Matematiikkaa kaikille
Syksy 2016
Pienryhmätehtävät 1 - **Ohjaajille**

Tehtävä

(1) (a) Ota pari ja valitse kaksi mustaa ja yksi punainen pelikortti. Käännä kortit selkäpuoli ylöspäin, sekoita ne ja aseta riviin. Parisi tavoitteena on saada punainen kortti. Hän valitsee pöydällä olevista korteista yhden, mutta ei vielä käännä sitä ympäri. Tarkista kaksi muuta korttia ja käännä ympäri yksi musta kortti. Kannattaako parisi vaihtaa valitsemaansa korttia?

(b) Merkitkää oheiseen taulukkoon ryhmänne onnistumiset ja epäonnistumiset sekä niissä tilanteissa, jossa valittua korttia vaihdetaan, että tilanteissa, jossa korttia ei vaihdeta. Mitä huomaatte?

	Voitto!	Häviö.
Korttia vaihdettiin		
Korttia ei vaihdettu		

Huomioita

Kyseessä on perinteinen Monty Hall -tehtävä. Aiheesta löytyy sekä suomen- että englanninkielinen Wikipedia-sivu (Monty Hallin ongelma / Monty Hall problem). Yleensä intuitio sanoo, ettei vaihtamisesta ole mitään etua – todennäköisyydet ovat $50 - 50$. Tämä ei kuitenkaan pidä paikkansa. Vaihtamalla korttia todennäköisyys saada punainen kortti on $2/3$. Yksi mahdollienn perustelu on seuraava: todennäköisyys sille, että osut ensimmäisellä kerralla oikeaan on $1/3$. Näin ollen punainen kortti on $2/3$ todennäköisyydellä jompi kumpi kahdesta muusta kortista. Kun toinen näistä korteista paljastetaan mustaksi kortiksi, todennäköiösyys "siirtyy" kokonaan jäljelle jääneelle kortille. Näin ollen vaihtamalla paranee.

B-kohdan tulosten pitäisi suunnilleen noudattaa tätä todennäköisyyttä. Todennäköisyyksien mukaan taulukko näyttää siis seuraavalta:

	Voitto!	Häviö.
Korttia vaihdettiin	67%	33%
Korttia ei vaihdettu	33%	67%

Jos opiskelijat eivät usko (ja muutenkin), voitte miettiä vastaavaa tilannetta sadalla kortilla. Tässä tapauksessa tilanne on seuraava: sinulla on 99 mustaa ja yksi punainen pelikortti pöydällä selkäpuoli ylöspäin. Parisi valitsee niistä yhden. Tämän jälkeen avaat 98 mustaa korttia. Kannattaako parisi vaihtaa ensin valitsemansa kortti toiseen selkäpuoli ylöspäin olevaan korttiin? Tottakai kannattaa - todennäköisyydet ovat tässä 99% vastaan 1%.

Kuvien piirtäminen tilanteista on erittäin suositeltavaa. Aiheeseen liitty myös mielenkiintoinen historiallinen aspekti. Marilyn von Savant, jolla on korkein koskaan mitattu älykkyysosamäärä, selitti ongelman ratkaisun Parade-sanomalehden "Ask Marilyn" -palstalla vuonna 1990. Hän sai siitä käsittämättömän määrän kritiikkiä niskaansa, myös koulutetuilta ihmisiltä ja matemaatikoilta. Ohessa vielä erään vakuuttuneen tohtorin kommentti: "You blew it, and you blew it big! Since you seem to have difficulty grasping the basic principle at work here, I'll explain. After the host reveals a goat, you now have a one-in-two chance of being correct. Whether you change your selection or not, the odds are the same. There is enough mathematical illiteracy in this country, and we don't need the world's highest IQ propagating more. Shame!" – Scott Smith, Ph.D. University of Florida.

