

## PRIMJENA STATISTIČKIH METODA U ISTRAŽIVANJU ODLAGANJA, ČUVANJA I ZAŠTITE DOKUMENTACIJE

---

*Problem odlaganja, čuvanja i zaštite dokumentacije, a s time i informacijskog sadržaja, zahtijeva mnogobrojna istraživanja. Među njima javljaju se i statistička istraživanja. Dominantna područja ovih istraživanja kreću se u nekoliko pravaca. Tu svakako ulazi problem stratifikacije promatranog skupa, projektiranja opsega uzoraka po stratumima, proračun odgovarajućih karakteristika te postavljanje regresijskih modela karakteristika dokumentacije.*

*Stratifikacija; uzorak; karakteristika; regresija.*

---

### 1. UVOD

Odlaganju, čuvanju i zaštiti dokumentacije (informacijskog sadržaja) može se pristupiti s više stajališta. S jedne strane, ovu klasu problema susreće se u povijesnom, privrednom, uslužnom, vojnom, a i u nekim drugim kompleksima. S druge strane, odlaganje, čuvanje i zaštita dokumentacije prvenstveno je u domeni informatičkih problema i metoda njihova rješavanja.

Postaviti modele utvrđivanja karakteristika dokumentacije znači izvesti prikupljanje informacija o stanju dokumentacije u poslovnim subjektima. Nadalje, izvodi se istraživanje relevantnih svojstava i karakteristika dokumentacije te postavljanje odgovarajućih modela. Modelima se generiraju informacije relevantne za projektiranje i izgradnju sustava odlaganja, čuvanja i zaštite dokumentacije u okviru zadanih kriterija.

S obzirom na činjenicu da se u modeliranju uvodi i statistika i da se radi o skupovima s velikim brojem elemenata, neophodno je izvesti stratifikaciju, odnosno podjelu skupa na stratum ili klase. Stratifikacija, ukoliko je dobro izvedena, treba doprinijeti postavljanju konzistentnih modela čija rješenja osiguravaju visoku vjerojatnost generiranja karakteristika dokumentacije.

## 2. STRATIFIKACIJA

U mnogim praktičnim situacijama javljaju se skupovi koji, s obzirom na svojstva elemenata, imaju veoma jaki varijabilitet te kao takvi predstavljaju heterogene skupove. Analizom ovih skupova dobivaju se karakteristike koje nisu dovoljno reprezentativne i kao takve ne mogu se uzeti kao opća ili zajednička svojstva elemenata. Da se dobiju reprezentativnije karakteristike, takvi skupovi se dijele na homogenije dijelove, tj. stratume. Na taj način, prilikom izbora elemenata, eliminira se onaj dio varijabiliteta koji se pojavljuje između stratuma, a koji može biti vrlo značajan.

Stratifikacija se provodi prema vrsti svojstva koje se izučava kod elemenata, i to najčešće u onim slučajevima kada se procjenjuju ili donose sudovi o nepoznatim karakteristikama. Tada se iz svakog stratuma izabire uzorak na osnovi stratificiranog izbora. Stratificirani izbor je plan kod kojega se skup dijeli na  $M$  stratuma ili klasa te se uzorak uzima iz svake klase kao da je ona odvojeni skup. Prema tome cilj je stratificiranja da se skup "razbije" na klase koje se međusobno značajno razlikuju u odnosu na svoje karakteristike (prosjeck, varijacija) ili nivo u kojem je zastupljeno neko kvalitativno obilježje.

Osnovni problem koji se javlja kod stratificiranog uzorka je u pitanju broja stratuma i načina stratifikacije. Naime, skup treba tako podijeliti da se osigura najmanja greška ocjene neke karakteristike. Mali broj stratuma može dovesti do značajne varijabilnosti, što utječe na značajnost odgovarajućih parametara (karakteristika). Velik broj stratuma otežava rad i znatno povećava troškove istraživanja. Jasno je da se određena varijabilnost u stratumu mora tolerirati i da je ni jedna shema stratifikacije ne može posve eliminirati.

