

PODSTAWY ANALIZY DANYCH POMIAROWYCH

Niepewności pomiarowe

Niepewność wyniku pomiaru odzwierciedla brak dokładnej wiedzy na temat wartości wielkości mierzonej. Wynik pomiaru jest jedynie oszacowaniem (estymatą) wartości wielkości mierzonej z powodu niepewności wynikającej z efektów losowych i niedoskonałej korekty wyniku dla efektów systematycznych. Dlatego też podając wynik pomiaru wielkości fizycznej, należy koniecznie podać także pewną ilościową informację o jakości tego wyniku, tak aby korzystający z tego wyniku mógł ocenić jego wiarygodność. Bez takiej informacji wyniki pomiarów nie mogą być porównywane ani między sobą, ani z wartościami odniesienia. Jako estymator niepewności pomiarowej najczęściej przyjmuje się odchylenie standardowe wyników pomiarów. Niepewność względna pomiaru jest niepewnością pomiaru podzieloną przez wartość bezwzględną wartości mierzonej, gdy wartość mierzona nie jest równa zero.

Wartość oczekiwana i odchylenie standardowe rozkładu prawdopodobieństwa

W przypadku zmiennej losowej dyskretnej przyjmującej N wartości x_i , każdej otrzymanej z prawdopodobieństwem p_i , wartość oczekiwana rozkładu wyrażona jest

$$\mu = \sum_{i=1}^N p_i x_i, \quad (1)$$

zaś za miarę rozrzutu uzyskanych wartości przyjmujemy odchylenie standardowe (σ). Wariancja (kwadrat odchylenia kwadratowego), zdefiniowana jest poprzez zależność:

$$\sigma^2 = \sum_{i=1}^N p_i (x_i - \mu)^2. \quad (2)$$

W przypadku zmiennej losowej ciągłej wartość oczekiwana rozkładu wyrażona jest

$$\mu = \int_{-\infty}^{\infty} x f(x) dx, \quad (3)$$

gdzie $f(x)$ jest funkcją gęstości prawdopodobieństwa, a wariancja wyrażona jest poprzez zależność:

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx. \quad (4)$$

Niepewności statystyczne pomiarów

Ponieważ pomiar niesie informację tylko o części obserwacji z populacji (N pomiarów) za **ocenę prawdziwej** wartości wielkości fizycznej, którą mierzymy, najczęściej przyjmuje się **średnią arytmetyczną**:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i. \quad (5)$$

Poszczególne obserwacje różnią się co do wartości z powodu przypadkowych zmian wielkości wejściowych lub oddziaływań przypadkowych. **Odchylenie standardowe eksperymentalne** s_x obserwacji, estymujące odchylenie standardowe rozkładu prawdopodobieństwa wielkości x , jest zdefiniowane poprzez zależność:

$$s_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2. \quad (6)$$

Inna często spotykana nazwa tej wielkości to **odchylenie standardowe z próby** lub też **statystyczna niepewność standardowa pojedynczego pomiaru**.

Za miarę **niepewności standardowej średniej** przyjmujemy wielkość $s_{\bar{x}}$, zdefiniowaną jako:

$$s_{\bar{x}}^2 = \frac{1}{N(N-1)} \sum_{i=1}^N (x_i - \bar{x})^2. \quad (7)$$

Niepewność średniej arytmetycznej zwana jest również **odchyleniem standardowym średniej**.

Korelacje między wielkościami mierzonymi

W przypadku gdy pomiar określa jednocześnie więcej niż jedną wielkość mierzoną, estymaty tych wielkości i ich niepewności są tylko częścią wiedzy, którą musimy osiąść. Wiedzę na temat zależności między mierzonymi wielkościami możemy uzyskać badając tzw. momenty mieszane, czyli kowariancje i korelacje. Estymatorem kowariancji wielkości x i y jest

$$S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}). \quad (8)$$

$S_{xy} > 0$ oznacza, że wraz ze wzrostem wartości jednej z wielkości, wzrastają wartości drugiej, $S_{xy} < 0$ oznacza, że wraz ze wzrostem wartości jednej z wielkości, maleją wartości drugiej. Wartość kowariancji równa zero świadczy o braku zależności między wielkościami. Ponieważ wartość kowariancji zależy od wartości badanych wielkości, do dyskusji wzajemnych zależności między zmiennymi często wygodniej jest użyć unormowanego i bezwymiarowego współczynnika korelacji zdefiniowanego jako

$$r_{xy} = \frac{S_{xy}}{s_x s_y}. \quad (9)$$

Współczynnik korelacji może przyjmować wartości z przedziału $\langle -1, 1 \rangle$, przy czym $r_{xy} > 0$ oznacza korelację dodatnią (wraz ze wzrostem wartości jednej z wielkości, wzrastają wartości drugiej), $r_{xy} < 0$ oznacza korelację ujemną (wraz ze wzrostem wartości jednej z wielkości, maleją wartości drugiej), $r_{xy} = 0$ oznacza brak zależności, a $r_{xy} = \pm 1$ oznacza dokładną zależność liniową.

Niepewności pomiarowe przyrządu

W specyfikacji technicznej większości przyrządów pomiarowych zazwyczaj podawana jest **rozszerzona niepewność standardowa** (Δ). Jest to wielkość podawana przez producenta przyrządu, który gwarantuje, że jeśli w wyniku pomiaru otrzymamy wartość x , to prawdziwa (dokładna) wartość wielkości mierzonej, z bardzo dużym prawdopodobieństwem, mieści się w przedziale $[x - \Delta, x + \Delta]$. Inne często spotykane nazwy tej wielkości to **dopuszczalny graniczny błąd wskazania** oraz **dokładność**.

Zakładając, że prawdopodobieństwo, że prawdziwa wartość wielkości mierzonej, mieści się w przedziale $[x - \Delta, x + \Delta]$ wynosi 1, a każda z wartości w tym przedziale jest równie prawdopodobna, związana z tym **niepewność standardowa przyrządu** wyraża się wzorem:

$$\sigma_{\Delta} = \frac{\Delta}{\sqrt{3}}. \quad (10)$$

W przypadku dobrej klasy przyrządów pomiarowych w specyfikacji technicznej urządzenia jest bezpośrednio podane odchylenie standardowe pomiaru (np. w przypadku wag używanych na pracowni jest ono podane pod nazwą **powtarzalności**) i to ono powinno być użyte jako wartość niepewności standardowej przyrządu.

