Solution Exam 2016

ANYM 20191129

Problem 1

Question 1.1

From the text we identify this as a one-sided (one-way) Manova,

This is confirmed from the output

```
Title 'GLM Analysis on Wine Data Set';
proc glm data=wine plots=none;
class type;
model CollInt Hue Flavan Proline=type/nouni;
manova h=_all_ / printe printh;
run;
```

We note that we only have a single class variable on the right hand isde in the model statement.

We see

Theorem 4.25

The ratio test for the test of the hypothesis H_0 against H_1 is given by the critical region

$$\{y_{11}, \ldots, y_{kn_k} \left| \frac{\det(\mathbf{w})}{\det(\mathbf{t})} \leq U(p, k-1, n-k)_\alpha \right.\}.$$

We have four variables p=4, three classes of grapes k=3, and from the output

Number of Observations Used 178

n=178. Then the distribution is

$$U(p, k-1, n-k) = U(4, 3-1, 178-3) = U(4,2,175)$$

The answer is 3

Question 1.2

We now need to convert the above distribution to an F-distribution. We use

Theorem 4.22

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

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

Then

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

is approximately distributed as

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

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

We have s=4, r=2, and n-k=175

We insert

$$t = \sqrt{\frac{4^2 \cdot 2^2 - 4}{4^2 + 2^2 - 5}} = \sqrt{\frac{16 \cdot 4 - 4}{16 + 4 - 5}} = \sqrt{\frac{60}{15}} = 2$$

$$v = \frac{1}{2}(2(175) + 2 - 4 - 1) = \frac{1}{2} \cdot 347$$

We insert into the F-distribution

$$F\left(4 \cdot 2, \frac{1}{2} \cdot 347 \cdot 2 + 1 - \frac{1}{2} \cdot 4 \cdot 2\right) = F(8{,}344)$$

The answer is 4

Question 1.3

We want to test the equality of 3 mean values for ColInt for each type. This is a 1-way ANOVA.

We get the relevant SS from the diagonals of the two matrices:

E = Error SSCP Matrix				
	ColInt	Hue	Flavan	Proline
ColInt	399.86153923	-7.174341392	50.168702876	11916.394286
Hue	-7.174341392	4.2853382151	-0.736111231	785.92797709
Flavan	50.168702876	-0.736111231	48.073815009	599.49720811
Proline	11916.394286	785.92797709	599.49720811	5198844.3273

H = Type III SSCP Matrix for Type				
	CollInt	Hue	Flavan	Proline
CollInt	551.41600164	-41.76718541	-120.8215497	28929.449698
Hue	-41.76718541	4.9620198299	22.698619657	2223.1115622
Flavan	-120.8215497	22.698619657	128.52239005	26914.708915
Proline	28929.449698	2223.1115622	26914.708915	12353664.645

We can the either get to the correct degrees of freedom by using

Theorem 2.21

Let the situation be as above. Then the likelihood ratio test at level α 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

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

Where the number of observation n=178, we have k=3 effective parameters in our full model (an intercept + a parameter for each type leads to a rank of 3), and only 1 parameter in the reduced mode (an intercept). Thus

$$F = \frac{551.4/(3-1)}{399.9/(178-3)} = \frac{551.4/2}{399.9/175}$$

Alternatively, if one is confused about the degrees of freedom, one can use what is learned in previous courses: https://02402.compute.dtu.dk/enotes/chapter8-StatisticsMultigroupANOVA

Theorem 8.6

Under the null hypothesis

$$H_0 : \quad \alpha_i = 0, \quad i = 1, 2, \ldots, k, \tag{8-18}$$

the test statistic

$$F = \frac{SS(Tr)/(k-1)}{SSE/(n-k)}, \tag{8-19}$$

follows an F-distribution with $k-1$ and $n-k$ degrees of freedom.

The answer is 2

Question 1.4

We compare 2 groups with the Mahalanobis distance and identify it as a two-sample Hotellings T^2-test

We get from

> **Theorem 4.9**
>
> We use the same notation as given above. Now, let
>
> $$T^2 = \frac{nm}{n+m}(\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T \mathbf{S}^{-1}(\bar{\mathbf{X}} - \bar{\mathbf{Y}}).$$
>
> Then the critical region for a test of H_0 against H_1 at level α is equal to
>
> $$C = \{x_1, \ldots, x_n, y_1, \ldots, y_m | \frac{n+m-p-1}{(n+m-2)p} t^2 > \mathrm{F}(p, n+m-p-1)_{1-\alpha}\}$$
>
> Here t^2 is the observed value of T^2.

