

TALLINNA TEHNIKAÜLIKOOL  
Majandusteaduskond  
Majandusanalüüsi ja rahanduse instituut

Kelly Künnapas  
**SOOLINE JA RASSILINE PALGALÕHE AMEERIKA  
ÜHENDRIIKIDES**

Bakalaureusetöö  
Õppekava TAAB, peeriala majandusanalüüs

Juhendaja: Heili Hein, MA

Tallinn 2021

Deklareerin, et olen koostanud lõputöö iseseisvalt ja olen viidanud kõikidele töö koostamisel kasutatud teiste autorite töödele, olulistele seisukohtadele ja andmetele, ning ei ole esitanud sama tööd varasemalt ainepunktide saamiseks. Töö pikkuseks on 8053 sõna sissejuhatusest kuni kokkuvõtte lõpuni.

Kelly Künnapas .....

(allkiri, kuupäev)

Üliõpilase kood: 186028TAAB

Üliõpilase e-posti aadress: [kelly.kunnapas@gmail.com](mailto:kelly.kunnapas@gmail.com)

Juhendaja: Heili Hein:

Töö vastab kehtivatele nõuetele

.....

(allkiri, kuupäev)

Kaitsmiskomisjoni esimees:

Lubatud kaitsmisele

.....

(nimi, allkiri, kuupäev)

## SISUKORD

LÜHIKOKKUVÕTE.....	4
SISSEJUHATUS .....	5
1. TEOREETILISED KÄSITLUSED JA VARASEM KIRJANDUS.....	7
1.1. Soolise palgalõhe ajalugu ja teoreetilised käsitlused.....	7
1.2. Rassilise ja etnilise palgalõhe ajalugu ja teoreetilised käsitlused .....	10
1.3. Varasemad empiirilised uuringud .....	13
2. METOODIKA JA ANDMED .....	16
2.1. Andmed ja mudeli kirjeldus .....	16
2.2. Kirjeldav statistika.....	17
2.3. Mudel.....	20
3. EMPIIRILINE ANALÜÜS .....	22
3.1. Esialgse mudeli tulemuste analüüs .....	22
3.2. Modifitseeritud mudeli tulemuste analüüs .....	23
3.3. Valgenahaliste valimiga mudeli tulemuste analüüs.....	25
3.4. Afroameeriklaste valimiga mudeli tulemuste analüüs.....	27
3.5. Järeldused .....	28
KOKKUVÕTE .....	31
SUMMARY.....	33
KASUTATUD ALLIKATE LOETELU .....	35
LISAD .....	38
Lisa 1. Esialgne OLS mudel.....	38
Lisa 2. Esialgse OLS mudeli multikollineaarsuse aruanne .....	41
Lisa 3. Andmete korrelatsioonmaatriks.....	44
Lisa 4. Modifitseeritud OLS mudel.....	45
Lisa 5. Modifitseeritud OLS mudeli multikollineaarsuse aruanne.....	47
Lisa 6. Valgenahaliste OLS mudel.....	48
Lisa 7. Valgenahaliste OLS mudeli multiokollineaarsuse aruanne.....	50
Lisa 8. Afroameeriklaste OLS mudel.....	51
Lisa 9. Afroameeriklaste OLS mudeli multikollineaarsuse aruanne.....	53
Lisa 10. Lihtlitsents .....	54

## LÜHIKOKKUVÕTE

Käesolev bakalaureusetöö keskendub soolisele ning rassilisele palgalõhele Ameerika Ühendriikides. Töö eesmärgi täitmiseks on autor püstitanud järgmised uurimisküsimused:

- 1) Kuidas mõjutab sugu palka, kui muude tunnuste mõju palgale on samuti arvesse võetud?
- 2) Kuidas mõjutab rass palka, kui muude tunnuste mõju palgale on samuti arvesse võetud?
- 3) Kas sooline palgalõhe erineb eri rasside lõikes?

Uurimisküsimustele vastuste leidmiseks koostatakse regressioonanalüüs, kus sõltuvaks tunnuseks on tunnipalk, kuhu on sisse arvatud ka preemiad, ning sõltumatuteks tunnusteks on sugu, rass, vanus, haridustase, töötundide arv nädalas, töökoha sektor, lugemis- ning kirjutamisoskus töö ja elupaik. Regressioonanalüüsis kasutatakse PIAAC (*Programme for the International Assessment of Adult Competencies*) uuringu andmeid. Andmed on kvantitatiivsed ning pärinevad aastast 2017. Andmete puhastamine ning korrelatsioonanalüüs viiakse läbi programmis RStudio ning regressioonanalüüs programmis Gretl. Bakalaureusetöö jaguneb kolmeks peatükiks. Esimeses peatükis tutvustatakse soolise ja rassilise palgalõhe ajaloolist tausta ning teoreetilisi käsitlusi, ühtlasi tehakse ülevaade varasemast empiirilisest kirjandusest. Teises peatükis tutvustatakse meetodikat, andmeid ja mudelit. Töö viimases peatükis viiakse läbi regressioonanalüüs, kirjeldatakse seda ning tehakse järeldused.

Regressioonanalüüsi tulemuste alusel, erinevate kontrolltunnuste mõju arvesse võttes, kaasneb naiseks olemisega 16,4% madalam tunnipalk ning afroameeriklaseks olemisega seostub 11,3% madalam palk võrreldes valgenahalistega. Sooline palgalõhe erineb eri rasside lõikes. Valgenahaliste regressioonanalüüsi kohaselt (eri kontrolltunnuste mõjut arvesse võttes) teenivad naised 17,5% madalamat tunnipalka ning afroameeriklaste regressioonanalüüsi kohaselt teenivad naised (eri kontrolltunnuste mõju arvesse võttes) 14,7% madalamat tunnipalka. Seega sooline palgalõhe on kõrgem valgenahaliste seas.

Võtmesõnad: sooline palgalõhe, rassiline palgalõhe, sooline ebavõrdsus, rassiline ebavõrdsus

## SISSEJUHATUS

Sooline ebavõrdsus on olnud päevakajaline teema juba väga pikka aega nii meedias kui ka akadeemilises kirjanduses, ent tänapäevani ei ole suudetud soolisele ebavõrdsusele täielikku lahendust leida. Sooline ebavõrdsus seisneb ressursside ebavõrdsusest jaotusest meeste ja naiste vahel, käesolevas töös keskendutakse just sissetuleku ebavõrdsusele. Kui ajalooliselt on olnud perele raha teenimise roll pigem meeste kanda ning naise roll on olnud kodu ja pere eest hoolitsemine, siis tänapäeva maailmapildis on rollid võrdsustunud ning naiste osalus tööjõus märkimisväärselt kasvanud. Üha enam soovivad naised olla iseseisvad ja sõltumatud. Vaatamata sellele on püsima jäänud erinevused meeste ja naiste palkade vahel ning see mõjutab perekondi tervikuna. Lisaks soolisele ebavõrdsusele on tähelepanu keskmes olnud väga palju ka rassiline ebavõrdsus. Rassilise ebavõrdsusega on nii ajalooliselt kui ka tänapäeval silma paistnud Ameerika Ühendriigid. Rassismivastase liikumise „*Black Lives Matter*“ kulgu, mis protesteerib afroameeriklaste diskrimineerimise vastu, on korduvalt kajastatud ka Eesti meedias. Nii sooline kui ka rassiline palgalõhe võib olla tingitud mitmetest erinevatest aspektidest – näiteks töökogemus, hariduse tase, oskused, isikuomadused ja palju muud. Sooline ja rassiline palgalõhe võib paraku olla tingitud ka diskrimineerimisest – palganumber otsustatakse osaliselt töötaja soo või rassi alusel, võtmata arvesse objektiivseid tegureid töötaja oskuste ja võimekuse kohta.

Käesoleva bakalaureusetöö eesmärgiks on hinnata soolist ning rassilist palgalõhet Ameerika Ühendriikides. Eesmärgini jõudmiseks on autor püstitanud järgnevad uurimisküsimused:

- 1) Kuidas mõjutab sugu palka, kui muude tunnuste mõju palgale on samuti arvesse võetud?
- 2) Kuidas mõjutab rass palka, kui muude tunnuste mõju palgale on samuti arvesse võetud?
- 3) Kas sooline palgalõhe erineb eri rasside lõikes?

Töös kasutatakse PIAAC (*Programme for the International Assessment of Adult Competencies*) uuringu raames kogutud andmeid. Uuringu läbi viimiseks kasutatakse värskemaid kättesaadavaid andmeid, mis pärinevad aastast 2017. Kasutatavad andmed on kvantitatiivsed ning tegu on ristanndmetega. Andmete puhastamine ning töötlemine viiakse läbi programmiga RStudio.

Uurimisküsimustele vastuste leidmiseks viiakse läbi regressioonanalüüs kasutades harilikku vähimruutude meetodit. Regressioonanalüüs koostatakse programmiga Gretl.

Bakalaureusetöö jaguneb kolmeks peatükiks. Esimeses peatükis tutvustatakse lugejale soolise ja rassilise palgalõhe käsitlusi ning vähesel määral ka ajaloolist tausta. Samuti tuuakse välja ülevaade varasemalt tehtud empiirilisest kirjandusest. Teises peatükis selgitatakse lugejale töös kasutatavat meetodikat ning kirjeldatakse andmeid ja mudelit. Töö kolmandas peatükis viiakse läbi regressioonanalüüs ning selgitatakse selle tulemusi ja tehakse järeldused.

# 1. TEOREETILISED KÄSITLUSED JA VARASEM KIRJANDUS

Sissetulekute ebavõrdsus nii sugude kui erinevate rassiliste ja etniliste rühmade vahel on viimaste aastakümnete jooksul olnud püsivalt päevakajaline teema, ent täielikku lahendust ei ole suudetud sellele probleemile tänapäevani leida. Käesolevas peatükis keskendutakse nii soolist kui ka rassilist ja etnilist palgalõhet mõjutavatele teguritele ning tutvustatakse lugejale vähesel määral ka ajaloolist tausta. Ühtlasi tehakse ülevaade varasemalt tehtud empiirilisest kirjandusest.

## 1.1. Soolise palgalõhe ajalugu ja teoreetilised käsitlused

Palgalõhe olemuse peale mõeldes meenub tõenäoliselt esimesena sooline palgalõhe, sest seda on ühiskonnas väga palju kajastatud nii meedias kui ka akadeemilises kirjanduses. Kuigi soolist palgalõhet on aastakümneid intensiivselt uuritud, on ruumi jäetud ka uutele innovaatilistele uuringutele (Blau, Kahn 2016). Sooline ebavõrdsus seisneb resursside ebavõrdsusest jaotumisest meeste ja naiste vahel, käesolevas uurimistöös keskendutakse just palkade ebavõrdsusele. Keskmist soolist palgalõhet väljendatakse üldiselt naiste keskmise tunnipalga protsentuaalse suhtena meeste keskmisesse tunnipalka, sealjuures ei ole arvestatud naiste ja meeste erinevusi ametis, tegevusalas ning tööajas. (Anspal *et al.* 2009)

Tänapäeva maailmapildis, mis on üha enam võrdsustuv, on naistel tööjõuturul märkimisväärselt suurem roll kui seda on olnud ajalooliselt. Abielus naiste tööjõus osalemise määr tõusis järsult, olles 1890. aastal 4,6% ning 2001. aastal 61,4% - see on märkimisväärne tööturu trend USA ajaloos viimase sajandi jooksul (Polachek 2004). Goldin (2006) on nimetanud seda vaikseks revolutsiooniks. Ta selgitab seda majanduslikult läbi kolme faasi. Esimese faasi ajal oli täiskasvanud naiste osakaal tööjõus väga väike ning naiste tööjõupakkumise sissetulekuelastsus tugevalt negatiivne. Enamasti ümbrises naisi sotsiaalne stigma töö iseloomu tõttu – töö oli tihtipeale räpane, ohtlik ning ajamahukas. Esimeselt faasilt teisele faasile üleminek toimus erinevate tegurite koosmõjul. Üheks murdepunktiks oli infotehnoloogia areng 1900. aastate alguses, millega koos tõusis nõudlus kontoritöötajate järele. Samuti tõusis aastatel 1910-1930 keskkooli astunute ning lõpetanute arv – üleriigiliselt kasvas koolilõpetajate määr, olles 1910.

aastal 9% ning 1928. aastal 27%. See andis võimaluse noortel naistel enne abiellumist tööturule siseneda ning töötada lühema tööajaga ning soliidsematel töökohtadel. Kolmanda faasi alguseks, 1960. aastatel, oli naiste tööjõupakkumine pigem elastne ning kogunõudlus kasvanud. Üha enam kasvas täiskasvanud abielus naiste töös osalemise määr. Feminismi taaselustumine andis noortele naistele tõuke iseseisvalt toimetulekuks. Kaks olulist mõjutegurit naiste iseseisvumisel oli kontratseptiivide kasutuselevõtmine ning lahutuste kasv 1960. aastatel. Kontraseptiivide kasutuselevõtt aitas noortel naistel paremini oma karjääri planeerida, samal ajal pereplaneerimise tulevikku jätta. Selle tulemusena panustati rohkem hariduse omandamisele, enesearengusse ja eriala õpingutesse. Naised olid suurema silmaringiga ning prioritseerisid karjääri rohkem kui nende varasemad põlvkonnad. 1980. ja 1990. aastateks peeti varasemalt mittetraditsioonilisi ameteid naiste seas traditsioonilisteks. (*Ibid*)

Sellegipoolest on endiselt püsima jäänud erinevused palkade vahel. Hoolimata võrdseid võimalusi käsitlevate õigusaktide kehtestamisest ning feministlike liikumiste suurenemisest on USAs püsinud märkimisväärne palgaerinevus aastakümneid – 1955. aastatel oli see 31%, 1970. aastate alguseks juba 35-37% ning langes siis 1982. aastaks 33%-ni (O'Neill 1985).