Uogólniona niepewność standardowa

Dla pomiarów bezpośrednich, w przypadku występowania wielu niepewności definiuje się uogólnioną niepewność pomiarową jako pierwiastek sumy kwadratów niepewności składowych. W przypadku oceny wartości mierzonej na podstawie N pomiarów przy wykorzystaniu przyrządu o dokładności Δ , **uogólnioną niepewność standardową** (u) wielkości mierzonej bezpośrednio wyznaczamy z zależności

$$u^2 = s_{\bar{x}}^2 + \sigma_{\Delta}^2 = s_{\bar{x}}^2 + \frac{1}{3} \Delta^2. \quad (11)$$

Złożona niepewność standardowa

Jeśli szukana wielkość nie jest wyznaczona w pomiarze bezpośrednio, tylko za pomocą pomiarów pośrednich, jej niepewność możemy wyznaczyć poprzez obliczenie **złożonej niepewności standardowej**. Przyjmijmy, że wielkość y wyznaczamy pośrednio, korzystając z zależności matematycznej $y = f(x_1, x_2, \dots, x_n)$, gdzie wielkości x_i są mierzone bezpośrednio. Ocenę wielkości y otrzymujemy podstawiając uzyskane oceny wartości wielkości x_i . **Jeśli wielkość x_i wyznaczamy na podstawie serii pomiarów, to do zależności matematycznej podstawiamy wartość średnią serii**. Niepewność u_y oceny y obliczana jest za pomocą wzoru na przenoszenie niepewności

$$u_y^2 = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} u_i \right)^2, \quad (12)$$

gdzie u_i są niepewnościami x_i .

Średnia ważona arytmetyczna

Jeżeli ta sama wielkość została zmierzona N niezależnymi metodami, w wyniku czego otrzymano N wartości x_i wraz z niepewnościami u_i , wówczas najlepszą oceną poszukiwanej wielkości jest **średnia ważona arytmetyczna** \bar{x}_w , zaś ocenę niepewności dokonujemy na podstawie niepewności wewnętrznej u_{int} i zewnętrznej u_{ext} , średniej ważonej gdzie:

$$\bar{x}_w = \frac{\sum_{i=1}^N \frac{x_i}{u_i^2}}{\sum_{i=1}^N \frac{1}{u_i^2}}, \quad (13)$$

$$u_{int}^2 = \frac{1}{\sum_{i=1}^N \frac{1}{u_i^2}}, \quad (14)$$

$$u_{ext}^2 = \frac{u_{int}^2}{N-1} \sum_{i=1}^N \left(\frac{x_i - \bar{x}_w}{u_i} \right)^2. \quad (15)$$

Praktyka nakazuje unikać zaniżania niepewności i w sytuacji gdy niepewności zewnętrzna i wewnętrzna są różne za niepewność średniej należy przyjąć większą z obliczonych wartości. Warto jednak zwrócić uwagę, że jeżeli niepewność zewnętrzna jest znacząco mniejsza od wewnętrznej może to oznaczać, że użyte niepewności cząstkowych wyników są przeszacowane, natomiast jeśli niepewność zewnętrzna jest znacznie większa od wewnętrznej może to sugerować niedoszacowanie tych niepewności.

Rozkład normalny

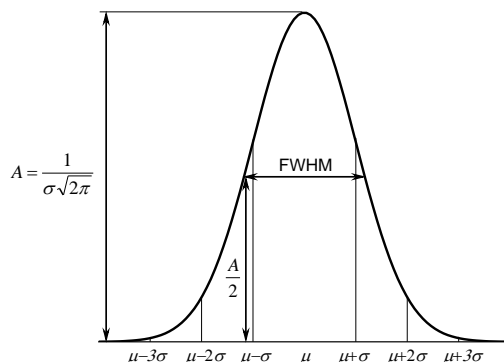
Jednym z najważniejszych rozkładów prawdopodobieństwa jest rozkład normalny. Jest on bardzo często obserwowany, gdyż, jeśli jakaś wielkość jest sumą lub średnią bardzo dużej liczby zmiennych losowych, to niezależnie od rozkładu każdego z tych czynników jej rozkład będzie zbliżony do normalnego. W szczególności wielokrotne powtarzanie tego samego pomiaru daje wyniki rozrzucone wokół określonej wartości. Jeśli wyeliminujemy wszystkie większe przyczyny błędów, zakłada się, że pozostałe mniejsze błędy muszą być rezultatem dodawania się do siebie dużej liczby niezależnych czynników, co daje w efekcie rozkład normalny. Odchylenia od rozkładu normalnego rozumiane są jako wskazówka, że zostały pominięte błędy systematyczne. To stwierdzenie jest głównym założeniem teorii niepewności pomiarowych. Choć rozkład normalny jest często stosowanym założeniem, nigdy nie jest ściśle realizowany. Ma on bowiem niezerową gęstość prawdopodobieństwa dla dowolnej wartości zmiennej losowej, podczas gdy w rzeczywistości zmienne są zawsze ograniczone, a często nieujemne. Mimo to rzeczywisty rozkład jest często bardzo zbliżony do normalnego, stąd zwykle zakłada się, że zmienna ma rozkład normalny. Nie należy jednak robić tego bez sprawdzenia jak wielkie są rozbieżności. Rozkłady dalekie od normalnego (np. z elementami odstającymi) mogą sprawić, że wyniki metod statystycznych będą mylnie interpretowane. Przykładem są tu metody regresji liniowej oraz korelacji Pearsona, które, choć zdefiniowane dla dowolnych rozkładów, mają sensowną interpretację tylko dla wielowymiarowego rozkładu normalnego wektora próbek. Jeśli w próbce występują na przykład elementy odstające, co jest szczególnym przypadkiem rozkładu dalekiego od normalnego, korelacja może przyjąć dowolną wartość między -1 a $+1$, bez względu na rzeczywistą zależność między zmiennymi losowymi. Także regresja będzie dawała błędne rezultaty.

Funkcja gęstości prawdopodobieństwa rozkładu normalnego ze średnią μ i odchyleniem standardowym σ jest przykładem funkcji Gaussa opisanej wzorem:

$$f = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (16)$$

Wykres funkcji prawdopodobieństwa tego rozkładu (Rysunek 1) jest krzywą w kształcie dzwonu, symetryczną względem wartości średniej rozkładu, której szerokość połówkowa (FWHM – *Full Width at Half Maximum*) wynosi $2\sigma\sqrt{\ln 4}$. Punkt przegięcia krzywej znajduje się w odległości jednego odchylenia standardowego od średniej.

Około 68,3% pola pod wykresem krzywej znajduje się w odległości jednego odchylenia standardowego od średniej, około 95,5% w odległości dwóch odchylenia standardowych i około 99,7% w odległości trzech.



Rysunek 1. Wykres gęstości prawdopodobieństwa rozkładu normalnego.

Jeśli $\mu = 0$ i $\sigma = 1$, to rozkład ten nazywa się standardowym rozkładem normalnym z funkcją gęstości prawdopodobieństwa opisywaną przez

$$f = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}. \quad (17)$$

Oznacza to, że jeśli x jest zmienną losową o rozkładzie normalnym ze średnią μ i odchyleniem standardowym σ , to $z = \frac{x-\mu}{\sigma}$ jest zmienną losową o standardowym rozkładzie normalnym.

Standardowy rozkład normalny został stabilizowany i inne rozkłady normalne są prostymi transformacjami rozkładu standardowego. W ten sposób możemy używać tablic dystrybuanty standardowego rozkładu normalnego do wyznaczenia wartości dystrybuanty rozkładu normalnego o dowolnych parametrach.

