V. Transformacije kvantitativnih podataka

Originalne (sirove) podatke koje dobijamo posmatranjem ili merenjem vrednosti koje jedinice posmatranja imaju u pogledu nekog kvantitativnog obeležja često je neophodno iskazati na drugačiji način, tj. transformisati ih. Transformisanje kvantitativnih podataka je veoma često potrebno iz nekoliko razloga:

a) radi boljeg razumevanja i lakšeg tumačenja značenja rezultata;
b) radi promene oblika raspodele rezultata kako bi se ispunili uslovi za primenu određenih statističkih postupaka;
c) radi pojednostavljenja statističke analize podataka.

Tako, na primer, sirovi rezultat na nekom upitniku ličnosti predstavlja prosto broj poena koje je ispitanik postigao na tom upitniku koji se izračunava u skladu sa ključem za dati upitnik. Međutim, sirovi broj poena sam po sebi ne daje nikakvu informaciju o izraženosti crte ličnosti kod datog ispitanika u odnosu na ostale članove populacije. Ukoliko, pak, rezultat ispitanika iskažemo kao odstupanje rezultata od aritmetičke sredine populacije (koja se ocenjuje na osnovu tzv. normativnog uzorka koji dobro predstavlja populaciju) samim tim ćemo bolje razumeti taj rezultat. Naime, ukoliko je ovo odstupanje pozitivno to znači da ispitanik ima nadprosečno izraženu ovu crtu ličnosti. Ukoliko je odstupanje sirovog rezultata od ocenjene aritmetičke sredine populacije negativno onda već znamo da ispitanik ima ispodprosečno izraženu datu crtu ličnosti, a ako je odstupanje jednako nuli to znači da ispitanik ima prosečno izraženu datu crtu ličnosti. Transformacija koju smo naveli u ovom primeru naziva se <u>centriranjem</u> rezultata. U nastavku teksta ova transformacija će biti podrobnije razmotrena. Ukoliko u navedenom primeru rezultat ispitanika iskažemo u obliku percentilnog ranga, takva transformacija će nam pomoći da odmah uočimo od kolikog procenta ispitanika dati ispitanik ima veći rezultat.

Mnogi statistički postupci podrazumevaju da su ispunjeni određeni uslovi (između ostalog i u pogledu <u>oblika</u> distribucije rezultata) za njihovu adekvatnu primenu. Na primer, za primenu određenih statističkih postupaka neophodno je da distribucija rezultata bude simetrična (ili, čak, normalna).[1] Ukoliko je distribucija sirovih podataka asimetrična (ili odstupa od normalne) moguće je primenom određenih transformacija distribuciju simetrizovati ili normalizovati.

U statističkoj analizi veza među varijablama mogu se koristiti linearni i nelinearni statistički modeli. Linearni statistički modeli su znatno jednostavniji od nelinearnih, ali zahtevaju da se veze među varijablama mogu dobro aproksimirati pravom linijom. Ukoliko to nije slučaj, umesto korišćenja složenijih nelinearnih statističkih modela moguće je varijable transformisati tako da je upotreba jednostavnijih linearnih modela opravdana.

[1] O normalnoj raspodeli ili funkciji gustine može se pročitati u Priručniku na strani 19 Dodatka 2.

Dakle, transformacije sirovih, izvornih rezultata omogućuju da se ti podaci izraze u obliku koji u izvesnom smislu predstavlja poboljšanje u odnosu na izvorni oblik rezultata.[2]

Transformacije kvantitativnih podataka možemo svrstati u dve osnovne grupe: algebarske i površinske transformacije (cf. Horst, 1966, str. 70).

1. Algebarske transformacije

Ove se transformacije izvode primenom određene matematičke funkcije

$$f\colon x_i \rightarrow T(x_i)$$

pri čemu je x_i oriiginalni, sirovi rezultat za *i*-tu jedinicu posmatranja, $T(x_i)$ je transformisani rezultat, a *f* je funkcija koja služi za transformaciju.
Zavisno od prirode funkcije *f* algebarske transformacije mogu biti nelinearne i linearne.

