MANOVA - Spectral Reflectance

Greg Johnson

8/14/2017

We have replicated data on the spectral reflectance of different animal species over time and over two measurements: 560nm and 720nm.

```
spectral <- read.table("data/spectral_data.txt", header = FALSE, col.names = c("560nm",
    "720nm", "Species", "Time", "Replication"))
p <- 2
g <- 3
n <- nrow(spectral)
```

We're interested in differences in the two reflectance variables across our three species. The analysis we'll pursue is the one-way MANOVA, to evaluate multivariate differences in reflectance between our groups.

One-Way MANOVA

We want to test the research question that the three Species don't differ in their spectral reflectance, as measured at two different wavelengths, 560nm and 720nm. The corresponding statistical hypothesis is $H_0 : \tau_1 = tau_2$.

```
JL<-as.matrix(spectral[spectral$Species=="JL",c(1,2)])
LP<-as.matrix(spectral[spectral$Species=="LP",c(1,2)])
SS<-as.matrix(spectral[spectral$Species=="SS",c(1,2)])
spect<-rbind(JL,LP,SS)

xbar<-apply(spect,2,mean)
xbarJL<-apply(JL,2,mean)
xbarLP<-apply(LP,2,mean)
xbarSS<-apply(SS,2,mean)

B<-12*(xbarJL-xbar)%*%t(xbarJL-xbar)
B<-B+12*(xbarLP-xbar)%*%t(xbarLP-xbar)
B<-B+12*(xbarSS-xbar)%*%t(xbarSS-xbar)

W<-matrix(0,2,2)
Tot<-matrix(0,2,2)
for(i in 1:36){
  temp<-spect[i,]
  Tot<-Tot+(temp-xbar)%*%t(temp-xbar)
  if(i %in% 1:12){W<-W+(temp-xbarJL)%*%t(temp-xbarJL)}
  if(i %in% 13:24){W<-W+(temp-xbarLP)%*%t(temp-xbarLP)}
  if(i %in% 25:36){W<-W+(temp-xbarSS)%*%t(temp-xbarSS)}
}
```

Rencher (2002) suggests that when the outcome variables are highly correlated, Roy's Largest Root is the most powerful MANOVA test statistic. The correlation between the two wavelengths is $r = 0.81$, so the Largest Root should be more powerful than Wilk's Lambda.

$$\Lambda_{Roy} = max(\lambda_i) = ||BW^{-1}||_\infty$$

```
Roy<-max(eigen(B%*%solve(W))$values)
pf((n-2-1)*Roy/2,2,n-2-1,lower.tail=FALSE)
```

```
## [1] 0.002189691
```

Our analysis rejects the null hypothesis; the reflectance mean vector is significantly different between the three populations or in other words, at least one mean vector is different from the rest (Roy's Largest Root $= 0.45$, p-value $< .05$).

Just for corroboration, we compute Wilk's Lambda:

```
Lambda<-det(W)/det(B+W) #Wilk's Lambda

pf((n-g-1)/(g-1)*(1-sqrt(Lambda))/sqrt(Lambda),2*(g-1),2*(n-g-1),lower.tail=FALSE)
```

```
## [1] 0.0129992
```

```
#p-value less than alpha
```

Roy's Greatest Root and Wilk's Lambda agree, the three species significantly differ in spectral reflectance.

