

Olimpiāžu pārskatos izmantotie apzīmējumi

Anotācija

Izmantojot izklājlapas, ko publisko LU Neklātienes Matemātikas Skola un dažus publiski pieejamus demogrāfisku datu apkopojumus (IZM - skolēnu skaits reģionos, PMLP - iedzīvotāju vecumstruktūra pašvaldībās), izveidoti olimpiāžu rezultātu skaitliski apkopojumi. Šobrīd analizējam vienīgi atklātās olimpiādes.

1 Kur atrast pārskatus

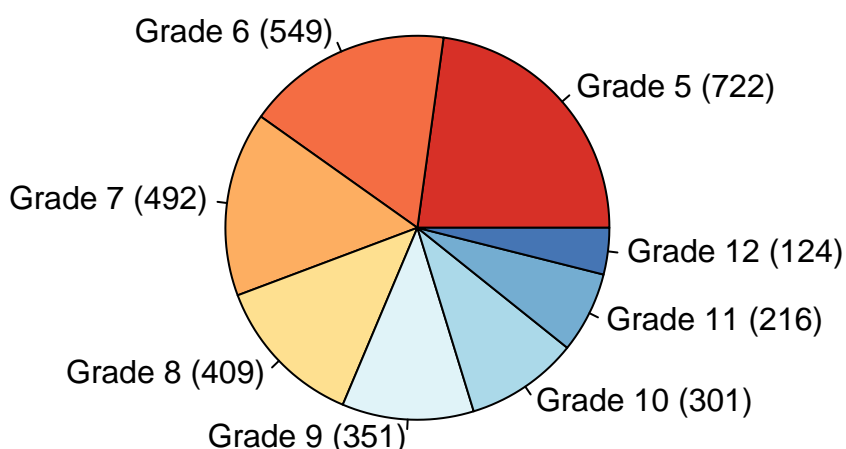
2 Dalībnieku aktivitāte

Šajā sadaļā atbildēsim uz jautājumu, kāda daļa no katrai klasei atbilstošās vecuma grupas skolēniem piedalījās 42. AMO. Dati par skolēnu skaitu pa reģioniem, klasēm un mācību valodām ņemti no IZM publiskotās statistikas — <http://izm.gov.lv/lv/publikacijas-un-statistika/statistika-par-visparejo-izglitiba/2014-2015-m-g>. Dati apkopoti par 9 lielajām pilsētām kā arī par reģioniem, kuros nav ietvertas lielās pilsētas. Ar *reģioniem* domāti NUTS 3 reģioni — sk. http://en.wikipedia.org/wiki/Statistical_regions_of_Latvia - Kurzeme, Latgale, Pierīga Rīga, Vidzeme, Zemgale.

Atklātā matemātikas olimpiāde ir sacensība starp individuāliem risinātājiem, tomēr to var izmantot arī dažādu novadu, skolu un skolotāju vai pulciņu vadītāju darba salīdzināšanai. Viens no svarīgākajiem rādītājiem ir aktivitāte - kāda daļa no attiecīgās skolas vai pašvaldības audzēkņiem piedalījās olimpiādē. Precīzu datu par skolēnu skaitu skolās mums nav, tomēr reizēm var izmantot tuvinājumus. Audzēkņu skaitu 5.-9.klašu grupās uzzinājām tieši no VIIS (Valsts izglītības informatizācijas sistēmas - datu kopa atspoguļo stāvokli 2014.g. 1.septembrī). Par 10.-12.klašu grupām mums šādu datu nav, tādēļ tai vietā skatījāmies uz audzēkņu skaitu, kuri kārtēja obligātos centralizētos eksāmenus matemātikā un reizinājām viņu skaitu ar 3 (vidusskolas vecāko klašu skaitu). Šādā gadījumā jābalstās uz pieņēmumu, ka skolēnu skaits skolā nav strauji mainījies un 10.-12.klasēs joprojām mācās aptuveni tikpat bērnu, cik skolu 2014.g. pavasarī pabeidza.

