Exam 2014, 9. December

1.1	1.2	1.3	1.4	1.6	1.5	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	3.4
2	2	3	4	2	4	2	1	4	4	2	1	4	5	1
3.5	4.1	4.2	4.3	4.4	5.1	5.2	5.3	5.4	5.5	6.1	6.2	6.3	6.4	6.5
3	5	2	5	5	3	1	5	4	4	2	4	1	3	1

Problem 1

We consider the model

$$\begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ X_3 & Y_3 & Z_3 \\ X_4 & Y_4 & Z_4 \\ X_5 & Y_5 & Z_5 \end{bmatrix} = \begin{bmatrix} 1 & -2 & 2 \\ 1 & -1 & -1 \\ 1 & 0 & -2 \\ 1 & 1 & -1 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} \alpha_x & \alpha_y & \alpha_z \\ \beta_x & \beta_y & \beta_z \\ \gamma_x & \gamma_y & \gamma_z \end{bmatrix} + \begin{bmatrix} \varepsilon_1 & \delta_1 & \varphi_1 \\ \varepsilon_2 & \delta_2 & \varphi_2 \\ \varepsilon_3 & \delta_3 & \varphi_3 \\ \varepsilon_4 & \delta_4 & \varphi_4 \\ \varepsilon_5 & \delta_5 & \varphi_5 \end{bmatrix}$$

Where the error terms $[\varepsilon_i \quad \delta_i \quad \varphi_i]'$, $i = 1,2,3,4,5,$ are independent and normally distributed $N_3(\mathbf{0}, \mathbf{\Sigma})$, and where $\mathbf{\Sigma}$ is the unknown dispersion matrix

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{xz} & \sigma_{yz} & \sigma_z^2 \end{bmatrix}$$

We assume that we obtained the following observations

$$\begin{bmatrix} 2 & 4 & 6 \\ 1 & 4 & 5 \\ 3 & 2 & 2 \\ 4 & 3 & 5 \\ 0 & 2 & 2 \end{bmatrix}$$

With the usual notation we have

$$\boldsymbol{x}'\boldsymbol{x} = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 14 \end{bmatrix}$$

Question 1.1-2

<u>*The maximum likelihood estimator for α_x becomes:*</u>

From Theorem 4.14:

Theorem 4.14

We consider the above mentioned situation. If the observations $\boldsymbol{Y}_i$ are normally distributed the maximum likelihood estimate of $\boldsymbol{\theta}$ is given by

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

we can simply implement:

$$X = \begin{bmatrix} 1 & -2 & 2 \\ 1 & -1 & -1 \\ 1 & 0 & -2 \\ 1 & 1 & -1 \\ 1 & 2 & 2 \end{bmatrix}$$

And $X^T X$=$\begin{bmatrix} 5 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 14 \end{bmatrix}$

$$Y = \begin{bmatrix} 2 & 4 & 6 \\ 1 & 4 & 5 \\ 3 & 2 & 2 \\ 4 & 3 & 5 \\ 0 & 2 & 2 \end{bmatrix}$$

So $X^T Y$=$\begin{bmatrix} 10 & 15 & 20 \\ -1 & -5 & -8 \\ -7 & 1 & 2 \end{bmatrix}$

So, the maximum likelihood estimator is:

$(X^T X)^{-1} \cdot X^T Y$=

From the table it can be seen that the value we are looking for is 2.

$$\begin{bmatrix} \alpha_x & \alpha_y & \alpha_z \\ \beta_x & \beta_y & \beta_z \\ \gamma_x & \gamma_y & \gamma_z \end{bmatrix}$$

The answer is 2

Question 1.2-2

The covariance between the maximum likelihood estimators for αx and βx becomes:

From the Theorem 4.18

Theorem 4.18

We consider the situation from theorem 4.14. Then the maximum likelihood estimate for $\mathbf{\Sigma}$ equals

$$\begin{aligned} \hat{\mathbf{\Sigma}}^* &= \frac{1}{n}\sum_{i=1}^{n}(\mathbf{Y}_i - \hat{\boldsymbol{\theta}}^T\mathbf{x}_i)(\mathbf{Y}_i - \hat{\boldsymbol{\theta}}^T\mathbf{x}_i)^T \\ &= \frac{1}{n}(\mathbf{Y} - \mathbf{x}\hat{\boldsymbol{\theta}})^T(\mathbf{Y} - \mathbf{x}\hat{\boldsymbol{\theta}}) \\ &= \frac{1}{n}[\mathbf{Y}^T\mathbf{Y} - (\mathbf{x}\hat{\boldsymbol{\theta}})^T(\mathbf{x}\hat{\boldsymbol{\theta}})]. \end{aligned}$$

The (i, j)'th element can also be written

$$\hat{\sigma}_{ij}^* = \frac{1}{n}(\mathbf{Y}_{i|} - \mathbf{x}\hat{\boldsymbol{\theta}}_{i|})^T(\mathbf{Y}_{j|} - \mathbf{x}\hat{\boldsymbol{\theta}}_{j|}).$$

We can directly see from the given , that is zero.

The answer is 2.

Question 1.3-3

The covariance between the maximum likelihood estimators for αx and αy becomes:

According to what it has been mentioned before, value 1/5 from $(X^T X)^{-1}$ is multiplied with the term σ_{xy} to find the covariance between the likelihood estimators for a_x and a_y.

The answer is 3.

