



Resultado parciais TCC

Autor

Iago Marçal Costa dos Santos

Requerente

Brasília, XXXX de 202X

Sumário

1	Introdução	2
2	Correções ao relatório parcial de TCC 1	2
2.1	Esfericidade	2
2.2	Correção nos graus de liberdade	2
3	Modelagem	3
4	Resultados	3
4.1	Dados brutos	3
4.2	Ranques	7

1 Introdução

Neste documento serão apresentados os resultados obtidos para as Análises de Variância por medidas repetidas para os dados de Moscas Brancas em armadilhas e Plantas.

2 Correções ao relatório parcial de TCC 1

2.1 Esfericidade

A esfericidade parte do princípio de que uma matriz é esférica se ela for do tipo H, ou seja,

$$\Sigma_{n \times n} = A_{n \times n} + A'_{n \times n} + \lambda I_{n \times n}.$$

Como demonstrado em uma das minhas referências (Huynh 1970), ao multiplicar Σ por uma matriz de contrastes normalizados C , na forma $C\Sigma C'$, $AC' = CA' = 0$.

Assim, em caso de esfericidade $C\Sigma C' = \lambda I$.

A hipótese do teste de Mauchly para esfericidade é $H_0)C\Sigma C' = \lambda I$

No relatório de TCC1 eu defini a estatística do teste de Mauchly para esfericidade como

$$W = \frac{(m-1)^{m-1} \times |CSC'|}{tr(CSC')^{s-1}},$$

Quando na verdade esse é o critério de Mauchly (m é o número de medidas repetidas).

$C_{8 \times 9}$ é uma matriz de contrastes ortogonais e o livro que uso como base informa que os contrastes adequados para dados temporais são contrastes polinomiais (a matriz C que anexe no github). Obtive esses contrastes pelo SAS (O SAS dá o nome de "matriz M").

$S_{9 \times 9}$ é a matriz de variâncias e covariâncias estimadas, ou seja, $e'e$, a matriz de resíduos transposta multiplicada pela matriz de resíduos.

Portanto, $CSC'_{8 \times 8}$ é idêntica a "matriz E" fornecida pelo SAS. Para conseguir essa matriz basta fazer $M(e'e)M'$ com a matriz de resíduos e a matriz de contrastes.

A estatística do teste de esfericidade é na verdade

$$\chi^2 = -\gamma \times \ln(W),$$

onde γ é igual a

$$\gamma = DFE - \frac{2m^2 - 3m + 3}{6(m-1)}, \quad (1)$$

onde DFE é o número de graus de liberdade do erro da análise de entre sujeitos. Para armadilhas, $DFE = 8$ e para plantas $DFE = 84$.

E tem distribuição χ^2 com $\frac{m(m-1)}{2} - 1$ graus de liberdade sob a hipótese nula.

2.2 Correção nos graus de liberdade

Citei que a correção de Greenhouse-Geisser utiliza os autovalores da matriz CSC' , quando na verdade são os valores da diagonal principal.

$$\hat{\epsilon}_{gg} = \frac{\left(\sum_{i=1}^{m-1} a_{ii} \right)^2}{(m-1) \sum_{i=1}^{m-1} \sum_{j=1}^{m-1} a_{ij}^2}$$

Por fim, descobri que a Correção de Huynh e Feldt é aplicada apenas quando não existem fatores entre sujeitos, tais como o tratamento e os blocos, apenas as medidas repetidas.

A solução é uma modificação que Lacoutre fez, então a correção adequada para os dados é a correção de Huynh-Feldt e Lacoutre

$$\hat{\epsilon}_{hfl} = \frac{(DFE + 1) \times (m-1) \hat{\epsilon}_{gg} - 2}{(m-1) \times (DFE - (m-1) \hat{\epsilon}_{gg})}.$$

3 Modelagem

A princípio, eu estava trabalhando com um modelo Split-plot, tal como sugeriu o livro do KUEHL, na forma:

$$Y_{ijk} = \mu + \tau_i + \beta_j + d_{ij} + m_k + (m\tau)_{ik} + (m\beta)_{jk} + \varepsilon_{ijk},$$

para as armadilhas.

