

Tartu Ülikool

Matemaatika-informaatikateaduskond

Matemaatilise statistika instituut

Toitumise seosed lipoproteiinidega
Tartu Ülikooli Eesti Geenivaramu andmestikus:
korreleeritud andmete mitmemõõtmeline analüüs

Bakalaureusetöö

Koostaja: Kristi Helekivi

Juhendaja: Krista Fischer

Tartu 2013

Sisukord

Sissejuhatus	4
1. Ülevaade andmetest	5
1.1. Andmete pärinevus ja liigitus	5
1.2. Andmete kirjeldus.....	6
1.2.1. Taustatunnused	6
1.2.2. Toitumist iseloomustavad tunnused	7
1.2.3. Lipoproteiinid	8
2. Statistiline metoodika	12
2.1. Korrelatsioonistruktuuri uurimine	12
2.1.1. Seosed kahe tunnuse vahel	12
2.1.2. Korrelatsioonimaatriks	13
2.2. Lineaarne regressioonanalüüs.....	13
2.3. Mitmemõõtmeline analüüs	14
2.3.1. Peakomponentanalüüs	15
2.3.2. Faktoranalüüs	16
2.4. Struktuurivõrrandite mudelid	18
3. Andmete uurimine	20
3.1. Korrelatsiooniseosed lipoproteiinide vahel	20
3.2. Korrelatsiooniseosed toiduainete vahel	22
3.3. Lineaarsed seosed toitumistunnuste ja lipoproteiinide vahel	23
3.4. Peakomponentanalüüs	28
3.4.1. Lipoproteiinide peakomponentanalüüs	28
3.4.2. Toiduainete seosed lipoproteiinide peakomponentidega.....	32
3.5. Faktoranalüüs.....	37
3.5.1. Lipoproteiinide faktoranalüüs	37
3.5.2. Toiduainete seosed lipoproteiinide faktoritega	40
4. Struktuurivõrrandite mudelid	44
Kokkuvõte	50
Summary.....	51
Kasutatud kirjandus	52

Lisad	53
Lisa 1. Lipoproteiinid koos inglisekeelse nimetusega ja lipoproteiinide jaotust kirjeldavad histogrammid	53
Lisa 2. Kasutatud statistilistel meetodikatel saadud tulemused toiduainete lõikes	70
Lisa 3. Lipoproteiinide korrelatsioonid esimese kuue peakomponendiga	75
Lisa 4. Lipoproteiinide korrelatsioonid kuue faktoriga	77

Sissejuhatus

Antud bakalaureusetöös on Tartu Ülikooli Eesti Geenivaramu andmete põhjal vaadeldud võimalikke seoseid inimeste toitumisharjumuste ja metabooloomika-tunnuste vahel. Metaboliidid on metabolismi ehk ainevahetuse käigus tekkinud ühendid. Tavaliselt mõeldakse metaboliidi all madalmolekulaarseid ühendeid [1].

Tartu Ülikooli Eesti Geenivaramu andmebaasis on ligikaudu 10 000 geenidoonori kohta olemas detailsed metabooloomika-andmed, mille hulgas on üle 70 tunnuse, mis iseloomustavad eri tüüpi lipoproteiinide ja neis sisalduvate lipiidide kontsentratsiooni. Lisaks on teada andmed geenidoonorite toitumiseelistuste kohta ning doonorite sugu, vanus ja kehamassiindeks.

Töö esimeses peatükis on kirjeldatud kasutatavate andmete tausta ning on antud ülevaade olulisemate tunnuste jaotusest, kasutades graafilisi meetodeid. Teises peatükis on toodud ülevaade kasutatavast statistilisest meetodikast. Kolmandas ning neljandas peatükis rakendatakse metoodikat Tartu Ülikooli Eesti Geenivaramu andmetele.

Lipoproteiinide andmestiku dimensionaalsuse vähendamiseks rakendatakse andmetele esmalt peakomponentanalüüsi ning seejärel faktoranalüüsi. Samuti uuritakse toitumistunnuste seoseid nii üksikute lipoproteiinidega kui ka lipoproteiinide analüüsimisel saadud peakomponentide ja faktoritega. Lõpuks on katsetatud struktuurivõrrandite mudelite metoodikat, mis võimaldab ühe protseduuri käigus nii andmestiku dimensionaalsuse vähendamist kui ka seosestruktuuri uurimist.

Bakalaureusetöö kirjutamiseks on kasutatud tekstitöötlusprogrammi Microsoft Word 2010. Analüüsid on läbi viidud statistikapaketiga R. Joonised on tehtud programmidega Microsoft Excel 2010 ja Graphviz 2.30.1. Viited kasutatud allikatele on nurksulgudes.

Autor tänab käesoleva bakalaureusetöö juhendajat Krista Fischerit rohkete nõuannete ja näpunäidete eest.

1. Ülevaade andmetest

1.1. Andmete pärinevus ja liigitus

Tartu Ülikooli Eesti Geenivaramu (edaspidi TÜ geenivaramu) on teadus- ja arendusasutus, mille eesmärk on edendada geeniuringute arengut, koguda teavet Eesti rahvastiku terviseandmete ja päriliku informatsiooni kohta ning rakendada uusimaid uuringutulemusi rahva tervise parandamiseks. [2]

TÜ geenivaramu andmebaasis on ligikaudu 52 000 vabatahtliku geenidoonori andmed, millest varaseimad pärinevad 2002. aasta oktoobrist. Et oleks võimalik uurida seoseid geenide ja haiguste vahel, täidavad kõik geenidoonorid küsimustiku, mis sisaldab isikuandmeid, tervise seisundi kirjeldust ja sugupuuandmeid. Lisaks võetakse igalt geenidoonorilt vereproov, millest eraldatakse DNA, vereplasma ja valged verelibled, mida säilitatakse vedelas lümfosüstis. [3, lk 5]

Käesolevas töös analüüsitakse 9884 geenidoonori andmeid. Antud valim (juhuvalim kõigist TÜ geenivaramu geenidoonoritest) koosneb indiviididest, kelle vereplasma proove analüüsiti Soomes, prof Mika Ala-Korpela laboris, tuumamagnetresonants-meetodil (NMR – *Nuclear Magnetic Resonance*). Tulemusena määrati plasmas 112 metaboliidi kontsentratsioonid. Neist 74 lipoproteiini andmed on kasutusel käesolevas töös.

Lisaks 74 metaboliidile on analüüsitava andmetabelisse valitud veel 21 tunnust geenidoonorite poolt täidetud küsimustikust. Neist 18 on toitumis- ja 3 taustatunnused.

Taustatunnustena on andmetabelis geenidoonori sugu, vanus geenidoonoriks liitumisel ja kehamassiindeks, mis on arvatud kaalu ja pikkuse ruudu suhtena.

Geenidoonorite toitumisharjumusi kirjeldab 14 toiduaine – kartuli, riisi-makaronide, pudru-müsli-helveste, piimatoodete, kala, liha, lihaproduktide (vorstid, viinerid), värskete juurviljade, keedetud juurviljade, värskete puuviljade-marjade, kompottide-keediste, maiustuste, karastusjookide, munade – tarbimine ühe nädala jooksul, päevade lõikes. Doonoritelt küsiti, kui tihti olid nad viimasel nädalal tarbinud nimetatud toite. Vastusevariantideks olid „mitte kordagi“, „1-2 päeval“, „3-5 päeval“ ja „6-7 päeval“.

Lisaks kirjeldavad toitumismustrit kohvi ja tee joomine ning leiva ja saia söömine päevas. Viimatinimetatud neli tunnust on kodeeritud järgnevalt: produkti mittetarbimine – „1“,

produkti tarbimine ühe korra (üks tass, üks viil) päevas – „2“, produkti tarbimine kaks kuni neli korda (viilu) päevas – „3“, produkti tarbimine viis või rohkem korda (tassi, viilu) päevas – „4“.

1.2. Andmete kirjeldus

1.2.1. Taustatunnused

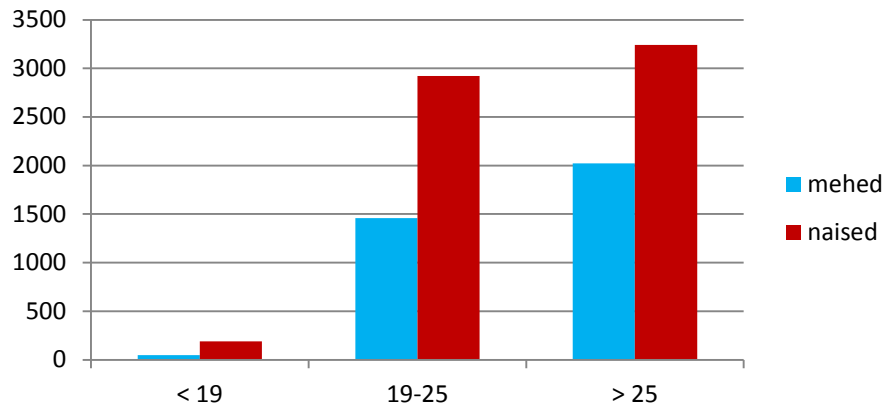
TÜ geenivaramu andmestikus, mille põhjal hakkasin uurima metaboliitide mõju tervisele, on kokku 9884 indiviidi. Neist mehi 3528, mis moodustab 35,7% kõigist analüüsitavas andmetabelis olevatest geenidoonoritest, ning naisi 6356 ehk 64,3% koguarvust (tabel 1).

Tabel 1. Sagedustabel soo ja vanuse järgi

	Sugu		Vanus liitumisel (aastates)							
	Mees	Naine	17-26	27-36	37-46	47-56	57-66	67-76	77-86	87-103
Indiviidide arv	3528	6356	1786	1716	1794	1815	1358	1000	383	32
Osakaal (protsentides)	35,7	64,3	18,1	17,4	18,1	18,4	13,7	10,1	3,9	0,3
Kokku	9884		9884							

Noorim geenidoonor oli andmete kogumise hetkel 17-aastane, vanim 103-aastane. Kusjuures nii noorim kui ka vanim geenidoonor olid naissoost. Andmetabelis olevate geenidoonorite keskmine vanus on 46 aastat (meestel 45 aastat, naistel 46 aastat).

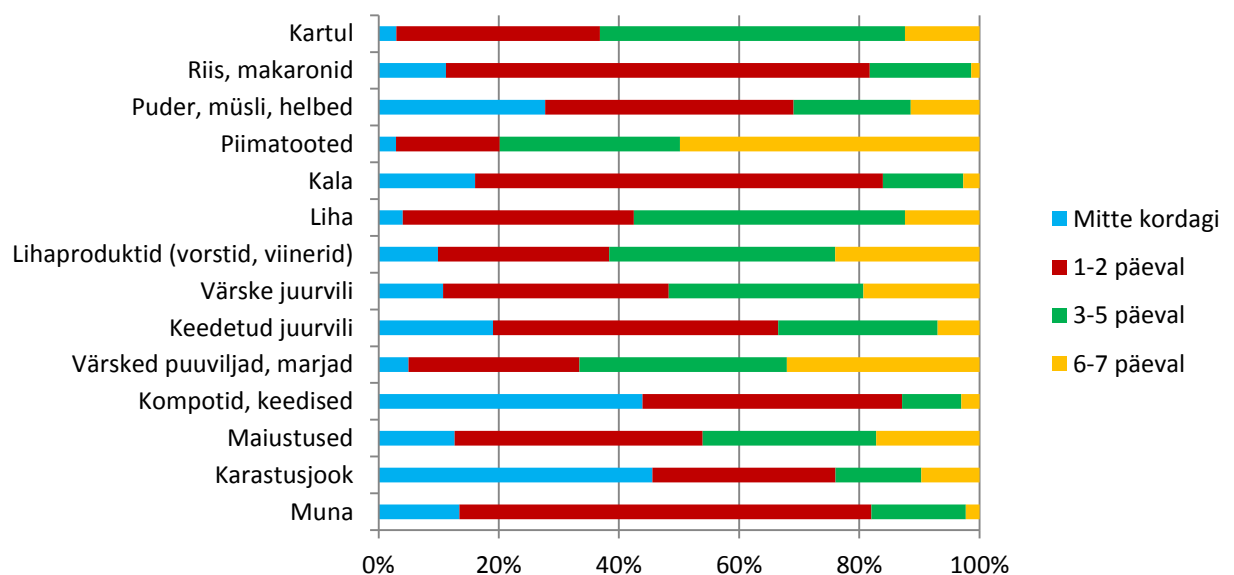
Geenidoonorite keskmine kehamassiindeks (KMI) oli 26,5 ühikut, mis kaldub pigem ülekaalu kui normaalkaalu poole [4]. Üle poolte (5263 indiviidi ehk 53,2%) andmestikus olevate geenidoonorite KMI ületab soovitusliku normaalkaalu piiri (25 ühikut). Alakaalulisi (KMI alla 18,5 ühiku) geenidoonoreid oli andmestikus 239 indiviidi ehk 2,4% kõigist geenidoonoritest. Normaalkaalu piiridesse jäävaid isikuid oli 4382 ehk 44,3% geenidoonorite KMI jääb soovitusliku 18,5–25 ühiku piiresse. Nii meeste kui ka naiste keskmine KMI on sarnane ning sarnaneb ka geenidoonorite keskmise KMI-ga, olles meeste puhul 26,6 ühikut ning naiste korral 26,4 ühikut (joonis 1).



Joonis 1. Geenidoonorite kehamassiindeksi jaotus soo järgi

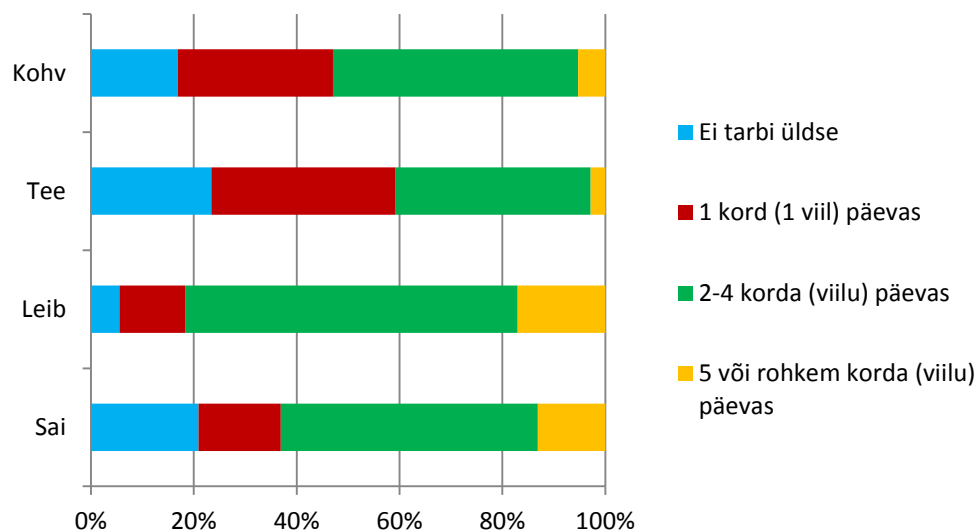
1.2.2. Toitumist iseloomustavad tunnused

Joonisel 2 on näha erinevate toiduainete tarbimise sagedus ühe nädala jooksul. Selgub, et geenidoonorid söövad kartulit 3–5 päeval (50,8%) ning riisi või makarone 1–2 päeval (70,5%) nädalas. Kala süüakse 1–2 päeval nädalas (67,8%). Piimatoodete söömine on geenidoonorite hulgas samuti levinud – vaid 2,9% andmetabelis olevatest isikutest ei tarbi piimatooted üldse. Võrreldes teiste tabelis olevate toiduainetega tarbitakse harvemini kompotte ja keediseid (vastusevariant „mitte kunagi“ oli 43,9% vastanutel) ning karastusjooke (vastusevariant „mitte kunagi“ 45,6% vastanutel).



Joonis 2. 14 toiduaine tarbimise protsendiline jaotus nädala lõikes

Joonis 3 annab ülevaate, kui sageli tarbisid geenidoonorid kohvi, teed, leiba ja saia ühe päeva jooksul. Ilmneb, et kohvi ei tarbi üldse 16,8% geenidoonoritest, seevastu 77,9% ehk üle kolmveerandi geenidoonoritest joob üks kuni neli tassi kohvi päevas. Tee mittetarbijaid oli 23,4%. Üle poolte geenidoonoritest (64,6%) söövad päevas 2–4 viilu leiba. 5,5% ei tarbi leiba üldse. Saia ei tarbi üldse viiendik (20,5%) vastanutest. Täpselt pooled (50,0%) söövad saia 2-4 viilu päevast.



Joonis 3. Protsendiline jaotus kohvi, tee, leiva ja saia tarbimisest ühe päeva jooksul

1.2.3. Lipoproteiinid

TÜ geenivaramu metabooloomika-andmete hulgas on 74 tunnust, mis iseloomustavad eri tüüpi lipoproteiinide kontsentratsiooni. Lipoproteiinid on valgu ehk proteiini ja lipiidi kompleksid, mis tagavad mitmete erinevate rasvamolekulide (sh kolesterooli) liikumise vereringesse [5].

Suuruse järjekorras, suuremast ja hajusamast väiksemaks ja tihkemaks jagunevad antud töös käsitletavatid tunnused nelja suurde gruppi – väga madala tihedusega lipoproteiinid (VLDL – *very low-density lipoprotein*), vahepealse tihedusega lipoproteiinid (IDL – *intermediate-density lipoprotein*), madala tihedusega lipoproteiinid (LDL – *low-density lipoprotein*) ja suure tihedusega lipoproteiinid (HDL – *high-density lipoprotein*). Gruppide sees on lipoproteiinid omakorda liigitunud nii osakese suuruse (XXL – erakordselt suured, XL – väga suured, L – suured, M – keskmise suurusega, S – väiksed, XS – väga väiksed) kui ka

täpsemate omaduste järgi. Näiteks „L_VLDL_C“ (*total cholesterol in large VLDL*) tähistab kogukolesterooli kontsentratsiooni suurtes VLDL-lipoproteiinides, „M_HDL_FC“ (*free cholesterol in medium HDL*) aga nn vaba kolesterooli taset keskmise suurusega HDL-lipoproteiinides.

VLDL-lipoproteiinid on lipiidide transportijateks inimese kehas [6]. TÜ geenivaramu andmetabelis on VLDL-tüüpi lipoproteiinide liike kõige enam – 31 tükki kuues erinevas suurusklassis (XXL, XL, L, M, S, XS). VLDL-lipoproteiine loetakse halva kolesterooli näitajaks, kuna nende kõrget taset seostatakse kõrgema südamehaiguste- ja insuldiriskiga [6].

IDL-kolesterool ei ole tavaliselt veres tuvastatav ning seda tüüpi lipoproteiine peetakse n-ö vaheühendiks VLDL- ja LDL-lipoproteiinide vahel [7]. Ka analüüsitavas andmetabelis on IDL-osakesi vaid nelja liiki ning grupisiselt neid suuruse järgi ei eristata.

Terminiga „halb kolesterool“ viidatakse tavaliselt eelkõige LDL-kolesteroolile, kuna just seda seostatakse koronaarse ateroskleroosiga, mille tagajärjeks on pärgarterite lubjastumine ja kitsenemine [8]. Andmetabelis on LDL-tüüpi kolesterooli osakesi 14 liiki kolmes erinevas suuruses (L, M, S).

HDL-lipoproteiine, mida on andmetabelis 22 liiki neljas erinevas suuruses (XL, L, M, S), loetakse seevastu „heaks kolesterooliks“, kuna nende kõrge tase tundub aitavat ennetada südame-veresoonkonnahaigusi. [8]

Lisaks on andmetabelis olemas ka tunnused „IDL_C“, „LDL_C“ ja „HDL_C“, millega on tähistatud vastavat tüüpi lipoproteiini kogukolesterooli kontsentratsioon. Näiteks tunnus „ILD_C“, kui IDL-tüüpi kolesterooli kontsentratsioon, sisaldab informatsiooni selle grupi nelja tunnuse kohta. VLDL-lipoproteiinidest on kogukolesterool eraldi välja toodud suurusklassides L, M ja S (tunnused „L_VLDL_C“, „M_VLDL_C“ ja „S_VLDL_C“).

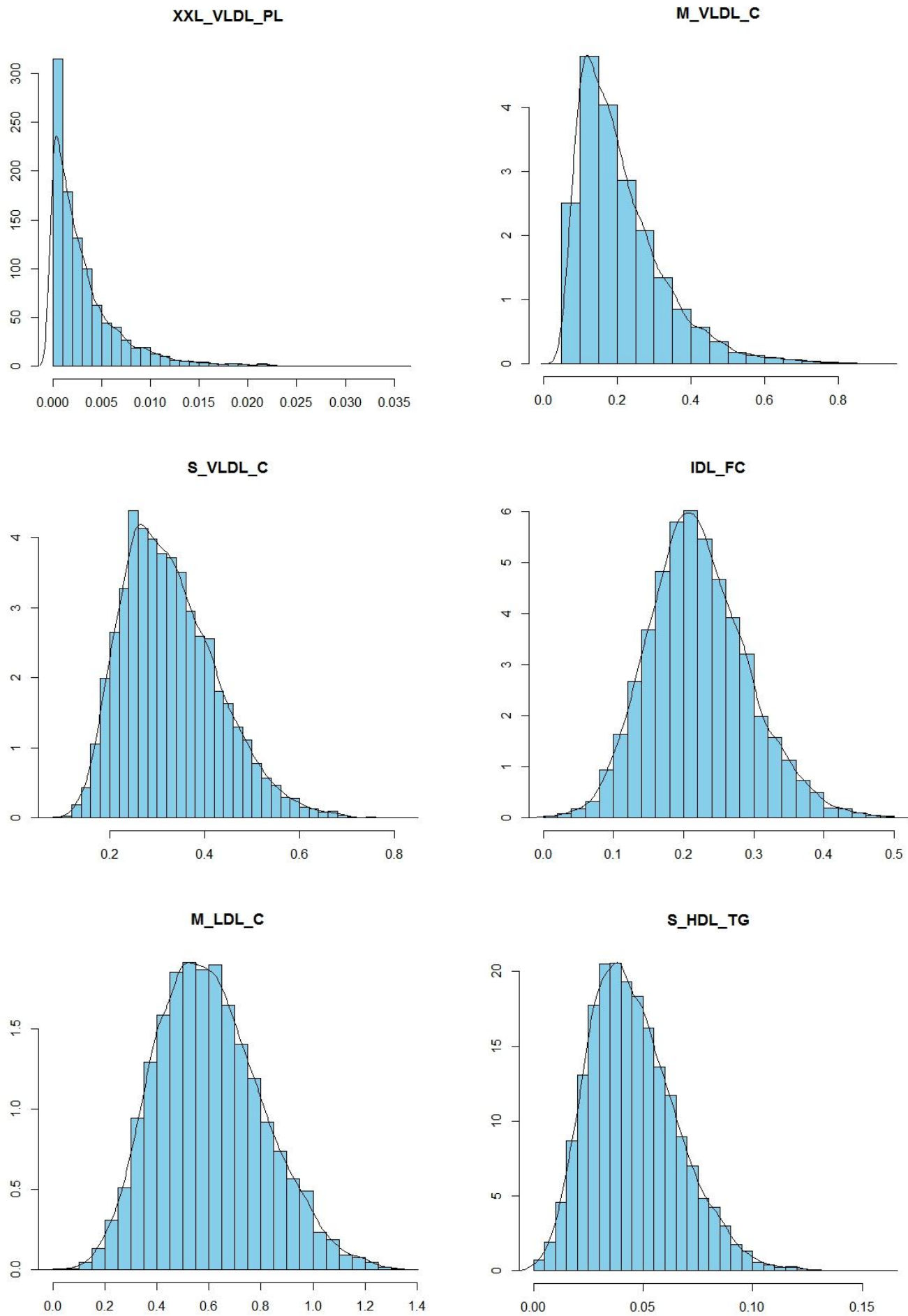
Tunnuse „kogukolesterool“ olemasolu annab meile võimaluse võrrelda, kas gruppidesisene jagunemine suuruse ja omaduste järgi annab lipoproteiinide kirjeldavusele midagi juurde või suudab kogukolesterooli tunnus ise lipoproteiinide grupi omadused piisavalt hästi ära kirjeldada.

Järgnevalt on välja toodud lipoproteiinide jaotusi kirjeldavad histogrammid (joonis 4). Joonisel on kujutatud lipoproteiini jaotuse histogramm koos vastava tihedusfunktsiooniga. Vertikaalsel teljel on tihedus ning horisontaalsel lipoproteiini suurusjärk. Kõik lipoproteiinid koos jaotuse histogrammiga ja inglisekeelse nimetusega on välja toodud lisas 1.

Erakordselt suurte, väga suurte ja suurte VLDL-lipoproteiinide jaotuse kuju sarnaneb eksponentsiaaljaotusega (tunnus „XXL_VLDL_PL“ joonisel 4), keskmise suurusega, väikeste ja väga väikeste VLDL-lipoproteiinide kuju aga Poissoni jaotusega (tunnused „M_VLDL_FC“ ja „S_VLDL_FC“ joonisel 4). Poissoni jaotusega sarnanevad ka kõik HDL-lipoproteiinide suurusklassid (tunnus „S_HDL_TG“ joonisel 4), välja arvatud keskmise suurusega HDL-lipoproteiinid.

IDL-lipoproteiinide, LDL-lipoproteiinide, keskmise suurusega HDL-lipoproteiinide ning kokkuvõtvate tunnuste „IDL_C“, „LDL_C“ ja „HDL_C“ jaotuse kuju sarnaneb normaaljaotusega (tunnused „IDL_FC“ ja „M_LDL_C“ joonisel 4).

Jaotuste erinev kuju tuleneb sellest, et normaaljaotusega tunnus „kogukolesterool“ avaldub teiste lipoproteiinide summana. Sisuliselt on kõigi lipoproteiinide kontsentratsioonid saadud sarnase mõõtmise tulemusena ning andmete teisendamise, et viia kõik jaotused samale kujule, ei ole siinkohal põhjendatud.



Joonis 4. Lipoproteiinide jaotusi kirjeldavad histogrammid

2. Statistiline metoodika

2.1. Korrelatsioonistruktuuri uurimine

2.1.1. Seosed kahe tunnuse vahel

Kahe tunnuse vahelise statistilise sõltuvuse kirjeldamiseks kasutatakse enamasti korrelatsioonikordajaid. Sobivaima korrelatsioonikordaja valik sõltub tunnuste tüübist ja sellest, millist tüüpi seose olemasolu soovitakse testida. Et analüüsida erinevate lipoproteiinide kui arvuliste tunnuste vahelisi seoseid ja nende tugevust, kasutame nendevaheliste seoste kirjeldamiseks lineaarset ehk Pearsoni korrelatsioonikordajat. Toitumistunnuste puhul on tegemist järjestustunnustega, seega nendevahelise seose leidmisel on antud töös kasutusel Spearmani korrelatsioonikordaja.

Lineaarne ehk Pearsoni korrelatsioonikordaja, mis kirjeldab kahe juhusliku suuruse X ja Y vahelise lineaarse seose tugevust, määratakse valemiga

$$r_p = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}},$$

kus n on objektide arv andmetabelis, X_i tunnuse X i -s mõõtmistulemus, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ tunnuse X aritmeetiline keskmine, Y_i tunnuse Y i -s mõõtmistulemus ning $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ tunnuse Y aritmeetiline keskmine. [9]

Spearmani korrelatsioonikordaja, mis mõõdab monotoonse suunatud sõltuvuse tugevust ning lähtub mõõtmistulemuste astakutest, arvutatakse valemiga

$$r_s = 1 - 6 \frac{\sum_{i=1}^n (p_i - q_i)^2}{n(n^2 - 1)},$$

kus n on objektide arv andmetabelis ning p_i ja q_i on vastavate tunnuste, mille vahel seost arvutatakse, i -nda objekti väärtuse astak. [9]

Sellisel määratud korrelatsioonikordaja r_{XY} väärtus asub -1 ja 1 vahel, $-1 \leq r \leq 1$, kusjuures tunnustevahelise kasvava seose korral on korrelatsioonikordaja positiivne ($r > 0$) ning kahaneva seose korral negatiivne ($r < 0$). Kui tunnused x ja y on funktsionaalses seoses kujul $y = a + bx$, siis on nendevahelise korrelatsioonikordaja absoluutväärtus võrdne ühega,

$|r| = 1$. Kui tunnused on sõltumatud, on nende korrelatsioonikordaja võrdne nulliga, $r = 0$. Valimi korrelatsioonikordaja r on hinnanguks üldkogumi korrelatsioonikordajale ρ , mis iseloomustab lineaarse seose tugevust üldkogumis

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

[9]

2.1.2. Korrelatsioonimaatriks

Soovides uurida seoseid korruga rohkem kui kahe tunnuse vahel, leiame iga kahe tunnuse vahelise Pearsoni korrelatsioonikordaja ning koondame tulemused ühisesse tabelisse, mida nimetatakse korrelatsioonimaatriksiks. [9]

Korrelatsioonimaatriksit kasutades on võimalik välja selgitada juhusliku vektori komponente ja komponentide rühmi, mis on ülejäänutega nõrgalt korreleeritud ning leida üksteist „dubleerivad“, st tugevasti korreleeritud komponente. [10, lk 69]

Teoreetiline korrelatsioonimaatriks p -dimensionaalse vektori $(X_1, \dots, X_p)^T$ jaoks on esitatav kujul

$$P = \begin{pmatrix} \rho_{X_1X_1} & \cdots & \rho_{X_1X_p} \\ \vdots & \ddots & \vdots \\ \rho_{X_pX_1} & \cdots & \rho_{X_pX_p} \end{pmatrix}.$$

Empiiriline korrelatsioonimaatriks, mis on arvutatav vaatlustulemuste põhjal, esitub kujul

$$R = \begin{pmatrix} r_{X_1X_1} & \cdots & r_{X_1X_p} \\ \vdots & \ddots & \vdots \\ r_{X_pX_1} & \cdots & r_{X_pX_p} \end{pmatrix}.$$

[11, lk 86-87]

2.2. Lineaarne regressioonanalüüs

Iseloomustamaks, kas ja kuidas argumenttunnuste muutumine toob kaasa funktsioontunnuse muutumise, kasutatakse sageli lineaarset mudelit. Tunnustevahelise seose olemasolu korral on lähtetunnuste abil võimalik prognoosida funktsioontunnuse oodatavat väärtust. [12]

Lineaarne regressioonanalüüsi mudel i -ndale elemendile y_i esitub kujul

$$y_i = \beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + \dots + \beta_k \cdot x_{ki} + \varepsilon_i,$$

kus $x_{1i}, x_{2i}, \dots, x_{ki}$ on mittejhuslikud argumenttunnused, $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ mudeli kordajad ning $\varepsilon_i \sim N(0, \sigma^2)$ i -nda elemendi juhuslik viga (sõltumatud $\forall i$ korral). Tundmatute parameetrite vektori $\vec{\beta} = (\beta_0, \beta_1, \dots, \beta_k)$ väärtust saab hinnata, kasutades vähimruutude meetodit, see tähendab, minimiseerides tekkivate prognoosivigade ruutude summat. Vähimruutude meetodil saadakse hinnangud $\widehat{\beta}_1, \widehat{\beta}_2, \dots, \widehat{\beta}_k$ mudeli parameetritele ja nende standardvigadele: $\hat{s}(\widehat{\beta}_1), \hat{s}(\widehat{\beta}_2), \dots, \hat{s}(\widehat{\beta}_k)$. Iga parameetri β_j ($j = 1, \dots, k$) korral saab testida hüpoteese

$$\begin{cases} H_0: \beta_j = 0 \\ H_1: \beta_j \neq 0 \end{cases}$$

kasutades teststatistikuna nn t-statistikut $\hat{t} = \frac{\widehat{\beta}_j}{\hat{s}(\widehat{\beta}_j)}$, mis on asümptootilise t-jaotusega. [12]

2.3. Mitmemõõtmeline analüüs

Tihti on objektidel mõõdetud palju omadusi, mida kõiki soovitakse uurida. Leitud tunnused on tähtsad objekti kirjeldajad, kuid kuna neid on palju, ei ole lihtne objekte nende vahel omavahel võrrelda. Mitmemõõtmeline analüüs on oluline statistikavaldkond, mida kasutatakse, kui on vaja analüüsida üheaegselt paljudest tunnustest koosnevat tunnusterühma. Sageli on mitmemõõtmelise analüüsi eesmärgiks andmete dimensionaalsuse vähendamine.

