Solutions to Assignment #1

1. (a) The marginal distributions are normal with means and variances given by elements of $\boldsymbol{\mu}$ and the diagonal elements of C, respectively. Therefore, we have $X_1 \sim \mathcal{N}(1, 4.5)$, $X_2 \sim \mathcal{N}(2, 4.0)$, $X_3 \sim \mathcal{N}(1, 7.5)$, $X_4 \sim \mathcal{N}(0, 8.0)$, and $X_5 \sim \mathcal{N}(0, 5.5)$.
(b) To obtain the conditional distribution of (X_1, X_2) given $X_3 = 2$, $X_4 = 3$ and $X_5 = -1$, we partition μ and C as follows:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

where

$$\boldsymbol{\mu}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\mu}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

and

$$C_{11} = \begin{pmatrix} 4.5 & -2.0 \\ -2.0 & 4.0 \end{pmatrix}, \quad C_{12} \begin{pmatrix} -1.5 & -1.0 & -0.5 \\ 3.0 & 2.0 & 1.0 \end{pmatrix}, \quad C_{22} = \begin{pmatrix} 7.5 & 5.0 & 2.5 \\ 5.0 & 8.0 & 4.0 \\ 2.5 & 4.0 & 5.5 \end{pmatrix}$$

with $C_{21} = C_{12}^T$.
Then the conditional distribution is normal with mean

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + C_{12}C_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2) = \begin{pmatrix} 0.8 \\ 2.4 \end{pmatrix}$$

and covariance matrix

$$C_{1|2} = C_{11} - C_{12}C_{22}^{-1}C_{21} = \begin{pmatrix} 4.2 & -1.4 \\ -1.4 & 2.8 \end{pmatrix};$$

$\boldsymbol{\mu}_{1|2}$ and $C_{1|2}$ are evaluated in R as follows:

```
> C <- matrix(c(4.5,-2.0,-1.5,-1,-0.5,-2.0,4.0,3.0,2.0,1.0,-1.5,3.0,7.5,5.0,
+  2.5,-1.0,2.0,5.0,8.0,4.0,-0.5,1.0,2.5,4.0,5.5), ncol=5,byrow=T)
> C11 <- C[1:2,1:2]
> C12 <- C[1:2,3:5]
> C22 <- C[3:5,3:5]
> mu1 <- c(1,2)
> mu2 <- c(1,0,0)
> x2 <- c(2,4,-1)
```

```
> mu1 + C12%*%solve(C22,x2-mu2)
     [,1]
[1,]  0.8
[2,]  2.4
> C11 - C12%*%solve(C22)%*%t(C12)
     [,1] [,2]
[1,]  4.2 -1.4
[2,] -1.4  2.8
```

(c) The inverse of C can be evaluated using the following R code (using the matrix `C` given in part (b)):

```
> round(solve(C),5)
        [,1]     [,2]     [,3]     [,4]     [,5]
[1,] 0.28571  0.14286  0.00000  0.00000  0.00000
[2,] 0.14286  0.42857 -0.14286  0.00000  0.00000
[3,] 0.00000 -0.14286  0.28571 -0.14286  0.00000
[4,] 0.00000  0.00000 -0.14286  0.28571 -0.14286
[5,] 0.00000  0.00000  0.00000 -0.14286  0.28571
```

(The rounding to 5 decimal places shows the exact zeroes in C^{-1}, which can be obscured by roundoff error.) In fact, we can write the inverse of C as

$$C^{-1} = \begin{pmatrix} 2/7 & 1/7 & 0 & 0 & 0 \\ 1/7 & 3/7 & -1/7 & 0 & 0 \\ 0 & -1/7 & 2/7 & -1/7 & 0 \\ 0 & 0 & -1/7 & 2/7 & -1/7 \\ 0 & 0 & 0 & -1/7 & 2/7 \end{pmatrix}.$$