Question 1.4-4

We now want to test the hypothesis

$$H_0 : \begin{bmatrix} \beta_y & \beta_z \\ \gamma_y & \gamma_z \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

against all alternatives. This hypothesis may also be written

$$H_0\text{: } \boldsymbol{A} \begin{bmatrix} \alpha_x & \alpha_y & \alpha_z \\ \beta_x & \beta_y & \beta_z \\ \gamma_x & \gamma_y & \gamma_z \end{bmatrix} \boldsymbol{B}' = \boldsymbol{C} \text{ against } H_1\text{: } \boldsymbol{A} \begin{bmatrix} \alpha_x & \alpha_y & \alpha_z \\ \beta_x & \beta_y & \beta_z \\ \gamma_x & \gamma_y & \gamma_z \end{bmatrix} \boldsymbol{B}' \neq \boldsymbol{C}$$

Here the matrix $\boldsymbol{A}$ is:

From Theorem 4.21

|||| **Theorem 4.21**

We consider the above mentioned situation including the assumption of the normality of the observations. Furthermore we consider the hypothesis

$$H_0 : \mathbf{A}\,\boldsymbol{\theta}\,\mathbf{B}^T = \mathbf{C} \quad \text{against} \quad H_1 : \mathbf{A}\,\boldsymbol{\theta}\,\mathbf{B}^T \neq \mathbf{C},$$

where $\mathbf{A}(r \times k)$, $\mathbf{B}(s \times p)$ and $\mathbf{C}(r \times s)$ are given (known) matrices. We introduce

$$\begin{aligned} \boldsymbol{\Delta} &= \mathbf{A}\,\hat{\boldsymbol{\theta}}\,\mathbf{B}^T - \mathbf{C} \\ \mathbf{R} &= n\hat{\boldsymbol{\Sigma}}^* = (\mathbf{Y} - \mathbf{x}\hat{\boldsymbol{\theta}})^T(\mathbf{Y} - \mathbf{x}\hat{\boldsymbol{\theta}}) = \mathbf{Y}^T\mathbf{Y} - \hat{\boldsymbol{\theta}}^T(\mathbf{x}^T\mathbf{x})\hat{\boldsymbol{\theta}} \end{aligned}$$

and

$$\begin{aligned} \mathbf{E} &= \mathbf{B}\,\mathbf{R}\,\mathbf{B}^T \\ \mathbf{H} &= \boldsymbol{\Delta}^T[\mathbf{A}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{A}^T]^{-1}\boldsymbol{\Delta}. \end{aligned}$$

The likelihood ratio test for testing H_0 against H_1 is equivalent to the test given by the critical region

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

where $\mathrm{U}(s, r, n - k)_\alpha$ is the α quantile in the null-hypothesis distribution of the test statistic (see below).

A selects the rows and we need to select the second and the third row so:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The answer is 4.

Question 1.5-2

__The matrix B is:__

B selects the columns so we have to select the second and the third column, and in this case the dimensions are 2x3:

$$B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The answer is 2.

Question 1.6-4

__If the hypothesis H_0 is true, then the distribution of the usual test statistic is:__

From Theorem 4.21,

From the answers 1.5 and 1.4 it can be seen that the dimensions of A and B matrices are (r x k) and (s x p) respectively.

So r = 2, k = 3, s = 2, p = 3 and n the number of the observations n = 5. Then the Usual test statistic is U(2,2,5-3)

The answer is 4.

Problem 2

We consider a normally distributed random variable

$$\begin{bmatrix} X \\ Y \end{bmatrix}$$

that represents a measurement of a two-dimensional property on subjects that belong to one of two populations π_1 or π_2. The dispersion matrix is equal to

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

The mean value depends on which population the subject belongs to:

$$E\left(\begin{bmatrix} X \\ Y \end{bmatrix}\right) = \begin{cases} \boldsymbol{\mu}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \ if\ \pi_1\ is\ true \\ \\ \boldsymbol{\mu}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \ if\ \pi_2\ is\ true \end{cases}$$

We finally assume that the prior probabilities of belonging to either population is 0.5. We immediately obtain

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{bmatrix}$$

$$\boldsymbol{\mu}_1'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_1 = \frac{4}{3} \quad , \quad \boldsymbol{\mu}_2'\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_2 = \frac{64}{3} \quad , \quad (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\mathbf{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = 12$$

Question 2.1-2

The linear discriminant function Z for distinguishing between the two populations is:

From Theorem 5.4:

Theorem 5.4

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

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

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

Where X=$\begin{bmatrix} X \\ Y \end{bmatrix}$, Thus the Linear Discriminant Function becomes:

$\left[\begin{bmatrix} X \\ Y \end{bmatrix}^T \Sigma^{-1}\mu_1 - \frac{1}{2}\mu_1^T \Sigma^{-1}\mu_1\right] - \left[\begin{bmatrix} X \\ Y \end{bmatrix}^T \Sigma^{-1}\mu_2 - \frac{1}{2}\mu_2^T \Sigma^{-1}\mu_2\right]$, After substitution:

$$\left[\begin{bmatrix} X \\ Y \end{bmatrix}^T \begin{bmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{1}{2}\frac{4}{3}\right] - \left[\begin{bmatrix} X \\ Y \end{bmatrix}^T \begin{bmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 4 \\ 4 \end{bmatrix} - \frac{1}{2}\frac{64}{3}\right] = -2X - 2Y + 10$$

The answer is 2.

Question 2.2-1

If π_2 *is true then the mean of* Z *is*:

From Theorem 5.8:

Theorem 5.8

We consider the random variable defined by the linear discriminator (omitting the term $-\log c$), i.e.

$$Z = \mathbf{X}^T \boldsymbol{\Sigma}^{-1}(\mu_1 - \mu_2) - \frac{1}{2}\mu_1^T \boldsymbol{\Sigma}^{-1}\mu_1 + \frac{1}{2}\mu_2^T \boldsymbol{\Sigma}^{-1}\mu_2\,.$$

Then

$$Z \sim \begin{cases} \mathrm{N}\left(+\frac{1}{2}\|\mu_1 - \mu_2\|^2_{\Sigma^{-1}},\ \|\mu_1 - \mu_2\|^2_{\Sigma^{-1}}\right) & \text{if } \pi_1 \text{ is true} \\ \mathrm{N}\left(-\frac{1}{2}\|\mu_1 - \mu_2\|^2_{\Sigma^{-1}},\ \|\mu_1 - \mu_2\|^2_{\Sigma^{-1}}\right) & \text{if } \pi_2 \text{ is true} \end{cases}.$$

The mean is:

$$-\frac{1}{2}\|\mu_1 - \mu_2\|^2_{\Sigma^{-1}} = -\frac{1}{2}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) = -\frac{1}{2}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \end{bmatrix}\right)^T \begin{bmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{bmatrix} \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \end{bmatrix}\right)$$

$$= -\frac{1}{2}12 = -6$$

The answer is 1.