Edasises teoorias eeldame, et soovime uusi tunnuseid konstrueerida, lähtudes m algsest tunnusest. Sellisel juhul on üldkogumimudeliks juhuslik vektor $\vec{X} = (X_1, \dots, X_m)^T$ oma mitmemõõtmelise jaotusega. Andmestik, st m tunnuse vaatlusandmed, on vektori \vec{X} realisatsioonid. Vektori \vec{X} jaotust iseloomustavate karakteristikutena kasutame keskvaartusvektorit $E\vec{X}$ ning dispersioonimaatriksit

$$D\vec{X} = \Sigma = E \left[(\vec{X} - E\vec{X})(\vec{X} - E\vec{X})^T \right], \quad m \times m. \quad [13]$$

Sageli tehakse andmete dimensionaalsuse vähendamiseks olemasolevate tunnuste baasil üks või paar uut tunnust ning püütakse objekte uurida (üksteisest eristada) uute tunnuste valguses. [13]

Võimalike seoste kirjeldamiseks ja selgitamiseks lipoproteiinide vahel on antud töös järgnevalt läbi viidud peakomponent- ning faktoranalüüs.

2.3.1. Peakkomponentanalüüs

Peakkomponentide analüüsis leitakse uued tunnused, peakkomponendid, lähtetunnuste lineaarkombinatsioonina. Andmestikust, milles on m tunnuse vaatlusandmed, leitakse peakomponendid P_i ($i = 1, \dots, m$) nii, et komponendil P_1 on maksimaalne võimalik dispersioon, komponendil P_2 suuruselt järgmine dispersioon jne. Saavutatakse olukord, kus väike arv esimesi peakomponente kirjeldab ära peaaegu kogu lähtetunnuste summaarse dispersiooni. [13]

Juhusliku vektori $\vec{X} = (X_1, \dots, X_m)^T$ peakomponentide vektori $\vec{P} = (P_1, \dots, P_m)^T$ esimene element avaldub lineaarkombinatsioonina $P_1 = \vec{\alpha}^T \vec{X}$, kus $\vec{\alpha} = (\alpha_1, \dots, \alpha_m)^T$ valitakse selliselt, et DP_1 on maksimaalne. Nõutud on kordajate normeeritus: $\vec{\alpha}^T \vec{\alpha} = 1$. [13]

Olgu dispersioonimaatriksi Σ omaväärtused $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ ja vastavad omavektorid $\vec{\gamma}_1, \vec{\gamma}_2, \dots, \vec{\gamma}_m$. Teooriast on teada, et esimest peakomponenti määravaks kordajate vektoriks $\vec{\alpha}$ on dispersioonimaatriksi Σ suurimale omaväärtusele λ_1 vastav omavektor $\vec{\gamma}_1$. [13]

Teine peakomponent on defineeritud kui $P_2 = \vec{\gamma}_2^T \vec{X}$. Mittekorreleeritus esimese peakomponendiga tagatakse nõudega, et omavektorid on ortogonaalsed. $DP_2 = \lambda_2$ on maksimaalne, kui peakomponentide defineerimiseks kasutada omavektoreid. Lähtetunnuste kogudispersioon on võrdne peakomponentide kogudispersiooniga:

$$DX_1 + DX_2 + \dots + DX_m = \text{tr}(\Sigma) = \sigma_{11} + \sigma_{22} + \dots + \sigma_{mm} = \lambda_1 + \lambda_2 + \dots + \lambda_m,$$

iga järgmise peakomponendi dispersioon on maksimaalne võimalik. Seega valem i -nda peakomponendi defineerimiseks on

$$P_i = \vec{\gamma}_i^T \vec{X},$$

ja tema dispersioon on

$$DP_i = \vec{\gamma}_i^T \Sigma \vec{\gamma}_i = \lambda_i.$$

Peakomponendi tähtsuse näitajaks on osakaal

$$\frac{\lambda_i}{\sum_{i=1}^m \lambda_i}.$$

Analüüsis kasutatavaks peakomponentide arvu määramiseks vaadatakse omaväärtusi – ühest suuremad või omavahel tugevalt erinevad omaväärtused viitavad sobivale arvule peakomponentidele. [13]

2.3.2. Faktoranalüüs

Faktoranalüüsi üheks eesmärgiks on samuti andmete dimensionaalsuse vähendamine – väikese arvu mittemõõdetavate tunnuste, faktorite, leidmine nii, et lähtetunnuseid saab ligikaudselt, aga võimalikult hästi faktorite kaudu esitada. Kui peakomponentanalüüsis defineeriti peakomponendid lähtetunnuste lineaarsete kombinatsioonidena, siis faktoranalüüsis eeldatakse, et iga lähtetunnus on esitatav väikese arvu faktorite lineaarse kombinatsioonina. Kuna ühised faktorid ei kirjelda lähtetunnuseid täielikult, lisatakse kasutuselevõetavasse mudelisse ka juhuslikud vead, mida nimetatakse omapärafaktoriteks. [13]

Faktoranalüüsi mudel esitatakse kujul

$$\vec{X} = A\vec{F} + \vec{\varepsilon},$$

kus \vec{X} : $m \times 1$ on lähtetunnuste vektor;

A : $m \times k$ ($k < m$) on lineaarne teisendusmaatriks, mida nimetatakse faktorkaalude maatriksiks ehk faktorlaadungite maatriksiks ehk faktormaatiksiks;

\vec{F} : $k \times 1$ on uute mittevaadeldud tunnuste ehk nn ühiste faktorite vektor (mille korral faktortunnused on standardiseeritud ja sõltumatud);

$\vec{\varepsilon}$: $m \times 1$ on juhuslike vigade ehk omapärade vektor (mille korral omapärad on sõltumatud nii omavahel kui ka sõltumatud faktoritest). Omapära on see osa hajuvusest, mis jääb faktorite poolt kirjeldamata. [13]

Seda osa tunnuse varieeruvusest, mis on kirjeldatud ühiste faktorite poolt ehk tunnuse hajuvuse süstemaatilist osa, mis on määratud iga tunnuse jaoks, nimetatakse tunnuse kommunaliteediks

$$h_i^2 = \sum_{j=1}^k a_{ij}^2.$$

Iga üksiku faktori mõju tunnuse hajuvusele on määratud faktorkaalu ruudu abil. [13]

Tunnuse omapäraks ψ_i nimetatakse seda osa tunnuse varieeruvusest, mis jääb ühiste faktorite poolt kirjeldamata. Seega avaldub tunnuse dispersioon kahe liidetava (süstemaatilise faktorite poolt kirjeldatud osa ning kirjeldamata omapära) abil

$$DX_i = h_i^2 + \psi_i.$$

Tunnuste kommunaliteetide summa annab ühiste faktorite poolt kirjeldatud osa üldvarieeruvusest. Faktormudeli kasulikkust võimaldab kirjeldada selle summa ja üldvarieeruvuse suhe (mis esitatakse tavaliselt protsentides):

$$\frac{\sum_{i=1}^m h_i^2}{tr\Sigma}.$$

[13]

Kasutatavate faktorite arvu k määramiseks on erinevaid kriteeriumeid – näiteks juba peakomponentanalüüsi juures mainitud omaväärtuste kriteerium, mille korral valitakse faktorite arvuks korrelatsioonimaatriksi ühest suuremate omaväärtuste arv. Kui omaväärtused võrduksid kõik ühega, siis kirjeldaksid faktorid samapalju kui tunnused ja selline mudel ei omaks mõtet. Kasutatakse ka nn nõlvakudiagrammi, mille puhul joonistatakse graafik, kus y-teljel on omaväärtused ning x-teljel faktori numbrid. [13]

Faktorlahendi leidmisel on antud töös kasutusel iteratiivne peafaktorite meetod. Nimetatud meetodi rakendamisel valitakse kommunaliteetide alghinnangu määramiseks h_i^2 hinnanguks i -nda tunnuse determinatsioonikordaja (SMC – *Squared Multiple Correlation with other variables*) ehk mitmese korrelatsioonikordaja ruut, mis saadakse, lähendades teda ülejäänud tunnuste lineaarse funktsiooniga. [13]

Iteratiivne peafaktorite meetod lähtub redutseeritud korrelatsioonimaatriksist. Redutseerituks nimetatakse korrelatsioonimaatriksit $\tilde{\rho}$, mille peadiagonaalil on ühed asendatud tunnuste kommunaliteetidega. Kuna kommunaliteetid on tundmatud, kasutatakse korrelatsioonimaatriksi redutseerimiseks nende eelhinnanguid. [13]

Leitakse faktorkaalude hinnangud ning arvutatakse uued kommunaliteetide hinnangud. Seejärel redutseeritakse korrelatsioonimaatriksi uuesti, kasutades peadiagonaalil uusi kommunaliteetide hinnanguid ning leitakse uue maatriksi peakomponentide abil uued faktorkaalude hinnangud. Saadud hinnagute põhjal saab ümber arvutada kommunaliteetid. Saadakse uus redutseeritud korrelatsioonimaatriks. Nii jätkatakse, kuni teatud sammust alates faktorkaalud enam oluliselt ei muutu (eeldusel, et iteratsiooniprotsess koondus). [13]

Praktikas on kasulik, et faktormudeli esitus ei ole ühene – erinevad faktorlahendid võimaldavad lähtetunnuste sisemist struktuuri kirjeldada erineva nurga alt. Selgemalt või huvitavamalt interpreteeritavate faktorite jaoks kasutatakse faktorite pööramist. Pööramise abil soovitakse tihti saavutada, et iga faktoriga oleks seotud ainult teatud hulk omavahel

tugevamalt seotud tunnuseid. Tuntuimaks ortogonaalseks pööramismeetodiks, mida kasutatakse, on *Varimax*. [13]

2.4. Struktuurivõrrandite mudelid

Peakomponent- ja faktoranalüüs võimaldab asendada otseselt mõõdetavad või üldiselt vaadeldavad tunnused mittemõõdetavate ehk latentsete tunnustega, mis avalduvad eelnevate lineaarkombinatsioonidena. Uute tekitatud tunnuste puhul on küsimus nende sisu avamises ja interpreteerimises. [13]

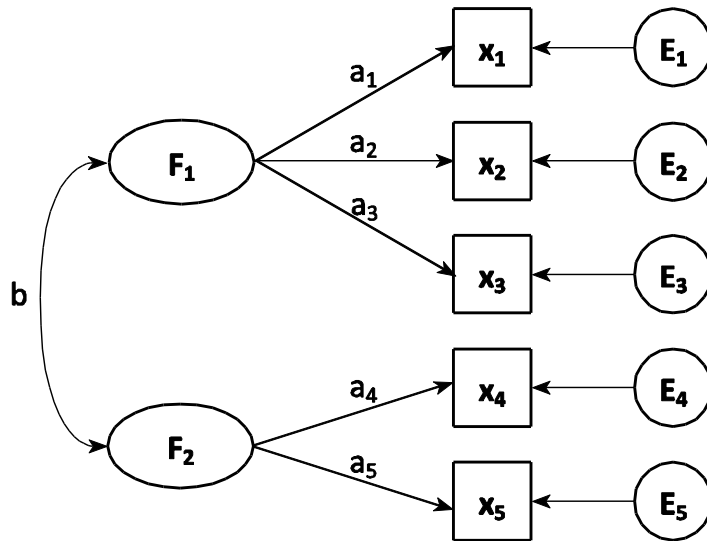
Struktuurivõrrandite mudelite analüüs on mitmese statistilise analüüsi tehnika, mille kasutamisel on kaks põhilist eesmärki – selgitada mudelisse kaasatud tunnuste vahelisi sõltuvusseoseid ning kirjeldada mudeli abil nende tunnuste hajuvusest võimalikult suur osa. [13]

Struktuurivõrrandite mudelite analüüs lähtub teoreetilistel kaalutlustel esitatud hüpoteetilisest struktuurist. Diagramm, mille abil see struktuur esitatakse, ning seosed, mida mudelisse kaasatavate tunnuste vahel arvatakse olevat, on paika pandud enne analüüsi. Vaid seoste tugevused ei ole teada. Parameetrite hinnangute saamiseks rakendatakse suurima tõepära meetodit vaadeldud tunnuste dispersioonimaatriksile. [13]

Struktuurivõrrandite mudelite kirjeldamine algab tavaliselt diagrammi abil, kus näidatakse ära mudelis osalevad tunnused ja nendevahelised seosed. Mudeli esitamiseks kasutatakse kolme liiki elemente: ringe või ovaale, mis kujutavad latentseid ehk mittemõõdetavaid tunnuseid (faktorid, vead); ristkülikuid, mis esindavad mõõdetuid tunnuseid (annavad latentsetele tunnustele sisu) ja nooli, mis eelnevalt nimetatud geomeetrilisi kujundeid ühendavad (määravad struktuuri). Ühendavatele nooltele kantakse kordajad, millest mõned võivad olla enne analüüsi uurija poolt fikseeritud. [13]

Kuna faktorid ei määra tunnuseid täielikult, tuleb skeemile lisada ka tunnuste juhuslikud vead ehk omapärafaktorid. [13]

Joonisel 5 on kujutatud mudel, kus on kaks faktorit (F_1 , F_2) ning viis mõõdetud tunnust (x_1 , x_2 , x_3 , x_4 , x_5). Iga mõõdetud tunnust mõjutab ka juhuslik viga (E_1 , E_2 , E_3 , E_4 , E_5). Faktorite otseseid mõjusid mõõdetavatele tunnustele kujutavad ovaalidest ristkülikutesse suunatud nooled. Joonisel kujutatud viis mõõdetud tunnust on omavahel korreleeritud, kuna nad sõltuvad kahest omavahel korreleeritud latentsest tunnusest.



Joonis 5. Kahe faktoriga struktuurivõrrandite mudel viie mõõdetud tunnusega ning tunnustele vastavate juhuslike vigadega

Teades kordajate a_1, \dots, a_5 ja b väärtuseid saame välja kirjutada joonisele 5 vastavad võrrandid ning leida vastavalt teadaolevatele lähtetunnustele hinnangud latentsetele tunnustele.

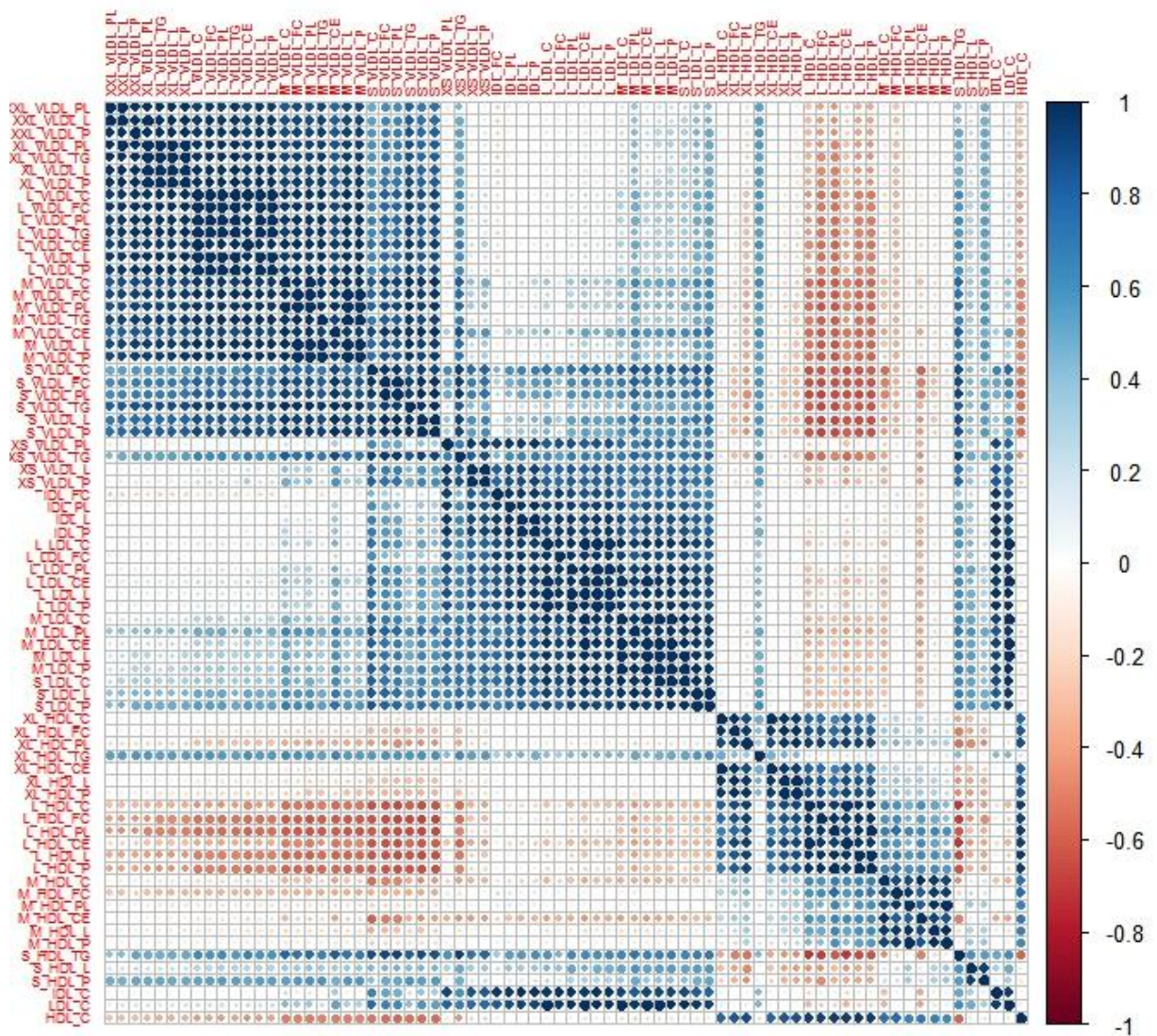
3. Andmete uurimine

3.1. Korrelatsiooniseosed lipoproteiinide vahel

Joonisel 6 on kujutatud lipoproteiinidevahelisi korrelatsiooniseoseid. Tugevad positiivsed korrelatsioonid on märgitud tumesinisega, tugevad negatiivsed korrelatsioonid tumepunasega ning mida nullilähedasem korrelatsioon, seda heledamalt on see kujutatud graafikul.

Vaadates korrelatsioonikordajate väärtusi, näeme, et nelja grupi (VLDL, IDL, LDL, HDL) lipoproteiinide grupisisesed korrelatsioonid on väga tugevad, olles enamasti üle 0,90, sageli isegi üle 0,99. Kõrgeimad korrelatsioonid esinevadki ühe grupi sees sama suurusega lipoproteiinide vahel. Näiteks tunnuste „XL_VLDL_L“ (lipiidide koguarv väga suurtes VLDL-lipoproteiinides) ja „XL_VLDL_TG“ (triglütseriidide arv väga suurtes VLDL-lipoproteiinides) vaheline korrelatsioon on $r_p = 0,9996$. Kõik grupisisesed seosed on positiivsed, mis tähendab, et seosed on kasvavad ehk samapidised.

Seevastu lipoproteiinide nn gruppidevahelised korrelatsioonid on sageli nullilähedased või negatiivselt korreleeritud. Suurim negatiivne korrelatsioon $r_p = -0,7148$ esineb tunnuste „S_HDL_TG“ (triglütseriidid väikestes HDL-lipoproteiinides) ja „L_HDL_C“ (kogukolesterooli kontsentratsiooni suurtes HDL-lipoproteiinides) vahel. Absoluutväärtuselt vähim korrelatsioon $r_p = 0,0003$ on tunnuste „XL_HDL_L“ ja „S_LDL_L“ vahel. Nullilähedane korrelatsioon on oodatav, kuna nagu peatükis 1.2.3 mainitud, siis HDL-tüüpi lipoproteiine loetakse hea kolesterooli näitajateks, LDL-lipoproteiine seostatakse sageli halva kolesterooliga. Kõigi grupivaheliste elementide korrelatsioonid jäävadki küllaltki väikesteks, olles absoluutväärtuselt alla 0,75, enamasti tunduvalt alla 0,5.

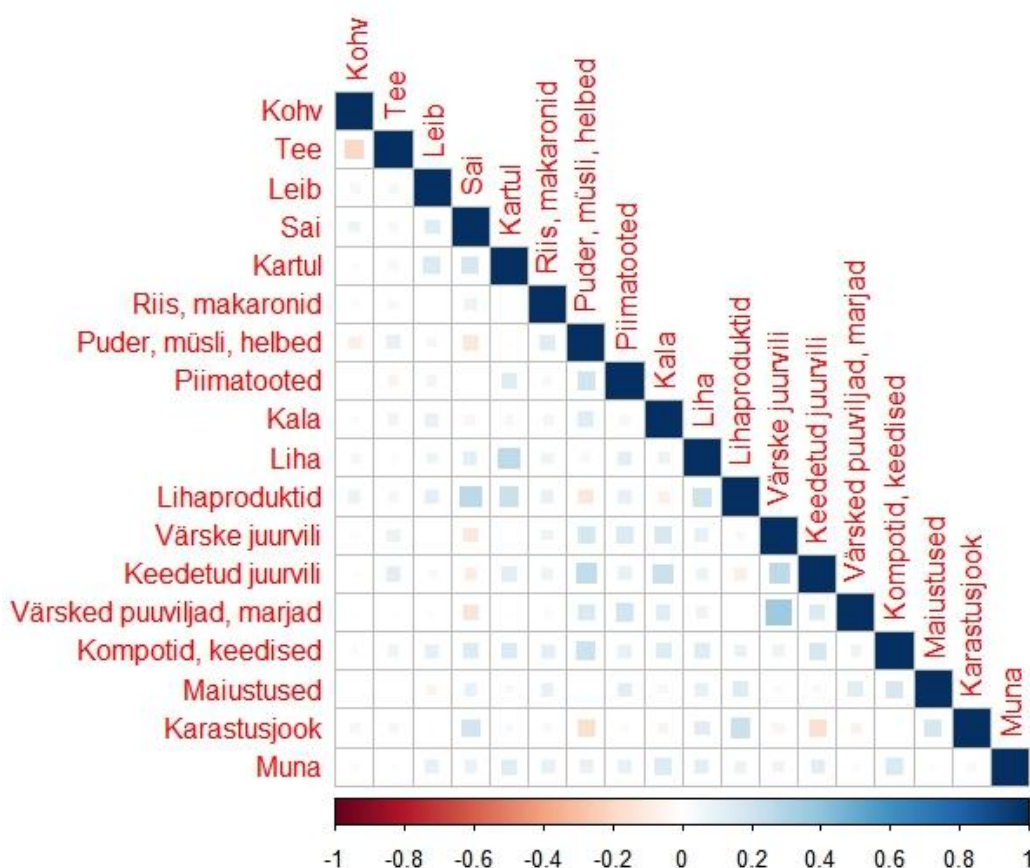


Joonis 6. Lipoproteiinide korrelatsioonid. Tugevad positiivsed korrelatsioonid on märgitud tumesinisega, tugevad negatiivsed korrelatsioonid tumepunasega.

Grupisisesed tunnustevahelised kõrged korrelatsioonid viitavad sellele, et uurides erinevate lipoproteiinide seoseid indiviidi toitumise ja taustatunnuste vahel võib tekkida olukord, kus samast grupist pärit lipoproteiinid üksteist „dubleerivad“ ning seega ei anna analüüsile midagi juurde. Parema interpreteeritavuse huvides oleks mõistlik metabloomika-andmed grupeerida, võttes kasutusele latentsed ehk mittemõõdetavad tunnused, mida kasutades saame tulemusi üldistada kõigile vaatusse all olevatele lipoproteiinidele.

3.2. Korrelatsiooniseosed toiduainete vahel

Erinevate toiduainegruppide vahelised korrelatsiooniseosed (joonis 7) osutusid kõik väga väikesteks, olles valdavalt absoluutväärtuselt alla 0,2. Kõige kõrgem seos esines värskete juurviljade ja värskete puuviljade tarbimise sageduse vahel ($r_s = 0,36$). Seega võib eeldada, et inimesed, kes tarbivad värskaid juurvilju, söövad rohkem ka värskaid puuvilju.



Joonis 7. Toiduainete vahelised korrelatsiooniseosed

Teistest kõrgemad seosed esinesid ka saia ja lihaproduktide ($r_s = 0,27$), värskete ja keedetud juurviljade ($r_s = 0,25$), liha ja kartuli ($r_s = 0,25$), teraviljatoodete (puder-müsli-helbed) ja keedetud juurviljade ($r_s = 0,24$), vorsti ja karastusjookide ($r_s = 0,22$), kala ja keedetud juurviljade ($r_s = 0,22$), lihaproduktide ja liha ($r_s = 0,21$), lihaproduktide ja kartuli ($r_s = 0,21$) ning teraviljatoodete ja moosi ($r_s = 0,21$) söömise sageduse vahel. Kõik mainitud seosed on positiivsed, mis tähendab, et nimetatud seosed on kasvavad ehk samapidised. Seega eelpool mainitud toitade korral toob ühe toidu sagedasem tarbimine kaasa ka teise toidu sagedama tarbimise.

Kõik need seosed on loogilised ja intuiitiivselt tajutavad – kartuleid süüakse sageli koos liha või lihaproduktidega (vorstid, viinerid), kala kõrvale sobivad keedetud juurviljad ning hommikusele pudrule lisatakse moosi. Ometi on seosed statistilise korrelatsiooniseose mõttes väga nõrgad. Võib eeldada, et kasutades neid mitmemõõtmelisel analüüsil toimub kadu informatsioonis ning peakomponent- ja faktoranalüüs ei suuda olemasolevat infot täielikult (mõistliku arvu peakomponentide või faktoritega) ära kirjeldada.

3.3. Lineaarsed seosed toitumistunnuste ja lipoproteiinide vahel

Järgnevalt on kasutatud lineaarset regressioonanalüüsi, et hinnata iga toiduaine tarbimise mõju igale lipoproteiinile.

