Written examination, date: 6th of December 2021 *Page 1 of 34 pages* *Enclosure: XX pages*
Course name: Multivariate Statistics
Course number: 02409
Aids allowed: All
Exam duration: 4 hours
Weighting: The questions are given equal weight

This exam is answered by:

_____ _____ _____
(name) (signature) (study no.)

There is a total of 30 questions for the 6 problems. The answers to the 30 questions must be written into the table below.

Problem	1	1	1	1	2	2	2	2	2	2
Question	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	2.5	2.6
Answer	3	5	3	4	3	4	5	4	3	2

Problem	3	3	3	3	3	4	4	4	4	4
Question	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4	4.5
Answer	3	4	4	1	5	2	4	5	4	2

Problem	5	5	5	6	6	6	6	6	6	6
Question	5.1	5.2	5.3	6.1	6.2	6.3	6.4	6.5	6.6	6.7
Answer	5	2	5	3	3	5	1	4	1	3

The possible answers for each question are numbered from 1 to 6. If you enter a wrong number, you may correct it by crossing the wrong number in the table and writing the correct answer immediately below. If there is any doubt about the meaning of a correction then the question will be considered not answered.

Only the front page must be returned. The front page must be returned even if you do not answer any of the questions or if you leave the exam prematurely. Drafts and/or comments are not considered, only the numbers entered above are registered.

A correct answer gives 5 points, a wrong answer gives – 1 point. Unanswered questions or a 6 (corresponding to "don't know") give 0 points. The total number of points needed for a satisfactorily answered exam is determined at the final evaluation of the exam. Especially note that the grade 10 may be given even if only one answer is wrong or unanswered.
Remember to write your name, signature, and study number on the front page.

Problem 1.

Enclosure A with SAS program and SAS output belongs to this problem.

We consider the relation between education and voting pattern. For each of the 98 Danish municipalities, we have data for the highest level of education each person and how many votes each party got, at the municipial election in 2017. The data is from Danmark Statistik.

The following variables are included in our analysis. It is *not* important to understand the meaning of the variables in details.

Educational variables	Meaning
H10	9 years mandatory school
H20	High School
H30	Vocational School
H35	Admittance education for universities
H40	Short higher education
H50	Medium higher education
H60	Bachelor
H70	Long higher education
H80	PhD
H90	Unknown

Political variables	Meaning
A	Socialdemokraterne
B	Radikale Venstre
C	Konservative
D	Nye Borgerlige
F	Socialistisk Folkeparti
I	Liberal Alliance
O	Dansk Folkeparti
V	Venstre
Ø	Enhedslisten
Å	Alternativet

We perform a canonical correlation analysis.

Question 1.1.

The 3rd set of canonical variates explain the following fraction of variation between the two datasets

We find in the output

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation	Eig
1	0.999133	0.998980	0.000176	0.998267	5
2	0.986890	0.984828	0.002645	0.973952	
3	0.857140	0.832918	0.026938	0.734690	
4	0.734708	0.696747	0.046727	0.539796	
5	0.559780	0.497739	0.069718	0.313354	
6	0.376799	0.264090	0.087119	0.141977	
7	0.299828	0.255518	0.092407	0.089897	
8	0.141449	.	0.099503	0.020008	
9	0.087767	.	0.100752	0.007703	
10	0.020094	.	0.101494	0.000404	

Refer to page 29

and the *squared coefficient of correlation represents the reduction in variance. i.e. the fraction of Y's variance, which can be explained by X,* since

$$\rho^2 = \frac{\mathrm{V}(Y) - \mathrm{V}(Y|X = x)}{\mathrm{V}(Y)}.$$

ANSWER 3

Question 1.2.

How many of the eigenvalues of $\boldsymbol{E}^{-1}\boldsymbol{H}$ are significantly different from zero at the 5%-level

We refer to

Theorem 6.16

Testing whether the canonical correlations are zero is equivalent to test whether the eigenvalues of $\boldsymbol{E}^{-1}\boldsymbol{H}$ are zero.

On page 389.

We find in the output:

Test of H0: The canonical correlations in the current row and all that follow are zero				
Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
0.00000287	28.19	100	570.32	<.0001
0.00165765	10.86	81	519.51	<.0001
0.06363783	4.48	64	467.92	<.0001
0.23986182	2.75	49	415.65	<.0001
0.52120804	1.61	36	362.85	0.0167
0.75906374	0.95	25	309.83	0.5290
0.88466634	0.66	16	257.26	0.8336
0.97205047	0.27	9	207.02	0.9822
0.99189626	0.18	4	172	0.9509
0.99959623	0.04	1	87	0.8517

And see that five have a p-value less than 0.05.

ANSWER 5

Question 1.3.

What fraction of variance in the WITH variable 'C' is explained by V1 and V2 together

Since the individual canonical variates are uncorrelated, we can just add the variance explained
We find in the output

Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables										
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
A	0.9185	0.2896	0.0040	-0.1143	-0.0939	-0.0137	-0.0259	-0.0001	0.0058	0.0011
B	0.9573	-0.1281	0.1672	-0.0195	0.0493	-0.0139	0.0188	-0.0096	-0.0080	0.0004
C	0.4969	-0.0969	0.3973	0.4616	-0.0245	-0.0364	0.0282	0.0177	0.0270	-0.0011
D	0.8279	0.0525	0.0962	0.0398	-0.1482	0.1231	0.0473	0.0241	-0.0209	0.0018
F	0.9475	-0.0275	0.1041	-0.1046	0.0439	0.0344	0.0253	-0.0002	-0.0035	-0.0041
I	0.9517	-0.0235	0.1281	-0.0109	0.0404	0.0363	-0.0414	-0.0213	-0.0101	-0.0001
O	0.8305	0.3237	-0.2137	0.0091	-0.0590	0.0322	0.0573	-0.0389	-0.0078	-0.0005
V	0.6434	0.5680	-0.1957	0.0360	0.1913	0.0124	-0.0102	0.0306	-0.0147	0.0015
Ø	0.9406	-0.3236	-0.0462	-0.0078	-0.0158	-0.0095	-0.0059	-0.0035	-0.0013	-0.0004
Å	0.9359	-0.3333	-0.0079	-0.0400	0.0169	0.0186	0.0080	-0.0047	0.0014	-0.0000

$$0.4969^2 + (-0.0969)^2 = 0.2563$$

Answer 3

Question 1.4.

