



Õpilaste Euroopa
statistikavõistlus 2019

Statistikauurimuste head ja vead

Ülevaade 2. vooru töödest

Ülle Kikas
žürii liige

Võistlusvoorude ülevaade

- Esimene voor: veebitestide lahendamine

- A (vanem vanuserühm) **84** võistkonda
- B (noorem vanuserühm) **100** võistkonda

- Teine voor: uurimistöö

- vooru pääses 35 (A) ja 40 (B) võistkonda
- **uurimuse esitas 27 (A) ja 32 (B) võistkonda**

Uurimus põhines Statistikaameti koostatud andmefailil Eesti leibkondade kulutuste ja tarbimise kohta (allikas: leibkonna eelarve uuring)

- Uurimistöö hindamine

- Vormistus (20%); Eesmärgid (20%); Analüüs (40%); Järeldused (20%)

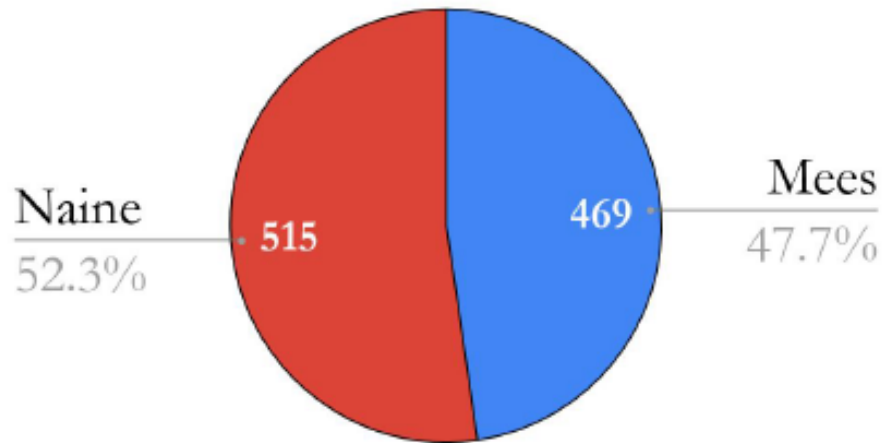
Vormistus

- Reeglina hea, nõutud ettekande osad olemas
 - Mõned tööd vormistatud nagu küsitluse kokkuvõtted
 - Teised vormistatud nagu uurimused
 - Viitamist vähe!
- Palju ilusaid esitlusi!



Tulemused

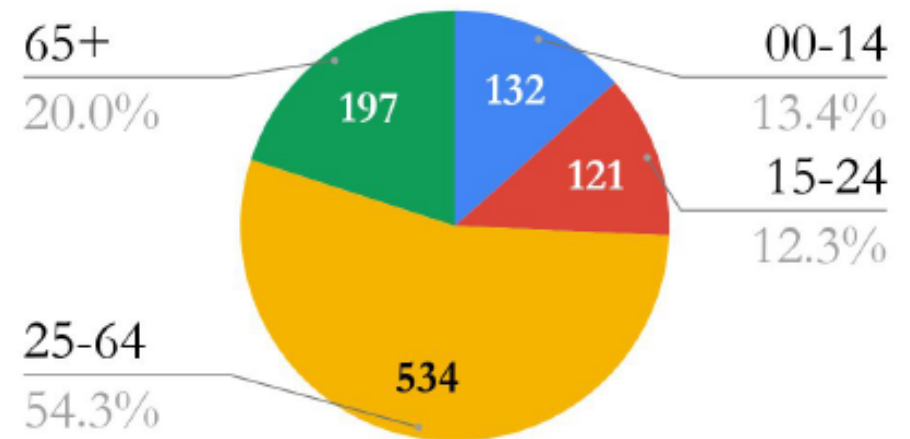
Küsitletute sugu %



Mehed ja naised on küsitluses (protsentuaalselt) üsnagi võrdselt esindatud.

