que, a variável resposta é dita pertencente a uma determinada categoria quando essa categoria é a que possui a maior dessas probabilidades. É importante ressaltar que para alguns valores da(s) variável(is) explicativa(s) a probabilidade de pertencer a uma determinada categoria é muito próxima da de pertencer a outra, o que dificulta a classificação. Esses valores formam o que se denomina região de incerteza. Assim, um dos focos deste trabalho foi a diminuição da região de incerteza, de modo que o intervalo de valores da(s) variável(is) explicativa(s) para qual a classificação é incerta seja a menor possível.

2.1 Avaliação da textura de arroz

A avaliação da textura de arroz é feita separadamente para os dois tipos de terreno de plantio que estão sendo levados em consideração, sendo esses, Terras Altas e Terrenos Irrigados.

2.1.1 Avaliação sensorial da textura de arroz

A expansão da indústria de alimentos processados e industrializados na segunda metade do século XX elevou a preocupação com a qualidade do produto que é fornecido ao consumidor. Isso acarretou em um crescimento da avaliação sensorial, que é a ciência que avalia a qualidade de um produto por meio da interação desse com o ser humano [11, p. 1]. Essa avaliação, também conhecida como análise sensorial, se torna cada vez mais importante para a sobrevivência e crescimento de uma indústria no competitivo mercado alimentício. Stone e Sidel (2004) definem avaliação sensorial como um método científico que evoca, mensura, analisa, e interpreta as respostas a um produto que são percebidas pelos sentidos da visão, do olfato, do paladar, da audição e do tato [20, 11, p. 2].

No Brasil, a avaliação sensorial surgiu em 1967, no Instituto Agrônomo de Campinas. Utilizando métodos de diferença, foi inicialmente realizada apenas para grãos de café [15, p. 17]. As técnicas que compreendem a avaliação sensorial passaram por grandes mudanças e evoluções desde a sua criação até os dias de hoje. Devido a intensidade dessas mudanças, alguns autores dividem a evolução da avaliação sensorial em quatro fases distintas [15, p. 16:18].

- *Primeira (Até 1940)* : É caracterizada pela fase de produção artesanal dos alimentos dentro da indústria. Foi nessa época que se começou a pensar em avaliar um produto por meio dos sentidos humanos. Quando uma avaliação sensorial era feita, ela acontecia da forma mais conveniente possível. Como por exemplo, o avaliador era o dono ou um funcionário qualquer da indústria.
- *Segunda (1940-1950)* : É determinada pelo início da expansão da indústria de alimentos. Nessa época foi introduzido o conceito de controle de processo e de qualidade de um produto. Avaliações sensoriais começaram a ocorrer com mais frequência do que na primeira fase, porém ainda eram pouco planejadas e possuíam fraco ou quase nenhum embasamento científico.

- *Terceira (1950-1970)*: É caracterizada por um grande desenvolvimento e avanço em métodos de avaliação sensorial. Nessa fase foram propriamente definido os órgãos utilizados na avaliação e os atributos que podem ser percebidos por esses, como por exemplo o olho, esse órgão pode determinar atributos como o tamanho e a coloração do alimento. Percebeu-se uma necessidade de organização e buscou-se uma padronização das amostras apresentadas e das perguntas feitas ao avaliador, e do local da avaliação.
- *Quarta (Desde 1970)* : Nessa fase busca-se rever o conceito de qualidade sensorial de um alimento. Os procedimentos de avaliação sensorial de melhor aceitação tem sido descritos, além da escolha das técnicas estatísticas que melhor representam os dados provenientes dessa avaliação.

É necessário ressaltar que a indústria alimentícia de alguns países não passaram por essas fases nos anos citados. Apesar da maioria já se encontrar na quarta fase, ainda existem indústrias que estão na primeira, na segunda ou na terceira fase [15, p. 16:18].

Por ser uma avaliação feita por seres humanos, a avaliação sensorial é influenciada por fatores de personalidade, psicológicos e fisiológicos, dentre outros fatores que afetam diretamente e indiretamente a performance do avaliador. Um exemplo de fator de personalidade do avaliador é a intrepidez inerente a esse indivíduo, de modo que, o avaliador cauteloso informa apenas as diferenças das quais possui certeza, enquanto que o avaliador ousado informa todas as diferenças que ele acredita existir na amostra, até mesmo aquelas sobre as quais ele não está seguro[15, p. 20].

Os fatores fisiológicos são referentes a adaptação do organismo a um estímulo. Como avaliador, o indivíduo é exposto a um mesmo estímulo continuamente, o que pode acarretar mudança ou diminuição na sensibilidade podendo levar a avaliações errôneas. Dentre os fatores psicológicos que podem ocorrer, um deles tem relação com a motivação do avaliador, ou seja, se o indivíduo está devidamente motivado ou não para a realização da avaliação[15, p. 20:21].

A fim de melhorar os resultados obtidos por meio da avaliação sensorial, algumas medidas devem ser tomadas. A primeira delas diz respeito ao local de realização dos testes. Esse local deve ser desprovido de odor, de barulho excessivo e de estímulos visuais como texturas e cores fortes nas paredes. O local deve proporcionar isolamento e comodidade ao

avaliador, oferecendo uma cadeira suficientemente confortável e espaço satisfatório para o manuseamento da amostra e preenchimento do questionário[15, p. 23:25].

A segunda medida tem relação com a padronização da amostra dos alimentos. Deve-se uniformizar a apresentação das amostras de alimento, colocando sempre a mesma quantidade de alimento em recipientes uniformes entre si. Também é de extrema importância manter uma constância entre as temperaturas das amostras. Se necessário, utilizam-se luzes coloridas para camuflar uma característica da amostra a fim de avaliar um determinado atributo do alimento. Por exemplo, a utilização da luz vermelha para mascarar a coloração de um alimento quando essa influenciar indiretamente na resposta do avaliador a um determinado atributo do alimento [15, p. 29:31].

Avaliação sensorial da textura de arroz no Projeto QualiArroz

Na Embrapa Arroz e Feijão, os avaliadores recebem treinamento próprio para realizar a devida avaliação do arroz por meio do tato. As amostras de arroz são apresentadas a eles de forma monádica (uma de cada vez) e sequencial (uma após a outra). É possível visualizar na Figura (2.1) o local de condução do teste que é constituído de seis cabines, uma para cada avaliador. Essas cabines são brancas e possuem três luzes, uma de coloração branca, outra vermelha e outra azul. O preparador da amostra entrega a amostra para o avaliador por meio de um escotilha que se encontra na cabine. A comunicação entre esses dois indivíduos se dá por meio de uma luz que se encontra do lado de fora da cabine, de modo que, o avaliador aperta o interruptor que está dentro da cabine quando ele termina de analisar uma amostra, para que o preparador possa recolhê-la e entregar-lhe uma nova amostra.

Após avaliarem o arroz, os avaliadores preenchem um questionário com dois itens sobre a textura do arroz avaliado. Esses itens apresentam uma escala de 7 pontos que variam gradativamente nos atributos pegajosidade e dureza do arroz. Para o caso da dureza, as sete categorias são: Extremamente Macio, Macio, Ligeiramente Macio, Macio com Centro Firme, Levemente Firme, Muito Firme e Extremamente Firme. Para a pegajosidade, as sete categorias são: Extremamente Solto, Muito Solto, Solto, Ligeiramente Solto, Pegajoso, Muito Pegajoso, Extremamente Pegajoso. Na avaliação desses dois atributos utiliza-se somente a amostra de arroz sob a luz de coloração branca.



Figura 2.1 – Local de condução da avaliação sensorial na Embrapa Arroz e Feijão

A avaliação sensorial ainda é o método mais adequado para a avaliação da textura de arroz, visto que não existem equipamentos que possuam a sensibilidade do ser humano. Entretanto, esse método possui algumas desvantagens, como a necessidade de se encontrar avaliadores com aptidão e disponibilidade de tempo para a grande quantidade de testes que são necessários para a Embrapa. Outra desvantagem é a limitação da quantidade de amostras que podem ser analisadas por dia, a fim de não se causar fadiga ao avaliador.

2.1.2 Medidas Instrumentais da textura de arroz

Assim como outros atributos do arroz que são medidos por meio da avaliação sensorial, a textura não é uma característica intrínseca do grão, mas sim resultado da interação desse com o ser humano. Devido a necessidade de um quantidade considerável de avaliadores e do tempo gasto na realização de uma boa avaliação sensorial, já foram criados equipamentos que buscam determinar a textura, tarefa essa que por definição somente o homem é capaz de realizar. A utilização desses equipamentos permite que um maior número de testes sejam feitos por dia, obtendo maior repetibilidade e precisão dos dados.

Um desses equipamentos é o Texturômetro que analisa os dois parâmetros principais da textura (pegajosidade e dureza) por meio da resistência à deformação que um alimento apresenta. A utilização desse equipamento mais analítico e preciso procura extin-

guir os fatores subjetivos pertinentes ao ser humano, como é o caso dos fatores psicológicos e fisiológicos [4, p. 2].

Medidas instrumentais da textura de arroz no Projeto QualiArroz

Em particular, a Embrapa Arroz e Feijão utiliza o procedimento otimizado descrito por Sesmat & Meullennet (2001) para a determinação dos parâmetros de dureza e pegajosidade de arroz cozido em Texturômetro (TA.XT.plus, Stable Micro Systems, Godalming, Surrey, UK) [19, 4, p. 2]. O texturômetro faz a avaliação de compressão uniaxial utilizando uma sonda que comprime a amostra de alimento, registrando em um gráfico a extensão da deformação e/ou resistência do material analisado. Para essa avaliação é necessário que o diâmetro da sonda seja igual ou superior ao diâmetro da amostra [4, p. 2].

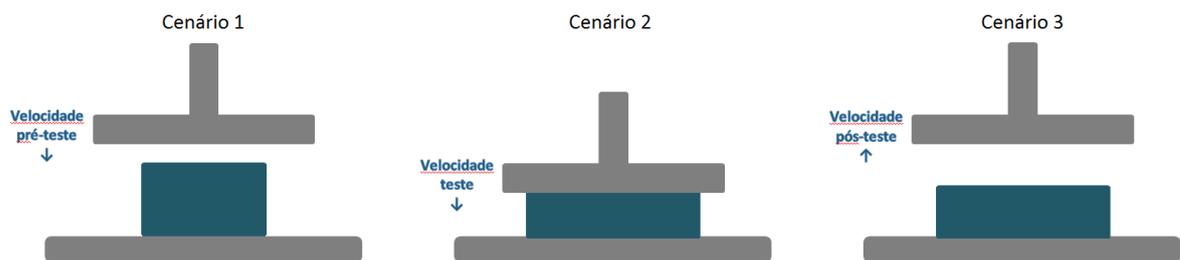


Figura 2.2 – análise de compressão

No cenário 1, a sonda que inicialmente se encontra a uma velocidade de pré-teste vai abaixando em direção a plataforma de avaliação do Texturômetro onde se encontra a amostra de arroz. Isso ocorre até que a sonda atinja uma força chamada “trigger” que é a evidência de que a sonda entrou em contato com a amostra. Após o registro da “trigger” a sonda comprime a amostra até determinada altura quando muda da velocidade de compressão pré-teste para a velocidade de teste. De modo que, a dureza instrumental é a força máxima, medida em Newton, registrada durante a análise de compressão no cenário 2. Já a pegajosidade instrumental é dada pela energia de adesão medida após a compressão de uma amostra, durante a volta da sonda à sua posição inicial no cenário 3 [4, p. 2:3]. No banco de dados e durante este trabalho, a dureza instrumental foi abreviada por DUREZAT. Já a pegajosidade instrumental foi abreviada por PEGAJT.

2.1.3 Medidas Instrumentais de viscosidade

O amido corresponde a até 95% da matéria prima do arroz sem casca, por isso ele é o componente mais explorado no estudo do arroz. Comparado a outros grãos como o milho e o trigo, a disponibilidade de amido no arroz é pequena devido ao tamanho do grão e a tecnologia de custo elevado para a extração de amido desse [3, p. 17].

Grande parte das propriedades físicas do arroz, principalmente seu comportamento no cozimento, pode ser explicada pela amilose, que corresponde de 20 a 30% do amido encontrado no grão [3, p. 18]. As medidas instrumentais de viscosidade buscam medir as características essenciais da amilose.

Medidas de viscosidade no Projeto QualiArroz

O processamento de amostragem dos grãos de arroz é feito logo após a colheita, debulha e secagem natural dos grãos. Aqueles grãos que ainda permanecerem com a casca são processados no moinho de provas, marca Suzuki, modelo MT 10mill (Santa Cruz do Rio Pardo, São Paulo, Brazil)[16, p. 202]. De modo que os grãos estejam prontos para a mensuração de características inerentes a esse.

A primeira medida de viscosidade presente no banco de dados é a TAAFIA, que é o teor de amilose aparente dos grãos determinada pelo Sistema FIA (Análise por Injeção de Fluxo) da Foss Tecator (FIASStar 5000, Dinamarca). As amostras de arroz, previamente moídas em moinho de facas Perten Laboratory Mill 3100, foram injetadas após completa dispersão e gelatinização em solução alcalina, e a absorbância do complexo formado com solução de iodo foi determinada em espectrofotômetro Uv-Visível através de um detector digital “Dual-Wavelength (DDW)” a 720 nm. O conteúdo de amilose das amostras foi calculado por meio de uma curva de calibração, preparada com cultivares padrão de arroz pré-selecionadas com teores de amilose conhecidos e determinados previamente por Cromatografia de Permeação em Gel (SEC/GPC) (FITZGERALD; McCOUCH; HALL, 2009) pelo IRRI (International Rice Research Institute) [6].

Outra medida de viscosidade é a TAASEC, que é o teor de amilose absoluto dos grãos determinada pelo Sistema de Cromatografia Líquida de Alta Eficiência (HPLC) Prominence (Shimadzu, Kyoto, Japão) acoplado com o detector de Índice de Refração, conforme metodologia de FITZGERALD, McCOUCH e HALL, (2009) [6]. As farinhas de arroz liofilizadas (50,0 mg) foram inicialmente gelatinizada, e depois adicionou 206 μ L de

solução tampão (5,0 mL de acetato de sódio 0,2 Mol/L, pH 4,0 e 180 mL de ácido acético glacial). A desramificação enzimática foi realizada conforme Lisle, Martin e Fitzgerald (2000) e Batey e Curtin (1996) [2]. Os parâmetros da fase móvel para separação foram constituídos pelo eluente A (acetato de amônio 0,05 mol/L + 0,05% Azida sódica, pH 4,75, filtrada), eluente B (solução azida sódica 0,05%, filtrada), eluente C (água Milli-Q filtrada) e eluente D (água Milli-Q filtrada). Utilizou-se um conjunto de guarda coluna e coluna UH250 da Waters (ultrahydrogel, SEc 250), calibrado com padrões de pululanos (Shodex Corporation, Kyoto, Japão), sendo o fluxo da coluna de 0,5 mL/min. Injetou 40 μ L de cada amostra no injetor automático SIL-20A, diminuindo-se o fluxo da coluna para 0,02 mL/min. Os dados foram coletados e analisados usando-se o programa LCsolution e, por meio de batch Processing, foi gerado o perfil cromatográfico de cada amostra.

A medida de viscosidade TG, que é a temperatura de gelatinização determinada indiretamente pelo teste de dispersão alcalina que é conduzido de acordo com metodologia apresentada por Martinez e Cuevas (1989) [13, p. 202]. Dez grãos (inteiros, sadios e polidos) de cada amostra foram distribuídos uniformemente em uma placa plástica de 4,8 cm de diâmetro, contendo 10 mL da solução de hidróxido de potássio (KOH) 1,7%. As placas foram tampadas e incubadas em estufa (FISHER, modelo 255G, Waltham, USA), à 30 °C, por 23 horas. A dispersão na amostra dos grãos é classificada em uma escala numérica de 1 ao 7. Após o teste, se obtém a temperatura de gelatinização de cada amostra multiplicando-se o número de grãos de arroz apresentados naquela amostra pelo seu correspondente nível de dispersão alcalina [16, p. 202].

As propriedades de pasta (perfil viscoamilográfico) das amostras das farinhas de arroz foram determinadas em Rapid Visco Analyser (RVA) (RVA4, Newport Scientific, Austrália), no qual foi utilizada uma suspensão da amostra moída (3g em 25 mL) corrigida para 14% de umidade e analisada de acordo com o seguinte regime de tempo/temperatura: 25°C por 2 minutos, aquecimento de 25°C a 95°C a uma taxa de 14°C/min, manutenção da pasta a 95°C por 3 minutos e resfriamento de 95°C a 25°C a uma taxa de 14°C/min, num total de 12,5 minutos de análise. Essas propriedades são definidas por (Teba, C.S) [21, p. 413] e são apresentadas na Figura (2.3). Das cinco medidas de perfil viscoamilográfico apresentadas na Figura (2.3), quatro estão no banco de dados e serão utilizadas neste trabalho. São essas PEAK, BREAKDOWN, FINAL e SETBACK.

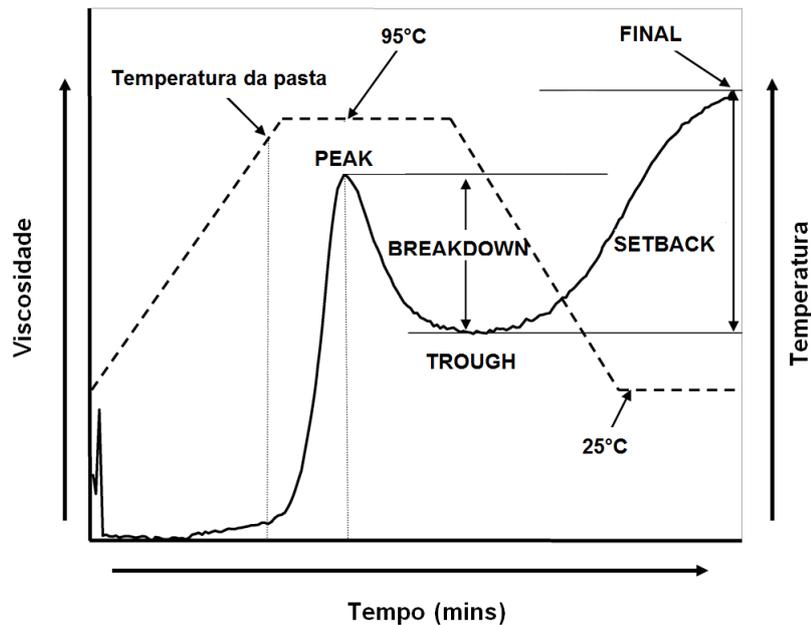


Figura 2.3 – Medidas de perfil viscoamilográfico

Nota: Baseado em Saunders, 2010 [18]

- *Viscosidade de pasta máxima (PEAK)*: é o maior valor da viscosidade durante o ciclo de aquecimento, que é obtido no ponto máximo da curva apresentada na Figura (2.3)
- *Viscosidade de pasta mínima à quente (TROUGH)*: é o menor valor da viscosidade durante os 3 minutos em que a temperatura é mantida constante a 95°C.
- *Quebra de viscosidade (BREAKDOWN)*: é a diferença entre a viscosidade de pasta máxima e a viscosidade de pasta mínima à quente.
- *Viscosidade final (FINAL)*: é o valor final da viscosidade durante o ciclo de resfriamento, que se dá a 25°C.
- *Tendência à retrogradação (SETBACK)*: é a diferença entre a viscosidade final e a viscosidade de pasta mínima à quente.

2.2 Componentes Principais

Um conjunto de dados onde cada coluna representa uma variável e cada linha se refere a um mesmo indivíduo ou uma mesma observação, pode ser visto como uma matriz

com m colunas e n linhas, onde m é o número de variáveis do banco de dados e n é o número de linhas, ou seja, o número de observações para cada variável. Essa matriz é dada por:

$$X_{n \times m} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nm} \end{bmatrix} = [X_1 \quad X_2 \quad \cdots \quad X_m], \quad (2.1)$$

onde X_j é o vetor que constitui a j -ésima variável, para $j = 1, \dots, m$ [8, p. 5].

A partir dessa matriz, é possível obter o vetor de média desses dados:

$$\bar{X} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_m \end{bmatrix}, \quad (2.2)$$

onde

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij}, \quad (2.3)$$

para $j = 1, \dots, m$ [8, p. 6:7].

Esse vetor de médias constitui uma outra medida resumo importante, que mostra a dispersão entre os valores de uma variável (variância) e entre os pares de valores entre 2 variáveis (covariância). A matriz de variância-covariância é dada por:

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & s_{22} & \cdots & s_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nm} \end{bmatrix}, \quad (2.4)$$

onde para $a \neq b$,

$$s_{ab} = s_{ba} = \frac{1}{n} \sum_{i=1}^n (X_{ij} - \bar{X}_a)(X_{ij} - \bar{X}_b) \quad (2.5)$$

para $a, b = 1, \dots, m$. Em particular, para $a = b$,

$$s_{ab} = s_a^2 = s_b^2 = \frac{1}{n} \sum_{i=1}^n (X_{ij} - \bar{X}_a)^2 = \frac{1}{n} \sum_{i=1}^n (X_{ij} - \bar{X}_b)^2, \quad (2.6)$$

[8, p. 7:8].

A fim de padronizar as escalas nas quais as diferentes variáveis são medidas, calcula-se a matriz de correlação R dividindo cada s_{ab} por suas respectivas variâncias s_a^2 e s_b^2 . De modo que:

$$R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{bmatrix}, \quad (2.7)$$

onde para $a \neq b$,

$$r_{ab} = \frac{s_{ab}}{s_a^2 s_b^2}. \quad (2.8)$$

Já para o caso $a = b$,

$$r_{ab} = \frac{s_a^2}{s_a^2} = \frac{s_b^2}{s_b^2} = 1, \quad (2.9)$$

[8, p. 8].

Essa padronização possibilita uma comparação direta entre os pares de variáveis, considerando que o par cujo coeficiente de correlação r_{ab} está mais perto de 1 possui duas variáveis que estão mais associadas entre si, em relação a pares que possuem r_{ab} mais perto de 0.

Utilizando a propriedade de simetria da correlação, que afirma que $r_{ab} = r_{ba}$, é possível observar que a matriz de correlação R é simétrica e quadrada pois tem dimensão $n \times m$. Isso permite o cálculo dos denominados autovalores para cada variável, que são representados pelo vetor λ que satisfaz a equação característica:

$$|R - \lambda I| = 0, \quad (2.10)$$

onde R é a matriz de correlação,

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{bmatrix} \quad e \quad I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \quad (2.11)$$

[8, p. 97].

Dada a matriz dos dados $X = [X_1 \ X_2 \ \dots \ X_m]$, que possui matriz de correlação R e cujos autovalores estão organizados de modo que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$, é possível calcular as seguintes combinações lineares (Y_j) das variáveis X_j , para $j = 1, \dots, m$:

$$Y_1 = c_1'X = c_{11}X_1 + c_{12}X_2 + \dots + c_{1m}X_m, \quad (2.12)$$

$$Y_2 = c_2'X = c_{21}X_1 + c_{22}X_2 + \dots + c_{2m}X_m, \quad (2.13)$$

$$\vdots \quad (2.14)$$

$$Y_m = c_m'X = c_{m1}X_1 + c_{m2}X_2 + \dots + c_{mm}X_m, \quad (2.15)$$

[8, p. 431].

Por conseguinte, as componentes principais são os vetores Y_j cujos c_i 's maximizam a variância e covariância dessas combinações lineares, que são obtidas por:

$$Var(Y_i) = c_i' R c_i, \quad (2.16)$$

$$Cov(Y_a, Y_b) = c_a' R c_b, \quad (2.17)$$

[8, p. 431].

Em outras palavras, as componentes principais são combinações lineares das variáveis originais, obtidas rotacionando-se os eixos dessas variáveis de modo a maximizar a variabilidade das componentes. Logo, a primeira componente explica uma maior variação dos dados da matriz X do que a segunda componente e assim por diante. As componentes principais permitem reduzir a dimensão do estudo, pois possibilitam a utilização de p componentes sendo $p < m$, onde m é o número de variáveis originais [8, p. 430:431].

2.3 Regressão Logística

Para entender um determinado evento e definir as possíveis circunstâncias que estão associadas a esse; primeiramente é necessário ajustar um modelo estatístico que melhor represente as possíveis relações entre uma variável resposta, que corresponde ao evento, e suas covariáveis, que são as circunstâncias. Os modelos estatísticos denominados modelos lineares generalizados podem ser divididos em três componentes apresentadas abaixo [1, p. 66].

- *Componente aleatória:* identifica a variável resposta Y , ou seja, a variável de interesse do problema. Determina a que distribuição de probabilidade essa variável pertence ou a qual ela melhor se ajusta.
- *Componente sistemática:* define as covariáveis do problema, ou seja, as variáveis que possivelmente estão associadas com a variável de interesse. A combinação linear dessas ocupa o lado direito da equação do modelo e é chamada de **preditor linear**. Considerando m variáveis explicativas (x_i 's), o preditor linear é dado por:

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m, \quad (2.18)$$

onde β_0 é o intercepto e β_i para $i = 1, \dots, m$ representa o efeito causado pela covariável x_i no modelo. Ao estimar os parâmetros β_i 's, o software R utiliza um preditor linear alternativo, dado por:

$$\beta_0 - \beta_1 x_1 - \beta_2 x_2 - \cdots - \beta_m x_m, \quad (2.19)$$

portanto durante este trabalho, utilizar-se-á a versão alternativa do preditor linear na definição e análise de modelos.

- *Função de ligação:* especifica a função matemática que melhor relaciona a média da variável resposta, que é dada pela esperança dessa: $\mu = E(Y)$, com o preditor linear. Logo o modelo generalizado linear é dado por:

$$g(\mu) = \beta_0 - \beta_1 x_1 - \beta_2 x_2 - \cdots - \beta_m x_m, \quad (2.20)$$

onde $g(\cdot)$ é a função de ligação que une a componente aleatória a sistemática.

O modelo linear generalizado mais difundido é aquele que assume distribuição normal para a variável resposta Y . A componente sistemática é composta de covariáveis quantitativas contínuas e a função de ligação é a identidade, ou seja, $g(\mu) = \mu$, então o modelo estatístico é dado por:

$$\mu = \beta_0 - \beta_1 x_1 - \beta_2 x_2 - \cdots - \beta_m x_m \quad (2.21)$$

[1, p. 67]. Esse modelo é conhecido como modelo ordinário de regressão linear e é de fácil estimação e interpretação, pois os β_i 's podem ser estimados pelo método dos mínimos quadrados.

O modelo linear generalizado que será usado nesse trabalho é o modelo de regressão logística que possui as seguintes componentes:

- *Componente aleatória:* formada por uma variável resposta Y que segue uma distribuição binomial. Em outras palavras, considerando n realizações, a probabilidade de ocorrer y sucessos no evento Y é dada por:

$$P(Y = y) = \frac{n!}{y!(n-y)!} \pi(y)^y (1 - \pi(y))^{n-y}. \quad (2.22)$$

para $y = 1, \dots, n$. Onde $\pi(y)$ é a probabilidade de ocorrer o evento, ou seja, obter sucesso no evento considerado.

- *Componente sistemática:* constituída de covariáveis que podem ser quantitativas contínuas ou categorizadas.
- *Função de ligação:* é a função logito de $\pi(y)$, ou seja $g[\pi(y)] = \log\left(\frac{\pi(y)}{1+\pi(y)}\right)$ [1, p. 71].

Então, o modelo de regressão logística é dado pela equação:

$$\log\left(\frac{\pi(y)}{1 + \pi(y)}\right) = \beta_0 - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_m x_m. \quad (2.23)$$

É possível que nem todas as m covariáveis definidas no modelo sejam utilizadas. Para averiguar quais delas possuem efeito significativo na estimação da variável resposta, realiza-se o teste de hipótese Wald descrito a seguir, para cada uma das $i = 1, \dots, m$ covariáveis [1, p. 84].

Teste de Wald:

- Hipóteses

As hipóteses do teste de Wald são:

$$\begin{cases} H_0) & \beta_i = 0 \\ H_1) & \beta_i \neq 0 \end{cases}$$

que podem ser reescritas como:

$$\left\{ \begin{array}{l} H_0) \text{ A covariável } x_i \text{ não exerce influencia significativa no modelo, mantidas} \\ \text{ constantes as outras covariáveis.} \\ H_1) \text{ A covariável } x_i \text{ exerce influencia significativa no modelo, mantidas} \\ \text{ constantes as outras covariáveis.} \end{array} \right.$$

- Estatística do teste

A estatística do teste de Wald é dada por:

$$z^2 = \frac{\hat{\beta}_i^2}{\hat{\sigma}_i^2} \quad (2.24)$$

onde $\hat{\sigma}_i^2$ é a estimativa da variância do valor estimado do parâmetro β_i ($\hat{\beta}_i$). A estatística z^2 possui uma distribuição aproximadamente qui-quadrada com 1 grau de liberdade.

- Decisão

A fim de aceitar ou não a hipótese nula (H_0), calcula-se o p-valor da estatística do teste, que é equivalente a probabilidade de se obter um valor mais extremo do que a estatística do teste z^2 , ou seja:

$$p - \text{valor} = P(\chi_1 > z^2) \quad (2.25)$$

Para um nível de significância geral α previamente definido, obtém-se um nível de significância particular α/m para cada teste de cada β_i . Se ($p - \text{valor} > \alpha/m$), então não há evidências suficientes para rejeitar a hipótese nula e a covariável x_i não deve ser considerada no modelo. Caso contrário, se ($p - \text{valor} < \alpha/m$), rejeita-se a hipótese nula e a covariável x_i deve ser considerada no modelo.

Para exemplificar o modelo de regressão logística, considere o estudo da pegajosidade do arroz com apenas duas categorias (pegajoso e solto) podemos definir a variável de interesse Y como sendo **arroz solto (1-sim ou 0-não)**. Então, o evento Y segue uma distribuição binomial e ocorre sucesso quando uma observação é definida como solto.

Para o arroz de terras altas, a componente sistemática será dada pelo *Score1*, que é a variável formada através da primeira componente principal das variáveis de perfil viscoamilográfico. O modelo considerado é dado por:

$$\log\left(\frac{\pi(y)}{1-\pi(y)}\right) = \beta_0 - \beta_1 \text{Score1}. \quad (2.26)$$

Logo, a probabilidade estimada de se obter um arroz solto é dada pela equação:

$$\hat{\pi}(y) = \left(\frac{e^{\hat{\beta}_0 - \hat{\beta}_1 \text{Score1}}}{1 + e^{\hat{\beta}_0 - \hat{\beta}_1 \text{Score1}}}\right). \quad (2.27)$$

Para estimar o β_1 utiliza-se o método de máxima verossimilhança. A função de máxima verossimilhança para o modelo de regressão logística é dado por:

$$L = \prod_{i=1}^l \pi(y) \prod_{i=n-l+1}^n [1 - \pi(y)], \quad (2.28)$$

onde l é o número de observações classificadas como solto e n é o número total de observações.

Aplicando logaritmo se obtém:

$$\mathfrak{l} = \sum_{i=1}^l \log[\pi(y)] + \sum_{i=n-l+1}^n \log[1 - \pi(y)]. \quad (2.29)$$

Diferentemente do modelo ordinário de regressão linear, os β_i 's não podem ser estimados através de uma fórmula fechada. Então após derivar a Equação ((2.29)) e igualar a zero é necessário a utilização de um processo de estimação iterativo, que utiliza o método de cálculo numérico Newton-Raphson [1, p. 88].

O ajuste de uma reta de regressão de um modelo ordinário de regressão linear para os dados do exemplo apresentados na Figura 2.4 seria impróprio, pois isso causaria uma inflação dos resíduos. Além disso haveria a possibilidade da probabilidade estimada ultrapassar o intervalo (0, 1) [1, p. 68]. Por isso se utiliza a função logito na hora de prever os valores de uma regressão logística.

Na Figura 2.5, mostra-se que os valores preditos da probabilidade do evento Y ajustados através de uma regressão logística têm formato de s e não ultrapassam os valores

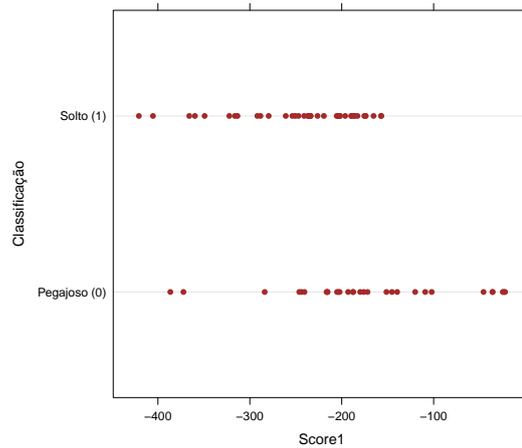


Figura 2.4 – Gráfico de dispersão da classificação de pegajosidade de acordo com o *Score1*, que é o escore formado pela primeira componente das variáveis de perfil viscoamilográfico.

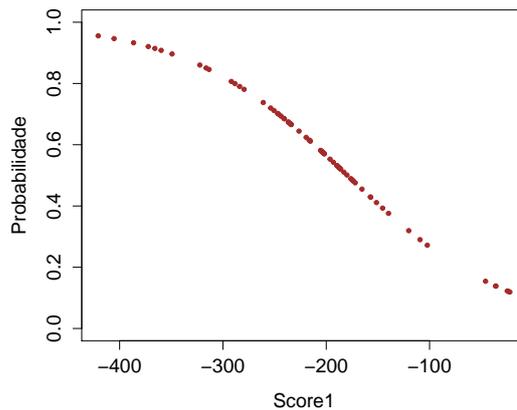


Figura 2.5 – Gráfico da probabilidade de classificar o arroz como solto de acordo com o *Score1*, que é o escore formado pela primeira componente das variáveis de perfil viscoamilográfico

previstos para uma probabilidade que é entre (0, 1). Na Figura 2.5 também revela-se que quanto menor o valor do *Score1*, maior a probabilidade do arroz ser considerado solto.

2.3.1 Politômica

As variáveis resposta dureza e pegajosidade possuem inicialmente sete categorias cada. Por esse motivo, as contagens para as categorias dessas variáveis tem distribuição multinomial, ao invés do modelo binário mais simplista que considera apenas duas

categorias. Posteriormente, algumas dessas sete categorias irão se fundir buscando uma melhor adequabilidade do modelo, de modo que, havendo duas categorias utilizar-se-á a regressão logística binária, e para o caso onde há mais de duas categorias utilizar-se-á a regressão logística politômica.

Hosmer e Lemeshow (2000) apresentam os três tipos de modelos mais utilizados quando as categorias da variável resposta apresentam uma certa ordem, ou seja, são quantitativas ordinais [7, p. 288-291]. Logo como as variáveis resposta desse estudo apresentam categorias que aumentam o nível do atributo, dureza ou pegajosidade, de forma gradual, logo elas são quantitativas ordinais. Os três tipos de modelo são:

- *Modelo da categoria adjacente:* compara cada categoria k com a categoria anterior $k - 1$, o modelo é dado por:

$$\log \left(\frac{\pi_k(y)}{\pi_{k-1}(y)} \right) = \tau_k + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m, \quad (2.30)$$

onde τ_k é o intercepto da k -ésima categoria e β_i para $i = 1, \dots, m$ representa o efeito causado pela covariável x_i no modelo.

