

Voiko turkulaisten kirjoittamista artikkeleista yli 100 % olla kvantitatiivisia?

Vetosuhteen (odds ratio) ja vedon (odds) tulkintaa¹

Hannu Rita: FT, lehtori, Metsävarojen käytön laitos, Helsingin yliopisto; erityisasiantuntija, Suomen ympäristökeskus (SYKE)

Pertti Töttö: YTT, professori, Sosiaalityön ja sosiaalipedagogiikan laitos, Kuopion yliopisto

Marja Alastalo: YTT, tutkijatohjuri, Sosiologian ja sosiaalipsykologian laitos, Tampereen yliopisto
hannu.rita@helsinki.fi, totto@messi.uku.fi, marja.alastalo@uta.fi

Janus vol. 16 (1) 2008, 72-80

Jani Erola ja Pekka Räsänen kirjoittavat suomalaisen sosiaalityönteiden menetelmätrendeistä (Janus 4/2007). He käyttävät empiirisiä tuloksia tulkitessaan odds ratio – suuretta, joka heidän mukaansa "[t]arkkaan ottaen -- on kahden mahdollisen toteutuman välinen suhde" (alaviite 9). Tieteellisten artikkeleiden alaviitteet ovat usein vastauksia käsikirjoituksen arvioitsijoille. Alaviite 9 vaikuttaa reaktiolta arvioitsijan (PT) seuraavaan huomautukseen:

"Isompi juttu: sivuilla 14 – 15 esiintyy useamman kerran ilmaus, että jollakin on x-kertainen todennäköisyys johonkin toiseen verrattuna. Tässä viitataan taulukon 3 vedonlyöntisuhteisiin. Vedonlyöntisuhde (odds ratio) ei kuitenkaan ole todennäköisyyksien suhde vaan kahden vedon (oddsin) suhde eli $(p/1-p)/(p'/1-p')$, missä p ja p' ovat todennäköisyyksiä."

Koska kirjoittajat eivät näytä ymmärtäneen, mitä arvioitsija tarkoitti, esitämme seuraavassa saman huomautuksen perusteellisemmin. Ensiksi kuitenkin lyhin versio: Erolan ja Räsänen tulkinta odds ratiolle on virheellinen.

Erola ja Räsänen käyttävät odds ratiolle (hyvää) suomennosta vedonlyöntisuhde. Hannu Rita (2003 ja 2004) on ehdottanut lyhyempää käännöstä 'vetosuhte', josta termille odds seuraava luonteva käännös, 'veto'. Käytämme seuraavassa näitä käännöksiä ja suosittelemme samaa muillekin. Näin vältetään siltä terminologiselta ilotulitukselta, joka tuoreen menetelmäoppaan (Jokivuori & Hietala 2007, 70 – 73) sivuilta räis-

kyy: riski, suhteellinen riski, riskiluku, riskitaso ja riskikerroin. Paitsi että riskin sekoittaminen odds ratiioon on riskialtista, on vaikea erottaa, milloin puhutaan oddsista, milloin odds ratiosista. Ne ovat eri käsitteitä ja tarvitsevat kumpikin oman käännöksensä.

Osoitamme ensiksi, että Erolan ja Räsänen tulkinnasta seuraa mahdottomia. Esitämme sen jälkeen muutaman muun samaa virhetulkintaa toistavan esimerkin, jotta kävisi selväksi, että huomautukseemme on laajempaa tarvetta. Sittem käymme läpi osuuksien kuvaamisen peruskäsitteitä, joita ovat veto, vetosuhte sekä riski ja riskisuhde. Lopuksi teemme karun päätelmän: osuuksien vertailun tulokset on vain opittava lausumaan vetojen avulla, niitä ei voi sellaiseenaan palauttaa todennäköisyyksiin.

Tässä oikaistaan...

Erola ja Räsänen tulkitsevat (s. 309):

"Näillä [turkulaisten kirjoittamilla] artikkeleilla on yli kahdeksankertainen todennäköisyys olla kvantitatiivinen tai sekä kvantitatiivinen suhteessa muihin artikkelityyppeihin."