Testy statystyczne

Ważnym elementem analizy danych doświadczalnych jest sprawdzanie zgodności wyników doświadczeń z przewidywaniami teoretycznymi lub też sprawdzanie wzajemnej zgodności wyników różnych pomiarów, a więc, mówiąc ogólnie, testowanie hipotez. Testy statystyczne pozwalają oszacować prawdopodobieństwo spełnienia pewnej hipotezy statystycznej w populacji na podstawie próby losowej z tej populacji.

W naukach empirycznych najczęściej spotyka się obecnie metody częstościowe testowania, z użyciem p -wartości – służące kontroli błędów decyzyjnych, tak aby w długim horyzoncie czasowym spodziewać się, że nie popełnimy ich częściej, niż założyliśmy (według przyjętego poziomu istotności, np. w 5% przypadków).

Standardowy przebieg procedury weryfikacyjnej zakłada następujące kroki:

1. Sformułowanie hipotezy zerowej i alternatywnej

Hipoteza zerowa (H_0) – jest to hipoteza poddana procedurze weryfikacyjnej, w której zakładamy, że różnica między analizowanymi parametrami lub rozkładami wynosi zero.

Hipoteza alternatywna (H_1) – hipoteza przeciwstawna do weryfikowanej.

2. Wybór statystyki testowej

Budujemy pewną statystykę W , która jest funkcją wyników z próby losowej i wyznaczamy jej rozkład przy założeniu, że hipoteza zerowa jest prawdziwa. Funkcję W nazywa się statystyką testową lub funkcją testową.

3. Określenie poziomu istotności α

Przyjmujemy maksymalne dopuszczalne prawdopodobieństwo popełnienia **błędu pierwszego rodzaju**, który polega na odrzuceniu hipotezy zerowej wtedy, gdy jest ona prawdziwa. Prawdopodobieństwo to jest oznaczane symbolem α i nazywane poziomem istotności. Wybór poziomu istotności zależy od badacza, natury problemu i od tego jak dokładnie chce on weryfikować swoje hipotezy. Na ogół przyjmujemy prawdopodobieństwo bliskie zeru, ponieważ

chcemy, aby ryzyko popełnienia błędu było jak najmniejsze. Najczęściej przyjmuje się poziom istotności 0,05, 0,03 lub 0,01. Z poziomem istotności α związany jest **poziom ufności** zdefiniowany jako $1 - \alpha$.

UWAGA NOMENKLATUROWA!

Klasyczna statystyka matematyczna rozróżnia dwa rodzaje testów: parametryczne i nieparametryczne. Przedmiotem pierwszych jest weryfikacja wartości parametru (np. test 3σ , test t-Studenta) w kontekście których używa się terminu „**istotność**”. Są to więc testy istotności na zadanym poziomie istotności. W teście nieparametrycznym przedmiotem testu jest zależność matematyczna (np. test χ^2 Pearsona). W odniesieniu do takich testów używany jest termin „**zgodność**”, a więc mówimy o teście zgodności na zadanym poziomie zgodności.

4. Wyznaczenie obszaru krytycznego testu

Określamy obszar wartości rozkładu funkcji testowej, których wystąpienie, przy założeniu prawdziwości hipotezy zerowej (H_0), jest wystarczająco mało prawdopodobne (zgodnie z przyjętym poziomem istotności), żeby (empiryczna) realizacja zmiennej losowej mieszcząca się w obszarze krytycznym pozwalała na odrzucenie tej hipotezy. Obszar krytyczny znajduje się zawsze na krańcach rozkładu. Jeżeli obliczona przez nas wartość statystyki testowej znajdzie się w tym obszarze, to weryfikowaną przez nas hipotezę H_0 odrzucamy. Wielkość obszaru krytycznego wyznacza dowolnie mały poziom istotności α , natomiast jego położenie określone jest przez hipotezę alternatywną. Obszar krytyczny od pozostałej części rozkładu statystyki oddzielony jest przez tzw. **wartości krytyczne** testu, czyli wartości odczytane z rozkładu statystyki przy danym α , tak aby spełniona była relacja zależna od sposobu sformułowania H_1 . Z obszarem krytycznym związany jest **przedział ufności**, czyli taki przedział badanego parametru w którym na $(1 - \alpha) \cdot 100\%$ leży jego rzeczywista wartość. Im większy poziom ufności, tym szerszy przedział ufności, a więc tym mniej godzimy się na popełnienie błędu pierwszego rodzaju.

5. Obliczenie p -wartości na podstawie próby

Wyniki próby opracowujemy w odpowiedni sposób, zgodnie z procedurą wybranego testu i są one podstawą do obliczenia wartości statystyki testowej (p -wartości).

6. Podjęcie decyzji

Wyznaczoną na podstawie próby wartość statystyki (p -wartość) porównujemy z wartością krytyczną testu. Jeżeli wartość ta znajdzie się w obszarze krytycznym, to hipotezę zerową należy odrzucić jako nieprawdziwą. Stąd wniosek, że prawdziwa jest hipoteza alternatywna. Jeżeli natomiast wartość ta znajdzie się poza obszarem krytycznym, oznacza to, że brak jest podstaw do odrzucenia hipotezy zerowej. Stąd wniosek, że hipoteza zerowa może, ale nie musi, być prawdziwa, a postępowanie nie dało żadnych dodatkowych informacji uprawniających do podjęcia decyzji o przyjęciu lub odrzuceniu hipotezy zerowej. **Należy podkreślić, że w przypadku, gdy test nie odrzuca hipotezy, nie oznacza to, że udowodniliśmy jej słuszność, a jedynie godzimy się z nią, gdyż nie jest sprzeczna z danymi.**

Powyższa standardowa procedura wymaga przyjęcia arbitralnego poziomu istotności α , a wynikiem weryfikacji jest odpowiedź binarna – albo statystyka testowa mieści się w przedziale ufności, albo nie. **Alternatywnym i nowocześniejszym podejściem jest nie przyjmowania a priori żadnych wartości α (pominięcie punktu 3 i 4), a obliczenie zamiast tego surowej p -wartości i podawanie jej jako wyników weryfikacji. p -wartość odpowiada takiemu poziomowi istotności, przy którym dla zaobserwowanej wartości statystyki testowej zmienia się konkluzja testu.** Pozwala to na przykład na porównywanie istotności różnych konkurencyjnych hipotez statystycznych.

Ważnym parametrem podczas stosowania testów statystycznych jest **moc testu** czyli prawdopodobieństwo uniknięcia **błędu drugiego rodzaju** – przyjęcia hipotezy zerowej, gdy w rzeczywistości jest ona fałszywa. Im większe jest to prawdopodobieństwo, tym lepszy jest dany test jako narzędzie do różnicowania między hipotezą prawdziwą i fałszywą. Moc można wyrazić jako

dopełnienie prawdopodobieństwa popełnienia błędu drugiego rodzaju (β), czyli $1 - \beta$. Moc zależy bezpośrednio i przede wszystkim od wielkości próby użytej w badaniu, rzeczywistej wielkości efektu na tle losowej zmienności w populacji oraz przyjętego poziomu istotności α .