- *Modelo da razão contínua:* compara cada categoria k com todas as categorias anteriores, o modelo é dado por:

$$\log \left(\frac{P(Y = k | x_1, x_2, \dots, x_m)}{P(Y < k | x_1, x_2, \dots, x_m)} \right) = \log \left(\frac{\pi_k(y)}{\pi_1(y) + \pi_2(y) + \cdots + \pi_{k-1}(y)} \right) = \theta_k + \beta_1 x_1 + \beta_2 x_1 + \cdots + \beta_m x_m, \quad (2.31)$$

onde θ_k é o intercepto da k -ésima categoria e β_i para $i = 1, \dots, m$ representa o efeito causado pela covariável x_i no modelo.

- *Modelo de chances proporcionais:* compara todas as categorias anteriores e equivalente a categoria k com todas as categorias acima, o modelo é dado por:

$$\log \left(\frac{P(Y \leq k | x_1, x_2, \dots, x_m)}{P(Y > k | x_1, x_2, \dots, x_m)} \right) = \log \left(\frac{\pi_1(y) + \pi_2(y) + \cdots + \pi_k(y)}{\pi_{k+1}(y) + \pi_{k+2}(y) + \cdots + \pi_K(y)} \right) = \beta_{0k} - \beta_1 x_1 - \beta_2 x_1 - \cdots - \beta_m x_m, \quad (2.32)$$

onde β_{0k} é o intercepto da k -ésima categoria e β_i para $i = 1, \dots, m$ representa o efeito causado pela covariável x_i no modelo. K é o número total de categorias do modelo.

Este último modelo apresentado também chamado de modelo logito cumulativo será o tipo utilizado neste trabalho. Exponencializando a Equação (2.32) obtém-se a probabilidade estimada da variável resposta pertencer a categoria k ou a uma categoria inferior a essa:

$$P(Y \leq k | x_1, x_2, \dots, x_m) = \frac{\exp(\beta_{0k} - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_m x_m)}{1 + \exp(\beta_{0k} - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_m x_m)}. \quad (2.33)$$

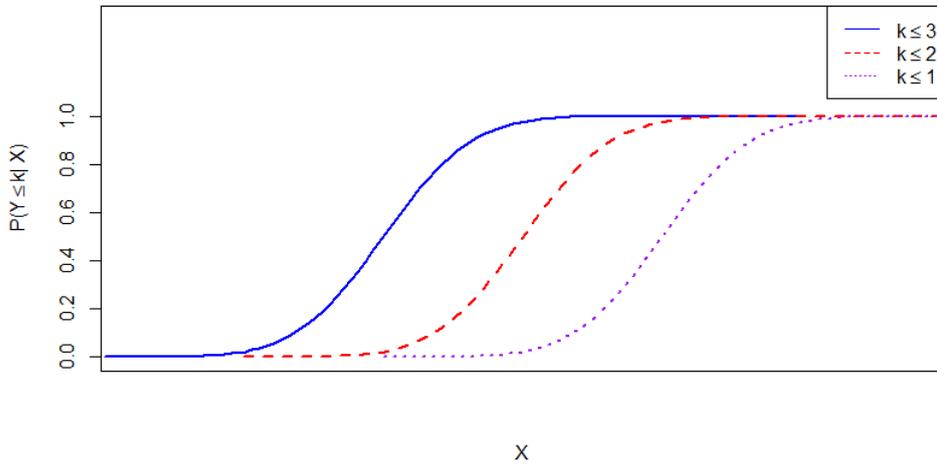


Figura 2.6 – Probabilidade estimada de se obter a variável resposta pertencente a categoria k ou a uma categoria inferior a essa, considerando diferentes valores da variável explicativa X .

Nota: Baseado em Agresti, 2007 [1, p. 181]

A Figura 2.6 é uma representação da Equação (2.33) para um exemplo onde o número total de categorias é $K = 4$ e só se tem uma variável explicativa X . É possível notar nessa figura que a inclinação das três curvas é a mesma, indicando que β_1 permanece o mesmo independente do k considerado.

Obtém-se a probabilidade específica de cada categoria do exemplo da Figura 2.6 através das Equações (2.34, 2.35, 2.36, 2.37)

$$P(Y = 1 | X) = P(Y \leq 1 | X), \quad (2.34)$$

$$P(Y = 2|X) = P(Y \leq 2|X) - P(Y \leq 1|X), \quad (2.35)$$

$$P(Y = 3|X) = P(Y \leq 3|X) - P(Y \leq 2|X), \quad (2.36)$$

$$P(Y = 4|X) = 1 - P(Y \leq 2|X). \quad (2.37)$$

As Equações (2.34,2.35,2.36,2.37) são representadas na Figura 2.7.

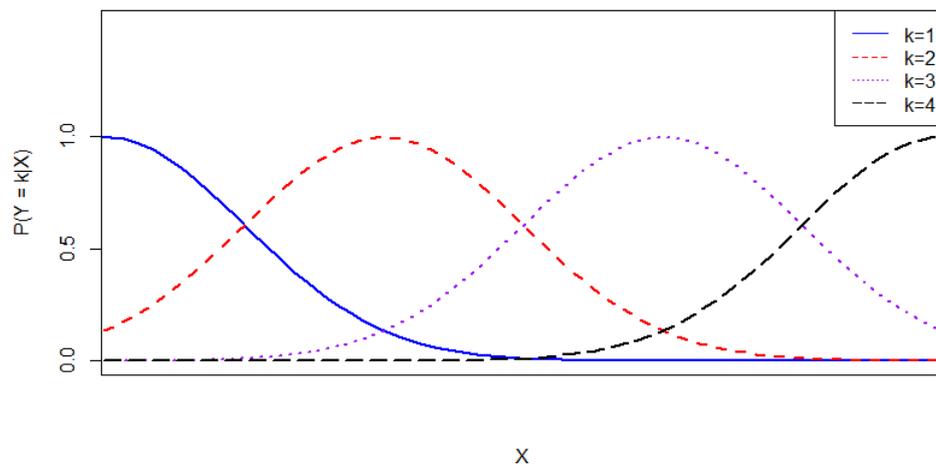


Figura 2.7 – Probabilidade estimada de se obter a variável resposta pertencente a categoria k , considerando diferentes valores da variável explicativa X .

Nota: Baseado em Agresti, 2007 [1, p. 181]

2.4 Classificação e Discriminação

Classificação e Discriminação são técnicas multivariadas, que buscam separar grupos de observações distintos entre si e alocar novas observações em grupos previamente definidos [8, p. 575].

O modelo da regressão logística pode ser considerado uma função de classificação, função utilizada para separar as observações em grupos pré-determinados. É através desse modelo que se calculam os valores preditos dados por $P(Y_i = k|x_{i1}, x_{i2}, \dots, x_{im})$ para cada observação i . Esses valores são os utilizados para separar os grupos.

Para o caso da regressão binária, há duas categorias possíveis, como por exemplo A e B , sendo que A representa a ocorrência de um certo evento e B a não ocorrência do

mesmo. Em primeiro lugar, determina-se um ponto de corte c , ou seja, um valor entre 0 e 1 para o qual valores preditos abaixo de c sejam classificados como B e acima como A . Em geral, utiliza-se 0,5, mas se há conhecimento prévio de que esse número deva ser maior ou menor, c pode variar, sendo por exemplo 0,2 ou 0,6. Ao longo desse trabalho utiliza-se $c = 0,5$ [9, p. 348].

Para o caso da regressão politômica, há K categorias possíveis, como por exemplo A,B,C e D para $K = 4$. Para cada observação tem-se quatro valores preditos, um para cada categoria. Portanto, a fim de classificar cada observação não é necessário um ponto de corte, é necessário apenas verificar qual das categorias apresenta o maior valor predito e classifica-lá como pertencente a essa.

Entretanto, observando a Figura 2.7 pode se visualizar que essa escolha da categoria nem sempre é fácil ou exata. Para alguns valores da variável explicativa, perto do pico de cada curva é possível escolher com certeza a categoria. No entanto, perto do local onde duas curvas se cruzam, é difícil escolher uma das duas categorias com certeza, afinal a diferença entre os valores preditos das duas é bem pequena.

Se a verdadeira origem da população for conhecida, ou seja, se antes da modelagem estava definido a qual categoria em que a observação pertence, é possível saber o quão eficiente a função de classificação está sendo em classificar as observações em suas verdadeiras categorias. Para tal, calcula-se a taxa do erro de classificação por validação cruzada, que é a porcentagem de observações classificadas em categorias erradas, que não correspondem a sua população de origem, sobre o número total de observações classificadas.

Considerando um número total de observações n , o termo validação cruzada se refere ao método de utilizar $n - 1$ observações na criação de um modelo preditivo e utilizar esse modelo para prever a observação que não havia sido utilizada na criação do modelo. Repete-se essa operação n vezes até que todas as observações sejam previstas.

2.4.1 Curva ROC

A curva ROC (Receiver Operating Characteristic) é uma representação gráfica da eficiência preditiva do modelo utilizado para a classificação de observações, em outras palavras, a curva ROC permite avaliar o quanto a previsão do modelo utilizado está sendo acurada.

Primeiramente, apresentam-se os conceitos necessários para a construção dessa curva para o caso da regressão logística binária. Dada um ponto de corte c , pode-se observar uma tabela de classificação como a apresentada na Tabela 2.1[9].

Tabela 2.1 – Classificação real, observada, versus a classificação sensorial prevista através dos modelos de regressão logística

	Classificação prevista		
	c	$y = 1$	$y = 0$
Classificação real	$y = 1$	n_{PR}	n_{PF}
	$y = 0$	n_{NF}	n_{NR}
		n_1	n_0

Defini-se como “categoria principal” a categoria de interesse, ou seja quando $y = 1$, e como “categoria secundária” quando $y = 0$. Na Tabela 2.1 mostra-se que os n_{PR} são os casos positivos reais, aqueles casos que pertencem a “categoria principal” e foram classificados sendo pertencentes a essa. Os n_{PF} são os casos positivos falsos que são os casos que pertencem a “categoria principal”, porém foram classificados como pertencentes a “categoria secundária”. Já os n_{NF} são os casos negativos falsos, referentes aos casos que pertencem a “categoria secundária”, mas foram classificados como pertencentes a “categoria principal”. Os n_{NR} são os casos negativos reais que são os casos que pertencem a “categoria secundária” e foram classificados como pertencentes a essa. Por fim, n_1 é a quantidade total de casos classificados como pertencentes a “categoria principal” e n_0 é a quantidade total de casos classificados como pertencentes a “categoria secundária”.

Da Tabela 2.1 pode-se definir duas medidas importantes para a avaliação da eficiência preditiva de um modelo:

- *Sensibilidade (Se)*: a proporção de casos positivos reais, sobre o número total de casos classificados como pertencentes a “categoria principal”, que pode ser obtida pela razão n_{PR}/n_1 .
- *Especificidade (Es)*: a proporção de casos negativos reais, sobre o número total de casos classificados como pertencentes a “categoria secundária”, que pode ser obtida pela razão n_{NR}/n_0 .

Outra medida muito utilizada para a construção da curva ROC é $[1 - (Es)]$ que é a proporção de casos positivos falsos, sobre o número total de casos classificados como pertencentes a “categoria secundária”, que pode ser obtida pela razão n_{PF}/n_0 .

Em uma situação ideal, seria esperado que (Se) e (Es) fossem iguais a 1 e consequentemente $[1 - (Es)]$ fosse igual a 0 para todo c , de modo que todos os casos teriam sido classificados corretamente. Essa situação pode ser visualizada pela curva vermelha na Figura 2.8. Essas três medidas (Se) , (Es) e $[1 - (Es)]$ variam bastante de acordo com o ponto de corte c utilizado, o que pode ser demonstrado pela curva verde na Figura 2.8 que mostra um exemplo comum de uma curva ROC.

De acordo com Kleinbaum e Klein [9], ROC é uma representação de *sensibilidade* (Se) contra $1 - \textit{especificidade}$ $[1 - (Es)]$ obtido através de diversos pontos de corte para os valores preditos.

A área sob a curva ROC denominada AUC (Area Under the Curve) mede a discriminação, que é a habilidade do modelo de classificar corretamente aqueles pertencentes a “categoria principal” e a “categoria secundária”. Para que se considere que um modelo possui uma boa discriminação é esperado que o (Se) seja maior do que $[1 - (Es)]$ para todos os pontos de corte. Em outras palavras, que a proporção de casos classificados corretamente na “categoria principal” (em relação ao total de casos classificados como pertencentes a “categoria principal”) ultrapasse a proporção de casos classificados erroneamente na “categoria secundária” (em relação ao total de casos classificados como pertencentes a “categoria secundária”)[9].

Consequentemente, pode-se classificar um modelo quanto a sua discriminação da seguinte forma:

- discriminação excelente: AUC de 0,9 até 1.
- discriminação boa: AUC de 0,8 até 0,9.
- discriminação razoável: AUC de 0,7 até 0,8.
- discriminação ruim: AUC de 0,6 até 0,7.
- discriminação péssima: AUC de 0,5 até 0,6.
- discriminação negativa: AUC de 0,0 até 0,5.

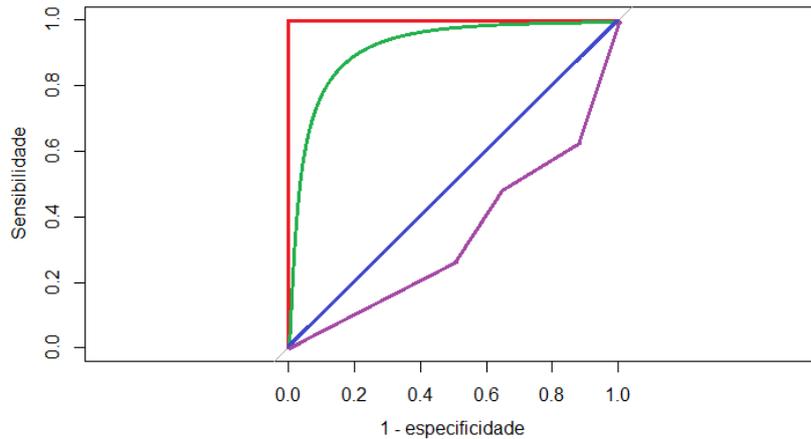


Figura 2.8 – Exemplos de Curva ROC

A reta azul na Figura 2.8 mostra um caso onde AUC é de 0,5, ou seja, tem discriminação péssima. Para esse caso, o modelo utilizado não está fazendo diferença na discriminação, pois classificar uma observação utilizando tal modelo é equivalente a jogar uma moeda e classificar a “categoria principal” quando se obtiver cara e classificar como “categoria secundária” quando se obtiver coroa, ou vice-versa [9, p. 357]. Já a discriminação negativa pode ser visualizada pela linha roxa da Figura 2.8, quando o modelo está prejudicando uma discriminação correta, ou seja, está prevendo mais observações erroneamente do que o número de observações que estão sendo corretamente previstas.

Para o caso politômico, obter-se-á uma curva ROC para cada categoria. A curva ROC que possuir o maior AUC representa a categoria que o modelo está prevendo com maior eficiência quando comparado as outras categorias.

3 Resultados

3.1 Análise descritiva dos dados para o ano de 2013

Na Tabela 3.1 mostra-se que os valores de classificação da variável dureza estão muito concentrados em uma única categoria, “Macio”, que está representando 73,54% dos dados. A categoria “Extremamente macio” é a que possui a segunda maior frequência, porém só representa 15,87% dos dados. As outras categorias de dureza apresentam frequências bem inexpressivas, sendo que nenhum valor foi classificado como “Muito firme” ou “Extremamente Firme”.

Tabela 3.1 – Classificação sensorial de dureza de arroz cozido

Dureza	Frequência	Porcentagem
Extremamente firme	0	0%
Muito firme	0	0%
Levemente firme	1	0,53%
Macio com centro firme	14	7,41%
Ligeiramente macio	5	2,65%
Macio	139	73,54%
Extremamente macio	30	15,87%

De acordo com a Tabela 3.2, 3a classificação dos valores para a variável pegajosidade está bem mais distribuída entre as categorias do que para a variável dureza. A categoria “Levemente solto” é a que representa a maior parte dos dados com 37,57%, seguida da categoria “Pegajoso” com 33,33%, em terceiro lugar vem a categoria “Muito pegajoso” com 22,75%, e por sua vez, a categoria “Solto” representa apenas 6,35% dos dados. Apesar da variável pegajosidade apresentar uma melhor distribuição na classificação dos dados, ainda há três categorias nas quais nenhum valor foi classificado, são essas “Extremamente solto”, “Muito solto”, e “Extremamente pegajoso”.

Tabela 3.2 – Classificação sensorial de pegajosidade de arroz cozido

Pegajosidade	Frequência	Pegajosidade
Extremamente pegajoso	0	0%
Muito pegajoso	43	22,75%
Pegajoso	63	33,33%
Levemente solto	71	37,57%
Solto	12	6,35%
Muito solto	0	0%
Extremamente solto	0	0%

Observando a classificação especificamente de acordo com o tipo de terreno, pode-se observar na Figura (3.1) que a concentração na classificação para a variável dureza é menor para as terras altas do que para os terrenos irrigados.

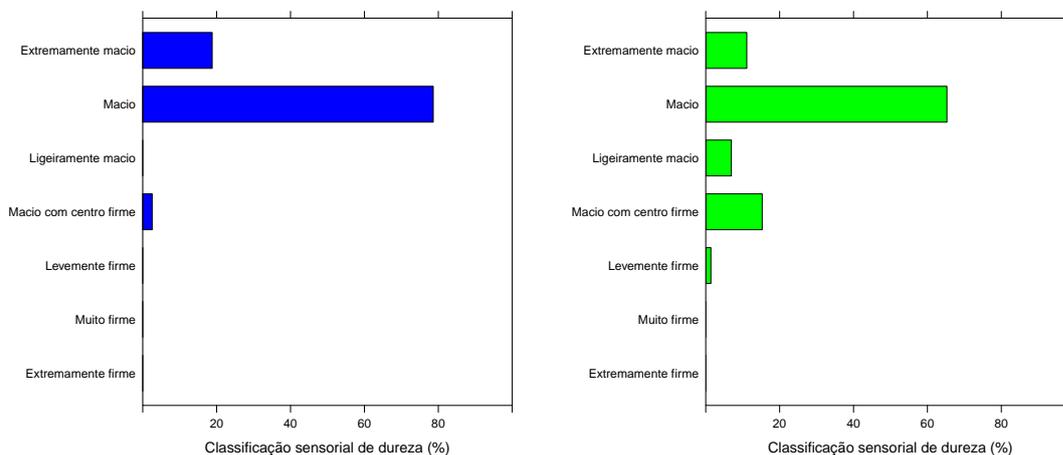


Figura 3.1 – Gráfico de barras da classificação de dureza para o arroz de Terrenos Irrigados (esquerda) e para o arroz de Terras Altas (direita)

Na Figura (3.2) mostra-se que a classificação dos valores para a variável pegajosidade está mais bem distribuída entre as categorias para os terrenos irrigados do que para as terras altas. No caso de terrenos irrigados, a categoria que apresenta maior frequência é “Pegajoso”, enquanto que para terras altas é “Levemente solto”.

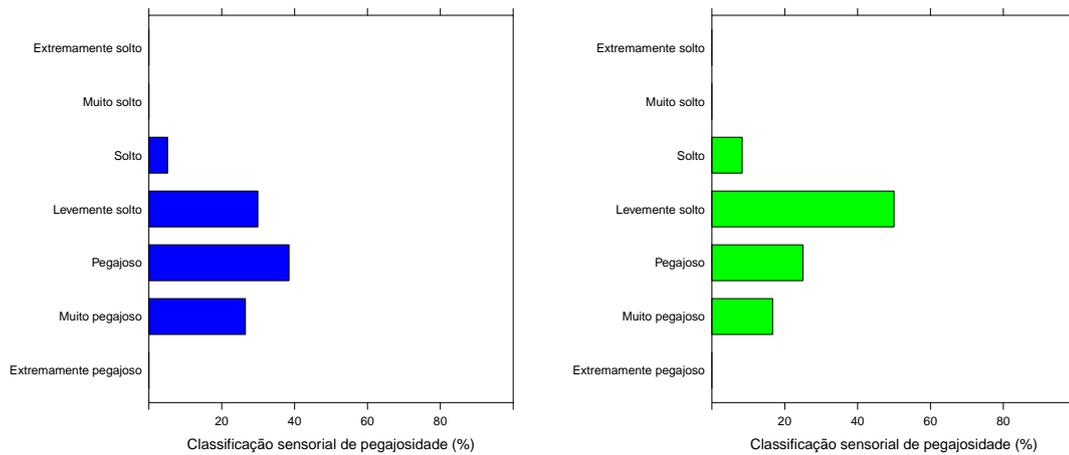


Figura 3.2 – Gráfico de barras da classificação de pegajosidade para o arroz de Terrenos Irrigados (esquerda) e para o arroz de Terras Altas (direita)

De acordo com a Tabela 3.3 mostra-se que as variáveis quantitativas de interesse possuem uma alta variação, principalmente para as variáveis PEAK (5538,056), BREAKDOWN (2240,554) e FINAL (13486,89).

Tabela 3.3 – Medidas resumo das variáveis quantitativas da qualidade do arroz

	Média	Variância	Mínimo	1° Quartil	Mediana	3° Quartil	Máximo
TAAFIA	15,15	52,11761	0	11,44	19	20,3	23,8
TAASEC	16,03	67,18892	0,06	10,28	19,96	22,79	25,25
TG	4,68	2,564881	2	3	4,4	6,4	7
PEAK	230,19	5538,056	53,58	179,17	247,58	289,17	349,58
BREAKDOWN	90,38	2240,554	10,5	53,67	80,25	127,83	205,5
FINAL	280,57	13486,89	21,83	209,58	264	385,92	571,33
SETBACK	140,76	4961,736	6,17	99,83	131,75	197,58	339,17
DUREZAT	140,85	453,8328	72,27	125,84	143,07	156,77	188,12
PEGAJT	-9,128	30,59415	-24,72	-13,01	-7,35	-4,46	-2,22

A Tabela 3.4 apresenta o teste de hipótese para o coeficiente de correlação Spearman, tendo como hipótese nula que $\rho = 0$ e hipótese alternativa que $\rho \neq 0$. Considerando um $\alpha = 0,01$ observou-se que a variável PEAK não possui correlação linear com a TAAFIA ($p\text{-valor} = 0,0124$), a TAASEC ($p\text{-valor} = 0,0488$), a DUREZAT ($p\text{-valor} = 0,2189$), nem com a PEGAJT ($p\text{-valor} = 0,0956$). O mesmo ocorre com a variável BREAKDOWN e a variável e as variáveis FINAL ($p\text{-valor} = 0,2321$), SETBACK ($p\text{-valor} = 0,6784$) e a DUREZAT ($p\text{-valor} = 0,0331$).

Tabela 3.4 – Matriz de correlação entre as variáveis quantitativas seguido do seu p-valor para a hipótese nula $\rho = 0$

	TAAFIA	TAASEC	TG	PEAK	BREAKDOWN	FINAL	SETBACK	DUREZAT	PEGAJT
TAAFIA	1	0,9725	0,389	0,1815	-0,2562	0,6889	0,7718	0,5243	0,9011
TAASEC	0,9725	1	<0,0001	0,0124	0,0003	<0,0001	<0,0001	<0,0001	<0,0001
TG	<0,0001	<0,0001	1	0,4752	-0,3119	0,6879	0,7728	0,5285	0,9095
PEAK	0,389	0,4752	<0,0001	1	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001
BREAKDOWN	0,1815	0,1436	-0,2979	<0,0001	1	0,0023	0,0007	0,0063	<0,0001
FINAL	0,0124	0,0488	<0,0001	0,7436	<0,0001	1	0,5902	0,0898	0,1216
SETBACK	-0,2562	-0,3119	-0,643	0,7436	1	0,0873	0,0304	-0,0712	-0,3676
DUREZAT	0,0003	<0,0001	<0,0001	<0,0001	<0,0001	0,2321	0,6784	0,03306	<0,0001
PEGAJT	0,6889	0,6879	0,2199	0,6957	0,0873	1	0,9724	0,3116	0,6867
	<0,0001	<0,0001	0,0023	<0,0001	0,2321	<0,0001	<0,0001	<0,0001	<0,0001
	0,7718	0,7728	0,2452	0,5902	0,0304	0,9724	1	0,3711	0,7567
	<0,0001	<0,0001	0,0007	<0,0001	0,6784	<0,0001	<0,0001	<0,0001	<0,0001
	0,5243	0,5285	0,1979	0,0898	-0,0712	0,3116	0,3711	1	0,3766
	<0,0001	<0,0001	0,0063	0,2189	0,0331	<0,0001	<0,0001	<0,0001	<0,0001
	0,9011	0,9095	0,4363	0,1216	-0,3676	0,6867	0,7567	0,3766	1
	<0,0001	<0,0001	<0,0001	0,0956	<0,0001	<0,0001	<0,0001	<0,0001	<0,0001

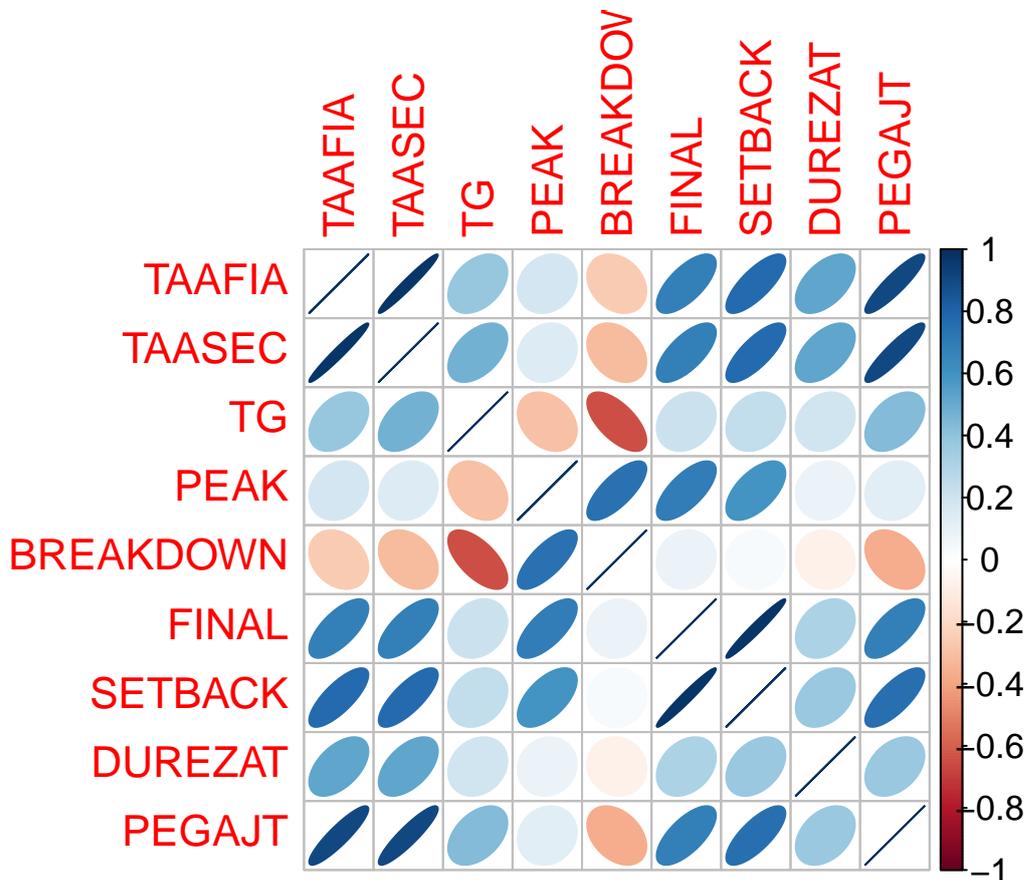


Figura 3.3 – Gráfico de calor para a matriz de correlação entre as variáveis

Na Figura (3.3) e na Tabela 3.4 mostra-se que as variáveis TAAFIA e TAASEC são altamente correlacionadas positivamente ($r = 0,9725$). A mesma coisa ocorre para as variáveis FINAL e SETBACK ($r = 0,9724$). A variável PEGAJT apresenta alta correlação com a variável TAAFIA ($r = 0,9011$) e com a variável TAASEC ($r = 0,9095$), seguidas da correlação com a variável SETBACK ($r = 0,7567$). A variável SETBACK

também é positivamente correlacionada com a TAAFIA ($r = 0,7718$) e a variável TAA-SEC ($r = 0,7728$). Já as variáveis TG e BREAKDOWN apresentam correlação negativa ($r = -0,643$).

De acordo com a Figura (3.4), a variável TAAFIA possui valores mais altos e mais concentrados para terrenos irrigados do que para terras altas. A variável TAASEC apresenta comportamento similar porém com uma maior dispersão dos valores para ambos os tipos de terreno. A variável TG apresenta valores dispersos e bem maiores do que aqueles visualizados em terras altas. As variáveis PEAK, BREAKDOWN, FINAL e SETBACK possui valores concentrados em torno de um mesmo valor para os dois tipos de terreno. Porém os terrenos irrigados nas variáveis PEAK e FINAL apresentam uma maior dispersão dos dados em terras altas, enquanto que as variáveis FINAL e SETBACK apresentam pouca variação de valores para terras altas. A DUREZAT e a PEGAJT apresentaram valores mais altos para os terrenos irrigados e dispersão parecida dos valores para ambos os tipos de terreno.

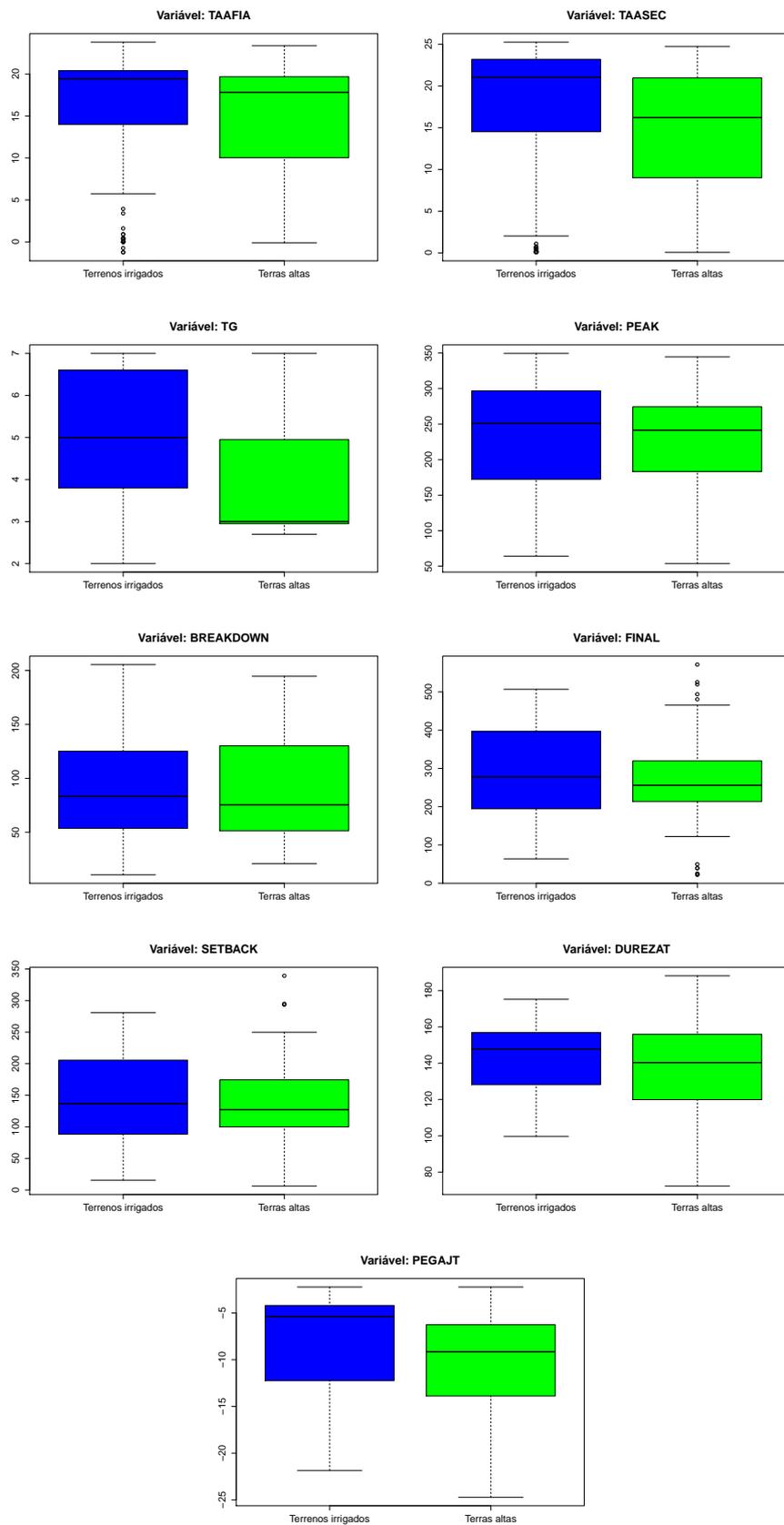


Figura 3.4 – Boxplot das variáveis quantitativas de acordo com o tipo de plantio de arroz (terrenos irrigados ou terras altas)

3.2 Análise de Componentes Principais para o ano de 2013

3.2.1 Terrenos Irrigados

Verificando a possibilidade de redução de dimensão do estudo, realizou-se uma análise de componentes principais com todas as medidas de viscosidade: TAAFIA, TAASEC, TG, PEAK, BREAKDOWN e FINAL. Essa análise de componentes principais juntamente com a análise de correlação apresentada no capítulo anterior mostraram que as variáveis TAASEC e SETBACK podem ser retiradas do estudo devido a alta correlação com TAAFIA e FINAL, respectivamente. Os resultados da nova análise de componentes principais sem a TAASEC e a SETBACK se encontram na Figura (3.5) e nas Tabelas 3.5 e 3.6 .

Pela Tabela 3.5 e pela Figura (3.5) é possível observar que a primeira componente só explica 50,03% da variâncias dos dados, por isso se faz necessário mais uma componente para resumir os dados. Como 86,21% da variância total dos dados pode ser explicada pelas duas primeiras componentes parece apropriado reter e utilizar somente essas duas componentes na análise dos dados.

Tabela 3.5 – Variância das componentes principais para arroz de Terrenos Irrigados e porcentagem da contribuição de cada uma dessas variâncias para a variância total.