- τ_i é o efeito dos tratamentos
- β_j o efeito dos blocos
- d_{ij} o erro 1
- m_k o efeito das medidas repetidas
- $(m\tau)_{ik}$ a interação entre as medidas repetidas e os tratamentos
- $(m\beta)_{jk}$ a interação entre as medidas repetidas e os blocos
- ε_{ijk} o erro 2.

Não botei interação entre tratamentos e blocos pois o modelo para armadilhas fica saturado e não tem soma de quadrados.

Esse tipo de modelo não funcionou. Ao fazer a estimativa de parâmetros tal qual eu fazia na aula de delineamento, utilizando *apply*, muitas vezes eu obtenho valores preditos negativos, o que não é condizente com a natureza do estudo.

Então eu passei a trabalhar com a ideia de um modelo politômico, em que cada medida repetida é uma variável resposta.

$$Y_{ijk} = m_k + (m\tau)_{ik} + (m\beta)_{jk} + \varepsilon_{ijk}.$$

Nesse caso, não existe uma média geral e sim 9 médias semanais para as armadilhas e 7 para as plantas.

No caso dos ranques, como eu ranqueio todas as observações dentro das medidas repetidas, não há diferenças semanais. Assim, o modelo é:

$$Y_{ijk} = (m\tau)_{ik} + (m\beta)_{jk} + \varepsilon_{ijk}.$$

A minha dúvida é se posso utilizar essa ideia de modelo politômico porque a minha tabela da ANOVA e soma de quadrados continua igual ao do modelo Split-plot, como vou mostrar mais pra frente.

4 Resultados

4.1 Dados brutos

- Matriz de resíduos e de contrastes fornecidas pelo SAS em anexo (preditos-resíduos-armadilhas)
- Critério de Mauchly $W = 8.828 \times 10^{-11}$ (pode ser confirmado pelo código R)
- $\gamma = 6.125$ e estatística do teste de Mauchly $\chi^2 = 118,646$ e p-valor < 0.001 (pode ser confirmado pelo R), portanto não há esfericidade na matriz de covariâncias estimada. É necessário usar correções ao teste F na análise dentre sujeitos. O valor das correções também pode ser confirmado pelo R.

Sphericity Tests				
Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
Orthogonal Components	35	8.828E-11	118.64607	<.0001

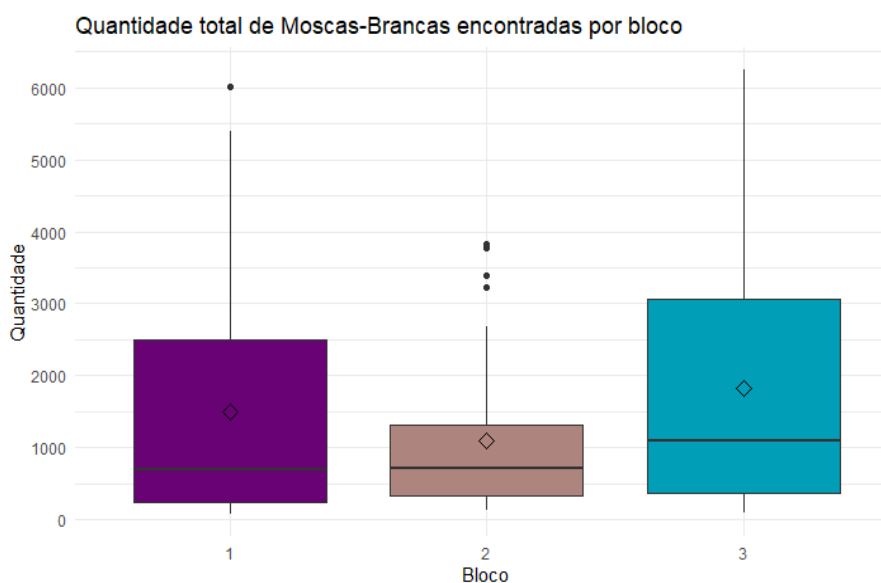
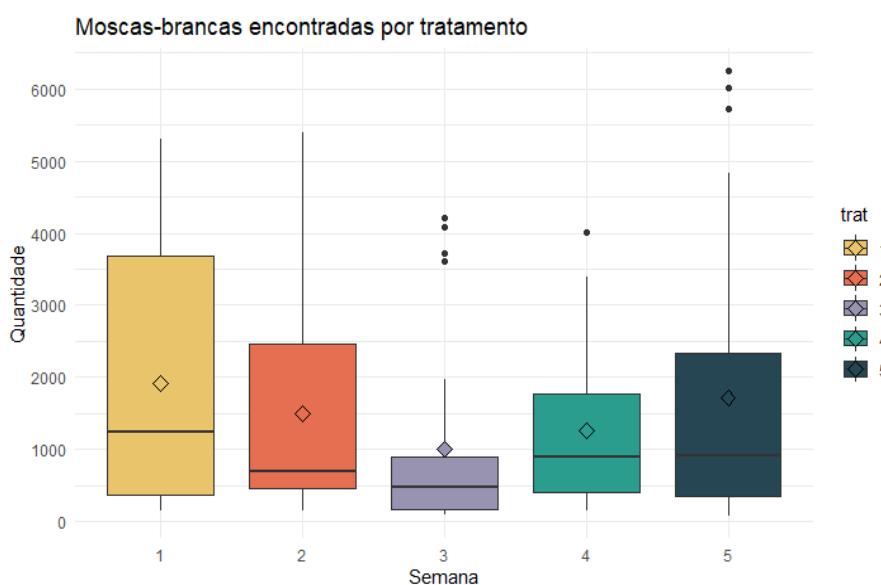
Correções	
Greenhouse-Geisser Epsilon	0.2778
Huynh-Feldt-Lecoutre Epsilon	0.3894

