

Laboratório 03 - Planejamento e Análise de Experimentos (MAE0316)

Caio M. de Almeida - 15444560 Eduardo Yukio G. Ishihara - 15449012
Gustavo S. Garone - 15458155 Ian B. Loures - 15459667
João Victor G. de Sousa - 15463912

26 de novembro de 2025

Neste laboratório, usaremos “.” como separador decimal, quatro dígitos decimais e a *seed* 123.

Exercício 1

Item a

A unidade experimental são as folhas, enquanto a unidade observacional são seus quadrantes.

Item b

Um DCA (Delineamento Completamente Aleatorizado) realiza a alocação aleatoriamente dos tratamentos às unidades experimentais, enquanto um DABC (Delineamento Aleatorizado em Blocos Completos) designa para cada bloco pelo menos uma repetição de cada tratamento.

Tabela 1: ANOVA para o modelo DCA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Trat	2	941.4	470.7000	5.2914	0.0115
Residuals	27	2401.8	88.9556	NA	NA

Tabela 2: ANOVA para o modelo DABC

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Planta	4	1403.2	350.8000	8.0797	3e-04
Trat	2	941.4	470.7000	10.8413	5e-04
Residuals	23	998.6	43.4174	NA	NA

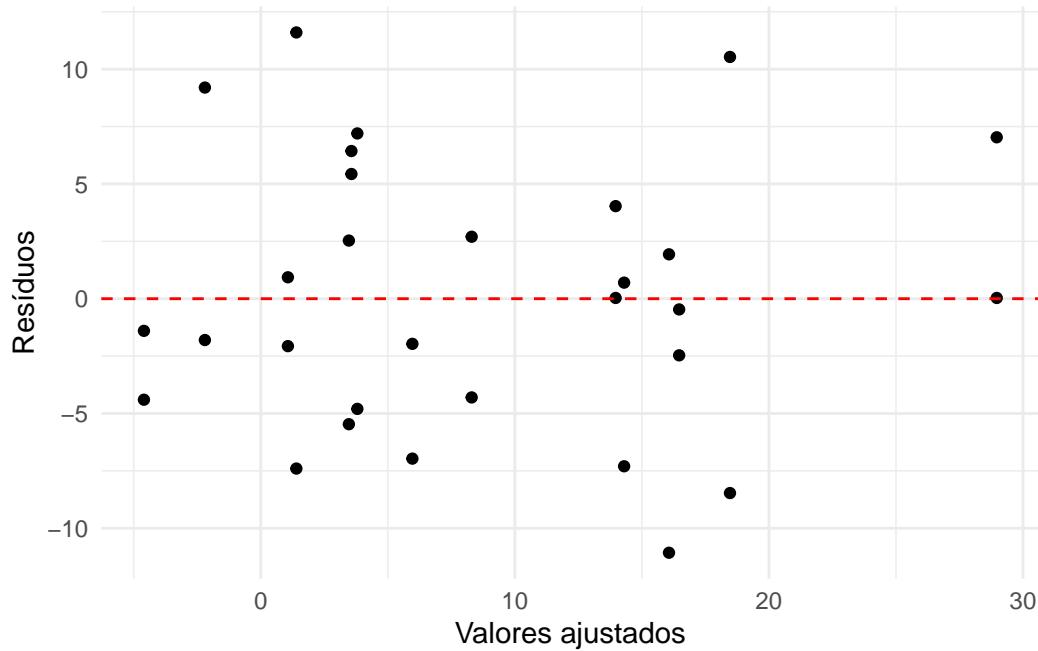
Vemos efeito significativo dos tratamentos em ambos modelos, como indica a Tabela 2 e a Tabela 1.

Porém, o modelo DABC efeito significativo para o fator Planta, o que indica que o modelo DABC é mais adequado para esses dados.

