

14. RACHUNEK BŁĘDÓW*

Błędy, które pojawiają się w czasie doświadczenia mogą mieć własne źródła. Są nimi błędy związane z błędną kalibracją torów pomiarowych, szumy, czas reagowania przyrządu, ograniczenia konstrukcyjne, wahania zasilania, zniekształcenia przekazywanych sygnałów, zużycie układu pomiarowego, wpływ otoczenia, nieprawidłowe interpretacje wyników, błąd w odczycie.

Szumem określamy sygnał nie niosący użytecznej informacji. Szumy mogą powstawać w elemencie pomiarowym, wówczas ulegają one wzmocnieniu we wzmacniaczach. Szumy mogą również powstawać w połączeniach pomiędzy czujnikiem a wzmacniaczem lub wzmacniaczem a komputerem. Należą do nich sygnały elektryczne, które powstają pomiędzy przewodami a izolacją w wyniku zjawisk elektrostatycznych i elektromagnetycznych. Podstawową właściwością szumów jest wzmocnienie ich razem z mierzonym sygnałem.

Czas reagowania układu odgrywa istotną rolę w przypadku przebiegów sygnałów, które zmieniają się w czasie np. w badaniach zmęczenia. Wskazania sygnału zależą od bezwładności układu. Reakcja czujnika (np. siły, przemieszczenia uchwytu mocującego próbkę) w dowolnej chwili jest rezultatem procesów poprzednich i aktualnych. Wielkość mierzona zależy nie tylko od rodzaju sygnału, ale i od charakterystyk zmian tego sygnału.

W czasie *przekazywania sygnału* od czujnika do komputera mogą następować jego zmiany np. przez zakłócenia w sieci (spadek napięcia) bądź przez skażenie sygnału na skutek rezonansu. W tych przypadkach sygnał rejestrowany np. przez komputer różni się od sygnału mierzonego przez czujnik.

Zużycie elementu pomiarowego także wpływa na wartości błędów. Przykładem jest stopień końcówek ekstensometru mierzącego odkształcenie w statycznej próbie rozciągania i możliwe ich poślizgi czy też zmiana oporu układu na skutek korozji.

Wpływ ośrodka na układ pomiarowy związany może być z temperaturą, która zawsze wpływa na dokładność przyrządu (może wpłynąć na sztywność sprężyn, lepkość smaru w łożyskach, luzy w mechanizmach). Na dokładność przyrządu może mieć wpływ wilgotność, ciśnienie czy też obecność kurzu w powietrzu.

Błąd w odczycie i w interpretacji danych są błędami subiektywnymi. Zależą od osoby przeprowadzającej doświadczenie. Związane to może być z oczekiwaniem jakiegoś wyniku przez osobę dokonującą pomiarów. Należy też zwrócić uwagę na błąd paralaksy, który powstaje w przypadku, gdy dokonujemy odczytu pod kątem innym niż prosty do tarczy przyrządu np. przy odczycie wartości przemieszczenia przy wykorzystaniu czujnika zegarowego (dotyczy to tylko takich urządzeń, gdzie sygnał nie jest rejestrowany w postaci cyfrowej).

Zwykle rozróżniamy dwie klasy błędów: systematyczne i przypadkowe.

*Opracował: Robert Molasy

Błędy systematyczne są to błędy powtarzające się w czasie kolejnych doświadczeń. Zaliczamy do nich błędy wynikające z niedokładnego wzorcowania czujnika lub jego zużycia. Błędy systematycznego nie możemy oszacować przez powtarzanie doświadczenia w identycznych warunkach. Zwykle dla oszacowania tego rodzaju błędów stosuje się pomiary kalibrujące znanych przebiegów lub pomiary różnymi przyrządami i przy zastosowaniu (jeśli jest to możliwe) różnych metod.

Błędy przypadkowe są to błędy, których wielkości i znaku nie możemy przewidzieć. W czasie pomiarów wyniki zależą od wielu czynników, które nazywamy parametrami pomiarów. W przypadku idealnych pomiarów wartości parametrów są stałe i wtedy wielkość mierzona jest określona jednoznacznie. Przy kolejnym pomiarze parametry mogą ulec zmianie a otrzymany wynik będzie się różnił od poprzedniego. Te nieznane i niekontrolowane parametry oczywiście zmniejszają dokładność pomiarów. Jeśli założymy, że te parametry wpływają zawsze w podobny sposób, to dla wyznaczenia wartości oczekiwanej można posłużyć się teorią prawdopodobieństwa.