Prilikom stratifikacije skupa treba voditi računa da razlike između stratuma budu što veće. U praksi se podjela skupa na stratume uvijek dovodi u kontekst s izučavanom situacijom. Poželjno je, da bi se odredio broj stratuma i način stratifikacije, prethodno prikupljanje informacija o strukturi izučavanog skupa. Informacije se odnose na ona obilježja ili svojstva koja će se ispitivati i čije karakteristike će se procjenjivati ili testirati.

Za potrebe modeliranja sustava odlaganja, čuvanja i zaštite dokumentacije stratifikaciji poslovnih subjekata može se prići s gledišta privrednih i neprivrednih subjekata. Unutar privrednih subjekata pojedine stratume mogu činiti djelatnosti kao što su: industrija i rudarstvo, poljoprivreda i ribarstvo, šumarstvo, vodoprivreda, građevinarstvo, promet i veze, trgovina, ugostiteljstvo i turizam, obrtništvo i osobne usluge, stambeno-komunalne djelatnosti i uređenje naselja i prostora, financijske, tehničke i poslovne usluge. Istraživanja trebaju, nadalje, pokazati da li je i unutar ovih stratuma moguće izvesti još precizniju stratifikaciju. Tako npr. unutar stratuma industrije

i rudarstva mogu se privredni subjekti klasificirati prema vrsti industrijske djelatnosti. Takva podjela mogla bi dovesti do homogenijih stratuma i statistički preciznijih rezultata.

Kod subjekata s područja neprivrede stratifikacija bi se mogla izvesti kako slijedi: prosvjeta, znanost, zdravstvena i socijalna zaštita, kulturno-socijalna djelatnost te državni organi i društvene organizacije. Unutar ovakve stratifikacije moguća je i uža podjela, što treba pokazati empirijsko istraživanje. Dubina stratificiranja treba ići do tih granica dok se ne dobiju grupe ili klase koje osiguravaju dovoljnu reprezentativnost karakteristika s obzirom na istraživana obilježja elemenata u stratumima.

Na ovom mjestu treba navesti da se u poslovnim subjektima susreće vrlo različita poslovna dokumentacija. Određena istraživanja (3) pokazuju da se u stratumu "industrija" dokumentacija može kategorizirati u velik broj grupa. Govori se o proizvodnoj, komercijalnoj, računovodstveno-financijskoj, analitičko-planskoj, kadrovskoj i sl. dokumentaciji. Istraživanjem treba utvrditi da li bi i unutar ovog segmenta trebalo formirati grupe ili klase kako bi se dobile što preciznije karakteristike prilikom izgradnje modela odlaganja, čuvanja i zaštite dokumentacije.

Predložena stratifikacija mogla bi se izvesti s obzirom i na neko drugo obilježje poslovnih subjekata. Tako npr. privredni subjekti mogu se grupirati ili klasificirati prema broju zaposlenih, visini prihoda ili troškova, dobiti i sl. Istraživanjem treba dati odgovor koji oblik stratifikacije je najprikladniji sa stajališta reprezentativnosti očekivanih karakteristika.

### 3. UZIMANJE UZORKA

Izgradnja modela odlaganja, čuvanja i zaštite dokumentacije poslovnih subjekata zahtijeva istraživanja subjekata po njihovim karakterističnim svojstvima. Tako npr. želi li se utvrditi opseg dokumentacije koji se javlja u tim subjektima, neophodno je istraživanjem obuhvatiti te subjekte. U svakom slučaju, s obzirom na geografsku, vremensku i prostornu definiciju takvog skupa, javlja se značajan broj poslovnih subjekata. Sveobuhvatno istraživanje svih subjekata dovelo bi prvenstveno do ogromnih troškova istraživanja te dugotrajnog procesa promatranja i analize rezultata istraživanja.

