

Homework 6

Tyler Poelking

3/31/2017

```
#Import data
bone_length <- read.table("http://www.stat.osu.edu/~pfc/teaching/3302/datasets/bone_length.txt", header=TRUE)

#remove 'individual' feat
bl = bone_length[, -1]
bl[, 1]

##      [1] 47.8 46.4 46.3 45.1 47.6 52.5 51.2 49.8 48.1 45.0 51.2 48.5 52.1 48.2
##     [15] 49.6 50.7 47.2 53.3 46.2 46.3

names(bl) = c('8', '8.5', '9', '9.5')
# (a) Find y-bar, S, and R
yBar = colMeans(bl)
yBar

##           8           8.5           9           9.5
## 48.655 49.625 50.570 51.445

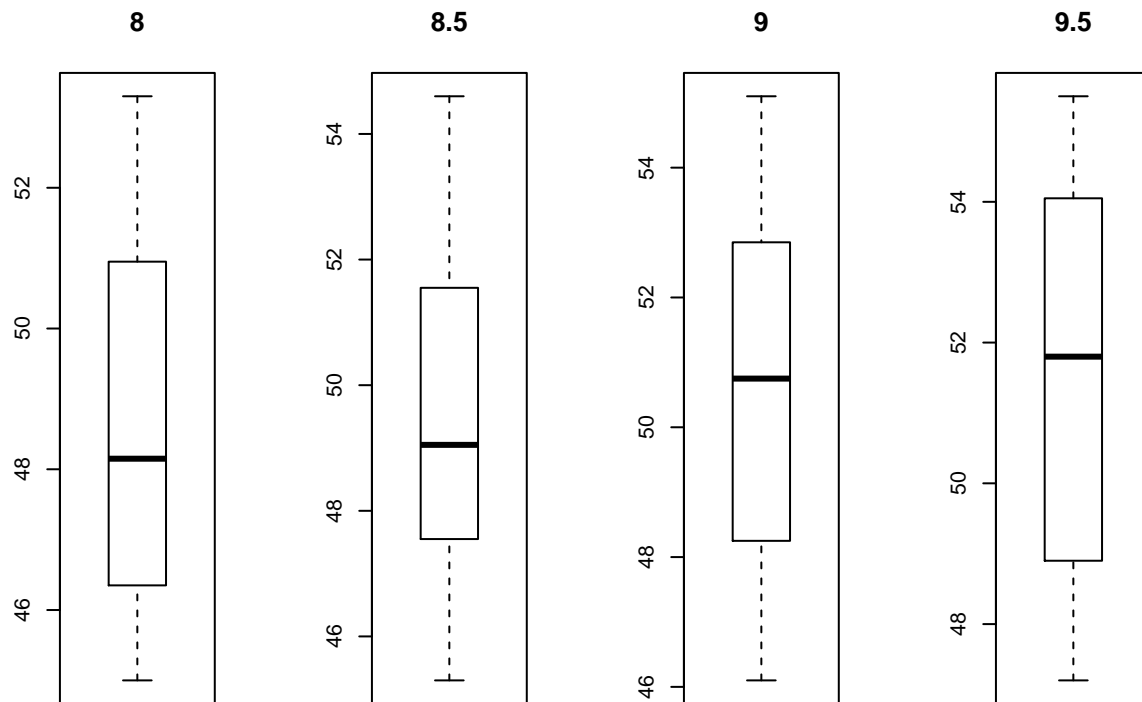
#covariance matrix
S = round(cov(bl), 2)
S

##           8  8.5    9  9.5
## 8    6.33 6.19 5.78 5.53
## 8.5  6.19 6.45 6.15 5.91
## 9    5.78 6.15 6.92 6.93
## 9.5  5.53 5.91 6.93 7.43

#Correlation matrix
R = round(cor(bl), 2)
R

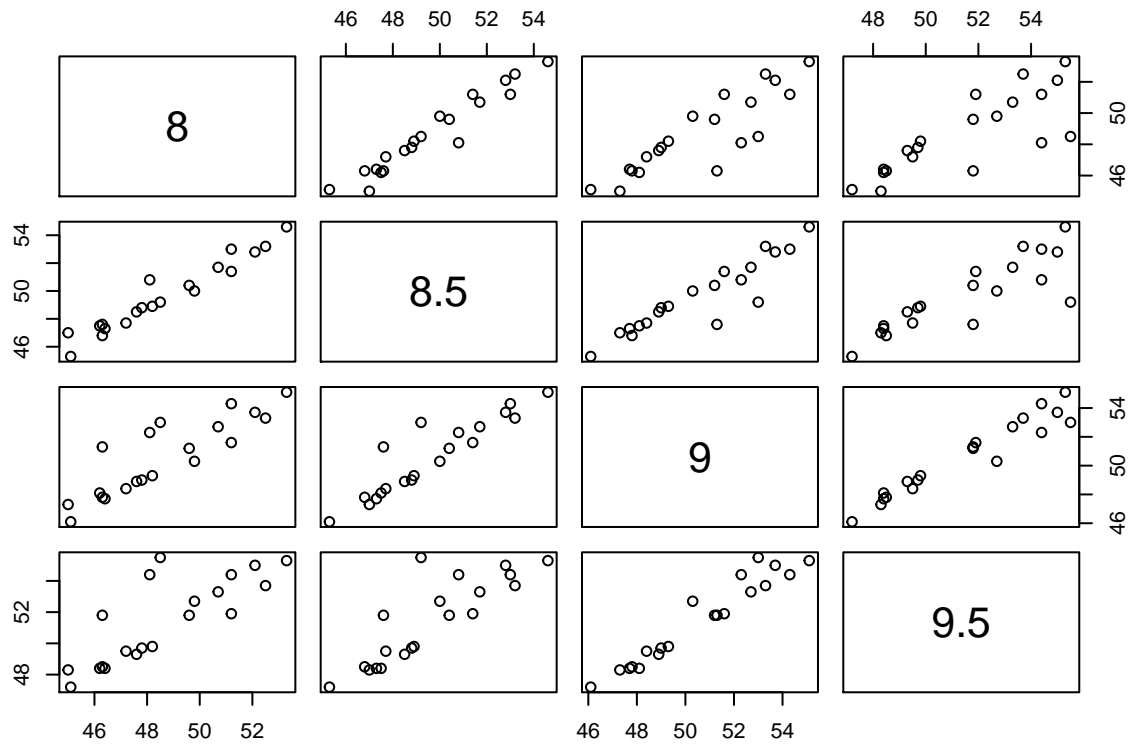
##           8  8.5    9  9.5
## 8    1.00 0.97 0.87 0.81
## 8.5  0.97 1.00 0.92 0.85
## 9    0.87 0.92 1.00 0.97
## 9.5  0.81 0.85 0.97 1.00

#marginal summaries
par(mfrow=c(1,4))
sapply(1:4, function(j) boxplot(bl[,j], main=names(bl)[j]))
```



```
##      [,1]      [,2]      [,3]      [,4]
## stats Numeric,5 Numeric,5 Numeric,5 Numeric,5
## n      20      20      20      20
## conf  Numeric,2 Numeric,2 Numeric,2 Numeric,2
## out   Numeric,0 Numeric,0 Numeric,0 Numeric,0
## group Numeric,0 Numeric,0 Numeric,0 Numeric,0
## names ""      ""      ""      ""
```

