

Praktyczne aspekty doboru próby

Dariusz Przybysz
Warszawa, 2 czerwca 2015

Określenie populacji

- Przed przystąpieniem do badania, wybraniem sposobu doboru próby konieczne jest precyzyjne określenie populacji, którą zamierzamy badać;
- Ma to istotne znaczenie dla określenia sposobu doboru próby, wskazania operatu losowania i innych szczegółowych kwestii powiązanych z realizacją badania.

Określenie populacji

Przykładowo populacją mogą być:

- Osoby mieszkające w Polsce, w wieku powyżej 18 roku życia (trzeba rozstrzygnąć wiele szczegółowych kwestii, np. , czy osoby mieszkające tymczasowo za granicą wchodzą w skład badanej zbiorowości),
- Gospodarstwa domowe w Polsce,
- Przedsiębiorstwa w Polsce określonego typu (np. z określonej branży lub określonej wielkości),
- Przedstawiciele elity biznesu,

Określenie populacji

- Osoby bezrobotne mieszkające w Polsce,
- Absolwenci uczelni wyższej,
- Radni gmin w Polsce,
- Szkoły,
- Uczniowie w szkołach,
- Czytelnicy określonej gazety lub czasopisma,
- Prace magisterskie złożone (obronione) w określonym roku

Operat losowania

W odniesieniu do określonej populacji należy poszukać adekwatnego operatu losowania, przykładowo:

- W odniesieniu do mieszkańców Polski najbardziej odpowiednia wydaje się baza PESEL
- W odniesieniu do przedsiębiorstw można się posłużyć bazą REGON
- W przypadku badań edukacyjnych niezbędny jest spis szkół wraz z ich charakterystykami (wielkość szkoły, typ szkoły: publiczna, społeczna, prywatna)

Operat losowania

W przypadku pewnych populacji nie istnieją adekwatne operaty losowania, przykładowo jeśli chcemy przebadać czytelników określonej gazety operat taki nie istnieje.

W takich sytuacjach jedynym rozwiązaniem jest wybranie próby kwotowej.

Operat losowania

Możliwość skorzystania z danego operatu losowania może napotykać na pewne ograniczenia:

- Operat może być niepełny, zawierać usterki , może być nieaktualny,
- Możliwość wykorzystania danego operatu może wiązać się z określonymi kosztami finansowymi,
- Wylosowanie próby na podstawie danego operatu lub opracowanie danego operatu może być czasochłonne. Należy to uwzględnić planując badanie, jego harmonogram.

Losowanie próby a dokładność wyników

- Celem losowania próby jest uzyskanie możliwie dokładnych wyników przy minimalizacji kosztów realizacji badania,
- Badacz musi wziąć pod uwagę takie czynniki jak możliwości logistyczne realizacji badania w określonym czasie, aspekty zwiększające szansę na wysoką realizację próby.

Dokładność oszacowania wyników

Przez dokładność oszacowania rozumiemy długość przedziału ufności. Przykładowo: chcemy oszacować średnią wzrostu **w populacji**. Przypuśćmy, że w oparciu o wyniki na podstawie **próby losowej** 900-osobowej możemy sformułować komunikat:

Szacujemy, że średnia wzrostu w populacji wynosi 168 +/- 2cm. Szacujemy ją na poziomie ufności 0,95.

- W powyższym komunikacie dokładność oszacowania to 2 cm. Poziom ufności wskazuje, że przyjęta metoda działa poprawnie z prawdopodobieństwem 0,95, tj. losując 100 prób 900 osobowych z tej samej populacji w przypadku 95 prób nasze oszacowanie (przedział ufności) będzie poprawne tj. będzie obejmowało średnią w populacji.

Dokładność oszacowania frakcji

Podobnie szacujemy odsetki (frakcję). Przykładowo: chcemy oszacować odsetek osób **w populacji** znających język angielski. Przypuśćmy, że na podstawie **próby losowej** możemy sformułować komunikat:

Szacujemy, że odsetek osób znających język angielski w populacji wynosi 30% +/- 3%. Szacujemy ten odsetek na poziomie ufności 0,99.

- W powyższym komunikacie dokładność oszacowania to 3%. Poziom ufności wskazuje, że nasza metoda działa poprawnie z prawdopodobieństwem 0,99.

Próba prosta a dokładność wyników

Zakładając, że jednostki losujemy w sposób prosty, tj. prawdopodobieństwo wylosowania każdej osoby z populacji jest takie samo (i losujemy każdą osobę oddzielnie z całej populacji), dokładność oszacowania średniej danej zmiennej (cechy) np. zarobków zależy od:

- Poziomu ufności, tj. prawdopodobieństwa, że procedura którą stosujemy daje poprawne oszacowanie: im wyższy poziom ufności tym dokładność oszacowania mniejsza.