Prilikom različitih istraživanja, a pogotovo kod tzv. "terenskih istraživanja", istraživači dolaze u poziciju da iz jednog relativno velikog skupa odaberu uzorak koji će poslužiti za procjenu karakteristika skupa ili pak za testiranje hipoteze o nepoznatim karakteristikama skupa. Principijelno je važno da uzorak bude što veći, no u mnogim slučajevima treba ga ograničiti po veličini. Razlozi ovome su u povećavanju troškova zbog veličine uzorka, a u nekim slučajevima i u uništavanju elemenata s obzirom na činjenicu da ispitivanje dovodi do lomova ili sličnih oštećenja elemenata.



Da se dođe do ispravnih zaključaka o skupu pomoću uzorka, on mora biti reprezentativan. Po strukturi uzorak predstavlja osnovni skup u malom. Ovakav zahtjev nameće potrebu da se planu izbora uzorka pristupi vrlo ozbiljno. Reprezentativan uzorak nije uvijek lako izabrati. Prilikom istraživanja u vrlo mnogo slučajeva premalo se poznaje skup koji treba istražiti. No, ovdje se može pojaviti kao otežavajuća okolnost i činjenica da su elementi prostorno vrlo disperzirani, da je varijabilnost promatranog obilježja velika i tome slično.

Kada se govori o varijabilnosti obilježja elemenata skupa, treba razlikovati homogene i heterogene skupove, a što može utjecati na samu veličinu odabranog uzorka. Poznata je činjenica da se iz homogenijih skupova metodom uzorka dobivaju preciznije procjene. Kada je u istraživanju broj elemenata nekog skupa poznat, tada se govori o konačnom skupu. Ukoliko se broj jedinica u skupu ne može ni približno odrediti, ili ako u taj skup neprekidno pristižu novi elementi, tada se takav skup smatra beskonačnim skupom. Ove konstatacije treba uzimati u obzir prilikom odabira elemenata u uzorak.

Elementi u uzorak mogu biti odabrani na više načina: namjernim, prigodnim, slučajnim i sustavnim odabirom elemenata. U velikom broju statističkih istraživanja koristi se slučajni odabir, tj. onaj u kojem svaki element skupa ima jednaku vjerojatnost da bude odabran u uzorak. Slučajni uzorci su jedini koji imaju poznatu raspodjelu, i to većinom normalnu. Slučajni uzorak sastavljen je prema određenim principima koji odgovaraju zakonu slučaja. Najbolji način kod sastavljanja uzorka sastoji se u upotrebi Tablice slučajnih brojeva.

Za potrebe izgradnje modela odlaganja, čuvanja i zaštite dokumentacije poslovnih subjekata koristit će se stratificirani slučajni uzorak. Stratificirani slučajni uzorak daje informacije ne samo o karakteristikama cijelog skupa već i o njegovim sastavnim dijelovima (stratumima, klasama ili grupama). Model stratificiranog uzorka sastoji se u slijedećem (6):

- skup se podijeli u grupe ili stratume, i to tako da se u svakoj grupi nalaze jedinice koje su što homogenije s obzirom na promatrano obilježje. Pri tome je važno da svaka jedinica skupa mora biti uključena samo u jednu od naznačenih grupa ili stratuma,
- iz svakog pojedinog stratuma posebno se izabire uzorak čija veličina zavisi o nizu činitelja te su te veličine međusobno i različite,
- potrebno je za svaki uzorak, izabran iz stratuma, izračunati karakteristiku koja će poslužiti u daljnjoj obradi za procjenu karakteristike skupa i
- na osnovi izračunatih karakteristika uzoraka svakog stratuma procjenjuje se karakteristika skupa.