1.1. <u>Nelinearne transformacije</u>

Nelinearne transformacije su one u kojima se kao funkcija za transformisanje koristi neka nelinearna funkcija. Najčešća familija ovih transformacija su tzv. stepene transformacije:

$$T_p(x_i) = \begin{cases} a x_i^p + b & (p \neq 0) \\ c \log x_i + d & (p = 0) \end{cases}$$

pri čemu a, b, c, d $\in R$. Pored toga, za p > 0 treba da bude i a > 0, a za p < 0, treba da bude i a < 0. Ovaj uslov je neophodan da bi transformacija imala određena poželjna svojstva, a pre svega da očuva redosled podataka, tj. da bude ispunjen sledeći uslov:

$$x_r > x_s \Rightarrow T_p(x_r) > T_p(x_s).$$

U drugom pogledu a, b, c i d su arbitrarni, a p se određuje tako da što više olakša analizu podataka.

Za određene vrednosti a, b i p nelinearne transformacije dobijaju određeni oblik. Najčešće korišćene nelinearne transfrmacije su sledeće:

a) Recipročna transformacija (a = 1, b = 0, p = -1):

$$T_{(p=-1)}(x_i) = \frac{1}{x_i}$$

Ovom transformacijom se često vreme reagovanja ispitanika u eksperimentima zamenjuje brzinom reagovanja.

---

[2] U transformacije rezultata o kojima će biti reči u ovom delu teksta ne spada kategorizacija kvantitativnih podataka. Na primer, iskazivanje uzrasta ispitanika u vidu kategorija (kategorija 1= od 0 do 4 godine, kategorija 2 = od 5 do 9 godina, kategorija 3 = od 10 do 14 godina, itd.) ne spada u transformacije kvantitativnih podataka o kojima će biti reči u ovom odeljku teksta.

b) Kvadratna transformacija (a = 1, b = 0, p = 2):

$$T_{(p=2)}(x_i) = x_i^2$$

c) Kvadratnokorenska transformacija (a = 1, b = 0, p = 1/2):

$$T_{(p=1/2)}(x_i) = x_i^{1/2} = \sqrt{x_i}$$

Ova transformacija pogodna je za transformaciju retkih učestalosti (na primer broj grešaka u eksperimentima).

d) Logaritamska (a = 1, b = 0, p = 0):

$$T_{(p=0)}(x_i) = \log x_i \text{ ili } T_{(p=0)}(x_i) = \ln x_i$$

Logaritamska transformacija se koristi kao zamena za stepenu jer kada bi se za p = 0 koristila stepena transformacija svi rezultati bi bili zamenjeni jedinicom. S druge strane opravdanje za ovu zamenu je i sledeća propozicija: kada p teži nuli tada je

$$\frac{x_i^p - 1}{p} = \ln x_i$$

<u>Zapamtite:</u>

- **Stepene transformacije mogu se korisno upotrebiti za simetrizovanje asimetričnih distribucija. Da bi se negativno asimetrična distribucija (tj. distribucija na čijem levom kraju je znatno manje rezultata, a većina rezultata su zbijeni u desnom kraju raspodele) simetrizovala potrebno je koristiti stepene transformacije u kojima je p > 1. Radi simetrizovanja pozitivno asimetrične raspodele (raspodele u kojoj su niži rezultati mnogo brojniji od viših) mogu se koristiti stepene transformacije u kojima je p < 1.**
- **Stepene transformacije najkorisnije su za podatke u kojima je količnik maksimalne i minimalne vrednosti 10 i više od toga.**

Stepene transformacije podrazumevaju da su originalni podaci pozitivni i da nemaju ograničenje na gornjem (desnom) kraju raspodele.

