

Ocena skuteczności projektowanych działań marketingowych

W przeważającej mierze ocena skuteczności projektowanych działań marketingowych dokonuje się na podstawie testów rynkowych, czyli badań realizowanych w formie eksperymentu. Zestawione w części II pracy możliwe schematy eksperymentalne dają się zamknąć w dwa bloki: badania realizowane na próbach losowych (tzw. schematy prawdziwe) oraz pozostałe. Podział ten jest zasadny, zważywszy na metody analizy wyników eksperymentu. W schematach prawdziwych do oceny istotności wyników eksperymentu konieczne jest wykorzystanie procedur wnioskowania statystycznego. W pozostałych schematach prowadzi się jedynie analizę porównawczą wyników rejestrowanych w wydzielonych grupach badawczych (eksperymentalnych i kontrolnych). W obu podejś-

Tablica 11.1. Najpopularniejsze sposoby oceny wyników eksperymentu w schematach prawdziwych

Skala pomiaru wyników eksperymentu (typ zmiennej zależnej)	Pomiar niezależny	Pomiar zależny
Nominalna	test U dla dwóch frakcji, test niezależności chi-kwadrat, ryzyko względne	test McNemara, test Cochra
Porządkowa	test serii, test Walda-Wolfowitza, test Kołmogorowa-Smirnowa, test Kruskala-Wallisa	test Friedmana
Przedziałowa lub ilorazowa	test U lub t dla dwóch średnich, analiza wariancji dla doboru całkowicie losowego	test t , analiza wariancji dla schematów blokowych, analiza <i>conjoint</i>

ciach przedmiotem analizy są wyniki eksperymentu rejestrowane na skalach nominalnych, porządkowych, przedziałowych lub ilorazowych. W przypadku skal nominalnych i porządkowych ocenie podlegają wskaźniki struktury oraz wartość dominanty i ewentualnie mediany. W przypadku skal przedziałowych i ilorazowych ocenie podlegają przede wszystkim wartość średnia oraz wariancja lub odchylenie standardowe. W schematach prawdziwych konieczne jest również rozróżnienie, czy wyniki pomiarów pierwotnego i końcowego dotyczą odmiennych grup, tj. badanej i kontrolnej, czy też są rejestrowane w ramach tej samej próby. W pierwszym przypadku mamy do czynienia z pomiarami niezależnymi, w drugim zaś z pomiarami zależnymi. Ma to wpływ na dobór właściwych testów statystycznych służących do oceny istotności wyników eksperymentu. Zakres najpopularniejszych procedur w tym przypadku zestawia tablica 11.1.

11.1. Test U dla dwóch frakcji

Test ten umożliwia ocenę eksperymentu, który opiera się na schemacie dwóch niezależnych pomiarów, a jego wyniki są rejestrowane na dychotomicznej skali nominalnej. Dotyczy on eksperymentu, w którym sprawdza się wpływ działania na zmienną zależną jednorodnego czynnika kontrolowanego, przy czym działaniu czynnika kontrolowanego poddaje się tylko jedną grupę jednostek. Chodzi tu o porównanie dwóch frakcji w populacjach, które reprezentuje każda z prób. Wymagana jest znaczna liczebność każdej z prób, tzn. co najmniej 120 jednostek; w każdej z nich odnotowuje się pewną liczbę elementów wyróżnionych, odpowiednio m_1 oraz m_2 . Sprawdzianem przypuszczenia, że obie frakcje są równe, jest statystyka U , którą definiuje wzór:

$$U = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}},$$

gdzie:

$$n = \frac{n_1 n_2}{n_1 + n_2}, \quad \bar{p} = \frac{m_1 + m_2}{n_1 + n_2};$$

m_1 – liczba elementów wyróżnionych w pierwszej próbie (grupa kontrolna); m_2 – liczba elementów wyróżnionych w drugiej próbie (grupa badana); n_1 – liczebność pierwszej próby; n_2 – liczebność drugiej próby.

Jeśli przypuszczenie o równości frakcji jest prawdziwe, to wartość bezwzględna statystyki U powinna być mała, tzn. nie powinna przekraczać wartości krytycznej odczytanej z tablic dystrybucyj rozkładu normalnego dla $F(u_\alpha) = 1 - \alpha/2$, gdzie α to ryzyko pomyłki (zwane poziomem istotności), na ogół wynoszące¹ 0,05.

Przykład 11.1

Ocena perswazyjności przygotowanego filmu reklamowego była przedmiotem badania eksperymentalnego w formie schematu symulowanego z pomiarem pierwotnym i końcowym (por. p. 6.3.1). W celu oceny, jak przygotowany film reklamowy zmienia nastawienie do produktu, zaproszono na seans filmowy losowo wybranych 300 konsumentów z rynku docelowego i następnie losowo rozdzielono ich na dwie równoliczne grupy widzów, kolejno oglądających film. Widzów z pierwszej grupy (grupa kontrolna) poproszono o wskazanie marki, która w danej klasie produktów oferuje ich zdaniem produkt najbardziej ekologiczny. W tym pomiarze pierwotnym odnotowano 39 respondentów, którzy jako markę najbardziej ekologiczną wskazali tę, która była przedmiotem badania. Widzom z drugiej grupy, nie sondując przed seans filmowym opinii na wskazany temat, w trakcie seansu filmowego (grupa badana) wyemitowano reklamę. Po seansie z kolei widzów z tej grupy poproszono o wskazanie marki ich zdaniem oferującej produkt najbardziej ekologiczny. W tym pomiarze końcowym odnotowano 63 osoby, które wyróżniły markę będącą przedmiotem badania. Przedmiotem analizy były różnice w percepcji w przypadku obu grup.

Zgodnie z przedstawionymi wyżej oznaczeniami otrzymujemy:

$$n_1 = 150, m_1 = 39, m_1/n_1 = 0,26,$$

$$n_2 = 150, m_2 = 63, m_2/n_2 = 0,42,$$

$$n = \frac{n_1 n_2}{n_1 + n_2} = \frac{150 \cdot 150}{300} = 75, \quad \bar{p} = \frac{m_1 + m_2}{300} = 0,34,$$

$$U = \frac{\frac{m_1 - m_2}{n_1 - n_2}}{\sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}} = \frac{0,26 - 0,42}{\sqrt{\frac{0,34 \cdot (1 - 0,34)}{75}}} \approx -2,93.$$

Wartość bezwzględna statystyki U obliczona na podstawie wyników eksperymentu, czyli 2,93, przekracza wartość krytyczną odczytaną z tablic rozkładu normalnego dla ryzyka pomyłki 0,05, czyli liczbę 1,96. Oznacza to, że wyniki pomiaru pierwotnego, w grupie kontrolnej, oraz pomiaru końcowego, w grupie badanej, istotnie się różnią. Przeprowadzony eksperyment wskazuje, iż przygotowany film reklamowy istotnie zmienia wyobrażenie o marce.

Ocena, czy film reklamowy istotnie zwiększa przekonanie o ekologicznym charakterze produktu sprzedawanego pod badaną marką, jest badaniem jednostronnym polegającym na sprawdzeniu, czy równość frakcji może być przeciwstawiona relacji wskazującej, że frakcja przekonanych pod wpływem reklamy jest istotnie większa od frakcji przekonanych wśród tych, którzy reklamy

¹ Wówczas $u_\alpha = 1,96$, co oznacza, że wartość bezwzględna statystyki U obliczona według podanego wzoru nie powinna przekroczyć liczby 1,96.

nie oglądają. Sprawdzenie takiego układu warunków sprowadza się do wyznaczenia wartości statystyki U według wzoru:

$$U = \frac{\frac{m_2 - m_1}{n_2 - n_1}}{\sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}} = \frac{0,42 - 0,26}{\sqrt{\frac{0,34 \cdot (1 - 0,34)}{75}}} \approx 2,93$$

oraz odczytania z tablic dystrybucyj rozkładu normalnego wartości krytycznej dla $F(u_\alpha) = 1 - \alpha$. Dla $\alpha = 0,05$ jest to wartość 1,64. Wynik wskazuje, iż z ryzykiem pomyłki 5% można uznać, że film reklamowy istotnie zwiększa zainteresowanie produktem.

