

Analiza dokładności ocen wartości średnich cech małych firm¹

W niniejszej pracy przedstawiono na odpowiednim materiale statystycznym praktyczny sposób racjonalnego wyznaczania liczebności prób z warstw. Przedstawiono także krótką analizę efektywności proponowanych rozwiązań.

Dane o działalności małych firm pochodzą z badania reprezentacyjnego oznaczanego symbolem DG-3 przeprowadzonego przez Wojewódzki Urząd Statystyczny w Katowicach w 1993r. Niniejszą analizę ograniczono do cech, których symbole niżej wyjaśniono. Badana populacja składała się z 167094 firm. Wylosowana próba liczyła 6656 firm, co stanowi 4% liczby wszystkich elementów populacji.

- d1p1ar1 - liczba właścicieli i współwłaścicieli
- d1p1ar2 - liczba właścicieli i współwłaścicieli-kobiet
- d1p1br1 - uczniowie ogółem
- d1p1br2 - uczennice
- d1p1cr1 - osoby, z którymi zawarto umowę agencyjno-prowizyjną
- d1p1cr2 - kobiety, z którymi zawarto umowę agencyjno-prowizyjną
- d1p1dr1 - osoby zatrudnione na podstawie umowy o pracę
- d1p1dr2 - kobiety zatrudnione na podstawie umowy o pracę
- d1p1er1 - razem zatrudnieni
- d1p1er2 - razem zatrudnione kobiety
- d1p2- liczba osób wykonująca prace na umowę zlecenie
- d1p3- liczba zatrudnionych w przeliczeniu na etaty
- d1p4- kwota wynagrodzeń brutto
- d1p5a- największe wynagrodzenie brutto
- d1p5a- najmniejsze wynagrodzenie brutto
- d1p5- średnie wynagrodzenie brutto

Celem analizy jest ocena dokładności estymacji wartości średnich wymienionych zmiennych. Z metodologicznego punktu widzenia gromadzenie danych odbywało się poprzez bezzwrotne losowanie prób prostych z warstw, przy czym warstwy były tworzone na podstawie cech określających rodzaj działalności firmy zgodny z Europejską Klasyfikacją Danych (EKD).

Wykorzystując ogólnie dostępne estymatory wartości średnich i ich błędów szacunku wyliczono na podstawie próby warstwowej oceny średnich badanych cech oraz rząd ich odchyłeń od nieznanych średnich w populacji². Dodajmy, że w tym podejściu nie uwzględniliśmy tzw. błędów nielosowych wynikających z braku odpowiedzi na ankietę wysyланą do przedsiębiorstwa.

¹Praca obejmuje część rezultatów otrzymanych w wyniku realizacji projektu badawczego nr 1 H02B 015 10 pt. *Metady powtarzalnych badań cech populacji* finansowanego przez Komitet Badań Naukowych.

²Wiele ciekawych własności procedur wnioskowania na podstawie prób warstwowych znajdziemy w pracy Steczkowskiego (1995).

Wartość średnia cechy w populacji jest oceniana za pomocą następującego estymatora:

$$\bar{X}_w = \sum_{h=1}^H w_h \bar{X}_h$$

gdzie:

$$w_h = \frac{N_h}{N}, \quad N = \sum_{h=1}^H N_h$$

Przez H oznaczono liczbę warstw, liczebność h -tej warstwy przez N_h . Średnią z n_h -elementowej próby losowanej bezzwrotnie z h -tej warstwy określa wzór:

$$\bar{X}_h = \sum_{i=1}^{n_h} X_{hi}$$

Symbolem X_{hi} oznaczono i -tą obserwację zmiennej w h -tej warstwie. Wiadomo, że średnia z próby warstwowej \bar{X}_w daje nieobciążone oceny średniej zmiennej w populacji. Wariancja estymatora \bar{X}_w jest oceniana za pomocą następującej statystyki:

$$\hat{V} = \sum_{h=1}^H w_h^2 \frac{N_h - n_h}{N_h n_h} S_h^2$$

gdzie:

$$S_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (X_{hi} - \bar{X}_h)^2$$