Jedna od veoma čestih nelinearnih transformacija u psihologiji je i <u>arkussinusna transformacija</u> sledećeg oblika:[3]

[3] O osnovnom obliku ove funkcije smo već pisali u delu teksta o osnovnim matematičkim pojmovima. $f(x) =$ arcsin x je funkcija za koju je sin $f(x) = x$. Definicioni skup funkcije je [-1, 1], a skup vrednosti segment [-π/2, π/2]. Funkcija je monotono rastuća.

$$T(x_i) = 2\arcsin\sqrt{x_i}$$

Ova transformacija se najčešće koristi za transformaciju proporcije tačnih odgovora na nekom testu znanja.

Zapamtite: Nelinearne transformacije rezultata menjaju oblik distribucije rezultata.

1.2. Linearne transformacije

Kao što im i samo ime kaže linearne transformacije su transformacije u kojima je *f* linearna funkcija. U opštem obliku ove transformacije mogu se zapisati na sledeći način:

$$T(x_i) = a + bx_i$$

pri čemu je a aditivna konstanta (konstanta koja menja početnu, nultu tačku) a b multiplikativna (skalirajuća) konstanta (konstanta koja menja skalu).
Linearne transformacije se daleko češće koriste u analizi podataka u psihologiji od nelinearnih.
U psihologiji je često potrebno izvorne rezultate linearno transformisati na skalu koja ima određenu vrednost aritmetičke sredine i standardne devijacije. Na primer, rezultati sa testa inteligencije ponekad se, radi lakše interpretacije i poređenja, prevode u skalu čija je aritmetička sredina 100, a standardna devijacija 10. Isto tako, veoma se često rezultati sa upitnika ličnosti prevode u skalu T-skorova, tj. u skalu čija je aritmetička sredina 50, a standardna devijacija 10. Za rešavanje problema transformisanja izvornih rezultata u takav oblik, kojim se postiže to da rezultati u novom, transformisanom obliku imaju odredjenu željenju vrednost aritmetičke sredine i standardne devijacije potrebno je odrediti samo vrednosti aditivne (a) i multiplikativne (b) konstante u linearnoj transformaciji. Jednostavan način da se ovaj problem reši u opštem slučaju je da se postave dve nezavisne jednačine koje govore o tome kako aditivna i multiplikativna konstanta linearne transformacije menjaju aritmetičku sredinu i standardnu devijaciju (kao što je objašnjeno u poglavlju **IV** Priručnika, str 26 i 30):

$$(1)\ M_t = a + b*M$$
$$(2)\ S_t = b*S$$

U ovim jednačinama M i S predstavljaju aritmetičku sredinu i standardnu devijaciju izvornih rezultata, a M_t i S_t označavaju aritmetičku sredinu i standardnu devijaciju rezultata koji su iz izvornih dobijeni linearnom transformacijom. Budući da su nam u gornjim dvema jednačinama poznate aritmetička sredina i standardna devijacija izvornih rezultata, a znamo i kakve treba da budu aritmetička sredina i standardna devijacija transformisanih rezultata, imamo dve nepoznate: a (aditivna konstanta) i b (multiplikativna konstanta). Iz jednačine (2) možemo odrediti rešenje za multiplikativnu konstantu:

$$b = \frac{S_t}{S}$$

a, zatim, ako u jednačini (1) konstantu b zamenimo izrazom $\frac{S_t}{S}$, možemo dobiti rešenje za aditivnu konstantu:

$$a = M_t - \frac{S_t}{S} * M$$

(prema Stockburger, 2000).

Ako želimo da, primenjujući dobijeno rešenje, rezultate na nekoj varijabli linearno transformišemo tako da aritmetička sredina transformisanih rezultata bude jednaka nuli, a da standardna devijacija ostane ista, primenili bismo linearnu transformaciju u kojoj bi multiplikativna konstanta b bila jednaka 1, dok bi aditivna konstanta a bila jednaka aritmetičkoj sredini izvornih rezultata sa negativnim predznakom, tj. a = 0 -1*M = -M. Zamenom ovih konstanti u linearnoj funkciji za transformaciju izvorne varijable x u transformisanu varijablu d dobijamo: d = x - M. Na taj način postižemo tzv. **centriranje** varijable.