11.2. Test niezależności chi-kwadrat

Test ten umożliwia ocenę eksperymentu, który opiera się na schemacie dwóch niezależnych pomiarów, a jego wyniki są rejestrowane na skali nominalnej co najmniej z dwiema kategoriami. Test ten dotyczy eksperymentu, w którym sprawdza się wpływ na zmienną zależną jednorodnego czynnika kontrolowanego, przy czym działaniu czynnika kontrolowanego poddaje się jedną grupę jednostek. Procedura testu obejmuje klasyfikację wyników w tablicy kontyngencji².

Tablica 11.2. Schemat klasyfikacji wyników eksperymentu w tablicy kontyngencji

Grupa	Wynik pomiaru dotyczący występowania poszczególnych kategorii zmiennej zależnej				Ogółem
	X_1	X_2	...	X_k	
Badana (1)	n_{11}	n_{12}	...	n_{1k}	$n_{1.}$
Kontrolna (2)	n_{21}	n_{22}	...	n_{2k}	$n_{2.}$
Ogółem	$n_{.1}$	$n_{.2}$	...	$n_{.k}$	n

Sprawdzianem przypuszczenia, że obie grupy charakteryzuje ten sam rozkład zmiennej zależnej (co oznacza brak wpływu czynnika kontrolowanego na zmienną zależną), jest statystyka χ^2 , którą definiuje wzór:

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}},$$

² Zaleca się, by tak klasyfikować informacje w tablicy, aby liczebności cząstkowe wynosiły co najmniej 5.

gdzie:

$$\hat{n}_{ij} = \frac{n_{i.} \cdot n_{.j}}{n},$$

przy czym i, j to numery kolejnych wierszy i kolumn w tablicy kontyngencji.

Jeśli wartość statystyki χ^2 przekracza wartość odczytaną z tablic rozkładu chi-kwadrat (por. aneks, tablica 4) dla $\alpha = 0,05$ i $\nu = (k-1)(l-1)$, gdzie k i l to odpowiednio liczba wierszy i kolumn w tablicy kontyngencji (dla tak zorganizowanego eksperymentu $k = 2$), to można uznać, że wyniki eksperymentu potwierdzają istotność wpływu zmiennej niezależnej (czynnika kontrolowanego) na nominalną zmienną zależną.

Przykład 11.2

W celu oceny wpływu informacji marketingowej na wybór miejsca zakupu przeprowadzono eksperyment rynkowy według schematu symulowanego z pomiarem pierwotnym i końcowym (por. p. 6.3). Rozkłady losowo wybranych dwóch grup respondentów według preferowanych miejsc zakupu w grupie kontrolnej (pomiar pierwotny, bez informacji marketingowej) oraz w grupie badanej (pomiar końcowy, pod wpływem informacji marketingowej) kształtowały się jak w tablicy 11.3.

Tablica 11.3. Wyniki obserwacji

Grupa	Preferowane miejsca zakupu				Ogółem
	supermarket	sklep branżowy	sklep osiedlowy	targowisko	
Badana (1)	300	120	50	230	700
Kontrolna (2)	260	240	10	90	600
Ogółem	560	360	60	320	1300

Dla wyników empirycznych otrzymujemy wartości statystyki χ^2 :

$$\begin{aligned} \chi^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} &= \frac{\left(300 - \frac{700 \cdot 560}{1300}\right)^2}{\frac{700 \cdot 560}{1300}} + \frac{\left(120 - \frac{700 \cdot 360}{1300}\right)^2}{\frac{700 \cdot 360}{1300}} + \frac{\left(50 - \frac{700 \cdot 60}{1300}\right)^2}{\frac{700 \cdot 60}{1300}} + \\ &+ \frac{\left(230 - \frac{700 \cdot 320}{1300}\right)^2}{\frac{700 \cdot 320}{1300}} + \frac{\left(260 - \frac{600 \cdot 560}{1300}\right)^2}{\frac{600 \cdot 560}{1300}} + \frac{\left(240 - \frac{600 \cdot 360}{1300}\right)^2}{\frac{600 \cdot 360}{1300}} + \frac{\left(10 - \frac{600 \cdot 60}{1300}\right)^2}{\frac{600 \cdot 60}{1300}} + \\ &+ \frac{\left(90 - \frac{600 \cdot 320}{1300}\right)^2}{\frac{600 \cdot 320}{1300}} \approx 123,8141. \end{aligned}$$

Dla $\alpha = 0,05$ i $\nu = (k-1)(l-1) = (2-1)(4-1) = 3$ wartość krytyczna statystyki χ^2 odczytana z tablic rozkładu chi-kwadrat wynosi 7,815. Wartość obliczona na podstawie wyników eksperymentu znacznie przekracza wartość krytyczną, co pozwala uznać wpływ informacji marketingowej na wybór miejsca zakupu produktu za istotny.

Jeśli zmienna zależna jest dychotomiczna, to wygodną formułą oceny wyników eksperymentu jest wskaźnik ryzyka względnego.

11.3. Wskaźnik ryzyka względnego oraz iloraz szans

Wskaźnik ryzyka względnego oraz iloraz szans pozwalają określić siłę i kierunek zależności między dwiema zmiennymi o charakterze nominalnym, dychotomicznym. Możliwe są tu dwa podejścia:

- badanie kohortowe, czyli badanie panelowe (eksperymentalne), dla którego wyznacza się wskaźnik ryzyka względnego;
- porównawcze badanie przypadków, czyli badanie retrospektywne, dla którego wyznacza się iloraz szans.

W badaniach przekrojowych badane zbiorowości konstruuje się na podstawie obserwacji zmiennej zależnej, w badaniach kohortowych zaś na podstawie obserwacji zmiennej niezależnej.

Przykład 11.3

Badanie kohortowe (panelowe)

100 losowo wybranym osobom udostępniono materiał reklamowy produktu, podczas gdy drugiej grupie 100 losowo wybranych osób reklamy nie pokazano. Spośród tych osób, którym udostępniono reklamę produktu, odnotowano 40, które w minionym tygodniu dokonały zakupu produktu, a wśród 100 losowo wybranych osób, którym nie udostępniono reklamy, odnotowano 20 przypadków zakupu produktu w tym czasie.

Badanie porównawcze przypadków

Wśród 100 losowo wybranych osób, które dokonały zakupu w minionym tygodniu, zidentyfikowano 70, które spotkały się z reklamą produktu w codziennej prasie, podczas gdy wśród 100 losowo wybranych osób, które nie dokonały zakupu w tym czasie, takich osób odnotowano 40.

Wyniki prowadzonych obserwacji klasyfikowane są w tablicy dwudzielnej w każdym z typów analizy inaczej, gdyż inaczej są definiowane grupy badana i kontrolna. W badaniach kohortowych grupy wyróżnia oddziaływanie lub brak oddziaływania czynnika kontrolowanego, zaś w porównawczych badaniach przypadków grupy wyróżnia występowanie lub brak występowania badanego zjawiska.

Tablica 11.7. Wyniki eksperymentu

Zachowania	Czynnik		Rozkład brzegowy znajomości treści rekl-
	osoby znające treść reklamy	osoby nie- znające treści reklamy	
Osoby, które dokonały zakupu	70	30	100
Osoby, które nie dokonały zakupu	40	60	100
Rozkład brzegowy decyzji o zakupie	110	90	200

Przedstawione wyniki są przykładem porównawczego badania przypadków na prób. Mogą one być podstawą oszacowania względnej możliwości (szansy) dokonania zakupu przeprowadzenia kampanii reklamowej. Zgodnie z przedstawioną regułą otrzymujemy wskaźnika z próby:

$$\hat{\psi} = \frac{a}{b} = \frac{70 \cdot 60}{30 \cdot 40} = 3,5,$$

co oznacza, że znajomość reklamy zwiększa ryzyko dokonania zakupu o 250%, gdyż ($= 250$).