Model Evaluation

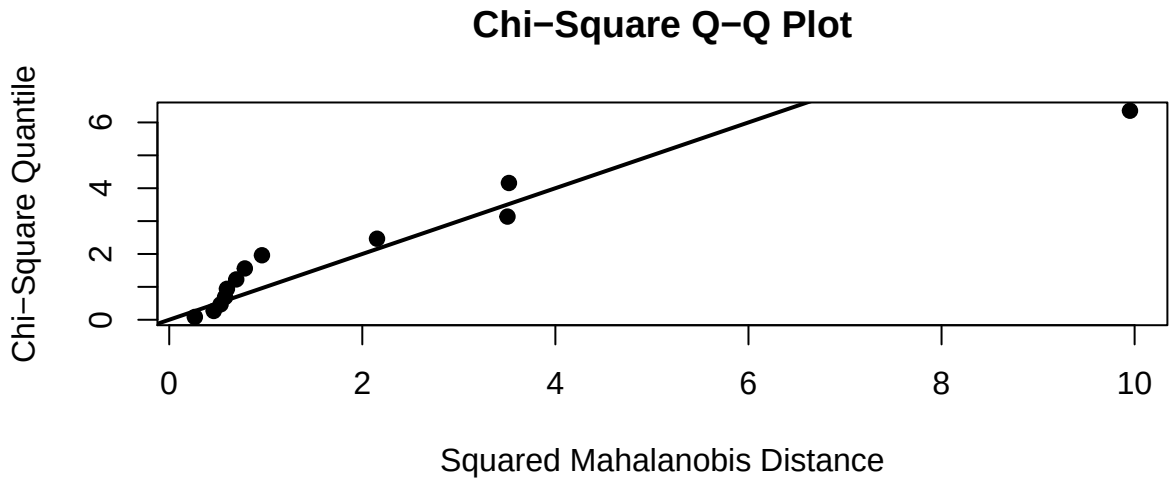
To evaluate the quality of our model we analyze the residuals and our MANOVA assumptions.

Recall that the MANOVA population model is:

$$X_l = \mu + \tau_l + e_l$$

where $e_l \sim N_p(0, \Sigma)$ for $l = 1, ..., g$. To check the distributional assumption of our population model, we can check the 3 groups of residuals in our sample.

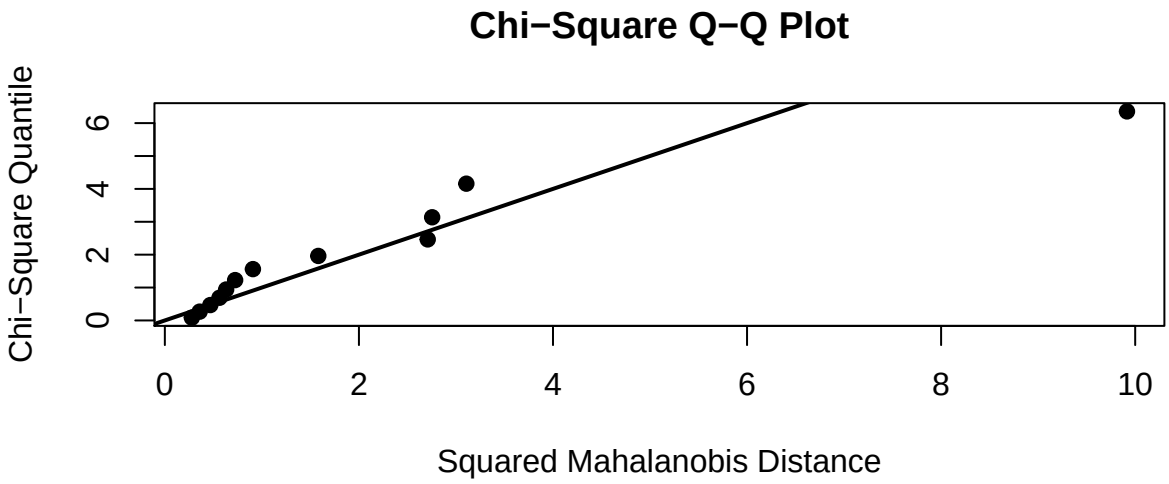
```
ehatJL<-JL-rep(1,12)%*%t(xbarJL)
ehatLP<-LP-rep(1,12)%*%t(xbarLP)
ehatSS<-SS-rep(1,12)%*%t(xbarSS)
ehat<-rbind(ehatJL,ehatLP,ehatSS)

mardiaTest(ehatJL,qqplot=TRUE)
```




```
##    Mardia's Multivariate Normality Test
## ---------------------------------------
```