Question 2.3-4

<u>The variance of Z is:</u>

$$\|\mu_1 - \mu_2\|^2_{\Sigma^{-1}} = (\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2) = \left(\begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}4\\4\end{bmatrix}\right)^T \begin{bmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{bmatrix} \left(\begin{bmatrix}1\\1\end{bmatrix} - \begin{bmatrix}4\\4\end{bmatrix}\right) = 12$$

The answer is 4.

Question 2.4-4

<u>Let Φ(x) be the cumulative distribution function of a N(0,1)-distributed random variable. If</u> π_2 *is true then the probability of misclassification is*:

The prior probabilities of belonging to either population is 0.5. From Theorem 5.1:

Theorem 5.1

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

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

Question 2.5-2

Problem 3

Enclosure A belongs to this problem. The data are the first 20 records of gene expression data taken from www.biostat.jhsph.edu/~ririzarr/Teaching/649/. It is not necessary to be familiar with gene expression data in order to solve the problems. The variables are called X and Y and we wish to see how well ln(Y) may be predicted based on ln(X). We do a regression analysis with ln(Y) = lny as dependent variable and ln(X) = lnx as independent variable

Question 3.1-1

<u>The 95% confidence interval for the coefficient to lnx is:</u>

From Theorem 2.15:

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

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

For a 95% confidence interval $a = 0.05 \Rightarrow 1 - a = 0.975$, and $\hat{\theta}_{io}$ is the coefficient in the regression model:

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.94929	0.22328	4.25	0.0005
lnx	1	0.96429	0.03336	28.91	<.0001

From the Table below it can be seen that n-k for the model is 18

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8.47517	8.47517	835.78	<.0001
Error	18	0.18253	0.01014		
Corrected Total	19	8.65770			

and $\sqrt{\hat{\theta}_{io}}$ is standard error of the coefficient of the above table. So the answer is

$$0.96429 - t(18)_{0.975} \times 0.3336 + t(18)_{0.975} \times 0.3336$$

The answer is 1.

Question 3.2-4

__The length of the 95% prediction interval for the last observation is:__

According to the Theorem 2.15:

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

$$[u - \mathrm{t}(n-k)_{1-\frac{\alpha}{2}} s\sqrt{c},\quad u + \mathrm{t}(n-k)_{1-\frac{\alpha}{2}} s\sqrt{c}].$$

For a 95% confidence interval $a = 0.05 \Rightarrow 1 - a = 0.975$, and u is the predicted value of the observations of last $u = 1185.78$ from the table below.

Output Statistics										
Obs	Dep. Variable	Pred. Value	Std Err. Mean Predict	Residual	Std Err. Residual	Student Residual	-2-1 0 1 2	Cook's D	Hat Diag H	DFFITS
1	6.9206	6.7342	0.0315	0.1864	0.0956	1.949	|***	0.206	0.0979	0.7023
2	7.4827	7.4960	0.0229	-0.0133	0.0981	-0.135	|	0.001	0.0518	-0.0308
3	7.5165	7.6248	0.0242	-0.1084	0.0978	-1.108	**|	0.038	0.0576	-0.2758
4	7.5384	7.6478	0.0245	-0.1094	0.0977	-1.120	**|	0.039	0.0590	-0.2825
5	7.8167	7.9019	0.0290	-0.0852	0.0964	-0.883	*|	0.035	0.0832	-0.2645
6	6.5770	6.6671	0.0332	-0.0901	0.0951	-0.948	*|	0.055	0.1085	-0.3297
7	6.4485	6.5542	0.0361	-0.1057	0.0940	-1.125	**|	0.094	0.1288	-0.4360
8	6.6869	6.7706	0.0306	-0.0837	0.0959	-0.872	*|	0.039	0.0926	-0.2767
9	6.6002	6.4097	0.0402	0.1904	0.0923	2.062	|****	0.402	0.1591	0.9975
10	7.0512	7.0293	0.0254	0.0219	0.0974	0.225	|	0.002	0.0638	0.0572
11	6.7760	6.7285	0.0316	0.0476	0.0956	0.498	|	0.014	0.0988	0.1612
12	7.4753	7.3886	0.0225	0.0868	0.0981	0.884	|*	0.021	0.0500	0.2016
13	7.6346	7.5255	0.0231	0.1091	0.0980	1.113	|**	0.035	0.0528	0.2647
14	9.1361	9.0159	0.0612	0.1202	0.0800	1.503	|***	0.661	0.3691	1.1945
15	7.6768	7.7025	0.0253	-0.0257	0.0975	-0.264	|	0.002	0.0629	-0.0665
16	8.4411	8.3905	0.0418	0.0506	0.0916	0.553	|*	0.032	0.1726	0.2473
17	7.4072	7.4857	0.0229	-0.0785	0.0981	-0.800	*|	0.017	0.0515	-0.1845
18	7.9695	8.0330	0.0321	-0.0635	0.0954	-0.665	*|	0.025	0.1016	-0.2201
19	6.7552	6.8176	0.0296	-0.0624	0.0963	-0.648	*|	0.020	0.0862	-0.1958
20	7.5167	7.5039	0.0230	0.0128	0.0980	0.130	|	0.000	0.0521	0.0297

It is show in the question 3.1 that n-k is 18 and

s is Std Error from the above table and c is given by the HATDiagH value of the above table. So the interval is:

$$7.5039 - t(18)_{0.975} \times \sqrt{0.0230^2 \cdot 0.0521}, 7.5039 + t(18)_{0.975} \times \sqrt{0.0230^2 \cdot 0.0521}$$

And the length of this interval:

$$2 \cdot t(18)_{0.975}\sqrt{0.0230^2 \cdot 0.0521}$$

The answer is 4.