W przypadku badania istotności wskaźnika ryzyka względnego, manego wyniku z próby wyznacza się przedział wartości, zgodnie przedziałowej oceny wyników z prób losowych. Dokładność oszacowania standardowym błędem szacunku, wygodnie jest analizować w logarytmicznym. Wariancję logarytmu ryzyka względnego wskaźnika w badaniu kohortowym definiuje wzór:

$$\sigma^2(\ln \hat{\psi}) = \frac{c}{a(a+c)} + \frac{d}{d(b+d)},$$

w porównawczym badaniu przypadków zaś, dla ilorazu szans, wzór:

$$\sigma^2(\ln \hat{\psi}) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}.$$

Maksymalny błąd dla logarytmu wskaźnika wynosi⁴ $1,96\sqrt{\sigma^2(\ln \hat{\psi})}$. (połowę rozpiętości przedziału dopuszczalnych wartości logarytmu ryzyka względnego, czyli przedziału:

⁴ Przy 5-procentowym ryzyku przekroczenia tej wartości.

$$\left[\ln \hat{\psi} - 1,96\sqrt{\sigma^2(\ln \hat{\psi})}, \ln \hat{\psi} + 1,96\sqrt{\sigma^2(\ln \hat{\psi})} \right].$$

Krańce tego przedziału należy następnie antylogarytmować.

Jeśli tak wyznaczony przedział nie zawiera wartości 1, to wynik uznaje się za istotny statystycznie. Jeśli zaś zawiera wartość 1, to uznaje się, że otrzymany wynik z próby jest nieistotny.

Przykład 11.6

Dla ilorazu szans wyznaczonego w przykładzie 11.5 standardowy błąd szacunku logarytmu wskaźnika wynosi:

$$\sigma^2(\ln \hat{\psi}) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} = \frac{1}{70} + \frac{1}{30} + \frac{1}{40} + \frac{1}{60} \approx 0,0893.$$

Stąd: $\sigma(\ln \hat{\psi}) \approx 0,299$.

Krańce przedziału dopuszczalnych wartości przed antylogarytmowaniem wynoszą odpowiednio:

$$\ln 3,5 - 1,96 \cdot 0,299 \approx 0,667,$$

$$\ln 3,5 + 1,96 \cdot 0,299 \approx 1,839.$$

Przedział dla ilorazu szans dokonania zakupu w wyniku kontaktu z materiałem reklamowym, po antylogarytmowaniu krańców, jest następujący:

$$[1,948, 6,289].$$

Przedział ten nie zawiera wartości 1 i tym samym wpływ reklamy na możliwość dokonania zakupu przez klienta należy uznać za istotny statystycznie. Przeprowadzone badanie wskazuje, że wzrost sprzedaży mógł być wynikiem kampanii reklamowej.

11.4. Test McNemara⁵

Test ten umożliwia ocenę eksperymentu, który opiera się na schemacie dwóch zależnych pomiarów, a jego wyniki są rejestrowane na dychotomicznej skali nominalnej. Dotyczy on eksperymentu, w którym sprawdza się wpływ na zmienną zależną jednorodnego czynnika kontrolowanego. Działaniu czynnika kontrolowanego poddaje się próbę po przeprowadzeniu pomiaru początkowego. Test ten odnosi się do schematów z podwójnym pomiarem w jednej grupie badawczej o liczebności co najmniej 20 jednostek. Wyniki pomiaru o charakterze nominalnym zestawia się w tablicy czteropolowej (tablica 11.8, gdzie a, b, c, d – liczby przypadków).

⁵ Por. S. Mynarski, *Praktyczne metody analizy danych rynkowych i marketingowych*, Zarządzanie-Ekonomia-Marketing, Kantor Wydawniczy ZAKAMYCZE, Zakamycze 2000, s. 107.

Tablica 11.8. Schemat tablicy czteropolowej zestawiającej wyniki eksperymentu

Pomiar	Badane zjawisko	
	występuje	nie występuje
Początkowy	<i>a</i>	<i>b</i>
Końcowy	<i>c</i>	<i>d</i>

Na podstawie otrzymanych wyników wyznacza się statystykę χ^2 o postaci:

$$\chi^2 = \frac{(|a - d| - 1)^2}{a + d}.$$

Jeśli wartość statystyki χ^2 przekracza wartość odczytaną z tablic rozkładu chi-kwadrat (por. aneks, tablica 4) dla $\alpha = 0,05$ i $\nu = 1$, to można uznać, że wyniki eksperymentu potwierdzają istotność wpływu zmiennej niezależnej (czynnika kontrolowanego) na nominalną zmienną zależną.

Przykład 11.7

W celu porównania zachowań nabywczych pod wpływem informacji marketingowej dokonano badania eksperymentalnego w formie schematu jednogrupowego z pomiarem początkowym i końcowym. W losowo wybranej grupie respondentów przeprowadzono ocenę zainteresowania produktem przed przekazaniem przygotowywanej informacji oraz po przekazaniu tej informacji. Wyniki eksperymentu zestawia tablica 11.9.

Tablica 11.9. Wyniki eksperymentu jednogrupowego z pomiarem początkowym i końcowym

Pomiar	Zainteresowanie zakupem	
	występuje	nie występuje
Początkowy	30	120
Końcowy	50	100

Obliczona wartość statystyki χ^2 to:

$$\chi^2 = \frac{(|a - d| - 1)^2}{a + d} = \frac{(|30 - 100| - 1)^2}{30 + 100} = 36,623.$$

Wartość odczytana z tablic rozkładu chi-kwadrat dla $\alpha = 0,05$ i $\nu = 1$ to 3,841. Znaczna przewaga wartości obliczonej nad wartością krytyczną pozwala uznać za istotny wpływ informacji marketingowej na deklarowane zachowania nabywcze.

11.5. Test Cochran⁶

Test ten umożliwia ocenę eksperymentu, który opiera się na schemacie kilku zależnych pomiarów, a jego wyniki są rejestrowane na dychotomicznej skali nominalnej. Dotyczy on eksperymentu, w którym sprawdza się wpływ na zmienną zależną czynnika kontrolowanego, który może przybierać kilka kategorii, przy czym działaniu czynnika kontrolowanego poddaje się za każdym razem tę samą grupę jednostek. Test Cochran jest uogólnieniem testu McNemara na większą liczbę pomiarów, k , realizowanych w grupie $n \geq 20$ respondentów, dla których prowadzi się test porównawczy. Przyjmujemy następujące oznaczenia: $n_1, n_2, \dots, n_i, \dots, n_k$ – liczba elementów wyróżnionych odpowiednio w pierwszym pomiarze, w drugim i w kolejnych pomiarach wśród wszystkich respondentów; $m_1, m_2, \dots, m_j, \dots, m_n$ – liczba elementów wyróżnionych odpowiednio przez pierwszego respondenta, drugiego i kolejnych respondentów w ramach wszystkich pomiarów.

Statystyka testu na równość frakcji ma postać:

$$\chi^2 = \frac{k(k-1) \left[k \sum_{i=1}^k n_i^2 - \left(\sum_{i=1}^k n_i \right)^2 \right]}{k \sum_{j=1}^n m_j - \sum_{j=1}^n m_j^2}.$$

Jeśli wartość statystyki χ^2 przekracza wartość odczytaną z tablic rozkładu chi-kwadrat (por. aneks, tablica 4) dla $\alpha = 0,05$ i $\nu = k - 1$, to można uznać, że wyniki eksperymentu potwierdzają istotność różnic między frakcjami elementów wyróżnionych w kolejnych porównywanych rozwiązaniach.

Przykład 11.8

W celu sprawdzenia, czy kampania reklamowa zmienia zainteresowanie produktem w losowo wybranej grupie 25 konsumentów, zaproponowano czterokrotnie zakup produktu: przed kampanią reklamową oraz w kolejnych trzech tygodniach prowadzenia kampanii. Liczba zainteresowanych w kolejnych pomiarach wynosiła odpowiednio: 5, 8, 9 i 9. Z kolei, obserwując zainteresowanie produktem poszczególnych respondentów w trakcie kampanii reklamowej, odnotowano następujące łączne liczby deklaracji zakupu składanych przez każdego z nich⁷:

0, 0, 0, 1, 1, 1, 1, 2, 0, 2, 1, 1, 0, 2, 3, 3, 1, 1, 1, 2, 1, 3, 3, 1, 0.

⁶ Na podstawie: Cz. Domański, *Testy statystyczne*, PWE, Warszawa 1990, s. 156–158.

⁷ Zero oznacza, że pierwszy respondent ani razu w kolejnych czterech pomiarach nie wyraził chęci nabycia produktu.