Nutiajud

Vanusevahemikud

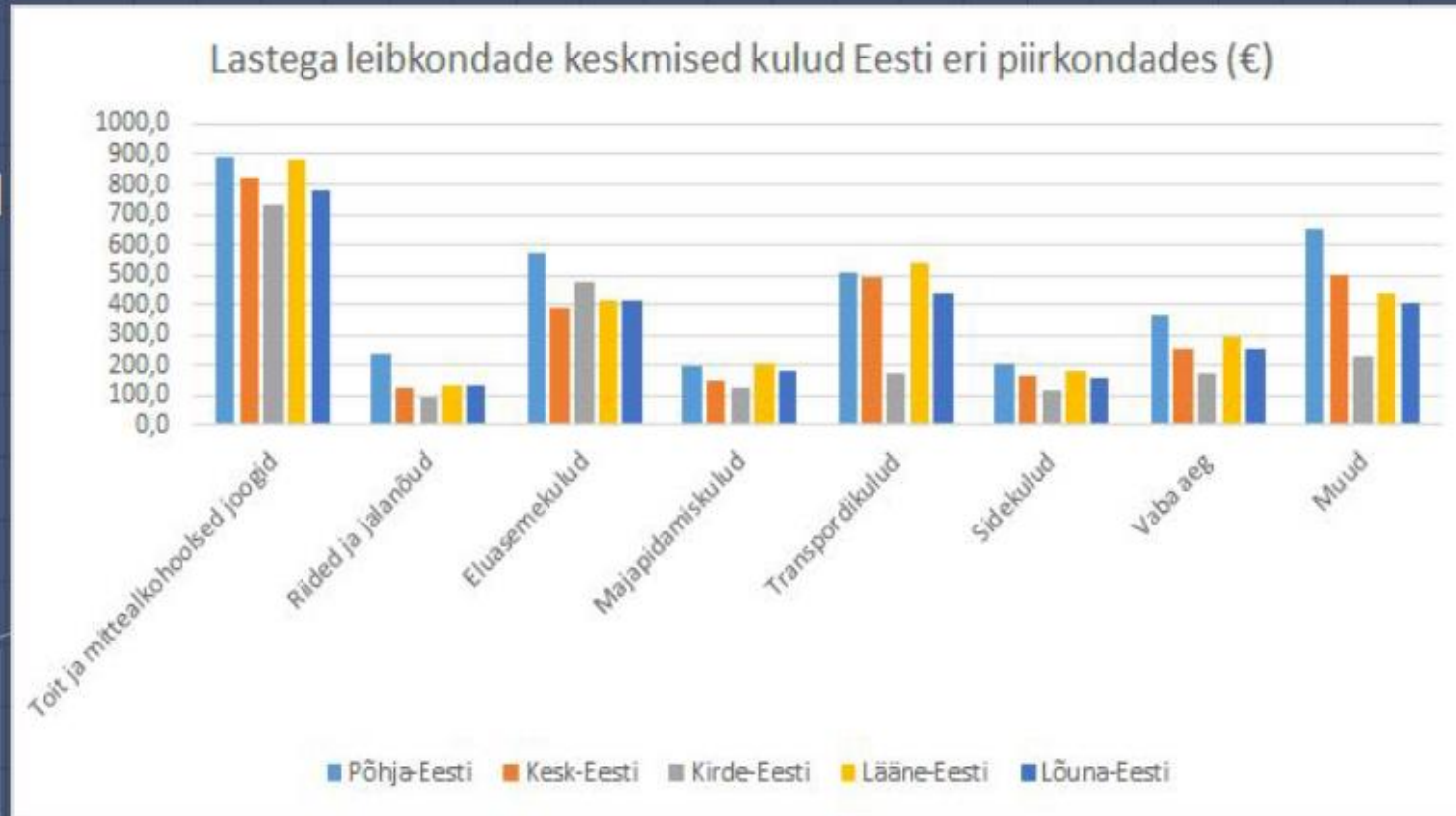


Enamik küsitletutest on **25 kuni 64 aastat** vana.



