

UFC/Universidade Federal do Ceará

Depto. de Estatística e Matemática Aplicada/CC - 2º AP - 2022/II

Disciplina: Análise Multivariada - Curso: Estatística/DEMA -

Profª: Sílvia Freitas - Aluna: Yasmin Miranda - Matrícula: 472393

01) Pesquisadores interessados em avaliar a função pulmonar em populações não patológicas pediram que um grupo de voluntários corresse em uma esteira até a exaustão. Amostras de ar foram coletadas em intervalos definidos e os conteúdos de gás analisados. Os resultados, em 4 medidas de consumo de oxigênio, para 25 homens e 25 mulheres são dados na Tabela abaixo

X1 = resting volume O₂ (L/min)

X2 = resting volume O₂ (mL/kg/min)

X3 = maximum volume O₂ (L/min)

X4 = maximum volume O₂ (mL/kg/min)

a) Encontre o vetor de médias amostral e a matriz de correlação amostral considerando toda a amostra e depois por cada grupo de sexo.

Vetor de médias:

	Média Global	Média Feminina	Média Masculina
X1	0.3554	0.3136	0.3972
X2	5.2542	5.1788	5.3296
X3	3.0014	2.3152	3.6876
X4	43.7876	38.1548	49.4204

Correlação Amostral Global:

	X1	X2	X3	X4
X1	1.0000	0.8121	0.5338	0.3897
X2	0.8121	1.0000	0.0872	0.2346
X3	0.5338	0.0872	1.0000	0.8302
X4	0.3897	0.2346	0.8302	1.0000

Correlação Amostral Feminina:

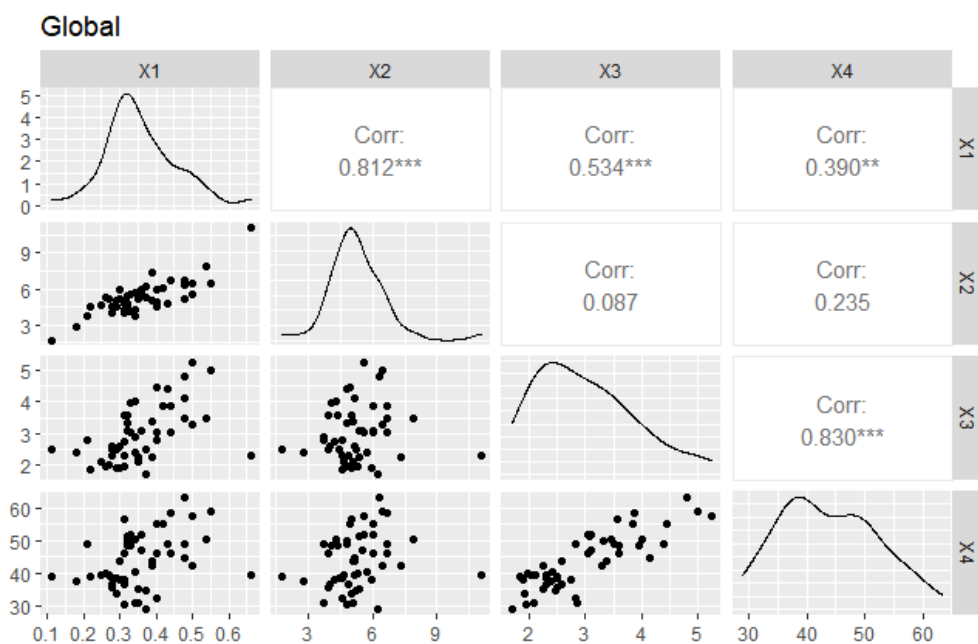
	X1	X2	X3	X4
X1	1.0000	0.9367	0.1217	0.0625
X2	0.9367	1.0000	-0.0681	0.1592
X3	0.1217	-0.0681	1.0000	0.6559

X4 0.0625 0.1592 0.6559 1.0000

Correlação Amostral Masculina:

	X1	X2	X3	X4
X1	1.0000	0.7755	0.5519	0.2401
X2	0.7755	1.0000	0.2045	0.4315
X3	0.5519	0.2045	1.0000	0.6591
X4	0.2401	0.4315	0.6591	1.0000