14.1. OCENA WYNIKÓW POMIARÓW

Pomiar odbywa się przy użyciu przyrządów o określonej dokładności. Pomiary kilku wielkości niezależnych używane są do wyznaczenia jednej, która nas interesuje. Przykładem może być wyznaczenie wytrzymałości na rozciąganie σ_m lub wydłużenia względnego A w statycznej próbie rozciągania. W pierwszym przypadku grubość próbki okrągłej d_0 mierzymy mikrometrem z dokładnością do setnych milimetra $d_0 = 10 \pm 0.01$ (mm) a wartość siły F_m z dokładnością do 200 N $F_m = 80000 \pm 200$ (N). Nominalny przekrój próbki wynosi $S_0 = 78,5$ (mm²), a nominalna wartość $\sigma_m = 1019$ (MPa). Zakładając, że popełniamy największe możliwe błędy odczytów, otrzymujemy dwie skrajne wartości wytrzymałości na rozciąganie:

$$\sigma_{m(\max)} = \frac{80200(N)}{78,19(mm^2)} = 1026(MPa) \quad \sigma_{m(\min)} = \frac{79800(N)}{78,66(mm^2)} = 1015(MPa)$$

różniące się od wartości nominalnej o $\Delta\sigma_{m1} = 7$ (MPa) oraz $\Delta\sigma_{m2} = 4$ (MPa). Pomiary były wykonane z dokładnością do 0,29% dla przekroju próbki i 0,25% dla wartości siły, a uzyskaliśmy wynik z mniejszą dokładnością, bo odpowiednio 0,69% dla $\Delta\sigma_{m1}$ oraz 0,39% dla $\Delta\sigma_{m2}$.

W drugim przypadku przy wyznaczaniu wydłużenia względnego długość bazy pomiarowej L_0 oraz przyrost wydłużenia (przy wykorzystaniu ekstensometru) w czasie próby ΔL mierzymy z dokładnością do setnych milimetra $L_0 = 50 \pm 0,02$ (mm) i $\Delta L = 10 \pm 0,01$ (mm). Nominalne wydłużenia względne wynosi

$A_0 = 20\%$. Zakładając, że popełniamy największe możliwe błędy odczytów, otrzymujemy dwie skrajne wartości wydłużenia względnego:

$$L_{0(\max)} = \frac{10,01}{49,99} \cdot 100(\%) = 20,03(\%)$$

$$L_{0(\min)} = \frac{9,99}{50,02} \cdot 100(\%) = 19,97(\%)$$

różniące się od wartości nominalnej o $\Delta L_{0l} = 0,03 (\%)$ oraz $\Delta L_{0l} = 0,03 (\%)$. Pomiar były wykonane z dokładnością do 0,04% dla długości bazy i 0,1% dla przyrostu wydłużenia. W tym przypadku także uzyskaliśmy wynik z mniejszą dokładnością, bo ok. 0,15%. Takie szacowanie błędów pomiaru jest zbyt pesymistyczne. Dla bardziej realistycznego oszacowania błędów zostały opracowane specjalne metody.

14.2. SZACOWANIE BŁĘDÓW DLA POJEDYNCZEGO POMIARU

Przy wykorzystaniu matematycznych metod opracowania i analizy wyników doświadczenia [1] możemy przyjąć następujące zasady postępowania:

a) *błąd bezwzględny* wielkości, która jest sumą algebraiczną wielu liczb przybliżonych, jest równy sumie błędów bezwzględnych jej poszczególnych składników. Jeśli $W = W_1 + W_2 + \dots + W_n$ to

$$\Delta W = \Delta W_1 + \Delta W_2 + \dots + \Delta W_n \quad (14.1)$$

Uwzględniając częściową kompensację błędów o różnych znakach otrzymujemy:

$$\Delta W = \sqrt{(\Delta W_1)^2 + (\Delta W_2)^2 + \dots + (\Delta W_n)^2} \quad (14.2)$$

a) *błąd względny* przybliżonej wartości W nazywamy stosunek błędu bezwzględnego ΔW do wartości bezwzględnej liczby W :

$$\delta W = \frac{\Delta W}{|W|} \quad (14.3)$$

Błąd względny różnicy liczb dodatnich jest większy od błędów względnych tych liczb, zwłaszcza gdy liczby te różnią się bardzo mało między sobą.

Przykład 14.1

Jeśli $W = W_1 - W_2$, $W_1 = 50 \pm 0,5$, $W_2 = 40 \pm 0,4$, to błędy względne tych liczb wynoszą $\delta W_1 = \delta W_2 = 1\%$. Natomiast błąd względny różnicy wynosi

$$\delta W = \frac{\Delta W_1 + \Delta W_2}{W_1 - W_2} \cdot 100(\%) = \frac{0,5 + 0,4}{10} \cdot 100(\%) = 10(\%)$$

i jest dziesięć razy większy od błędów liczb W_i .