Jedno od vrlo važnih i čestih pitanja postavlja se u vezi kako velik mora biti uzorak. Ovo pitanje je vrlo složeno te u nekim specijalnim slučajevima postoje i gotova pravila, tablice i nomogrami. No, u mnogobrojnim pak slučajevima na ovo pitanje nije moguće odgovoriti jednim konkretnim brojem. Naime, potrebna veličina uzorka prvenstveno zavisi o varijabilnosti obilježja jedinica skupa koji se istražuje. Isto tako, veličina ovisi i o

preciznosti s kojom se pojava želi istraživati. U svezi s ovim pojave s malom varijabilnosti obilježja elemenata pokazuju da je za potrebe ili procjene ili testiranja hipoteza nepoznatih karakteristika skupa dovoljan i mali uzorak. Jaka varijabilnost elemenata s obzirom na ispitivano obilježje zahtijeva velik uzorak. Isto tako, velika preciznost zahtijeva velik uzorak, a manja preciznost manji uzorak.

Za uzorak je važnije da bude reprezentativan nego da bude velik. Reprezentativnost uzorka postiže se pažljivim slučajnim izborom elemenata u uzorak. U nekim specifičnim slučajevima može se "grubo" predvidjeti u kojem postotku je neko svojstvo zastupljeno u skupu. Veličina uzorka tada se može donekle odrediti tako da se taj postotak pomnoži s postotkom koji nedostaje do 100% (5).

U istraživanju koje je potrebno provesti za konstrukciju modela odlaganja, čuvanja i zaštite dokumentacije velik broj karakteristika svodi se na proračun srednje vrijednosti (aritmetičke sredine). U takvom slučaju za veličinu uzorka mogu se koristiti ovi obrasci (6):

$$n' = \left[ \frac{z * \sigma}{G_a} \right]^2, \quad (1)$$

ili

$$n' = \left[ \frac{z * V}{G_r} \right]^2, \quad (2)$$

gdje je:

$z$  - koeficijent pouzdanosti koji se dobije iz tablice površina ispod normalne krivulje za zadanu statističku vjerojatnost,

$\sigma$  - standardna devijacija skupa,

$G_a$  - apsolutna greška koja se može tolerirati prilikom procjene karakteristike skupa,

$V$  - koeficijent varijacije koji se računski dobije omjerom standardne devijacije i aritmetičke sredine skupa i

$G_r$  - relativna greška koja se može tolerirati prilikom procjene karakteristike skupa.

Korištenjem obrazaca (1) i (2) može se dobiti da je frakcija izbora  $f = n/N$  ( $n$  - opseg uzorka;  $N$  - opseg skupa) veća od 0,05. U takvom slučaju može se uzeti u obzir faktor korekcije što znači da je veličina uzorka izračunata naznačenim obrascima tek prethodni rezultat. Konačan broj elemenata koje treba istraživanjem obuhvatiti uzorkom proizlazi tada iz obrasca:

$$n = \frac{n'}{1 + \frac{n'}{N}}. \quad (3)$$

Na ovom mjestu može se postaviti jedno praktično pitanje. Ono se sastoji u činjenici da se u obrascima (1) i (2) traži poznavanje standardne devijacije osnovnog skupa. U načelu, o toj karakteristici na početku israživanja nema se nikakvih informacija, a do nje



se ne može doći ni pomoću podataka iz uzorka jer se ulazi tek u fazu projektiranja uzorka. Ta nezgoda rješava se ili grubom procjenom standardne devijacije skupa kada se koristi svako znanje o tom skupu. Ako se u toj gruboj procjeni uzme veća standardna devijacija od stvarne, posljedica rezultira većim uzorkom od potrebnog što daje precizniju procjenu karakteristike. No, ukoliko se o skupu nema nikakvih informacija, odnosno ako se s njime do sada nije susretalo, tada se može gruba ocjena standardne devijacije skupa dobiti preko tzv. pivot uzorka (6).

