

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

U bazi se nalaze, između ostalih, podaci o prosečnom postignuću zemalja na testovima matematičke, čitalačke i naučne pismenosti na PISA2022 testiranju. Napomena: proseci matematičke, čitalačke i naučne pismenosti za svaku od 81 zemlje koja je učestvovala u testiranju, dakle, nisu sirovi podaci u pravom smislu te reči. Sve informacije o testiranju, uzorku, varijablama, zadacima i ostalim materijalima nalaze se na sledećem linku:

<https://www.oecd.org/en/data/datasets/pisa-2022-database.html>. Izveštaj se nalazi na sledećem linku: https://www.oecd.org/en/publications/pisa-2022-results-volume-i_53f23881-en/full-report.html

Zadatak 1:

◆ Izračunati aritmetičku sredinu, standardnu devijaciju, varijansu, skjunis (koeficijent asimetrije) i kurtozis (koeficijent izduženosti) za raspodelu varijable matematičke pismenosti (**Mmata**). Koristićemo se komandom **Analyze → Descriptive Statistics → Descriptives → Options** gde ćemo, klikom pored imena svake od njih, označiti tražene statistike i zatražiti prikazivanje njihovih vrednosti u ispisu.

◆ Centrirati i standardizovati rezultate na varijabli **Mmata** korišćenjem komande **COMPUTE**. Centrirane rezultate sačuvati u polju **Target variable** kao varijablu **CMmata**, a standardizovane rezultate u varijabli **ZMmata**;

Centriramo rezultate sa varijable **Mmata** služeći se sledećim obrascem:

$$d_i = x_i - M$$

Standardizujemo rezultate sa varijable **Mmata** služeći se sledećim obrascem:

$$z_i = (x_i - M)/S$$

x_i je u našem slučaju varijabla **Mmata**, a vrednosti za statistike **M** i **S** naći u ispisu prve tačke ovog zadatka. Da biste dobili preciznu vrednost, vrednost sa više od dve decimale, tabelu stisnite dvostrukim klikom. Kada se tabela otvori u novom prozoru, dvostrukim klikom na vrednost Mean ili Sd otvara se više decimala.

◆ Izračunati aritmetičku sredinu, standardnu devijaciju, varijansu, skjunis (koeficijent asimetrije) i kurtozis (koeficijent izduženosti) za varijable **CMmata** i **ZMmata**;

*Uporedite izračunate statistike za varijable **Mmata**, **CMmata** i **ZMmata**; Šta se dešava sa vrednostima **M** i **S** nakon centriranja, a šta nakon standardizacije? Šta se dešava sa vrednostima skjunisa i kurtozisa? Šta zaključujemo o obliku raspodele podataka nakon linearnih transformacija koje se izvode nad podacima?*

Zadatak 2:

- ◆ Grafički prikazite raspodelu varijable **Mmata** pomoću histograma.

Histogram napraviti komandom **Graphs** → **Legacy Dialogs** → **Histogram**, ubacivanjem varijable **Mmata** u prozor **Variable:** pa klikom na **OK**. Komanda ima mogućnost automatskog pravljenja grupisane raspodele, ali mi želimo grupisanu raspodelu u kojoj je veličina intervala razreda jednaka 50 i to možemo podesiti nakon što program isporuči *default* histogram. U prozoru za ispis (engl. *Output*), dvostrukim klikom na dobijeni histogram ulazimo u **Chart Editor**. Potom, isto dvostrukim klikom na bilo koji stubić, pozivamo dijaloški prozor pod imenom **Properties**, gde u **Binning/X Axis/Custom/Interval width** upisujemo 50.

Još jedan način grafičkog prikazivanja kvantitativne varijable jeste pomoću **kutijastog dijagrama**. Odlaskom na **Graphs** → **Legacy Dialogs** → **Boxplot** pozivamo dijalog **Boxplot** gde biramo **Simple**, i označavamo **Data in Chart Are: Summaries of separate variables**. Željenu varijablu prebacujemo u polje **Boxes Represent:** pa klikom na **OK** generišemo dijagram.

- ◆ Prikazite i jediničnu raspodelu **Mmata** komandom **Analyze** → **Descriptive Statistics** → **Frequencies**, uključivanjem **Display frequency tables**.

- ◆ *Uporedimo grafički prikaz i jediničnu raspodelu. Šta nam daje više informacija o obliku raspodele varijable **Mmata**? Koja vrednost Matematičke pismenosti je modalna vrednost?*

- ◆ Grafički prikazite raspodelu varijable **Mmata** pomoću kutijastog dijagrama (engl. *boxplot*).

Dvostrukim klikom na dijagram u prozoru za ispis ulazimo u **Chart Editor**, prozor za uređivanje, u kojem možemo promeniti orijentaciju "kutije" pomoću podmenija **Options** komande **Transpose Chart**. Pre toga treba isključiti natpis *Matematička pismenost - prosek zemlje* dvostrukim klikom na njega i isključivanjem opcije **Display labels**. Natpis nam sada nije potreban, a njegovo uklanjanje doprinosi boljoj preglednosti.