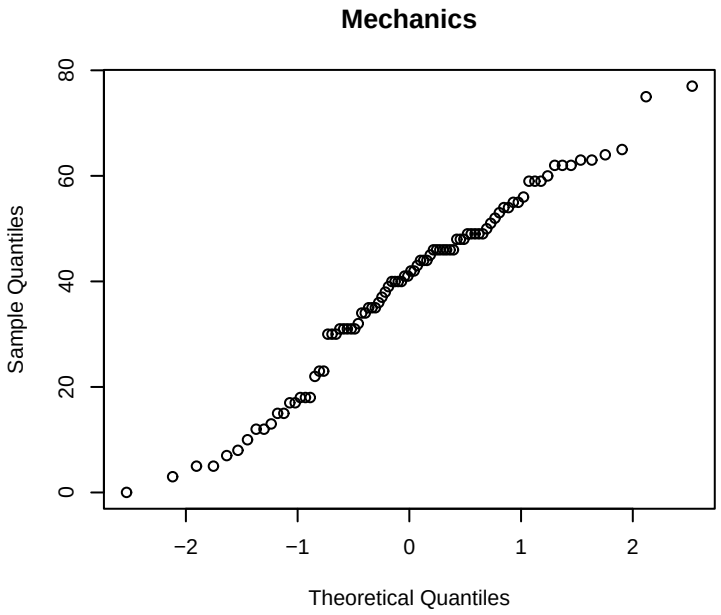
The graph structure of the dependence of $\boldsymbol{X}$ is

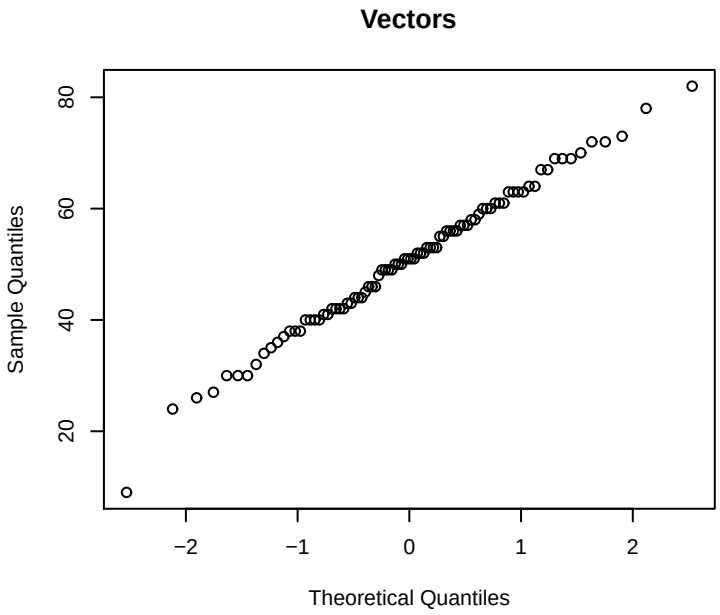
$$\mathbf{1} \longleftrightarrow \mathbf{2} \longleftrightarrow \mathbf{3} \longleftrightarrow \mathbf{4} \longleftrightarrow \mathbf{5}$$

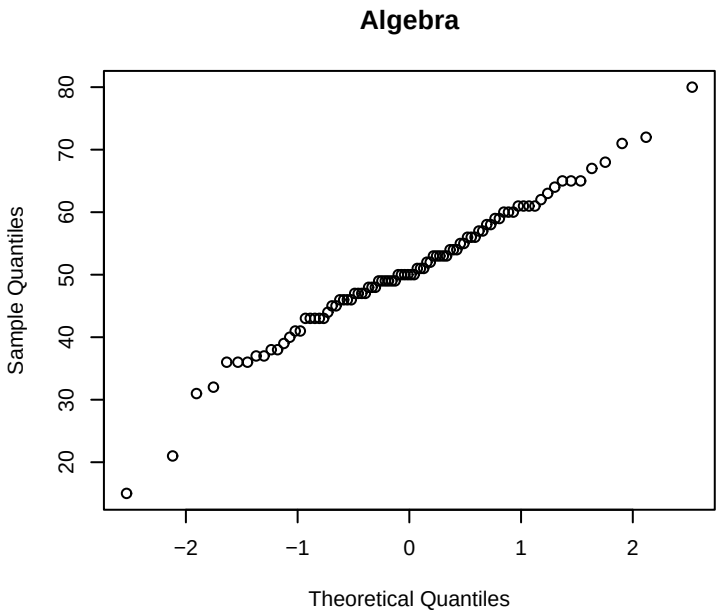
Thus, for example, variables 1 and 3 are conditionally independent given variable 2 while variables 1 and 5 are conditionally independent given variables 2,3, and 4.

