Romuald Stupnicki

ANALIZA I PREZENTACJA DANYCH ANKIETOWYCH

WYDAWNICTWA
AKADEMII
WYCHOWANIA
FIZYCZNEGO

Skrypt

Wydanie I (ISBN 83-87210-29-3) - 2003

Wydanie II poprawione i uzupełnione ISBN 978-83-61830-97-9

Redakcja i korekta techniczna - Joanna Kłyszejko
Projekt okładki - W. Kosiński

Wydawnictwo AWF Warszawa 2015 Wydanie II (poprawione i uzupełnione)
Objętość 4,21 a.w. E-książka Format B-5

Od autora

Do napisania tej książki, podobnie jak do wcześniej wydanej „Biometrii", skłoniły mnie zajęcia dydaktyczne ze studentami warszawskiej AWF. Ankiety, sondaże i obserwacje są często stosowane w badaniach, a o posługiwaniu się tymi technikami traktują liczne publikacje, w tym podręcznikowe. Znacznie uboższe natomiast jest piśmiennictwo poświęcone analizie wyników, przy czym analizy takie są często dość powierzchowne.

Niniejsza książka jest próbą wypełnienia tej luki nie pretendując do wyczerpującego ujęcia metodologii badań ankietowych. Nie jest ona podręcznikiem metod statystycznych. Nieco obszerniej niż w podręcznikach potraktowano tu praktyczną analizę liczebności (test chi-kwadrat), jednak bez wchodzenia w teoretyczne podstawy rachunku. Zakładam, że czytelnik zna podstawowe pojęcia i procedury rachunku statystycznego oraz wie, jak posługiwać się komputerem, zwłaszcza programem Excel.

Opisane sposoby analizy mogą być stosowane nie tylko do wyników badań ankietowych, ale również do wszelkich zestawień liczbowych. Może się zdarzyć, że w niektórych sytuacjach opisane metody okażą się niewystarczające, dlatego będę wdzięczny za wszelkie uwagi krytyczne i wskazanie takich mniej typowych rodzajów zestawień. Pozwoli to uzupełnić i poprawić następne wydanie.

ANALIZA I PREZENTACJA DANYCH ANKIETOWYCH

Spis treści

WPROWADZENIE

Techniki ankietowe są jednym ze sposobów badania opinii respondentów na określony temat. Są one szeroko stosowane, jednak uzyskane z ich pomocą wyniki nie zawsze dają miarodajne odpowiedzi. Składa się na to wiele przyczyn, wśród których można wymienić braki w trafności, rzetelności i reprezentatywności, nie licząc powierzchownego często bądź wręcz niewłaściwego opracowania wyników. Pojęcia te zostaną pokrótce omówione w rozdziale 1.

Ankiety są zazwyczaj tworzone dla potrzeb określonego zagadnienia, są więc narzędziem układanym *ad hoc*. Odróżnia to je od np. kwestionariuszy testów psychometrycznych, które są narzędziem standardowym, przeznaczonym do pomiaru pewnych zdefiniowanych cech. Aby wyniki badań psychometrycznych lub socjometrycznych były trafne i rzetelne, kwestionariusze są standaryzowane, czyli poddawane procedurom sprawdzającym. Standaryzowane, obiektywne narzędzie charakteryzuje się m.in. następującymi cechami:

- pytania zawarte w kwestionariuszu nie mogą być modyfikowane,
- do kwestionariusza dołączona jest instrukcja, a wyniki ocenia się według klucza,
- wynik testu jest w zasadzie niezależny od warunków badania, dzięki czemu uzyskuje się powtarzalne wyniki.

W dalszych rozważaniach termin „kwestionariusz" będzie użyty tylko w odniesieniu do standaryzowanych narzędzi badawczych, natomiast w odniesieniu do ankiety będzie stosowany termin „arkusz pytań", a nie często spotykany „kwestionariusz ankiety". Jakkolwiek ankiety bywają weryfikowane, np. przez różne zespoły badaczy, nie są one standaryzowane tak jak kwestionariusze, dlatego zagadnienie to nie będzie tu omawiane.

Inną kategorię stanowią sondaże i tzw. ankiety konsumenckie. Sondaże opinii publicznej, często przeprowadzane telefonicznie, składają się zwykle z kilku zwięzłych pytań, np. na temat preferencji politycznych; przykładem mogą być tzw. *exit polls*, czyli pytania kierowane do wychodzących z lokali wyborczych, na kogo oddali głosy. Ankiety konsumenckie adresowane są do rozmaitych usługobiorców – klientów hoteli, przedsiębiorstw usługowych itp. i zawierają pytania na temat jakości świadczonych usług, kultury obsługi itp. Sondaże i ankiety konsumenckie mają zwykle na celu jedynie uzyskanie odpowiedzi na określone pytanie, nie będą tu zatem omawiane.

Jak wspomniano wyżej, badania ankietowe często mają na celu uzyskanie danych dotyczących wąskiego zagadnienia, dlatego zestaw pytań należy przygotować niezwykle starannie. Ankieta nie może być za obszerna zarówno objętościowo, jak i tematycznie, bo przeciętny respondent nie będzie wystarczająco umotywowany, żeby na ankietę rzetelnie odpowiedzieć. Niniejszy podręcznik ma na celu przede wszystkim zapoznanie czytelnika z

zasadami analizy zebranego materiału, a metodologia samych badań ankietowych została przedstawiona tylko w stopniu koniecznym do realizacji tego celu.

Należy także wspomnieć o etyczno-prawnym aspekcie badań ankietowych. Wszelkie badania ankietowe powinny uzyskać akceptację Komisji Etycznej, nawet jeżeli nie jest to przewidziane statutem lub zarządzeniem. Badania ankietowe muszą być zawsze zgodne z przepisami o ochronie danych osobowych.

Przed przystąpieniem do zaplanowania badania ankietowego należy starannie rozważyć poniższe pytania, a odpowiedzi na nie zamieścić w uzasadnieniu i opisie badań:

- **Jaki jest główny temat i cel ankiety?** (Po co przeprowadza się ankietę; do czego potrzebna jest informacja o wynikach badań)
- **Do kogo adresowana jest ankieta?** (Zdefiniować populację).
- **Czy badania mają być reprezentatywne, czy wyczerpujące?**
- **Jeśli reprezentatywne, to ile osób ma być przebadanych i jak te osoby będą wybrane?**
- **Do kogo adresowane są wyniki ankiety?** (Wyniki nie powinny iść „do szuflady", tylko czemuś służyć).
- **Jakie korzyści mogą przynieść ankietowanym wyniki badań?** (Bardzo ważny czynnik motywacyjny).

Podane tu terminy i pojęcia zostaną wyjaśnione i omówione w dalszych rozdziałach. Pokazane w tekście przykłady okien dialogowych funkcji Excel odnoszą się do nowszych wersji Microsoft Office (2007 – 2013), lecz są podobne zarówno do starszych wersji (XP, 2003), jak i programu Open Office.

1. DOBÓR PRÓBY REPREZENTATYWNEJ

Przystępując do badań za pomocą techniki ankietowej należy przede wszystkim określić populację (zbiorowość), której mają dotyczyć badania. Możemy tu wyróżnić dwie ewentualności: *badania wyczerpujące*, gdy chcemy poznać opinie na określone tematy pewnej zamkniętej grupy, np. personelu zakładu, i kierujemy ankietę do wszystkich członków tej grupy bądź *badania reprezentatywne*, gdy na podstawie wyników uzyskanych w badanej grupie chcemy wyciągnąć wnioski odnoszące się do większej populacji. Aby badana grupa była reprezentatywna dla określonej zbiorowości, musi zostać wybrana z tej zbiorowości w sposób losowy. Wynika z tego, że badania wyczerpujące nie mogą prowadzić do wniosków uogólniających, gdyż badana zbiorowość, jako z założenia nielosowa, nie może być reprezentatywna. Pewne pozorne wyjątki od tej zasady zostaną omówione dalej.

Innym typem badań są np. subiektywne oceny jakości produktów, preferencji konsumenckich itp., w których warunek reprezentatywności nie musi być konieczny. Jako przykład może służyć organoleptyczna ocena produktów spożywczych przeprowadzana w różnych warunkach przez odpowiednio przeszkolone osoby; stosuje się tu zasady badań eksperymentalnych, według których niezbędne jest losowe przydzielanie osób do poszczególnych grup. Zagadnienia te nie będą tu omawiane, nie muszą bowiem być rozwiązywane techniką ankietową, mimo że można do nich stosować przedstawiane tu zasady zestawień i analizy typowe dla ocen subiektywnych.

Poniżej będą rozważane wyłącznie zagadnienia dotyczące badań reprezentatywnych, a więc zdefiniowanie badanej populacji, ustalenie struktury tej populacji, wybór planu badań i sposobu losowego wyboru próby reprezentatywnej.

Zdefiniowanie badanej populacji

Nierzadko zdarza się, że chcąc poznać np. opinie studentów na jakiś temat, przeprowadza się badania ankietowe wśród studentów jednej uczelni, a wyniki opisuje się jako „opinie studentów ... na przykładzie takiej to a takiej uczelni". Zastanówmy się nad tym przykładem. Trzeba zacząć od tego, że pojęcie „studenci" jest niejasne, bo studenci nie stanowią jednorodnej grupy. Mogą to być np. studenci studiów dziennych lub zaocznych, uczelni państwowych lub prywatnych, odbywający studia stacjonarne w kampusie uczelnianym lub studiujący w prowincjonalnych punktach konsultacyjnych, studenci jednej uczelni lub wszystkich uczelni danego typu, słuchacze studiów podyplomowych itp., nie mówiąc o kierunkach studiów i ich stopniu (licencjacki, magisterski, doktorancki). Określenie „zaoczni studenci kierunku wychowanie fizyczne, studia licencjackie, wszystkie typy szkół" może być dobrą definicją badanej populacji. Zbiorowość taka będzie niejednorodna ze względu na lokalizację uczelni, rok studiów, płeć, stan cywilny, miejsce zamieszkania,

źródło utrzymania, status społeczny, majątkowy, rodzinny itp., a więc będzie się składała z pewnych kategorii, co może zostać uwzględnione w tzw. metryczce ankiety.

Innym przykładem mogą być „wyniki badania losowej próby 1033 dorosłych Polaków", często publikowane w dziennikach. Poza niepodaniem definicji terminów „losowa próba" (jak losowana?) i „Polacy" (obywatele polscy? osoby zamieszkałe w Polsce?), brak tu informacji o sposobie ankietowania. Często stosuje się w takich wypadkach wywiad telefoniczny, a wówczas definicję badanej populacji należy uzupełnić. Badana będzie bowiem populacja osób posiadających telefony i skłonnych do udzielenia odpowiedzi. Nawet jeżeli taki sondaż nie będzie traktowany jako reprezentatywny, należy wiedzieć, do jakiej zbiorowości ma on zastosowanie.

Dodatkowym elementem definicji jest przedział czasu, do którego odnosi się pojęcie populacji. W ujęciu statystycznym każda populacja niezdefiniowana czasowo jest nieskończenie wielka, obejmuje bowiem nie tylko te elementy, które istnieją obecnie, lecz także te, które należały do tej populacji kiedyś i które będą do niej należeć w przyszłości. Umiejscowienie populacji w czasie już na poziomie definicji pozwoli zatem na poprawniejsze wnioskowanie. Jeżeli na przykład przedstawia się wyniki badań przeprowadzonych 20 lat temu, to wyniki takie i wynikające z nich wnioski mogą mieć znaczenie jedynie historyczne, bo zmieniły się warunki, w których przeprowadzano badania.

Podane w tym podrozdziale przykłady skierowane są zwłaszcza do początkujących badaczy, którzy często nie zdają sobie sprawy z konieczności ścisłego definiowania pojęć, w tym także przedmiotu badań. Przystępując do badań należy zatem zacząć od możliwie dokładnego zdefiniowania badanej populacji i wyróżnienia w niej pewnych kategorii. Następnie można przystąpić do ułożenia planu badań, który będzie obejmował następujące punkty:

- oszacowanie całkowitej liczebności badanej populacji,
- określenie całkowitej liczby ankiet, jaka może być zbadana,
- ustalenie podziału populacji na warstwy i kategorie wg pewnej hierarchii,
- ustalenia zasad losowania w obrębie warstw,
- ustalenie sposobu przeprowadzenia badań.

Punkty te zostaną omówione bardziej szczegółowo poniżej, a w dalszej części tego rozdziału rozpatrzymy fikcyjną sytuację i odnoszący się do niej tok rozumowania.

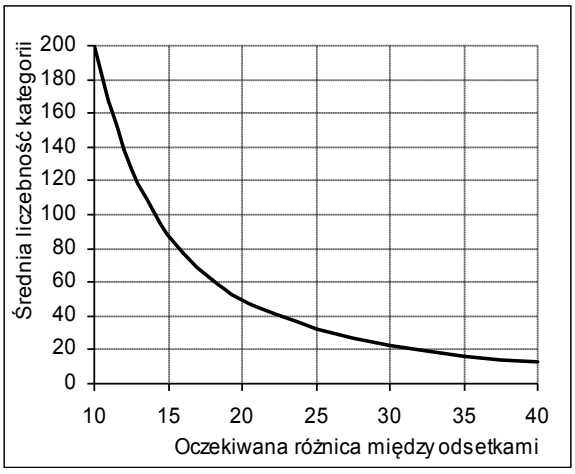
Oszacowanie liczebności populacji

Mówiąc o oszacowaniu liczebności populacji, ma się oczywiście na myśli liczebność w okresie prowadzenia badań, a więc pewien stan chwilowy, który też może być płynny, stąd termin „oszacowanie". Źródłem oszacowania może być np. rocznik statystyczny, publikacje prasowe, dane urzędowe (np. ministerialne, regionalne) itp. Jeżeli oszacowana wielkość zdefiniowanej populacji wynosi np. 20 000, a maksymalna liczba ankiet np. 800, to znaczy,

że ma zostać zbadane 4% populacji. Jeżeli populacja jest dzielona na warstwy (patrz niżej), powinno być zbadane ok. 4% osób z każdej warstwy.

Określenie całkowitej liczby ankiet

Jest to bardzo podstawowy wymóg. Od tego, ile ankiet badacz jest w stanie przygotować, rozprowadzić i opracować, zależy m.in. sposób podziału populacji na kategorie (patrz niżej). To z kolei może wpłynąć na zakres badań, a zatem także na definicję populacji. Trzeba się liczyć z wymogiem, aby na każdą kombinację kategorii populacji (np. uczelnia - rok studiów - płeć) wypadło przeciętnie nie mniej niż 30 - 50 ankiet. Może się okazać, że zaplanowane badanie obejmie np. 10 obiektów, w każdym wyróżni się 3 grupy wiekowe i dwie płcie; daje to łącznie 60 kategorii, a więc co najmniej 1800 zebranych ankiet. Taka liczba może się okazać nierealna, bo badacz może nie być w stanie zebrać i opracować więcej niż 800 ankiet, zatem plan będzie wymagał znacznych modyfikacji, np. redukcji liczby obiektów.



Postulowane tu minimum 30 – 50 ankiet jest dość arbitralne. W celu dokładniejszego oszacowania tego minimum można użyć wzoru: $n \approx 10^{(4.3 - 2 \log \Delta\%)}$, gdzie $\Delta\%$ jest wymaganą istotną różnicą w punktach procentowych. Graficzny obraz tego wzoru pokazano na wykresie obok.

Ryc. 1.1. Średnia liczba ankiet na kategorię wymagana do osiągnięcia pożądanej znamiennej różnicy między kategoriami

Podział populacji na warstwy i kategorie

Jak wspomniano wyżej, każda populacja jest niejednorodna, zwykle ze względu na wiele różnych cech. Można tu wyróżnić np. region, z którego pochodzi dana osoba, typ szkoły, do której uczeń uczęszcza, poziom wykształcenia rodziców, status majątkowy, płeć, i wiele innych. Niektóre cechy, takie jak region, typ szkoły, kategoria miejscowości (wieś, miasto), są łatwe do zdefiniowania i wyodrębnienia. Mogą one stanowić tzw. **warstwy** populacji. Jeżeli populacja jest liczna, rzędu wielu tysięcy lub milionów, losowy wybór osób byłby technicznie niewykonalny. Podział populacji na warstwy bardzo ułatwia przeprowadzenie reprezentatywnych badań, pod warunkiem wyboru podobnego odsetka osób z poszczególnych warstw i przestrzegania zasad losowego doboru w obrębie warstw.

Cechy stanowiące o niejednorodności populacji są rejestrowane w tzw. metryczce ankiety. Niektóre z nich, np. wiek, poziom wykształcenia, miejsce zamieszkania itp. są dzielone na **kategorie**. Dokonując podziału populacji na warstwy, trzeba pamiętać o ograniczeniach wynikających z całkowitej możliwej liczby ankiet (patrz wyżej). Mając np. 3 warstwy – terytorialną z ośmioma kategoriami (np. województwa), wieku z trzema kategoriami i dwie kategorie płci, otrzymamy łącznie 8×3×2 = 48 kategorii. Minimalna liczba ankiet wyniesie ok. 1500. Metryczka (zob. s. 20) może zawierać jeszcze inne cechy, np. poziom wykształcenia, miejsce zamieszkania itp. Jeśli zechce się scharakteryzować każdą kombinację, może się okazać, że dla niektórych kategorii jest zbyt mało ankiet. Dlatego podział na warstwy oraz wybór cech rejestrowanych w metryczce i ich podział na kategorie muszą być zrobione oszczędnie.

Losowy wybór obiektów i osób

Po ustaleniu warstw dokonuje się wyboru obiektów (np. szkół, szpitali), w których będą prowadzone badania. Jeżeli obiektów składających się na daną populację jest niewiele (kilka), lepiej objąć badaniami wszystkie, bo różnice między nimi mogą być duże. Jeżeli natomiast obiekty są liczne, wówczas należy wylosować odpowiednią ich liczbę w każdej warstwie. Należy najpierw oszacować liczbę wszystkich obiektów na badanym obszarze, a następnie liczby obiektów w poszczególnych warstwach, analogicznie jak opisano wyżej (zob. „oszacowanie liczebności populacji"). Odsetek losowanych obiektów (osób) powinien być w każdej warstwie podobny.

Schemat ten napotyka w praktyce na znaczne trudności. Jeżeli na przykład zgodnie z planem badań wylosowano w danym województwie 10 szkół, nie wszystkie szkoły mogły wyrazić zgodę na przeprowadzenie badań. Praktycznym wyjściem z tej sytuacji będzie wylosowanie większej liczby szkół, np. 15, aby zwiększyć szanse uzyskania zgody w 10 z nich. Pozostając przy przykładzie ze szkołami, może być konieczne losowanie klas w obrębie danej szkoły, a dopiero w wylosowanych klasach – losowanie dzieci.

Losując obiekty na danym obszarze, np. szkoły, należy dysponować pełną ich listą i każdej szkole przypisać kolejny numer. Losowanie polega wówczas na posłużeniu się tablicami liczb losowych bądź stworzenia takiej tablicy w arkuszu Excela: należy zaznaczyć pierwszą komórkę w wolnej kolumnie i wybrać funkcję LOS. W sąsiedniej komórce (np. B1) wpisać formułę: `=100*A1` (jeżeli lista zawiera więcej niż 100 obiektów) lub `=10*A1` (jeżeli jest mniej niż 100 obiektów), nacisnąć ENTER i zmniejszyć liczbę miejsc dziesiętnych zostawiając tylko liczby całkowite. Następnie zaznaczyć obie komórki (A1 i B1) i przeciągnąć myszką w dół o znacznie więcej pozycji, niż zawiera lista obiektów.

Na ryc. 1.1 pokazano fragment ciągu liczb losowych dla listy liczącej np. 260 obiektów. Liczby nieprzekraczające wartości 260 będą numerami wylosowanych obiektów; w podanym przykładzie są to: 149, 224, 7 i 217.

	A	B
1	0.317	317
2	0.149	149
3	0.748	748
4	0.366	366
5	0.224	224
6	0.966	966
7	0.594	594
8	0.007	7
9	0.513	513
10	0.353	353
11	0.217	217

Ryc. 1.1. Fragment tabeli liczb losowych generowanych funkcją LOS

Kolumna A – liczba losowa; kolumna B – liczba losowa ×100

W podobny sposób można losować osoby w obiekcie, można jednak użyć prostszego sposobu – z alfabetycznej listy wybrać tyle kolejnych nazwisk, ile przypada według planu na dany obiekt. Nie ma znaczenia, czy będą to kolejne nazwiska od początku listy, czy od końca, bądź ze środka. Można przyjąć, że uporządkowanie według pierwszej litery nazwiska jest już uporządkowaniem losowym.

Przeprowadzenie badań

Ankietę można przeprowadzać w różny sposób. Najczęściej stosowane są trzy:

a. Przesłanie arkuszy ankiety drogą korespondencyjną; zwrot wypełnionych arkuszy następuje tą samą drogą. Sposób ten umożliwia dystrybucję arkuszy do bardzo dużej liczby respondentów, ma jednak wady – niemożliwe jest wyjaśnienie ewentualnych wątpliwości, bardzo trudno jest kontrolować rzetelność odpowiedzi, a odsetek zwrotów nie przekracza na ogół 30%. W pewnych badaniach żywieniowych rozesłano pocztą ponad 10 000 ankiet, z czego odesłano ok. 2500; z tej liczby ok. 35% zawierało zupełnie nieprawdopodobne dane, a z faktu, że ok. 65% ankiet zawierało prawdopodobne dane, wcale nie wynika, że były one rzetelne. Przykład ten pokazuje trudności związane z pozyskiwaniem odpowiedzi drogą pocztową.

b. Ankieta audytoryjna: ankieter przychodzi do grupy respondentów i asystuje przy wypełnianiu arkuszy. Sposób ten zapewnia lepszą rzetelność, bowiem ankieter najpierw może wyjaśnić cel badań i udzielić ew. wskazówek co do sposobu wypełniania arkuszy. Wadą natomiast jest duże zaangażowanie czasowe ankietera, więc przy większej liczbie respondentów konieczny jest udział kilku lub nawet wielu odpowiednio przeszkolonych ankieterów.

c. W pewnych wypadkach konieczne jest ustne zadawanie pytań przez ankietera i wpisywanie przez niego odpowiedzi do arkusza. Sytuacja taka, częsta w badaniach socjologicznych, może mieć również miejsce przy ankietowaniu młodszych dzieci, osób w podeszłym wieku, niepełnosprawnych itp. Sposób ten jest właściwie odmianą opisanego wyżej (punkt *b*), z jego zaletami i wadami.

Oprócz powyższych możliwe są inne sposoby – np. za pośrednictwem internetu lub telefonu. Ten ostatni sposób nadaje się raczej do sondażu, a nie do ankiet zawierających wiele pytań. Ponadto oba wymienione sposoby zawężają badaną populację do posiadaczy łącza internetowego lub telefonu, co powinno być uwzględnione w definicji populacji.

Na koniec rozpatrzmy względnie prosty przykład zaplanowania badań ankietowych. Załóżmy, że badaną populacją mają być studenci zaoczni akademii wychowania fizycznego (uczelni państwowych), 3-letnich studiów licencjackich kierunku wychowania fizycznego. Uczelni tych jest 8, w jednej nie ma studiów zaocznych na tym kierunku. Przybliżone liczby studentów na poszczególnych latach przedstawiono w tabeli 1.1.