Componentes Principais	Desvio Padrão	Proporção da variância	Proporção acumulada da variância
Primeira	1,5817	0,5003	0,5003
Segunda	1,3449	0,3617	0,8621
Terceira	0,67040	0,08989	0,95196
Quarta	0,47609	0,04533	0,99729
Quinta	0,11630	0,00271	1

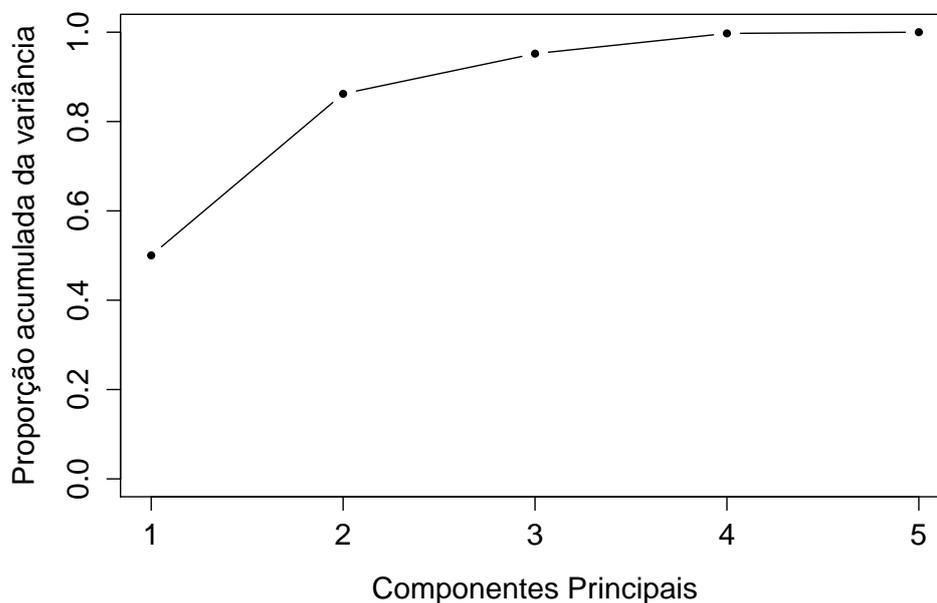


Figura 3.5 – Proporção acumulada da variância explicada por cada componente principal para arroz de Terrenos Irrigados.

Tabela 3.6 – Contribuição de cada variável nas duas primeiras componentes principais para arroz de Terrenos Irrigados e coeficiente de correlação entre as variáveis dos dados e as componentes principais selecionadas.

	Componentes principais	
	Coeficiente de correlação	
	Primeira componente	Segunda componente
TAAFIA	0,00002	-0,67056
	0,00003	-0,902
TG	-0,45450	-0,34056
	-0,719	-0,458
PEAK	0,59955	-0,15141
	0,948	-0,204
BREAKDOWN	0,55585	0,25831
	0,879	0,347
FINAL	0,35356	-0,58712
	0,559	-0,790

A Tabela 3.6 mostra que as variáveis PEAK, BREAKDOWN e FINAL estão contribuindo de maneira similar a primeira componente, opondo-se a contribuição da variável TG nessa componente. A variável TAAFIA é a que menos contribui na primeira compo-

nente, mas é também a que apresenta a maior contribuição na segunda componente. A segunda variável que mais contribui na segunda componente é a FINAL, seguida da TG. A variável BREAKDOWN contribui de maneira oposta a das outras variáveis na segunda componente.

3.2.2 Terras Altas

Verificando a possibilidade de redução de dimensão do estudo, realizou-se uma análise de componentes principais com todas as medidas de viscosidade: TAAFIA, TAASEC, TG, PEAK, BREAKDOWN e FINAL. Essa análise de componentes principais juntamente com a análise de correlação apresentada no capítulo anterior mostraram que as variáveis TAASEC e SETBACK podem ser retiradas do estudo devido a alta correlação com TAAFIA e FINAL, respectivamente. Os resultados da nova análise de componentes principais sem a TAAFIA e FINAL se encontram na Figura (??) e nas Tabelas 3.7 e 3.8.

Pela Tabela (3.7) e pela Figura ?? é possível observar que a primeira componente só explica 49,98% da variâncias dos dados, por isso se faz necessário mais uma componente para resumir os dados. Como 90,22% da variância total dos dados pode ser explicada pelas duas primeiras componentes parece apropriado reter e utilizar somente essas duas componentes na análise dos dados.

Tabela 3.7 – Variância das componentes principais para arroz de Terras Altas e porcentagem da contribuição de cada uma dessas variâncias para a variância total

Componentes Principais	Desvio Padrão	Proporção da variância	Proporção acumulada da variância
Primeira	1,5808	0,4998	0,4998
Segunda	1,4184	0,4023	0,9022
Terceira	0,59350	0,07045	0,97261
Quarta	0,35011	0,02451	0,99712
Quinta	0,11991	0,00288	1

Tabela 3.8 – Contribuição de cada variável nas duas primeiras componentes principais para arroz de Terras Altas e coeficiente de correlação entre as variáveis dos dados e as componentes principais selecionadas

	Componentes principais	
	Coeficiente de correlação	
	Primeira componente	Segunda componente
TAAFIA	-0,56335	-0,01402
	-0,8906	-0,0199
TG	-0,47905	0,36254
	-0,757	0,514
PEAK	-0,26428	-0,62648
	-0,418	-0,889
BREAKDOWN	0,18215	-0,65922
	0,288	-0,935
FINAL	-0,59171	-0,20328
	-0,935	-0,288

Na Tabela 3.8 mostra-se que as variáveis TAAFIA, TG e FINAL estão contribuindo de maneira similar a primeira componente, se opondo a contribuição da variável BREAKDOWN nessa componente. A variável BREAKDOWN é a que menos contribui na primeira componente, mas é também a que apresenta a maior contribuição na segunda componente. A segunda variável que mais contribui na segunda componente é a PEAK, seguida da TG. A variável TG contribui de maneira oposta a das outras variáveis na segunda componente.

3.3 Resultados da Regressão Logística

Neste capítulo são apresentados em detalhes os melhores modelos obtidos para os anos de 2013 e 2014 e um resumo dos demais modelos analisados. Tabelas e Figuras mais detalhadas sobre os demais modelos podem ser obtidos no Apêndice.

3.3.1 Para o ano de 2013

Predição da avaliação sensorial da dureza por medidas instrumentais de textura para arroz de Terrenos Irrigados

Devido ao resultado observado na Figura 3.1, optou-se por recodificar as categorias da avaliação sensorial de dureza em apenas 3, Extremamente Macio (EM), Macio (M) e Ligeiramente Macio ou inferior (LML = LM + MCF + LF).

O modelo ajustado que propõe a avaliação sensorial de dureza por meio da dureza instrumental $DUREZAT_i$ de arroz de Terrenos Irrigados para o ano de 2013, segue o modelo logito cumulativo:

$$P(Y_i \leq k | DUREZAT_i) = \frac{\exp(\hat{\beta}_{0k} - \hat{\beta} DUREZAT_i)}{1 + \exp(\hat{\beta}_{0k} - \hat{\beta} DUREZAT_i)}, \quad k = 1, 2, 3 \quad (3.1)$$

aonde $k=1$ indica a categoria Ligeiramente macio ou inferior (LML = LM + MCF + LF), $k=2$ indica a categoria Macio (M) e $k=3$, a categoria Extremamente Macio (EM). O i -ésimo valor da variável dureza instrumental é dado por $DUREZAT_i$ e $\hat{\beta}$ representa o efeito dessa variável na classificação da avaliação sensorial.

Para um nível de significância de 5%, a medida de dureza instrumental apresentou efeito significativo nesse modelo (p -valor $< 0,00001$), logo houve ajustamento dos dados ao modelo proposto. Os modelos logito cumulativo com as estimativas dos parâmetros são:

$$P(Y_i \leq 1 | DUREZAT_i) = \frac{\exp[-12,56599 - (-0,05920)DUREZAT_i]}{1 + \exp[-12,56599 - (-0,05920)DUREZAT_i]}, \quad (3.2)$$

$$P(Y_i \leq 2 | DUREZAT_i) = \frac{\exp[-6,67342 - (-0,05920)DUREZAT_i]}{1 + \exp[-6,67342 - (-0,05920)DUREZAT_i]}. \quad (3.3)$$

A probabilidade de pertencer a uma determinada categoria da avaliação sensorial de dureza, para cada valor da dureza instrumental é dado por:

$$P(Y_i = 1 | DUREZAT_i) = P(Y_i \leq 1 | DUREZAT_i), \quad (3.4)$$

$$P(Y_i = 2 | DUREZAT_i) = P(Y_i \leq 2 | DUREZAT_i) - P(Y_i = 1 | DUREZAT_i), \quad (3.5)$$

$$P(Y_i = 3 | DUREZAT_i) = 1 - P(Y_i \leq 2 | DUREZAT_i). \quad (3.6)$$

Utilizando o método de validação cruzada, obtém-se a matriz de classificação sensorial observada versus a classificação sensorial prevista por meio dos modelos de regressão logística (3.4), (3.5) e (3.6) (Tabela 3.9). Através da Tabela 3.9, calcula-se a taxa do erro de classificação, que é de 20,51%. Nota-se que nenhuma observação foi prevista como Ligeiramente macio ou inferior (LML).

Tabela 3.9 – Classificação da avaliação sensorial de dureza para o ano de 2013 versus a classificação prevista, por meio do modelo de regressão logística utilizando a dureza instrumental de arroz de Terrenos Irrigados

		Classificação prevista		
		LML	M	EM
Classificação real	LML	0	2	1
	M	0	89	3
	EM	0	18	4

Outra alternativa para avaliar a qualidade de classificação do modelo proposto é a curva ROC. De acordo com essa curva apresentada na Figura 3.6, mostra-se que as categorias Macio (M) e Extremamente Macio (EM) estão sendo discriminadas com maior precisão pelo modelo utilizado do que a categoria Ligeiramente macio ou inferior (LML). Isso pode ser melhor visualizado pela respectiva AUC (área sob a da curva) de cada categoria, que foi de 0,82275 na categoria Extremamente Macio (EM); 0,81335 na categoria Macio (M) e 0,64048 na categoria Ligeiramente macio ou inferior (LML).

De acordo com a Figura (3.7), a medida que a dureza instrumental aumenta, a probabilidade do arroz receber avaliação sensorial como Extremamente Macio (EM) diminui gradativamente, enquanto que a probabilidade de receber avaliação sensorial como Ligeiramente Macio ou inferior (LML) aumenta. A região de incerteza do gráfico mostra que quando o valor da dureza instrumental está em torno de 113 não é possível concluir com segurança que a avaliação sensorial seria preferencialmente do tipo Extremamente Macio (EM) ou Macio (M). Assim para valores de dureza instrumental ao redor do intervalo 90 a 106 o arroz seria classificado sensorialmente como Extremamente Macio (EM), já para valores pertencentes ao intervalo de 120 a 180 o arroz seria classificado como Macio (M).

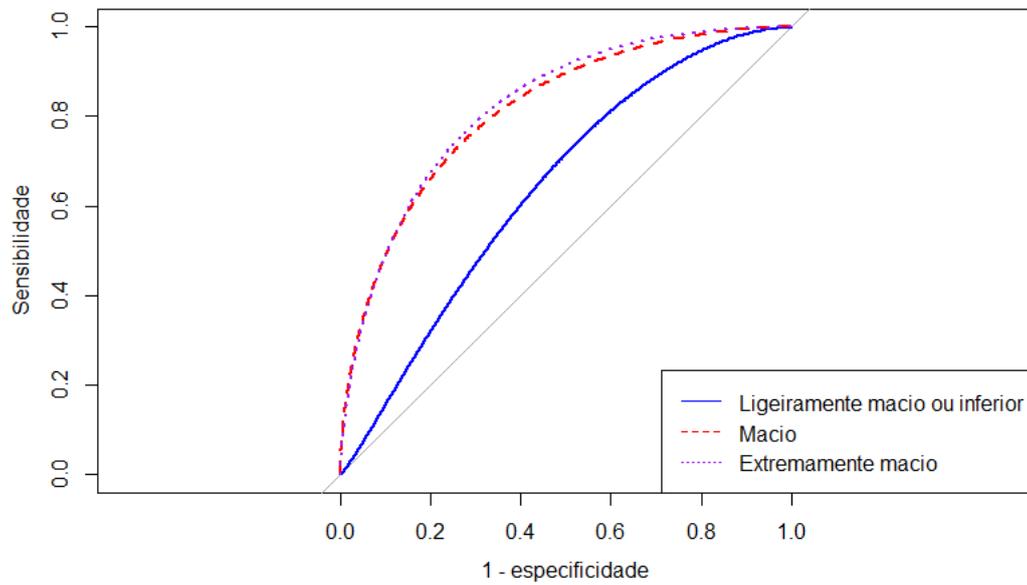


Figura 3.6 – Curva de classificação ROC da avaliação sensorial de dureza para o ano de 2013, prevista por meio do modelo de regressão logística utilizando a dureza instrumental de arroz de Terrenos Irrigados

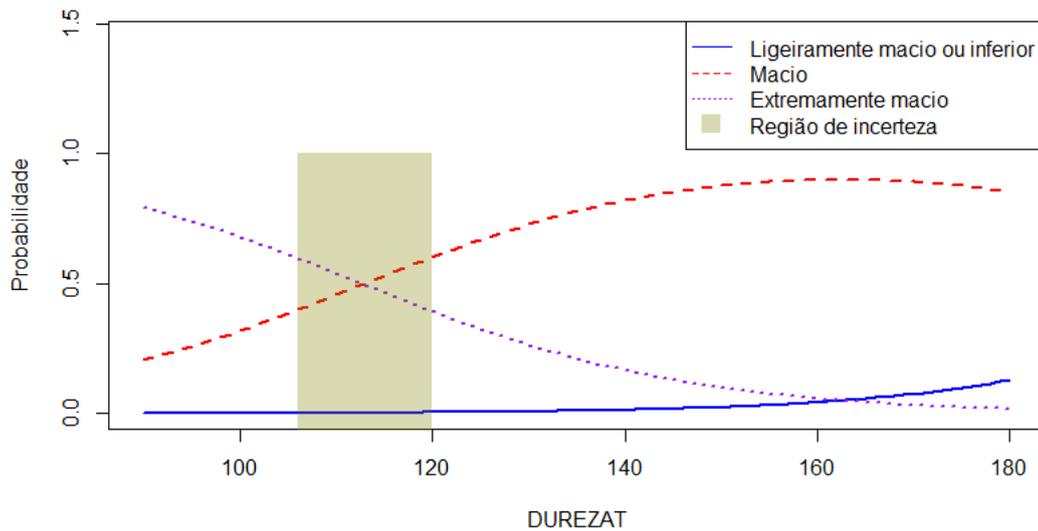


Figura 3.7 – Probabilidades das categorias de avaliação sensorial de dureza considerando diferentes valores da dureza instrumental de arroz de Terrenos Irrigados para o ano de 2013

Predição da avaliação sensorial da dureza por medidas de viscosidade para arroz de Terrenos Irrigados

Na Tabela 3.6 apresentam-se os pesos de cada medida de viscosidade na construção das duas primeiras componentes principais para terrenos irrigados. Multiplicando-se estes pesos por cada valor das variáveis originais obtêm-se novas variáveis que são utilizadas como variáveis explicativas no modelo de regressão logística politômica. Estas variáveis são denominadas de $C1$ e $C2$ por representarem combinações lineares das medidas de viscosidade e serem formadas pelos coeficientes da primeira e segunda componentes principais.

O modelo ajustado que propõe a avaliação sensorial de dureza por meio das variáveis $C1$ e $C2$ de arroz de Terrenos Irrigados para o ano de 2013, segue o modelo logito cumulativo:

$$P(Y_i \leq k | C1_i, C2_i) = \frac{\exp(\hat{\beta}_{0k} - \hat{\beta}_1 C1_i - \hat{\beta}_2 C2_i)}{1 + \exp(\hat{\beta}_{0k} - \hat{\beta}_1 C1_i - \hat{\beta}_2 C2_i)}, \quad k = 1, 2, 3 \quad (3.7)$$

aonde $k=1$ indica a categoria Ligeiramente macio ou inferior (LML = LM + MCF), $k = 2$ indica a categoria Macio (M) e $k = 3$, a categoria Extremamente Macio (EM). O i -ésimo valor da variável $C1$ é dado por $C1_i$ e $\hat{\beta}_1$ representa o efeito dessa variável na classificação da avaliação sensorial. Já o i -ésimo valor da variável $C2$ é dado por $C2_i$ e $\hat{\beta}_2$ representa o efeito dessa variável na classificação da avaliação sensorial.

Dado um nível de significância de 2,5% para cada variável, as variáveis $C1$ e $C2$ apresentaram efeito significativo nesse modelo, pois para $C1$ o p-valor foi de 0,00253 e para $C2$ o p-valor foi menor do que 0,00001, ou seja ambos os p-valores foram menores do que 0,025 indicando que houve ajustamento dos dados ao modelo proposto.

Os modelos logito cumulativo com as estimativas dos parâmetros são:

$$P(Y_i \leq 1 | C1_i, C2_i) = \frac{\exp(-5,49880 - 0,01067C1_i - 0,02265C2_i)}{1 + \exp(-5,49880 - 0,01067C1_i - 0,02265C2_i)}, \quad (3.8)$$

$$P(Y_i \leq 2 | C1_i, C2_i) = \frac{\exp(0,52778 - 0,01067C1_i - 0,02265C2_i)}{1 + \exp(0,52778 - 0,01067C1_i - 0,02265C2_i)}, \quad (3.9)$$

A probabilidade de pertencer a uma determinada categoria da avaliação sensorial de dureza, para cada valor de $C1$ e $C2$ é dado por:

$$P(Y_i = 1|C1_i, C2_i) = P(Y_i \leq 1|C1_i, C2_i), \quad (3.10)$$

$$P(Y_i = 2|C1_i, C2_i) = P(Y_i \leq 2|C1_i, C2_i) - P(Y_i = 1|C1_i, C2_i), \quad (3.11)$$

$$P(Y_i = 3|C1_i, C2_i) = 1 - P(Y_i \leq 2|C1_i, C2_i). \quad (3.12)$$

Utilizando o método de validação cruzada, obtém-se a matriz de classificação sensorial observada versus a classificação sensorial prevista por meio dos modelos de regressão logística estimados (3.10),(3.11),(3.12) (Tabela 3.10). Com base na Tabela 3.10, calcula-se a taxa do erro de classificação, que é de 21,37%. Nota-se que nenhuma observação foi prevista como Ligeiramente macio ou inferior (LML).

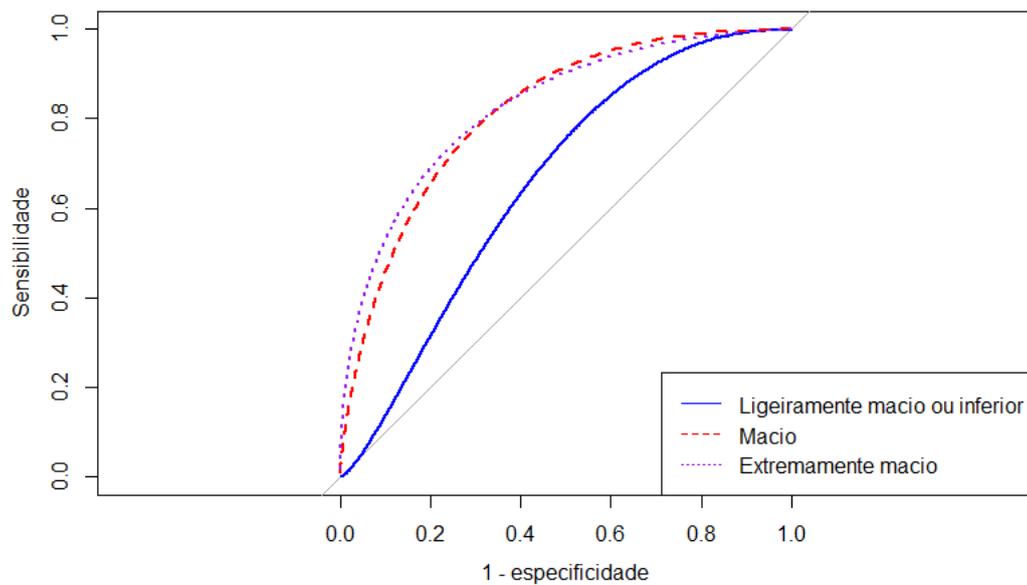


Figura 3.8 – Curva de classificação ROC da avaliação sensorial de dureza para o ano de 2013, prevista por meio do modelo de regressão logística utilizando componentes principais de arroz de Terrenos Irrigados

De acordo com a curva ROC apresentada na Figura 3.8, mostra-se que as categorias Macio (M) e Extremamente Macio (EM) estão sendo discriminadas com maior precisão pelo modelo utilizado do que a categoria Ligeiramente macio ou inferior (LML). Isso pode ser melhor visualizado pela respectiva AUC (área sob a da curva) de cada categoria, que foi de 0,82617 na categoria Extremamente Macio (EM);0,81544 na categoria Macio (M) e 0,65666 na categoria Ligeiramente macio ou inferior (LML).

Tabela 3.10 – Classificação da avaliação sensorial de dureza para o ano de 2013 versus a classificação prevista, por meio do modelo de regressão logística utilizando componentes principais de arroz de Terrenos Irrigados

	Classificação prevista			
	LML	M	EM	
Classificação real	LML	0	3	0
	M	0	88	4
	EM	0	18	4

Por meio dos modelos estimados (3.10), (3.11) e (3.12) também é possível obter a Figura (3.9) que relaciona os diferentes valores das variáveis $C1$ e $C2$ com a chance do arroz ser considerado pertencente a uma determinada categoria de avaliação sensorial de dureza. Na Figura 3.9 mostra-se que para valores elevados das variáveis $C1$ e $C2$ o arroz seria classificado como Extremamente Macio (EM) e para valores extremamente baixos de $C1$ e $C2$ o arroz seria classificado como Ligeiramente Macio ou inferior (LML). Para valores intermediários de $C1$ e $C2$ a avaliação sensorial seria preferencialmente do tipo Macio (M).

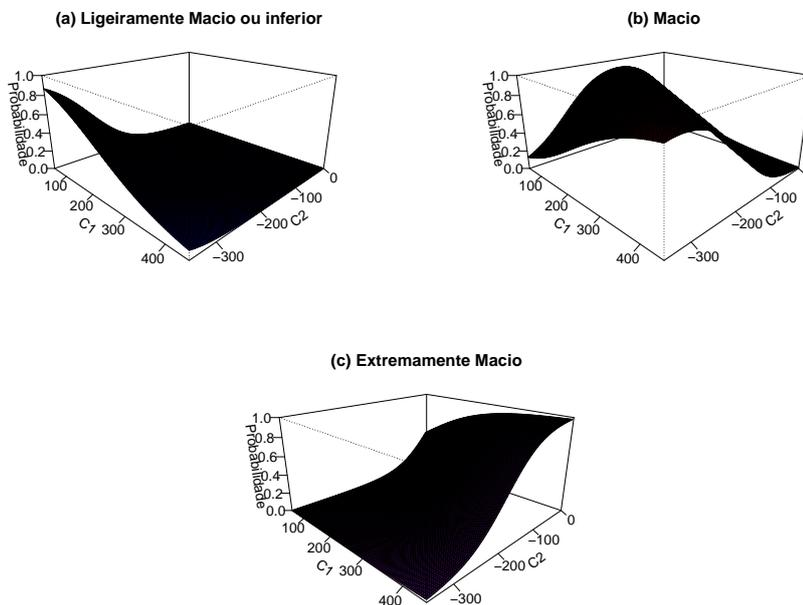


Figura 3.9 – Probabilidades das categorias de avaliação sensorial de dureza considerando diferentes valores das variáveis $C1$ e $C2$ de arroz de Terrenos Irrigados para o ano de 2013

Comparação da predição da avaliação sensorial da dureza por medidas instrumentais de textura e medidas de viscosidade para arroz de Terrenos Irrigados

Ao se utilizar as medidas de viscosidade obtém-se um erro de classificação de 21,37%, que é apenas um pouco maior do que o obtido utilizando as medidas instrumentais de textura (20,51%). Isso indica que é possível obter modelos com capacidade discriminatória praticamente tão boa aplicando-se medidas de viscosidade quanto utilizando medidas instrumentais de textura.

3.3.2 Para o ano de 2014

Predição da avaliação sensorial da pegajosidade por medidas instrumentais de textura para arroz de Terras Altas

Na Tabela 3.11 mostra-se que os valores da variável pegajosidade foram classificados em cinco das sete categorias possíveis. No entanto, devido a baixa frequência observada na categoria Muito Pegajoso (MP), será ajustado um modelo de regressão logística politômica considerando quatro categorias, a nova categoria Muito Pegajoso (MP = EP + MP) que abrange as observações classificadas nas antigas categorias Extremamente Pegajoso (EP) e Muito Pegajoso (MP), a categoria Pegajoso (P), a categoria Levemente Solto (LS) e a categoria Solto (S).

Tabela 3.11 – Classificação sensorial de pegajosidade de arroz de Terras Altas para o ano de 2014

Categoria	Frequência	Pegajosidade
Extremamente pegajoso	11	15,27%
Muito pegajoso	2	2,77%
Pegajoso	21	29,17%
Levemente solto	35	48,61%
Solto	3	4,17%
Muito solto	0	0%
Extremamente solto	0	0%

O modelo ajustado que propõe a avaliação sensorial de pegajosidade por meio da pegajosidade instrumental $PEGAJT_i$ de arroz de Terras Altas para o ano de 2014, segue o modelo logito cumulativo:

$$P(Y_i \leq k | PEGAJT_i) = \frac{\exp(\hat{\beta}_{0k} - \hat{\beta} PEGAJT_i)}{1 + \exp(\hat{\beta}_{0k} - \hat{\beta} PEGAJT_i)}, \quad k = 1, 2, 3, 4 \quad (3.13)$$

aonde $k = 1$ indica a categoria Muito pegajoso (MP = EP + MP), $k = 2$ indica a categoria Pegajoso (P), $k = 3$ indica a categoria , (LS) e $k = 4$, a categoria Solto (S). O i -ésimo valor da variável pegajosidade instrumental é dado por $PEGAJT_i$ e $\hat{\beta}$ representa o efeito dessa variável na classificação da avaliação sensorial.

Para um nível de significância de 5%, a medida de pegajosidade instrumental apresentou efeito significativo nesse modelo (p-valor $< 0,00001$), logo houve ajustamento dos dados ao modelo proposto. Os modelos logitos cumulativos com as estimativas dos parâmetros são:

$$P(Y_i \leq 1|PEGAJT_i) = \frac{\exp(-7,69236 - 0,46297PEGAJT_i)}{1 + \exp(-7,69236 - 0,46297PEGAJT_i)}, \quad (3.14)$$

$$P(Y_i \leq 2|PEGAJT_i) = \frac{\exp(-4,88303 - 0,46297PEGAJT_i)}{1 + \exp(-4,88303 - 0,46297PEGAJT_i)}, \quad (3.15)$$

$$P(Y_i \leq 3|PEGAJT_i) = \frac{\exp(0,14126 - 0,46297PEGAJT_i)}{1 + \exp(0,14126 - 0,46297PEGAJT_i)}. \quad (3.16)$$

A probabilidade de pertencer a uma determinada categoria da avaliação sensorial de pegajosidade, para cada valor da pegajosidade instrumental é dado por:

$$P(Y_i = 1|PEGAJT_i) = P(Y_i \leq 1|PEGAJT_i), \quad (3.17)$$

$$P(Y_i = 2|PEGAJT_i) = P(Y_i \leq 2|PEGAJT_i) - P(Y_i = 1|PEGAJT_i), \quad (3.18)$$

$$P(Y_i = 3|PEGAJT_i) = P(Y_i \leq 3|PEGAJT_i) - P(Y_i \leq 2|PEGAJT_i), \quad (3.19)$$

$$P(Y_i = 4|PEGAJT_i) = 1 - P(Y_i \leq 3|PEGAJT_i). \quad (3.20)$$

Utilizando-se o método de validação cruzada, obtém-se a matriz de classificação sensorial observada versus a classificação sensorial prevista por meio dos modelos de regressão logística estimados (3.17), (3.18), (3.19) e (3.20) (Tabela 3.12). Baseado na Tabela 3.12, calcula-se a taxa do erro de classificação, que é de 23,61%. Nota-se que nenhuma observação foi prevista como Solto (S).

Tabela 3.12 – Classificação da avaliação sensorial de pegajosidade para o ano de 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de Terras Altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	6	5	2	0
	P	3	17	1	0
	LS	1	2	32	0
	S	0	0	3	0

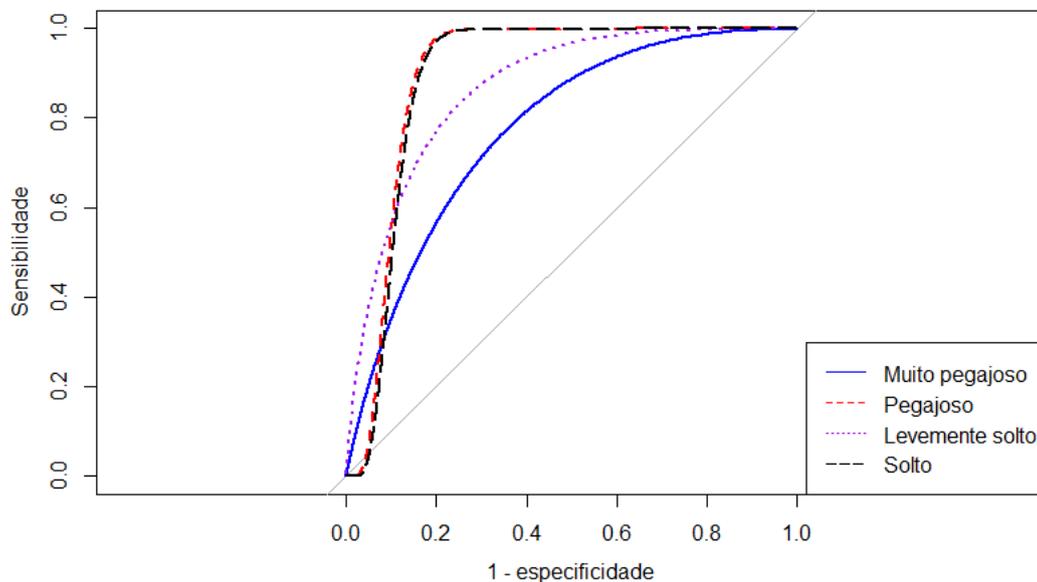


Figura 3.10 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2014, prevista por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de Terras Altas

De acordo com a curva ROC apresentada na Figura 3.10, mostra-se que as categorias Pegajoso (P) e Solto (S) estão sendo discriminadas com maior precisão pelo modelo utilizado do que as categorias , (LS) e Muito Pegajoso (MP). Isso pode ser bem visualizado pela respectiva AUC (área sob a da curva) de cada categoria, que foi de 0,89877 na categoria Pegajoso (P); 0,89134 na categoria Solto (S); 0,86812 na categoria Levemente Solto (LS) e 0,77644 na categoria Muito Pegajoso (MP).

Observa-se pela Figura 3.11 que a medida que a pegajosidade instrumental aumenta a probabilidade do arroz receber avaliação sensorial como Muito Pegajoso (MP)

diminui gradativamente, enquanto que a probabilidade de receber avaliação sensorial como Solto (S) aumenta. A região de incerteza à esquerda do gráfico mostra que quando o valor da pegajosidade instrumental está entre o intervalo -17 a $-15,5$ não é possível concluir com segurança que a avaliação sensorial seria preferencialmente do tipo Muito Pegajoso (MP) ou Pegajoso (P). Já a região de incerteza central do gráfico mostra que quando o valor da pegajosidade instrumental está entre o intervalo $-11,5$ a -10 não é possível concluir com segurança que a avaliação seria preferencialmente do tipo Pegajoso (P) ou Levemente Solto (LS). A região de incerteza à direita do gráfico mostra que quando o valor da pegajosidade instrumental está em torno de 0 não é possível concluir com segurança que a avaliação sensorial seria preferencialmente do tipo Levemente Solto (LS) ou Solto (S). Assim para valores de pegajosidade instrumental pertencentes ao intervalo -25 a -17 o arroz seria classificado sensorialmente como Muito Pegajoso (MP) e para valores pertencentes ao intervalo de $-15,5$ a $-11,5$ o arroz seria classificado como Pegajoso (P). Já para os valores mais elevados da pegajosidade instrumental, entre -10 até -1 o arroz seria classificado como Levemente Solto (LS).

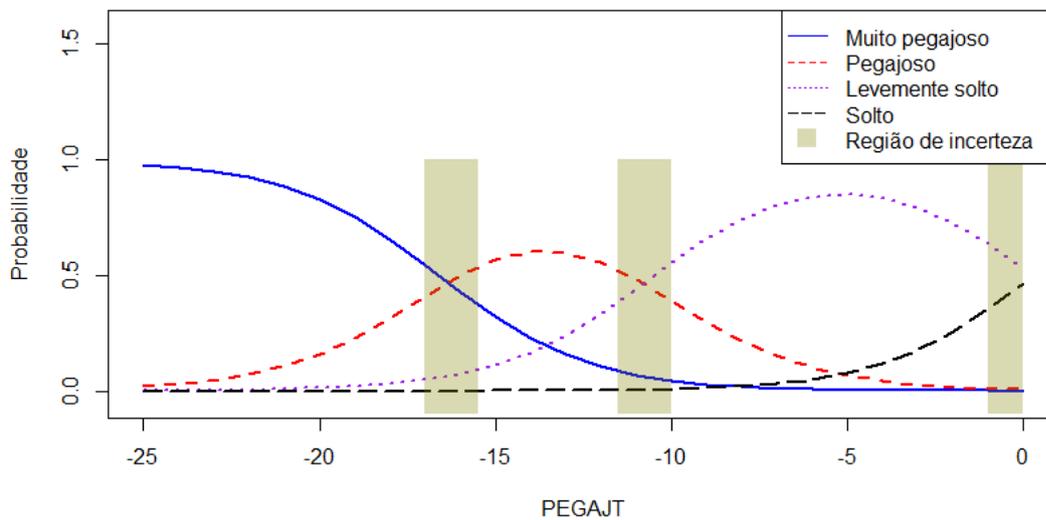


Figura 3.11 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores da pegajosidade instrumental de arroz de Terras Altas para o ano de 2014

Regressão Logística Binária

A fim de diminuir a porcentagem de observações que estão sendo classificadas erroneamente (23,61%) na avaliação de pegajosidade considerando quatro categorias: MP, P, LS e S, decidiu-se realizar uma análise que considera apenas duas categorias (Pegajoso e Solto), esperando que essa forneça classificações mais precisas.

