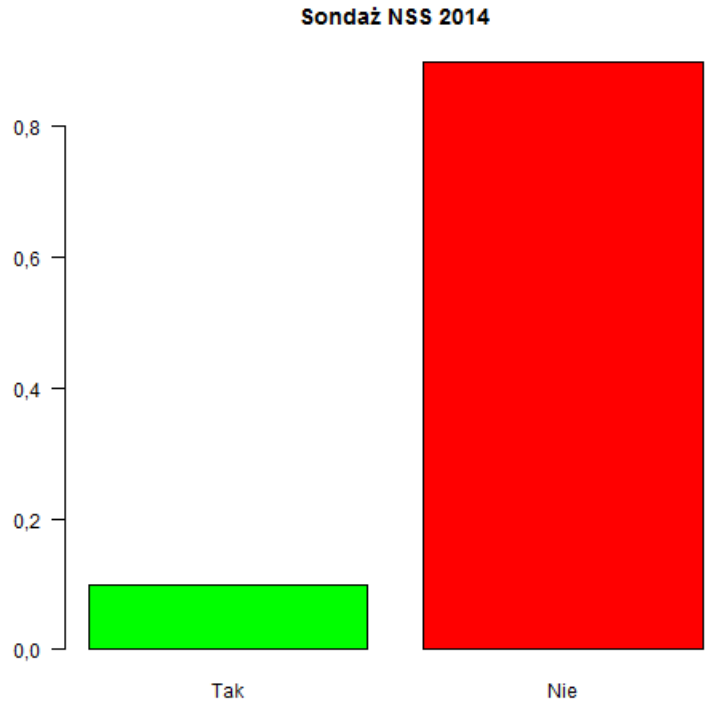
“PogRomcy Danych” feat. *Na Straży Sondaży*

1. Wstęp

Z badań przeprowadzonych przez zespół Na Straży Sondaży w 2014 r. wynika, że 95% Polaków nie wie jak powstają sondaże!



Jak interpretować ten wynik? Co on oznacza? Odpowiedź jest prosta. Praktycznie nic, bo ta informacja nie ma żadnej wartości. Dlaczego? Nie tylko dlatego, że jest zmyślona. Podstawowy problem stanowi brak jakiejkolwiek noty metodologicznej: nie powiedziałem dokładnie kiedy zorganizowano badanie, jaką techniką je przeprowadzono (telefonicznie czy bezpośrednio), jak zadano pytanie - co to znaczy, że ktoś wie jak powstają sondaże, a także nie zdefiniowałem kim są Polacy - czy są to obywatele, czy mieszkańcy Polski, w jakim byli wieku. Nie podałem również jakim błędem mogą być obarczone wyniki (tzw. błąd statystyczny), ani ile osób wzięło udział w badaniu. Codziennie w prasie i innych mediach pojawiają się podobne “dane sondażowe”. Czy mają one jakąkolwiek wartość? Czy można im zaufać? Jak odróżnić “dobry” sondaż od “złego”.

Na te oraz inne pytania postaramy się udzielić odpowiedzi w trakcie naszego kursu. Pokażemy na czym polega sondaż, z jakich elementów się składa, jakie są jago ograniczenia - czego nie powie nam nawet najlepsze badanie. Zaczniemy od przykładów łatwych, a skończymy na bardziej zaawansowanych próbując wcielić się na chwilę w rolę “sondażyst”. W imieniu zespołu Na Straży Sondaży zapraszamy do odkrywania niezwykłych możliwości, ale

także pewnych ograniczeń metody zwanej reprezentacyjną, metody która stoi za wszystkimi sondażami.

Zadania 1:

1. Wymień trzy artykuły prasowe z 2014 r., w których powołano się na wyniki badań społecznych (sondaży). Jeżeli w artykule podano taką informację to napisz ile osób brało udział w badaniu (ewentualnie ile firm lub innych instytucji jeżeli badanie nie dotyczyło postaw ludzi)

L.p.	Tytuł artykułu	Link do strony z artykułem	Czy podano wielkość próby? Zapisz ją jeżeli została podana	Czy podano dokładną datę realizacji badania? Jaką?
1.				
2.				
3.				

2. Metoda reprezentacyjna

Na pierwszy rzut oka badanie sondażowe, badanie na próbie, może się wydawać zadaniem karkołomnym. Oto na podstawie niewielkiej liczby obserwacji np. 1000 respondentów, staramy się opisać dużo większą, czasami nawet o kilka rzędów, populację. Czy ma to jakikolwiek sens? Czy da się wyznaczyć na tej podstawie przeciętną wagę, wzrost, miesięczne wydatki, liczbę przeczytanych książek albo czas spędzany dziennie na facebooku przez dorosłych mieszkańców Polski. Trzeba wiedzieć, że oficjalnie mieszka około 31 mln ludzi w wieku 18 i więcej lat. W tej sytuacji 1000 osób stanowi w zaokrągleniu trzy dziesięciotysięczne PROCENTA populacji (1 000 /31 000 000 = 0,000032)!!! To bardzo mało. Trudno uwierzyć, że taka garstka obserwacji może dostarczyć nam wiarygodnych informacji o całej badanej zbiorowości. Dla porównania zastanówmy się, czy na podstawie jednego kilometra drogi da się powiedzieć, jak będzie wyglądała cała podróż mierząca 300 000 km (Ziemia w obwodzie liczy tylko 40 000 km). Na pierwszy rzut oka nie. Okazuje się jednak, że

nauka, pod postacią statystyki, daje nam pewne narzędzia, które pozwalają trafnie wnioskować o dużych “obiektach” nawet na podstawie ich niewielkiego wycinka. W przypadku podróży kluczem do sukcesu byłoby umiejętne wybranie takich małych odcinków z całej drogi, które ułożą się w próbny kilometr.
Statystycy i badacze społeczni wiedzą doskonale, że dobrze dobrana próba stanowi świetny opis całej populacji. Jak to możliwe? Co trzeba zrobić, żeby przy użyciu małego kamyka dowiedzieć się czegoś o wielkiej "górze"?

Zadania 2:

1. Wymień trzy badania na próbach przeprowadzone przez instytucje państwowe i podaj link do ich wyników lub raportu. Mogą to być badania z roku 2014, ale także wcześniejszych lat.

L.p.	Nazwa instytucji	Tytuł badania	Link do strony z raportem
1.			
2.			
3.			

.

3. Przykłady prób

Zacznijmy od naszych codziennych doświadczeń. Wbrew pozorom większość naszej wiedzy o świecie czerpiemy z prób. I nie chodzi tu o metodę prób i błędów. Przyjrzyjmy się naszemu zdrowiu. Czasami zdarza się, że lekarz każe nam zrobić badanie krwi. Ale czy to oznacza, że trzeba zbadać całą krew w organizmie, wszystkie komórki? Na szczęście nie. Wystarczy mała próbka. Lekarz pobiera od nas zaledwie 10 ml krwi. W całym organizmie mamy jej aż 4,5l (4500ml). Tak więc próba stanowi 1/450. całej objętości krwi. To bardzo bardzo mało. Mimo to lekarz potrafi określić, co dzieje się w całym organizmie, a nie tylko w pobranej próbce.

Weźmy inny, mniej dramatyczny przykład. Wyobraźmy sobie, że chcemy ugotować zupę i lubimy, gdy jest ona odpowiednio słona. Jak to sprawdzamy? Czy musimy wypić całą zupę z garnka? Absolutnie nie. Wystarczy jedna łyżeczka, która zawiera 15ml zupy i stanowi zaledwie 3/1000 pięciolitrowego garnka. Znowu dobraliśmy niewielką próbkę, żeby zbadać większą całość. Ale przypadek zupy jest szczególny. Żeby przekonać się, czy zupa jest odpowiednio słona, musimy ją najpierw dobrze WYMIESZAĆ. Tylko wtedy proporcja soli w łyżce zupy, będzie taka sama jak proporcja w całym garnku - łyżka zupy będzie dobrze reprezentować całą zupę. Kluczową kwestią jest więc REPREZENTATYWNOŚĆ PRÓBY. Na pewno niektórzy słyszeli już to pojęcie. Zapamiętajmy je na chwilę, chociaż później będziemy musieli z niego zrezygnować. Reprezentatywność można rozumieć na różne sposoby. Często mówi się, że tak jak w przypadku łyżki zupy, próba musi być “miniaturą” populacji. Innymi słowy powinna odtwarzać strukturę i zależności obserwowane w całej zbiorowości. W przypadku badania krwi lub zupy brzmi to sensownie. Zauważmy jednak, że badane substancje są “jednorodne”. Każda porcja zupy czy krwi jest właściwie identyczna (dla uproszczenia, bo specjaliści z pewnością powiedzą, że to nie jest takie proste). A co jeśli badana zbiorowość nie jest i nie może być jednorodna? To problem, z którym bardzo często mierzą się nauki społeczne.

Zadania 3:

1. Podaj przykład z życia codziennego badania na próbie (np. badanie ilości soli w zupie).

L.p.	Czego dotyczy badanie?	Co jest próbą?
1.		
2.		
3.		

4. Sondaż

Przejdźmy do badań społecznych i tzw. sondaży politycznych. Ich wyniki często pojawiają się w prasie i mają duże znaczenie dla polityków oraz pewnie trochę mniejsze dla wyborców. Wyobraźmy sobie, że chcemy zmierzyć poziom poparcia dla wybranej partii P w wyborach do parlamentu. Dla naszych celów odsłońmy kilka faktów dotyczących badanej zbiorowości:

Tabela 1. POPULACJA WYBORÓW (dane w procentach)

Popieram partię P/ Miejsce zamieszkania	Nie biorę udziały w wyborach	NIE	Tak	Razem
Miasto	30	5	25	60
Wieś	20	10	10	40
RAZEM	50	15	35	100

W powyższej tabeli umieściliśmy procentowy rozkład dwóch cech w populacji pełnoletnich mieszkańców Polski (dane fikcyjne). Pierwszą cechą jest miejsce zamieszkania (w wierszach), a drugą poparcie dla partii P (w kolumnach). Widzimy, że w miastach mieszka 60% ludności, a na wsi 40%. Łącznie w całej zbiorowości (RAZEM) 35% obywateli zagłosowałoby na partię P, 15% na inną partię, a 50% w ogóle nie wzięłoby udziału w wyborach. Możemy również powiedzieć, że osoby mieszkające w mieście i popierające partię P stanowią 25% ogółu uprawnionych do głosowania, a także że osoby które mieszkają na wsi i nie biorą udziału w wyborach stanowią 20% populacji. Oczywiście, w normalnych warunkach tego typu informacje są niedostępne dla badacze. My je “odsłaniamy” potrzeby kursu.