W przypadku, gdy wynik pomiarów jest funkcją zmiennych niezależnych podlegających mnożeniu lub dzieleniu, to błędy względne tych liczb się dodaje. Błąd względny wyrażenia

$$f = \frac{W_1 \cdot W_2 \cdot \dots \cdot W_m}{Z_1 \cdot Z_2 \cdot \dots \cdot Z_n} \quad (14.4)$$

oznacza się

$$\delta f = \delta W_1 + \delta W_2 + \dots + \delta W_m + \delta Z_1 + \delta Z_2 + \dots + \delta Z_n \quad (14.5)$$

Dla dużej liczby $m+n$ czynników, uwzględniając kompensację błędów o różnych znakach, możemy przyjąć:

$$\delta f = \sqrt{(\delta W_1)^2 + (\delta W_2)^2 + \dots + (\delta W_m)^2 + (\delta Z_1)^2 + (\delta Z_2)^2 + \dots + (\delta Z_n)^2} \quad (14.6)$$

Rozpatrzmy dla przykładu wyznaczenie granicy wytrzymałości na rozciąganie σ_m . Zakładając, iż wielkości zostały wyznaczone z błędem względnym $\delta S_0 = 0,29\%$ i $\delta F_m = 0,25\%$ oraz stosując wyrażenia (14.5) i (14.6) odpowiednio otrzymamy:

$$\delta f = \delta S_0 + \delta F_m = 0,54\% \quad \delta f = \sqrt{(\delta S_0)^2 + (\delta F_m)^2} = 0,38\%$$

Błąd bezwzględny funkcji zmiennych $f(W_i)$, $i = 1, 2, \dots, n$, różniczkowalnej w pewnym obszarze, wywołany przez małe błędy argumentów oceniamy jako:

$$\Delta W = \left| \frac{\partial W}{\partial W_1} \right| \Delta W_1 + \left| \frac{\partial W}{\partial W_2} \right| \Delta W_2 + \dots + \left| \frac{\partial W}{\partial W_n} \right| \Delta W_n \quad (14.7)$$

Przykład 14.2

Metodę różniczki zupełnej zastosujemy do wyznaczenia współczynnika intensywności naprężeń K_Q dla próbki trójpunktowo zginanej. Przyjmijmy następujące dane: $P_{max} = 4200 \cdot 10^{-6}$ MN, $W = 0,02$ m, $B = 0,015$ m, $S = 0,08$ m, $f(a/W) = 2,663$ i założmy, że zostały one zmierzone z dokładnością 1%. Maksymalne błędy

bezwzględne wynoszą: $\Delta P_{max} = \pm 42 \cdot 10^{-6}$ N, $\Delta W = \pm 2 \cdot 10^{-4}$ m, $\Delta B = \pm 1,5 \cdot 10^{-4}$ m, $\Delta S = \pm 8 \cdot 10^{-4}$ m, $\Delta f(a/W) = \pm 0,02663$. Współczynnik intensywności naprężeń wynosi

$$K_Q = \frac{P_Q \cdot S}{B \cdot W^{3/2}} \cdot f\left(\frac{a}{W}\right) = \frac{4200 \cdot 10^{-6} \cdot 0,08}{0,015 \cdot 0,02^{3/2}} \cdot 2,663 = 21,09 \text{ MPa}\sqrt{\text{m}}$$

Pochodne funkcji względem poszczególnych zmiennych wynoszą:

$$\frac{\partial K_Q}{\partial P_Q} = \frac{S}{B \cdot W^{3/2}} \cdot f\left(\frac{a}{W}\right) = \frac{0,08}{0,015 \cdot 0,02^{3/2}} \cdot 2,663 = 5021,4$$

$$\frac{\partial K_Q}{\partial S} = \frac{P_Q}{B \cdot W^{3/2}} \cdot f\left(\frac{a}{W}\right) = \frac{4200 \cdot 10^{-6}}{0,015 \cdot 0,02^{3/2}} \cdot 2,663 = 263,6$$

$$\frac{\partial K_Q}{\partial f(a/W)} = \frac{P_Q \cdot S}{B \cdot W^{3/2}} = \frac{4200 \cdot 10^{-6} \cdot 0,08}{0,015 \cdot 0,02^{3/2}} = 7,92$$

$$\frac{\partial K_Q}{\partial B} = -\frac{P_Q \cdot S}{B^2 \cdot W^{3/2}} \cdot f\left(\frac{a}{W}\right) = -\frac{4200 \cdot 10^{-6} \cdot 0,08}{0,015^2 \cdot 0,02^{3/2}} \cdot 2,663 = -1406,0$$

$$\frac{\partial K_Q}{\partial W} = -1,5 \cdot \frac{P_Q \cdot S}{B \cdot W^{5/2}} \cdot f\left(\frac{a}{W}\right) = -1,5 \cdot \frac{4200 \cdot 10^{-6} \cdot 0,08}{0,015 \cdot 0,02^{5/2}} \cdot 2,663 = -1581,7$$

