



VARIABILIDADE DE RESPOSTA AO NEMÁTODE DA MADEIRA DO PINHEIRO, *Bursaphelenchus xylophilus*, DE PLANTAS JOVENS DE PINHEIRO BRAVO, *Pinus pinaster*, PROVENIENTES DE UMA POPULAÇÃO DE MELHORAMENTO

Bruno Emanuel Duarte Ribeiro

Dissertação para obtenção do Grau de Mestre em
Gestão e Conservação de Recursos Naturais

Orientador: Doutora Maria Isabel Carrasquinho de Freitas

Co-orientadores: Professora Doutora Ana Maria Amorim Sampaio da Silva

Professora Doutora Manuela Rodrigues Branco Simões

Júri:

Presidente: - Doutora Maria Teresa Marques Ferreira da Cunha Cardoso, Professora Associada do Instituto Superior de Agronomia da Universidade Técnica de Lisboa

Vogais: - Doutora Maria Helena Reis de Noronha Ribeiro de Almeida, Professora Associada do Instituto Superior de Agronomia da Universidade Técnica de Lisboa;

- Doutora Manuela Rodrigues Branco Simões, Professora Auxiliar do Instituto Superior de Agronomia da Universidade Técnica de Lisboa

- Doutor Arlindo Lima, Professor Auxiliar do Instituto Superior de Agronomia da Universidade Técnica de Lisboa

- Doutora Maria Isabel Carrasquinho de Freitas, Investigadora Auxiliar do Instituto Nacional de Investigação Agrária e Veterinária, I.P.

Lisboa, 2012

Agradecimentos

Este trabalho não teria sido realizado sem a contribuição de algumas pessoas que, de várias formas, o tornaram possível:

À Doutora Isabel Carrasquinho, minha orientadora, pelo apoio a nível científico, pelas horas de revisão e crítica, pela calma demonstrada nos momentos de maior tensão, pelos conhecimentos transmitidos, pela confiança, e principalmente por ter estado presente em todas as fases do trabalho. Também lhe agradeço a nível pessoal pela amizade, pela preocupação, pela simpatia e pelos conselhos. Aprendi muito nestes três anos de trabalho, e a maioria devo-o a ela.

À Professora Ana Sampaio, minha co-orientadora, pelo apoio a nível científico e pessoal, a sua simpatia, conhecimentos de estatística, críticas construtivas e disponibilização de bibliografia, ajudaram a tornar esta tese num trabalho melhor.

À Professora Manuela Branco, minha co-orientadora, pela disponibilidade, pela bibliografia e pela sua capacidade de leccionar. As suas aulas de ecologia forneceram-me conhecimentos muito importantes para a realização desta tese.

À Doutora Filomena Nóbrega, pelos conhecimentos partilhados e pelo constante incentivo, e principalmente por ser a pessoa mais simpática e bondosa que conheço.

Ao Professor Manuel Mota, pela constante colaboração em trabalhos e pela disponibilidade total para esclarecer dúvidas e enviar bibliografia.

Ao Engenheiro Alexandre Aguiar pela bibliografia de melhoramento florestal e pelos conhecimentos transmitidos.

Ao Doutor Edmundo Sousa pela confiança depositada e pela garantia que nada me faltaria durante a realização dos trabalhos.

À Doutora Rita Costa pelas palavras de encorajamento.

À Direcção do Instituto Nacional de Investigação Agrária e Veterinária, por me ter proporcionado a possibilidade de elaboração deste trabalho.

À Direcção da Herdade da Comporta, por permitir a seleção de árvores na sua propriedade e nos dar acesso à mesma.

Aos funcionários do Centro Experimental do Escaroupim pela ajuda no processamento das sementes.

Aos funcionários do Viveiro da Mata Nacional de Valverde em Alcácer do Sal, em especial à Dona Armandina, pela sementeira, por cuidarem das plantas usadas neste ensaio, pelos almoços e pela simpatia e boa disposição.

A todos os colegas do ICAAM em Évora, especialmente à colega Margarida Espada, pela disponibilidade em todos os assuntos relacionados com o nemátode.

Ao pessoal do laboratório de nematologia, principalmente à colega Margarida Fontes, pela disponibilização de inóculo e por me ensinarem os procedimentos necessários à identificação e manipulação do nematode.

À colega Lurdes Inácio pelas revisões ao manuscrito, pela simpatia, pelo constante incentivo, e principalmente por me levantar a moral cada vez que passo por ela.

À colega Joana Godinho, pela partilha de conhecimentos e simpatia.

Ao colega João Pessoa por ter sido mais do que um companheiro de trabalho, e por me ter ajudado a compreender a floresta que não vem nos livros.

À colega Teresa Salema pela simpatia, ajuda nas saídas de campo, e ajuda durante a inoculação e monitorização das plantas.

Ao colega António Saraiva, pela partilha de conhecimentos e boa disposição.

Aos colegas Luis Bonifácio e Pedro Naves, pela ajuda nos assuntos relacionados com o inseto e com o nematode, e pela boa disposição.

Ao colega Paulo Godinho pela simpatia e disponibilidade para ajudar.

À colega Inês Portugal e Castro, pela ajuda durante o ensaio de inoculação e pela simpatia e disponibilidade.

Ao colega Adérito Matos, pela ajuda no campo.

À colega Patricia Ribeiro, pela ajuda nos trabalhos de campo.

À minha colega e amiga Carmen Santos, pelas longas conversas, pelos conselhos, pela amizade, pelos conhecimentos, pela disponibilidade e por ter estado presente em todas as fases do processo, dentro e fora do local de trabalho.

Aos meus colegas de mestrado Daniela Costa, Diana Ratola, Mário Carmo, Lia Godinho e Ana Beatriz Lacerda, que iluminaram os meus dias de mestrado, ajudaram-me nos trabalhos, deram-me ideias, divertiram-me, e principalmente porque nunca tinha visto pessoas tão diferentes entre si formarem um grupo tão perfeito.

A todos os professores da Universidade de Évora e do Instituto Superior de Agronomia, pelos conhecimentos partilhados e disponibilidade para esclarecer dúvidas.

A todos os colegas, ex-colegas e amigos não mencionados anteriormente a certeza que todos tiveram um papel determinante na minha formação como biólogo e como homem.

À minha família, especialmente aos meus pais por tornarem tudo isto possível e à minha irmã pelo apoio incondicional.

Finalmente, à Inês Silva, que tornou este trabalho realidade pela única razão que é o combustível que me faz mover.

RESUMO

A doença da murchidão do pinheiro, causada pelo nemátode da madeira do pinheiro (NMP), *Bursaphelenchus xylophilus*, é uma das mais devastadoras doenças florestais, afetando, em Portugal, Pinheiro bravo (*Pinus pinaster*).

Dado que uma das estratégias de controlo é a plantação de pinheiros resistentes, foi iniciado em Portugal um programa pioneiro de melhoramento para a resistência a esta doença. Paralelamente, e considerando que os esforços para o desenvolvimento de programas de melhoramento genético para a resistência a doenças devem ser focados também em populações melhoradas para outras características, considerou-se um estudo à variabilidade genética de resposta à inoculação com NMP numa população já melhorada para o volume e forma do tronco.

Neste estudo efetuou-se um ensaio de inoculação em estufa, de diferentes concentrações de NMP, em plantas de Pinheiro bravo com dois anos, provenientes da população já melhorada. Os resultados permitiram confirmar a existência de variabilidade na resposta das famílias testadas, assim como os efeitos das diferentes concentrações de inóculo e características biométricas das plantas no resultado final. Permitiu também obter maiores conhecimentos quanto às técnicas e duração de ensaios de inoculação, que irão ser usadas nos testes às descendências do programa de melhoramento para a resistência já iniciado.

Palavras-chave: Pinheiro bravo, *Pinus pinaster*, melhoramento florestal, nemátode da madeira do pinheiro, *Bursaphelenchus xylophilus*, variabilidade de resposta.

ABSTRACT

Pine wilt disease, caused by the pine wood nematode (PWN), *Bursaphelenchus xylophilus*, is one of the most threatening forest diseases. The only known host in Portugal is the maritime pine, *Pinus pinaster*.

One of the strategies to control this disease is the establishment of resistant hosts. Thus, a pioneer breeding programme for resistance to this disease has already been initiated in Portugal. Simultaneously, and considering that the breeding efforts for disease resistance should also be focused in populations improved for other traits, a study of the variability in response to the PWN inoculation in a Portuguese population already improved for volume and stem form, was considered.

Inoculating different PWN concentrations in two year-old maritime pines, from the already improved population, was the methodology used in this study. The results confirmed the existence of variability in response of the tested families, the effects of plant height and stem diameter, and the effects of the different inoculum concentrations applied in this study. The results obtained allowed also increased knowledge of the inoculation techniques. This knowledge will be used to test the progenies from the breeding programme for resistance, already initiated.

Keywords: Maritime pine, *Pinus pinaster*, forest tree breeding, pine wood nematode, *Bursaphelenchus xylophilus*, variability in response.

EXTENDED ABSTRACT

Pine wilt disease (PWD) is caused by the pinewood nematode (PWN), *Bursaphelenchus xylophilus*, and vectored by the insect *Monochamus galloprovincialis*. Introduced in Portugal in 1999, and recently detected in Spain, this pathogen is listed as a quarantine pest in Europe. This disease results from a complex system of three closely interacting elements: a host, a pathogen and a vector. The failure of any of these elements to operate will break the infection cycle and prevent new infections. So far, the only known PWN host in Portugal is maritime pine (*Pinus pinaster*), one of the three main Portuguese forest species, representing 27% of the total forest area.

The establishment of a maritime pine improvement programme for resistance/tolerance to PWD is one of the several strategies to control PWD. In fact, data reported by Japanese breeding programmes for PWD resistance showed that family variation after artificial inoculation was highly significant and family mean heritability exceeded 0.8. According to some authors, pathogen artificial inoculation of 2-year-old seedlings has been the principal method of assessing resistance in forest breeding programmes. External symptoms, assessed after artificial inoculation, can be used as a tolerance indicator.

Taking that into account, a pioneer maritime pine breeding programme for resistance/tolerance to PWD has already been initiated in Portugal with the selection of candidate to plus trees at Herdade da Comporta. Since breeding efforts for resistance/tolerance should also be focused on populations already improved for other traits, evaluation of the genetic effect on the resistance/tolerance to PWD has been considered for an existing maritime pine improvement population, from Mata Nacional do Escaroupim, from the improvement programme for volume and stem form.

From this population, nine families were chosen to make a preliminary study on their variability in response to the PWN. This study was performed in a greenhouse through an inoculation procedure, with controlled temperature and watering. After inoculation, external symptoms were assessed and divided in four levels: 1 – to 25% of the plant symptomatic; 2 – from 26 to 50% of the plant symptomatic; 3 – from 51 to 75% of the plant symptomatic; 4 – more than 76% of the plant symptomatic. The division in symptom levels was expressed as a percentage of visible symptoms in the plant's needles. The plants were monitored weekly until 14 weeks after inoculation.

In order to perform the necessary analysis, the independent variables considered were, CONCENTRAÇÃO (inoculum concentration, ordinal variable) and FAMILIA (plant

family, nominal variable). Due to the fact that the independent variables ALTURA (plant height) and DIÂMETRO (stem diameter) were not controlled during the experimental set up, they were considered as covariates. The dependent variables considered were ESTADO DA PLANTA (mortality, binary, divided in: 0-healthy plant; 1-dead plant) and SINTOMATOLOGIA (symptomatology, ordinal, divided in the four response levels explained before). The experimental set-up used was a split-plot on a completely randomized block design. The main plot considered was CONCENTRAÇÃO, and the sub-plot considered was FAMÍLIA. This split-plot design was divided in four blocks. The statistical programme used was SPSS (v. 19, SPSS Inc., Chicago).

There were several statistical methods used in this study, among them were, descriptive statistics, non-parametric statistics and categorical regression. The non-parametric statistics used were Kolmogorov-Smirnov's test, Levene's test, Chi-square's test, Kruskal-Wallis's test, McNemar's test and Wilcoxon's test. The categorical regression was divided in logistic regression (for the binary dependent variable), multinomial and ordinal regression (for the ordinal dependent variable).

The first result obtained allowed the choice of four observation dates to be analyzed. The dates chosen were: 14, 35, 47 and 92 days after inoculation (DAI). Using descriptive statistics, the results obtained showed that there was variability in response at all observation dates, except 14 DAI. Plants from family 34 were considered the most resistant to the inoculations and on the other hand, family 58 was considered to be the least resistant, either considering ESTADO DA PLANTA and SINTOMATOLOGIA. Using non parametric statistics it was also possible to verify that stem diameter was not evenly distributed per concentration and family, but plant height was evenly distributed per concentration. The final results showed that categorical regression allowed the modelling for each one of the observation dates, except for 14 DAI.

The conclusions obtained with this study are: artificial inoculation of two-year-old maritime pine seedlings with PWN allowed the detection of variability in response in the nine tested families (from these families, family 34 was considered the most resistant, and family 58 was considered the least resistant); dividing the response in four levels gives more accurate results than dividing the response in dead or alive; all inoculum concentrations caused deaths, but the one that caused the greatest mortality and a huge amount of severe symptoms was 500 PWN/ml; symptoms and mortality started to become significantly different only 21 DAI; and different heights and stem diameters

influence the response (since the greater the diameter or height, fewer symptoms and less mortality occurred).

The main conclusion of this study however was that it appears to be an advantage to conjugate the objectives from the improvement programme for the volume and stem form, with a resistance breeding objective. New trials should be developed, and in the future, progenies from the breeding programme for resistance already initiated at Herdade da Comporta, should be tested for their variability as well. Progeny inoculation trials may be a good way to obtain fast and accurate results and control the evolution of this hazardous disease.

INDÍCE

1.INTRODUÇÃO	1
1.1.A doença da murchidão do pinheiro	1
1.1.1.A doença da murchidão do pinheiro em Portugal	4
1.1.2.Estratégias de controlo da doença da murchidão do pinheiro	5
1.2.Melhoramento genético para a resistência à doença da murchidão do pinheiro	6
1.3.Objetivos do trabalho	9
1.4.Plano do trabalho	10
2.MATERIAL E MÉTODOS	11
2.1.Material vegetal	11
2.2.Agente patogénico	11
2.3.Ensaio de inoculação	11
2.4.Planeamento experimental.....	12
2.4.1.Descrição das variáveis de estudo.....	12
2.4.2.Desenho experimental.....	13
2.5.Metodologias estatísticas.....	13
2.5.1. Análise exploratória.....	13
2.5.2. Análise não-paramétrica	14
2.5.3. Ajustamento de modelos de regressão categorial.....	17
2.5.3.1.Regressão logística.....	17
2.5.3.2.Regressão ordinal.....	22
2.5.3.3.Regressão multinomial.....	25
3.RESULTADOS.....	27
3.1.Seleção de datas de observação.....	27
3.2.Análise exploratória	30
3.3.Ajustamento de modelos de regressão categorial.....	40
3.3.1.Regressão logística.....	40
3.3.2.Regressão ordinal	48
3.3.3.Regressão multinomial	53
4.DISSCUSSÃO, CONCLUSÕES E PERSPETIVAS FUTURAS.....	58

5.REFERÊNCIAS BIBLIOGRÁFICAS	62
Anexo I.....	70

LISTA DE QUADROS

Quadro 2.1. Recomendações de funções de ligação a usar na regressão ordinal, adaptado de Maroco (2007).	24
Quadro 3.1. Resultados dos testes de Qui-Quadrado para o ESTADO DA PLANTA, considerando as variáveis FAMÍLIA e CONCENTRAÇÃO isoladamente, para todas as observações realizadas durante a monitorização do ensaio.	27
Quadro 3.2. Resultados dos testes de Kruskal-Wallis para a SINTOMATOLOGIA, considerando as variáveis FAMÍLIA e CONCENTRAÇÃO isoladamente, para todas as observações realizadas durante a monitorização do ensaio.	28
Quadro 3.3. Resultados do teste de McNemar para a variável binária ESTADO DA PLANTA, considerando os pares de dias após inoculação (DAI).	29
Quadro 3.4. Teste de Wilcoxon para a variável ordinal SINTOMATOLOGIA, considerando os pares de dias após inoculação (DAI).	30
Quadro 3.5. Contagem do número de plantas por FAMÍLIA para os níveis de ESTADO DA PLANTA, aos 14, 35, 47 e 92 dias após inoculação (DAI) e respetivas percentagens.	31
Quadro 3.6. Contagem do número de plantas por CONCENTRAÇÃO para os níveis de ESTADO DA PLANTA, aos 14, 35, 47 e 92 dias após inoculação (DAI), e respetivas percentagens.	33
Quadro 3.7. Contagem do número de plantas por FAMÍLIA para os níveis de SINTOMATOLOGIA (SIN, 1 a 4), aos 14, 35, 47 e 92 dias após inoculação (DAI), e respetivas percentagens.	34
Quadro 3.8. Contagem do número de plantas por CONCENTRAÇÃO para os níveis de SINTOMATOLOGIA (SIN, 1 a 4), aos 14, 35, 47 e 92 dias após inoculação (DAI), e respetivas percentagens.	36
Quadro 3.9. Análise descritiva para as variáveis ALTURA e DIÂMETRO.	38
Quadro 3.10. Teste de Kruskal-Wallis à uniformidade de distribuição das alturas e diâmetros das plantas pelas diferentes variáveis independentes CONCENTRAÇÃO e FAMÍLIA.	39
Quadro 3.11. Avaliação da linearidade das variáveis explicativas contínuas relativamente à variável dependente ESTADO DA PLANTA, para as 4 datas de observação após inoculação (DAI): 14, 35, 47 e 92, testando a interação entre a variável explicativa e o seu logaritmo neperiano.	41
Quadro 3.12. Coeficientes Logit do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 35 DAI.	43
Quadro 3.13. Coeficientes Logit do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 47 DAI.	45
Quadro 3.14. Coeficientes Logit do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, e DIÂMETRO, aos 92 DAI.	47
Quadro 3.15. Coeficientes <i>Logit</i> do modelo de regressão ordinal dos diferentes níveis da variável SINTOMATOLOGIA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 47 DAI.	50
Quadro 3.16. Coeficientes <i>Logit</i> do modelo de regressão ordinal dos diferentes níveis da variável SINTOMATOLOGIA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 92 DAI.	52
Quadro 3.17. Coeficientes <i>Logit</i> do modelo de regressão multinomial dos diferentes níveis da variável SINTOMATOLOGIA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 35 DAI.	55

LISTA DE FIGURAS

Figura 1.1. Interação entre os diferentes agentes envolvidos na doença da murchidão dos pinheiros (Bonifácio 2009)	1
Figura 2.1. Procedimento de inoculação de pinheiros jovens em estufa.	12
Figura 3.1. Gráficos de barras para a variável FAMILIA relativamente aos níveis de ESTADO DA PLANTA, aos 14, 35, 47 e 92 dias após inoculação (DAI).	32
Figura 3.2. Gráficos de barras para a variável CONCENTRAÇÃO relativamente aos níveis de ESTADO DA PLANTA, aos 14, 35, 47 e 92 dias após inoculação (DAI).	33
Figura 3.3. Gráficos de barras para a variável FAMILIA relativamente aos níveis de SINTOMATOLOGIA, aos 14, 35, 47 e 92 dias após inoculação (DAI).	35
Figura 3.4. Gráficos de barras para a variável CONCENTRAÇÃO relativamente aos níveis de SINTOMATOLOGIA, aos 14, 35, 47 e 92 dias após inoculação (DAI).	37
Figura 3.5. Caixas de bigodes das distribuições das ALTURAS e DIÂMETROS das plantas por FAMÍLIA e por CONCENTRAÇÃO.	40
Figura 3.6. Curva ROC da relação entre a sensibilidade e a especificidade aos 35 DAI, para a variável dependente ESTADO DA PLANTA.	42
Figura 3.7. Curva ROC da relação entre a sensibilidade e a especificidade aos 47 DAI, para a variável dependente ESTADO DA PLANTA.	44
Figura 3.8. Curva ROC da relação entre a sensibilidade e a especificidade aos 92 DAI, para a variável dependente ESTADO DA PLANTA.	47
Figura 3.9. Representação gráfica das frequências relativas aos diferentes níveis da variável resposta SINTOMATOLOGIA aos 14 dias após inoculação (DAI).	49
Figura 3.10. Representação gráfica das frequências relativas aos diferentes níveis da variável resposta SINTOMATOLOGIA aos 35 dias após inoculação (DAI).	49
Figura 3.11. Representação gráfica das frequências relativas aos diferentes níveis da variável resposta SINTOMATOLOGIA aos 47 dias após inoculação (DAI).	50
Figura 3.12. Representação gráfica das frequências relativas aos diferentes níveis da variável resposta SINTOMATOLOGIA aos 92 dias após inoculação (DAI).	52

LISTA DE ABREVIATURAS

AFN – Autoridade Florestal Nacional

CV – Coeficiente de Variação

DAI – Dias Após Inoculação

DAP – Diâmetro à Altura do Peito

H₀ – Hipótese Nula

INRB – Instituto Nacional de Recursos Biológicos

K-S – Kolmogorov-Smirnov

MNE – Mata Nacional do Escaroupim

NMP – Nemátode da Madeira do Pinheiro

OR – Odds Ratio

PWD – Pine Wilt Disease

RAPD-PCR - Random Amplification of Polymorphic DNA - Polymerase Chain Reaction

RFLP – Restriction Fragment Length Polymorphism

SPSS – Statistical Package for the Social Sciences

UISPF – Unidade de Investigação em Silvicultura e Produtos Florestais

1. INTRODUÇÃO

1.1. A doença da murchidão do pinheiro

A doença da murchidão dos pinheiros, *pine wilt disease (PWD)* em inglês, é causada pelo nemátode da madeira do pinheiro (**NMP**), *Bursaphelenchus xylophilus*. É uma doença que afeta maioritariamente espécies do género *Pinus*, que apresentam diferentes níveis de resistência/tolerância. Entre estas espécies destacam-se *P. halepensis*, *P. pinea* e *P. radiata* com suscetibilidade intermédia, e *P. nigra*, *P. pinaster*, *P. thunbergii*, *P. densiflora* e *P. sylvestris* como suscetíveis. *B. xylophilus* requer um vetor para infectar novos hospedeiros, papel desempenhado por diversas espécies do género *Monochamus*, nomeadamente *M. alternatus*, *M. carolinensis* e *M. galloprovincialis* (Evans et al. 1996).

Os principais sintomas desta doença consistem num rápido e generalizado enfraquecimento da árvore com morte total ou parcial da copa. Durante este processo, as agulhas tornam-se amareladas e acastanhadas começando nas mais jovens, sendo no entanto capazes de se manter na árvore durante um longo período de tempo (Rodrigues et al. 2009). Até à data, não há nenhum método que permita distinguir visualmente entre árvores que estão a morrer devido à PWD ou a qualquer outro agente (Schröder et al. 2009), como por exemplo algumas espécies de escolítídeos (Bonifácio 2009).

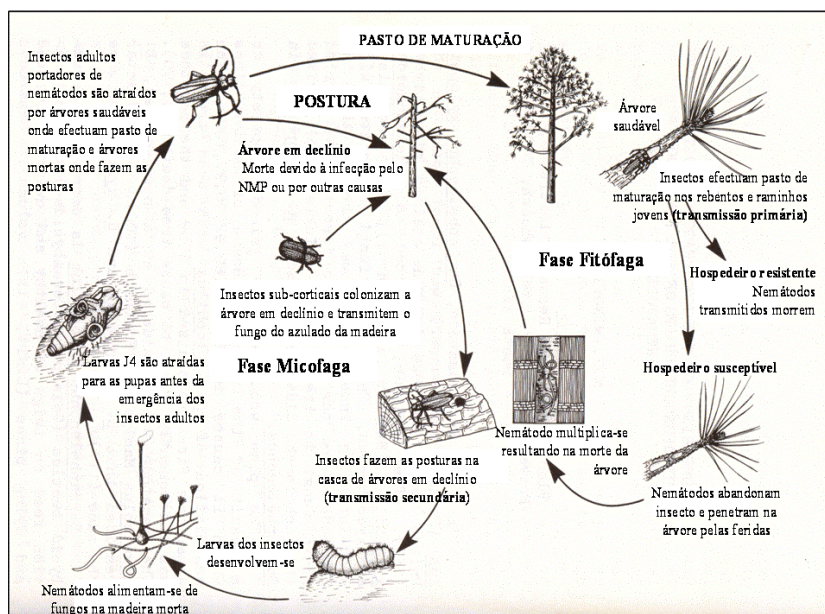


Figura 1.1 – Interação entre os diferentes agentes envolvidos na doença da murchidão dos pinheiros (Bonifácio 2009)

A complexa interação entre patógeno, vetor e hospedeiro vegetal está representada na Figura 1.1. Aproveitando o facto de que a exsudação de resina cessa quando um hospedeiro morre e as plantas deixam de ter um mecanismo de defesa contra os insetos, as espécies europeias deste inseto preferem fazer a oviposição em árvores enfraquecidas, depositando os ovos nos ramos pequenos da copa até à base do tronco (Schröder et al. 2009). Seis a nove dias depois, dependendo da temperatura ambiental, a larva eclode, e inicia a construção de uma galeria no tecido subcortical. Após a alimentação subcortical, as larvas escavam uma galeria centrípeta para o interior da madeira, que termina na câmara pupal. Caso este hospedeiro esteja infetado, os nemátodes aí presentes rodeiam as câmaras pupais construídas pelas larvas dos insetos (Bonifácio 2009). Quando os insetos emergem, transportam já os nemátodes nas suas traqueias, podendo infetar novas árvores (Evans et al. 1996; Gordillo e Kim 2012). O processo de infeção ocorre quando os insetos provenientes de madeira infetada, transportando o NMP no seu sistema traqueal, voam para pinheiros saudáveis para se alimentarem. A infeção do hospedeiro ocorre então através das feridas feitas pelos insetos vetores durante o pasto de maturação, pois os insetos provocam feridas no floema e xilema, deixando esses tecidos vulneráveis à penetração de *B. xylophilus* (Kuroda et al. 1991). Ainda de acordo com estes autores, a partir desses tecidos vasculares, os nemátodes chegam aos canais de resina, tanto radiais como laterais, conseguindo mover-se cerca de 25-50cm por dia. Imediatamente após a entrada no hospedeiro, os nemátodes começam a alimentar-se de células da planta, a reproduzir-se rapidamente e a depositar ovos, acabando por destruir os vasos, o que acabará por provocar a morte do hospedeiro. O NMP apresenta duas diferentes vias no seu ciclo de vida, a propagativa e a dispersiva (Wingfield 1987). Durante a fase propagativa de rápida multiplicação, a população de NMP é composta por adultos (machos e fêmeas) e quatro estádios juvenis. A multiplicação e distribuição dos nemátodes vai depender maioritariamente do grau de suscetibilidade da planta hospedeira, da sua condição fisiológica, e particularmente da temperatura ambiental (Rutherford e Webster 1987; Rutherford et al. 1990; Evans et al. 1996). A morte da árvore pode advir num período de cerca de 3 semanas (Rodrigues et al. 2009). A certa altura do processo invasivo, a população de NMP deixa de se multiplicar e começa a diminuir. Assim que isto acontece, a população inicia a sua fase dispersiva. O inseto, aproveitando o facto de que o hospedeiro se encontra infetado, efetua a oviposição, continuando assim o ciclo (Evans et al. 1996). Para que o ciclo de vida de *B. xylophilus* se complete, é necessário que certas condições estejam reunidas, nomeadamente, o hospedeiro ser suscetível, encontrar-se em condições desfavoráveis (como stress hídrico) concretizar-se a atração do hospedeiro vegetal ao vetor,

a temperatura ambiental ser a ideal à reprodução e multiplicação do agente patogénico (Evans et al. 1996). De acordo com Vicente et al. (2011) e Rutherford et al. (1990) as temperaturas médias do ar de aproximadamente 20 °C ou superiores são as ideais para o desenvolvimento da doença. Segundo alguns autores, o défice de água no solo ou temperaturas médias do ar de cerca de 27 °C, aumentam o rápido progresso dos sintomas (Li 2008; Ikeda e Kiyohara 1995). Árvores mais velhas parecem ser mais suscetíveis a manifestar a doença causada pelo NMP, que geralmente não se observa em árvores de idades inferiores a 5/6 anos (Rodrigues et al. 2009).

