Regressio ja diskreetti todennäköisyysjakauma lukion pitkässä matematiikassa

Pro gradu -tutkielma

Saara-Sofia Jauho
Matemaattisten tieteiden tutkinto-ohjelma
Oulun yliopisto

2022

Sisällys

Johdanto

Nykyinen lukion opetussuunnitelma (LOPS 2021) pohjautuu oppimiskäsitykseen, jonka mukaan oppiminen tapahtuu opiskelijan aktiivisen ja tavoitteelisen toiminnan kautta [13]. Lukion opetussuunnitelman mukaan oppimisprosessissa opiskelija tulkitsee, analysoi ja arvioi hänelle esitettyä tietoa aikaisempien kokemustensa ja tietojensa pohjalta sekä luo näistä uudenlaisia kokonaisuuksia [13]. Nykyinen oppimiskäsitys siis korostaa opiskelijalähtöistä ja tutkivaa oppimista, jota käytössä olevien opetusmenetelmien sekä tarjolla olevien oppimateriaalien tulisi tukea.

Tämä pro gradu -tutkielma on osa Oulun yliopiston *avoin oppikirja* -projektia, jonka tarkoituksena on tarjota kaikille avointa oppimateriaalia lukion matematiikan kursseille. Oppimateriaalit tehdään osana opettajaopiskelijoiden pro gradu -töitä ja ne perustellaan tieteellisillä tutkimuksilla sekä voimassa olevilla lukion opetussuunnitelman perusteilla [14]. Projektin tavoitteena on siis luoda jokaisen pro gradu -työn tekohetkellä voimassa olevan lukion opetussuunnitelman mukaista matematiikan oppimateriaalia vapaaseen käyttöön.

Kyseessä on pitkä projekti, jonka aikana voimassa oleva lukion opetussuunnitelma voi ehtiä vaihtumaan useampaankin otteeseen. Tämän vuoksi oppikirjaprojektiin on valittu kaksi yhteistä tieteellistä artikkelia, joita hyödynnetään jokaisen oppimateriaalin suunnittelussa. Nämä artikkelit ovat *Habits of Mind: An Organizing Principle for Mathematics Curricula* [3] ja *Collaborative Learning in Mathematics: A Challenge to Our Beliefs and Practices* [15]. Ensimmäiseksi mainitusta artikkelista jokaisen oppikirjan tekijäryhmä valitsee oman kirjansa käsittelemään sisältöön parhaiten sopivat matemaattiset ajattelutavat, joita kyseisen kirjan oppimateriaaleilla halutaan tavoitella, sekä jälkimmäiseksi mainitusta tehtävätyypit, jotka tukevat näiden ajattelutapojen saavuttamista sekä kirjan käsittelemien asioiden oppimista.

Tämän lisäksi oppikirjaprojektissa on päätetty jokaisen oppimateriaalin noudattavan keskenään samanlaista etenemistapaa, jossa teoria käydään läpi tieteellisiin artikkeleihin sekä sen hetkiseen opetussuunnitelmaan pohjautuvien pohdintatehtävien avulla. Kirjasarjassa pohdintatehtävien tarkoituksena on nykyisen oppimiskäsityksen mukaisesti aktivoida opiskelijoita ja saada heidät yhdistämään uusi tieto aiemmin opittuun.

Tässä pro gradu -tutkielmassa käsitellään regression sekä diskreetin todennäköisyysjakauman aihealueita, jotka ovat osa nykyisten lukion opetussuunnitelman perusteiden mukaista pitkän matematiikan moduulia *Tilastot ja todennäköisyys* (MAA8) [13]. Kyseiselle moduulille tarkoitettu koko oppikirja koostuu viiden eri pro gradu -työn sisältämistä oppimateriaaleista. Oppikirja on suunniteltu niin, että ensin käsitellään tilastoihin liittyvät oppimateriaalit ja tämän jälkeen todennäköisyyteen liittyvät oppimateriaalit. Tästä johtuen tämän tutkielman sisältämä oppimateriaali sijoittuu varsinaisen oppikirjan eri puolille niin, että regression osuus päättää tilastoja käsittelevän osuuden ja diskreetin todennäköisyysjakauman osuus todennäköisyyttä käsittelevän osuuden.

Tutkielma rakentuu viidestä eri osasta. Ensimmäinen osa käsittelee oppimateriaalin yleisiä tavoitteita ja materiaalissa hyödynnettäviä tehtävätyyppejä, jotka pohjautuvat lukion opetussuunnitelman perusteisiin [13] ja koko avoimessa oppikirjasarjassa hyödynnettäviin kahteen yhteiseen artikkeliin [3, 15]. Toinen osa sisältää tuotetun oppimateriaalin tarkemman perustelun, joka pohjautuu yhteisten artikkeleiden ja opetussuunnitelman lisäksi myös muihin tieteellisiin artikkeleihin.

Kolmas osa taas muodostuu itse oppimateriaalista. Oppimateriaali koostuu pohdintatehtävien lisäksi teoriaosuuksista, erilaisista esimerkki- ja mallitehtävistä sekä myös harjoitustehtävistä, joiden avulla opiskelijat voivat harjoitella opeteltavaa asiaa. Määrällisesti oppimateriaali sisältää vähemmän harjoitustehtäviä kuin perinteisissä oppikirjoissa on totuttu näkemään. Tämä johtuu siitä, että harjoitustehtävät on pyritty suunnittelemaan niin, että ne eivät ole keskenään lähes samoja tehtäviä eri lukuarvoilla, vaan ne lähestyvät harjoiteltavaa asiaa eri näkökulmista. Tällöin opiskelijoiden tulee harjoiteltua opiskeltavaa asiaa monipuolisemmin.

Tutkielman kahtena viimeisenä osana ovat opettajan opas – joka sisältää ajankäyttösuunnitelman, oppimistavoitteet sekä vinkkejä ja vastauksia pohdintatehtäviin – ja harjoitustehtävien vastaukset.

1 Oppimateriaalin tavoitteet

1.1 Lukion opetussuunnitelman tavoitteet

Tämän pro gradu -työn sisältämä oppimateriaali on tehty noudattaen tutkielman tekohetkellä voimassa olevia vuonna 2019 julkaistuja lukion opetussuunnitelman perusteita [13]. Lukion opetussuunitelman perusteet määrittävät valtakunnalliset tavoitteet lukiokoulutukselle niin yleisellä kuin oppiaine- ja moduulikohtaisella tasolla. Moduulilla nykyisessä lukion opetussuunnitelmassa tarkoitetaan tietyn opintopistemäärän laajuista opintojaksokokonaisuutta [13].

Tämän tutkielman oppimateriaali on tarkoitettu kahden opintopisteen laajuiselle *Tilastot ja todennäköisyys* (MAA8) moduulille ja se kattaa regression sekä diskreetin todennäköisyysjakauman aihepiirien osat tästä moduulista. Näihin aihepiireihin liittyviä lukion opetussuunnitelman perusteissa määriteltyjä moduulin MAA8 tavoitteita, joihin tässä oppimateriaalissa siis keskitytään, ovat, että opiskelija

- osaa havainnollistaa kahden muuttujan yhteisjakaumaa sekä määrittää korrelaatiokertoimen ja regressiokäyrän
- osaa käyttää ohjelmistoja digitaalisessa muodossa olevan datan hakemisessa, käsittelyssä ja tutkimisessa sekä tilastollisen tiedon esittämisessä
- ymmärtää diskreetin todennäköisyysjakauman käsitteen ja oppii määrittämään jakauman odotusarvon sekä tulkitsemaan sitä
- osaa hyödyntää ohjelmistoja diskreetin todennäköisyysjakauman havainnollistamisessa, tunnuslukujen määrittämisessä sekä todennäköisyyksien laskemisessa [13].

Lukion opetussuunnitelman perusteissa moduulille MAA8 määritellyistä keskeisistä sisällöistä tässä oppimateriaalissa keskitytään korrelaatioon ja lineaariseen regressioon, diskreettiin todennäköisyysjakaumaan ja sen odotusarvoon sekä binomijakaumaan konkreettisena esimerkkinä diskreetistä todennäköisyysjakaumasta [13]. Lisäksi MAA8 moduulin koko oppikirjan tekijäryhmän kanssa on kirjan suunnitteluvaiheessa tulkittu lukion opetussuunnitelman perusteissa yhtenä moduulin keskeisenä sisältönä mainittuja keskilukuja ja keskihajontaa niin, että ne koskevat myös diskreettiä todennäköisyysjakaumaa [13]. Tämän vuoksi oppimateriaalissa käsitellään myös diskreetin todennäköisyysjakauman ja binomijakauman keskihajonta, pääpainon kuitenkin pysyessä odotusarvossa. Muita tähän oppikirjan osaan liittyviä keskeisiä käsitteitä on lukion opetussuunnitelman sekä MAOL:in lukion opetussuunnitelman perusteiden 2019 tukimateriaalin pohjalta tulkittu olevan hajontakuvio, selittävä ja selitettävä muuttuja, satunnaismuuttuja sekä pistetodennäköisyys ja myös nämä käsitteet käydään läpi tässä oppimateriaalissa [13, 11].

Lukion opetussuunnitelman perusteissa on määritelty moduulikohtaisten tavoitteiden lisäksi myös yleisempiä tavoitteita ja tehtäviä lukion matematiikan opetukselle [13]. Tutkielman sisältämä oppimateriaali on suunniteltu noudattamaan myös näitä tavoitteita ja tehtäviä.

Matematiikan yhdeksi tehtäväksi lukion opetussuunnitelman perusteissa on määritelty, että se antaa opiskelijalle valmiuksia ymmärtää, soveltaa, tuottaa ja arvioida matemaattisesti esitettyä tietoa [13]. Tämän tutkielman oppimateriaalien avulla opiskelijat oppivat luomaan kahden muuttujan yhteisjakaumia sekä soveltamaan näitä jakaumia muuttujien välisten riippuvuuksien tutkimiseen. Oppimateriaali pyrkii myös kiinnittämään opiskelijoiden huomion siihen, millaisia päätelmiä riippuvuuksista voidaan jakauman avulla tehdä, mikä tukee matemaattisesti esitetyn tiedon arvioinnin tehtävää. Lisäksi oppimateriaali auttaa opiskelijoita ymmärtämään diskreetin todennäköisyysjakauman käsitteen sekä soveltamaan sitä ja jakaumaan liittyviä tunnuslukuja diskreettiin todennäköisyyteen liittyvissä laskuissa.

Lukion opetussuunnitelman perusteissa matematiikan tehtäväksi on määritelty myös ohjaaminen matematiikan merkityksen ymmärtämiseen, jota on tässä oppimateriaalissa pyritty toteuttamaan luomalla tehtäviä, jotka liittyvät arkielämän tilanteisiin ja ovat mahdollisimman lähellä opiskelijan omaa elämää [13]. Esimerkiksi Pohdinnassa A.1 ja Esimerkissä A.3 käsitellään tilanteita, jotka ovat tuttuja varmasti lähes jokaiselle opiskelijalle heidän omasta elämästään.

Lisäksi opetussuunnitelman perusteissa on asetettu matematiikan yhdeksi tehtäväksi kehittää opiskelijan tietokoneohjelmistojen hyödyntämisen taitoja sekä yhdeksi yleiseksi tavoitteeksi, että opiskelija ymmärtää, ettei ohjelmiston tulos yksinään riitä väitteen osoitukseksi, todistukseksi tai perusteluksi [13]. Näitä matematiikan tehtävää ja tavoitetta oppimateriaalissa on pyritty noudattamaan luomalla materiaaliin harjoitustehtäviä, joiden ratkaisemisessa opiskelijan tulee hyödyntää ohjelmistoja. Ohjelmistojen hyödyntämistä regression ja diskreetin todennäköisyysjakauman aihepiireihin liittyen käydään kuitenkin läpi myös itse oppimateriaalissakin – muun muassa Pohdinnoissa A.7 ja A.21 – ja näissä oppimateriaalin kohdissa on pyritty korostamaan sitä, että ohjelmistojen hyödyntäminen sekä niistä saatavat tiedot ja tulokset ovat vain osa koko ratkaisua.

Lukion opetussuunnitelman perusteissa on määritelty myös nykyinen oppimiskäsitys, joka korostaa opiskelijan aktiivista roolia ja tavoitteellisuutta oppimisessa [13]. Opetussuunnitelman mukaan oppimisprosessin aikana opiskelija tulkitsee, analysoi ja arvioi hänelle esitettyä tietoa aikaisempien kokemustensa ja tietojensa pohjalta sekä luo näistä uusia kokonaisuuksia yhdistäen uudet tiedot osaksi aiemmin oppimaansa [13]. Tämän tutkielman sisältämä oppimateriaali ja erityisesti siihen kuuluvat pohdintatehtävät on pyritty suunnittelemaan niin, että ne tukevat tätä opetussuunnitelman perusteissa määriteltyä nykyistä oppimiskäsitystä aktivoiden opiskelijoita ja saaden heidät yhdistämään uuden tiedon aiemmin opittuun.

Lukion opetussuunnitelman mukaan oppiminen tapahtuu myös vuorovaikutuksessa muun muassa opettajien ja muiden oppilaiden kanssa [13]. Tämän vuoksi oppimateriaali on pyritty suunnittelemaan myös niin, että se lisäisi matematiikan oppimisen ja opettamisen sosiaalista luonnetta. Erityisesti pohdintatehtävät on ajateltu sellaisiksi tehtäviksi, että ne voidaan käydä oppituntien aikana yhdessä läpi joko opettajan tai muiden oppilaiden kanssa miettien ja myös harjoitustehtäviä on mahdollista laskea ja pohdiskella yhdessä. Kuitenkin tehtäviä laatiessa on haluttu antaa mahdollisuus myös itsenäiseen opiskeluun ja oppimateriaalissa pystyykin etenemään halutessaan myös omaan tahtiin.

1.2 Oppikirjan tekijöiden valitsemat tavoitteet ja tehtävätyypit

Lukion opetussuunnitelman perusteiden [13] lisäksi tutkielman sisältämän oppimateriaalin suunnittelussa on käytetty erityisesti artikkeleita *Habits of Mind: An Organizing Principle for Mathematics Curricula* [3] ja *Collaborative Learning in Mathematics: A Challenge to Our Beliefs and Practices* [15]. Ensimmäiseksi mainitusta artikkelista MAA8 moduulille oppikirjan tehneen ryhmän kanssa on kirjan yhteisessä suunnitteluvaiheessa valittu muutama matemaattinen ajattelutapa, joita oppikirjan materiaaleilla halutaan tavoitella. Tämän lisäksi jälkimmäiseksi mainitusta artikkelista on tekijäryhmän kanssa valittu muutama tehtävätyyppi, jotka tukevat näiden valittujen matemaattisten ajattelutapojen saavuttamista ja oppikirjan käsittelemien asioiden oppimista.

1.2.1 Habits of mind -artikkelin tavoitteet

Se, millaisia matemaattisia ongelmia opiskelijat tulevat valmistumisensa jälkeen kohtaamaan, muuttuu ajan myötä siinä missä maailmakin [3]. Esimerkiksi viisitoista vuotta sitten ohjelmointia ei pidetty kovinkaan tärkeänä matemaattisena taitona, eikä se sisältynyt peruskoulun tai lukion opetussuunnitelmiin. Tilanne nykypäivänä on kuitenkin täysin toinen ja ohjelmoinnin opettaminen aloitetaan nykyään jo ala-asteella.

Tällainen jatkuva maailman muutos aiheuttaa haasteita opetukseen erityisesti matematiikan osalta. Emme voi tietää, mitkä matemaattiset sisällöt ja aihealueet ovat keskiössä sillä hetkellä, kun tänä vuonna koulunsa aloittavat oppilaat valmistuvat ja siirtyvät työelämään, joten miten voimme valita, mitä sisältöjä ja aihealueita heidän matematiikan opetuksessaan tulisi painottaa [3]?

Artikkelissa *Habits of Mind: An Organizing Principle for Mathematics Curricula* erilaisia matemaattisia ajattelutapoja (*habits of mind*) ehdotetaan ratkaisuksi tähän ongelmaan. Cuocon, Golbergin ja Markin mukaan matematiikan opetuksessa tulisi keskittyä auttamaan opiskelijoita omaksumaan erilaisia matemaattisia ajattelutapoja sen sijaan, että keskityttäisiin pelkästään saamaan opiskelijat oppimaan erilaisia ratkaisutapoja lukuisiin erilaisiin matemaattisiin ongelmiin. Tällä tavoin heidän mukaansa opetuksessa pystytään antamaan opiskelijoille kokonaisvaltaisempaa matemaattista osaamista, jonka avulla opiskelijat pystyvät tulevaisuudessa selviämään myös sellaisista matemaattisista ongelmista, joita nykypäivänä ei välttämättä vielä edes ole olemassa. Artikkelin mukaisten matemaattisten ajattelutapojen omaksuminen auttaa opiskelijoita ajattelemaan kuin matemaatikko, mikä taas antaa suoraan työkaluja erilaisten matemaattisten ongelmien ratkaisemiseen [3].

Artikkelissa Cuoco ym. esittelevät kahdeksan erilaista matemaattista ajattelutapaa, joista seuraavaksi esiteltäviä kolmea ajattelutapaa on hyödynnetty tämän oppimateriaalin suunnittelussa.

Säännönmukaisuuksien etsijät

Matematiikan opetuksen pitäisi kehittää opiskelijoita säännönmukaisuuksien etsijöiksi (*pattern sniffers*). Tätä ajattelutapaa käyttävä opiskelija etsii ja hyödyntää säännönmukaisuuksia tai malleja matemaattisten ongelmien ratkaisemisessa. Opiskelijoiden tulisi tottua etsimään säännönmukaisuuksia aina matemaattisia ongelmia ratkaistessaan, olivat ongelmat sitten heidän itse keskimiään tai jonkun muun heille ratkaistavaksi antamia. [3].

Matematiikassa kaikki laskusäännöt pohjautuvat loppujen lopuksi johonkin säännönmukaisuuteen. Tämän huomatessaan ja oppiessaan rutiininomaisesti etsimään säännönmukaisuuksia matemaattisista ongelmista opiskelijat oppivat myös samalla luomaan uutta matematiikkaa tai johtamaan jo luotua matematiikkaa ilman, että jokaista tulosta, kaavaa tai laskusääntöä tarvitsee opetella ulkoa.

Oppimateriaalissa opiskelijoiden säännönmukaisuuksien etsintää pyritään kehittämään muun muassa pohdintatehtävien avulla. Esimerkiksi Pohdinnoissa A.8 ja A.18 tarkoituksena on, että opiskelija joutuu ensin itse pohtimaan ja miettimään opittavaan asiaan liittyvää tehtävää sekä tutkimaan, liittyykö asiaan säännönmukaisuutta, ennen kuin oppikirja paljastaa säännönmukaisuudet oppimateriaalin teoriaosuuksissa.

Kuvailijat

Matematiikan opetuksen pitäisi opettaa opiskelijoita myös kuvailijoiksi (*describers*). Matematiikkaa voidaan pitää omana kielenään ja opiskelijoiden tulisi osata käyttää tätä kieltä niin kirjallisesti kuin suullisestikin. Erityisesti opiskelijoiden tulisi oppia kuvailemaan täsmällisesti omien ratkaisujensa vaiheita käyttäen matemaattista kieltä, sillä jonkun asian selittäminen tai kuvaileminen toiselle ymmärrettävästi syventää myös omaa ymmärrrystä kyseisestä asiasta. [3].

Opiskelijoita tulisi opettaa käyttämään matemaattista kieltä myös väittelemisessä eli harjaannuttaa heidän taitojaan aukottomasti perustella matemaattisia väitteitä tosiksi. Tällä tavoin opiskelijat tottuvat tuottamaan matemaattisia todistuksia myös itse sen sijaan, että he osaisivat ainoastaan ymmärtää muiden tekemiä todistuksia. [3].