Com essas informações, partimos para a Análise de variância. A análise entre sujeitos é a que estamos mais interessados, visto que o objetivo do trabalho é apontar se existe alguma diferença entre as variedades de tomateiro.

A análise entre sujeitos consiste em ignorar o efeito das medidas repetidas e verificar diferenças nas condições do experimento, tais como tratamentos e blocos.

Análise entre sujeitos					
Source	DF	Sum Square	Mean Square	F Value	Pr >F
trat	4	14144350.93	3536087.73	1.00	0.4609
bloco	2	11941969.30	5970984.65	1.69	0.2445
Error	8	28289707.51	3536213.44		

Com esses valores, não é possível rejeitar a hipótese de que os tratamentos e blocos influenciem na quantidade de moscas brancas apanhadas pelas armadilhas, o que pode ser confirmado pelos gráficos abaixo:

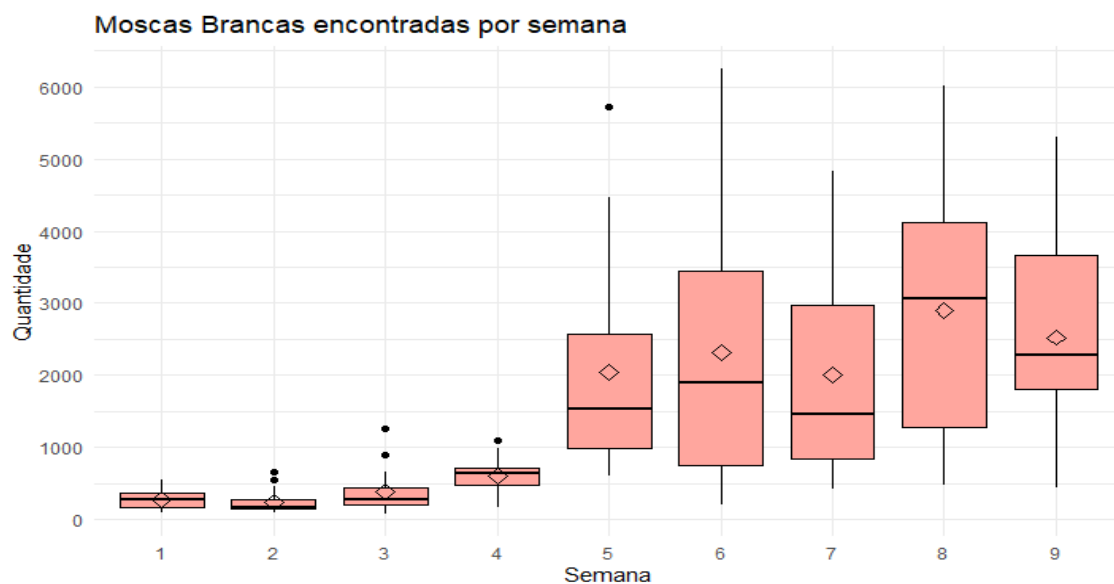
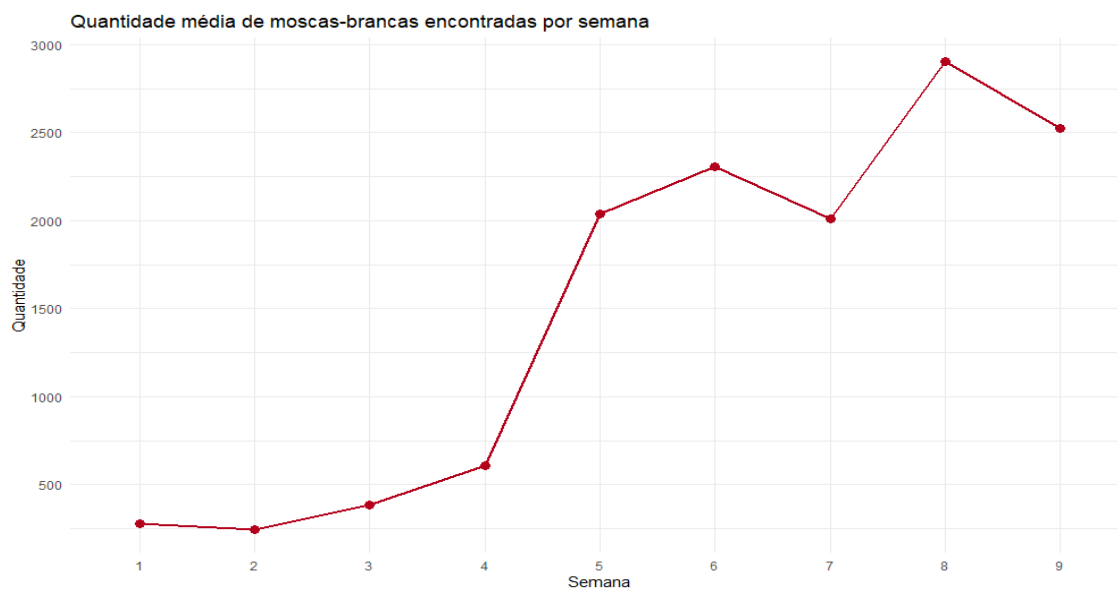


A análise dentro os sujeitos investiga a relação entre medidas repetidas, ou seja como cada sujeito se comporta no decorrer do estudo.