And are asked for the T-test statistic. We use all four variables so p=4. We find in the output

Class Level Information					
Type	Variable Name	Frequency	Weight	Proportion	Prior Probability
Type1	Type1	59	59.0000	0.331461	0.333333
Type2	Type2	71	71.0000	0.398876	0.333333
Type3	Type3	48	48.0000	0.269663	0.333333

So n=71 and m=48. We then just need the Mahalanobis distance

Generalized Squared Distance to Type		
From Type	Type2	Type3
Type2	0	19.02435
Type3	19.02435	0

i.e. $(\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T \mathbf{S}^{-1}(\bar{\mathbf{X}} - \bar{\mathbf{Y}})$ = 19.02

We can now insert in $\frac{n+m-p-1}{(n+m-2)p} t^2$

$$F = \frac{71 + 48 - 7 - 1}{(71 + 48 - 2)4} \cdot \frac{71 \cdot 48}{71 + 48} 19.02$$

The answer is 5

Question 1.5

We need test whether Flavan and Proline yield further information. We use from

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 need the number of observation for type 2 and 3.

Class Level Information					
Type	Variable Name	Frequency	Weight	Proportion	Prior Probability
Type1	Type1	59	59.0000	0.331461	0.333333
Type2	Type2	71	71.0000	0.398876	0.333333
Type3	Type3	48	48.0000	0.269663	0.333333

We need the Mahalanobis for all four variables d^2

Linear Discriminant Analysis on Types 2 and 3 from Wine Data Set
Variables used are ColInt, Hue, Flavan, Proline

The DISCRIM Procedure

Generalized Squared Distance to Type		
From Type	Type2	Type3
Type2	0	19.02435
Type3	19.02435	0

And for ColInt and Hue only d_1^2

Linear Discriminant Analysis on Types 2 and 3 from Wine Data Set
Variables used are ColInt, Hue

The DISCRIM Procedure

Generalized Squared Distance to Type		
From Type	Type2	Type3
Type2	0	9.57363
Type3	9.57363	0

We can now collect everything and insert in $\frac{n_1+n_2-p-1}{p-q} \frac{d^2-d_1^2}{(n_1+n_2)(n_1+n_2-2)/(n_1n_2)+d_1^2}$

$$F = \frac{71+48-4-1}{2} \cdot \frac{19.02-9.57}{\frac{(71+48)(71+48-2)}{71 \cdot 48} + 9.57}$$
$$= \frac{71+48-5}{2} \cdot \frac{71 \cdot 48 \cdot (19.02-9.57)}{(71+48)(71+46) + 71 \cdot 48 \cdot 9.57}$$

The answer is 5

Question 1.6

The observation is:

Obs	Type	Colint	Hue	Flavan	Proline
142	Type3	5.6000	0.700	0.50	780

We look at the LDF's

Linear Discriminant Analysis on Types 2 and 3 from Wine Data Set
Variables used are Colnt, Hue

The DISCRIM Procedure

Linear Discriminant Function for Type		
Variable	Type2	Type3
Constant	-24.27489	-23.28922
Colint	2.11025	3.50475
Hue	39.79642	30.25656

We have

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 subtract the two and get

$$[Colint \quad Hue]\begin{bmatrix} 2.11 - 3.50 \\ 39.80 - 30.26 \end{bmatrix} + (-24.27 + 23.29) > 0$$

Since we have equal priors and losses, we just 0 on the right hand side.

The answer is 3.

Question 1.7

We now need to change c. We have

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 the losses are still equal, we have $\log \frac{2/3}{1/3} = -0.6931$. The answer is 2.

Question 1.8

We find the number of misclassifications in the output, where they are off the diagonal. We note that we should look at the tables for cross validation.

LDA page 7 in output

Classification Summary for Calibration Data: WORK.WINE
Cross-validation Summary using Linear Discriminant Function

Number of Observations and Percent Classified into Type				
From Type	Type1	Type2	Type3	Total
Type1	56	3	0	59
	94.92	5.08	0.00	100.00
Type2	5	65	1	71
	7.04	91.55	1.41	100.00
Type3	0	1	47	48
	0.00	2.08	97.92	100.00
Total	61	69	48	178
	34.27	38.76	26.97	100.00
Priors	0.33333	0.33333	0.33333	

A total of 3+5+1+1=10 misclassifications. We find the QDA on page 9 in the output.

Classification Summary for Calibration Data: WORK.WINE
Cross-validation Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into Type				
From Type	Type1	Type2	Type3	Total
Type1	58	1	0	59
	98.31	1.69	0.00	100.00
Type2	3	67	1	71
	4.23	94.37	1.41	100.00
Type3	0	1	47	48
	0.00	2.08	97.92	100.00
Total	61	69	48	178
	34.27	38.76	26.97	100.00
Priors	0.33333	0.33333	0.33333	

A total of 1+3+1+1=6 misclassifications.

