# Solution Exam 2017

ANYM 20191130

## Problem 1

### Question 1.1

Both from the text and from the SAS-code we identify the problem as a 2-sided (2-way) MANOVA (section 4.3.2), as we have two categorical variables on the right hand side in the model.

```
proc glm data=pendling;
class city distance;
model independent
spouces
topleaders
salaryHigh
salaryMedium
salaryLow
salaryOther
salaryUnknown = city distance;
manova h=_all_/printe printh;
run;
```

We go to the output and find the test for 'city' effect.

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall city Effect
H = Type III SSCP Matrix for city
E = Error SSCP Matrix

S=8 M=44.5 N=338.5

| Statistic | Value | F Value | Num DF | Den DF | Pr > F |
|---|---|---|---|---|---|
| NOTE: F Statistic for Roy's Greatest Root is an upper bound. | | | | | |
| Wilks' Lambda | 0.07372765 | 2.68 | 784 | 5435.1 | <.0001 |
| Pillai's Trace | 1.99721058 | 2.33 | 784 | 5488 | <.0001 |
| Hotelling-Lawley Trace | 3.60827056 | 3.12 | 784 | 4916.4 | <.0001 |
| Roy's Greatest Root | 1.57639735 | 11.03 | 98 | 686 | <.0001 |

The answer is 4

## Question 1.2

We have identified the problem as a 2-sided (2-way) MANOVA (section 4.3.2). The relevant test statistic is given in

**Theorem 4.26**

The ratio test at level $\alpha$ for test of $H_0$ against $H_1$ is given by the critical region

$$\{y_{11},\ldots,y_{km} \mid \frac{\det(\mathbf{q}_1)}{\det(\mathbf{q}_1+\mathbf{q}_2)} \leq U(p,k-1,(k-1)(m-1))_\alpha\}.$$

The ratio test at level $\alpha$ for test of $K_0$ against $K_1$ is given by the critical region

$$\{y_{11},\ldots,y_{km} \mid \frac{\det(\mathbf{q}_1)}{\det(\mathbf{q}_1+\mathbf{q}_3)} \leq U(p,m-1,(k-1)(m-1))_\alpha\}.$$

where *p*=8 is the number of jobs, *k*=99 is the number of cities, and *m*=8 is the number of distances. Inserting $U(p,k-1,(k-1)(m-1)) = U\big(8,99-1,(99-1)(8-1)\big) = U(8,98,686)$

The answer is 1

## Question 1.3

We need to find an F-approximation for the U-distribution, and are asked whether it is exact or approximate. We use

**Theorem 4.22**

Let $U$ be U(s,r,n-k)-distributed and let

$$\begin{aligned} t &= \begin{cases} 1 & s^2+r^2=5 \\ \sqrt{\frac{s^2r^2-4}{s^2+r^2-5}} & s^2+r^2\neq 5 \end{cases} \\ v &= \frac{1}{2}(2(n-k)+r-s-1). \end{aligned}$$

Then

$$\mathrm{F} = \frac{1-U^{\frac{1}{t}}}{U^{\frac{1}{t}}} \cdot \frac{vt+1-\frac{1}{2}sr}{sr}$$

is approximately distributed as

$$\mathrm{F}(sr, vt+1-\frac{1}{2}sr).$$

If either $s$ or $r$ are equal to 1 or 2, then the approximation is exact.

And can see, that we only need to consider the first 2 parameters. We again use theorem 4.26 (see previous question), and find $U(p, m-1, (k-1)(m-1)) = U\big(8,8-1,(99-1)(8-1)\big) = U(8,7,686)$. I.e. In the question p=8, and q=7. The answer 5 follows immediately.

## Question 1.4

We are asked for the test-statistic for an ANOVA. We consider the T, Q1, Q2, Q3 matrices on page 305 (denoted SSCP matrices in the SAS output). The diagonals in the matrices can be used for univariate tests (see e.g. slides lecture K).

We use the test in

**Theorem 2.21**

Let the situation be as above. Then the likelihood ratio test at level $\alpha$ of testing

$$H_0 : \boldsymbol{\mu} \in H \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \in M \backslash H,$$

is equivalent to the test given by the critical region

$$C_\alpha = \{(y_1, \ldots, y_n) | \frac{\|p_M(\boldsymbol{y}) - p_H(\boldsymbol{y})\|^2/(k-r)}{\|\boldsymbol{y} - p_M(\boldsymbol{y})\|^2/(n-k)} > \mathrm{F}(k-r, n-k)_{1-\alpha}\}.$$

We thus only need to find the SS for the city and the SSres. Returning to the definition of the T, Q1, Q2, Q3 matrices on page 305 we see that Q1 has the residual or error SS in the diagonal, and that Q2 has the SS for the city. We find them in the output

H = Type III SSCP Matrix for city

| | independent | spouces | topleaders | salaryHigh | salaryMedium | salaryLow | salaryOther | salaryUnknown |
|---|---|---|---|---|---|---|---|---|
| independent | 0.01711 | 0.00059 | -0.00212 | -0.01228 | -0.01076 | 0.00012 | -0.00162 | 0.00895 |
| spouces | 0.00059 | 0.00003 | -0.00008 | -0.00063 | -0.00038 | 0.00008 | 0.00003 | 0.00035 |
| topleaders | -0.00212 | -0.00008 | 0.00065 | 0.00218 | 0.00176 | -0.00121 | -0.00010 | -0.00108 |
| salaryHigh | -0.01228 | -0.00063 | 0.00218 | 0.03716 | 0.01171 | -0.02145 | -0.00742 | -0.00927 |
| salaryMedium | -0.01076 | -0.00038 | 0.00176 | 0.01171 | 0.01004 | -0.00558 | -0.00030 | -0.00650 |
| salaryLow | 0.00012 | 0.00008 | -0.00121 | -0.02145 | -0.00558 | 0.02159 | 0.00498 | 0.00149 |
| salaryOther | -0.00162 | 0.00003 | -0.00010 | -0.00742 | -0.00030 | 0.00498 | 0.00465 | -0.00021 |
| salaryUnknown | 0.00895 | 0.00035 | -0.00108 | -0.00927 | -0.00650 | 0.00149 | -0.00021 | 0.00627 |