The interpretation - when using the appropriate correlations - of V4 vs W4 is broadly

We use the correlations

Correlations Between the VAR Variables and Their Canonical Variables										
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
H10	0.9679	0.1833	-0.1527	0.0253	-0.0443	0.0385	0.0429	-0.0134	0.0037	-0.0108
H20	0.9878	-0.0419	0.1045	-0.0777	-0.0402	-0.0023	-0.0329	-0.0299	0.0011	-0.0441
H30	0.9208	0.3283	-0.2035	0.0236	-0.0193	0.0261	-0.0119	0.0292	0.0106	0.0158
H35	0.7340	0.4685	0.3317	-0.3175	-0.0524	0.0704	-0.0976	-0.1169	0.0074	0.0077
H40	0.9944	0.0610	0.0061	0.0012	0.0323	0.0282	-0.0136	0.0418	-0.0265	-0.0545
H50	0.9978	0.0338	0.0395	-0.0097	0.0050	0.0086	0.0114	0.0065	-0.0339	0.0153
H60	0.9636	-0.2392	0.0668	-0.0685	-0.0099	-0.0094	-0.0396	-0.0380	0.0318	-0.0298
H70	0.9535	-0.2732	0.1138	0.0313	0.0174	0.0157	-0.0215	0.0050	0.0348	-0.0105
H80	0.9430	-0.2385	0.2190	0.0087	-0.0007	-0.0020	0.0092	0.0322	0.0667	-0.0134
H90	0.9727	-0.1812	-0.1076	0.0279	-0.0628	0.0542	0.0315	0.0061	0.0037	-0.0266

Correlations Between the WITH Variables and Their Canonical Variables										
	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10
A	0.9193	0.2934	0.0046	-0.1556	-0.1677	-0.0365	-0.0863	-0.0009	0.0661	0.0569
B	0.9581	-0.1298	0.1951	-0.0265	0.0880	-0.0369	0.0628	-0.0679	-0.0912	0.0204
C	0.4974	-0.0982	0.4635	0.6283	-0.0438	-0.0967	0.0942	0.1252	0.3075	-0.0553
D	0.8286	0.0532	0.1122	0.0541	-0.2648	0.3268	0.1577	0.1702	-0.2378	0.0880
F	0.9483	-0.0278	0.1215	-0.1424	0.0784	0.0914	0.0845	-0.0016	-0.0403	-0.2042
I	0.9525	-0.0238	0.1494	-0.0149	0.0722	0.0962	-0.1382	-0.1505	-0.1153	-0.0062
O	0.8312	0.3280	-0.2493	0.0124	-0.1055	0.0855	0.1910	-0.2752	-0.0890	-0.0251
V	0.6439	0.5756	-0.2283	0.0490	0.3417	0.0330	-0.0342	0.2162	-0.1677	0.0751
Ø	0.9415	-0.3279	-0.0539	-0.0106	-0.0283	-0.0253	-0.0196	-0.0250	-0.0149	-0.0218
Å	0.9367	-0.3377	-0.0092	-0.0544	0.0303	0.0493	0.0267	-0.0335	0.0164	-0.0012

V4 is mainly negatively correlated with admittance education (H35), and W4 is a contrast between C vs. (A, F)

ANSWER 4

Problem 2.

We are given the following data – and we use SAS to analyse them

Obs H10 H20 H30 H35 H40 H50 H60 H70 H80 H90 A
1 9924 1866 12385 11 1752 4083 186 1022 50 330 7378
2 17360 6147 20447 96 3655 9997 545 3340 167 701 20321
3 8674 1870 10783 26 1041 2882 139 687 38 309 8649
4 3891 1172 5379 21 797 2456 177 1061 81 160 3432
5 19682 6115 24697 63 3927 8898 616 3059 203 709 20396
6 600 146 961 2 100 360 37 120 4 42 1037
7 16593 4904 21397 62 3657 10467 673 4276 261 556 15612
8 10041 3015 12891 38 2536 6973 540 4332 360 330 12793
9 7744 2123 10371 26 1354 4276 325 1773 136 289 6274
10 46415 45525 49222 774 12692 41916 13028 37456 4160 2377 71595

11 17754 5572 21482 79 3597 8841 664 2588 139 612 8489
12 11440 3453 14022 56 1918 6517 255 1712 55 358 14574
13 9036 2016 10543 15 1441 3325 184 1149 91 343 5178
14 4438 865 5352 9 484 1793 59 445 34 161 2640
15 11943 2367 15435 22 1654 4949 193 1392 51 424 6945
16 10059 2327 11942 16 1464 4286 132 979 43 365 8421
17 4867 895 5167 21 559 1970 88 484 26 159 4276

To get the data into SAS, we use the following code:

```
data problem2;
input Obs H10 H20 H30 H35 H40 H50 H60 H70 H80 H90 A;
datalines;
1 9924 1866 12385 11 1752 4083 186 1022 50 330 7378
2 17360 6147 20447 96 3655 9997 545 3340 167 701 20321
3 8674 1870 10783 26 1041 2882 139 687 38 309 8649
4 3891 1172 5379 21 797 2456 177 1061 81 160 3432
5 19682 6115 24697 63 3927 8898 616 3059 203 709 20396
6 600 146 961 2 100 360 37 120 4 42 1037
7 16593 4904 21397 62 3657 10467 673 4276 261 556 15612
8 10041 3015 12891 38 2536 6973 540 4332 360 330 12793
9 7744 2123 10371 26 1354 4276 325 1773 136 289 6274
10 46415 45525 49222 774 12692 41916 13028 37456 4160 2377 71595
11 17754 5572 21482 79 3597 8841 664 2588 139 612 8489
12 11440 3453 14022 56 1918 6517 255 1712 55 358 14574
13 9036 2016 10543 15 1441 3325 184 1149 91 343 5178
14 4438 865 5352 9 484 1793 59 445 34 161 2640
15 11943 2367 15435 22 1654 4949 193 1392 51 424 6945
16 10059 2327 11942 16 1464 4286 132 979 43 365 8421
17 4867 895 5167 21 559 1970 88 484 26 159 4276
;
```

Question 2.1.

Question 2.1

We now perform a linear regression with A as the dependent variable and including all educational levels dependent variables as well as an intercept, i.e., a model of the form

$$A = \mu + \beta_1 H10 + \beta_2 H20 + \beta_3 H30 + \beta_4 H35 + \beta_5 H40 + \beta_6 H50 + \beta_7 H60 + \beta_8 H70 + \beta_9 H80 + \beta_{10} H90 + \epsilon$$

Where μ is the intercept and ϵ is the error term.

The variable with the largest tolerance in this model is:

We use this code

```
proc reg data=problem2;
model A = H10 H20 H30 H35 H40 H50 H60 H70 H80 H90 / r tol;
run;
```

and find in the output

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance
Intercept	1	2062.29879	2854.29868	0.72	0.4972	.
H10	1	0.80477	2.61286	0.31	0.7685	0.00062204
H20	1	9.84557	4.56448	2.16	0.0744	0.00019291
H30	1	0.17346	1.64540	0.11	0.9195	0.00136
H35	1	-78.14837	143.78732	-0.54	0.6064	0.00065624
H40	1	-5.53851	3.75468	-1.48	0.1906	0.00379
H50	1	-2.52655	4.33675	-0.58	0.5814	0.00026661
H60	1	-26.14736	11.02361	-2.37	0.0554	0.00038268
H70	1	9.80680	12.44719	0.79	0.4608	0.00003753
H80	1	-34.95016	102.97256	-0.34	0.7459	0.00004310
H90	1	-29.12479	32.69846	-0.89	0.4074	0.00152

ANSWER 3, H40

Question 2.2.

