

Code ▼

Project 5

JONATHAN FENG

PROBLEM 11.10

Use the glucose data of Table 3.9.

Hide

```
glucose <- read.csv(file="C:\\Users\\Taterthot\\Desktop\\da 410\\a6\\blood.csv", header = TRUE,
  fileEncoding = 'UTF-8-BOM')
glucose
```

y1 <int>	y2 <int>	y3 <int>	x1 <int>	x2 <int>	x3 <int>
60	69	62	97	69	98
56	53	84	103	78	107
80	69	76	66	99	130
55	80	90	80	85	114
62	75	68	116	130	91
74	64	70	109	101	103
64	71	66	77	102	130
73	70	64	115	110	109
68	67	75	76	85	119
69	82	74	72	133	127

1-10 of 34 rows

Previous
1
2
3
4
Next

(a) Find the canonical correlations between (y1,y2,y3) and (x1,x2,x3)

Standardize and Separate Y and X

Hide

```
glucose.std<- sweep(glucose, 2, sqrt(apply(glucose,2,var)), FUN="/")

glucose.y <- glucose.std[,1:3]
glucose.x <- glucose.std[,4:6]

glucose.y
```

y1 <dbl>	y2 <dbl>	y3 <dbl>
6.317396	10.213946	7.874857
5.896236	7.845495	10.669161
8.423195	10.213946	9.653050
5.790946	11.842257	11.431244
6.527976	11.102116	8.636940
7.791455	9.473805	8.890968
6.738556	10.510003	8.382912
7.686165	10.361975	8.128885
7.159716	9.917890	9.526037
7.265006	12.138313	9.399023

1-10 of 34 rows

Previous 1 2 3 4 Next

glucose.x

x1 <dbl>	x2 <dbl>	x3 <dbl>
3.046496	2.994099	4.255785
3.234939	3.384633	4.646623
2.072874	4.295881	5.645430
2.512574	3.688382	4.950607
3.643233	5.641056	3.951801
3.423383	4.382666	4.472917
2.418353	4.426059	5.645430
3.611826	4.773201	4.733476
2.386946	3.688382	5.167739
2.261317	5.771234	5.515150

1-10 of 34 rows

Previous 1 2 3 4 Next

Finding correlation blocks and correlation matrix.

From the previous problem in 3.9 for R12 we did the combination of 11,12,21,22 but in this problem we have to do all correlation combinations up to 33 for R12.

Hide

```

R11 <- cor(glucose.y)
R22 <- cor(glucose.x)

R12 <- c(
  cor(glucose.y[,1], glucose.x[,1]),
  cor(glucose.y[,1], glucose.x[,2]),
  cor(glucose.y[,1], glucose.x[,3]),
  cor(glucose.y[,2], glucose.x[,1]),
  cor(glucose.y[,2], glucose.x[,2]),
  cor(glucose.y[,2], glucose.x[,3]),
  cor(glucose.y[,3], glucose.x[,1]),
  cor(glucose.y[,3], glucose.x[,2]),
  cor(glucose.y[,3], glucose.x[,3])
)

R12 <- matrix(R12, ncol=ncol(R22), byrow=T)
R21 <- t(R12)

R11

```

	y1	y2	y3
y1	1.0000000	0.1591649	0.2011594
y2	0.1591649	1.0000000	0.2159339
y3	0.2011594	0.2159339	1.0000000

Hide

R22

	x1	x2	x3
x1	1.0000000	0.4843682	0.2850466
x2	0.4843682	1.0000000	0.1826615
x3	0.2850466	0.1826615	1.0000000

Hide

R12

	[,1]	[,2]	[,3]
[1,]	0.28036101	0.14811604	0.2650217
[2,]	-0.05564890	0.35192166	-0.1422024
[3,]	0.03515924	-0.01276187	0.1655364

Hide

R21

```

      [,1]      [,2]      [,3]
[1,] 0.2803610 -0.0556489  0.03515924
[2,] 0.1481160  0.3519217 -0.01276187
[3,] 0.2650217 -0.1422024  0.16553643

```

Finding E1 E2 matrices.