We reduce the number by 10-6=4. The answer is 4

Problem 2

Question 2.1

We use from page 34

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

We insert

$$\rho_{24|3} = \frac{\rho^2 - \rho \cdot \rho}{\sqrt{(1-\rho^2)(1-\rho^2)}} = 0$$

The answer is 1

Question 2.2

We use

Theorem 1.40

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

Then

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

If we let

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

then

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

We insert

$$\rho_{1|23}^2 = 1 - \frac{\det \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}}{1 \cdot \det \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}} = 1 - \frac{1}{1-\rho^2}(1 + \rho^4 + \rho^4 - \rho^4 - \rho^2 - \rho^2) = 1 - \frac{1 + \rho^4 - 2\rho^2}{1 - \rho^2}$$
$$= 1 - \frac{(1-\rho^2)(1-\rho^2)}{1-\rho^2} 1 - (1-\rho^2) = \rho^2$$

The answer is 3

Question 2.3

We use

> **Theorem 1.40**
>
> We consider the situation above. Let σ_i be the i'th column in $\mathbf{\Sigma}_{xy}$, i.e. σ_i^T is the i'th row in $\mathbf{\Sigma}_{yx}$. Further, let σ_{ii} denote the i'th diagonal element, i.e. the variance of Y_i
> Then
> $$\rho_{y_i|x} = \frac{\sqrt{\sigma_i^T \mathbf{\Sigma}_{xx}^{-1} \sigma_i}}{\sqrt{\sigma_{ii}}}.$$
> If we let
> $$\mathbf{\Sigma}_i = \begin{bmatrix} \sigma_{ii} & \sigma_i^T \\ \sigma_i & \mathbf{\Sigma}_{xx} \end{bmatrix},$$
> then
> $$1 - \rho_{y_i|x}^2 = \frac{\det \mathbf{\Sigma}_i}{\sigma_{ii} \det \mathbf{\Sigma}_{xx}} = \frac{\mathrm{V}(Y_i|\mathbf{X})}{\mathrm{V}(Y_i)},$$

We insert

$$\rho_{2|34}^2 = 1 - \frac{\det \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}}{1 \cdot \det \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}} = 1 - \frac{1}{1-\rho^2}(1 + \rho^4 + \rho^4 - \rho^4 - \rho^2 - \rho^2) = 1 - \frac{1 + \rho^4 - 2\rho^2}{1-\rho^2}$$
$$= 1 - \frac{(1-\rho^2)(1-\rho^2)}{1-\rho^2} 1 - (1-\rho^2) = \rho^2$$

The answer is 3

Question 2.4

We have the hint to calculate it directly in the conditional distribution. Lets us try that. We start by using

> **Theorem 1.26**
>
> If $\mathbf{X}_2$ is regularly distributed, i.e. if $\mathbf{\Sigma}_{22}$ has full rank, then the distribution of $\mathbf{X}_1$ conditioned on $\mathbf{X}_2 = \mathbf{x}_2$ is again a normal distribution, and the following holds
> $$\begin{aligned} \mathrm{E}(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) &= \boldsymbol{\mu}_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ \mathrm{D}(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) &= \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}. \end{aligned}$$
> If $\mathbf{\Sigma}_{22}$ does not have full rank then the conditional distribution is still normal and $\mathbf{\Sigma}_{22}^{-1}$ in the above equations should be substituted by a generalised inverse $\mathbf{\Sigma}_{22}^{-}$.

$$D\left(\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \middle| \begin{bmatrix} X_3 \\ X_4 \end{bmatrix}\right) = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} - \begin{bmatrix} \rho^2 & \rho^3 \\ \rho & \rho^2 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho^2 & \rho \\ \rho^3 & \rho^2 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} - \frac{1}{1-\rho^2} \begin{bmatrix} \rho^2(\rho^2 - \rho^4) & \rho^2(\rho - \rho^3) \\ \rho(\rho^2 - \rho^4) & \rho(\rho - \rho^3) \end{bmatrix}$$
$$= \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} - \frac{1}{1-\rho^2} \begin{bmatrix} \rho^4(1-\rho^2) & \rho^3(1-\rho^2) \\ \rho^3(1-\rho^2) & \rho^2(1-\rho^2) \end{bmatrix} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} - \begin{bmatrix} \rho^4 & \rho^3 \\ \rho^3 & \rho^2 \end{bmatrix} = \begin{bmatrix} 1-\rho^4 & \rho - \rho^3 \\ \rho - \rho^3 & 1 - \rho^2 \end{bmatrix}$$

We now have the conditional dispersion or variance-covariance matrix. To calculate the correlation we use

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

$$\rho_{12|34} = \frac{\rho - \rho^3}{\sqrt{(1-\rho^4)\cdot(1-\rho^2)}} = \frac{\rho(1-\rho^2)}{\sqrt{(1-\rho^2)\cdot(1+\rho^2)\cdot(1-\rho^2)}} = \frac{\rho}{\sqrt{1+\rho^2}}$$

