

Veikkauksen pelaajien pelaaminen suhteessa
ulkomaisten vedonlyöntiyhtiöiden
todennäköisyysarvioihin

Jussi Vesanen

Tilastotieteen pro gradu -tutkielma

Jyväskylän yliopisto
Matematiikan ja tilastotieteen laitos
6. joulukuuta 2016

JYVÄSKYLÄN YLIOPISTO

Matematiikan ja tilastotieteen laitos

Vesanen, Jussi: Veikkauksen pelaajien pelaaminen suhteessa ulkomaisten vedonlyöntiyhtiöiden todennäköisyysarvioihin

Tilastotieteen pro gradu -tutkielma, 30 sivua

6. joulukuuta 2016

Tiivistelmä

Tällä tutkielmalla on kaksi päätavoitetta. Ensimmäinen niistä on muodostaa lineaarinen regressiomalli, jonka avulla ennustetaan neljän ulkomaisen vedonlyöntiyhtiön arvioita jalkapallo-otteluiden lopputuloksien todennäköisyyksille. Tämän lisäksi mallin avulla ennustetaan myös Veikkauksen pelaajien pelaamien rahojen jakautumista eri lopputulosvaihtoehtojen kesken. Selittävinä muuttujina tässä työssä lähtökohtaisesti on ajateltu tarvittavan seuraavia muuttujia: vedonlyöntiyhtiö, ottelun ennakoitu tasaisuus sekä ottelun odotettu maalimäärä. Suurin mielenkiinto tässä kohdistuu siihen, onko tarvetta muuttujien välisille interaktioille. Tutkielman toinen tavoite on yksinkertaisesti selvittää, minkä tulosvaihtoehtojen kohdilla eroja on ja mihin suuntaan. Eroilla tarkoitetaan eroja nimenomaan Veikkauksen pelaajien pelaamisessa verrattuna ulkomaisten vedonlyöntiyhtiöiden todennäköisyysarvioihin.

Tämän työn aineisto koostuu Veikkauksen ja neljän ulkomaisen vedonlyöntiyhtiön tulosvetokertoimista vuoden 2014 jalkapallon MM-kisojen kaikista 64 ottelusta. Veikkauksen kohdalla kertoimista saadaan osuus, millä kutakin tulosvaihtoehtoa on pelattu. Ulkomaisten vedonlyöntiyhtiöiden tapauksessa puolestaan kertoimista saadaan muodostettua yhtiöiden todennäköisyysarviot kullekin tulosvaihtoehdolle. Näille tehdään vielä logit-muunnos, minkä jälkeen muodostetaan lineaarinen regressiomalli ennustamaan todennäköisyysarvioita ja Veikkauksen pelaajien pelaamista.

Tutkielman tulokset kertovat siitä, että kaikkien tulosvaihtoehtojen kohdalla Veikkauksen pelaajien pelaaminen ei noudata samaa linjaa ulkomaisten vedonlyöntiyhtiöiden todennäköisyysarvioiden kanssa. Tässä työssä oli käsiteltävänä yhteensä 12 tulosvaihtoehtoa, joista peräti kahdeksan kohdalla saadaan viitteitä siitä, että eroja on. Näistä tulosvaihtoehdoista kuusi ovat sellaisia, joissa Veikkauksen pelaajat pelasivat kyseistä tulosvaihtoehtoa vähemmän kuin niitä muiden vedonlyöntiyhtiöiden todennäköisyysarvioiden mukaan tulisi pelata. Nämä kuusi tulosvaihtoehtoa ovat suosikin 1–0-, 2–0- ja 2–1-voitot, 0–0- ja 1–1-tasapelit sekä altavastajaan 0–1-voitto. Suosikin 3–1-voittoa ja tietyin ehdoin myös 3–2-voittoa Veikkauksen pelaajat puolestaan pelasivat enemmän kuin niitä muiden vedonlyöntiyhtiöiden arvioiden mukaan tulisi pelata. Vain neljän tulosvaihtoehdon tapauksessa Veikkauksen pelaajien pelaaminen noudatti samaa linjaa muiden vedonlyöntiyhtiöiden todennäköisyysarvioiden kanssa.

Avainsanat: Veikkaus, vedonlyönti, tulosveto, tulosvaihtoehto, kerroin, ulkomainen vedonlyöntiyhtiö, lineaarinen regressiomalli

Sisältö

1	Johdanto	1
2	Aineisto ja tutkimusongelma	3
2.1	Kertoimien kerääminen ja niiden muunnokset	3
2.2	Tutkimusongelman esittely	7
3	Lineaarinen regressiomalli	9
3.1	Mallien vertailu	10
4	Aineiston analyysi	12
4.1	Mallien vertailu	12
4.2	Mallien diagnostiikka	12
4.3	Tulokset tulosvaihtoehdoittain	16
5	Yhteenveto	27
6	Lähteet	30

1 Johdanto

Tulosveto on vedonlyöntimuoto, jossa yritetään veikata kohteena olevan ottelun täsmällistä lopputulosta. Tulosveto kuuluu niin kutsuttuihin vedonlyöntipeleihin, joihin katsotaan kuuluvan lisäksi esimerkiksi pitkäveto ja moniveto. Terveiden ja hyvinvoinnin laitoksen vuonna 2015 tehdyn raportin mukaan näitä vedonlyöntipelejä pelasi 21 % suomalaisista miehistä ja 3 % naisista (Salonen & Raisamo, 2015). Tulosvedon kohdalla useimmiten vedonlyöntiyhtiö määrää itse kertoimet, joilla kutakin vaihtoehtona olevaa tulosta voi veikata. Tässä tapauksessa vedonlyöntiyhtiöiden tavoitteena on asettaa kertoimet vastaamaan lopputuloksen todennäköisyyttä mahdollisimman tarkasti (Rautee & Merikallio, 2009). Toinen, joskin huomattavasti harvinaisempi tapa on, että vedonlyöntiyhtiöllä on vain vedonvälittäjän rooli. Tällöin kunkin tulosvaihtoehdon lopullinen kerroin määräytyy sen mukaan, kuinka paljon vedonlyöjät ovat siihen panostaneet kyseisen vedonlyöntiyhtiön kautta. Mitä enemmän johonkin lopputulokseen on rahaa panostettu, sen pienempi kerroin tälle lopulta muodostuu. Vedonvälittäjän roolissa toimiva vedonlyöntiyhtiö ottaa itselleen ennalta määrätyn suuruisen osuuden summasta, jonka vedonlyöjät panostavat yhteensä kohteeseen.

Monista ulkomaisista vedonlyöntiyhtiöistä poiketen Veikkaus ei tarjoa tulosvetokertoimia, vaan se toimii vain vedonvälittäjänä henkilöille, jotka haluavat veikata jonkin ottelun lopputulosta. Tällä tavalla Veikkaus varmistaa, ettei se voi jäädä tappiolle, päättyy ottelu miten tahansa. Tulosvetojen kohdalla Veikkaus ottaa itselleen niin sanottua vedonvälityspalkkiota 20 % summasta, jolla kutakin tulosvetokohdetta eli ottelua pelataan (<https://www.veikkaus.fi/fi/tulosveto#!/ohjeet> Marraskuu 2014). Vedonlyöjän kannalta tämäläinen vedonlyöntimuoto luo epävarmuutta siitä, kuinka kannattavaa jonkin tuloksen pelaaminen on. Epävarmuus johtuu siitä, että lopulliset kertoimet kullekin tulokselle selviävät vasta, kun kohde on sulkeutunut ja on tiedossa, kuinka suurella rahamäärällä mitään tulosta pelattiin.

Paljon yleisemmin käytössä oleva tapa on, että vedonlyöntiyhtiö itse tarjoaa kullekin tulokselle kertoimen. Silloin kun kertoimet asetetaan vedonlyöntiyhtiön toimesta, voi vedonlyöntiyhtiö jäädä tappiolle, jos oikean lopputuloksen kerroin on liian suuri suhteessa siihen, kuinka paljon sitä on pelattu. Tämän välttääkseen vedonlyöntiyhtiöt voivat muuttaa antamiaan kertoimia, mutta jokainen vedonlyöjä saa oikein veikatessaan rahansa niin moninkertaisena takaisin kuin vedonlyöntihetkellä tuloksen kerroin oli. Vedonlyöjälle tämä tapa on siinä mielessä turvallisempi, että hän tietää tarkalleen, kuinka moninkertaisena hän saa rahansa takaisin, jos veto osuu oikeaan.

Veikkauksen harjoittamassa pelimuodossa kertoimet muodostuvat pelkääntään sen perusteella, miten paljon pelaajat kullekin kohteelle rahaa panostavat. Tämä aiheuttaa sen, että jokaisen kohdetta veikkaavan panostus vaikuttaa lopullisiin kertoimiin. Michael Walker kertoo kirjassaan *The Psychology of Gambling*, että uhkapelaamisen motiiveina on yleensä joko raha tai jännityksen hakeminen tai näiden yhdistelmä (Walker, 1995). Tämä näkyy myös Veikkauksen pelaajissa. On oletettavaa, että osa pelaajista asettaa rahansa sellaiseen tulokseen, johon uskoo ottelun päättyvän välittämättä kertoimesta, sillä se voi muuttua paljonkin ennen kohteen sulkeutumista. Osa kohdetta pelaavista sen sijaan kuitenkin tarkistaa sen hetkiset kertoimet ja sijoittaa rahansa sen perusteella parhaaksi katsomiinsa tuloksiin. Suomalaisten vedonlyöjien pelaamista kuvaa hyvin se, että Urheilulehden Jorma Vuoksenmaata koskevassa artikkelissa Vuoksenmaa to-

teaa, että suomalaiset ovat maailman parhaita, kun puhutaan järkipelaamisesta (Nevalainen, 2016). Koska luultavasti osa kuitenkin pelaa rahansa kertoimista välittämättä, niin pelattavaan kohteeseen sijoitettavat rahat eivät välttämättä jakaudu samalla tavalla, kuin jos kertoimet olisivat vedonlyöntiyhtiön toimesta valmiiksi määrätty. Tästä ei ole olemassa aikaisempaa tutkimustietoa, jossa tätä olisi tilastollisilla menetelmillä tutkittu. Näin ollen ei ole täyttä varmuutta siitä, kuinka suuressa osassa nämä kertoimista välittämättä tehtävät vedot ovat Veikkauksen pelitavassa.

Tämän työn tavoitteena on selvittää, poikkeako rahojen jakautuminen eri tulosten kesken, kun käytössä ovat edellä mainitut kaksi erilaista vedonlyöntimuotoa. Tavoitteen saavuttamiseksi tässä työssä konstruoidaan lineaarinen regressiomalli, jolla pyritään ennustamaan sekä Veikkauksen pelaajien pelaaamista että ulkomaisten vedonlyöntiyhtiöiden todennäköisyysarvioiden muodostumista. Mallin avulla etsitään tulosvaihtoehtoja, joiden kohdalla on eroja. Lineaarisen regressiomallin sijaan olisi voitu käyttää myös sekamallia. Sen avulla voidaan mallintaa aineistoa, jossa havainnot eivät ole riippumattomia (Brown & Prescott, 1999). Sekamallin tapauksessa olisi näin ollen huomioitu ottelun aiheuttama riippuvuus. Koska eri tulosvaihtoehtoja on kuitenkin runsaasti, niin yksinkertaisuuden vuoksi tässä työssä käytetään lineaarista regressiomallia sekamallin sijaan ja analyysit suoritetaan tulosvaihtoehto kerrallaan. Lineaarisen mallin etuna on sen helppo ymmärrettävyys ja tulkittavuus moniin muihin malleihin verrattuna (Sengupta & Jammalamadaka, 2003).

Työn rakenne on seuraava: 2. luvussa esitellään tutkimusongelmat ja aineisto sekä käydään tarkemmin läpi aineiston kerääminen ja muokkaaminen. Luvussa 3 syvennytään lineaarisen regressiomallin teoriaan ja mallien vertailuun. Tämän jälkeen luvussa 4 esitellään aineiston analysointia ja saatuja tuloksia. Lopuksi luvussa 5 on yhteenveto ja pohdinta tuloksista.

