2. Do this problem with R. Use the Pollution Data we used in Introduction.

2.1 First do log-transformation on the three variables SO2, Population and Manufacture and also change their variable names to reflex this change.

2.2 Run PCA and extract all PCs. From the output of PCA, find the eigenvalues. What is the proportion of total variances explained by the first PC? What is the proportion of total variances explained by the first two PCs together?

* 1. For each of them, construct a QQ plot. Do you see any deviations from normality? Note that you should focus on systematic deviations rather than minor deviations that could be due to sampling variability.

2.4 Run Shapiro-Wilk test on the PCs. What are the conclusions with Holm’s correction?

2.5 Run parallel analysis with 1, 000 replications and show error bar (using 1 SE). How many PCs does parallel analysis suggest?