E = Error SSCP Matrix

| | independent | spouces | topleaders | salaryHigh | salaryMedium | salaryLow | salaryOther | salaryUnknown |
|---|---|---|---|---|---|---|---|---|
| independent | 0.072405 | 0.002343 | 0.006323 | 0.030663 | 0.012864 | 0.045904 | 0.005506 | 0.016977 |
| spouces | 0.002343 | 0.000144 | 0.000170 | 0.000743 | 0.000361 | 0.000656 | 0.000125 | 0.000418 |
| topleaders | 0.006323 | 0.000170 | 0.004038 | 0.020799 | 0.007197 | 0.024251 | 0.002766 | 0.006113 |
| salaryHigh | 0.030663 | 0.000743 | 0.020799 | 0.173111 | 0.047890 | 0.148231 | 0.015724 | 0.028085 |
| salaryMedium | 0.012864 | 0.000361 | 0.007197 | 0.047890 | 0.019857 | 0.046981 | 0.006808 | 0.008552 |
| salaryLow | 0.045904 | 0.000656 | 0.024251 | 0.148231 | 0.046981 | 0.321032 | 0.049171 | 0.066990 |
| salaryOther | 0.005506 | 0.000125 | 0.002766 | 0.015724 | 0.006808 | 0.049171 | 0.016025 | 0.009173 |
| salaryUnknown | 0.016977 | 0.000418 | 0.006113 | 0.028085 | 0.008552 | 0.066990 | 0.009173 | 0.024238 |

We insert, where *r*=8 is the number of distances, *k*=rk(x)=106, and *n*=792 the number of observations

$$\frac{\left|\left|P_M(y)-P_H(y)\right|\right|^2/(k-r)}{\left|\left|y-P_M(y)\right|\right|^2/(n-k)} = \frac{0.02159/(106-8)}{0.321032/(792-106)} = \frac{0.02159/98}{0.321032/686}$$

This test can also be written (https://02402.compute.dtu.dk/enotes/chapter8-StatisticsMultigroupANOVA)

**Theorem 8.22**

Under the null hypothesis

$$H_{0,Tr}: \quad \alpha_i = 0, \quad i = 1,2,\ldots,k, \tag{8-44}$$

the test statistic

$$F_{Tr} = \frac{SS(Tr)/(k-1)}{SSE/((k-1)(l-1))}, \tag{8-45}$$

follows an $F$-distribution with $k-1$ and $(k-1)(l-1)$ degrees of freedom. Further, under the null hypothesis

$$H_{0,Bl}: \quad \beta_j = 0, \quad j = 1,2,\ldots,l, \tag{8-46}$$

the test statistic

$$F_{Bl} = \frac{SS(Bl)/(l-1)}{SSE/((k-1)(l-1))}, \tag{8-47}$$

follows an $F$-distribution with $l-1$ and $(k-1)(l-1)$ degrees of freedom.

We insert

$$F = \frac{SS(city)/(k-1)}{SS(error)/((k-1)(m-1))} = \frac{0.02159/98}{0.321032/686}$$

The answer is 4

## Problem 2

### Question 2.1

What fraction of the variance, does the 3 factors describe. We need to now what the total variance is. Recall that factor analysis is always performed on the correlation matrix. See page 404

Furthermore, we assume that the observations are standardised in such a way that $\mathrm{V}(X_i) = 1, \forall i$ i.e. that the variance-covariance matrix for $\boldsymbol{X}$ is equal to its correlation matrix which is denoted

$$\mathrm{D}(\boldsymbol{X}) = \mathbf{R} = \begin{pmatrix} 1 & \cdots & r_{1k} \\ \vdots & & \vdots \\ r_{k1} & \cdots & 1 \end{pmatrix}.$$

Since each variable now have variance 1, the total variance will be the number of variables, i.e. 8.

The fraction explained is then the variance explained of each factor, divided with the total variance.

We find the variance explained by each factor in the output

| Variance Explained by Each Factor | | |
|---|---|---|
| Factor1 | Factor2 | Factor3 |
| 4.6570564 | 2.0695673 | 0.7242992 |

And divide with 8

$$\frac{4.6570564}{8} + \frac{2.0695673}{8} + \frac{0.7242992}{8} = 0.5821 + 0.2587 + 0.0905$$

The answer is 5

## Question 2.2

**Theorem 6.8**

If we are using the estimated *variance-covariance matrix* $\hat{\Sigma}$, the test statistic for testing the hypothesis above becomes

$$Z_1 = -n^* \log \frac{\det \hat{\Sigma}}{\hat{\lambda}_1 \cdot \ldots \cdot \hat{\lambda}_m \cdot \hat{\lambda}_*^{k-m}} = -n^* \log \frac{\hat{\lambda}_{m+1} \cdot \ldots \cdot \hat{\lambda}_k}{\hat{\lambda}_*^{k-m}},$$

where

$$n^* = n - m - \frac{1}{6}(2(k-m) + 1 + \frac{2}{k-m}),$$

and

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

The critical region using a test at level $\alpha$ is approximately

$$\{(x_1, \ldots, x_n) | z_1 > \chi^2(\frac{1}{2}(k-m+2)(k-m-1))_{1-\alpha}\}.$$

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 $\alpha$ becomes approximately equal to

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

However, it should be noted that this approximation is far worse than the corresponding approximation for the variance-covariance matrix.

We want to test for equality of the last 3 eigenvalues. We use theorem 6.8.

Since the eigenvalues are based on the correlation matrix (see previous question), we use the $Z_2$.

We find the number of observation

| Input Data Type | Raw Data |
|---|---|
| Number of Records Read | 792 |
| Number of Records Used | 792 |
| N for Significance Tests | 792 |

the and the eigenvalues in the output

| | |
|---|---|
| 6 | 0.08534339 |
| 7 | 0.06477790 |
| 8 | 0.03807943 |

And insert in the test

$$Z_2 = -792 \log \frac{0.08534339 \cdot 0.06477790 \cdot 0.03807943}{\left[\frac{(0.08534339 + 0.06477790 + 0.03807943)}{3}\right]^3} = 126.2$$

The answer is 2

## Question 2.3

The uniqueness is described on page 404

$$V(X_j) = a_{j1}^2 + \cdots + a_{jm}^2 + \delta_j = 1.$$

Here we introduce the notation

$$h_j^2 = a_{j1}^2 + \cdots + a_{jm}^2, \qquad j = 1, \ldots, k.$$

These quantities are called ***communalities*** and $h_j^2$ describes how large a proportion of $X_j$'s variance is due to the $m$ common factors. Correspondingly $\delta_j$ gives the ***uniqueness*** in $X_j$'s variance. i.e. the proportion of $X_j$'s variance which is not due to the $m$ common factors.

