

## Exercises: Statistical Inference for Two Populations

### A: Difference between Means

- 1) We are interested in determining whether the position of Coca-Cola in stores affects sales. Specifically does Coca-Cola sell better when placed in the front or the middle of the aisle. Sales from a random sample of 10 stores when Coca-Cola was placed in the front of the aisle, and from another random sample of 10 stores when Coca-Cola was placed in the middle of the aisle were recorded. It is assumed that the sales of Coca-Cola in stores that display it at the front of the aisle and in stores that display it in the middle of the aisle are normally distributed. The data is in file [stores.csv](#).
  - a) Test the hypothesis that the position of Coca-Cola in stores affects sales. Use the 5% level of significance.
  - b) Obtain a 95% confidence interval for the difference in the mean sales of Coca-Cola in stores that display it at the front of the aisle, and in stores and that display it in the middle of the aisle.
  - c) Perform diagnostic checks for outliers and to assess the validity of the normality assumptions.
  
- 2) A new flight booking system which has been implemented at ABC Airways is designed to increase the number of passengers an operator can service in a given period of time. The number of passengers that a random sample of fifteen operators have serviced, in a one-hour period before and after the implementation of the new system, is given in the file [flights.csv](#).

Let  $X_1$  and  $X_2$  be the number of passengers serviced before and after the implementation of the new flight booking system, respectively, and let  $D = X_1 - X_2$ .

- a) Can we conclude at the 5% level of significance that the new flight booking system has resulted in an increase in the mean number of passengers being serviced? Explain your answer.
- b) Explain why a paired-samples analysis is relevant here.
- c) Obtain the 95% confidence interval for the difference between the mean number of passengers serviced before and after the implementation of the new flight booking system. Report and interpret the confidence interval.
- d) Perform diagnostic checks for outliers and to assess the assumption of normality of the underlying population distribution of the differences between the number of passengers, serviced before and after the implementation of the new flight booking system.

## B: Equality of Variances

- 1) An experiment is conducted to compare the consistency of two brands (A and B) of mercury detectors in measuring mercury concentration in the air. At midday in a downtown city area, 10 measurements of mercury concentration were made with the Brand A detector and 9 measurements of mercury concentration were made with the Brand B detector. It is assumed that the mercury measurement concentrations from each detector is normally distributed. The data is in the file [mercury.csv](#).
  - a) Does the data provide sufficient evidence that the Brand B detector measures mercury concentration in the air more consistently than the Brand A detector. Use the 1% level of significance.
  - b) Obtain and report a 99% confidence interval estimate of the ratio of variances of measurements made with the Brand A and Brand B detectors.
  - c) Perform diagnostic checks for outliers and to assess the validity of the normality assumptions.
  
- 2) For Exercise 1 in **A: Difference between Means**, using the data file [stores.csv](#):
  - a) Test for equality of variances of sales of Coca-Cola for stores that display it at the front of the aisle and for stores that display it in the middle of the aisle. Use the 5% level of significance.
  - b) Based on the outcome for Part (a), test the hypothesis that the position of Coca-Cola in stores affects sales. Use the 5% level of significance.
  - c) If outliers were previously detected, remove the outliers, perform the diagnostic check again; then perform the tests in Parts (a) and (b) again and comment on any differences.

### **C: Difference between Proportions**

- 1) Is there a difference in the proportion of construction workers who are under 35 years of age and the proportion of telephone repair workers who are under 35 years of age? Suppose a study is conducted using random samples of 338 construction workers and 218 telephone repair workers. The sample of construction workers and the sample of telephone repair workers include 297 and 192 people, respectively, under 35 years of age.
  - a) Obtain and report a 95% confidence interval to estimate the difference in proportions of people under 35 years of age among construction workers and telephone repair workers.
  - b) Test if there is a significant difference in proportions of people under 35 years of age among construction workers and telephone repair workers. Use the 5% level of significance.
  
- 2) According to a survey, 9% of 780 small businesses use telecommuting of workers in an effort to reduce unscheduled absenteeism. For 915 large companies, this proportion is 6%.
  - a) Is the proportion significant higher for small businesses than for large companies on this issue? Use the 5% level of significance.
  - b) Obtain and report a 95% confidence interval to estimate the difference in proportions for small and large companies.