2 Aineisto ja tutkimusongelma

Tässä tutkimuksessa käytetty aineisto sisältää vuoden 2014 jalkapallon MM-kisoissa pelattujen otteluiden tulosvetokertoimia viideltä eri vedonlyöntiyhtiöltä. Näihin viiteen yritykseen lukeutuvat Veikkaus sekä neljä ulkomaista vedonlyöntiyhtiötä, jotka ovat William Hill, Ladbrokes, Unibet ja Nordichet.

2.1 Kertoimien kerääminen ja niiden muunnokset

Jokaiselta neljältä tutkimuksessa mukana olevalta ulkomaiselta yhtiöltä on kerätty talteen niiden tarjoamat kertoimet eri tulosvaihtoehdoille. Näin on toimitu jokaisen MM-kisoissa pelatun ottelun kohdalla. Kustakin ottelusta on yksi vedonlyöntiyhtiö tarjonnut yleensä noin 40 eri tulosvaihtoehdolle jonkin kertoimen. Se, kuinka monelle tulokselle kerroin on asetettu, riippuu pitkälti vedonlyöntiyhtiöstä ja kohteena olevasta ottelusta. Kaikkia todennäköisimpiä tuloksia pystyi pelaamaan jokaisen yhtiön kautta, mutta sen sijaan epätodennäköisimmille lopputuloksille eivät kaikki yritykset tarjonneet kertoimia. Näistä neljästä yhtiöstä kaikkein niukimmin kertoimia tarjosi Unibet, joka asetti ottelusta riippuen noin 30 todennäköisimmälle lopputulokselle kertoimen. Eniten kohteena olevia tulosvaihtoehtoja puolestaan oli useimmiten Nordichetillä, joka asetti parhaimmillaan lähes sadalle tulokselle kertoimen. Mitä pidemmälle turnaus eteni, sitä useammalle tulosvaihtoehdolle yhtiöt yleensä asettivat kertoimia. Tässä tutkimuksessa keskitytään kuitenkin vain kaikkein yleisimpiin tulosvaihtoehtoihin. Näin ollen tässä työssä mukana oleville tulosvaihtoehdoille on jokainen yhtiö asettanut kertoimen jokaiseen otteluun.

Tulosvetokertoimien kerääminen jokaisen neljän ulkomaisen vedonlyöntiyhtiön tapauksessa tapahtui samana päivänä, kuin kyseinen ottelu pelattiin. Melkein jokaisen ottelun kertoimet kerättiin 1–2 tuntia ennen kuin ensimmäinen ottelu sinä päivänä alkoi. Ottelupäivinä pelattiin yhdestä neljään ottelua. Näin ollen kunkin ottelun kertoimet otettiin talteen pääpiirteittäin 1–8 tuntia ennen kyseisen ottelun alkua. Joissakin tapauksissa joko osa tai kaikki neljä aineistossa mukana olevaa ulkomaista vedonlyöntiyhtiötä muuttivat hieman joidenkin tulosvaihtoehtojen kertoimia päivän edetessä. Tämä johtuu siitä, että yritykset reagoivat tilanteisiin, jotka muuttavat ottelun voimasuhteita. Esimerkiksi jonkin pelaajan loukkaantuminen voi olla tällainen tilanne. Jos jotakin tulosvaihtoehtoa pelataan enemmän kuin kyseinen yhtiö on arvioinut, on vaara, että ottelun päättyessä kyseiseen tulokseen yhtiö jää tappiolle. Myös tällöin yrityksen on syytä tarkistaa kertoimiaan. Tässä työssä ei kertoimien muutosta ole kuitenkaan tarpeen huomioida.

Neljän edellä mainitun yhtiön lisäksi tätä työtä varten kerättiin kertoimia Veikkaukselta, jonka tulosvetokertoimet muodostuvat eri tavalla kuin tutkimuksen muilla yhtiöillä. Veikkauksen kertoimet muodostuvat sen mukaan, kuinka suurella rahamäärällä mitään tulosvaihtoehtoa pelataan. Mitä enemmän jotakin tulosta on pelattu, sen pienempi kyseisen tuloksen kertoimesta lopulta muodostuu. Lopulliset kertoimet kullekin tulosvaihtoehdolle selviävät siten vasta, kun pelattava kohde on sulkeutunut. Tämän vuoksi Veikkauksen tulosvetokertoimet otettiin talteen vasta, kun kertoimet olivat lopullisia, eli kohde oli sulkeutunut. Kerääminen tapahtui Ylen teksti-tv:n kautta, jossa jokaisesta tulosvedon kohteena olevasta ottelusta näkyy noin 40 eri tulosvaihtoehtoa ja niiden kertoimet. Nämä tulosvaihtoehdot muodostuvat kaikista pelatuista tasapelituloksista,

16 eniten pelatusta kotivoittoon päättyvästä tuloksesta sekä 16 eniten pelatusta vierasvoittoon päättyvästä tuloksesta. Tässä työssä tyydytään tarkastelemaan 12 tulosvaihtoehtoa, joten myös Veikkauksen kohdalla saatiin jokaisesta ottelusta tarvittavat kertoimet kerättyä. Tulosvaihtoehtojen rajausta perustuu siihen, että mukaan on otettu kaikki yleisimmät lopputulokset. Tutkimusta varten otettiin talteen myös jokaisen ottelun vaihto, joka kertoo, kuinka suurella rahamäärällä kunkin kohteen lopputulosta veikattiin Veikkauksen kautta. Taulukossa 1 on esitetty ottelun Brasilia–Kroatia tässä työssä mukana olevat tulosvetokertoimet kaikilta viideltä yhtiöltä.

Taulukko 1: Tutkimuksessa mukana olevien yhtiöiden tulosvetokertoimia Brasilia–Kroatia-ottelun eri tulosvaihtoehtoilta.

Brasilia–Kroatia		Vaihto: 125 243 euroa			
	William Hill	Ladbrokes	Unibet	Nordicbet	Veikkaus
1–0	5.50	5.50	5.00	5.15	6.20
2–0	5.50	5.50	4.75	5.00	6.14
2–1	10.00	10.00	9.00	9.50	10.78
3–0	8.00	7.50	7.00	7.70	7.79
3–1	15.00	15.00	14.00	14.00	10.20
3–2	51.00	51.00	46.00	48.00	33.83
	William Hill	Ladbrokes	Unibet	Nordicbet	Veikkaus
0–0	10.00	9.00	10.50	11.00	12.73
1–1	10.00	10.00	9.50	10.75	12.59
2–2	34.00	34.00	28.00	38.00	28.87
	William Hill	Ladbrokes	Unibet	Nordicbet	Veikkaus
0–1	29.00	21.00	20.00	22.00	26.96
0–2	67.00	51.00	56.00	80.00	58.82
1–2	34.00	34.00	33.00	45.00	30.22

Taulukossa 1 olevat kertoimet kertovat sen, kuinka moninkertaisena kukin yhtiö palauttaa vedonlyöjän tulokselle asettaman panoksen, jos veikattava ottelu päättyy kyseiseen tulokseen. Jokainen tutkimuksessa mukana oleva yhtiö tarjosi kertoimet tuossa muodossa, minkä vuoksi ne ovat sellaisena myös alkuperäisessä aineistossa. Tulkittavuuden kannalta on kuitenkin helpompaa, jos kertoimet korvataan niiden käänteisluvuilla, jolloin niiden arvojoukko vaihtelee lukujen 0 ja 1 välillä (Taulukko 2). Nyt jokaista kerrointa vastaa luku, joka ilmoittaa, kuinka suurella osuudella kutakin tulosta korkeintaan saa pelata, jotta vedonlyöntitoimisto ei jää tuloksen osuessa tappiolle. Veikkauksen kohdalla luvut ovat 1.25-kertaisia suhteessa siihen, kuinka suuri osa kohteeseen panostetuista rahoista on pelattu juuri kyseiselle tulokselle. Tämä johtuu siitä, että Veikkaus ottaa 20 % kohteen kokonaisvaihdosta itselleen. Nyt näitä kertoimista saatuja osuuksia voidaan vertailla eri vedonlyöntiyhtiöiden kesken. Vedonlyöjän kannalta järkevintä on pelata rahat haluamalleen tulokselle sen yhtiön kautta, jonka kohdalla luku on pienin, sillä silloin kerroin on suurin.

Taulukko 2: Brasilia–Kroatia-ottelun tulosvetokertoimien käänteisluvut eri vedonlyöntiyhtiöissä.

Brasilia–Kroatia					Vaihto: 125 243 euroa
	William Hill	Ladbrokes	Unibet	Nordicbet	Veikkaus
1–0	0.182	0.182	0.200	0.194	0.161
2–0	0.182	0.182	0.211	0.200	0.163
2–1	0.100	0.100	0.111	0.105	0.093
3–0	0.125	0.133	0.143	0.130	0.128
3–1	0.067	0.067	0.071	0.071	0.098
3–2	0.020	0.020	0.022	0.021	0.030
	William Hill	Ladbrokes	Unibet	Nordicbet	Veikkaus
0–0	0.100	0.111	0.095	0.091	0.079
1–1	0.100	0.100	0.105	0.093	0.079
2–2	0.029	0.029	0.036	0.026	0.035
	William Hill	Ladbrokes	Unibet	Nordicbet	Veikkaus
0–1	0.034	0.048	0.050	0.045	0.037
0–2	0.015	0.020	0.018	0.013	0.017
1–2	0.029	0.029	0.030	0.022	0.033

Nämä kertoimien käänteisluvut kerrotaan vielä lisäksi kunkin vedonlyöntiyhtiön arvioidulla palautusprosentilla. Palautusprosentilla tarkoitetaan sitä, kuinka monta prosenttia kohteeseen pelatuista rahoista keskimäärin palautuu voittoina takaisin oikein veikanneille pelaajille. Jokaiselle tutkimuksen ulkomaiselle vedonlyöntiyhtiölle on saatu kertoimien perusteella arvio niiden käyttämästä palautusprosentista. Jos yhtiön palautusprosentti olisi 100, niin silloin kaikkien mahdollisten tulosvaihtoehtojen kertoimien käänteislukujen summaksi muodostuisi luku 1. Koska vedonlyöntiyhtiöiden tavoitteena on kuitenkin kerätä itselleen voittoa, niin niiden tulee käyttää pienempiä kertoimia. Tällöin kaikkien tulosvaihtoehtojen kertoimien käänteislukujen summaksi tulee lukua 1 suurempi luku. Arvio kunkin yhtiön käyttämälle palautusprosentille saadaankin jakamalla luku 1 luvulla, joka muodostuu kaikkien mahdollisten tulosvaihtoehtojen kertoimien käänteislukujen summasta. William Hillille ja Nordicbetille saatiin tätä tapaa käyttäen palautusprosentiksi 82, Ladbrokesille 80 ja Unibetille 76. Veikkauksen palautusprosentti on aiemmin mainittu 80. Kertomalla kertoimien käänteisluvut jokaisen yhtiön omalla palautusprosentilla saadaan jokaiselle tulosvaihtoehdolle luku, joka kertoo vedonlyöntiyhtiön arvion kyseisen tulosvaihtoehdon todennäköisyydelle. Veikkauksen kohdalla kyseinen luku kertoo sen, kuinka suuri osa kohteeseen sijoitetusta rahamäärästä on kullekin tulosvaihtoehdolle pelattu. Seuraavassa on havainnollistava esimerkki siitä, miten kertoimista saadaan vedonlyöntiyhtiöiden todennäköisyysarviot (Taulukko 3).

Taulukko 3: Havainnollistava esimerkki, miten Brasilia–Kroatia-ottelun 1–0-tuloksen kertoimista saadaan todennäköisyysarviot.