Oceny wartości średnich i błędów ich estymacji znajdujemy w tabelicy 1. W jej trzeciej kolumnie zapisano ocenę wariancji estymatora średniej. Z kolei w następnej kolumnie zapisano jej pierwiastek, który jest oceną przeciętnego błędu szacunku średniej. W końcu w ostatniej kolumnie tabelicy 1 znajdujemy ocenę błędu względnego szacunku średniej. Przykładowo, średnia wynagrodzeń brutto (zmienna d1p4) wynosi 2060.45 zł i odchyła się ta wartości od nieznannej szacowanej wartości średniej wynagrodzeń brutto przeciętnie o 137.73 zł in plus bądź in minus. Błąd ten stanowi 6.68% oceny wartości średniej. Należy uznać, że błąd ten jest zbyt duży, bowiem zwykle przyjmuje się poziom dopuszczalnego błędu nie powinien być większy od 5%.

Estymacja wartości średnich była prowadzona na podstawie prób o liczebnościach ustalanych przez Główny Urząd Statystyczny. Rozmiary tych prób bynajmniej nie są proporcjonalne do liczebności warstw, co zaleca się jako najprostszy zabieg prowadzący zwykle do podniesienia dokładności estymacji. Proporcjonalny sposób wyliczania warstw wyjaśnia następujące wyrażenie:

$$n_{p,h} = N w_h, \quad h=1, \dots, H$$

Na podstawie danych z tabelicy 1 wyliczono oceny błędów szacunku średniej przy założeniu, że liczebności losowanych prób byłyby proporcjonalne do rozmiarów warstw. Wyniki

obliczeń zamieszczono w tablicy 2, przy czym w jej drugiej kolumnie zamieszczono ocenę przeciętnego błędu szacunku, a w trzeciej kolumnie ocenę względnego błędu szacunku.

Tablica 1. Oceny dokładności estymacji średnich.

zmienna	średnia	wariancja estymatora	błąd przeciętny	błąd względny
d1p1ar2	.51	.0002	.0143	2.82
d1p1ar1	1.40	.0002	.0152	1.09
d1p1br2	.08	.0001	.0122	15.79
d1p1br1	.14	.0003	.0164	11.75
d1p1cr2	.02	.0000	.0060	33.66
d1p1cr1	.04	.0002	.0130	30.08
d1p1dr2	.42	.0006	.0254	6.10
d1p1dr1	.78	.0018	.0420	5.40
d1p1er2	1.02	.0014	.0373	3.66
d1p1er1	2.36	.0029	.0542	2.30
d1p2	.17	.0003	.0186	10.83
d1p3	.74	.0016	.0404	5.49
d1p4	2060.46	18971.0481	137.7354	6.68
d1p5a	914.53	1971.7152	44.4040	4.86
d1p5b	552.81	529.8657	23.0188	4.16
d1p5	874.08			

Źródło: obliczenia własne.

Tablica 2. Ocena błędów estymacji średnich dla liczebności prób ustalanych w sposób proporcjonalny i optymalny.

zmienna	dobór prób proporcjonalny		dobór prób optymalny	
	błąd przeciętny	błąd względny	błąd przeciętny	błąd względny
d1p1ar2	.0074	1.5	.0083	1.6
d1p1ar1	.0080	.6	.0088	.6
d1p1br2	.0058	7.5	.0070	9.0
d1p1br1	.0082	5.8	.0095	6.8
d1p1cr2	.0028	15.9	.0032	18.0
d1p1cr1	.0069	15.9	.0070	16.1
d1p1dr2	.0167	4.0	.0136	3.3
d1p1dr1	.0248	3.2	.0205	2.0
d1p1er2	.0217	2.1	.0205	2.0
d1p1er1	.0307	1.3	.0294	1.2
d1p2	.0148	8.6	.0175	10.2
d1p3	.0240	3.3	.0212	2.9
d1p4	84.0264	4.1	71.3326	3.5
d1p5a	22.6592	2.5	23.3806	2.6
d1p5b	11.5441	2.1	12.5627	2.3

Źródło: obliczenia własne.