Bezwzględna wartość błędu współczynnika intensywności naprężeń wyznaczona ze wzoru (14.7) wynosi:

$$\begin{aligned} \Delta K_Q &= \left| \frac{\partial K_Q}{\partial P_Q} \right| \Delta P_Q + \left| \frac{\partial K_Q}{\partial S} \right| \Delta S + \left| \frac{\partial K_Q}{\partial f(a/W)} \right| \Delta f(a/W) + \left| \frac{\partial K_Q}{\partial B} \right| \Delta B + \left| \frac{\partial K_Q}{\partial W} \right| \Delta W = \\ &= 0,21089 + 0,21088 + 0,21091 + 0,2109 + 0,31634 \approx 1,16 \text{ MPa}\cdot\text{m}^{1/2} \end{aligned}$$

Błąd względny wynosi:

$$\delta K_Q = \frac{\Delta K_Q}{|K_Q|} \cdot 100 = \frac{1,16}{21,09} \cdot 100 = 5,5\%$$

Powyższa metoda wyolbrzymia znaczenie błędów poszczególnych pomiarów. W literaturze [2] spotyka się inne metody, które uwzględniają kompensację poszczególnych błędów. Wyznaczona wartość błędu w oparciu o zależność (14.6) dla rozpatrywanego przypadku powyżej wynosi:

$$\Delta K_Q = \sqrt{\left(\frac{\partial K_Q}{\partial P_Q} \Delta P_Q\right)^2 + \left(\frac{\partial K_Q}{\partial S} \Delta S\right)^2 + \left(\frac{\partial K_Q}{\partial f} \Delta f\right)^2 + \left(\frac{\partial K_Q}{\partial B} \Delta B\right)^2 + \left(\frac{\partial K_Q}{\partial W} \Delta W\right)^2}$$

$$= (0.21089^2 + 0.21088^2 + 0.21091^2 + 0.2109^2 + 0.31634^2)^{1/2} \approx 0.53 \text{ Mpa m}^{1/2}$$

a błąd względny wynosi:

$$\delta K_Q = \frac{\Delta K_Q}{|K_Q|} \cdot 100 = \frac{0,53}{21,5} \cdot 100 = 2,5\%$$

14.3. SZACOWANIE BŁĘDÓW DLA LICZNYCH POMIARÓW

Do szacowania błędów wyników uzyskanych na podstawie liczyń pomiarów wykorzystuje się metody statystyczne. Na podstawie ograniczonej liczby danych pozwalają one określić wartość najbardziej prawdopodobną szukaną wielkość oraz prawdopodobny błąd poszczególnego pomiaru.

Pojęcie *statystycznej próby losowej* jest podstawowym pojęciem statystyki matematycznej. Statystyka zajmuje się metodami opisu i analizy liczbowych prawidłowości występujących w tzw. populacjach statystycznych, tj. zbiorowościach elementów zróżnicowanych ze względu na badane cechy. Statystyczne badania populacji dzielą się na pełne i częściowe. W odróżnieniu od pełnych badania częściowe obejmują pewną określoną część populacji zwaną *próbą* (próbką). Wyniki wchodzące do próby będziemy nazywać *elementami próby*, a liczbę tych elementów – *liczebnością* (*licznością*) *próby*.

Potrzeba stosowania badań częściowych wynika z następujących przyczyn:

- populacja statystyczna jest nieskończona lub bardzo duża,
- badanie poszczególnego elementu jest zawsze niszczące,
- badanie poszczególnego elementu jest kosztowne,
- istnieje potrzeba zbadania populacji w bardzo krótkim czasie.

W przypadku, gdy próba jest mała, tzn. gdy jej liczebność n wynosi kilka lub kilkanaście to wyniki próby prezentuje się zazwyczaj w postaci wektora wierszowego liczb $(x_1, x_2, \dots, x_n)$. Dla dużej próby (n – kilkadziesiąt lub więcej) taki długi wierszowy wektor wyników próby wygodnie jest pogrupować (według wielkości) w pewne klasy, tworząc tzw. *szereg rozdzielczy*. Szereg rozdzielczy otrzymuje się z pojedynczych wyników próby przez zaliczenie ich do przyjętych klas wielkości

(lub przedziałów) i podanie liczności odpowiadającej każdej klasie. Prezentacja wyników dużej próby za pomocą szeregu rozdzielczego ma postać tablicy o dwóch kolumnach. Pierwszą kolumną w tablicy są wartości badanej cechy pogrupowane w k klas, a drugą kolumną są liczebności n_j zaobserwowanych w próbie wyników należących do poszczególnych klas o numerze j ($j = 1, 2, \dots, k$).

