

QUESTÕES

1. Um pesquisador realizou um experimento com quatro grupos de sujeitos e obteve os seguintes resultados:

Tabela de ANOVA - Pesquisador 1

Source	DF	SS	MS
Grupo	3	345,0	115,0
Error	16	402,8	25,2
Total	19	747,8	

Um segundo pesquisador replicou o estudo usando uma outra amostra de sujeitos da mesma população. Seus resultados foram os seguintes:

Tabela de ANOVA - Pesquisador 2

Source	DF	SS	MS
Grupo	3	107,7	35,9
Error	16	358,8	22,4
Total	19	466,6	

- Calcule as estatísticas F em cada caso e interprete os resultados. Qual hipótese pode ser testada através destas estatísticas?
- Que suposições devem estar satisfeitas para a validade do teste realizado em a)?
- Compare os Quadrados Médios do resíduo. O que você recomendaria aos pesquisadores?
- Por que um Teste de Aleatorização do efeito dos grupos poderia ser indicado aos pesquisadores? Como esse teste poderia ser realizado?

2. A tabela a seguir é baseada em dados da imunização de duas vacinas de acordo com a presença de um alelo alvo em voluntários do estudo clínico.

Vacina 1			Vacina 2		
Imunizados	Não	Sim	Imunizados	Não	Sim
Alelo-	160	16252	Alelo-	90	15870
Alelo+	51	41236	Alelo+	120	38492

- A razão de chances (também conhecida como odds ratio, OR) é uma medida definida, neste caso, como:

$$\left[\frac{P(\text{Sim}|\text{Alelo}+)}{P(\text{Não}|\text{Alelo}+)} \right] / \left[\frac{P(\text{Sim}|\text{Alelo}-)}{P(\text{Não}|\text{Alelo}-)} \right]$$

Para os dados de cada vacina, calcule o OR. Interprete.

- O teste Qui-Quadrado destes dados forneceu os seguintes resultados:

Vacina 1: $\text{Chi-Sq} = 166.562 + 0.611 + 66.210 + 0.243 = 233.626$

Vacina 2: $\text{Chi-Sq} = 13.303 + 0.051 + 5.499 + 0.021 = 18.875$

Que teste está sendo realizado? Calcule o nível descritivo (nominal) do teste e interprete.

Um modelo de regressão logística foi usado na análise dos dados de cada vacina. Os resultados estão apresentados a seguir:

Vacina 1:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.62080	0.07945	58.16	<2e-16
as.factor(alel) 1	2.07444	0.16107	12.88	<2e-16

Vacina 2:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.1724	0.1057	48.931	< 2e-16
as.factor(alel) 1	0.5983	0.1398	4.281	1.86e-05

- c) Com base nestas análises, comente a afirmação: “Estar vacinado protege da doença. Mas algumas formas de imunização por vacina podem ser mais eficazes que outras”

3. Duas espécies (A e B) de peixes foram avaliadas em um experimento para verificar o efeito da insulina na produção de glicogênio muscular. De cada espécie, 18 peixes foram aleatorizados aos tratamentos, um grupo para receber a solução salina (placebo) e o outro a solução com insulina. Após 24 horas os animais foram sacrificados e mediu-se a quantidade de glicogênio muscular (mg/g). Medidas descritivas dos dados obtidos são apresentadas a seguir.

Espécie	Tratamento	Média	Desvio Padrão
A	Salina	1,200	0,309
	Insulina	0,793	0,231
B	Salina	1,200	1,190
	Insulina	0,776	0,800

- a) Calcule os coeficientes de variação e comente sobre o padrão de variação da resposta desses grupos.
- b) Construa o gráfico de perfis de médias. Visualmente parece existir efeito de interação entre os fatores espécie e tratamento? Justifique.
- c) A seguinte tabela de ANOVA foi obtida na análise destes dados. Complete e conclua sobre o efeito da insulina.

FV	GL	SQ	QM	F	P
Espécie	_____	0,00065	_____	_____	_____
Trat	_____	1,55376	_____	_____	_____
E*Trat	_____	_____	1,55506	_____	_____
Error	_____	5,92000	_____	_____	_____
Total	_____	_____	_____	_____	_____

- d) Obtenha a tabela de ANOVA para estes dados considerando um único fator em 4 níveis. Interprete os resultados.
- e) Compare as análises realizadas em c), d).

4. Um pesquisador estudou os efeitos de três dietas experimentais na variação do nível de lipídeos no plasma. Quinze indivíduos masculinos, com pesos corporais dentro da normalidade, foram agrupados em cinco blocos de acordo com a idade. Dentro de cada bloco três dietas, controlando o conteúdo lipídico (D1=muito baixo, D2=moderado e D3=baixo), foram aleatoriamente atribuídas aos três indivíduos. Após um período de tempo fixado foi mensurada a redução do nível de lipídeos plasmático. Os dados estão apresentados a seguir.

Blocos	FxIdade	D1	D2	D3
1	15-24	0,73	0,67	0,15
2	25-34	0,86	0,75	0,21
3	35-44	0,94	0,81	0,26
4	45-54	1,40	1,32	0,75
5	55-64	1,62	1,41	0,78

- Que tipo de delineamento experimental foi adotado? Justifique.
- Há indicação de efeito de interação entre os fatores FxIdade e Dieta? Construa o gráfico de perfis individuais de resposta para avaliar isso.
- Se essa interação for significativa, o que deveria ser recomendado ao pesquisador?
- Interprete a tabela de análise de variância destes dados apresentada a seguir.