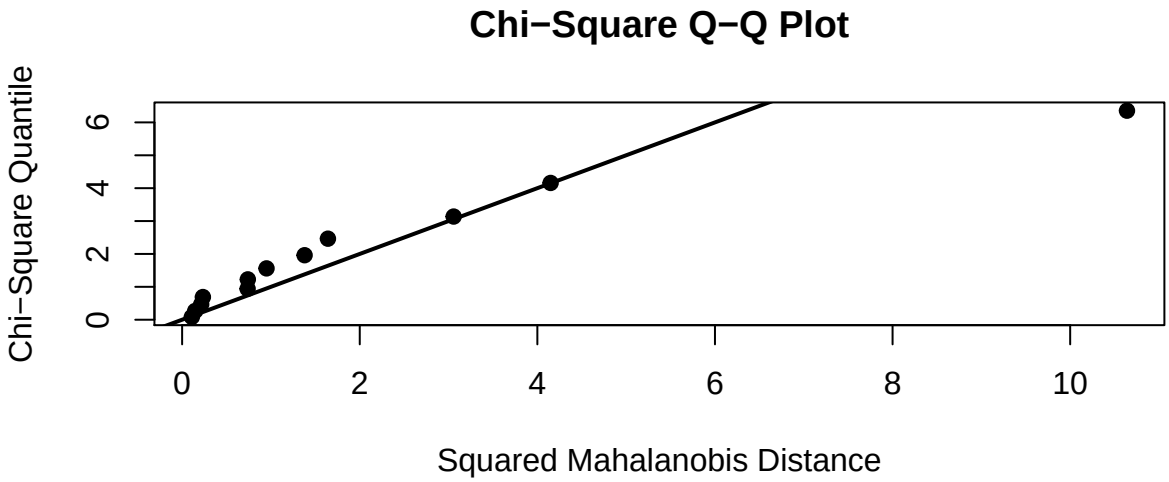
```
##    data : ehatJL
##
##    g1p            : 7.501824
##    chi.skew       : 15.00365
##    p.value.skew   : 0.004693655
##
##    g2p            : 10.96766
##    z.kurtosis     : 1.285037
##    p.value.kurt   : 0.1987795
##
##    chi.small.skew : 22.16448
##    p.value.small  : 0.0001858624
##
##    Result          : Data are not multivariate normal.
## ---------------------------------------
```

```
mardiaTest(ehatLP,qqplot=TRUE)
```




```
##    Mardia's Multivariate Normality Test
## ---------------------------------------
##    data : ehatLP
##
##    g1p            : 7.013966
##    chi.skew       : 14.02793
##    p.value.skew   : 0.007206441
##
##    g2p            : 10.65753
##    z.kurtosis     : 1.150742
##    p.value.kurt   : 0.2498382
##
##    chi.small.skew : 20.72308
##    p.value.small  : 0.0003593167
##
##    Result          : Data are not multivariate normal.
## ---------------------------------------
```

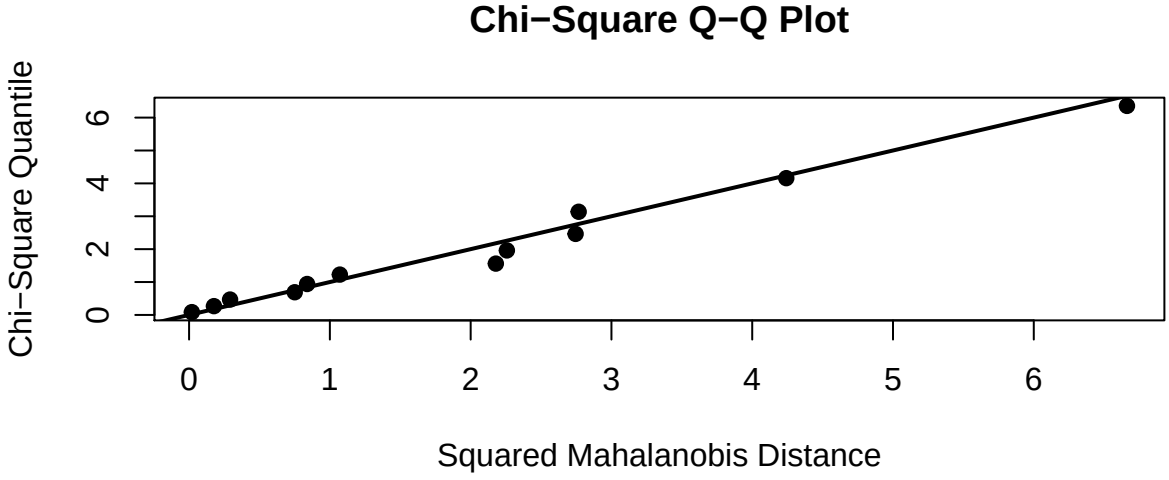
```
mardiaTest(ehatSS,qqplot=TRUE)
```




```
##    Mardia's Multivariate Normality Test
## ---------------------------------------
##    data : ehatSS
##
##    g1p            : 9.462258
##    chi.skew       : 18.92452
##    p.value.skew   : 0.0008132405
##
##    g2p            : 12.20773
##    z.kurtosis     : 1.822
##    p.value.kurt   : 0.06845498
##
##    chi.small.skew : 27.95667
##    p.value.small  : 1.272769e-05
##
##    Result          : Data are not multivariate normal.
## ---------------------------------------
```