Kako bismo lakše očitali vrednosti statistika koje prikazuje kutijasti dijagram, potrebno je da prvo podesimo **Minor ticks** (crtice na apscisi), a potom i rešetku. Dvostrukim klikom na apscisu (tj. na brojeve ispod) uključuje se dijaloški prozor. U tom dijaloškom prozoru, **Minor ticks** se uključuju u podmeniju **Labels and Thicks**, odeljku **Minor Ticks**, stiskom na **Display Ticks**, i unosom **Number of minor ticks per major ticks**. U našem slučaju to ima smisla da bude broj 4. Rešetka se uključuje u prozoru **Chart Editor** pomoću podmenija **Options** i komande **Show grid lines**, odabirom **Both major and minor ticks**.

- ◆ *Šta nam boxplot govori o merama centralne tendencije i merama raspršenja varijable **Mmata**?*

Zadatak 3:

◆ Sortirajte podatke na varijabli **Mmata** komandom **Sort Cases** tako da ispitanik sa najnižim rezultatom bude prvi; Sortiranje podataka može se obaviti dvojako: ili preko menija **Data → Sort Cases** ili, što je brže i praktičnije, desni klik na ime varijable po kojoj želimo da izvršimo sortiranje pa izborom **Sort Ascending** ili **Sort Descending** već prema tome želimo li rastući ili opadajući niz (u našem slučaju rastući).

◆ Rangujte rezultate na varijabli **Mmata** komandom **Rank Cases** tako da najniži rezultat dobije rang 1; Rangiranje obavljamo pomoću **Transform → Rank Cases**.

◆ *Posmatranjem fajla sa podacima posle svake od ovih komandi, ustanovite u čemu je razlika između komandi **Sort Cases** i **Rank Cases**.*

Tips & tricks: kada poredimo dve ili više varijabli radi veće preglednosti možemo da ih fizički smestimo jednu do druge. Varijable se mogu premeštati prevlačenjem, tako što nakon klika na ime varijable i selekcije iste, ponovo kliknemo na njeno ime posle čega se pored kursora pojavljuje bleđa kontura pravougaonika ili crni precrtani krug. Držeći sve vreme pritisnuto levo dugme miša pomeramo kursor u željenom smeru pri čemu nam pomaže crvena linija vodilja. Kada smo varijabli locirali novo mesto (biće smeštena iza crvene linije) pustimo taster i to je to.

◆ *Pregledom podataka na novoj varijabli **RMmata**, ustanovite kako se dele tzv. vezani rangovi, tj. kako se dodeljuju rangovi zemljama (osobama ili nekoj trećoj jedinici posmatranja nekog istraživanja) koji imaju iste rezultate na varijabli.*

Zadatak 4:

◆ Prevedite izvorne rezultate na varijablama Matematička pismenost - prosek zemlje (**Mmata**), Čitalačka pismenost - prosek zemlje (**Mcitanje**), Naučna pismenost – prosek zemlje (**Mnauka**) u standardizovane rezultate, tj. z vrednosti korišćenjem procedure **Analyze→Descriptive Statistics→Descriptives**. Prevođenje rezultata na varijablama u standardizovane vrednosti sem komandom **Transform →Compute** (kao u prethodnim zadacima) može se izvesti i pomoću **Analyze →Descriptive Statistics →Descriptives** tako što ćemo uključiti opciju **Save standardized values as variables**.

◆ Prevedite izvorne rezultate na varijablama Matematička pismenost - prosek zemlje (**Mmata**) i Čitalačka pismenost - prosek zemlje (**Mcitanje**) i Naučna pismenost – prosek zemlje (**Mnauka**) u percentilne rangove. **Percentilne rangove** računamo za svaku iz običnih rangova (dobijenih komandom **Rank cases**) njihovim deljenjem sa ukupnim brojem slučajeva i množenjem sa 100 (da bismo dobili procenete) u komandi **Compute**. Broj slučajeva u našem zadatku je ukupan broj zemalja koji je učestvovao u PISA testiranju 2022, odnosno 81. Varijablama sa percentilnim rangovima dajte imena **PRMmata**, **PRMcitanje** i **PRMnauka**.

◆ *Korišćenjem standardnih, tj. z vrednosti ustanovite da li zemlja sa rednim brojem 40 u odnosu na ostale članove uzorka ima izraženiji rezultat na Matematičkoj (**Mmata**), Čitalačkoj pismenosti (**Mcitanje**) ili Naučnoj pismenosti (**Mnauka**).*

◆ *Korišćenjem percentilnih rangova ustanovite da li zemlja sa rednim brojem 40 u odnosu na ostale zemlje ima izraženiji rezultat na Matematičkoj (**Mmata**), Čitalačkoj (**Mcitanje**) ili Naučnoj pismenosti (**Mnauka**). A zemlja sa rednim brojem 19?*

◆ *Na kojoj od ovih varijabli zemlja ima viši sirovi, tj. izvorni rezultat?*