FV	GL	SQ	QM	F	P
trat	2	1,32028	0,66014	273,35	0,000
bloco	4	1,41896	0,35474	146,89	0,000
Error	8	0,01932	0,00242		
Total	14	2,75856			

5. Um estudo considera o efeito da acupuntura (A1=ausente A2=presente) e de um analgésico (B1=ausente B2=presente) na redução da dor pós-cirúrgica. 32 pacientes foram agrupados, em grupos de tamanho quatro, de acordo com uma avaliação inicial de seu nível de tolerância à dor.

- O nível inicial de tolerância à dor foi considerado como variável de bloqueio no experimento, por quê? Como deve ter sido feita a aleatorização dos tratamentos aos pacientes?
- A tabela de ANOVA da análise destes dados é dada a seguir. Que suposições estão sendo feitas aos dados para a validade desta análise? Se for adotado um nível de significância de 5%, há efeito de interação? E se for de 10%?

FV	GL	SQ	QM	F	P
bloco	7	5,5988	0,7998	55,30	0,000
A	1	2,3112	2,3113	159,79	0,000
B	1	3,3800	3,3800	233,68	0,000
A*B	1	0,0450	0,0450	3,11	0,092
Error	21	0,3037	0,0145		
Total	31	11,6388			

- Obtenha a tabela de ANOVA do modelo reduzido, sem o efeito de interação. Há efeito da acupuntura na redução esperada da dor? Há efeito do analgésico?
- Resultados de Testes t estão apresentados a seguir. Interprete os resultados do teste e os intervalos de confiança. Que tipo de Teste t foi realizado, para amostras independentes ou pareadas? Há evidência de diferenças nos efeitos esperados desses dois tratamentos? Justifique.

T-Test - Acupuntura

	N	Mean	StDev	SE Mean
A1	16	0,887	0,515	0,129
A2	16	1,425	0,597	0,149
Difference	16	-0,5375	0,1708	0,0427
95% CI for mean difference: (-0,6285; -0,4465)				
T-Test of mean difference=0 (vs not = 0): T-Value=-12,59 P-Value=0,000				

T-Test - Analgésico

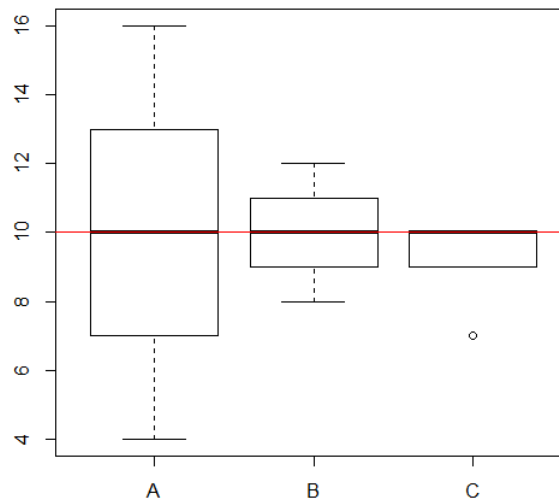
	N	Mean	StDev	SE Mean
B1	16	0,831	0,503	0,126
B2	16	1,481	0,546	0,136
Difference	16	-0,6500	0,1966	0,0492

95% CI for mean difference: (-0,7548; -0,5452)
T-Test of mean difference=0 (vs not = 0): T-Value=-13,22 P-Value=0,000

6. Um laboratório clínico precisa decidir por um dentre três instrumentos (A, B e C) que será utilizado para fazer dosagens químicas no sangue. Foram preparadas soluções contendo uma concentração conhecida (10 mg/ml) da substância a ser dosada. Os resultados obtidos com cada instrumento seguem abaixo.

Instr.	Medidas										Média	dp
A:	5	10	7	15	16	12	4	8	10	13	10	4.06
B:	11	10	11	10	12	9	10	8	9	10	10	1.15
C:	10	10	9	9	10	10	10	11	7	10	9.5	0.97

- a) O boxplot destes dados está apresentado a seguir. Comente sobre o padrão de variação das medidas feitas por esses instrumentos. O valor 10mg/ml representa que quantil da distribuição dos dados?



- b) Para cada instrumento calcule o intervalo ($\text{média} \pm 2 \cdot \text{erro padrão}$). Interprete.

Em medidas clínicas, três critérios são utilizados frequentemente: precisão, não-vício e exatidão.

Precisão: refere-se à dispersão de um conjunto de observações. Quanto menor a variabilidade maior a precisão.

Não-vício: refere-se à tendência de um conjunto de medidas ser igual a um verdadeiro valor.

Para um instrumento ser **Exato**, suas leituras precisam ser tanto precisas quanto não-viciadas.

- c) Descreva os instrumentos em termos destas definições.
d) Qual instrumento você recomendaria ao laboratório? Justifique.

7. A tabela a seguir reproduz os resultados de um experimento que visa verificar a eficácia de uma nova vacina.

Grupo	Doente	Não Doente	Total
Vacinado	11	54	65
Não vacinado	25	03	28
Total	36	57	93

- Se na coleta destes dados foi usado um estudo observacional prospectivo, como os dados foram coletados?
- E se foi usado um estudo retrospectivo?
- E se foi usado um estudo transversal?
- Sob cada um destes casos, calcule a estatística razão de chances (odds ratio, OR). Interprete.

