

Departamento de Estatística - Universidade de Brasília

Aplicação do Método Scott-Knott em Estudo de Brusone no Trigo

Nelson Oliveira Pinheiro

Brasília
02/2017

Nelson Oliveira Pinheiro

Aplicação do Método Scott-Knott em Estudo de Brusone no Trigo

Projeto apresentado para obtenção do título
de Bacharel em Estatística ao Departamento
de Estatística da Universidade de Brasília

Orientador: Prof. Dr. Eduardo Monteiro de Castro Gomes

Brasília
02/2017

Sumário

| | | |
|------------|---|-----------|
| 1 | RESUMO | 5 |
| 2 | INTRODUÇÃO | 7 |
| 3 | REVISÃO DE LITERATURA | 9 |
| 3.1 | Definição do Modelo | 9 |
| 3.2 | Estimação dos parâmetros (ANOVA) | 10 |
| 3.3 | Tabela da Análise de Variância | 10 |
| 4 | MÉTODO DE SCOTT-KNOTT | 11 |
| 4.1 | Definição do Método Scott-Knott | 11 |
| 4.1.1 | Exemplo Simplificado de Aplicação | 12 |
| 4.1.2 | Exemplo - Amostra Balanceada | 12 |
| 5 | PACOTE SCOTTKNOTT NO R | 19 |
| 5.1 | Aplicando o Pacote ScottKnott | 19 |
| 5.1.1 | Pacote ScottKnott - Amostra Balanceada | 19 |
| 5.1.2 | Pacote ScottKnott - Amostra Desbalanceada | 20 |
| 6 | ADAPTAÇÃO DO PACOTE SCOTTKNOTT PARA DADOS DESBALANCEADOS | 23 |
| 7 | RESULTADOS E DISCUSSÕES | 25 |
| 7.1 | Análise Descritiva | 25 |
| 7.1.1 | Aplicação da Análise de Variância aos Dados | 28 |
| 7.2 | Análise Scott-Knott | 29 |
| 7.3 | Análise Scott-Knott por Ciclos | 33 |
| 7.3.1 | Ciclo Curto | 33 |
| 7.3.2 | Ciclo Médio | 36 |
| 7.3.3 | Ciclo Longo | 38 |
| 8 | CONCLUSÃO | 41 |
| | REFERÊNCIAS | 43 |
| | ANEXO A – R-SCRIPT | 45 |

1 Resumo

A doença Brusone vem se alastrando nos trigais brasileiros. Com o objetivo de encontrar genótipos resistentes a essa doença, a Embrapa realizou um experimento, em que foram coletadas parcelas de três metros cada, de forma completamente casualizada no período de três anos.

Para encontrar genótipos resistentes, é necessário realizar um agrupamento deles, considerando seu ciclo e ano. Assim, é possível realizar a comparação entre as médias.

O método de Scott-Knott se adequa ao estudo, pois trata-se de comparações múltiplas. Ele realiza a comparação de médias por meio de grupos que as classificam de forma homogênea, minimizando a soma de quadrados dentro dos grupos e maximizando entre eles.

O pacote ScottKnott, disponibilizado pelo software R, é essencial para a aplicabilidade do método, por se tratar de um banco de dados considerado como grande, assim dificultando a aplicação manual do método. Mas o pacote apresenta um viés na estatística do teste para o caso em que os dados são desbalanceados como ocorre no estudo. A adaptação ao pacote foi realizada com o objetivo de minimizar esse problema. O pacote foi aplicado aos dados e foram realizados testes para verificar a eficiência desta adaptação.

Por fim, foi realizado um estudo separando os tratamentos por ciclos, com o objetivo de minimizar a perda de genótipos que são considerados como bons de acordo com seus respectivos ciclos.

2 Introdução

Inicialmente, discutiremos sobre a brusone, que é também chamada de branqueamento de espiga. Conforme estudos realizados na Empresa Brasileira de Pesquisa Agropecuária (Embrapa), constatou-se que a brusone é uma doença causada pelo fungo *Pyricularia Grisea*, que recentemente se alastrou nos trigais brasileiros. Também possui como característica a penetração nas ramificações concentradas nas folhas da planta (trigo), impedindo o enchimento dos grãos, assumindo coloração esbranquiçada e secando as espigas rapidamente.

Uma vez que o trigo é atacado pela doença, ele se torna inútil para indústria de alimentos e pode ser utilizado na produção de derivados de baixo valor de mercado. Desde que a doença foi encontrada no Brasil pela primeira vez, no estado do Paraná, em 1985, buscaram-se genótipos resistentes à doença, que nem sempre surtem o efeito esperado. A busca intensa por materiais resistentes à brusone poucas vezes produzem o efeito esperado (sucesso), daí a necessidade de uma análise de experimentos para obter-se êxito.

A Embrapa do Distrito Federal realiza anualmente o plantio de trigo no cerrado no período de fevereiro a abril. Para o estudo, temos 406 tipos de genótipos (tratamento) no total. O experimento completamente casualizado foi realizado em cada repetição. Observou-se também o número de dias até que a planta espigasse. Por conta disso, deriva-se a variável ciclo, que é definida por curto, médio e longo. Por fim, o experimento foi feito no período de três anos (2011, 2012 e 2013), ocasionando o desbalanceamento dos dados, já que um tipo de genótipo pode não ter sequência no plantio do referente período analisado.

Por se tratar de um estudo de comparações múltiplas de médias, cujo objetivo é encontrar genótipos resistentes à brusone, a aplicação do método de Scott-Knott aos dados será fundamental. O método, separa as médias em grupos homogêneos, assim minimizando a soma de quadrados dentro dos grupos e maximizando entre eles.

A análise dos dados foi feita com base nas médias das replicações aplicada em diferentes cenários possíveis. O modelo de Análise de Variância não é muito eficiente para o experimento, posto que rejeitamos facilmente a hipótese nula do mesmo, ou seja, há ao menos um par de médias diferentes entre si. Utilizamos o método Scott-Knott, pois ele é eficiente na comparação das médias a um número alto de tratamentos.

O método Scott-Knott separa as médias dos tratamentos em grupos homogêneos, assim minimizando soma de quadrados dentro e maximizando entre os grupos e sem sobrepor esses grupos. Segundo Ramalho *et al.* (2000), o método Scott-Knott tem por intuito observar o quão é importante a divisão de k tratamentos em grupos homogêneos, baseando-se na razão de verossimilhança, implicando assim na maximização da soma de quadrados entre os grupos.

O software R possui um pacote definido como "ScottKnott", em que foi implementado por JELIHOVSKI *et al.* (2014) e executa o método de Scott-Knott para casos em que os dados são balanceados. Para utilizar o pacote em dados desbalanceados, é necessário realizar uma adaptação nas funções do pacote.

Por fim, iremos sugerir com base na análise de dados e nas especificações do método aplicado quais os tipos de genótipos são mais resistentes à doença brusone no geral e também separando-os pela variável ciclo.

3 Revisão de Literatura

O objetivo do experimento é a comparação das médias entre as repetições dos tratamentos. Quando se trata de comparação de médias, a análise de variância é importante. A modelagem dos dados deve ser feita para que possamos prosseguir com a análise de variância (ANOVA), em que é representada da seguinte maneira:

Tabela 1 – Modelagem dos dados obtidos pela Embrapa

| Tratamentos (Genótipo) | Repetições | Médias de Tratamentos |
|------------------------|------------------------------------|-----------------------|
| 1 | $y_{11} \ y_{12} \ y_{13}$ | $\bar{y}_{1.}$ |
| 2 | $y_{21} \ y_{22} \ y_{23}$ | $\bar{y}_{2.}$ |
| 3 | $y_{31} \ y_{32} \ y_{33}$ | $\bar{y}_{3.}$ |
| \vdots | $\vdots \quad \vdots \quad \vdots$ | \vdots |
| k | $y_{k1} \ y_{k2} \ y_{k3}$ | $\bar{y}_{k.}$ |

Para averiguar melhor o problema de variação dos fatores, pode-se realizar uma Análise de Variância que, segundo Montgomery (1997), é definida por k tratamentos de um único fator que desejamos comparar. A variável resposta de cada tratamento é aleatória e os dados podem ser organizados como na Tabela 1.

3.1 Definição do Modelo

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \text{ onde } \begin{cases} i = 1, 2, \dots, k \\ j = 1, 2, \dots, k. \end{cases}$$

Assumindo que ε_{ij} (erro aleatório iid) $\sim N(0, \sigma^2)$, temos que, μ é a média da variável resposta y_{ij} e τ_i é o efeito do i-ésimo tratamento. O objetivo é testar a igualdade das médias dos k tratamentos. Uma maneira equivalente para isso é testar as seguintes hipóteses:

$$\begin{cases} H_0 : \tau_1 = \tau_2 = \dots = \tau_k = 0 \\ H_1 : \tau_i \neq 0 \text{ para ao menos um } i \text{ onde } i=1, \dots, k. \end{cases}$$

Ou seja, na hipótese nula, temos que os efeitos dos tratamentos é igual a 0, que implica na igualdade das médias dos tratamentos. Já na hipótese alternativa, temos que ao menos um efeito é diferente de zero, o que implica na diferença de ao menos um par das médias dos tratamentos entre si.

3.2 Estimação dos parâmetros (ANOVA)

De acordo com Montgomery (1997), assumindo ε_{ij} (erro aleatório iid) $\sim N(0, \sigma^2)$, o modelo de Análise de Variância é definido da seguinte maneira:

Soma de Quadrado.

$$SQ_{Total} = \sum_{i=1}^k \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{N},$$

$$SQ_{Tratamento} = \sum_{i=1}^k y_{i.}^2 - \frac{y_{..}^2}{N}$$

$$SQ_{Residuo} = SQ_{Total} - SQ_{Tratamento}$$

Quadrado Médio.

$$QM_{Tratamento} = \frac{SQ_{Total}}{k-1}.$$

$$QM_{Residuo} = \frac{SQ_{Residuo}}{N-k}.$$

Estatística do Teste.

$$F_0 = \frac{QM_{Tratamento}}{QM_{Residuo}}.$$

Assumindo H_0 verdadeira, temos que $F_0 \sim F_{\alpha; k-1, N-k}$.

3.3 Tabela da Análise de Variância

Tabela 2 – Análise de Variância

| Causas de Variação | SQ | GL | QM | F |
|--------------------|-------------------|-----|-------------------|--|
| Tratamento | $SQ_{Tratamento}$ | k-1 | $QM_{Tratamento}$ | $\frac{QM_{Tratamento}}{QM_{Residuo}}$ |
| Erro | $SQ_{Residuo}$ | N-k | $QM_{Residuo}$ | |
| Total | SQ_{Total} | N-1 | | |

GL - Grau de Liberdade.

SQ - Soma de Quadrados.

QM - Quadrado Médio.

O teste se conclui com base em F_0 , pois caso $|F_0| > F_{crítico}$, então rejeita-se H_0 , isto é, dentre os pares de médias dos tratamentos, ao menos um par deles pode ser diferente. Caso contrário, não há evidências para rejeitar H_0 , ou seja, os pares de médias dos tratamentos são iguais. Outra maneira de analisarmos é através do p-valor: se ele for menor que o α estabelecido, rejeitamos H_0 . Para prosseguir na aplicação do método Scott-Knott, temos que obter F_0 pertencente à região crítica, de tal forma a rejeitarmos a hipótese de que as médias são iguais entre si.

4 Método de Scott-Knott

O método de Scott-Knott é eficiente nas aplicações em experimentos em que queremos comparar diferentes tratamentos. Seu intuito é separar as médias dos tratamentos em grupos homogêneos, assim minimizando soma de quadrados dentro; e maximizando-a entre os grupos, sem sobrepor-los. Para isso, é necessário ordenar as médias para que suas respectivas classificações sejam possíveis, determinar o número de partições possíveis, com a finalidade de realizar essa maximização entre grupos.

Segundo Ramalho *et al.* (2000), o método de Scott-Knott consiste em testar o quão significativo é a divisão de k tratamentos em dois grupos que maximizem a soma de quadrados entre eles utilizando a razão de verossimilhança.

Para Fisher (1958), é necessário considerar as $(k-1)$ partições possíveis, uma vez que quando as médias são ordenadas, não é necessário comparar a \bar{k}_{n+1} com a \bar{k}_n , pois é fundamental apenas uma comparação entre elas para determinar a sua classificação.

4.1 Definição do Método Scott-Knott

Segundo Scott Knott (1974), o método é definido da seguinte maneira:

Hipóteses.

$$\begin{cases} H_0 : G_1 = G_2 \\ H_1 : G_1 \neq G_2 \end{cases}$$

Assumindo G_1 e G_2 como sendo os totais dos dois grupos de médias com k_1 e k_2 tratamentos cada, ou seja, os totais de tratamentos em cada grupo, podemos definir a estatística do teste da seguinte maneira:

$$\lambda = \frac{\pi}{2(\pi-2)} * \frac{B_0}{\sigma_0^2}.$$

Assumindo H_0 verdadeira, temos que $\lambda \sim \chi^2_{(\frac{k}{\pi-2})}$

Em que π é o número irracional de valor aproximado 3,1415926 e B_0 representa a soma de quadrado. É necessário considerar o valor máximo entre as partições dos grupos e pode ser definida da seguinte maneira:

$$B_0 = \frac{T_1^2}{k_1} + \frac{T_2^2}{k_2} - \frac{(T_1+T_2)^2}{k_1+k_2}.$$

Onde, $T_1 = \sum_{i=1}^{k_1} y_{i.}$ e $T_2 = \sum_{i=k_1+1}^{k_2} y_{i.}$

σ_0^2 é o Estimador de Máxima Verossimilhança de σ_y^2 , definido da seguinte forma:

$$\sigma^2_0 = \frac{1}{k+v} \left[\sum_{i=1}^k (y_{i.} - y_{..})^2 + v s_{\bar{y}}^2 \right].$$

Considerando que, $s_{\bar{y}}^2 = \frac{QM_{Residuo}}{r}$ seja o estimador não viesado de $\sigma_{\bar{y}}^2$, r representa o número de repetições realizadas aos tratamentos e v os graus de liberdade do resíduo. O teste se conclui com base no λ obtido, ou seja, se $\lambda \leq \lambda_{crítico}$. Não há evidências para rejeitar a hipótese nula H_0 que consiste na igualdade entre grupos, caso contrário, rejeita-se H_0 . Então, tem-se indícios de que os grupos são diferentes.