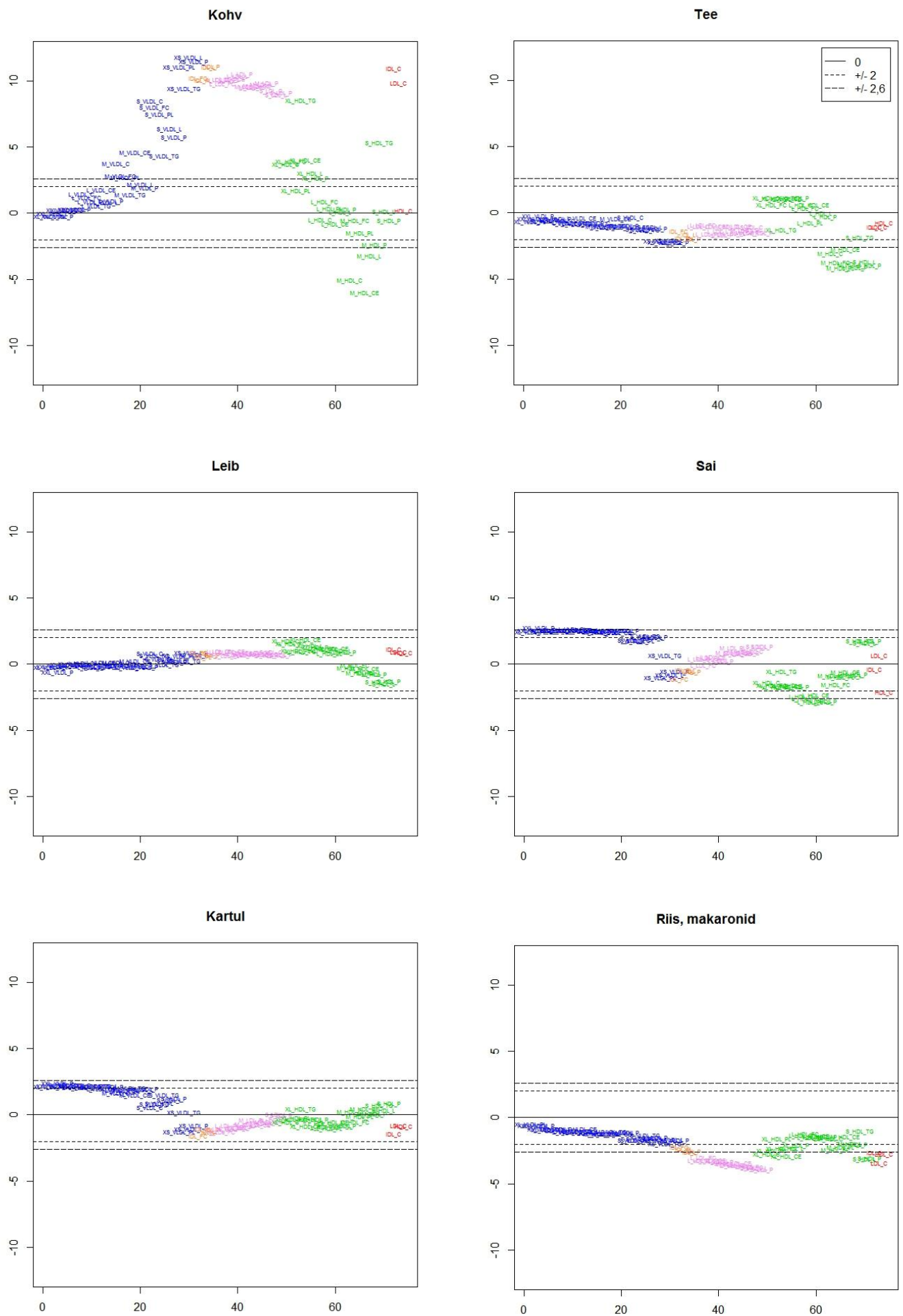
Arvestatud on ka soo, vanuse ja kehamassiindeksi mõju, seega i -ndale lipoproteiinile nmr_i vastav mudel, kus on uuritud näiteks kohvi mõju, on kirjutatav kujul

$$nmr_i = \beta_0 + \beta_1 \cdot kohv + \beta_2 \cdot sugu + \beta_3 \cdot vanus + \beta_4 \cdot KMI + \varepsilon_i,$$

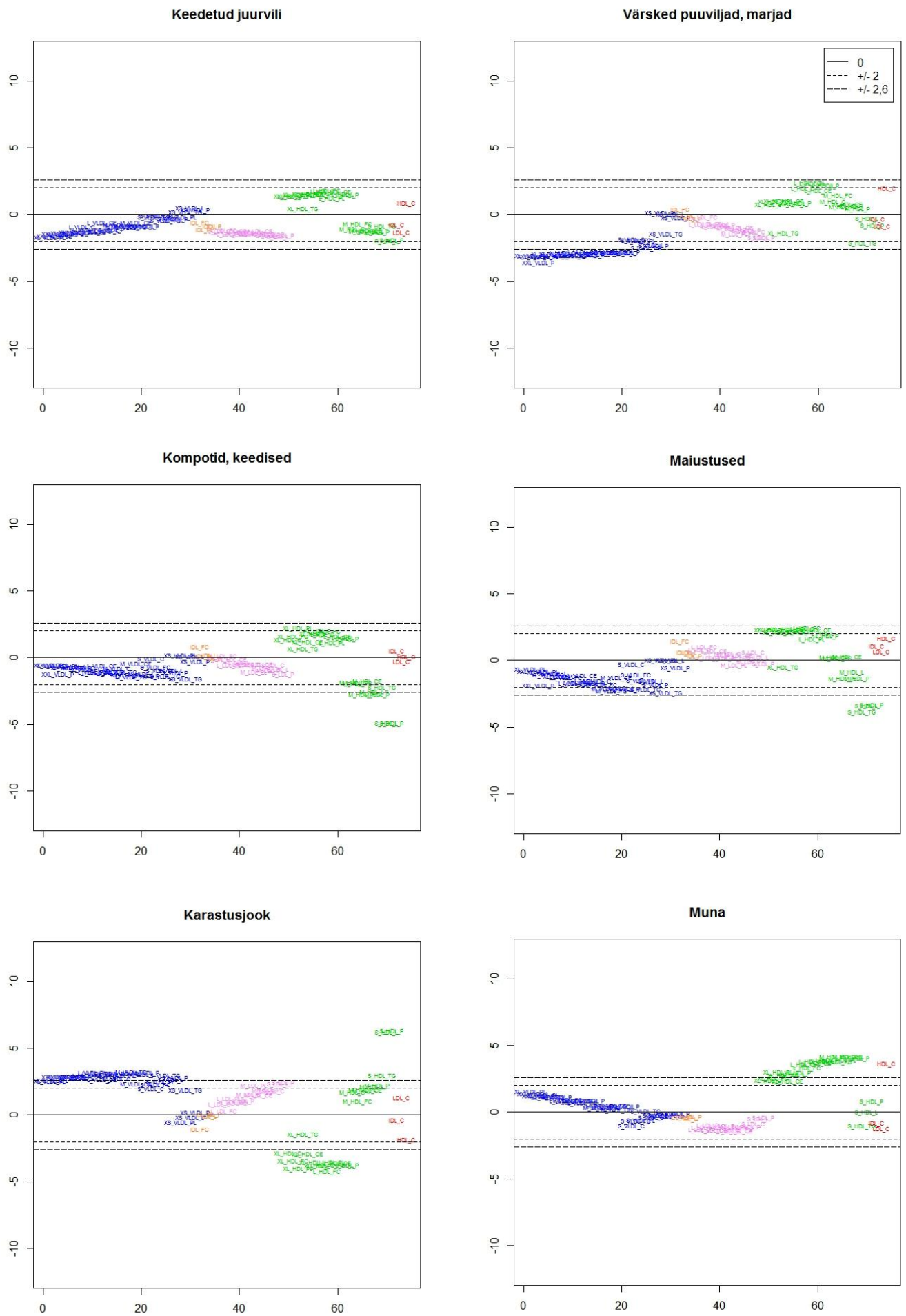
kus $\beta_0, \beta_1, \beta_2, \beta_3$ ja β_4 on mudeli kordajad ning ε_i on i -nda elemendi juhuslik viga. Mudeli hindamisel on salvestatud parameetritele β_1 vastava t -statistiku väärtus iga lipoproteiini korral. Nii on iga toiduaine jaoks saadud 74 t -statistiku väärtust, mis on kantud joonistele 8a–8c.

Horisontaalse musta punktiirjoonega on tähistatud piirid -2 ja $+2$ ning $-2,6$ ja $+2,6$, mis vastavad normaaljaotuse vastavalt 97,5 ja 99,5 kvantiilidele. Seega olulisusenivool $\alpha = 0,05$ loetakse statistiliselt mitteoluliseks seosed lipoproteiinidega, mis jäävad sisemiste piirjoonte vahele. Olulisusenivool $\alpha = 0,01$ on statistiliselt mitteolulised välimiste piirjoonte vahele jäävad seosed. Välja on toodud ka nullpunkt.

Sinise värviga on tähistatud VLDL-, kollasega IDL-, lillaga LDL- ja rohelisega HDL-tüüpi lipoproteiinide grupp. Kokkuvõtvad tunnused IDL_C, LDL_C ja HDL_C on kõigil joonistel paremas servas punase värviga. Lisaks joonistele on tekkinud seosed toitumistunnuste kaupa välja toodud ka lisas 2.



Joonis 8a. Lineaarne regressioonseos lipoproteiini ja toitumistunnuse vahel



Joonis 8c. Lineaarne regressioonseos lipoproteiini ja toitumistunnuse vahel

Joonistelt 8a–8c on näha, et sellised toiduainete grupid nagu leib, kartul ning värsked ja keedetud juurviljad osutusid lipoproteiinide ja toiduainete vahelise lineaarse seose kirjeldamisel olulisusnivool $\alpha = 0,01$ statistiliselt mitteoluliseks – t-statistiku väärtused on kõigil nendel juhtudel absoluutväärtuselt väiksemad kui 2,6. Samuti on näha, et ei esine toiduainet, mis mõjutaks korruga kõiki lipoproteeine.

Kõige tugevamad lineaarsed seosed esinesid tunnuse „kohv“ juures. Nimetatud tunnus on negatiivselt seotud kolme HDL-tüüpi lipoproteiiniga, ent positiivne seos esineb sel tunnusel keskmise, väikse ja väga väikse suurusega VLDL-lipoproteiinidega, kõigi IDL- ja LDL-tüüpi lipoproteiinidega ning mõningate HDL-lipoproteiinidega. Seega kohvi mõjul suurenevad nn halva kolesterooli näitajad VLDL ja LDL ning nendevaheliseks kesktasemeks loetav IDL. Mõju positiivseks loetavalt HDL-lipoproteiinidele on aga negatiivne ehk kahandav.

HDL-tüüpi lipoproteiinidele avaldavad negatiivset mõju ka sai, karastusjoogid, tee, kompotid-keedised ja maiustused. Neist esimesed kaks on samaaegselt positiivses seoses ka VLDL-lipoproteiinidega. Saia, karastusjooke, kompotte-keediseid ja maiustusi peetakse sageli ebatervislikuks nende suure suhkruisalduse ja/või tühjade süsivesikute sisalduse tõttu. Paljud inimesed joovad ka teed suhkruga, see võiks seletada tunnuse „tee“ ja positiivse HDL-kolesterooli vahelist negatiivset seost.

HDL-lipoproteiinidega seostuvad positiivselt eelkõige toitumistunnused „kala“, „liha“ ja „muna“. Lisaks mõjutab tunnus „kala“ samaaegselt VLDL-tüüpi kolesterooli negatiivselt. Ka „värsked puuviljad“ seostusid VLDL-kolesterooliga negatiivselt. Seega võib neid toiduaineid lugeda kasulikuks, kuna nad vähendavad halvaks kolesterooliks loetava VLDL-tüüpi ja/või suurendavad heaks kolesterooliks peetava HDL-tüüpi lipoproteiinide kontsentratsiooni inimese kehas. Lihaproductid mõjutasid head tüüpi HDL-kolesterooli samaaegselt nii positiivses kui ka negatiivses suunas.

LDL-tüüpi kolesterooliga negatiivselt seostuvad ehk nn halva kolesterooli mõju vähendavad „riis-makaronid“, „puder-müsli-helbed“ ja „piimatooted“. Seejuures on kõik kolm tunnust negatiivses seoses ka HDL-lipoproteiinidega, eriti tugevad seosed on teraviljatoodetel (puder-müsli-helbed). Põhjus võib peituda selles, et nimetatud toiduainegrupid on küllalt laialdased – näiteks hommikune piimaga keedetud kiudaineterikas puder on kasulik ning vähendab negatiivse kolesterooli mõju, ent suhkrurikkad hommikusöögihelbed ei tõsta head tüüpi kolesterooli taset.

Lineaarsel regressioonil saadud tulemused kinnitavad arusaamu, et kala ja liha söömine on tervisele kasulikud, maiustusi ja karastusjooke võib seevastu pidada pigem ebatervislikeks toiduaineteks. Huvitav on, et tervislikeks peetavate värskete puu- ja juurviljade ning keedetud juurviljade mõju lipoproteiinidele on statistiliselt ebaoluline või mitte nii tugev kui teistel toiduainetel. Samuti ilmnes, et muna, mida sageli seostatakse kolesterooliga halvas mõttes, suurendab vaid HDL-lipoproteiinide taset. HDL-kolesterooli kõrge tase aitab aga, nagu peatükis 1.2.3 mainitud, ennetada südame-veresoonkonnahaigusi.

3.4. Peakomponentanalüüs

3.4.1. Lipoproteiinide peakomponentanalüüs

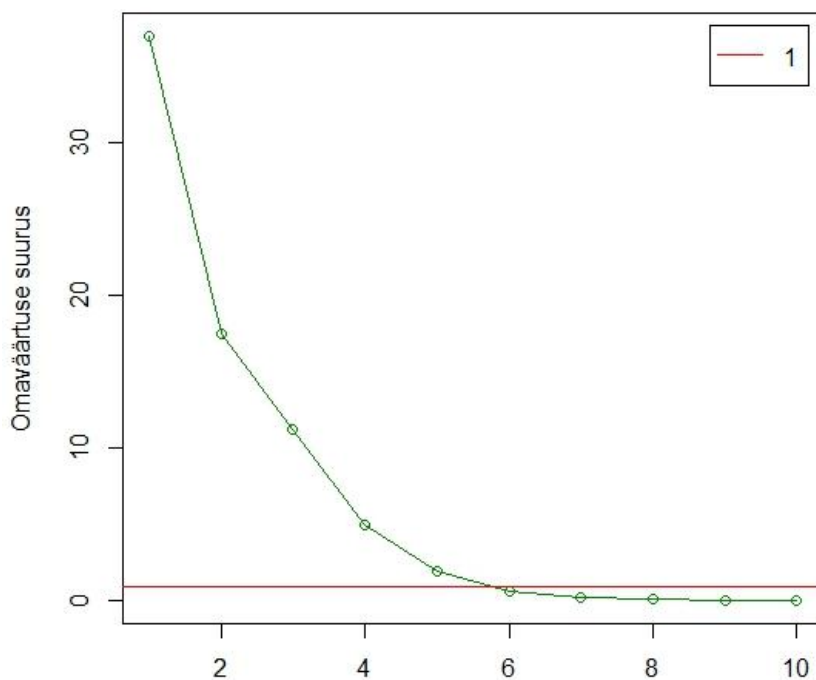
Rakendades peatükis 2.3.1 kirjeldatud peakomponentanalüüsi TÜ geenivaramu lipoproteiinide andmetele, saame tulemuseks, et viiele peakomponendile $P_1 - P_5$ vastavad omaväärtused on suuremad kui 1 (tabel 2). Lipoproteiinide andmestiku esimesed kolm peakomponenti kirjeldavad ära 88,9% koguarieeruvusest, esimesed neli 95,7% ning esimesed viis koguni 98,3% kogudispersioonist.

Kuigi kuues peakomponent P_6 , millele vastav omaväärtus oli väiksem kui üks, ei anna kirjeldavusele enam palju juurde, on ta välja toodud järgnevatel joonistel ning tabelites, et näidata mõju olulisuse muutumist.

Tabel 2. Lipoproteiinide kuus esimest peakomponenti, koos vastava omaväärtuse, standardhälbe, kirjeldavuse osakaalu ja kumulatiivse kirjeldavuse osakaaluga

	P_1	P_2	P_3	P_4	P_5	P_6
Omaväärtus	37,00	17,52	11,23	5,03	1,94	0,66
Standardhälve	6,08	4,19	3,35	2,24	1,39	0,81
Peakomponendi dispersioon (%)	0,5500	0,2368	0,1518	0,0680	0,0262	0,0089
Kumulatiivne osakaal (%)	0,5500	0,7368	0,8887	0,9567	0,9829	0,9917

Joonisel 9 on nn nõlvakudiagrammil kujutatud esimest kümnet omaväärtust. On näha, et esimeste omaväärtuste vahelised vahed on suured, kuid teatud hetkest alates vahed stabiliseeruvad. Graafikul tekib nn „küünarnukk“, millest alates on omaväärtused peaaegu võrdsed väikesed arvud.



Joonis 9. Lipoproteiinide peakomponentide esimesed kümme omaväärtust

Peakomponentide korrelatsioonid lähtetunnustega on aluseks peakomponentide interpreteerimisele. TÜ geenivaramu andmestikus olevate lipoproteiinide korrelatsioonid esimese kuue peakomponendiga on graafiliselt välja toodud joonisel 10 ning tabelis arviliselt esitatud lisa 3.

Kõigil viiel graafikul on horisontaalsel teljel esimene peakomponent ning vertikaalsel teljel järjest peakomponendid $P_2 - P_6$. Sinise värviga on tähistatud VLDL-, kollasega IDL-, lillaga LDL- ja rohelisega HDL-tüüpi lipoproteiinide grupid. Kokkuvõtavad tunnused „IDL_C“, „LDL_C“ ja „HDL_C“ on kujutatud punase värviga.

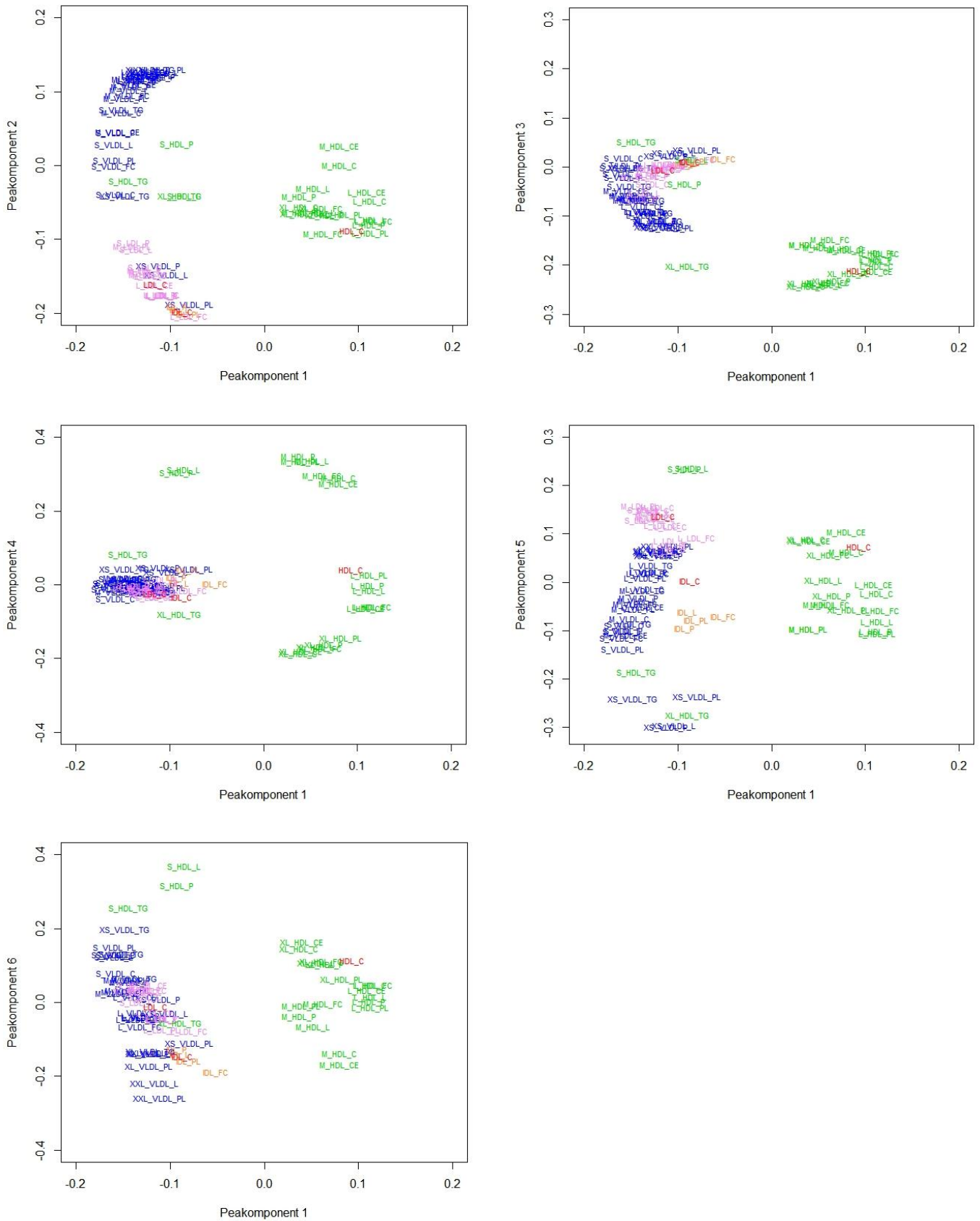
Joonistelt on näha, et esimese peakomponendi määravad lipoproteiinid, kus HDL-tüüpi lipoproteiinide väärtus on suur ning VLDL- ja LDL-lipoproteiinide väärtus väike. Seega esimese peakomponendi kõrge väärtus on geenidoonoritel, kelle puhul positiivseks peetav kolesterool HDL on suur ning negatiivne LDL-tüüpi kolesterool väike. Kõrge HDL-

lipoproteiinide ja madal LDL-lipoproteiinide tase seostuvad väiksema südamehaiguste riskiga. Seega peakomponent iseloomustab terviseriske – geenidoonorid, kelle puhul esimese peakomponendi väärtus on kõrge, on väiksema riskiga; peakomponendi madalama väärtuse korral on inimesel aga suurem risk südamehaigustele.

Teise peakomponendi juures on määravaks suurte VLDL-lipoproteiinide ning väiksed IDL- ja LDL-lipoproteiinide väärtused. HDL-tüüpi lipoproteiinide väärtused teist peakomponenti oluliselt ei mõjuta. Kolmanda peakomponendi juures on määravaks eelkõige HDL-kolesterooli väikesed näitajad (välja arvatud S-suurusega HDL-lipoproteiinid). Neljas peakomponent seostub taas HDL-kolesterooliga, jagades HDL-lipoproteiinid n-õ pooleks – määravaks on keskmiste ja väikese suurusega HDL-lipoproteiinide suured ning ülejäänud HDL-lipoproteiinide väikesed väärtused.

Viienda peakomponendi juures on tähtsad suure väärtusega S-HDL- ja LDL-lipoproteiinid ning väikse väärtusega XS-VLDL-lipoproteiinid. Lisaks jagab see peakomponent VLDL-tüüpi tunnused pooleks, aga väga selgeid seoseid enam esile tuua ei saa. Viies peakomponent annab ka kirjeldavusele juurde vaid 1,2% kogudispersioonist.

Tänu peakomponentide kasutamisele on toimunud informatsiooni kokkusurumine, kus mitme lähtetunnuse info avaldub ühes peakomponendis, andes sellele nime.



Joonis 10. Lipoproteiinide peakkomponentide korrelatsioonid esimese kuue peakomponendiga

3.4.2. Toiduainete seosed lipoproteiinide peakomponentidega

Uurides lipoproteiinide peakomponentanalüüsil saadud viie esimese peakomponendi lineaarseid seoseid toiduainetega, ilmneb, et statistiliselt mitteolulisteks (olulisusnivool 0,01) osutuvad toiduainetegrupid „leib“, „sai“, „kartul“, „värsked juurviljad“, „keedetud juurviljad“ ja „maiustused“ (joonised 11a–11c).

Teiste toiduainete ja lipoproteiinide vaheliste seoste selgitamisel lähtume eelmises peatükis saadud tulemustest, et esimene peakomponent, mille määramist mõjutas kõige rohkem kõrge HDL- ja madal LDL-lipoproteiinide tase, iseloomustab terviseriske. Teine peakomponent seostus vaid nn halba tüüpi VLDL-, IDL- ja LDL-kolesterooliga. Esimesed kaks peakomponenti kirjeldavad ära 73,7% koguvarieeruvusest.

Kõige tugevamad seosed esinevadki esimese ja teise peakomponendi ning tunnuse „kohv“ vahel. Selle tunnuse puhul on seosed kahe esimese peakomponendiga negatiivsed. Seega on rohkem kohvi joovatel indiviididel madal HDL-kolesterooli tase ning kõrgem IDL-, LDL- ja VLDL-kolesterooli tase. Tuginedes eelnevale, saab kohvi tarbimist seeläbi seostada südameveresoontehaiguste riskitegurina.

Esimene peakomponent seostub lisaks kohvile negatiivselt ka karastusjookidega, positiivselt mõjutab seda peakomponenti pudru-müsli-helveste, kala ja värskete puuviljade tarbimine.

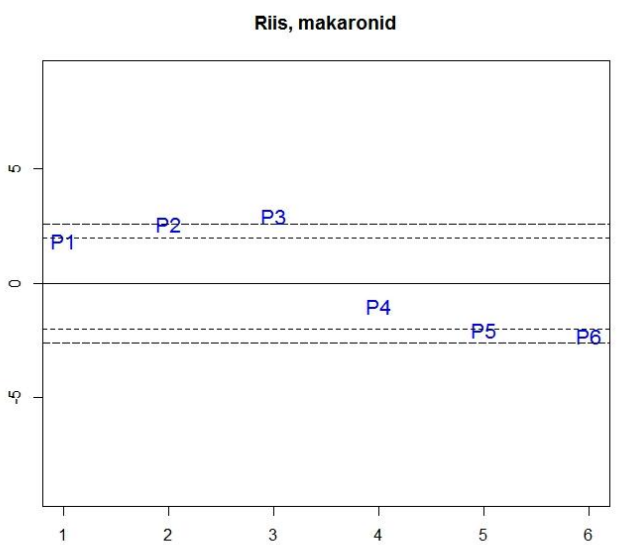
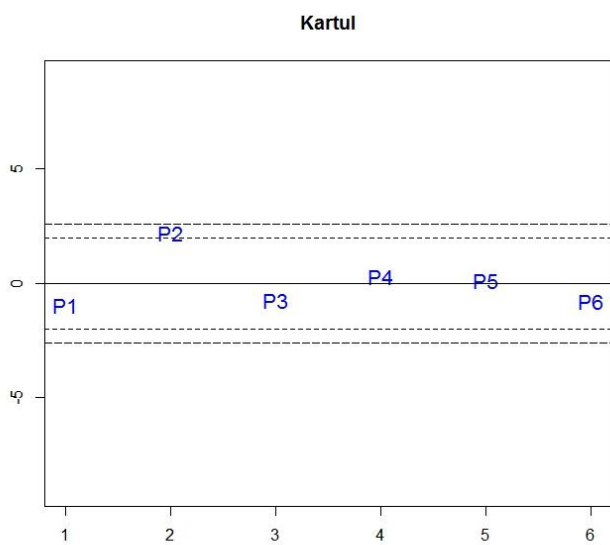
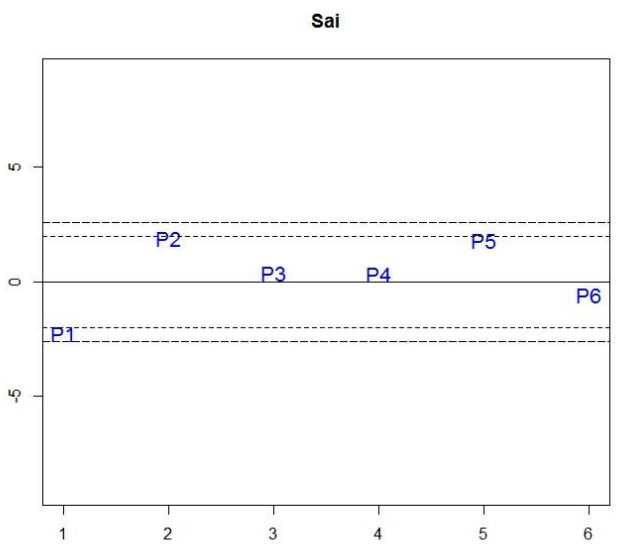
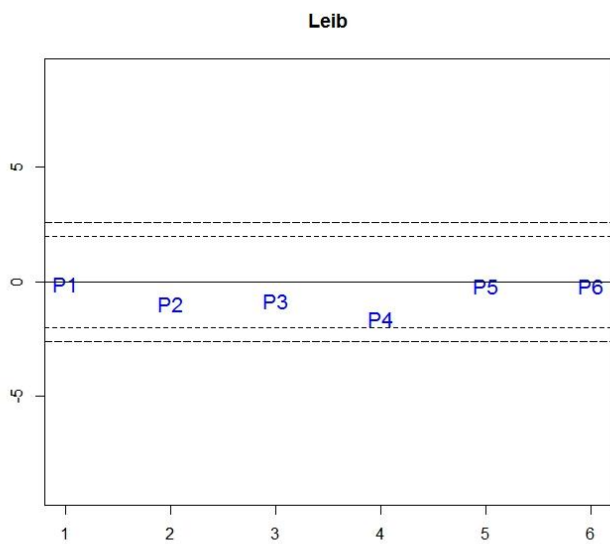
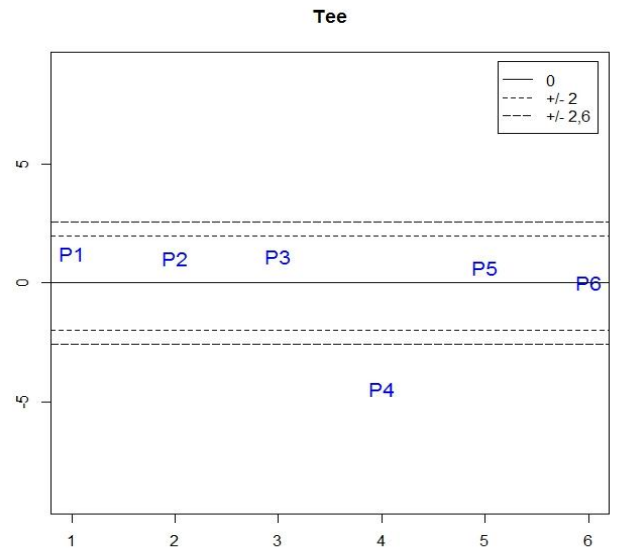
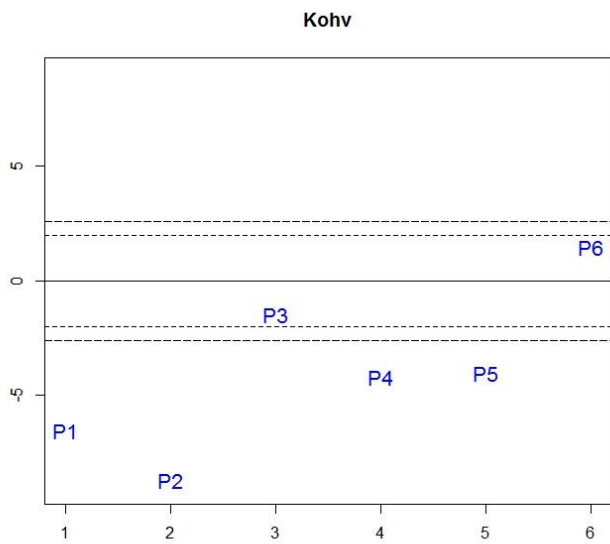
Teine peakomponent on lisaks tunnusele „kohv“ oluline ka „kala“ korral. Mõlema tunnusega seostub peakomponent negatiivselt. Positiivne seos on teisel peakomponendil tunnustega „riis-makaronid“ ja „puder-müsli-helbed“.