Logo, foi ajustado um novo modelo que considera uma nova categoria Solto (S^*), que abrange as observações classificadas nas antigas categorias , (LS) e Solto (S) e uma nova categoria Pegajoso (P^*), que abrange as observações classificadas nas antigas categorias Muito Pegajoso (MP) e Pegajoso (P).

Dado um nível de significância de 5%, a medida de pegajosidade instrumental apresentou efeito significativo nesse modelo (p-valor $< 0,00001$), logo houve ajustamento dos dados ao modelo logito binário proposto.

Através desse modelo, a chance de avaliação sensorial na categoria Pegajoso (P^*) é dada por:

$$P(Y_i = 0|PEGAJT_i) = \frac{\exp(6,5287 - 0,6515PEGAJT_i)}{1 + \exp(6,5287 - 0,6515PEGAJT_i)}, \quad (3.21)$$

aonde o i -ésimo valor da variável pegajosidade instrumental é dado por $PEGAJT_i$ e $\hat{\beta} = 0,6515$ representa o efeito dessa variável na classificação da avaliação sensorial. Já a probabilidade de pertencer a categoria Solto (S^*) é dado por:

$$P(Y_i = 1|PEGAJT_i) = 1 - P(Y_i = 0|PEGAJT_i). \quad (3.22)$$

Utilizando-se o método de validação cruzada, obtém-se a matriz de classificação sensorial observada versus a classificação sensorial prevista por meio dos modelos de regressão logística estimados (3.21) e (3.22) (Tabela 3.13). Baseado na Tabela 3.13, calcula-se a taxa do erro de classificação, que é de 8,33%. Esse erro é bem menor do que o observado na classificação da pegajosidade politômica indicando que realizar uma análise sensorial considerando uma quantidade menor de categorias pode aumentar a precisão na previsão da classificação por meio de modelos estatísticos, pois a distinção entre categorias é maior quanto menos categorias o modelo possuir.

Tabela 3.13 – Classificação da avaliação sensorial de pegajosidade para o ano de 2014 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de Terras Altas

	Classificação prevista	
	P*	S*
Classificação real	P*	31
	S*	3

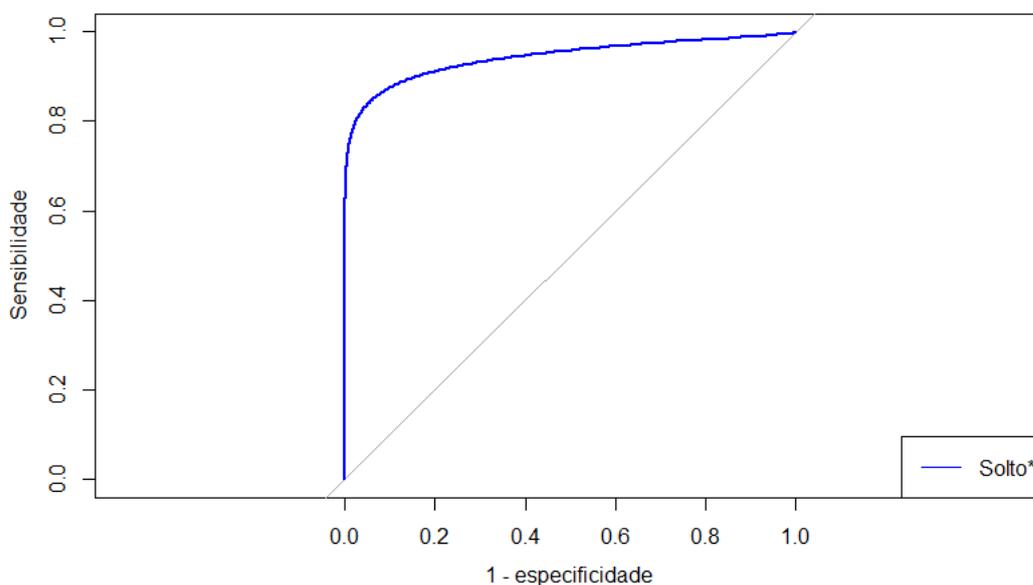


Figura 3.12 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2014, prevista por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de Terras Altas

Observa-se da curva ROC da Figura 3.12 que o modelo binário apresentou discriminação excelente na classificação das categorias Pegajoso (P*) e Solto (S*). Isso também pode ser visualizado por sua AUC (área sob a curva) de 0,943, que está bem próximo de 1, valor que representa a AUC da situação perfeita, quando todos os indivíduos são classificados corretamente por meio do modelo utilizado.

A Figura 3.13 representa a Equação (3.22) relacionando os diferentes valores da pegajosidade instrumental ($PEGAJT_i$'s) com a chance do arroz ser considerado pertencente a categoria Solto (S*). De acordo com a Figura 3.13, à medida que a pegajosidade instrumental aumenta a probabilidade do arroz receber avaliação sensorial como Solto (S*) também aumenta. Assim os valores de pegajosidade instrumental no intervalo de

−25 a −10,5 seriam classificados por meio da avaliação sensorial como Pegajoso (P*) e valores no intervalo −9,5 a 0 como Solto (S*). Essa nova classificação aumentou a certeza na classificação, pois passou-se a ter apenas uma pequena região de incerteza, valores em torno de −10, na qual não é possível concluir com segurança que a avaliação sensorial seria preferencialmente do tipo Pegajoso (P*) ou Solto (S*).

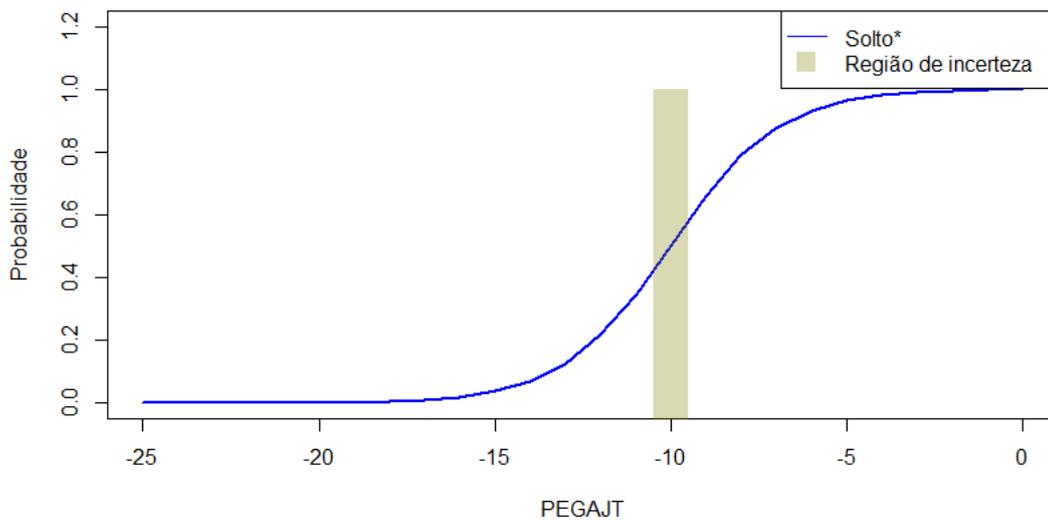


Figura 3.13 – Probabilidade do arroz receber avaliação sensorial como Solto (S*) considerando diferentes valores da pegajosidade instrumental de arroz de Terras Altas para o ano de 2014

Predição da avaliação sensorial da pegajosidade por medidas de viscosidade para arroz de Terras Altas

O modelo ajustado que propõe a avaliação sensorial de pegajosidade por meio das variáveis $C1$ e $C2$ de arroz de Terras Altas para o ano de 2014, segue o modelo logito cumulativo:

$$P(Y_i \leq k | C1_i, C2_i) = \frac{\exp(\hat{\beta}_{0k} - \hat{\beta}_1 C1_i - \hat{\beta}_2 C2_i)}{1 + \exp(\hat{\beta}_{0k} - \hat{\beta}_1 C1_i - \hat{\beta}_2 C2_i)}, \quad k = 1, 2, 3, 4 \quad (3.23)$$

aonde, $k = 1$ indica a categoria Muito pegajoso ($MP = EP + MP$), $k = 2$ indica a categoria Pegajoso (P), $k = 3$ indica a categoria , (LS) e $k = 4$, a categoria Solto (S). O i -ésimo valor da variável $C1$ é dado por $C1_i$ e $\hat{\beta}_1$ representa o efeito dessa variável na

classificação da avaliação sensorial. Já o i -ésimo valor da variável $C2$ é dado por $C2_i$ e $\hat{\beta}_2$ representa o efeito dessa variável na classificação da avaliação sensorial.

Para um nível de significância de 2,5% para cada variável, a variável $C1$ apresentou efeito significativo (p-valor $< 0,00001$) enquanto que a variável $C2$ não apresentou efeito significativo (p-valor = 0,13). Por essa razão, o modelo logito cumulativo passa a ser:

$$P(Y_i \leq k | C1_i, C2_i) = \frac{\exp(\hat{\beta}_{0k} - \hat{\beta}_1 C1_i)}{1 + \exp(\hat{\beta}_{0k} - \hat{\beta}_1 C1_i)}, \quad k = 1, 2, 3, 4 \quad (3.24)$$

aonde $k = 1$ indica a categoria Muito pegajoso (MP = EP + MP), $k = 2$ indica a categoria Pegajoso (P), $k = 3$ indica a categoria , (LS) e $k = 4$, a categoria Solto (S). O i -ésimo valor da variável $C1$ é dado por $C1_i$ e $\hat{\beta}_1$ representa o efeito dessa variável na classificação da avaliação sensorial. Esse novo modelo está se ajustando bem aos dados, pois o p-valor do efeito da variável $C1$ nesse caso é menor que 0,00001 que é menor que o nível de significância de 5%.

Os modelos logitos cumulativos com as estimativas dos parâmetros são:

$$P(Y_i \leq 1 | C1_i, C2_i) = \frac{\exp[2,22402 - (-0,04406C1_i)]}{1 + \exp[2,22402 - (-0,04406C1_i)]}, \quad (3.25)$$

$$P(Y_i \leq 2 | C1_i, C2_i) = \frac{\exp[4,86121 - (-0,04406C1_i)]}{1 + \exp[4,86121 - (-0,04406C1_i)]}, \quad (3.26)$$

$$P(Y_i \leq 3 | C1_i, C2_i) = \frac{\exp([11,50592 - (-0,04406C1_i)])}{1 + \exp[11,50592 - (-0,04406C1_i)]}. \quad (3.27)$$

A probabilidade de pertencer a uma determinada categoria da avaliação sensorial de pegajosidade, para cada valor de $C1$ e $C2$ é dado por:

$$P(Y_i = 1 | C1_i, C2_i) = P(Y_i \leq 1 | C1_i, C2_i), \quad (3.28)$$

$$P(Y_i = 2 | C1_i, C2_i) = P(Y_i \leq 2 | C1_i, C2_i) - P(Y_i = 1 | C1_i, C2_i), \quad (3.29)$$

$$P(Y_i = 3 | C1_i, C2_i) = P(Y_i \leq 3 | C1_i, C2_i) - P(Y_i \leq 2 | C1_i, C2_i), \quad (3.30)$$

$$P(Y_i = 4 | C1_i, C2_i) = 1 - P(Y_i \leq 3 | C1_i, C2_i). \quad (3.31)$$

Utilizando-se o método de validação cruzada, obtém-se a matriz de classificação sensorial observada versus a classificação sensorial prevista por meio dos modelos de regressão logística estimados (3.28), (3.29), (3.30) e (3.31) (Tabela 3.14). Pela Tabela 3.14,

calcula-se a taxa do erro de classificação, que é de 23,61%. Nota-se que nenhuma observação foi prevista como Solto (S).

Tabela 3.14 – Classificação da avaliação sensorial de pegajosidade para o ano de 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando componentes principais de arroz de Terras Altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	6	7	0	0
	P	0	18	3	0
	LS	0	4	31	0
	S	0	0	3	0

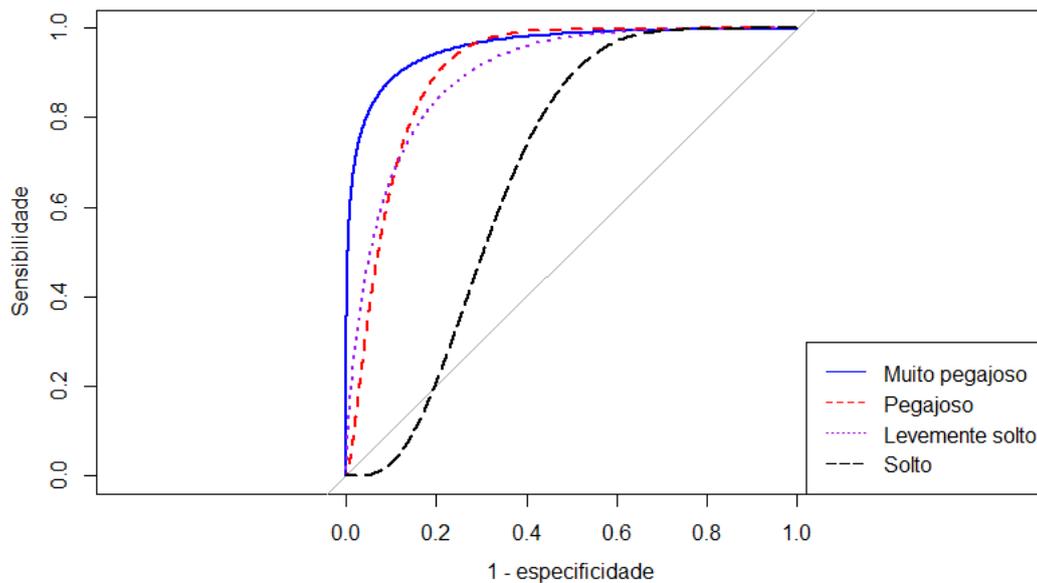


Figura 3.14 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2014, prevista por meio do modelo de regressão logística utilizando componentes principais de arroz de Terras Altas

De acordo com a curva ROC apresentada na Figura 3.14, mostra-se que a categoria Muito Pegajoso (MP) está sendo discriminada com maior precisão pelo modelo utilizado do que as categorias Pegajoso (P) e Levemente Solto (LS). Já a categoria Solto (S) apresenta uma discriminação pobre pois apenas três observações pertencem verdadeiramente a essa categoria, o que afeta em muito a precisão da classificação nessa categoria. Isso pode ser melhor visualizado pela respectiva AUC (área sob a da curva) de cada categoria,

que foi de 0,95985 na categoria Muito Pegajoso (MP); 0,90570 na categoria Pegajoso (P); 0,89842 na categoria Levemente Solto (LS) e 0,68244 na categoria Solto (S).

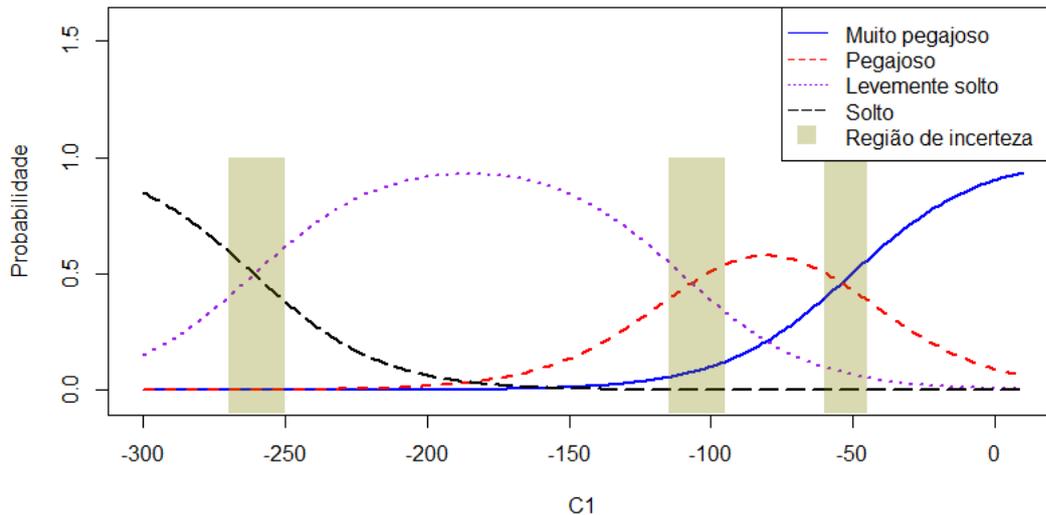


Figura 3.15 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores da variável $C1$ de arroz de Terras Altas para o ano de 2014.

Por meio dos modelos estimados (3.28), (3.29), (3.30) e (3.31) é possível obter a Figura 3.15 que relaciona os diferentes valores da variável $C1$ com a chance do arroz ser considerado pertencente a uma determinada categoria de avaliação sensorial de pegajosidade. Na Figura 3.15 mostra-se que à medida que a variável $C1$ aumenta, a probabilidade do arroz receber avaliação sensorial como Muito Pegajoso (MP) aumenta gradativamente, enquanto que a probabilidade de receber avaliação sensorial como Solto (S) diminui. A região de incerteza à esquerda do gráfico mostra que quando o valor da variável $C1$ está no intervalo -270 a -250 não é possível concluir com segurança que a avaliação sensorial seria preferencialmente do tipo Solto (S) ou Levemente Solto (LS). Já a região de incerteza central do gráfico mostra que quando o valor da variável $C1$ está no intervalo -115 a -95 não é possível concluir com segurança que a avaliação seria preferencialmente do tipo Pegajoso (P) ou Levemente Solto (LS). A região de incerteza à direita do gráfico mostra que quando o valor da variável $C1$ está no intervalo de -60 a -45 não é possível concluir com segurança que a avaliação sensorial seria preferencialmente do tipo Pegajoso (P) ou Muito Pegajoso (MP). Assim para valores da variável $C1$ pertencentes ao intervalo

−300 a −270 o arroz seria classificado por meio da avaliação sensorial como Solto (S) e para valores pertencentes ao intervalo de −250 a −115 o arroz seria classificado como Levemente Solto (LS). Já para os valores mais elevados de $C1$, entre −95 até −60 o arroz seria classificado como Pegajoso (P). E para valores de extrema direita de $C1$ acima de −45 o arroz seria classificado como Muito Pegajoso (MP).

Regressão Logística Binária

A fim de diminuir a porcentagem de observações que estão sendo classificadas erroneamente (23,61%) na avaliação de pegajosidade considerando quatro categorias: MP, P, LS e S, decidiu-se realizar uma análise que considera apenas duas categorias (Pegajoso e Solto), esperando que essa forneça classificações mais precisas.

Logo, foi ajustado um novo modelo que considera uma nova categoria Solto (S^*), que abrange as observações classificadas nas antigas categorias Levemente Solto (LS) e Solto (S) e uma nova categoria Pegajoso (P^*), que abrange as observações classificadas nas antigas categorias Muito Pegajoso (MP) e Pegajoso (P).

O novo modelo que é logito binário ajustou-se bem aos dados, pois ambas as variáveis $C1$ e $C2$ apresentaram efeito significativo nesse modelo. O p-valor de $C1$ foi menor do 0,00001 e o de $C2$ foi menor do que 0,003562, ou seja ambos os p-valores foram menores do que o nível de significância de 0,025 para cada variável.

Por meio desse modelo, a chance de avaliação sensorial na categoria Pegajoso é (P^*) é dada por:

$$P(Y_i = 0|C1_i, C2_i) = \frac{\exp[-2,62649 - (-0,08905)C1_i - 0,02109C2_i]}{1 + \exp[-2,62649 - (-0,08905)C1_i - 0,02109C2_i]}, \quad (3.32)$$

aonde o i -ésimo valor da variável $C1$ é dado por $C1_i$ e $\hat{\beta}_1 = -0,08905$ representa o efeito dessa variável na classificação da avaliação sensorial. Já o i -ésimo valor da variável $C2$ é dado por $C2_i$ e $\hat{\beta}_2 = 0,02109$ representa o efeito dessa variável na classificação da avaliação sensorial. A probabilidade de pertencer a categoria Solto (S^*) é dado por:

$$P(Y_i = 1|C1_i, C2_i) = 1 - P(Y_i = 0|C1_i, C2_i). \quad (3.33)$$

Utilizando-se o método de validação cruzada, obtém-se a matriz de classificação sensorial observada versus a classificação sensorial prevista por meio do modelo de regres-

são logística estimado (3.32) e (3.33) (Tabela 3.15). Baseado na Tabela 3.15, calcula-se a taxa do erro de classificação, que é de 5,56%. Esse erro é bem menor do que o observado na classificação da pegajosidade politômica sugerindo que realizar uma análise sensorial considerando uma quantidade menor de categorias pode aumentar a precisão na previsão da classificação por meio de modelos estatísticos, pois a distinção entre categorias é maior quanto menos categorias o modelo possuir.

Tabela 3.15 – Classificação da avaliação sensorial de pegajosidade para o ano de 2014 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando componentes principais de arroz de Terras Altas

	Classificação prevista		
	P*	S*	
Classificação real	P*	33	1
	S*	3	35

Oberva-se na curva ROC da Figura 3.16 que o modelo binário apresentou discriminação excelente na classificação das categorias Pegajoso (P*) e Solto (S*). Isso também pode ser visualizado por sua AUC (área sob a da curva) de 0,9565, que está bem próximo de 1, valor que representa a AUC da situação perfeita, quando todos os indivíduos são classificados corretamente por meio do modelo utilizado.

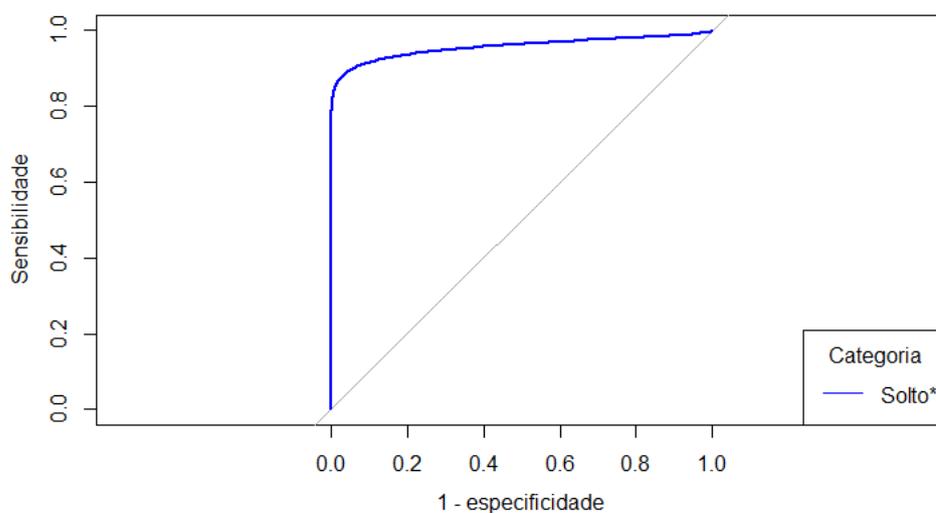


Figura 3.16 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2014, prevista por meio do modelo de regressão logística binário utilizando componentes principais de arroz de Terras Altas

Na Figura 3.17, observa-se que à medida que a variável $C1$ aumenta e a variável $C2$ diminui, a probabilidade do arroz receber avaliação sensorial como Solto (S^*) diminui. Assim, para valores baixos de $C1$ e altos de $C2$ o arroz seria classificado como Solto (S^*) e para valores elevados de $C1$ e baixos de $C2$ o arroz seria classificado como Pegajoso (P^*). Para valores intermediários de $C1$ e $C2$ não é possível concluir com segurança que a avaliação sensorial seria preferencialmente do tipo Pegajoso (P^*) ou Solto (S^*).

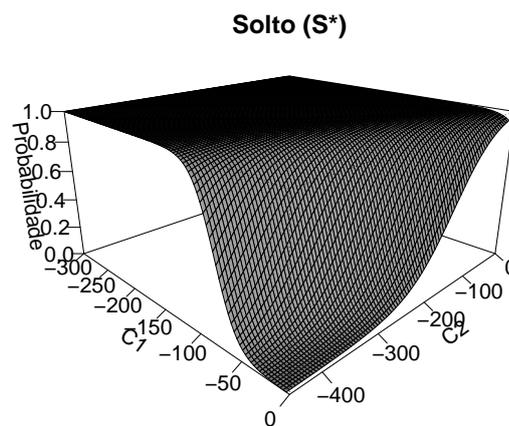


Figura 3.17 – Probabilidade do arroz receber avaliação sensorial como Solto (S^*) considerando diferentes valores das variáveis $C1$ e $C2$ de arroz de Terras Altas para o ano de 2014

Comparação da predição da avaliação sensorial da pegajosidade por medidas instrumentais de textura e medidas de viscosidade para arroz de Terras Altas

Ao se utilizarem as medidas de viscosidade considerando quatro categorias obtém-se um erro de classificação de 23,61% que é igual ao obtido por meio das medidas instrumentais de textura (23,61%). Já considerando duas categorias, o erro de classificação foi menor utilizando-se as medidas de viscosidade (5,56%) do que as medidas instrumentais de textura (8,33%). Isso sugere que para esse caso, foi possível obter um modelo com capacidade discriminatória igual ou superior utilizando as medidas de viscosidade ao invés das medidas instrumentais de textura. Isso é almejado para todos os casos, pois o custo financeiro da obtenção das medidas de viscosidade é inferior ao da obtenção das medidas instrumentais de textura.

3.3.3 Resumo dos resultados para os modelos analisados

Nas Tabelas (3.16), (3.17) e (3.18) apresentam-se resultados para todos os modelos analisados. A coluna **Ano** faz referência a quatro tipos de modelos; 2013, que se refere aos modelos obtidos utilizando as observações desse ano; 2014, que se refere aos modelos obtidos utilizando as observações desse ano; 2013 + 2014 que se refere aos modelos obtidos utilizando a totalidade das observações referentes a esses dois anos; 2013 \Rightarrow 2014, que se refere as previsões feitas para o ano de 2014 utilizando os modelos obtidos com as observações de 2013. A coluna **T.T.** se refere ao tipo de terreno, Terrenos Irrigados (Ir.) ou Terras Altas (Al.). A coluna **Med.** se refere ao tipo de medidas, instrumentais de textura (I.) e de viscosidade (V.). A coluna **Par.** se refere aos parâmetros estimados. As caselas simbolizadas pelo sinal de – indicam que o parâmetro não foi significativo naquele modelo, que a categoria em questão não foi definida para o modelo por falta de observações ou que não houve ajustamento dos dados ao modelo proposto.

Tabela 3.16 – Parâmetros estimados dos modelos logito cumulativo e binários

Ano	T.T.	Med.	Par.	Categorias						
				Dureza		Pegajosidade				
				LML	M	EP	MP	P	LS	P*(b)
2013	Ir.	I.	β_{0k}	-12,56599	-6,67342	-	-3,87249	-1,38460	1,17035	1,34115
			β	-0,05920	-0,05920	-	0,29158	0,29158	0,29158	0,28442
		V.	β_{0k}	-5,49880	0,52778	-	-0,83092	1,43170	3,97964	-0,86819
			β_1	0,01067	0,01067	-	-0,01570	-0,01570	-0,01570	-0,01453
	Al.	I.	β_{0k}	-	-	-	-3,99311	-2,42016	0,72436	2,57558
			β	-	-	-	-0,19884	-0,19884	-0,19884	-0,21135
		V.	β_{0k}	-	-	-	0,61931	2,21787	5,28512	-2,28400
			β_1	-	-	-	-0,01172	-0,01172	-0,01172	-0,01272
2014	Ir.	I.	β_{0k}	-	16,969*	-	-12,75267	-6,20096	-2,39945	6,722
			β	-	-0,133*	-	0,95451	0,95451	0,95451	1,044
		V.	β_{0k}	-	-1,07070*	-	-2,89439	-0,07179	3,98516	0,77983
			β_1	-	-0,01405*	-	0,03058	0,03058	0,03058	0,03863
	Al.	I.	β_{0k}	-	-	-	-7,69236	-4,88303	0,14126	6,5287
			β	-	-	-	0,46297	0,46297	0,46297	0,6515
		V.	β_{0k}	-	-	-	2,22402	4,86121	11,50592	-2,62649
			β_1	-	-	-	-0,04406	-0,04406	-0,04406	-0,08905
2013 + 2014	Ir.	I.	β_{0k}	-15,34694	-8,61320	-8,06987	-5,10620	-2,34159	0,41059	2,4837
			β	-0,07289	-0,07289	0,38302	0,38302	0,38302	0,38302	0,4105
		V.	β_{0k}	-5,94534	0,63767	-3,20004	-1,13468	1,14473	3,93851	-0,46913
			β_1	-0,01085	-0,01085	0,01759	0,01759	0,01759	0,01759	0,02080
	Al.	I.	β_{0k}	-4,70859	-0,75618	-6,43751	-5,20910	-3,20310	0,56986	3,8381
			β	-0,02166	-0,02165986	0,28460	0,28460	0,28460	0,28460	0,3496
		V.	β_{0k}	-	-	-0,19304	1,07409	2,96376	6,65150	-1,82311
			β_1	-	-	-0,01839	-0,01839	-0,01839	-0,01839	-0,03238
2013 + 2014	Ir.	I.	β_{0k}	-15,34694	-8,61320	-8,06987	-5,10620	-2,34159	0,41059	2,4837
			β	-0,07289	-0,07289	0,38302	0,38302	0,38302	0,38302	0,4105
		V.	β_{0k}	-5,94534	0,63767	-3,20004	-1,13468	1,14473	3,93851	-0,46913
			β_1	-0,01085	-0,01085	0,01759	0,01759	0,01759	0,01759	0,02080
	Al.	I.	β_{0k}	-4,70859	-0,75618	-6,43751	-5,20910	-3,20310	0,56986	3,8381
			β	-0,02166	-0,02165986	0,28460	0,28460	0,28460	0,28460	0,3496
		V.	β_{0k}	-	-	-0,19304	1,07409	2,96376	6,65150	-1,82311
			β_1	-	-	-0,01839	-0,01839	-0,01839	-0,01839	-0,03238

¹As caselas com uma * mostram os parâmetros de uma regressão logística binária considerando apenas duas categorias: Macio (M) e Extremamente Macio (EM). Esse modelo foi ajustado porque a frequência das outras categorias de dureza era nula.

²A coluna P*(b) mostra os parâmetros de uma regressão logística binária ao se recategorizar a variável pegajosidade em duas categorias: Pegajoso (P*) e Solto (S*).

Tabela 3.17 – Taxa do erro de classificação

Ano	T.T.	Med.	Dureza	Pegajosidade	
				Politômica	Binária
2013	Ir.	I.	20,51%	55,56%	36,75%
		V.	21,37%	47,86%	39,32%
	Al.	I.	-	44,44%	33,33%
		V.	-	44,44%	30,56%
2014	Ir.	I.	14,67%	24%	12%
		V.	20%	37,33%	12%
	Al.	I.	-	23,61%	8,33%
		V.	-	23,61%	5,56%
2013 +	Ir.	I.	17,71%	47,4%	23,96%
		V.	18,23%	46,35%	26,56%
	Al.	I.	26,39%	39,58%	18,75%
		V.	-	50%	17,36%
2013 ↓ 2014	Ir.	I.	17,33%	29,33%	22,67%
		V.	16%	46,67%	28%
	Al.	I.	-	30,56%	13,89%
		V.	-	44,44%	33,33%

¹As taxas do erro de classificação fazem referência as tabelas de classificação de avaliação sensorial versus a classificação prevista, que podem ser encontradas no Apêndice.

Tabela 3.18 – Área sob a curva ROC

Ano	T.T.	Med.	Categorias								
			Dureza			Pegajosidade					
			LML	M	EM	EP	MP	P	LS	S	S*
2013	Ir.	I.	0,64048	0,81335	0,82275	-	0,85834	0,56518	0,70220	0,74606	0,7353
		V.	0,65666	0,81544	0,82617	-	0,81864	0,60462	0,70869	0,63813	0,7417
	Al.	I.	-	-	-	-	0,78504	0,65410	0,72502	0,64209	0,7565
		V.	-	-	-	-	0,93098	0,51115	0,72102	0,50791	0,7285
2014	Ir.	I.	-	-	0,9032	-	1	0,89080	0,77216	0,64561	0,9301
		V.	-	-	0,7741	-	0,95667	0,77855	0,82397	0,74120	0,9276
	Al.	I.	-	-	-	-	0,77644	0,89877	0,86812	0,89134	0,943
		V.	-	-	-	-	0,95985	0,90570	0,89842	0,68244	0,9565
2013 +	Ir.	I.	0,66751	0,66751	0,84861	0,94191	0,79933	0,65662	0,75747	0,70067	0,7922
		V.	0,68029	0,80988	0,80159	0,84899	0,81465	0,64318	0,76064	0,68758	0,8114
	Al.	I.	0,63433	0,70999	0,68101	0,74376	0,76933	0,78394	0,79758	0,70052	0,8512
		V.	-	-	-	0,91830	0,89837	0,61683	0,81231	0,52021	0,8476
2013 ↓ 2014	Ir.	I.	-	0,89776	0,90316	-	1	0,87827	0,85745	0,64561	0,9301
		V.	-	0,85013	0,79937	-	0,97175	0,74987	0,82125	0,67320	0,924
	Al.	I.	-	-	-	-	0,77644	0,93212	0,85033	0,89134	0,943
		V.	-	-	-	-	0,97445	0,61472	0,82760	0,61589	0,8062

¹As áreas sob a curva ROC fazem referência as curvas de classificação ROC, que podem ser encontradas no Apêndice.

4 Conclusão

Este trabalho avaliou a possível substituição da avaliação sensorial da textura do arroz, a partir de medidas instrumentais de textura e de medidas de viscosidade. A avaliação sensorial é o método mais apropriado e indicado para a análise da textura de arroz cozido, entretanto sua substituição é almejada devido ao seu alto custo financeiro e necessidade de treinamento, aptidão e disponibilidade de pessoas.

A análise de Componentes Principais mostrou-se eficiente na redução da dimensão do estudo. De modo que, apenas 1 ou 2 componentes principais foram suficientes para explicar 80% ou mais da variância dos dados das 5 medidas de viscosidade.

Na análise dos dados, foram levados em consideração dois tipos de terreno: Terrenos Irrigados e Terras Altas. De modo geral, parece não haver grandes diferenças entre os resultados desses dois tipos de Terreno. Salvo para alguns casos específicos, como por exemplo, foi possível obter modelos preditivos para a dureza sensorial de arroz de Terrenos Irrigados, enquanto que para a dureza de arroz de Terras Altas isso não foi possível na maioria das vezes. Essa diferença está possivelmente relacionada ao fato de uma melhor distribuição das observações entre as categorias da dureza para arroz de Terrenos Irrigados.

Foram realizadas análises para o ano de 2013 e de 2014 separadamente. Os modelos de 2014 apresentaram taxas de erros de classificação consideravelmente menores que os de 2013, indicando que foi possível obter uma melhor discriminação com os dados de 2014. No entanto, ressalta-se que esse resultado pode ser apenas uma consequência de haver sido coletado uma menor quantidade de observações para o ano de 2014 do que para o ano de 2013.