Oppimateriaalissa pyritään kehittämään opiskelijoita kuvailijoiksi mallintamalla matemaattisen kielen käyttöä teoriaosuuksissa sekä malli- ja esimerkkitehtävissä. Lisäksi pohdinta- ja harjoitustehtävissä opiskelijoita pyydetään usein perustelemaan joko omia ratkaisujaan tai tehtävässä esitettyjä ratkaisuvaihtoehtoja. Näin on esimerkiksi Pohdinnassa A.10 ja Harjoitustehtävässä 4. Myös erilaiset väitteiden arviointitehtävät, kuten Pohdinta A.1 ja Harjoitustehtävä 6, ohjaavat opiskelijoita harjoittelemaan kuvailemista, kun heidän tulee mahdollisimman aukottomasti perustella, milloin väitteet pitävät paikkaansa vai pitävätkö koskaan.

Visualisoijat

Matematiikan opetuksen tulisi harjaannuttaa opiskelijoista myös visualisoijia (*visualizers*). Matematiikka sisältää monenlaista visualisointia, jonka tarkoituksena on muun muassa auttaa matemaattisten prosessien ymmärtämistä. Esimerkiksi binomin toiseen potenssiin korottamista matemaattisena prosessina voi visualisoida pinta-alojen avulla, mikä auttaa ymmärtämään kyseistä laskusääntöä. Matematiikan opetuksessa tulisikin pyrkiä saamaan opiskelijat visualisoimaan mahdollisimman paljon eri tyyppistä matematiikkaa, kuten dataa, geometrisia asioita, matemaattisia riippuvuuksia ja prosesseja sekä ilmiöiden muutoksia. [3].

Oppimateriaalissa opiskelijoiden visualisointia pyritään lisäämään keskittymällä datan ja matemaattisten riippuvuuksien visualisointiin. Regression osiossa opiskelijat oppivat hyödyntämään hajontakuviota ja regressiokäyriä kahden muuttujan datan visualisoinnissa sekä käyttämään näitä visuaalisia komponentteja muuttujien välisten riippuvuuksien tutkimisessa. Diskreetin todennäköisyysjakauman osiossa taas opiskelijat pääsevät luomaan taulukoita ja kuvaajia diskreettien todennäköisyysjakaumien visualisoinniksi.

Tämän lisäksi erityisesti erilaisten esitystapojen tulkintatehtävät, kuten Harjoitustehtävä 3 ja 9, tukevat opiskelijoiden kehittymistä visualisoijiksi, sillä päästessään vertailemaan jonkun matemaattisen asian monta eri esitystapaa yhtä aikaa he saavat yhdistettyä kaikki nämä saman asian erilaiset visualisoinnit toisiinsa ja oppivat visualisoimaan matemaattisia asioita monipuolisemmin.

1.2.2 Tehtävätyypit

Artikkelissa *Collaborative Learning in Mathematics: A Challenge to Our Beliefs and Practices* Malcolm Swan esittelee erilaisia tehtävätyyppejä, joiden tarkoituksena on saada opiskelijoita omaksumaan aktiivisempia lähestymistapoja opiskeluun. Sen sijaan, että opiskelijat näkisivät matematiikan vain sarjana toisiinsa liittymättömiä ulkoa muistettavia erilaisten tehtävien ratkaisutapoja, pitäisi heidät saada keskustelemaan ja selittämään matemaattisia ideoita toisilleen, haastamaan ja opettamaan toisiaan sekä luomaan ja ratkaisemaan toistensa keksimiä tehtäviä myös yhdessä tavanomaisen yksin puurtamisen lisäksi [15].

Artikkelissa esitellyistä tehtävätyypeistä kolmea on hyödynnetty tämän oppimateriaalin tehtävien ideoimisessa ja nämä kolme tehtävätyyppiä esitellään seuraavaksi tarkemmin.

Erilaisten ratkaisutapojen vertailu

Tämän tehtävätyypin tehtävissä painopiste siirretään oikean vastauksen saamisesta siihen, miten vastaukseen on päädytty. Tarkoituksena on saada opiskelijat huomaamaan, että matemaattisia ongelmia voidaan yleensä ratkaista useammalla eri tavalla, sillä monet opiskelijat eivät voi edes yrittää ratkaista matemaattista ongelmaa, jos he kokevat etteivät tiedä ongelman oikeaa ratkaisutapaa. Toisena ongelmana ovat opiskelijat, jotka omaksuvat vain yhden mahdollisista ratkaisutavoista ja saavat sillä kyllä oikeita vastauksia, mutta tehottomasti. [15]

Kun opiskelijat pääsevät harjoittelemaan tehtäviä, joissa pyydetään ratkaisemaan sama ongelma useammalla eri tavalla tai vertailemaan jo annettuja saman tehtävän eri ratkaisuja sekä perustelemaan, mikä näistä ratkaisuista olisi tehokkain tapa ratkaista tehtävä, saavat opiskelijat enemmän itsevarmuutta matematiikan suhteen sekä oppivat käyttämään joustavammin eri ratkaisutapoja [15].

Väitteiden arviointi

Näissä tehtävissä opiskelijoiden tulee päättää, onko jokin tietty matemaattinen väite aina, joskus vai ei koskaan tosi. Lisäksi opiskelijoita voidaan pyytää miettimään, millaisilla ehdoilla tai muutoksilla väite olisi aina tosi. Väitteiden arviointitehtävät kehittävät erityisesti opiskelijoiden matemaattista argumentointia ja perustelua. Ne myös rohkaisevat opiskelijoita keksimään heidän omaa päättelyään tukevia esimerkkejä ja vastaesimerkkejä. Väitteiden arviointitehtävät voidaan lisäksi suunnitella niin, että niissä käsitellään aiheeseen liittyviä yleisiä väärinymmärryksiä tai hankaluuksia. [15].

Erilaisten esitystapojen tulkinta

Matemaattisia asioita voidaan esittää monella eri tavalla. Yhden ja saman asian esittämiseen voidaan käyttää esimerkiksi sanoja, kaavoja, diagrammeja, taulukoita tai vaikka kuvaajia. Tämän tehtävätyypin mukaisten tehtävien tarkoituksena on saada opiskelijat yhdistämään samaan matemaattiseen ideaan liittyvät erilaiset esitystavat toisiinsa. Opiskelijat voivat esimerkiksi ideoida yhdelle asialle mahdollisimman monta erilaista esitystapaa yhdessä tai vertailla yhden asian erilaisia esitystapoja toisiinsa samalla perustellen, miksi nämä esitystavat ovat samat. [15].

2 Oppimateriaalin perustelut

Tutkielman sisältämä oppimateriaali sisältyy moduulille *Tilastot ja todennäköisyys* (MAA8) suunniteltuun oppikirjaan. Oppimateriaali sisältää oppikirjan aihepiirit *Regressio* ja *Diskreetti todennäköisyysjakauma*, joista regressio päättää tilastoja käsittelevän oppikirjan osan ja diskreetti todennäköisyysjakauma todennäköisyyksiä käsittelevän osan.

Oppimateriaali koostuu siis kahden hieman erillisen aihepiirin oppimateriaaleista, jotka molemmat ovat oppikirjan pääaihepiirien viimeisinä opiskeltavia asioita, joten oppimateriaaleissa hyödynnetään myös oppikirjassa jo aiemmin läpikäytyjä asioita.

Tutkielman tässä osassa perustellaan lukion opetussuunnitelman perusteiden ja tieteellisten artikkeleiden avulla oppimateriaalien suunnittelussa tehdyt valinnat. Itse oppimateriaalit löytyvät tutkielman lopusta liitteestä A.

2.1 Regressio

Lukion opetussuunnitelman perusteissa regression aihepiiriin liittyväksi moduulin MAA8 yhdeksi tavoitteeksi on määritelty, että opiskelija osaa havainnollistaa kahden muuttujan yhteisjakaumaa sekä määrittää korrelaatiokertoimen ja regressiokäyrän [13]. Tämän vuoksi regression osuuden oppimateriaalissa lähdetään liikkeelle kahden muuttujan yhteisjakauman havainnollistamisesta, johon liittyviä peruskäsitteitä ovat hajontakuvio, selittävä muuttuja ja selitettävä muuttuja. Tämän jälkeen oppimateriaalissa käsitellään regressioanalyysin, regressioyhtälön, lineaarisen regressiosuoran ja regressiokäyrän käsitteet sekä lopuksi korrelaation ja korrelaatiokertoimen käsitteet.

Käsitteiden läpikäyntijärjestys pohjautuu opetussuunnitelman perusteiden lisäksi Stephanie Caseyn artikkeliin [2], jonka mukaan artikkelissa esitelty hypoteettinen oppimisen kehityskaari (*HLT*) toimii peräkkäisen ja asteittaisen ajattelun mallina, joka opiskelijoiden tulee käydä läpi ymmärtääkseen opiskeltavan aiheen syvällisesti. Artikkelissa esitelty malli lineaarisen regression oppimiselle koostuu viidestä eri vaiheesta, joista jokainen keskittyy yhden lineaariseen regressioon liittyvän asian oppimiseen ja liittää uutta tietoa aina edellisessä vaiheessa opittuun. Erityisesti mallin mukaan lineaarisen regression oppimisen tulisi lähteä kahden muuttujan yhteisjakauman havainnollistamisesta ja regressioyhtälön käsite tulisi oppia ennen korrelaation opettamista [2].

Kahden muuttujan yhteisjakaumaa käytetään kahden muuttujan välisen riippuvuuden tutkimiseen, jonka vuoksi Pohdinnassa A.1 opiskelijoita johdatellaan riippuvuuksien aihepiirin pariin. Tehtävässä opiskelijat pääsevät pohtimaan kahden muuttujan välisiä riippuvuuksia kolmessa arkielämään liittyvässä väittämässä, eli tehtävä on Swanin artikkelin mukainen väitteiden arviointitehtävä [15]. Pohdintatehtävän tarkoituksena on antaa opiskelijoille kolme esimerkkiä tilanteista, joissa muuttujien välisen riippuvuuden olemassaolosta voidaan suoraan todeta jotain ja tällä tavoin johdatella opiskelijat riippuvuuksien pariin. Pohdinta A.1 kehittää opiskelijoita samalla myös Cuoco ym. artikkelin mukaisiksi kuvailijoiksi, sillä opiskelijoita pyydetään perustelemaan vastauksensa tehtävään [3]. Pohdinnan A.1 jälkeen oppimateriaalissa on pieni teksti, joka johdattelee opiskelijoita kahden muuttujan riippuvuuksien tutkimisen pariin sekä esittelee kahden muuttujan yhteisjakauman termin.

Caseyn mukaan tutkimuksissa on havaittu, että opiskelijoilla on hankaluuksia tunnistaa, kumpi kahdesta muuttujasta on selittävä muuttuja ja kumpi selitettävä muuttuja, jos he joutuvat itse muuttamaan johonkin tilastolliseen tutkimukseen liittyvät kahden muuttujan taulukoidut arvot hajontakuvioksi, eli päättämään kumman muuttujan arvot tulevat hajontakuviossa x-akselille (selittävä) ja kumman y-akselille (selitettävä) [2]. Tämä on kuitenkin aina ensimmäinen vaihe kahden muuttujan yhteisjakauman havainnollistamisessa, joten opiskelijoille pyritään luomaan hyvä ymmärrys näistä peruskäsitteistä heti oppimateriaalin alussa.

Määritelmässä A.2 opiskelijoille annetaan selkeä, tilastokeskuksen sivuilta katsottu määritelmä selitettävän ja selittävän muuttujan käsitteille sekä samalla myös hajontakuvion käsitteelle [16]. Hajontakuvion määritelmässä hyödynnetään havaintoaineiston käsitettä, joka on käyty läpi oppikirjan aiemmissa tilastoja käsittelevissä kappaleissa. Määritelmässä A.2 otetaan mukaan myös selitettävän ja selittävän muuttujan toiset nimitykset eli riippuva ja riippumaton muuttuja, jotta nämä molemmat yleisesti käytössä olevat nimitykset tulevat opiskelijoille tutuiksi.

Määritelmän A.2 jälkeen Esimerkki A.3 selventää määriteltyjen käsitteiden käyttöä ja yhdistää ne konkreettiseen arkielämän esimerkkiin. Oppimateriaalissa arkielämään liittyvät esimerkit ja tehtävät tukevat matematiikan merkityksen ymmärtämistä, mikä on määritelty lukion opetussuunnitelmassa yhdeksi matematiikan opetuksen tehtäväksi [13]. Lisäksi Esimerkki A.3 tukee Habits of mind -artikkelista valittua visualisoinnin tavoitetta [3].

Caseyn sekä Engelin ja Sedlmeierin mukaan monella opiskelijalla on hankaluuksia ymmärtää, että tilastotieteessä kahden muuttujan dataa kuvaavassa hajontakuviossa samaa x-akselin muuttujan arvoa saa vastata useampi y-akselin muuttujan arvo, sillä funktioiden kanssa he ovat tottuneet siihen, ettei tällainen ole sallittua [2, 4]. Lisäksi Engelin ja Sedlmeierin artikkelin mukaan jotkut opiskelijat tulkitsevat tällaiset tilanteet niin, ettei muuttujien välillä tällöin ole riippuvuutta ollenkaan [4]. Jotta opiskelijat tottuisivat tähän tilastotieteessä hyvin yleiseen funktioista poikkeavaan asiaan, on Esimerkki A.3 suunniteltu niin, että hajontakuviossa useammassa eri kohdassa sama x-akselin muuttujan arvo saa useamman y-akselin muuttujan arvon.

Caseyn mukaan opiskelijoilla on huomattu olevan vaikeuksia myös sen kanssa, että lineaarisista funktioista poiketen hajontakuviossa datapisteet eivät välttämättä sijoitu kasvavaan tai vähenevään muotoon, mutta muuttujien välillä voi silti olla lineaarista riippuvuutta [2]. Tämän vuoksi Esimerkin A.3 hajontakuvio on suunniteltu myös niin, että muuttujien välillä on lineaarista riippuvuutta, vaikka havaintopisteet eivät asetu monotonisesti, jotta opiskelijat tottuvat tilastotieteessä myös tällaisen olevan sallittua.

Esimerkin A.3 jälkeen Pohdinnassa A.4 opiskelijat pääsevät heti soveltamaan uusia tietojaan. Pohdintatehtävässä on annettu kolmen eri tilastollisen tutkimuksen tutkimuskysymykset sekä tutkimuksiin liittyvät selitettävät ja selittävät muuttujat sekaisin lueteltuina. Opiskelijoiden tulisi nyt osata soveltaa tietojaan erityisesti selitettävään muuttujaan liittyen ja valita luettelosta annettuja tutkimuskysymyksiä vastaavat selitettävät muuttujat. Koska muuttujiksi ei ole lueteltu muita ylimääräisiä muuttujia kuin tutkimuksiin liittyvät selittävät muuttujat, antaa Pohdinta A.4 opiskelijoille tilaisuuden soveltaa myös tietojaan selittävään muuttujaan liittyen poissulkumenetelmän kautta. Habits of mind -artikkelin tavoitteista tässä pohdintatehtävässä toteutuu opiskelijoiden kehittyminen säännönmukaisuuksien etsijöiksi, sillä tehtävän ratkaisemisessa voi hyödyntää sitä säännönmukaisuutta, jonka mukaan selitettävä muuttuja on aina se muuttuja, jonka arvojen vaihtelua tilastollisessa tutkimuksessa pyritään selittämään jollain toisella muuttujalla [3].

Hueyn ja Bakerin artikkelin mukaan kahden muuttujan välisen lineaarisen riippuvuuden olemassaoloa määrittäessään opiskelijoiden tulisi tulkita muuttujista kerätystä datasta muodostettua hajontakuviota ja olla hajontakuvion tarkastelun pohjalta vakuuttuneita siitä, että muuttujien välillä näkyy riippuvuutta ja että tämä riippuvuus on juuri lineaarista [9]. Tätä opiskelijat pääsevät harjoittelemaan Pohdinnassa A.5, jossa tehtävänä on tarkastella neljää eri hajontakuviota ja miettiä, kannattaisiko näiden hajontakuvioiden pistepilviin sovittaa suora vai ei. Hajontakuvioista yksi on esimerkki tilanteesta, jossa muuttujien välillä ei näy lainkaan riippuvuutta, yksi tilanteesta, jossa riippuvuus on ennemminkin toisen asteen polynomin muotoista kuin lineaarista ja kaksi tilanteista, joissa riippuvuutta voidaan kuvata suoralla.

Pohdinta A.5 kehittää kaikkia Habits of mind -artikkelista valittuja tavoitteita opiskelijoissa, sillä tehtävässä tulee perustella omat vastaukset, etsiä pistepilvistä säännönmukaisuuksia sekä tarkastella annettujen muuttujaparien välisiä riippuvuuksia visuaalisesti [3]. Pohdinnan A.5 jälkeen oppimateriaalissa kerrotaan hieman regressioanalyysistä ja käydään läpi regressioyhtälön, regressiokäyrän sekä lineaarisen regressiosuoran käsitteet. Tiedot tähän oppimateriaalin kohtaan on katsottu tilastokeskuksen sivuilta [16].

Caseyn artikkelissa esitellyn hypoteettisen oppimisen kehityskaaren mallin mukaan ennen virallisen regressiosuoran sovittamisen opettamista, opiskelijoiden tulisi vapaamuotoisemmin tarkastella, millainen suora parhaiten sopii annettuun pistepilveen [2]. Tätä opiskelijat pääsevät tekemään Pohdinnassa A.6, joka on samantyylinen tehtävä kuin Caseyn artikkelissa esitelty lineaarisen regression oppimisen tutkimuksessa hyödynnetty tehtävä [2].

Pohdinnassa A.6 on annettu kaksi Esimerkin A.3 pistepilveen sovitettua suoraa, joista toinen on määritetty pienimmän neliösumman menetelmällä, eli niin, että suora kulkee mahdollisimman läheltä kaikkia havaintopisteitä ja toinen niin, että suora kulkee mahdollisimman monen havaintopisteen kautta. Opiskelijoiden tehtävänä on vertailla suoria keskenään ja tarkastella, miten suorat asettuvat havaintopisteisiin nähden. Tämän jälkeen opiskelijoita pyydetään perustelemaan, kumpi suora kuvaa pistepilveä paremmin.

Caseyn artikkelissa esitellyssä tehtävässä kolmasosa tehtävän tehneistä opiskelijoista oli ollut sitä mieltä, että havaintopisteisiin suoraa sovitettaessa on tärkeämpää saada suora kulkemaan mahdollisimman monen pisteen kautta kuin mahdollisimman läheltä kaikkia pisteitä [2]. Siksi tässäkin pohdintatehtävässä voidaan olettaa usean opiskelijan olevan aluksi sitä mieltä, että useamman pisteen kautta kulkeva suora kuvaa pistepilveä paremmin, jolloin tähän harhakäsitykseen voidaan puuttua, kun pohdintatehtävää käydään yhdessä läpi. Pohdinnan A.6 päätarkoituksena onkin tarjota tilaisuus korjata tämä harhakäsitys niiltä opiskelijoilta, jotka ovat sen omaksuneet. Samalla tehtävä kehittää Cuocon ym. artikkelin tavoitteiden mukaisesti opiskelijoista kuvailijoita ja visualisoijia, sillä tehtävässä opiskelijoiden tulee perustella vastauksensa sekä visuaalisesti tutkia kumpi suora kuvaa pistepilveä paremmin [3].