Lausuman takana on Erolan ja Räsänen suorittama Januksessa ja Sosiologiassa julkaistujen artikkelien jako kahteen ryhmään. Ensimmäiseen kuuluvat paitsi puhtaasti kvantitatiiviset myös sekä kvantitatiivista että kvalitatiivista otetta

käyttäneet artikkelit, toiseen ryhmään kuuluvat kaikki muut, mm. teoreettiset tai vain kvalitatiivista otetta käyttäneet artikkelit. Kutsumme jatkossa ensimmäisen ryhmän artikkeleita lyhyesti kvantitatiivisiksi. Erolan ja Räsäsen tavoitteena on selvittää kvantitatiivisten artikkelien osuuksien eroa turkulaisten ja muiden kuin turkulaisten kirjoittamien artikkelien joukossa.

Tulkinnassa esiintyvä määre 'yli kahdeksankertainen' on peräisin Erolan ja Räsäsen Taulukossa 3 (s. 309) esiintyvistä odds ration lukuarvosta 8.2. Sille esitetty tulkinta todennäköisyyksien suhteena on kuitenkin virheellinen: kyseessä ei ole todennäköisyyksien, vaan vetojen suhde.

... ja täällä ...

Helsingin Sanomat julkaisi 17.9.2007 pääuutisena jutun Osmo Kivisen tutkijaryhmän tuoreen artikkelin pohjalta. Uutisessa kerrottiin, miten akateemisten vanhempien lapset valtaavat yhä yliopistopaikat. Tekstin mukaan heillä on kahdeksankertainen "todennäköisyys", "mahdollisuus" tai "suhteellinen mahdollisuus" päätyä yliopistoon verrattuna ei-akateemisten vanhempien lapsiin.

Luku kahdeksan on pyöristetty Acta Sociologica julkaisun artikkelin Taulukosta I (Kivinen, Hedman & Kaipainen 2007, 238). Kyseessä on odds ratio (vetosuhte), jolle on taulukossa kirjattu arvoksi 8.2. Käyttämällä taulukossa annettuja osuuksia (60 % akateemisten ja 15 % ei-akateemisten lapsilla) vetosuhteeksi tulee 8.5; jos käytetään annettuja frekvenssejä saadaan 8.4. Käytämme jatkossa viimeksi mainittua arvoa, koska siihen pyöristysvirheet eivät ole vaikuttaneet. Kirjoittajat sanovat tulkitsevansa laskemansa vetosuhteen siten, että se kertoo kuinka moninkertaiset ovat akateemisten vanhempien lasten "mahdollisuudet" (chances tai opportunities) päätyä yliopistoon ei-akateemisten vanhempien lasten mahdollisuuksiin verrat-

tuna (Kivinen, Hedman & Kaipainen 2007, 239 ja 245).

Artikkelin Taulukosta I voi laskea, että ei-akateemisten vanhempien lapsista yliopistoon päätyi 15.3 prosenttia. Jos akateemisten lapset ovat 8.4-kertaisesti onnekkaampia, heistä pääsee yliopistoon 128,5 prosenttia! Odds ration tulkitseminen todennäköisyyksien suhteeksi on siten ilmeisen virheellistä.

... ja vielä täälläkin

Tieteellisistä julkaisuista löytyisi lisää vastaavia esimerkkejä, mutta löytyy niitä oppikirjoistakin: samaa lipsahdusta opetetaan tuleville tutkijapolville. Monessa suhteessa erinomainen oppikirja (Nummenmaa 2007) opastaa SPSS-ohjelmiston logistisen regression tulosteiden tulkinnassa näin (s. 327):

"Jos esimerkiksi äidin ikä –muuttujan $Exp(B)$ -kerroin on 1.05, niin jokainen äidin ikävuosi kasvattaa 5 %lla turvallisen kiintymyksen todennäköisyyttä."

Mainittu $Exp(B)$ kuvaa kvantitatiivisen selittävän muuttujan (äidin ikä, vuosissa) vaikutuksen suuruutta lapsen turvallisen kiintymyksen kehittymisen todennäköisyyteen. Tässäkin on kyseessä vetosuhte, jota ei voi suoraan tulkita todennäköisyyksien suhteena.

Näyttää siltä, että on syytä palata perusteisiin ja katsoa, millaiseen tulkintaan vetosuhteen määrittämisestä lähtien päädytään. Teemme tarkastelun Erolan ja Räsäsen kvantitatiivisten artikkelien osuuksien vertailun kontekstissa.