Ajalooliselt on olnud naistel suuresti kodu- ja perehoidja roll ning mehe roll on olnud sissetuleku teenimine. Tänapäeval on rollid üha enam võrdsustunud, kuid sellegipoolest seisavad naised situatsiooni ees, kus nad peavad tegema valikuid töö ja pere loomise vahel. 20. sajandi teises pooles, kui naiste tööjõus osalemise määr tõusis märgatavalt, esines muutusi ka sündimuses. Kui viljakuse määr oli 1960. aastal 2,48, siis 2000. aastaks oli see langenud 2,13ni. Samuti esines muutusi ka esmaste sünnitajate seas – 1970. aastal oli 19 protsenti esmasünnitajatest 25 aastased või vanemad, 2000. aastaks oli see protsent tõusnud 50ni. (Amuedo-Dorantes, Kimmel 2003) Vaatamata esmasünnitajate tõusvale vanusele on naisel elus etapp, kus luuakse pere. Rasedus, sünnitus ja sünnitusjärgne periood vähendab ajutiselt tööjõupakkumist ning produktiivsust – töölt eemal olek mõjutab ka ametiga seotud oskusi, mis hiljem võivad viia madalama palgani. Isegi kui töölt eemalolek oskusi ei mõjuta, siis sellegipoolest vähendab ajutine töölt eemalolek tööstaaži ning kogemust, ühtlasi kaovad mõneks ajaks võimalused tööalaseks arenguks ja ametikõrgenduseks. Naistel on võimalus töötada peale lapse saamist ka osalise tööajaga, ent sellisel juhul saadakse ka madalamat töötasu. (Loughran, Zissimopoulos 2007)

Mõistagi mängivad inimeste kognitiivsed oskused ning vaimne võimekus olulist rolli tööturul edu saavutamisel. Sama olulist rolli mängivad ka mittekognitiivsed omadused. Suure viisiku mudeli



kohaselt on nendeks omadusteks ekstravertsus, meelekindlus, koostöövalmidus, vaimne stabiilsus ja avatus kogemustele. Mudel aitab mõista, mil määral on sooline palgalõhe ja vaimne võimekus tingitud meheline ja naiseline isikuomaduste erinevustest, sest teadaolevalt on edu tööturul suures osas seotud iseloomu ja käitumisega. Näiteks on mehed altimad pidama ülemusega palgaläbirääkimisi ning on valmis küsima kõrgemat palka kui naised. (Mueller, Plug 2006) Dittrich *et al.* (2014) leiavad oma uurimuses, et meeste ja naiste palgaläbirääkimiste erinevus ei sõltu tingimata ainult läbirääkimisoskusest, vaid ka käitumisest. Nad leiavad, et kui naised läheneksid läbirääkimisi pidades enesekindlamalt ja nõudlikumalt ning ei nõustuks esimese pakkumisega, oleksid palgaläbirääkimised tulemuslikumad ning see aitaks omakorda vähendada soolist palgalõhet. Samuti on isikuomadused tihedalt seotud ka ametivalikuga – Cable ja Judge (1997) on oma uuringus leidnud, et empaatilisus on otseses positiivses seoses eelistusega töötada organisatsioonis, mille organisatsioonikultuuri märksõnadeks on toetus ja kokkuhoidvus ning negatiivne seos leiti eelistusega töötada agressiivses ja rangelt eesmärkidele orienteeritud keskkonnas. Samuti võivad naised suurema riskitundlikkuse tõttu eelistada vähem riskantseid ameteid, mille tulemusena teenitakse tihtipeale ka madalamat töötasu.

Üks osa soolisest palgalõhest võib olla põhjustatud ka diskrimineerimisest. Sooline diskrimineerimine tähendab, et sama haridusliku tausta, kogemuse ning isikuomadustega naisele ja mehele makstakse erinevat palka. (Anspal *et al.* 2009) Inimkapitali teooria kohaselt on palk seotud töötaja tootlikkusega. Teooria kohaselt peaksid meeste ja naiste vahelised palga erinevused olema tingitud tootlikkuse erinevusest, mitte aga diskrimineerimisest. (Gardeazabal, Ugidos 2003) Sellegipoolest esineb situatsioone, kus värbamine võrdsete kogemuste korral võib sõltuda soost või rassist ning sama tootlikkusega naised ja mehed saavad erinevat palka. (Anspal *et al.* 2009) Sooline diskrimineerimine võib alata juba enne tööturule sisenemist, sest õppeainete valik koolis võib suuresti määrata tuleviku ametivaliku ning selle läbi ka palga. Lahterdamise ja eelarvamuste tõttu ei pruugi naine valida maskuliinsema iseloomuga ameteid, mistõttu võidakse valida stereotüüpsem amet näiteks teenindussektoris, kus palgatase on madalam. (Mueller, Plug 2006) Tööturul jaguneb diskrimineerimine majandusteooria järgi kaheks: eelistustepõhine diskrimineerimine ja statistiline diskrimineerimine. Eelistustepõhine diskrimineerimine tähendab, et tööandja eelistab värbamisel töötajat sөө, rahvuse või rassi alusel. Statistiline diskrimineerimine tähendab, et tööandja lähtub kandideerijat iseloomustavatest keskmistest tunnustest – tööandja eeldab, et inimesel on kindlad soole omaseks peetavad tunnused ning otsus tehakse nende järlduste põhjal. Eelistustepõhist diskrimineerimist eristab statistilisest diskrimineerimisest asjaolu, et otsus

tehakse pelgalt eelistuse põhjal, viimase puhul aga püüab tööandja kandideerija omadusi ennustada grupi keskmistele omadustele toetudes. (Anspal *et al.* 2009)

## 1.2. Rassilise ja etnilise palgalõhe ajalugu ja teoreetilised käsitlused

Peale teist maailmasõda on Ameerika Ühendriikides rassilise ebavõrdsuse suhtes toimunud palju muutusi ent ebavõrdsus on sellegipoolest püsima jäänud tänase päevani. Sõjajärgne kiire majanduskasv tõi kaasa ka muutusi rassilises ebavõrdsuses, kuid 1960. aastatel rauges majanduskasvuga koos ka edu rassilise võrdõiguslikkuse suunas. (Eschbach, Waters 1995) 1980. aastast hakkas uuesti tööalane segregatsioon vähenema, mis võimaldas afroameeriklastel siseneda sektorisse, kus tavapäraselt olid domineerinud valgenahalised. Vaatamata tööalase segregatsiooni vähenemisele püsis sissetulekute erinevus. (Grodsky, Pager 2001) Suur roll rassilise ebavõrdsuse süvenemisel on olnud ka immigratsioonil. 1990. aastaks oli 7,9% ehk 19,8 miljonit inimest USA elanikkonnast sündinud välismaal. (Eschbach, Waters 1995). Tänapäeval ei ole enam demograafiliste muutuste põhiliseks aspektiks sisseränne – 2008. aastal sõlmiti 14,6% abieludest erineva rassilise taustaga inimeste vahel, vähemusrahvuste sünd on tõusutrendis ning 2014. aastast on vähemusrahvustest lapsed moodustanud suurema osa avalike koolide lastest. Ameerika Ühendriikide ühiskond on majanduslike ja demograafiliste muutuste keskel – 2043. aastaks ennustatakse, et vähemusrahvused moodustuvad suurema osa USA elanikkonnast. Ameerika Ühendriikide ühiskonnas etnilise koosseisu muutudes kerkib üha enam esile sissetulekute erinevuse probleem. (Noguera 2017)

Diskrimineerimine on oluline tegur rassiliste palgaerinevuste selgitamisel. 1940. aastal teenisid tumedanahalised naised vaid 50% valgenahaliste naiste palgast, kusjuures võrreldavad grupid olid samas vanuses ning sama haridustasemega. Alles 1980. aastaks olid palgad võrdsed. Paraku hakkas palgalõhe peale 1980. aastat uuesti suurenema. (Anderson, Shapiro 1996) 1964. aastal võeti vastu kodanikuõiguse seadus, mille kohaselt keelati tööandjal diskrimineerimine nahavärvuse, usutunnistuse, soo või rahvusliku päritolu alusel. (Coleman 2003) Selle tulemusena vähenes afroameeriklasest meeste keskmise nädalapalga erinevus oluliselt, olles 1967. aastal 50% ning 1974. aastal 30%. (Couch, Daly 2002) Sellegipoolest kohtab diskrimineerimist endiselt. Jencksi ja Petersoni (1991) kirjutatud raamatus on leitud, et väga paljud tööandjad diskrimineerivad ning omavad eelarvamusi noorte tumedanahaliste meeste suhtes. Wilson (1996) toob oma raamatus välja, et tervelt 74% küsitatud tööandjatest väljendas negatiivset suhtumist tumedanahaliste

suhtes. Tihtipeale tehakse tööandja poolt otsus pelgalt eelarvamuse põhjal – Darity Jr. ja Mason (1998) on oma uuringus leidnud, et afroameeriklastele omaste nimedega tööle kandideerimise avaldused saavad tihedamini eitava vastuse. Donohue III ja Siegelmani (1991) tehtud uuringus selgus, et ainult 1% töötajatest, kes on kogenud tööalast diskrimineerimist, on pöördunud õiguskaitseorganite poole. Põhjuseks suured kulutused seoses kohtusse kaebamisega. Mõned teadlased väidavad, et diskrimineerimine ei ole peamine rassilise palgalõhe põhjus, vaid selleks on oskuste puudujääk (Coleman 2003). 1980. aastatel hakati väärtustama üha enam oskustega töölisi. Lundberg ja Startz (1983) tehtud uuringust tuleb välja, et tumedanahalised võivad oma oskustesse vähem panustada, sest nende tasuvus oskuste omandamisel on madalam kui valgenahalistel. Ühtlasi leiavad nad, et rassiline palgalõhe esineb, sest tumedanahalised sisenevad tööturule vähemate oskustega, eeldusel, et oskuste omandamisest saadav tulu on madal ning see ei tasu ära (*Ibid*).

Ajalooliselt on muutusi rassilises erinevuses mõõdetud ka erinevate saavutuslike testidega. Üheks neist on *Armed Forces Qualifications Test (AFQT)*, mis võeti kasutusele esimese maailmasõja ajal. Sõltumatud uuringud on näidanud, et *AFQT* test on rassiliselt erapooletu ning see aitab ennustada tegelikku töö tulemuslikkust (Johnson, Neal 1996). Seda kasutatakse, et mõõta sobivust ajateenistusse astumisel. Test hõlmas nii verbaalseid- kui ka loogika- ja matemaatikaoskusi, ühtlasi peegeldas see hariduse kvaliteeti ning kodust kasvatust. Uuringud on näidanud, et testi tulemused tumedanahaliste ja valgenahaliste vahel erinevad silmnähtavalt – tumedanahalised saavutavad madalama taseme olles samas klassis. Ühtlasi on leitud tugev korrelatsioon testi tulemuste ja sissetulekute vahel. (O’Neill 1990)

Üheks suurimaks teguriks palga määramisel on haridus, sest haridus annab eelise saada tööd kõrgema palgaga ametikohal. Modernses ühiskonnas on haridus avalik hüvis, mida riik peaks inimestele tagama. Paraku on haridus aga mõnele ühiskonnagrupile rohkem kättesaadav kui teistele. Kõige äärmuslikum oli rassiline ebavõrdsus hariduses Ameerika Ühendriikides 19. sajandil kodusõja alguses, mida nimetatakse ka antebellumi ajastuks. Afroameeriklastele keelati kooliharidus ning neile lugemise ja kirjutamise õpetamine muutus ebaseaduslikuks. Alles kodusõja lõppedes avanes afroameeriklastel võimalus uuesti haridust omandada, ühtlasi said afroameeriklased kodusõja lõppedes ka valimisõiguse, mille tulemusena oli neil võimalik oma poliitilist võimu hariduse teemadel rakendada. (Walters 2001) Selle tulemusena oli afroameeriklastest laste kooli astunute osakaal aastal 1870 10% ning 1880. aastaks oli osakaal 34% (O’Neill 1990). 19. sajandi lõpus ja 20. sajandi alguses rakendasid valgenahalised eliitgrupid

jõuliselt oma poliitilise huve, mis hõlmas ressurside ebaühtlast jaotamist afroameeriklaste ja valgenahaliste koolide vahel. 20. sajandi kahel esimesel aastakümnel avati Ameerika Ühendriikide lõunapoolsetes osariikides arvukalt avalikke koole valgenahalistele ent afroameeriklastele vaid mõned üksikud. Peale varasemalt mainitud kodanikuõiguse seaduse vastuvõtmist kadus valgetenahalistel eliitgruppidel võim rakendada avalikus sektoris oma poliitilisi huve ning neil oli võimalus saata lapsed enda ressursidega erakoolidesse. (Walters 2001)

Vahemikul 1916 kuni 1970 toimus Ameerika Ühendriikides suur migratsioon, kus afroameeriklased liikusid lõunast üle riigi. Selle tulemusena hakkas alates 1940. aastast üha enam haritud tumedanahalisi tööturule sisenema, mis aitas kaasa rassilise lõhe vähenemisele. Kuigi tumedanahaliste haridustase ning võimalused olid kasvanud, püsisid sellegipoolest erinevused hariduslike saavutuste vahel. Erinevus võib olla põhjustatud koolide taseme, kui ka perekondliku tausta erinevusest. (O'Neill 1990) Paljudes Ameerika Ühendriikide linnaosades, kus elavad valdavalt vähemused, kasvavad lapsed üles vaesuse ja vägivallaga. Ühtlasi on nendes piirkondades koolide tase väga madal. Dallases läbiviidud uuringu kohaselt on 143000 lapsest suurem osa vaesest perekonnast, 85% neist on afroameeriklased või mehhiklased. Suurel osal neist on lugemisoskus alla keskmise. Enamikul püsib selline tase keskkoolini, mistõttu on neil raske õppekavaga toime tulla. See ühtlasi tõstab koolist välja kukkunute määra ning vähendab ka ülikooli astunute arvu. (Farkas, Vicknair 1996) Aastal 1988 oli ülikooli lõpetanute osakaal 25-34 aastaste afroameeriklasest meeste seas 15,5%, valgenahaliste meeste ülikooli lõpetanute osakaal aga 27,9% (O'Neill 1990).

