

Fajl koji ćemo koristiti u svim zadacima je **PISA2022zemlje.sav**.

U bazi se nalaze, između ostalih, podaci o udelu učenika svake zemlje čije je najviše postignuće određeni nivo matematičke, čitalačke i naučne pismenosti na PISA2022 testiranju.

Sve informacije o testiranju, uzorku, varijablama, zadacima i ostalim materijalima PISA2022 nalaze se na sledećem linku: <https://www.oecd.org/en/data/datasets/pisa-2022-database.html>.

Zadatak 1:

◆ Napravite raspodelu učestalosti koristeći komande: **Analyze → Descriptive Statistics → Frequencies**, uključivanjem **Display frequency tables**, za sledeće varijable:

var. **mat_maxL** (Matematička pismenost - nivo na kom se nalazi najviše učenika/ca u zemlji)

var. **cit_maxL** (Čitalačka pismenost - nivo na kom se nalazi najviše učenika/ca u zemlji).

◆ *Više informacija o nivoima postignuća kojima ćemo se bavimo na ovom času nalaze se na sledećem linku: https://www.oecd.org/en/publications/pisa-2022-results-volume-i_53f23881-en/full-report/component-10.html-chapter-d1e4472-d219c91287: I to u tabelama:*

Table I.3.1. Description of the eight levels of mathematics proficiency in PISA 2022

Table I.3.3. Description of the eight levels of reading proficiency in PISA 2022 [1/2]

Table I.3.3. Description of the eight levels of reading proficiency in PISA 2022 [2/2].

Po završetku izrade ove tačke, možete kratko pogledati linkovani segment izveštaja i dati smisao svojim rezultatima.

◆ Iskoristite štapićasti dijagram da prikazete grafički raspodele učestalosti na varijabli sa oznakom Svetska regija kojoj zemlja pripada (var. **region**). Komande su sledeće **Graphs → Legacy Dialogs → Bar Charts**. Označite **Simple** i **Summaries for groups of cases**. Varijablu prebacite u polje **Category Axis**.

Iskoristite pitasti dijagram (engl. *pie*) da prikazite grafički raspodele učestalosti na varijabli **region**. Komande su sledeće: **Graphs → Legacy Dialogs → Pie Charts**. Označite **Simple** i **Summaries for groups of cases**. Kada se otvori novi prozor, označiti **Slices represent: N of Cases**, a onda prebaciti varijablu u **Define slices by**. Dakle, svaki isečak pite će predstavljati jednu svetsku regiju, a njegova veličina će predstavljati broj zemalja. Pita dijagram se može uređivati nakon dvostrukog klika na dijagram, kada se otvori Chart Editor. Najpre treba razmisliti o ubacivanju oznaka za veličinu isečaka, ali i ime isečka (da li će stajati u legendi ili kod samog isečka). To se radi u podmeniju **Element**, odabirom **Show data labels**, nakon čega se otvara dijaloški prozor u kome se mogu uređivati po želji. Isprobajte sve opcije.

◆ *Uporedimo prikaze, Šta pie chart u odnosu na bar lakše prikazuje kao informaciju, a šta pomalo skriva, i obrnuto?*

Zadatak 2:

◆ Napravite raspodelu učestalosti koristeći komande: **Analyze → Descriptive Statistics → Frequencies**, uključivanjem **Display frequency tables**, za sledeće varible:

Članstvo u OECD (The Organisation for Economic Co-operation and Development) (var. **OECD**),
Članstvo u EU (Evropskoj Uniji) (var. **EU**).

◆ Izračunajte koja od dveju varijabli **EU** i **OECD** nosi više informacija za opis zemalja koje su učestvovala u testiranju PISA 2022, tj. koja od ovih dveju varijabli ima veću *entropiju*. Koristite za računanje komandu **COMPUTE** i odaberite iz Function Group: **Arithmetic**: a zatim iz Functions and Special Variables: **Ln**, prirodni logaritam, tj. logaritam sa osnovom e.

Entropija (H), odnosno **Target Variable**: koji pravimo, računa se prema sledećem obrascu:

$$H = - \sum p_i * \log p_i$$

Verovatnoće za kategorije, proporcije (p_i), obe varijable se nalaze u ispisu prve tačke ovog zadatka, ali u formatu procenta, tako da ih za potrebe računanja entropije, koja koristi proporcije, delimo sa 100.