Zastanówmy się jak z powyższej populacji dobrać próbę reprezentatywną, aby móc trafnie na jej podstawie ocenić, jaki procent mieszkańców Polski popiera partię P?
I tu pojawia się pierwszy problem. Nie da się bowiem “wymieszać” obywateli tak jak zupy. Zbiorowość nie jest jednorodna. Preferencje wśród mieszkańców miast i wsi nie są identyczne. Nie ma więc gwarancji, że jeśli idąc ulicą w mieście lub na wsi zapytamy dowolnych 10 osób o ich preferencje partyjne to będziemy mogli powiedzieć jakie jest poparcie w całym kraju. W uproszczeniu w mieście 4 na 10 osób zagłosowałoby na partię P (bo 25%/60%≈0,4), a na wsi 1 na 4 (bo 10%/40%=0,25). W pierwszym przypadku poparcie będzie zawyżone, a w drugim zaniżone, w stosunku do ogólnokrajowych wyników (35%). Widać więc, że nie każda próba będzie “REPREZENTATYWNA”. I tu zatrzymajmy się znowu przy definicji REPREZENTATYWNOŚCI. Zauważmy, że zależy nam na tym, aby wynik z próby był zgodny z tym co obserwujemy w populacji. Czyli w zasadzie nie musi być tak, że próba jest miniatura populacji. Ważniejsze jest to, że na jej podstawie prawidłowo odgadujemy wyniki dla całej zbiorowości.

Żeby przeprowadzić nasz polityczny sondaż musimy zrealizować badanie zarówno na wsi jak i w mieście. Każda osoba należąca do populacji musi mieć szansę znalezienia się w próbie. To bardzo ważne. Sposób dobierania próby, zwany schematem, nie może uniemożliwiać nikomu znalezienia się w próbie.
W dalszej części kursu pokażemy jak prawidłowo zdefiniować populację, jak błędy w kwestionariuszu mogą wpłynąć na zachowanie respondentów, jak dobrać, a dokładnie wylosować próbę reprezentatywną,

Zadania 4:

1. Czy w przypadku badania populacji mieszkańców Polski, wyniki sondy ulicznej przeprowadzonej w Warszawie, Krakowie lub Poznaniu można uznać za wiarygodne (dające się uogólnić na całą populację)?
2. Na podstawie danych z Tabeli 1. POPULACJA WYBORÓW odpowiedz na poniższe pytania:
    a. Jaki procent mieszkańców wsi popiera partię **P**?
    b. Jaki procent mieszkańców miasta nie pójdzie na wybory?
    c. Jaki procent osób które nie popierają partii **P** mieszka w mieście?

5. Problem badawczy: populacja i próba

W reprezentacyjnych badaniach sondażowych kluczowymi terminami są ***populacja*** i ***próba***.
Populacja jest to zbiorowość, o której badacz chce zdobyć informacje na podstawie sondażu. Przykładami populacji mogą być mieszkańcy Krakowa, Gdyni, uczniowie wszystkich szkół gimnazjalnych lub pełnoletni mieszkańcy Polski, nie tylko obywatele.
Próba jest to zbiór jednostek wybranych z populacji, o której badacz chce zdobyć informacje. Problem ze zdefiniowaniem tych dwóch pojęć dobrze ilustruje przykład z historii. W 1936 roku „Literary Digest", popularny magazyn informacyjny przeprowadził w USA badanie przedwyborcze. Do ludzi wybranych z książek telefonicznych i list rejestracyjnych samochodów wysłano DZIESIĘĆ MILIONÓW kart pocztowych, pytając, na kogo zamierzają oddać głos w wyborach prezydenckich – na republikanina Alfa Landona czy demokratę Franklina Roosevelta? Odpowiedziało ponad dwa miliony ludzi, wskazując że nowym prezydentem wybrany zostanie Alf Landon (57%), a nie Franklin Roosevelt (43%). Mogłoby się wydawać, że przebadanie tylu osób jest dużo bardziej wiarygodne i miarodajne niż przeprowadzenie badania na niewielkim wycinku populacji. Nic bardziej mylnego. Realne wybory dość drastycznie zweryfikowały wnioski z tych badań – nowym prezydentem został Franklin Roosevelt, mając największą przewagę głosów w historii – otrzymał 61%.
Dla porównania w tym samym czasie przedwyborczy sondaż, przeprowadził również George Gallup, który trafnie przewidział wyniki wyborów. W swoim badaniu posłużył się on jednak niewielką próbą kwotową (czyli opartą na znajomości określonych cech populacji, np. płeć, dochód, wiek, miejsce zamieszkania itp.).
Na czym więc polegał problem? Na całkowitym braku kontroli nad badaną próbą. Na pytanie zadane przez „Literary Digest" odpowiedziało zaledwie 22% wszystkich zapytanych osób. Jak się okazało karty w większości odsyłali republikanie. Drugi problem polegał na nieprawidłowym zdefiniowaniu populacji. Respondenci do badania zostali dobrani na

podstawie spisu abonentów telefonicznych i właścicieli samochodów. Taka konstrukcja próby dawała nadreprezentację zamożnych wyborców, czyli pominięcie ludzi biednych, którzy w większości głosowali na „New deal" Roosevelta.
Przykład ten ilustruje, jak dobór nawet dużej próby z niepoprawnie zdefiniowanej populacji może doprowadzić do zupełnie nietrafnych wniosków. Analizując wyniki badań sondażowych należy zawsze zwracać uwagę na sposób definiowania populacji. Definicja zbiorowości generalnej powinna zawierać informacje o:

- położeniu w przestrzeni zbiorowości (mieszkańcy Polski, mieszkańcy Poznania itp.)
- wieku respondentów (pełnoletni Polacy, osoby w wieku 15 i więcej ukończonych lat itp.)
- inne cechy (osoby posiadające obywatelstwo polskie, osoby z wykształceniem wyższym itp.)

Zdania 5:

1. Wymień trzy firmy zajmujące się badaniem rynku i opinii społecznej, które w 2014 r. prowadziły badania sondażowe w Polsce i podaj link do strony z raportami badawczymi tej firmy:

L.p.	Nazwa firmy badawczej	Link do strony z raportem
1.		
2.		
3.		

1. Masz następujący problem badawczy. Musisz zbadać poziom czytelnictwa tygodnika Na Straży Sondaży. Ukazuje się wyłącznie w formie drukowanej w miastach wojewódzkich (siedzibach wojewody). Pismo jest przeznaczone dla wszystkich, niezależnie od wykształcenia czy wieku.

a. Jak zdefiniujesz czytelnictwo? (Pytania pomocnicze: Kiedy ktoś staje się czytelnikiem? Jak często trzeba czytać żeby stać się czytelnikiem? Ile trzeba przeczytać żeby stać się czytelnikiem?)
b. Jak zdefiniujesz populację czytelników? (Pytania pomocnicze: jaki jest minimalny wiek czytelnika? czy czytelnicy mieszkają tylko w miastach wojwódzkich czy także w innych miejscach?)
c. Zapisz pytanie o czytelnictwo (maksymalnie dwa zdania):.....
d. Czy twoim zdaniem pytanie o czytelnictwo może być uznane za drażliwe?

2. W 2013 r. Dom Badawczy Maison przygotowała na zlecenie Polskiej Fundacji Pomocy Dzieciom „Maciuś" raport pt.:"*Głód i niedożywienie dzieci w Polsce".* Został on przygotowany na podstawie badanie przeprowadzono telefonicznie (technika CATI - Computer Assisted Telephone Interview) na ogólnopolskiej, reprezentatywnej próbie 800 przedstawicieli instytucji zajmujących się dziećmi i ich sytuacją życiową. Z raportu wynika, że pracownicy szkół i pracownicy OPS/PCPR szacują, że co dziesiąte dziecko z klas 1–3 dotyka problem niedożywienia.
    a. Zdefiniuj badaną populację.
    b. Podaj liczbę uczniów klas 1-3 w Polsce w 2013 r.(mogą to być dane zgodne ze stanem na 1 czerwca 2013 r. lub dla późniejszej daty przed końcem 2013 r.)
    c. Maksymalnie w trzech zdaniach napisz co rozumiesz przez niedożywienie
    d. Czy z danych przedstawionych w raporcie może wynikać, że w Polsce z głodu cierpi około 800 000 dzieci? Odpowiedź uzasadnij.

http://nastrazysondazy.uw.edu.pl/czy-800-000-dzieci-w-naszym-kraju-cierpi-glod/

6. Kwestionariusz

Dlaczego dziennikarze powinni informować o sposobie zadania pytania (z przytoczeniem dokładnego jego brzmienia, ale to oddzielny temat) pokazuje klasyczne już badanie opisywane przez Schumana [3], zrealizowane w 1986 roku w Stanach Zjednoczonych. Badacze z Uniwersytetu Michigan zapytali o najważniejsze wydarzenia lub zmiany, jakie zaszły w ostatnich 50 latach i wydają się respondentom najbardziej istotne. Połowa ankietowanych miała do dyspozycji kafeterię: II wojna światowa, podbój kosmosu, zabójstwo J. F. Kennedy'ego, wynalezienie komputera, wojna w Wietnamie, inne, nie wiem. Druga połowa otrzymała pytanie otwarte, a więc sama proponowała odpowiedzi. Wśród tej grupy wynalezienie komputera zostało wskazane jako najbardziej istotne wydarzenie lub zmiana przez 1% badanych, podczas gdy w pierwszej połowie ankietowanych była to najczęściej

wybierana odpowiedź (30%). Tak ogromna różnica wydaje się wystarczająco dobitnie pokazywać, jakie znaczenie ma forma zadanego pytania.