Tak więc szereg rozdzielczy wyników próby ma postać:

x_j	n_j	lub	$[x_{0j}, x_{1j})$	n_j
x_1	n_1		$x_{o1} - x_{11}$	n_1
x_2	n_2		$x_{o2} - x_{12}$	n_2
$\dots$	$\dots$		$\dots$	$\dots$
x_k	n_k		$x_{ok} - x_{1k}$	n_k

Pierwsza postać szeregu rozdzielczego dotyczy przypadku cechy skokowej, a druga cechy ciągłej. W przypadku cechy ciągłej zaleca się tak podzielić cały zakres zmienności, by liczba k klas (przedziałów klasowych) nie była zbyt mała ani zbyt duża (zwykle $5 \leq k \leq 15$).

Przykład szeregu rozdzielczego wyników próby $n = 40$ próbek badanych ze względu na twardość HRC:

$[x_{0j}, x_{1j})$	n_j
24 – 25	2
25 – 26	8
26 – 27	18
27 – 28	7
28 – 29	5

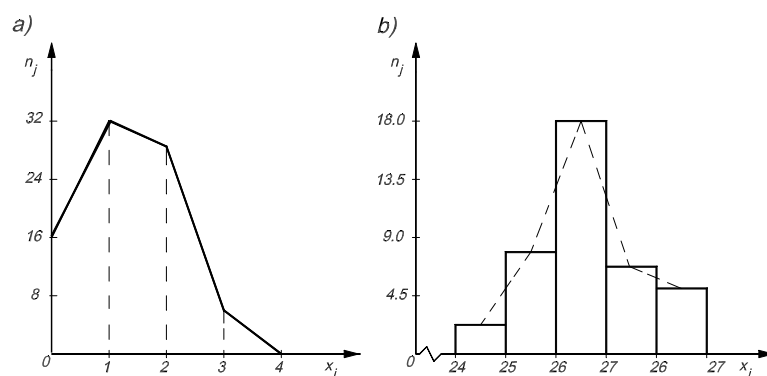
Szereg rozdzielczy wyników próby nazywa się często *rozkładem empirycznym*, gdyż można z niego uzyskać rozkład procentowy (tzw. rozkład częstości) wyników próby rozdzielonych na k klas. Jak również tzw. *szereg skumulowanych liczebności*, który jest podstawą tzw. dystrybuanty empirycznej określanej jako:

$$F_n(x_{1r}) = \frac{1}{n} \sum_{j=1}^r n_j \quad \text{dla} \quad r = 1, 2, \dots, k \quad (14.8)$$

Dla podanego powyżej przykładu szeregu rozdzielczego próby $n = 40$ próbek, rozkład empiryczny (częstości) i jego dystrybuanta empiryczna są następujące:

j	x_j	n_j/n	$F_n(x_{1j})$
1	25	0,050	0,050
2	26	0,200	0,250
3	27	0,450	0,700
4	28	0,175	0,875
5	29	0,125	1,000

Prezentacja wyników próby może ponadto mieć postać graficzną. Dla cechy skokowej szereg rozdzielczy może być podstawą wykresu statystycznego zwanego *wielobokiem liczebności*, który jest graficznym obrazem empirycznego rozkładu cechy skokowej (rys. 14.1a). Szereg graficzny wyników próby dla cechy ciągłej graficznie przedstawia się w postaci wykresu statystycznego zwanego *histogramem*. Jest to wykres w kształcie prostokątów dla poszczególnych przedziałów klasowych w ten sposób, że wysokość każdego prostokąta jest wprost proporcjonalna do liczebności n_j elementów danej j -tej klasy. Łącząc odcinkami środki górnych podstaw prostokątów histogramu można zbudować wykres zwany *diagramem* (rys. 14.1b).



Rys. 14.1. Graficzna prezentacja wyników próby:
a) wielobok liczebności, b) histogram i diagram

Podstawą wnioskowania o populacji na podstawie wyników próby są wartości pewnych charakterystyk próby, zwanych *statystykami z próby*.

Jeżeli n -elementową próbę (losową) oznaczmy jako wektor losowy przez $\mathbf{X} = (X_1, X_2, \dots, X_n)$, a realizację próby (wektor liczbowych wyników próby) przez $\mathbf{x} = (x_1, x_2, \dots, x_n)$, to *statystyką* Z_n nazywamy dowolną, byle o wartościach rzeczywistych funkcję próby $\mathbf{X}$, tzn. $Z_n = g(\mathbf{X})$.

Dla różnych celów i rodzajów wnioskowania z próby o populacji posługujemy się w statystyce matematycznej różnymi statystykami. Wygodnie jest podzielić statystyki z próby na:

- miary skupienia (tendencja centralna),
- miary rozproszenia (rozrzut, dyspersja),
- miary współzależności (korelacja).