The answer is 5

Problem 3

Question 3.1+3.2

We find the values α and β by using

Theorem 2.3

Let $\mathbf{x}$ and $\boldsymbol{\theta}$ be given as in the preceding section and let $Y \sim \mathrm{N}_n(\mathbf{x}\,\boldsymbol{\theta}, \sigma^2\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is positive definite. Then the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ is given by $\mathbf{x}\,\hat{\boldsymbol{\theta}}$ being the projection (with respect to $\boldsymbol{\Sigma}$) onto M, $\hat{\boldsymbol{\theta}}$ is a solution to the so-called *normal equation(s)*

$$(\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x})\hat{\boldsymbol{\theta}} = \mathbf{x}^T\boldsymbol{\Sigma}^{-1}y.$$

If $\mathbf{x}$ has full rank k, then

$$\hat{\boldsymbol{\theta}} = (\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x})^{-1}\mathbf{x}^T\boldsymbol{\Sigma}^{-1}Y,$$

Since the errors are normally distributed with common variance we get $\Sigma = I$. We then have

$$x^T x = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 7 & 0 \\ 0 & 28 \end{bmatrix}$$

$$(x^T x)^{-1} = \begin{bmatrix} 1/7 & 0 \\ 0 & 1/28 \end{bmatrix}$$

$$x^TY = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 21 \end{bmatrix}$$

$$\theta = (x^Tx)^{-1}x^TY = \begin{bmatrix} 1/7 & 0 \\ 0 & 1/28 \end{bmatrix} \begin{bmatrix} 2 \\ 21 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 21/28 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 3/4 \end{bmatrix}$$

α is thus 2/7 and β is ¾

The answer is thus 1 and 4 respectively.

Question 3.3

We find the variance of a predicted value page 123

$$U = (z_1, \ldots, z_k) \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \end{bmatrix} = \mathbf{z}^T \hat{\boldsymbol{\theta}}$$

as our predictor.

We have that $\mathrm{E}(U) = \mathrm{E}(Y)$ and that

$$\begin{aligned} V(U) &= \mathbf{z}^T \mathrm{D}(\hat{\boldsymbol{\theta}}) \mathbf{z} \\ &= \sigma^2 \mathbf{z}^T (\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1} \mathbf{z} \\ &= \sigma^2 c, \end{aligned}$$

where

$$c = (z_1, \ldots, z_k)(\mathbf{x}^T \boldsymbol{\Sigma}^{-1} \mathbf{x})^{-1} \begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix}.$$

We start by finding σ^2.

σ^2 is estimated using

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-k}\|Y - \mathrm{x}\hat{\theta}\|^2$$
$$= \frac{1}{n-k}(Y - \mathrm{x}\hat{\theta})^T\Sigma^{-1}(Y - \mathrm{x}\hat{\theta}).$$

We have 7 observations n=7 and 2 parameters in our model k=2. We insert

$$\sigma^2 = \frac{1}{7-2}\cdot 3\cdot\frac{19}{28} = \frac{1}{5}\cdot 3\cdot\frac{19}{28}$$

Where $3\cdot\frac{19}{28}$ is given in the problem.

We now estimate *c*

$$c = \begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}1/7 & 0\\ 0 & 1/28\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} = 1/7$$

The variance is then

$$V(U) = \sigma^2 c = \frac{1}{5}\cdot 3\cdot\frac{19}{28}\cdot\frac{1}{7} = \frac{1}{35}\cdot 3\cdot\frac{19}{28}$$

The answer is 5

Question 3.4

The confidence interval for a mean predicted value is given in

> **Theorem 2.15**
>
> Let the situation be as above. Then the $(1-\alpha)$-*confidence interval* for the expected value of a new observation Y will be
>
> $$[u - \mathrm{t}(n-k)_{1-\frac{\alpha}{2}} s\sqrt{c}, \quad u + \mathrm{t}(n-k)_{1-\frac{\alpha}{2}} s\sqrt{c}].$$

We see that the degree of freedom for the t-distribution is

$$t(n-k) = t(7-2) = 5$$

Problem 4

Question 4.1

We use

Theorem 6.3

The principal components are uncorrelated and the variance of the i'th component is λ_i i.e. the i'th largest eigenvalue.

We find the first eigenvalue in the output

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	9.01626784	7.38629813	0.7514	0.7514

The answer is 3

Question 4.2

We use

Remark 6.7

From the theorem we have that if we seek the linear combination of the original variables which explains most of the variation in these, then the first principal component is the solution. If we seek the m variables which explain most of the original variation, then the solution is the m first principal components. A measure of how well these describe the original variation is found by means of theorems 6.3 and 6.5 which show that the m first principal components describe the fraction

$$\frac{\lambda_1 + \cdots + \lambda_m}{\lambda_1 + \cdots + \lambda_m + \cdots + \lambda_k}$$

of the original variation.

