Exercices 2

Computing descriptive statistics from a detection experiment

In a signal detection experiment, a faint stimulus (e.g. a faint sound or a faint visual target) is presented or not at each trial and the participant must indicate whether he has perceived it or not. There are four possible outcomes for each trial:

1. A *hit* is when the participant correctly detects the target.
2. A *miss* is when the target was there but the participant did not detect it.
3. A *false alarm* is when the participant reports the presence of the target when it was not actually there.
4. A *correct rejection* is when the participant correctly reports that the target was not present.

   One defines;

- The *hit rate* , equal to #hits / (#hits + #misses)
- The *False alarm rate*, equal to #false alarms / (#false alarms + # correct rejections)

1. Let us first suppose that the data from a participant is represented as a string. This string represents a series of trials, each trial being represented by two characters indicating the trial type (1=target present, 0=target absent) and the participant's response (Y=target perceived, N=No target perceived). For example:

   _____
   data = "0Y,0N,1Y,1Y,0N,0N,0Y,1Y,1Y"
   _____

   Write a function that, given such a string, returns the Hit rate and the False rate (hint: use the function `split()`)

2. Suppose the results from different participants are stored in different files `subj*.dat` (download)

   Write a script that computes the hit rates and false alarms for each subject, andisplays the group averages and standard deviations.

3. use `matplotlib.pyplot.plot` to display each participant as a dot on a graphics with False alarm rate on the X-axis and Hit Rate on the Y-axis.

Read the section on reading comma separated value ('.csv") files from `http://automatetheboringstuff.com/chapter14/`

google ngrams

- Read Michel, Jean-Baptiste, Yuan Kui Shen, Aviva P. Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, et al. 2010. "Quantitative Analysis of Culture Using Millions of Digitized Books." Science, December. `https://doi.org/10.1126/science.1199644`. (use scholar.google.com to find a pdf copy)

- Go to the ngram viewer `https://books.google.com/ngrams`.

- At the bottom of the page, there is a message "Raw data is available for download here"". Follow the"here" link. download the 1-grams file 'z' for the dataset "English Version 20120701"". Uncompress it (it is in the .gzip format). It is a text file listing words starting with a 'z'.

  Zipfian