A PWD é nativa da América do Norte, nomeadamente do Canadá e Estados Unidos da América (Dropkin e Foudin 1979; Robbins 1982; Bowers et al. 1992; Sutherland 2008), com um registo também no México (Dwinell 1993). Nestas zonas, por norma não causa murchidão em coníferas, pois tanto o NMP como o seu hospedeiro têm uma história de co-evolução que permitiu que as árvores se tornassem tolerantes a este organismo (Kiyohara e Bolla 1990; Dwinell 1997). Assim, excetuando algumas plantações exóticas de pinheiro, na América do Norte, este nemátode apenas coloniza coníferas fragilizadas por outros agentes bióticos ou stresses abióticos (Evans et al. 1996; Naves 2008).

A atividade humana ao longo dos anos tem ajudado à dispersão da doença para fora da região de origem, principalmente através da importação/exportação de madeira infetada (Evans et al. 1996). Esta doença foi já introduzida no Japão (Mamiya 1983; Mamiya 1984; Kobayashi 1988; Mamiya 1988; Kishi 1995), China (Cheng et al. 1983; Yang e Qouli 1989; Yang 2004), Taiwan (Tzean e Tang 1985), Coreia do Sul (Yi et al. 1989; La et al. 1999; Shin 2008), Portugal continental (Mota et al. 1999) e Madeira (Fonseca et al. 2010), e Espanha (Abelleira et al. 2011; Robertson et al. 2011). Fora da região de origem, o NMP mostra-se muito agressivo e causador de elevadas mortalidades em pinheiros nativos.

A dispersão de *B. xylophilus* é uma preocupação global, dado que o impacto da sua presença se pode tornar preocupante num quadro de alterações climáticas, que irá aumentar a vulnerabilidade das florestas de pinheiros em termos mundiais, tornando esta doença um problema não exclusivo dos países onde já foi detetada até à data. Entre esses países destaca-se a Austrália, na qual existe um milhão de hectares de florestas plantadas, sendo que 80% são florestas de espécies de *Pinus*, e também a Nova Zelândia com os seus 1.6 milhões de hectares de florestas plantadas, e 89% de floresta comercial de *P. radiata*. À Nova Zelândia acresce o problema dos 200 000 m³/ano de comércio de madeira que passa nos seus portos (Mota & Vieira 2008). A problemática estende-se também a outros países europeus, como a França, que tem mais de metade da área total mundial de Pinheiro bravo (4,4 milhões de hectares), produzindo 15 milhões de metros cúbicos de madeira por ano, na

região da Aquitânia (Sanz *et al.* 2006). Por fim, há a destacar o facto de que as florestas do Norte da Europa são maioritariamente compostas de coníferas, especialmente espécies de *Pinus* como o *P. sylvestris*, que são de enorme importância económica para esses países (Vicente *et al.* 2011), e bastante suscetíveis à PWD (Evans *et al.* 1996).

1.1.1. A doença da murchidão do pinheiro em Portugal

Em 1999 *B. xylophilus* foi associado à morte de Pinheiro bravo (*P. pinaster*) na região de Marateca/Pegões, concelho de Palmela, em Portugal (Mota *et al.* 1999). Foi identificada uma maior proximidade genética entre as populações de NMP portuguesas e as asiáticas (através da análise dos padrões de restrição, RFLP) (Metge e Burgermeister 2005), o que coloca a madeira proveniente da Ásia como o possível ponto de infeção. Vieira *et al.* (2007), usando RAPD-PCR, estudaram 24 isolados da área inicialmente afetada (Península de Setúbal), mostrando a quase inexistente diversidade genética entre isolados, e a sua proximidade com os isolados chineses. Mais recentemente, e atendendo à dispersão da doença por todo o continente (Mota *et al.* 2009) e ilha da Madeira (Fonseca *et al.* 2010), Valadas *et al.* (2012) mostraram que a diversidade genética é agora maior, sugerindo múltiplas introduções a partir de diferentes origens.

O único vetor de *B. xylophilus* conhecido em Portugal é o cerambicídeo *Monochamus galloprovincialis* (Sousa *et al.* 2001), de nome comum longicórnio-do-pinheiro. A emergência dos imagos desta espécie, na região da península de Setúbal, dá-se desde finais de Maio até início de Setembro, com o pico em Julho (Naves 2007). Este autor concluiu que, caso as condições climáticas sejam favoráveis (temperatura do ar amena e pouca precipitação), existem insetos em voo até Outubro, com o máximo em Junho e Julho.

Quanto ao hospedeiro, o Pinheiro bravo é o único identificado até à data no nosso país (Bonifácio 2009). A distribuição deste pinheiro em Portugal corresponde maioritariamente à área litoral que vai desde a bacia do Tejo ao rio Minho, caminhando para o interior nas regiões Norte e Centro. Vegeta até altitudes de 700-900m, preferencialmente em encostas expostas à influência marítima, de Sudoeste a Norte (Oliveira *et al.* 2000). Esta espécie está também espalhada pela faixa atlântica do Sudoeste da Europa (França, Espanha e Portugal) e em menor escala no Mediterrâneo Ocidental (Itália, Sardenha, Sicília, Córsega, Marrocos, Argélia, Tunísia e Sul de França) (Pico 2011). Para além do vasto património genético presente em *P. pinaster* (Baradat e Marpeau 1988), em Portugal esta espécie abrange 27% da área florestal nacional (AFN 2010), representando também o maior volume, biomassa e carbono armazenado. No entanto, a comparação dos dados de inventário

actuais com anteriores, de 1995 a 1998, mostra que esta espécie teve uma redução de área ocupada de cerca de 9% em 10 anos, ou seja, perdeu cerca de $9,1 \times 10^4$ ha (AFN 2010; Pico 2011). Ainda de acordo com AFN (2010), o Pinheiro bravo ocupa atualmente uma área de $8,8 \times 10^5$ ha, em diversos tipos de povoamentos (puros, mistos dominantes, jovens puros/dominantes e dispersos), e representa um volume de $8,5 \times 10^7$ m³ de madeira em pé. A fileira florestal do Pinheiro bravo é de extraordinária importância para a economia nacional, assentando sobretudo na produção de madeira e resina. Contribuem também decisivamente para estabilização de dunas (Mota e Vieira 2008). Na floresta nacional de pinheiros, esta espécie é a mais abundante (62.5%), seguida de *P. pinea* e de *P. sylvestris* (Rodrigues 2008). As florestas de pinhal bravo são também zonas onde se podem desenvolver atividades de apicultura, caça, produção de cogumelos, ou resinagem. São também importantes pelo facto de serem uma fonte de biodiversidade, onde um enorme número de espécies se alberga. A expansão da PWD no nosso país afeta assim em grande medida a economia portuguesa.

1.1.2. Estratégias de controlo da doença da murchidão dos pinheiros

Dado que esta doença resulta de uma interação complexa entre nemátode, inseto e árvore hospedeira, basta que haja uma quebra neste circuito, para interromper o ciclo de infeção, sendo então possível controlar a doença. As estratégias de controlo, apesar de variadas podem incidir em cada um dos elementos desta interação, individualmente ou em conjunto. Entre elas destacam-se as culturais (remoção de árvores mortas ou sintomáticas antes da emergência do vetor, abate e sua destruição e/ou queima), as químicas (injeção de nematicidas em troncos e aplicação aérea de inseticidas) as biotécnicas (armadilhas para captura do inseto vetor), as biológicas (parasitóides, predadores, microrganismos entomopatogénicos e nematodes não patogénicos) e o melhoramento genético da árvore hospedeira (através do aumento da resistência dos hospedeiros à doença, plantação de árvores resistentes) (Kishi 1995; Evans et al. 1996; Shin 2008; Nose e Shiraishi 2008; Gordillo & Kim 2012).

As medidas culturais têm sido consideradas, até à data, o único método de controlo prático e com alguns resultados satisfatórios. É, no entanto, importante que a gestão se realize durante todas as fases do processo, tendo especial cuidado com o transporte da madeira proveniente de pinheiros infetados, que deve ser efetuada em contentores fechados, ou coberta por rede tratada com insecticida. Tal já constitui matéria da legislação portuguesa com o Decreto-Lei nº95/2011, de 8 de agosto, que determina que as madeiras de coníferas

para exportação têm de obrigatoriamente ser submetidas a choque térmico (56°C/30min.). De salientar que, todas as técnicas de gestão estão condicionadas por constrangimentos económicos, implementação prática e também conveniência e oportunidade política (Gordillo e Kim 2012).

Um outro aspecto a ter em conta na gestão desta doença é a possibilidade de existirem, nos povoamentos, árvores assintomáticas. Estas, embora se encontrem infectadas pelo nemátode, não apresentam sintomas visíveis. A ausência de sinais de doença pode ser motivada por condições ambientais e/ou fisiológicas não favoráveis ao NMP, como por exemplo, a infeção de uma árvore no final do período de voo do inseto (Setembro/Outubro), em que as temperaturas começam a decrescer, impossibilitando a evolução da doença. Contudo, os nemátodes podem permanecer no interior da planta, esperando pelas condições adequadas da próxima época que permitam o seu desenvolvimento (Futai, 2003).

1.2. Melhoramento genético para a resistência à doença da marchitão dos pinheiros

Como referido, o melhoramento genético é também uma das várias estratégias para controlar a doença da marchitão dos pinheiros, através da obtenção e produção de plantas que apresentem algum grau de resistência/tolerância a esta doença. É uma estratégia indirecta de combate e de desenvolvimento numa perspectiva de médio e longo prazo. Em nematologia, a resistência não significa imunidade (ausência de infeção ou de mortalidade na descendência), mas sim a capacidade de um hospedeiro de suprimir o desenvolvimento e reprodução do nemátode. De acordo com este conceito, um hospedeiro suscetível permite que o nemátode se reproduza livremente. Desta forma, a resistência pode ser avaliada através da capacidade de reprodução e desenvolvimento do NMP dentro da planta. Por outro lado, a tolerância é, tal como a resistência, um conceito relativo, descrevendo a sensibilidade de um hospedeiro ao parasitismo ou quantidade de danos que esse hospedeiro consegue suportar, e pode ser avaliado em termos de supressão em rendimento. Tolerância e resistência são qualidades independentes de um hospedeiro, e a seleção para ambas as características pode ser feita em conjunto ou em separado num programa de melhoramento genético (Boerma e Hussey 1992).

O melhoramento genético para uma espécie florestal consiste no desenvolvimento de um conjunto de ações que visam analisar a variabilidade existente para uma determinada característica, juntar a componente genética dessa variabilidade em indivíduos superiores, e por fim produzi-los em massa de forma a obter o máximo de ganho, aliado a uma elevada base genética (Zobel e Talbert 1984; Roldão 2000). O melhoramento genético florestal

baseia-se na teoria da genética quantitativa clássica, que assume que as características são herdadas através da ação de vários genes em diferentes loci, cada um produzindo um pequeno efeito (efeitos genéticos “aditivos”). O fenótipo é determinado pela soma dos efeitos genético e ambiental (Carson e Carson 1989). Sendo o melhoramento genético uma estratégia que implica manipular a variabilidade das características com o intuito de a reduzir e simultaneamente a conservar, tanto quanto possível, de forma a assegurar a elasticidade de adaptação a diversas condições ambientais, torna-se necessário criar uma ampla base genética de plantas melhoradas. A avaliação do potencial sucesso de uma seleção fenotípica é realizada através do conceito heritabilidade, que corresponde à proporção da variância fenotípica total que é variância genética. Uma das grandes vantagens das espécies florestais é, pois, a existência de uma enorme variabilidade genética, pouco alterada pela ação do Homem.

Inicialmente, muitos programas de melhoramento genético no campo florestal desenvolveram-se visando a produção de madeira, nomeadamente, o aumento em volume, a melhoria da forma do tronco e das propriedades da madeira, especialmente da densidade. Contudo, alguns programas também se focaram na resistência a doenças (Carson e Carson 1989). Para Danti et al. (2008), o melhoramento genético florestal com vista à resistência a factores bióticos ou abióticos adversos, justifica-se apenas em espécies de elevado valor silvícola e ecológico, ornamental ou tecnológico (comercial), quando os meios de luta à disposição não são aplicáveis em larga escala, ou têm eficácia limitada, e são de elevado custo. A maioria dos programas de melhoramento genético corresponde a um trabalho de décadas que deve ser estabelecido cuidadosamente, pois a oportunidade de modificar características de seleção ou outras circunstâncias é difícil (Carson e Carson 1989). Devido ao facto deste processo ser moroso, implicando não só as limitações de orçamento como factores biológicos, nomeadamente atingir as idades reprodutivas que permitam o cruzamento entre indivíduos com características desejadas, a maioria dos programas de melhoramento para a resistência encontra-se hoje em dia em fases iniciais (Sniezko 2006).

O início de um programa de melhoramento para a resistência envolve a seleção fenotípica de árvores saudáveis em florestas naturais, plantações, ou parques urbanos, onde a doença afetou um grande número de indivíduos (Sniezko 2006). Em algumas espécies, a resistência ao agente patogénico é o único critério de seleção, enquanto que noutras características adicionais fazem também parte da seleção. No entanto, quanto mais características são incluídas no programa, mais intensiva e dispendiosa será a actividade de melhoramento (Eriksson et al. 2006). Na sequência da seleção fenotípica, a avaliação genética em ambientes uniformes, quer através de testes clonais, ou de descendências, é o passo que

se segue no melhoramento para a resistência, usando normalmente inoculações artificiais como forma de detectar não só o grau, mas também o tipo de resistência/tolerância (Sniezko 2006). Estas experiências são realizadas, na maioria das vezes, com plantas jovens, contudo, os resultados não são totalmente representativos da resposta que as mesmas plantas teriam numa fase mais adulta. De acordo com Daub (2007) a inoculação é um válido preditor de sensibilidade ou tolerância em espécies de árvores e útil para descortinar patogenicidade ou ausência dela no NMP. Para Sniezko (2006), um aspecto fundamental num programa de melhoramento para a resistência passa necessariamente pelo estabelecimento de ensaios de campo, como forma de examinar a longevidade dessa resistência/tolerância em diferentes condições ambientais.

Dados reportados por programas de melhoramento para a resistência ao NMP realizados no Japão para as espécies *Pinus densiflora* e *P. thunbergii*, comparando rácios de sobrevivência usando famílias selecionadas e não selecionadas (Kato et al. 1995; Toda e Kurinobu 1998; Toda e Kurinobu 2001; Toda e Kurinobu 2002), mostraram que a variação familiar após inoculação artificial foi altamente significativa e que a heritabilidade média familiar excedeu 0,8 (Kurinobu 2008). Estes programas de melhoramento contribuíram para a selecção de algumas árvores resistentes e para o estabelecimento de pomares clonais com vista à produção de semente (Toda e Kurinobu 2002).

Em 2009, em Portugal iniciou-se um programa de melhoramento genético para a resistência/tolerância à PWD com a selecção de cerca de 500 árvores adultas de Pinheiro bravo candidatas a “plus”, numa área que apresenta uma das mais altas incidências da doença em Portugal. O critério de selecção destas árvores candidatas foi a dominância (árvores dominantes), idade (árvores adultas), diâmetro à altura do peito (DAP) (maior que 20-25 cm) e ausência de sintomas externos. Durante cerca de dois anos estas árvores foram sendo monitorizadas como forma de obter informação relevante quanto ao grau de tolerância e resistência à doença, ajudando na selecção das melhores famílias (Ribeiro et al. 2012). A selecção de árvores aparentemente saudáveis em zonas altamente fustigadas pela doença, mantê-las sob observação, cloná-las e analisar estes indivíduos e as suas descendências para a tolerância e resistência à doença constitui a metodologia de acção do programa de melhoramento para esta espécie. Pretende-se obter o núcleo da atividade de melhoramento, que é a população de melhoramento. Kurinobu (2008) descreveu critérios semelhantes para a selecção de árvores candidatas a “plus” e também para a sua avaliação, através de descendências de polinização aberta e testes clonais. O estabelecimento de um ensaio de descendências desta população de melhoramento é um projeto ambicionado para o futuro, tendo já sido iniciada a sementeira dessas mesmas descendências.

De acordo com Carson e Carson (1989), os esforços para o desenvolvimento de programas de melhoramento genético para a resistência a doenças devem ser focados também em populações melhoradas para outras características. A existência de um programa de melhoramento genético para o Pinheiro bravo que visa o aumento do volume e a melhoria da forma do tronco, estabelecido desde 1988 (Roulund et al, 1988), e já com resultados de avaliações genéticas, possibilita o estudo da hipótese de junção de objectivos de melhoramento para a produção de madeira de qualidade e para a resistência ao NMP.

De facto, foi iniciado na primeira metade da década de sessenta do século passado, em Portugal, um programa de seleção de árvores superiores na Mata Nacional de Leiria. A seleção efetuada, entre 1963 e 1965, pelo técnico Australiano D. H. Perry, consistiu essencialmente na escolha de 85 árvores superiores, seleccionadas para as características: altura, diâmetro à altura do peito (DAP), forma do tronco, fio espiralado, vigor e forma da copa (Perry e Hopkins 1967). Esta ação marcou o início do programa de melhoramento de Pinheiro bravo Australiano e em Portugal do estabelecimento do primeiro Pomar Clonal de Semente da Mata Nacional do Escaroupim (MNE) com 60 clones (Aguiar et al, 2011). Posteriormente, em 1978, finalizou-se um segundo Pomar Clonal de sementes com 49 destes clones representados. Mais tarde, estas árvores vieram a constituir parte da população de melhoramento do programa de melhoramento genético para o volume e forma do tronco (Roulund et al 1988). Em 1987, foi estabelecido, em três locais diferentes, um teste de descendências com 49 clones. Os resultados dos estudos genéticos realizados nestes campos de ensaio permitiram iniciar, em 2000, o estabelecimento de um pomar clonal de elite, também na MNE (hoje designado por Centro Experimental do Escaroupim). Este pomar de elite é constituído pelos 17 melhores clones (em termos do seu valor reprodutivo) dos 49 existentes no Pomar Clonal de Sementes da MNE, e permitem a obtenção de semente com ganhos genéticos de 21% em volume e 17% na retidão do fuste. Atualmente a produção deste pomar é superior a 100 kg de semente da categoria de Material Florestal de Reprodução Testado (Aguiar et al. 2003; 2011).

1.3. Objetivos do trabalho

O objetivo geral consiste em avaliar a variabilidade na resposta ao nemátode da madeira do pinheiro de plantas jovens de Pinheiro bravo provenientes de famílias testadas, resultantes do programa de melhoramento genético do Pinheiro bravo para o volume e forma do tronco. Mais concretamente, pretende-se explorar em diferentes seções temporais, a associação e a correlação entre os fatores família, concentração de inóculo aplicada, características

biométricas da planta, nomeadamente o diâmetro e a altura, e a resposta das plantas, avaliada em termos qualitativos, variáveis binárias e ordinais. Pretende-se ainda avaliar o efeito isolado e conjunto dos factores da família e concentração de nemátode inoculado, na resposta das plantas.

1.4. Plano do trabalho

Após a introdução teórica, no capítulo 1, onde se apresentam questões relacionadas com a doença da murchidão do pinheiro, segue-se o capítulo 2 onde são descritos os materiais utilizados (vegetal e agente patogénico), e as diversas metodologias aplicadas, nomeadamente a inoculação, o planeamento experimental adoptado, as diferentes variáveis em estudo, as análises estatísticas aplicadas. No capítulo 3 são apresentados os resultados da análise estatística exploratória e de ajustamento de modelos. Finalmente, são incluídas, no capítulo 4, a discussão, as conclusões e as perspetivas futuras.

2. MATERIAL E MÉTODOS

2.1. Material Vegetal

As plantas de Pinheiro bravo (*Pinus pinaster*) utilizadas neste estudo são plantas com dois anos, produzidas no viveiro da Mata Nacional de Valverde (Alcácer do Sal, Portugal), e provenientes de nove famílias de meios-irmãos existentes no pomar clonal testado da Mata Nacional do Escaroupim (Salvaterra de Magos, Portugal). Estas nove famílias (correspondentes aos números 8, 14, 15, 16, 17, 31, 34, 48, e 58) foram selecionadas de entre as 17 famílias testadas, obtidas no desenvolvimento de um programa de melhoramento genético para o volume e forma do tronco (Aguiar et al. 2011), pela sua disponibilidade à data do estudo.

2.2. Agente Patogénico

O isolado de *Bursaphelenchus xylophilus* selecionado (Bx011.001) foi disponibilizado pelo Laboratório de Nematologia do Instituto Nacional de Investigação Agrária e Veterinária (INIAV, Oeiras, Portugal). Foi obtido de madeira infetada proveniente de uma zona do concelho de Marinha Grande, distrito de Leiria, Portugal (N39° 43, 338" W9° 01, 557"). Dado que isolados com muito tempo de laboratório são passíveis de perder a sua patogenicidade, o isolado de referência usado nos testes de inoculação foi extraído e selecionado em Abril de 2011, mantendo-se assim apenas quatro meses em condições controladas de laboratório. Este mesmo isolado foi identificado morfológicamente como *B. xylophilus*, seguido de uma confirmação molecular, usando o método descrito por Ribeiro et al. (2012). A extração, manutenção e multiplicação do isolado, e a preparação do inóculo foram realizados no referido laboratório de nematologia.

2.3. Ensaio de inoculação

O ensaio em análise foi realizado na estufa existente no campus do INIAV, em Oeiras, tendo sido iniciado em 4 de Agosto de 2011. Esta estufa encontra-se equipada com um sistema de arrefecimento tipo *cooling*. A temperatura média do ar, registada diariamente durante o período em que decorreu o ensaio, foi de 25,8°C, com um desvio padrão de 4,3°C. As regas foram realizadas manualmente providenciando uma disponibilidade de água regular.

A metodologia de inoculação aplicada foi adaptada do método de Futai (1980) (Figura 2.1), através de um corte na casca, seguido de inoculação sem esterilização prévia da superfície. O procedimento experimental realizado iniciou-se retirando agulhas no terço superior da planta, seguido de uma incisão longitudinal recorrendo a um bisturi. Injectou-se 1 ml de

suspensão de NMP contendo diferentes concentrações de nemátodes e selou-se a zona de inoculação com algodão e parafilme para evitar a sua desidratação.



Figura 2.1 – Procedimento de inoculação de pinheiros jovens em estufa.

Após inoculação, a avaliação dos sintomas de murchidão foi baseada em quatro níveis de sintomatologia, adaptado de Daub (2007): 1 – até 25% da planta sintomática; 2 – de 26% a 50% da planta sintomática; 3 – de 51 a 75% da planta sintomática; 4 – mais de 76% da planta sintomática. A divisão em níveis/classes de murchidão foi expressa como uma percentagem de cobertura de sintomas relativamente à totalidade das agulhas, e associado com a vitalidade da planta. Este tipo de monitorização foi realizado semanalmente para todas as plantas presentes neste estudo até à quase total mortalidade das plantas, correspondendo a 14 semanas de observações.

2.4. Planeamento experimental

2.4.1. Descrição das variáveis de estudo

Estudaram-se dois factores: a concentração de nemátode (número de nemátodes por ml de água) utilizada no inóculo (CONCENTRAÇÃO) e a família testada da população de melhoramento para o volume e forma do tronco (FAMILIA). Os factores em estudo são caracterizados da seguinte forma:

- CONCENTRAÇÃO – variável ordinal, em que se consideraram 4 níveis (A – inoculação água, B – 250 NMP/ml, C – 500 NMP/ml, D – 750 NMP/ml);
- FAMILIA – variável nominal, com 9 níveis (famílias de meios-irmãos testadas do pomar clonal de sementes testado da Mata Nacional do Escaroupim);

A estas variáveis independentes foram adicionadas duas variáveis contínuas, a ALTURA e o DIÂMETRO, como covariáveis, dado que os dados correspondentes às mesmas não foram controlados no delineamento experimental, sendo apenas características das plantas usadas neste estudo. A altura foi medida desde a raiz ao topo superior e, dada a pequena dimensão das plantas, o diâmetro foi medido em três secções diferentes e obtida a média dessas secções.

Quanto às variáveis dependentes consideraram-se dois tipos de variáveis qualitativas: binária e ordinal. Como variável qualitativa binária considerou-se a divisão dos sintomas em dois estados (ESTADO DA PLANTA): 0 – planta viva; 1 – planta morta. Considerou-se planta viva quando esta apresentava até 50% das agulhas com sintomas, e planta morta quando mais de 50% da planta apresentava sintomas da doença. Quanto à variável qualitativa ordinal foi realizada a avaliação dos sintomas de murchidão (SINTOMATOLOGIA) em quatro níveis de resposta: 1 – até 25% da planta sintomática; 2 – de 26% a 50% da planta sintomática; 3 – de 51 a 75% da planta sintomática; 4 – mais de 76% da planta sintomática.

2.4.2. Desenho experimental

Dado a ocorrência de um gradiente de temperatura e humidade entre o painel húmido e a ventoinha extratora da estufa do sistema *cooling*, foi necessário o estabelecimento do ensaio utilizando um desenho experimental conjugado de parcelas subdivididas (*split-plot*) em blocos casualizados completos. Neste desenho experimental, o material experimental encontra-se dividido em blocos, sendo que cada bloco corresponde a uma única réplica. Este desenho requer também que cada bloco contenha o mesmo número de unidades experimentais que o seguinte. As unidades experimentais são divididas em parcelas principais e parcelas secundárias (Hoshmand 1997). Foi considerado como parcela principal a variável CONCENTRAÇÃO, e como parcela secundária a variável FAMILIA. Como é natural neste tipo de desenho experimental, os tratamentos da parcela secundária não estão dispostos aleatoriamente em todo o bloco, mas apenas nas parcelas principais (Hoshmand 1997). O número total de plantas estudadas foi de 432, sendo a disposição efetuada com 4 blocos ou repetições ($r=4$) de 108 plantas, cada bloco dividido aleatoriamente em 4 parcelas principais (correspondentes à variável CONCENTRAÇÃO), de 27 plantas cada. Por fim, cada parcela principal foi dividida em 9 parcelas secundárias (correspondentes à variável FAMILIA), com um total de 3 plantas por família.

O programa estatístico utilizado na análise dos dados foi o SPSS (v. 19, SPSS Inc., Chicago).

2.5. Metodologias estatísticas

2.5.1. Análise exploratória

De acordo com Pestana e Gageiro (2008), a estatística descritiva centra-se no estudo de características não uniformes das unidades observadas ou experimentadas. Aplica-se para reduzir, descrever e apresentar os dados, através de indicadores estatísticos e

representações gráficas e tabulares apropriadas. Neste estudo, utilizaram-se medidas descritivas de localização (tendência central e tendência não central), de dispersão (absoluta e relativa), assimetria e achatamento. Efetuou-se também um estudo descritivo univariado, em variáveis nominal, ordinais e quantitativas

2.5.2. Análise não paramétrica

Sempre que as condições de aplicação dos testes paramétricos não se verificaram, ou seja, quando se detectou violação dos pressupostos de normalidade e de homogeneidade das variâncias (Maroco 2007), ou no caso de as variáveis serem do tipo qualitativo (em escala pelo menos ordinal), utilizaram-se testes não-paramétricos.

Os testes não-paramétricos usados neste estudo foram os seguintes: teste de Kolmogorov-Smirnov, teste de Levene, teste do Qui-Quadrado, teste de Kruskal-Wallis, teste de McNemar e teste de Wilcoxon.

Para avaliar a normalidade de uma distribuição, o teste mais utilizado é o de Kolmogorov-Smirnov (K-S). Este teste não-paramétrico avalia o ajustamento ou aderência à normalidade da distribuição de uma variável de nível ordinal ou superior, através da comparação das frequências relativas acumuladas observadas $F(x_i)$, com as frequências relativas acumuladas esperadas $F_0(x_i)$. O valor do teste é a maior diferença entre ambas (D), e é dado pela equação: $D = \max [\max(|F(x_i) - F_0(x_i)|); \max(|F(x_{i-1}) - F_0(x_i)|)]$, em que $F_0(X) \sim N(\mu, \sigma)$. Tem como hipótese nula (H_0) que a variável de nível ordinal ou superior tem distribuição normal. Rejeita-se H_0 , se $p \leq \alpha$ (Pestana e Gageiro 2008).

