

POLITECHNIKA ŁÓDZKA

TOMASZ W. WOJTATOWICZ

**METODY ANALIZY DANYCH
DOŚWIADCZALNYCH**

Wybrane zagadnienia

ŁÓDŹ 1998

Przedślowie

Specyfiką teorii pomiarów jest jej wtórny charakter w stosunku do metod badawczych stosowanych w różnych naukach, a szczególnie w fizyce. Różnorodność zadań, pomysłowość i subtelność metod i technik badawczych, powoduje, że nie ma i nie może być jednej, uniwersalnej teorii pomiarów. Podręcznik ten ma za zadanie zapoznanie Czytelnika z wybranymi zagadnieniami, które może napotkać na swej drodze eksperymentatora, nie wyczerpuje jednak wszystkich znanych metod i technik opracowywania danych doświadczalnych.

Autor starał się zilustrować przykładami każde zagadnienie, nie wszystkie przykłady dotyczą bezpośrednio eksperymentów fizycznych, można w tym podręczniku znaleźć przykłady z zakresu inżynierii i technologii, a także biologii i agrotechniki, daje to przy okazji informację, w jak wielu dziedzinach stosowane są te same techniki analizy danych.

Tak jak w wielu innych podręcznikach, tak i w tym, Czytelnik spotka terminologię i interpretację niektórych pojęć i metod odbiegającą od tej, występującej np. w literaturze podanej na końcu podręcznika, wynika to z wielu względów, głównie z powodu wprowadzania metod i terminologii oddzielnie w każdej z dziedzin i opracowywania terminologii na podstawie tłumaczeń literatury z wielu języków obcych. Autor starał się podać terminy stosowane w różnych dziedzinach oraz angielskie odpowiedniki naszych terminów i ich skróty (często stosowane w angielskojęzycznych programach matematycznych i statystycznych).

1. WPROWADZENIE

- *wnioskowanie w nauce: dedukcja i indukcja,*
- *projektowanie i organizacja badań naukowych,*
- *logika eksperymentu,*
- *zdrowy rozsądek w eksperymentach.*

1.1. Wnioskowanie dedukcyjne i indukcyjne.

Wnioskowanie dedukcyjne polega na rozumowaniu od stwierdzeń ogólnych do szczególnych. Tzn. mamy kilka ogólnych praw, lub tylko jedno, a za zadanie mamy określić, co wydarzy się w przypadku wystąpienia pewnego specyficznego zestawu warunków. Lepiej uzmysłowią nam to następujące przykłady:

- Mamy wzór na pole powierzchni koła $S = \pi r^2$, jakie jest pole powierzchni koła o promieniu 2 cm?
- Mamy prawa Boyle'a-Mariotta i Charlesa dotyczące przemian gazu doskonałego, i na ich podstawie powinniśmy określić, jak będzie zmieniała się objętość gazu podczas zmian ciśnienia i temperatury tego gazu?
- Mamy zestaw parametrów fizyko-chemicznych charakteryzujących półprzewodniki oraz kilka próbek nieznanymi materiałami, naszym zadaniem jest określenie, które z nich są półprzewodnikami.

Prawie wszystkie problemy, z którymi macie Państwo do czynienia podczas studiów, są tego typu (dobrze oparte na podstawach naukowych). Oznacza to także, że powinni Państwo posiadać duży zapas wiadomości podstawowych i łatwość wnioskowania dedukcyjnego o szczególnych przypadkach.

Drugi sposób wnioskowania jest przeciwny do tego. Mamy zespół poszczególnych przypadków i na tej podstawie mamy odpowiedzieć na pytanie o ogólne prawo rządzące tymi przypadkami (wszystkimi elementami klasy wydarzeń do których należą rozpatrywane przypadki). Wnioskowanie od szczegółu do ogółu nazywane jest wnioskowaniem indukcyjnym. Wyjaśnię to na kilku przykładach:

- Mamy pola powierzchni oraz promienie kilku kół, na tej podstawie mamy napisać ogólny wzór na pole powierzchni koła o dowolnym promieniu.
- Mamy serię obserwacji objętości gazu doskonałego w różnych warunkach temperatury i ciśnienia, oczywiście powinniśmy na tej podstawie podać równanie stanu gazu doskonałego.

Zwróćmy przy tym uwagę na to, że wszystkie te problemy opierają się na grupach obserwacji (pomiarów). Choć czasem wnioskowanie odbywa się w oparciu o obserwację zjawiska bezpośrednio występującego w naturze (bez opisu ilościowego). Na przykład: mamy kilka próbek nieznanymi gatunków chwastów, na tej podstawie mamy określić cechy charakterystyczne klasy chwastów i ich położenie w systematyce roślin.

Zwykle wszystkie obserwacje wykonuje się w ściśle określonych warunkach, natomiast badane parametry zmienia się w uporządkowany sposób, poprzez odpowiednie oddziaływania. Inne parametry mogące wpływać na obserwacje są praktycznie minimalizowane. Czynność tę nazywamy eksperymentem. Jakikolwiek eksperyment zaprojektujemy, jego celem jest przeprowadzenie obserwacji (pobranie próbek losowych), które mogą być użyte do posiadającego pozory prawdopodobieństwa uogólnienia na temat badanego zjawiska. Dokonywanie takich uogólnień jest typowym zadaniem wnioskowania indukcyjnego.

Nie należy jednak odnieść wrażenia, że wnioskowanie indukcyjne stanowi zupełnie niezależny od wnioskowania dedukcyjnego sposób myślenia. Bowiem, wnioski wynikające z indukcji muszą być zawsze weryfikowane przy pomocy precyzyjnych metod dedukcyjnych. Jeżeli uważnie przyjrzymy się przykładom, zauważymy pomiędzy nimi dosyć istotne różnice.

W przypadku obliczania pola powierzchni koła nie ma niepewności co do wyniku. Dla każdego danego promienia jest tylko jedna odpowiedź (o ile nie weźmiemy pod uwagę dokładności wyznaczenia liczby π).

Inny charakter ma problem rzutu monetą. Podstawowym założeniem jest, że moneta jest jednorodna, a więc prawdopodobieństwa wyrzucenia orła i reszki są jednakowe. Wówczas wynik pojedynczego rzutu jest niemożliwy do przewidzenia, oba wyniki mogą być otrzymane z prawdopodobieństwem równym 0,5. Gdy postawimy sobie pytanie co się stanie jeżeli rzucimy monetą 10 razy, jeszcze trudniej jest dać jednoznaczną odpowiedź, ponieważ mamy wówczas aż 11 możliwości o zróżnicowanym prawdopodobieństwie. Oczywiście mogą wystąpić błędy próbkowania, bowiem w tym wypadku nie ma jednoznacznego związku przyczyna-skutek.

Termin szansa (prawdopodobieństwo) jest trudny do zdefiniowania, ale myślę, że jest zrozumiały nawet bez precyzyjnej definicji. Kiedy w problemie pojawia się element szansy, pojawiają się wtedy dla eksperymentatora duże

trudności. Szczególnie duże w przypadku wnioskowania indukcyjnego. Rozpatrzmy dedukcyjny problem dziesięciokrotnego rzucania nieobciążoną monetą. Poprzez dedukcję możemy obliczyć dla wszystkich 11 możliwych wyników prawdopodobieństwo ich wystąpienia. Na przykład prawdopodobieństwo wystąpienia 5 reszek i 5 orłów wynosi 0.246 czyli 24.6%. Jeśli zmodyfikujemy początkowe założenia (na przykład obciążając ją lekko z jednej strony), obliczenia staną się bardziej pracochłonne, ale dalej będą wykonalne i proste. Na szczęście teoria prawdopodobieństwa jest już dobrze opracowana, istnieją tablice i metody skrócone (oraz bardzo rozbudowane programy komputerowe), zmniejszające ilość koniecznej pracy do minimum.

Teraz rozpatrzmy ten sam problem indukcyjnie. Jeżeli rzucimy 10 razy otrzymamy 5 razy orła i 5 razy reszkę, co możemy powiedzieć o jednorodności monety? Z całkowitą pewnością możemy jedynie stwierdzić, że moneta ta nie ma po obydwu stronach reszki (zarówno jak i orła). Jeśli moneta nie jest obciążona, takiego wyniku możemy spodziewać się z prawdopodobieństwem równym około 25%. Z dużym prawdopodobieństwem, że jest to prawda, możemy stwierdzić, że kostka nie jest mocno obciążona. Musimy jednak pamiętać, że nigdy takie stwierdzenie nie jest całkowicie pewne. Nawet w przypadku silnie obciążonej monety (gdy np. prawdopodobieństwo wyrzucenia reszki jest równe 90%) istnieje pewne, niewielkie prawdopodobieństwo pojawienia się 5 reszek i 5 orłów. Nie zdefiniowaliśmy do tej pory co oznacza niewielkie lub duże obciążenie, jednak metody statystyczne pozwalają wyznaczyć zakres obciążeń które możemy uważać za pomijalnie małe. Osoby przyzwyczajone do precyzyjnych odpowiedzi dedukcyjnej matematyki mogą być rozczarowane niejasnością odpowiedzi.

1.2. Projektowanie badań naukowych

Odkrycie możliwości wyrażenia każdej wartości interesującej nas własności (mierzonej wielkości fizycznej) za pomocą liczby, jest bardzo ważnym osiągnięciem myśli ludzkiej. Otrzymanie na drodze doświadczalnej związku ilościowego pomiędzy określoną wielkością fizyczną, a jej wartością przyjętą za jednostkę odniesienia, nazywamy pomiarem. Należy jednak pamiętać, że każdemu pomiarowi towarzyszy występowanie błędów. Ponieważ mierzona wielkość fizyczna (wraz z błędami) może być traktowana jako proces losowy, statystyka matematyczna daje nam również ważne wskazówki co do projektowania poprawnego eksperymentu:

1. Powtarzalność - oznacza zastosowanie oddziaływania dwa i więcej razy w celu doświadczalnego oszacowania błędów oraz poprawy precyzji pomiaru (obserwacji) skutków tego oddziaływania. Ilość koniecznych powtórzeń zależy od wielkości różnic (odchyłek) które chcemy wykryć oraz zmienności wielkości, którą chcemy zbadać. Pamiętając o tych obu rzeczach na początku eksperymentu zmniejszamy ilość stresów towarzyszących badaniom.
2. Przypadkowość - jest to przyporządkowanie oddziaływań do przyrządów badawczych w taki sposób, że wszystkie mają jednakową szansę otrzymania oddziaływania.
3. Lokalna kontrola - ma na celu zmniejszenie błędów pomiarowych, na przykład pomiary grupujemy w bloki i po wykonaniu każdego bloku analizujemy jego wyniki.

Uniwersalna procedura przeprowadzenia badań naukowych jest trudna do zdefiniowania, jednak zwykle wymienia się następujące elementy:

1. Sformułowanie hipotezy - próbne, nieobowiązujące rozwiązanie, wyjaśnienie.
2. Zaplanowanie eksperymentu obiektywnie testującego tę hipotezę.
3. Skrupulatne wykonanie obserwacji i zebranie danych doświadczalnych podczas eksperymentu.
4. Interpretacja wyników eksperymentu - rozważenie otrzymanych danych w kontekście innych znanych faktów i danych mogących potwierdzić lub zaprzeczyć naszym wynikom i postawionej na wstępie hipotezie.
5. Zakres stosowalności wniosków - oczywiście powinien być jak najszerszy. Eksperyment powtarzalny w czasie i przestrzeni zwiększa zakres stosowalności wniosków wypływających z niego. Inną drogą poszerzenia zakresu jest eksperyment współczynnikowy, w którym efekty działania jednego z czynników badane są w funkcji zmieniających się pozostałych możliwych czynników.
6. Obliczenie wielkości błędów pomiarowych - w każdym eksperymencie istnieje pewien element niepewności co do ważności uzyskanych wyników. Doświadczenie powinno być tak zaprojektowane, aby można było oszacować (obliczyć) wielkość błędu.

Można także określić najważniejsze kroki eksperymentatora:

1. Zdefiniowanie problemu - pierwszym krokiem na drodze do rozwiązania jest prawidłowe (przejrzyste, jasne) sformułowanie problemu. Jeśli nie możemy zdefiniować problemu mamy małą szansę na rozwiązanie tego problemu. Jeżeli problem jest zrozumiały, powinniśmy potrafić postawić pytania, na które odpowiedź przybliży nas do rozwiązania.
2. Zestawienie celów - może być w formie pytań na które należy odpowiedzieć, hipotez które należy przetestować lub zjawisk które należy zbadać. Cele te powinny być dobrze sprecyzowane, bowiem tylko takie postawienie celów

- umożliwia eksperymentatorowi prawidłowe i efektywne zaprojektowanie doświadczenia. Jeżeli mamy więcej niż jeden cel, cele powinny być uporządkowane pod względem ważności, i w takiej kolejności uwzględnione w projekcie. Przy określaniu celów nie należy być ani zbyt ambitnym ani zbyt ostrożnym.
3. Wybór oddziaływań - sukces eksperymentu zależy w dużej mierze od skrupulatnego doboru oddziaływań (zmiennych parametrów), opracowania metodyki postępowania, które pozwolą odpowiedzieć na postawione pytania.
 4. Wybór badanego materiału - należy uwzględnić cele eksperymentu oraz wielkość populacji o której wnioski chcemy wyciągnąć. Materiał powinien stanowić próbę reprezentatywną tego materiału.
 5. Wybór układu doświadczalnego - tutaj ponownie należy rozważyć cele, ale podstawową zasadą jest wybranie najprostszego układu spełniającego wymagania co do dokładności pomiaru.
 6. Wybór obserwowanej wielkości i ilości powtórzeń.
 7. Kontrola wzajemnego wpływu obserwowanych wielkości - zwykle stosuje się tu wartości graniczne lub obróbkę statystyczną.
 8. Wstępna, teoretyczna analiza przydatności wyników - zebrane dane powinny prawidłowo opisywać skutki oddziaływania będące celem eksperymentu.
 9. Przeprowadzenie analizy statystycznej i zsumowanie wyników - opisanie źródeł błędów i określenie stopni swobody dla analizy wariancji. Należy zaplanować zastosowanie różnych testów F, zaplanować jak otrzymane wyniki będą zastosowane oraz przygotować odpowiednie tabele lub wykresy przedstawiające spodziewany efekt pomiarów (które należy porównać z założonymi celami). W tym miejscu dobrze jest dać nasze plany do przejrzenia kolegom, mogą oni zauważyć błędy, których my nie zauważyliśmy.
 10. Przeprowadzenie doświadczenia - przeprowadzając eksperyment staraj się zachować obiektywizm. Zorganizuj tak zapisywanie danych aby łatwo je było potem analizować. Jeśli konieczne jest kopiowanie (przepisywanie) danych nie zapnij porównać ze sobą oba egzemplarze!
 11. Analiza danych i interpretacja wyników - Wszystkie dane doświadczalne powinny zostać zanalizowane w zaplanowany sposób a wyniki zinterpretowane w świetle warunków doświadczenia, hipotezy powinny zostać przetestowane.
 12. Przygotowanie kompletnego, czytelnego i poprawnego raportu badań.

1.3. Hipotezy

Hipoteza jest to próbna teoria dotycząca natury i powiązań poszczególnych obserwacji. Hipotezy różnią się swą subtelnością i w związku z tym źródłem swego powstania. Prosta hipoteza może być na przykład tylko uogólnieniem obserwacji, hipoteza bardziej złożona może postulować istnienie powiązań między zdarzeniami lub skomplikowanych łańcuchów przyczynowo-skutkowych. Analogia jest tu bardzo potężnym narzędziem, największe znaczenie ma jednak wyobraźnia. Możliwość budowania hipotez opiera się na założeniu, że w naturze istnieje pewien ład, nie jest to jednak równoznaczne ze stwierdzeniem, że wszystkie części natury są uporządkowane. Jeżeli dwie hipotezy pasują do zaobserwowanych faktów, a jedna z nich jest prostsza od drugiej, to zazwyczaj przyjmuje się tę prostszą, do czasu, gdy dalsze fakty nie spowodują jej odrzucenia.

Trzeba zawsze pamiętać o tym, że hipoteza jest tylko próbnym pomysłem, sugestią - która jeżeli nie została sprawdzona nie może być traktowana jako prawo. Zdarza się, szczególnie na pograniczu nauki, że hipotezy przyjmuje się bez dostatecznego sprawdzenia. Prawdopodobieństwo nie może jednak zastąpić dowodów. W celu sprawdzenia hipotezy można dokonać dodatkowych obserwacji lub przeprowadzić doświadczenie sprawdzające. Nie ma ścisłego rozgraniczenia między doświadczeniem a prostą obserwacją, lecz w doświadczeniu obserwator zwykle ingeruje w pewnym stopniu i stwarza warunki lub wywołuje wydarzenia korzystne dla swego celu. Zespół warunków wymaganych dla danego zdarzenia (nazywanych zmiennymi) może być zazwyczaj ograniczony do skończonej liczby, wystarczającej do celów praktycznych. W idealnym doświadczeniu wszystkim tym zmiennym można nadać wartości żądane przez eksperymentatora. Hipoteza powinna nie tylko pasować do faktów, które wywołały jej stworzenie i nowych obserwacji, ale również wykazywać zgodność z pozostałymi częściami nauki.

W fizyce istnieją hipotezy trudne do sprawdzenia drogą bezpośrednich doświadczeń, zamiast tego sprawdza się następstwa wydedukowane z tych hipotez. Trudno jest np. wymyślić doświadczenie bezpośrednio sprawdzające równanie falowe Schrodingera w mechanice kwantowej, a przecież zostały potwierdzone tysiące następstw wyprowadzonych z niego drogą dedukcji matematycznej. Ważne jest rozróżnianie warunku koniecznego dla prawdziwości danego twierdzenia, od warunku dostatecznego.

Jedna z reguł postępowania eksperymentatora mówi, że w przypadku, gdy wydaje nam się, iż w aparaturze występuje jakaś symetria, czyli odwrócenie pewnej wielkości lub przestawienie dwu elementów nie powinno dawać żadnego efektu (lub efekt powinien być przewidywalny), to taką zamianę należy koniecznie wykonać.

Przykład 1.

Podczas wyznaczania współczynnika przewodnictwa cieplnego przy pomocy aparatu Christiansena¹, mierzy się temperaturę w trzech różnych punktach, przy pomocy trzech jednakowych termometrów. Ze względu na symetrię układu zamiana miejscami dwu termometrów nie powinna wpłynąć na wynik pomiaru. Gdy zamienimy termometry, najczęściej stwierdzamy, że wystąpiła różnica wskazań, bowiem każdy z termometrów ma inny błąd systematyczny, a każdy z pomiarów obciążony jest błędem przypadkowym. Zamieniając termometry i obliczając średnią z kilku pomiarów, znacznie redukujemy całkowity błąd pomiaru. Jeżeli różnice temperatur $T_2 - T_1$ i $T_3 - T_2$ są małe, wskazane wydaje się tutaj zastąpienie zwykłych termometrów, platynowymi termometrami oporowymi włączonymi w ramiona dwóch mostków Wheatstone'a.

¹ Sz. Szczeniowski, *Fizyka doświadczalna. II. Ciepło i fizyka cząsteczkowa*, PWN Warszawa 1976, str. 144.

2. ELEMENTY STATYSTYKI

- rozkłady prawdopodobieństwa;
- hipotezy statystyczne;
- estymacja.

2.1. Pojęcia podstawowe

Wynik każdego pomiaru obarczony jest przypadkowym błędem (jest przesunięty w stosunku do rzeczywistej wartości o pewną wartość, której nie znamy), jest zatem zmienną losową, której rozkład powinniśmy znać.

Zagadnienie polegające na wydaniu orzeczenia, czy obserwowany rozkład można aproksymować przez pewien określony rozkład teoretyczny, jest jednym z podstawowych zadań statystyki matematycznej. Rozkład prawdopodobieństwa zmiennej losowej służy jako teoretyczny model rozkładu wartości badanego parametru (cechy) w populacji generalnej, z której pobieramy próbkę. Wnioskowanie statystyczne będące przedmiotem statystyki matematycznej może występować w zależności od potrzeb praktycznych w dwojakim rodzaju:

- a) estymacji - szacowania parametrów rozkładu badanej cechy w populacji;
- b) testowania hipotez statystycznych - dotyczących rozkładu badanej cechy w populacji. Podstawowe metody i narzędzia są dla tych dwóch rodzajów wnioskowania odmienne.

Przystępując do badań statystycznych w mniejszym lub większym stopniu nie znamy rozkładu interesującego nas parametru (cechy). Możemy jednak (na ogół) ustalić klasę rozkładów, które mogą być brane pod uwagę jako ewentualne rozkłady tego parametru. Podstawowym pytaniem jest w takim przypadku pytanie: Czy wskazany rozkład może być uznany, czy też nie, za rozkład badanego parametru?

Każdej wartości zmiennej losowej przypisujemy prawdopodobieństwo jej wystąpienia. Dla zmiennej losowej dyskretnej definicja prawdopodobieństwa jest intuicyjna:

$$P = \frac{n(x)}{N} \quad (1)$$

gdzie $n(x)$ jest liczbą przypadków wystąpienia wartości x zmiennej losowej w N próbach ($N \in \mathbb{N}$). Dla zmiennej losowej ciągłej definicja ta wymaga określenia niewielkiego przedziału zmienności Δx ($\Delta x \neq 0$), dla którego określane jest prawdopodobieństwo:

$$P(x_0 < x < x_0 + \Delta x) = \frac{n(x_0 < x < x_0 + \Delta x)}{N} \quad (2)$$

Dla zmiennej losowej ciągłej korzystniej jest stosować gęstość prawdopodobieństwa:

$$f(x_0) = \frac{P(x_0 < x < x_0 + \Delta x)}{\Delta x} \quad (3)$$

Z punktu widzenia eksperymentatora bardzo ważnym parametrem rozkładu prawdopodobieństwa jest wartość oczekiwana (wartość średnia) zmiennej losowej, przyjmowana najczęściej jako rzeczywista wartość wielkości fizycznej, która opisuje ten rozkład:

$$E(x) = \sum_{i=1}^k x_i P(x_i) \quad (4a)$$

$$E(x) = \int_a^b x f(x) dx \quad (4b)$$

Istotnym parametrem rozkładu prawdopodobieństwa jest także dyspersja (odchylenie standardowe), stanowiące podstawę do wyznaczania błędu mierzonej wielkości fizycznej.

$$\sigma(x) = \sqrt{V(x)} \quad (5)$$

gdzie V oznacza wariancję, którą w zależności od typu rozkładu można obliczyć korzystając z następujących zależności:

$$V(x) = \sum [x_i - E(x)]^2 P(x_i) \quad (6a)$$

$$V(x) = \int_a^b [x - E(x)]^2 f(x) dx \quad (6b)$$

Zapoznamy się teraz z wybranymi rozkładami zmiennych losowych.

2.2. Rozkład dwumianowy.

Rozważmy populację, którą możemy podzielić na dwie frakcje, w jednej z frakcji występuje pewna cecha, nie występująca w drugiej. Jeżeli z tej populacji pobierzemy losowo jeden element, to cechuje go prawdopodobieństwo p posiadania tej cechy i prawdopodobieństwo $1-p$, że cecha ta nie wystąpi. Jeśli po zwróceniu pierwszego elementu dokonamy ponownego losowania, to prawdopodobieństwo posiadania cechy (np. posiadania barwy zielonej) będzie równe p . Zakładamy tu, że wylosowany element ma prawdopodobieństwo p posiadania koloru zielonego, bez względu na wyniki poprzednich losowań. Po pobraniu dwu elementów są trzy możliwości: oba elementy były zielone (prawdopodobieństwo p^2), jeden zielony drugi innego koloru (prawdopodobieństwo $2p(1-p)$), brak zielonego (prawdopodobieństwo $(1-p)^2$). Prawdopodobieństwo wylosowania k elementów zielonych przy n -krotnym losowaniu ze zwracaniem będzie równe:

$$P_n^k = \frac{n!}{k! (n-k)! p^k (1-p)^{n-k}} \quad (7)$$

Rozkład wartości P_n^k nazywany jest rozkładem dwumianowym lub rozkładem Bernoulliego. Przy dużej liczbie próbek wartość przeciętna:

$$\bar{r} = \sum_{k=0}^n (k P_n^k) = np \quad (8)$$

Wartość najbardziej prawdopodobna, czyli wartość k' , która wystąpi najczęściej przy dużej liczbie próbek, jest największa z liczb całkowitych spełniająca nierówność: $k' \leq (n+1)p$ (gdy $(n+1)p$ jest liczbą całkowitą, wówczas wystąpią dwie najbardziej prawdopodobne wartości $(n+1)p$ i $(n+1)p-1$ mające to samo prawdopodobieństwo). Odchylenie standardowe σ od wartości przeciętnej dla każdej populacji określa wzór:

$$\sigma = \sum_{k=0}^n \left[(k - \bar{k})^2 P_n^k \right]^{1/2} \quad (9)$$

zaś wariancja σ^2 wynosi: $\sigma^2 = np(1-p)$.

Jest to użyteczna miara "rozrzutu" wartości k w różnych próbkach. Część próbek, w których wartość k odchyła się od wartości przeciętnej $\bar{k} = np$ o więcej niż dwie lub trzy wielokrotności σ jest nieznaczna. Dla dużych n i p niezbyt bliskich 0 lub 1 można zastosować przybliżenie:

$$P_n^k = \frac{1}{\sqrt{2\pi\sigma}} \exp(-0.5y^2) \left[1 + \frac{2p-1}{\sigma} \left(\frac{y}{2} - \frac{y^3}{\sigma} \right) \right] \quad (10)$$

gdzie $y = \frac{k - np}{\sqrt{np(1-p)}} = (k - k') / \sigma$.

Rozkład dwumianowy ma różne zastosowania, po pierwsze można przy jego pomocy przewidywać rozkład próbek w populacji, w której występuje dwumianowy rozkład pewnej cechy, po drugie możemy posłużyć się nim przy sprawdzaniu hipotez statystycznych, a także możemy wykorzystać do estymacji.

Przykład 2.

Załóżmy, że energia pochodząca z jednego źródła ma być z przerwami zużywana przez 5 przyrządów. Aby otrzymać choćby tylko z grubsza oszacowanie zapotrzebowania na energię, musimy dodatkowo przyjąć, że w każdej chwili prawdopodobieństwo p zapotrzebowania każdego przyrządu na energię jest takie samo, przyrządy pracują niezależnie od siebie, każdy z przyrządów korzysta z energii 12 minut w ciągu godziny. Niech X oznacza liczbę przyrządów korzystających z energii w danym momencie czasu, zadaniem naszym jest obliczenie prawdopodobieństwa, że liczba przyrządów korzystających z energii w danym momencie jest nie większa niż 2. Zmienna X ma rozkład dwumianowy, w którym $n = 5$, $p = 12/60 = 0.2$, czyli prawdopodobieństwa:

$$\begin{aligned} P(X=0) &= 0.8^5 \approx 0.33 \\ P(X=1) &= 5 \cdot 0.8^4 \cdot 0.2 \approx 0.41 \\ P(X=2) &= 10 \cdot 0.8^3 \cdot 0.2^2 \approx 0.20 \\ P(X=3) &= 10 \cdot 0.8^2 \cdot 0.2^3 \approx 0.05 \\ P(X=4) &= 5 \cdot 0.8 \cdot 0.2^4 \approx 0.01 \\ P(X=5) &= 0.2^5 \approx 0.00 \end{aligned}$$

Prawdopodobieństwo tego, że liczba przyrządów pobierających energię równocześnie jest nie większa od 2 jest równa sumie prawdopodobieństw $P(X=0) + P(X=1) + P(X=2) \approx 0.94$. Wartość najbardziej prawdopodobna $[(n+1)p] = [1,2] = 1$, co jest wynikiem logicznym, jeśli popatrzymy na założenia. Był to przykład przewidywania rozkładu próbek. Przykładem testowania hipotezy, jest sprawdzanie hipotezy, że pewna klasa w szkole stanowi próbkę losową pod względem liczby dzieci leworęcznych, pobraną z populacji zawierającej 5% dzieci leworęcznych. Czy hipotezę tę należy odrzucić, gdy okaże się, że wszystkie 20 dzieci w klasie będzie leworęcznych? Wynik taki może wystąpić z prawdopodobieństwem 10, czyli hipotezę należy odrzucić.

Przykład 3.

Rozważmy teraz przypadek 4 dzieci leworęcznych w grupie 20 dzieci. Łączne prawdopodobieństwo zdarzeń: w klasie jest nie mniej niż 4 osoby leworęczne jest równe 0.017, zatem reguła: "odrzućmy za każdym razem hipotezę, gdy w grupie 20 dzieci będzie 4 lub więcej leworęcznych" może nie być prawdziwa w 1.7% przypadków. W takim przypadku można wnioskować, że albo w grupie dzieci jest coś szczególnego (dzięki czemu nie jest ona próbą losową), albo dzieci wybrano z populacji zawierającej więcej niż 5% leworęcznych albo też obie te możliwości występują razem.

Metody sprawdzania hipotez mogą odpowiedzieć na pytanie: "czy ta przyczyna powoduje rzeczywiste skutki?". Natomiast na pytanie: "jak wielki jest skutek tej przyczyny?" odpowiada nam teoria estymacji, podając albo zakres w którym domyślamy się istnienia rzeczywistej wartości, albo pojedynczą liczbę jako przypuszczalnie trafną ocenę rzeczywistej wartości.

2.3. Rozkład Poissona.

Istnieją pewne zagadnienia (np. przy korzystaniu ze znaczników promieniotwórczych o długim okresie rozpadu) gdy zliczamy kolejne przypadkowe zdarzenia, aby w ten sposób otrzymać pewne wartości średnie charakteryzujące zjawisko.

Mówimy, że zmienna losowa X ma rozkład Poissona gdy:

$$P(X = k) = \frac{\exp(-\lambda)\lambda^k}{k!} \quad (11)$$

gdzie $\lambda = \frac{1}{n} \sum_{k=0}^n kn_k > 0$, zaś $k = 0, 1, 2, \dots$.

Rozkład Poissona można stosować do aproksymacji rozkładu dwumianowego, gdy prawdopodobieństwo p jest małe, a liczba doświadczeń n duża, przyjmujemy wtedy, że $\lambda = np$. Także wiele istniejących w praktyce rozkładów może być w dobrym przybliżeniu aproksymowanych przez rozkład Poissona.

Przykład 4.

Rozpatrzmy dane liczbowe uzyskane przy badaniu rozpadu promieniotwórczego. Przeprowadzono $n = 2608$ pomiarów trwających po 7,5 sekundy każdy, polegających na zliczaniu przez licznik scyntylacyjny liczby dochodzących do niego cząstek. Dane liczbowe przedstawia tabela 1.

Tabela 1.

Wyniki badania rozpadu promieniotwórczego.

Liczba cząstek k	Liczba doświadczeń n_k	n_k/n	P(k) $\lambda = 3.85$
0	57	0.022	0.021
1	203	0.078	0.081
2	383	0.147	0.156
3	525	0.201	0.201
4	532	0.204	0.195
5	408	0.156	0.151
6	273	0.105	0.097
7	139	0.053	0.054
8	45	0.017	0.026
9	27	0.010	0.014
10	16	0.006	0.007
Razem	2608	0.999	1.000

Prawdopodobieństwa podane w czwartej kolumnie tej tabeli zostały policzone na podstawie wzoru dla $\lambda = 3,85$. Z danych liczbowych wynika, że różnice pomiędzy zaobserwowanymi częstościami empirycznymi a prawdopodobieństwami teoretycznymi w rozkładzie Poissona są rzędu 10^{-3} .

2.4. Rozkład gamma.

Zmienna losowa ciągła ma rozkład gamma, wtedy, gdy gęstością prawdopodobieństwa jest następująca funkcja:

$$f(x) = \begin{cases} 0 & \text{dla } x \leq 0 \\ \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} & \text{dla } x > 0 \end{cases} \quad (12)$$

przy czym $a > 0$ i $p > 0$ są parametrami tego rozkładu. Funkcja Γ występująca we wzorze jest nazywana funkcją gamma Eulera, która dla $p > 0$ opisana jest zależnością:

$$\Gamma(p) = \int_0^{+\infty} \lambda^{p-1} e^{-x} dx \quad (13)$$

Wartość przeciętna zmiennej losowej X jest równa:

$$\bar{x} = p/a \quad (14)$$

zaś wariancja tej zmiennej losowej jest równa:

$$\sigma = p/a^2 \quad (15)$$

Szczególnym przypadkiem rozkładu gamma, gdy $p = 1$, jest rozkład wykładniczy, z którym mamy do czynienia w zagadnieniach ruchu na liniach telefonicznych, problemach czasu obsługi i czasu oczekiwania na obsługę, czy to w przypadku sklepu, czy też w przypadku obsługi maszyn, w problemach czasu eksploatacji elementów i przyrządów.

Szczególne znaczenie ma tu wykładnicze prawo niezawodności. Pod pojęciem niezawodności rozumie się prawdopodobieństwo bezawaryjnej pracy w ciągu czasu t , czyli tego, że urządzenie wykona zamierzone czynności w określonym przedziale czasu i w określonych warunkach. Stwierdzono, że dobrą aproksymacją niezawodności N jest funkcja:

$$N(t) = \exp(-\lambda t) \quad \text{dla } t > 0 \quad (16)$$

lub

$$N(t) = 1 - F(t) \quad (17)$$

gdzie $F(t)$ jest dystrybuantą (prawdopodobieństwem skumulowanym) w punkcie t zmiennej losowej T o rozkładzie wykładniczym. T oznacza tutaj czas poprawnej pracy, a własność tej zmiennej opisaną pierwszym wzorem nazywamy właśnie wykładniczym prawem niezawodności.

2.5. Rozkład Weibulla

Zmienna ma rozkład Weibulla, jeżeli jej gęstość prawdopodobieństwa opisana jest wzorem:

$$f(x) = \begin{cases} 0 & \text{dla } x < 0 \\ \lambda p x^{p-1} e^{-\lambda p x} & \text{dla } x \geq 0 \end{cases} \quad (18)$$

o parametrach p i $\lambda > 0$. Wartość przeciętna zmiennej losowej X jest równa:

$$\bar{x} = \lambda^{-1/p} \Gamma\left(\frac{1}{p} + 1\right) \quad (19)$$

zaś wariancja tej zmiennej losowej opisana jest zależnością:

$$\sigma^2 = \lambda^{-2/p} \left\{ \Gamma\left(\frac{2}{p} + 1\right) - \left[\Gamma\left(\frac{1}{p} + 1\right) \right]^2 \right\} \quad (20)$$

a funkcja Γ jest dana wzorem (13).

2.6. Rozkład Erlanga.

Dla rozkładu Erlanga funkcja gęstości prawdopodobieństwa ma postać:

$$f(x) = \begin{cases} 0 & \text{dla } x \leq 0; \\ \frac{1}{(k-1)lm^k} e^{-x/m} x^{k-1} & \text{dla } x > 0 \end{cases} \quad (21)$$

Wartość przeciętna zmiennej losowej X jest równa:

$$\bar{x} = km \quad (22)$$

zaś wariancja tej zmiennej losowej jest równa:

$$\sigma^2 = km^2 \quad (23)$$

2.7. Rozkład normalny.

Z innym problemem mamy do czynienia, gdy składniki pewnej populacji różnią się między sobą ilościowo (a nie jakościowo jak w rozkładzie dwumianowym). W zależności od populacji, wartości te mogą rozkładać się rozmaicie, istnieje jednak pewien typ rozkładu, nazywany rozkładem normalnym, zajmujący wysoką pozycję ze względu na matematyczną prostotę, częste występowanie jako graniczna postać innych rozkładów teoretycznych oraz jako przybliżenie rozkładów rzeczywistych. Wszędzie tam, gdzie wielkość danej cechy kształtuje się pod wpływem dużej liczby czynników i żaden z nich nie góruje nad pozostałymi, można spodziewać się występowania rozkładu normalnego. Mówimy, że zmienna losowa X ma rozkład normalny, jeśli gęstość prawdopodobieństwa f tej zmiennej opisuje zależność:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\} \quad (24)$$

Gęstość prawdopodobieństwa jest funkcją, która w przypadku zmiennej losowej X o dystrybuancie F typu ciągłego, spełnia warunek:

$$F(X) = \int_{-\infty}^{+\infty} f(u) du \quad (25)$$

Przypomnijmy, że dystrybuanta jest funkcją prawdopodobieństwa określoną wzorem: $F(x) = P((-\infty, x))$. Wartość przeciętna dla tego rozkładu jest równa m zaś wariancja σ^2 . Do obliczania prawdopodobieństw postaci $P(a < X < b) = F(b) - F(a)$ korzystamy ze stabilizowanej dla $x \geq 0$ funkcji

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{u^2}{2}\right) du \quad (26)$$

i zależności:

$$F(x) = \begin{cases} 0.5 + \Phi(x) & \text{dla } x \geq 0 \\ 0.5 - \Phi(x) & \text{dla } x < 0 \end{cases} \quad (27)$$

Przykład 5.

Wytrzymałość lin stalowych produkowanych w pewnej fabryce jest zmienną losową o rozkładzie normalnym $N(100 \text{ MPa}, 5 \text{ MPa})$ czyli o wartości przeciętnej $m = 100 \text{ MPa}$ i wariancji $\sigma^2 = 5 \text{ MPa}$. Mamy obliczyć ile przeciętnie lin spośród 1000 ma wytrzymałość mniejszą niż 90 MPa, oraz co która przeciętnie lina ma wytrzymałość mniejszą niż 90 MPa?

Przyjmijmy, że częstość przyjmowania wartości z przedziału $(-\infty, 90)$ jest równe prawdopodobieństwu przyjmowania wartości z tego przedziału. Odpowiedź na pierwsze pytanie uzyskujemy obliczając prawdopodobieństwo tego, że wytrzymałość jest mniejsza od 90 MPa, czyli korzystając ze wzoru: $P(X < 90) = P(X - 100/5 < 90 - 100/5) = P(Y < -2) = 0.5 - \Phi(2) = 0.02275$. Natomiast odpowiedź na drugie pytanie uzyskamy obliczając

odwrotność prawdopodobieństwa $P(X < 90)$, czyli $K = 1/0.023$, 43, a więc co 43 lina ma wytrzymałość mniejszą niż 90 MPa.

Przykład 6.