4.1.1 Exemplo Simplificado de Aplicação

Considere um experimento com $k = 5$ tratamentos com médias ordenadas da seguinte maneira: $\bar{y}_1, \bar{y}_2, \bar{y}_3, \bar{y}_4, \bar{y}_5$, existem $(2^{(k-1)} - 1) = 15$ partições possíveis destas médias em dois grupos (G_1 e G_2). É importante considerar as $(5-1) = 4$ partições das médias ordenadas em dois grupos:

partição (1): \bar{y}_1 *versus* $\bar{y}_2 \bar{y}_3 \bar{y}_4 \bar{y}_5$
 partição (2): $\bar{y}_1 \bar{y}_2$ *versus* $\bar{y}_3 \bar{y}_4 \bar{y}_5$
 partição (3): $\bar{y}_1 \bar{y}_2 \bar{y}_3$ *versus* $\bar{y}_4 \bar{y}_5$
 partição (4): $\bar{y}_1 \bar{y}_2 \bar{y}_3 \bar{y}_4$ *versus* \bar{y}_5

Suponha que rejeitamos H_0 apenas para a partição (2), daí realizamos a partição dentro dos grupos da seguinte maneira:

subpartição(1.1): \bar{y}_1 *versus* \bar{y}_2 , se for significativo.
 subpartição(2.1): \bar{y}_3 *versus* $\bar{y}_4 \bar{y}_5$
 subpartição(2.2): $\bar{y}_3 \bar{y}_4$ *versus* \bar{y}_5

4.1.2 Exemplo - Amostra Balanceada

Com o objetivo de aplicabilidade do método aos dados disponibilizados pela Embrapa, selecionou-se uma amostra de tamanho $k=5$, assim como no exemplo 3.1.1. Foram aplicados os testes F (Anova) e o de Tukey antes de adentrar ao método Scott-Knott, a fim de verificar se havia a necessidade de aplicar o Scott-Knott para comparar as médias entre as repetições dos tratamentos da amostra. Após isso, notou-se que o método de Scott-Knott é necessário e assim ele foi aplicado de maneira manual, isto é, sem a utilização do pacote ScottKnott.

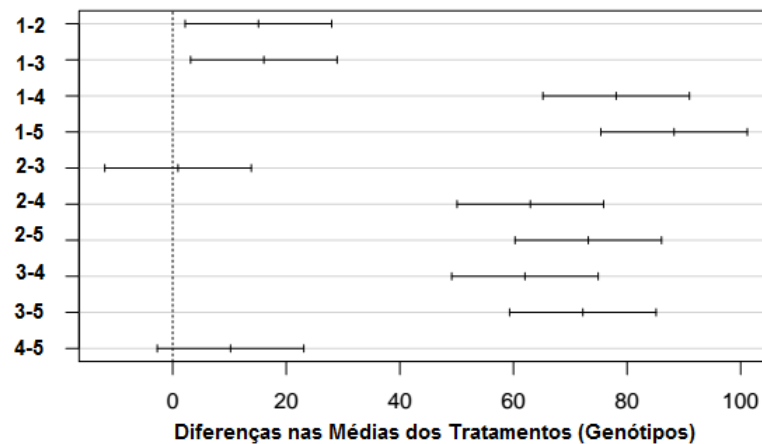
Tabela 3 – Amostra dos dados da Embrapa.

| | Genotipo | espigamento | Média | rep1 | rep2 | rep3 | ciclo | ano |
|---|--------------------|-------------|-------|-------|-------|-------|-------|------|
| 1 | M6 Sinthetic | 60.50 | 25.40 | 32.14 | 15.79 | 28.26 | curto | 2011 |
| 2 | CASS03GH00084S | 48.50 | 97.62 | 97.98 | 98.26 | 96.61 | curto | 2011 |
| 3 | CPAC_05320 | 45.00 | 87.42 | 84.83 | 88.24 | 89.19 | curto | 2012 |
| 4 | OPATA_85 | 56.00 | 24.50 | 25.69 | 27.40 | 20.24 | curto | 2012 |
| 5 | CBFUSARIUM_ENT_006 | 74.00 | 9.34 | 4.00 | 12.90 | 11.11 | medio | 2013 |

Tabela 4 – Tabela de resultados ANOVA (Amostra).

| Causas de Variação | SQ | GL | QM | F | P-Valor |
|--------------------|--------|----|-------|-----|------------|
| Tratamento | 19.717 | 4 | 4.929 | 214 | 1.22e(-09) |
| Erro | 230 | 10 | 23 | | |

A tabela de ANOVA mostra que as diferenças entre as médias são altamente significativas, já que rejeitamos a hipótese nula de que as médias são iguais entre si, pois obtivemos um $p - \text{valor} < \alpha$ (estabelecendo $\alpha = 0.05$). Podemos concluir que há uma diferença significativa entre o percentual médio de perda do tratamento (genótipo) por conta do brusone. Entretanto, a análise não terminou ainda, e para verificar entre exatamente quais pares de genótipos ocorrem diferenças significativas, usamos o teste de Tukey HSD.



Analisando o gráfico de Tukey HSD temos que as diferenças entre 1-2 (OPATA_85 - CBFUSARIUM_ENT_006), 1-3 (M6 Sinthetic - CBFUSARIUM_ENT_006), 1-4 (CPAC_05320 - CBFUSARIUM_ENT_006), 1-5 (CASS03GH00084S - CBFUSARIUM_ENT_006), 2-4 (CPAC_05320 - OPATA_85), 2-5 (CASS03GH00084S - OPATA_85), 3-4 (CPAC_05320 - M6 Sinthetic) e 3-5 (CASS03GH00084S - M6 Sinthetic) são significativas para $\alpha = 0,05$.

Logo, estes contribuem para a rejeição de H_0 na ANOVA.

Já que rejeitamos H_0 na Análise de Variância, uma opção para dar continuidade na análise é o teste de Scott-Knott. Para o teste, sabemos que sob H_0 a estatística do teste $\sim \chi^2(\alpha, \frac{k}{(\pi-2)})$, em que $\alpha = 0,05$ (pré estabelecido, mas pode ser alterado), k representa o número de tratamentos contidos nos grupos a serem comparados e $\pi \approx 3,1415926$. Teremos um número real no segundo fator da χ^2 , então uma opção é arredondá-lo para o maior número inteiro próximo a ele, assim obtemos uma variância mais conservadora.

Após realizar todos os procedimentos citados sobre o método, ordenando as médias dos tratamentos, obteve-se através do software R Core Team (2016) os seguintes resultados para as respectivas estatísticas do teste referente as 4 (k-1) partições dos grupos de médias:

- $\lambda_1 = 6,055361 < 11,070 \approx \chi^2_{(0.05,5)}$
- $\lambda_2 = 10,56301 < 11,070 \approx \chi^2_{(0.05,5)}$
- $\lambda_3 = 19,73621 > 11,070 \approx \chi^2_{(0.05,5)}$
- $\lambda_4 = 9,230370 < 11,070 \approx \chi^2_{(0.05,5)}$

A análise do teste consiste que se $\lambda \geq \chi^2(\alpha, \frac{k}{(\pi-2)})$, então rejeitamos a hipótese de que os grupos de médias são iguais. Podemos notar que para λ_3 foi utilizado o valor máximo das soma de quadrado dos possíveis pares de grupos a serem testados, assim temos as seguinte hipóteses:

$$\begin{cases} H_0 : G1(media_1, media_2, media_3) = G2(media_4, media_5) \\ H_1 : G1(media_1, media_2, media_3) \neq G2(media_4, media_5) \end{cases}$$

Rejeitamos a hipótese nula, ou seja, $G1(media_1, media_2, media_3)$ e $G2(media_4, media_5)$ são grupos de médias dos tratamentos que diferem entre si, em outros termos, a soma de quadrados B0 para este caso foi significativamente diferente. Devemos aplicar novamente o teste, realizando todo o procedimento para os respectivos subgrupos de λ_3 e seus respectivos dados.

Primeiramente, será trabalhado sobre o $G1(media_1, media_2, media_3)$, faremos a análise de variância, teste de TukeyHSD, subpartição do grupo e aplicação do método Scott-Knott através das possíveis 2 partições referente ao novo $k=3$. Assim, temos (k-1) partições, conforme Fisher (1958).

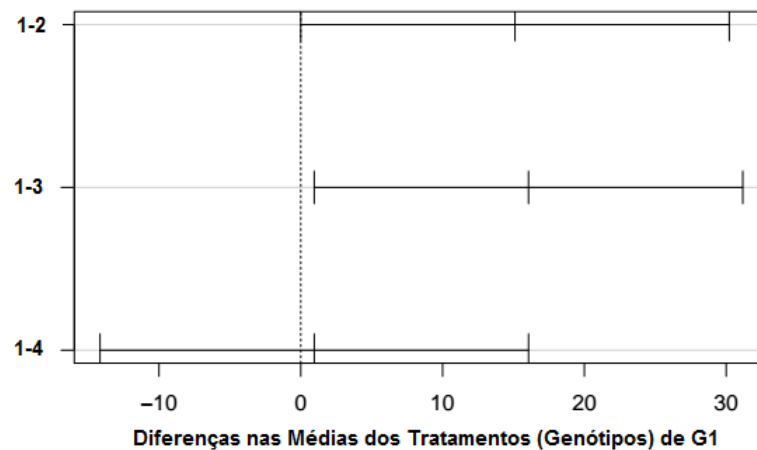
Tabela 5 – Resultados da ANOVA (Grupo de Médias G1)

| Causas de Variação | SQ | GL | QM | F | P-Valor |
|--------------------|-------|----|-------|-------|---------|
| Tratamento | 487,0 | 2 | 243,5 | 6,694 | 0,0296 |
| Erro | 218,3 | 6 | 36,38 | | |

$$\begin{cases} H_0 : \text{As médias são iguais entre si.} \\ H_1 : \text{Ao menos um par de médias difere entre si.} \end{cases}$$

Rejeitamos a hipótese nula, pois $0,0296 < 0,05$, isto é, as médias do grupo G1 diferem entre si para ao menos um para delas. O teste de TukeyHSD pode nos apontar quais são.

Figura 1 – Gráfico: Tukey HSD (Grupo de Médias G1)



O gráfico mostra que a diferença entre 1-3 (M6 Synthetic - CBFUSARIUM_ENT_006) é significativa para $\alpha = 0,05$. Logo, este contribui para a rejeição de H_0 na ANOVA. Assim, foi aplicado o método de Scott-Knott a G1 e obteve-se as seguintes estatísticas do teste:

- $\lambda_{1.1} = 8,527174 > 7,815 \approx \chi^2_{0.05,3}$
- $\lambda_{1.2} = 2,540287 < 7,815 \approx \chi^2_{0.05,3}$

A análise do teste consiste que se $\lambda \geq \chi^2(\alpha, \frac{k}{(\pi-2)})$, então rejeitamos a hipótese de que os grupos de médias são iguais. Podemos notar que para $\lambda_{1.1}$ foi utilizado o valor máximo das soma de quadrado dos possíveis pares de grupos a serem testados, logo temos as seguinte hipóteses:

$$\begin{cases} H0 : G1.1(media_1) = G1.2(media_2, media_3) \\ H1 : G1.1(media_1) \neq G1.2(media_2, media_3) \end{cases}$$

Podemos notar que para $\lambda 1.1$ rejeitamos a hipótese nula, ou seja, $G1.1(media_1)$ e $G1.2(media_2, media_3)$ são grupos de médias dos tratamentos que diferem entre si. Em outras palavras, a soma de quadrados B0 para este caso foi significativamente diferente. Devemos aplicar novamente o teste, realizando todo o procedimento para os respectivos subgrupos de $\lambda 1.1$ e seus respectivos dados.

Nesse caso, é necessário trabalhar somente com $G1.2(media_2, media_3)$, pois $G1.1(media_1)$ difere das demais médias de $G1.2$, logo ela já pertence a um grupo específico que será determinado mais a frente após determinar os outros grupos de todas as médias. Prosseguiremos então com a análise sobre $G1.2(media_2, media_3)$.

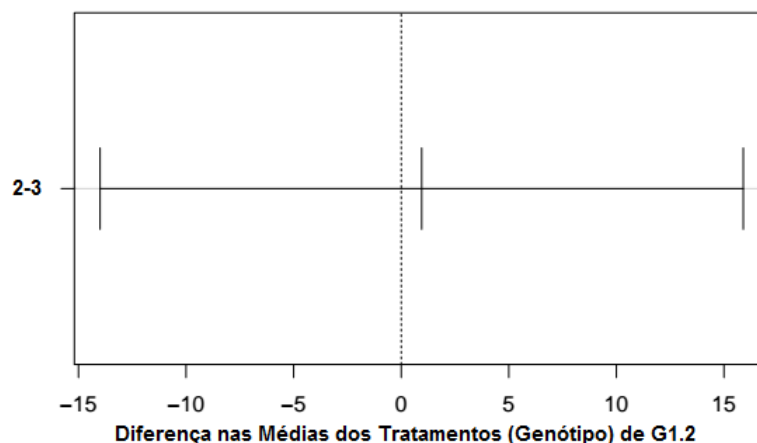
Tabela 6 – Resultados ANOVA (Grupo de Médias G1.2)

| Causas de Variação | SQ | GL | QM | F | P-Valor |
|--------------------|--------|----|-------|-------|---------|
| Tratamento | 1,36 | 1 | 1,36 | 0,031 | 0,868 |
| Erro | 173,92 | 4 | 43,48 | | |

$$\begin{cases} H0 : media_2 = media_3 \\ H1 : media_2 \neq media_3 \end{cases}$$

Não rejeitamos a hipótese nula de que $media_2$ e $media_3$ são iguais entre si, já que $0,868 > 0,05$. Poderíamos concluir que essas médias pertencem ao mesmo grupo com base na ANOVA, mas iremos prosseguir com os passos seguintes para termos certeza.