Badacze społeczni doskonale wiedzą, że umieszczanie jakiejkolwiek nazwy instytucji czy organizacji w treści pytania znacznie zniekształca wyniki. Jest to potwierdzone wieloma testami i badaniami socjologicznymi oraz psychologicznymi, a studia przypadków są szeroko opisywane w literaturze naukowej[2]. Formułowane pytania powinny odnosić się do rzeczywistości jak najbardziej neutralnie,

Zadania 6:

1. Odpowiedz na pytania związane z następującym badaniem: Firmę Research.NK przygotowała w 2013 r. raport „Prezentacja treści seksualnych przez młodzież poprzez wideoczaty" dla Naukowej Akademickiej Sieci Komputerowej. Dane zostały zebrane od respondentów przez Internet. Próba liczyła 976 nastolatków - osób w wieku 13-16 lat. Spośród nich 528 zdeklarowało, że korzysta z wideo rozmów. W tej grupie 10 osób zadeklarowało, że rozbiera się w sieci na żywo (Pytanie brzmiało: „Czy zdarzyło Ci się rozbierać się lub prezentować zachowania seksualne podczas wideorozmowy?"). Wiadomo również, że w badaniu wzięło udział 157 szesnastolatków, z czego 96 z nich korzysta z wideo rozmów, a 5 deklaruje rozbieranie się w czasie .
    a. Czy uważasz, że respondenci generalnie (a wiec przytłaczająca większość z nich) udzielali szczerych (zgodnych ze stanem faktycznym), odpowiedzi na pytanie o to, czy rozbierali się w trakcie wideoczatów - zarówno Ci którzy przyznali się do tego typu zachowań jak i ci którzy ich nie potwierdzili?
        i. Tak - generalnie odpowiadali szczerze;
        ii. Nie - generalnie odpowiadali nieszczerze;
        iii. Nikt tego nie wie. Może część tak, a część nie.
        iv. Jestem dzielna/y i napiszę jakie jest moje własne zdanie na ten temat : ...
    b. Czy uważasz, że badanie przez Internet daje większe poczucie anonimowości niż klasyczne badania prowadzone przez telefon lub twarzą w twarz z ankieterem? Uzasadnij maksymalnie w trzech zdaniach.
        i. Tak, ponieważ ...
        ii. Nie. ponieważ ...
    c. Czy w badaniu przeprowadzonym przez internet byłabyś / byłbyś skłonny odpowiadać szczerze na pytania dotyczące seksualności, chorób intymnych lub łamania prawa? Dlaczego? (Pytania pomocnicze: zastanów się, czy ważna w tych kwestiach jest anonimowość i poufność danych, a także, czy w ogóle mówienie na ten temat sprawia Ci jakieś problemy):
        i. Tak, ponieważ ...
        ii. Nie, ponieważ ...
    d. Jaki procent nastolatków w powyższym badaniu zadeklarował, że korzysta z wideo rozmów i rozbiera się w ich trakcie?

e. Jaki procent osób które zadeklarowały, że rozbierają się w trakcie wideoczatów do 16 latkowie?
f. Jaki procent nastolatków, którzy korzystają z wideo rozmów zadeklarował, że się rozbiera w ich trakcie?
g. Czy znając wyniki badania zgodziłabyś/zgodziłbyś się z następującymi określeniami:
    i. „Nowa plaga w sieci. Rozbierają się na żywo" - TAK / NIE
    ii. „Nagie gimnazjalistki w sieci. Nowa plaga w internecie" - TAK / NIE
    iii. "Nowe zjawisko w internecie. Gizmazjaliści rozbierają się w sieci" - TAK / NIE
    iv. "Uwaga na wideoczaty. Niektóre nastolatki występują nago"

http://nastrazysondazy.uw.edu.pl/naga-prawda-o-badaniu-zachowan-seksualnych-mlodziezy-na-wideoczatach/

2. Poniżej znajdują się dwa sondaże. Wypełnił je i odpowiedz na pytania:
    a. *Sondaż 1:*
        i. *Czy gdyby wybory odbyły się w najbliższą niedzielę to wziąłbyś(ęłaby) Pan(i) w nich udział?*
        ii. *Jeżeli tak, to na jaką jedną partię odałby Pan(i) głos?*
            1. *oddam pusty/nieważny głos;*
            2. *partię P;*
            3. *partię Przyjaciół Demokracji;*
            4. *partię Przyjaciół Otwartości;*
            5. *partię Przyjaciół Społeczeństwa;*
            6. *partię Przyjaciół Środowiska;*
            7. *partię Przyjaciół Uczciwości;*
    b. *Sondaż 2.*
        i. *Proszę określić czy zgadza się Pan/Pani z następującymi stwierdzeniami:*
            1. *Podatki w Polsce są za wysokie i należy je niezwłocznie obniżyć*
            2. *ZUS jest organizacją drogą, nieefektywną i marnującą publiczne pieniądze*
            3. *Obywatele lepiej będą zarządzać swymi pieniędzmi niż urzędnicy w ich imieniu*
            4. *każda rodzina powinna móc liczyć na wsparcie ze strony państwa.*
        ii. *Czy słyszał(a) Pan(i) o powstaniu nowej partii "P", której programem jest m. in. obniżenie podatków, ograniczenie obciążeń biurokratycznych, zmniejszenie liczby urzędników oraz wsparcie dla rodzin?*
        iii. *Czy gdyby partia "P" brała udział w najbliższych wyborach do Sejmu to jaką jedną partię oddałby Pan(i) głos?*
            1. *oddam pusty/nieważny głos;*
            2. *partię P;*

3. *partię Przyjaciół Demokracji;*
4. *partię Przyjaciół Otwartości;*
5. *partię Przyjaciół Społeczeństwa;*
6. *partię Przyjaciół Środowiska;*
7. *partię Przyjaciół Uczciwości;*

c. Pytanie do sondaży 1. oraz 2.:
   i. *Czy* w obu sondażach udzieliłaś/eś takiej samej odpowiedzi?
   ii. Czy twoim zdaniem kolejność pytań w sondażu 1. 2. może mieć wpływ na odpowiedzi respondentów?

http://nastrazysondazy.uw.edu.pl/wiplerowski-cud-sondazowy/

7. Dobór próby

Przejdź do najbardziej technicznej części sondażu, a wiec doboru próby i realizacji badania. W branży sondażowej korzysta się z różnych rozwiązań i nie zawsze są to próby losowe mimo iż wiadomo, że są one obarczone pewnymi błędami. Dlatego my zajmiemy się metodą reprezentatywną, a więc próbami losowymi.

Wróćmy do naszej populacji, która pojawiła się w rozdziale 4. Sondaż. Załóżmy, że badana zbiorowość składa się z 20 osób wśród których preferencje i miejsce zamieszkania rozkładają się w następujący sposób (rozkład procentowy jest dokładnie taki jak w Tabeli 1. POPULACJA WYBORÓW (dane w procentach)):

Tabela 7.1. POPULACJA WYBORÓW (liczba obserwacji)

Popieram partię P/ Miejsce zamieszkania	Nie biorę udziały w wyborach	NIE	Tak	Razem
Miasto	6	1	5	12
Wieś	4	2	2	8
RAZEM	10	3	7	20

Z powyższej tabeli wynika, że w miastach mieszka 12 osób, a na wsi 8. Poparcie dla partii P deklaruje 7 osób, poparcie dla innych partii 3, a na wybory w ogóle nie pójdzie 10 osób. W załączonym pliku znajduje się lista zawierająca dane o wszystkich osobach należących do populacji:
https://docs.google.com/spreadsheets/d/1iSt2ZD9F8DhEh8UonnYqZ71wG7gqm5MSlpiZM2vL-Gw/pubhtml?gid=1189066294&single=true
Kolejne kolumny w pliku to:

- L.p.- liczba porządkowa
- Miasto_wies - miejsce zamieszkania
  - M - miasto

    - W - wieś
- Glosowanie - czy pójdziesz na wybory i na kogo zagłosujesz
    - -1 - nie pójdę na wybory
    - 0 - pójdę na wybory, ale NIE zagłosuję na partię P
    - 1 - pójdę na wybory i zagłosuję na partię P.

Przejdźmy do badania sondażowego. Załóżmy, że chcemy oszacować nieznany nam odsetek obywateli nie pójdzie na wybory (na chwilę zapomnijmy, że wynosi on 50%). Interesują nas wartości -1 w kolumnie “Glosowanie”. la uproszczenia, losujemy dwie różne osoby spośród wszystkich 20 obywateli. Dobór próby przebiega wg. następującego schematu:

1. Do próby losujemy jedną osobę spośród wszystkich 20. Każdy obywatel ma takie samo prawdopodobieństwo znalezienia się w próbie - wynosi ono 1/20;
2. Spośród pozostałych 19 osób znowu losujemy jednego respondenta. Ponownie w zbiorowości z której dobierany jest respondent każdy ma takie samo prawdopodobieństwo znalezienia się w próbie - wynosi on tym razem 1/19;
3. Otrzymujemy próbę, w której znajdują się dwie osoby. Poza próbą zostaje 18 obywateli.