Para testar a homogeneidade de variâncias usou-se o teste de Levene (Levene 1960). Este teste analisa a homogeneidade das variâncias, sendo bastante robusto em condições de desvio da normalidade. A sua estatística é dada por:

$$W = \frac{(N - k)}{(k - 1)} \cdot \frac{\sum_{i=1}^k n_i (\bar{Z}_i - \bar{Z})^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2}$$

em que n_i é a dimensão de cada uma das k amostras ($i = 1, \dots, k$) e N é a dimensão da amostra global. A variável Z pode definir-se como $Z_{ij} = |X_{ij} - \bar{X}_i|$ ($i = 1, \dots, k$ e $j = 1, \dots, n_i$), em que X_{ij} é a observação j da amostra i e \bar{X}_i é a média da amostra i . Por fim, \bar{Z}_i é a média de Z_{ij} na amostra i e \bar{Z} a média de Z_{ij} na amostra global. W tem uma distribuição F (Fisher-Snedcor) com $k-1$ e $N-k$ graus de liberdade e esta estatística tem como H_0 que as variâncias

das k -populações estimadas entre amostras são homogêneas ($\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$). Rejeita-se H_0 , se $p \leq \alpha$ (Maroco 2007).

O teste do qui-quadrado (X^2) estuda se duas ou mais populações independentes diferem relativamente a uma determinada característica. Os dados podem ser organizados em tabelas de frequências absolutas (tabelas de contingência). Este teste tem como hipótese nula a não existência de diferenças entre as amostras relativamente à distribuição nas classes da variável. A estatística deste teste é dada pela fórmula:

$$X^2 = \sum_{i=1}^L \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

em que C é o número de colunas e L o número de linhas da tabela de contingência. As frequências esperadas são dadas por:

$$E_{ij} = \frac{L_i \times C_j}{N}$$

onde

$$L_i = \sum_{j=1}^C O_{ij} \text{ e } C_j = \sum_{i=1}^L O_{ij}$$

se referem, respectivamente, aos totais em linha e em coluna (Maroco 2007).

O teste de Kruskal-Wallis (Kruskal e Wallis 1952) pode ser considerado como a alternativa não-paramétrica à ANOVA a um factor. Pode ser usado para definir se k amostras independentes provêm da mesma população ou de populações diferentes. A sua estatística é definida por:

$$H = \frac{\frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)}{1 - \frac{\sum_{i=1}^g (t_i^3 - t_i)}{N^3 - N}}$$

em que R_j representa a soma das ordens de cada uma das j amostras e N a dimensão da amostra global. O denominador é uma correção necessária caso existam mais de dois

grupos de empates, em que t é definido pelo número de observações em cada grupo de empates. H tem uma distribuição de qui-quadrado e este teste tem como hipótese nula que a distribuição dos valores da variável dependente é idêntica nas k populações, relacionando as medianas ($H_0: \theta_1 = \theta_2 = \dots = \theta_k$). A rejeição de H_0 significa que pelo menos uma das amostras provém de uma população diferente, contudo, não a identifica. Este teste exige mensuração no mínimo ao nível ordinal, e deve ser utilizado quando sujeitos diferentes são distribuídos por três ou mais situações (Siegel 1975).

Quanto ao teste de McNemar, é usado em experiências do tipo “Antes” vs. “Depois”, de forma a testar contagens ou proporções em populações de onde foram extraídas duas amostras relacionadas. A mensuração é feita ao nível nominal dicotômico, sendo cada indivíduo usado como o seu próprio controlo. O teste é efetuado usando uma tabela de contingência 2x2, em que A representa a mudança de estado de “+” para “-”, e B a mudança de “-” para “+”. A estatística do teste é dada por:

$$X^2 = \frac{(|A - B| - 1)^2}{A + B} \sim \chi^2(1)$$

A hipótese nula deste teste considera que as duas probabilidades marginais para cada uma das respostas é igual ($H_0: P(A) = P(B)$). Rejeita-se H_0 , se $p \leq \alpha$ (Maroco 2007).

Por fim, o teste de Wilcoxon usa-se para comparar a medida de tendência central da população sob estudo com um determinado valor teórico (Maroco 2007). É usado para comparar duas amostras relacionadas, ou medidas repetidas de uma mesma amostra, testando se a sua ordenação das médias é igual. Utiliza informações sobre o sentido da diferença entre cada par ($|D_i| = |Y_i - X_i|$), assim como o valor dessa diferença. Tem como pressupostos que os dados estão emparelhados e provêm da mesma população, que cada par é escolhido aleatoriamente e é independente, e que os dados são medidos numa escala intervalar. Tem como hipótese nula que a diferença das medianas entre os pares é zero, ou que o valor esperado de X é igual ao de Y ($H_0: E(X) = E(Y)$) (Siegel 1975). A estatística do teste quando existem empates é dada por:

$$T = \frac{W^+ - W^-}{\sqrt{\sum_{i=1}^n R_i^2}}$$

onde W^+ é a soma das ordens que correspondem a diferenças positivas, W^- é a soma das ordens que correspondem a diferenças negativas, e R_i é a ordem atribuída a $|D_i|$. Para amostras grandes, $n \geq 20$ a distribuição de T tende para normal, sendo a estatística do teste quando não existem empates (Pestana e Gageiro 2008):

$$T = \frac{W^+ - \frac{n(n+1)}{4}}{\sqrt{n(n+1)(2n+1)/24}}$$

Este teste tem ainda uma distribuição Z (N(0,1)).

2.5.3. Ajustamento de modelos de regressão categorial

A regressão define um conjunto vasto de técnicas estatísticas usadas para modelar relações entre variáveis e prever o valor de uma variável dependente (ou de resposta), a partir de um conjunto de preditores independentes, ou variáveis explicativas. Quando a variável dependente é de natureza qualitativa e assume apenas valores nas classes discretas e mutuamente exclusivas, deve ser usada a regressão categorial. A regressão categorial tem características que se destacam, tais como, o facto de considerar variáveis predictoras do tipo qualitativo e quantitativo, não assumir relação linear entre a variável dependente e os preditores, não exigir que os preditores apresentem distribuição normal, e ser menos sensível a observações extremas. No entanto, este tipo de regressão é designada de forma diferente consoante o tipo de variável dependente que se pretende modelar. Quando esta é de natureza nominal dicotómica, designa-se por regressão logística, quando é nominal policotómica, designa-se de regressão multinomial, e quando as classes desta variável podem ser ordenadas, diz-se regressão ordinal (Maroco 2007).

Deve-se no entanto ter em conta, antes de iniciar a análise de uma regressão categorial, que o objetivo final é sempre o de encontrar o melhor modelo, mais parcimonioso, no entanto, biologicamente razoável e que descreva a relação entre um grupo de variáveis (Hosmer e Lemeshow 2000).

2.5.3.1. Regressão logística

Tendo em conta a variável resposta ESTADO PLANTA, do tipo nominal dicotómico, a técnica mais apropriada de forma a modelar a ocorrência, em termos de probabilidades, de uma das duas realizações das classes da variável ($P[Y=1|X]$), é a regressão logística.

A função usada na regressão logística para estimar a probabilidade de uma determinada realização j ($j=1, \dots, n$) da variável dependente ser bem sucedida (sucesso), $P[Y_j=1]=\hat{\pi}_j$, é a função logística cuja forma genérica, para uma única variável independente X é:

$$\hat{\pi}_j = \frac{e^{\beta_0 + \beta_1 X_j}}{1 + e^{\beta_0 + \beta_1 X_j}}$$

Em que β_1 representa o coeficiente da variável independente em estudo, e β_0 o valor da probabilidade na ausência do efeito da variável independente.

A função logística para mais de uma variável independente (X_1, \dots, X_m), representa-se por:

$$\hat{\pi}_j = \frac{e^{\beta_0 + \beta_1 X_{1j} + \dots + \beta_m X_{mj}}}{1 + e^{\beta_0 + \beta_1 X_{1j} + \dots + \beta_m X_{mj}}}$$

Apesar destes modelos poderem ser ajustados recorrendo à regressão não-linear, a solução passa usualmente pela linearização desta função, através da transformação *Logit* (π), ou função de ligação:

$$\text{Logit}(\hat{\pi}_j) = \text{Ln}\left(\frac{\hat{\pi}_j}{1 - \hat{\pi}_j}\right)$$

Neste caso, o rácio $\hat{\pi}/(1-\hat{\pi})$ é designado por Rácio de Verosimilhança, Chances ou *odds*, e traduz a razão entre a probabilidade de sucesso (π) e a probabilidade de insucesso ($1-\pi$).

Dado que:

$$\text{Logit}(\hat{\pi}_j) = \beta_0 + \beta_1 X_j$$

Em que β_0 é o valor de $\text{Ln}\left(\frac{\hat{\pi}_j}{1-\hat{\pi}_j}\right)$ quando $Y_j = 0$ e β_1 representa o coeficiente *Logit* quando X_j

sofre um acréscimo unitário. A dificuldade interpretativa dos coeficientes do modelo de regressão logística com apenas uma variável independente (simples) ou com mais do que uma variável independente (múltipla), é usualmente superada com a interpretação da exponencial dos coeficientes. Assim, $\text{Exp}(\beta_i)$ representa o rácio das chances ou *odds ratio* (OR) do “sucesso” vs. “insucesso” quando a variável independente X_i aumenta uma unidade relativamente às chances do “sucesso” vs. “insucesso” quando X_i permanece constante. A percentagem do rácio de chances pode ser calculado usando a exponencial dos coeficientes do modelo a multiplicar por cem ($\% \text{Rácio Chances} = 100 \times [\text{exp}(\beta_j) - 1]$). Para um valor de

$Exp(\beta_i)$ superior a 1 ($\beta > 0$) ocorre um acréscimo das chances, enquanto que para um valor de $Exp(\beta_i)$ inferior a 1 ($\beta < 0$) ocorre um decréscimo das chances, quando a variável independente varia uma unidade (para variáveis independentes quantitativas) ou quando se passa da classe de referência para a classe em teste (para as variáveis independentes qualitativas). No caso de $Exp(\beta_i) = 1$, então a variável independente não afecta as chances (Maroco 2007).

Ainda de acordo com este autor, o modelo estatístico de regressão categorial deve cumprir os pressupostos de linearidade e aditividade, de proporcionalidade, de constância de efeito, os seus erros devem ser independentes, apresentar distribuição binomial, e por fim os preditores não devem ser multicolineares.

Antes de analisar as chances deve ser efetuado o ajustamento do modelo estatístico. A metodologia usada para ajustar os modelos na regressão logística é o método da Máxima Verosimilhança (*Maximum Likelihood*). Este método estima os coeficientes de regressão que maximizam a probabilidade de encontrar as realizações da variável dependente amostrada (Hosmer e Lemeshow 2000).

Uma vez ajustado o modelo de regressão logística é necessário avaliar a significância do modelo ajustado (teste ao rácio de verosimilhanças), a significância do ajustamento global (teste de Hosmer & Lemeshow), e a significância individual dos coeficientes (teste de Wald), sendo que as duas primeiras avaliações são de âmbito global e a terceira de âmbito individual.

A estatística do teste do rácio de verosimilhanças (G^2) testa a hipótese nula, $H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$, comparando a verosimilhança de um modelo só com a constante (L_0 ; modelo nulo ou reduzido), com a verosimilhança do modelo com as variáveis independentes (L_c ; modelo completo) (Agresti 2002):

$$G^2 = X_0^2 - X_c^2 = -2LnL_0 - (-2LnL_c) = -2Ln\left[\frac{L_0}{L_c}\right] \sim \chi^2(p)$$

em que X_0^2 é igual a $-2LnL_0$; e X_c^2 é $-2LnL_c$. O $-2LnL$ (*-2 Log Likelihood*) é usado como indicador de mediocridade do ajustamento do modelo aos dados. Considera-se que quanto maior o $-2LnL$, pior é o ajustamento. Para $-2LnL=0$ o ajustamento é perfeito. O modelo apenas é estatisticamente significativo quando a adição de uma ou mais variáveis independentes ao modelo, reduz significativamente o $-2LnL$. Em caso de rejeição de H_0 , conclui-se que pelo menos uma das variáveis independentes do modelo consegue prever

o *Logit* (π). No entanto, concluindo que o modelo completo é significativo, permite apenas afirmar que pelo menos uma das variáveis independentes do modelo completo influencia significativamente a variável dependente, não significando necessariamente que o ajustamento seja bom (Maroco 2007).

O teste de Hosmer & Lemeshow (2000) é baseado na comparação entre valores observados e as estimativas das probabilidades de sucesso de cada uma das n observações. A estatística deste teste segue uma distribuição do Qui-Quadrado com $K-2$ graus de liberdade e é calculada a partir de uma tabela de contingência. A estatística do teste é dada por:

$$X_{HL}^2 = \sum_{i=1}^g \frac{(O_i - E_i)^2}{E_i}$$

em que: O_i são os valores observados e E_i os valores esperados no grupo i . Para que ocorra ajustamento do modelo completo, os valores observados devem ser suficientemente próximos dos esperados (Maroco 2007). De acordo com este autor, sendo o modelo ajustado significativo, implica que existe pelo menos uma variável independente linearmente relacionada com o *Logit* (π_j).

O teste de Wald permite identificar quais as variáveis independentes que influenciam significativamente o *Logit*. Este teste avalia se determinado coeficiente é nulo, condicionado pelos valores estimados dos outros coeficientes, ou seja, a hipótese nula é $H_0: \beta_i = 0 | \beta_0, \beta_1, \beta_{i-1}; \beta_{i+1}; \beta_p$, sendo que a estatística do teste é dada por:

$$T_{Wald_i} = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$$

onde $\hat{\beta}_i$ é o estimador de β_i e $SE(\hat{\beta}_i)$ é o estimador do erro-padrão de β_i . Esta estatística tem distribuição *t-student*, que se aproxima assintoticamente da distribuição normal reduzida quando a dimensão da amostra é grande.

A procura do modelo mais parcimonioso que explique os dados leva a que a escolha das variáveis que irão fazer parte desse modelo seja alvo de determinados critérios (Hosmer e Lemeshow 2000). Na regressão logística múltipla é possível a utilização de algoritmos de seleção de variáveis com poder preditor. Dos diferentes métodos de seleção possíveis utilizou-se o método *Forward LR*. Neste método, a entrada de uma variável no modelo é

feita em função da significância da estatística "Score" e a remoção feita a partir da significância do teste do rácio de verosimilhança baseado nas estimativas parciais de máxima verosimilhança do modelo. De modo geral, a significância para adição e remoção de uma variável ao modelo é de $\alpha \leq 0,05$ e $\alpha > 0,10$, respectivamente (Maroco 2007).

De realçar que, para as variáveis contínuas o procedimento de ajustamento difere ligeiramente do efetuado para as nominais e ordinais. Neste caso, o modelo de regressão logística assume uma relação linear entre este tipo de variáveis explicativas e a transformação do *logit* da variável de resposta, ou seja, para cada medida da variável quantitativa ocorre um incremento ou diminuição na probabilidade do acontecimento. É necessário confirmar se o pressuposto de linearidade se aplica. Caso esta linearidade seja violada, então a regressão logística subestima o grau da relação das variáveis explicativas podendo gerar β não significativos quando de facto o são. Neste caso deve-se transformar a variável contínua numa variável categórica, a partir de pontos de corte de significado biológico ou usando os quartis. O teste à linearidade é efetuado aplicando o procedimento de Box-Tidwell, seguido do estudo da interação entre a variável contínua e o seu logaritmo, a partir de uma regressão logística. Considera-se então como hipótese nula a relação linear entre a variável contínua e a transformação do *Logit* da variável resposta (Agresti 2002; Pestana e Gageiro 2009).

Por fim, para avaliar a qualidade da classificação feita por um modelo ajustado, é usual comparar a percentagem global de classificações correctas obtidas com o modelo, com a percentagem proporcional de classificações correctas obtidas aleatoriamente. Esta percentagem é calculada a partir do número de sujeitos observados em cada uma das k classes da variável dependente (C_i), sobre o total (N), pela expressão:

$$\text{Classificação correcta por acaso (\%)} = 100 \times \sum_{i=1}^k \left(\frac{C_i}{N}\right)^2$$

Se a percentagem de casos classificados correctamente pelo modelo for superior em pelo menos 25% à percentagem de classificação proporcional por acaso, considera-se que o modelo tem boas propriedades classificativas (Maroco 2007). De acordo com este autor, a sensibilidade e especificidade permitem também avaliar a eficiência classificativa do modelo, sendo a sensibilidade, a percentagem de classificações correctas na classe de referência "sucesso" da variável dependente, e a especificidade, a percentagem de classificações correctas na classe "insucesso" do modelo. Considera-se um modelo com boas capacidades

preditivas quando este apresenta sensibilidade e especificidade superiores a 80%. Para percentagens entre 50 e 80% o modelo tem capacidades preditivas razoáveis. Abaixo dos 50% as capacidades preditivas são medíocres.

Outra medida de classificação é a área sob a curva ROC (*Receiver Operating Characteristic*), que varia entre 0 e 1. Para uma área igual a 0,5 o modelo não consegue discriminar os sujeitos “com” contra os “sem” a característica, melhor do que a escolha por mero acaso. Por outro lado, quanto mais próximo o valor for de 1, maior a capacidade do modelo para discriminar os indivíduos que apresentam a característica de interesse dos que não a apresentam. Contudo, é importante compreender que esta estatística avalia apenas se a probabilidade de ocorrência do grupo “sucesso” é ou não superior à do grupo “insucesso”. Um modelo pode apresentar um valor de ROC elevado e ainda assim prever de forma medíocre as probabilidades de sucesso observadas (Maroco 2007; Pestana e Gageiro 2009).

2.5.3.2. Regressão ordinal

Para a avaliação da variável dependente ordinal SINTOMATOLOGIA, considerou-se a regressão ordinal. Dada a relação de ordem entre as classes da variável dependente, a modelação deve ser efetuada em termos de probabilidades acumuladas (Maroco 2007; Pestana e Gageiro 2009).

No modelo de probabilidades acumuladas, a probabilidade de se observar uma classe inferior ou igual a k do conjunto das K classes da variável dependente, para um determinado vector de observações das variáveis independentes X , é dada por $(P(Y_j \leq k|x) = \pi_1 + \pi_2 + \dots + \pi_k \ (j = 1, \dots, n; k = 1, \dots, K)$ onde $\pi_1 = P(Y_j = 1), \pi_2 = P(Y_j = 2), \dots, \pi_k = P(Y_j = k)$. A ordem das classes é considerada no modelo, no entanto, este não engloba a última classe uma vez que o valor desta é redundante, dado ser a classe de referência. Em modelos ordinais, o “sucesso” é referido como acontecimento ($P(Y_j \leq k)$) e o seu complementar, “insucesso”, referido como não acontecimento ($1 - P(Y_j \leq k) = P(Y_j > k)$). Para linearizar o modelo de probabilidade cumulativa não linear, toma-se o logaritmo neperiano (Ln) do rácio das chances acumuladas $P(Y_j \leq k)/P(Y_j > k)$, ou $Logit [P(Y_j \leq k)]$, que é:

$$Logit[P(Y_j \leq k|x)] = Ln\left(\frac{P(Y_j \leq k|X)}{1 - P(Y_j \leq k|X)}\right) = Ln\left(\frac{P(Y_j \leq k|X)}{P(Y_j > k|X)}\right) = \alpha_k + X^*\beta$$

$$(k = 1, \dots, K - 1)$$

em que α_k representa o parâmetro de localização para as $k = 1, \dots, K-1$ classes da variável dependente, β é o vector dos coeficientes de regressão e X^* é a matriz das variáveis independentes (Maroco 2007). De acordo com este autor, é possível estimar o logaritmo da probabilidade da variável dependente apresentar valores de classes inferiores ou iguais a k , comparativamente com a probabilidade de tomar valores de classes superiores. Para $\beta > 0$, quando X aumenta, Y diminui (aumenta a probabilidade de a variável dependente ter níveis de ordem inferiores). No caso de $\beta < 0$, quando X aumenta, Y aumenta (aumenta a probabilidade de a variável dependente ter níveis de ordem superiores). Assim, sendo $\beta > 0$ é necessário re-escrever o modelo como:

$$\text{Logit}[P(Y_j \leq k|x)] = \text{Ln}\left(\frac{P(Y_j \leq k|x)}{P(Y_j > k|x)}\right) = \alpha_k - \beta X_j$$

A estatística do rácio de chances acumuladas, igual para todas as classes, é dado por:

$$OR_k = \frac{\frac{P(Y \leq k|x = x + d)}{P(Y > k|x = x + d)}}{\frac{P(Y \leq k|x = x)}{P(Y > k|x = x)}} = e^{-d\beta} = \frac{1}{e^{d\beta}}$$

Sendo d a distância entre os dois pontos da variável independente, então se $d = 1$, as chances de observar uma classe inferior ou igual a k , comparativamente com as chances de observar uma classe superior, variam e^β por cada unidade da variável independente (mantendo constantes as restantes variáveis independentes).

O modelo de variável latente assume a existência de uma variável latente contínua (η) impossível de medir directamente, sendo a variável manifesta (Y) resultante do corte da variável latente em K -classes, ordinais e mutuamente exclusivas. Dado que a probabilidade de se observar uma determinada classe para um determinado valor de X é dada pela área da curva entre dois pontos de corte, então a probabilidade de se observar uma determinada classe de Y é dada por $P(Y_j = k|x_j) = P(\alpha_{k-1} \leq \eta \leq \alpha_k|x_j)$. Sendo F a função de distribuição dos erros do modelo estrutural, substituindo η_j por $\beta_0 + \beta X_j + \varepsilon_j$ e rearranjando os termos, pode-se escrever: $P(Y_j = k|x_j) = F(\alpha_k - \beta X_j) - F(\alpha_{k-1} - \beta X_j)$ eliminando a constante β_0 que é redundante com α . O inverso da função F , designa-se por função de

ligação, dado que promove a ligação linear entre a parte aleatória do modelo e a parte sistemática (Maroco 2007). De acordo com este autor, no programa SPSS estão disponíveis cinco funções de ligação (quadro 2.1), cuja utilização depende do tipo de distribuição de probabilidades que as classes da variável dependente apresentam. Uma má escolha de função de ligação pode comprometer a significância do modelo e a sua capacidade preditiva. Esta escolha pode ser feita recorrendo a um histograma das frequências da variável resposta.

Quadro 2.1 – Recomendações de funções de ligação a usar na regressão ordinal, adaptado de Maroco (2007).

	Função Link (F^{-1})	Usar quando
<i>Logit</i>	$\text{Ln}\left[\frac{P(Y \leq k)}{P(Y > k)}\right]$	As classes de Y apresentam distribuição uniforme
<i>Log-log Complementar</i>	$\text{Ln}(-\text{Ln}(1 - P[Y \leq k]))$	As classes de Y de maior ordem são as mais frequentes
<i>Log-log Negativo</i>	$-\text{Ln}(-\text{Ln}(P[Y \leq k]))$	As classes de Y de menor ordem são as mais frequentes
<i>Cauchit</i>	$\text{Tan}(\pi(P[Y \leq k] - 0.5))$	As classes de Y de menores e maiores ordens são as mais frequentes
<i>Probit</i>	$\Phi^{-1}(P[Y \leq k])$, onde Φ é a função de distribuição $N(0,1)$	A variável latente é do tipo normal (assumpção)

Quanto à avaliação da qualidade do modelo, esta é feita de forma equivalente à efetuada para a regressão logística. A significância é avaliada pelo teste do rácio de verosimilhanças. A avaliação do tipo de modelo (*Logit*, *Probit*, etc.) pode fazer-se comparando o $-2\text{Ln}L$ dos dois modelos em causa. O melhor modelo será o que apresentar menor $-2\text{Ln}L$ (Maroco 2007).

O modelo de regressão ordinal assume que a influência das variáveis independentes sobre a função de ligação é igual para todas as K classes da variável dependente (paralelismo das linhas da função de ligação). Logo, deve-se usar o teste de linhas paralelas, em que $H_0: \beta_1 = \beta_2 = \dots = \beta_{k-1}$. A estatística de teste usada é o rácio de verosimilhança de dois modelos ordinais, o primeiro assumindo que os declives são iguais e o segundo assumindo que são diferentes. É representada por:

$$X_{LP}^2 = -2LL_{H_1} - (-2LL_{H_0}) = -2\text{Ln}\left[\frac{L_{H_1}}{L_{H_0}}\right] \sim \chi_{(k-2)p}^2$$

Para um $p - value = P(\chi^2 \geq X_{Lp}^2)$ do teste muito pequeno, rejeita-se H_0 e conclui-se que os declives não são homogêneos. Pretende-se no entanto não rejeitar a hipótese nula e, caso essa rejeição ocorra, pode ser associada a uma má escolha da função de ligação. Sempre que o teste da homogeneidade das variâncias não permita a utilização da regressão ordinal deve-se considerar o uso da regressão multinomial, não contemplando a relação de ordem entre as classes da variável resposta.

2.5.3.3. Regressão multinomial

Este tipo de regressão é normalmente aplicado quando a variável dependente é policotômica nominal (apresentando mais de duas classes mutuamente exclusivas) (Norusis 2008).

Ajustando a metodologia definida por Maroco (2007) a uma variável dependente de quatro classes, a probabilidade da variável dependente Y tomar o valor de qualquer uma das classes, considerando a normalização do sistema relativamente a uma categoria da variável dependente, e igualando o coeficiente β_0 a zero, vem:

$$P(Y = 0|X) = \frac{1}{1 + e^{\beta_{10} + \beta_{11}X_1 + \dots + \beta_{1p}X_p} + e^{\beta_{20} + \beta_{21}X_1 + \dots + \beta_{2p}X_p} + e^{\beta_{30} + \beta_{31}X_1 + \dots + \beta_{3p}X_p}}$$

$$P(Y = 1|X) = \frac{e^{\beta_{10} + \beta_{11}X_1 + \dots + \beta_{1p}X_p}}{1 + e^{\beta_{10} + \beta_{11}X_1 + \dots + \beta_{1p}X_p} + e^{\beta_{20} + \beta_{21}X_1 + \dots + \beta_{2p}X_p} + e^{\beta_{30} + \beta_{31}X_1 + \dots + \beta_{3p}X_p}}$$

$$P(Y = 2|X) = \frac{e^{\beta_{20} + \beta_{21}X_1 + \dots + \beta_{2p}X_p}}{1 + e^{\beta_{10} + \beta_{11}X_1 + \dots + \beta_{1p}X_p} + e^{\beta_{20} + \beta_{21}X_1 + \dots + \beta_{2p}X_p} + e^{\beta_{30} + \beta_{31}X_1 + \dots + \beta_{3p}X_p}}$$

$$P(Y = 3|X) = \frac{e^{\beta_{30} + \beta_{31}X_1 + \dots + \beta_{3p}X_p}}{1 + e^{\beta_{10} + \beta_{11}X_1 + \dots + \beta_{1p}X_p} + e^{\beta_{20} + \beta_{21}X_1 + \dots + \beta_{2p}X_p} + e^{\beta_{30} + \beta_{31}X_1 + \dots + \beta_{3p}X_p}}$$

As chances de ocorrer uma das classes da variável dependente relativamente à classe de referência 0 são então:

$$\frac{P(Y = 1|X)}{P(Y = 0|X)} = e^{\beta_{10} + \beta_{11}X_1 + \dots + \beta_{1p}X_p}$$

$$\frac{P(Y = 2|X)}{P(Y = 0|X)} = e^{\beta_{20} + \beta_{21}X_1 + \dots + \beta_{2p}X_p}$$

$$\frac{P(Y = 3|X)}{P(Y = 0|X)} = e^{\beta_{30} + \beta_{31}X_1 + \dots + \beta_{3p}X_p}$$

sendo o modelo *Logit*:

$$\text{Ln}\left[\frac{P(Y = 1|X)}{P(Y = 0|X)}\right] = \beta_{10} + \beta_{11}X_1 + \dots + \beta_{1p}X_p$$

$$\text{Ln}\left[\frac{P(Y = 2|X)}{P(Y = 0|X)}\right] = \beta_{20} + \beta_{21}X_1 + \dots + \beta_{2p}X_p$$

$$\text{Ln}\left[\frac{P(Y = 3|X)}{P(Y = 0|X)}\right] = \beta_{30} + \beta_{31}X_1 + \dots + \beta_{3p}X_p$$

Assim, considerando uma variável dependente de k classes são necessários $k-1$ chances relativas à classe de referência da variável dependente. Estas $k-1$ equações constituem o modelo multinomial. Este modelo é ajustado usando o método da máxima verossimilhança e as medidas de qualidade do ajustamento já referidos para as outras regressões categoriais.