Test 3σ

Jednym z najprostszych testów stosowanych do badania zgodności wyników jest tzw. test 3σ , spotykany w dwóch typach zagadnień:

- Hipoteza teoretyczna głosi, że wielkość mierzona ma wartość μ , a wynik pomiaru x tej wielkości jest wartością zmiennej losowej o wartości oczekiwanej μ i dyspersji σ , gdzie σ jest pierwiastkiem kwadratowym z wariancji. Test prowadzimy w ten sposób, że wyznaczamy wartość $|x - \mu|$ i sprawdzamy, jak uzyskana wartość ma się do wartości 3σ . Jeśli $|x - \mu| > 3\sigma$, to odrzucamy hipotezę o wartości μ wielkości mierzonej, jeśli zaś $|x - \mu| \leq 3\sigma$ to wnioskujemy, że hipoteza nie jest sprzeczna z danymi.
- Hipoteza teoretyczna głosi, że dwa pomiary uzyskane różnymi metodami (w różnych warunkach) są pomiarami tej samej wielkości. Niech wynik x uzyskany jedną metodą będzie wartością zmiennej losowej o dyspersji σ_x , zaś wynik y uzyskany drugą metodą będzie wartością zmiennej losowej o dyspersji σ_y . Test prowadzimy w ten sposób, że wyznaczamy wartość $|x - y|$ i sprawdzamy, jak wartość ta ma się do wartości 3σ ; gdzie $\sigma^2 = \sigma_x^2 + \sigma_y^2$. Ponownie jeśli $|x - y| > 3\sigma$, to odrzucamy hipotezę, że oba pomiary dotyczyły tej samej wielkości (odrzucamy hipotezę o równości wartości oczekiwanych zmiennych x i y). Jeśli zaś $|x - y| \leq 3\sigma$, to wnioskujemy, że hipoteza nie jest sprzeczna z danymi.

UWAGA: W praktyce na ogół nie znamy wartości dyspersji σ , a jedynie jej oszacowanie u , czyli niepewność standardową całkowitą wyniku pomiaru.

Jeśli pomiary opisywane się rozkładem Gaussa, to testowi można nadać interpretację probabilistyczną: dopuszczamy odrzucenie prawdziwej hipotezy nie częściej niż 3 razy na 1000 decyzji (poziom istotności testu wynosi 0,003). Zastąpienie testu 3σ analogicznym testem 2σ oznacza odrzucanie prawdziwej hipotezy nie częściej niż 1 raz na 20 decyzji. **Należy podkreślić, że test 3σ jest testem, w którym podejmuje się wyłącznie decyzję odrzucenia hipotezy, którą się sprawdza, bądź stwierdza się brak podstaw do odrzucenia tej hipotezy.**

Test t-Studenta

Test t-Studenta jest jednym z mniej skomplikowanych i bardzo często wykorzystywanych testów statystycznych używanych do weryfikacji hipotez gdy mamy do czynienia z małymi próbkami (najczęściej przyjmuje się, że próba jest mała gdy jej liczebność jest mniejsza niż 30, przy większych liczebnościach przyjmuje się, że zbiega on do rozkładu normalnego – Rysunek 2 a). Jest na przykład bardzo często wykorzystywany do porównania dwóch średnich między sobą.

1. **Test t-Studenta dla jednej próby** określa, czy średnia z próbki pobranej z populacji o normalnym rozkładzie jest zgodna z wartością hipotetyczną dla danego poziomu istotności. Jeśli zmienne losowe X_1, X_2, \dots, X_n mają jednakowy rozkład prawdopodobieństwa, który jest rozkładem normalnym o średniej μ i odchyleniu standardowym σ , wówczas zmienna:

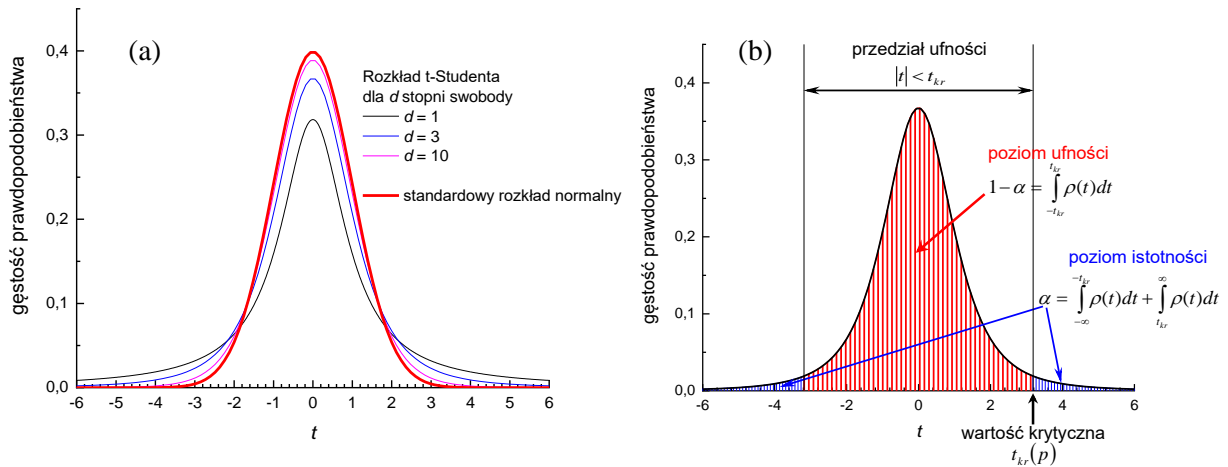
$$t = \frac{\bar{X} - \mu}{s_X \sqrt{\frac{1}{n}}} \quad (18)$$

ma rozkład t-Studenta o $d = n - 1$ stopni swobody. \bar{X} jest wartością średnią z próby, natomiast s_X – odchyleniem standardowym z eksperymentalnym.

2. **Test t-Studenta dla dwóch prób niezależnych** określa, czy średnie z dwóch niezależnych próbek o rozkładzie normalnym są równe, czy też różnią się od siebie i określa poziom istotności różnicy średnich. Jeśli dwie próby o liczebności n_1 i n_2 , wartości średnich \bar{X}_1 i \bar{X}_2 i odchyleniach standardowych eksperymentalnych s_1 i s_2 zostały wylosowane z populacji mających taki sam rozkład normalny, wówczas zmienna:

$$t = \frac{\hat{X}_1 - \hat{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{19}$$

ma rozkład t-Studenta o $d = n_1 + n_2 - 2$ stopni swobody, gdzie $s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$.



Rysunek 2. (a) Porównanie standardowego rozkładu normalnego z rozkładem t-Studenta w zależności od liczby stopni swobody. (b) Rozkład t-Studenta dla trzech stopni swobody z zaznaczoną wartością krytyczną, przedziałem ufności oraz współczynnikiem ufności, dla poziomu istotności $\alpha = 0,05$.

