Ocenjevanje starosti odraslih indijcev iz ortopantomografije.

Urh Peček in Alen Kahteran

03 januar 2021

1. Predstavitev raziskovalnega problema

Pogosto je zaradi nekaterih vprašanj v forenzičnem kontekstu potrebno oceniti starost posameznika. Čeprav je za oceno starosti mogoče uporabiti več delov telesa so zobje tisti del telesa na podlagi katerega se najpogosteje uporablja identifikacija in ocena starosti. Pri otrocih je določanje starosti na podlagi zob razmeroma preprost in natančen postopek, vendar pa pri odraslih to predstavlja izziv. Nekatere metode določanja starosti na podlagi zob so zelo zapletene in uničujoče, zato se običajno ne uporabljajo. Preučevanje rentgenskih slik zob je neuničujoč postopek, ki ga je mogoče uporabiti tako pri živih kot pokojnih posameznikih in je metoda, ki je v nasprotju z drugimi dolgotrajnimi, dražjimi in bolj uničljivimi metodami. Poleg tega se izogiba pristranskosti, ki je značilna za subjektivnega opazovalca in tako izboljša zanesljivost in natančnost.

Za oceno starosti lahko uporabimo katerikoli zob, vendar so podočniki dober kandidat za oceno starosti saj so pogosto prisotni tudi pri starejših ljudeh in jih je zaradi največje površine pulpe najlažje analizirati.

Cilj študije je bil razviti metodo za ocenjevanje kronološke starosti odraslih indijcev na podlagi razmerja med starostjo in različnimi morfološkimi spremenljivkami prednjih zob.

2. Opis podatkov in njihove pridobitve

Za dosego potrebovane pomembnosti in moči študije so bile v zobozdravstvenem centru v Indiji 60 moškim in 60 ženskam, odvzete rentgenske slike prednjih zob. Preiskovanci so bili razdeljeni v štiri skupine glede na starost: 21-30 let, 31-40 let, 41-50 let in 51-60 let. Sodelovanje v študiji je bilo prostovoljno in pacienti so zadoščali naslednjim kriterij:

- Starost od 21 do 60 let
- Popolnoma izraščen desni maksilarni podočnik
- Popolnoma oblikovana korenina desnega maksilarnega podočnika
- Desni maksilarni podočnik je brez kakršne koli patologije
- Odsotnost restavracije in endodontskega polnjenja desnega maksilarnega podočnika
- Odsotnost neuravnanega ali zasukanega desnega podočnika

Upoštevana starost je bila izračunana z razliko med datumom radiografije in datumom rojstva.

Ortopantomografi izbranih bolnikov so bili pridobljeni z uporabo optičnega bralnika, slike pa obdelane z uporabo računalniško podprtega programa (CAD). Nato so bile na podlagi radiografskih slik ocenjene meritve območja podočnika. Na podlagi slik zob so bila preučena naslednje lastnosti oziroma morfološke spremenljivke:

- AR = razmerje med površino pulpe in zoba
- p = razmerje med dolžino pulpe in korenine
- r = razmerje med dolžino pulpe in zoba
- a = razmerje širine pulpe in korenine na ravni CEJ
- c = razmerje med širino pulpe in korenine na ravni srednjega korena
- b = razmerje med širino pulpe in korenine na srednji ravni med nivojem CEJ in nivojem korenov

Preiskovanci so bili razdeljeni v štiri skupine glede na starost: 21-30 let, 31-40 let, 41-50 let in 51-60 let.

3. Uporabljene statistične metode

Statistična analiza je bila izvedena s pomočjo programa SPSS. Kot statistično pomembna vrednost pa je bila obravnavana vrednost $p < 0.05$.

Pearsonov koeficient korelacije

Ena izmed najosnovnejših statističnih metod je Pearsonov koeficient korelacije. Ta se uporablja za merjenje moči povezanosti med dvema spremenljivkama. Odgovori nam na dve vprašanji in sicer ali med spremenljivka obstaja linearna povezanost in kako močna je le ta. V našem primeru je bil Pearsonov koeficient korelacije uporabljen za izračun povezanosti med opazovano starostjo in morfološkimi napovednimi spremenljivkami.

Pogoj za izračun Pearsonovega koeficienta korelacije je linearna odvisnost obeh vpletenih spremenljivk, medtem ko se za določanje povezanosti spremenljivk, ki niso povezane linearno, običajno uporablja Spearmanov koeficient korelacije.

