



Sondaż. Między próbą a populacją

- [Wprowadzenie](#)
- [Badanie 1](#)
- [Film](#)
- [Audiobook](#)
- [Badanie 2](#)
- [Dla nauczyciela](#)



Sondaż. Między próbą a populacją

Źródło: domena publiczna.

Materiał jest częścią serii „**Statystycznie rzecz biorąc**”.

„Badania zostały przeprowadzone na próbie 1230 mieszkańców małych miast” – z takimi lub podobnymi sformułowaniami często spotykamy się, analizując badania opinii społecznej. Ale czym w zasadzie jest ta próba? I czy odpowiedzi grupy ludzi mogą świadczyć o opinii całej populacji?

Twoje cele

- Scharakteryzujesz relacje między próbą a populacją i określisz parametry poprawnie dobranej próby do badań sondażowych.
- Zaprojektujesz próbę statystyczną do sondażu.
- Przeprowadzisz krótki sondaż o tematyce politycznej na wybranej samodzielnie próbie uczniów swojej szkoły.

Dla zainteresowanych



**Badania opinii publicznej:
historia, metody i ośrodki**

Badanie opinii publicznej

**Kształtowanie się opinii
publicznej**



**Wpływ opinii publicznej
na decyzje polityczne**



**Opinia publiczna
a polityka zagraniczna**



**Badania opinii publicznej:
historia, metody i ośrodki**



Badanie opinii publicznej



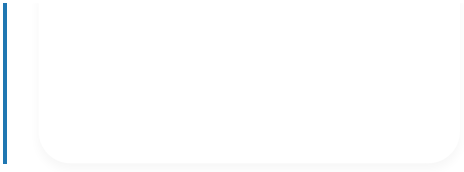
**Rzeczywistość sondaży
a rzeczywistość
w sondażach**



**Sondaż. Między prawdą
a pół...**



**Sondaż: między pytaniem
a odpowiedzią**



Badanie 1

Eksperyment: w celu przeprowadzenia pierwszego sondażu z klasy wybieramy tylko trzy osoby, którym zostanie zadane pytanie. Wyniki zapisujemy.

Odpowiedz na pytanie, wybierając jedną z zaproponowanych odpowiedzi. Wariant „trudno powiedzieć” wybierz tylko wtedy, gdy rzeczywiście nie potrafisz się zdecydować na żadną z pozostałych odpowiedzi.

Ćwiczenie 1

Jak myślisz, czy najbliższe wybory prezydenckie w Polsce zostaną rozstrzygnięte w pierwszej turze?

☐ Zdecydowanie tak – zostaną rozstrzygnięte w pierwszej turze.

☐ Najprawdopodobniej zostaną rozstrzygnięte w pierwszej turze.

☐ Nie, będzie potrzebna druga tura.

☐ Trudno powiedzieć.

Ćwiczenie 2

Rozważ następujące problemy i odpowiedz na pytania.

Co jest potrzebne do poznania opinii jakiejś zbiorowości (klasy, szkoły, mieszkańców miasta, społeczeństwa, na przykład polskiego)? Czy koniecznie trzeba pytać wszystkich członków danej zbiorowości?

Jeśli dojdiesz do wniosku, że nie trzeba pytać wszystkich, aby się zorientować – w tym wypadku – w opiniach na temat spodziewanego przebiegu wyborów, to minimalnie ile osób musi odpowiedzieć na pytanie sondażowe? Dlaczego nie wystarczy zapytać jednej osoby?

Jakie konsekwencje dla wyników sondażu prezydenckiego może mieć fakt przeprowadzenia badania na przykład w trakcie parady równości, wśród jej uczestników? Przy czym załóż, że pytanie sondażu zostanie zadane dokładnie takiej liczbie respondentów, jaką uznasz za wystarczającą.

A gdyby zapytać wszystkich mających zamiar głosować, na kogo oddadzą swój głos w wyborach, to czy uzyskamy pewność, że rzeczywisty wynik głosowania będzie dokładnie taki jak deklaracje? Dlaczego tak lub dlaczego nie?