De acordo com Maroco (2007), os rácios de chances calculam-se por:

$$OR(c, 0|X_i) = \text{Exp}(\beta_{ci}) = \frac{P(Y = c|X_i = x_i + 1)/P(Y = 0|X_i = x_i + 1)}{P(Y = c|X_i = x_i)/P(Y = 0|X_i = x_i)}$$

sendo sempre relativos à classe de referência. O rácio de chances de uma classe 2 em relação a uma classe 1 pode ser obtido pelo quociente das probabilidades de se observarem as duas classes de interesse. Considerando um modelo com apenas uma variável independente:

$$\text{Ln}\left[\frac{P(Y = 2|X)}{P(Y = 1|X)}\right] = (\beta_{02} - \beta_{01}) + (\beta_{21} - \beta_{11})X$$

Para a exponencial, o rácio de chances fica então:

$$\exp(\beta_{21} - \beta_{11}) = \exp(\beta_{21}) / \exp(\beta_{11})$$

Assim, o rácio de chances de duas classes é simplesmente o rácio das chances de cada uma dessas classes relativamente à classe de referência.

3.RESULTADOS

No sub-capítulo 3.1. apresentam-se os resultados referentes à seleção das datas de estudo. No sub-capítulo 3.2. é efetuada a análise exploratória dos dados, tendo em conta a variável resposta binária, ESTADO DA PLANTA, a ordinal, SINTOMATOLOGIA, e as covariáveis, ALTURA e DIÂMETRO. Por fim, no sub-capítulo 3.3. estão representados os resultados do ajustamento dos modelos estatísticos, para as datas de observação selecionadas, relativos às regressões logística, ordinal e multinomial.

3.1. Seleção de datas de observação

As 14 observações semanais permitiram obter um conjunto alargado de dados durante o ensaio, para os dois tipos de variável resposta qualitativa, ESTADO DA PLANTA e SINTOMATOLOGIA. De forma a proceder à exploração dos resultados obtidos, considerando diferentes seções temporais independentes, optou-se por escolher quatro datas de análise selecionadas a partir destas 14 observações. Consoante a variável resposta, ESTADO DA PLANTA ou SINTOMATOLOGIA, este estudo foi realizado, respetivamente, através dos testes de Qui-Quadrado e Kruskal-Wallis.

Relativamente à variável resposta binária, ESTADO DA PLANTA, o quadro 3.1 apresenta os resultados do teste de ajustamento do Qui-Quadrado, considerando cada uma das variáveis independentes, FAMÍLIA e CONCENTRAÇÃO, isoladamente, e para cada uma das datas de observação. A hipótese nula, H_0 , considera a frequência de mortes uniforme nas 9 famílias e nas 4 concentrações, para probabilidades de erro tipo I (α) de 0,05.

Quadro 3.1 – Resultados dos testes de Qui-Quadrado para o ESTADO DA PLANTA, considerando as variáveis FAMÍLIA e CONCENTRAÇÃO isoladamente, para todas as observações realizadas durante a monitorização do ensaio.

DAI	Família			Concentração		
	χ^2	g.l.	p	χ^2	g.l.	p
7	8,02	8	0,432	3,01	3	0,391
14	8,02	8	0,432	3,01	3	0,391
21	10,64	8	0,223	11,95	3	0,008
28	19,04	8	0,015	68,00	3	<0,001
35	28,60	8	<0,001	106,38	3	<0,001
40	28,58	8	<0,001	147,22	3	<0,001
47	26,03	8	0,001	160,52	3	<0,001
55	24,12	8	0,002	164,78	3	<0,001
61	22,33	8	0,004	166,17	3	<0,001
69	17,07	8	0,029	166,67	3	<0,001
75	15,96	8	0,043	171,20	3	<0,001
85	13,98	8	0,082	190,13	3	<0,001
92	13,53	8	0,095	197,98	3	<0,001
97	12,98	8	0,113	198,30	3	<0,001

Quando considerado o comportamento da variável dependente ESTADO DA PLANTA face ao efeito isolado da variável FAMÍLIA, verifica-se que até aos 21 DAI não existem evidências que permitam concluir que a proporção das plantas que morrem varie com a família, dado que não ocorre rejeição de H_0 ($X^2(8)=10,64$; $p=0,223$; $N=432$). No entanto, a partir de 28 DAI ($X^2(8)=19,04$; $p=0,015$; $N=432$), até aos 75 DAI ($X^2(8)=15,96$; $p=0,043$; $N=432$) rejeita-se H_0 , para um nível de significância $\alpha=0,05$. O nível de significância para a rejeição de H_0 passa de altamente significativo ($p\leq 0,001$) aos 35 DAI ($X^2(8)=28,60$; $p<0,000$; $N=432$) para significativo ($p\leq 0,05$) aos 69 DAI ($X^2(8)=17,07$; $p=0,029$; $N=432$). Dos 85 DAI ($X^2(8)=13,98$; $p=0,082$; $N=432$) aos 92 DAI ($X^2(8)=13,53$; $p=0,095$; $N=432$), apenas ocorreria rejeição de H_0 para um nível de significância de $\alpha=0,10$. Aos 97 DAI ($X^2(8)=12,98$; $p=0,113$; $N=432$) não ocorre rejeição de H_0 para qualquer nível de significância do analista.

Relativamente à variável CONCENTRAÇÃO, é possível verificar que até aos 14 DAI não existem evidências para considerar que as frequências de mortes sejam uniformes nas 4 concentrações ($X^2(3)=3,01$; $p=0,391$; $N=432$). Contudo, a partir dos 28 DAI até aos 97 DAI, rejeita-se H_0 para qualquer nível de significância do analista.

Quanto à variável resposta ordinal SINTOMATOLOGIA, o quadro 3.2 apresenta os resultados obtidos com o teste de Kruskal-Wallis para se avaliar o efeito isolado das variáveis independentes FAMÍLIA e CONCENTRAÇÃO em que a hipótese nula, H_0 , considera a distribuição dos valores da variável dependente idênticos nas diferentes classes da variável independente selecionada (medianas iguais), para probabilidades de erro tipo I (α) de 0,05.

Quadro 3.2 – Resultados dos testes de Kruskal-Wallis para a SINTOMATOLOGIA, considerando as variáveis FAMÍLIA e CONCENTRAÇÃO isoladamente, para todas as observações realizadas durante a monitorização do ensaio.

DAI	Família			Concentração		
	X^2	g.l.	P	X^2	g.l.	P
7	7,02	8	0,535	2,00	3	0,571
14	7,22	8	0,513	1,57	3	0,666
21	17,75	8	0,023	26,00	3	<0,001
28	25,86	8	0,001	86,80	3	<0,001
35	31,05	8	<0,001	123,50	3	<0,001
40	27,23	8	0,001	156,46	3	<0,001
47	26,99	8	0,001	167,83	3	<0,001
55	25,08	8	0,002	175,40	3	<0,001
61	23,03	8	0,003	175,74	3	<0,001
69	20,36	8	0,009	176,57	3	<0,001
75	18,21	8	0,020	185,98	3	<0,001
85	16,18	8	0,040	197,38	3	<0,001
92	16,47	8	0,036	188,66	3	<0,001
97	15,70	8	0,047	192,35	3	<0,001

Quando considerado o efeito isolado da variável FAMÍLIA, constata-se que até aos 14 DAI não existem evidências que permitam concluir que as plantas apresentem diferentes níveis de SINTOMATOLOGIA ($\chi^2_{KW}(8)=7,22$; $p=0,513$; $N=432$), para qualquer nível de significância do analista. Todavia, a partir dos 21 DAI até aos 97 DAI rejeita-se H_0 para um nível de significância de $\alpha=0,05$, sendo no entanto este nível de significância decrescente a partir dos 40 DAI.

Quanto à variável CONCENTRAÇÃO, existem evidências que permitem considerar que a partir dos 21 dias após inoculação a resposta das plantas a diferentes concentrações de inóculo apresenta diferenças altamente significativas ($p\leq 0,001$), até ao fim do ensaio.

Em sumula, para ambas as variáveis, FAMÍLIA e CONCENTRAÇÃO, as respostas, quer em termos de ESTADO DA PLANTA, quer em termos de SINTOMATOLOGIA, variam a partir dos 21 DAI. Optou-se então por selecionar quatro datas distintas: 14 DAI (última data sem rejeição de H_0 para ambas as variáveis), 35 e 47 DAI (a meio do ensaio), e 92 DAI (no final do ensaio).

Tendo em conta que as variáveis de resposta podem ser comparadas entre datas de observação e não apenas entre os níveis de resposta em cada data, foram efetuados testes de comparação para as quatro datas atrás referidas. Para a variável resposta binária aplicou-se o teste de McNemar (Siegel 1975), que considera a hipótese nula da igualdade entre as probabilidades marginais de cada um dos binómios das datas de observação (quadro 3.3). Para a variável ordinal, aplicou-se o teste de Wilcoxon, em que H_0 considera que as diferenças das medianas entre os pares das medidas repetidas da mesma amostra são iguais (quadro 3.4). Para ambos os testes considera-se o nível de significância (α) de 0,05.

Para a variável ESTADO DA PLANTA, a resposta das plantas, entre os pares de datas considerados, apresenta diferenças altamente significativas ($p<0,001$), ou seja, a proporção de plantas mortas altera-se entre as duas datas.

Quadro 3.3 – Resultados do teste de McNemar para a variável binária ESTADO DA PLANTA, considerando os pares de dias após inoculação (DAI).

Pares DAI	ESTADO DA PLANTA	
	χ^2	P
14 e 35	174,01	<0,001
35 e 47	46,17	<0,001
47 e 92	44,18	<0,001
14 e 92	272,00	<0,001

Relativamente à variável resposta ordinal, SINTOMATOLOGIA, o teste de Wilcoxon indica que para as quatro datas consideradas, quando comparadas duas a duas, as respostas diferem também entre si ($p < 0,001$).

Quadro 3.4 – Teste de Wilcoxon para a variável ordinal SINTOMATOLOGIA, considerando os pares de dias após inoculação (DAI).

Pares DAI	SINTOMATOLOGIA	
	Z	P
14 e 35	-12,98	<0,001
35 e 47	-8,83	<0,001
47 e 92	-8,59	<0,001
14 e 92	-16,33	<0,001

Em resumo, as quatro datas selecionadas para seções temporais representam respostas globais diferentes.

3.2. Análise exploratória

De seguida são apresentados os resultados da análise exploratória aos dados, considerando as quatro datas de observação selecionadas anteriormente, 14, 35, 47 e 92 DAI.

VARIÁVEL ESTADO DA PLANTA

Os resultados obtidos para a variável resposta binária, ESTADO DA PLANTA, por cada uma das variáveis independentes, FAMÍLIA e CONCENTRAÇÃO, foram analisados através de tabelas de contingência e de gráficos de barras.

Assim, de acordo com o quadro 3.5, aos 14 DAI a resposta observada para a variável FAMILIA é de um modo geral constante, apresentando na sua maioria plantas saudáveis (99,8%).

A partir dos 35 DAI, verifica-se uma variação de resposta das famílias. A família 34 é a que apresenta um menor número de plantas mortas até final das observações (45,8%). Por outro lado, as três famílias, 8, 17 e 58, são as que apresentam níveis mais elevados de mortalidade aos 35 DAI (56,2%, 54,2% e 54,2%, respectivamente). Contudo, aos 92 DAI, a família 58 é a que apresenta uma mortalidade mais elevada (77,1%).

Quadro 3.5 – Contagem do número de plantas por FAMÍLIA para os níveis de ESTADO DA PLANTA, aos 14, 35, 47 e 92 dias após inoculação (DAI) e respectivas percentagens.

ESTADO DA PLANTA									
FAM	14 DAI		35 DAI		47 DAI		92 DAI		Total
	Saudável	Morta	Saudável	Morta	Saudável	Morta	Saudável	Morta	
8	48 (100%)	0 (0,0%)	21 (43,8%)	27 (56,2%)	16 (33,3%)	32 (66,7%)	15 (31,2%)	33 (68,8%)	48
14	48 (100%)	0 (0,0%)	27 (56,2%)	21 (43,8%)	22 (45,8%)	26 (54,2%)	18 (37,5%)	30 (62,5%)	48
15	48 (100%)	0 (0,0%)	36 (75,0%)	12 (25,0%)	28 (58,3%)	20 (41,7%)	19 (39,6%)	29 (60,4%)	48
16	48 (100%)	0 (0,0%)	28 (58,3%)	20 (41,7%)	24 (50,0%)	24 (50,0%)	20 (41,7%)	28 (58,3%)	48
17	47 (97,9%)	1 (2,1%)	22 (45,8%)	26 (54,2%)	16 (33,3%)	32 (66,7%)	13 (27,1%)	35 (72,9%)	48
31	48 (100%)	0 (0,0%)	27 (56,2%)	21 (43,8%)	27 (56,2%)	21 (43,8%)	17 (35,4%)	31 (64,6%)	48
34	48 (100%)	0 (0,0%)	39 (81,2%)	9 (18,8%)	32 (66,7%)	16 (33,3%)	26 (54,2%)	22 (45,8%)	48
48	48 (100%)	0 (0,0%)	33 (68,8%)	15 (31,2%)	26 (54,2%)	22 (45,8%)	18 (37,5%)	30 (62,5%)	48
58	48 (100%)	0 (0,0%)	22 (45,8%)	26 (54,2%)	14 (29,2%)	34 (70,8%)	11 (22,9%)	37 (77,1%)	48
Total	431 (99,8%)	1 (0,2%)	255 (59,0%)	177 (41,0%)	205 (47,4%)	227 (52,6%)	157 (36,3%)	275 (63,7%)	432

Os gráficos que se seguem dizem respeito à distribuição percentual de plantas mortas e saudáveis por data de observação (figura 3.1).

Como se pode verificar, a família 34 predomina em termos de proporção de plantas saudáveis aos 35, 47 e 92 DAI. Aos 14 DAI não se detetaram diferenças na mortalidade. Por outro lado, verifica-se que a família 58 é a que apresenta uma maior mortalidade na última data de observação.

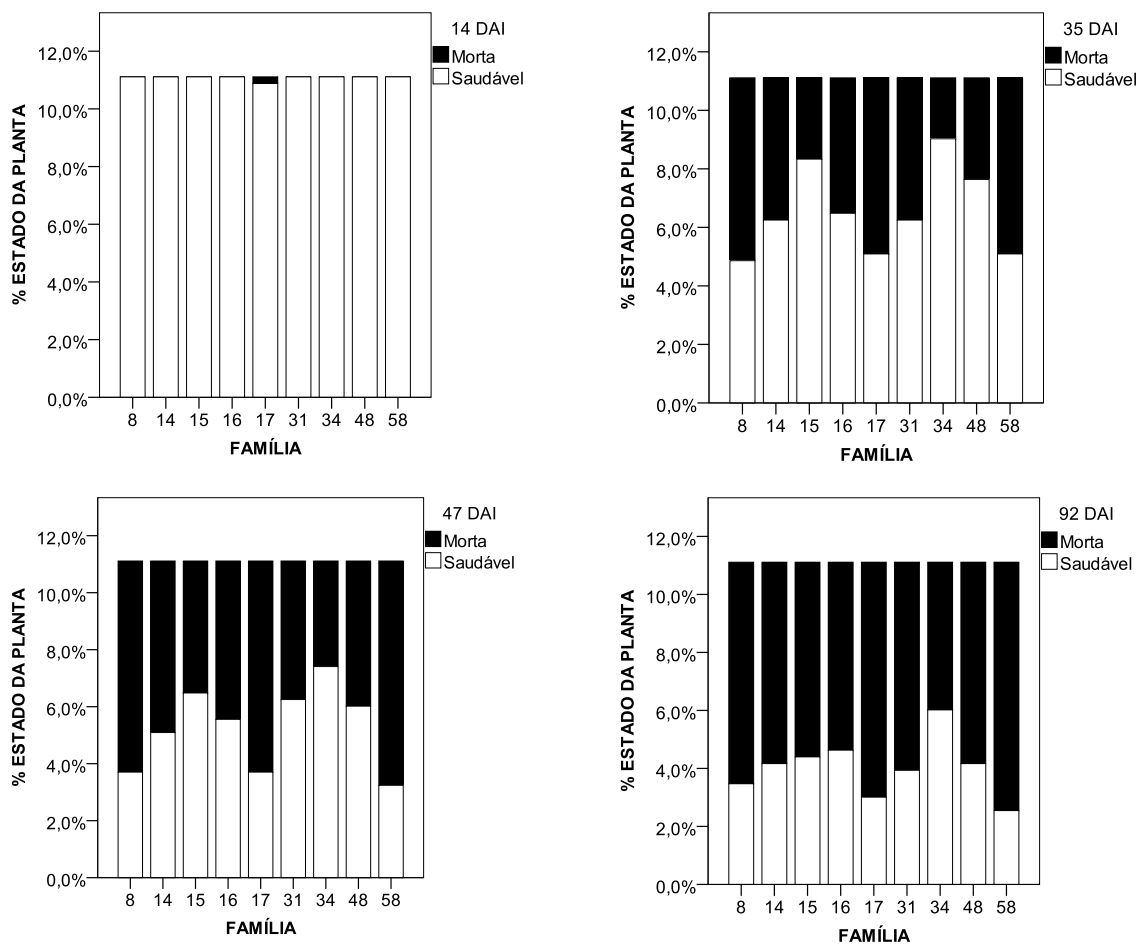


Figura 3.1 – Gráficos de barras para a variável FAMÍLIA relativamente aos níveis de ESTADO DA PLANTA, aos 14, 35, 47 e 92 dias após inoculação (DAI).

Relativamente à variável CONCENTRAÇÃO, o quadro 3.6 indica que, aos 14 DAI, a resposta é constante, sendo a maioria das plantas saudáveis (99,8%).

A partir dos 35 DAI, a tendência é para que a inoculação com 500 NMP/ml (C) seja a que causa maior mortalidade (35 DAI: 66,7%; 47 DAI: 79,6%), seguida da inoculação com 750 NMP/ml (D) (35 DAI: 50,9%; 47 DAI: 66,7%). Por fim, aos 92 DAI, as mortalidades são mais elevadas para a inoculação de 500 NMP/ml (86,1%).

Quadro 3.6 – Contagem do número de plantas por CONCENTRAÇÃO para os níveis de ESTADO DA PLANTA, aos 14, 35, 47 e 92 dias após inoculação (DAI), e respectivas percentagens.

ESTADO DA PLANTA									
CON	14 DAI		35 DAI		47 DAI		92 DAI		Total
	Saudável	Morta	Saudável	Morta	Saudável	Morta	Saudável	Morta	
A	107 (99,1%)	1 (0,9%)	107 (99,1%)	1 (0,9%)	107 (99,1%)	1 (0,9%)	100 (92,6%)	8 (7,4%)	108
B	108 (100%)	0 (0,0%)	59 (54,6%)	49 (45,4%)	40 (37,0%)	68 (63,0%)	22 (20,4%)	86 (79,6%)	108
C	108 (100%)	0 (0,0%)	36 (33,3%)	72 (66,7%)	22 (20,4%)	86 (79,6%)	15 (13,9%)	93 (86,1%)	108
D	108 (100%)	0 (0,0%)	53 (49,1%)	55 (50,9%)	36 (33,3%)	72 (66,7%)	20 (18,5%)	88 (81,5%)	108
Total	431 (99,8%)	1 (0,2%)	255 (59,0%)	177 (41,0%)	205 (47,4%)	227 (52,6%)	157 (36,3%)	275 (63,7%)	432

Os gráficos seguintes apresentam a distribuição percentual de plantas mortas e saudáveis para as quatro datas de observação selecionadas (figura 3.2).

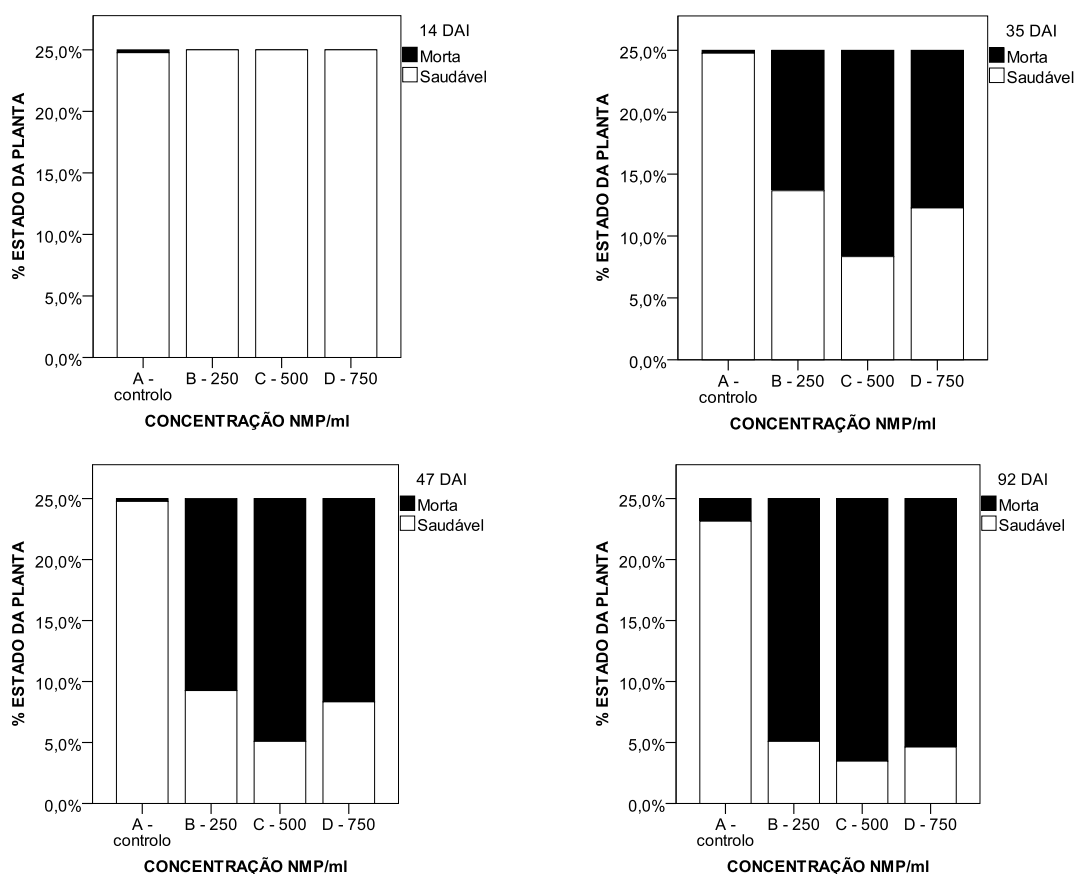


Figura 3.2 – Gráficos de barras para a variável CONCENTRAÇÃO relativamente aos níveis de ESTADO DA PLANTA, aos 14, 35, 47 e 92 dias após inoculação (DAI).

A análise dos gráficos permite verificar que a concentração 500 NMP/ml (C) é a que causa maior mortalidade aos 35, 47 e 92 DAI, seguida da concentração 750 NMP/ml e por fim da concentração 250 NMP/ml. Aos 14 DAI não se detetam diferenças.

VARIÁVEL SINTOMATOLOGIA

De seguida foi efetuada a análise exploratória, para as quatro datas selecionadas, da variável resposta ordinal, SINTOMATOLOGIA, tendo em conta cada uma das variáveis independentes, FAMILIA e CONCENTRAÇÃO.

No quadro 3.7, verifica-se que aos 14 DAI a resposta observada para a variável FAMILIA é de um modo geral semelhante em todas as famílias, apresentando uma maioria de sintomas nível 1 (até 25% da planta sintomática).

Quadro 3.7 – Contagem do número de plantas por FAMILIA para os níveis de SINTOMATOLOGIA (SIN, 1 a 4), aos 14, 35, 47 e 92 dias após inoculação (DAI), e respetivas percentagens.

	SIN	FAMILIA									
		8	14	15	16	17	31	34	48	58	Total
14 DAI	1	48 (100%)	48 (100%)	48 (100%)	47 (97,9%)	47 (97,9%)	46 (95,8%)	47 (97,9%)	46 (95,8%)	48 (100%)	425 (98,4%)
	2	0 (0,0%)	0 (0,0%)	0 (0,0%)	1 (2,1%)	0 (0,0%)	2 (4,2%)	1 (2,1%)	2 (4,2%)	0 (0,0%)	6 (1,4%)
	3	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)
	4	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	1 (2,1%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	1 (0,2%)
35 DAI	1	16 (33,3%)	23 (47,9%)	31 (64,6%)	26 (54,2%)	17 (35,4%)	27 (56,2%)	33 (68,8%)	31 (64,6%)	17 (35,4%)	221 (51,2%)
	2	5 (10,4%)	4 (8,3%)	5 (10,4%)	2 (4,2%)	5 (10,4%)	0 (0,0%)	6 (12,5%)	2 (4,2%)	5 (10,4%)	34 (7,9%)
	3	6 (12,5%)	8 (16,7%)	4 (8,3%)	1 (2,1%)	6 (12,5%)	2 (4,2%)	4 (8,3%)	3 (6,2%)	5 (10,4%)	39 (9,0%)
	4	21 (43,8%)	13 (27,1%)	8 (16,7%)	19 (39,6%)	20 (41,7%)	19 (39,6%)	5 (10,4%)	12 (25,0%)	21 (43,8%)	138 (31,9%)
47 DAI	1	16 (33,3%)	22 (45,8%)	27 (56,2%)	23 (47,9%)	14 (29,2%)	25 (52,1%)	28 (58,3%)	22 (45,8%)	12 (25,0%)	189 (43,8%)
	2	0 (0,0%)	8 (16,7%)	1 (2,1%)	1 (2,1%)	2 (4,2%)	2 (4,2%)	4 (8,3%)	4 (8,3%)	2 (4,2%)	16 (3,7%)
	3	1 (2,1%)	1 (2,1%)	3 (6,2%)	2 (4,2%)	2 (4,2%)	1 (2,1%)	4 (8,3%)	3 (6,2%)	4 (8,3%)	21 (4,9%)
	4	31 (64,6%)	25 (52,1%)	17 (35,4%)	22 (45,8%)	30 (62,5%)	20 (41,7%)	12 (25,0%)	19 (39,6%)	30 (62,5%)	206 (47,7%)
92 DAI	1	11 (22,9%)	17 (35,4%)	15 (31,2%)	14 (29,2%)	12 (25,0%)	16 (33,3%)	23 (47,9%)	10 (20,8%)	9 (18,8%)	127 (29,4%)
	2	4 (8,3%)	1 (2,1%)	4 (8,3%)	6 (12,5%)	1 (2,1%)	1 (2,1%)	3 (6,2%)	8 (16,7%)	2 (4,2%)	30 (6,9%)
	3	0 (0,0%)	0 (0,0%)	6 (12,5%)	1 (2,1%)	1 (2,1%)	6 (12,5%)	0 (0,0%)	1 (2,1%)	0 (0,0%)	15 (3,5%)
	4	33 (68,8%)	30 (62,5%)	23 (47,9%)	27 (56,2%)	34 (70,8%)	25 (52,1%)	22 (45,8%)	29 (60,4%)	37 (77,1%)	260 (60,2%)
Total	48	48	48	48	48	48	48	48	48	432	

A partir dos 35 DAI, é possível verificar uma variação na resposta das famílias. A família 34 é a que apresenta um maior número de plantas com SINTOMATOLOGIA mais baixa (nível

1), mantendo-se esta tendência até aos 92 DAI (47,9%). Por outro lado, as famílias 8, 17 e 58 apresentam aos 35 DAI uma maior percentagem de respostas de nível 3 e 4 de SINTOMATOLOGIA. A família 58 destaca-se pela elevada percentagem de observações no nível mais severo de sintomas da doença, ou seja nível 4 (mais de 76% da planta sintomática) aos 92 DAI (77,1%).

As representações gráficas da distribuição percentual de plantas mortas e saudáveis por data de observação encontram-se na figura 3.3.