We still consider the data from problem 2.1

When only including four dependent variables, we attain the highest R^2 by including

We can find that by using Rsquare selection

```
proc reg data=problem2;
model A = H10 H20 H30 H35 H40 H50 H60 H70 H80 H90 / selection=Rsquare;
run;
```

3	0.9648	H35 H60 H80
3	0.9642	H20 H35 H80
3	0.9638	H10 H30 H90
3	0.9596	H20 H35 H70
4	0.9868	H20 H40 H60 H70
4	0.9865	H20 H40 H60 H80
4	0.9849	H20 H60 H80 H90
4	0.9849	H20 H50 H60 H70
4	0.9845	H10 H20 H60 H80
4	0.9843	H20 H30 H60 H80

ANSWER 4, H20 H40 H60 H70

Question 2.3.

We still consider the data from Problem 2

We now perform a linear regression with A as the dependent variable and H20 H60 H80 as the independent variables with an intercept, i.e., a model of the form:

$A = \mu + \gamma_1 H20 + \gamma_2 H60 + \gamma_3 H80 + \epsilon$

Where μ is the intercept and ϵ is the error term.

The observation with the *lowest* leverage is:

We specify the new model

```
proc reg data=problem2;
model A = H20 H60 H80 /influence;
run;
```

And add `influence` to get influence diagnostics

Obs	Residual	RStudent	Hat Diag H
1	1586	0.7089	0.0944
2	951.7722	0.4893	0.3299
3	2376	1.0833	0.0829
4	-856.1903	-0.3834	0.1235
5	1286	0.6434	0.2822
6	582.6095	0.2778	0.2316
7	310.5249	0.1393	0.1363
8	-831.3296	-0.7329	0.7666
9	-721.0576	-0.3160	0.0886
10	152.7914	2.1738	0.9988
11	-5658	-4.5898	0.2729
12	3811	1.9297	0.1126
13	-2521	-1.1511	0.0755
14	-1112	-0.4984	0.1184
15	-663.7988	-0.2890	0.0775
16	120.5160	0.0520	0.0675
17	1185	0.5392	0.1408

ANSWER 5, obs 16

Question 2.4.

We still consider the data from Problem 2

We still perform a linear regression with A as the dependent variable and H20 H60 H80 as the independent variables with an intercept, i.e., a model of the form:
$A = \mu + \gamma_1 H20 + \gamma_2 H60 + \gamma_3 H80 + \epsilon$
Where μ is the intercept and ϵ is the error term.

The observation with largest impact on the parameter estimate for H80 is:

We need the DFBETAS and can reuse the code from before

```
proc reg data=problem2;
model A = H20 H60 H80 /influence;
run;
```

We find in the output

Output Statistics									
						DFBETAS			
Obs	Residual	RStudent	Hat Diag H	Cov Ratio	DFFITS	Intercept	H20	H60	H80
1	1586	0.7089	0.0944	1.2907	0.2288	0.1991	-0.1047	0.1225	-0.1022
2	951.7722	0.4893	0.3299	1.8993	0.3433	-0.1669	0.3053	-0.1381	0.0259
3	2376	1.0833	0.0829	1.0341	0.3257	0.2653	-0.1159	0.1473	-0.1287
4	-856.1903	-0.3834	0.1235	1.4967	-0.1439	-0.1339	0.0973	-0.0378	0.0027
5	1286	0.6434	0.2822	1.6755	0.4034	-0.1936	0.3573	-0.1960	0.0717
6	582.6095	0.2778	0.2316	1.7472	0.1525	0.1521	-0.1254	0.0982	-0.0623
7	310.5249	0.1393	0.1363	1.5845	0.0554	-0.0115	0.0337	-0.0396	0.0321
8	-831.3296	-0.7329	0.7666	4.9535	-1.3283	0.0883	-0.1528	1.1106	-1.2621
9	-721.0576	-0.3160	0.0886	1.4620	-0.0985	-0.0771	0.0438	-0.0001	-0.0188
10	152.7914	2.1738	0.9988	305.3033	62.8308	0.0317	-3.6572	5.4093	-1.7067
11	-5658	-4.5898	0.2729	0.0329	-2.8119	0.0009	-1.1871	-1.0428	1.8018
12	3811	1.9297	0.1126	0.5266	0.6873	0.1181	0.2535	0.1490	-0.3011
13	-2521	-1.1511	0.0755	0.9799	-0.3288	-0.2041	0.0652	0.0551	-0.0915
14	-1112	-0.4984	0.1184	1.4394	-0.1826	-0.1743	0.1222	-0.0656	0.0255
15	-663.7988	-0.2890	0.0775	1.4522	-0.0838	-0.0559	0.0124	-0.0332	0.0353
16	120.5160	0.0520	0.0675	1.4758	0.0140	0.0075	0.0004	0.0020	-0.0028
17	1185	0.5392	0.1408	1.4568	0.2183	0.2144	-0.1551	0.1229	-0.0795

ANSWER 4, obs 11

Question 2.5.

We still consider the data from Problem 2

We still perform a linear regression with A as the dependent variable and H20 H60 H80 as the independent variables with an intercept, i.e., a model of the form:

$A = \mu + \gamma_1 H20 + \gamma_2 H60 + \gamma_3 H80 + \epsilon$

Where μ is the intercept and ϵ is the error term.

