Mingjie Zhao HW8

**a** H01: chd does not have significant effect on the dependent variables(age, ht, wt, SBP)

p-value is smaller than 0.01, so reject.

H02: behtype does not have significant effect on the dependent variables(age, ht, wt, SBP)

p-value is smaller than 0.01, so reject.

**Codes**

wcgs1= read.table("wcgs22.dat",header=T)

wcgs= as.matrix(wcgs1)

Y=wcgs[,1:4]

chd=wcgs[,5]

behtype=wcgs[,6]

> fit.main = manova(Y ~ chd+behtype)

> summary(fit.main)

Df Pillai approx F num Df den Df Pr(>F)

chd 1 0.0292368 23.7024 4 3148 < 2.2e-16 \*\*\*

behtype 1 0.0098296 7.8127 4 3148 2.912e-06 \*\*\*

Residuals 3151

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**b**

H01: chd does not have significant effect on the dependent variables(age, ht, wt, SBP)

p-value is smaller than 0.01, so reject.

H02: behtype does not have significant effect on the dependent variables(age, ht, wt, SBP)

p-value is smaller than 0.01, so reject.

**c**

H0: the interaction of chd and behtype does not have significant effect on the dependent variables(age, ht, wt, SBP)

p-value is larger than 0.01, so fail to reject.

**codes**

> fit = manova(Y ~ chd\*behtype)

> summary(fit, test="Wilks") # ANOVA table of Wilks' lambda

Df Wilks approx F num Df den Df Pr(>F)

chd 1 0.97076 23.6950 4 3147 < 2.2e-16 \*\*\*

behtype 1 0.99017 7.8104 4 3147 2.924e-06 \*\*\*

chd:behtype 1 0.99909 0.7155 4 3147 0.5812

Residuals 3150

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

2.

**a** first principal component: linear combination that maximizes Var() subject to =1



**b**

Proportion of Variance explained by pc1 to pc4 is: 0.3998 0.2958 0.2003 0.1042

**codes**

> pca1 = prcomp(Y,scale=T) # Talk about the scale option

>

> # results deal use scale =T option

>

> summary(pca1)

Importance of components:

PC1 PC2 PC3 PC4

Standard deviation 1.2646 1.0877 0.8950 0.6455

Proportion of Variance 0.3998 0.2958 0.2003 0.1042

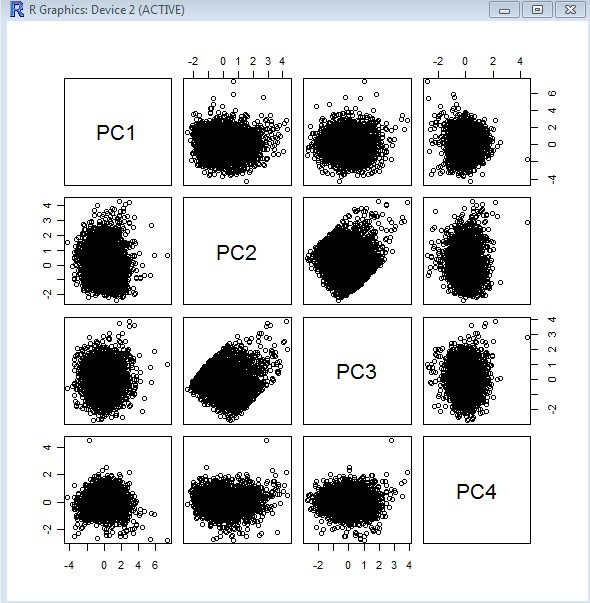
Cumulative Proportion 0.3998 0.6956 0.8958 1.0000

**c** PC1 could explain 39.98% of the total variance, PC3 could explain 20.03% of the total variance

PC1 is the linear combination of Y ( denoted as  ) that maximize var() subject to 

PC3 is the linear combination of Y ( denoted as  ) that maximize var() subject to , and cov(,)=0, where i=1,2.

**d**



Codes

pairs(pca1$x)

**e**

H0: CHD status does not have significant effect on the variance explained by each principal component

According to the p-values, reject H0 of PC1 and 2, fail to reject PC3 and 4. And conclude that CHD has significant effect on PC1 and PC2, but CHD has no significant effect on PC3 and PC4.

Code

glmall=glm(wcgsmatrix[,5]~pca1$x[,1]+pca1$x[,2]+pca1$x[,3]+pca1$x[,4],family=binomial)

> summary(glmall)

Call:

glm(formula = wcgsmatrix[, 5] ~ pca1$x[, 1] + pca1$x[, 2] + pca1$x[,

3] + pca1$x[, 4], family = binomial)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.0743 -0.4415 -0.3591 -0.2951 2.6631

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -2.560724 0.072591 -35.276 < 2e-16 \*\*\*

pca1$x[, 1] 0.191567 0.050918 3.762 0.000168 \*\*\*

pca1$x[, 2] 0.466038 0.055995 8.323 < 2e-16 \*\*\*

pca1$x[, 3] -0.068281 0.067887 -1.006 0.314505

pca1$x[, 4] 0.005301 0.096805 0.055 0.956332

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1781.2 on 3153 degrees of freedom

Residual deviance: 1696.6 on 3149 degrees of freedom

AIC: 1706.6

Number of Fisher Scoring iterations: 5