UNIVERSIDAD DE LAS AMÉRICAS PUEBLA
O22 LDS1061 Selected Topics 1
Dr. Miguel Reyes
2022.09.25 1st Task

1. So far we have reviewed two multivariate techniques to help describe multivariate data: Principal Components Analysis (PCA) and Correspondence Analysis (CA). The first deals with vectors whose components are numerical random variables, while the second helps to visualize relationships or dependencies between categorical variables (also known as factors with different levels). In this first activity you are going to look for a case study where these techniques can be applied; that is, you are going to look for data sets where 1) PCA can be applied; 2) CA can be applied.
   You are going to make a report where you explain the following:
   — 1) What the data set is about.
   — 2) Given the context of the data, explain why it is relevant to apply PCA or CA in each case. For this, **it would be reasonable that there were starting questions** (e.g., think of the example of the Bumpus' sparrows, in which the underlying question was whether there were morphological differences between the surviving sparrows and those that died).
   — 3) Discuss each step of the procedure (PCA or CA) in detail (e.g., in PCA, mention what the eigenvalues mean, the scree-plot, how to decide a reasonable number of principal components, how reliable is the representation of the multivariate data in the principal components space, etc).
   — 4) Answer the starting questions and **mention any new discovery or insight resulting from the visualization of the data**.

Specifics:

- A) This activity can be done in groups of no more than 2 people.
- B) Write a report describing the solution to each exercise. If using Word, consider the following characteristics: **Format:** US Letter; **Font:** Calibri, 11; **Margins:** 1 in. up, down, left and right; **Line spacing:** 1.5. If using other editors, try a similar format.
  - B1) Include a cover with general information: e.g., team members, subject, task, etc. In your report, **follow a structure where it is very clear that you are following steps 1 to 4 above.**
  - B2) **Include in your report all the necessary to support your arguments**, so that your report is fully understandable (graphs, tables, etc.). Besides, include your R Script in a separate file.
  - B3) Do not skimp on writing, but write with meaning. Avoid wordiness or a gimmicky writing style. Be formal, but clear and direct.
  - B4) Be careful with your writing. At this level you are expected to be able to clearly express your ideas in writing, following the spelling and syntax rules of the language.
- C) Upload your report to Blackboard (each member of the team).

– C1) Deadline: upload your document no later than Monday 03 OCT, 13:59. Considering the same deadline, a printed version of your report should be in my mailbox (located at CN205) or you can drop it off at my office as well (at CN208J).

– C2) Reports submitted after the deadline will be penalized with 10% per day of delay (e.g., submitting from Monday 03, 14:00 to Tuesday 04, 13:59, your report's maximum possible grade will be 9; submitting on Tuesday 04, 14:00 to Wednesday 05, 13:59, your report's maximum possible grade will be 8, and so forth).