Dataset (s) used

Part I: "ketosis.csv" is a hypothetical cohort study dataset from Dohoo,I., Martin, W. & Stryhn,H., 2010. Veterinary Epidemiologic Research, Second Edition. Our research question is: is there an effect of the condition of a cow at calving on the risk of clinical ketosis?

Part II: The Oswego dataset from the CDC (http://www.cdc.gov/eis/casestudies/xoswego.401-303.student.pdf) contains information on the records of 75 persons under investigation for the cause of acute food poisoning after a dinner party. This is a case-control study & we want to know whether the consumption of chocolate at the dinner party was linked to the food poisoning symptoms experienced by some guests.

Your tasks###

PART I

Set the working directory, open the "ketosis" dataset

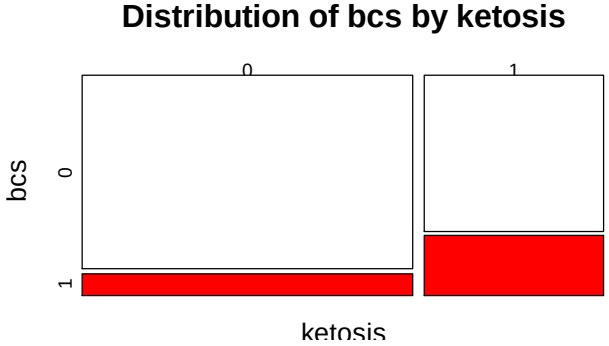
Load the "epiDisplay" package

1. *Working with 2*2 tables*

From within the "base" package in R, tables can be created using the *table()* function. By default, the output will not come with row total or column total, but you can add these by using *addmargins().* Create a 2*2 table for "ketosis" and "bcs".

```
m<-table(cow_ket$ketosis, cow_ket$bcs)
m #table without row sum or column sum
addmargins(m)
```

As an alternative, the function *tabpct()* from the "epiDisplay" package also returns a frequency table. In addition, it also gives column and row percent cross-tabulation as well as mosaic plot. Try it!



The width of the bar in the plot denotes the relative proportion of the row variable. Inside each bar, the relative proportion denotes the distribution of column variables within each row variable.

2. *Estimating probabilities and CI*

Estimate the proportion of cows in your population having ketosis, and the 95%CI for that estimate using the function *ci.binomial().* How would you estimate a 99% CI?

3. *Chi-square test*

Start by writing down the H0 for the Chi-square test before running it using *chisq.test().* What do you conclude?

HELP

If you wanted to see what the expected proportions were under H0, you can retrieve them:

```
mod<-chisq.test(m)
mod
mod$expected #expected proportions if the H0 were true
```

4. *Risk ratio*

The package "epiDisplay" offers lots of interesting functions to calculate and display RR, OR based on the type of study your data come from (case-control, cohort etc…). For more information, type *?cs*

Since our data come from a cohort study, we will estimate a RR and 95%CI using

```
cs(cow_ket$ketosis, cow_ket$bcs)
```

PART II

1. *Creating a 2*2 table*

Imagine that you did not have a .csv file to import into R but only a summary of the data, e.g.:

```
#           chocolate
#    ill      FALSE TRUE
#     FALSE      7   22
#     TRUE      20   25
```

How do you enter these data into R to make a 2*2 table you can do a chi-square test on? You can use the *make2x2()* from the "epiDisplay" package as follows:

```
table1 <- make2x2(25,22,20,7) #from epiDisplay
table1
```

2. *Odds ratio*

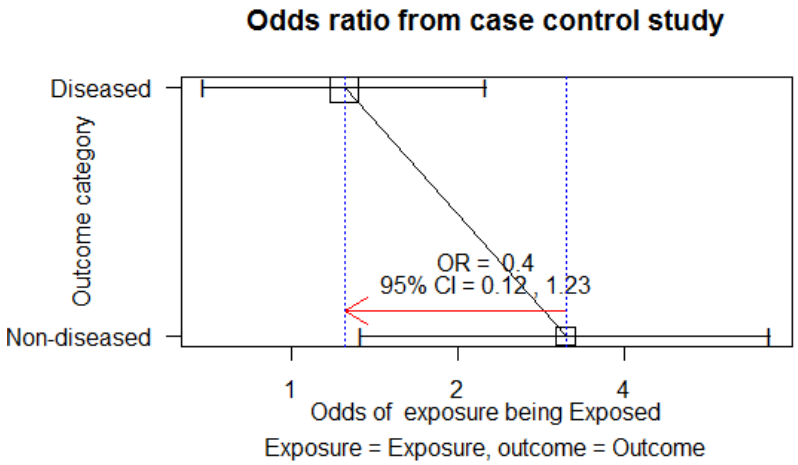
We can use the package "epiDisplay" to calculate the OR from this case-control study:

```
cc(outcome=NULL, exposure=NULL, cctable=table1 ,graph=T, design="case-control")
```

HELP

This function even draws a graph comparing the odds of exposure between cases and controls for you. However, to get it to display properly, you may need to increase the graph margins as follows:

```
par(mar=c(5, 6, 4, 2) + 0.1)#to increase left margin of graph
```



Other useful resources###

Analysis of Epidemiological Data Using R and EpiDisplay: https://cran.r-project.org/web/packages/epiDisplay/epiDisplay.pdf

R graphics basics: http://research.stowers-institute.org/mcm/efg/R/Graphics/Basics/mar-oma/index.htm

Aviva Petrie & Paul Watson 2013: Statistics for Veterinary and Animal Science, 3rd Edition, Wiley-Blackwell. -> Chapter 9: the Chi-squared test.