Aby rozstrzygnąć czy wartość t-Studenta jest istotna statystycznie, to znaczy wskazuje na istotną statystycznie różnicę między \hat{X}_1 i \hat{X}_2 (czy \hat{X} i μ), należy określić wymagany poziom istotności prowadzonego testu (Rysunek 2 b). Hipotezy przyjmujemy jako niesprzeczną z danymi, ze współczynnikiem ufności $1 - \alpha$, jeśli obliczona wartość t mieści się w przedziale ufności. Wartości krytyczne, tzn. granice przedziałów ufności dla zadanej liczby stopni swobody (d) oraz poziomu istotności (α) są stabelaryzowane (Tabela 1). Aby sprawdzić, czy obliczona wartość t-Studenta wskazuje na istotną statystycznie zależność, musimy sprawdzić czy znajduje się w przedziale ufności, tzn. porównać wartość bezwzględną otrzymanej wartości t z wartościami krytycznymi rozkładu dla zadanej liczby stopni swobody (d) oraz poziomu istotności (α). Jeżeli wartość bezwzględna obliczonej wartości testu t-Studenta w naszym badaniu jest większa od wartości krytycznej oznacza to, że istotność statystyczna testu jest mniejsza niż przyjęty poziom istotności co uprawnia do odrzucenia hipotezy (na zadanym poziomie istotności). Jeżeli zaś wartość bezwzględna obliczonej wartości testu t-Studenta w naszym badaniu jest mniejsza od wartości krytycznej nie mamy podstaw do odrzucenia hipotezy.

Tabela 1. Wartości krytyczne rozkładu t-Studenta dla zadanej liczby stopni swobody d i wybranych poziomów istotności α .

$\alpha \backslash d$	1	2	3	4	5	6	7	8	9	10
0,01	63,6567	9,9248	5,8409	4,6041	4,0321	3,7074	3,4995	3,3554	3,2498	3,1693
0,05	12,7062	4,3029	3,1824	2,7764	2,5706	2,4469	2,3646	2,306	2,2622	2,2281
$\alpha \backslash d$	11	12	13	14	15	16	17	18	19	20
0,01	3,1058	3,0545	3,0123	2,9768	2,9467	2,9208	2,8982	2,8784	2,8609	2,8453
0,05	2,201	2,1788	2,1604	2,1448	2,1314	2,1199	2,1098	2,1009	2,093	2,086

Jeśli dysponujemy oprogramowaniem, który umożliwia wyznaczenie dystrybuanty rozkładu (całki rozkładu prawdopodobieństwa), lepiej jest, w miejsce wartości krytycznej, podawać

tzw. prawdopodobieństwo testowe (p -wartość, p -value) – prawdopodobieństwo kumulatywne wylosowania próby takiej lub bardziej skrajnej jak zaobserwowana. Jeśli p -wartość jest niższa, niż przyjęty z góry poziom istotności statystycznej można postępować tak, jakby hipoteza została odrzucona.

Ponownie należy podkreślić, że **test t-Studenta jest testem, w którym na podstawie wyników próby losowej podejmuje się wyłącznie decyzję odrzucenia hipotezy, którą się sprawdza, bądź stwierdza się brak podstaw do odrzucenia tej hipotezy.**

Metoda najmniejszych kwadratów

Załóżmy, że wielkości fizyczne x i y wiąże zależność $y = f(x, a_1, a_2, \dots, a_n)$, gdzie a_k są nieznanymi nam parametrami. Dla $N > n$ różnych wartości x_i mierzymy odpowiadające im wartości y_i ($i = 1, 2, \dots, N$). Metoda najmniejszych kwadratów pozwala na wyznaczenie wartości parametrów a_k oraz ich niepewności na podstawie tych pomiarów.

Wyznaczenie wartości parametrów a_k odbywa się poprzez żądanie minimalizacji funkcji χ^2 , która mierzy odchylenie zadanej zależności funkcyjnej od punktów doświadczalnych, tzw. reszt (ε_i) w stosunku do odchylenia standardowego (σ_i):

$$\chi^2 = \sum_{i=1}^N \frac{\varepsilon_i^2}{\sigma_i^2} = \sum_{i=1}^N \frac{(y_i - f(x_i, a_1, a_2, \dots, a_n))^2}{\sigma_i^2} \quad (20)$$

względem parametrów a_1, a_2, \dots, a_n .

Oznacza to rozwiązanie układu n równań na n współczynników:

$$\frac{\partial \chi^2}{\partial a_k} = 0.$$

Stosując metodę najmniejszych kwadratów wykorzystującą odchylenie standardowe σ_i , na ogół nie znamy jej wartości i dlatego w praktyce w ich miejsce stosujemy oszacowania u_i , czyli niepewności standardowe wyników pomiaru. Dodatkowo metoda ta zakłada, że dla **dokładnie** ustalonej wartości zmiennej niezależnej x_i , wykonywany jest pomiar zmiennej zależnej, w którego wyniku otrzymujemy wartość y_i z niepewnością u_i . W praktyce najczęściej spotykamy problemy, w których obie zmienne są wyznaczane z niepewnościami. Jeśli chcemy uzyskać wyniki analityczne w formie zamkniętej, to nadal stosujemy najprostszą formę metody najmniejszych kwadratów, a za zmienną niezależną (x) przyjmujemy wielkość znaną dokładniej.

W celu obliczenia niepewności uzyskanych wartości współczynników a_1, a_2, \dots, a_n korzysta się ze wzoru na błąd pośredni funkcji zależnej od parametrów $f(y_i)(a_1(y_i), a_2(y_i), \dots, a_n(y_i))$ przyjmując, że niepewność pomiarowa wynika tylko z niepewności zmiennej y

$$\sigma_f^2 = \sum_{i=1}^N \sigma_i^2 \left(\frac{\partial f}{\partial y_i} \right)^2. \quad (21)$$

W przypadku gdy niepewności pomiarowe są nieznanne, to przyjmuje się że są takie same i szacuje się je z rozrzutu danych pomiarowych, zakładając, że są równe odchyleniu standardowemu reszt (s) opisanego zależnością:

$$s^2 = \frac{1}{d} \sum_{i=1}^N (y_i - f(x_i, a_1, \dots, a_n))^2 = \frac{1}{d} \sum_{i=1}^N \varepsilon_i^2. \quad (22)$$

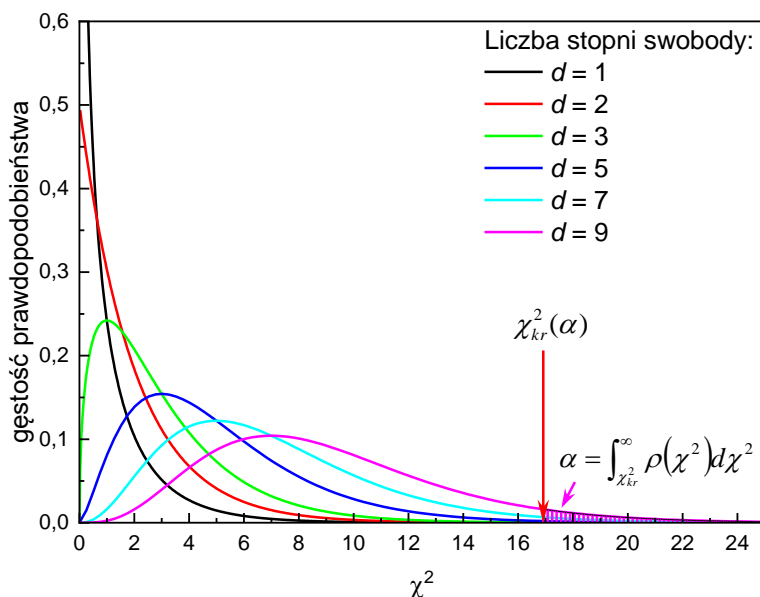
gdzie $d = N - n$ to liczba stopni swobody (liczba niezależnych wyników obserwacji (N) pomniejszonej o liczbę związków (n), które łączą te wyniki ze sobą). Jest to równoważne z zapostulowaniem, że $\chi^2 = N - n$, czyli wartości oczekiwanej rozkładu.