Kada je skup stratificiran, tada se broj jedinica iz svakog stratuma, koje će ući u uzorak, uzima na osnovi navedenih obrazaca. Međutim, tu su mogući i drugačiji pristupi, kao što su proporcionalni stratificirani izbor, sistemski izbor i sl.

#### 4. REGRESIJSKI MODELI

Modeli odlaganja, čuvanja i zaštite dokumentacije poslovnih subjekata mogu se temeljiti i na regresijskoj analizi. Ovom analizom istražuje se varijabilitet dviju ili više pojava od kojih neke mogu biti zavisne a neke nezavisne. Za svaku od pojava unaprijed se zna ili pretpostavlja da može biti nezavisna ili zavisna varijabla. Prema tome, zadatak je regresijske analize da otkrije funkcionalni oblik, tj. regresijski model kojemu se najviše približava kvantitativno slaganje varijacija promatranih pojava. Nadalje, on treba da pokaže kako se zavisne varijable mijenjaju u ovisnosti o nezavisnim te da na osnovi stupnja slaganja njihovih varijacija omogući ocjenu i predviđanje ponašanja zavisnih varijabli. U istraživanju odlaganja, čuvanja i zaštite dokumentacije pokazalo se da se može koristiti mnogostruka ili multipla linearna regresija (3). Model multiple linearne regresije može se postaviti na slijedeći način (7):

$$Y' = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p, \quad (4)$$

gdje su  $\beta_1, \beta_2, \dots, \beta_p$  parametri, a  $X_1, X_2, \dots, X_p$  nezavisne varijable. Za određivanje vrijednosti parametara primjenjuje se metoda najmanjih kvadrata, dok su varijable modela izražene raznim kvantitativnim svojstvima dokumentacije. Parametrom  $\beta_1$  mjeri se prosječna ili očekivana promjena u  $Y'$  kada se  $X_1$  povećava (umanjuje) za 1 jedinicu, a ostali  $X_i$  ostaju nepromijenjeni. Ista analogija vrijedi i za ostale parametre. Zbog toga se parametri  $\beta_i$  nazivaju koeficijentima parcijalne regresije  $Y'$  u zavisnosti o  $X_i$ . Za razmatranje značajnosti jednadžbi multiple regresije, kada su poznate vrijednosti drugih varijabli, bitna je standardna greška procjene, koeficijent multiple determinacije i koeficijent multiple korelacije. Ove karakteristike mogu se izračunati korištenjem ovih obrazaca (6):

$$\sigma_y = \left[ \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N} \right]^{1/2}, \quad (5)$$

$$R^2 = \frac{\sum_{i=1}^N (Y' - \bar{Y})^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2}, \quad (6)$$

$$R = \left[ \frac{\sum_{i=1}^N (Y' - \bar{Y})^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \right]^{1/2}. \quad (7)$$

Koeficijentom multiple korelacije R izražava se smjer i jakost linearne veze između pojave Y i pojava X<sub>i</sub>. Vrijednosti ovog koeficijenta kreću se u granicama ±1. Ukoliko su vrijednosti R bliže krajnjim granicama, tada je veza među pojavama statistički vrlo značajna. U protivnom govori se o labavoj statističkoj vezi. Naime, kada je R=0, tada između pojave Y i pojava X<sub>i</sub> neme statističke veze.

Prilikom korištenja multiple linearne regresije često se može postaviti i pitanje koje su varijable X najznačajnije u određivanju varijable Y'. Može se reći da nema jedinstvenog ili potpuno odgovorajućeg odgovora te se ovom problemu može prići na nekoliko načina. Na osnovi modela (4) može se napisati slijedeći izraz (7):

$$\sigma_{Y'}^2 = \beta_1^2 \sigma_1^2 + \beta_2^2 \sigma_2^2 + \dots + \beta_p^2 \sigma_p^2 + \sigma^2 \quad (8)$$