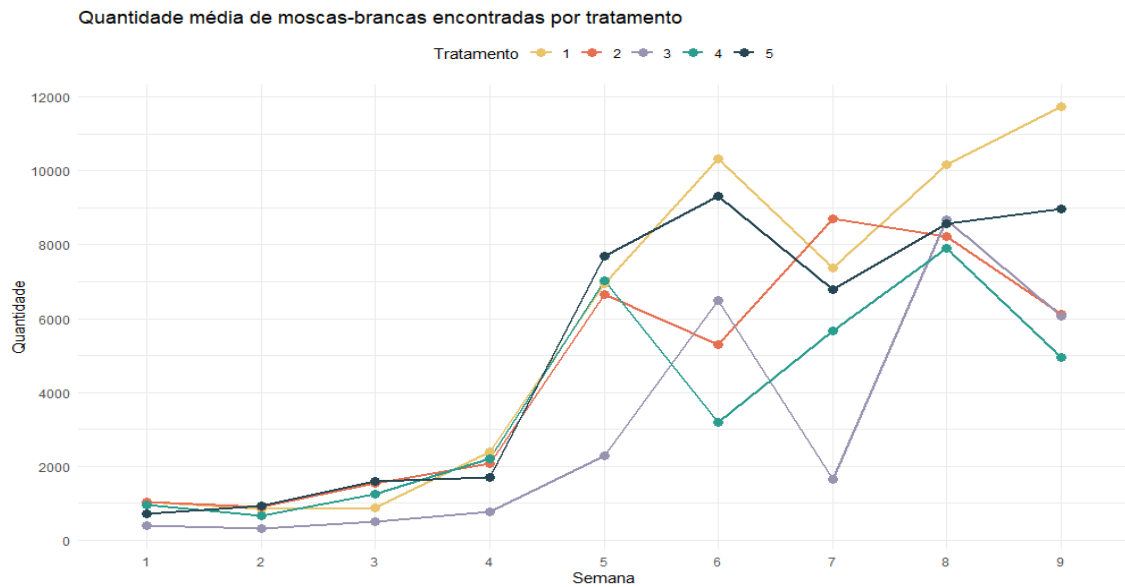
Análise dentre sujeitos

Source	DF	Sum Square	Mean Square	F Value	Pr >F	Adj Pr >F	
						G - G	H-F-L
semana	8	140090485.9	17511310.7	20.36	<.0001	<.0001	<.0001
semana*trat	32	25450001.6	795312.6	0.92	0.5867	0.5264	0.5405
semana*bloco	16	42111529.4	2631970.6	3.06	0.0008	0.0402	0.0210
Error(semana)	64	55044915.2	860076.8				

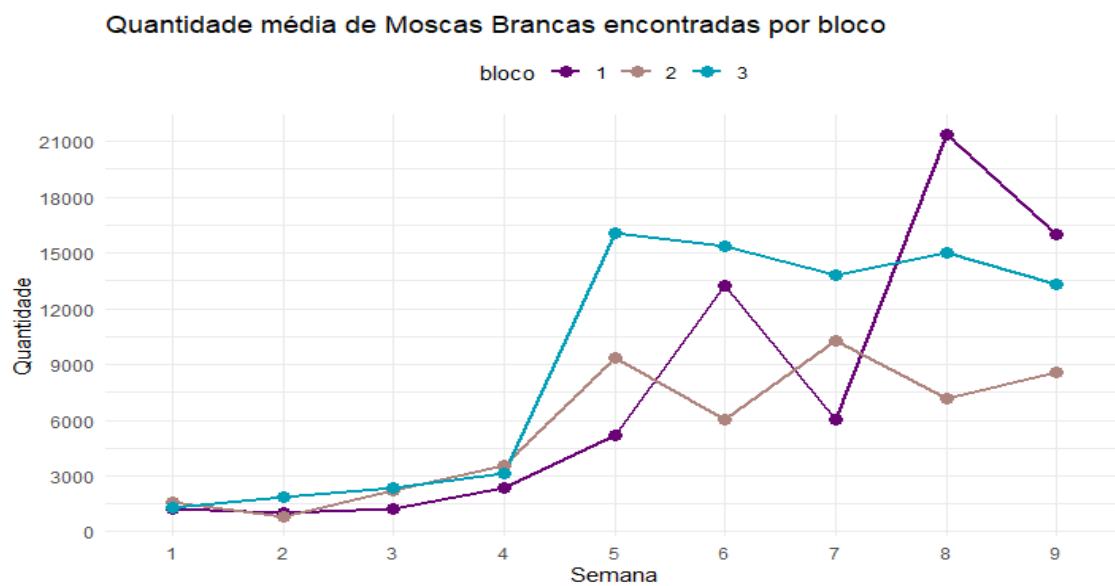
Pela tabela, existem evidências para rejeitar a hipótese de igualdade entre as semanas. Pelo gráfico boxplot abaixo, podemos verificar que com o decorrer do estudo a quantidade de moscas brancas apreendidas cresce.



A interação entre o tempo e o tratamentos não é significativa, ou seja, no decorrer do estudo, as quantidades de moscas brancas apreendidas cresce igualmente entre as armadilhas de cada uma das variedades de tomateiro.



Já a interação entre o tempo e os blocos é significativa, portanto a quantidade de moscas apreendidas cresce de forma diferente entre as armadilhas de cada bloco.



Falta fazer a Análise de comparações multiplas.