1° Do *miar skupienia* wyników próby zaliczamy takie statystyki jak momenty z próby (przede wszystkim średnia arytmetyczna) i niektóre statystyki pozycyjne (mediana, kwartyle).

Statystykę zwaną momentem zwykłym rzędu r z próby określamy jako

$$A_r = \overline{X^r} = \frac{1}{n} \sum_{i=1}^n X_i^r \quad (14.9)$$

Najczęściej używanym w praktyce statystycznej momentem z próby jest średnia arytmetyczna, którą określamy jako

$$A_1 = \overline{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (14.10)$$

Średnia arytmetyczna jest podstawą wnioskowania o wartości oczekiwanej (średniej) rozkładu populacji. Dla konkretnych wyników próby $\mathbf{x} = (x_1, x_2, \dots, x_n)$ wartość statystyki $\overline{X}$, tj. średniej arytmetycznej z próby, oblicza się zwykle jako

$$\overline{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (14.11)$$

Dla dużych prób, których wyniki pogrupowane są w szereg rozdzielczy o k klasach, których środkami są liczby x_j^0 a liczebnościami n_j , wartość średniej arytmetycznej z próby wygodnie jest obliczać jako *średnią ważoną* środków klas:

$$\overline{x} = \frac{1}{n} \sum_{j=1}^k x_j^0 n_j \quad (14.12)$$

Wartość średniej arytmetycznej według powyższego wzoru dla wyników $n = 40$ elementów pogrupowanych w szereg rozdzielczy zawartych w poniższej tabeli wynosi:

$$\overline{x} = \frac{1}{40} \sum_{j=1}^5 x_j^0 n_j = \frac{1065}{40} \approx 26,6$$

$[x_{0j}, x_{1j})$	n_j	x_j^0	$x_j^0 n_j$
24 – 25	2	24,5	49
25 – 26	8	25,5	204
26 – 27	18	26,5	477
27 – 28	7	27,5	192,5
28 – 29	5	28,5	142,5
Σ	40		1065,0

Ważną własnością wartości średniej arytmetycznej jest, że suma odchyleń od niej dla wszystkich wyników x_i w próbie wynosi zero:

$$\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n x_i - n\bar{x} = \sum_{i=1}^n x_i - \sum_{i=1}^n x_i = 0$$

Jako miary tendencji centralnej empirycznego rozkładu uzyskanego w próbie występują też (zwłaszcza dla małych prób) tzw. *statystyki pozycyjne*. Przed określeniem statystyk pozycyjnych wyniki należy uporządkować w niemalejący ciąg liczb

$$x_1 \leq x_2 \leq \dots \leq x_n$$

Statystyką pozycyjną rzędu r nazywamy taką zmienną losową $Z_r^{(n)}$, której wartości zajmują r -te miejsce w uporządkowanym, niemalejącym ciągu wyników n -elementowej próby. Liczbę r nazywamy *rzędem* lub *rangą statystyki pozycyjnej* $Z_r^{(n)}$. Na przykład $Z_1^{(n)}$ jest statystyką pozycyjną rzędu 1 (lub pierwszą) o wartościach najmniejszych w n -elementowej próbie.

Szczególnie ważną rolę odgrywa statystyka pozycyjna $Z_{n/2}^{(n)}$ dla parzystej liczby obserwacji lub $Z_{(n+1)/2}^{(n)}$ dla nieparzystej liczby obserwacji n w próbie. Jest to środkowa statystyka pozycyjna zwana *medianą z próby* (kwantyl z próby rzędu 1/2). Medianę z próby oznaczamy przez *me*. Mediana z próby ma szczególne zastosowanie dla małych prób n . W dużej próbie, pogrupowanej w szereg rozdzielczy, wartość mediany z próby wyznaczamy jako

$$me = x_{0m} + \frac{h}{n_m} \left(\frac{n}{2} - \sum_{j=1}^{r_m-1} n_j \right) \quad (14.13)$$

gdzie:

- r_m – numer przedziału klasowego, w którym znajduje się *me*,
- x_{0m} – początek przedziału klasowego, w którym znajduje się *me*,
- n_m – liczebność przedziału r_m , w którym znajduje się *me*,
- h – rozpiętość (długość) przedziałów klasowych.

Przykład 14.3

Wyniki próby $n = 5$ –elementowej są następujące: 1010, 1100, 1050, 1080, 1060
Po uporządkowaniu w ciąg malejący otrzymujemy:
1010, 1050, 1060, 1080, 1100.

Zatem wartość mediany z próby wynosi $me = Z_{(5+1)/2}^{(5)} = Z_3^{(5)} = 1060$.

