35459 Multivariate Statistics

Week 6 Seminar - MANOVA and Multivariate Regression

Two-way Multivariate ANOVA

Consider the Plastic film data in Example 6.13 of Johnson and Wichern. In this experiment, two factors were tested (Change in the rate of extrusion, amount of additive) and three response variables were measured (tear resistance, gloss, opacity).

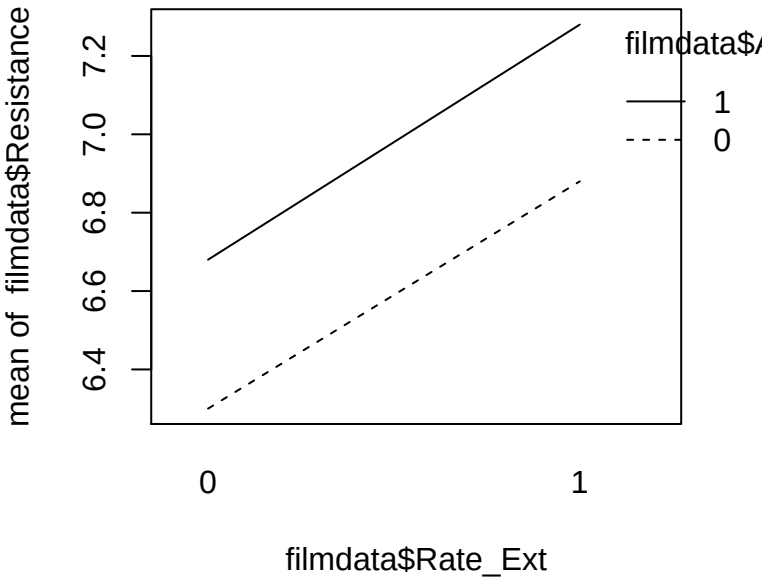
The model that we are fitting is

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\beta}_j + \boldsymbol{\tau\beta}_{ij} + \boldsymbol{\epsilon}_{ij}$$

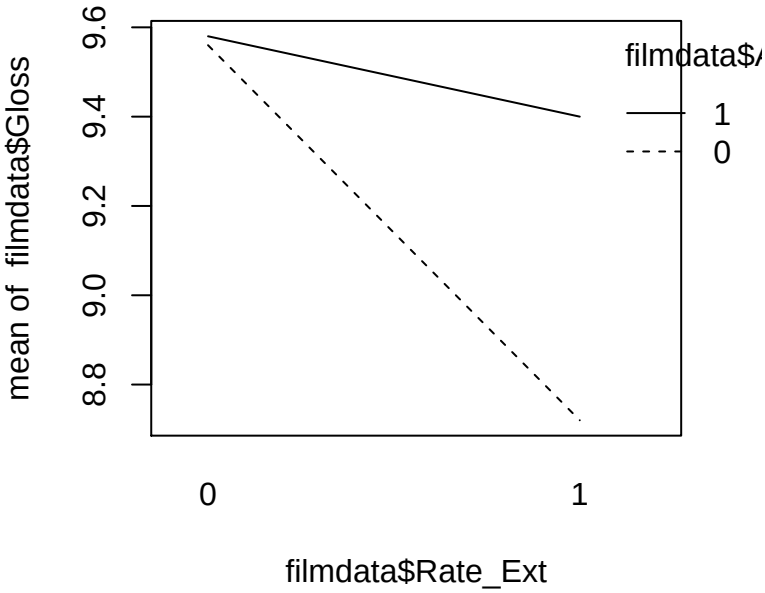
where $i = 1, 2$ and $j = 1, 2$

```
library(MVN)
library(VGAM)
library(car)
```

