Case Studies

"Data Analytics"

Topic

Summer Term 2013

FirstName LastName

July 10, 2013

Contents

1 Introduction

```
data = Glass[,-10]
par(mfrow = c(3, 3))
for(i in 1:9){
        newdata = (data[,i]-mean(data[,i]))/(sqrt(var(data[,i])))
        qqnorm(newdata,
                        main = paste("QQ-Plot of ", colnames(data)[i], sep = ""),
                        las = 1,
                        cex.main = 1, cex.lab = 1, cex.axis = 1)
        abline(0,1, col="red")
}

x = Glass[,-10]
cm = colMeans(x)
S = cov(x)
d = apply(x,1,function(x) t(x-cm) %*% solve(S) %*% (x-cm))
plot(qc <- qchisq((1:nrow(x)-1/2)/nrow(x), df=9),
     sd <- sort(d), xlab=expression(paste(chi[9]^2,'Quantile')),
     ylab='Ordered Distances', xlim=range(qc)*c(1,1.1),
     pch=19,cex.lab=2,cex.axis=2,cex=2)
out <- which(rank(abs(qc-sd), ties='random') > nrow(x)-3)
text(qc[out], sd[out]-1.5, names(out),cex=2,col='blue')
abline(a=0,b=1,col='red',lwd=2)

x = Glass.type7
cm = colMeans(x)
S = cov(x)
d = apply(x,1,function(x) t(x-cm) %*% solve(S) %*% (x-cm))
plot(qc <- qchisq((1:nrow(x)-1/2)/nrow(x), df=9),
     sd <- sort(d), xlab=expression(paste(chi[9]^2,'Quantile')),
     ylab='Ordered Distances', xlim=range(qc)*c(1,1.1),
     pch=19,cex.lab=2,cex.axis=2,cex=2)
out <- which(rank(abs(qc-sd), ties='random') > nrow(x)-3)
text(qc[out], sd[out]-1.5, names(out),cex=2,col='blue')
abline(a=0,b=1,col='red',lwd=2)
```