Przykład 14.4

Dla wyników próby $n = 40$ pogrupowanych w szereg rozdzielczy

j	$[x_{0j}, x_{lj})$	n_j	$\sum_{i=1}^j n_i$
1	24 – 25	2	2
2	25 – 26	8	10
3	26 – 27	18	28
4	27 – 28	7	35
5	28 – 29	5	40

mamy

$$n/2 = 20, \quad r_m = 3, \quad x_{0m} = 26, \quad n_m = 18, \quad h = 1, \quad \sum_{j=1}^{r_m-1} n_j = 10$$

Wartość mediany wynosi: $me = 26 + 1/18 (20 - 10) = 26,55$

2° Najczęściej używanymi statystykami z próby jako *miara rozproszenia* (rozrzutu) wyników z próby są: rozstęp, wariancja z próby, odchylenie standardowe z próby, współczynnik zmienności z próby.

a) *rozstęp* jako miara rozrzutu wyników z próby używany jest jedynie dla małych prób i określony jest za pomocą pierwszej i ostatniej statystyki pozycyjnej jako:

$$R = Z_n^{(n)} - Z_1^{(n)} = X_{\max} - X_{\min} \quad (14.14)$$

b) podstawową miarą rozrzutu wyników próby jest *wariancja z próby* S^2 i arytmetyczny pierwiastek kwadratowy z wariancji, tj. *odchylenie standardowe z próby* S , które określamy jako:

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2} \quad (14.15)$$

Dla małych prób używa się następującego wzoru:

$$\hat{S}^2 = \frac{n}{n-1} \cdot S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (14.16)$$

Dla dużych prób, których wyniki pogrupowane są w szereg rozdzielczy o k klasach, których środkami są liczby x_j^0 , wartość s^2 wariancji z próby oblicza się ze wzoru:

$$s^2 = \frac{1}{n} \sum_{j=1}^k (x_j^0 - \bar{x})^2 n_j \quad (14.17)$$

Od wartości tej zaleca się odejmowanie tzw. *poprawki na grupowanie*, tj. liczby $1/12h^2$ w przypadku bardzo małej liczby przedziałów, blisko 5.

c) *współczynnik zmienności w próbie* określa się jako

$$V = \frac{S}{\bar{X}} \quad (14.18)$$

Jest to względna (niemianowana) miara rozrzutu wyników z próby.

Przykład 14.5

Dla wyników próby $n = 7$

x_i	1010	1100	1050	1020	1080	1060	1030	$\Sigma = 7350$
$(x_i - \bar{x})^2$	1600	2500	0	900	900	100	400	$\Sigma = 6400$

obliczamy:

$$R = 1100 - 1010 = 90, \quad \bar{x} = \frac{7350}{7} = 1050,$$

$$s^2 = \frac{6400}{7} = 914,3 \quad s = \sqrt{s^2} = 30,24$$

$$\hat{s}^2 = \frac{6400}{6} = 1066,7 \quad \hat{s} = \sqrt{\hat{s}^2} = 32,66$$

$$v = \frac{s}{\bar{x}} = \frac{30,24}{1050} = 0,03,$$

W przypadku, gdy mamy do czynienia z wielkościami o różnym rzędzie wielkości np. granica plastyczności i wydłużenie względne, jednakowy względny rozrzut wyników obu wielkości oznacza jednakową ich regularność w czasie doświadczeń.

Im mniejsza wartość rozstępu, odchylenia standardowego i współczynnika zmienności to wyznacznane wielkości cechują się większą dokładnością.

Przykład 14.6

Dla wyników próby $n = 40$ pogrupowanych w szereg rozdzielczy

$[x_{0j}, x_{1j})$	n_j	x_j^0	$(x_j^0 - \bar{x})^2$	$(x_j^0 - \bar{x})^2 n_j$
24 – 25	2	24,5	4,41	8,82
25 – 26	8	25,5	1,21	9,68
26 – 27	18	26,5	0,01	0,18
27 – 28	7	27,5	0,81	5,67
28 – 29	5	28,5	3,61	18,05
Σ	40			42,40

obliczamy:

$$s^2 = \frac{42,40}{40} - \frac{1^2}{12} = 1,06 - 0,083 = 0,977 \qquad s = \sqrt{0,977} = 0,988$$

3° Najczęściej używaną statystyką z próby jako *miarą korelacji* w dwuwymiarowym rozkładzie empirycznym jest tzw. *współczynnik korelacji z próby*, którego wartość obliczamy jako:

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x \cdot s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (14.19)$$

Wartość r spełnia nierówność $|r| \leq 1$.