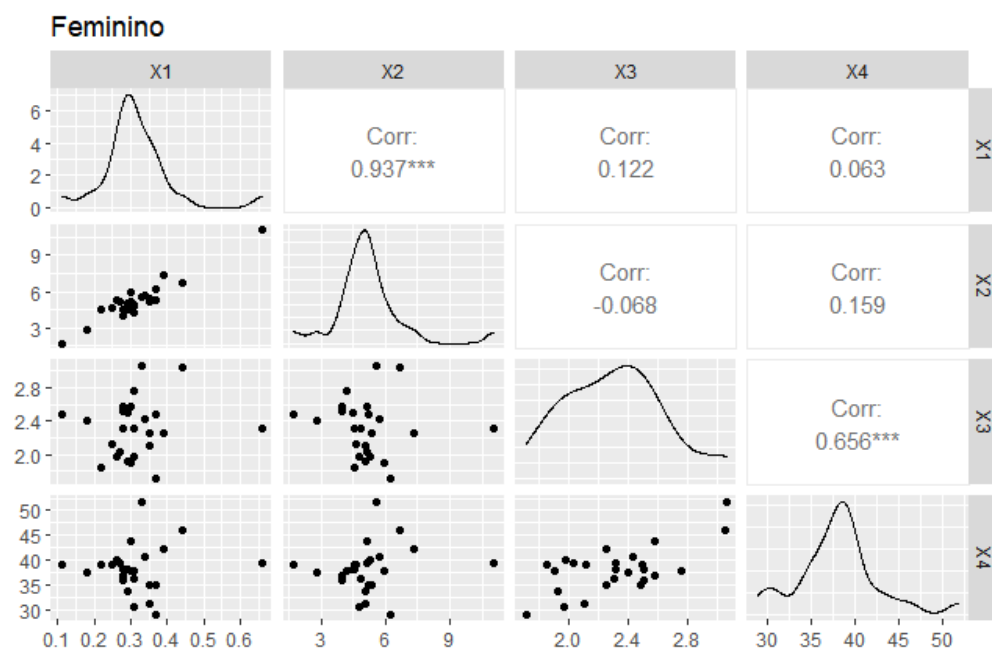
b)Faça uma representação gráfica dos dados globalmente e por sexo.



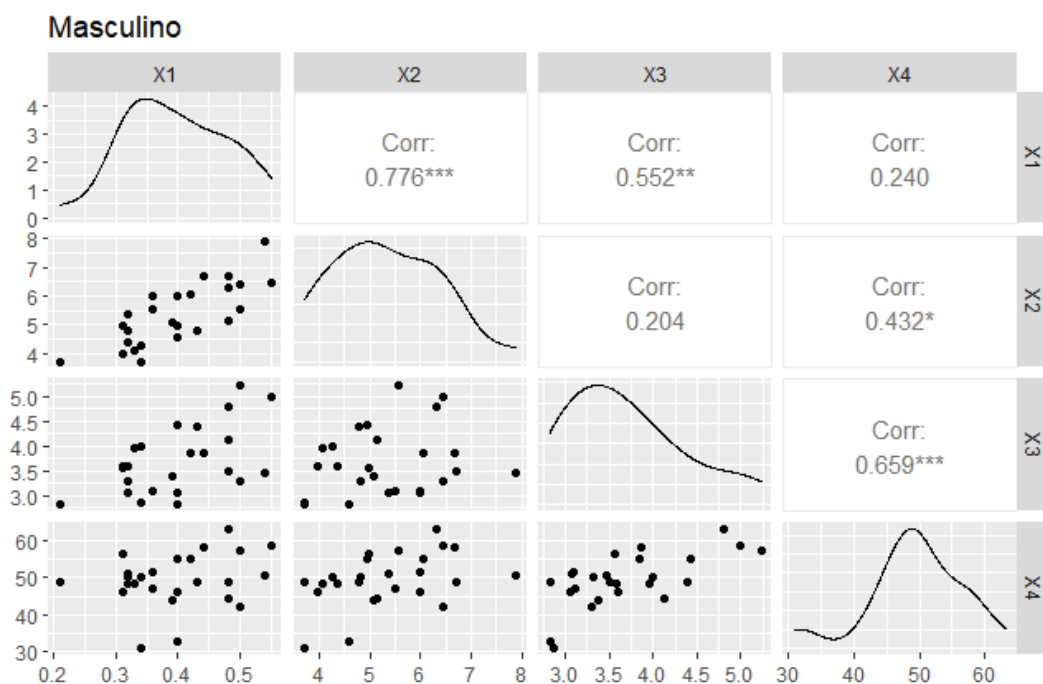
Notamos que a variável X1 tem forte correlação com a X2 e razoável correlação com a X3, enquanto com a X4 a correlação é relativamente baixa.