I.e. we look at the cumulative fraction of the eigenvalues which we find in the output

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	9.01626784	7.38629813	0.7514	0.7514
2	1.62996970	0.75537952	0.1358	0.8872
3	0.87459018	0.72547429	0.0729	0.9601
4	0.14911589	0.06258133	0.0124	0.9725
5	0.08653455	0.00778286	0.0072	0.9797
6	0.07875169	0.01712586	0.0066	0.9863
7	0.06162583	0.02716816	0.0051	0.9914
8	0.03445767	0.00126063	0.0029	0.9943
9	0.03319704	0.01084316	0.0028	0.9970
10	0.02235388	0.01438586	0.0019	0.9989
11	0.00796802	0.00280031	0.0007	0.9996
12	0.00516771		0.0004	1.0000

We thus need 3 components and the answer is 3.

Question 4.3

We find the principal components in the output and look for a contrast between the X and Y variables

Eigenvectors							
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
X1	0.265382	-.344758	0.341909	-.542694	0.225310	0.417245	0.262426
X2	0.304236	0.048394	0.337080	-.044573	-.761246	0.263464	-.249231
X3	0.274342	0.434404	0.000887	-.001131	-.037508	-.140833	-.208649
X4	0.293249	0.356342	0.052250	-.070095	0.164756	-.083087	-.294205
X5	0.305738	0.264857	0.089019	-.182918	0.413683	0.072865	-.178829
X6	0.253710	0.477851	-.088451	-.036464	-.098089	-.056597	0.767310
Y1	0.265175	-.157106	0.531964	0.690452	0.195213	-.045306	0.112292
Y2	0.291196	-.328656	0.100210	-.109384	-.140987	-.682585	0.168448
Y3	0.294705	-.235797	-.354534	-.104648	-.119393	-.184028	-.044614
Y4	0.310159	-.197091	-.247105	-.107818	0.015822	-.089014	-.167410
Y5	0.312982	-.161072	-.213937	0.187527	0.272425	0.088748	-.153441
Y6	0.285988	-.107409	-.474810	0.342581	-.117167	0.451223	0.153737

The best candidate is Prin2, while not perfect because of the sign on X1.

The answer is 2.

Question 4.4

We look for an overall mean of the variables. This is best reflected by Prin1: Approximately same values and the same sign.

The answer is 1

Question 4.5

We use 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

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

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

We find in the output

Observations	1000
Variables	12

And the eigenvalues

	Eigenvalue
1	9.01626784
2	1.62996970
3	0.87459018
4	0.14911589
5	0.08653455
6	0.07875169
7	0.06162583
8	0.03445767
9	0.03319704
10	0.02235388
11	0.00796802
12	0.00516771

We insert

$$Z_2 = -1000 \log \frac{0.08653455 \cdot 0.07875169 \cdot 0.06162583 \cdot 0.03445767 \cdot 0.03319704 \cdot 0.02235388 \cdot 0.00796802 \cdot 0.00516771}{[(0.08653455 + 0.07875169 + 0.06162583 + 0.03445767 + 0.03319704 + 0.02235388 + 0.00796802 + 0.00516771)/8]^8}$$
$$= 2943.6$$

The answer is 3

Question 4.6

We use 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

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

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

And insert

$$\chi^2\left(\frac{1}{2}(12-4+2)(12-4-1)\right) = \chi^2\left(\frac{1}{2} \cdot 10 \cdot 7\right) = \chi^2(35)$$

The answer is 5

Problem 5

Question 5.1

We find it in the output

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation
1	0.953951	0.953511	0.002847	0.910022
2	0.645910	0.641986	0.018439	0.417200
3	0.399448	0.390246	0.026590	0.159559
4	0.316757	0.313096	0.028464	0.100335
5	0.226966	.	0.030009	0.051514
6	0.029380	.	0.031611	0.000863

The answer is 3

Question 5.2

We find the test 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
1	0.03757517	134.00	36	4341.4	<.0001
2	0.41760357	38.96	25	3675.5	<.0001
3	0.71654635	21.80	16	3025.1	<.0001
4	0.85258356	18.15	9	2412	<.0001
5	0.94766752	13.51	4	1984	<.0001
6	0.99913684	0.86	1	993	0.3546

We see 5 is significantly different.

The answer is 2

Question 5.3

We find the second canonical variates in the output

Standardized Canonical Coefficients for the 1986-06-27 data						
	CV1986_1	CV1986_2	CV1986_3	CV1986_4	CV1986_5	CV1986_6
X1	0.5514	-0.5409	1.9129	1.2145	-0.5961	-0.7176
X2	0.3166	-0.7202	-2.6187	0.1623	0.8722	0.7393
X3	-0.2869	-0.4019	1.9721	2.6552	3.2173	-7.8240
X4	0.5800	0.8112	-2.1606	-2.2725	-8.1019	4.9352
X5	0.0354	0.8211	0.7829	-3.0269	4.3987	-0.1936
X6	-0.1694	0.3592	0.5710	2.0738	0.2119	2.8311

Standardized Canonical Coefficients for the 2005-08-18 data						
	CV2005_1	CV2005_2	CV2005_3	CV2005_4	CV2005_5	CV2005_6
Y1	0.3993	-0.0204	-1.6169	0.3869	-0.3325	1.4917
Y2	0.3149	-1.6782	0.9617	0.6863	1.3566	-2.5848
Y3	-0.6950	-1.9933	0.7574	0.7115	0.4530	7.1496
Y4	1.1925	2.0726	-0.5733	-0.6626	-5.3062	-4.8381
Y5	-0.0346	1.0631	1.3199	-3.3716	3.0067	0.9448
Y6	-0.1711	0.6832	-0.9417	2.7726	0.9864	-1.8341

We see that they are a contrast between the first 3 variables and the last 3.

