Final Exam

Spencer Swartz

May 4, 2017

## Problem 1

### Part 1a

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1a -|-|-|-|-|-|-|-|-|-|-|-

require(DS705data)

## Loading required package: DS705data

load("careerbarrier.rda")  
require(psych)

## Loading required package: psych

## Warning: package 'psych' was built under R version 3.3.3

mat <- cor(careerbarrier)  
cortest.bartlett(mat,n=76)

## $chisq  
## [1] 287.4985  
##   
## $p.value  
## [1] 6.756016e-19  
##   
## $df  
## [1] 105

1. Null Hypothesis: The correlation matrix is the identity matrix. Alternative Hypothesis: The correlation matrix is not the identity matrix.
2. Reject the null hypothesis (p=0) at a 5% level of significance. The correlation matrix is not the identity matrix.
3. p=0

### Part 1b1

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1b1 -|-|-|-|-|-|-|-|-|-|-|-

KMO(mat)

## Kaiser-Meyer-Olkin factor adequacy  
## Call: KMO(r = mat)  
## Overall MSA = 0.68  
## MSA for each item =   
## money lazy law noexp math support health reading   
## 0.36 0.67 0.67 0.61 0.67 0.66 0.67 0.73   
## english aoda grades disc social relatshp looks   
## 0.69 0.82 0.67 0.66 0.75 0.72 0.58

1. MSA = 0.68
2. Yes
3. Yes
4. money

### Part 1b2

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1b2 -|-|-|-|-|-|-|-|-|-|-|-

mat <- cor(careerbarrier[,2:15])  
KMO(mat)

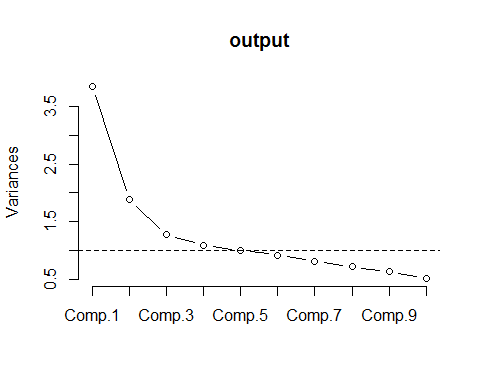
## Kaiser-Meyer-Olkin factor adequacy  
## Call: KMO(r = mat)  
## Overall MSA = 0.7  
## MSA for each item =   
## lazy law noexp math support health reading english   
## 0.68 0.67 0.61 0.68 0.66 0.68 0.74 0.70   
## aoda grades disc social relatshp looks   
## 0.82 0.67 0.70 0.77 0.71 0.62

1. MSA = 0.7
2. Yes
3. No
4. none

### Part 1c

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1c -|-|-|-|-|-|-|-|-|-|-|-

output <- princomp(careerbarrier[,2:15], cor=TRUE)  
plot(output,type="lines") # scree plot   
abline(h=1,lty=2)



output

## Call:  
## princomp(x = careerbarrier[, 2:15], cor = TRUE)  
##   
## Standard deviations:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7   
## 1.9629161 1.3732443 1.1271731 1.0418512 1.0003923 0.9586789 0.8989851   
## Comp.8 Comp.9 Comp.10 Comp.11 Comp.12 Comp.13 Comp.14   
## 0.8453388 0.7924190 0.7173535 0.6392802 0.5998518 0.5417063 0.5080299   
##   
## 14 variables and 76 observations.

1. 3
2. 5

### Part 1d

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1d -|-|-|-|-|-|-|-|-|-|-|-

fan <- principal(careerbarrier[,2:15],nfactors=5,rotate="varimax")  
print(fan,cut=.5,sort=TRUE)