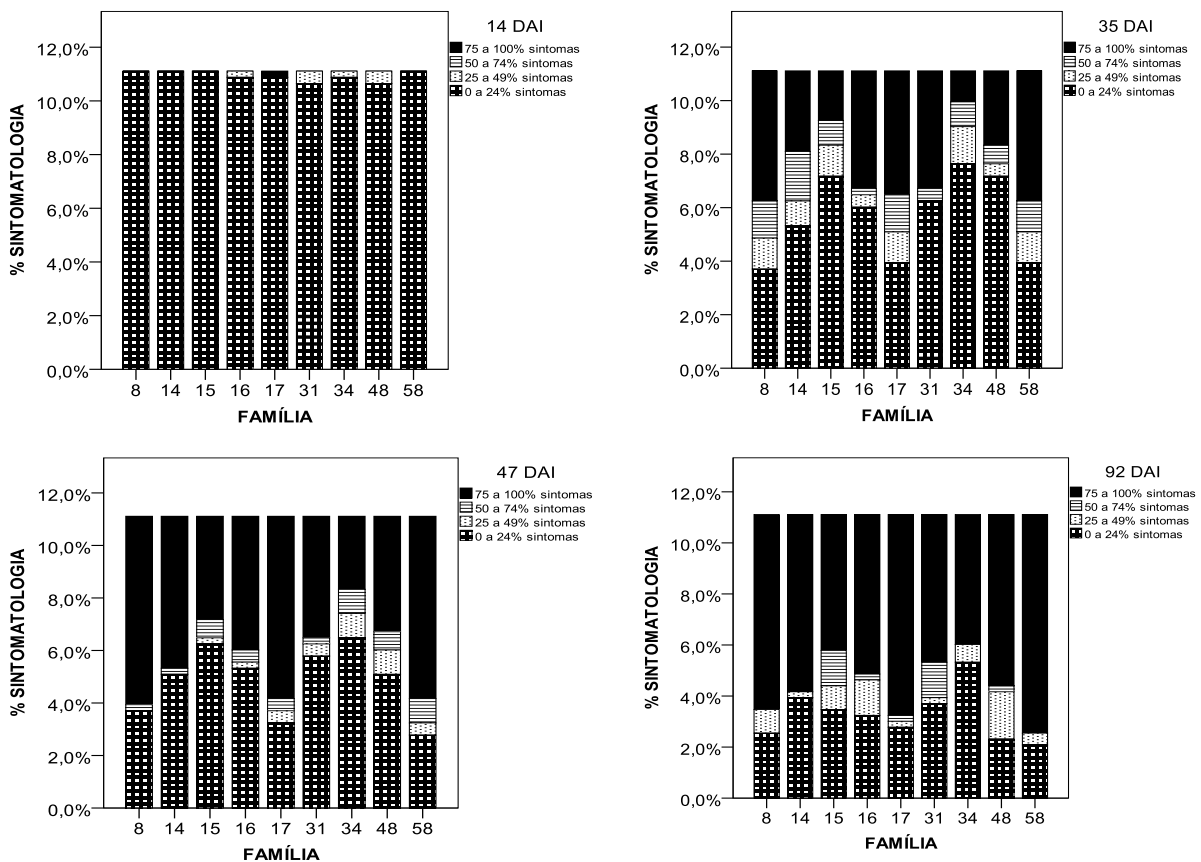


Figura 3.3 – Gráficos de barras para a variável FAMÍLIA relativamente aos níveis de SINTOMATOLOGIA, aos 14, 35, 47 e 92 dias após inoculação (DAI).

Como se pode constatar pelos gráficos, para os 35, 47 e 92 DAI, a família 34 é aquela que apresenta uma maior percentagem de observações relativas ao menor nível de sintomatologia (nível 1). Por outro lado, a família 8 é a que apresenta maior percentagem de observações nível 4 (nível mais alto de sintomas) aos 47 DAI, sendo substituída nesta posição aos 92 DAI, pela família 58. Aos 14 DAI, não se descortinam diferenças relevantes.

Relativamente à variável CONCENTRAÇÃO, de acordo com o quadro 3.8, aos 14 DAI as respostas das plantas mantêm-se na sua quase totalidade até 25% do total da planta sintomática (nível de 1 de SINTOMATOLOGIA).

Quadro 3.8 – Contagem do número de plantas por CONCENTRAÇÃO para os níveis de SINTOMATOLOGIA (SIN, 1 a 4), aos 14, 35, 47 e 92 dias após inoculação (DAI), e respetivas percentagens.

	SIN	CONCENTRAÇÃO				Total
		A	B	C	D	
14 DAI	1	107 (99,1%)	105 (97,2%)	106 (98,1%)	107 (99,1%)	425 (98,4%)
	2	0 (0,0%)	3 (2,8%)	2 (1,9%)	1 (0,9%)	6 (1,4%)
	3	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	0 (0,0%)
	4	1 (0,9%)	0 (0,0%)	0 (0,0%)	0 (0,0%)	1 (0,2%)
35 DAI	1	106 (98,1%)	45 (41,7%)	30 (27,8%)	40 (37,0%)	221 (51,2%)
	2	1 (0,9%)	14 (13,0%)	6 (5,6%)	13 (12,0%)	34 (7,9%)
	3	0 (0,0%)	8 (7,4%)	13 (12,0%)	18 (16,7%)	39 (9,0%)
	4	1 (0,9%)	41 (38,0%)	59 (54,6%)	37 (34,3%)	138 (31,9%)
47 DAI	1	105 (97,2%)	36 (33,3%)	17 (15,7%)	31 (28,7%)	189 (43,8%)
	2	2 (1,9%)	4 (3,7%)	5 (4,6%)	5 (4,6%)	16 (3,7%)
	3	0 (0,0%)	8 (7,4%)	3 (2,8%)	10 (9,3%)	21 (4,9%)
	4	1 (0,9%)	60 (55,6%)	83 (76,9%)	62 (57,4%)	206 (47,7%)
92 DAI	1	84 (77,8%)	19 (17,6%)	9 (8,3%)	15 (13,9%)	127 (29,4%)
	2	16 (14,8%)	3 (2,8%)	6 (5,6%)	5 (4,6%)	30 (6,9%)
	3	0 (0,0%)	9 (8,3%)	2 (1,9%)	4 (3,7%)	15 (3,5%)
	4	8 (7,4%)	77 (71,3%)	91 (84,3%)	84 (77,8%)	260 (60,2%)
Total		108	108	108	108	432

A partir dos 35 DAI a tendência é para que a inoculação com 500 NMP/ml (C) seja a que conduz a maior percentagem de plantas nível 4 (35 DAI: 54,6%; 47 DAI: 76,9%; 92 DAI: 84,3%).

Os gráficos seguintes mostram a distribuição percentual de plantas mortas e saudáveis por data de observação (figura 3.4).

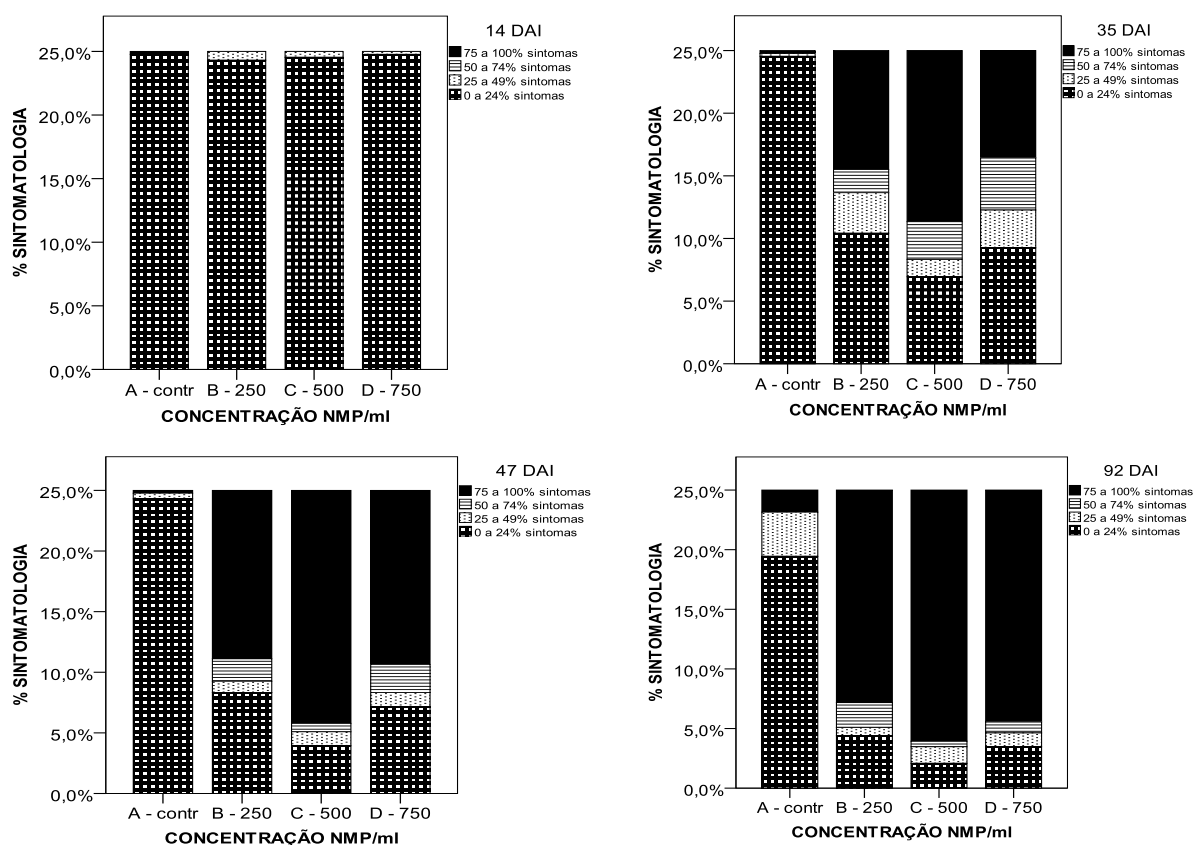


Figura 3.4 – Gráficos de barras para a variável CONCENTRAÇÃO relativamente aos níveis de SINTOMATOLOGIA, aos 14, 35, 47 e 92 dias após inoculação (DAI).

Estes gráficos permitem constatar que aos 14 DAI não ocorrem diferenças visíveis entre as diferentes concentrações. É possível também verificar que a maior percentagem de observações do nível mais baixo de sintomas (nível 1), é sempre apresentado pela concentração B (250 NMP/ml), seguida da concentração 750 NMP/ml e da concentração 500 NMP/ml. De realçar o facto de que aos 35 DAI, apesar de uma maior percentagem de plantas nível 1 na concentração 250 NMP/ml relativamente à concentração 750 NMP/ml, esta concentração também apresenta uma maior percentagem de plantas com sintomas nível 4.

COVARIÁVEIS

De seguida apresentam-se os resultados das análises exploratórias às covariáveis ALTURA e DIÂMETRO.

No quadro 3.9 é possível verificar que a média de ALTURA das plantas usadas no ensaio é de 32,53 cm, o desvio padrão é de 7,10 cm e a mediana é 32,20 cm. É possível verificar também que a amplitude entre o valor mínimo e máximo desta variável é de 34,50 cm, com

um mínimo de 16,50 cm e um máximo de 51,00 cm. A variável DIÂMETRO, apresenta um valor médio de 0,32 cm, o desvio padrão é de 0,09 cm e uma mediana de 0,31 cm. A amplitude do DIÂMETRO das plantas usadas no ensaio é de 0,51 cm, com um mínimo de 0,08 cm e um máximo de 0,58 cm.

Através do coeficiente de variação, verifica-se que a dispersão relativa dos valores da variável DIÂMETRO é superior relativamente à distribuição dos valores da variável ALTURA, tomando os valores de 27,85% e 21,83% respectivamente. De acordo com Pestana e Gageiro (2008), considera-se que os coeficientes apresentam uma dispersão média, logo as médias são representativas.

Relativamente ao estudo da assimetria das variáveis, esta é avaliada pelo quociente entre o coeficiente de enviesamento e o seu erro padrão, verificando-se uma simetria dos dados para ambas as variáveis, dado que os valores obtidos para a ALTURA e para o DIÂMETRO são respectivamente 0,08 e 1,68, e se encontram compreendidos entre -1,96 e 1,96 ($\alpha=0,05$).

Por fim, analisando o achatamento, verifica-se que para a variável ALTURA, o valor do quociente entre o coeficiente de curtose pelo seu erro padrão é de -3,04 (inferior a -1,96, $\alpha=0,05$), logo a distribuição diz-se platicúrtica, o que significa que é mais achatada que a distribuição normal. Já para o DIÂMETRO, sendo o valor do quociente de -0,380 (inferior a 1,96 e superior a -1,96, $p=0,05$), considera-se esta variável como mesocúrtica. Deste modo, visto que para esta variável a assimetria é moderada, pode utilizar-se a média aritmética para interpretar o DIÂMETRO das plantas inoculadas.

Quadro 3.9 – Análise descritiva para as variáveis ALTURA e DIÂMETRO

	Altura (cm)	Diâmetro (cm)
Média Aritmética	32,53	0,32
Erro Padrão	0,34	0,00
Mediana	32,20	0,31
Moda	27,00 ^a	0,30 ^a
Desvio Padrão	7,10	0,09
Coeficiente de Variação (%)	21,83	27,85
Variância	50,45	0,01
Enviesamento	0,01	0,20
Erro Padrão do Enviesamento	0,12	0,12
Curtose	-0,70	-0,09
Erro Padrão da Curtose	0,23	0,23
Amplitude	34,50	0,51
Mínimo	16,50	0,08
Máximo	51,00	0,58
Percentis	25	27,12
	50	32,20
	75	37,68

^a. Existe moda múltipla. O menor valor é mostrado.

Os resultados do teste de Kolmogorov-Smirnov para a variável ALTURA relativamente à sua distribuição pela variável FAMÍLIA, suporta a rejeição da hipótese nula para $\alpha=0,05$, ou seja da existência de uma distribuição normal ($KS(48)_{família14}=0,13$; $p=0,033$). Contudo tal não se verifica para a variável ALTURA relativamente à CONCENTRAÇÃO, dado que não ocorre rejeição da hipótese nula para nenhuma classe desta variável. Por outro lado, com o teste de Levene é rejeitada a hipótese nula da homogeneidade das variâncias para a variável FAMÍLIA para $\alpha=0,05$ ($F(8,423)=2,69$; $p=0,007$), não a rejeitando para a variável CONCENTRAÇÃO ($F(3,428)=2,06$; $p=0,105$). Esta violação dos pressupostos de normalidade e de homogeneidade das variâncias impede a utilização do teste ANOVA a um factor para testar se as variáveis ALTURA e DIÂMETRO se encontram uniformemente distribuídas pelas variáveis independentes em estudo (teste à unidade experimental). Como tal, utilizou-se o teste não-paramétrico de Kruskal-Wallis (quadro 3.10) considerando como H_0 que a distribuição dos valores das covariáveis, ALTURA e DIÂMETRO, é idêntica nos diferentes níveis das variáveis independentes selecionadas, CONCENTRAÇÃO e FAMÍLIA (medianas iguais), para uma probabilidade de erro tipo I (α) de 0,05, e consideradas separadamente na análise.

Quadro 3.10 – Teste de Kruskal-Wallis à uniformidade de distribuição das alturas e diâmetros das plantas pelas diferentes variáveis independentes CONCENTRAÇÃO e FAMÍLIA.

	FAMÍLIA		CONCENTRAÇÃO	
	Altura	Diâmetro	Altura	Diâmetro
χ^2	94,00	19,72	6,13	22,06
<i>g.l.</i>	8	8	3	3
<i>p</i>	<0,001	0,011	0,105	<0,001

Para ambas as covariáveis, ALTURA e DIÂMETRO, a partir do teste de Kruskal-Wallis, são rejeitadas ambas as hipóteses nulas de que ocorre distribuição uniforme das plantas pelas diferentes famílias (ALTURA: $\chi^2_{KW}(8)=94,00$; $p<0.001$ e DIÂMETRO: $\chi^2_{KW}(8)=19,72$; $p=0.011$). Os resultados permitem ainda verificar que a hipótese de distribuição uniforme de diâmetros das plantas é rejeitada para as diferentes concentrações de nemátode inoculadas ($\chi^2_{KW}(3)=22,06$; $p<0.001$), mas não existem evidências para rejeitar a uniformidade da altura das plantas ($\chi^2_{KW}(3)=6,13$; $p=0.105$).

De seguida, apresentam-se, na figura 3.5., as caixas de bigodes das distribuições das alturas e diâmetros das plantas pelas diferentes famílias e concentrações.

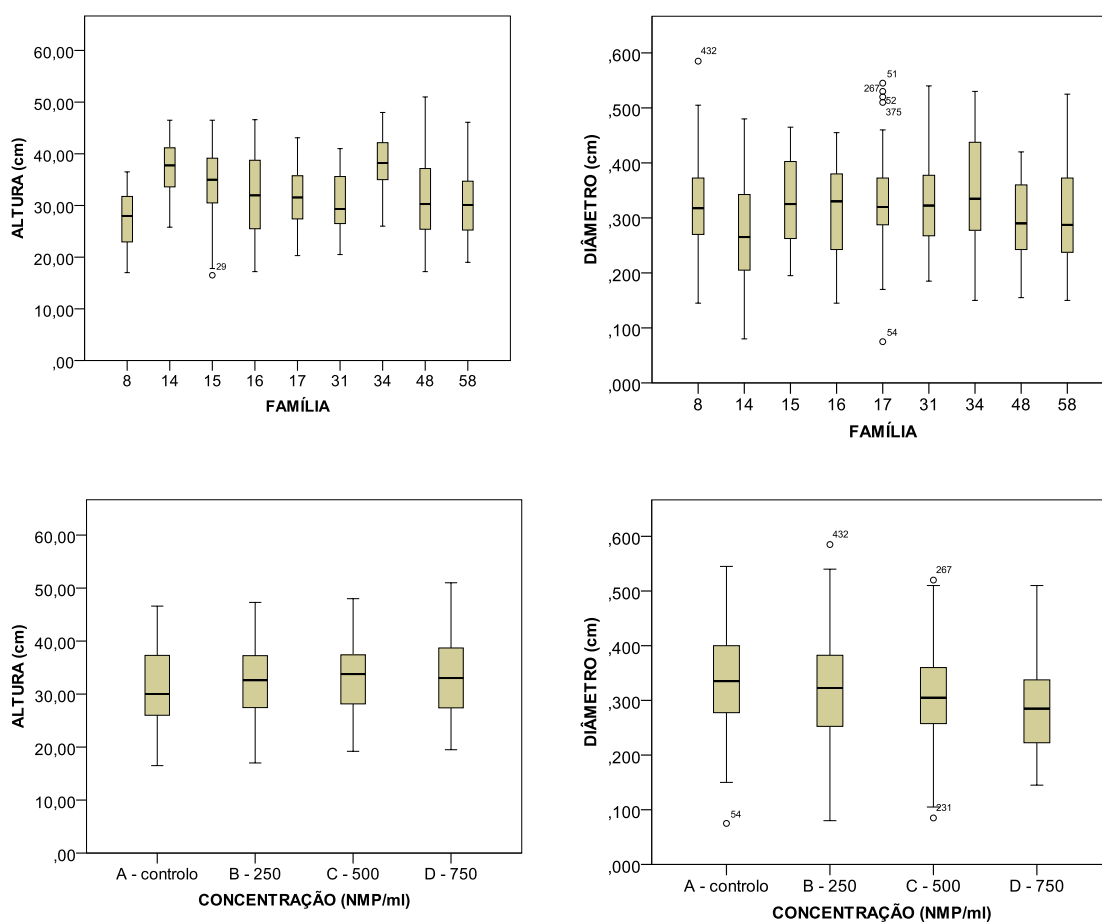


Figura 3.5 – Caixas de bigodes das distribuições das ALTURAS e DIÂMETROS das plantas por FAMÍLIA e por CONCENTRAÇÃO.

Estes resultados sugerem que existem evidências que permitem concluir que as alturas das plantas estão distribuídas uniformemente pelas diferentes concentrações, mas não os diâmetros. No que respeita às famílias, as alturas e diâmetros das plantas não estão distribuídas uniformemente.

3.3. Ajustamento de modelos de regressão categorial

Os resultados dos ajustamentos dos modelos de regressão categorial, ou seja, modelos de regressão logística, ordinal e multinomial, para cada uma das quatro datas de observação, são apresentados de seguida.

3.3.1. Regressão logística

Os resultados da regressão logística da interação entre as variáveis explicativas contínuas, ALTURA e DIÂMETRO, e o seu \ln , que permite a avaliação do pressuposto da existência

de uma relação linear entre estas variáveis e a transformação *Logit* da variável de resposta, ESTADO DA PLANTA, encontram-se no quadro 3.11.

Quadro 3.11 – Avaliação da linearidade das variáveis explicativas contínuas relativamente à variável dependente ESTADO DA PLANTA, para as 4 datas de observação após inoculação (DAI): 14, 35, 47 e 92, testando a interação entre a variável explicativa e o seu logaritmo neperiano.

	B	X^2_{Wald}	g.l.	p
DIAMETRO*Ln(DIÂMETRO)				
14 DAI	3245,37	0,004	1	0,947
35 DAI	-5,53	0,384	1	0,535
47 DAI	-12,15	2,024	1	0,155
92 DAI	-3,07	0,104	1	0,747
ALTURA*Ln(ALTURA)				
14 DAI	-2205,53	0,736	1	0,391
35 DAI	-0,28	5,139	1	0,023
47 DAI	-0,44	13,582	1	<0,001
92 DAI	-0,35	10,018	1	0,002

A linearidade entre a variável explicativa ALTURA e a transformação logística da variável resposta é violada, para as datas de observação 35, 47 e 92 DAI, dado que usando o teste de Wald (X^2_{Wald}) se rejeitar H_0 (35 DAI: $X^2_{Wald}(1)=-0,28$; $p=0,023$; 47 DAI: $X^2_{Wald}(1)=-0,44$; $p<0,001$; 92 DAI: $X^2_{Wald}(1)=-0,35$; $p=0,002$). Relativamente à variável DIÂMETRO, não se rejeita a hipótese nula para todas as datas de observação, concluindo-se que a relação entre a variável DIÂMETRO e a variável resposta é linear.

Para o estudo da variável ALTURA no modelo final procedeu-se assim à sua transformação para ordinal, considerando quatro níveis obtidos a partir dos quartis.

Os resultados da regressão logística *Forward:LR* revelaram que as interações não eram passíveis de ser usadas nos modelos, logo as mesmas não foram consideradas para nenhuma das datas de observação.

14 DAI

Para esta data, usando os testes do rácio de verosimilhança (G^2), de Hosmer & Lemeshow (X^2_{HL}), verifica-se que de acordo com o *Logit* ajustado, tanto a CONCENTRAÇÃO ($G^2(3)=2,78$; $p=0,427$; $X^2_{HL}(1)=0,00$; $p=1,000$), como a FAMÍLIA ($G^2(8)=4,41$; $p=0,818$; $X^2_{HL}(1)=0,00$; $p=1,000$), como a ALTURA ($G^2(1)=2,76$; $p=0,097$; $X^2_{HL}(2)=0,00$; $p=1,000$), apresentam testes de rácio de verosimilhanças que não permitem concluir que para esta data nenhuma das variáveis independentes adicionadas ao modelo tem poder preditivo sobre a variável ESTADO DA PLANTA. A covariável DIÂMETRO, apesar de não rejeitar a

hipótese nula do ajustamento do modelo ($G^2(1)=14,13$; $p<0,001$; $X^2_{HL}(3)=0,00$; $p=1,000$) para esta data, não apresenta um efeito estatisticamente significativo sobre o *Logit* da probabilidade de morte das plantas ($X^2_{Wald}(1)=0,02$; $p=0,881$), para um coeficiente β , $b_{diâmetro}=-3530,32$.

35 DAI

O modelo ajustado para esta data de observação é estatisticamente significativo ($G^2(1)=65,58$; $p<0,001$; $X^2_{HL}(8)=6,53$; $p=0,589$) e a presença de cinco observações extremas (plantas: 136, 175, 176, 269 e 309) não afectam as significâncias do ajustamento e dos coeficientes do modelo final (anexo I).

O modelo simplificado classifica 82,9% dos dados correctamente, valor consideravelmente superior à percentagem proporcional de classificações correctas por acaso (51,63%). A especificidade é 0,808, ou seja, o modelo classifica correctamente 80,8% dos indivíduos que morreram, e a sensibilidade é de 0,843, ou seja, o modelo classifica correctamente 84,3% dos indivíduos que sobreviveram. O modelo tem ainda uma capacidade discriminante excelente (ROC $c=0,920$; $p<0,001$) (figura 3.6).

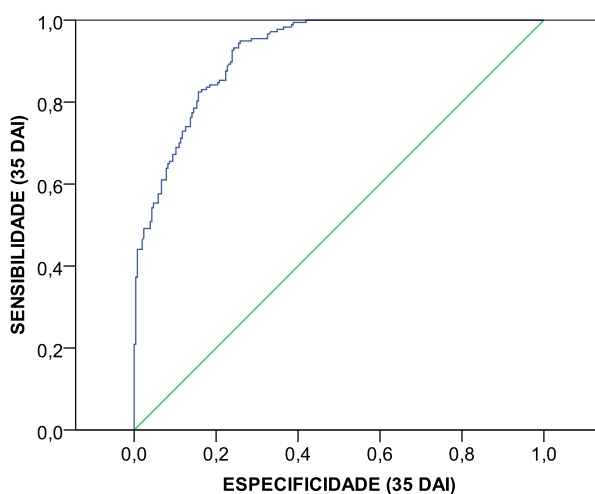


Figura 3.6 – Curva ROC da relação entre a sensibilidade e a especificidade aos 35 DAI, para a variável dependente ESTADO DA PLANTA.

As estimativas dos coeficientes do modelo para as variáveis FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO são dadas no quadro 3.12. A regressão logística revelou que as variáveis FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO apresentaram um efeito estatisticamente significativo sobre o *Logit* da probabilidade de ocorrer morte das plantas, de acordo com o modelo *Logit* ajustado. Pelo contrário, as famílias: 8, 15, 16, 17 e 48 não

apresentam um efeito estatisticamente significativo relativamente à classe de referência (família 58), para um $\alpha=0,05$. Quando considerado um nível de significância $\alpha=0,10$, apenas as famílias 16, 17 e 48 não apresentam um efeito estatisticamente significativo. A concentração B, relativamente à classe de referência (concentração D), também não apresenta efeito estatisticamente significativo, sobre o Logit da probabilidade de ocorrer morte das plantas. Por fim, para $\alpha=0,001$, apenas as variáveis CONCENTRAÇÃO, ALTURA e DIÂMETRO, assim como a concentração controlo e C, apresentam um efeito estatisticamente significativo sobre o *Logit* da probabilidade de ocorrer morte da planta.

Quadro 3.12 – Coeficientes Logit do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 35 DAI.

Variável	B	S.E.	X ² _{Wald}	g.l.	p-value	OR	I.C. a 95% para Exp(B)
FAMÍLIA			20,32	8	0,009		
8	-1,25	0,73	2,92	1	0,088	0,29]0,07; 1,20 [
14	-1,52	0,67	5,08	1	0,024	0,22]0,06; 0,82 [
15	-1,09	0,66	2,73	1	0,099	0,34]0,09; 1,23 [
16	-0,16	0,64	0,06	1	0,801	0,85]0,24; 3,01 [
17	-0,98	0,67	2,16	1	0,142	0,38]0,10; 1,39 [
31	-1,55	0,71	4,75	1	0,029	0,21]0,05; 0,86 [
34	-2,30	0,66	11,98	1	0,001	0,10]0,03; 0,37 [
48	-0,50	0,66	0,56	1	0,453	0,61]0,16; 2,23 [
CONCENTRAÇÃO			41,61	3	<0,001		
A – controlo	-5,91	1,17	25,51	1	<0,001	0,00]0,00; 0,03 [
B – 250 NMP/ml	0,16	0,37	0,20	1	0,655	1,18]0,57; 2,42 [
C – 500 NMP/ml	1,32	0,38	12,34	1	<0,001	3,74]1,79; 7,80 [
ALTURA	-0,66	0,16	16,09	1	<0,001	0,52]0,38; 0,71 [
DIÂMETRO	-16,18	2,39	46,01	1	<0,001	0,00]0,00; 0,00 [
Constante	7,43	0,98	57,53	1	<0,001	1690,26	

O modelo final para os 35 DAI que permite estimar a probabilidade ($\hat{\pi}$) de ocorrer morte nas plantas é então, de acordo com o quadro 3.12:

$$\hat{\pi} = \frac{1}{1 + e^{\left[-7,43 - 1,25FAM(8) - 1,52FAM(14) - 1,09FAM(15) - 0,16FAM(16) - 0,98FAM(17) - 1,55FAM(31) - 2,30FAM(34) - 0,50FAM(48) - 5,91CONC(A) + 0,16CONC(B) + 1,32CONC(C) - 0,66ALT - 16,18DIAM \right]}}$$

De acordo com este modelo, o rácio de chances de ocorrer morte ($Y=1$) é maior para a família 58 (a família de referência), sendo o seu valor 15% superior ao da família 16, a segunda a causar maior mortalidade. Por outras palavras, as chances de ocorrer morte para a família 58 relativamente à família 16 são de 1:0,85, tendo as restantes famílias um valor de chance ainda inferior a este. Por outro lado, o rácio de chances de ocorrer morte é menor

para a família 34, relativamente à família de referência, sendo o seu valor 90% inferior, ou seja, tem uma chance de morte de 0,1:1.