Verificação das assunções

Primeiramente, verificaremos as assunções do modelo com blocos (DABC)

Figura 1: Gráfico Resíduos x Ajustados para modelo DABC

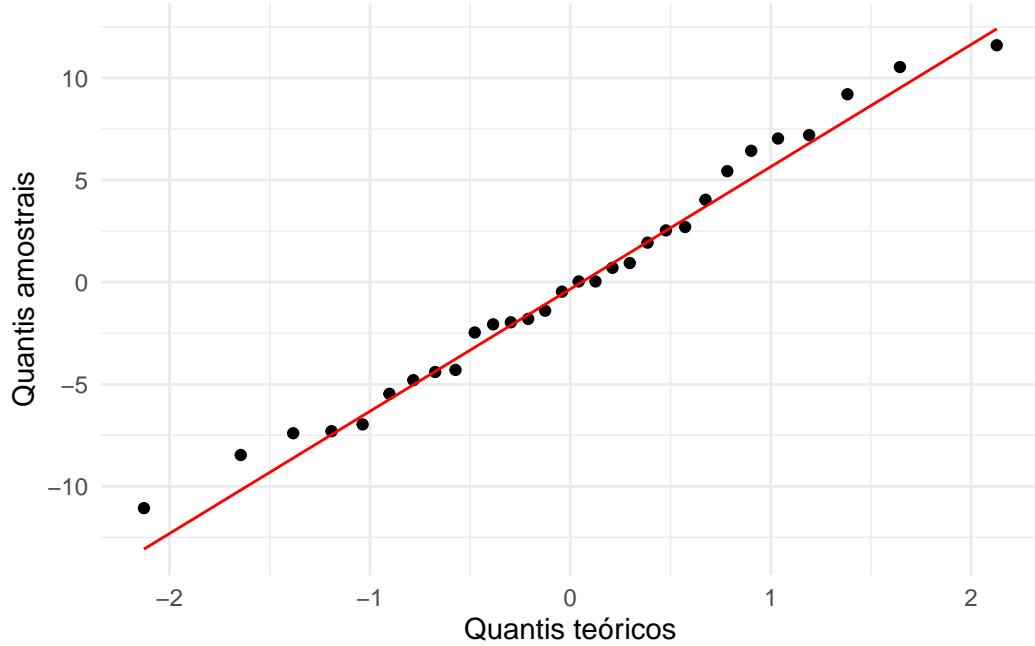


O gráfico de resíduos vs ajustados na Figura 1 não apresenta padrões evidentes, sugerindo homoscedasticidade dos resíduos. Prosseguiremos com o teste de Shapiro-Wilk para normalidade dos resíduos, com resultados na Tabela 3, e um QQ-plot para visualização, na Figura 2.

Tabela 3: Teste de Shapiro-Wilk para os resíduos do modelo DABC

Estatística	Valor-p
0.9799	0.8235

Figura 2: QQ-Plot dos resíduos do modelo DABC



Finalmente, com normalidade não rejeitada, aplicamos o teste de Bartlett para verificar homocedasticidade, com resultados na Tabela 4.