Kolejne składniki wzoru pozwalającego wyznaczyć wartość statystyki χ^2 to odpowiednio:

$$\sum_{i=1}^4 n_i = 5 + 8 + 9 + 9 = 31,$$

$$\sum_{i=1}^4 n_i^2 = 5^2 + 8^2 + 9^2 + 9^2 = 251,$$

$$\sum_{j=1}^{25} m_j = 0 + 0 + 0 + 1 + 1 + 1 + 1 + 2 + 0 + 2 + 1 + 1 + 0 + 2 + 3 + 3 + 1 + 1 + 1 + 2 + 1 + 3 + 3 + 1 + 0 = 31,$$

$$\sum_{j=1}^{25} m_j^2 = 0^2 + 0^2 + 0^2 + 1^2 + 1^2 + 1^2 + 1^2 + 2^2 + 0^2 + 2^2 + 1^2 + 1^2 + 0^2 + 2^2 + 3^2 + 3^2 + 1^2 + 0^2 + 2^2 + 3^2 + 3^2 + 1^2 + 0^2 = 63.$$

Wartość statystyki χ^2 obliczona na podstawie przeprowadzonego eksperymentu wynosi zatem:

$$\chi^2 = \frac{k(k-1) \left[k \sum_{i=1}^k n_i^2 - \left(\sum_{i=1}^k n_i \right)^2 \right]}{k \sum_{j=1}^n m_j - \sum_{j=1}^n m_j^2} = \frac{4 \cdot 3 [4 \cdot 251 - (31)^2]}{4 \cdot 31 - 63} \approx 9,197.$$

Wartość odczytana z tablic rozkładu chi-kwadrat dla $\alpha = 0,05$ i $\nu = 4 - 1 = 3$ to 7,815. Przewaga wartości obliczonej nad wartością krytyczną pozwala uznać za istotny wpływ realizowanej kampanii reklamowej na deklarowane zachowania nabywcze.

11.6. Test Walda–Wolfowitza⁸

Test ten umożliwia ocenę eksperymentu, który opiera się na schemacie dwóch niezależnych pomiarów, a jego wyniki są rejestrowane na skali co najmniej porządkowej. W eksperymencie sprawdza się wpływ na zmienną zależną jednorodnego czynnika kontrolowanego. Działaniu czynnika kontrolowanego poddaje się tylko jedną z prób. Test ten pozwala ocenić, czy wyniki niezależnych pomiarów, początkowego i końcowego, rejestrowane na skali porządkowej, wskazują na występowanie istotnego wpływu czynnika kontrolowanego na zmienną zależną. Wyniki obu pomiarów łączy się w jeden zbiór i porządkuje w kolejności niemalejącej. W utworzonym ciągu można wyodrębnić serie wyników pochodzących z pomiaru początkowego oraz serie utworzone z wyników pomiaru końcowego. Jeśli liczba serii nie jest zbyt mała, to należy uznać, że pomiędzy pomiarem początkowym i końcowym brak statystycznie istotnych różnic. W przeciwnym wypadku zrealizowany eksperyment wskazuje na istotne rozbieżności. Procedura testu serii została szczegółowo opisana w p. 5.4.

⁸ Na podstawie: *ibid.*, s. 138–140.

Przykład 11.9

W monadycznym teście produktu poproszono dwie losowo wybrane grupy po 15 gospodyń domowych (stali użytkownicy oraz użytkownicy po raz pierwszy) o ocenę na skali od 1 (ocena całkowicie negatywna) do 5 (ocena całkowicie pozytywna) jego walorów użytkowych. Otrzymano następujące wyniki.

Stali użytkownicy: 2, 3, 4, 2, 3, 1, 5, 2, 3, 2, 3, 1, 2, 3, 2.

Użytkownicy po raz pierwszy: 3, 2, 4, 4, 2, 5, 4, 3, 5, 4, 5, 2, 3, 2, 4.

Po połączeniu obu zbiorów wyników i uporządkowaniu według wartości niemalejących otrzymuje się następujący ciąg:

1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 5.

W ciągu tym występuje 8 serii, które tworzą wyniki pochodzące ze wspólnego pomiaru. Dla $n_1 = n_2 = 15$ oraz $\alpha = 0,05$ z tablic rozkładu liczby serii odczytuje się wartości krytyczne: $k_{\alpha/2} = 10$ oraz $k_{1-\alpha/2} = 20$. Zaobserwowana liczba serii kształtuje się poniżej pierwszej z odczytanych wartości, co skłania do uznania, że wyniki obu pomiarów istotnie się różnią.

W przypadku gdy liczba jednostek objętych badaniem w jednej grupie przekracza 20, należy skorzystać ze zbieżności rozkładu serii do rozkładu normalnego i posłużyć się statystyką U zdefiniowaną w p. 5.4.

11.7. Test Kołmogorowa–Smirnowa⁹

Test ten umożliwia ocenę eksperymentu, który opiera się na schemacie dwóch niezależnych pomiarów, a jego wyniki są rejestrowane na skali co najmniej porządkowej. W eksperymencie sprawdza się wpływ na zmienną zależną jednorodnego czynnika kontrolowanego. Działaniu czynnika kontrolowanego poddaje się tylko jedną z prób. Test zgodności Kołmogorowa–Smirnowa pozwala rozstrzygnąć, czy istnieje zgodność rozkładów rang nadanych przez dwie różne populacje, reprezentowane odpowiednio przez grupę kontrolną i grupę badaną. Podstawą oceny jest zakres rozbieżności pomiędzy strukturami obu grup, opisywanymi za pomocą dystrybuant, czyli w formie skumulowanej. Jako miarę rozbieżności przyjmuje się największą z zaobserwowanych różnic, czyli:

$$D = \sup_x |G_1(x) - G_2(x)|,$$

gdzie: $G_1(x)$ – wartości dystrybuanty empirycznej wyników grupy kontrolnej;
 $G_2(x)$ – wartości dystrybuanty empirycznej wyników grupy badanej.

Statystyką oceniającą zakres rozbieżności jest statystyka λ o postaci:

⁹ Por. M. Rószkiewicz, *Statystyka. Kurs podstawowy*, EFEKT, Warszawa 2002, s. 120–122.

$$\lambda = D\sqrt{n},$$

gdzie:

$$n = \frac{n_1 n_2}{n_1 + n_2},$$

przy czym: n_1 – liczebność próby kontrolnej; n_2 – liczebność grupy badanej.

Wartości krytyczne tej statystyki, świadczące o znacznych rozbieżnościach rozkładów i tym samym przeczące zgodności obu struktur, odczytuje się z tablic rozkładu λ Kołmogorowa (por. aneks, tablica 6) dla przyjętego poziomu ryzyka pomyłki (poziomu istotności) α .

Przykład 11.10

W celu oceny wpływu informacji marketingowej na zakres akceptacji marki przeprowadzono eksperyment rynkowy według schematu symulowanego z pomiarem pierwotnym i końcowym na dwóch losowo wybranych grupach respondentów. Rozkłady ocen, wyrażonych za pomocą rang w skali od 1 (całkowity brak akceptacji) do 5 (całkowita akceptacja) zestawiają tablice 11.10 i 11.11.

Tablica 11.10. Wyniki pomiaru pierwotnego

Ranga	Liczba respondentów	Odsetek respondentów
1	80	0,26667
2	130	0,43333
3	50	0,16667
4	30	0,10000
5	10	0,03333
Razem	300	1,00000

Przekształcając oba rozkłady w rozkłady skumulowane (dystrybuanty empirycznej), otrzymujemy wartości podane w tablicy 11.12.

Tablica 11.12. Wartości dystrybuanty empirycznej dla rozkładów

Rangi (x_i)	Dla grupy kontrolnej $G_1(x_i)$	Dla grupy badanej $G_2(x_i)$	$G_1(x_i) - G_2(x_i)$
1	0,26667	0,06666	0,20001
2	0,70000	0,56666	0,13334
3	0,86667	0,73333	0,13334
4	0,96667	0,90000	0,06667
5	1,00000	1,00000	0,00000

Największa co do modułu różnica to: $D = \sup |G_1(x) - G_2(x)| = 0,2001$. Tym samym statystyka:

$$\lambda = D \cdot \sqrt{n} = 0,2001 \sqrt{150} \approx 2,45, \text{ gdyż: } n = \frac{n_1 n_2}{n_1 + n_2} = \frac{300 \cdot 300}{300 + 300} = 150.$$

Porównując wartość obliczoną dla wyników empirycznych z wartością krytyczną odczytaną z tablic dla $\alpha = 0,05$, czyli z wartością $\lambda_\alpha = 1,358$, należy uznać, iż rozkłady w obu populacjach są różne, co pozwala twierdzić, że informacja marketingowa istotnie zmienia zakres akceptacji marki.