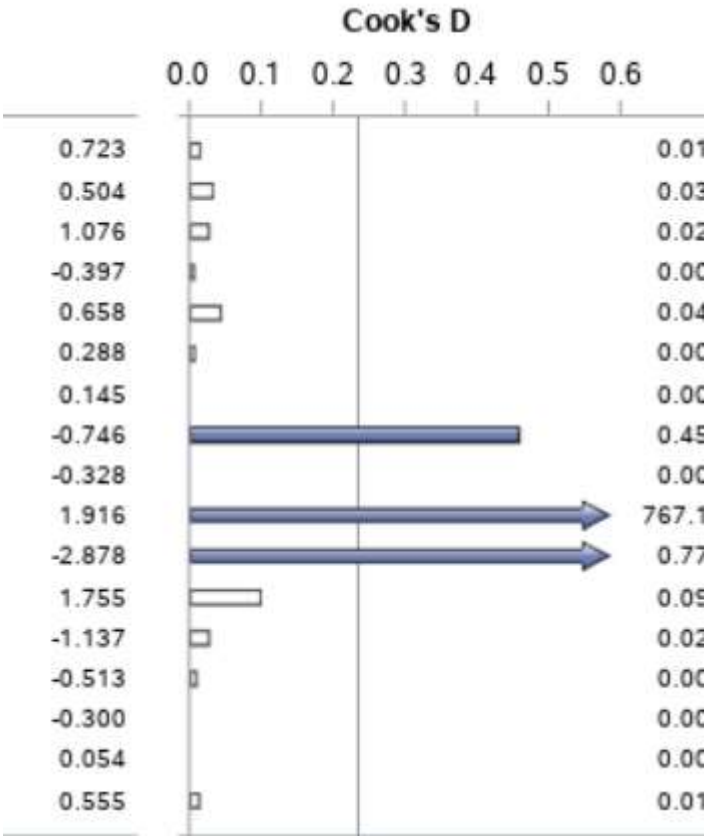
The number of observations with a Cook's D larger than 0.235 is:

.We specify residual diagnostics in addition to influence

```
proc reg data=problem2;
model A = H20 H60 H80 /r influence;
run;
```

We can then either find it in the figure or the table

in the table



Cook's D
0.014
0.031
0.026
0.006
0.043
0.006
0.001
0.457
0.003
767.109
0.777
0.098
0.026
0.009
0.002
0.000
0.013

ANSWER 3, 3

Question 2.6.

We still consider the data from problem 2.1

We still perform a linear regression with A as the dependent variable and H20 H60 H80 as the independent variables with an intercept, i.e., a model of the form:

$A = \mu + \gamma_1 H20 + \gamma_2 H60 + \gamma_3 H80 + \epsilon$

Where μ is the intercept and ϵ is the error term.

The two parameter estimates with the numerically largest covariance are:

We use

```
proc reg data=problem2;
model A = H20 H60 H80 /covb;
run;
```

Covariance of Estimates				
Variable	Intercept	H20	H60	H80
Intercept	1180056.9412	-438.2618633	2773.3037227	-4172.481194
H20	-438.2618633	0.2372283395	-1.25712203	1.4305039757
H60	2773.3037227	-1.25712203	18.916131442	-45.94214095
H80	-4172.481194	1.4305039757	-45.94214095	129.05574313

ANSWER 2, Intercept and H80

Problem 3.

Enclosure B with SAS program and SAS output belongs to this problem.

We now look at relation between different glass types and the mineral content of the glass and the refractive index. The data is from https://archive.ics.uci.edu/ml/datasets/glass+identification

and was collected by *B. German, Central Research Establishment, Home Office Forensic Science Service, Aldermaston, Reading, Berkshire RG7 4PN*

We consider the following types and mineral contents. A detailed understading of the types and mineral content is not necessary.

The six (ID) different types of glass are

Glass ID	Type
1	Building window, float processed
2	Building window, non float processed
3	Vehicle window, float processed
5	Containers
6	Table ware
7	Headlamps

The classification variables are

Classification variable	Meaning
RI	Refractive Index
Na	Sodium
Mg	Magnesium
Al	Aluminium
Si	Silicon
K	Potassium
Ca	Calcium
Ba	Barium
Fe	Iron

We will try to distinguish between the different glass types by means of discriminant analyses.

Question 3.1.

The number of resubstitution misclassifications - when going from Quadratic Discriminant Analysis to Linear Discriminant Analysis - increases by:

We find the resubstitution tables in the Enclosure

The DISCRIM Procedure
Classification Summary for Calibration Data: HOME.GLASS
Resubstitution Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into Type							
From Type	1	2	3	5	6	7	Total
1	61 87.14	5 7.14	4 5.71	0 0.00	0 0.00	0 0.00	70 100.00
2	43 56.58	25 32.89	6 7.89	0 0.00	1 1.32	1 1.32	76 100.00
3	0 0.00	0 0.00	16 94.12	0 0.00	1 5.88	0 0.00	17 100.00
5	0 0.00	1 7.69	0 0.00	12 92.31	0 0.00	0 0.00	13 100.00
6	0 0.00	0 0.00	0 0.00	0 0.00	9 100.00	0 0.00	9 100.00
7	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	29 100.00	29 100.00
Total	104 48.60	31 14.49	26 12.15	12 5.61	11 5.14	30 14.02	214 100.00
Priors	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	

The DISCRIM Procedure
Classification Summary for Calibration Data: HOME.GLASS
Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into Type							
From Type	1	2	3	5	6	7	Total
1	46 65.71	14 20.00	10 14.29	0 0.00	0 0.00	0 0.00	70 100.00
2	16 21.05	41 53.95	12 15.79	4 5.26	3 3.95	0 0.00	76 100.00
3	3 17.65	3 17.65	11 64.71	0 0.00	0 0.00	0 0.00	17 100.00
5	0 0.00	2 15.38	0 0.00	10 76.92	0 0.00	1 7.69	13 100.00
6	1 11.11	1 11.11	0 0.00	0 0.00	7 77.78	0 0.00	9 100.00
7	0 0.00	1 3.45	1 3.45	2 6.90	1 3.45	24 82.76	29 100.00

Number of Observations and Percent Classified into Type							
From Type	1	2	3	5	6	7	Total
Total	66 30.84	62 28.97	34 15.89	16 7.48	11 5.14	25 11.68	214 100.00
Priors	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	

QDA row-wise:
5+4+
43+6+1+1+
1+
1+
0+
0= 62

LDA row-wise:
14+10+
16+12+4+3+
3+3+
2+1+
1+1+
1+1+2+1 = 75

ANSWER 3, 75-62 = 13

Question 3.2.

The two most different classes as measured by Mahalanobis distances (based on the pooled dispersion matrix) are:

We find the LDA in the output and the table:

Generalized Squared Distance to Type						
From Type	1	2	3	5	6	7
1	0	1.23907	1.40159	17.49199	15.72381	37.11137
2	1.23907	0	2.01470	11.35021	13.57536	32.09556
3	1.40159	2.01470	0	18.88398	16.68870	41.34160
5	17.49199	11.35021	18.88398	0	16.08992	22.73775
6	15.72381	13.57536	16.68870	16.08992	0	13.94608
7	37.11137	32.09556	41.34160	22.73775	13.94608	0

ANSWER 4, Type 3 and 7

Question 3.3.

Regardless of the previous question we now test if type 1 and type 3 have different mean values, when using all nine classification variables.