Warto zwrócić uwagę, że metoda najmniejszych kwadratów jest bardzo czuła na punkty najbardziej oddalone od średniej, które mogą wprowadzić największy błąd. Jeśli mamy w danych pojedynczą zakłócającą obserwację bardzo oddaloną od reszty, przyciągnie ona do siebie linię trendu. Takie zjawisko jest niestety częste w realnych danych, nie należy więc stosować metody najmniejszych kwadratów bez sprawdzenia (choćby na wykresie rozrzutu) braku elementów odstających i ich usunięcia.

Test χ^2 Pearsona

Mając dwie serie pomiarów x i y , przyjmując, że wartość x jest mierzona dokładnie, a σ_i jako dyspersję wartości y_i , wartość χ^2 zdefiniowana wzorem (1) poprzez porównanie z modelowym rozkładem χ^2 dla określonej liczby stopni swobody pozwala osądzić zasadność wyboru zależności teoretycznej $y = f(x)$.

Rozkład χ^2 o liczbie stopni swobody równej d (Rysunek 3), to rozkład zmiennej losowej, która jest sumą d kwadratów niezależnych zmiennych losowych o standardowym rozkładzie normalnym. Wartość oczekiwana (średnia) tego rozkładu jest równa d natomiast jego odchylenie standardowe wynosi $\sqrt{2d}$.



Rysunek 3. Rozkład χ^2 w zależności od liczby stopni swobody (d) z zaznaczonym dla $d=9$, poziomem zgodności $\alpha = 0,05$ i odpowiadającą mu wartością krytyczną χ_{kr}^2 .

Jednym ze sposobów na walidację stosowanego modelu jest porównanie otrzymanej wartości χ^2 z wartością krytyczną (χ_{kr}^2) rozkładu dla zadanej liczby stopni swobody d i poziomu zgodności α (Tabela 2). Jeśli $\chi^2 > \chi_{kr}^2$ to odrzucamy model jako sprzeczny z danymi na zadanym poziomie ufności. Jeśli $\chi^2 < \chi_{kr}^2$ to przyjmujemy, że zaproponowany model jest niesprzeczny z naszymi danymi. Warto zauważyć, że jeśli otrzymana wartość χ^2 jest wyraźnie mniejsza od liczby stopni swobody (wartości oczekiwanej rozkładu), to najprawdopodobniej pomiary nie były niezależne lub przeceniliśmy wartości niepewności pomiarowych. Wynika to z faktu, że dane podlegają pewnemu rozkładowi, więc bardzo mało prawdopodobne jest, aby wszystkie otrzymane wyniki tak dobrze pasowały do modelu teoretycznego.

Jeśli dysponujemy oprogramowaniem, który umożliwia wyznaczanie dystrybuanty rozkładu (całki rozkładu prawdopodobieństwa), lepiej jest, w miejsce wartości krytycznej, podawać tzw. prawdopodobieństwo testowe (p -wartość), prawdopodobieństwo że zmienna rozkładu χ^2 przyjmie wartość większą niż wartość χ^2 uzyskaną z danych. Jeśli p -wartość jest niższa, niż przyjęty z góry poziom istotności statystycznej można postępować tak, jakby hipoteza została odrzucona. Podejście to dostarczając czytelnikowi informację o poziomie zgodności (α) jaki należy dopuścić aby móc przyjąć prawdziwość badanego modelu, umożliwia mu wyrażenie własnej opinii w tej kwestii.

Ponieważ wartość χ^2 zależy od liczby stopni swobody, czyli liczby danych pomiarowych, często wygodniej jest do walidacji modelu użyć wartości zredukowanej $\widehat{\chi^2}$, czyli wartości χ^2 przypadającej na stopień swobody:

$$\widehat{\chi^2} = \frac{1}{d} \chi^2, \quad (23)$$

gdzie, podobnie jak poprzednio, $d = N - n$ to liczba stopni swobody. Oczekiwana wartość $\widehat{\chi^2}$ jest bliska 1, więc doświadczalne wartości wokół wartości 1 oznaczają dobre dopasowanie modelu do danych doświadczalnych, i oznacza to, że różnice między danymi i modelem są rzędu niepewności. Jeśli niepewności są przeszacowane wartość $\widehat{\chi^2}$ będzie znacznie mniejsza od jeden, zaś w przypadku niedopasowania modelu lub niedoszacowanych niepewności, wartość $\widehat{\chi^2}$ będzie znacznie większa niż 1. Warto jednak zauważyć, że termin „bliska 1”, oznacza dla poziomu zgodności $\alpha = 0,05$ i dla $d = 1$ wartość $\widehat{\chi^2} = 3,84$, dla $d = 10$ wartość $\widehat{\chi^2} = 1,83$, a dla $d = 30$ wartość $\widehat{\chi^2} = 1,46$.

Ważne jest aby pamiętać, że stosując test χ^2 Pearsona nie uważamy, że udowodniliśmy słuszność badanego modelu, lecz jedynie z pewnym prawdopodobieństwem stwierdzamy jego niesprzeczność z obserwowanymi danymi.

Tabela 2. Wartości krytyczne rozkładu χ^2 dla zadanej liczby stopni swobody d i wybranych poziomów zgodności α .

Poziom zgodności α	Wartości krytyczne χ_{kr}^2 , dla podanych liczb d stopni swobody									
	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$	$d = 7$	$d = 8$	$d = 9$	$d = 10$
0,010	6,63	9,21	11,35	13,28	15,08	16,81	18,47	20,09	21,67	23,21
0,050	3,84	5,99	7,82	9,49	11,07	12,59	14,07	15,51	16,92	18,31
	$d = 11$	$d = 12$	$d = 13$	$d = 14$	$d = 15$	$d = 16$	$d = 17$	$d = 18$	$d = 19$	$d = 20$
0,010	24,73	26,22	27,69	29,14	30,58	32,00	33,41	34,81	36,19	37,57
0,050	19,68	21,03	22,36	23,68	25,00	26,30	27,59	28,87	30,14	31,41
	$d = 21$	$d = 22$	$d = 23$	$d = 24$	$d = 25$	$d = 26$	$d = 27$	$d = 28$	$d = 29$	$d = 30$
0,010	38,93	40,29	41,64	42,98	44,31	45,64	46,96	48,28	49,59	50,89
0,050	32,67	33,92	35,17	36,42	37,65	38,89	40,11	41,34	42,56	43,77

Warto podkreślić, że do zastosowania testu χ^2 Pearsona niezbędna jest wiedza na temat niepewności pomiarowych zmiennej zależnej. Niestety w praktyce fizyka eksperymentatora nie zawsze jest łatwo poprawnie oszacować niepewności pomiarowe. Dlatego testowi χ^2 Pearsona zawsze powinna towarzyszyć refleksja, czy na przykład duża wartość χ^2 wynika nie tyle z nieprawidłowego modelu, a raczej z niedoszacowania niepewności pomiarowych. Często spotykaną praktyką jest więc przeskalowanie przyjętych niepewności pomiarowych zgodnie z wartością oczekiwaną χ^2 i postawienie sobie pytania, czy tak przyjęte niepewności miałyby uzasadnienie w wykonywanym doświadczeniu.