Powyższy schamat losowania nazwiemy fachowo: losowaniem prostym bez zwracania 2 osób z populacji 20. Losowanie jest proste ponieważ na każdym etapie wszyscy obywatele pozostający w populacji mają takie samo prawdopodobieństwo dostania się do próby. Bez zwracania ponieważ po wylosowaniu jednej osoby do próby nie wraca ona do puli z której losujemy następną. Ten bardzo prosty schemat losowania próby pozwali nam przeanalizować pewne podstawowe zagadnienia związane z metodą reprezentacyją. Zastanówmy się przede wszystkim, jak będą wyglądały dobierane próby. Jako identyfikatora osób w próbie wykorzystamy liczby z kolumny L.p. z listy zawierającej wszystkich obywateli naszej populacji. Zapis (1,2) oznaczać będzie, że do próby wylosowano najpierw osobę o liczbie porządkowej 1, a następnie osobę o liczbie porządkowej 2. Próby możemy więc rozpisać wg. prostej reguły. Jeżeli do próby w pierwszym kroku dobierzemy osobę o liczbie porządkowej 1 to w drugim kroku do pary możemy dobrać osoby z liczbą 2, 3, 4, 5, …lub 20. W ten sposób otrzymamy próby: (1,2), (1,3), (1,4), (1,5), ... lub (1,20). Jeżeli do próby w pierwszym kroku dobierzemy osobę o liczbie porządkowej 2 to w drugim kroku do pary możemy dobrać osoby z liczbą 1, 3, 4, 5, …lub 20. W ten sposób otrzymamy próby: (2,1), (2,3), (2,4), (2,5), ... lub (2,20). Widzimy więc, że dla każdej z 20 osób w populacji możemy dobrać 19 różnych par. Łącznie możemy więc dobrać 20x19=380 dwuosobowych prób. Rozpiszmy je:
(1,2); (1,3); (1,4); (1,5); (1,6); (1,7); (1,8); (1,9); (1,10); (1,11); (1,12); (1,13); (1,14); (1,15); (1,16); (1,17); (1,18); (1,19); (1,20); (2,1); (2,3); (2,4); (2,5); (2,6); (2,7); (2,8); (2,9); (2,10); (2,11); (2,12); (2,13); (2,14); (2,15); (2,16); (2,17); (2,18); (2,19); (2,20); (3,1); (3,2); (3,4); (3,5); (3,6); (3,7); (3,8); (3,9); (3,10); (3,11); (3,12); (3,13); (3,14); (3,15); (3,16); (3,17); (3,18); (3,19); (3,20); (4,1); (4,2); (4,3); (4,5); (4,6); (4,7); (4,8); (4,9); (4,10); (4,11); (4,12); (4,13); (4,14); (4,15); (4,16); (4,17); (4,18); (4,19); (4,20); (5,1); (5,2); (5,3); (5,4); (5,6); (5,7); (5,8); (5,9); (5,10); (5,11); (5,12); (5,13); (5,14); (5,15); (5,16); (5,17); (5,18); (5,19); (5,20); (6,1); (6,2); (6,3); (6,4); (6,5); (6,7); (6,8); (6,9); (6,10); (6,11); (6,12); (6,13); (6,14); (6,15); (6,16); (6,17); (6,18); (6,19); (6,20); (7,1); (7,2); (7,3); (7,4); (7,5); (7,6); (7,8); (7,9); (7,10); (7,11); (7,12); (7,13); (7,14); (7,15); (7,16); (7,17); (7,18); (7,19); (7,20); (8,1); (8,2); (8,3); (8,4); (8,5);

(8,6); (8,7); (8,9); (8,10); (8,11); (8,12); (8,13); (8,14); (8,15); (8,16); (8,17); (8,18); (8,19); (8,20); (9,1); (9,2); (9,3); (9,4); (9,5); (9,6); (9,7); (9,8); (9,10); (9,11); (9,12); (9,13); (9,14); (9,15); (9,16); (9,17); (9,18); (9,19); (9,20); (10,1); (10,2); (10,3); (10,4); (10,5); (10,6); (10,7); (10,8); (10,9); (10,11); (10,12); (10,13); (10,14); (10,15); (10,16); (10,17); (10,18); (10,19); (10,20); (11,1); (11,2); (11,3); (11,4); (11,5); (11,6); (11,7); (11,8); (11,9); (11,10); (11,12); (11,13); (11,14); (11,15); (11,16); (11,17); (11,18); (11,19); (11,20); (12,1); (12,2); (12,3); (12,4); (12,5); (12,6); (12,7); (12,8); (12,9); (12,10); (12,11); (12,13); (12,14); (12,15); (12,16); (12,17); (12,18); (12,19); (12,20); (13,1); (13,2); (13,3); (13,4); (13,5); (13,6); (13,7); (13,8); (13,9); (13,10); (13,11); (13,12); (13,14); (13,15); (13,16); (13,17); (13,18); (13,19); (13,20); (14,1); (14,2); (14,3); (14,4); (14,5); (14,6); (14,7); (14,8); (14,9); (14,10); (14,11); (14,12); (14,13); (14,15); (14,16); (14,17); (14,18); (14,19); (14,20); (15,1); (15,2); (15,3); (15,4); (15,5); (15,6); (15,7); (15,8); (15,9); (15,10); (15,11); (15,12); (15,13); (15,14); (15,16); (15,17); (15,18); (15,19); (15,20); (16,1); (16,2); (16,3); (16,4); (16,5); (16,6); (16,7); (16,8); (16,9); (16,10); (16,11); (16,12); (16,13); (16,14); (16,15); (16,17); (16,18); (16,19); (16,20); (17,1); (17,2); (17,3); (17,4); (17,5); (17,6); (17,7); (17,8); (17,9); (17,10); (17,11); (17,12); (17,13); (17,14); (17,15); (17,16); (17,18); (17,19); (17,20); (18,1); (18,2); (18,3); (18,4); (18,5); (18,6); (18,7); (18,8); (18,9); (18,10); (18,11); (18,12); (18,13); (18,14); (18,15); (18,16); (18,17); (18,19); (18,20); (19,1); (19,2); (19,3); (19,4); (19,5); (19,6); (19,7); (19,8); (19,9); (19,10); (19,11); (19,12); (19,13); (19,14); (19,15); (19,16); (19,17); (19,18); (19,20); (20,1); (20,2); (20,3); (20,4); (20,5); (20,6); (20,7); (20,8); (20,9); (20,10); (20,11); (20,12); (20,13); (20,14); (20,15); (20,16); (20,17); (20,18); (20,19);

Powyżej znajdują się wszystkie możliwe **380 dwuosobowe próby otrzymane wg. wcześniej opisanego schematu.** Widzimy, że niektóre pary w próbach się powtarzają. Możemy bowiem wylosować najpierw osobę o nr 1, a potem osobę numer 2. Otrzymujemy wtedy próbę (1,2). Lub na odwrót. Najpierw wylosujemy osobę numer 2. a potem osobę nr 1. Otrzymujemy wtedy próbę (2,1). Nasuwa się pytanie, w ilu próbach znajdzie się każdy obywatel.

Tabela 7.2.

L.p.	Liczba wystąpień w próbach
1	38
2	38
3	38
4	38
5	38
6	38
7	38
8	38

9	38
10	38
11	38
12	38
13	38
14	38
15	38
16	38
17	38
18	38
19	38
20	38

Odpowiedź znajduje się powyżej. Każdy obywatel może znaleść się w 38 próbach. Dlaczego? Bo tworzy 19 różnych par z każdym innym obywatelem, a dodatkowo w każdej parze może być pierwszy lub drugi. Dlatego liczba wystąpień wyniesie 19x2=38.
Jak policzyć ogólna liczbę prób które możemy wylosowanć. Bardzo prosto. W populacji jest 20 osób. Zgodnie z naszymi założeniami losowanie dwuelemntowej próby przebiega w dwóch etapach. Najpierw losujemy pierwszą osobę spośród 20 dostępnych. Następnie z pozostałych 19 osób losujemy drugą osobę. Czyli próbę możemy wylosować na 20x19=380 sposobów. Widzimy więc, że przyjęty schemat losowania próby z danej populacji generuje pewną skończoną i policzalną zbiorowość prób. To bardzo ważne. Sposób w jaki generujemy próby decyduje o tym, co w nich się znajdzie. Zobaczmy więc co generuje nasz schemat.
Policzmy dla wszystkich wymienionych wyżej prób odsetek osób, które deklarują udziała w wyborach. Ponieważ próba liczy tylko dwie osoby więc możliwe są tylko trzy rodzaje wyników:

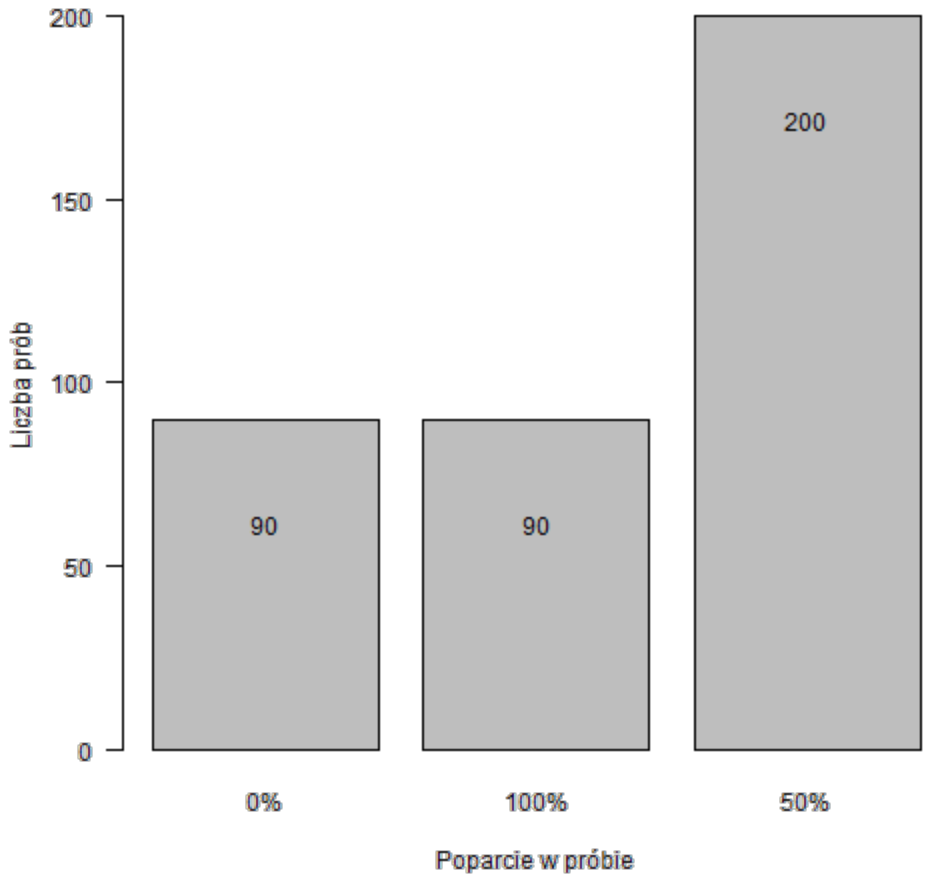
- (ABSENCJA, ABSENCJA) = (100%) - żadna z dwóch osób NIE zamierza wziąć udziału w wyborach;
- (GŁOSOWANIE, GŁOSOWANIE) = (0%) - obie osoby zamierzają wziąć udziału w wyborach;
- (ABSENCJA, GŁOSOWANIE) lub (GŁOSOWANIE, ABSENCJA) = (50%) - jedna z dwóch osób zamierza wziąć udział w wyborach.

