

Primena metoda uzorkovanja sa verovatnoćom u funkciji kvantifikacije poslovnih aktivnosti

Application of sampling methods with probability as a function of business activities quantification

dr Dragana Radojičić¹
Univerzitet u Beogradu
Ekonomski fakultet

Sadržaj – Glavni cilj tehnika uzorkovanja zasnovanih na verovatnoći je da se dobije visoko reprezentativan uzorak. U okviru ovog rada na izabranu bazu podataka, primenicemo nekoliko planova uzorkovanja: prost slučajni uzorak bez ponavljanja i sistematski uzorak, kao i količniko i regresiono ocenjivanje u okviru tih planova. Implementirani su planovi uzorkovanja kako bi odredili ocene sredine i totala posmatranog obeležja populacije, u cilju krajnjeg poređenja dobijenih ocena. Posebno, upoređujemo tehnike uzorkovanja da bi smanjili varijansu u proceni srednje vrednosti. Dodatno, korsiteći pomoćnu promenljivu za koju se pretpostavlja da je u korelacionoj vezi sa veličinom koja se ispituje, primenjuje se količniko ocenjivanje i regresiono ocenjivanje kako bi se potencijalno povećala preciznost ocene.

Abstract - The main goal of probability sampling techniques is to obtain a highly representative sample. Within this work, on the selected database we will apply several sampling plans: a simple random sample without repetition and a systematic sample, as well as ratio and regression estimations within these plans. Sampling plans are implemented in order to determine the mean and total scores of the observed characteristics of the population, with the aim of ultimately comparing the obtained scores. In particular, we compare sampling techniques to reduce the variance in the estimate of the mean. In addition, using an auxiliary variable that is assumed to be correlated with the quantity under investigation, quantitative scoring, and regression scoring are applied to potentially increase the accuracy of the score.

1 MOTIVACIJA

Cilj svake proizvodnje je da svoje aktivnosti pretvori u novac, tj. da realizuje što veći obim proizvodnje uz poštovanje zahtevanog kvaliteta, asortimana i dinamike proizvodnje, naravno uz što manja ulaganja. Svaki poslovno-proizvodni sistem ima za cilj da ostvari što bolje ekonomsko proizvodne rezultate, pre svega, što veću ekonomičnost, rentabilnost i produktivnost proizvodnje, odnosno što bolji poslovni uspeh.

Poslovno proizvodni sistemi, kao otvoreni sistemi su u stalnoj interakciji sa okolinom. Kompanije rade u nekom makrookruženju kojeg oblikuju različiti uticaji. Ukoliko kompanija nastoji da ostvari dugoračan poslovni uspeh i veću konkurentnost, treba konstantno unapređuje menja da bi adekvatno odgovorila na dinamiku promena koje nameću složeni globalni uslovi. Organizaciono okruženje svake kompanije čine brojni uticajni faktori, koji deluju

izvan i unutar same organizacije, a imaju različite uticaje na ciljeve, funkcionisnje organizacije, poslovne aktivnosti i uspešnost kompanije. Proces prodaje predstavlja strukturiran redosled međusobno povezanih aktivnosti koje kompanija kao prodavac preduzima da bi potencijalnim klijentima plasirala svoje i/ili tuđe proizvode, odnosno da bi potencijalne klijente pretvorila u kupce.

Prodajni procesi i organizaciona struktura kompanije treba da su usaglašeni i prilagodljivi savremenom dinamičkom okruženju. Neophodno je da tržišno orijentisane kompanije vrši analizu procesa prodaje u cilju iznalaženja načina za organizaciona poboljšanja i unapređenja poslovnih aktivnosti, odnosno poboljšanje rezultata poslovanja. Neki od najvažnijih zadataka funkcije prodaje su: prikupljanje podataka i informacija na tržištu prodaje u zemlji i inostranstvu, razvoj i primena različitih prognostičkih metoda, participacija u izradi osnovnog plana poslovanja i proizvodnje kompanije, organizacija prodaje, permanentno unapređivanje sopstvenog rada i saradnje sa istraživačkim organizacijama, povraćaj uloženi sredstava u poslovanje i proizvodnju itd.

Osnovni cilj analize prodaje je da se na bazi relevantnih pokazatelja ostvarenog nivoa prodaje u prethodnom periodu i nivoa sadašnje prodaje predvidi nivo buduće prodaje. Efikasnost ovih analiza moguće je poboljšati korišćenjem metoda teorije uzorkovanja i primenom informacionih tehnologija što je prikazano u ovom radu.

