MANOVA

KEY

2/18/2020

## MANOVA of shell measurements by shell type

Import the sheet “shells” from “multivariate\_data\_fixed.xlsx”, into a data frame called “shells”:

library(readxl)  
data.frame(read\_excel("multivariate\_data\_fixed.xlsx", "shells")) -> shells

Make a vector with the names of the response variables in it:

shell.variables <- c("major.axis","minor.axis","surface.len","height","depth","ln.weight")

Conduct the MANOVA:

manova(as.matrix(shells[shell.variables]) ~ shell.type, data = shells) -> shells.manova  
summary(shells.manova)

## Df Pillai approx F num Df den Df Pr(>F)   
## shell.type 1 0.97442 717.38 6 113 < 2.2e-16 \*\*\*  
## Residuals 118   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

**Question: are the shell types different? How do you know?**

Yes they are, the p-value is less than 0.05, and there are only two shell types.

**Question: can you tell from this MANOVA table if the shells are different on every variable, or is it possible that they are only different on some variables but not all of them?**

No, you can only tell that they are on at least one variable.

## MANOVA of the LandSat data by cover type

Import the LandSat data.

data.frame(read\_excel("multivariate\_data\_fixed.xlsx", "landsat")) -> landsat

Transform all of the bands except band4 and band6:

landsat <- transform(landsat, ln.band1 = log(band1), ln.band2 = log(band2), ln.band3 = log(band3), ln.band5 = log(band5), ln.band7 = log(band7))

Make a list of response variables for the landsat dataset:

landsat.variables <- c("ln.band1","ln.band2","ln.band3","band4","ln.band5","band6","ln.band7")

Run the MANOVA and get the MANOVA table:

manova(as.matrix(landsat[landsat.variables])~cover.type, data = landsat) -> landsat.manova  
  
summary(landsat.manova)

## Df Pillai approx F num Df den Df Pr(>F)   
## cover.type 3 0.66146 7.7583 21 576 < 2.2e-16 \*\*\*  
## Residuals 196   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

**Question: are cover types different in their reflectances across these bands? How do you know?**

Yes, the p-value is less than 0.05.

**Question: can you tell which cover types are different from these results?**

You cannot tell, since there are more than 2 of them.

## Paired comparison of bird sexual size dimorphism

Import the birds data:

data.frame(read\_excel("multivariate\_data\_fixed.xlsx", "bird\_ssd")) -> birds

Make lists of the variables for each sex:

male.variables <- c("m.mass","m.wing","m.tarsus","m.bill","m.tail")  
female.variables <- c("f.mass","f.wing","f.tarsus","f.bill","f.tail")

Make a matrix of differences:

birds[male.variables] - birds[female.variables] -> diffs

Make a list of names for the diffs:

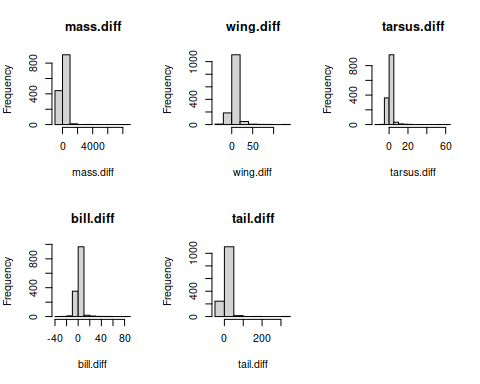
diff.names <- c("mass.diff","wing.diff","tarsus.diff","bill.diff","tail.diff")

And then assign the names of the columns to diffs:

colnames(diffs) <- diff.names

Make histograms of the diffs:

par(mfrow = c(2,3))  
lapply(diff.names, FUN = function(x) hist(diffs[,x], main = x, xlab = x))



**Question: based on the histograms would we meet the (multivariate) normality assumption? How can you tell?**

No, the differences are right skewed.

Calculate the log of the ratios of male to female sizes:

log(birds[male.variables]/birds[female.variables]) -> bird.log.ratios

Make a list of names for the log ratios:

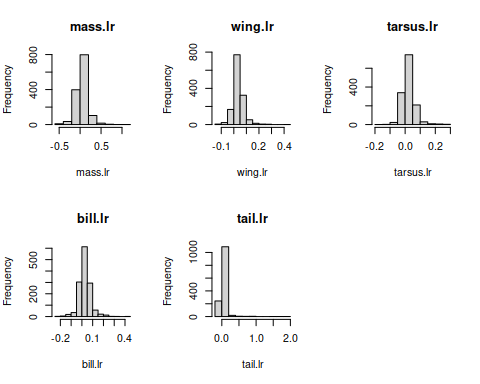
lr.names <- c("mass.lr","wing.lr","tarsus.lr","bill.lr","tail.lr")

Rename the columns in bird.log.ratios:

colnames(bird.log.ratios) <- lr.names

Plot the histograms of bird.log.ratios:

par(mfrow = c(2,3))  
lapply(lr.names, FUN = function(x) hist(bird.log.ratios[,x], main = x, xlab = x))



**Question: are the distributions of log ratios more bell-shaped than the distribution of differences?**

Yes, they are better, although tail.lr is still pretty right skewed.

**Question: why couldn’t we just use the log of the differences to fix the right-skew?**

Many of the differences are 0, and the log of 0 is undefined.

**Question: are males always bigger in birds, according to these histograms? How do you know?**

No, because there are some negative log ratios.

Conduct the paired analysis as an intercept-only model - this will test the means of the log-ratios against 0, which will give us the multivariate paired analysis we are looking for:

manova(as.matrix(bird.log.ratios) ~ 1) -> bird.paired.manova

Get the paired analysis test output:

anova(bird.paired.manova)

## Analysis of Variance Table  
##   
## Df Pillai approx F num Df den Df Pr(>F)   
## (Intercept) 1 0.43711 211.38 5 1361 < 2.2e-16 \*\*\*  
## Residuals 1365   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

**Question: are the sexes the same size on average? How do you know?**

No, because the p-value is less than 0.05.

Back-transform the mean differences:

exp(bird.paired.manova$coefficients)

## mass.lr wing.lr tarsus.lr bill.lr tail.lr  
## (Intercept) 1.050797 1.037295 1.023999 1.030456 1.050778

**Question: which sex is generally bigger? How do you know?**

The males are generally bigger, because the ratios are greater than 1 on average.