I.e. *communality+uniqueness = 1*. We then just need to find the variable with the lowest communality

| Final Communality Estimates: Total = 7.450923 | | | | | | | |
|---|---|---|---|---|---|---|---|
| independent | spouces | topleaders | salaryHigh | salaryMedium | salaryLow | salaryOther | salaryUnknown |
| 0.95685759 | 0.94505116 | 0.90671210 | 0.89657028 | 0.93199361 | 0.97191100 | 0.91217650 | 0.92965065 |

i.e. salaryHigh. The answer is 3

## Question 2.4

Since we use the correlation matrix in factor analysis (see question 2.1) the total variance is 8.

We find the variance explained by rotated factor 1

| Variance Explained by Each Factor | | |
|---|---|---|
| Factor1 | Factor2 | Factor3 |
| 2.7503238 | 2.6559237 | 2.0446753 |

And the answer is then $2.7503238/8$ i.e option 1

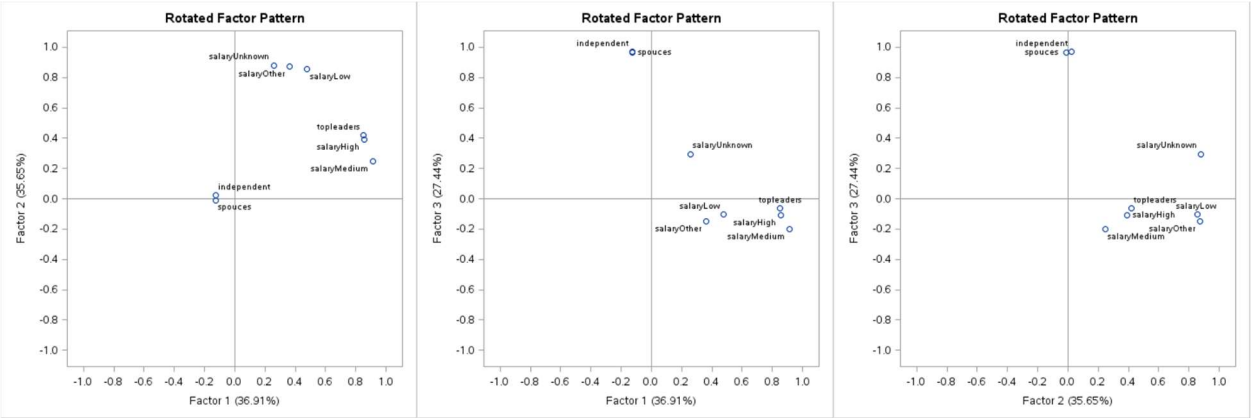
## Question 2.5

There are two components in this question. First we need to understand the meaning of the factors.

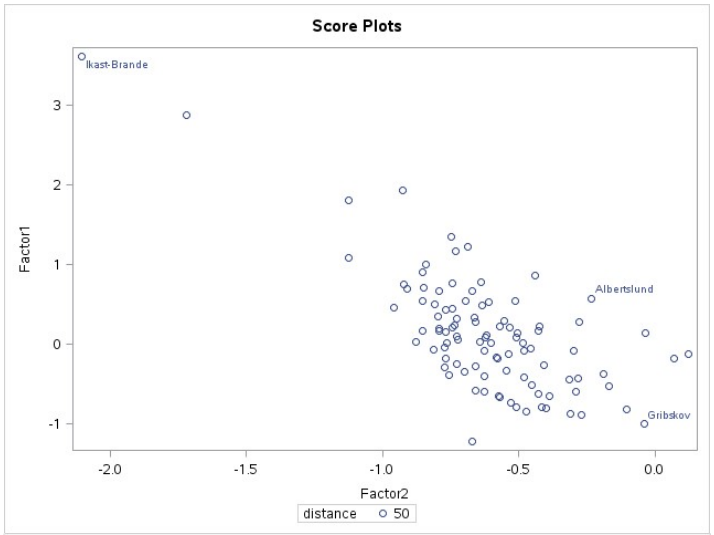
We have

A: *The proportion of high- and medium-salaried as well as of topleaders*
B: *The proportion of independent and of spouces*

If we consider the factor patterns



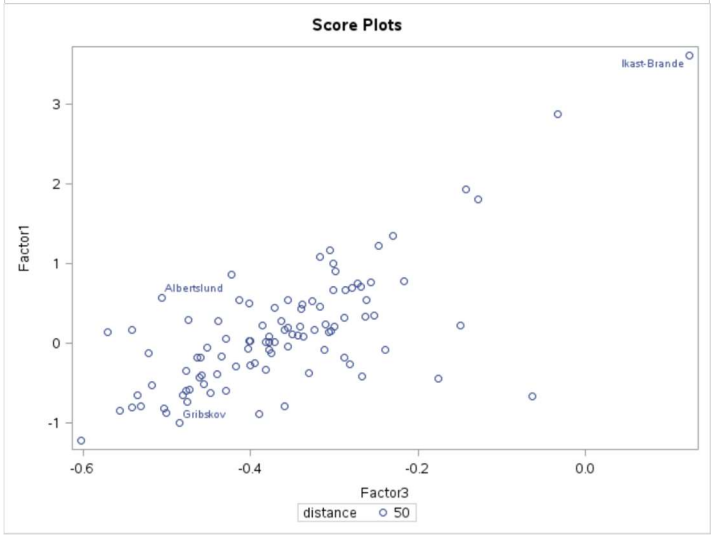

We see that A is described by factor 1, and that B is described by factor 3. We then consider the score plots

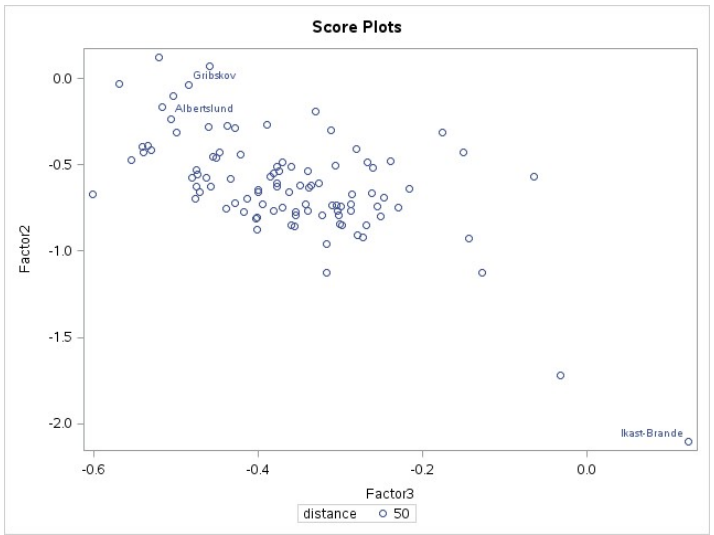



We then want to sort the 32 cities after the proportion of A and B. We start with A, i.e. factor 1. We see that Ikanst Brande has by far the highest loading, followed by Albertslund and Gribskov.