The answer is 4

Problem 6

Question 6.1

We find the usual test

Test statistic for $H_0: E(\boldsymbol{Y}) \in H$ against $H_1: E(\boldsymbol{Y}) \in M \backslash H$:

$$\frac{\|p_M(Y) - p_H(Y)\|^2/(k-r)}{\|Y - p_M(Y)\|^2/(n-k)} = \frac{(SS_{res}(Hyp) - SS_{res}(Mod))/(DF_{res}(Hyp) - DF_{res}(Mod))}{SS_{res}(Mod)/DF_{res}(Mod)}$$

We find the relevant tables in the output for model M

Model M

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	12	0.36471758	0.03039313	399.06	<.0001
Error	32	0.00243716	0.00007616		
Corrected Total	44	0.36715474			

And for model H

Model H

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.36071708	0.36071708	2409.39	<.0001
Error	43	0.00643765	0.00014971		
Corrected Total	44	0.36715474			

We insert

$$F = \frac{(0.00643765 - 0.00243716)/(43-32)}{0.00243716/32}$$

The answer is 4

Question 6.2

We find the test statistic

Theorem 2.21

Let the situation be as above. Then the likelihood ratio test at level α 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 insert $F(k-r, n-k) = F(43-32, 32) = F(11,32)$

The answer is 4

Question 6.3

We use

Remark 2.24

The estimated standard deviation $(\hat{V}(\hat{\theta}_{i_0}))^{1/2}$ of $\hat{\theta}_{i_0}$ is often provided by standard software as *'standard error of estimate'* or similar. It is thus straight forward to compute the critical limits. This result may also be used in setting up confidence limits for θ_{i_0}. More specifically, a $(1-\alpha)$ confidence interval becomes

$$\left[\hat{\theta}_{i_0} - t(f)_{1-\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{\theta}_{i_0})},\ \ \hat{\theta}_{i_0} + t(f)_{1-\frac{\alpha}{2}}\sqrt{\hat{V}(\hat{\theta}_{i_0})}\right]$$

Where *f* is $f = n - \mathrm{rk}(\boldsymbol{x})$

We have n=45 observations, and rk(x) is 2 (intercept and γ), leading to f=43

We find the output for model H

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.0039255847	0.00267248	1.47	0.1491
corrDose	0.0110837163	0.00022580	49.09	<.0001

The standard error is $\sqrt{\hat{V}(\hat{\theta}_{i_0})}$

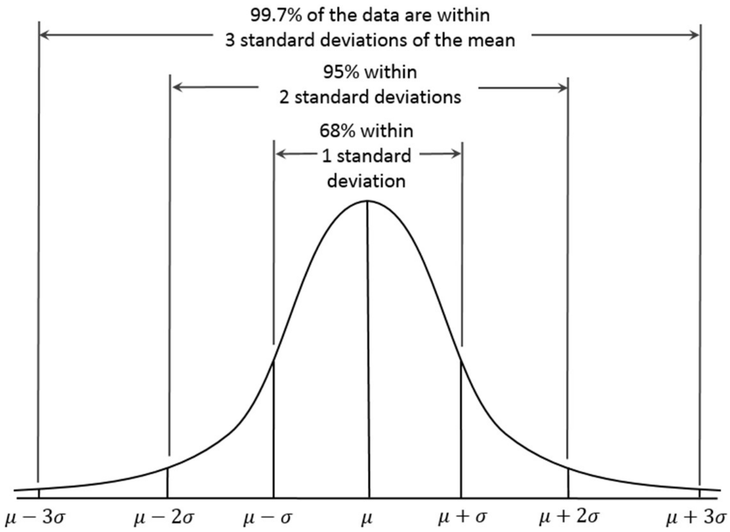
We insert

$$0.0110837163 \pm t(43)_{0.975} \cdot 0.00022580$$

The answer is 1

Question 6.4

We assume the residuals follow a standardized normal distribution, i.e. variance equal 1. We now seek residuals that have a probability of 5% or smaller. Since we assume they follow a standardized normal distribution, that means we seek residuals smaller than -2 or larger than 2.