2 Uvod

Grupa podataka je celokupan skup jedinica ili elemenata u kojoj svaka jedinica ima određena obeležja tj. karakteristike. Izučavanje grupe podataka u celini je moguće ukoliko veličina tog skupa nije velika i ne zahteva velike troškove. Češći je slučaj da je skup podataka velikog obima i da je njegovo izučavanje u celini dugotrajno, zahteva velika sredstva, skupo ili je čak principijalno nemoguće. Iz tog razloga se koristi uzorkovanje. Dakle bira se podskupa elemenata, uzorak, iz osnovnog skupa na kome se vrši posmatranje ili istraživanje u cilju opštavanje rezultata cele grupe podataka koju reprezentuje izabrani uzorak. U okviru ovog istraživanja primenicemo dva tipa verovatnosnih uzoraka: prost slučajni uzorak bez ponavljanja i sistematski uzorak. Zadatak ovog istraživanja je da odredimo ocene totala i srednje vrednosti obelezja kod posmatrane grupe podataka u okviru različitih planova uzorkovanja, sa akcentom na nepristrasnim ocenama i ocenama varijanse dobijenih ocena. Osim toga, u okviru prostog slučajnog uzorka radi potencijalnog poboljšanja ocena uključeno je i kolicnicko i regresiono ocenjivanje. Neki od najvažnijih zadataka funkcije prodaje su: prikupljanje podataka i informacija na tržištu prodaje u

zemlji i inostranstvu, razvoj i primena različitih prognostičkih metoda, participacija u izradi osnovnog plana poslovanja i proizvodnje kompanije, organizacija prodaje, permanentno unapređivanje sopstvenog rada i saradnje sa istraživačkim organizacijama, povraćaj uloženi sredstava u poslovanje i proizvodnju itd. Realna i dobro definisana politika prodajnih cena omogućava kompaniji ostvarenje profita i proširivanje na nova tržišta. Treba imati u vidu da na formiranje prodajnih cena utiču različiti interni i eksterni faktori, [1].

Za formiranje prodajnih cena koriste se različite metode a sam izbor metodologije koja će se primeniti u kompaniji zavisi, pre svega, od uslova u kojima kompanija posluje, kao i od znanja i iskustva rukovodstva kompanije. Ostvareni profit kompanije je izvor sredstava za finansiranje njenog rasta i daljeg razvoja. U tom smislu kompanije teže maksimiziranju profita, što većoj profitabilnosti, odnosno što boljim ekonomsko-proizvodnim rezultatima.



Slika 1. Zastupljenost posmatranih proizvoda

3 DESKRIPTIVNE STATISTIKE

U nastavku biće prikazani rezultate deskriptivnih statistika, kao i osnovni parametri dobijeni na osnovu baze podataka. U ovom istraživanju koristi se baza „coffee“ koja je dostupna na sajtu Kaggle¹ i sadrži 4248 observacije.

U bazi se nalaze različiti tipovi podataka: kategorički i numerički. Kategorički podaci su kvalitativni podaci koji razdvajaju ispitanike u grupe prema nekim karakteristikama, i ovom istraživanju kategoričke promenljive koje ćemo posmatrati su: market (tržište), region (region), product_category (kategorija proizvoda), product (proizvod). Informacije o kategoričkim podacima možemo dobiti iz njihovog grafičkog prikazivanja. Slika 1 predstavlja zastupljenost različitih proizvoda koji se nalaze u posmatranoj bazi podataka, dok Slika 2 prezentuje zastupljenost različitih klasa proizvoda. Numerički podaci nastaju kao rezultat merenja i izražavaju se kvantitativno. Kada su u pitanju numeričke promenljive, u ovom istraživanju posmatramo: cost (cena proizvoda), inventory (inventar), net_profit (neto dobit), sales (prodaja).