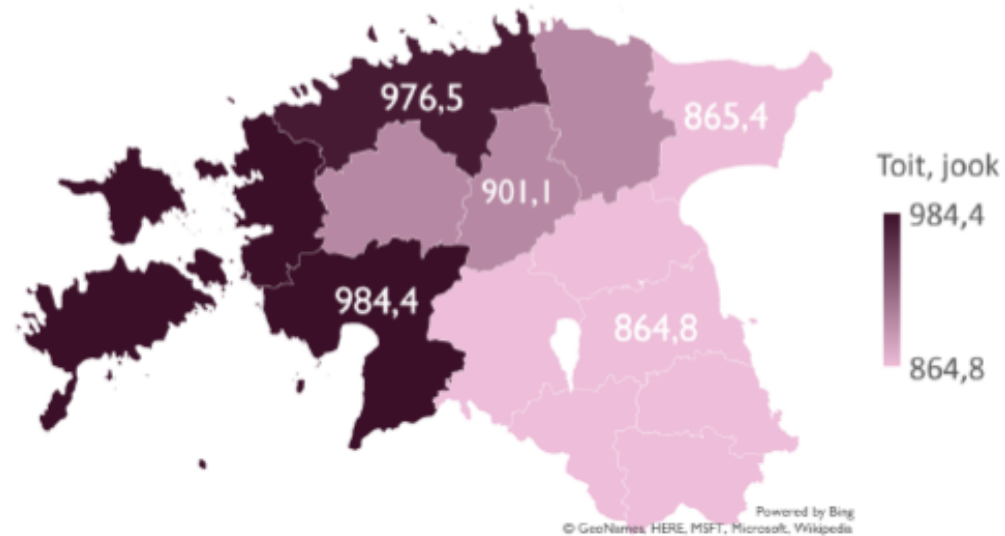
Keskmete kulude võrdlus piirkonniti lastega peredes

- Enim kulub raha toidule;
- suurimad kulutused on peamiselt Põhja-Eestis;
- kõige väiksemad kulud on Kirde-Eestis.



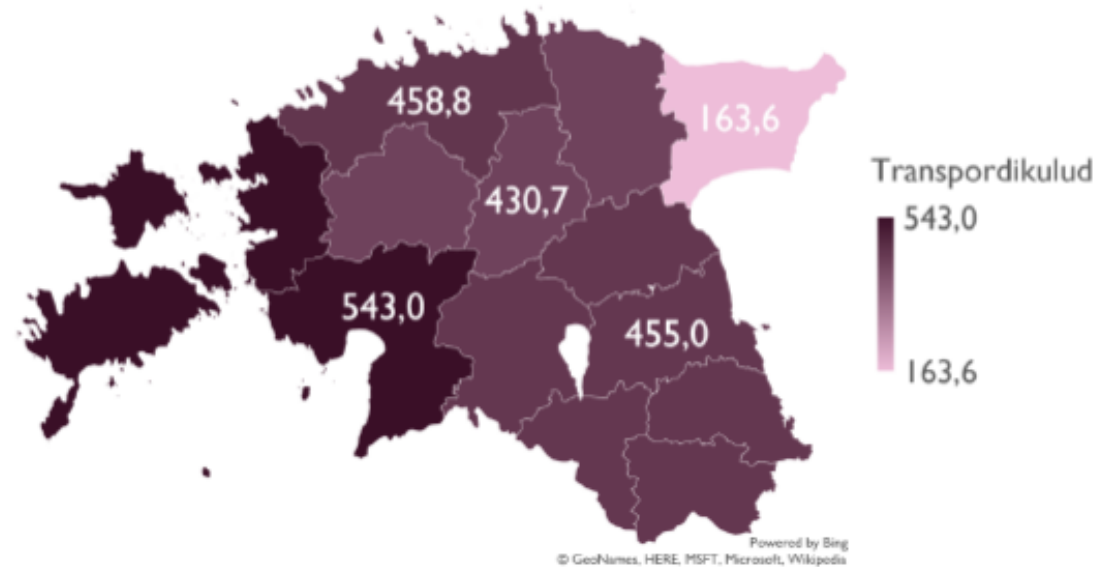


Kulutused toidule ja mittealkohoolsetele jookidele piirkonniti (€)



Kõige enam kulutavad toidule raha Lääne-Eestis elavad inimesed. Saartel on toidukaubad kallimad, kuna nende transpordiks kulub rohkem ressursse. Toidule kulutatakse kõige vähem Kagu-Eestis, kuna seal asub palju talusid ning seetõttu tarbitakse rohkem omatoodangut.

Kulutused transpordile piirkonniti (€)



Enim kulutakse raha transpordile Lääne-Eestis. Seal on asustus hajus ning igapäevane transport pikemate vahemaadega. Saarte elanikel kulub raha ka praamidele. Kõige vähem kulutatakse transpordiks Ida-Virumaal, kuna seal elavad inimesed peamiselt linnades ja ühistransport on hästi korraldatud.