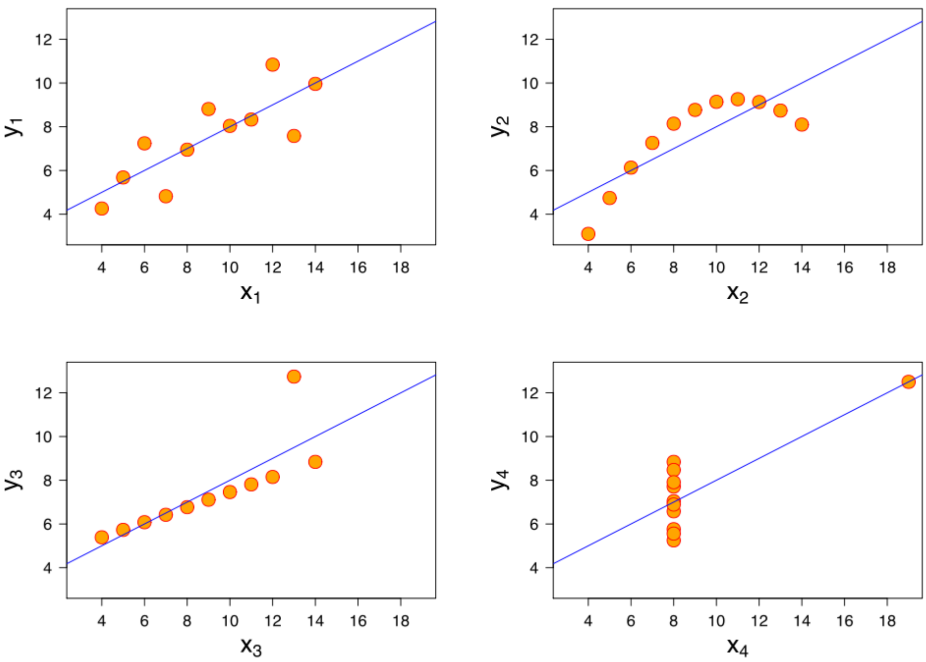
Tehtävä

(2) Ohessa on taulukko neljästä eri tilanteesta, joissa kaikissa on kaksi muuttujaa x ja y. Laske kullekin tilanteelle muuttujien x ja y keskiarvot.

Data 1		**Data 2**		**Data 3**		**Data 4**	
x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Taulukko 1: Anscomben kvartetti; (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4).

Huomioita



Kuva 1: Taulukon 1 datasetit graafisesti esitettynä.

Kyseessä on Anscomben kvartetti. Näille neljälle datasetille yleisimmät tilastolliset tunnusluvut ovat samat. Pelkästään siis niiden perusteella datasetit vaikuttavat identtisiltä. Kun ne kuitenkin esitetään graafisessa muodossa, huomataan, että ne kuvaavat täysin erilaisia ilmiöitä. Tästä syystä aina kun mahdollista, tilastodataa kannattaa tarkastella myös graafisesti!

Tähän tehtävään liittyy myös Excel-pohja, jonka avulla voi laskea helposti keskiarvojen lisäksi muitakin tilastollisia tunnuslukuja.

Tehtävän datasetti ei ole mitenkään uniikki; samankaltaisia settejä on helppo generoida tietokoneohjelmien avulla.

Tehtävä

(3) Pohdi seuraavia tilanteita. Havaitsetko niissä jotain omituista?

(a) Juulia ja Henkka ovat ostaneet tosi hienot hatut. He eivät muista hattujensa hintoja, joten heille tulee sanaharkkaa siitä, kumpi teki paremmat kaupat eli kumman hattu maksoi vähemmän. He päättävät mennä takaisin hattukauppaan selvittämään asiaa. Sitä ennen he lyövät vetoa, että se, kumman hattu on kalliimpi, joutuu antamaan hattunsa toiselle. Juulia ajattelee: "Voittaminen ja häviäminen ovat yhtä todennäköisiä. Hävitessäni menetän rahaa hattuni hinnan verran, mutta voittaessani voitan enemmän kuin hattuni hinnan verran. Tästä syystä minun kannattaa osallistua kilpailuun." Henkka tekee täsmälleen samanlaiset johtopäätökset. Näin ollen sekä Juulian että Henkan kannattaa osallistua vedonlyöntiin.

