

Epi Lab color code

Software/Packages/Add-ins required
 Software/Packages/Add-ins recommended
 Description text
 R code to copy/paste into console
 R code to copy/paste into console that needs adjustment to your personal workspace
 Websites where you can download requirements

Lab #8 requirements

- R - <http://cran.r-project.org/bin/windows/base/>
- R Studio - www.rstudio.com/ide/download/desktop
- Internet connection

Inter-lab proficiency testing

We were asked by our laboratory to assist them in evaluating some data for inter-lab proficiency testing. Unfortunately there were just too few comparative tests done to do much with the information but it allowed us to plan ahead for the next time these proficiency tests are done. A few years back we did a similar project as part of an ELISA approval study and a subset of the data has been taken and used here to illustrate how to start proficiency testing using R. This month's lab is just a start to this, we'll be getting to the more detailed analysis at a later stage, this month is very much based on raw agreement calculations where 3 different labs tested the same sample and returned a positive or negative (dichotomous) result.

We've also found it necessary to summarise some of the new functions included in these labs to help when searching for a previously used function or code example.

Functions and Code covered - Lab 8

allocation of row names; find and replace missing information in dataset; table function; targeting individual values in dataset using [x,y]; array function intro with allocation of column and row names (dimnames); drop residual level in summary

The code

#In this lab we have a sample dataset for 3 labs that tested serum collected from a field trial of an avian species evaluating an ELISA which was being validated for South African conditions. We pre-removed any sample which was just tested at one lab. The purpose is to evaluate the agreement between the labs (raters) given that the ELISA result in this case is either POSITIVE or NEGATIVE. For this session we'll just be looking at RAW agreement only.

```
rm(list = ls()) #removes (almost) everything from the working environment so that you start with a clean sheet. Read in the data:
elisadata<-read.csv("http://www.jdata.co.za/backpagelabs/backpagelabs_jdg_agreementraw.csv")
#Overview of the data
head(elisadata, 50)
summary(elisadata)
#to clean our table a bit lets make the row names for each line equal the sample ID
rownames(elisadata)<- elisadata[,1]; head(elisadata)
#now we remove the first SampleID column
elisadata<- elisadata[,-1];head(elisadata)
summary(elisadata)
elisadata
#if you scroll through the data you'll notice that not every sample was tested in all 3 labs. This can also be noticed through the summary function. Lets replace the missing data with "NA"
elisadata[elisadata==""]<-NA; summary(elisadata)
#we need to drop the residual "" level
elisadata<-droplevels(elisadata); summary(elisadata)
#OK, we are ready to start with a relatively clean dataset
#While we are not going to go into the details of rater (in this case lab) agreement (see http://john-uebersax.com/stat/agree.htm for an excellent overview) the following were decisions we took for the data we received.
#Our data here is in nominal dichotomous scale with a result either being positive or negative or NA (not available - lets assume here there are no suspect cases). This means for agreement stats we cannot use tests that rely on ordered, interval or ratio data (for example inter class correlation or ANOVA), so the basic options for dichotomous inter-rater reliability are: 1. Raw agreement; 2. Cohen's kappa (for two raters); 3. Fleiss kappa (if there
```

are more than 2 raters)

#We'll deal with latter two in a future lab, lets just look at RAW AGREEMENT which is essentially proportions. We'll look at each lab compared to each other (i.e. Lab1 vs Lab2; Lab1 vs Lab3 and Lab2 vs Lab3) and split the raw analysis into overall agreement and categorised agreement.

###1. PROPORTION OF OVERALL AGREEMENT###

#Here agreement is where both labs detected negative results and where both detected positive results on the same sample. Raw agreement can be determined by the sum of agreements divided by the total number of samples evaluated

###1.1 Lab1 versus Lab2##

```
#First make a 2X2 table of the data
lab1versuslab2_raw<-table
(elisadata$lab1,elisadata$lab2,dnn=c(colnames(elisadata
[1:2]))); lab1versuslab2_raw
#note the table function - to include the column names so that you know which way the table is orientated you add the dnn part of the table function indicating that you want to label the table with the first two column names of the dataset. NOTE: this is just a label so you need to get the order right, its not automated.
#OK to get the raw agreement which is the agreed results divided by the total evaluated (by both labs (excluding NULL)
l1l2raw<-(lab1versuslab2_raw[1,1]+lab1versuslab2_raw
[2,2])/sum(lab1versuslab2_raw); l1l2raw
#so the agreement here was 91.9%
```

###1.2 Lab1 versus Lab3##

```
lab1versuslab3_raw<-table
(elisadata$lab1,elisadata$lab3,dnn=c(colnames(elisadata
[1]),colnames(elisadata[3]))); lab1versuslab3_raw
l1l3raw<-(lab1versuslab3_raw[1,1]+lab1versuslab3_raw
[2,2])/sum(lab1versuslab3_raw);l1l3raw
#so the agreement here was 98.07%
```

###1.3 Lab2 versus Lab3##

```
lab2versuslab3_raw<-table
(elisadata$lab2,elisadata$lab3,dnn=c(colnames(elisadata
[2:3])));lab2versuslab3_raw
```

```
l2l3raw<- (lab2versuslab3_raw[1,1]+lab2versuslab3_raw
[2,2])/sum(lab2versuslab3_raw);l2l3raw
#so the agreement here was 92.61%
```

#There must be an easier way to do this but it would be nice to have the results in a matrix to look at quickly...so here is the known long way. Make an array of the results which will be 3 rows by 3 columns

```
arrayRAW<-array(c
(NA,l1l2raw,l1l3raw,l1l2raw,NA,l2l3raw,l1l3raw,l2l3raw,NA),c(3,3));arrayRAW
#assign row and column names (which in this case will be the same)
dimnames(arrayRAW)<-list(c(colnames(elisadata[1:3])),c
(colnames(elisadata[1:3]))); arrayRAW
```

###2. POSITIVE AND NEGATIVE AGREEMENT

#A subcategory of RAW agreement is to sub categorise the agreement into categorised (in our case positive and negative results). The rationale for this may be that you may have a situation much like ours here, where the number of negative agreement is skewing your overall agreement making it look better than it really is.