Kolmas peakomponent seostub negatiivselt tunnustega „kala“ ja „muna“. Positiivne seos on sellel peakomponendil taas tunnustega „riis-makaronid“ ja „puder-müsli-helbed“. Kolmas peakomponent eristas HDL-tüüpi lipoproteiine, seega seostuvad neli viimatinimetatud toiduainet peakomponentanalüüsist lähtudes hea kolesterooliga.

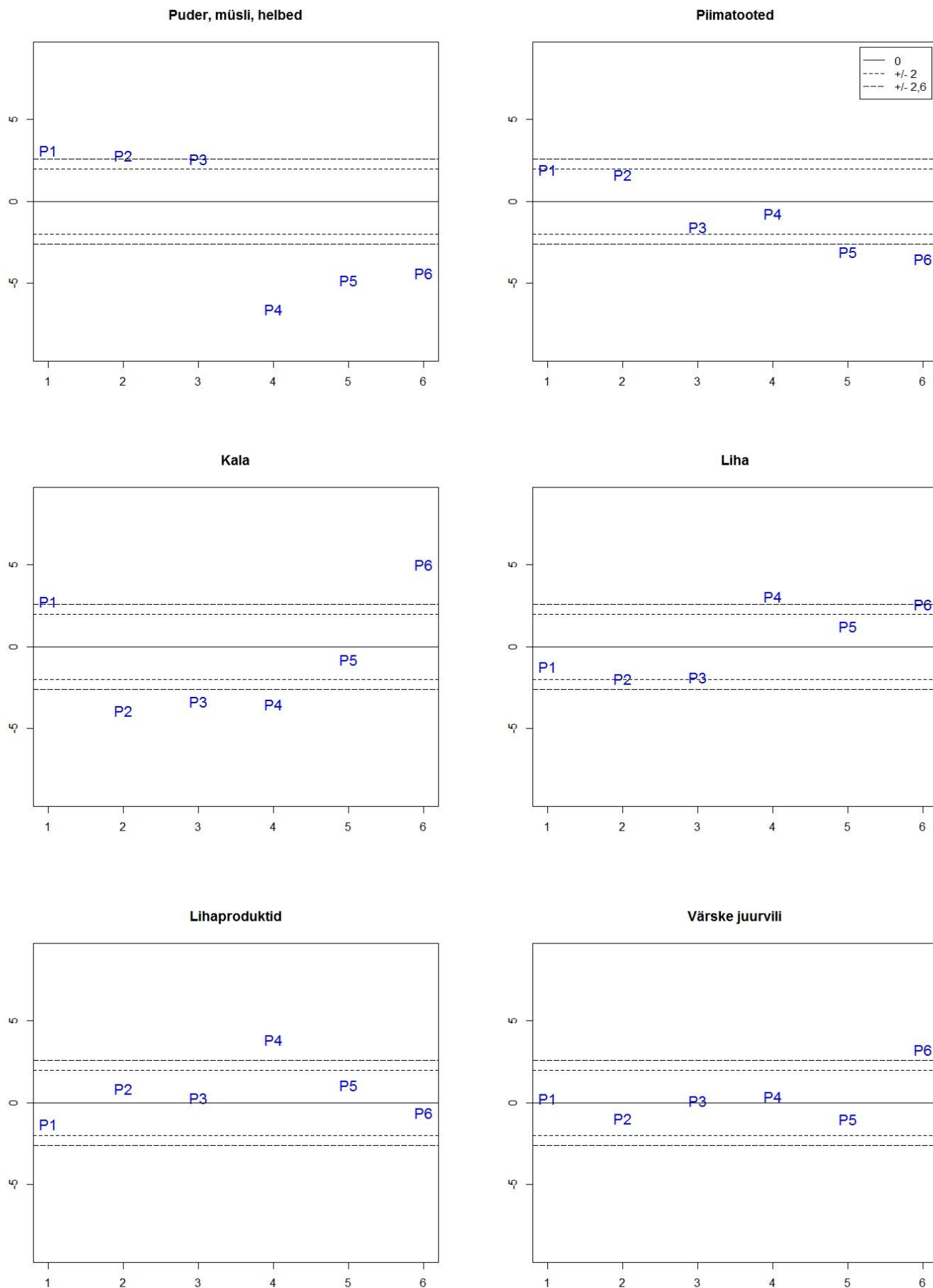
Neljandal peakomponendil, mille määrasid samuti HDL-tüüpi lipoproteiinid, esineb enim statistiliselt olulisi seoseid toiduainetega – see peakomponent on oluline tunnuste „kohv“, „tee“, „puder-müsli-helbed“, „kala“, „kompotid-keedised“ (kõigiga negatiivne seos), „liha“, „lihaproductid“, „karastusjoogid“ (positiivne seos) korral.

Viienda peakomponendiga seostuvad tunnused „kohv“, „puder-müsli-helbed“, „piimatooted“ negatiivselt ning „karastusjoogid“ positiivselt.

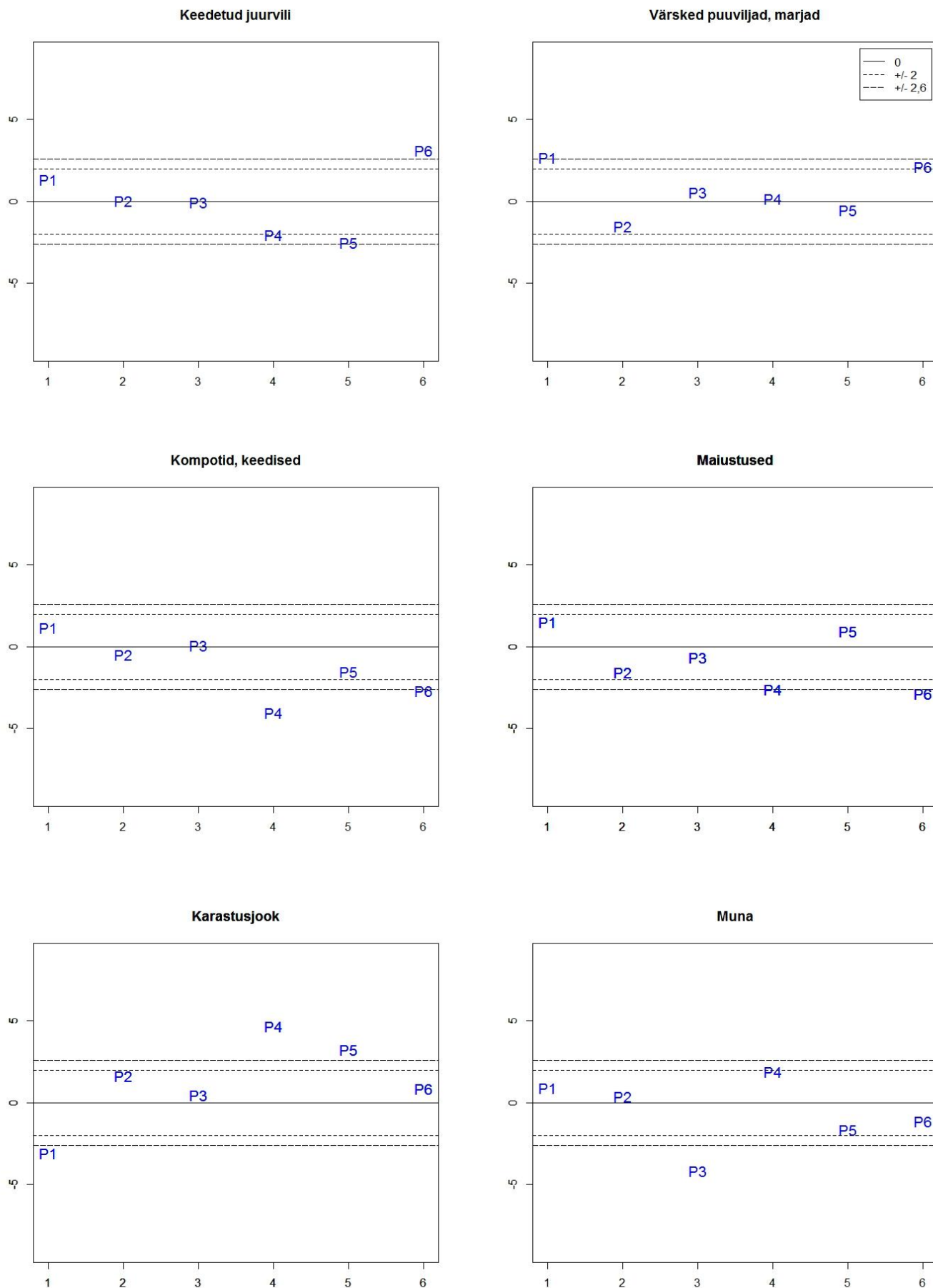
Olulisusnivool 0,01 on statistiliselt olulised peakomponendid toiduainete lõikes välja toodud ka lisas 2.



Joonis 11a. Toitumistunnuste seosed lipoproteiinide peakomponentidega



Joonis 11b. Toitumistunnuste seosed lipoproteiinide peakomponentidega



Joonis 11c. Toitumistunnuste seosed lipoproteiinide peakomponentidega

3.5. Faktoranalüüs

3.5.1. Lipoproteiinide faktoranalüüs

Uurimaks, kuidas lipoproteiinid ühiste tunnuste põhjal faktoritesse jagunevad, kasutame faktoranalüüsi, teostades *Varimax*-pööramise. Esimesele kuuele omaväärtusele vastava faktori kirjeldavusprotsent koos kumulatiivse kogukirjeldavusega on välja toodud tabelis 3. Näeme, et esimesed 5 faktorit kirjeldavad ära 97,7% koguarieeruvusest. Lisades kuuenda faktori tõuseb kirjeldavuse protsent veelgi – koguarieeruvusest on kirjeldatud 98,9%. Samas on kuuenda faktori kirjeldamisvõime vaid 1,2%.

Tabel 3. Faktorite kirjeldamisvõime

	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Faktor 5	Faktor 6
Faktori dispersioon	0,363	0,321	0,173	0,092	0,028	0,012
Kumulatiivne dispersioon	0,363	0,684	0,857	0,949	0,977	0,989

Kasutatav faktorite arv on määratud, lähtudes omaväärtuste kriteeriumist ja rakendustarkvara R võimalustest. Korrelatsioonimaatriksilt saadud omaväärtusi, mis oleksid suuremad kui üks oli viis tükki (joonis 9), ent R-i abil ei olnud võimalik 74 lipoproteiini informatsiooni suruda kokku viide faktorisse. Kaasates analüüsi aga seitse faktorit, muutus seitsmenda faktori kirjeldavus väga väikeseks (0,4% koguarieeruvusest).

Kuue faktori korrelatsioonid andmetabelist olevate lipoproteiinidega on esitatud joonisel 12 ning lisas 4.

Kõigil viiel graafikul on horisontaalsel teljel esimene faktor ning vertikaalsel teljel järjest faktorid $F_2 - F_6$. Sinise värviga on tähistatud VLDL-, kollasega IDL-, lillaga LDL- ja rohelisega HDL-tüüpi lipoproteiinide grupid. Kokkuvõtavad tunnused „IDL_C“, „LDL_C“ ja „HDL_C“ on kujutatud punase värviga.

Joonise 12 graafikutelt on näha, et esimene faktor on tugevamalt korreleeritud lipoproteiinidega, kus VLDL- ja LDL-tüüpi lipoproteiinide väärtused on suured ning HDL-kolesterooli tase väike. Seega on esimese faktori kõrged väärtused geenidonoril, kellel head tüüpi kolesteroolitase on madal ning halvaks peetava kolesterooli näitajad kõrged.

Teise faktori määramisel on olulised suured IDL- ja LDL- ning väiksed VLDL-tüüpi lipoproteiinide väärtused. Järelikult seostub see faktor vaid nn halba tüüpi kolesterooliga. Kolmanda faktori määravad ära HDL-lipoproteiinide ehk positiivse kolesterooli suured väärtused. Neljas faktor on positiivselt korreleeritud keskmise ja väikese suurusega (M ja S) HDL-lipoproteiinidega. Viies faktor jagab kaheks VLDL-lipoproteiine. Kuuenda faktori puhul aga selget gruppi, mida see faktor iseloomustama sobiks, enam ei eristu.

3.5.2. Toiduainete seosed lipoproteiinide faktoritega

Sarnaselt lipoproteiinide peakomponentide ja toiduainete vaheliste lineaarsete seoste puhul, osutusid ka lipoproteiinide faktorite ja toiduainete seoseid uurides olulisusnivool 0,01 statistiliselt ebaolulisteks toiduainetegruppideks leib, sai, kartul, piimatooted, värsked juurviljad, värsked puuviljad ning maiustused (joonised 13a–13c). Väga nõrgalt on seotud „riis-makaronid“.

Huvitav on, et esimene faktor ei osutunud oluliseks ühegi toiduprodukti juures.

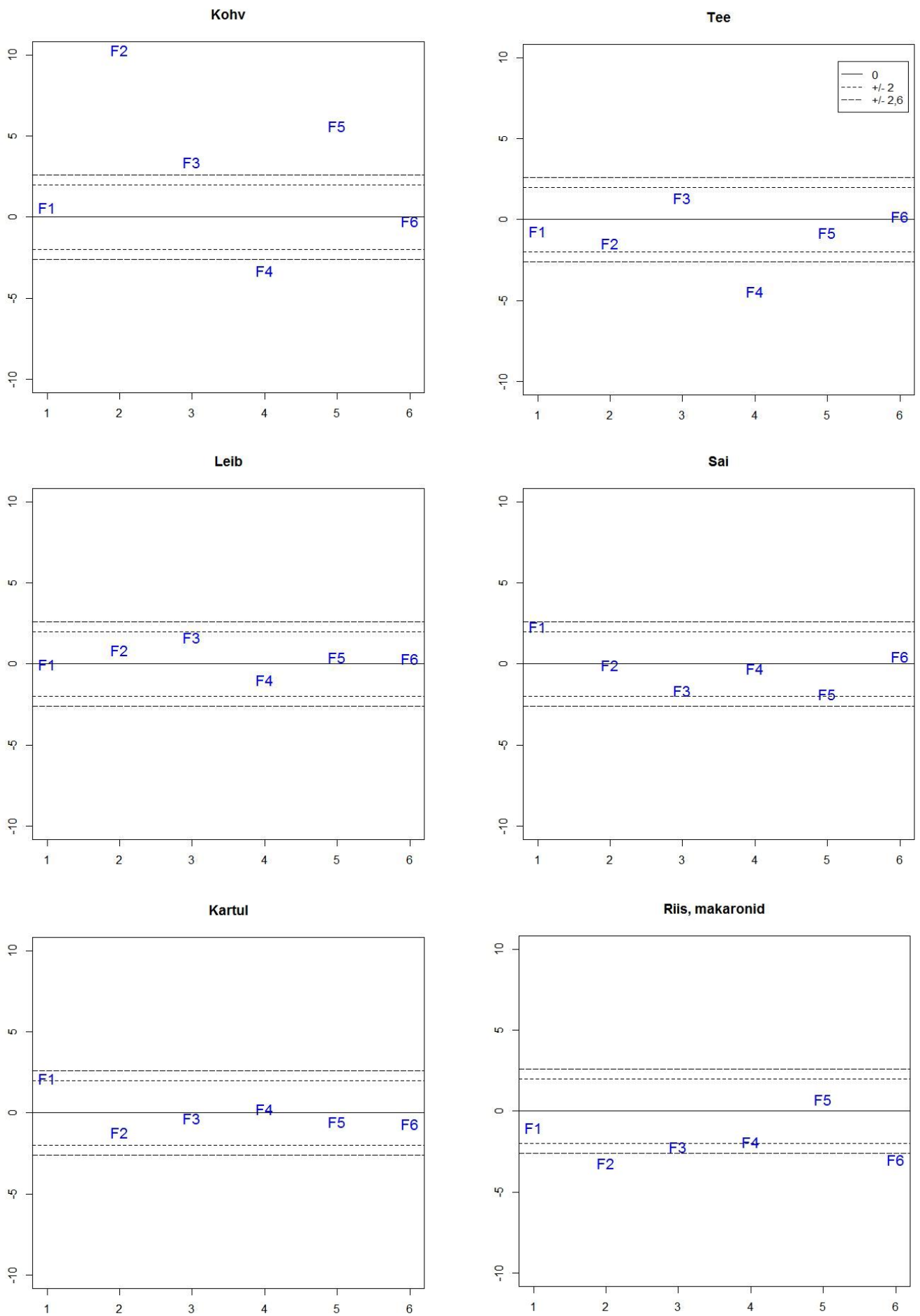
Kõige tugevam seos esines tunnuse „kohv“ juures teise faktoriga. Positiivne seos tähendab, et rohkem kohvi joovatel inimestel on kõrgem LDL-tüüpi lipoproteiinide tase ja madalam HDL-lipoproteiinide tase. Erinevalt peakomponentanalüüsist ei iseloomusta see tulemus kohvijoomine ja VLDL-lipoproteiinide vahelist seost. Teine faktor oli seotud ka tunnustega „riis-makaronid“ ja „puder-müsli-helbed“ (mõlemaga negatiivselt). Seega seosed kahe viimatinimetatud tunnuse ja kohvi vahel olid erisuunalised.

Kolmas faktor seostus küllaltki tugevalt tunnusega „kala“. Lisaks veel kohvi, karastusjookide ja munaga. Tunnuste „kala“, „kohv“ ja „muna“ korral olid seosed positiivsed, tunnuse „karastusjook“ puhul aga negatiivne. Kolmanda faktori puhul olid olulised HDL-lipoproteiinide suured väärtused. Järelikult, lähtudes kolmandat faktorit määravatest lipoproteiinidest, suurendavad „kala“, „kohv“ ja „muna“ hea kolesterooli taset inimese kehas.

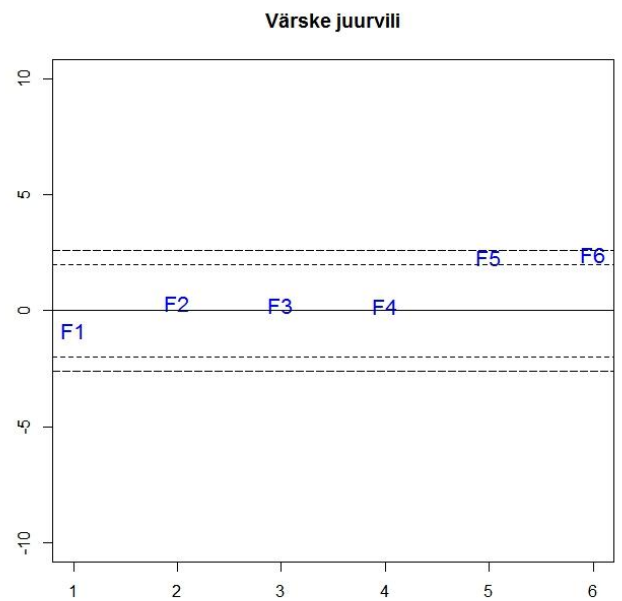
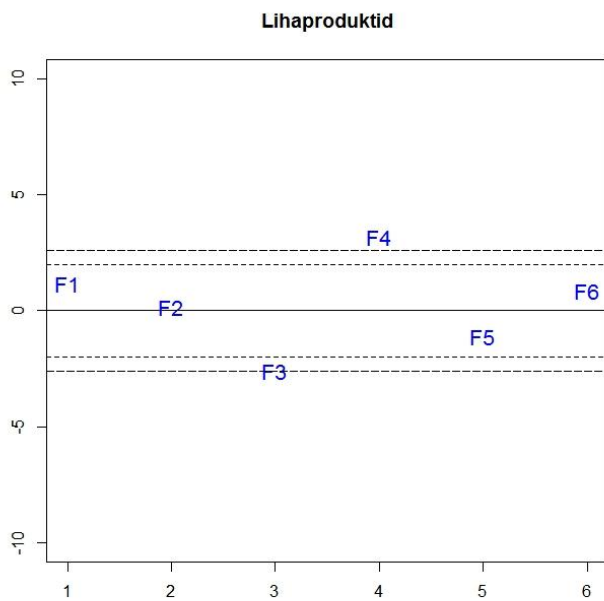
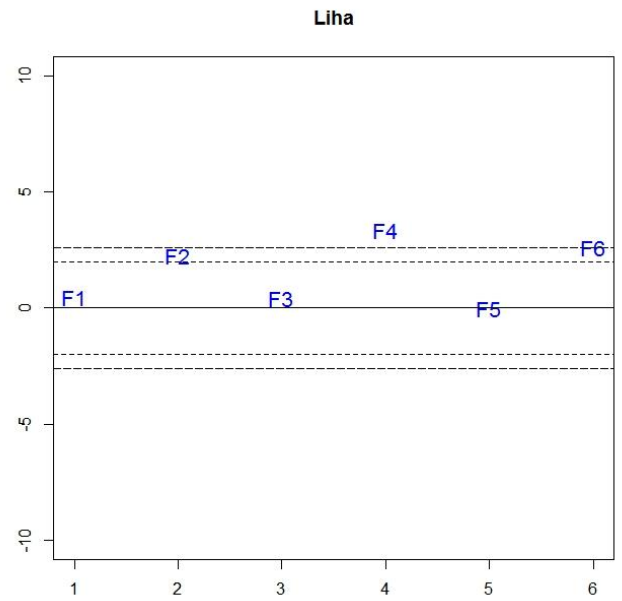
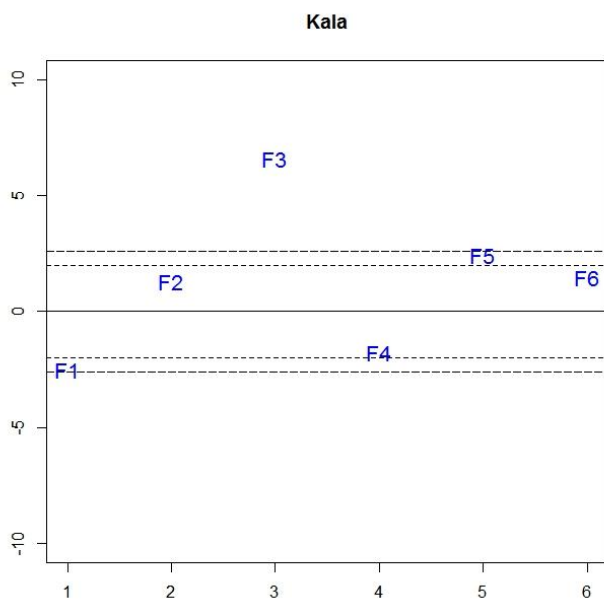
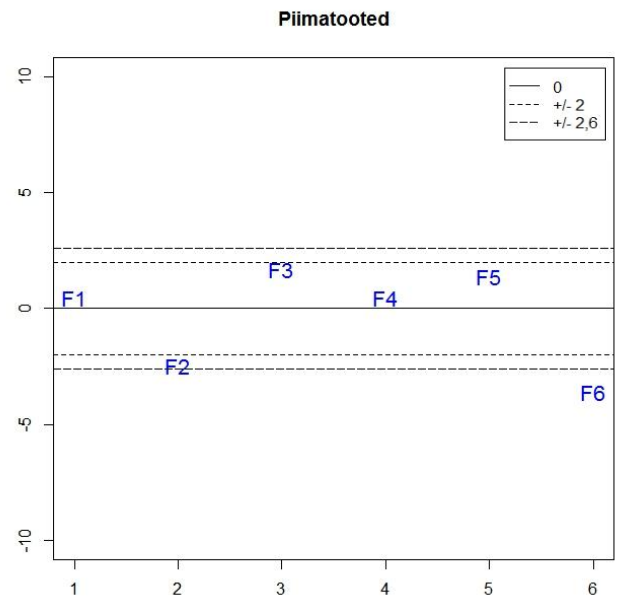
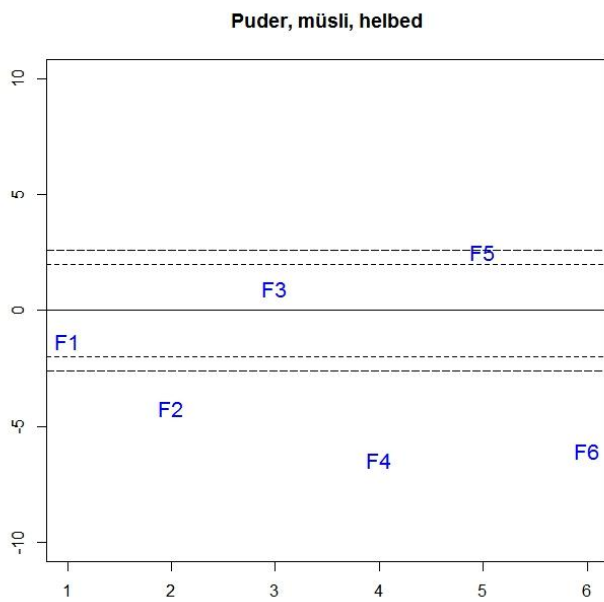
Neljas faktor oli oluline toiduainete „kohv“, „tee“, „puder-müsli-helbed“, „kompotid-keedised“ (negatiivne seos nende kolme tunnusega), „liha“, „lihaproductid“, „karastusjook“ ja „muna“ (positiivne seos) juures. Ka neljanda faktori määramisel oli oluline positiivset tüüpi HDL-kolesterool.

Viies faktor seostus positiivselt kohvi ja keedetud juurviljadega. Viies faktor seostus halba tüüpi VLDL-lipoproteiinidega. Seega võib eeldada, et kohv ja keedetud juurviljad mõjutavad mõlemad negatiivseks peetavat kolesterooli.

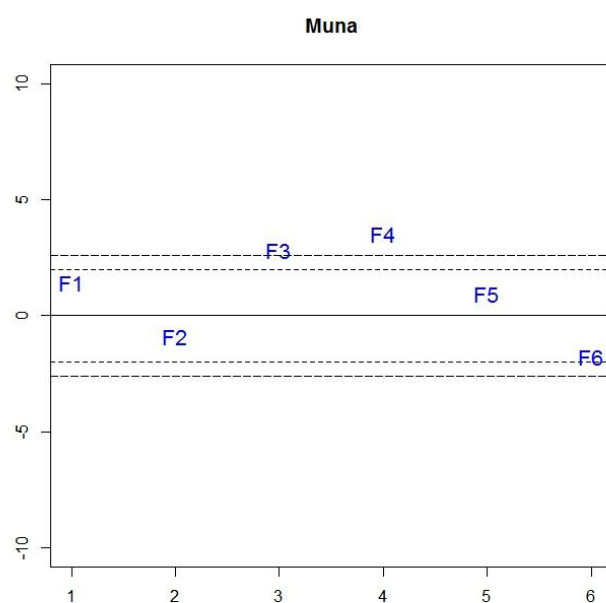
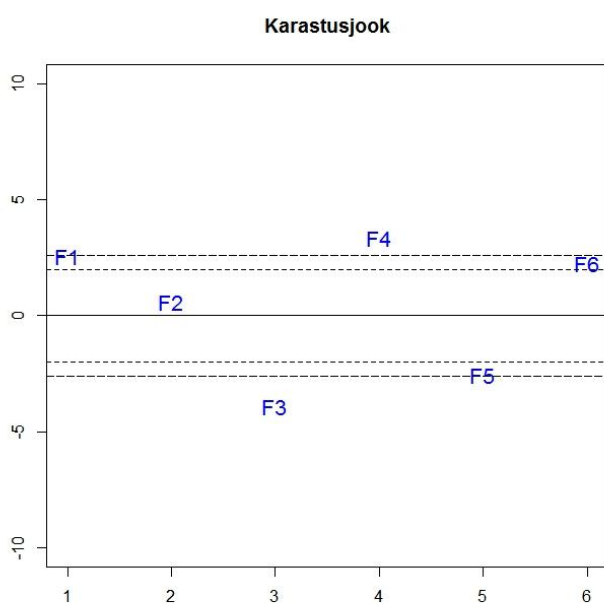
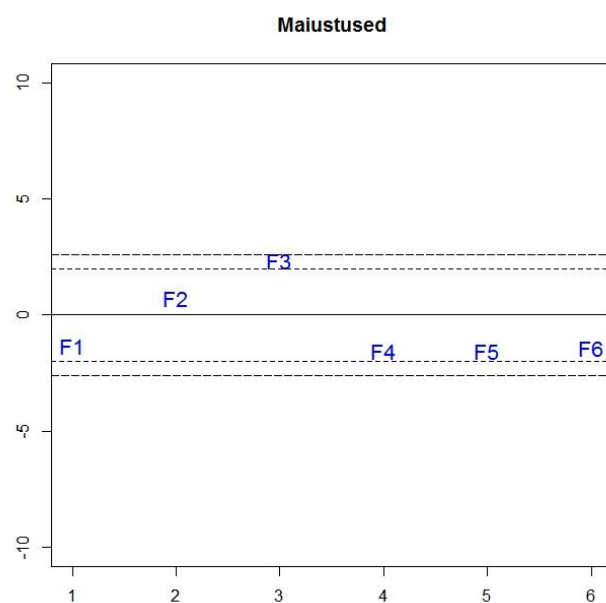
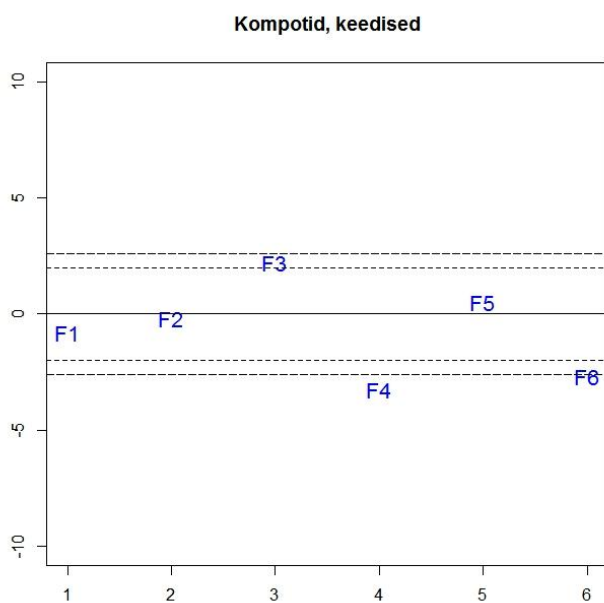
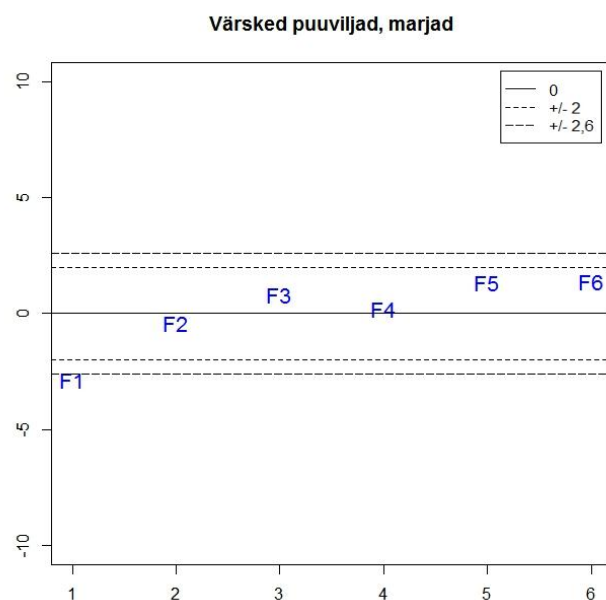
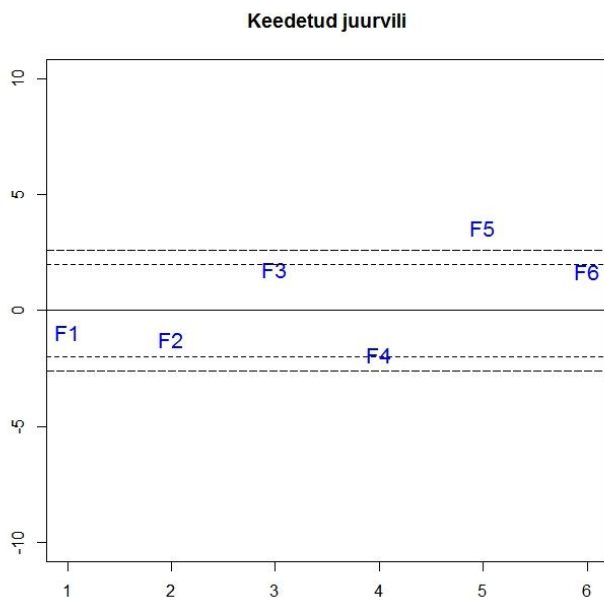
Olulisusnivool 0,01 on statistiliselt olulised faktorid toiduainete lõikes välja toodud ka lisas 2.