11.8. Test Kruskala–Wallisa¹⁰

Test ten umożliwia ocenę eksperymentu, który opiera się na schemacie kilku niezależnych pomiarów, a jego wyniki są rejestrowane na skali co najmniej porządkowej. W eksperymencie sprawdza się wpływ na zmienną zależną czynnika kontrolowanego, który może przybierać kilka kategorii, przy czym działaniu czynnika kontrolowanego poddaje się za każdym razem inne grupy jednostek. Zbiór wszystkich obserwacji z wszystkich prób porządkuje się, nadając coraz wyższym wartościom zmiennej zależnej kolejne rangi. Wartość statystyki testu oceniającego zgodność efektów działania czynnika kontrolowanego określa reguła:

$$\chi^2 = \frac{12}{n \cdot (n+1)} \sum_{j=1}^k \frac{\left(\sum_{i=1}^{n_j} r_{ij} \right)^2}{n_j} - 3 \cdot (n+1),$$

gdzie: k – liczba kategorii czynnika kontrolowanego; n – liczba jednostek we wszystkich próbach łącznie; n_j – liczba jednostek w j -tej próbie, dla $j = 1, \dots, k$; r_{ij} – ranga nadana i -tej obserwacji zmiennej zależnej z j -tej próby.

Jeśli wartość statystyki χ^2 przekracza wartość odczytaną z tablic rozkładu chi-kwadrat (por. aneks, tablica 4) dla $\alpha = 0,05$ i $\nu = k - 1$, to można uznać, że wyniki eksperymentu potwierdzają istotność różnic między efektami działania różnych poziomów czynnika kontrolowanego.

Przykład 11.11

W losowo wybranych 5 grupach, liczących po 19 konsumentów, przeprowadzono test monadyczny (każda grupa oceniała tylko jedną markę), oceniający walory smakowe 5 najpopularniejszych marek jogurtu. W ocenie walorów smakowych respondenci posługiwali się skalą od 1 (całkowicie niezadowolająca) do 10 (całkowicie zadowolająca). Wyniki wraz z rangami nadanymi

¹⁰ Por. Cz. Domański, *Testy statystyczne*, PWE, Warszawa 1990, s. 147–148.

ocenom wystawionym przez kolejnych respondentów zestawia tablica 11.13. Dane zostały posortowane, gdyż łatwiej jest wówczas nadać rangi kolejnym wartościom.

Tablica 11.13. Wyniki testu monadycznego w 5 grupach konsumentów

Oceniana marka									
A		B		C		D		E	
wynik	ranga	wynik	ranga	wynik	ranga	wynik	ranga	wynik	ranga
1	3,5	2	15,5	4	40,5	1	3,5	5	53
1	3,5	2	15,5	4	40,5	1	3,5	5	53
1	3,5	3	31	5	53	1	3,5	5	53
2	15,5	3	31	5	53	2	15,5	5	53
2	15,5	3	31	5	53	2	15,5	5	53
2	15,5	3	31	5	53	2	15,5	6	67
2	15,5	4	40,5	5	53	2	15,5	6	67
2	15,5	5	53	5	53	2	15,5	6	67
2	15,5	5	53	6	67	2	15,5	7	76,5
3	31	5	53	6	67	2	15,5	7	76,5
3	31	5	53	6	67	2	15,5	7	76,5
3	31	6	67	6	67	2	15,5	7	76,5
3	31	6	67	7	76,5	2	15,5	8	87
3	31	7	76,5	7	76,5	3	31	8	87
3	31	7	76,5	7	76,5	3	31	8	87
4	40,5	7	76,5	8	87	3	31	8	87
4	40,5	8	87	8	87	4	40,5	8	87
5	53	8	87	8	87	5	53	8	87
5	53	9	94	9	94	5	53	9	94
Razem	476,5	×	1039	×	1251,5	×	405	×	1388

Wartość statystyki χ^2 obliczona na podstawie przeprowadzonego eksperymentu wynosi:

$$\chi^2 = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{\left(\sum_{i=1}^{n_j} r_{ij} \right)^2}{n_j} - 3(n+1) =$$

$$= \frac{12}{95(95+1)} \left[\frac{(476,5)^2}{19} + \frac{(1039)^2}{19} + \frac{(1251,5)^2}{19} + \frac{(405)^2}{19} + \frac{(1388)^2}{19} \right] - 3(95+1) = 55,73.$$

Wartość odczytana z tablic rozkładu chi-kwadrat dla $\alpha = 0,05$ i $\nu = k - 1 = 4$ to 9,488. Wartość obliczona przekracza wartości krytyczną, co wskazuje na istotność różnic w postrzeganiu walorów smakowych najpopularniejszych na rynku marek jogurtów.

11.9. Test Friedmana¹¹

Test ten umożliwia ocenę eksperymentu, który opiera się na schemacie kilku zależnych pomiarów, a wyniki eksperymentu są rejestrowane na skali co najmniej porządkowej. W eksperymencie sprawdza się wpływ na zmienną zależną czynnika kontrolowanego, który może przybierać kilka kategorii, przy czym działaniu czynnika kontrolowanego poddaje się za każdym razem te same jednostki. W przyjętym schemacie eksperymentu następuje zatem wielokrotny pomiar tych samych jednostek. Są to pomiary zależne. Wyniki pomiarów rejestrowane dla każdej jednostki porządkuje się w kolejności niemalejącej i nadaje im kolejne rangi. Wartość statystyki testu, oceniającego zgodność efektów działania czynnika kontrolowanego, określa reguła:

$$\chi^2 = \frac{12}{k(k+1)n} \sum_{j=1}^k \left(\sum_{i=1}^n r_{ij} \right)^2 - 3(k+1)n,$$

gdzie: k – liczba kategorii czynnika kontrolowanego; n – liczba jednostek w próbie; r_{ij} – ranga nadana j -tej obserwacji zmiennej zależnej u i -tej jednostki, dla kolejnych $i = 1, \dots, n$, poprzez wszystkie $j = 1, \dots, k$.

Jeśli wartość statystyki χ^2 przekracza wartość odczytaną z tablic rozkładu chi-kwadrat (por. aneks, tablica 4) dla $\alpha = 0,05$ i $\nu = k - 1$, to można uznać, że wyniki eksperymentu potwierdzają istotność różnic między efektami działania różnych poziomów czynnika kontrolowanego.

Przykład 11.12

W losowo wybranej grupie 20 konsumentów przeprowadzono test porównawczy dotyczący walorów smakowych 5 najpopularniejszych marek jogurtu. W ocenie walorów smakowych poszczególnych marek respondenci posługiwali się skalą od 1 (całkowicie niezadowolająca) do 10 (całkowicie zadowolająca). Wyniki, wraz z rangami nadanymi ocenom wystawionym przez każdego respondenta oddzielnie (w wierszach), zestawia tablica 11.14.

Wartość statystyki χ^2 obliczona na podstawie przeprowadzonego eksperymentu wynosi:

$$\chi^2 = \frac{12}{k(k+1)n} \sum_{j=1}^k \left(\sum_{i=1}^n r_{ij} \right)^2 - 3(k+1)n =$$

$$= \frac{12}{5(5+1)20} [(35)^2 + (65,5)^2 + (73)^2 + (29,5)^2 + (82)^2] - 3(5+1)20 = 8,77.$$

¹¹ Por. *ibid.*, s. 162.

