

ЗАДАЦИ

1. Популацију чини база података **trees**, која садржи податке о 31 стаблу.
 - а) Изабрати прост случајан узорак обима $n = 10$ помоћу функције **sample()** и оценити средњу вредност обележја **Volume** користећи добијени узорак.
 - б) Изабрати узорак са понављањем са вероватноћама избора пропорционалним обиму стабла (обележје **Girth**) и оценити средњу вредност обележја **Volume** користећи добијени узорак.
2. У индустријском граду у јужној Индији већина породица се бави текстилном индустријом. У фајлу **baza1.txt** дати су подаци о броју машина за ткање у свакој од 51 улице у том месту. Одредити укупан број машина за ткање у граду користећи прост случајан узорак обима $n = 8$.
3. У фајлу **baza2.txt** дати су подаци о броју радника и производњи у 10 фабрика у индустријској зони. Из ове популације изабрати узорак обима 3 са понављањем са вероватноћама избора пропорционалним броју радника у фабрици. Користећи добијени узорак одредити *Hansen – Hurvitz*-ову оцену укупне производње.
4. У фајлу **baza3.txt** дати су подаци о броју посета позоришту за ученике неке школе, као и стратум ком сваки ученик припада. Ученици су подељени у 3 стратума према успеху у школи. Изабрати стратификовани случајни узорак обима 30, користећи пропорционални распоред, а затим оценити просечан број посета позоришту за ученике те школе, дисперзију оцене и оцену те дисперзије.
5. Из популације обима 75 изабран је прост случајан узорак. У фајлу **baza4.txt** дате су вредности обележја x (обележје које се испитује) и y (помоћно обележје) елемената који су изабрани у узорак. Одредити линеарну регресиону оцену средње вредности и укупне суме обележја x , ако је познато да је $\bar{y} = 5475$.
6. Популацију чини 900 студената. Како изабрати узорак обима 120 користећи системски узорак?
7. База података **Womenlf** (пакет "**car**") састоји се од 263 опсервације и 4 колоне. У колонама су дате вредности следећих обележја, редом: учешће радне снаге (узима вредности: "**fulltime**", ако жена ради пуно радно време, "**not.work**", ако не ради ван куће и "**parttime**", ако ради део времена), приходи мужа (вредности у хиљадама), присуство деце (вредности: "**absent**", ако нема деце у породици, "**present**", ако има деце у породици) и регија ("**Atlantic**", "**BC**", "**Ontario**", "**Prairie**", "**Quebec**").
 - а) Нека је радни статус жена обележје које се испитује, а присуство деце у породици обележје у односу на које се врши стратификација. Изабрати стратификован случајан узорак без понављања обима 100, користећи пропорционални распоред. Оценити проценат жена запослених пуно радно време.
 - б) Нека сваки од региона представља један кластер. Изабрати прост случајан узорак без понављања који чине 3 кластера и на основу тог узорка наћи оцену просечних прихода мушкараца у популацији.
 - в) Ако популацију чине само опсервације из региона **Ontario**, изабрати систематски узорак обима 12, израчунати оцену укупних прихода мушкараца у тој регији и дисперзију те оцене.
8. База података **Mroz** (пакет "**car**") има 753 врсте и 8 колона. У колонама су дате вредности следећих обележја, редом: **lfp** (запосленост), **k5** (број деце од 5 или мање година), **k618** (број деце која имају од 6 до 18 година), **age** (старост жена), **ws** (завршене високе студије супруге), **hc** (завршене високе студије супруга), **lwg** (логаритам очекиваног износа плате - за запослене жене је у питању логаритам износа плате, а за незапослене приписана вредност заснована на осталим променљивама) и **inc** (породично финансијско стање).

- а) Ако је запосленост жена обележје које се испитује, изабрати прост случајан узорак без понављања обима 100, оценити проценат запослених жена и наћи 90% интервал поверења за проценат запослених жена.
 - б) Изабрати узорак са понављањем са вероватноћама избора пропорционалним величини обима 100, где величина сваке јединице представља старост жена (**age**). На основу добијеног узорака одредити *Hansen – Hurwitz*-ову оцену укупне суме обележја **lwg** и оцену дисперзије те оцено.
 - в) Ако популацију чине жене које немају децу узраста од 6 до 18 година, изабрати систематски узорак обима 43, израчунати оцену просечног финансијског стања жена и дисперзију те оцено.
9. База података **Womenlf** (пакет "**car**") састоји се од 263 опсервације и 4 колоне. У колонама су дате вредности следећих обележја, редом: учешће радне снаге (узима вредности: "**fulltime**", ако жена ради пуно радно време, "**not.work**", ако не ради ван куће и "**parttime**", ако ради део времена), приходи мужа (вредности у хиљадама), присуство деце (вредности: "**absent**", ако нема деце у породици, "**present**", ако има деце у породици) и регион ("**Atlantic**", "**BC**", "**Ontario**", "**Prairie**", "**Quebec**").
- а) Ако је присуство деце у породици обележје које се испитује, изабрати прост случајан узорак са понављањем обима 50, оценити проценат породица које имају децу, одредити дисперзију те оцено и оцену те дисперзије.
 - б) Нека су приходи мушкараца обележје које се посматра, а присуство деце у породици обележје у односу на које се врши стратификација. Извести формулу за обиме узорака по стратумима користећи оптимални распоред. Изабрати стратификован случајан узорак без понављања користећи оптимални распоред ако је познато да су трошкови за избор елемената у одговарајући стратум: $c_1 = 100$ и $c_2 = 400$, фиксни трошкови $c_0 = 150$ и укупни (унапред фиксирани) трошкови $C = 8000$. Користећи добијени узорак наћи оцену укупних прихода мушкараца у популацији.
 - в) Нека сваки од региона представља један кластер. Изабрати прост случајан узорак без понављања који чине 2 кластера, одредити дисперзију оцено просечних прихода мушкараца у популацији и на основу узорака наћи оцену те дисперзије.
10. База података **Mroz** (пакет "**car**") има 753 врсте и 8 колона. У колонама су дате вредности следећих обележја, редом: **lfp** (запосленост), **k5** (број деце од 5 или мање година), **k618** (број деце која имају од 6 до 18 година), **age** (старост жена), **ws** (завршене високе студије супруге), **hc** (завршене високе студије супруга), **lwg** (логаритам очекиваног износа плате - за запослене жене је у питању логаритам износа плате, а за незапослене приписана вредност заснована на осталим променљивама) и **inc** (породично финансијско стање).
- а) Извести формулу за одређивање обима простог случајног узорака без понављања за оцењивање средње вредности обележја популације. Одредити обим узорака за оцењивање средње вредности обележја **inc** ако је из претходних испитивања познато да је дисперзија приближно једнака $s^2 \approx 1.2$. Прихватљиво одступање је 7% са ризиком 4%. Изабрати прост случајан узорак без понављања тог обима и наћи 90% интервал поверења за обележје **inc**.
 - б) Нека је породично финансијско стање **inc** обележје које се посматра, а завршене високе студије супруге **ws** променљива у односу на коју се врши стратификација. Изабрати стратификовани случајан узорак обима 150 користећи пропорционални распоред. Наћи оцену средине обележја **inc**, дисперзију оцено и оцену те дисперзије.
 - в) Из популације изабрати прост случајан узорак без понављања обима 120. Наћи количничку оцену средине обележја **lwg** ако је **inc** помоћно обележје, израчунати дисперзију те оцено и оцену те дисперзије.