Yhtiö	Kerroin	1/kerroin	Todennäköisyysarvio
William Hill	5.50	$1/5.50 = 0.182$	$0.82 \cdot 0.182 = 0.149$
Ladbrokes	5.50	$1/5.50 = 0.182$	$0.80 \cdot 0.182 = 0.145$
Unibet	5.00	$1/5.00 = 0.200$	$0.76 \cdot 0.200 = 0.152$
Nordicbet	5.15	$1/5.15 = 0.194$	$0.82 \cdot 0.194 = 0.159$
Veikkaus	6.20	$1/6.20 = 0.161$	$0.80 \cdot 0.161 = 0.129$

Tämän työn yhtenä tavoitteena on mallintaa vedonlyöntiyhtiöiden todennäköisyysarvioiden yhteyttä selittäviin muuttujiin lineaarisella regressiomallilla. Todennäköisyysarvio on luku välillä $[0,1]$. Jotta vaste voi saada arvoja koko reaaliakselilta, tehdään todennäköisyysarviolle logit-muunnos seuraavasti:

$$y = \text{logit}(\alpha) = \log\left(\frac{\alpha}{1-\alpha}\right),$$

missä α on kunkin tuloksen todennäköisyysarvio. Taulukossa 4 on esitetty vedonlyöntiyhtiöiden todennäköisyysarvioille tehdyt logit-muunnokset.

Taulukko 4: Brasilia–Kroatia-ottelun luvut logit-muunnoksen jälkeen.

Brasilia–Kroatia		Vaihto: 125 243 euroa			
	William Hill	Ladbrokes	Unibet	Nordicbet	Veikkaus
1–0	-1.742	-1.771	-1.719	-1.664	-1.910
2–0	-1.742	-1.771	-1.658	-1.629	-1.898
2–1	-2.415	-2.442	-2.383	-2.359	-2.524
3–0	-2.170	-2.125	-2.105	-2.127	-2.168
3–1	-2.850	-2.876	-2.858	-2.777	-2.464
3–2	-4.114	-4.139	-4.086	-4.052	-3.721
	William Hill	Ladbrokes	Unibet	Nordicbet	Veikkaus
0–0	-2.415	-2.327	-2.551	-2.519	-2.702
1–1	-2.415	-2.442	-2.442	-2.494	-2.690
2–2	-3.700	-3.726	-3.579	-3.814	-3.558
	William Hill	Ladbrokes	Unibet	Nordicbet	Veikkaus
0–1	-3.537	-3.229	-3.231	-3.252	-3.487
0–2	-4.391	-4.139	-4.286	-4.570	-4.284
1–2	-3.700	-3.726	-3.748	-3.987	-3.605

Tulosvetokertoimien lisäksi aineisto sisältää jokaisesta 64 ottelusta pitkävetokertoimet jokaiselta viideltä vedonlyöntiyhtiöltä (Taulukko 5). Nämä kertoimet kertovat sen, kuinka todennäköisinä eri yhtiöt pitävät ottelun päättymistä kotivoittoon, tasapeliin ja vierasvoittoon. Ne antavat osviittaa siitä, kuinka tasainen veikkattava ottelu ennakkoon on ja kumpi ottelussa on suosikki.

Taulukko 5: Brasilia–Kroatia-ottelun pitkävetokertoimet.

	1	X	2
William Hill	1.30	5.00	11.00
Ladbrokes	1.30	5.00	13.00
Unibet	1.32	5.00	12.00
Nordicbet	1.31	5.20	11.85
Veikkaus	1.27	4.80	12.50

Aineistoon on kerätty myös kertoimet sille, tuleeko ottelussa korkeintaan kaksi vai vähintään kolme maalia (Taulukko 6). Nämäkin kertoimet on kerätty jokaiselta viideltä vedonlyöntiyhtiöltä ja ne ovat suuntaa antavia sen suhteen, kuinka runsasmaalinen ottelu on kulloinkin odotettavissa. Toisin kuin tulosvedon kohdalla, Veikkaus asettaa muiden vedonlyöntitoimistojen tapaan edellä mainitut kertoimet itse.

Taulukko 6: Brasilia–Kroatia-ottelun alle 2.5 maalia ja yli 2.5 maalia -kertoimet.

	Alle 2.5	Yli 2.5
William Hill	1.73	2.00
Ladbrokes	1.75	1.95
Unibet	1.75	2.08
Nordicbet	1.78	2.09
Veikkaus	1.80	1.90

2.2 Tutkimusongelman esittely

Tämän tutkimuksen tavoitteena on selvittää, minkä tulosvaihtoehtojen kohdalla on eroa ulkomaisten vedonlyöntiyhtiöiden lopputulosten todennäköisyysarvioiden ja Veikkauksen pelaajien pelaamisen kesken. Halutaan löytää sellaisia tulosvaihtoehtoja, joissa ulkomaisten vedonlyöntiyhtiöiden todennäköisyysarvioilla ja Veikkauksen pelaajien pelaamisella on eroja veikkattavasta ottelusta riippumatta. Kiinnostuksen kohteena on myös se, onko veikkattavan ottelun joukkueiden välisellä ennakoidulla tasoerolla vaikutusta siihen, minkä tuloksien kohdalla eroja on. Myös siitä ollaan kiinnostuneita, vaikuttaako ottelun odotetun maalimäärän suuruus siihen, miten Veikkauksen pelaajien pelaaminen eroaa muiden yhtiöiden todennäköisyysarvioista. Jos Veikkauksen kautta rahansa pelaavien vedonlyöjien pelaamat rahat jakautuvat eri tulosvaihtoehtojen kesken niin kuin muilla vedonlyöntiyhtiöillä, niin ero kertoimissa eri yhtiöiden välillä on systemaattista ja se johtuu vain siitä, että yritykset toimivat eri riskitasolla.

Taulukossa 7 on esitetty regressiomallin muuttujat, joiden avulla tuloskohtaisia eroja lopputulosten todennäköisyysarvioissa eri yhtiöiden kesken mallinnetaan. Vastemuuttujana on todennäköisyysarviosta tehty logit-muunnos. Yhtiömuuttuja sisältää kaikki viisi tutkimuksessa mukana olevaa vedonlyöntiyhtiötä. Tarkoituksena on verrata tulosvaihtoehto (esim. 1–0) kerrallaan Veikkauksen pe-

laajien pelaamien rahojen jakautumista muiden yritysten lopputulosten todennäköisyysarvioihin. Tutkimuksessa on mukana yhteensä 12 eri tulosvaihtoehtoa. Ne ovat suosikin 1-0-, 2-0-, 2-1-, 3-0-, 3-1- ja 3-2-voitot, altavastajaan 0-1-, 0-2- ja 1-2-voitot sekä 0-0-, 1-1- ja 2-2-tasapelit. Ne eivät välttämättä joka ottelussa olleet kaikista todennäköisimmät tulosvaihtoehdot, mutta pääpiirteittäin ne sitä ovat ja sen takia ne ovat valikoituneet tähän työhön mukaan.

Tasoero-muuttuja on jaettu kolmeen osaan sen mukaan, kuinka suuri tasoero vastakkain pelaavilla joukkueilla on ennen ottelua arvioitu olevan. Tämän muuttujan arvot perustuvat tutkimuksessa mukana olevien vedonlyöntiyhtiöiden arvioihin joukkueiden välisistä voimasuhteista. Jokaisen ottelun kohdalla muodostettiin pitkävetokertoimet tutkimuksessa mukana olevien yhtiöiden pitkävetokertoimien pohjalta niin, että jokaiselle merkille (1, X ja 2) otettiin mediaani kaikkien viiden yhtiön kertoimista kyseiselle merkille. Esimerkiksi Taulukosta 5 nähdään, että Brasilia–Kroatia-ottelun pitkävetokertoimiksi muodostui 1.3 Brasilian voitolle, 5 tasapelille ja 12 Kroatian voitolle mediaania käytettäessä. Näin muodostettiin jokaiselle ottelulle uudet pitkävetokertoimet, joiden pohjalta ottelut jaettiin kolmeen eri tasoeroluokkaan. Luokassa 1 ennakoitu tasoero on suurin. Tähän luokkaan kuuluvat kaikki ottelut, joissa muodostetuissa pitkävetokertoimissa altavastajaan kerroin on 6 tai enemmän. Luokassa 3 ennakoitu tasoero puolestaan on pienin ja siihen kuuluvat ne ottelut, joissa muodostetuissa kertoimissa suosikin kerroin on 2 tai enemmän. Kaikki muut ottelut luokitellaan tasoeroluokkaan 2.

Maalimäärä-muuttuja on jaettu kahteen osaan sen perusteella, kuinka suuri ottelun odotettu maalimäärä on. Nämäkin arvot pohjautuvat kaikkien viiden vedonlyöntiyhtiön arvioihin ottelun runsasmaalisuudesta. Myös tässä tapauksessa jokaisen ottelun kohdalla muodostettiin uudet kertoimet vaihtoehdoille alle 2.5 maalia ja yli 2.5 maalia. Menetelmä oli täsmälleen samanlainen kuin tasoero-muuttujan tapauksessa. Luokkaan 1 luokitellaan ottelut, joissa muodostetuissa kertoimissa yli 2.5 maalia saa kertoimeksi vähintään 2 ja luokkaan 2 ottelut, joissa vastaava kerroin on vähemmän kuin 2.

Taulukko 7: Muuttujien arvot ja niiden selitykset.

Muuttuja	Arvo
Yhtiö	1 = Veikkaus
	2 = William Hill
	3 = Ladbrokes
	4 = Unibet
	5 = Nordicbet
Tasoero	1 = suuri
	2 = keskisuuri
	3 = pieni
Maalimäärä	1 = pieni
	2 = suuri

3 Lineaarinen regressiomalli

Seuraava lineaarisen regressiomallin esitys perustuu teokseen Davison (2003). Regressiomalli kuvaa, kuinka vastemuuttujan arvot muuttuvat selittävien muuttujien arvojen muuttuessa. Regressioanalyysi on paljon käytetty menetelmä tilastotieteessä, sillä muuttujien selittäminen toisilla muuttujilla on keskeistä monissa tutkimuksissa. Muiden muuttujien arvojen vaihtelun avulla saadaan tietoa vastemuuttujan suhteesta kyseisiin muuttujiin tai voidaan ennustaa tulevia havaintoja. Monissa tutkimuksissa tämä on olennaista. Useimmissa tapauksissa vasteena on yksi muuttuja, jota selitetään usealla muuttujalla. Tällöin lineaarinen regressiomalli on muotoa

$$y_j = \beta_0 + \beta_1 x_{j1} + \varepsilon_j, \quad j = 1, \dots, n,$$

missä y_j on vaste, β_0 vakio, β_1 selittävän muuttujan x_{j1} kerroin ja ε_j on jäännöstermi. Tämä yleistyy useammalle selittävälle muuttujalle seuraavasti:

$$y_j = \beta_1 x_{j1} + \dots + \beta_p x_{jp} + \varepsilon_j = x_j^T \beta + \varepsilon_j,$$

missä $x_j^T = (x_{j1}, \dots, x_{jp})$ on $1 \times p$ -vektori selittäviä muuttujia, jotka selittävät vastetta y_j , β on $p \times 1$ -vektori tuntemattomia parametreja ja ε_j on havaitsematon virhemuuttuja. Matriisimerkinnöin

$$y = X\beta + \varepsilon,$$

missä y on muotoa $n \times 1$ oleva vektori, jonka j :s tekijä on y_j , X on $n \times p$ -vektori, jonka j :s rivi on x_j^T ja ε on muotoa $n \times 1$ oleva vektori, jonka j :s tekijä on ε_j .