Joonis 13a. Toitumistunnuste seosed lipoproteiinide faktoritega



Joonis 13b. Toitumistunnuste seosed lipoproteiinide faktoritega

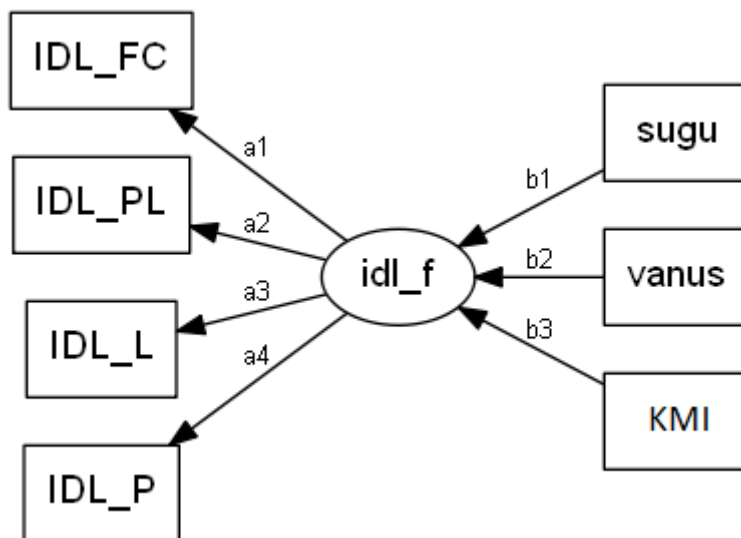


Joonis 13c. Toitumistunnuste seosed lipoproteiinide faktoritega

4. Struktuurivõrrandite mudelid

Seosestruktuuri leidmisel struktuurivõrrandite mudelite abil lähtume esialgu vaid IDL-lipoproteiinide grupist, mille osakesi oli analüüsitavas andmetabelis nelja liiki (tunnused „IDL_FC“, „IDL_PL“, „IDL_L“, „IDL_P“). Struktuurivõrrandite mudelite kasutamine võimaldab üheaegselt uurida seoseid toitumis- ja/või taustatunnuste ning lipoproteiinide vahel vähendades samaaegselt andmete dimensionaalsust, võttes kasutusele latentse faktortunnuse, mis kirjeldab korraga ära neli üksikut lipoproteiini.

Joonisel 14 on toodud struktuurivõrrandite mudel, kus IDL-tüüpi 4 lipoproteiini sõltuvad läbi latentse faktortunnuse „idl_f“ soost, vanusest ja kehamassiindeksist. Parameetrite hinnangud on välja toodud tabelis 4. Tunnuste „sugu“, „vanus“ ja „KMI“ dispersioon on enne analüüsi fikseeritud võrdseks ühega.



Joonis 14. Struktuurivõrrandite mudel IDL-ide ning soo, vanuse ja KMI vahel

Tabel 4. IDL-lipoproteiinide jaoks soo, vanuse ja KMI põhjal hinnatud struktuurivõrrandite mudeli parameetrite hinnangud

Parameeter	Hinnang	Standardviga	p-väärtus
a_1 (IDL_FC \leftarrow idl_f)	0,8522	0,0066	<.0001
a_2 (IDL_FC \leftarrow idl_f)	0,8819	0,0064	<.0001
a_3 (IDL_FC \leftarrow idl_f)	0,8879	0,0063	<.0001
a_4 (IDL_FC \leftarrow idl_f)	0,8877	0,0063	<.0001
b_1 (idl_f \leftarrow sugu)	0,1407	0,0101	<.0001
b_2 (idl_f \leftarrow vanus)	0,4392	0,0112	<.0001
b_3 (idl_f \leftarrow KMI)	0,1186	0,0108	<.0001
V[IDL_FC] (IDL_FC \leftrightarrow IDL_FC)	0,0811	0,0012	<.0001
V[IDL_PL] (IDL_PL \leftrightarrow IDL_PL)	0,0159	0,0002	<.0001
V[IDL_L] (IDL_L \leftrightarrow IDL_L)	0,0025	0,0001	<.0001
V[IDL_P] (IDL_P \leftrightarrow IDL_P)	0,0030	0,0001	<.0001

Joonisele 14 vastavad võrrandid, lähtudes tabelis 4 olevatest parameetrite hinnangutest, on järgmised:

$$IDL_f = 0,1407 * sugu + 0,4392 * vanus + 0,1186 * KMI$$

$$IDL_{FC} = 0,8521 * idl_f + 0,0811$$

$$IDL_{PL} = 0,8819 * idl_f + 0,0159$$

$$IDL_L = 0,8879 * idl_f + 0,0025$$

$$IDL_P = 0,8877 * idl_f + 0,0030$$

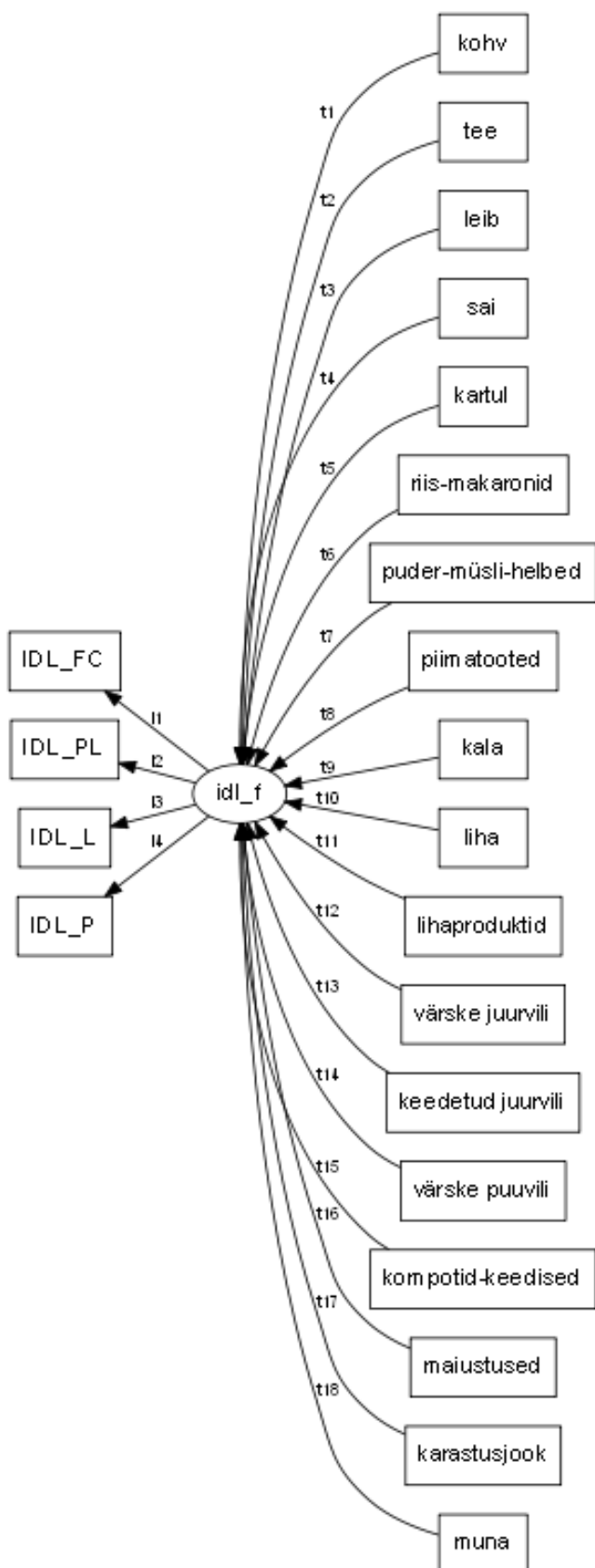
Seega saame kokkuvõtliku faktori „idl_f“ abil ära kirjeldada lipoproteiinide grupi IDL tunnused ning leida vastavalt geenidoonori soole, vanusele ja kehamassiindeksile selle grupi lipoproteiinide tasemete väärtused.

Joonisel 15 on kujutatud, kuidas latentne tunnus „idl_f“ avaldub 18 uuritud toiduainegrupi kaudu. Kasutades tabelis 5 välja toodud parameetrite kordajaid, on võimalik välja kirjutada võrrandid ning leida geenidoonori IDL-tüüpi lipoproteiinide „IDL_FC“, „IDL_PL“, „IDL_L“ ja „IDL_P“ kontsentratsioonid vastavalt sellele kui sageli ja missuguseid toiduaineid uuritav isik tarbib. Toiduainete dispersioonid on enne analüüsi fikseeritud võrdseks ühega.

Lineaarsel regressioonil ilmnesid IDL-lipoproteiinide ja toitumistunnuste vahel statistiliselt olulised seosed vaid tunnuste „kohv“ (tugev positiivne seos), „riis-makaronid“ ja „puder-müsli-helbed“ (nõrgad negatiivsed seosed) juures. Struktuurivõrrandite mudelite korral on olulisi seoseid tunduvalt rohkem.

Olulisusnivool $\alpha = 0,01$ on statistiliselt mitteolulisteks toiduaineteks „tee“, „leib“, „kartul“, „puder-müsli-helbed“, „liha“, „lihaproductid“, „värske juurvili“, „värske puuvili“ ja „muna“. Kõige tugevamad seosed esinevad tunnuste „kohv“ ja „karastusjoogid“ juures, kusjuures seosed on erisuunalised.

Samas sarnanevad tulemused eelneval analüüsil saadud tulemustega – kohviga on seosed väga tugevad ning statistiliselt mitteoluliseks osutunud toiduained on samuti suuresti kattuvad.



Joonis 15. Struktuurivõrrandite mudel IDL-ide ja toiduainete vahel

Tabel 5. IDL-lipoproteiinide jaoks toiduainete põhjal hinnatud struktuurivõrrandite mudeli parameetrite hinnangud

Parameeter	Hinnang	Standardviga	p-väärtus
i_1 (IDL_FC \leftarrow idl_f)	0,9306	0,0072	<.0001
i_2 (IDL_PL \leftarrow idl_f)	0,9630	0,0070	<.0001
i_3 (IDL_L \leftarrow idl_f)	0,9696	0,0069	<.0001
i_4 (IDL_P \leftarrow idl_f)	0,9693	0,0069	<.0001
t_1 (idl_f \leftarrow kohv)	0,1067	0,0104	<.0001
t_2 (idl_f \leftarrow tee)	0,0002	0,0104	0,9865
t_3 (idl_f \leftarrow leib)	0,0080	0,0104	0,4434
t_4 (idl_f \leftarrow sai)	-0,0411	0,0109	0,0002
t_5 (idl_f \leftarrow kartul)	0,0102	0,0108	0,3452
t_6 (idl_f \leftarrow riis-makaronid)	-0,0727	0,0103	<.0001
t_7 (idl_f \leftarrow puder-müsli-helbed)	0,0211	0,0111	0,0565
t_8 (idl_f \leftarrow piimatooted)	-0,0372	0,0063	0,0005
t_9 (idl_f \leftarrow kala)	0,0716	0,0106	<.0001
t_{10} (idl_f \leftarrow liha)	-0,0082	0,0107	0,4449
t_{11} (idl_f \leftarrow lihaproduktid)	-0,0105	0,0111	0,3430
t_{12} (idl_f \leftarrow värsked juurvili)	-0,0080	0,0112	0,4749
t_{13} (idl_f \leftarrow keedetud juurvili)	0,0547	0,0111	<.0001
t_{14} (idl_f \leftarrow värsked puuviljad)	0,0155	0,0111	0,1619
t_{15} (idl_f \leftarrow kompotid-keedised)	0,0305	0,0108	0,0049
t_{16} (idl_f \leftarrow maiustused)	-0,0850	0,0106	<.0001
t_{17} (idl_f \leftarrow karastusjook)	-0,1125	0,0107	<.0001
t_{18} (idl_f \leftarrow muna)	0,0224	0,0105	0,0322
V[IDL_FC] (IDL_FC \leftrightarrow IDL_FC)	0,0809	0,0012	<.0001
V[IDL_PL] (IDL_PL \leftrightarrow IDL_PL)	0,0159	0,0002	<.0001
V[IDL_L] (IDL_L \leftrightarrow IDL_L)	0,0025	0,0001	<.0001
V[IDL_P] (IDL_P \leftrightarrow IDL_P)	0,0030	0,0001	<.0001

Peatükis 3 läbi viidud peakomponent- ja faktoranalüüs näitasid, et kogu TÜ geenivaramu metaboliitide andmestikus oleva informatsiooni kokkusurumiseks sobiks kõige paremini viis kuni kuus latentset ehk otseselt mittemõõdetavat tunnust. Need tunnused kirjeldaksid esialgsed andmed ära ~98% ulatuses ning kuna neid on tunduvalt vähem kui esialgseid tunnuseid (74 lipoproteiini) suureneks ka ülevaatlikkus.

Kasutades struktuurivõrrandite mudeleid, ei õnnestunud läbi viia mitmemõõtmelist analüüsi ülejäänud kolme lipoproteiinide grupi (VLDL-, LDL- ja HDL-lipoproteiinid) peal, kuna tarkvaral tekkisid mudelite hindamisel probleemid. Hindamise lihtsustamiseks proovisin näiteks võrdsustada nende lipoproteiinide omavahelisi korrelatsioone, mis peatüki 2.2 põhjal ületasid suurusjärku 0,99. Samuti katsetasin teiste hinnatavate parameetrite fikseerimist, ent kahjuks tulemuseta.

IDL-lipoproteiinide näite põhjal oli näha, et struktuurivõrrandite mudelite meetod võimaldab üheaegselt nii andmestiku dimensionaalsuse vähendamist kui ka tunnustevahelise seosestruktuuri uurimist. Seega on struktuurivõrrandite mudelid suure andmestiku korral heaks meetodiks mudelisse kaasatud tunnuste vaheliste sõltuvusseoste kirjeldamisel ning nimetatud meetodit tasuks edaspidi TÜ geenivaramu andmestiku kirjeldamiseks kindlasti lähemalt uurida.

Kokkuvõte

Bakalaureusetöö eesmärgiks oli mitmemõõtmelise analüüsi meetodite uurimine andmete dimensionaalsuse vähendamiseks (kas niimoodi tekib andmete osas selgus või kaob osa informatsioonist ära) ning soov leida seoseid erinevate toiduainete ja eri tüüpi lipoproteiinide (ning seeläbi hea ja halva kolesterooli) vahel.

Kasutatavad andmed pärinevad Tartu Ülikooli Eesti Geenivaramust. Valim koosnes 9884 indiviidist, kellel oli mõõdetud 74 tunnust, mis iseloomustavad eri tüüpi kolesterooli kontsentratsiooni ning jagunevad nelja suurde gruppi (VLDL-, IDL-, LDL- ja HDL-lipoproteiinid). Samuti oli teada nende inimeste sugu, vanus ja kehamassiindeks ning 18 erineva toiduprodukti tarbimise sagedus. Analüüsimeetoditena kasutati korrelatsioonikordajate uurimist, lineaarset regressiooni, peakomponentanalüüsi ja faktoranalüüsi. Lisaks vaadeldi IDL-tüüpi kolesterooli seoseid struktuurivõrrandite mudelite abil.

Korrelatsioone uurides ilmnas, et vaadeldavad lipoproteiinid on omavahel grupiti tugevalt seotud, kuid korrelatsioonid toiduainete vahel on pigem madalad. Hiljem kinnitasid seda peakomponent- ja faktoranalüüs – 74 lipoproteiini on hästi kirjeldatavad 5-6 peakomponendi või faktori abil. Ilmnas, et mõningate toiduainete (näiteks kohvi, teraviljatoodete ja kala) tarbimine seostub teatavat tüüpi lipoproteiinidega väga tugevalt. Eelkõige olid seosed näha lineaarse regressioonanalüüsi tulemuste juures, kust oli näha kuidas kindel toiduaine mingit üksikut lipoproteiini mõjutab.

Peakomponent- ja faktoranalüüs lähtuvad metaboliitide üldisest korrelatsioonistruktuurist. Kuigi faktorite ja esimeste peakomponentide kirjeldusvõime on TÜ geenivaramu metaboliitide andmestiku korral väga kõrge, ei kandu see täielikult üle lineaarsetele seostele. Peakomponendid ja faktorid annavad seose üldise lipoproteiinide profiiliga, aga iga toiduaine tarbimine omakorda on tõenäoliselt iga üksiku lipoproteiini konkreetse taseme mõjutajaks.

Tulevikus on võimalik tööd edasi arendada eelkõige struktuurivõrrandite mudelite osas – käesolevas töös uuriti nimetatud mudeleid vaid IDL-tüüpi lipoproteiine kasutades. Ilmnas, et nimetatud mudelite abil on võimalik prognoosida inimese lipoproteiini kontsentratsiooni taset, teades isiku toitumisharjumusi. Samuti saab töö edasiarenduses lisaks uurida haiguste ja metabooloomika vahelisi mitmesuunalisi seoseid.

**Connections between eating patterns and lipoproteins
based on The Estonian Genome Center data:
Multivariate Analysis for correlated data**

Bachelor thesis

Kristi Helekivi

Summary

The purpose of this thesis was to find possible relations between eating habits and lipoprotein levels and investigate advantages of multivariate analysis. The analysis is based on data from The Estonian Genome Center, University of Tartu.

The analysis was carried out using data of 9884 individuals, whose plasma samples were analyzed using the NMR-technology to determine the concentrations of 74 lipoprotein metabolites. Their eating patterns and background characteristics sex, age and body mass index were also known. Analysis was carried out using Linear Regression, Principal Component Analysis (PCA), Factor Analysis (FA) and Structural Equation Modeling (SEM). Within 4 main lipoprotein subgroups (VLDL, IDL, LDL, HDL) the correlations between measurements were very high but correlations between food products were rather low. Later, also PCA and FA confirmed presumptions – the lipoprotein measurements can be described by a small number of principal components or factors.

Consumption of some food items (e.g., coffee, cereals, and fish) was significantly associated with some lipoprotein groups, whereas some other food items did not indicate associations. The main findings were the same across different analysis methods.

Unfortunately, it was not possible to investigate all the wished-for connections with SEM due to computer software problems. Analyses were only carried out with one type of lipoproteins (Intermediate-density lipoproteins). The outcome was that SEM is a suitable method for predicting people lipoprotein-levels based on their food consumption, but it needs further investigation.

Kasutatud kirjandus

- [1] Metaboliidid. Wikipedia, URL (vaadatud 11.02.2013)
<http://en.wikipedia.org/wiki/Metabolite>
- [2] Geenivaramu põhikiri. Tartu Ülikooli Eesti Geenivaramu kodulehekül, URL (vaadatud 11.02.2013) <http://www.geenivaramu.ee/info/ametlik-info/geenivaramu-pohikiri.html>
- [3] *Estonian Genome Center*. Aastaraamat 2001-2011. Tartu: Estonian Genome Center, University of Tartu 2011, lk 5
- [4] Kehamassiindeks. Wikipedia, URL (vaadatud 11.02.2013)
http://en.wikipedia.org/wiki/Body_mass_index
- [5] Lipoproteiinid. Wikipedia, URL (vaadatud 11.02.2013)
<http://en.wikipedia.org/wiki/Lipoprotein>
- [6] VLDL-i testimine. MedlinePlus, URL (vaadatud 30.04.2013)
<http://www.nlm.nih.gov/medlineplus/ency/article/003494.htm>
- [7] Vahepealse tihedusega lipoproteiinid. Wikipedia, URL (vaadatud 11.02.2013)
http://en.wikipedia.org/wiki/Intermediate-density_lipoprotein
- [8] Lipoproteiinid. Hea kolesterool (HDL), halb kolesterool (LDL). *Scientific Psychic*, URL (vaadatud 11.02.2013) <http://www.scientificpsychic.com/health/lipoproteins-LDL-HDL.html>
- [9] Traat I. (2012). Matemaatiline statistika II. Käsikirjaline loengukonspekt. Tartu: Tartu Ülikool, matemaatilise statistika instituut.
- [10] Tiit, E. & Parring, A. (1976). MATEMAATILINE STATISTIKA III. Tartu: Tartu Riiklik Ülikool, lk 69
- [11] Härdle, W. & Simar, L. (2003). Applied Multivariate Statistical Analysis. Berliin: Springer, lk 86-87
- [12] Möls, M. (2013). Lineaarsed mudelid. Käsikirjaline loengukonspekt. Tartu: Tartu Ülikool, matemaatilise statistika instituut.
- [13] Traat I. (2013). Mitmemõõtmeline analüüs. Käsikirjaline loengukonspekt. Tartu: Tartu Ülikool, matemaatilise statistika instituut.

Lisad

Lisa 1. Lipoproteiinid koos inglisekeelse nimetusega ja lipoproteiinide jaotust kirjeldavad histogrammid

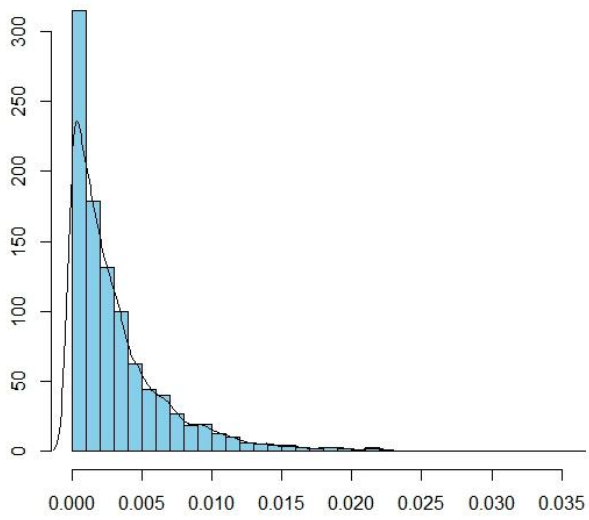
Tunnuse lühend	Inglisekeelne nimetus
XXL_VLDL_PL	<i>Phospholipids in chylomicrons and extremely large VLDL</i>
XXL_VLDL_L	<i>Total lipids in chylomicrons and extremely large VLDL</i>
XXL_VLDL_P	<i>Concentration of chylomicrons and extremely large VLDL particles</i>
XL_VLDL_PL	<i>Phospholipids in very large VLDL</i>
XL_VLDL_TG	<i>Triglycerides in very large VLDL</i>
XL_VLDL_L	<i>Total lipids in very large VLDL</i>
XL_VLDL_P	<i>Concentration of very large VLDL particles</i>
L_VLDL_C	<i>Total cholesterol in large VLDL</i>
L_VLDL_FC	<i>Free cholesterol in large VLDL</i>
L_VLDL_PL	<i>Phospholipids in large VLDL</i>
L_VLDL_TG	<i>Triglycerides in large VLDL</i>
L_VLDL_CE	<i>Cholesterol esters in large VLDL</i>
L_VLDL_L	<i>Total lipids in large VLDL</i>
L_VLDL_P	<i>Concentration of large VLDL particles</i>
M_VLDL_C	<i>Total cholesterol in medium VLDL</i>
M_VLDL_FC	<i>Free cholesterol in medium VLDL</i>
M_VLDL_PL	<i>Phospholipids in medium VLDL</i>
M_VLDL_TG	<i>Triglycerides in medium VLDL</i>
M_VLDL_CE	<i>Cholesterol esters in medium VLDL</i>
M_VLDL_L	<i>Total lipids in medium VLDL</i>
M_VLDL_P	<i>Concentration of medium VLDL particles</i>

S_VLDL_C	<i>Total cholesterol in small VLDL</i>
S_VLDL_FC	<i>Free cholesterol in small VLDL</i>
S_VLDL_PL	<i>Phospholipids in small VLDL</i>
S_VLDL_TG	<i>Triglycerides in small VLDL</i>
S_VLDL_L	<i>Total lipids in small VLDL</i>
S_VLDL_P	<i>Concentration of small VLDL particles</i>
XS_VLDL_PL	<i>Phospholipids in very small VLDL</i>
XS_VLDL_TG	<i>Triglycerides in very small VLDL</i>
XS_VLDL_L	<i>Total lipids in very small VLDL</i>
XS_VLDL_P	<i>Concentration of very small VLDL particles</i>
IDL_FC	<i>Free cholesterol in IDL</i>
IDL_PL	<i>Phospholipids in IDL</i>
IDL_L	<i>Total lipids in IDL</i>
IDL_P	<i>Concentration of IDL particles</i>
L_LDL_C	<i>Total cholesterol in large LDL</i>
L_LDL_FC	<i>Free cholesterol in large LDL</i>
L_LDL_PL	<i>Phospholipids in large LDL</i>
L_LDL_CE	<i>Cholesterol esters in large LDL</i>
L_LDL_L	<i>Total lipids in large LDL</i>
L_LDL_P	<i>Concentration of large LDL particles</i>
M_LDL_C	<i>Total cholesterol in medium LDL</i>
M_LDL_PL	<i>Phospholipids in medium LDL</i>
M_LDL_CE	<i>Cholesterol esters in medium LDL</i>
M_LDL_L	<i>Total lipids in medium LDL</i>
M_LDL_P	<i>Concentration of medium LDL particles</i>
S_LDL_C	<i>Total cholesterol in small LDL</i>

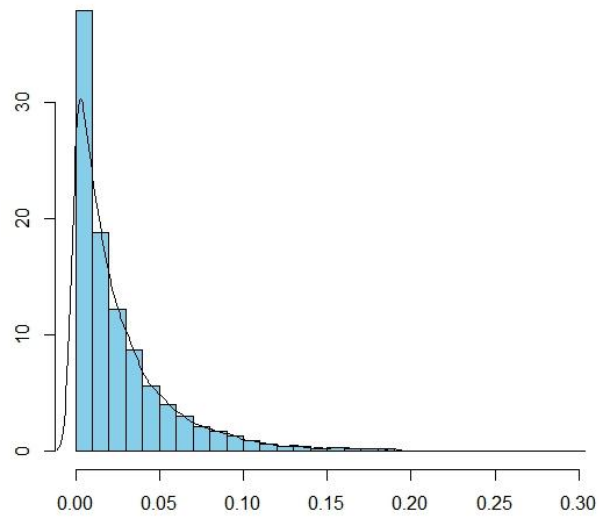
S_LDL_L	<i>Total lipids in small LDL</i>
S_LDL_P	<i>Concentration of small LDL particles</i>
XL_HDL_C	<i>Total cholesterol in very large HDL</i>
XL_HDL_FC	<i>Free cholesterol in very large HDL</i>
XL_HDL_PL	<i>Phospholipids in very large HDL</i>
XL_HDL_TG	<i>Triglycerides in very large HDL</i>
XL_HDL_CE	<i>Cholesterol esters in very large HDL</i>
XL_HDL_L	<i>Total lipids in very large HDL</i>
XL_HDL_P	<i>Concentration of very large HDL particles</i>
L_HDL_C	<i>Total cholesterol in large HDL</i>
L_HDL_FC	<i>Free cholesterol in large HDL</i>
L_HDL_PL	<i>Phospholipids in large HDL</i>
L_HDL_CE	<i>Cholesterol esters in large HDL</i>
L_HDL_L	<i>Total lipids in large HDL</i>
L_HDL_P	<i>Concentration of large HDL particles</i>
M_HDL_C	<i>Total cholesterol in medium HDL</i>
M_HDL_FC	<i>Free cholesterol in medium HDL</i>
M_HDL_PL	<i>Phospholipids in medium HDL</i>
M_HDL_CE	<i>Cholesterol esters in medium HDL</i>
M_HDL_L	<i>Total lipids in medium HDL</i>
M_HDL_P	<i>Concentration of medium HDL particles</i>
S_HDL_TG	<i>Triglycerides in small HDL</i>
S_HDL_L	<i>Total lipids in small HDL</i>
S_HDL_P	<i>Concentration of small HDL particles</i>
IDL_C	<i>Total cholesterol in IDL</i>
LDL_C	<i>Total cholesterol in LDL</i>