Alates 2000. aastast kuni 2017. aastani on Ladina-Ameerika päritolu (*hispanic*) inimeste arv tõusnud 66,9%. Tänapäeval moodustavad Ladina-Ameerika päritolu elanikud umbkaudu 18% Ameerika Ühendriikide elanikest. Seega ka nende seas on välja joonistunud oluline ebavõrdsus tööturul. Ladina-Ameerika päritolu naiste palgalõhe on oluliselt suurem kui Ladina-Ameerika päritolu meeste palgalõhe võrreldes valgenahaliste meestega. Ladina-Ameerika päritolu naiste osakaal tööturul on Ameerika Ühendriikides üha kasvamas, kuid palgalõhet pole suudetud oluliselt vähendada alates 2000. aastast, mistõttu on see probleem aina enam süvenev. (Davila, Mora 2018)

### 1.3. Varasemad empiirilised uuringud

Soolist ja etnilist palgalõhet on eelnevalt teiste autorite poolt uuritud korduvalt. O'Neill (1985) on oma uuringus leidnud, et Ameerika Ühendriikides on palgalõhe olnud aastate lõikes muutuses, kuid mitte märkimisväärselt. Kui 1955. aastal oli palgalõhe 31%, siis 1970. aastaks oli see kasvanud 37%-ni ning 1982. aastal taaskord kahanenud 33%-ni. Ta leiab, et üheks põhjuseks palgalõhe kõikumise puhul on muutused oskustes, mis on omandatud töökogemuse ning haridusega. 1950. aastal olid tööjõus osalevad naised omandanud haridust 1,6 aastat rohkem kui mehed, kuid 1979. aastaks oli see erinevus kadunud. Autor leiab, et erinevuse kadumine üksinda oleks palgalõhet suurendanud 7 protsendipunkti võrra. Kui lisada haridustee pikkuse erinevusele veel töökogemuse erinevus 1952. ja 1960. aastate vahel, oleks see omakorda suurendanud palgalõhet veel 2 protsendipunkti võrra. Nende järelduste kohaselt oleks pidanud palgalõhe olema sellel perioodil suurem, kui ta tegelikult oli. Autor eeldab, et selle põhjuseks on diskrimineerimise vähenemine antud perioodil. Üheks palgalõhet mõjutavaks teguriks on ka töötus, mis 1970. aastatel oli väga kõrge. Ühtlasi mõjutab palgalõhet negatiivselt ka vähema kogemusega tööjõu palkade langus. Autor toob välja ka palgalõhe erinevuse rassi järgi. Ta leiab, et 1955. aastast hakkas afroameeriklasest naiste ja meeste palgalõhe järsult vähenema, ta ei seosta seda haridusega, vaid suurenenud abielulahutuste määraga. Selle tulemusena võis suureneda afroameeriklasest naiste vajadus otsida kõrgemapalgalist tööd. (*Ibid*)

Fields ja Wolff (1995) on uurinud tööstusharude vahelist soolist palgalõhet Ameerika Ühendriikides, kasutades 1988. aasta andmeid. Nende mudelis on sõltuvaks muutujaks tunnipalk ning sõltumatuteks muutujateks haridus, töökogemus aastates, residentsus, regioon, perekonnasseis, rass, elukutse ning tööstusharu. Oma uuringu tulemusena leidsid nad, et tööstusharude vaheline sooline palgalõhe on märkimisväärne. Nad toovad välja, et üheks põhjuseks tööstusharude vahelise soolise palgalõhe tekkeks on asjaolu, et ettevõtted peavad mehi rohkem püsivaks tööjõuks kui naisi, mistõttu investeeritakse nende koolitamisse rohkem ning makstakse neile kõrgemat preemiat. (*Ibid*)

Varasemad uuringud on näidanud, et rassiline palgalõhe on meeste seas suurem kui naiste seas. 1990. ja 1991. aasta andmete põhjal tehtud uuringust selgub, et afroameeriklasest mehed teenivad valgenahalistest meestest tervelt 24,4% vähem palka, samal ajal kui afroameeriklaste ja valgenahaliste naiste palgaerinevus on 18,5% (Johnson, Neal 1996) Bayard *et al.* (1998) on oma uurimuses sellele küsimusele põhjuseid otsinud. Autorid on omad uuringus kasutanud samuti

1990. aasta andmeid. Autor on leidnud oma kirjeldava statistika tabelis tõepoolest, et keskmine tunnipalga erinevus on meeste seas suurem kui naiste seas. Ta toob välja, et üheks palgalõhe põhjuseks võiks olla asjaolu, et valgenahalised on statistiliselt rohkem abielus, saavad vähem lapsi ning omavad kõrgemat haridust. Kõik need indikaatorid on seotud kõrgema palgaga. Ühtlasi tuleb välja, et haridustaseme erinevus on meeste seas suurem kui naiste seas, mis selgitab ka asjaolu, et palgalõhe on meeste seas suurem võrreldes naistega. Regressioonanalüüsis selgus, et afroameeriklaste ning valgenahaliste naiste palgaerinevus on 2%, samasugune erinevus leiti ka valgenahaliste naiste ja Ladina-Ameerika päritolu naiste vahel. Afroameeriklaste ja valgenahaliste meeste palgaerinevus tuli aga märkimisväärtelt suurem – 12%. Valgenahaliste meeste ja Ladina-Ameerika päritolu meeste vaheline palgaerinevus on analüüsi tulemusel 7%. (*Ibid*)

Coleman (2003) on oma artiklis uurinud afroameeriklasest meeste sissetuleku ebavõrdsuse seost diskrimineerimise ja tööalaste oskustega. Autor kasutab oma mudelis andmeid uuringust, mis on läbi viidud *MCSUI (Multi-City Study of Urban Inequality)* teadlaste poolt. 1992.-1994. aastatel läbi viidud uuringus osales 3000 tööandjat, kellest 720 küsitletu vastused kaasati uuringusse. Tööandjatele esitati küsimusi viimaste värvatud töötajate kohta ametikohale, mis ei nõudnud kõrgharidust. Autor on oma mudelisse lisaks rassi komponendile arvestanud ka oskuste, inimkapitali ning tööturu mõju palgale ning püstitanud hüpoteesi, et kui diskrimineerimine on rassilise palgalõhe peamiseks põhjuseks, siis peaks rassi komponent olema mudelis statistiliselt oluline, isegi kui eelnevalt mainitud teiste komponentide mõju palgale on mudelis arvesse võetud. Kui aga rassi komponent mudelis osutub statistiliselt mitteoluliseks, võib rassilise palgalõhe põhjuseks olla vähemusrasside madalam haridus- või oskustetase. Autor leidis, et kui mudelile ainult rass arvesse võtta, teenivad afroameeriklased tervelt 19% vähem sissetulekut. Kui mudelile lisada inimkapitali komponent, teenivad afroameeriklased 11% vähem sissetulekut. Kui autor lisas mudelisse oskuste komponendi, sai ta taaskord tulemuseks, et afroameeriklased teenivad 11% vähem sissetulekut. See tähendab, et sama inimkapitali ning oskuste taseme juures teenivad afroameeriklased endiselt vähem palka. Autor järeldab sellest, et palkade erinevus on siiski põhjustatud diskrimineerimisest. (*Ibid*) Oskuste seost rassilise palgalõhega on uurinud ka Petre (2019). Ta leidis, et mittekognitiivsed oskused selgitavad 2-4% rassilisest palgalõhest afroameeriklaste ja valgenahaliste vahel ning kognitiivsed oskused selgitavad 8-70% rassilisest palgalõhest afroameeriklaste ja valgenahaliste vahel. Ta järeldab, et rassilise palgalõhe püsimisele on kaks selgitust – kas vähemusrassid saavadki sama võimekuse juures väiksemat palka diskrimineerimise tõttu või vähemusrassid pakuvadki tööturul väiksemat võimekust ja oskusi, mistõttu saavad nad madalamat töötasu. (*Ibid*)

Avery *et al.* (2019) on uurinud asjaolu, et tumedanahalised töötajad peavad vähem palgaläbirääkimisi kui nende valgenahalised kolleegid ning palgaläbirääkimiste tulemusena saavad madalama palga kui valgenahalised töötajad, eriti kui tööandja on rassiliste eelarvamustega. Nad uurisid rassipõhiseid stereotüüpe ning nende mõju läbirääkimistele. Nende hüpotees, et ülemuse rassilised eelarvamusel mõjutavad läbirääkimiste tulemusi, leidis kinnitust. Ühtlasi leidis kinnitust ka hüpotees, et ülemuse äripartneri rassilised eelarvamusel mõjutavad läbirääkimiste tulemusi. Kolmas hüpotees leidis samuti kinnitust, milles ennustati, et tumedanahalised töötajad, kes kipuvad rohkem palgaläbirääkimisi pidama, saavad tulemusena madalama palga kui need, kes peavad palgaläbirääkimisi vähem. (*Ibid*)

Vaatamata ajaloolistele edusammudele on tänapäeval paraku rassiline palgalõhe samal tasemel või ühtlasi isegi suurenenud kui see oli 1979. aastatel. Wilson ja Rodgers III (2016) on oma aruandes võrrelnud valgenahaliste meeste keskmist tunnipalka tumedanahaliste meeste ning tumedanahaliste naiste keskmise tunnipalgaga - 2015. aastal teenisid afroameeriklasest mehed 22% vähem tunnipalka ning afroameeriklasest naised 34,2% vähem tunnipalka, kusjuures hariduse ning kogemuse tase oli võrreldavatel gruppidel sama. Afroameeriklasest naiste ning valgenahaliste naiste tunnipalga lõhe oli sealjuures 11,7%. Ühtlasi teenivad afroameeriklasest naised 12,2% väiksemat tunnipalka kui afroameeriklasest mehed, mis toob välja ka märkimisväärse soolise palgalõhe. Autorid on välja toonud ka asjaolu, et olulisemad faktorid üha enam suureneva rassilise palgalõhe juures ei ole madal haridustase või vähene kogemus, vaid suurimaks põhjuseks on siiski diskrimineerimine. (*Ibid*)

## 2. METOODIKA JA ANDMED

Bakalaureuse töö teises peatükis tutvustab autor töö empiirilises osas kasutatavaid andmeid ning ökonomeetrilisi meetodeid. Lisaks tehakse andmetest kirjeldav statistika. Viimaks kirjeldatakse bakalaureusetöö empiirilises osas kasutatavat mudelit.

### 2.1. Andmed ja mudeli kirjeldus

Käesoleva töö eesmärgini jõudmiseks on autor püstitanud järgmised uurimisküsimused:

- 1) Kuidas mõjutab sugu palka, kui muude tunnuste mõju palgale on samuti arvesse võetud?
- 2) Kuidas mõjutab rass palka, kui muude tunnuste mõju palgale on samuti arvesse võetud?
- 3) Kas sooline palgalõhe erineb eri rasside lõikes?

Eelmainitud uurimisküsimuste vastuseni jõudmiseks kasutatakse PIAAC (*Programme for the International Assessment of Adult Competencies*) uuringu raames kogutud andmeid. Bakalaureusetöös kasutatavad kvantitatiivsed andmed pärinevad aastast 2017. Tegu on ristanndmetega, mis tähendab erinevate objektide võrdlemist samal ajamomendil (Sauga 2021).

Esimeseks koostatava mudeli sõltuvaks muutujaks on palk, kuhu on sisse arvatud preemiad. Mudeli sõltumatuteks tunnusteks on rass, sugu, vanus, laste olemasolu, lugemisoskus tööl, kirjutamisoskus tööl, amet, tegevusala, sektor, hariduse tase, tervislik seisund, töötunnid nädalas, elupaik, regioon, kas vastaja on USA-s sündinud, inglise keele kirjaoskus ning inglise keele lugemisoskus. Rass on andmetes jaotatud nelja gruppi: valgenahalised, Ladina-Ameerika päritolu (*hispanic*), afroameeriklased ning muud rassid. Ametid on jaotatud järgmiselt: kaitsevägi, õigus ja kõrgemad ametnikud, spetsialistid, tehnikud, ametnikud, teenindustöötajad, põllumajandus- ja kalandustöötajad, käsitöö, tehase- ja masinaoperaatorid ning lihttöölised. Tegevusalad on jaotatud järgmiselt: põllumajandus, kaevandus, tootmine, elektri- ja gaasitootmine, kanalisatsioon ja jäätmekäitlus, ehitus, hulgi- ja jaekaubandus ning mootorsõidukite remont, transport ja ladustamine, majutus- ja toitlustusteenused, informatsiooni- ja kommunikatsioonitehnoloogia,



finants- ja kindlustustegevus, kinnisvara, kutse- ja teadustegevus, haldus- ja tugiteenused, avalik haldus, haridus, sotsiaaltöö, kunst ja meelelahutus, teenused. Sektor on jaotatud kolmeks: erasektor, avalik sektor ning mittetulundusühing. Haridustaset eristatakse järgmiselt: põhiharidus või madalam, keskkharidus, kutseharidus, rakenduskõrgharidus, kõrgharidus – bakalaureusekraad, kõrgharidus – magistrikraad või doktorikraad. Tervislikku seisundit on vastajad saanud hinnata kas suurepäraseks, väga heaks, heaks, rahuldavaks või halvaks. Elupaik on jaotatud neljaks: suurlinn, äärelinn, väikelinn, maakoht. Regiooniks saavad vastajad valida kas riigi kirde-, kesk- lääne-, lõuna- või põhjaosa. Inglise keele kirja- ning lugemisoskuse saavad vastajad hinnata kas väga heaks, heaks, halvaks või puudulikuks. Laste olemasolul arvestatakse, kas lapsi on või ei ole.

Palga ja sõltumatute pidevate ja järjestatud kategooriatunnuste vahelist statistilist seost kontrollitakse korrelatsioonanalüüsiga, mis viiakse läbi statistikatarkvaraga RStudio. Palga ja sõltumatute tunnuste vahelise täpsema seose uurimiseks tehakse regressioonanalüüs kasutades harilikku vähimruutude meetodit statistikaprogrammis Gretl. Enne mudeli tulemuste tõlgendamist kontrollitakse, kas vajalikud eeldused on täidetud: heteroskedastiivsuse puudumine, multikollineaarsuse puudumine, mudeli kuju õigsus ning jääkliikmete alluvus normaaljaotusele. Sõltumatute tunnuste hulgas esineb mitmeid kategooriatunnuseid ning nende mudelisse lisamiseks kasutatakse fiktiivseid tunnuseid. Kategooriatunnuste hulka kuuluvad rass, tegevusala, amet, sektor, hariduse tase, tervislik seisund, elupaik, regioon, inglise keele kirjaoskus ning inglise keele lugemisoskus. Fiktiivsete tunnuste loomisel luuakse igale kategooriatunnusele baastunnus, millega teisi tunnuseid võrreldakse. Rassi baastunnuseks on valgenahalised, tegevusala baastunnuseks põllumajandus, ameti baastunnuseks kaitsevägi, sektori baastunnuseks erasektor, hariduse taseme baastunnuseks põhiharidus või madalam, tervisliku seisundi baastunnuseks vastus „suurepärase“, elupaiga baastunnuseks suurlinn, regiooni baastunnuseks riigi kirdeosa, inglise keele kirjaoskuse ning inglise keele lugemisoskuse baastunnuseks vastus „väga hea“. Vanus, lugemisoskus tööl, kirjutamisoskus tööl ning töötunnid nädalas lisatakse mudelisse pidevate tunnustena.

## **2.2. Kirjeldav statistika**

Andmete töötlemine ning puhastamine viidi läbi statistikatarkvaraga RStudio. Esialgses andmete valimis oli vaatlusi 3660. Osad uuringus osalenud inimestest ei käinud palgatööl (näiteks õppurid, koduperenaised, pensionärid, töötud) või olid enda tunnipalga, rassi või soo, mis on käesoleva töö

jaoks ühed olulisimad tunnused, avaldamata jätnud, mistõttu tuli need vaatlused andmetest eemaldada. Andmete puhastamise tulemusena eemaldati 1659 vaatlust. Seejärel kontrolliti andmetes ka erindite olemasolu. Erindid tunnipalga näol tähendavad kas väga suuri või väga väikeseid väärtusi. Erindite eemaldamise tulemusena eemaldati andmetest 1 ülemine ja 1 alumine protsentiil ehk kokku 2%. Erindite tõttu eemaldati valimist 42 vaatlust. Andmete puhastamise tulemusena jäi valimisse 1959 vaatlust. Analüüsis kasutatavad lõplikud puhastatud andmed on avalikult nähtava failina lingi olemasolul välja toodud kasutatud kirjanduse loetelus (Analüüsis kasutatavad andmed 2021)

Tabelis 1 on välja toodud töö valimit kirjeldav statistika tunnipalga ning nädalas töötatud tundide kohta. Kirjeldav statistika hõlmab tunnipalga ning nädalas töötatud tundide keskmist, standardhälvet ning maksimum- ja miinimumväärtust. Tunnipalk on esitatud dollarites.