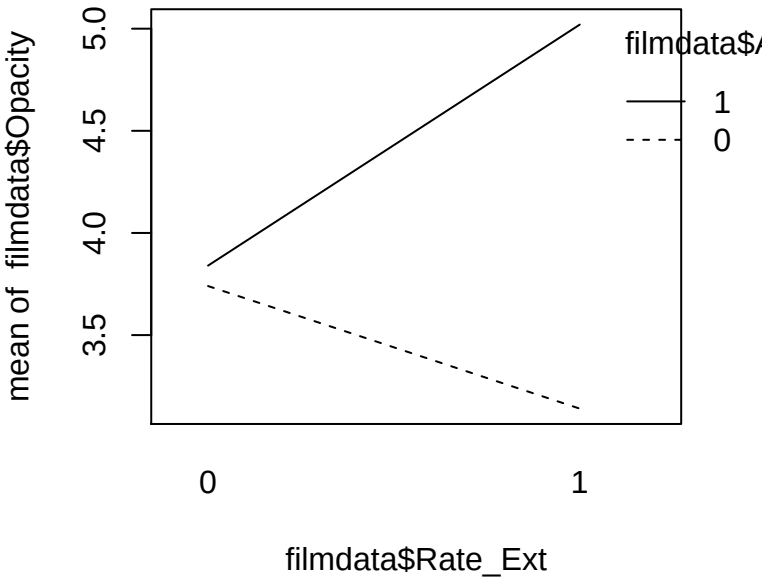
```
filmdata<-read.csv("C:/Documents/Film.csv")

filmdata$Rate_Ext<-as.factor(filmdata$Rate_Ext)
filmdata$Additive<-as.factor(filmdata$Additive)

interaction.plot(filmdata$Rate_Ext,filmdata$Additive,filmdata$Resistance)
```




```
interaction.plot(filmdata$Rate_Ext,filmdata$Additive,filmdata$Gloss)
```




```
interaction.plot(filmdata$Rate_Ext,filmdata$Additive,filmdata$Opacity)
```




```
filmmodel<-manova(cbind(Resistance,Gloss,Opacity)~Rate_Ext*Additive,data=filmdata)
print(filmmodel)

## Call:
##    manova(cbind(Resistance, Gloss, Opacity) ~ Rate_Ext * Additive,
##     data = filmdata)
##
## Terms:
##                  Rate_Ext Additive Rate_Ext:Additive Residuals
## resp 1               1.74     0.76              0.00      1.76
## resp 2               1.30     0.61              0.54      2.63
## resp 3               0.42     4.90              3.96     64.92
## Deg. of Freedom         1        1                 1        16
##
## Residual standard errors: 0.332 0.4053 2.014
## Estimated effects may be unbalanced

summary.aov(filmmodel)

##  Response Resistance :
##                    Df Sum Sq Mean Sq F value Pr(>F)
## Rate_Ext            1  1.740   1.740    15.8 0.0011 **
## Additive            1  0.760   0.760     6.9 0.0183 *
## Rate_Ext:Additive   1  0.001   0.001     0.0 0.9471
## Residuals          16  1.764   0.110
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response Gloss :
##                   Df Sum Sq Mean Sq F value Pr(>F)
## Rate_Ext           1  1.301   1.301    7.92  0.012 *
## Additive           1  0.612   0.612    3.73  0.071 .
## Rate_Ext:Additive  1  0.544   0.544    3.32  0.087 .
## Residuals         16  2.628   0.164
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response Opacity :
##                   Df Sum Sq Mean Sq F value Pr(>F)
## Rate_Ext           1    0.4    0.42    0.10   0.75
## Additive           1    4.9    4.90    1.21   0.29
## Rate_Ext:Additive  1    4.0    3.96    0.98   0.34
## Residuals         16   64.9    4.06
```

```
summary(filmmodel,intercept=TRUE,test="Wilks")
```

```
##                   Df Wilks approx F num Df den Df Pr(>F)
## (Intercept)        1 0.001     5951      3     14 <2e-16 ***
## Rate_Ext           1 0.382        8      3     14  0.003 **
## Additive           1 0.523        4      3     14  0.025 *
## Rate_Ext:Additive  1 0.777        1      3     14  0.302
## Residuals         16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(filmmodel,intercept=TRUE, test="Roy")
```

```
##                   Df  Roy approx F num Df den Df Pr(>F)
## (Intercept)        1 1275     5951      3     14 <2e-16 ***
## Rate_Ext           1    2        8      3     14  0.003 **
## Additive           1    1        4      3     14  0.025 *
## Rate_Ext:Additive  1    0        1      3     14  0.302
## Residuals         16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(filmmodel,intercept=TRUE, test="Hotelling-Lawley")
```

```
##                   Df Hotelling-Lawley approx F num Df den Df Pr(>F)
## (Intercept)        1             1275     5951      3     14 <2e-16 ***
## Rate_Ext           1                2        8      3     14  0.003 **
```

```
## Additive           1                 1        4      3     14  0.025 *
## Rate_Ext:Additive  1                 0        1      3     14  0.302
## Residuals         16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fitIII <- lm(cbind(Resistance,Gloss,Opacity) ~ Rate_Ext*Additive,
             data=filmdata, contrasts=list(Rate_Ext=contr.sum,
                           Additive=contr.sum))
filmmodel <- Manova(fitIII, type="III")
summary(filmmodel, multivariate=TRUE)
```

```
##
## Type III MANOVA Tests:
##
## Sum of squares and products for error:
##            Resistance  Gloss Opacity
## Resistance      1.764  0.020  -3.070
## Gloss           0.020  2.628  -0.552
## Opacity        -3.070 -0.552  64.924
##
## ----------------------------------------
##
## Term: (Intercept)
##
## Sum of squares and products for the hypothesis:
##            Resistance  Gloss Opacity
## Resistance      920.7 1264.0   534.0
## Gloss          1264.0 1735.4   733.1
## Opacity         534.0  733.1   309.7
##
## Multivariate Tests: (Intercept)
##                  Df test stat approx F num Df den Df Pr(>F)
## Pillai            1         1     5951      3     14 <2e-16 ***
## Wilks             1         0     5951      3     14 <2e-16 ***
## Hotelling-Lawley  1      1275     5951      3     14 <2e-16 ***
## Roy               1      1275     5951      3     14 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ----------------------------------------
##
## Term: Rate_Ext
##
## Sum of squares and products for the hypothesis:
```

```
##            Resistance   Gloss Opacity
## Resistance     1.7405 -1.5045  0.8555
## Gloss         -1.5045  1.3005 -0.7395
## Opacity        0.8555 -0.7395  0.4205
##
## Multivariate Tests: Rate_Ext
##                  Df test stat approx F num Df den Df  Pr(>F)
## Pillai            1    0.6181    7.554      3     14 0.00303 **
## Wilks             1    0.3819    7.554      3     14 0.00303 **
## Hotelling-Lawley  1    1.6188    7.554      3     14 0.00303 **
## Roy               1    1.6188    7.554      3     14 0.00303 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ----------------------------------------
##
## Term: Additive
##
## Sum of squares and products for the hypothesis:
##            Resistance  Gloss Opacity
## Resistance     0.7605 0.6825   1.931
## Gloss          0.6825 0.6125   1.732
## Opacity        1.9305 1.7325   4.901
##
## Multivariate Tests: Additive
##                  Df test stat approx F num Df den Df Pr(>F)
## Pillai            1    0.4770    4.256      3     14 0.0247 *
## Wilks             1    0.5230    4.256      3     14 0.0247 *
## Hotelling-Lawley  1    0.9119    4.256      3     14 0.0247 *
## Roy               1    0.9119    4.256      3     14 0.0247 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## ----------------------------------------
##
## Term: Rate_Ext:Additive
##
## Sum of squares and products for the hypothesis:
##            Resistance  Gloss Opacity
## Resistance     0.0005 0.0165  0.0445
## Gloss          0.0165 0.5445  1.4685
## Opacity        0.0445 1.4685  3.9605
##
## Multivariate Tests: Rate_Ext:Additive
```

```
##                   Df test stat approx F num Df den Df Pr(>F)
## Pillai             1    0.2229    1.339      3     14  0.302
## Wilks              1    0.7771    1.339      3     14  0.302
## Hotelling-Lawley   1    0.2868    1.339      3     14  0.302
## Roy                1    0.2868    1.339      3     14  0.302
```