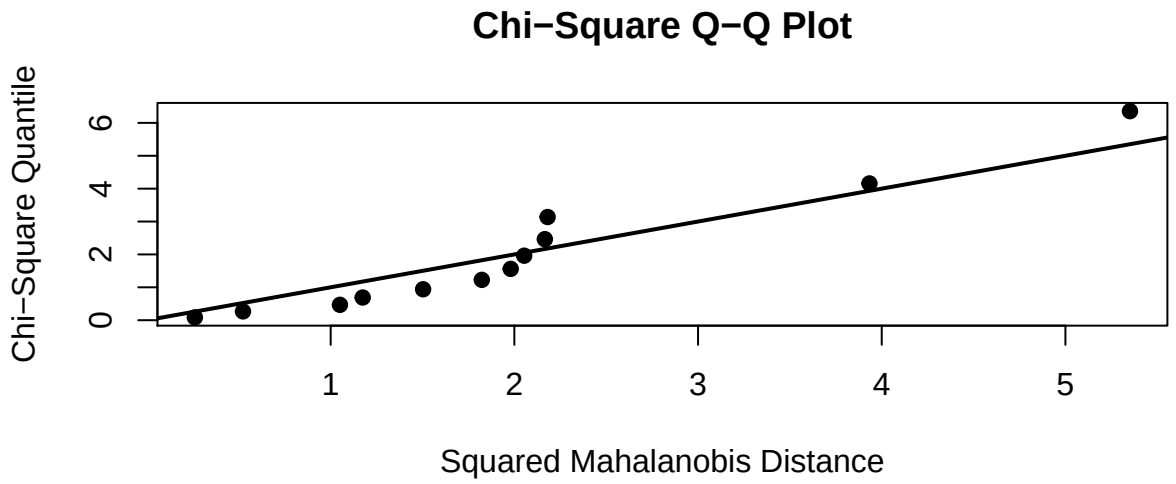
Our groups do not follow a multivariate normal distribution. One method of solution is to apply a transformation uniformly to the three groups.

```
Spec<-rep(c("JL","LP","SS"),each=12)
manovafit_log<-manova(I(1/(spect))~Spec)
mardiaTest(resid(manovafit_log)[1:12,],qqplot=TRUE)
```



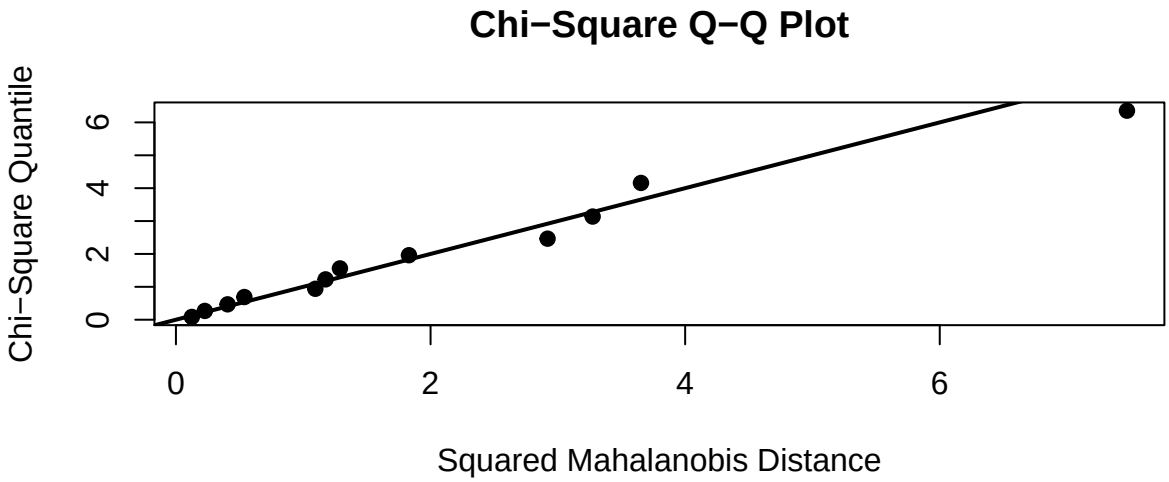
```
##    Mardia's Multivariate Normality Test
## ---------------------------------------
##    data : resid(manovafit_log)[1:12, ]
##
##    g1p            : 1.324434
##    chi.skew       : 2.648868
##    p.value.skew   : 0.6181912
##
##    g2p            : 7.495484
##    z.kurtosis     : -0.2184618
##    p.value.kurt   : 0.8270693
##
##    chi.small.skew : 3.9131
##    p.value.small  : 0.4178942
##
##    Result         : Data are multivariate normal.
## ---------------------------------------
```

```
mardiaTest(resid(manovafit_log)[13:24,],qqplot=TRUE)
```




```
##    Mardia's Multivariate Normality Test
## ---------------------------------------
##    data : resid(manovafit_log)[13:24, ]
##
##    g1p            : 0.6750515
##    chi.skew       : 1.350103
##    p.value.skew   : 0.8528193
##
##    g2p            : 5.841884
##    z.kurtosis     : -0.9344917
##    p.value.kurt   : 0.3500503
##
##    chi.small.skew : 1.99447
##    p.value.small  : 0.736776
##
##    Result         : Data are multivariate normal.
## ---------------------------------------
```

```
mardiaTest(resid(manovafit_log)[25:36,],qqplot=TRUE)
```