Figura 2 – Gráfico: Tukey HSD (Grupo de Médias G1.2)



O gráfico nos dá outra evidência de que a diferença entre 2-3 (M6 Synthetic - OPATA_85) não é significativa para $\alpha = 0,05$. Assim, foi aplicado o método de Scott-Knott a G1.2 e obteve-se as seguintes estatísticas do teste:

- $\lambda_{1.1.2} = 0,06399444 < 5,991 \approx \chi_{0.05,2}^2$

$$\begin{cases} H_0 : G_{1.2.1}(media_2) = G_{1.2.2}(media_3) \\ H_1 : G_{1.2.1}(media_2) \neq G_{1.2.2}(media_3) \end{cases}$$

Podemos finalmente concluir que para $\lambda_{1.1.2}$ não rejeitamos H_0 , ou seja, $G_{1.2.1}(media_2)$ e $G_{1.2.2}(media_3)$ são iguais entre si, logo essas duas médias pertencem ao mesmo grupo. Agora será feita a análise referente a $G_2(media_4, media_5)$. E a análise de variância, teste de TukeyHSD, subpartição do grupo e aplicação do método Scott-Knott, através das possíveis 1 partições referente ao novo $k=2$, assim temos $(k-1)$ partições, segundo Fisher (1958).

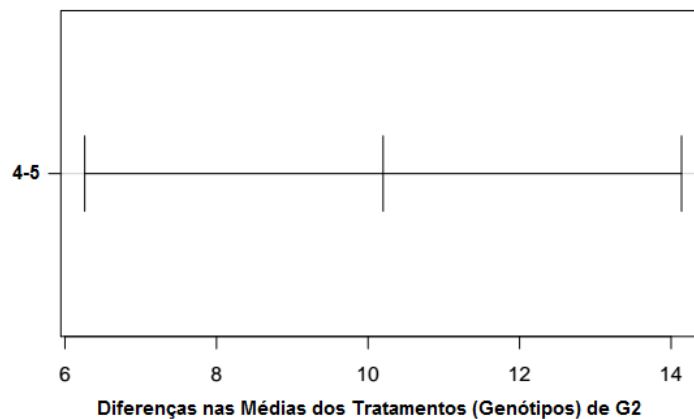
Tabela 7 – Resultados da ANOVA (Grupo de Médias G2)

| Causas de Variação | SQ | GL | QM | F | P-Valor |
|--------------------|--------|----|--------|-------|---------|
| Tratamento | 156,04 | 1 | 156,04 | 51,69 | 0,00198 |
| Erro | 12,07 | 4 | 3,02 | | |

$$\begin{cases} H_0 : media_4 = media_5 \\ H_1 : media_4 \neq media_5 \end{cases}$$

Rejeitamos a hipótese nula, pois $0,00198 < 0,05$, isto é, as médias do grupo G2 diferem entre si. O teste de TukeyHSD pode nos dar uma visão melhor sobre isso.

Figura 3 – Gráfico: Tukey HSD (Grupo de Médias G2)



O gráfico mostra que a diferença entre 4-5 (M6 CASS03GH00084S-CPAC_05320) é significativa para $\alpha = 0,05$. Logo, este contribui para a rejeição de H_0 na ANOVA. Assim, foi aplicado o método de Scott-Knott a G_1 e obteve-se as seguintes estatísticas do teste:

- $\lambda_{2.1} = 7.662603 > 5,991 \approx \chi^2_{0.05,2}$

$$\begin{cases} H_0 : G_{2.1}(media_4) = G_{2.2}(media_5) \\ H_1 : G_{2.1}(media_4) \neq G_{2.2}(media_5) \end{cases}$$

Rejeitamos a hipótese nula de que os subgrupos de G_2 são iguais, ou seja, $G_{2.1}(media_4)$ pertence a um grupo e $G_{2.2}(media_5)$ a outro. Agora podemos classificar todas as médias em seus respectivos grupos de acordo com a classificação de agrupamento do método Scott-Knott e a ordenação das médias. Assim, foi obtido a seguinte classificação das médias dos tratamentos (genótipos) por grupos:

Tabela 8 – Resultado da aplicação do Método Scott-Knott manual.

| Tratamentos | Valor | Média | Grupo |
|--------------------|-------|-----------|-------|
| CASS03GH00084S | 97,62 | $media_5$ | A |
| CPAC_05320 | 87,42 | $media_4$ | B |
| M6 Sinthetic | 25,40 | $media_3$ | C |
| OPATA_85 | 24,45 | $media_2$ | C |
| CBFUSARIUM_ENT_006 | 9,34 | $media_1$ | D |

5 Pacote ScottKnott no R

Segundo JELIHOVSKI *et al.* (2014), o algoritmo de agrupamento "ScottKnott" realiza a aplicação do método de Scott, A.J. e Knott, M. para distinguir as médias dos tratamentos que causam variação na variável resposta após a análise de variância.

O algoritmo apresenta uma característica que o define diferente dos demais métodos de comparação de diversas médias, com um teste de razão de verossimilhança para determinar quando parar o processo de agrupamento para criar subconjuntos não-sobrepostos.

Por conta disso, ocorre o problema maior para a aplicação do algoritmo aos dados fornecidos pela Embrapa, pois eles são desbalanceados, isto é, esse estimador varia de acordo com o número de repetições do tratamento, assim alterando a estatística do teste. Inferiu-se, após testes com dados desbalanceados, que o pacote não obtém êxito para esses casos, pois ele desconsidera a variação do estimador de Máxima Verossimilhança, ou seja, considera apenas o caso em que os tratamentos apresentam número de repetições iguais entre todos.

5.1 Aplicando o Pacote ScottKnott

Executando o pacote, foram considerados os dados dos exemplos do capítulo 3, com o objetivo de verificar se o pacote adere a teoria e vice-versa. Para os casos em que os dados são balanceados, isto é, quando os tratamentos possuem número de repetições iguais e para os casos em que os dados são desbalanceados, em que no cenário há parcelas faltantes.

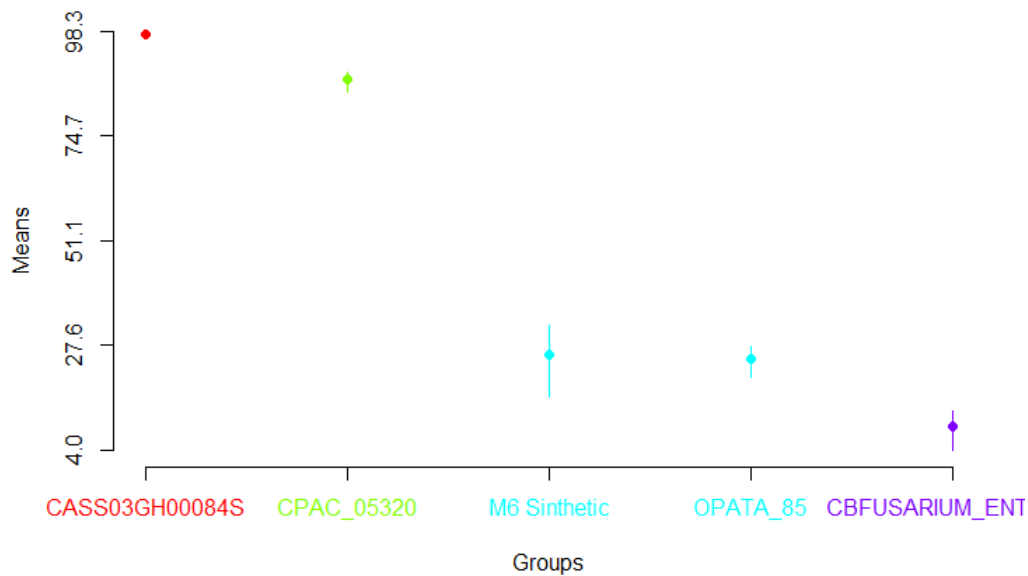
5.1.1 Pacote ScottKnott - Amostra Balanceada

Tabela 9 – Resultado do Pacote ScottKnott (Amostra Balanceada).

| Tratamentos | Médias | SK(5%) |
|--------------------|--------|--------|
| CASS03GH00084S | 97.62 | a |
| CPAC_05320 | 87.42 | b |
| M6 Sinthetic | 25.40 | c |
| OPATA_85 | 24.44 | c |
| CBFUSARIUM_ENT_006 | 9.34 | d |

Nota-se que obtemos os mesmos grupos de médias do exemplo. O pacote também nos proporciona uma análise gráfica, como podemos observar na figura 4.

Figura 4 – Gráfico: Médias agrupadas por cores (Amostra Balanceada)



Pode-se inferir que, o pacote funciona de maneira analoga à teoria de Scott Knott (1974), ambos para o caso em que os dados estão balanceados.

5.1.2 Pacote ScottKnott - Amostra Desbalanceada

Figura 5 – Mensagens de aviso na aplicação de SK na amostra desbalanceada.

```
warning messages:
1: In if ((lam < valchisq) | (ord1 == k)) { :
  the condition has length > 1 and only the first element will be used
2: In if (lam > valchisq) { :
  the condition has length > 1 and only the first element will be used
3: In if (lam < valchisq) { :
  the condition has length > 1 and only the first element will be used
4: In if ((lam < valchisq) | (ord1 == k)) { :
  the condition has length > 1 and only the first element will be used
5: In if (lam > valchisq) { :
  the condition has length > 1 and only the first element will be used
6: In if (lam < valchisq) { :
  the condition has length > 1 and only the first element will be used
```

Ao aplicarmos o pacote para o caso em que os dados são desbalanceados, obtemos as seguintes mensagens de aviso, conforme mostrado acima, que indicam a não aderência ao desbalanceamento, ou seja, no cálculo de σ_0^2 depende do estimador da variância das médias dos tratamentos $s_y^2 = \frac{QM_{Residuo}}{r}$ e consequentemente esse estimador depende das repetições dos tratamentos. Para o caso balanceado isso não é um problema, pois r é sempre o mesmo e como no caso desbalanceados esse r varia de acordo com o número de

repetições que há no tratamento, logo obtemos um vetor com valores diferentes de σ_0^2 e consequentemente o mesmo ocorre na estatística do teste λ , pois é dada em função de σ_0^2 e B0.

Nota-se que o pacote não consegue distinguir os demais valores de λ e utiliza apenas o primeiro para realizar o procedimento de inferência sobre a estatística do teste. Uma adaptação para que o pacote selecione o $s_{\bar{y}}^2$ que representa o ponto de corte do grupo seria suficiente para resolver essa questão de reconhecimento dos elementos, mas não estaria correto, pois estaríamos viesando a estatística do teste, uma vez que esse $s_{\bar{y}}^2$ não representa todos os $s_{\bar{y}}^2$ de fato.

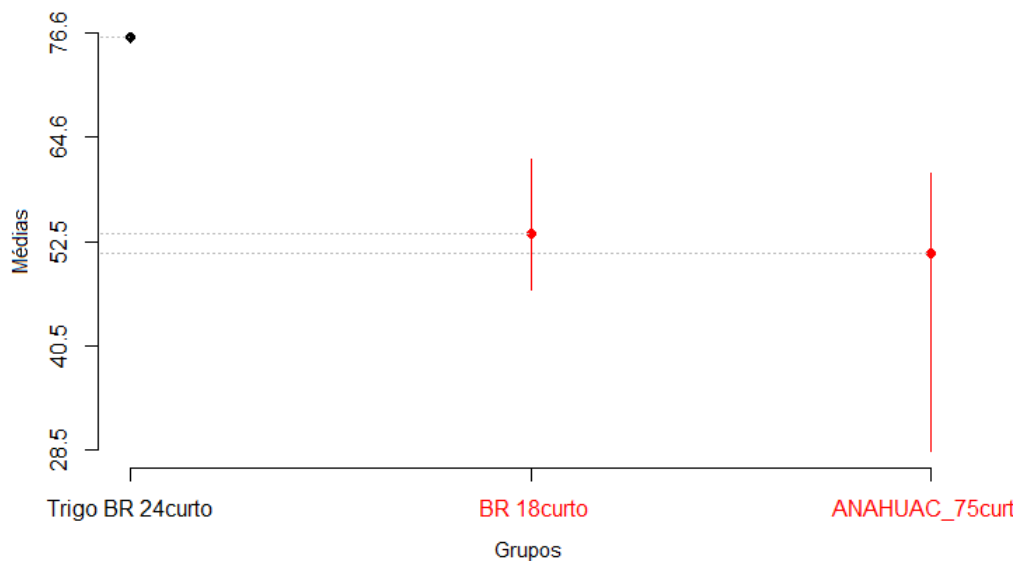
Mesmo com essa complicação, o pacote seleciona o primeiro elemento de λ e faz o processo, gerando os resultados a seguir.

Tabela 10 – Resultado do Pacote ScottKnott (Amostra Desbalanceada).

| Tratamentos | Médias | SK(5%) |
|------------------|--------|--------|
| Trigo BR 24curto | 76.20 | a |
| BR 18curto | 53.44 | b |
| ANAHUAC_75curto | 51.14 | b |

Nota-se que obtemos os mesmos grupos de médias do exemplo. Isso pode ter ocorrido pelo fato de que se trata de uma amostra pequena e a variabilidade de σ_0^2 para as respectivas repetições dos tratamentos é pequena, logo não afeta muito na estatística do teste. O pacote também nos proporciona uma análise gráfica, como podemos observar na figura 6.