Eesmärgid – huvitavad, mitmekesised, motiveerivad

- Kuidas erinevad lastega ja lasteta perede toimetulek ja kulutused;
- Kuidas erinevad toimetuleku hinnangud piirkonniti;
- Kuidas toimetuleku hinnangud seostuvad tegelike kulutustega;
- Kui palju kulutatakse kokku igas piirkonnas;
- Kuidas jaotuvad leibkonnad kuluklassidesse;
- Kuidas kodune keel ja eluase mõjutavad toimetulekut ja kulutusi;
- Kui suured on leibkonna kulutused eluasemele eri piirkondades;
- Millised leibkonna kulutused sõltuvad / ei sõltu toimetuleku hinnangust;
- Kas Põhja-Eesti leibkonnad kulutavad vabale ajale rohkem kui teistes piirkondades; ...jne

Uurimisküsimuste sõnastamine

Korrektne teaduslik stiil

- Kas ja kuidas mõjutab leibkonna elukoht Eestis leibkonna toimetulekut, kulusid, kulude struktuuri ning eluruumi tüüpi ja eluruumi kasutamise alust.

Vabam, pigem ajakirjanduslik stiil

- Kui palju kulutavad eestlased ja venelased aastas raha?
- Kuidas hindavad eestlased ja venelased oma toimetulekut?
- Kuidas erinevad eestlaste ja venelaste kulutused?
 - Terminid ja mõisted olgu täpsed - „kodune keel“, „kulutused leibkonnaliikme kohta“, et vältida valesti tõlgendatavaid järeldusi!
 - Kui on valitud vaba küsimuste stiil, siis tuleb töös mõisted lahti seletada!

Kas andmed sobivad püsitatud eesmärkide uurimiseks?

- Alati ei sobinud!
- Vähene süvenemine andmete kirjeldusse:
 - „Andmestik sisaldab
 - * **isikut** iseloomustavaid tunnuseid (sugu, vanus)
 - * **leibkonda**, kuhu isik kuulub, iseloomustavaid tunnuseid (leibkonna tüüp, leibkonna toimetulek, eluruumi tüüp jm)
 - * **kulutusi leibkonnaliikme kohta** kululiikide järgi (toit, eluase, vaba aeg jm)“
- Selle andmestiku alusel **EI SAA UURIDA**:
 - mingi vanuserühma (laste, vanurite) kulutusi; tüdrukute ja poiste kulutusi.
 - SAAB UURIDA selle leibkonna kulutusi (leibkonnaliikme kohta), millesse poiss või tüdruk, laps või vanur kuulub.

Analüüs

- Kasutati koolis õpitud statistikafunktsioone ja diagramme:
 - Summa, keskmine, mood, mediaan, sagedustabel, korrelatsioonikordaja, variatsioonikordaja, hajuvus
 - Tulpdiagramm, sektordiagramm, joondiagramm, histogramm, karpdiagramm, korrelatsioonidiagramm (Python);
 - Valdavalt kirjeldav statistika, vähe seoste loomist
- Tarkvara: Excel, Google arvutustabelid, IBM SPSS Statistica viewer, Python,
- Heades töödes: diagrammid hästi valitud; selgitavad tekstid on informatiivsed ja hästi seotud eesmärgiga; arutelu on seotud kontekstiga, tekst ja diagrammid toetavad lõppjäreldesteni jõudmist;
- Kehvemates töödes: diagrammides või tabelite juures on selgitavat teksti vähe (loodetakse, et lugeja saab ise aru), diagrammide pealkirjad on formaalsed, analüüs kirjeldab ainult matemaatilist tulemust, ei ole seotud kontekstiga, jm.