Para as observações conjuntas dos anos de 2013 e de 2014 e de forma isolada, o poder preditivo dos modelos que utilizam medidas de viscosidade é similar a dos modelos que utilizam medidas instrumentais de textura. Porém, ao usar os modelos obtidos com as

observações de 2013 para prever as observações de 2014, melhores previsões foram obtidas para modelos que utilizam medidas instrumentais de textura do que para modelos que utilizam medidas de viscosidade.

5 Considerações finais

A avaliação sensorial da textura de arroz cozido resulta da avaliação de duas medidas essenciais de textura: a dureza e a pegajosidade do arroz. Cada uma dessas medidas são definidas para sete categorias. Todos os modelos considerados, seja de arroz de terrenos irrigados ou de terras altas, possuíam no mínimo duas categorias nas quais nenhuma observação da amostra havia sido classificada na avaliação sensorial. Para alguns modelos, chegou-se a ter até cinco categorias nulas. Além disso, algumas das categorias que não eram nulas apresentavam pouquíssimas observações. A falta de observações em categorias impossibilitou consideravelmente a melhora das técnicas de classificação, o que dificulta a substituição da avaliação sensorial da textura do arroz, a partir de medidas instrumentais de textura e de medidas de viscosidade.

Para a obtenção de modelos preditivos futuros, com uma alta taxa de acerto de classificação, se faz necessário uma redução das categorias levadas em consideração nas medidas sensoriais de dureza e pegajosidade. Uma outra alternativa, para solucionar o problema de categorias nulas, é a aplicação de algum método de cozimento do arroz que viabilize a avaliação de todas as categorias possíveis pelos avaliadores da avaliação sensorial.

Uma medida que pode atenuar o problema de categorias com poucas observações é a utilização de informações a priori de anos anteriores ao que se quer prever, através de técnicas de estatística bayesiana.

Referências Bibliográficas

- [1] AGRESTI, A. **An introduction to categorical data analysis**. 2. ed. Florida: Wiley, 2007. 15, 17, 19, 22, 23
- [2] BATEY, L.; CURTIN, B. M. **Measurement of amylose/amylopectin ratio by high-performance liquid chromatography**. *Starch/Starke*, 48. p. 338-344, 1996. 11
- [3] BUENO, P.D.F. **Viscoamilografia na estimativa do teor de amilose e características de consumo de arroz**. Universidade Federal de Pelotas. Pelotas, 2008. 10
- [4] CARVALHO, R.N. ; Bassinello, P.Z. et. al. **Pegajosidade arroz polido cozido em texturômetro**. Embrapa Arroz e Feijão. Comunicado Técnico. 6 p. Santo Antônio de Goiás, Goiânia. 9
- [5] FITZGERALD, M.A. ; MCCOUCH, S.R. ; HALL, R.D. **Not just a grain of rice: the quest for quality**. *Trends in Plant Science*. Londres,2008. 1
- [6] FITZGERALD, M. A.; McCOUCH, S. R.; HALL, R. D. **Not just a grain of rice: the quest for quality**. *Trends in Plant Science*. Oxford, vol. 14. n.3 p.133-139, 2009. 10
- [7] HOSMER, D.W. ; LEMESHOW, S. **Applied logistic regression**. 2. ed. John Wiley & Sons, 2000. 21
- [8] JOHNSON, R.A. ; WICHERN, D. W. **Applied multivariate statistical analysis**. 6. ed. Prentice Hall, 2002. 3, 4, 13, 14, 15, 23
- [9] KLEINBAUM, D.G. ; KLEIN, M. **Logistic regression: A Self-Learning Text**. 3. ed. Springer. New York, 2010. 24, 25, 26, 27
- [10] ISO. **Sensory analysis : Vocabulary international organization for standartization ISSO 5492: 1992**. 1992. 1

- [11] LAWLESS, H.T.; HEYMANN, H. **Sensory evaluation of food: principles and practices**. 2. ed. New York. Springer, 619 p. New York, 2010. 5
- [12] LISLE, A. J.; MARTIN, M.; FITZGERALD, M. A. **Chalky and translucent rice grains differ in starch composition and structure and cooking properties**. Cereal Chem, 77. p. 627-632, 2000.
- [13] MARTINÉZ, C. ; CUEVAS, F. **Evaluation of culinary and milling quality of rice**. 3. ed. Centro Internacional de Agricultura Tropical. Cali, 1989. 11
- [14] MCCULLAGH, P. ; NELDER J. A. **Generalized linear models** 2. ed. Chapman e Hall/CRC, 1989.
- [15] MINIM, P. V. R. **Análise sensorial: estudos com consumidores**. 2. ed. Viçosa: UFV, 2006. 5, 6, 7
- [16] MOREIRA, A.M. **Proposed methodology for quality pre-Selection of rice populations**. Cereal Chemistry. 91. vol. 2. n.2 p.201-206, 2014. 10, 11
- [17] SANTOS, T.P.B. et al. **Efeito dos grãos gessados nos teores de amilose e propriedades de pasta do arroz**. Anais da 63ª Reunião Anual da SBPC. Goiânia, 2011. 1
- [18] Saunders J. **Physicochemical properties of select western canadian wheat starches and their relationship to liquefaction and fermentative bio-ethanol performance**. University of Manitoba, Winnipeg, Manitoba, Canada, (2010). 12
- [19] SESMAT, A.; MEULLENET, J.F. **Prediction of rice sensory texture attributes from a single compression test, multivariate regression, and a stepwise model optimization method**. Journal of Food Science, vol. 66. n.1 p.124- 131, 2001. 9
- [20] STONE, H.; SIDEL, J.L. **Sensory evaluation practices**. 3. ed. Academic Press. Orlando, 2004. 5
- [21] TEBA, C.S. et al. **Efeito dos parâmetros de extrusão sobre as propriedades de pasta de massas alimentícias pré-cozidas de arroz e feijão**. Alimentos e Nutrição Araquara. 20. vol. 3. n. 2009. 11
- [22] VON BORRIES, G.F.; BASSINELLO, P.Z.; RIOS, E.S. et al. **Modelos estatísticos para avaliação da qualidade culinária de arroz: textura e propriedades viscoamilográficas**. XII Conferência Internacional do Arroz para a América Latina e Caribe; 2015 Fev 23-26. Porto Alegre, RS, Brasil. 1

- [23] VON BORRIES, G.F.; BASSINELLO, P.Z.; RIOS, E.S. et al. **Aplicação de modelo de regressão logística para predição da qualidade culinária de arroz**. 9º Seminário Jovens Talentos; 2015, Jun 29-30. Santo Antônio de Goiás, GO, Brasil. 1

Apêndice

5.1 Tabelas e Figuras da Regressão Logística para o ano de 2013

5.1.1 Avaliação sensorial por medidas instrumentais em Terrenos Irrigados

Pegajosidade

Tabela 5.1 – Classificação da avaliação sensorial de pegajosidade para o ano de 2013 versus a classificação prevista, por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de arroz de terras altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	17	11	3	0
	P	7	16	22	0
	LS	0	16	19	0
	S	0	1	5	0

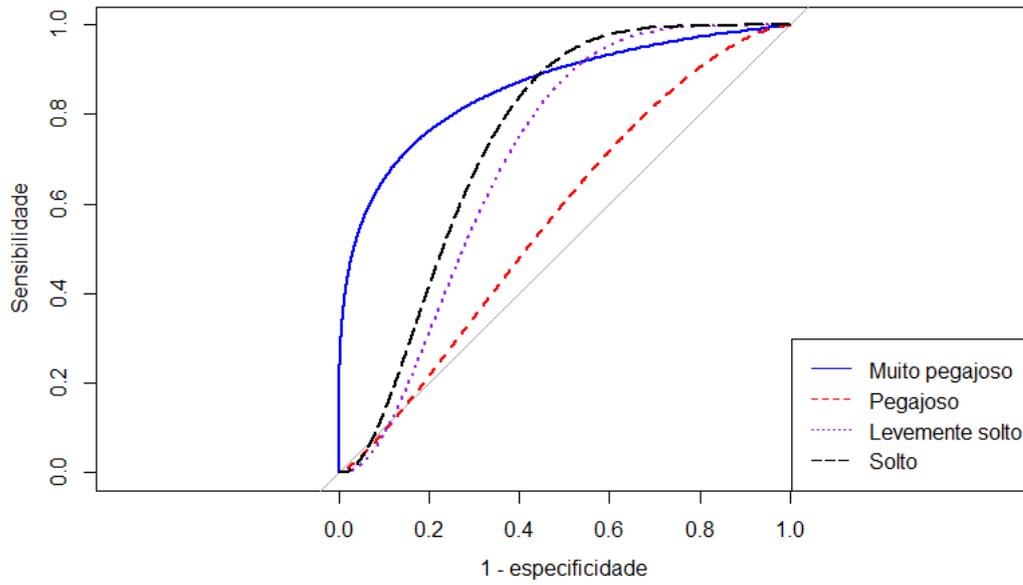


Figura 5.1 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2013, prevista por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

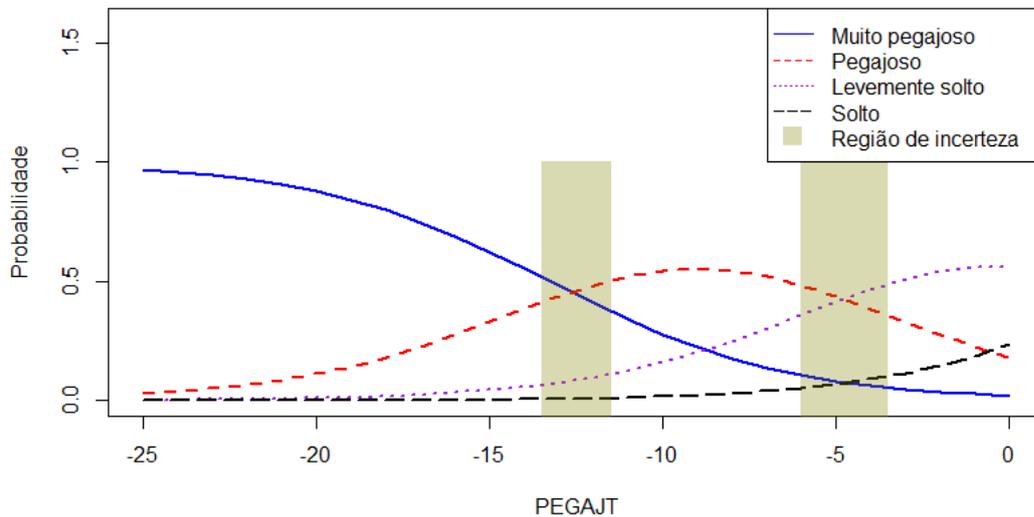


Figura 5.2 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores da pegajosidade instrumental de arroz de arroz de arroz de Terras Altas para o ano de 2013

Pegajosidade binária

Tabela 5.2 – Classificação da avaliação sensorial de pegajosidade para o ano de 2013 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

	Classificação prevista		
	P*	S*	
Classificação real	P*	51	25
	S*	18	23

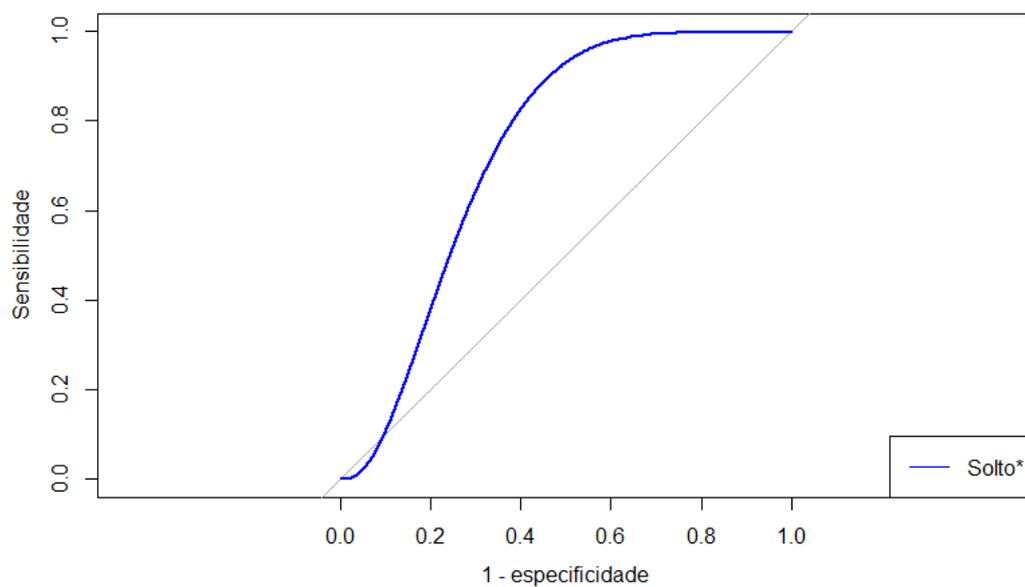


Figura 5.3 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2013, prevista por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

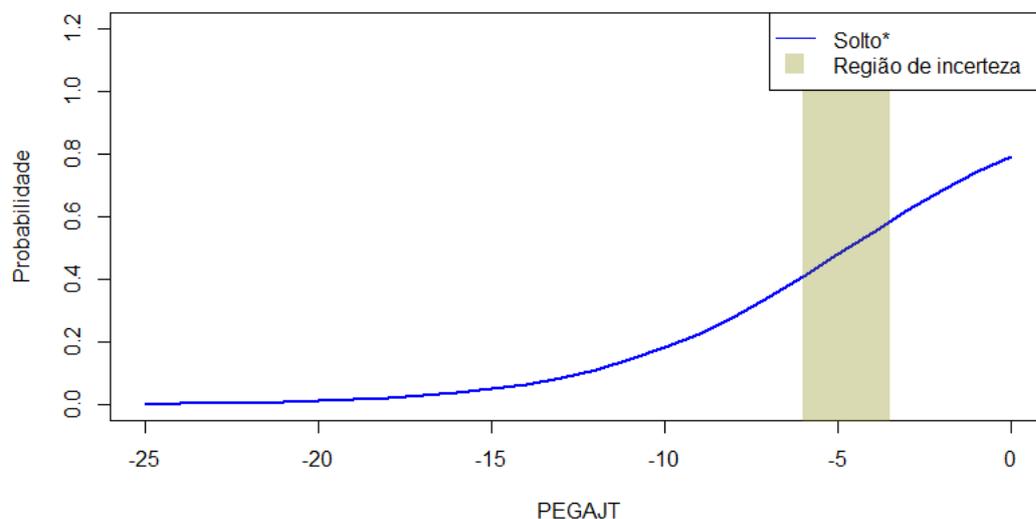


Figura 5.4 – Probabilidade do arroz receber avaliação sensorial como Solto (S^*) considerando diferentes valores da pegajosidade instrumental de arroz de arroz de arroz de Terras Altas para o ano de 2013

5.1.2 Avaliação sensorial por medidas do perfil viscoamilográfico em Terrenos Irrigados

Pegajosidade

Tabela 5.3 – Classificação da avaliação sensorial de pegajosidade para o ano de 2013 versus a classificação prevista, por meio do modelo de regressão logística utilizando componentes principais de arroz de arroz de arroz de Terras Altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	22	6	3	0
	P	7	25	13	0
	LS	2	19	14	0
	S	0	4	2	0

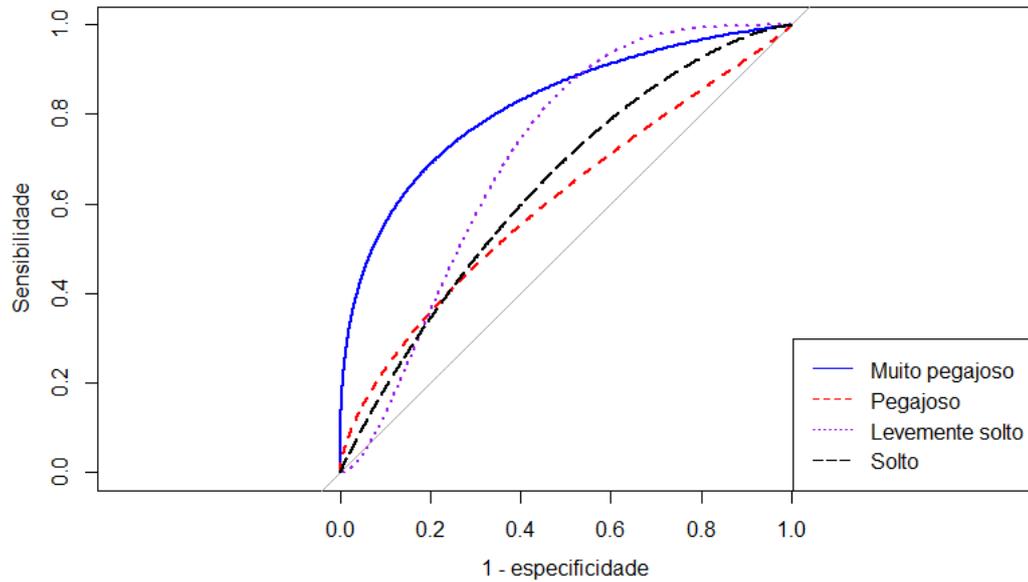


Figura 5.5 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2013, prevista por meio do modelo de regressão logística utilizando componentes principais de arroz de arroz de Terras Altas

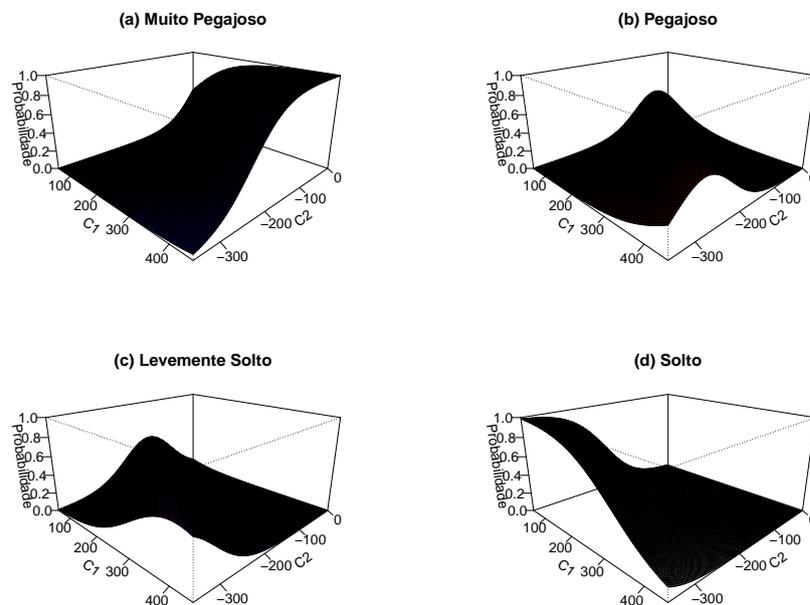


Figura 5.6 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores das variáveis C_1 e C_2 de arroz de arroz de Terras Altas para o ano de 2013

Pegajosidade binária

Tabela 5.4 – Classificação da avaliação sensorial de pegajosidade para o ano de 2013 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando componentes principais de arroz de arroz de arroz de Terras Altas

	Classificação prevista	
	P*	S*
Classificação real	P*	61
	S*	31

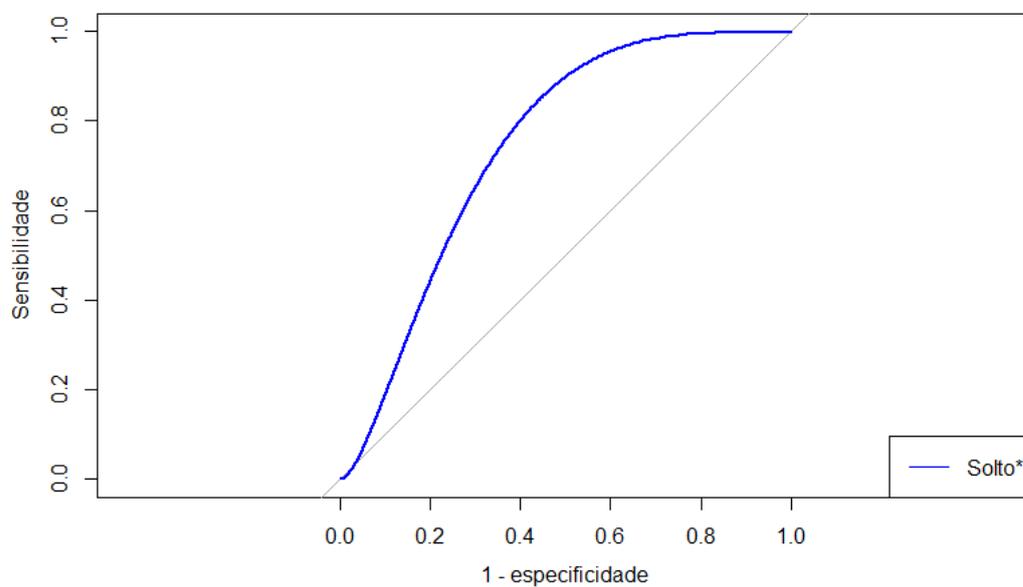


Figura 5.7 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2013, prevista por meio do modelo de regressão logística binário utilizando componentes principais de arroz de arroz de arroz de Terras Altas

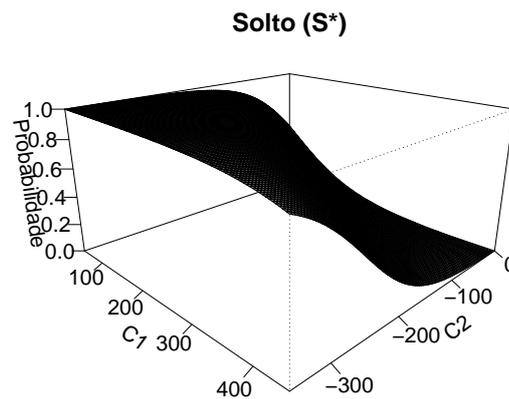


Figura 5.8 – Probabilidade do arroz receber avaliação sensorial como Solto (S^*) considerando diferentes valores das variáveis $C1$ e $C2$ de arroz de arroz de arroz de Terras Altas para o ano de 2013

5.1.3 Avaliação sensorial por medidas instrumentais em de arroz de Terras Altas

Pegajosidade

Tabela 5.5 – Classificação da avaliação sensorial de pegajosidade para o ano de 2013 versus a classificação prevista, por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de Terras Altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	4	1	7	0
	P	2	4	12	0
	LS	1	3	32	0
	S	0	0	6	0

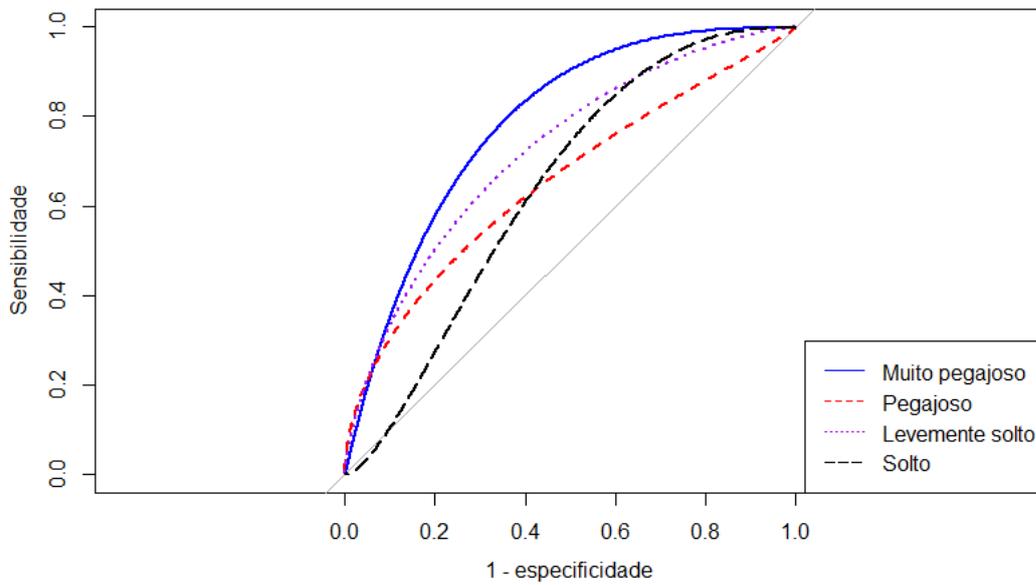


Figura 5.9 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2013, prevista por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de Terras Altas

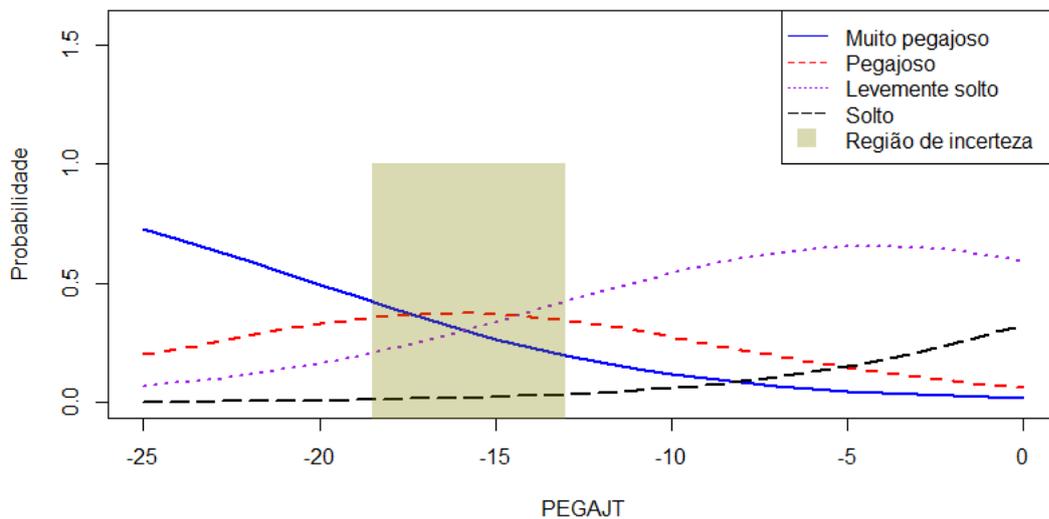


Figura 5.10 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores da pegajosidade instrumental de arroz de Terras Altas para o ano de 2013

Pegajosidade binária

Tabela 5.6 – Classificação da avaliação sensorial de pegajosidade para o ano de 2013 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de Terras Altas

	Classificação prevista	
	P*	S*
Classificação real	P*	18
	S*	12
		30

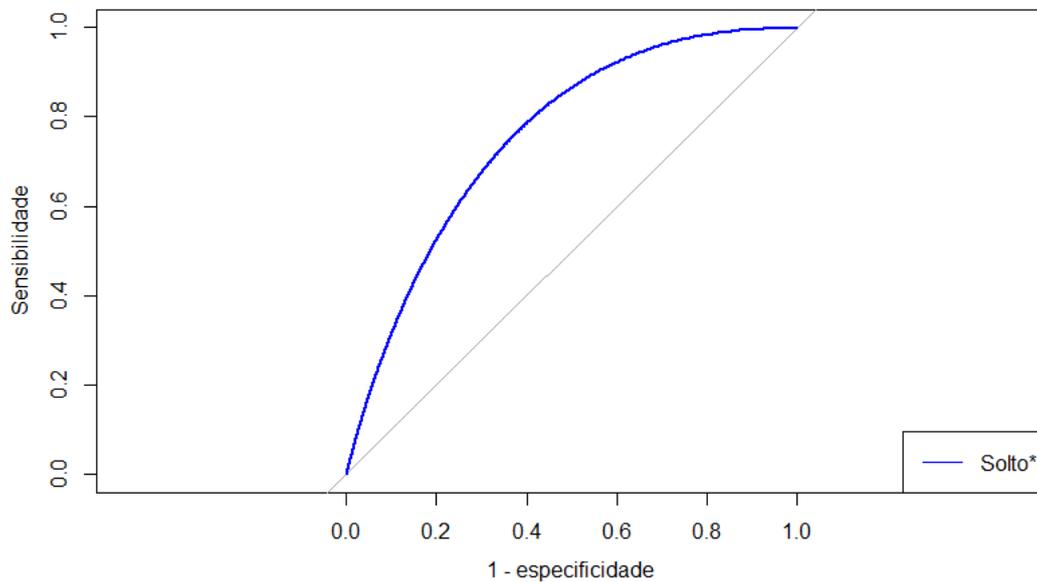


Figura 5.11 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2013, prevista por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de Terras Altas

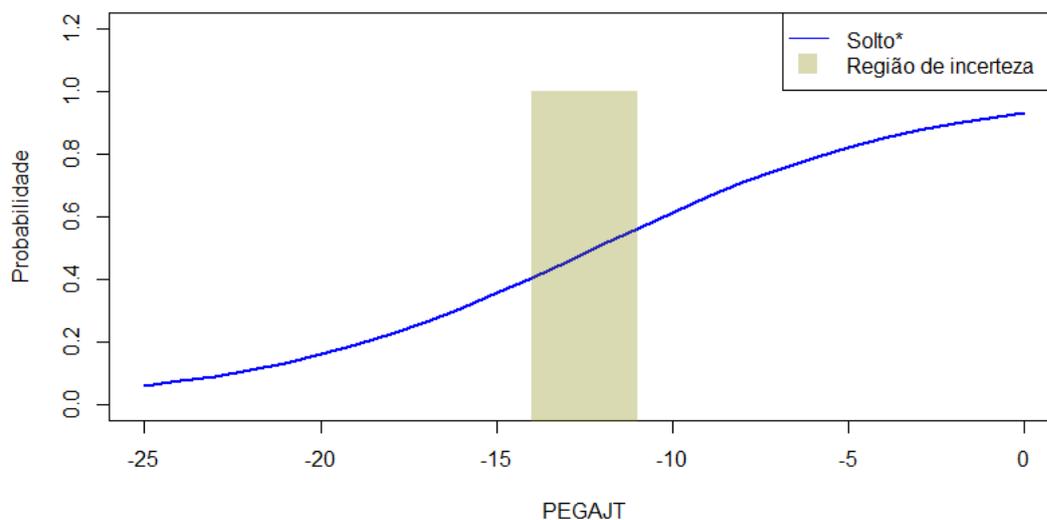


Figura 5.12 – Probabilidade do arroz receber avaliação sensorial como Solto (S*) considerando diferentes valores da pegajosidade instrumental de arroz de Terras Altas para o ano de 2013

5.1.4 Avaliação sensorial por medidas do perfil viscoamilográfico em de arroz de Terras Altas

Pegajosidade

Tabela 5.7 – Classificação da avaliação sensorial de pegajosidade para o ano de 2013 versus a classificação prevista, por meio do modelo de regressão logística utilizando componentes principais de arroz de Terras Altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	5	5	2	0
	P	1	1	16	0
	LS	0	2	34	0
	S	0	0	6	0

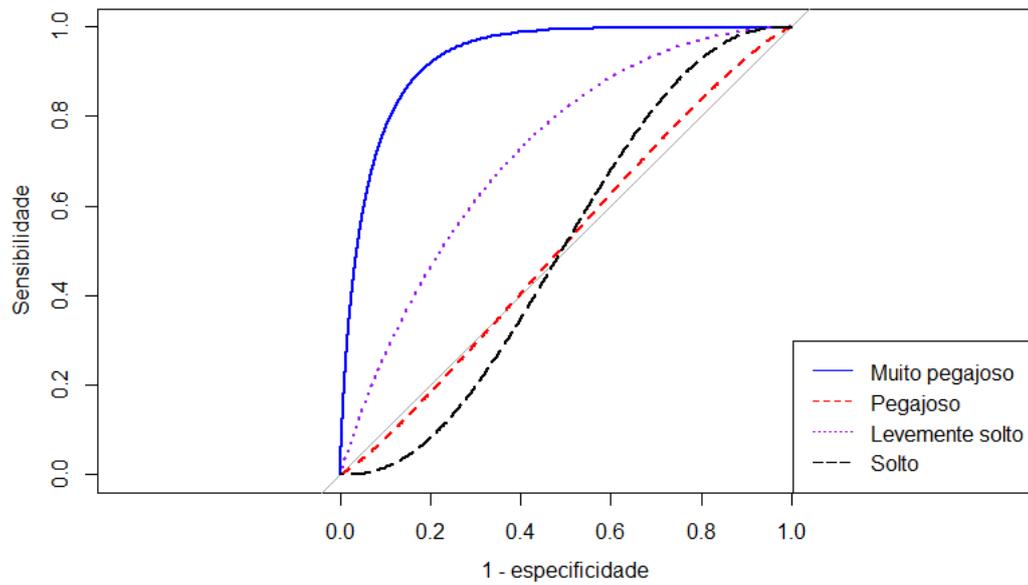


Figura 5.13 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2013, prevista por meio do modelo de regressão logística utilizando componentes principais de arroz de Terras Altas

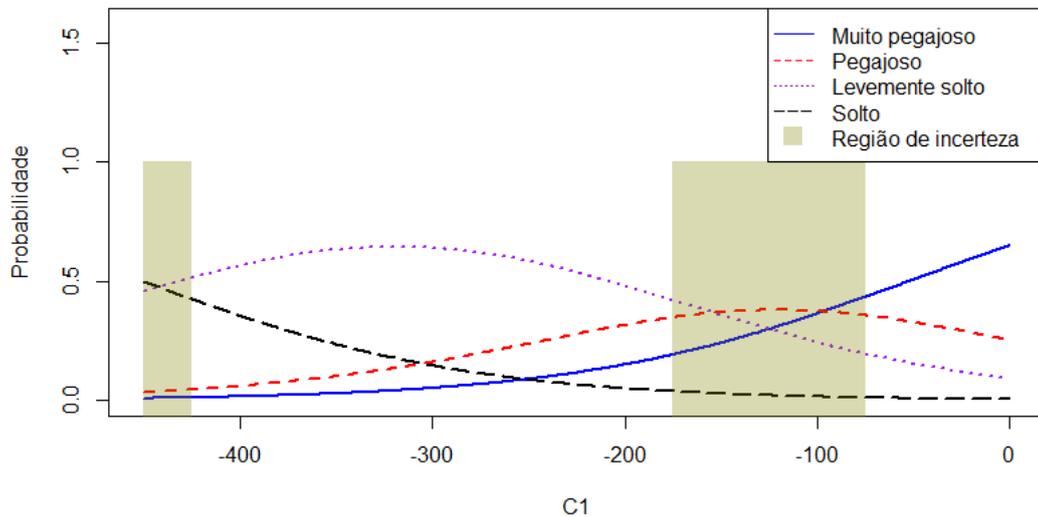


Figura 5.14 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores das variável C1 de arroz de Terras Altas para o ano de 2013

Pegajosidade binária

Tabela 5.8 – Classificação da avaliação sensorial de pegajosidade para o ano de 2013 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando componentes principais de arroz de Terras Altas