Figura 6 – Gráfico: Médias agrupadas por cores (Amostra Desbalanceada)



É necessário realizar uma adaptação ao pacote para que ele seja aplicado em casos que os dados apresentam desbalanceamento.

6 Adaptação do Pacote ScottKnott para Dados Desbalanceados

Como já mencionado, faz-se necessário uma adaptação ao método Scott-Knott. Vimos também que a estatística do teste é definida da seguinte maneira:

$$\lambda = \frac{\pi}{2(\pi-2)} * \frac{B0}{\sigma_0^2} \sim \chi^2\left(\frac{k}{(\pi-2)}\right).$$

Há um viés nessa estatística do teste quando os dados apresentam desbalanceamento, visto que o número de repetições varia de acordo com cada tratamento e isso interfere no cálculo da estimativa da variância das médias, que é definida da seguinte maneira:

$$s_{\bar{y}}^2 = \frac{QM_{Residuo}}{r}$$

De acordo com CONRADO *et al.* (2017), essa estatística do teste é válida apenas sob um número igual de observações para cada tratamento ($n1 = n2 = \dots = nk$), ou seja, para o caso balanceado em que o número de repetições é idêntico para todos os tratamentos. Além disso, sob um design experimental equilibrado, $s_{\bar{y}}^2$ tem o mesmo valor para cada tratamento e nos leva à seguinte equação:

$$s_c^2 = \frac{1}{k} \left[\sum_{i=1}^k s_{\bar{y}_i} \right]$$

Trata-se do ajuste proposto para adequar o método Scott-Knott à questão do desbalanceamento, isto é, a estimativa da variância das médias dos tratamentos é definida pela sua média, mas variando o número de repetições de cada tratamento. Logo, temos a seguinte expressão para o novo Estimador de Máxima Verossimilhança de $\sigma_{\bar{y}}^2$:

$$\sigma_{0c}^2 = \frac{1}{k+v} \left[\sum_{i=1}^k (y_{i.} - y_{..})^2 + v s_c^2 \right].$$

À medida que o número de repetições por tratamento diminui, o σ_{0c}^2 aumenta, assim diminuindo o valor da estatística do teste, de modo a ter uma probabilidade menor de rejeitar a hipótese nula, isto é, a probabilidade de rejeitar H_0 dado que H_0 é verdadeira se torna pequena. Nota-se que o ajuste é adequado para o caso balanceado, uma vez que a estimativa da variância terá apenas um valor e sua média será representada por essa mesma estimativa.

Para validar a adaptação para casos desbalanceados, CONRADO *et al.* (2017) realizou método de Monte Carlo (Metropolis e Ulan, 1949), em que foram simulados mais de 40 milhões de experiências para vários níveis desequilibrados combinados com vários valores de α .

Concluiu-se que o ajuste proposto para dados desbalanceados apresentou desempenho semelhante ao procedimento original sob casos desbalanceados, mantendo um controle satisfatório do erro de tipo I do experimento e desempenho melhorado em $\alpha \geq 0.05$. Esse ajuste aumenta o espectro para o uso do teste, proporcionando ao pesquisador uma alternativa ao MCP (Procedimento de Comparações Múltiplas), mesmo sob experimentos com parcelas faltantes (desbalanceamento).

7 Resultados e Discussões

7.1 Análise Descritiva

O banco de dados disponibilizado pela Embrapa contém 490 observações no total, 413 tratamentos e 6 variáveis. Com o objetivo de encontrar genótipos resistentes a essa doença, a Embrapa realizou um experimento, onde foram coletados o percentual de perda da planta em relação a doença, nas parcelas de três metros obtidas de forma completamente casualizada em cada ano. No ano de 2011 sobreviveram 146 tipos de tratamentos, já no ano de 2012 esse número foi de 194 e por fim no ano de 2013 tivemos 149 tipos sobreviventes. A variável "repetição" que caracteriza o percentual de perda da planta em função da doença é observada em três momentos, conforme podemos observar na figura 1. Na base de dados, há uma observação que não foi coletada os dados referentes à variável "repetição". Logo, o genótipo *CIGM89.559-1B* foi removido do banco de dados, pois o objetivo é a análise de comparação de médias dos tratamentos e sem os valores das repetições, o tratamento não é significativo para o experimento.

Figura 7 – Sumário dos dados disponibilizados pela Embrapa.

| Genotipo | espigamento | rep1 | rep2 |
|----------------|---------------|------------------|----------------|
| AFRICA_43 : 3 | Min. :41.50 | Min. : 0.6623 | Min. : 0.00 |
| ANGAS : 3 | 1st Qu.:50.00 | 1st Qu.: 35.4286 | 1st Qu.: 36.88 |
| BH_1146 : 3 | Median :57.00 | Median : 58.6716 | Median : 57.57 |
| BRS_179 : 3 | Mean :59.31 | Mean : 59.6839 | Mean : 60.43 |
| BRS_220 : 3 | 3rd Qu.:66.00 | 3rd Qu.: 92.3664 | 3rd Qu.: 97.13 |
| BRS_ANGICO: 3 | Max. :95.50 | Max. :100.0000 | Max. :100.00 |
| (Other) :472 | | NA's :1 | NA's :2 |
| rep3 | ciclo | ano | |
| Min. : 0.00 | curto:321 | Min. :2011 | |
| 1st Qu.: 42.22 | longo: 45 | 1st Qu.:2011 | |
| Median : 58.53 | medio:124 | Median :2012 | |
| Mean : 60.71 | | Mean :2012 | |
| 3rd Qu.: 94.66 | | 3rd Qu.:2013 | |
| Max. :100.00 | | Max. :2013 | |
| NA's :1 | | | |

Foram desconsiderados os casos em que o percentual de perda do trigo em relação à brusone foi de 100% nas três repetições em ao menos um ano, pois nesses casos temos um trigo totalmente vulnerável à doença e queremos encontrar tratamentos resistentes a ela. Foi aplicado uma concatenação aos tratamentos e ciclo, de forma a distinguir quando um tratamento é aplicado em mais de um tipo de ciclo, e obtivemos 330 níveis. Na figura 8, podemos observar a forma final do banco de dados, a fim de dar início ao experimento.

| | | |
|--------------------------|-----------------------|----------------|
| Sonalikacurto | TRIGO_BR_23curto | PF_100348curto |
| Precoz ParanÃ; Intacurto | FUNDACEP_30curto | PF_100368curto |
| M6 Sintheticcurto | BRS_TangarÃ;curto | PF_100399curto |
| Pampeano Intacurto | PF_93159curto | PF_100409curto |
| Bul Bulcurto | _307curto | PF_100410curto |
| Embrapa 22curto | AFRICA_43curto | PF_100413curto |
| LR 12 (RL 6011)curto | PF_022203medio | PF_100415curto |
| Trigo BR 33curto | BRS_COMBOATAmidio | PF_100437curto |
| ANGAScurto | BRS_GUABIJUmedio | PF_100480curto |
| BRS Guamirimcurto | G_1179-37medio | PF_100508curto |
| Gigas 1381curto | MAYA_74medio | PF_100538curto |
| IAC 24curto | KENYA_FARMERmedio | PF_100544curto |
| IAC 5curto | WUHAN_3medio | PF_100605curto |
| PF 89326curto | GREECE_78310-A-3medio | PF_100618curto |
| Caillouxcurto | MELCHIORmedio | PF_100634curto |
| PF 92393curto | MAYA_74medio | PF_100706curto |
| BRS TimbaÃ°vacurto | GRANAROLOlongo | PF_100729curto |
| RS - 1 Fenixcurto | ANGAScurto | PF_100742curto |
| PF 010255curto | BRS_GUAMIRIMcurto | PF_100744curto |
| PF 0262curto | BUL_BULcurto | PF_100754curto |
| BRS 194medio | CAILLOUXcurto | PF_100774curto |
| BRS ComboatÃ£medio | CHHOTI_LERMAcurto | PF_100785curto |
| Relinmedio | HUANCAcurto | PF_100798curto |
| Narino 59longo | IAC_5curto | PF_100833curto |
| Snogglongo | MGS1_ALIANÃ‡Acurto | PF_100838curto |
| AFRICA_43curto | PF_050633curto | PF_100857curto |
| IAC_5curto | PF_050667curto | PF_100860curto |
| JANZcurto | Salcantaycurto | PF_100861curto |
| PF_050757curto | Vulture_SIBcurto | PF_100886curto |
| PF_050759curto | PF_090358curto | PF_100902curto |
| IPF_71349curto | PF_090371curto | PF_100910curto |
| CD_105curto | PF_090386curto | PF_100915curto |
| CNT_10curto | PF_100234curto | PF_100917curto |
| JACUIcurto | PF_100242curto | PF_100919curto |
| PF_815299curto | PF_100344curto | PF_100920curto |
| PF_100936curto | PF_100336medio | |

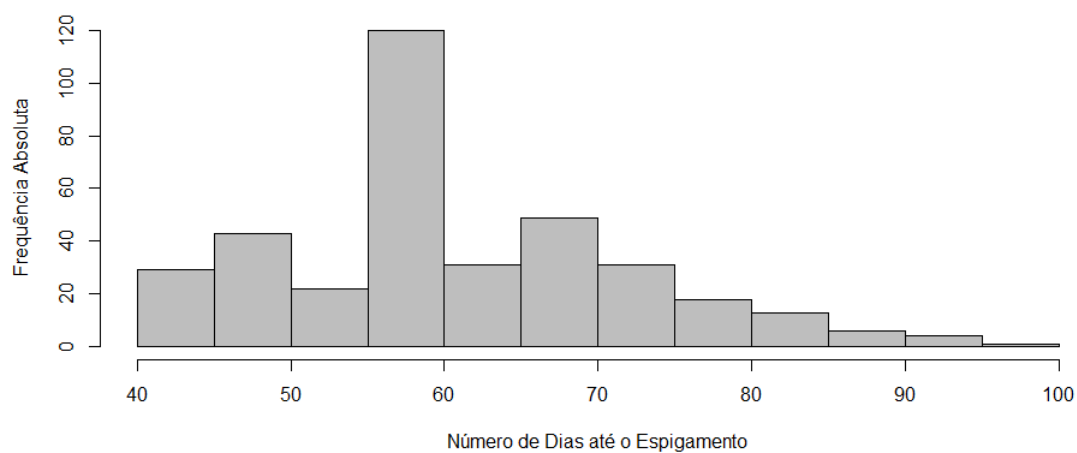
No quadro acima, estão listados os 107 tratamentos que foram excluídos por apresentarem perda de 100% nas três repetições em ao menos um ano. Ao ocorrer esse evento foi excluído o tratamento dos anos que não apresentou a perda de 100% em outro ano, pois quando o mesmo é totalmente perdido em um ano, ao ser considerado como perda média ou baixa em outro pode significar apenas um fenômeno não controlado ("sorte").

Figura 8 – Sumário dos dados disponibilizados pela Embrapa Limpos.

| Genotipo | espigamento | rep1 | rep2 |
|------------------|------------------|------------------|----------------|
| Length:367 | Min. :42.50 | Min. : 0.6623 | Min. : 0.00 |
| Class :character | 1st Qu.:55.00 | 1st Qu.: 30.5957 | 1st Qu.: 30.16 |
| Mode :character | Median :59.00 | Median : 50.2463 | Median : 50.00 |
| | Mean :60.96 | Mean : 48.5857 | Mean : 49.69 |
| | 3rd Qu.:67.00 | 3rd Qu.: 65.8659 | 3rd Qu.: 67.31 |
| | Max. :95.50 | Max. :100.0000 | Max. :100.00 |
| | | | NA's :1 |
| rep3 | ciclo | ano | |
| Min. : 0.00 | Length:367 | Min. :2011 | |
| 1st Qu.: 32.54 | Class :character | 1st Qu.:2011 | |
| Median : 51.97 | Mode :character | Median :2012 | |
| Mean : 49.86 | | Mean :2012 | |
| 3rd Qu.: 64.29 | | 3rd Qu.:2012 | |
| Max. :100.00 | | Max. :2013 | |

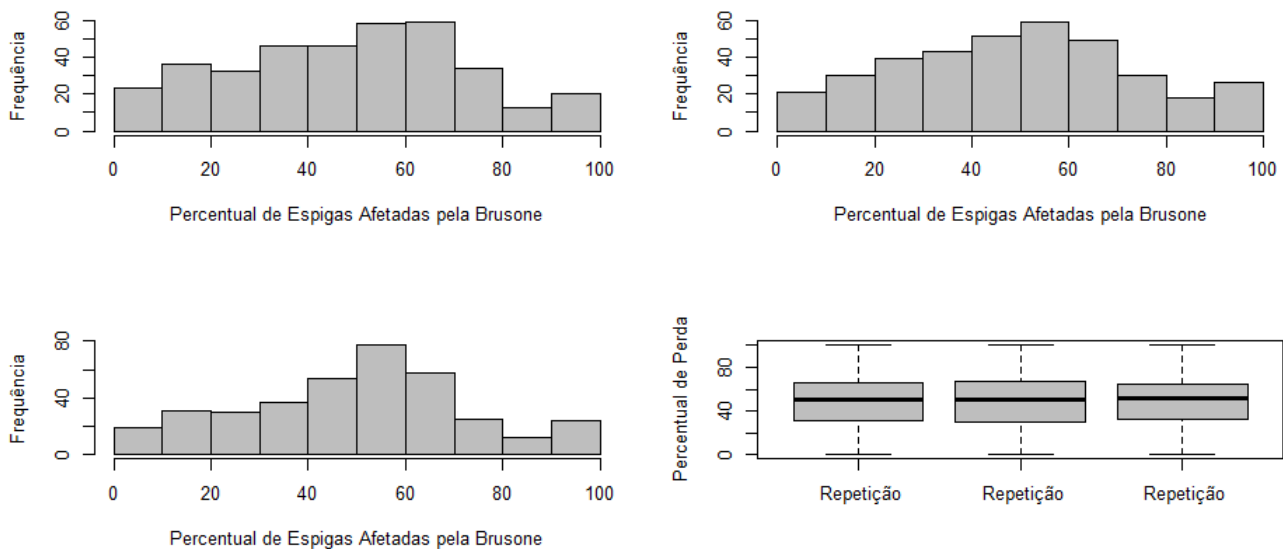
A representação gráfica e análise de cada variável é dada da seguinte forma:

Figura 9 – Gráfico: Número de dias até o espigamento do trigo.