Jak wygląda rozkład poparcia w poszczególnych próbach (zachowano wcześniejszą kolejność):
(0%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (0%);

(0%); (0%); (0%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (0%); (0%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (0%); (0%); (0%); (50%); (50%); (50%); (50%); (50%); (50%); (50%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (50%); (50%); (100%); (100%); (100%); (100%); (100%); (100%); (50%); (50%); (50%); (50%); (100%); (100%); (100%);

Na pierwszy rzut oka widać, że nie wszystkie wyniki są poprawne. W wielu próbach szacowana ABSENCJA jest zaniżona (0%) lub zawyżona (100%). Nie powinno to nas jednak dziwić ponieważ wyniki są ściśle związane z tym w jaki sposób dobieraliśmy próby. Policzmy dokładnie, ile wyników każdego rodzaju daje nam przyjęty schemat losowania:

Wykres 7.1. Rozkład oszacowań z prób



Na powyższym wykresie widzimy, że w 90 próbach absencja wynosi 0%, w 90 próbach 100%, a w 200 próba 50%. To bardzo logiczny rezultat. Jeżeli połowa obywateli będzie głosować w wyborach, a połowa nie, to oczywiste jest, że najczęściej jedna z dwóch osób w próbie będzie podwyższać frekwencję, a druga obniżać. Wiemy więc dokładnie czego możemy się spodziewać po naszym schemacie losowania próby - jakie wynika pojawią się najczęściej, a jakie najrzadziej. Zastanówmy, co by było gdybyśmy doświadczenie z losowaniem próby powtórzyli wielokrotnie. Załóżmy, że 1000 razy dobieramy, a potem zwracamy do naszej 20-osobowej populacji dwie osoby i za każdym razem notujemy poziom absencji. Wokół jakiego odsetka będą oscylować wyniki? Możesz się domyślać, że na 1000 przypadków w około 237 (90/380*1000) absencja wyniosłaby 0%, podobnie w 237 (90/380*1000) 100%, a w 526 (200/380*1000) 50%. Czyli rozkład wyników będzie podobny do tego na na wykresie słupkowym powyżej.
Wiemy już, że nie wszystkie próby odwzorowują poziom absencji w populacji. Wprawdzie dla ponad połowy wszystkich możliwch prób wynik jest prawidłowy niemniej jednak wciąż istnieje

spore ryzyko, że się pomylimy. Czy to znaczy, że nasz schemat losowania jest “zły”? Jak w ogóle sprawdzić, czy jest “dobry”? Że nasze wnioski z badania mogą być trafne? Spójrzmy na nasz problem w następujący sposób: wiadomo, że poszczególne próby dobierane do badania sondażowego mogą dawać trochę inne wyniki. Najlepiej gdyby oszacowania te, jeżeli nie trafiały w punkt, to przynajmniej oscylowały wokół prawidłowego wyniki. Schemat doboru próby powinien być tak zaprojektowany, aby wyniki z generowanych przy jego użyciu prób “ciąyły” w kierunku wartości obserwowanej w populacji. Co to znaczy? Zastanówmy się jaki wynik przeciętnie dają próby. Zastosujemy zasadę często stosowaną w szkole. Gdy chcemy się czegoś dowiedzieć o wynikach ucznia z całego roku liczmy średnią arytmetyczną jego ocen. Podobnie zróbmy z naszymi próbami. Policzmy czego możemy się po nich przeciętnie spodziewać. W tym celu dodajemy do siebie wszystkie 380 wyników ze wszystkich prób, a następnie dzielimy je przez liczbę wszystkich prób, czyli 380 - zupełnie jak średnią ocen w szkole. Możemy sobie uprościć to zadanie i zsumować wyniki w następujący sposób: 0%x90+100%x90+50%x200=19000%. Chwilowo rezultat jest absurdalny, ale to minie gdy podzielimy go przez liczbę prób: 19000% / (90+90+200) = 19000% / 380 =50% !!! I tu docieramy do sedna sprawy. Oto okazało się, że “przeciętnie” na próbę przypada 50% absencja**!!! To jest dokładnie tyle ile wynosi ono w całej populacji!!!** Oto podstawa całej metody reprezentacyjnej - **przeciętny wynik z próby powinien być równy wynikowi dla całe popualcji**. Jest to najważniejsza zasada badań sondażowych, ale także ogólnie wszyskich badań prowadzonych na próbach. Dzięki temu wiemy, że przeciętnie trafiamy w punkt - próba nie jest obciążona.

Tabela 7.3. Podsumowanie obliczeń

Absencja w próbie (A)	Liczba prób z daną absencją (B)	Iloczyn absencji i liczby prób z daną absencją (A*B)
0%	90	0%*90=0%
50%	200	50%*200=10000%
100%	90	100%*90=9000%
RAZEM	380	19000%

W następnych rozdziałach przyjrzymy się temu jak realizowane sa badania sondażowe oraz w jaki sposób wnioskuje się o populacji na podstawie próbu

zadanie: Wylosuj próbę z podanej populacji (PRÓBA A)

Zadania 7:

1. Odpowiedz na pytania związane z następującym schematem losowania próby z omawianej populacj (załączonej do rozdziału): ***losowanie proste bez zwracania 4 obywateli z popualcji 20*** i szacowania poziomu absencji na podstawie próby
    a. pamiętając, że do próby dobierane są 4 osoby napisz jakie możliwe kombinacje wyników sa możliwe do osiągnięcia
    b. napisz ile różnych prób można wylosować na podstawie podanego schamatu losowania
    c. napisz w ilu różnych próbach może się pojawić każdy obywatel
    d. jaki będzie przeciętny poziom absencji w próbie
    e. czy próba otrzymana z podanego schamatu jest obciążona
2. Wylosuj próbę 14 osób z populacji załączonej do rozdziału:
    a. zapisz L.p. osób wybranych do próby
    b. napisz jaki poziom absencji otrzymasz z wylosowanej próby
    c. jaki jest przeciętnych poziom absencji w próbie losowanej wg. takiego schematu jak ta wylosowana przez Ciebie?
    d. Napisz o ile punktów procentowych różni się wynik z twojej próby od wyniku w całej popualcji (50%)

8. Technika realizacji badania

Gdy już wiemy kogo dokładnie chcemy zbadać, musimy zastanowić się nad tym w jaki sposób dotrzeć do wybranych respondentów. Istnieją trzy podstawowe rodzaje technik realizowania badań sondażowych:

1. Wywiady bezpośrednie;
2. Wywiady telefoniczne;
3. Ankiety internetowe.

Do pierwszej grupy należą badania realizowane przy pomocy ankiety papierowej lub na komputerze przy udziale ankietera. Druga grupa badania, w których ankieterzy dzwonią do respondentów i zbierają od nich informacje przez telefon. Ostatni rodzaj to techniki polegające na zbieraniu informacji od respondentów przez internet bez udziału ankietera. W poniżej tabeli znajduje się, krótki opis wybranych technik badawczych.

Tabela 8.1. Techniki realizacji badań

Rodzaj techniki	Wywiad bezpośredni		Wywiad telefoniczny	Ankieta Internetowa
Technika	**PAPI (Paper And Pencil Interviewing)**	**CAPI (ComputerAssisted Personal Interviewing)**	**CATI (Computer Assisted Telephone Interviewing)**	**CAWI (Computer Assisted Web Interviewing)**
Opis	ankieta papierowa, wypełniana przez ankietera lub	konieczny udział ankietera, który sprawuje nadzór techniczny nad realizacją ankiety	dane są zbierane przez ankieterów przez telefon	dane są zbierane przez internet bez pośrednictwa, czy wsparcia ankieterów

	samodzielnie przez respondenta			
Zalety	- możliwośc dotarcia do każdej osoby znajdującej sie w populacji - brak ograniczeń technicznych	- możliwośc dotarcia do każdej osoby znajdującej sie w populacji - brak ograniczeń technicznych; - mozliwość zapisywania wyników od razu w bazie komputerowej	- szybka realizacja	-niewielki koszt - stosunkowo szybka realizacja
Wady	realizacja długa i kosztowna ze względu na rozproszenie respondentów w przestrzeni i konieczność podjęcia bezpśredniego kontaktu z każdym z nich	realizacja długa i kosztowna ze względu na rozproszenie respondentów w przestrzeni i konieczność podjęcia bezpśredniego kontaktu z każdym z nich	ograniczenie do osób posiadających telefon komórkowy	- ograniczenie do osób korzystajacych z internetu

Badania realizowane technikami bezpośrednimi są realizowane w “terenie”, czyli najczęściej w miejscu zamieszkania respondenta. Ankieterzy muszą więc wiedzieć, gdzie mogą zastać respondenta. Aby posługiwać się tą metodą konieczne są dane o miejscach zamieszkania osób należących do populacji. Na szczęście w Polsce istnieją przynajmniej dwa bardzo dobrze rozwinięte bazy danych, które umożliwiają pozyskanie tego rodzaju danych. Są to rejestr PESEL (baza ewidencji ludności w Polsce prowadzona przez MSW) oraz rejestr TERYT (baza adresów wszyskich mieszkań w Polsce prowadzona przez GUS). W dwóch najpopularniejszych wariantach techniki bezpośredniej odpowiedzi respondentów są zapisywane przez ankietera na papierowym kwestionariuszu (technika PAPI - Pencil and Paper Interview), bądź na laptopie lub innym urządzeniu mobilnym (technika CAPI - Computer Assisted Personal Interview). To drugie, częściej stosowane rozwiązanie eliminuje mozolny proces kodowania odpowiedzi, czyli przenoszenia ich z papieru do komputera. Techniki bezpośrednie są często stosowane do realizacji ogólnopolskich badań sondażowych, ale nie tylko. Cechują się one relatywnie niskim odsetkiem odmów wzięcia udziału w badaniu. Respondentowi trudniej odmówić jest wzięcia udziału w badaniu, jeżeli ankieter odwiedza go osobiście, niż w przypadku, gdy kontaktuje się z nim na przykład przez telefon. Tego rodzaju techniki pozwalają również na dotarcie do największej grupy ludności, nieosiągalnej poprzez połączenie telefoniczne, bądź przez internet. Jakość ma jednak swoją cenę. Ze względu na konieczność poniesienia kosztów pracy i podróży ankieterów (którzy w przypadku badań ogólnopolskich muszą przecież odwiedzić miejscowości rozrzucone po całym kraju) techniki oparte na bezpośrednim kontakcie ankietera z respondentem są najdroższe ze wszytkich technik badawczych. Ten sposób prowadzenia badań wiąże się również z dość długim okresem gromadzenia danych.
Kolejną często wykorzystywana technika polega na przeprowadzeniu wywiadów przez telefon (CATI - computer assisted telephone interview). Ta metoda jest dużo szybsza i tańsza od wywiadów face-to-face (eliminuje się w niej koszt oraz czas potrzebny na dotarcie ankietera