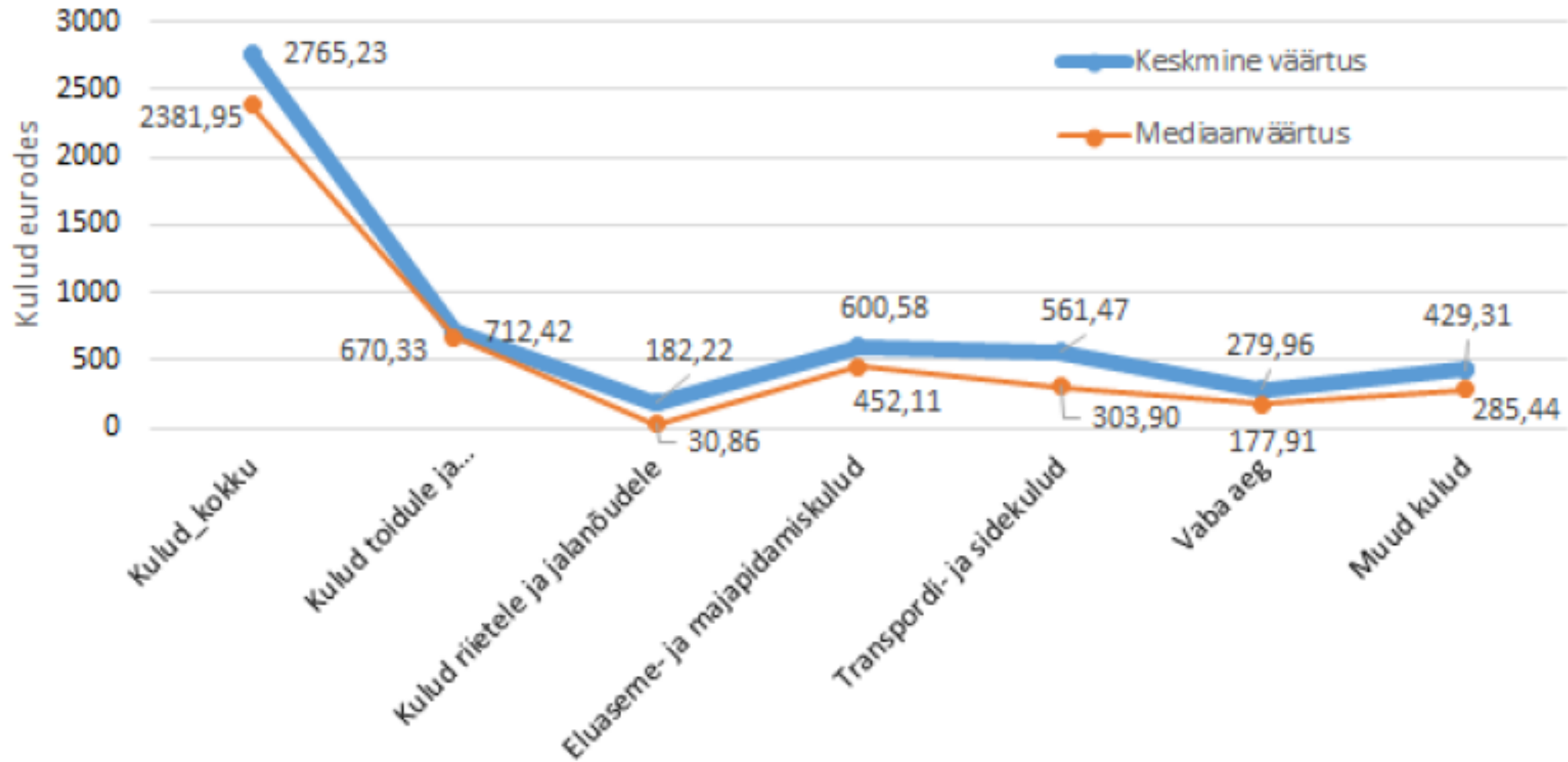
Näiteid valesti valitud diagrammi tüübist või esitusviisist

- Kulutuste kogusumma ja üksikud kulutuste liigid ühel diagrammil - **liiga lai skaala ulatus**, üksikute liikide erinevused halvasti nähtavad!
- Virnastatud tulpdiagramm **ei sobi** erinevates elamutüüpides toidule tehtud kulutuste kirjeldamiseks – need kulutused ei liitu!
- Sektordiagramm **ei sobi** piirkondade keskmiste kulutuste näitamiseks: need kulutused ei liitu ega moodusta tervikut!

Valitud on vale diagrammitüüp.