Analisando o histograma para número de dias até que ocorra o espigamento do trigo, podemos perceber que o ciclo curto. Com a ocorrência do espigamento entre 40 e 60 dias ocorre com mais frequência que os demais ciclos.

Figura 10 – Gráfico: Comportamento das repetições



Analisando os histogramas da figura 10, podemos observar que nas repetições há uma concentração no percentual de perda da planta entre 40 e 60 %. Isso pode ser devido a qualidade da coleta dos dados, por exemplo, quando se demora a coletar o percentual de um genótipo que não foi tão afetado pela doença no ato da colheita. Por conta disso, a doença pode se propagar após a seleção.

Analisando o box plot comparativo na mesma figura, nota-se um paralelismo entre as medianas, ou seja, medianas de valores próximos. Isso confirma o que foi apontado na análise dos histogramas, já que os valores são aproximadamente entre 40 e 60 %.

7.1.1 Aplicação da Análise de Variância aos Dados

Utilizando o software R, foi obtido a seguinte tabela de análise de variância:

Tabela 11 – Resultados ANOVA - Dados EMBRAPA

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|------------|-----|-----------|---------|---------|-------------|
| Tratamento | 310 | 278350.67 | 897.91 | 1.92 | 3.27e-13*** |
| Resíduo | 789 | 368549.00 | 467.11 | | |

Estabelecendo $\alpha = 0,05$, temos que $3.27e - 13 * * * < \alpha$, ou seja, rejeitamos a hipótese de que as médias dos tratamentos são iguais entre si, logo as médias diferem para ao menos um par.

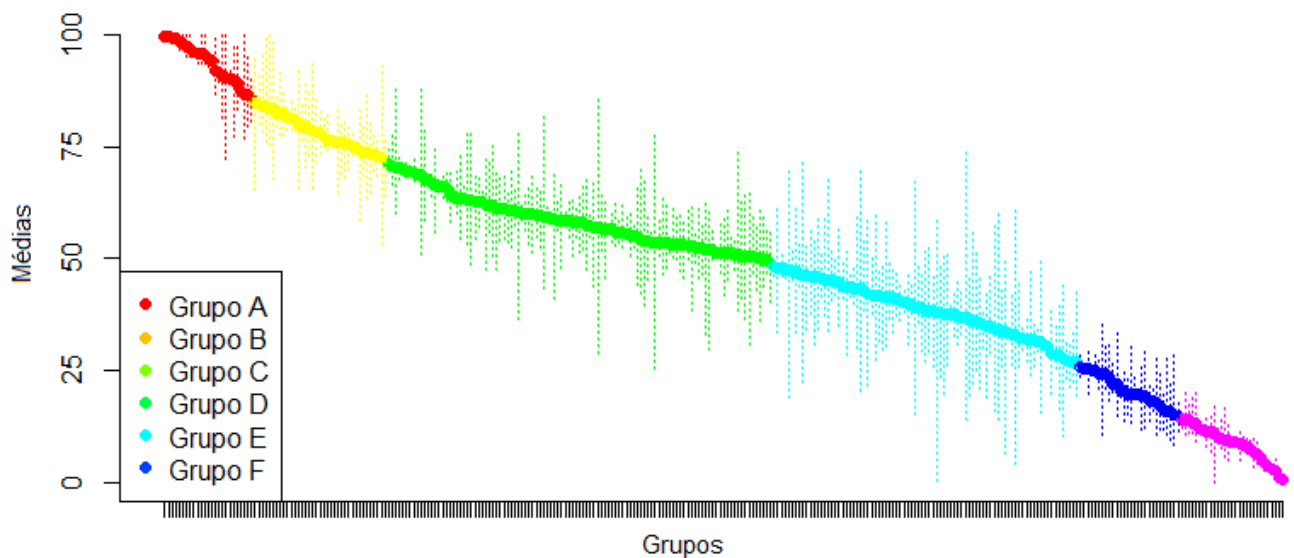
7.2 Análise Scott-Knott

A tabela abaixo sumariza o resultado do experimento, em que aplicado o método de Scott-Knott utilizando o software R Core Team (2016). Foram analisados as hipóteses de que os pares de grupos eram iguais ou se diferiam entre si, de maneira a agrupá-los no fim do experimento.

Tabela 12 – Sumário - Scott-Knott Aplicado aos Dados

| | Médias | SK(0.5%) |
|----------|---------|----------|
| Min. : | 0.6808 | a: 25 |
| 1st Qu.: | 33.8348 | b: 37 |
| Median : | 51.3846 | c:108 |
| Mean : | 50.1648 | d: 84 |
| 3rd Qu.: | 65.8013 | e: 28 |
| Max. : | 99.5614 | f: 29 |

Figura 11 – Gráfico: Médias agrupadas por cores (Dados Embrapa)



A forma de sumário para a apresentação dos resultados é adequada por se tratar de um experimento cujo número de tratamentos é considerado grande. O esboço de uma tabela que apresentasse todos os grupos e seus referentes percentuais de perda em relação à brusone seria algo que tomaria não somente espaço na questão de apresentação, mas também tempo no ato de inferência sobre o objetivo em questão.

Posto que nosso objetivo no experimento é identificar os genótipos do trigo que apresentam maior resistência à brusone, então é significativo esboçarmos o resultado contendo

os grupos de maior significância em relação a menor perda percentual do trigo quando afetado pela doença.

Assim, a fim de obter sucesso, estabelecemos para esse caso um corte em aproximadamente 14% de perda do trigo de tal maneira que apresenta-se no resultado apenas do grupo "f", apresentados na tabela abaixo:

Tabela 13 – Genótipos de menor percentual de perda em relação à brusone

| Tratamentos | Médias | SK(5%) | Quantidade de Anos |
|-------------------------|--------|--------|--------------------|
| PF 020458curto | 13.99 | f | 2011 |
| CASS03GH00077Slongo | 13.99 | f | 2012 |
| Har 604longo | 13.91 | f | 2011 |
| Lovrin 13longo | 13.47 | f | 2012 |
| PF_020062curto | 13.24 | f | 2012 |
| CBFUSARIUM_ENT_025curto | 12.11 | f | 2012,2013 |
| PF 010069longo | 11.85 | f | 2011 |
| BRS 220medio | 11.54 | f | 2012,2013 |
| CASW02GH00045Slongo | 11.22 | f | 2012 |
| CIGM921696medio | 10.85 | f | 2011 |
| BRS 177longo | 10.21 | f | 2012 |
| BRS Angicomedio | 9.73 | f | 2011 |
| PF_070475longo | 9.49 | f | 2012 |
| CBFUSARIUM_ENT_006medio | 9.34 | f | 2011 |
| CASW02GH00002Slongo | 9.18 | f | 2011 |
| Frontanamedio | 8.98 | f | 2011 |
| Klein Lucerolongo | 8.76 | f | 2011 |
| PF_926medio | 8.28 | f | 2012,2013 |
| CASW00GH00065Smedio | 7.96 | f | 2012,2013 |
| CBFusarium_ENT_016curto | 7.66 | f | 2012,2013 |
| Granarololongo | 7.02 | f | 2012,2013 |
| BRS 179curto | 6.45 | f | 2011 |
| CBFUSARIUM_ENT_014medio | 5.69 | f | 2012 |
| Melchiormedio | 4.57 | f | 2012 |
| Safiramedio | 3.84 | f | 2013 |
| Trintecincolongo | 3.02 | f | 2011 |
| SAFIRAlongo | 2.76 | f | 2012 |
| JesuÃtamedio | 1.09 | f | 2013 |
| CASW94Y00116Smedio | 0.68 | f | 2011 |

Como mencionado anteriormente, foi realizado um corte, com a finalidade de obter-se um resultado que apresentasse sucesso e dinamismo na apresentação. Mas ao realizarmos este procedimento em relação ao percentual de perda, estamos desprezando outra característica que apresenta significância, que é o ciclo dos genótipos.

Nesse caso, estamos suprimindo informação, pois há o descarte de genótipos que possuem perda percentual baixa de determinado ciclo. Isso significa uma perda grande, já que produtores rurais que optam por um tipo de ciclo no plantio do seu trigo.

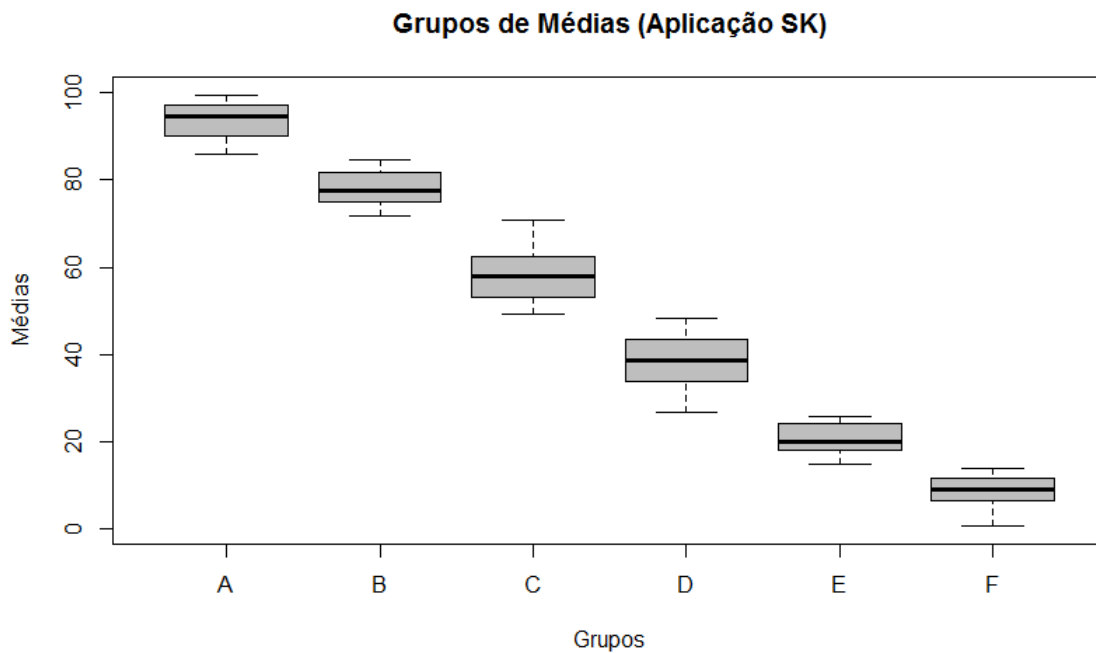
Logo, torna-se necessário uma análise do experimento para cada subgrupo dos dados em relação ao ciclo.

Tabela 14 – Comparação através de min e máx (Dados Embrapa)

| (a) Dados Originais | | | (b) Média dos Dados | | |
|---------------------|-------|--------|---------------------|-------|-------|
| | Min | Máx | | Min | Máx |
| Grupo_A | 70.87 | 100.00 | GrupoM_A | 86.00 | 99.56 |
| Grupo_B | 52.71 | 100.00 | GrupoM_B | 71.93 | 84.67 |
| Grupo_C | 24.35 | 87.93 | GrupoM_C | 49.16 | 70.87 |
| Grupo_D | 0.00 | 73.68 | GrupoM_D | 26.85 | 48.27 |
| Grupo_E | 7.89 | 35.48 | GrupoM_E | 14.92 | 25.84 |
| Grupo_F | 0.00 | 20.32 | GrupoM_F | 0.68 | 13.99 |

Na tabela 14 (a), temos o mínimo e o máximo de cada grupo referente às repetições. Nota-se que os grupos apresentam intersecção nas suas respectivas amplitudes, o que é algo normal, visto que estamos tratando de dados observacionais.

Figura 12 – Gráfico: Box-Plots - Grupos resultantes do Teste de Scott-Knott.



Podemos observar no gráfico acima que a evidência de que o pacote é eficiente levantada pela tabela 14 (a) se confirma novamente, pois os box-plots apresentam uma forma de "escada", em que cada grupo de média não invade a amplitude do outro. Assim, temos as médias classificadas adequadamente em seus respectivos grupos. Para finalizar, foi aplicado o Teste de Tukey, que consiste em definir a menor diferença significativa. Tal procedimento utiliza a amplitude da distribuição studentizada.

Tabela 15 – Sumário: P-valores de Tukey nos grupos de médias Scott-Knott (Local)

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|----------------|------|---------|--------|------|---------|------|
| Tukey Local GA | 0.53 | 0.99 | 1.00 | 0.97 | 1.00 | 1.00 |
| Tukey Local GB | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Tukey Local GC | 0.42 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Tukey Local GD | 0.96 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Tukey Local GE | 0.88 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Tukey Local GF | 0.01 | 0.91 | 1.00 | 0.85 | 1.00 | 1.00 |

Na tabela 15, temos o sumário dos P-valores de Tukey, aplicado em cada grupo de médias, isto é, foi aplicado o Tukey de maneira local. Nota-se para o grupo F, considerando um $\alpha = 0.05$, algumas diferenças de pares de médias influenciam para rejeitamos a hipótese de que as médias dos tratamentos pertencem ao mesmo grupo. Já temos fortes indícios de que a adaptação está sendo eficiente no método, mas para finalmente ter certeza o próximo passo é uma aplicação do método de Tukey aos grupos considerando todos os outros. Na

Tabela 16 – Sumário: P-valores de Tukey nos grupos de médias Scott-Knott (Global)

| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-----------------|------|---------|--------|------|---------|------|
| Tukey Global GA | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Tukey Global GB | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Tukey Global GC | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Tukey Global GD | 0.95 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Tukey Global GE | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Tukey Global GF | 0.95 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

tabela 16, temos o sumário dos P-valores de Tukey aplicado em cada grupo de médias, mas considerando o âmbito global, ou seja, considerando toda a informação. Percebe-se que nenhum grupo é rejeitado a hipótese de que as médias de fato pertencem aos mesmos.