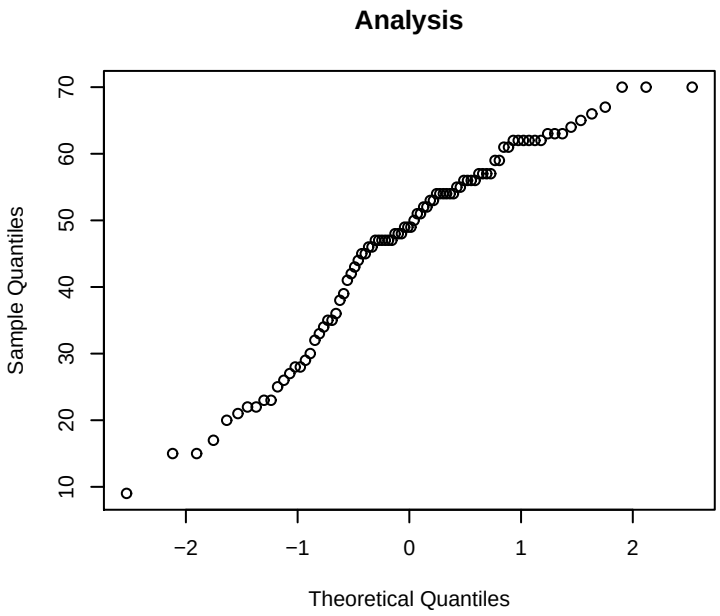
2. (a) The normal qq plots (shown on the following page) and Shapiro-Wilk tests are generated as follows:

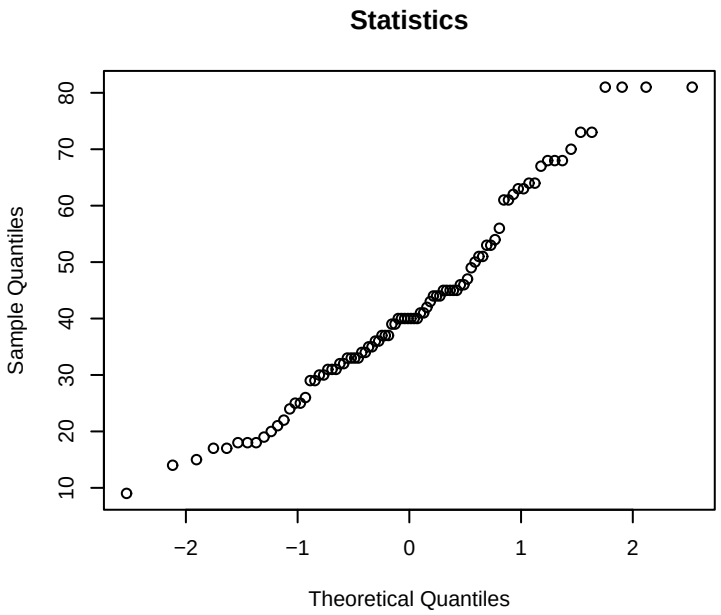
```
> qqnorm(mec,main="Mechanics")
> shapiro.test(mec)
W = 0.9724, p-value = 0.05708

> qqnorm(vec,main="Vectors")
```