Kolejny przykład dotyczy pomiaru odległości. Pomiar ten jest obarczony błędem systematycznym $b = -50$ mm (polegającym na podawaniu odległości mniejszej od rzeczywistej) oraz błędem przypadkowym X (który jest zmienną o rozkładzie normalnym $N(0 \text{ mm}, 100 \text{ mm})$). Błąd całkowity Y jest sumą tych błędów. Obliczyć prawdopodobieństwo tego, że $|Y| < 100$ mm oraz odczytany wynik pomiaru nie przekracza rzeczywistej wartości mierzonej wielkości. Zmienna losowa $Y = X + b$ ma rozkład normalny $N(b, 100)$, zatem prawdopodobieństwo $P(|Y| < 100) = P(-100 + 50/100 < Y + 50/100 < 100 + 50/100) = P(-1/2 < Y + 50/100 < 3/2) = \Phi(3/2) + \Phi(1/2)$, 0.625. Natomiast fakt, że wynik pomiaru nie przekracza rzeczywistej wartości mierzonej wielkości jest równoważny temu, że błąd Y jest nie większy od zera, czyli należy obliczyć $P(Y < 0) = P(Y + 50/100 < 50/100) = P(Y + 50/100 < 1/2) = 0.5 + \Phi(1/2) = 0.691$.

2.8. Rozkład chi-kwadrat.

W statystyce matematycznej bardzo często występują rozkłady prawdopodobieństwa, które mówimy teraz szczegółowo. Pierwszym z tych rozkładów jest rozkład chi-kwadrat o k stopniach swobody, który jest rozkładem zmiennej losowej ciągłej o funkcji gęstości prawdopodobieństwa określonej wzorem:

$$f(\chi^2) = \begin{cases} 0 & \text{dla } x \leq 0 \\ \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2} & \text{dla } x > 0 \end{cases} \quad (28)$$

Wartość przeciętna zmiennej losowej jest równa liczbie stopni swobody n , zaś wariancja wynosi $2n$. Rozkład chi-kwadrat jest szczególnym przypadkiem rozkładu gamma. Dla $n > 30$ rozkład chi-kwadrat można z bardzo dobrą dokładnością aproksymować rozkładem normalnym.

W tablicach (np. w książce [15]) podane są prawdopodobieństwa $P(\chi_n^2 = \chi_\alpha^2) = \alpha$ dla wybranych wartości α i $n \leq 30$.

2.9. Rozkład t Studenta

Rozkład t Studenta o k stopniach swobody jest rozkładem zmiennej losowej ciągłej t_n postaci:

$$t_n = \frac{X}{\sqrt{\chi_n^2/n}} \quad (29)$$

gdzie zmienna losowa X ma rozkład normalny $N(0, 1)$ a χ_n^2 na rozkład chi-kwadrat o n stopniach swobody i zmienne te są od siebie niezależne.

Gęstość prawdopodobieństwa zmiennej losowej t_n określona jest zależnością:

$$f(x) = \begin{cases} 0 & \text{dla } x \leq 0 \\ \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{t^2}{n}\right)^{(n+1)/2}} & \text{dla } x > 0 \end{cases} \quad (30)$$

Wykres gęstości prawdopodobieństwa rozkładu Studenta jest symetryczny względem prostej $t = 0$ i jest podobny do wykresu gęstości prawdopodobieństwa rozkładu normalnego. Wartość przeciętna istnieje tylko dla $n > 1$ i jest równa zero, natomiast wariancja zmiennej losowej t istnieje tylko dla $n > 2$ i jest równa:

$$\sigma^2 = n / (n-2).$$

Dla dużych n ($n > 30$) rozkład Studenta można aproksymować rozkładem normalnym.

2.10. Rozkład F Snedecora.

Zmienna losowa F jest stosunkiem dwóch niezależnych zmiennych losowych X i Y o rozkładach chi-kwadrat z odpowiednio m i n stopniami swobody:

$$F = \frac{m}{n} \frac{X}{Y} \quad (31)$$

Gęstość prawdopodobieństwa zmiennej losowej F dana jest wzorem:

$$f_F(z) = \begin{cases} 0 & \text{dla } z \leq 0 \\ \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{n}{m}\right)^{m/2} \frac{z^{(n/2-1)}}{(z+m/n)^{(m+n)/n}} & \text{dla } z > 0 \end{cases} \quad (32)$$

Wartość przeciętna zmiennej losowej istnieje dla $n > 2$ i wynosi:

$$z = n/(n-2),$$

natomiast wariancja istnieje dla $n > 4$ i jest równa:

$$\sigma^2 = 2n(m+n-2) / m(n-2)(n-4) \quad (33)$$

Rozkład F Snedecora jest stabilizowany, podawane są liczby $F_{\alpha/2}$ takie, że $P(F > F_{\alpha/2}) = \alpha$ dla wybranych wartości α .

Zmienna F jest stosunkiem pomiędzy dwoma wariancjami i jest stosowana do zbadania, czy dwa niezależne estymatory wariancji mogą być przyjęte do estymacji tej samej wariancji. Nazwa testu pochodzi od nazwiska Ronalda A. Fishera, a wprowadzona została przez Georga W. Snedecora. Statystyka F jest szeroko stosowana w analizie wariancji do testowania hipotez statystycznych.

2.11. Hipotezy statystyczne.

Hipoteza statystyczna jest to jakiegokolwiek przypuszczenie co do rozkładu populacji generalnej (dotycząca parametrów lub postaci rozkładu określonego zbioru). Hipotezy te mogą być dwojakiemu rodzaju:

- hipotezy parametryczne - precyzujące wartość parametru w rozkładzie populacji generalnej znanego typu;
- hipotezy nieparametryczne - precyzujące jedynie typ rozkładu populacji generalnej.

W statystyce przeważnie nie mamy absolutnej pewności co do słuszności pewnej hipotezy, a osiągnięcie takiej całkowitej pewności często jest nieopłacalne lub wręcz niemożliwe.

Konieczne jest jeszcze wprowadzenie w tym miejscu dwóch pojęć:

Hipoteza zerowa - podstawowa hipoteza statystyczna sprawdzana przy pomocy testu.

Hipoteza alternatywna - hipoteza konkurencyjna w stosunku do hipotezy zerowej w tym sensie, że jeżeli odrzuca się hipotezę zerową, to do testowania przyjmuje się hipotezę alternatywną. Hipotez alternatywnych może być więcej niż jedna. Do odrzucenia hipotezy zerowej wystarczy, że którakolwiek hipoteza alternatywna ma większe od niej prawdopodobieństwo.

Proces sprawdzania postawionej hipotezy zerowej ma zwykle następujący przebieg: stawiamy pewną hipotezę odnośnie całej populacji, pobieramy próbę, badamy ją i na tej podstawie akceptujemy lub odrzucamy postawioną hipotezę.

Weryfikacja hipotezy statystycznej odbywa się poprzez zastosowanie specjalnego narzędzia, zwanego testem statystycznym. Jest to reguła postępowania, która każdej możliwej próbie losowej przyporządkowuje decyzję przyjęcia lub odrzucenia sprawdzanej hipotezy. Przyjęcie i odrzucenie hipotezy w teście statystycznym nie jest jednak równoznaczne z logicznym udowodnieniem jej prawdziwości lub fałszywości. Należy bowiem pamiętać, że odrzucając sprawdzaną hipotezę w teście statystycznym, kierujemy się jedynie tym, że dane liczbowe wynikające z pomiarów dają nam małą szansę potwierdzenia prawdziwości tej hipotezy. Możliwy jest jednak przypadek, gdy hipoteza jest prawdziwa, natomiast nasze dane pomiarowe są złe lub po prostu mało prawdopodobne przy tej hipotezie.

Wnioskowanie statystyczne w metodach testowania hipotez statystycznych opiera się głównie na własnościach rozkładu normalnego.

2.12. Estymacja.

Podstawowe pojęcia z jakimi spotykamy się w teorii estymacji:

Estymator - dowolna funkcja służąca do oszacowania nieznanego parametru populacji generalnej;

Estymator nieobciążony - estymator dla którego wartość przeciętna jest równa zero, tzn. estymator szacujący parametr rozkładu bez błędu systematycznego;

Estymator efektywny - estymator o możliwie małej wariancji;

Estymator zgodny - estymator który jest stochastycznie zbieżny do parametru, czyli estymator podlegający działaniu prawa wielkich liczb (stosowanie większych prób poprawia dokładność szacunku);

Estymator wystarczający - estymator skupiający w sobie wszystkie informacje o badanym parametrze zawarte w próbie losowej;

Estymacja punktowa - metoda szacunku nieznanego parametru polegająca na tym, że jako wartość parametru przyjmuje się wartość estymatora tego parametru otrzymaną z n-elementowej próby losowej;

Estymacja przedziałowa - estymacja polegająca na budowie przedziału ufności dla tego parametru. Przedział ufności jest przedziałem losowym wyznaczonym za pomocą rozkładu estymatora, a mający tę własność, że pokrywa wartość parametru z góry zadanym prawdopodobieństwem, zapisujemy go zwykle w postaci $P(a < X < b) = 1 - \alpha$.

3. GENEROWANIE LICZB LOSOWYCH

- *Liczby losowe;*
- *Tablice liczb losowych;*
- *Generatory liczb losowych;*
- *Testy losowości.*

3.1. Liczby losowe.

Liczne decyzje oparte są na losowym wyborze, często dokonywanym w oparciu o liczby losowe, które znajdują się w tablicach lub są otrzymywane przy pomocy specjalnych generatorów. Liczby losowe mogą mieć różne postacie i wielkości. Zasadniczą cechą liczb losowych jest to, że znajomość liczb występujących w przeszłości nie wpływa na skuteczność przewidywania liczb przyszłych, czyli prawdopodobieństwo uzyskania określonej liczby przy kolejnej próbie nie ulega zmianom, zdarzenia są zatem niezależne.

Podstawowymi generatorami liczb losowych są generatory fizyczne, takie jak moneta, urna o odpowiedniej zawartości, kostka do gry lub generatory wykorzystujące losowy przebieg zjawisk fizycznych (rozpady promieniotwórcze, termiczny szum w półprzewodnikach).

Standardowym modelem liczb losowych jest jednostajny rozkład prawdopodobieństwa. Wygodnie jest generować liczby losowe z przedziału $<0, 1>$, bowiem liczby z tego przedziału można wygodnie i prosto przekształcić na elementy dowolnego przedziału. Natomiast generowanie liczb losowych o innych rozkładach sprowadza się do wykonania odpowiednich rachunków na liczbach o rozkładzie równomiernym.

Równomierność (jednostajność) rozkładu nie jest wystarczającym warunkiem losowości, ponieważ należy jeszcze zapewnić, aby poszczególne liczby nie pojawiały się w sposób okresowy.

Istnieje szereg standardowych testów losowości. Niektóre z nich sprawdzają własności zbiorów poszczególnych cyfr dziesiętnych pod względem następujących wymagań:

- a) losowe cyfry dziesiętne powinny pojawiać się w odstępach wynoszących średnio 10 cyfr - test na odstęp;
- b) w małych grupach cyfr losowych układy cyfr powinny powtarzać się z częstotliwością zgodną z rachunkiem prawdopodobieństwa - test pokerowy;
- c) każda kombinacja dwóch kolejnych cyfr losowych powinna pojawiać się z jednakową częstotliwością - test seryjny.

Inne testy sprawdzają zbiory ciągłych liczb dziesiętnych ze względu na następujące wymagania:

- a) test na korelację - między kolejnymi liczbami nie występuje korelacja;
- b) test na wariancję - średnia odległość dwóch losowych punktów na powierzchni kwadratu jednostkowego wynosi $1/2$.

W przypadku liczb otrzymywanych z generatorów programowanych na komputerach z góry wiadomo, że liczby te są pseudolosowe, bowiem otrzymywane kolejno liczby są zdeterminowane przez wybór stałych początkowych dla generatora.

3.2. Tablice liczb losowych.

Przy pomocy generatorów liczb losowych otrzymuje się tablice liczb losowych które znajdują szerokie zastosowanie w badaniach reprezentacyjnych (np. przy pobieraniu próbek losowych), w projektowaniu doświadczeń itp. Przykład tablicy liczb losowych stanowi tabela 2.

Tabela 2.
Tablica liczb losowych.

1	2	3	4	5	6	7	8	9	10
1534	7106	2836	7873	5574	7545	7590	5574	1202	7712
6128	8993	4102	2551	0330	2358	6427	7067	9325	2454
6047	8566	8644	9343	9297	6751	3500	8754	2913	1258
0806	5201	5705	7355	1448	9562	7514	9205	0402	2427
9915	8274	4525	5695	5752	9630	7172	6988	0227	4264
2882	7158	4341	3463	1178	5789	1173	0670	0820	5067
9213	1223	4388	9760	6691	6861	8214	8813	0611	3131
8410	9836	3899	3883	1253	1683	6988	9978	8026	6751
9974	2362	2103	4326	3825	9079	6187	2721	1489	4216
3402	8162	8226	0782	3364	7871	4500	5598	9424	3816
8188	6569	1492	2139	8823	6878	0613	7161	0241	3834
3825	7020	1124	7483	9155	4919	3209	5959	2364	2555
9801	8788	6338	5899	3309	0807	0968	0539	4205	8257

3.3. Generatory liczb losowych o równomiernym rozkładzie prawdopodobieństwa

Zazwyczaj liczby losowe są generowane przez komputer na podstawie pewnego wzoru matematycznego, realizowanego rekurencyjnie z wykorzystaniem poprzednich wyników. Dla zadanej wartości początkowej, generator zawsze wytworzy ten sam ciąg liczb losowych. Znając ten ciąg, można dokładnie przewidzieć liczbę następną. To samo odnosi się również do opublikowanych tablic liczb losowych, bowiem jeśli osoba losująca zna tablice to liczby zawarte w nich przestają być losowe. Liczby losowe, których dokładny wykaz można poznać, noszą nazwę liczb pseudolosowych.

Najprostszym przykładem generatora liczb losowych jest generator zbudowany w oparciu o odcinek jednostkowy. Poszczególne segmenty odcinka o jednostkowej długości określono jako możliwe zdarzenia, długości poszczególnych fragmentów zależą od ich prawdopodobieństwa (rysunek 3.1). Musimy zwrócić przy tym uwagę na to, aby wykorzystać cały odcinek jednostkowy i by poszczególne fragmenty nie zachodziły na siebie. Następnie należy jedynie generować losowe punkty na tym odcinku i przetwarzać je na wyniki w zależności od tego, na który segment przypadają.

Do generowania ciągu cyfr losowych o rozkładzie równomiernym można wykorzystać rozwinięcie dziesiętne dowolnej liczby niewymiernej (przykład wykorzystania liczby π do tego celu przedstawiony jest w książce Zielińskiego²).

Inna metoda generowania ciągów liczb losowych polega na wykorzystaniu ciągu liczb pochodzącego z generatora fizycznego i dopasowaniu ich do naszych potrzeb za pomocą operacji arytmetycznych

² R. Zieliński, Generatory liczb losowych. Programowanie i testowanie na maszynach cyfrowych, WN-T, Warszawa 1979, str. 22.

Podstawową rolę w realizacji programowanych generatorów liczb pseudolosowych o różnych rozkładach prawdopodobieństwa odgrywają algorytmy wytwarzania liczb pseudolosowych o równomiernym rozkładzie prawdopodobieństwa w przedziale $\langle 0,1 \rangle$. Do najbardziej rozpowszechnionych metod otrzymywania takich rozkładów należą metody kongruentne, dzielące się na dwie klasy: metody multiplikatywne generowania liczb pseudolosowych i metody addytywno-multiplikatywne (mieszane) oraz generatory Fibonacciego. W metodach mieszanych generowanie liczb x_i mniejszych niż dany dodatni moduł m poczynając od dowolnej nieujemnej liczby $x_0 < m$ polega na obliczaniu kolejnych wartości wyrażenia:

$$x_i = [a \cdot x_{i-1} + c] \bmod m \quad (34)$$

gdzie $i = 1, 2, 3, \dots$, zaś stałe spełniają warunki: $0 < a < m$, $0 \leq c < m$. Stała m najczęściej oznacza zakres liczb całkowitych, które mogą być reprezentowane w słowie maszynowym o długości b bitów ($m = 2^b$). Generator ten jest generatorem okresowym, przy czym pełny okres równy m mamy wtedy, gdy:

c jest liczbą pierwszą względem m ;

$a = k \cdot p + 1$ dla każdego czynnika pierwszego p liczby m (k jest dowolną liczbą całkowitą);

$a = k \cdot p + 4$, jeżeli 4 jest dzielnikiem liczby m .

Dla $c = 0$ generator nazywamy multiplikatywnym generatorem kongruentnym. Generator ten jest także generatorem okresowym, maksymalny jego okres jest równy 2^b . Generator multiplikatywny osiąga ten okres, gdy x_0 jest liczbą nieparzystą oraz gdy $c = k \cdot 8 + 3$ lub $c = k \cdot 8 + 5$ dla dowolnego całkowitego k .

Przykład 7.

Procedura generatora addytywno-multiplikatywnego:

```
Procedure Rand(iv : integer; f: real; var iy : integer);
```

```
const
```

```
maxint=65539;
```

```
randint=2147483647;
```

```
randre=0.4656613E-9;
```

```
BEGIN
```

```
iy:= iv*maxint;
```

```
IF iy < 0 THEN iy:=iy+randint+1;
```

```
f:=iy;
```

```
f:=f*randre;
```

```
END;
```

```
Function Rownom(a,b,iv): real;
```

```
var
```

```
x : real;
```

```
y : integer;
```

```
BEGIN
```

```
Rand(iv,x,y);
```

```
Rownom:=a+(b-a)*x;
```

```
END;
```

Istnieją też generatory addytywne (nazywane też uogólnionymi generatorami Fibonacciego), o ogólnej postaci:

$$x_{i+1} = [a_0 x_i + a_1 x_{i-1} + \dots + a_k x_{i-k} + b] \bmod m \quad (35)$$

gdzie a_i są zawsze równe zeru lub jedności, najprostsze generatory Fibonacciego korzystają z zależności:

$$x_{i+1} = [x_i + x_{i-1}] \text{ mod } m \quad (36)$$

Generatory te dobrze spełniają testy równomierności rozkładu, gorzej natomiast spełniają testy niezależności. Duża ich atrakcyjność polega na szybkości działania takich generatorów.

Z praktycznego punktu widzenia metody multiplikatywne są prostsze w realizacji (mniejsza liczba operacji arytmetycznych) i charakteryzują się wyższą wiarygodnością statystyczną (spełniają więcej testów losowych). Natomiast metody mieszane zapewniają dłuższy cykl sekwencji (dłuższy okres powtarzalności liczb losowych). Aby otrzymać ciąg liczb o rozkładzie równomiernym na przedziale (0,1) musimy liczby otrzymane przy pomocy tych generatorów podzielić przez m .

Otrzymane w ten sposób ciągi nie są w pełni losowe, lecz mają własności "pseudolosowe", takie jak: rozkład jednostajny w przedziale (0, m), korelację zero między różnymi x , losowo pojawiające się ciągi liczb parzystych lub nieparzystych itp. Jednostajność rozkładu można testować testem Chi-kwadrat. Liczby pseudolosowe o rozkładach różnych od jednostajnego można łatwo otrzymać jako funkcje $F(x_i)$, gdzie x_i są liczbami pseudolosowymi o rozkładzie jednostajnym.

3.4. Generatory liczb losowych o dowolnych rozkładach prawdopodobieństwa.

Generator liczb losowych o rozkładzie wykładniczym (z wartością oczekiwaną równą 1) można otrzymać z generatora o rozkładzie równomiernym obliczając logarytmy wylosowanych liczb (jest to tzw. metoda odwracania dystrybuanty). Metoda ta wymaga jednak dużo czasu i pamięci, dlatego opracowano szereg innych metod (np. metodę przybliżania rozkładu wykładniczego przy pomocy rozkładu dyskretnego).

Najprostszy sposób generowania liczb losowych o rozkładzie normalnym $N(k,1)$ polega na posumowaniu kolejnych $2k$ liczb losowych o rozkładzie jednostajnym.

Natomiast generator liczb losowych o rozkładzie Erlanga o k stopniu swobody otrzymujemy stosując sumowanie k liczb losowych o rozkładzie wykładniczym. Realizuje to poniższa procedura:

```
...
a:=0;
for i:=1 to k do
a:=a+Rand(a,b,xs);
erl:=-M*log(a);
...

```

Istnieje kilka ogólnych metod otrzymywania generatorów liczb losowych o dowolnym rozkładzie (np. takim jak na rysunku 4.2). Jeżeli znamy funkcję gęstości rozkładu, możemy zastosować pierwszy sposób (tzw. metodę eliminacji), polegający na:

- 1°. wygenerowaniu liczby losowej z_1 o rozkładzie jednostajnym z przedziału $\langle x_p, x_k \rangle$;
- 2°. wygenerowaniu liczby losowej z_2 o rozkładzie jednostajnym z przedziału $\langle 0, y \rangle$;
- 3°. sprawdzeniu czy $z_2 \leq f(z_1)$. Jeżeli warunek jest spełniony liczba z_1 jest liczbą losową o wymaganym rozkładzie (jeżeli nie - wracamy do punktu 1°).

W tym samym przypadku możemy też zastosować drugi sposób (tzw. superpozycji rozkładów). Jeżeli $g_y(x)$ jest gęstością pewnego rozkładu prawdopodobieństwa zależna od pewnego parametru y będącego zmienną losową o gęstości h a gęstość rozkładu zmiennej losowej który chcemy wygenerować jest równa:

$$f(x) = \int_{-\infty}^{+\infty} g_y(x)h(y)dy \quad (37)$$

wówczas:

- 1°. generujemy liczbę losową y zgodnie z rozkładem o gęstości h ;
- 2°. generujemy liczbę losową x zgodnie z rozkładem o gęstości g_y z parametrem y wylosowanym w punkcie 1°.

Natomiast gdy znamy dystrybuantę $h(x)$ generowanego rozkładu możemy zastosować metodę odwracania dystrybuanty. W tym przypadku:

- 1°. generujemy liczbę losową z_1 o rozkładzie jednostajnym z przedziału $\langle x_p, x_k \rangle$;
- 2°. obliczamy liczbę losową x na podstawie wzoru:

$$x = x_p + (x_{i+1} - x_i) \frac{h(z) - h(x_i)}{h(x_{i+1}) - h(x_i)} \quad (38)$$

przy czym $x_p \leq x_i \leq z \leq x_{i+1} < x_k$, natomiast dystrybuanta jest reprezentowana przez tablicę swych wartości w punktach x_i :

$$h(x_i) \leq h(z) \leq h(x_{i+1}) \quad (39)$$

3.5. Testy na losowość (nieprzypadkowość).

Ważną cechą liczb losowych otrzymanych przy pomocy generatora jest ich całkowita przypadkowość. Jeżeli nie jesteśmy w stanie zdefiniować przypadkowości tych liczb ilościowo pozostaje nam analizowanie trendów występujących w nich. Najprostszym sposobem analizy jest wykonanie wykresu i poszukiwanie dużych odchylenia od przypadkowości, jest to jednak decyzja subiektywna. Istnieją również testy statystyczne (niestety są one skuteczne w pełni dopiero w przypadku dużych zestawów danych), w większości przypadków odnoszące się do liczb losowych o rozkładzie równomiernym.

3.5.1. Serie.

Serie składają się z następujących po sobie punktów pomiarowych (liczb losowych), które w wyraźny sposób ukazują tendencję do wzrostu lub malenia wartości. Na wykresie będą one widoczne jako dodatnio lub ujemnie nachylone zbocze. Ponieważ wystąpienie ograniczonej ilości serii o niewielkiej długości jest dopuszczalne w przypadku liczb losowych, należy odróżnić te serie wynikłe z fluktuacji prawdopodobieństwa od innych serii. Aby określić, czy serie występujące w zbiorze danych są losowe należy obliczyć wartość średnią dla tego zbioru a następnie wszystkim wartościom nie większym niż średnia przyporządkować znak minus a pozostałym znak plus. Ilość serii N_{ob} jest równa liczbie zmian znaku powiększonej o 1. Przewidywana ilość serii w zbiorze n liczb losowych dana jest zależnością:

$$N_{sp} = 1/3 (2n-1) \quad (40)$$

zaś odchylenie standardowe dla liczby serii jest równe:

$$S = \sqrt{(16n - 29) / 90} \quad (41)$$

Następnie obliczamy stosunek:

$$r = \frac{N_{ob} - N_{sp}}{s} \quad (42)$$

Ogólnie możemy stwierdzić, że $r > 2$ wskazuje na niskie prawdopodobieństwo wystąpienia takiej ilości serii w zbiorze liczb losowych, $r > 3$ oznacza tak niskie prawdopodobieństwo wystąpienia serii, równoważne wystąpieniu elementu zakłócającego losowość (błąd w generatorze). Szczegółowe analizy w przypadku $n > 19$ wymagają zastosowania rozkładu chi-kwadrat, zaś dla mniejszych zbiorów liczb należy zastosować rozkład t Studenta. Liczba stopni swobody jest równa $n - 1$.

Przykład 8.

Weźmy następujący zbiór liczb losowych: 11; 12; 13; 14; 15; 10; 9; 8. Średnia wartość dla tego zbioru wynosi 11,5 a przypisane im znaki są następujące: - + + + - - -, czyli mamy dwie zmiany znaku i $N_{ob} = 3$. $N_{sp} = (2 \cdot 8 - 1)/3 = 5$ oraz $S = \sqrt{(16 \cdot 8 - 29) / 90} = 1.05$, w konsekwencji parametr $r = (3 - 5) / 1.05 = 1.90$. Jest on mniejszy od 2, a liczb losowych jest 8, zatem stosujemy rozkład t dla liczby stopni swobody równej 7. Odczytana z tablicy 3.5 wartość krytyczna z poziomem istotności 90% wynosi 1,47.

3.5.2. Trendy i nachylenia.

Aby zbadać, czy w zbiorze liczb losowych występują jakieś trendy musimy określić nachylenie i odchylenie standardowe dla tego zbioru. W tym celu metodą najmniejszych kwadratów prowadzimy przez punkty odpowiadające tym liczbom prostą o nachyleniu a . Następnie obliczamy stosunek:

$$t = \frac{a - 0.000}{\sigma^2} \quad (43)$$

i porównujemy z wartościami rozkładu t Studenta dla $(n - 2)$ stopni swobody (n oznacza liczbę punktów). Znak stosunku określa nam z jakim trendem mamy do czynienia.

Przykład 9.

Zbiór $n = 20$ liczb ma nachylenie $a = -2,50$ z odchyleniem standardowym $\sigma = 1,45$. Dla takich danych parametr $t = 2,50/1,45 = 1,72$. Natomiast parametr t z rozkładu Studenta dla poziomu ufności 10% i liczby stopni swobody $k = 18$ jest równy 1,73. Zatem możemy sądzić, że nachylenie a może wystąpić w tym zbiorze z prawdopodobieństwem równym 10%.

3.5.3. Średni kwadrat kolejnych różnic (MSSD).

Innym sposobem sprawdzenia poprawności działania generatora liczb losowych jest policzenie różnic pomiędzy kolejnymi n liczbami losowymi i podstawienie tych różnic do wzoru:

$$\text{MSSD} = \frac{\sum_{i=1}^n (x_{i+1} - x_i)^2}{n - 1} \quad (44)$$

oraz obliczenie wariancji σ^2 tego zbioru liczb i podzielenie przez nią parametru MSSD. Dla n dążącego do nieskończoności stosunek MSSD/σ^2 dąży do 2. Większe wartości stosunku oznaczają, że generator liczb losowych źle działa. Dla wartości mniejszych porównujemy otrzymaną liczbę z wartościami krytycznymi przedstawionymi w tabeli 3.

Tabela 3.

Wybrane wartości krytyczne dla stosunku MSSD/σ^2 .

n	α		
	0.1%	1.0%	5.0%
4	0.5898	0.6256	0.7805
5	0.4161	0.5379	0.8204
6	0.3634	0.5615	0.8902
7	0.3695	0.6140	0.9359
8	0.4036	0.6628	0.9825
9	0.4420	0.7088	1.0244
10	0.4816	0.7518	1.0623
20	0.7852	1.0406	1.2996
40	1.0850	1.2934	1.4921
60	1.2349	1.4144	1.5814
∞	2.000	2.000	2.000

4. TESTOWANIE HIPOTEZ STATYSTYCZNYCH

- *testy parametryczne;*
- *testy nieparametryczne.*

4.1. Wstęp.

Jak już wspomniałem poprzednio (patrz Rozdział 2.8) hipoteza statystyczna jest hipotezą dotyczącą parametrów (jest to tzw. hipoteza parametryczna) lub postaci rozkładu określonego zbioru (hipoteza nieparametryczna). W statystyce przeważnie nie mamy absolutnej pewności co do słuszności pewnej hipotezy, a osiągnięcie całkowitej pewności często jest nieopłacalne lub nawet z różnych względów niemożliwe. Proces sprawdzania hipotezy ma zwykle następujący przebieg: stawiamy pewną hipotezę odnośnie całej populacji, pobieramy próbę, badamy ją i na tej podstawie akceptujemy lub odrzucamy postawioną hipotezę.

Weryfikacja hipotezy statystycznej odbywa się przez zastosowanie specjalnego narzędzia, zwanego testem statystycznym. Jest to reguła postępowania, która każdej możliwej próbie losowej przyporządkowuje decyzję przyjęcia lub odrzucenia sprawdzanej hipotezy. Należy jednak podkreślić, że przyjęcie lub odrzucenie hipotezy w teście statystycznym nie jest równoznaczne z logicznym udowodnieniem jej prawdziwości lub fałszywości. Należy bowiem pamiętać, że odrzucając sprawdzaną hipotezę w teście statystycznym, kierujemy się jedynie tym, że dane liczbowe wynikające z pomiarów dają nam małą szansę prawdziwości tej hipotezy. Możliwy jest jednak przypadek, gdy hipoteza zerowa jest prawdziwa, natomiast nasze dane pomiarowe są złe lub, po prostu, mało prawdopodobne przy tej hipotezie.

W praktyce najczęściej stosuje się tzw. testy istotności, w których na podstawie wyników próby losowej podejmuje się jedynie decyzję odrzucenia sprawdzanej hipotezy lub stwierdza się, że brak jest podstaw do jej odrzucenia.

4.2. Testy parametryczne.

W praktycznych zastosowaniach statystyki bardzo często zachodzi konieczność sprawdzenia hipotez dotyczących wartości średnich w dwóch populacjach. Wprawdzie można porównywać ze sobą całe rozkłady, niejednokrotnie wystarcza nam jednak ograniczenie się jedynie do podstawowego parametru tych rozkładów, jakim jest wartość średnia populacji, czasem trzeba też skorzystać z innych parametrów rozkładu (przeważnie wariancji). Poznamy teraz wybrane testy parametryczne (testy istotności).

4.2.1. Test zgodności średniej próby ze średnią populacji (Test t Studenta).

Przy pomocy tego testu możemy sprawdzić hipotezę zerową postaci:

$$H_0 : \mu = \mu_0$$

natomiast hipoteza alternatywna ma postać:

$$H_1 : \mu \neq \mu_0$$

W praktyce rzadko znamy wartość średnią i odchylenie standardowe populacji generalnej, musimy więc zadowolić się szacunkiem tych wielkości przy pomocy najczęściej stosowanych estymatorów - wartości średniej z próby:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (45)$$

i odchylenia standardowego na podstawie próby obliczonego w oparciu o wzór:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (46)$$

wówczas statystyka:

$$t = \frac{x - \mu}{s / \sqrt{n}} \quad (47)$$

ma rozkład t Studenta o $n - 1$ stopniach swobody (n oznacza liczbę próbek), pod warunkiem, że populacja ma rozkład normalny lub bardzo zbliżony do niego.

Zatem, gdy chcemy sprawdzić hipotezę zerową o równości średniej wartości dla próby ze średnią wartością dla populacji, korzystamy z tablic rozkładu t Studenta i dla założonego poziomu ufności α odczytujemy wartość krytyczną t_α , taką, że:

$$P\{|t| \geq t_\alpha\} = 1 - \alpha \quad (48)$$

Następnie porównujemy wartość krytyczną z obliczoną wartością t i jeżeli:

- $|t| \geq t_\alpha$ wówczas odrzucamy hipotezę zerową;
- $|t| < t_\alpha$ wtedy nie ma podstaw do odrzucenia hipotezy zerowej.

Przykład 10.

Wiemy, że średni czas świecenia żarówki wynosi $\mu_0 = 1059$ godzin. Po wprowadzeniu zmian w technologii postanowiono sprawdzić, czy zmiany te nie skróciły czasu świecenia. Hipoteza zerowa ma zatem postać $H_0 : \mu_1 = \mu_0$, czyli: średni czas świecenia żarówki nie uległ zmianie. Do badania pobrano losowo próbę 10 żarówek, wyniki tych badań przedstawia tabela 4.

Tabela 4.

Wyniki badań losowych żarówek.

Czas pracy żarówki x [h]	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1015	-33	1089
1017	-31	961
1045	-3	9
1058	10	100
1147	99	9801
987	-61	3721
1295	247	61009
1054	6	36
884	-164	26896
978	-70	4900
10480	0	108522

Obliczony na podstawie próby średni czas świecenia żarówki wynosi:

$$\bar{x} = \frac{10480}{10} = 1048h$$

natomiast odchylenie standardowe:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{108522}{9}} = 109.8$$

obliczona na tej podstawie wartość statystyki Studenta:

$$t = \frac{1048 - 1059}{109.8} \sqrt{10} = -0,317$$

Odczytana z tablic dla poziomu ufności 0,95 wartość krytyczna $t_{\alpha} = 1,833$, zatem nie ma podstaw, aby hipotezę postawioną zerową odrzucić.

4.2.2. Test dla wariancji populacji generalnej (Test chi-kwadrat).

Wariancja jako miara rozproszenia (rozrzutu) badanej cechy bywa szczególnie często w różnych badaniach naukowych wykorzystywana do oceny stopnia jednorodności albo powtarzalności wyników liczbowych uzyskiwanych w eksperymentach naukowych. W szczególności ocena dokładności przyrządu pomiarowego, za pomocą którego mierzy się wyniki eksperymentu, wymaga często sprawdzania hipotez o wariancji σ^2 populacji. W praktyce wygodniejszą miarą rozproszenia jest odchylenie standardowe, ale średnie odchylenie kwadratowe s będące estymatorem odchylenia standardowego ma bardziej skomplikowany rozkład.

Rozkład chi-kwadrat, z którego korzysta się w tym teście, jest stabilizowany tylko dla liczby stopni swobody od 1 do 30, dla liczby stopni swobody większej od 30 korzystamy z rozkładu normalnego, do którego zbliżony jest ten rozkład, ale dla zmiennej $\sqrt{2\chi^2} - \sqrt{2k-1}$, gdzie k jest liczbą stopni swobody. Aby przeprowadzić test należy:

1. obliczyć wartość średnią \bar{x} oraz odchylenie kwadratowe s^2 korzystając ze wzoru:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2; \quad (49)$$

2. obliczyć wartość statystyki

$$\chi^2 = n s^2 / \sigma_0^2 \quad (50)$$

gdzie σ_0^2 jest hipotetyczną wartością wariancji;

3. dla poziomu ufności α i liczby stopni swobody $k = n - 1$ odczytujemy z tablicy rozkładu chi-kwadrat taką wartość krytyczną, aby spełniona była równość:

$$P\{\chi^2 \geq \chi_{\alpha}^2\} = 1 - \alpha. \quad (51)$$

Nierówność $\chi^2 \geq \chi_{\alpha}^2$ określa prawostronny obszar krytyczny, tzn. gdy z porównania wartości obliczonej z próby z wartością krytyczną znajdzie nierówność $\chi^2 \geq \chi_{\alpha}^2$ hipotezę zerową należy odrzucić. W przeciwnym wypadku, nie ma podstaw do odrzucenia hipotezy zerowej.

Przykład 11.

Dokonano 11 niezależnych pomiarów średnicy odlewanych rur i otrzymano następujące wyniki:

$$50.2, 50.4, 50.6, 50.5, 49.9, 50.0, 50.3, 50.1, 50.0, 49.6, 50.6 \text{ mm.}$$

Na poziomie ufności $\alpha = 95\%$ mamy zweryfikować hipotezę, że wariancja uzyskiwanych średnic rur jest równa 0,04 mm.

Wartość przeciętna $\bar{x} = 50,2$, natomiast odchylenie standardowe $s = 0,3162$, odchylenie kwadratowe $s^2 = 0,1$, zatem statystyka $\chi^2 = 11 \cdot 0,1 / 0,04 = 27,5$. Dla przyjętego poziomu ufności $\alpha = 0,95$ i dla liczby stopni swobody $k = n - 1 = 10$ odczytana z tablic wartość krytyczna $\chi_{\alpha}^2 = 18,307$, jest więc mniejsza od obliczonej, zatem hipotezę zerową należy odrzucić.

4.2.3. Test dla dwóch średnich wartości prób.

W praktycznych zastosowaniach zachodzi często potrzeba sprawdzenia hipotez dotyczących równości wartości średnich w dwóch populacjach normalnych. Wprawdzie można porównywać całe rozkłady badanych

populacji, ale ze względu na zalety i ważność parametru jakim jest wartość średnia populacji, często wystarczy porównać tylko te parametry. Test ten stosuje się na przykład w celu porównania dwóch przyrządów, nowej metody pomiarowej ze starą itp. Badamy dwie populacje o rozkładach normalnych $N(m_1, \sigma_1)$ i $N(m_2, \sigma_2)$. Odchylenia standardowe σ_1 i σ_2 tych populacji są znane. Na podstawie dwu niezależnych prób losowych o liczebnościach n_1 i n_2 sprawdzamy hipotezę zerową $H_0: m_1 = m_2$, jako hipotezę alternatywną przyjmując hipotezę $H_1: m_1 \neq m_2$. Z wyników pomiarów obliczamy wartości średnie, a następnie wartość statystyki U według wzoru:

$$U = (\bar{x}_1 - \bar{x}_2) / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}. \quad (52)$$

Z tablicy rozkładu normalnego $N(0,1)$ należy dla przyjętego poziomu ufności α odczytać wartość krytyczną u_α taką, aby spełniona była równość:

$$P\{|U| \geq u_\alpha\} = \alpha$$

- Gdy zachodzi nierówność $|u| \geq u_\alpha$, należy odrzucić hipotezę zerową na korzyść hipotezy alternatywnej.
- Gdy $|u| < u_\alpha$ nie mamy podstaw do odrzucenia hipotezy zerowej.

Gdy nie znamy wartości odchyłeń standardowych porównywanych populacji, a wiemy jedynie, że są one równe, stosuje się do weryfikacji hipotezy zerowej rozkład t Studenta, korzystając przy tym z dwóch małych prób o liczebnościach n_1 i n_2 . Po obliczeniu wartości średnich \bar{x}_1 i \bar{x}_2 oraz wariancji s_1^2 i s_2^2 obliczamy wartość statystyki t korzystając ze wzoru:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (53)$$

a następnie dla $k = n_1 + n_2 - 2$ stopni swobody i poziomu ufności α odczytujemy z tablicy rozkładu t Studenta wartość krytyczną t_α .