8. É conhecido que o tratamento com aspirina induz micro sangramentos no tecido gastrointestinal. Um pesquisador está interessado em comparar a perda média de sangue provocada pelo uso de dois medicamentos: aspirina e um novo agente R-803. No experimento conduzido o efeito de um placebo foi também observado. Além disso, cada um dos pacientes do estudo recebeu todos os tratamentos, sendo a ordem aleatorizada e respeitando um período de limpeza do organismo entre as intervenções. Os resultados da análise destes dados estão apresentados a seguir:

Source	DF	SS	MS	F	P
Paciente	_____	58.571			
Trat	_____	257.275			
Error	_____				
Total	26	429.034			

- Supondo um delineamento aleatorizado em blocos completos, preencha a tabela de ANOVA. Que suposições estão envolvidas nesta análise?
- Existe diferença na perda média de sangue entre os tratamentos? Justifique.
- Por que “Paciente” está sendo analisado como uma fonte de variação?
- Por que foi introduzido no estudo o tratamento Placebo?
- Os resultados a seguir foram obtidos da aplicação do teste de comparações múltiplas de Tukey. Conclua sobre o efeito dos tratamentos.

Difference	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
R803-Placebo	-0.1322	1.254	-0.1055	0.9939
Aspir-Placebo	6.4811	1.254	5.1691	0.0003
Aspir-R803	6.613	1.254	5.275	0.0002

9. Proponha um delineamento experimental para as situações a seguir. Indique a resposta sob estudo, os fatores, possíveis blocos ou variáveis de controle, a amostragem de unidades amostrais, as unidades experimentais e procedimentos de aleatorização que podem estar envolvidos. Em cada caso, apresente também as fontes de variação e os graus de liberdade de uma tabela de ANOVA correspondente ao ajuste de um modelo linear aos dados.

Um dermatologista quer estudar a eficiência de quatro fórmulas diferentes de loção usadas no tratamento de uma alergia que causa lesão principalmente nos membros dos indivíduos. Medidas como a área da lesão, bem como o grau de vermelhidão e a temperatura na região da lesão são indicadores do comprometimento da alergia. Para o estudo, ele poderá dispor de pacientes de um centro especializado no atendimento desse tipo de alergia, mas desconfia que algumas fontes de variação deverão ser controladas, por exemplo:

- Existe variabilidade individual na resposta dos pacientes ao tratamento.

- b) Além disso (a), é conhecido que existe diferenças entre os tratamentos devido ao sexo e faixa etária do paciente (jovem, adulto ou idoso).
- c) Além destes (a e b), o efeito do tratamento pode variar de acordo com a ancestralidade do paciente.
- d) Além destes (a, b e c), o efeito do tratamento pode variar de acordo com o genótipo de dois marcadores moleculares conhecidos, SNP1 e SNP2.

10. Um estudo foi conduzido para avaliar o efeito do genótipo de um marcador (G1, G2 e G3) no tempo de reabilitação de pacientes que sofreram acidente com lesão grave. 24 registros de pacientes masculinos, com idades entre 18 e 30 anos, foram amostrados de um Centro de Reabilitação e foi então obtido o tempo (meses) que foi necessário para a reabilitação do paciente. Além disso, foi obtido o consentimento dos mesmos para a genotipagem do marcador. O tempo médio é apresentado a seguir.

	G1	G2	G3
Tamanho Amostral	8	10	6
Tempo Médio (dias)	38	32	24

- a) Que tipo de estudo foi realizado? Se uma análise de variância será conduzida, apresente o modelo a ser adotado e as suposições envolvidas.
- b) Preencha convenientemente a seguinte tabela de análise de variância desses dados.

Fonte Variação	Grau Liberdade	Soma Quadrados	Quadrado Médio	Estatística F
Grupo				
Resíduo				
Total		1088,0		

- c) Qual é a hipótese de interesse sob teste? Calcule e interprete o nível descritivo do teste. Qual é a conclusão?
- d) Neste caso é possível construir um Teste de Aleatorização da hipótese de interesse? Justifique. (Pense em usar o princípio da Aleatorização Mendeliana)

11. Considere um marcador molecular com classes genotípicas dadas por aa=0, Aa=1 e AA=2, de acordo com o número de alelos de risco que o indivíduo carrega. Se o efeito deste marcador será avaliado na altura dos indivíduos de uma população por meio de um modelo linear, descreva como podemos definir as variáveis preditoras (X) e os parâmetros (β) do modelo em cada caso a seguir:

- a) efeito aditivo (linear) e de dominância (desvio da linearidade) do marcador.
- b) efeito linear do marcador.
- c) efeito dominante do marcador: basta uma cópia do alelo A para um possível efeito.
- d) efeito recessivo do marcador: é preciso 2 cópias do alelo A para um possível efeito.

Os resultados a seguir correspondem às questões 12 e 13:

Considere os dados “Express” que contém medidas da expressão gênica de quatro regiões do cromossomo 6 (E1, E2, E3, E4) associadas a uma resposta imunológica, a qual foi avaliada em 150 pacientes submetidos ao transplante de medula e classificados de acordo com o genótipo de um SNP alvo do estudo (aa, Aa e AA). A seguir são apresentados alguns resultados da análise destes dados.