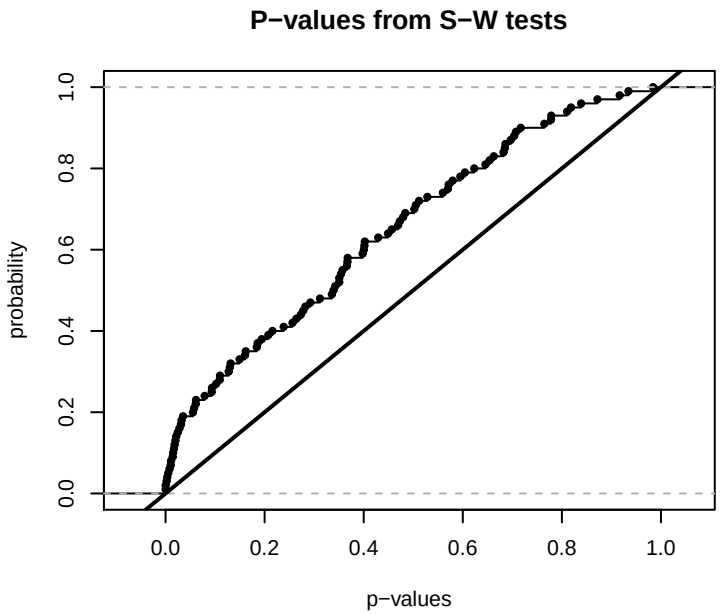

Mechanics
Sample Quantiles
Theoretical Quantiles


Vectors
Sample Quantiles
Theoretical Quantiles


Algebra
Sample Quantiles
Theoretical Quantiles


Analysis
Sample Quantiles
Theoretical Quantiles


Statistics
Sample Quantiles
Theoretical Quantiles


P−values from S−W tests
probability
p−values

```
> shapiro.test(vec)
W = 0.9931, p-value = 0.9276

> qqnorm(alg,main="Algebra")
> shapiro.test(alg)
W = 0.9821, p-value = 0.2637

> qqnorm(ana,main="Analysis")
> shapiro.test(ana)
W = 0.9425, p-value = 0.0006896

> qqnorm(sta,main="Statistics")
> shapiro.test(sta)
W = 0.9645, p-value = 0.01633
```

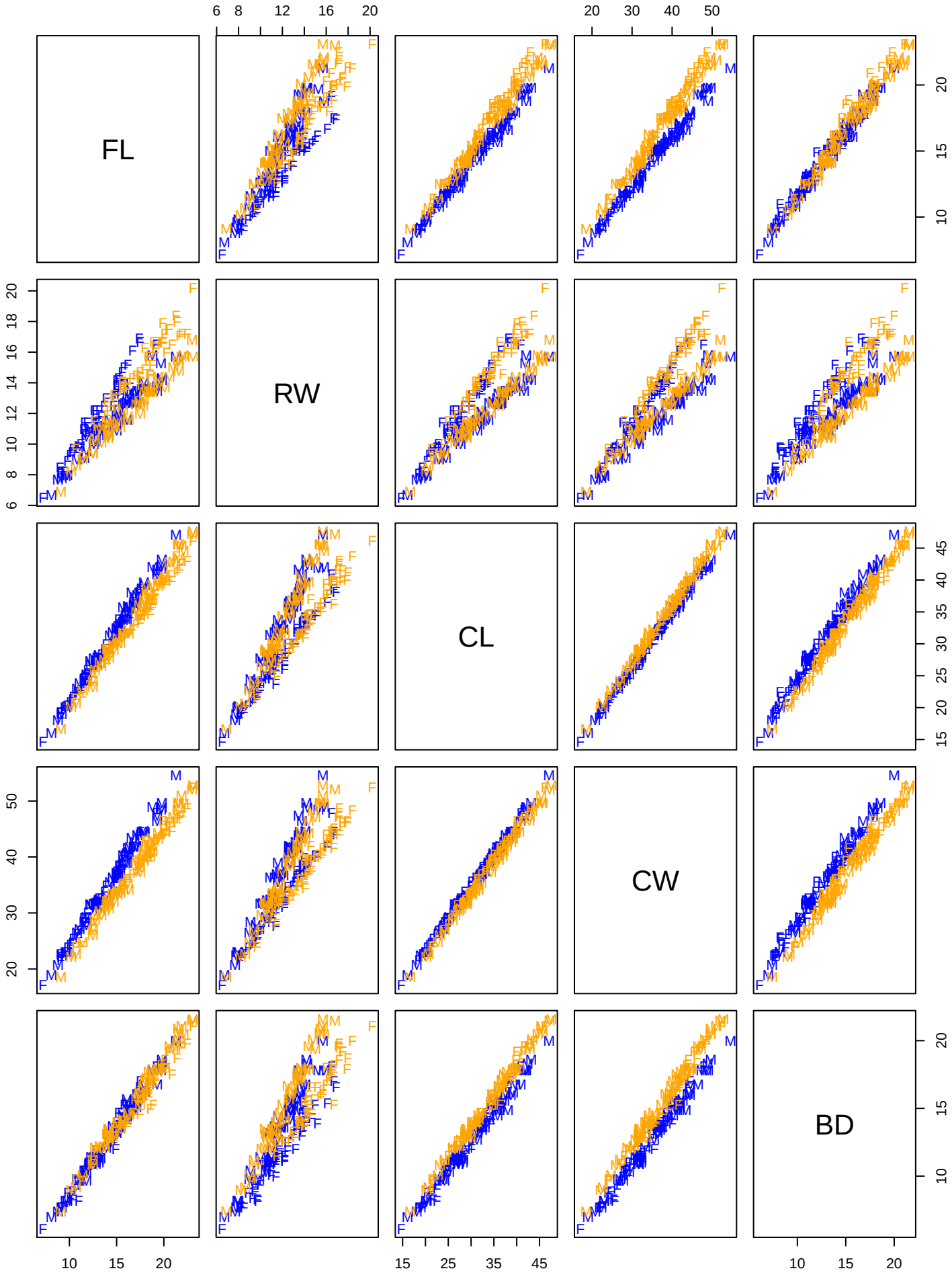
Of the 5 exam marks, only "Analysis" is clearly non-normal although the S-W test does cast some doubt on "Mechanics" and "Statistics".
(b) The output from `qqmultinorm` is given on the previous page. Note that the p-values lie above the 45 degree line, which indicates that for the 100 projections, we are seeing more small p-values than we would expect to see if the data came from a multivariate normal distribution. This suggests that a multivariate normal is probably not 100% appropriate for these data.

