

## **1. Ideia inicial**

Meu projeto de mestrado foca na análise da resistência à murcha de *Fusarium* em uma população F<sub>8</sub> de feijoeiro comum envolvendo parentais contrastantes. No momento, ainda estou realizando os experimentos e não tenho resultados. Porém, planejei desenvolver uma análise fenotípica simulada para criar visualizações que representassem a distribuição da severidade da doença entre diferentes genótipos, comparando especificamente os parentais (IAC-Tybatã, resistente, e Branquinho, suscetível) com suas linhagens recombinantes (RILs).

A ideia inicial era criar um boxplot mostrando a severidade da doença na população, outro boxplot mostrando a progressão temporal da doença ao longo do experimento usando AACPD/AUCPD e um gráfico mostrando a distribuição geral da resistência na população. Além disso, está previsto no projeto análises estatísticas, que tentei implementar (mas com menos sucesso), e a parte de análise genotípica, que não é aplicável já que esse é um passo futuro na pesquisa.

## **2. Resultados**

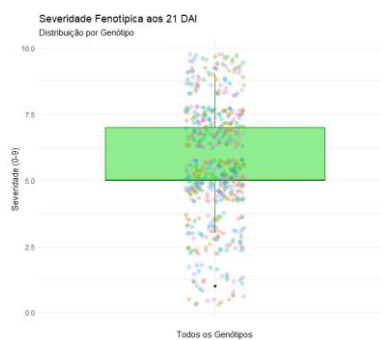
Desenvolvi um script, que depois adaptei para uma apresentação em Quarto. Para o estabelecimento inicial dos dados, pensei em criar um data frame novo com os 180 genótipos do trabalho + os 2 parentais, mas eu queria simular os dados com as identificações corretas de cada genótipos. Então, reutilizei um arquivo Excel pré-existente em formato longo, contendo a identificação dos genótipos, os dias de avaliação (7, 14 e 21 DAI), as repetições (1-3) e a nota de severidade vazia.

Para simular as notas, optei por criar uma coluna “perfil” e randomizar os genótipos predestinados a serem resistentes, moderados ou suscetíveis à doença. Com isso, foi possível distribuir o progresso de notas sem correr o risco de um genótipo receber uma nota alta no início do experimento, mas ganhar uma nota baixa no final, o que seria inconsistente. Nessa etapa, if/else ficou ruim, então descobri e apliquei o case\_when.

Com os dados simulados, fiz as análises descritivas, observando o desempenho da população com a média e desvio padrão dos resultados. Como a primeira coluna “perfil” foi feita apenas para amarrar os genótipos à um padrão

de progresso de notas, descartei ela e fiz uma nova coluna, agora com base no desempenho de cada genótipo pós-simulação.

Usei o ggplot para a visualização desses dados e, como esperado, a distribuição constante seguiu o padrão de muitos genótipos moderados, poucos suscetíveis e poucos resistentes (Figura 1). Admito que não entendo muito de estatística básica, não tive isso na graduação e farei uma disciplina a respeito apenas semestre que vem. Segui as análises sugeridas no meu projeto, porém, por algum motivo que ainda não entendi, a ANOVA deu um p extremamente alto. Na minha concepção, não faz muito sentido, já que os dados simulados em teoria seguem o padrão da doença, tendo genótipos em grupos diferentes de acordo com o perfil de resistência. Certamente é algo que irei estudar melhor no futuro.



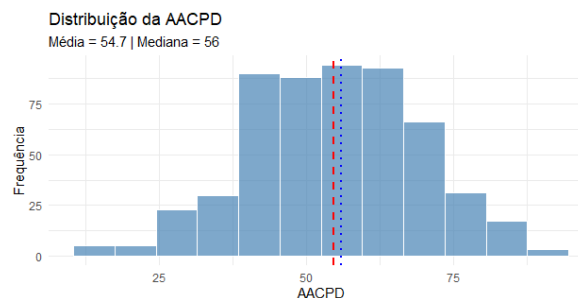
**Figura 1.** Distribuição da severidade da doença.

Assim como o processamento da análise descritiva, fiz o cálculo da Área Abaixo da Curva de Progresso da Doença (AACPD) e a visualização desses dados em boxplot (Figura 2). Tive que criar um novo data frame sem a coluna de perfil já que a função aucpd do agricolae não funcionou com o data frame original. Essa etapa foi bem-sucedida e com resultados esperados. Para testar, fiz ANOVA com os dados do AACPD, mas também teve um p gigante.



**Figura 2.** Distribuição de AACPD na população testada.

Por último, fiz a análise por genótipo, junto com a classificação deles de acordo com a média de notas e visualização com um histograma. (Figura 3). Testei um boxplot de todos os genótipos, mas ficou horrível então não utilizei ele na apresentação. Por último, fiz um boxplot apenas com os parentais e para a surpresa de ninguém, o parental resistente realmente foi resistente e o suscetível foi suscetível.



**Figura 3.** Histograma da distribuição de AACPD na população testada.

Um fenômeno que não entendi muito bem é que, ao criar tipples com os melhores e os piores genótipos, o apenas um teve o perfil “Altamente Resistente”, que foi o parental resistente. Mas o outro parental não foi “Altamente Suscetível”. Também não acho que a distribuição de todos os perfis foi correta, portanto, atribuo isso à simulação inicial dos dados.

### 3. Autorreflexão e sugestões

A experiência geral da disciplina foi positiva no geral. Consegui compreender bem o funcionamento do software e os pacotes apresentados certamente serão utilizados no futuro durante o meu desenvolvimento acadêmico. Para alguém que nunca teve contato com isso, foi uma surpresa conseguir acompanhar tão bem.

Entretanto, enfrentei dificuldades significativas na análise estatística, tanto na aula e no projeto. Mas atribuo essa dificuldade à minha falta de experiência e dificuldade pessoal.

Outro ponto “negativo” foi a organização das aulas à tarde e o ritmo geral das aulas teóricas. Em relação às teóricas, as aulas bem estruturadas (com destaque às ministradas pelo Prof. Maurício), eram perfeitamente compreensíveis em relação aos passos a serem seguidos. Mas, por ser uma aula que *\*requer\** um certo nivelamento, alguns alunos se distraíram/dispersavam e acabavam atrapalhando com conversar tangenciais.

Em alguns momentos, avancei sozinho no script da aula enquanto o professor atendia alunos que estavam com dificuldades básicas.

Já na aula prática, meu problema era o barulho desordenado e a falta de uma correção dos exercícios. Como prefiro trabalhar em casa, onde tenho ambiente controlado, perdi oportunidades de tirar dúvidas com o professor, já que cada semana abordava novo conteúdo sem espaço para revisão do anterior.

Como sugestão, futuramente as aulas práticas poderiam focar na discussão e correção das tarefas anteriores e até um nivelamento com mais foco em alunos que precisam de mais compreensão em coisas básicas, assim evitando dispersão na parte teóricas e atraso do fluxo da aula.

Agradeço pelo aprendizado.