If then consider B, i.e. factor 3, we again see that Ikast-Brande is the highest. The second score-plot is a bit ambigiuos about the order of Gribskov and Albertslund and we consider the third score plot where it is clear that the order is Ikast-Brande, Gribskov, Albertslund.

The answer is then 3



# Problem 3

## Question 3.1

We need to find Hotellings $T^2$ for comparing two groups, i.e. a two-sample situation.

**Theorem 4.9**

We use the same notation as given above. Now, let

$$T^2 = \frac{nm}{n+m}(\bar{X}-\bar{Y})^T \mathbf{S}^{-1}(\bar{X}-\bar{Y}).$$

Then the critical region for a test of $H_0$ against $H_1$ at level $\alpha$ is equal to

$$C = \{x_1,\ldots,x_n,y_1,\ldots,y_m \mid \frac{n+m-p-1}{(n+m-2)p}t^2 > \mathrm{F}(p,n+m-p-1)_{1-\alpha}\}$$

Here $t^2$ is the observed value of $T^2$.

We further know that the $T^2$ and Mahalanobis' distance is closely related

$$D^2 = \| \hat{\mu}_1 - \hat{\mu}_2 \|^2_{\hat{\Sigma}^{-1}} = (\hat{\mu}_1 - \hat{\mu}_2)^T \hat{\Sigma}^{-1} (\hat{\mu}_1 - \hat{\mu}_2)$$

This is equal to the *generalized squared distance* if the priors are equal and the variances are assumed equal, i.e. Linear Discriminant Analysis

**Definition 5.15**

Assuming that the hypothesis $H_0 : \Sigma_1 = \cdots = \Sigma_k$ is true, we define *the squared generalized distance from* $\hat{\mu}_j$ *to population* $\pi_i$ as

$$D_i^2(\hat{\mu}_j) = \begin{cases} (\hat{\mu}_i - \hat{\mu}_j)^T \hat{\Sigma}^{-1} (\hat{\mu}_i - \hat{\mu}_j) & \text{if the priors are equal} \\ (\hat{\mu}_i - \hat{\mu}_j)^T \hat{\Sigma}^{-1} (\hat{\mu}_i - \hat{\mu}_j) - 2\log p_i & \text{if the priors are not all equal} \end{cases}$$

If the hypothesis is *not* true, we define *the squared generalized distance from* $\hat{\mu}_j$ *to population* $\pi_i$ as

$$D_i^2(\hat{\mu}_j) = \begin{cases} (\hat{\mu}_i - \hat{\mu}_j)^T \hat{\Sigma}_i^{-1} (\hat{\mu}_i - \hat{\mu}_j) & \text{if the priors are equal} \\ (\hat{\mu}_i - \hat{\mu}_j)^T \hat{\Sigma}_i^{-1} (\hat{\mu}_i - \hat{\mu}_j) + \log\det\hat{\Sigma}_i - 2\log p_i & \text{if the priors are not all equal} \end{cases}$$

This we can find in the output

| Generalized Squared Distance to quality | | |
|---|---|---|
| From quality | 3 | 8 |
| 3 | 0 | 32.48466 |
| 8 | 32.48466 | 0 |

Together with number of observations

| Class Level Information | | | | | |
|---|---|---|---|---|---|
| quality | Variable Name | Frequency | Weight | Proportion | Prior Probability |
| 3 | _3 | 10 | 10.0000 | 0.454545 | 0.500000 |
| 8 | _8 | 12 | 12.0000 | 0.545455 | 0.500000 |

We can now insert in the expression from theorem 4.9

$$T^2 = \frac{nm}{n+m} D^2 = \frac{10 \cdot 12}{10+12} 32.48466 = 177.1891$$

The answer is 3

## Question 3.2

We have the classification with the 3 variables in the output, however we now need to change the prior probability so that poor wine is twice as probable as good wine, i.e. $p_{bad} = 2/3$ and $p_{good} = 1/3$

Since we want to classify as poor, if the function is larger zero, we subtract good from poor. The situation is the same as in

**Theorem 5.4**

Let $\pi_1 \sim N(\mu_1, \Sigma)$ and $\pi_2 \sim N(\mu_2, \Sigma)$. Then we have

$$\frac{f_1(x)}{f_2(x)} \geq c \Leftrightarrow x^T \Sigma^{-1}(\mu_1 - \mu_2) - \frac{1}{2}\mu_1^T \Sigma^{-1}\mu_1 + \frac{1}{2}\mu_2^T \Sigma^{-1}\mu_2 \geq \log c$$

$$\Leftrightarrow \left[x^T \Sigma^{-1}\mu_1 - \frac{1}{2}\mu_1^T \Sigma^{-1}\mu_1\right] - \left[x^T \Sigma^{-1}\mu_2 - \frac{1}{2}\mu_2^T \Sigma^{-1}\mu_2\right] \geq \log c\,.$$

We look at the output

| Linear Discriminant Function for quality | | |
|---|---|---|
| Variable | 3 | 8 |
| Constant | -176.84275 | -156.35796 |
| pH | 130.26761 | 86.97796 |

| Linear Discriminant Function for quality | | |
|---|---|---|
| Variable | 3 | 8 |
| sulphates | 33.05130 | 61.98355 |
| alcohol | -10.82904 | -1.39863 |

Where the individual discriminant functions are of the form

$$S_i^L(x) = x^T \Sigma^{-1}\mu_i - \frac{1}{2}\mu_i^T \Sigma^{-1}\mu_i + \log p_i$$

i.e. the minus sign is included in the constant term.