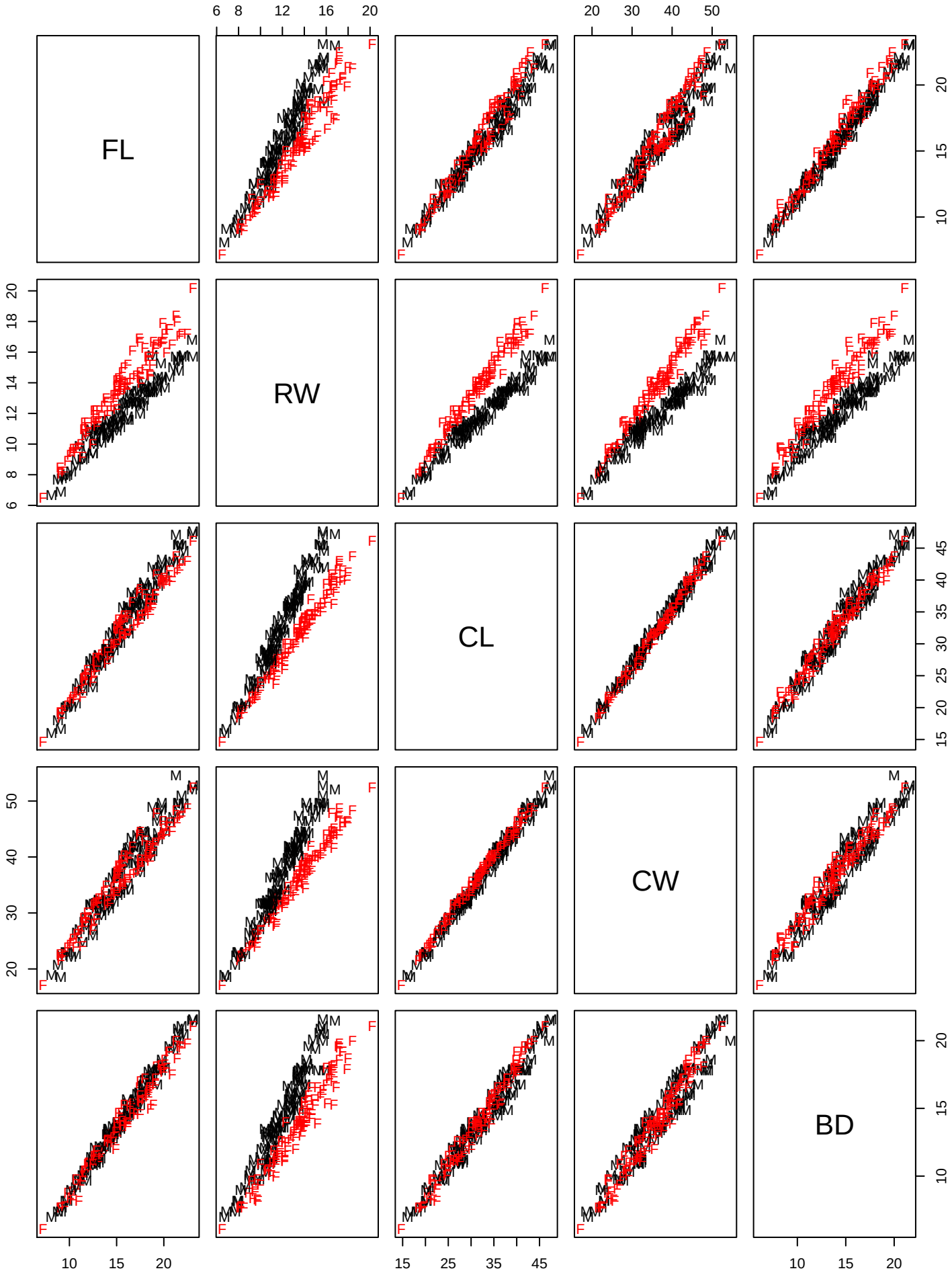
3. (a) The pairwise scatterplots (for parts (a) and (b)) are given on the follow pages. We are looking for scatterplots with a clear separation of the two colours; this seems fairly clear in the plots of CW vs FL and CD vs BD, for example.
(b) To facilitate the interpretation, we will assign colours to the two sexes:

```
> coloursex <- ifelse(sex=="M","black","red")
> pairs(cbind(FL,RW,CL,CW,BD),pch=sex,col=coloursex)
```