Os parâmetros do modelo são:

Par.	Semana 1	Semana 2	Semana 3	Semana 4	Semana 5	Semana 6	Semana 7	Semana 8	Semana 9
Par.	221.200	434.467	617.933	591.933	3739.800	3873.467	3016.200	2959.067	3136.733
trat 1	104.000	-24.000	-237.667	234.333	-243.000	337.000	193.333	533.000	921.000
trat 2	101.667	-4.667	-19.667	123.333	-348.667	-1343.667	641.667	-115.333	-950.667
trat 3	-106.667	-199.000	-366.000	-315.000	-1803.333	-938.333	-1717.667	33.333	-975.333
trat 4	77.000	-91.667	-114.333	164.667	-223.000	-2038.333	-375.333	-220.333	-1341.667
bloco 1	-3.200	-166.000	-227.600	-157.000	-2181.600	-434.800	-1556.400	1268.400	528.600
bloco 2	63.600	-214.400	-27.200	81.200	-1351.800	-1868.600	-703.200	-1567.600	-956.800

No R, basta apenas tirar o valor do tratamento 5 e do bloco 3 dos outros parâmetro para obter os valores desta tabela.

Falta o teste de normalidade multivariada para justificar a técnica não paramétrica.

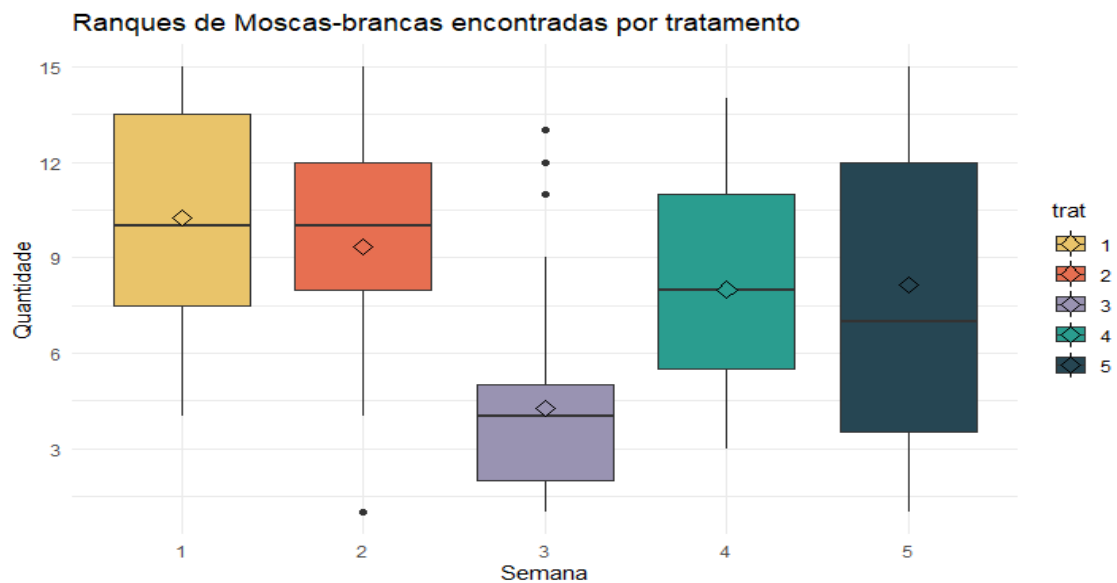
4.2 Ranques

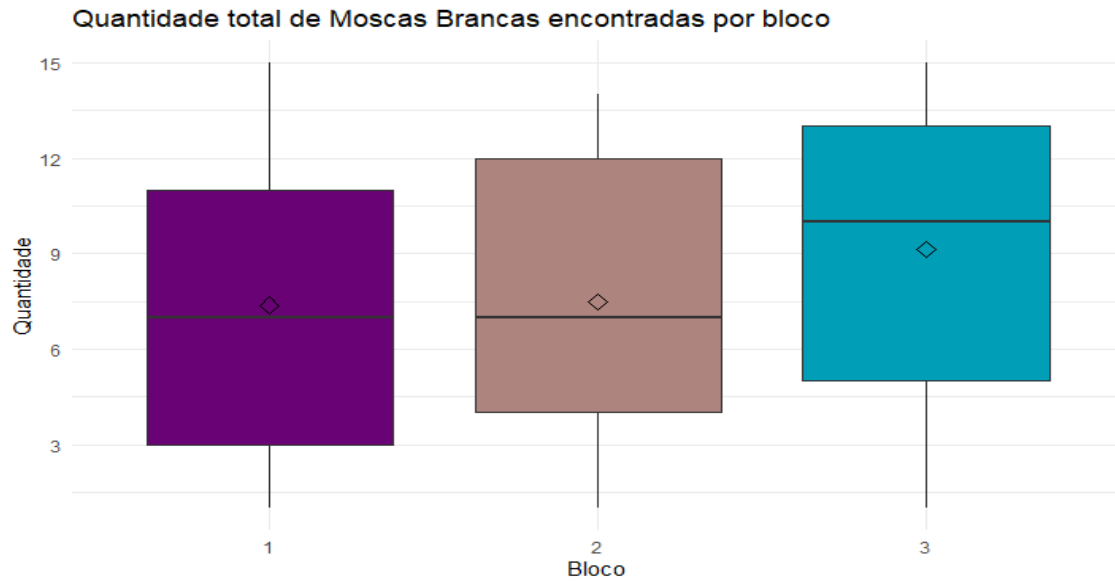
- Matriz de resíduos e de contrastes fornecidas pelo SAS em anexo (preditos-resíduos-armadilhas)
- Critério de Mauschly $W = 0.7719$ (pode ser confirmado pelo código R)
- $\gamma = 6.125$ e estatística do teste de Mauschly $\chi^2 = 21.174$ e p-valor < 0.387 (pode ser confirmado pelo R). Não há esfericidade na matriz de covariâncias estimada. É necessário usar correções ao teste F na análise dentre sujeitos.

Sphericity Tests				
Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
Orthogonal Components	35	7.55×10^{-10}	107.647	<.0001

Análise entre sujeitos					
Source	DF	Sum Square	Mean Square	F Value	Pr > F
trat	4	562.1296296	140.5324074	2.28	0.1491
bloco	2	86.8777778	43.4388889	0.71	0.5223
Error	8	492.7148148	61.5893519		

Novamente, não existem evidências para afirmar que exista algum tratamento ou bloco que seja diferente dos demais.