Estatísticas descritivas (independente das classes genotípicas)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
E1	1.000	1.600	4.350	3.758	5.100	6.900
E2	4.300	5.100	5.800	5.843	6.400	7.900
E3	0.100	0.300	1.300	1.199	1.800	2.500
E4	2.000	2.800	3.000	3.057	3.300	4.400

Centróides das classes genotípicas

	aa	Aa	AA
E1	1.462	4.260	5.552
E2	5.006	5.936	6.588
E3	0.246	1.326	2.026
E4	3.428	2.770	2.974

Matriz de covariância (independente de classe genotípica): S

	E1	E2	E3	E4
E1	3.12	1.27	1.30	-0.33
E2	1.27	0.69	0.52	-0.04
E3	1.30	0.52	0.58	-0.12
E4	-0.33	-0.04	-0.12	0.19

Autovalores de S: 4.23 0.24 0.08 0.02

Autovetores de S:

	[,1]	[,2]	[,3]	[,4]
[1,]	0.86	-0.17	0.08	0.48
[2,]	0.36	0.66	-0.58	-0.32
[3,]	0.36	-0.08	0.55	-0.75
[4,]	-0.08	0.73	0.60	0.32

Matriz de correlação (independente de classe genotípica): R

	E1	E2	E3	E4
E1	1.00	0.87	0.96	-0.43
E2	0.87	1.00	0.82	-0.12
E3	0.96	0.82	1.00	-0.37
E4	-0.43	-0.12	-0.37	1.00

Autovalores de R: 2.92 0.91 0.15 0.02

Autovetores de R:

	[,1]	[,2]	[,3]	[,4]
[1,]	-0.58	-0.02	-0.14	0.80
[2,]	-0.52	-0.38	0.72	-0.26
[3,]	-0.56	-0.07	-0.63	-0.52
[4,]	0.27	-0.92	-0.24	0.12

Sb: matriz de covariância ENTRE as classes genotípicas

	E1	E2	E3	E4
E1	437.10	165.25	186.77	-57.24
E2	165.25	63.21	71.28	-19.95
E3	186.77	71.28	80.41	-22.93
E4	-57.24	-19.95	-22.93	11.34

Sw: matriz de covariância DENTRO das classes genotípicas (sob homocedasticidade)

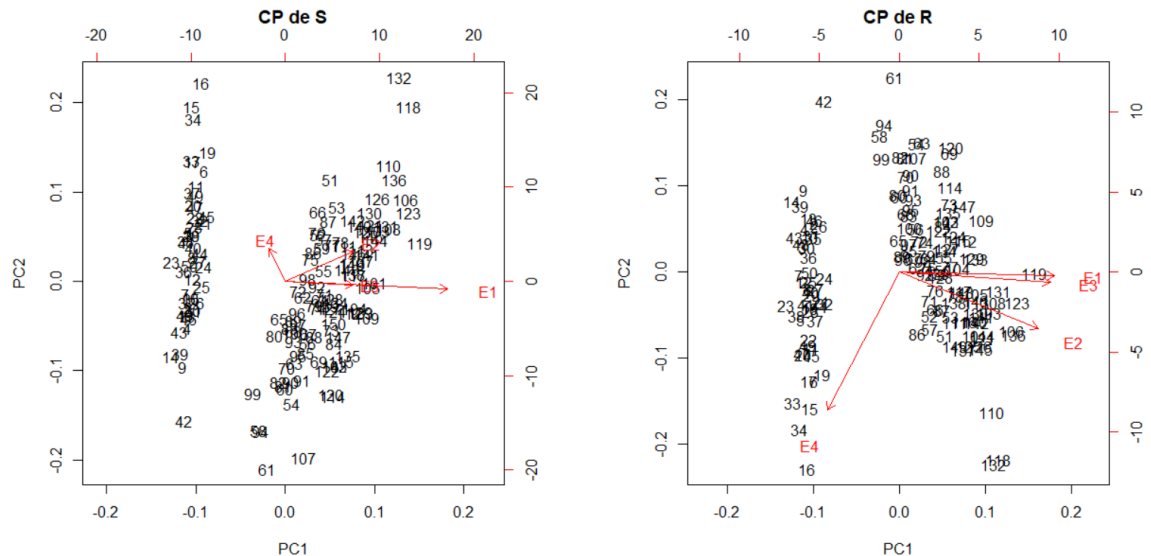
	E1	E2	E3	E4
E1	27.22	24.62	6.27	8.12
E2	24.62	38.96	5.65	13.63
E3	6.27	5.65	6.16	4.81
E4	8.12	13.63	4.81	16.96

Autovalores de $(S_w)^{-1}S_b$: 32.19 0.285

Autovetores de $(S_w)^{-1}S_b$:

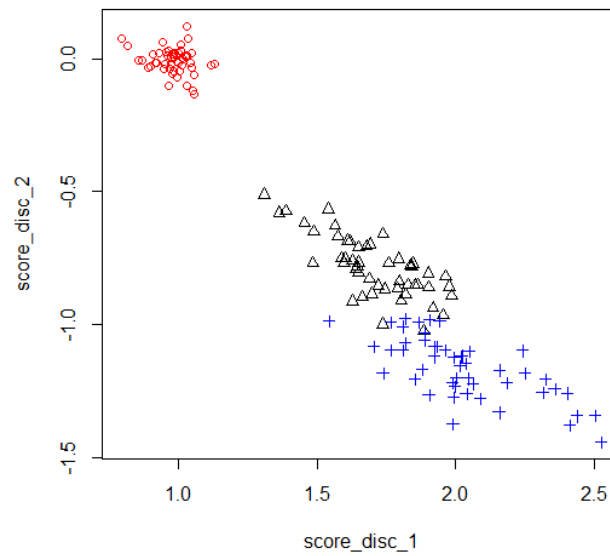
	[,1]	[,2]
[1,]	0.1815	0.0768
[2,]	-0.0684	-0.0019
[3,]	0.2318	-0.2341
[4,]	-0.1265	-0.1785

12. As figuras a seguir mostram o biplot da análise de Componentes Principais (CP) dos dados originais e normalizados, respectivamente.