Pearsonov koeficient korelacije dveh številskih spremenljivk X in Y se izračuna kot razmerje med kovarianco in produktom obeh standardnih odklonov

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma(X)\,\sigma(Y)}\,.$$

Cenilka za populacijski korelacijski koeficient je vzorčni korelacijski koeficent:

$$r = \frac{S_{XY}}{S_X S_Y}\,.$$

Kovarianca je mera linearne povezanosti med spremenljivkama in njeno cenilko izračunamo po formuli

$$S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})\cdot(Y_i - \overline{Y})$$

kjer sta S_X^2 in S_Y^2 vzorčni varianci

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2\,.$$

Tu je n velikost vzorca ter $\overline{X}$ in $\overline{Y}$ vzorčni povprečji posamezne spremenljivke.

Pearsonov koeficient korelacije vrne vrednosti med -1 in 1 tj. $\rho_{XY} \in [-1,1]$, kjer 1 pomeni popolno pozitivno povezanost in -1 popolno negativno povezanost. V primeru $\rho_{XY} = 1$ med X in Y obstaja linearna funkcijska zveza $Y = a + bX$. Kadar ne gre niti za pozitivno niti za negativno povezanost rečemo, da spremenljivki med seboj nista linearno povezani oziroma nekorelirani in tedaj je koeficient blizu 0. Če je $\rho_{XY} = 0$ sta spremenljivki porazdeljeni normalno, potem sta tudi neodvisni. Če je $\rho_{XY} \neq -1, 0, 1$ potem med X in Y ni linearne povezanosti. V praktičnem preizkušanju odvisnosti je skoraj nemogoče izračunati popolno odvisnost -1 ali 1, saj na posamezno odvisno spremenljivko praviloma vpliva več dejavnikov, med njimi tudi slučajni vplivi.

Pri statističnem sklepanju o nekoreliranosti spremenljivk X in Y, oziroma v primeru normalne porazdeljenosti neodvisnosti, postavimo ničelno domnevo $H_0 : \rho_{XY} = 0$ (spremenljivki nista linearno povezani) naproti alternativni domnevi $H_A : \rho \neq 0$ (spremenljivki sta linearno povezani).

Za preverjanje te domneve uporabimo testno statistiko T, ki jo izračunamo kot

$$T = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}\,,$$

kjer je r vzorčni korelacijski koeficient in n velikost vzorca. Ta se porazdeljuje po studentovi oziroma t porazdelitvi z $n-2$ stopinjami prostosti. Test je dvostranski in pri stopnji značilnost α izberemo tak t_α, da velja

$$P(|T| < t_\alpha) = 1 - \alpha\,.$$

Če je $|T| > t_\alpha$ ničelno hipotezo zavrnemo in potrdimo, da sta spremenljivki linearno povezani oziroma korelirani. Če je $|T| < t_\alpha$ ničelno hipotezo obdržimo. Seveda se lahko o hipotezi odločimo tudi na podlagi vrednosti p in če je $p < \alpha$ ničelno hipotezo zavrnemo.

V našem primeru spremenljivka Y predstavlja opazovano starost pacientov, spremenljivka X pa vrednost ene izmed izbranih morfoloških napovednih spremenljivk.

Poudariti velja, da Pearsonov koeficient korelacije govori o povezanosti dveh spremenljivk, ne pa tudi o vplivu ene spremenljivke na drugo. Poleg tega Pearsonov koeficient korelacije ne more ugotoviti razlike med odvisnimi in neodvisnimi spremenljvkami, zato se moramo kot raziskovalec zavedati podatkov. Pearsonov koeficient korelacije nam prav tako ne bo dal nobenih podatkov o naklonu premice na grafu, ki opisuje linearno povezanost spremenljivk.

Uporaba pearsonovega koeficienta korelacije je najpogostejša v linearni regresiji in enačbo, ki najbolje opisuje linearno odvisnost obeh spremenljivk, je moč izračunati z le to. Ta enačba je ob visokih vrednostih Pearsonovega koeficienta najbolj točna.