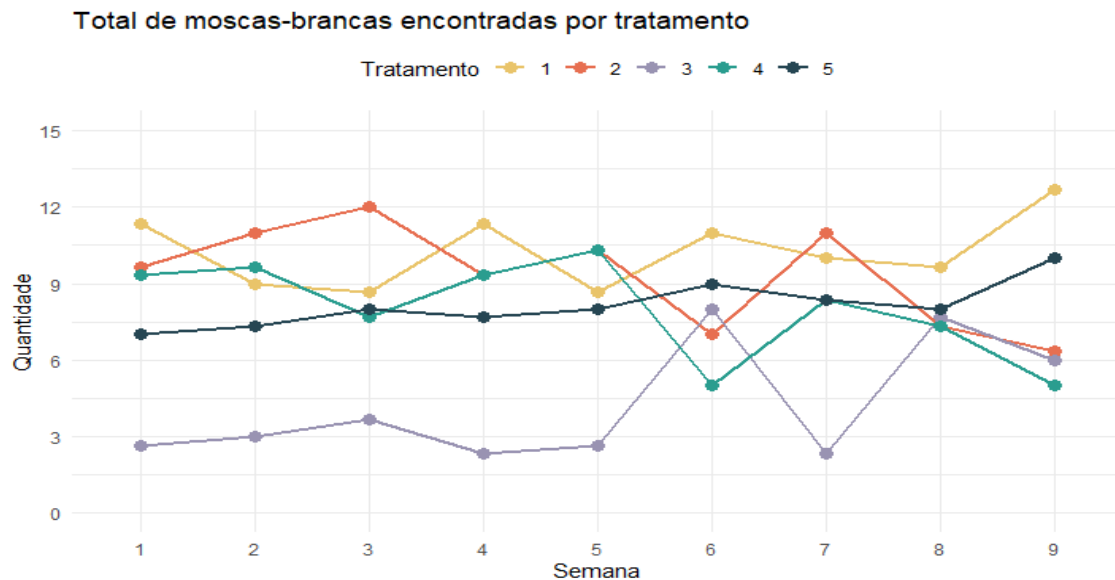




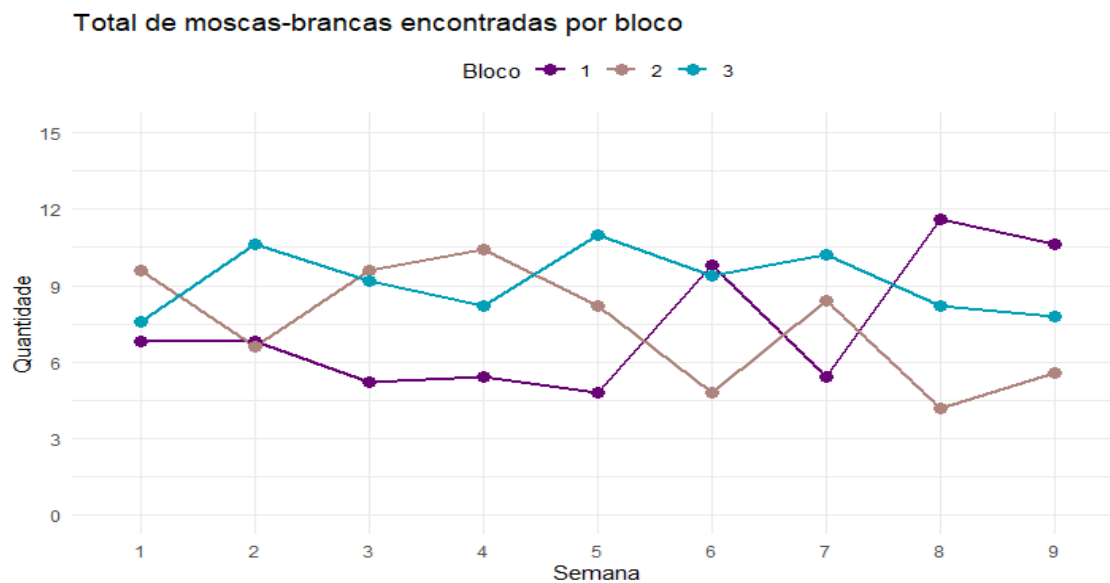
Para a análise dentre os sujeitos não existe o efeito das medidas repetidas, visto que semanalmente todos os ranques variam de 1 a 15.

Source	DF	Type III SS	Mean Square	F Value	Pr >F	Adj Pr >F	
						G - G	H-F-L
moscab	8	0.0000000	0.0000000	0.00	1.0000	1.0000	1.0000
moscab*trat	32	384.3703704	12.0115741	1.69	0.0370	0.1593	0.1194
moscab*bloco	16	539.4222222	33.7138889	4.75	<.0001	0.0064	0.0016
Error(moscab)	64	453.9851852	7.0935185				

Note que com as correções de esfericidade, o p-valor cresceu a ponto de não rejeitar a hipótese nula para interação entre tempo e tratamentos. Portanto, não existem evidências para rejeitar a hipótese de igualdade entre a interação do tempo e dos tratamentos, ou seja, os tratamentos mantêm a mesma média de ranks por todo o estudo.



Já a hipótese de igualdade entre a interação dos blocos é rejeitada, portanto existe algum bloco cuja média de ranks varia de forma significativa durante o decorrer do estudo.



Parâmetros do modelo de ranks

Par.	semana1	semana2	semana3	semana4	semana5	semana6	semana7	semana8	semana9
Intercept	6.60	10.10	9.20	7.867	11.00	10.40	10.53	8.20	9.80
trat 1	4.33	1.33	0.67	3.67	0.67	2.00	1.67	1.67	2.67
trat 2	2.67	3.50	4.0	1.67	2.33	-2.00	2.67	-0.67	-3.67
trat 3	-4.33	-4.50	-4.33	-5.33	-5.33	-1.00	-6.00	-0.33	-4.00
trat 4	2.33	2.167	-0.33	1.67	2.33	-4.00	-0.00	-0.67	-5.00
bloco 1	-0.80	-3.70	-4.00	-2.80	-6.20	0.40	-4.80	3.40	2.80
bloco 2	2.00	-4.10	0.40	2.20	-2.80	-4.60	-1.80	-4.00	-2.20

Retirar o valor do tratamento 5 e bloco 3 para cada semana para obter esses valores.