Exercise

An experiment was conducted for comparing 2 methods (1 and 2) of teaching shorthand computer programing to 60 students. Students were allocated to three groups, whose 12 hours of instruction were broken into different lengths (strategy)

1. 2 hours of instruction/day for 6 weeks
2. 3 hours of instruction/day for 4 weeks
3. 4 hours of instruction/day for 3 weeks

Two response variables were measured, Y_1=speed, and Y_2=accuracy. The data are in the file Shorthand.csv

Perform a two–way MANOVA to determine whether the methods and strayegies had any effect on speed or accuracy.

Multivariate Regression

Amitirptyline is prescribed by some physicians as an antidepressant. However, there are also conjectured side effects that seem to be related to the use of the drug, such as irregular heartbeat, abmormal blood pressure and irregular waves on and ECG. Data gathered from 17 patients who were asmitted to the hospital after an amitriptyline overdose had the following respnse variables measured

- Total TCAD plasma level (y_1)
- Amount of amitriptyline present in TCAD plasma level (y_2)

and the following predictive variables measured

- Gender (1 if male, 0 if female) (x_1)
- Amount of antidepressants taken at time of overdose (x_2)
- PR wave measurement (x_3)

- Diastolic blood pressure (x_4)
- QRS wave measurement (x_5)

The model that we will fit is

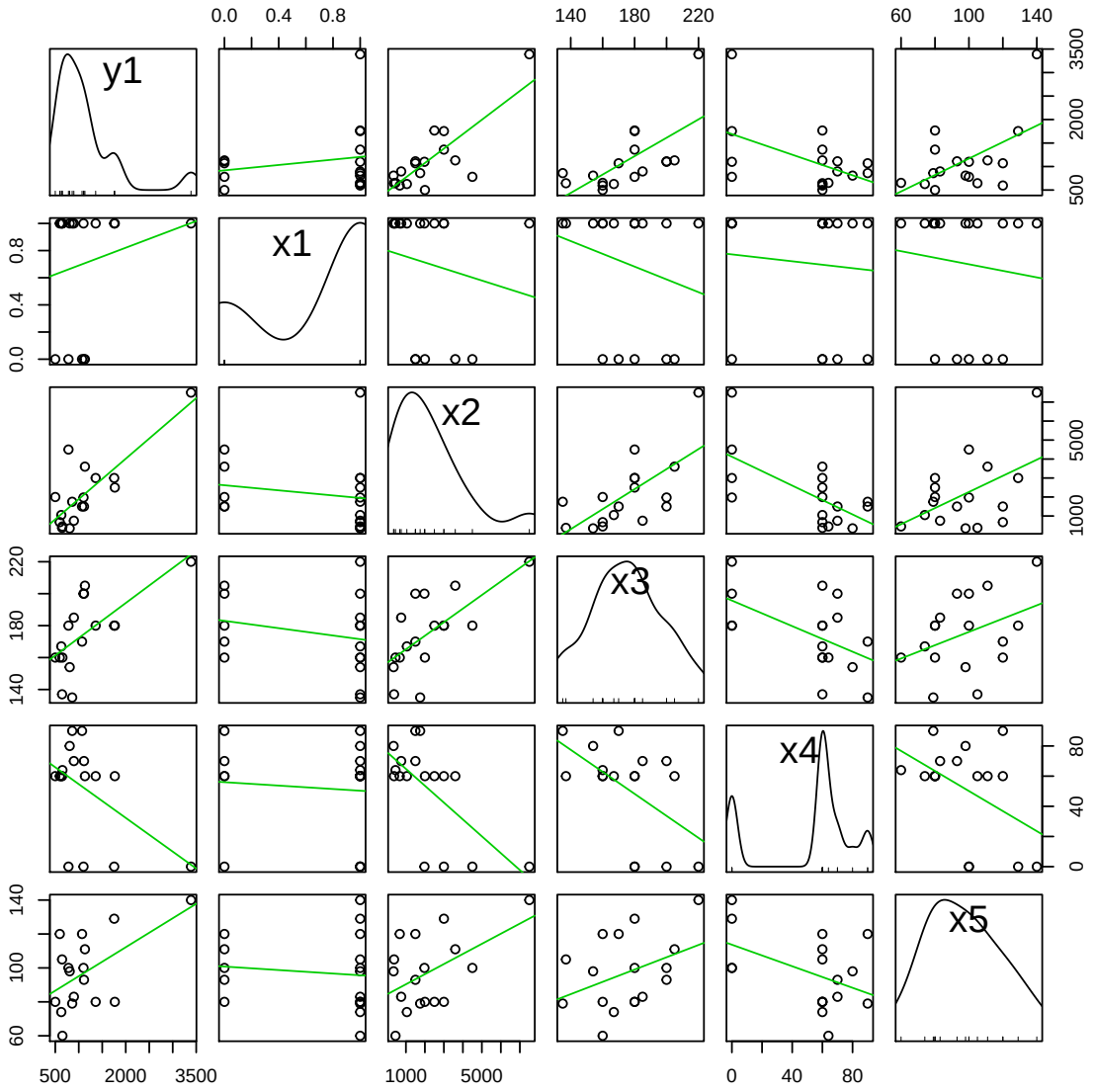
$$\boldsymbol{Y} = \boldsymbol{\beta}^T \boldsymbol{X} + \boldsymbol{\epsilon}$$

