This project corresponds to the Statistics subject of the Specialization Career in Data Science of the Technological Institute of Buenos Aires

This Project corresponds to Statistics subject

Fusible inner linings with increasing strength are being explored to support the outer fabric and improve the shape and drape of various pieces of clothing. The article “Compability of outer an fusible interliningfabrics in tailored garments(Textile Res J.,1997:137-142) gave the attached data of extensibility(%) at 100 g/cm for both high-density(A) and specimens of low-density fabrics(B).

A. 1.2 0.9 0.7 1.0 1.7 1.7 1.1 0.9 1.7 1.9 1.3 2.1 1.6 1.8 1.4 1.3 1.9 1.6 0.8 2.0 1.7 1.6 2.3 2.0

B. 1.6 1.5 1.1 2.1 1.5 1.3 1.0 2.6

The following exercise is from Jay Devore's book, Probability and Statistic for Engineering and the Sciences, 8th Edition, Cengage Learning.

You are asked to do the following.

1. Begin by indicating the variables to be analyzed and for which one or more samples of observations are provided.

2. Descriptive measures of each data set.

3. The following plots for each data set: histogram, stem leaf (back to back, if applicable), boxplots if applicable.

4. Draw the appropriate conclusions from points 1 and 2.

5. If the data sets contain at least 15 data points, perform a suitable QQ-plot. To make conclusions.

6. Apply the Shapiro-Wilk test to each data set (it can be applied even in small samples) and indicate the hypotheses and conclusions that the results lead to together with point 4.

7. Indicate those measurements, atypical points, symmetry, asymmetry, etc. that you can observe after performing the previous points. Make a global conclusion with everything analyzed.

8. If the data is not normal, try a transformation of the same with logarithm and/or square root and repeat points 5, 6 and 7 above with the new set of transformed data. (If the data is normal it is not necessary!).

9. If you have more than one sample, say whether they are independent or paired samples. Compare both samples and indicate the most notable differences or similarities between them.

1. Variable: extensibility (%) at 100g/cm

Groups A: high quality fabrics (24 samples)

Group B: low quality fabrics (8 samples)

2. Descriptive measures summary measures mean(group\_a) 1.508333median(group\_a) 1.6 mean(group\_b) 1.5875 median(group\_b) 1.5 mode(group\_a) 1.7 mode(group\_b) 1.5 Measures of dispersion Sample Range(group\_a) 1.6 Sample range(group\_b) 1.6 Interquartile distance(group\_a) 0.65 Interquartile distance(group\_b) 0.475 Standard deviation(group\_a) [0.4442059 Standard deviation(group\_b) 0.5303301 Variance(group\_a) 0.1973188 Variance(group\_b) 0.28125

From the summary and deviation measures, it can be seen that the median and average are not the same in both groups. In the boxplot, it can be seen that the median is not exactly in the middle of the box as it should be if there is symmetry.

6. In both cases, p-value is greater than alpha (significance level) therefore there is no evidence enough to reject the null hypothesis that the distribution of the variable is normal.

7. From the box-plot, histogram and stem-leaf graphs it was possible to determine the asymmetry, since that the median in the boxplots is not exactly in the middle of the boxes. Also, the presence of outliers could not be observed in the graphs. Finally I know concludes that the variables group\_a and group\_b have a normal distribution.

8. Since both groups have a normal distribution, no transformations are made. 9. These are independent samples. The comparison of both samples can be understand from the measures of summary and dispersion.

The null hypothesis that both populations have the same variance couldn´t be rejected.