Multipla linearna regresija

Multipla linearna regresija, znana tudi kot linearna regresija je statistična tehnika, ki uporablja več pojasnjevalnih spremenljivk za napovedovanje izida odzivne spremenljivke. Je metoda oziroma funkcija, ki analitiku omogoča napovedovanje vrednosti ene spremenljivke na podlagi informacij, ki so znane o drugih spremenljivkah. Cilj multiple linearne regresije je modeliranje linearne povezanosti med pojasnjevalnimi (neodvisnimi) spremenljivkami in odzivno (odvisno) spremenljivko.

Formulo, ki predstavlja multiplo linearno regresijo lahko zapišemo kot

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \epsilon_i\,,$$

kjer je y_i odzivna spremenljivka, x_i pa pojasnjevalna oziroma napovedna spremenljivka. β_j so parametri oziroma koeficienti modela in ϵ_i modelska napaka oziroma ostanek.

V našem primeru je bil model multiple linearne regresije uporabljen za oceno starosti v odvisnosti od morfoloških napovednih spremenljivk. V začetku smo model zapisali kot

$$Starost_i = \beta_0 + \beta_1 \cdot AR_i + \beta_2 \cdot p_i + \beta_3 \cdot r_i + \beta_4 \cdot a_i + \beta_5 \cdot c_i + \beta_6 \cdot b_i + \epsilon_i$$

za $i = 1, \ldots, n = 120$. Kasneje pa je bil model multiple linearne regresije za oceno starosti razvit z izbiro tistih morfoloških spremenljivk, ki so pomembno prispevale k oceni starosti in zapisan kot

$$Starost_i = \beta_0 + \beta_1 \cdot AR_i + \beta_2 \cdot c_i + \epsilon_i$$

za $i = 1, \ldots, n = 120$.

Model multiple linearne regresije temelji na naslednjih predpostavkah:

- Linearnost odvisnosti odzivne spremenljivke od napovednih spremenljivk. V našem primeru to pomeni, da je opazovana starost pacienti v linearni zvezi z izmerjenimi vrednostmi morfoloških napovednih spremenlj.
- Varianca napak oziroma odzivne spremenljivke pogojno na napovedne spremenljivke je konstantna. Torej je razlika med opazovano starostjo in njeno napovedjo izračunano na podlagi modela neodvisna od napovednih spremenljivk ali katerih drugih dejavnikov, ki v model niso vključeni.
- Pričakovana vrednost napake je 0, kar pomeni, da je vsota odstopanj opazovane od ocenjenih starosti približno 0.
- Napake so med seboj neodvisne, kar je v težko preveriti in je predvsem odvisno od kakovosti vzorčenja, kar smo pri nas predpostavili, da je bilo kakovostno.

Statistična mera na podlagi katere določamo kakovost linearnega regresijskega modela je koeficient multiple determinacije, ki ga označimo z R^2. Pove nam kolikšen delež variabilnosti odzivne spremenljivke, torej starosti smo uspeli razložiti z napovednimi v našem primeru izmerjenimi morfološkimi spremenljivkami. R^2 se vedno poveča z dodajanjem napovednih spremenljivk v model, čeprav nekatere niso nujno povezane z odzivno spremenljivko. Koeficienta determinacije samega po sebi tako ni mogoče uporabiti za ugotavljanje, katere napovedne spremenljivke je potrebno vključiti v model in katere morda izključiti. Njegova vrednost se giblje med 0 in 1, pri čemer vrednost 0 pomeni, da izida na podlagi modela ni mogoče napovedati, vrednost 1 pa, da je izid mogoče napoveda brez napak. R^2 je definiran kot

$$R^2 = \frac{SS_{model}}{SS_{yy}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}.$$

V nadaljevanju študije je bil za natančnost napovedi razvit model s katerim lahko ločeno napovemo starosti moških in žensk. To je bilo storjeno z vključitvijo interakcije med spremenljivkama, kjer je bila kot napovedna spremenljivka dodana spol, ki ima dve vrednosti in sicer moški in ženski. Recimo, da jo označimo z w in da pri moških zavzame vrednost 0 ter pri ženskah vrednost 1. Končni model lahko zapišemo kot

$$Starost_i = \beta_0 + \beta_1 \cdot AR_i + \beta_2 \cdot c_i + \beta_3 \cdot AR_i \cdot w_i + \beta_4 \cdot c_i \cdot w_i + \beta_5 \cdot w_i + \epsilon_i$$

za $i = 1, \ldots, n = 120$.