Slika 2. Zastupljenost različitih klasa proizvoda

	cost	inventory	net_profit	sales
count	4248.000000	4248.000000	4248.000000	4248.000000
mean	84.433145	749.381356	61.097693	192.987524
std	67.249769	661.031896	101.708546	151.133127
min	0.000000	-3534.000000	-638.000000	17.000000
25%	43.000000	432.000000	17.000000	100.000000
50%	60.000000	619.000000	40.000000	138.000000
75%	100.000000	910.500000	92.000000	230.000000
max	364.000000	8252.000000	778.000000	912.000000

Tabela 1. Deskriptivne statistike numeričkih obeležja prisutnih u bazi

Tabela 1 prikazuje vrednosti osnovnih deskriptivnih statistika numeričkih obeležja kao što su aritmetička sredina, standardna devijacija, donji i gornji kvartil, medijana, minimalna prisutna vrednost, maksimalna prisutna vrednost.

S obzirom da je obeležje 'neto profit' target varijabla koju ocenjujemo u ovom istraživanju, naš prevashodni cilj jeste i pronalazjenje tzv. pomoćnog obeležja za formiranje količničke i regresione ocene. Ideja je da značajan deo varijacije od target varijable može da se objasni preko pomoćne varijable. Zbog toga, pomoćna varijabla treba da bude u što većoj meri korelisana sa neto profitom. Na osnovu izračunatih korelacija između posmatranih obeležja prisutnih u bazi koje su prikazane na Slici 3 i u okviru Tabele 2, vidimo međusobnu korelaciju promenljivih i da je ona upravo najveća između neto profita i prodaje (sales), i iznosi 0,797331. Interpretirajući visinu koeficijenta korelacije kao što je istaknuto u [2], zaključujemo da postoji jaka korelaciona veza između promenljive neto profit i promenljive prodaja. Dakle, prilikom dalje analize kao pomoćnu promenljivu posmatraćemo prodaju.

¹ <https://www.kaggle.com/>



Slika 3. Korelacija između posmatranih obeležja

	cost	inventory	net_profit	sales
cost	1.000000	0.621086	0.464838	0.886651
inventory	0.621086	1.000000	-0.091545	0.325934
net_profit	0.464838	-0.091545	1.000000	0.797331
sales	0.886651	0.325934	0.797331	1.000000

Tabela 2. Korelacija između posmatranih obeležja

Formule koje su korišćene za implementaciju planova uzorkovanja i za statistička izračunavanja mogu se pronaći u sledećim knjigama [3], [4], [5].

Pre početka praktičnog rada, izdvojićemo statistike od interesa (koje će u nastavku biti predmet ocenjivanja):

- Total obeležja u posmatranj bazi podataka:

$$Y = \sum_{i=1}^N Y_i = 259543 \quad (1)$$

- Sredina obeležja:

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i = 61,1 \quad (2)$$

- Varijansa obeležja:

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2 = 101,71 \quad (3)$$

Različite metode uzorkovanja iz konačne populacije, kao i primene tih metoda, i ranije su privlačile pažnju istraživača, pogledajte [6], [7], [8], [9], itd. U novije vreme različite metode uzorkovanja u kombinaciji sa tehnikama mašinskog učenja su predmet izučavanja istraživača. U radu [10] prikazan je novi pristup analizi pouzdanosti koji kombinuje uzorkovanje značaja vektorske mašine relevantnosti i simulacije podskupa. Autori u radu [11] koriste metoda uzorkovanja da bi redukovali broj izračunavanja udaljenosti. Stratifikacija latinske hiperkocke, kao tehnika smanjenja varijanse koja omogućava kontrolu varijacija po stratumima je prikazana u [12].

4 Prost slučajni uzorak

A. Određivanje obima uzorka

Osnovna motivacija teorije uzorkovanja je da se se izdvoji dobar reprezentativni uzorak, na osnovu koga je moguće generalizovati rezultate o celom posmatranom skupu. Pri samom odabira jedinica uzorka ključno je odrediti veličinu uzorka i tehniku odabira jedinica uzorka. Kako bi odredili veličinu uzorka n koristimo rezultate dobijene u [3] i [13]

$$n = \frac{N^2 z^2 S^2}{d^2 + N z^2 S^2}, \quad (4)$$

gde je N obim celog skupa, d najveća dozvoljena razlika između ocene i prave vrednosti, z vrednost koju citamo iz tablica za normalnu raspodelu za 95% nivo poverenja. Na osnovu zadate formule dobijena je veličina uzorka od 338 jedinica. Definisanjem veličine uzorka kao $n = 337$, i s obzirom da je obim celog skupa $N = 4248$, dobijamo da je frakcija uzorka (ili stopa izbora uzorka) $f = 0,079$.