## Principal Components Analysis  
## Call: principal(r = careerbarrier[, 2:15], nfactors = 5, rotate = "varimax")  
## Standardized loadings (pattern matrix) based upon correlation matrix  
## item RC2 RC1 RC3 RC4 RC5 h2 u2 com  
## looks 14 0.79 0.67 0.33 1.2  
## relatshp 13 0.76 0.65 0.35 1.2  
## disc 11 0.71 0.53 0.47 1.1  
## social 12 0.59 0.50 0.50 1.9  
## reading 7 0.82 0.78 0.22 1.3  
## english 8 0.74 0.67 0.33 1.4  
## health 6 0.63 0.59 0.41 2.0  
## aoda 9 0.58 0.59 0.41 2.2  
## lazy 1 0.81 0.77 0.23 1.4  
## law 2 0.76 0.65 0.35 1.3  
## grades 10 0.65 0.53 0.47 1.5  
## support 5 0.82 0.71 0.29 1.1  
## math 4 0.64 0.60 0.40 2.0  
## noexp 3 0.89 0.85 0.15 1.1  
##   
## RC2 RC1 RC3 RC4 RC5  
## SS loadings 2.28 2.28 2.11 1.27 1.15  
## Proportion Var 0.16 0.16 0.15 0.09 0.08  
## Cumulative Var 0.16 0.33 0.48 0.57 0.65  
## Proportion Explained 0.25 0.25 0.23 0.14 0.13  
## Cumulative Proportion 0.25 0.50 0.73 0.87 1.00  
##   
## Mean item complexity = 1.5  
## Test of the hypothesis that 5 components are sufficient.  
##   
## The root mean square of the residuals (RMSR) is 0.08   
## with the empirical chi square 96.17 with prob < 1.4e-08   
##   
## Fit based upon off diagonal values = 0.89

1. 65%
2. yes >60%

## Problem 2

### Part 2a

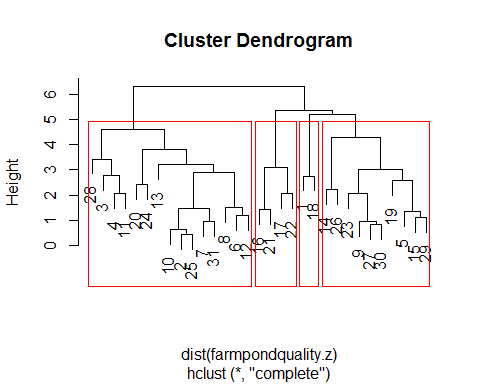
### -|-|-|-|-|-|-|-|-|-|-|- Answer 2a -|-|-|-|-|-|-|-|-|-|-|-

load("farmpondquality.rda")  
farmpondquality.z <- scale(farmpondquality)

### Part 2b

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2b -|-|-|-|-|-|-|-|-|-|-|-

output <- hclust(dist(farmpondquality.z))  
plot(output)  
rect.hclust(output,h=5)



1. 4

### Part 2c

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2c -|-|-|-|-|-|-|-|-|-|-|-

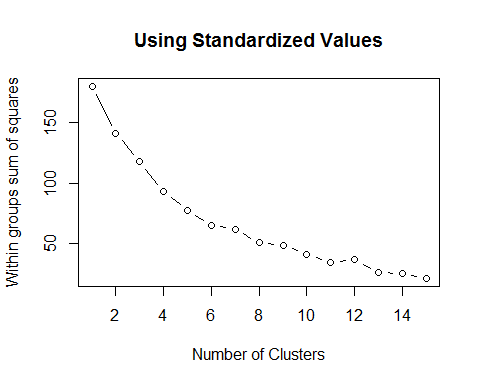
farmpondquality$clusternumber <- cutree(output, h=5)  
summary(as.factor(farmpondquality$clusternumber))

## 1 2 3 4   
## 2 15 10 4

round(aggregate(farmpondquality,by=list(farmpondquality$clusternumber),FUN=mean),2)

## Group.1 nitro turb phos disoxy temp cond clusternumber  
## 1 1 0.79 13.77 2.51 9.51 20.62 346.75 1  
## 2 2 0.30 28.45 1.88 9.22 23.08 149.63 2  
## 3 3 0.18 9.09 1.63 11.51 20.41 206.72 3  
## 4 4 0.18 9.18 0.93 8.55 22.57 488.56 4

wss <- (nrow(farmpondquality.z)-1)\*sum(apply(farmpondquality.z,2,var))  
for (i in 2:15) wss[i] <- sum(kmeans(farmpondquality.z,  
 centers=i)$withinss)  
plot(1:15, wss, type="b", xlab="Number of Clusters",  
 ylab="Within groups sum of squares",  
 main="Using Standardized Values")