The separation between the two sexes seems clear, for example, in the plots of CL vs RW and CW vs RW.
(c) Because the scatterplots given by `scatterplot3d` are static, they don't really convey much information about triples (triplets) of variables that separate colour and sex.

4. (a) The plots of p-values for 20 and 1000 projections are given on the last page. These p-values are consistent with the projections being normally distributed even though the co-ordinate wise qq plots are clearly non-normal.
(b) This is not easy but here's a fairly rigorous proof. Effectively, we need to show that $\max_{1\leq j\leq p} a_j^2/\boldsymbol{a}^T\boldsymbol{a}$ is small for almost all projections and so when p is large $\boldsymbol{a}^T\boldsymbol{X} = \sum_{j=1}^p a_j X_j$ is approximately normally for almost all projections.


FL
RW
CL
CW
BD
6 8 12 16 20
20 30 40 50
10 15 20
6 8 10 12 14 16 18 20
15 20 25 30 35 40 45
20 30 40 50
10 15 20
15 25 35 45
10 15 20
M
F


FL
RW
CL
CW
BD
F
M

We can generate "uniformly distributed" projections by taking $Z_1, \cdots, Z_p$ independent $\mathcal{N}(0,1)$ random variables and defining

$$a_j = \frac{Z_j}{\sqrt{Z_1^2 + \cdots + Z_p^2}} = \frac{Z_j/\sqrt{p}}{\sqrt{(Z_1^2 + \cdots + Z_p^2)/p}}$$

so that $\boldsymbol{a}^T\boldsymbol{a} = 1$. When p is large,

$$\frac{Z_1^2 + \cdots + Z_p^2}{p} \approx 1$$

by the Strong Law of Large Numbers. Therefore

$$\max_{1\le j\le p} a_j^2 = \max_{1\le j\le p} \frac{a_j^2}{\boldsymbol{a}^T\boldsymbol{a}} \approx \max_{1\le j\le p} \frac{Z_j^2}{p}$$

for large p. Now

$$P\left(\max_{1\le j\le p} \frac{Z_j^2}{p} > \epsilon\right) \le pP(\frac{Z_j^2}{p} > \epsilon) = pP(|Z_j| > \sqrt{\epsilon p}).$$