Ir svarīgs ne tikai dalībnieku skaits, bet arī viņu sagatavotības līmenis. Šajā grafikā ikviena olimpiādes dalībnieka rezultātam ir aprēķināta z -normalizētā vērtība jeb z -score, t.i. no iegūtā punktu skaita jeb *raw score* atņem attiecīgās klases aritmētisko vidējo un izdala ar attiecīgās klases standartnovirzi. Pēc tam katrā reģionā un katrā klašu grupā atsevišķi rēķina šo z -normalizēto vērtību aritmētisko vidējo. Kā redzams diagrammā, vislabākie vērtējumi olimpiādēs ir Latgalē (zilais grafiks) un Rīgā (sarkanais grafiks). Jaunāko klašu grupās Latgales skolnieku rezultāti pārsniedz valstī vidējo par aptuveni 0.3-0.5 standartnovirzēm. (Tipiski jebkurā klašu grupā punktu skaita standartnovirze ir $\sigma \in [8, 11]$, t.i. Latgales skolēnu rezultāti vidēji ir par 3-4 punktiem augstāki nekā citur. Tomēr salīdzināt iegūto punktu starpības nav visai korekti, jo katrā olimpiādes gadā un klašu grupā uzdevumu grūtība un tāad arī punktu izkliede var jūtami atšķirties.)

Participants per Grade



2.1 Dalība un sociāli-ekonomiskie rādītāji

Šeit varētu ievietot diagrammas pa novadiem vai novadu grupām, kas parāda divu parametru attiecību (varētu būt runa par burbulišu diagrammām, ko zīmē divās dimensijās; turklāt burbulīša laukums ir aptuveni proporcionāls skolēnu skaitam olimpiādē).

- Sociāli-ekonomisko rādītāju — bezdarbu, IIN uz 1 iedzīvotāju, pašvaldības izdevumus uz 1 skolēnu vai skolēnu skaitu skolā.
- Dalībnieku aktivitāti (dalībnieku attiecību pret visiem skolēniem novadā) kā arī olimpiādes summāro rezultātu (punktu summas attiecību pret visiem skolēniem novadā).

Šādas diagrammas palīdzētu saprast, kādi sociālie priekšnoteikumi veicina interesi par olimpiādēm, kāda izglītības politika (piemēram, mazo skolu saglabāšana vai slēgšana; lielāki vai mazāki izdevumi par vienu skolēnu) varētu pozitīvi iespaidot olimpiāžu rezultātus.

3 Vidējie rezultāti dalībnieku kategorijām

3.1 “Nevienlīdzība” un Džini koeficienti

Viegli redzēt, ka skolēnu izredzes olimpiādē būtiski atšķiras no izvēlētā matemātikas skolotāja, skolas un arī no pašvaldības. Šajā apakšnodaļā mēģināsim saprast, cik lielā mērā varbūtība piedalīties olimpiādē, izredzes iegūt godalgotu vietu (mūsu aprēķinos — atrašanās augšējā kvartilē) un arī iegūtais punktu skaits ir sadalīti nevienlīdzīgi starp skolotāju audzēkņiem. Ņemot vērā to, ka labākie olimpiāžu rezultāti ir sakoncentrēti nedaudzās skolās, šādu nevienlīdzību nevar pilnībā izskaidrot ar skolotāju prasmēm vien, bet gan ar bērnu atlasī.