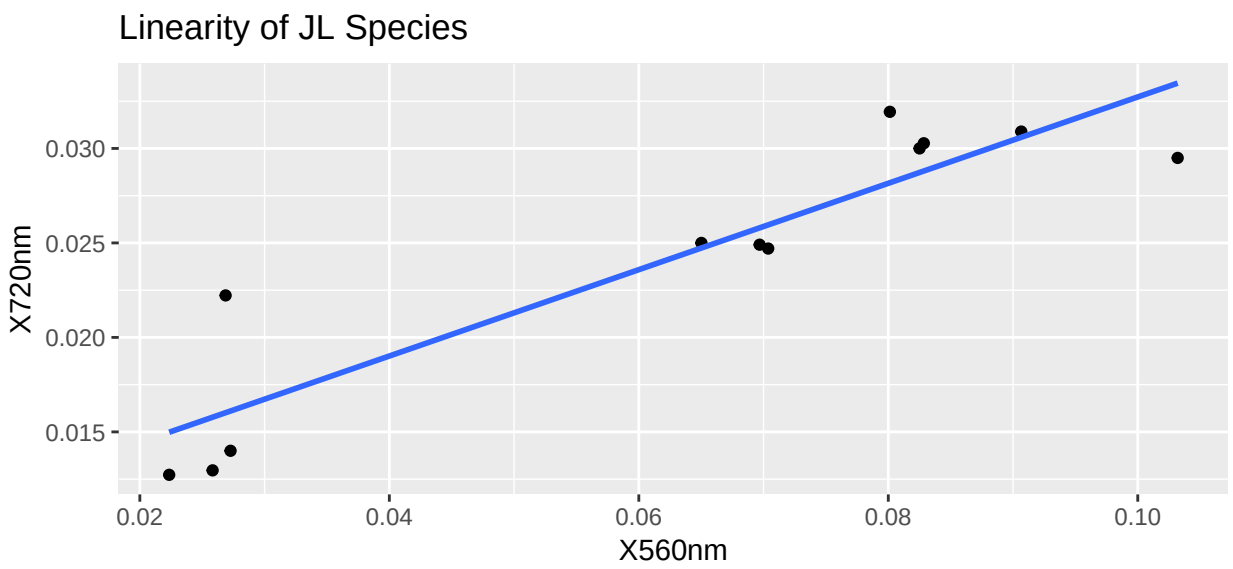



```
##    Mardia's Multivariate Normality Test
## ---------------------------------------
##    data : resid(manovafit_log)[25:36, ]
##
##    g1p            : 3.52256
##    chi.skew       : 7.045121
##    p.value.skew   : 0.1335229
##
##    g2p            : 8.042106
##    z.kurtosis     : 0.01823238
##    p.value.kurt   : 0.9854535
##
##    chi.small.skew : 10.40756
##    p.value.small  : 0.03409436
##
##    Result          : Data are not multivariate normal.
## ---------------------------------------
```