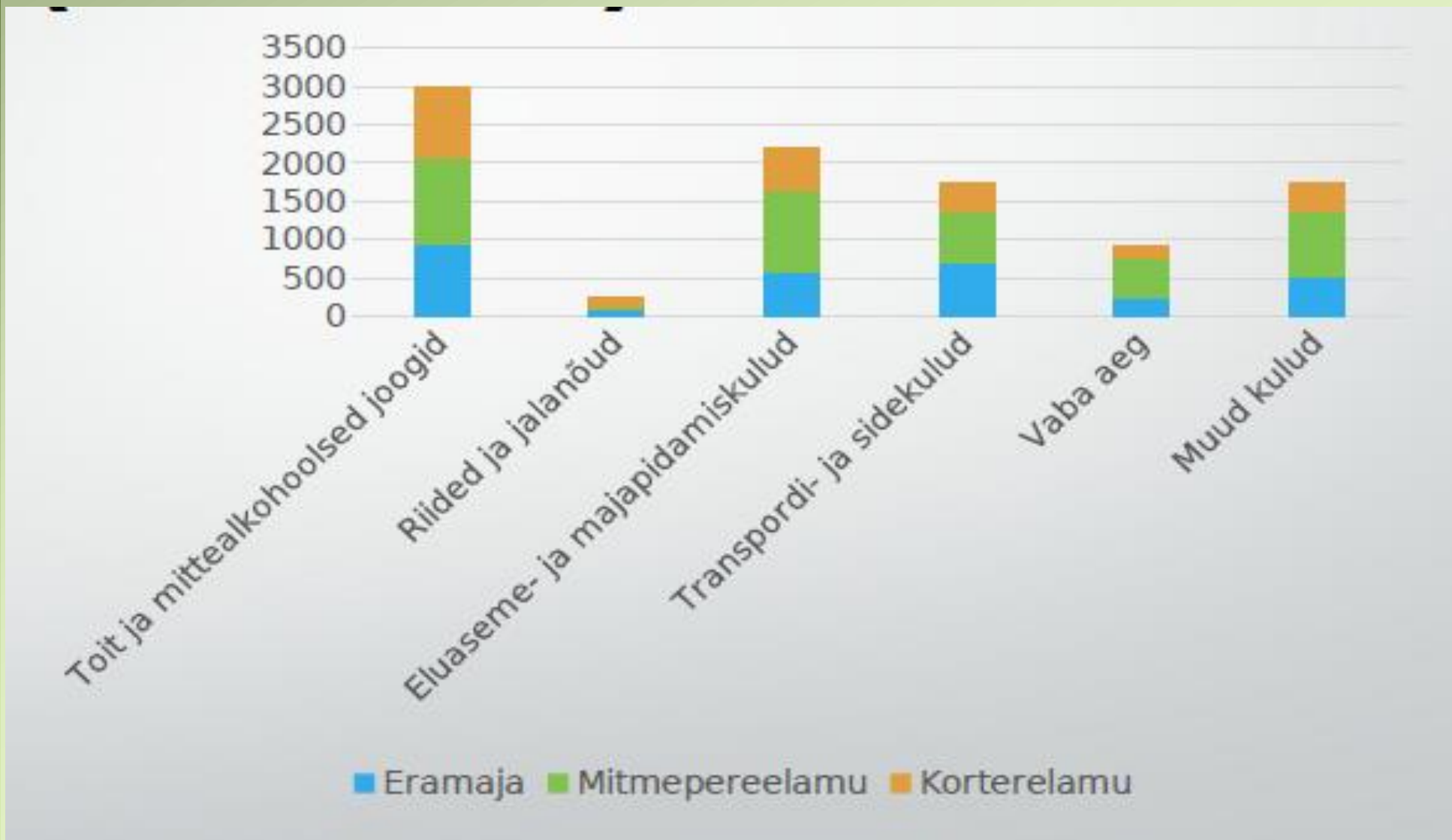
Kulude jaotus liigiti 00-14 aastaste hulgas