Centriranje rezultata

Centriranje rezultata na varijabli sastoji se, dakle, u tome da se svaki rezultat iskaže kao odstupanje od aritmetičke sredine svih rezultata:

$$d_i = x_i - M$$

Devijaciona mera ili **centrirani skor** (d_i) je, dakle, rezultat iskazan kao odstupanje izvornog, netransformisanog ili sirovog skora od aritmetičke sredine niza kojem skor pripada.

Zapamtite: **Centrirani** rezultati imaju **aritmetičku sredinu** jednaku **0 a standardnu devijaciju (odnosno varijansu)** jednaku **standardnoj devijaciji (odnosno varijansi) izvornih rezultata.**

Ukoliko aritmetička sredina transformisanih rezultata treba da bude jednaka nuli, a standardna devijacija jednaka jedinici, tada su neophodne vrednosti konstanti za transformaciju sledeće:

$$a = 0 - \frac{1}{S} * M = -\frac{M}{S}$$

$$b = \frac{1}{S}$$

Zamenom ovih konstanti u linearnoj funkciji za transformaciju izvorne varijable x u transformisanu varijablu z dobijamo:

$$z = -\frac{M}{S} + \frac{1}{S}x = \frac{x - M}{S}$$

Na taj način napravili smo transformaciju koja se uobičajeno zove **standardizacija** varijable x.

<u>Standardizacija</u>

Standardizacija je transformacija rezultata u oblik u kojem podaci ne zavise od mernih jedinica:

$$z_i = \frac{x_i - M}{S}$$

Za standardizovane mere važe sledeće osobine:

1. $\sum_{i=1}^{n} z_i = 0$ jer je $\sum_{i=1}^{n} (x_i - M) = 0$

2. $\frac{\sum_{i=1}^{n} z_i^2}{n} = 1 \quad \Leftrightarrow \quad \sum_{i=1}^{n} z_i^2 = n$

<u>**Zapamtite:**</u> **Standardizovani** rezultati imaju **aritmetičku sredinu** jednaku **0,** a **varijansu (pa i standardnu devijaciju)** jednaku **1.**

<u>Transformacija rezultata u standardni normalni oblik</u>:

Standardni normalni oblik rezultata dobija se deljenjem standardizovanih rezultata kvadratnim korenom veličine uzorka, tj. broja rezultata:

$$z_{ni} = \frac{z_i}{\sqrt{n}}$$

Zbir kvadriranih rezultata koji su transformisani u standardni normalni oblik, tj.

dužina vektora varijable u kojoj su ovakvi rezultati jednaka je jedinici:[4]

$$\sum_{i=1}^{n} z_{ni}^{2} = 1$$

<u>Zapamtite:</u>

Rezultati u **standardnom normalnom** obliku imaju **aritmetičku sredinu** jednaku **0,** a **varijansu** jednaku **1/n.**

****

Centriranje, standardizovanje varijable i transformacija u standardni normalni oblik u SPSS programu se mogu izvesti naredbom **Compute** (meni **Transform**), a standardizacija se osim na taj način može automatski napraviti i uključivanjem opcije **Save standardized values as variables** (klikom na kvadratić ispred te opcije) u proceduri **Descriptives** menija **Descriptive statistics**. Na kraju ispisa iz procedure **Descriptives** (ako je ukljucena opcija **Save standardized values as variables**) program obaveštava korisnika o imenima standardizovanih varijabli koje su upisane na kraju aktivnog fajla sa podacima.

****

Na kraju, veoma čest slučaj u psihologiji je takav postupak linearnog transformisanja kojim se postiže da aritmetička sredina transformisanih rezultata bude 50, a standardna devijacija 10. U tom slučaju vrednosti konstanti za transformisanje su sledeće:

$$a = 50 - \frac{10}{S} * M$$

$$b = \frac{10}{S}$$

Na ovaj način dobijamo obrazac za pretvaranje izvornih rezultata na varijabli x u takozvane T-vrednosti:

$$T = (50 - \frac{10}{S} * M) + \frac{10}{S} x$$

---

[4] Dužina vektora je kvadratni koren zbira kvardriranih elemenata tog vektora. Budući da je u ovom slučaju zbir kvadriranih elemenata vektora jednak jedinici, to je i njegova dužina jednaka jedinici.