Ponadto w ostatnich dwóch kolumnach tablicy 2 znajdujemy oceny błędów oceny średnich przy założeniu, że liczebności prób losowanych z warstw były wyznaczone w sposób optymalny. Polega to na tym, że rozmiar próby losowanej z danej warstwy jest proporcjonalny do iloczynu frakcji elementów warstwy w populacji i oceny odchylenia standardowego zmiennej w warstwie, co wyjaśnia następujący wzór:

$$n_{o,h} = \frac{w_h S_h}{\sum_{k=1}^H w_k S_k}$$

Przyjęto, że optymalne liczebności prób wyznaczamy w stosunku do ocen odchyłeń standardowych w warstwach wynagrodzeń brutto w firmach (zmienna d1p4). Z tablicy 2 wynika m.in., że ocena względnego błędu estymacji średnich wynagrodzeń brutto jest już niższa od dopuszczalnego poziomu 5%, bo wynosi 4.1% dla proporcjonalnej lokalizacji prób i 3.5% dla optymalnej lokalizacji prób w warstwach.

Tablica 3: Współczynniki względnej efektywności.

zmienna	metoda ustalenia liczebności prób		
	proporcjonalna EFP_O	optymalna EFZ_O	rzeczywista EFZ_P
d1plar2	-10.7	71.6	92.2
d1plar1	-9.0	72.4	89.4
d1p1br2	-16.4	76.0	110.4
d1p1br1	-14.1	72.8	101.1
d1p1cr2	-11.7	86.9	111.5
d1p1cr1	-,8	87.1	88.6
d1p1dr2	22.7	86.2	51.8
d1p1dr1	11.7	88.9	69.1
d1p1er2	5.5	81.8	72.2
d1p1er1	4.4	84.1	76.4
d1p2	-15.3	6.1	25.2
d1p3	13.1	90.5	68.4
d1p4	17.8	93.1	63.9
d1p5a	-3.1	89.9	96.0
d1p5b	-8.1	83.2	99.4

Źródło: obliczenia własne.

W tablicy 3 dokonano porównania dokładności ocen średnich przy analizowanych trzech wariantach ustalania liczebności prób w warstwach. Obliczono względne odchylenia błędów szacunku dla branych pod uwagę metod wyznaczania liczebności prób. Oznaczając przez BP, BO i BZ oceny błędów średnich szacunku przeciętnej w populacji na podstawie średniej z prób losowanych z warstw o liczebnościach odpowiednio proporcjonalnych do rozmiarów warstw, optymalnych i rzeczywistych. Zamieszczone w tablicy 3 wielkości obliczono według następujących wzorów:

$$EFP_O=100\%(BP-BO)/BO,$$

$$EFZ_P=100\%(BZ-BP)/BP$$

$$EFZ_O=100\%(BZ-BO)/BO$$

Z analizy danych w tablicy 3 wynika, że zarówno warianty proporcjonalny jak i optymalny ustalania liczebności warstw prowadzi do znacznie większej dokładności estymacji przeciętnych badanych zmiennych niż to miało miejsce przy rzeczywistych rozmiarach losowanych prób. Zaznaczmy, że oprócz dwóch zmiennych (dip2 i d1p1dr2) zysk na dokładności estymacji przy stosowaniu wariantu proporcjonalnego ustalania liczebności prób wynosi co najmniej 60% w stosunku do rzeczywiście otrzymanej dokładności estymacji. W przypadku optymalnego sposobu wyznaczania rozmiarów prób otrzymujemy podobne wyniki.

Z przeprowadzonej analizy wynika jednoznacznie, że istnieje realna możliwość podniesienia dokładności ocen wartości średnich badanych cech charakteryzujących działalność małych firm. Mamy dwie możliwości: bądź ustalać liczebności prób losowanych z warstw proporcjonalnie do rozmiarów odpowiednich warstw, bądź wyznaczać je w sposób optymalny. Ze względu na prostotę należy wybrać pierwszy z wymienionych wariantów wyznaczania liczebności prób. Należy tu podkreślić, że proponowany sposób korekty sposobu wyznaczania liczebności losowanych prób bynajmniej nie prowadzi do wzrostu kosztów badania, bo suma liczebności prób losowanych z warstw pozostaje ta sama.

Określony sposób wyznaczania liczebności prób może prowadzić do wyliczenia ułamkowych rozmiarów prób. Wówczas należy zaokrąglić z nadmiarem otrzymane liczby. Ponadto należy również zalecić aby losować przynajmniej dwa elementy z każdej warstwy, co umożliwi wyliczenie oceny wariancji estymatora.