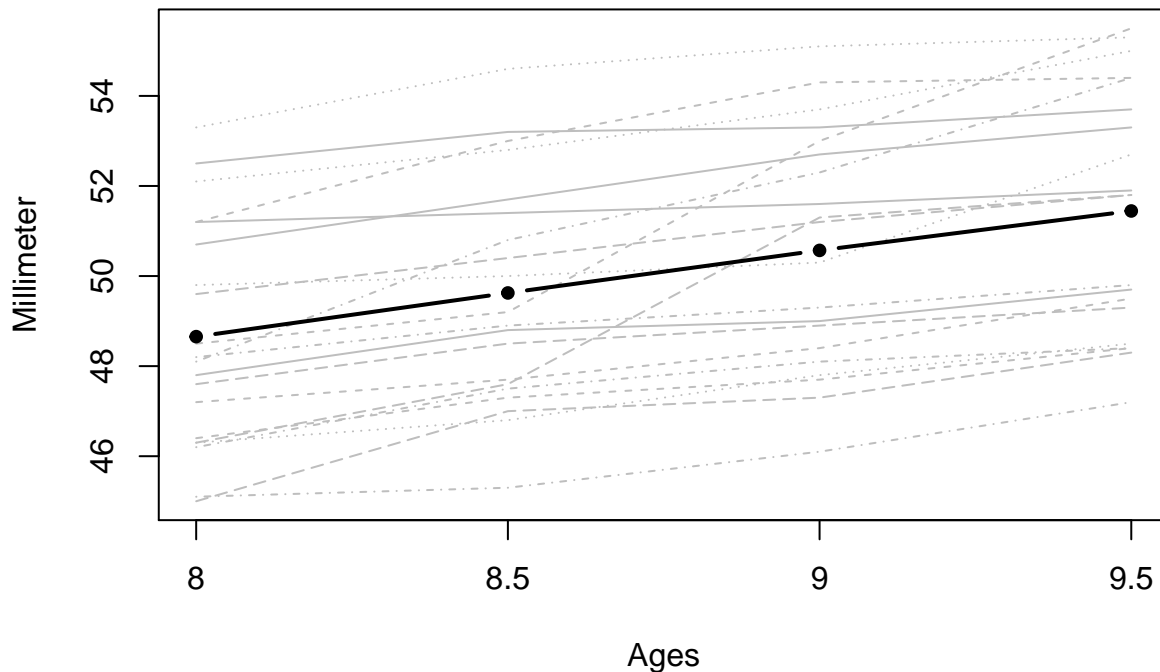
```
## Produce a pairs plots
par(mfrow=c(1,1))
pairs(bl)
```