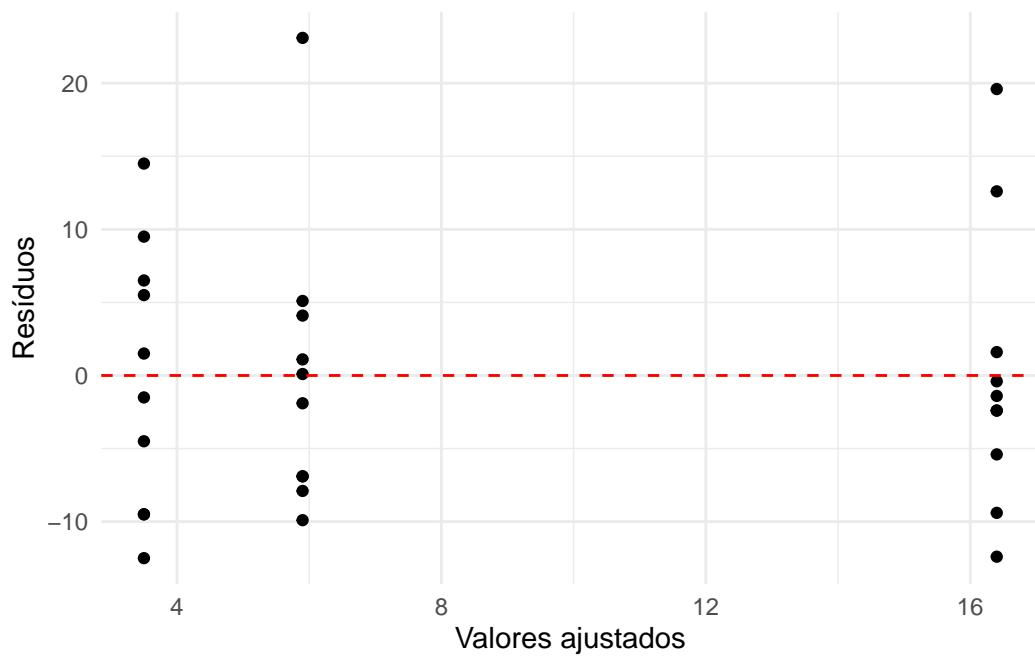
Tabela 4: Teste de Bartlett para os resíduos do modelo DABC

Estatística	Valor-p
2.6201	0.2698

Deste diagnóstico, concluímos que o modelo DABC satisfaz as assunções de normalidade e homocedasticidade dos resíduos. Dessa forma, por incluir **Planta** no modelo, o DABC é um modelo mais completo para esses dados. Sendo assim, poderíamos parar nossa análise por aqui, mas seguiremos com diagnóstico do modelo DCA.

Agora, verificaremos as assunções do modelo sem blocos (DCA)

Figura 3: Gráfico Resíduos x Ajustados para modelo DCA

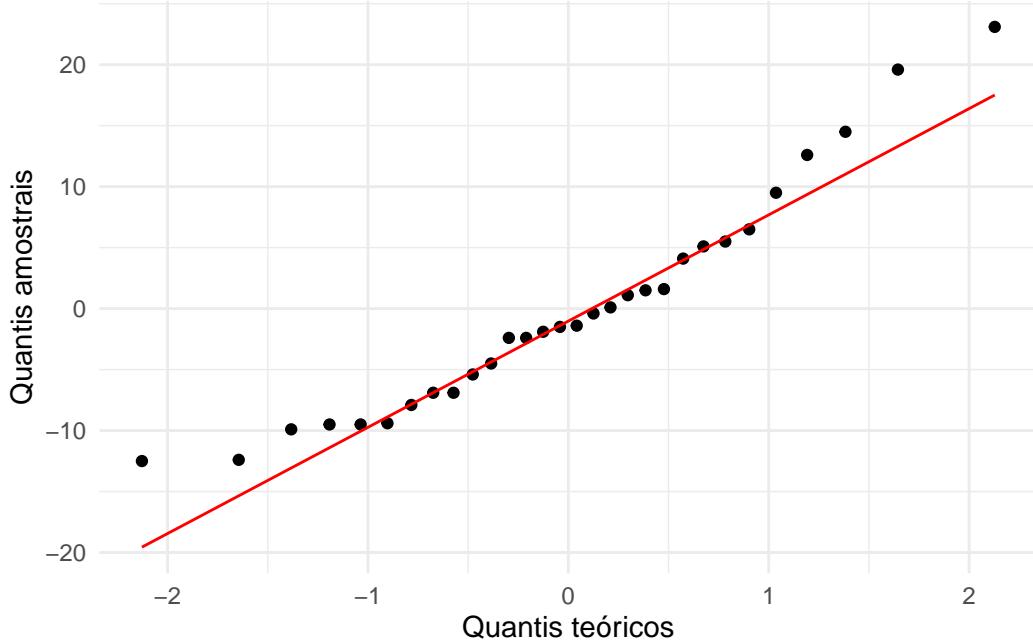


O gráfico de resíduos vs ajustados na Figura 3 pode indicar heterocedasticidade dos resíduos. Prosseguiremos com o teste de Shapiro-Wilk para normalidade dos resíduos, com resultados na Tabela 5, e um QQ-plot para visualização, na Figura 4.