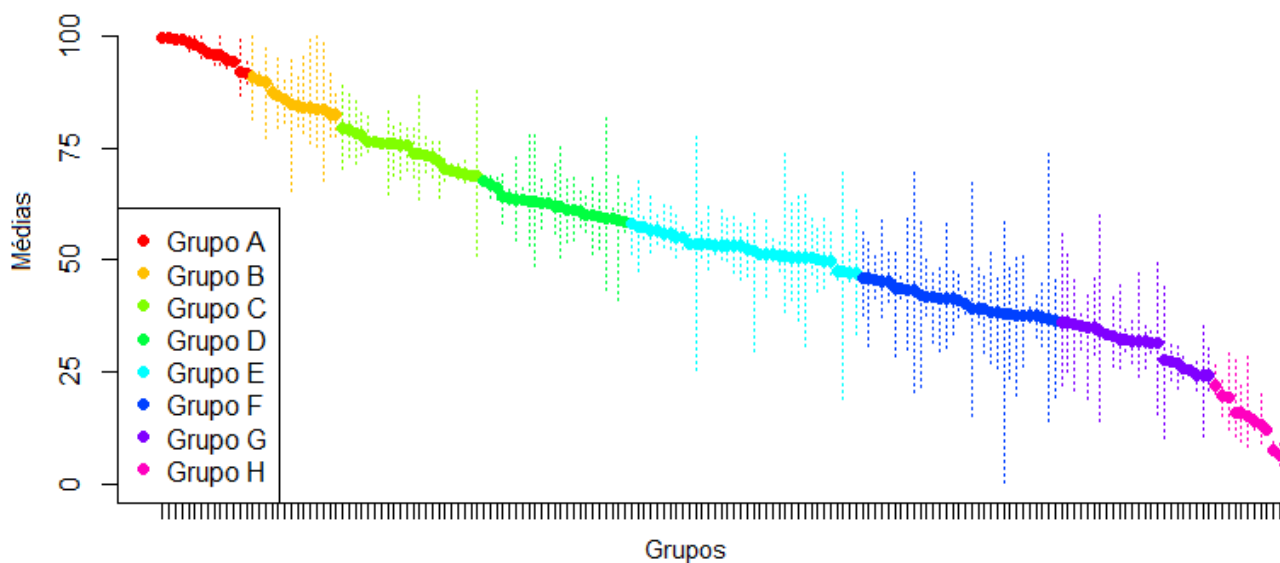
7.3 Análise Scott-Knott por Ciclos

7.3.1 Ciclo Curto

Tabela 17 – Sumário - Scott-Knott aplicado aos dados (Ciclo Curto)

| Médias | SK(5%) |
|----------------|--------|
| Min. : 6.452 | a :14 |
| 1st Qu.:38.295 | b :14 |
| Median :53.193 | c :22 |
| Mean :55.244 | d :23 |
| 3rd Qu.:71.115 | e :36 |
| Max. :99.561 | f :31 |
| | g :24 |
| | h :11 |

Figura 13 – Gráfico: Médias agrupadas por cores (Ciclo Curto)



A figura 13 apresenta o gráfico sobre as médias dos tratamentos agrupadas por cores em relação ao ciclo curto. Os grupos G e H apresentam a menor perda percentual em relação à brusone, considerando o percentual de aproximadamente 36% como corte de sucesso no experimento. Na tabela 18, são representados os genótipos que compõem ambos os grupos, a fim de observarmos a estatística do nosso interesse.

Tabela 18 – Genótipos de menor percentual de perda em relação à brusone (Ciclo Curto)

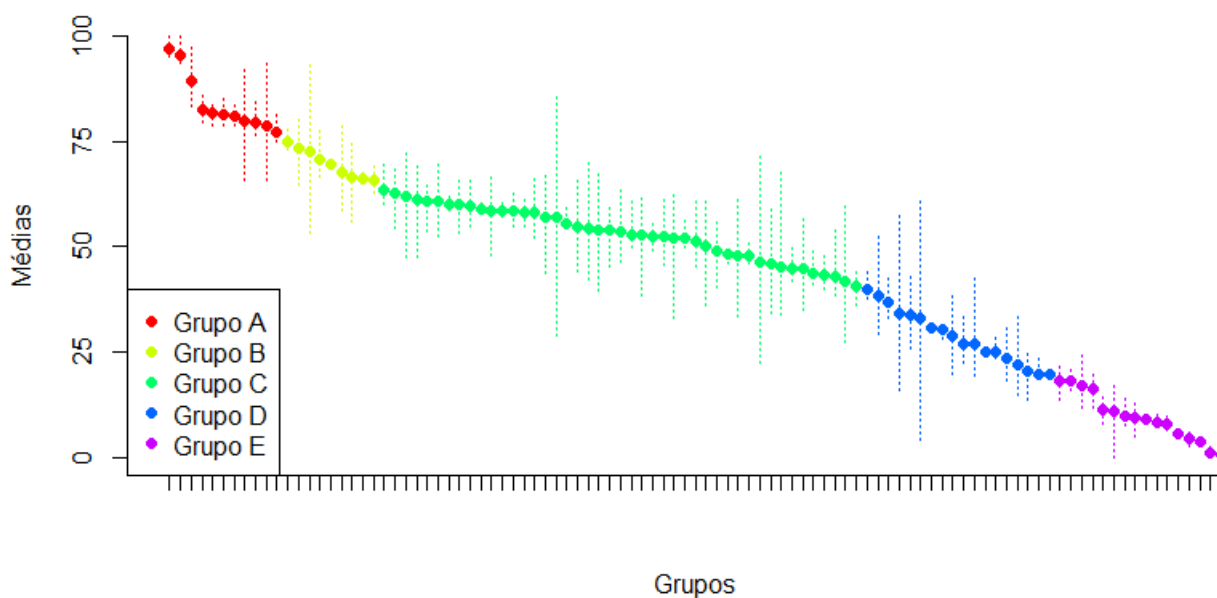
| | Means | SK(5%) | Quantidade em anos |
|----------------------------|-------|--------|--------------------|
| CIGM93275curto | 36.23 | g | 2012,2013 |
| BRS_LOUROcurto | 36.01 | g | 2012,2013 |
| EMBRAPA_27curto | 35.86 | g | 2012,2013 |
| PF_9127curto | 35.36 | g | 2013 |
| MENCEKIcurto | 34.95 | g | 2013 |
| CPAC_07340curto | 34.88 | g | 2012 |
| CPAC_0787curto | 34.17 | g | 2012 |
| Embrapa 27curto | 33.38 | g | 2012 |
| PF_030027/1curto | 33.15 | g | 2013 |
| CNT 10curto | 32.42 | g | 2012 |
| IPF_82880curto | 32.38 | g | 2012 |
| Onixcurto | 32.00 | g | 2012 |
| CIGM87.2771-1B-0PR-0Bcurto | 31.97 | g | 2012 |
| TRIGO_BR_24curto | 31.78 | g | 2012 |
| CPAC_0544curto | 31.63 | g | 2013 |
| TRIGO_BR_18curto | 31.45 | g | 2012 |
| BRS 208curto | 27.74 | g | 2011 |
| CPAC_07407curto | 27.33 | g | 2012 |
| PF_940110curto | 27.03 | g | 2011 |
| T 50130curto | 25.84 | g | 2012 |
| CIGM921698curto | 25.51 | g | 2012 |
| OPATA_85curto | 24.44 | g | 2011 |
| IPF 71349curto | 24.39 | g | 2012,2013 |
| Embrapa 40curto | 24.17 | g | 2012 |
| BRS_264curto | 22.17 | h | 2011 |
| PF 990606curto | 19.94 | h | 2012 |
| CPAC_07434curto | 19.37 | h | 2011 |
| GIZAcurto | 16.02 | h | 2011 |
| T_50130curto | 15.98 | h | 2011 |
| CASW96Y00538Scurto | 15.05 | h | 2012,2013 |
| PF 020458curto | 13.99 | h | 2012,2013 |
| PF_020062curto | 13.24 | h | 2012,2013 |
| CBFUSARIUM_ENT_025curto | 12.11 | h | 2012,2013 |
| CBFusarium_ENT_016curto | 7.66 | h | 2011 |
| BRS 179curto | 6.45 | h | 2012 |

7.3.2 Ciclo Médio

Tabela 19 – Sumário - Scott-Knott aplicado aos dados (Ciclo Médio)

| Médias | SK(5%) |
|-----------------|--------|
| Min. : 0.6808 | a:11 |
| 1st Qu.:27.8680 | b: 9 |
| Median :51.3388 | c:45 |
| Mean :46.5622 | d:18 |
| 3rd Qu.:60.8315 | e:16 |
| Max. :96.7728 | |

Figura 14 – Gráfico: Médias agrupadas por cores (Ciclo Médio)



A figura 14 apresenta o gráfico em que são mostradas as médias dos tratamentos agrupadas por cores em relação ao ciclo médio. Os grupos que apresentam a menor perda percentual em relação à brusone são os grupos D e E, considerando o percentual de aproximadamente 40 % como corte de sucesso no experimento. Na tabela 20, são representados os genótipos que compõem ambos os grupos, a fim de observarmos a estatística do nosso estudo.

Tabela 20 – Genótipos de menor percentual de perda em relação à brusone (Ciclo Médio)

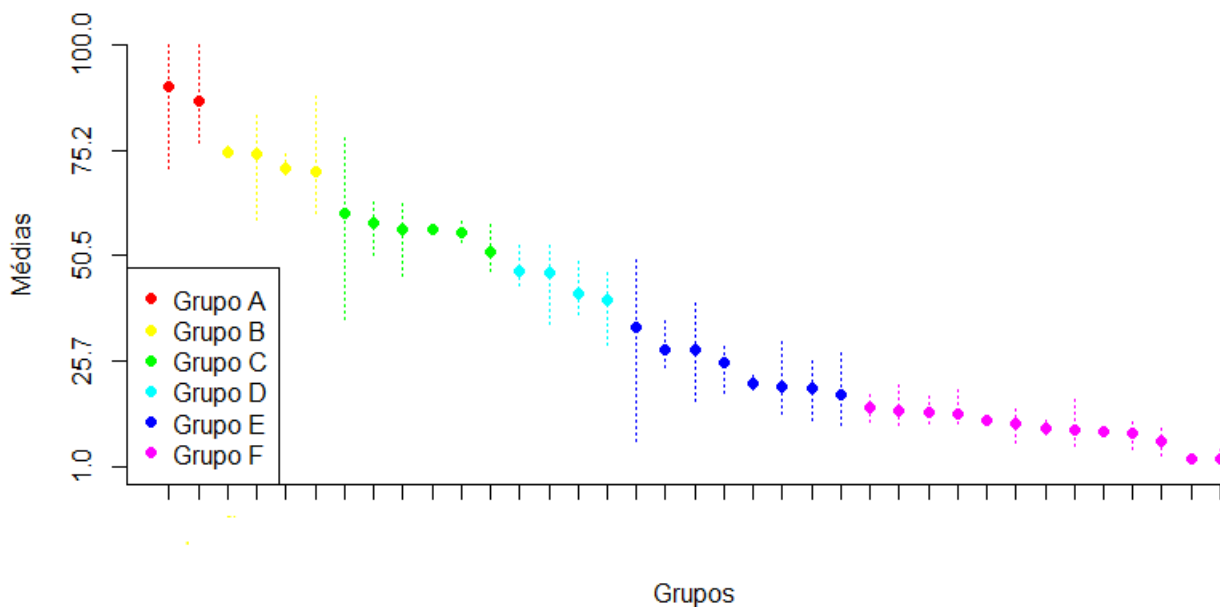
| | Means | SK(5%) | Quantidade em anos |
|-------------------------|-------|--------|--------------------|
| PATRIARCAmedio | 39.87 | d | 2011 |
| KLEIN_LUCEROmedio | 38.53 | d | 2012 |
| CIGM921666medio | 36.99 | d | 2011 |
| PF 990283medio | 34.29 | d | 2011 |
| CEP_24medio | 33.88 | d | 2011 |
| SUMAI_3medio | 33.01 | d | 2011 |
| CIGM89567medio | 30.91 | d | 2011 |
| CD_118medio | 30.22 | d | 2011 |
| CASW94Y00063Smedio | 28.82 | d | 2011 |
| BRS_ANGICOmedio | 26.91 | d | 2011 |
| Opata 85medio | 26.85 | d | 2012 |
| WeeBill1medio | 25.26 | d | 2012 |
| IPR_130medio | 25.08 | d | 2012 |
| BRS_229medio | 23.65 | d | 2012 |
| PF_050757medio | 21.87 | d | 2013 |
| PF 9052medio | 20.60 | d | 2013 |
| IPR_144medio | 19.83 | d | 2013 |
| LAGOA_VERMELHAMedio | 19.70 | d | 2013 |
| BRS_220medio | 18.22 | e | 2013 |
| OC_8154medio | 18.15 | e | 2012 |
| PF 89156medio | 17.27 | e | 2012 |
| Maya 74medio | 16.36 | e | 2013 |
| BRS 220medio | 11.54 | e | 2012 |
| CIGM921696medio | 10.85 | e | 2013 |
| BRS Angicomedio | 9.73 | e | 2011 |
| CBFUSARIUM_ENT_006medio | 9.34 | e | 2012 |
| Frontanamedio | 8.98 | e | 2012,2013 |
| PF_926medio | 8.28 | e | 2011 |
| CASW00GH00065Smedio | 7.96 | e | 2012 |
| CBFUSARIUM_ENT_014medio | 5.69 | e | 2011 |
| Melchiormedio | 4.57 | e | 2012 |
| Safiramedio | 3.84 | e | 2013 |
| JesuÃtamedio | 1.09 | e | 2013 |
| CASW94Y00116Smedio | 0.68 | e | 2011 |

7.3.3 Ciclo Longo

Tabela 21 – Sumário - Scott-Knott aplicado aos dados (Ciclo Longo)

| Médias | SK(5%) |
|----------------|--------|
| Min. : 2.761 | a: 2 |
| 1st Qu.:13.469 | b: 4 |
| Median :28.443 | c: 6 |
| Mean :35.266 | d: 4 |
| 3rd Qu.:56.520 | e: 8 |
| Max. :90.291 | f:13 |

Figura 15 – Gráfico: Médias agrupadas por cores (Ciclo Longo)



A figura 15 apresenta o gráfico em que são mostradas as médias dos tratamentos agrupadas por cores em relação ao ciclo longo. Os grupos que apresentam a menor perda percentual em relação à brusone são os grupos D, E e F, considerando o percentual de aproximadamente 47 % como corte de sucesso no experimento. Na tabela 22, são representados os genótipos que compõem ambos os grupos, a fim de observarmos a estatística do nosso interesse.