A variável X2 possui baixa correlação com as demais variáveis. Ela e X1 indicam, pelo gráfico de densidade da diagonal, seguirem distribuição normal.

A variável X3 mostra alta correlação com a X4, sendo as mais correlacionadas globalmente. Ao contrário das duas variáveis citadas no parágrafo anterior, seus gráficos de densidade indicam que elas são assimétricas.



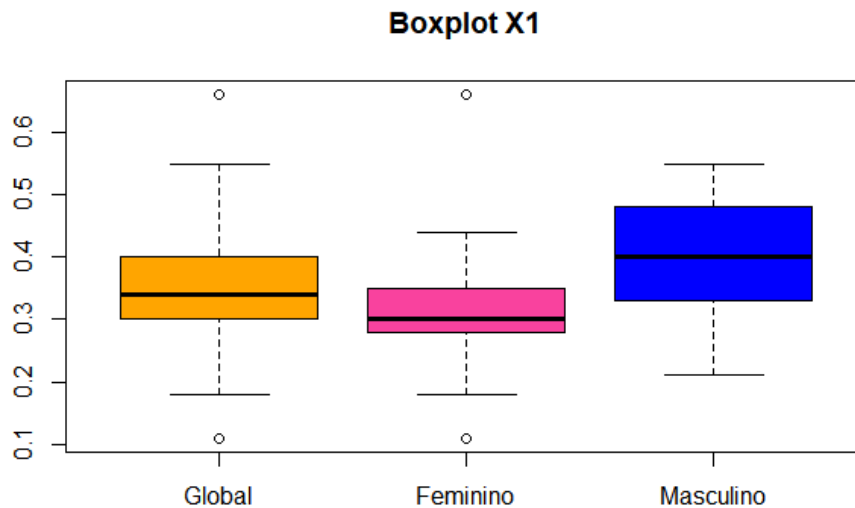
Ao analisar as variáveis pelo sexo dos indivíduos, notamos mudanças em relação aos padrões globais. Aqui, as variáveis mais correlacionadas são X1 e X2, seguidas por X3 e X4. Também notamos uma correlação negativa fraca entre X3 e X2. Quanto aos gráficos de densidade, percebemos uma grande diferença nos dois últimos: X3 e X4. Anteriormente, a maior parte dos valores de X3 estavam mais concentrados à esquerda, agora, estão mais dispersos. X4 por sua vez possui curva similar à normal, com densidade crescente até um ponto de máximo e depois decrescente.



Os dados de indivíduos do sexo masculino mostram que, nesse grupo, houve uma certa discrepância entre as observações, resultando em gráficos de densidade com curvas assimétricas para as 4 variáveis. A correlação entre as mesmas é mais parecida com as

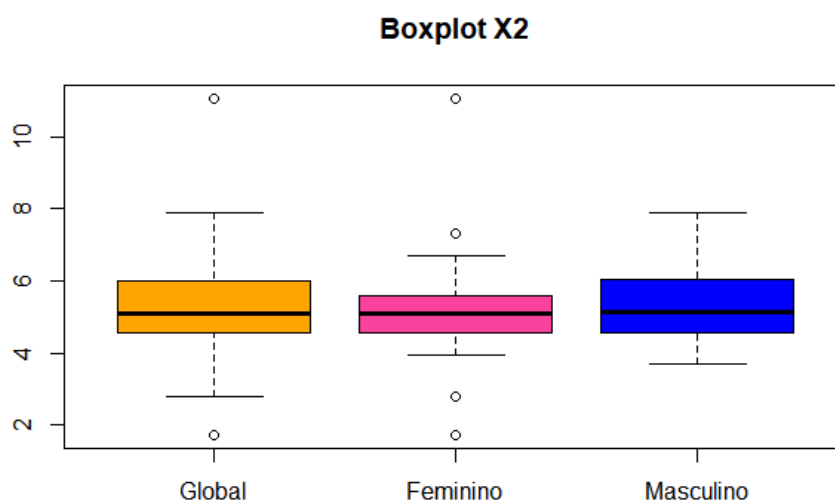
correlações globais, sendo todas positivas. Aqui, a variável X1 e X2 também são as com maior correlação, mas seu valor é de 0.77.