Veto (odds) ja vetosuhte (odds ratio)

Olkoon p_T todennäköisyys sille, että turkulaisten (alaindeksi T viittaa siis Turkuun) kirjoittama

artikkeli päättyy Erolan ja Räsänen luokituksessa tyyppiin 'kvantitatiiviset tai sekä kvanti- että kvalitatiiviset artikkelit'. Vastaavalle todennäköisyydelle muiden kuin turkulaisten kirjoittamien artikkelien kohdalla käytettäkään merkintää p_M (alaindeksi M viittaa muihin kuin turkulaisiin). Erolan ja Räsänen ylle kirjattu vetosuhteen tulkinta näyttää näillä merkinnöillä tällaiselta:

$$p_T = 8.2 \times p_M$$

Oletetaanpa, että muiden kuin turkulaisten kirjoittamista artikkeleista 25 % luokiteltiin tyyppiin 'kvantitatiiviset tai sekä kvanti- että kvalitatiiviset artikkelit'. Tämän mukaan p_M on 25 %. Jos tämä todennäköisyys kerrotaan 8.2:lla, saadaan p_T :lle arvio, joka ylittää 200 %! Kvantitatiivinen ote näyttää todella lyöneen läpi Turussa! Itse asiassa se näyttää lyöneen hiukan ylikin, sillä osuudet ovat lukuja, jotka ovat 0 %:n ja 100 % välissä. Asianmukainen tulkinta vetosuhteelle (odds ratio) ei tietenkään koskaan saa johtaa osuuksiin, jotka ovat suurempia kuin 100 %.

Veto (odds)

Suhteellisesta osuudesta tai todennäköisyydestä p saadaan veto (odds) $v(p)$ kaavalla

$$v(p) = p / (1 - p) \quad (1)$$

Osuutta 50 % vastaa veto $v(0.50) = 0.50 / (1 - 0.50) = 1$. Arvoa 50 % pienemmät osuudet johtavat vetoihin, jotka ovat ykköstä pienempiä (mutta kuitenkin positiivisia). Esimerkiksi edellä esitettyyn kvantitatiivisten artikkelien osuuteen 25 % liittyy veto

$$v(0.25) = 0.25 / (1 - 0.25) = 0.333.$$

Osuuden 10 % veto on 0.111. Arvon 50 % ylittävät osuudet johtavat puolestaan ykköstä suurempiin, periaatteessa rajattoman suuriinkin vedon arvoihin. Esimerkiksi $v(99.9) \approx 1\,000$ ja

$v(99.999) \approx 100\,000$. Arvon 50 % suhteen symmetrisesti sijaitsevat osuudet (esimerkiksi 10 ja 90 %) saavat vedoikseen käänteisluvut (9.000 ja 0.111). Tämän kanssa käy yksiin se, että osuutta 100 % vastaava veto on ääretön ja osuutta nolla vastaava nolla. Muutamia osuuksiin liittyviä vedon arvoja on koottu Taulukkoon 1.

Tulkinta vedonlyöntikontekstissa

Ajatellaan arpajaisia, joissa voittoarvalla saa takaisin arvan hinnan sekä yhden euron voittoa. On ilmeistä, että jos voittoarpoja on suhteellisesti ottaen paljon eli voittoarvan saamisen todennäköisyys on suuri, uskaltaa arvasta maksaa paljonkin, koska arvan hinnan suuria menetyksiä - tyhjiä arpoja - tulee harvoin. Mutta jos voittoarpoja on niukasti, eivät arvat ole paljonkaan arvoisia, sillä hinnan menetyksiä tulee usein.

Jos voittoarpojen osuus on esimerkiksi 25 %, arvan ostaja menettää arvan hinnan keskimäärin kolmella kerralla neljästä. Jos arpa maksaa 1/3 euroa eli reilut 33 senttiä, hän menettää näillä kolmella kerralla yhteensä yhden euron. 'Neljännellä' kerralla hän osuu voittoarpaan, saa takaisin arvan hinnan ja voittona yhden euron, joka kompensoi kolmen tyhjän arvan aiheuttaman menetyksen. Edellä laskettiin, että $v(0.25) = 0.333$. Voittotodennäköisyyden (25 %) veto antaa siis reilun pelin mukaisen arvan hinnan (33.3 snt). Arpojen ostaminen hinnalla, joka ylittää 33 senttiä, on pitkällä tähtäimellä tappiollista; jos hinta jää alle 33 sentin, on aika lähteä kioskille.

Veto-suure antaa kvantitatiivisen täsmennyksen edellä esitetylle, arvan hintaa koskevalle heuristiselle päättelylle: reilun pelin mukainen arvan hinta saadaan laskemalla voittoarpojen osuuden veto. Pelaaja ei pitkällä tähtäimellä joudu tappiolle, ellei maksa arvoista sitä enempää.