Obliczaliśmy χ^2 jako

$$\chi^2 = \sum_{i=1}^N \frac{\varepsilon_i^2}{\sigma_i^2} = \sum_{i=1}^N \frac{(y_i - f(x_i, a_1, a_2, \dots, a_n))^2}{\sigma_i^2}.$$

Przyjmijmy że rzeczywiste niepewności są równe $u_i = \alpha \sigma_i$. Chcemy tak dobrać współczynniki skalujące niepewności (α) aby χ^2 było równe jego wartości oczekiwanej:

$$\chi_{skal}^2 = \sum_{i=1}^N \frac{(y_i - f(x_i, a_1, a_2, \dots, a_n))^2}{u_i^2} = d.$$

Ponieważ

$$\chi_{skal}^2 = \sum_{i=1}^N \frac{(y_i - f(x_i, a_1, a_2, \dots, a_n))^2}{\alpha^2 \sigma_i^2} = \frac{1}{\alpha^2} \sum_{i=1}^N \frac{(y_i - f(x_i, a_1, a_2, \dots, a_n))^2}{\sigma_i^2} = \frac{1}{\alpha^2} \chi^2$$

mamy

$$\alpha = \sqrt{\frac{\chi^2}{\chi_{skal}^2}} = \sqrt{\frac{\chi^2}{d}} = \sqrt{\widehat{\chi^2}}.$$

Czyli czynnikiem skalującym niepewności pomiarowe jest pierwiastek z wartości zredukowanej χ^2 . Jeśli tak przeskalowane niepewności, zostaną w **uzasadniony** sposób przyjęte przez badacza, to oznacza, że nie mamy podstaw aby odrzucić badany model. Zastosowanie tej procedury może być również przydatne, gdy wyznaczona wartość χ^2 jest zaskakująco mała, co może świadczyć na przykład o pierwotnym przeszacowaniu niepewności pomiarowych.

W przypadku braku informacji o niepewnościach pomiarowych, metoda najmniejszych kwadratów może również posłużyć do ich wyznaczenia (przy założeniu poprawności modelu opisującego dane doświadczalne). Zakłada się wówczas niepewności równe średniej niepewności przypadającej na stopień swobody:

$$s^2 = \frac{1}{d} \sum_{i=1}^N \varepsilon_i^2. \quad (24)$$

Wówczas

$$\widehat{\chi^2} = \frac{1}{d} \sum_{i=1}^N \frac{\varepsilon_i^2}{s^2} = \frac{1}{d} \frac{1}{s^2} \sum_{i=1}^N \varepsilon_i^2 = \frac{1}{d} \frac{1}{\frac{1}{d} \sum_{i=1}^N \varepsilon_i^2} \sum_{i=1}^N \varepsilon_i^2 = 1$$

czyli jest równie jego wartości oczekiwanej.

Istotność parametrów dopasowania

Ważnym elementem analizy dopasowanego modelu jest sprawdzanie jak istotne są parametry, którymi jest opisany. Wyobraźmy sobie doświadczenie, którego wyniki są modelowo opisywane zależnością liniową $y = Ax$. W praktyce, bardzo często badacz nie wie jaki model opisuje jego dane, a jedynie go postuluje. Może się więc zdarzyć, że postulowany przez badacza model będzie postaci $y = Ax + B$. Model ten będzie również zgodny z danymi ale spodziewamy się, że otrzymane oszacowanie parametru B będzie bliskie zeru. Dlatego po każdym dopasowaniu, należy przeanalizować otrzymane oszacowania parametrów pod kątem ich istotności w zastosowanym modelu. Do tego celu stosuje się najprostsze testy (test 3σ , test t-Studenta), które w szybki sposób porównują otrzymane oszacowanie parametru (wraz z jego niepewnością) z zerem i odpowiadzą na pytanie jak prawdopodobne jest, że parametr ten jest istotny w dopasowaniu.

Ważne jest podkreślenia, że należy zachować szczególną ostrożność przy określaniu ilości parametrów. W szczególności należy unikać przeparametryzowania modelu, nieistotność parametru nie znaczy jednak automatycznie, że zmienną należy usunąć z modelu. W praktyce przy doborze parametrów objaśniających model należy zawsze kierować się zdrowym rozsądkiem i teorią dotyczącą badanego zagadnienia.

Badanie zależności parametrów dopasowania

Przypominamy, że w przypadku dopasowania modelu postaci $y(x, a_1, a_2, \dots, a_n)$ metodą najmniejszych kwadratów konstruujemy funkcję

$$\chi^2 = \sum_{i=1}^N \frac{(y_i - y(x_i, a_1, a_2, \dots, a_n))^2}{u_i^2},$$

a następnie minimalizujemy ją względem wszystkich parametrów modelu $\frac{\partial \chi^2}{\partial a_k} = 0$, wyznaczając w ten sposób ocenę parametrów modelu a_1, a_2, \dots, a_n .

Jak łatwo sobie wyobrazić w przypadku dopasowania modelu z więcej niż jednym parametrem pomiędzy poszczególnymi parametrami mogą występować zależności. Aby je badać oblicza się macierz kowariancji parametrów dopasowania, która stanowi uogólnienie pojęcia wariancji i kowariancji na przypadek wielowymiarowy. Macierz ta ma postać:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}, \quad (25)$$

gdzie σ_k^2 jest wariancją k -tego parametru, a σ_{kl} kowariancją między k -tym i l -tym parametrem. Warto zauważyć, że macierz Σ jest macierzą symetryczną, której wyznacznik jest nieujemny.