Tabel 1. Valimit kirjeldav statistika, tunnipalk ja nädalas töötatud tunnid

	Tunnipalk (\$)	Nädalas töötatud tunnid
Keskmine	23,184	39,710
Miinimumväärtus	5,828	3,000
Maksimumväärtus	106,923	112,000
Standardhälve	15,560	12,327

Allikas: PIAAC (2021), autori arvutused

Tabelist 1 on näha, et käesoleva bakalaureusetöö valimi keskmiseks tunnipalgaks on 23,184 dollarit, miinimumväärtus ja maksimumväärtus vastavalt 5,828 dollarit ja 106,923 dollarit ning standardhälve 15,560. Nädalas töötatud tunde on keskmiselt 39,710. Minimaalselt tehakse tööd 3 tundi nädalas ning maksimaalselt 112 tundi nädalas. Nädalas töötatud tundide standardhälve on 12,327.

Tabelis 2 on välja toodud käesoleva bakalaureusetöö valimi keskmised tunnipalgad vastavalt soole ning rassile.

Tabel 2. Valimit kirjeldav statistika, keskmine tunnipalk vastavalt soole ning rassile

Sugu	Rass	Keskmine tunnipalk (\$)	N
Mees	valgenahaline	27,22	630
Mees	Ladina-Ameerika päritolu	22,84	106
Mees	afroameeriklane	19,13	126
Mees	teistest rassidest	28,53	79

Sugu	Rass	Keskmine tunnipalk (\$)	N
Naine	valgenahaline	21,68	647
Naine	Ladina-Ameerika päritolu	19,07	142
Naine	afroameeriklane	17,34	140
Naine	teistest rassidest	22,73	89

Allikas: PIAAC (2021), autori arvutused

Tabeli 2 veerust „N“ on näha, kui suure osa valimist iga kategooriagrupp moodustab. Valimis on 630 valgenahalist meest, 106 Ladina-Ameerika päritolu meest, 126 afroameeriklasest meest ning 79 meest on teistest rassidest. Valgenahalisi naisi on valimis 647, Ladina-Ameerika päritolu naisi 142, afroameeriklasest naisi 140 ning teistest rassidest on 89 naist. Tabelist 2 on näha, et kõrgeim keskmine tunnipalk on teistest rassidest meestel ning madalaim tunnipalk afroameeriklasest naistel – vastavalt 28,53 dollarit ning 17,34 dollarit. Võrreldes meeste keskmist tunnipalka rasside lõikes, on näha, et kõrgeimat tunnipalka teenivad teistest rassidest mehed – 28,53 dollarit ning madalaimat tunnipalka afroameeriklasest mehed – 19,13 dollarit. Naiste hulgas on rasside lõikes kõrgeim tunnipalk samuti teistest rassidest naistel ning madalaim tunnipalk afroameeriklasest naistel – vastavalt 22,73 dollarit ning 17,34 dollarit. Käesolevast tabelist on näha ka soolist palgalõhet – kõikide rasside lõikes teenivad naised vähem palka kui mehed. Sooline palgalõhe on suurim ehk 25,52% teistest rassidest meeste ja naiste vahel, kus mehed teenivad keskmiselt 28,53 dollarit tunnis ning naised 22,73 dollarit tunnis. Väikseim sooline palgalõhe ehk 10,32% esineb afroameeriklasest meeste ja naiste vahel. Rassilist palgalõhet sugude lõikes võrreldes esineb suurem palgalõhe meeste seas.

Andmeanalüüsi ühe osana viidi läbi korrelatsioonanalüüs, et uurida muutujatevahelisi seoseid. Korrelatsioonanalüüsi läbi viimiseks kasutati Spearmani korrelatsiooni, sest see sobib nii pidevate kui ka järjestatud ja kaheväärtuseliste kategooriatunnuste jaoks. Korrelatsioonimaatriks on esitatud lisas 3. Korrelatsioonanalüüsi koefitsiendi statistiline olulisus on tabelis märgitud tärnidega. Üks tärn tähistab statistilist olulisust olulisuse nivool 10%, kaks tärna olulisuse nivool 5% ning kolm tärna olulisuse nivool 1%. Muutujate vahel ei esine tugevat korrelatsiooni. Mõõdukas positiivne korrelatsioon esineb tööl lugemisoskuse ning tunnipalga vahel. Samuti esineb mõõdukas positiivne korrelatsioon ka tööl kirjutamisioskuse ning tunnipalga vahel. Ühtlasi esineb mõõdukas positiivne korrelatsioon ka tunnipalga ja vanuse vahel. Viimaks esineb mõõdukas positiivne korrelatsioon ka tunnipalga ning nädalas töötatud tundide vahel. Antud seosed on loogilised, sest kõrgemad oskused toovad mõistagi kaasa suurema tunnipalga ning samuti teenivad vanemad inimesed tõenäolisemalt suuremat palka. Niisamuti teenivad suuremat tunnipalka inimesed, kes teevad nädalas rohkem tööd. Nõrk negatiivne korrelatsioon esineb

tunnipalga ja soo vahel, mis tähendab, et naised teenivad võrreldes meestega madalamat tunnipalka. Nõrk negatiivne korrelatsioon esineb ka laste olemasolu ja tunnipalga vahel, mis tähendab, et laste olemasolu mõjutab tunnipalka negatiivselt.

### 2.3. Mudel

Andmeanalüüsi teise osana viib autor läbi regressioonanalüüsi kasutades harilikku vähimruutude meetodit, mis viiakse läbi statistikaprogrammis Gretl. Autor on koostanud mudeli kirjeldamiseks log-level regressioonvõrrandi. Log-level võrrand tähendab, et sõltuv muutuja mudelis on logaritmitud ning tänu sellele on mudeli tulemusi parem tõlgendada. Regressioonmudel põhineb võrrandil 1.

$$\ln(y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots \beta_n X_n + \varepsilon \quad (1)$$

kus

$\ln(y)$  – logaritmitud sõltuv muutuja ehk tunnipalk, kuhu on sisse arvestatud preemiad

$\beta_0$  – vabaliige ehk konstant

$\beta_1, \dots, \beta_n$  – mudeli parameetrid ehk koefitsendid

$X_1, \dots, X_n$  – mudeli sõltumatud muutujad

Mudeli koefitsientide tõlgendamisel saab lähtuda võrrandist 2.

$$\% \Delta y = 100 * \beta_n \quad (2)$$

See tähendab, et kui muuta mudeli sõltumatuid muutujaid ühe võrra, muutub sõltuv muutuja ligikaudu  $y$   $100 * \beta_n$  võrra. Valemit 2 võib kasutada tõlgendamisel juhul, kui sõltumatu tunnuse koefitsient jääb vahemikku -0,1 kuni 0,1. Kui koefitsiendi absoluutväärtus on suurem kui 0,1, ei anna valem 2 piisavalt täpset tulemust. Sel juhul tuleb mudeli tõlgendamisel kasutada võrrandit 3.

$$\% \Delta y = 100 * (e^{\beta_n} - 1) \quad (3)$$

Mudeli usaldusväarsuse kontrollimiseks viiakse läbi vajalikud testid. Esmalt kontrollitakse jääkliikmete alluvust normaaljaotusele. Jääkliikmete alluvust normaaljaotusele kontrollitakse

statistikaprogrammis Gretl Doorik-Hanseni testiga. Testi nullhüpotees on, et jääkliikmed alluvad normaaljaotusele ning sisukaks hüpoteesiks, et jääkliikmed ei allu normaaljaotusele. Nullhüpotees tuleb vastu võtta juhul, kui testastikule vastav olulisuse tõenäosuse  $p$  väärtus on suurem kui valitud olulisuse nivoo (selles töös 0,05). Juhul, kui võetakse vastu sisukas hüpotees, tähendab see, et parameetrite hinnangud ei ole mõjusad. Käesoleva bakalaureusetöö valimi puhul ei oma aga sisukas hüpotees suurt tähtsust, sest valim on piisavalt suur ning piisavalt suure valimi korral on parameetrite hinnangud mõjusad ka siis, kui jääkliikmed normaaljaotusele ei allu.

Teiseks kontrollitakse tunnuste vahelist multikollineaarsust samuti statistikaprogrammiga Gretl. Multikollineaarsus tähendab, et regressioonmudelis olevad regressorid on omavahel tugevalt korreleeritud, mille tulemusena on raske eristada eri tunnuste mõju sõltuvale tunnusele ja parameetrite standardvigade hinnangud on suured. (Sauga 2021) Multikollineaarsuse kontrollimiseks kontrollitakse VIF väärtuseid, mis peavad olema väiksemad või võrdsed väärtusega 10. Multikollineaarsus mõnede fiktiivsete tunnuste vahel (näiteks multikollineaarsus eri ametialade fiktiivsete tunnuste vahel) ei ole oluliseks probleemiks, sest sisuliselt on tegemist ühe ja sama tunnusega.

Kolmandaks kontrollitakse, kas mudelis esineb heteroskedastiivsus. Heteroskedastiivsuse esinemist kontrollitakse White'i testiga. Heteroskedastiivsuse esinemisel juhuslike liikmete dispersioon ei ole konstantne (Sauga 2021). Nullhüpotees tähendab heteroskedastiivsuse puudumist, mis tähendab, et vealiikmete dispersioon on konstantne. Sisukas hüpotees tuleb vastu võtta juhul, kui testi olulisuse tõenäosuse väärtus  $p$  on väiksem kui valitud olulise nivoo (selles töös 0,05). Sellisel juhul tuleb kasutada kohandatud standardvigu ning seejärel võib mudelit tõlgendada.

Viimaks kontrollitakse mudeli kuju õigsust. Mudeli kuju õigsust kontrollitakse Ramsey RESET testi abil. Nullhüpotees võetakse vastu, kui testi olulisuse tõenäosuse väärtus  $p$  on suurem kui valitud olulisuse nivoo (selles töös 0,05). See tähendab, et mudeli kuju on õige ning mudelit võib tõlgendada. Vastasel juhul võetakse vastu sisukas hüpotees, mis tähendab, et mudeli kuju ei ole korrektne ning selle tulemusi pole kohane tõlgendada. Sisuka hüpoteesi korral peab mudelit modifitseerima, eri tunnuseid logaritmides, ruutu võttes, eemaldades või lisades.

### 3. EMPIIRILINE ANALÜÜS

Käesolevas peatükis esitletakse lugejale empiirilise analüüsi tulemused. Empiiriline analüüs koostatakse statistikaprogrammis Gretl ning teostatakse hariliku vähimruutude meetodi abil. Esmalt analüüsitakse esialgse mudeli tulemusi ning seejärel korrigeeritud mudeli tulemusi. Regressioonanalüüsi tulemusena tahetakse välja selgitada, kuidas mõjutavad tunnipalka autori poolt välja valitud sõltumatud tunnused. Ühtlasi leitakse vastused ka töös esitatud uurimisküsimustele.

#### 3.1. Esialgse mudeli tulemuste analüüs

Esialgne mudel (vt Lisa 1) koostati statistikaprogrammis Gretl kasutades harilikku vähimruutude meetodit. Esialgse mudeli sõltuvaks muutujaks on logaritmitud tunnipalk, kuhu on sisse arvestatud ka preemiad. Mudeli sõltumatuteks tunnusteks on sugu, vanus, rass, laste olemasolu, lugemisoskus tööl, kirjutamisoskus tööl, töötunnid nädalas, kas vastaja on USA-s sündinud, amet, tegevusala, sektor, hariduse tase, tervis, elupaik, inglise keele lugemisoskus, inglise keele kirjutamisoskus ning regioon. Rass, amet, tegevusala, sektor, hariduse tase, tervis, elupaik, inglise keele kirjutamisoskus, inglise keele lugemisoskus ning regioon on kategooriatunnused ning neist on tehtud fiktiivsed tunnused. Mudelist jättis programm automaatselt välja puudevate tunnustega vaatlused, seega koosneb mudeli valim 1684 vaatlusest. Mudeli  $p$ -väärtus on  $9,3 \times 10^{-211}$ , ehk mudel tervikuna on statistiliselt oluline olulisuse nivool 1%. Mudeli determinatsioonikordaja ( $R^2$ ) on 0,5175 ehk mudeli selgitusvõime on 51,75%.

Mudeli usaldusväärsuse kontrollimiseks viidi esmalt läbi test, et näha, kas jääkliikmed alluvad normaaljaotusele. Testi  $p$ -väärtuseks tuli  $4,5935 \times 10^{-7}$ , mis tähendab, et vastu tuleb võtta sisukas hüpotees ehk jääkliikmed ei allu normaaljaotusele. Antud juhul ei mängi see suurt rolli, sest mudeli valim on piisavalt suur. Teiseks kontrolliti sõltumatute tunnuste vahelist multikollineaarsust. Esialgse mudeli multikollineaarsuse aruanne on nähtav lisa 2. Multikollineaarsus näitab, kui tugevalt on tunnused omavahel korreleeritud. Aruandest on näha, et VIF väärtus on suurem kui kümme ainult ameti fiktiivsetel tunnustel, kuid tegemist on ühe ja sama tunnusega, mistõttu ei

tekita see probleemi. Teiste tunnuste VIF väärtused on kõik väiksemad kui kümme, seega tunnuste vahel ei esine tugevat korrelatsiooni. Kolmandaks kontrolliti mudelis heteroskedastiivsuse esinemist White'i testiga. Testi  $p$ -väärtus tuli 0,0082, mis on väiksem kui 0,05 ehk vastu tuleb võtta sisukas hüpotees. Sisukas hüpotees tähendab, et mudelis esineb heteroskedastiivsus ning juhuslike liikmete dispersioon ei ole konstantne. Seega tuleb mudelis kasutada kohandatud standardvigu. Viimaks testiti, kas mudeli kuju on õige. Selle jaoks tehti Ramsey RESET test. Ramsey RESET testi  $p$ -väärtus tuli 0,0002, mistõttu tuleb vastu võtta sisukas hüpotees ehk mudeli kuju ei saa õigeks lugeda ning seda ei või tõlgendada. Testi sisuka hüpoteesi tõttu tuleb mudelit modifitseerida, kuni leitakse õige kujuga mudel.