	Classificação prevista	
	P*	S*
Classificação real	P*	14
	S*	6

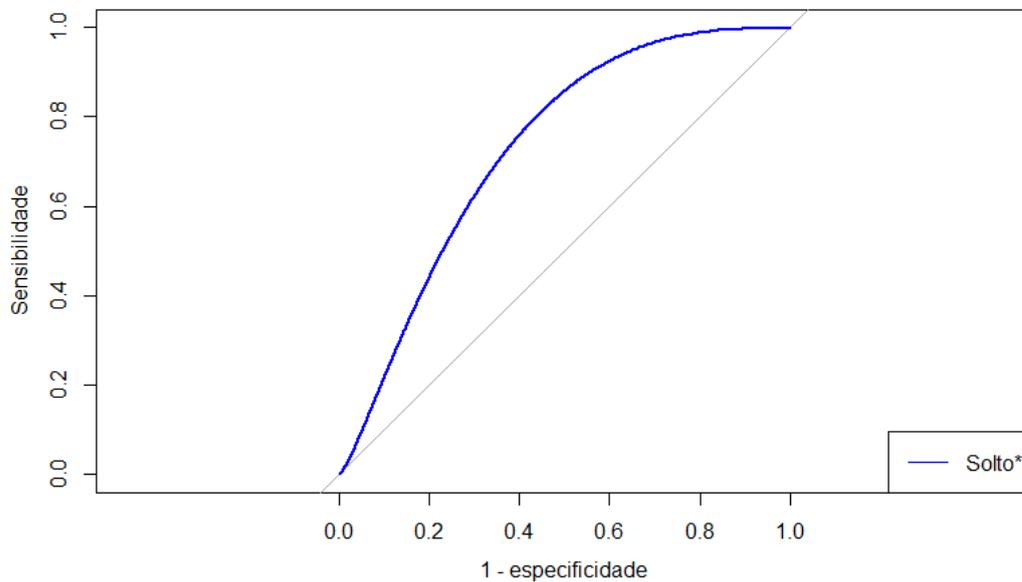


Figura 5.15 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2013, prevista por meio do modelo de regressão logística binário utilizando componentes principais de arroz de Terras Altas

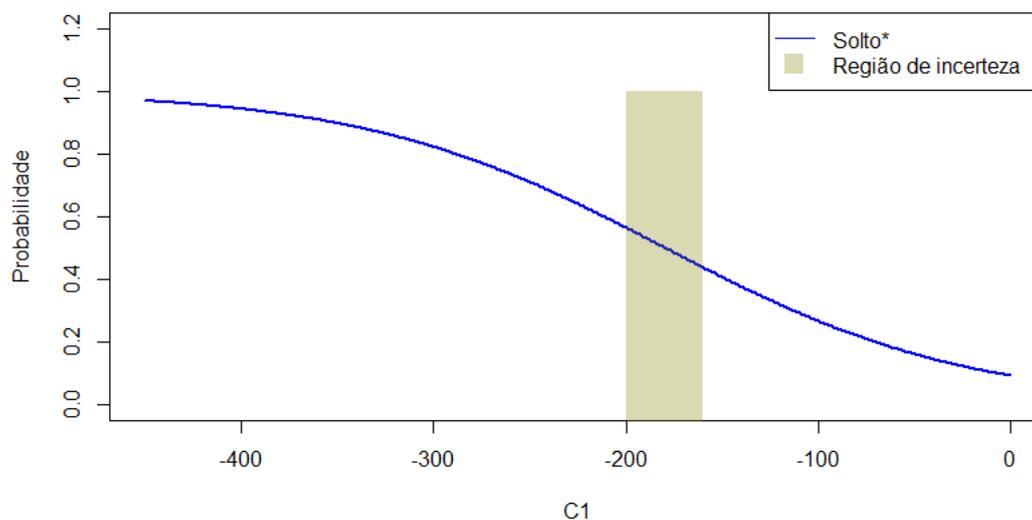


Figura 5.16 – Probabilidade do arroz receber avaliação sensorial como Solto (S^*) considerando diferentes valores da variável $C1$ de arroz de Terras Altas para o ano de 2013

5.2 Tabelas e Figuras da Regressão Logística para o ano de 2014

5.2.1 Avaliação sensorial por medidas instrumentais em Terrenos Irrigados

Dureza

Tabela 5.9 – Classificação sensorial de dureza de arroz de arroz de Terras Altas para o ano de 2014

Dureza	Frequência	Porcentagem
Extremamente firme	0	0%
Muito firme	0	0%
Levemente firme	0	0%
Macio com centro firme	0	0%
Ligeiramente macio	0	0%
Macio	62	82,67%
Extremamente macio	13	17,33%

Tabela 5.10 – Classificação da avaliação sensorial de dureza para o ano de 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando a dureza instrumental de arroz de arroz de arroz de Terras Altas

		Classificação prevista	
		M	EM
Classificação real	M	58	4
	EM	7	6

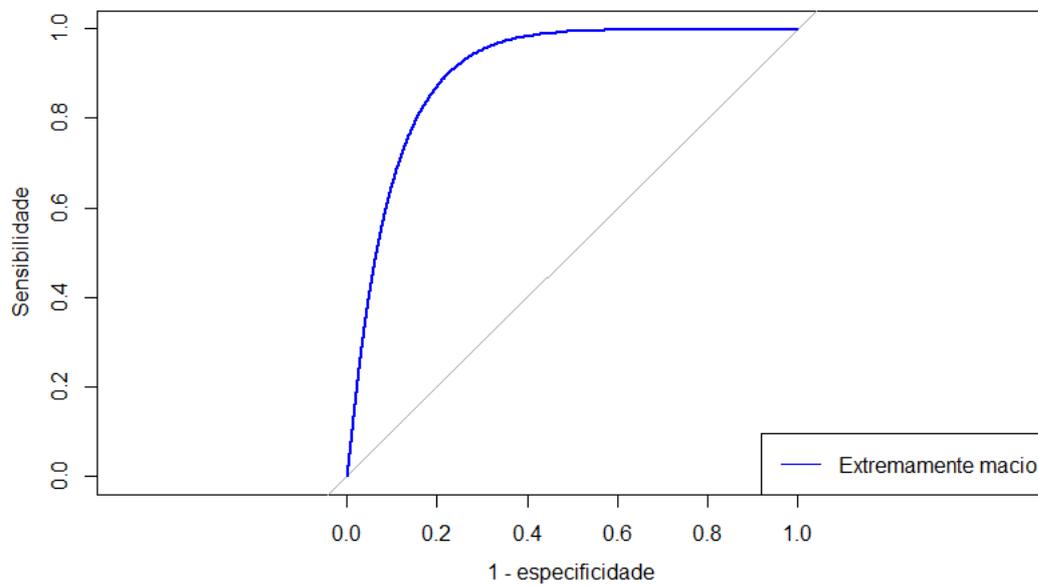


Figura 5.17 – Curva de classificação ROC da avaliação sensorial de dureza para o ano de 2014, prevista por meio do modelo de regressão logística utilizando a dureza instrumental de arroz de arroz de arroz de Terras Altas

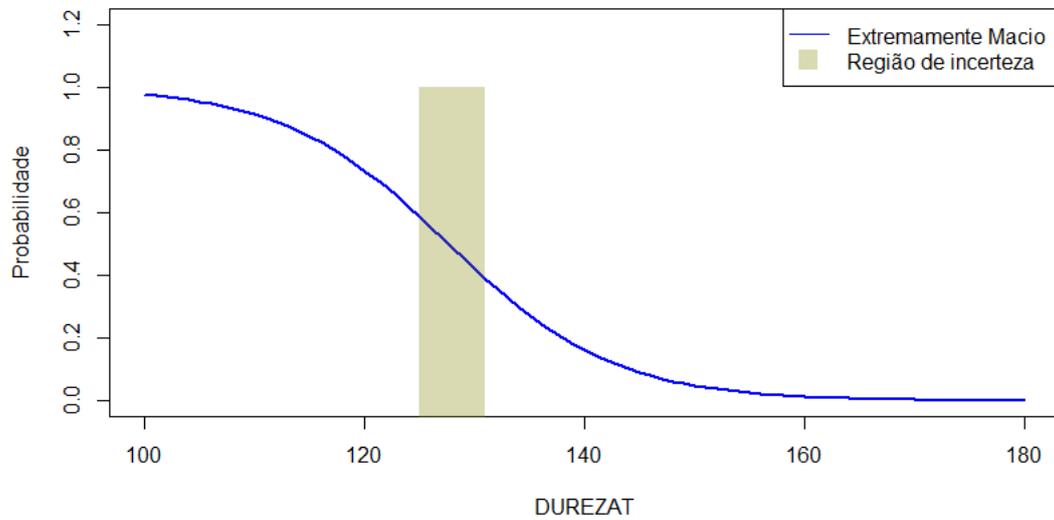


Figura 5.18 – Probabilidades das categorias de avaliação sensorial de dureza considerando diferentes valores da dureza instrumental de arroz de arroz de Terras Altas para o ano de 2014

Pegajosidade

Tabela 5.11 – Classificação sensorial de pegajosidade arroz de arroz de arroz de Terras Altas para o ano de 2014

Pegajosidade	Frequência	Pegajosidade
Extremamente pegajoso	11	14,67%
Muito pegajoso	1	1,33%
Pegajoso	19	25,33%
Levemente solto	36	48,00%
Solto	8	10,67%
Muito solto	0	0%
Extremamente solto	0	0%

Tabela 5.12 – Classificação da avaliação sensorial de pegajosidade para o ano de 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	11	1	0	0
	P	0	12	7	0
	LS	0	2	34	0
	S	0	0	8	0

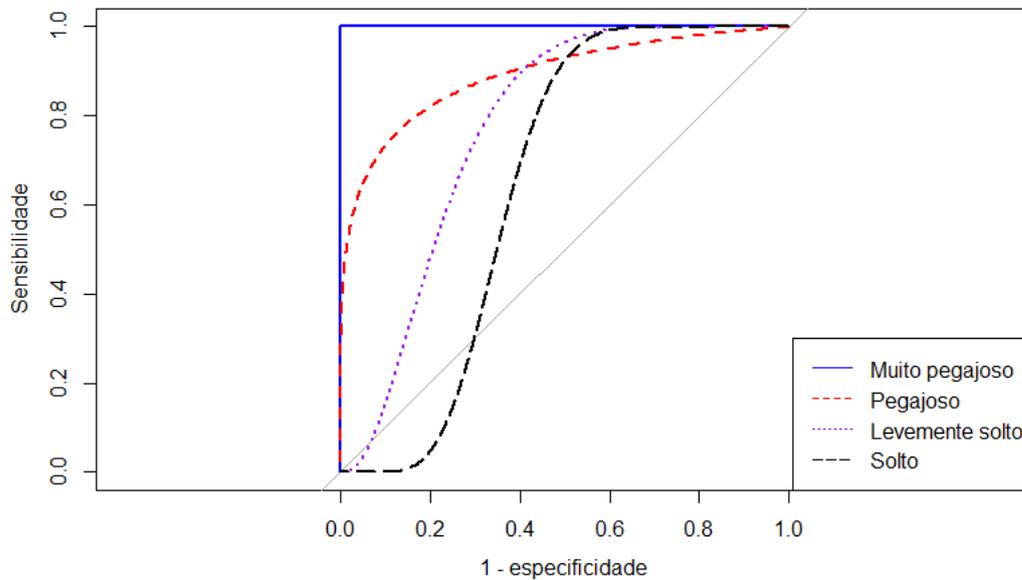


Figura 5.19 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2014, prevista por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

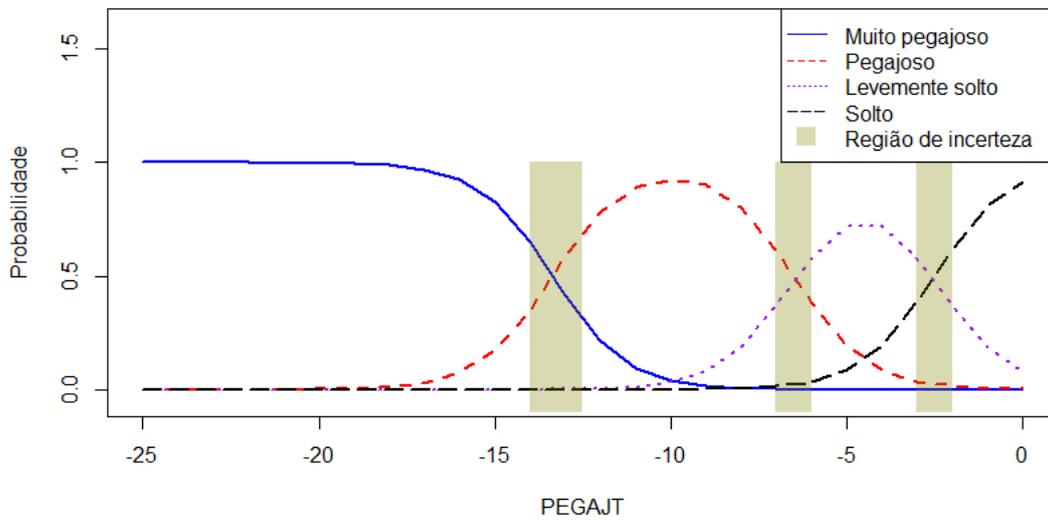


Figura 5.20 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores da pegajosidade instrumental de arroz de arroz de Terras Altas para o ano de 2014

Pegajosidade binária

Tabela 5.13 – Classificação da avaliação sensorial de pegajosidade para o ano de 2014 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

		Classificação prevista	
		P*	S*
Classificação real	P*	24	7
	S*	2	42

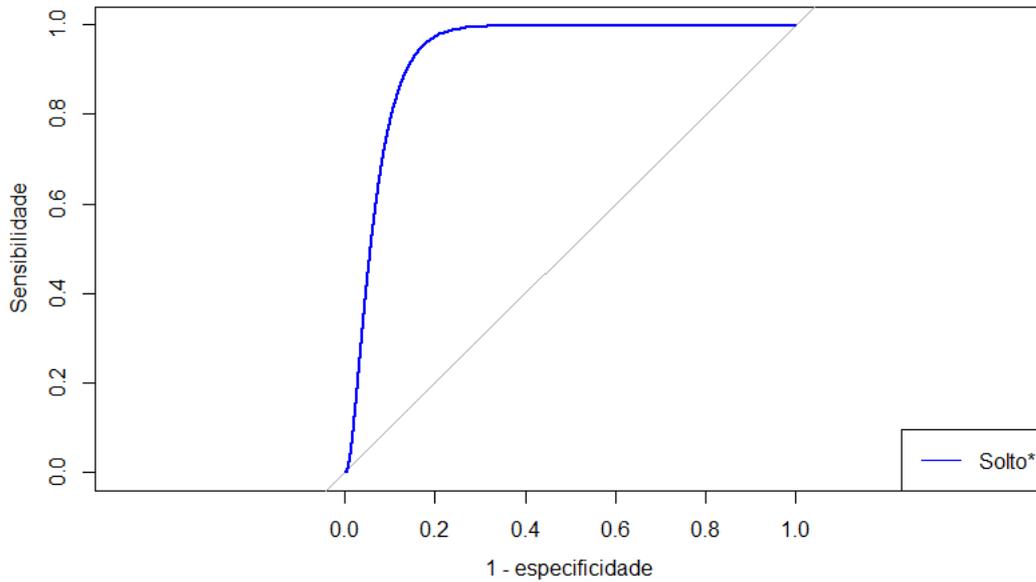


Figura 5.21 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2014, prevista por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

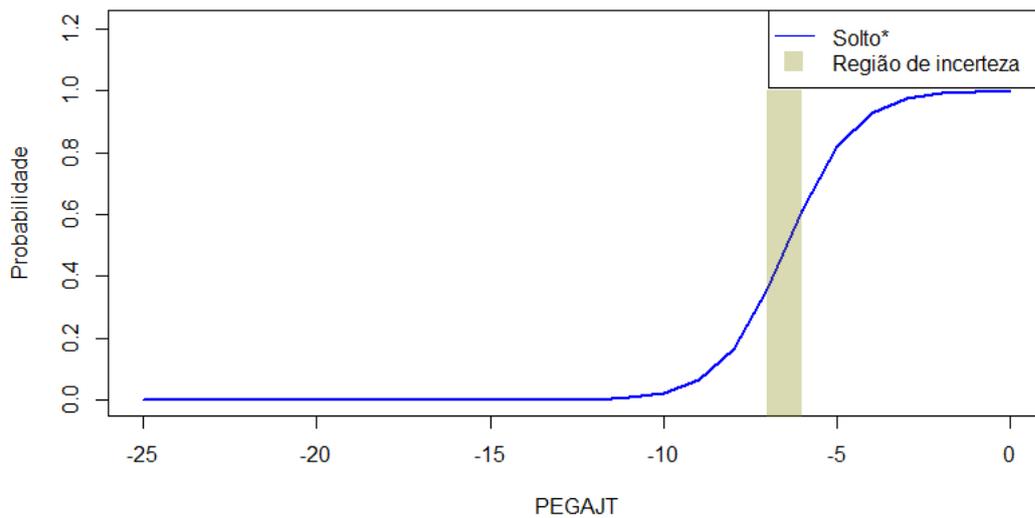


Figura 5.22 – Probabilidade do arroz receber avaliação sensorial como Solto (S*) considerando diferentes valores da pegajosidade instrumental de arroz de arroz de arroz de Terras Altas para o ano de 2014

5.2.2 Avaliação sensorial por medidas do perfil viscoamilográfico em Terrenos Irrigados

Tabela 5.14 – Contribuição de cada variável nas duas primeiras componentes principais para Terrenos Irrigados e coeficiente de correlação entre as variáveis dos dados e as componentes principais selecionadas

	Componentes principais	
	Coeficiente de correlação	
	Primeira componente	Segunda componente
TAAFIA	-0,13875	0,66958
	-0,227	0,871
TG	0,44403	0,38332
	0,727	0,498
PEAK	-0,58554	0,07867
	-0,959	0,102
BREAKDOWN	-0,50216	-0,37619
	-0,822	-0,489
FINAL	-0,43425	0,50696
	-0,711	0,659

Dureza

Tabela 5.15 – Classificação da avaliação sensorial de dureza para o ano de 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando componentes principais de arroz de arroz de arroz de Terras Altas

		Classificação prevista	
		M	EM
Classificação real	M	58	4
	EM	11	2

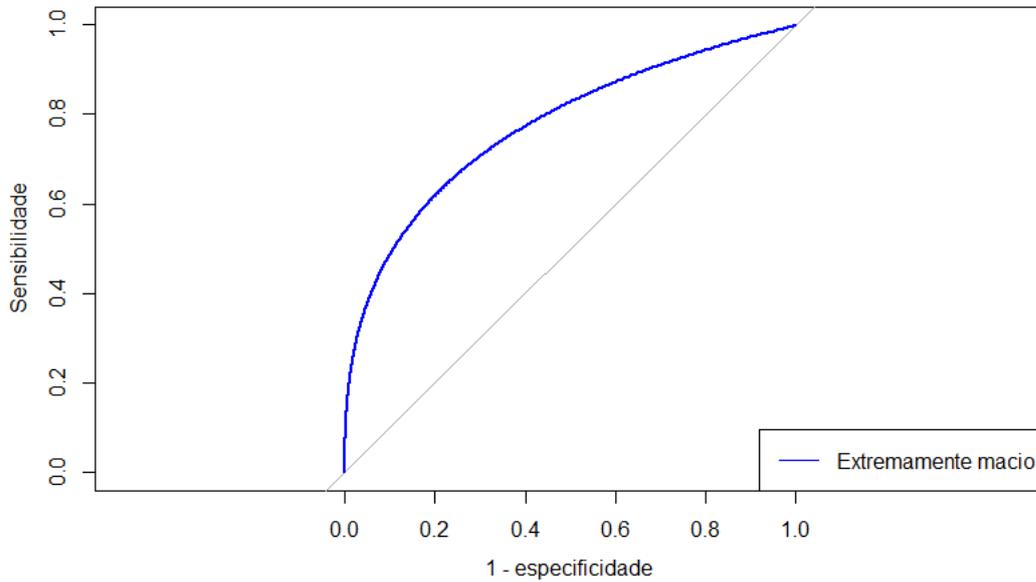


Figura 5.23 – Curva de classificação ROC da avaliação sensorial de dureza para o ano de 2014, prevista por meio do modelo de regressão logística utilizando componentes principais de arroz de arroz de arroz de Terras Altas

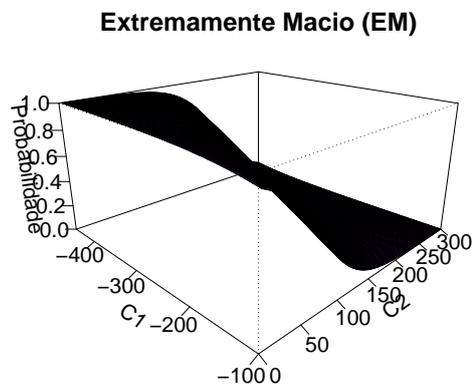


Figura 5.24 – Probabilidade do arroz receber avaliação sensorial como Extremamente Macio (EM) considerando diferentes valores das variáveis C1 e C2 de arroz de arroz de Terras Altas para o ano de 2014

Pegajosidade

Tabela 5.16 – Classificação da avaliação sensorial de pegajosidade para o ano de 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando componentes principais de arroz de arroz de arroz de Terras Altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	5	7	0	0
	P	3	8	8	0
	LS	0	3	33	0
	S	0	0	7	1

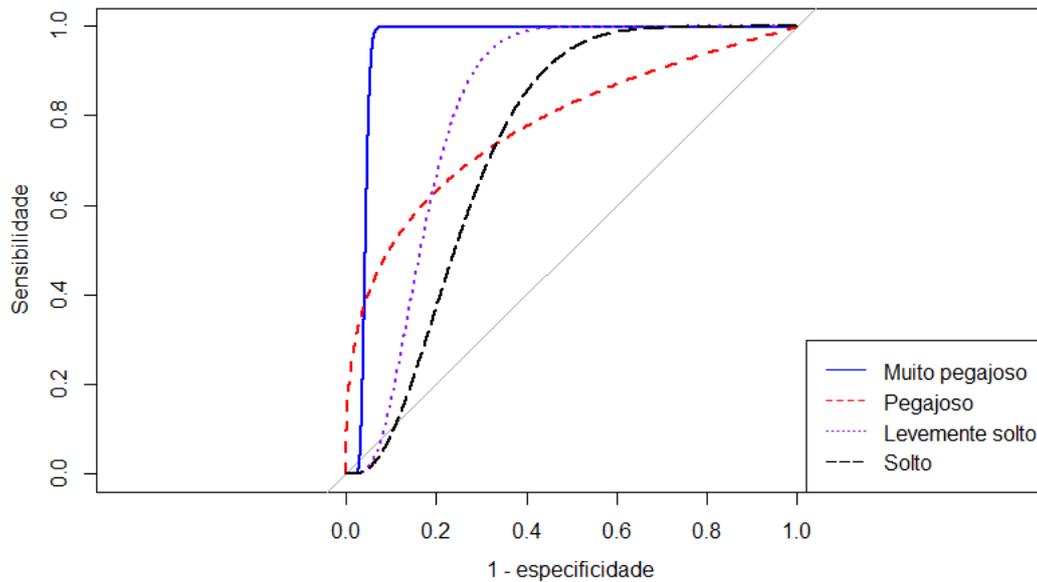


Figura 5.25 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2014, prevista por meio do modelo de regressão logística utilizando componentes principais de arroz de arroz de arroz de Terras Altas

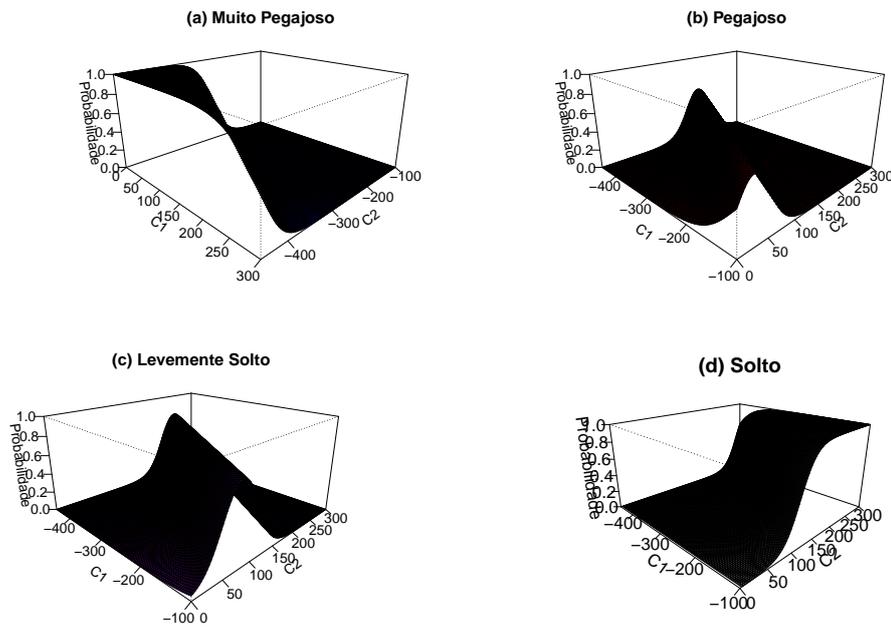


Figura 5.26 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores das variáveis C_1 e C_2 de arroz de arroz de Terras Altas para o ano de 2014

Pegajosidade binária

Tabela 5.17 – Classificação da avaliação sensorial de pegajosidade para o ano de 2014 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando componentes principais de arroz de arroz de Terras Altas

	Classificação prevista	
	P*	S*
Classificação real	P*	5
	S*	40

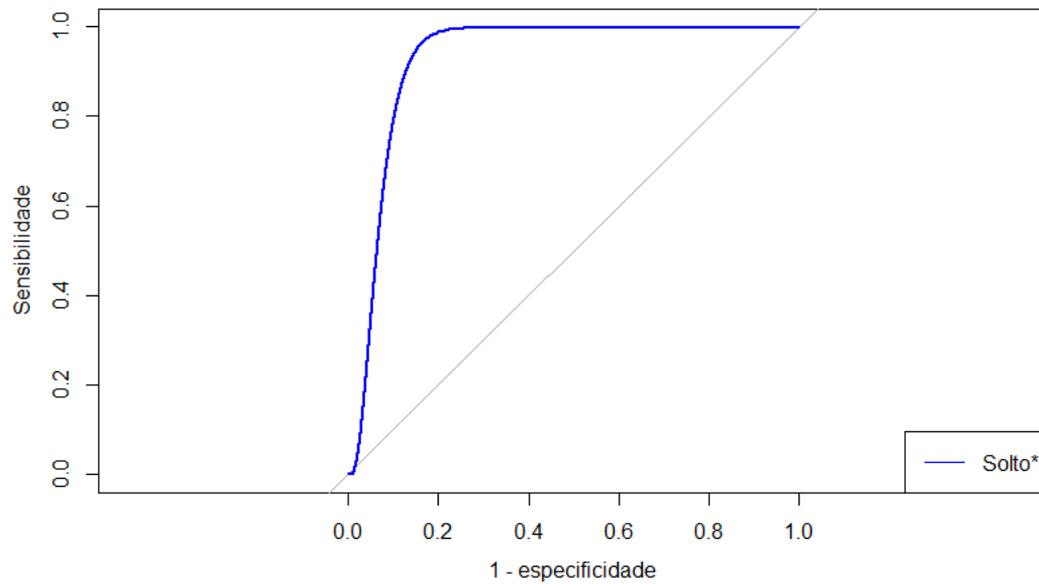


Figura 5.27 – Curva de classificação ROC da avaliação sensorial de pegajosidade para o ano de 2014, prevista por meio do modelo de regressão logística binário utilizando componentes principais de arroz de arroz de Terras Altas

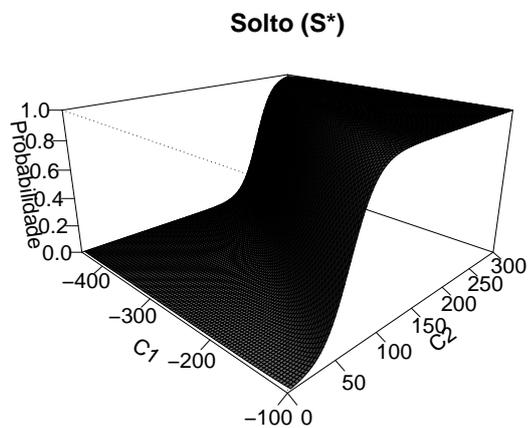


Figura 5.28 – Probabilidade do arroz receber avaliação sensorial como Solto (S^*) considerando diferentes valores das variáveis $C1$ e $C2$ de arroz de arroz de Terras Altas para o ano de 2014

5.2.3 Avaliação sensorial por medidas instrumentais em de arroz de Terras Altas

Dureza

Tabela 5.18 – Classificação sensorial de dureza de arroz de Terras Altas para o ano de 2014

Dureza	Frequência	Porcentagem
Extremamente firme	0	0%
Muito firme	0	0%
Levemente firme	0	0%
Macio com centro firme	6	8,33%
Ligeiramente macio	0	0%
Macio	59	81,95%
Extremamente macio	7	9,72%

5.3 Tabelas e Figuras da Regressão Logística para o ano de 2013 unido ao de 2014

5.3.1 Avaliação sensorial por medidas instrumentais em Terrenos Irrigados

Dureza

Tabela 5.19 – Classificação sensorial de dureza de arroz de arroz de arroz de Terras Altas para os anos 2013 e 2014

Dureza	Frequência	Porcentagem
Extremamente firme	0	0%
Muito firme	0	0%
Levemente firme	0	0%
Macio com centro firme	3	1,56%
Ligeiramente macio	0	0%
Macio	154	80,21%
Extremamente macio	35	18,23%

Tabela 5.20 – Classificação da avaliação sensorial de dureza para os anos 2013 e 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando a dureza instrumental de arroz de arroz de arroz de Terras Altas

	Classificação prevista			
	LML	M	EM	
Classificação real	LML	0	2	1
	M	0	148	6
	EM	0	25	10

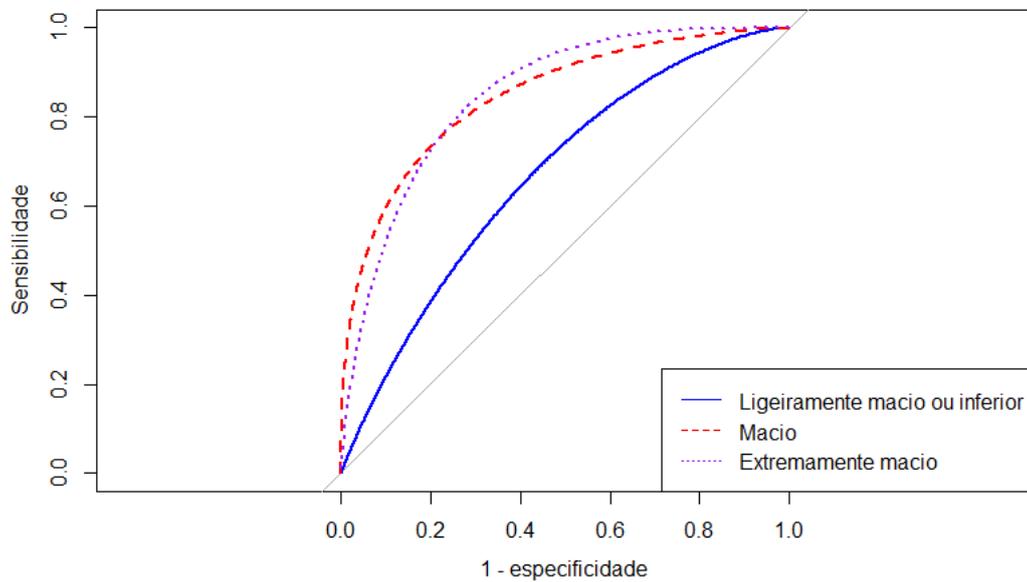


Figura 5.29 – Curva de classificação ROC da avaliação sensorial de dureza para os anos 2013 e 2014, prevista por meio do modelo de regressão logística utilizando a dureza instrumental de arroz de arroz de arroz de Terras Altas

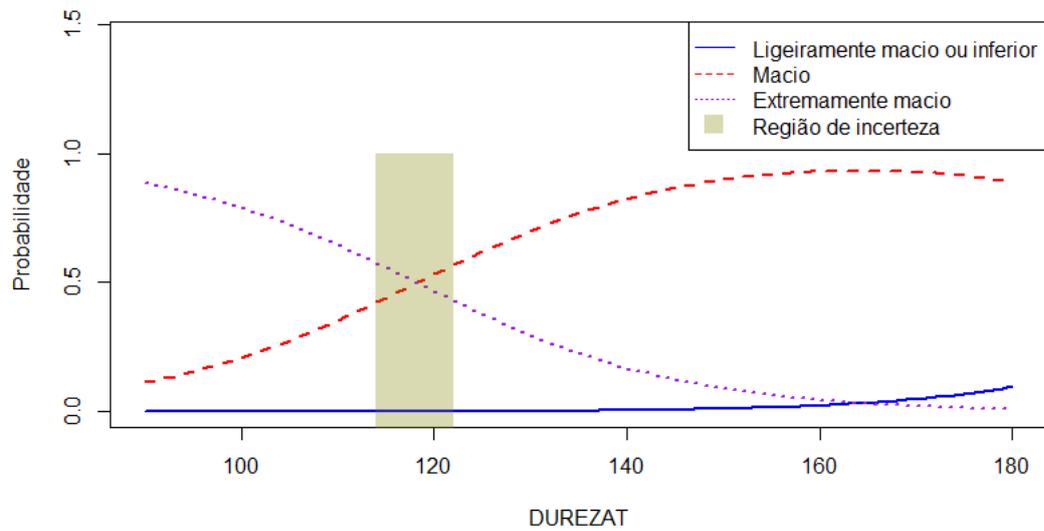


Figura 5.30 – Probabilidades das categorias de avaliação sensorial de dureza considerando diferentes valores da dureza instrumental de arroz de arroz de Terras Altas para os anos 2013 e 2014

Pegajosidade

Tabela 5.21 – Classificação sensorial de pegajosidade arroz de arroz de arroz de Terras Altas para os anos 2013 e 2014

Pegajosidade	Frequência	Pegajosidade
Extremamente pegajoso	11	5,73%
Muito pegajoso	32	16,67%
Pegajoso	64	33,33%
Levemente solto	71	36,98%
Solto	14	7,29%
Muito solto	0	0%
Extremamente solto	0	0%

Tabela 5.22 – Classificação da avaliação sensorial de pegajosidade para os anos 2013 e 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

		Classificação prevista				
		EP	MP	P	LS	S
Classificação real	EP	0	11	0	0	0
	MP	4	12	12	4	0
	P	0	4	28	32	0
	LS	0	0	10	61	0
	S	0	0	0	14	0

