

Визуелизација и анализа података

Други колоквијум

25. децембар 2020.

1. Написати *R* скрипту која формира вектор *p* целих бројева од 31 до 60 (укључујући и те бројеве), а затим од тог вектора формира две матрице: матрицу *A* димензије 5×6 , која се попуњава по колонама, матрицу *B* димензије 5×6 , која се попуњава по врстама.
 - а. Одредити вектор *k* свих средњих вредности по колонама матрице *A*. Формирати матрицу *D* тако што се добијени вектор *k* прикључи као шеста врста матрици *A*.
 - б. Да ли матрична једначина $DX = d$, где је *d* вектор-колона који садржи суме по колонама матрице *B*, има решење? Образложити.
2. Уграђена база података *Chickweight* садржи податке о телесној тежини пилића на различитим врстама исхране.
 - а. Формирати *boxplot* за тежину пилића, посебно за сваку од 4 врсте исхране. *Boxplot*-ове представити заједно, на истом графику.
 - б. На основу графика под а. закључити да ли у некој од група има нетипичних вредности. Навести у којој групи и колико их је, уколико постоје. У којој групи је интерквартилни распон највећи, а у којој најмањи? Образложити.
3. Уграђена база података *airquality* садржи податке о дневном квалитету ваздуха у Њујорку од маја до септембра 1973. године.
 - а. Направити нову подбазу података која ће садржати податке о данима у септембру. Ту подбазу користити у следећим тачкама задатка.
 - б. Одредити средњу јачину ветра за дане од 1. септембра закључно са 15. септембром.
 - в. Поделити податке из **целог септембра** у категорије на основу података о дневној температури и то на следећи начин:
 - i. Прва категорија обухвата дане у којима је температура била до првог квартила.
 - ii. Друга категорија обухвата дане у којима је температура била између првог квартила и трећег квартила.
 - iii. Трећа категорија обухвата преостале дане.Резултате представити на *barchart*-у. Различитим бојама представити сваку од категорија. Додати легенду.
4. У фајлу *Pritisak.csv* дати су подаци о притиску и пулсу пацијената.
 - а. Одредити јачину и смер линеарне повезаности између дијастолног и систолног притиска на основу Пирсоновог коефицијента корелације. Уколико је корелација јака, одредити линеарни регресиони модел којим се помоћу вредности дијастолног притиска рачунају вредности систолног притиска. На графику представити податке и регресиону праву. Графику додати наслов.
 - б. Помоћу тог модела одредити предвиђену вредност систолног притиска за пацијента чији је дијастолни притисак 80.
 - в. Пронаћи пар варијабли за које нема сврхе фомирирати линеарни регресиони модел. Образложити избор.