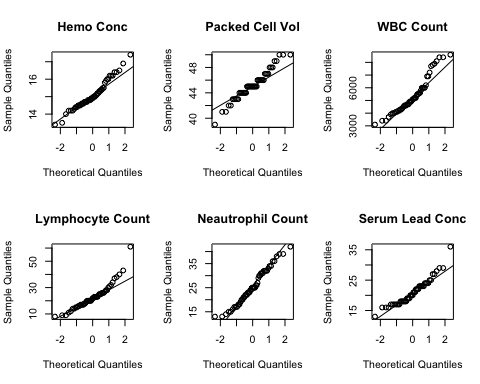
Homework 7

Tyler Poelking

4/5/2017

knitr::opts\_chunk$set(echo = TRUE)  
library(mvtnorm)  
  
plot.bivariate.normal <- function (mu, Sig, ...) {  
  
 x1 <- sqrt(Sig[1,1]) \* seq(-2.5, 2.5, length=50) + mu[1]  
 x2 <- sqrt(Sig[2,2]) \* seq(-2.5, 2.5, length=50) + mu[2]  
   
 xs <- expand.grid(x1, x2)  
  
 zs <- dmvnorm(xs, mu, Sig)  
   
 contour(x1, x2, matrix(zs, length(x1)),  
 xlab=expression(z[1]),  
 ylab=expression(z[2]), ...)  
  
 abline(v=mu[1], lty=2, col="blue")  
 abline(h=mu[2], lty=2, col="blue")  
}  
  
  
hemo <- read.table("http://www.stat.osu.edu/~pfc/teaching/3302/datasets/six\_hematology\_variables.txt")  
  
names(hemo) = c('Hemo Conc', 'Packed Cell Vol', 'WBC Count', 'Lymphocyte Count', 'Neautrophil Count', 'Serum Lead Conc')  
  
par(mfrow=c(2,3))  
  
for (j in 1:length(names(hemo))) {  
 qqnorm(hemo[,j], main=names(hemo)[j])  
 qqline(hemo[,j])  
}

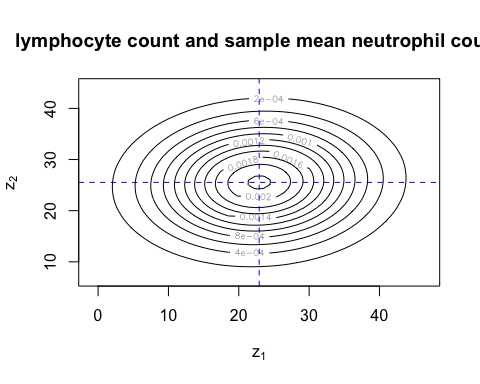


1. The above qq-plot is used to determines how closely the various variables in the data follow a normal distribution. The hematology variable most closely following a normal appears to be the Neatrophil Count, deviating fairly at the far tail ends of the theoretical quantiles. Hemo Count, White Blood Cell Count, and Lymphocyte Count follow the normal well until the larger theoretical quantiles are reached, where their sample quantiles tend to be larger than the theoretical. The qq-plot corresponding Packed Cell Count has a staggered, stair like structure, because it has less factor levels. It deviates in the sense that is has a larger slope than a theoretical normal distribution would have. Lastly, Serum Lead Concentration deviates from the normal on its tail ends, it's sample quantiles being larger on both the lower and upper theoretical quantile values.
2. The approximate joint distribution of the lymphocyte count and sample mean neutrophil count:

par(mfrow=c(1,1))  
colMeans(hemo)[c(4,5)]

## Lymphocyte Count Neautrophil Count   
## 22.90196 25.52941

hemN = hemo[4]  
hemN = cbind(hemN,hemo[5])  
covTab = cov(hemN)  
  
  
  
mu = c(colMeans(hemo)[c(4,5)][[1]], colMeans(hemo)[c(4,5)][[2]])  
sigma = matrix(c(covTab[1],covTab[2],covTab[3],covTab[4]), nrow = 2)  
  
plot.bivariate.normal(mu, sigma, main="lymphocyte count and sample mean neutrophil count")



The distribution appears to be uniform, since the oval is on a horizonal diagonal and symetrical across the blue axis's, with more variance spanning across z1 (neutrophil count).