B. Prost slučajni uzorak ocenjivanje karakteristika

Prost slučajni uzorak (PSU) (eng. Simple Random Sample - SRS) je najjednostavniji tip verovatnosnog uzorkovanja. U okviru ovog način uzorkovanja n različitih jedinica uzorka bira se tako da svaka jedinica iz populacije ima istu verovatnoću da bude uključena u uzorak, tj. svaka moguća kombinacija od n jedinica ima istu šansu da bude uzorak iz populacije velicine N .

Iz prostog slučajnog uzorka izdvajamo i važne deskriptivne statistike, koje će biti od posebnog značaja prilikom ocenjivanja parametra. Formule koje su korišćene za ocenjivanje parametra kod PSU izvedene su u poglavlju 2 knjige [3].

- Uzoračka varijanca u PSU-u:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = 9425,49 \quad (5)$$

- Prosečan neto profit u uzorku:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 68,67 \quad (6)$$

- Ocena varijanse sredine neto profita:

$$\hat{V}(\bar{y}) = \frac{s^2}{n} (1 - f) = 25,749 \quad (7)$$

- Ocena totala obeležja neto profit:

$$\hat{Y} = N * \bar{y} = 259552,8 \quad (8)$$

C. Količničko ocenjivanje kod prostog slučajnog uzorka

Količničko ocenjivanje je tehnika ocenjivanja koja se sprovodi da bi se potencijalno poboljšala preciznost ocena. Primenjuje se uz pomoć dodatnih informacija tj. uvodjenjem pomoćne promenljive X koja je u korelaciji sa promenljivom Y . Da bi se sprovela ova metoda pomoćna promenljiva treba da bude poznata na svim elementima uzorka, ili treba da bude poznat njen total. U nastavku izvršili smo količničko ocenjivanje nad PSU, posmatrajuć promenljivu net profit kao glavno obeležje, dok je pomoćno obeležje promenljiva sales. Najpre ćemo ustanoviti nekoliko varijabli koje će biti korišćene pri količničkom ocenjivanju.

- Količnik obeležja populacije:

$$R = \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N X_i} = \frac{Y}{X} = 0,31659 \quad (9)$$

- Količnik uzorka:

$$\hat{R} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = \frac{y}{x} = 0.35084 \quad (10)$$

Primenjivanjem količničkog ocenjivanja na PSU i korišćenjem formula iz poglavlja 6 iz knjige [3] dobijeni su sledeći rezultati.

- Količnička ocena totala populacije:

$$\widehat{Y}_R = \hat{R}X = \frac{y}{x} X = 235755.84 \quad (11)$$

- Količnička ocena sredine populacije:

$$\underline{\widehat{Y}}_R = \underline{\hat{R}}\underline{X} = \frac{y}{x} \underline{X} = 55.49808 \quad (12)$$

Izračunate su i varijansa količnika i i aproksimacija varijanse dobijene ocene količničke ocene sredine populacije:

- $V(\widehat{Y}_R) \approx \frac{1-f}{n(N-1)} \sum_{i=1}^N (y_i - Rx_i)^2 = 13,314 \quad (13)$

- $V(\underline{\widehat{R}}) \approx \frac{1-f}{n\underline{x}^2(N-1)} \sum_{i=1}^N (y_i - Rx_i)^2 = 0,00036 \quad (14)$

D. Regresiono ocenjivanje kod prostog slučajnog uzorka

Glavni cilj količničkog ocenjivanja, kao i regresionog ocenjivanja je da poveća preciznost ocene koristeći pomoćnu promenljivu. Regresiona analiza se često koristi za predviđanja, kao i za identifikaciju trendova ili obrazaca u podacima. Obeležje koje se uzima za pomoćno obeležje X (sales) je u jakoj korelacionoj vezi sa veličinom koja se ispituje Y (neto profit). Ukoliko je veza između Y i X linearna, koristi se ocenjivanje zasnovano na linearnoj regresiji Y po X. Formule koje su korišćene pri regresionom ocenjivanju mogu se pronaći u poglavlju 7 u [3], kao i u knjigama Petrovic i Sampling. Linearna regresionu ocenu populacijske sredine označavamo sa y_{lr} gde indeks 'lr' označava linearnu regresiju, a računamo je koristeći sledeću formulu:

$$\underline{y}_{lr} = \underline{y} + b(\underline{X} - \underline{x}), \quad (15)$$

a parametar b je koeficijent linearne regresije. Parametar b modelira nagib prave i meri promenu target obeležja Y kada se pomoćno obeležje X promeni za jednu jedinicu. S obzirom da posmatramo relaciju target varijable sa jednom pomoćnom promenljivom, koristi se prosta linearna regresija. U okviru ove metode da bi se odredile ocene parametara koristi se metoda najmanjih kvadrata koja minimizira sumu kvadrata reziduala. U nastavku su istaknute formule koje su korišćene kod regresionog ocenjivanja, kao i dobijene vrednosti regresionog ocenjivanja kod PSU.

- Koeficijent linearne regresije:

$$b = \frac{\sum_{i=1}^n (y_i - \underline{y})(x_i - \underline{x})}{\sum_{i=1}^n (x_i - \underline{x})^2} = 0.563 \quad (16)$$

- Varijansa koeficijenta linearne regresije je 0.00137

Rezultati regresionog modela sa deskriptivnim statistikama su prikazani na Slici 4.

OLS Regression Results						
Dep. Variable:	net_profit	R-squared:	0.643			
Model:	OLS	Adj. R-squared:	0.642			
Method:	Least Squares	F-statistic:	603.6			
Date:	Mon, 17 Apr 2023	Prob (F-statistic):	6.16e-77			
Time:	13:47:09	Log-Likelihood:	-1872.1			
No. Observations:	337	AIC:	3748.			
Df Residuals:	335	BIC:	3756.			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-43.9004	5.671	-7.741	0.000	-55.056	-32.745
sales	0.5631	0.023	24.568	0.000	0.518	0.608
Omnibus:	133.510	Durbin-Watson:	1.950			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6850.322			
Skew:	-0.797	Prob(JB):	0.00			
Kurtosis:	25.030	Cond. No.	410.			

Slika 4. Rezultati regresionog ocenjivanja kod PSU

Dobijena regresiona ocena populacijske sredine y_{lr} je 63.769, dok je varijansa dobijene ocene 17,34.

5 Sistematski uzorak

Sistematski uzorak obima n sa korakom k se bira tako što se prvo izabere jedna jedinica i od prvih k jedinica, a zatim biramo svaku k -tu jedinicu. Korak k , koji se u literaturi često naziva i interval, određuje se celobrojnim deljenjem broja elemenata u populaciji N sa veličinom uzorka n . Dakle, formira se uzorak čiji su elementi sa sledećim indeksima:

$$i, i + k, i + 2k, \dots, i + (n-1)k.$$

Sistematski uzorak se primenjuje kad su elementi populacije poredani u nekom poretku (proizvodi na traci, knjige na polici, itd.).

A. Količničko ocenjivanje kod sistematskog uzorka

Ocene dobijene primenom količničkog ocenjivanja kod sistematskog uzorka su:

- Količnik obeležja populacije:

$$R = \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N X_i} = \frac{Y}{X} = 0.31659 \quad (17)$$

- Količnik uzorka:

$$\hat{R} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = \frac{y}{x} = 0.3537 \quad (18)$$

- Količnička ocena totala populacije:

$$\widehat{Y}_R = \hat{R}X = 259543 \quad (19)$$

- Količnička ocena sredine populacije:

$$\underline{\widehat{Y}}_R = \underline{\hat{R}}\underline{X} = 61.09769 \quad (20)$$

Izračunate su i aproksimacije varijanse količnika i varijanse dobijene ocene količničke ocene sredine populacije:

- $V(\widehat{Y}_R) \approx \frac{1-f}{n(N-1)} \sum_{i=1}^N (y_i - Rx_i)^2 = 12.62 \quad (21)$

- $$V(\hat{R}) \approx \frac{1-f}{n\bar{X}^2(N-1)} \sum_1^N (y_i - Rx_i)^2 = 0.00034$$
 (22)

B. Regresiono ocenjivanje kod sistematskog uzorka

Ocene dobijene primenom regresionog ocenjivanja kod sistematskog uzorka su:

- Koeficijent linearne regresije:

$$b = \frac{\sum_1^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_1^n (x_i - \bar{x})^2} = 0.54287 \quad (16)$$