We look in the output

Obs	scan	Posi-tion	gel	Corr-Dose	FBP	Residual	Rstudent	CooksD	DFFITS
1	1	1	1	2.0186	0.01367	0.007048	0.94786	0.02647	0.58567
2	1	2	2	1.9866	0.02543	0.001404	0.18616	0.00105	0.11483
3	1	3	3	1.9965	0.02454	-0.007504	-1.02920	0.03498	-0.67496
4	1	4	4	2.0397	0.01659	0.000341	0.04606	0.00007	0.03032
5	1	5	5	0.0000	-0.01325	-0.016442	-2.16650	0.06007	-0.93328
6	2	1	1	2.0190	0.02909	0.014385	2.02970	0.11019	1.25387
7	2	2	2	1.9871	0.02252	-0.009598	-1.30637	0.04885	-0.80562
8	2	3	3	1.9970	0.03902	-0.001118	-0.15089	0.00078	-0.09894
9	2	4	4	2.0402	0.02545	0.001115	0.15067	0.00078	0.09917
10	2	5	5	0.0000	0.02176	0.010488	1.32391	0.02444	0.57031
11	3	1	6	2.0144	0.02901	0.001400	0.19059	0.00131	0.12865
12	3	2	7	1.9840	0.03208	0.003853	0.52637	0.00990	0.35468
13	3	3	8	1.9884	0.02083	-0.000144	-0.01953	0.00001	-0.01315
14	3	4	9	2.0253	0.02047	-0.001676	-0.22834	0.00189	-0.15437
15	3	5	5	0.0000	-0.00033	-0.005359	-0.66395	0.00655	-0.28915
16	1	1	6	9.8833	0.11614	0.002191	0.29802	0.00317	0.20009
17	1	2	7	9.7431	0.11877	0.005442	0.74582	0.01956	0.50076
18	1	3	8	9.7201	0.10500	-0.000773	-0.10506	0.00039	-0.07054
19	1	4	9	9.9245	0.10739	-0.001437	-0.19527	0.00136	-0.13112
20	1	5	5	0.0000	0.00765	0.004461	0.55045	0.00442	0.23712
21	2	1	6	9.8833	0.11987	-0.002162	-0.29414	0.00309	-0.19749
22	2	2	7	9.7431	0.11064	-0.010771	-1.51655	0.07664	-1.01823
23	2	3	8	9.7201	0.11972	0.005864	0.80484	0.02271	0.54037
24	2	4	9	9.9245	0.11255	-0.004360	-0.59561	0.01256	-0.39992
25	2	5	5	0.0000	0.00288	-0.008388	-1.04815	0.01563	-0.45152
26	3	1	1	9.8745	0.08778	-0.008724	-1.17737	0.03903	-0.71662
27	3	2	2	9.7305	0.10687	-0.005781	-0.77063	0.01714	-0.46910
28	3	3	3	9.7352	0.12881	0.008195	1.12322	0.04035	0.72725
29	3	4	4	9.9704	0.10293	-0.004037	-0.54498	0.00979	-0.35285
30	3	5	5	0.0000	0.01054	0.005510	0.68292	0.00692	0.29741
31	1	1	1	20.6310	0.19444	-0.020758	-3.18228	0.23801	-1.99416
32	1	2	2	20.4520	0.24854	0.017584	2.57796	0.17013	1.61305

Obs	scan	Posi-tion	gel	Corr-Dose	FBP	Residual	Rstudent	CooksD	DFFITS
33	1	3	3	20.4120	0.24268	0.004261	0.57990	0.01163	0.38484
34	1	4	4	21.0180	0.23534	0.006414	0.88120	0.02693	0.58957
35	1	5	5	0.0000	0.00096	-0.002232	-0.27445	0.00111	-0.11823
36	2	1	1	20.5970	0.23095	0.008050	1.09161	0.03574	0.68366
37	2	2	2	20.4180	0.23505	-0.003608	-0.48173	0.00715	-0.30125
38	2	3	4	20.3780	0.20731	-0.003834	-0.77440	0.10030	-1.13470
39	2	4	3	20.9830	0.26776	-0.003834	-0.77440	0.10030	-1.13470
40	2	5	5	0.0000	0.01904	0.007771	0.96868	0.01342	0.41729
41	3	1	6	20.6480	0.23500	-0.001428	-0.19615	0.00147	-0.13603
42	3	2	7	20.4780	0.23695	0.001477	0.20272	0.00156	0.14043
43	3	3	8	20.3710	0.22203	-0.004947	-0.68348	0.01748	-0.47275
44	3	4	9	20.9250	0.24142	0.007472	1.04478	0.04059	0.72747
45	3	5	5	0.0000	0.00922	0.004190	0.51768	0.00400	0.22545

I.e. 4 observations.

The answer is 4.