The p-value for the usual test of difference in means is:

We have from the note:

Theorem 5.12

Using the significance level α, the critical area for a test of the hypothesis $\mu_1 = \mu_2$ against all alternatives becomes

$$C = \left\{ x_{11}, \cdots, x_{1n_1}, x_{21}, \cdots, x_{2n_2} \;\middle|\; \frac{n_1+n_2-p-1}{p(n_1+n_2-2)} \cdot \frac{n_1 n_2}{n_1+n_2} d^2 > F(p, n_1+n_2-p-1)_{1-\alpha} \right\}$$

Here d^2 is the observed value of D^2.

We find in the output

Class Level Information

Type	Variable Name	Frequency	Weight	Proportion	Prior Probability
1	1	70	70.0000	0.804598	0.500000
3	3	17	17.0000	0.195402	0.500000

Generalized Squared Distance to Type

From Type	1	3
1	0	2.16348
3	2.16348	0

We find n_1=70, n_2=17, p=9, d^2 =2.16348

And insert into the equation

$$\frac{n_1+n_2-p-1}{p(n_1+n_2-2)} \cdot \frac{n_1 n_2}{n_1+n_2} d^2 = \frac{70+17-9-1}{9(70+17-2)} \cdot \frac{70 \cdot 17}{70+17} \cdot 2.16348 = 2.9786$$

We find the degrees of freedom

$$F(p, n_1+n_2-p-1)_{1-\alpha} = F(9, 70+17-9-1)_{1-\alpha} = F(9, 77)_{1-\alpha}$$

We find the P-value using SAS

```
data test;
P = 1-cdf('F',1.9296,9,77);
;
```

0.0042855281

ANSWER 4, 0.0043

Question 3.4.

Considering the Linear Discriminant Analysis, we now test whether Ca, Fe, and Si contribute to the discrimination between type 1 and type 2. The usual test statistic for this is:

We use

Theorem 5.21

The critical region for testing the hypothesis that the last $p-q$ variables do not contribute to the discrimination between the populations π_1 and π_2, i.e. the hypothesis that $\Delta^2_{(2|1)} = 0$ against all alternatives is

$$\left\{x_{11}, \cdots, x_{2n_2} \;\middle|\; \frac{n_1+n_2-p-1}{p-q} \frac{d^2-d_1^2}{(n_1+n_2)(n_1+n_2-2)/(n_1n_2)+d_1^2} > F(p-q, n_1+n_2-p-1)_{1-\alpha}\right\}$$

Here d^2 and d_1^2 are the observed values of D^2 and D_1^2.

We find in the output for the full model for type 1 and 3

Generalized Squared Distance to Type		
From Type	1	3
1	0	2.16348
3	2.16348	0

And for the reduced model

Generalized Squared Distance to Type		
From Type	1	3
1	0	1.11352
3	1.11352	0

We have n1=70 and n2 = 17

Class Level Information					
Type	Variable Name	Frequency	Weight	Proportion	Prior Probability
1	1	70	70.0000	0.804598	0.500000
3	3	17	17.0000	0.195402	0.500000

The full model has p=9 variables and the reduced has q=6

We insert

$$\frac{70+17-9-1}{9-6} \cdot \frac{2.16348 - 1.11352}{(70+17)(70+17-2)/(70 \cdot 17) + 1.11352}$$

ANSWER 1

Question 3.5.

We now consider the Quadratic Discriminant Analysis. When going from the full LDA to the QDA the resubstitution specificity of class 3 is increased by:

We find in the note page 351

Specificity (SPC), **true negative rate** (TNR)	$SPC = TNR = \frac{tn}{fp+tn} = \frac{tn}{NN}$	Conditional probability of being classified negative given true class is negative.

We can use the relation
tn +tp +fn + fp = TN,
where tp, fn, fp, and TN are easy to count, to find tn.
We then use
fp + tn = NN
to find NN
We find in the output

The DISCRIM Procedure
Classification Summary for Calibration Data: HOME.GLASS
Resubstitution Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into Type							
From Type	1	2	3	5	6	7	Total
1	61 87.14	5 7.14	4 5.71	0 0.00	0 0.00	0 0.00	70 100.00
2	43 56.58	25 32.89	6 7.89	0 0.00	1 1.32	1 1.32	76 100.00
3	0 0.00	0 0.00	16 94.12	0 0.00	1 5.88	0 0.00	17 100.00
5	0 0.00	1 7.69	0 0.00	12 92.31	0 0.00	0 0.00	13 100.00
6	0 0.00	0 0.00	0 0.00	0 0.00	9 100.00	0 0.00	9 100.00
7	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	29 100.00	29 100.00
Total	104 48.60	31 14.49	26 12.15	12 5.61	11 5.14	30 14.02	214 100.00
Priors	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	

The DISCRIM Procedure
Classification Summary for Calibration Data: HOME.GLASS

Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into Type							
From Type	1	2	3	5	6	7	Total
1	46 65.71	14 20.00	10 14.29	0 0.00	0 0.00	0 0.00	70 100.00
2	16 21.05	41 53.95	12 15.79	4 5.26	3 3.95	0 0.00	76 100.00
3	3 17.65	3 17.65	11 64.71	0 0.00	0 0.00	0 0.00	17 100.00
5	0 0.00	2 15.38	0 0.00	10 76.92	0 0.00	1 7.69	13 100.00
6	1 11.11	1 11.11	0 0.00	0 0.00	7 77.78	0 0.00	9 100.00
7	0 0.00	1 3.45	1 3.45	2 6.90	1 3.45	24 82.76	29 100.00
Total	66 30.84	62 28.97	34 15.89	16 7.48	11 5.14	25 11.68	214 100.00
Priors	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	

QDA:
fp = 10
fn = 1
tp = 16
TN = 214
tn = 214 - 10 - 1 - 16 = 187
NN = fp + tn = 10 + 187 = 197
SPC = 187 / 197 = 0.949238578680203

LDA:
fp = 23
fn = 6
tp = 11
TN = 214
tn = 214 - 23 - 6 - 11 = 174
NN = fp + tn = 23 + 174 = 197
SPC = 174 / 197 = 0.883248730964467

Increase:
(187 - 174)/197 = 13/197 = 0.065989847715736

ANSWER 5

Problem 4.

Enclosure C with SAS program and SAS output belongs to this problem.

We still consider the glass data introduced in Question 3. However, we now only consider the five types 1, 2, 3, 5, and 7.

Further, we have also omitted the refractive index (RI) from the analysis.

We will now investigate the relationship between the mineral contents.

Question 4.1.