Oletetaan, että virhetermit ε_j ovat riippumattomia ja noudattavat normaali-jakaumaa keskiarvolla 0 ja varianssilla σ^2 . Silloin vasteen y_j arvot ovat riippumattomia ja normaalisti jakautuneita keskiarvolla $x_j^T \beta$ ja varianssilla σ^2 . Uskottavuusfunktio parametreille β ja σ^2 on muotoa

$$L(\beta, \sigma^2) = \prod_{j=1}^n \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} (y_j - x_j^T \beta)^2 \right\},$$

jolloin log-uskottavuudeksi saadaan

$$\ell(\beta, \sigma^2) = -\frac{1}{2} \left\{ n \log \sigma^2 + \frac{1}{\sigma^2} \sum_{j=1}^n (y_j - x_j^T \beta)^2 \right\}.$$

Riippumatta parametrin σ^2 arvosta log-uskottavuus maksimoituu, kun β saa arvon, joka minimoi neliösumman

$$SS(\beta) = \sum_{j=1}^n (y_j - x_j^T \beta)^2 = (y - X\beta)^T (y - X\beta).$$

Suurimman uskottavuuden estimaatti parametrille β saadaan ratkaisemalla seuraavat yhtälöt:

$$\frac{\partial SS(\beta)}{\partial \beta_r} = 2 \sum_{j=1}^n x_{jr} (y_j - \beta^T x_j) = 0, \quad r = 1, \dots, p.$$

Matriisimuodossa voidaan kirjoittaa

$$X^T(y - X\beta) = 0,$$

mikä tarkoittaa, että $(X^T X)\beta = X^T y$. Jos syntynyt matriisi $X^T X$ on täysasteinen ja siten kääntyvä, pienimmän neliösumman estimaattori parametrille β on

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

Suurimman uskottavuuden estimaattori parametrille σ^2 saadaan sen profiiluskottavuudesta

$$\ell_p(\sigma^2) = \max_{\beta} \ell(\beta, \sigma^2) = -\frac{1}{2} \left\{ n \log \sigma^2 + \frac{1}{\sigma^2} (y - X\hat{\beta})^T (y - X\hat{\beta}) \right\},$$

josta seuraa, että suurimman uskottavuuden estimaattori parametrille σ^2 on

$$\hat{\sigma}^2 = n^{-1} (y - X\hat{\beta})^T (y - X\hat{\beta}) = n^{-1} \sum_{j=1}^n (y_j - x_j^T \hat{\beta})^2.$$

Estimaattori $\hat{\sigma}^2$ on kuitenkin harhainen. Harhaton estimaattori parametrille σ^2 on muotoa

$$s^2 = \frac{1}{n-p} (y - X\hat{\beta})^T (y - X\hat{\beta}) = \frac{1}{n-p} \sum_{j=1}^n (y_j - x_j^T \hat{\beta})^2.$$

Luottamusvälit parametrin β komponenteille saadaan estimaattorien $\hat{\beta}$ ja s^2 jakaumista. Parametrille β_r pätee

$$\hat{\beta}_r \sim N(\beta_r, \sigma^2 \nu_{rr}),$$

missä ν_{rr} on matriisin $(X^T X)^{-1}$ r :s diagonaalialkio ja $\hat{\beta}$ on riippumaton estimaattorista s^2 , joka noudattaa jakaumaa $(n-p)^{-1} \sigma^2 \chi_{n-p}^2$. Näin ollen

$$T = \frac{\hat{\beta}_r - \beta_r}{\sqrt{s^2 \nu_{rr}}} \sim t_{n-p}.$$

3.1 Mallien vertailu

Usein regressiomalleissa on tärkeää tietää, onko selittävillä muuttujilla vaikutusta vasteeseen eli mitkä selittävästä muuttujista ovat tarpeellisia. Seuraava esitys perustuu teokseen Davison (2003). Oletetaan lineaarinen malli

$$y = X\beta + \varepsilon = (X_1, X_2) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \varepsilon = X_1 \beta_1 + X_2 \beta_2 + \varepsilon,$$

missä X_1 on $n \times q$ -matriisi, X_2 on $n \times (p-q)$ -matriisi, $q < p$, sekä β_1 ja β_2 ovat vastaavasti q - ja $p-q$ -pituisia vektoreita. Oletamme, että X on p -asteinen ja X_1 q -asteinen matriisi. Selittävät muuttujat X_2 ovat tarpeettomia, jos $\beta_2 = 0$, jolloin on järkevämpää käyttää mallia $y = X_1 \beta_1 + \varepsilon$.

Sovitettu arvo $\hat{y}_1 = X_1(X_1^T X_1)^{-1} X_1^T y$ on vasteen y ortogonaaliprojektio. Jäännösvektori $y - \hat{y}_1 = \{I_n - X_1(X_1^T X_1)^{-1} X_1^T\}y$ koostuu kahdesta ortogonaalivektorista $y - \hat{y}$ ja $\hat{y} - \hat{y}_1$ seuraavasti:

$$y - \hat{y}_1 = (y - \hat{y}) + (\hat{y} - \hat{y}_1),$$

missä $(y - \hat{y})^T(\hat{y} - \hat{y}_1) = 0$. Vektori $y - \hat{y}$ on monimutkaisemman mallin jäännös ja vektori $\hat{y} - \hat{y}_1$ kertoo muutoksen sovitetuissa arvoissa, kun X_2 lisätään design-matriisiin. Edellä mainittujen vektoreiden ollessa normaalisti jakautuneen vektorin y ortogonaalisia lineaarifunktioita, ovat ne toisistaan riippumattomia. Pythagoraan lauseesta seuraa, että

$$(y - \hat{y}_1)^T(y - \hat{y}_1) = (y - \hat{y})^T(y - \hat{y}) + (\hat{y} - \hat{y}_1)^T(\hat{y} - \hat{y}_1),$$

joka on vastaavasti

$$SS(\hat{\beta}_1) = SS(\hat{\beta}) + \{SS(\hat{\beta}_1) - SS(\hat{\beta})\}.$$

Täten jäännöseliösumma yksinkertaisemmalle mallille koostuu kahdesta riippumattomasti jakautuneesta osasta, jotka ovat monimutkaisemman mallin jäännöseliösumma $SS(\hat{\beta})$ ja eliösummien erotus $SS(\hat{\beta}_1) - SS(\hat{\beta})$, kun matriisiin X_2 sarakkeet lisätään design-matriisiin. Jos yksinkertaisempi malli on oikea, niin myös monimutkaisempi malli on, sillä β_2 saa tarkan arvon nolla. Tässä tapauksessa $SS(\hat{\beta}_1) \sim \sigma^2 \chi_{n-q}^2$ ja $SS(\hat{\beta}) \sim \sigma^2 \chi_{n-p}^2$. Koska $SS(\hat{\beta}_1) - SS(\hat{\beta})$ on riippumaton jäännöseliösummasta $SS(\hat{\beta})$, niin kun $\beta_2 = 0$, noudattaa $SS(\hat{\beta}_1) - SS(\hat{\beta})$ jakaumaa $\sigma^2 \chi_{p-q}^2$. Näin ollen

$$F = \frac{\{SS(\hat{\beta}_1) - SS(\hat{\beta})\}/(p-q)}{SS(\hat{\beta})/(n-p)} \sim F_{p-q, n-p}. \quad (1)$$

Jos β_2 poikkeaa nolasta, on eliösummien erotus suurempi kuin jos $\beta_2 = 0$. Näin ollen, jos $\beta_2 \neq 0$, niin F saa suuria arvoja koskien $F_{p-q, n-p}$ -jakaumaa. F -testiä käyttämällä voidaan siten testata yksinkertaisemman mallin riittävyyttä suurien arvojen osoittaessa, että $\beta_2 \neq 0$. F -arvo perustuu kahden mallin uskottavuussuhteen vertailuun. Kun X_2 koostuu yhdestä kovariaatista, on β_2 skalaari ja testit sekä luottamusvälit sille saadaan sovittamalla monimutkaisempi malli ja laskemalla $T = (\hat{\beta}_2 - \beta_2)/(s\nu_{rr}^{1/2})$. Tässä s^2 on parametrin σ^2 estimaatti monimutkaisemmasta mallista tunnusluvun T nollahypoteesin ollessa voimassa. Tällöin testattaessa hypoteesia $\beta_2 = 0$ saadaan, että $F = T^2 = \hat{\beta}_2^2/(s^2\nu_{rr})$.

4 Aineiston analyysi

4.1 Mallien vertailu

Tämän työn tekninen toteutus on suoritettu R-ohjelmistolla (R Core Team, 2013). Tarkoituksena on löytää malli, jota käytetään ennustamaan kertoimille tehtyjä logit-muunnoksia eri tulosvaihtoehdoille kunkin tutkimuksessa mukana olevan vedonlyöntiyhtiön kohdalla. Jokaiselle tulosvaihtoehdolle etsitään parhaiten sopiva malli. Jos kuitenkin käy niin, että lopulliset mallit eivät selittävien muuttujien suhteen juurikaan poikkea toisistaan, niin harkitaan samojen muuttujien käyttämistä jokaisen tulosvaihtoehdon kohdalla. Ensin tutkitaan kaikista yksinkertaisinta mallia, jossa vasteena on kertoimesta saadulle todennäköisyyssarviolle tehty muunnos ja selittävänä muuttujana on vedonlyöntiyhtiö. Tämän jälkeen malliin otetaan yksitellen mukaan selittäviksi muuttujiksi sekä ottelun arvioitu tasaisuus että ottelun odotettu maalimäärä.

Varianssianalyysiin perustuvalla testillä (1) saadaan selville, onko näiden kahden selittävän muuttujan mukanaolo mallissa kannattavaa vai ei. Tämä testataan jokaiselle tulosvaihtoehdolle erikseen. Testin tuloksista huomataan, että vain tulosvaihtoehtojen 2-1 ja 3-0 kohdalla parhaiten sopiva malli olisi sellainen, jossa selittävinä muuttujina ovat vain vedonlyöntiyhtiö ja ottelun arvioitu tasaisuus. Suosikin 2-1-voiton tapauksessa maalimäärä-muuttujan lisääminen ei testin perusteella ($p = 0.092$) ole välttämätöntä. Suosikin 3-0-voitossa testi ($p = 0.682$) tukee vielä heikommin maalimäärä-muuttujan lisäämistä malliin. Muiden tulosvaihtoehtojen kohdalla paras malli sisältää myös maalimäärä-muuttujan.

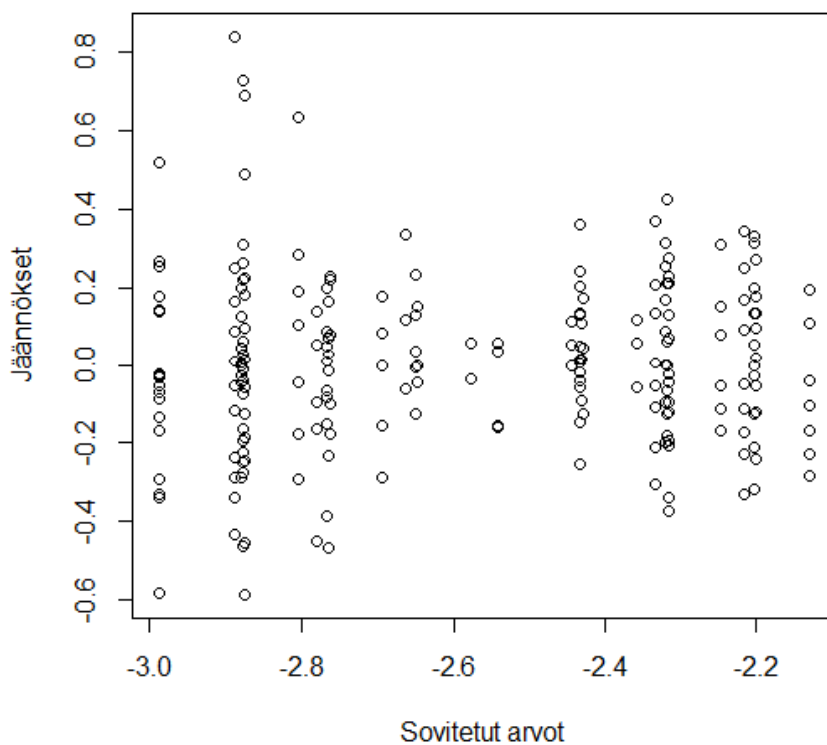
Tämän jälkeen otetaan pohdintoihin mukaan mahdolliset interaktiot. Varianssianalyysiin perustuen todetaan, että ainoastaan tulosvaihtoehdon 3-2 kohdalla vedonlyöntiyhtiön ja odotetun maalimäärän välinen interaktio tuo malliin sen verran lisäarvoa ($p = 0.013$), että kyseisen interaktion mukaanottamista kannattaa harkita. Muiden tulosvaihtoehtojen kohdalla ei mikään selittävien muuttujien välinen interaktio paranna mallia niin, että interaktioita olisi mielekästä sisällyttää malliin. Näin ollen suurimmalle osalle tulosvaihtoehdoista saadaan malli, jossa selittävinä muuttujina ovat vedonlyöntiyhtiö, ottelun arvioitu tasaisuus ja ottelun odotettu maalimäärä. Tulosvaihtoehdolle 3-2 malliin voidaan lisätä vedonlyöntiyhtiön ja odotetun maalimäärän välinen interaktio. Tulosvaihtoehdoille 2-1 ja 3-0 puolestaan parhaiten sopii malli, jossa selittävinä muuttujina ei ole kuin vedonlyöntiyhtiö ja ottelun ennakoitu tasaisuus.