<u>Zapamtite</u>:

- Linearne transformacije **ne menjaju oblik distribucije** rezultata. One, takodje, **ne menjaju nivo merne skale** sa koje potiču podaci**.** Na primer, ako izvorni podaci potiču sa intervalne merne skale i podaci koji su linearno transformisani zadržavaju svojstva intervalne skale.
- **I aditivna i multiplikativna** konstanta linearne transformacije **utiču na promenu aritmetičke sredine**, dok na **promenu standardne devijacije** ili varijanse utiče **samo multiplikativna, skalirajuća konstanta**.

****

Algebarske transformacije se u programu SPSS mogu izvesti korišćenjem komande **Compute** (koja je objašnjena u Priručniku u odeljku **IV.3.** na strani 23). U polje **Target Variable** ove komande upisuje se ime (**Name**) nove varijable u kojoj će biti sačuvane transformisane vrednosti, dok se transformacija definiše u polju **Numeric Expression**.

Primeri:

1. Izvorna varijabla EPQN sadrži broj poena, tj. skor na varijabli ekstraverzija. Izračunali smo aritmetičku sredinu i standardnu devijaciju izvornih rezultata i one iznose 15 i 5. U novoj varijabli C_EPQN treba da budu centrirani rezultati ili devijacioni skorovi ispitanika.
U polje **Target Variable** upisaćemo **C_EPQN**, a u polje **Numeric Expression** upisaćemo **EPQN-15**
Kada kliknemo na dugme **OK**, program će napraviti novu varijablu na kraju (tj. "desnom kraju") aktivnog fajla sa podacima.

2. Želimo da standardizujemo rezultate na varijabli EPQN. U polje **Target Variable** upisaćemo **Z_EPQN**, a u polje **Numeric Expression** upisaćemo (**EPQN-15)/5**

3. Želimo da pretvorimo izvorne podatke iz varijable BRGR koja sadrže broj grešaka u eksperimentu korišćenjem kvadratno-korenske transformacije. U polje **Target Variable** upisaćemo **KK_BRGR**, a u polje **Numeric Expression** ubacićemo iz spiska **Functions** funkciju **SQRT(?)** i zameniti znak pitanja imenom izvorne varijable, u ovom slučaju **BRGR**.

4. Želimo da pretvorimo izvorne podatke iz varijable VR koja sadrži vremena reagovanja ispitanika u eksperimentu korišćenjem logaritamske transformacije. U polje **Target Variable** upisaćemo **L_VR**, a u polje **Numeric Expression** ubacićemo iz spiska **Functions** funkciju **LN(?)** i zameniti znak pitanja imenom izvorne varijable, u ovom slučaju **VR**.

5. Želimo da podatke na varijabli EPQE, koja sadrži izvorne rezultate sa upitnika za varijablu ekstraverzija, pretvorimo u skalu čija će aritmetička sredina biti 50, a standardna devijacija 10. Izračunali smo prethodno da je aritmetička sredina izvorne varijable EPQE jednaka 14, a standardna devijacija jednaka 2. U polje **Target Variable** upisaćemo **T_EPQE**, a u polje **Numeric Expression** upisaćemo **(10/2)*EPQE + (50 - 10*14/2)** ili, jednostavnije, **5*EPQE -20**.