Przedyskutuj swoje odpowiedzi z koleżankami i kolegami z klasy. Przeanalizuj, czy bardzo się różnicie w swoich odpowiedziach. Ustalcie wspólnie, z czego to wynika.

Film

Polecenie 1

Zapoznaj się z filmem i oceń w jego kontekście wartość poznawczą sondażu przeprowadzonego w Badaniu 1.

Film dostępny pod adresem <https://zpe.gov.pl/a/D11bL7cE>

Źródło: Englishsquare.pl sp. z o.o., licencja: CC BY-SA 3.0.

Film nawiązujący do treści materiału Sondaż - między próbą a populacją.

Ćwiczenie 1

Określ parametry próby dla sondażu przeprowadzanego w twojej klasie, przy założeniu, że próba ma stanowić 60% całej klasy i ma być reprezentatywna ze względu na płeć. Dodatkowo opisz sposób doboru jednostek empirycznych do badań, który pozwoli ci uniknąć błędu systematycznego.

Polecenie 1

Zapoznaj się z audiobookiem i rozważ sposób skonstruowania próby do sondażu na temat wyborów prezydenckich, jaki można by przeprowadzić w twojej szkole za zgodą dyrektora placówki.

Audiobook można wysłuchać pod adresem: <https://zpe.gov.pl/b/PvnObVWOp>

Próba: reprezentatywność i losowość a liczebność

Czym jest reprezentatywność próby i jaką odgrywa rolę w badaniach empirycznych wspartych statystyką, łatwo dostrzec, analizując jej zaprzeczenie. Korzystając na co dzień z zasobów sieciowych, wiecie, że w różnych miejscach internetu możecie się natknąć na pytania sondażowe poświęcone... właściwie wszystkiemu. To bardzo łatwy „(...) i bardzo tani sposób by zadać użytkownikom pytania. Ale kluczem jest to, że te pytania zadajemy właśnie internautom. Tymczasem zgodnie z wynikami najnowszej »Diagnozy Społecznej 2013« 75,7% Polaków posiada w domu dostęp do internetu, z czego 60,8% z niego korzysta. Pomijanie tak dużej grupy osób, rzędu 30–40%, zakłamuje nasze wyniki. Także dlatego, że dostęp do internetu rzadziej mają osoby mieszkające na wsi (ze względu na miejsce zamieszkania) czy starsze (ze względu na niższe kompetencje komputerowe), które to osoby jednocześnie mogą reprezentować elektorat określonych partii. Tym samym nasze badanie przeszacowuje wyniki jednych i nie doszacowuje wyników innych partii politycznych, ponieważ nie uwzględnia zjawiska wykluczenia cyfrowego. Podobne problemy mogą wystąpić, gdy korzystamy z badań telefonicznych (brak telefonu) czy ankiet pocztowych (utrudniony dostęp do placówek pocztowych). Klasycznym przykładem błędnych danych wynikających ze złego doboru próby jest sondaż magazynu informacyjnego »Literary Digest« z 1936 r. Dobór próby oparto na książkach telefonicznych i listach rejestracyjnych samochodów. Do respondentów wysyłano pocztą papierowy kwestionariusz. Wyniki sondażu wskazywały na przegraną Roosevelta (43%), podczas gdy w rzeczywistych wyborach uzyskał największą w historii przewagę głosów nad kontrkandydatem (61% poparcia). Tak znaczący błąd wynikał właśnie z doboru próby – w sondażu wzięło udział zbyt wiele osób zamożnych, czyli posiadających samochód lub telefon”.

Czym jest zatem reprezentatywność?