Poszczególne uczelnie stanowią naturalną pierwszą warstwę. Drugą warstwę może stanowić rok studiów, a trzecią – płeć, otrzymamy zatem $7\times3\times2 = 42$ grupy, co wymaga zebrania co najmniej $42\times30 = 1260$ ankiet. Należy to skonfrontować z możliwością przeprowadzenia takich badań; badania będą wykonywane przez dwie osoby metodą grupową (audytoryjną; *b*). Dobre przygotowanie badań (uzgodnienie terminów – dni i godzin) może umożliwić ankietowanie ok. 60 osób w ciągu jednego dnia, przy czym możliwych będzie 10 osobowyjazdów (np. 3 wyjazdy dwuosobowe i 4 jednoosobowe), a więc może się udać zebranie ok. $60\times10 = 600$ ankiet, czyli około połowy z wyliczonych 1260.

Tab. 1.1. Zestawienie orientacyjnych liczb studentów obu płci na poszczególnych latach studiów zaocznych WF w 8 uczelniach AWF i planowane liczby ankietowanych studentów

Nr uczelni	Orientacyjne liczby studentów				Planowana liczba ankiet			
	I rok	II rok	III rok	∑	I rok	II rok	III rok	Σ
1	110	80	60	250				
2	200	120	120	440	140	85	85	310
3	80	40	50	170	55	30	35	120
4	150	110	110	370				
5	50	50	30	130				
6	70	90	90	250	50	60	60	170
7	80	50	50	180				
∑	740	540	510	1790				600

Biorąc pod uwagę minimalną liczbę ankiet w grupie (30), liczba grup nie może przekroczyć 20, konieczna jest zatem weryfikacja planu badań. Możliwe są następujące modyfikacje:

- objęcie badaniami tylko niektórych, wylosowanych uczelni,
- rezygnacja z podziału na lata studiów,
- rezygnacja z podziału na płeć.

W odpowiedziach na pytania ankiety mogą wystąpić istotne różnice tak między płciami, jak i między latami studiów, zatem najrozsądniejsze wydaje się zmniejszenie liczby badanych uczelni drogą losowania do trzech, co redukuje liczbę grup do $3 \times 3 \times 2 = 18$. Powiedzmy, że wylosowano uczelnie nr 2, 3 i 6, w których jest łącznie 860 studentów. Należy zatem objąć ankietowaniem ok. 70% (600/860) studentów z poszczególnych lat tych trzech uczelni; planowane liczby ankietowanych studentów zamieszczono w tabeli. Planowane liczby badanych wynoszą od 67 do 75% całkowitych liczebności poszczególnych lat studiów, stałość odsetkowa jest więc wystarczająca. Dane te odnoszą się jednak do wszystkich studentów danego roku i uczelni bez podziału na płeć, a np. w uczelni nr 3 planowane liczby ankietowanych studentów są za małe, żeby uwzględnić dwie płcie. Może to uniemożliwić przeprowadzenie pełnej analizy dla trzech warstw populacji. Sytuacja taka zostanie omówiona w rozdziale 4.

Badania pilotażowe

Sondaże pilotażowe są często stosowane w badaniach ankietowych. Celem badań pilotażowych jest wstępne zweryfikowanie opracowanej ankiety (lub innej formy sondażu) pod względem trafności i rzetelności. Dotyczy to wszystkich omówionych zagadnień – liczby i układu pytań, treści pytań i ich sformułowań z punktu widzenia logiki i komunikatywności itp. Często się zdarza, że w wyniku pilotażu ankietę nieco się modyfikuje. Z tego punktu widzenia, wspomniany wcześniej wymóg krytycznego przejrzenia ankiety przez różne osoby można traktować jak swojego rodzaju „działania przedpilotażowe". W każdym razie, wszystkie te zabiegi mają na celu optymalizację ankiety pod względem treści, układu, odniesienia do badanej zbiorowości itp.

Badania pilotażowe przeprowadza się zwykle na małą skalę, rzędu kilkudziesięciu ankiet, i nie musi to być próba rygorystycznie losowa. Chodzi bowiem o uzyskanie orientacyjnego, przybliżonego obrazu zagadnienia na tle wyników ankiety. Wyniki takich badań są często prezentowane na konferencjach naukowych, zwykle z adnotacją „badania wstępne" lub „badania pilotażowe", należy jednak pamiętać, że są to badania niereprezentatywne, których wyniki mogą jedynie stanowić pewną wskazówkę do właściwych badań.

2. RODZAJE PYTAŃ I ODPOWIEDZI

Najogólniej mówiąc, można wyróżnić dwie grupy pytań: zamknięte i otwarte. Pytania zamknięte cechują się tym, że udzielane odpowiedzi mieszczą się w ściśle określonym zakresie lub skali, natomiast odpowiedzi na pytania otwarte respondent musi sam sformułować. O ile więc odpowiedzi na pytania zamknięte są określone w kategoriach ilościowych bądź jakościowych, o tyle odpowiedzi na pytania otwarte trzeba sklasyfikować na odpowiednie kategorie dopiero po zebraniu wypełnionych arkuszy, bo nie można z góry przewidzieć odpowiedzi. Pytanie otwarte może stanowić jedną z opcji wyboru w pytaniu zamkniętym (np. „inne – jakie?"), co zostanie omówione w podrozdziale „Pytania otwarte".

Pytania zamknięte z odpowiedziami kategorialnymi

Pytania zamknięte zawierają polecenie udzielenia określonej odpowiedzi (jednej lub więcej). Najprostszym przykładem, w którym odpowiedź wynika wprost z treści pytania, jest *odpowiedź dychotomiczna* lub *zero-jedynkowa*:

Czy masz przyjaciół? Nie (0) – Tak (1)

Odpowiedzi te są rozłączne, nie można bowiem wybrać obu jednocześnie.

Do pytania często jest dołączona lista wyborów, zawierająca więcej niż dwie opcje odpowiedzi. Lista taka, niezależnie od jej długości, nazywana jest przez licznych autorów **kafeterią**. W niniejszym opracowaniu pojęcie kafeterii jest zawężone do listy wyborów nieilościowych. Np. w pytaniu:

Z kim najchętniej spędzasz wakacje (wybierz jedną odpowiedź):

(a) sam
(b) z rodziną
(c) z partnerem (partnerką)
(d) w grupie

mamy typową kafeterię kategorialną (nieilościową), bowiem poszczególne kategorie nie tworzą ciągu dającego się uporządkować ilościowo; inaczej mówiąc, opcje kafeterii można ułożyć w dowolnej kolejności. Ponadto, powyższe kategorie nie są ściśle rozłączne, bowiem można sobie wyobrazić wybór więcej niż jednej odpowiedzi. Poniżej zostaną omówione najważniejsze rodzaje odpowiedzi na pytania z kafeterią.

- *Wybór pojedynczy*. Z podanej kafeterii należy wybrać tylko jedną, najważniejszą dla respondenta odpowiedź.
- *Wybór wielokrotny ograniczony*. Z podanej kafeterii należy wybrać określoną liczbę odpowiedzi – np. 3. Odpowiedzi mogą być tylko wskazane (zaznaczone) bądź też należy również wpisać kolejność ważności wybranych odpowiedzi (rangowanie).
- *Wybór wielokrotny nieograniczony*. Jak wyżej, ale liczba odpowiedzi jest dowolna.

– *Odpowiedzi rangowane.* Należy podać kolejność opcji kafeterii (dotyczy wyboru wielokrotnego ograniczonego lub całkowitego, a więc wszystkich opcji odpowiedzi).

Odpowiedzi zawierające kafeterię są zwykle ułożone liniowo, tzn. jest tylko jeden ciąg kategorii. Niekiedy jednak, choć rzadko, do pytania dołączona jest tabelka z koniecznością wyboru w dwóch kierunkach (tab. 2.1).

Tab. 2.1. Pytanie z dwiema kafeteriami (dwukierunkowe odpowiedzi): „Jakie rodzaje sportów uprawiasz i kto cię do nich zachęcił?

Kto zachęcił Rodzaj sportu	Instruktor WF	Koledzy	Rodzice	Media
Biegi				
Sporty siłowe	×	×		
Sporty walki				
Gry zespołowe		×	×	×

Przy tak skonstruowanym pytaniu należy unikać wielokrotnego wyboru w obu kierunkach, jak pokazano powyżej, bowiem uniemożliwia to analizę wyników, a przynajmniej znacznie ją utrudnia. W takich wypadkach trzeba przestrzegać zasady, że ew. wielokrotny wybór może dotyczyć tylko jednej zmiennej (np. rodzaju sportu), dla drugiej zmiennej (kto zachęcił) zaś pozostawić wybór tylko jednej opcji.

Pytania zamknięte z odpowiedziami ilościowymi

Odpowiedzi na tę klasę pytań są stopniowane, tworząc określony ciąg logiczny, nie stanowią zatem kafeterii w powyższym rozumieniu. Można tu podać takie przykłady:

Czy należy uprawiać sport? Nie ☐ – Jest to obojętne ☐ – Tak ☐

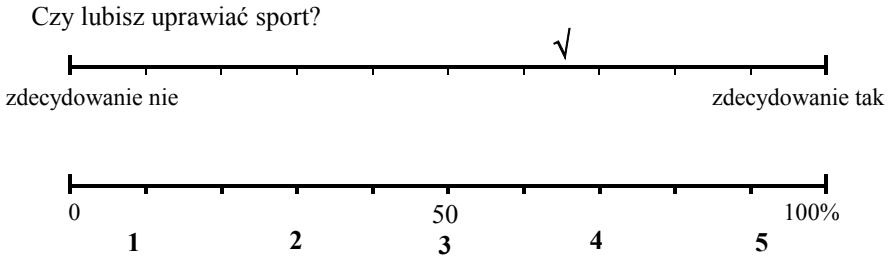
Czy lubisz uprawiać sport? Zdecydowanie nie ☐ – Trochę ☐ – Średnio ☐ – Bardzo ☐ – zdecydowanie tak ☐ (zaznacz właściwe pole)

Odpowiedzi na te pytania oparte są na *skali porządkowej*, którą można traktować jak ilościową, wyraża bowiem natężenie oceny. Skale porządkowe zazwyczaj obejmują 3 – 5 kategorii. Poszczególne kategorie są rozłączne, prawdziwa może być bowiem tylko jedna odpowiedź.

Odpowiedzi zaznaczone w odpowiednich polach należy przenieść do zestawienia w postaci **kodów** liczbowych, np. od 1 (zdecydowanie nie) do 5 (zdecydowanie tak). Kodów tych nie należy umieszczać w arkuszu pytań, aby respondent nie sugerował się wartościującą oceną. Kody można traktować jak wartości zmiennej, jednak pod warunkiem, że odstępy (interwały) między poszczególnymi kategoriami opisowymi są takie same jak między warto-

ściami kodów. Znaczy to, że różnica w stopniu oceny między np. „zdecydowanie nie" i „trochę" powinna być taka sama jak np. między „średnio" i „bardzo", ponieważ różnice między odpowiednimi wartościami kodów są stałe i wynoszą jeden. Warunek ten jest w praktyce trudny do spełnienia, gdyż odczuwane natężenie poszczególnych ocen przez różne osoby może być dalece niejednakowe. Poza tym, respondent często chciałby wybrać odpowiedź pośrednią między dwoma narzuconymi stopniami skali. Jeżeli jednak można przyjąć, że odstępy między poszczególnymi stopniami są podobne, wówczas można przypisać poszczególnym kategoriom wartości punktowe, np. od 1 do 5 i traktować je jak wartości cechy ciągłej, podobnie jak np. oceny szkolne. Należy jednak zwrócić uwagę, że w odróżnieniu od ocen szkolnych, punkty przypisywane poszczególnym kategoriom nie mają znaczenia wartościującego, a tylko porządkujące, ułatwiające wyznaczenie wartości przeciętnej.

Alternatywą do skali ilościowej jest skala ciągła. Zamiast wybierać jeden z 5 wymienionych stopni, można zaznaczyć natężenie oceny na tzw. *skali analogowej*, czyli kresce o długości np. 10 cm. Respondent zaznacza na niej punkt odpowiadający natężeniu jego oceny. Skala analogowa jest pozbawiona wad skali porządkowej, może jednak wymagać pouczenia respondenta o posługiwaniu się nią. Poniżej (ryc. 2.1) pokazano odpowiedź na omawiane już pytanie wyrażoną w skali analogowej – można łatwo zauważyć, że zaznaczony punkt mieści się między kategoriami „średnio" i „bardzo".



Ryc. 2.1. Przykłady skali analogowej z pomocniczą podziałką. Na dolnej skali zaznaczono wartości punktowe mające odpowiadać kategoriom skali porządkowej.

Niekiedy może być korzystne naniesienie pomocniczej podziałki na skalę analogową, użycie jej należy jednak pozostawić wyczuciu badacza. Oceny respondentów naniesione na skalę analogową można łatwo zmierzyć (linijką) i wyrazić liczbowo (zob. przykład w rozdziale 4, sekcja „Zestawianie wyników ankiety"), zatem wyniki takie można opracowywać tak, jak pomiary wyrażone w skali zamkniętej (0 – 100%). W opinii autora należy zalecać posługiwania się skalą analogową zamiast porządkowej wszędzie tam, gdzie to jest możliwe. Skale analogowe są w wielu krajach powszechnie stosowane w badaniach samooceny, np. w badaniach klinicznych.

Wymienione w poprzednim podrozdziale odpowiedzi rangowane należy zaliczyć do ilościowych, bowiem rangi są wyrażone w skali porządkowej i można je analizować statystycznie. Metoda rangowania ma jednak pewne wady. Rozpatrzmy następujące pytanie: „Jakie formy aktywności fizycznej uważasz za istotne dla zdrowia? Uszereguj wszystkie odpowiedzi od najbardziej do najmniej ważnej." Opcje odpowiedzi (kafeteria): „a" gimnastyka poranna, „b" jogging, „c" ćwiczenia na siłowni, „d" aerobik, „e" kolarstwo. Powiedzmy, że respondent podał taką kolejność: d, b, a, e, c, czyli wymienia aerobik na pierwszym miejscu, a ćwiczenia na siłowni na ostatnim. Trudność polega na tym, że gdyby respondent uważał ćwiczenia na siłowni za w ogóle nieważne dla zdrowia, to nie ma możliwości poinformowania o tym, bo należało podać rangi dla wszystkich opcji. Ponadto, skala porządkowa nie uwzględnia odległości między rangami, podobne jak to ma miejsce dla pytań z odpowiedziami stopniowanymi. Wyjściem z tego jest punktowa ocena poszczególnych opcji odpowiedzi. Ocenę taką można przeprowadzić na dwa sposoby:

(a) Stałą liczbę punktów, np. 10, należy dowolnie rozdzielić między opcje odpowiedzi, a więc można np. przypisać po kilka punktów różnym opcjom lub nawet przydzielić wszystkie 10 punktów jednej opcji, a pozostałym zero;
(b) Każdą opcję odpowiedzi ocenić punktowo, np. od 0 do 5. Poniżej pokazano przykład odpowiedzi na powyższe pytanie stosując omówione sposoby oceny.

Kafeteria	Ranking	Punktacja (a)	Punktacja (b)
Gimnastyka poranna	3	1	2
Jogging	2	3	5
Ćwiczenia na siłowni	5	0	1
Aerobik	1	5	5
Kolarstwo	4	1	3
Suma	15	10	16

Tab. 2.2. Rodzaje odpowiedzi na pytanie: „Jakie formy aktywności fizycznej uważasz za istotne dla zdrowia?"
(a) dowolny podział 10 punktów między opcje odpowiedzi
(b) punktowanie każdej opcji odpowiedzi w skali 0 – 5

Jak łatwo zauważyć, punktowa ocena typu (b) daje najwięcej informacji. Wydaje się przy tym, że ocena taka nie sprawi respondentom większych trudności niż rangowanie, a może być nawet łatwiejsza w odbiorze.

Przykłady najważniejszych odpowiedzi kategorialnych i ilościowych i sposób ich zestawiania podano w rozdziale 4.

Pytania otwarte

O ile w pytaniach zamkniętych respondent ma za zadanie wybrać jedną lub więcej gotowych opcji odpowiedzi, o tyle pytania otwarte wymagają odpowiedzi (opisowej lub enumeratywnej), którą respondent musi sam sformułować. Przykłady pytań otwartych:

Jak Pan(i) ocenia relacje szkoła – rodzice? Proszę podać i uzasadnić ocenę .....
.....
(*odpowiedź opisowa*)

Proszę wymienić trzy najważniejsze motywy uprawiania sportu: .....
..... (*odpowiedzi enumeratywne*)

W pierwszym pytaniu badacza interesuje nie tylko sama ocena, ale również, a może przede wszystkim, jej uzasadnienie. W drugim pytaniu chodzi tylko o wyliczenie deklarowanych motywów. W żadnym z tych pytań nie ma zawartej podpowiedzi.

Pytanie otwarte może stanowić również jedną z opcji odpowiedzi na pytanie zamknięte. Na przykład kafeteria pytania nr 4 (s. 24) zawiera pytanie otwarte:

Z kim najchętniej spędzasz wakacje (wybierz jedną odpowiedź):

(a) sam
(b) z rodziną
(c) z partnerem (partnerką)
(d) w grupie
(e) inne – z kim?

Jeżeli preferencje respondenta nie mieszczą się w zamkniętej kafeterii, może on wybrać opcję (e) i wymienić, z kim najchętniej spędza wakacje. Pytanie z taką kafeterią określa się jako półotwarte.

Aby odpowiedzi na pytania otwarte można było opracować ilościowo, należy je uprzednio poklasyfikować. Badacz nie jest w stanie przewidzieć, jakich odpowiedzi na dane pytanie respondent może udzielić, dlatego klasyfikacji dokonuje się po zebraniu wszystkich ankiet. Liczba kategorii klasyfikacji nie powinna być duża – co najwyżej kilka. Jeżeli pozostaje niewielka liczba odpowiedzi, których nie udaje się sklasyfikować, można je ująć w kategorię „inne" i ewentualnie opisać werbalnie w prezentacji wyników, jeśli jest to uzasadnione wagą pytania i odpowiedzi.

Należy w tym miejscu przestrzec czytelnika przed nadużywaniem pytań otwartych. O ile otwarta *opcja* w kafeterii może być celowa, o tyle *pytanie* otwarte może utrudnić ocenę wyników ankiety. Lepiej wyróżnić spodziewane kategorie (np. po przeprowadzeniu badań pilotażowych) w kafeterii, bo właściwa odpowiedź na pytanie otwarte może na przykład nie przyjść respondentowi do głowy. Podsumowując – pytania otwarte stosować tylko wówczas, gdy jest to niezbędne w celu uzyskania właśnie swobodnych wypowiedzi.

3. KONSTRUKCJA ARKUSZA PYTAŃ

Układając arkusz pytań należy uwzględnić czynniki mające decydujący wpływ na uzyskiwane wyniki. Najważniejszym z nich jest subiektywizm respondentów, na który składa się chwilowa dyspozycja, zrozumienie treści pytania, nastawienie do ankiety itp. Ta sama osoba może różnie odpowiedzieć na to samo pytanie w innym czasie lub okolicznościach. Ponadto, jeżeli ankieta będzie zawierała zbyt wiele pytań, należy liczyć się ze zmęczeniem respondenta, trudnością jego skupienia się na treści pytania, a w efekcie z odpowiadaniem w sposób machinalny lub niedbały. Tak wypełniona ankieta nie będzie zatem mogła być uznana za rzetelną (patrz niżej). Przede wszystkim należy pamiętać o tym, że wyniki ankiety odzwierciedlają tylko **deklarowane** poglądy, opinie, zachowania itp., w tym także często chęć pokazania się w jak najlepszym świetle (mimo anonimowości ankiety), a nie obiektywne fakty. Ma to szczególne znaczenie przy wyciąganiu wniosków z uzyskanych danych.

Anonimowość ankiet. Ankiety są z założenia anonimowe. Niekiedy jednak zachodzi konieczność ograniczenia anonimowości, gdy wyniki uzyskane z ankiet mają być porównywane z innymi informacjami, np. pomiarami, lub gdy ankieta ma zostać po pewnym czasie powtórzona, a wyniki porównane indywidualnie. Można wówczas prosić respondenta o wpisanie do arkusza swojego nazwiska po uprzednim dokładnym wyjaśnieniu celu odstąpienia od anonimowości; ankietowani muszą na to wyrazić zgodę. Można również prosić o wpisanie do ankiety dowolnego kodu i zapamiętanie go, aby ten sam kod został wpisany do arkusza pomiarów lub przy następnym ankietowaniu. Należy pamiętać o tym, że dla opublikowania tak uzyskanych danych może być konieczna zgoda właściwej Komisji Etyki, uzyskana przed podjęciem badań. W każdym wypadku respondenci muszą mieć jednak poczucie, że ich deklaracje nie zostaną wykorzystane imiennie w jakichkolwiek celach.

Poniżej omówiono pokrótce trzy najistotniejsze parametry ankiety: trafność (uwarunkowana przygotowaniem arkusza pytań), rzetelność (zależna od sposobu przeprowadzenia badań) i reprezentatywność (zależna od założonego planu badań).

Trafność. Pod pojęciem trafności rozumiemy zgodność wyników ankiety z założonym celem. Jeżeli zatem pytania zawarte w ankiecie zostały właściwie dobrane, sformułowane i rozmieszczone, odpowiedzi na te pytania będą zawierały interesującą nas informację, a nie jakąś inną. Decydujące znaczenie mają tu owe trzy wspomniane elementy - dobór pytań, ich sformułowanie i rozmieszczenie wśród innych pytań.

Rzetelność. Rzetelność wyników ankiety lub inaczej - ich wiarogodność, jest warunkiem poprawności wyciąganych wniosków i uogólnień. Istotne znaczenie ma tutaj zaangażowanie respondenta, a więc chęć udzielenia prawdziwych odpowiedzi, dlatego należy zadbać o to, by okoliczności odpowiadania na pytania zawarte w ankiecie były sprzyjające. Jednym z

podstawowych warunków uzyskania rzetelnych odpowiedzi jest anonimowość ankiety. Respondenci muszą mieć pewność, że nie zostaną zidentyfikowani mimo niepodawania nazwiska, dlatego arkusze nie mogą być numerowane (wyjąwszy numerację stron). Innym koniecznym warunkiem jest oczywiście omówiona wyżej trafność ankiety. Wydaje się, że najkorzystniejsze warunki będą wówczas, gdy respondenci odpowiadają na pytania grupowo w obecności ankietera, który może udzielić ewentualnych wyjaśnień i odpowiednio zachęcić do rzetelnych odpowiedzi.

Jeżeli respondenci wypełniają ankiety bez nadzoru, należy liczyć się z niską rzetelnością. Można temu zapobiec, przynajmniej częściowo, ograniczając liczbę pytań do niezbędnego minimum oraz przez włączenie do arkusza tzw. **skali kłamstwa**.

Aby utworzyć skalę kłamstwa, należy kilka pytań powtórzyć stosując inne sformułowania i rozmieścić je w różnych miejscach ankiety. Najlepsze wyniki osiąga się, gdy pytania ankiety nie są grupowane tematycznie, wówczas bowiem najłatwiej wprowadzić powtórki. Na przykład pytanie „czy lubisz wycieczki krajoznawcze" (odpowiedzi w skali stopniowanej lub ciągłej) można również sformułować: „czy zwiedzanie kraju uważasz za interesujące". Te dwa pytania tworzą parę.