The usual test statistic for the correlation between Mg and Ba conditioned on Al being different from zero is:

We have from the note, page 34

$$\rho_{y_1y_2|x} = \frac{\rho_{y_1y_2} - \rho_{y_1x}\rho_{y_2x}}{\sqrt{(1-\rho_{y_1x}^2)(1-\rho_{y_2x}^2)}}.$$

We find in Enclosure C

Pearson Correlation Coefficients, N = 205									
	Na	Mg	Al	Si	K	Ca	Ba	Fe	Type
Na	1.00000	-0.21301	0.22953	-0.19598	-0.23884	-0.29280	0.38487	-0.22332	0.46359
Mg	-0.21301	1.00000	-0.53020	-0.12761	-0.02797	-0.45236	-0.52505	0.06075	-0.74197
Al	0.22953	-0.53020	1.00000	0.04263	0.33338	-0.28562	0.49080	-0.08111	0.66040
Si	-0.19598	-0.12761	0.04263	1.00000	-0.18026	-0.19165	-0.09596	-0.08019	0.11453
K	-0.23884	-0.02797	0.33338	-0.18026	1.00000	-0.31928	-0.05532	-0.02796	0.04428
Ca	-0.29280	-0.45236	-0.28562	-0.19165	-0.31928	1.00000	-0.11116	0.13615	-0.01943
Ba	0.38487	-0.52505	0.49080	-0.09596	-0.05532	-0.11116	1.00000	-0.06848	0.63422
Fe	-0.22332	0.06075	-0.08111	-0.08019	-0.02796	0.13615	-0.06848	1.00000	-0.15832
Type	0.46359	-0.74197	0.66040	0.11453	0.04428	-0.01943	0.63422	-0.15832	1.00000

We insert

$$\frac{-0.52505 - (-0.53020 \cdot 0.49080)}{\sqrt{(1 - 0.53020^2)(1 - 0.49080^2)}} = -0.3585$$

The test-statistic for this is from page 39

Theorem 1.37

Let $R = R_{ij|m+1\ldots p}$ be the empirical partial correlation coefficient between Z_i and Z_j conditioned on (or: for given) $Z_{m+1,\ldots,Z_p}$. It is assumed to be computed from the unbiased estimates of the variance-covariance matrix and from n observations. Then

$$\frac{R}{\sqrt{1-R^2}}\sqrt{n-2-(p-m)} \sim \mathrm{t}(n-2-(p-m)),$$

if $\rho_{ij|m+1,\ldots,p} = 0$.

We find the number of observations

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Na	205	13.35346	0.76125	2737	10.73000	15.79000
Mg	205	2.74507	1.42743	562.74000	0	4.49000
Al	205	1.44834	0.49715	296.91000	0.29000	3.50000
Si	205	72.62654	0.75264	14888	69.81000	75.18000
K	205	0.51888	0.65783	106.37000	0	6.21000
Ca	205	8.93941	1.42300	1833	5.43000	16.19000
Ba	205	0.18273	0.50668	37.46000	0	3.15000
Fe	205	0.05951	0.09881	12.20000	0	0.51000
Type	205	2.63902	2.03558	541.00000	1.00000	7.00000

We insert

$$\frac{-0.3585}{\sqrt{1-0.3585^2}}\sqrt{205-2-1} = -5.4580$$

ANSWER 2, -5.4580

Question 4.2.

When performing a PCA on the correlation matrix of the data, we need the following number of components to account for at least 90 % of the variance in the data

We find in the enclosure

Eigenvalues of the Correlation Matrix: Total = 8 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.27031849	0.68215593	0.2838	0.2838
2	1.58816256	0.28458435	0.1985	0.4823
3	1.30357821	0.13932153	0.1629	0.6453
4	1.16425668	0.26001291	0.1455	0.7908
5	0.90424377	0.44447192	0.1130	0.9038
6	0.45977184	0.15182670	0.0575	0.9613
7	0.30794515	0.30622185	0.0385	0.9998
8	0.00172330		0.0002	1.0000

We need 5

ANSWER 4, 5 components

Question 4.3.

We still consider a PCA on the correlation matrix of the data.

The usual test-statistic for the last two eigenvalues being equal is:

We find on page 375, theorem 6.8

If we instead are using the estimated *correlation matrix* $\hat{\mathbf{R}}$ we get the criterion

$$Z_2 = -n \log \frac{\det \hat{\mathbf{R}}}{\hat{\lambda}_1 \cdot \ldots \cdot \hat{\lambda}_m \cdot \hat{\lambda}^{k-m}} = -n \log \frac{\hat{\lambda}_{m+1} \cdot \ldots \cdot \hat{\lambda}_k}{\hat{\lambda}_*^{k-m}},$$

where

$$\hat{\lambda}_* = (k - \hat{\lambda}_1 - \ldots - \hat{\lambda}_m)/(k-m) = (\hat{\lambda}_{m+1} + \ldots + \hat{\lambda}_k)/(k-m).$$

The critical region for a test at level α becomes approximately equal to

$$\{x_1, \ldots, x_n | z_2 > \chi^2(\frac{1}{2}(k-m+2)(k-m-1))_{1-\alpha}\}.$$

Eigenvalues of the Correlation Matrix: Total = 8 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	2.27031849	0.68215593	0.2838	0.2838
2	1.58816256	0.28458435	0.1985	0.4823
3	1.30357821	0.13932153	0.1629	0.6453
4	1.16425668	0.26001291	0.1455	0.7908
5	0.90424377	0.44447192	0.1130	0.9038
6	0.45977184	0.15182670	0.0575	0.9613
7	0.30794515	0.30622185	0.0385	0.9998
8	0.00172330		0.0002	1.0000

We have 205 observations and insert

$$Z_2 = -205 \log \frac{0.30794515 \cdot 0.00172330}{((0.30794515 + 0.00172330)/(8-6))^{8-6}} = 781.16$$

ANSWER 5

Question 4.4.

We now consider a factor analysis on the data.

Unrotated factor 1 and 2 combined, explains the following fraction of the variance in *Al*

We find in the output

Factor Pattern				
	Factor1	Factor2	Factor3	Factor4
Na	0.57646	0.07840	-0.65066	-0.27143
Mg	-0.71334	0.56202	-0.27346	-0.18368
Al	0.80937	0.20078	0.34076	0.09827
Si	-0.02664	0.04681	-0.03730	0.97108
K	0.14168	0.56154	0.70389	-0.17491
Ca	-0.19021	-0.91427	0.19794	-0.09887
Ba	0.80830	-0.14404	-0.07658	-0.10002
Fe	-0.25253	-0.22735	0.38396	-0.23201

$0.80937^2+0.20078^2 = 0.6954$

ANSWER 4

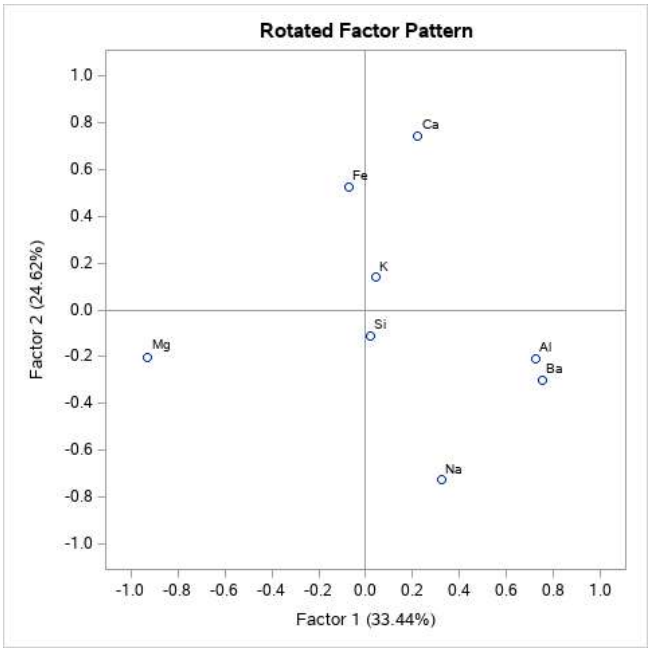
Question 4.5.