- Em cada caso, os dois primeiros CP explicam quanto da variabilidade total dos dados? Justifique. Como decidir quantos componentes devem ser usados na redução de dimensionalidade dos dados?
- Como decidir qual análise usar para a redução de dimensionalidade dos dados, aquela baseada em S ou em R ?
- Como estes componentes principais estão definidos em termos das cargas que são impostas às variáveis do estudo?
- Hipoteticamente, considere um vetor de dados observados e calcule o escore do CP, nos dois casos (da análise via S e via R).

13. Há interesse em verificar se há diferença entre as classes genotípicas. A função discriminante linear de Fisher foi construída e obteve-se o gráfico a seguir com a dispersão das observações de acordo com os dois primeiros escores discriminantes ($aa=o$, $Aa=\Delta$, $AA=+$).



- a) A solução de Fisher se baseia na decomposição espectral de qual matriz? Quais são as suposições envolvidas? Essa solução é ótima em que sentido?
- b) Escreva as duas primeiras funções discriminantes cujos escores estão indicados no gráfico acima. Quais variáveis são mais importantes para a discriminação? Para este problema, quantas funções discriminantes (de Fisher) podem ser construídas.
- c) Com base na matriz de classificação (empírica) apresentada abaixo, o que se pode dizer sobre o classificador de Fisher? Como um método de validação cruzada poderia ser usado neste caso?

Grupo predito	Grupo genotípico		
	aa	Aa	AA
aa	50	0	0
Aa	0	48	2
AA	0	1	49

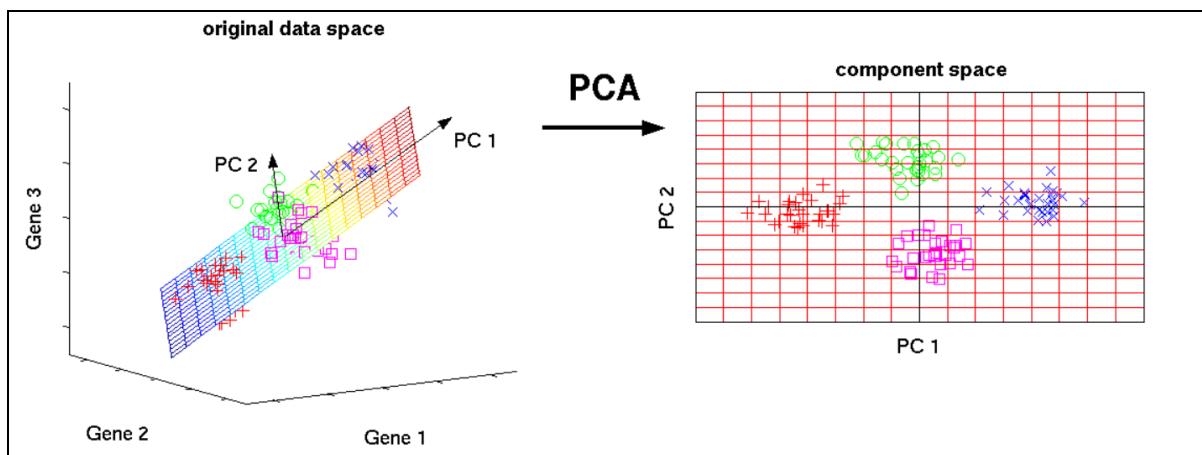
- d) O centroide dos escores das duas primeiras funções discriminantes para os três grupos é:

	LD1	LD2
aa:	0.9821	-0.0083
Aa:	1.7079	-0.7835
AA:	2.0232	-1.1614

Qual é o grupo predito para a observação com valores de expressão gênica dado por (6.3, 2.7, 4.9, 1.8)? Justifique.

14. 2. A visualização de dados é uma das etapas mais importante, se não a mais importante, na análise de dados. Contudo, para dados de alta dimensão, por exemplo, os dados de expressão gênica, já conhecidos de vocês, de dimensão $n=189$ por $p=22.215$, poderiam ser construídos 17.766 gráficos de dispersão para pares de indivíduos, ou um número ainda maior para pares de variáveis. Como é possível representar dados de alta dimensão em um único gráfico de dispersão? Que propriedades ótimas das linhas bem como das colunas da matriz de dados podem estar preservadas nessas representações?

15. A figura a seguir permite a visualização de um conjunto de dados originais de expressão de três genes (Genes 1, 2 e 3) bem como dos correspondentes dois primeiros componentes principais.



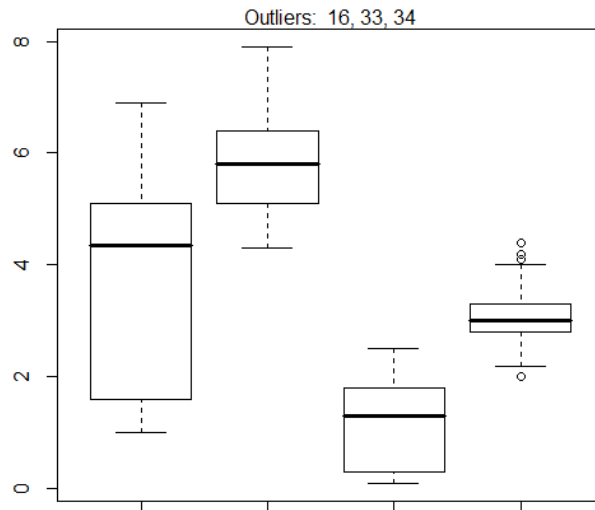
- Os componentes principais satisfazem quais propriedades? O que garante que dois componentes reduzem bem os dados?
- Se $n \ll p$ (problema big-p), qual a recomendação para o cálculo dos CP? Quais são os problemas encontrados na análise de tais dados (big-p) e como isso pode ser resolvido?
- Para garantir a separação dos grupos, que análise pode ser recomendada na redução de dimensionalidade? Nesse caso, como duas direções ótimas poderiam ser obtidas e sob quais suposições?
- Com o objetivo de separação dos grupos, neste caso, qual é o número máximo de possíveis eixos que podem ser extraídos na análise dos dados? Como escolher quantos eixos devem ser usados na redução dos dados? Justifique.