****

2. Površinske transformacije

Površinske transformacije sastoje se u prevodjenju sirovih rezultata u transformisane vrednosti koje govore o tome koliki deo od ukupne "površine" distribucije (kada se distribucija grafički prikaže) je ispod ili iznat date vrednosti, tj. koliki procenat rezultata u distribuciji leži ispod ili iznad tako dobijene vrednosti. Najpoznatije površinske transformacije u psihologiji su prevodjenje sirovih rezultata u <u>percentile</u>, <u>percentilne</u> <u>rangove</u> i <u>normalizovane</u> vrednosti.[5] (Percentili i percentilni rangovi su definisani u Priručniku, odeljak *IV.6. Percentili i percentilni rangovi*). Normalizovane vrednosti koje odgovaraju određenom izvornom rezultatu dobijaju se tako što se izračunaju kumulativne relativne frekvence (kumulativne frekvence u procentima ili kumulativne proporcije) za svaki sirovi rezultat i zatim se sirovom rezultatu pridružuje onaj normalizovani rezultat koji na normalnoj raspodeli odgovara njegovoj kumulativnoj relativnoj frekvenci.. Pri prevodjenju izvornih rezultata u normalizovane vrednosti treba imati na umu da ovakva transformacija ima smisla **samo ako se može pretpostaviti** da se varijabla čiji su to rezultati normalno ditribuira u populaciji i ako možemo biti uvereni da su aritmetička sredina i standardna devijacija koje smo dobili na uzorku dobra ocena populacijske aritmetičke sredine i standardne devijacije (ocenjivanje parametara objašnjeno je u Priručniku u odeljku *VI.1. Statističko ocenjivanje parametara*, str. 49 do 55)

****

Računanje percentilnih rangova u SPSS programu može se izvesti tako što se rezultati na varijabli ranguju (procedura **Transform/Rank Cases**), a zatim se na osnovu varijable sa rangovima (čije ime se dobija tako što program dodaje slovo **r** ispred imena izvorne varijable) računaju percentilni rangovi komandom **Transform/Compute**. To se radi tako što se od ranga koji odgovara određenoj meri oduzme 0.5, i zatim ova razlika podeli ukupnim brojem rezultata, tj. ispitanika čije rezultate imamo. Dobijenu vrednost moguće je pomnožiti sa 100 da bi se percentilni rang mere iskazao procentom. Na primer, ako su u varijabli REXTRA iz prethodnog primera rangovi za varijablu ekstroverzija, komandom

COMPUTE PR_EXTR = [100*(REXTRA - 0.5)]/302

dobili bismo percentilne rangove za sve rezultate (ispitanike) koji bi bili upisani u varijabli PREKSTRA na kraju aktivnog fajla sa podacima..

[5] <u>Normalizacija</u> rezultata je transformacija koja se koristi za prevođenje distribucije rezultata u normalnu raspodelu.

Pretvaranje izvornih rezultata u normalizovane skorove može se izvesti korišćenjem procedure **Transform/Rank Cases**. U polje **Variable(s)** prozora za dijalog **Rank Cases** ubaci se ime varijable koja sadrži izvorne podatke, zatim klikne na dugme **Rank Types.** U novom prozoru **Rank Cases: Types** treba kliknuti na dugme **More** i uključiti opciju **Normal Scores**. Uobičajenu opciju (koja je automatski uključena u okviru **Proportion Estimation Formula**) treba menjati samo kada smo sigurni da Blumov postupak želimo da zamenimo nekim od preostalih načina ocene kumulativne proporcije u distribuciji koja odgovara nekom rangu. Program će automatski napraviti novu varijablu koja će u svom imenu imati ispred imena izvorne varijable slovo **n** ili, pak, slova **nor** ispred nekog broja (zavisno od toga koliko takvih varijabli već ima u aktivnom fajlu) i u takvu varijablu upisaće normalizovane vrednosti.
****

<u>Reference koje su korišćene u pisanju ovog teksta</u>

Emerson, J. D., & Stoto, M. A. (1983). Transforming data. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Understanding robust and exploratory data analysis* (pp.97-128). New York: Wiley.

Horst, P. (1966). *Psychological measurement and prediction*. Belmont, CA: Wadswortth publishing company, inc.

Stockburger, D.W. (2000). Linear transformations, WWW dokument, Skinuto 18.2. 2001. sa adrese http://www.psychstat.smsu.edu/introbook/sbk15m.htm.