Osuuden laskeminen vedosta

Jos veto v tunnetaan, on sitä vastaava osuus tai todennäköisyys p helposti laskettavissa:

$$p = v / (1 + v) \quad (2)$$

Jos esimerkiksi $v(p) = 2.73$, on $p = 2.73 / (1 + 2.73) = 73.2\%$. Tulos on helppo tarkistaa:

$$v(73.2) = 0.732 / (1 - 0.732) = 2.73.$$

Jokaista osuutta vastaa siis tietty veto, jokaista vetoa tietty osuus. Jos arvan hinta tunnetaan, saadaan siitä laskettua reilun pelin mukainen voittoarpojen osuus. Jos voittoarpojen osuus on sitä pienempi, on arvonta pitkällä tähtäimellä ostajalle epäedullinen (ja myyjälle edullinen). Urheiluviedonlyönissä ammattipelaajat etsivät ns. ylikertoimia. Niitä syntyy, kun vedon tarjoaja - ainakin pelaajan mielestä - on arvioinut voittotodennäköisyyden alakanttiin ja myy siis 'arpojaan' liian halvalla oikeaan todennäköisyyteen verrattuna. Silloin on panostettava heti, eikä vasta nyt!

Odds ratio eli vetosuhde

Veto vaikuttaa hankalalta tavalta kuvata osuuksia tai todennäköisyyksiä. Sinänsä ei olekaan mitään tarvetta siirtyä käyttämään vetoa tilanteissa, jossa halutaan ilmaista itse osuuksien suuruuksia. Vedonlyöntihenkinen lähestymistapa voidaan huoletta jättää säväyttämään laimeaa oluttaan särpiviä brittejä. Meille on tutumpaa puhua osuudesta 25 % kuin vedosta 0.333 (tai sen käänteisluvusta 3).

Mutta tilanne muuttuu, kun halutaan verrata kahta osuutta. Osuuksien olemusta rajojen 0 ja 100 % rajoittamina lukuina kunnioittava tapa verrata kahta osuutta on laskea niitä vastaavien vetojen osamäärä. Näin syntyy vetojen suhde eli vetosuhde (odds ratio). Sen avulla kuvataan

sitä, kuinka paljon kaksi todennäköisyyttä tai suhteellista osuutta poikkeavat toisistaan, kuinka erilaisia ne ovat. Vetosuhteella operointi säilyttää osuudet osuuksina, eikä koskaan johda välin $[0, 100]$ ulkopuolella oleviin, osuuksiksi kelpaamattomiin arvoihin.

Vetosuhde (VS), joka vertaa osuutta p_T osuuteen p_M , määritellään näin:

$$VS = v(p_T) / v(p_M)$$

Esimerkiksi osuuksien 73.2 % ja 25.0 % välinen vetosuhde on

$$VS = v(73.2) / v(25.0) = 8.2$$

siillä $v(73.2) = 2.73$ ja $v(25.0) = 0.333$. Toisin kirjoitettuna tämä on

$$v(73.2) = 8.2 \times v(25.0).$$

Osuuden 73.2 % veto on siten 8.2-kertainen osuuden 25.0 % vetoon verrattuna.

Koska 73.2 % ei osuuteen 25.0 % verrattuna ole edes kolminkertainen, ei vetosuhde todellakaan ilmaise osuuksien tai todennäköisyyksien suhdetta. Se kertoo – nimensä ja määritelmän mukaisesti – vetojen suhteen eli

$$v(p_T) = VS \times v(p_M) \quad (3)$$

missä VS on osuuksien p_T ja p_M välinen vetosuhde. Vetosuhteen arvojen tulkinnan on perustuttava tähän yhteyteen.

Osuuksien suhdetta kutsutaan riskisuhdeksi (risk ratio)

Kun osuuksia verrataan suoraan niiden suhteen avulla, puhutaan riskisuhdeesta (risk ratio). Sen käyttö on tavallista lääketieteessä, erityisesti epidemiologiassa. Pohjois-Karjala-projektissa

arvioitiin, että sydän- ja verisuonitautiperäisen kuoleman riski tupakoitsijoilla on 2.1-kertainen tupakoimattomiin verrattuna. Tämä merkitsee sitä, että tupakoitsijan sydänkuoleman todennäköisyys on 2.1-kertainen tupakoimattomaan verrattuna.