Earl Babbie w książce pod tytułem „Badania społeczne w praktyce” uważa za reprezentatywną próbę taką, której zagregowane cechy „ściśle odzwierciedlają te same zagregowane cechy populacji. Jeśli na przykład 50% populacji stanowią kobiety, to próba musi zawierać »w przybliżeniu« 50% kobiet, aby była reprezentatywna (...). Zauważmy, że próby nie muszą być reprezentatywne pod każdym względem; reprezentatywność ogranicza się do tych cech, które są związane z głównymi celami badania. Jednak nie zawsze z góry wiemy, które cechy są istotne”.

Próba nie jest zatem reprezentatywna „po prostu”, „ale zawsze ze względu na coś”. Może być reprezentatywna ze względu na rozkład cech (a więc strukturę populacji). I to niekoniecznie wszystkich. Przy doborze kwotowym, na przykład, konstruowanym ze względu na takie cechy jak: płeć, wiek i wykształcenie, z pewnością będzie reprezentatywna tylko na owe cechy (płeć, wiek i wykształcenie). Nie musi, choć może, być reprezentatywna ze względu na pozostałe cechy badanych osób. Ponadto może być reprezentatywna ze względu na zależności pomiędzy cechami lub na wartości poszczególnych cech. W tym miejscu wypada przypomnieć nasz pierwszy sondaż. Owa trzy osobowa próba z pewnością nie mogła być reprezentatywna w tym znaczeniu (ze względu na wartości poszczególnych cech). Konstruując pytanie badające taką cechę, jak opinia w sprawie wyborów prezydenckich, założyliśmy, że w populacji będzie ona miała cztery wartości (którym odpowiadały cztery odpowiedzi na pytanie). W trzyosobowej próbie natomiast mogą zostać wybrane maksymalnie trzy różne odpowiedzi. Czy zatem badacz nie ma szans uzyskać próby reprezentatywnej tak „w ogóle”? Odpowiadając na to pytanie, musimy przejść do zagadnienia losowości.

„Podstawowa zasada doboru losowego jest następująca: próba jest reprezentatywna dla populacji, z której jest dobierana, jeśli wszyscy członkowie tej populacji mają takie same szanse, że zostaną wybrani do tej próby (...). [Takie próby jednak] rzadko, jeśli w ogóle kiedykolwiek, reprezentują w sposób doskonały populację, z których zostały dobrane. Mimo to dobór losowy ma dwie szczególne zalety. Po pierwsze, próby losowe, choć nigdy nie są doskonale reprezentatywne, są z reguły bardziej reprezentatywne niż inne rodzaje prób, ponieważ są wolne od omówionych powyżej obciążeń. Praktycznie rzecz biorąc, bardziej prawdopodobne jest, że próba z jakiejś populacji będzie dla niej reprezentatywna, gdy jest to próba losowa, niż gdy jest nielosowa. Po drugie, co ważniejsze, rachunek prawdopodobieństwa pozwala nam oszacować reprezentatywność próby losowej. Można sobie wyobrazić sytuację, w której niezbyt kompetentny badacz mógłby zupełnie przypadkowymi metodami dobrać próbę reprezentującą niemal doskonale jakąś większą populację. Jednak szanse powodzenia takiego przedsięwzięcia są niewielkie. Nie bylibyśmy w stanie oszacować, na ile

prawdopodobne jest, że taka próba jest reprezentatywna. Natomiast osoba dobierająca próbę losową może podać dokładne oszacowanie, na ile jej się to udało”.

Wyobraźcie sobie takie losowanie próby do sondażu prezydenckiego w waszej szkole, wśród waszych koleżanek i kolegów? Jak mógłby on przebiegać? Najpierw musielibyście stworzyć operat. Operat to spis wszystkich jednostek empirycznych składających się na populację. W tym wypadku spis wszystkich uczniów, na przykład w kolejności alfabetycznej, ponumerowanych kolejno. Powiedzmy, że jest ich 596. Aby z takiego operatu wybrać do badań 250-osobową próbę, wystarczy wskazać losowo (można do tego celu użyć tak zwanych tablic losowych: 250 numerów z zakresu od 1 do 596) i z wylosowanymi osobami przeprowadzić badanie, zadając im pytania sondażu.