1. It looks to me like the knee is at 4 so we will use 4 clusters.

set.seed(0)  
fit <- kmeans(farmpondquality.z,centers = 4,nstart = 25)  
fit

## K-means clustering with 4 clusters of sizes 13, 4, 7, 7  
##   
## Cluster means:  
## nitro turb phos disoxy temp cond  
## 1 -0.3730094 -0.2546998 -0.6019295 0.4118078 -0.14663186 -0.63960559  
## 2 0.5698533 2.1170857 0.1146791 -1.6561973 0.36646764 -0.51357760  
## 3 0.2259235 -0.5603824 -0.5365443 -0.4606866 -0.02940373 1.40363912  
## 4 0.1411777 -0.1763670 1.5888824 0.6422992 0.09230995 0.07767275  
##   
## Clustering vector:  
## [1] 4 1 2 2 1 4 1 4 1 1 2 4 4 4 1 3 3 3 3 3 3 3 1 1 1 4 1 2 1 1 1  
##   
## Within cluster sum of squares by cluster:  
## [1] 23.43302 12.77028 31.40550 24.51009  
## (between\_SS / total\_SS = 48.8 %)  
##   
## Available components:  
##   
## [1] "cluster" "centers" "totss" "withinss"   
## [5] "tot.withinss" "betweenss" "size" "iter"   
## [9] "ifault"

round(aggregate(farmpondquality,by=list(fit$cluster),FUN=mean),2)

## Group.1 nitro turb phos disoxy temp cond clusternumber  
## 1 1 0.21 14.50 1.13 10.81 21.71 128.40 2.54  
## 2 2 0.39 54.28 1.83 6.22 22.72 147.33 2.00  
## 3 3 0.32 9.38 1.20 8.87 21.94 435.39 3.14  
## 4 4 0.31 15.82 3.25 11.32 22.18 236.16 2.14

## Problem 3

### Part 3a

### -|-|-|-|-|-|-|-|-|-|-|- Answer 3a -|-|-|-|-|-|-|-|-|-|-|-

Naus <- c(98,110,161)  
NoNaus <- c(264,321,280)  
NausTable <- rbind(Naus,NoNaus)  
dimnames(NausTable) <- list( pheno = c("Nausea","No Nausea"),type =c ("Front","Middle","Rear") )  
out <- chisq.test(NausTable,correct=FALSE)  
out

##   
## Pearson's Chi-squared test  
##   
## data: NausTable  
## X-squared = 14.509, df = 2, p-value = 0.000707

min(out$expected)

## [1] 108.248

1. : Nausea and Seat Position are independent.

: Nausea and Seat Position are associated.

1. x-sqr = 14.509
2. df = 2
3. p = 0.0007
4. Reject at a 5% level of significance. There is sufficient evidence to conclude that Nausea and Seat Position are associated.

### Part 3b

### -|-|-|-|-|-|-|-|-|-|-|- Answer 3b -|-|-|-|-|-|-|-|-|-|-|-

prop.test(98,362,conf.level=.9,correct=FALSE)

##   
## 1-sample proportions test without continuity correction  
##   
## data: 98 out of 362, null probability 0.5  
## X-squared = 76.122, df = 1, p-value < 2.2e-16  
## alternative hypothesis: true p is not equal to 0.5  
## 90 percent confidence interval:  
## 0.2341110 0.3107273  
## sample estimates:  
## p   
## 0.2707182

1. 0.234
2. 0.311
3. With 95% confidence, the population proportion of people that sit in the front and report nausea is between 0.234 and 0.311.

That is, between 23.4% and 31.1% of people that sit in the front report nausea, with 95% confidence.

### Part 3c

### -|-|-|-|-|-|-|-|-|-|-|- Answer 3c -|-|-|-|-|-|-|-|-|-|-|-

(161/98)/(280/264)

## [1] 1.54898

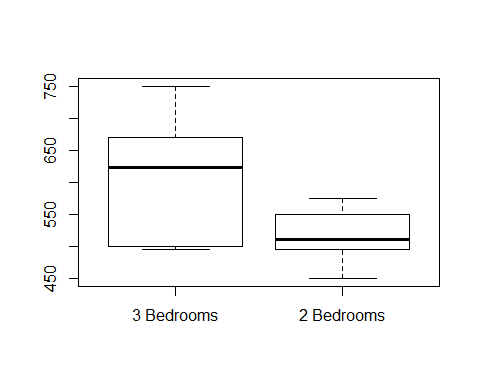
1. 1.549
2. The odds of having nausea for someone who is sitting in the rear are 1.549 times as high as for someone who is sitting in the front.