which can be written as

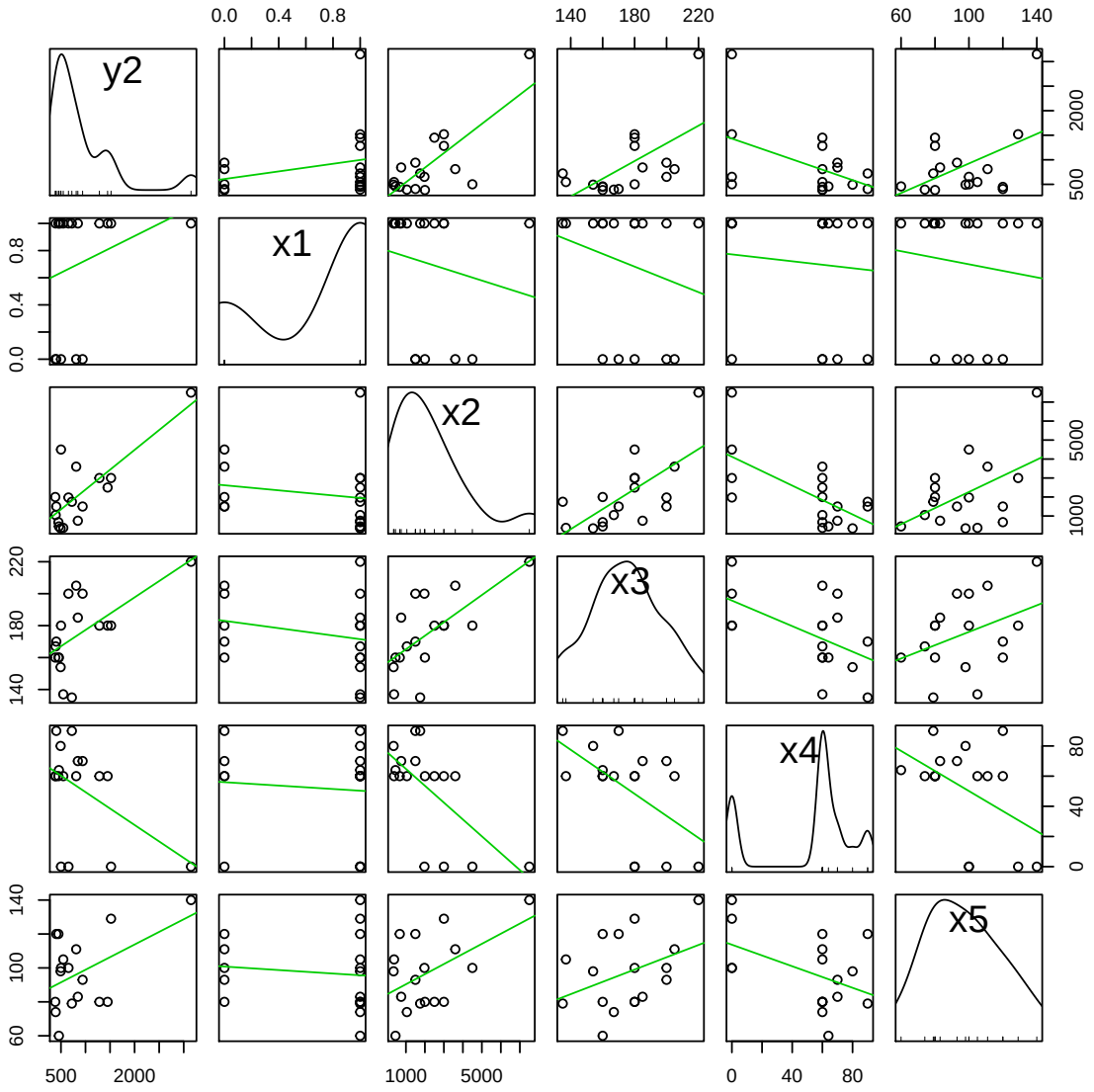
$$\begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix} = \begin{bmatrix} \beta_{01} & \beta_{11} & \beta_{21} & \beta_{31} & \beta_{41} & \beta_{51} \\ \beta_{02} & \beta_{12} & \beta_{22} & \beta_{32} & \beta_{42} & \beta_{52} \end{bmatrix} \times \begin{bmatrix} 1 \\ x_{1i} \\ x_{2i} \\ x_{3i} \\ x_{4i} \\ x_{5i} \end{bmatrix} + \begin{bmatrix} e_{1i} \\ e_{2i} \end{bmatrix}$$