4.2 Mallien diagnostiikka

Lineaarisen regressiomallin käytettävyyden pohjautuu siihen, että normaalisuusoletukset ovat voimassa. Eroavaisuudet aineiston ja mallin välillä voivat olla joko yksittäistapauksia tai systemaattisia tai sekä että. Yksittäiset erot voivat johtua esimerkiksi poikkeavista havainnoista. Systemaattista eroa esiintyy puolestaan erityisesti silloin, kun joko vasteelle tai kovariaatille joudutaan tekemään muunnoksia, korreloivat virhetermit oletetaan virheellisesti riippumattomiksi tai jokin termi on jätetty virheellisesti pois. Tämänäylyisiä ongelmia voidaan tutkia monin eri tavoin. Jäännöksen graafinen tarkastelu on usein käytetty metodi. Lineaarisen regressiomallin taustalla on joitain oletuksia, joiden tulee olla voimassa, jotta tulokset eivät ole harhaisia. Ensinnäkin vastemuuttujan täytyy olla line-

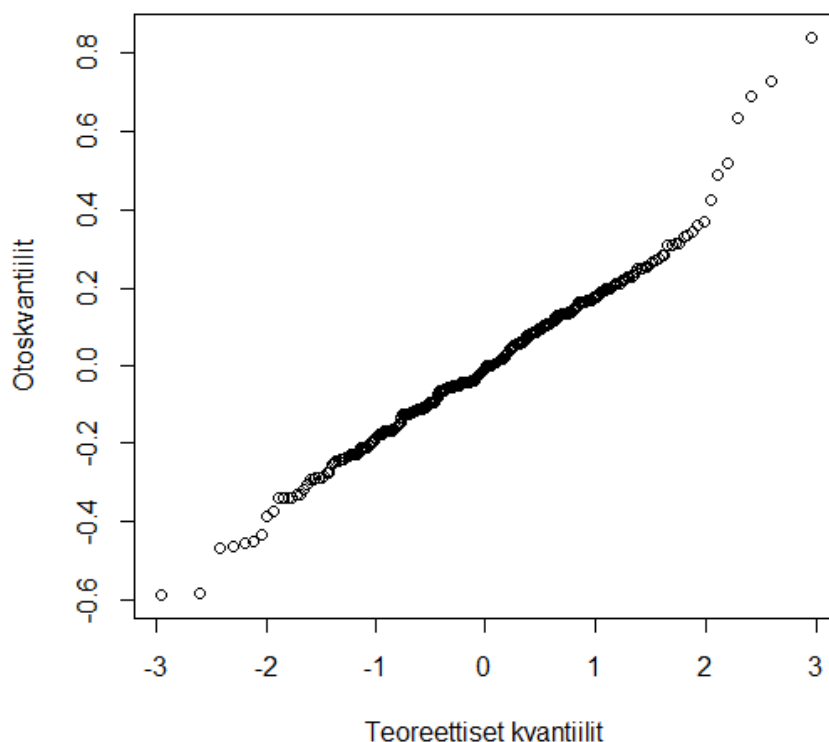
aarisesti riippuva jokaisesta selittävästä muuttujasta. Toiseksi varianssin tulee olla samansuuruista riippumatta vasteen arvoista. Tämän lisäksi jäännösten on oltava keskenään riippumattomia ja normaalisti jakautuneita. (Davison, 2003)

Seuraavassa käydään läpi joitain tulosvaihtoehtoja oletusten voimassaolon suhteen. Käytännössä oletusten voimassaoloa tarkastellaan seuraavalla tavalla. Kuvassa 1 on tulosvaihtoehdon 0–0 jäännökset ja sovitetut arvot. Kuvasta nähdään, että pisteet ovat jakautuneet melko satunnaisesti keskiviivan molemmille puolille, joten riippuvuutta ei juurikaan ole. Tarkasti katsottuna aivan kuvan vasemmassa yläkulmassa on muutama piste, jotka ovat hieman muita kauempana nollatasosta. Tämä selittyy alkulohkon päätöskierroksen ottelulla USA–Saksa, jossa molemmat joukkueet olisivat varmistaneet tasapelillä jatkoonpääsytään ja näin ollen ottelu poikkesi muista otteluista lähtökohdiltaan. Tämä luonnollisesti vaikutti sekä vedonlyöntiyhtiöiden todennäköisyysarvioihin että Veikkauksen pelaajien pelaamiseen.



Kuva 1: Tulosvaihtoehdon 0–0 jäännösten ja sovitettujen arvojen yhteisjakauma.

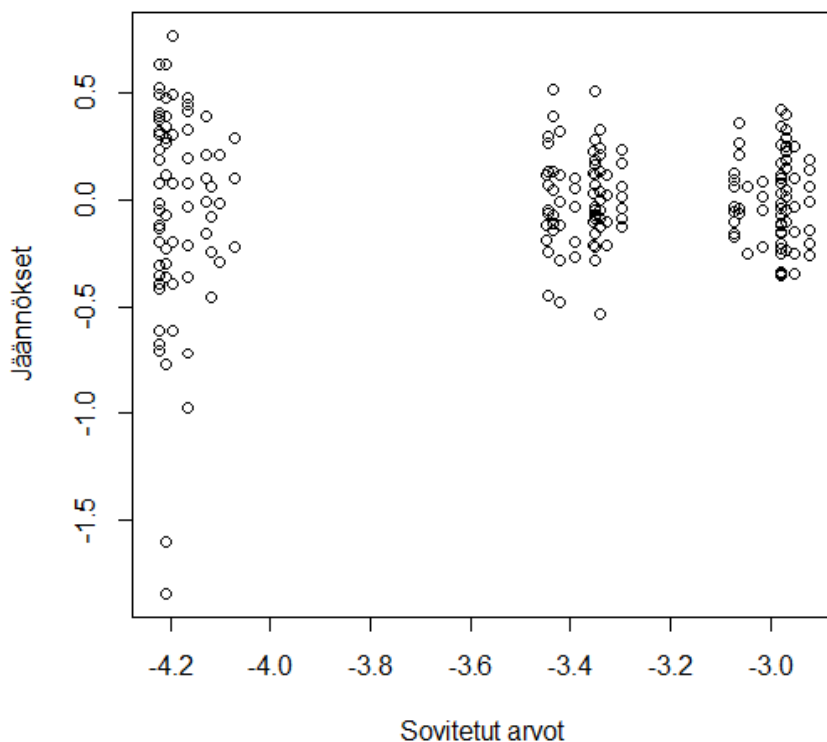
Kuvasta 2 nähdään, kuinka hyvin jäännökset noudattavat normaalijakaumaa. Pisteet poikkeavat suoralta jonkin verran molemmissa päissä, mutta pääosin ne kulkevat suoralla. Tulosvaihtoehdon 0–0 kohdalla voidaan siten todeta normaalisuusoletuksien pätevän.



Kuva 2: Tulosvaihtoehdon 0–0 jäännösten normaalisuus.

Seuraavana tarkastellaan altavastajaan 0–2-voittoa. Kuvasta 3 nähdään, että tässäkin tapauksessa sovitteiden ja jäännösten välillä ei ole havaittavissa riippuvuutta. Pisteet jakautuvat satunnaisesti keskiviivan molemmille puolille muutamaa poikkeavaa havaintoa lukuun ottamatta. Tässä tapauksessa eniten nollatasosta poikkeavat pisteet ovat aivan kuvan vasemmassa alakulmassa. Nämä johtuvat otteluista Argentiina–Iran ja Kamerun–Brasilia. Näissä kahdessa ottelussa joukkueiden välinen ennakoitu tasoero oli huomattavasti suurempi kuin missään muussa MM-kisojen ottelussa. Näin ollen näiden otteluiden kohdalla altavastajaan 0–2-voitto oli erittäin epätodennäköinen, mikä aiheutti poikkeavia havaintoja Kuvaan 3. Myös muiden tulosvaihtoehtojen kohdalla nämä kaksi ottelua aiheuttavat hieman hajontaa vastaavissa tarkasteluissa, mutta eivät läheskään yhtä voimakasta kuin altavastajaan 0–2-voiton kohdalla. Esimerkiksi Kuvassa 1 kauimpana alhaalla olevat pisteet kuuluvat jompaankumpaan näistä kahdesta ottelusta. Kuvassa 3 puolestaan nollatason yläpuolella olevat kauim-

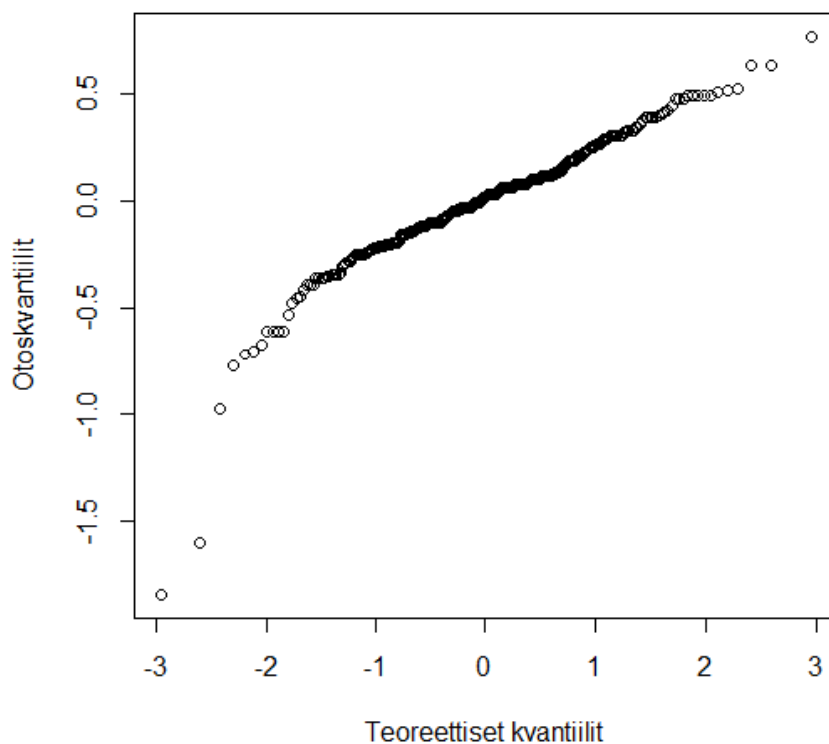
mat pisteet johtuvat tulosvaihtoehdon 0–0 tavoin ottelusta USA–Saksa.



Kuva 3: Altavastajaan 0–2-voiton jäännösten ja sovitettujen arvojen yhteisjakauma.

Kuvasta 4 nähdään tulosvaihtoehdon 0–2 kohdalla, kuinka hyvin jäännökset noudattavat normaalijakaumaa. Monien muiden tulosvaihtoehtojen tapaan myös tässä pisteet hieman karkaavat suoralta molemmissa päissä, mutta normalisuusetusten voidaan silti sanoa olevan kunnossa.

Muiden tulosvaihtoehtojen jäännöstarkastelut antavat hyvin samanlaisia tuloksia kuin edellä esitettyjen kahden tulosvaihtoehdon kuvat näyttävät. Tämän vuoksi niitä ei esitetä tässä työssä sen tarkemmin. Osalla tulosvaihtoehtoja kuvat näyttävät jopa hieman paremmilta oletusten voimassaoloa ajatellen. Suurimmat poikkeamat kuvissa aiheutuvat samoista kolmesta ottelusta, jotka erottuivat jonkin verran tulosvaihtoehtojen 0–0 ja 0–2 kohdalla. Kaiken kaikkiaan normalisuusetusten voidaan sanoa olevan kunnossa jokaisen tulosvaihtoehdon kohdalla.