Caseyn mukaan monilla opiskelijoilla, ja opettajillakin, on harhakäsitys siitä, että lineaarisen regressiosuoran ylä- ja alapuolelle jää aina yhtä monta havaintopistettä, mutta mitään tällaista matemaattista sääntöä ei ole olemassa, vaikka välillä regressiosuorat asettuvatkin tällä tavoin havaintopisteisiin nähden [2]. Jotta opiskelijoille ei pääsisi syntymään tällaista harhakäsitystä, on Esimerkin A.3 hajontakuvion havaintopisteet suunniteltu niin, että molemmissa Pohdinnan A.6 tilanteissa suoran ylä- ja alapuolille jää eri määrät havaintopisteitä.

Pohdinnan A.6 jälkeen Pohdinnassa A.7 opiskelijat pääsevät harjoittelemaan GeoGebralla virallisen regressiosuoran sovittamista sekä samalla hajontakuvion tekemistä itse, sillä lukion opetussuunnitelman mukaan opiskelijoiden tulisi osata havainnollistaa kahden muuttujan yhteisjakaumaa sekä käyttää havainnollistamisessa apuna myös ohjelmistoja [13]. Ennen tätä oppimateriaalissa kuitenki kerrotaan lyhyesti regressioyhtälöstä ja pienimmän neliösumman menetelmästä.

Pienimmän neliösumman menetelmän esittely on otettu mukaan oppimateriaaliin, koska lukion opetussuunnitelman perusteiden mukaan opiskelijoiden tulisi osata määrittää regressiokäyrä sekä hyödyntää tähän myös ohjelmistoja [13]. Tämän vuoksi oppimateriaalissa kerrotaan lyhyesti, millä periaatteella regressiokäyrä matemaattisesti määritettäisiin sekä annetaan Pohdinnassa A.7 tarkemmat ohjeet regressiomallin sovittamiseen GeoGebralla. Tässä oppimateriaalissa opiskelijoita pyritään ohjaamaan juuri GeoGebran hyödyntämiseen kaikista tarjolla olevista ohjelmistoista, sillä Paul Hewsonin mukaan GeoGebra on tarjolla olevista ohjelmistoista kaikista mielenkiintoisin tilastotieteen opettamisen kannalta [8]. Hewson muun muassa toteaa GeoGebran olevan siinä mielessä arvokas ohjelmisto, että sitä pystyy hyvin intuitiivisesti hyödyntämään matematiikassa keskeisten aiheiden tutkimiseen [8].

Pohdinnassa A.7 opiskelijoille jätetään itse selvitettäväksi, mistä regressioyhtälön saa GeoGebralla määritettyä, koska GeoGebra näyttää regressioyhtälön automaattisesti regressiomallin sovittamisen yhteydessä, jolloin opiskelijoilta ei tässä vaadita muuta kuin regressioyhtälön termin ymmärtämistä. Pohdinnan A.7 lopussa on lisäksi kysymys, joka ohjaa opiskelijoita huomaamaan, että regressiomallin avulla voidaan tehdä arvioita ja ennusteita selitettävän muuttujan käyttäytymisestä. Habits of mind -artikkelin tavoitteista tämän pohdintatehtävän tavoitteena on tukea opiskelijoiden kehitystä visualisoijiksi [3]. Pohdinnan A.7 jälkeen oppimateriaalissa kerrotaan myös tarkemmin ennustamisesta regression perusteella sekä esitellään myös lineaarisen regressiosuoran yhtälö ja sen kertoimet. Tiedot oppimateriaalin tähän kohtaan on katsottu tilastokeskuksen tilastokoulun sivuilta [17].

Seuraavaksi oppimateriaalissa käsitellään korrelaatiota. Opiskelijoille annetaan oppimateriaalissa tiedot siitä, mitä korrelaation sekä korrelaatiokertoimen käsitteillä tarkoitetaan. Nämä tiedot on katsottu tilastokeskuksen sivuilta [16]. Tämän jälkeen Pohdinnassa A.8 opiskelijat pääsevät itse tutkimaan eri suuruisia korrelaatiokertoimia sekä niihin liittyviä hajontakuvioita ja regressiosuoria. Pohdinnan A.8 tarkoituksena on tukea Habits of mind -artikkelin tavoitteista opiskelijoiden kehittymistä säännönmukaisuuksien etsijöiksi [3]. Pohdintatehtävässä opiskelijat pääsevät nimittäin ensin itse tutkimaan korrelaatiokertoimeen liittyviä säännönmukaisuuksia, kuten minkä suuruisia arvoja korrelaatiokerroin saa annetuissa tilanteissa, milloin kerroin on positiivinen ja milloin negatiivinen sekä milloin korrelaatiokertoimen itseisarvo on suurin ja milloin pienin, ja vasta tämän jälkeen oppimateriaalissa kerrotaan tarkemmin korrelaatiokertoimesta sekä sen ominaisuuksista. Tietoja korrelaatiokertoimen ominaisuuksista on katsottu Studeo Oy:n lyhyen matematiikan Tilastot ja todennäköisyys -oppikirjasta [7]. Pohdinnassa A.8 opiskelijat pääsevät myös liittämään korrelaatiokertoimen eri suuruusluokkia tietynlaisiin hajontakuvioihin sovitettuihin suoriin eli pohdintatehtävä kehittää opiskelijoita myös visualisoijiksi Habits of mind -artikkelin tavoitteista [3].

Lukion opetussuunnitelman mukaan opiskelijoiden tulee osata määrittää korrelaatiokerroin sekä hyödyntää määrittämisessä myös ohjelmistoja [13]. Tämän vuoksi Pohdinnan A.8 jälkeen oppimateriaalissa annetaan lyhyesti tietoa siitä, miten korrelaatiokerroin matemaattisesti määritettäisiin sekä tarkemmat ohjeet korrelaatiokertoimen määrittämiseen GeoGebralla. Korrelaatiokertoimen matemaattisesti määrittämisen osuudessa hyödynnetään keskiarvon ja keskihajonnan käsitteitä, jotka ovat tuttuja opiskelijoille jo oppikirjan aiemmista tilastoja käsittelevistä kappaleista. Kaava Pearsonin korrelaatiokertoimelle on katsottu tilastokeskuksen tilastokoulun sivuilta [17].

Hueyn ja Bakerin mukaan opiskelijoiden tulisi myös oppia arvioimaan mahdollisten poikkeavien havaintopisteiden vaikutusta regressioanalyysin tuloksiin ennen analyysin toteuttamista ja korrelaatiokertoimen määrittämistä [9]. Tähän opiskelijoita pyritään opettamaan Pohdinnalla A.10 ja Huomautuksella A.11. Lisäksi Pohdinnalla A.10 ja Huomautuksella A.11 pyritään puuttumaan Engelin ja Sedlmeierin esille tuomaan yleiseen harhaluuloon, jonka mukaan vahva korrelaatio kahden muuttujan välillä automaattisesti tarkoittaisi, että muuttujia voidaan kuvata lineaarisella regressiomallilla [4].

Pohdinnassa A.10 on neljä hajontakuviota, joihin kaikkiin on regressioanalyysillä sovitettu lineaarinen regressiosuora sekä määritetty korrelaatiokerroin ja kaikissa neljässä hajontakuviossa nämä ovat samat. Opiskelijoiden tehtävänä on huomata, että vaikka kaikissa tilanteissa korrelaatiokerroin on lähes yksi, ei lineaarinen regressiosuora ole sopiva matemaattinen malli kuvaamaan kaikkia annettuja hajontakuvioita. Lisäksi Pohdinnassa A.10 opiskelijoiden tulee tarkastella kahta hajontakuvioista tarkemmin, joissa molemmissa on yksi poikkeava havaintopiste ja pohtia, millainen vaikutus kyseisen havaintopisteen poistamisella voisi olla sovitettuun regressiosuoraan. Arvot Pohdinnan A.10 hajontakuvioiden havaintopisteisiin on katsottu Engelin ja Sedlmeierin artikkelissa esitellyistä hajontakuvioista [4].

Huomautuksessa A.11 tiivistetään lopuksi Pohdinnassa A.10 käsitellyt asiat. Lisäksi Huomautuksessa A.12 kiinnitetään opiskelijoiden huomio siihen, ettei regressioanalyysin tai korrelaatiokertoimen määrityksen perusteella voida tehdä päätelmiä muuttujien välisistä syy-seuraussuhteista, sillä Bataneron, Estepan ja Godinon mukaan moni opiskelija ei ymmärrä, että vahva korrelaatio muuttujien välillä ei automaattisesti tarkoita myös syy-seuraussuhdetta muuttujien välillä [1]. Tukemaan opiskelijoiden ymmärrystä tästä asiasta Huomatukseen A.12 on otettu mukaan arkielämään liittyvä ja helposti ymmärrettävä konkreettinen esimerkki tilanteesta, jossa muuttujien välillä on korrelaatiota, mutta kumpikaan ei ole seurausta toisesta.

Regression osuuden oppimateriaali päättyy harjoitustehtäviin, jotka on suunniteltu niin, että ne auttavat opiskelijoita ymmärtämään oppimateriaalissa käsitellyt asiat paremmin sekä saavuttamaan lukion opetussuunnitelmassa määritellyt oppimisen tavoitteet regression aihepiirin osalta [13]. Harjoitustehtävien 1 ja 2 tavoitteena on syventää opiskelijoiden ymmärrystä selitettävän ja selittävän muuttujan käsitteistä, mikä tukee opetussuunnitelmassa määritettyä tavoitetta kahden muuttujan yhteisjakauman havainnollistamisen osaamisesta [13]. Harjoitustehtävässä 1 opiskelijat pääsevät itse miettimään, kumpi annetuista kahdesta muuttujasta kannattaisi valita selitettäväksi muuttujaksi ja kumpi selittäväksi muuttujaksi, jos muuttujien välistä riippuvuutta lähdettäisiin tilastollisesti tutkimaan. Harjoitustehtävässä 2 opiskelijoille on taas annettu valmiita tutkimuskysymyksiä, joiden perusteella heidän pitää päätellä, kumpi kysymyksessä esiintyvistä muuttujista on selitettävä ja kumpi selittävä muuttuja.

Caseyn mukaan opiskelijat tarvitsevat kokemusta myös tilanteista, joissa selittävän ja selitettävän muuttujan määrittäminen on epäselvempää ymmärtääkseen, että aina muuttujia ei voida selkeästi jakaa selitettävään ja selittävään muuttujaan [2]. Myös Engel ja Sedlmeier toteavat, että selitettävän ja selittävän muuttujan erottaminen toisistaan voi olla vaikeaa, sillä jos muuttujien välillä on tilastollista riippuvuutta, niin se ei vielä välttämättä tarkoita, että muuttujien välillä olisi myös selkeä syy-seuraussuhde [4]. Tämän vuoksi Harjoitustehtävien 1 ja 2 viimeiset kohdat liittyvät toisiinsa. Harjoitustehtävän 1 c-kohdan tarkoituksena on toimia esimerkkinä tilanteesta, jossa selitettävän ja selittävän muuttujan valitseminen ei ole niin selkeää ja Harjoitustehtävän 2 c-kohdan tarkoituksena on havainnollistaa sitä, että selitettävä ja selittävä muuttuja riippuvat loppujen lopuksi siitä, millaisella tutkimuskysymyksellä muuttujien välistä riippuvuutta halutaan lähteä tutkimaan, eli kummalla muuttujalla halutaan yrittää selittää toista.

Harjoitustehtävä 3 on Swanin artikkelin mukainen erilaisten esitystapojen tulkintatehtävä ja myös se käsittelee selitettävää ja selittävää muuttujaa [15]. Siinä opiskelijoiden tulisi yhdistää annetut tutkimuskysymykset oikeisiin hajontakuvioihin eli tehtävän tarkoituksena on edistää opiskelijoiden visualisointia Habits of mind -artikkelin tavoitteista [3]. Erityisesti tehtävän tarkoituksena on saada opiskelijat visualisoimaan selitettävä muuttuja aina y-akselille ja selittävä muuttuja aina x-akselille hajontakuviossa.

Harjoitustehtävä 4 on Swanin artikkelin mukainen erilaisten ratkaisutapojen vertailutehtävä ja siinä opiskelijat pääsevät vertailemaan kahdella eri tavalla samalle aineistolle toteutettua regressioanalyysiä sekä tutkimaan poikkeavan havainnon vaikutuksia regressioanalyysin tuloksiin ja korrelaatiokertoimeen [15]. Tehtävässä opiskelijat pääsevät myös harjoittelemaan ennustamista regressiomallin perusteella sekä miettimään perusteita sille, missä tilanteissa toinen toteutetuista regressioanalyyseistä voisi olla parempi kuin toinen. Harjoitustehtävä 4 tukee opiskelijoiden kehitystä kuvailijoiksi Habits of mind -artikkelin tavoitteista, sillä tehtävässä tulee myös perustella omat vastaukset [3].

Harjoitustehtävä 5 on Swanin artikkelin mukainen erilaisten esitystapojen tulkintatehtävä ja sen tarkoituksena on saada opiskelijat yhdistämään korrelaatiokertoimen eri suuruusluokat tietyn näköisiin pistepilviin ja niihin sovitettuihin suoriin [15]. Harjoitustehtävän 5 tarkoituksena on kehittää Habits of mind -artikkelin tavoitteista opiskelijoiden visualisointia korrelaatiokertoimen käsitteen paremmin ymmärtämisen lisäksi [3].

Myös Harjoitustehtävä 6 liittyy korrelaatikertoimeen ja se on Swanin artikkelin mukainen väitteiden arviointitehtävä [15]. Harjoitustehtävän 6 tarkoituksena on kiinnittää opiskelijoiden huomiota siihen, millaisia tulkintoja korrelaatiokertiomen perusteella voidaan tehdä. Tehtävän viimeisen väittämän tarkoituksena on puuttua Engelin ja Sedlmeierin artikkelissa esille tuotuun harhakäsitykseen, jonka mukaan korrelaatiokertoimen arvo tarkoittaisi suoraan, että muuttujien välillä ei ole lainkaan tilastollista riippuvuutta, vaikka tällaisessa tilanteessa muuttujien välillä voi edelleen olla eilineaarista riippuvuutta [4]. Samalla Harjoitustehtävä 6 kehittää opiskelijoita Cuocon ym. artikkelin mukaisiksi kuvailijoiksi, sillä opiskelijoiden tulee perustella vastauksensa tehtävään [3].

Engel ja Sedlmeier ehdottavat, että korrelaation ja regression opetuksessa tulisi hyödyntää myös todellista dataa, sillä opiskelijoille mielenkiintoiseen aiheeseen liittyvän todellisen datan analysoiminen motivoi ja rohkaisee projektityöskentelyä sekä tieteellisen tutkimuksen tekemistä [4]. Tämän vuoksi harjoitustehtävä 7 on koko regressioanalyysin oppimateriaalin käsittelemät asiat kokoava harjoitustehtävä, jossa opiskelijat pääsevät toteuttamaan tilastollisen tutkimuksen täysin itse alusta loppuun saakka. Tehtävässä opiskelijat saavat valita itseään kiinnostavan tutkimuskohteen, kerätä tutkimusta varten havaintoaineiston esimerkiksi luokkakavereiltaan, muodostaa havaintoaineistosta hajontakuvion, suorittaa aineistolle regressioanalyysin sekä määrittää korrelaatiokertoimen, jos he toteavat valitsemiensa muuttujien välillä vallitsevan lineaarista riippuvuutta. Tehtävässä opiskelijoita pyydetään myös tekemään johtopäätöksiä muuttujien välisestä riippuvuudesta tutkimuksensa tulosten perusteella sekä hyödyntämään havaintoaineiston käsittelyssä GeoGebraa.

2.2 Diskreetti todennäköisyysjakauma

Lukion opetussuunnitelman perusteissa diskreetin todennäköisyysjakauman aihepiiriin liittyväksi moduulin MAA8 yhdeksi tavoitteeksi on määritetty, että opiskelija ymmärtää diskreetin todennäköisyysjakauman käsitteen ja oppii määrittämään jakauman odotusarvon sekä tulkitsemaan sitä [13]. Tämän vuoksi diskreetin todennäköisyysjakauman oppimateriaalissa lähdetään liikkelle diskreetin todennäköisyysjakauman käsitteen opettelusta sekä jakauman odotusarvon määrittämisestä ja tulkitsemisesta. Tätä varten opiskelijoiden tulee kuitenkin ymmärtää myös diskreetin satunnaismuuttujan käsite. Pohdinnassa A.13 opiskelijoita johdatellaankin kaikkien näiden kolmen käsitteen pariin sekä samalla myös diskreetin todennäköisyysjakauman odotusarvon tulkintaan. Tehtävään on otettu mallia Sanoma Pron pitkän matematiikan Todennäköisyys ja tilastot -oppikirjan tehtävästä [6].

Pohdinnassa A.13 opiskelijoille on annettu kaksi taulukkoa, joista toinen kuvaa kahden nelisivuisen nopan samanaikaisessa heitossa saatavien silmälukujen tulon eri arvojen todennäköisyyksiä ja toinen arviota saatavista tulon eri arvojen lukumääristä, kun noppien heittoa toistetaan 3200 kertaa. Tehtävänä Pohdinnassa A.13 opiskelijoilla on tarkastella annettuja taulukoita sekä vastata niihin liittyviin kysymyksiin. Ensimmäisestä taulukosta opiskelijoiden tulee tutkia, mikä on todennäköisin silmälukujen tulon arvo yhdellä noppien heitolla ja jälkimmäisestä taulukosta, mikä on saatujen tulon arvojen keskiarvo, kun noppia heitetään 3200 kertaa. Lopuksi opiskelijoiden tulee vielä pohtia, miten tämä arvo voitaisiin määrittää suoraan ensimmäisestä taulukosta ilman, että noppien heittojen lukumäärää kiinnitetään.

Pohdinnassa A.13 kysymysten tarkoituksena on johdatella opiskelijoita diskreetin todennäköisyysjakauman odotusarvon käsitteeseen sekä sen tulkintaan ja saada heidät myös huomaamaan, että odotusarvo ei aina ole sama asia kuin todennäköisin lopputulos. Samalla Pohdinta A.13 antaa opiskelijoille esimerkin diskreetistä satunnaismuuttujasta sekä diskreetistä todennäköisyysjakaumasta, jonka jälkeen oppimateriaalissa käsitellään myös näitä käsitteitä tarkemmin. Habits of mind -artikkelin tavoitteista Pohdinta A.13 kehittää opiskelijoista visualisoijia sekä säännönmukaisuuksien etsijöitä, sillä pohdintatehtävä auttaa opiskelijoita visualisoimaan diskreettiä todennäköisyysjakaumaa sekä tutkimaan diskreetin todennäköisyysjakauman odotusarvoon liittyvää säännönmukaisuutta [3].

Marcos Magalhãesin mukaan yleinen satunnaismuuttujan käsitteeseen liittyvä väärinymmärrys opiskelijoilla on, että satunnaismuuttuja on täysin ennustamattomissa oleva muuttuja, sillä opiskelijat assosioivat satunnaisuuden villin tai kontrolloimattoman kanssa [10]. Tällöin opiskelijat ajattelevat, että satunnaismuuttuja voi saada mitä arvoja tahansa ja ilman minkäänlaista säännönmukaisuutta [10]. Tällaisen väärinymmärryksen syntymisen välttämiseksi oppimateriaalissa otetaan ensin Pohdinnassa A.13 käsittelyyn konkreettinen esimerkki diskreetistä satunnaismuuttujasta, eli tässä tapauksessa kahden nelisivuisen nopan silmälukujen tulo, ja vasta tämän jälkeen oppimateriaalissa määritellään Määritelmässä A.14 satunnaismuuttujan ja erityisesti diskreetin satunnaismuuttujan käsitteet. Tällä tavoin opiskelijoilla on jo olemassa oikeanlainen mielikuva satunnaismuuttujasta ennen käsitteen nimityksen esille tuomista, jolloin käsitteessä mukana olevan satunnaisuuden ei pitäisi saada opiskelijoille aikaan tällaisia väärinymmärryksiä. Lisäksi Määritelmässä A.14 määriteltyjen käsitteiden käyttöä selvennetään vielä Esimerkissä A.15 liittämällä ne tähän Pohdinnassa A.13 esille tuotuun esimerkkitapaukseen.

