

۱. اندازه گیری و تحلیل پراکندگی داده ها

فرض کنید درصد های کنکور 12 دانش آموز در درس ریاضیات به شرح زیر باشد:

89,50,49,11,44,98,45,57,5,65,62,57

A. با استفاده از خلاصه پنج عددی (Min, Q1, median, Q3, Max) Boxplot را برای این داده ها

رسم کرده و همچنین مقادیر پرت (Outliers) را در این مجموعه داده را اعلام کنید. (باید مقادیر

پرت " $1.5 * IQR$ " را از Boxplot حذف کنید)

B. با توجه به مقادیر میانگین، مد و میانه این مجموعه داده در مورد نوع تقارن داده ها

(positively or negatively skewed) توضیح دهید.

۲. محاسبه شباهت و یا عدم شباهت داده ها

در یک نظرسنجی مربوط به استفاده از خدمات "های وب" 3 کاربر در مورد ویژگیهایی از این شرکت

نظر داده اند. برای هر ویژگی دو حالت وجود دارد: کاربر یا آن ویژگی را دوست دارد (1) یا دوست

ندارد (0) جدول به شکل زیر است:

ویژگی ها	سرعت اینترنت	قیمت مناسب	پشتیبانی	تنوع بسته ها
کاربر A	0	1	1	0
کاربر B	1	1	0	1
کاربر C	0	0	1	0

با توجه به نوع داده ها میزان عدم شباهت بین این سه کاربر را با استفاده از contingency table محاسبه و

ضمناً مشخص کنید هر کدام از این ویژگی ها از چه نوعی (asymmetric OR symmetric binary) هستند؟