Quanto à variável CONCENTRAÇÃO, o rácio de chances de ocorrer morte é maior para a concentração B relativamente à concentração de referência (D), sendo o seu valor 17,9% superior, e tendo chances de 1,18:1, sendo no entanto ainda maior para a concentração C, dado o aumento de 273,9% nas chances relativamente à concentração D, 3,74:1.

Relativamente às variáveis contínuas, o rácio de chances de ocorrer mortes diminui 48,3% por cm de ALTURA, e diminui 100% por cada cm de DIÂMETRO.

47 DAI

Para esta data, o modelo ajustado é estatisticamente significativo ($G^2(1)=84,23$; $p<0,001$; $X^2_{HL}(8)=5,32$; $p=0,723$), com a presença de 11 observações extremas (plantas: 130, 153, 175, 176, 190, 193, 218, 297, 374, 379, e 422) que não afectam o ajustamento e a significância dos coeficientes do modelo final, conforme se pode constatar no anexo I.

Quanto às classificações, o modelo simplificado classifica bem 89,1% dos dados, valor superior à percentagem proporcional de classificações correctas por acaso (50,13%). A especificidade do modelo é de 93,4% dos indivíduos que morreram, e a sensibilidade é de 84,4% dos indivíduos que sobreviveram. O modelo final tem ainda uma capacidade discriminante excelente (ROC $c=0,952$; $p<0,001$) (figura 3.7).

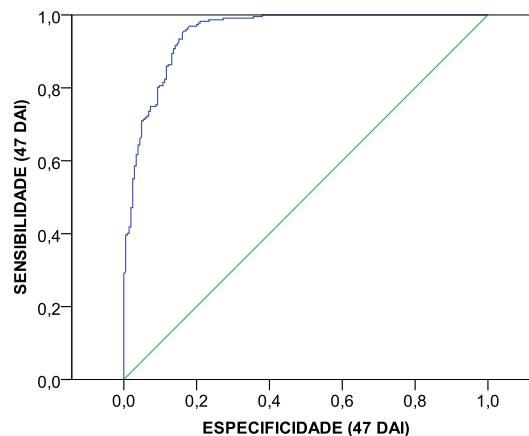


Figura 3.7 – Curva ROC da relação entre a sensibilidade e a especificidade aos 47 DAI, para a variável dependente ESTADO DA PLANTA.

As estimativas dos coeficientes do modelo para as variáveis FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO são dadas no quadro 3.13. Todas as variáveis presentes na regressão logística apresentaram um efeito estatisticamente significativo sobre o *Logit* da

probabilidade de ocorrer morte nas plantas, respectivamente: FAMÍLIA, CONCENTRAÇÃO, ALTURA, e DIÂMETRO. Pelo contrário, as famílias: 16 e 48, relativamente à classe de referência família 58, assim como a concentração B, relativamente à classe de referência concentração D, não apresentam efeito estatisticamente significativo para qualquer nível de significância do analista, sobre o *Logit* da probabilidade de ocorrer morte das plantas. Considerando um $\alpha=0,001$, apenas as variáveis CONCENTRAÇÃO, ALTURA e DIÂMETRO, a família 34 e a concentração controlo, apresentam efeito estatisticamente significativo.

Quadro 3.13 – Coeficientes Logit do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 47 DAI.

Variável	B	S.E.	X ² _{Wald}	g.l.	p-value	OR	I.C. a 95% para Exp(B)
FAMÍLIA			27,78	8	0,001		
8	-2,18	0,96	5,23	1	0,022	0,11]0,02; 0,73 [
14	-1,93	0,84	5,32	1	0,021	0,15]0,03; 0,75[
15	-2,07	0,85	5,91	1	0,015	0,13]0,02; 0,67 [
16	-0,38	0,94	0,16	1	0,686	0,68]0,11; 4,28[
17	-2,93	0,88	11,06	1	0,001	0,05]0,01; 0,30[
31	-1,81	0,87	4,36	1	0,037	0,16]0,03; 0,90[
34	-3,05	0,86	12,50	1	<0,001	0,05]0,01; 0,26[
48	0,12	0,99	0,02	1	0,900	1,13]0,16; 7,83 [
CONCENTRAÇÃO			47,86	3	<0,001		
A – controlo	-8,34	1,38	36,73	1	<0,001	0,00]0,00; 0,00 [
B – 250 NMP/ml	0,48	0,43	1,22	1	0,269	1,61]0,69; 3,76[
C – 500 NMP/ml	1,46	0,45	10,57	1	0,001	4,28]1,78; 10,30[
ALTURA	-0,75	0,20	14,74	1	<0,001	0,47]0,32; 0,69 [
DIÂMETRO	-20,80	2,92	50,90	1	<0,001	0,00]0,00; 0,00 [
Constante	10,84	1,41	59,41	1	<0,001	51220,44	

O modelo final para os 47 DAI que permite estimar a probabilidade ($\hat{\pi}$) de ocorrer morte nas plantas é então, de acordo com o quadro 3.13:

$$\hat{\pi} = \frac{1}{1 + e^{\left[-10,84 - 2,18FAM(8) - 1,93FAM(14) - 2,07FAM(15) - 0,38FAM(16) - 2,93FAM(17) - 1,81FAM(31) - 3,05FAM(34) + 0,12FAM(48) - 8,34CONC(A) + 0,48CONC(B) + 1,46CONC(C) - 0,75ALT - 20,80DIAM \right]}}$$

O quadro 3.13 permite ainda calcular o rácio de chances (OR) de ocorrer morte (Y=1), que é maior para a família 48, sendo o seu valor 13,2% superior ao da família 58 (família de referência). Por outras palavras, as chances de ocorrer morte para a família 48 relativamente à família 58 são de 1,132:1, tendo as restantes famílias um valor de chance ainda inferior a este. Por outro lado, o rácio de chances de ocorrer morte é menor para a

família 34, relativamente à família de referência, sendo o seu valor 95,3% inferior, ou tem uma chance de morte de 0,047:1.

Quanto à variável CONCENTRAÇÃO, o rácio de chances de ocorrer morte é maior para a concentração B relativamente à concentração de referência D, sendo o seu valor 61,1% superior, e tendo chances de 1,611:1, sendo no entanto ainda maior para a concentração C, dado o aumento de 328,4% nas chances relativamente à concentração D, 4,284:1. A concentração com menor rácio de chances de ocorrer morte é a concentração controlo, dado que tem 0,000:1, ou 0% de chances de ocorrer relativamente à concentração de referência.

Relativamente às variáveis contínuas, o rácio de chances de ocorrer mortes diminui 52,9% por cm de ALTURA, e diminui 100% por cada cm de DIÂMETRO.

92 DAI

Na análise aos 92 DAI, a regressão logística Forward: LR revelou que a variável ALTURA ($b_{\text{altura}}=-0,10$; $X^2_{\text{Wald}}(1)=0,29$; $p=0,589$; OR=0,90), não apresenta um efeito estatisticamente significativo sobre o *Logit* da probabilidade de ocorrer morte. Pelo contrário, as variáveis FAMÍLIA ($X^2_{\text{Wald}}(8)=15,81$; $p=0,045$), CONCENTRAÇÃO ($X^2_{\text{Wald}}(3)=88,78$; $p<0,001$) e DIÂMETRO ($b_{\text{diâmetro}}=-20,61$; $X^2_{\text{Wald}}(1)=57,42$; $p<0,001$; OR=0,00), apresentaram um efeito estatisticamente significativo sobre o *Logit* da probabilidade de ter um acidente de acordo com o modelo Logit ajustado ($G^2(1)=90,62$; $p<0,001$; $X^2_{\text{HL}}(8)=9,90$; $p=0,272$).

Assim, recorrendo ao mesmo método ajustou-se um novo modelo, estatisticamente significativo ($G^2(1)=104,24$; $p<0,001$; $X^2_{\text{HL}}(8)=9,75$; $p=0,283$), com apenas as variáveis FAMÍLIA, CONCENTRAÇÃO e DIÂMETRO. A presença de 12 observações extremas (plantas: 43, 85, 87, 95, 190, 218, 304, 349, 379, 392, 393 e 414) não afectam as significâncias do ajustamento e os coeficientes do modelo final, e foram mantidas na análise (anexo I).

O modelo simplificado classifica bem 89,4% dos dados, valor superior à percentagem proporcional de classificações correctas por acaso (53,73%). A especificidade do modelo é de 93,1% dos indivíduos que morreram, e a sensibilidade é de 82,8% dos indivíduos que sobreviveram. O modelo final tem ainda uma capacidade discriminante excelente (ROC $c=0,954$; $p<0,001$) (figura 3.8).

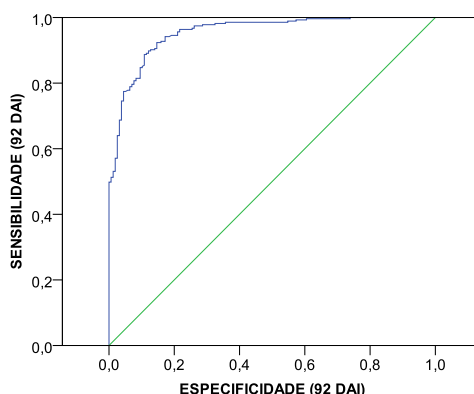


Figura 3.8 – Curva ROC da relação entre a sensibilidade e a especificidade aos 92 DAI, para a variável dependente ESTADO DA PLANTA.

As estimativas dos coeficientes do modelo para as variáveis FAMÍLIA, CONCENTRAÇÃO, e DIÂMETRO são dadas no quadro 3.14. Aos 92 DAI, todas as variáveis presentes na regressão logística apresentaram um efeito estatisticamente significativo sobre o *Logit* da probabilidade de ocorrer morte nas plantas, respectivamente: FAMÍLIA, CONCENTRAÇÃO, e DIÂMETRO. Pelo contrário, as famílias: 14, 16, 17 e 48, relativamente à classe de referência família 58, assim como a concentração B, e a concentração C, relativamente à classe de referência concentração D, não apresentam efeito estatisticamente significativo, para $\alpha=0,05$, sobre o Logit da probabilidade de ocorrer morte das plantas. Quando considerado um $\alpha=0,10$, ocorre rejeição da hipótese nula para a concentração B e C. Por fim, para um $\alpha=0,001$, ocorre rejeição da hipótese nula para as variáveis CONCENTRAÇÃO e DIÂMETRO, assim como para a concentração controle.

Quadro 3.14 – Coeficientes Logit do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, e DIÂMETRO, aos 92 DAI.

Variável	B	S.E.	X ² _{Wald}	g.l.	p-value	OR	I.C. a 95% para Exp(B)
FAMÍLIA			18,27	8	0,019		
8	-1,94	0,84	5,36	1	0,021	0,14]0,03; 0,74[
14	-1,02	0,72	1,98	1	0,159	0,36]0,09; 1,49[
15	-1,56	0,74	4,44	1	0,035	0,21]0,05; 0,90[
16	0,39	0,86	0,20	1	0,655	1,47]0,27; 7,98 [
17	-0,47	0,79	0,36	1	0,551	0,62]0,13; 2,94[
31	-1,55	0,71	4,76	1	0,029	0,21]0,05; 0,85[
34	-1,62	0,77	4,44	1	0,035	0,20]0,04; 0,89 [
48	0,56	0,88	0,40	1	0,529	1,74]0,31; 9,81 [
CONCENTRAÇÃO			90,69	3	<0,001		
A – controle	-5,10	0,62	68,55	1	<0,001	0,01]0,00; 0,02 [
B – 250 NMP/ml	0,79	0,47	2,87	1	0,090	2,21]0,88; 5,54 [
C – 500 NMP/ml	0,79	0,48	2,79	1	0,095	2,21]0,87; 5,61[
DIÂMETRO	-20,95	2,65	62,35	1	<0,001	0,00]0,00; 0,00[
Constante	9,20	1,16	63,11	1	<0,001	9862,47	

O modelo final para os 92 DAI que permite estimar a probabilidade ($\hat{\pi}$) de ocorrer morte nas plantas é então, de acordo com o quadro 3.14:

$$\hat{\pi} = \frac{1}{1 + e^{[-9,20 - 1,94FAM(8) - 1,02FAM(14) - 1,56FAM(15) + 0,39FAM(16) - 0,47FAM(17) - 1,55FAM(31) - 1,62FAM(34) + 0,56FAM(48) - 5,10CONC(A) + 0,79CONC(B) + 0,79CONC(C) - 20,95DIAM]}}$$

Quanto ao rácio de chances de ocorrer morte ($Y=1$), é maior para a família 48, sendo o seu valor 74,3% superior ao da família 58 (família de referência). Por outras palavras, as chances de ocorrer morte para a família 48 relativamente à família 58 são de 1,743:1, tendo as restantes famílias um valor de chance ainda inferior a este. Por outro lado, o rácio de chances de ocorrer morte é menor para a família 8, relativamente à família de referência, sendo o seu valor 86% inferior, ou tem uma chance de morte de 0,14:1.

Quanto à variável CONCENTRAÇÃO, o rácio de chances de ocorrer morte é maior para a concentração B e C de igual forma, relativamente à concentração de referência D, sendo o seu valor 121,2% superior, e tendo chances de 2,121:1. A concentração com menor rácio de chances de ocorrer morte é a concentração controlo, dado que tem 0,006:1, ou 0,6% de chances de ocorrer relativamente à concentração de referência.

Relativamente à variável contínua, o rácio de chances de ocorrer mortes diminui 100% por cada cm de DIÂMETRO.

3.3.2. Regressão ordinal

A avaliação do efeito das variáveis explicativas FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO sobre a probabilidade de resposta da variável ordinal SINTOMATOLOGIA, foi realizado nas quatro datas de observação recorrendo à regressão ordinal.

14 DAI

Através da observação do histograma de distribuição de frequências para os níveis da variável dependente SINTOMATOLOGIA para esta data (figura 3.9), seria de prever que a escolha da função de ligação recaísse sobre a função *Log-log* Negativo, no entanto, a escolhida foi a função *Cauchit*, por apresentar melhor significância estatística.

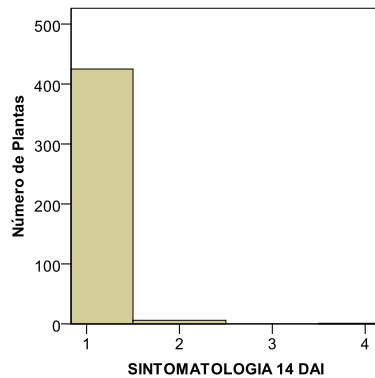


Figura 3.9 – Representação gráfica das frequências relativas aos diferentes níveis da variável resposta SINTOMATOLOGIA aos 14 dias após inoculação (DAI).

Embora o pressuposto do modelo da homogeneidade dos declives tenha sido validado, já que H_0 não foi rejeitada ($\chi^2_{LP}(13)=14,22$; $p=0,359$), o ajustamento do modelo não foi estatisticamente significativo ($G^2(13)=18,31$; $p=0,146$).

35 DAI

De acordo com Maroco (2007), para esta data, dada a distribuição de frequências da variável resposta, a função de ligação aconselhada seria a *Log-log* Negativo (figura 3.10).

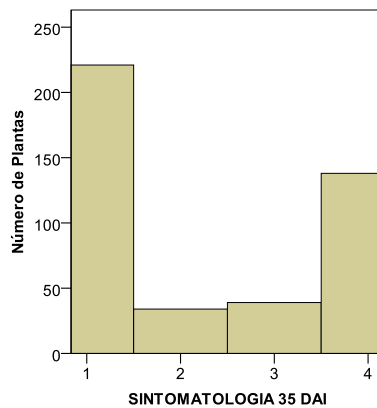


Figura 3.10 – Representação gráfica das frequências relativas aos diferentes níveis da variável resposta SINTOMATOLOGIA aos 35 dias após inoculação (DAI).

No entanto, apesar de todas as funções de ligação terem sido testadas, nenhuma validou o pressuposto do modelo da homogeneidade dos declives. Assim, para esta data ajustou-se uma regressão multinomial com as mesmas variáveis em estudo (sub-capítulo 3.3.3).

47 DAI

Para esta data a função de ligação considerada seria a *Cauchit* (figura 3.11), no entanto foi escolhida a *Logit*, dado que foi a que apresentou melhor significância estatística.

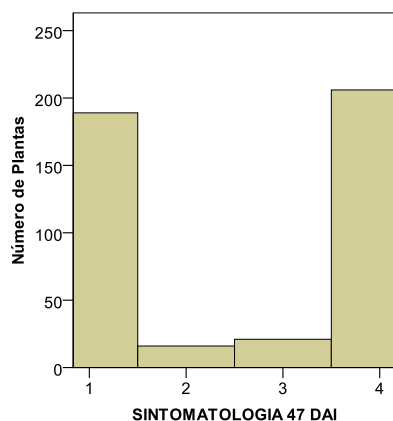


Figura 3.11 – Representação gráfica das frequências relativas aos diferentes níveis da variável resposta SINTOMATOLOGIA aos 47 dias após inoculação (DAI).

Através da regressão ordinal, verifica-se que o pressuposto do modelo da homogeneidade dos declives foi validado ($\chi^2_{LP}(26)=13,66$; $p=0,977$), já que não existem evidências para se rejeitar H_0 . O modelo foi considerado estatisticamente significativo ($G^2(13)=408,60$; $p<0,001$).

Os coeficientes e a significância do modelo ordinal ajustado estão descritos no quadro 3.15.

Quadro 3.15 - Coeficientes *Logit* do modelo de regressão ordinal dos diferentes níveis da variável SINTOMATOLOGIA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 47 DAI.

Parâmetros		B	S.E.	χ^2_{Wald}	g.l.	p-value	I.C. a 95% para Exp(B)
Pontos de corte	Nível 1	-10,10	1,05	93,36	1	<0,001]-12,15; -8,05[
	Nível 2	-9,66	1,03	88,19	1	<0,001]-11,67; -7,64[
	Nível 3	-9,12	1,01	81,68	1	<0,001]-11,10; -7,14[
Localização	Diâmetro	-18,38	2,27	65,57	1	<0,001]-22,82; -13,97[
	Altura	-0,82	0,17	24,32	1	<0,001]-1,14; -0,49[
	Família						
	8	0,27	0,72	0,14	1	0,710]-1,15; 1,69[
	14	-1,46	0,70	4,36	1	0,037]-2,83; -0,09[
	15	-1,44	0,62	5,33	1	0,021]-2,66; -0,22[
	16	-1,67	0,65	6,66	1	0,010]-2,94; -0,40[
	17	0,00	0,70	0,00	1	0,994]-1,36; 1,37[
	31	-2,11	0,67	9,96	1	0,002]-3,42; -0,80[
	34	-1,07	0,64	2,75	1	0,097]-2,33; 0,20[
	48	-2,28	0,65	12,40	1	<0,001]-3,55; -1,01[
	Concentração						
	A - controle	-6,62	0,83	63,73	1	<0,001]-8,24; -4,99 [
B – 250 NMP/ml	0,35	0,36	0,94	1	0,333]-0,35; 1,04[
C – 500 NMP/ml	1,63	0,38	18,31	1	<0,001]0,88; 2,38[

O modelo obtido através dos coeficientes é o seguinte:

$$\begin{aligned} \ln \left[\frac{P(Y \leq k)}{P(Y > k)} \right] = & \alpha_k - (-18,38\text{DIAM} - 0,82\text{ALT} + 0,27\text{FAM}(8) - 1,46\text{FAM}(14) - \\ & -1,44\text{FAM}(15) - 1,67\text{FAM}(16) + 0,00\text{FAM}(17) - 2,11\text{FAM}(31) - 1,07\text{FAM}(34) - \\ & -2,28\text{FAM}(48) - 6,62\text{CONC}(A) + 0,35\text{CONC}(B) + 1,63\text{CONC}(C)) \end{aligned}$$

ou em probabilidade linear não acumulada:

$$P(Y \leq k) = \frac{1}{1 + e^{-\left[\alpha_k - 18,38\text{DIAM} - 0,82\text{ALT} + 0,27\text{FAM}(8) - 1,46\text{FAM}(14) - 1,44\text{FAM}(15) - 1,67\text{FAM}(16) + 0,00\text{FAM}(17) - 2,11\text{FAM}(31) - 1,07\text{FAM}(34) - 2,28\text{FAM}(48) - 6,62\text{CONC}(A) + 0,35\text{CONC}(B) + 1,63\text{CONC}(C) \right]}}$$

De acordo com o modelo, à medida que o DIÂMETRO aumenta, diminui a probabilidade de se observarem níveis de ordem superior ($b_{\text{diâmetro}} = -18,38$; $p < 0,001$), sendo o seu OR=96013561.012, o que significa que por cm de aumento no DIÂMETRO, o rácio de probabilidade de se observarem níveis de menor ordem comparativamente à probabilidade de se observarem níveis de maior ordem aumenta 9601356001,2%. O mesmo acontece com a ALTURA ($b_{\text{altura}} = -0,82$; $p < 0,001$), sendo o seu OR igual a 2,270, ou seja a sua probabilidade de observar níveis de menor ordem é de 127% com o aumento de cm de ALTURA. Quanto à variável FAMÍLIA, verifica-se que o Ln das probabilidades de se observar um nível de ordem inferior, relativamente a um nível de ordem superior diminui 0,27 para a família 8 e é 0,00 para a família 17 relativamente à FAMÍLIA de referência (família 58), no entanto nenhuma delas tem um resultado estatisticamente significativo para $\alpha = 0,05$ ($b_{\text{família8}} = 0,27$; $p = 0,710$, e $b_{\text{família17}} = 0,00$; $p = 0,994$, respectivamente). O rácio de chances de se observar um nível de menor ordem relativamente a um de maior ordem é igual a 0,763 e 1 respectivamente, diminuindo 23,7% quando se passa da família de referência para a família 8 e mantendo-se inalterado para a família 17. Relativamente à variável CONCENTRAÇÃO, verifica-se que para a concentração C as probabilidades de se observar um nível de ordem inferior, relativamente a um nível de ordem superior diminui 1,63 ($b_{\text{concentraçãoC}} = 1,63$; $p < 0,001$), relativamente à concentração de referência (D). O rácio de chances respectivo de se observar um nível de menor ordem relativamente a um de maior ordem é igual a 0,196, diminuindo 80,4% quando se passa da concentração de referência para a concentração C.

A função de ligação a considerar de acordo com a figura 3.12 seria a *Log-log* Complementar, no entanto, a função de ligação considerada foi a *Logit*, dado que foi esta a apresentar melhor significância estatística.

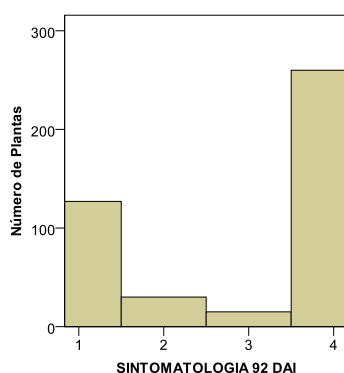


Figura 3.12 – Representação gráfica das frequências relativas aos diferentes níveis da variável resposta SINTOMATOLOGIA aos 92 dias após inoculação (DAI).

A regressão ordinal permite verificar que o pressuposto do modelo da homogeneidade dos declives foi validado ($X^2_{LP}(26)=8,52$; $p=1,000$), e demonstrou-se a impossibilidade de adicionar interações ao modelo, dado que piorava as significâncias das variáveis. O modelo foi considerado estatisticamente significativo ($G^2(13)=335,48$; $p<0,001$).

Os coeficientes e a significância do modelo ordinal ajustado estão dispostos na quadro 3.16.

Quadro 3.16 - Coeficientes *Logit* do modelo de regressão ordinal dos diferentes níveis da variável SINTOMATOLOGIA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 92 DAI.

Parâmetros		B	S.E.	X^2_{Wald}	g.l.	p-value	I.C. a 95% para Exp(B)
Pontos de corte	Nível 1	-8,86	0,91	95,35	1	<0,001]-10,64; -7,08 [
	Nível 2	-8,09	0,88	84,20	1	<0,001]-9,82; -6,36 [
	Nível 3	-7,72	0,87	78,75	1	<0,001]-9,43; -6,02[
Localização	Diâmetro	-16,86	1,95	74,88	1	<0,001]-20,68; -13,04 [
	Altura	-0,02	0,15	0,01	1	0,919]-0,30; 0,28[
	Família						
	8	-0,32	0,64	0,25	1	0,614]-1,59; 0,94[
	14	-1,90	0,68	7,72	1	0,005]-3,24; -0,56[
	15	-1,63	0,61	7,19	1	0,007]-2,82; -0,44[
	16	-1,56	0,62	6,32	1	0,012]-2,78; -0,34 [
	17	-0,37	0,66	0,32	1	0,569]-1,66; 0,91 [
	31	-1,45	0,63	5,32	1	0,021]-2,69; -0,22[
	34	-1,82	0,61	8,85	1	0,003]-3,02; -0,62[
	48	-1,28	0,64	3,93	1	0,046]-2,53; -0,02[
	Concentração						
	A - controlo	-3,75	0,41	84,07	1	<0,001]-4,56; -2,95 [
B – 250 NMP/ml	0,32	0,37	0,78	1	0,377]-0,37; 1,05[
C – 500 NMP/ml	0,88	0,40	4,98	1	0,026]0,11; 1,66[

O modelo pode ser exposto da seguinte forma:

$$\begin{aligned} \ln \left[\frac{P(Y \leq k)}{P(Y > k)} \right] = & \alpha_k - (-16,86\text{DIAM} - 0,02\text{ALT} - 0,32\text{FAM}(8) - 1,90\text{FAM}(14) - \\ & -1,63\text{FAM}(15) - 1,56\text{FAM}(16) - 0,37\text{FAM}(17) - 1,45\text{FAM}(31) - 1,82\text{FAM}(34) - \\ & -1,28\text{FAM}(48) - 3,75\text{CONC}(A) + 0,32\text{CONC}(B) + 0,88\text{CONC}(C)) \end{aligned}$$

ou em probabilidade linear não acumulada:

$$P(Y \leq k) = \frac{1}{1 + e^{-\left[\alpha_k - 16,86\text{DIAM} - 0,02\text{ALT} - 0,32\text{FAM}(8) - 1,90\text{FAM}(14) - 1,63\text{FAM}(15) - 1,56\text{FAM}(16) - 0,37\text{FAM}(17) - 1,45\text{FAM}(31) - 1,82\text{FAM}(34) - 1,28\text{FAM}(48) - 3,75\text{CONC}(A) + 0,32\text{CONC}(B) + 0,88\text{CONC}(C) \right]}}$$

De acordo com o modelo, à medida que o DIÂMETRO aumenta, diminui a probabilidade de se observarem níveis de ordem superior ($b_{\text{diâmetro}} = -16,86$; $p < 0,001$), sendo o seu OR = 20999307,102, logo, por cada cm de aumento no DIÂMETRO, o rácio de probabilidade de se observarem níveis de menor ordem comparativamente à probabilidade de se observarem níveis de maior ordem aumenta 2099930610,2%. O mesmo acontece com a ALTURA ($b_{\text{altura}} = -0,02$; $p = 0,919$), não sendo no entanto este resultado estatisticamente significativo para $\alpha = 0,05$. O OR para esta variável é igual a 1,020, ou seja a sua probabilidade de observar níveis de menor ordem é de 2,0% com o aumento de cm de ALTURA. Quanto à variável FAMÍLIA, verifica-se que a família 58 (de referência) tem a tendência a apresentar níveis de ordem superior dado que nenhuma das outras famílias apresenta um b positivo. Relativamente às concentrações, as probabilidades de se observar um nível de ordem inferior, relativamente a um nível de ordem superior diminui 0,88 ($b_{\text{concentraçãoC}} = 0,88$; $p < 0,001$), relativamente à concentração de referência. O rácio de chances respectivo de se observar um nível de menor ordem relativamente a um de maior ordem é igual a 0,415, diminuindo 58,5% quando se passa da concentração de referência para a concentração C.

3.3.3. Regressão multinomial

A regressão multinomial foi apenas realizada para a data de observação 35 DAI, dada a impossibilidade de validação do pressuposto de homogeneidade dos declives na regressão ordinal.