Tablica 11.14. Wyniki testu porównawczego

Respondent	Oceniana marka									
	A		B		C		D		E	
	wynik	ranga	wynik	ranga	wynik	ranga	wynik	ranga	wynik	ranga
1	3	3	2	1,5	5	4	2	1,5	6	5
2	2	1	4	3	6	4	3	2	8	5
3	4	2	5	3,5	5	3,5	2	1	8	5
4	3	2,5	3	2,5	4	4	2	1	7	5
5	5	3,5	3	2	8	5	1	1	5	3,5
6	2	2,5	2	2,5	7	5	1	1	6	4
7	3	2,5	3	2,5	7	4	2	1	9	5
8	1	1	3	2	5	3,5	5	3,5	8	5
9	4	2	5	3	6	4	2	1	8	5
10	2	1,5	6	4	9	5	2	1,5	5	3
11	2	1	9	5	8	4	3	2	5	3
12	3	1,5	8	5	5	3	3	1,5	7	4
13	5	3	5	3	5	3	2	1	7	5
15	2	1,5	5	4,5	4	3	2	1,5	5	4,5
16	3	2	8	5	7	4	2	1	6	3
17	3	1	7	4	5	3	4	2	8	5
18	1	1,5	7	4	6	3	1	1,5	8	5
19	1	1	7	4,5	6	3	2	2	7	4,5
20	2	1	6	4	8	5	5	2,5	5	2,5
Razem	×	35,0	×	65,5	×	73,0	×	29,5	×	82,0

Wartość odczytana z tablic rozkładu chi-kwadrat dla $\alpha = 0,05$ i $k - 1 = 4$ to 9,448. Wartość obliczona nie przekracza wartości krytycznej, co wskazuje na brak istotnych różnic w postrzeganiu walorów smakowych najpopularniejszych na rynku marek jogurtów.

11.10. Test U lub t dla dwóch średnich

Test ten umożliwia ocenę eksperymentu, który opiera się na schemacie dwóch pomiarów, a jego wyniki są rejestrowane na skali przedziałowej lub ilorazowej. W eksperymencie sprawdza się wpływ na zmienną zależną jednorodnego czynnika kontrolowanego. W zależności od sposobu zorganizowania eksperymentu jego analiza może przebiegać dwiema drogami.

1. Jeśli działaniu czynnika kontrolowanego poddaje się jedną z dwóch grup jednostek, czyli w eksperymencie występuje grupa badana i grupa kontrolna, to obowiązuje formuła testu dla pomiarów niezależnych. Dla wartości obserwowanych w pomiarze początkowym oraz końcowym wyznacza się wartości średnie oraz odchylenia standardowe. Postać statystyki testu uzależniona jest od warunków dodatkowych badania, czyli od zakresu wiedzy o zmiennej zależnej:

a. Jeśli zmienna zależna ma rozkład normalny o znanych odchyleniach standardowych, odpowiednio w grupie badanej σ_1 i w grupie kontrolnej σ_2 (znajomość kształtu rozkładu oraz wartości odchylenia standardowego może pochodzić z wcześniejszych badań tej zbiorowości), to obowiązuje następująca postać statystyki testu:

$$U = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$

gdzie: $\bar{x}_1, \bar{x}_2$ – wartości średnie zmiennej zależnej odpowiednio w grupie badanej i kontrolnej; σ_1, σ_2 – odchylenia standardowe zmiennej zależnej w populacjach reprezentowanych odpowiednio przez grupę badaną i grupę kontrolną; n_1, n_2 – liczba jednostek odpowiednio w grupie badanej i grupie kontrolnej.

Jeśli przypuszczenie o braku wpływu czynnika kontrolowanego na zmienną zależną jest prawdziwe, to wartość bezwzględna statystyki U powinna być mała, tzn. nie powinna przekraczać wartości krytycznej odczytanej z tablic dystrybucji rozkładu normalnego dla $F(u_\alpha) = 1 - \alpha / 2$, gdzie α to ryzyko pomyłki (poziom istotności), na ogół wynoszące¹² 0,05.

b. Jeśli zmienna zależna ma rozkład normalny o nieznanach odchyleniach standardowych, w grupie badanej σ_1 i w grupie kontrolnej σ_2 , to przy wnioskowaniu na podstawie tzw. „małych” prób ($n < 30$) obowiązuje następująca postać statystyki testu:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}},$$

gdzie: S_1^2, S_2^2 – wariancje zmiennej zależnej odpowiednio w grupie badanej i grupie kontrolnej.

Jeśli przypuszczenie o braku wpływu czynnika kontrolowanego na zmienną zależną jest prawdziwe, to wartość bezwzględna statystyki t powinna być mała, tzn. nie powinna przekraczać wartości krytycznej odczytanej z tablic wartości krytycznych w rozkładzie t Studenta dla ustalonego ryzyka pomyłki (poziomu istotności) α i $\nu = n_1 + n_2 - 2$.

c. Jeśli zmienna zależna ma dowolny rozkład o nieznanach odchyleniach standardowych, w grupie badanej σ_1 i w grupie kontrolnej σ_2 , to przy wnioskowaniu

¹² Wówczas $u_\alpha = 1,96$, co oznacza, że wartość statystyki U obliczona według powyższego wzoru nie powinna przekroczyć liczby 1,96.

waniu na podstawie tzw. „dużych” prób ($n > 30$) obowiązuje następująca postać statystyki testu:

$$U = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}.$$

Jeśli przypuszczenie o braku wpływu czynnika kontrolowanego na zmienną zależną jest prawdziwe, to wartość bezwzględna statystyki U powinna być mała, tzn. nie powinna przekraczać wartości krytycznej odczytanej z tablic dystrybucyjny rozkładu normalnego dla $F(u_\alpha) = 1 - \alpha / 2$, gdzie α to ryzyko pomyłki (poziom istotności).

Przykład 11.13

Wiadomo, że przy ustalonej cenie rozkład wielkości miesięcznych zakupów badanego produktu jest normalny. Aby ocenić wpływ zmiany ceny na zachowania nabywcze konsumentów w dwóch losowo wybranych grupach konsumentów, przeprowadzono eksperyment, jednej z nich proponując produkt o nowej cenie jednostkowej. W grupie kontrolnej, której oferowano produkt po starej cenie, uzyskano dla 10 konsumentów średnią wielkość zakupu w okresie miesiąca na poziomie $\bar{x}_1 = 11,04$ szt., z przeciętnym zróżnicowaniem 0,45 szt. W grupie badanej zaś, której proponowano produkt po nowej cenie jednostkowej, średnia ta dla 15 konsumentów wyniosła $\bar{x}_2 = 11,15$ szt., z przeciętnym zróżnicowaniem 0,35 szt.

Brak informacji o odchyleniu standardowym wielkości miesięcznego zakupu w populacjach, które reprezentowały obie grupy, oraz mała ich liczebność sprawiają, że procedura jest realizowana według wariantu b, czyli:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{11,04 - 11,15}{\sqrt{\frac{10(0,45)^2 + 15(0,35)^2}{10 + 15 - 2} \left(\frac{1}{10} + \frac{1}{15} \right)}} = -0,66.$$

Dla $\alpha = 0,05$ oraz $\nu = 10 + 15 - 2 = 23$ wartość krytyczna w rozkładzie t Studenta wynosi 2,0687. Ponieważ wartość bezwzględna statystyki t obliczona na podstawie wyników eksperymentu nie przekracza wartości krytycznej, nie ma podstaw, by uznać, że zmiana ceny istotnie zmienia zachowania nabywcze konsumentów.

2. Jeśli działaniu czynnika kontrolowanego poddaje się jedną grupę jednostek, uprzednio przeprowadziwszy w niej pomiar początkowy, czyli w eksperymencie występuje tylko jedna grupa, to wówczas obowiązuje formuła testu dla pomiarów zależnych. Z obserwowanych w pomiarach początkowym oraz końcowym wartości zmiennej zależnej wyznacza się różnice d_i (dla $i = 1, \dots, n_d$) między wynikami pomiarów dla każdej jednostki oddzielnie, a następnie oblicza się wartość średnią oraz odchylenie standardowe otrzymanych różnic. Po-

stać statystyki testu uzależniona jest od liczby jednostek objętych badaniem. Dla małej próby ($n < 30$) statystyka testu przyjmuje postać:

$$t = \frac{\bar{d}}{S_d} \sqrt{n_d},$$

gdzie: n_d – liczba obliczonych różnic; $\bar{d}$ – średnia z obliczonych różnic:

$$\bar{d} = \frac{1}{n_d} \sum_{i=1}^{n_d} d_i;$$

S_d – odchylenie standardowe obliczonych różnic:

$$S_d = \sqrt{\frac{1}{n_d - 1} \sum_{i=1}^{n_d} (d_i - \bar{d})^2}.$$

Dla dużej grupy objętej badaniem korzysta się ze zbieżności rozkładu t Studenta do rozkładu normalnego i według powyższego wzoru określa się wartość statystyki U .