- Gdy $|t| \geq t_\alpha$ hipotezę zerową należy odrzucić.
- Gdy $|t| < t_\alpha$ nie ma podstaw do odrzucenia hipotezy zerowej.

4.2.4. Test z (zgodności średniej próby ze średnią populacji).

Średnia populacji generalnej równa się średniej średnich wszystkich możliwych prób:

$$\mu = \frac{1}{k} \sum_{i=1}^k (\bar{x}_i) \quad (54)$$

gdzie k jest liczbą prób. Odchylenie standardowe średnich prób równa się odchyleniu standardowemu populacji generalnej podzielonemu przez pierwiastek z liczby prób.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (55)$$

Jeżeli populacja generalna ma rozkład normalny, średnie prób też będą miały rozkład normalny o średniej równej średniej populacji generalnej mi i odchyleniu standardowym danym zależnością (55).

Statystyka z ma wówczas postać:

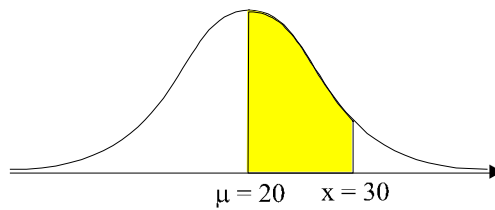
$$z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} \quad (56)$$

Statystyka ta jest odpowiada polu powierzchni zawartej pod krzywą rozkładu normalnego, które odpowiada prawdopodobieństwu w granicach $\langle \mu, x \rangle$.

Na przykład dla $x = 30$ i populacji generalnej o $\mu = 20$ i $\sigma = 5$ (rysunek 4.1) statystyka $z = 5$, co odpowiada polu powierzchni $S_x = 0,4772$ (całe pole powierzchni pod krzywą jest równe 1), czyli $P\{20 \leq x \leq 30\} = 0,4772$.

Stawiana przez nas hipoteza zerowa ma postać:

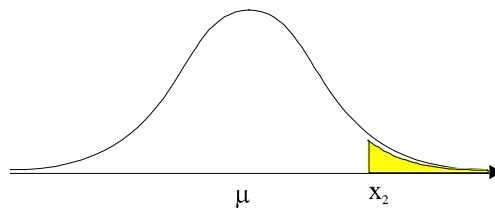
$$H_0: \mu_1 = \mu_0$$



Rys. 4.1. Rozkład normalny $N(20,5)$.

W teście tym możemy postępować na dwa sposoby. Pierwszy z nich jest następujący:

- 1°. z tablic parametru z rozkładu normalnego możemy odczytać odpowiednią wielkość z_{kr} dla pola pod krzywą odpowiadającego prawdopodobieństwu równemu założonemu poziomowi istotności α (rysunek 4.2);



Rys. 4.2. Ilustracja testu z przeprowadzonego pierwszym sposobem.

2. obliczamy $x_1 = \mu_0 - z_{kr} \sigma_{\bar{x}}$ oraz $x_2 = \mu_0 + z_{kr} \sigma_{\bar{x}}$;
 3°. Hipotezę zerową odrzucamy, gdy $\bar{x} < x_1$ lub gdy $\bar{x} > x_2$.

Możemy też postąpić w inny sposób:

- 1°. obliczamy statystykę z zgodnie ze wzorem (66);
 2°. w tablicach znajdujemy wartość z_{kr} dla założonego poziomu istotności α .
 • Jeżeli $z \geq z_{kr}$ hipotezę zerową należy odrzucić.
 • Jeżeli $z < z_{kr}$ nie ma podstaw do odrzucenia hipotezy zerowej.

Przykład 12.

W celu zbadania, czy wydajność pracy robotników pewnego zakładu jest zgodna z normą, przeprowadzono obserwacje pracy 100 tynkarzy i otrzymano rezultaty przedstawione w tabeli 5.

Tabela 5.
 Wyniki pomiarów otynkowanej powierzchni.

Otynkowano m ²	Liczba robotników
13	4
14	8
15	20
16	35
17	21
18	7
19	5
	100

Norma przewiduje otynkowanie 17 m² sciany na dzień. Stawiamy zatem hipotezę zerową

$$H_0 : \mu_1 = \mu_0$$

gdzie $\mu_0 = 17 \text{ m}^2$.

Na podstawie poprzednio prowadzonych obszernych badań ustalono odchylenie standardowe dla tego typu prac $\sigma = 2 \text{ m}^2$.

Wobec tego odchylenie standardowe średnich prób (wzór 55) wynosi

$$\sigma_{\bar{x}} = \frac{2}{\sqrt{100}} = 0,2 \text{ m}^2$$

Jeżeli zastosujemy test z przy poziomie istotności 0,05, to z tablic odczytamy wartość krytyczną $z_{kr} = 1,96$.

Wobec tego hipotezę zerową odrzucimy wtedy, gdy średnia próby o liczebności 100 będzie mniejsza od

$$\mu_0 - z_{kr} \sigma_{\bar{x}} = 17 - 1,96 \cdot 0,2 = 16,608 \text{ m}^2$$

lub większa od

$$\mu_0 + z_{kr} \sigma_{\bar{x}} = 17 + 1,96 \cdot 0,2 = 17,392 \text{ m}^2$$

Na podstawie próby obliczyliśmy średnią arytmetyczną, która wynosi $\bar{x} = 16,02 \text{ m}^2$. Jest ona mniejsza od dolnej granicy, wobec tego odrzucamy hipotezę zerową.

4.2.5. Test Bartletta.

Przy pomocy tego testu możemy zweryfikować hipotezę o równości wariancji wielu prób o równej liczebności:

$$H_0 : \sigma_1 = \sigma_2 = \dots = \sigma_k$$

Algorytm postępowania jest następujący:

1. obliczamy wartości średnie dla każdej z prób korzystając ze wzoru (45) oraz średnie odchylenia standardowe przy pomocy wzoru:

$$s_i^2 = \frac{1}{n_i - 1} \sum (x_{ij} - \bar{x}_i)^2 \quad (57)$$

2. obliczamy parametr

$$c = 1 + \frac{1}{3(k-1)} \left(\sum_i^k \frac{1}{n_i - 1} - \frac{1}{n-k} \right) \quad (58)$$

gdzie $n = \sum_i n_i$;

3. obliczamy statystykę:

$$\chi_{obl}^2 = \frac{2,303}{c} \left[\log \left(\frac{\sum (n_i - 1) s_i^2}{n - k} \right) - (n_i - 1) \log s_i^2 \right]. \quad (59)$$

Z tablic rozkładu chi-kwadrat dla założonego poziomu ufności α i $k - 1$ stopni swobody odczytujemy wartość krytyczną.

Jeżeli wartość obliczona jest nie mniejsza niż wartość krytyczna wtedy hipotezę zerową należy odrzucić. W przeciwnym przypadku nie ma podstaw do odrzucenia hipotezy zerowej.

4.3. Testy nieparametryczne.

Omawiane poprzednio testy statystyczne pozwalały formułować szereg wniosków dotyczących różnych parametrów statystycznych. Badanie zjawisk w drodze obliczania wybranych parametrów jest bardzo efektywnym

sposobem poznania, wynika to ze zwięzłej i precyzyjnej formy opisu. Jednak testy parametryczne, mimo swej różnorodności, nie dają odpowiedzi na wszystkie istotne pytania, głównie dlatego, że testy te mogą być stosowane w przypadku, gdy badana wielkość (populacja) ma rozkład normalny lub bardzo zbliżony do niego. Ponadto testy parametryczne, jak sama nazwa wskazuje, opisują pewną właściwość badanego zjawiska (wyników pomiarów), nie dając dostatecznych podstaw do formułowania wniosków ogólnych. Testy nieparametryczne są uniezależnione od rozkładu badanej cechy, mogą być więc stosowane także w przypadku dowolnych rozkładów, niekoniecznie zbliżonych do normalnego.

Testy nieparametryczne możemy podzielić na dwie grupy: pierwsza z nich to tzw. testy zgodności, pozwalające na sprawdzenie hipotezy, że populacja ma określony typ rozkładu, a druga to testy dla hipotezy, że dwie próby pochodzą z jednej populacji (czyli, że dwie populacje mają ten sam rozkład).

4.3.1. Test zgodności chi-kwadrat.

Jest to jeden z najstarszych testów statystycznych, pozwalający na sprawdzenie hipotezy, że populacja ma określony typ rozkładu (opisany pewną dystrybucją w postaci funkcji), przy czym może to być zarówno rozkład ciągły lub skokowy. Jedynym ograniczeniem jest to, że próba musi być duża, zawierająca co najmniej kilkadziesiąt próbek, bowiem wyniki jej musimy podzielić na pewne klasy wartości. Klasy te nie powinny być zbyt mało liczne, do każdej z nich powinno wpadać przynajmniej po 8 wyników. W przypadku, gdy w rozkładzie empirycznym występuje klasa o liczebności mniejszej od 8, należy klasę tę połączyć z sąsiednią. Dla każdej klasy z rozkładu hipotetycznego oblicza się liczebności teoretyczne, które porównuje się z empirycznymi za pomocą odpowiedniej statystyki chi-kwadrat. Gdy rozbieżności pomiędzy liczebnościami empirycznymi a teoretycznymi są zbyt duże, hipoteza, że populacja ma zakładany rozkład teoretyczny, musi być odrzucona. Sposób postępowania jest następujący:

1. Wyniki dzielimy na r rozłącznych klas o liczebnościach n_i , przy czym liczebność próby $n = \sum_i^r n_i$ otrzymując w ten sposób rozkład empiryczny.
2. Formułujemy hipotezę zerową, że badana populacja ma rozkład o dystrybucji należącej do pewnego zbioru rozkładów o określonym typie postaci funkcyjnej dystrybucji;
3. Z hipotetycznego rozkładu obliczamy dla każdej z r klas wartości badanej zmiennej losowej X prawdopodobieństwa p_i , że zmienna losowa przyjmie wartości należące do klasy o numerze i ($i = 1, 2, \dots, r$);
4. Obliczamy liczebności teoretyczne np_i , które powinny wystąpić w klasie i , gdyby populacja miała założony rozkład;
5. Ze wszystkich liczebności empirycznych n_i oraz hipotetycznych np_i wyznaczmy wartość statystyki:

$$\chi^2 = \sum_i^r \frac{(n_i - np_i)^2}{np_i} \quad (60)$$

która, przy założeniu prawdziwości hipotezy zerowej, ma rozkład chi-kwadrat o $r - 1$ stopniach swobody lub o $r - k - 1$ stopniach swobody, gdy z próby oszacowano k parametrów rozkładu;

6. Z tablicy rozkładu chi-kwadrat dla ustalonego poziomu ufności odczytuje się taką wartość krytyczną aby zachodziło $P(\chi^2 < \chi^2_\alpha) = 1 - \alpha$.
7. Porównujemy obie wartości i jeśli zachodzi nierówność $\chi^2 \geq \chi^2_\alpha$, to hipotezę należy odrzucić. W przeciwnym przypadku, gdy $\chi^2 < \chi^2_\alpha$, nie ma podstaw do odrzucenia hipotezy zerowej, nie oznacza to jednak, że możemy ją przyjąć.

Test chi-kwadrat jest bowiem tak zbudowany, że im bliższa zeru jest wartość χ^2 , tym hipoteza zerowa jest bardziej wiarygodna.

Przykład 13.

W pewnym doświadczeniu fizycznym mierzy się czas rozbłysku. Przeprowadzono $n = 1000$ niezależnych doświadczeń nad tym efektem i zbiór pogrupowanych wyników jest taki jak w tabeli 6.

Na poziomie ufności 99% należy zweryfikować hipotezę, że czas występowania badanego w tych doświadczeniach efektu świetlnego ma rozkład normalny. Z treści zadania nie wynikają parametry rozkładu hipotetycznego. Nasza hipoteza zerowa zatem będzie brzmiała: $F(x) \in \Omega$ gdzie Ω jest klasą wszystkich dystrybucji

normalnych. Dwa parametry rozkładu, średnią wartość m i odchylenie standardowe σ , szacujemy z próby za pomocą estymatorów $\bar{x} = 0.67$ i $s = 0.30$. Dalsze wyniki zestawiamy w tabeli 7, gdzie $F(u_i)$ jest wartością dystrybuanty rozkładu normalnego $N(0,1)$ w punkcie $u_i = (x_i - \bar{x}) / s$, który jest standaryzowaną wartością prawego końca przedziału klasowego.

Tabela 6.
Wyniki pomiaru czasu trwania rozblýsku.

Czas trwania [s]	Liczba pomiarów
0.0 - 0.2	50
0.2 - 0.4	128
0.4 - 0.6	245
0.6 - 0.8	286
0.8 - 1.0	134
1.0 - 1.2	90
1.2 - 1.4	67

Tabela 7.
Wyniki obliczeń.

x_i	n_i	$(x_i - \bar{x})/s$	$F(u_i)$	p_i	np_i	χ^2
0.2	50	-1.567	0.058	0.058	58	1.10
0.4	128	-0.900	0.184	0.126	126	0.03
0.6	245	-0.233	0.410	0.229	229	1.12
0.8	286	0.433	0.666	0.253	253	4.30
1.0	134	1.100	0.864	0.198	198	20.62
1.2	90	1.676	0.962	0.098	98	0.65
1.4	67	2.433	0.992	0.030	30	45.63
	1000			0.992	992	73.52

Liczba stopni swobody $k = 7 - 2 - 1 = 4$, gdyż na podstawie próby losowej zostały policzone dwa parametry: wartość średnia i odchylenie standardowe. Z tablic rozkładu chi-kwadrat, dla poziomu istotności 0,01, znajdujemy wartość krytyczną $\chi^2_{\alpha} = 13,277$. Wartość krytyczna jest mniejsza od obliczonej statystyki χ^2 równej 73,52, zatem hipotezę o normalności rozkładu należy odrzucić.

4.3.2. Test zgodności λ Kołmogorowa.

W teście zgodności λ Kołmogorowa, dla zweryfikowania hipotezy, że populacja ma określony typ rozkładu, nie rozpatruje się, jak w teście chi-kwadrat, liczebności szeregu empirycznego i porównuje z liczebnościami szeregu hipotetycznego, ale porównuje się dystrybuantę empiryczną i hipotetyczną. Bowiem, gdy populacja ma rozkład zgodny z hipotezą, to wartości dystrybuanty empirycznej i hipotetycznej powinny być we wszystkich badanych punktach zbliżone. Test rozpoczynamy od zanalizowania różnic między tymi dwoma dystrybuantami, największa z nich posłuży następnie do budowy statystyki lambda, której rozkład nie zależy od postaci dystrybuanty

hipotetycznej. Rozkład ten określa wartości krytyczne w tym teście. Jeżeli maksymalna różnica w pewnym punkcie obszaru zmienności badanej cechy jest zbyt duża, to hipotezę, że rozkład populacji ma taką dystrybuantę jak przypuszczamy, należy odrzucić.

Stosowanie tego testu jest jednak ograniczone, dystrybuanta hipotetyczna musi bowiem być ciągła, w zasadzie powinniśmy też znać parametry tego rozkładu, jednak w przypadku dużych prób możemy je szacować na podstawie próby.

Sposób postępowania w teście Kołmogorowa jest następujący:

1. porządkujemy wyniki w kolejności rosnącej lub grupujemy je w stosunkowo wąskie przedziały, o prawych końcach x_i i odpowiadających im liczebnościach n_i ;
2. wyznaczamy dla każdego x_i wartość empirycznej dystrybuanty $F_n(x)$ korzystając ze wzoru:

$$F_n(x_k) = \frac{1}{n} \sum_{i=1}^k n_i \quad (61)$$

3. z rozkładu hipotetycznego wyznaczamy dla każdego x_i wartość teoretycznej dystrybuanty $F(x)$;
4. dla każdego x_i obliczamy wartość bezwzględną różnicy $F_n(x) - F(x)$;
5. obliczamy wartość statystyki $D = \sup|F_n(x) - F(x)|$ oraz wartość statystyki:

$$\lambda = D\sqrt{n} \quad (62)$$

która, przy prawdziwości hipotezy zerowej, powinna mieć rozkład Kołmogorowa.

6. dla ustalonego poziomu ufności α odczytujemy z granicznego rozkładu Kołmogorowa wartość krytyczną spełniającą warunek $P\{\lambda \geq \lambda_{kr}\} = 1 - \alpha$.

Gdy $\lambda \geq \lambda_{kr}$ hipotezę zerową należy odrzucić, w przeciwnym wypadku nie ma podstaw do odrzucenia hipotezy zerowej.

Przykład 14.

Przebadano próbkę o liczebności $n = 1000$, a wyniki, pogrupowane w 10 wąskich klasach, zawarto w tabeli 8. Naszym zadaniem jest wysunąć sensowną hipotezę zerową dotyczącą rozkładu i zweryfikować ją na poziomie ufności 95%.

Tabela 8.

Przykładowe wyniki pomiarów.

Granice klas	Liczebność	$S_n(x_i)$	$(x_i - \bar{x})/s$	$F(x_i)$	$ S_n(x_i) - F(x_i) $
- 63.0	25	0.025	-2.0	0.0228	0.0022
63.0 - 63.5	65	0.090	-1.5	0.0668	0.0232
63.5 - 64.0	88	0.178	-1.0	0.1587	0.0193
64.0 - 64.5	131	0.309	-0.5	0.3085	0.0005
64.5 - 65.0	163	0.472	0.0	0.5000	0.0280
65.0 - 65.5	208	0.680	0.5	0.6915	0.0115
65.5 - 66.0	149	0.829	1.0	0.8413	0.0123
66.0 - 66.5	98	0.927	1.5	0.9332	0.0062
66.5 - 67.0	54	0.981	2.0	0.9772	0.0038
67.0 -	19	1.000	$+\infty$	1.0000	0.0000
	1000				

Rozkład liczebności jest zbliżony do symetrycznego, maksimum ma w jednej ze środkowych klas, co nasuwa hipotezę, że rozkład badanej cechy jest rozkładem normalnym $N(m, \sigma)$. Jeśli w wysuniętej hipotezie przyjąć $m =$

65, to w przedziale $\langle 63.0, 67.0 \rangle$, a więc o długości 4, mieściłoby się $1000 - (25 + 19) = 956$ wyników, co stanowi 95.6%. Z własności rozkładu normalnego wiemy, że prawdopodobieństwo przyjęcia wartości z przedziału o końcach $u - 1.96\sigma$ i $u + 1.96\sigma$ wynosi 95%, więc dla próby o liczebności 1000 w przedziale tym powinno się znaleźć 950 wyników, a więc niewiele mniej niż 956. Długość przedziału wynosi 3.92σ , co odpowiada w zadaniu wartości 4, zatem sensowną hipotezą wydaje się być $\sigma = 1$, czyli nasza hipoteza zerowa $H_0: N(65, 1)$.

W trzeciej kolumnie umieszczamy wartości dystrybuanty empirycznej dla zgrupowanych danych obliczone według wzoru:

$$S_n(x_i) = \begin{cases} 0 & \text{dla } x < x_0 \\ n_1/n & \text{dla } x_0 \leq x < x_1 \\ (n_1 + n_2)/n & \text{dla } x_1 \leq x < x_2 \\ \dots\dots\dots & \dots\dots\dots \\ 1 & \text{dla } x \geq x_k \end{cases} \quad (63)$$

W czwartej kolumnie umieszczamy standaryzowane prawe końce klas $(x - m)/\sigma$, w piątej kolumnie odczytane z tablicy wartości dystrybuanty $F(x_i)$ rozkładu $N(0, 1)$, a w ostatniej wartości bezwzględne różnic między dystrybuantami, z których największa jest $d_4 = 0.0280$. Następnie obliczamy $\sqrt{n} \cdot d_n = \sqrt{1000} \cdot 0.0280 = 0.886$. Dla poziomu ufności 0,95 odczytujemy z tablic rozkładu Kołmogorowa wartość krytyczną $\lambda_{kr} = 1,354$. Jest ona większa od obliczonej, zatem wyniki próby nie przeczą hipotezie zerowej, że rozkład populacji generalnej jest normalny $N(65, 1)$.

4.3.3. Test Kołmogorowa-Lillieforsa.

Gdy na podstawie n -elementowej ($n > 30$) próbki oszacujemy nieznanne parametry m i σ rozkładu przy pomocy estymatorów \bar{x} (danego wzorem (57)) oraz

$$s = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (64)$$

do weryfikacji hipotezy zerowej, że badana cecha ciągła ma rozkład normalny $N(\bar{x}, s)$ stosujemy test Kołmogorowa-Lillieforsa.

Algorytm postępowania jest następujący:

1. obliczamy na podstawie próby \bar{x} i s^2 ;
2. wyznaczamy dystrybuantę empiryczną na podstawie wzoru:

$$S_n(x) = \begin{cases} 0 & \text{dla } x < 0 \\ k_n & \text{dla } x_k \leq x < x_{k+1}, 1 \leq k \leq n-1 \\ 1 & \text{dla } x \geq x_n \end{cases} \quad (65)$$

3. obliczamy wartości bezwzględne różnic $|F(x_i) - S_n(x_i)|$ dla $i = 1, \dots, n$, gdzie $F(x)$ jest dystrybuantą rozkładu $N(\bar{x}, s)$. Stąd mamy:

$$F(x_i) = P(X < x_i) = P\{X - \bar{x} / s < x_i - \bar{x} / s\} = P(U < u_i) = F_0(u_i),$$

gdzie $u_i = (x_i - \bar{x}) / s$, a F_0 jest dystrybuantą rozkładu $N(0, 1)$.

4. spośród wszystkich obliczonych różnic wybieramy największą:
 $d_n' = \max |F(x_i) - S_n(x_i)|$
5. porównujemy obliczoną wartość d_n' z wartością krytyczną $k_n(\alpha)$ odczytaną z tablicy testu Kołmogorowa-Lillieforsa przy danym n oraz przyjętym poziomie ufności α .

Jeżeli $d_n' \in \langle k_n(\alpha), 1 \rangle$ to hipotezę odrzucamy, natomiast gdy $d_n' \in \langle 0, k(\alpha) \rangle$ próbka nie przeczy hipotezie, że pochodzi ona z populacji mającej rozkład $N(\bar{x}, s)$.

Przykład 15.

W wyniku pomiaru ładunków elektrycznych metodą Millikana otrzymano 40 następujących wyników (tabela 9):

Testem Kołmogorowa-Lillieforsa mamy zweryfikować hipotezę, że rozkład ładunków elektronów jest $N(\bar{x}, s)$. Na podstawie próbki obliczamy $\bar{x} = 1,59997$ i $s = 0,0032086$ (wobec tego $s^2 = 0,00001030$). Hipotetycznie zatem

rozkład jest $N(1,600, 0,00321)$. W celu obliczenia teoretycznych wartości dystrybuanty musimy przeprowadzić standaryzację, obliczone wartości zawarte są w tabeli 9. Skoki dystrybuanty obliczone dla $n_i = 2$ są równe $S_{40} = 2 \cdot \frac{1}{40}$, dla $n_i = 4$ mamy $S_{40} = 0,10$, natomiast dla $n_i = 6$ mamy $S_{40} = 0,15$.

Tabela 9.
Wyniki pomiarów ładunków elektrycznych metodą Millikana.

$x_i \cdot 10^{-19} \text{ C}$	n_i	$x_i \cdot 10^{-19} \text{ C}$	n_i
1.5951	4	1.5998	2
1.5961	2	1.6002	2
1.5968	2	1.6008	2
1.5975	2	1.6012	2
1.5878	4	1.6035	6
1.5985	2	1.6042	2
1.5988	2	1.6045	2
1.5992	2	1.6055	2

Jak widać w tabeli 10 największa różnica $d_n' = 0,106$. Dla $n = 40$ i poziomu ufności $1 - \alpha = 0,95$ odczytujemy z tablic testu Kołmogorowa-Lillieforsa wartość krytyczną $k_n(0,95) = 0,1401$. Ponieważ $d_{40} \in <0, k_n(\alpha) \equiv <0, 1,1401)$ zatem próbka nie przeczy normalności rozkładu przy poziomie ufności 95%.

4.3.4. Test Shapiro-Wilka.

Poznaliśmy już podobieństwa i różnice testów zgodności chi-kwadrat i Kołmogorowa. Obecnie zwróćmy uwagę na fakt, że żaden z dotychczas poznanych testów nie jest oparty na pełnej informacji, jaką można otrzymać z próbki. W teście chi-kwadrat strata informacji polega na grupowaniu obserwacji w klasy oraz na nieuwzględnianiu znaków różnic $n_i - np_i$. Natomiast test Kołmogorowa oparty jest tylko na jednej, największej różnicy d_n . Testem wykorzystującym pełną informację z próbki jest test Shapiro-Wilka odnoszący się jednak tylko do normalności rozkładu.

Jako statystykę testową przyjmuje się w nim zmienną losową:

$$W = \frac{a_i(n) (X_{n-i+1} - X_i)^2}{(X_j - \bar{X})^2} \quad (66)$$

gdzie $a_i(n)$ są stałymi zależnymi zarówno od liczności próbki n oraz od i (są one stabilizowane), a tzw. quasi-rozstępny rzędu i :

$$X_{n-i+1} - X_i = \begin{cases} i = 1, \dots, n/2, & \text{gdy } n \text{ parzyste} \\ i = 1, \dots, n-1/2, & \text{gdy } n \text{ nieparzyste} \end{cases}$$

Hipotezę o normalności rozkładu odrzuca się, gdy wartość W_d statystyki W obliczona na podstawie niezgrupowanej próbki leży poza przedziałem $(W(\alpha/2, n), W(1-\alpha/2, n))$, którego końcami są wartości krytyczne W podane w tablicy dla wartości α najczęściej używanych w zastosowaniach.

Przykład 16.

Pobrano próbkę o liczności $n = 19$, wyniki uporządkowano według wielkości: 12,4, 14,2, 14,9, 15,6, 16,1, 16,8, 17,3, 17,9, 18,2, 18,6, 19,3, 19,7, 20,4, 21,9, 22,8, 23,7, 25,2, 25,9, 27,4.

Tabela 10.
Wyniki obliczeń dla danych z tabeli 9.

u_i	$F(u_i)$	$S_n(x_i)$	$ S_n(x_i - F(u_i)) $
-1.518	0.062	0.10	0.038
	0.111	0.15	0.039
	0.156	0.20	0.044
	0.212	0.25	0.038
	0.246	0.35	0.104
	0.316	0.40	0.084
	0.356	0.45	0.094
	0.394	0.50	0.106
	0.480	0.55	0.070
	0.520	0.60	0.080
	0.603	0.65	0.047
	0.645	0.70	0.055
	0.867	0.85	0.017
	0.908	0.90	0.008
	0.924	0.95	0.026
	0.960	1.00	0.040

Na poziomie ufności $\alpha = 0,90$ należy zweryfikować testem Shapiro-Wilka hipotezę o normalności rozkładu badanej cechy w populacji generalnej.

Najpierw powinniśmy obliczyć kolejne różnice $x_{19} - x_1$, $x_{18} - x_2$, ..., $x_{11} - x_9$, wyniki tych obliczeń zawiera tabela 11.

W trzeciej kolumnie tej tabeli wypisane są odczytane z tablic przy $n = 19$ wartości a_i , natomiast w czwartej kolumnie umieszczamy iloczyny liczb z kolumn drugiej i trzeciej. Suma liczb z czwartej kolumny podniesiona do kwadratu i równa 305,52 jest licznikiem obliczanej statystyki W . Mianownik tego wyrażenia obliczamy korzystając z zależności:

$$\sum_{j=1}^n (x_j - \bar{x})^2 = \sum_{j=1}^n x_j^2 - n\bar{x}^2$$

Po podstawieniu do wzoru otrzymujemy:

$$\sum_{j=1}^{19} x_j^2 = 7869,77 \text{ oraz } \bar{x} = 19,3842,$$

zatem mianownik statystyki W_d jest równy:

$$7869,77 - 19 \cdot (19,3842)^2 = 7869,77 - 7139,20 = 730,57$$

czyli nasza obliczona statystyka jest równa $305,52 / 730,57 = 0,418$.

Z tablicy odczytujemy wartości kwantyli $W(\alpha/2, n) = W(0,95, 19) = 0,901$ i $W(1-\alpha/2, n) = W(0,95, 19) = 0,982$. Ponieważ obliczona wartość $W = 0,418$ leży poza przedziałem $(0,901, 0,982)$, więc hipotezę o normalności badanej cechy należy odrzucić przy poziomie ufności 90%.

Tabela 11.
Wyniki obliczeń dla testu Shapiro-Wilka.

i	$x_{n-i+1} - x_i$	$a_i(n)$	
1	$27,4 - 12,4 = 15,0$	0,4808	7,21200
2	$25,9 - 14,2 = 11,7$	0,3232	3,78144
3	$25,2 - 14,9 = 10,3$	0,2561	2,63783
4	$23,7 - 15,6 = 8,1$	0,2059	1,66779
5	$22,8 - 16,1 = 6,7$	0,1641	1,09947
6	$21,9 - 16,8 = 5,1$	0,1271	0,64821
7	$20,4 - 17,3 = 3,1$	0,0932	0,28892
8	$19,7 - 17,9 = 1,8$	0,6120	0,11016
9	$19,3 - 18,2 = 1,1$	0,0303	0,03333
			17,47915

4.3.5. Test niezależności chi-kwadrat.

Jest to test istotności pozwalający na sprawdzenie czy dwie badane cechy są niezależne. Próba musi być duża. Wyniki próby klasyfikujemy w tabelicę o r wierszach i s kolumnach, czyli dzielimy próbę na r grup ze względu na wartości cechy X i na s grup ze względu na wartości cechy Y. Wnętrze tablicy wypełniają liczebności tak utworzonych "dwuwymiarowych" klas n_{ij} , które muszą być nie mniejsze niż 8. Sumując wiersze i kolumny tablicy otrzymujemy liczebności brzegowe $n_{i\bullet}$ i $n_{\bullet j}$, takie, że:

$$n_{i\bullet} = \sum_j^s n_{ij} \quad (67)$$

$$n_{\bullet j} = \sum_i^r n_{ij} \quad (68)$$

zaś:

$$n = \sum_i^s \sum_j^r n_{ij} = \sum_i^s n_{i\bullet} = \sum_j^r n_{\bullet j} \quad (69)$$

Hipoteza zerowa może być sformułowana w następujący sposób:

$$H_0: P\{X = x_i, Y = y_j\} = P\{X = x_i\} \cdot P\{Y = y_j\}$$

a algorytm testowania tak postawionej hipotezy jest następujący:

1°. Obliczamy prawdopodobieństwa brzegowe:

$$p_{i\bullet} = \frac{n_{i\bullet}}{n} \quad (70)$$

$$p_{\bullet j} = \frac{n_{\bullet j}}{n} \quad (71)$$

2°. dla każdej kratki tablicy obliczamy prawdopodobieństwo hipotetyczne:

$$P_{ij} = p_{i\bullet} \cdot p_{\bullet j} \quad (72)$$

przy czym:

$$\sum_i^s \sum_j^r p_{ij} = 1$$

3°. mnożąc macierz prawdopodobieństw hipotetycznych przez liczebność próby n otrzymujemy macierz liczebności teoretycznych np_{ij} ;

4°. konstruujemy statystykę:

$$\chi^2 = \sum_i^s \sum_j^r \frac{(n_{ij} - np_{ij})^2}{np_{ij}} \quad (73)$$

5°. obliczamy ilość stopni swobody $k=(r-1)(s-1)$ i odczytujemy z tablic rozkładu chi-kwadrat wartość krytyczną χ_α^2 dla poziomu istotności $1-\alpha$ i liczby stopni swobody k , spełniającą zależność:

$$P\{\chi^2 \geq \chi_\alpha^2\} = 1 - \alpha.$$

6°. obliczoną wartość statystyki porównujemy z wartością krytyczną:

- Jeżeli $\chi^2 \geq \chi_\alpha^2$ hipotezę o niezależności badanych cech należy odrzucić;
- Jeżeli $\chi^2 < \chi_\alpha^2$ nie ma podstaw do odrzucenia hipotezy zerowej.

Przykład 17.

Stawiamy następującą hipotezę zerową: studentki lepiej zdają egzaminy niż studenci.

W celu przetestowania tej hipotezy z populacji generalnej studentów pewnej uczelni pobrano próbe losową o liczebności $n = 180$. Otrzymane wyniki przedstawia tabela 12.

Tabela 12.

Wyniki zdawania egzaminów przez studentki i studentów.

sesja	studentki	studenci	
zaliczona	75	25	100
niezaliczona	55	25	80
	130	50	180

Korzystając ze wzorów (70) – (72) obliczamy prawdopodobieństwa hipotetyczne (patrz tabela 12a) oraz odpowiadające im liczebności hipotetyczne a następnie korzystając z zależności (73) obliczamy statystykę χ^2 .

Tabela 12a.

Poszczególne etapy obliczeń w teście niezależności χ^2

Prawdopodobieństwa hipotetyczne

sesja	studentki	studenci	
zaliczona	0,40	0,16	0,56
niezaliczona	0,32	0,12	0,44
	0,72	0,28	1,00

Liczebności hipotetyczne

sesja	studentki	studenci	
zaliczona	72	29	
niezaliczona	58	18	

Obliczona wartość statystyki $\chi^2 = 3,554$, natomiast wartość krytyczna dla poziomu ufności $\alpha = 0,95$ i liczby stopni swobody $k = 1$, odczytana z tablic, wynosi $\chi_{\alpha}^2 = 3,841$ i jest większa od wartości obliczonej, zatem nie ma podstaw do odrzucenia hipotezy zerowej.

4.3.6. Test Wilcoxona

Dla dwóch populacji, w których interesująca nas cecha jest zmienną losową typu ciągłego możemy sprawdzić hipotezę, że w obu populacjach cecha ta ma taką samą dystrybucję. W tym celu z obu populacji pobieramy próby losowe: z pierwszej populacji o licznosci $k - \{x_1, x_2, \dots, x_k\}$ i z drugiej populacji o licznosci $l - \{y_1, y_2, \dots, y_l\}$. Aby można było przeprowadzić test Wilcoxona próbki muszą mieć przynajmniej po 4 elementy, a obie próbki w sumie muszą mieć przynajmniej 20 elementów.

Algorytm postępowania jest następujący:

1° . uporządkowujemy łącznie obie próbki według wzrastających wartości cechy.

2° . obliczamy ilość inwersji U w uporządkowanych próbkach. Para elementów (x_i, y_j) tworzy inwersję, gdy $y_j < x_i$

dla dowolnych i, j . Statystyka U ma rozkład normalny $N\left(0.5kl, \sqrt{\frac{1}{12}kl(k+l-1)}\right)$.

3° . W tablicach rozkładu normalnego znajdujemy wartość krytyczną u_{kr} , taką, że: $P\{|X| > u_{kr}\} = 1-\alpha$.

• Jeżeli $|u-m| > u_{kr}\sigma$ hipotezę zerową należy odrzucić.

• W przeciwnym wypadku nie ma podstaw do odrzucenia hipotezy zerowej.

5. ANALIZA WARIANCJI

- eksperyment jednoczynnikowy - podział według jednego kryterium;
- eksperyment wieloczynnikowy - podział według dwóch i więcej kryteriów;
- schemat kwadratu łacińskiego.

5.1. Wprowadzenie

W rozwiniętych badaniach empirycznych podstawową metodą sprawdzania hipotez naukowych jest eksperyment. Z najprostszym typem doświadczeń mamy do czynienia wtedy, gdy interesuje nas (jak to było dotychczas) wpływ jednego czynnika na przebieg eksperymentu. Bardziej złożonymi są doświadczenia z wieloma oddziaływującymi czynnikami. W doświadczeniu takim obiekty odpowiadające wartościom (poziomom) badanego czynnika i przedmioty eksperymentu (np. próbki) mogą być w różny sposób wzajemnie przyporządkowane, zgodnie z ustalonym kryterium tworząc określone typy układów doświadczalnych: układy proste (np. układ kompletnie zrandomizowany), układy blokowe (np. układ bloków zrandomizowanych kompletnych) i układy kolumnowo-wierszowe (np. układ kwadratu łacińskiego). W każdym z tych układów sposób przyporządkowania próbek i wartości czynników jest całkowicie losowy.

Nowa metodologia badań eksperymentalnych, a dokładniej planowania eksperymentu opartego na analizie wariancji, zaproponowana przez Ronalda A. Fishera, wykorzystywana była początkowo w rolnictwie. Pozwala ona manipulować więcej niż jedną zmienną niezależną jednocześnie, umożliwia to znaczne rozszerzenie zasięgu generalizacji wniosków eksperymentalnych. Najistotniejsze jednak jest to, że metoda ta pozwala uwzględnić efekt łącznego oddziaływania dwóch lub więcej zmiennych niezależnych na zmienną zależną.

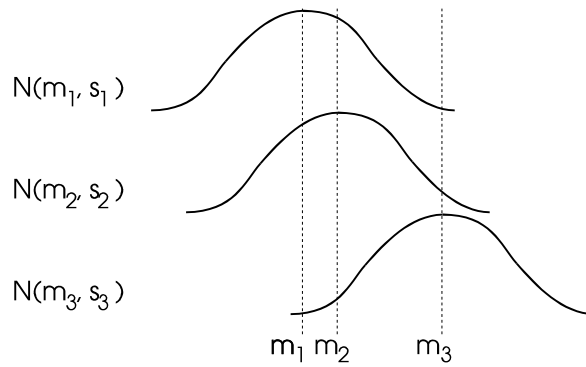
Jednym z częściej rozwiązywanych przy jej pomocy problemów jest analiza czynników zewnętrznych wpływających na wynik przeprowadzanego doświadczenia. W przypadku, gdy na przykład obserwujemy ilość substancji wydzielonej podczas przeprowadzanego doświadczenia chemicznego przy różnych temperaturach, mamy do czynienia z klasyfikacją jednoczynnikową (czynnikiem brany pod uwagę podczas analizy, jest w tym wypadku temperatura). Można także stosować klasyfikację według dwóch lub więcej kryteriów (np. możemy we wspomnianym poprzednio eksperymencie oprócz temperatury uwzględnić stężenia reagujących substancji, szybkość ich mieszania, zastosowany katalizator itd.) i wtedy mamy klasyfikację wieloczynnikową.

5.2. Eksperyment jednoczynnikowy.

Poprzednio rozważaliśmy zastosowanie statystyki t Studenta do porównywania średnich wartości dwóch prób (paragraf 4.1.3). Gdy musimy porównać ze sobą średnie więcej niż dwóch prób, musimy porównywać wszystkie średnie parami, co jest bardzo kłopotliwe i pracochłonne. Innym sposobem jest zastosowanie statystyki F. Dane z kilku prób lub dane z jednej próby pogrupowane według pewnego kryterium oznaczamy za pomocą x_{ij} , gdzie i oznacza numer populacji (grupy) a j kolejny numer obserwacji (pomiaru), liczba elementów w każdej grupie może być inna. Stawiamy hipotezę zerową, że między średnimi populacji z których zostały pobrane próby nie ma różnic: $H_0: m_1 = m_2 = \dots = m_n$.