Question 3.3-5

We now remove each of the observations one at a time. The observation that - when removed- will cause the numerically least change in its predicted value is no:

Low DFFITS value indicates that omitting that observation would not significantly affect the predicted values. We consider the absolute value of DFFITS and Cooks D when evaluating the impact of an observation on the predicted values. From the table observation 20 has lowest DFFIT value.

Cook's D

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

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

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 $\pm 2\sqrt{p/(n-p)}$.

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

Output Statistics										
Obs	**Dep. Variable**	**Pred. Value**	**Std Err. Mean Predict**	**Residual**	**Std Err. Residual**	**Student Resi-dual**	-2-1 0 1 2	**Cook's D**	**Hat Diag H**	**DFFITS**
1	6.9206	6.7342	0.0315	0.1864	0.0956	1.949	|***	0.206	0.0979	0.7023
2	7.4827	7.4960	0.0229	-0.0133	0.0981	-0.135	|	0.001	0.0518	-0.0308
3	7.5165	7.6248	0.0242	-0.1084	0.0978	-1.108	**|	0.038	0.0576	-0.2758
4	7.5384	7.6478	0.0245	-0.1094	0.0977	-1.120	**|	0.039	0.0590	-0.2825
5	7.8167	7.9019	0.0290	-0.0852	0.0964	-0.883	*|	0.035	0.0832	-0.2645
6	6.5770	6.6671	0.0332	-0.0901	0.0951	-0.948	*|	0.055	0.1085	-0.3297
7	6.4485	6.5542	0.0361	-0.1057	0.0940	-1.125	**|	0.094	0.1288	-0.4360
8	6.6869	6.7706	0.0306	-0.0837	0.0959	-0.872	*|	0.039	0.0926	-0.2767
9	6.6002	6.4097	0.0402	0.1904	0.0923	2.062	|****	0.402	0.1591	0.9975
10	7.0512	7.0293	0.0254	0.0219	0.0974	0.225	|	0.002	0.0638	0.0572
11	6.7760	6.7285	0.0316	0.0476	0.0956	0.498	|	0.014	0.0988	0.1612
12	7.4753	7.3886	0.0225	0.0868	0.0981	0.884	|*	0.021	0.0500	0.2016
13	7.6346	7.5255	0.0231	0.1091	0.0980	1.113	|**	0.035	0.0528	0.2647
14	9.1361	9.0159	0.0612	0.1202	0.0800	1.503	|***	0.661	0.3691	1.1945
15	7.6768	7.7025	0.0253	-0.0257	0.0975	-0.264	|	0.002	0.0629	-0.0665
16	8.4411	8.3905	0.0418	0.0506	0.0916	0.553	|*	0.032	0.1726	0.2473
17	7.4072	7.4857	0.0229	-0.0785	0.0981	-0.800	*|	0.017	0.0515	-0.1845
18	7.9695	8.0330	0.0321	-0.0635	0.0954	-0.665	*|	0.025	0.1016	-0.2201
19	6.7552	6.8176	0.0296	-0.0624	0.0963	-0.648	*|	0.020	0.0862	-0.1958
20	7.5167	7.5039	0.0230	0.0128	0.0980	0.130	|	0.000	0.052	0.0297

The answer is 5.

Question 3.4-1

The number of observations that have as well an extreme RStudent residual as an extreme (i.e. large) leverage is:

Output Statistics										
Obs	Dep. Variable	Pred. Value	Std Err. Mean Predict	Residual	Std Err. Residual	Student Resi-dual	-2-1 0 1 2	Cook's D	Hat Diag H	DFFITS
1	6.9206	6.7342	0.0315	0.1864	0.0956	1.949	|***	0.206	0.0979	0.7023
2	7.4827	7.4960	0.0229	-0.0133	0.0981	-0.135	|	0.001	0.0518	-0.0308
3	7.5165	7.6248	0.0242	-0.1084	0.0978	-1.108	**|	0.038	0.0576	-0.2758
4	7.5384	7.6478	0.0245	-0.1094	0.0977	-1.120	**|	0.039	0.0590	-0.2825
5	7.8167	7.9019	0.0290	-0.0852	0.0964	-0.883	*|	0.035	0.0832	-0.2645
6	6.5770	6.6671	0.0332	-0.0901	0.0951	-0.948	*|	0.055	0.1085	-0.3297
7	6.4485	6.5542	0.0361	-0.1057	0.0940	-1.125	**|	0.094	0.1288	-0.4360
8	6.6869	6.7706	0.0306	-0.0837	0.0959	-0.872	*|	0.039	0.0926	-0.2767
9	6.6002	6.4097	0.0402	0.1904	0.0923	2.062	|****	0.402	0.1591	0.9975
10	7.0512	7.0293	0.0254	0.0219	0.0974	0.225	|	0.002	0.0638	0.0572
11	6.7760	6.7285	0.0316	0.0476	0.0956	0.498	|	0.014	0.0988	0.1612
12	7.4753	7.3886	0.0225	0.0868	0.0981	0.884	|*	0.021	0.0500	0.2016
13	7.6346	7.5255	0.0231	0.1091	0.0980	1.113	|**	0.035	0.0528	0.2647
14	9.1361	9.0159	0.0612	0.1202	0.0800	1.503	|***	0.661	0.3691	1.1945
15	7.6768	7.7025	0.0253	-0.0257	0.0975	-0.264	|	0.002	0.0629	-0.0665
16	8.4411	8.3905	0.0418	0.0506	0.0916	0.553	|*	0.032	0.1726	0.2473
17	7.4072	7.4857	0.0229	-0.0785	0.0981	-0.800	*|	0.017	0.0515	-0.1845
18	7.9695	8.0330	0.0321	-0.0635	0.0954	-0.665	*|	0.025	0.1016	-0.2201
19	6.7552	6.8176	0.0296	-0.0624	0.0963	-0.648	*|	0.020	0.0862	-0.1958
20	7.5167	7.5039	0.0230	0.0128	0.0980	0.130	|	0.000	0.0521	0.0297

We can see that obs 9 with extreme RStudent residual equal to 2.062 has not extreme leverage, same for obs 1. And the obs 14 with high Leverage does not have extreme RStudent Residual.