Atskaites punkts: Skaitliska “vienlīdzības” simulācija. Protams, pilnīga vienlīdzība nav nedz sasniedzama, nedz arī vēlama. Lai Džini koeficientu aprēķinam rastos kāds atskaites punkts, izdarīsim virkni pieņēmumu par to, kā izglītība ir organizēta hipotētiskā valstī Aizspogulijā, kur ir līdzīga skolēnu un

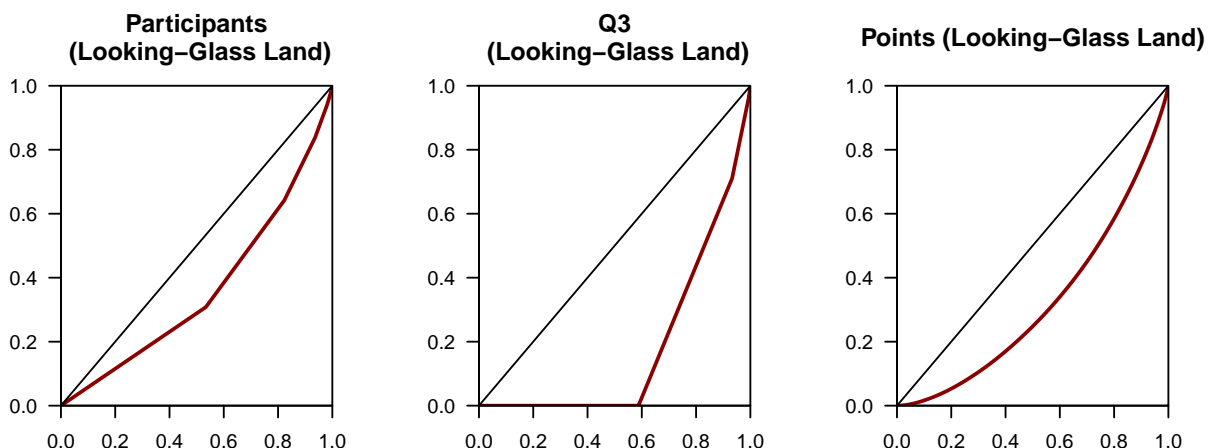
skolotāju demogrāfija, bet olimpiādes rezultāts atspoguļo vienīgi skolēnu matemātiskās spējas. Ja mums rastos iespēja salīdzināt Latvijā pastāvošo rezultātu nevienlīdzību ar kādu reālu valsti, tad Aizspoguliju varētu aizstāt ar kādu reālu piemēru.

- Aizspogulijā dzīvo 120000 skolēnu, kas mācās no 5. līdz 12. klasei.
- Katra Aizspogulijas skolotāja vai pulciņa vadītāja pārraudzībā esošo skolēnu skaits ir nejaušs, vienmērīgi sadalīts vesels skaitlis no 0 līdz 100. (Pavisam Aizspogulijā ir 2400 matemātikas skolotāju — vidēji pa vienam uz katrām 50 skolēniem.)
- Aptuveni 2% no Aizspogulijas skolēniem vēlas piedalīties matemātikas olimpiādē.
- Visā Aizspogulijā skolēniem ir līdzīgas iespējas sagatavoties olimpiādei, olimpiādes rezultāts atspoguļo nevis skolotāja, skolas vai pašvaldības izvēli, bet ir atkarīgs no skolēna dotumiem (matemātiska apdāvinātība, spēja pierakstīt risinājumus, mērķtiecīga gatavošanās, utml.) Olimpiādē saņemtie punkti ir sadalīti ar *nogrieztu normālo sadalījumu* (*truncated normal distribution*) ar vidējo vērtību $\mu = 15$, standartnovirzi $\sigma = 10$ un vērtībām intervālā $[0, 50]$.
- Jebkuram skolēnam ir vienādas izredzes nokļūt pie jebkura skolotāja; nenotiek skolēnu stratifikācija atkarībā no mācību rezultātiem.

Šajā piemērā Aizspogulijas skaitliskie lielumi (skolēnu un matemātikas skolotāju skaits; punktu sadalījums olimpiādē) aptuveni atbilst Latvijas situācijai, tomēr nepastāv skolēnu šķirošana. Aizspogulijas rādījumus Džini koeficientam ņemsim par atskaites punktu, lai varētu salīdzināt, par kādu daļu Latvijā pastāvošās iespējas gatavoties olimpiādēm ir nevienlīdzīgākas. (Nebūtu jēgas salīdzināt ar Džini koeficientu 0, jo tas nozīmētu olimpiāžu dalībnieku un viņu punktu absolūti vienādu sadalījumu starp skolotājiem, kas ir mazticami pat ar Aizspogulijas pieņēmumiem.)