Kuva 4: Altavastajaan 0–2-voiton jäännösten normalisuus.

4.3 Tulokset tulosvaihtoehtoin

Tulosvaihtoehdot analysoidaan yksi kerrallaan. Kaikille tulosvaihtoehdoille käytettävä malli on muotoa

$$y_{ijkl} = \beta_0 + \beta_{1i} + \beta_{2j} + \beta_{3k} + \varepsilon_{ijkl}, \quad (2)$$

missä $i = 1, 2, 3, 4, 5$ viittaa yhtiöön, $j = 1, 2, 3$ tasoeroon ja $k = 1, 2$ maalimäärään ja $l = 1, \dots, n$, missä n on otteluiden määrä. Tulosvaihtoehtojen 2–1 ja 3–0 kohdalla sovitetaan myös malli, jossa ei maalimäärä-muuttujaa ole ollenkaan. Suosikin 3–2-voiton kohdalla puolestaan sovitetaan ylläolevan mallin lisäksi malli, jossa on mukana myös yhtiön ja maalimäärän välinen interaktio-termi.

Suosikin 1–0-voiton kohdalla vakiotermin estimaatti on -2.009 ja keskivirhe 0.023. Vertailutilanteessa yhtiönä on Veikkaus, arvioitu tasoero on suuri ja odotettu maalimäärä on pieni. Tällöin $\beta_{1i} = 0$, $\beta_{2j} = 0$ ja $\beta_{3k} = 0$. Näihin parametrien arvoihin vertautuvat muiden yhtiöiden sekä tasoero- ja maalimääräluokkien parametrien estimaatit. Taulukosta 8 nähdään nämä parametrien estimaatit eri yhtiöille, tasoeroille ja maalimäärälle suosikin 1–0-voiton kohdalla.

Taulukko 8: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen suosikin 1–0-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.154 (0.023)	< 0.001
Ladbrokes (β_{13})	0.123 (0.023)	< 0.001
Unibet (β_{14})	0.167 (0.023)	< 0.001
Nordicbet (β_{15})	0.287 (0.023)	< 0.001
Tasoero 2 (β_{22})	-0.146 (0.019)	< 0.001
Tasoero 3 (β_{23})	-0.251 (0.020)	< 0.001
Maalimäärä 2 (β_{32})	-0.301 (0.017)	< 0.001

Taulukon 8 lukujen tarkempi tulkinta ei ole kovin mielekästä, sillä ne eivät suoraan kerro muuta kuin sen minkäsuuntainen ero Veikkaukseen kullakin yhtiöllä on sekä erojen suhteellisen suuruuden. Mitä kauempana luku on nollassa, niin sitä suurempi on ero Veikkaukseen. Jokaisen ulkomaisen vedonlyöntiyhtiön kohdalla ero on tilastollisesti merkitsevä. Tässä työssä käytetään tilastollisen merkitsevyyden rajana 0.05 riskitasoa. Myös tasoero- ja maalimäärä-muuttujien kohdalla p-arvot ovat selkeästi alle 0.05. Tämä merkitsee sitä, että suosikin 1–0-voiton tapauksessa sekä tasoerojen että maalimäärien eri luokilla on eroja keskenään. Nämä muuttujat eivät kuitenkaan sen erityisemmin ole tässä työssä kiinnostuksen kohteena, joten niiden analysointiin ei syvennyttä sen tarkemmin.

Erojen konkretisoimiseksi havainnollistetaan tämän tulostulovaihtoehdon kohdalla, miten eri vedonlyöntiyhtiöiden todennäköisyysarviot poikkeavat toisistaan. Tämä toteutetaan palauttamalla muunnosten arvot todennäköisyyksiksi, jotka kertovat suoraan erojen todellisen suuruuden. Suoritetaan tämä ottelulle, jossa arvioitu tasoero on suuri ja odotettu maalimäärä on pieni. Tällöin Veikkauksen pelaajien todennäköisyysarvion estimaatti saadaan ratkaisemalla α yhtälöstä $\log\left(\frac{\alpha}{1-\alpha}\right) = -2.009$. Tästä saadaan, että $\alpha = 0.118$. Muiden yhtiöiden kohdalla lisätään niitä vastaavan parametrin estimaatti vakiotermin estimaattiin. Näin ollen esimerkiksi William Hillin todennäköisyysarvion estimaatti α saadaan yhtälöstä $\log\left(\frac{\alpha}{1-\alpha}\right) = -2.009 + 0.154$. Tästä saadaan, että $\alpha = 0.135$. Täten William Hillin todennäköisyysarvio on Veikkaukseen verrattuna 1.7 prosenttiyksikköä suurempi.

Vertailtaessa muita yhtiöitä Veikkaukseen voidaan nähdä, että suosikin 1–0-voitolle saadaan käytettävällä mallilla selkeitä viitteitä siihen suuntaan, että Veikkauksen pelaajien ja muiden vedonlyöntiyhtiöiden todennäköisyysarvioissa on eroa. Veikkauksen pelaajat arvioivat ottelusta riippumatta suosikin 1–0-voiton todennäköisyyden pienemmäksi kuin ulkomaiset vedonlyöntiyhtiöt. Tämä näkyy suoraan Taulukosta 9 ja Taulukossa 8 se nähdään siinä, että jokaisen yhtiön kohdalla parametrin estimaatti on positiivinen luku.

Taulukko 9: Parametrien estimaatit ja muunnosten arvot yhtiöittäin sekä niiden perusteella lasketut todennäköisyysarviot suosikin 1–0-voiton tapauksessa, kun kyseessä on ottelu, jossa arvioitu tasoero on suuri ja odotettu maalimäärä pieni.

	parametrin (β_{1i}) estimaatti	$\log\left(\frac{\alpha}{1-\alpha}\right)$	tn-arvio (α)
Veikkaus	0.000	-2.009	0.118
William Hill	0.154	-1.855	0.135
Ladbrokes	0.123	-1.886	0.132
Unibet	0.167	-1.842	0.137
Nordicbet	0.287	-1.722	0.152

Tulosvaihtoehdon 2–0 kohdalla tulokset ovat lähestulkoon identtiset 1–0-tuloksen tulosten kanssa. Tässäkin tapauksessa Veikkauksen pelaajat arvioivat suosikin 2–0-voiton todennäköisyyden pienemmäksi kuin ulkomaiset yhtiöt. Erot Veikkauksen ja muiden vedonlyöntiyhtiöiden välillä eivät ole kuitenkaan yhtä selkeät kuin 1–0-tuloksen tapauksessa. Suosikin 2–0-voiton kohdalla vakiotermin estimaatti on -2.006 ja keskivirhe 0.026. Taulukossa 10 on esitetty tulokset suosikin 2–0-voitolle.

Taulukko 10: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen suosikin 2–0-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.063 (0.025)	0.013
Ladbrokes (β_{13})	0.061 (0.025)	0.015
Unibet (β_{14})	0.078 (0.025)	0.002
Nordicbet (β_{15})	0.187 (0.025)	< 0.001
Tasoero 2 (β_{22})	-0.354 (0.021)	< 0.001
Tasoero 3 (β_{23})	-0.652 (0.022)	< 0.001
Maalimäärä 2 (β_{32})	-0.140 (0.018)	< 0.001

Tulosvaihtoehtojen 2–1 ja 3–0 tapauksissa analyysit tehdään kahdella eri mallilla. Toinen käytettävä malli on muuten samanlainen kuin malli, jota käytetään kaikkien tulosvaihtoehtojen kohdalla, mutta siinä ei ole mukana maalimäärämuuttujaa. Suosikin 2–1-voiton kohdalla vakiotermin estimaatti on -2.423 ja keskivirhe 0.021, kun käytetään mallia, jossa maalimäärämuuttuja on mukana. Tällä mallilla saadaan suosikin 2–1-voitolle samansuuntaisia tuloksia suosikin 1–0- ja 2–0-voittojen kanssa. Nyt erot Veikkauksen ja muiden yhtiöiden välillä eivät kuitenkaan ole yhtä selviä. Itse asiassa Veikkauksen ja Ladbrokesin välillä ei edes ole tilastollisesti merkitsevää eroa. Muiden yhtiöiden kanssa erot ovat kuitenkin niin suuria, että ne ovat tilastollisesti merkitseviä. Tulosvaihtoehdon 2–1 tapauksessakin siis Veikkauksen pelaajien arviot lopputuloksen todennäköisyydestä ovat pienempiä kuin kolmella muulla ulkomaisella yhtiöllä. Ainoastaan Ladbrokesin ja Veikkauksen välillä eroa ei ole.

Suosikin 3–0-voiton kohdalla saadaan vakiotermin estimaatiksi -2.418 ja keskivirheeksi 0.038. Tässä tapauksessa eroja Veikkauksen ja muiden yhtiöiden välillä ei juurikaan ole. Ainoastaan Unibetiin verrattuna Veikkauksen pelaajien arvio lopputuloksen todennäköisyydestä on pienempi. Näin ollen Veikkauksen pelaajien pelaaminen noudattaa pitkälti samaa linjaa muiden yhtiöiden todennäköisyysarvioiden kanssa. Tulokset suosikin 2–1- ja 3–0-voitoille on esitetty Taulukoissa 11 ja 12.

Taulukko 11: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen suosikin 2–1-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.059 (0.020)	0.004
Ladbrokes (β_{13})	-0.013 (0.020)	0.523
Unibet (β_{14})	0.044 (0.020)	0.031
Nordicbet (β_{15})	0.073 (0.020)	< 0.001
Tasoero 2 (β_{22})	0.034 (0.017)	0.045
Tasoero 3 (β_{23})	-0.069 (0.018)	< 0.001
Maalimäärä 2 (β_{32})	0.025 (0.015)	0.092

Taulukko 12: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen suosikin 3–0-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.039 (0.037)	0.294
Ladbrokes (β_{13})	< 0.001 (0.037)	0.994
Unibet (β_{14})	0.096 (0.037)	0.010
Nordicbet (β_{15})	0.063 (0.037)	0.089
Tasoero 2 (β_{22})	-0.563 (0.031)	< 0.001
Tasoero 3 (β_{23})	-0.988 (0.033)	< 0.001
Maalimäärä 2 (β_{32})	0.011 (0.027)	0.681

Tulosvaihtoehtojen 2–1 ja 3–0 kohdalla voidaan käyttää myös mallia, joka on muotoa

$$y_{ijl} = \beta_0 + \beta_{1i} + \beta_{2j} + \varepsilon_{ijl}, \quad (3)$$

missä erona aikaisemmin käytettyyn malliin on maalimäärä-muuttujan puuttuminen. Tällä mallilla vakiotermin estimaatiksi saadaan 2–1-tulokselle -2.402 ja keskivirheeksi 0.017. Suosikin 3–0-voiton tapauksessa vakiotermin estimaatti on -2.409 ja keskivirhe 0.031. Kummankaan tulosvaihtoehdon kohdalla maalimäärä-muuttujan poistaminen mallista ei vaikuta eroihin Veikkauksen ja muiden yhtiöiden välillä, vaan tulkinnat ovat täysin samanlaiset kuin aiemmin. Taulukois-

sa 13 ja 14 on esitetty tulokset suosikin 2–1- ja 3–0-voitoille, kun on käytetty mallia, jossa ei ole maalimäärä-muuttujaa mukana.