Hide

```

E1 <- solve(R11) %*% R12 %*% solve(R22) %*% R21
E2 <- solve(R22) %*% R21 %*% solve(R11) %*% R12

print(E1)

```

```

      [,1]      [,2]      [,3]
y1  0.12005804 -0.06140349  0.04336961
y2 -0.06516628  0.24219472 -0.05616142
y3  0.03094742 -0.08212428  0.03278953

```

Hide

```
print(E2)
```

```

      [,1]      [,2]      [,3]
x1  0.09160344 -0.06689565  0.1163865
x2 -0.05023836  0.18672327 -0.1171695
x3  0.07596807 -0.05451490  0.1167156

```

Eigen Vectors

Hide

```
eigen(E1)
```

```

eigen() decomposition
$values
[1] 0.2922871 0.0922920 0.0104632

$vectors
      [,1]      [,2]      [,3]
[1,] 0.3886112 -0.9225340 -0.2916344
[2,] -0.8641174 -0.3828454  0.1470526
[3,] 0.3198163  0.0485839  0.9451587

```

Hide

```
eigen(E2)
```

```
eigen() decomposition
$values
[1] 0.2922871 0.0922920 0.0104632

$vectors
      [,1]      [,2]      [,3]
[1,] -0.5055211 0.4182371 0.7395554
[2,] 0.7375469 0.7864504 -0.2135379
[3,] -0.4477421 0.4545037 -0.6383254
```

The canonical correlations

Hide

```
canon.corr <- sqrt(eigen(E1)$values)
canon.corr
```

```
[1] 0.5406358 0.3037960 0.1022898
```

Hide

```
canon.corr2 <- sqrt(eigen(E2)$values)
canon.corr2
```

```
[1] 0.5406358 0.3037960 0.1022898
```

Characteristic equations:

$a_1 = (0.389, -0.864, 0.320)$ $b_1 = (-0.506, 0.738, -0.448)$
 $a_2 = (-0.923, -0.383, 0.049)$ $b_2 = (0.418, 0.786, 0.455)$
 $a_3 = (-0.292, 0.147, 0.945)$ $b_3 = (0.740, -0.214, -0.638)$

Hide

```
u1 = 0.389*glucose.y + 0.864*glucose.y + 0.320*glucose.y
v1 = -0.506*glucose.x + 0.738*glucose.x - 0.448*glucose.x

u2 = -0.923*glucose.y - 0.383*glucose.y + 0.049*glucose.y
v2 = 0.418*glucose.x + 0.786*glucose.x + 0.455*glucose.x

u3 = -0.292*glucose.y + 0.147*glucose.y + 0.945*glucose.y
v3 = 0.740*glucose.x - 0.214*glucose.x - 0.638*glucose.x

u1
```

y1	y2	y3
<dbl>	<dbl>	<dbl>
9.937264	16.06654	12.38715

y1 <dbl>	y2 <dbl>	y3 <dbl>
9.274780	12.34096	16.78259
13.249686	16.06654	15.18425
9.109159	18.62787	17.98135
10.268506	17.46363	13.58591
12.255959	14.90230	13.98549
10.599748	16.53223	13.18632
12.090338	16.29939	12.78674
11.262233	15.60084	14.98446
11.427854	19.09357	14.78466
1-10 of 34 rows	Previous 1 2 3 4 Next	

Hide

v1

x1 <dbl>	x2 <dbl>	x3 <dbl>
-0.6580432	-0.6467253	-0.9192496
-0.6987469	-0.7310808	-1.0036705
-0.4477407	-0.9279102	-1.2194128
-0.5427161	-0.7966906	-1.0693312
-0.7869383	-1.2184680	-0.8535889
-0.7394506	-0.9466559	-0.9661501
-0.5223642	-0.9560287	-1.2194128
-0.7801543	-1.0310114	-1.0224307
-0.5155803	-0.7966906	-1.1162317
-0.4884445	-1.2465865	-1.1912725
1-10 of 34 rows	Previous 1 2 3 4 Next	