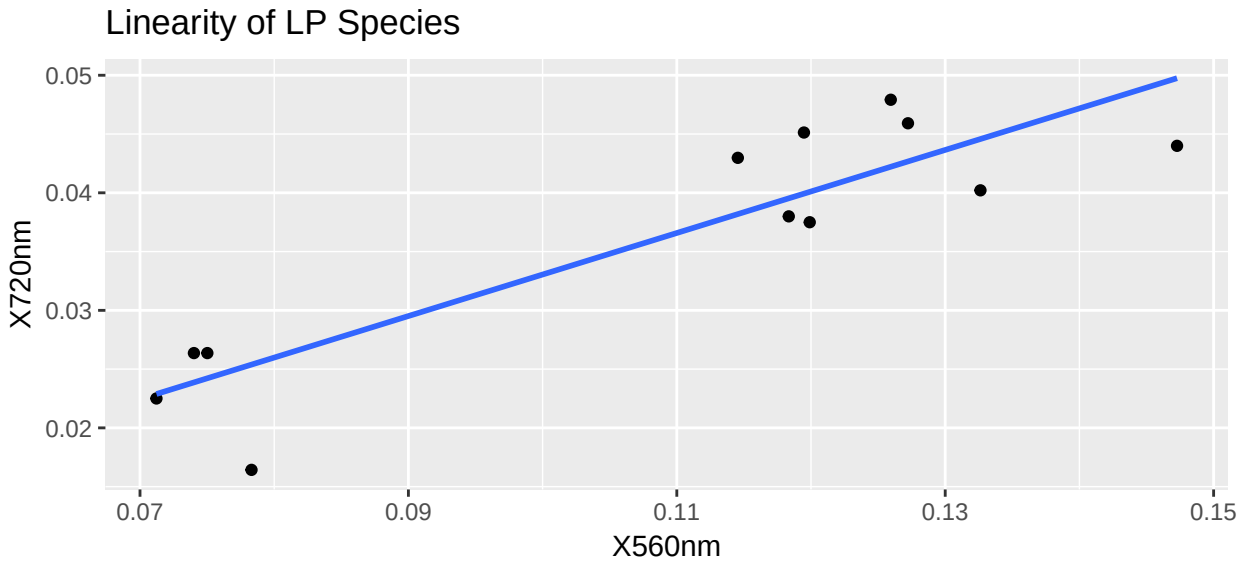
It looks like the inverse transformation works well for our data. Although the data for our SS group aren't quite normal, it's good enough for our application.

Our second assumption to check is linearity of the two spectral outcomes within each of the three groups. We can check this visually by constructing fitting linear regression lines between the DV's and assessing how well the lines fit the scatter.