So the answer is 0 observations

The answer is 1.

Question 3.5-3

Student Residual describes the standardized form of the residuals. It's a measure of how much each data point's observed value differs from the predicted value, taking into account the estimated standard error of the residuals.

Output Statistics										
Obs	Dep. Variable	Pred. Value	Std Err. Mean Predict	Residual	Std Err. Residual	Student Resi-dual	-2-1 0 1 2	Cook's D	Hat Diag H	DFFITS
1	6.9206	6.7342	0.0315	0.1864	0.0956	1.949	|***	0.206	0.0979	0.7023
2	7.4827	7.4960	0.0229	-0.0133	0.0981	-0.135	|	0.001	0.0518	-0.0308
3	7.5165	7.6248	0.0242	-0.1084	0.0978	-1.108	**|	0.038	0.0576	-0.2758
4	7.5384	7.6478	0.0245	-0.1094	0.0977	-1.120	**|	0.039	0.0590	-0.2825
5	7.8167	7.9019	0.0290	-0.0852	0.0964	-0.883	*|	0.035	0.0832	-0.2645
6	6.5770	6.6671	0.0332	-0.0901	0.0951	-0.948	*|	0.055	0.1085	-0.3297
7	6.4485	6.5542	0.0361	-0.1057	0.0940	-1.125	**|	0.094	0.1288	-0.4360
8	6.6869	6.7706	0.0306	-0.0837	0.0959	-0.872	*|	0.039	0.0926	-0.2767
9	6.6002	6.4097	0.0402	0.1904	0.0923	2.062	|****	0.402	0.1591	0.9975
10	7.0512	7.0293	0.0254	0.0219	0.0974	0.225	|	0.002	0.0638	0.0572
11	6.7760	6.7285	0.0316	0.0476	0.0956	0.498	|	0.014	0.0988	0.1612
12	7.4753	7.3886	0.0225	0.0868	0.0981	0.884	|*	0.021	0.0500	0.2016
13	7.6346	7.5255	0.0231	0.1091	0.0980	1.113	|**	0.035	0.0528	0.2647
14	9.1361	9.0159	0.0612	0.1202	0.0800	1.503	|***	0.661	0.3691	1.1945
15	7.6768	7.7025	0.0253	-0.0257	0.0975	-0.264	|	0.002	0.0629	-0.0665
16	8.4411	8.3905	0.0418	0.0506	0.0916	0.553	|*	0.032	0.1726	0.2473
17	7.4072	7.4857	0.0229	-0.0785	0.0981	-0.800	*|	0.017	0.0515	-0.1845
18	7.9695	8.0330	0.0321	-0.0635	0.0954	-0.665	*|	0.025	0.1016	-0.2201
19	6.7552	6.8176	0.0296	-0.0624	0.0963	-0.648	*|	0.020	0.0862	-0.1958
20	7.5167	7.5039	0.0230	0.0128	0.0980	0.130	|	0.000	0.0521	0.0297

So for the largest value of Student Residual, we have the worst prediction of lny.

This observation is observation 9.

The answer is 3.

Problem 4

Enclosure B belongs to this problem. The data are measurements of the average number of miles per gallon gasoline a number of cars could drive under controlled circumstances. The original data source is R.M. Heavenrich, J.D. Murrell, and K.H. Hellman, Light Duty Automotive Technology and Fuel Economy Trends Through 1991, U.S. Environmental Protection Agency, 1991 (EPA/AA/CTAB/91-02). The number of cases is 82 and the variables measured were:

1. vol: Cubic feet of cabin space

2. hp: Engine horsepower

3. mpg: Average miles per gallon

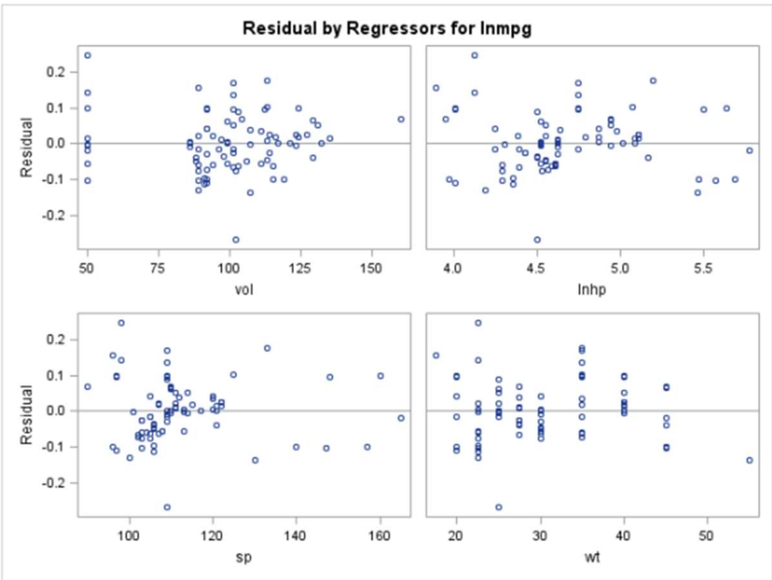
4. sp: Top speed (mph)

5. wt: Vehicle weight (100 lb)

The fuel consumption (mpg) is the independent variable, and in order to improve linearity we have taken the logarithm of mpg. Furthermore, we have added the logarithm and the square of the independent variables to the data set. Firstly, we consider the model that contains ‘vol lnhp sp wt’.