Próba prosta a dokładność wyników

- Zróżnicowania (wariancji) analizowanej zmiennej (cechy). Im wariancja zmiennej większa (np. zróżnicowanie zarobków), tym oszacowanie jest mniej dokładne,
- Wielkości próby, im próba większa tym wyniki bardziej dokładne.

Wiązkowanie a koszty i dokładność wyników

- Losując próbę w sposób prosty, narażamy się na wysokie koszty docierania ankieterów do poszczególnych respondentów. Z tego względu często losuje się wiązki osób w ramach pewnych miejscowości;
- Ma to jednak wpływ na dokładność oszacowania, np. wylosowanie 10 osób mieszkających na wsi niezależnie od siebie, daje na ogół bardziej dokładne wyniki niż wylosowanie 2 wsi i 5 osób w ramach każdej z tych wsi; z drugiej strony zastosowanie pierwszej strategii jest droższe.

Wiązkowanie a koszty i dokładność wyników

- Wiązkowanie pogarsza dokładność oszacowania tym bardziej, im bardziej homogeniczni są mieszkańcy tych miejscowości (w odniesieniu do analizowanych zmiennych)
- W praktyce okazuje się jednak, że wiązkowanie może być opłacalne: homogeniczność wiązek można zrekompensować zwiększeniem liczebności próby (na co można sobie pozwolić w ramach tego samego budżetu, jeśli mamy znaczącą redukcję kosztów docierania do respondentów przez ankieterów w przypadku wiązek)

Warstwowanie a zwiększenie dokładności

- Można zwiększyć dokładność oszacowania nie zwiększając równocześnie wielkości próby przez tzw. warstwowanie;
- Populację dzieli się na możliwe homogeniczne podzbiorowości (warstwy) ze względu na cechy, które analizujemy; warstwy te powinny się różnić między sobą ze względu na te cechy;

Warstwowanie a zwiększenie dokładności

Potem określamy ile osób ma być wylosowanych z danej warstwy. Jedną strategią jest losowanie z każdej z tych warstw wielkości próby proporcjonalnie do liczby osób w populacji w danej warstwie;

Dodatkowo można poprawić dokładność oszacowania losując relatywnie więcej osób z warstw bardziej heterogenicznych a mniej osób z warstw bardziej homogenicznych (relatywnie w stosunku do alokacji proporcjonalnej). Określają to odpowiednie formuły statystyczne.

Warstwowanie - - problemy w praktyce

- Na ogół nie dysponujemy bardzo dokładną wiedzą na temat zróżnicowania cech w badanej populacji,
- Przykładowo, w bazie PESEL mamy informacje o płci, wieku i charakterystyce miejsca zamieszkania (wieś-miasto; wielkość miejscowości)
- Warstwowanie może w praktyce opierać się tylko na tych zmiennych; warstwowanie może jednak być efektywne jeśli cechy, które analizujemy są skorelowane z powyższymi informacjami dostępnymi w bazie PESEL.

Europejski Sondaż Społeczny

runda 7 – przykład doboru próby

- W polskiej edycji schemat doboru próby został opracowany przez Zbigniewa Sawińskiego i w kolejnych edycjach ulega nieznacznym modyfikacjom
- Wymogiem dla każdego kraju uczestniczącego w ESS jest realizacja próby efektywnej o wielkości 1500 osób. Tj. próby prostej liczącej 1500 osób lub próby losowanej wg. innego schematu (wiązkowanej, warstwowanej itd.) o innej liczebności, która pozwala na uzyskanie takiej samej dokładności jaką uzyskalibyśmy losując 1500 osób w sposób prosty.

ESS runda 7 –dobór próby

- Zgodnie z założeniami badania liczebność próby losowanej została wyznaczona w taki sposób, aby zapewnić przynajmniej 1756 zrealizowanych wywiadów (co przekłada się na liczebność efektywną 1500 zrealizowanych wywiadów)
- Oszacowując liczebność próby założono, że poziomem realizacji (response rate) – tj. stosunek liczby wywiadów zakończonych sukcesem do liczby wylosowanych osób, które należą do badanej populacji – wyniesie w przybliżeniu 70%.

ESS runda 7 –dobór próby

- Dodatkowo uwzględniono, że operat PESEL obejmuje osoby, które nie wchodzą w zakres badanej populacji (tzw. kategorii *ineligibles*, m. in. osoby mieszkające w momencie badania za granicą, osoby zmarłe pomiędzy momentem losowania próby a rozpoczęciem badania).
- Na podstawie dostępnych źródeł i poprzednich edycji ESS ich odsetek został oszacowany na 7,1%. Z tych względów (przewidywanego 70% odsetka realizacji i uwzględnienia *ineligibles*) liczebność losowanej próby została ustalona na 2716 osób.