#Parallel coordinate plot

```
matplot(t(bl), type="l", xaxt="n", col="gray",
        xlab="Ages", ylab="Millimeter", main = 'Ramus (Jaw) Bone Length at Four Ages For 20 Boys')
## Now we correctly label the x axis.
axis(side=1, at=1:4, names(bl))
## create and connect length means for each age
points(1:4, colMeans(bl), cex=1, pch=20, type="b", lwd=2)
```

Ramus (Jaw) Bone Length at Four Ages For 20 Boys



In the boxplots, the spread on all 4 numerical age group levels are relatively even, ranging from around 55 at the high end and 46 at the low end. The quartile deviations from the mean are around the same as well, except jaw length group corresponding to age 8.5's median is further away from its third quartile than it is from its first quartile. The other age group's median distances from their quartiles are uniform. The box plot for each age group contain no outliers. And most importantly, the medians increase with each increase in age group. The median jaw length (in millimeters) for ages 9.5 is larger than the median jaw length for ages 9, which is larger than the median jaw length for ages 8.5, and so on. This, of course, is evidence for jaw length increasing with age.

The scatter plot matrix shows the relationship between the jaw lengths, in millimeters, of each feature (in this case just age group). Each plot suggests a positive correlation between all age groups. The S and R matrix shows this to be true. Not surprisingly, correlation strength decreases the further away the age groups are from each other. The strength of positive relationship between age 8 and 8.5, for example, is stronger than the positive correlation between age 8 and 9 or age 8 and 9.5. So the jaw length at age 8 more accurately predicts the length at 8.5 than it does at 9 or 9.5.

The summaries and plot above attest to the fact that a boy's average ramus (jaw) length increases with age. As boys get older, this expected mm increase in jaw length every half year decreases. The average length for boys of age 8.5 is 0.97 mm larger than boys of 8. The average length for boys of age 9 is 0.945 mm larger than boys of 8.5. Finally, the average length for boys of age 9.5 is 0.875 mm larger than boys of 9.

```
blood_glu <- read.table("http://www.stat.osu.edu/~pfc/teaching/3302/datasets/blood_glucose.txt", header = TRUE)
```

```
#Mean Vector
```

```
yxBar = colMeans(blood_glu)
```

```
yxBar
```

```
##      y1      y2      y3      x1      x2      x3
```

```
## 70.08 73.54 75.10 109.68 104.24 109.98
```

```
#Covariance Matrix
```

```
S = round(cov(blood_glu),2)
```