Enosmerna analiza variance - ANOVA

Analiza variance (ANOVA) je podoben statistični test kot test t za neodvisne vzorce, le da se pri analizi variance primerja povprečja treh ali več skupin med seboj. V primeru samo dveh skupin oziroma vzorcev data test t in ANOVA enake rezultate. V primeru analize več kot dveh vzorcev pa test t ne bo več zanesljiv saj ima je za testno statistiko s katero preverjamo ničelne domneve potrebno uporabiti multivariatno t-statistiko, ki ima vpliv na stopnjo napak rezultata.

V naši analizi je bil uporabljen enosmerni ANOVA test, poznamo pa tudi dvosmernega. Ime se nanaša na število neodvisnih spremenljivk v analizi testa variance. Enosmerna ANOVA ocenjuje vpliv samo enega dejavnika na spremenljivko samo enega odziva in določa ali so vzorci enaki. Dvosmerna ANOVA pa vključuje dve neodvisni spremenljivki, ki vplivata na odvisno spremenljivko. Uporablja se za opazovanje interakcije med dvema dejavnikoma in hkrati preizkuša učinek več dejavnikov.

Pri nas je bila enosmerna analiza variance uporabljena za primerjavo razlik med opaženo in ocenjeno starostjo v različnih starostnih skupinah.

Pogoj za uporabo ANOVE je, da so podatki porazdeljeni normalno, kar lahko preverimo z različnimi grafičnimi in statističnimi testi. Temeljna načina sta histogram in Q-Q diagram podatkov. Najbolj znana statistična testa za preverjanje normalne porazdeljenosti podatkov sta Shapiro-Wilkov test, Kolmogorov-Smirnov test. V naši analizi to pomeni, da so razlike med opazovanimi in ocenjenimi starostmi porazdeljene normalno.

Drugi pogoj za uporabo ANOVE je homoskedastičnost med skupinami, torej morajo biti variance proučevanih populacij za vsako skupino enake, med njimi ne sme obstajati statistično značilno pomembna razlika. V našem primeru morajo biti torej variabilnosti razlik opazovanih in ocenjenih starosti enake med vsemi starostnimi skupinami.

Pri izračunu ANOVE uporabljamo dve vrsti povprečnih vrednosti, to so vzorčna povprečja $\mu_1, \mu_2, \ldots, \mu_m$, katere pri nas predstavljajo povprečne razlike opazovane in ocenjene starosti za $m = 4$ starostne skupine in "veliko" povprečje μ, ki je povprečje vseh opazovanj skupaj, torej povprečna razlika ne glede na starostno skupino.

ANOVA z F-testom primerja del variabilnosti, ki je pojasnjena z razlikami med skupinami in del variabilnosti, ki je prisotna znotraj skupin. Uporablja se lahko samo na številskih razmernostnih spremenljivkah, pri nas razlikah opazovanih in ocenjenih starostih. Pove nam ali se tri ali več skupin, pri nas so to starostne skupine, med seboj statistično značilno razlikuje v določeni lastnosti, v našem primeru v razliki med opazovano in ocenjeno starostjo. Ne pove pa nam kateri dve ali več teh skupin se razlikujejo, za ta podatek bi bilo potrebno uporabiti post-hoc test.

Ničelna hipoteza v ANOVI velja, če so vsa vzorčna povprečja enaka oziroma med njimi ni statistično značilne razlike, torej med starostnimi skupinami ni statistično značilne razlike med povprečnimi razlikami opazovane in ocenjene starosti. Na drugi strani alternativna hipoteza predvideva, da obstaja vsaj eno vzorčno povprečje, ki je statistično značilno različno od ostalih.

$$H_0 : \mu_1 = \mu_2 = \ldots \mu_m = \mu$$

in

$$H_1 : \mu_i \neq \mu_j,\ \exists\, i \neq j\,.$$

Za izračun ANOVE je najprej potrebno izračunati variabilnost med skupinami. Potrebno je upoštevati tehtano odstopanje glede na velikost vzorca, torej skupine. Odstopanje dobi večjo težo, če gre za večji vzorec. Vsako kvadratno odstopanje pomnožimo z velikostjo vzorca in jih seštejemo. To se imenuje vsota kvadratov za spremenljivost med skupinami.