Test analizy wariancji można stosować wówczas, gdy rozkłady populacji są normalne lub zbliżone do normalnego oraz mają jednakowe wariancje. Może bowiem zdarzyć się tak, że wszystkie populacje mają rozkłady normalne i jednakowe wariancje, ale różnią się wartościami średnimi (rysunek 7.1).

Istotą analizy wariancji jest bowiem rozbicie na addytywne składniki (których liczba wynika z potrzeb eksperymentu) sumy kwadratów całego zbioru wyników. Porównanie poszczególnej wariancji wynikającej z działania danego czynnika oraz tak zwanej wariancji resztowej, czyli wariancji mierzącej błąd losowy (przy zastosowaniu testu F Snedecora) daje odpowiedź czy dany czynnik odgrywa istotną rolę w kształtowaniu wyników eksperymentu.



Rys. 7.1. Wszystkie populacje mają rozkłady normalne i jednakowe wariancje ($\sigma_1=\sigma_2=\sigma_3$), ale różnią się wartościami średnimi ($m_1 \neq m_2 \neq m_3$).

Wariancja i-tej próby o liczebności n_i dana jest wzorem:

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_j)^2 \quad (74)$$

Wariancję ogólną można oszacować jako średnią z wariancji prób. Wobec tego wariancja oszacowana na podstawie wariancji prób (tzw. wariancja wewnątrz grup) jest opisana zależnością:

$$\sigma_w^2 = \frac{\sum_{j=1}^{n_1} (x_{1j} - \bar{x}_j)^2 + \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_j)^2 + \dots + \sum_{j=1}^{n_k} (x_{kj} - \bar{x}_j)^2}{\sum_{i=1}^k n_i - k} \quad (75)$$

i ma $N-k$ stopni swobody.

Dla potrzeb praktyki statystyka eksperymentalna wypracowała już bardzo wiele metod planowania doświadczeń, jak na przykład bloki losowe, kwadraty łacińskie, analiza czynnikowa. Każdy rodzaj doświadczenia ma odrębny schemat analizy wariancji.

Test analizy wariancji zwykle przeprowadza się według określonego schematu, ujętego w postaci tzw. tablicy analizy wariancji, mającej różną liczbę wierszy w zależności od konkretnego schematu, ale kolumny zawsze są takie jak przedstawione w tabeli 13.

Wariancja między grupami dana jest wzorem:

$$\sigma_m^2 = \frac{\sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2}{k - 1} \quad (76)$$

Statystykę F oblicza się tak, aby była ona większa od jedności, czyli dla $\sigma_w^2 > \sigma_m^2$ mamy $F = \frac{\sigma_w^2}{\sigma_m^2}$ natomiast w

przypadku gdy $\sigma_m^2 > \sigma_w^2$ wzór ma postać $F = \frac{\sigma_m^2}{\sigma_w^2}$.

Tabela 13.

Tablica analizy wariancji.

źródło zmienności	suma kwadratów	stopnie swobody	wariancja	statystyka
miedzy populacjami	$\sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2$	k-1	σ_m^2	$F = \frac{S_1^2}{S_2^2}$
wewnątrz populacji	$\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$	n-k	σ_w^2	

5.2.1. Hipotezy zerowe i alternatywne w jednoczynnikowej analizie wariancji.

Analiza wariancji (w skrócie określana często skrótem ANOVA) umożliwia ocenę prawdopodobieństwa tego, że różnice między średnimi wartościami wyników dla kilku ($k > 2$) prób losowych są dziełem przypadku. Podaną w poprzednim paragrafie ogólną hipotezę zerową (wraz z hipotezą alternatywną $H_1 : \sim H_0$) można zastąpić innymi zestawami hipotez.

Hipotezy te mogą być sformułowane w następujący sposób:

$$\begin{aligned}
 1^\circ. H_0 : \bigwedge_i (m_i - m) = 0 & \quad H_1 : \bigvee_j (m_j - m) \neq 0; \\
 2^\circ. H_0 : \sum_{i=1}^k (m_i - m)^2 = 0 & \quad H_1 : \sum_{i=1}^k (m_i - m)^2 \neq 0; \\
 3^\circ. H_0 : \bigwedge_i \alpha_i = 0 & \quad H_1 : \bigvee_j \alpha_j \neq 0
 \end{aligned}$$

gdzie α_i oznacza i-ty efekt czynnika A wpływającego na badaną cechę X i $\alpha_i = (m_i - m)$;

$$4^\circ. H_0 : \sum_{i=1}^k \alpha_i^2 = 0 \quad H_1 : \sum_{i=1}^k \alpha_i^2 \neq 0;$$

5.2.2. Weryfikacja hipotezy o równości wartości przeciętnych w przypadku klasyfikacji jednoczynnikowej.

W teście tym wymaga się aby badana cecha X miała w każdej z k populacji taki sam rozkład $N(m_i, s)$. Jeżeli nie znamy wartości odchylenia standardowego lub nie mamy pewności co do równości odchyłeń standardowych we wszystkich rozkładach, musimy estymować ten parametr na podstawie prób losowych i udowodnić hipotezę o równości wszystkich wariancji przy pomocy testu parametrycznego Bartletta opisanego w paragrafie 4.2.5 (gdy próby są różnej liczebności) lub Cochran-Coxa albo Hartleya (gdy próby są równoliczne) przeprowadzanego w sposób zbliżony do testowania zbiorów danych opisanego w rozdziale 4. Analiza wariancji nie wymaga równej liczebności prób.

Z każdej z badanych populacji pobieramy próbkę o liczebności n_i ($i = 1, \dots, k$). Zatem x_{ij} będzie j-tym wynikiem w i-tej próbce, średnia wartość i-tej próbki jest równa:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \tag{77}$$

natomiast średnia ogólna:

$$\bar{x} = 1/n \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij} = 1/n \sum_{i=1}^k \bar{x}_i n_i \tag{78}$$

przy czym liczba wszystkich próbek $n = \sum_{i=1}^k n_i$

Sumę kwadratów odchyłeń poszczególnych obserwacji x_{ij} od średniej ogólnej \bar{x} (oznaczaną dalej symbolem q) można przedstawić w postaci sumy dwóch składników:

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

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

Pierwszy składnik (oznaczany symbolem q_R) nazywany jest sumą odchyłeń wewnątrz grup (albo resztkową sumą kwadratów), drugi zaś (oznaczany q_G) nazywany jest sumą kwadratów pomiędzy grupami.

Odpowiadające im zmienne losowe Q , Q_R i Q_G , po podzieleniu przez odpowiednie stopnie swobody są estymatorami nieznannej wariancji σ^2 . Natomiast zmienna losowa $F = \frac{Q_G / k - 1}{Q_R / n - k}$ ma rozkład F Snedecora o $(k-1, n-k)$

k) stopniach swobody.

Ze względu na to, że rozpatrywany model wymaga równości wariancji badanej cechy we wszystkich k populacjach, to w przypadku braku takiej informacji weryfikujemy najpierw hipotezę o równości wszystkich wariancji jednym z trzech testów: Bartletta, Cochran-Coxa lub Hartleya.

Do weryfikacji hipotezy zerowej o równości wartości średnich wszystkich prób (wobec hipotezy alternatywnej, że nie wszystkie wartości przeciętne są równe) stosujemy test oparty na zdefiniowanej powyżej statystyce F . Za zbiór wartości krytycznych przyjmujemy przedział $<F(1-\alpha, k-1, n-k), +\infty)$, gdzie $F(1-\alpha, k-1, n-k)$ jest kwantylem rozkładu F Snedecora.

Jeżeli obliczona wartość statystyki F należy do przedziału krytycznego wówczas hipotezę zerową o równości wartości przeciętnych należy odrzucić. Natomiast jeżeli obliczona wartość statystyki F nie należy do przedziału krytycznego wtedy nie ma podstaw do odrzucenia hipotezy zerowej.

Przykład 18.

Zmierzono długości świecenia trzech typów żarówek, otrzymując następujące czasy w godzinach:

typ 1: 1802, 1992, 1854, 1880, 1761, 1900;

typ 2: 1664, 1755, 1823, 1862;

typ 3: 1877, 1710, 1882, 1720, 1950.

Z poziomem ufności $\alpha = 95\%$ należy zweryfikować hipotezę, że wartości przeciętne czasów świecenia żarówek wszystkich typów są jednakowe (hipotezą alternatywną jest, że wartości te nie są jednakowe).

Najpierw należy przy pomocy testu Bartletta (gdyż próby są o różnej liczebności) zweryfikować hipotezę, że równe są wariancje dla trzech prób (test jest opisany w rozdziale 4.2.1).

Z obliczeń testu Bartletta otrzymujemy, że przy poziomie ufności $\alpha = 95\%$ nie ma powodu do odrzucenia hipotezy o równości wariancji. Zakładając więc prawdziwość hipotezy $\sigma_1 = \sigma_2 = \sigma_3$ możemy przejść do weryfikacji hipotezy $H_0: m_1 = m_2 = m_3$.

Z obliczeń otrzymujemy: $\bar{x}_1 = 1864.8$, $\bar{x}_2 = 1776.0$ a $\bar{x}_3 = 1827.8$, a co za tym idzie $\bar{x} = 1828.8$ oraz

$$\sum_{i=1}^3 (\bar{x}_i - \bar{x})^2 n_i = 18932.36 \text{ a zatem obliczona statystyka } F = \frac{1/2 \cdot 18932.36}{8405.5} = 1.126.$$

Z tablic rozkładu Snedecora odczytujemy $F(\alpha, k-1, n-k) = F(0.95, 2, 12) = 3.88$, zatem przedziałem krytycznym jest $<3.88, +\infty)$. Obliczona wartość nie należy do tego przedziału. Nie ma więc podstaw do odrzucenia weryfikowanej hipotezy o równości wartości średnich tych trzech prób.

5.2.3. ANOVA dla dwóch prób.

W celu lepszego zrozumienia koncepcji analizy wariancji prześledźmy przykład pomiarów przeprowadzonych dla dwóch próbek, za każdym razem powtórzonych pięciokrotnie. Wyniki przedstawia tabela 14.

Tabela 14.
Przykład wyników pomiarów.

Zmienna	Powtórzenie	Y_i	\bar{Y}_i	
1	19 14 15 17 20	85	17	\bar{Y}_1
2	23 19 19 21 18	100	20	\bar{Y}_2
		185	18.5	$\bar{Y}_{..}$

5.3. Weryfikacja hipotez dotyczących wartości przeciętnych w przypadku klasyfikacji podwójnej.

Z zagadnieniem tym mamy do czynienia np. w przypadku badania twardości stopu w skład którego wchodzi dwa metale A i B a ich zawartość w stopie decyduje o twardości. Możemy zatem podzielić nasze obserwacje na r klas ze względu na wartość cechy A oraz na p klas ze względu na wartość cechy B. Wszystkie obserwacje są zatem podzielone na rp grup. Ograniczymy się w naszych rozważaniach do przypadku, gdy w każdym z wariantów dokonano takiej samej liczby l pomiarów badanej cechy X. Badana cecha w każdej z rp populacji ma rozkład $N(m, \sigma^2)$ o nieznanymi parametrach. Weryfikacja hipotez dotyczących wartości przeciętnych m w oparciu o wyniki prób pobranych z każdej populacji. Wyniki tych prób można przedstawić w tabeli (tabela 15), przy czym:

$$\bar{x}_{i.} = \frac{1}{lp} \sum_{j=1}^p \sum_{k=1}^l x_{ijk} \quad \text{dla } i = 1, \dots, r.$$

$$\bar{x}_{.j} = \frac{1}{lr} \sum_{i=1}^r \sum_{k=1}^l x_{ijk} \quad \text{dla } j = 1, \dots, p.$$

$$\bar{x} = \frac{1}{rpl} \sum_{j=1}^p \sum_{k=1}^l \sum_{i=1}^r x_{ijk}$$

$$m_{i.} = \frac{1}{p} \sum_{j=1}^p m_{ij} \quad \text{dla } i = 1, \dots, r.$$

$$m_{.j} = \frac{1}{r} \sum_{i=1}^r m_{ij} \quad \text{dla } j = 1, \dots, p.$$

$$m = \frac{1}{rpl} \sum_{i=1}^r \sum_{j=1}^p m_{ij}$$

Tabela 15.
Tabela wyników dla klasyfikacji podwójnej

Czynnik	A_1	A_2	...	A_r	średnia
B_1	x_{111}, \dots, x_{1l1}	x_{211}, \dots, x_{2l1}	...	x_{r11}, \dots, x_{rl1}	$\bar{x}_{.1}$
B_2	x_{121}, \dots, x_{1l2}	x_{221}, \dots, x_{2l2}	...	x_{r21}, \dots, x_{rl2}	$\bar{x}_{.2}$
...
B_p	x_{1p1}, \dots, x_{1lp}	x_{2p1}, \dots, x_{2lp}	...	x_{rp1}, \dots, x_{rlp}	$\bar{x}_{.p}$
średnia	$\bar{x}_{1.}$	$\bar{x}_{2.}$...	$\bar{x}_{r.}$	\bar{x}

W przypadku tego modelu możemy zweryfikować hipotezy:

1. o równości wartości średnich we wszystkich rp populacjach:

$$H_0: m_{ij} = m \text{ dla } i = 1, \dots, r; j = 1, \dots, p.$$

2. o równości wszystkich wartości przeciętnych m_i badanej cechy poddanej działaniu czynnika A w r wariantach, bez uwzględniania wpływu czynnika B:

$$H_0: m_{1.} = \dots = m_{r.} \text{ dla } i = 1, \dots, r.$$

3. o równości wszystkich wartości przeciętnych m_j badanej cechy poddanej działaniu czynnika B w p wariantach, bez uwzględniania wpływu czynnika A:

$$H_0: m_{.1} = \dots = m_{.p} \text{ dla } j = 1, \dots, p.$$

4. że odchylenie wartości przeciętnej m_{ij} od ogólnej wartości przeciętnej m jest równe sumie efektów czynnika A i czynnika B:

$$H_0: m_{ij} - m = (m_{i.} - m) + (m_{.j} - m).$$

Gdy hipoteza ta nie jest spełniona, czyli gdy nie zachodzi zjawisko addytywności oddziaływania efektów rozpatrywanych czynników, wtedy mówimy, że zachodzi współdziałanie tych czynników, a wyrażenie $m_{ij} - m_{i.} - m_{.j} + m$ jest miarą tego współdziałania.

Podobnie jak w przypadku klasyfikacji pojedynczej sumę kwadratów odchyleń od ogólnej średniej q rozkładamy na sumę czterech składników:

$$\begin{aligned} q &= \sum_i^p \sum_j^r \sum_k^1 (x_{ijk} - \bar{x})^2 = 1p \sum_i^p (\bar{x}_{i.} - \bar{x})^2 + 1r \sum_j^r (\bar{x}_{.j} - \bar{x})^2 + \\ &+ 1 \sum_i^p \sum_j^r (\bar{x}_{ij.} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2 + \sum_i^p \sum_j^r \sum_k^1 (x_{ijk} - \bar{x}_{ij.})^2 = \\ &= q_A + q_B + q_{AB} + q_R \end{aligned}$$

Analogicznie również zmienne losowe $Q_A / (r-1)$, $Q_B / (p-1)$, $Q_{AB} / (r-1)(p-1)$, $Q_R / rp(l-1)$, $Q / (rp(l-1))$ są nieobciążonym estymatorem nieznannej wariancji σ^2 , natomiast zmienne Q_A / σ^2 , Q_B / σ^2 , Q_{AB} / σ^2 , Q_R / σ^2 , Q / σ^2 są niezależne i mają rozkłady χ^2 .

Hipotezę H_{01} (93) dotyczącą równości wartości przeciętnych badanej cechy we wszystkich populacjach weryfikujemy na podstawie rp niezależnych próbek, każda o liczebności 1, stosując test analizy wariancji w przypadku klasyfikacji pojedynczej.

Do weryfikacji hipotezy zerowej H_{02} (94) stosujemy test oparty na statystyce zdefiniowanej następująco:

$$F_A = \frac{Q_A / r - 1}{Q_R / rp(l-1)} \quad (79)$$

która ma rozkład Snedecora o $r-1$ i $rp(l-1)$ stopniach swobody.

Do weryfikacji hipotezy zerowej H_{03} stosujemy statystykę:

$$F_B = \frac{Q_B / p - 1}{Q_R / rp(l-1)} \quad (80)$$

o $p-1$ i $rp(l-1)$ stopniach swobody.

Natomiast do weryfikacji hipotezy H_{04} stosujemy statystykę:

$$F_{AB} = \frac{Q_{AB} / (r-1)(p-1)}{Q_R / rp(l-1)} \quad (81)$$

o $(r-1)(p-1)$ i $rp(l-1)$ stopniach swobody.

Zbiorami krytycznymi tych testów są przedziały $<F_{kr}(1-\alpha, w, v), +\infty)$, gdzie w i v oznaczają odpowiednie stopnie swobody.

Przykład 19.

Z trzech różnych wydziałów pewnej uczelni wylosowano po $l = 4$ studentów z każdego roku studiów i obliczono średnią ocen uzyskaną przez każdego studenta w ostatnim semestrze. Uzyskano rezultaty przedstawione w tabeli 16.

Tabela 16.

Średnie ocen wybranych studentów.

Rok studiów	Wydział		
	A	B	C
I	3.6, 4.1, 3.1, 2.4	3.1, 2.5, 3.3, 3.8	2.7, 4.2, 2.9, 3.7
II	2.8, 4.3, 3.8, 3.0	3.9, 2.6, 3.2, 3.3	3.0, 4.4, 3.9, 3.1
III	3.2, 4.1, 4.8, 4.0	3.4, 2.9, 4.1, 2.8	4.0, 3.3, 3.4, 3.0
IV	3.2, 3.9, 4.2, 3.6	3.6, 4.4, 2.8, 2.9	3.7, 5.0, 2.6, 3.4
V	4.0, 4.0, 3.5, 3.8	4.0, 3.0, 4.5, 3.7	3.0, 3.8, 4.8, 3.5

Zakładając, że średnie uzyskiwanych ocen mają rozkłady normalne o tej samej wariancji na poziomie ufności $\alpha = 95\%$ zweryfikować następujące hipotezy:

- wartości przeciętne średnich ocen dla studentów różnych wydziałów są jednakowe;
- wartości przeciętne średnich ocen dla różnych lat studiów są jednakowe;
- wartości przeciętne ocen średnich dla pierwszych dwóch lat są jednakowe.

W tym przypadku mamy $r = 3$ (wydziały) oraz $p = 5$ (liczba lat studiów). Po obliczeniach otrzymujemy wyniki przedstawione w tabeli 17.

Tabela 17.

Wyniki obliczeń dla danych z tabeli 16.

p	r			Razem
	1	2	3	
1	3.300	3.175	3.375	3.283
2	3.475	3.259	3.600	3.342
3	4.025	3.300	3.425	3.583
4	3.725	3.675	3.674	3.692
5	3.825	3.800	3.775	3.800
Razem	3.670	3.440	3.568	3.550

Następnie obliczamy sumy kwadratów odchyień:

$$q_A = 0.5365, \text{ stopni swobody mamy } df = 2, \text{ zatem } q_A / df = 0.26815$$

$$q_B = 2.3797 \quad df = 4 \quad q_B / df = 0.59492$$

$$q_{AB} = 0.06980 \quad df = 8 \quad q_{AB} / df = 0.00872$$

$$q_R = 18.4050 \quad df = 45 \quad q_R / df = 0.4090$$

$$q = 21.3908 \quad df = 59$$

Obliczone na tej podstawie statystyki F mają następujące wartości:

$$F_A = 0.26815 / 0.4090 = 0.6556$$

$$F_B = 0.59492 / 0.4090 = 1.4546$$

i mają odpowiednio 2 i 45 oraz 4 i 45 stopni swobody. Suma kwadratów wszystkich odchyleń $q = 21.3908$ ma 59 stopni swobody. Po podzieleniu przez odpowiednie stopnie swobody otrzymujemy średnie kwadratów równe $q_A / r - 1 = 0.26815$, $q_B / p - 1 = 0.59492$, $q_{AB} / (r-1)(p-1) = 0.008725$, $q_R / rp(1-1) = 0.409$.

Aby zweryfikować hipotezę o równości wartości przeciętnych średnich ocen dla studentów różnych wydziałów należy wykorzystać statystykę F_A . Wartością krytyczną jest dla tej statystyki $F(\alpha, r-1, rp(1-1)) = F(0.95, 2, 45) = 3.21$. Ponieważ obliczona wartość statystyki jest mniejsza od wartości krytycznej, zatem brak jest podstaw do odrzucenia tej hipotezy zerowej.

Dla zweryfikowania hipotezy o równości wartości przeciętnych średnich ocen dla różnych lat studiów korzystamy ze statystyki $F_B = 1.4546$. Wartością krytyczną jest dla tej statystyki $F(\alpha, p-1, rp(1-1)) = F(0.95, 4, 45) = 2.59$. Ponieważ i tym razem obliczona wartość statystyki jest mniejsza od wartości krytycznej, więc również brak jest podstaw do odrzucenia hipotezy zerowej.

Dla zweryfikowania trzeciej hipotezy zerowej o równości wartości przeciętnych średnich ocen dla pierwszych dwóch lat studiów musimy policzyć wartości przeciętne dla tych lat $\bar{x}_1 = 3.3$ i $\bar{x}_2 = 3.5$ oraz odchylenia standardowe dla tych wartości $s_1 = 0.082$ i $s_2 = 0.125$, musimy bowiem najpierw dla tych rozkładów udowodnić hipotezę o równości wariancji stosując jeden z testów parametrycznych. W tym przypadku zastosujemy test Cochra. W tym celu obliczamy statystykę:

$$G = \max(s_i^2) / \sum_{i=1}^n s_i^2 = 0.015625 / 0.022349 = 0.6991.$$

Wartość krytyczna tej statystyki odczytana z tablic testu Cochra wynosi 0.9057 i jest większa od obliczonej, zatem nie ma podstaw do odrzucenia hipotezy o równości wariancji rozkładów ocen średnich dla dwóch pierwszych lat studiów.

Możemy zatem przystąpić do weryfikowania hipotezy zerowej. Wartość przeciętnej oceny dla tych lat studiów $\bar{x} = 3.4$, obliczone dla tego przypadku sumy kwadratów odchyleń: $q = 0.24$ (stopień swobody jest równy 1), $q = 7.95$ (liczba stopni swobody 22), wobec tego wartość statystyki F dla tej hipotezy:

$$F_A = 0.2400 / 0.3614 = 0.6641$$

Krytyczna wartość statystyki $F_{A_{kr}} = F(0.95, 1, 22) = 4.30$ i jest większa od wartości obliczonej, zatem wniosek brzmi: nie ma podstaw do odrzucenia hipotezy zerowej (czyli możemy przyjąć, że wartości przeciętne ocen dla dwóch pierwszych lat studiów są jednakowe).

5.4. Schemat kwadratu łacińskiego LQ-r (n=1) III

Zapoznamy się teraz z analizą wariancji dla trzech efektów mieszanych. Zastosowany w tym przypadku schemat kwadratu łacińskiego (Latin sQuare - LQ) wymaga co najmniej tylu powtórzeń ile jest czynników wpływających na wynik i nie jest wygodny w przypadku dużej liczby takich czynników. Najczęściej stosuje się go w przypadku 4 - 8 czynników z pojedynczym pomiarem w każdej komórce tabeli. Na przykład przy sprawdzaniu poprawności obliczeń wykonywanych przy pomocy trzech kalkulatorów (oznaczonych literami A, B, C) równocześnie przez trzech różnych "operatorów" nasza tabela może wyglądać tak jak tabela 18.

Tabela 18.

Schemat kontroli obliczeń (numer rzędu odpowiada kolejnemu wykonaniu obliczeń).

		Kolumna ("operator")		
		I	II	III
	1	B	A	C
R	2	C	B	A
z	3	A	C	B
ą	4	B	C	A
d	5	C	A	B
	6	A	B	C

Kolumny i rzędy w tym eksperymencie mogą odnosić się także do kolejności w jakiej wykonuje się pomiary i położenia układów doświadczalnych. W naszym przykładzie możemy wyeliminować wpływ czasu (a tym samym zmęczenia "operatora") jak i samego operatora na błąd eksperymentu (pozostawiając jedynie błędy wnoszone przez same kalkulatory). Jest to przykład podwójnego kwadratu łacińskiego 3x6.

Każdy eksperyment, który ma być analizowany przy pomocy schematu kwadratu łacińskiego rozpoczyna się od losowego tworzenia odpowiedniego kwadratu łacińskiego dla przewidywanej liczby oddziaływań (czynników) i układów pomiarowych zastosowanych do przeprowadzenia eksperymentu (najczęściej liczba kolumn, wierszy i oddziaływań jest w tego typu eksperymencie jednakowa).

Stopnie swobody dla wyników tak zaprojektowanego doświadczenia są następujące:

całkowity	$df_{tot} = pr - 1$
wierszy	$df_{row} = r - 1$
kolumn	$df_{col} = p - 1$
czynników	$df_{tr} = n - 1$
błędu (reszty)	$df_{error} = (r-1)(p-1) - (n-1) =$ $df_{tot} - df_{row} - df_{col} - df_{tr}$

Algorytm przeprowadzania analizy jest następujący:

1°. Obliczamy współczynnik korekcyjny:

$$C = \frac{\left(\sum_{ijk} x_{ijk} \right)^2}{rp} \quad (82)$$

2°. Obliczamy sumę kwadratów dla wierszy:

$$SSR = \frac{\sum_i x_{i..}^2}{p} - C \quad (83)$$

oraz średni kwadrat dla wierszy:

$$MSR = \frac{SSR}{df_{row}} \quad (84)$$

następnie obliczamy sumę kwadratów dla kolumn:

$$SSC = \frac{\sum_j x_{.j.}^2}{r} - C \quad (85)$$

oraz średni kwadrat dla kolumn:

$$MSC = \frac{SSC}{df_{col}} \quad (86)$$

aż wreszcie obliczamy sumę kwadratów dla czynników:

$$SST = \frac{\sum_k x_{...k}^2}{r} - C \quad (87)$$

oraz średni kwadrat dla czynników:

$$MST = \frac{SST}{df_{tr}} \quad (88)$$

3°. obliczamy całkowitą sumę kwadratów:

$$SS = \sum_{ijk} x_{ijk}^2 - C \quad (89)$$

oraz sumę kwadratów resztkową:

$$SSE = SS - SSR - SSC - SST \quad (90)$$

i średni kwadrat resztkowy:

$$MSE = \frac{SSE}{df_{error}} \quad (91)$$

4°. Obliczamy statystyki:

dla wierszy:

$$F_{row} = \frac{MSR}{MSE} \quad (92)$$

dla kolumn:

$$F_{col} = \frac{MSC}{MSE} \quad (93)$$

dla czynników:

$$F_{tr} = \frac{MST}{MSE} \quad (94)$$

i porównujemy je z odpowiednimi wartościami krytycznymi dla założonego poziomu ufności.

Przykład 20.

W doświadczeniu nad nawożeniem pól zastosowano następujące czynniki (nawozy): A - $(\text{NH}_4)_2\text{SO}_4$, B - NH_4NO_3 , C - $\text{CO}(\text{NH}_2)_2$, D - $\text{Ca}(\text{NO}_3)_2$, E - NaNO_3 , F - NoN (brak nawożenia). Nawozy stosowane są w jednakowych dawkach (w kg/ha).

W pierwszym etapie przeprowadzono losowanie kwadratu łacińskiego 6x6, wynik przedstawia tabela 19.

Tabela 19.

Kwadrat łaciński 6x6.

	4	2	5	1	3	6
1	B	D	E	F	A	C
3	C	E	A	D	F	B
5	A	F	C	B	E	D
4	D	A	F	C	B	E
2	F	B	D	E	C	A
6	E	C	B	A	D	F

Wyniki tak zaplanowanego eksperymentu (osiągnięte plony buraków cukrowych w q/ha) przedstawia tabela 20. Na tej podstawie możemy przedstawić sumaryczne i średnie wyniki (plony) dla poszczególnych czynników - zawiera je tabela 21.

Tabela 20.
Wyniki eksperymentu.

	I	II	III	IV	V	VI	$\sum_j x_{ij}$
I	F 28.2	D 29.1	A 32.1	B 33.1	E 31.1	C 32.4	186.0
II	E 31.0	B 29.5	C 29.4	F 24.8	D 33.0	A 30.6	178.3
III	D 30.6	E 28.8	F 21.7	C 30.8	A 31.9	B 30.1	173.9
IV	C 33.1	A 30.4	B 28.8	D 31.4	F 26.7	E 31.9	182.3
V	B 29.9	F 25.8	E 30.3	A 30.3	C 33.5	D 32.3	182.1
VI	A 30.8	C 29.7	D 27.4	E 29.1	B 30.7	F 21.4	169.1
$\sum_i x_{ij}$	183.6	173.3	169.7	179.5	186.9	178.7	$\sum_{ij} x = 1071.7$

Tabela 21.
Plony dla poszczególnych nawozów.

Czynnik	A	B	C	D	E	F
$\sum_k x_{\bullet\bullet k}$	186.1	182.1	188.9	183.8	182.2	148.6
$\bar{x}_{\bullet\bullet k}$	31.0	30.4	31.5	30.6	30.4	24.8

Stopnie swobody: całkowity $df_{tot} = pr - 1 = 35$
wierszy $df_{row} = r - 1 = 5$
kolumn $df_{col} = p - 1 = 5$
czynników $df_{tr} = n - 1 = 5$
błędu (reszty) $df_{error} = (r-1)(p-1) - (n-1) = 35 - 5 - 5 - 5 = 5 \cdot 5 - 5 = 20$.

Współczynnik korekcyjny $C = (1071)^2 / 36 = 31\ 903.91$, natomiast suma kwadratów dla wierszy $SSR = \frac{186.0^2 + \dots + 169.1^2}{6} - 31903.91 = 32.19$ i średni kwadrat dla wierszy $MSR = 32.19 / 5 = 6.438$, zaś dla kolumn

$SSC = \frac{183.6^2 + \dots + 178.7^2}{6} - 31903.91 = 33.67$ oraz $MSC = 33.67 / 5 = 6.734$. Dla czynników odpowiednio suma

$SST = 185.77$ a średnia $MST = 37.154$.

Całkowita suma kwadratów $SS = 28.2^2 + \dots + 29.1^2 - 31903.91 = 32185.79 - 31903.91 = 281.88$ a resztkowa suma kwadratów $SSE = 30.25$ i resztkowy średni kwadrat $MSE = 30.25 / 20 = 1.513$. Obliczone na tej podstawie statystyki:

dla wierszy - $F_{row} = 6.438 / 1.513 = 4.26$

dla kolumn - $F_{col} = 6.734 / 1.513 = 4.45$

dla czynników - $F_{tr} = 37.154 / 1.513 = 24.56$

Wszystkie te statystyki mają (5, 20) stopni swobody, więc dla poziomu ufności $\alpha = 95\%$ znaleziona w tablicach wartość krytyczna $F_{kr} = F(0.95, 5, 20) = 2.71$.

Wszystkie wartości obliczone są większe od wartości krytycznej, co oznacza, że wszystkie trzy źródła wpływu na badaną cechę populacji generalnej należy uważać za bardzo istotne. Na podstawie tego możemy wysunąć wniosek, że istnieją zasadnicze różnice w wynikach występujące pomiędzy wierszami, kolumnami i czynnikami.

Dalszym etapem analizy naszych danych jest rozdzielenie zmiennych (średnich). Na podstawie wyniku naszego eksperymentu możemy odpowiedzieć na szereg pytań:

- 1) czy nawożenie wpływa na wzrost plonów (wydzielamy czynnik F) ?;
- 2) czy nawozy organiczne są lepsze od nieorganicznych ?;
- 3) czy $\text{NH}_4\text{-N}$ jest lepszy niż $\text{NO}_3\text{-N}$?;
- 4) czy $(\text{NH}_4)_2\text{SO}_4$ jest lepszy niż NH_4NO_3 ?;
- 5) czy $\text{Ca}(\text{NO}_3)_2$ jest lepszy niż NaNO_3 ?.

Pytań takich może być oczywiście więcej, w zależności od tego jakie czynniki czy grupy czynników chcemy ze sobą porównywać.

Po postawieniu pytań tworzymy tabelę zawierającą współczynniki oddziaływania c_i oraz sumaryczne wartości plonów $x_{..k}$ osiągnięte dla poszczególnych czynników (tabela 21), będącą podstawą dalszej analizy. Współczynniki oddziaływania posumowane w każdym wierszu powinny dawać zero.

Następnie korzystając z zależności (100), lub w przypadku $df=1$ ze wzoru:

$$SS_{obs} = \frac{\left(\sum_i c_i x_{..k} \right)^2}{r \sum_i c_i^2} \quad (95)$$

obliczamy sumy odchyłeń dla poszczególnych czynników, np. dla przypadku porównywania nawożenia z jego brakiem ($df=1$) obliczamy:

$$SS_{obs}(\text{NoN} - \text{N}) = \frac{(5 \cdot 148.6 - 186.1 - 182.1 - 188.9 - 183.8 - 182.2)^2}{6 \cdot (25 + 1 + 1 + 1 + 1 + 1)} = 180.200$$

dzieląc tę liczbę przez odpowiednią ilość stopni swobody otrzymujemy średnie kwadraty resztkowe MS_{obs} , a na tej podstawie obliczamy wartość statystyki korzystając ze wzoru:

$$F = \frac{MS_{obl}(\text{źródło})}{MS_{obl}(\text{reszta})} \quad (96)$$

Tabela 22.

Wyniki badań po rozdzieleniu wartości.

	Czynnik	NoN	$(\text{NH}_4)_2\text{SO}$	NH_4NO_3	$\text{CO}(\text{NH}_2)_2$	$\text{CA}(\text{NO}_3)_2$	NaNO_3
Porównanie	$x_{..k}$	148.6	186.1	182.1	188.9	183.8	182.2
NoN - N		+5	-1	-1	-1	-1	-1
Organiczny-Nieorgan.		0	-1	-1	+4	-1	-1
$\text{NH}_4\text{-N} - \text{NO}_3\text{N}$		0	+1	+1	0	-1	-1
$(\text{NH}_4)_2\text{SO}_4 - \text{NH}_4\text{NO}_3$		0	+1	-1	0	0	0
$\text{Ca}(\text{NO}_3)_2 - \text{NaNO}_3$		0	0	0	0	+1	-1

czyli w naszym przypadku:

$$F(\text{NoN} - \text{N}) = 180.20 / 1.513 = 119.10$$

wyniki pozostałych obliczeń zawiera tabela 23.

Znajdujemy teraz wartości krytyczne odpowiednich statystyk F:

$$F_{kr} = F(0.95, 1, 20) = 4.35$$

$$F_{kr} = F(0.95, 5, 20) = 2.71$$

Tylko w przypadku porównywania wyników dla testu "bez nawozu - z nawozem" wartość obliczona jest większa od wartości krytycznej, a zatem nie ma podstaw do odrzucenia takiej hipotezy. W pozostałych przypadkach wybór źródła odchyleń nie ma istotnego znaczenia.

Tabela 23.

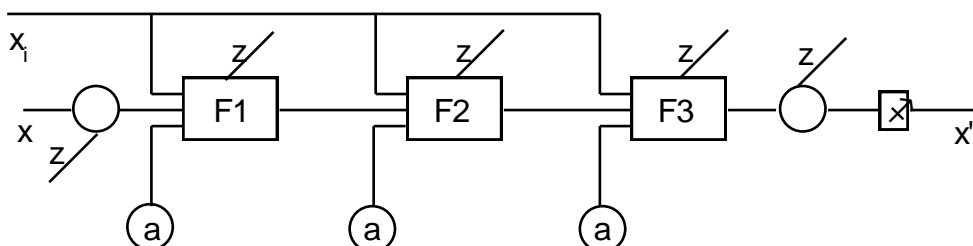
Wyniki obliczeń.

Źródło odchyleń	df	SS _{obs}	MS _{obs}	F	F _{kr}
czynniki	5	185.77	37.15	24.56	2.71
NoN-N	1	180.20	180.20	119.10	4.35
Organ. - Nieorgan.	1	3.816	3.816	2.52	4.35
NH ₄ -N - NO ₃ N	1	0.202	0.202	0.13	4.35
(NH ₄) ₂ SO ₄ - NH ₄ NO ₃	1	1.334	1.334	0.88	4.35
Ca(NO ₃) ₂ - NaNO ₃	1	0.213	0.213	0.14	4.35
reszta	20	30.25	1.513		

6. TEORIA BŁĘDÓW

- rodzaje błędów;
- charakterystyka metrologiczna;
- klasa niedokładności;
- wnioskowanie w teorii błędów;
- sposoby redukcji błędów arytmetycznych;
- obliczanie błędów w praktyce.

Kiedy opublikujemy jakiś uzyskany przez siebie wynik eksperymentalny, staje się on własnością publiczną i może być wykorzystany przez innych do ich własnych celów: do sprawdzenia przewidywań teoretycznych, zaprojektowania eksperymentu czy też zbudowania jakiegoś przyrządu. Osoba korzystająca z naszych wyników musi jednak wiedzieć, czy są one wystarczająco dokładne do jego zastosowań. Również wyciągając na podstawie takich danych wnioski musimy wiedzieć, jak dalece możemy mieć do nich zaufanie. Każdy eksperyment, każdy pomiar i prawie każda operacja składowa pomiaru daje wyniki obciążone różnymi typami błędów (na rys. 6.1 schematycznie przedstawione są błędy wpływające na wynik pomiaru pewnej wielkości). Właśnie z tego powodu, podstawowym obowiązkiem eksperymentatora jest ocena błędów przeprowadzonego pomiaru i podanie rzetelnej informacji o ich wielkości (bez przesady w jedną, jak i w drugą stronę). Tym bardziej informacja taka powinna być rzetelna, że na podstawie samej oceny błędów pomiarów (a nie tylko danych eksperymentalnych), można wyciągać również pewne istotne wnioski. Należy przy tym pamiętać o tym, że cele doświadczalne często określają dopuszczalne granice błędu.



Rys. 6.1. Źródła błędów w układzie pomiarowym (x - wielkość mierzona, x_i - wielkość wpływająca, a - parametry konstrukcyjne, z - zakłócenia, ϕ - procedura analizy danych).