## Problem 4

### Part 4a

### -|-|-|-|-|-|-|-|-|-|-|- Answer 4a -|-|-|-|-|-|-|-|-|-|-|-

load("rent.rda")  
boxplot(monthlyrent$rent~monthlyrent$type)



1. The two groups look to have fairly normally shapes and distributions. There are no outliers in either group. That being said it looks like the range of rent between the two groups vary, the 3 bedroom apartments range from ~$500 to ~$750, where as the 2 bedrooms range from ~$450 to ~$575.

### Part 4b

### -|-|-|-|-|-|-|-|-|-|-|- Answer 4b -|-|-|-|-|-|-|-|-|-|-|-

b2 <- monthlyrent$rent[monthlyrent$type=="2 Bedrooms"]  
b3 <- monthlyrent$rent[monthlyrent$type=="3 Bedrooms"]  
shapiro.test(b2)

##   
## Shapiro-Wilk normality test  
##   
## data: b2  
## W = 0.96512, p-value = 0.8423

shapiro.test(b3)

##   
## Shapiro-Wilk normality test  
##   
## data: b3  
## W = 0.90679, p-value = 0.2596

1. For both populations we do not reject the null hypothesis (b2-pval = 0.8423, b3-pval = 0.2596), The distributions of the two different appartment styles seem to be normal.

### Part 4c

### -|-|-|-|-|-|-|-|-|-|-|- Answer 4c -|-|-|-|-|-|-|-|-|-|-|-

require(car)

## Loading required package: car

## Warning: package 'car' was built under R version 3.3.2

##   
## Attaching package: 'car'

## The following object is masked from 'package:psych':  
##   
## logit

test <- leveneTest(monthlyrent$rent~monthlyrent$type)  
test

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 1 7.5154 0.01342 \*  
## 18   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

1. Let = population variance in the rent for two bedroom apartments. Let = population variance in the rent for three bedroom apartments.
2. p = 0.0134
3. Reject at (P = 0.0134) indicating that there is evidence to show that the population variances, of rent, are different for two and three bedroom apartments.

### Part 4d

### -|-|-|-|-|-|-|-|-|-|-|- Answer 4d -|-|-|-|-|-|-|-|-|-|-|-

t.test(monthlyrent$rent~monthlyrent$type, var.equal=F, conf.level=0.9)

##   
## Welch Two Sample t-test  
##   
## data: monthlyrent$rent by monthlyrent$type  
## t = 3.0105, df = 12.32, p-value = 0.01057  
## alternative hypothesis: true difference in means is not equal to 0  
## 90 percent confidence interval:  
## 37.65164 146.34836  
## sample estimates:  
## mean in group 3 Bedrooms mean in group 2 Bedrooms   
## 608 516

1. t = 3.0105
2. df = 12.32
3. p-value = 0.01057
4. Reject at (P = 0.0106). There is significant evidence to suggest that population mean rent is different for two and three bedroom apartments.
5. We are 90% confident that the population mean rent for two bedroom apartments is between $37.65 and $146.35 different than the population mean rent of three bedroom apartments.

## Problem 5

### Part 5a

### -|-|-|-|-|-|-|-|-|-|-|- Answer 5a -|-|-|-|-|-|-|-|-|-|-|-

load("gpa7th.rda")  
fit <- lm(y~x1+x2+x3+x1:x3+x2:x3,data=gpa7th)  
summary(fit)