In general, there are m response variables, r parameters per response variable, and n observations. Here, $m = 2$, $p = 6$, and $n = 17$.

```
y1<-c(3389, 1101, 1131, 596, 896, 1767, 807, 1111, 645, 628, 1360,
      652, 860, 500, 781, 1070, 1754)
y2<-c(3149, 653, 810, 448, 844, 1450, 493, 941, 547, 392, 1283, 458,
      722, 384, 501, 405, 1520)
x1<-c(1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1)
x2<-c(7500, 1975, 3600, 675, 750, 2500, 350, 1500, 375, 1050, 3000,
      450, 1750, 2000, 4500, 1500, 3000)
x3<-c(220, 200, 205, 160, 185, 180, 154, 200, 137, 167, 180, 160, 135,
      160, 180, 170, 180)
x4<-c(0, 0, 60, 60, 70, 60, 80, 70, 60, 60, 60, 64, 90, 60, 0, 90, 0)
x5<-c(140, 100, 111, 120, 83, 80, 98, 93, 105, 74, 80, 60, 79, 80, 100,
      120, 129)

scatterplotMatrix(cbind(y1,x1,x2,x3,x4,x5),smoother=FALSE)
```




```
scatterplotMatrix(cbind(y2,x1,x2,x3,x4,x5),smoother=FALSE)
```