Quanto ao comportamento de cada variável, seguem os boxplots:



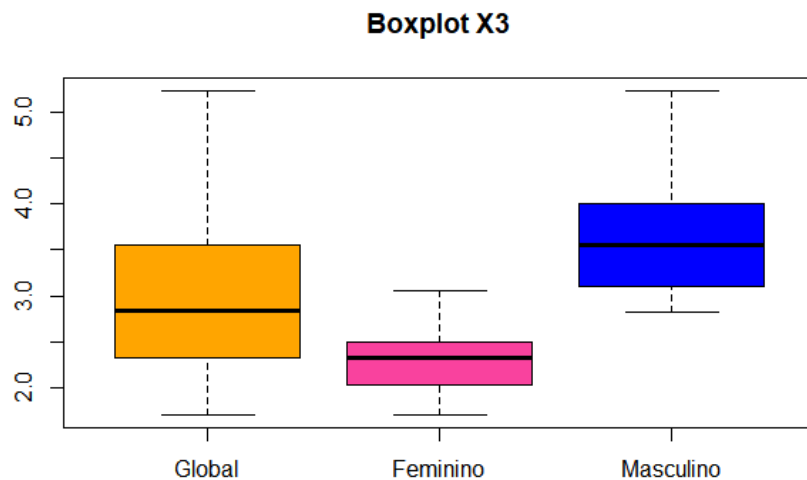
Notamos que os valores masculinos para essa variável são, em sua maioria, maiores que os femininos e mais homogêneos. Existem 2 mulheres que tiveram medidas atípicas, sendo bem maior do que o comum e bem menor. O grupo de mulheres possui uma dispersão maior dessas observações, algo que podemos observar pela distância da linha dentro da caixa rosa (segundo quartil) em relação às fronteiras da caixa (primeiro e terceiro quartil).

Desse modo, vemos que o boxplot global condensa esses dois comportamentos e mantém as duas mulheres como outliers.

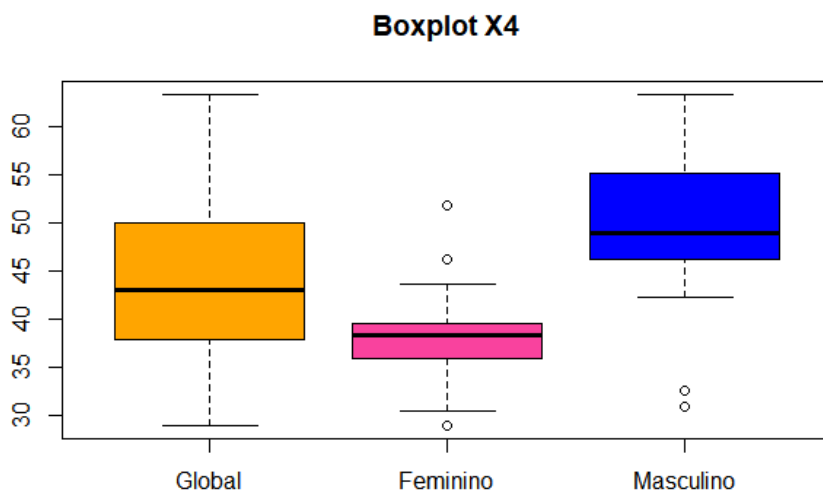


A variável X2 tem comportamento similar em ambos os grupos, com mudança sutil dos valores mínimos e mais acentuada do valor máximo. A mediana é aproximadamente igual, mas no grupo feminino existem 4 observações atípicas, resultando em 2 outliers no

boxplot global. Notamos que no grupo feminino a variável parece ser simétrica, enquanto no masculino ela parece ter assimetria mais à direita.