Riskisuhde soveltuu kuitenkin vain pienten osuukien vertailuun, sillä - kuten aiemmin nähtiin - kertolasku johtaa helposti sataa prosenttia suurempiin 'osuuksiin'. Pienillä osuuksilla riskisuhteen ja vetosuhteen lukuarvot ovat lähellä toisiaan. Osuukien 8 % ja 4 % riskisuhde on 2.000 ja vetosuhde 2.087; osuukien 80 % ja 40 % riskisuhde on edelleen 2.000, mutta vetosuhde jo 6.000; osuuksilla 95 % ja 47.5 % vetosuhde on peräti 21. Kun lukuarvot voivat olla näin erilaisia, on tärkeää olla selvillä siitä, kumpaan vertailun tapaan on käytetty.

Riskisuhteen arvosta ei voi laskea vastaavaa vetosuhdetta (eikä tietysti toisinkaan päin) ilman, että jompikumpi vertailtavista osuuksista tunnetaan. Riskisuhteen ja vetosuhteen yhteyden kaavan antaa esimerkiksi Rita ja Komonen (2008, sivu 71, kaava 4).

Osuukien vertailu on eri asia kuin niiden suuruudet

Vetosuhde vertaa vetoja ja sitä kautta osuukia. Pelkästä vetosuhteen arvosta, esimerkiksi 8.2:stä, ei voi päätellä vertailun kohteena olleiden osuukien suuruuksia. Osuukien 25.0 % ja 73.2 % välinen vetosuhde on 8.2. Muita pareja, joista laskettu vetosuhde on yhtä suuri, ovat esimerkiksi 1.0 % ja 7.6 %; 10.0 % ja 47.7 %; 50.0 % ja 89.1 % sekä 90.0 % ja 98.7 %². Kaikissa näissä pareissa verrattavat kaksi osuutta ovat vetosuhteella mitattuna yhtä kaukana toisistaan, yhtä erilaisia. (Kaikissa pareissa sekä osuukien suhde että niiden erotus on erilainen.)

Lukijaa kiinnostaa kuitenkin paitsi se, kuinka erilaisia kvantitatiivisten artikkelien osuudet turkulaisten ja muiden kirjoittamien artikkelien joukossa ovat, myös se, kuinka yleisiä kyseiset artikkelit kummassakin kirjoittajaryhmässä ovat. Tällaista tietoa ei Erolan ja Räsäsen artikkelista löydy, sillä Taulukossa I (s. 303) on annettu artikkelityyppien prosenttijakauma lehdittäin, mutta ei yliopistoittain. Tämän vuoksi olemme edellä joutuneet arvaamaan kvantitatiivisten artikkelien osuuden muiden kuin turkulaisten kirjoittamien artikkelien joukossa (arvasimme 25.0 %), ja laskeneet ilmoitetun vetosuhteen 8.2 avulla, kuinka suuri vastaava osuus turkulaisten kirjoittajien joukossa on (saimme 73.2 %).

Osuukien suuruuksia koskeva kuvaileva tieto on ilmiön ymmärtämisen kannalta arvokasta. Koulutusmahdollisuuksien luokkaeroja tutkittaessa (Kivinen et al. 2007) on eron suuruuden lisäksi kiinnostavaa tietää sekin, kuinka suuri osuus akateemisten ja ei-akateemisten vanhempien lapsista päätyi yliopistoon. Koska vetosuhteen arvo kirjaa nimenomaan osuukien vertailun tuloksen, on itse osuudet ilmoitettava erikseen.

Osuukien ilmoittaminen on erityisen tärkeää silloin, kun tutkimus käsittelee ilmiöaluetta, jota on tutkittu vähän tai ei lainkaan tai jossa voi olettaa tapahtuneen muutoksia. Tällöin – kuten Erolan ja Räsäsen tapauksessa – osuukien suuruuksia ei voi arvioida aikaisemman tutkimuksen perusteella. Osuukien puuttuminen tekee vaikeaksi myös vaikkapa kvantitatiivisten menetelmien käytön sukupuolittuneisuutta koskevien tulosten kansainvälisen vertailun (esim. Grant & Ward 1987 ja Platt 2006).

Erolan ja Räsäsen päättelyn läpinäkyvyyttä heikentää vielä se, että sen enempää aineiston kuvauksen yhteydessä kuin artikkelin kuvioissa tai taulukoissa ei anneta tietoja havaintojen määrästä. Niiden avulla lukija voisi edes itse yrittää muodostaa käsityksen annettujen veto-

suhteiden täsmällisyydestä (precision), kun sen enempää keskivirheitä kuin luottamusvälejäkään ei löydy.