##   
## Call:  
## lm(formula = y ~ x1 + x2 + x3 + x1:x3 + x2:x3, data = gpa7th)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.20709 -0.30514 0.03727 0.25057 1.15223   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.963561 1.221225 -2.427 0.0209 \*   
## x1 0.040486 0.008774 4.614 5.72e-05 \*\*\*  
## x2 0.031657 0.011863 2.669 0.0117 \*   
## x3Auditory 2.074237 1.652304 1.255 0.2182   
## x1:x3Auditory -0.034406 0.015210 -2.262 0.0304 \*   
## x2:x3Auditory 0.002440 0.015614 0.156 0.8768   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4805 on 33 degrees of freedom  
## Multiple R-squared: 0.6653, Adjusted R-squared: 0.6146   
## F-statistic: 13.12 on 5 and 33 DF, p-value: 4.572e-07

1. Yes
2. x3 and x2x3

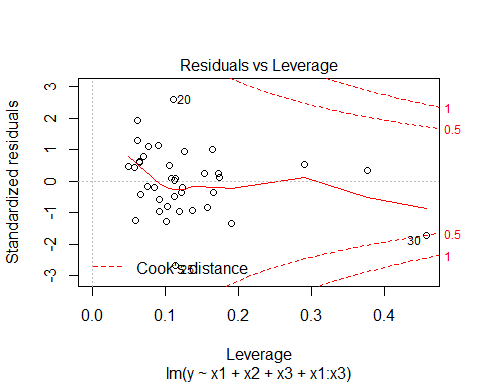
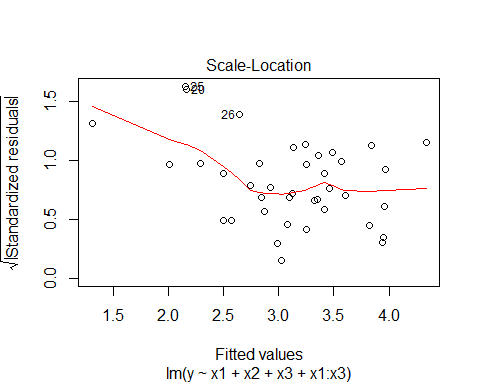
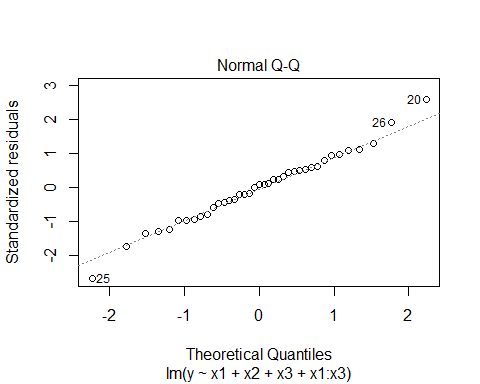
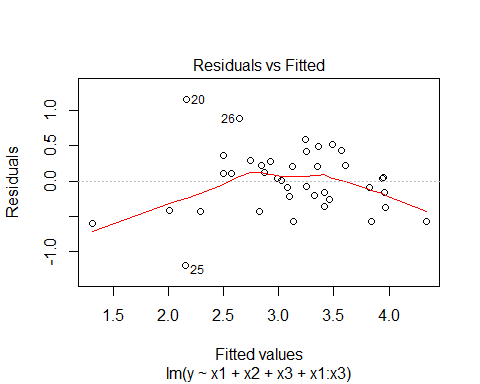
### Part 5b

### -|-|-|-|-|-|-|-|-|-|-|- Answer 5b -|-|-|-|-|-|-|-|-|-|-|-

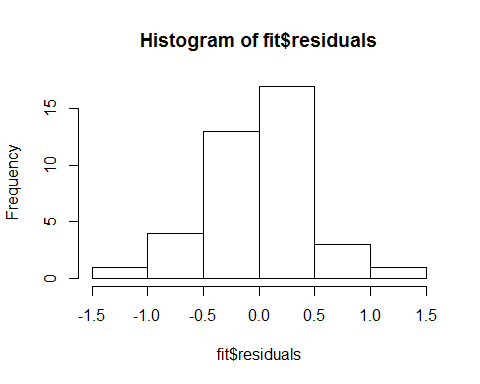
fit <- lm(y~x1+x2+x3+x1:x3,data=gpa7th)  
summary(fit)