Tabela 5: Teste de Shapiro-Wilk para os resíduos do modelo DCA

Estatística	Valor-p
0.9366	0.0738

Figura 4: QQ-Plot dos resíduos do modelo DCA



O QQ-plot na Figura 4 sugere que os resíduos podem não seguir uma distribuição normal, além disso, o teste de Shapiro-Wilk na Tabela 5 apresentou um **valor-p** bem mais baixo que do modelo DABC, indicando uma maior evidência contra a normalidade dos resíduos. Ainda assim, não rejeitou a normalidade ao nível de 5%.

Com desconfiança da normalidade, usaremos o teste de Levene para verificar homocedasticidade, com resultados na Tabela 6.

Tabela 6: Teste de Levene para os resíduos do modelo DCA

Estatística	Valor-p
0.0852	0.9186

O teste de Levene na Tabela 6 não rejeita a homocedasticidade dos resíduos. Ainda assim, o modelo DCA apresenta mais indícios de violação da normalidade que o modelo DABC, que também modela melhor os dados. Desta forma, escolheríamos o modelo DABC para descrever esses dados. Por curiosidade, verificamos também os R^2 ajustados dos dois modelos, na Tabela 7.

Tabela 7: R^2 ajustado dos modelos DCA e DABC

Modelo	R^2 Ajustado
DCA	0.2284
DABC	0.6234

Tabela 8: ANOVA por permutação para o modelo DABC

Analysis of Variance Table

```
Response: Obser
          Df R Sum Sq R Mean Sq Iter Pr(Prob)
Planta      4 1403.2    350.80 5000 < 2.2e-16 ***
Trat       2  941.4    470.70 5000 < 2.2e-16 ***
Residuals 23  998.6     43.42
---
Signif. codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tabela 10: Teste de aleatorização com lmPerm

Item c

Realizaremos o teste de aleatorização para um modelo DABC, considerando que as assunções de normalidade e homocedasticidade foram satisfeitas.

De acordo com a Tabela 8, o teste de aleatorização também indica efeito significativo dos tratamentos. O **valor-p** do fator de tratamentos do teste de aleatorização global é 0 (menor que 0.0001).

Exercício 2

Conferimos os resultados obtidos pela geração de dadaos e implementação da aleatorização “Manual” no R

Tabela 9: Teste de aleatorização passo-a-passo

Comparacao	Dif_observada	p_raw	p_ajustado_Holm
A vs B	1.2962	0.0032	0.0096
A vs C	0.4258	0.3970	0.3970
B vs C	0.8704	0.0744	0.1488

Vamos comparar os resultados obtidos no passo-a-passo, na Tabela 9, com os obtidos pelo pacote **lmPerm**:

```
[1] "Settings: unique SS "
```

Analysis of Variance Table

```
Response: y
          Df R Sum Sq R Mean Sq Iter Pr(Prob)
bloco      5 0.4039   0.08078 5000      0.01 **
trat       2 5.2382   2.61912 5000    <2e-16 ***
Residuals 10 3.2218   0.32218
---
Signif. codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tabela 13: ANOVA por permutação para ‘EIF4’

```
[1] "Settings: unique SS "
Analysis of Variance Table

Response: EIF4
  Df R Sum Sq R Mean Sq Iter Pr(Prob)
Fase      2   8.581   4.2906 1000     0.02 *
Residuals 69  72.954   1.0573
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exercício 3

Item a

Vamos comparar os grupos do gene **EIF4** e do miRNA **miR_150** primeiramente por medidas-resumos. Abaixo temos as tabelas com as estatísticas.

Tabela 11: Medidas-resumo para o gene ‘EIF4’

Gene	Grupo	Min	Med	Mean	Max	DP	EP
EIF4	CB (n = 14)	-0.3231	0.1326	0.6980	6.3798	1.8775	0.5018
EIF4	DS (n = 8)	-0.0639	0.0261	0.0640	0.3226	0.1242	0.0439
EIF4	FC (n = 50)	-0.3406	-0.3261	-0.1866	4.9320	0.7426	0.1050

Tabela 12: Medidas-resumo para o gene ‘miR_150’

Gene	Grupo	Min	Med	Mean	Max	DP	EP
miR_150	CB (n = 14)	-0.6604	-0.3949	0.2524	5.3202	1.5974	0.4269
miR_150	DS (n = 8)	-0.0337	0.9459	1.2883	3.9279	1.1818	0.4178
miR_150	FC (n = 50)	-0.6724	-0.4595	-0.2714	1.5428	0.5299	0.0749

Ao analisar a Tabela 11 e a Tabela 12, apesar do alto desvio-padrão, podemos perceber que tanto em **EIF4** e **miR_150**, as medidas quantílicas e a média amostral diferem entre si. Iremos testar a hipótese de que não há diferença entre os grupos usando um teste de permutação, com **maxIter** = 1000 para ser igual ao trabalho do CEA.

Como podemos perceber pelas tabelas, ao nível de significância de 0.05, há evidências para rejeitarmos ambas as hipóteses nulas iniciais, separadamente. Ou seja, há evidências para rejeitar que não há diferença entre os grupos para **EIF4** e **miR_150**.

Tabela 14: ANOVA por permutação para ‘miR_150’

```
[1] "Settings: unique SS "
Analysis of Variance Table

Response: miR_150
    Df R Sum Sq R Mean Sq Iter Pr(Prob)
Fase      2   17.852    8.9262 1000 < 2.2e-16 ***
Residuals 69   56.709    0.8219
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Item b

Primeiramente, aplicaremos a PERMANOVA usando o pacote **adonis2** para as 3 situações:

- Situação 1: 12 variáveis dependentes (os 8 genes e os 4 miRNAs) e 1 variável independente (classificação dos grupos DS, FC, e CB).
- Situação 2: 8 variáveis dependentes (os 8 genes) e 1 variável independente (a classificação dos grupos: DS, FC, e CB).
- Situação 3: 4 variáveis dependentes (os 4 miRNAs) e 1 variável independente (a classificação dos grupos: DS, FC, e CB).

Adotaremos um nível de significância $\alpha = 0.05$.

Abaixo, mostramos as saídas da PERMANOVA para as 3 situações, respectivamente.

Tabela 13: PERMANOVA para a situação 1

	Df	SumOfSqs	R2	F	Pr(>F)
Model	2	171.6775	0.1909	8.1379	0.001
Residual	69	727.8141	0.8091	NA	NA
Total	71	899.4916	1.0000	NA	NA

Tabela 14: PERMANOVA para a situação 2

	Df	SumOfSqs	R2	F	Pr(>F)
Model	2	83.7435	0.1358	5.4194	0.002
Residual	69	533.1174	0.8642	NA	NA
Total	71	616.8608	1.0000	NA	NA

Tabela 15: PERMANOVA para a situação 3

	Df	SumOfSqs	R2	F	Pr(>F)
Model	2	87.9340	0.3111	15.5818	0.001
Residual	69	194.6968	0.6889	NA	NA
Total	71	282.6308	1.0000	NA	NA

No teste, a hipótese nula é de igualdade dos grupos por variável dependente (a depender das 3 situações acima mencionadas), para todas variáveis dependentes.