Diskreetin ja jatkuvan muuttujan ero on käsitelty oppikirjassa jo aiemmin tilastollisen muuttujan yhteydessä, joten tässä oppimateriaalissa siihen ei Määritelmää A.14 tarkemmin enää puututa, harjoitustehtävää 1 lukuunottamatta. Pohdinnassa A.13 käsiteltävä kahden nopan yhtä aikaa heittämisen tilanne on myös jo oppikirjan aiemmista kappaleista tuttu satunnaisilmiö opiskelijoille ja tämän vuoksi Pohdinnassa A.13 on annettu valmiina noppien silmälukujen tulon eri arvojen todennäköisyydet, sillä näiden todennäköisyyksien määrittäminen opiskelijoiden tulisi jo osata.

Magalhãesin mukaan toinen satunnaismuutujan käsitteeseen liittyvä yleinen harhakäsitys opiskelijoilla on, että satunnaismuuttujan kaikki arvot ovat aina keskenään yhtä todennäköisiä. Tämä johtuu Magalhãesin mukaan siitä, että satunnaismuuttujan käsitteen opettaminen on aloitettu sellaisista esimerkeistä, joissa kaikki satunnaismuuttujan arvot ovat yhtä todennäköisiä, jolloin joillekin opiskelijoille syntyy harhakäsitys siitä, että tällainen pätee aina [10]. Tämän vuoksi Pohdintaan A.13 on valittu sellainen esimerkki satunnaismuuttujasta, että sen kaikki arvot eivät ole yhtä todennäköisiä.

Magalhãesin mukaan joillakin opiskelijoilla on myös hankaluuksia erottaa teoreettinen todennäköisyysjakauma ja frekvenssijakauma toisistaan [10]. Tämän vuoksi Pohdintaan A.13 on otettu mukaan esimerkki näistä molemmista jakaumista. Lisäksi Magalhãesin mukaan joillain opiskelijoilla on hankaluuksia ymmärtää, että teoreettinen frekvenssijakauma ei välttämättä takaa sitä, että satunnaismuuttuja käyttäytyy tismalleen jakauman mukaisesti myös todellisuudessa [10]. Tämän vuoksi Pohdinnan A.13 jälkimmäisen taulukon kuvauksessa on pyritty korostamaan sitä, että taulukon kuvaama jakauma noppien silmälukujen tulon eri arvojen lukumääristä on teoreettisena frekvenssijakaumana vain todennäköisyyksien avulla tehty arvio mahdollisesti 3200 heitolla saatavista tulon eri arvojen lukumääristä, jolloin, jos todellisuudessa tällainen koe toteutettaisiin, saattaisivat arvojen lukumäärät vaihdella tehtävässä annetuista lukumääristä. Tämän enempää tätä asiaa ei kuitenkaan tässä oppimateriaalissa käsitellä, sillä tilastollinen todennäköisyys on käsitelty jo oppikirjan aiemmissa osissa.

Satunnaismuuttujan käsitteen määrittelemisen ja siihen liittyvän esimerkin jälkeen opiskelijat pääsevät heti soveltamaan uusia tietojaan Pohdinnassa A.16. Pohdintatehtävässä on annettu kolme eri satunnaismuuttujaa sekä niiden arvojoukot sekaisin lueteltuina. Opiskelijoiden tulisi nyt osata oppimateriaalissa heille annettujen tietojen pohjalta yhdistää satunnaismuuttujat niitä vastaaviin arvojoukkoihin, eli Pohdinta A.16 on Swanin artikkelin mukainen erilaisten esitystapojen tulkintatehtävä [15]. Habits of mind -artikkelin tavoitteista Pohdinta A.16 kehittää opiskelijoita säännönmukaisuuksien etsijöiksi, sillä tehtävän ratkaisemisessa voi hyödyntää oppimateriaalissa annettuja satunnaismuuttujaan liittyviä säännönmukaisuuksia sekä visualisoijiksi, sillä tehtävä auttaa opiskelijoita yhdistämään satunnaismuuttujien sanalliset kuvaukset niitä vastaaviin arvojoukkoihin [3].

Memnunin, Ozbilenin ja Dincin mukaan todennäköisyyslaskennan oppimisen kannalta on tärkeää, että opiskelijoille tarjotaan esimerkkejä myös sellaisista todennäköisyyksiin liittyvistä tilanteista, jotka he pystyvät liittämään omiin elämiinsä [12]. Tämän vuoksi Pohdinnan A.16 satunnaismuuttujat on pyritty valitsemaan hieman arkisemmista tilanteista kuin Pohdintojen A.13 ja A.18 sekä Esimerkkien A.15 ja A.17 kahden nelisivuisen nopan samanaikaisen heiton tilanne, sillä osa opiskelijoista ei välttämättä ole koskaan edes törmännyt harvemmin käytettävään nelisivuiseen noppaan. Kuitenkin tavallista korttipakkaa tai arpakuutiota voi olettaa suurimman osan opiskelijoista joskus käyttäneen, jonka vuoksi Pohdinnan A.16 satunnaismuuttujat on suunniteltu näihin liittyviksi.

Pohdinnan A.16 jälkeen oppimateriaalissa kerrotaan tarkemmin diskreetistä todennäköisyysjakaumasta ja pistetodennäköisyydestä. Tämän jälkeen Esimerkki A.17 liittää opiskelijoille nämä käsitteet Pohdinnassa A.13 esitettyyn ensimmäiseen taulukkoon. Esimerkki A.17 tukee siten opiskelijoiden kehitystä Cuoco ym. artikkelin mukaisiksi visualisoijiksi [3].

Koska Pohdinnan A.13 ensimmäisessä taulukossa on kyseessä diskreetti todennäköisyysjakauma, on satunnaismuuttujan pistetodennäköisyyksien summa siinä yksi. Tämän huomaamiseen opiskelijoita johdatellaan Pohdinnalla A.18. Pohdinta A.18 kehittää opiskelijoista Cuoco ym. artikkelin mukaisia säännönmukaisuuksien etsijöitä, sillä opiskelijat pääsevät ensin itse tutkimaan diskreetin satunnaismuuttujan pistetodennäköisyyksiin liittyvää säännönmukaisuutta, ja vasta tämän jälkeen oppimateriaalin tekstissä kerrotaan tämä säännönmukaisuus opiskelijoille [3]. Samalla oppimateriaalissa myös kerrotaan, mistä tämä säännönmukaisuus johtuu, jonka miettimiseen opiskelijoita on myös johdateltu Pohdinnassa A.18.

Seuraavaksi oppimateriaalissa määritellään Määritelmissä A.19 ja A.20 diskreetin satunnaismuuttujan odotusarvo ja keskihajonta. Tämän jälkeen Pohdinnalla A.21 pyritään toteuttamaan lukion opetussuunnitelmassa määriteltyä tavoitetta, jonka mukaan opiskelijoiden tulisi osata hyödyntää myös ohjelmistoja diskreetin todennäköisyysjakauman havainnollistamisessa ja tunnuslukujen määrittämisessä [13]. Tehtävänä opiskelijoilla on kopioida Pohdinnan A.13 ensimmäisen taulukon arvot GeoGebran "Taulukkolaskenta" -ohjelmistoon ja harjoitella näiden arvojen sekä Pohdinnassa A.21 annettujen ohjeiden avulla diskreetin todennäköisyysjakauman havainnollistamista taulukon lisäksi myös pylväsdiagrammina. Lopuksi Pohdinnassa A.21 opiskelijat pääsevät itse tutkimaan, miten GeoGebralla saadaan "Näytä tilastot" -osiosta diskreetin satunnaismuuttujan odotusarvo ja keskihajonta selville. Tämä ei vaadi opiskelijoilta muuta kuin sen hoksaamista, että GeoGebra ei erikseen näytä odotusarvoa, vaan se löytyy samasta kohtaa kuin keskiarvo. Tämä johtuu siitä, että GeoGebran "Taulukkolaskenta" -ohjelmisto on pääsääntöisesti tarkoitettu tilastollisten jakaumien tunnuslukujen määritykseen, mutta kuten opiskelijoita oppimateriaalissa pyritään saamaan huomaamaan, diskreetin todennäköisyysjakauman odotusarvo voidaan rinnastaa diskreetin tilastollisen jakauman keskiarvoon, vaikka kyseessä ei olekaan sama asia.

Oppimateriaalissa hyödynnettäväksi ohjelmistoksi on todennäköisyyslaskentaan kuuluvaan osuuteenkin valittu GeoGebra, sillä Vágován ja Kmetován mukaan GeoGebra on ohjelmisto, joka parantaa opiskelijoille matematiikan oppimisessa välttämätöntä taitoa eli visualisointia [18]. Näin ollen GeoGebran hyödyntäminen oppimateriaalissa tukee samalla myös opiskelijoiden kehittymistä visualisoijiksi oppikirjaan valituista Habits of mind -artikkelin yleisistä tavoitteista [3].

Pohdinnan A.21 jälkeen oppimateriaalissa siirrytään käsittelemään toistokokeiden todennäköisyyksiä. KimberLeigh Hadfieldin mukaan todennäköisyyslaskenta on monelle opiskelijalle vaikeaa, sillä todennäköisyyteen liittyvien matemaattisten ongelmien ratkaiseminen koostuu useasta eri vaiheesta. Hadfieldin mukaan tämä monivaiheinen prosessi osoittautuu monelle opiskelijalle liian työlääksi kognitiiviseksi prosessiksi ja todennäköisyyslaskennan opetuksessa tulisikin pyrkiä keventämään tätä kognitiivista prosessia opiskelijoilta. Hadfieldin esittelemän todennäköisyyslaskennan ongelmien ratkaisemisesta aiheutuvan kognitiivisen prosessin yksi vaihe on tunnistaa, mitä todennäköisyyden laskusääntöjä tehtävän ratkaisemisessa tarvitaan [5]. Pohdinnalla A.22 opiskelijoita pyritään johdattelemaan vaiheittaisesti toistokokeen todennäköisyyksien laskemiseen. Tehtävästä aiheutuvaa kognitiivisen prosessin taakkaa on pyritty oppimateriaalissa keventämään antamalla opiskelijoille ennen Pohdintaa A.22 tieto siitä, mitä oppikirjassa aiemmin jo käsiteltyjä todennäköisyyslaskennan laskusääntöjä kyseisen tehtävän ratkaisemisessa tarvitaan.

Memnunin ym. artikkelin mukaan opiskelijoilla on todettu olevan hankaluuksia määrittää vastatapahtuman todennäköisyyttä [12]. Tämän vuoksi, vaikka vastatapahtuman todennäköisyys on käsitelty jo oppikirjan aiemmissa osissa, Pohdinnassa A.22 opiskelijoiden johdatteleminen toistokokeen todennäköisyyden laskemiseen aloitetaan sillä, että opiskelijoiden tulee ensin miettiä Pohdinnan A.22 satunnaisilmiön kerran tapahtumisen sekä kerran tapahtumatta jäämisen todennäköisyydet. Tällä tavoin, kun tapahtuman ja sen vastatapahtuman todennäköisyyksien määrittäminen on heti tehtävän alussa omana kohtanaan, pystyvät myös opettajat helpommin huomaamaan, jos jotkut opiskelijat tarvitsevat kyseisen asian oppimisessa lisää tukea.

Seuraavaksi opiskelijoille annetaan Pohdinnassa A.22 kaksi eri tilannetta, joissa heidän pitää palauttaa mieleen riippumattomien tapahtumien kertolaskusäännön soveltaminen. Tämän jälkeen Pohdinta A.22 auttaa opiskelijoita palauttamaan mieleen binomikertoimen käytön ja viimeisenä erillisten tapahtumien yhteenlaskusäännön käytön. Pohdinnan A.22 tarkoituksena on auttaa opiskelijoita huomaamaan ne säännönmukaisuudet, joiden perusteella toistokokeen todennäköisyyden laskukaava on määritetty eli tehtävä tukee Habits of mind -artikkelin tavoitteista opiskelijoiden kehitystä säännönmukaisuuksien etsijöiksi [3]. Tämän jälkeen oppimateriaalissa määritellään toistokokeen laskukaava Määritelmässä A.23. Pohdintaan A.22 on otettu mallia Sanoma Pron pitkän matematiikan Todennäköisyys ja tilastot -oppikirjan tehtävästä [6]. Seuraavaksi oppimateriaalissa käsitellään vielä binomijakauma sekä Määritelmässä A.24 binomijakaumaa noudattavan satunnaismuuttujan odotusarvo ja keskihajonta.

Lukion opetussuunnitelman perusteissa diskreetin todennäköisyysjakauman aihepiiriin liittyväksi yhdeksi tavoitteeksi on määritelty, että opiskelija osaa hyödyntää ohjelmistoja diskreetin todennäköisyysjakauman havainnollistamisessa, tunnuslukujen määrittämisessä sekä todennäköisyyksien laskemisessa [13]. Oppimateriaalin viimeisessä pohdintatehtävässä opiskelijat pääsevät harjoittelemaan näitä tavoitteita GeoGebralla binomijakauman näkökulmasta.

Memnunin ym. artikkelin mukaan opiskelijoilla on todettu olevan hankaluuksia myös satunnaisilmiön perusjoukon määrittämisen kanssa [12]. Tämän vuoksi Pohdinnassa A.25 lähdetään liikkeelle siitä, että opiskelijoiden tulee ensin ilman GeoGebraa määrittää toistokokeessa tarkasteltava tapahtuma sekä tapahtuman toteutumisen todennäköisyys. Tällä tavoin, kun tarkasteltavan tapahtuman toteutumisen todennäköisyyden laskeminen (joka siis sisältää yhtenä vaiheena satunnaisilmiön perusjoukon määrittämisen) on tehtävän alussa omana kohtanaan, pystyvät opettajat helpommin tarjoamaan lisää tukea todennäköisyyslaskennan peruskäsitteiden oppimista varten niille opiskelijoille, jotka vielä tässä vaiheessa oppikirjaa kokevat hankaluuksia näiden asioiden kanssa.

Tämän jälkeen Pohdinnassa A.25 siirrytään harjoittelemaan itse GeoGebran hyödyntämistä. Tarkoituksena Pohdinnan A.25 c- ja d-kohdissa on, että opiskelijat saavat itse tutkia, mistä kohtaa GeoGebralla löytää kohdissa käsiteltävät tiedot. Tämä tukee opiskelijoiden kehittymistä Habits of mind -artikkelin mukaisiksi säännönmukaisuuksien etsijöiksi [3]. Tehtävän e- ja f-kohdissa opiskelijoita ohjataan tutkimaan sitä, miten GeoGebralla pystyy määrittämään binomitodennäköisyyksiä erilaisissa tilanteissa. Pohdinnan A.25 viimeisessä kohdassa oletetaan, että opiskelijat osaavat jo laskea ”korkeintaan”, ”vähintään” yms. tilanteisiin liittyviä todennäköisyyksiä, sillä nämä asiat on käsitelty jo oppikirjan aiemmissa osissa.

Myös diskreetin todennäköisyysjakauman osuuden oppimateriaali päättyy harjoitustehtäviin, jotka on suunniteltu niin, että ne auttavat opiskelijoita ymmärtämään oppimateriaalissa käsitellyt asiat paremmin sekä saavuttamaan lukion opetussuunnitelmassa määritellyt oppimisen tavoitteet diskreetin todennäköisyysjakauman aihepiirin osalta [13]. Harjoitustehtävän 8 tavoitteena on syventää opiskelijoiden ymmärrystä diskreetin satunnaismuuttujan käsitteestä. Samalla Harjoitustehtävä 8 kehittää opiskelijoista Cuoco ym. artikkelin mukaisia kuvailijoita, sillä tehtävässä opiskelijoiden tulee perustellusti valita annetuista satunnaismuuttujista dikreetit satunnaismuuttujat [3].

Harjoitustehtävä 9 on Swanin artikkelin mukainen erilaisten esitystapojen tulkintatehtävä, jossa opiskelijoille on annettu kahden eri diskreetin satunnaismuuttujan todennäköisyysjakaumat sekaisin sekä taulukkoina että pylväsdiagrammeina ja tehtävänä opiskelijoilla on yhdistää saman todennäköisyysjakauman eri esitystavat toisiinsa [15]. Opiskelijoita pyydetään tehtävässä myös perustelemaan vastauksensa eli Habits of mind -artikkelin tavoitteista Harjoitustehtävä 9 tukee opiskelijoiden kehitystä kuvailijoiksi ja visualisoijiksi [3]. Lukion opetussuunnitelman perusteissa määritellyistä oppimisen tavoitteista tässä harjoitustehtävässä keskitytään diskreetin todennäköisyysjakauman käsitteen paremmin ymmärtämiseen [13].

Harjoitustehtävä 10 on diskreetin todennäköisyysjakauman oppimateriaalin alkupuolen aiheet kokoava harjoitustehtävä, jossa opiskelijat pääsevät hyödyntämään myös GeoGebraa. Habits of mind -artikkelin tavoitteista tässä harjoitustehtävässä toteutuu opiskelijoiden kehittyminen visualisoijiksi [3]. Harjoitustehtävä 11 taas keskittyy diskreetin todennäköisyysjakauman oppimateriaalin alkupuolen aiheista tarkemmin odotusarvon määrittämiseen ja sen tulkintaan. Tehtävän on suunniteltu kehittävän opiskelijoista kuvailijoita Habits of mind -artikkelin tavoitteiden mukaisesti [3]. Harjoitustehtävä 12 on koko binomijakauman aihepiirin kokoava harjoitustehtävä, jonka ratkaisemisessa opiskelijoilla on jälleen mahdollisuus hyödyntää halutessaan myös GeoGebraan. Habits of mind -artikkelin tavoitteista siinä toteutuu kuvaileminen [3].

Lähdeluettelo

[1] Batanero, C., Estepa, A. & Godino, J. (1997). Evolution of Students' Understanding of Statistical Sssociation in a Computer-based Teaching Environment. Teoksessa J. Garfield & G. Burrill (toim.), *Research on Teaching Statistics and New Technologies* (s. 191–206). International Statistical Institute.

[2] Casey, S. A. (2014). Teachers' Knowledge of Students' Conceptions and Their Development When Learning Linear Regression. Teoksessa K. Makar (toim.), *Sustainability in Statistics Education: Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, Arizona, USA* International Statistical Institute.

[3] Cuoco, A., Goldenberg, E. P. & Mark, J. (1996). Habits of Mind: An Organizing Principle for Mathematics Curricula. *The Journal of Mathematical Behavior, 15*(4), 375–402.

[4] Engel, J. & Sedlmeier, P. (2013). Correlation and Regression in the Training of Teachers. Teoksessa C. Batanero, G. Burrill & C. Reading (toim.), *Teaching Statistics in School Mathematics — Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study: The 18th ICMI Study* (s. 247–258). Springer.

[5] Hadfield, K. F. (2021). Providing Ability to Probability: Reducing Cognitive Load Through Worked-out Examples. *Teaching Statistics, 43*(1), 28-35.

[6] Heiskanen, P., Kaakinen, P., Lehtonen, J., Leikas, M. & Tahvanainen, J. (2020). *Tekijä Pitkä matematiikka 10: Todennäköisyys ja tilastot (LOPS 2016)*. Sanoma Pro.

[7] Hellsten, H., Hellsten, L., Nurmi, K., Oikarinen, J. & Salminen-Saari, J. (2017). *MAB5 Tilastot ja todennäköisyys (LOPS2016)*. Studeo Oy.

[8] Hewson, P. (2009). GeoGebra for Mathematical Statistics. *International Journal for Technology in Mathematics Education, 16*(4).