HDL_C	<i>Total cholesterol in HDL</i>
-------	---------------------------------

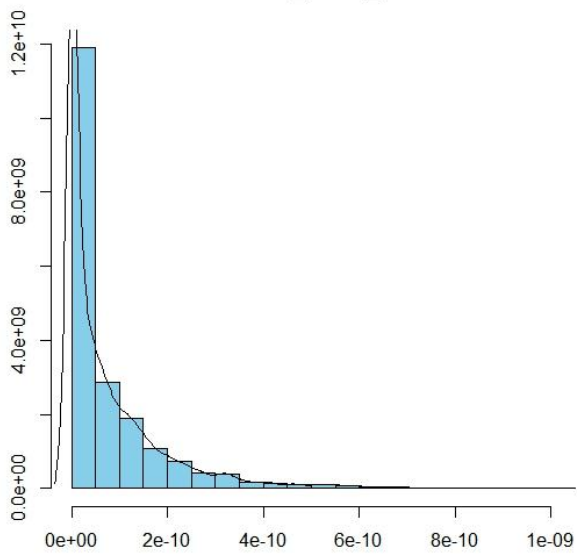
XXL_VLDL_PL



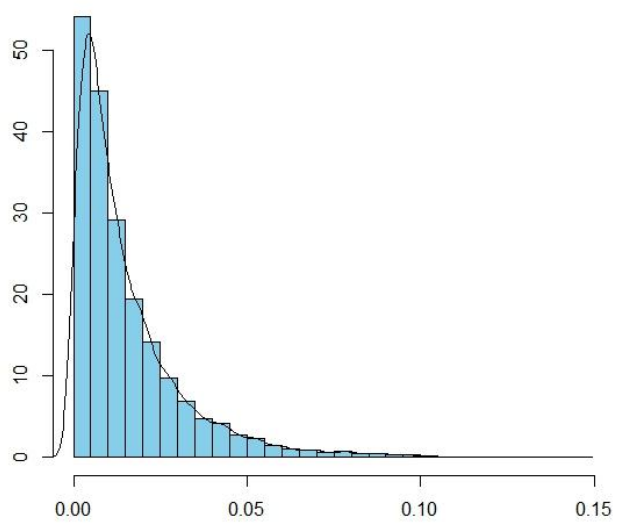
XXL_VLDL_L



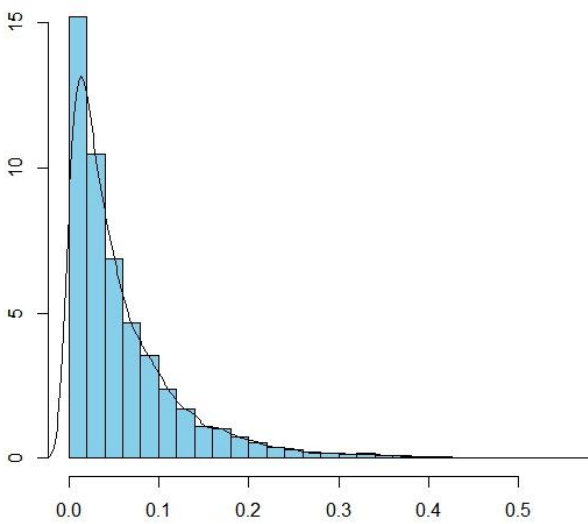
XXL_VLDL_P



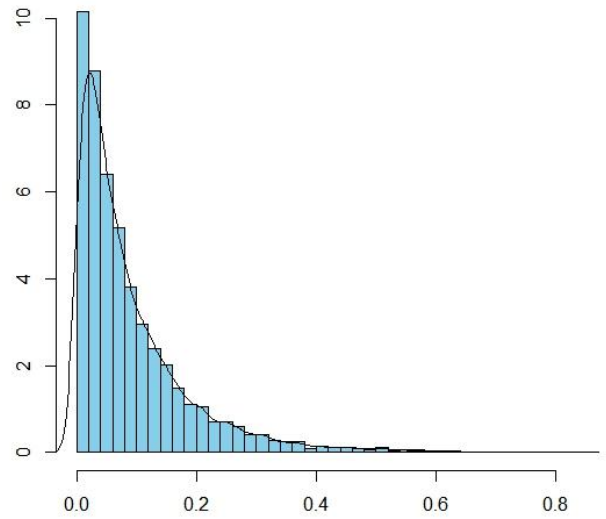
XL_VLDL_PL



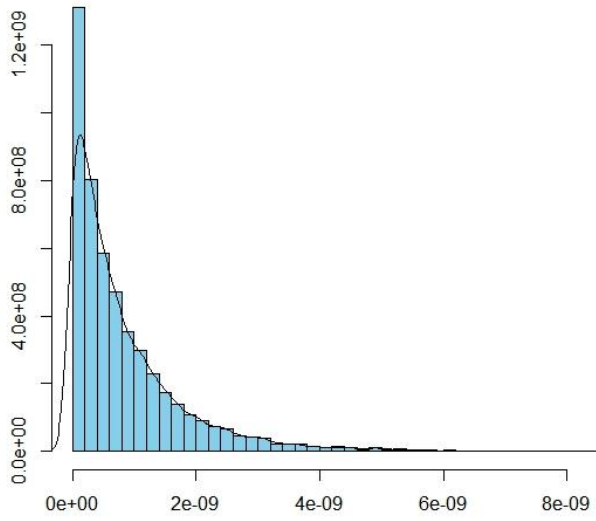
XL_VLDL_TG



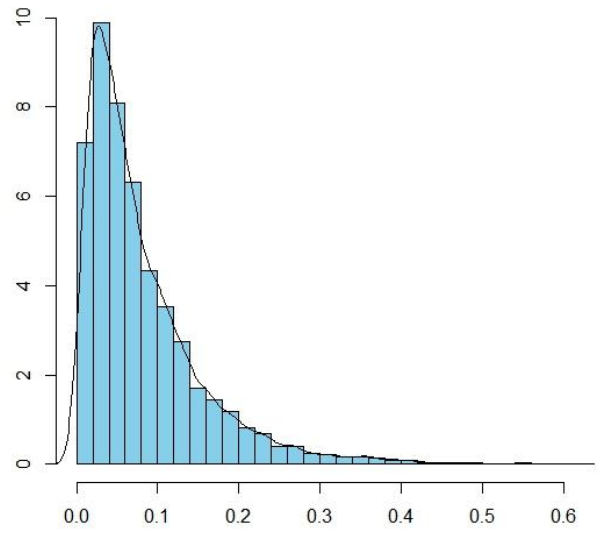
XL_VLDL_L



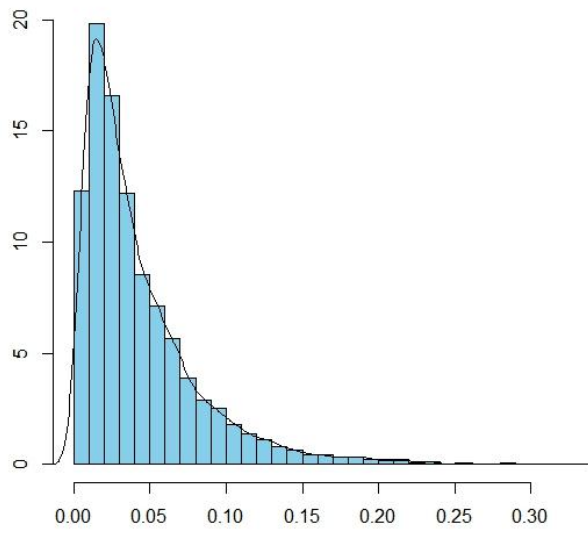
XL_VLDL_P



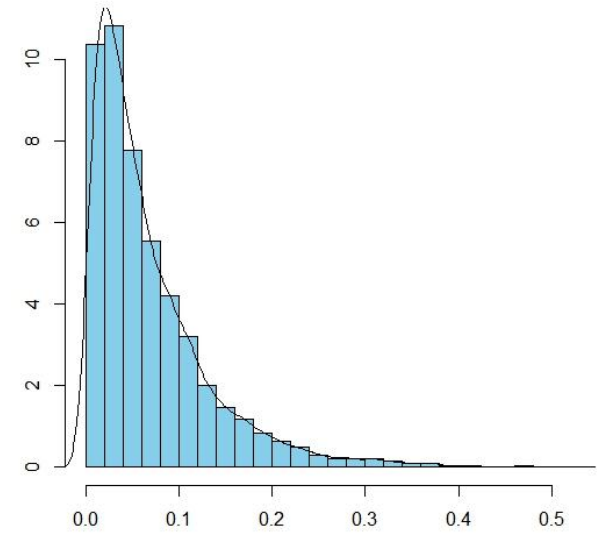
L_VLDL_C



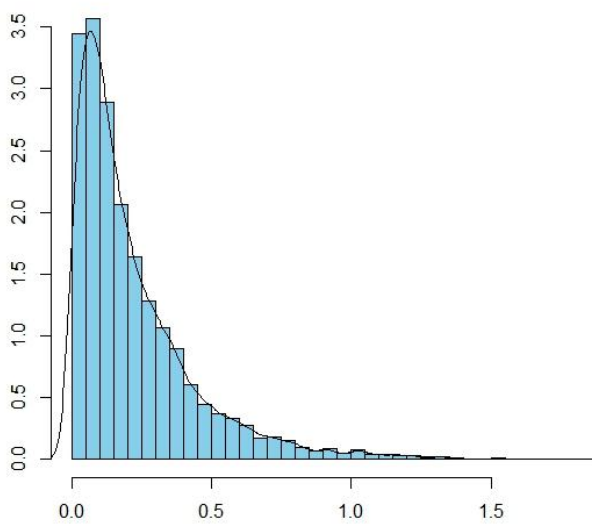
L_VLDL_FC



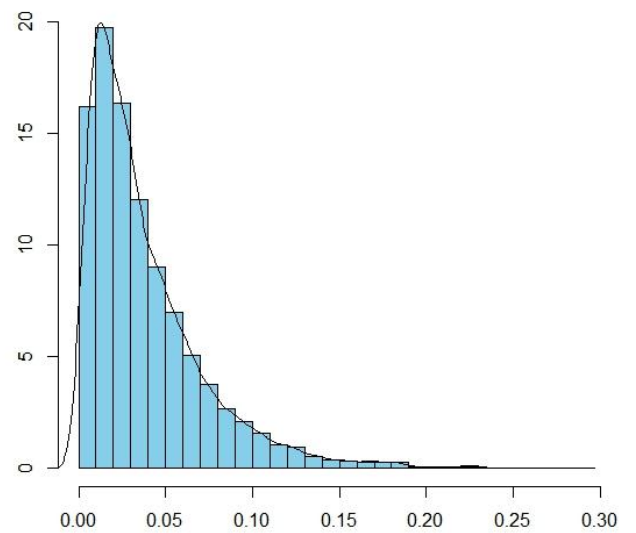
L_VLDL_PL



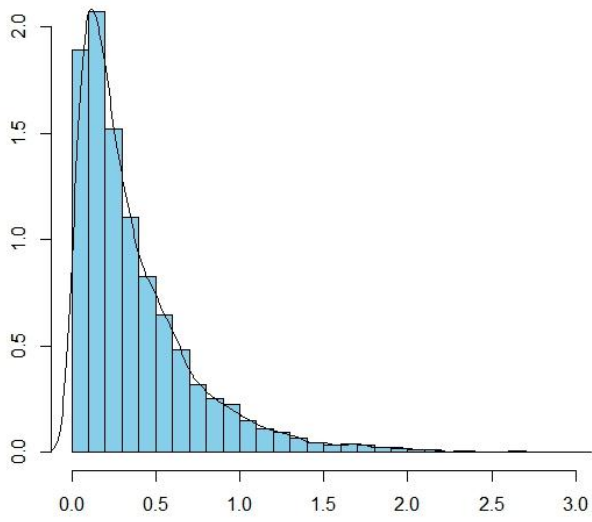
L_VLDL_TG



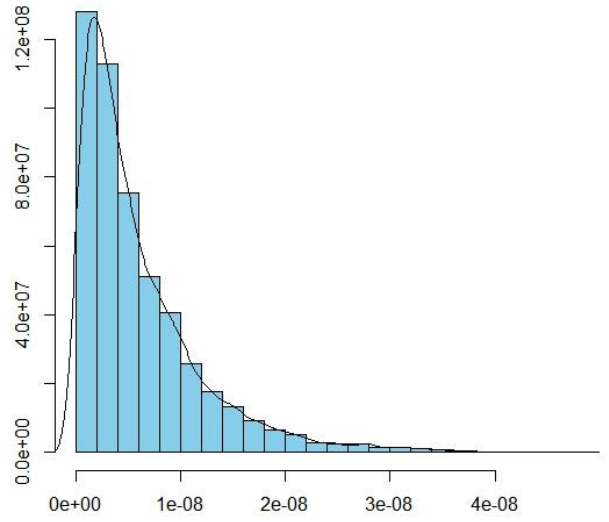
L_VLDL_CE



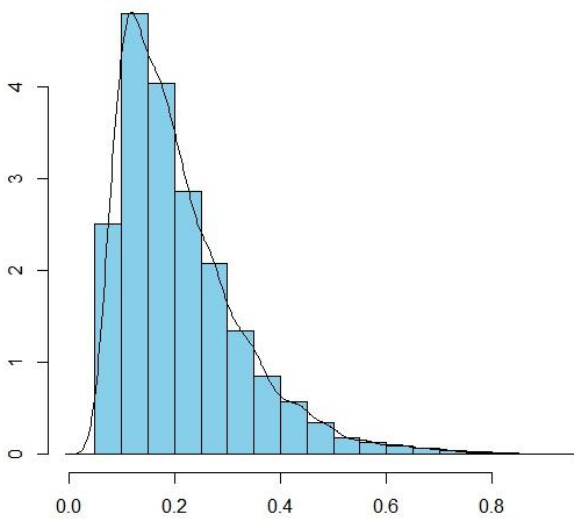
L_VLDL_L



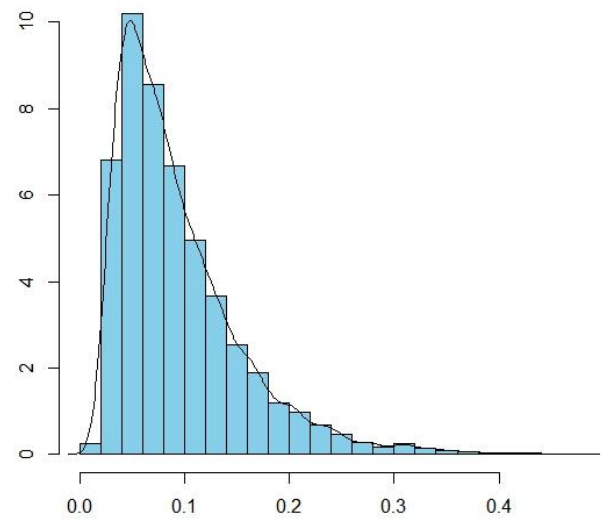
L_VLDL_P



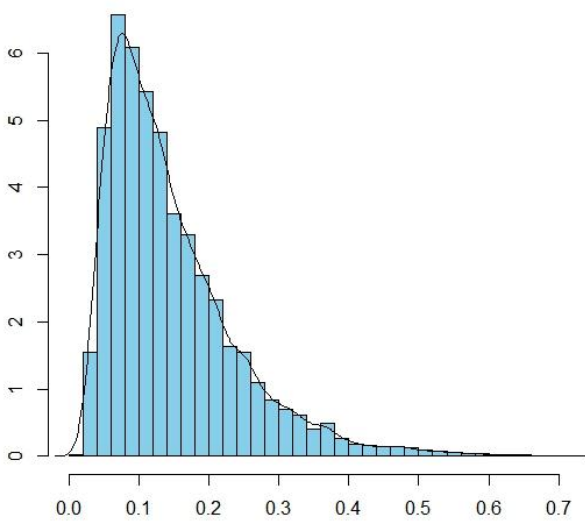
M_VLDL_C



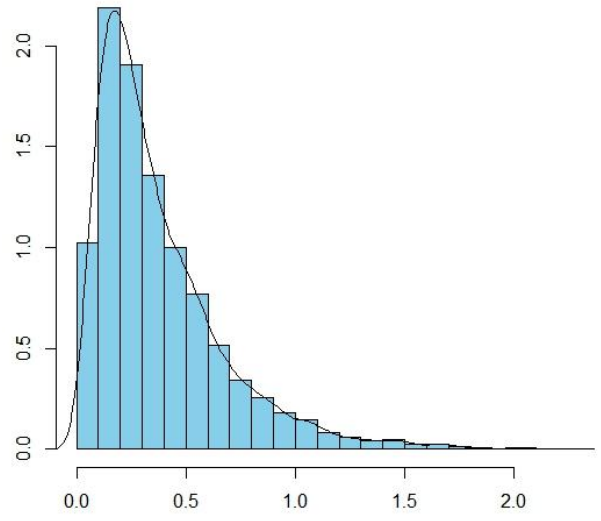
M_VLDL_FC



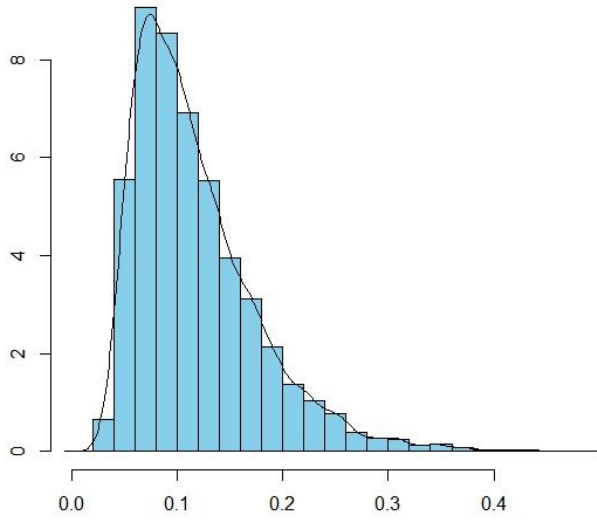
M_VLDL_PL



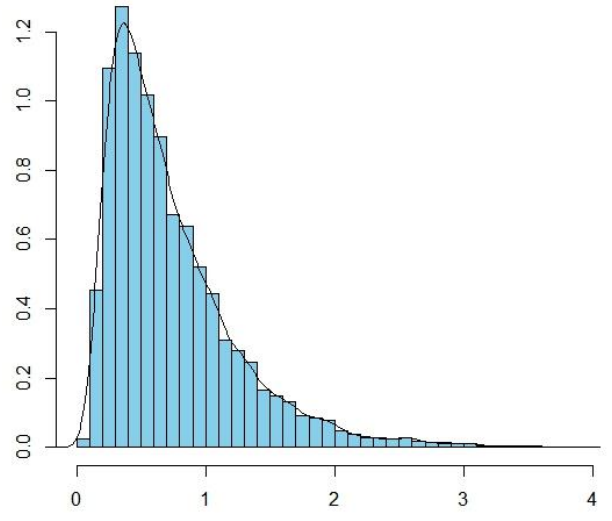
M_VLDL_TG



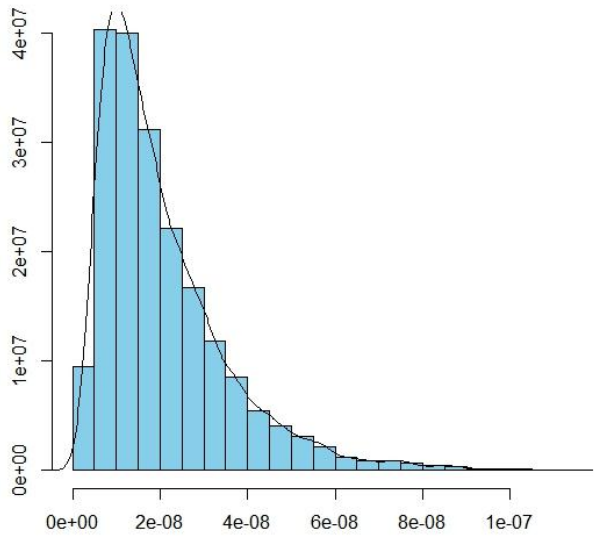
M_VLDL_CE



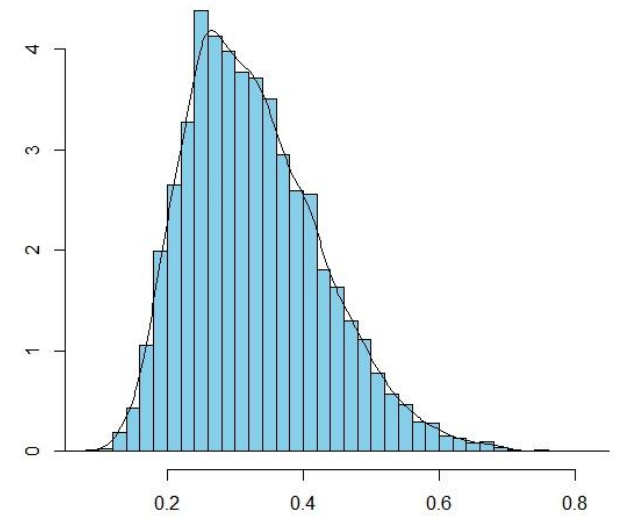
M_VLDL_L



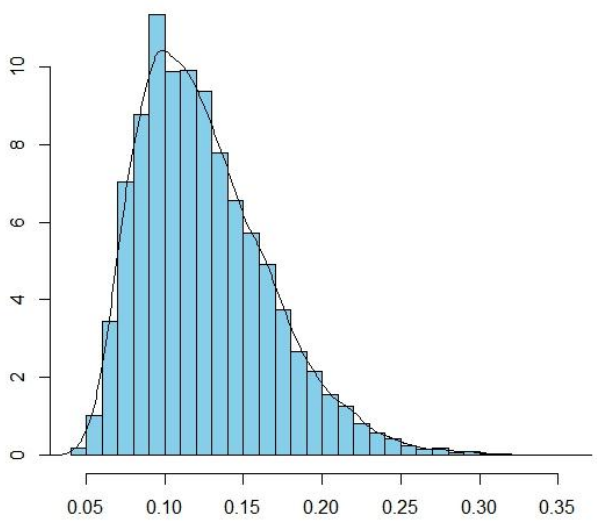
M_VLDL_P



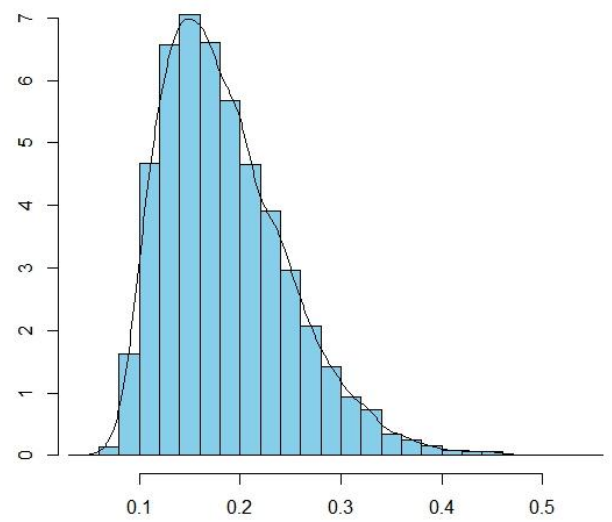
S_VLDL_C



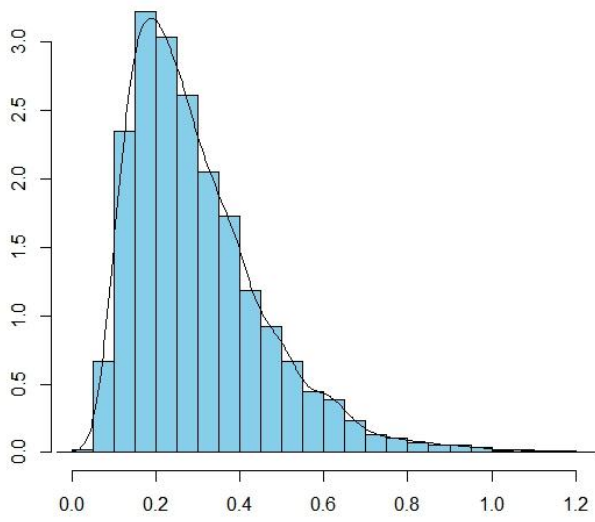
S_VLDL_FC



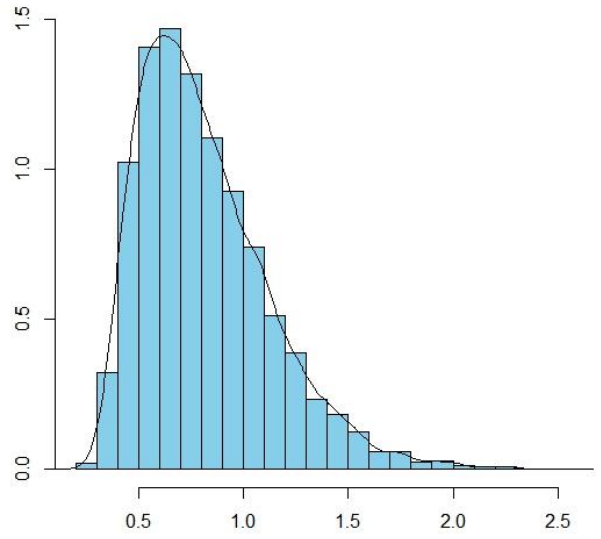
S_VLDL_PL



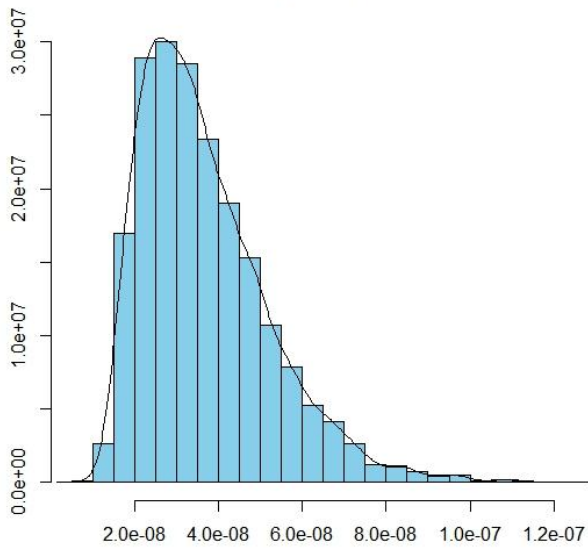
S_VLDL_TG



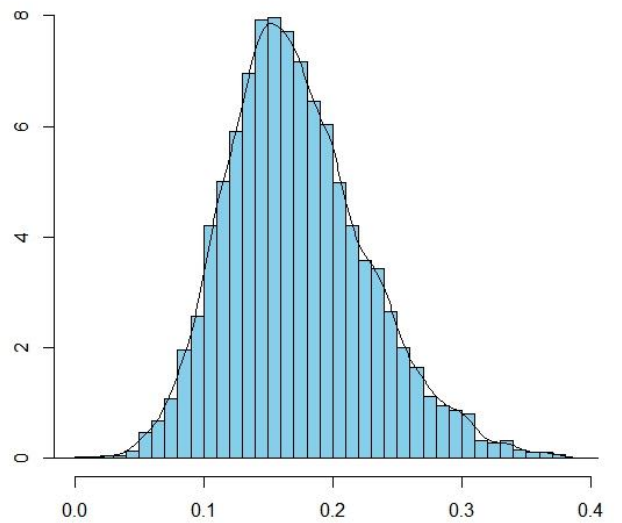
S_VLDL_L



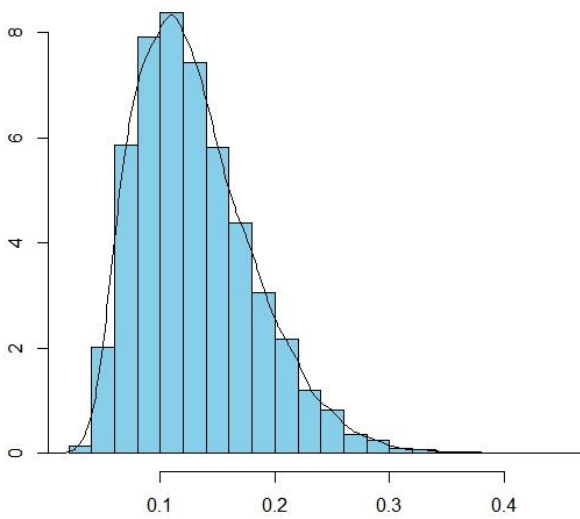
S_VLDL_P



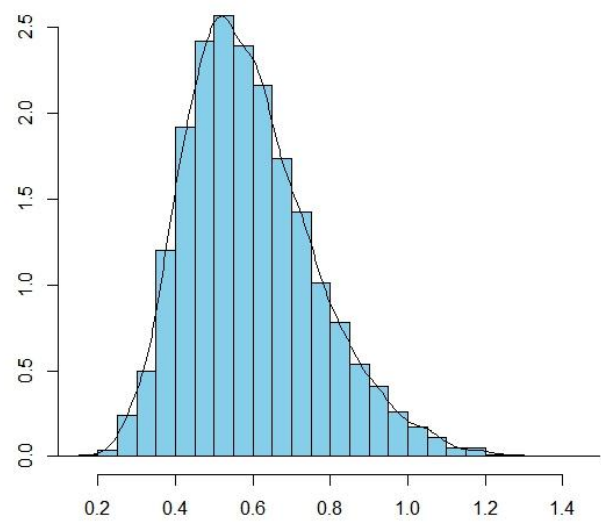
XS_VLDL_PL



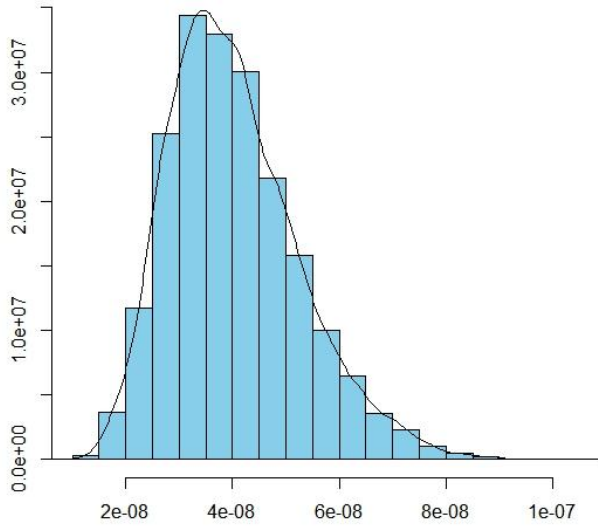
XS_VLDL_TG



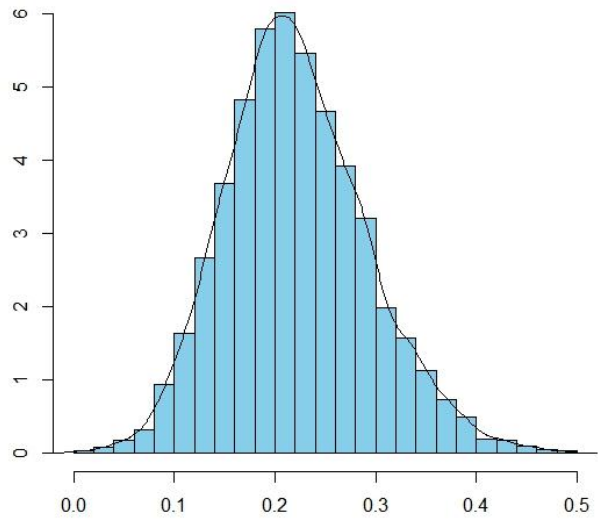
XS_VLDL_L



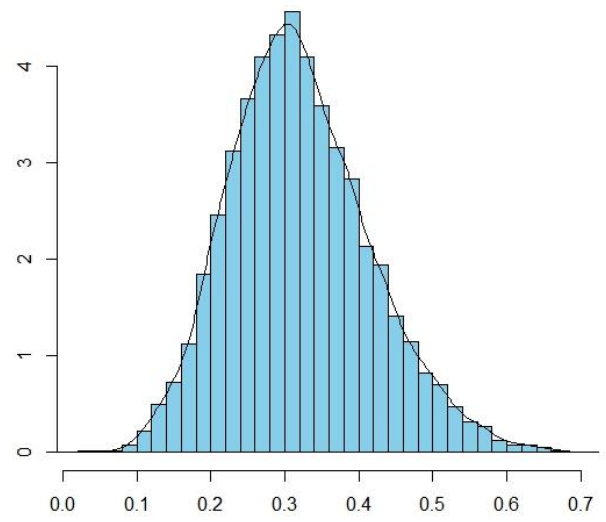
XS_VLDL_P



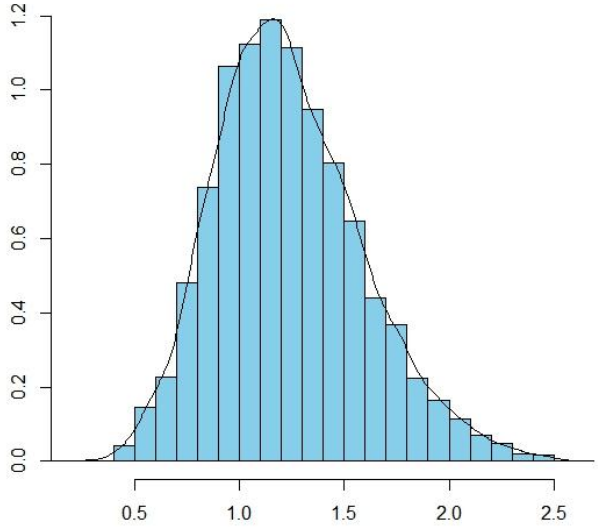
IDL_FC



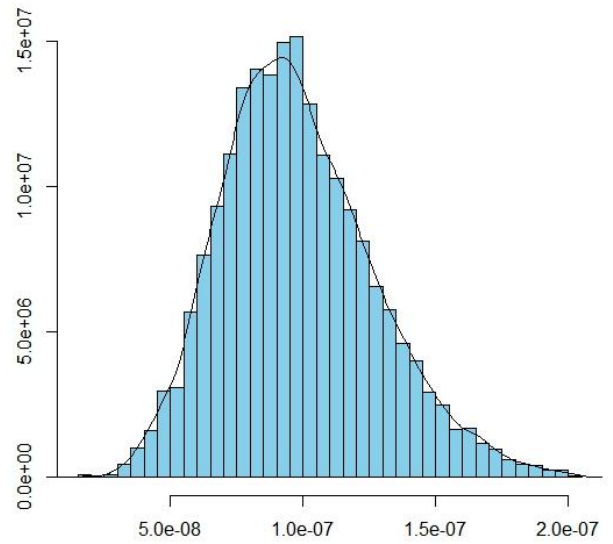
IDL_PL



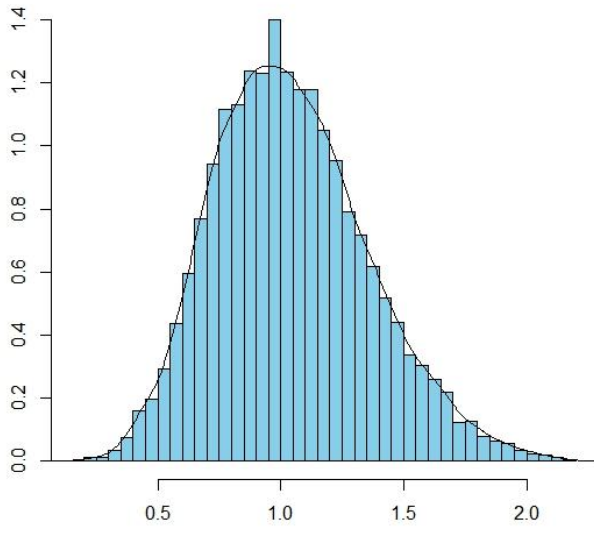
IDL_L



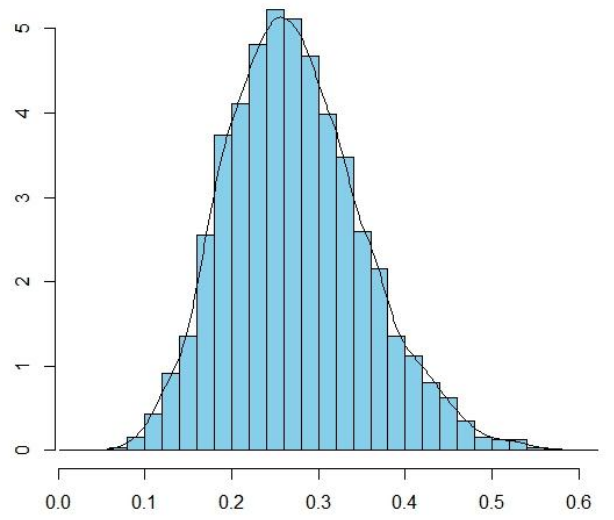
IDL_P



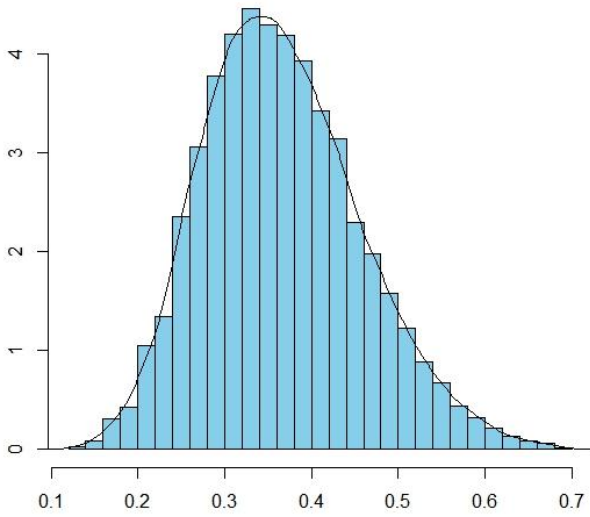
L_LDL_C



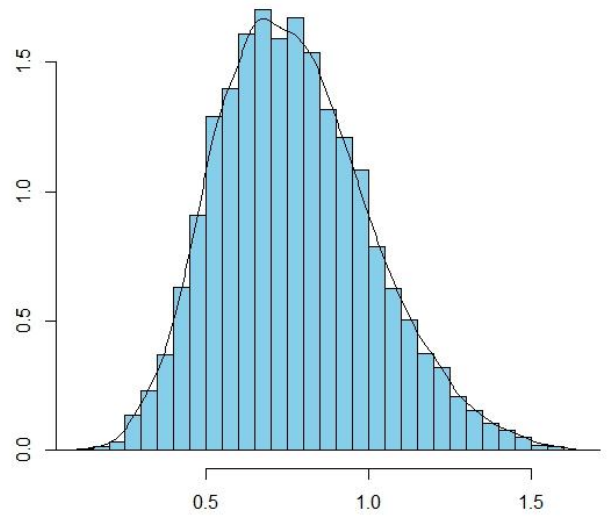
L_LDL_FC



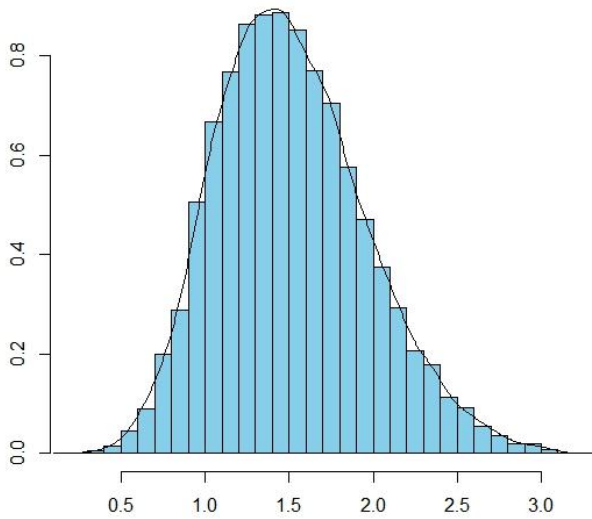
L_LDL_PL



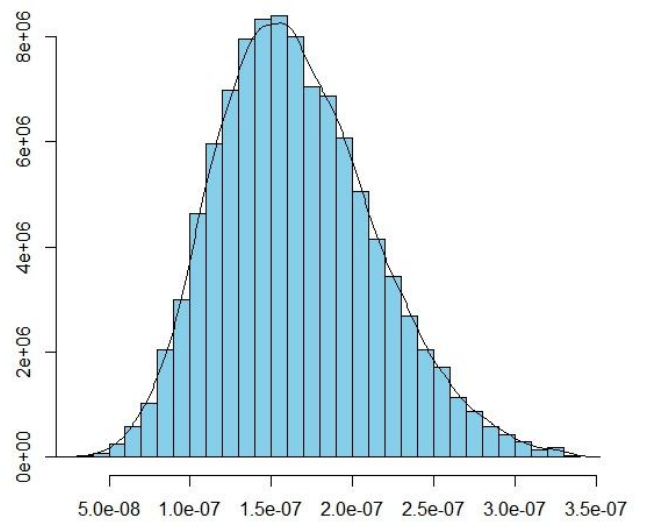
L_LDL_CE



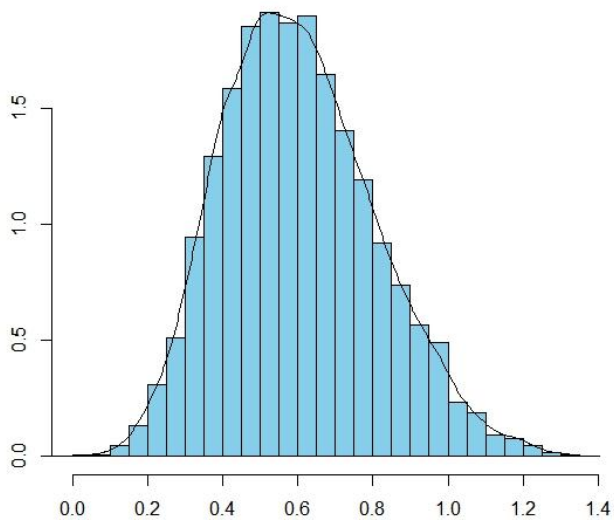
L_LDL_L



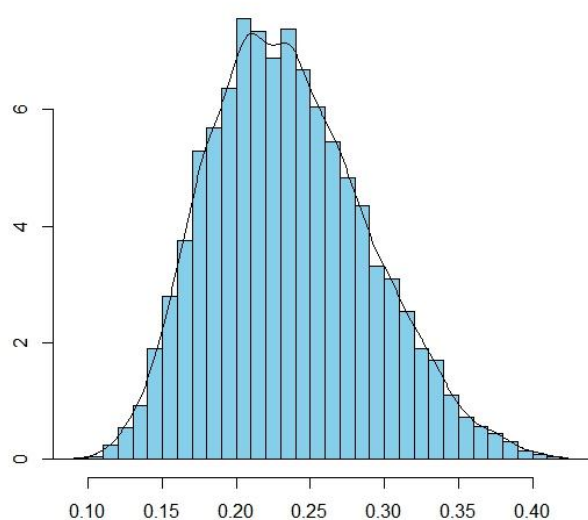
L_LDL_P



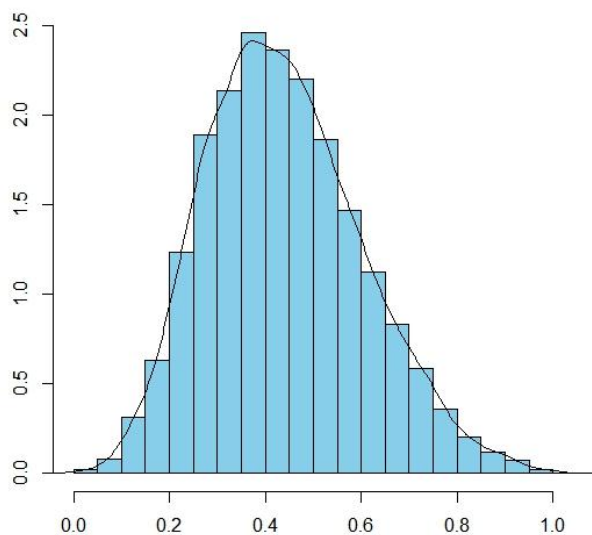
M_LDL_C



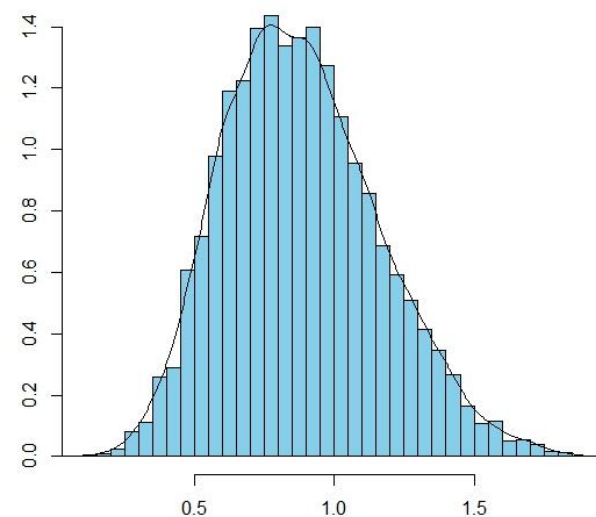
M_LDL_PL



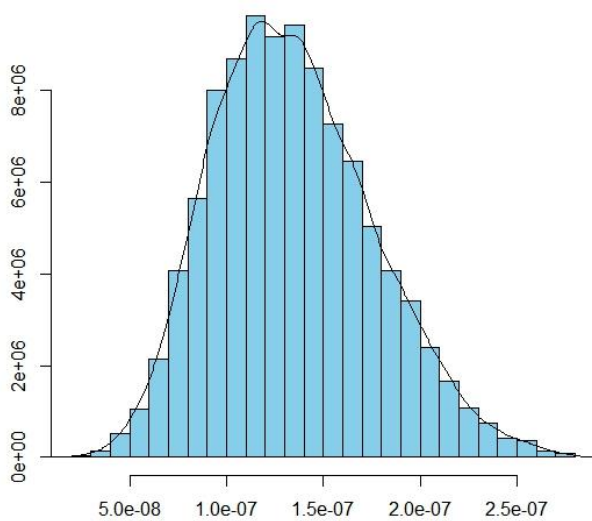
M_LDL_CE



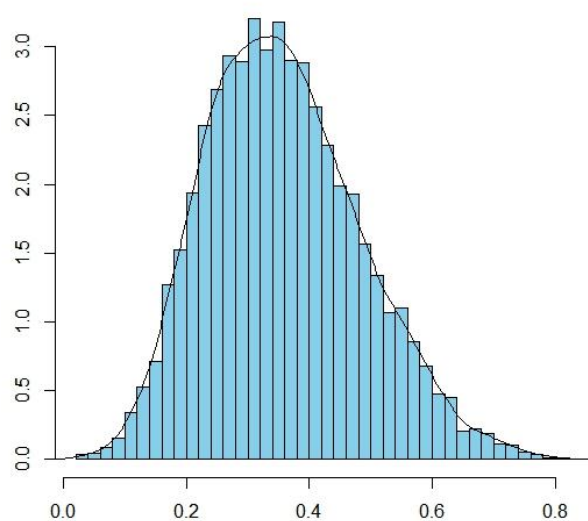
M_LDL_L



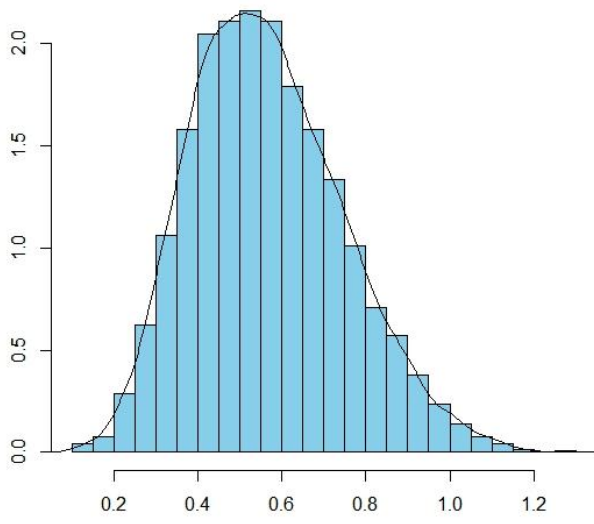
M_LDL_P



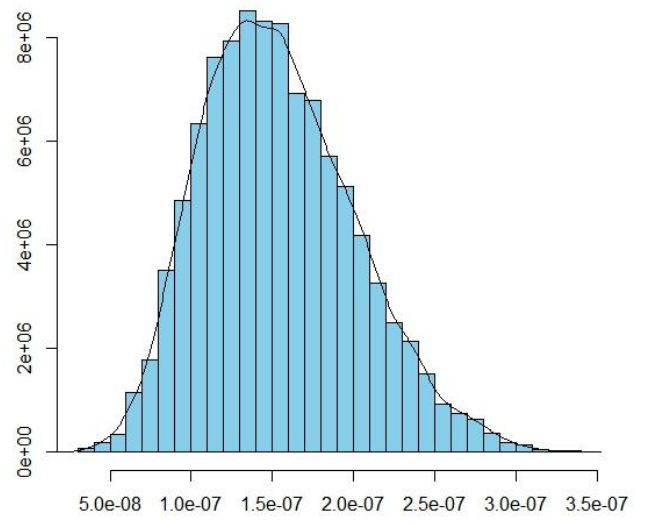
S_LDL_C



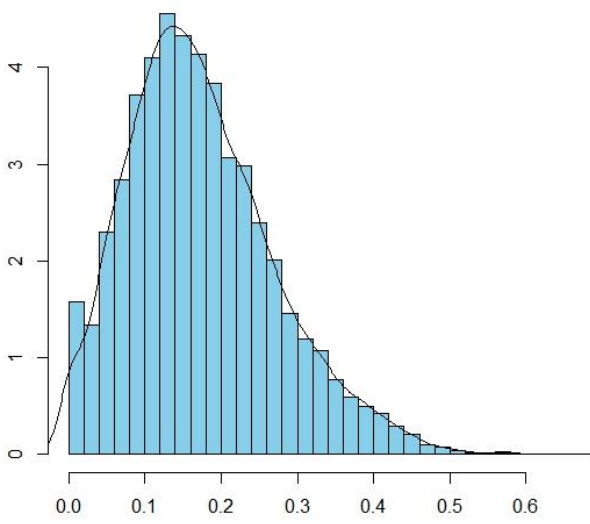
S_LDL_L



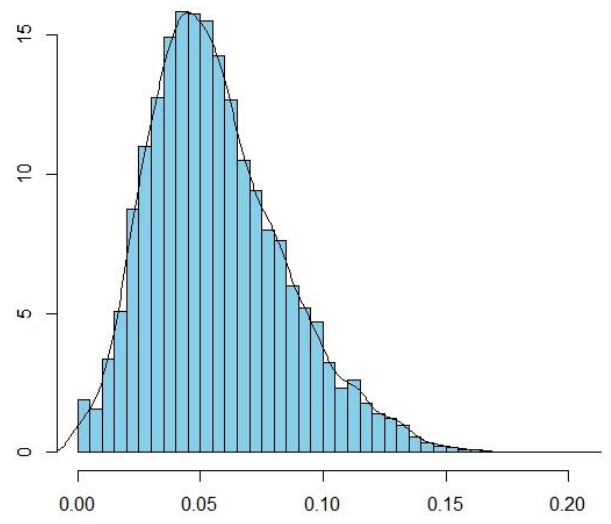
S_LDL_P



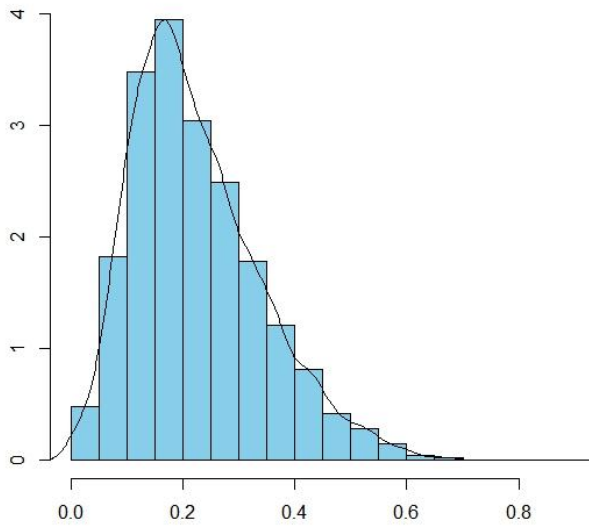
XL_HDL_C



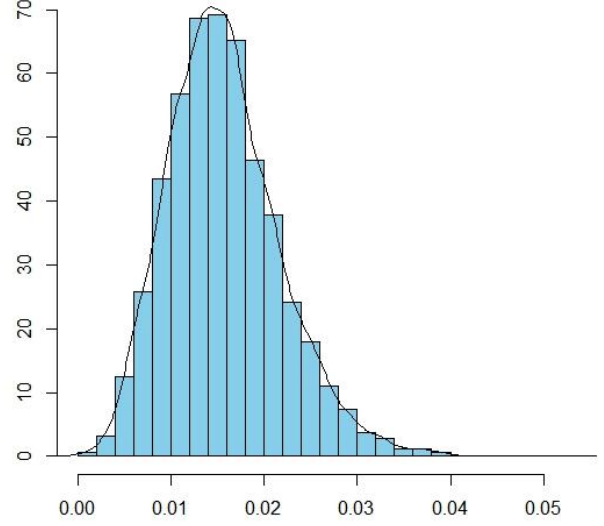
XL_HDL_FC



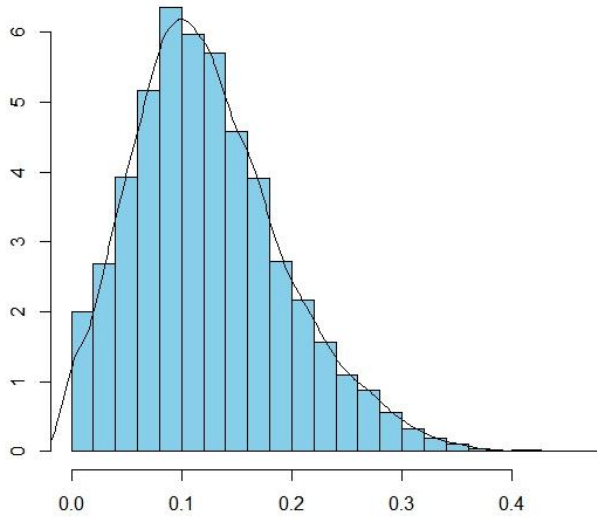
XL_HDL_PL



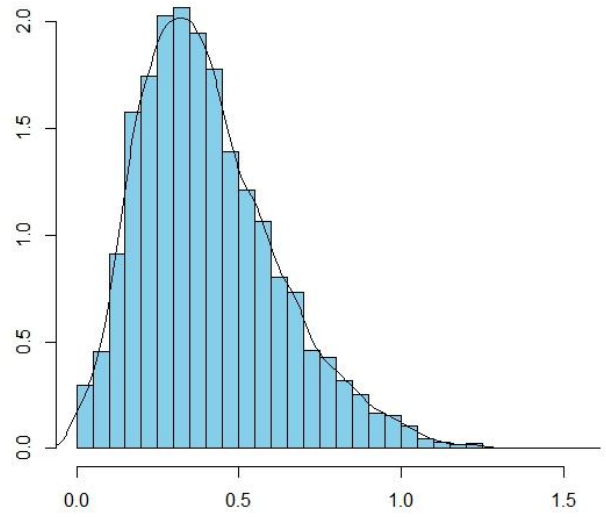
XL_HDL_TG



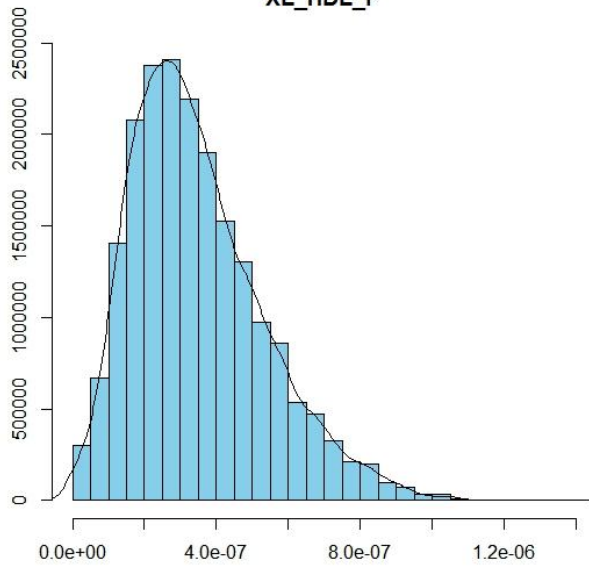
XL_HDL_CE



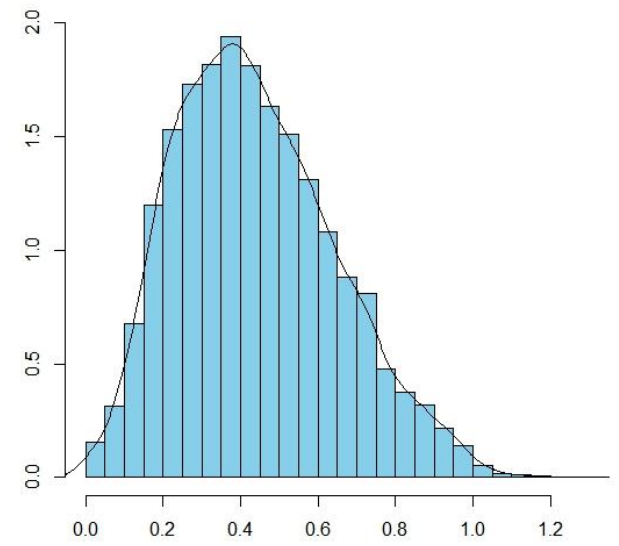
XL_HDL_L



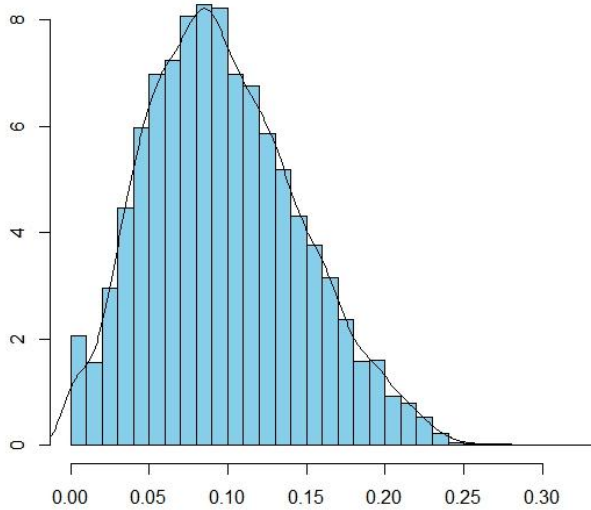
XL_HDL_P



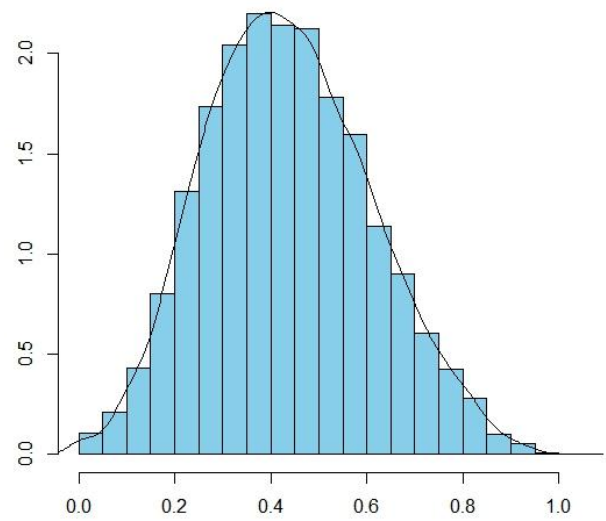
L_HDL_C



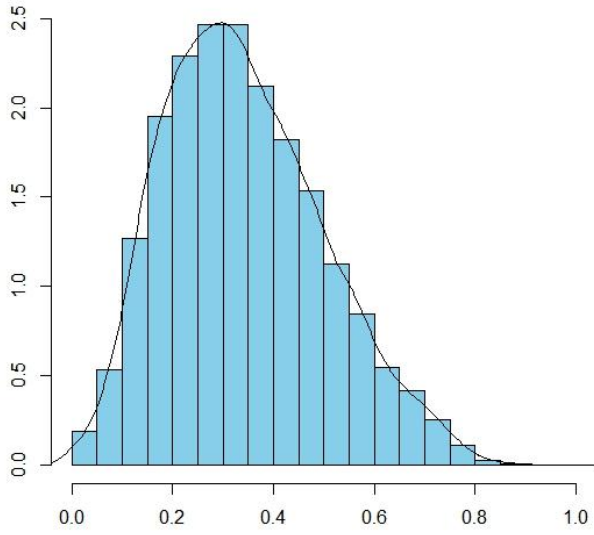
L_HDL_FC



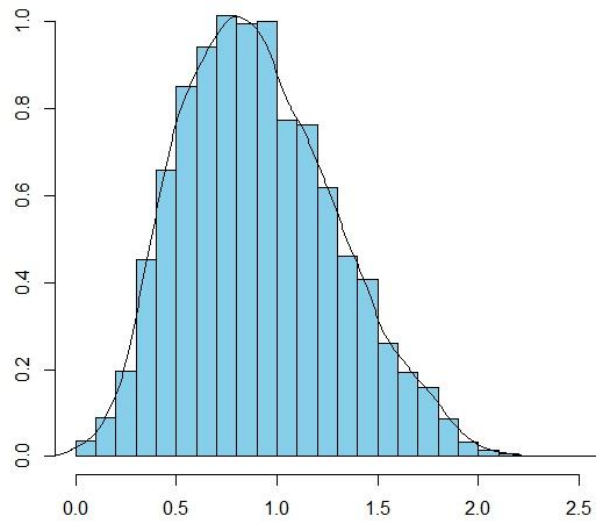
L_HDL_PL



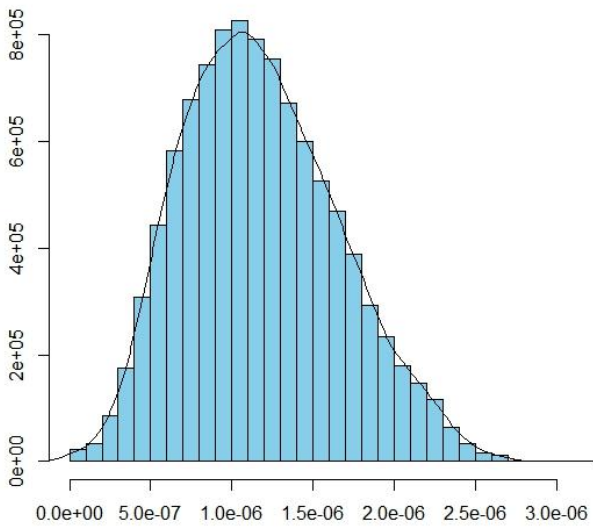
L_HDL_CE



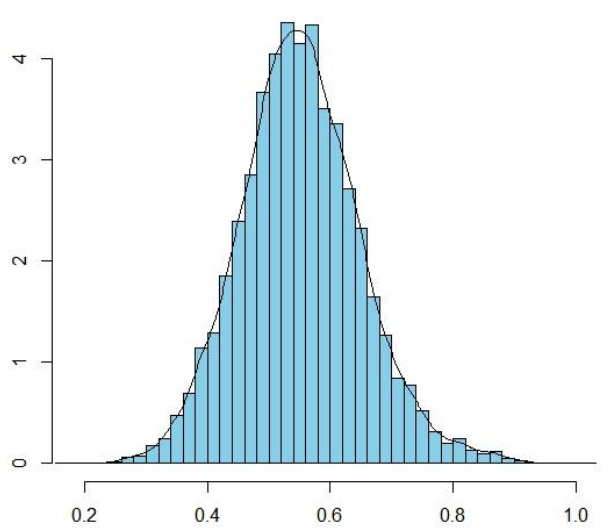
L_HDL_L



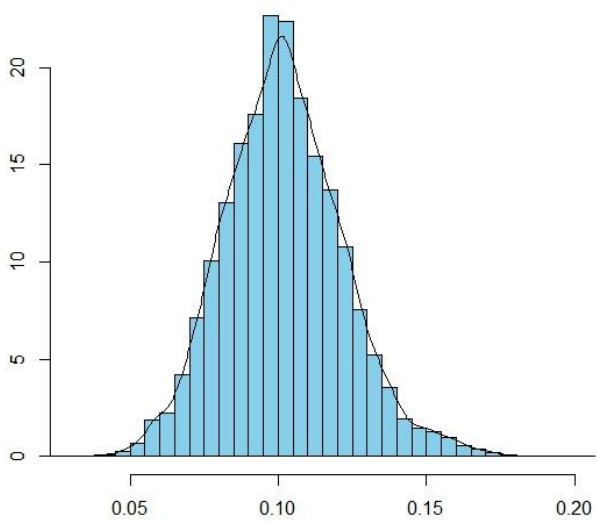
L_HDL_P



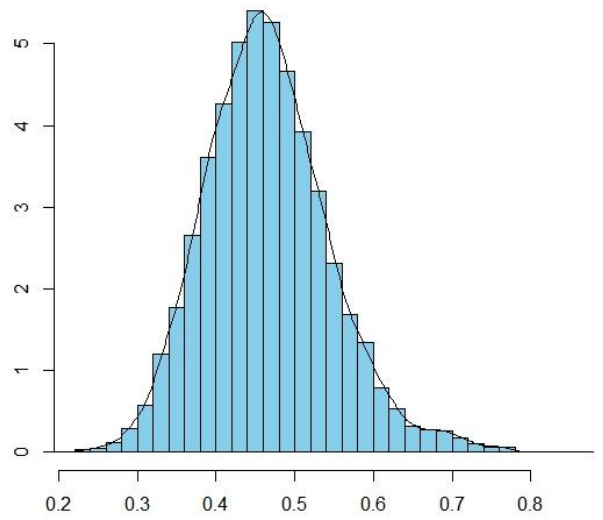
M_HDL_C



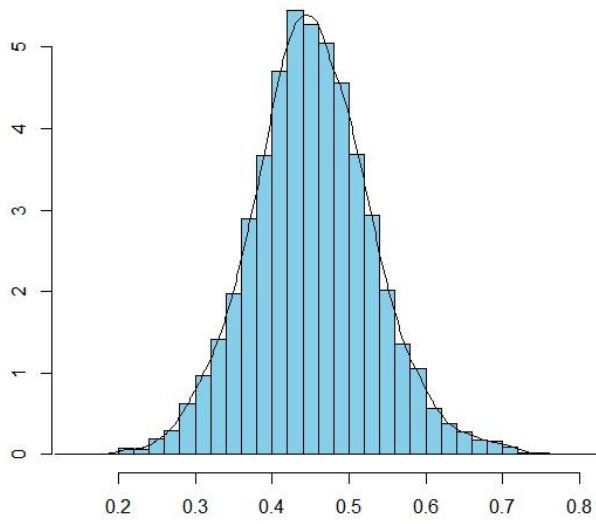
M_HDL_FC



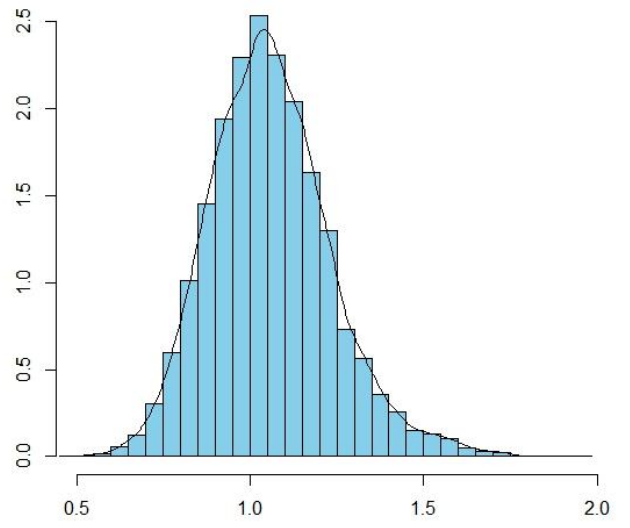
M_HDL_PL



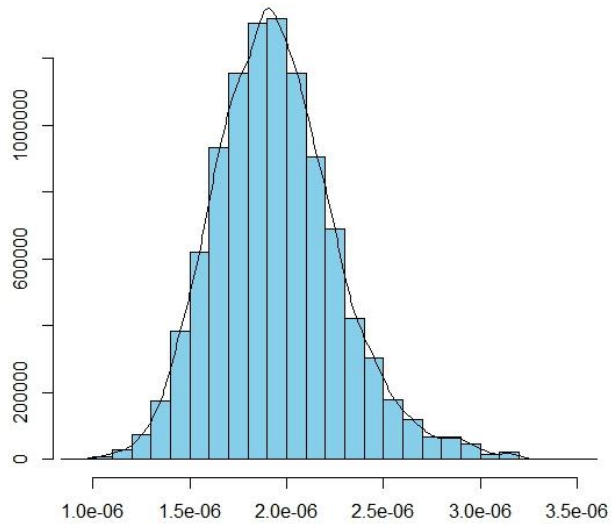
M_HDL_CE



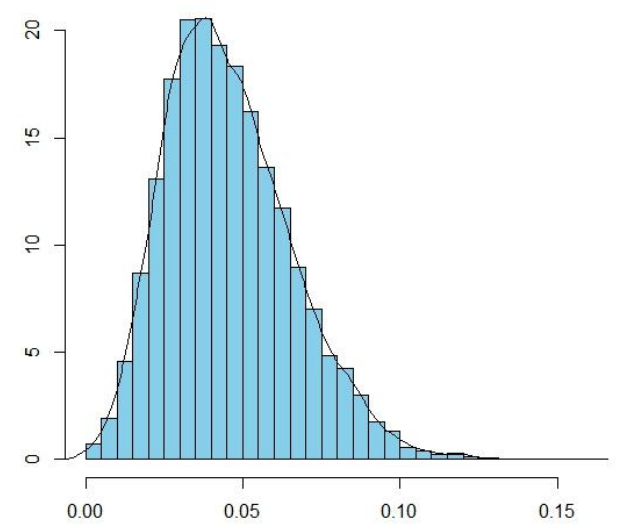
M_HDL_L



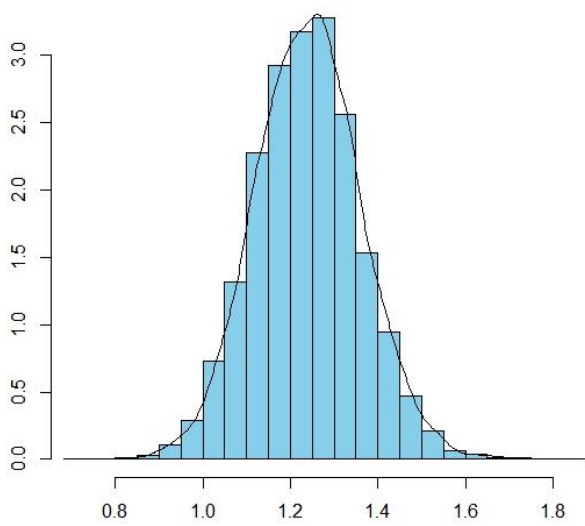
M_HDL_P



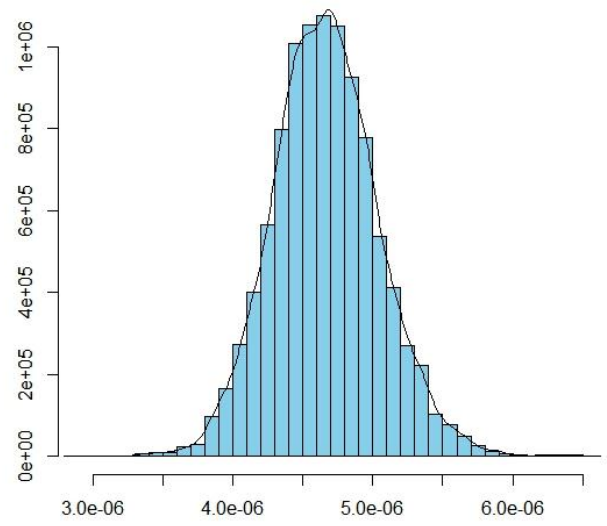
S_HDL_TG



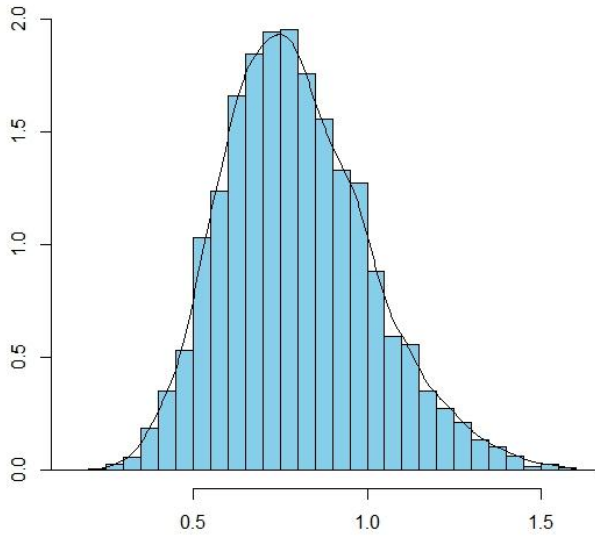
S_HDL_L



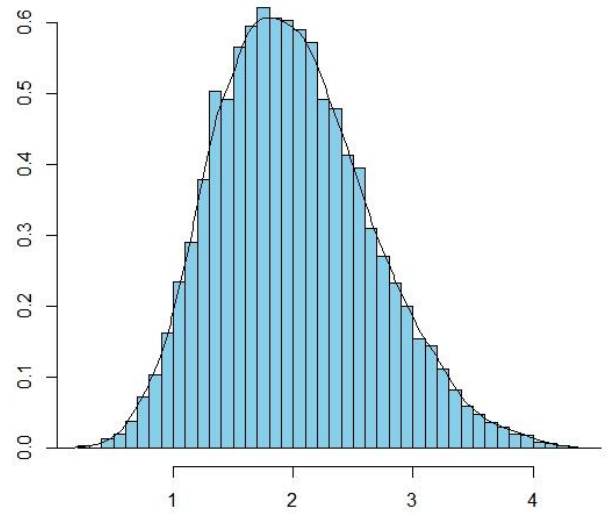
S_HDL_P



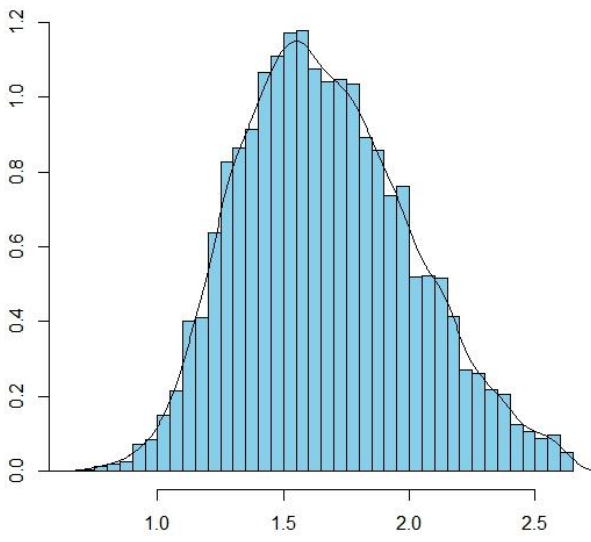
IDL_C



LDL_C



HDL_C



Lisa 2. Kasutatud statistilistel meetodikatel saadud tulemused toiduainete lõikes