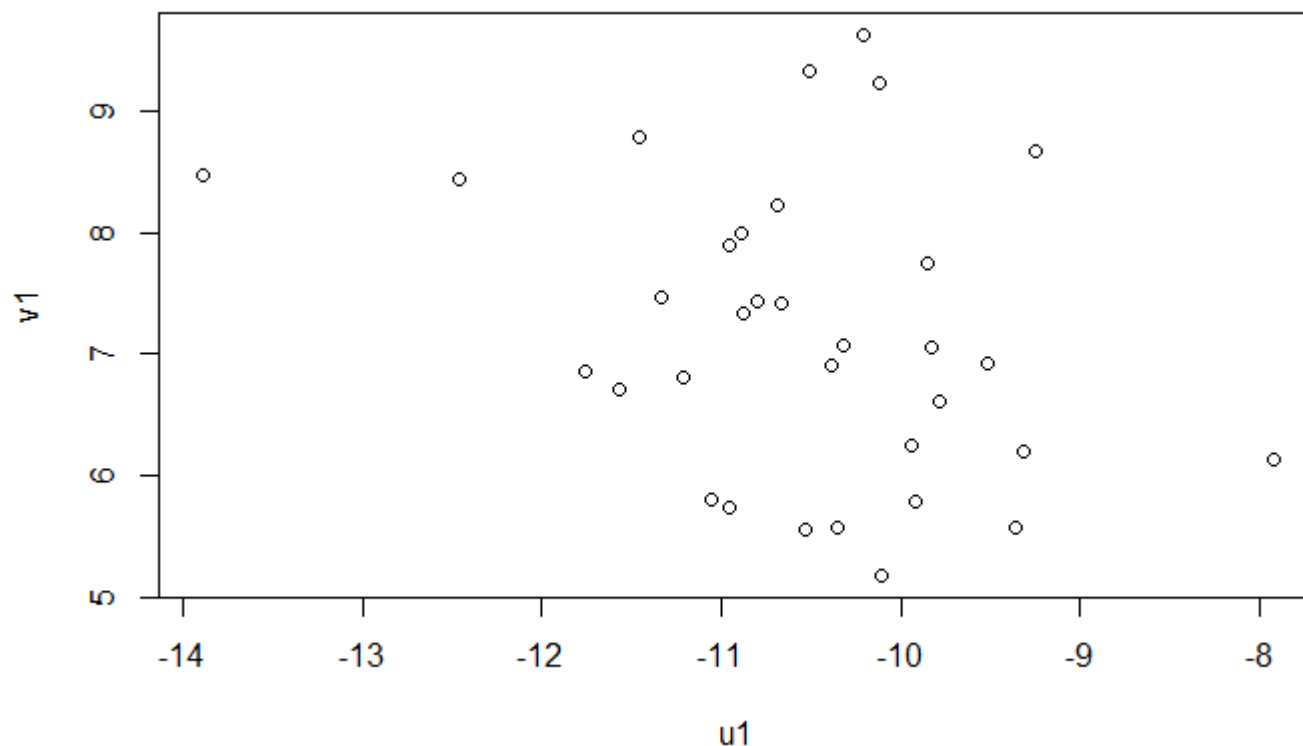
Hide

NA

plotting first set canonical:

Hide

```
u1 <- as.matrix(glucose.std[,1:3]) %*% as.matrix(eigen(E1)$vectors[,2])
v1 <- as.matrix(glucose.std[,4:6]) %*% as.matrix(eigen(E2)$vectors[,2])
plot(u1,v1)
```