Jeśli tak obliczona wartość bezwzględna statystyki testu przekracza wartość krytyczną odczytaną z tablic rozkładu t Studenta przy ustalonym poziomie α ryzyka pomyłki (poziomie istotności) oraz dla $\nu = n_d - 1$, to wpływ czynnika kontrolowanego na poziom zmiennej zależnej należy uznać za istotny.

Przykład 11.14

W teście mającym na celu porównanie walorów smakowych jogurtu przedstawiono pięciu losowo wybranym konsumentom ofertę dwóch marek. Respondenci wyrażali swoje poglądy na skali przedziałowej od 0 do 100. Wyniki wraz z wyznaczonymi różnicami ocen zestawia tablica 11.15.

Tablica 11.15. Wyniki testu sekwencyjnego dla dwóch marek jogurtów

Respondent	Ocena marki A	Ocena marki B	Różnica ocen d
1	73	68	5
2	85	79	6
3	64	60	4
4	90	84	6
5	69	65	4

Wartość statystyki testu:

$$t = \frac{\bar{d}}{S_d} \sqrt{n_d} = \frac{5}{0,89} \sqrt{5} \approx 12,56.$$

Dla $\alpha = 0,05$ oraz $\nu = 5 - 1 = 4$ wartość krytyczna w rozkładzie t Studenta wynosi 2,7765. Ponieważ wartość bezwzględna statystyki t obliczona na podstawie wyników eksperymentu przekracza wartość krytyczną, są podstawy, by uznać, że testowane marki jogurtów są różnie postrzegane pod względem walorów smakowych.

11.11. Analiza wariancji

Test analizy wariancji umożliwia ocenę eksperymentu, który opiera się na schemacie kilku niezależnych pomiarów, a jego wyniki są rejestrowane na skali przedziałowej lub ilorazowej. W eksperymencie sprawdza się wpływ na zmienną zależną czynnika kontrolowanego, który może przybierać kilka kategorii, przy czym działaniu czynnika kontrolowanego poddaje się za każdym razem inną grupę jednostek. W przyjętym schemacie eksperymentu następuje zatem kilka pomiarów niezależnych.

Ocena wyników eksperymentu za pomocą analizy wariancji polega na ocenie oddziaływania czynnika kontrolowanego na rozkład zmiennej zależnej o charakterze przedziałowym lub ilorazowym. Jeśli eksperyment polega na kontrolowanym oddziaływaniu tylko jednego czynnika, to jako metodę oceny efektów eksperymentu stosuje się analizę wariancji w klasyfikacji pojedynczej. Jeśli zaś eksperyment polega na kontrolowanym oddziaływaniu większej liczby czynników, to narzędziem oceny efektów takiego eksperymentu może być analiza wariancji w klasyfikacji wielorakiej.

Przykład 11.15

W firmie oferującej soki owocowe na rynki lokalne w celu zwiększenia wielkości sprzedaży podjęto próbę ustalenia, czy kolor opakowania soku ma wpływ na wielkość sprzedaży. W ustalonym dniu w wybranych 21 punktach sprzedaży zorganizowano eksperyment polegający na sprzedaży soku w opakowaniu ustalonego koloru: w 7 punktach sprzedawano soki tylko w opakowaniach niebieskich, w 7 punktach sprzedawano soki tylko w opakowaniach zielonych, a w 7 punktach sprzedawano soki tylko w opakowaniach białych. Zaobserwowane ilości sprzedanych opakowań przedstawia tabela 11.16.

Tabela 11.16. Wyniki eksperymentu

Kolor opakowania	Wielkość sprzedaży w poszczególnych punktach						
Zielony	18	22	22	23	22	23	19
Niebieski	20	18	19	21	20	20	18
Biały	10	11	12	11	11	11	10

W przeprowadzonym eksperymencie obserwowano oddziaływanie na wielkość sprzedaży tylko jednego czynnika, którym był kolor opakowania towaru. Ocena efektów tego eksperymentu pole-

gała na określeniu, czy istotnie kolor opakowania ma wpływ na wielkość sprzedaży. Ocena ta może być przeprowadzona metodą analizy wariancji w klasyfikacji pojedynczej.

Przykład 11.16

W celu ustalenia, czy forma reklamy powinna być dostosowana do segmentów geograficznych rynku, zorganizowano eksperyment polegający na przeprowadzeniu w trzech regionach kolejno czterech form reklamy: ulicznej, prasowej, telewizyjnej i radiowej. Po zakończeniu kampanii reklamowych obserwowano wielkość sprzedaży w trzech miastach każdego regionu. Ustalono następujące wyniki eksperymentu.

Tabela 11.17. Wyniki eksperymentu

Region	Reklama uliczna			Reklama prasowa			Reklama telewizyjna			Reklama radiowa		
1	60	50	50	80	90	70	100	110	110	150	100	110
2	20	50	30	40	30	10	80	70	60	140	100	90
3	20	20	10	30	30	10	60	40	20	120	90	70

W przeprowadzonym eksperymencie obserwowano wpływ na wielkość sprzedaży dwóch czynników: lokalizacji sprzedaży oraz formy reklamy poprzedzającej sprzedaż. Ocena efektów tego eksperymentu polega na określeniu, czy wpływ rodzaju reklamy na wielkość sprzedaży jest zróżnicowany regionalnie. Ocena ta może być przeprowadzona metodą analizy wariancji w klasyfikacji podwójnej.

Ocena wyników eksperymentu w przypadku działania różnorodnych czynników polega na porównaniu rozkładów zmiennej zależnej w grupach warunkowanych oddziaływaniem kontrolowanych czynników. Jeśli obserwowane wartości zmiennej zależnej zostaną oznaczone jako x_{ij} , to i ($i = 1, \dots, k$) określa rodzaj kategorii czynnika kontrolowanego i oznacza przynależność jednostki do i -tej grupy, będącej pod wpływem działania i -tej kategorii czynnika kontrolowanego, zaś j ($j = 1, \dots, n_i$) określa numer jednostki w i -tej grupie.

Porównanie rozkładów w utworzonych grupach dotyczy ustalenia, czy wartości średnie tych rozkładów są jednakowe, co sugerowałoby brak różnic w kształtowaniu się zmiennej zależnej pod wpływem zmiennego oddziaływania kontrolowanych czynników. Gdyby zaś wartości średnie różniły się istotnie, to można by sądzić, że kontrolowany czynnik ma wpływ na występowanie zmiennej zależnej.

W analizie wariancji przyjmuje się dwa założenia¹³. Po pierwsze, uznaje się, że w każdej z k grup objętych badaniem, będących pod wpływem poszczególnych

¹³ Por. J. Steczkowski, A. Zeliaś, *Analiza wariancji i kowariancji w badaniach ekonomicznych*, PWN, Warszawa 1983, s. 98–112.

kategorii czynnika kontrolowanego, zmienna zależna ma rozkład normalny $N(\mu_i, \sigma_i)$ dla $i = 1, 2, \dots, k$. Po drugie, przyjmuje się, że wariancje w tych rozkładach są identyczne, czyli zakłada się, że zachodzi równość: $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$. Założenie dotyczące równości wariancji może być weryfikowane m.in. testem Bartletta, testem Cochran-Coxa, testem Levene'a lub testem Hartleya¹⁴.

W teście Bartletta ocenie podlega wartość statystyki chi-kwadrat, wyznaczona według wzoru:

$$\chi^2 = \frac{2,303}{c} [(n-k) \log \frac{\sum_{i=1}^k (n_i - 1) \hat{S}_i^2}{n-k} - \sum_{i=1}^k (n_i - 1) \log \hat{S}_i^2].$$

Niezbędne są przy tym następujące obliczenia, przeprowadzone dla uzyskanych wyników eksperymentu:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij},$$

gdzie: $\bar{x}_i$ – wartość średnia zmiennej zależnej w i -tej grupie, określonej przez i -tą kategorię czynnika kontrolowanego;

$$\hat{S}_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2,$$

gdzie: $\hat{S}_i^2$ – wariancja zmiennej zależnej w i -tej grupie, określonej przez i -tą kategorię czynnika kontrolowanego;

$$c = 1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{n-k} \right),$$

gdzie: c – parametr testu Bartletta; $n = \sum_{i=1}^k n_i$ – liczba wszystkich obserwacji, k – liczba kategorii czynnika kontrolowanego i tym samym liczba badanych grup.