To nevienlīdzību (pareizāk sakot - galaiznākumu nevienādību), kas rodas Aizspogulijā varbūtisko sadalījumu ieviesto nejaušību rezultātā iegūsim, izmantojot skaitlisku olimpiādes rezultātu simulāciju. Katram no 2400 skolotājiem piešķiram nejaušu skolēnu skaitu no 0 līdz 100; katrs skolēns piedalās Bernulli eksperimentā (ar varbūtību 2% piedalās olimpiādē); visbeidzot katrs olimpiādes dalībnieks iegūst punktu kopskaitu atbilstoši nogrieztajam normālajam sadalījumam. Šādi iegūtajiem datiem aprēķinām tās pašas Lorenca līknes un Džini koeficientus.

Lorenca līkne dalībnieku skaita, augšējās kvartiles dalībnieku skaita un punktu summas sadalījumam starp skolotājiem Aizspogulijā (olimpiādes rezultātu vietā skaitliska simulācija):



| Measurement Type | Gini for Looking-Glass Land |
|------------------|-----------------------------|
| Participants | 0.277 |
| Q3 | 0.655 |
| Points | 0.370 |

Aizspoguļijas simulācijā ir ievērojama nevienlīdzība starp to, cik potenciālos olimpiāžu laureātus (risinātājus, kuru rezultāts ir augšējā kvartilē) sagatavotu katrs skolotājs. Ja skolotāju iesaiste olimpiādē ir līdzīgāka un vidējais audzēkņu skaits uz vienu skolotāju ir mazāks, tad ir pašsaprotami, ka liela daļa no viņiem nesagatavos nevienu laureātu. No šejienes izriet secinājums, ka skolu darbu nebūtu pareizi salīdzināt pēc sagatavoto olimpiāžu laureātu skaita — it īpaši, ja mērķis ir izglītības iespēju vienlīdzība. Savukārt, salīdzinājums pēc olimpiādes dalībnieku skaita un arī olimpiādē savāktu punktu skaita ir pilnīgi adekvāts: šie ir rādītāji, kurus ikviena skola var uzlabot neatkarīgi no citām. Pat ar pieņēmumu, ka tikai aptuveni 2% skolēnu vēlas piedalīties atklātajās matemātikas olimpiādēs.

4 Atsevišķo uzdevumu vērtējumi

4.1 Vidējie vērtējumi

Vidējo vērtējumu uzdevumam $X = \{x_i\}$ aprēķina pēc formulas:

$$E(X) = \frac{\sum_{i=1}^n x_i}{n},$$

kur x_i ir i -tā skolēna vērtējums par uzdevumu X , bet n ir visu attiecīgās klases darbu skaits: $n = |X|$.

4.2 Šenona entropija

Entropiju uzdevuma X vērtējumiem aprēķina pēc formulas:

$$H(X) = - \sum_{i=0}^{10} p_i \cdot \log_2 p_i,$$

kur $p_i = \frac{|\{x \in X | x=i\}|}{|X|}$ ir varbūtība saņemt par uzdevumu X vērtējumu i (piemēram p_0 - vērtējumu "0 punkti" vai "uzdevums nav risināts" dalījums ar visu attiecīgās klases darbu skaitu). Ja kāds no vērtējumiem i par attiecīgo uzdevumu nav sastopams, tad attiecīgo saskaitāmo entropijas formulā izlaiž.