[9] Huey, M. E. & Baker, D. L. (2015). Tackling Misconceptions about Linear Associations. *The Mathematics Teacher, 109*(1), 46–53.

[10] Magalhães, M. N. (2014). Challenges for Learning About Distributions in Courses for Future Mathematics Teachers. Teoksessa K. Makar (toim.), *Sustainability in Statistics Education: Proceedings of the Ninth International Conference on Teaching Statistics, Flagstaff, Arizona, USA* International Statistical Institute.

[11] MAOL. (17.1.2020). *LOPS 2019 Pitkän matematiikan tukimateriaalia.* https://maol.fi/app/uploads/2020/01/LOPS2019_MAA_MAOL.pdf

[12] Memnun, D. S., Ozbilen, O. & Dinc, E. (2019). A Qualitative Research on the Difficulties and Failures about Probability Concepts of High School Students. *Journal of Educational Issues, 5*(1), 1-19.

[13] Opetushallitus. (7.11.2019). *Lukion opetussuunnitelman perusteet 2019.* https://www.oph.fi/sites/default/files/documents/lukion_opetussuunnitelman_perusteet_2019.pdf

[14] Oulun yliopisto. (julkaisuaika tuntematon). *Avoin oppikirja.* Haettu 5.5.2022 osoitteesta https://www.oulu.fi/fi/projektit/avoin-oppikirja

[15] Swan, M. (2006). *Collaborative Learning in Mathematics: A Challenge to Our Beliefs and Practices.* National Institute of Adult Continuing Education.

[16] Tilastokeskus. (julkaisuaika tuntematon). *Käsitteet.* Haettu 23.5.2022 osoitteesta https://www.stat.fi/meta/kas/index.html

[17] Tilastokeskus. (julkaisuaika tuntematon). *Pearsonin korrelaatiokerroin*. Haettu 11.6.2022 osoitteesta https://tilastokoulu.stat.fi/verkkokoulu_v2.xql?course_id=tkoulu_tilaj&lesson_id=4&subject_id=4&page_type=sisalto

[18] Vágová, R. & Kmetová, M. (2019). GeoGebra, a Tool to Improve Students' Visual Imaging. *Acta Didactica Napocensia*, *12*(2), 225-237.

A Regressio ja diskreetti todennäköisyysjakauma

A.1 Regressio

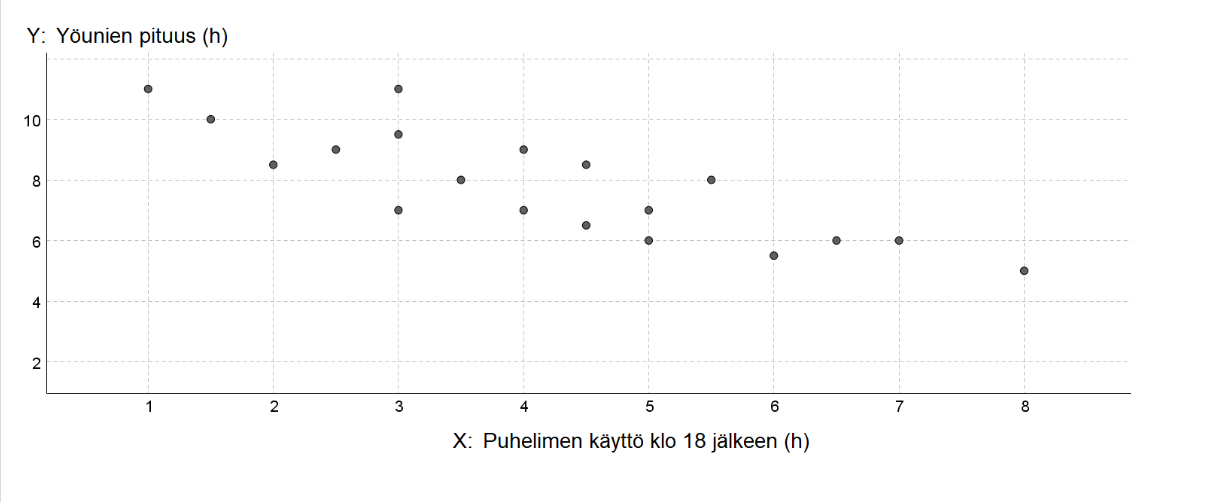
Pohdinta A.1 Pitävätkö seuraavat väittämät aina, joskus vai ei koskaan paikkaansa? Perustele vastauksesi.

1. Ihmisen paino riippuu siitä, kuinka pitkä ihminen on.
2. Opiskelijan silmien väri riippuu siitä, kuinka hyvin tilastotiedettä opiskelija osaa.
3. Ostaessasi kaupasta irtokarkkeja, joiden kilohinta on vakio, karkkien lopullinen hinta riippuu ostettavien karkkien määrästä.

Usein asioiden välisen riippuvuuden olemassaoloa ei voida suoraan tietää tai päätellä, kuten Pohdinnassa A.1, vaan riippuvuuden olemassaoloa, muotoa ja voimakkuutta täytyy tutkia tilastollisella tutkimuksella. Tilastotieteessä kahden asian välistä riippuvuutta voidaan tutkia *kahden muuttujan yhteisjakauman* avulla. Määritellään seuraavaksi muutama tähän liittyvä käsite.

Määritelmä A.2 *Hajontakuvio* on pistepilvi, joka syntyy, kun kahdesta muuttujasta samaan havaintoaineistoon kerätyt havaintoarvot sijoitetaan näiden kahden muuttujan muodostamaan koordinaatistoon. Hajontakuviossa y-akselille sijoitetaan tutkimuksen pääkohteena oleva muuttuja eli *selitettävä muuttuja* (käytetään myös nimitystä *riippuva muuttuja*) ja x-akselille *selittävä muuttuja* (käytetään myös nimitystä *riippumaton muuttuja*), jolla selitettävän muuttujan arvojen vaihtelua pyritään selittämään.

Esimerkki A.3 Elisa halusi selvittää, voiko yöunien pituus riippua siitä, mihin aikaan illasta puhelimen käytön lopettaa. Hän teetti luokkakavereilleen kyselyn, jossa kysyttiin puhelimen käytön lopettamisajankohta edellisenä iltana ja viimeisimpien yöunien pituus, molemmat puolen tunnin tarkkuudella. Alla on hajontakuvio Elisan teettämän kyselyn tuloksista. Hänen tutkimuksessaan yöunien pituus on selitettävä muuttuja ja puhelimen käytön lopettamisajankohta selittävä muuttuja.



Pohdinta A.4 Alla on lueteltuina muuttujia sekä eri tilastollisten tutkimusten tutkimuskysymyksiä. Yhdistä tilastollisen tutkimuksen tutkimuskysymys (1-3) ja tutkimuksessa pääkohteena oleva muuttuja eli tutkimuksen selitettävä muuttuja (a-f) toisiinsa. Huomaa, että muuttujia jää myös yli.

1. Miten opiskelijoiden matemaattinen osaaminen vaikuttaa opiskelijoiden fysikaaliseen osaamiseen?
2. Riippuuko liikenteessä tapahtuvien liikenneonnettomuuksien lukumäärä vuorokauden ajasta?
3. Vaikuttaako ikä siihen, kuinka paljon kierrättää?

a) Opiskelijoiden matemaattinen osaaminen

b) Liikenteessä tapahtuvien liikenneonnettomuuksien lukumäärä

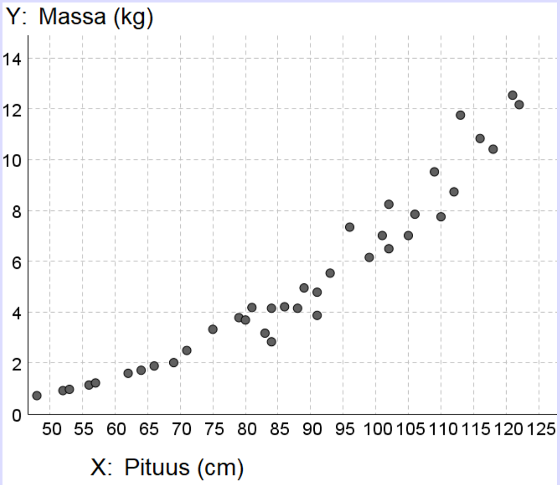
c) Vuorokauden aika

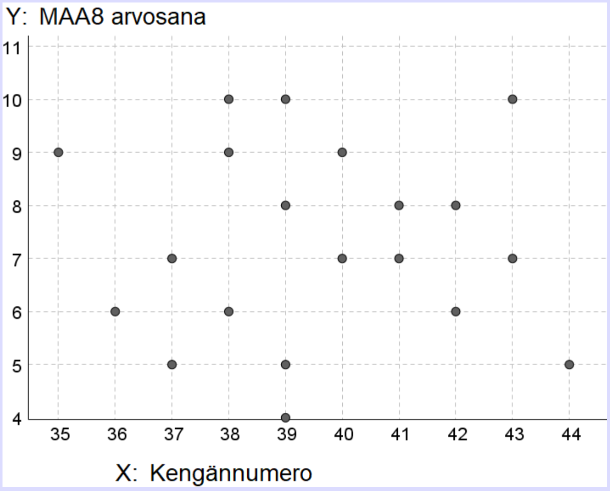
d) Ikä

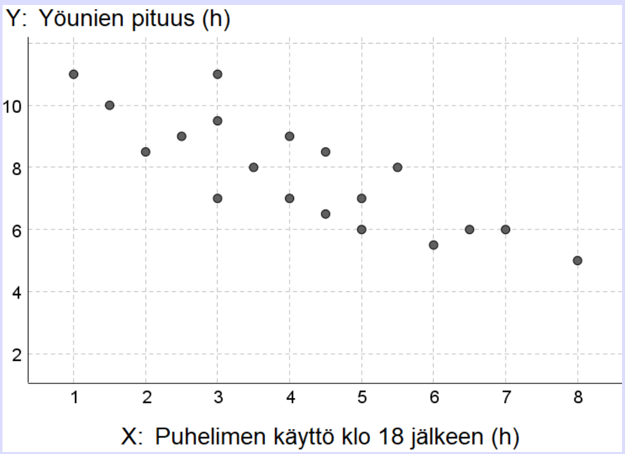
e) Opiskelijoiden fysikaalinen osaaminen

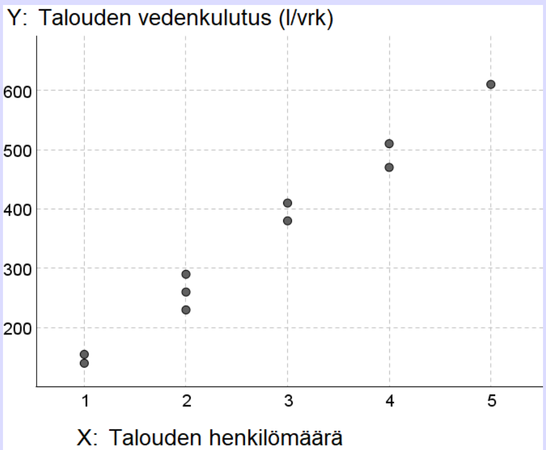
f) Kierrättämisen määrä

Pohdinta A.5 Alla on neljä hajontakuviota, joista ensimmäinen esittää eräällä kalastusalueella tutkittujen haukien painoja ja pituuksia, toinen eräiden opiskelijoiden MAA8 moduulin arvosanoja ja kengännumeroita, kolmas Esimerkin A.3 opiskelijoiden yöunien pituuksia ja puhelimen käytön lopettamisajankohtia sekä neljäs erään taloyhtiön kotitalouksien keskimäärin vuorokaudessa kuluttamia vesimääriä ja kotitalouksissa asuvien henkilöiden määriä. Hajontakuvioiden muuttujien välistä riippuvuutta pitäisi pistepilvien lisäksi kuvata nyt myös jollakin matemaattisella mallilla. Tarkastele annettuja hajontakuvioita ja mieti, voitaisiinko hajontakuvioiden pistepilviä mallintaa suoralla. Perustele vastauksesi.



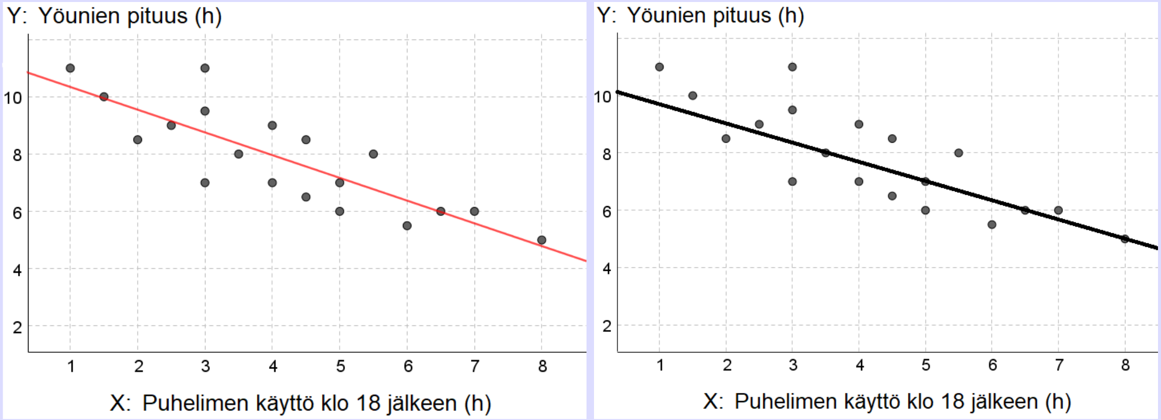






Tilastotieteessä *regressioanalyysillä* voidaan mallintaa muuttujien välisiä riippuvuussuhteita. Regressioanalyysillä saadaan aikaan *regressioyhtälö*, eli matemaattinen regressiomalli, joka kuvaa yhden selitettävän muuttujan ja yhden tai useamman selittävän muuttujan välistä tilastollista yhteyttä. Yksinkertaisimmassa tapauksessa, eli silloin, kun muuttujien välistä tilastollista riippuvuutta voidaan kuvata suoralla, regressioanalyysi tuottaa *lineaarisen regressiosuoran*. Muissa tapauksissa puhutaan *regressiokäyrästä*.

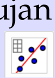
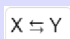
Pohdinta A.6 Alla on kaksi Esimerkin A.3 hajontakuvion pistepilveen sovitettua suoraa. Tarkastele, miten suorat asettuvat Esimerkissä A.3 esitellyssä tutkimuksessa kerättyihin havaintopisteisiin nähden ja perustele, kumpi suorista kuvaa havaintopisteitä paremmin.



Regressioyhtälöllä tarkoitetaan sellaisen kuvaajan yhtälöä, joka on mahdollisimman lähellä jokaista havaintopistettä, joihin kyseinen kuvaaja on sovitettu. Matemaattisesti tällainen yhtälö saadaan määritettyä laskemalla niiden pisteiden pystysuuntaiset etäisyydet regressiokäyrästä, joiden kautta kuvaaja ei kulje ja minimoimalla näiden etäisyyksien neliöiden summan. Tällaisesta menetelmästä käytetään myös nimitystä *pienimmän neliösumman menetelmä*.

Regressioyhtälön määrittäminen käsin on hyvin työlästä, joten sen määrittämisessä hyödynnetään usein erilaisia ohjelmistoja, kuten GeoGebran ”Taulukkolaskenta” -ohjelmistoa. Seuraavassa pohdintatehtävässä harjoitellaan GeoGebran käyttöä.

Pohdinta A.7 Tehtävän lopussa on Esimerkissä A.3 Elisan keräämä havaintoaineisto taulukoidussa muodossa. Avaa aineisto ja harjoittele hajontakuvion tekemistä sekä regressiosuoran sovittamista GeoGebralla seuraavien ohjeiden avulla.

- Hajontakuviota varten maalaa molempien muuttujien arvot ja valitse vasemmassa yläkulmassa olevasta painikkeesta  aukeavasta valikosta ”Kahden muuttujan regressioanalyysi”. Puhelimella tai tabletilla valitse aukeavasta valikosta .
- Hajontakuvion asetukset voit muokata mieluisiksi hajontakuvion yläpuolella olevasta painikkeesta  aukeavasta ”Kuvaaja” -valikosta (”Automaattinen mittasuhde” -kohdan ☑ tulee ottaa pois ennen asetusten muokkaamista).
- Oikeassa yläkulmassa olevasta painikkeesta  voi tarvittaessa vaihtaa x- ja y-akseleille sijoitetut muuttujat toisinpäin.
- Regressiosuoran havaintopisteisiin saa sovitettua hajontakuvion alapuolella olevasta ”Regressiomalli” -valikosta.

Vastaa lopuksi seuraaviin kysymyksiin.

1. Mikä on tekemäsi regressioanalyysin avulla aikaan saatu regressioyhtälö?
2. Ennusta regressioyhtälön avulla, minkä pituiseksi Elisan Esimerkissä A.3 tekemän tutkimuksen perusteella yöunien voidaan keskimäärin arvioida jäävän, jos puhelimen käyttöä jatkaa vielä yhdeksän tuntia klo 18 jälkeen.

https://www.geogebra.org/classic/r7g2qmje

Regressioyhtälön avulla voidaan ennustaa, minkä arvon selitettävä muuttuja *keskimäärin* saa tietyllä selittävän muuttujan arvolla.

Lineaarisen regression tapauksessa kahden muuttujan välistä riippuvuutta kuvaava regressioyhtälö on aina muotoa

$$y = bx + a,$$

jossa b on regressiosuoran kulmakerroin ja a regressiosuoran vakiokerroin. Regressiosuoran kulmakerroin kuvaa sitä, kuinka paljon selitettävä muuttuja keskimäärin muuttuu selittävän muuttujan arvon kasvaessa yhdellä yksiköllä ja regressiosuoran vakiokerroin sitä, minkä arvon selitettävä muuttuja keskimäärin saa selittävän muuttujan arvon ollessa nolla.

Regressiomallin ennustamiskyky on sitä parempi, mitä lähempänä regressiokäyrää havaintopisteet sijaitsevat. Myös muuttujien välinen tilastollinen riippuvuus on sitä voimakkaampaa, mitä lähempänä regressiokäyrää havaintopisteet sijaitsevat.

Kahden muuttujan välisen tilastollisen riippuvuuden voimakkuutta voidaan myös mitata. Tällöin puhutaan *korrelaatiosta*, joka on kahden muuttujan välistä tilastollista riippuvuutta kuvaava mitta ja *Pearsonin korrelaatiokertoimesta*, joka taas on kahden muuttujan välisen *lineaarisen* riippuvuuden mitta.