Vetosuhde ja kvantitatiiviset selittäjät

Nummenmaan (2007, s. 327) esimerkissä tulkittiin vetosuhteen arvo 1.05 näin:

"-- jokainen äidin ikävuosi kasvattaa 5 %:lla turvallisen kiintymyksen todennäköisyyttä."

Vetosuhdetta on tässä käytetty kuvaamaan kvantitatiivisen suureen, äidin iän, vaikutuksen suuruutta lapsen turvallisen kiintymyksen kehittymisen todennäköisyyteen. Kuten jo aiemmin kirjoitetusta voi päätellä, tämäkin vetosuhteen tulkinta ei ole kunnossa. Vetosuhdetta voi kuitenkin mielekkäästi käyttää – ja tulkita - kvantitatiivisten selittäjien yhteydessä. Ongelmallista on ensi silmäyksellä se, että vuosissa mitatulla äidin iällä voi olla yli kolmekymmentä eri arvoa, joiden kesken turvallisen kiintymyksen kehittäneiden lasten osuuksia vertaillaan. Tilanne voidaan helposti palauttaa edellä esitettyyn kahden osuuden vertaamiseen. Jos aineiston nuorimman äidin ikä oli 19 vuotta, verrataan ensin turvallisen kiintymyksen kehittäneiden lasten osuutta 20-vuotiailla äideillä vastaavaan osuuteen 19-vuotiailla äideillä. Seuraavaksi tehdään sama vertailu 22- ja 21-vuotiaiden äitien kohdalla. Näin jatketaan, kunnes saavutetaan aineiston iäkkäimmät äidit, joita verrataan vuotta nuorempiin. Tuloksena saadaan joukko vetosuhdetta, joille on yhteistä se, että ne vertaavat havaittuja lapsen turvallisen kiintymyksen kehittymisen osuuksia sellaisten äitien ryhmien kesken, joiden välillä on yhden vuoden ikäero.

län vaikutusta voidaan luonnehtia yhdellä luvulla, jos oletetaan, että yhden vuoden lisäys äidin iässä vaikuttaa turvallisen kiintymyksen kehittymiseen samalla tavalla riippumatta siitä, mistä iästä lähdetään liikkeelle. (län vaikutus ajatellaan

siis lineaariseksi.) Tehty oletus johtaa laskemaan 'keskiarvon' kaikista havaituista yhden lisävuoden vaikutusta kuvaavista vetosuhdeista. Teknisesti asia ei tilasto-ohjelmassa tapahdu ihan näin, mutta tulostuksessa olevan $\text{Exp}(B)$ -suureen tulkinnan kannalta kysymys on tästä.

Nummenmaan (2007) tulkinnan 'vetomuotoilu' menisikin tähän tapaan:

turvallisen kiintymyksen veto kasvaa 5 %:lla jokaista äidin ikävuotta kohti

Jos vetopuhe tuntuu abstraktilta tai vielä vieraalta, voi iän vaikutusta kuvaavan vetosuhteen arvoa havainnollistaa antamalla esimerkkejä osuuksista, joiden välinen vetosuhte on 1.05:

Vetosuhteen arvo 1.05 merkitsee sitä, että jos lapsen turvallisen kiintymyksen kehittyminen tietyn ikäisillä äideillä on 50 %, on se vuotta vanhemmilla 51.2 %.

Lähtöarvoksi voi ottaa systemaattisesti juuri osuuden 50 %, sillä sen veto on 1, mikä helpottaa laskutoimituksia. Tässä tarvitaan kyllä myös sisällöllistä harkintaa, sillä lukija tulkitsee annetut osuudet lähes varmasti ilmauksena ilmiön yleisyydestä. Tässä tapauksessa käytetty arvo 50 % antaneekin aliarvion turvallisen kiintymyksen kehityksen yleisyydestä, sillä koko aineistossa tuo osuus on noin 70 % (29/41). Vetosuhteen arvon tulkintaan saattaakin olla järkevää yhdistää tieto havaitusta osuuden suuruudesta. Jos todennäköisyys on 30-vuotiailla äideillä 66 %, voidaan vetosuhteen arvoa 1.05 havainnollistaa sanomalla:

Lapsen turvallisen kiintymyksen kehittymisen todennäköisyys 30-vuotiailla äideillä on 66.0 % ja 31-vuotiailla äideillä 67.1 %