We then subtract good (8) from poor (3).

$$[pH \quad sulphates \quad alcohol] \begin{bmatrix} 130.26761 - 86.97796 \\ 33.05130 - 61.98355 \\ -10.82904 + 1.39863 \end{bmatrix} + (-176.84275 + 156.35796) > \log c$$

$$[pH \quad sulphates \quad alcohol]\begin{bmatrix}43.29\\-28.93\\-9.43\end{bmatrix} - 20.4848 > \log c$$

We then just need to adjust the constant term for the change in prior probability. We use

**Theorem 5.1**

The *Bayes solution* to the classification problem is given by the region

$$R_1 = \left\{ x \;\middle|\; \frac{f_1(x)}{f_2(x)} \geq \frac{L_{21}p_2}{L_{12}p_1} \right\}.$$

Since we no nothing about losses we will use equal losses. Yielding

$$[pH \quad sulphates \quad alcohol]\begin{bmatrix}43.29\\-28.93\\-9.43\end{bmatrix} - 20.4848 > \log\frac{p_2}{p_1}$$

$$[pH \quad sulphates \quad alcohol]\begin{bmatrix}43.29\\-28.93\\-9.43\end{bmatrix} - 20.4848 > \log\frac{1/3}{2/3}$$

$$[pH \quad sulphates \quad alcohol]\begin{bmatrix}43.29\\-28.93\\-9.43\end{bmatrix} - 20.4848 > -0.6931$$

$$[pH \quad sulphates \quad alcohol]\begin{bmatrix}43.29\\-28.93\\-9.43\end{bmatrix} - 19.7917 > 0$$

The answer is 5

## Question 3.3

We test for additional information using

**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 $\pi_1$ and $\pi_2$, 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 the number of observations in the output

Class Level Information

| quality | Variable Name | Frequency | Weight | Proportion | Prior Probability |
|---|---|---|---|---|---|
| 3 | _3 | 10 | 10.0000 | 0.454545 | 0.500000 |
| 8 | _8 | 12 | 12.0000 | 0.545455 | 0.500000 |

We have $n_1$=10, $n_2$=12. Further that $p$=3 and $q$=1.

We then need Mahalanobis' distance. We find $d^2$

| Generalized Squared Distance to quality | | |
|---|---|---|
| From quality | 3 | 8 |
| 3 | 0 | 32.48466 |
| 8 | 32.48466 | 0 |

And we find $d_1^2$

| Generalized Squared Distance to quality | | |
|---|---|---|
| From quality | 3 | 8 |
| 3 | 0 | 4.70685 |
| 8 | 4.70685 | 0 |

We can then insert

$$\frac{n_1 + n_2 - p - 1}{p - q} \cdot \frac{d^2 - d_1^2}{(n_1 + n_2)(n_1 + n_2 - 2)/(n_1 n_2) + d_1^2}$$

$$\frac{10 + 12 - 3 - 1}{3 - 1} \cdot \frac{34.48466 - 4.70685}{(10 + 12)(10 + 12 - 2)/(10 \cdot 12) + 4.70685}$$

And the answer is 1.

# Problem 4

## Question 4.1

For a matrix to be positive-definite either the eigenvalues must be positive

**Theorem A.36**

The symmetrical matrix $\boldsymbol{A}$ is positive definite respectively semi-definite, if all $\boldsymbol{A}$ 's eigenvalues are positive respectively non-negative.

This can easily be done in e.g. Maple.

Or we can test it via the determinant of all its principal minors, which must be strictly greater than zero.

**Theorem A.37**

A symmetrical $n \times n$ matrix $\boldsymbol{A}$ is positive definite if the determinants of all ***principal minors***

$$d_i = \det \begin{bmatrix} a_{11} & \cdots & a_{1i} \\ \vdots & & \vdots \\ a_{i1} & \cdots & a_{ii} \end{bmatrix}, \qquad i = 1, \ldots, n,$$

are positive.

We use theorem A.37 and the expression given in the problem and consider the minors

$$\det(\rho_4) = (1-\rho)^{4-1}[1+(4-1)\rho] = (1-\rho)^3(1+3\rho)$$

$$\det(\rho_3) = (1-\rho)^{3-1}[1+(3-1)\rho] = (1-\rho)^2(1+2\rho)$$

$$\det(\rho_2) = (1-\rho)^{2-1}[1+(2-1)\rho] = (1-\rho)(1+\rho)$$

$$\det(\rho_1) = (1-\rho)^{1-1}[1+(1-1)\rho] = 1$$

For $\rho_4$:

For the first parenthesis to be positive, $\rho$ must be less than 1

For the second parenthesis to be positive, $\rho$ must be greater than $-\frac{1}{3}$

$$-\frac{1}{3} < \rho < 1$$

We see that this is also fulfilled by the minors.

The answer is 3

## Question 4.2

To find the correlation coefficient we can use

**Theorem 6.12**

Let the situation be given in the above mentioned definition and let $\mathrm{D}(\mathbf{Z}) = \boldsymbol{\Sigma}$ be partitioned analogously

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix}$$

Then the $r$'th canonical correlation is equal to the $r$'th largest root $\varrho_r$ of

$$\det \begin{bmatrix} -\rho\boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & -\rho\boldsymbol{\Sigma}_{xx} \end{bmatrix} = 0$$

and the coefficients in the $r$'th pair of canonical variables satisfies

$$\text{(i)} \quad \begin{bmatrix} -\varrho_r\boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & -\varrho_r\boldsymbol{\Sigma}_{xx} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_r \\ \boldsymbol{b}_r \end{bmatrix} = \mathbf{0}$$

$$\text{(ii)} \quad \boldsymbol{a}_r^T\boldsymbol{\Sigma}_{yy}\boldsymbol{a}_r = 1$$

$$\text{(iii)} \quad \boldsymbol{b}_r^T\boldsymbol{\Sigma}_{xx}\boldsymbol{b}_r = 1$$

However, we then need to remember to square it afterwards. Or we can use

**Theorem 6.13**

Let the situation be as in the previous theorem. Then we have

$$(\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} - \varrho_r^2\boldsymbol{\Sigma}_{yy})\boldsymbol{a}_r = \mathbf{0}$$

$$\det(\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} - \varrho_r^2\boldsymbol{\Sigma}_{yy}) = 0$$

respectively

$$(\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx} - \varrho_r^2\boldsymbol{\Sigma}_{xx})\boldsymbol{b}_r = \mathbf{0}$$

$$\det(\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx} - \varrho_r^2\boldsymbol{\Sigma}_{xx}) = 0$$