We still consider a factor analysis.

The interpretation of the rotated factors is:

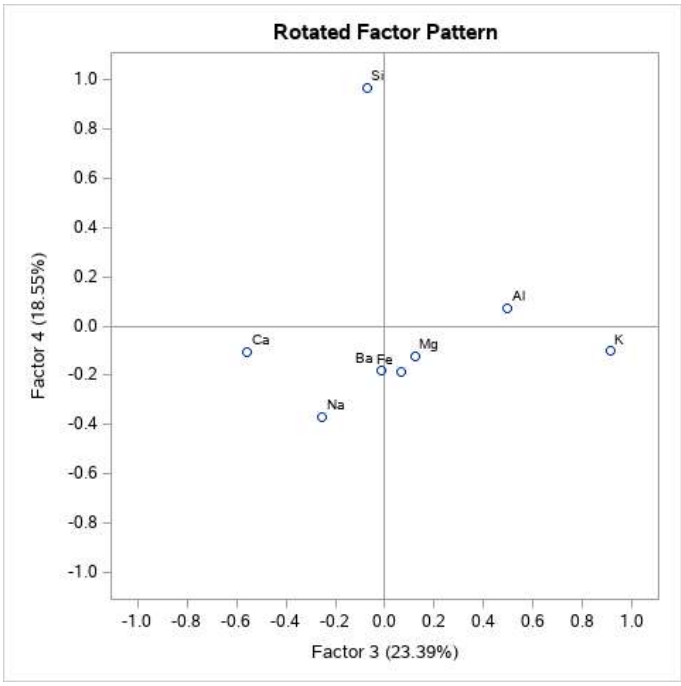
F1 is mainly a contrast between Al and Ba vs Mg

F2 is mainly a contrast between Ca and Fe vs Na



F3 is mainly a contrast between K and Ca

F4 is mainly Si



ANSWER 2

Problem 5

Enclosure D with SAS program and SAS output belongs to this problem.

We now consider high-school students and the connection between 1) the parents' educational level and 2) free vs. payed lunch on the grades in math, reading and writing.

The data are from https://www.kaggle.com/roshansharma/student-performance-analysis

We only consider the first 100 observations. It is not important to understand the variables in detail.

The parental levels are

Name	Meaning - Parents highest level of educations is an
associat	associate degree
bachelor	bachelor degree
college	college degree
highscho	High school degree
master	Masters degree

We consider a model of the form

$$[\text{math reading writing}] = \mu + \text{parents}_k + \text{lunch}_m$$

Question 5.1.

We now consider the parental effect on grades. The p-value for the usual test for parental effect falls in the interval:

We find in the output

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall parental Effect H = Type III SSCP Matrix for parental E = Error SSCP Matrix S=3 M=0 N=45					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
NOTE: F Statistic for Roy's Greatest Root is an upper bound.					
Wilks' Lambda	0.85459909	1.24	12	243.7	0.2544
Pillai's Trace	0.14753888	1.22	12	282	0.2716
Hotelling-Lawley Trace	0.16764634	1.27	12	156.74	0.2392
Roy's Greatest Root	0.15164396	3.56	4	94	0.0094

And see that it falls in the last interval
ANSWER 5

Question 5.2.

Using the usual notation, the matrix Q1 is given by

We find on page 305-306, that Q1 is the residual variation, i.e. error. We find in the output

E = Error SSCP Matrix			
	math	**reading**	**writing**
math	20363.539584	17783.049007	18667.912643
reading	17783.049007	20911.688847	20941.914527
writing	18667.912643	20941.914527	22742.518098

ANSWER 2

Question 5.3.

We now consider the lunch effect. The usual test statistic for lunch effect, if we only consider the variables *math* and *reading,* is:

We find **Q1**

E = Error SSCP Matrix			
	math	reading	writing
math	20363.539584	17783.049007	18667.912643
reading	17783.049007	20911.688847	20941.914527
writing	18667.912643	20941.914527	22742.518098

And Q3

H = Type III SSCP Matrix for lunch			
	math	reading	writing
math	3074.7226868	2903.6887224	2722.2525588
reading	2903.6887224	2742.1686622	2570.8250336
writing	2722.2525588	2570.8250336	2410.1877629

The test-statistic is: $\frac{\det(\boldsymbol{q1})}{\det(\boldsymbol{q1}+\boldsymbol{q3})} = \frac{\det\left(\begin{bmatrix} 20363.539584 & 17783.049007 \\ 17783.049007 & 20911.688847 \end{bmatrix}\right)}{\det\left(\begin{bmatrix} 20363.539584 & 17783.049007 \\ 17783.049007 & 20911.688847 \end{bmatrix} + \begin{bmatrix} 3074.7226868_8 & 2903.6887224 \\ 2903.6887224 & 2742.1686622 \end{bmatrix}\right)} = 0.8666$

ANSWER 5

Problem 6.

We consider a normally distributed random variable

$$\begin{bmatrix} \boldsymbol{Y} \\ \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ X_1 \\ X_2 \end{bmatrix}$$

with expectation vector and dispersion matrix equal to

$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & 4 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 1 & 1 & 4 \end{bmatrix}$$

In the sequel you may find the following expressions useful

$$\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}^{-1} = \frac{1}{7}\begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} - \frac{1}{7}\begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix} = \frac{1}{7}\begin{bmatrix} 10 & 8 \\ 8 & 26 \end{bmatrix}$$

Question 6.1.

The squared correlation between Y_1 and Y_2 is

We start by finding the correlation. For that we use from page 8 in the notes

$$\rho_{ij} = \mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{V}(X_i)\mathrm{V}(X_j)}}.$$

Inserting, we get $\frac{1}{\sqrt{2 \cdot 4}}$, which we then square and get option 3: $\frac{1}{8}$

ANSWER 3, $\frac{1}{8}$

Question 6.2.