Nie tylko końcowe wyniki doświadczenia powinny być uzyskiwane z odpowiednią precyzją, dotyczy to także (o może przede wszystkim) wszystkich wielkości pośrednich uzyskiwanych w trakcie eksperymentu, bowiem zwykle dokonuje się pomiaru wielu różnych wielkości pierwotnych, i dopiero drogą ich kombinacji uzyskuje wynik końcowy, przy czym jego dokładność określona jest poprzez błędy wielkości pierwotnych, czyli tak zwane błędy cząstkowe. Widzimy zatem, że pojęcie błędu pomiaru jest bardzo istotne dla eksperymentatora.

6.1. Rodzaje błędów.

Definicja błędu może być sformułowana w różny sposób:

- Błąd jest to różnica między stanem danej wielkości a stanem rzeczywistym jej wielkości - definicja ta wyraża błąd w dziedzinie rzeczywistości powstający przy tworzeniu obrazu rzeczywistości za pomocą pomiarów;
- Jeżeli wartość wielkości wynosi x , a przyjęto wartość x' , to różnica $\Delta = x' - x$ jest błędem - jest to najlepsza definicja w przypadku eksperymentów myślowych;
- Błąd jest to różnica między wartością wielkości a wartością poprawną tej samej wielkości - definicja ta wyraża błąd w dziedzinie abstrakcji.

Pojęcie błędu nie jest związane z odwzorowaniem stanu wielkości w wartość, lecz wyraża istnienie pewnej odległości między dwoma stanami lub wartościami bez określenia przyczyny zaistnienia tej różnicy. Zakładamy, że odwzorowaniem stanu rzeczywistego jest $|$ wartość poprawna $|$ wielkości. Założenie to powoduje najwięcej komplikacji w teorii błędów, gdyż stan rzeczywisty wielkości nie jest tak naprawdę znany, a więc nieznaną jest i wartość poprawna.

Nie można udowodnić, że wartość poprawna jest funkcją wartości rzeczywistej, dlatego oba pojęcia służą do zdefiniowania błędu. W konsekwencji błędy pomiaru można wyrazić w dziedzinie stanów lub w dziedzinie ich obrazów. W dalszych rozważaniach nie będziemy określać, czy błąd jest wyrażony w dziedzinie stanów czy w dziedzinie obrazów, sam błąd określając jako różnicę między wynikiem pomiaru a poprawnym stanem mierzonej wielkości czyli pomiędzy stanem na wejściu układu pomiarowego a wartością na wyjściu tego układu.

Rozróżnia się trzy rodzaje miar błędów: błędy prawdziwe, błędy umownie prawdziwe i błędy graniczne. Każdy z nich może być przedstawiany w trzech postaciach: błędu bezwzględnego, błędu względnego i błędu unormowanego (czyli zakresowego lub sprawdzonego)

W modelach matematycznych układów pomiarowych (metod pomiarowych) każde źródło błędu występuje jako pewna zmienna losowa lub proces losowy o znanych charakterystykach (rysunek 1).

Podział błędów dokonywany jest ze względu na:

- 1) charakter błędu - wyróżnia się tu błąd systematyczny i błąd przypadkowy (błąd losowy);
- 2) warunki pomiaru - w warunkach odniesienia błąd nazywa się błędem podstawowym, w innych warunkach występują ponadto błędy dodatkowe;
- 3) charakter mierzonej wielkości - według tego kryterium błędy pomiarowe dzielimy na błędy statyczne i błędy dynamiczne;
- 4) fizyczne przyczyny powstawania błędu - wyróżnia się tu m.in. błąd wzorcowania, błąd niestałości, błędy kwantowania, błędy próbkowania, błędy zliczania itd.

Definicje błędów systematycznego i przypadkowego należą w metrologii do elementarnych pojęć i zapewne spotkali się z nimi państwo wcześniej. Przypomnimy sobie zatem te pojęcia w skrócie.

Błąd systematyczny jest to błąd, który przy wielokrotnym wykonywaniu pomiaru tej samej wielkości w tych samych warunkach ma wartość stałą lub zmienia się według znanego prawa. Wszystkie pozostałe błędy określa się jako przypadkowe.

Błąd systematyczny jest wartością oczekiwaną błędu przy pomiarze tej samej wartości wielkości mierzonej w tych samych warunkach. Warunkiem określenia wartości błędu systematycznego jest znajomość wartości wielkości wpływających na pomiar, wartości wielkości mierzonej oraz czasu T , który upłynął od wzorcowania układu. Fizyczny sens tego rodzaju błędu nie ulega zmianie w porównaniu z definicją klasyczną: są to błędy powstające z przyczyn losowych, mające jednak jednakowy znak lub niesymetryczny względem zera rozkład prawdopodobieństwa. Na przykład błąd koincydencji ("spiętrzanie" wyników pomiarów) przy zliczaniu impulsów za pomocą licznika Geigera-Müllera ma zawsze znak ujemny, natomiast błąd tła przy pomiarach izotopowych ma znak dodatni. Wartość oczekiwana tych błędów jest różna od zera i wartość tę należy nazwać błędem systematycznym.

Błędy przypadkowe zawsze występują w eksperymencie i powodują rozrzut kolejnych odczytów wokół rzeczywistej wartości mierzonej wielkości (oczywiście, gdy występuje błąd systematyczny, pomiary układają się wokół pewnej, przesuniętej względem rzeczywistej, wartości). Błędy przypadkowe można wykryć drogą powtarzania pomiarów, przy okazji poprawiając precyzję pomiarów, korzystając ze średniej wartości serii pomiarów.

Wartość oczekiwana błędu przypadkowego jest równa zeru, właściwość ta nie zawsze jest zgodna z sensem fizycznym błędów przypadkowych.

Przy omawianiu błędów wygodnie jest wprowadzić na wzór terminologii stosowanej w literaturze zachodniej rozróżnienie pomiędzy pojęciami dokładność i precyzja. Wynik pomiaru określamy wówczas jako dokładny, gdy jest on wolny od błędów systematycznych, natomiast jako precyzyjny, gdy jego błąd przypadkowy jest bardzo mały.

Błąd podstawowy jest to błąd występujący, gdy układ pomiarowy znajduje się w warunkach odniesienia (wielkości wpływające mają wówczas wartości stałe). Składnik systematyczny błędu podstawowego nazywa się błędem wzorcowania, a składnik przypadkowy - błędem niestałości.

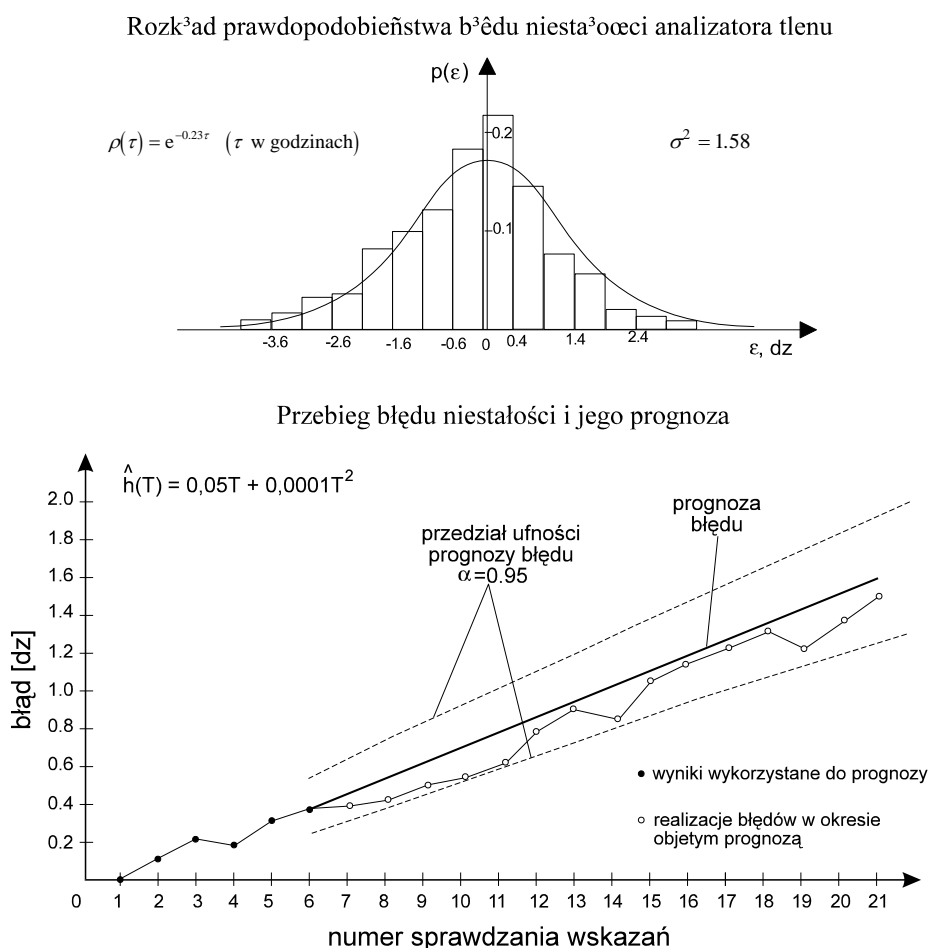
Inaczej patrzy się na błąd wzorcowania przy wyznaczaniu tego błędu w sposób teoretyczny, na podstawie danych wzorca, warunków i sposobu wzorcowania. Wyznacza się wówczas graniczne wartości tego błędu. Przez dobór odpowiedniej procedury wzorcowania możliwe jest zatem zmniejszenie błędu wzorcowania poprzez uśrednienie błędów (lub wygładzanie), natomiast w wyniku aproksymacji zmierzonej charakterystyki do prostszego lub wygodniejszego z jakichś względów kształtu, możemy zwiększyć błąd.

Błąd dodatkowy jest spowodowany odmiennością warunków w jakich dokonuje się pomiaru, od warunków w jakich przeprowadzono wzorcowanie układu pomiarowego. Można go także zdefiniować jako błąd spowodowany zastosowaniem uproszczonej procedury wzorcowania w stosunku do procedury "naturalnej", czyli procedury z korekcją błędów systematycznych (błąd ten jest więc błędem metody wzorcowania).

Błąd dynamiczny można zdefiniować dwójako:

- a) jest to błąd spowodowany odmiennymi niż idealne właściwościami dynamicznymi układu pomiarowego;
 b) jest to błąd spowodowany zastosowaniem statycznej procedury wzorcowania dla układu przy pomocy którego dokonujemy pomiarów dynamicznych mierzonej wielkości.

Każdy układ pomiarowy (i wzorzec) powinien posiadać swoją charakterystykę metrologiczną. Pod tym pojęciem rozumie się ogół wiadomości o błędach układu pomiarowego przedstawionych w pewien uporządkowany sposób. Różnorodność przyrządów i różnorodność ich zastosowań powodują, że nie istnieje jednolity sposób opracowywania charakterystyk metrologicznych. Istnieją co prawda pewne zalecenia normatywne, producenci aparatury korzystają jednak z różnych wariantów charakterystyk, upraszczając je w dość dowolny sposób. Na przykład dla termomagnetycznego analizatora tlenu charakterystyka (zaczepnięta z pracy J. Piotrowskiego [3]) wygląda tak jak na rysunku 2.



Rys. 2. Charakterystyka metrologiczna termomagnetycznego analizatora tlenu.

6.2. Klasa niedokładności.

Klasa niedokładności przyrządu (systemu pomiarowego) określa nam zakres (przedział), którego nie może przekroczyć błąd podstawowy w całym zakresie pomiarowym. Wyrażenie własności metrologicznych za pomocą klasy niedokładności jest najbardziej zbliżone do miary Lebesgue'a zbioru wartości wielkości mierzonej w pewnym stanie o wartości b najlepiej odwzorowującej stan a, równej wskazaniu przyrządu, z tą różnicą, że definicja klasy niedokładności nie zawiera ustalenia poziomu ufności. Brak ten można uzupełnić na podstawie metodyki sprawdzania wskaźników odpowiednio interpretując wyniki sprawdzania. Sprawdzanie wskaźników przeprowadza się w warunkach odniesienia w punktach skali przyrządu opisanych cyframi. Punkty te oznaczają się x_j , $j = 1, \dots, k$. Wynikiem sprawdzania jest zbiór błędów $\{\Delta_j\}$. Układ pomiarowy spełnia wymagania klasy niedokładności, gdy $|\Delta_j| \leq \Delta_{\text{dop}}$ dla $j = 1, \dots, k$. Zakłada się następujące własności błędów:

- 1) kolejne wyniki sprawdzania wskazań są zrealizowane niezależnie;
- 2) prawdopodobieństwo, że błąd Δ_j przekroczy wartość dopuszczalną wynosi P;
- 3) błąd w każdym z k punktów sprawdzania nie przekracza przedziału $2 \varepsilon = 2 \Delta_{\text{dop}}$;
- 4) poza przedziałem błędu dopuszczalnego może się znaleźć dopiero $k + l - y$ wynik sprawdzania ($l \geq 1$), przy czym punkt ten nie jest w procesie sprawdzania wskazań realizowany.

Wyniki sprawdzania można opisać za pomocą rozkładu geometrycznego, którego gęstość prawdopodobieństwa wyraża wzór:

$$p(U = k+1) = (1-P)^k P \quad (97)$$

Wartość przeciętna zmiennej losowej U wynosi $E(U) = 1 / P$, a wariancja $\text{var}(U) = (1-P) / P^2$. Najbardziej efektywnym nieobciążonym estymatorem jest:

$$P' = 1 / (k+1) \quad (98)$$

a gdy $l > 1$, to $P' < 1/(k+1)$. Liczba punktów wskazań przyrządów pomiarowych podlegających sprawdzeniu przy kontroli wymagań klasy niedokładności jest z góry znana. Zwykle jest ich 6-16, czyli prawdopodobieństwo zawiera się w granicach $\langle 0.059, 0.200 \rangle$. Z drugiej strony wyznaczając wartość P według wzoru (2) wariancję estymowanego parametru, możemy obliczyć korzystając z twierdzenia Cramera-Rao o dolnym ograniczeniu wariancji:

$$\text{var}(P') = P^2(1-P) / (k+1) \quad (99)$$

Tak więc cechy konstrukcyjne przyrządu i metodyka sprawdzania wskazań pozwalają ustalić prawdopodobieństwo przekroczenia przez błąd wartości dopuszczanej przez klasę niedokładności oraz ocenić prawdopodobieństwo, że przy znamiennej ufności $1-P$ w k punktach sprawdzania wskazań wartość dopuszczalna błędu nie zostanie przekroczona.

6.3. Wnioskowanie w teorii błędów.

Myślą przewodnią teorii błędów jest znalezienie sposobów wykonywania pomiarów i obliczania błędów w warunkach fizycznych ograniczeń powodujących niemożliwość równoczesnego wyznaczania błędu i stosowania układu pomiarowego do właściwych pomiarów. Rozdzielenie tych czynności w czasie powoduje przypadkową zmienność właściwości układu pomiarowego, występowanie stanów nieustalonych itp., zatem na wykonującym badania spoczywa obowiązek przewidywania zmian błędów w czasie i w zmieniających się warunkach pomiarów.

Naturalnym sposobem postępowania jest w tym wypadku wydzielenie niezmiennego składnika błędu (lub łatwego do przewidywania np. zmieniającego się według znanej zależności, prawa) oraz reszty błędu stanowiącej składnik losowy. Idea ta leży u podstaw podziału na błędy systematyczne i przypadkowe. Systematyczny błąd charakteryzuje się tym, że jego wartość w chwili pomiaru jest znana, bądź jako wartość stała, bądź jako wartość do obliczenia. Z oczywistych względów staramy się, aby jak największa część błędu miała znane, powtarzalne właściwości. I w ten sposób dochodzi do wyróżnienia błędu wzorcowania, błędów dodatkowych, błędów dynamicznych, bowiem prawa zmienności tych błędów mają inne uwarunkowania, inaczej się te błędy wyznacza i inaczej oblicza. Natomiast od sposobu wykonywania pomiarów zależy udział błędów przypadkowych (przypadkowych nie tylko z natury).

Jedną z idei sformułowanych przez Gaussa jest wykorzystanie właściwości błędu przypadkowego: jednakowego prawdopodobieństwa błędów dodatnich i ujemnych, co przy sumowaniu powoduje redukcję błędów. Konsekwencją tego jest zmniejszenie błędu przypadkowego przez wykonanie serii pomiarów i wzięcie jako wyniku końcowego pomiarów wartości średniej:

$$\Delta_{\text{śr}} = \frac{1}{n} \sum \Delta_i \quad \Rightarrow \quad \sigma_n^2 = \frac{1}{n^2} \sum \sigma_i^2 = \frac{1}{n} \sigma^2$$

Rozszerzeniem tej idei jest stosowanie rozmaitych filtrów, głównie w przypadku zmieniającej się wartości wielkości mierzonej.

Inną ideą ograniczania błędów przypadkowych jest adiustacja układu. Adiustacją nazywa się czynności regulacji układu pomiarowego sprowadzające błąd do zadanej wartości (najczęściej do zera). Oczywiście, zamiast adiustacji można przeprowadzić wzorcowanie.

Można dokonać adjustacji według wartości chwilowej błędu - w chwili $t = t_0$ dokonuje się korekcji wskazań układu pomiarowego, tak aby wartość błędu po korekcji wynosiła zero. Adjustacja według wartości chwilowej błędu niestałości powoduje zmniejszenie błędu przypadkowego przez okres, dla którego $r(t-t_0) > 0.5$.

Przykład 21.

Gdy współczynnik korelacji $\rho(\tau) = \exp(-a \tau)$ okres ten wynosi $T < 1/a \ln 2 = 0.693 / a$. Dla analizatora tlenu, dla którego $a = 0,23$ okres $T < 3$ godzin.

Możemy przeprowadzić adjustację według wartości średniej błędu. Przed adjustacją należy wyznaczyć wartości błędu podstawowego w kolejnych momentach t_i ($i = 1, 2, \dots, N$), potem obliczyć wartość średnią i dokonać adjustacji

$$r_j = -\frac{1}{N} \sum_{i=1}^N [g_0(t_i)\varepsilon(t_i) + h_0(t_i)]$$

gdzie g_0, h_0, ε wyrażają błąd podstawowy $\{\Delta\} = g(x_1, \dots, x_m, T)\{\varepsilon\} + h(x_1, \dots, x_m, T)$. Już dla niewielkich N wariancja błędu przypadkowego nieznacznie przewyższa wariancję składnika losowego błędu (np. dla $N = 4$ mamy $\sigma_{\Delta} = 1.12 g_0$). Zatem wystarczy kilka sprawdzeń wskazań przeprowadzonych co $T > 3 / a$ (czyli dla analizatora tlenu odstęp czasu powinien być większy od 12 godzin).

Innym sposobem adjustacji jest adjustacja z predykcją błędów. Na podstawie pomiaru błędów przed rozpoczęciem właściwych pomiarów, wyznaczamy zależność błędu od czasu i ekstrapolujemy go na czas pomiaru, tak aby w chwili pomiaru błąd był równy zero.

Tutaj nasuwa się uwaga: Przy planowaniu eksperymentu nie należy do końca opierać się na wnioskach dostarczanych przez statystykę matematyczną, gdyż podaje ona swoje reguły w oparciu o idealne przypadki (założenia), nie znajdujące nigdy w pełni odbicia w rzeczywistości. Statystyka jest także bezradna jeśli chodzi o analizę błędów systematycznych.

6.4. Praktyczne obliczanie błędów.

Przejdźmy teraz do praktycznego obliczania błędów. Najczęściej nasze obliczenia zaczynamy od błędu standardowego, który dla serii n pomiarów wyraża się wzorem:

$$\sigma = \sqrt{\frac{\sum_i x_i^2 - \frac{1}{n} \left(\sum_i x_i \right)^2}{n(n-1)}} \quad (100)$$

W przypadku wyniku końcowego Z będącego kombinacją wyników pierwotnych A i B musimy zastosować następujące reguły działania na błędach:

W przypadku, gdy $Z = A + B$ lub $Z = A - B$, błąd względny ΔZ wartości końcowej obliczamy jako:

$$(\Delta Z)^2 = (\Delta A)^2 + (\Delta B)^2 \quad (101)$$

W przypadku, gdy $Z = A \cdot B$ lub $Z = A / B$, błąd względny ΔZ wartości końcowej obliczamy korzystając z zależności:

$$\left(\frac{\Delta Z}{Z} \right)^2 = \left(\frac{\Delta A}{A} \right)^2 + \left(\frac{\Delta B}{B} \right)^2 \quad (102)$$

lub według niektórych podręczników dla iloczynu korzystając ze wzoru:

$$\left(\frac{\Delta Z}{Z} \right) = \left(\frac{\Delta A}{A} \right) + \left(\frac{\Delta B}{B} \right) \quad (103)$$

zaś dla ilorazu ze wzoru:

$$\left(\frac{\Delta Z}{Z}\right) = \left(\frac{\Delta A}{A}\right) - \left(\frac{\Delta B}{B}\right) \quad (104)$$

Jeżeli $Z = A^n$ zależność dla błędów względnych jest następująca:

$$\frac{\Delta Z}{Z} = n \frac{\Delta A}{A} \quad (105)$$

Natomiast, gdy $Z = \ln A$ mamy zależność:

$$\Delta Z = \frac{\Delta A}{A} \quad (106)$$

a dla dowolnego logarytmu ($Z = \log_b A$) korzystamy ze wzoru:

$$\Delta Z = \frac{1}{\ln(b)} \frac{\Delta A}{A} \quad (107)$$

W przypadku $Z = \exp A$ stosujemy wzór:

$$\frac{\Delta Z}{Z} = \Delta A \quad (108)$$

Obliczenia błędów dla funkcji nie wymienionych powyżej wykonuje się poprzez rozwinięcie tej funkcji w szereg Taylora. Jeśli funkcja będzie postaci $Z = f(x_1, x_2, \dots, x_n)$, gdzie x_1, x_2, \dots, x_n są wynikami naszych pomiarów obciążonych odpowiednio błędami $\Delta x_1, \Delta x_2, \dots, \Delta x_n$, to błąd bezwzględny ΔZ obliczmy korzystając z zależności:

$$|\Delta Z| = \sum_{i=1}^n \left| f'(x_1, x_2, \dots, x_n) \Delta x_i \right| \quad (109)$$

Gdy mamy funkcję wielu zmiennych: $Z = f(x_1, x_2, \dots, x_n)$ wówczas funkcję tę rozwijamy w szereg Taylora i uwzględniając dwa pierwsze wyrazy otrzymujemy, że:

$$|\Delta Z| = \left[\frac{\partial}{\partial x_1} \Delta x_1 + \frac{\partial}{\partial x_2} \Delta x_2 + \dots + \frac{\partial}{\partial x_n} \Delta x_n \right] Z \quad (110)$$

Niektórzy eksperymenciści, po wyznaczeniu błędu w zwykły sposób, zwiększają go o pewien arbitralny czynnik, aby uwzględnić wszystkie nieznanne źródła błędów. Takie subiektywne zawyżanie jest mało przydatne dla innych, może spowodować zamieszanie lub przesłonić rzeczywiste różnice (i ich przyczyny!).

Przyjęte jest podawanie absolutnej wartości błędu końcowego. Końcową wartość mierzonej wielkości i jej błąd należy podawać z jednakową liczbą cyfr ułamka dziesiętnego, których nie powinno być więcej niż cyfr znaczących. Przy czym wszystkie obliczenia błędów powinny być wykonywane z dokładnością do jednej lub (co najwyżej) dwóch cyfr znaczących.

W ogólności eksperyment należy tak zaplanować, aby żadna z mierzonych wielkości nie dawała większego wkładu do błędu od pozostałych.

7. ZARZĄDZANIE ZBIORAMI DANYCH

- *obserwacje nietypowe,*
- *wyniki odbiegające - sposoby wykrywania i eliminacji,*
- *obróbka zestawów danych,*
- *porównywanie danych z różnych laboratoriów.*

7.1. Wprowadzenie

Posługiwanie się prostymi zestawami danych jest bardzo ważną operacją dla wielu inżynierów i naukowców. Dane powinny być przeglądane i weryfikowane pod względem występujących w nich błędów i zgodności wartości. W wielu przypadkach konieczna jest odpowiedź na pytanie czy wszystkie otrzymane wartości są jednakowe, czy też występują jakieś wyniki mocno odbiegające od pozostałych. Sposoby łączenia ze sobą danych pochodzących z różnych źródeł lub otrzymanych w różnym czasie przez to samo źródło (ten sam układ pomiarowy) stanowią dla badacza istotny problem. Ten rozdział opisuje techniki, które stosuje się do tych celów.

7.2. Obserwacje nietypowe.

Problem danych odbiegających znacznie od pozostałych często prześladowuje zarówno ludzi wykonujących pomiary, jak i późniejszych użytkowników tych danych. Mając zestaw danych składający się z grup pomiarów które mają idealnie tę samą wartość, trudno jest twierdzić, jak bardzo powinny różnić się pojedyncze wartości, aby mogły być uznane za "obce", a nie za ekstremalne odchylenia od wartości średniej.

Jak wspomniano na początku, podejrzane jest, gdy po uporządkowaniu od najmniejszych do największych wartości, jedna lub obie skrajne wartości znacznie odbiegają od średniej. Podobna sytuacja jest, gdy na wykresie zauważymy punkty znacznie odbiegające od gładkiej krzywej. Nasuwa się tu pytanie, dlaczego przywiązujemy taką wagę do nietypowych obserwacji? Otóż są dwie ważne przyczyny związane ze statystyką: po pierwsze - średnia obliczona dla zestawu danych zawierających "obce" wartości jest błędna (jest to szczególnie ważne dla małych zestawów danych). Niestety, w przypadku małych zbiorów danych, bez znajomości charakterystyki tych danych pochodzącej z innych źródeł, usuwanie błędnych wartości przynosi mniejsze efekty. Po drugie - znacznie odbiegające wartości wpływają w znaczny sposób na obliczoną wartość odchylenia standardowego, znów szczególnie w przypadku małych zbiorów danych, oraz wtedy gdy do estymacji odchylenia standardowego zastosuje się granice.

Odbiegające wartości pomiarów są jeszcze interesujące z innego powodu - mogą wskazać ludziom wykonującym pomiary zarówno błędy, niezręczności jak i awarie sprzętu pomiarowego. Na tej podstawie można określić w jaki sposób można usprawnić układ pomiarowy lub metodykę pomiaru. Każde pojawienie się odbiegającej wartości powinno wywoływać krytyczne przejrzenie całego procesu pomiarowego, który dostarczył błędny wynik. W pierwszym rzędzie należy oczywiście sprawdzić obliczenia, później błędy transkrypcji i dekodowania a dopiero na końcu poszukiwać ewentualnego uszkodzenia w układzie pomiarowym. Gdy i tę możliwość odrzucimy, pozostaje nam tylko jakiś rodzaj niewytłumaczalnego błędu, lub... zmierzylimy nie tę próbkę, którą zamierzaliśmy. Istnieje szereg procedur pozwalających, w oparciu o prawa statystyczne, podjąć decyzję o odrzuceniu lub pozostawieniu podejrzanego wyniku pomiaru.

7.3. Prawo Grubego Błędu.

Jeżeli wiemy jaka powinna być wartość odchylenia standardowego dla naszych wyników pomiarów możemy w prosty sposób określić, który z pomiarów należy odrzucić korzystając z obliczenia dla tego punktu wartości wyrażenia:

$$M = \frac{|x_i - \bar{x}|}{\sigma^2} \quad (111)$$

gdzie x_1 - "podejrzana wartość". Jeżeli otrzymamy $M > 4$, to punkt taki należy odrzucić.

Prawo grubego błędu jest uproszczonym (zgrubnym) testem t, dla którego poziom ufności może być oszacowany jako nie mniejszy niż 0.999 w przypadku prawidłowego określenia odchylenia standardowego, w innych przypadkach będzie bliższy 0.95. Ten test jest rzadko stosowany, przeważnie gdy nie można zastosować innych procedur. Często odchylenie standardowe w takim przypadku oblicza się po odrzuceniu wszystkich podejrzanych punktów, a dopiero potem przeprowadza ich weryfikację. Test ten jest szczególnie przydatny w przypadku, gdy przygotowujemy wykres zbioru wyników pomiarowych. Po odrzuceniu złych punktów możemy garficznie lub metodą najmniejszych kwadratów poprowadzić krzywą najlepiej odwzorowującą położenia punktów pomiarowych. Odchylenia standardowe dla narysowanych punktów lub poprowadzonej pomiędzy nimi krzywej powinny pozwolić na określenie, czy odrzucone punkty mieszczą się w zakresie $M < 4$. Jeżeli tak jest, musimy włączyć je do zbioru i ponownie poprowadzić krzywą. W praktyce można zastąpić odchylenie standardowe średnim błędem kwadratowym, bez specjalnego pogorszenia dokładności testu. Na przykład, podczas skalowania przyrządu otrzymano zestaw danych, które naniesiono na wykres i ręcznie poprowadzono przez nie krzywą. Zauważono wówczas, że jeden z punktów leży znacznie dalej od poprowadzonej linii niż pozostałe. Odchylenie od wykreślonej linii jest równe 0.06, natomiast średni błąd kwadratowy wynosi 0.012, zatem $M = 5$ i zgodnie z prawem grubego błędu pomiar ten należy odrzucić.

7.4. Test Dixona.

Test ten służy również do znajdowania nietypowych wartości pomiarowych w seriach pomiarów i jest łatwy w użyciu ze względu na prostotę obliczeń koniecznych do wykonania. Zakłada się, że nie znamy wartości średniej zmierzonej serii ani odchylenia standardowego dla niej, a zbiór zmierzonych wartości jest dla nas jedynym źródłem informacji. Przyjmujemy jednak założenie, że wartości te mają rozkład normalny. Aby przeprowadzić test Dixona punkty pomiarowe układa się od najmniejszego do największego, następnie dla założonego poziomu ufności oblicza się krytyczny iloraz Dixona r , przy czym w zależności od liczby punktów pomiarowych w serii stosuje się różne wzory (patrz tabela 24). Na przykład dla $n = 8..10$, gdy podejrzany jest x_n obliczamy r_{11} , które jest równe $(x_n - x_{n-1}) / (x_n - x_2)$. Następnie w tablicach sprawdzamy wartości krytyczną r_{kr} dla założonego poziomu ufności, i jeżeli obliczona wartość jest większa niż wartość krytyczna przedstawiona w tabeli, usuwamy punkt pomiarowy.

Tabela 24.

Wzory stosowane w teście Dixona.

n	dla podejrzanego	wzór
1 .. 7	x_1	$r_{10} = \frac{x_2 - x_1}{x_n - x_1}$
	x_n	$r_{10} = \frac{x_n - x_{n-1}}{x_n - x_1}$
8 .. 10	x_1	$r_{11} = \frac{x_2 - x_1}{x_{n-1} - x_1}$
	x_n	$r_{11} = \frac{x_n - x_{n-1}}{x_n - x_2}$

Przykład 22.

Otrzymano następującą serię pomiarów pewnej wielkości x : 9, 12, 12, 13, 13, 14, 15. Wartość $x_1 = 9$ jest podejrzana w tej serii. Obliczamy r_{10} , gdyż seria składa się z 7 pomiarów i otrzymujemy $r_{10} = 3/6 = 0.5$. Dla poziomu ufności 95% z tablic odczytujemy $r_{kr} = 0.507$, zaś dla poziomu ufności 90% mamy $r_{kr} = 0.434$, co oznacza, że dla pierwszego poziomu ufności punkt pomiarowy należy pozostawić, zaś w drugim przypadku należy go odrzucić. Gdy zamiast $x_1 = 9$, będziemy mieli $x_1 = 8$, wtedy wartość ilorazu $r_{10} = 4/7 = 0.571$ i również dla poziomu ufności 95% ten punkt pomiarowy należałoby odrzucić. Należy przy tym koniecznie pamiętać, że przyjmując poziom ufności 95% dopuszczamy ryzyko wystąpienia 5% mylnych decyzji!

Po odrzuceniu punktu należy sprawdzić, czy jeszcze jakiś punkt nie jest podejrzany, choć dwukrotne wystąpienie błędnych punktów w serii pomiarowej jest mało prawdopodobne. Wystąpienie większej ilości punktów znacznie odbiegających od pozostałych wyników pomiarów zgodnie z testem Dixona wskazuje na zastosowanie złej metody

pomiarowej. Należy jednak zawsze pamiętać, że dane reprezentują czyjś czas i pieniądze, i nie można ich niefrasobliwie odrzucać, mimo że stosowanie testów jest łatwe i szybkie.

Dobry analityk (naukowiec, inżynier) powinien wykorzystywać dane pomiarowe takie jakie są, choć "obce" dane powinny być ostrzeżeniem, że układ pomiarowy ma problemy, które powinny być rozwiązane innymi sposobami niż poprzez odrzucanie części wyników pomiarów.

7.5. Test Grubbsa.

Innym szeroko stosowanym testem jest test Grubbsa, który wymaga obliczenia odchylenia standardowego, a zatem jest trochę bardziej pracochłonny niż test Dixona. Jednak nawet stosunkowo proste kalkulatory mogą wykonać te obliczenia. Sposób postępowania jest następujący:

- 1) porządkujemy punkty pomiarowe od najmniejszego do największego;
- 2) decydujemy, który punkt: pierwszy czy ostatni jest podejrzany;
- 3) obliczamy średnią wartość próby \bar{x} oraz odchylenie standardowe s używając wszystkie dane;
- 4) obliczamy parametr T w następujący sposób:
 - a) gdy podejrzany jest punkt x_1

$$T = (\bar{x} - x_1) / s \quad (112)$$

- b) gdy podejrzany jest punkt x_n

$$T = (x_n - \bar{x}) / s \quad (113)$$

- 5) wybieramy poziom ufności dla testu i porównujemy obliczoną wartość T z wartością krytyczną T_{kr} podaną w tabeli 25.

Jeżeli obliczona wartość przekracza poziom krytyczny należy usunąć punkt pomiarowy z serii.

Przykład 23.

Weźmy serię pomiarów składającą się z liczb 9, 12, 12, 13, 13, 14, 15 i podobnie jak poprzednio weźmy pod uwagę liczbę 9 jako podejrzany punkt. Obliczmy $X = 12.57$, $s = 1.90$, $T = 1.87$. Dla poziomu ufności 95% krytyczna wartość $T_{kr} = 1.938$, czyli jest większa od obliczonej. Zatem punkt należy pozostawić. Dla pierwszego punktu pomiarowego równego 8 mielibyśmy $\bar{x} = 12.43$, $s = 2.22$ i $T = 1.99$, a więc ten punkt należałoby odrzucić.

Test ten możemy zastosować, gdy podejrzana jest jedna z wartości, najmniejsza lub największa, natomiast w przypadku, gdy podejrzane są obie wartości x_1 i x_n należy posłużyć się inną wersją tego testu.

Tabela 25.

Wartości krytyczne T dla próby o wątpliwej obserwacji skrajnej przy poziomie istotności 0.05.

n	T	n	T	n	T	n	T
3	1.15	11	2.23	19	2.53	35	2.82
4	1.46	12	2.29	20	2.56	40	2.87
5	1.67	13	2.33	21	2.58	45	2.92
6	1.82	14	2.37	22	2.60	50	2.96
7	1.94	15	2.41	23	2.62	60	3.03
8	2.03	16	2.44	24	2.64	70	3.09
9	2.11	17	2.47	25	2.66	80	3.14
10	2.18	18	2.50	30	2.75	90	3.18
						100	3.21

Zamiast parametru T obliczamy rozstęp $R = x_n - x_1$ oraz odchylenie standardowe i porównujemy z odpowiednimi tablicami (tabela 26) wartości krytycznych iloraz R/σ . Jeżeli obliczona wartość jest większa od krytycznej, należy odrzucić obie podejrzane wartości.

Przykład 24.

Otrzymano w trakcie pomiarów pewnej wielkości $n = 10$ następujących wartości uporządkowanych od najmniejszej do największej: -20, -5, -2, 2, 5, 5, 5, 6, 6, 28. Podejrzane jest $x_1 = -20$ i $x_{10} = 28$, zatem rozstęp jest równy $R = 48$, zaś odchylenie standardowe dla tego zbioru wynosi $\sigma = 11.9$, obliczony parametr R/σ jest równy 4.03. Dla poziomu ufności 95% odczytana z tablicy wartość krytyczna $(R/\sigma)_{kr} = 3.680$, zatem jest mniejsza od obliczonego stosunku, czyli obie wartości należy odrzucić.

Tabela 26.

Wartości krytyczne R/σ przy poziomie istotności 0.05.

n	R/σ	n	R/σ	n	R/σ	n	R/σ
3	2.00	11	3.80	19	4.43	150	6.18
4	2.43	12	3.91	20	4.49	200	6.38
5	2.75	13	4.00	30	4.89	500	6.94
6	3.01	14	4.09	40	5.15	1000	7.33
7	3.22	15	4.17	50	5.35		
8	3.40	16	4.24	60	5.50		
9	3.55	17	4.31	80	5.73		
10	3.68	18	4.38	100	5.90		

W przypadku gdy po uporządkowaniu stwierdzimy, że podejrzane są dwie skrajne wartości, możemy posłużyć się jeszcze inną odmianą testu Grubbsa. Tym razem obliczamy sumę kwadratów odchyleń od średniej dla całej próby, razem z wartościami wątpliwymi:

$$S^2 = n \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \quad (114)$$

oraz bez nich:

$$S_{1,2}^2 = (n-2) \sum x^2 - \frac{1}{n-2} \left(\sum x \right)^2 \quad (115)$$

i obliczamy stosunek tych wielkości:

$$S_{1,2}^2/S^2$$

a otrzymany wynik porównujemy z odpowiednimi tablicami (na przykład z tabelą 27). Tym razem, obie wartości należy wyeliminować, wtedy gdy wartość krytyczna (dla danego poziomu ufności) jest mniejsza aniżeli wartość obliczona.

Przykład 25.

Otrzymano następujący zestaw wyników: 2.02, 2.22, 3.04, 3.23, 3.59, 3.73, 3.94, 4.05, 4.11, 4.13. Podejrzane wydają się być dwie najmniejsze wartości. Po podstawieniu do wzorów otrzymujemy, że: $S^2 = 5.351$, zaś $S_{1,2}^2 = 1.197$, wobec tego ich stosunek będzie równy $S_{1,2}^2/S^2 = 0.224$.

Dla poziomu ufności $\alpha = 95\%$ i $n = 10$ pomiarów odczytujemy z tablic wartość krytyczną równą 0.2330. Ponieważ wartość krytyczna jest większa od obliczonej, zatem obie najmniejsze wartości należy wyeliminować.