gdje se sa  $\sigma_i^2$  označava varijanca varijabli X<sub>i</sub>. Omjer

$$\tau = \frac{\beta_i^2 \sigma_i^2}{\sigma_{Y'}^2} \quad (9)$$

izražava frakciju varijance  $Y'$  koja se pripisuje njezinoj linearnoj regresiji u odnosu na  $X_i$ . Ova frakcija može se sa sigurnošću smatrati relativnim značenjem  $X_i$ . Kod slučajno odabranog uzorka parametri  $\beta_i$  mijenjaju se s  $b_i$  te omjeri  $b_i^2 \sum x_i^2 / \sum y_i^2$  predstavljaju procjene iz uzorka tih frakcija.

U izgradnji modela odlaganja, čuvanja i zaštite dokumentacije nezavisne varijable  $X_i$  mogu imati različita značenja. Tako npr. one mogu biti količina dokumenata po vrsti, količina dokumenta s jednom ili obje strane zapisa, količina dokumenta po određenim veličinama formata itd. Nadalje, nezavisne varijable mogu također biti i broj zaposlenih, troškovi poslovanja, dobit isl.

## 5. ZAKLJUČAK

Da se postavi valjani model odlaganja, čuvanja i zaštite dokumentacije poslovnih subjekata, neophodna su opsežna, dugotrajna i s troškovnog stajališta skupa istraživanja. Svakako, bilo bi pretenciozno da se u istraživanje uključe svi poslovni subjekti s područja privrede i neprivrede. Zato, neophodno je prvenstveno sve privredne subjekte podijeliti na stratume ili klase te iz njih korištenjem metode uzoraka dobiti tražene karakteristike. Model koji će osigurati generiranje karakteristika za racionalno odlaganje, čuvanje i zaštitu dokumentacije temelji se na regresijskoj analizi. Za konstrukciju regresijskih modela potrebno je definirati zavisne i nezavisne varijable te postaviti funkcionalne relacije kao i izvesti proračun svih parametara. Ukoliko se preko mjera jakosti veze ustanovi statistička značajnost tih veza, tada se regresijski modeli mogu staviti u implementaciju.

## LITERATURA

1. Čaval J., STATISTIČKE METODE, u privrednim i društvenim istraživanjima, Sveučilište u Rijeci, Rijeka, 1981.
2. Hemmerle J.W., STATISTICAL COMPUTATIONS ON A DIGITAL COMPUTER, Blaisdell Publishing Company, Toronto-London, 1967.
3. Hutinski Ž., MODEL PRIMJENE MIKROFILMA U ZAŠTITI DOKUMENTACIJE PODUZEĆA, doktorska disertacija, Fakultet organizacije i informatike, Varaždin, 1991.
4. Hutinski Ž., KOMPARATIVNE PREDNOSTI MIKROFILMA U UREDSKOM POSLOVANJU, IBI br.33, Zagreb, 1991, 2-6.



5. Petz B., OSNOVNE STATISTIČKE METODE ZA NEMATEMATIČARE, SNL, Zagreb, 1985.
6. Serdar V., Šošić I., UVOD U STATISTIKU, Školska knjiga, Zagreb, 1981.
7. Snedecor G.W., Cochran W.G., STATISTICAL METHODS, The Yowa State University Press, Yowa, 1967.

Primljeno: 1991-10-22

*Bojanić M., Die Anwendung der statistischen Methoden bei der Forschung der Ablage, der Aufbewahrung und des Schutzes der Dokumentation*

#### ZUSAMMENFASSUNG

*Der Aufbau des Modelles der Ablage, der Aufbewahrung und des Schutzes der Dokumentation von geschäftlichen Subjekten verlangt unter anderem die Aufstellung der regressiven Modelle. Durch Definieren von abhängigen und unabhängigen Variablen kann man Modelle sichern, auf Grund deren man über die Ablage, die Aufbewahrung und den Schutz der Dokumentation sprechen kann.*

(Prijevod: Vesna Šimunić)