Należy jeszcze ustalić kryterium rzetelności - może to być np. zgodność 3 par pytań na 4 tworzące skalę kłamstwa. Przy jednej parze niezgodnych odpowiedzi uznajemy zatem ankietę za jeszcze rzetelną, natomiast dwie lub więcej par niezgodnych dyskwalifikuje ankietę danego respondenta.

Należy podkreślić, że w odróżnieniu od badań np. demograficznych, gdzie arkusze pytań mają na celu ustalenie stanu faktycznego (np. posiadanych dóbr), skala kłamstwa w badaniach **opinii** respondentów ma stwierdzić stopień spójności deklarowanych stwierdzeń. Spójność ta może nie mieć związku z pozytywnym zaangażowaniem respondenta. Tak stosowana skala kłamstwa ma szczególne znaczenie w badaniach psychologicznych.

Reprezentatywność. Reprezentatywność jest związana z poprawnym zdefiniowaniem populacji, która ma być poddana badaniu oraz z takim zaplanowaniem badań, które zapewni losowy dobór osób do badanej próby. Szczegóły dotyczące planowania badań i losowego doboru omówiono w rozdziale 1.

Schemat ankiety

Arkusz pytań składa się zazwyczaj z dwóch części – metryczki i zestawu pytań. Niekiedy, zwłaszcza gdy ankieta jest rozsyłana korespondencyjnie, dołączana jest jeszcze instrukcja wypełnienia arkusza.

Metryczka. Jest to wydzielona, informacyjna część ankiety, umieszczana najczęściej na końcu arkusza. Zawiera pytania dotyczące podstawowych informacji o respondencie, zazwyczaj jest to wiek (lub określony przedział wiekowy), płeć, rodzaj działalności zawodowej, wykształcenie (rodziców, gdy ankietuje się uczniów), miejsce zamieszkania (wg kate-

gorii miejscowości – np. wieś, miasta do 50 000 mieszkańców, miasta 50 000 – 500 000, miasta powyżej 500 000) itp. Przy zbieraniu materiałów naukowych stosowana jest wprawdzie zasada gromadzenia nadmiaru informacji, bo „wszystko może się przydać", jednak w wypadku ankiet należy zadawać tylko te pytania, które będą wykorzystane w opracowaniu. Każde bowiem dodatkowe pytanie zwiększa objętość ankiety, a dodatkowe pytanie w metryczce może wzbudzać opór respondentów.

Zestaw pytań. Jest to zasadnicza część ankiety zawierająca pytania na interesujący badacza temat. Pytania powinny być formułowane jasno i ściśle, a przy tym przystępnym językiem, aby były jednoznaczne i zrozumiałe bez potrzeby dodatkowych wyjaśnień. Zestaw pytań powinien być dany do krytycznego przejrzenia wielu osobom, w tym potencjalnym respondentom, aby pytania możliwie dokładnie spełniały wspomniane warunki. Jeszcze raz należy podkreślić, że ankieta powinna być możliwie zwięzła, a jej rozmiar dostosowany nie tylko do potrzeb tematu badawczego, ale również do wielkości próby, która ma być przebadana i do liczby wyróżnianych kategorii. Związane z tym szczegóły zostały omówione w rozdziale 1

Układ pytań

Jeżeli mamy już zestaw jednoznacznych, poprawnych językowo i komunikatywnych pytań spełniających kryteria trafności, a przy tym zestaw ten jest oszczędny, a więc zawiera tylko to, co potrzebne, to pozostaje jeszcze jedna kwestia – właściwego rozmieszczenia pytań w arkuszu. Zwykle stosowanym, niejako naturalnym sposobem jest układ tematyczny, w którym kolejne pytania tworzą logiczny ciąg. Układ taki jest poprawny i skuteczny przy założeniu pełnego zaangażowania i rzetelności respondentów. Ponieważ w życiu bywa inaczej, układ pytań może być uzależniony od spodziewanej rzetelności odpowiedzi, a więc m.in. od stopnia kontroli przez ankietera. Można zaryzykować tezę, że im mniejsza kontrola (jak np. w przypadku ankiet korespondencyjnych), tym układ pytań może być bardziej bezładny. Wydaje się, że logiczny układ tematyczny bardziej sprzyja udzielaniu odpowiedzi „poprawnych", a ponadto umożliwia respondentowi łatwą kontrolę nad odpowiedziami na poprzednie pytania, a zatem nad spójnością wszystkich odpowiedzi. Jeżeli pytania nie są ułożone tematycznie, respondent musi wyszukiwać w arkuszu pytania o podobnej treści, a to wymaga dodatkowego wysiłku. Jak wspomniano pod hasłem „skala kłamstwa", nietematyczny układ pytań może umożliwić lepszą kontrolę rzetelności ankiety. Z drugiej strony, bezładny układ może sprawić na respondencie niekorzystne wrażenie, co zniechęci go do udzielania rzetelnych odpowiedzi. Jeżeli ankieta jest obszerna i np. podzielona na rozdziały, układ tematyczny może być układem z wyboru. Układ nietematyczny byłby wówczas znacznym utrudnieniem nie tylko dla respondenta, ale przede wszystkim dla badacza.

Pewną odmianą odstępstwa od logicznego układu pytań jest „odwrócenie kierunku" części pytań. Dotyczy to przede wszystkim pytań z odpowiedziami stopniowanymi (w skali porządkowej). Weźmy jako przykład nieco zmodyfikowane pytanie z rozdziału 1:

Jak bardzo lubisz uprawiać sport? (odpowiedzi w skali od 1 do 5 punktów)

Kierunek odpowiedzi jest dodatni: od oceny najniższej do najwyższej. Jeżeli ankieta zawiera więcej pytań z odpowiedziami stopniowanymi, może być niekiedy korzystne odwrócenie kierunku niektórych pytań. Powyższy przykład można sformułować inaczej:

Czy masz niechęć do uprawiania sportu? (odpowiedzi w skali od 1 do 5 punktów)

W tym wypadku kierunek jest odwrotny – im mniej punktów (niższa ocena) tym bardziej respondent lubi uprawiać sport. Oczywiście, jeżeli ankieta zawiera niejednakowo zorientowane pytania, zwłaszcza oceniane w skali punktowej lub analogowej, badacz powinien dysponować tzw. kluczem, wskazującym zorientowanie poszczególnych pytań. Sposób ten jest często wykorzystywany w kwestionariuszach psychometrycznych. W tabeli 3.1. przedstawiono dwa pytania pochodzące z kwestionariusza Spielbergera samooceny lęku. Pytania zostały tak skonstruowane, by suma punktów uzyskanych ze wszystkich odpowiedzi była miernikiem poziomu lęku (cechy lub stanu). Odwrotne zorientowanie pytań utrudnia „machinalne" udzielanie odpowiedzi.

Tab. 3.1. Przykład pytań zorientowanych odwrotnie względem siebie (z kwestionariusza Spielbergera samooceny lęku)

	Zdecydowanie nie	Raczej nie	Raczej tak	Zdecydowanie tak
Czuję się nadmiernie podniecony	1	2	3	4
Jestem radosny	4	3	2	1

Powyższa punktacja nie jest oczywiście zamieszczona w kwestionariuszu, tylko w szablonie (kluczu), badany ma jedynie zaznaczyć właściwe pole.

Odwrotne zorientowanie nie musi być oczywiście ograniczone do pytań z odpowiedziami ilościowymi, można je stosować również do pytań wymagających odpowiedzi kategorialnych. Szczególnie przydatne jest wykorzystanie odwrotnie zorientowanych pytań w skali kłamstwa. Odwrotne zorientowanie jednego z pytań w parze (zob. s. 21) narzuca inne ich sformułowania, co stanowi dodatkowy „kamuflaż".

Stosowanie odwrotnie zorientowanych pytań musi być jednak dobrze przemyślane i oszczędne, żeby uniknąć błędnych odpowiedzi wynikających z nieuwagi respondenta. Dotyczy to szczególnie ankiet korespondencyjnych. Należy się tu kierować podobną ostrożnością jak w wypadku pytań powtarzanych, odmiennie sformułowanych (zob. uwagi o skali kłamstwa).

4. SPORZĄDZANIE ZESTAWIEŃ

Analiza wyników zestawień liczbowych jest znacznie ułatwiona wówczas, gdy zestawienia takie są właściwie ułożone, co umożliwia porządkowanie i klasyfikację wyników według określonych kryteriów. Bardzo dobrym narzędziem jest arkusz kalkulacyjny Excel, który umożliwia prosty i przejrzysty sposób rejestracji wyników, ich sortowanie, przeprowadzenie niezbędnych obliczeń i przygotowanie wykresów do prezentacji. Maksymalna pojemność jednego arkusza wynosi 65 536 wierszy i 256 kolumn. Poniżej zostaną przedstawione zasady sporządzania zestawień.

Zebrane ankiety należy najpierw dokładnie przejrzeć. Ankiety osób niespełniających kryteriów wyboru do badań należy odrzucić. Na przykład: Badano studentki i studentów I i II roku studiów, ale przez omyłkę ankietowano również osoby z III roku studiów; te ankiety należy odrzucić. Inny przykład – wśród 74 zebranych ankiet 5 miało braki w odpowiedziach, a pozostałe 69 zawierało komplet odpowiedzi. Przed przystąpieniem do analizy należy te 5 ankiet odrzucić, podając tylko odsetek odrzuconych: 5/74 = 7%, a wszystkie zestawienia i obliczenia przeprowadzić tylko na kompletnych ankietach (n = 69).

Zestawianie wyników indywidualnych ankiet

Ogólną zasadą jest rejestracja wyników danej ankiety w ten sposób, aby w jednym wierszu mieściły się wszystkie wyniki jednego respondenta, poszczególne pytania zaś – w kolejnych kolumnach. Należy jedynie ustalić sposób zapisywania odpowiedzi na poszczególne pytania w zależności od ich typu. Poniżej pokazano fragment fikcyjnej ankiety zawierającej omówione rodzaje pytań i odpowiedzi oraz sposób ich zapisu w arkuszu Excela.

Nr ankiety 56

Metryczka

Płeć	K M√
Rok studiów	I II √
Miejsce zamieszkania	(a) Wieś (b) Miasto do 100 tys. (c) Miasto powyżej 100 tys. √

Pytania

Nr	Pytanie	Opcje odpowiedzi	
1.	Czy uczęszczasz na zajęcia WF	TAK √ NIE	
2.	Stosunek do zajęć WF w szkole średniej	Pozytywny √ Obojętny Negatywny	
3.	Jak dużo wysiłku wkładasz w przygotowanie się do egzaminu (zaznacz na skali poniżej) wcale — v — maksymalnie		
4.	Z kim najchętniej spędzasz wakacje (wybierz jedną odpowiedź)	(a) sam (b) z rodziną (c) z partnerem (partnerką) (d) z grupą √ (e) inne (jakie)	
5.	Gdzie najchętniej spędzasz wakacje (wybierz dowolną liczbę odpowiedzi)	(a) w górach (b) nad morzem √ (c) w lesie √ (d) w mieście (e) nad jeziorem √ (f) inne (jakie)	
6.	Jakie zajęcia ruchowe uważasz za istotne dla utrzymania dobrej formy fizycznej (podaj kolejność wszystkich wymienionych obok, począwszy od najważniejszego)	(a) spacery (b) siłownia (c) jogging (d) rower (e) pływanie (f) inne (jakie): *gry zesp.*	4 2 5 3 1 6

	A	B	C	D	E	F	G	H	I	J	K
1	nr	płeć	r.stud.	m.zam.	1	2	3	4	5	6	5n
2	56	M	II	c	1	1	6,5	d	bce	ebdacf	3
3	14	K	I	b	0	0	8,2	c	ac	aedcfb	2
4	17	K	I	c	1	1	6,8	b	abe	ecdabf	3
5	38	K	II	a		0	9,0	b	d		1

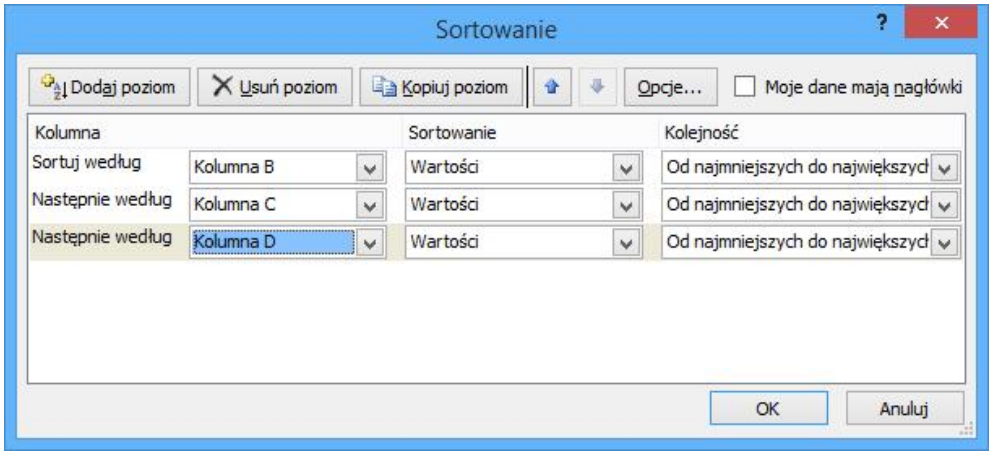
Ryc. 4.1. Przykład wpisywania wyników ankiety do arkusza EXCEL
Pełne zestawienie wyników dla jednej grupy studentek przedstawiono w aneksie

Powyżej pokazano zestawienie kilku ankiet w arkuszu Excela zgodnie z zasadą, że wszystkie dane z określonej ankiety zebrane są w jednym wierszu. Jak widać, ankiety nie były w żaden sposób uporządkowane. W podobny sposób wpisuje się do arkusza wyniki

wszystkich ankiet. W razie braku odpowiedzi na jakieś pytanie, należy zostawić **puste pole**, jak w ankiecie nr 38 (E5 i J5). W kolumnie G (pytanie 3) wpisano zmierzone odległości (w cm) od początku skali do „ptaszka".

Zanim przystąpi się do analizy danych, należy sporządzić kopię pliku (pod inną nazwą), która będzie służyła jako „rezerwa" w razie np. utraty części lub wszystkich danych, pomieszania danych przy pomyłkach w sortowaniu itp.

Pierwszym krokiem w przygotowaniu danych do analizy będzie odpowiednie posortowanie ankiet wierszami według żądanych kryteriów, np. płci, a w obrębie płci – roku studiów czy miejsca zamieszkania. W tym celu należy zaznaczyć obszar zawierający dane: kliknąć myszą w numer wiersza „2", następnie przejść **pionowym suwakiem** na koniec obszaru danych i trzymając wciśnięty klawisz „Shift", kliknąć w numer ostatniego wiersza z danymi. Zostanie zaczerniony cały obszar. Następnie kliknąć prawym przyciskiem, wybierz „Sortuj", potem „Sortowanie niestandardowe", potem „Dodaj poziom". Wyłączyć okienko „Moje dane mają nagłówki". Kliknąć dwa razy zakładkę „Dodaj poziom" i wybrać litery odpowiadające kolumnom danych. W podanym przykładzie (ryc. 4.2) sortowanie nastąpi według płci (kolumna B), w obrębie płci według roku studiów (kolumna C), a w obrębie każdego roku studiów według miejsca zamieszkania (kolumna D).



Ryc. 4.2. Okno dialogowe polecenia „Sortowanie"

Po tym zabiegu ankiety będą podzielone na 12 grup: 2 płcie, 2 lata studiów, 3 kategorie według miejsca zamieszkania. Poszczególne grupy należy od siebie oddzielić pustymi wierszami w liczbie 3 + maksymalna liczba opcji odpowiedzi. W podanym przykładzie będzie to 9 wierszy. Można również dane każdej grupy umieścić w oddzielnych arkuszach, ale może to utrudnić zestawianie i analizę danych.

	A	B	C	D	E	F	G	H	I	J	K
1	nr	płeć	r.stud.	m.zam.	1	2	3	4	5	6	5n
76		K	I	wieś	73	73	74	74	74	72	
77					71		6,20				
78							1,73				
79	*a*	-1	1			5		6	45		2
80	*b*	0	2			32		16	42		38
81	*c*	1	3			36		31	29		21
82	*d*		4					19	17		9
83	*e*		5					2	51		3
84	*f*		6						14		1

Ryc. 4.3. Zestawienie wyników ankiet studentek I roku zamieszkałych na wsi

W wyniku segregacji otrzymano np. 74 ankiety (wiersze 2 – 75) studentek I r. studiów, mieszkanek wsi. Rozpatrzmy sporządzenie zestawienia odpowiedzi na poszczególne pytania tych ankiet. Pierwsze 3 wiersze (76 – 78) rezerwuje się na zestawienia ogólne, dalszych 6 (79 – 84) – na szczegółowe (ryc. 4.3).

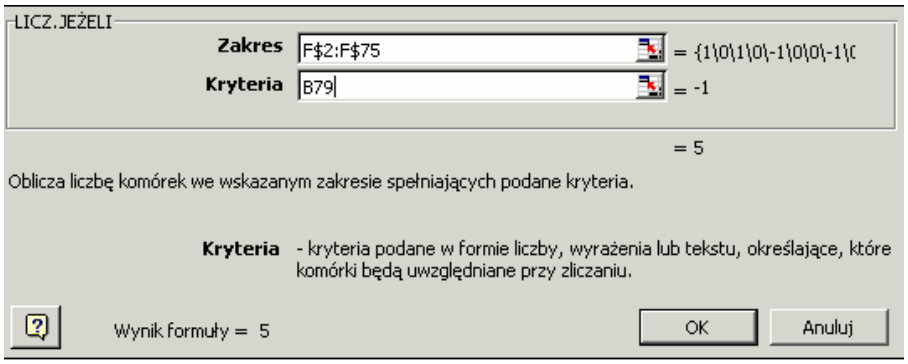
W pierwszym wierszu (76) należy wstawić liczby odpowiedzi: zaznaczyć komórkę E76 i wywołać funkcję ILE.NIEPUSTYCH, a w pasku okna dialogowego wpisać ręcznie zakres danych E2:E75 lub przeciągnąć ten zakres myszką. Otrzymany wynik 73 mówi, że jedna studentka nie odpowiedziała na pierwsze pytanie (wiersz 5, ankieta nr 38). Po to właśnie pozostawia się pustą komórkę, jeżeli nie ma odpowiedzi. Następnie należy naprowadzić kursor myszy na prawy, dolny róg komórki zawierającej wynik 73 i trzymając naciśnięty lewy klawisz myszy przeciągnąć w prawo do komórki J76. Okazuje się, że na pytanie 2 były również 73 odpowiedzi, na pytania 3 – 5 odpowiedziały wszystkie studentki z tej grupy, a na pytanie 6 nie odpowiedziały dwie.

Pytanie 1. Na pytanie to możliwe były dwie odpowiedzi – tak lub nie, kodowane jako 1 lub 0, dlatego wystarczy obliczyć sumę liczb w kolumnie E, a więc liczbę jedynek. Należy zaznaczyć komórkę E77, kliknąć w ikonkę SUMA i wpisać ręcznie zakres lub przeciągnąć myszką, jak dla funkcji ILE.NIEPUSTYCH. Wynik sumowania (71) świadczy o tym, że na pytanie 1 twierdząco odpowiedziało 71 studentek, a zatem dwie odpowiedziały negatywnie.

Pytanie 2. Możliwe były 3 rodzaje odpowiedzi, kodowane jako –1, 0, 1, trzeba zatem zliczyć każdy rodzaj. W kolumnie B, w wierszach 79 – 81, wpisać kody odpowiedzi*. Następnie zaznaczyć komórkę F79, wywołać funkcję LICZ.JEŻELI i wpisać adresy, jak pokazano na ryc. 4.4. Konieczne jest wpisanie stałych adresów zakresu F$2:F$75, aby nie ulegały one zmianie przy kopiowaniu formuły („przeciąganiu” myszką). Po kliknięciu w pole

* Miejsce wpisania kodów jest oczywiście dowolne

OK otrzyma się w komórce F79 wynik 5. Należy teraz przejść do komórki F79 i przeciągnąć do pozycji F81. Otrzyma się liczby poszczególnych kodów odpowiedzi (zob. ryc. 4.3).



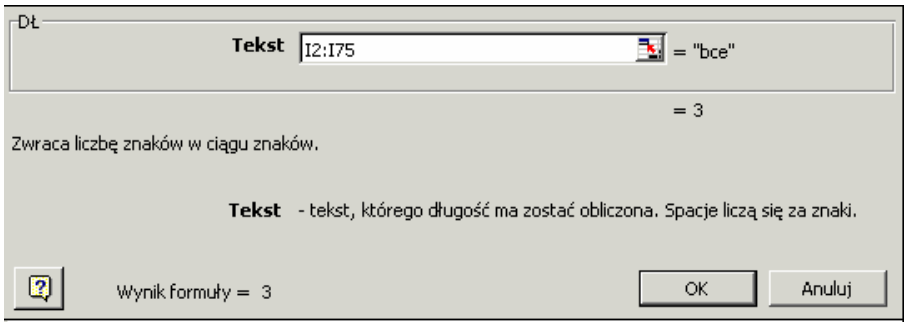
Ryc. 4.4. Okno dialogowe funkcji LICZ.JEŻELI

Pytanie 3. Zmierzone odległości odpowiadają procentom maksymalnego deklarowanego wysiłku włożonego w przygotowanie do egzaminów. Należy zaznaczyć komórkę G77, wywołać funkcję ŚREDNIA, wpisać zakres G2:G75 i nacisnąć klawisz ENTER. W zaznaczonej komórce G78 wywołać funkcję ODCH.STANDARDOWE, wpisać zakres G2:G75, nacisnąć ENTER (zob. ryc. 4.3).

Pytanie 4. W wierszach A79 – A84 wpisać kody odpowiedzi (od „a" do „f", bo jest to największa liczba opcji odpowiedzi w ankiecie; przed i po literze należy wstawić gwiazdkę, żeby były policzone wszystkie znaki)[*]. Zaznaczyć komórkę H79, wywołać funkcję LICZ.JEŻELI i postępować jak w pytaniu 2. Przeciągnąć myszką do pozycji H83, bo w tym pytaniu najwyższym kodem odpowiedzi jest „e". Jak łatwo sprawdzić (ryc. 4.3), suma odpowiedzi wynosi 74, bo tyle ankiet zawierało odpowiedź na to pytanie. Pominiemy na razie odpowiedź *e* jako otwartą.

Pytanie 5. Postępować jak w pytaniu 4. Gdyby nie było gwiazdek przy kodach w kolumnie A, liczone byłyby tylko komórki zawierające jeden kod literowy. Ponieważ można było wybrać dowolną liczbę odpowiedzi, suma wszystkich wyborów (198) jest większa niż liczba ankiet (74). Średnia liczba wyborów wynosi zatem 198 / 74 = 2,68. Interesujące może być dodatkowe zestawienie: ile osób wybrało jedną tylko odpowiedź, ile dwie, trzy itd. Trzeba zatem policzyć, ile znaków zawierają poszczególne komórki zakresu I2:I75. W tym celu należy zaznaczyć komórkę K2 i wywołać funkcję DŁ (długość ciągu znaków). Po przeciągnięciu myszką otrzyma się żądane wartości (ryc. 4.1 i 4.3, kolumna K). Aby zestawić liczby wyborów należy wpisać te liczby, a więc od 1 do 6, np. w komórkach C79 – C84, zaznaczyć komórkę K79, wstawić funkcję LICZ.JEŻELI i postępować jak w pytaniu 2. Jak widać, najwięcej studentek wybrało dwie odpowiedzi.