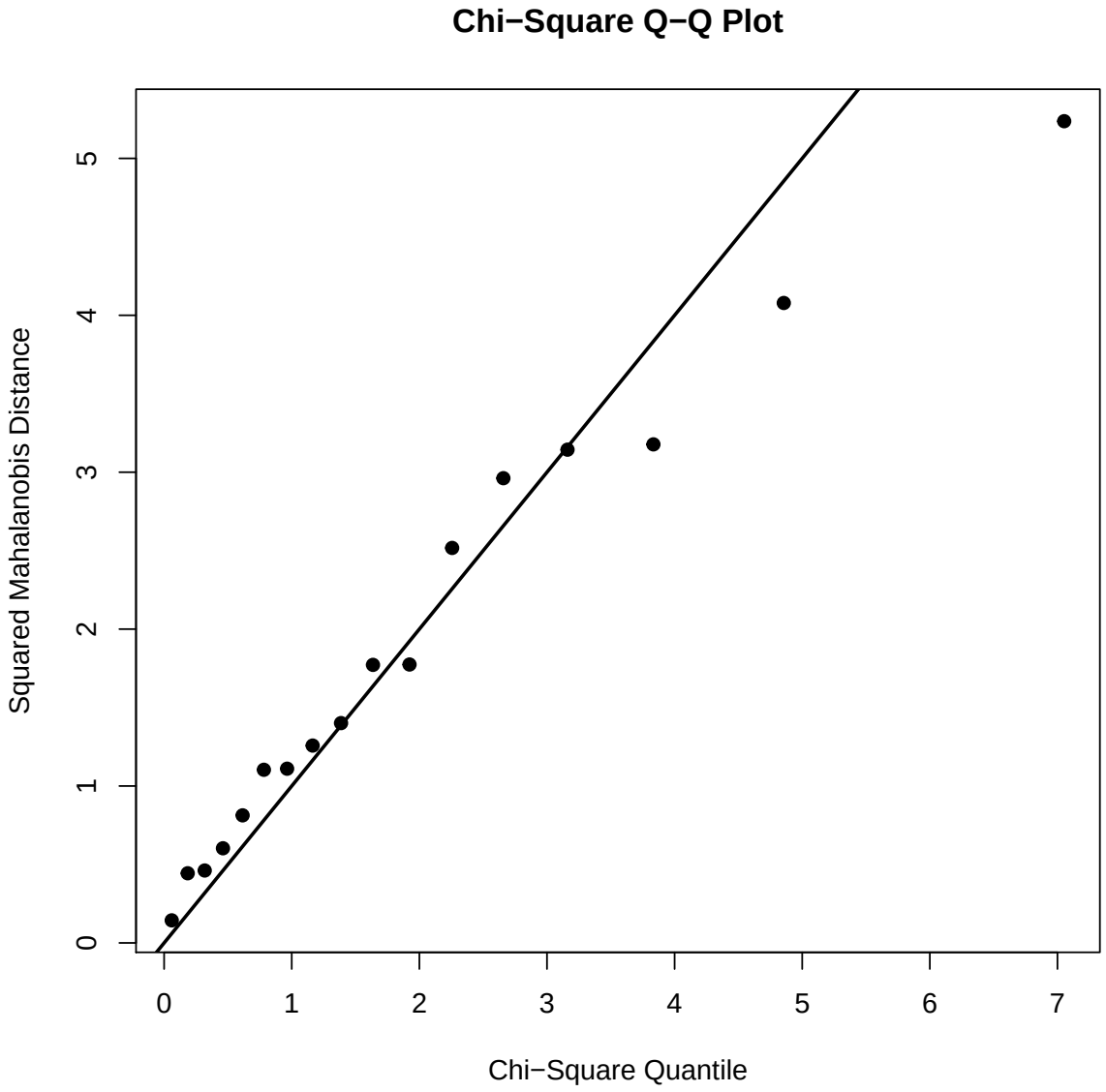



```
ami_model<-lm(cbind(y1,y2)~x1+x2+x3+x4+x5)
summary(ami_model)

## Response y1 :
##
## Call:
## lm(formula = y1 ~ x1 + x2 + x3 + x4 + x5)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -399.2 -180.1    4.5  164.1  366.8
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.88e+03   8.93e+02   -3.22  0.00811 **
## x1           6.76e+02   1.62e+02    4.17  0.00156 **
## x2           2.85e-01   6.09e-02    4.68  0.00068 ***
## x3           1.03e+01   4.25e+00    2.41  0.03436 *
## x4           7.25e+00   3.23e+00    2.25  0.04603 *
## x5           7.60e+00   3.85e+00    1.97  0.07401 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 281 on 11 degrees of freedom
## Multiple R-squared:  0.887,Adjusted R-squared:  0.836
## F-statistic: 17.3 on 5 and 11 DF,  p-value: 6.98e-05
##
##
## Response y2 :
##
## Call:
## lm(formula = y2 ~ x1 + x2 + x3 + x4 + x5)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -373.9 -247.3  -83.7  217.1  462.7
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.73e+03   9.29e+02   -2.94  0.01350 *
## x1           7.63e+02   1.69e+02    4.53  0.00086 ***
## x2           3.06e-01   6.33e-02    4.84  0.00052 ***
## x3           8.90e+00   4.42e+00    2.01  0.06952 .
## x4           7.21e+00   3.35e+00    2.15  0.05478 .
## x5           4.99e+00   4.00e+00    1.25  0.23862
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 292 on 11 degrees of freedom
## Multiple R-squared:  0.876,Adjusted R-squared:  0.82
## F-statistic: 15.6 on 5 and 11 DF,  p-value: 0.000113

Manova(ami_model)

##
## Type II MANOVA Tests: Pillai test statistic
##    Df test stat approx F num Df den Df Pr(>F)
## x1  1     0.655     9.50      2     10 0.0049 **
```

```
## x2  1     0.691    11.18      2     10 0.0028 **
## x3  1     0.346     2.65      2     10 0.1192
## x4  1     0.324     2.39      2     10 0.1414
## x5  1     0.292     2.06      2     10 0.1781
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
roystonTest(residuals(ami_model),qqplot=TRUE)
```




```
##   Royston's Multivariate Normality Test
## ---------------------------------------------
##   data : residuals(ami_model)
##
```

```
##   H       : 0.7135
##   p-value : 0.6317
##
##   Result  : Data is multivariate normal.
## ------------------------------------------
```

The $(1-\alpha)\%$ simultaneous prediction interval for Y_{0i} are

$$\boldsymbol{z}_0^T\widehat{\boldsymbol{\beta}_{(i)}} \pm \sqrt{\frac{m(n-r-1)}{n-r-m} \times F_{m,n-r-m}(\alpha)} \times \sqrt{(1+\boldsymbol{z}_0^T(\boldsymbol{Z}^T\boldsymbol{Z})^{-1}\boldsymbol{z}_0) \times \frac{n\widehat{\sigma_{ii}}}{n-r-1}}$$

```
newdata<-data.frame(x1=1,x2=1200,x3=140,x4=70,x5=85)
predict(ami_model,newdata,interval="prediction")

##      y1    y2
## 1 729.5 575.7

beta<-coefficients(ami_model)
z0<-matrix(c(1,1,1200,140,70,85),ncol=1)
Zmat<-cbind(c(rep(1,17)),x1,x2,x3,x4,x5)
sigmat<-cov(residuals(ami_model))
m<-2
n<-17
r<-6

for(i in 1:2){
lower<-t(z0)%*%beta[,i]-sqrt((m*(n-r-1))/(n-r-m)*qf(0.95,m,n-r-m))*
  sqrt((1+t(z0)%*%solve(t(Zmat)%*%Zmat)%*%z0)*(n/(n-r-1)*sigmat[i,i]))
upper<-t(z0)%*%beta[,i]+sqrt((m*(n-r-1))/(n-r-m)*qf(0.95,m,n-r-m))*
  sqrt((1+t(z0)%*%solve(t(Zmat)%*%Zmat)%*%z0)*(n/(n-r-1)*sigmat[i,i]))
  cat(paste("The 95% CI for y",i," is: (",lower,", ",upper,")\n"))
}

## The 95% CI for y 1  is: ( -310.068821144514 ,  1769.11836536906 )
## The 95% CI for y 2  is: ( -505.284056515704 ,  1656.7349910547 )
```