(b) Juulia ja Henkka ostavat joka päivä uudet hatut, unohtavat lounaaseen mennessä niiden hinnan ja riitelevät iltapäivällä siitä, kumpi teki paremmat kaupat. He lyövät vetoa ja lähtevät selvittämään asiaa. Koska molempien kannattaa osallistua vedonlyöntiin, pitkällä aikavälillä molemmat odottavat jäävänsä voitolle.

Huomioita

Kyseessä on muunnelme Necktie paradiksina tunnetusta ongelmasta. Paradoksi piilee Juulian päättelyketjussa – hän käyttää "hattuni hintaa" kahdessa eri kontekstissä. Vähän sama, kuin määriteltäisiin muuttuja x tarkoittamaan kahta eri asiaa ja päädytään mielenkiintoisiin lopputulemiin.

Päättelyn virheellisyys tulee esille esimerkiksi seuraavanlaisen esimerkin kautta. Oletetaan yksinkertaisuuden vuoksi, että hatut voivat maksaa vain joko 20 € tai 40 €. Nyt seuraavat neljä tapausta ovat yhtä todennäköisiä:

Tapaus	Juulian hatun hinta	Henkan hatun hinta	Juulian odotusarvo
1	20 €	20 €	0 €
2	20 €	40 €	$+40$ €
3	40 €	20 €	-40 €
4	40 €	40 €	0 €

Näistä kaksi keskimmäistä ovat mielenkiintoisimmat. Huomaa, että euromääräisesti tapauksen 3 "hattuni hinta" on sama asia kuin tapauksen 2 "enemmän kuin hattuni hinta".

Ongelmaan liittyy myös kahden kirjekuoren ongelma ("Two envelopes problem", katso vaikka Wikipedia-artikkelia). Ongelmassa näytetään, miten rahaa saadaan nyhjästyä tyhjästä. Ongelma on seuraava:

Saat kaksi täysin samanlaista kirjekuorta. Tiedät, että molemmissa on jokin positiivinen summa rahaa, ja että toisessa kirjeessä on kaksinkertainen määrä rahaa kuin toisessa. Tehtävänäsi on valita toinen kirjekuorista, jolloin saat pitää sen sisältämän rahasumman. Valitset kuoren. Ennen kuin avaat sen, saat vielä mahdollisuuden vaihtaa kuorta toiseen.

Merkitään kirjaimella x valitsemasi kirjekuoren sisältämää rahasummaa. Todennäköisyys sille, että toisessa kuoressa on enemmän rahaa on $1/2$, ja todennäköisyys sille, että toisessa kuoressa on vähemmän rahaa on sama $1/2$. Näin ollen toinen kirjekuori sisältää joko $2x$ tai $x/2$ verran rahaa. Jos valitsemassasi kirjekuoressa on vähemmän rahaa kuin toisessa, toinen kuori sisältää rahaa $2x$ verran. Jos valitsemassasi kirjekuoressa on enemmän rahaa kuin toisessa, toinen kuori sisältää rahaa $x/2$ verran. Odotusarvo toisen kuoren sisältämälle rahasummalle on siis

$$\frac{1}{2}(2x) + \frac{1}{2}\left(\frac{x}{2}\right) = \frac{5}{4}x.$$

Tämä on suurempi summa kuin valitsemassasi kuoressa oleva rahasumma x, joten sinun kannattaa vaihtaa kuorta. Tutkitaan nyt tämän uuden valitsemasi kuoren sisältämää rahasummaa ja merkitään sitä vaikka kirjaimella y. Käytyämme läpi täsmälleen saman päättelyketjun voimme todeta, että toisessa kuoressa on odotusarvoisesti rahaa $5/4y$ verran. Sinun kannattaa siis jälleen vaihtaa kuorta. Itse asiassa sinun kannattaa jatkaa valitsemasi kuoren vaihtamista ikuisesti samalla, kun odotusarvo niiden sisältämille rahasummille kasvaa kasvamistaan. Matematiikka ei toimi näin; mikä meni siis pieleen?

Tehtävä

(4) Haluat selvittää, kuinka monta oppilasta on keskimäärin Helsingin peruskoulujen luokilla.