16. a) Na análise de redução de dimensionalidade de uma matriz de dados $Y_{n \times p}$, os Componentes Principais satisfazem quais propriedades? O que garante que dois componentes reduzem bem os dados?

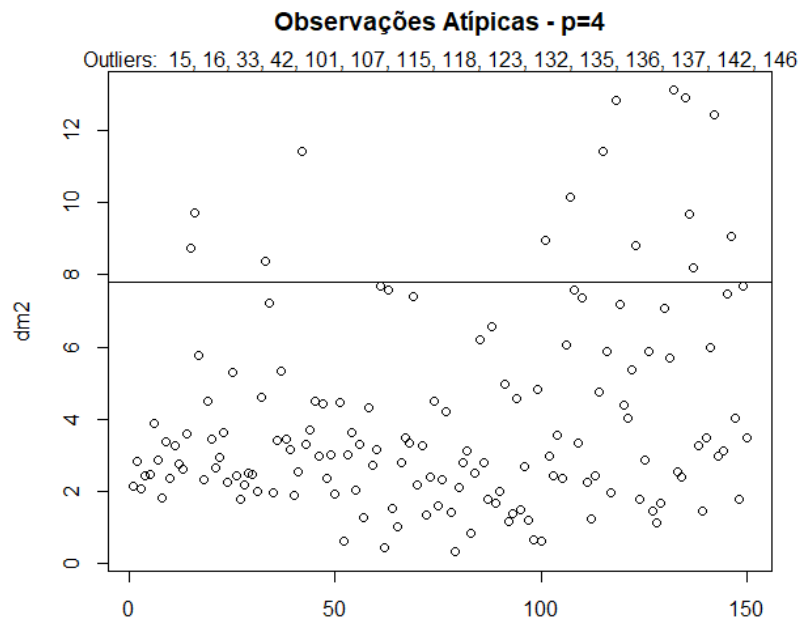
b) Na análise de redução de dimensionalidade de uma matriz de dados $Y_{n \times p}$, em que, para G grupos, $n = n_1 + n_2 + \dots + n_G$, os Eixos Discriminantes da Solução Linear de Fisher satisfazem quais propriedades? O que garante que dois eixos discriminantes reduzem bem os dados?

c) Na análise de redução de dimensionalidade de uma matriz de dados $Y_{n \times p}$, com $p = p_1 + p_2$, os Eixos Canônicos da Correlação Canônica satisfazem quais propriedades? O que garante que o primeiro par desses eixos reduzem bem os dados?

17. a) O boxplot para uma amostra avaliada em 4 variáveis é apresentado a seguir. Há indicação de observações atípicas (outliers)? Qual critério de identificação de observações atípicas é adotado?



b) O gráfico a seguir mostra a distância de Mahalanobis calculada para os mesmos dados do boxplot anterior. Há indicação de observações atípicas (outliers)? Qual critério de identificação de observações atípicas é adotado neste caso?



c) Compare os dois critérios de identificação de *outliers*.

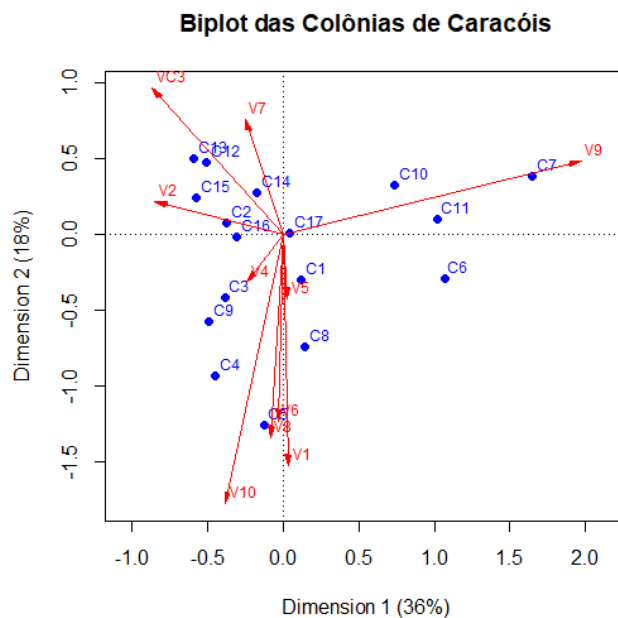
18. Em Manly (2007) é apresentada uma tabela de frequências com a distribuição de 17 colônias de caracóis (C1 a C17) de acordo com 10 formatos de conchas (V1 a V10):