Tabela 27.Wartości krytyczne $S_{1,2}^2/S^2$ przy poziomie istotności 0.05.

n	$S_{1,2}^2/S^2$	n	$S_{1,2}^2/S^2$	n	$S_{1,2}^2/S^2$
4	0.0008	10	0.2305	16	0.4048
5	0.0376	11	0.2666	17	0.4259
6	0.0565	12	0.2996	18	0.4455
7	0.1020	13	0.3295	19	0.4636
8	0.1478	14	0.3568	20	0.4804
9	0.1909	15	0.3818		

7.6. Test Youdena.

Każde laboratorium dąży do tego, aby jego wyniki pomiarowe były porównywalne z podobnymi wynikami otrzymanymi w innych dobrych laboratoriach. Youden wiele lat temu postawił zagadnienie: jak można osądzić równość serii pomiarów ze względu na błąd systematyczny. Inne pytanie postawił Cochran: jak osądzić zestawy danych doświadczalnych ze względu na precyzję pomiarów, tym zagadnieniem zajmiemy się w dalszej części.

Test Youdena polega na ocenieniu wyników osiągniętych w różnych laboratoriach na podstawie wyników kilku testów karuzelowych (cykliczne, przemienne wykorzystanie zasobów danych). Mamy przy tym kilka możliwości: laboratoria otrzymują ten sam materiał do badań i mają zmierzyć określoną wielkość taką samą ilość razy (możliwy jest także tylko jeden pomiar), laboratoria dostają jednakowe zestawy materiałów i dokonują pomiarów w tym samym czasie lub wreszcie materiał krąży określoną ilość razy pomiędzy laboratoriami.

Dla każdego materiału laboratorium uzyskujące najwyższy rezultat otrzymuje jeden punkt, następne otrzymuje dwa punkty itd. Punkty sumuje się i porównuje z tablicami rozkładów prawdopodobieństwa (wszystkie laboratoria powinny wykonać te samą liczbę pomiarów). Oczywiście, jeśli laboratorium ciągle otrzymuje największe lub najmniejsze wyniki, należy wątpić czy jest ono w ogóle wiarygodne. Co jednak należy sądzić o laboratorium, które stosunkowo często dostarcza takich wyników? Youden w tablicach zestawiał zakresy punktów, które powinny być przewidywane w wyniku takiego rankingu przy założonym poziomie ufności. Oczywiście zakres punktów zależy od ilości laboratoriów uwzględnianych w teście oraz ilości materiałów dla których policzono punktacje.

Tabela 28.

Przykładowe wyniki pomiarów.

Lab	p	q	r	s	t
A	11.6	15.3	21.1	19.2	13.4
B	11.0	14.8	20.8	19.3	12.8
C	11.3	15.2	21.0	18.9	12.8
D	10.8	15.0	20.6	19.0	13.3
E	11.5	15.1	20.8	18.6	12.7
F	11.1	14.7	20.5	18.7	13.0
G	11.2	14.9	20.7	18.8	13.2
H	10.9	14.6	20.9	19.1	13.1
I	11.4	14.8	20.9	18.5	12.9
J	11.0	15.0	21.0	18.9	13.3

Przykład 26.

Weźmy pod uwagę laboratoria, które oznaczymy kolejnymi literami alfabetu od A do J. Wykonały one pomiary 5 wielkości, które oznaczymy małymi literami p - t. Wyniki tych pomiarów przedstawia tabela 28, natomiast punktację dla tego zestawu danych przedstawia tabela 29.

Tabela 29.

Punktacja dla wyników z tabeli 28.

Lab	p	q	r	s	t	suma
A	1	1	1	2	1	6
B	7	8	6	1	8	30
C	4	2	2	5	9	22
D	10	4	9	4	2	29
E	2	3	7	9	10	31
F	6	9	10	8	6	39
G	5	6	8	7	4	30
H	9	10	4	3	5	31
I	3	7	5	10	7	32
J	8	5	3	6	3	25

Z tablic odczytujemy, że dla poziomu ufności 95% największe prawdopodobieństwo prawidłowych wyników jest, gdy punkty należą do przedziału od 15 do 40. Zatem istnieje tylko 5% szansa, że laboratoria które mają mniej niż 15 punktów przeprowadziły prawidłowo pomiary, podobnie gdy mają więcej niż 40 punktów. Zatem nasz test wskazuje, że wyniki pochodzące z laboratorium A nie są wystarczająco wiarygodne.

Podobnie jak w przypadku innych testów statystycznych trudności pojawiają się (ze względów probabilistycznych) w przypadku małych zbiorów danych doświadczalnych. Oczywiście najlepiej, gdy oceny dokonuje się na podstawie dużych zbiorów, choć najczęściej nie jest to możliwe. Odbiegające wyniki przeważnie oznaczają błąd lub wykonanie znacznie odbiegające od pozostałych. Choć z doświadczeń wynika, że odstające laboratorium może być jedynym, które otrzymało prawidłowy wynik. Dlatego, podobnie jak w przypadku zbiorów punktów, trzeba dokładnie sprawdzić przyczyny odbiegania wyników. Jedynym sposobem sprawdzenia, gdzie otrzymywane są prawidłowe wyniki, jest przeprowadzenie pomiarów dla materiałów testowych o dobrze znanych wartościach mierzonych wielkości.

7.7. Test Cochra.

Test ten ma za zadanie ocenę precyzji pomiarów pochodzących z różnych laboratoriów. Test F, o którym mówiliśmy poprzednio, jest użyteczny w przypadku gdy dwie wariancje są znacznie różniące się. Test Cochra dla ekstremalnych wartości wariancji stosuje się gdy w grupie wyników pomiarowych jedna z wariancji w sposób ekstremalny odbiega od pozostałych. Jedynym ograniczeniem stosowania tego testu, jest aby każda wariancja musi się opierać na tej samej liczbie stopni swobody. Sposób postępowania jest następujący:

1. obliczamy wariancje i uporządkowujemy je od najmniejszej do największej. Tylko największa wariancja będzie nas dalej interesować.
2. obliczamy stosunek:

$$C = \frac{\max(s_i^2)}{\sum_i^n s_i^2} \quad (116)$$

3. porównujemy otrzymaną wartość z odpowiednimi tablicami kwantyli G (np. z tablicą rzędu 0.95 przedstawioną w tabeli 30), jeśli jest ona większa, przyjmujemy, że przy poziomie ufności 95% jest to więcej, niż maksymalne dopuszczalne odchylenie.

Należy zauważyć, że wartość krytyczna zależy nie tylko od liczby wariacji branych pod uwagę, ale też od liczby powtarzających się wartości użytych do obliczenia każdej z wariacji (trzeba szczególnie pamiętać, że próby muszą być równoliczne).

Tabela 30.

Kwantyle $G(p, k, n)$ rzędu $p = 0.95$ statystyki G Cochran.

n	k										
	1	2	3	4	5	6	7	8	9	10	16
2	0,9985	0,9750	0,9392	0,9057	0,8772	0,8534	0,8332	0,8159	0,8010	0,7880	0,7341
3	,9669	,8709	,7977	,7457	,7071	,6771	,6530	,6333	,6167	,6025	,5466
4	,9065	,7679	,6841	,6287	,5985	,5598	,5365	,5175	,5017	,4884	,4366
5	,8412	,6838	,5938	,5440	,5063	,4783	,4387	,4387	,4241	,4118	,3645
6	0,7808	0,6161	0,5321	0,4803	0,4447	0,4184	0,3980	0,3817	0,3682	0,3568	0,3135
7	,7271	,5612	,4800	,4307	,3974	,3726	,3535	,3384	,3259	,3154	,2756
8	,6798	,5157	,4377	,3910	,3595	,3362	,3185	,3043	,2926	,2829	,2462
9	,6385	,4775	,4027	,3584	,3286	,3067	,2901	,2768	,2659	,2568	,2226
10	,6020	,4450	,3733	,3311	,3029	,2823	,2666	,2541	,2439	,2353	,2032
12	0,5410	0,3924	0,3264	0,2880	0,2624	0,2439	0,2299	0,2187	0,2098	0,2020	0,1737
15	,4709	,3346	,2758	,2419	,2195	,2034	,1911	,1815	,1736	,1671	,1429
20	,3894	,2705	,2205	,1921	,1735	,1602	,1501	,1422	,1357	,1303	,1108
24	,3434	,2354	,1907	,1656	,1493	,1374	,1286	,1216	,1160	,1113	,0942
30	,2929	,1980	,1593	,1377	,1237	,1137	,1061	,1002	,0958	,0921	,0771
40	0,2370	0,1576	0,1259	0,1082	0,0969	0,0887	0,0827	0,0780	0,0745	0,0713	0,0595
60	,1737	,1131	,0895	,0765	,0682	,0623	0,583	,0552	,0520	,0497	,0411
120	,0998	,0632	,0495	,0419	,0371	,0337	,0312	,0292	,0279	,0266	,0218
∞	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0000	,0000

Przykład 27.

Grupa 5 laboratoriów (oznaczonych literami A - E) wykonała po 3 pomiary dla tej samej próbki otrzymując wyniki przedstawione w tabeli 31.

Tabela 31.

Wyniki pomiarów pewnej wielkości fizycznej.

Lab	A	B	C	D	E
s	1.25	1.33	1.05	2.75	1.10
s ²	1.56	1.77	1.10	7.56	1.21

Patrząc na drugi wiersz tabeli 31 widzimy, że wartość wariacji danych z laboratorium D odbiega od pozostałych. Obliczona na podstawie tych danych liczba Cochran $C = 7.56/13.20 = 0.5727$. Natomiast odczytana z tablic liczba krytyczna $C_{kr} = G(0.95, 5, 3) = 0.6838$.

Zatem wartość wariancji znaleziona przez laboratorium D jest zbyt odbiegająca od pozostałych. Gdyby w laboratorium tym $s = 4.0$, wtedy można by uznać ich wyniki za maksymalne dopuszczalne odchylenie. Natomiast od $s = 3.0$ nie możemy tak uważać.

7.8. Test Hartleya.

W przypadku, gdy liczności n próbek pobranych w k laboratoriach są równe, i nie mniejsze niż 5, do weryfikacji wyników możemy zastosować test Hartleya, w którym po obliczeniu wariancji S i uporządkowaniu ich od najmniejszej do największej, wyznaczamy wartość statystyki:

$$H = \max(S_i^2) / \min(S_i^2) \quad (117)$$

W tablicach (tabela 32) dla danych k i n znajdujemy, dla żądanego poziomu ufności, krytyczną wartość parametru $H(p, k, n)$ i porównujemy ją z wartością obliczoną. Podobnie jak w teście Cochran, jeśli obliczona statystyka jest większa od krytycznej, przyjmujemy, że wariancje odbiegają od siebie więcej niż jest to dopuszczalne.

Przykład 28.

Korzystając z danych poprzedniego przykładu (tabela 12), znajdujemy maksymalną wariancję $S_{\max} = 2.75$ zaś minimalną $S_{\min} = 1.05$, czyli liczba Hartleya obliczona według wzoru (59) jest równa $H = 6.86$. Odczytana dla poziomu ufności 95% wartość krytyczna $H(0,95, 5, 3)$ jest równa 50.7 i jest większa od obliczonej wartości H , co oznacza, że maksymalna wariancja w sposób znaczny odbiega od pozostałych. Wniosek jest zatem identyczny jak w przypadku testu Cochran.

Tabela 32.

Kwantyle $H(p, k, n)$ rzędu p statystyki H Hartleya dla poziomu istotności 0.05.

n	k										
	2	3	4	5	6	7	8	9	10	11	12
2	39,0	87,5	142	202	266	333	403	475	550	626	704
3	15,4	27,8	39,2	50,7	62,0	72,9	83,5	93,9	104	114	124
4	9,60	15,5	20,6	25,2	29,5	33,6	37,5	41,1	44,6	48,6	51,4
5	7,15	10,8	13,7	16,3	18,7	20,8	22,9	24,7	26,5	28,2	29,9
6	5,82	8,38	10,4	12,1	13,7	15,0	16,3	17,5	18,6	19,7	20,7
7	4,99	6,94	8,44	9,70	10,8	11,8	12,7	13,5	14,3	15,1	15,8
8	4,43	6,00	7,18	8,12	9,03	9,78	10,5	11,1	11,7	12,2	12,7
9	4,03	5,34	6,31	7,11	7,80	8,41	8,95	9,45	9,91	10,3	10,7
10	3,72	4,85	5,67	6,34	6,92	7,42	7,87	8,28	8,66	9,01	9,34
12	3,28	4,16	4,79	5,30	5,72	6,09	6,42	6,72	7,00	7,25	7,48
15	2,86	3,54	4,01	4,37	4,68	4,95	5,19	5,40	5,59	5,77	5,93
20	2,46	2,95	3,29	3,54	3,76	3,94	4,10	4,24	4,37	4,49	4,59
30	2,07	2,40	2,61	2,78	2,91	3,02	3,12	3,21	3,29	3,36	3,39
60	1,67	1,85	1,96	2,04	2,11	2,17	2,22	2,26	2,30	2,33	2,36
∞	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

8. POPRAWIANIE PRECYZJI POMIARÓW

- sposoby poprawiania precyzji pomiarów;
- analiza kowariancji;
- naturalne ograniczenia możliwości pomiarów.

8.1. Pojęcia podstawowe

Dokonując pomiarów obserwator dąży do uzyskania informacji, które rozszerzą jego wiedzę o badanym zjawisku, procesie czy materiale. Przed przystąpieniem do wykonywania pomiarów ma on zwykle pewne wstępne dane o mierzonej wielkości fizycznej, charakterze badanego procesu (np. o rozkładzie prawdopodobieństwa jakiegoś podlega mierzona wielkość).

Posiadanie informacji o procesie oznacza (pod względem matematycznym) znajomość wielowymiarowego rozkładu gęstości prawdopodobieństwa (wszystkich momentów tego rozkładu). Uzyskany na podstawie wykonanych pomiarów rozkład gęstości prawdopodobieństwa zależy od poziomu organizacji eksperymentu, od stopnia niedokładności pomiaru (wywołanej zarówno czynnikami zewnętrznymi jak i szumami aparatury) oraz czynnikami dezinformacji związanymi z mikroskopową strukturą wielkości mierzonej a także zniekształceniem wielkości mierzonej wywołanym wprowadzeniem do układu czujnika pomiarowego.

Pomiar jest eksperymentem wykonywanym odpowiednimi metodami, za pomocą odpowiednich narzędzi, zorganizowanych w odpowiedni system. Pomiar może być też traktowany jako proces uzyskiwania informacji o obiekcie mierzonym. Informacja ta jest przenoszona przez sygnały, którymi mogą być zarówno zjawiska fizyczne jak i obiekty fizyczne (informacja zawarta jest w ich cechach). Musimy przy tym pamiętać o trzech aspektach sygnału: treści sygnału (niesionej informacji), nośniku sygnału (jest to wspomniane uprzednio zjawisko lub obiekt) oraz o kodzie sygnału (czyli sposobie przyporządkowania informacji przejawom cech nośnika).

Jako precyzję pomiarów będziemy tu rozumieć zdolność pomiarów do wykrycia rzeczywistych efektów oddziaływań. Ogólnie możemy stwierdzić, że im eksperyment jest precyzyjniejszy, tym mniejsze różnice w efektach oddziaływania potrafi on wykryć. Im większa jest zmienność w mierzonej wielkości w wyniku tego samego oddziaływania, tym większy jest błąd związany z różnicą pomiędzy dwoma średnimi i mniej precyzyjny będzie eksperyment pod względem wykrywania różnic w mierzonej wielkości wywołanych oddziaływaniami. Na drugim wykładzie mówiliśmy już, że precyzja, do jakiej należy dążyć w eksperymencie zależy od jego celu. W ogólności to twierdzenie jest prawdziwe, ale w wielu eksperymentach fizycznych zwłaszcza przy pomiarze wielkości fundamentalnych, nie możemy przewidzieć jaka precyzja będzie ostatecznie wystarczająca. Należy wówczas dążyć do osiągnięcia największej dokładności na jaką pozwala badane zjawisko i dostępna technika pomiarowa.

To ostatnie stwierdzenie przypomina nam o istnieniu takiego pojęcia jak dokładność pomiaru. Rozumiemy przy tym, że dokładność pomiaru jest tym większa, im mniejszy jest względny lub bezwzględny graniczny błąd pomiaru (ang. uncertainty). Błąd graniczny definiuje się jako połowę przedziału, w którym prawdopodobieństwo znalezienia się prawdziwej wartości jest nie mniejsze od z góry ustalonej wartości.

Standardowy błąd różnicy pomiędzy dwoma wartościami średnimi wzrasta ze wzrostem różnic między wartościami standardowych odchyień s i maleje ze wzrostem liczby powtórzeń n :

$$s_d = \sqrt{2\sigma^2/n} \quad (118)$$

Zatem metodą poprawienia precyzji pomiarów może być zmniejszenie zmienności (rozrzutu) wewnątrz serii pomiarów mierzonej wielkości lub zwiększenie efektywnej liczby pomiarów (powtórzeń).

Precyzję pomiarów można poprawić poprzez:

1. zwiększenie liczby pomiarów;
2. staranny dobór oddziaływań;
3. doskonalenie techniki pomiarowej;
4. wybór materiału doświadczalnego;
5. wybór przyrządów pomiarowych;
6. wykonując dodatkowe pomiary;

7. zaplanowanie eksperymentów wstępnych i grupowych.

W dalszej części tego rozdziału po kolei omówię te sposoby.

Należy przy tym pamiętać, że istnieją naturalne ograniczenia w pomiarach. Wydawać by się mogło, że gdy używamy odpowiednio czułych przyrządów pomiarowych i zachowujemy odpowiednią staranność, możemy wykonywać pomiary z dowolną precyzją. Nie jest to jednak prawda, bowiem między innymi z praw fizyki (zasada nieokreśloności) i konstrukcji przyrządów (istnieje szereg zjawisk powodujących wzrost przypadkowych fluktuacji - szumów) wynikają granice, poza które nie możemy wykroczyć. Również rozważaniom na ten temat poświęcę końcowy fragment tego rozdziału.

8.2. Zwiększenie liczby pomiarów.

Precyzję pomiarów zawsze można zwiększyć poprzez wydłużenie serii pomiarowych, jednak stopień poprawy zmniejsza się szybko ze wzrostem liczby pomiarów. Na przykład, gdy wykonaliśmy 4 pomiary, aby zwiększyć dwukrotnie precyzję pomiarów (przy założeniu, że obliczymy dwie średnie), należy wykonać aż 16 pomiarów. Wynika to stąd, że poziom ufności (ang. level of significance = LSD)

$$\text{LSD} = t \sqrt{2s^2 / n},$$

a statystyka t maleje ze wzrostem liczby powtórzeń, powodując malenie tempa wzrostu precyzji. Przy planowaniu eksperymentu trzeba być pewnym, że założona liczba powtórzeń pozwoli nam wykryć różnice o interesującej nas amplitudzie. Nie należy wykonywać eksperymentów w przypadku, gdy nie możemy zwiększyć liczby pomiarów w wystarczający sposób, ani nie mamy innego sposobu poprawy dokładności a prawdopodobieństwo uzyskania poprawnych wyników jest zbyt niskie.

Opracowano specjalne tabele³ pozwalające oszacować minimalną liczbę pomiarów koniecznych do wykrycia założonych wielkości różnic. Opierają się one na zależności:

$$r \geq 2[(CV)^2 / D^2](t_1 + t_2)^2 \quad (119)$$

gdzie CV jest współczynnikiem wariancji:

$$CV = s(100)/Y_{\text{sr}} \quad (120)$$

a D jest różnicą, którą chcemy wykryć; t_1 jest tablicową wartością t dla założonego poziomu ufności i stopnia swobody odpowiadającego stopniowi swobody błędu doświadczalnego, zaś t_2 jest wartością tablicową statystyki t Studenta dla liczby stopni swobody dla błędu oraz poziomu ufności równego prawdopodobieństwu $2(1-P)$, gdzie P jest prawdopodobieństwem wykrycia znaczącego wyniku w danej serii pomiarowej.

Aby zastosować tę nierówność należy wpięrow określić liczbę koniecznych zdaniem eksperymentatora pomiarów, i na tej podstawie określić parametr r , a następnie na podstawie tego parametru rozwiązać nierówność ponownie, biorąc najbliższą liczbę całkowitą większą od obliczonego r jako liczbę niezbędnych pomiarów.

Przykład 29.

Chcemy przeprowadzić eksperyment stosując w sposób losowy sześć oddziaływań (hartujemy stal stosując różne szybkości chłodzenia). Zakładamy poziom istotności eksperymentu równy 5% i chcemy wykryć z prawdopodobieństwem 80% zmiany rzędu 10% procent wartości średniej. Z innych doświadczeń otrzymaliśmy wskazówkę, że dobrze przeprowadzony eksperyment powinien mieć współczynnik wariancji równy około 5%. Przyjmujemy, że 6 powtórzeń jest wystarczającą liczbą. Wobec tego dla pierwszych obliczeń $r = 6$, liczba oddziaływań $n = 6$, liczba stopni swobody błędu $df = (r-1)(n-1) = 5 \cdot 5 = 25$.

Odczytane z tablic wartości statystyki t są odpowiednio równe $t_1 = 2.060$ i $t_2 = 0.856$. Podstawiając te wartości do wzoru otrzymujemy nierówność:

$$r \geq 6 \cdot 2 (5 / 10) (2.060 + 0.856) = 4.25$$

Zatem do następnych obliczeń bierzemy $r = 5$, wobec tego $df = (5-1)(6-1) = 20$. Odpowiednie wartości statystyki t z tablic są równe: $t = 2.086$ i $t = 0.860$ zatem rozwiązując ponownie równanie otrzymujemy:

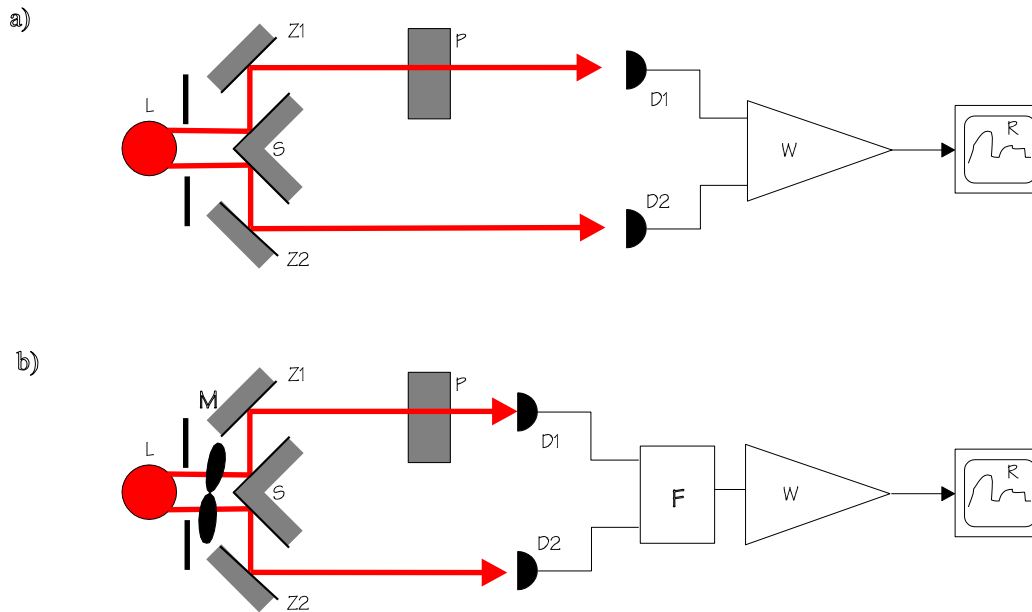
$$r \geq 5 \cdot 2 (5 / 10) (2.086 + 0.860) = 4.34.$$

Wobec tego dla poziomu istotności eksperymentu równego 5% i z prawdopodobieństwem 80% wykrywania zmian rzędu 10% procent wartości średnich wystarczająca liczba powtórzeń pomiarów jest równa 5.

³W.G. Cochran, G.M. Cox, *Experimental Design*, J.Wiley & Sons, Inc., New York 1964, p. 617.

8.3. Dobór oddziaływań.

Skrupulatny dobór oddziaływań (i sposobu ich stosowania) jest nie tylko ważny ze względu na osiągnięcie zamierzonego celu, może mieć również wpływ na precyzję pomiarów. Na przykład rozważmy spektrofotometr którego uproszczony schemat przedstawia rysunek 8.1a. Zastosowanie źródła ciągłego światła monochromatycznego powoduje, że fotodetektor rejestruje także światło pochodzące z innych źródeł światła oraz wszelkiego rodzaju szumy. Gdybyśmy na drodze światła wstawili modulator (np. tarczę ze szczelinami - rysunek 8.1b),



Rys. 8.1. Schemat spektrofotometru a) układ "klasyczny", b) układ z detekcją fazoczułą.

wówczas modulowana wiązka po przejściu przez próbkę zostaje wykryta przez układ fazoczuły i oddzielona od innych promieni świetlnych (pochodzących np. od odbić światła od elementów układu). Układ fazoczuły wyeliminuje także szumy generowane w fotodetektorze.

8.4. Doskonalenie techniki pomiarowej.

Technika eksperymentu obejmuje:

- metodykę projektowania eksperymentu;
- sposoby celowego oddziaływania na badane zjawisko;
- metodykę prowadzenia obserwacji;
- technikę mierzenia;
- metodykę modelowania matematycznego;
- sposoby oceny wiarygodności eksperymentu.

Tak więc technika eksperymentu w naturalny sposób zawiera w sobie technikę mierzenia, jako że mierzenie jest szczególnym przypadkiem eksperymentu, a każdy eksperyment ilościowy opiera się na pomiarach. Zarówno technika eksperymentu, jak i technika mierzenia, czerpią swe uzasadnienie z teorii modelowania, a przede wszystkim modelowania matematycznego. Ostatecznym bowiem celem eksperymentu jest nadanie uzyskanej nowej wiedzy formy modelu matematycznego. Oprócz tego w metrologii model matematyczny występuje w podwójnej roli: 1) jako model narzędzia pomiarowego; 2) jako model obiektu mierzonego. Zaś sam model narzędzia pomiarowego może być zarówno przedmiotem identyfikacji (gdy eksperyment ma na celu określenie jego właściwości) jak i przedmiotem syntezy (gdy budujemy narzędzie o określonych z góry właściwościach).

Zła technika pomiarowa może powiększyć błąd pomiarowy oraz zakłócić efekty oddziaływań. Dobra technika pomiarowa powinna:

- 1) odznaczać się powtarzalnością wyników;
- 2) umożliwiać rzetelny i nieobciążony pomiar efektów oddziaływań;
- 3) zabezpieczać przed grubymi błędami;

- 4) pozwalać na kontrolę zewnętrznych wpływów, w taki sposób aby wszystkie pomiary były jednakowo obciążone.

Wiele czasu pochłania eksperymentatorowi poszukiwanie przyczyn zakłóceń w urządzeniach. Nagromadzone doświadczenie wykazuje, że racjonalne postępowanie przy poszukiwaniu źródeł zakłóceń jest znacznie skuteczniejsze niż przypadkowy dobór prób. Przy nieprawidłowościach wyposażenia użyteczna bywa również jeszcze jedna zasada: po umiejscowieniu zakłóceń zanotować w sposób łatwo dostępny dla innych poszczególne wyróżniające objawy, po których można rozpoznać zakłócenie, gdy wystąpi ono ponownie, najlepsze metody jego umiejscowienia oraz zabiegi, które pozwoliły na jego wyeliminowanie. Inna zasada mówi: po znalezieniu i usunięciu źródła zakłóceń upewnić się, że jego usunięcie jest trwałe. Poszukiwanie źródeł zakłóceń może być ułatwione przez odpowiednie zaprojektowanie samego urządzenia badawczego, zapewniającego jego dostępność, łatwość rozbierania, odpowiednie wyposażenie w przyrządy, punkty kontrolne itp. (patrz punkt 8.6)

8.5. Wybór materiału doświadczalnego.

Dla niektórych typów badań pożądanym jest skrupulatnie dobrany, jednorodny materiał doświadczalny. Przy dokonywaniu selekcji należy jednak pamiętać o populacji generalnej (i jej własnościach), z której wybierany jest materiał, aby poprzez nieodpowiedni dobór materiału nie wpłynąć na osiągnięty wynik.

Byłoby bardzo pożądanym, aby udało się obmyśleć metodę pobierania próbek zapewniającą, że próbka będzie zawierała poszczególne odmiany elementów w tych samych proporcjach, w których występuje ona w całej klasie. Wiele myślno nad tym zagadnieniem, ale nie osiągnięto jego pełnego rozwiązania. Kolejne próbki pobierane z tej klasy będą się z reguły różniły zarówno między sobą, jak i od klasy macierzystej. Różnice te określa się jako odmienność próbki, należy je brać pod uwagę, gdy wyciąga się wnioski na podstawie przebadanej próbki.

W pewnych przypadkach można otrzymać bardziej reprezentatywną próbkę dzieląc klasę na części, tzw. podklasy, na podstawie jakiejś innej znanej cechy, która może mieć wpływ na cechę przez nas badaną, a następnie pobierając próbkę z każdej podklasy. Liczebność każdej podklasy zależy od zmienności badanej właściwości w ramach każdej podklasy. Ta metoda wartszowego pobierania próbek tylko wówczas jest lepsza od postępowania losowego, gdy wiadomo, że elementy każdej podklasy są bardziej jednorodne względem badanej cechy niż cała klasa.

Nie tylko zwiększenie liczby powtórzeń, ale i zwiększenie ilości badanych próbek wpływa na poprawę precyzji pomiarów. Jednak, gdy próbki będą badane pojedynczymi pomiarami, da to mniejszą niż spodziewana poprawę dokładności.

8.6. Wybór i konstruowanie przyrządów pomiarowych.

Przed rozpoczęciem systematycznych pomiarów eksperymentator powinien wiedzieć jak działają poszczególne elementy aparatury, a także powinien upewnić się, że wie co czym steruje i w jaki sposób.

Wybierając przyrząd pomiarowy musimy zwrócić uwagę na następujące wskazówki:

1. W miarę możliwości należy dążyć do jak najdalej idącego uniezależnienia pracy przyrządu od wpływu czynników zakłócających
2. Dogodne i przyjemne warunki pracy operatora wywierają często rozstrzygający wpływ na samo doświadczenie.
3. Przyrząd musi być łatwo dostępny i dawać się zdemontować (zarówno w celu przeprowadzenia naprawy jak i modernizacji).
4. Badaczowi opłaca się zazwyczaj sprawdzanie podstawowych koncepcji w wersji wstępnej przyrządu, a następnie konstruowanie aparatury w wersji ostatecznej.
5. Z jednej strony szaleństwem jest poświęcanie miesięcy lub lat na budowę wyszukanego zestawu aparatury jedynie po to, aby w końcu stwierdzić, że pewna maleńka, kluczowa część nie może być uruchomiona, przez co cała aparatura staje się bezwartościowa – należało tę kluczową część zbudować i przetestować najpierw. Z drugiej strony, równie bezsensowne jest poświęcanie zbyt wiele czasu na kontrole każdej części składowej w trakcie jej budowy, ponieważ czas ten będzie stracony w przypadku, gdy zawiedzie następny element. Najrozsądniejszą metodą wydaje się szukanie części najmniej pewnych i budowanie i sprawdzanie ich w pierwszej kolejności.

8.7. Wybór schematu doświadczalnego - eksperymenty wstępne i grupowe.

Dużą część rozdziałów 1 i 5 poświęciłem wyborowi najlepszemu schematu doświadczalnego. Oczywiście nie byłem w stanie omówić wszystkich znanych i stosowanych schematów badawczych. Coraz większe zainteresowanie badaczy budzą tzw. eksperymenty grupowe (patrz rozdz. 5 - Analiza wariancji) - którym starałem się poświęcić stosunkowo dużo miejsca. Próbowałem także podkreślić znaczenie wyboru odpowiedniego schematu badawczego dla osiągniętej precyzji i wiarygodności pomiarów. Literatura tego tematu jest bardzo bogata, osoby zainteresowane wyborem innych niż przedstawione przeze mnie schematy odsyłam zatem do niej.

W rzeczywistym eksperymencie, w odróżnieniu od pomiarów wykonywanych w laboratoriach studenckich, prawie zawsze wykonuje się pomiary próbne. Mają one na celu:

- 1) skontrolowanie poprawności działania poszczególnych elementów aparatury;
- 2) określenie odpowiedniego zakresu wartości dla mierzonej wielkości oraz stosowanych oddziaływań;
- 3) ocenienie błędów mierzonej wielkości;
- 4) nauczenie badacza techniki (procedury) eksperymentu - znalezienie najlepszego sposobu przeprowadzania pomiaru i zapisywania wyników.

Ogólnie biorąc, wszędzie gdzie to możliwe, wskazane jest dokonywanie obserwacji, bądź pomiarów porównawczych, zamiast polegania na pomiarach bezwzględnych. W większości wypadków porównanie to jest faktycznym przedmiotem doswiadczenia i zazwyczaj lepiej jest przeprowadzić bezpośrednio, niż usiłować dokonać dwóch równoległych pomiarów bezwzględnych.

Chociaż z drugiej strony, korzystnie jest również nadać pomiarom charakter bezwzględny, wyrażając ich wyniki w odniesieniu do powszechnie przyjętych wzorców, ponieważ umożliwia to posłużenie się wynikami uzyskanymi przez różnych eksperymentatorów, bądź to do nowych porównań, bądź do konfrontacji z teorią.

8.8. Analiza kowariancji.

Jedną z technik pozwalającą na zmniejszenie błędów doświadczalnych jest zmniejszenie zmienności zmiennej Y (mierzonej wielkości) związanej z niezależną zmienną X (oddziaływaniem). Technika ta jest nazywana analizą kowariancji. Na przykład w eksperymencie agrotechnicznym występują znaczne różnice w drzewostanie pomiędzy poszczególnymi działkami doświadczalnymi. Jeżeli moglibyśmy w sensowny sposób oszacować jaki plon dawałaby działka, gdyby każda miała jednakowy drzewostan, wzrosła by precyzja pomiaru wpływu różnych czynników na osiągnięte plony. Oszacowanie opierające się na założeniu, że plon jest wprost proporcjonalny do drzewostanu, nie jest sensowne, gdyż wprowadza się obciążenie faworyzujące działki z rzadkim drzewostanem.

Pojęcie kowariancji jest skomplikowane zarówno z punktu widzenia obliczeń koniecznych do przeprowadzenia, jak i z punktu widzenia interpretacji otrzymanych wyników. Algorytm prostych obliczeń jest następujący:

1. przeprowadzamy wstępną analizę wariancji (patrz wykład 8) obliczając: odpowiednie stopnie swobody df oraz sumy kwadratów SSX i SSY , średnie sumy kwadratów MSX i MSY , a także wartość statystyki F ;
2. obliczamy sumy dla poszczególnych oddziaływań (T_{tx} i T_{ty}) i bloków (T_{bx} i T_{by}), oraz współczynnik korekcyjny:

$$C = \frac{\sum X \sum Y}{rn} \quad (121)$$

3. na tej podstawie obliczamy sumy iloczynów dla bloków:

$$SXYB = \frac{\sum T_{bx} T_{by}}{n} - C \quad (122)$$

dla oddziaływań:

$$SXYT = \frac{\sum T_{tx} T_{ty}}{p} - C \quad (123)$$

sumę całkowitą:

$$SXY = \sum X \cdot Y - C \quad (124)$$

oraz sumę resztkową:

$$SXYE = SXY - SXYB - SXYT \quad (125)$$

Aby zrozumieć skąd pochodzi błąd sumy całkowitej i sumy resztkowej należy odwołać się do wykładu na temat analizy wariancji. W podobny sposób możemy wyodrębnić efekty poszczególnych oddziaływań oraz ogólne wartości średnie, pozostawiając tylko błędy systematyczne. Pozostaje nam jeszcze obliczenie odchylenia współczynnika regresji liniowej między zmiennymi X i Y:

$$SSTE = SSYE - SXYE^2 / SSXE$$

które określa nam ile wynosiłaby suma kwadratów Y po usunięciu wpływu X na Y, i ma o jeden stopień swobody mniej niż błąd.. Stopnie swobody dla "oddziaływań z błędem" otrzymujemy dodając do siebie odpowiednie stopnie swobody dla poszczególnych oddziaływań i dla błędu. Wartość sumy kwadratów dla oddziaływań pozbawionych błędu otrzymuje się również poprzez odejmowanie odpowiednich sum kwadratów. Po podzieleniu przez ilość stopni swobody otrzymujemy wartości średnie MSE i MST, a na ich podstawie obliczamy statystykę F:

$$F = MST / MSE \quad (126)$$

Wartość statystyki F będzie nam wzrastać wraz z doskonaleniem techniki pomiarowej. Interpretacja wyników zależy jednak od tego, jak silnie wpływaliśmy na wartości zmiennej niezależnej X w naszym doświadczeniu.

Jeżeli potrafimy wartości X zmieniać tylko w wąskim zakresie, zaś przed wprowadzeniem zmian obserwowaliśmy bardzo duży zakres zmian zmiennej Y, który uległ znacznemu zmniejszeniu po wprowadzeniu zmian, oznacza to, że zmienność Y została wyolbrzymiona w wyniku losowości, a więc zmiany Y muszą być interpretowane bardzo ostrożnie.

Analiza kowariancji, jak widać, jest jeszcze trudniejsza w interpretacji niż analiza wariancji. Zapewne dlatego jest rzadko stosowana w planowaniu i weryfikacji eksperymentu.

8.9. Graniczne możliwości pomiarów.

Podstawowy etap pomiaru - oddziaływanie wzajemne elementu pomiarowego (czujnika) z badanym procesem fizycznym jest związane nieodłącznie z niepełnym odwzorowaniem właściwości procesu i zaburzeniem (w mniejszym lub większym stopniu) przebiegu samego procesu, jego równowagi termodynamicznej, kształtu pól itp., co się wiąże ze stratami informacji. Dalsze straty informacji związane są z formowaniem i przetwarzaniem sygnału pomiarowego przez przyrząd pomiarowy (kanał pomiarowy mikrokomputera - o czym będziemy mówić przy okazji przetwarzania analogowo-cyfrowego). Na przykład, mierząc zwykłym termometrem temperaturę wody w kalorymtrze, wprowadzamy znaczące zmiany w układzie termodynamicznym, naruszamy jego zamkniętość. Niech c - ciepło właściwe wody, T - temperatura początkowa wody, T - temperatura termometru, C - pojemność cieplna termometru. Z bilansu cieplnego mamy, że końcowa temperatura wody z termometrem:

$$T = (c m T + C T) / (C + c m) \quad (127)$$

$$D = T - T = C (T - T) / (C + c m) \quad (128)$$

Wariancję tej wielkości można zapisać w postaci:

$$S = S / 1 + a \quad (129)$$

gdzie $a = (c m) / C$.