### 3.2. Modifitseeritud mudeli tulemuste analüüs

Eelnevas alapeatükis tutvustati lugejale esialgset koostatud mudelit. Ramsey RESET testi tulemuseks oli sisukas hüpotees ehk esialgset mudelit ei saanud õigeks lugeda, mistõttu ei võinud seda ka tõlgendada. Mudeli korduva modifitseerimise tulemusena leiti lõplik mudel. Mudelit korrigeerides jäeti mudelist välja järgnevad tunnused: laste olemasolu, kas vastaja on USA-s sündinud, amet, tegevusala, tervislik seisund, inglise keele kirjutamis- ja lugemisoskus ning region. Uue mudeli sõltuvaks tunnuseks on endiselt logaritmitud tunnipalk, kuhu on sisse arvestatud ka preemiad. Mudeli sõltumatuteks tunnusteks on sugu, vanus, rass, hariduse tase, töötunnid nädalas, töötunnid nädalas ruutu võetuna, sektor, lugemisoskus tööl, kirjutamisoskus tööl ning elupaik. Fiktiivsed tunnused on võetud rassist, hariduse tasemest, sektorist ning elupaigast. Fiktiivsete tunnuste baastunnused on samad, mis eelnevalt mainitud andmete ja mudeli kirjelduse alapeatükis. Vanus, töötunnid nädalas ning lugemis- ja kirjutamisoskus tööl on mudelisse lisatud pideva tunnuseks. Modifitseeritud mudel on välja toodud lisas 4. Mudelist jättis programm taaskord välja puuduvate tunnustega vaatlused, mille tulemusena koosneb mudeli valim 1691 vaatlusest. Uue mudeli  $p$ -väärtus on  $9,3 \times 10^{-198}$ , ehk mudel on statistiliselt oluline olulisuse nivool 1%. Mudeli determinatsioonikordaja on 0,4121 ehk mudeli seletusvõime on 41,21%. See on küll madalam kui esialgsel mudelil, kuid endiselt arvestatav.

Uue mudeli usaldusväärsuse kontrollimiseks viidi läbi samad testid, mis esialgse mudeli korral. Esmalt kontrolliti, kas jääkliikmed alluvad normaaljaotusele. Testi  $p$ -väärtuseks tuli  $2,945 \times 10^{-5}$  ehk vastu tuleb võtta sisukas hüpotees, mis tähendab, et mudeli jääkliikmed ei allu normaaljaotusele. Taaskord ei oma see suurt tähtsust, sest mudeli valim on piisavalt suur.

Järgmiseks kontrolliti mudeli tunnuste vahelist multikollineaarsust. Multikollineaarsuse aruanne on välja toodud lisas 5. VIF väärtused on suuremad kui kümme ainult töötunnid nädalas ning töötunnid nädalas ruudu tunnuste vahel. See ei tekita probleemi, sest tegu on ühe ja sama tunnusega. Kõigi teiste tunnuste VIF väärtused on väiksemad kui kümme, mis tähendab, et tunnuste vahel ei esine tugevat korrelatsiooni. Kolmandaks kontrolliti heteroskedastiivsuse esinemist mudelis White'i testiga. Testi  $p$ -väärtuseks tuli  $5,9 \times 10^{-7}$ , mis tähendab, et vastu tuleb võtta sisukas hüpotees ehk mudelis esineb heteroskedastiivsus, mistõttu juhuslike liikmete dispersioon ei ole konstantne. Ka see pole probleem, sest mudelis on kasutatud kohandatud standardvigu. Viimaks kontrolliti mudeli kuju õigsust Ramsey RESET testi abil. Testi  $p$ -väärtuseks tuli 0,1084, mis tähendab, et vastu tuleb võtta nullhüpotees. Nullhüpotees antud testi puhul tähendab, et mudeli kuju on õige ning seda võib tõlgendada.

Modifitseeritud mudelis on soo tunnuse  $p$ -väärtus  $3,9 \times 10^{-93}$  ehk sugu on statistiliselt oluline olulisuse nivool 1%. Tunnuse koefitsient on -0,1794, seega mudeli kohaselt on naiseks olemine seotud 16,4% madalama palga teenimisega, kusjuures eri kontrolltunnuste mõju palgale on arvesse võetud. Selgub, et Ladina-Ameerika päritolu on mudelis statistiliselt ebaoluline, omades  $p$ -väärtust 0,3085. Samuti on statistiliselt mitteoluline ka tunnus teised rassid, mille  $p$ -väärtus on 0,3338. Afroameeriklased teenivad mudeli kohaselt (muude tunnuste mõju arvesse võttes) võrreldes valgenahalistega 11,3% väiksemat tunnipalka. Vastava tunnuse koefitsient on -0,1198 ning  $p$ -väärtus 0,0002 ehk kehtib statistiline olulisus olulisuse nivool 1%.

Hariduse tase mõjutab tunnipalka positiivselt ning mida kõrgem haridus, seda kõrgem tunnipalk. Keskkharidusega inimesed saavad võrreldes baasgrupiga ehk põhihariduse või madalama haridusega inimestega 20,4% kõrgemat tunnipalka. Kutseharidusega inimesed saavad võrreldes baasgrupiga 34,49% kõrgemat tunnipalka. Rakenduskõrgharidusega inimesed saavad võrreldes baasgrupiga 39,23% kõrgemat palka. Bakalaureusekraadiga inimesed teenivad 70,31% kõrgemat palka kui baasgrupp ning magistri- või doktorikraadiga inimesed tervelt 107,97% kõrgemat tunnipalka. Vanuse  $p$ -väärtus on mudelis  $9,85 \times 10^{-31}$ , mis tähendab, et tunnus on statistiliselt oluline olulisuse nivool 1%. Vanuse koefitsient on positiivne, seega kõrgem vanus on seostatud kõrgema palgaga.

Töötatud tunnid nädalas mõjutavad palka positiivselt (antud muutuja koefitsient on positiivne ja statistiliselt oluline nivool 1%), mis tähendab, et suurem töötundide arv on seotud kõrgema tunnipalgaga. Töötatud tundide ruutu võetud tunnuse koefitsient on aga negatiivne, mis tähendab,



et lisatundide tegemine on efektiivne ning positiivse mõjuga palgale ainult mingi piirini ning seejärel positiivne mõju aeglustub. Avaliku sektori tunnus ei ole mudelis statistiliselt oluline, tunnuse  $p$ -väärtus on 0,2519. Mittetulundusühingutes töötavad inimesed teenivad aga võrreldes erasektoriga 7,6% madalamat tunnipalka, tunnuse koefitsient on -0,0764 ning  $p$ -väärtus 0,05, mis tähendab statistilist olulisust olulisuse nivool 10%.

Lugemisoskus tööl ning kirjutamisoskus tööl on mõlemad palgaga positiivselt seotud. Lugemisoskus on mudelis oluline olulisuse nivool 1% ning kirjutamisoskus oluline olulisuse nivool 5%. Maakohas elamine ei oma mudelis statistilist olulisust. Äärelinnas elavad inimesed teenivad võrreldes suurlinnas elavate inimestega aga 6,3% kõrgemat tunnipalka. Tunnus on mudelis statistiliselt oluline olulisuse nivool 5%. Väikelinnas elavad inimesed teenivad võrreldes suurlinnas elavate inimestega 13,7% madalamat palka ning tunnus on oluline olulisuse nivool 1%.

### **3.3. Valgenahaliste valimiga mudeli tulemuste analüüs**

Käesoleva bakalaureusetöö uurimisküsimustele vastuste leidmiseks moodustati veel kaks mudelit, et välja selgitada kas sooline palgalõhe on suurem valgenahaliste või afroameeriklaste seas. Esmalt kitsendati valimit, valides mudelisse ainult valgenahalised inimesed. Valimis on 1277 valgenahalist inimest, kuid puuduvate vastuste tõttu teatud tunnuste puhul jäeti valimist välja 137 inimest ehk lõplikus mudelis on 1140 vaatlust.

Valgenahaliste regressioonmudeli sõltumatuteks muutujateks valiti samad tunnused, mis modifitseeritud mudeli korral: sugu, vanus, hariduse tase, töötunnid nädalas, töötunnid nädalas ruut, sektor, lugemisoskus- ja kirjutamisoskus tööl ning elupaik. Sõltuvaks tunnuseks on taaskord logaritmitud tunnipalk, kuhu on sisse arvestatud preemiad. Mudel on nähtav lisas 6. Mudeli  $p$ -väärtuseks on  $2,1 \times 10^{-139}$  ehk mudel on statistiliselt oluline olulisuse nivool 1%. Mudeli determinatsioonikordaja on 0,4015 ehk mudeli selgitusvõime on 40,2%. Mudeli usaldusväarsuse kontrollimiseks viidi läbi vajalikud testid. Esmalt testiti jääkliikmete allumist normaaljaotusele. Testi  $p$ -väärtuseks tuli 0,024 ehk vastu tuleb võtta sisukas hüpotees, mis tähendab, et jääkliikmed ei allu normaaljaotusele. Siinkohal ei ole see probleemiks, sest mudeli valimi maht on piisavalt suur. Teiseks kontrolliti sõltumatute tunnuste vahelist multikollineaarsust. Multikollineaarsuse aruanne on nähtav lisas 7. Aruandes on näha, et ainukesed tunnused, mille VIF väärtused on suuremad kui kümme, on töötunnid nädalas ja selle ruut. Nende tunnuste vaheline

multikollineaarsus ei ole oluline, sest tegu on ühe ja sama tunnusega nagu eelnevates mudeliteski. Kõikide teiste tunnuste VIF väärtused jäävad alla kümne, mis tähendab, et tunnuste vahel ei esine multikollineaarsust. Järgmiseks viidi läbi White'i test, et kontrollida heteroskedastiivsuse olemasolu mudelis. White'i testi  $p$ -väärtuseks tuli  $6,94 \times 10^{-9}$ . Selle kohaselt tuleb vastu võtta sisukas hüpotees ehk mudelis esineb heteroskedastiivsus, mille tõttu tuleb mudelis kasutada robustseid standardvigu. Viimaks kontrolliti mudeli kuju õigsust Ramsey RESET testi abil. Testi  $p$ -väärtus tuli 0,2827 ehk vastu tuleb võtta nullhüpotees, seega on mudeli kuju korrektne ning seda võib tõlgendada.

Mudelis on sugu statistiliselt oluline olulisuse nivool 1%. Tunnuse koefitsient on -0,1929 ehk mudeli kohaselt teenivad naised (muude tunnuste mõju samuti arvesse võttes) 17,5% madalamat tunnipalka kui mehed. Vanus mõjutab tunnipalka positiivselt, tunnus on mudelis statistiliselt oluline olulisuse nivool 1% ning koefitsient on positiivne. Samuti mõjutab tunnipalka positiivselt hariduse tase – mida kõrgem haridus, seda kõrgem palk. Kõik hariduse tasemed on mudelis statistiliselt olulised olulisuse nivool 1%. Kesklaridusega inimesed teenivad võrreldes põhihariduse või madalama haridusega inimestest (baasgrupp) 34% kõrgemat tunnipalka. Kutseharidusega inimesed teenivad võrreldes baasgrupiga 59,8% kõrgemat tunnipalka. Rakenduskõrgharidusega inimesed teenivad baasgrupist 58,2% kõrgemat tunnipalka. Bakalaureusekraadiga inimesed teenivad baasgrupiga võrreldes 87,34% kõrgemat tunnipalka ning magistri- või doktorikraadiga inimesed tervelt 134,4% kõrgemat tunnipalka, mis on märkimisväärselt rohkem.

Töötundide arv nädalas mõjutab tunnipalka samuti positiivselt – pikem töönael on seotud kõrgema tunnipalgaga. Lisatundide tegemine on sarnaselt eelnevale mudelile efektiivne ainult mingi piirini, seda näitab töötunnid nädalas tunnuse ruut, mille koefitsient on negatiivne. Avaliku sektori tunnus on mudelis statistiliselt oluline olulisuse nivool 10% ning selle koefitsient on 0,0612 ehk avalikus sektoris töötajad teenivad võrreldes erasektoriga 6,12% madalamat tunnipalka (muude tunnuste mõju arvesse võttes). Mittetulundusühingu tunnus on mudelis statistiliselt mitteoluline. Lugesioskus tööol on mudelis statistiliselt oluline olulisuse nivool 1% ning tunnipalgaga positiivselt seotud. Kirjutamisoskus tööol ei ole mudelis statistiliselt oluline. Äärelinnas elavad inimesed teenivad võrreldes suurlinnas elavate inimestega 7,8% kõrgemat tunnipalka ning väikelinnas elavad inimesed teenivad võrreldes suurlinnas elavate inimestega 13,3% madalamat tunnipalka (muude tunnuste mõju arvesse võttes). Maakohas elamine ei oma mudelis statistilist olulisust.

### 3.4. Afroameeriklaste valimiga mudeli tulemuste analüüs

Viimaseks analüüsitavaks mudeliks on tunnipalka mõjutavate tegurite regressioon afroameeriklaste seas. Valimis oli esialgselt 266 vaatlust ent teatud tunnuste puuduvate väärtuste tõttu eemaldati mudelist 57 väärtust ning mudeli lõplik vaatluste arv on 209. Antud mudeli sõltumatuteks tunnusteks valiti samad tunnused, mis valgenahaliste valimiga mudeli korral. Sõltuvaks tunnuseks on endiselt logaritmitud tunnipalk, kuhu on sisse arvestatud preemiad. Mudel on nähtav lisas 8.

Mudeli  $p$ -väärtuseks tuli  $1,67 \times 10^{-13}$  ehk mudel on statistiliselt oluline olulisuse nivool 1%. Mudeli determinatsioonikordaja on 0,3842 ehk mudeli kirjeldusvõime on 38,4%. Mudeli usaldusväarsuse kontrollimiseks viidi taaskord läbi samad vajalikud testid, mis eelnevate mudelite korral. Esmalt kontrolliti jääkliikmete alluvust normaaljaotusele. Testi  $p$ -väärtuseks tuli 0,0012 ehk vastu tuleb võtta sisukas hüpotees – see tähendab, et jääkliikmed ei allu normaaljaotusele. Järgmiseks kontrolliti sõltumatute tunnuste vahelist multikollineaarsust. Multikollineaarsuse aruanne on nähtav lisas 9. Taaskord esineb multikollineaarsus ainult töötunnid nädalas tunnuse ning selle ruudu vahel, nende VIF väärtus on suurem kui kümme. Nagu eelnevalt mainitud, ei tekita see mudelis probleemi, sest tegemist on ühe ja sama tunnusega. Teiste tunnuste VIF väärtused on alla kümne, mis tähendab, et tunnuste vahel ei esine multikollineaarsust. Järgmiseks kontrolliti heteroskedastiivsuse esinemist White'i testiga. Testi  $p$ -väärtuseks tuli 0,884016, mis tähendab, et vastu tuleb võtta nullhüpotees ehk mudelis ei esine heteroskedastiivsust ehk vealiikmete dispersioon on konstantne. Viimaks kontrolliti mudeli kuju õigsust Ramsey RESET testiga. Testi  $p$ -väärtuseks tuli 0,5856 ehk vastu tuleb võtta nullhüpotees – mudel on õige kujuga ning seda võib tõlgendada.