S

```
##      y1    y2    y3    x1    x2    x3
## y1 95.54 17.61 12.18 60.52 23.00 62.84
## y2 17.61 73.19 14.25  5.73 61.28 -1.66
## y3 12.18 14.25 76.17 46.75 32.77 69.84
## x1 60.52  5.73 46.75 808.63 320.59 227.36
## x2 23.00 61.28 32.77 320.59 505.86 167.35
## x3 62.84 -1.66 69.84 227.36 167.35 508.71

bg_1st = blood_glu[,c(1,4)]
bg_2nd = blood_glu[,c(2,5)]
bg_3rd = blood_glu[,c(3,6)]

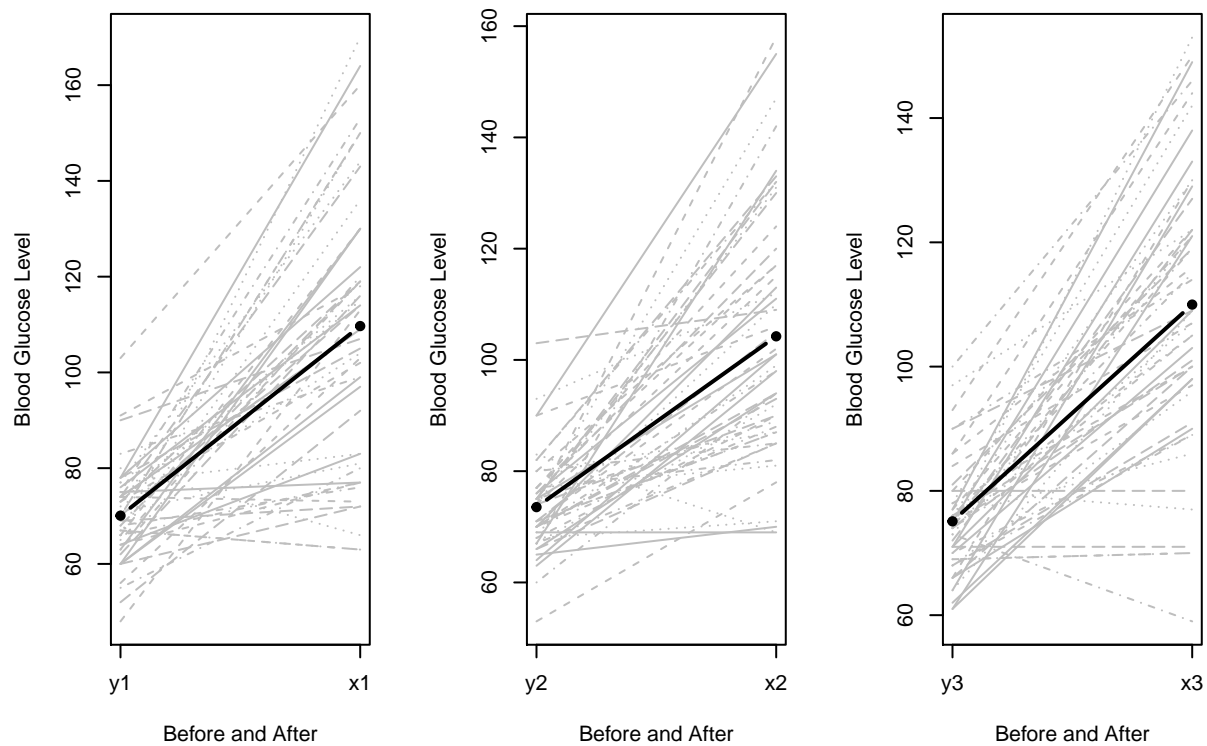
par(mfrow=c(1,3))

matplot(t(bg_1st), type="l", xaxt="n", col="gray",
        xlab="Before and After", ylab="Blood Glucose Level", main = 'First Pregnancy: Blood Glucose Level')
## Now we correctly label the x axis.
axis(side=1, at=1:2, names(bg_1st))
## create and connect length means for each age
points(1:2, colMeans(bg_1st), cex=1, pch=20, type="b", lwd=2)

matplot(t(bg_2nd), type="l", xaxt="n", col="gray",
        xlab="Before and After", ylab="Blood Glucose Level", main = 'Second Pregnancy: Blood Glucose Level')
## Now we correctly label the x axis.
axis(side=1, at=1:2, names(bg_2nd))
## create and connect length means for each age
points(1:2, colMeans(bg_2nd), cex=1, pch=20, type="b", lwd=2)

matplot(t(bg_3rd), type="l", xaxt="n", col="gray",
        xlab="Before and After", ylab="Blood Glucose Level", main = 'Third Pregnancy: Blood Glucose Level')
## Now we correctly label the x axis.
axis(side=1, at=1:2, names(bg_3rd))
## create and connect length means for each age
points(1:2, colMeans(bg_3rd), cex=1, pch=20, type="b", lwd=2)
```