$$SS_{between} = n_1(\overline{x_1} - \overline{x_G})^2 + n_2(\overline{x_2} - \overline{x_G})^2 + \ldots + n_m(\overline{x_m} - \overline{x_G})^2\,.$$

Da dobimo dobro mero variabilnosti med skupinami moramo tehtano vsoto kvadriranih odklonov deliti še s stopinjami prostosti $m - 1$,

$$MS_{between} = \frac{n_1(\overline{x_1} - \overline{x_G})^2 + n_2(\overline{x_2} - \overline{x_G})^2 + \ldots + n_m(\overline{x_m} - \overline{x_G})^2}{m-1}\,.$$

Variabilnost znotraj skupin izmerimo tako, da izračunamo kako se posamezna vrednost v posamezni skupini razlikuje od povprečja skupine. Naredimo vsoto kvadriranih odklonov od povprečja znotraj skupine.

$$SS_{within} = \sum(x_{i1} - \overline{x_1})^2 + \sum(x_{i2} - \overline{x_2})^2 + \ldots + \sum(x_{im} - \overline{x_m})^2\,.$$

Tako kot pri variabilnosti med skupinami moramo vsoto kvadriranih odklonov znotraj skupine deliti s stopinjami prostosti. Tokrat stopinje prostosti predstavlja vsota velikosti skupin $N = n_1 + n_2 + \ldots n_m$ minus število skupin m,

$$MS_{within} = \frac{\sum(x_{i1} - \overline{x_1})^2 + \sum(x_{i2} - \overline{x_2})^2 + \ldots + \sum(x_{ik} - \overline{x_k})^2}{N - m} \, .$$

Statistika s katero preverjamo ničelno domnevo pri testu ANOVA je

$$F = \frac{MS_{between}}{MS_{within}}$$

in se pri ničelni domnevi porazdeljujejo po porazdelitvi F z $m-1$ in $N-m$ stopinjami prostosti. Ima enorepo zavrnitveno območje in velja, da manjša kot je vrednost F bolj podobna so vzorčna povprečja med skupinami in težje zavrnemo ničelno hipotezo o enakih povprečnih vrednostih. Ničelno hipotezo pri stopnji značilnosti α zavrnemo v primeru

$$F > F_\alpha(m-1,\ N-m)\,,$$

sicer ničelno hipotezo obdržimo.

Studentov test t

Studentov test t je eden najpogosteje uporabljenih testov s katerimi preverjamo ničelne hipoteze pri majhnih vzorcih. Je vsak test pri katerem se testna statistika pod ničelno domnevo porazdeljuje po studentovi oziroma t porazdelitvi. Poznamo tri osnovne teste:

- Test t za en vzorec pri katerem primerjamo povprečje v populaciji μ. Velja:

  $$H_0 : \mu = \mu_0\,,$$
  $$H_A : \mu \neq \mu_0$$

  in testna statistika je enaka

  $$T = \frac{\overline{x} - \mu_0}{s\sqrt{n}}$$

  in se pri ničelni hipotezi porazdeljujejo po porazdelitvi t z $n-1$ stopinjami prostosti.

- Test t za parne meritve oziroma odvisna vzorca pri katerem na eni obravnavani enoti primerjamo 2 spremenljivki. Primerjamo povprečji prve in druge spremenljivke. Velja:

  $$H_0 : \mu_A = \mu_B\,,$$
  $$H_A : \mu_A \neq \mu_B$$

  in testna statistika je enaka

  $$T = \frac{\overline{x_A} - \overline{x_B}}{s_r\sqrt{n}}$$

  in se pri ničelni hipotezi porazdeljujejo po porazdelitvi t z $n-1$ stopinjami prostosti. s_r označuje vzorčno standardno napako izračunano na razliki spremenljivk.

- Test t za neodvisna vzorca, kjer predpostavimo enaki varianci. Uporabimo ga v primeru, ko na dveh populacijah obravnavamo eno spremenljivko, pri čemer ima prva populacija n_1 enot in druga populacija n_2 enot. Primerjamo povprečji prve in druge populacije. Velja:

$$H_0 : \mu_1 = \mu_2 \,,$$
$$H_A : \mu_1 \neq \mu_2$$

in testna statistika je enaka

$$T = \frac{\overline{x_1} - \overline{x_2}}{s_r \sqrt{1/n_1 + 1/n_2}} \,,$$

kjer je

$$s_r^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}$$

in se pri ničelni hipotezi porazdeljujejo po porazdelitvi t z $n_1 + n_2 - 2$ stopinjami prostosti.