[*] W wypadku Open Office, zamiast ***a*** należy wstawić **.*a.***



Ryc. 4.5. Okno dialogowe funkcji DŁ

Pytanie 6. Tu sprawa jest bardziej skomplikowana, bowiem trzeba ustalić częstość kolejności poszczególnych wyborów, chyba że wystarczy ustalenie częstości tylko pierwszego (ew. tylko pierwszego i ostatniego wyboru). Rozpatrzymy kolejno obie sytuacje, tj. częstość pierwszego i ostatniego wyboru oraz częstości wszystkich wyborów.

Pierwszy wybór. Należy wpisać kody z jedną gwiazdką, jak na ryc. 4.7, kolumna D, a w komórce L79 wstawić funkcję LICZ.JEŻELI i przeciągnąć myszką do L84. Ponieważ przed kodem nie ma gwiazdki, będą zliczane tylko kody na pierwszej pozycji ciągu. Jeżeli gwiazdka poprzedzi kod (kolumna T), można wstawić tę funkcję np. w komórkę M79 i otrzyma się częstości *ostatniego wyboru*.

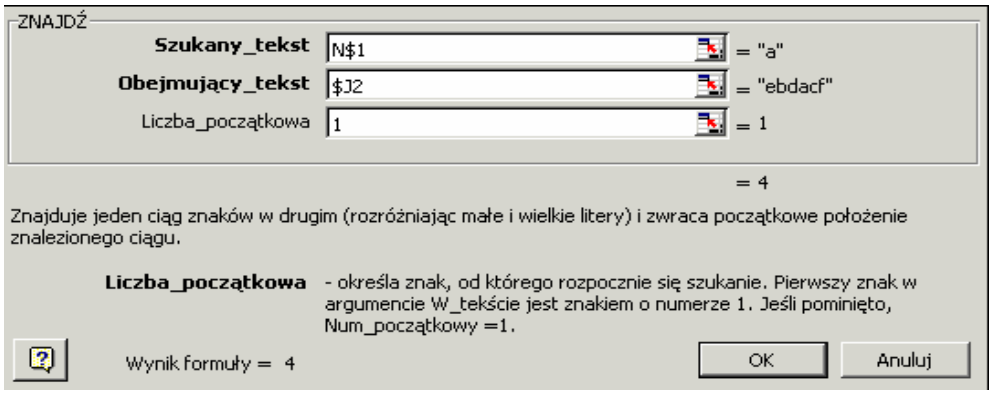
	A	B	C	D		J	M	N	O	P	Q	R	S
1	nr	płeć	r.stud.	m.zam.		6		a	b	c	d	e	f
2	56	M	II	c		ebdacf		4	2	5	3	1	6
3	14	K	I	b		aedcfb		1	6	4	3	2	5
4	17	K	I	c		ecdabf		4	5	2	3	1	6
5	38	K	II	a									

Ryc. 4.6. Kolejność wyborów odpowiedzi na pytanie nr 6

	A	B	C	D		L	M	N	O	P	Q	R	S	T
1	nr	płeć	r.stud.	m.zam.		"I"	ost.	a	b	c	d	e	f	
76		K	I	wieś										
77														
78														
79	*a*	-1	1	a*		8	10	8	4	7	19	24	10	*a
80	*b*	0	2	b*		3	25	3	3	7	26	8	25	*b
81	*c*	1	3	c*		31	0	31	23	14	3	1	0	*c
82	*d*		4	d*		12	2	12	15	24	11	8	2	*d
83	*e*		5	e*		17	0	17	27	20	3	5	0	*e
84	*f*		6	f*		1	35	1	0	0	10	26	35	*f

Ryc. 4.7. Zestawienie częstości pierwszego i ostatniego wyboru odpowiedzi na pytanie 6

Wszystkie wybory. Zestawienie wszystkich wyborów wymaga podania kolejności poszczególnych 6 kodów odpowiedzi na pytanie nr 6 dla każdej ankiety. Wymaga to dołączenia sześciu kolumn, jak pokazano na ryc. 4.6 (kolumny N – S) i wpisania kodów od „a" do „f" w pierwszym wierszu tych kolumn. Następnie w komórkę N2 należy wstawić funkcję ZNAJDŹ (ryc. 4.8), w polu „Szukany tekst" wstawić adres kodu *a* (N$1), w polu „Obejmujący tekst" adres pierwszej komórki zakresu danych ($J2), a w polu „Liczba początkowa" liczbę 1. Liczby zawarte w tych komórkach (wiersz 2, ankieta nr 56) oznaczają, że kod „a" znajduje się na 4. miejscu, kod „b" na 2., kod „c" na 5. itd. Jeżeli teraz zaznaczy się komórki N2 – S2 i przeciągnie myszką do wiersza 75, otrzyma się kolejności wyborów dla poszczególnych ankiet.



Ryc. 4.8. Okno dialogowe funkcji ZNAJDŹ

Aby zestawić kolejności, należy znowu posłużyć się funkcją LICZ.JEŻELI, ale zamiast kodów literowych użyć liczb wyznaczających kolejność, a więc od 1 do 6 (ryc. 4.7., kolumna C). Wyniki będą zawarte w komórkach od N79 do S84, przy czym w kolumnie N będą zawarte częstości pierwszego wyboru (jak w kolumnie L), a w kolumnie S – ostatniego (jak w kolumnie M). Liczby w każdym wierszu od 79. do 84. (kolumny N do S), jak również liczby w każdej kolumnie od N do S (wiersze 79 – 84) muszą się oczywiście sumować do liczby ankiet zawierających odpowiedź na pytanie nr 6, a więc do 72.

Jeżeli liczba wyborów z rangowaniem jest ograniczona np. do trzech, postępuje się jak wyżej. W wielu komórkach (po 4 w każdym wierszu) wystąpią komunikaty #ARG, a sumy w zestawieniu (kolumny N – S) będą mniejsze niż liczba ankiet.

Pytania 4, 5 i 6 zawierały w kafeterii po jednej odpowiedzi otwartej. W powyższych zestawieniach odpowiedzi otwarte były ujęte jako jedna kategoria, chociaż w rzeczywistości rozmaitość odpowiedzi może być duża. Odpowiedzi otwarte na poszczególne pytania należy pogrupować, aby ułatwić analizę wyników. Liczba kategorii nie może być duża, najlepiej od 2 do 4, gdyż odpowiedzi ulegną zbytniemu rozdrobnieniu. Na przykład, na pytanie 5

(sposób spędzania wakacji) było 14 otwartych odpowiedzi (zob. ryc. 4.3): u rodziny, z bratem, samotna włóczęga, w domu, na wsi, pracując na wsi, z rodzicami, w domu, u ciotki, u brata za granicą, oprowadzając wycieczki, u rodziny, na wsi, u rodziny. Można tu wyróżnić odpowiedzi w kategorii „z rodziną" (łącznie 9), „na wsi" (3) i dwie różne. Dane dotyczyły tylko jednej grupy (studentki I roku, mieszkanki wsi), podziału odpowiedzi na kategorie należy zaś dokonać biorąc pod uwagę wszystkie grupy łącznie.

Zestawianie wyników zbiorczych

Wprowadzenie indywidualnych danych z ankiet do arkusza Excela, czyli zestawienie wyników poszczególnych ankiet, jest pierwszym krokiem do analizy danych. Koniecznym dalszym etapem jest podsumowanie wyników, czyli sporządzenie zestawienia zbiorczego dla każdej grupy ankiet. W omawianym tu (fikcyjnym) przykładzie jest 12 grup (2 płcie × dwa lata studiów × trzy kategorie miejsca zamieszkania).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	płeć	r.stud.	zam.		1	2	3	4	5	5n	6	a	b	c	d	e	f	
2	**K**	**I**	**a**	n	73	73	74	74	74		72							
3	N	74		1	71													
4				śr.	(5n)		6,20		2,68									(4-6)
5				SD	*1*		1,73	6	45	2		8	4	7	19	24	10	*a*
6					*2*			16	42	38		3	3	7	26	8	25	*b*
7				(2)	*3*			31	29	21		31	23	14	3	1	0	*c*
8				*-1*	*4*	5		19	17	9		12	15	24	11	8	2	*d*
9				*0*	*5*	32		2	51	3		17	27	20	3	5	0	*e*
10				*1*	*6*	36			14	1		1	0	0	10	26	35	*f*
11								Σ	198									
12	**K**	**II**	**a**	n	52	52	52	52	52		52							
13	N	52		1	47													
14				śr.	(5n)		7.11		2.81									(4-6)
15				SD	*1*		1.57	2	36	5		2	2	3	21	17	7	*a*
16					*2*			12	41	15		2	1	3	22	14	10	*b*
17				(2)	*3*			30	30	25		25	16	5	2	2	2	*c*
18				*-1*	*4*	5		7	12	2		7	11	19	2	5	8	*d*
19				*0*	*5*	20		1	21	2		16	18	6	2	10	0	*e*
20				*1*	*6*	27			6	3		0	4	16	3	4	25	*f*
21								Σ	146									

Ryc. 4.9. Zbiorcze zestawienie danych dla dwóch grup: studentek I i II roku, mieszkanek wsi. W nawiasach podano numery pytań, do których odnoszą się oznaczenia kursywą.

Na poprzednich stronach omówiono sposoby zestawiania danych z różnego typu pytań i przedstawiono na rycinach 4.3 i 4.7, są to jednak tylko zestawienia robocze, a zawarte w nich dane nie są *liczbami*, tylko *wynikami formuł*, które nie mogą być bezpośrednio kopiowane. Zbiorcze zestawienie najlepiej sporządzić w osobnym arkuszu lub pliku przez przekopiowanie do niego sumarycznych danych z poszczególnych grup, jak pokazano na rycinie 4.9. Porównując powyższą tabelę z zestawieniami na ryc. 4.3 i 4.7, można łatwo zidentyfikować poszczególne pozycje i sposób ich rozmieszczenia.

Kopiowanie bloków danych: zaznaczyć obszar z danymi, kliknąć w ikonkę „kopiuj", następnie zaznaczyć miejsce, do którego dane będą kopiowane (tylko pierwszą komórkę), wywołać polecenie Edycja – Wklej specjalnie – *Wartości* i ew. zmniejszyć liczbę miejsc dziesiętnych. W wierszu nagłówkowym (1) podane są numery pytań (E1 – Q1), w pierwszym wierszu zestawienia – liczby odpowiedzi udzielone na dane pytanie (E2 – K2 i E12 – K12), a w komórkach B3 i B13 – liczby ankiet w danej grupie. Kolumna G zawiera średnie i odchylenia standardowe, w komórkach I4 i I14 podane są średnie liczby wyborów z kafeterii.

W taki sam sposób należy sporządzić zbiorcze zestawienia dla wszystkich grup. Umożliwi to sporządzenie zestawień analitycznych, niezbędnych do przeprowadzenia analizy danych, jak opisano w następnym rozdziale.

5. ANALIZA ZESTAWIEŃ

W analizie danych ankietowych mamy do czynienia głównie z analizą zestawień, chociaż niekiedy analizuje się również indywidualne dane, np. gdy w tabelach są dane ilościowe, które koreluje się między sobą.

Analiza danych, nawet tak stosunkowo niewielkiej ankiety jak omawiano w powyższym, fikcyjnym przykładzie, wymaga zestawiania danych w rozmaitych kombinacjach, dlatego praca z zestawieniem zbiorczym (ryc. 4.9) byłaby bardzo skomplikowana. Ponadto, analizie poddaje się nie całe zestawienie zbiorcze, a dane dla określonego pytania z poszczególnych grup. Praca będzie zatem znacznie ułatwiona, gdy z zestawień zbiorczych sporządzi się zestawienia analityczne dla każdego pytania oddzielnie. Poniżej zostaną omówione zestawienia analityczne odpowiedzi na niektóre pytania.

Zestawienia analityczne

W tabeli poniżej (ryc. 5.1) pokazano analityczne zestawienie wyników wszystkich grup dla pytania nr 1. Dla ułatwienia późniejszej analizy obliczono odsetki odpowiedzi „nie" (0) dla obu lat studiów łącznie, ponieważ liczby odpowiedzi negatywnych są bardzo małe. Przy tylko dwóch opcjach odpowiedzi – 0 lub 1, nie potrzeba obliczać odsetków odpowiedzi „tak", bo stanowią one uzupełnienie do 100%.

Tab. 5.1. Analityczne zestawienie danych wszystkich grup dla pytania nr 1. BO – brak odpowiedzi na pytanie; N – liczba odpowiedzi; 0 (%) – odsetki odpowiedzi „nie" dla obu lat studiów łącznie. Całkowita liczba ankiet: N = 74.

płeć	**K**	Pytanie 1						**M**						
m.zam.	a		b		c			a		b		c		
rok st.	I	II	I	II	I	II	Σ	I	II	I	II	I	II	Σ
BO	1	0	2	0	1	0	4	0	0	1	1	0	2	4
N	73	52	45	56	88	103	417	46	38	56	61	82	112	395
Nie (0)	2	5	7	6	11	17	48	5	2	4	8	5	11	35
Tak (1)	71	47	38	50	77	96	379	41	36	52	53	77	101	360
0 (%)	5,6		12,9		14,7			8,3		10,3		8,2		

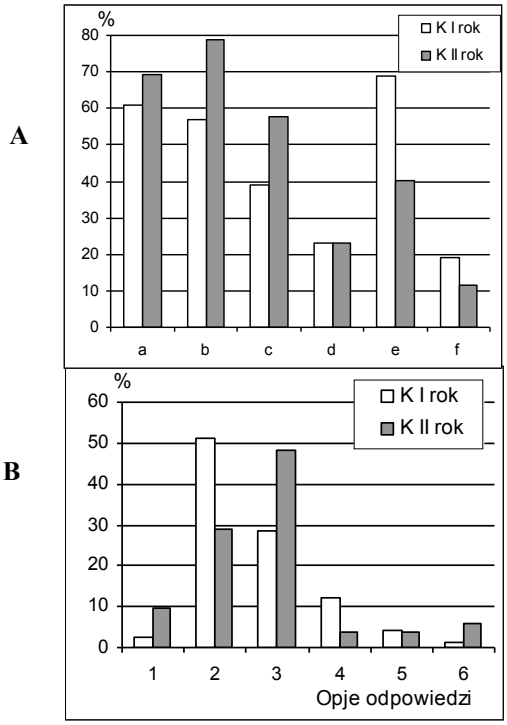
W podobny sposób można układać tabele zestawień analitycznych dla innych pytań (cech) z tym, że jeżeli dla danego pytania są więcej niż dwie opcje odpowiedzi, trzeba obliczyć odsetki dla wszystkich opcji i ewentualnie sporządzić wykres (kolumnowy lub liniowy). Wykres taki, będący obrazem rozkładu częstości dla danego pytania (zmiennej), ułatwi znalezienie różnic, które należy zweryfikować rachunkiem statystycznym. Rozkłady odsetkowe należy odnosić do liczby odpowiedzi na dane pytanie, a nie do całkowitej liczby ankiet i

uzupełnić informacją o odsetkowym braku odpowiedzi na to pytanie (w stosunku do całkowitej liczby ankiet). Na przykład, całkowita liczba odpowiedzi studentek na pytanie 1 (tab. 5.1) wynosi 417, a 4 studentki nie odpowiedziały na to pytanie, razem było zatem 421 ankiet. Odsetek braków odpowiedzi wynosi zatem 100· 4 / 421 = 0,95%.

Dla pytania nr 5, w którym można było wybrać dowolną liczbę opcji z kafeterii, sporządzono dwa zestawienia: jedno zawiera całkowite liczby wyborów poszczególnych opcji (a więc ich suma jest większa niż liczba ankiet), a drugie – sumę częstości wyborów, a więc podaje ilu respondentów wybrało tylko jedną opcję, ilu dwie itd. (por. ryc. 4.3 i 4.9). Zestawienie analityczne dla tego pytania będzie więc wyglądało następująco:

płeć	K	Pytanie 5	
m.zam.	a		
rok st.	I	II	
BO	0	0	
N	74	52	
śr.	2,68	2.81	
a	45	36	
b	42	41	
c	29	30	
d	17	12	
e	51	21	
f	14	6	
	198	146	
1	2	5	
2	38	15	
3	21	25	
4	9	2	
5	3	2	
6	1	3	

Ryc. 5.1. Zestawienie analityczne dla pytania nr 5 (fragment)



Ryc. 5.2. Wykresy odsetkowe do zestawienia analitycznego na rycinie obok

W tabeli (ryc. 5.1) pokazano tylko fragment danych obejmujący pierwsze dwie grupy. Odpowiadające tym danym wartości odsetkowe pokazano na wykresach (ryc. 5.2). Z wykresu „A" wynika, że analizie warto poddać tylko dane dla opcji odpowiedzi „b", „c" i „e", natomiast z wykresu „B" – tylko częstości dwóch i trzech wyborów. Pozostałe różnice są zbyt małe, ale dla wprawy warto to sprawdzić rachunkowo.

Powyższe przykłady ukazują ogólnie zasadę tworzenia zestawień analitycznych. Podobne zestawienia dla innych typów pytań będą pokazane poniżej w podrozdziale dotyczącym analizy danych.

Zdecydowaną większość danych ankietowych stanowią liczebności. Należy tu wyróżnić dwie kategorie.

1. Odnoszące się do wyborów rozłącznych (dane sumują się do liczby ankiet lub do 100%),
2. Odnoszące się do wyborów wielokrotnych (sumy mogą być większe niż liczba ankiet; dane nie sumują się do 100%).

Statystyczna analiza liczebności

Do analizy liczebności służy funkcja χ^2 (chi-kwadrat) w postaci logarytmicznej (tzw. funkcja G; zob. Literatura uzupełniająca, poz. 7). Jej stosowanie pokazane jest na poniższym przykładzie. Poniższa tabela 5.3 jest fragmentem danych pokazanych w tabeli 5.1. Liczby znajdujące się w pogrubionych ramkach (2; 5; 71; 47; 125) tworzą tzw. **obszar A**, a liczby w zacienionych polach (7; 118; 73; 52), czyli tzw. sumy brzegowe – **obszar B**. Obszary A i B można wyróżnić w każdej tabeli, niezależnie od jej rozmiaru.

	I	II	Σ
Nie (0)	2	5	7
Tak (1)	71	47	118
Σ	73	52	125

Ryc. 5.3. Fragment danych z tabeli 5.1 (K, opcja odpowiedzi *a*). Liczby w grubych ramkach – obszar A, w polach zacieniowanych – obszar B.

Analiza polega na porównaniu względnych (procentowych) częstości występowania odpowiedzi „nie" i „tak" na I i II roku studiów z tzw. częstościami oczekiwanymi, obliczonymi przy założeniu tzw. hipotezy zerowej. Hipoteza zerowa mówi, że względne częstości na obu latach studiów są identyczne. Względną częstością oczekiwaną odpowiedzi „nie" jest (procentowy) stosunek *wszystkich* odpowiedzi „nie" do *wszystkich* odpowiedzi w ogóle, czyli 100×(7 / 125) = 5,6%. Częstość odpowiedzi „tak" wynosi odpowiednio 100×(118 / 125). Zgodnie z hipotezą zerową częstość odpowiedzi „nie" na obu latach studiów będzie taka sama i równa 5,6%. W rzeczywistości, na I roku częstość ta wynosi 100×(2 / 73) = 2,7%, a na II roku - 100×(5 / 52) = 9,6%. Różnica między nimi wydaje się znaczna, ale dane procentowe oparte są na bardzo małych liczebnościach (2 i 5), dlatego należy zweryfikować hipotezę zerową za pomocą rachunku. W tej konkretnej sytuacji nie można użyć tzw. klasycznej analizy chi-kwadrat wykorzystującej liczebności rzeczywiste i oczekiwane, bowiem „klasyczny" rachunek wymaga, aby żadna liczebność oczekiwana nie była mniejsza od 5, a według niektórych statystyków nawet od 8. Dlatego do analizy liczebności będzie stosowana wspomniana wyżej funkcja G, która jest wolna od tego rodzaju ograniczeń.

Obliczenie wartości chi-kwadrat w postaci funkcji G polega na tym, że dla wszystkich liczebności (n_i) z tabelki na rycinie 5.3 oblicza się wyrażenie $\boxed{n_i \cdot \ln(n_i)}$, a następnie sumuje

się otrzymane wartości oddzielnie dla obszaru A i B. Ostatecznie otrzymujemy wartość funkcji z wzoru: $\chi^2 \cong G = 2 \cdot (A - B)$. Wartość tę porównuje się z wartością tablicową dla żądanego prawdopodobieństwa (zwykle 0,05 lub mniej) i liczby stopni swobody równej $df = (w - 1) \cdot (k - 1)$, gdzie w – liczba wierszy, a k – liczba kolumn w analizowanej tabeli (tylko dane, bez sum). Tablicę wartości funkcji chi-kwadrat zamieszczono na stronie 61, a wartości wyrażenia $n \cdot \ln(n)$ dla n = 0 do 459 na stronie 62.

Wykonanie rachunków w arkuszu Excela (ryc. 5.4): po wpisaniu liczebności z obszarów A (komórki A2 – A6) i B (komórki A8 – A11), w komórce B2 sąsiadującej z pierwszą wartością wpisuje się formułę: `=A2*ln(A2)`, a następnie przeciąga myszką. Komórki B7 i B12 zawierają sumy dla obszarów A i B, a komórka B13 wartość funkcji G. Analizowana tabela ma 2 wiersze i 2 kolumny (ryc. 5.3), a zatem 1 stopień swobody. Tablicowa wartość χ^2 dla df = 1 i p = 0,05 wynosi 3,84. Otrzymana wartość (2,74) jest mniejsza od tablicowej, zatem nie ma podstaw do odrzucenia hipotezy zerowej. Wniosek: badane grupy (mieszkanki wsi, studentki I i II roku) nie różnią się znamiennie pod względem częstości odpowiedzi „tak" lub „nie" na pytanie nr 1.

	A	B
1	n	ln(n)
2	2	1.39
3	5	8.05
4	71	302.65
5	47	180.96
6	125	603.54
7	**ΣA**	**1096.59**
8	7	13.62
9	118	562.94
10	73	313.20
11	52	205.46
12	**ΣB**	**1095.22**
13	G	2.74

Ryc. 5.4. Obliczanie wartości funkcji χ^2 z danych na ryc. 5.3

W tej sekcji opisano tylko rachunkową technikę obliczania funkcji G dla jednej tabeli czteropolowej (2 wiersze, 2 kolumny), którą można stosować do tabeli dowolnej wielkości. Tabele większe niż 2×2 będą miały więcej niż jeden stopień swobody, a ich analiza będzie bardziej złożona, co zostanie omówione w dalszej części.

Niekiedy zachodzi potrzeba porównania dwu liczebności lub wartości procentowych bez możliwości ułożenia tabeli. Przykład: spośród 464 osób 227 zakwalifikowano do kategorii A, a 175 do kategorii B. Odpowiednie odsetki wynoszą 37.9 i 48.9%. Czy ta różnica jest znaczna? Najprostszy dokładny test: $t = \frac{227 - 175}{\sqrt{(227 - 175)}} = 2,54$ a więc różnica jest znamienna ($p<0.05$).