Järgnevas tabelis on välja toodud olulisusnivool 0,01 lineaarsel regressioonanalüüsil statistiliselt oluliseks osutunud lipoproteiinid ning peakomponent- ja faktoranalüüsil saadud lipoproteiinide peakomponentide ja faktorite numbrid toiduainete lõikes.

Lineaarse regressioonanalüüsil saadud seosed on järjestatud t-statistikute väärtuste järgi.

	Lineaarne regressioon		Peakomponentanalüüs		Faktoranalüüs	
	Negatiivne seos	Positiivne seos	Negatiivne seos	Positiivne seos	Negatiivne seos	Positiivne seos
Kohv	M_HDL_CE M_HDL_C M_HDL_L	XL_HDL_P M_VLDL_PL M_VLDL_FC XL_HDL_L XL_HDL_C M_VLDL_C XL_HDL_FC XL_HDL_CE S_VLDL_TG M_VLDL_CE S_HDL_TG S_VLDL_P S_VLDL_L S_VLDL_PL S_VLDL_FC S_VLDL_C XL_HDL_TG S_LDL_L S_LDL_P S_LDL_C XS_VLDL_TG M_LDL_PL M_LDL_C M_LDL_CE M_LDL_L L_LDL_FC LDL_C M_LDL_P IDL_PL L_LDL_PL L_LDL_C L_LDL_CE IDL_FC L_LDL_L	1, 2, 4, 5		4	2, 3, 5

		L_LDL_P IDL_C IDL_L XS_VLDL_PL IDL_P XS_VLDL_P XS_VLDL_L				
Tee	M_HDL_PL M_HDL_P S_HDL_P M_HDL_L M_HDL_FC S_HDL_L M_HDL_C M_HDL_CE		4		4	
Leib						
Sai	L_HDL_PL L_HDL_P L_HDL_L L_HDL_FC	XL_VLDL_PL L_VLDL_PL XXL_VLDL_P				
Kartul						
Riis, makaronid	S_LDL_P S_LDL_L S_LDL_C M_LDL_P M_LDL_L M_LDL_PL M_LDL_C LDL_C M_LDL_CE L_LDL_P L_LDL_L L_LDL_CE L_LDL_PL L_LDL_C S_HDL_L S_HDL_P L_LDL_FC XL_HDL_CE HDL_C XL_HDL_C IDL_C IDL_L			2, 3	2, 6	
Puder, müsli, helbed	S_HDL_L S_HDL_P M_HDL_PL M_HDL_P M_HDL_L S_LDL_P M_HDL_FC M_LDL_PL		4, 5, 6	1, 2, 3	2, 4, 6	

	M_LDL_P S_LDL_L M_LDL_L S_LDL_C M_LDL_CE M_LDL_C LDL_C M_HDL_C L_LDL_P L_LDL_PL L_LDL_L L_LDL_CE L_LDL_C M_HDL_CE L_LDL_FC S_HDL_TG IDL_P IDL_L XS_VLDL_TG IDL_PL HDL_C IDL_C					
Piima- tooted	S_HDL_L S_HDL_P M_LDL_PL M_LDL_CE M_LDL_C M_LDL_L M_LDL_P L_LDL_CE LDL_C L_LDL_C S_LDL_C S_LDL_P L_LDL_L L_LDL_P S_HDL_TG		5, 6		6	
Kala	L_VLDL_PL M_VLDL_PL M_VLDL_TG M_VLDL_P M_VLDL_L M_VLDL_FC L_VLDL_L L_VLDL_TG L_VLDL_P L_VLDL_FC L_VLDL_C L_VLDL_CE M_VLDL_C XL_VLDL_PL XXL_VLDL_P	L_HDL_PL XL_HDL_PL L_HDL_P L_HDL_FC L_HDL_L XL_HDL_FC HDL_C L_HDL_C L_HDL_CE XL_HDL_P XL_HDL_L XL_HDL_C XL_HDL_CE	2, 3, 4	1, 6		3

	XL_VLDL_TG XL_VLDL_L XXL_VLDL_L XL_VLDL_P XXL_VLDL_PL S_HDL_TG M_VLDL_CE S_VLDL_TG					
Liha		S_LDL_P M_HDL_L M_HDL_FC M_HDL_P M_HDL_PL S_HDL_P S_HDL_L		4, 6		4, 6
Liha- produktid	XL_HDL_CE XL_HDL_C	S_HDL_P S_HDL_L		4		4
Värske juurvili				6		
Keedetud juurvili				6		5
Värsked puuviljad, marjad	XXL_VLDL_P XXL_VLDL_L XL_VLDL_PL L_VLDL_PL XL_VLDL_P XXL_VLDL_PL L_VLDL_P XL_VLDL_L L_VLDL_TG L_VLDL_L XL_VLDL_TG L_VLDL_FC L_VLDL_C M_VLDL_PL M_VLDL_C M_VLDL_CE M_VLDL_FC M_VLDL_L M_VLDL_P M_VLDL_TG L_VLDL_CE			1		
Kompotid, keedised	S_HDL_L S_HDL_P M_HDL_PL M_HDL_P		4, 6		4, 6	
Maiustused	S_HDL_TG S_HDL_L S_HDL_P		6			
Karastus- jook	L_HDL_FC XL_HDL_PL XL_HDL_P	L_VLDL_CE S_VLDL_L M_VLDL_C	1	4, 5	3	4

	L_HDL_C L_HDL_L L_HDL_P L_HDL_CE XL_HDL_L L_HDL_PL XL_HDL_FC XL_HDL_CE XL_HDL_C	S_VLDL_P L_VLDL_C XL_VLDL_L XXL_VLDL_P XL_VLDL_PL XL_VLDL_P XL_VLDL_TG L_VLDL_FC S_VLDL_TG M_VLDL_FC S_HDL_TG L_VLDL_P L_VLDL_L M_VLDL_L L_VLDL_TG M_VLDL_PL M_VLDL_P L_VLDL_PL M_VLDL_TG S_HDL_L S_HDL_P				
Muna		XL_HDL_TG XL_HDL_L XL_HDL_P XL_HDL_PL L_HDL_FC L_HDL_C HDL_C L_HDL_CE M_HDL_FC L_HDL_PL L_HDL_L L_HDL_P M_HDL_PL M_HDL_P M_HDL_CE M_HDL_L M_HDL_C	3			3, 4

Lisa 3. Lipoproteiinide korrelatsioonid esimese kuue peakomponendiga

	P1	P2	P3	P4	P5	P6
XXL_VLDL_PL	-0.111	0.129	-0.124			-0.260
XXL_VLDL_L	-0.117	0.125	-0.120			-0.220
XXL_VLDL_P	-0.121	0.119	-0.123			-0.138
XL_VLDL_PL	-0.123	0.121	-0.115			-0.173
XL_VLDL_TG	-0.121	0.130	-0.111			-0.134
XL_VLDL_L	-0.123	0.127	-0.111			-0.139
XL_VLDL_P	-0.124	0.124	-0.109			-0.141
L_VLDL_C	-0.135	0.114				
L_VLDL_FC	-0.132	0.117				
L_VLDL_PL	-0.135	0.115				
L_VLDL_TG	-0.129	0.126				
L_VLDL_CE	-0.138	0.109				
L_VLDL_L	-0.131	0.122				
L_VLDL_P	-0.132	0.120				
M_VLDL_C	-0.152					
M_VLDL_FC	-0.146					
M_VLDL_PL	-0.148					
M_VLDL_TG	-0.138	0.116				
M_VLDL_CE	-0.156				-0.110	
M_VLDL_L	-0.144	0.102				
M_VLDL_P	-0.142	0.106				
S_VLDL_C	-0.158					
S_VLDL_FC	-0.160				-0.116	0.128
S_VLDL_PL	-0.159				-0.140	0.148
S_VLDL_TG	-0.152					0.130
S_VLDL_L	-0.160				-0.102	0.121
S_VLDL_P	-0.158				-0.101	0.128
XS_VLDL_PL		-0.188			-0.238	-0.111
XS_VLDL_TG	-0.148				-0.242	0.196
XS_VLDL_L	-0.104	-0.149			-0.297	
