**APPLIED STATISTICS EXAM**

**DATE:** 12/07/2022

**STUDENT:** FILIPPO CIPRIANI

**PERSONA CODE:** 10956877

**EXERCISE NUMBER N**

We take a look at the data and see that maybe there are [N] clusters of data.

We compute the distance matrix for the data using [DISTANCE] distance, then using a [LINKAGE] linkage we build the dendrogram here:

[DENDROGRAM PLOT]

The clustering is not bad, we have a cophenetic coefficient of [COPH. COEF] and visualizing the datas we can see that it performs well in separing the clusters we have seen before:

[PLOT OF COLOURED CLUSTERS]

Clusters have dimensions [DIM. OF CLUSTERS]. The centroids are [(Length ; Width)] :

[CENTROIDS]

We check that the datas belonging to a certain cluster are normal bivariate by performing a shapiro test for each cluster of datas, we have pvalues:

[P-VALUES]

P- values high enough to confirm the hypothesis of normality. We also assume the same covariance structures between clusters.

We perform a Manova using the cluster labels: the Wilks statistic suggests us that there’s a sensible difference between the groups, the pvalue is [P-VALUE]. So the membership to a cluster rather than another makes the difference.

We have to build [N g\*p ad esempio 3 gruppi \* 2 variabili=6] Bonferroni interval for the differences in mean of global confidence [1-ALPHA]%, to do so we extract the W matrix from the Manova values and compute the sample means of each cluster. We computer the quantile of the t of confidence 1-0.1/2\*6 and n – g degrees of freedom.

[BONFERRONI INTERVALS]

We can see that the first two clusters (red and black) are not distinguishable by [VARIABLE 1] (we cannot assume the difference of their means to be different from 0) as we previously see in the picture. ….For all the other differences of values instead we can see a clear difference from 0, only the difference mean1 – mean3 is a little closer to zero in the Width but still we can say that is different from 0, since its interval does not even contain the zero…..

These intervals confirms the output from the Manova and also our previous visive analysis. Giving a glance at the coloured plot of the clusters we can see how this intervals are coherent with the distances among clusters that we can see there.

**POINT A)**

**POINT B)**

**POINT C)**

**POINT D)**