Para as 3 situações, a um nível de significância α , há evidências para rejeitar a hipótese nula, ou seja, caso considerarmos um espaço de hipóteses fechado, há evidências de que existe diferença em pelo menos um grupo dentro de uma variável dependente, pelo menos em uma variável dependente.

Aplicaremos testes post hoc para entendermos quais grupos diferem, em especial, o post hoc de Bonferroni. Usaremos a função **pairwise.adonis**.

Abaixo, mostramos as saídas do post hoc para as 3 situações, respectivamente.

Tabela 16: Post Hoc para a situação 1

pairs	Df	SumsOfSqs	F.Model	R2	p.value	p.adjusted	sig
FC vs CB	1	127.6318	11.1454	0.1524	0.001	0.003	*
FC vs DS	1	44.8480	8.2711	0.1287	0.001	0.003	*
CB vs DS	1	72.3255	3.2727	0.1406	0.005	0.015	.

Tabela 17: Post Hoc para a situação 2

pairs	Df	SumsOfSqs	F.Model	R2	p.value	p.adjusted	sig
FC vs CB	1	59.8248	7.0551	0.1022	0.001	0.003	*
FC vs DS	1	22.6261	6.1441	0.0989	0.004	0.012	.
CB vs DS	1	40.3831	2.4162	0.1078	0.055	0.165	

Tabela 18: Post Hoc para a situação 3

pairs	Df	SumsOfSqs	F.Model	R2	p.value	p.adjusted	sig
FC vs CB	1	67.8070	22.8164	0.2690	0.001	0.003	*
FC vs DS	1	22.2219	12.7739	0.1857	0.001	0.003	*
CB vs DS	1	31.9424	5.9307	0.2287	0.007	0.021	.

Para a situação 1 (Tabela 16) as comparações entre os pares FC vs CB (valor-p = 0.001, valor-p ajustado = 0.003), FC vs DS (valor-p = 0.001, valor-p ajustado = 0.003) e CB vs DS (valor-p = 0.005, valor-p ajustado = 0.015) foram todas significativas estaticamente, portanto havendo evidências de que os grupos apresentam diferenças entre si.

Para a situação 2 (Tabela 17) as comparações entre os pares FC vs CB (valor-p = 0.001, valor-p ajustado = 0.003) e FC vs DS (valor-p = 0.004, valor-p ajustado = 0.012) foram significativas estaticamente, portanto havendo evidências de que o grupo FC difere de CB e que o grupo FC difere de DS. A comparação CB vs DS (valor-p = 0.055, valor-p ajustado = 0.165) não foi estatisticamente significante, não havendo evidências para rejeitar que os grupos CB e DS possuam igualdades.

Para a situação 3 (Tabela 18) as comparações entre os pares FC vs CB (valor-p = 0.001, valor-p ajustado = 0.003), FC vs DS (valor-p = 0.001, valor-p ajustado = 0.003) e CB vs DS (valor-p = 0.007, valor-p ajustado = 0.021) foram todas significativas estaticamente, portanto havendo evidências de que os grupos apresentam diferenças entre si.

Adicionalmente, iremos aplicar os testes **nonpartest** sobre os dados. É uma coleção de testes não-paramétricos multivariados, que, especialmente, tem uma interpretação bem mais fácil do que o **adonis**: adquire-se o efeito de tratamento “k” caso o que pode ser definido como a probabilidade de que um sujeito escolhido aleatoriamente do tratamento “k”, que neste estudo não é outra coisa que o grupo, exiba um nível de expressão mais alto do que um sujeito escolhido aleatoriamente de qualquer um dos grupos de tratamento, incluindo o tratamento “k”. Essa probabilidade é mostrada nas tabelas ‘releffects’ a seguir.

Abaixo, mostramos as saídas do `nonpartest` para as 3 situações, respectivamente, com 2 tabelas por situação.