Taulukko 13: Mallin (3) parametrien estimaatit keskivirheineen ja p-arvoineen suosikin 2–1-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.059 (0.020)	0.004
Ladbrokes (β_{13})	-0.013 (0.020)	0.524
Unibet (β_{14})	0.044 (0.020)	0.032
Nordicbet (β_{15})	0.073 (0.020)	< 0.001
Tasoero 2 (β_{22})	0.023 (0.016)	< 0.001
Tasoero 3 (β_{23})	-0.083 (0.016)	< 0.001

Taulukko 14: Mallin (3) parametrien estimaatit keskivirheineen ja p-arvoineen suosikin 3–0-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.039 (0.037)	0.293
Ladbrokes (β_{13})	< 0.001 (0.037)	0.994
Unibet (β_{14})	0.096 (0.037)	0.010
Nordicbet (β_{15})	0.063 (0.037)	0.089
Tasoero 2 (β_{22})	-0.567 (0.029)	< 0.001
Tasoero 3 (β_{23})	-0.995 (0.029)	< 0.001

Suosikin 3–1-voiton kohdalla vakiotermin estimaatiksi saadaan -2.599 ja keskivirheeksi 0.027. Tuloksista nähdään, että Veikkauksen pelaajat arvioivat suosikin 3-1-voiton todennäköisyyden selvästi suuremmaksi kuin ulkomaiset yhtiöt. Erot ovat jokaisen yhtiön kohdalla tilastollisesti merkitsevät. Tulosvaihtoehdon 3–2 kohdalla tulokset ovat hyvin samankaltaiset tulosvaihtoehdon 3–1 kanssa. Myös tämän lopputuloksen todennäköisyydet Veikkauksen pelaajat arvioivat selvästi suuremmaksi kuin ulkomaiset yhtiöt. Erot ovat jokaisen yhtiön kohdalla tilastollisesti merkitsevät. Suosikin 3–2-voiton tapauksessa vakiotermin estimaatti on -3.774 ja keskivirhe 0.037. Tulokset suosikin 3-1-voitolle on esitetty Taulukossa 15 ja Taulukossa 16 on esitetty tulokset suosikin 3–2-voitolle, kun käytössä on malli, jota käytetään kaikkien tulosvaihtoehtojen kohdalla.

Taulukko 15: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen suosikin 3–1-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	-0.216 (0.026)	< 0.001
Ladbrokes (β_{13})	-0.298 (0.026)	< 0.001
Unibet (β_{14})	-0.277 (0.026)	< 0.001
Nordicbet (β_{15})	-0.252 (0.026)	< 0.001
Tasoero 2 (β_{22})	-0.190 (0.022)	< 0.001
Tasoero 3 (β_{23})	-0.433 (0.023)	< 0.001
Maalimäärä 2 (β_{32})	0.174 (0.019)	< 0.001

Taulukko 16: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen suosikin 3–2-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	-0.101 (0.037)	0.006
Ladbrokes (β_{13})	-0.199 (0.037)	< 0.001
Unibet (β_{14})	-0.143 (0.037)	< 0.001
Nordicbet (β_{15})	-0.242 (0.037)	< 0.001
Tasoero 2 (β_{22})	0.177 (0.031)	< 0.001
Tasoero 3 (β_{23})	0.099 (0.032)	0.002
Maalimäärä 2 (β_{32})	0.319 (0.027)	< 0.001

Vaihtoehtoinen malli, jota voidaan käyttää tulosvaihtoehdolle 3–2, on muotoa

$$y_{ijkl} = \beta_0 + \beta_{1i} + \beta_{2j} + \beta_{3k} + \beta_{1i} : \beta_{3k} + \varepsilon_{ijkl}, \quad (4)$$

missä $\beta_{1i} : \beta_{3k}$ on yhtiön ja maalimäärän välinen interaktiotermin. Tällä mallilla vakiotermin estimaatti on -3.709 ja keskivirhe 0.043. Taulukosta 16 nähdään, että kun odotettu maalimäärä on pieni, niin mallin antamat tulokset ovat samankaltaisia kuin aiemmin käytetyn mallin tapauksessa. Toisin sanoen Veikkauksen pelaajat pelaavat järjestään tätä tulosvaihtoehtoa enemmän kuin niiden todennäköisyydet ovat muiden vedonlyöntiyhtiöiden arvioiden mukaan, kun tarkastelun kohteena on ennakkoon vähämaalinen ottelu. Taulukossa 17 on esitetty tulokset suosikin 3–2-voitolle, kun käytetään mallia, joka sisältää yhtiön ja maalimäärän välisen interaktion ja ottelun odotettu maalimäärä on pieni.

Taulukko 17: Mallin (4) parametrien estimaatit keskivirheineen ja p-arvoineen suosikin 3–2-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	-0.188 (0.050)	< 0.001
Ladbrokes (β_{13})	-0.261 (0.050)	< 0.001
Unibet (β_{14})	-0.264 (0.050)	< 0.001
Nordicbet (β_{15})	-0.300 (0.050)	< 0.001
Tasoero 2 (β_{22})	0.177 (0.030)	< 0.001
Tasoero 3 (β_{23})	0.099 (0.032)	0.002
Maalimäärä 2 (β_{32})	0.184 (0.053)	< 0.001
William Hill:Maalimäärä 2 ($\beta_{12};\beta_{32}$)	0.179 (0.072)	0.014
Ladbrokes:Maalimäärä 2 ($\beta_{13};\beta_{32}$)	0.128 (0.072)	0.078
Unibet:Maalimäärä 2 ($\beta_{14};\beta_{32}$)	0.249 (0.072)	< 0.001
Nordicbet:Maalimäärä 2 ($\beta_{15};\beta_{32}$)	0.121 (0.072)	0.094

Edellisestä taulukosta ei suoraan selviä, onko Veikkauksen ja muiden yhtiöiden välillä eroa myös tilanteessa, jossa odotettu maalimäärä on suuri. Tämän selvittämiseksi tutkitaan tarkemmin tilannetta, kun maalimäärä-muuttuja saa arvon 2. Tilanteessa, jossa odotettu maalimäärä on suuri, saadaan estimaatit jokaiselle vedonlyöntiyhtiölle Taulukosta 17 lisäämällä kunkin yhtiön estimaattiin kyseisen yhtiön ja maalimäärän välisen interaktion estimaatti. Jotta saadaan selville, onko tässä tilanteessa muiden yhtiöiden ja Veikkauksen välillä eroa, käytetään Waldin testiä (Harrell, 2001). Testin avulla saadaan selville, että odotetun maalimäärän ollessa suuri, vain Ladbrokesin ja Nordicbetin kohdalla on eroa Veikkaukseen. Tämä tarkoittaa sitä, että Ladbrokesiin ja Nordicbetiin verrattuna Veikkauksen pelaajat pelaavat suosikin 3–2-voittoa enemmän kuin näiden yhtiöiden mukaan tulisi pelata. Sen sijaan William Hillin ja Unibetin kohdalla näin tapahtuu vain silloin, kun kyseessä on ennakkoon vähämaalinen ottelu. Ennakkoon runsasmaalisen ottelun kohdalla näiden kahden yhtiön todennäköisyysarvioissa ei ole tilastollisesti merkitsevää eroa verrattuna Veikkauksen pelaajien pelaamiseen. Taulukossa 18 on esitetty estimaatit ja p-arvot, kun käytetään mallia, jossa on yhtiön ja maalimäärän välinen interaktio mukana ja ottelun odotettu maalimäärä on suuri.

Taulukko 18: Mallin (4) yhtiökohtaiset parametrien estimaatit keskivirheineen ja p-arvoineen suosikin 3-2-voiton kohdalla odotetun maalimäärän ollessa suuri.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	-0.009 (0.050)	0.861
Ladbrokes (β_{13})	-0.134 (0.050)	0.010
Unibet (β_{14})	-0.015 (0.050)	0.779
Nordicbet (β_{15})	-0.179 (0.050)	< 0.001

Tässä työssä tutkittavien tasapelitulovaihtoehtojen kohdalla on myös joi-
tain eroja Veikkauksen pelaajien ja ulkomaisten vedonlyöntiyhtiöiden todennä-
köisyysarvioissa. Ainoastaan 2–2-tuloksen kohdalla eroja ei ole suuntaan eikä
toiseen. Sen sijaan 0–0- ja 1–1-tulosten kohdalla eroja on ja ne ovat tilastol-
lisesti merkitseviä. Molemmilla tulovaihtoehtoilla erot ovat samansuuntaisia,
eli näissä kummassakin tapauksessa Veikkauksen pelaajat arvioivat lopputulos-
ten todennäköisyydet pienemmiksi kuin ulkomaiset yhtiöt. Tulovaihtoehdon 0–
0 kohdalla vakiotermin estimaatti on -2.541 ja keskivirhe 0.036. Lopputuloksen
1–1 tapauksessa vakiotermin estimaatti on -2.506 ja keskivirhe 0.031. Tulovaih-
toehdon 2–2 kohdalla vakiotermin estimaatiksi saadaan -3.503 ja keskivirheeksi
0.042. Taulukoissa 19, 20 ja 21 on esitetty tulokset jokaiselle tasapelitulokselle.

Taulukko 19: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen
tulovaihtoehdon 0–0 tapauksessa. Vertailukohteena on ottelu, jossa arvioitu
tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.184 (0.036)	< 0.001
Ladbrokes (β_{13})	0.099 (0.036)	0.006
Unibet (β_{14})	0.114 (0.036)	0.002
Nordicbet (β_{15})	0.111 (0.036)	0.002
Tasoero 2 (β_{22})	0.109 (0.030)	< 0.001
Tasoero 3 (β_{23})	0.226 (0.031)	< 0.001
Maalimäärä 2 (β_{32})	-0.447 (0.026)	< 0.001

Taulukko 20: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen
tulovaihtoehdon 1–1 tapauksessa. Vertailukohteena on ottelu, jossa arvioitu
tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.200 (0.030)	< 0.001
Ladbrokes (β_{13})	0.139 (0.030)	< 0.001
Unibet (β_{14})	0.163 (0.030)	< 0.001
Nordicbet (β_{15})	0.221 (0.030)	< 0.001
Tasoero 2 (β_{22})	0.266 (0.025)	< 0.001
Tasoero 3 (β_{23})	0.361 (0.026)	< 0.001
Maalimäärä 2 (β_{32})	-0.106 (0.022)	< 0.001

Taulukko 21: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen tulosvaihtoehdon 2–2 tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.026 (0.041)	0.524
Ladbrokes (β_{13})	-0.047 (0.041)	0.260
Unibet (β_{14})	0.071 (0.041)	0.089
Nordicbet (β_{15})	-0.057 (0.041)	0.167
Tasoero 2 (β_{22})	0.381 (0.034)	< 0.001
Tasoero 3 (β_{23})	0.462 (0.036)	< 0.001
Maalimäärä 2 (β_{32})	0.207 (0.030)	< 0.001

Veikkauksen pelaajien todennäköisyysarviot ovat muiden yhtiöiden arvioita pienempiä myös altavastajaan 0–1-voitolle. Sen sijaan tulosvaihtoehdoissa, joissa altavastaja voittaa ottelun 0–2 tai 1–2, ei ole tilastollisesti merkitsevää eroa Veikkauksen pelaajien ja ulkomaisten yhtiöiden todennäköisyysarvioiden kesken. Altavastajaan 0–1-voiton tapauksessa vakiotermin estimaatti on -3.275 ja keskivirhe 0.038. Altavastajaan 0–2-voiton kohdalla vakiotermin estimaatti on -4.128 ja keskivirhe 0.053. Altavastajaan 1–2-voiton tapauksessa vakiotermin estimaatiksi saadaan -3.558 ja keskivirheeksi 0.043. Tulokset näille tulosvaihtoehdoille on esitetty Taulukoissa 22, 23 ja 24.