ESS runda 7 –dobór próby

- W ESS uwzględnia się niejednakowe współczynniki realizacji w miejscowościach o różnej wielkości.
- Ich estymacji dokonano na podstawie analizy pracy sieci ORBS IFiS PAN w ostatnio realizowanych badaniach, a także na podstawie informacji o trudnościach w realizacji badań występujących w innych sieciach ankieterskich.
- Liczba osób losowanych w poszczególnych klasach wielkości miejscowości została przemnożona przez odwrotność przewidywanego poziomu realizacji.

ESS runda 7 –dobór próby

Specyfiką doboru próby do badania Europejski Sondaż Społeczny jest podział populacji ludności Polski na dwie części, w ramach których losowanie przebiegało według odmiennych schematów.

- Pierwsza część objęła populację osób mieszkających w miastach o wielkości ponad 50 tys. mieszkańców,
- Druga część objęła osoby mieszkające w pozostałych miejscowościach (miasta poniżej 50 tys. mieszkańców oraz wsie).

ESS runda 7 –dobór próby

Dobór próby w miastach ponad 50 tys. mieszkańców

Próba w tej części populacji została dobrana jako **próba prosta losowana bezzwrotnie w każdym mieście osobno**. Mówiąc inaczej, próba w tej części miała charakter **warstwowy**, przy czym *każde miasto stanowiło odrębną warstwę*. Osoby losowane były bezpośrednio spośród wszystkich mieszkańców danego miasta, z pominięciem fazy wiązkania. Schemat prostej próby losowej zastosowano z tego względu, że jest on bardziej efektywny niż wiązkanie (daje większą dokładność) a w przypadku dużych miast nie jest wyraźnie droższy.

ESS runda 7 –dobór próby

Dobór próby w miastach poniżej 50 tys. mieszkańców oraz na wsi

Zastosowano odmienny schemat losowania próby, a mianowicie dwustopniowy losowy dobór warstwowy. W przypadku rozważanej warstwy w pierwszej fazie losowano nie pojedyncze osoby lecz miejscowości, z których następnie losowano wiązki 4-osobowe.

ESS runda 7 –dobór próby

Dobór próby w miastach poniżej 50 tys. mieszkańców oraz na wsi

Dobór miejscowości miał charakter doboru warstwowego, przy czym warstwy wyróżniono ze względu na (i) kategorię wielkości miejscowości (wieś, miasto do 10 tys. mieszkańców, miasto 10-20 tys. mieszkańców, miasto 20-50 tys. mieszkańców) oraz (ii) województwo.

ESS runda 7 –dobór próby

Dobór próby w miastach poniżej 50 tys. mieszkańców oraz na wsi

Łącznie wyróżniono więc 64 warstwy. W obrębie każdej warstwy miejscowości zostały dobrane metodą losowania ze zwracaniem, przy czym prawdopodobieństwa losowania poszczególnych miejscowości przyjęto jako proporcjonalne do ich wielkości (tj. proporcjonalnie do liczby mieszkańców w wieku 15 lub więcej).

ESS runda 7 –dobór próby

**Dobór próby w miastach poniżej 50 tys. mieszkańców
oraz na wsi**

W kolejnym etapie w ramach każdej z wylosowanych miejscowości (miasta lub wsi) losowana była wiązka 4 osobowa tj. losowano 4 osoby w sposób prosty bezzwrotny.

Europejski Sondaż Społeczny

runda 7 – przykład doboru próby

- Aby określić liczbę osób losowanych z poszczególnych miejscowości wykorzystuje się informacje z danych oficjalnych Głównego Urzędu Statystycznego
- Jednostką o której posiadamy informacje jest gmina
- Losujemy gminy z prawdopodobieństwami proporcjonalnymi do wielkości gmin. Specyfikując parametry próby z bazy PESEL wskazujemy, że z danej gminy ma być wylosowana jedna miejscowość (z prawdopodobieństwem proporcjonalnym do wielkości miejscowości w gminie)

Europejski Sondaż Społeczny

runda 7 – przykład doboru próby

- Bardzo istotne jest prawidłowe określenie zasad doboru próby we wniosku składanym do Departamentu Spraw Obywatelskich MSW. Badacz może wylosować gminy i określić zasady doboru próby, natomiast z powodów oczywistych nie losuje konkretnych osób z bazy PESEL.
- Bardzo istotną kwestią jest wysoka realizacja próby. Niski odsetek realizacji może pogarszać wyniki w sposób trudny do przewidzenia, nie mamy bowiem informacji o osobach nieprzebadanych. W wielu sytuacjach bardziej celowe może być wylosowanie mniejszej próby i przeznaczenie środków finansowych na to zwiększyć szansę na wysoką realizację próby.