Hide

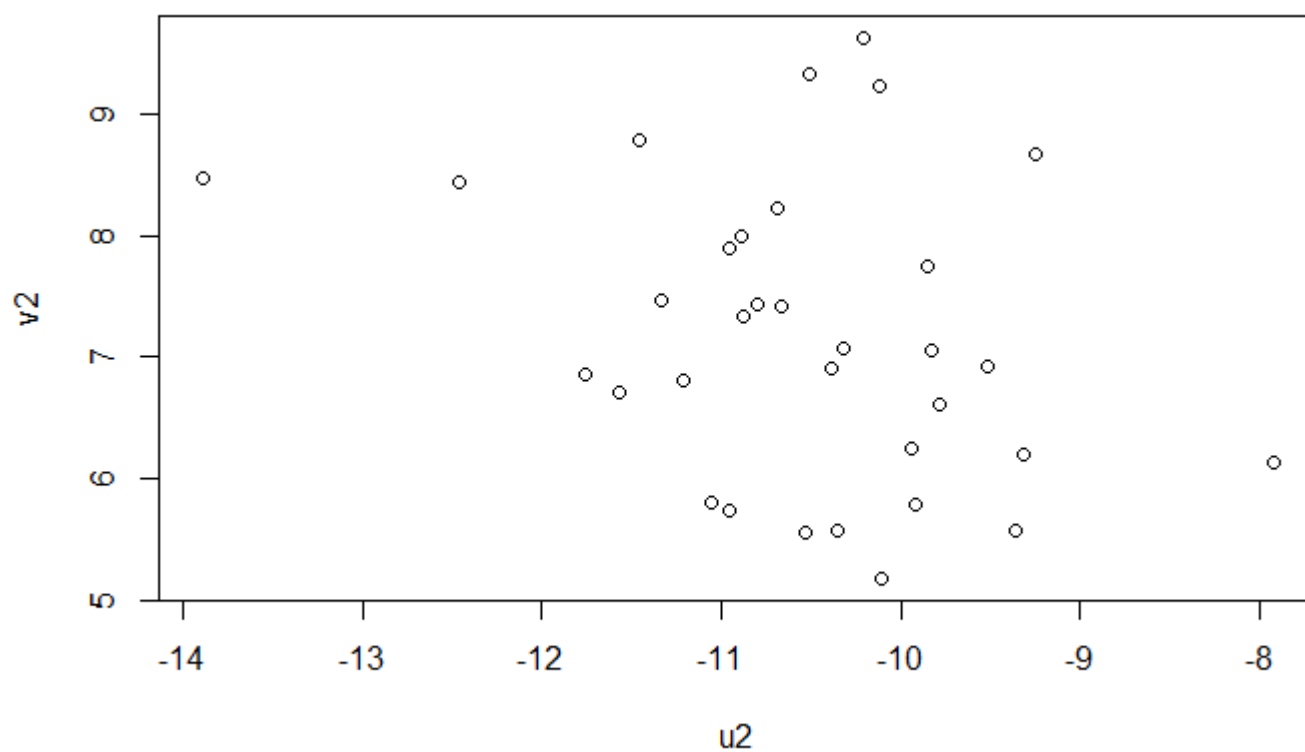
```
cor(u1,v1)
```

```
      [,1]
[1,] -0.303796
```

plotting second set canonical:

Hide

```
u2 <- as.matrix(glucose.std[,1:3]) %*% as.matrix(eigen(E1)$vectors[,2])
v2 <- as.matrix(glucose.std[,4:6]) %*% as.matrix(eigen(E2)$vectors[,2])
plot(u2,v2)
```

[Hide](#)