Zakładając, że temperatura T_k ma rozkład normalny oraz gęstość prawdopodobieństwa temperatury początkowej jest także normalny, mamy entropię różniczkową temperatury (bezwarunkową):

$$H(T_w) = \log_2 \sqrt{2 \pi e} (S_{T_w}) \quad (130)$$

a entropię warunkową po przeprowadzeniu pomiaru:

$$H(T_w / T_k) = \log_2 \sqrt{2 \pi e} a S_{T_k} \quad (131)$$

Zatem ilość informacji zawartej w wielkości T_k o temperaturze wody możemy obliczyć ze wzoru:

$$I(T_k, T_w) = \log_2 a S_{T_w} / S_{T_r} \quad (132)$$

czyli ilość informacji jest tym większa im większa jest wariancja wielkości mierzonej oraz im mniejsza jest pojemność cieplna termometru. Ze wzrostem nieokreśloności temperatury początkowej termometru maleje ilość uzyskiwanej informacji.

Inną przyczyną strat ponoszonych w początkowym etapie przeprowadzania pomiaru jest skończony czas trwania pomiaru. W praktyce żaden przyrząd nie mierzy wartości chwilowej, lecz wartość uśrednioną w przedziale czasu odpowiadającym czasowi trwania pomiaru.

Straty są tym większe im szybciej zmienia się proces a także im większy jest czas uśrednienia i wariancja procesu. Jeżeli jednak uwzględnimy w wyniku pomiaru błędy addytywne to straty informacji są mniejsze.

Stwierdzenie, że ilość informacji otrzymanych z pomiaru maleje ze wzrostem błędu pomiaru nie jest jednoznaczne ze stwierdzeniem, że istnieje jakiś próg mierzalności dla danego procesu.

Termin małe wielkości określa wielkości mierzone, które są współmierne z wartością błędu lub mniejsze od tej wartości. Przy pomiarach przyrostów, z błędem porównuje się zmianę wielkości mierzonej, a nie jej wartość bezwzględną.

Dokonując pomiarów małych wielkości stosuje się jedną z istniejących metod akumulacji błędów (uśredniania). Poza uśrednianiem wymuszonym, stosuje się również inne metody uśredniania (najprostsza polega na obliczaniu średniej arytmetycznej). Niektóre z nich przedstawię pokrótce. Na pierwszy rzut oka wzór na ilość informacji

$$I = \log_2 \sqrt{\frac{\sigma_x^2}{\sigma_y^2} + 1} \quad (133)$$

(σ_x^2 i σ_y^2 są wariancjami błędu sumarycznego na wejściu i wyjściu bloku - patrz rys. 2) wskazuje, że wielokrotny pomiar stałej (niezmiennej) wielkości fizycznej jest bezsensowny, bowiem wielkość mierzona nie zmienia się, a błąd jest inny przy każdym pomiarze, może wydawać się, że $\sigma_x/\sigma_y = 0$, w wyniku czego otrzymujemy zerową ilość informacji. Twierdzenie takie jest jednak nieprawdziwe, bowiem wariancja σ_x^2 jest obliczana w zbiorze wyników pomiarów, nadto każdy pomiar małej wielkości dostarcza jednak pewną ilość informacji. W przypadku uśredniania polegającego na obliczaniu średniej arytmetycznej (którą możemy stosować gdy rzeczywista wartość wielkości mierzonej nie zmienia się podczas pomiarów) wielkość błędu uśredniania dąży ze wzrostem liczby pomiarów do pewnej wartości oczekiwanej, przy czym żądana dokładność przybliżenia α osiągnana jest dla liczby n pomiarów, którą można oszacować z nierówności:

$$P\{|y_n - \bar{Y}| < \alpha\} > 1 - \frac{\sigma_x^2}{n\alpha^2} \quad (134)$$

Z zależności tej wynika, że zwiększając n możemy otrzymać dowolnie małe odchylenie, w praktyce dokładność przybliżenia jest ograniczona przez błędy obliczeniowe lub w wyniku naruszenia warunku stałości mierzonej wielkości w czasie trwania serii pomiarów.

9. KORELACJA I REGRESJA

- *pojęcia podstawowe;*
- *diagram korelacyjny i tablica korelacyjna;*
- *korelacja liniowa;*
- *regresja dla dwóch zmiennych;*
- *korelacja i regresja dla wielu zmiennych.*

9.1. **Pojęcia podstawowe.**

Przy badaniu populacji generalnej równocześnie ze względu na dwie lub więcej cech mierzalnych posługujemy się pojęciami regresji i korelacji. Oba te pojęcia dotyczą zależności między zmiennymi, przy czym korelacja zajmuje się siłą tej zależności, natomiast regresja jej kształtem.

Korelacja jest definiowana jako współzależność statystyczna wyników pomiarów różnych zjawisk, zależnych od wspólnej przyczyny lub pozostających ze sobą w bezpośrednim związku przyczynowo-skutkowym, na przykład współpowiązanie wzrostu i wagi u dzieci w określonych grupach wieku, współpowiązanie ciśnienia i temperatury gazu zamkniętego w zbiorniku itp. Skrajnym przypadkiem skorelowania jest współzależność liniowa zmiennych losowych.

O korelacji mówimy, że jest prosta lub dodatnia wtedy, gdy ze wzrostem jednej zmiennej rośnie także druga. Natomiast gdy wzrostowi jednej zmiennej towarzyszy malenie drugiej mamy do czynienia z korelacją odwrotną lub ujemną.

Natomiast regresja w statystyce matematycznej oznacza empirycznie wyznaczoną zależność funkcyjną między skorelowanymi zmiennymi losowymi. Po ustaleniu, że między badanymi cechami istnieje niezbyt słaba korelacja, przystępuje się do znalezienia funkcji regresji, która pozwala na przewidywanie wartości jednej cechy przy założeniu, że druga cecha przyjęła określoną wartość.

W praktyce największe znaczenie ma regresja liniowa, odpowiadająca liniowej zależności pomiędzy rozpatrywanymi zmiennymi losowymi. Wprawdzie regresja liniowa rzadko występuje w praktyce w postaci "czystej", stanowi jednak wygodne narzędzie do otrzymywania przybliżonych zależności. Przy bardziej skomplikowanych współzależnościach stosuje się regresję nieliniową, np. kwadratową. Rozróżnia się przy tym dwa modele danych: model I, w którym wartości zmiennej losowej są znane (dobrze określone) i model II, w którym zmienna losowa jest przypadkowa lub obciążona błędem.

Należy jednak pamiętać, że pojęcie korelacji różni się zarówno od związku przyczynowego, jak i od pojęcia zależności stochastycznej między zmiennymi losowymi.

9.2. **Diagram korelacyjny i tablica korelacyjna.**

Dana jest populacja generalna, w której dwie mierzalne cechy X i Y są zmiennymi losowymi. Jeżeli nieznanne są pewne parametry rozkładu dwuwymiarowej zmiennej (X, Y) , to powstaje problem wyznaczenia ich oszacowań na podstawie próby losowej n par liczb (x_i, y_i) . Traktując x_i i y_i jako współrzędne punktu na płaszczyźnie, można próbkę przedstawić graficznie w postaci tzw. diagramu korelacyjnego (rysunek 9.1).

W przypadku prób o liczebności większej od 30 buduje się zwykle tzw. tablicę korelacyjną. Aby sporządzić taką tablicę należy dla każdej z cech zbudować szereg rozdzielczy, obliczając rozstęp:

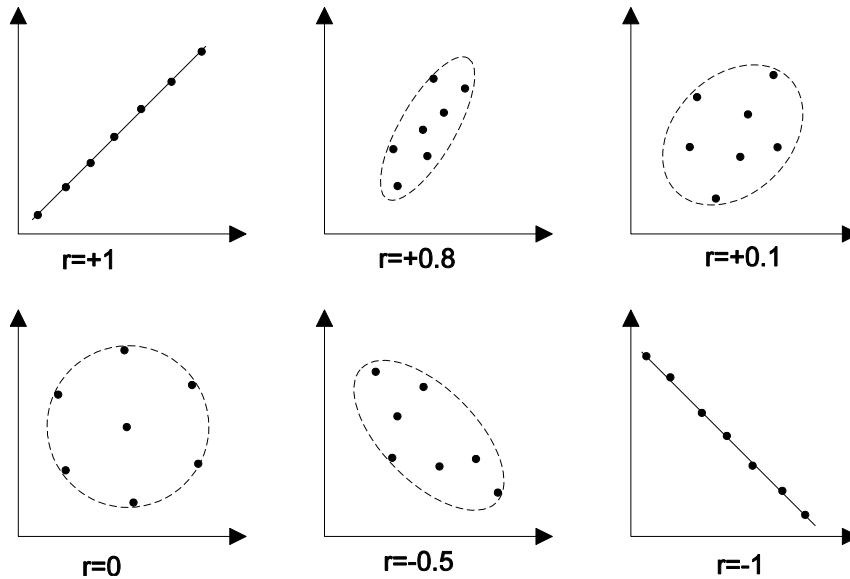
$$\begin{aligned} R_x &= x_{\max} - x_{\min} \\ R_y &= y_{\max} - y_{\min} \end{aligned} \quad (135)$$

a następnie na podstawie liczebności próby n przyjmujemy odpowiednią liczbę klas k i obliczamy długość klasy:

$$d_x = R_x / k \quad (136)$$

$$d_y = R_y / k$$

Jako dolną granicę pierwszej klasy dla zmiennej X przyjmujemy wartość niewiele mniejszą niż x_{min} , podobnie dla zmiennej Y wartość niewiele mniejszą niż y_{min} .



Rys. 9.1. Różne typy diagramów korelacyjnych wraz z odpowiadającymi im współczynnikami korelacji.

Klasyfikację przeprowadzamy w tabelcy ze względu na obie cechy równocześnie i otrzymujemy w ten sposób tabelcę korelacyjną - wierszom przypisujemy wartości klas pierwszej zmiennej, zaś kolumnom wartości klas drugiej zmiennej, natomiast w polach tabelcy umieszczamy liczebności powstałych w ten sposób klas dwuwymiarowych.

Liczebności poszczególnych dwuwymiarowych klas tabelcy oznaczamy przez n_{ij} , przy czym musi zachodzić równość:

$$\sum_{i=1}^k \sum_{j=1}^k n_{ij} = n \quad (137)$$

oraz

$$\sum_{j=1}^k n_{ij} = n_{i\bullet} \quad (138)$$

$$\sum_{i=1}^k n_{ij} = n_{\bullet j}$$

Przez \bar{x}_i i \bar{y}_j oznaczymy odpowiednie środki klas. Liczby $n_{i\bullet}$ są licznosciami klas przy badaniu ze względu na ceche X bez uwzględnienia cechy Y, podobnie $n_{\bullet j}$ są licznosciami klas przy badaniu ze względu na ceche Y bez uwzględnienia cechy X.

9.3. Korelacja liniowa z próbki.

Gdy zależność między dwiema badanymi cechami jest liniowa, to najlepszym miernikiem korelacji między nimi jest współczynnik korelacji ρ zdefiniowany w następujący sposób:

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \quad (139)$$

gdzie $\text{cov}(X,Y)$ oznacza kowariancję X i Y. Dla danych niezgrupowanych kowariancję z próbki oblicza się ze wzoru:

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} \quad (140)$$

Współczynnik korelacji zmienia się w granicach od -1 do +1. Gdy $\rho = -1$ lub $\rho = +1$, wtedy między zmiennymi X i Y istnieje ścisła zależność w postaci liniowej. Gdy $\rho = 0$, wtedy zmienne są zupełnie nieskorelowane. Im $|\rho|$ jest bliższy 1, tym korelacja jest mocniejsza.

Estymatorem współczynnika korelacji ρ między dwoma badanymi cechami X i Y w populacji jest współczynnik korelacji z próby, obliczony na podstawie n par (x_i, y_i) wyników przy pomocy wzoru:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2\right) \left(\frac{1}{n} \sum_{j=1}^n y_j^2 - \bar{y}^2\right)}} \quad (141)$$

Dla danych zgrupowanych w tabelicę korelacyjną współczynnik korelacji z próby obliczamy korzystając z zależności:

$$r = \frac{\frac{1}{n} \sum_{i=1}^n \bar{x}_i \bar{y}_i n_{ij} - \bar{x} \bar{y}}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n \bar{x}_i^2 n_{i\bullet} - \bar{x}^2\right) \left(\frac{1}{n} \sum_{j=1}^n \bar{y}_j^2 n_{\bullet j} - \bar{y}^2\right)}} \quad (142)$$

$$= \frac{\frac{1}{n} \sum_{i=1}^n \bar{x}_i \left(\sum_{j=1}^n \bar{y}_j n_{ij} \right) - \bar{x} \bar{y}}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^n \bar{x}_i^2 n_{ij} - \bar{x}^2\right) \left(\frac{1}{n} \sum_{j=1}^n \bar{y}_j^2 n_{ij} - \bar{y}^2\right)}} \quad (143)$$

Duże wartości bezwzględne współczynnika korelacji świadczy o dużej współzależności liniowej między cechami X i Y, nie może być jednak dowodem związku przyczynowego pomiędzy tymi wielkościami (cechami).

Współczynnik r nazywany współczynnikiem determinacji (coefficient of determination) o $(n-1)$ stopniach swobody może być estymatorem wariancji.

Rozkład estymatora r parametru ρ jest na ogół bardzo skomplikowany. Przy założeniu, że populacja generalna ma dwuwymiarowy rozkład normalny z parametrem $\rho = 0$, rozkład współczynnika korelacji z próby r jest prostszy i sprowadza się do rozkładu t Studenta. Gdy próba jest duża korzysta się oczywiście z granicznego rozkładu normalnego. Pozwala na sprawdzenie hipotezy dla wartości współczynnika korelacji w populacji.

W przypadku gdy mierzalne cechy X i Y mają w populacji generalnej rozkłady normalne, bądź bardzo zbliżone do normalnego, a z populacji wylosowano dużą liczbę elementów (kilkaset), to przybliżony wzór na przedział ufności dla współczynnika korelacji r jest wtedy następujący:

$$P\{r - u_\alpha(1-r^2) / \sqrt{n} < \rho < r + u_\alpha(1-r^2) / \sqrt{n}\} = 1 - \alpha$$

gdzie u_α jest wartością standaryzowanej zmiennej normalnej, którą odczytujemy z tablicy rozkładu $N(0,1)$ dla ustalonego z góry poziomu ufności α w taki sposób, aby $P\{-u_\alpha < U < u_\alpha\} = \alpha$. W takim wypadku wygodnie jest korzystać z wyników pogrupowanych w tablicy korelacyjnej.

W przypadku, gdy z populacji pobrana jest losowo niezbyt duża próba, należy sprawdzić hipotezę zerową, że zmienne X i Y są nieskorelowane, $H_0: \rho = 0$ (wobec hipotezy alternatywnej $H_0: \rho \neq 0$). Test istotności dla tej hipotezy przeprowadzamy w oparciu o statystykę:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad (144)$$

Statystyka ta, ma (przy założeniu o prawdziwości hipotezy H_0 !) rozkład t Studenta z $n-2$ stopniami swobody. Wobec tego korzystając z tablicy rozkładu t dla ustalonego poziomu ufności α odczytujemy wartość krytyczną t_α taką, że:

$$P\{|t| \geq t_{\alpha}\} = 1 - \alpha.$$

Postawioną przez nas hipotezę zerową należy odrzucić, gdy $|t| \geq t_{\alpha}$.

Możemy również testować hipotezę, że współczynnik korelacji ma w populacji określoną wartość $H_0: \rho = \rho_0$. Po obliczeniu z próby wartości współczynnika korelacji r obliczamy następnie wartość statystyki U :

$$u = \left(1.1513 \log \frac{1+r}{1-r} - 1.1513 \log \frac{1+\rho_0}{1-\rho_0} - \frac{\rho_0}{2(n-1)} \right) \sqrt{n-3} \quad (145)$$

W przypadku prawdziwości hipotezy H_0 statystyka ta ma rozkład normalny $N(0,1)$. Z tablicy rozkładu normalnego odczytujemy krytyczną wartość u spełniającą warunek $P\{|U| \geq u_{\alpha}\} = 1 - \alpha$. Jeżeli obliczona wartość statystyki spełnia nierówność $|u| \geq u_{\alpha}$ hipotezę zerową należy odrzucić, w przeciwnym wypadku nie ma podstaw do odrzucenia hipotezy.

Przykład 30.

Dokonano $n = 50$ pomiarów wymiarów odlewów wyniki zestawiając w tabeli 33. Na poziomie ufności $\alpha = 95\%$ zweryfikować hipotezę, że istnieje korelacja pomiędzy wymiarami odlewów.

Tabela 33.

Wyniki pomiarów wymiarów odlewów.

i	x_i	y_i	i	x_i	y_i
1	38.5	5.5	26	34.2	3.6
2	41.1	4.8	27	39.1	5.1
3	37.8	5.0	28	37.5	4.9
4	36.0	4.9	29	35.5	5.0
5	32.2	5.1	30	36.6	4.1
6	36.8	4.3	31	40.5	5.5
7	33.5	4.5	32	37.2	5.0
8	35.3	3.8	33	34.5	4.8
9	31.1	3.4	34	38.5	4.5
10	42.5	5.7	35	34.0	4.1
11	39.5	5.4	36	33.5	4.0
12	42.1	5.2	37	32.5	4.5
13	38.0	5.2	38	36.4	4.5
14	36.5	5.1	39	37.5	5.6
15	40.0	4.5	40	41.4	5.3
16	36.5	4.4	41	39.5	6.0
17	34.0	4.4	42	38.1	3.9
18	34.5	3.9	43	35.7	4.6
19	44.5	6.6	44	39.5	6.0
20	38.0	5.9	45	35.5	4.6
21	40.0	5.7	46	40.5	6.1
22	36.5	5.4	47	37.5	4.3
23	38.8	5.1	48	33.5	5.2
24	34.5	4.6	49	42.5	6.6
25	36.1	4.2	50	38.0	4.4

Diagram korelacyjny dla tych danych przedstawia rysunek 9.2 Porównując go z rysunkiem 9.1 można przewidzieć, że współczynnik korelacji liniowej z próby będzie się zawierał w granicach od 0.5 do 0.8.

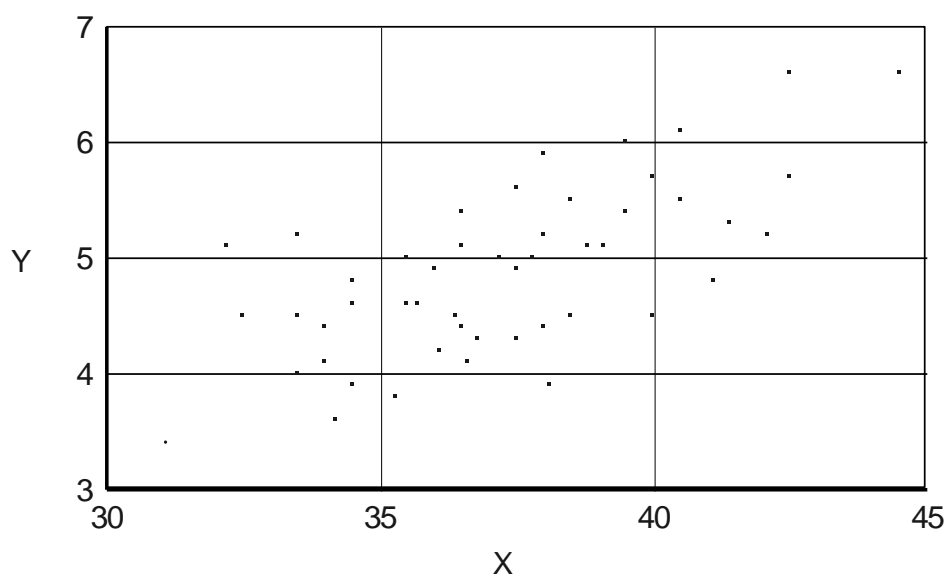


Diagram korelacyjny

Rys. 9.2. Diagram korelacyjny dla danych z tabeli 33.

Obliczamy rozstępy:

$$R_x = 44.5 - 31.1 = 13.4 \text{ i } R_y = 6.6 - 3.4 = 3.2$$

Ponieważ liczba pomiarów $n = 50$ przyjmujemy liczbę klas k równą 7. Zatem długości klas są równe: dla cechy X (wymiaru) $d_x = R_x / k = 13.4 / 7 \cong 2$ i dla cechy Y $d_y = 3.2 / 7 \cong 0.5$.

Jako dolną granicę dla cechy X przyjmujemy $x = 31.0$ a dla cechy Y wartość $y = 3.25$. Otrzymujemy zatem tablicę korelacyjną którą przedstawia tabela 34.

Tabela 34.

Tablica korelacyjna dla danych z tabeli 33.

i		1	2	3	4	5	6	7
		X						
Y		31-33	33-35	35-37	37-39	39-41	41-43	43-45
1	3.25-3.75	1	1	-	-	-	-	-
2	3.75-4.25	1	3	3	1	-	-	-
3	4.25-4.75	1	3	5	3	1	-	-
4	4.75-5.25	1	2	3	5	2	-	-
5	5.25-5.75	-	-	1	2	3	2	-
6	5.75-6.25	-	-	-	1	2	1	-
7	6.25-6.75	-	-	-	-	-	-	-
n_{\cdot}		4	9	12	12	8	4	1

Średnie wartości zmiennych $\bar{x} = 37.273$ i $\bar{y} = 5.19$ oraz średnie odchylenia kwadratowe $s_x^2 = 8.5136$ i $s_y^2 = 0.4544$, zatem odpowiednie odchylenia standardowe $s_x = 2.9178$ zaś $s_y = 0.6741$.

Podstawiając do wzoru na współczynnik korelacji liniowej z próby otrzymujemy:

$$r = \frac{\frac{1}{50} \sum_{j=1}^k \bar{y}_j \left(\sum_{i=1}^k \bar{x}_i n_{ij} \right) - \bar{x} \bar{y}}{s_Y s_X} = \frac{0.02 \cdot 9041 - 37.273 \cdot 5.19}{2.9178 \cdot 0.6741} = 0.6878$$

Dwie ostatnie kolumny i dwa ostatnie wiersze tablicy z tabeli 35 zawierają pomocnicze rachunki pokazujące jak w dwojaki sposób można wykonać obliczenia wartości wyrażenia $\sum \sum \bar{x}_i \bar{y}_i n_{ij}$ (oczywiście do rozwiązania wystarczy obliczenia wykonać jednym z tych sposobów).

9.4. Regresja dla dwóch zmiennych - proste regresji.

Dana jest populacja generalna, w której cechy (X, Y) mają pewien dwuwymiarowy rozkład. Prosta regresji drugiego rodzaju cechy Y względem cechy X ma równanie:

$$y = ax + b \quad (146)$$

gdzie:

$$a = p \frac{\sigma_X}{\sigma_Y} \quad (147)$$

nazywany jest współczynnikiem regresji liniowej cechy Y względem X , zaś

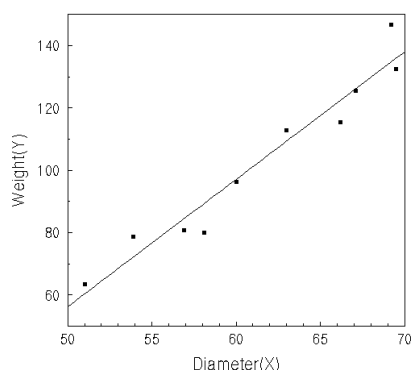
$$b = EY - p \frac{\sigma_X}{\sigma_Y} EX \quad (148)$$

jest współczynnikiem przesunięcia lub wyrazem wolnym.

Jeżeli rozkład cech jest nieznan, parametry a i b szacuje się na podstawie próby metodą najmniejszych kwadratów (rysunek 9.3). Prosta $y = Ax + B$ jest oszacowaniem metodą najmniejszych kwadratów prostej regresji cechy Y względem cechy X na podstawie próby, gdy $A = r S_Y / S_X$ i $B = \bar{Y} - A \bar{X}$. Oznacza to, że funkcja:

$$f(\alpha, \beta) = \sum_{i=1}^n [Y_i - (\alpha X_i + \beta)]^2 = \sum_{i=1}^n d_i^2 \quad (149)$$

osiąga minimum, gdy $\alpha = A$ i $\beta = B$.



Rys. 9.3. Oszacowanie parametrów a i b przy pomocy metody najmniejszych kwadratów.

Tabela 35.

Tablica zawierająca wyniki obliczeń.

	i	1	2	3	4	5	6	7						
k	\bar{y}_k	\bar{x}_i							n _k	$\bar{y}_k \cdot n_k$	\bar{y}_k^2	$\bar{y}_k^2 \cdot n_k$	$\sum_i \bar{x}_i \cdot n_{ik}$	$\bar{y}_k \sum_i \bar{x}_i \cdot n_{ik}$
		32	34	36	38	40	42	44						
1	3.5	1	1	-	-	-	-	-	2	7,0	12,25	24,50	66	231
2	4.0	1	3	3	1	-	-	-	8	32,0	16,0	128,00	280	1 120
3	4.5	1	3	5	3	1	-	-	13	58,5	20,25	263,25	468	2 106
4	5.0	1	2	3	5	2	1	-	14	70,0	25,00	350,00	520	2 600
5	5.5	-	-	1	2	3	2	-	8	44,0	30,25	242,00	316	1 738
6	6.0	-	-	-	1	2	1	-	4	24,0	36,00	144,00	160	960
7	6.5	-	-	-	-	-	4	1	1	6,5	42,25	42,25	44	286
$n_{i.}$		4	9	12	12	8	4	1	50	242,0		1 194,00		9 041
$\bar{x}_i \cdot n_i$		128	306	432	456	320	168	44	1854					
\bar{x}_i^2		1024	1156	1296	1296	1600	1764	1936						
$\bar{x}_i^2 \cdot n_i$		4096	10404	15552	17328	12800	7056	1936	69172					
$\sum_k \bar{y}_k \cdot n_{ik}$		17	39	55	595	43	22	6.5						
$\bar{x}_i \sum_k \bar{y}_k \cdot n_{ik}$		544	1326	1980	2261	1720	924	286	9041					

W przypadku prostej regresji cechy X względem cechy Y współczynniki oblicza się korzystając ze wzorów:

$$A = \frac{1}{R} \frac{S_Y}{S_X} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \quad (150)$$

$$B = \bar{Y} - A\bar{X} \quad (151)$$

W tym wypadku funkcja

$$f(\alpha, \beta) = \sum_{i=1}^n [X_i - (\alpha Y_i + \beta)]^2 \quad (152)$$

ma minimum dla $\alpha = A$ i $\beta = B$.

Jeżeli prostą regresji wyznacza się według danych z tablicy korelacyjnej, to średnie występujące we wzorach należy zastąpić średnimi ważonymi.

Korzystając ze współczynnika korelacji możemy określić wartość doświadczalną statystyki Snedecora dla 1 i n-2 stopni swobody:

$$F = \frac{r^2(n-2)}{1-r^2} \quad (153)$$

a to pozwala nam określić poziom ufności z jakim wyznaczyliśmy prostą regresji.

Możemy także określić przedział ufności dla współczynnika A jako $(-ts_A, +ts_A)$, gdzie t jest statystyką Studenta o n-2 stopniach swobody i założonym poziomie ufności α , zaś

$$s_A^2 = \frac{(1-r^2) \sum_i y_i^2}{\sum_i x_i^2} \quad (154)$$

Oprócz omówionych dotychczas prostych regresji stosuje się także tzw. prostą regresji ortogonalnej (wyznaczaną również przy pomocy metody najmniejszych kwadratów). Dla tej prostej funkcja

$$f(\alpha, \beta) = \sum_{i=1}^n \frac{(Ax_i - y_i + B)^2}{A^2 + 1} = \sum_{i=1}^n d_i^2 \quad (155)$$

jest minimalna gdy $a = A$ i $\beta = B$. Równanie prostej regresji ortogonalnej wyznaczonej na podstawie próby jest postaci:

$$y = \frac{s_Y^2 - s_X^2 + \sqrt{(s_Y^2 - s_X^2)^2 + 4 \text{cov}^2(x, y)}}{2 \text{cov}(x, y)} (x - \bar{x}) + \bar{y} \quad (156)$$

Przykład 31.

Na podstawie tablicy korelacyjnej z poprzedniego przykładu obliczyć równania prostych regresji oraz kąt pomiędzy nimi.

Aby wyznaczyć współczynniki prostych regresji należy najpierw policzyć wartości przeciętne jako klasyczne wartości średnie arytmetyczne, otrzymując dla cechy X wartość $\bar{x} = 37.273$ i dla cechy Y wartość $\bar{y} = 5.19$ oraz estymować wariancje $s_X^2 = 8.5136$ i $s_Y^2 = 0.4544$, co wykonano poprzednio. Następnie należy, korzystając ze wzoru (140), obliczyć kowariancję $\text{cov}(x, y)$:

$$\text{cov}(x, y) = \frac{1}{50} \cdot 9041 - 37.273 \cdot 5.19 = -12.6269.$$

Podstawiając do wzorów (147) i (148) otrzymujemy:

$$a = r S_Y / S_X = \text{cov}(x, y) / S_X S_Y = 0.1589$$

$$a' = 1/r S_x/S_y = 0.3359$$

$$b = y - a x = 5.19 - 0.1589 \cdot 37.273 = -0.7327$$

$$b' = y - a' x = 5.19 - 0.3359 \cdot 37.273 = -7.3300.$$

Zatem prosta regresji Y względem X dana jest równaniem:

$$y = 0.1589 x - 0.7327$$

natomiast prosta regresji X względem Y ma równanie:

$$y = 0.3359 x - 7.3300$$

9.5. Korelacja i regresja dla wielu zmiennych.

W przypadku korelacji więcej niż dwóch zmiennych należy zdefiniować dodatkowo następujące pojęcia:

Korelacja prosta (całkowita) jest to korelacja pomiędzy dwoma zmiennymi (bez uwzględnienia pozostałych zmiennych).

Korelacja cząstkowa jest to korelacja dla dwóch zmiennych w przypadku, gdy pozostałe zmienne utrzymywane są na stałym poziomie.

Korelacja wielokrotna jest to połączona korelacja pomiędzy wieloma zmiennymi, które zmieniają się równocześnie.

Dla korelacji prostej obliczamy współczynnik korelacji liniowej:

$$r_{yx_l}^2 = \frac{\left(\sum_{i=1}^n x_{li} y_i \right)^2}{\sum_{i=1}^n x_{li}^2 \sum_{i=1}^n y_i^2} \quad (157)$$

gdzie l - numer zmiennej niezależnej, względem której obliczamy współczynnik korelacji.

Dla korelacji cząstkowej musimy obliczyć korelację prostą pomiędzy X_1 i X_2 :

$$r_{x_1 x_2}^2 = \frac{\left(\sum x_1 x_2 \right)^2}{\sum x_1^2 \sum x_2^2} \quad (158)$$

i na tej podstawie dla ustalonego X_2 możemy obliczyć:

$$r_{y x_1 | x_2}^2 = \frac{\left(r_{y x_1} - r_{y x_2} r_{x_1 x_2} \right)^2}{\left(1 - r_{y x_2}^2 \right) \left(1 - r_{x_1 x_2}^2 \right)} \quad (159)$$

Dla korelacji wielokrotnej badamy związek Y z połączonymi X_1 i X_2 przy pomocy współczynnika korelacji:

$$R_{Y|x_1 x_2}^2 = \frac{r_{y x_1}^2 + r_{y x_2}^2 - 2 r_{y x_1} r_{y x_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} \quad (160)$$

przy czym $R \in \langle 0, 1 \rangle$.

W przypadku większej ilości zmiennych niezależnych:

$$r_{y x_1 | x_2, x_3, \dots, x_m}^2 = \frac{\left(r_{y x_1 | x_3, x_4, \dots, x_m} - r_{y x_2 | x_3, x_4, \dots, x_m} r_{x_1 x_2 | x_3, x_4, \dots, x_m} \right)^2}{\left(1 - r_{y x_2 | x_3, x_4, \dots, x_m}^2 \right) \left(1 - r_{x_1 x_2 | x_3, x_4, \dots, x_m}^2 \right)} \quad (161)$$

czyli, dla obliczenia współczynnika korelacji cząstkowej dowolnego rzędu, konieczna jest znajomość trzech odpowiednich współczynników cząstkowych rzędu o 1 niższego.

Współczynnik korelacji wielokrotnej dany jest zależnością:

$$1 - R_{y|x_1 \dots x_m}^2 = (1 - r_{yx_1}^2)(1 - r_{yx_2|x_1}^2)(1 - r_{yx_3|x_1 x_2}^2) \dots (1 - r_{yx_m|x_1 \dots x_{m-1}}^2) \quad (162)$$

9.6. Krzywe regresji.

Krzywe regresji mają równanie ogólne postaci:

$$y = a + b_1 x_1 + b_2 x_2 \quad (163)$$

gdzie b_i jest współczynnikiem cząstkowej regresji rzędu i -tego.

W tym przypadku, aby znaleźć wartości a , b_1 i b_2 należy rozwiązać następujący układ równań nieliniowych:

$$\begin{aligned} a n + b_1 \sum x_1 + b_2 \sum x_2 &= \sum y \\ a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 &= \sum x_1 y \\ a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 &= \sum x_2 y \end{aligned} \quad (164)$$

lub rozwiązać układ:

$$\begin{aligned} b_1 \sum x_1 + b_2 \sum x_1 x_2 &= \sum x_1 y \\ b_1 \sum x_1 x_2 + b_2 \sum x_2^2 &= \sum x_2 y \end{aligned} \quad (165)$$

a współczynnik $a = y - b_1 x_1 - b_2 x_2$.

10. METODY ESTYMACJI PARAMETRYCZNEJ

- matematyczny model zjawiska;
- podstawowa zasada metody najmniejszych kwadratów;
- metoda największej wiarygodności;
- analiza dyspersyjna.

W przypadku występowania pomiędzy poszczególnymi wielkościami ukrytych związków statystycznych przed badaczem staje zadanie znalezienia ogólnej tendencji właściwej danemu procesowi lub zjawisku, przy czym najczęściej zachodzi potrzeba przedstawienia tej zależności w postaci formuły matematycznej. Przy rozwiązywaniu zadań tego typu w statystyce (i nie tylko) stosuje się bardzo wygodną metodę najmniejszych kwadratów. Metodę tę można stosować niezależnie od tego, czy znane są błędy pomiarów lub rozkład, któremu podlegają pomiary. Jednak, gdy chcemy wnioskować o błędach dopasowywanych parametrów i jakości dopasowania krzywej do wartości eksperymentalnych, niezbędna jest znajomość zarówno błędów jak i rozkładów.

Natomiast najbardziej popularną metodą estymacji nieznanych parametrów rozkładu populacji jest metoda największej wiarygodności. Metoda ta pozwala na znalezienie estymatorów nieznanych parametrów w takich rozkładach populacji, w których znana jest ich postać funkcyjna. Estymatory uzyskane metodą największej wiarygodności mają wiele pożądanych własności. Trzy najważniejsze ze względów praktycznych to:

1. Dla dużej liczby pomiarów estymator podlega rozkładowi normalnemu;
2. Wariancja estymatora, czyli ocena dokładności wyznaczenia wartości prawdziwej, jest najlepsza jaką można osiągnąć w danej sytuacji (optymalna);
3. Estymator uzyskany tą metodą nie zależy od tego, czy maksimum wiarygodności wyznaczymy dla estymowanego parametru, czy też dla dowolnej jego funkcji.

10.1. Matematyczny model zjawiska.

W przypadku opracowywania danych statystycznych metodą najmniejszych kwadratów stosuje się zazwyczaj przy modelowaniu pewną formę standardową, najczęściej przedstawiając poszukiwaną zależność w formie wielomianu określonego stopnia. Problem polegający na tym ile wyrazów tego szeregu należy uwzględnić w określonym przypadku rozwiązuje się w oparciu o ogólne wyobrażenia o charakterze badanego zjawiska lub w oparciu o wyniki innych metod analizy danych. W najprostszym przypadku, gdy przyrost y jest proporcjonalny do przyrostu x można skorzystać z zależności liniowej:

$$y = a_0 + a_1 x \quad (166)$$

Jeśli są podstawy aby oczekiwać, że wartość y będzie progresywnie zmieniać się ze wzrostem x oraz jeżeli spodziewamy się, że w granicach zmiany x wystąpi ekstremum wartości y należy dołączyć wyraz $a_2 x^2$.

Gdyby równanie drugiego stopnia okazało się niewystarczające, możemy zastosować wielomian trzeciego stopnia. W niektórych przypadkach może zająć konieczność zastosowania paraboli wyższego rzędu, choć w praktyce przypadki tego rodzaju występują bardzo rzadko.

Natomiast czasem zamiast paraboli zachodzi konieczność zastosowania hiperboli (symetrycznej względem osi OX). Równanie przyjmuje wówczas postać:

$$y = a_0 + a_1 \frac{1}{x} \quad (167)$$

lub wyrażenia wyższego rzędu:

$$y = a_0 + a_1 \frac{1}{x} + a_2 \frac{1}{x^2} \quad (168)$$

Zmienną w takim wielomianie asymptotycznym może być nie tylko $1/x$, lecz także każda inna funkcja x o określonych parametrach np. $z = \log x$ lub $z = \sqrt{x}$.

W przypadku badania zjawisk w których zachodzą statystyczne związki dwóch, trzech lub większej ilości cech stosujemy wielomiany odpowiedniego stopnia względem tych zmiennych np. wielomian:

$$y = k + a x + b z \quad (169)$$

W tych przypadkach, gdy matematyczny model zjawiska powinien uwzględniać wpływ czynników powiązanych ze sobą, do równania należy wprowadzić parzyste iloczyny odpowiednich zmiennych pomnożone przez odpowiednie współczynniki, np.:

$$y = k + ax + bxy + cxz \quad (170)$$

Jeżeli w związku z obecnością kilku czynników wiadomo, że ich oddziaływanie nie ma charakteru prostoliniowego, to do równania wprowadza się wyrazy odpowiednich zmiennych w kwadracie lub w wyższych stopniach, np.:

$$y = k + a_1 x + a_2 x^2 + b_1 z + b_2 z^2 + cxz. \quad (171)$$

Jeśli można się spodziewać wystąpienia wzajemnych związków wyższego rzędu, to oprócz iloczynu xz do równania można włączyć wyrazy x^2z , xz^2 oraz x^2z^2 . Należy to jednak czynić tylko w szczególnych przypadkach, gdyż w znacznym stopniu komplikują one matematyczny model zjawiska, a w małym stopniu uściślają równanie.

Oczywiście także w przypadku wielu zmiennych można do równania włączyć funkcje tych zmiennych.