##   
## Call:  
## lm(formula = y ~ x1 + x2 + x3 + x1:x3, data = gpa7th)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.18700 -0.30971 0.03875 0.24889 1.15463   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.097275 0.858682 -3.607 0.000983 \*\*\*  
## x1 0.040035 0.008166 4.903 2.3e-05 \*\*\*  
## x2 0.033065 0.007602 4.350 0.000118 \*\*\*  
## x3Auditory 2.296831 0.825052 2.784 0.008710 \*\*   
## x1:x3Auditory -0.033512 0.013890 -2.413 0.021377 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4736 on 34 degrees of freedom  
## Multiple R-squared: 0.665, Adjusted R-squared: 0.6256   
## F-statistic: 16.88 on 4 and 34 DF, p-value: 1.035e-07

plot(fit)



hist(fit$residuals)



require(lmtest)

## Loading required package: lmtest

## Warning: package 'lmtest' was built under R version 3.3.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 3.3.3

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

bptest(fit)

##   
## studentized Breusch-Pagan test  
##   
## data: fit  
## BP = 9.0755, df = 4, p-value = 0.05924

1. No striking patterns seem to be present in the residual plots to indicate that the model assumptions of independence, homogeneity of variance, and normality are not met.
2. At Alpha= 0.05, There is insufficient evidence to claim the residuals do not have equal variance (p=0.05924)

### Part 5c

### -|-|-|-|-|-|-|-|-|-|-|- Answer 5c -|-|-|-|-|-|-|-|-|-|-|-

fit <- lm(y~x1+x2+x3+x1:x3,data=gpa7th)  
newdata <- data.frame(x1=50,x2=105,x3="Auditory")  
predict.lm(fit,newdata,interval="confidence")

## fit lwr upr  
## 1 2.997526 2.696134 3.298919

1. 2.696
2. 3.299
3. With 95% confidence, the confidence interval for GPA of a randomly selected student with x1=50,x2=105, and x3=Auditory is between 2.696 and 3.299.

### Part 5d

### -|-|-|-|-|-|-|-|-|-|-|- Answer 5d -|-|-|-|-|-|-|-|-|-|-|-

confint(fit)

## 2.5 % 97.5 %  
## (Intercept) -4.84232757 -1.352222970  
## x1 0.02343918 0.056630368  
## x2 0.01761615 0.048513730  
## x3Auditory 0.62012444 3.973538015  
## x1:x3Auditory -0.06173917 -0.005284324

With 95% confidence, the students gpa increases by 0.018 to 0.049 points for each 1 unit increase in x2.

### Part 5e

### -|-|-|-|-|-|-|-|-|-|-|- Answer 5e -|-|-|-|-|-|-|-|-|-|-|-

out <- glm(x3~y+x1+x2,data=gpa7th,family="binomial")  
summary(out)

##   
## Call:  
## glm(formula = x3 ~ y + x1 + x2, family = "binomial", data = gpa7th)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.6142 -1.0610 -0.4663 1.0956 1.6013   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 4.86116 3.88495 1.251 0.2108   
## y 1.53479 0.80143 1.915 0.0555 .  
## x1 -0.03143 0.03968 -0.792 0.4282   
## x2 -0.07116 0.04236 -1.680 0.0930 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 53.834 on 38 degrees of freedom  
## Residual deviance: 48.659 on 35 degrees of freedom  
## AIC: 56.659  
##   
## Number of Fisher Scoring iterations: 4

newdata <- data.frame(y=3.5,x1=50,x2=105)  
predict(out, newdata, type="response")

## 1   
## 0.7665911

## Problem 6

### Part 6a

### -|-|-|-|-|-|-|-|-|-|-|- Answer 6a -|-|-|-|-|-|-|-|-|-|-|-

require(MVN)

## Loading required package: MVN

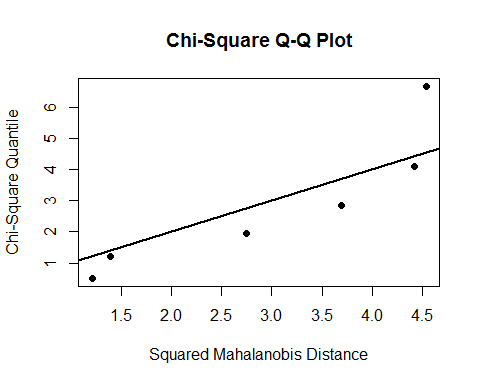
## Warning: package 'MVN' was built under R version 3.3.3