Exercise

The data in Liver.csv was collected during an investigation of the presence of diploid, tetraploid, and octaploid nuclei in liver cells. Measurements of the abundance of these nuclei were measured for ten patients, as well as their age and sex (1=female, 0=male). Determine whether there is a relationship between age, sex, and the abundance of these cells. Calculate a simultaneous prediction interval for a 40 year old male.

Multivariate GLM

We use generalised linear models to describe responses that are not normally distributed, such as binary outcomes or counts. Examples of GLMs are logistic regression, Poisson regression, Inverse Gaussian and negative binomial regression. In a GLM, we define:

1. A *response distribution*. The response distribution is a member of the exponential family of distributions that is thought to describe the distribution of the response. The Poisson, Binomial, Bernoulli, Exponential, Gamma, Negative Binomial, Normal, log Normal and inverse Gaussian distributions are all members of the exponential family. A distribution which is part of this family is defined in terms of a set of parameters $\boldsymbol{\theta}$ such that the PDF decomposes in the following way

$$f(y|\theta,\psi,\omega) = \exp\left(\frac{y\theta - b(\theta)}{\psi}\omega + c(y,\psi,\omega)\right)$$

2. A *link function*. This is a bijective (one to one) function that links the expected response to the linear combinations of random predictor variables

$$\mu_i = h(\eta_i) = h(\boldsymbol{x}_i^T\boldsymbol{\beta}).$$

   A link function that allows $\theta = \eta$, where θ is the parameter(s) in the relationship above.

We can extend the idea of multivariate regression to models with non–normal responses. For example, the class of generalised linear models can be extended to multivariate GLMs. In these models, each response is fitted as a generalised linear model with its own link function and response distribution. The benefit of using a multivariate model instead of a sequence of univariate GLMs is that the correlation between response variables can be incorporated into the inference.

An example of a multivariate GLM is multinomial logistic regression. The multinomial logistic regression model is used to model a categorical outcome with three or more categories (say, m in total). Since the outcome is one of the m categories, we define $m-1$ indicator variables in the following way

$$y_{ij} = \begin{cases} 1 & \text{if } Y_i = j; \\ 0 & \text{otherwise,} \end{cases}$$

for $j = 1,\ldots,m-1$. If $Y_i = m$, then $y_{ij} = 0$ for $j = 1,\ldots,m-1$. The link function for each of these response variables is the logistic link

$$g(\pi_{i1},\pi_{i2},\ldots,\pi_{im}) = \log\left(\frac{\pi_{ij}}{\pi_{i1}+\pi_{i2}+\ldots+\pi_{im}}\right).$$

The response distribution is the Bernoulli distribution, where $P(Y_i = j) = \pi_{ij}$. The linear predictor can be expressed as

$$\log\left(\frac{P(Y_i=j)}{P(Y_i=m)}\right) = \boldsymbol{x}_i^T\boldsymbol{\beta}_j$$

for $j = 1, \ldots, m-1$. In R, Thomas Yee's VGAM package allows us to fit multivariate GLMs using the *vglm* function (vector GLM).

Example

Mukhopadhyay and Khuri (2008) describe an experiment that was conducted by Gennings, Carter and Martin on the ability of two different drugs, morphine sulphate and Δ^9–tetrahydro–cannebinal (Δ^9–THC),to relieve pain in mice. The authors of the latter paper varied the dose of the two drugs and measured two responses, whether there was pain relief and whether there were side effects. Pain relief was successful if it took longer than eight seconds for the mouse to react when its tail was placed under a heat lamp, and side effects were said to have been observed if the body temperature of the mouse fell below 35 °C. Then we have four response categories, no relief or side effects, both relief and side effects, relief but no side effects, and no relief but with side effects.

The following model is set up:

$$\begin{aligned} \eta_1(\boldsymbol{x}) &= \beta_1 + \beta_2 x_1 + \beta_3 x_2, \\ \eta_2(\boldsymbol{x}) &= \beta_4 + \beta_5 x_1 + \beta_6 x_2, \\ \eta_3(\boldsymbol{x}) &= \beta_7 + \beta_8 x_1 + \beta_9 x_2, \\ \eta_4(\boldsymbol{x}) &= 0, \end{aligned} \tag{1}$$

