STAT 210
Applied Statistics and Data Analysis:
Homework 3
Solution

Due on Oct. 22/2020

Question 1

The data set `Pima.tr` in the package `MASS` has measurements taken on a sample of Pima Indian women in a study related to diabetes. Using the help in `R` read about the variables included in the study.

(a) Use this data set to evaluate the hypothesis that the population mean of diastolic blood pressure for Pima Indian women is not 70. Use all the procedures reviewed in class. In each case state clearly the assumptions underlying your procedures and check whether they are verified. Comment on your results.

(b) Use the same data set to test whether the diastolic blood pressure is lower for non-diabetic women, when compared with diabetic women. Use all the procedures reviewed in class. In each case state clearly the assumptions underlying your procedures and check whether they are verified. Comment on your results.

Question 2

An experiment was performed where ten mice were subject to a treatment. The data correspond to the weight of each mice before and after treatment. We want to determine whether the treatment had an impact on the weight of the mice. For this question we will use the data set `mice2` in the package `datarium`.

What test or tests are adequate in this situation? State clearly the hypothesis you want to test in this case. What are the assumptions underlying your procedures? Check whether they are verified. Carry out all the tests suitable in this case. Comment on your results.

Question 3

The data set `properties` in the `datarium` package has information about real estate properties purchased and buyers. Properties are divided into four categories 'flat', 'bungalow' (i.e., a one-storey home), 'detached house' and 'terrace', while buyers are divided into 'single male', 'single female', 'married couple', and 'family'.

(a) Produce a contingency table for this data. Calculate row and column totals.

(b) Represent this data as a mosaic plot.

(c) Test whether these two variables are independent. State clearly the hypothesis that you are testing and verify that the conditions required for the test you choose are satisfied. (Note: If you want a table of expected values, you can get one from the output of the `chisq.test`. To see how, look at the help for this function.)

Question 4

For this question use the data set `PlantGrowth` in the `base` package. The data set compares yields –as measured by dry weight of plants– from a control and two treatments. There are 30 observations on 2 variables:

`weight` (g) and `group` with three levels: `ctrl`, `trt1` and `trt2`. Read the help file for more information.

Do a complete analysis of variance for this set. Determine whether the treatments have an effect of the response by means of a hypothesis test. Plot the diagnostic charts and comment on them. Use also Levene's test and Shapiro-Wilk. Use Tukey's HSD procedure to make pairwise comparisons and comment on the results.