

STAT 847: Reading Assignment 1

Q1. What's the name of the test for outliers used in this chapter?

The test for outliers is called Tietjen-Moore test

Q2. What does the variable 'pdays' represent?

The variable 'pdays' in bank.csv represents number of days that passed after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)

Q3. How many rows are there in the bank marketing data?

We know from the dimensions that there are 11162 rows of data

Q4. What does each row represent in the bank marketing data?

The rows of the dataset represent a list of customers from the bank participating in the a marketing campaign undertaken by a financial institution that assists in analyzing future strategies with a view to improving future marketing campaigns for the bank.

Q5. (Challenge) What is two sample t-test actually comparing in the “the t-test in R” page?
The two sample t-test compares the numeric population means of the two samples.

Q6. What makes a model parsimonious?

Parsimonious models are simple models with great explanatory predictive power which usually explain data with a minimum number of parameters, or predictor variables.

Q7. Almost every named distribution (e.g., the normal, the uniform) has a function that calculates its cumulative distribution function. What is the letter that all such functions start with?

All such functions start with ‘p’ similar to pnorm to find the cumulative distribution function.

Q8. According to the Shapiro-Wilk test, are bank balances normally distributed?

The bank balances are not normally distributed. For data ranging 1:10, p value is 0.01549, while for data ranging 1:4000, the p value $< 2.2e-16$. Therefore, since p value is consistently lesser than 0.05 it is always not normally distributed.