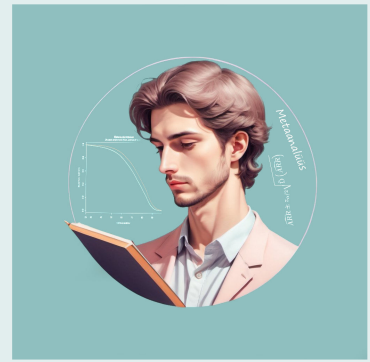




Biostatistika

kodune ülesanne 2



Indialaste puuduvad andmed

Andmete taust.

Käesoleva sajandi alguses oli teada, et geeni SLC24A5 mutatsioon positsioonis rs1426654 muudab inimese nahavärvi (kui omad antud kohas DNAs tähte G, oled tumedama nahaga, kui aga kannad mutatsiooni ehk tähte A siis oled heledama nahavärviga – mutatsioon häirib tõsiselt nahale tumeda tooni andva melaniini tootmist). Aafrikast väljarännanud inimestel tekkisid mitmed mutatsioonid, mis aitasid neil uue keskkonnaga kohaneda ja muutsid nende nahavärvi heledamaks, nende seas ka kõnealune mutatsioon mis lööb rivist välja geeni SLC24A5 töö. Aga peale Aafrikast väljarändamist suundusid osad inimesed tagasi lõunapoolsetesse piirkondadesse, nagu näiteks lõuna-Indiasse ja neil osutus hädavajalikuks oma nahka kaitseks intensiivse päikesekiirguse vastu taas tumedamaks muuta. Aga kuidas see muutus toimus? Kas leiutati uus mehhanism, mis võimaldas vaatamata rikkis SLC24A5 geenile ikka piisavalt palju melaniini toota (st need inimesed oleksid sõltumata vaatluseluse mutatsiooni olemasolust või mitteolemasolust ikka tumeda nahaga) või tuleneb lõuna-indialaste tumedam nahavärv ikkagi sellest, et nende sekka on hiljem sattunud algset geenivarianti omavaid inimesi (Aafrikasi?), kelle geneetiline panus on aidanud kohalikel uuesti kohaneda eluks intensiivse päikesegea piirkondades?

Selle uurimiseks otsustati ühes India maapiirkonnas, kus teadaolevalt esines nii mutatsiooni kandjaid kui ka mittekanjdajaid, uurida mutatsiooni mõju (kas see on samasugune kui Ameerikas elavatel aafrika/euroopa päritolu segaverelistel).

Kuna geenimutatsioonide määramine oli tol ajal veel kallis, sooviti hakkama saada nii, et võimalikult vähestel inimestel mõõdetakse genotüüp (kas nad kannavad ühte, kahte või mitteühtegi muteerunud geenivarianti). Sestap otsustati mutatsiooni mõju uurimiseks genotüpiseerida antud maapiirkonnast vaid kõige heledama ja kõige tumedama nahavärviga inimesi (nahavärvi mõõtmine vastava aparaadiga Indias oli äärmiselt väikeste tööjõukuludega ettevõtmine). Saadi andmestik, kus on kirjas inimese nahavärv (kui tumeda nahaga ta on, tunnus *melaniin*, mõõdetud viisil mis ei lase end mõjutada päevitusest tingitud naha tumenemisest), tema *sugu* ja osadel inimestel ka tema rs1426654 genotüüp (AA/AG/GG). Saadi tulemuseks, et genotüübi ja nahavärvi vahel ikkagi eksisteerib seos:

```
> print(load(url("https://www-1.ms.ut.ee/mart/biostat/puuduvad.RData")))
> m1=lm(melaniin~factor(genotyyp), data=andmed)
> drop1(m1, test="F")
              Df Sum of Sq      RSS      AIC F value    Pr(>F)
<none>                7301.5  830.47
factor(genotyyp)  2      9175.7 16477.2 1023.43  150.17 < 2.2e-16 ***
```

Paraku tekitas saadud tulemus otsekohe uusi küsimusi: kas antud mõju on sama tugev kui teistel (kellel pole ehk Indias „väljatootatud“ mutatsiooni mõju nõrgendavat uut mutatsiooni)? Kui mingit erilist

mutatsiooni vastu kaitsvat lisamehhanismi pole, siis muudab kahelt vanemalt muteerunud geenivariandi pärimine inimese nahavärvhi hinnanguliselt 7,6..11,4 ühikut heledamaks. Kas Indialastel on antud muutus samasuur või ehk väiksem? Proovi seda uurida esmalt kasutades lihtsat analüüsi:

```
m1=lm(melaniin~factor(genotyypp), data=andmed)
summary(m1)
```

ja vasta järgmistele küsimustele:

Ülesanded

1. Milline on puuduvate andmete tekkemehhanism antud uuringus (MCAR/MAR/NMAR). Põhjenda oma arvamust!
2. Arvestades väljapakutud puuduvate andmete tekkemehhanismi – kas väljapakutud lihtne analüüs võiks viia objektiivsete hinnanguteni? Kas võiksime kasutada mitmest imputeerimist? Põhjenda/arutle!
3. Sõltumata sellest, milline oli sinu vastus 2. küsimusele, hinda genotüübi mõju melaniinile kasutades mitmest imputeerimist. Milliste tulemusteni jõuad?
4. Anna vastus ka uurimisküsimusele – kas on põhjust arvata, et lõuna-India elanikud on evolutsiooni käigus hankinud endale mingi erilise uue võtte, mis aitab neil SLC24A5 mutatsiooni mõju vähendada (ja seeläbi saavutada nahavähi eest kaitsvat tumedamat nahavärvi)?