The squared correlation coefficient can then e.g. be found using a symbolic solver, or by fairly tedious calculations, e.g. using

$$\begin{aligned} \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} - x^2\boldsymbol{\Sigma}_{yy} &= \tfrac{\rho^2}{1-\rho^2}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}1 & -\rho\\ -\rho & 1\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} - x^2\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix} \\ &= \tfrac{2\rho^2}{1+\rho}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} - x^2\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix} \\ &= \tfrac{1}{1+\rho}\begin{bmatrix}2\rho^2 - x^2(1+\rho) & 2\rho^2 - x^2\rho(1+\rho)\\ 2\rho^2 - x^2\rho(1+\rho) & 2\rho^2 - x^2(1+\rho)\end{bmatrix} \end{aligned}$$

Or directly on D(z)

The answer is 1.

## Question 4.3

Since our matrix D(Z) is completely symmetric, our first canonical variable must be of the form $V_1 = Y_1 + Y_2$, reducing out options to answers 2, 3 or 5. We then need to figure out the scaling to give it unit variance.

Using Theorem 6.12 (ii) we have

$$[a \quad a]\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\begin{bmatrix}a\\ a\end{bmatrix} = 1$$

$$2a^2 + 2a^2\rho = 1$$

$$a = \pm\frac{1}{\sqrt{2(1+\rho)}}$$

The answer is 2.

If we do not realise that relation, we could start by finding the canonical correlation, when we have that we can find the canonical variables, and finally we can standardize them using 6.12 (ii) or (iii).

By hand it is doable but a bit tedious. With a symbolic solver e.g. Maple or using 'syms' in matlab it is fairly straight forward.

## Problem 5

We start by copying the data from the table into SAS, and at the same time we generate the variables needed in the model.

```
data exam2017;
input obs x1 x2 y;
x1sq = x1**2;
x2sq = x2**2;
x1x2 = x1*x2;
datalines;
1 1 1 370
2 1 2 452
3 1 3 273
4 1 4 422
5 2 1 526
6 2 2 624
7 2 3 513
8 2 4 1059
9 3 1 980
10 3 2 1200
11 3 3 995
12 3 4 1751
;
```

### Question 5.1

We run the regression model in SAS

```
proc reg data=exam2017;
model y = x1 x1sq x2 x2sq x1x2;
run;
```

and find $R^2$ in the output

$$R^2 = \frac{\text{SSTot} - \text{SSRes}}{\text{SSTot}},$$

i.e. the squared multiple correlation coefficient, $R^2$ ***(R-square)*** can also be expressed as the part of the total variation in the $Y$'s which are explained using the independent variables.

| **Root MSE** | 177.28578 | **R-Square** | 0.9102 |
|---|---|---|---|

| | | | |
|---|---|---|---|
| **Dependent Mean** | 763.75000 | **Adj R-Sq** | 0.8355 |
| **Coeff Var** | 23.21254 | | |

The answer is 1

## Question 5.2

We perform backwards elimination

```
proc reg data=exam2017;
model y = x1 x1sq x2 x2sq x1x2 / selection=backward;
run;
```

And find the answer in the output

**Backward Elimination: Step 1**
**Variable x1 Removed: R-Square = 0.9021 and C(p) = 4.5441**

The answer is 5

## Question 5.3

We run the following code

```
proc reg data=exam2017;
model y = x1 x1sq x2 x2sq x1x2 / vif tol;
run;
```

and find the answer in the output

**Parameter Estimates**

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Tolerance | Variance Inflation |
|---|---|---|---|---|---|---|---|
| **Intercept** | 1 | 1000.66667 | 553.57468 | 1.81 | 0.1207 | . | 0 |
| **x1** | 1 | -339.75000 | 460.60197 | -0.74 | 0.4886 | 0.01852 | 54.00000 |
| **x1sq** | 1 | 124.87500 | 108.56493 | 1.15 | 0.2938 | 0.02041 | 49.00000 |

| Parameter Estimates | | | | | | | |
|---|---|---|---|---|---|---|---|
| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| | Tolerance | Variance Inflation |
| x2 | 1 | -531.91667 | 283.10269 | -1.88 | 0.1093 | 0.02614 | 38.25000 |
| x2sq | 1 | 87.58333 | 51.17800 | 1.71 | 0.1379 | 0.03101 | 32.25000 |
| x1x2 | 1 | 106.55000 | 56.06269 | 1.90 | 0.1061 | 0.08333 | 12.00000 |

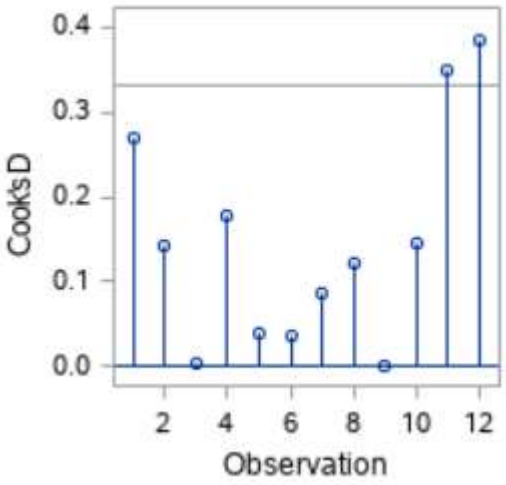
The answer is 5

## Question 5.4

We run the following code

```
proc reg data=exam2017;
model y = x1 x1sq x2 x2sq x1x2 / influence;
run;
```

And find Cook's D in the output




Observation 12

The answer is 5

## Question 5.5

We use the same code as in 5.4

```
proc reg data=exam2017;
model y = x1 x1sq x2 x2sq x1x2 / influence;
run;
```

and consider DFBETAS

While DFFITS measures changes in the prediction of an observation corresponding to changes in all parameter estimates, then ***DFBETAS*** simply measures the change in each individual parameter estimate. As a rule of thumb they should lie within say $\pm 2$. A rule adjusted for number of observations says within $\pm 2/\sqrt{n}$.