Analiza wyborów rozłącznych

Jako przykład wyborów rozłącznych może służyć zestawienie pokazane w tabeli 5.1. Na początek można sprawdzić, czy poszczególne lata studiów różnią się częstością odpowiedzi „tak" lub „nie". Obliczenia takie zostały przeprowadzone dla studentek mieszkanek wsi (m.zam. *a*; ryc. 5.3 i 5.4), należy zatem wykonać obliczenia dla pozostałych 5 zespołów (studentki, *b* i *c*, studenci, *a* – *c*). Obliczone wartości G wynoszą kolejno 2,74; 0,52; 0,27; 0,89; 1,15; 0,89. Żadna wartość nie przekracza 3,84, a więc można uznać, że między I i II

rokiem brak statystycznie znamiennych różnic, co upoważnia do potraktowania obu lat studiów łącznie. Pozostaje sprawdzić, czy występują różnice między kategoriami miejsca zamieszkania, a także między płciami. W tym celu trzeba utworzyć nowe zestawienia analityczne (tab. 5.2).

Analizę rozpoczniemy od zbadania, czy częstości odpowiedzi na pytanie nr 1 zależą od płci. Z sum liczb odpowiedzi dla każdej płci (tab. 5.2) tworzymy nową tabelkę (tab. 5.3), dla której należy wykonać obliczenia, jak pokazano wyżej, otrzymując wartość G = 1,29. Można stwierdzić, że w odpowiedziach na pytanie nr 1 nie ma znamiennej różnicy między studentkami i studentami globalnie.

Tab. 5.2. Zestawienie częstości odpowiedzi na pytanie nr 1 według miejsca zamieszkania i płci dla I i II roku studiów łącznie

	Studentki (G = 6,41)				Studenci (G = 0,39)			
m.zam.	a	b	c	Σ	a	b	c	Σ
Nie (0)	7	13	28	48	7	12	16	35
Tak (1)	118	88	173	379	77	105	178	360
Σ	125	101	201	427	84	117	194	395

Tab. 5.3. Zestawienie globalnych częstości odpowiedzi na pytanie nr 1 według płci (G = 1,29)

	K	M	Σ
Nie (0)	48	35	83
Tak (1)	379	360	739
Σ	427	395	822

Z kolei należy dokonać analizy według miejsca zamieszkania – oddzielnie dla studentek i studentów (w tabeli 5.2 zaznaczono obszary B). Otrzymane wartości funkcji G wynoszą 6,41 (studentki) i 0,39 (studenci). Każda z tych tabelek ma 2 stopnie swobody (bo każda ma 2 wiersze i 3 kolumny), a wartość χ^2 dla p = 0,05 i df = 2 wynosi 5,99 (zob. s. 62).

Gdy tabela danych ma jeden stopień swobody, wystarczy posłużyć się jedną wartością tablicową χ^2 = 3,84: jeżeli obliczona wartość G jest większa lub równa 3,84, odrzucamy hipotezę zerową i wnioskujemy, że dane w tabeli są znamiennie zróżnicowane; w przeciwnym wypadku uznajemy, że w tabeli brak różnic znamiennych. Jeżeli natomiast tabela ma dwa lub więcej stopni swobody, wówczas możliwe są trzy sytuacje:

1. Obliczona wartość G jest mniejsza niż **3,84** (wartość dla **jednego** stopnia swobody); wynik taki oznacza, że w tabeli danych brak jakichkolwiek znamiennych zróżnicowań.
2. Obliczona wartość G jest zawarta między 3,84 a wartością tablicową dla danej liczby stopni swobody; wynik taki oznacza, że niektóre dane w tabeli **mogą być** znamiennie zróżnicowane.

3. Obliczona wartość G jest większa lub równa wartości tablicowej dla danej liczby stopni swobody (np. dla df = 2 będzie to 5,99); wynik taki oznacza, że niektóre dane w tabeli **na pewno są** znamiennie zróżnicowane.

W sytuacjach (2) i (3) konieczna jest dalsza analiza w celu wykrycia konkretnych zróżnicowań, nie wystarczy bowiem stwierdzenie, że dane w tabeli wykazują znamienne zróżnicowanie, bo nie wiadomo które – czy wszystkie ze wszystkimi? To może nie być (i najczęściej nie jest) prawdą.

Wracając do omawianego przykładu, nie stwierdzono zróżnicowania odpowiedzi w zależności od miejsca zamieszkania studentów (G = 0,39 < 3,84), natomiast w wypadku studentek zależność taka jest znamienna (G = 6,41 > 5,99). Należy zatem przystąpić do dalszej, szczegółowej analizy, aby stwierdzić, jaki jest związek odpowiedzi z miejscem zamieszkania studentek.

Analiza polega na porównywaniu ze sobą poszczególnych grup (*a/b, a/c, b/c*) bądź też na kolejnym porównaniu danych z kolumn *a, b* i *c* z sumą pozostałych kolumn.

Porównywanie poszczególnych grup ma miejsce wówczas, gdy badania są typu przekrojowego, a więc grupy można uważać za niezależne. Jeżeli natomiast badania są powtarzane w różnym czasie i chce się ocenić ich dynamikę, bądź gdy dotyczy to wyborów wielokrotnych, właściwy będzie drugi sposób. Zostanie on omówiony w następnej sekcji.

W tabeli 5.4 powtórzono dane dla studentek z tabeli 5.2 oraz dane do porównania kolumny *a* z *b* (środkowa część tabeli). Wartość funkcji G dla tego porównania (df = 1) wynosi 3,66, jest więc bliska znamienności. Wartości funkcji G dla pozostałych porównań (*a/c* i *b/c*) wynoszą odpowiednio 6,07* i 0,06 (gwiazdka * oznacza poziom znamienności statystycznej; zob. tablica funkcji χ^2 na końcu książki). Ponieważ nie ma różnicy *b/c*, dane z tych kolumn można połączyć (prawa strona tabeli). Różnica między wariantem *a* oraz *b+c* łącznie jest znamienna (G = 6,35*). Można wyciągnąć wniosek, że studentki obu lat studiów, mieszkanki wsi, znamiennie częściej niż mieszkanki miast uczęszczają na zajęcia WF; odpowiednie częstości odsetkowe wynoszą 118/125 = 94,4% i 261/302 = 86,4%. Żadnych różnic zależnych od miejsc zamieszkania nie stwierdzono u studentów (G = 0,39; tab. 5.2), można zatem przyjąć jedną wartość odsetkową (360/395 = 91,1%) dla wszystkich studentów.

Tab. 5.4. Zestawienie częstości odpowiedzi na pytanie nr 1 według miejsca zamieszkania studentek (powtórzenie tab. 5.2) i sprawdzenie różnicy między wariantem *a* i *b*, oraz między *a* i sumą pozostałych (*b* + *c*)

m.zam.	a	b	c	Σ	a	b	Σ	a	b+c	Σ
Nie (0)	7	13	28	48	7	13	20	7	41	48
Tak (1)	118	88	173	379	118	88	206	118	261	379
Σ	125	101	201	427	125	101	226	125	302	427
			G	6,41*		G	3,66		G	6,35*

Jak pokazano w tabeli 5.3, nie ma różnicy między studentkami i studentami w globalnych częstościach odpowiedzi. Zależne od miejsca zamieszkania różnice wśród studentek każą jednak spojrzeć na sprawę nieco inaczej, bowiem średni odsetek studentów uczęszczających na zajęcia WF plasuje się między studentkami – mieszkankami wsi a mieszkankami miast. Można by porównać studentów – mieszkańców wsi z taką samą grupą studentek oraz mieszkańców i mieszkanki miast, jednak bardzo duża jednorodność studentów (G = 0,39 !) pozwala na porównanie wszystkich studentów łącznie najpierw z mieszkankami wsi, a potem z mieszkankami miast (tab. 5.5).

Tab. 5.5. Porównanie studentów ze studentkami pod względem częstości uczestnictwa w zajęciach WF

Pytanie nr 1	Studenci	Studentki	
	a + b + c	a	b + c
Nie (0)	7	13	28
Tak (1)	118	88	173
Σ	125	101	201
% (1)	91,1	94,4	86,4
G		1,47	3,88*

Zanalizujemy jeszcze wybory rozłączne na innym przykładzie przedstawionym w tabeli 5.6. W pewnej gminie przeprowadzono dwukrotnie sondaż popularności trzech kandydatów na wójta. Aby stwierdzić, czy wystąpiły znamienne zmiany popularności w tych dwóch okresach, należy porównywać dane dla poszczególnych kandydatów względem sum dla pozostałych kandydatów (A względem B + C itd.). Wyniki analizy są następujące: wartość G dla całej tabeli (df = 2) wynosi 5.17, nie świadczy więc o znamienności (5,17<5,99), ale większa niż 3,84, ma więc tutaj zastosowanie sytuacja (2) opisana na stronie 36. Wartości funkcji G dla poszczególnych kandydatów wynoszą 2,94, 5,07* i 0,34. Wynika z tego, że kandydat B znamiennie stracił na popularności, natomiast w wypadku kandydata A wystąpiła **tendencja** wzrostowa: wartość G = 2,94 jest mniejsza niż 3,84, ale większa niż 2,71 (df = 1; p<0.10). Można sądzić, że gdyby w sondażach brała udział większa liczba osób, wzrost popularności kandydata A okazałby się statystycznie znamienny.

Kandydat	maj	wrzesień	Σ
A	35	44	79
B	43	26	69
C	18	20	38
Σ	96	90	186

Tab. 5.6. Wyniki dwukrotnego sondażu popularności trzech kandydatów na wójta

Podane przykłady ilustrują sposoby przeprowadzenia pełnej analizy. Najpierw oblicza się wartość funkcji G dla całej tabeli i ocenia wynik: jeżeli mamy sytuację (1), nie prowadzi

się dalszej analizy, jeżeli natomiast mamy sytuację (2) lub (3), redukuje się tabelę do mniejszych tabelek przez zestawianie ze sobą różnych elementów tabeli lub przez komasowanie kolejnych kolumn lub wierszy.

Możliwość dowolnego zestawiania danych – np. według miejsca zamieszkania w obrębie tej samej płci jest istotną cechą analizy wyborów rozłącznych. Pełna analiza jest w każdym razie konieczna, gdyż poprzestanie tylko na obliczeniu funkcji G dla całej tabeli o wymiarach większych niż 2×2 (df = 1) może dać jedynie odpowiedź, że dane w tabeli są znamiennie zróżnicowane. Jest to niewystarczająca informacja, bowiem trzeba wskazać, które konkretnie elementy są znamiennie zróżnicowane.

Analiza wyborów wielokrotnych

Charakterystyczną cechą wyborów wielokrotnych jest to, że sumy wszystkich wyborów są większe niż liczba ankiet. Podstawą odniesienia jest liczba ankiet zawierających odpowiedzi na dane pytanie (N), dlatego też analiza zestawienia musi być prowadzona inaczej niż w wypadku odpowiedzi dychotomicznych lub z jednym wyborem. Zasadą jest rozpatrywanie każdej opcji odpowiedzi oddzielnie w układzie dychotomicznym:

(1) - liczba ankiet zawierających wybór danej opcji (n),

(0) - liczba ankiet nie zawierających tego wyboru (N – n).

W tabeli 5.7 zestawiono częstości wyboru dwóch opcji odpowiedzi na pytanie nr 5 (zob. s. 24). W wierszach (0) podano liczby ankiet, w których nie wybrano danej opcji: jest to np. wartość 29 = 74 – 45. Analizę tej opcji („a") przeprowadza się analogicznie do pytania nr 1, a więc porównując ze sobą lata studiów dla każdego wariantu miejsca zamieszkania.

Tab. 5.7. Zestawienie częstości wyborów opcji *a* i *d* (pytanie nr 5) według miejsca zamieszkania i płci dla I i II roku studiów

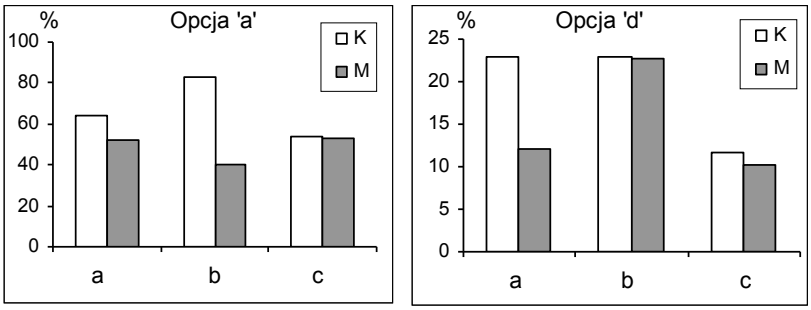
płeć	**K**	Pytanie 5							**M**						
m.zam.	a		b		c		Σ	a		b		c		Σ
rok st.	I	II	I	II	I	II		I	II	I	II	I	II	
BO	0	0	3	1	1	0	5	1	0	0	1	1	1	4
N	74	52	44	56	89	101	416	45	38	57	62	82	112	396
a (0)	29	16	6	11	39	49	150	22	18	36	35	39	52	202
a (1)	45	36	38	45	50	52	266	23	20	21	27	43	60	194
% (1)	61	69	81	80	57	50	63	50	53	38	44	52	53	49
d (0)	57	40	33	44	80	88	342	40	33	46	46	73	101	339
d (1)	17	12	11	12	9	13	74	5	5	11	16	9	11	57
% (1)	23	23	23	21	10	13	18	11	13	20	26	11	10	14

W żadnym wypadku nie stwierdzono znamiennej różnicy między latami studiów (wartości G dla wszystkich porównań są <1), można zatem porównać poszczególne warianty miejsca zamieszkania dla obu lat studiów łącznie, zarówno dla każdej płci oddzielnie, jak i ew. między płciami (porównanie np. wariantu *a* u studentek i studentów).

W tabeli 5.8 zestawiono te same dane, ale dla obu lat studiów łącznie, natomiast na rycinie 5.5 pokazano odsetkowe wartości wyborów. Na podstawie wykresu można się zorientować, które dane warto poddać analizie, a których nie. Dotyczy to zarówno porównań między wariantami miejsca zamieszkania w obrębie płci, jak też porównań między płciami dla poszczególnych miejsc zamieszkania.

Tab. 5.8. Zestawienie częstości wyborów opcji „a" i „d" (pytanie nr 5) według miejsca zamieszkania (a, b, c) i płci (K, M) dla obu lat studiów łącznie

płeć	**K**				**M**			
m.zam.	*a*	*b*	*c*	Σ	*a*	*b*	*c*	Σ
N	126	100	190	416	83	119	194	396
a (0)	45	17	88	150	40	71	91	202
a (1)	81	83	102	266	43	48	103	194
% (1)	64	83	54	63,9	52	40	53	49,0
d (0)	97	77	168	342	73	92	174	339
d (1)	29	23	22	74	10	27	20	57
% (1)	23	23	12	17,8	12	23	10	14,4



Ryc. 5.5. Odsetkowe częstości wyborów odpowiedzi na pytanie nr 5 (opcje „a" i „d") dla obu lat studiów łącznie w zależności od miejsca zamieszkania studentek (K) i studentów (M)

Wyniki analizy są następujące: opcja odpowiedzi „a", studentki – dla całej tabeli (df = 2) G = 26,1***; częstość wyboru dla miejsca zamieszkania *b* jest znamiennie większa od *a* (G = 10,14**) i od *c* (G = 26,1***), natomiast różnica *a/c* nie jest znamienna (G = 3,52). Studenci – dla całej tabeli G = 5,16 (<5,99; nieznamienne, ale >3,84); miejsca zamieszkania

a i *c* nie różnią się (G = 0,04), wariant *b* zaś jest mniejszy od *a*+*c* łącznie (G = 5,12*). Opcja odpowiedzi „d": studentki – dla całej tabeli G = 9,51*; miejsca zamieszkania *a* i *b* nie różnią się (G < 0,01), wariant *c* jest znamiennie mniejszy od *a*+*b* łącznie (G = 9,51**). Studenci: dla całej tabeli G = 9,08*; warianty *a* i *c* nie różnią się (G = 0,18), wariant *b* jest od nich obu łącznie znamiennie mniejszy (G = 8,90**). Na koniec można zbadać różnice między studentkami i studentami dla wariantów *a* i *b* (opcja odpowiedzi „a") oraz dla wariantu *a* w opcji odpowiedzi „d"; odpowiednie wartości G wynoszą 3,22, 43,4*** i 4,15*.

Aby uzupełnić ocenę odpowiedzi na pytanie nr 5, należy jeszcze zanalizować częstości wyboru jednej opcji, dwóch opcji itd. przedstawione na rycinie 5.2 B. Jak wynika z wykresu, można oczekiwać znamiennych różnic między studentkami I i II roku (mieszkanki wsi) w wyborze dwóch lub trzech opcji odpowiedzi. Obliczenia należy przeprowadzić jak w przykładzie analizy popularności (s. 36), tzn. oddzielnie dla 2 i 3 wyborów. Odpowiednie wartości G wynoszą 6,47* i 5,09*, a więc studentki II roku, mieszkanki wsi, znamiennie rzadziej niż studentki I roku wybierały dwie opcje odpowiedzi, a znamiennie częściej trzy. Analogicznie należy zbadać liczebności wyborów studentek z pozostałych miejsc zamieszkania, a następnie studentów (nie pokazano tu odpowiednich danych).

Te dwa warianty odpowiedzi („a" i „d") wybrano tylko jako przykłady; w identyczny sposób analizuje się pozostałe warianty pytania nr 5.

Analiza danych uporządkowanych (rangowanych)

Nieco inaczej przygotowuje się do analizy dane uporządkowane, jak w pytaniu nr 6. Jako przykład podano w tabeli 5.9 część danych z ryciny 4.9 odnoszącą się do studentek I roku, mieszkanek wsi, jak również odpowiadające liczebnościom punkty, a w tabeli 5.10 – podstawowe parametry obliczone z tych danych.

Tab. 5.9. Częstości wyboru kolejności poszczególnych opcji odpowiedzi (a – f) na pytanie nr 6 (por. ryc. 4.9); studentki I roku, mieszkanki wsi.

Rangi	6	5	4	3	2	1	Iloczyny liczebności i rang						Σ	N - Σ	SR
Kolejność	*1*	*2*	*3*	*4*	*5*	*6*	*1*	*2*	*3*	*4*	*5*	*6*			
a	8	4	7	19	24	10	48	20	28	57	48	10	211	221	2.93
b	3	3	7	26	8	25	18	15	28	78	16	25	180	252	2.50
c	31	23	14	3	1	0	186	115	56	9	2	0	368	64	5.11
d	12	15	24	11	8	2	72	75	96	33	16	2	294	138	4.08
e	17	27	20	3	5	0	102	135	80	9	10	0	336	96	4.67
f	1	0	0	10	26	35	6	0	0	30	52	35	123	309	1.71

Objaśnienia: n = 72; N = 72 · 6 = 432; SR (średnia ranga) = Σ/n

Suma każdego wiersza zestawienia (n) w lewej stronie tabeli 5.9 dla tej grupy wynosi 72, natomiast maksymalna suma punktów możliwa do osiągnięcia (N) wynosi 432 (gdyby we wszystkich ankietach badanej grupy dana opcja wyboru została wymieniona na pierwszym miejscu). Te dwie wartości (n i N) są stałe dla wszystkich opcji odpowiedzi w danej grupie.

Aby można było ustalić przeciętne kolejności opcji odpowiedzi, należy kolejnościom przypisać punkty rangowe: najwięcej pierwszej, a ostatniej jeden. Opcji odpowiedzi jest 6, więc maksymalna ranga wyniesie 6. W prawej części tabeli 5.9 podano punkty będące iloczynami rang i częstości podanych w lewej części tabeli oraz sumy punktów (Σ) dla poszczególnych wierszy. Sumy te podzielone przez wartość n dają w wyniku średnią rangę (pozycję) każdej opcji odpowiedzi. Różnice między maksymalną liczbą punktów a sumami punktów (Σ) odpowiadają liczbie „punktów niezdobytych"; są one konieczne do dalszej analizy. Takie zestawienia i obliczenia należy sporządzić dla wszystkich badanych grup.

Podobnie jak w analizie poprzednich pytań należy zacząć od zbadania ewentualnych różnic między latami studiów, stosując test chi-kwadrat. Jako przykład pokazano dane dla studentek I i II roku, mieszkanek wsi. Dane odnoszą się do częstości kolejności wyboru odpowiedzi „a" (spacery).

Tab. 5.10. Zestawienie częstości kolejności wyboru odpowiedzi „a" na pytanie nr 6; studentki I i II roku, mieszkanki wsi. G = 2,01.

Rok	n	SR	Σ	N - Σ	N
I	72	2,93	211	221	432
II	52	2,65	138	174	312
		Σ	349	395	744

Otrzymana wartość G jest mniejsza niż 3,84, a więc różnica między I i II rokiem jest nieznamienna. W taki sam sposób należy sprawdzić różnice między I i II rokiem studiów dla poszczególnych opcji odpowiedzi i wszystkich wariantów miejsca zamieszkania. Jeżeli okaże się, że nie ma różnic, można będzie obliczyć odpowiednie sumy dla obu lat i dalszą analizę prowadzić dla miejsc zamieszkania i płci, podobnie jak w poprzednich pytaniach. Należy jednak zwrócić uwagę, że podstawą obliczeń nie są liczby ankiet, a liczby punktów rangowych, a więc zmienna, którą można analizować jak zmienną ciągłą (patrz niżej).

Analiza danych ilościowych

W badaniach ankietowych spotykamy najczęściej trzy rodzaje danych ilościowych:

- wartości rangowe, w tym omówione wyżej punkty rangowe,
- wartości punktowe odpowiadające stopniowanym ocenom werbalnym w skali porządkowej (s. 15),
- wartości wyrażone w ciągłej skali analogowej (s. 16).

W analizie wartości rangowych stosuje się popularny test *t* Studenta oraz analizę wariancji, co zostanie pokazane dla danych z poprzedniego przykładu

Tab. 5.11. Zestawienie danych dla studentek I i II roku, mieszkanek wsi (pytanie nr 6, opcja „a") oraz obliczone z tych danych proporcje (p) i ich wariancje (s^2)

Rok	Σ	n	N	p	s^2
I	211	72	432	0,488	0,250
II	138	52	312	0,442	0,247
Σ	349	124	744	0,469	0,249

Do obliczenia wartości testu *t* potrzebne są następujące wielkości:
Proporcje (np. 211/432 = 0,488) i wariancje (s^2; np. $0,488 \cdot (1 - 0,488) = 0,250$
Różnica między proporcjami $\Delta = 0,488 - 0,442 = 0,046$
Proporcja średnia (średnia ranga) $p = 349/744 = 0,469$
Wariancja średnia $s^2 = 0,469 \cdot (1 - 0,469) = 0,249$ (df = 744 – 2 = 742)
Błąd różnicy $s_\Delta = \sqrt{0,249\left(\frac{1}{432} + \frac{1}{312}\right)} = 0,0371$
$t = 0,046/0,0371 = 1,24$

Otrzymana wartość *t* jest mniejsza od tablicowej (zob. s. 64) równej 1,96, a więc różnica między I i II rokiem jest nieznamienna. W taki sam sposób należy sprawdzić różnice między I i II rokiem studiów dla poszczególnych opcji odpowiedzi i wszystkich wariantów miejsca zamieszkania. Jeżeli okaże się, że nie ma różnic, można będzie obliczyć średnie wartości dla obu lat i dalszą analizę prowadzić dla miejsc zamieszkania i płci, podobnie jak w poprzednich pytaniach.