O modelo ajustado é estatisticamente significativo ($G^2(39)=391,53$; $p<0,001$), e os testes ao rácio de verosimilhança, demonstram que todas as variáveis: DIÂMETRO ($G^2(3)=81,15$; $p<0,001$), ALTURA ($G^2(3)=28,59$; $p<0,001$), FAMÍLIA ($G^2(24)=56,02$; $p<0,001$) e CONCENTRAÇÃO ($G^2(9)=205,28$; $p<0,001$), têm um efeito estatisticamente significativo, para $\alpha=0,05$, sobre o *Logit* da probabilidade de ocorrerem níveis superiores de SINTOMATOLOGIA, sendo o nível de referência a SINTOMATOLOGIA nível 1.

Este modelo apresenta ainda 88,2% de classificações correctas para a SINTOMATOLOGIA nível um, 0% para o nível dois, 7,7% para o nível três, e 79,7% para o nível quatro. No global, o modelo ajustado classifica correctamente 71,3% dos sujeitos, sendo a percentagem proporcional de classificações correctas por acaso igual a 37,8%. Assim, o modelo apresenta valores de classificações correctas muito superiores ao que se obteria por mero acaso. As estimativas dos coeficientes do modelo para as variáveis independentes FAMÍLIA, CONCENTRAÇÃO, DIÂMETRO e ALTURA, e para os níveis da variável resposta SINTOMATOLOGIA, relativamente à classe de referência “nível 1” são dadas na quadro 3.17.

De acordo com o modelo ajustado, a passagem da classe de referência “nível 1” para as classes “nível 2” e “nível 3” não é significativa para a variável ALTURA, nem para a concentração C, para $\alpha=0,05$. Considerando o mesmo nível de significância, não há efeito estatisticamente significativo em todos os níveis para a concentração B, e para a maioria dos níveis da variável FAMÍLIA. A passagem da classe de referência “nível 1” para as restantes classes é afectada significativamente pelas restantes variáveis. Pode-se assim concluir que por cada aumento de 1 cm no DIÂMETRO, as chances de ocorrer aumento de nível de SINTOMATOLOGIA decresce 100% para cada classe da variável resposta relativamente à classe de referência. Quanto à variável ALTURA, as chances de ocorrer aumento de nível por decréscimo de cm são maiores para a classe “nível 2” quando comparado com a classe de referência (não sendo o efeito do teste estatisticamente significativo), que corresponde a uma chance de 1:1,19, correspondente a um valor percentual de -16% de chance de aumentar de nível 1 para o 2 de SINTOMATOLOGIA, por cada cm de ALTURA.

Quanto à variável independente FAMÍLIA, é de destacar que a família 8 tem sempre as maiores chances de ter níveis de ordem superior relativamente à classe de referência (nível 1 de SINTOMATOLOGIA) e à família de referência (58). Este valor de chance aumenta 70,6% do nível 2 para o nível 3, decrescendo 40,8% do nível 3 para o 4. Por outro lado, a família 48 tem sempre as menores chances de ter níveis de ordem superior quando

comparada com a família de referência. Este valor decresce 12,5% do nível 2 para o nível 3, e 14,3% do nível 3 para o 4.

Quadro 3.17 – Coeficientes *Logit* do modelo de regressão multinomial dos diferentes níveis da variável SINTOMATOLOGIA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 35 DAI.

Variáveis	B	S.E.	X ² _{Wald}	g.l.	p-value	OR	I.C. a 95% para Exp(B)
Nível 2							
B	3,93	1,26	9,77	1	0,002		
Diâmetro	-10,87	3,02	12,97	1	<0,001	0,00]0,00; 0,01[
Altura	-0,17	0,26	0,45	1	0,502	0,84]0,51; 1,39[
Família							
8	0,62	0,92	0,46	1	0,496	1,87]0,31; 11,34[
14	-1,25	0,92	1,84	1	0,175	0,29]0,05; 1,75[
15	-1,13	0,82	1,89	1	0,169	0,32]0,06; 1,62[
16	-1,79	0,99	3,28	1	0,070	0,17]0,02; 1,16[
17	0,14	0,88	0,03	1	0,872	1,15]0,20; 6,53[
31	-21,27	0,00		3		0,00]0,00; 0,00[
34	-0,67	0,81	0,69	1	0,407	0,51]0,10; 2,51[
48	-2,59	0,99	6,89	1	0,009	0,08]0,01; 0,52[
Concentração							
A - controlo	-4,45	1,14	15,52	1	<0,001	0,01]0,00; 0,10[
B – 250 NMP/ml	0,28	0,49	0,32	1	0,570	1,32]0,50; 3,49[
C – 500 NMP/ml	-0,16	0,59	0,07	1	0,789	0,86]0,27; 2,71[
Nível 3							
B	6,86	1,34	26,08	1	<0,001		
Diâmetro	-20,52	3,56	33,25	1	<0,001	0,00]0,00; 0,00[
Altura	-0,19	0,25	0,57	1	0,451	0,83]0,51; 1,35[
Família							
8	1,16	0,98	1,39	1	0,238	3,19]0,46; 21,84[
14	-1,11	0,93	1,44	1	0,230	0,33]0,05; 2,02[
15	-1,42	0,92	2,40	1	0,121	0,24]0,04; 1,45[
16	-2,53	1,26	4,02	1	0,045	0,08]0,01; 0,95[
17	0,62	0,92	0,45	1	0,504	1,86]0,30; 11,34[
31	-2,09	1,06	3,87	1	0,049	0,12]0,02; 0,99[
34	-1,02	0,93	1,22	1	0,269	0,36]0,06; 2,21[
48	-2,61	0,96	7,46	1	0,006	0,07]0,01; 0,48[
Concentração							
A - controlo	-23,29	6640,60	0,00	1	0,997	0,00]0,00; - [
B – 250 NMP/ml	-0,43	0,56	0,58	1	0,446	0,65]0,22; 1,95[
C – 500 NMP/ml	0,54	0,52	1,08	1	0,299	1,72]0,62; 4,77[
Nível 4							
B	8,86	1,15	59,77	1	<0,001		
Diâmetro	-19,37	2,80	47,86	1	<0,001	0,00]0,00; 0,00[
Altura	-0,92	0,20	21,15	1	<0,001	0,40]0,27; 0,59[
Família							
8	0,64	0,84	0,57	1	0,451	1,89]0,36; 9,90[
14	-1,39	0,81	2,98	1	0,084	0,25]0,05; 1,21[
15	-1,58	0,76	4,33	1	0,038	0,21]0,05; 0,91[
16	-0,98	0,73	1,78	1	0,182	0,38]0,09; 1,58[
17	0,29	0,78	0,14	1	0,706	1,34]0,29; 6,13[
31	-1,28	0,74	2,96	1	0,085	0,28]0,06; 1,20[
34	-1,57	0,82	3,72	1	0,054	0,21]0,04; 1,03[
48	-2,73	0,75	13,35	1	<0,001	0,06]0,02; 0,28[
Concentração							
A – controlo	-6,74	1,29	27,40	1	<0,001	0,00]0,00; 0,02[
B – 250 NMP/ml	0,58	0,44	1,69	1	0,193	1,78]0,75; 4,26 [
C – 500 NMP/ml	1,66	0,45	13,70	1	<0,001	5,28]2,19; 12,76[

Por fim, para a variável independente CONCENTRAÇÃO, destaca-se que, para o nível 2, a concentração B é a teria maiores chances de apresentar níveis de SINTOMATOLOGIA de ordem superior, relativamente ao nível 1 (de referência) e à concentração de referência (D). No entanto, este teste não é estatisticamente significativo ($p=0,570$). Para os outros níveis, 3 e 4, é a concentração C a que tem chances maiores, aumentando 207% do nível 3 para o 4. A probabilidade de observar cada um dos níveis em função das variáveis independentes é dada por:

Nível 1

$$P(Y = 1) = \frac{1}{1 + e^{\left[\begin{array}{l} 3,93 - 10,87DIAM - 0,17ALT + 0,62FAM(8) - \\ -1,25FAM(14) - 1,13FAM(15) - 1,79FAM(16) + \\ + 0,14FAM(17) - 21,27FAM(31) + 0,67FAM(34) - \\ - 2,59FAM(48) - 4,45CONC(A) + 0,28CONC(B) - 0,16CONC(C) \end{array} \right]} + e^{\left[\begin{array}{l} 6,86 - 20,52DIAM - 0,19ALT + 1,16FAM(8) - \\ -1,11FAM(14) - 1,42FAM(15) - 2,53FAM(16) + \\ + 0,62FAM(17) - 2,09FAM(31) - 1,02FAM(34) - \\ - 2,61FAM(48) - 23,29CONC(A) - 0,43CONC(B) + 0,54CONC(C) \end{array} \right]} + e^{\left[\begin{array}{l} 8,86 - 19,37DIAM - 0,92ALT + 0,64FAM(8) - 1,39FAM(14) - 1,58FAM(15) - 0,98FAM(16) + \\ + 0,29FAM(17) - 1,28FAM(31) - 1,57FAM(34) - 2,73FAM(48) - 6,74CONC(A) + 0,58CONC(B) + 1,66CONC(C) \end{array} \right]}}$$

Nível 2

$$P(Y = 2) = \frac{e^{\left[\begin{array}{l} 3,93 - 10,87DIAM - 0,17ALT + 0,62FAM(8) - \\ -1,25FAM(14) - 1,13FAM(15) - 1,79FAM(16) + \\ + 0,14FAM(17) - 21,27FAM(31) + 0,67FAM(34) - \\ - 2,59FAM(48) - 4,45CONC(A) + 0,28CONC(B) - 0,16CONC(C) \end{array} \right]}}{1 + e^{\left[\begin{array}{l} 3,93 - 10,87DIAM - 0,17ALT + 0,62FAM(8) - \\ -1,25FAM(14) - 1,13FAM(15) - 1,79FAM(16) + \\ + 0,14FAM(17) - 21,27FAM(31) + 0,67FAM(34) - \\ - 2,59FAM(48) - 4,45CONC(A) + 0,28CONC(B) - 0,16CONC(C) \end{array} \right]} + e^{\left[\begin{array}{l} 6,86 - 20,52DIAM - 0,19ALT + 1,16FAM(8) - \\ -1,11FAM(14) - 1,42FAM(15) - 2,53FAM(16) + \\ + 0,62FAM(17) - 2,09FAM(31) - 1,02FAM(34) - \\ - 2,61FAM(48) - 23,29CONC(A) - 0,43CONC(B) + 0,54CONC(C) \end{array} \right]} + e^{\left[\begin{array}{l} 8,86 - 19,37DIAM - 0,92ALT + 0,64FAM(8) - 1,39FAM(14) - 1,58FAM(15) - 0,98FAM(16) + \\ + 0,29FAM(17) - 1,28FAM(31) - 1,57FAM(34) - 2,73FAM(48) - 6,74CONC(A) + 0,58CONC(B) + 1,66CONC(C) \end{array} \right]}}$$

Nível 3

$$P(Y = 3) = \frac{e^{\left[\begin{array}{l} 6,86 - 20,52DIAM - 0,19ALT + 1,16FAM(8) - \\ -1,11FAM(14) - 1,42FAM(15) - 2,53FAM(16) + \\ + 0,62FAM(17) - 2,09FAM(31) - 1,02FAM(34) - \\ - 2,61FAM(48) - 23,29CONC(A) - 0,43CONC(B) + 0,54CONC(C) \end{array} \right]}}{1 + e^{\left[\begin{array}{l} 3,93 - 10,87DIAM - 0,17ALT + 0,62FAM(8) - \\ -1,25FAM(14) - 1,13FAM(15) - 1,79FAM(16) + \\ + 0,14FAM(17) - 21,27FAM(31) + 0,67FAM(34) - \\ - 2,59FAM(48) - 4,45CONC(A) + 0,28CONC(B) - 0,16CONC(C) \end{array} \right]} + e^{\left[\begin{array}{l} 6,86 - 20,52DIAM - 0,19ALT + 1,16FAM(8) - \\ -1,11FAM(14) - 1,42FAM(15) - 2,53FAM(16) + \\ + 0,62FAM(17) - 2,09FAM(31) - 1,02FAM(34) - \\ - 2,61FAM(48) - 23,29CONC(A) - 0,43CONC(B) + 0,54CONC(C) \end{array} \right]} + e^{\left[\begin{array}{l} 8,86 - 19,37DIAM - 0,92ALT + 0,64FAM(8) - 1,39FAM(14) - 1,58FAM(15) - 0,98FAM(16) + \\ + 0,29FAM(17) - 1,28FAM(31) - 1,57FAM(34) - 2,73FAM(48) - 6,74CONC(A) + 0,58CONC(B) + 1,66CONC(C) \end{array} \right]}}$$

Nível 4

$$\begin{aligned}
 P(Y = 4) = & \\
 & e^{\left[\begin{array}{l} 8,86 - 19,37DIAM - 0,92ALT + 0,64FAM(8) - 1,39FAM(14) - 1,58FAM(15) - 0,98FAM(16) + \\ + 0,29FAM(17) - 1,28FAM(31) - 1,57FAM(34) - 2,73FAM(48) - 6,74CONC(A) + 0,58CONC(B) + 1,66CONC(C) \end{array} \right]} \\
 1 + & e^{\left[\begin{array}{l} 3,93 - 10,87DIAM - 0,17ALT + 0,62FAM(8) - \\ - 1,25FAM(14) - 1,13FAM(15) - 1,79FAM(16) + \\ + 0,14FAM(17) - 21,27FAM(31) + 0,67FAM(34) - \\ - 2,59FAM(48) - 4,45CONC(A) + 0,28CONC(B) - 0,16CONC(C) \end{array} \right]} + e^{\left[\begin{array}{l} 6,96 - 20,52DIAM - 0,19ALT + 1,16FAM(8) - \\ - 1,11FAM(14) - 1,42FAM(15) - 2,53FAM(16) + \\ + 0,62FAM(17) - 2,09FAM(31) - 1,02FAM(34) - \\ - 2,61FAM(48) - 23,29CONC(A) - 0,43CONC(B) + 0,54CONC(C) \end{array} \right]} + \\
 & + e^{\left[\begin{array}{l} 8,86 - 19,37DIAM - 0,92ALT + 0,64FAM(8) - 1,39FAM(14) - 1,58FAM(15) - 0,98FAM(16) + \\ + 0,29FAM(17) - 1,28FAM(31) - 1,57FAM(34) - 2,73FAM(48) - 6,74CONC(A) + 0,58CONC(B) + 1,66CONC(C) \end{array} \right]}
 \end{aligned}$$

6. DISCUSSÃO, CONCLUSÕES E PERSPETIVAS FUTURAS

A inoculação artificial com NMP em plantas de Pinheiro bravo com dois anos de idade permitiu detetar variabilidade na resposta das nove famílias de elite da população de melhoramento para o volume e forma da árvore, existentes no Centro Experimental do Escaroupim. Estes resultados, embora preliminares, são promissores. Apesar das limitações inerentes ao estudo de resistência a doenças baseado em plantas jovens e inoculações artificiais aplicadas em ambiente controlado, esta tem sido a principal forma de o fazer em programas de melhoramento (Sniezko 2006). De facto, McNamara (2004) refere que diferenças na origem, estádios de vida, frequência de cultura do inóculo, concentrações e técnicas de inoculação, possível contaminação das populações de NMP, assim como idade das plantas inoculadas e condições ambientais, podem afetar os resultados dos testes de inoculação artificial.

Neste estudo, a inoculação de um isolado obtido recentemente de uma zona altamente afetada pela doença foi a forma de garantir a manutenção do grau de virulência e agressividade do inóculo em condições o mais próximo possível das encontradas nos povoamentos afetados. Como Boerma e Hussey (1992) referem, a utilização de um isolado agressivo é importante para detetar genótipos com o máximo grau de resistência.

Um aspeto importante a considerar no estudo da variação de resposta à PWD diz respeito à sua discriminação em termos de categorias/níveis consideradas. A divisão em níveis de sintomatologia utilizada neste estudo, comparativamente ao binómio planta morta/viva, parece permitir detetar mais cedo, nas plantas, os efeitos da inoculação quer na família quer na concentração de nemátode inoculada. De facto, a comparação das significâncias das respostas entre concentrações e entre famílias pela variável resposta binária (viva e morta) e a ordinal (sintomatologia) mostrou que, nesta última, aos 21 DAI, já se verificaram diferenças significativas nas respostas, refletindo a sua maior capacidade discriminante.

Daub (2007) refere que a mortalidade das plantas infetadas com NMP pode ser alterada pela temperatura ambiente. Rutherford et al. (1990), acrescentam que o desenvolvimento dos sintomas da PWD necessita de uma temperatura média de Verão de pelo menos 20°C. Por outro lado, Yang et al. (2002) afirmam que se tal temperatura média não for atingida poderá ocorrer apenas uma infeção latente, sem sintomas visíveis imediatos. De forma a garantir que, durante o ensaio realizado neste estudo, todas as plantas pudessem estar em igualdade de circunstâncias no que respeita ao desenvolvimento dos sintomas, procurou-se

que as condições ambientais se mantivessem controladas, garantindo temperaturas médias acima dos 20°C.

De acordo com Sniezko (2006), em melhoramento florestal para a resistência a doenças, o objetivo é o de manter uma diversidade genética que permita adaptabilidade. Os resultados de diferenças de mortalidade e sintomatologia obtidos na realização deste estudo permitem verificar a existência de variabilidade nas famílias analisadas. A análise dos resultados observados permitiu identificar a família 34 como a menos suscetível e a família 58 como a mais vulnerável, para todas as datas de observação, exceto aos 14 DAI, data em que ainda não eram observadas diferenças significativas na mortalidade e sintomatologia. Por outro lado, a regressão categorial, para além de confirmar as diferenças de mortalidade e sintomatologia entre estas duas famílias, permitiu a construção de modelos em que se prevê a probabilidade de morte ou acréscimo de sintomas para as famílias, considerando simultaneamente a concentração de nemátode aplicada, a altura e o diâmetro da planta. De facto, as chances de mortalidade aos 35 DAI são 90% menores para a família 34 relativamente à família 58. Aos 47 DAI as chances aumentam para 95,3% e aos 92 DAI, são 80,2% menores, para a família 34, relativamente à família 58. Considerando a variável resposta sintomatologia, a regressão ordinal permite confirmar as diferenças entre estas duas famílias, dado que aos 47 DAI, as chances da família 34 apresentarem sintomas de ordem inferior é 191,5% maior para a família 34 relativamente à 58. O mesmo acontece aos 92 DAI, em que as chances são ainda mais díspares, com uma diferença de 517,2% nas chances de apresentar sintomas de ordem inferior. Por fim, para os 35 DAI, a regressão multinomial permite verificar que a família 34 tem sempre menores chances de ter níveis de ordem superior quando comparada com a família 58. Assim sendo, e tendo em conta, que de acordo com Li (2008), o método mais eficaz de controlo da doença da marchidão do pinheiro é a plantação de *Pinus* spp. resistentes, então a família 34 seria uma família candidata a testes mais aprofundados sobre as suas características e capacidade de resistência à PWD. Aliás, no programa de melhoramento japonês, Toda e Kurinobu (1998, 2001, 2002) já tinham verificado que era possível encontrar árvores tolerantes à doença da marchidão do pinheiro através de ensaios de inoculação, assim como Woo et al. (2007), que verificaram que é possível encontrar variação na suscetibilidade das plantas ao NMP em ensaios de estufa. Os resultados obtidos devem no entanto ser avaliados no campo, e a variabilidade encontrada deve ser testada, tal como indicado por Sniezko (2006).

Considerar a inoculação em plantas jovens implica necessariamente decidir a concentração de nemátodes a aplicar. Braasch (1997) considera que a quantidade inicial de inóculo deve conter o número suficiente de nemátodes capazes de sobreviver e construir uma população

estável. Peng e Moen (2003) acrescentam que, para a avaliação da resistência, a quantidade de inóculo aplicada deve ser suficiente para o tipo de experiência que se pretende, mas limitada de forma a não provocar demasiado dano ou morte mecânica que vá mascarar material genético potencialmente resistente. Os resultados obtidos neste estudo, mostraram que diferentes concentrações de NMP provocaram diferentes dinâmicas no desenvolvimento dos sintomas e mortalidades das plantas, algo já indicado por Li (2008) para plantas de 3-4 anos de várias espécies de *Pinus*. Este autor mostrou que a aplicação de baixas densidades de nemátodes (menos de 200 NMP) pode não provocar a morte da planta, dado os nemátodes ficarem presos na superfície resinosa da ferida e, conseqüentemente morrerem. Considerando que o insecto vetor transmite diariamente cerca de 350 NMP em condições naturais, Li (2008) sugere que densidades de inóculo de cerca de 200 NMP por planta possa ser o número mais apropriado para estudar a resposta com inoculações artificiais. Este autor considera ainda que a inoculação em números de milho é menos convincente. Neste estudo, a concentração 250 NMP/ml já foi suficiente para causar mortalidade, contudo a inoculação de 500 NMP/ml foi a que causou a maior mortalidade. De realçar, no entanto, que esta concentração provocou maior mortalidade que a de 750 NMP/ml. Este facto poderá ser devido a que, como referido por Daub (2007), concentrações mais elevadas possam provocar menor taxa de reprodução do nemátode no interior da planta.

A diferenciação na sintomatologia das plantas começou a ser significativamente diferente, entre concentrações e entre famílias, aos 21 DAI. A idade das plantas usadas nos estudo tem influência na rapidez em que se tornam visíveis os sintomas. Para Akiba et al. (2012), sintomas em *P. thunbergii*, de idade inferior a um ano, iniciaram-se 7 DAI, enquanto que, para Daub (2007), a ocorrência dos primeiros sintomas em plantas de *P. sylvestris* com 3-4 de idade, foi visível quatro semanas após inoculação, independentemente da densidade de inóculo.

Os resultados deste estudo demonstram que as variáveis altura e diâmetro influenciam a resposta das plantas ao inóculo, verificando-se que quanto maiores são os diâmetros e as alturas, menores os níveis de sintomatologia ou de mortalidade. Estes resultados estão parcialmente de acordo com os obtidos por Mori et al. (2008) em plantas de *P. thunbergii* e *P. densiflora*. Os autores constataram que o diâmetro influenciava a resposta, mas o mesmo não se verificou para a altura.

É possível concluir através deste estudo que a conjugação dos objetivos de melhoramento genético, resistência ao NMP e volume e forma do tronco, parece ser vantajosa para o Pinheiro bravo. Um estudo mais aprofundado das diferenças de resistência presentes nas

famílias testadas da população de melhoramento para o volume e forma do tronco, assim como o estudo da variabilidade genética presente nesta população, deve ser considerado num futuro próximo, já que os ganhos genéticos em 21% em volume e 17% em retidão de fuste obtidos num programa desenvolvido há mais de 50 anos, assim o justificam. A doença da marchidão do pinheiro é, de facto, uma doença devastadora para esta espécie florestal. Este ensaio forneceu indicações importantes para o estabelecimento de ensaios de inoculação artificial em plantas jovens de Pinheiro bravo. Através de análises estatísticas apropriadas ao tipo de dados em análise (variáveis de resposta binária e ordinal), demonstrou-se a importância da variabilidade de resposta de diferentes famílias condicionada pela concentração de inóculo e do factor tempo de resposta. Há no entanto a destacar que a presença de uma variável quantitativa que avaliasse a quantidade de NMP na planta após a morte da planta era essencial para a determinação da resistência da mesma. Tal não foi possível dada a escolha efetuada inicialmente de avaliar a sintomatologia das plantas até aos 97 dias após inoculação e só depois se proceder ao abate. Esta observação da capacidade reprodutiva e de desenvolvimento do NMP recorrendo a uma variável quantitativa é pertinente e será considerada em ensaios de inoculação futuros.

Como perspectiva futura, há a considerar a importância deste estudo no desenrolar do programa de melhoramento para a resistência ao nemátode, iniciado com a seleção de árvores na Herdade da Comporta. Esta população, cujos indivíduos adultos continuam a ser monitorizados, tem já, em preparação, um ensaio de descendências com cerca de 100 famílias, que, no espaço de dois anos, poderá ser analisado quanto à variabilidade genética de resposta à doença e quanto à possibilidade de existência de indivíduos resistentes. O cruzamento destes dados com os dados obtidos na monitorização das árvores adultas no campo permitirá certamente encurtar o tempo necessário para cumprir com sucesso este programa de melhoramento. Tal como Sniezko (2006) refere, múltiplos ciclos de seleção e melhoramento serão geralmente necessários para desenvolver resistência a agentes patogénicos não nativos. Zobel e Talbert já em 1984 indicam que, a maior desvantagem dos programas de melhoramento é o tempo, devendo-se obter o maior ganho genético possível por unidade de tempo. Ensaio de inoculação simultaneamente em descendências e em clones poderão ser uma forma de encurtar esse tempo e obter resultados sólidos que permitam controlar a evolução desta doença.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- Abelleira, A., Picoaga, A., Mansilla, J. P. e Aguin, O., 2011. Detection of *Bursaphelenchus xylophilus*, causal agent of pine wilt disease in northwestern Spain. *Plant Disease* 95: 776.
- AFN, 2010. 5º Inventário Florestal Nacional – Relatório final. Direção Geral de Gestão florestal. Autoridade Florestal Nacional.
- Agresti, A., 2007. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Inc. Segunda Edição. 372p.
- Aguiar, A., Almeida, M. H. e Borralho, N., 2003. Genetic control of growth, wood density and stem characteristics of *Pinus pinaster* in Portugal. *Silva Lusitana* 11: 1-9.
- Aguiar, A., Carrasquinho, I. e Alpuim, M., 2011. Melhoramento genético em pinheiro bravo. *In: Paula Coelho e Pedro Reis (eds.), Agrorural: contributos científicos*. INRB e INCM, Lisboa: pp. 1013-1024.
- Akiba, M.; Ishihara, M.; Sahashi, N. e Nakamura, K., 2012. Virulence of *Bursaphelenchus xylophilus* isolated from naturally infested pine forests to five resistant families of *Pinus thunbergii*. *Plant Disease* 96:249–252
- Baradat, P. e Marpeau, A., 1988. Le pin maritime *Pinus pinaster* Ait. : Biologie et génétique des terpènes pour la connaissance et l'amélioration de l'espèce. PhD Thesis, Université de Bordeaux, Bordeaux, France.
- Boerma, H. R. e Hussey, R. S., 1992. Breeding plants for resistance to nematodes. *Journal of Nematology*. 24 : 242-252.
- Bolla, R. I.; Winter, R. E. K.; Fitzsimmons, K. e Linit, M. J., 1986. Pathotypes of the pinewood nematode *Bursaphelenchus xylophilus*. *Journal of Nematology* 18: 230-238
- Bonifácio, L. F. P., 2009. Impacte e evolução da doença da murchidão dos pinheiros (Pine Wilt Disease) na zona afectada a sul do rio Tejo. PhD Thesis, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal.
- Bowers, W., Hudak, J., Raske, A., Magasi, L., Myren, D., Lachance, D., Cerezke, H. e Sickle, G., 1992. Host and vector surveys for the pinewood nematode, *Bursaphelenchus xylophilus* (Steiner and Buhner) Nickle (Nematoda: Aphelenchoididae) in Canada. Information Report Newfoundland and Labrador Region, Forestry Canada.
- Braasch, H., 1997. Host and pathogenicity tests with pine wood nematode (*Bursaphelenchus xylophilus*) from North America under Central European weather conditions. *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes* 49 : 209-214.