Koska vetosuhte 1.05 kuvaa vain yhden lisävuoden vaikutusta, se jää lukuarvoltaan pieneksi. Paremmin hahmottuvia lukuja saadaan, kun efekti lasketaan pidempää ajanjaksoa kohti. Jos äitien ikäero on 10 vuotta, kasvaa turvallisen kiintymyksen kehittymisen veto 1.05^{10} – eli 1.63-kertaiseksi³, ts. 63 prosenttia. (Viiden prosentin vuosikasvu tuo kymmenessä vuodessa korkoa korolle peräti 63 %.) Tätäkin arvoa voi havainnollista edellä olevaan tapaan:

Lapsen turvallisen kiintymyksen kehittymisen todennäköisyys 30-vuotiailla äideillä on 66.0 %, kun se 40-vuotiailla on 76.0 %

Vaarana tällaisessa kikkaillussa on, että osuukseen arvojen ja niiden vertailun periaatteellinen ero hämärtyy.

Lopputulema

Sanotaan vetosana niin kuin se on

Yllä virheelliseksi osoitettu tulkinta ilmentää pyrkimystä selkeyttää vetosuhteen käsitettä ja sen tulkintaa palauttamalla se tutumpaan todennäköisyyden käsitteeseen. Osuukseen ja todennäköisyyksien vertailu eroaa väistämättä 'tavallisten' lukujen vertailusta siksi, että osuudet eivät voi saada negatiivisia tai yli 100 %:n meneviä arvoja. Tämän vuoksi niiden vertailua ei voi rajoituksitta perustaa suhteisiin (tai erotuksiin), vaan joudutaan kulkemaan vetojen (odds) kautta.

Käsityksemme mukaan on syytä olla potkimatta tulkintaa vastaan ja ottaa lusikka kauniiseen kätehen. Hyväksytään se karu tosiasia, että vetosuhteen selkeä havainnollistaminen todennäköisyyksien tai osuukseen tasolla ei ole mahdollista ja kirjataan vetosuhte sellaisena kuin se määritelmästä nousee:

Lapsen turvallisen kiintymyksen veto kasvaa 63 % aina kun äidin ikä kasvaa kymmenen vuotta

On vain opittava elämään vetosuhteiden kanssa ja puhumaan tutujen todennäköisyyksien sijaan vedoista. Erola ja Räsänen olisivat voineet kirjoittaa:

Kvantitatiivisen artikkelin kirjoittamisen veto on turkulaisilla yli kahdeksankertainen muihin verrattuna

tai

Kvantitatiivisen artikkelin kirjoittamisen vetosuhte on yli kahdeksan turkulaisten hyväksi

Helsingin Sanomien Osmo Kivisen tutkijaryhmän tuloksia esittelevästä uutisesta olisimme voineet lukea:

Yliopisto-opintojen veto on akateemisten vanhempien lapsilla 8.4-kertainen muiden kuin akateemisten vanhempien lapsiin verrattuna

tai

Yliopisto-opintojen vetosuhte on 8.4 akateemisten vanhempien lasten hyväksi

Lopuksi

Yhteiskuntatieteissä tutkimusmenetelmien suosio vaihtelee. Saattaa käydä niin, että vuonna 2030 muiden kuin turkulaisten Januksessa ja Sosiologiassa julkaistuista artikkeleista vain 20 % on kvantitatiivisia, ja turkulaisillakin osuus on enää 59 %. Lukija pohtikoon, onko ero kvantitatiivisen otteen suosiossa turkulaisten ja muiden kesken reilussa 20 vuodessa kasvanut vai vähentynyt. Tarpeelliset työkalut löytyvät edeltä. Lisä-

pohdinnan tarvetta saattaa syntyä, jos paikalliset trendiheilahdukset ovat olleet niin erilaisia, että tilanne on kääntynyt toisin päin, ts. osuus 20 % koskeekin turkulaisia ja 59 % muita.

Viitteet

¹ Kiitämme Tilastollisen työkalun teroitus - kurssin osanottajia kirjoituksen selkeyttä parantaneista oivallisista huomautuksista.

² Pareja, joissa vetosuhte on juuri 8,2, on äärettömän monta: kun yhtälöön (3) sijoitetaan mikä tahansa arvo osuudelle p_H , voidaan laskea veto $v(p_H)$ ja siitä edelleen osuus p_T kaavaa (2) käyttäen.