Taulukko 22: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen altavastajaan 0–1-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.177 (0.037)	< 0.001
Ladbrokes (β_{13})	0.164 (0.037)	< 0.001
Unibet (β_{14})	0.184 (0.037)	< 0.001
Nordicbet (β_{15})	0.296 (0.037)	< 0.001
Tasoero 2 (β_{22})	0.512 (0.031)	< 0.001
Tasoero 3 (β_{23})	0.777 (0.033)	< 0.001
Maalimäärä 2 (β_{32})	-0.262 (0.027)	< 0.001

Taulukko 23: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen altavastajaan 0–2-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.025 (0.051)	0.624
Ladbrokes (β_{13})	-0.002 (0.051)	0.974
Unibet (β_{14})	0.055 (0.051)	0.283
Nordicbet (β_{15})	0.011 (0.051)	0.833
Tasoero 2 (β_{22})	0.776 (0.043)	< 0.001
Tasoero 3 (β_{23})	1.148 (0.045)	< 0.001
Maalimäärä 2 (β_{32})	-0.093 (0.037)	0.014

Taulukko 24: Mallin (2) parametrien estimaatit keskivirheineen ja p-arvoineen altavastajaan 1–2-voiton tapauksessa. Vertailukohteena on ottelu, jossa arvioitu tasoero on suuri, odotettu maalimäärä pieni ja käytetty yhtiö Veikkaus.

	parametrin estimaatti (s.e.)	p-arvo
William Hill (β_{12})	0.049 (0.042)	0.242
Ladbrokes (β_{13})	-0.032 (0.042)	0.440
Unibet (β_{14})	0.012 (0.042)	0.771
Nordicbet (β_{15})	0.021 (0.042)	0.611
Tasoero 2 (β_{22})	0.613 (0.035)	< 0.001
Tasoero 3 (β_{23})	0.875 (0.037)	< 0.001
Maalimäärä 2 (β_{32})	0.066 (0.031)	0.031

Näin ollen tässä tutkimuksessa käytetyn mallin perusteella peräti kuudella eri tulosvaihtoehdolla on sen suuntaista eroa havaittavissa, että Veikkauksen pelaajat arvioivat kyseisen lopputuloksen todennäköisyyden pienemmäksi kuin ulkomaisten vedonlyöntiyhtiöiden kertoimenasettajat. Tai ainakin Veikkauksen pelaajat pelaavat näitä tuloksia vähemmän kuin niitä tulisi pelata. Nämä tulosvaihtoehdot ovat suosikin 1–0-, 2–0- ja 2–1-voitot, 0–0- ja 1–1-tasapelit sekä altavastajaan 0–1-voitto. Suosikin 2–1-voittoa lukuun ottamatta näillä kaikilla tulosvaihtoehdoilla erot todennäköisyysarvioissa ovat tilastollisesti merkitsevät jokaisen vedonlyöntiyhtiön ja Veikkauksen pelaajien välillä. Suosikin 2–1-voitollakin erot ovat tilastollisesti merkitseviä Ladbrokesia lukuun ottamatta.

Tutkimuksessa mukana olleista tulosvaihtoehdoista kahdella erot ovat toiseen suuntaan. Nämä tulosvaihtoehdot ovat suosikin 3–1- ja 3–2-voitot. Näiden lopputulosten kohdalla Veikkauksen pelaajat pelaavat mallin perusteella enemmän kuin muiden vedonlyöntiyhtiöiden arvioiden mukaan tulisi pelata. Lopuille neljälle tulosvaihtoehdolle ei tässä käytetty malli eroja löydä, kun verrataan Veikkauksen pelaajien pelaamista ulkomaisten yhtiöiden todennäköisyysarvioihin. Nämä tulosvaihtoehdot ovat altavastajaan 0–2- ja 1–2-voitot, 2–2-tasapeli sekä suosikin 3–0-voitto. Ainoan poikkeuksen tähän tekee Unibet, kun kyseessä

on suosikin 3–0-voitto. Tällöin ero Veikkauksen pelaajien todennäköisyysarvioon on tilastollisesti merkitsevä. Tässä tapauksessa Veikkauksen pelaajat arvioivat todennäköisyyden pienemmäksi kuin Unibetin kertoimenasettajat.

Analysit suoritettiin kaikille tulosvaihtoehdoille myös ilman jäännöskuvis- sa hieman erottuneita otteluita. Kaikkien tulosvaihtoehtojen kohdalla tulkinnat pysyivät kuitenkin täysin samanlaisina, olivat nämä ottelut mukana tai ei. Näin ollen kaikki ottelut pidettiin tutkimuksessa mukana. Taulukossa 25 on esitetty saadut tulokset sen mukaan, miten eri tulosvaihtoehdot eroavat Veikkauksen pelaajien ja ulkomaisten vedonlyöntiyhtiöiden kesken. Mitä ylempänä tulosvaihtoehto taulukossa on, niin sitä vähemmän Veikkauksen pelaajat pelaavat kyseistä tulosta verrattuna muiden yhtiöiden todennäköisyysarvioihin. Ja kääntäen, mitä alempana tulosvaihtoehto taulukossa on, sitä enemmän Veikkauksen pelaajat sitä tulosta pelaavat suhteessa muiden yhtiöiden todennäköisyysarvioihin.

Taulukko 25: Tulosvaihtoehdot luokiteltuna sen mukaan, minkälaista Veikkauksen pelaajien pelaaminen niissä on verrattuna ulkomaisten vedonlyöntiyhtiöiden todennäköisyysarvioihin.

Tulosvaihtoehto	Lopputulema
1–0	Veikkauksen pelaajat
2–0	pelaavat
0–0	vähemmän kuin
1–1	vedonlyöntiyhtiöiden
0–1	todennäköisyysarviot
2–1	ovat
3–0	Veikkauksen pelaajien
0–2	pelaaminen samassa
1–2	linjassa vedonlyöntiyhtiöiden
2–2	todennäköisyysarvioiden kanssa
3–2	Veikkauksen pelaajat pelaavat enemmän kuin
3–1	vedonlyöntiyhtiöiden todennäköisyysarviot ovat

5 Yhteenveto

Tämän työn tavoitteena oli rakentaa yksinkertainen malli, jonka avulla voidaan ennustaa vedonlyöntiyhtiöiden todennäköisyysarvioita jalkapallo-otteluiden lopputuloksista sekä Veikkauksen pelaajien pelaamien rahojen jakautumista eri tulosvaihtoehdoille. Varsinaisena mielenkiinnon kohteena oli kuitenkin etsiä tämän mallin perusteella tulosvaihtoehtoja, joiden kohdalla vedonlyöntiyhtiöiden todennäköisyysarviot poikkesivat Veikkauksen pelaajien pelaamisesta. Lähes kaikkien tulosvaihtoehtojen kohdalla parhaaksi malliksi rakentui lopulta sellainen, jossa selittävinä muuttujina olivat vedonlyöntiyhtiön lisäksi ottelun ennakoitu tasaisuus sekä odotettu maalimäärä. Muuttujien välisiä interaktioita ei malliin sisälly, mikä kertoo siitä, että tulokset eivät juurikaan riipu siitä, kuinka epätasainen ja runsasmaalinen ottelu on etukäteen odotettavissa.

Saadut tulokset tulosvaihtoehtokohtaisista eroista Veikkauksen pelaajien pelaamisen ja vedonlyöntiyhtiöiden todennäköisyysarvioiden välillä olivat lopulta helposti jaettavissa kolmeen luokkaan. Tulosvaihtoehtoista suosikin 1-0-, 2-0- ja 2-1-voitot, 0-0- ja 1-1-tasapelit sekä altavastajaan 0-1-voitto olivat sellaisia, joissa havaittiin enemmän tai vähemmän eroa siihen suuntaan, että Veikkauksen pelaajat pelasivat näitä tuloksia vähemmän kuin niitä tulisi pelata. Suosikin 3-0-voiton, 2-2-tasapelin sekä altavastajaan 0-2- ja 1-2-voittojen kohdalla erot eivät olleet tilastollisesti merkitseviä. Kaksi muuta tutkielmassa käsiteltyä tulosvaihtoehtoa olivat suosikin 3-1- ja 3-2-voitot, joita Veikkauksen pelaajat puolestaan pääsääntöisesti pelasivat enemmän kuin niitä vedonlyöntiyhtiöiden todennäköisyysarvioiden perusteella tulisi pelata.

Tutkielman tuloksista voi tehdä erilaisia päätelmiä. Yksi merkille pantava seikka on, että vähämaaliset tulosvaihtoehdot ovat Veikkauksen pelaajien epäsuosiossa, kun taas runsasmaaliset vaihtoehdot, kuten 3-1- ja 3-2-tulokset olivat järjestään ylipelattuja Veikkauksen pelaajien toimesta. Tätä voi spekuloida esimerkiksi sillä, että Veikkauksen pelaajat panostavat mieluummin niihin tuloksiin, jotka ovat mahdollisia, vaikka ottelussa tulisi heti alkuun maali. Näin ollen jännitettävää riittää runsasmaalisemmilla lopputuloksilla kauemmin. Täytyy muistaa, että tässä työssä aineisto koostui jalkapallon MM-kisojen otteluisista, jotka kaikki näkyivät ilmaisilla TV-kanavilla Suomessa. Tämä varmasti lisää sellaisten veikkaajien määrää Veikkauksella, jotka haluavat maksimoida jännityksen vedoillaan samalla, kun katsovat ottelua televisiosta.

Tämän tutkielman tulokset ovat yleistettävissä tietyin ehdoin myös MM-kisojen ulkopuolisiin jalkapallo-otteluihin sekä muihin kuin tämän työn vedonlyöntiyhtiöihin. Tuloksien yleistämisessä on kuitenkin huomioitava se, että jalkapallon arvokisoissa ottelut pelataan puolueettomalla maaperällä, kun taas lähes tulkoon kaikki muut jalkapallo-ottelut ympäri maailmaa pelataan toisen joukkueen kotikentällä. Lisäksi kuten Nevalainen (2016) Urheilulehden artikkelissaan toteaa, niin jalkapallon arvoturnaukset poikkeavat seurajoukkuejalkapallosta myös siinä, että karsinnat kisoihin kestävät lähes kaksi vuotta, jonka aikana joukkueiden voimasuhteet voivat vaihdella paljonkin. Tästä johtuen joukkueista ei ole yhtä paljon luotettavaa informaatiota saatavilla kuin seurajoukkueista. Tämä voi omalta osaltaan aiheuttaa eroja sekä pelaajien että vedonlyöntitoimistojen toiminnassa verrattuna muihin otteluihin. Nämä edellä mainitut huomiot eivät kuitenkaan tarkoita sitä, että tämän tutkielman tulokset eivät olisi yleistettävissä muihin jalkapallo-otteluihin kuin arvokisoihin. Tulosten yleistettävyyden kannalta tärkeintä lienee, että pelattavaan otteluun panostettu rahamäärä on

riittävän suuri, jotta epävarmuus rahojen jakautumisesta eri tulosvaihtoehtojen kesken on mahdollisimman pientä. Kärjistäen voidaankin sanoa, että mitä suurempaa mielenkiintoa pelattava ottelu etukäteen herättää, niin sitä varmemmin tämän työn tulokset kyseiseen otteluun soveltuvat.

Tässä työssä kehitetty lähestymistapa ei ole sidoksissa työssä sovellettuun aineistoon. Esitettyä lähestymistapaa, lineaarista regressiomallia, voidaan soveltaa periaatteessa minkä tahansa lajin tulosvaihtoehdon tai minkä tahansa vedonlyöntiyhtiöiden todennäköisyysarvioiden keskinäiseen vertailuun.

Kiitokset

Haluan kiittää tutkielmani ohjaajaa FT Salme Kärkkäistä asiantuntevasta avusta ja kärsivällisyydestä koko projektin ajan. Lisäksi haluan esittää kiitokset tutkielman tarkastajalle lehtori Harri Högmanderille.

6 Lähteet

- Brown, H. & Prescott, R. (1999). *Applied Mixed Models in Medicine*. Chichester: Wiley.
- Davison, A. C. (2003). *Statistical Models*. New York: Cambridge University Press.
- Harrell, F. E., Jr (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer.
- Nevalainen, P. (2016). Kun rikkaus ei riitä. *Urheilulehti* 22/2016, 4–15.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rautee, R. & Merikallio, A. (2009). *Rahapelien kahdet kasvot*. Helsinki: Annukka Merikallio.
- Salonen, A. & Raisamo, S. (2015). *Suomalaisten rahapelaaminen 2015*. Terveystieteiden tutkimuskeskus (THL). Raportti 16/2015. Helsinki.
- Sengupta, D. & Jammalamadaka, S. R. (2003). *Linear Models: an integrated approach*. River Edge, N.J.: World Scientific.
- Walker, M. B. (1995). *The Psychology of Gambling*. Oxford: Butterworth-Heinemann.