Question 4.1-5

__The diagnostic plots show that one might consider adding two quadratic terms to the model. The variables whose squares should be added are:__

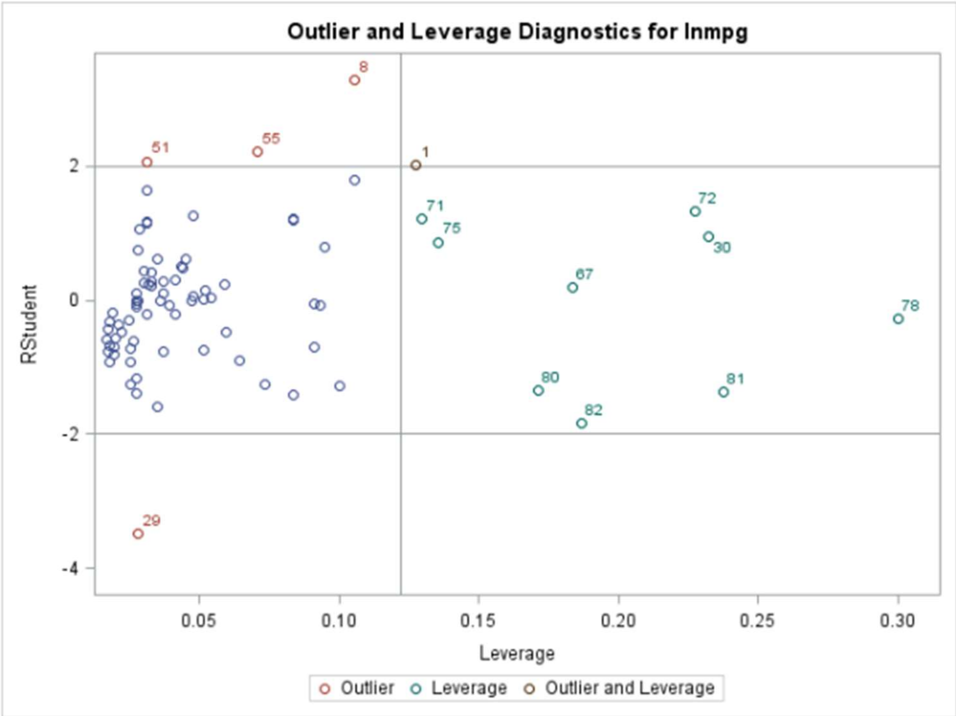


For lnhp and wt there are some patterns, some systematic structures for the residuals in the diagnostic plots. This indicates that a more complex model might need. So, adding quadratic terms for these two predictors could be a solution.

The answer is 5.

Question 4.2-2

__How many residuals (RStudent) fall outside the usual [-2, 2] interval:__



From the Plot it is obvious that 5 residuals fall outside the interval (Number of red observations: 51,55,8,1,29).

The answer is 5.

Question 4.3-5

We now consider the model that contains 'wt lnhp sqsp sqhp' and we want to make a single test in order to see whether we may assume that the coefficients to sqsp and sqhp can be assumed to be simultaneously equal to 0. The usual test statistic is:

The usual test statistic is given by :

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

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

And the values of each terms can be found in the tables below:

For the initial model:

And for final model:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	6.94151	3.47076	487.54	<.0001
Error	79	0.56239	0.00712		
Corrected Total	81	7.50390			

$$SSres(Mod) = 0.44873, \qquad SSres(Hyp) = 0.56239$$

$$and\ DFres(Mod) = 77\ and\ DFres(Hyp) = 79$$

$$\frac{0.56238 - 0.44873/_{79-77}}{0.44873/_{77}}$$

The answer is 5.

Question 4.4-5

__The distribution under the null hypothesis of the statistic above is:__

It is shown to the above question that k-r=2 and n-k=77 where

k is the rank of the full model (H1)

r is the rank of the simpler model (H0)

n is the number of observations.

From theorem 2.21 F(2,77)

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

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

The answer is 5.

Problem 5

Enclosure C with SAS program and SAS output belongs to this problem. The measurements are estimated correlations between 176 observed values of different properties of offspring of alfalfa (lucerne). The data is described in Example 5.4, p. 238 in the course book Ersbøll & Conradsen (2012). The variables measured are also given in the table below.

Variable no. & name	Unit of measure	Explanation
x1: Type of growth	Grade 1 - 9	1 = growth is lying down, 9 = growth is upright
x2: Regrowth after winter	Grade 1 - 9	1 = worst, 9 = best
x3: Ability to creep	Grade 1 - 9	1 = no runners, 9 = most runners
x4: Activity	Grade 1 - 9	1 = weakest, 9 = strongest
x5: Time of blooming	Grade 1 - 9	1 = latest blooming, 9 = earliest blooming
x6: Plant height	cm	
x7: Seed weight	g per plant	
x8: Plant weight	g per plant after drying	
x9: Percent seed	%	Calculated per plant by means of (7) and (8)

Question 5.1-3

How many principal components must we include if we want to explain at least 80% of the total variation?

From the Table given below, we can find how many principal components we need to include if we want at least 80% of the total variation, we look at the cumulative fraction of the eigenvalues (the sum of the proportions).

Eigenvalues of the Correlation Matrix: Total = 9 Average = 1				
	Eigenvalue	Difference	Proportion	Cumulative
1	3.12713262	0.78562655	0.3475	0.3475
2	2.34150607	1.12443569	0.2602	0.6076
3	1.21707038	0.37454816	0.1352	0.7429
4	0.84252222	0.15878292	0.0936	0.8365
5	0.68373929	0.38994087	0.0760	0.9124
6	0.29379842	0.04096571	0.0326	0.9451
7	0.25283270	0.03277307	0.0281	0.9732
8	0.22005963	0.19872097	0.0245	0.9976
9	0.02133867		0.0024	1.0000

We can see that including 4 Principal components the variation is achieved.