Sugu on antud mudelis statistiliselt oluline olulisuse nivool 1%. Tunnuse koefitsient on -0,1593, mis tähendab, et mudeli kohaselt teenivad naised, muude tunnuste mõju samuti arvesse võttes, 14,7% madalamat tunnipalka. Vanus on samuti statistiliselt oluline olulisuse nivool 1% ning palgaga positiivselt seotud.

Sarnaselt eelnevatele mudelitele on kõrgem haridustase seotud kõrgema palgaga. Antud mudelis on kutsehariduse tunnus küll statistiliselt mitteoluline, kuid positiivse koefitsiendiga. Keskkhariduse tunnus on mudelis statistiliselt oluline olulisuse nivool 5%. Keskkharidusega inimesed teenivad võrreldes põhihariduse või madalama haridusega inimestest (baasgrupp) 34%

kõrgemat tunnipalka. Rakenduskõrgharidus on mudelis samuti statistiliselt oluline olulisuse nivool 5% ning nemad teenivad võrreldes baasgrupiga 38,3% kõrgemat tunnipalka. Bakalaureusekraad ning magistri- või doktorikraad on mudelis statistiliselt olulised olulisuse nivool 1% ning nemad teenivad baasgrupiga võrreldes vastavalt 71,2% ning 103,4% kõrgemat tunnipalka.

Töötundide tunnus on antud mudelis statistiliselt oluline olulisuse nivool 10% ning tunnuse koefitsient on positiivne ehk keskmiselt ning muude tunnuste mõju arvesse võttes on pikem töönael seotud kõrgema tunnipalgaga. Erinevalt eelnevatest mudelitest ei esine antud mudelis tendentsi, et lisatundide tegemine on efektiivne ainult mingi piirini, sest töötunnid nädalas tunnuse ruut on mudelis statistiliselt mitteoluline, kuigi sarnaselt teistele mudelitele on selle tunnuse koefitsient negatiivne. See võib olla põhjustatud asjaolust, et afroameeriklaste mudel on tunduvalt väiksema valimi mahuga kui ülejäänud kaks mudelit, mistõttu on statistiliselt olulisi tulemusi keerulisem leida. Samuti on mudelis statistiliselt mitteoluline sektori tunnus, elupaiga tunnus ning lugemisoskus tööl. Parem kirjutamisoskus on tunnipalgaga positiivselt seotud.

### **3.5. Järeldused**

Käesoleva bakalaureusetöö eesmärgiks on välja selgitada, kui suur on rassiline ning sooline palgalõhe USAs ning kas sooline palgalõhe erineb eri rasside lõikes. Selle jaoks koostati mudel programmis Gretl hariliku vähimruutude meetodi abil. Mudeli sõltuvaks tunnuseks valiti tunnipalk, kuhu on sisse arvestatud preemiad. Sõltumatud tunnused valiti mudelisse vastavalt sellele, mis võiksid potentsiaalselt tunnipalka mõjutada. Samuti oli arvestatud ka muutujatega, mida olid oma mudelites kasutanud varasemate empiiriliste uuringute autorid. Mudeli korrigeerimise tulemusena eemaldati mudelist mõned välja valitud sõltumatud muutujad.

Valimi kirjeldavast statistikast on näha keskmisi tunnipalkasid sugude ja rasside lõikes. Võrreldes valimi keskmisi palkasid valgenahaliste meeste ja naiste vahel, on näha, et mehed teenivad 25,6% kõrgemat tunnipalka. Afroameeriklasest naise ja mehe palgaerinevus on 10,32%. Kirjeldava statistika järgi on valgenahaliste töötajate palkade erinevus naiste ja meeste vahel märkimisväärselt suurem kui afroameeriklaste vahel. Valgenahalise mehe ja afroameeriklasest naise tunnipalka võrreldes on erinevus veelgi suurem – tervelt 57,88%. Wilson ja Rodgers III (2016) on oma töös leidnud sarnase tulemuse valgenahaliste meeste ja afroameeriklasest naiste tunnipalga vahel. Vaadeldes tunnipalkasid eri rassidest meeste vahel, on näha, et valgenahalised mehed teenivad

42,29% kõrgemat tunnipalka kui afroameeriklasest mehed. Taaskord on Wilson ja Rodgers III (2016) oma töös leidnud sarnase erinevuse. Eri rassidest naiste tunnipalkasid võrreldes on samuti näha üsna suurt tunnipalga erinevust – valgenahaliste naiste keskmine tunnipalk on afroameeriklasest naiste tunnipalgast 25,03% kõrgem. Valimi kirjeldava statistika põhjal on rassiline keskmise tunnipalga erinevus suurem meeste seas. Sarnasele tulemusele on oma töös jõudnud ka Johnson ja Neal (1996), Bayard *et al.* (1998) ning Wilson ja Rodgers III (2016).

Eelnevalt võrreldud keskmiste tunnipalkade erinevused on vaadeldud valimi kirjeldava statistika tabelist ning seal ei ole arvesse võetud erinevaid palka mõjutavaid tegureid. Hariliku vähimruutude meetodi abil moodustatud regressioonmudelil on sõltumatute muutujatena arvesse võetud järgmised tunnused: sugu, vanus, hariduse tase, töötundide arv nädalas, sektor, lugemis- ja kirjutamisoskus ning elupaik. Sõltumatuid tunnuseid arvesse võttes teenivad naised 16,4% madalamat tunnipalka kui mehed. Ühtlasi teenivad mudeli kohaselt afroameeriklasest töötajad 11,3% madalamat tunnipalka. Vaadeldes regressioonmudelit, kus valimiks on ainult valgenahalised, on näha, et naised teenivad võrreldes meestega 17,5% madalamat tunnipalka. Afroameeriklastest koosneva valimi puhul esineb sarnane tendents – naised teenivad meestest 14,7% madalamat tunnipalka. Sooline tunnipalga erinevus on suurem siiski valgenahaliste seas.

Tunnipalka mõjutavad mudelis lisaks soo ning rassile erineval määral ka teised tegurid. Suur mõju tunnipalgale on haridustasemel. Modifitseeritud mudelis on kõige suurem positiivne mõju magistri- või doktorikraadi tunnusel, mille korral on tunnipalk võrreldes põhihariduse või madalama haridusega inimese tunnipalgast 107,97% kõrgem. Valgenahaliste regressioonmudelil on see vastavalt 134,4% ning afroameeriklaste regressioonmudelil 103,4%. Modifitseeritud mudeli kohaselt saab keskharidusega inimene võrreldes baasgrupiga 20,4% kõrgemat palka, valgenahaliste ja afroameeriklaste mudeli kohaselt on see erinevus 34%. Seosed on loogilised, sest iga kõrgem hariduse tase annab inimesele rohkem teadmisi ja oskusi ning tänu sellele ka kõrgema tunnipalga.

Tunnipalka mõjutab positiivselt ka töötatud tundide arv nädalas. Antud tunnuse positiivne mõju pole aga lineaarne ning alates teatud piirist lisanduvate töötundide positiivne mõju palgale aeglustub. Valgenahaliste ning afroameeriklaste mudelite kohaselt on positiivne mõju tunnipalgale sama. Avalikus sektoris töötamine ei oma modifitseeritud mudelis statistilist olulisust. Mittetulundusühingutes töötavad inimesed teenivad võrreldes erasektoriga 7,6% madalamat palka. Valgenahaliste mudelis teenivad avaliku sektori töötajad võrreldes erasektoriga 6,12% madalamat

tunnipalka. Mittetulundusühingu tunnus ei oma mudelis statistilist olulisust. Afroameeriklaste mudelis on kõik sektorid statistiliselt mitteolulised.

Positiivse mõjuga tunnipalgale on ka oskused. Modifitseeritud mudeli kohaselt on lugemis- ning kirjutamisoskus palgaga positiivselt seotud. Valgenahaliste valimiga mudelis on lugemisoskus samuti palgaga positiivses seoses. Kirjutamisoskus ei oma antud mudelis statistilist olulisust. Afroameeriklaste mudelis vastupidiselt on kirjutamisoskus palgaga positiivses seoses ning lugemisoskus ei ole statistiliselt oluline. Viimaks mõjutab tunnipalka ka inimese elukoht. Nii modifitseeritud mudelis kui ka valgenahaliste mudelis teenivad äärelinnas elavad inimesed võrreldes suurlinnas elavate inimestega kõrgemat palka, vastavalt 6,3% ning 7,8%. Selline seos on loogiline, sest suurema tõenäosusega elavad äärelinnades jõukamad inimesed, kes omavad näiteks eramaja, kuid võivad töötada sellegipoolest suurlinnas. Väikelinnas elavad inimesed teenivad võrreldes suurlinnas elavate inimestega modifitseeritud mudeli kohaselt 13,7% madalamat tunnipalka ning valgenahaliste mudeli kohaselt 13,3% madalamat tunnipalka. Mõlemas mudelis on maakohas elamine statistiliselt ebaoluline ning statistilist olulisust ei oma elukoht ka afroameeriklaste mudelis.

Oluline on välja märkida, et eelpool mainitud palka mõjutavad tegurid on kehtivad välja arvatud protsendi juures vaid juhul, kui teised muutujad jäävad samaks. Samuti mõjutavad peale mudelisse sisse arvatud muutujate tunnipalka veel väga palju erinevaid tegureid, seega täpsema soolise ning rassilise palgalõhe uurimiseks tuleks mudelisse hõlmata veelgi rohkem sõltumatuid muutujaid.

## KOKKUVÕTE

Käesoleva bakalaureusetöö raames taheti välja selgitada soo ning rassi mõju tunnipalgale Ameerika Ühendriikides. Samuti oli üheks eesmärgiks võrrelda tunnipalkade erinevusi eri rasside lõikes. Töö eesmärkide saavutamiseks püstitas autor järgnevad uurimisküsimused:

- 1) Kuidas mõjutab sugu palka, kui muude tunnuste mõju palgale on samuti arvesse võetud?
- 2) Kuidas mõjutab rass palka, kui muude tunnuste mõju palgale on samuti arvesse võetud?
- 3) Kas sooline palgalõhe erineb eri rasside lõikes?

Püstitatud uurimisküsimustele vastuste leidmiseks koguti andmeid PIAAC (*Programme for the International Assessment of Adult Competencies*) uuringust ning nendega viidi läbi regressioonanalüüs programmis Gretl. Töös kasutati ka programmi RStudio, millega tehti andmete töötlus ja puhastus ning korrelatsioonanalüüs. Regressioonanalüüsi sõltuvaks muutujaks valiti tunnipalk, kuhu on sisse arvestatud preemiad. Sõltumatuteks tunnusteks valiti lõplikku mudelisse sugu, rass, vanus, hariduse tase, töötundide arv nädalas, sektor, lugemis- ning kirjutamisoskus tööl ja elupaik. Kokku koostati kolm erinevat regressioonmudelit. Esimene modifitseeritud mudel hõlmas kõiki rasse. Teine mudel viidi läbi ainult valgenahaliste valimiga ning kolmas mudel afroameeriklaste mudeliga. Kõikides mudelites kasutati samu kontrollmuutujaid. Sõltumatud muutujad valiti vastavalt sellele, mis võiks kõige enam potentsiaalselt tunnipalka mõjutada, ühtlasi vaadati ka varasemas empiirilises kirjanduses kasutatavaid sõltumatuid muutujaid.

Töös koostati kirjeldava statistika tabel, kust on näha valimi keskmiseid tunnipalku soo ning rasside lõikes. Tabelist selgus, et valgenahaliste meeste keskmine tunnipalk on valgenahaliste naiste keskmisest tunnipalgast 25,6% kõrgem. Afroameeriklasest mehe keskmine tunnipalk on afroameeriklasest naise keskmisest tunnipalgast 10,32% kõrgem. Valgenahalised mehed teenivad aga afroameeriklasest meestega võrreldes 42,29% kõrgemat keskmist tunnipalka. Valgenahaliste naiste keskmine tunnipalk on afroameeriklasest naiste keskmisest tunnipalgast 25,03% kõrgem. Võrreldud keskmised tunnipalkade erinevused on pärit kirjeldava statistika tabelist ning seal ei ole arvesse võetud tunnipalka mõjutavaid tegureid.

Bakalaureusetöö empiirilises osas koostati hariliku vähimruutude meetodi abil regressioonanalüüs, kuhu on arvesse võetud sõltumatud muutujad, mis palka mõjutavad. Kõikide tunnuste protsentuaalsed muutused on arvestatavad vaid juhul, kui teised tunnused jäävad samaks. Mudelis, kus valimis on kõik rassid, teenivad naised 16,4% madalamat tunnipalka kui mehed. Ainult valgenahaliste valimiga mudelis on see erinevus 17,5% ning afroameeriklaste valimiga mudelis 14,7%. Kõikide rassidega mudelis teenivad afroameeriklasest töötajad võrreldes valgenahalistega 11,3% madalamat tunnipalka.

Lisaks soole ja rassile oli regressioonanalüüsi sisse arvestatud veel mitmeid erinevaid kontrollmuutujaid. Võrreldes baasgrupiga (põhiharidus või madalam) mõjutab haridustaseme poolest palka positiivselt kõige enam magistri- või doktorikraad ning kõige vähem keskharidus. Positiivse mõjuga palgale on ka töötundide arv nädalas, aga seda ainult mingi kindla piirini. Peale seda ei ole lisa töötundide tegemine enam efektiivne ega positiivse mõjuga. Kirjutamis- ning lugemisoskuste omandamine mõjutab tunnipalka samuti positiivselt. Äärelinnas elavad inimesed teenivad mudeli kohaselt võrreldes suurlinnas elavate inimestega suuremat palka ning väikelinnas elavad inimesed väiksemat palka.

Käesoleva bakalaureusetöö eesmärgiks oli uurida soolist ning rassilist palgalõhet Ameerika Ühendriikides. Uurimuse käigus leiti vastused kõikidele uurimisküsimustele, seega sai bakalaureusetöö eesmärk täidetud. Täpsemate tulemuste leidmiseks saaks koostada veelgi spetsiifilisema mudeli, kaasates rohkem erinevaid kontrollmuutujaid.



# **SUMMARY**

## **GENDER AND RACIAL WAGE GAP IN THE UNITED STATES**

Kelly Künnapas

The goal of this bachelor's thesis was to find out the effect of gender and race on hourly wages in the United States. Another goal was to compare the differences in hourly wages between different races. In order to find the purpose of the thesis, author has raised the following reasearch questions:

- 1) How does gender affect wage if the effect of other characteristics on wage is also taken into account?
- 2) How does race affect wage if the effect of other characteristics on wage is also taken into account?
- 3) Does the gender pay gap differ between races?

In order to answer to the research questions, data was collected from the PIAAC study and regression analysis was performed in the Gretl programme. RStudio was also used during work for data processing and correlation analysis. The dependent variable of the regression was the hourly wage, which includes bonuses. Independent variables in the regression were gender, race, age, level of education, number of hours worked per week, sector, literacy and reading skills at work and urbanization. A total of three different regression models were constructed. The first modified model included all races. The second model was performed with a white sample only and the third model with an African American sample. The same control variables were used in all models. Independent variables were selected according to what could most potentially affect hourly wages, as well as independent variables used in previous empirical literature.