Augsta entropija nozīmē augstu nenoteiktību. Ja visi vērtējumi par attiecīgo uzdevumu būtu vienādi, tad entropija ir 0. Ja visi vērtējumi no 0 līdz 11 ir vienādi bieži sastopami, tad entropija ir $\log_2(11) \approx 3.46$. Pārāk zema entropija (piemēram tuvu 1 vai mazāka) nozīmē to, ka pats uzdevums vai tā vērtēšanas sistēma nav bijuši pārāk noderīgi olimpiādes dalībnieku punktu skaita diferencēšanai, jo pārāk daudzi vērtējumi ir vienādi.

4.3 Korelācija ar pārējiem uzdevumiem

Izvēļu testu (*multiple choice exams*) analizē bieži izmanto *biseriālo korelācijas koeficientu* — kāda ir korelācija starp eksāmena kopīgo rezultātu un atbildēm (pareizas/nepareizas) uz konkrēto uzdevuma jautājumu. Ar šo skaitlisko kritēriju var atrast tos testu jautājumus, kuri varētu būt mulsinoši formulēti vai arī nemēra tās pašas prasmes, ko citi šī paša testa uzdevumi.

Protams, olimpiāde nav izvēļu tests (un par katru no jautājumiem ir vairāk vērtējumu nekā tikai 0 vai 1). Tomēr arī šajā gadījumā korelācija var noderēt. Ja uzskatām, ka olimpiāde kopumā mēra noteikta veida matemātiskas prasmes, tad par katru uzdevumu var uzdot jautājumu: Cik labi šis uzdevums palīdz mērīt to pašu, ko olimpiādes uzdevumu komplekts kopumā? Ja konkrētais uzdevums labi "iederas" starp citiem, tad korelācija būs augsta, ja tas mēra kādas stipri atšķirīgas prasmes nekā citi tās pašas klases uzdevumi, tad korelācija būs mazāka. Korelāciju rēķina pēc šādas formulas:

$$\text{cor}(X, Y) = \frac{\sum_{i=1}^n x_i y_i - n \cdot E(X) \cdot E(Y)}{\sqrt{\sum_{i=1}^n x_i^2 - n \cdot E(X)^2} \cdot \sqrt{\sum_{i=1}^n y_i^2 - n \cdot E(Y)^2}},$$

kur x_i ir i -tā dalībnieka vērtējums par uzdevumu X , y_i ir vērtējumu summa par visiem 4 atlikušajiem uzdevumiem, n — darbu skaits attiecīgajā klasē, $E(X)$ apzīmē x_i aritmētisko vidējo.

Korelācijas koeficients vienmēr ir intervālā $[-1, 1]$. Teorētiski varētu gadīties arī negatīva korelācija, t.i. tāds uzdevums, kuru veiksmīgākie olimpiādes dalībnieki risināja sliktāk nekā mazāk veiksmīgie. Tad paša uzdevuma vai vērtēšanas sistēmas korektums radītu nopietnas šaubas. Olimpiāžu praksē tomēr negatīvas korelācijas nav vērojamas. Salīdzinoši zemas uzdevuma vērtējuma korelācijas ar citiem uzdevumiem ļauj atrast tos uzdevumus, kuru tēma vai vērtēšanas kritēriji ir būtiski atšķirušies no citiem uzdevumiem tajā pašā klases 5 uzdevumu komplektā.

4.4 Vērtējumu atšķirības zēniem un meitenēm

Vērtējumu atšķirību konkrētas klases uzdevumam X aprēķina kā divu vidējo vērtību starpību:

$$\Delta_{\text{gender}}(X) = E(X_{\text{male}}) - E(X_{\text{female}}),$$

kur X_{male} ir visi attiecīgās klases zēnu vērtējumi un X_{female} ir attiecīgās klases meiteņu vērtējumi, un $E(X)$ - skaitļu virknes X aritmētiskais vidējais.

Katram no uzdevumiem atrasta zēnu un meiteņu vidējo vērtējumu starpība. Pozitīvs skaitlis nozīmē to, ka zēnu vērtējums bija augstāks, negatīvs skaitlis — to, ka meiteņu vērtējums bija augstāks.