The answer is 3.

Question 5.2-1

The degrees of freedom for the usual test statistic for testing the hypothesis that the smallest 3 eigenvalues are equal against all alternatives is:

From Theorem 6.8:

Theorem 6.8

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

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

where

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

and

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

The critical region using a test at level α is approximately

$$\{(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) | z_1 > \chi^2(\frac{1}{2}(k-m+2)(k-m-1))_{1-\alpha}\}.$$

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 α becomes approximately equal to

$$\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n | z_2 > \chi^2(\frac{1}{2}(k-m+2)(k-m-1))_{1-\alpha}\}.$$

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

We can find in the output that the number of the observations is n=176 and k=9 and m=9-3=6 since we want to test the hypothesis that the last three eigenvalues are equal to zero.

The degrees of freedom $\frac{1}{2}(k-m+2)(k-m-1) = \frac{1}{2}(9-6+2)(9-6-1) = 5$

The answer is 1.

Question 5.3-5

We now consider a factor analysis with three factors of the data considered above. Consider the following statements on a plant having a large value of an arbitrary factor.

A. A small plant with an upright growth.

B. A plant that immediately looks healthy: Early blooming, good height and weight, lot of seeds - absolute and relative.

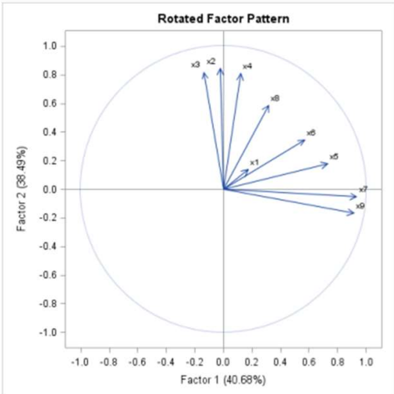
C. A plant with very little seed.

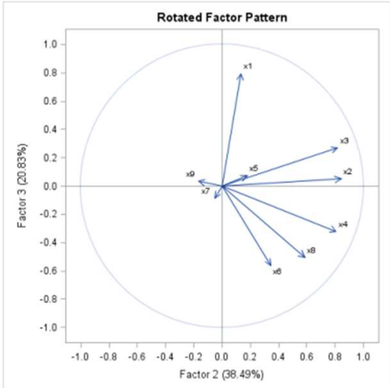
D. A plant with good ‘dynamic’ properties: A heavy plant that grows fine after winter, that has many runners and shows good activity.

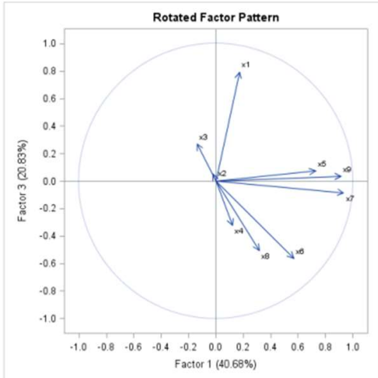
E. A plant that irrespective of whether it is upright or lying down looks good and has good ‘dynamic’ properties.

For (VARIMAX rotated factor 1, VARIMAX rotated factors 2, VARIMAX rotated factors 3) the following characterization is adequate:

From the plots:







Factor 1 describes a plant with high x9, x7, x5 and average x6,x8, which means a lot of seeds which are heavy and early blooming, and are relatively high and heavy after drying. These characteristics match with the statement B: *A plant that immediately looks healthy: Early blooming, good height and weight, lot of seeds - absolute and relative.*

Factor 2 describes a plant with high x2, x3, x4 and x8 which means best regrowth after winter, high ability to creep, strong activity and high plant weight after drying which matches with the statement D: *A plant with good ‘dynamic’ properties: A heavy plant that grows fine after winter, that has many runners and shows good activity*.

Factor 3 describes a plant with growth lying upright but corresponds to negative values for x4, x6 and x8. Which means weak activity, short plant and low weight after drying, matching statement A: A *small plant with an upright growth.*

Thus (VARIMAX rotated factor 1, VARIMAX rotated factors 2, VARIMAX rotated factors 3)=(B,D,A)

The answer is 5.

Question 5.4-4

What fraction of the total variance will be explained by the first VARIMAX rotated factor?

From the results of the varimax rotation method:

Variance Explained by Each Factor		
Factor1	Factor2	Factor3
2.7197196	2.5733320	1.3926576

Be careful that we are asked for the variance explained and not for the fraction of the variance.

The answer is 4.

Question 5.5-4

What fraction of the variation of the first variable x1 is explained by the third VARIMAX rotated factor?

Rotated Factor Pattern			
	Factor1	Factor2	Factor3
x1	0.17187	0.13527	0.79486
x2	-0.02264	0.84248	0.05167
x3	-0.13662	0.81816	0.26946
x4	0.12117	0.80725	-0.32194
x5	0.72812	0.17583	0.07362
x6	0.56859	0.34586	-0.56079
x7	0.93267	-0.05229	-0.08533
x8	0.31511	0.58534	-0.50358
x9	0.91307	-0.16832	0.03382

We are looking for the fraction of the variance. Thus from the above Table we can use the square values to find the answer: 0.79486^2.

The answer is 4.

Problem 6

Enclosure D belongs to this problem. The data are due to Littell, Freund and Spector (1991). They are measurements on the effect of three different weightlifting programs. There were taken measurements at 7 different time points, the same for the three treatments. They were

- RI: The number of repetitions of weightlifting was increased as subjects became stronger.
- WI: The amount of weight was increased as subjects became stronger.
- CONT: Control group with no training.