Tabela 19: Tabela ‘results’: nonpartest para a situação 1

	Test Statistic	df1	df2	P-value	Permutation Test p-value
ANOVA type test p-value	18.244	8.815	165.1599	0	0
McKeon approx. for the Lawley Hotelling Test	14.140	24.000	96.4111	0	0
Muller approx. for the Bartlett-Nanda-Pillai Test	9.261	24.972	121.0743	0	0
Wilks Lambda	11.589	24.000	116.0000	0	0

Tabela 20: Tabela ‘releffects’: nonpartest para a situação 1

	TP53	CDK	PTEN	MDM4	MYB	SHIP	EIF4	CEBPB	155	150	34a	175p
CB	0.2297	0.6508	0.2455	0.6419	0.6369	0.5719	0.7748	0.2932	0.8492	0.5293	0.8214	0.3229
DS	0.6927	0.6701	0.8889	0.8941	0.2457	0.6155	0.8246	0.6450	0.6163	0.8724	0.4774	0.1337
FC	0.5449	0.4306	0.5090	0.3972	0.5024	0.4614	0.3711	0.5347	0.3836	0.4322	0.4136	0.6082

Obs: a Tabela 20 tem alguns nomes de coluna alterados para a visualização. A 2^a coluna corresponde ao gene CDKN1A e as últimas 4 colunas correspondem aos miRNAs, respectivamente: miR_155, miR_150, miR_34a, miR_17_5p

Tabela 21: Tabela ‘results’: nonpartest para a situação 2

	Test Statistic	df1	df2	P-value	Permutation Test p-value
ANOVA type test p-value	16.570	6.695	125.4337	0	0
McKeon approx. for the Lawley Hotelling Test	14.084	16.000	97.9292	0	0
Muller approx. for the Bartlett-Nanda-Pillai Test	10.584	16.666	129.4215	0	0
Wilks Lambda	12.325	16.000	124.0000	0	0

Tabela 22: Tabela ‘releffects’: nonpartest para a situação 2

	TP53	CDKN1A	PTEN	MDM4	MYB	SHIP	EIF4	CEBPB
CB	0.2297	0.6508	0.2455	0.6419	0.6369	0.5719	0.7748	0.2932
DS	0.6927	0.6701	0.8889	0.8941	0.2457	0.6155	0.8246	0.6450
FC	0.5449	0.4306	0.5090	0.3972	0.5024	0.4614	0.3711	0.5347

Tabela 23: Tabela ‘results’: nonpartest para a situação 3

	Test Statistic	df1	df2	P-value	Permutation Test p-value
ANOVA type test p-value	22.258	6.658	124.7402	0	0
McKeon approx. for the Lawley Hotelling Test	13.118	8.000	91.9913	0	0
Muller approx. for the Bartlett-Nanda-Pillai Test	11.879	8.341	137.7690	0	0
Wilks Lambda	12.542	8.000	132.0000	0	0

Tabela 24: Tabela ‘releffects’: nonpartest para a situação 3

	miR_155	miR_150	miR_34a	miR_17_5p
CB	0.8492	0.5293	0.8214	0.3229
DS	0.6163	0.8724	0.4774	0.1337
FC	0.3836	0.4322	0.4136	0.6082

Ao olhar a Tabela 19, a Tabela 21 e a Tabela 23, podemos perceber que em todas as situações o valor-p e o valor-p de permutação foram iguais a 0 entre os grupos, para todos os testes multivariados de igualdade de grupos. Ou seja, nas 1000 permutações, nenhuma contribuiu para H_0 .

Se desejarmos, podemos também analisar a Tabela 20, a Tabela 22 e a Tabela 24 para avaliar os efeitos relativos entre cada variável dependente e seus grupos. Acreditamos que os testes anteriores já são suficientes.

Conclusão: Foi possível verificar que as expressões dos genes e dos miRNAs apresentaram diferenças significativas entre os grupos estudados.