#Positive agreement can be defined as the number of times two raters (labs) agreed on positive findings compared to the total number of positive findings detected (either by 1st rater or 2nd rater or both)

##2.1.1 Lab 1 versus Lab 2 positive agreement##

```
lab1versuslab2_pos<-table(elisadata$lab1,elisadata$lab2,
dnn=c(colnames(elisadata[1:2])));lab1versuslab2_pos
```

```
l1l2pos<-lab1versuslab2_pos[2,2]/(lab1versuslab2_pos
[2,2]+lab1versuslab2_pos[2,1]+lab1versuslab2_pos
[1,2]);l1l2pos
```

#0.55 - 55% positive agreement

##2.1.2 Lab 1 versus Lab 3 positive agreement##

```
lab1versuslab3_pos<-table(elisadata$lab1,elisadata$lab3,
dnn=c(colnames(elisadata[1]),colnames(elisadata
[3])));lab1versuslab3_pos
```

```
l1l3pos<-lab1versuslab3_pos[2,2]/(lab1versuslab3_pos
[2,2]+lab1versuslab3_pos[2,1]+lab1versuslab3_pos
[1,2]);l1l3pos
```

#0.7 - 70% positive agreement

##2.1.3 Lab 2 versus Lab 3 positive agreement##

```
lab2versuslab3_pos<-table
(elisadata$lab2,elisadata$lab3,dnn=c(colnames(elisadata
[2:3])));lab2versuslab3_pos
```

```
l2l3pos<-lab2versuslab3_pos[2,2]/(lab2versuslab3_pos
[2,2]+lab2versuslab3_pos[2,1]+lab2versuslab3_pos
[1,2]);l2l3pos
```

#0.527 - 52.7% positive agreement

#To make an array of positive agreement results

```
arraypos<-array(c
(NA,l1l2pos,l1l3pos,l1l2pos,NA,l2l3pos,l1l3pos,l2l3pos,NA),c(3,3)); arraypos
#assigns row and column names (which in this case will be the same)
dimnames(arraypos)<-list(c(colnames(elisadata[1:3])),c
(colnames(elisadata[1:3]))); arraypos
```

##2.2.1 Lab 1 versus Lab 2 negative agreement##

```
lab1versuslab2_neg<-table(elisadata$lab1,elisadata$lab2,
dnn=c(colnames(elisadata[1:2])));lab1versuslab2_neg
```

```
l1l2neg<-lab1versuslab2_neg[1,1]/(lab1versuslab2_neg
[1,1]+lab1versuslab2_neg[1,2]+lab1versuslab2_neg
[2,1]);l1l2neg
```

#0.9102 - 91% negative agreement

##2.2.2 Lab 1 versus Lab 3 negative agreement##

```
lab1versuslab3_neg<-table(elisadata$lab1,elisadata$lab3,
dnn=c(colnames(elisadata[1]),colnames(elisadata
[3])));lab1versuslab3_neg
```

```
l1l3neg<-lab1versuslab3_neg[1,1]/(lab1versuslab3_neg
[1,1]+lab1versuslab3_neg[1,2]+lab1versuslab3_neg[2,1]);
l1l3neg
```

#0.0.979 - 97.9% negative agreement

##2.2.3 Lab 2 versus Lab3 negative agreement##

```
lab2versuslab3_neg<-table(elisadata$lab2,elisadata$lab3,
dnn=c(colnames(elisadata[2:3])));lab2versuslab3_neg
```

```
l2l3neg<-lab2versuslab3_neg[1,1]/(lab2versuslab3_neg
[1,1]+lab2versuslab3_neg[1,2]+lab2versuslab3_neg
[2,1]);l2l3neg
```

#0.9195 - 91.9% negative agreement

#To make an array of negative agreement results

```
arrayneg<-array(c
(NA,l1l2neg,l1l3neg,l1l2neg,NA,l2l3neg,l1l3neg,l2l3neg,NA),c(3,3));arrayneg
#assigns row and column names (which in this case will be the same)
dimnames(arrayneg)<-list(c(colnames(elisadata[1:3])),c
(colnames(elisadata[1:3]))); arrayneg
```

#So in summary - the RAW agreement stats which is the start of the process with this inter lab proficiency data are:

```
arrayRAW
arraypos
arrayneg
```

#In a future lab we'll be looking at interpreting this information but as a start we can start to think that lab 1 and 3 have got slightly better raw agreements compared to 1 versus 2 and 3 versus 2 and this was true for the overall raw agreement and the positive and negative raw agreement. There also seems to be an issue with positive result agreement although maybe because its an ELISA which is being evaluated (which is generally highly sensitive) this is not a problem... maybe let us know if you have an opinion on this.

The result

	lab1	lab2	lab3
lab1	NA	0.9191617	0.9807939
lab2	0.9191617	NA	0.9261364
lab3	0.9807939	0.9261364	NA

Overall raw agreement

	lab1	lab2	lab3
lab1	NA	0.5500000	0.7000000
lab2	0.55	NA	0.5272727
lab3	0.70	0.5272727	NA

Positive raw agreement

	lab1	lab2	lab3
lab1	NA	0.9102990	0.9798928
lab2	0.9102990	NA	0.9195046
lab3	0.9798928	0.9195046	NA

Negative raw agreement

References

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>