(a) Valitset satunnaisesti sopivan määrän oppilaita ja lähetät heille kyselylomakkeen. Vastausten perusteella lasket luokkakoon keskiarvoksi 28 oppilasta.

(b) Kysyt luokkien kokoa vielä kaikkien Helsingin peruskoulujen rehtoreilta. Vastausten perusteella lasket luokkakoon keskiarvoksi 23 oppilasta.

Mistä keskiarvojen ero johtuu? Keksi tilanne, jossa haluat erityisesti käyttää tapaa (a). Minkälaisissa muissa tilanteissa tässä tehtävässä esille tullut ilmiö esiintyy?

Huomioita

Ero keskiarvoissa johtuu siitä, että satunnaisotoksessa isoissa luokissa olevat oppilaat tulevat yliedustetuiksi. Tätä voi havainnollistaa esimerkiksi piirtämällä kaksi ympyrää (=luokkaa), joista toisen sisällä on 10 pistettä (=oppilasta) ja toisen sisällä 30 pistettä. Otat 10 oppilaan satunnaisotoksen tästä 40 oppilaan kokonaismäärästä. Valitset todennäköisesti isomman luokan oppilaita, sillä heitä on enemmän kaikkien oppilaiden joukossa. Tällainen lähestymistapa on hyödyllista kuitenkin joissain tilanteissa, esimerkiksi silloin, jos haluat tutkia oppilaiden kokemuksia. "Oikea" keskimääräinen oppilasmäärä luokilla on kuitenkin rehtorien ilmoittama 23 oppilasta.

Tehtävä on yksi esimerkki "Inspection bias" -ilmiöstä. Se esiintyy monessa muussakin paikassa. Esimerkiksi lentoyhtiöt saattavat valittaa siitä, että koneet lentävät tyhjinä eikä matkustajia riitä. Matkustajat kuitenkin valittavat täyteen ahdetuista koneista. Tämä johtuu siitä, että vain muutama lentomatkustaja kokee tyhjän koneen mutta moni matkustaja kokee täyteen ahdetun koneen.

Voit miettiä myös vaikkapa lyhyttä kolmen asiakkaan taksijonoa keskiviikkona klo 13 ja pitkää 30 asiakkaan taksijonoa lauantaina klo 04. Olet mielivaltaisella ajanhetkellä taksijonossa. Millä todennäköisyydellä olet pitkässä jonossa verrattuna lyhyeen jonoon? Käykö sinulle aina niin, että aina kun olet ottamassa taksia, eteen sattuu pitkä jono? Onko universumi sinua vastaan vai olisiko niin, että maailma vain toimii todennäköisyyksien mukaan?

Sama pätee esimerkiksi metrojen vuoroväleihin. Oletko aina asemalle siten, että joudut odottamaan pitkän aikaa? Jos metrolla on kaksi vuoroväliä, pitkä ja lyhyt, ja tulet asemalle sattumanvaraisena ajanhetkenä, on todennäköisempää, että osut pidempään vuoroväliin koska se on, no, pidempi.

Jos valitset mielivaltaisesti yhden Facebookin käyttäjän ja yhden hänen kavereistaan, on todennäköistä, että kaverilla on enemmän kavereita kuin ensimmäiseksi valitulla käyttäjällä. Valinnan ensimmäinen vaihe laittaa kaikki Facebookin käyttäjät samalle viivalle. Vääristymä tapahtuu, kun valitaan tämän käyttäjän kavereista yksi; on todennäköisempää valita henkilö, jolla on paljon kavereita koska hänellä on paljon kavereita.

Jos olet vankilassa, voit ihmetellä sitä, että muilla vangeilla on niin pitkiä tuomioita. Oletetaan, että vangilla A on lyhyt tuomio ja vangilla B pitkä tuomio. Joudut vankilaan mielivaltaisella ajanhetkellä. On todennäköisempää, että tämä ajanhetki osuu yksiin vangin B pitkän tuomion kanssa.

Esimerkkejä voi keksiä paljon lisää. Kuulisin mielelläni teidän ja opiskelijoiden ideoita.