Do zbadania zróżnicowań między miejscami zamieszkania w obrębie płci nie można użyć pokazanego tu testu *t* Studenta bez uprzedniej analizy wariancji, bo mamy więcej niż dwa warianty miejsca zamieszkania. Przykład analizy wariancji pokazano w tabeli 5.12.

Wynik analizy świadczy o tym, że między trzema badanymi proporcjami występuje znamienne zróżnicowanie. W podanym przykładzie średnia proporcja odpowiedzi (opcja *a*) jest tym większa, im wyższa jest kategoria miejsca zamieszkania. Można poprzestać na tym stwierdzeniu, ale wobec stosunkowo niedużej różnicy między kategoriami *a* i *b* może być interesujące zbadanie, czy różnica między nimi jest znamienna.

Ocenę tej różnicy przeprowadza się testem *t,* jak opisano powyżej, z tym, że jako średnią wariancję należy wziąć wartość 0,293 z analizy wariancji (tab. 5.12). Testu *t* można użyć w taki sposób tylko do oceny różnic między tymi średnimi, które sąsiadują ze sobą przy uszeregowaniu od najmniejszej do największej, a więc między *b* i *a* oraz *c* i *b*, natomiast nie między średnimi *c* i *a*. W tym ostatnim wypadku należy użyć tzw. testów wielokrotnego rozstępu, np. testu Scheffé'go. Zainteresowanych odsyłam do „Biometrii".

Wartość funkcji *t* dla różnicy *b* – *a* wynosi 1,60, jest więc nieznamienna.

Tab. 5.12. Analiza wariancji dla miejsc zamieszkania (*a, b, c*) studentek obu lat studiów łącznie. Pytanie nr 6, opcja odpowiedzi „a".

	Σ	N - 1	N	p	s^2	S_{xx}	"grupy"
a	349	743	744	0,469	0,249	185,04	163,71
b	313	605	606	0,517	0,250	151,09	161,67
c	703	1145	1146	0,613	0,237	271,52	431,25
Σ	1365	2493	2496	0,547		607,64	756,62

Analiza wariancji

poprawka	746,48	S_{xx}	df	s^2	F
grupy	756,62	10,14	2	5,069	17,33***
reszta		607,64	2077	0,293	

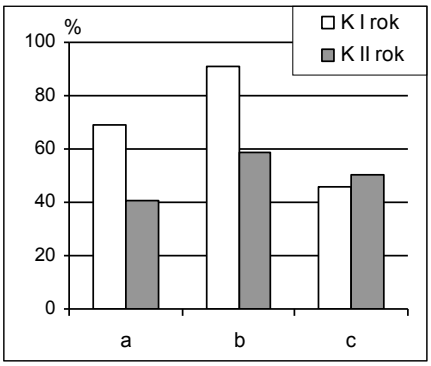
Objaśnienia: $S_{xx} = s^2 \cdot (N - 1)$; „grupy" = Σ^2/N; „poprawka" = $1365^2/2496$; s^2 (w dolnej tabelce – „Analiza wariancji") = S_{xx}/df; F = 5,069/0,293.

Pozostaje jeszcze interpretacja wyniku analizy; wydawałoby się, że należy badać różnice między poszczególnymi kategoriami zmiennej (tu – miejsca zamieszkania). Wyobraźmy sobie jednak, że kategorii takich jest więcej, np. do 10 000 (wieś), 10 – 50 tys., 50 – 100 tys., 100 – 500 tys., 500 tys. – milion, powyżej miliona), wówczas żadne dwie kolejne kategorie mogą się nie różnić. Kategorie te tworzą ponadto ciąg ilościowy i każde inne ich uszeregowanie będzie nielogiczne. Jeżeli badana cecha (tu – wybór kolejności opcji *a* pytania nr 6) będzie miała tendencję do jednokierunkowych zmian (tu – wzrostu) ze wzrostem kategorii miejsca zamieszkania, to **można** poprzestać na takim stwierdzeniu i nie prowadzić bardziej szczegółowej analizy.

Rozpatrzmy jeszcze analizę częstości wyboru odpowiedzi „e" na pytanie nr 5, bo tu wyniki układają się odmiennie niż w poprzednich przykładach. W tabeli 5.13 pokazano zestawienie analityczne dla studentek, a na rycinie 5.6 odsetkowe częstości wyborów tej odpowiedzi („wakacje nad jeziorem").

Tab. 5.13. Zestawienie częstości wyborów opcji „e" (pytanie nr 5) według miejsca zamieszkania studentek I i II roku

m.zam.	a		b		c	
rok	I	II	I	II	I	II
N	74	52	44	56	89	101
e (0)	23	31	4	23	58	50
e (1)	51	21	40	33	41	51
% (1)	69	40	91	59	46	50



Ryc. 5.6. Odsetkowe częstości wyborów opcji „e" (pytanie nr 5) według miejsca zamieszkania (a, b, c) studentek I i II roku

Można zauważyć, że:

1. studentki I roku, mieszkanki wsi (*a*) i miast do 100 000 mieszkańców (*b*) częściej wybierały tę odpowiedź niż studentki II roku,
2. brak takiej różnicy u mieszkanek dużych miast (*c*),
3. studentki I roku ze średnich miast (*b*) częściej wybierały tę odpowiedź niż studentki I roku, mieszkanki wsi (*a*),
4. częstości wyboru tej odpowiedzi przez studentki II roku, mieszkanek wsi i średnich miast (*a, b*) i studentki z dużych miast (*c*), obu lat studiów, wydają się podobne.

Przedstawiona poniżej analiza wyników tego ostatniego przypadku odpowie na pytanie, czy przypuszczenie to może być uznane za uzasadnione.

Tab. 5.14. Analiza wariancji opcji odpowiedzi „e" pytania nr 5 dla studentek II roku miejsc zamieszkania *a, b* i *c* oraz studentek I roku *c* (przypuszczenie 4 powyżej)

	Σ	N - 1	N	p	s^2	S_{xx}	"grupy"
a, II	21	51	52	0,404	0,241	12,28	8,48
b, II	33	55	56	0,589	0,242	13,31	19,45
c, I	41	88	89	0,461	0,248	21,86	18,89
c, II	51	100	101	0,505	0,250	25,00	25,75
Σ	146	294	298	0,490		72,45	72,57

Analiza wariancji

poprawka	71,53	S_{xx}	df	s^2	F
grupy	72,57	1,04	3	0,347	1,41
reszta		72,45	294	0,246	

Zróżnicowanie czterech badanych grup okazało się nieznamienne ($F = 1,41 < 2,60$; por. tablica funkcji F, s. 65), zatem nie ma podstaw do odrzucenia przypuszczenia 4.

W znany już sposób można ocenić różnice między pierwszymi latami studiów miejscowości *a* i *b* (G = 8,40**; przypuszczenie 3), między I i II rokiem dla kategorii miejscowości *a* (G = 10,2**) i *b* (G = 14,01***; przypuszczenie 1) oraz między I i II rokiem w kategorii *c* (G = 0,37; przypuszczenie 2). Interpretacja tych wyników zostanie przedstawiona w rozdziale 6.

Pozostają jeszcze do omówienia inne rodzaje odpowiedzi ilościowych. Pytania wymagające stopniowanych odpowiedzi w skali porządkowej (s. 15) można analizować jak zmienną ilościową, jednak pod warunkiem, że postulat równości odstępów między kategoriami odpowiedzi można uznać za spełniony (zob. s. 15). W przeciwnym wypadku można analizować po prostu częstości poszczególnych odpowiedzi (zob. Analiza wyborów rozłącznych) bądź przypisać poszczególnym kategoriom odpowiednie wartości liczbowe, inne niż podano na stronie 14, np.:

Czy lubisz uprawiać sport?

Zdecydowanie nie	Trochę	Średnio	Bardzo	Zdecydowanie tak (s. 15)
0	2	3	5	6

Przypisanie kategoriom odpowiedzi takich wartości punktowych jest oczywiście subiektywne i wymaga uzasadnienia.

Zastąpienie skali porządkowej skalą analogową (s. 16) pozwala traktować dane odczytane ze skali jak pomiary i analizować je za pomocą testów parametrycznych (analiza wariancji, test *t*, i inne). Mogą tu wystąpić pewne trudności związane z tym, że proponowana skala analogowa jest zamknięta, tzn. wartości zawarte są w przedziale 0 – 10. Rozkłady takich wartości można przyjąć za normalne, jeżeli mieszczą się przedziale wartości od ok. 3 do ok. 7. Poza tym przedziałem rozkłady będą skośne i konieczne jest stosowanie odpowiedniego przekształcenia zmiennej. Zainteresowanych odsyłam do „Biometrii", gdzie opisano zarówno przekształcanie zmiennej, jak i wykonanie różnych rodzajów analizy wariancji. Przykładem zastosowania skali ciągłej jest odpowiedź na pytanie nr 3 (zob. s. 24) wyrażona w zestawieniu zbiorczym (ryc. 4.3) przez średnią i odchylenie standardowe. Taki sam sposób należy zastosować do wielokrotnych wyborów z punktowanymi odpowiedziami (zob. s. 16 i tab. 2.2).

Zależności między badanymi zmiennymi

Mówiąc o zależnościach (czy też raczej współzależnościach) między zmiennymi ma się najczęściej na myśli zmienne ilościowe ciągłe (pomiary) i zachodzące między nimi korelacje. Pojęcie współzależności odnosi się jednak również do zmiennych dyskretnych (liczebności), a stosowany powszechnie do analizy tych zmiennych test chi-kwadrat jest często nazywany testem niezależności. Biorąc za przykład dane z tabeli 5.1 widać, że mamy tu do czynienia z czterema zmiennymi: płcią (K, M), rokiem studiów (I, II), miejscem zamieszkania (*a, b, c*) i odpowiedzią (0, 1 lub nie, tak). W dalszych tabelach częstości zestawiane są

w różnych kombinacjach zmiennej, np. w tabeli 5.2 zestawiono kombinacje odpowiedzi i miejsca zamieszkania studentek.

Obliczona dla tego przykładu wartość funkcji χ^2 (w postaci funkcji G) ma interesującą właściwość: pierwiastek z wartości G podzielonej przez całkowitą liczebność analizowanej tabeli jest równoważny współczynnikowi korelacji, a ściślej mówiąc – tzw. stosunkowi korelacyjnemu. Na przykład, w tabeli 5.4 przedstawiono zależność między rodzajem odpowiedzi (tak – nie), a miejscem zamieszkania studentek. Wartość G wyniosła 6,41, a całkowita liczebność – 427. Podstawiając te dane do powyższego wzoru otrzymamy r = 0,123. Inaczej mówiąc, korelacja między miejscem zamieszkania a rodzajem odpowiedzi wyniosła dla studentek 0,123, przy czym znamienność tego współczynnika korelacji jest taka sama jak dla wartości funkcji G ($p<0,05$).

$$\sqrt{\frac{G}{N}} = r$$

Należy zwrócić uwagę, że we wszystkich rozpatrywanych dotąd przykładach odpowiedzi były zestawiane ze zmiennymi zawartymi w metryczce (płeć, rok studiów, miejsce zamieszkania). Podobnie można jednak zestawiać ze sobą („korelować") częstości odpowiedzi na **różne** pytania, np. nr 1 i 2 z przykładu podanego na s. 23. Zestawienie analityczne pokazano w tabeli 5.13. Należy podkreślić, że korelowanie częstości odpowiedzi na różne pytania jest zasadne tylko wtedy, gdy ma to sens i wiąże się z badanym zagadnieniem.

Otrzymany wynik (r = 0,495***) świadczy o wysokiej zależności między stosunkiem do zajęć WF w szkole średniej a uczęszczaniem na zajęcia WF na uczelni – kierunek korelacji jest tu dodatni. Oczywiście, żeby sporządzić to zestawienie, należy w zestawieniach zbiorczych dla poszczególnych miejsc zamieszkania połączyć dane studentek I i II roku i posortować według odpowiedzi na pytanie nr 1, a następnie nr 2 (por. ryc. 4.3, s. 25).

pyt. 1 / pyt. 2	0	1	Σ
–1	6	4	10
0	1	51	52
1	0	63	63
Σ	7	118	125

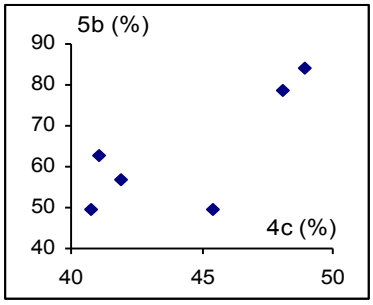
Tab. 5.13. Zestawienie częstości odpowiedzi na pytania nr 1 i 2 (studentki I i II roku łącznie, mieszkanki wsi)
G = 30,61***; r = 0,495

Podobnie postępuje się z odpowiedziami ilościowymi stopniowanymi (w skali porządkowej). Jeżeli odległości między kategoriami są podobne, wówczas korelacje między takimi zmiennymi można obliczać parametrycznie, a więc traktując wartości punktowe przypisane kategoriom odpowiedzi jak zmienną ciągłą. To samo dotyczy oczywiście danych wyrażonych w skali analogowej, uwzględniając wspomniane wyżej zastrzeżenie dotyczące rozkładów takich zmiennych. Współczynnik korelacji między wartościami zmiennych ciągłych należy oczywiście obliczać z danych zawartych w indywidualnych ankietach.

Można również zestawić na wykresie punktowym (tzw. opcja XY w Excelu) średnie wartości odsetkowe odpowiedzi w badanych grupach, co ułatwia zorientowanie się, czy istnieje jakaś współzależność, jak pokazano na rycinie 5.7, na której zestawiono częstości wyboru opcji „c" odpowiedzi na pytanie nr 4 (wakacje z partnerem) z częstościami opcji „b" odpowiedzi na pytanie nr 5 (wakacje nad morzem). Ze średnich danych nie oblicza się współczynnika korelacji; aby obliczyć taki współczynnik należałoby segregować indywidualne dane w arkuszach poszczególnych grup według odpowiedzi na pytanie nr 4 (pierwszy wybór; por. ryc. 4.2) i pytanie 5 (drugi wybór) i zestawić następującą tabelę dla poszczególnych grup:

Tab. 5.14. Zestawienie łącznych wyborów odpowiedzi 4„c" i 5„b". Studentki, mieszkanki wsi, oba lata studiów łącznie.

4 c / 5 b	Wybrane	Niewybrane	Σ
Wybrane	**51**	32	83
Nie wybr.	5	**38**	43
Σ	56	70	126



Ryc. 5.7. Zależność między średnimi proporcjami (*p*) odpowiedzi na pytanie 4„c" i 5„b" we wszystkich grupach studentek

Wartość funkcji G dla tej tabeli wynosi 31,5, zatem współczynnik korelacji jest równy 0,500. Jak wynika z układu tabeli kierunek korelacji jest dodatni. Tabelę taką można ułożyć dla wszystkich studentek łącznie. Porównanie współczynników korelacji z różnych grup nie będzie tu omawiane; zainteresowanych odsyłam do „Biometrii".

6. PREZENTACJA WYNIKÓW

Dobrze zaplanowane i właściwie przeprowadzone badania, których wyniki wyczerpująco zestawiono i zanalizowano, powinny jeszcze zostać odpowiednio zaprezentowane. Wyniki prezentowane są tabelarycznie i/lub graficznie, a ich omówienie kończy się wnioskami. Poniżej zostaną omówione podstawowe sposoby graficznego przedstawienia wyników oraz zasady opisu badań techniką ankietową. Zasady te stosują się zarówno do raportów, rozpraw (magisterskich, doktorskich), prac przeznaczonych do publikacji, jak i danych prezentowanych na konferencjach. Klasyczny układ wspomnianych wyżej rodzajów prezentacji obejmuje następujące części: wstęp-wprowadzenie (zawierający przegląd piśmiennictwa), cel pracy oraz hipotezy i pytania badawcze, materiał i metody, wyniki, dyskusja, wnioski, piśmiennictwo i ew. aneks. Poniżej zostaną omówione tylko niektóre z nich.

Formy prezentacji

Najważniejsze formy prezentacji graficznej to wykres kolumnowy („słupki"), kołowy („pomarańcza"), liniowy („wykres") i punktowy („korelogram"). Przedstawiane na wykresach odsetki należy odnosić do liczby odpowiedzi na dane pytanie, a nie do liczby ankiet, a w tekście lub w opisie ilustracji należy podać liczbę odpowiedzi lub odsetek braku odpowiedzi.

Wykresy kolumnowe. Programy graficzne (np. Excel) oferują różne formy wykresów kolumnowych. Najlepiej stosować najprostsze, płaskie, bo łatwo na nich umieścić dodatkowe informacje, np. słupki błędu. Kolumny trójwymiarowe (perspektywiczne, przestrzenne) pogarszają czytelność obrazu. Wykresy kolumnowe są wskazane wówczas, gdy trzeba pokazać kilka kategorii i grup jednocześnie, jak na przykład na rycinie 6.2. Dalej pokazano sposób tworzenia i formatowania wykresu kolumnowego.

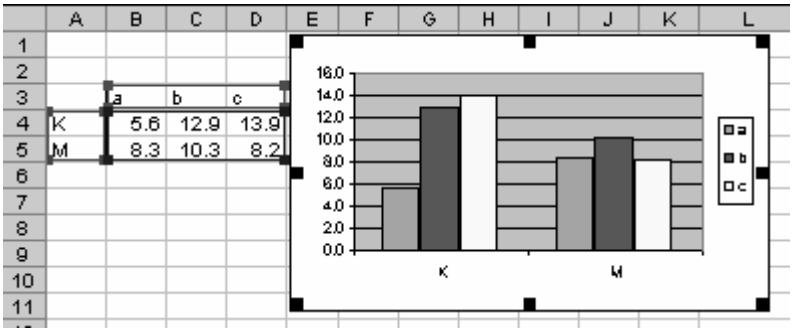
Wykresy kołowe. Wykresy kołowe płaskie lub przestrzenne nadają się do przedstawienia procentowych udziałów różnych kategorii w jednej grupie danych (suma udziałów stanowi 100%). Przykład wykresu kołowego – rycina 6.3.

Wykresy liniowe. Służą do tworzenia profilów grupowych dynamicznych (zmiany w czasie) lub statycznych. Mogą zastępować wykresy kolumnowe (ryc. 6.4 i 6.5), dzięki czemu uzyskuje się nieraz większą przejrzystość, zwłaszcza w ocenie różnic.

Wykresy punktowe. Wykresy punktowe (XY) służą przede wszystkim do prezentacji współzależności między zmiennymi; dotyczy to nie tylko wartości indywidualnych, ale także średnich grupowych. W omawianych tu przykładach tylko jedna zmienna jest wyrażona w skali ciągłej (pytanie nr 3), więc indywidualnych danych nie ma z czym korelować. Przykładem korelogramu średnich wyborów odpowiedzi na pytania nr 4 i 5 jest rycina 5.7.

Inne formy prezentacji to zestawienia tabelaryczne stosowane przede wszystkim w tekście, a także prezentacje multimedialne; te ostatnie nie będą tu omawiane. Zarówno wykresy, jak i tabele muszą być „samowyjaśniające się", tzn. winny zawierać wszelkie istotne informacje bez potrzeby odwoływania się do tekstu. Jeżeli pewien typ wykresu lub tabeli powtarza się, a zmienia się tylko ich zawartość, można nie powtarzać objaśnień np. symboli za każdym razem, tylko umieścić pod ilustracją adnotację „Objaśnienia oznaczeń jak do ryciny XX" lub podobną. Taką adnotację umieszczono przy rycinie 6.4, odsyłając do liczebności poszczególnych grup podanych w tabeli 5.4, bowiem umieszczenie tych informacji na rycinie lub w jej opisie byłoby bardzo kłopotliwe.

Poniżej przedstawiono technikę sporządzania wykresu w Excelu, jego sformatowanie, uzupełnienie objaśnieniami i przeniesienie do dokumentu (pliku Worda).



Ryc. 6.1. Tworzenie wykresu w arkuszu Excel – dane liczbowe i surowy wykres (aktywny); po zaznaczeniu obszaru kliknąć w zakładkę *Wstawianie* i wybrać typ wykresu (*Kolumnowy*, *2W*). W starszych wersjach Excela po zaznaczeniu obszaru kliknąć w ikonkę *Kreator wykresów*, wybrać *Dalej* i wybrać *Kolumny* i *Zakończ*.

W celu sformatowania wykresu do postaci przedstawionej na rycinie 6.1 należy wykonać podane poniżej czynności (opcjonalne).

- Przenieść ramkę legendy: kliknąć w tło, złapać myszką środkowy uchwyt ramki tła i przesunąć w prawo do prawej krawędzi ramki legendy; kliknąć wewnątrz ramki legendy, trzymając naciśnięty klawisz myszki przesunąć ramkę w żądane miejsce;
- Zlikwidować szare tło: kliknąć w tło i nacisnąć DELETE;
- Zlikwidować poziome linie: kliknąć w jedną z linii i nacisnąć DELETE;
- Zmiana proporcji wykresu: uaktywnić wykres (kliknąć koło ramko wewnątrz wykresu), złapać myszką prawy środkowy uchwyt (kursor zmieni się na poziomą podwójną strzałkę) i przesunąć w lewo do żądanego formatu; UWAGA: zmienia się wielkość czcionki! mając stale aktywny wykres, wybrać żądaną wielkość czcionki (np. 10);
- W opisie osi pionowej zlikwidować miejsca dziesiętne i zmienić skalowanie: kliknąć w oś (na jej końcach pojawią się „uchwyty"), wybrać polecenie Format – Zaznaczona oś –

Liczby – Miejsca dziesiętne (ustawić na zero); przejść do zakładki Skala, w okienku Maksimum wpisać 15, a w okienku Jednostka główna wpisać 5, potem OK;

- Zmienić kolory słupków: kliknąć w słupek (zostaną zaznaczone wszystkie słupki tego koloru, czyli cała seria), wybrać polecenie Format – Zaznaczone serie danych – Desenie i wybrać żądany kolor;
- Wstawić dodatkowe informacje: kliknąć w ikonkę Rysowanie, uaktywnić wykres, w pasku narzędzi Rysowanie kliknąć w ikonkę Pole tekstowe (kursor myszki zmienia kształt) i w żądanym miejscu wykresu „narysować" odpowiednie pole tekstowe i wpisać tam np. gwiazdkę; kliknąć w ikonkę Zapisywanie pliku;
- Przeniesienie wykresu do pliku tekstowego: w pliku tekstowym kliknąć w miejsce, gdzie ma być wstawiony wykres; przejść do arkusza Excel, uaktywnić wykres, kliknąć w ikonkę Kopiuj, przejść do pliku tekstowego, wybrać polecenie Edycja – Wklej specjalnie – Obraz (Format – enhanced metafile) – OK; jeżeli uchwyty ramki wykresu mają postać pustych kwadracików należy tylko ustawić żądany rozmiar.

  Jeżeli natomiast uchwyty są pełnymi kwadracikami, a sam wykres nie jest widoczny, należy uaktywnić wykres, wybrać polecenie Format – Obraz – Układ – Ramka; jeżeli tekst ma się znajdować tylko nad i pod wykresem, a nie dookoła, należy w dalszym ciągu wybrać Zaawansowane – Zawijanie tekstu – Góra i dół.