V naši analizi je bil za primerjavo morfoloških spremenljivk moških z morforoliškimi spremenljivkami žensk uporabljen studentov test t za neodvisna vzorca. Za primerjavo opazovane in ocenjene starosti pa je bil uporabljen studentov test t za parne meritve oziroma odvisna vzorca.

Kappa statistika

Cohenov kappa koeficient κ je merilo zanesljivosti ujemanja med dvema ali več ocenjevalci (in tudi zanesljivosti znotraj ocenjevalca). Upošteva tudi možnost, da do dogovora pride po naključju. Če se ocenjevalci popolnoma strinjajo je $\kappa = 1$ in če med ocenjevalci ni ujemanja, razen tistega kar bi po naključju pričakovali, je $\kappa = 0$. Če je statistika negativna, to pomeni, da med ocenjevalci ni učinkovitega ujemanja ali pa je ujemanje slabše od naključnega.

Definicija kappe je

$$\kappa = \frac{p_0 - p_e}{1 - p_e} = 1 - \frac{1 - p_0}{1 - p_e} \,,$$

kjer je p_0 relativno opazovano ujemanje med ocenjevalci (natančnost), p_e pa hipotetična vrednost naključnega ujemanja.

S kappa statistiko je bila v naši analizi izmerjena vrednost ujemanja med ocenjevalci. Torej njihovo strinjanje glede uporabljenih metod in zaključnih rezultatov.

4. Predstavitev rezultatov

- Vrednost kappa statistike $\kappa = 0.94$ je pokazala, da je bilo ujemanje med ocenjevalci skoraj popolno.
- Morfološke spremenljivke niso pokazale statistično pomembne razlike med spoloma ($p > 0.05$).
- Med vsemi vključenimi morfološkimi spremenljivkami, sta samo AR (razmerje med površino pulpe in zoba) ter c (razmerje med širino pulpe in korenine na ravni srednjega korena) pokazali statistično značilno pomembne rezultate. Zato je bil kasneje uporabljen zgoraj opisani regresijski model v katerem kot napovedni spremenljivki nastopata zgolj AR in c.
- Pearsonovi korelacijski koeficienti med opazovanimi starostmi in morfološkimi napovednimi spremenljivkami pri moških, ženskah in v celotni populaciji vzorca so bili visoko statistično značilni. Torej obstaja linearna povezanost med starostjo in morfološkimi napovednimi spremenljivkami pri obeh spolih ter neglede na spol. Korelacijski koeficienti opazovane starosti so bili obratno korelirani z AR (razmerje med površino pulpe in zoba) in c (razmerje med širino pulpe in korenine na ravni srednjega

korena). Velja torej, da večja kot je vrednost AR oziroma c, manjša je starost, pri upoštevanju ostalih spremenljivk v modelu ($p < 0.001$).

- Kljub temu, da spol nima statistično pomembnega vpliva na oceno starosti, so bile razvite ločene formule s katerimi lahko čim bolje napovemo starosti moških in žensk. Zgoraj opisan model v katerem z multiplo linearno regresijo modeliramo odvisnost starosti glede na spol od napovednih spremenljivk AR in c nam da naslednje rezultate:
  Za osebe neznanega spola: Starost = 72,48 - 203,74 (AR) - 51,69 (c)
  Za moške: Starost = 73,33 - 209,97 (AR) - 54,06 (c)
  Za ženske: Starost = 72,12 - 179,12 (AR) - 68,66 (c)
- Razlika med opaženo in ocenjeno starostjo v skupni populaciji se je gibala od -2.2 do $+1.5$ let, pri čemer ni bilo statistično značilne razlike med opazovano in ocenjeno starostjo v katerikoli izmed starostnih skupin. (21-30 let - $p = 0.790$, 31-40 let - $p = 0.952$, 41-50 let - $p = 0.865$, 51-60 let - $p = 0.667$, skupno - $p = 0.921$)
- Na podlagi vrednosti ANOVA = 2.295 in vrednosti $p = 0.085$ je bilo zaključeno, da med opazovanimi in ocenjenimi starostmi znotraj skupin ni statistično značilnih razlik med skupinami.
- Razlika med opaženo in ocenjeno starostjo pri moških se je gibala od -0.9 do $+0.8$ let, pri ženskah pa od -1.0 do $+0.8$ let. Med opazovano in ocenjeno starostjo ni bilo statistično značilne razlike v nobeni izmed skupin (ženske - $p = 0.951$, moški - $p = 0.937$, skupno - $p = 0.921$).