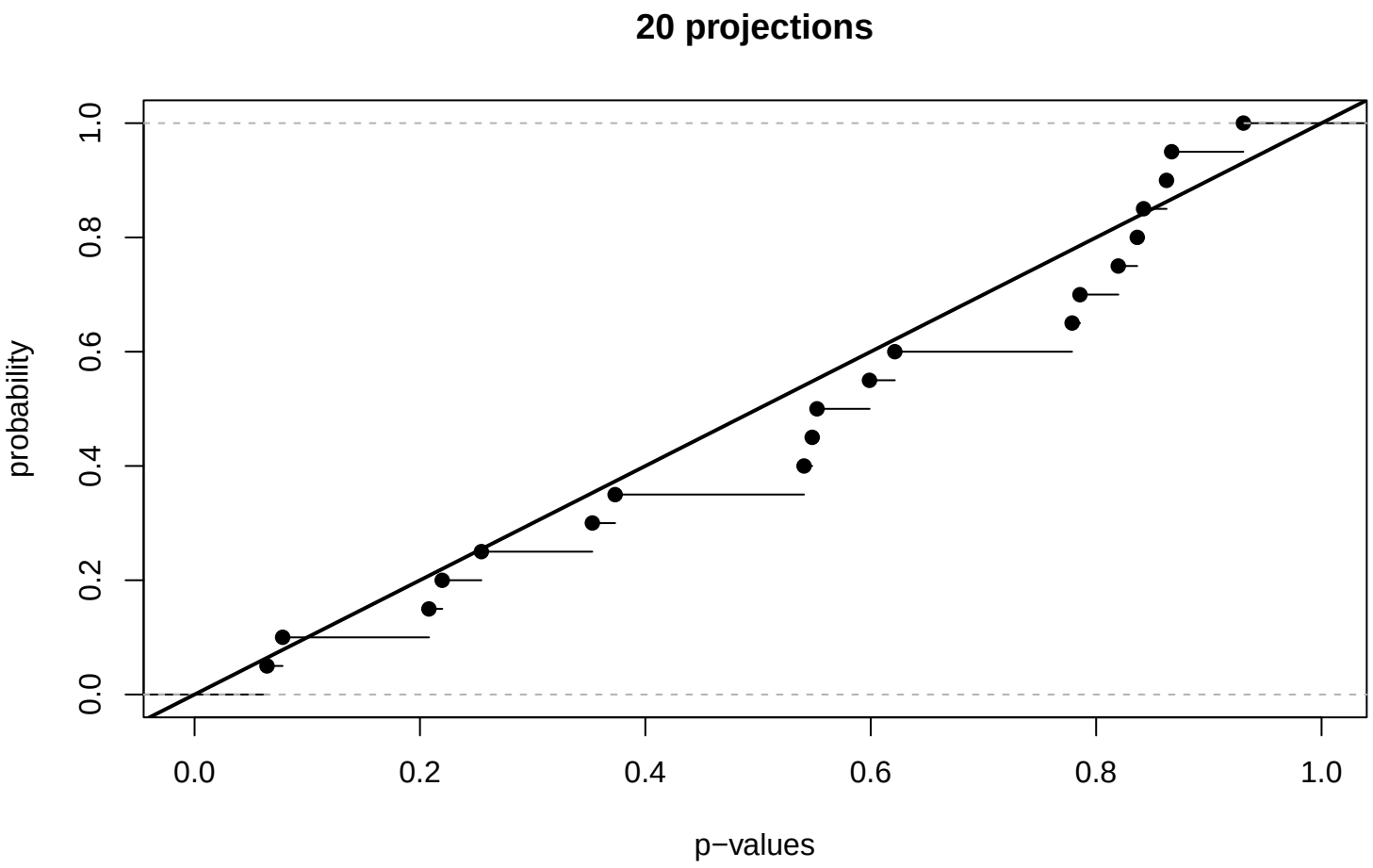
Since Z_j is $\mathcal{N}(0,1)$,

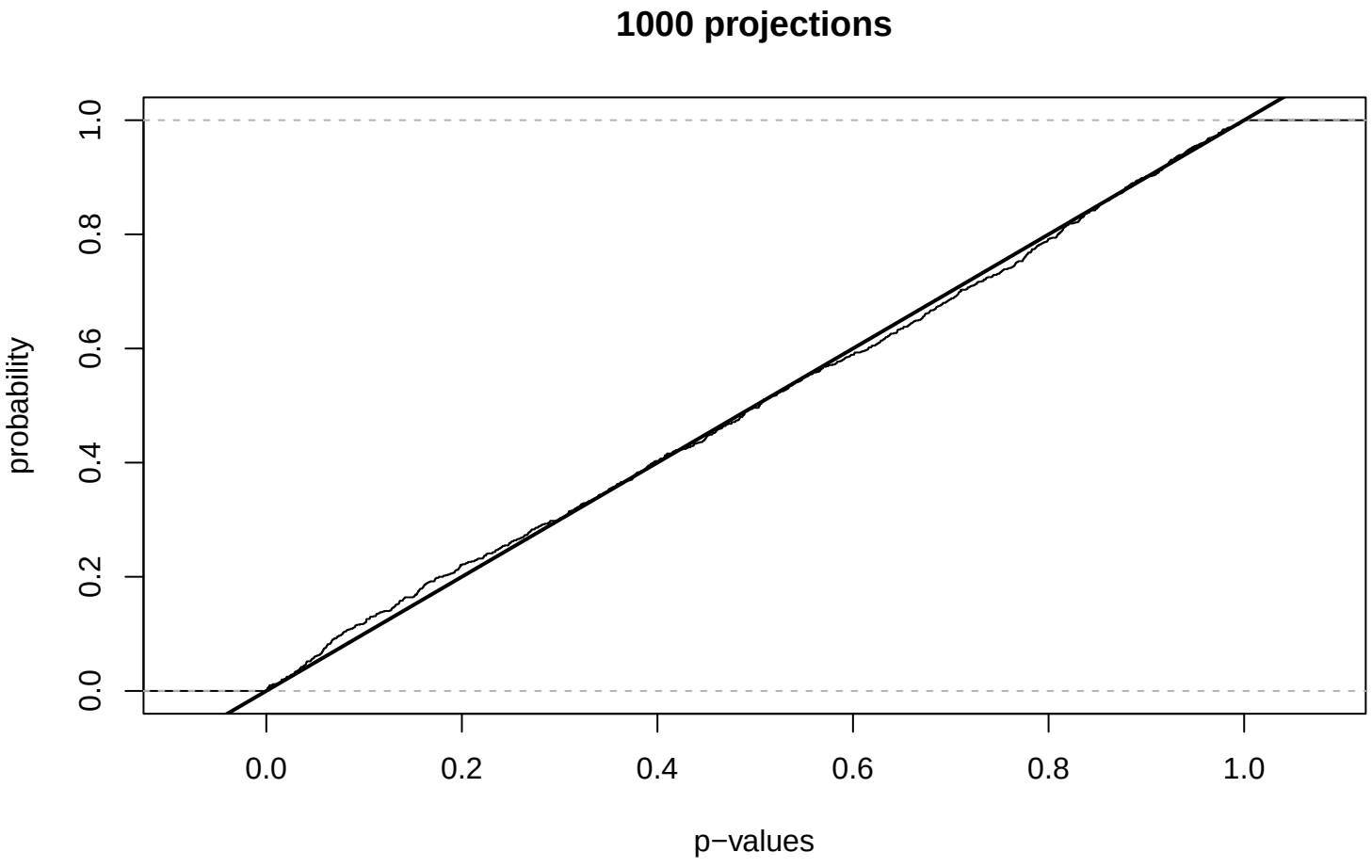
$$P(|Z_j| > \sqrt{\epsilon p}) \approx \frac{1}{\sqrt{2\pi\epsilon p}} \exp\left(-\frac{1}{2}\epsilon p\right)$$

and so

$$pP(|Z_j| > \sqrt{\epsilon p}) \approx \frac{\sqrt{p}}{\sqrt{2\pi\epsilon}} \exp\left(-\frac{1}{2}\epsilon p\right) \to 0$$

as $p \to \infty$ for any $\epsilon > 0$.


20 projections
probability
0.0
0.2
0.4
0.6
0.8
1.0
p−values


1000 projections
probability
0.0
0.2
0.4
0.6
0.8
1.0
p−values