Tak jak w przypadku analizy zależności dwóch zmiennych losowych, często wygodniejsza jest analiza macierzy korelacji. Macierz ta ma postać:

$$corr = \begin{bmatrix} 1 & \frac{\sigma_{12}}{\sigma_1\sigma_2} & \cdots & \frac{\sigma_{1n}}{\sigma_1\sigma_n} \\ \frac{\sigma_{21}}{\sigma_2\sigma_1} & 1 & \cdots & \frac{\sigma_{2n}}{\sigma_2\sigma_n} \\ \vdots & \cdots & \ddots & \vdots \\ \frac{\sigma_{n1}}{\sigma_n\sigma_1} & \frac{\sigma_{n2}}{\sigma_n\sigma_2} & \cdots & 1 \end{bmatrix}. \quad (26)$$

Wówczas, wartości wszystkich elementów macierzy należą do przedziału $\langle -1,1 \rangle$ (ponieważ są współczynnikami korelacji), wszystkie elementy leżące na diagonalu macierzy równe są 1 (określa to stopień skorelowania i -tego parametru z nim samym), a wyznacznik tej macierzy należy do przedziału $\langle 0,1 \rangle$. Wyznacznik macierzy korelacji jest miernikiem współliniowości (skorelowania) zmiennych objaśniających. Im bliższy 1, tym niższy jest stopień wzajemnego skorelowania zmiennych objaśniających. Im bliższy 0, tym siła tej korelacji większa. Wartości wyznacznika bliskie 0 świadczą o złym doborze parametrów objaśniających. Jeśli wyznacznik jest zbyt mały, musimy modyfikować model – powinniśmy z niego eliminować zmienne współliniowe. Istnienie silnej korelacji między zmiennymi objaśniającymi negatywnie wpływa na efektywność estymatorów parametrów modelu.

Jako ocenę wariancji parametru a_k przyjmujemy jego kwadrat niepewności. W celu obliczenia tej niepewności przyjmujemy, że wynika ona tylko z niepewności zmiennej y i korzystamy ze wzoru na niepewność pośrednią funkcji $a_k(y_i)$

$$u_{a_k}^2 = \sum_{i=1}^N u_{y_i}^2 \left(\frac{\partial a_k}{\partial y_i} \right)^2. \quad (27)$$

Jako ocenę kowariancji dwóch parametrów przyjmujemy

$$s_{a_k a_l} = \sum_{i=1}^N u_{y_i}^2 \left(\frac{\partial a_k}{\partial y_i} \right) \left(\frac{\partial a_l}{\partial y_i} \right), \quad (28)$$

a ich współczynnika korelacji:

$$r_{a_k a_l} = \frac{s_{a_k a_l}}{u_{a_k} u_{a_l}}. \quad (29)$$

Iteracyjna metoda najmniejszych kwadratów

Gdy postać funkcji $f(x_i, a_1, \dots, a_n)$ opisującej model nie pozwala na analityczne zminimalizowanie funkcji χ^2 w celu wyznaczenia wartości parametrów modelu (a_1, \dots, a_n) konieczne są przybliżenia numeryczne. Wykorzystując różne algorytmy numeryczne np.: metodę Newtona, metodę największego spadku, algorytm Levenberga-Marquardta można startując z danego punktu początkowego (początkowego przybliżenia), poprzez kolejne, iteracyjnie uzyskiwane, przybliżenia uzyskać zbieżność do optymalnej estymaty parametrów (a_1, \dots, a_n) według metody najmniejszych kwadratów. W dalszym ciągu zakładamy, że najlepszym dopasowaniem modelu $f(x_i, a_1, \dots, a_n)$ do serii danych pomiarowych $(x_i, y_i \pm \sigma_i)$ jest to minimalizujące

$$\chi^2 = \sum_{i=1}^N \frac{(y_i - f(x_i, a_1, \dots, a_n))^2}{\sigma_i^2}.$$

Procedura optymalizacji polega na wykonaniu kolejnych kroków:

1. Zdefiniowaniu początkowych wartości parametrów modelu (a_1^0, \dots, a_n^0) tak, aby były one bliskie rzeczywistym wartościom parametrów (a_1, \dots, a_n) i umożliwiały otrzymanie zbieżności algorytmu;
2. Określeniu tolerancji $\delta\chi^2$ zmian funkcji χ^2 ;
3. Obliczeniu funkcji χ^2 dla parametrów początkowych: $\chi^2(\mathbf{a}^0) = \sum_{i=1}^N \frac{(y_i - f(x_i, a_1^0, \dots, a_n^0))^2}{\sigma_i^2}$;
4. Obliczenie „lokalnie” optymalnej korekcji estymat oraz iteracyjna zmiana parametrów (a_1^i, \dots, a_n^i) , do momentu spełnienia warunku, że kolejne wartości funkcji χ^2 nie różnią się od siebie o więcej niż założona tolerancja.

W celu obliczenia niepewności uzyskanych wartości współczynników (a_1, \dots, a_n) ponownie korzysta się z propagacji niepewności pomiarowych przyjmując, że niepewność pomiarowa wynika tylko z niepewności zmiennej y .

Istnieje jednak szereg przykładów funkcji, w których parametry pojawiają się w formie nieliniowej, ale po wykonaniu zamiany zmiennych, zależności te można przekształcić do postaci liniowej funkcji szukanych parametrów – być może kosztem przedefiniowania niektórych z nich. Po wyznaczeniu ocen parametrów zależności liniowej oraz ich niepewności i oceny kowariancji, odwracamy transformację i uzyskujemy oceny parametrów zależności oryginalnej, a ich niepewności i ocenę ich kowariancji wyznaczamy na podstawie wzoru na propagację niepewności pomiarowych. Jeśli niepewności zmiennej zależnej są na tyle małe, że przybliżenie propagacji niepewności pomiarowych jest wystarczająco dokładne, to oceny wartości parametrów uzyskane na podstawie zlinearyzowanej zależności są bardzo bliskie wartościom wynikającym z metody najmniejszych kwadratów zastosowanej do oryginalnego problemu.

Niepewności oraz korelacje funkcji

Dla funkcji $f(A_1, A_2, \dots, A_n)$ zmiennych A_i , której rozwinięcie:

$$f(A_1, A_2, \dots, A_n) \approx f(\mu_1, \mu_2, \dots, \mu_n) + \sum_{i=1}^n \frac{\partial f}{\partial A_i} (A_i - \mu_i) \quad (30)$$

do wyrazów liniowych wokół punktu $(\mu_1, \mu_2, \dots, \mu_n)$ dostarcza jej akceptowalnej aproksymacji w przestrzeni zmiennych A_i w hiperkostce o rozmiarach kilku niepewności u_i , wzór na niepewność wartości tej funkcji przyjmuje postać:

$$u_f^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial A_i} \right)^2 u_i^2 + \sum_{i=1, j \neq i}^n \frac{\partial f}{\partial A_i} \frac{\partial f}{\partial A_j} c_{ij}. \quad (31)$$

W przypadku dwóch funkcji $f(A_1, A_2, \dots, A_n)$ oraz $g(A_1, A_2, \dots, A_n)$ zmiennych A_i o niepewnościach u_i , do których to funkcji mają zastosowanie założenia dotyczące wyznaczania niepewności wielkości pośrednio mierzonej (przenoszenia niepewności), ocena kowariancji między tymi funkcjami przyjmuje postać:

$$c_{fg} = \sum_{i=1}^n \frac{\partial f}{\partial A_i} \frac{\partial g}{\partial A_i} u_i^2 + \sum_{i=1, j \neq i}^n \frac{\partial f}{\partial A_i} \frac{\partial g}{\partial A_j} c_{ij}. \quad (32)$$

Należy zwrócić uwagę na następujący fakt: nawet gdy zmienne A_i są statystycznie niezależne, funkcje f i g pozostają skorelowane – fluktuacje wartości zmiennych wymuszają odpowiednią współzmiennność wartości tych funkcji.