Metoda jest skomplikowana pod względem organizacyjnym. Ale biegłość w posługiwaniu się narzędziami komputerowymi (arkusze kalkulacyjne, bazy danych) i odrobina wyobraźni z pewnością umożliwi wam przeniesienie tego skomplikowanego procesu do środowiska cyfrowego.

Badacze społeczni nie mają jednak tak łatwego życia. Po pierwsze, ustawa o ochronie danych osobowych uniemożliwia tworzenie operatów imiennych. Po drugie, dostęp do innych operatów jest niezwykle drogi. Studiując metodologię badań społecznych, poznacie jednak z pewnością różne sposoby dobierania prób losowych, quasi-losowych i celowych (czyli nielosowych – bo i takie, w niektórych sytuacjach stosowane są w socjologii i badaniach rynkowych).

Wracając do wielkości próbki: minimalna oznacza w naszym przypadku – optymalna. O tym, że zbyt mała próbka może się okazać niereprezentatywna z jakiegoś istotnego dla badania punktu widzenia, mówiliśmy przed chwilą. O tym, jak skonstruowane są wzory mówić tutaj nie będziemy, bo temat przekracza zakresem kilka takich lekcji.

Optymalizacja polega na tym, aby właściwa próbka była nie tylko nie za mała, ale i nie za duża. Zwiększanie jej liczebności ponad pewien próg podnosi tylko koszty. Nie wpływa już na pewność wnioskowania. Podam przykład:

Powiedzmy, że chcecie dokonać szacowania wielkości parametru w populacji (odsetka pragnących głosować na kandydata X) na podstawie próby, tak by z prawdopodobieństwem 95% popełnić co najwyżej 5-procentowy błąd przy szacowaniu. Wielkości prób, jakie w tym celu musielibyście dobrać dla populacji określonej wielkości wynoszą: dla populacji 6 osób – próba 6 osób, dla populacji 300 osób – próba 168 (a więc ponad 50% populacji). Dla populacji 600 000 (duże miasto) – próba 384 osoby i uwaga, dla populacji 30 000 000 też... 384 osoby. Po co więc zwiększać próbę? Tylko po to, by zmniejszyć błąd szacowania. Ale uwaga, jeśli badając

600-tysięczną populację, będziemy chcieli zmniejszyć ewentualny błąd z 5% do 2%, wówczas będziemy potrzebowali do próby nie 384 osoby, ale 2391 osób. Tylko kto udźwignie koszt takich badań?

Źródła cytatów: Wojciech Walczak, Joanna Suchomska, Beata Bielska, *Ani fakt, ani fikcja. Dlaczego sondaże mówią nie to, co komentatorzy?*, 13.10.2013, nowyobywatel.pl; Earl Babbie, *Badania społeczne w praktyce*, Warszawa 2004, s. 210–211.

Źródło: Englishsquare.pl sp. z o.o., licencja: CC BY-SA 3.0.

Ćwiczenie 1

Zdecyduj, które stwierdzenia są prawdziwe, a które fałszywe.