na miejsce wywiadu), jednakże jest to obwarowane pewnymi wadami. Przede wszystkim, metoda ta nie powinna być wykorzystywana w przypadku badań o charakterze ogólnopolskim. W Polsce wciąż nie wszyscy są posiadaczami telefonów (według badań 89% Polaków posiada telefon komórkowy). Powoduje to, że żadne badanie prowadzone tą metodą nie będzie w pełni obejmowało wszystkich mieszkańców Polski. Dodatkowo, w przypadku badań telefonicznych liczba odmów wzięcia udziału w badania jest wyższa, niż w przypadku badań face-to-face. Technika ta sprawdza się natomiast w sytuacjach, gdy do zbadania jest mniejsza populacja, do której członków posiadamy kontakt telefoniczny.
Ostatnią techniką są badania internetowe (CAWI - Computer Assisted Web Interview). W przypadku tej metody respondenci wypełniają ankietę w specjalnie do tego celu przygotowanym serwisie internetowym.
Sposób pozyskiwania informacji od respondentów powinien być dostosowany do populacji, kwestionariusza (pytań badawczych), harmonogramu realizacji oraz co bardzo istotne, budżetubadania.Istnieją cztery podstawowe techniki zbierania danych, których najważniejsze wady i zalety
Szczególnie istotną cechą różnicującą opisane powyżej techniki jest tak zwany poziom realizacji próby (nazywany również poziomem response-rate), czyli odsetek osób wytypowanych przez nas do udziału w badaniu, z którymi faktycznie udało się przeprowadzić wywiad. Trzeba pamiętać, że praktycznie nigdy nie występuje sytuacja, w której badaczom udaje się zapytać o opinię wszystkie osoby włączone przez nich do próby. Dzieje się tak z kilku powodów. Po pierwsze, nie do wszystkich osób można dotrzeć. Zdarzają się osoby do których dotarcie jest całkowicie niemożliwe (na przykład żołnierze pełniący czynną służbę na okręcie podwodnym przez dłuższy czas), a także takie, do których dotarcie jest bardzo trudne (ze względu na absorbujące życie zawodowe, częste podróże i tym podobne czynniki). Ponadto, część osób, do których uda się dotrzeć z różnych przyczyn odmawia wzięcia udziału w badaniu. Oba te czynniki są dodatkowym źródłem błędu w badaniu i celem każdego dobrego badacza jest ich zminimalizowanie. Okazuje się, że różne sposoby dotarcia do respondentów cechują się innym średnim wskaźnikiem realizacji próby.
Drugą ważną kwestią, decydującą o wyborze metody realizacji badania jest badana populacja. Nie każdą populację można zbadać dowolnie wybraną techniką, w niektórych przypadkach wybór techniki odpowiednio dobranej do badanej populacji pozwala skutecznie zminimalizować błędy realizacji.

Zadania 8.

1. Uniwersytecki Zespół Na Straży Sondaży zamówił badanie dotyczące popularności strony internetowej “www.nastrazysondazy.uw.edu.pl”. Chodziło oszacowanie odsetka osób w wieku 18-35 lat zamieszkałych w Polsce, które w ciągu ostatniego miesiąca zapoznały się z treścią (przeczytały cały lub prawie cały) przynajmniej jednego artykułu na stronie. Badanie zostało przeprowadzone metoda CAWI przez firmę “Polski Panel Internetowy” na próbie 917 osób w wieku 18-35 lat spośród 50 tys. osób które dobrowolnie zarejestrowały się do bazy internetowej firmy i za drobna opłatą zgadzają się odpowiadać na pytania w różnych ankietach. Wiadomo również, ze

osoby do badania zostały dobrane w ten sposób, aby rozkład płci wieku oraz wielkości miejscowości deklarowenego zamieszkania był zgodny z danymi podawanymi przez GUS na temat mieszkańców Polski. Odpowiedz na pytania związane z tym badaniem::

a. Czy populacja osób z posiadających dostęp do internetu (korzystających z internetu do celów prywatnych w domu, bibliotece, pracy lub szkole/uczelni) zawiera wszystkich mieszkańców Polski?
    i. Tak
    ii. Nie - Nie wszyscy mieszkańcy Polski mają dostęp do Internetu - sprwdż dane GUS
b. Czy próba badawcza 917 osób została dobrana z populacji polskich internautów - osób korzystających z internetu do celów prywatnych (nie związanych z praca zarobkowa, ale np. komunikacją ze znajomymi i rodziną, nauką, czytaniem prasy, graniem, oglądaniem fimów, robieniem zakupów itp.) w domu, kawiarence internetowej, bibliotece, pracy lub szkole/uczelni?
    i. Tak
    ii. Nie - próba została dobrana spośród 50 tys. osób, które same zgłosiły się do udzielania odpowiedzi na pytania ankietowe przez Internet. Choć jest to mało prawdopodobne można sobie wyobraźić, że w próbie mogł znaleźć się kogoś kto korzysta z Internetu tylko w celach zarobkowych, w tym do wypełniania ankiet za pieniądze. Zgodnie z nasza definicją taka osoba nie jest internautą.
c. Czy badanie zlecone przez Na Straży Sondaży obejmuje popluację polskich internautów.
    i. Tak - badanie dotyczy oszacowania odsetka wśród mieszkańców Polski ogółem nie tylko internautów. Musimy jednak pamiętać, że Polscy internauci nie maja ponieważ pytanie dotyczy odwiedziania strony “Na strazy sondaży”. Nie mozna odwiedzić strony internetowej nie będąc
    ii. Nie
d. Czy badanie zrealizowne przez “Polski Panel Internetowy” obejmuję populacje
e. Czy rozkład płci, wieku i wielkości miejscowości zamieszkania w próbie jest zgodny z rozkladem tych cech w populacji mieszkańców Polski?
    i. Tak
    ii. Nie - rozklad jest zgodny z danymi GUS dotyczącymi populacji mieszkańców Polski, a nie polskich internautów.

9. Wnioskowanie i błedy w oszacowaniach

Wróćmy do naszej populacji

Tabela 9.1. POPULACJA WYBORÓW (liczba obserwacji)

Popieram partię P/ Miejsce zamieszkania	Nie biorę udziały w wyborach	NIE	Tak	Razem
Miasto	6	1	5	12
Wieś	4	2	2	8
RAZEM	10	3	7	20

Tym razem znajmiemy się poparciem dla partii P. W stosunku do badania absencji wyborczej niewiele sie zmieni. Po prostu zamiast wartości -1 w kolumnie “Glosowanie” w naszym zbiorze danych teraz będziemy analizowali występowanie wartości 1, czyli glosowanie na partię P.
Zmienimy trochę nasz schemat losowanie. Tym razem dobierać będziemy po 6 obywateli do próby, ale samo jak wcześniej losujemy ich w sposób prosty bez zwracania.
Zobaczmy jakie próby wygeneruje nasz schemat dobporu próby. Ogólnie możliwych są następujące wyniki:

- (TAK, TAK, TAK, TAK, TAK, TAK) = 100%
- (TAK, TAK, TAK, TAK, TAK, NIE) = 83%
- (TAK, TAK, TAK, TAK, NIE, NIE) = 67%
- (TAK, TAK, TAK, NIE, NIE, NIE) = 50%
- (TAK, TAK, NIE, NIE, NIE, NIE) = 33%
- (TAK, NIE, NIE, NIE, NIE, NIE) = 17%
- (NIE, NIE, NIE, NIE, NIE, NIE) = 0%

Różnych prób w których żaden obywatel się nie powtórzy otrzymamy 20*19*18*17*16*15=27907200. Trudno taką ilość rozpisać. Co możemy więc zrobić. Przede wszystkim możemy wykonać proste, ale trochę żmudne obliczenia, które pozwolą nam określić w ilu próba pojawi się każda obserwacja. Przypomnijmy sobie więc, co zrobiliśmy w przypadku dwu-osobowej próby. Policzyliśmy, że każdy obywatel może utworzyć 19 par z innymi obywatelami i dodatkowo może być wylosowany jako pierwszy lub drugi. W konsekwencji liczba prób, w których może występić wynosiła 19*2=38. A co w przypadku trzyosobowych prób? Wtedy każdy obywatel mógłby dobrać najpierw 1 z 19 pozostałych osób, a potem 2 z 18 pozostałych osób. Sam natomiast mógłby zostać wylosowany jako pierwszy, drugi lub trzeci. W efekcie każdy obywatel występowałby w 19*18*3=1026 trójkach. Analogicznie dla sześcioosobowej próby, każdy obywatel może utworzyć szóstkę najpierw z jednym z 19, potem jednym z 18, potem jednym z 17 itd. aż wreszcie jednym z 15 obywateli. Co więcej może on zostać wylosowany jako 1,2,3,4,5 lub 6 do próby. Ostatecznie każdy obywatel występuje wiec w 19*18*17*16*15=1395360 sześcioosobowych próbach.