XS_VLDL_P	-0.113	-0.136			-0.300	
IDL_FC		-0.221				-0.190
IDL_PL		-0.202				-0.161
IDL_L		-0.196				-0.141
IDL_P		-0.192				-0.128
L_LDL_C	-0.108	-0.175			0.114	
L_LDL_FC		-0.205				
L_LDL_PL	-0.109	-0.175				
L_LDL_CE	-0.116	-0.162			0.117	
L_LDL_L	-0.109	-0.177				
L_LDL_P	-0.109	-0.177				
M_LDL_C	-0.122	-0.149			0.153	
M_LDL_PL	-0.140	-0.110			0.157	
M_LDL_CE	-0.124	-0.149			0.130	
M_LDL_L	-0.126	-0.145			0.145	
M_LDL_P	-0.128	-0.142			0.133	
S_LDL_C	-0.127	-0.139			0.148	
S_LDL_L	-0.136	-0.114			0.149	
S_LDL_P	-0.140	-0.105			0.128	
XL_HDL_C			-0.243	-0.189		0.145

XL_HDL_FC		-0.237	-0.174		0.109
XL_HDL_PL		-0.218	-0.147		
XL_HDL_TG		-0.203		-0.276	
XL_HDL_CE		-0.237	-0.182		0.161
XL_HDL_L		-0.241	-0.171		0.106
XL_HDL_P		-0.234	-0.165		0.101
L_HDL_C	0.113	-0.203			
L_HDL_FC	0.116	-0.178			
L_HDL_PL	0.112	-0.176		-0.107	
L_HDL_CE	0.109	-0.213			
L_HDL_L	0.112	-0.193			
L_HDL_P	0.111	-0.189		-0.103	
M_HDL_C		-0.167	0.289		-0.139
M_HDL_FC		-0.149	0.294		
M_HDL_PL		-0.160	0.332		
M_HDL_CE		-0.170	0.273	0.102	-0.170
M_HDL_L		-0.166	0.335		
M_HDL_P		-0.160	0.347		
S_HDL_TG	-0.145			-0.186	0.254
S_HDL_L			0.310	0.234	0.366
S_HDL_P			0.302	0.233	0.314
IDL_C		-0.198			-0.148
LDL_C	-0.116	-0.162		0.135	
HDL_C			-0.213		0.111

Lisa 4. Lipoproteiinide korrelatsioonid kuue faktoriga

	F1	F2	F3	F4	F5	F6
XXL_VLDL_PL	0.965				-0.173	-0.146
XXL_VLDL_L	0.976				-0.150	-0.115
XXL_VLDL_P	0.981				-0.107	
XL_VLDL_PL	0.988				-0.123	
XL_VLDL_TG	0.990				-0.115	
XL_VLDL_L	0.991				-0.105	
XL_VLDL_P	0.990				-0.109	
L_VLDL_C	0.986		-0.110			
L_VLDL_FC	0.991					
L_VLDL_PL	0.987		-0.124			
L_VLDL_TG	0.990		-0.110			
L_VLDL_CE	0.973	0.107	-0.129			
L_VLDL_L	0.990		-0.112			
L_VLDL_P	0.990		-0.107			
M_VLDL_C	0.921	0.283	-0.184		0.153	
M_VLDL_FC	0.954	0.186	-0.183		0.104	
M_VLDL_PL	0.939	0.204	-0.217		0.129	
M_VLDL_TG	0.966		-0.194			0.102
M_VLDL_CE	0.868	0.381	-0.181	-0.128	0.206	
M_VLDL_L	0.956	0.152	-0.199		0.101	
M_VLDL_P	0.959	0.129	-0.199			
S_VLDL_C	0.592	0.674	-0.262	-0.226	0.235	
S_VLDL_FC	0.714	0.544	-0.273	-0.168	0.271	0.100
S_VLDL_PL	0.720	0.512	-0.304	-0.141	0.304	0.102
S_VLDL_TG	0.896	0.263	-0.255		0.200	0.113
S_VLDL_L	0.796	0.448	-0.276	-0.143	0.238	0.100
S_VLDL_P	0.832	0.388	-0.274	-0.122	0.233	0.106
XS_VLDL_PL	-0.102	0.890	-0.106		0.377	-0.186
XS_VLDL_TG	0.570	0.620	-0.259		0.456	
XS_VLDL_L	0.104	0.832	-0.163		0.476	-0.164
XS_VLDL_P	0.173	0.817	-0.185		0.491	-0.126
IDL_FC	-0.267	0.932			0.135	-0.162
IDL_PL		0.964			0.156	-0.141
IDL_L		0.979			0.144	-0.128
IDL_P		0.970			0.190	-0.133
L_LDL_C		0.988				
L_LDL_FC	-0.108	0.992				
L_LDL_PL		0.990				
L_LDL_CE	0.159	0.975		-0.107		
L_LDL_L		0.993				
L_LDL_P	0.105	0.992				
M_LDL_C	0.222	0.954				0.144
M_LDL_PL	0.409	0.883				0.125
M_LDL_CE	0.226	0.954				0.146
M_LDL_L	0.261	0.949				0.133
M_LDL_P	0.282	0.946				0.125
S_LDL_C	0.310	0.933				0.139
S_LDL_L	0.434	0.884				0.123
S_LDL_P	0.483	0.860				0.139
XL_HDL_C			0.968		-0.109	

XL_HDL_FC			0.985			
XL_HDL_PL	-0.199		0.961	0.106		
XL_HDL_TG	0.610	0.426	0.501		0.367	-0.178
XL_HDL_CE			0.959			
XL_HDL_L			0.992			
XL_HDL_P			0.988			
L_HDL_C	-0.319	-0.234	0.875	0.268		
L_HDL_FC	-0.429	-0.159	0.844	0.251		
L_HDL_PL	-0.453		0.756	0.435		-0.105
L_HDL_CE	-0.261	-0.259	0.885	0.268		
L_HDL_L	-0.387	-0.153	0.828	0.360		
L_HDL_P	-0.403	-0.125	0.808	0.386		
M_HDL_C	-0.127	-0.268	0.312	0.881	-0.158	
M_HDL_FC	-0.284		0.324	0.892		
M_HDL_PL	-0.101		0.239	0.952	0.115	
M_HDL_CE		-0.351	0.308	0.841	-0.229	
M_HDL_L			0.243	0.963		
M_HDL_P			0.198	0.971	0.104	
S_HDL_TG	0.517	0.544	-0.448		0.386	0.120
S_HDL_L	0.241	0.448	-0.446	0.504	-0.131	0.415
S_HDL_P	0.503	0.238	-0.387	0.521	-0.165	0.384
IDL_C		0.986				-0.101
LDL_C	0.173	0.973				
HDL_C	-0.317		0.806	0.476	-0.101	

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, **Kristi Helekivi** (sünnikuupäev: 29.07.1991)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose „Toitumise seosed lipoproteiinidega Tartu Ülikooli Eesti Geenivaramu andmestikus: korreleeritud andmete mitmemõõtmeline analüüs“, mille juhendaja on Krista Fischer,
 - 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 06.05.2013