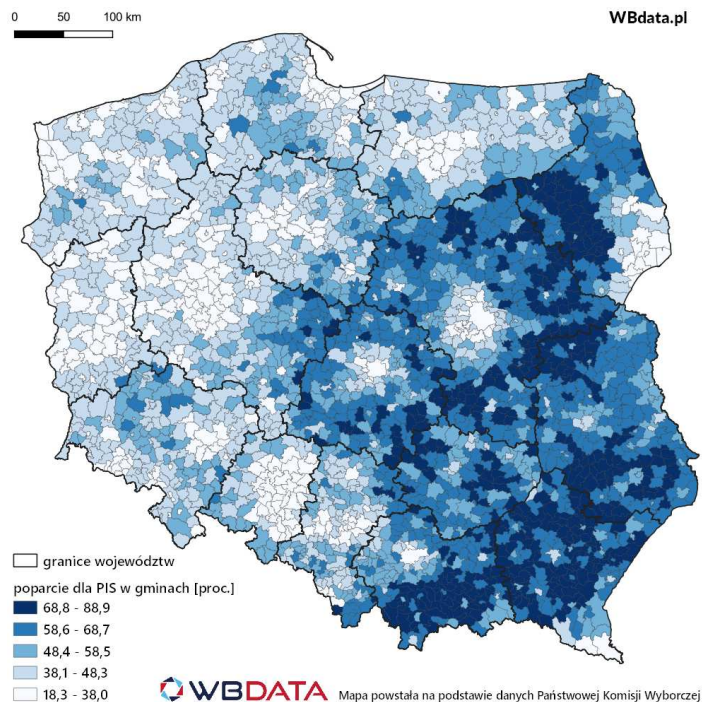
Stwierdzenie	Prawda	Fałsz
Próba losowa to próba przypadkowa.	☐	☐
Operat próby to nazwa listy zawierającej wszystkich respondentów wylosowanych do próby.	☐	☐
Nie można ustalić minimalnej wielkości próby dla 4-osobowej populacji.	☐	☐
Socjologowie stosują w badaniach zarówno próby losowe, jak i nielosowe.	☐	☐
Reprezentatywność próby w żadnym ze znaczeń nie zależy od jej liczebności.	☐	☐
Dobór losowy to taki dobór, w którym wszyscy członkowie populacji mają takie same szanse, że zostaną wybrani do próby.	☐	☐

Ćwiczenie 2

Zapoznaj się z poniższą tabelą oraz grafiką i wykonaj zadanie.

województwo	liczba ludności	udział w ludności trzech województw	liczebności w próbie
mazowieckie	5 391 813	66%	660
zachodniopomorskie	1 703 009	21%	210
lubuskie	1 015 440	13%	130
razem	8 110 262	100%	1000

Źródło: polskawliczbach.pl [dostęp: 02.04.2020 r.].



Wyniki wyborów do sejmiku 2019 na mapach

Źródło: www.wbdata.pl [dostęp: 02.04.2020 r.].

Rozważ przydatność następującego algorytmu doboru próbki 1000 respondentów do badania sondażowego mającego odpowiedzieć na pytanie o poziom poparcia dla PiS w Polsce pół roku po wyborach.

Proponowany **algorytm**: W pierwszym kroku losowo zostały wybrane trzy z 16 województw: mazowieckie, zachodniopomorskie i lubuskie. Następnie losowo dobierani są respondenci z bazy PESEL w obrębie województw. Liczba respondentów z każdego województwa w próbie jest proporcjonalna do liczebności ludności w danym województwie, uzyskanej z aktualnych danych GUS.

Uzasadnij swoje rozważania, wskazując na ewentualny wynik takiego badania.

Słownik

algorytm

sposób postępowania składający się zazwyczaj ze zbioru ściśle określonych kroków o ustalonych zależnościach, który prowadzi do rozwiązania problemu

opinia publiczna

reakcja zbiorowości na zdarzenia, zjawiska wywołane działaniem władz, instytucji wchodzących w skład szeroko rozumianej sfery publicznej; najogólniej oznacza aprobatę lub dezaprobatę dla widocznych w przestrzeni publicznej stanowisk, zachowań artykułowaną przez jakąś część społeczeństwa, zbiorowości

punkt procentowy (p.p.)

jednostka różnicy między dwiema wartościami wyrażanymi w procentach; wykorzystywana w porównaniach, np. jeśli udział wyborców partii X w styczniu wyniósł 20%, a w lutym – 35%, to ten udział wzrósł o 15 punktów procentowych (a nie o 15%)

respondent

podmiot (najczęściej osoba, ale także podmiot prawny), który przekazuje dane do celów statystycznych; w badaniu sondażowym osoba wylosowana z populacji, która odpowiada na pytania wywiadu zadawane przez ankietera lub pytania ankiety (przez samodzielne wypisywanie odpowiedzi w formularzu)

różnica procentowa

różnica między dwiema wielkościami wyrażonymi w procentach, podawana w jednostkach punkty procentowe (p.p.)