The descriptive statistics table, which shows the average hourly wages of the sample by gender and race, shows that the average hourly wage of white men is 25,6% higher than the average hourly wage of white woman. The average hourly wage of an African American man is 10,32% higher

than the average hourly wage of an African American woman. White men earn 42,29% higher average hourly wages than African American men. The average hourly wage of white women is 25,03% higher than the average hourly wage of African American women. The compared average differences in hourly wages are from the table of descriptive statistics and do not take into account the factors that affect hourly wages.

The regression model constructed takes into account independent variables that affect wages. The percentage change in all variables is only relevant if the other variables remain the same. In the model with all races, women earn 16,4% lower hourly wage than men. In the model with only white people, this difference is 17,5% and with African American people, the difference is 14,7%. In the model with all races, African American workers earn 11,3% less per hour than white workers.

In addition to gender and race, there are other factors affecting income. Compared to the base group, a master's or doctor degree has the biggest positive effect on salary in terms of educational level. Secondary education has the least positive effect. The number of working hours per week also has a positive effect on wages, but only up to a certain point. After that, working extra hours is no longer effective. Acquiring writing and reading skills at work also has a positive effect on hourly wages. According to the model, people living in the suburbs earn higher wages than people living in the city. People living in the small towns earn less than people living in the city.

The purpose of this bachelor's thesis was to study the gender and racial wage gap in the United States. In the thesis all the answers were found to the research questions, so the goal of the bachelor's thesis was fulfilled. To find more accurate results, an even more specific model could be done by including more different control variables.

## KASUTATUD ALLIKATE LOETELU

Amuedo-Dorantes, C., Kimmel, J. (2005). The Motherhood Wage Gap for Women in the United States: The Importance of College and Fertility Delay. *Review of Economics of the Household*, 3, 17–48.

Analüüsis kasutatavad andmed. Kättesaadav: [https://livettu-my.sharepoint.com/:x:/g/personal/kekunn\\_ttu\\_ee/Ef4v6u7EXu1JuFJtMmsovN4B5ZANbGxLSYmkVQVF0trHBw?rttime=jrZWpNoL2Ug3](https://livettu-my.sharepoint.com/:x:/g/personal/kekunn_ttu_ee/Ef4v6u7EXu1JuFJtMmsovN4B5ZANbGxLSYmkVQVF0trHBw?rttime=jrZWpNoL2Ug3), 30. aprill 2021.

Anderson, D., Shapiro, D. (1996). Racial Differences in Access to High-Paying Jobs and the Wage Gap between Black and White Women. *ILR Review*, 49 (2), 273–286.

Anspal, S., Biin, H., Kallaste, E., Karu, M., Kraut, L. (2009). *Sooline palgalõhe*. Kättesaadav: <http://praxis.ee/wp-content/uploads/2014/03/2010-Sooline-palgalohe-kirjandus.pdf>, 21. märts 2021.

Avery, D. R., Hernandez, M., Kaiser, C. R., Volpone, S. D. (2019). Bargaining While Black: The Role of Race in Salary Negotiations. *Journal of Applied Psychology*, 104 (4), 581–592.

Blau, F. D., Kahn, L. M. (2016). The Gender Wage Gap: Extent, Trends, and Explanations. *Nber Working Paper Series*, No. 21913.

Bayard, K., Hellerstein, J., Neumark, D., Troske, K. (1998). Why are Racial and Ethnic Wage Gaps Larger for Men than for Women? Exploring the Role of Segregation Using the New Worker-Establishment Characteristics Database. *Contributions to Economic Analysis*, 241, 175–203.

Cable, D. M., Judge, T. A. (1997). Applicant Personality, Organizational Culture, and Organization Attraction. *Personnel Psychology*, 50 (2), 359–395.

Coleman, M. G. (2003). Job Skill and Black Male Wage Discrimination. *Social Science Quarterly*, 84 (4), 892–905.

Couch, K., Daly, M. (2002). Black-White Wage Inequality in the 1990s: a Decade of Progress. *Economic Inquiry*, 40 (1), 31–41.

Darity Jr., W. A., Mason, P. L. (1998). Evidence on Discrimination in Employment: Codes of Color, Codes of Gender. *Journal of Economic Perspectives*, 12 (2), 63–90.

Davila, A., Mora, M. T. (2018). *The Hispanic-white wage gap has remained wide and relatively steady*. Kättesaadav: <https://files.eric.ed.gov/fulltext/ED593384.pdf>, 01.mai 2021.

- Dittrich, M., Knabe, A., Leipold, K. (2014). Gender Differences in Experimental Wage Negotiations. *Economic Inquiry*, 52 (2), 862–873.
- Donohue III J. J., Siegelman, P. (1991). The Changing Nature of Employment Discrimination Litigation. *Stanford Law Review*, 43 (5), 983–1033.
- Eschbach K., Waters M. C. (1995). Immigration and Ethnic and Racial Inequality in the United States. *Annual Review of Sociology*, 21, 419–446.
- Farkas, G., Vicknair, K. (1996). Appropriate Tests of Racial Wage Discrimination Require Controls for Cognitive Skill: Comment on Cancio, Evans, and Maume. *American Sociological Review*, 61 (4), 557–560.
- Fields, J., Wolff, E. N. (1995). Interindustry Wage Differentials and the Wage Gap. *ILR Review*, 49, (1), 105–120.
- Gardaezabal, J., Ugidos, A. (2003). Gender Wage Discrimination at Quantiles. *Journal of Population Economics*, 18, 165–179.
- Goldin, C. (2006). The Quiet Revolution that Transformed Women’s Employment, Education, and Family. *Nber Working Paper Series*, No. 11953.
- Grodsky, E., Pager, D. (2001). The Structure of Disadvantage: Individual and Occupational Determinants of the Black-White Wage Gap. *American Sociological Review*, 66, 542–567.
- Jencks, C., Peterson, P. E. (koost) (1991). *The Urban Underclass*. Washington, D. C.: The Brookings Institution.
- Johnson, W., Neal, D. A. (1996). The Role of Premarket Factors in Black-White Wage Differences. *Journal of Political Economy*, 104 (5), 869–895.
- Loughran, D. S., Zissimopoulos, J. M. (2009). Why Wait? The Effect of Marriage and Childbearing on the Wages of Men and Women. *The Journal of Human Resources*, 44 (2), 326–349.
- Lundberg, S. J., Startz, R. (1983). Private Discrimination and Social Intervention in Competitive Labor Market. *The American Economic Review*, 73 (3), 340–347.
- Mueller, G., Plug, E. (2006). Estimating the Effect of Personality on Male and Female Earnings. *Industrial and Labor Relations Review*, 60 (1), 3–22.
- Noguera, P. A. (2017). Introduction to „Rachial Inequality and Education: Patterns nad Prospects for the Future. *The Educational Forum*, 81, 129–135.
- O’Neill, J. (1985). The Trend in the Male-Female Wage Gap in the United States. *Journal of Labor Economics*, 3 (1), S91–S116.
- O’Neill, J. (1990). The Role of Human Capital in Earnings Differences Between Black and White Men. *Journal of Economic Perspectives*, 4 (4), 25–45.

- Petre, M. (2019). Contribution of Skills to the Racial Wage Gap. *Journal of Human Capital*, 13 (3), 479–518.
- Polachek, S. W. (2004). How the Human Capital Model Explains Why the Gender Wage Gap Narrowed. *IZA Discussion Papers*, No. 1102.
- Rodgers III, W. M., Wilson, V. (2016). *Black-white wage gaps expand with rising wage inequality*. Kättesaadav: <https://files.epi.org/pdf/101972.pdf>, 01. aprill 2021.
- Sauga, A. *Vabalt kasutatav ökonomeetriapakett Gretl*. Kättesaadav: <https://www.sauga.pri.ee/gretl/index.html>, 22. aprill 2021.
- Walters, P. B. (2001). Educational Access and the State: Historical Continuities and Discontinuities in Racial Inequality in American Education. *Sociology of Education*, 74, 35-49.
- Wilson, W. J. (koost) (1996). *When Work Disappears: The World of the New Urban Poor*. New York: A Division of Random House, Inc.

# LISAD

## Lisa 1. Esialgne OLS mudel

	Koefitsient	Standardviga	P-väärtus	
Konstant	3.00188	0.215284	8.06e-42	***
Sugu	-0.152210	0.0231405	6.43e-11	***
Vanus	0.0478977	0.00428999	6.14e-28	***
Laste olemasolu	-0.0156068	0.0240456	0.5164	
Lugemisoskus tööl	0.0479871	0.0142508	0.0008	***
Kirjutamisoskus tööl	0.00293662	0.0119615	0.8061	
Töötunnid nädalas	0.00174376	0.000966485	0.0714	*
Kas USAs sündinud	-0.0366674	0.0373779	0.3267	
<u>Rass:</u>				
Ladina-Ameerika	0.00998615	0.0349498	0.7751	
Afroameeriklane	-0.0504743	0.0330149	0.1265	
Teisest rassist	0.0492145	0.0392733	0.2103	
<u>Ametid:</u>				
Õigus ja kõrgemad ametnikud	-0.176053	0.188012	0.3492	
Spetsialistid	-0.127268	0.187023	0.4963	
Tehnikud	-0.327148	0.187119	0.0806	*
Ametnikud	-0.483458	0.190084	0.0111	**
Teenindustöötajad	-0.496475	0.187573	0.0082	***
Põllumajandus- ja kalandustöötajad	-0.496256	0.246684	0.0444	**
Käsitöö	-0.475719	0.192694	0.0137	**
Tehase- ja masinaoperaatorid	-0.523302	0.193456	0.0069	***
Lihttöölised	-0.631551	0.192951	0.0011	***
<u>Tegevusalad:</u>				
Kaevandus	0.155564	0.0454225	0.0006	***
Tootmine	0.0210490	0.0956323	0.8258	
Elektri- ja gaasitootmine	-0.220184	0.0424716	2.44e-07	***
Kanaliseerimine ja jäätmeäritlus	-0.0376558	0.0431262	0.3827	
Ehitus	-0.0809456	0.0523816	0.1225	
Hulgi- ja jaekaubandus ning mootorsõidukite remont	0.0306787	0.132320	0.8167	
Transport ja ladustamine	0.191410	0.0547141	0.0005	***
Majutus- ja toitlustusteenused	0.00950178	0.0601070	0.8744	
Informatsiooni- ja kommunikatsioonitehnoloogia	-0.248630	0.262433	0.3436	
Finants- ja kindlustustegevus	0.0875051	0.0560644	0.1188	

Kinnisvara	0.145768	0.0531147	0.0061	***
Kutse- ja teadustegevus	0.211947	0.0532186	7.12e-05	***
Haldus- ja tugiteenused	-0.172683	0.0717345	0.0162	**
Avalik haldus	0.0949362	0.0584632	0.1046	
Haridus	-0.228940	0.0776570	0.0032	***
Sotsiaaltöö	0.151210	0.103584	0.1445	
Kunst ja meelelahutus	0.243185	0.0584583	3.35e-05	***
Teenused	1.59900	0.410636	0.0001	***
Kodumajapidmised tööandjana või toodete/teenuste tootmine enda tarbeks	0.226095	0.148143	0.1272	
Ekstraterritoriaalsete organisatsioonide tegevus	0.226235	0.410013	0.5812	
<u>Sektorid:</u>				
Avalik sektor	0.00409198	0.0300119	0.8916	
Mittetulundühing	-0.0285510	0.0408211	0.4844	
<u>Hariduse tase:</u>				
Keskharidus	0.111459	0.0497514	0.0252	**
Kutseharidus	0.204424	0.0589690	0.0005	***
Rakenduskõrgharidus	0.232789	0.0563353	3.78e-05	***
Kõrgharidus - bakalaureusekraad	0.324946	0.0542483	2.58e-09	***
Kõrgharidus – magistri- või doktorikraad	0.520480	0.0602319	1.31e-17	***
<u>Tervislik seisund:</u>				
Väga hea	0.0198221	0.0274288	0.4700	
Hea	-0.0219036	0.0287891	0.4469	
Rahuldav	-0.0700608	0.0410275	0.0879	*
Halb	-0.124373	0.0854569	0.1458	
<u>Elupaik:</u>				
Äärelinn	0.0383230	0.0239711	0.1101	
Väikelinn	-0.114656	0.0370590	0.0020	***
Maakoht	-0.0152142	0.0316480	0.6308	
<u>Inglise kl kirjaoskus:</u>				
Hea	-0.000971150	0.0467653	0.9834	
Halb	-0.112871	0.0925412	0.2228	
Puudulik	-0.363298	0.256543	0.1569	
<u>Inglise kl lugemisoskus:</u>				
Hea	-0.0352440	0.0566819	0.5342	
Halb	0.0143274	0.147678	0.9227	
Puudulik	0.285650	0.327032	0.3825	
<u>Regioon:</u>				
Kesk-lääneosa	-0.0839632	0.0360404	0.0199	**
Lõunaosa	-0.0986090	0.0359775	0.0062	***

Põhjaosa	0.0227438	0.0399301	0.5690
----------	-----------	-----------	--------

Determinatsioonikordaja:	0.517531
--------------------------	----------

Mudeli P-väärtus	9.3e-10
------------------	---------

White's test for heteroskedasticity -

Null hypothesis: heteroskedasticity not present

Test statistic: LM = 96.6474

with p-value =  $P(\text{Chi-square}(66) > 96.6474) = 0.00829287$

Test for normality of residual -

Null hypothesis: error is normally distributed

Test statistic: Chi-square(2) = 29.1869

with p-value = 4.59351e-07

RESET test for specification -

Null hypothesis: specification is adequate

Test statistic:  $F(2, 1619) = 8.51556$

with p-value =  $P(F(2, 1619) > 8.51556) = 0.000209439$

Allikas: PIAAC (2021), autori poolt koostatud



## Lisa 2. Esialgse OLS mudeli multikollineaarsuse aruanne

Sugu	1.384
Vanus	1.404
Laste olemasolu	
Lugemisoskus tööl	1.304
Kirjutamisoskus tööl	1.214
Töötunnid nädalas	1.259
Kas USAs sündinud	1.363
<u>Rass:</u>	1.615
Ladina-Ameerika	1.534
Afroameeriklane	1.375
Teisest rassist	1.498
<u>Ametid:</u>	
Õigus ja kõrgemad ametnikud	41.337
Spetsialistid	68.162
Tehnikud	61.559
Ametnikud	23.470
Teenindustöötajad	52.753
Põllumajandus- ja kalandustöötajad	2.986
Käsitöö	19.696
Tehase- ja masinaoperaatorid	19.440
Lihttöölised	17.273
<u>Tegevusalad:</u>	
Kaevandus	2.094
Tootmine	1.114
Elektri- ja gaasitootmine	2.060
Kanalisatsioon ja jäätmekäitlus	2.046
Ehitus	1.650
Hulgi- ja jaekaubandus ning mootorsõidukite remont	1.072
Transport ja ladustamine	1.456