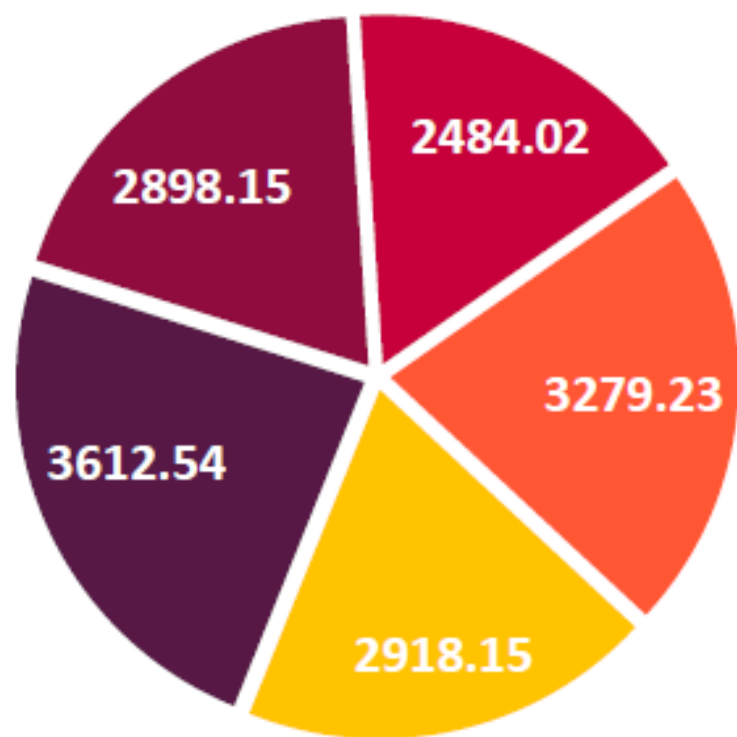
Valitud on vale diagrammitüüp



Leibkonnaliikme kulutused Kesk-Eestis eluruumi tüübi järgi



Kogukulutused leibkonnaliikme kohta aastas piirkonniti (€)



■ 1 Põhja-Eesti ■ 2 Kesk-Eesti ■ 3 Kirde-Eesti
■ 4 Lääne-Eesti ■ 5 Lõuna-Eesti

Valitud on vale diagrammitüüp



Analüüs: Kuidas tõlgendada suuremaid või väiksemaid kulutusi?

TULEMUSI ANALÜÜSIDES MÕTLE OLUKORD HÄSTI LÄBI JA ARVESTA, et

- kulutuste tegemise võimalust piirab leibkonna sissetulek;
- kulutused on esitatud leibkonnaliikme kohta;
- osa kululiike (toit, eluase) on nn sundkulud ja muid kulutusi saab teha siis, kui sundkulud on kaetud.

Näited ebaõigetest järeldustest

- „kogukulutused 0-14-aastastele lastele on kõige väiksemad võrreldes teiste vanuserühmadega“. Tulemus väljendab pigem seda, et peredes, kus on lapsi, on rohkem pereliikmeid, mistõttu iga pereliikme kohta tehtavad kulud on väiksemad, st lastega perede toimetulek on keerulisem.
 - **Õigem oleks:** „peredes, kus on 0-14-aastaseid lapsi, on leibkonnaliikme kulutused (sõltumata vanusest) kõige väiksemad.
- Lastega leibkondadel on võrreldes lasteta leibkondadega vähem kulutusi.
 - **Õigem oleks:** lastega leibkondades tehakse ühe liikme kohta vähem kulutusi kui lasteta leibkondades.

Järeldused

- Järeldused esitati iga alaeesmärgi juures või esitluse lõpus.
- Enamasti olid asjakohased ja seotud kontekstiga.
- Vähesed tõid välja järeldusi piiravaid tingimusi või asjaolusid.
- Sõnastus: esmatähtis olgu järeldus ise, alles seejärel järeldus hüpoteesi täitumise kohta!
 - Halb näide: „esimene hüpotees ei leidnud kinnitust, teine hüpotees leidis osaliselt kinnitust“
- Andmete vale kasutamine/tõlgendamine tõi kaasa valesid järeldusi.

Andmeanalüütiku vastutus

- Andmed kirjeldavad reaalse elu protsesse – ühiskond vajab nende kohta infot;
- Andmeanalüütik on tõlgendaja, kes loob andmetest teadmist;
- On võimalik toota „andmepõhiseid valeuudiseid“.
 - Statistikas on põhimõtteliselt ruumi subjektiivseks otsustamiseks
- Usaldusväärne andmeanalüütik
 - tunneb hästi ja kasutab õigesti matemaatilisi tööriistu;
 - ei taotle kallutatud eesmärke, süveneb andmete konteksti;
 - kasutab täpseid mõisteid ja termineid;
 - teeb järeldusi, mis kajastavad objektiivselt andmeanalüüsi tulemusi.

Õnnitlen
Euroopa statistikavõistluse 2019
Eesti võitjaid!