Dodatek

Prezentowane rezultaty były wyliczone za pomocą programu napisanego w języku macierzowym znanego pakietu statystycznego SPSS. Po to aby przeprowadzić te obliczenia należy dysponować dwoma modułami tego pakietu, a są nimi: BASE-SYSTEM i ADVANCED STATISTICS³. List prezentowanego programu pozwala wyliczyć wektor średnich z próby warstwowej i ocenia jego macierz wariancji i kowariancji.

```
SET MXLOOP 6700.
MATRIX.
GET XX /VARIABLES=d1p1ar1,d1p1ar2,d1p1br1,d1p1br2,d1p1cr1.
GET W/VARIABLES=warstwa2. /* zmienna warswująca
COMPUTE LZ=NCOL(XX). /* liczba zmiennych
GET NW/FILE='c:\wus\dg3_93\nrwar.sav'/VARIABLES=licz_war.
/* NW - wektor liczebności warstw
COMPUTE LW=NROW(NW). /* liczba warstw
COMPUTE LPOP=CSUM(NW). /* liczebność populacji
COMPUTE F=NW/LPOP. /* frakcje liczebności warstw
COMPUTE LP=NROW(XX). /* liczebność próby
COMPUTE V=MAKE(LZ,LZ,0).
COMPUTE SX=MAKE(1,LZ,0).
```

³ Bliższe informacje o innych modułach pakietu SPSS oraz o jego możliwościach użycia do różnych analiz statystycznych udziela firma COMPANION Oprogramowanie, ul. Królewska 57, 30-081 Kraków, tel.: (012)369680, tel/fax (012)360791, e-mail: mail@companion.krakow.pl

```

COMPUTE WLP=MAKE(1,LW,0). /*wektor liczebności prób losowanych z warstw
LOOP #I=1 TO LW.
  LOOP #K=1 TO LP.
    DO IF (W(#K)=#I).
      COMPUTE WLP(#I)=WLP(#I)+1.
    END IF.
  END LOOP.
END LOOP.
LOOP #I=1 TO LW.
  DO IF (WLP(#I)>0).
    COMPUTE H=0.
    COMPUTE M=MAKE(WLP(#I),1,0).
    LOOP #K=1 TO LP.
      DO IF (W(#K)=#I).
        COMPUTE H=H+1.
        COMPUTE M(H)=#K.
      END IF.
    END LOOP.
    COMPUTE X=XX(M,:).
    COMPUTE SR=CSUM(X)/H.
    COMPUTE SX=SX+SR*F(#I).
    DO IF (H>1).
      COMPUTE J=MAKE(H,1,1).
      COMPUTE X=X-J*SR.
      COMPUTE C=T(X)*X/(H-1).
      /*MSAVE C/TYPE=COV/VARIABLES=d1p1er1,d1p1er2,d1p4.
      COMPUTE V=V+C*F(#I)*F(#I)*(NW(#I)-H)/(NW(#I)*H).
    END IF.
  END IF.
END LOOP.
MSAVE SX/TYPE=MEAN/OUTFILE='c:\wus\dg3_93\SRED.sav'/VARIABLES=d1p1ar1,d1p1ar2.
COMPUTE SR=T(SQRT(DIAG(V))).
MSAVE SR/TYPE=MEAN/OUTFILE='c:\wus\dg3_93\SRED.sav'/VARIABLES=d1p1ar1,d1p1ar2.
MSAVE V/TYPE=COV/OUTFILE='c:\wus\dg3_93\SRED.sav'/VARIABLES=d1p1ar1,d1p1ar2.
COMPUTE b1=sqrt(TRACE(V)).
COMPUTE b2=DET(V).
COMPUTE L=EVAL(V).
COMPUTE L=T(L).
SAVE {b1,b2,L}/OUTFILE='c:\wus\dg3_93\PARAM.sav'/VARIABLES=tr,det,l1,l2,l3.
END MATRIX.

```

Bibliografia

Steczkowski J. (1995): Metoda reprezentacyjna w badaniach zjawisk ekonomiczno-społecznych. PWN, Warszawa.