◆ *Koja varijabla ima veću entropiju, kakve su entropije?*

Kada entropiju iskazujemo u prirodnim jedinicama (koristimo logaritam sa osnovom e) maksimalna entropija je:

- a) za dihotomnu (binarnu) kategoričku varijablu: $H = -(0.5 * \ln(0.5) + 0.5 * \ln(0.5)) = 0.69$
- b) za kategoričku varijablu sa tri kategorije: $H = -(0.333 * \ln(0.333) + 0.333 * \ln(0.333) + 0.333 * \ln(0.333)) = 1.10$
- c) za kategoričku varijablu sa četiri kategorije: $H = -(0.25 * \ln(0.25) + 0.25 * \ln(0.25) + 0.25 * \ln(0.25) + 0.25 * \ln(0.25)) = 1.39$ ^I

Šta to znači za opis zemalja koje učestvuju u PISA testiranju 2022? Da li biste koristili entropiju ubuduće pre za varijablu sa dve ili za varijablu sa tri i više kategorija?

Zašto nismo uzeli za poređenje entropija tri varijable pismenosti? Jesu li one čist slučaj kategoričke varijable ili neki specijalni slučaj?

Zadatak 3:

- ◆ Izračunajte aritmetičku sredinu za varijablu **OECD** bez obzira na to što je kategorička varijabla;
- ◆ *Pogledajte ponovo raspodelu učestalosti za varijablu **OECD** i pokušajte da ustanovite čemu je u stvari jednaka aritmetička sredina na ovoj varijabli; Da li je to uvek slučaj, bez obzira na to o kakvim se kategoričkim varijablama radi i kako su kodirane kategorije tih varijabli?*

Zadatak 4:

◆ Napravite raspodelu učestalosti varijable **Matematička pismenost - nivo na kom se nalazi najviše učenika/ca u zemlji (mat_maxL)** ali samo u grupi zemalja koja pripada OECD, tj zemalja označenih cifrom 1 na varijabli **OECD**.

Prvo, za odabir ispitanika koristiti komandu **Data → Select Cases**. U dijaloškom prozoru Select Cases, označimo **If condition is satisfied** i definišemo uslov tako što prvo stisnemo dugme **If...** i onda unesemo: **OECD=1**.

Zatim, ponovo uraditi: **Analyze → Descriptive Statistics → Frequencies**, uključivanjem **Display frequency tables**, za varijablu **mat_maxL** (Matematička pismenost - nivo na kom se nalazi najviše učenika/ca u zemlji). Pregledajte raspodelu.

◆ Ovo iznad je bila jednostavna selekcija, sa samo jednim uslovom. Ako bismo želeli da, na primer, selektujemo OECD zemlje, ali samo van Evrope, uneli bismo u polje: **region~="Europe" & OECD=1**. Znak **~=** označava da je nešto: različito od. Navodnici (kao u slučaju "Europe") se unose kad je u pitanju varijabla **String** tipa, kao što je varijabla **region**.

Pravljenje ovih uslova, osim da demonstriramo mogućnosti opcije **Select Cases**, može da nam pomogne i da lakše pregledmo tabelu podatak, pa da i pregledamo ko su članice OECD van Evrope. Uradićemo to u **Data View**, komandom **Sort Descending** po generisanoj **filter_\$** varijabli (pojavila se s desne strane naših kolona u **Data View**).

◆ Isključite **Select cases** iz prethodne tačke time što ćete označiti u istom prozoru **All cases**.

Sada ponovo napravite raspodelu učestalosti **mat_maxL**, ali odvojeno za zemlje koje pripadaju OECD i za zemlje koje ne pripadaju OECD. Za rešavanje zadatka prvo iskoristite komandu **Data → Split File**, stiskom na dugme **Organize output by groups** i ubacivanjem željene varijable u prozor. Zatim napravite raspodelu učestalosti **mat_maxL**.

Zadatak 5:

◆ Kodirajte varijablu **region** u kompletni disjunktni oblik, tj. napravite selektorsku ili indikatorsku matricu za ovu varijablu. **Transform** → **Create dummy variables**:

The screenshot shows the 'Create Dummy Variables' dialog box. On the left, a list of variables is shown, with 'zemlja' selected. On the right, the 'Create Dummy Variables for:' box contains 'region'. The 'Main Effect Dummy Variables' section has 'Create main-effect dummies' checked, and the 'Root Names (One Per Selected Variable):' field contains 'region_'. The 'Value Order' section has 'Ascending' selected. The 'Macros' section has 'Omit first dummy category from macro definitions' unchecked. The 'Measurement Level Usage' section has 'Do not create dummies for scale variable values' selected. The dialog also includes sections for 'Two-Way Interactions' and 'Three-Way Interactions', both with their respective checkboxes unchecked. At the bottom, there are buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.

Pogledajte novonastale varijable s desne strane, u **Data View**. Premestite varijablu **region** blizu njih radi bolje preglednosti. Da li je ovo “dummy” o kojoj smo učili?

Kako bismo ovo peške izveli u komandi **Recode into different variables**?