Tabela 9.1. LICZBA PRÓB, W KTÓRYCH WYSTĘPUJE KAŻDY OBYWATEL

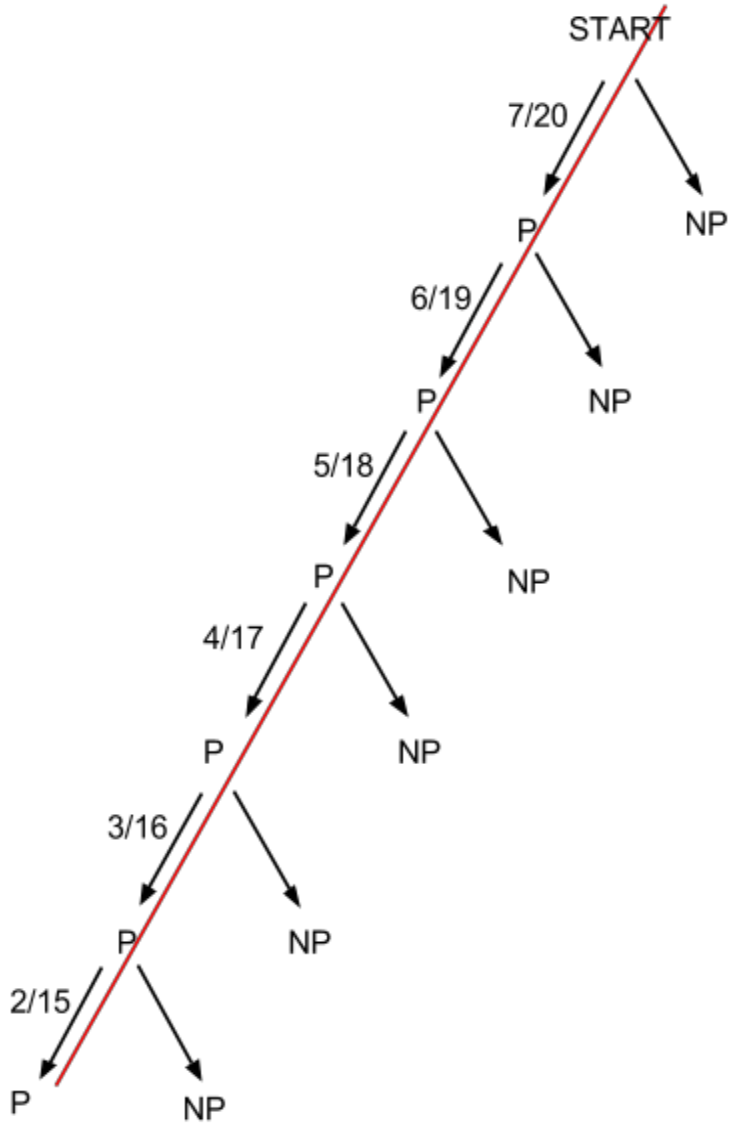
L.p.	Liczba wystąpień w próbach
1	8372160
2	8372160
3	8372160
4	8372160
5	8372160
6	8372160
7	8372160
8	8372160
9	8372160
10	8372160
11	8372160
12	8372160
13	8372160
14	8372160
15	8372160
16	8372160
17	8372160
18	8372160
19	8372160
20	8372160

Wiemy już w ilu próba wystąpi każdy obywatel: w ośmiu milionach trzysta siedemdziesięciu dwóch tysiącach stu sześćdziesięciu. Dużo!

Zastanówmy się teraz, z jaką częstotliwością będą występowały poszczególne wyniki z próby, czyli ile jest takich prób, w których poparcie dla partii P wyniesie 100%, ile takich, w których poparcie wyniesie ⅝ itd. Odpowiedź na to pytanie można uzyskać dwoma metodami. Albo stosując żmudne obliczenia na papierze albo stosując żmudne obliczenia na komputerze. Przyjrzyjmy się pierwszej metodzie.

Wyobraźmy sobie proces dobierania próby.. Składa się on z 6 następujących po sobie oddzielnych losowań. Każde z nich ma wpływ na następne. Zacznijmy od rozważenia sytuacji, w której losujemy do próby tylko osoby, które zagłosują na partię P (poparcie w próbie 100%). Zaczynamy od dobrania naszego pierwszego respondenta. Kto nim będzie? Prawdopodobieństwo, że wylosujemy osobę popierającą P wynosi 7/20, a niepopierającą P 13/20. Załóżmy, że wylosowaliśmy obywatela popierającego P. Teraz losujemy drug osobę. W populacji pozostało 19 osób z czego 6 popiera P. W związku z tym prawdopodobieństwo, że wylosujemy obywatela popierającego P wynosi 6/19, a niepopierającego P 13/19. Potem losujemy trzeciego, czwartego, piątego i wreszcie szóstego respondenta. Za każdym razem szanse wylosowania obywatela popierającego P są mniejsze. Możemy sobie rozpisać kolejne etapy losowania jako drzewko.

Rysunek 8.1 "Drzewko" ilustrujące losowanie próby, w której poparcie dla P wynosi 100%



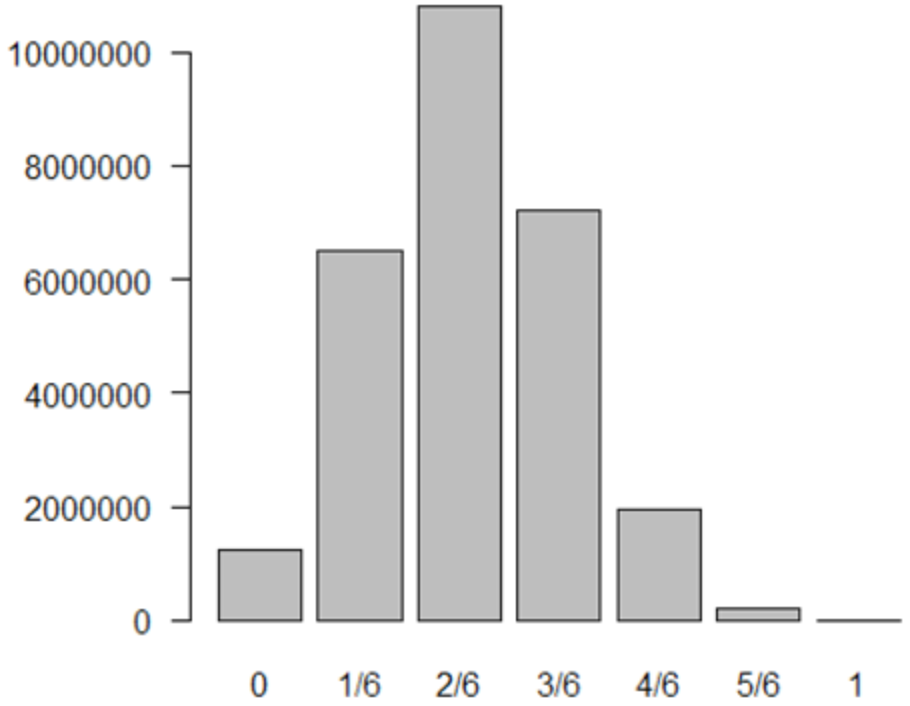
Mając rozpisane wszystkie kroki losowania możemy popoliczyć wreszcie, ile wynosi prawdopodobieństwo otrzymania próby, w której poparcie dla partii P wynosi 100%. W tym celu musimy wyznaczyć iloczyn (przemnożyć) wszystkie wartości z powyżego "drzewka" z "gałązek" oznaczonych jako P: 7/20*6/19*5/18*4/17*3/16*2/15=7/38760 (0,00018). Tą wartość interpretujemy jako prawdopodobieństwo wylosowania sześcioosobowej próby składającej się wyłącznie z obywateli popierających partię P. Co więcej, to jest też prawdopodobieństwo wylosowania takiej próby spośród wszystkich 27907200 prób. Skoro tak, to znaczy, że liczba prób, w których są sami zwolennicy parti P jest równa 27907200*7/38760=5040. W ten sposób udało nam się policzyć pierwszy, stosunkowo prosty przypadek. Co z resztą? Żeby policzyć prawdopodobieństwo dla innych wyników z próby

należałoby rozpisać “drzewko” dla wszystkich możliwych wyników losowania. Łącznie jest ich 64, bowiem na każdy wynik losowania w jednym etapie przypadają dwa wyniki następne w kolejnym. W związku z tym otrzymujemy 2*2*2*2*2*2 =64 różnych wyników. Duża część z nich opisuje podobne zdarzenia np. prawdopodobieństwa wylosowania prób (P, P, P, P, P, nP) oraz (P, P, P, P, nP, P) policzymy oddzielnie na naszym “drzewku”, ale będziemy musieli je zsumować, bo opisują jeden przypadek, gdy poparcie w próbie dla partii P wynosi ⅚.
Oczywiście to samo możemy policzyć też inaczej, zlecając większośc pracy komputerowci. To rozwiązanie nie wymaga od nas rozumienia procesu losowania i dlatego jest mniej “eleganckie”, ale za to pozwoli szybko otrzymać interesujące nas wyniki. Przy pomocy programu R udalo nam sie otrzymać następujące rozwiązanie dla próby sześcioosobowej::

Tabela 9.1. Liczba prób 6-osobowych z danym poparciem

Poparcie	Liczba prób
0	1235520
1/6	6486480
2/6	10810800
3/6	7207200
4/6	1965600
5/6	196560
1	5040
RAZEM	27907200

Tabela 9.2. Poparcie dla partii P w próbach sześcioosobowych



Widzimy jak wiele różnych możliwości możemy uzyskać z naszych prób. Przyjrzyjmy się wynikom.. Najwięcej prób zawiera poparcie dla partii P na poziomie 2/6. Dużo mniej ⅙ i 3/6. Wyniki 0, 4/6 można nazwać rzadkimi, a ⅚ i 1 bardzo rzadkimi. Zauważmy również, że poparcia dla partii P nigdy nie jest równe 35%. Czy to oznacza, że nasz schemat jest wadliwy. Jak wiemy to, czy próba jest “trafna” określa się na podstawie przeciętnej wartości oszacowania w próbie. Gdy mierzyliśmy ten parametr dla sondażu dotyczącego absencji w wyborach, wiedzieliśmy, że przynajmniej część prób “trafia w punkt”. Tym razem tak nie jest. Żadna próba nie daje wyniku 35% poparcia dla partii P. Ale czy rzeczywiście schemat doboru respondentów jest wadliwy. Jeżeli sprawdzimy przeciętny wynik z próby to okaże się, że wynosi on dokładnie 35%, a więc dokładnie tyle ile wynosi on w populacji. To efekt zastosowanego schamatu losowania próby. Zgdonie z tym, co mówiliśmy wcześniej oznacza to, że jest on poprawny. Ale co z tego skoro nie portafi dokładnie “trafić” w wynik! Potrzebujemy dodatkowej miary trafności wyników. Musi ona określać skalę błąd popełnianego przy korzystaniu z konkretnych schematów losowania. Dzięki temu będziemy porównywać schmaty, a przede wszystkim kreślić, jak duże jest ryzyko, że się myslimy. Posłużymy się przy tym rozwiązaniem wymyślonym przez statystyków. Policzmy jak duzy jest rozstęp między wynikiem z próby, a prawdziwym wynikiem w populacji (35%). Wszystkie obliczenia znajdują się w poniższej tabeli