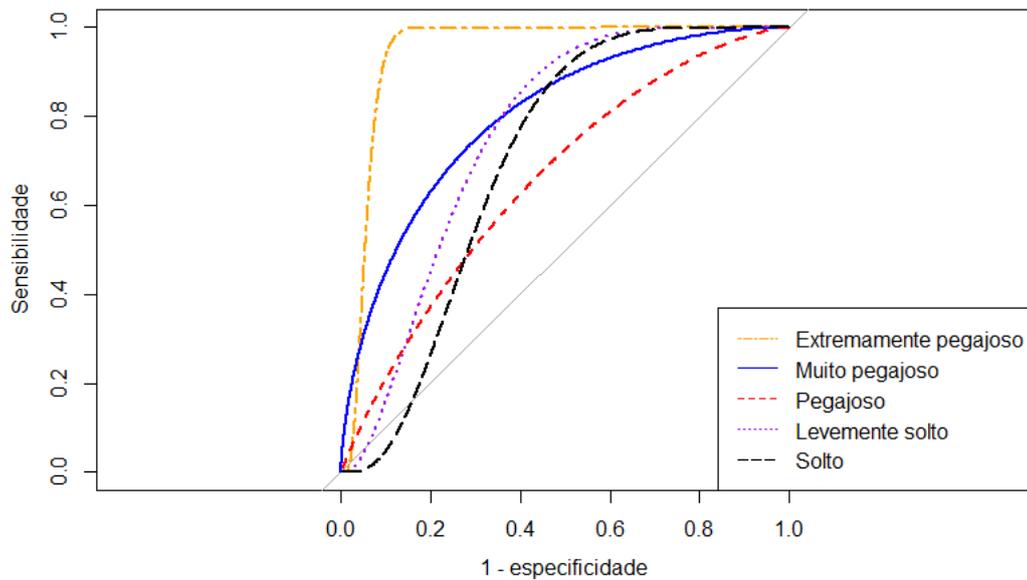


Figura 5.31 – Curva de classificação ROC da avaliação sensorial de pegajosidade para os anos de 2013 e 2014, prevista por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

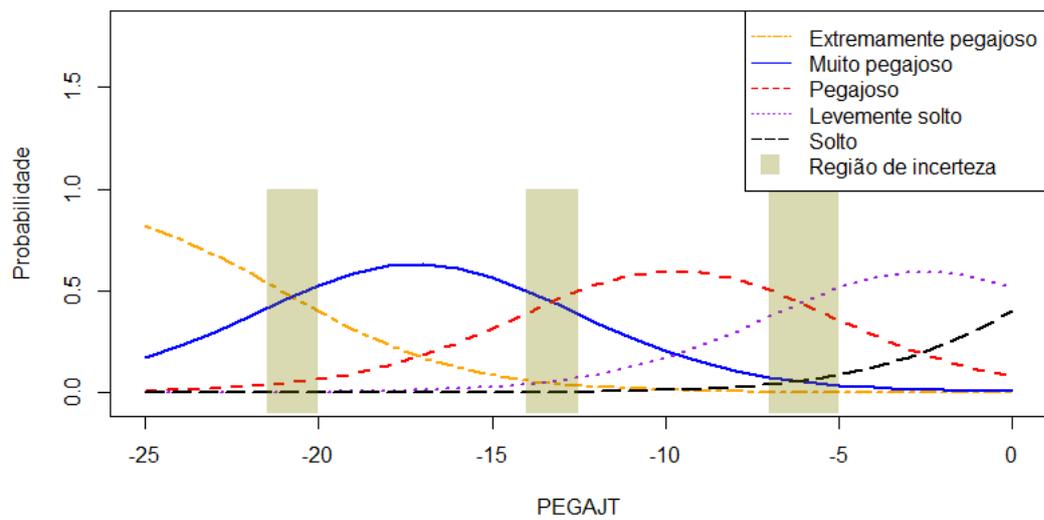


Figura 5.32 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores da pegajosidade instrumental de arroz de arroz de Terras Altas para os anos 2013 e 2014

Pegajosidade binária

Tabela 5.23 – Classificação da avaliação sensorial de pegajosidade para os anos 2013 e 2014 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

	Classificação prevista		
	P*	S*	
Classificação real	P*	71	36
	S*	10	75

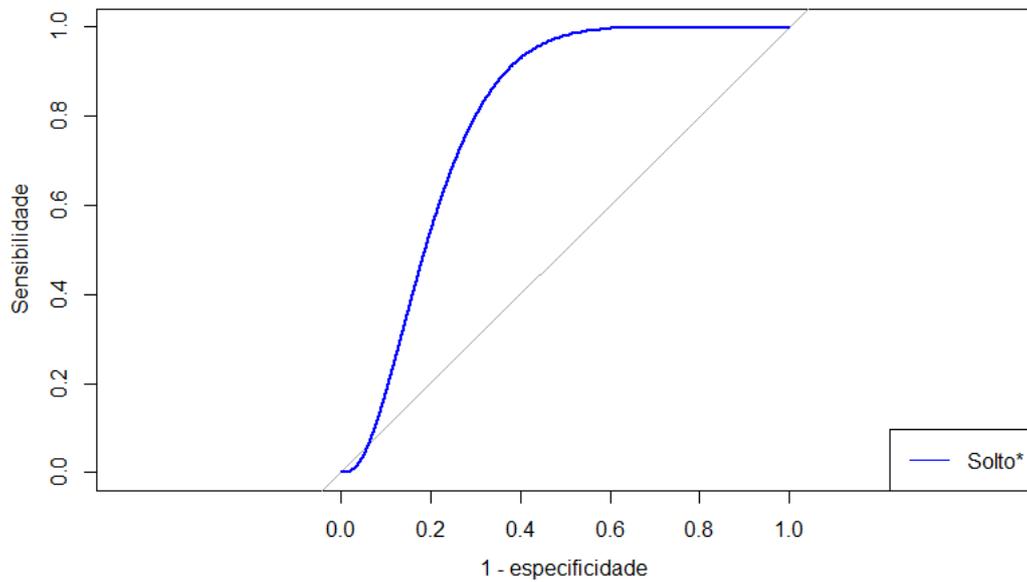


Figura 5.33 – Curva de classificação ROC da avaliação sensorial de pegajosidade para os anos 2013 e 2014, prevista por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

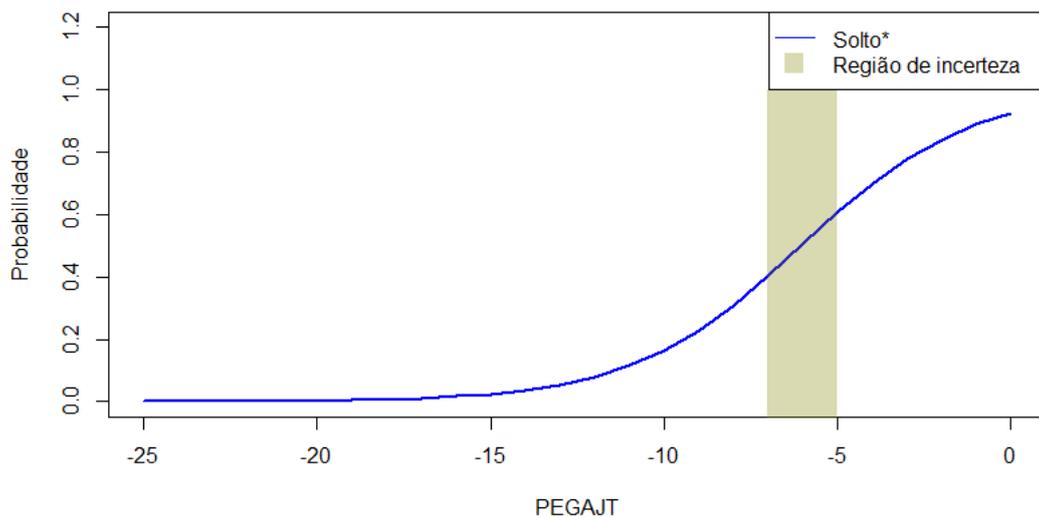


Figura 5.34 – Probabilidade do arroz receber avaliação sensorial como Solto (S^*) considerando diferentes valores da pegajosidade instrumental de arroz de arroz de arroz de Terras Altas para os anos 2013 e 2014

5.3.2 Avaliação sensorial por medidas do perfil viscoamilográfico em Terrenos Irrigados

Tabela 5.24 – Contribuição de cada variável nas duas primeiras componentes principais para Terrenos Irrigados e coeficiente de correlação entre as variáveis dos dados e as componentes principais selecionadas

	Componentes principais	
	Coeficiente de correlação	
	Primeira componente	Segunda componente
TAAFIA	-0,04510	0,67071
	-0,0719	0,8978
TG	0,45217	0,35459
	0,721	0,475
PEAK	-0,59783	0,12831
	-0,953	0,172
BREAKDOWN	-0,53934	-0,29960
	-0,860	-0,401
FINAL	-0,38108	0,56408
	-0,608	0,755

Dureza

Tabela 5.25 – Classificação sensorial de dureza de arroz de Terras Altas para os anos 2013 e 2014

Dureza	Frequência	Porcentagem
Extremamente firme	0	0%
Muito firme	0	0%
Levemente firme	1	0,69%
Macio com centro firme	17	11,81%
Ligeiramente macio	5	3,47%
Macio	106	73,61%
Extremamente macio	15	10,42%

Tabela 5.26 – Classificação da avaliação sensorial de dureza para os anos 2013 e 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando componentes principais de arroz de arroz de arroz de Terras Altas

		Classificação prevista		
		MLM	M	EM
Classificação real	MLM	0	3	0
	M	0	149	5
	EM	0	27	8

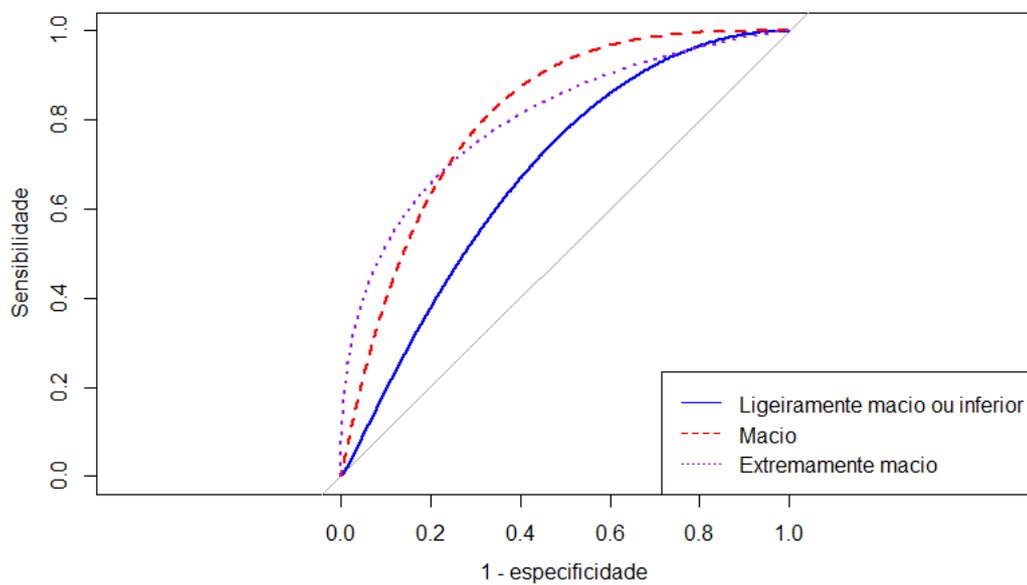


Figura 5.35 – Curva de classificação ROC da avaliação sensorial de dureza para os anos 2013 e 2014, prevista por meio do modelo de regressão logística utilizando componentes principais de arroz de arroz de arroz de Terras Altas

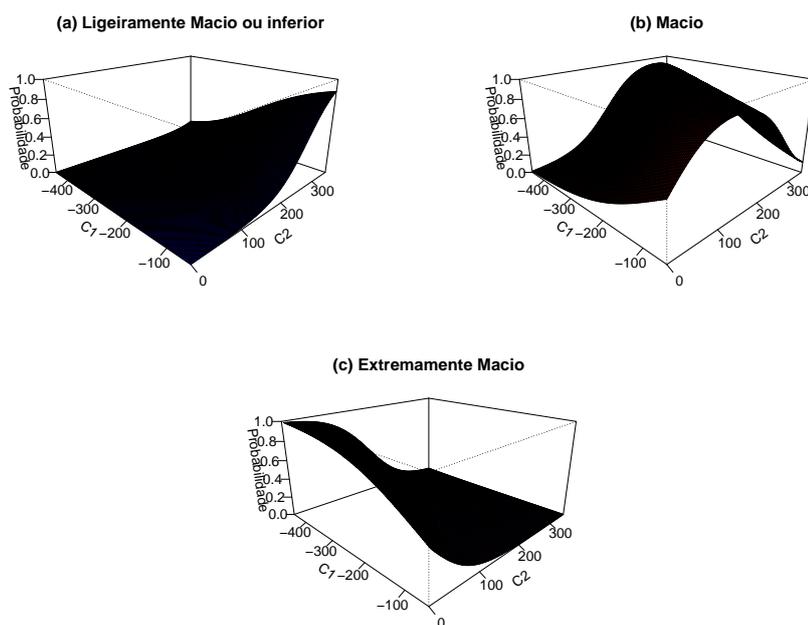


Figura 5.36 – Probabilidades das categorias de avaliação sensorial de dureza considerando diferentes valores das variáveis C_1 e C_2 de arroz de arroz de Terras Altas para os anos 2013 e 2014

Pegajosidade

Tabela 5.27 – Classificação sensorial de pegajosidade arroz de Terras Altas para os anos 2013 e 2014

Pegajosidade	Frequência	Pegajosidade
Extremamente pegajoso	11	7,64%
Muito pegajoso	14	9,72%
Pegajoso	39	27,08%
Levemente solto	71	49,31%
Solto	9	6,25%
Muito solto	0	0%
Extremamente solto	0	0%

Tabela 5.28 – Classificação da avaliação sensorial de pegajosidade para os anos 2013 e 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando componentes principais de arroz de arroz de arroz de Terras Altas

		Classificação prevista				
		EP	MP	P	LS	S
Classificação real	EP	0	4	7	0	0
	MP	2	17	9	4	0
	P	0	9	30	25	0
	LS	0	0	15	56	0
	S	0	0	0	14	0

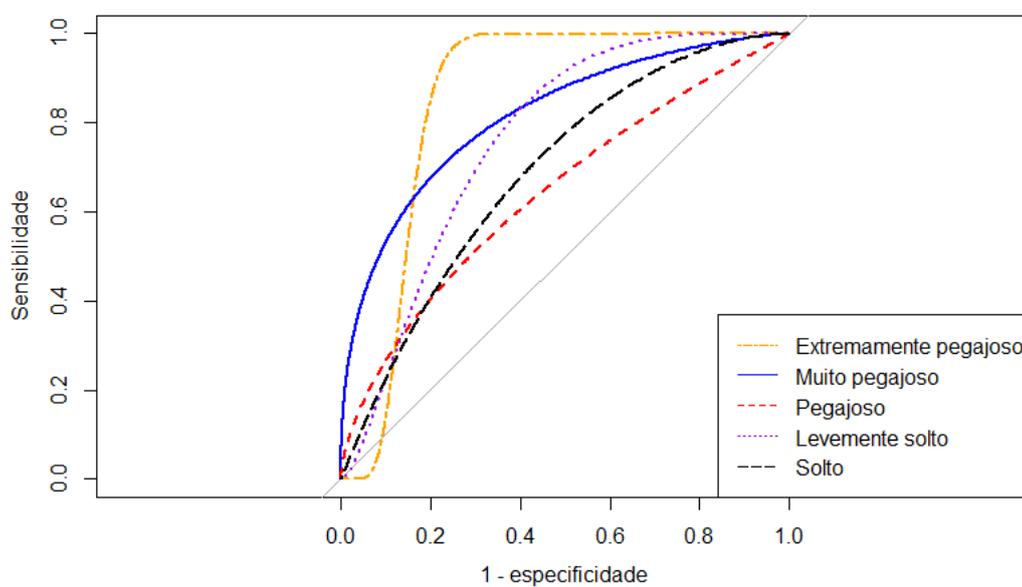


Figura 5.37 – Curva de classificação ROC da avaliação sensorial de pegajosidade para os anos 2013 e 2014, prevista por meio do modelo de regressão logística utilizando componentes principais de arroz de arroz de arroz de Terras Altas

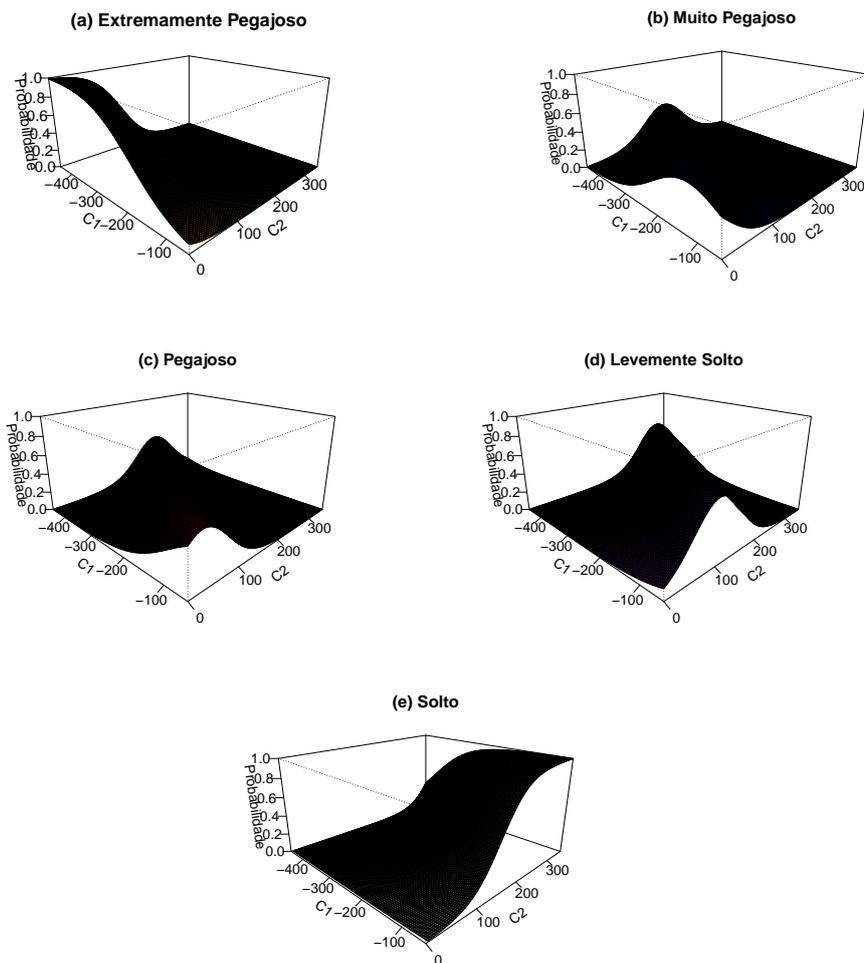


Figura 5.38 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores das variáveis C_1 e C_2 de arroz de arroz de arroz de Terras Altas para os anos 2013 e 2014

Pegajosidade binária

Tabela 5.29 – Classificação da avaliação sensorial de pegajosidade para os anos 2013 e 2014 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando componentes principais de arroz de arroz de arroz de Terras Altas

	Classificação prevista	
	P*	S*
Classificação real	P*	80
	S*	24
		27
		61

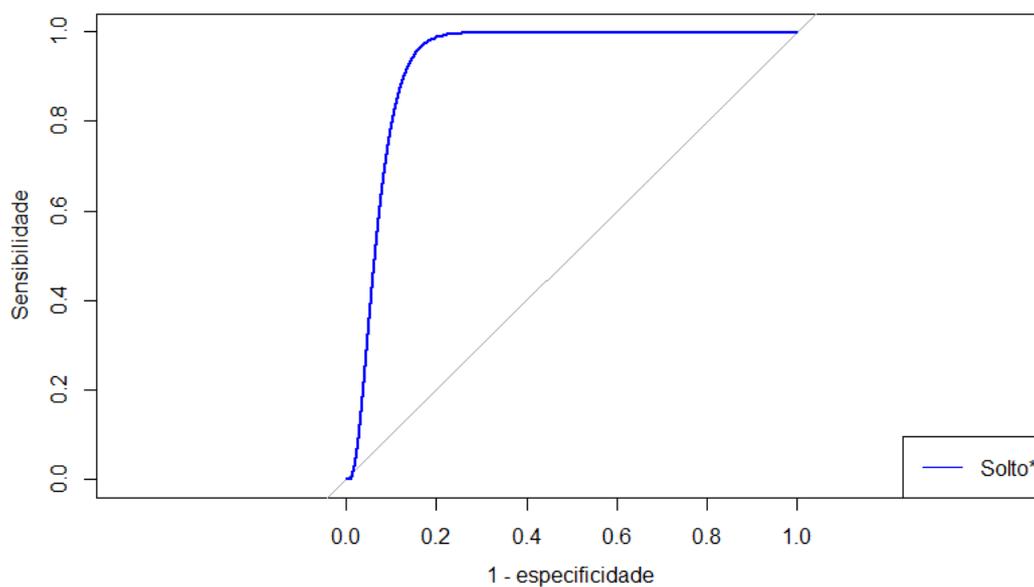


Figura 5.39 – Curva de classificação ROC da avaliação sensorial de pegajosidade para os anos 2013 e 2014, prevista por meio do modelo de regressão logística binário utilizando componentes principais de arroz de arroz de Terras Altas

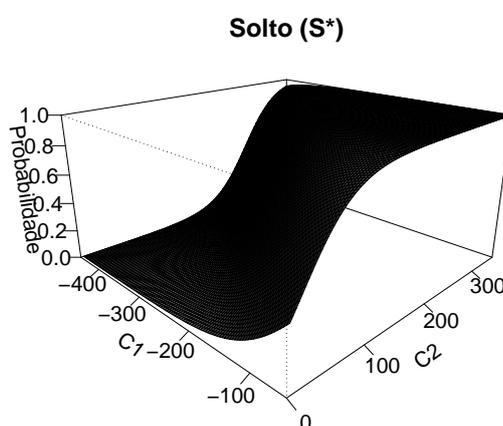


Figura 5.40 – Probabilidade do arroz receber avaliação sensorial como Solto (S^*) considerando diferentes valores das variáveis C_1 e C_2 de arroz de arroz de Terras Altas para os anos 2013 e 2014

5.3.3 Avaliação sensorial por medidas instrumentais em de arroz de Terras Altas

Dureza

Tabela 5.30 – Classificação sensorial de dureza de arroz de Terras Altas para os anos 2013 e 2014

Dureza	Frequência	Porcentagem
Extremamente firme	0	0%
Muito firme	0	0%
Levemente firme	1	0,69%
Macio com centro firme	17	11,81%
Ligeiramente macio	5	3,47%
Macio	106	73,61%
Extremamente macio	15	10,42%

Tabela 5.31 – Classificação da avaliação sensorial de dureza para os anos 2013 e 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando a dureza instrumental de arroz de Terras Altas

		Classificação prevista		
		LML	M	EM
Classificação real	LML	0	23	0
	M	0	106	0
	EM	0	15	0

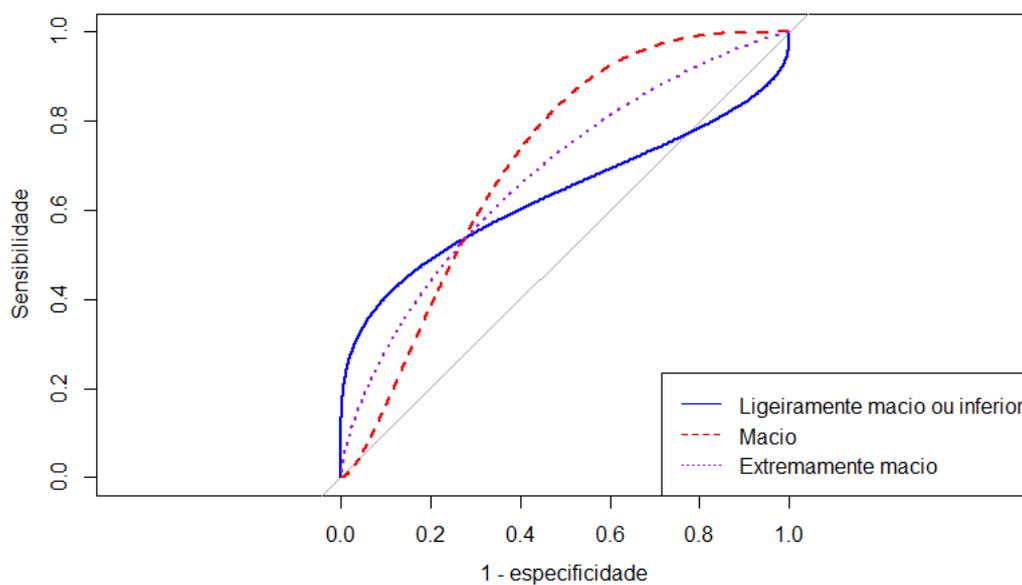


Figura 5.41 – Curva de classificação ROC da avaliação sensorial de dureza para os anos 2013 e 2014, prevista por meio do modelo de regressão logística utilizando a dureza instrumental de arroz de Terras Altas

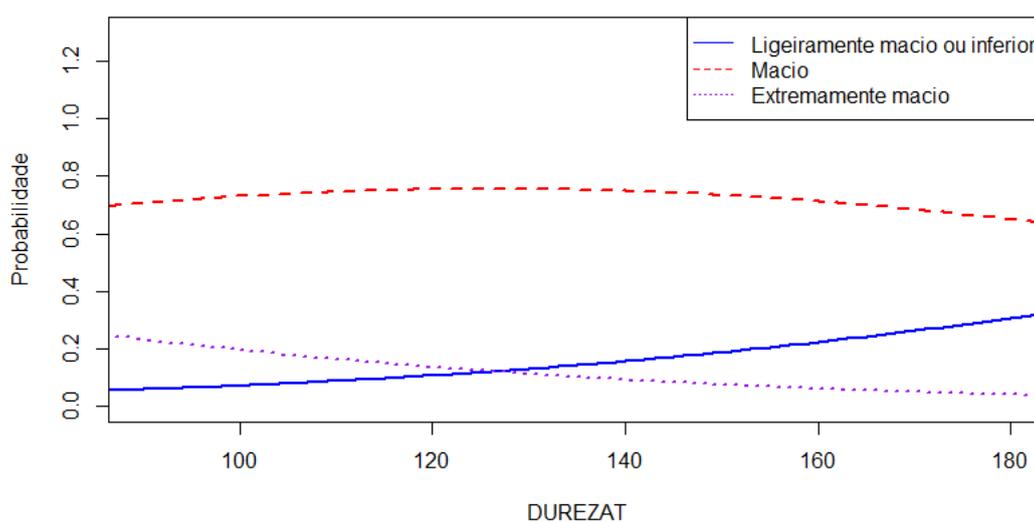


Figura 5.42 – Probabilidades das categorias de avaliação sensorial de dureza considerando diferentes valores da dureza instrumental de arroz de Terras Altas para os anos 2013 e 2014

Pegajosidade

Tabela 5.32 – Classificação sensorial de pegajosidade arroz de Terras Altas para os anos 2013 e 2014

Pegajosidade	Frequência	Pegajosidade
Extremamente pegajoso	11	7,64%
Muito pegajoso	14	9,72%
Pegajoso	39	27,08%
Levemente solto	71	49,31%
Solto	9	6,25%
Muito solto	0	0%
Extremamente solto	0	0%

Tabela 5.33 – Classificação da avaliação sensorial de pegajosidade para os anos 2013 e 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de Terras Altas

		Classificação prevista				
		EP	MP	P	LS	S
Classificação real	EP	4	0	3	4	0
	MP	3	0	7	4	0
	P	1	0	26	12	0
	LS	0	0	14	57	0
	S	0	0	1	8	0

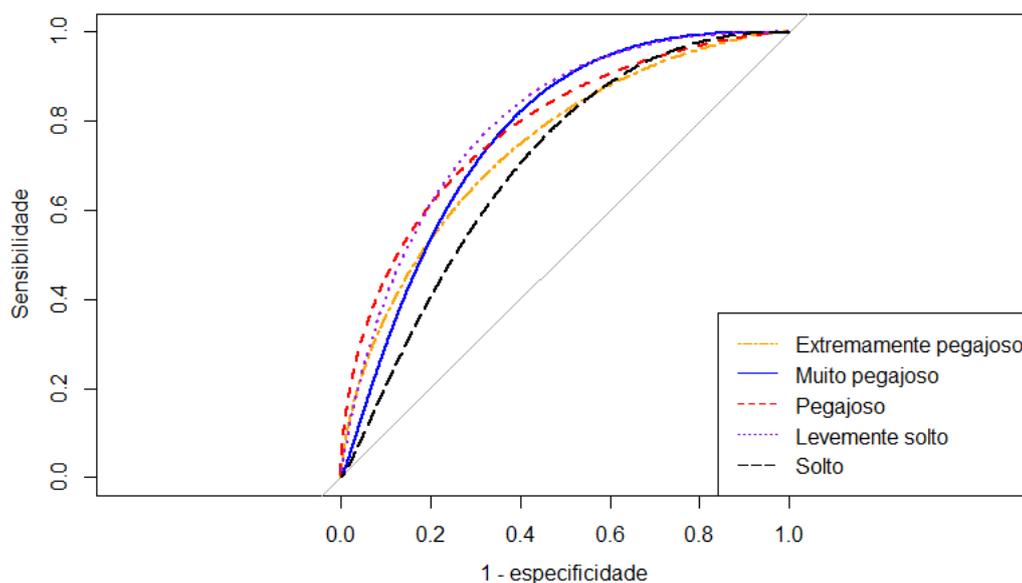


Figura 5.43 – Curva de classificação ROC da avaliação sensorial de pegajosidade para os anos 2013 e 2014, prevista por meio do modelo de regressão logística utilizando a pegajosidade instrumental de arroz de Terras Altas

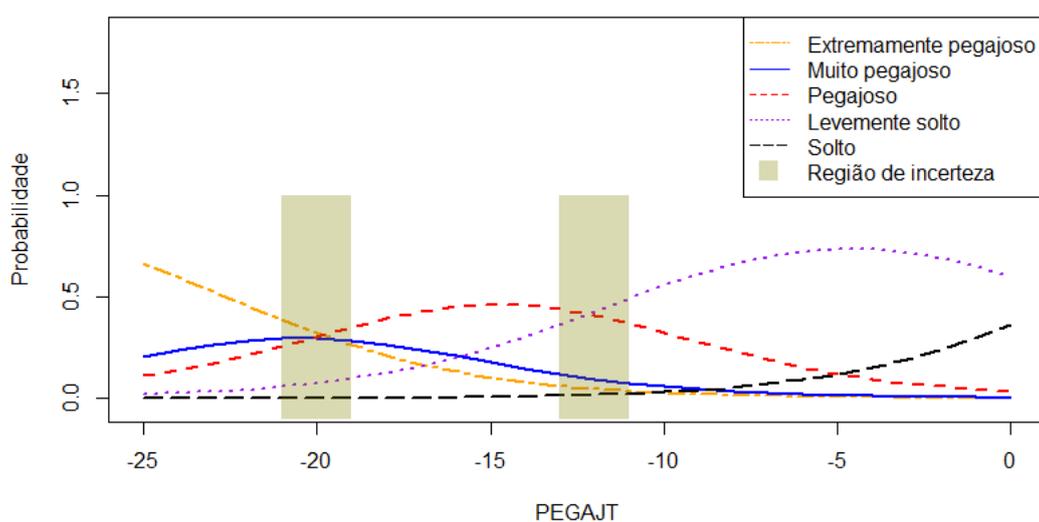


Figura 5.44 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores da pegajosidade instrumental de arroz de Terras Altas para os anos 2013 e 2014

Pegajosidade binária

Tabela 5.34 – Classificação da avaliação sensorial de pegajosidade para os anos 2013 e 2014 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de Terras Altas

	Classificação prevista	
	P*	S*
Classificação real	P*	52
	S*	15

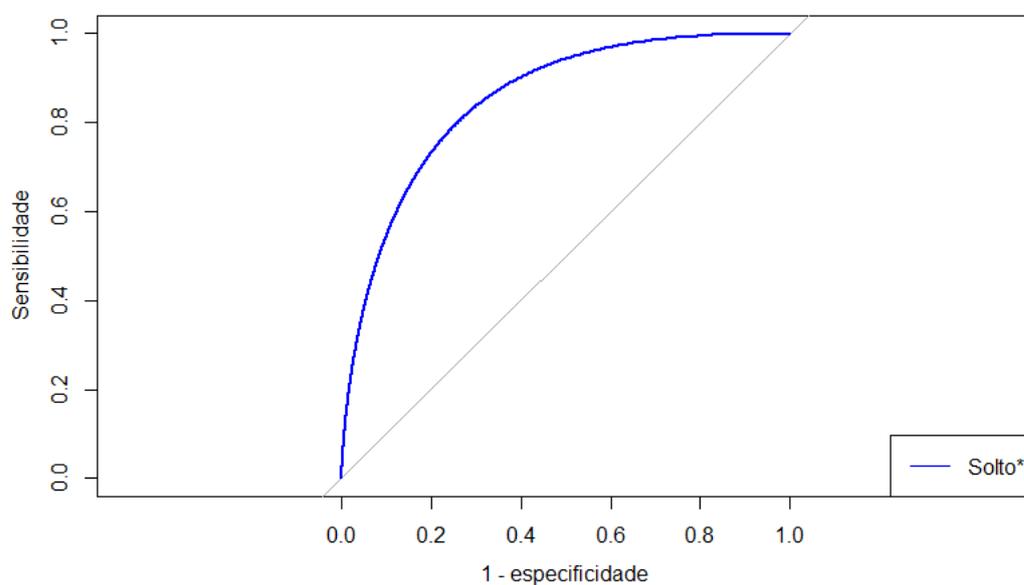


Figura 5.45 – Curva de classificação ROC da avaliação sensorial de pegajosidade para os anos 2013 e 2014, prevista por meio do modelo de regressão logística binário utilizando a pegajosidade instrumental de arroz de Terras Altas