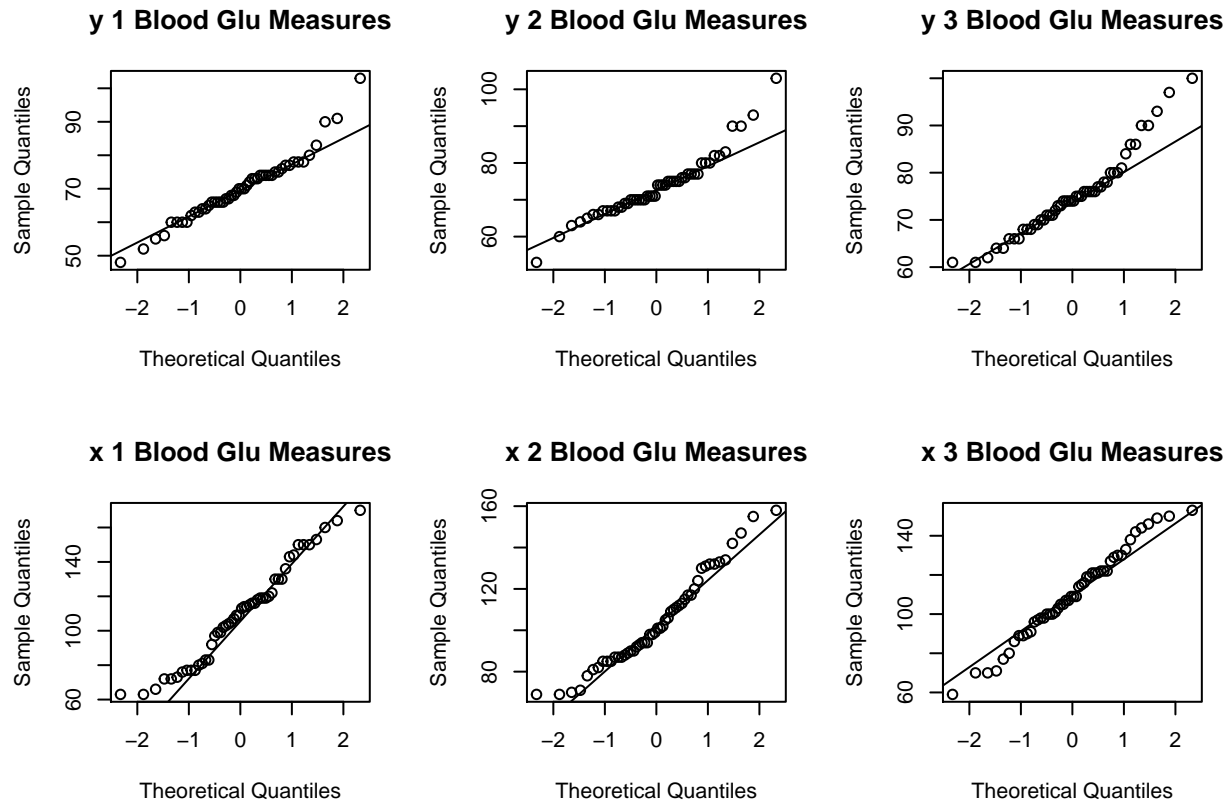
: Blood Glucose Level before ancy: Blood Glucose Level Before a/: Blood Glucose Level Before an



```
par(mfrow=c(2,3))

for (j in 1:3) {
  qqnorm(blood_glu[,j], main=sprintf("y %d Blood Glu Measures", j))
  qqline(blood_glu[,j])
}

for (j in 4:6) {
  qqnorm(blood_glu[,j], main=sprintf("x %d Blood Glu Measures", j-3))
  qqline(blood_glu[,j])
}
```



The summary statistics along with the graphical summaries above indicate that blood glucose level rises and hour after sugar intake for pregnant women. The mean vector shows that the base blood glucose level (the y's) before sugar intake increases with each pregnancy, since the average y2 is greater than the average y1, and the average y2 is greater than the average y3. The pairs plot indicates that there is a rise in average blood glucose level after sugar intake. The rise is greatest in the first pregnancy (39.61), least in the second pregnancy (30.7), and somewhere in between (34.88) in the last pregnancy.

A qqplot for each measurement/feature indicates how well each measurement of blood glucose level follows a normal distribution. The baseline blood glucose level measurements for the first and second pregnancies are similar. Both follow a normal distribution fairly well until the later theoretical quantiles are reached, at which point the dots fall above the normal line. The baseline blood glucose level measurement for the third pregnancy has this feature too, but the level in which the dots deviate from the normal line is further exaggerated (tails off even more).

The tail ends of all three post sugar intake qqplots deviate from the normal qqline. For these post sugar intake plots, the points at the lower theoretical quantiles for the first two pregnancies deviate above the normal qqline, where in the last plot corresponding to the third pregnancy, the points at the lower theoretical quantiles deviate below the normal qqline. This means that, for the third pregnancy, women whose blood level increases below the average increases even lower than if the measurement followed a normal distribution. This is of course reversed for women in their first and second pregnancy.