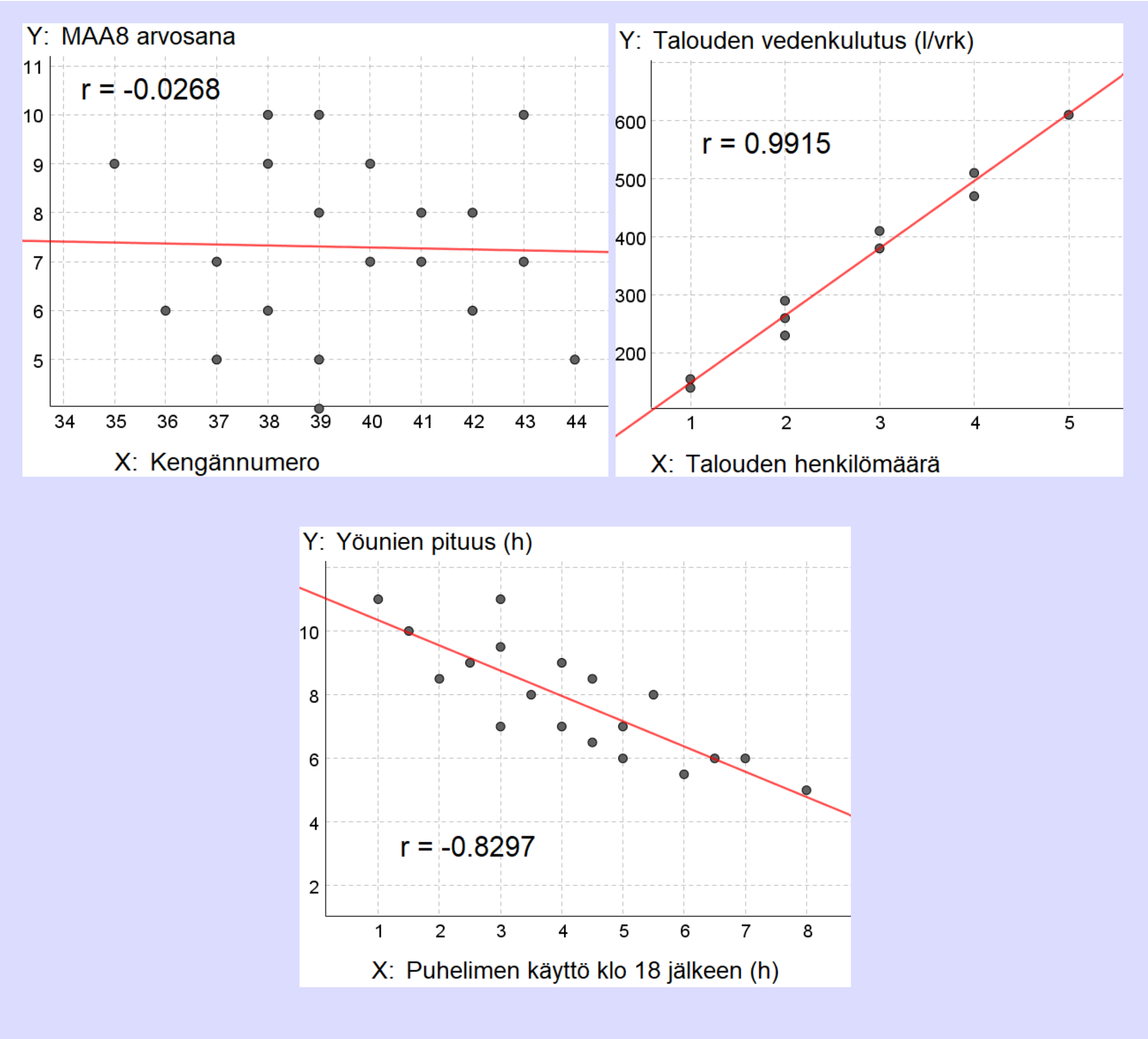
Puhekielessä sekä korrelaatiolla että korrelaatiokertoimella viitataan usein Pearsonin korrelaatiokertoimeen ja jatkossa myös tässä oppikirjassa korrelaatiokertoimella tarkoitetaan juuri Pearsonin korrelaatiokerrointa. Korrelaatiokerrointa merkitään tilastotieteessä kirjaimella r.

Pohdinta A.8 Alla on kolme hajontakuviota, joihin on GeoGebralla sovitettu lineaarinen regressiosuora sekä määritetty korrelaatiokerroin r. Tutki kuvia ja vastaa seuraaviin kysymyksiin.

a) Minkä suuruisia arvoja korrelaatiokerroin r saa näissä hajontakuvioissa?

b) Milloin korrelaatiokerroin on positiivinen ja milloin negatiivinen?

c) Milloin korrelaatiokertoimen itseisarvo $|r|$ on suurin ja milloin pienin?



Korrelaatiokerroin r voi saada arvoja vain väliltä [-1, 1]. Mitä lähempänä lukua 1 korrelaatiokertoimen itseisarvo $|r|$ on, sitä voimakkaampaa muuttujien välinen lineaarinen riippuvuus on. Jos korrelaatiokertoimen itseisarvo on tasan 1, sijoittuvat havaintopisteet täydellisesti regressiosuoralle.

Jos havaintopisteitä voidaan mallintaa nousevalla suoralla, puhutaan *positiivisesta korrelaatiosta*, sillä tällöin $r > 0$. Jos taas havaintopisteitä voidaan mallintaa laskevalla suoralla, puhutaan *negatiivisesta korrelaatiosta*, sillä tällöin $r < 0$.

Huomautus A.9 Korrelaatiokerroin ja lineaarisen regressiosuoran kulmakerroin saavat aina molemmat joko positiivisen arvon tai molemmat negatiivisen arvon, sillä laskevan regressiosuoran kulmakerroin $b < 0$ ja nousevan regressiosuoran kulmakerroin $b > 0$. Korrelaatiokerroin ja regressiosuoran kulmakerroin eivät kuitenkaan ole sama asia, sillä korrelaatiokerroin voi saada vain arvoja väliltä [-1, 1] ja regressiosuoran kulmakerroin taas mitä tahansa reaalilukuarvoja. Lisäksi matemaattisesti lineaarisen regressiosuoran kulmakerroin voidaan laskea kaavalla

$$b = r\frac{s_y}{s_x}$$

ja vakiokerroin kaavalla

$$a = \bar{y} - b\bar{x},$$

missä r on korrelaatiokerroin, s_x ja s_y muuttujien x ja y otoskeskihajonnat ja $\bar{x}, \bar{y}$ muuttujien x ja y keskiarvot.

Kahden muuttujan välisen lineaarisen riippuvuuden voimakkuutta voidaan karkeasti jaotella seuraavan asteikon mukaisesti.

korrelaatio	merkityksetön	kohtalainen	huomattava	voimakas
$\lvert r\rvert$	$0 - 0.3$	$0.3 - 0.6$	$0.6 - 0.8$	$0.8 - 1$

Matemaattisesti korrelaatiokerroin määritetään kaavalla

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y},$$

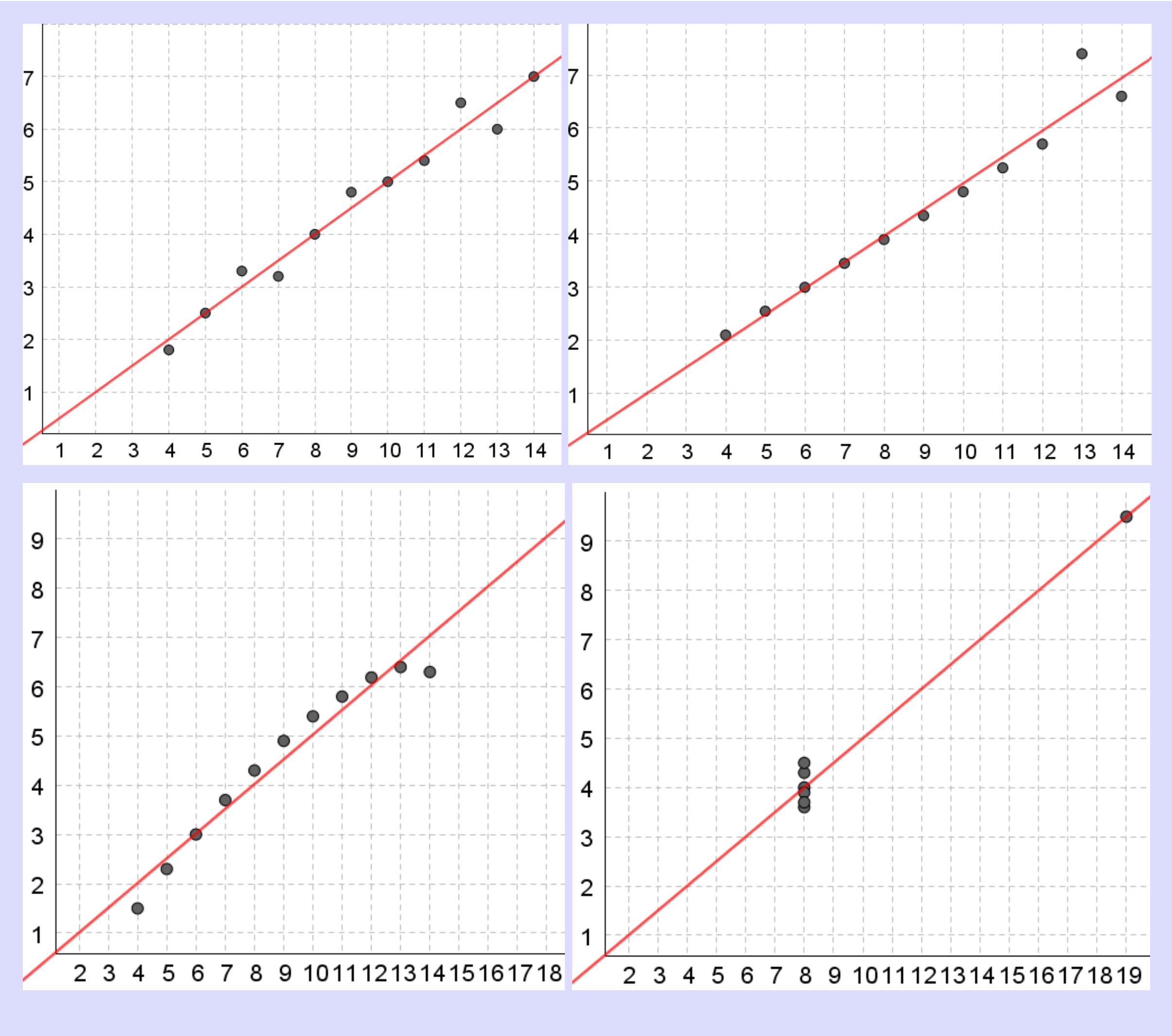
missä n on kahden muuttujan havaintoaineistossa havaintopisteiden (x_i, y_i) lukumäärä, s_x, s_y muuttujien x ja y otoskeskihajonnat ja $\bar{x}, \bar{y}$ muuttujien x ja y keskiarvot.

Korrelaatiokertoimen määritys tehdään kuitenkin yleensä jollakin ohjelmistolla, kuten GeoGebran "Taulukkolaskenta" -ohjelmistolla. GeoGebrassa korrelaatiokerroin löytyy painikkeesta Σx aukeavasta "Näytä tilastot" -osiosta sen jälkeen, kun havaintopisteisiin on ensin sovitettu regressiosuora. Harjoittele korrelaatiokertoimen määritystä GeoGebralla Pohdinnan A.7 aineiston avulla.

Pohdinta A.10 Alla on neljä hajontakuviota, joihin on GeoGebralla sovitettu lineaarinen regressiosuora sekä määritetty korrelaatiokerroin r. Kaikissa neljässä tilanteessa regressioyhtälöksi saatiin $y = 0.5x$ ja korrelaatiokertoimeksi $r = 0.99$. Tarkastele hajontakuvioita ja vastaa seuraaviin kysymyksiin. Perustele vastauksesi.

a) Kuvaako lineaarinen regressiosuora jokaista hajontakuviota yhtä hyvin?

b) Tarkastele oikeanpuoleisia hajontakuvioita. Poikkeaako jokin havaintopiste selkeästi muista? Muuttuisiko havaintopisteisiin sovitettu regressiosuora, jos tämä muista mahdollisesti poikkeava havaintopiste poistettaisiin havaintoaineistosta?



Huomautus A.11 Vaikka korrelaatiokerroin kuvaa kahden muuttujan välisen lineaarisen riippuvuuden voimakkuutta, ei vahva korrelaatio pelkästään vielä takaa sitä, että muuttujien välinen riippuvuus on juuri lineaarista, kuten Pohdinnassa A.10 huomattiin. Tämän vuoksi kahden muuttujan välistä riippuvuutta tuleekin aina ensin tutkia hajontakuvion avulla. Hajontakuvio myös yleensä paljastaa, jos havaintoaineistoon on päässyt mukaan esimerkiksi virheellinen mittaustulos, joka poikkeaa huomattavasti muusta havaintoaineistosta. Tällaiset *poikkeavat havainnot* voidaan poistaa havaintoaineistosta ennen regressioanalyysin suorittamista, kunhan niiden virheellisyydestä voidaan olla varmoja.

Huomautus A.12 Regressioanalyysillä saadaan tietoa muuttujien välisestä tilastollisesta riippuvuudesta ja korrelaatiokertoimen määrityksellä tietoa muuttujien välisen lineaarisen riippuvuuden voimakkuudesta. Kummankaan perusteella ei kuitenkaan voida tehdä päätelmiä muuttujien välisistä syy-seuraussuhteista. Esimerkiksi aurinkorasvan myyntilukujen ja ihonsa polttaneiden ihmisten välillä voidaan todeta olevan positiivista korrelaatiota, mutta kumpikaan ei ole seurausta toisesta.

Harjoitustehtävät

1. Haluat tutkia, onko seuraavien muuttujien välillä tilastollista riippuvuutta. Kumman muuttujista valitsisit selitettäväksi muuttujaksi ja kumman selittäväksi muuttujaksi? Perustele vastauksesi.

   a) Taloudessa vuorokauden aikana kulutettu sähkömäärä ja taloudessa asuvien henkilöiden määrä.

   b) Auton nastarenkailla ajettu matka ja auton nastarenkaista irronneiden nastojen lukumäärä.

   c) Tuotteen kysyntä ja tuotteen hinta.

2. Alla on muutaman tilastollisen tutkimuksen tutkimuskysymys. Kumpi kysymyksien muuttujista on selitettävä muuttuja ja kumpi selittävä muuttuja?

   a) Vaikuttaako vuodenaika alkoholijuomien kulutukseen?

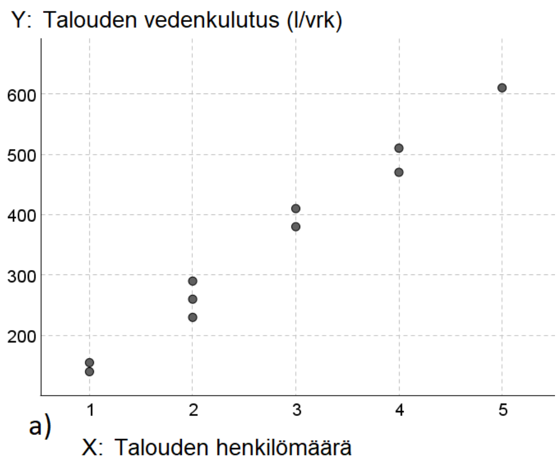
   b) Miten poissaolojen lukumäärä vaikuttaa matematiikan kurssin arvosanaan?

   c) Miten tuotteen kysyntä riippuu tuotteen hinnasta?

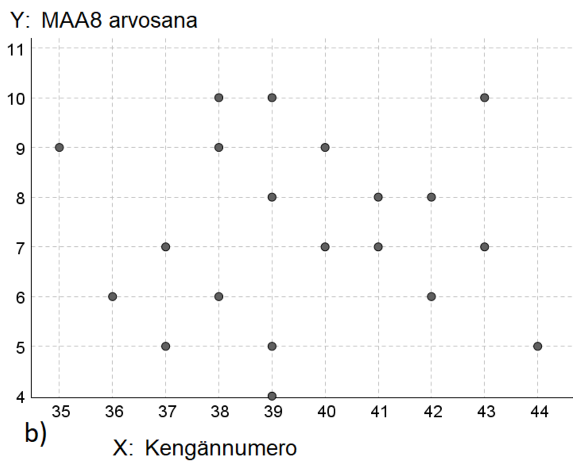
3. Alla on kahden tilastollisen tutkimuksen tutkimuskysymykset sekä neljä eri hajontakuviota. Yhdistä tutkimuskysymykset (1-2) tutkimuksissa saatujan tuloksia kuvaaviin hajontakuvioihin (a-d). Huomaa, että hajontakuvioita jää myös yli.

   1. Miten opiskelijoiden kengännumerot riippuvat opiskelijoiden MAA8 moduulin arvosanoista?

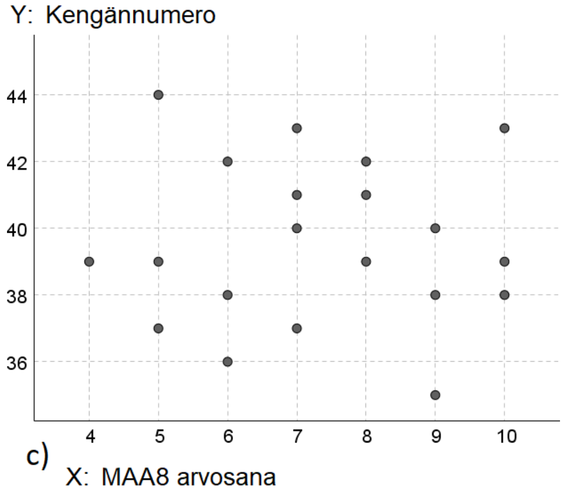
   2. Vaikuttaako kotitaloudessa asuvien henkilöiden lukumäärä kotitaloudessa keskimäärin vuorokaudessa kulutettuun vesimäärään?

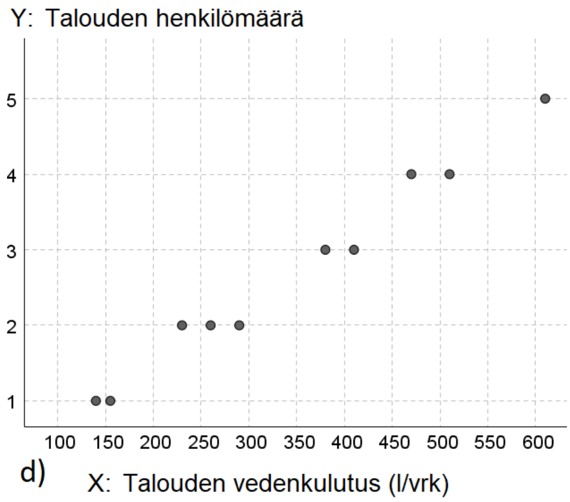


a)



b)





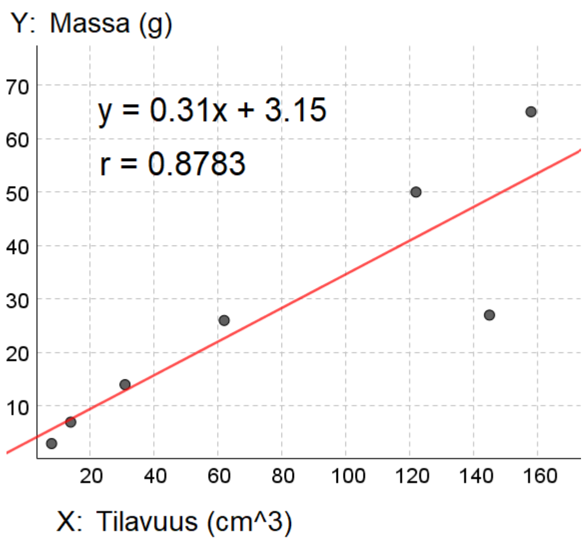
4. Opiskelijat tutkivat puupalojen massan riippuvuutta tilavuudesta. Alla on Alinan (vas.) ja Matin (oik.) tekemät regressioanalyysit samasta havaintoaineistosta. Tarkastele analyysien tuloksia ja vastaa seuraaviin kysymyksiin. Perustele vastauksesi.