The conditional mean $E(Y_1|X_2 = x_2)$

We use from page 23

Theorem 1.27

If X_2 is regularly distributed, i.e. if Σ_{22} has full rank, then the distribution of X_1 conditioned on $X_2 = x_2$ is again a normal distribution, and the following holds

$$\begin{aligned} \mathrm{E}(X_1|X_2 = x_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \mathrm{D}(X_1|X_2 = x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

If Σ_{22} does not have full rank then the conditional distribution is still normal and Σ_{22}^{-1} in the above equations should be substituted by a generalised inverse Σ_{22}^{-}.

We insert

$$E(Y_1|X_2 = x_2) = 0 + 0 \cdot 2^{-1}(x_1 - 1) = 0$$

ANSWER 3

Question 6.3.

The conditional mean $E(Y_1|X_1 = x_1)$ is

We use from page 23

Theorem 1.27

If X_2 is regularly distributed, i.e. if Σ_{22} has full rank, then the distribution of X_1 conditioned on $X_2 = x_2$ is again a normal distribution, and the following holds

$$\begin{aligned} \mathrm{E}(X_1|X_2 = x_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \mathrm{D}(X_1|X_2 = x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

If Σ_{22} does not have full rank then the conditional distribution is still normal and Σ_{22}^{-1} in the above equations should be substituted by a generalised inverse Σ_{22}^{-}.

We insert

$$E(Y_1|X_1 = x_1) = 0 + 1 \cdot 2^{-1}(x_1 - 1) = \frac{1}{2}x_1 - \frac{1}{2}$$

ANSWER 5

Question 6.4.

The conditional mean $E(Y_1|\boldsymbol{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix})$ is

We use from page 23

Theorem 1.27

If X_2 is regularly distributed, i.e. if Σ_{22} has full rank, then the distribution of X_1 conditioned on $X_2 = x_2$ is again a normal distribution, and the following holds

$$\begin{aligned} \mathrm{E}(X_1|X_2 = x_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \mathrm{D}(X_1|X_2 = x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

If Σ_{22} does not have full rank then the conditional distribution is still normal and Σ_{22}^{-1} in the above equations should be substituted by a generalised inverse Σ_{22}^{-}.

We insert

$$D\begin{bmatrix} Y_1 \\ \boldsymbol{X} \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 4 \end{bmatrix}$$

$$E(Y_1 \left| \boldsymbol{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = E(Y_1) + (1 \quad 0)\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}^{-1}\begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix}$$

$$= \frac{1}{7}[1 \quad 0]\begin{bmatrix} 4 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix}$$

$$= \frac{1}{7}[4 \quad -1]\begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix}$$

$$= \frac{4}{7}x_1 - \frac{1}{7}x_2 - \frac{3}{7}$$

ANSWER 1

Question 6.5.

The squared partial correlation $\rho^2_{Y_1Y_2|X_1}$ between Y_1 and Y_2 given X_1 is

We use from page 23

Theorem 1.27

If X_2 is regularly distributed, i.e. if Σ_{22} has full rank, then the distribution of X_1 conditioned on $X_2 = x_2$ is again a normal distribution, and the following holds

$$\begin{aligned} \mathrm{E}(X_1|X_2 = x_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \mathrm{D}(X_1|X_2 = x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

If Σ_{22} does not have full rank then the conditional distribution is still normal and Σ_{22}^{-1} in the above equations should be substituted by a generalised inverse Σ_{22}^{-}.

We insert

$$D(\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}|X_1 = x_1) = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot 2^{-1} \cdot [1 \quad 0] = \begin{bmatrix} 1.5 & 1 \\ 1 & 4 \end{bmatrix}$$

We find the correlation. For that we use from page 8 in the notes

$$\rho_{ij} = \mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{V}(X_i)\mathrm{V}(X_j)}}.$$

Inserting, we get $\frac{1}{\sqrt{1.5 \cdot 4}}$, which we then square and get

ANSWER 4: $\frac{1}{6}$

Question 6.6.

The squared partial correlation $\rho^2_{Y_1Y_2|X_1X_2}$ between Y_1 and Y_2 given $\boldsymbol{X}$ is

We use from page 23

Theorem 1.27

If X_2 is regularly distributed, i.e. if Σ_{22} has full rank, then the distribution of X_1 conditioned on $X_2 = x_2$ is again a normal distribution, and the following holds

$$\begin{aligned} \mathrm{E}(X_1|X_2 = x_2) &= \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \\ \mathrm{D}(X_1|X_2 = x_2) &= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}. \end{aligned}$$

If Σ_{22} does not have full rank then the conditional distribution is still normal and Σ_{22}^{-1} in the above equations should be substituted by a generalised inverse Σ_{22}^{-}.

We insert

$$D(\begin{bmatrix}Y_1\\Y_2\end{bmatrix}|\boldsymbol{X}=\boldsymbol{x}) = \begin{bmatrix}2 & 1\\1 & 4\end{bmatrix} - \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\cdot\begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}^{-1}\cdot\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} = \frac{1}{7}\begin{bmatrix}10 & 8\\8 & 26\end{bmatrix}$$

We find the correlation. For that we use from page 8 in the notes

$$\rho_{ij} = \mathrm{Corr}(X_i, X_j) = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{V}(X_i)\mathrm{V}(X_j)}}.$$

Inserting, we get $\frac{8}{\sqrt{10\cdot 26}}$, which we then square and get

ANSWER 1: $\frac{64}{260}$

Question 6.7.

The squared multiple correlation $\rho^2_{Y_1|X_1X_2}$ between Y_1 and $[X_1 \;\; X_2]^T$ is

We use from page 42

Theorem 1.42

We consider the situation above. Let σ_i be the i'th column in Σ_{xy}, i.e. σ_i^T is the i'th row in Σ_{yx}. Further, let σ_{ii} denote the i'th diagonal element, i.e. the variance of Y_i

Then

$$\rho_{y_i|x} = \frac{\sqrt{\sigma_i^T\Sigma_{xx}^{-1}\sigma_i}}{\sqrt{\sigma_{ii}}}.$$

If we let

$$\Sigma_i = \begin{bmatrix}\sigma_{ii} & \sigma_i^T\\ \sigma_i & \Sigma_{xx}\end{bmatrix},$$

then

$$1-\rho^2_{y_i|x} = \frac{\det\Sigma_i}{\sigma_{ii}\det\Sigma_{xx}} = \frac{\mathrm{V}(Y_i|X)}{\mathrm{V}(Y_i)},$$

We insert

$$1 - \frac{\det\begin{bmatrix}2 & 1 & 0\\1 & 2 & 1\\0 & 1 & 4\end{bmatrix}}{2\cdot\det\begin{bmatrix}2 & 1\\1 & 4\end{bmatrix}} = 0.2857 = \frac{2}{7}$$

ANSWER 3