Tabela 9.2. Podsumowanie obliczeń

Poparcie dla partii P (A)	Liczba prób z daną absencją (B)	Różnica między wynikiem w próbie, a poparciem w populacji (A -35%)	Kwadrat różnicy (A-35%)^ 2	Iloczyn liczby prób ikwadratu błędu ((A-35%)^2)*B
0	1235520	-0,3500	0,1225	151351,2
1/6	6486480	-0,1833	0,0336	218017,8
1/3	10810800	-0,0167	0,0003	3003
1/2	7207200	0,1500	0,0225	162162
2/3	1965600	0,3167	0,1003	197106
5/6	196560	0,4833	0,2336	45918,6
1	5040	0,6500	0,4225	2129,4
RAZEM	27907200	---	---	779688

Zaczynamy od rozpisania wszystkich możliwych wyników z próby (kolumna A). Nasępnie dopisujemy, ile prób w naszym schemacie daje konkretny wynik (kolumna B). Obliczenia błędy zaczniemy od wyznaczenia różnicy między wynikiem w próbie i wynikiem w populacji (A - 35%). Następnie wyznaczymy kwadrat tego błędu ((A-35%)^2). W ten sposób otrzymaliśmy kwadrat błędu popełnianego w danym rodzaju próby. Nas interesuje nie błąd dla konkretnej próby (choć to też jest interesujące), ale błąd przeciętnie popełniany dla danego schematu losowania prób. Czyli chcemy poprostu policzyć średnią z naszego błędu - średnią arytmetyczną. Potrzebujemy do tego iloczynu kwadratów błędow w poszczególnych rodzajach prób i liczby prób każdego rodzaju (((A-35%)^2)*B). Następnie sumujemy otrzymane wartości (779688) i dzielimy przez ogólna liczbe prób. W efekcie otrzymujemy 779688/27907200 = 0,02793859649 Tyle wynosi przeciętny błąd kwadratowy (statystycy nazywaja go “Wariancją”). Czy to dużo? Trudno powiedzieć. Dlatego policzymy pierwiastek tej liczby. W ten sposób otrzymamy pierwiastek średniego kwadratu błędu, czyli jak gdyby przeciętny błąd jaki popełniamy stosując nasz schemat losowania próby. Wynosi on 0,02793859649^(1/2) = 0,1671484265. W ten sposób otrzymujemy miarę, którą nazywamy odchyleniem standardowym. Trzeba podkreślić, że nie jest przeciętnym błąd popełniany przez próby tylko pierwiastkiem przeciętnego kwadratu błędu. Mimo to zazwyczaj traktuje się go jako przeciętny błąd popełniany przez próbę. Wiemy więc, że po naszej próbie możemy się spodziewać błędu na poziomie prawie 17% (0,167). To bardzo dużo biorąc pod uwagę, że

rzeczywiste poparcie w próbie wynosi 35%. Nasz miernik pokazuje, że przeciętnie mylimy się o około połowę w naszych oszacowaniach!!!
Czy jest jakiś sposób, żeby temu zaradzić? Tak, i to dosyć prosty - możemy zwiększyć próbę. Każdy intuicyjnie wie, że większa próba to większa precyzja. Ale dlaczego tak jest?
Poniżej znajdują się trzy wykresy dla trzech schematów losowania i policzonych dla nich błędów standardowych oszacowania poparcia dla partii P. .
Pierwszy wykres ilustruje wyniki dla próby składającej się z pięciu respondentów, drugi dla 10, a trzeci dla 15. Ze względu na różną liczebność prób inne są możliwe oszacowania poparcia dla partii P i różny jest też ich rozkład. Widzimy jednak, że wyniki zaczynają się "skupiać" wokół rzeczywistego wyniku (czerwona pionowa linia). To właśnie metoda poprawiająca dokładność oszacowania.
Dlaczego tak się dzieje? Ponieważ w małych próbach stosunkowo łatwiej może się zdarzyć, że wszyscy będą za lub przeciw partii P. Innymi słowy łatwiej o skrajny wynik. Zwiększając próbę zwięszkamy szanse na to, że wyniki się "ustabilizują" - Przy większej liczbie obserwacji jest mniejsza szansa, ze próba zostanie zdominowana przez jedną frakcję.
Przejdźmy do kwestii wnioskowania. Prawdziwie badania sondażowe polegają na wnioskowaniu na podstawie pojedynczej próby. Nie ma więc możliwości analizowania wyników z ze wszyskich prób

10. Błędy systematyczne

Do tej pory poznaliśmy sposoby wnioskowania z próby na podstawie wybranego schematu losowania. Musimy jednak pamiętać, że samo prawdziwe badanie sondażowe nie składa się wyłącznie z losowania próby, ale przede wszystkim z "terenowej" realizacji. W zasadzie jest to główna część badania. Od niej wszystko zależy. Mamy nadzieję, że
Wyobraźmy sobie nasza próbę i sytuację, gdy niespodziewane na wywiady, pomimo bardzo wielu prób podjęcia kontaktu uda się zrealizować wyłącznie z osobami zamieszkałymi w mieście

Wnioskowanie
Do tej pory analizowaliśmy rozkłady wszystkich możliwych wyników z próby. Potrafimy sobie wyobrazić co nas czeka. Badanie sondażowe polega jednak na dobraniu tylko jednej próby osób które potrafimy zidentyfikować, ale dla których nie znamy rozkładu preferencji politycznych czy innych badanych cech. W związku z tym w prawdziwym badaniu nie znamy błędu schematu z góry. Tak samo jak poziom poparcia dla partii P musimy go oszacować na podstawie próby. Jak to zrobimy
No i tutaj też mam problem, w jaki sposób podjąć ten temat, aby był on "strawny". Jeżeli nic nie
wymyślę, to po prostu kursant policzy średnią z próby (punktowy estymator odsetka w populacji), uwzględniając tylko osoby, które „udzieliły odpowiedzi". Na końcu kursant porównuje wyniki przeprowadzonego przez siebie „sondażu", z parametrem w populacji i myśli „Ojej, ale fajne te sondaże!".

O podstawie procentowania (ostatnio w dzienniku gazecie prawnej podali wyniki ogółem i w rozbiciu na wszystko, bez informowania, że dane w rozbiciu są niedokładne)

Tu trzeba dodać fragment o

O błędzie:

Celem badań sondażowych jest zdobycie informacji na temat wybranej zbiorowości, zwanej w badaniach populacją. Sondaże są najczęściej realizowane tylko na stosunkowo niewielkiej część osób z populacji, na temat której badacz chce zdobyć informacje. Specjalne metody doboru respondentów do badania pozwalają uogólniać wyniki uzyskane w próbie na całą zbiorowość. Niestety dziennikarze często zapominają, że takie uogólnienie odbywa się z określoną dokładnością – nazywaną najczęściej błędem statystycznym.

O systematycznym błędzie - wieś i miasto

Problem występuje nagminnie w badaniach poparcia dla partii politycznych [1] [2]. Gdy czyta się w gazetach, że określony odsetek Polaków popiera daną partię polityczną, trzeba pamiętać, że na podstawie sondażu taki wynik można podać tylko z określonym przybliżeniem.

O ile jeden odsetek wystarczy traktować jako przybliżony, o tyle poważny problem pojawia się w sytuacji, kiedy w gazetach porównuje się odsetek osób popierających jakąś partię z odsetkiem popierających ją w poprzednim miesiącu [3]. W standardowych badaniach społecznych (próba około 1000 respondentów) przyjmuje się najczęściej, że błąd statystyczny wynosi +/-3% [4]. Oznacza to, że jeśli w styczniu daną partię popierało 25% respondentów, a w lutym 22% respondentów, to nie można powiedzieć, że poparcie dla tej partii w populacji spadło, gdyż różnica ta nie jest istotna statystycznie.

Wyjaśniając, na czym polega owa istotność, należy powiedzieć, że błąd statystyczny w powyższym przypadku oznacza, że skoro w próbie ze stycznia 25% osób popierało rząd, to z prawdopodobieństwem 95% w całej populacji popierało go od 22% do 28%. Gdy w próbie z lutego 22% respondentów wyraziło poparcie dla rządu, to wśród wszystkich dorosłych Polaków mogło go popierać od 19% do 25%. Jak widać przedziały dla stycznia i lutego się na siebie nakładają, przez co różnica w poparciu między tymi miesiącami może nie świadczyć o spadku poparcia dla rządu.

Ciekawym publicystycznym rozwiązaniem tego problemu jest pisanie w artykule tylko o odsetku respondentów, którzy wzięli udział w badaniu (a więc przytaczaniu wyniku dla próby), a nie o populacji, której badani są tylko reprezentantami. W artykule *Marna oferta wyborcza* [5] autorzy pisząc na temat przyczyn niskiej aktywności politycznej Polaków ani razu nie wypowiedzieli się na temat wszystkich Polaków, opisują jedynie wyniki dla przebadanej próby. Nie jest to już niepoprawne, tak jak w poprzednim przypadku, ale przerzuca odpowiedzialność na czytelnika, który sam musi ocenić, na ile dokładnie w danej próbie odzwierciedlone są poglądy wszystkich Polaków.

Najbardziej profesjonalnym rozwiązaniem byłoby jednak pisanie, w jakim przedziale mieści się badana cecha (np. poparcie dla partii) w całej zbiorowości, a nie wśród przebadanych respondentów. Jak widać bowiem, przy założeniu poprawności realizacji badania, wynik ten z

pewnym przybliżeniem oddaje rzeczywisty odsetek dla całej populacji. Jeśli nie podaje się przedziału tego przybliżenia, wówczas lepiej jest napisać o dokładnych wynikach uzyskanych dla przebadanych respondentów, niż dla całej zbiorowości, z której zostali oni wybrani.

zadanie: Z wylosowanej próby wyłącz niektóre obserwacje (PRÓBA B)