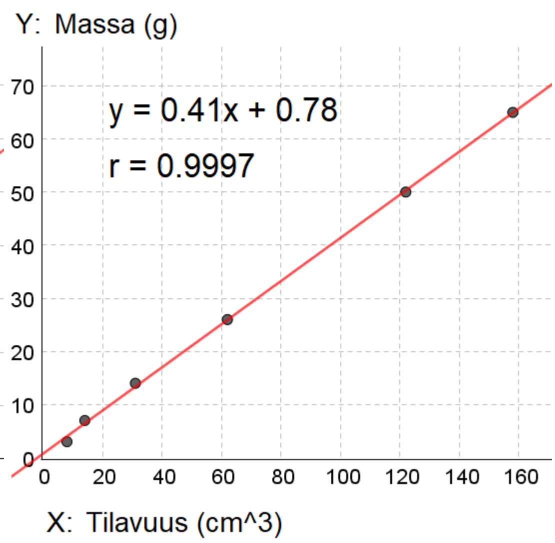
   a) Miten puupalan massa riippuu puupalan tiheydestä?

   b) Mitä eroa Alinan ja Matin regressioanalyysien tuloksissa on? Mistä ero johtuu?

   c) Missä tilanteessa Matin regressioanalyysin tuottama regressiomalli kuvaa paremmin puupalojen massan riippuvuutta puupalojen tilavuudesta? Entä missä tilanteessa taas Alinan tuottama regressiomalli riippuvuudelle on parempi?

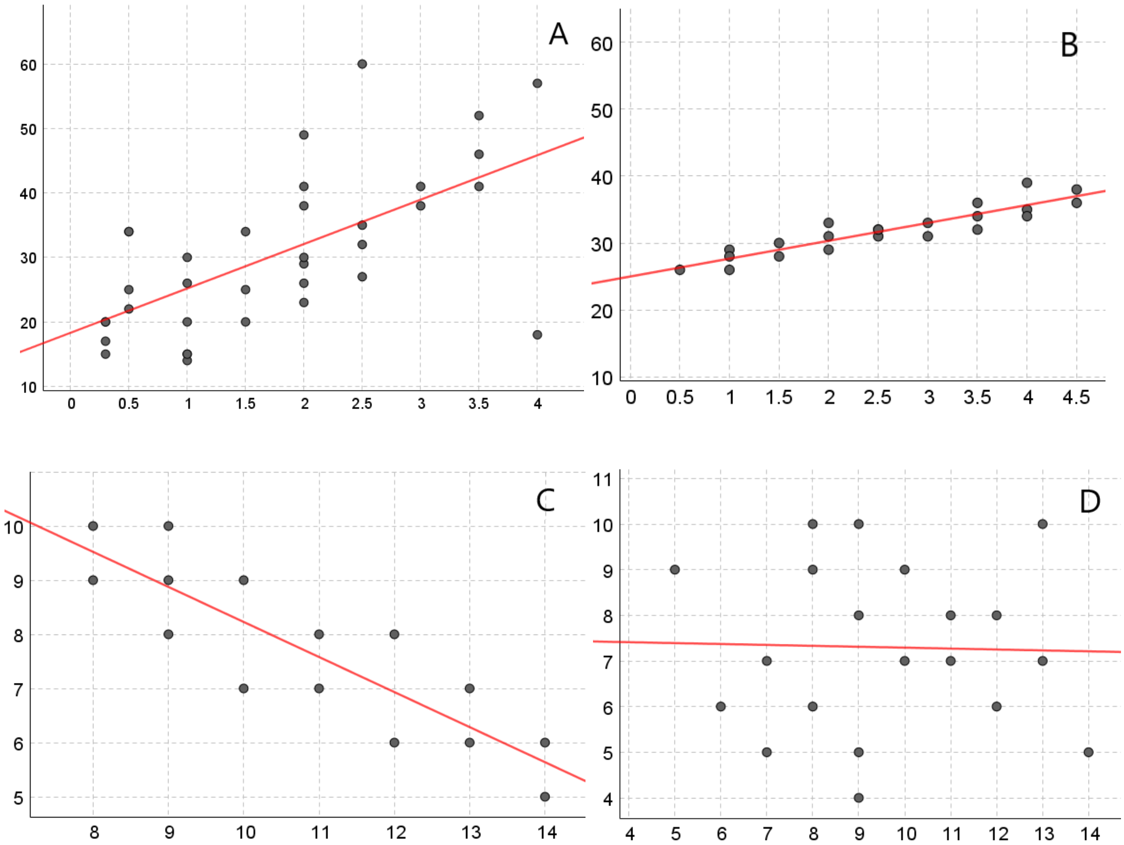
   d) Ennusta molempien regressiomallien avulla, paljonko tilavuudeltaan $170\,\text{cm}^3$ oleva puupala keskimäärin painaisi.





5. Alla on neljään eri hajontakuvioon sovitetut lineaarisen regressiosuorat sekä korrelaatiokertoimen arvoja lueteltuna. Yhdistä korrelaatiokerroin (1-4) sitä vastaavaan hajontakuvioon sovitettuun suoraan (A-D).

1. $r = -0.0268$
2. $r = 0.9254$
3. $r = 0.6324$
4. $r = -0.8665$



6. Pitävätkö seuraavat väittämät aina, joskus vai ei koskaan paikkaansa? Jos vain joskus, niin milloin? Perustele vastauksesi.

   a) Korrelaatiokertoimen arvo on yhtä suuri kuin lineaarisen regressiosuoran kulmakerroin.

   b) Jos kahden muuttujan välillä on negatiivista korrelaatiota, niin tällöin toisen muuttujan arvojen kasvaessa myös toisen muuttuja arvot kasvavat.

   c) Jos korrelaatiokerroin on nolla, kahden muuttujan välillä ei ole lainkaan tilastollista riippuvuutta.

7. Valitse itseäsi kiinnostava asia ja tutki tämän asian riippuvuutta jostakin toisesta asiasta. Tutkimuksesi tulisi sisältää seuraavat vaiheet:

   - Havaintoaineiston kerääminen
   - Valittujen kahden muuttujan jakauman havainnollistaminen
   - Regressioyhtälön määrittäminen
   - Korrelaatiokertoimen määrittäminen

- Johtopäätösten tekeminen

Hyödynnä havaintoaineiston käsittelyssä GeoGebraa ja pohdi johtopäätöksiä tehdessäsi erityisesti, onko valitsemiesi muuttujien välillä tilastollista riippuvuutta sekä millaista ja kuinka voimakasta mahdollinen riippuvuus on.

A.2 Diskreetti todennäköisyysjakauma

Pohdinta A.13 Kun heitetään yhden kerran kahta tetraedrin muotoista nelisivuista noppaa – joissa molemmissa on siis silmäluvut $1, 2, 3$ ja 4 – ja kerrotaan saadut silmäluvut keskenään, todennäköisyys saada tietty silmälukujen tulon arvo noudattaa alla olevaa taulukkoa.

Tulon arvo	Arvon todennäköisyys
1	1/16
2	1/8
3	1/8
4	3/16
6	1/8
8	1/8
9	1/16
12	1/8
16	1/16

Mikä silmälukujen tulon arvo yhdellä kahden nelisivuisen nopan heitolla on todennäköisintä saada?

Toistetaan kahden nelisivuisen nopan heitto 3200 kertaa. Tällöin noppien silmälukujen tulon eri arvoja 3200 heiton joukossa voidaan lukumäärällisesti arvioida saatavan seuraavan taulukon mukaisesti.

Tulon arvo	Arvon lukumäärä 3200 heiton joukossa
1	$1/16 \cdot 3200 = 200$
2	$1/8 \cdot 3200 = 400$
3	400
4	600
6	400
8	400
9	200
12	400
16	200

Mikä on kahden nelisivuisen nopan silmälukujen tulon saatujen arvojen keskiarvo, kun kahden nopan heitto toistetaan 3200 kertaa ja oletetaan, että heittojen arvot jakautuvat edellisen taulukon mukaisesti?

Pohdi myös, miten tämän tuloksen voisi saada suoraan tehtävän alussa annetun noppien silmälukujen tulon arvojen todennäköisyyksien taulukosta ilman, että kiinnitetään heittojen lukumäärää.

Määritelmä A.14 *Satunnaismuuttuja* on muuttuja, jonka arvot määräytyvät jonkin satunnaisilmiön mukaan. Satunnaismuuttujan arvot ovat reaalilukuja. Satunnaismuuttujaa merkitään yleensä isoilla kirjaimilla. *Diskreetti satunnaismuuttuja* voi saada vain tiettyjä erillisiä arvoja $x_1, x_2, \ldots$, kun taas *jatkuva satunnaismuuttuja* voi saada mitä tahansa arvoja joltain reaalilukuväliltä.

Esimerkki A.15 Olkoon X kahden nelisivuisen nopan silmälukujen tulo. Tällöin kahden nelisivuisen nopan heitto samanaikaisesti on satunnaisilmiö, joka määrää satunnaismuuttujan X arvot.

Satunnaismuuttujan X kaikki mahdolliset arvot ovat 1, 2, 3, 4, 6, 8, 9, 12 ja 16. Kyseessä on siis diskreetti satunnaismuuttuja.

Pohdinta A.16 Alla on lueteltuina muutama satunnaismuuttuja sekä niiden arvojoukot. Yhdistä satunnaismuuttujat (1-3) niitä vastaaviin arvojoukkoihin (a-c).

1. X : ”kahden tavallisen arpakuution heitossa saatavien silmälukujen summa”
2. Y : ”kahden tavallisen arpakuution heitossa saatavien silmälukujen erotus, kun suuremmasta silmäluvusta vähennetään aina pienempi”
3. Z : ”tavallisesta korttipakasta kahden kortin nostossa saatavien herttojen lukumäärä”

a) $\{0, 1, 2, 3, 4, 5\}$

b) $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

c) $\{0, 1, 2\}$

Todennäköisyyden jakautumista satunnaismuuttujan arvojen kesken tutkitaan *todennäköisyysjakaumien* avulla. Tässä oppikirjassa keskitytään *diskreettiin todennäköisyysjakaumaan*, joka kuvaa todennäköisyyden jakautumista diskreetin satunnaismuuttujan arvojen kesken.

Todennäköisyyttä sille, että diskreetti satunnaismuuttuja X saa arvon x_k, merkitään $p_k = p(x_k) = P(X = x_k)$. Näitä todennäköisyyksiä sanotaan satunnaismuuttujan X *pistetodennäköisyyksiksi*.

Esimerkki A.17 Pohdinnan A.13 ensimmäinen taulukko on diskreetin satunnaismuuttujan X arvoja ja pistetodennäköisyyksiä kuvaava diskreetti todennäköisyysjakauma, kun X on kahden nelisivuisen nopan silmälukujen tulo.

Pohdinta A.18 Olkoon X kahden nelisivuisen nopan silmälukujen tulo. Tarkastele Pohdinnan A.13 ensimmäistä taulukkoa ja siinä esitettyjä satunnaismuuttujan X pistetodennäköisyyksiä sekä erityisesti näiden pistetodennäköisyyksien summaa. Mitä huomaat? Mistä tämä voisi johtua?

Diskreetin satunnaismuuttujan arvoja vastaavat tapahtumat ovat erillisiä ja täyttävät koko satunnaismuuttujaa määrittävän satunnaisilmiön perusjoukon. Tämän vuoksi diskreetin satunnaismuuttujan kaikkien arvojen pistetodennäköisyyksien summa on aina yksi.

Oppikirjassa aiemmin käsitellyn diskreetin tilastollisen jakauman tapaan myös diskreettiä todennäköisyysjakaumaa voidaan kuvata tunnusluvuilla. Todennäköisyyslaskennassa diskreetin tilastollisen jakauman keskiarvoa vastaa *diskreetin satunnaismuuttujan odotusarvo*.

Määritelmä A.19 Oletetaan, että satunnaismuuttujan X arvot ovat $x_1, x_2, \ldots, x_n$ ja arvojen todennäköisyydet $p_1, p_2, \ldots, p_n$. Tällöin *diskreetin satunnaismuuttujan X odotusarvo* on

$$E(X) = \sum_{i=1}^{n} p_i x_i = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n.$$

Satunnaismuuttujan X odotusarvon merkintä $E(X)$ tulee englannin kielen sanoista *expected value* eli odotettu arvo. Satunnaismuuttujan odotusarvolle käytetään myös merkintää μ.

Diskreetin tilastollisen jakauman tapaan myös diskreetin todennäköisyysjakauman kohdalla ollaan kiinnostuneita jakauman hajonnasta. *Diskreetin satunnaismuuttujan keskihajonnalla* voidaan kuvata satunnaismuuttujan arvojen vaihtelun suuruutta satunnaismuuttujan odotusarvosta.

Määritelmä A.20 Oletetaan, että satunnaismuuttujan X arvot ovat $x_1, x_2, \ldots, x_n$, arvojen todennäköisyydet $p_1, p_2, \ldots, p_n$ ja odotusarvo μ. Tällöin *diskreetin satunnaismuuttujan X keskihajonta* on

$$D(X) = \sqrt{\sum_{i=1}^{n} p_i(x_i - \mu)^2} = \sqrt{p_1(x_1 - \mu)^2 + \ldots + p_n(x_n - \mu)^2}.$$

Satunnaismuuttujan X keskihajonnan merkintä $D(X)$ tulee englannin kielen sanoista *standard deviation* eli keskihajonta. Satunnaismuuttujan keskihajonnalle käytetään myös merkintää σ.

Yleensä diskreettejä todennäköisyysjakaumia havainnollistetaan taulukoiden lisäksi myös muun muassa pylväsdiagrammien avulla. Seuraavassa pohdinnassa harjoitellaan, miten tähän ja diskreetin satunnaismuuttujan tunnuslukujen määrittämiseen voi hyödyntää GeoGebraa.

Pohdinta A.21 Kopioi Pohdinnan A.13 ensimmäisen taulukon arvot GeoGebran "Taulukkolaskenta" -ohjelmistoon kahteen vierekkäiseen sarakkeeseen siten, että vasemman puoleisessa sarakkeessa on satunnaismuuttujan X : "kahden nelisivuisen nopan silmälukujen tulo" arvot ja oikeanpuoleisessa sarakkeessa satunnaismuuttujan X arvoja vastaavat pistetodennäköisyydet. Tämän jälkeen harjoittele seuraavien ohjeiden avulla, miten kyseisen diskreetin todennäköisyysjakauman saa esitettyä GeoGebralla taulukon lisäksi myös pylväsdiagramminna.

- Maalaa molempien muuttujien arvot ja valitse vasemmassa yläkulmassa olevasta painikkeesta  aukeavasta valikosta "Yhden muuttujan analyysi". Puhelimella tai tabletilla valitse aukeavasta valikosta .
- Pylväsdiagrammin asetukset voit muokata mieluisiksi diagrammin yläpuolella olevasta painikkeesta  aukeavasta "Kuvaaja" -valikosta ("Automaattinen mittasuhde" -kohdan ☑ tulee ottaa pois ennen asetusten muokkaamista).

Tutki lopuksi pylväsdiagrammin yläpuolella olevasta painikkeesta 

aukeavasta "Näytä tilastot" -osiosta, miten saat GeoGebralla määritettyä diskreetin satunnaismuuttujan X odotusarvon ja keskihajonnan.

Seuraavaa pohdintaa varten palauta mieleesi oppikirjassa aiemmin käsitelty binomikerroin sekä riippumattomien tapahtumien kertolaskusääntö ja erillisten tapahtumien yhteenlaskusääntö.

Pohdinta A.22 Monivalintakokeessa on kuusi kysymystä ja jokaisessa kysymyksessä on kolme vastausvaihtoehtoa, joista vain yksi on oikein.

a) Tarkastellaan aluksi vain yhtä monivalintakokeen kysymystä. Millä todennäköisyydellä arvaamalla saa kysymyksen oikein? Entä millä todennäköisyydellä arvaamalla saa kysymyksen väärin?

b) Tarkastellaan sitten koko monivalintakoetta. Millä todennäköisyydellä arvaamalla saa monivalintakokeesta täsmälleen ensimmäisen kysymyksen oikein ja viisi muuta kysymystä väärin?

c) Millä todennäköisyydellä arvaamalla saa monivalintakokeesta täsmälleen kaksi ensimmäistä kysymystä oikein ja neljä muuta kysymystä väärin?

d) Kuinka monella eri tavalla monivalintakokeessa voi saada kaikista kuudesta kysymyksestä täsmälleen kaksi oikein?

e) Millä todennäköisyydellä arvaamalla saa monivalintakokeesta täsmälleen kaksi kysymystä oikein?

Kun samaa satunnaiskoetta toistetaan useamman kerran ja seurataan jokaisella toistolla, tapahtuuko tapahtuma A, on kyseessä *toistokoe*, jos lisäksi tapahtuman A todennäköisyys ei riipu edeltävien toistojen tuloksista.

Kun toistokoe toistetaan n kertaa, tasan r onnistumista voidaan saada $\binom{n}{r}$ eri tavalla. Koska eri tavat saada n:ssä toistossa tasan r onnistumista ovat erillisiä tapahtumia ja tapahtuman A toteutuminen yhdellä toistokerralla ei riipu edeltävien toistojen tuloksista, voidaan toistokokeen todennäköisyys erillisten tapahtumien yhteenlaskusäännön ja riippumattomien tapahtumien kertolaskusäännön nojalla laskea seuraavasti.

Määritelmä A.23 Oletetaan, että tapahtuma A tapahtuu yksittäisessä satunnaiskokeessa todennäköisyydellä p. Toistetaan satunnaiskoetta n kertaa. Tällöin

$$P(''A \text{ tapahtuu } n \text{ toistossa tasan } r \text{ kertaa}'') = \binom{n}{r} p^r (1-p)^{n-r}.$$

Toistokokeen todennäköisyyttä sanotaan myös *binomitodennäköisyydeksi*, sillä sen laskemisessa hyödynnetään binomikerrointa. Käydään läpi seuraavaksi tähän liittyen *binomijakauma*.

Olkoon X tapahtuman A toteutumisten lukumäärä n:nnän toiston toistokokeessa. Olkoon lisäksi tapahtuman A toteutumisen todennäköisyys yksittäisessä satunnaiskokeessa p. Tällöin satunnaismuuttujaa X vastaavat pistetodennäköisyydet voidaan laskea Määritelmän A.23 kaavalla ja sanotaan, että satunnaismuuttuja X noudattaa *binomijakaumaa* parametrein n ja p. Tätä merkitään $X \sim Bin(n, p)$.

Määritelmä A.24 Oletetaan, että satunnaismuuttuja $X \sim Bin(n, p)$. Tällöin satunnaismuuttujan X odotusarvo on

$$E(X) = np$$

ja keskihajonta

$$D(X) = \sqrt{np(1-p)}.$$

Seuraavassa pohdinnassa harjoitellaan GeoGebran hyödyntämistä binomitodennäköisyyksien laskemisessa sekä binomijakauman havainnollistamisessa ja binomijakaumaa noudattavan satunnaismuuttujan tunnuslukujen määrittämisessä.

Pohdinta A.25 Olkoon X parittomien silmälukujen lukumäärä, kun tavallista arpakuutiota heitetään neljä kertaa.

a) Mikä tässä toistokokeessa on toivottu tapahtuma A?

b) Mikä tapahtuman A toteutumisen todennäköisyys on yksittäisessä satunnaiskokeessa? Entä kuinka monta kertaa satunnaiskoe toistetaan?

Seuraavasta linkistä pääset GeoGebran ”Todennäköisyyslaskuri” -ohjelmistoon. `https://www.geogebra.org/classic#probability`

Avaa ohjelmisto ja etsi jakaumavalikosta binomijakauma. Syötä sitten GeoGebraan b-kohdassa määrittämäsi binomijakauman parametrit ja vastaa seuraaviin kysymyksiin ohjelmiston näytölle ilmaantuvien tietojen avulla.

c) Määritä satunnaismuuttujan X todennäköisyysjakauma niin taulukkona kuin pylväsdiagrammina.

d) Määritä satunnaismuuttujan X odotusarvo ja keskihajonta.

e) Määritä todennäköisyys tapahtumalle B = ”pariton silmäluku saadaan tasan kerran, kun tavallista arpakuutiota heitetään neljä kertaa”.

f) Tutki, miten saisit ohjelmiston vasemmassa alalaidassa olevien painikkeiden  avulla seuraavien tapahtumien C ja D todennäköisyydet selville. C = ”pariton silmäluku saadaan korkeintaan kaksi kertaa, kun tavallista arpakuutiota heitetään neljä kertaa” ja D = ”pariton silmäluku saadaan vähintään kerran, kun tavallista arpakuutiota heitetään neljä kertaa”.