W przypadku dużej próby z dwuwymiarowej populacji pogrupowane w tablicę korelacyjną to wartość współczynnika korelacji z próby obliczamy ze wzoru:

$$r = \frac{\sum_{i=1}^k \sum_{j=1}^l (x_i^0 - \bar{x})(y_j^0 - \bar{y}) \cdot n_{ij}}{n \cdot s_x \cdot s_y} \quad (14.20)$$

gdzie:

n_{ij} – liczebność w kratkach tablicy korelacyjnej dla $i = 1, 2, \dots, k$ oraz $j = 1, 2, \dots, l$.

Przykład 14.7:

Dla wyników pogrupowanych w tabeli korelacyjnej

x_j	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
2	35	2,6	3,2	8,32	6,76	10,24
1	26	3,6	-5,8	-20,88	12,96	33,64
5	28	-0,4	-3,8	1,52	0,16	14,44
6	39	-1,5	7,2	-10,8	2,25	51,84
9	31	-4,4	0,8	-3,52	19,36	0,64
$\Sigma = 23$	$\Sigma = 159$			$\Sigma = -25,36$	$\Sigma = 41,49$	$\Sigma = 110,8$

obliczamy: $\bar{x} = \frac{23}{5} = 4,6$ $\bar{y} = \frac{159}{5} = 31,8$ $r = \frac{-25,36}{\sqrt{41,49 \cdot 110,8}} = -0,37$

Statystyką wyrażającą w dwuwymiarowym rozkładzie empirycznym *funkcję liniowej regresji Y względem X* opisujemy

$$y = bx + a \quad (14.21)$$

gdzie *współczynnik liniowej regresji z próby* wynosi

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = r \cdot \frac{s_y}{s_x} \quad (14.22)$$

natomiast

$$a = \bar{y} - b \cdot \bar{x} \quad (14.23)$$

Dla przykładu 14.7: $b = \frac{-25,36}{41,49} = -0,61$ $a = 31,8 + 0,611 \cdot 4,6 = 34,61$

a funkcja regresji liniowej z próby ma postać: $y = -0,61x + 34,61$

Należy pamiętać, że po obliczeniu wyników z próby jakiegokolwiek statystyki, iż ta statystyka jest zmienną losową o określonym rozkładzie prawdopodobieństwa. Dlatego też wnioskowanie z wartości statystyki otrzymanej z wyników jednej konkretnej próby należy opierać na rozkładzie prawdopodobieństwa tej statystyki.

14.4. ZNACZĄCE CYFRY W WYNIKACH POMIARÓW

W czasie opracowywania wyników należy dbać, by zarejestrowane cyfry zawierały wszystkie otrzymane informacje. Odrzucić natomiast należy te wszystkie cyfry, które przy danej metodzie pomiarów nie niosą informacji.

Pomiaru nie można uznać za zakończony, jeśli nie przeprowadzono oceny dokładności pomiaru. Wartość pracy eksperymentatora znacznie maleje, jeśli zaokrąglając wyniki liczbowe gubi informacje otrzymane w wyniku pomiarów potrzebne do wyznaczenia dokładności pomiarów.

Nie ma ogólnej reguły, można natomiast przyjąć, że ostatnia cyfra wyniku liczbowego powinna wyrażać wahania poszczególnych pomiarów. Zaokrąglając wyniki, należy ostatnią z pozostających cyfr zwiększyć o jedność, jeśli pierwsza z odrzuconych cyfr jest równa lub większa od 5. Jeśli odrzucona cyfra jest mniejsza od 5 to pozostawimy ostatnią znaczącą cyfrę nie zmienioną, np. jeśli wartość 105,152 zaokrąglać do trzech znaczących cyfr, należy pozostawić 105,2. Wartość 105,148 zaś zaokrąglamy 105,1. Gdy odrzucona cyfra była 5 bez cyfr dalszych, często pozostałość zaokrągla się do najbliższej liczby parzystej, np. 105,35 do 105,4 a 105,25 do 105,2.

Mnożąc lub dzieląc powinniśmy zachowywać jedynie te cyfry, których dokładność odpowiada dokładności składników. Iloczyn 123,14 i 5,2 należy zapisać jako $123 \cdot 5,1 = 1117$, gdyż mnożnik 5,1 jest podany z dokładnością 1%, dlatego mnożnik 123,14 też należy rozpatrywać z taką dokładnością.

Przy dodawaniu postępujemy w podobny sposób tzn. $14,4 + 3,203$ należy zapisać $14,4 + 3,2 = 17,6$

Literatura

1. RUMSZYSKI L. Z.: Matematyczne opracowanie wyników eksperymentu. Warszawa: WNT 1973
2. GREŃ J.: Statystyka matematyczna. Warszawa: PWN 1987
3. DIETRICH M.: Podstawy konstrukcji maszyn, tom 1. Warszawa: PWN 1986