Majutus- ja toitlustusteenused	1.391
Informatsiooni- ja kommunikatsioonitehnoloogia	1.271
Finants- ja kindlustustegevus	1.317
Kinnisvara	1.802
Kutse- ja teadustegevus	1.393
Haldus- ja tugiteenused	1.178
Avalik haldus	1.662
Haridus	1.167
Sotsiaaltöö	1.113
Kunst ja meelelahutus	1.566
Teenused	1.038
Kodumajapidmised tööandjana või toodete/teenuste tootmine enda tarbeks	1.077
Ekstraterritoriaalsete organisatsioonide tegevus	1.035
<u>Sektorid:</u>	
Avalik sektor	1.785
Mittetulundühing	1.335
<u>Hariduse tase:</u>	
Keskharidus	5.606
Kutseharidus	2.625
Rakenduskõrgharidus	3.521
Kõrgharidus - bakalaureusekraad	6.038
Kõrgharidus – magistri- või doktorikraad	4.961
<u>Tervislik seisund:</u>	
Väga hea	1.843
Hea	1.843

Rahuldav	1.392
Halb	1.108
<u>Elupaik:</u>	
Äärelinn	1.401
Väikelinn	1.307
Maakoht	1.465
<u>Inglise kl kirjaoskus:</u>	
Hea	2.113
Halb	2.061
Puudulik	1.618
<u>Inglise kl lugemisoskus:</u>	
Hea	2.358
Halb	1.998
Puudulik	1.316
<u>Regioon:</u>	
Kesk-lääneosa	2.994
Lõunaosa	3.145
Põhjaosa	2.534

Allikas: PIAAC (2021), autori poolt koostatud

### Lisa 3. Andmete korrelatsioonimaatriks (Spearmani korrelatsioonikordajad)

	Tunnipalk	Sugu	Vanus	Laste olemasolu	Lugemis oskus tööl	Kirjutamis oskus	Tervis	Tunnid nädalas	Kas USAs sündinud	Inglise kl kirjaoskus	Inglise kl lugemisoskus
Tunnipalk	<b>1.00</b>	-	-	-	-	-	-	-	-	-	-
Sugu	-0.14 ***	<b>1.00</b>	-	-	-	-	-	-	-	-	-
Vanus	0.32 ***	0.03	<b>1.00</b>	-	-	-	-	-	-	-	-
Laste olemasolu	-0.14 ***	-0.08 ***	0.46 ***	<b>1.00</b>	-	-	-	-	-	-	-
Lugemis oskus tööl	0.47 ***	-0.02	0.18 ***	-0.14 ***	<b>1.00</b>	-	-	-	-	-	-
Kirjutamis oskus tööl	0.30 ***	-0.03	0.06 **	-0.10 ***	0.51 ***	<b>1.00</b>	-	-	-	-	-
Tervis	-0.13 ***	0.04	0.06 ***	-0.04	-0.11 ***	0.07 ***	<b>1.00</b>	-	-	-	-
Tunnid nädalas	0.35 ***	-0.23 ***	0.11 ***	-0.14 ***	0.33 ***	0.30 ***	0.02	<b>1.00</b>	-	-	-
Kas USAs sündinud	0.02	-0.03	0.04 **	-0.02	0.03	0.03	0.03	0.01	<b>1.00</b>	-	-
Inglise kl kirjaoskus	-0.12 ***	-0.11 ***	0.04	-0.04 **	-0.12 ***	0.08 ***	0.10 ***	-0.01	0.37 ***	<b>1.00</b>	-
Inglise kl lugemisoskus	-0.11 ***	-0.09 ***	0.03	-0.05 **	-0.12 ***	0.07 ***	0.11 ***	0.00	0.36 ***	0.78 ***	<b>1.00</b>

Allikas: PIAAC (2021), autori poolt koostatud

## Lisa 4. Modifitseeritud OLS mudel

	Koefitsient	Standardviga	P-väärtus	
Konstant	1.99973	0.0916140	3.90e-93	***
sugu	-0.179412	0.0226586	4.36e-15	***
vanus	0.0494546	0.00420538	9.85e-31	***
<u>Rass:</u>				
Ladina-Ameerika	-0.0378710	0.0371789	0.3085	
Afroameeriklane	-0.119871	0.0320892	0.0002	***
Teisest rassist	0.0346048	0.0357953	0.3338	
<u>Hariduse tase:</u>				
Keskharidus	0.185655	0.0440355	2.62e-05	***
Kutseharidus	0.296299	0.0571877	2.47e-07	***
Rakenduskõrgharidus	0.330992	0.0517128	2.01e-10	***
Kõrgharidus - bakalaureusekraad	0.532472	0.0481150	1.60e-27	***
Kõrgharidus – magistri- või doktorikraad	0.732224	0.0534428	1.43e-40	***
Töötunnid nädalas	0.0182353	0.00361309	4.98e-07	***
Töötunnid nädalas ruut	-0.0001722	4.18971e-05	4.15e-05	***
<u>Sektorid:</u>				
Avalik sektor	-0.0306919	0.0267757	0.2519	
Mittetulundusühing	-0.0764349	0.0390248	0.0503	*
Lugemisoskus tööl	0.0742802	0.0159658	3.54e-06	***
Kirjutamisoskus tööl	0.0325534	0.0131063	0.0131	**
<u>Elupaik:</u>				
Äärelinn	0.0630914	0.0259164	0.0150	**
Väikelinn	-0.137077	0.0349489	9.13e-05	***
Maakoht	-0.0460272	0.0346278	0.1840	

Determinatsioonikordaja: 0,412189

Mudeli p-väärtus: 9,3e-198

Kasutatud on kohandatud standardvigu

RESET test for specification -

Null hypothesis: specification is adequate

Test statistic:  $F(2, 1669) = 2.22445$

with p-value =  $P(F(2, 1669) > 2.22445) = 0.108448$

White's test for heteroskedasticity -

Null hypothesis: heteroskedasticity not present

Test statistic: LM = 283.922

with p-value =  $P(\text{Chi-square}(177) > 283.922) = 5.90046e-07$

Test for normality of residual -

Null hypothesis: error is normally distributed

Test statistic: Chi-square(2) = 20.8656

with p-value = 2.94508e-05

Allikas: PIAAC (2021), autori poolt koostatud

## Lisa 5. Modifitseeritud OLS mudeli multikollineaarsuse aruanne

sugu	1.096
vanus	1.075
<u>Rass:</u>	
Ladina-Ameerika	1.115
Afroameeriklane	1.101
Teisest rassist	1.056
<u>Hariduse tase:</u>	
Keskharidus	5.153
Kutseharidus	2.445
Rakenduskõrgharidus	3.150
Kõrgharidus - bakalaureusekraad	5.089
Kõrgharidus – magistri- või doktorikraad	3.990
Töötunnid nädalas	12.814
Töötunnid nädalas ruut	12.545
<u>Sektorid:</u>	
Avalik sektor	1.141
Mittetulundühing	1.094
Lugemisoskus tööl	1.494
Kirjutamisoskus tööl	1.423
<u>Elupaik:</u>	
Äärelinn	1.326
Väikelinn	1.221
Maakoht	1.332

Allikas: PIAAC (2021), autori poolt koostatud

## Lisa 6. Valgenahaliste OLS mudel

	Koefitsient	Standardviga	P-väärtus	
Konstant	1.96989	0.100510	8.53e-74	***
sugu	-0.192900	0.0280473	1.01e-11	***
vanus	0.0472345	0.00503006	3.22e-20	***
<u>Hariduse tase:</u>				
Keskharidus	0.292847	0.0480479	1.50e-09	***
Kutseharidus	0.468448	0.0660875	2.39e-12	***
Rakenduskõrgharidus	0.458901	0.0578038	4.91e-15	***
Kõrgharidus - bakalaureusekraad	0.627800	0.0508578	6.41e-33	***
Kõrgharidus – magistri- või doktorikraad	0.851717	0.0598460	2.22e-42	***
Töötunnid nädalas	0.0170278	0.00380854	8.58e-06	***
Töötunnid nädalas ruut	-0.000159171	4.27563e-05	0.0002	***
<u>Sektorid:</u>				
Avalik sektor	-0.0612015	0.0331764	0.0653	*
Mittetulundühing	-0.0713634	0.0510288	0.1622	
Lugemisoskus töö	0.0874747	0.0201567	1.55e-05	***
Kirjutamisoskus töö	0.0110012	0.0163840	0.5021	
<u>Elupaik:</u>				
Äärelinn	0.0788755	0.0331735	0.0176	**
Väikelinn	-0.142346	0.0429014	0.0009	***
Maakoht	-0.0409663	0.0414727	0.3235	

Determinatsioonikordaja: 0,401548

Mudeli p-väärtus: 2,1e-139

Kasutatud on kohandatud standardvigu

RESET test for specification -

Null hypothesis: specification is adequate

Test statistic:  $F(2, 1121) = 1.26475$

with p-value =  $P(F(2, 1121) > 1.26475) = 0.282712$



White's test for heteroskedasticity -

Null hypothesis: heteroskedasticity not present

Test statistic: LM = 237.768

with p-value =  $P(\text{Chi-square}(126) > 237.768) = 6.94188e-09$

Test for normality of residual -

Null hypothesis: error is normally distributed

Test statistic: Chi-square(2) = 7.44356

with p-value = 0.0241908

Allikas: PIAAC (2021), autori poolt koostatud

## Lisa 7. Valgenahaliste OLS mudeli multiokollineaarsuse aruanne

sugu	1.122
vanus	1.047
<u>Hariduse tase:</u>	
Keskharidus	6.958
Kutseharidus	2.975
Rakenduskõrgharidus	3.991
Kõrgharidus - bakalaureusekraad	7.194
Kõrgharidus – magistri- või doktorikraad	5.432
Töötunnid nädalas	12.501
Töötunnid nädalas ruut	12.062
<u>Sektorid:</u>	
Avalik sektor	1.157
Mittetulundühing	1.116
Lugemisoskus tööl	1.520
Kirjutamisoskus tööl	1.429
<u>Elupaik:</u>	
Äärelinn	1.432
Väikelinn	1.269
Maakoht	1.417

Allikas: PIAAC (2021), autori poolt koostatud

## Lisa 8. Afroameeriklaste OLS mudel

	Koefitsient	Standardviga	P-väärtus	
Konstant	1.73531	0.262533	3.70e-10	***
sugu	-0.159322	0.0611196	0.0099	***
vanus	0.0554969	0.0113085	1.96e-06	***
<u>Hariduse tase:</u>				
Keskharidus	0.292654	0.123394	0.0187	**
Kutseharidus	0.235651	0.151181	0.1207	
Rakenduskõrgharidus	0.324413	0.141884	0.0233	**
Kõrgharidus - bakalaureusekraad	0.537714	0.143138	0.0002	***
Kõrgharidus – magistri- või doktorikraad	0.710179	0.152543	6.02e-06	***
Töötunnid nädalas	0.0176784	0.00976883	0.0719	*
Töötunnid nädalas ruut	-0.000183171	0.000123326	0.1391	
<u>Sektorid:</u>				
Avalik sektor	0.0175042	0.0684207	0.7984	
Mittetulundühing	-0.109197	0.107576	0.3113	
Lugemisoskus töö	0.0463375	0.0438945	0.2925	
Kirjutamisoskus töö	0.102651	0.0354019	0.0042	***
<u>Elupaik:</u>				
Äärelinn	0.00599327	0.0693014	0.9312	
Väikelinn	-0.110727	0.0913076	0.2267	
Maakoht	-0.0735420	0.137929	0.5945	

Determinatsioonikordaja: 0,384273

Mudeli p-väärtus: 1,67e-13

RESET test for specification -

Null hypothesis: specification is adequate

Test statistic:  $F(2, 190) = 0.536776$

with p-value =  $P(F(2, 190) > 0.536776) = 0.585514$

Test for normality of residual -

Null hypothesis: error is normally distributed

Test statistic:  $\text{Chi-square}(2) = 13.3582$

with p-value = 0.0012569

White's test for heteroskedasticity -

Null hypothesis: heteroskedasticity not present

Test statistic:  $LM = 102.739$

with p-value =  $P(\text{Chi-square}(121) > 102.739) = 0.884016$

Allikas: PIAAC (2021), autori poolt koostatud

## Lisa 9. Afroameeriklaste OLS mudeli multikollineaarsuse aruanne

sugu	1.090
vanus	1.112
<u>Hariduse tase:</u>	
Keskharidus	4.403
Kutseharidus	2.332
Rakenduskõrgharidus	2.999
Kõrgharidus - bakalaureusekraad	3.052
Kõrgharidus – magistri- või doktorikraad	2.687
Töötunnid nädalas	13.064
Töötunnid nädalas ruut	12.944
<u>Sektorid:</u>	
Avalik sektor	1.183
Mittetulundühing	1.181
Lugemisoskus tööl	1.566
Kirjutamisoskus tööl	1.561
<u>Elupaik:</u>	
Äärelinn	1.262
Väikelinn	1.275
Maakoht	1.214

Allikas: PIAAC (2021), autori poolt koostatud

## Lisa 10. Lihtlitsents

### Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks<sup>1</sup>

Mina Kelly Künnapas

1. Annan Tallinna Tehnikaülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Sooline ja rassiline palgalõhe Ameerika Ühendriikides“,

mille juhendaja on Heili Hein,

1.1 reprodutseerimiseks lõputöö säilitamise ja elektroonse avaldamise eesmärgil, sh Tallinna Tehnikaülikooli raamatukogu digikogusse lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2 üldsusele kättesaadavaks tegemiseks Tallinna Tehnikaülikooli veebikeskkonna kaudu, sealhulgas Tallinna Tehnikaülikooli raamatukogu digikogu kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. Olen teadlik, et käesoleva lihtlitsentsi punktis 1 nimetatud õigused jäävad alles ka autorile.

3. Kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest ning muudest õigusaktidest tulenevaid õigusi.

---

11.05.2021

---

<sup>1</sup> Lihtlitsents ei kehti juurdepääsupiirangu kehtivuse ajal vastavalt üliõpilase taotlusele lõputööle juurdepääsupiirangu kehtestamiseks, mis on allkirjastatud teaduskonna dekaani poolt, välja arvatud ülikooli õigus lõputööd reprodutseerida üksnes säilitamise eesmärgil. Kui lõputöö on loonud kaks või enam isikut oma ühise loomingulise tegevusega ning lõputöö kaas- või ühisautor(id) ei ole andnud lõputööd kaitsvale üliõpilasele kindlaksmääratud tähtajaks nõusolekut lõputöö reprodutseerimiseks ja avalikustamiseks vastavalt lihtlitsentsi punktidele 1.1. jq 1.2, siis lihtlitsents nimetatud tähtaja jooksul ei kehti.