## sROC 0.1-2 loaded

##   
## Attaching package: 'MVN'

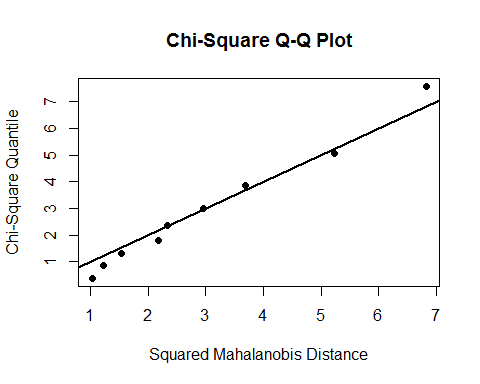
## The following object is masked from 'package:psych':  
##   
## mardia

load("beef.rda")  
store10 <- as.matrix(beef[beef$STORAGE==10,2:4])  
store40 <- as.matrix(beef[beef$STORAGE==40,2:4])  
store80 <- as.matrix(beef[beef$STORAGE==80,2:4])  
store120 <- as.matrix(beef[beef$STORAGE==120,2:4])  
hzTest(store10,qqplot=TRUE)



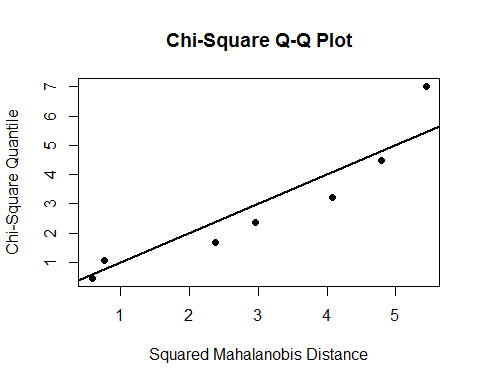
## Henze-Zirkler's Multivariate Normality Test   
## ---------------------------------------------   
## data : store10   
##   
## HZ : 0.5267688   
## p-value : 0.2292891   
##   
## Result : Data are multivariate normal.   
## ---------------------------------------------

hzTest(store40,qqplot=TRUE)



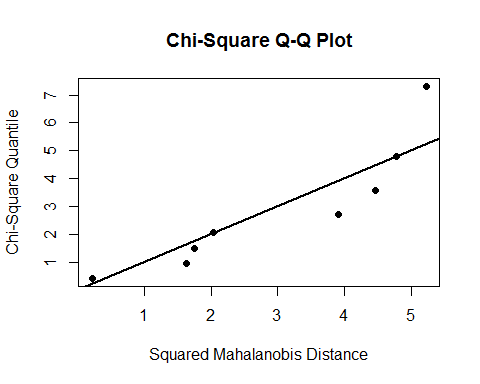
## Henze-Zirkler's Multivariate Normality Test   
## ---------------------------------------------   
## data : store40   
##   
## HZ : 0.6218614   
## p-value : 0.1493898   
##   
## Result : Data are multivariate normal.   
## ---------------------------------------------

hzTest(store80,qqplot=TRUE)



## Henze-Zirkler's Multivariate Normality Test   
## ---------------------------------------------   
## data : store80   
##   
## HZ : 0.4524749   
## p-value : 0.4813074   
##   
## Result : Data are multivariate normal.   
## ---------------------------------------------

hzTest(store120,qqplot=TRUE)



## Henze-Zirkler's Multivariate Normality Test   
## ---------------------------------------------   
## data : store120   
##   
## HZ : 0.4840055   
## p-value : 0.4271044   
##   
## Result : Data are multivariate normal.   
## ---------------------------------------------

1. All of the hypothesis tests indicate that the data could be sampled from multivariate normal distributions. The qqplots seem somewhat linear, Given the robustness of MANOVA it is probably safe to assume multivariate normality.

### Part 6b

### -|-|-|-|-|-|-|-|-|-|-|- Answer 6b -|-|-|-|-|-|-|-|-|-|-|-

1. No

source('BoxMTest.R')  
out<-BoxMTest(as.matrix(beef[,2:4]),beef$STORAGE)

##   
## ------------------------------------------------------------  
## MBox F df1 df2 P  
## ------------------------------------------------------------  
## 40.7426 1.7548 18 1901 0.0255  
## ------------------------------------------------------------  
## Covariance matrices are significantly different.