Tabela 22 – Genótipos de menor percentual de perda em relação à brusone (Ciclo Médio)

| | Means | SK(5%) | Quantidade em anos |
|-------------------------|-------|--------|--------------------|
| BRS_UMBULongo | 46.72 | d | 2011 |
| TRINTECINCOLongo | 46.52 | d | 2011 |
| 1855_83longo | 41.46 | d | 2012 |
| Karimlongo | 40.04 | d | 2011 |
| GALEGO_RAPADOLongo | 33.79 | e | 2011 |
| FRONTANALongo | 28.62 | e | 2012 |
| Shangailongo | 28.44 | e | 2011,2012 |
| Greece 78310-A-3longo | 25.36 | e | 2011 |
| CBFUSARIUM_ENT_007longo | 20.68 | e | 2011 |
| Embrapa 16longo | 19.87 | e | 2011 |
| CASW02GH00005Slongo | 19.38 | e | 2011 |
| CASW00GH00019Slongo | 17.94 | e | 2011 |
| Frondosolongo | 14.92 | f | 2011 |
| CASS03GH00077Slongo | 13.99 | f | 2011 |
| Har 604longo | 13.91 | f | 2011 |
| Lovrin 13longo | 13.47 | f | 2012 |
| PF 010069longo | 11.85 | f | 2013 |
| CASW02GH00045Slongo | 11.22 | f | 2013 |
| BRS 177longo | 10.21 | f | 2011 |
| PF_070475longo | 9.49 | f | 2012 |
| CASW02GH00002Slongo | 9.18 | f | 2013 |
| Klein Lucerolongo | 8.76 | f | 2011 |
| Granarololongo | 7.02 | f | 2012 |
| Trintecincolongo | 3.02 | f | 2011 |
| SAFIRALongo | 2.76 | f | 2012 |

8 Conclusão

Foi realizada uma análise descritiva aos dados, organizando-os, descrevendo-os e nos dando dessa forma um caminho a ser seguido na análise dos tratamentos (genótipos). A partir do conhecimento da base de dados, notou-se que se tratava de uma análise de comparação de médias. Daí veio a aplicação do método ANOVA, que foi eficiente dentro de suas limitações e nos mostrou a necessidade de aplicar outro método de comparação de médias. Assim, o método de Scott-Knott foi definido como o método a ser empregado, por ser eficiente na comparação de diversas médias.

A aplicação manual do teste de Scott-Knott foi utilizada para amostras simples, tanto para o caso balanceado quanto para o desbalanceado. Isso foi fundamental para o aprendizado do método. Logo, foi possível identificar quais seriam as possíveis dificuldades na aplicação dos métodos com a base de dados que contém desbalanceamento e saber o que seria necessário para poder suprir essas barreiras.

O pacote ScottKnott foi essencial para a análise do experimento, já que é muito eficiente para comparação de várias médias, além de proporcionar uma análise gráfica. Contudo, fez-se necessária uma pequena adaptação para adequá-lo ao caso em que o número de repetições dos tratamentos é desigual, ou seja, no caso em que ocorre o desbalanceamento dos dados. Após a adaptação, foi possível separar as médias dos tratamentos em grupos homogêneos e realizar uma análise sobre os grupos de médias.

Por fim, o estudo nos proporcionou vários tratamentos de diferentes ciclos, que apresentam alta resistência à brusone em relação aos demais. A partir daí, a Embrapa terá um amplo resultado para distinguir o que proporciona melhor eficiência. Para uma produção em massa, mas de maneira dispersa quanto aos tratamentos e seus ciclos, pode-se apostar em vários genótipos e de diferentes ciclos dentre aqueles que apresentaram melhor desempenho contra a doença no experimento. Dessa maneira, defendemos que os agricultores poderão investir mais no plantio do trigo, aumentando a oferta no mercado, gerando empregos e assim melhorar a economia que envolve toda a produção do trigo.

Referências

CONRADO, T. V.; FERREIRA, D. F.; SCAPIM, C. A.; MALUF, W. R. **Adjusting the Scott-Knott cluster analyses for unbalanced designs**. Crop Breed. Appl. Biotechnol. vol.17 no.1 Viçosa Jan./Mar. 2017

FISHER, W. K. **On grouping for maximum homogeneity**. Journal of the American Statistical Association. Vol. 53, No. 284 (Dec., 1958), pp. 789-798.

JELIHOVSCHI, E.G.; FARIA, J.C.; Allaman, I.B. **ScottKnott: a package for performing the Scott-Knott clustering algorithm in R**. TEMA (São Carlos) vol.15 no.1 São Carlos Jan./Apr. 2014.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. John Wiley and Sons: New York, p.66-76, 1997.

R Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing: Vienna, Austria, 2016. Disponível em: <https://www.rproject.org/>. Acesso em: 3. out. 2017.

RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. **Experimentação em genética e melhoramento de plantas**. Lavras: UFLA, p.87-134, 2000.

SCOTT, A. J.; KNOTT, M. A. **A cluster analysis method for grouping means in the analysis of variance**. Biometrics. Raleigh, v.30, n.3, p.507-512, Sept. 1974.

Silva, Cristiane Mariana Rodrigues da. **Uso do teste de Scott-Knott e da análise de agrupamentos, na obtenção de grupos de locais para experimentos com cana-de-açúcar**. 2007. 50 f. Dissertação (Mestrado em Agronomia). Universidade de São Paulo Escola Superior de Agricultura “Luiz de Queiroz”, São Paulo.

ANEXO A – R-Script

```
##### Pacotes e Bibliotecas #####

install.packages("ScottKnott")
install.packages("xtable")
library(ScottKnott)
library(xtable)
library(readxl)

## Entrada dos Dados Desbalanceados ##
x <- read_excel("C:/Users/NELSON/Downloads/dadosDez.xlsx")
x <- data.frame(x)
head(x)
levels(factor(x$Genotipo))

exclude <- subset(x,x$rep1==100 & x$rep2==100 & x$rep3==100)

k <-exclude[c(1:35),1]
k2 <-exclude[c(36:70),1]
k3 <- exclude[c(71:105),1]
tabelaex <- data.frame(k,k2,k3)
xtable(tabelaex)

exclude <- c(exclude$Genotipo)

out <- function(x, y) x[!x %in% y]
gen <- out(x$Genotipo,exclude )
x <- subset(x, x$Genotipo %in% gen)

x2 <- NULL
x1 <- rep(x$Genotipo,each =3)
for (i in 1:(length(x1)/3)) {
  x2 <- c(x2, as.matrix(x[i, (4:6)]))
}
dados <- data.frame(x1,x2)
head(dados)
```

```
##### Função SK adaptada para o caso desbalanceado #####
sk.default <- function (x, y = NULL, model, which, id.trim = 3, error,
                        sig.level = 0.05, dispersion = c("mm", "s", "se"), ...)
{
  if (is.data.frame(y))
    y <- as.matrix(y[, 1])
  else stopifnot(is.atomic(y))
  if (is.matrix(x) || is.atomic(x))
    x <- as.data.frame(x)
  if (!is.null(y))
    dat <- as.data.frame(cbind(x, y))
  else dat <- x
  av <- eval(substitute(aov(fo, dat), list(fo = formula(model))))
  if (class(av)[1] == "aov")
    res <- sk.aov(x = av, which = which, id.trim = id.trim,
                  sig.level = sig.level, dispersion = dispersion)
  else res <- SK.aovlist(x = av, which = which, id.trim = id.trim,
                        error = error, sig.level = sig.level, dispersion = dispersion)
  invisible(res)
}

sk.aov <- function (x, which = NULL, id.trim = 3, sig.level = 0.05,
                    dispersion = c("mm", "s", "se"), ...)
{
  if (is.null(which))
    which <- names(x$model)[2]
  mt <- model.tables(x, "means")
  if (is.null(mt$n))
    stop("No factors in the fitted model!")
  r <- mt$n[names(mt$tables)][[which]]
  MSE <- deviance(x)/df.residual(x)
  m <- as.vector(mt$tables[[which]])
  nms <- names(mt$tables[[which]])
  ord <- order(m, decreasing = TRUE)
  m.inf <- m.inf.1a(x, which, dispersion)
  rownames(m.inf) <- nms
  m.inf <- m.inf[order(m.inf[, 1], decreasing = TRUE), ]
  mMSE <- MSE/r
  dfr <- x$df.residual
  g <- nrow(m.inf)
}
```

```

groups <- maxvalue(g, m.inf[, 1], mMSE, dfr, sig.level = sig.level,
                  1, rep(0, g), 0, rep(0, g))
res <- list(av = x, groups = groups, nms = nms, ord = ord,
           m.inf = m.inf, sig.level = sig.level)
class(res) <- c("SK", "list")
invisible(res)
}

maxvalue <- function (g, means, mMSE, dfr, sig.level, k, group, ngroup,
                     markg, g1 = g, sqsum = rep(0, g1))
{
  for (k1 in k:(g - 1)) {
    t1 <- sum(means[k:k1])
    k2 <- g - k1
    t2 <- sum(means[(k1 + 1):g])
    sqsum[k1] <- t1^2/(k1 - k + 1) + t2^2/k2 - (t1 + t2)^2/(g -
                                                    k + 1)
  }
  ord1 <- order(sqsum, decreasing = TRUE)[1]
  b0 <- max(sqsum)
  si02 <- (1/((g - k + 1) + dfr)) * (sum((means[k:g] - mean(means[k:g]))^2) +
                                     dfr * mean(mMSE[k:g]))
  lam <- (pi/(2 * (pi - 2))) * b0/si02
  valchisq <- qchisq((sig.level), lower.tail = FALSE, df = (g - k + 1)/(pi - 2))
  if ((lam < valchisq) | (ord1 == k)) {
    if (lam > valchisq) {
      ngroup <- ngroup + 1
      group[k] <- ngroup
      k <- ord1 + 1
    }
    if (lam < valchisq) {
      ngroup <- ngroup + 1
      group[k:g] <- ngroup
      if (prod(group) > 0)
        return(group)
      k <- g + 1
      g <- markg[g]
    }
  }
  while (k == g) {
    ngroup <- ngroup + 1

```

```

    group[g] <- ngroup
    if (prod(group) > 0)
      return(group)
    k <- g + 1
    g <- markg[g]
  }
}
else {
  markg[ord1] <- g
  g <- ord1
}
maxvalue(g, means, mMSE, dfr, sig.level, k, group, ngroup,
          markg)
}
desbalanceada <- function(x){
  sk.default(x,x=x2,y = x1,which = 'x1', model = 'x2~x1')
}
dadosdesb <- desbalanceada(dados)
sumdadosdesb <- summary(dadosdesb)

anos <- as.data.frame(matrix(NA,311,1))
geno <- c(levels(factor(x$Genotipo)))
for(ge in 1:length(levels(factor(x$Genotipo)))){
  anos[ge,] <- paste(sort(unique(x[x$Genotipo==geno[ge],]$ano)),collapse = ",")
}
colnames(anos)<- c("Qtd em anos")
fim <- cbind(sumdadosdesb,anos)
xtable(summary(sumdadosdesb))
xtable(fim[283:311,c(2,3,4)])
plot(dadosdesb,col=rainbow(max(dadosdesb$groups)),mm.lty=3,id.las=2,
      rl=FALSE,title='factor levels')
legend("bottomleft", pch = c(19, 19, 19, 19,19, 19),
      col = rainbow(max(dadoscurto1$groups)),
      legend = c("Grupo A","Grupo B","Grupo C","Grupo D","Grupo E","Grupo F"))

##### Tabela Min e Max dados originais por grupo #####