Wstawianie wykresu do tekstu poleceniem Wklej specjalnie – Obraz (Format – enhanced metafile) ma tę zaletę, że taki plik tekstowy ma małą objętość, natomiast tę wadę, że w razie konieczności poprawienia wykresu trzeba tego dokonać w Excelu, bo edycja wykresu w tekście jest niemożliwa.

Opis technik badawczych

Materiał i metody. Powinny tu się znaleźć trzy podrozdziały:

- Opis badanej zbiorowości,
- Stosowane metody i techniki badawcze,
- Analiza danych.

Opis badanej zbiorowości (populacji) powinien zawierać dokładną definicję tej populacji i jej przybliżoną liczebność, podział na warstwy, zasady losowego wyboru, liczebność i skład badanej próby reprezentatywnej. Należy opisać sposób przeprowadzenia badań i zastosowane techniki badawcze; jeżeli np. ankieta była wzorowana na innej, należy to zaznaczyć i podać źródło. Jeżeli ankieta została ułożona przez autora i nie była publikowana, należy ją zamieścić w aneksie, chyba że redakcja zaleci inaczej. W opisie analizy danych należy podać metody klasyfikacji danych, stosowane testy statystyczne i przyjęty minimalny poziom znamienności (zwykle $p \leq 0{,}05$). Należy przestrzegać zasady, że nie opisuje się metod zawartych w podręcznikach, a odsyłacz do źródła podaje się tylko dla metod nietypowych, rzadko stosowanych. Koniecznym uzupełnieniem jest podanie definicji (lub źródła) tych

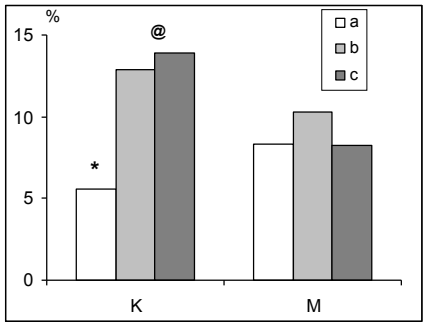
stosowanych terminów, które mogą być niejednoznaczne i zamieszczenie ich na wstępie opracowania, w rozdziale „Materiał i metody" bądź w aneksie.

Prezentacja i omówienie wyników

Przyjęte jest, że opis i omówienie uzyskanych wyników tylko wyjątkowo łączy się w jednym rozdziale, tak więc oddzielnie powinno się przedstawić i omówić wyniki, a oddzielnie przeprowadzić dyskusję. Dyskusja obejmuje konfrontację z danymi z piśmiennictwa, spekulacje dotyczące np. przyczyn uzyskania takich, a nie innych wyników, ich znaczenia teoretycznego i praktycznego, wskazówek co do bardziej szczegółowego zbadania pewnych zagadnień itp.

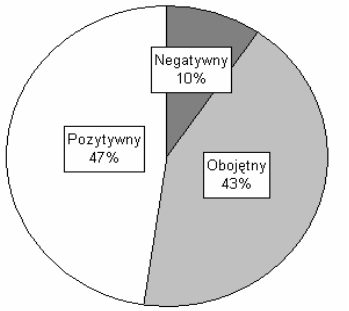
Przedstawienie wyników, zwłaszcza jeżeli jest ich wiele, powinno być podzielone na zagadnienia. Wyniki przedstawia się zazwyczaj w formie ilustracji (tabele, wykresy) i opisu, przy czym opis werbalny może dotyczyć przedstawionego materiału ilustracyjnego bądź stanowić jedyną formę prezentacji pewnych danych. Tam, gdzie to możliwe, powinno się stosować prezentację graficzną, jeżeli jednak materiał dotyczący danego pytania ankiety jest złożony i wymagałby wielu wykresów, korzystniejsze może się okazać ujęcie wyników w tabeli. Możliwa jest również forma mieszana – pokazanie w tabeli zestawienia wszystkich wyników, a tylko ciekawsze z nich pokazać na wykresie. Należy jednak przestrzegać zasady, że jeżeli w tabeli zamieszcza się dane bezwzględne (liczebności), to na wykresie te same dane powinny być ujęte odsetkowo. Nie należy danych z tabeli przedstawiać w identycznej postaci graficznie. Poniżej będą przedstawione przykłady prezentacji na podstawie wyników omawianych w rozdziale 4.

Pytanie nr 1. Odpowiedź na to pytanie jest dychotomiczna (tak – nie), więc wystarczy przedstawić odpowiedzi „nie", bo jest ich mniej, więc ewentualne różnice będą lepiej widoczne. Jak opisano w podrozdziale „Analiza wyborów rozłącznych" (s. 35), w żadnej grupie studentek bądź studentów nie stwierdzono znamiennych różnic w częstościach odpowiedzi „nie" między I i II rokiem studiów, dlatego zamiast podawania danych dla poszczególnych lat wystarczy odpowiednia wzmianka w tekście. Odsetkowe wyniki odpowiedzi na to pytanie przedstawiono na rycinie 6.2. Jak widać, studenci stanowią względnie jednorodną grupę, brak bowiem między nimi znamiennych różnic, podobnie jak między studentkami z grup *b* i *c*. Odsetek odpowiedzi „nie" plasuje studentów bliżej studentek – mieszkanek wsi niż pozostałych, u których odsetek tych odpowiedzi jest znamiennie wyższy.



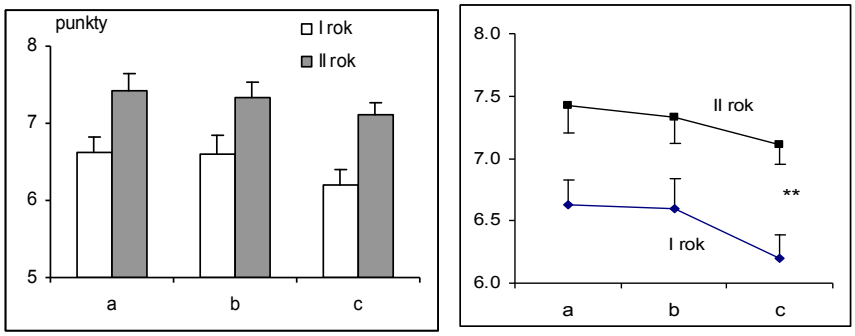
Ryc. 6.2. Odsetki odpowiedzi „nie" na pytanie nr 1 w grupach studentek (K) i studentów (M), mieszkańców wsi (a), miast do 100 000 (b) i powyżej 100 000 mieszkańców (c). Liczebności grup podano w tabeli 5.4.
* Znamiennie mniej niż w grupach *b* i *c* łącznie ($p<0,05$); @ Znamiennie więcej niż u wszystkich studentów łącznie ($p<0,05$)

Pytanie nr 2. Należy przedstawić częstości wszystkich trzech opcji odpowiedzi. Jeżeli zestawia się tylko wyniki w jednej grupie bez porównywania z innymi, najkorzystniej jest pokazać odsetkowe częstości na wykresie kołowym, bo na kolumnowym nie widać, że odsetki sumują się do 100%. Chcąc natomiast porównać różne grupy, lepiej użyć wykresu kolumnowego, na którym można zestawić częstości w poszczególnych grupach.



Ryc. 6.3. Odsetkowe częstości deklarowanego stosunku do zajęć z wychowania fizycznego. Studentki, mieszkanki wsi, I i II rok łącznie (n = 126).

Pytanie nr 3. Odpowiedzi są ilościowe, wyrażone w ciągłej skali analogowej od 0 do 10. Rycina 6.4 przedstawia wysiłek wkładany w przygotowanie do egzaminów deklarowany przez studentki I i II roku z różnych kategorii miejsca zamieszkania.



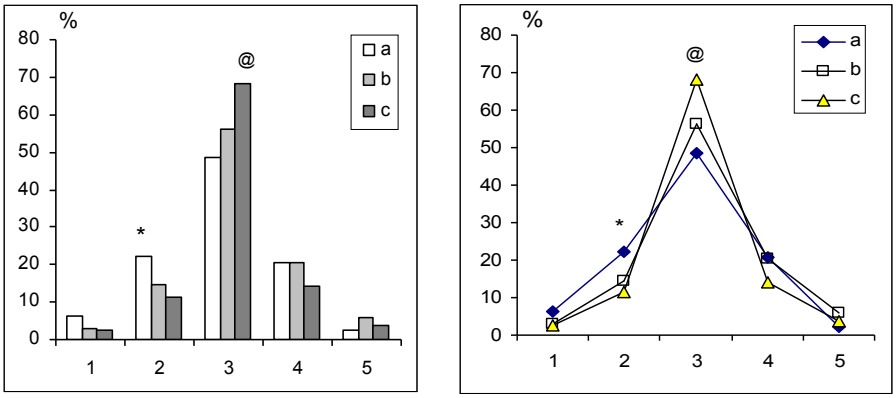
Ryc. 6.4. Deklarowany wysiłek wkładany w przygotowanie do egzaminów przez studentki I i II roku z różnych kategorii miejsca zamieszkania (a, b, c) przedstawiony na wykresie kolumnowym i liniowym (średnie ± SE).
** Znamienna różnica między I i II rokiem ($p<0,01$).Liczebności grup podano w tabeli 5.4.

Wykres kolumnowy jest metodologicznie bardziej poprawny niż liniowy, bowiem w tym ostatnim linie łączące punkty sugerują ciągłość zmian w zależności od kategorii miejsca zamieszkania, co nie musi być prawdą. Niemniej jednak, wykres liniowy jest w tym wypadku bardziej przejrzysty i pozwala łatwiej ocenić różnice i trendy. Ocenę różnic ułatwia naniesienie na wykres nie odchyleń standardowych (SD), lecz błędów standardowych (SE). Fakt, że kreski odpowiadające błędom standardowym nie stykają się ze sobą, wskazuje na znamienność różnic między odpowiadającymi sobie punktami. Wyniki analizy wariancji

(nie przedstawione tutaj) wskazują, że nie ma różnic między kategoriami miejsca zamieszkania, natomiast wyniki dla II roku są znamiennie wyższe niż dla I roku. Może to być związane z większym obciążeniem egzaminami na II roku; wniosek taki byłby bardziej uzasadniony, gdyby podobne wyniki uzyskano również dla studentów (nie przedstawiono tu danych) bądź podano zestawienie porównujące sesje egzaminacyjne I i II roku.

Pytanie nr 4. Wykresy (ryc. 6.5) przedstawiają częstości wyboru jednej opcji odpowiedzi z kafeterii, dlatego odsetki wyborów w poszczególnych grupach sumują się do 100%. Wykres kolumnowy jest poprawną ilustracją odpowiedzi na to pytanie; wykres liniowy jest tu użyty jako profil wielocechowy, przydatny dla roboczej oceny wyników.

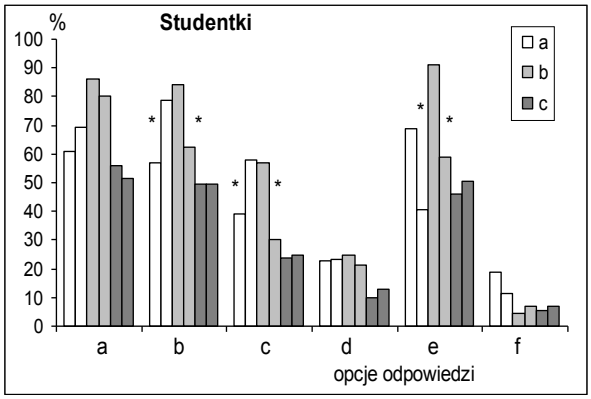
W żadnej kategorii nie obserwowano znamiennych różnic między I i II rokiem studiów, dlatego podano odsetki wyborów dla obu lat łącznie. Znikomy odsetek studentek deklarowało chęć spędzania wakacji samotnie lub w kategorii „inne" (do 6%); najwięcej studentek wybierało opcję „z partnerem", przy czym studentki z dużych miast wybierały tę opcję znamiennie częściej niż studentki z pozostałych dwóch kategorii miejsca zamieszkania. Odpowiednie odsetki wahały się od 48 do 68%. Wybory opcji „b" („z rodziną") i „d" („w grupie") były na podobnym poziomie (11 – 22%), przy czym mieszkanki wsi znamiennie częściej niż inne studentki deklarowały chęć spędzania wakacji z rodziną.



Ryc. 6.5. Częstości wyboru opcji odpowiedzi (1 – 5) na pytanie nr 4. Studentki obu lat studiów łącznie z różnych kategorii miejsca zamieszkania.
* Znamiennie wyższe niż w grupie *c* ($p<0,05$); @ Znamiennie wyższe niż w grupach *a* i *b* ($p<0,05$). Liczebności grup podano w tabeli 5.4.

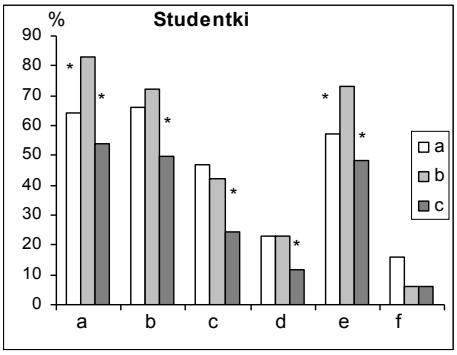
Pytanie nr 5. Jak wynika z ryciny 6.7, różnice między I i II rokiem studiów mogą być znamienne tylko dla kategorii *a* i *b* miejsca zamieszkania i opcji odpowiedzi „b", „c" i „e". Wszystkie te różnice okazały się znamienne, ale ponieważ są one różnokierunkowe, trudno

wyciągać na ich podstawie uogólniające wnioski co do preferencji studentek. O różnicach tych wystarczy wspomnieć w tekście, ewentualnie zamieścić tabelę wyników.

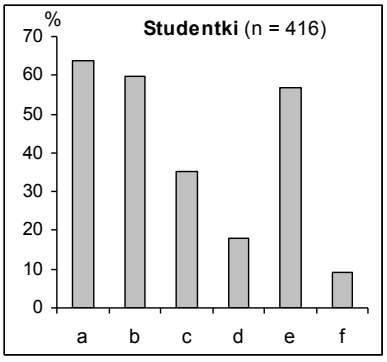


Ryc. 6.6. Odsetkowe wybory opcji odpowiedzi (a – f) na pytanie nr 5 zanotowane w grupach studentek z różnych kategorii miejsca zamieszkania (a – wieś; b – miasta <100 000; c – miasta >100 000 mieszkańców). I i II rok zaznaczone tym samym kolorem.
* Znamienna różnica między I i II rokiem ($p<0,05$)

Widoczna różnokierunkowość preferencji studentek I i II roku (opcje „b", „c", „e") może być podstawą do potraktowania obu lat studiów łącznie i zbadania różnic między kategoriami miejsca zamieszkania (ryc. 6.7).



Ryc. 6.7. Odsetkowe wybory opcji odpowiedzi (a – f) na pytanie nr 5 zanotowane w grupach studentek według miejsca zamieszkania
* Znamienna różnica między sąsiednimi kategoriami ($p<0,05$)

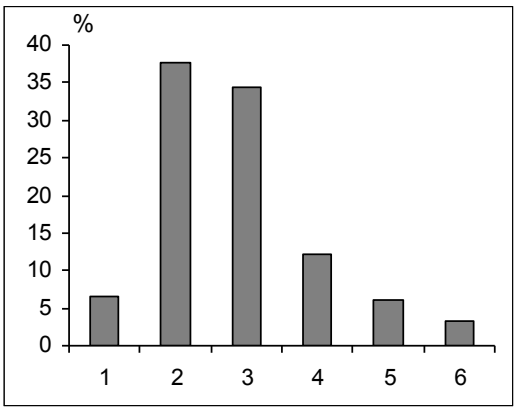


Ryc. 6.8. Odsetkowe wybory opcji odpowiedzi na pytanie nr 5; wszystkie studentki łącznie

Mieszkanki wsi rzadziej niż mieszkanki średnich miast wybierały opcje „a" i „e", natomiast mieszkanki dużych miast rzadziej wybierały wszystkie opcje odpowiedzi (z wyjątkiem „f" – inne; tu różnica była bliska znamienności) niż mieszkanki średnich miast (ryc. 6.7).

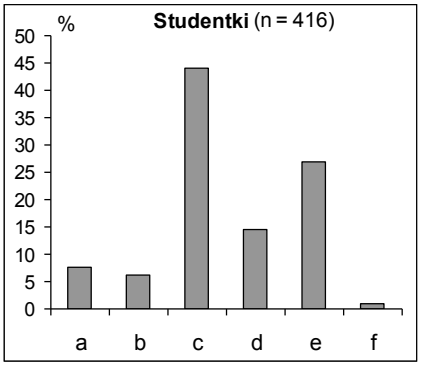
Globalne częstości poszczególnych opcji odpowiedzi studentek pokazano na rycinie 6.8. Częstości wyboru opcji „a”, „b” i „e” były podobne i wynosiły około 60%, najrzadziej wybierano opcję „f” (ok. 9%).

Na rycinie 6.9 przedstawiono odsetkowy rozkład częstości wyborów odpowiedzi na pytanie nr 5 w ankietach wszystkich badanych studentek łącznie. Jak widać, większość studentek wybierała 2 lub 3 odpowiedzi na to pytanie.

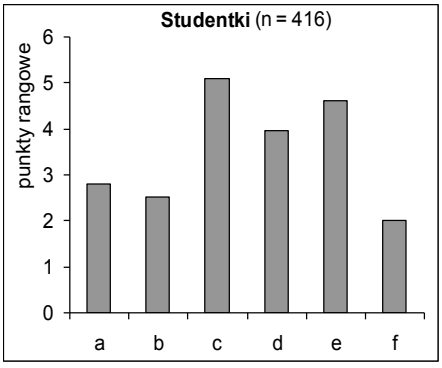


Ryc. 6.9. Odsetkowy rozkład częstości wyborów odpowiedzi na pytanie nr 5; wszystkie studentki łącznie (n = 416)

Pytanie nr 6. Na rycinie 6.10 przedstawiono odsetki wszystkich studentek, które wymieniły poszczególne opcje odpowiedzi na pierwszym miejscu. Jak widać, blisko połowa studentek na pierwszym miejscu wymieniła odpowiedź „c” (jogging), a ponad jedna czwarta – „e” (pływanie).



Ryc. 6.10. Odsetkowy udział poszczególnych opcji odpowiedzi (a – f) wymienionych na pierwszym miejscu; wszystkie studentki łącznie



Ryc. 6.11. Średnie punkty rangowe opcji odpowiedzi (a – f) na pytanie nr 6; wszystkie studentki łącznie

Na rycinie 6.11 pokazano średnie punkty rangowe poszczególnych opcji odpowiedzi dla wszystkich studentek łącznie. Jak wynika z analizy (chi-kwadrat lub test *t*; nie przedstawio-

no tu wyników), tylko opcje „a" i „b" nie różnią się znamiennie, zatem średnia kolejność odpowiedzi układa się następująco: c, e, d, (a, b), f. Dane te obrazują przeciętną ważność poszczególnych opcji wynikającą z ich rankingu. Duże rozbieżności między rycinami 6.10 i 6.11 wynikają z częstego umieszczania np. opcji „d" na drugim miejscu, co wpłynęło na jej wysoką wagę, tylko o jeden punkt niższą niż opcji „c".

Wnioski

O ile w częściach „wyniki" i „dyskusja" przedstawia się i omawia rezultaty przeprowadzonych badań, o tyle we „wnioskach" powinna być zawarta kwintesencja tego, co z tych badań **wynika**. Wnioski nie mogą zatem stanowić streszczenia pracy, a tak się, niestety, często dzieje. Jako przykład mogą posłużyć stwierdzenia dotyczące pytania nr 3 (s. 53), chociaż wyniki dotyczą tylko jednej grupy studentek:

1. Studentki II. roku deklarują większy wysiłek wkładany w przygotowanie się do egzaminów w porównaniu ze studentkami I roku.
2. Większy wysiłek wkładany w przygotowanie się do egzaminów przez studentki II roku w porównaniu ze studentkami I roku może być związany z różnicami w obciążeniu sesji egzaminacyjnych I i II roku, co należy uwzględnić w dalszych badaniach.

Pierwsze zdanie jest stwierdzeniem faktu, a nie wnioskiem. Drugie zdanie jest wnioskiem, wymienia bowiem prawdopodobną przyczynę zaobserwowanego faktu i co z tego wynika.

Bardzo często wnioski dzieli się na poznawcze i aplikacyjne. Badania przeprowadza się po to, aby odpowiedzieć na pytania badawcze będące rozwinięciem hipotezy lub hipotez badawczych, dąży się zatem do stwierdzenia faktów przemawiających za lub przeciw postawionym hipotezom. Bezpośrednimi przesłankami „wniosków poznawczych" są wyniki badań, bo w nich zawarta jest informacja. Natomiast „wnioski aplikacyjne" są postulatami (rekomendacjami) formułowanymi nie tylko na podstawie uzyskanych wyników i wysnutych wniosków, ale również na podstawie przesłanek prakseologicznych.

7. PODSUMOWANIE

Przedstawione w tym opracowaniu zagadnienia są dalekie od wyczerpującego omówienia tematu. Można mieć jednak nadzieję, że omawiane techniki zestawieniowe i obliczeniowe z jednej strony pozwolą na lepsze wykorzystanie informacji zawartych w odpowiedziach na ankietę, z drugiej zaś zachęcą do dalszego rozwijania tych technik. Spróbujmy zatem dokonać krótkiego przeglądu poszczególnych rozdziałów.

Dobór próby reprezentatywnej. Jak wspomniano we wstępie, omówienie ograniczono do prób reprezentatywnych pobieranych warstwowo, bowiem badacz nie dysponujący odpowiednim zapleczem logistycznym może nie być w stanie przeprowadzić bezpośredniego losowania próby, zwłaszcza z dużej populacji. Pominięto zatem różne rodzaje badań niereprezentatywnych, a także nowoczesne metody sieciowe zbierania informacji.

Rodzaje pytań i odpowiedzi oraz *Konstrukcja arkusza pytań*. Ze względu na raczej elementarny poziom tego podręcznika omówiono tylko najważniejsze typy pytań i odpowiedzi, dając przy tym zbyt mało konkretnych przykładów ich zastosowania oraz bardziej szczegółowej analizy układu pytań w arkuszu. Przykłady takie można znaleźć w dostępnym piśmiennictwie fachowym.

Sporządzanie i analiza zestawień. Omówiono podstawowe techniki analizy danych oparte przede wszystkim na dwuwymiarowej funkcji chi-kwadrat, a w pewnym stopniu również na teście *t*, analizie wariancji i miarach korelacyjnych. Wynika z tego, że analiza dotyczyła prawie wyłącznie pojedynczych cech określonych pytaniami. Niewątpliwym brakiem jest pominięcie w omówieniu unormowanych profilów wielocechowych i ich klasyfikacji, wykraczałoby to jednak poza założony elementarny zakres.

Prezentacja wyników. W omówieniu elementów prezentacji graficznej ograniczono się do zagadnień typowych dla drukowanych tekstów – raportów, publikacji, rozpraw naukowych itp. Pominięto szczegółowe omówienie układu i zawartości tabel, jak również prezentacji multimedialnych. Omówienie wybranych części rozpraw naukowych (wstęp, materiał i metody, wyniki, dyskusja, wnioski) jest bardzo pobieżne, bowiem specyfika danej dziedziny wiedzy może wymagać nieco innego rozłożenia akcentów. Ta część ograniczona jest zatem do pewnych ogólnych wskazówek i do zwrócenia uwagi na najczęściej popełniane błędy.