1. Yes

### Part 6c

### -|-|-|-|-|-|-|-|-|-|-|- Answer 6c -|-|-|-|-|-|-|-|-|-|-|-

mv.out <- manova(cbind(beef$BEEFY,beef$BLOODY,beef$GRASSY)~beef$STORAGE)  
summary(mv.out,test="Wilks")

## Df Wilks approx F num Df den Df Pr(>F)   
## beef$STORAGE 3 0.1269 8.6897 9 58.56 3.816e-08 \*\*\*  
## Residuals 26   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

1. : The population mean vectors for beefy, bloody and grassy aromas are equal for the 4 storage times.

: At least one mean vector differs from another.

1. Reject at (p=3.816e-8). There is sufficient evidence in the sample to conclude that at least one mean vector differs from another.
2. 0

### Part 6d

### -|-|-|-|-|-|-|-|-|-|-|- Answer 6d -|-|-|-|-|-|-|-|-|-|-|-

summary.aov(mv.out)

## Response 1 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## beef$STORAGE 3 322.55 107.516 37.192 1.491e-09 \*\*\*  
## Residuals 26 75.16 2.891   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response 2 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## beef$STORAGE 3 91.744 30.5814 33.948 3.837e-09 \*\*\*  
## Residuals 26 23.422 0.9008   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response 3 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## beef$STORAGE 3 51.727 17.2422 45.563 1.713e-10 \*\*\*  
## Residuals 26 9.839 0.3784   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

beefy.out <- aov(BEEFY~STORAGE,data = beef)  
TukeyHSD(beefy.out,"STORAGE")

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = BEEFY ~ STORAGE, data = beef)  
##   
## $STORAGE  
## diff lwr upr p adj  
## 40-10 -1.171283 -3.629593 1.2870267 0.5668196  
## 80-10 -6.212302 -8.807288 -3.6173169 0.0000033  
## 120-10 -7.881942 -10.400960 -5.3629235 0.0000000  
## 80-40 -5.041019 -7.391613 -2.6904251 0.0000189  
## 120-40 -6.710658 -8.977108 -4.4442085 0.0000001  
## 120-80 -1.669639 -4.083652 0.7443732 0.2537189

bloody.out <- aov(BLOODY~STORAGE,data=beef)  
TukeyHSD(bloody.out,"STORAGE")

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = BLOODY ~ STORAGE, data = beef)  
##   
## $STORAGE  
## diff lwr upr p adj  
## 40-10 -0.5138894 -1.886182 0.8584034 0.7351889  
## 80-10 -3.2559536 -4.704542 -1.8073650 0.0000092  
## 120-10 -4.1458363 -5.552018 -2.7396546 0.0000001  
## 80-40 -2.7420641 -4.054227 -1.4299012 0.0000279  
## 120-40 -3.6319468 -4.897138 -2.3667554 0.0000001  
## 120-80 -0.8898827 -2.237447 0.4576821 0.2907963

grassy.out <- aov(GRASSY~STORAGE,data=beef)  
TukeyHSD(grassy.out,"STORAGE")

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = GRASSY ~ STORAGE, data = beef)  
##   
## $STORAGE  
## diff lwr upr p adj  
## 40-10 0.6527761 -0.2366643 1.542217 0.2090224  
## 80-10 2.3472207 1.4083298 3.286112 0.0000016  
## 120-10 3.3472213 2.4358160 4.258626 0.0000000  
## 80-40 1.6944446 0.8439769 2.544912 0.0000556  
## 120-40 2.6944451 1.8744216 3.514469 0.0000000  
## 120-80 1.0000005 0.1265874 1.873414 0.0203952

1. There is not enough evidence to conclude that differences in population mean beefy aroma as well as bloody aroma exist among storage times 40~10 and 120~80. In addtion there is not enough evidence to conclude that differences in population mean grassy aroma exist among the storage times 40~10. But with 95% confidence, we can see that the shorter the storage time the higher the score in relation to aroma between the remaining storage pairs.