- Carson, S. D. e Carson, M. J., 1989. Breeding for resistance in forest trees – a quantitative genetic approach. *Annual Review of Phytopathology* 27:373-95.
- Cox, D. R. e Snell, E. K., 1989. *Analysis of Binary Data*. Taylor & Francis. Segunda Edição. 240p.
- Chang, R., 1999. Effects of inoculation method and inoculum concentration on the occurrence of pine wilt disease. *Taiwan Journal of Forest Science* 14:409-417.
- Cheng, H. R., Lin, M., Li, W., e Fang, Z., 1983. The occurrence of a pine wilting disease caused by a nematode found in Nanjing. *Forest Pest and Disease* 4: 1-5.
- Danti, R., Raddi, P. Panconesi, A., Di Lonardo, V., Nembi, V., Rocca, G. D. e Torraca, G., 2008. O melhoramento genético do cipreste para a resistência ao cancro cortical. (<http://www.arsia.toscana.it/filfor/cypmed-arsia/download/algarve%20brochure%20pdf/11-melhorament-28-35.pdf>). Acesso em 10/08/2012.
- Daub, M., 2007. Investigations on pathogenicity, invasion biology and population dynamics of the pine wood nematode *Bursaphelenchus xylophilus* (Steiner und Buhner 1934) Nickle 1970 in European conifers. Dissertation. University of Bonn, Germany.
- Decreto-Lei nº95/2011, de 8 de Agosto, Diário da República nº 151/2011 – I Série. Ministério da Agricultura, do Mar, do Ambiente e do Ordenamento do Território. Lisboa. pp.4202-4219.
- Dropkin, V. H., e Foudin, A. S., 1979. Report of the occurrence of *Bursaphelenchus lignicolus* induced pine wilt disease in Missouri. *Plant Disease Reporter* 63: 904-905.
- Dwinell, L. D., 1993. First report of pinewood nematode (*Bursaphelenchus xylophilus*) in Mexico. *Plant Disease* 77: 846.
- Dwinell, L.D., 1997. The pinewood nematode: regulation and mitigation. *Annual Review of Phytopathology* 35: 153-166.
- Eriksson, G., Ekberg, I. e Clapham, D., 2006. *An Introduction to Forest Genetics*. Genetic Center SLU, Uppsala, Sweden. Segunda Edição. 185p.
- Evans, H., McNamara, D., Braasch, H., Chadoeuf, J. e Magnusson, C., 1996. Pest Risk Analysis (PRA) for the territories of the European Union (as PRA area) on *Bursaphelenchus xylophilus* and its vectors in the genus *Monochamus*. *European Plant Protection Organization Bulletin* 26: 199-249.
- Fonseca, L., Lopes, A., Cardoso, J., Pestana, M., Abreu, F., Nunes, N., Mota, M. e Abrantes, I., 2010. The pinewood nematode, *Bursaphelenchus xylophilus*, in Madeira Island. 30th International Symposium of European Society of Nematologists, Vienna, Austria, p.176. [Abstr.]

- Futai, K., 1980. Population dynamics of *Bursaphelenchus lignicolus* (Nematoda: Aphelenchoididae) and *B. mucronatus* in pine seedlings. Japanese Society of Applied Entomology and Zoology 15: 458-464.
- Futai, K., 2003. Role of asymptomatic carrier trees in epidemic spread of pine wilt disease. Journal Forest Research 8:253-260. DOI: 10.1007/s10310-003-0034-2.
- Gordillo, L. F. e Kim, Y., 2012. A simulation of the effects of early eradication of nematode infected trees on spread of pine wilt disease. European Journal Plant Pathology 132: 101-109.
- Hoshmand, A. R., 1997. Statistical Methods for Environmental and Agricultural Sciences. Taylor & Francis. Segunda Edição. 464p.
- Hosmer, D. W., e Lemeshow, S., 2000. Applied Logistic Regression. John Wiley & Sons, Inc. Segunda Edição. 375p.
- Ikeda, T., e Kiyohara, T., 1995. Water relations, xylem embolism and histological features of *Pinus thunbergii* inoculated with virulent or avirulent pine wood nematode *Bursaphelenchus xylophilus*. Journal Experimental Botany 46: 441-449.
- Kato, K., Kawamura, K. e Ueki, C., 1995. Resistance to the pinewood nematode in seedlings obtained from resistant clonal seed orchards. Japanese Journal of Nematology 25: 52-55.
- Kishi, Y., 1995. The pine wood nematode and the Japanese pine sawyer. Thomas Company Limited, Tokyo, Japan. 302 pp.
- Kiyohara, H. e Bolla, R. I., 1990. Pathogenic variability among populations of the pinewood nematode, *Bursaphelenchus xylophilus*. Forest Science 36: 1061-1076.
- Kobayashi, F., 1988. The Japanese pine sawyer. In: Berryman AA (ed.), Dynamics of Forest Insect Populations: Patterns, Causes, Implications. pp. 431-454, New York: Plenum Press.
- Kruskal, W. H. e Wallis, W. A., 1952. Use of ranks in one-criterion variance analysis. Journal of the American Statistic Association 47: 583-621.
- Kurinobu, S., 2008. Current Status of resistance breeding of Japanese Pine species to Pine Wilt Disease. Forest Science and Technology 4: 83-94.
- Kuroda, K., Yamada, T. e Ito, S., 1991. *Bursaphelenchus xylophilus* induced pine wilt: factors associated with resistance. European Journal of Forest Pathology 21: 430-438.
- La, Y., Moon, Y., Yeo, W., Shin, S. e Bak, W., 1999. Recent status of pine wilt disease in Korea. In: Futai, K., Togashi, K. and Ikeda, T. (Eds). Sustainability of pine forests in

- relation to pine wilt and decline. Proceedings of the Symposium, Tokyo, Japan, 26-30 October 1998. Kyoto, Japan, Shokado Shoten, pp. 239-241.
- Levene, H., 1960. Robust Tests for Equality of Variances. *In: Contributions to Probability and Statistics*. Olkin, I. (ed.). Palo Alto, CA: Stanford University Press.
- Li, H., 2008. Identification and Pathogenicity of *Bursaphelenchus xylophilus* species (Nematoda: Parasitaphelenchidae). PhD Thesis, Ghent University, Ghent, Belgium.
- Mamiya, Y., 1983. Pathology of the pine wilt disease caused by *Bursaphelenchus xylophilus*. *Annual Review of Phytopathology* 21: 201-220.
- Mamiya, Y., 1984. Resistance of *Pinus* spp. against *Bursaphelenchus xylophilus* and *B. mucronatus* focusing on the host age (in Japanese). *Trans Meeting of Japanese Forest Society* 95:475–476.
- Mamiya, Y., 1988. History of pine wilt disease in Japan. *Journal of Nematology* 20: 219-226.
- Maroco, J., 2007. *Análise Estatística: com utilização do SPSS*. Edições Silabo. 3ª edição. 822p.
- Martins, A. P., 1974. *Iniciação à Estatística*. Porto Editora. 1ª Edição. Porto. 176p.
- McNamara, D., 2004. Quarantine concerns about the methods used to demonstrate pathogenicity of *Bursaphelenchus* spp. *In: Mota, M. e Vieira, P., eds. The pinewood nematode, Bursaphelenchus xylophilus*. Proceedings of an International Workshop, University of Evora, Portugal, 20-22 August 2001. Brill Academic Publishers, Leiden Netherlands: 187-197.
- Metge, K. e Burgermeister, W., 2005. Molecular identification and pathway analysis of the introduced pinewood nematode *Bursaphelenchus xylophilus*. *In: Proceedings of a Symposium of Plant Protection and Plant Health in Europe. Introduction and spread of invasive species*: 2pp.
- Mori, Y.; Miyahara, F.; Tsutsumi, Y. e Kondo, R., 2008. Relationship between resistance to pine wilt disease and the migration or proliferation of pine wood nematodes. *European Journal Plant Pathology* 122:529-538. DOI: 10.1007/s10658-008-9321-2.
- Mota, M. M., Braasch, H., Bravo, M. A., Penas, A. C., Burgermeister, W., Metge, K. e Sousa, E., 1999. First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. *Nematology* 1: 727-734.
- Mota, M. M. e Vieira, P., 2008. Pine wilt disease: a worldwide threat to forest ecosystems. Springer. pp. 405

- Mota, M. M., Futai, K. e Vieira, P., 2009. Pine wilt disease and the pinewood nematode, *Bursaphelenchus xylophilus*. In: Integrated management of fruit crops and forest nematodes, Edited by Ciancio, A. and Mukerji, K. G. Springer: 253-274.
- Nagelkerke, N. J. D., 1991. A Note on a General Definition of the Coefficient of Determination *Biometrika* 78 : 691–2. doi:10.1093/biomet/78.3.691
- Naves, P. M. P. G. D., 2007. Biology of *Monochamus galloprovincialis* (Coleoptera, Cerambycidae) and its role as vector of the pine wood nematode *Bursaphelenchus xylophilus* (Nematoda, Parasitaphelenchidae) in Portugal. PhD thesis. Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal.
- Naves, A. C. G. P., 2008. Morphobiometric and molecular characterization of *Bursaphelenchus xylophilus* Fuchs, 1937 (Nematoda: Aphelenchoididae) species associated with *Pinus pinaster* Aiton in Portugal. PhD thesis, Universidade de Évora, Évora, Portugal.
- Norusis, M. J., 2008. SPSS 16.0 Advanced Statistical Procedures Companion. Prentice Hall Inc.418p.
- Nose, M., e Shiraishi, S., 2008. Breeding for resistance to pine wilt disease. In: B. Guang Zhao, K. Futai, J. R. Sutherland & Y. Takeuchi, eds., Pine Wilt Disease (pp. 334-350). Tokio: Springer.
- Oliveira, A. C., Pereira, J. S. e Correia, A. V., 2000. A silvicultura do pinheiro bravo. Centro Pinus, Porto, 111pp.
- Peng, Y. e Moens, M., 2003. Host resistance and tolerance to migratory plant-parasitic nematodes. *Nematology* 5: 145-177.
- Perry, D. H. e Hopkins, E. R., 1967. Importation of breeding material of *Pinus pinaster* Ait. from Portugal. Forest Department Western Australia: Bulletin nº75.
- Pestana, M. H. e Gageiro, J. P., 2008. Análise de Dados para Ciências Sociais: a complementaridade do SPSS. Edições Silabo. 5ª Edição. 692 p.
- Pestana, M. H. e Gageiro, J. P., 2009. Análise Categórica, Árvores de Decisão e Análise de Conteúdo: em ciências sociais e da saúde com o SPSS. Lidel. Primeira Edição. 551p.
- Pico, F. C. M. de A., 2011. Efeito do nemátode da madeira do pinheiro no módulo de elasticidade da madeira de pinheiro-bravo. Dissertação para obtenção de Grau de Mestre em Engenharia Florestal e dos Recursos Naturais, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Lisboa, Portugal.

- Ribeiro, B., Espada, M., Vu, T., Nóbrega, F., Mota, M. e Carrasquinho, I., 2012. Pine wilt disease: detection of the pinewood nematode (*Bursaphelenchus xylophilus*) as a tool for a pine breeding program. *Forest Pathology*. (aceite para publicação).
- Robbins, K., 1982. Distribution of the pinewood nematode in the United States. *In*: Appleby, J.E. and Malek, R.B., eds. *Proceedings of the national pine wilt disease workshop*. III. Nat. Hist. Surv. Champaign, IL., pp. 3-6.
- Robertson, L., Cobacho Arcos, S., Escuer, M., Santiago Merino, R., Esparrago, G., Abelleira, A. e Navas, A., 2011. Incidence of the pinewood nematode *Bursaphelenchus xylophilus* Steiner & Buhner, 1934 (Nickle, 1970) in Spain. *Nematology* 13: 755-757.
- Rodrigues, A., Casquilho, M., Oliveira, H. e Bordalo, J., 2009. A statistical analysis of the impact of nematode attack symptomatology on the mechanical behaviour of *Pinus pinaster* Ait. wood. *European Journal Forest Research*. DOI: 10.1007/s10342-009-0307-1.
- Rodrigues, J., 2008. National eradication programme for the pinewood nematode. *In* *Pine Wilt Disease: A Worldwide Threat to Forest Ecosystems*, Edited by Mota, M. and Vieira, P. Springer. pp. 5-14.
- Roldão, M. I. C. F., 2000. Propagação vegetativa de sobreiros selecionados. PhD Thesis. Instituto Superior de Agronomia, Universidade Técnica de Lisboa, Lisboa, Portugal.
- Roulund, H., Alpuim, M., Varela, M. C. e Aguiar, A., 1988. A Tree Improvement plan for *Pinus pinaster* in Portugal. HEDESELSKABET.EFN. Lisboa
- Rutherford, T. A. e Webster, J. M., 1987. Distribution of pine wilt disease with respect to temperature in North America, Japan, and Europe. *Canadian Journal of Forest Research* 17: 1050-1059.
- Rutherford, T. A., Mamiya, Y. e Webster, J. M., 1990. Nematode-induced pine wilt disease: factors influencing its occurrence and distribution. *Forest Science* 36: 145-155.
- Sanz, F., Latour, S., Neves, M., Bastet, E., Pischedda, D., Piñero, G., Gauthier, T., Lesbats, J., Plantier, C., Marques, A., Lanvin, J. D., Santos, J. A., Touza, M., Pedras, F., Parrot, J., Reuling, D. e Faria, C., 2006. Industrial applications of *Pinus pinaster*. FFCI/DTG/FIBA/AIMMP/CTBA. Ourense. 256p.
- Siegel, S., 1975. *Estatística Não-Paramétrica (para as ciências do comportamento)*. McGraw-Hill. Brasil. 350p.
- Schröder, T., McNamara, D. G. e Gaar, V., 2009. Guidance on sampling to detect pine wood nematode *Bursaphelenchus xylophilus* in trees, wood and insects. *Bulletin OEPP/EPPO Bulletin* 39: 179-188.

- Shin, S-C., 2008. Pine wilt disease in Korea. *In*: Zhao, B. G., Futai, K., Sutherland, J. R., & Takeuchi, Y., eds. Pine Wilt Disease. Springer, pp. 26-32.
- Shtatland, E. S., Kleinman, K. e Cain, E. M., 2002. One more time about R^2 measures of fit in logistic regression. Proceedings of the NESUG. September 29-October 2, 2002, Buffalo, New York.
- Snieszko, R. A., 2006. Resistance breeding against nonnative pathogens in forest trees – current successes in North America. Canadian Journal Plant Pathology 28: S270-S279.
- Sousa, E., Bravo, M. A., Pires, J., Naves, P. M., Penas, A. C., Bonifácio, L. e Mota, M., 2001. *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae) associated with *Monochamus galloprovincialis* (Coleoptera: Cerambycidae) in Portugal. Nematology 3: 89-91.
- Sutherland, J. R., 2008. A brief overview of the pine wood nematode and pine wilt disease in Canada and the United States. *In*: Zhao, B. G., Futai, K., Sutherland, J. R., & Takeuchi, Y., eds. Pine Wilt Disease. Springer, pp. 13-17.
- Toda, T. e Kurinobu, S., 1998. Current status of resistance breeding of Japanese pine species to Pine Wilt Disease. Proceedings International Symposium SustainAbility. Pine Forests Relation Pine Wilt Decline: 208-212.
- Toda, T. e Kurinobu, S., 2001. Genetic Improvement in Pine Wilt Disease resistance in *Pinus thunbergii*: The effectiveness of pre-screening with an artificial inoculation at the nursery. Journal of Forest Research 6: 197-201.
- Toda, T. e Kurinobu, S., 2002. Realized genetic gains observed in progeny tolerance of selected red pine (*Pinus densiflora*) and black pine (*P. thunbergii*) to pine wilt disease. Silvae Genetica 51: 42-44.
- Tzean, S. S. e Tang, J. S., 1985. The occurrence of the pine wood nematode, *Bursaphelenchus xylophilus*, in Taiwan. Proceedings 6th ROC Symposium of Electron Microscopy, pp. 38–39.
- Valadas, V., Laranjo, M., Barbosa, P., Espada, M., Mota, M. e Oliveira, S., 2012. The pine wood nematode, *Bursaphelenchus xylophilus*, in Portugal: possible introductions and spread routes of a serious biological invasion revealed by molecular methods. Nematology 14: 899-911.
- Vicente, C., Espada, M., Vieira, P. e Mota, M., 2011. Pine wilt disease: a threat to European forestry. European Journal Plant Pathology 133:89-99. Doi: 10.1007/s10658-011-9924-x

- Vieira, P., Burgermeister, W., Mota, M., Metge, K. e Silva, G., 2007. Lack of genetic variation of *Bursaphelenchus xylophilus* in Portugal revealed by RAPD-PCR analyses. *Journal of Nematology* 25: 361-364.
- Wingfield, M. J., 1987. A comparison of the mycophagous and the phytophagous phases of the pine wood nematode. *In: Pathogenicity of the Pine Wood Nematode*. Wingfield, M.J, APS Press, St Paul (US) 8:1-90.
- Woo, K. S.; Kim, Y. S.; Koo, Y. B.; yeo, J. K. e Moon, Y. S., 2007. Variation in susceptibility of pine species seedlings with the pine wood nematode, *Bursaphelenchus xylophilus*, in greenhouse. *Plant Pathology Journal* 23:70-75.
- Yang, B., 2004. The history, dispersal and potential threat of pine wood nematode in China. *In: Mota, M. and Vieira, P., eds. The pinewood nematode, Bursaphelenchus xylophilus*. *Nematology Monographs and Perspectives*, volume 1. Leiden-Boston: Brill Academic Publishers, pp. 21-24.
- Yang, B. e Qouli, W., 1989. Distribution of the pinewood nematode in China and susceptibility of some Chinese and exotic pines to the nematode. *Canadian Journal of Forest Research* 19: 1527–1530.
- Yang, B.; Wang, L.; Zhao, W.; Xu, F.; Zhang, P. e Li, Z., 2002. The latent infection of *Bursaphelenchus xylophilus* and a new transmission way of PWN by *Monochamus alternatus*. *Forest Research, Beijing* 15: 251-255.
- Yi, C., Byun, B., Park, J., Yang, S. e Chang, K., 1989. First finding of the pine wood nematode, *Bursaphelenchus xylophilus* (Steiner & Buhrer) Nickle and its insect vector in Korea. *Research Reports of the Forestry Research Institute Seoul* 38: 141-149.
- Zobel, B. e Talbert, J., 1984. *Applied Forest Tree Improvement*, John Wiley & Sons. New York, 505 p.

ANEXO I

Estudo da regressão logística com e sem a presença de observações extremas (outliers)

35 dias após inoculação (DAI)

Ajustamento do modelo (**com** todas as observações):

$$G^2(1)=65,58; p<0,001;$$

$$X^2_{HL}(8)=6,53; p=0,589;$$

$$R^2_{CS}=0,490;$$

$$R^2_N=0,660$$

Ajustamento do modelo (**sem** a presença dos outliers):

$$G^2(1)=68,51; p<0,001;$$

$$X^2_{HL}(8)=9,64; p=0,291;$$

$$R^2_{CS}=0,496;$$

$$R^2_N=0,670$$

Coeficientes *Logit* do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 35 DAI. (**com** todas as observações).

Variável	B	S.E.	X ² _{Wald}	g.l.	p-value	OR	I.C. a 95% para Exp(B)
FAMÍLIA			20,32	8	0,009		
8	-1,25	0,73	2,92	1	0,088	0,29]0,07; 1,20 [
14	-1,52	0,67	5,08	1	0,024	0,22]0,06; 0,82 [
15	-1,09	0,66	2,73	1	0,099	0,34]0,09; 1,23 [
16	-0,16	0,64	0,06	1	0,801	0,85]0,24; 3,01 [
17	-0,98	0,67	2,16	1	0,142	0,38]0,10; 1,39 [
31	-1,55	0,71	4,75	1	0,029	0,21]0,05; 0,86 [
34	-2,30	0,66	11,98	1	0,001	0,10]0,03; 0,37 [
48	-0,50	0,66	0,56	1	0,453	0,61]0,16; 2,23 [
CONCENTRAÇÃO			41,61	3	<0,001		
A – controlo	-5,91	1,17	25,51	1	<0,001	0,00]0,00; 0,03 [
B – 250 NMP/ml	0,16	0,37	0,20	1	0,655	1,18]0,57; 2,42 [
C – 500 NMP/ml	1,32	0,38	12,34	1	<0,001	3,74]1,79; 7,80 [
ALTURA	-0,66	0,16	16,09	1	<0,001	0,52]0,38; 0,71 [
DIÂMETRO	-16,18	2,39	46,01	1	<0,001	0,00]0,00; 0,00 [
Constante	7,43	0,98	57,53	1	<0,001	1690,26	

Coefficientes *Logit* do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 35 DAI. (**sem** a presença dos *outliers*).

Variável	B	S.E.	X^2_{Wald}	g.l.	p-value	OR	I.C. a 95% para Exp(B)
FAMÍLIA			22,38	8	0,004		
8	-1,35	0,75	3,28	1	0,070	0,26]0,06; 1,12[
14	-1,73	0,70	6,05	1	0,014	0,18]0,04; 0,70[
15	-1,13	0,67	2,84	1	0,092	0,32]0,09; 1,20[
16	-0,20	0,66	0,10	1	0,756	0,82]0,23; 2,94[
17	-1,33	0,69	3,78	1	0,052	0,26]0,07; 1,01[
31	-1,61	0,72	4,94	1	0,026	0,20]0,05; 0,83[
34	-2,52	0,69	13,36	1	<0,001	0,80]0,02; 0,31[
48	-0,55	0,68	0,66	1	0,417	0,58]0,15; 2,17[
CONCENTRAÇÃO			41,55	3	<0,001		
A – controle	-6,06	1,20	25,66	1	<0,001	0,00]0,00; 0,02[
B – 250 NMP/ml	0,04	0,38	0,01	1	0,926	1,04]0,50; 2,16[
C – 500 NMP/ml	1,32	0,38	11,99	1	0,001	3,74]1,77; 7,88[
ALTURA	-0,66	0,17	15,18	1	<0,001	0,52]0,37; 0,72[
DIÂMETRO	-17,15	2,52	46,49	1	<0,001	0,00]0,00; 0,00[
Constante	7,83	1,03	57,92	1	<0,001	2505,48	

47 dias após inoculação (DAI)

Ajustamento do modelo (**com** todas as observações):

$$G^2(1)=84,23; p<0,001;$$

$$X^2_{\text{HL}}(8)=5,32; p=0,723;$$

$$R^2_{\text{CS}}=0,576;$$

$$R^2_{\text{N}}=0,769$$

Ajustamento do modelo (**sem** a presença dos *outliers*):

$$G^2(1)=83,55; p<0,001;$$

$$X^2_{\text{HL}}(8)=8,40; p=0,395;$$

$$R^2_{\text{CS}}=0,581;$$

$$R^2_{\text{N}}=0,774$$

Coefficientes *Logit* do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 47 DAI. (**com** todas as observações).

Variável	B	S.E.	X ² _{Wald}	g.l.	p-value	OR	I.C. a 95% para Exp(B)
FAMÍLIA			27,78	8	0,001		
8	-2,18	0,96	5,23	1	0,022	0,11]0,02; 0,73 [
14	-1,93	0,84	5,32	1	0,021	0,15]0,03; 0,75[
15	-2,07	0,85	5,91	1	0,015	0,13]0,02; 0,67 [
16	-0,38	0,94	0,16	1	0,686	0,68]0,11; 4,28[
17	-2,93	0,88	11,06	1	0,001	0,05]0,01; 0,30[
31	-1,81	0,87	4,36	1	0,037	0,16]0,03; 0,90[
34	-3,05	0,86	12,50	1	<0,001	0,05]0,01; 0,26[
48	0,12	0,99	0,02	1	0,900	1,13]0,16; 7,83 [
CONCENTRAÇÃO			47,86	3	<0,001		
A – controlo	-8,34	1,38	36,73	1	<0,001	0,00]0,00; 0,00 [
B – 250 NMP/ml	0,48	0,43	1,22	1	0,269	1,61]0,69; 3,76[
C – 500 NMP/ml	1,46	0,45	10,57	1	0,001	4,28]1,78; 10,30[
ALTURA	-0,75	0,20	14,74	1	<0,001	0,47]0,32; 0,69 [
DIÂMETRO	-20,80	2,92	50,90	1	<0,001	0,00]0,00; 0,00 [
Constante	10,84	1,41	59,41	1	<0,001	51220,44	

Coefficientes *Logit* do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 47 DAI. (**sem** a presença dos *outliers*).

Variável	B	S.E.	X ² _{Wald}	g.l.	p-value	OR	I.C. a 95% para Exp(B)
FAMÍLIA			28,03	8	<0,001		
8	-2,19	0,97	5,08	1	0,024	0,11]0,02; 0,75 [
14	-1,90	0,84	5,09	1	0,024	0,15]0,03; 0,78 [
15	-2,07	0,86	5,80	1	0,016	0,13]0,02; 0,68[
16	-0,42	0,95	0,20	1	0,657	0,66]0,10; 4,21 [
17	-3,33	0,91	13,38	1	<0,001	0,04]0,01; 0,21[
31	-1,83	0,88	4,28	1	0,039	0,16]0,03; 0,91[
34	-2,99	0,89	11,41	1	0,001	0,05]0,01; 0,28[
48	0,11	1,00	0,01	1	0,915	1,11]0,16; 7,87 [
CONCENTRAÇÃO			47,30	3	<0,001		
A – controlo	-8,54	1,41	36,73	1	<0,001	0,00]0,00; 0,00[
B – 250 NMP/ml	0,41	0,44	0,85	1	0,355	1,50]0,63; 3,58[
C – 500 NMP/ml	1,48	0,46	10,54	1	0,001	4,39]1,80; 10,74[
ALTURA	-0,83	0,20	16,65	1	<0,001	0,44]0,30; 0,65 [
DIÂMETRO	-21,23	3,03	49,07	1	<0,001	0,00]0,00; 0,00[
Constante	11,21	1,48	57,39	1	<0,001	74151,91	

92 dias após inoculação (DAI)

Ajustamento do modelo (**com** todas as observações):

$$G^2(1)=104,24; p<0,001;$$

$$X^2_{HL}(8)=9,75; p=0,283;$$

$$R^2_{CS}=0,544;$$

$$R^2_N=0,745$$

Ajustamento do modelo (**sem** a presença dos *outliers*):

$$G^2(1)=96,48; p<0,001;$$

$$X^2_{HL}(8)=6,39; p=0,603;$$

$$R^2_{CS}=0,588;$$

$$R^2_N=0,803$$

Coefficientes *Logit* do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, e DIÂMETRO, aos 92 DAI. (**com** todas as observações).

Variável	B	S.E.	X ² _{Wald}	g.l.	p-value	OR	I.C. a 95% para Exp(B)
FAMÍLIA			18,27	8	0,019		
8	-1,94	0,84	5,36	1	0,021	0,14]0,03; 0,74[
14	-1,02	0,72	1,98	1	0,159	0,36]0,09; 1,49[
15	-1,56	0,74	4,44	1	0,035	0,21]0,05; 0,90[
16	0,39	0,86	0,20	1	0,655	1,47]0,27; 7,98 [
17	-0,47	0,79	0,36	1	0,551	0,62]0,13; 2,94[
31	-1,55	0,71	4,76	1	0,029	0,21]0,05; 0,85[
34	-1,62	0,77	4,44	1	0,035	0,20]0,04; 0,89 [
48	0,56	0,88	0,40	1	0,529	1,74]0,31; 9,81 [
CONCENTRAÇÃO			90,69	3	<0,001		
A – controlo	-5,10	0,62	68,55	1	<0,001	0,01]0,00; 0,02 [
B – 250 NMP/ml	0,79	0,47	2,87	1	0,090	2,21]0,88; 5,54 [
C – 500 NMP/ml	0,79	0,48	2,79	1	0,095	2,21]0,87; 5,61[
DIÂMETRO	-20,95	2,65	62,35	1	<0,001	0,00]0,00; 0,00[
Constante	9,20	1,16	63,11	1	<0,001	9862,47	

Coefficientes *Logit* do modelo de regressão logística da variável ESTADO DA PLANTA, em função da FAMÍLIA, CONCENTRAÇÃO, ALTURA e DIÂMETRO, aos 35 DAI. (**sem** a presença dos *outliers*).

Variável	B	S.E.	X ² _{Wald}	g.l.	p-value	OR	I.C. a 95% para Exp(B)
FAMÍLIA			23,64	8	0,003		
8	-2,15	1,04	4,24	1	0,039	0,12]0,02; 0,90 [
14	-1,10	0,89	1,53	1	0,217	0,33]0,06; 1,91 [
15	-2,02	0,89	5,19	1	0,023	0,13]0,02; 0,75[
16	0,62	1,12	0,30	1	0,582	1,86]0,20; 16,83 [
17	-0,28	0,99	0,08	1	0,781	0,76]0,11; 5,30[
31	-1,42	0,91	2,42	1	0,120	0,24]0,04; 1,45[
34	-2,77	0,90	9,45	1	0,002	0,06]0,01; 0,37[
48	1,15	1,19	0,94	1	0,331	3,17]0,31; 32,38 [
CONCENTRAÇÃO			67,42	3	<0,001		
A – controlo	-7,16	0,96	55,48	1	<0,001	0,00]0,00; 0,00[
B – 250 NMP/ml	0,87	0,52	2,79	1	0,095	2,38]0,86; 6,61[
C – 500 NMP/ml	0,81	0,52	2,42	1	0,120	2,24]0,81; 6,20[
ALTURA	-0,23	0,22	1,11	1	0,292	0,79]0,52; 1,22 [
DIÂMETRO	-23,45	3,17	54,78	1	<0,001	0,00]0,00; 0,00[
Constante	11,01	1,46	56,80	1	<0,001	60389,62	