Varijansa koeficijenta linearne regresije je 0.00142

Rezultati regresionog modela sa deskriptivnim statistikama su prikazani na Slici 4.

```

=====
OLS Regression Results
=====
Dep. Variable: net_profit R-squared: 0.706
Model: OLS Adj. R-squared: 0.705
Method: Least Squares F-statistic: 846.3
Date: Mon, 17 Apr 2023 Prob (F-statistic): 1.17e-95
Time: 18:43:42 Log-Likelihood: -1952.4
No. Observations: 354 AIC: 3909.
DF Residuals: 352 BIC: 3917.
DF Model: 1
Covariance Type: nonrobust
=====
coef std err t P>|t| [0.025 0.975]
-----+-----+-----+-----+-----
const -40.6959 5.137 -7.922 0.000 -50.799 -30.592
sales 0.5429 0.019 29.091 0.000 0.506 0.580
=====
Omnibus: 100.711 Durbin-Watson: 2.161
Prob(Omnibus): 0.000 Jarque-Bera (JB): 4864.658
Skew: 0.115 Prob(JB): 0.00
Kurtosis: 21.159 Cond. No. 441.
=====

```

Slika 4. Rezultati regresionog ocenjivanja kod PSU

Dobijena regresiona ocena populacijske sredine y_{lr} je 64.07, dok je varijansa dobijene ocene 17,23.

6 ZAKLJUČAK

Koristeći različite metode teorije uzoraka (za konačnu populaciju) dobili smo ocene karakteristika od interesa putem statističkih proračuna. Preciznije, posmatrajući prost slučajni uzorak i sistematski uzorak, kao i metode količnikog i regresionog ocenjivanja, ocenili smo sredinu i total target varijable 'neto profit', kao i varijanse dobijenih ocena. Sagledavanjem rezultata primećujemo da se ocene mogu smatrati preciznim, s obzirom da su izračunate varijanse relativno male. Kao i kod prostog slučajnog uzorka i kod sistematskog uzorka najmanja varijansa ocene sredine obeležja dobijena je primenom količnikog ocenjivanja. Daljim sagledavanjem dobijenih ocena zaključujemo da je najmanja varijansa zabeležena regresionim ocenjivanjem kod sistematskog uzorka. Korišćenjem pomoćne varijable dobija se preciznija ocena, jer se regresiono ocenjivanje sprovodi uz pomoć dodatnih informacija. Dakle, sagledavanjem rezultata kao optimalan plan uzorkovanja biramo regresiono ocenjivanje kod sistematskog uzorka.

LITERATURA

- 1 Fitzsimmons, J., Steffens, P., & Douglas, E. (2005). Growth and profitability in small and medium sized Australian firms. Proceedings AGSE Entrepreneurship Exchange, Melbourne.
- 2 Evans, J. D. (1996). Straightforward statistics for the behavioral sciences. Thomson Brooks/Cole Publishing Co.
- 3 Cochran, W. G. (1977). Sampling techniques. John Wiley & Sons.
- 4 Petrović, Lj. (2007). Teorija uzoraka i planiranje eksperimenata. Centar za izdavačku delatnost Ekonomskog fakulteta.
- 5 Sampath, S. (2005). Sampling theory and methods. Alpha Science Int'l Ltd.
- 6 Durbin, J. (1953). Some results in sampling theory when the units are selected with unequal probabilities. Journal of the Royal Statistical Society: Series B (Methodological), 15(2), 262-269.
- 7 Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of the American statistical Association, 47(260), 663-685.
- 8 Madow, W. G. (1949). On the theory of systematic sampling, II. The Annals of Mathematical Statistics, 20(3), 333-354.
- 9 Murthy, M. N. (1967). Sampling theory and methods. Sampling theory and methods.
- 10 Xie, B., Peng, C., & Wang, Y. (2023). Combined relevance vector machine technique and subset simulation importance sampling for structural reliability. Applied Mathematical Modelling, 113, 129-143.
- 11 Ding, S., Li, C., Xu, X., Ding, L., Zhang, J., Guo, L., & Shi, T. (2023). A Sampling-Based Density Peaks Clustering Algorithm for Large-Scale Data. Pattern Recognition, 136, 109238.
- 12 SONG, C., & KAWAI, R. (2023). Dynamic finite-budget allocation of stratified sampling with adaptive variance reduction by strata. SIAM Journal on Scientific Computing.