Since there may be large individual differences between the strength of the subjects participating in the study we have normalized the last two measurements by dividing them with the first measurement yielding the new variables RS6 and RS7. We shall only consider these relative measures of the strength of the subjects.

Question 6.1-2

We now want to compare the three treatments with a multivariate analysis of variance of the relative strengths at times 6 and 7. Then the null hypothesis distribution of the usual test statistic Wilk's Lambda is:

From Theorem 4.21 the usual test statistic:

Theorem 4.21

We consider the above mentioned situation including the assumption of the normality of the observations. Furthermore we consider the hypothesis

$$H_0 : \mathbf{A}\,\boldsymbol{\theta}\,\mathbf{B}^T = \mathbf{C} \quad \text{against} \quad H_1 : \mathbf{A}\,\boldsymbol{\theta}\,\mathbf{B}^T \neq \mathbf{C},$$

where $\mathbf{A}(r \times k)$, $\mathbf{B}(s \times p)$ and $\mathbf{C}(r \times s)$ are given (known) matrices. We introduce

$$\begin{aligned} \boldsymbol{\Delta} &= \mathbf{A}\,\hat{\boldsymbol{\theta}}\,\mathbf{B}^T - \mathbf{C} \\ \mathbf{R} &= n\hat{\boldsymbol{\Sigma}}^* = (\mathbf{Y} - \mathbf{x}\,\hat{\boldsymbol{\theta}})^T(\mathbf{Y} - \mathbf{x}\,\hat{\boldsymbol{\theta}}) = \mathbf{Y}^T\mathbf{Y} - \hat{\boldsymbol{\theta}}^T(\mathbf{x}^T\mathbf{x})\hat{\boldsymbol{\theta}} \end{aligned}$$

and

$$\begin{aligned} \mathbf{E} &= \mathbf{B}\,\mathbf{R}\,\mathbf{B}^T \\ \mathbf{H} &= \boldsymbol{\Delta}^T[\mathbf{A}(\mathbf{x}^T\mathbf{x})^{-1}\mathbf{A}^T]^{-1}\boldsymbol{\Delta}. \end{aligned}$$

The likelihood ratio test for testing H_0 against H_1 is equivalent to the test given by the critical region

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

where $\mathrm{U}(s, r, n-k)_\alpha$ is the α quantile in the null-hypothesis distribution of the test statistic (see below).

We need to find first the dimensions of our parameter matrix. There are 2 independent variables and 3 dependent variables. Thus (3 x 2) and we want to extract the sixth and seventh row for testing:

C has the same dimensions as the parameters extracted $C(r\ x\ s) = C(2\ x\ 2), r = 2\ and\ s = 2$

We have $A(r\ x\ k) = \ A(2\ x\ 3)$ since we are only interested on 6th and 7th row and there are 3 treatments. And $B(s\ x\ p) = B(2\ x\ 3). So\ r = 2, k = 3, s = 2\ and\ p = 3$

The number of observations is $n = 57$

Total Sample Size	57	DF Total	56
Variables	2	DF Within Classes	54
Classes	3	DF Between Classes	2

$$U(s, r, n-k) = U(3{,}2{,}57-3)$$

The answer is 2.

Question 6.2-4

Under the usual assumptions the estimate for the variance of RS6 is equal to:

From the Table below the diagonal elements is the variance of rs6 and rs7 respectively divided to DF within the classes.

E = Error SSCP Matrix		
	rs6	rs7
rs6	0.0349065144	0.0320103992
rs7	0.0320103992	0.0389930628

The answer is 4.

Question 6.3-1

We now only consider the active treatment groups WI and RI, and we want to test whether the means of [RS6 RS7]' are the same for the two treatment groups. The usual test statistic becomes:

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

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

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

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*** π_i as

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

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

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

Since we are now interested in only two treatment groups we need to take into account the two model, $n = 16$ *and* $m = 21$, $p = 2$ and $(\bar{X} - \bar{Y})^T S^{-1}(\bar{X} - \bar{Y})$ is the generalized squared distance.

Class Level Information					
treatment	Variable Name	Frequency	Weight	Proportion	Prior Probability
ri	ri	16	16.0000	0.432432	0.500000
wi	wi	21	21.0000	0.567568	0.500000

Generalized Squared Distance to treatment		
From treatment	ri	wi
ri	0	0.16700
wi	0.16700	0

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

By substitution:

$$\frac{16+2-2-1}{(16+21-2)2}\frac{16\cdot 21}{16+21}0.16700$$

The answer is 1.

Question 6.4-3

The distribution under the null hypothesis of the test statistic above is:

The distribution under the null hypothesis is given by the Theorem 4.9:

$$F(p, n+m-p-1) = F(2{,}16+21-2-1)$$

The answer is 3.

Question 6.5-1

We are now interested in using the values of $[RS6\ RS7]'$ in distinguishing between the two training programs with a decision function of the form $[RS6\ RS7]'\begin{bmatrix} a \\ b \end{bmatrix} > c$. Rounded to integers the values of a, b and c should be chosen as

From Theorem 5.4

Theorem 5.4

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

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

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

From the Proc discrim analysis on the relative measurements for the two treatments WI and RI we have:

Linear Discriminant Function for treatment		
Variable	**ri**	**wi**
Constant	-748.96607	-754.69084
rs6	1125	1098
rs7	346.41638	378.60181

$[RS6\ RS7]'\begin{bmatrix} 1125-1098 \\ 346.42-378.60 \end{bmatrix} + (-748.97-(-754.69)) > log c$

For equal losses and prior $log c = log 1 = 0$

The above equation becomes:

$[RS6\ RS7]'\begin{bmatrix} 27 \\ -32 \end{bmatrix} + (-6) > 0 \Rightarrow [RS6\ RS7]'\begin{bmatrix} 27 \\ -32 \end{bmatrix} > 6$

So with substitution $a = 27, b = -32\ and\ c = 6$