$$\text{DFBETAS}_j = \frac{\hat{\theta}_j - \hat{\theta}(i)_j}{\hat{\sigma}(i)\sqrt{(\mathbf{x}^T\mathbf{x})^{-1}_{jj}}}.$$

We look in the output and find the highest numerical value for DFBETAS for the intercept

| Output Statistics | | | | | | |
|---|---|---|---|---|---|---|
| | DFBETAS | | | | | |
| Obs | Intercept | x1 | x1sq | x2 | x2sq | x1x2 |
| 1 | -1.0133 | 0.5621 | -0.2981 | 0.7468 | -0.4216 | -0.6927 |
| 2 | 0.3352 | -0.4532 | 0.3205 | 0.2785 | -0.4532 | 0.2482 |
| 3 | -0.0000 | 0.0397 | -0.0421 | -0.0710 | 0.0595 | 0.0326 |
| 4 | -0.0623 | 0.1124 | -0.2384 | 0.0122 | -0.3371 | 0.5539 |
| 5 | 0.0172 | -0.2485 | 0.2636 | 0.2089 | -0.1864 | 0.0000 |
| 6 | -0.2579 | 0.2861 | -0.3035 | 0.1784 | -0.2146 | -0.0000 |
| 7 | 0.4822 | -0.4636 | 0.4917 | -0.3394 | 0.3477 | 0.0000 |
| 8 | -0.2240 | 0.4615 | -0.4895 | -0.2378 | 0.3462 | -0.0000 |
| 9 | -0.0006 | 0.0000 | -0.0022 | 0.0015 | -0.0031 | 0.0051 |
| 10 | -0.1706 | -0.1538 | 0.3263 | 0.4838 | -0.4614 | -0.2527 |
| 11 | -0.0000 | 0.6413 | -0.6802 | -0.7304 | 0.9620 | -0.5269 |
| 12 | 0.6720 | -0.5192 | 0.3671 | -0.6946 | 0.5192 | 0.8531 |

The answer is 1

## Question 5.6

We run both the full M1 model and the reduced M2 model

```
title 'M1';
proc reg data=exam2017;
model y = x1 x1sq x2 x2sq x1x2;
run;
title 'M2';
proc reg data=exam2017;
model y = x1 x2;
run;
```

We get the ANOVA tables

**M1**

**The REG Procedure**
**Model: MODEL1**
**Dependent Variable: y**

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 5 | 1912595 | 382519 | 12.17 | 0.0043 |
| Error | 6 | 188581 | 31430 | | |
| Corrected Total | 11 | 2101176 | | | |

**M2**

**The REG Procedure**
**Model: MODEL1**
**Dependent Variable: y**

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 2 | 1665432 | 832716 | 17.20 | 0.0008 |
| Error | 9 | 435744 | 48416 | | |
| Corrected Total | 11 | 2101176 | | | |

We use theorem 2.21

See exercise 3.6 for how to use it

We can write is as either

$$\frac{\|p_M(\mathbf{y}) - p_H(\mathbf{y})\|^2/(k-r)}{\|\mathbf{y} - p_M(\mathbf{y})\|/(n-k)} =$$

$$\frac{\left[SS_{mod}(Mod) - SS_{mod}(Hyp)\right] / \left[DF_{mod}(Mod) - DF_{mod}(Hyp)\right]}{SS_{res}(Mod)/DF_{res}(Mod)}$$

$$F = \frac{(1912595 - 1665432)/(5-2)}{188581/6} = 2.6213$$

Or

$$\frac{\|p_M(\mathbf{y}) - p_H(\mathbf{y})\|^2/(k-r)}{\|\mathbf{y} - p_M(\mathbf{y})\|/(n-k)} = \frac{\left[SS_{res}(Hyp) - SS_{res}(Mod)\right] / \left[DF_{res}(Hyp) - DF_{res}(Mod)\right]}{SS_{res}(Mod)/DF_{res}(Mod)}$$

$$F = \frac{(435744 - 188581)/(9-6)}{188581/6} = 2.6213$$

The answer is 4

We could of course also have used the 'vanilla' F-test with

$$\frac{\|p_M(\mathbf{y}) - p_H(\mathbf{y})\|^2/(k-r)}{\|\mathbf{y} - p_M(\mathbf{y})\|/(n-k)} =$$

And found the rank of our system matrix in M1 which is k, and the rank of our system matrix in the reduced model M2 which is r.

## Question 5.7

We use the code from the previous question

```
title 'M2';
proc reg data=exam2017;
model y = x1 x2;
run;
```

**M2**

**The REG Procedure**
**Model: MODEL1**
**Dependent Variable: y**

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
| Model | 2 | 1665432 | 832716 | 17.20 | 0.0008 |
| Error | 9 | 435744 | 48416 | | |
| Corrected Total | 11 | 2101176 | | | |

And get the p-value of the relevant test.

That this is a test of all parameters except the intercept being equal to 0 can be seen from Explanation 2.33 GLM with intercept

If we in the above case compute

$$F = \frac{\mathrm{SS}(\mathrm{Model})/(\mathrm{rk}(\mathbf{x}) - 1)}{\mathrm{SSRes}(\mathrm{Model})/(n - \mathrm{rk}(\mathbf{x}))}$$

this will be the test statistic for the hypothesis that all parameters *except* the intercept are zero, i.e.

The answer is 2

# Problem 6

The problem we consider is treated in theorem 4.21

We first need to realise the dimensions of our parameter matrix. We have 4 dependent variables and 7 independent. That leads to $\theta(7\ x\ 4)$. We then want to extract the 3rd and 4th row for testing

## Question 6.1

C has the same dimensions as the parameters extracted, i.e. 2x4.