Question 6.5

The largest overall influence on the model parameters, can be measured by the Cooks D or the DFFITS. We have

Cook's D

A confidence region for the parameter $\boldsymbol{\theta}$ is all the vectors $\boldsymbol{\theta}^*$, which satisfy

$$\frac{1}{p\hat{\sigma}^2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)^T \mathbf{x}^T \mathbf{x}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \le F(p, n-p)_{1-\alpha}.$$

We use the left hand side as a measure of the distance between the parameter vector and $\hat{\boldsymbol{\theta}}$. We let $\hat{\boldsymbol{\theta}}(i)$ be the estimate, which corresponds to the deletion of the i'th observation

$$\mathbf{y}(i) = (y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_n)^T$$

and therefore have

$$\hat{\boldsymbol{\theta}}(i) = [\mathbf{x}(i)^T \mathbf{x}(i)]^{-1} \mathbf{x}(i)^T \mathbf{y}(i).$$

Cook's D then equals

$$\frac{1}{p\hat{\sigma}^2}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}(i))^T \mathbf{x}^T \mathbf{x}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}(i)).$$

And

DFFITS

DFFITS is - like Cook's distance - a measure of the total change when deleting one single observation. As a rule of thumb they should lie within say ± 2. A similar rule adjusted for number of observations says within $\pm 2\sqrt{p/(n-p)}$.

$$\text{DFFITS} = \frac{\hat{y}_i - \hat{y}(i)_i}{\hat{\sigma}(i)\sqrt{h_{ii}}} = \frac{\mathbf{x}_i[\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}(i)]}{\hat{\sigma}(i)\sqrt{h_{ii}}}.$$

We consult the output and find the observation with the largest values for Cooks D and DFFITS

Obs	scan	Posi-tion	gel	Corr-Dose	FBP	Residual	Rstudent	CooksD	DFFITS
1	1	1	1	2.0186	0.01367	0.007048	0.94786	0.02647	0.58567
2	1	2	2	1.9866	0.02543	0.001404	0.18616	0.00105	0.11483
3	1	3	3	1.9965	0.02454	-0.007504	-1.02920	0.03498	-0.67496
4	1	4	4	2.0397	0.01659	0.000341	0.04606	0.00007	0.03032
5	1	5	5	0.0000	-0.01325	-0.016442	-2.16650	0.06007	-0.93328
6	2	1	1	2.0190	0.02909	0.014385	2.02970	0.11019	1.25387
7	2	2	2	1.9871	0.02252	-0.009598	-1.30637	0.04885	-0.80562
8	2	3	3	1.9970	0.03902	-0.001118	-0.15089	0.00078	-0.09894
9	2	4	4	2.0402	0.02545	0.001115	0.15067	0.00078	0.09917
10	2	5	5	0.0000	0.02176	0.010488	1.32391	0.02444	0.57031
11	3	1	6	2.0144	0.02901	0.001400	0.19059	0.00131	0.12865
12	3	2	7	1.9840	0.03208	0.003853	0.52637	0.00990	0.35468
13	3	3	8	1.9884	0.02083	-0.000144	-0.01953	0.00001	-0.01315
14	3	4	9	2.0253	0.02047	-0.001676	-0.22834	0.00189	-0.15437
15	3	5	5	0.0000	-0.00033	-0.005359	-0.66395	0.00655	-0.28915
16	1	1	6	9.8833	0.11614	0.002191	0.29802	0.00317	0.20009
17	1	2	7	9.7431	0.11877	0.005442	0.74582	0.01956	0.50076
18	1	3	8	9.7201	0.10500	-0.000773	-0.10506	0.00039	-0.07054
19	1	4	9	9.9245	0.10739	-0.001437	-0.19527	0.00136	-0.13112
20	1	5	5	0.0000	0.00765	0.004461	0.55045	0.00442	0.23712
21	2	1	6	9.8833	0.11987	-0.002162	-0.29414	0.00309	-0.19749
22	2	2	7	9.7431	0.11064	-0.010771	-1.51655	0.07664	-1.01823
23	2	3	8	9.7201	0.11972	0.005864	0.80484	0.02271	0.54037
24	2	4	9	9.9245	0.11255	-0.004360	-0.59561	0.01256	-0.39992
25	2	5	5	0.0000	0.00288	-0.008388	-1.04815	0.01563	-0.45152
26	3	1	1	9.8745	0.08778	-0.008724	-1.17737	0.03903	-0.71662
27	3	2	2	9.7305	0.10687	-0.005781	-0.77063	0.01714	-0.46910
28	3	3	3	9.7352	0.12881	0.008195	1.12322	0.04035	0.72725
29	3	4	4	9.9704	0.10293	-0.004037	-0.54498	0.00979	-0.35285
30	3	5	5	0.0000	0.01054	0.005510	0.68292	0.00692	0.29741
31	1	1	1	20.6310	0.19444	-0.020758	-3.18228	0.23801	-1.99416
32	1	2	2	20.4520	0.24854	0.017584	2.57796	0.17013	1.61305

i.e observation 31. The answer is 2.