Tehtävä

(5) Vankilassa on 100 vankia ja heitä vartioi julma vanginvartija. Vartija kertoo vangeille, että hän antaa heille huomenna tehtävän, jonka perusteella kaikki vangit joko vapautetaan tai lukitaan vankilaan ikuisesti. Tehtävä on seuraava:

Vangit ovat numeroitu numeroilla $1, 2, \ldots, 100$. Vangeilla on rinnassaan lappu, jossa on heidän oma numeronsa. Vanginvartija käy tänään illalla viemässä erääseen huoneeseen sata numeroitua laatikkoa. Hän kirjoittaa sadalle lapulle numerot $1, 2, \ldots, 100$ ja laittaa ne sattumanvaraisesti laatikoihin (yksi lappu per laatikko). Huomenna aamulla vanginvartija vie vangit yksi kerrallaan laatikkohuoneeseen, jossa kukin vanki saa katsoa viiteenkymmeneen laatikkoon eli puoleen laatikoista. Jos vanki löytää lapun, jossa on sama numero kuin hänen rinnassaan, hän onnistui tehtävässään. Viidenkymmenen laatikon avaamisen jälkeen vanginvartija vie vangin kolmanteen huoneeseen odottamaan. Kun kaikki vangit ovat kolmannessa huoneessa, vanginvartija julistaa tuloksen. Mikäli yhdellekin vangille käy niin, että hän ei löytänyt omaa numeroaan laatikoista, kaikki vangit häviävät ja heidät lukitaan ikuisesti vankilaan.

Vangeilla on ilta aikaa suunnitella toimintastrategia. Millä strategialla heillä on 30% mahdollisuus onnistua? Vangit eivät voi kommunikoida sen jälkeen, kun tehtävän suorittaminen alkaa. Jatkokysymys: jos vangit saavat lahjottua siivoojan vaihtamaan kahden laatikon sisällön keskenään, vankien mahdollisuus onnistua on 100%. Miksi?

Jos sata vankia on paljon, niin kannattaa miettiä samaa ensin kymmenellä vangilla.

Huomioita

Jos vanki valitsee avaamansa 50 laatikkoa satunnaisesti, hän löytää oman numeronsa 50% todennäköisyydellä. Jos kaikki vangit toteuttavat tätä satunnaisuuden strategiaa, todennäköisyys sille, että kaikki vangit onnistuvat tehtävässään on

$$\left(\frac{1}{2}\right)^{100} \approx 10^{-31},$$

joka on tosi pieni luku. Jotain parempaa pitäisi siis keksiä.

Selviytymisstrategian keskeisin idea nojautuu permutaation käsitteeseen. Ratkaisustrategia on seuraava: vanki aloittaa laatikoiden avaamisen siitä laatikosta, jonka numero vastaa hänen rinnassaan olevaa numeroa. Vanki avaa laatikon, katsoo laatikon sisältämän numeron ja siirtyy numeroa vastaavan laatikon luo. Vanki avaa sen, katsoo numeron ja siirtyy numeroa vastaavan laatikon luo. Näin edeten vanki löytää lopulta laatikon, joka sisältää hänen oman numeronsa, ja joka ohjaisi hänet takaisin ensimmäisen laatikon luokse.

Voi olla, että kaikki laatikot ovat osa yhtä pitkää permutaatioketjua. Tällöin vanki ei löydä omaa laatikkoaan sallitulla 50 avauksella. Yli 50 mittaisia ketjuja voi kuitenkin olla vain yksi. Laskettu 30% todennäköisyys selviytymiselle perustuu siihen, miten todennäköisesti tällaisia yli 50 numeron mittaisia ketjuja syntyy. Lahjomalla siivooja vaihtamaan kahden laatikon sisällön keskenään saadaan ainoa mahdollinen yli 50 numeron mittainen ketju jaettua kahtia. Näin kaikki vangit pääsevät vapaiksi käyttämällä yo. strategiaa.

Tarkemmin ratkaisustrategiasta ja todennäköisyyden laskemisesta kannattaa katsoa `http://datagenetics.com/blog/december42014/index.html`.