

## VAJA 03 IZ EXCELA: OPISNA STATISTIKA. RAČUNANJE IN (STANDARDIZIRANA) NORMALNA PORAZDELITEV.

Datum: november 2011

Pripravil Emil Polajnar

S to vajo spoznamo računanje opisnih statistik in (standardizirano) normalno porazdelitev s pomočjo vgrajenih ukazov v programu Excel.

Podatki so že vnešeni v datoteko Vaje 03.xls.

Literatura: poglavji 4.4 in 5.1 v Kastelec, D., Košmelj, B. (2010): Osnove statistike z Excelom 2007. El. knjiga. Ljubljana: Biotehniška fakulteta.

### NAVODILO

Za 27 držav članic Evropske unije (EU27) imamo za leto 2007 podatke o količini odpadkov v kg na osebo na leto. Države članice so razdeljene glede na leto pristopa, in sicer na stare članice, ki so pristopile pred letom 2004, in na nove članice, ki so pristopile leta 2004 ali kasneje.

stare članice		nove članice	
država	odpadki [kg]	država	odpadki [kg]
Avstrija	597	Bolgarija	468
Belgija	492	Ciper	754
Danska	801	Češka republika	294
Finska	507	Estonija	536
Francija	541	Latvija	377
Grčija	448	Litva	400
Irska	788	Madžarska	456
Italija	550	Malta	652
Luksemburg	694	Poljska	322
Nemčija	564	Romunija	379
Nizozemska	630	Slovaška	309
Portugalska	472	Slovenija	441
Španija	588		
Švedska	518		
Velika Britanija	572		

Podatki so v Excelovi preglednici (datoteka Vaje 03.xls) zapisani v obliki podatkovne zbirke. Vsaka vrstica podatkovne zbirke predstavlja eno statistično enoto, medtem ko stolpci predstavljajo spremenljivke. Prva vrstica v podatkovni tabeli vsebuje imena polj (spremenljivk).

1. Vseh 27 držav obravnavamo kot eno populacijo.

- Kaj je enota analize? Katere spremenljivke so obravnavane? Kakšna je njihova merska lestvica?
- Kolikšna je količina odpadkov na osebo na leto, ki razdeli države EU27 na polovico? Uporabi funkcijo MEDIAN.
- Kolikšna je povprečna količina odpadkov na osebo na leto v državah EU27? Uporabi funkcijo AVERAGE.
- Glede na zadnja dva izračuna, ali je porazdelitev količine odpadkov simetrična ali asimetrična (in v katero smer)?
- Kolikšno je standardno odstopanje od povprečne vrednosti za posamezne države? Izračunaj populacijsko varianco s pomočjo funkcije VAR.P in populacijski standardni odklon s pomočjo funkcije STDEV.P.
- S pomočjo funkcije SKEW izračunaj koeficient asimetričnosti in interpretiraj rezultat. Opozorilo. Vgrajena funkcija izračuna koeficient asimetričnosti za vzorec in ne za populacijo, vendar ne storimo prevelike napake, kadar nas zanima samo smer asimetričnosti.

$$g_1 = \frac{n}{(n-1)(n-2)} \sum \left( \frac{x_i - \bar{x}}{s} \right)^3$$

- S pomočjo funkcije KURT izračunaj koeficient sploščenosti in interpretiraj rezultat. Opozorilo. Vgrajena funkcija izračuna koeficient sploščenosti za vzorec in ne za populacijo, vendar ne storimo prevelike napake, kadar nas zanima samo, ali je porazdelitev sploščena oziroma koničasta.

$$g_2 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left( \frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

2. Primerjava starih in novih članic držav EU27 med seboj. S pomočjo vrtilne tabele (*Pivot Table*) izračunaj osnovne statistične karakteristike za populacijo starih in novih članic EU. Izračunati želimo naslednje karakteristike: število držav, najmanjšo in največjo vrednost, aritmetično sredino, populacijsko varianco in populacijski standardni odklon.

V ta namen v podatkovni zbirki označi stolpca (vključno z naslovno vrstico), ki označujeta članstvo in količino odpadkov. Ustvari vrtilno tabelo (*Insert—Pivot Table*). Imena vrstic v vrtilni tabeli naj označujejo, ali gre za novo ali staro članico EU. To storiš tako, da ustrezno spremenljivko preneseš v območje *Row Labels*. Za vsako od zgoraj zahtevanih karakteristik preneseš ustrezno spremenljivko v območje *Values* in pod *Value Field Settings* vsakič izberi ustrezno karakteristiko (*Count, Min, Max, Average, Varp, StDevp*).

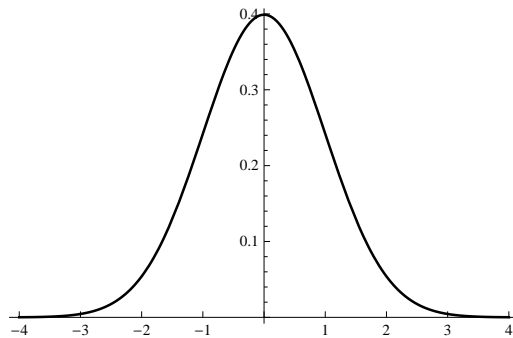
Po izpisu zahtevanih karakteristik za obe skupini držav odgovori na naslednji vprašanji.

- Ali je večja razpršenost vrednosti med starimi ali novimi članicami EU27?
- Ali ima relativno glede na svojo skupino več odpadkov na osebo na leto Slovenija ali Francija? Pomagaj si s funkcijo STANDARDIZE.

3. Standardizirana normalna porazdelitev. Uporabi vgrajeno pomoč za funkcije v Excelu in ugotovi, kaj izračunata funkciji NORM.S.DIST in NORM.S.INV?

- a.) Nariši standardizirano normalno porazdelitev za vrednosti od  $-4$  do  $4$ , kot je prikazano na spodnji sliki, ter za vse vrednosti izračunaj verjetnost  $P(Z \leq z)$ .

V ta namen ustvari stolpec za spremenljivko  $Z$ , ki naj zavzame vrednosti od  $-4$  do  $4$  s korakom  $0.1$ . V sosednjih stolpcih s pomočjo ustrezne funkcije izračunaj vrednosti verjetnostne gostote  $p(z_i)$  in verjetnosti  $P(Z \leq z_i)$  za standardizirano normalno porazdelitev v vseh točkah  $z_i$ . Verjetnostno gostoto prikaži s poligonom (uporabi graf *Scater*).



- b.) S pomočjo zgoraj omenjenih funkcij izračunaj naslednje vrednosti.

- $P(Z \leq 0.85) = ?$
- $P(Z \geq -1.65) = ?$
- $P(-2.58 \leq Z \leq -1.65) = ?$
- $P(Z \leq z) = 0.25 \quad z = ?$
- $P(Z \geq z) = 0.01 \quad z = ?$

4. Normalna porazdelitev. Uporabi vgrajeno pomoč za funkcije v Excelu in ugotovi, kaj izračunata funkciji NORM.DIST in NORM.INV? Z njuno pomočjo odgovori na naslednja vprašanja za populaciji držav EU27.

- Kolikšen delež držav EU27 proizvede manj kot 350 kg odpadkov na osebo na leto?
- Kolikšen delež držav EU27 proizvede več kot 600 kg odpadkov na osebo na leto?
- Izračunaj tretji kvartil in ga interpretiraj.
- Koliko odpadkov na osebo na leto ima tretjina držav z najmanjšo količino odpadkov?