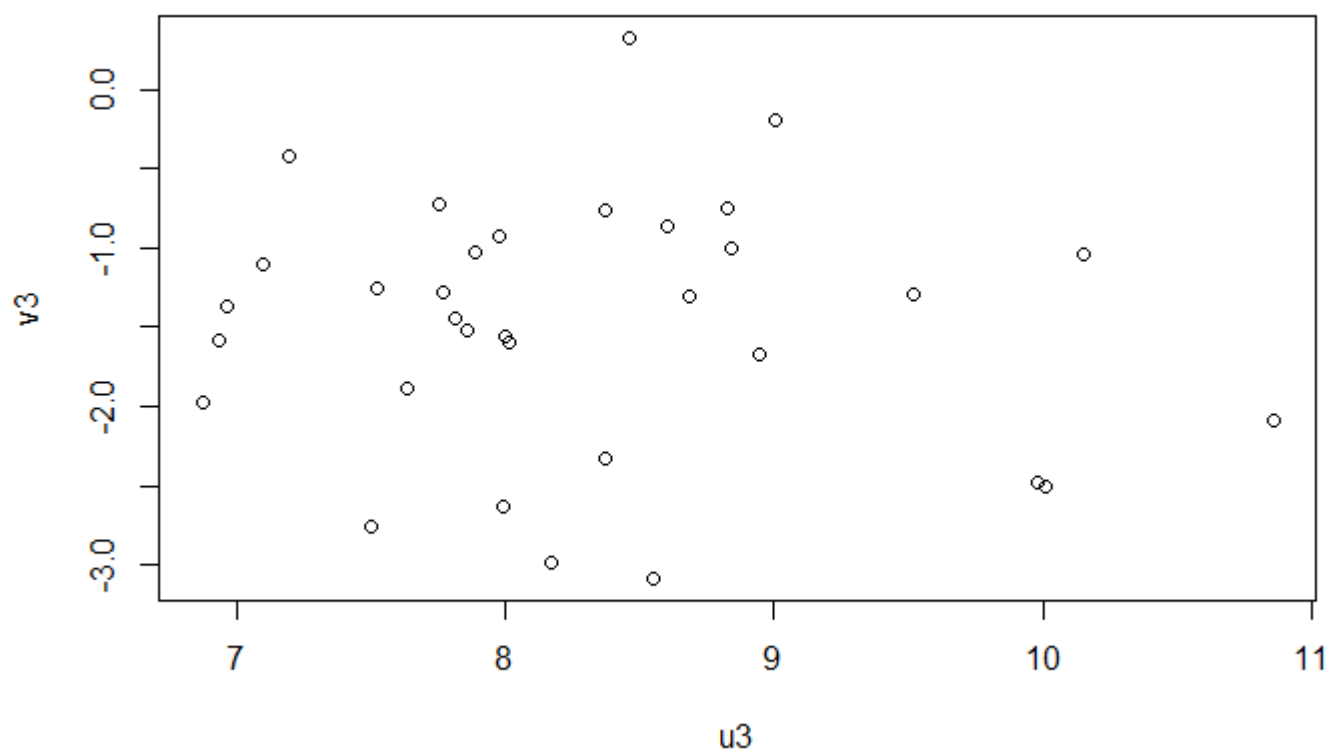
```
cor(u2,v2)
```

```
      [,1]  
[1,] -0.303796
```

plotting third set canonical:

[Hide](#)

```
u3 <- as.matrix(glucose.std[,1:3]) %*% as.matrix(eigen(E1)$vectors[,3])  
v3 <- as.matrix(glucose.std[,4:6]) %*% as.matrix(eigen(E2)$vectors[,3])  
plot(u3,v3)
```

Hide

```
cor(u3,v3)
```

```
      [,1]  
[1,] -0.1022898
```

(b) Find the standardized coefficients for the canonical variates.

Hide

```
library(CCA)  
s1 <- diag(sqrt(diag(cov(glucose.y))))  
cc1 <- cc(glucose.y, glucose.x)  
s1 %*% cc1$xcoef
```

```
      [,1]      [,2]      [,3]  
[1,] -0.4281702 -0.88510423 -0.3015242  
[2,]  0.9520809 -0.36731233  0.1520394  
[3,] -0.3523722  0.04661271  0.9772105
```

Hide

```
s2 <- diag(sqrt(diag(cov(glucose.x))))  
s2 %*% cc1$ycoef
```

	[,1]	[,2]	[,3]
[1,]	-0.6283733	-0.3351165	-0.9334710
[2,]	0.9167863	-0.6301509	0.2695288
[3,]	-0.5565529	-0.3641754	0.8056979

(c) Test the significance of each canonical correlation.