10.2. Podstawowa zasada metody najmniejszych kwadratów.

Rozważmy przypadek równania drugiego stopnia:

$$y = a_0 + a_1 x + a_2 x^2 \quad (172)$$

Zadanie sprowadza się tu do znalezienia wartości liczbowych a_0 , a_1 i a_2 . Można je rozwiązać posiadając szereg obserwacji par zmiennej zależnej y i zmiennej niezależnej x :

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n).$$

W przypadku obecności związku statystycznego pomiędzy y i x nie ma możliwości poprowadzenia krzywej przez wszystkie punkty pomiarowe, niektóre obliczone na podstawie wzoru wartości będą odbiegać od wartości empirycznych. Naszym celem jest zminimalizowanie tych odchyłeń, w tym celu należy ustalić matematyczne zasady pomiaru stopnia niezgodności rzeczywistych wartości z wyliczonymi. U podstawy metody najmniejszych kwadratów leży zasada zgodnie z którą stopień niezgodności jest mierzony sumą kwadratów odchyłeń wartości rzeczywistej y i obliczonej Y :

$$\sum (y - Y)^2 = \text{minimum.}$$

Rozwiązanie takiego zadania jest stosunkowo proste, mamy bowiem funkcję:

$$f(a_0, a_1, a_2) = \sum (y - a_0 - a_1 x - a_2 x^2)^2$$

dla której musimy znaleźć wartości a_0 , a_1 i a_2 dla których funkcja ta osiąga minimum. W tym celu wystarczy przyrównać poszczególne pochodne tych zmiennych do zera i rozwiązać otrzymany w ten sposób układ równań:

$$\partial f / \partial a_0 = -2 \sum (y - a_0 - a_1 x - a_2 x^2) = 0$$

$$\partial f / \partial a_1 = -2 \sum (y - a_0 - a_1 x - a_2 x^2) x = 0$$

$$\partial f / \partial a_2 = -2 \sum (y - a_0 - a_1 x - a_2 x^2) x^2 = 0.$$

Po wykonaniu prostych przekształceń otrzymujemy układ trzech równań, zwany układem równań normalnych:

$$\begin{aligned} a_0 n + a_1 \sum x + a_2 \sum x^2 &= \sum y \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum x^3 &= \sum xy \\ a_0 \sum x^2 + a_1 \sum x^3 + a_2 \sum x^4 &= \sum x^2 y \end{aligned} \quad (173)$$

Można ustalić regułę na podstawie której od razu, abstrahując od wszystkich poprzednich wyliczeń, potrafimy zapisać potrzebny układ równań normalnych, przy dowolnej ilości wyrazów w wyrażeniu wyjściowym. Dla równania:

$$Y = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$$

pierwsze z równań układu otrzymujemy poprzez pomnożenie równania wyjściowego przez 1 i posumowaniu (z zamianą Y na y) po wszystkich pomiarach.

$$\sum a_0 + \sum a_1 x + \sum a_2 x^2 = a_0 n + a_1 \sum x + a_2 \sum x^2 = \sum y$$

Dla drugiego równania przed posumowaniem mnożymy równanie wyjściowe przez x, dla kolejnych równań mnożymy przez x podniesione do potęgi o jeden mniejszej od numeru równania.

Układ równań normalnych najczęściej rozwiązuje się metodą wyznaczników. Obliczenia rozpoczynamy od wyznacznika głównego takiego układu równań:

$$W = \det \begin{vmatrix} n, & \sum x, & \sum x^2 \\ \sum x, & \sum x^2, & \sum x^3 \\ \sum x^2, & \sum x^3, & \sum x^4 \end{vmatrix} =$$

$$= n \sum x^2 \sum x^4 + 2 \sum x \sum x^2 \sum x^3 - (\sum x^2)^3 - n (\sum x^3)^2 - (\sum x)^2 \sum x^4$$

Jeżeli ilość par obserwacji nie jest mniejsza od ilości niewiadomych oraz jeśli obserwacje te są niezależne, to wyznacznik ten jest różny od zera, a układ równań posiada rozwiązanie.

Wyznacznik dla a_0 jest postaci:

$$W_0 = \det \begin{vmatrix} \sum y, & \sum x, & \sum x^2 \\ \sum xy, & \sum x^2, & \sum x^3 \\ \sum x^2 y, & \sum x^3, & \sum x^4 \end{vmatrix} =$$

$$= \sum y \sum x^2 \sum x^4 + 2 \sum xy \sum x^2 \sum x^3 - (\sum x^2)^2 \sum x^2 y - \sum x (\sum x^3)^2 - \sum x \sum xy \sum x^4.$$

Zatem

$$a_0 = \frac{W_0}{W} \tag{174}$$

$$= \frac{\sum y \sum x^2 \sum x^4 + 2 \sum xy \sum x^2 \sum x^3 - (\sum x^2)^2 \sum x^2 y - \sum x (\sum x^3)^2 - \sum x \sum xy \sum x^4}{n \sum x^2 \sum x^4 + 2 \sum x \sum x^2 \sum x^3 - (\sum x^2)^3 - n (\sum x^3)^2 - (\sum x)^2 \sum x^4}$$

W podobny sposób obliczamy pozostałe współczynniki wielomianu.

Suma kwadratów odchyłeń $SSD = S(y-Y)^2$ nazywana jest resztkową sumą kwadratów. Choć metoda najmniejszych kwadratów daje gwarancję, że układ odchyłeń y od Y jest w określonym sensie najlepszy, interesująca jest liczbowa wartość sumy kwadratów jako pewna miara rozrzutu. W przypadku niewielkiej wartości tej sumy tendencja scharakteryzowana równaniem dość ściśle odzwierciedla zmiany rzeczywistych y.

Resztkową sumę kwadratów można łatwo policzyć, korzystając przy tym z sum policzonych już uprzednio dla układu równań normalnych, obliczając tylko sumę y^2 , gdyż:

$$\sum (y-Y)^2 = \sum y^2 - a_0 \sum y - a_1 \sum xy - a_2 \sum x^2 y.$$

Resztkowa suma kwadratów jako taka nie może być traktowana jako wyczerpująca charakterystyka bliskości rzeczywistych wartości y z teoretycznymi, bowiem zależy ona silnie od ilości obserwacji. Nasuwa się tutaj użycie jako miary odchylenia resztkowej sumy kwadratów podzielonej przez liczbę pomiarów, jednak z punktu widzenia statystyki korzystniejsze jest zastosowanie w tym miejscu liczby stopni swobody, otrzymując w ten sposób średni kwadrat.

Na podstawie resztkowej sumy kwadratów możemy obliczyć błędy współczynników korzystając dla a_0 z zależności:

$$S_{a_0} = \frac{SSD / (n-2)}{\sqrt{n - \left(\sum_i x\right)^2 / \sum_i x^2}} \quad (175)$$

natomiast dla a z zależności:

$$S_{a_1} = \frac{SSD / n-2}{\sqrt{\sum_i x^2 - \left(\sum_i x\right)^2}} \quad (176)$$

W podobny sposób obliczamy błędy dla pozostałych współczynników wielomianu.

Jeśli nie wszystkie z zaobserwowanych wartości Y_i odznaczają się taką samą dokładnością, to średnie powinny być średnimi ważonymi, przy czym waga ω_i każdego z punktów jest odwrotnie proporcjonalna do kwadratu odchylenia standardowego. Równanie sumy kwadratów ma wówczas postać:

$$SSD = \sum \omega_i (y-Y)^2 \quad (177)$$

Często spotyka się przypadki, gdy względna dokładność ma wartość stałą, czyli $\sigma_i = c Y_i$ lub $\omega_i = (\sigma_i)^{-1} = (c^2 Y_i^2)^{-1}$.

Przykład 32.

Znaleźć równanie paraboli dla danych doświadczalnych przedstawionych w tabeli 36.

Tabela 36.

Wyniki pomiarów i wyniki wstępnych obliczeń potrzebnych do otrzymania współczynników równania paraboli.

i	x_i	y_i	x_i^2	$x_i y_i$	x_i^3	x_i^4	$x_i^2 y_i$
1	2.5	6.5	6.25	16.25	15.625	39.0625	40.625
2	3.0	9.4	9.00	28.20	27.000	81.000	84.600
3	3.5	12.7	12.25	44.45	42.875	150.0625	155.575
4	4.0	17.0	16.00	68.00	64.000	256.0000	272.000
5	4.5	20.8	20.25	93.60	91.125	410.0625	421.200
6	5.0	26.2	25.00	131.00	125.000	625.0000	655.000
7	5.5	30.9	30.25	169.96	166.375	915.0625	934.725
Σ	28.0	123.5	119.0	551.45	532.000	2476.25	2563.725

Na podstawie wzoru (174) i korzystając z odpowiednich policzonych sum zamieszczonych w tabeli 36 otrzymujemy:

$$a_0 = \frac{123.5 \cdot 119.0 \cdot 2476.25 + 2 \cdot 551.45 \cdot 119.0 \cdot 532.0 - (119.0)^2 \cdot 2563.725 - 28 \cdot (532.0)^2 - 28 \cdot 551.45 \cdot 2476.25}{7 \cdot 119.0 \cdot 2476.25 + 2 \cdot 28 \cdot 119.0 \cdot 532.0 - (119.0)^3 - 7 \cdot (532.0)^2 - (28.0)^2 \cdot 2476.25} = 92323.55$$

zaś policzone w podobny sposób

$$a_1 = \frac{7 \cdot 551.45 \cdot 2476.25 + 28 \cdot 2563.725 \cdot 119.0 + 119.0 \cdot 123.5 \cdot 532.0 - (119.0)^2 \cdot 551.45 - 7 \cdot 2563.725 \cdot 532.0 - 28 \cdot 123.5 \cdot 2476.25}{7 \cdot 119.0 \cdot 2476.25 + 2 \cdot 28 \cdot 119.0 \cdot 532.0 - (119.0)^3 - 7 \cdot (532.0)^2 - (28.0)^2 \cdot 2476.25}$$

$$= 1.1595$$

oraz

$$a_2 = \frac{7 \cdot 119.0 \cdot 2563.725 + 28 \cdot 532.0 \cdot 123.5 + 119.0 \cdot 28 \cdot 551.45 - 123.5 \cdot (119.0)^2 - 551.45 \cdot 532.0 \cdot 7 - 2563.725 \cdot (28.0)^2}{7 \cdot 119.0 \cdot 2476.25 + 2 \cdot 28 \cdot 119.0 \cdot 532.0 - (119.0)^3 - 7 \cdot (532.0)^2 - (28.0)^2 \cdot 2476.25}$$

$$= 0.8810$$

czyli poszukiwane równanie paraboli ma kształt:

$$y = 92323.55 + 1.1595 x + 0.8810 x^2.$$

10.3. Metoda największej wiarygodności.

Podstawowym pojęciem występującym w metodzie największej wiarygodności jest pojęcie wiarygodności próby. Wiarygodność (likelihood) n-elementowej próby prostej lub funkcja wiarygodności dana jest zależnością:

$$L(x, \theta) = \begin{cases} f(x_i, \theta) & \text{dla populacji o rozkładzie ciągłym} \\ p(x_i, \theta) & \text{dla populacji o rozkładzie skokowym} \end{cases} \quad (178)$$

gdzie $f(x_i, \theta)$ oznacza funkcję gęstości prawdopodobieństwa a $p(x_i, \theta)$ funkcję prawdopodobieństwa, zaś θ może być pojedynczym parametrem lub wektorem.

Jak wynika z definicji, wiarygodność próby prostej jest łącznym prawdopodobieństwem (gęstością prawdopodobieństwa), jakie dała próba, przy czym wiarygodność ta zależy od prawdziwej wartości szacowanego parametru θ . Dla ustalonego wektora wyników próby, wiarygodność jest jedynie funkcją wartości parametru θ .

Wiarygodność próby jest prawdopodobieństwem otrzymania takich wyników, jakie dała próba.

W praktyce, zadanie wyznaczania najwiarygodniejszego estymatora parametru θ , polegające na maksymalizacji funkcji wiarygodności L , zamieniamy na zadanie maksymalizacji $\ln L$, gdyż funkcja ta osiąga ekstremum w tym samym punkcie co funkcja L , a jest wygodniejsza do różniczkowania, bowiem

$$\ln L = \sum_i^n \ln f(x_i, \theta).$$

Algorytm znajdowania najwiarygodniejszego estymatora parametru θ jest następujący:

1. znajdujemy dla danego rozkładu populacji funkcję wiarygodności L ;
2. logarytmujemy funkcję wiarygodności otrzymując $\ln L$;
3. stosując warunek konieczny ekstremum rozwiązujemy równanie:

$$\ln(L) = 0 \quad (179)$$

otrzymując estymator $\hat{\theta} = g(X)$.

4. sprawdzamy warunek dostateczny na maksimum:

$$\ln(L) < 0 \quad (180)$$

Równanie (180) może być niekiedy trudne do analitycznego rozwiązania i wówczas stosujemy odpowiednie metody numeryczne.

W metodzie największej wiarygodności wprowadza się przedziały wiarygodności dla odpowiednich poziomów wiarygodności. Rozwiązanie równania:

$$\ln L(\theta) = \ln L(\hat{\theta}) - a$$

ze względu na θ dla $a = 0.5, 2, 4.5$ wyznacza przedziały wiarygodności odpowiadające poziomom wiarygodności 68%, 95% i 99.7%

Przykład 33.

Populacja generalna ma rozkład dwupunktowy zero-jedynkowy z nieznanym parametrem p . Znajdźmy najwiarygodniejszy estymator tego parametru z n -elementowej próby prostej.

Ponieważ prawdopodobieństwo w tym rozkładzie dane jest funkcją:

$$P(x_i, p) = p^{x_i} (1-p)^{1-x_i}$$

zatem funkcja wiarygodności ma postać:

$$L = \prod_{i=1}^n P(x_i, p) = p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} = p^m (1-p)^{n-m}$$

gdzie m oznacza liczbę sukcesów w próbie.

Po zlogarytmowaniu mamy:

$$\ln L = m \ln(p) + (n-m) \ln(1-p)$$

a różniczka tego wyrażenia wynosząca:

$$\frac{\partial(\ln L)}{\partial p} = \frac{m}{p} - \frac{n-m}{1-p} = \frac{m-pn}{p(1-p)}$$

jest równa zero wtedy, gdy:

$$\hat{p} = \frac{m}{n}$$

Druga pochodna logarytmu:

$$\frac{\partial^2(\ln L)}{\partial p^2} = \frac{m}{p^2} - \frac{(n-m)}{(1-p)^2}$$

jest mniejsza od zera dla $p = \hat{p}$, co oznacza, że funkcja wiarygodności ma w tym punkcie maksimum, a \hat{p} jest najwiarygodniejszym estymatorem parametru p .

W celu określenia przedziału wiarygodności dla poziomu wiarygodności 95% rozwiązujemy równanie:

$$\ln L(p) = \ln L(\hat{p}) - a$$

czyli

$$m \ln p + (n-m) \ln(1-p) = m \ln(m/n) + (n-m) \ln(1-m/n) - 2$$

11. ZAPISYWANIE I PREZENTACJA WYNIKÓW POMIARÓW

- *zapisywanie wyników eksperymentu;*
- *wykresy i rysunki poglądowe;*
- *programy graficzne i statystyczne.*

Rozdział ten zawiera szereg rad dotyczących prowadzenia zapisu wyników pomiarów i ich prezentacji na forum publicznym. Podane tutaj reguły i przykłady mają na celu uzmysłowienie państwu, że zapis wyników i inne notatki należy prowadzić w sposób dokładny, pełny i jasny, a przy tym z minimum wysiłku z naszej strony. Rady dotyczące graficznej prezentacji wyników mają pomóc czytelnikom w przygotowaniu przejrzystej i zrozumiałej dla innych prezentacji otrzymanych wyników badań.

11.1. Zapisywanie wyników eksperymentu.

W każdym eksperymencie ważne jest zapisywanie na bieżąco wszystkiego, co zostało zrobione, bez żadnej obróbki. Przed zapisaniem odczytanych wyników nie należy dokonywać (szczególnie w pamięci!), żadnych, nawet najprostszych obliczeń.

Zapisując wyniki należy czynić to w sposób najbardziej przejrzysty, tak, aby po upływie dowolnego okresu czasu można było ponownie z nich skorzystać bez większych trudności.

Najlepiej nie trzymać się sztywno jednej metody zapisywania wyników: albo zeszyt laboratoryjny albo luźne kartki, ale dostosowywać metodę do schematu eksperymentu. Niekiedy korzystna jest kombinacja obu tych metod, co obecnie w dobie skoroszytu nie stanowi specjalnego problemu. Warto mieć także dodatkowy zeszyt (notatnik) na notatki luźno związane z przeprowadzaniem eksperymentem (przypadkowe pomysły, dodatkowe pomiary itd.).

W czasie ręcznej rejestracji wyników pomiarów dobrze jest powtórnie odczytywać wskazania, aby sprawdzić poprawność naszego zapisu. Należy pamiętać o każdorazowym zapisywaniu jakich przyrządów się używało (włącznie z numerem fabrycznym lub innym charakterystycznym oznaczeniem) i jakie były wszystkie nastawienia (nawet te, wydawałoby się, mało istotne dla naszego pomiaru). Oczywiście wszystkie notatki powinny być datowane, zaś kartki ponumerowane (szczególnie te luźne, nie spięte czy zszyte).

Nie należy prowadzić notatek "na brudno", przepisywać "na czysto", a szczególnie nie wolno niszczyć oryginalnych zapisów. Postępowanie takie nie tylko zajmuje dużo czasu, ale może wprowadzić dodatkowe błędy, oczywiście jest jeszcze sprawa "poprawienia" otrzymanych wyników w trakcie przepisywania, która stanowi trudną do odparcia pokusę.

Gdy porównujemy ze sobą wyniki kilku pomiarów, korzystnie jest dokonywać tego zestawiając je w tabeli, bowiem dla naszego oka łatwiejsze jest porównywanie cyfr zapisanych w kolumnie. Oprócz tego, taki zapis jest bardziej zwarty i przejrzysty. Dla wygody należy tak dobierać wielokrotność jednostki miary, aby liczby zapisane w kolumnie mieściły się w zakresie od 0,1 do 100. Jednostka zmierzonej wielkości powinna znajdować się w nagłówku, a nie po każdej zmierzonej wartości (w ogóle należy unikać zbędnych powtórzeń).

11.2. Graficzna prezentacja wyników.

Trudno przecenić znaczenie rysunków w notatkach i publikacjach. Rysunek połączony z kilkoma słowami komentara stanowi przeważnie najprostszy i najbardziej efektywny sposób wyjaśnienia przebiegu eksperymentu, opisu układu doświadczalnego oraz wprowadzenia oznaczeń.

Rysunek nie musi być artystyczny, nie musi charakteryzować się fotograficzną dokładnością, powinien jednak być czytelny nie tylko dla autora. W przypadku schematów aparatury dobrze jest zachować

(choć w przybliżeniu) skalę, jeżeli jednak zniekształcenie proporcji może pomóc w jaśniejszym pokazaniu istotnych szczegółów, należy proporcje zmienić.

W fizyce doświadczalnej wykresy służą trzem celom:

- - do graficznego wyznaczenia wartości pewnej wielkości;
- - stanowią pogładową ilustrację;
- - do ustalania empirycznych (przybliżonych) zależności między dwoma wielkościami.

Zasadniczym obecnie celem sporządzania wykresów jest przedstawianie wyników, a zatem powinny one być możliwie jasne i zrozumiałe nie tylko dla autora.

Wskazówki dotyczące sporządzania wykresu::

- Przy doborze skali powinniśmy zwrócić szczególną uwagę na to, aby punkty pomiarowe nie leżały zbyt blisko siebie (najlepiej aby pokrywały cały diagram) oraz aby skala była prosta (szczególnie jest to istotne w przypadku rysunków z naniesioną siatką np. wykonanych na papierze milimetrycznym); czasem o doborze skali decydują względy teoretyczne, i wówczas poprzednie uwagi należy uwzględnić jedynie w miarę możliwości.
- Oś współrzędnych należy opisać za pomocą nazw lub symboli (lub obydwu naraz), oraz dobrać taki mnożnik, aby działości na skali były opisane liczbami 1, 2, 3, ... lub 10, 20, 30.
- Jeżeli na wykresie oprócz punktów pomiarowych nanosi się krzywą teoretyczną lub punkty wynikające z teorii, to należy to zrobić tak, aby punkty eksperymentalne były wyróżnione (poprzez wielkość lub kształt punktu).
- W przypadku, gdy na wykresie nie ma krzywej teoretycznej, dobrze jest poprowadzić przez punkty eksperymentalne "możliwie gładką" krzywą.
- Do rozróżnienia punktów eksperymentalnych pochodzących z różnych serii pomiarowych, wykonanych w różnych warunkach lub dla różnych próbek należy wykorzystywać różne symbole lub różne kolory.
- Gdy wykonujemy wykresy na papierze skalę na osiach i punkty pomiarowe powinniśmy nanosić najpierw ołówkiem, a dopiero po sprawdzeniu poprawić wszystko np. tuszem.
- Ponieważ wprowadzenie oznaczeń błędów stanowi dodatkową pracę i komplikuje wykres, błędy powinniśmy nanosić wtedy, gdy informacje o nich mogą mieć znaczenie przy interpretacji wyników lub, gdy są różne dla różnych punktów pomiarowych.

W chwili obecnej większość wykresów i ilustracji sporządza się korzystając z wyspecjalizowanych programów komputerowych, począwszy od najprostszych, wchodzących na przykład w skład pakietu Microsoft Office (MSGraph), arkusze kalkulacyjne (np. Excel), poprzez programy statystyczne (np. Statistica) aż po wyspecjalizowane programy do robienia wykresów (np. CoPlot z pakietu CoHort). Programy te mają różne możliwości i różne poziomy trudności użytkowania.

12. KOMPUTERY JAKO ŹRÓDŁO BŁĘDÓW W BADANIACH DOŚWIADCZALNYCH

- *system komputerowy jako narzędzie przetwarzania danych doświadczalnych;*
- *błędy kwantowania;*
- *oprogramowanie matematyczne i statystyczne i jego zastosowania w analizie danych doświadczalnych.*

12.1. Systemy komputerowe

Zastosowanie systemu komputerowego do gromadzenia i przetwarzania danych doświadczalnych wiąże się z możliwością wystąpienia dodatkowych błędów:

- * błędów dyskretyzacji (kwantowania) i zaokrągleń, które występują przy przekształcaniu danych analogowych w postać cyfrową i dalszej obróbce danych w postaci liczbowej;
- * błędów przekłamań urządzeń transmisji i przetwarzania danych cyfrowych, które spowodowane są przez zakłócenia zewnętrzne lub przez wadliwość działania przyrządów elektronicznych;
- * błędów spowodowanych wadliwą obsługą operatorską systemu pomiarowego, na przykład zbyt późną lub błędną reakcją na sygnały systemu komputerowego.

Wpływ wymienionych tu błędów można zmniejszyć, dzięki odpowiedniemu zaprojektowaniu systemu komputerowego. Oczywiście błędy operatora mogą być ograniczone poprzez przeszkolenie operatora (przećwiczenie wszystkich możliwych reakcji systemu pomiarowego i odpowiadających im zachowań operatora).

Aby uniknąć błędów przekłamań struktura logiczna systemu komputerowego powinna być rozszerzona o wbudowane w nią mechanizmy automatycznej kontroli (wykrywania) i korekcji błędów. Działanie takich mechanizmów jest oparte na wykorzystaniu ogólnych praw teorii informacji a zwłaszcza zależności pomiędzy możliwością wykrywania i korekcji błędów oraz nadmiarowością statystyczną i strukturalną zapisów informacyjnych.⁴

O nadmiarowości statystycznej możemy mówić wtedy, gdy rozkład prawdopodobieństwa sygnałów docierających do naszego czujnika pozostawia pewną niewykorzystaną rezerwę objętości sygnału (nie wykorzystujemy maksymalnej gęstości informacji).

O nadmiarowości strukturalnej mówimy wtedy, gdy do naszej informacji pochodzącej z czujnika wprowadzimy dodatkowe elementy weryfikujące poprawność transmisji i zapisu (np. bity parzystości, sumy kontrolne itp.)

12.2. Błędy kwantowania

⁴ W.W. Mitiugow, *Fizyczne podstawy teorii informacji*, PWN, Warszawa 1980.

13. OBLICZENIA

W wielu przypadkach celem eksperymentu jest podanie wartości liczbowej pewnej wielkości, poprawne obliczenie tej liczby jest równie ważne jak poprawne przeprowadzenie eksperymentu. Obliczenia możemy wykonywać przy pomocy kalkulatora, komputera oraz siebie samego (czyli „na pieczo”), każde z tych urządzeń ma swoją specyfikę i może być przyczyną różnych błędów. Przeprowadza się doświadczenia podczas których zbiera się a następnie interpretuje ogromne ilości danych, do czego niezbędne są ogromne komputery. W większości przypadków wystarczające obecnie są komputery osobiste, tym bardziej wygodne, że dostępne jest dużo sprawdzonego oprogramowania służącego obliczeniom matematycznym i analizie danych doświadczalnych. Niezastąpiony jest również kalkulator (szczególnie tzw. naukowy - pozwalający przeprowadzić obliczenia podstawowych funkcji oraz obliczenia statystyczne). W tym miejscu należy powrócić do zapisywania wyników. Większość kalkulatorów i programów komputerowych podaje liczby z dokładnością do 8, 10 a nawet 12 cyfr. W znacznej większości eksperymentów nie można uzyskać tylu cyfr znaczących, i należy unikać notowania wartości ze zbyt dużą ilością cyfr, tylko dlatego że widzimy je na wyświetlaczu. Eliminowanie cyfr nieznaczących ułatwia ocenę istotności (poprawności) otrzymanego wyniku oraz zmniejsza prawdopodobieństwo popełnienia błędu. Z drugiej strony jednak, należy pozostawiać jedną (a nawet czasem dwie) cyfrę nieznaczącą, bo być może nie stracimy w ten sposób istotnej informacji, szczególnie jeżeli eksperymentator nie jest pewien jaka dokładność jest osiągnąta.

Podczas wykonywania obliczeń należy pamiętać, aby:

- * unikać niepotrzebnych obliczeń. Gdy wielkości o znanych wartościach mamy podstawić do szeregu zależności, aby w końcu otrzymać interesującą nas wielkość, należy najpierw wykonać te operacje na symbolach, a podstawiać dopiero do wzoru końcowego (najlepiej po skontrolowaniu wymiarów). Powinniśmy tak postępować nawet wówczas, gdy interesują nas wielkości otrzymywane podczas pośrednich obliczeń.
- * postępować dokładnie i systematycznie. Wyniki należy zapisywać pozostawiając wolne miejsce na naniesienie ewentualnych poprawek (polecany jest zapis w tabeli, w taki sposób, że liczby umieszczone w danej kolumnie stanowią wyniki działań zapisanych w nagłówku na liczbach zawartych w poprzednich kolumnach).
- * weryfikować obliczenia po każdym etapie. Znalezienie błędu po zakończeniu obliczeń oznacza wykonanie ich od początku.
- * sprawdzić, czy przynajmniej 2/3 pomiarów leży w przedziale $\pm\sigma$ wokół wartości średniej, oraz czy błąd względny wartości końcowej jest większy od wszystkich błędów względnych wartości zmierzonych i parametrów użytych do obliczeń.

13.1. Wagi statystyczne wyników pomiarów.

Jeżeli pewna wielkość została zmierzona wielokrotnie w kilku oddzielnych seriach pomiarowych (np. w ciągu kilku dni) to wartość średnia otrzymana z wartości średnich poszczególnych serii w prosty sposób jest dobra w przypadku gdy serie składały się zawsze z tej samej liczby pomiarów i obarczone były tym samymi błędami. W przeciwnym wypadku należy wartość średnią obliczać uwzględniając wagi statystyczne serii w_i (w najprostszym przypadku jest to procentowy udział pomiarów danej serii w całkowitej liczbie pomiarów):

$$\bar{x} = \frac{\sum w_i \bar{x}_i}{\sum w_i} \quad (181)$$

W przypadku, gdy mamy poszczególne serie obarczone różnymi błędami: $x_1 \pm \Delta x_1$, $x_2 \pm \Delta x_2$, itd. jako wagi statystyczną należy nadać każdej zmierzonej serii:

$$w_i = \frac{\sigma^2}{(\Delta x_i)^2} \quad (182)$$

gdzie σ jest błędem standardowym, którym obarczony jest rozkład prawdopodobieństwa do którego należą wartości pomiarów.

Błąd standardowy średniej dla wszystkich pomiarów wynosi zatem:

$$\Delta \bar{z} = \sigma / \sqrt{\sum n_i} \quad (183)$$

zaś najlepszą wartość x i jej błąd standardowy można obliczyć z zależności:

$$\bar{x} = \frac{\sum (1/\Delta x_i)^2 x_i}{\sum (1/\Delta x_i)^2} \quad (184)$$
$$\Delta \bar{x} = \frac{1}{\sqrt{\sum (1/\Delta x_i)^2}}$$

Jak widać, nie zależą one od błędu standardowego rozkładu prawdopodobieństwa mierzonej wielkości.

14. LITERATURA

1. E. Bright Wilson jr., Wstęp do badań naukowych, PWN, Warszawa, 1968.
2. G. L. Squires, Praktyczna fizyka, PWN, Warszawa 1992.
3. H. Szydlowski (red.), Teoria pomiarów, PWN, Warszawa 1981.
4. J. Piotrowski, Teoria pomiarów. Pomiary w fizyce i technice, PWN, Warszawa, 1986.
5. G. I. Kawalerow, S. M. Mandelsztam, Wprowadzenie do teorii pomiarów, PWN, Warszawa, 1983.
6. H. Abramowicz, Jak analizować wyniki pomiarów?, PWN, Warszawa 1992.
7. J. Greń, Statystyka matematyczna. Podręcznik programowany, PWN, Warszawa, 1987.
8. J. Greń, Statystyka matematyczna. Modele i zadania, PWN, Warszawa, 1984.
9. J. Brzeziński, R. Stachowski, Zastosowanie analizy wariancji w eksperymentalnych badaniach psychologicznych, PWN, Warszawa, 1984.
10. W. Pieriegudow, Metoda najmniejszych kwadratów i jej zastosowanie, PWE, Warszawa, 1967.
11. J.W. Linnik, Metoda najmniejszych kwadratów i teoria opracowywania obserwacji, PWN, Warszawa, 1962.
12. T. M. Little, F.J. Hills, Agricultural experimentation. Design and analysis, Wiley and Sons, New York, 1987.
13. J. K. Taylor, Statistical techniques for data analysis, Lewis Publ., Inc., New York, 1990.
14. R. F. Barton, Wprowadzenie do symulacji i gier, WNT, Warszawa 1974.
15. A. Plucińska, E. Pluciński, Elementy probabilistyki, PWN, Warszawa 1979.
16. T. Puchalski, Statystyka. Wykład podstawowych zagadnień, PWN, Warszawa 1978.
17. P. Perkowski, Technika symulacji cyfrowej, WNT, Warszawa 1980.
18. W. Krysicki i in., Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, cz. II, Statystyka matematyczna, PWN, Warszawa 1986.
19. J. W. Tukey, Exploratory Data Analysis, Addison-Wesley Publ. Co., Reading, MA, USA, 1977.
20. R. Zieliński, Generatory liczb losowych, WNT, Warszawa 1979.
21. R. Wieczorkowski, R. Zieliński, Komputerowe generatory liczb losowych, WNT, Warszawa 1997.
22. J. L. Kulikowski, Komputery w badaniach doświadczalnych, PWN, Warszawa 1993.
23. J. M. Jaworski, R. Z. Morawski, J. S. Ołędzki, Wstęp do metrologii i techniki eksperymentu, WNT, Warszawa 1992.
24. L. Gajek, M. Kałużka, Wnioskowanie statystyczne. Modele i metody. WNT, Warszawa 1994.
25. W. Wagner, P. Błażczak, Statystyka matematyczna z elementami doświadczalnictwa, Akademia Rolnicza w Poznaniu, 1992.
26. E. Rafajłowicz, Algorytmy planowania eksperymentu z implementacjami w środowisku Mathematica, Akademińska Oficyna Wydawnicza PLJ, Warszawa 1996.
27. G. S. Kembrovkij (red), Fizicheskiy praktikum, Izd. „Universitetskoe”, Minsk 1986.

SPIS TREŚCI

1.	WPROWADZENIE	
5		
1.1.	WNOSKOWANIE DEDUKCYJNE I INDUKCYJNE	5
1.2.	PROJEKTOWANIE BADAŃ NAUKOWYCH	6
1.3.	HIPOTEZY	7
2.	ELEMENTY STATYSTYKI	
9		
2.1.	POJĘCIA PODSTAWOWE	9
2.2.	ROZKŁAD DWUMIANOWY	10
2.3.	ROZKŁAD POISSONA	11
2.4.	ROZKŁAD GAMMA	12
2.5.	ROZKŁAD WEIBULLA	13
2.6.	ROZKŁAD ERLANGA	13
2.7.	ROZKŁAD NORMALNY	14
2.8.	ROZKŁAD CHI-KWADRAT	15
2.9.	ROZKŁAD T STUDENTA	15
2.10.	ROZKŁAD F SNEDECORA	16
2.11.	HIPOTEZY STATYSTYCZNE	16
2.12.	ESTYMACJA	17
3.	GENEROWANIE LICZB LOSOWYCH	
18		
3.1.	LICZBY LOSOWE	18
3.2.	TABLICE LICZB LOSOWYCH	18
3.3.	GENERATORY LICZB LOSOWYCH O RÓWNOMIERNYM ROZKŁADZIE PRAWDOPODOBIENSTWA	19
3.4.	GENERATORY LICZB LOSOWYCH O DOWOLNYCH ROZKŁADACH PRAWDOPODOBIENSTWA	21
3.5.	TESTY NA LOSOWOŚĆ (NIEPRZYPADKOWOŚĆ)	22
3.5.1.	<i>Serie</i>	22
3.5.2.	<i>Trendy i nachylenia</i>	22
3.5.3.	<i>Średni kwadrat kolejnych różnic (MSSD)</i>	23
4.	TESTOWANIE HIPOTEZ STATYSTYCZNYCH	
24		
4.1.	WSTĘP	24
4.2.	TESTY PARAMETRYCZNE	24
96		

4.2.1. Test zgodności średniej próby ze średnią populacji (Test t Studenta).....	24
4.2.2. Test dla wariancji populacji generalnej (Test chi-kwadrat).	26
4.2.3. Test dla dwóch średnich wartości prób.....	26
4.2.4. Test z (zgodności średniej próby ze średnią populacji)......	27
4.2.5. Test Bartletta.	29
4.3. TESTY NIEPARAMETRYCZNE.	29
4.3.1. Test zgodności chi-kwadrat.....	30
4.3.2. Test zgodności λ Kołmogorowa.	31
4.3.3. Test Kołmogorowa-Lillieforsa.	33
4.3.4. Test Shapiro-Wilka	34
4.3.5. Test niezależności chi-kwadrat.....	36
4.3.6. Test Wilcoxon	38
5.....ANALIZA WARIANCJI	
39	
5.1. WPROWADZENIE	39
5.2. EKSPERYMENT JEDNOCZYNNIKOWY.	39
5.2.1. Hipotezy zerowe i alternatywne w jednoczynnikowej analizie wariancji.....	41
5.2.2. Weryfikacja hipotezy o równości wartości przeciętnych w przypadku klasyfikacji jednoczynnikowej. 41	
5.2.3. ANOVA dla dwóch prób.....	42
5.3. WERYFIKACJA HIPOTEZ DOTYCZĄCYCH WARTOŚCI PRZECIĘTNYCH W PRZYPADKU KLASYFIKACJI PODWÓJNEJ. 43	
5.4. SCHEMAT KWADRATU ŁACIŃSKIEGO LQ-R (N=1) III.....	46
6.....TEORIA BŁĘDÓW	
52	
6.1. RODZAJE BŁĘDÓW.....	52
6.2. KLASA NIEDOKŁADNOŚCI.	54
6.3. WNIOSKOWANIE W TEORII BŁĘDÓW.	55
6.4. PRAKTYCZNE OBLICZANIE BŁĘDÓW.	56
7.....ZARZĄDZANIE ZBIORAMI DANYCH	
58	
7.1. WPROWADZENIE	58
7.2. OBSERWACJE NIETYPOWE.	58
7.3. PRAWO GRUBEGO BŁĘDU.	58
7.4. TEST DIXONA.....	59

7.5. TEST GRUBBSA.....	60
7.6. TEST YAUDENA.....	62
7.7. TEST COCHRANA.....	63
7.8. TEST HARTLEYA.....	65
8.	POPRAWIANIE PRECYZJI POMIARÓW
66	
8.1. POJĘCIA PODSTAWOWE.....	66
8.2. ZWIĘKSZENIE LICZBY POMIARÓW.....	67
8.3. DOBÓR ODDZIAŁYWAŃ.....	68
8.4. DOSKONALENIE TECHNIKI POMIAROWEJ.....	68
8.5. WYBÓR MATERIAŁU DOŚWIADCZALNEGO.....	69
8.6. WYBÓR I KONSTRUOWANIE PRZYRZĄDÓW POMIAROWYCH.....	69
8.7. WYBÓR SCHEMATU DOŚWIADCZALNEGO - EKSPERYMENTY WSTĘPNE I GRUPOWE.....	70
8.8. ANALIZA KOWARIANCJI.....	70
8.9. GRANICZNE MOŻLIWOŚCI POMIARÓW.....	71
9.	KORELACJA I REGRESJA
73	
9.1. POJĘCIA PODSTAWOWE.....	73
9.2. DIAGRAM KORELACYJNY I TABLICA KORELACYJNA.....	73
9.3. KORELACJA LINIOWA Z PRÓBKII.....	74
9.4. REGRESJA DLA DWÓCH ZMIENNYCH - PROSTE REGRESJI.....	78
9.5. KORELACJA I REGRESJA DLA WIELU ZMIENNYCH.....	82
9.6. KRZYWE REGRESJI.....	83
10. METODY ESTYMACJI PARAMETRYCZNEJ	84
10.1. MATEMATYCZNY MODEL ZJAWISKA.....	84
10.2. PODSTAWOWA ZASADA METODY NAJMNIEJSZYCH KWADRATÓW.....	85
10.3. METODA NAJWIĘKSZEJ WIARYGODNOŚCI.....	88
11. ZAPISYWANIE I PREZENTACJA WYNIKÓW POMIARÓW	90
11.1. ZAPISYWANIE WYNIKÓW EKSPERYMENTU.....	90
11.2. GRAFICZNA PREZENTACJA WYNIKÓW.....	90
12. KOMPUTERY JAKO ŹRÓDŁO BŁĘDÓW W BADANIACH DOŚWIADCZALNYCH.....	92
12.1. SYSTEMY KOMPUTEROWE.....	92
12.2. BŁĘDY KWANTOWANIA.....	92

13. OBLICZENIA	93
13.1. WAGI STATYSTYCZNE WYNIKÓW POMIARÓW.	93
14. LITERATURA	95