statystyka indukcyjna

dział statystyki zwany również wnioskowaniem statystycznym; zajmuje się formułowaniem i weryfikacją hipotez statystycznych określających wartość parametru jakiejś cechy w populacji (np. średnie wydatki w złotych polskich przeznaczane na święta lub procent planujących wziąć udział w wyborach) na podstawie rozkładu tej cechy w próbie dobranej losowo z tejże populacji

tendencja

skłonność do zachowań określonego typu (niepojawiających się stale, ale co jakiś czas) albo zmieniające się stany rzeczy (np. pomiary wartości jakiejś cechy), które pozwalają prognozować stałą zmianę w czasie wartości tej cechy (np. tendencja malejąca lub rosnąca)

wizualizacja

ogólne określenie metod graficznej prezentacji informacji; metody te służą zarówno prezentacji abstrakcyjnych pojęć, jak i wartości liczbowych; powinny ułatwić odbiorcy zrozumienie przekazywanych treści

Badanie 2

Eksperyment: najpierw wybierzcie losowo spośród siebie próbę odpowiadającą 50% uczniów. Oni odpowiadają na pytanie jako pierwsi. Zapiszcie wyniki. Następnie na pytanie sondażowe odpowiada cała klasa i również zapisywane są wyniki.

Weź pod uwagę dotychczasowe rozważania z Badania 1, wykonanych ćwiczeń oraz przedstawionych materiałów i ponownie odpowiedz na pytanie, wybierając jedną z zaproponowanych odpowiedzi. Wariant „trudno powiedzieć” wybierz tylko wtedy, gdy rzeczywiście nie potrafisz się zdecydować na inną odpowiedź.

Ćwiczenie 1

Jak myślisz, czy najbliższe wybory prezydenckie w Polsce zostaną rozstrzygnięte w pierwszej turze?

☐ Zdecydowanie tak – zostaną rozstrzygnięte w pierwszej turze.

☐ Najprawdopodobniej zostaną rozstrzygnięte w pierwszej turze.

☐ Nie, będzie potrzebna druga tura.

☐ Trudno powiedzieć.

Porównanie wyników badań

Ćwiczenie 2

Przeanalizuj różnice w procentowych rozkładach odpowiedzi na pytanie sondażowe.

W pierwszym kroku porównaj odpowiedzi z sondażu zrealizowanego w Badaniu 1 z sondażem przeprowadzonym wśród wszystkich uczniów. Rozważ ewentualne przyczyny w różnicach poszczególnych wskaźników procentowych (jeśli uznasz je za istotne).

W drugim kroku porównaj różnice między rozkładem odpowiedzi uzyskanym od wszystkich uczniów a – osobno – rozkładami uzyskanymi z dwóch prób: pierwszej z udziałem trzech osób w Badaniu 1 i drugiej z sondażu przeprowadzonego na próbie 50% klasy. Rozważ ewentualne przyczyny w różnicach poszczególnych wskaźników procentowych (jeśli uznasz je za istotne).

Dla nauczyciela

Autorka: Paulina Wierzbińska

Przedmiot: Wiedza o społeczeństwie

Temat: Sondaż. Między próbą a populacją

Grupa docelowa:

III etap edukacyjny, liceum ogólnokształcące, technikum, zakres rozszerzony

Podstawa programowa:

Zakres rozszerzony

VI. Społeczeństwo obywatelskie i kultura polityczna.

Uczeń:

10) przygotowuje i przeprowadza w najbliższym otoczeniu krótki sondaż opinii publicznej na temat dotyczący polityki i interpretuje jego wyniki.

Kształtowane kompetencje kluczowe:

- kompetencje obywatelskie;
- kompetencje cyfrowe;
- kompetencje osobiste, społeczne i w zakresie umiejętności uczenia się;
- kompetencje w zakresie rozumienia i tworzenia informacji.

