Multivariate RMA - profile analysis

Dayna Denver

Wed Dec 7 12:47:58 2022

## Profile analysis - the multivariate approach to repeated measures analysis

### Import the data

Import the data:

library(readxl)  
  
data.frame(read\_excel("leech\_weights.xlsx")) -> weights

Load the car library:

library(car)

## Loading required package: carData

Make a variables list of the days (which are the response variables in our analysis):

days <- c("d0", "d1", "d2", "d3", "d4", "d5", "d6", "d8")

We need to describe the within-subjects design - we will do this by making an R factor out of the list of days:

time.factor <- factor(days)

Then we make a data frame out of the time.factor object:

time.frame <- data.frame(time.factor)

Fit a MANOVA model using the weights as the responses, and food as the predictor:

manova(as.matrix(weights[days]) ~ food, data=weights) -> weights.mlm

Now we have what we need to run the RMA using:

Anova(weights.mlm, idata=time.frame, idesign=~time.factor) -> weights.rma

Get the multivariate tests:

print(weights.rma)

##   
## Type II Repeated Measures MANOVA Tests: Pillai test statistic  
## Df test stat approx F num Df den Df Pr(>F)   
## (Intercept) 1 0.97309 686.94 1 19 2.224e-16 \*\*\*  
## food 1 0.82303 88.36 1 19 1.412e-08 \*\*\*  
## time.factor 1 0.99093 202.96 7 13 3.028e-12 \*\*\*  
## food:time.factor 1 0.95594 40.29 7 13 8.134e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### Question: which effects are signficant? How do you know?

all of the the values are significant ..

#### Question: which effect is the test of flatness?

the within-subjects main effect is to test flatness

#### Question: which effect is the test of parallelism?

the means of differences between the different food types tests for parallelism, between x interactions

## Post-hoc procedures

Now that we know we have an interaction between food and time, we need to find out a) on which days the foods differ, b) within a food type, which successive days differ, and c) on which days do the groups have different amounts of change (i.e. on which days are the changes not parallel)?

### A. On what days do the weights differ between food types?

If we look at one day at a time, we can test on which days the food types differ:

summary(aov(weights.mlm))

## Response d0 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 9.998 9.9978 4.9716 0.03803 \*  
## Residuals 19 38.209 2.0110   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response d1 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 25.093 25.0932 12.208 0.002428 \*\*  
## Residuals 19 39.053 2.0554   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response d2 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 62.916 62.916 28.793 3.535e-05 \*\*\*  
## Residuals 19 41.518 2.185   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response d3 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 117.927 117.927 79.292 3.295e-08 \*\*\*  
## Residuals 19 28.258 1.487   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response d4 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 136.734 136.734 133.69 4.841e-10 \*\*\*  
## Residuals 19 19.433 1.023   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response d5 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 140.511 140.511 143.26 2.706e-10 \*\*\*  
## Residuals 19 18.636 0.981   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response d6 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 157.05 157.052 204.11 1.293e-11 \*\*\*  
## Residuals 19 14.62 0.769   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response d8 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 221.136 221.136 292.58 5.351e-13 \*\*\*  
## Residuals 19 14.361 0.756   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### Question: what alpha level should you use for determining if these are statistically significant differences at each of the eight days?

0.05/8 = 0.0062

#### Question: which of these differences are statistically significant, using the adjusted alpha level you identified?

all but the first day are significant

### B. how is weight changing over time within each food type?

We could also be interested in how the weights change over time within each food type. We need paired t-tests for these comparisons because the same leeches are being measured repeatedly. To make this simpler to calculate for every pair of days, we can calculate the differences by hand, and then do a one-sample t-test of the differences against 0. First, we need to calculate the sequential differences:

weights[days[-8]] - weights[days[-1]] -> diffs  
  
colnames(diffs) <- paste(days[-8], days [-1], sep="-")

We want to know if there are differences between sequential time points within each of the food types, so we should split the data by food type:

split(diffs, weights$food) -> diffs.split

Conduct all 14 of the paired t-tests (that is, the one-sample t-tests of the differences against 0) using lapply():

lapply(diffs.split, FUN=function(x) apply(x, MARGIN=2, FUN = t.test)) -> diffs.split.ts

Get the p-values for the long list of t-tests in diffs.split.ts:

sapply(diffs.split.ts, FUN=function(x) sapply(x, FUN=function(y) y$p.value))

