

Beadandó feladat

Idősorok és többdimenziós statisztikai gyakorlat
2024/2025. őszi félév

Technikai tudnivalók

A futtatásokhoz a forráskódot R nyelven kell megírni. A megoldást E-mail-ben küldjétek el, csatoljátok a forráskódot a levélhez!

A megoldás elvárt formátumai:

- R-es forráskód .r (nem Word, txt stb.) kiterjesztésű fájlban;
- a szövegfájl pdf, Word vagy html formátumban. Ha Rmd-ben készíted, akkor az Rmd fájlt és a lefordított pdf vagy html fájlt is csatold!

A fájlok és a változók nevében ne használjatok magyar ékezetes betűket!

Beadási határidő: 2024. október 13., 18:00

Késve beadott megoldás nulla pontot ér.

Feladat leírása

A Morpho Stacey befektetési bank matematikai elemzője vagy. Azt a feladatot kapod, hogy vizsgáld az egyes munkakörök közti átjárást, a karrierutakat, illetve azt számszerűsítsd, a legfőbb versenytárs, a Golden Saints milyen mértékben szívta el a munkaerőt az elmúlt 30 évben. A HR részleg a fizetésekről az alábbi összesített értékeket adta a 2024. január 1-jei bérszámfejtés alapján (minden érték \$-ban, éves bérek mindenféle kiegészítő juttatással):

Rang/munkakör	Átlag	Szórás	Alsó kvartilis	Medián	Felső kvartilis
Analyst (An)	121 708	17 638	98 342	120 472	133 914
Associate (As)	169 743	31 943	144 421	168 962	192 031
Vice President (VP)	206 840	47 463	176 431	205 234	239 523
Executive Director (ED)	363 064	109 425	306 312	355 264	427 973
Managing Director (MD)	647 945	450 962	302 521	620 763	820 343

A táblázatban fentről lefelé folyamatosan nő az alkalmazottak rangja, a munkakör egyre több felelősséggel és egyben magasabb fizetéssel jár.

Mint a magánszférában általában, a vállalatok számára kulcsfontosságú, hogy a legjobbak dolgozzanak náluk, hozzásegítve őket piaci pozíciójuk megtartásához vagy javításához. Ez a helyzet azzal jár, hogy a munkavállalók sokszor megversenyeztetik a munkaadókat, ki kínálja a jobb lehetőségeket (érdekesebb munka, kedvezőbb kompenzációs csomag) és akár rövid idő után is egy másik vállalathoz mennek dolgozni. A honlapomon lévő `Bead24.Rdata` állományban található `rangok` nevű 3 dimenziós tömb az adott év során a munkakörök közötti átlépések számát, illetve a Morpho Stacey-től kilépők számát tartalmazza:

1. dimenzió: melyik munkakör kategóriából indul: An, As, VP, ED, MD
2. dimenzió: melyik kategóriába lép át: An, As, VP, ED, MD, El_GS és El_mas, ahol az El_GS arra utal, hogy a munkavállalók a legfőbb versenytárs Golden Sains kötelékébe léptek át, míg az El_mas azt takarja, hogy egy másik céghez mentek vagy bármi más okból léptek ki (például nyugdíjba vonultak). Nem különböztetjük meg azokat, kiket bocsátott el a bank, illetve kik mondtak fel maguktól.
3. dimenzió: idő (év), ami az átlépések (előléptetések/lefokozások) évére utal

Modellezzük a bank munkakörei közti átjárást/elvándorlást Markov-lánccal!

Kidolgozandó kérdések/feladatok:

- Melyek a Markov-lánc állapotai? Ábrázold a Markov-lánc gráfját és osztályozd az állapotokat!
- Becsüld meg az átmenetvalószínűségi mátrixot!

- c.) Adj 95%-os megbízhatóságú intervallumbecslést az egyes munkakörökből annak a valószínűségére, hogy a legfőbb versenytárhoz megy át egy adott év során egy alkalmazott!
- d.) Várhatóan hány év múlva lesz egy analyst-ből associate?
- e.) Melyik a valószínűbb – az, hogy egy analyst-et associate-ként előléptetnek, vagy az, hogy még analyst-ként elmegy a banktól?
- f.) Becsüld meg, hogy mennyi a vállalat 2024. évi éves bérköltsége! Ezt soknak/kevésnek tartod? Hasonlítsd össze Magyarország éves költségvetésének bevételi főösszegével!
- g.) Írj egy rövid, 3-6 mondatos összefoglalót a felső vezetés számára a munkakörök közti átlépésekről. Mindenképp térj ki arra, hogy mik a főbb megfigyeléseid arra vonatkozóan, a fő versenytárs mely munkakörökben dolgozókat szívja el leginkább. Próbálj javaslatokat tenni arra, hogy miként csökkenthető a cégtől az elvándorlás és a felső vezetésnek mi lenne a teendője ezzel kapcsolatban!

Szorgalmi (SZ) kérdés: Illessz az MD-k keresetére egy alkalmas vastag szélű eloszlást, ellenőrizd annak megfelelő illeszkedését, majd adj 95%-os intervallumbecslést az MD-k éves keresetére! Miért lenne biztosan helytelen az MD-k esetén például normális eloszlást illeszteni a fizetésekre?

Útmutatók:

- Az Rdata fájlok beolvasása a `load` paranccsal megy. Esetünkben a `load("BeadNagy.Rdata")` tölti be az Rdata fájl tartalmát a globális környezetbe. A beolvasott `rangok` nevű tömb dimenzióinak az elnevezéseit a `dimnames(rangok)` paranccsal tudjátok lekérdezni.
- Az `apply` függvénnyel gyorsan lehet sorösszegeket számolni, például az `x` mátrix sorösszegét az `apply(x, 1, sum)` paranccsal számolhatod ki. Itt az 1-es arra utal, hogy az 1. dimenziót rögzíted, és a másik dimenzió szerint végzed az összegzést. Az `apply` magasabb dimenzióban is így működik, például egy 3 dimenziós `y` tömbre használható az `apply(y, c(1, 2), mean)` függvény, ami a 3. dimenzió szerinti átlagolást fogja végrehajtani.
- Az intervallumbecslést legegyszerűbb a megfelelő tapasztalati kvantilisek segítségével konstruálni. Kvantiliseket a `quantile` függvénnyel lehet számolni, például egy `x` vektorra a `quantile(x, c(0.6, 0.8), type=6)` a 0,6-os és 0,8-as tapasztalati kvantilist adja meg. Mivel a tapasztalati kvantilisbecslés nem egyértelmű, több beépített módszer közül lehet választani. A statisztika kurzusok többségében oktatott képlet a `type=6` beállításnak felel meg.
- Az 1. oldalon lévő táblázat számai csak az f.) feladathoz és a szorgalmihoz szükségesek.

Pontozás:

Feladatok	a.)	b.)	c.)	d.)	e.)	f.)	g.)	Összesen	SZ	Mindösszesen
Maximális pontszám	5	5	5	3	4	4	4	30	4	34