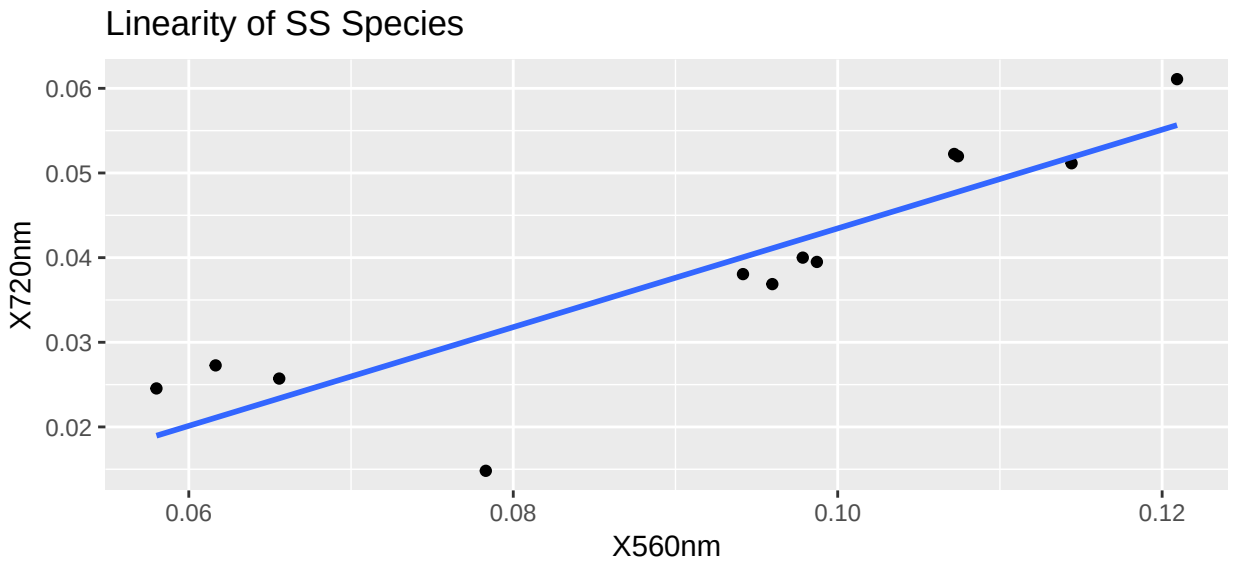
```
ggplot(aes(x = X560nm, y = X720nm), data = as.data.frame(1/JL)) + geom_point() +
    geom_smooth(method = "lm", se = FALSE) + labs(title = "Linearity of JL Species")
```




```
ggplot(aes(x = X560nm, y = X720nm), data = as.data.frame(1/LP)) + geom_point() +
    geom_smooth(method = "lm", se = FALSE) + labs(title = "Linearity of LP Species")
```




```
ggplot(aes(x = X560nm, y = X720nm), data = as.data.frame(1/SS)) + geom_point() +
    geom_smooth(method = "lm", se = FALSE) + labs(title = "Linearity of SS Species")
```



Linearity within groups checks out.

Our final assumption is homogeneity of variance-covariance matrices which we can check with Box's M test:

```
boxM(data=1/spectral[,c(1,2)],grouping=spectral[,3])
```

```
##
##  Box's M-test for Homogeneity of Covariance Matrices
##
## data:  1/spectral[, c(1, 2)]
## Chi-Sq (approx.) = 19.13, df = 6, p-value = 0.003949
```

Box's M is sensitive to non-normality and is known for being overly sensitive which is why it's recommended by some (Tabachnick & Fidell, 2001) to be evaluated at the $\alpha = 0.001$ level. By this criterion, our assumption is satisfied.

Follow-up Comparisons

Now that we know our species are different, we can use 95% simultaneous confidence intervals to determine exactly which mean components differ.

With confidence at least 95%,

$$\tau_{ki} - \tau_{li} \in \left(\bar{x}_{ki} - \bar{x}_{li} \pm t_{36-3}\left(\frac{0.05}{2 \cdot 3 \cdot 2}\right)\sqrt{\frac{w_{ii}}{n-g}\left(\frac{1}{n_k} + \frac{1}{n_l}\right)}\right)$$

$$\left(\bar{x}_{1i} - \bar{x}_{2i} \pm 2.81\sqrt{\frac{w_{ii}}{33}\left(\frac{1}{6}\right)}\right)$$

$$\left(\bar{x}_{1i} - \bar{x}_{3i} \pm 2.81\sqrt{\frac{w_{ii}}{33}\left(\frac{1}{6}\right)}\right)$$

$$\left(\bar{x}_{2i} - \bar{x}_{3i} \pm 2.81\sqrt{\frac{w_{ii}}{33}\left(\frac{1}{6}\right)}\right)$$

```
xbarJL-xbarLP-2.81*sqrt(diag(W)/(33*6))
```

```
##    X560nm    X720nm
##  2.494474 -1.764458
```

```
xbarJL-xbarLP+2.81*sqrt(diag(W)/(33*6))
```

```
##   X560nm   X720nm
## 21.00386 32.90946
```

```
xbarJL-xbarSS-2.81*sqrt(diag(W)/(33*6))
```

```
##     X560nm     X720nm
##  0.7569741 -1.0961248
```

```
xbarJL-xbarSS+2.81*sqrt(diag(W)/(33*6))
```

```
##   X560nm   X720nm
## 19.26636 33.57779
```

```
xbarLP-xbarSS-2.81*sqrt(diag(W)/(33*6))
```

```
##    X560nm    X720nm
## -10.99219 -16.66862
```

```
xbarLP-xbarSS+2.81*sqrt(diag(W)/(33*6))
```

```
##    X560nm    X720nm
##  7.517193 18.005292
```