Cele operacyjne:

Uczeń:

- charakteryzuje relacje między próbą a populacją i określa parametry poprawnie dobranej próby do badań sondażowych;
- projektuje próbę statystyczną do sondażu;
- przeprowadza krótki sondaż o tematyce politycznej na dobranej samodzielnie próbie uczniów swojej szkoły.

Strategie nauczania:

- konstruktywizm;
- konektywizm.

Metody i techniki nauczania:

- dyskusja;
- rozmowa nauczająca z wykorzystaniem ćwiczeń interaktywnych;
- eksperyment.

Formy pracy:

- praca indywidualna;
- praca w grupach;
- praca całego zespołu klasowego.

Środki dydaktyczne:

- komputery z głośnikami, słuchawkami i dostępem do internetu;
- zasoby multimedialne zawarte w e-materiale;
- tablica interaktywna/tablica, pisak/kreda;
- telefony z dostępem do internetu.

Przebieg lekcji

Faza wstępna:

1. Zalogowany na platformie nauczyciel wyświetla na tablicy interaktywnej (lub za pomocą rzutnika) pytanie sondażowe z sekcji „Badanie 1”. Uczniowie na telefonach komórkowych lub tabletach logują się do systemu, dołączają do lekcji.
2. Przedstawienie tematu i celów zajęć.
3. Nauczyciel prosi chętną/wybraną osobę o podanie trzech liczb z przedziału odpowiadającego liczbie uczniów na liście obecności w dzienniku.
4. Wybrane osoby udzielają odpowiedzi na pytanie sondażowe. Nauczyciel, korzystając ze swojego panelu użytkownika, wyświetla zestawienie odpowiedzi uczniów.

Faza realizacyjna:

1. Podział klasy na grupy. Każdy zespół rozważa problemy przedstawione w ćwiczeniu 2 w sekcji „Badanie 1”. Następnie osoby reprezentujące grupy przedstawiają efekty pracy, które są wspólnie dyskutowane przez uczniów.
2. Praca z filmem. Uczniowie w trakcie emisji wynotowują najważniejsze informacje. Dyskutują na temat wartości poznawczej sondażu przeprowadzonego w Badaniu 1 na próbie trzech osób.
3. Klasa wspólnie wykonuje ćwiczenie 1 dołączone do filmu.
4. Praca z audiobookiem. Uczniowie po zapoznaniu się z materiałem zastanawiają się nad sposobem skonstruowania próby do sondażu na temat wyborów prezydenckich, jaki można by przeprowadzić w ich szkole.
5. Nauczyciel prosi, by zespół klasowy wybrał losowo próbę odpowiadającą 50% uczniów. Ponownie wyświetla pytanie sondażowe (Badanie 2) i prosi o odpowiedź osoby

wybrane do próby. Wyświetla wyniki. Następnie na pytanie sondażowe odpowiada cała klasa i również wyświetlane są wyniki.

6. Podział na grupy. Każda z nich wykonuje ćwiczenie 2 w sekcji Badanie 2. Zespoły porównują wyniki sondażu przeprowadzonego w trzech badaniach. Przedstawiciele zespołów prezentują odpowiedzi, które są wspólnie omawiane i porównywane.

Faza podsumowująca:

1. W ramach podsumowania uczniowie oceniają wartość poznawczą wszystkich badań przeprowadzonych na zajęciach i określają parametry poprawnie dobranej próby do badań sondażowych.

Praca domowa:

- Uczniowie analizują materiał źródłowy i wykonują ćwiczenia 1 i 2 dołączone do audiobooka.

Materiały pomocnicze:

- Walter Lippmann, *Opinia publiczna*, Kraków 2020.

Wskazówki metodyczne opisujące różne zastosowania multimedium:

Uczniowie mogą wykorzystać film oraz audiobook do przygotowania własnej prezentacji multimedialnej na temat próby statystycznej.