where x_1 is the dose of morphine sulphate in mg/kg, x_2 is the dose of Δ^9–THC in mg/kg, and $\eta_1(\boldsymbol{x})$, $\eta_2(\boldsymbol{x})$, $\eta_3(\boldsymbol{x})$, and $\eta_4(\boldsymbol{x})$ are the linear predictors for Pain/Side Effects, Pain/No Side Effects, No Pain/Side Effects, and No Pain/No Side Effects, respectively.

```
GenningsRaw<-read.csv("C:/Documents/GenningsData.csv") GenningsFrame<-
data.frame(GenningsRaw)

linmod<-vglm(formula = cbind(Pain.and.Side.effect, Pain.and.No.side.effect,
          No.Pain.and.Side.effect, No.Pain.and.No.side.effect) ~ x1 + x2,
          family = multinomial, data = GenningsFrame, weights = m)
summary(linmod)

##
## Call:
## vglm(formula = cbind(Pain.and.Side.effect, Pain.and.No.side.effect,
##     No.Pain.and.Side.effect, No.Pain.and.No.side.effect) ~ x1 +
##     x2, family = multinomial, data = GenningsFrame, weights = m)
##
## Pearson residuals:
##                       Min    1Q Median     3Q Max
## log(mu[,1]/mu[,4]) -3.3 -1.18  -0.68 -0.345 5.3
## log(mu[,2]/mu[,4]) -5.0 -1.76   1.03  1.727 7.3
## log(mu[,3]/mu[,4]) -1.3 -0.63  -0.28 -0.085 9.2
```

```
##
## Coefficients:
##               Estimate Std. Error z value
## (Intercept):1    -5.29      0.317   -16.7
## (Intercept):2    -2.32      0.214   -10.8
## (Intercept):3    -5.64      0.639    -8.8
## x1:1              0.73      0.067    10.9
## x1:2              0.83      0.059    14.0
## x1:3              0.42      0.143     2.9
## x2:1              0.84      0.065    12.9
## x2:2              0.57      0.062     9.1
## x2:3              0.66      0.084     7.9
##
## Number of linear predictors:  3
##
## Names of linear predictors:
## log(mu[,1]/mu[,4]), log(mu[,2]/mu[,4]), log(mu[,3]/mu[,4])
##
## Dispersion Parameter for multinomial family:   1
##
## Residual deviance: 468.4 on 96 degrees of freedom
##
## Log-likelihood: -385.1 on 96 degrees of freedom
##
## Number of iterations: 6
```

```
intmod<-vglm(formula = cbind(Pain.and.Side.effect, Pain.and.No.side.effect,
            No.Pain.and.Side.effect, No.Pain.and.No.side.effect) ~ x1 * x2,
            family = multinomial, data = GenningsFrame, weights = m)
summary(intmod)
```

```
##
## Call:
## vglm(formula = cbind(Pain.and.Side.effect, Pain.and.No.side.effect,
##     No.Pain.and.Side.effect, No.Pain.and.No.side.effect) ~ x1 *
##     x2, family = multinomial, data = GenningsFrame, weights = m)
##
## Pearson residuals:
##                       Min    1Q Median    3Q Max
## log(mu[,1]/mu[,4]) -3.3 -1.15  -0.60 -0.25 5.3
## log(mu[,2]/mu[,4]) -5.2 -1.65   0.81  1.62 7.6
## log(mu[,3]/mu[,4]) -1.3 -0.59  -0.24 -0.11 9.1
##
## Coefficients:
##               Estimate Std. Error z value
```

```
## (Intercept):1   -4.205      0.375 -11.202
## (Intercept):2   -2.372      0.228 -10.407
## (Intercept):3   -5.830      0.798  -7.305
## x1:1             0.443      0.105   4.238
## x1:2             0.838      0.073  11.444
## x1:3             0.512      0.214   2.392
## x2:1             0.703      0.079   8.877
## x2:2             0.548      0.073   7.498
## x2:3             0.653      0.105   6.234
## x1:x2:1          0.045      0.039   1.141
## x1:x2:2          0.016      0.039   0.406
## x1:x2:3          0.002      0.047   0.043
##
## Number of linear predictors:  3
##
## Names of linear predictors:
## log(mu[,1]/mu[,4]), log(mu[,2]/mu[,4]), log(mu[,3]/mu[,4])
##
## Dispersion Parameter for multinomial family:   1
##
## Residual deviance: 451 on 93 degrees of freedom
##
## Log-likelihood: -376.5 on 93 degrees of freedom
##
## Number of iterations: 7

lrtest(intmod,linmod)

## Likelihood ratio test
##
## Model 1: cbind(Pain.and.Side.effect, Pain.and.No.side.effect, No.Pain.and.Side.effe
##     No.Pain.and.No.side.effect) ~ x1 * x2
## Model 2: cbind(Pain.and.Side.effect, Pain.and.No.side.effect, No.Pain.and.Side.effe
##     No.Pain.and.No.side.effect) ~ x1 + x2
##   #Df LogLik Df Chisq Pr(>Chisq)
## 1  93   -376
## 2  96   -385  3  17.4    0.00059 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exercise

A group of researcers studied infections from births by Caesarian section. The response variable had three outcomes, no infection, and infections of two types (imaginatively named type I and II). Three variables were thought to be related to the risk of infection, whether the Caesarian section was planned or not, whether there were risk factors present (such as excessive weight and diabeties), and whether antibiotics were given as a prophylaxis.

The data are in the dataset Caesarian.csv. Determine which of the factors are related to the risk of infection with the two types of infection.

Exercises

MANOVA - Johnson and Wichern Exercises 6.31, 6.32, 6.33, 6.41

Regression - Johnson and Wichern Exercises 7.15, 7.17, 7.18, 7.19, 7.21, 7.26