The answer is 3

## Question 6.2

We need to extract row 3 and 4. The answer is thus 4

## Question 6.3

We get the usual test statistic from theorem 4.21

$$\{\mathbf{y}|\frac{\det(\mathbf{e})}{\det(\mathbf{e}+\mathbf{h})} \leq \mathrm{U}(s,r,n-k)_{\alpha}\}$$

Where we have

$$\mathbf{A}(r \times k),\ \mathbf{B}(s \times p)\ \ \text{and}\ \mathbf{C}(r \times s)$$

We have A(r x k) = A(2 x 7) and C(r x s) = C(2 x 4). Further we have from Enclosure C that the number of observations is n=1572

We insert

$$U(s,r,n-k) = U(4,2,1572-7) = U(4,2,1565)$$

The answer is 4

# Problem 7

We can use

**Remark 1.10 Rules for computing moments of simple functions**

$$\mathrm{E}(a+bX) = a + b\,\mathrm{E}(X)$$
$$\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$$

$$\mathrm{V}(a+bX) = b^2\,\mathrm{V}(X)$$
$$\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y) + 2\mathrm{Cov}(X,Y)$$
$$= \mathrm{V}(X) + \mathrm{V}(Y) \quad \text{iff } X, Y \text{ independent}$$

$$\mathrm{Cov}(X,X) = \mathrm{V}(X)$$
$$\mathrm{Cov}(aX,bY) = ab\mathrm{Cov}(X,Y)$$
$$\mathrm{Cov}(X+U,Y) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(U,Y)$$
$$\mathrm{Cov}(X,Y+V) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,V)$$

$$\mathrm{E}(\mathbf{A}+\mathbf{X}) = \mathbf{A} + \mathrm{E}(\mathbf{X})$$
$$\mathrm{E}(\mathbf{AX}) = \mathbf{A}\,\mathrm{E}(\mathbf{X})$$
$$\mathrm{E}(\mathbf{XB}) = \mathrm{E}(\mathbf{X})\mathbf{B}$$
$$\mathrm{E}(\mathbf{X}+\mathbf{Y}) = \mathrm{E}(\mathbf{X}) + \mathrm{E}(\mathbf{Y})$$
$$\mathrm{D}(\mathbf{b}+\mathbf{X}) = \mathrm{D}(\mathbf{X})$$
$$\mathrm{D}(\mathbf{AX}) = \mathbf{A}\,\mathrm{D}(\mathbf{X})\mathbf{A}^T$$
$$\mathrm{D}(\mathbf{X}+\mathbf{Y}) = \mathrm{D}(\mathbf{X}) + \mathrm{D}(\mathbf{Y}) + \mathrm{C}(\mathbf{X},\mathbf{Y}) + \mathrm{C}(\mathbf{Y},\mathbf{X})$$
$$= \mathrm{D}(\mathbf{X}) + \mathrm{D}(\mathbf{Y}) \quad \text{iff } X, Y \text{ independent}$$
$$\mathrm{C}(\mathbf{X},\mathbf{X}) = \mathrm{D}(\mathbf{X})$$
$$\mathrm{C}(\mathbf{X},\mathbf{Y}) = \mathrm{C}(\mathbf{Y},\mathbf{X})^T$$
$$\mathrm{C}(\mathbf{AX},\mathbf{BY}) = \mathbf{A}\,\mathrm{C}(\mathbf{X},\mathbf{Y})\mathbf{B}^T$$
$$\mathrm{C}(\mathbf{X}+\mathbf{U},\mathbf{Y}) = \mathrm{C}(\mathbf{X},\mathbf{Y}) + \mathrm{C}(\mathbf{U},\mathbf{Y})$$
$$\mathrm{C}(\mathbf{X},\mathbf{Y}+\mathbf{V}) = \mathrm{C}(\mathbf{X},\mathbf{Y}) + \mathrm{C}(\mathbf{X},\mathbf{V})$$

For the first two questions

## Question 7.1

We have $\mathrm{E}(\mathrm{AX} - \mathrm{BY} + \mathrm{AX}) = \mathrm{E}(\mathrm{AX}) - \mathrm{E}(\mathrm{BY}) + \mathrm{E}(\mathrm{AX}) = \mathrm{AE}(\mathrm{X}) + \mathrm{AE}(\mathrm{X}) - \mathrm{BE}(\mathrm{Y}) = 2\mathrm{A}\mu_x - B\mu_y$

The answer is 1

## Question 7.2

We have $D(AX - BY) = D(AX) + D(BY) - C(AX, BY) - C(BY, AX) = AD(X)A^T + BD(Y)B^T - AC(X,Y)B^T - BC(Y,X)A^T = A\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}A^T + B\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}B^T - A\begin{bmatrix}\rho & \rho\\ \rho & \rho\end{bmatrix}B^T - B\begin{bmatrix}\rho & \rho\\ \rho & \rho\end{bmatrix}A^T$

The answer is 5

## Question 7.3

We use from page 34

$$\rho_{ij|k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{\sqrt{(1-\rho_{ik}^2)(1-\rho_{jk}^2)}}.$$

$$\rho_{x_1x_2|y_2} = \frac{\rho_{x_1x_2} - \rho_{x_1y_2}\rho_{x_2y_2}}{\sqrt{(1-\rho_{x_1y_2}^2)(1-\rho_{x_2y_2}^2)}} = \frac{\rho - \rho\rho}{\sqrt{(1-\rho^2)(1-\rho^2)}} = \frac{\rho - \rho\rho}{1-\rho^2}$$

The answer is 2

## Question 7.4

We use

**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 have 10 observation n=10, p = 4 (four variable x1, x2, y1, y2) and we only condition on one variable, i.e. m=3. We insert

$$t\big(n-2-(p-m)\big) = t\big(10-2-(4-3)\big) = t(7)$$

The answer is 3

## Question 7.5

We use

**Theorem 1.42**

We consider the situation above. Let $\sigma_i$ be the $i$'th column in $\boldsymbol{\Sigma}_{xy}$, i.e. $\sigma_i^T$ is the $i$'th row in $\boldsymbol{\Sigma}_{yx}$. Further, let $\sigma_{ii}$ denote the i'th diagonal element, i.e. the variance of $Y_i$

Then

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

If we let

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_{ii} & \sigma_i^T \\ \sigma_i & \boldsymbol{\Sigma}_{xx} \end{bmatrix},$$

then

$$1-\rho_{y_i|x}^2 = \frac{\det\boldsymbol{\Sigma}_i}{\sigma_{ii}\det\boldsymbol{\Sigma}_{xx}} = \frac{\mathrm{V}(Y_i|\boldsymbol{X})}{\mathrm{V}(Y_i)},$$

We insert and use the equation for determinant given in problem 4

$$\rho^2_{x_1|y_1y_2} = 1 - \frac{\det \Sigma_i}{\sigma_{ii} \det \Sigma_{xx}} = 1 - \frac{\det \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}}{1 \det \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}} = 1 - \frac{(1-\rho)^2(1+2\rho)}{1-\rho^2} = 1 - \frac{2\rho^3 - 3\rho^2 + 1}{1-\rho^2}$$

The answer is 3