Zaključek

Starosti so bile ocenjene z uporabo regresijskih enačb na podlagi razmerja med površino pulpe in zoba in razmerja med širino pulpe in korenine na ravni srednjega korena. Rezultati študije pa niso pokazali statistično značilne razlike med opaženimi in ocenjenimi starostmi. Z modelom je pojasnjeno 99.7% variabilnosti starosti, standardna napaka ocene regresije pa je znašala 0.60.

Rezultati nekaterih prejšnjih študij v kombinaciji z študijo so pokazali, da bi lahko rasni in kulturni dejavniki imeli pomembno vlogo pri ocenjevanju starosti, zato kljub temu, da so rezultati obetavni jih ni mogoče posplošiti na druge populacije. Študija je pokazala, da spol na oceno starosti upoštevajoč meritve zgornjih zob (podočnikov) nima statistično pomembnega vpliva, kar je podobno rezultatom prejšnjih študij.

Torej je iz študije mogoče sklepati, da obstaja linearno razmerje med širino pulpe in korenine na ravni srednjega korena in razmerje med površino pulpe in zoba, za podočnik na zgornji desni čeljusti, in starostjo med odraslo indijsko populacijo. Uporabljene metode ocenjevanja starosti ni mogoče uporabiti za večkoreninske zobe, saj je na njih težko izvesti točne meritve, poleg tega je pri neznanih osebah (rasna pripadnost) starost nemogoče oceniti z uporabo regresijskih enačb, ki temeljijo na omenjenih spremenljivkah.

Najina največja pomisleka sta se pojavila pri izbiri spremenljivk. AR in c sta se pokazala kot statistično značilna, vendar takrat ko so bile vključene vse spremenljivke. Seveda, če bi katero izmed izbranih spremenljivk odstranili, bi se lahko zgodilo da bi bile druge spremenljivke statistično značilne. Lahko bi se odločili za uporabo drugih kriterijev pri izbiri spremenljivk kot je npr. AIC. Poleg tega, ne piše če so preverjali morebitno kolinearnost med spremenljivkami, kar bi tudi lahko pokazalo težave v modelu. Manjšo težavo sva še videla da edina interakcija, ki so jo opravili je bila pri spolu, ne piše če so testirali tudi interakcije ostalih spremenljivk. Sicer predpostavljava da so to preverili, vendar v članku ni predstavljene analize ostankov, da bi se prepričali da so predpostavke linearnega modela izpolnjene.

Nadaljna raziskava bi morala biti usmerjena v analizo večjega vzorca in mora poleg starosti in spola vključevati tudi rasne in kulturne parametre.

Viri in literatura

ANOVA (Analiza Variance). n.d. https://www.statistik.si/anova-analiza-variance/.

Cohen's Kappa. n.d. https://en.wikipedia.org/wiki/Cohen%27s_kappa.

Glen, Stephanie. n.d. *Correlation Coefficient: Simple Definition, Formula, Easy Steps*. https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/.

Kenton, Will. 2019. *Analysis of Variance (Anova)*. https://www.investopedia.com/terms/a/anova.asp#:~:text=Analysis%20of%20variance%2C%20or%20ANOVA,the%20dependent%20and%20independent%20variables.

———. 2020. *Multiple Linear Regression (Mlr)*. https://www.investopedia.com/terms/m/mlr.asp.

Korelacijska Analiza. n.d. https://omr.fnm.um.si/wp-content/uploads/2018/04/6.pdf.

Oman, Špela. n.d. *Pearsonov Koeficient Korelacije*. https://www.benstat.si/blog/pearsonov-koeficient-korelacije/.

Saxena, Sudhanshu. 2011. "Age estimation of indian adults from orthopantomographs." *Brazilian Oral Research* 25 (June): 225–29. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-83242011000300006&nrm=iso.

Singh, Gurchetan. 2018. *A Simple Introduction to Anova (with Applications in Excel)*. https://www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance/.