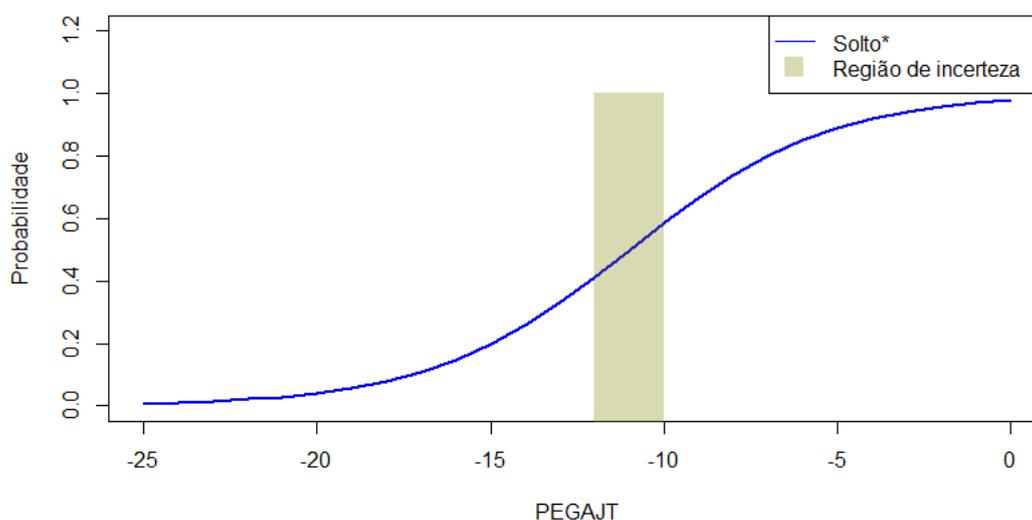


Figura 5.46 – Probabilidade do arroz receber avaliação sensorial como Solto (S*) considerando diferentes valores da pegajosidade instrumental de arroz de Terras Altas para os anos 2013 e 2014

5.3.4 Avaliação sensorial por medidas do perfil viscoamilográfico em de arroz de Terras Altas

Tabela 5.35 – Contribuição de cada variável nas duas primeiras componentes principais para de arroz de Terras Altas e coeficiente de correlação entre as variáveis dos dados e as componentes principais selecionadas

	Componentes principais	
	Coeficiente de correlação	
	Primeira componente	Segunda componente
TAAFIA	-0,56856	0,10708
	-0,897	0,154
TG	-0,51078	-0,31244
	-0,806	-0,449
PEAK	-0,15538	0,66146
	-0,245	0,950
BREAKDOWN	0,28061	0,60475
	0,443	0,868
FINAL	-0,55943	0,29608
	-0,883	0,425

Pegajosidade

Tabela 5.36 – Classificação da avaliação sensorial de pegajosidade para os anos 2013 e 2014 versus a classificação prevista, por meio do modelo de regressão logística utilizando componentes principais de arroz de Terras Altas

		Classificação prevista				
		EP	MP	P	LS	S
Classificação real	EP	5	0	6	0	0
	MP	6	0	7	1	0
	P	1	0	9	29	0
	LS	0	0	12	58	1
	S	0	0	2	7	0

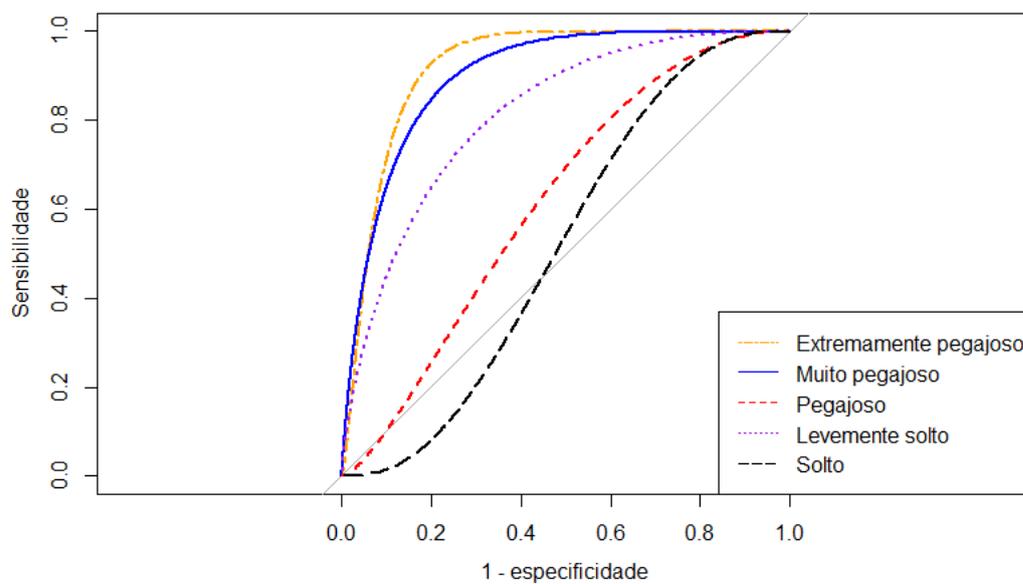


Figura 5.47 – Curva de classificação ROC da avaliação sensorial de pegajosidade para os anos 2013 e 2014, prevista por meio do modelo de regressão logística utilizando componentes principais de arroz de Terras Altas

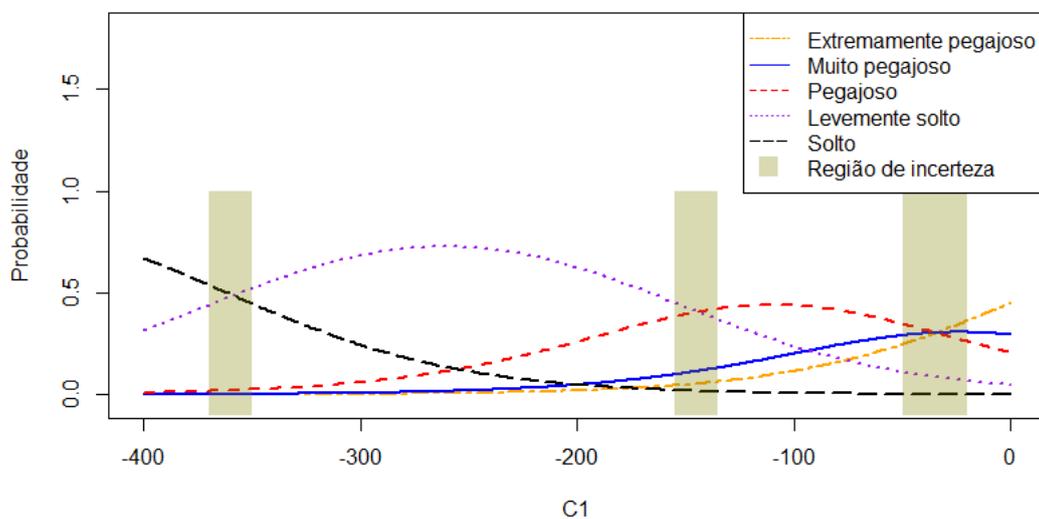


Figura 5.48 – Probabilidades das categorias de avaliação sensorial de pegajosidade considerando diferentes valores da variável *C1* de arroz de Terras Altas para os anos 2013 e 2014

Pegajosidade binária

Tabela 5.37 – Classificação da avaliação sensorial de pegajosidade para os anos 2013 e 2014 versus a classificação prevista, por meio do modelo de regressão logística binário utilizando componentes principais de arroz de Terras Altas

	Classificação prevista	
	P*	S*
Classificação real	P*	54
	S*	15

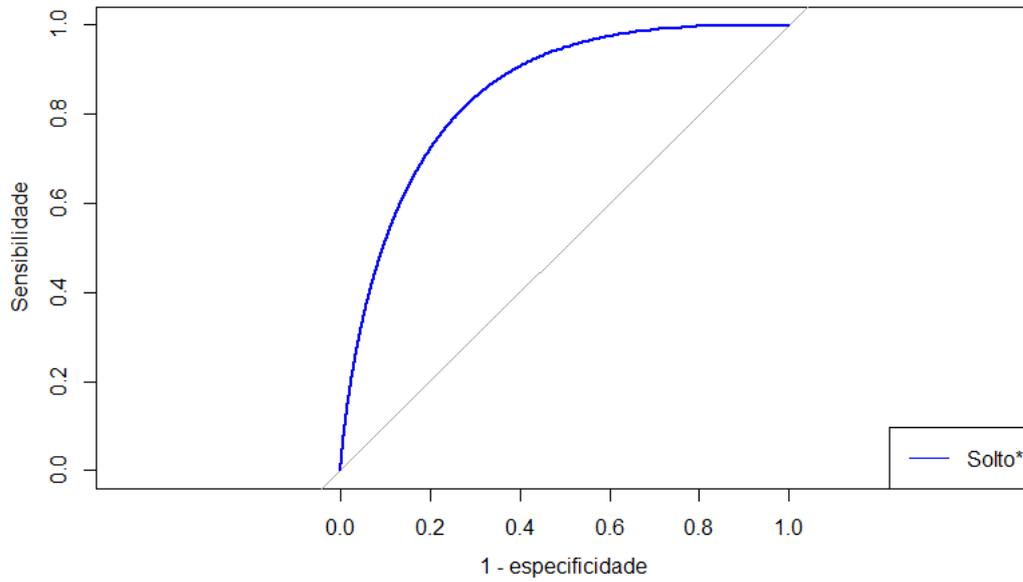


Figura 5.49 – Curva de classificação ROC da avaliação sensorial de pegajosidade para os anos de 2013 e 2014, prevista por meio do modelo de regressão logística binário utilizando componentes principais de arroz de Terras Altas

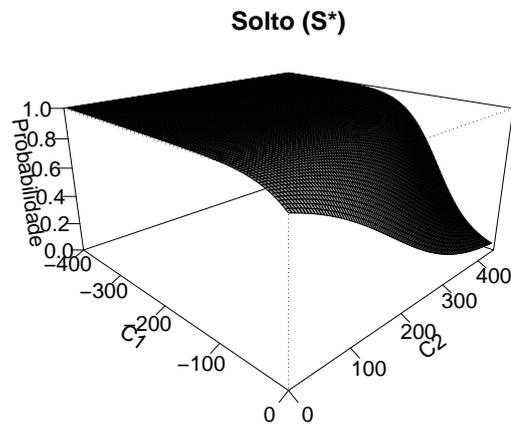


Figura 5.50 – Probabilidade do arroz receber avaliação sensorial como Solto (S^*) considerando diferentes valores das variáveis C_1 e C_2 de arroz de Terras Altas para os anos 2013 e 2014

5.4 Tabelas e Figuras da previsão para o ano de 2014 utilizando os modelos de Regressão Logística obtidos por meio das observações de 2013

5.4.1 Avaliação sensorial por medidas instrumentais em Terrenos Irrigados

Dureza

Tabela 5.38 – Classificação da avaliação sensorial de dureza para as observações de 2014 versus a classificação prevista, por meio dos modelos de regressão logística de 2013 utilizando a dureza instrumental de arroz de arroz de arroz de Terras Altas

		Classificação prevista		
		MLM	M	EM
Classificação real	MLM	0	0	0
	M	0	62	0
	EM	0	13	0

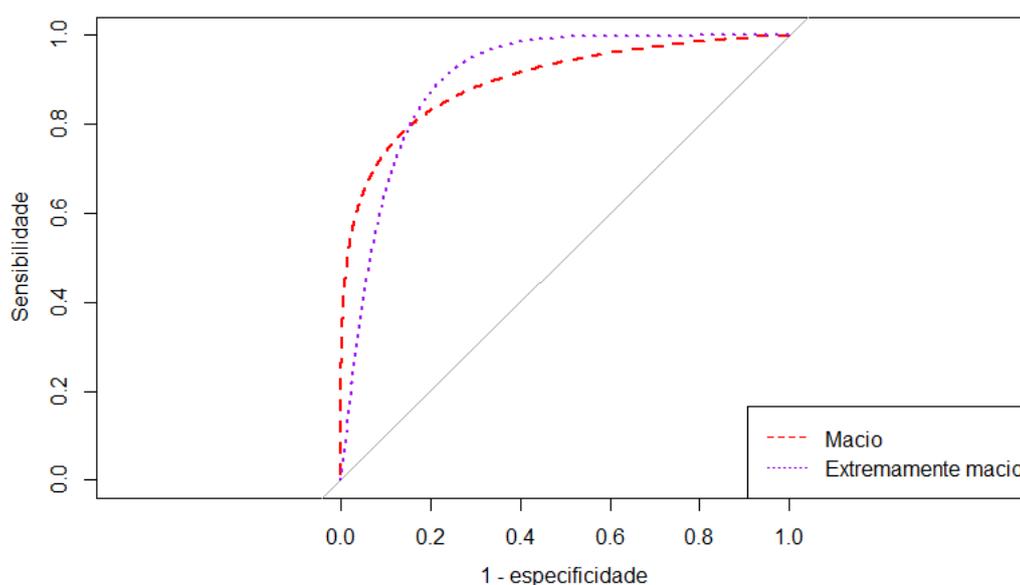


Figura 5.51 – Curva de classificação ROC da dureza sensorial para as observações de 2014 previstas, por meio dos modelos de regressão logística de 2013 utilizando a dureza instrumental de arroz de arroz de arroz de Terras Altas

Pegajosidade

Tabela 5.39 – Classificação da avaliação sensorial de pegajosidade para as observações de 2014 versus a classificação prevista, por meio do modelo de regressão logística de 2013 utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	11	1	0	0
	P	0	17	2	0
	LS	0	11	25	0
	S	0	3	5	0

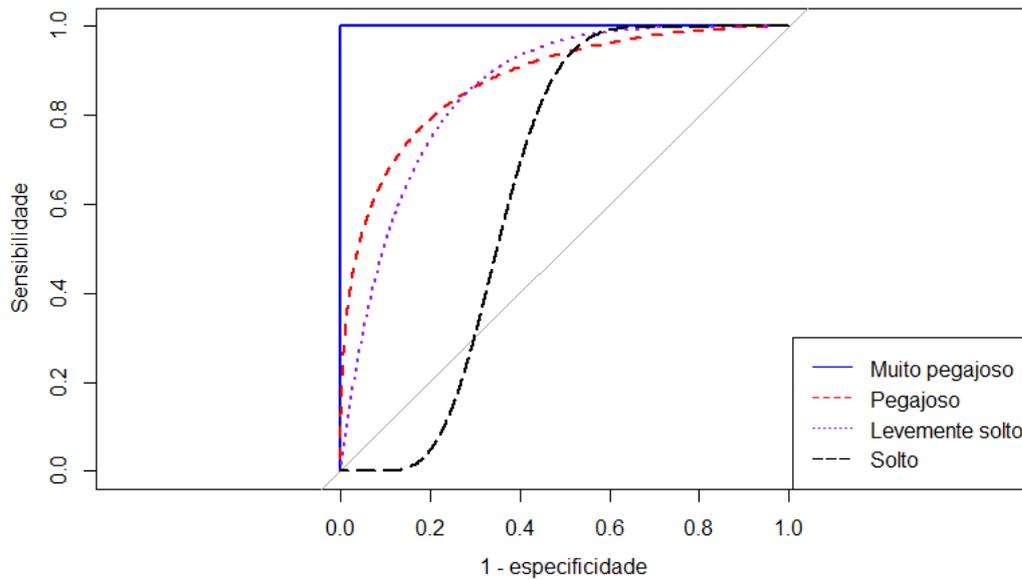


Figura 5.52 – Curva de classificação ROC da avaliação sensorial de pegajosidade para as observações de 2014, previstas por meio do modelo de regressão logística de 2013 utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

Pegajosidade binária

Tabela 5.40 – Classificação da avaliação sensorial de pegajosidade para as observações de 2014 versus a classificação prevista, por meio do modelo de regressão logística binário de 2013 utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

	Classificação prevista	
	P*	S*
Classificação real	P*	29
	S*	15

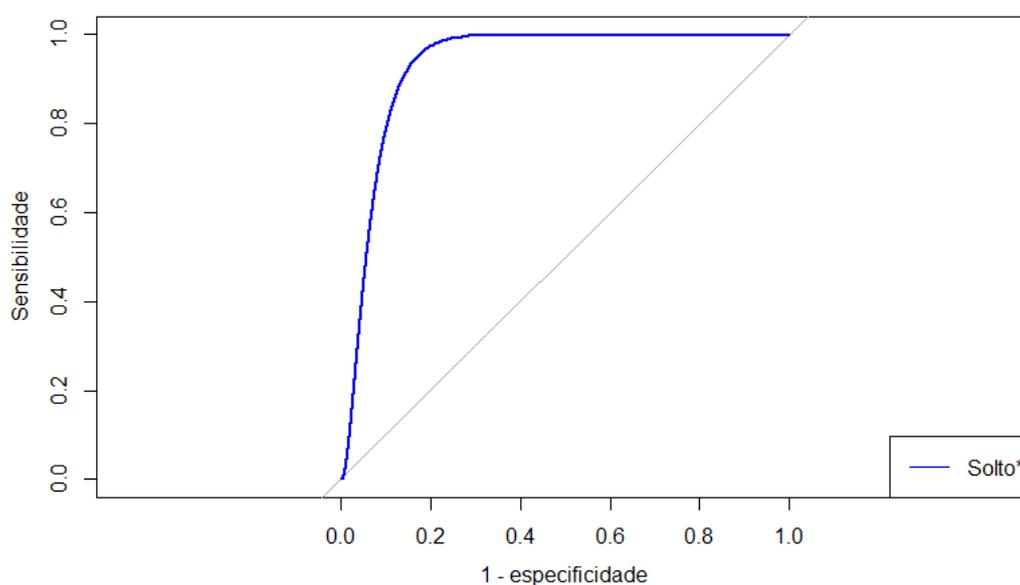


Figura 5.53 – Curva de classificação ROC da avaliação sensorial de pegajosidade para as observações de 2014, previstas por meio do modelo de regressão logística binário de 2013 utilizando a pegajosidade instrumental de arroz de arroz de arroz de Terras Altas

5.4.2 Avaliação sensorial por medidas do perfil viscoamilográfico em Terrenos Irrigados

Dureza

Tabela 5.41 – Classificação da avaliação sensorial de dureza para as observações de 2014 versus a classificação prevista, por meio do modelo de regressão logística de 2013 utilizando componentes principais de arroz de arroz de arroz de Terras Altas

		Classificação prevista		
		MLM	M	EM
Classificação real	MLM	0	0	0
	M	0	62	0
	EM	0	12	1

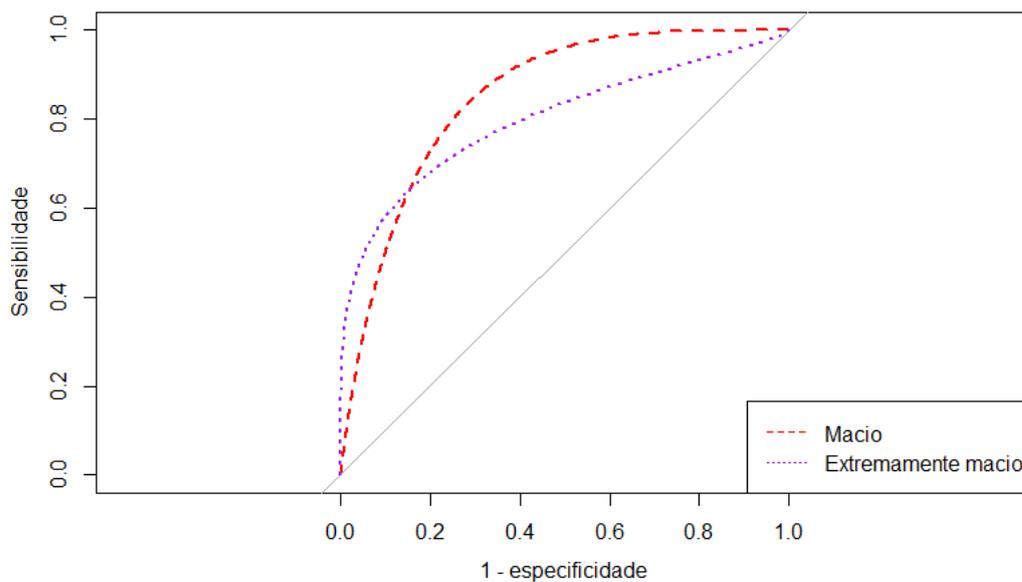


Figura 5.54 – Curva de classificação ROC da avaliação sensorial de dureza para as observações de 2014, previstas por meio do modelo de regressão logística de 2013 utilizando componentes principais de arroz de arroz de arroz de Terras Altas

Pegajosidade

Tabela 5.42 – Classificação da avaliação sensorial de pegajosidade para as observações de 2014 versus a classificação prevista, por meio do modelo de regressão logística de 2013 utilizando componentes principais de arroz de arroz de arroz de Terras Altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	9	3	0	0
	P	2	14	3	0
	LS	0	19	17	0
	S	0	3	5	0

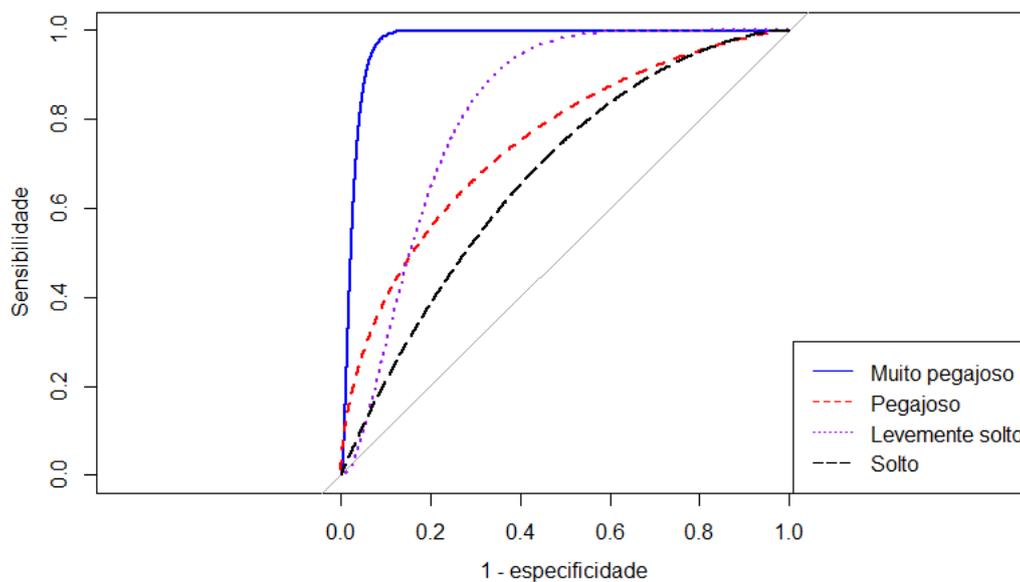


Figura 5.55 – Curva de classificação ROC da avaliação sensorial de pegajosidade para as observações de 2014, previstas por meio do modelo de regressão logística de 2013 utilizando componentes principais de arroz de arroz de arroz de Terras Altas

Pegajosidade binária

Tabela 5.43 – Classificação da avaliação sensorial de pegajosidade para as observações de 2014 versus a classificação prevista, por meio do modelo de regressão logística binário de 2013 utilizando componentes principais de arroz de arroz de arroz de Terras Altas

	Classificação prevista	
	P*	S*
Classificação real	P*	29
	S*	19

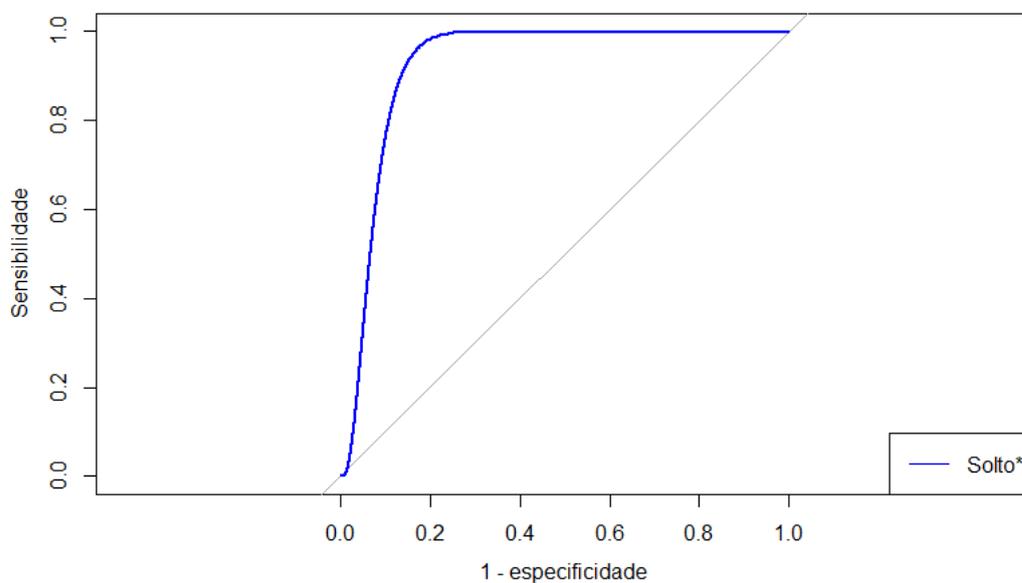


Figura 5.56 – Curva de classificação ROC da avaliação sensorial de pegajosidade para as observações de 2014, previstas por meio do modelo de regressão logística binário de 2013 utilizando componentes principais de arroz de arroz de arroz de Terras Altas

5.4.3 Avaliação sensorial por medidas instrumentais em de arroz de Terras Altas

Pegajosidade

Tabela 5.44 – Classificação da avaliação sensorial de pegajosidade para as observações de 2014 versus a classificação prevista, por meio do modelo de regressão logística de 2013 utilizando a pegajosidade instrumental de arroz de Terras Altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	6	0	7	0
	P	0	10	11	0
	LS	0	1	34	0
	S	0	0	3	0

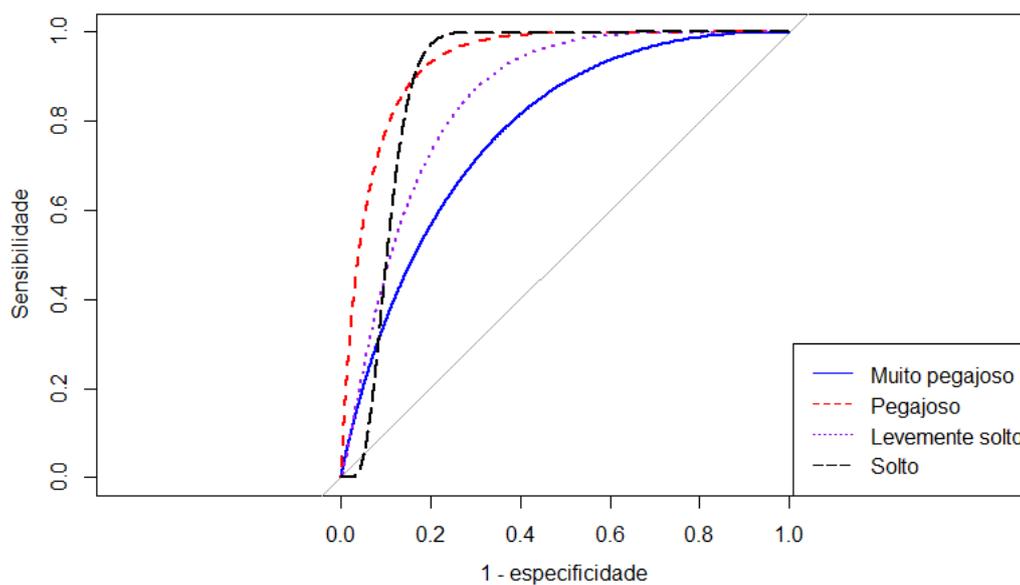


Figura 5.57 – Curva de classificação ROC da avaliação sensorial de pegajosidade para as observações de 2014, previstas por meio do modelo de regressão logística de 2013 utilizando a pegajosidade instrumental de arroz de Terras Altas

Pegajosidade binária

Tabela 5.45 – Classificação da avaliação sensorial de pegajosidade para as observações de 2014 versus a classificação prevista, por meio do modelo de regressão logística binário de 2013 utilizando a pegajosidade instrumental de arroz de Terras Altas

	Classificação prevista		
	P*	S*	
Classificação real	P*	27	7
	S*	3	35

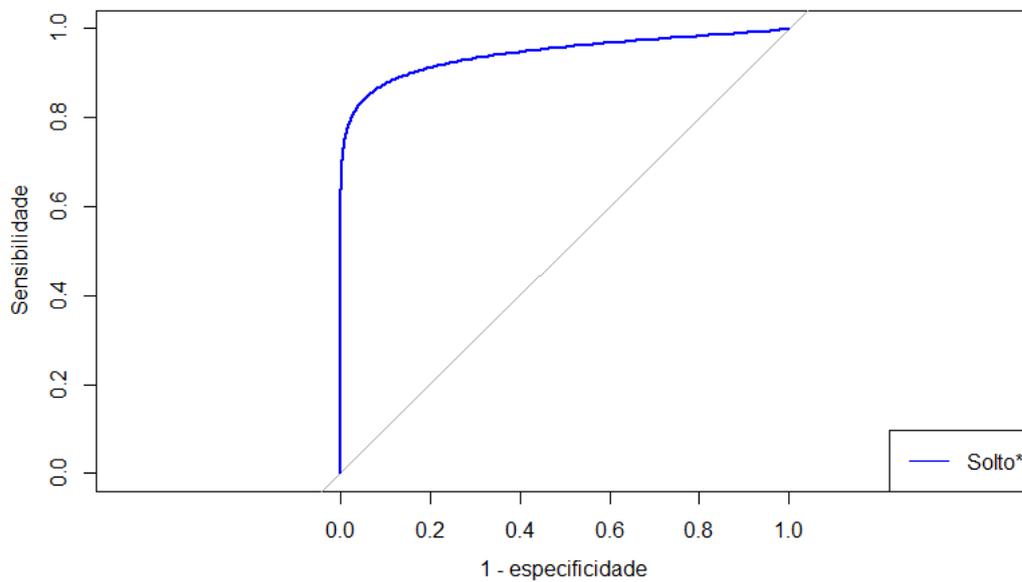


Figura 5.58 – Curva de classificação ROC da avaliação sensorial de pegajosidade para as observações de 2014, previstas por meio do modelo de regressão logística binário de 2013 utilizando a pegajosidade instrumental de arroz de Terras Altas

5.4.4 Avaliação sensorial por medidas do perfil viscoamilográfico em de arroz de Terras Altas

Pegajosidade

Tabela 5.46 – Classificação da avaliação sensorial de pegajosidade para as observações de 2014 versus a classificação prevista, por meio do modelo de regressão logística de 2013 utilizando componentes principais de arroz de Terras Altas

		Classificação prevista			
		MP	P	LS	S
Classificação real	MP	6	5	2	0
	P	0	0	21	0
	LS	0	1	34	0
	S	0	0	3	0

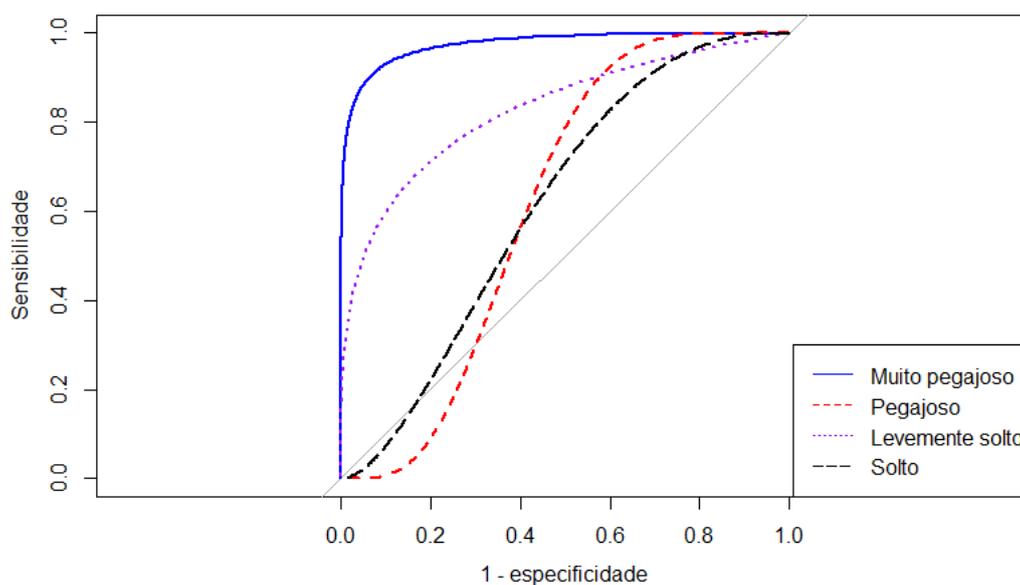


Figura 5.59 – Curva de classificação ROC da avaliação sensorial de pegajosidade para as observações de 2014, previstas por meio do modelo de regressão logística de 2013 utilizando componentes principais de arroz de Terras Altas

Pegajosidade binária

Tabela 5.47 – Classificação da avaliação sensorial de pegajosidade para as observações de 2014 versus a classificação prevista, por meio do modelo de regressão logística binário de 2013 utilizando componentes principais de arroz de Terras Altas

	Classificação prevista	
	P*	S*
Classificação real	P*	20
	S*	34

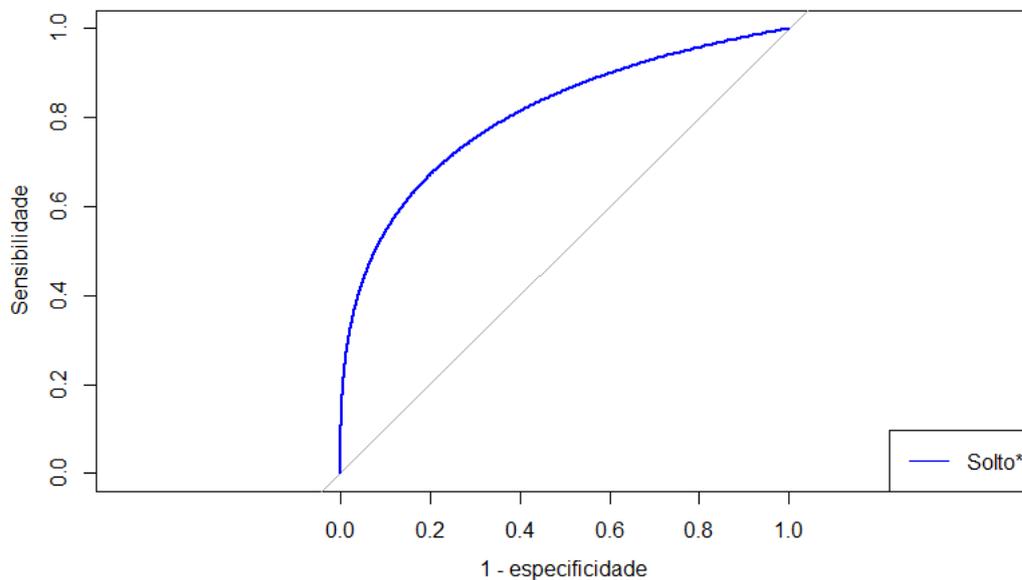


Figura 5.60 – Curva de classificação ROC da avaliação sensorial de pegajosidade para as observações de 2014, previstas por meio do modelo de regressão logística binário de 2013 utilizando componentes principais de arroz de Terras Altas