Tanto na variável X3 como na X4, o grupo masculino teve medidas maiores que o feminino. Na X3, porém, não foram computadas observações discrepante, enquanto na X4 ambos computaram, sendo valores abaixo do comum no grupo masculino e no feminino valores acima e abaixo do comum. A variável X3 aparenta ser simétrica no grupo masculino, e a X4 aparenta ser assimétrica nesse grupo. A variável X3 também aparenta ser simétrica no grupo feminino, ao contrário da X4. De maneira geral, o boxplot global é mais simétrico em X4 do que em X3.



c) Teste a normalidade univariada e multivariada dos dados, globalmente e por cada grupo de sexo.

Globalmente:

Aplicando o teste Anderson-Darling (AD), temos que X1 e X4 seguem distribuição normal, com nível descritivo igual a 0.0752 e 0.2733, respectivamente. Entretanto, X2 e X3 não seguem, com nível descritivo igual a 0.0098 e 0.0472, respectivamente.

A base de dados não segue distribuição normal multivariada, pelo teste de Royston, com nível descritivo igual a 0.00025.

Feminino:

Aplicando o teste AD, temos que X3 e X4 seguem distribuição normal, com nível descritivo igual a 0.5620 e 0.1194, respectivamente. Entretanto, X1 e X2 não seguem, com nível descritivo igual a 0.0042 e 0.0022, respectivamente.

A base de dados não segue distribuição normal multivariada, pelo teste de Royston, com nível descritivo igual a 8.686009e-05.

Masculino:

Aplicando o teste Anderson-Darling (AD), temos que X1, X2, X3 e X4 seguem distribuição normal, com nível descritivo igual a 0.03249, 0.7490, 0.1599 e 0.1326, respectivamente.

A base de dados segue distribuição normal multivariada, pelo teste de Royston, com nível descritivo igual a 0.1749.

d) Utilizando os dados dessa amostra, teste a hipótese de que o vetor de médias, em cada grupo, pode ser considerado como sendo $\mu = (0,40 ; 4,50 ; 3,52 ; 51,15)^T$, a um nível de 5% de significância, "supondo que os dados tenham normalidade multivariada".

Feminino:

Não, pois a média populacional de X1, X3 e X4 diferem da média populacional sob a hipótese nula.

Masculino:

Não, pois a média populacional de X2 difere da média populacional sob a hipótese nula

e) Construa os intervalos simultâneos de 95% de confiança para as médias individuais, em cada grupo.

Feminino:

IC(X1, 95%): [0.2425063 ; 0.3846937]

IC(X2, 95%): [3.9770980 ; 6.3805020]

IC(X3, 95%): [2.0650312 ; 2.5653688]

IC(X4, 95%): [34.679155; 41.6304444]

Masculino:

IC(X1, 95%): [0.3363874 ; 0.4580126]

IC(X2, 95%): [4.5587502 ; 6.1004498]

IC(X3, 95%): [3.2010284 ; 4.1741716]

IC(X4, 95%): [44.0636998 ; 54.7771002]

f) Construa os intervalos de bonferroni com 95% de confiança para as médias individuais, em cada grupo.

IC(X1, 95%): [0.03139572; 0.1358043]

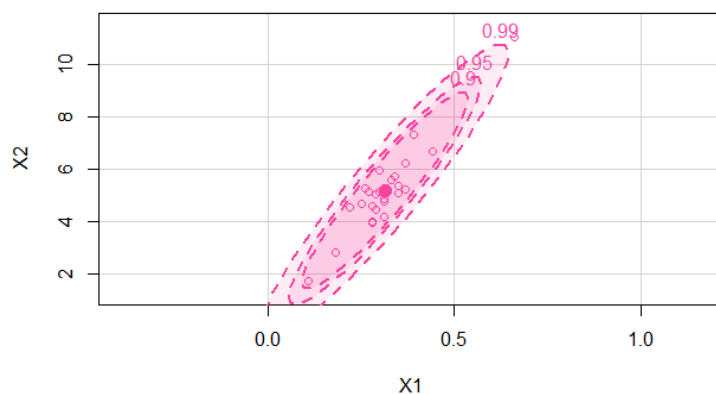
IC(X2, 95%): [-0.6458621; 0.9474621]

IC(X3, 95%): [1.067105 ; 1.677695]

IC(X4, 95%): [7.702452 ; 14.82875]

g) Considerando cada grupo, construa as elipses de confiança para as duas variáveis mais correlacionadas. Use 95% de confiança

Regiões de 95% Confiança Base Feminina: X1 e X2



Regiões de 95% Confiança Base Masculina