## arg blood  
## d0-d1 6.175098e-05 8.189817e-09  
## d1-d2 6.642295e-06 2.527432e-06  
## d2-d3 7.360390e-05 NaN  
## d3-d4 1.836634e-03 4.501287e-04  
## d4-d5 5.594553e-01 5.910512e-01  
## d5-d6 3.438646e-02 1.000000e+00  
## d6-d8 9.448570e-03 1.113472e-02

#### Question: what alpha level should you use to decide which of these p-values are statistically significant?

0.05/14 = 0.0036

#### Question: which of the sequential differences are statistically significant?

All of them except for (d4 - d5) and (d5 - d6)

#### Question: why did we need to use paired t-tests for these tests, but could use an un-paired post-hoc comparisons to test the effects of food?

we use a paired t-test when we want to compare time for the same set of leeches, we use unpaired when comparing the argenine and blood differences because the leeches would be independent

#### Question: why is the comparison between day 2 and day 3 for blood-fed leeches reported as “NaN”?

there is no difference between day 2 and 3 for weight, standard dev. for differences = 0, standard error =0

We may instead be more interested in when the weights start to differ from the first day (initial conditions). The analysis is exactly the same, but we would calculate differences of each day from 1 to 8 from the first day, 0.

weights[days[-1]] - replicate(7,weights$d0) -> diffs.initial

Split the differences from initial weights by food type:

colnames(diffs.initial) <- paste("d0", days [-1], sep = "-")  
split(diffs.initial, weights$food) -> diffs.initial.split

Run the paired t-tests:

lapply(diffs.initial.split, FUN = function(x) apply(x, MARGIN=2, FUN=t.test)) -> diffs.initial.split.ts

Get the p-values for the paired t-tests:

sapply(diffs.initial.split.ts, FUN=function(x) sapply(x, FUN = function(y) y$p.value))

## arg blood  
## d0-d1 6.175098e-05 8.189817e-09  
## d0-d2 1.286335e-06 6.461429e-09  
## d0-d3 5.017614e-08 6.461429e-09  
## d0-d4 6.323758e-09 1.287624e-08  
## d0-d5 2.920471e-08 1.159449e-08  
## d0-d6 3.278190e-09 9.153420e-09  
## d0-d8 2.711608e-10 2.864736e-08

#### Question: which days differ from initial conditions for arg-fed leeches? What about for blood-fed leeches?

all days

### C - at what days are the sequential differences different between food types?

To see if every pair of sequential differences is different between groups or not, first make a MANOVA of sequential differences compared between food groups:

lm(as.matrix(diffs) ~ food, data=weights) -> parallel.mlm

Compare the differences between food types each day:

summary(aov(parallel.mlm))

## Response d0-d1 :  
## Df Sum Sq Mean Sq F value Pr(>F)  
## food 1 3.4128 3.4128 2.1343 0.1604  
## Residuals 19 30.3817 1.5990   
##   
## Response d1-d2 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 8.5419 8.5419 38.411 5.916e-06 \*\*\*  
## Residuals 19 4.2252 0.2224   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response d2-d3 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 8.5699 8.5699 37.625 6.765e-06 \*\*\*  
## Residuals 19 4.3277 0.2278   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response d3-d4 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 0.69541 0.69541 5.5316 0.02961 \*  
## Residuals 19 2.38861 0.12572   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response d4-d5 :  
## Df Sum Sq Mean Sq F value Pr(>F)  
## food 1 0.02573 0.025733 0.2558 0.6188  
## Residuals 19 1.91118 0.100588   
##   
## Response d5-d6 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 0.46007 0.46007 5.3143 0.0326 \*  
## Residuals 19 1.64485 0.08657   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Response d6-d8 :  
## Df Sum Sq Mean Sq F value Pr(>F)   
## food 1 5.4692 5.4692 13.93 0.001412 \*\*  
## Residuals 19 7.4596 0.3926   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#### Question: what alpha level should you use for these comparisons?

0.05/7 = 0.007

#### Question: are the changes over time different between the groups between every day of measurement, or were the changes the same for both blood and arginine-fed leeches on some days? Be sure to refer to the p-values to support your answer.

Yes they appear to be based on the differences in p values. (d1 - d2) p=5.916e-06 (d2 - d3) p=6.765e-06 (d6 - d8) p=0.001412