Harjoitustehtävät

8. Alla on lueteltuna muutama satunnaismuuttuja. Mitkä annetuista satunnaismuuttujista ovat diskreettejä? Perustele vastauksesi.

   a) X : "parittomien silmälukujen lukumäärä, kun tavallista arpakuutiota heitetään kuusi kertaa".

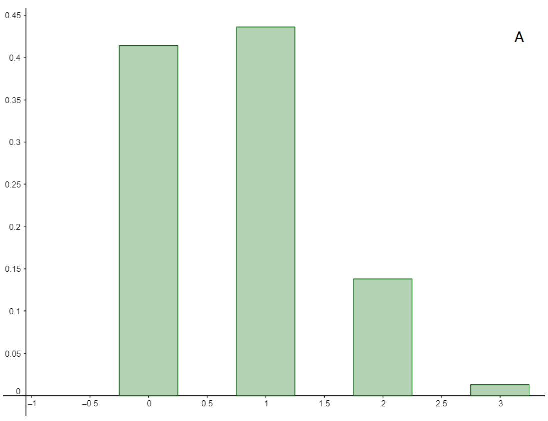
   b) Y : "heittoalustalle pudonneiden tavallisten arpakuutioiden etäisyys, kun kaksi arpakuutiota heitetään alustalle samanaikaisesti yhden kerran".

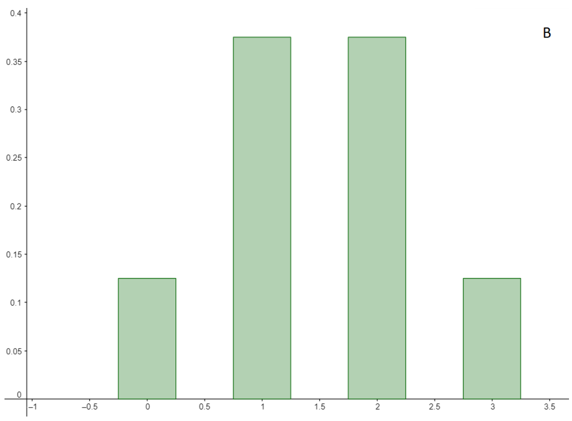
   c) Z : "silmälukujen summan neliö, kun kahta nelisivuista noppaa heitetään samanaikaisesti yhden kerran".

9. Alla on kahden eri diskreetin satunnaismuuttujan todennäköisyysjakaumat sekaisin sekä taulukkoina (1-2) että pylväsdiagrammeina (A-B). Yhdistä saman diskreetin todennäköisyysjakauman eri esitystavat toisiinsa. Perustele vastauksesi.

1. X	P
0	0.125
1	0.375
2	0.375
3	0.125

2. X	P
0	0.414
1	0.436
2	0.138
3	0.013





10. Tavallisesta korttipakasta nostetaan neljä korttia. Olkoon satunnaismuuttujaa X neljän kortin nostossa saatavien ruutujen lukumäärä. Määritä satunnaismuuttujan X todennäköisyysjakauma ja esitä se sekä taulukon että pylväsdiagrammin avulla. Apuna voit käyttää GeoGebraa. Määritä myös satunnaismuuttujan X odotusarvo ja keskihajonta.

11. Pelissä heitetään kahta tavallista arpakuutiota. Pelin panoksena on kaksi pelimerkkiä. Pelin säännöt ovat seuraavat:

   - Jos arpakuutioiden silmälukujen summa on vähintään 10, saat 10 pelimerkkiä.

- Jos arpakuutioiden silmälukujen summa on alle 10 ja parillinen, saat 7 pelimerkkiä.
- Jos arpakuutioiden silmälukujen summa on alle 10 ja pariton, menetät panoksesi.

Huomaa, että voitto tarkoittaa pelistä saatujen pelimerkkien määrän ja peliin panoksena sijoitettujen pelimerkkien määrän erotusta.

a) Laske pelaajan voiton odotusarvo yhdellä pelikerralla.
b) Arvioi yhden pelimerkin tarkkuudella pelaajan kokonaisvoitto, kun pelaaja pelaa peliä kymmenen kertaa.
c) Kannattaako peliä mielestäsi pelata? Perustele vastauksesi.

12. Olkoon X arpojen sisältämien voittojen lukumäärä. Arpajaisissa ilmoitetaan, että joka 15:sta arpa voittaa ja arpoja ostetaan 10 kappaletta. Määritä satunnaismuuttujan X todennäköisyysjakauma haluamassasi muodossa. Apuna voit käyttää myös GeoGebraa. Vastaa lopuksi seuraaviin kysymyksiin.

a) Mikä on satunnaismuuttujan odotusarvo? Entä varianssi?
b) Millä todennäköisyydellä tasan kahdessa arvassa on voitto?
c) Millä todennäköisuudellä korkeintaan kahdessa arvassa on voitto?
d) Millä todennäköisyydellä ainakin kahdessa arvassa on voitto?

B Opettajan opas

B.1 Ajankäyttösuunnitelma

Tämä oppimateriaali on tarkoitettu lukion pitkän matematiikan moduulille *Tilastot ja todennäköisyys* (MAA8). Oppimateriaali on vain osa koko moduulille tarkoitettua oppikirjaa ja se sisältää *Regression* sekä *Diskreetin todennäköisyysjakauman* aihepiirien osat oppikirjasta. Regression osuus on tarkoitettu käsiteltäväksi moduulin alkupuolen eli tilastoja käsittelevän osuuden päätteeksi ja diskreetin todennäköisyysjakauman osuus moduulin loppupuolen eli todennäköisyyttä käsittelevän osuuden päätteeksi.

Oppimateriaali on suunniteltu siten, että kokonaisuudessaan sen läpikäymiseen käytettäisiin kuusi 45 minuutin pituista tai neljä 75 minuutin pituista oppituntia. 45 minuuttisille ja 75 minuuttisille oppitunneille voisi käyttää esimerkiksi seuraavanlaista tuntijakoa.

	75 min	45 min
Regressio	2	3
Diskreetti todennäköisyysjakauma	2	3

B.2 Pohdintatehtävät

Tässä osiossa kerrotaan tarkemmin, mitä oppimistavoitteita oppimateriaalin sisältämillä pohdintatehtävillä pyritään saavuttamaan sekä annetaan vinkkejä opettajalle, kuinka opiskelijoita voi auttaa pääsemään näihin tavoitteisiin esimerkiksi johdattelevia kysymyksiä hyödyntämällä. Tässä osiossa annetaan myös eriyttämiseen liittyviä neuvoja, kuten minkälaisia vihjeitä tehtäviin voi opiskelijoille tarvittaessa antaa tehtävässä eteenpäin pääsemistä varten tai miten tehtäviä voi viedä vielä pidemmälle. Lisäksi tässä osiossa annetaan vastaukset pohdintatehtäviin, ja pohdintatehtävät olisikin hyvä käydä aina myös yhteisesti läpi, ettei opiskelijoille pääse jäämään asioista vääriä käsityksiä.

B.2.1 Regressio

Pohdinta A.1

Tämän pohdintatehtävän tarkoituksena on johdatella opiskelijoita riippuvuuksien aihepiiriin antamalla opiskelijoille kolme esimerkkiä tilanteista, joissa muuttujien välisen riippuvuuden olemassaolosta voidaan suoraan todeta jotain.

Jos opiskelijoilla on vaikeuksia hahmottaa, mitä A riippuu B:stä -väittämällä tarkoitetaan, heitä voi ohjeistaa miettimään, muuttuuko A, jos B:tä muutetaan.

Vastaukset:

1. Aina

2. Ei koskaan

3. Aina

Pohdinta A.4

Tämän pohdintatehtävän tarkoituksena on tukea opiskelijoiden selitettävän ja selittävän muuttujan käsitteiden oppimista antamalla opiskelijoille tilaisuus heti soveltaa oppimateriaalissa heille annettuja uusia tietojaan näihin käsitteisiin liittyen.

Lisäpohdintana tähän tehtävään opiskelijoita voi pyytää miettimään myös, mitkä ovat annettujen tutkimusten selittävät muuttujat. Tehtävää voi myös eriyttää ylöspäin poistamalla tehtävästä muuttujien listan kokonaan, jolloin opiskelijat joutuvat itse myös ensin miettimään, mitkä kaksi muuttujaa eri tutkimuskysymyksissä ovat tutkimuksen kohteina.

Vastaukset:

1e

2b

3f

Pohdinta A.5

Tämän pohdintatehtävän tarkoituksena on johdatella opiskelijoita regressioanalyysin ja lineaarisen riippuvuuden pariin antamalla opiskelijoiden tarkastella erilaisia hajontakuvioita ja miettiä, voisiko niitä mallintaa suoralla.

Lisäpohdintana tähän tehtävään opiskelijoita voi pyytää miettimään, millaisella matemaattisella mallilla niitä hajontakuvioita voitaisiin mallintaa, joihin suora ei sovi malliksi.

Vastaukset:

Vasenta yläkulmaan voidaan mallintaa toisen asteen polynomilla.

Oikeaa yläkulmaa ei voida mallintaa millään matemaattisella mallilla.

Alarivin hajontakuvioita voidaan mallintaa suoralla.

Pohdinta A.6

Tämän pohdintatehtävän tarkoituksena on johdatella opiskelijat lineaarisen regressiosuoran ominaisuuksiin antamalla opiskelijoiden vertailla kahta samaan hajontakuvioon eri tavalla sovitettua suoraa. Opiskelijoiden tulisi päättää ja perustella, kumpi sovitetuista suorista kuvaa havaintopisteitä paremmin. Lisäksi tämän pohdintatehtävän tarkoituksena on tarjota tilaisuus korjata niiltä opiskelijoilta harhakäsitys siitä, että suoran on tärkeämpää kulkea useamman havaintopisteen kautta kuin mahdollisimman läheltä kaikkia havaintopisteitä, jotka ovat sitä mieltä, että oikeanpuoleinen suora kuvaa pistepilveä paremmin.

Vastaukset:

Vasemmanpuoleinen suora kuvaa havaintopisteitä paremmin.

Pohdinta A.7

Tämän pohdintatehtävän tarkoituksena on antaa opiskelijoille tilaisuus harjoitella GeoGebran käyttöä hajontakuvion tekemiseen sekä regressiosuoran sovittamiseen ja regressioyhtälön määrittämiseen. Lisäksi pohdintatehtävän lopussa olevan kysymyksen tarkoituksena on ohjata opiskelijoita huomaamaan, että regressiomallin avulla voidaan ennustaa selitettävän muuttujan käyttäytymistä selittävän muuttujan arvojen vaihdellessa.

Vastaukset:

1. $y = -0.7933x + 11.1352$
2. Yöunien voidaan arvioida jäävän keskimäärin neljän tunnin pituisiksi.

Pohdinta A.8

Tämän pohdintatehtävän tarkoituksena on tutustuttaa opiskelijat korrelaatiokertoimen ominaisuuksiin antamalla heidän tutkia eri suuruisia korrelaatiokertoimia sekä niihin liittyviä hajontakuvioita ja regressiosuoria.

Vastaukset:

1. Korrelaatiokerroin saa arvot -0.0268, 0.9915 ja -0.8297.
2. Korrelaatiokerroin on positiivinen silloin, kun regressiosuora on nouseva ja negatiivinen silloin, kun regressiosuora on laskeva.
3. Korrelaatiokertoimen itseisarvo on suurin silloin, kun havaintopisteet ovat lähimpänä regressiosuoraa ja pienin silloin, kun havaintopisteet ovat kauimpana regressiosuorasta.

Pohdinta A.10

Tämän pohdintatehtävän tarkoituksena on saada opiskelijat huomaamaan, että kahden muuttujan välistä lineaarista riippuvuutta tutkittaessa pelkkä vahva korrelaatio ei vielä takaa lineaarisen riippuuvuuden olemassaoloa. Lisäksi tämän pohdintatehtävän tarkoituksena on havainnollistaa myös poikkeavan havainnon vaikutusta korrelaatiokertoimeen ja regressiosuoraan.

Vastaukset:

1. Lineaarinen regressiosuora ei kuvaa jokaista hajontakuviota yhtä hyvin.
2. Molemmissa hajontakuvioissa yksi havaintopiste poikkeaa selkeästi muista. Regressiosuora muuttuisi, jos tämä havaintopiste poistettaisiin havaintopisteistä.

B.2.2 Diskreetti todennäköisyysjakauma

Pohdinta A.13

Tämän pohdintatehtävän tarkoituksena on johdatella opiskelijoita diskreetin satunnaismuuttujan, diskreetin todennäköisyysjakauman ja diskreetin todennäköisyysjakauman odotusarvon käsitteisiin sekä odotusarvon tulkintaan.

Tehtävää voi eriyttää ylöspäin antamalla ensimmäisessä taulukossa vain esimerkiksi kahden ensimmäisen tulon arvon todennäköisyydet ja pyytämällä opiskelijoita ensin täydentämään taulukon itse loppuun. Tällöin opiskelijat voivat täydentää itse myös jälkimmäisen taulukon loppuosan. Apuna todennäköisyyksien laskemisessa opiskelijoita kannattaa ohjeistaa käyttämään 4×4 -ruudukkoa suotuisten alkeistapausten ja perusjoukon määrittämisessä.

Vastaukset:

Yhdellä kahden nelisivuisen nopan heitolla on todennäköisintä saada silmälukujen tuloksi 4.

3200 heiton sarjassa taulukon arvion mukaan kahden nelisivuisen nopan silmälukujen tulon arvojen keskiarvo on 6,25.

Saman arvon voi määrittää ensimmäisestä taulukosta suoraan käyttämällä tulon arvojen todennäköisyyksiä niiden frekvensseinä ja ykköstä havaintojen kokonaislukumääränä.

Pohdinta A.16

Tämän pohdintatehtävän tarkoituksena on syventää opiskelijoiden ymmärrystä satunnaismuuttujan käsitteestä sekä siitä, millaisia arvoja eri satunnaismuuttujat voivat saada antamalla opiskelijoille tilaisuuden soveltaa uusia tietojaan heti niiden omaksumisen jälkeen.

Vastaukset:

1b

2a

3c

Pohdinta A.18

Tämän pohdintatehtävän tarkoituksena on saada opiskelijat huomaamaan, että diskreetin satunnaismuuttujan arvoja vastaavien pistetodennäköisyyksien summa on aina yksi. Lisäksi tehtävän tarkoituksena on ohjata opiskelijat miettimään, mistä tämä voisi johtua.

Vastaukset:

Oppimateriaalin teksti heti pohdintatehtävän jälkeen vastaa tämän pohdintatehtävän kysymyksiin.

Pohdinta A.21

Tämän pohdintatehtävän tarkoituksena on antaa opiskelijoille tilaisuus harjoitella GeoGebran käyttöä diskreetin todennäköisyysjakauman havainnollistamisessa sekä diskreetin satunnaismuuttujan tunnuslukujen määrittämisessä.

Vastaukset:

Satunnaismuuttujan X odotusarvo on 6,25 ja keskihajonta 4,1458.

Pohdinta A.22

Tämän pohdintatehtävän tarkoituksena on vaiheittain johdatella opiskelijoita toistokokeen todennäköisyyden laskemiseen. Lisäksi tehtävän eri vaiheiden tavoitteena on auttaa opettajaa huomaamaan, jos opiskelijat tarvitsevat vielä lisää tukea tehtävän eri vaiheissa sovellettavien asioiden oppimiseen.

Vastaukset:

a) Kysymyksen saa arvaamalla oikein todennäköisyydellä $\frac{1}{3}$ ja väärin todennäköisyydellä $\frac{2}{3}$.

b) 4,39% todennäköisyydellä.

c) 2,19% todennäköisyydellä.

d) 15.

e) 32,93% todennäköisyydellä.

Pohdinta A.25

Tämän pohdintatehtävän tarkoituksena on antaa opiskelijoille tilaisuus harjoitella GeoGebran käyttöä binomijakauman havainnollistamisessa sekä binomijakaumaa noudattavan satunnaismuuttujan tunnuslukujen määrittämisessä ja erilaisten binomitodennäköisyyksien laskemisessa. Lisäksi tehtävän eri vaiheiden tavoitteena on auttaa opettajaa huomaamaan, jos opiskelijat tarvitsevat lisää tukea tehtävän eri vaiheissa sovellettavien asioiden oppimiseen.

Vastaukset:

a) Tapahtuma A on "tavallisen arpakuution heitossa saadaan pariton silmäluku".

b) $p = P(A) = \frac{1}{2}$ ja $n = 4$.

c) Pylväsdiagrammi GeoGebran "Todennäköisyyslaskurissa" löytyy keskeltä näyttöä ja taulukko näytön oikeasta ylänurkasta.

d) $\mu = 2$ ja $\sigma = 1$.

e) $P(B) = 0.25$.

f) $P(C) = 0.6875$ ja $P(D) = 0.9375$.

C Harjoitustehtävien vastaukset

1. a) Selitettäväksi muuttujaksi kannattaa valita taloudessa vuorokauden aikana kulutettu sähkömäärä ja selittäväksi muuttujaksi taloudessa asuvien henkilöiden määrä.
   b) Selitettäväksi muuttujaksi kannattaa valita auton nastarenkaista irronneiden nastojen lukumäärä ja selittäväksi muuttujaksi auton nastarenkailla ajettu matka.
   c) Selitettävän muuttujan ja selittävän muuttujan voi valita kummin päin tahansa.

2. a) Selitettävä muuttuja on alkoholijuomien kulutus ja selittävä muuttuja on vuodenaika.
   b) Selitettävä muuttuja on matematiikan kurssin arvosana ja selittävä muuttuja on poissaolojen lukumäärä.
   c) Selitettävä muuttuja on tuotteen kysyntä ja selittävä muuttuja on tuotteen hinta.

3. 1c ja 2a.

4. a) Puupalan massan ja tilavuuden välillä on lineaarinen riippuvuus.
   b) Alina ja Matti ovat saaneet tuloksiksi erilaiset regressiosuorat ja erisuuruiset korrelaatiokertoimet. Ero johtuu poikkeavasta havainnosta.
   c) Matin tekemän regressioanalyysin tuottama regressiomalli kuvaa paremmin puupalojen massan riippuvuutta puupalojen tilavuudesta, jos poikkeava havainto voidaan todeta virheelliseksi mittaustulokseksi. Alinan regressiomalli taas kuvaa riippuvuutta paremmin, jos poikkeavan havainnon kohdalla ei ole kyse mittausvirheestä ja jos tätä havaintopistettä vastaavan puupalan massan poikkeamiselle yleisestä trendistä ei löydy mitään muutakaan syytä, jonka perusteella havaintopisteen voisi poistaa havaintoaineistosta.
   d) Alinan regressiomallin avulla ennustettuna puupala painaisi keskimäärin $55,85\,\text{g}$ ja Matin regressiomallin avulla ennustettuna keskimäärin $70,48\,\text{g}$.

5. 1D, 2B, 3A ja 4C.

6. a) Joskus.
   b) Ei koskaan.
   c) Joskus.

7. -

8. Satunnaismuuttujat X ja Z.

9. 1B ja 2A.

10. $\mu = 1$ ja $\sigma = 0.8402$.

11. a) 2,38 pelimerkkiä.
    b) Pelaaja on voittanut noin 24 pelimerkkiä.
    c) Peliä kannattaa pelata.

12. a) $\mu = 0.6667$ ja $\sigma = 0.7888$.
    b) 11,52% todennäköisyydellä.
    c) 97,51% todennäköisyydellä.
    d) 14,01% todennäköisyydellä.