	V1	V2	VC3	V4	V5	V6	V7	V8	V9	V10
C1	15	0	51	2	40	17	1	9	42	34
C2	24	57	80	1	24	3	10	2	19	19
C3	0	9	7	3	6	8	3	5	0	6
C4	0	1	16	0	0	10	0	2	0	33
C5	76	4	0	33	0	23	0	6	0	31
C6	39	1	0	4	4	1	2	4	66	0
C7	0	1	6	11	2	7	2	1	185	0
C8	0	5	9	4	6	0	0	5	23	47
C9	1	22	6	8	5	3	1	3	0	21
C10	1	4	4	0	9	4	17	0	32	0
C11	17	1	3	1	5	7	25	0	82	6
C12	25	30	135	7	4	0	46	1	11	5
C13	0	17	165	33	0	12	26	1	8	13
C14	0	0	62	28	1	11	7	10	20	0
C15	41	24	138	15	4	8	27	2	0	7
C16	7	1	59	15	18	0	59	0	11	49
C17	0	0	0	54	14	2	61	1	21	16

a) O teste Qui-Quadrado desta tabela apresentou os resultados a seguir. Que teste foi realizado? Como o grau de liberdade foi calculado? Interprete os resultados.

X-squared = 3537.3, df = 144, p-value < 2.2e-16

b) Foi realizada uma Análise de Correspondência (rowGreen) destes dados. O biplot dos resultados estão apresentados a seguir. Os dados da tabela de contingência estão bem resumidos neste biplot? Justifique. Interprete os resultados.



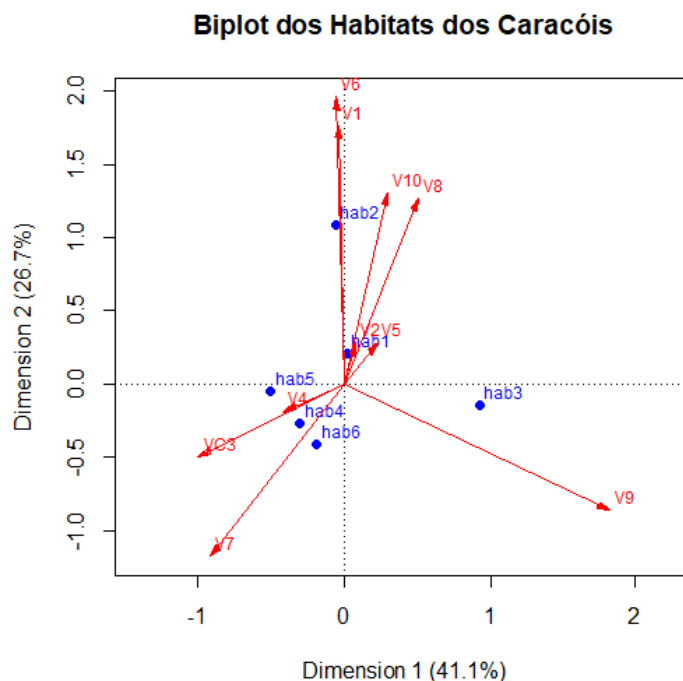
Os dados da tabela de frequências no nível de colônias apresentados anteriormente foram acumulados de acordo com 6 habitats dessas colônias (hab1 a hab6). A tabela resultante está apresentada a seguir.

	V1	V2	VC3	V4	V5	V6	V7	V8	V9	V10
hab1	39	57	131	3	64	20	11	11	61	53
hab2	76	14	23	36	6	41	3	13	0	70
hab3	41	33	25	27	26	15	22	13	306	68
hab4	42	48	365	69	10	30	104	12	121	24
hab5	48	25	197	30	22	8	86	2	11	56
hab6	0	0	0	54	14	2	61	1	21	16

c) O teste Qui-Quadrado dessa última tabela apresentou os resultados a seguir. Que teste foi realizado? Como o grau de liberdade foi calculado? Interprete os resultados.

X-squared = 1710.6, df = 45, p-value < 2.2e-16

d) Foi realizada uma Análise de Correspondência (rowGreen) destes dados. O biplot dos resultados estão apresentados a seguir. Os dados da tabela de contingência estão bem resumidos neste biplot? Justifique. Interprete os resultados.



19. Os dados a seguir correspondem a um escore da expressão gênica de 3 regiões cromossômicas (E1, E2 e E3) avaliados em trigêmeos (dois monozigotos, Mo1 e Mo2, e um dizigoto, Di) para 6 mães.

Mãe	Mo1			Mo2			Di		
	E1	E2	E3	E1	E2	E3	E1	E2	E3
1	1	6	7	3	6	7	2	5	7
2	5	3	2	4	4	3	4	4	4
3	6	1	1	7	1	1	5	2	1
4	7	1	2	2	2	2	7	2	1
5	2	5	4	2	6	6	3	5	6
6	3	4	4	1	7	5	3	5	4

Os resultados de uma Análise de Correlação Canônica (CCA) relacionando os trigêmeos (dois a dois) são mostrados a seguir:

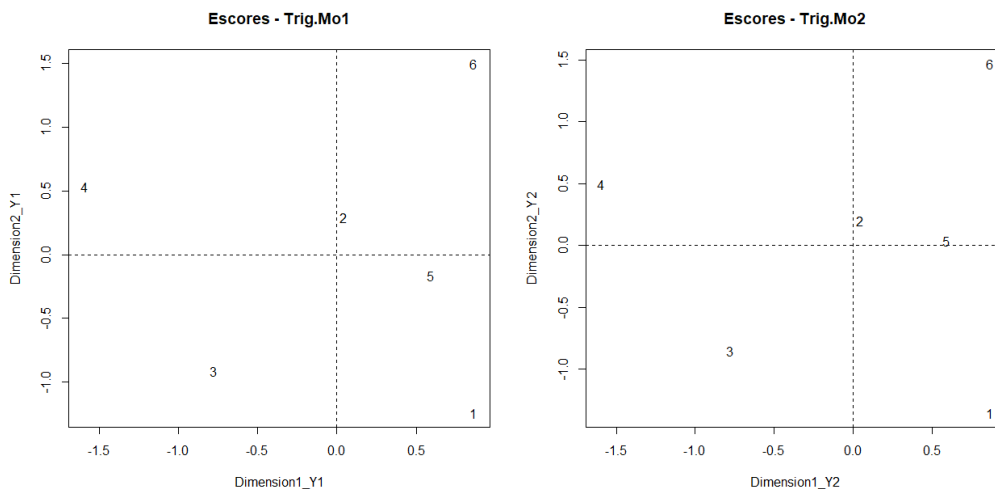
Coeficientes de correlação canônica

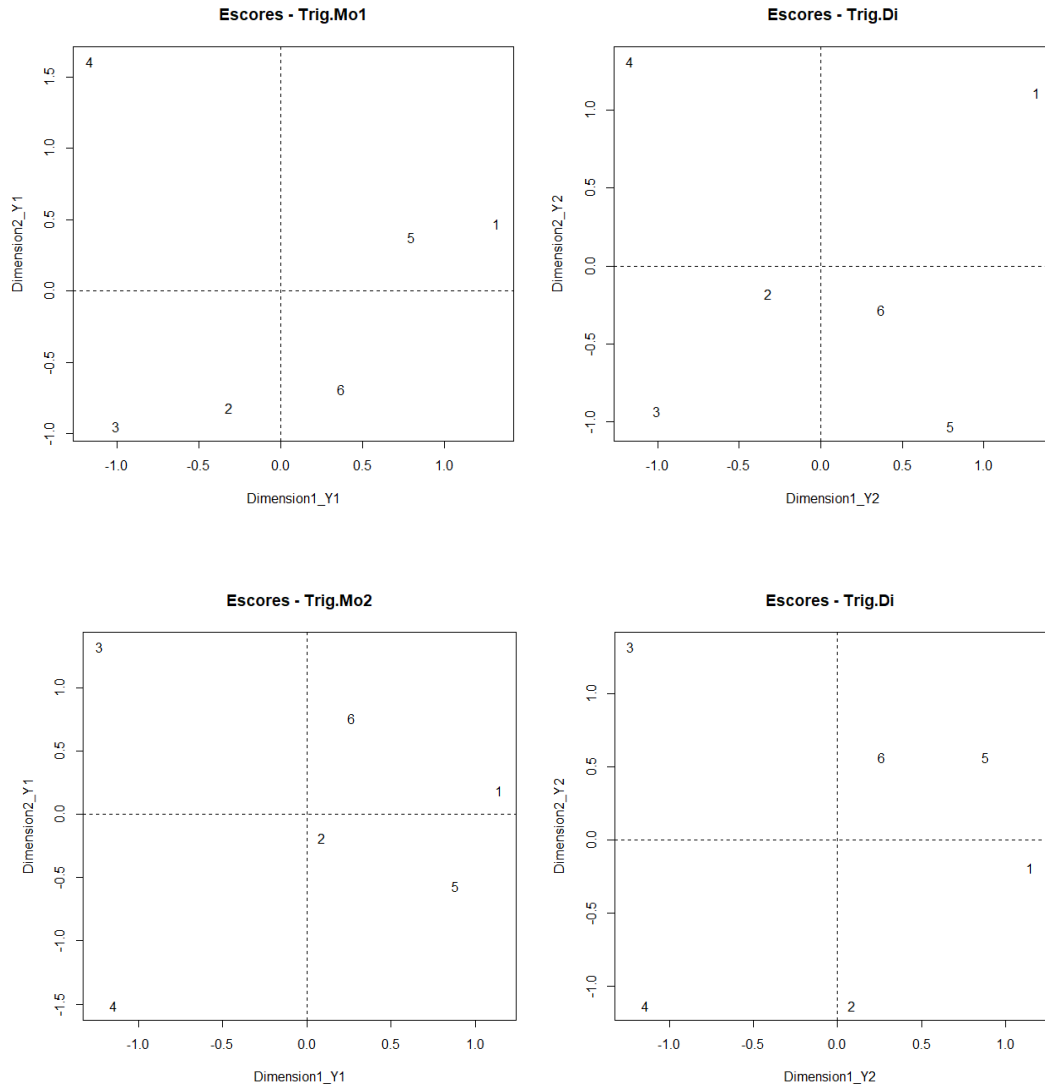
Mo1 e Mo2: 1.0000000 0.9934184 0.8437339

Mo1 e Di: 1.0000000 0.6947155 0.3173112

Mo2 e Di: 1.0000000 0.7500000 0.3223292

Representação dos 6 trigêmeos de acordo com os dois primeiros Escores obtidos da CCA relacionando Mo1 e Mo2 (primeira linha de gráficos), Mo1 e Di (segunda linha de gráficos) e Mo2 e Di (terceira linha de gráficos).





- b) Os gráficos com a distribuição dos escores para os 6 trigêmeos nos dois primeiros eixos canônicos de cada análise permitem identificar quais trigêmeos mais (e menos) contribuem para a correlação canônica?
- c) Como os eixos canônicos da Correlação Canônica estão relacionados aos eixos dos Componentes Principais da análise de cada conjunto de variáveis?

20. a) Prepare e responda uma questão livremente.
- b) O que mais aprendeu e o que menos aprendeu na disciplina?

Boa sorte 😊