³ Potenssiinkorotus on herkkä pyörästysvirheille: oikea vetosuhteen arvo saattaisi olla 1.045 (joka siis olisi pyörästynyt arvoon 1.05), jolloin kymmenen lisävuoden vaikutus vetosuhteeseen on vain 55 %; tai peräti 1.0549, jolloin saataisiin 71 %. Arvon yksi lähellä olevat vetosuhteen arvot onkin syytä antaa tavallista suuremmalla tarkkuudella. Vetosuhteen lukuarvo riippuu myös mittayksiköstä, sillä se kertoo vedon muuttumisen selittäjän yhden yksikön kasvua kohti. Jos äidin ikä olisi mitattu päivissä, olisi viiden prosentin vuosikasvu saavutettu vetosuhteella 1.00013. Tällaisilla yksikkömuunnoksilla ei ole vaikutusta vetosuhteen tilastolliseen merkitsevyyteen.

Kirjallisuus

Erola, Jani & Räsänen, Pekka (2007) Suomalaisen sosiaalitieteen menetelmätrendit 1990- ja 2000-luvuilla. *Janus* 15(4); 297–315.

Grant, Linda & Ward, Kathryn B. (1987) Is there an association between gender and methods in sociological research? *American Sociological Review* 52; 856–862.

Jokivuori, Pertti & Hietala, Risto (2007) Määrällisiä tarinoita. Monimuuttujamenetelmien käyttö ja tulkinta. WSOY: Helsinki.

Kivinen, Osmo, Hedman, Juha & Kaipainen, Päivi (2007) From Elite University to Mass Higher Education: Educational Expansion, Equality of Opportunity and Returns to University Education. *Acta Sociologica* 50(3): 231–247.

Nummenmaa, Lauri (2007) Käyttätymistieteiden tilastolliset menetelmät. 3. painos. Tammi: Helsinki.

Platt, Jennifer (2006) How distinctive are Canadian research methods? *Canadian Review of Sociology and Anthropology* 43(2); 205–231.

Rita, Hannu (2003) Arveita vedoista. *Yliopisto* 5–6.

Rita, Hannu (2004): Vetosuhte (odds ratio) ei ole todennäköisyyksien suhde. *Metsätieteen aikakauskirja* 2: 207 – 212.

Rita, Hannu & Komonen, Atte (2008) Odds ratio: an ecologically sound tool to compare proportions. *Annales Zoologici Fennici* 45: 66–72.

Taulukko I Vetosuhteen (odds ratio) ja sen laskemiseen tarvittavien vetojen (odds) arvoja, kun osuutta p (rivit) verrataan osuuteen r (sarakeet). Esimerkiksi rivillä 20 ja sarakeella 10 oleva vetosuhteen arvo 2.25 saadaan jakamalla osuuden 0.20 veto (0.25) osuuden 0.10 vedolla (noin 0.11) (Rita & Komonen 2008, muokattu)

	r (%)	5	10	20	30	40	50	60	70	80	90	95
p (%)	Veto	0.05	0.11	0.25	0.43	0.67	1.00	1.50	2.33	4.00	9.00	19.00
5	0.05	1.00	0.47	0.21	0.12	0.08	0.05	0.04	0.02	0.01	0.01	0.00
10	0.11	2.11	1.00	0.44	0.26	0.17	0.11	0.07	0.05	0.03	0.01	0.01
20	0.25	4.75	2.25	1.00	0.58	0.38	0.25	0.17	0.11	0.06	0.03	0.01
30	0.43	8.14	3.86	1.71	1.00	0.64	0.43	0.29	0.18	0.11	0.05	0.02
40	0.67	12.67	6.00	2.67	1.56	1.00	0.67	0.44	0.29	0.17	0.07	0.04
50	1.00	19.00	9.00	4.00	2.33	1.50	1.00	0.67	0.43	0.25	0.11	0.05
60	1.50	28.50	13.50	6.00	3.50	2.25	1.50	1.00	0.64	0.38	0.17	0.08
70	2.33	44.33	21.00	9.33	5.44	3.50	2.33	1.56	1.00	0.58	0.26	0.12
80	4.00	76.00	36.00	16.00	9.33	6.00	4.00	2.67	1.71	1.00	0.44	0.21
90	9.00	171.0	81.00	36.00	21.00	13.50	9.00	6.00	3.86	2.25	1.00	0.47
95	19.00	361.0	171.0	76.00	44.33	28.50	19.00	12.67	8.14	4.75	2.11	1.00