[Hide](#)

```
cancor2<-function(x,y,dec=4){
  x<-as.matrix(x);y<-as.matrix(y)
  n<-dim(x)[1];q1<-dim(x)[2];q2<-dim(y)[2];q<-min(q1,q2)
  S11<-cov(x);S12<-cov(x,y);S21<-t(S12);S22<-cov(y)
  E1<-eigen(solve(S11)%*%S12%*solve(S22)%*%S21);E2<-
eigen(solve(S22)%*%S21%*solve(S11)%*%S12)
  rsquared<-as.double(E1$values[1:q])
  LR<-NULL;pp<-NULL;qq<-NULL;tt<-NULL
  for (i in 1:q){
    LR<-c(LR,prod(1-rsquared[i:q]))
    pp<-c(pp,q1-i+1)
    qq<-c(qq,q2-i+1)
    tt<-c(tt,n-1-i+1)}
  m<-tt-0.5*(pp+qq+1);lambda<-(1/4)*(pp*qq-2);s<-sqrt((pp^2*qq^2-
4)/(pp^2+qq^2-5))
  F<-((m*s-2*lambda)/(pp*qq))*((1-LR^(1/s))/LR^(1/s));df1<-pp*qq;df2<-(m*s-2*lambda);pval<-1-pf(F
,df1,df2)
  outmat<-round(cbind(sqrt(rsquared),rsquared,LR,F,df1,df2,pval),dec)

  colnames(outmat)=list("R","RSquared","LR","ApproxF","NumDF","DenDF","pvalue")
  rownames(outmat)=as.character(1:q);xrels<-
round(cor(x,x%*%E1$vectors)[,1:q],dec)
  colnames(xrels)<-
apply(cbind(rep("U",q),as.character(1:q)),1,paste,collapse="")
  yrels<-round(cor(y,y%*%E2$vectors)[,1:q],dec)
  colnames(yrels)<-
apply(cbind(rep("V",q),as.character(1:q)),1,paste,collapse="")
  list(Summary=outmat,a.Coefficients=E1$vectors,b.Coefficients=E2$vectors,
XUCorrelations=xrels,YVCorrelations=yrels)
}
```

[Hide](#)

```
cancor2(glucose.y, glucose.x)
```

```

$Summary
      R  RSquared    LR  ApproxF  NumDF    DenDF  pvalue
1 0.5406    0.2923 0.6357  1.5527     9 68.2952 0.1475
2 0.3038    0.0923 0.8982  0.7996     4 58.0000 0.5304
3 0.1023    0.0105 0.9895  0.3172     1 30.0000 0.5775

$a.Coefficients
      [,1]      [,2]      [,3]
[1,] 0.3886112 -0.9225340 -0.2916344
[2,] -0.8641174 -0.3828454  0.1470526
[3,]  0.3198163  0.0485839  0.9451587

$b.Coefficients
      [,1]      [,2]      [,3]
[1,] -0.5055211 0.4182371  0.7395554
[2,]  0.7375469 0.7864504 -0.2135379
[3,] -0.4477421 0.4545037 -0.6383254

$XUCorrelations
      U1      U2      U3
y1 0.3475 -0.9342 -0.0807
y2 -0.8078 -0.4981  0.3151
y3  0.2329 -0.2107  0.9494

$YVCorrelations
      V1      V2      V3
x1 -0.3430 0.7441  0.5733
x2  0.5108 0.8590  0.0354
x3 -0.5682 0.5748 -0.5888

```

H0: the second set of variables is independent from the first set of variables.

Ha: the second set of variables is dependent on the first set of the variables.

Based on our significance for each of the canonical relationship pairs, the first relationship has a p-value higher than 0.05 alpha level thus we fail to reject the null hypothesis and say that the first and second set of variables are independent from each other. Our second and third pairs when comparing has a p-value much larger than our level of significance at 0.05 thus we fail to reject the null hypothesis and say that the first and second set of variables are independent from each other.

Thus to put it in simple terms, none of the canonical relationship pairs are dependent.