Comparing the JL and LP species:

$$\tau_{(JL,560nm)} - \tau_{(LP,560nm)} \in (2.49, 21.00)$$

$$\tau_{(JL,720nm)} - \tau_{(LP,720nm)} \in (-1.76, 32.91)$$

Comparing the JL and SS species:

$$\tau_{(JL,560nm)} - \tau_{(SS,560nm)} \in (0.76, 19.27)$$

$$\tau_{(JL,720nm)} - \tau_{(SS,720nm)} \in (-1.10, 33.58)$$

Comparing the LP and SS species:

$$\tau_{(LP,560nm)} - \tau_{(SS,560nm)} \in (-10.99, 7.52)$$

$$\tau_{(LP,720nm)} - \tau_{(SS,720nm)} \in (-16.67, 18.01)$$

Based on our Simultaneous Testing Procedure,

1. The JL and LP species differ on the 560nm spectral wavelength but not on the 720nm one.
2. The JL and SS species also differ on the 560nm but not on 720nm.
3. The LP and SS species do not differ on 560nm or 720nm.

In sum, the species don't differ in 720nm but they don't differ in 560nm. Curiously, in 560nm, JL and LP differ from each other; JL and SS differ from each other; but LP and SS don't differ.

d) Using the first model, compute 95% bootstrap simultaneous confidence intervals. Compare results.

```
nboot<-10000
xbarJLs<-matrix(NA,nboot,2)
xbarLPs<-matrix(NA,nboot,2)
xbarSSs<-matrix(NA,nboot,2)
for(i in 1:nboot){
  JL_boot<-JL[sample(1:12,replace=TRUE),]
  LP_boot<-LP[sample(1:12,replace=TRUE),]
  SS_boot<-SS[sample(1:12,replace=TRUE),]
  xbarJLs[i,]<-apply(JL_boot,2,mean)
  xbarLPs[i,]<-apply(LP_boot,2,mean)
  xbarSSs[i,]<-apply(SS_boot,2,mean)
}

#JL vs. LP for 560
quantile(xbarJLs[,1]-xbarLPs[,1],probs=c(.00417,1-.00417))
```

```
##    0.417%   99.583%
##  3.026632 22.230253
```

```
#JL vs. LP for 720
quantile(xbarJLs[,2]-xbarLPs[,2],probs=c(.00417,1-.00417))
```

```
##     0.417%    99.583%
##  0.2945694 32.3877119
```

```
#JL vs. SS for 560
quantile(xbarJLs[,1]-xbarSSs[,1],probs=c(.00417,1-.00417))
```

```
##    0.417%   99.583%
##  1.216531 20.280962
```

```
#JL vs. SS for 720
quantile(xbarJLs[,2]-xbarSSs[,2],probs=c(.00417,1-.00417))
```

```
##      0.417%     99.583%
##  0.02739575 32.69107999
```

```
#LP vs. SS for 560
quantile(xbarLPs[,1]-xbarSSs[,1],probs=c(.00417,1-.00417))
```

```
##    0.417%   99.583%
## -4.704021  1.295760
```

```
#LP vs. SS for 720
quantile(xbarLPs[,2]-xbarSSs[,2],probs=c(.00417,1-.00417))
```

```
##    0.417%   99.583%
## -13.82880  14.41307
```

Comparing the JL and LP species:

$$\tau_{(JL,560nm)} - \tau_{(LP,560nm)} \in (2.69, 22.12)$$

$$\tau_{(JL,720nm)} - \tau_{(LP,720nm)} \in (0.40, 32.03)$$

Comparing the JL and SS species:

$$\tau_{(JL,560nm)} - \tau_{(SS,560nm)} \in (1.01, 20.58)$$

$$\tau_{(JL,720nm)} - \tau_{(SS,720nm)} \in (-0.66, 33.67)$$

Comparing the LP and SS species:

$$\tau_{(LP,560nm)} - \tau_{(SS,560nm)} \in (-4.67, 1.09)$$

$$\tau_{(LP,720nm)} - \tau_{(SS,720nm)} \in (-13.24, 13.79)$$

The bootstrapped confidence intervals are tighter but the results are mostly the same as the theoretical confidence intervals. The only exception is that now JL and LP differ significantly on 560nm *and* 720nm.