Statystyka χ^2 ma rozkład chi-kwadrat i jeśli założenie o równości wariancji jest spełnione, to powinna przyjmować małe wartości. Jeśli zatem wartość tej statystyki nie przekracza wartości χ_{α}^2 , odczytanej z tablic rozkładu chi-kwadrat dla $\nu = k - 1$, to nie ma podstaw, by twierdzić, że założenie to nie jest spełnione.

Pozytywne rozstrzygnięcie testu o równości wariancji umożliwia przeprowadzenie procedury analizy wariancji, weryfikującej równość średnich w roz-

kładach warunkowanych kategoriami czynników kontrolowanych. Natomiast negatywne rozstrzygnięcie testu o równości wariancji uniemożliwia przeprowadzenie tej procedury. Porównanie rozkładów zmiennej zależnej między grupami, wyznaczonymi przez działanie poszczególnych kategorii czynnika kontrolowanego, jest wówczas możliwe za pomocą testu Kruskala-Wallisa (por. p. 11.8).

11.11.1. Analiza wariancji w klasyfikacji pojedynczej¹⁵

W przypadku gdy w eksperymencie występuje tylko jeden czynnik kontrolowany, o liczbie kategorii k , do oceny istotności jego wpływu na przebieg badanego zjawiska można wykorzystać metodę analizy wariancji w klasyfikacji pojedynczej.

Dla uzyskanych wyników eksperymentu definiuje się następujące wielkości:

$$q = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2,$$

gdzie: q – zmienność całkowita zmiennej zależnej; $\bar{x}$ – średnia ogólna zmiennej zależnej, wyznaczona dla wszystkich wyników eksperymentu;

$$q_G = \sum_{i=1}^k (\bar{x}_i - \bar{x})^2 \cdot n_i,$$

gdzie: q_G – zmienność międzygrupowa, opisująca różnicowanie między grupami, wyznaczonymi przez działanie kolejnych kategorii czynnika kontrolowanego;

$$q_R = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2,$$

gdzie: q_R – zmienność wewnątrzgrupowa (zwana również zmiennością resztową), opisująca różnicowanie wewnątrz grup wyznaczonych przez kolejne kategorie czynnika kontrolowanego.

Zachodzi następująca tożsamość:

$$q = q_G + q_R.$$

Jeśli oddziaływanie czynnika kontrolowanego na zmienną zależną nie występuje, to rozkłady tej zmiennej warunkowane czynnikiem kontrolowanym nie powinny różnić się bardzo między sobą, tzn. wartości średnie tych rozkładów po-

¹⁴ Testy te są szczegółowo omówione w: *ibid.*, s. 98–113.

¹⁵ Na podstawie: *ibid.*, s. 134–146.

winny być zbliżone lub równe. W takim przypadku zmienność międzygrupowa powinna być mniejsza niż zmienność wewnątrzgrupowa. Im wielkość q_G będzie mniejsza od wielkości q_R , tym mocniejsze będą podstawy, by twierdzić, że czynnik kontrolowany nie ma wpływu na zmienną zależną. W przeciwnym wypadku, gdy wielkość q_G będzie większa od wielkości q_R , można sądzić, że zależność ta istnieje. Porównanie wielkości q_G i q_R rozstrzyga o ocenie eksperymentu.

Statystyka F , zestawiająca wielkości q_G i q_R , wyznaczona według poniższego wzoru, ma rozkład F Snedecora o $(k-1)$ i $(n-k)$ stopniach swobody i powinna przyjmować niewielkie wartości.

$$F = \frac{\sum_{i=1}^k (\bar{x}_i - \bar{x})^2 \cdot n_i}{\sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2} \cdot \frac{n-k}{k-1}.$$

Jeśli zatem obliczona wartość przekracza wartość odczytaną z tablic rozkładu F Snedecora dla $\alpha = 0,05$ i $\nu_1 = k-1$ oraz $\nu_2 = n-k$, to wówczas wynik eksperymentu można uznać za istotny statystycznie.

Przykład 11.17

Na podstawie eksperymentu opisanego w przykładzie 11.15 należy sprawdzić, czy kolor opakowania ma wpływ na wielkość sprzedaży soków oferowanych przez firmę. Najpierw należy sprawdzić, czy zachowane jest założenie o równości wariancji rozkładów wielkości sprzedaży w podgrupach określonych przez kolory opakowań. Jeśli założenie o identyczności wariancji grupowych będzie spełnione, można przeprowadzić badanie wartości średnich grupowych za pomocą testu analizy wariancji.

Dla przedstawionych danych otrzymuje się następujące wartości:

$$\bar{x}_z \approx 21,28, \hat{S}_z^2 \approx 3,90,$$

$$\bar{x}_n \approx 19,43, \hat{S}_n^2 \approx 1,28,$$

$$\bar{x}_b \approx 10,86, \hat{S}_b^2 \approx 0,48.$$

Przyjmując $n = 21$ i $k = 3$, otrzymuje się:

$$c = 1 + \frac{1}{3(3-1)} \left(\frac{1}{7-1} + \frac{1}{7-1} + \frac{1}{7-1} - \frac{1}{21-3} \right) \approx 1,074074,$$

$$\chi^2 = \frac{2,303}{1,074074} \left\{ (21-3) \log \frac{(7-1) \cdot 3,90 + (7-1) \cdot 1,28 + (7-1) \cdot 0,48}{21-3} \right. \\ \left. - \left[(7-1) \log 3,90 + (7-1) \log 1,28 + (7-1) \log 0,48 \right] \right\} = 5,7895.$$

Dla $\alpha = 0,05$ oraz $\nu = 3 - 1$ wartością krytyczną odczytaną z tablic jest liczba 5,991. Obliczona wartość nie przekracza wartości krytycznej, co pozwala uznać, że założenie o równości wariancji wielkości sprzedaży w podgrupach utworzonych przez kolory opakowań jest spełnione.

Następnie weryfikacji podlega przypuszczenie o równości średnich, które jest głównym zagadnieniem w ocenie wyników eksperymentu. Dla otrzymanych wyników eksperymentu wyznaczono kolejno: $q_G \approx 433,2381$, $q_R \approx 34,0000$,

$$F = \frac{q_G}{q_R} \cdot \frac{(n-k)}{(k-1)} = \frac{433,2381}{34} \cdot \frac{18}{2} \approx 114,681.$$

Wartość statystyki F przekracza wartość 3,49, odczytaną z tablic dla $\alpha = 0,05$ oraz $\nu_1 = 2$ i $\nu_2 = 18$, zatem przeprowadzone testy pozwalają uznać, że kolor opakowania istotnie wpływa na wielkość sprzedaży soków oferowanych przez firmę.

Porównanie średnich w poszczególnych grupach pozwala wskazać, które kolory opakowania istotnie charakteryzują inne poziomy sprzedaży. Procedura ta, zwana analizą *post hoc*, może przebiegać według testu wielokrotnych porównań Tukeya.

W teście wielokrotnych porównań Tukeya ocenie podlegają różnice między wszystkimi możliwymi parami wartości średnich, które odnotowano dla poszczególnych grup. Za istotne uznaje się tylko te różnice, które przekraczają wartość krytyczną wyznaczoną według wzoru:

$$\text{limit} = q_{\alpha; n-k, k} \sqrt{\frac{q_R}{n_i}},$$

gdzie: $q_{\alpha; n-k, k}$ – wartość krytyczna dla testu Tukeya, odczytana z tablic statystycznych dla $\nu = n - k$ i k (por. aneks, tablica 9); n_i – liczebności w podgrupach, co do których wymaga się, by były jednakowe¹⁶.

Przykład 11.18

Dla wyników testu analizy wariancji dotyczących wpływu koloru opakowania na wielkość sprzedaży otrzymuje się:

$$\text{limit} = q_{\alpha; \nu, k} \sqrt{\frac{q_R}{n_i}} = 3,61 \sqrt{\frac{34}{18}} = 4,96.$$

Oto wszystkie możliwe porównania średnich:

¹⁶ Jeśli warunek ten nie jest spełniony, za n_i przyjmuje się najmniejszą z odnotowanych liczebności grup objętych badaniem.