Istotnym czynnikiem, który wpłynął na wspomniane wyżej braki niniejszej książki, była presja czasu – zależało mi na tym, aby książka jak najprędzej ukazała się jako pomoc dydaktyczna. Mam nadzieję, że następne wydanie zostanie odpowiednio uzupełnione.

LITERATURA UZUPEŁNIAJĄCA

1. Blalock H.M. Statystyka dla socjologów. PWN, Warszawa, 1975
2. Brzeziński J. Metodologia badań psychologicznych. PWN, Warszawa, 1999
3. Ferguson G.A., Takane Y. Analiza statystyczna w psychologii i pedagogice. PWN, Warszawa, 1999
4. Gajewski J. Edytor prezentacji multimedialnych – PowerPoint. W: Podstawy informatyki cz. II, R.Stupnicki i W.Malinowska (red.), Wydawnictwa AWF Warszawa, 2002
5. Iwańska E. Arkusz kalkulacyjny – Excel. W: Podstawy informatyki cz. I, R.Stupnicki i W.Malinowska (red.), Wydawnictwa AWF Warszawa, 2002
6. Pilch T., Bauman T. Zasady badań pedagogicznych (wyd. 2). Wydawnictwo Akademickie Żak, Warszawa, 2001
7. Stupnicki R. Biometria. Wydawnictwa AWF Warszawa, 2015

SŁOWNICZEK TERMINÓW

Poniżej podano znaczenia terminów stosowanych w niniejszym opracowaniu, aby uniknąć nieporozumień związanych z ewentualnymi innymi definicjami podawanymi w różnych źródłach.

Ankieta – termin używany w dwóch znaczeniach: 1) badanie określonej grupy respondentów w celu uzyskania informacji o deklarowanych poglądach, opiniach, postawach itp. (ang. *questionnaire studies*); 2) arkusz pytań służący do przeprowadzenia badań (ang. *questionnaire*)

Badania eksperymentalne – przeprowadzane w celu ustalenia ew. wpływu określonego czynnika (czynników) na badane cechy; zazwyczaj nie stosuje się kryterium reprezentatywnego doboru prób

Badania reprezentatywne – przeprowadzane na próbie reprezentatywnej (zob.)

Badania wyczerpujące – obejmują całą zdefiniowaną populację, a nie tylko jej reprezentację

Hipoteza zerowa – hipoteza statystyczna zakładająca, że np. nie ma różnic między średnimi dwu lub więcej grup, nie ma zależności między danymi zestawionymi w tabeli itp.

Kafeteria – zestaw odpowiedzi (*opcji odpowiedzi*) dołączony do danego pytania.

Kategoria – jednostka podziału większego obiektu na mniejsze, np. kategorie w obrębie warstwy populacji, kategorie jakościowe lub ilościowe) odpowiedzi na dane pytanie itp.

Kwestionariusz – arkusz pytań; w psychometrii odnosi się do standaryzowanego testu

Obiekt – jednostka poddana badaniu; może to być pojedyncza osoba, szkoła, określony zespół danych itp.

Odpowiedź otwarta – inaczej *opcja otwarta*, jedna z opcji odpowiedzi zawartych w kafeterii (np. *inne – jakie?*) do określenia przez respondenta.

Populacja – zespół jednostek (ludzi, zwierząt, roślin, przedmiotów, wyników pomiaru itp.) spełniający kryteria określone definicją; zbliżony termin: *zbiorowość*.

Próba losowa prosta – grupa osób bezpośrednio, losowo wybranych z odpowiednio zdefiniowanej całej populacji, nie dzielonej np. na warstwy

Próba reprezentatywna – grupa osób wybranych losowo z odpowiednio zdefiniowanej populacji lub jej części, np. warstwy lub kategorii, w założeniu stanowiąca odzwierciedlenie populacji na małą skalę

Rzetelność – (ang. *reliability*) wiarogodność, w tym wypadku – odpowiedzi na pytania ankiety; mała rzetelność może wynikać ze złej woli, niezrozumienia pytań itp.

Skala – zbiór kategorii opisowych (np. wcale – czasem – zawsze) lub liczbowych (np. 1 – 2 – 3 punkty) tworzących logicznie uporządkowany ciąg, służący do pomiaru danego zjawiska i obejmujący jego pełny zakres

Tendencja – prawdopodobieństwo hipotezy zerowej nie osiągnęło wprawdzie poziomu krytycznego (zwykle p = 0,05), ale jest mu bliskie; zwykle leży ono między 0,05 a 0,10. Opisując wyniki analizy statystycznej należy podać przyjęty poziom prawdopodobieństwa.

Trafność – (ang. *accuracy*) zgodność charakteru (nie nasilenia!) badanego zjawiska z założeniem, np. zakłada się pomiar masy ciała i mierzy się masę ciała, a nie np. masę ciała i ubrania

Warstwa – część zdefiniowanej populacji, dzielona dalej na kategorie (zob.); na przykład, warstwa terytorialna może zawierać kategorie administracyjne (województwa) lub geograficzne (regiony), warstwa edukacyjna – różne kategorie (typy) szkół itp.

Warstwowanie – (ang. *stratification*) podział zdefiniowanej populacji na części (warstwy) wg określonych kryteriów, np. terytorialnych,

Zbiorowość – (ang. *community*) lub społeczność, termin używany nieraz zamiennie z „populacją" (zob.), ale w odczuciu mniej precyzyjny

Zmienna – (ang. *variable*) w badaniach ankietowych – pytanie, a ściślej treść pytania; termin bliskoznaczny: *cecha*.

Znamienność – (ang. *significance*) równoważne terminy: *znamienność statystyczna, istotność statystyczna*. Oznacza, że został przekroczony pewien przyjęty poziom prawdopodobieństwa (zwykle p = 0,05) upoważniający do odrzucenia hipotezy zerowej (patrz wyżej). Opisując wyniki analizy statystycznej należy podać przyjęty poziom prawdopodobieństwa.

TABLICE STATYSTYCZNE

Wartości funkcji χ^2 (chi-kwadrat)

p / df	º	*	**	***
	0,10	**0,05**	**0,01**	**0,001**
1	2.71	3.84	6.63	10.8
2	4.61	5.99	9.21	13.8
3	6.25	7.81	11.4	16.3
4	7.78	9.49	13.3	18.5
5	9.24	11.1	15.1	20.5
6	10.64	12.6	16.8	22.5
7	12.02	14.1	18.5	24.3
8	13.36	15.5	20.1	26.1
9	14.68	16.9	21.7	27.9
10	15.99	18.3	23.2	29.6
11	17.3	19.7	24.7	31.3
12	18.6	21.0	26.2	32.9
13	19.8	22.4	27.7	34.5
14	21.1	23.7	29.1	36.1
15	22.3	25.0	30.6	37.7
16	23.5	26.3	32.0	39.3
17	24.8	27.6	33.4	40.8
18	26.0	28.9	34.8	42.3
19	27.2	30.1	36.2	43.8
20	28.4	31.4	37.6	45.3
25	34.4	37.7	44.3	52.6
30	40.3	43.8	50.9	59.7
40	51.7	55.8	63.7	73.4
50	63.1	67.5	76.2	86.7

Wartości funkcji $n \cdot \ln(n)$

W pierwszej kolumnie podane są dziesiątki, w pierwszym wierszu – jednostki.
Wartość $n \cdot ln(n)$ dla liczebności 23 wynosi 72,12.

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0	0	0	1,39	3,30	5,55	8,05	10,75	13,62	16,64	19,78
10	23,03	26,38	29,82	33,34	36,95	40,62	44,36	48,16	52,03	55,94
20	59,91	63,93	68,00	72,12	76,27	80,47	84,71	88,99	93,30	97,65
30	102,04	106,45	110,90	115,38	119,90	124,44	129,01	133,60	138,23	142,88
40	147,56	152,26	156,98	161,73	166,50	171,30	176,12	180,96	185,82	190,70
50	195,60	200,52	205,46	210,43	215,41	220,40	225,42	230,45	235,51	240,57
60	245,66	250,76	255,88	261,02	266,17	271,34	276,52	281,71	286,93	292,15
70	297,39	302,65	307,92	313,20	318,50	323,81	329,14	334,47	339,82	345,19
80	350,56	355,95	361,35	366,76	372,19	377,63	383,07	388,53	394,01	399,49
90	404,98	410,49	416,00	421,53	427,07	432,62	438,18	443,75	449,33	454,92
100	460,52	466,13	471,75	477,38	483,02	488,67	494,32	499,99	505,67	511,36
110	517,05	522,76	528,47	534,19	539,93	545,67	551,42	557,17	562,94	568,72
120	574,50	580,29	586,09	591,90	597,71	603,54	609,37	615,21	621,06	626,92
130	632,78	638,65	644,53	650,42	656,31	662,21	668,12	674,04	679,96	685,89
140	691,83	697,78	703,73	709,69	715,65	721,63	727,61	733,59	739,59	745,59
150	751,60	757,61	763,63	769,66	775,69	781,73	787,78	793,83	799,89	805,96
160	812,03	818,11	824,19	830,28	836,38	842,48	848,59	854,70	860,83	866,95
170	873,09	879,22	885,37	891,52	897,68	903,84	910,01	916,18	922,36	928,54
180	934,73	940,93	947,13	953,34	959,55	965,77	971,99	978,22	984,45	990,69
190	996,93	1003,18	1009,44	1015,70	1021,96	1028,23	1034,51	1040,79	1047,08	1053,37
200	1059,66	1065,96	1072,27	1078,58	1084,90	1091,22	1097,54	1103,87	1110,21	1116,55
210	1122,89	1129,24	1135,60	1141,96	1148,32	1154,69	1161,06	1167,44	1173,82	1180,21
220	1186,60	1192,99	1199,39	1205,80	1212,21	1218,62	1225,04	1231,46	1237,89	1244,32
230	1250,76	1257,20	1263,64	1270,09	1276,55	1283,00	1289,46	1295,93	1302,40	1308,87
240	1315,35	1321,84	1328,32	1334,81	1341,31	1347,81	1354,31	1360,82	1367,33	1373,85
250	1380,37	1386,89	1393,42	1399,95	1406,48	1413,02	1419,57	1426,11	1432,66	1439,22
260	1445,78	1452,34	1458,91	1465,48	1472,05	1478,63	1485,21	1491,80	1498,38	1504,98
270	1511,57	1518,17	1524,78	1531,39	1538,00	1544,61	1551,23	1557,85	1564,48	1571,11
280	1577,74	1584,38	1591,02	1597,66	1604,31	1610,96	1617,61	1624,27	1630,93	1637,60
290	1644,27	1650,94	1657,61	1664,29	1670,97	1677,66	1684,35	1691,04	1697,73	1704,43
300	1711,13	1717,84	1724,55	1731,26	1737,98	1744,70	1751,42	1758,14	1764,87	1771,60
310	1778,34	1785,08	1791,82	1798,56	1805,31	1812,06	1818,81	1825,57	1832,33	1839,10
320	1845,86	1852,63	1859,41	1866,18	1872,96	1879,74	1886,53	1893,32	1900,11	1906,90
330	1913,70	1920,50	1927,30	1934,11	1940,92	1947,73	1954,55	1961,37	1968,19	1975,01
340	1981,84	1988,67	1995,51	2002,34	2009,18	2016,02	2022,87	2029,72	2036,57	2043,42
350	2050,28	2057,14	2064,00	2070,86	2077,73	2084,60	2091,48	2098,35	2105,23	2112,11
360	2119,00	2125,88	2132,78	2139,67	2146,56	2153,46	2160,36	2167,27	2174,17	2181,08
370	2188,00	2194,91	2201,83	2208,75	2215,67	2222,60	2229,53	2236,46	2243,39	2250,33
380	2257,27	2264,21	2271,15	2278,10	2285,05	2292,00	2298,95	2305,91	2312,87	2319,83
390	2326,80	2333,76	2340,73	2347,71	2354,68	2361,66	2368,64	2375,62	2382,61	2389,60
400	2396,59	2403,58	2410,57	2417,57	2424,57	2431,57	2438,58	2445,59	2452,60	2459,61
410	2466,62	2473,64	2480,66	2487,68	2494,71	2501,74	2508,77	2515,80	2522,83	2529,87
420	2536,91	2543,95	2550,99	2558,04	2565,09	2572,14	2579,19	2586,25	2593,30	2600,37
430	2607,43	2614,49	2621,56	2628,63	2635,70	2642,78	2649,85	2656,93	2664,01	2671,10
440	2678,18	2685,27	2692,36	2699,45	2706,55	2713,64	2720,74	2727,84	2734,95	2742,05
450	2749,16	2756,27	2763,38	2770,50	2777,62	2784,74	2791,86	2798,98	2806,11	2813,23

Wartości testu *t* Studenta

Test dwustronny

p / df	°	*	**	***
	0,10	**0,05**	**0,01**	**0,001**
4	2.132	2.776	4.604	
5	2.015	2.571	4.032	6.859
6	1.993	2.447	3.707	5.959
7	1.895	2.365	3.499	5.405
8	1.860	2.306	3.355	5.041
9	1.833	2.262	3.250	4.781
10	1.812	2.228	3.169	4.587
11	1.796	2.201	3.106	4.437
12	1.782	2.179	3.055	4.318
13	1.771	2.160	3.012	4.221
14	1.761	2.145	2.977	4.140
15	1.753	2.131	2.947	4.073
16	1.746	2.120	2.921	4.015
17	1.740	2.110	2.898	3.965
18	1.734	2.101	2.878	3.922
19	1.729	2.093	2.861	3.883
20	1.725	2.086	2.845	3.850
21	1.721	2.080	2.832	3.820
22	1.717	2.074	2.819	3.792
23	1.714	2.069	2.808	3.768
24	1.711	2.064	2.797	3.745
25	1.708	2.060	2.788	3.725
26	1.706	2.056	2.779	3.707
27	1.703	2.052	2.771	3.690
28	1.701	2.048	2.763	3.674
29	1.699	2.045	2.756	3.659
30	1.697	2.042	2.750	3.646
40	1.684	2.021	2.704	3.551
60	1.671	2.000	2.660	3.460
90	1.662	1.987	2.632	3.402
120	1.658	1.980	2.617	3.373
∞	1.645	1.960	2.576	3.291

Wartości funkcji *F* Snedecora

p < 0,05 (*)

df	1	2	3	4	5	6	7	8	9	10
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93
125	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

p < 0,01 (**)

10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.79	2.70
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50
125	6.81	4.75	3.92	3.45	3.14	2.92	2.76	2.63	2.53	2.44
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

ANEKS

Dane z indywidualnych arkuszy wypełnionych przez studentki I roku wychowania fizycznego ze środowisk wiejskich. Wykaz pytań znajduje się na stronach 23 – 24.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	No.	Płeć	Rok stud.	M.zam	1	2	3	4	5	6	5n	a	b	c	d	e	f
2	6	F	I	a	1	0	5.5	a	abe	badecf	3	2	1	5	3	4	6
3	9	F	I	a	1	1	7.2	d	ae	afcbed	2	1	4	3	6	5	2
4	23	F	I	a	1	1	6.1	c	bcef	dfcbae	4	5	4	3	1	6	2
5	37	F	I	a		0	9.0	b	d	dfcbae	1	5	4	3	1	6	2
6	45	F	I	a	1	0	5.6	b	ab	dfbace	2	4	3	5	1	6	2
7	46	F	I	a	0	0	4.3	c	ae	edacbf	2	3	5	4	2	1	6
8	51	F	I	a	1	1	5.8	d	acd	dfcbae	3	5	4	3	1	6	2
9	52	F	I	a	1	0	6.8	c	be	afcbed	2	1	4	3	6	5	2
10	59	F	I	a	1	0	7.5	c	be	dfbace	2	4	3	5	1	6	2
11	63	F	I	a	1		8.1	d	abce	edacbf	4	3	5	4	2	1	6
12	77	F	I	a	0	-1	6.6	c	acde	dfcbae	4	5	4	3	1	6	2
13	79	F	I	a	1	1	4.5	b	ae	edacbf	2	3	5	4	2	1	6
14	95	F	I	a	1	0	5.0	a	abd	edacbf	3	3	5	4	2	1	6
15	98	F	I	a	1	1	5.5	c	ae	dfbace	2	4	3	5	1	6	2
16	102	F	I	a	1	1	6.1	c	ae	edacbf	2	3	5	4	2	1	6
17	106	F	I	a	1	0	7.3	c	ab	fcbade	2	4	3	2	5	6	1
18	121	F	I	a	1	0	4.2	b	aef	dfbace	3	4	3	5	1	6	2
19	125	F	I	a	1	0	6.6	c	be	dfcbae	2	5	4	3	1	6	2
20	130	F	I	a	1	1	4.4	c	be	edacbf	2	3	5	4	2	1	6
21	133	F	I	a	1	1	5.2	d	bcd	febacd	3	4	3	5	6	2	1
22	140	F	I	a	1	0	6.8	d	ae	dfcbae	2	5	4	3	1	6	2
23	168	F	I	a	1	1	8.1	c	ab	edacbf	2	3	5	4	2	1	6
24	169	F	I	a	1	1	4.1	c	abcdef	edacbf	6	3	5	4	2	1	6
25	184	F	I	a	1	0	5.1	b	ae	afcbed	2	1	4	3	6	5	2
26	189	F	I	a	1	1	4.0	d	bde	edacbf	3	3	5	4	2	1	6
27	191	F	I	a	1	1	4.2	b	acd		3						
28	196	F	I	a	1	1	6.3	c	ab	edacbf	2	3	5	4	2	1	6
29	212	F	I	a	1	1	3.1	c	ae	afcbed	2	1	4	3	6	5	2
30	215	F	I	a	1	1	5.5	d	be	edacbf	2	3	5	4	2	1	6
31	251	F	I	a	1	1	5.9	c	abcef	deacbf	5	3	5	4	1	2	6
32	256	F	I	a	1	0	6.3	d	be		2						
33	261	F	I	a	1	-1	7.0	a	abd	dcbeaf	3	5	3	2	1	4	6
34	270	F	I	a	1	-1	6.4	b	abe	fcbade	3	4	3	2	5	6	1
35	275	F	I	a	1	1	6.2	a	be	ceadbf	2	3	5	1	4	2	6
36	278	F	I	a	1	1	5.7	e	ac	edacbf	2	3	5	4	2	1	6
37	289	F	I	a	1	0	8.0	c	be	edacbf	2	3	5	4	2	1	6
38	291	F	I	a	1	1	6.6	c	be	febacd	2	4	3	5	6	2	1
39	296	F	I	a	1	0	6.2	d	bcef	edacbf	4	3	5	4	2	1	6
40	302	F	I	a	1	1	7.8	d	ac	fcbade	2	4	3	2	5	6	1
41	312	F	I	a	1	0	3.2	c	bc	febacd	2	4	3	5	6	2	1
42	324	F	I	a	1	1	4.3	d	ae	debacf	2	4	3	5	1	2	6
43	332	F	I	a	1	1	8.3	b	ae	edacbf	2	3	5	4	2	1	6
44	334	F	I	a	1	0	6.6	c	ae	cfbdae	2	5	3	1	4	6	2
45	335	F	I	a	1	0	7.1	c	bef	edacbf	3	3	5	4	2	1	6
46	344	F	I	a	1	0	8.9	b	be	dcbeaf	2	5	3	2	1	4	6
47	347	F	I	a	1	0	7.7	c	cdf	febacd	3	4	3	5	6	2	1
48	352	F	I	a	1	0	6.9	d	be	cfbdae	2	5	3	1	4	6	2

49	363	F	I	a	1	0	8.2	c	abe	edacbf	3	3	5	4	2	1	6
50	368	F	I	a	1	1	8.5	b	bce	dcbeaf	3	5	3	2	1	4	6
51	379	F	I	a	1	0	2.8	c	bdef	dcbeaf	4	5	3	2	1	4	6
52	388	F	I	a	1	1	6.3	b	ace	bfadce	3	3	1	5	4	6	2
53	391	F	I	a	1	1	6.8	a	ace	edacbf	3	3	5	4	2	1	6
54	396	F	I	a	1	1	7.2	d	abe	bfadce	3	3	1	5	4	6	2
55	411	F	I	a	1	1	6.1	b	acdef	fabdce	5	2	3	5	4	6	1
56	425	F	I	a	1	1	5.4	d	ace	edacbf	3	3	5	4	2	1	6
57	426	F	I	a	1	0	9.1	c	adef	dfabce	4	3	4	5	1	6	2
58	433	F	I	a	1	1	3.5	d	bce	bfadce	3	3	1	5	4	6	2
59	441	F	I	a	1	0	3.9	b	acdef	edacbf	5	3	5	4	2	1	6
60	452	F	I	a	1	1	8.8	b	ae	dfabce	2	3	4	5	1	6	2
61	455	F	I	a	1	0	8.7	c	ac	edacbf	2	3	5	4	2	1	6
62	459	F	I	a	1	0	3.4	d	aef	cfbdae	3	5	3	1	4	6	2
63	467	F	I	a	1	1	3.3	c	bc	fabdce	2	2	3	5	4	6	1
64	473	F	I	a	1	1	6.0	a	abf	edacbf	3	3	5	4	2	1	6
65	485	F	I	a	1	1	6.4	c	acde	abdecf	4	1	2	5	3	4	6
66	504	F	I	a	1	0	8.4	c	bef	edacbf	3	3	5	4	2	1	6
67	512	F	I	a	1	0	9.2	d	be	adcfbe	2	1	5	3	2	6	4
68	526	F	I	a	1	0	7.2	b	bc	cfbdae	2	5	3	1	4	6	2
69	533	F	I	a	1	1	3.3	d	c	cfebad	1	5	4	1	6	3	2
70	537	F	I	a	1	1	3.9	c	bd	ebcfad	2	5	2	3	6	1	4
71	542	F	I	a	1	0	7.0	c	bc	cedbaf	2	5	4	1	3	2	6
72	549	F	I	a	1	1	3.5	c	acde	adbecf	4	1	3	5	2	4	6
73	568	F	I	a	1	1	8.0	d	ac	dfabce	2	3	4	5	1	6	2
74	572	F	I	a	1	1	8.0	b	ab	fdceba	2	6	5	3	2	4	1
75	591	F	I	a	1	0	8.6	e	abce	abcdef	4	1	2	3	4	5	6
76					73	73	74	74	74	72							
77				Σ	71		6.20	śr.									
78							1.73	SD				1	2	3	4	5	6
79	*a*	-1	1			3		6	45		2	8	4	7	19	24	10
80	*b*	0	2			32		16	42		38	3	3	7	26	8	25
81	*c*	1	3			38		31	29		21	31	23	14	3	1	0
82	*d*		4					19	17		9	12	15	24	11	8	2
83	*e*		5					2	51		3	17	27	20	3	5	0
84	*f*		6						14		1	1	0	0	10	26	35

Funkcje EXCEL użyte w tej książce i ich angielskie odpowiedniki

DŁ	LEN
ILE.NIEPUSTYCH	COUNTA
LICZ.JEŻELI	COUNTIF
LOS	RAND
ODCH.STANDARDOWE	STDEV
SORTUJ	SORT
ŚREDNIA	AVERAGE
ZNAJDŹ	FIND

SKOROWIDZ