```

```

grupoA <- rownames(dadosdesb$m.inf)[dadosdesb$groups==1]
dados[is.element(dados$x1,grupoA),]
Grupo_A <- range(dados[is.element(dados$x1,grupoA),2])

grupoB <- rownames(dadosdesb$m.inf)[dadosdesb$groups==2]
dados[is.element(dados$x1,grupoB),]
Grupo_B <- range(dados[is.element(dados$x1,grupoB),2])

grupoC <- rownames(dadosdesb$m.inf)[dadosdesb$groups==3]
dados[is.element(dados$x1,grupoC),]
Grupo_C <- range(dados[is.element(dados$x1,grupoC),2])

grupoD <- rownames(dadosdesb$m.inf)[dadosdesb$groups==4]
dados[is.element(dados$x1,grupoD),]
Grupo_D <- range(dados[is.element(dados$x1,grupoD),2])

grupoE <- rownames(dadosdesb$m.inf)[dadosdesb$groups==5]
dados[is.element(dados$x1,grupoE),]
Grupo_E <- range(dados[is.element(dados$x1,grupoE),2])

grupoF <- rownames(dadosdesb$m.inf)[dadosdesb$groups==6]
dados[is.element(dados$x1,grupoF),]
Grupo_F <- range(dados[is.element(dados$x1,grupoF),2])

tabminmax <- t(data.frame(Grupo_A,Grupo_B,Grupo_C,Grupo_D,
                          Grupo_E,Grupo_F))
colnames(tabminmax) <- c("Min","Máx")
xtable(tabminmax)

##### Tabela Min e Max dados médias por grupo #####

GrupoM_A <- range(tapply(dados[is.element(dados$x1,grupoA),2],
                        dados[is.element(dados$x1,grupoA),1],mean),na.rm = T)
GrupoM_B <- range(tapply(dados[is.element(dados$x1,grupoB),2],
                        dados[is.element(dados$x1,grupoB),1],mean),na.rm = T)
GrupoM_C <- range(tapply(dados[is.element(dados$x1,grupoC),2],
                        dados[is.element(dados$x1,grupoC),1],mean),na.rm = T)
GrupoM_D <- range(tapply(dados[is.element(dados$x1,grupoD),2],
                        dados[is.element(dados$x1,grupoD),1],mean),na.rm = T)

```

```

GrupoM_E <- range(tapply(dados[is.element(dados$x1,grupoE),2],
                        dados[is.element(dados$x1,grupoE),1],mean),na.rm = T)
GrupoM_F <- range(tapply(dados[is.element(dados$x1,grupoF),2],
                        dados[is.element(dados$x1,grupoF),1],mean),na.rm = T)
GrupoM_G <- range(tapply(dados[is.element(dados$x1,grupoG),2],
                        dados[is.element(dados$x1,grupoG),1],mean),na.rm = T)

tabminmaxm <- t(data.frame(GrupoM_A,GrupoM_B,GrupoM_C,GrupoM_D,
                        GrupoM_E,GrupoM_F))
colnames(tabminmaxm) <- c("Min","Máx")
xtable(tabminmaxm)

##### Box-Plots dos Grupos de Médias #####

tabboxA <- data.frame(tapply(dados[is.element(dados$x1,grupoA),2],
                        dados[is.element(dados$x1,grupoA),1],mean))
boxA <- tabboxA[!(is.na(tabboxA$tapply.dados.is.element.dados.x1..
                        grupoA...2...dados.is.element.dados.x1..)),]

tabboxB <-data.frame(tapply(dados[is.element(dados$x1,grupoB),2],
                        dados[is.element(dados$x1,grupoB),1],mean))
boxB <- tabboxB[!(is.na(tabboxB$tapply.dados.is.element.dados.x1..
                        grupoB...2...dados.is.element.dados.x1..)),]

tabboxC <-data.frame(tapply(dados[is.element(dados$x1,grupoC),2],
                        dados[is.element(dados$x1,grupoC),1],mean))
boxC <- tabboxC[!(is.na(tabboxC$tapply.dados.is.element.dados.x1..
                        grupoC...2...dados.is.element.dados.x1..)),]

tabboxD <-data.frame(tapply(dados[is.element(dados$x1,grupoD),2],
                        dados[is.element(dados$x1,grupoD),1],mean))
boxD <- tabboxD[!(is.na(tabboxD$tapply.dados.is.element.dados.x1..
                        grupoD...2...dados.is.element.dados.x1..)),]

tabboxE <-data.frame(tapply(dados[is.element(dados$x1,grupoE),2],
                        dados[is.element(dados$x1,grupoE),1],mean))
boxE <- tabboxE[!(is.na(tabboxE$tapply.dados.is.element.dados.x1..
                        grupoE...2...dados.is.element.dados.x1..)),]

```

```

tabboxF <-data.frame(tapply(dados[is.element(dados$x1,grupoF),2],
                           dados[is.element(dados$x1,grupoF),1],mean))
boxF <- tabboxF[!(is.na(tabboxF$tapply.dados.is.element.dados.x1..
                       grupoF...2...dados.is.element.dados.x1..)),]

tabboxG <-data.frame(tapply(dados[is.element(dados$x1,grupoG),2],
                           dados[is.element(dados$x1,grupoG),1],mean))
boxG <- tabboxG[!(is.na(tabboxG$tapply.dados.is.element.dados.x1..
                       grupoG...2...dados.is.element.dados.x1..)),]

boxplot(boxA,boxB,boxC,boxD,boxE,boxF,col="grey",
        names=c("A","B","C","D","E","F"),
        main ="Grupos de Médias (Aplicação SK)",xlab = "Grupos",ylab = "Médias")

##### Análise de Variância nos Grupos (SK) #####

dadosA <- dados[is.element(dados$x1,grupoA),c(1,2)]
modeloA <- aov(dadosA$x2~dadosA$x1)
anovaA <- summary(modeloA)
xtable(anovaA)

dadosB <- dados[is.element(dados$x1,grupoB),c(1,2)]
modeloB <- aov(dadosB$x2~dadosB$x1)
anovaB <- summary(modeloB)
xtable(anovaB)

dadosC <- dados[is.element(dados$x1,grupoC),c(1,2)]
modeloC <- aov(dadosC$x2~dadosC$x1)
anovaC <- summary(modeloC)
xtable(anovaC)

dadosD <- dados[is.element(dados$x1,grupoD),c(1,2)]
modeloD <- aov(dadosD$x2~dadosD$x1)
anovaD <- summary(modeloD)
xtable(anovaD)

dadosE <- dados[is.element(dados$x1,grupoE),c(1,2)]
modeloE <- aov(dadosE$x2~dadosE$x1)
anovaE <- summary(modeloE)

```

```

xtable(anovaE)

dadosF <- dados[is.element(dados$x1, grupoF), c(1, 2)]
modeloF <- aov(dadosF$x2~dadosF$x1)
anovaF <- summary(modeloF)
xtable(anovaF)

dadostot <- dados[is.element(dados$x1, rownames(dadosdesb$m.inf)
                           [dadosdesb$groups==1 | dadosdesb$groups==2
                           | dadosdesb$groups==3 | dadosdesb$groups==4
                           | dadosdesb$groups==5 | dadosdesb$groups==6]), c(1, 2)]
modelo <- aov(dadostot$x2~dadostot$x1)
anovaG <- summary(modeloG)
xtable(anovaG)

##### Análise de Tukey nos Grupos (SK) Local #####

tukeyA <- TukeyHSD(modeloA, ordered = T)
t1 <- summary(tukeyA$`dadosA$x1`[, 4])
tukeyB <- TukeyHSD(modeloB, ordered = T)
t2 <- summary(tukeyB$`dadosB$x1`[, 4])
tukeyC <- TukeyHSD(modeloC, ordered = T)
t3 <- summary(tukeyC$`dadosC$x1`[, 4])
tukeyD <- TukeyHSD(modeloD, ordered = T)
t4 <- summary(tukeyD$`dadosD$x1`[, 4])
tukeyE <- TukeyHSD(modeloE, ordered = T)
t5 <- summary(tukeyE$`dadosE$x1`[, 4])
tukeyF <- TukeyHSD(modeloF, ordered = T)
t6 <- summary(tukeyF$`dadosF$x1`[, 4])

tukeyloc <- rbind(t1, t2, t3, t4, t5, t6)
xtable(tukeyloc)

##### Análise de Tukey nos Grupos (SK) Global #####

tukeyGeral <- TukeyHSD(modelo, ordered = T)
pares <- tukeyGeral$`dadostot$x1`[, 4]
nomespares <- rownames(data.frame(pares))

```

```
paresA <- tukeyA$'dadosA$x1'[,4]
nomespaesA <- rownames(data.frame(paresA))
parestukA <- subset(pares,nomespaes %in% nomespaesA)
sgA <-summary(parestukA)

paresB <- tukeyB$'dadosB$x1'[,4]
nomespaesB <- rownames(data.frame(paresB))
parestukB <- subset(pares,nomespaes %in% nomespaesB)
sgB <- summary(parestukB)

paresC <- tukeyC$'dadosC$x1'[,4]
nomespaesC <- rownames(data.frame(paresC))
parestukC <- subset(pares,nomespaes %in% nomespaesC)
sgC <- summary(parestukC)

paresD <- tukeyD$'dadosD$x1'[,4]
nomespaesD <- rownames(data.frame(paresD))
parestukD <- subset(pares,nomespaes %in% nomespaesD)
sgD <- summary(parestukD)

paresE <- tukeyE$'dadosE$x1'[,4]
nomespaesE <- rownames(data.frame(paresE))
parestukE <- subset(pares,nomespaes %in% nomespaesE)
sgE <- summary(parestukE)

paresF <- tukeyD$'dadosD$x1'[,4]
nomespaesF <- rownames(data.frame(paresF))
parestukF <- subset(pares,nomespaes %in% nomespaesF)
sgF <- summary(parestukF)

paresG <- tukeyG$'dadosG$x1'[,4]
nomespaesG <- rownames(data.frame(paresG))
parestukG <- subset(pares,nomespaes %in% nomespaesG)
sgG <- summary(parestukG)

xtable(rbind(sgA,sgB,sgC,sgD,sgE,sgF))
```

```
##### Análise por Ciclos #####
```

```
##### Ciclo Curto #####
curto <- subset(x,x$ciclo=='curto')
x2 <- NULL
x1 <- rep(curto$Genotipo,each =3)
for (i in 1:(length(x1)/3)) {
  x2 <- c(x2, as.matrix(curto[i, (4:6)]))
}
dadoscurto <- data.frame(x1,x2)
dadoscurto1 <- desbalanceada(dadoscurto)
sumdadoscurto1 <- summary(dadoscurto1)
anos <- as.data.frame(matrix(NA,175,1))
geno <- c(levels(factor(curto$Genotipo)))
for(ge in 1:length(levels(factor(curto$Genotipo)))){
  anos[ge,] <- paste(sort(unique(curto[curto$Genotipo==geno[ge],]$ano))
                    ,collapse = ",")
}
colnames(anos)<- c("Qtd em anos")
fim <- cbind(sumdadoscurto1,anos)
xtable(summary(sumdadoscurto1))
xtable(fim[141:175,c(2,3,4)])
plot(dadoscurto1,col=rainbow(max(dadoscurto1$groups)),mm.lty=3,id.las=0,
      rl=FALSE,ylab = 'Médias',xlab = 'Grupos')
legend("bottomleft", pch = c(19, 19, 19, 19,19, 19, 19, 19),
      col = rainbow(max(dadoscurto1$groups)),
      legend = c("Grupo A","Grupo B","Grupo C","Grupo D","Grupo E",
                  "Grupo F","Grupo G","Grupo H"))
##### Ciclo Médio #####
medio <- subset(x,x$ciclo=='medio')
x2 <- NULL
x1 <- rep(medio$Genotipo,each =3)
for (ii in 1:(length(x1)/3)) {
  x2 <- c(x2, as.matrix(medio[ii, (4:6)]))
}
dadosmedio <- data.frame(x1,x2)
dadosmedio1 <- desbalanceada(dadosmedio)
sumdadosmedio1 <- summary(dadosmedio1)
xtable(summary(sumdadosmedio1))
anos <- as.data.frame(matrix(NA,99,1))
geno <- c(levels(factor(medio$Genotipo)))
```

```

for(ge in 1:length(levels(factor(medio$Genotipo)))){
  anos[ge,] <- paste(sort(unique(medio[medio$Genotipo==geno[ge],]$ano))
                    ,collapse = ",")
}
colnames(anos)<- c("Qtd em anos")
fim <- cbind(sumdadosmedio1,anos)
xtable(fim[66:99,c(2,3,4)])
plot(dadosmedio1,col=rainbow(max(dadosmedio1$groups)),mm.lty=3,id.las=0,
      rl=FALSE,xlab = 'Grupos',ylab = 'Médias')
legend("bottomleft", pch = c(19, 19, 19, 19,19),
      col = rainbow(max(dadosmedio1$groups)),
      legend = c("Grupo A","Grupo B","Grupo C","Grupo D","Grupo E"))
##### Ciclo Longo #####
longo <- subset(x,x$ciclo=='longo')
x2 <- NULL
x1 <- rep(longo$Genotipo,each =3)
for (i in 1:(length(x1)/3)) {
  x2 <- c(x2, as.matrix(longo[i, (4:6)]))
}
dadoslongo <- data.frame(x1,x2)
dadoslongo1 <- desbalanceada(dadoslongo)
sumdadoslongo1 <- summary(dadoslongo1)
xtable(summary(sumdadoslongo1))
anos <- as.data.frame(matrix(NA,37,1))
geno <- c(levels(factor(longo$Genotipo)))
for(ge in 1:length(levels(factor(longo$Genotipo)))){
  anos[ge,] <- paste(sort(unique(longo[longo$Genotipo==geno[ge],]$ano))
                    ,collapse = ",")
}
colnames(anos)<- c("Quantidade em anos")
fim <- cbind(sumdadoslongo1,anos)
xtable(fim[13:37,c(2,3,4)])
plot(dadoslongo1,col=rainbow(max(dadoslongo1$groups)),mm.lty=3,id.las=0,
      rl=FALSE,xlab = 'Grupos',ylab = 'Médias')
legend("bottomleft", pch = c(19, 19, 19, 19,19, 19),
      col = rainbow(max(dadoslongo1$groups)),
      legend = c("Grupo A","Grupo B","Grupo C","Grupo D","Grupo E","Grupo F"))

```

