PODSTAWY ANALIZY DANYCH POMIAROWYCH CZ. 3

Test 3σ

Jednym z najprostszych testów 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.

Należy podkreślić, że w przypadku, gdy test 3σ nie odrzuca hipotezy, nie oznacza to, że udowodniliśmy jej słuszność, a jedynie godzimy się z nią, gdyż nie jest sprzeczna z danymi.

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 l raz na 20 decyzji.

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

Metoda najmniejszych kwadratów

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

Wyznaczenie wartości parametrów a_j 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):

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

względem parametrów $a_1,a_2,\dots,a_k$.

Oznacza to rozwiązanie układu k równań na k 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 $a_1, a_2, \dots, a_k$ korzysta się ze wzoru na błąd pośredni funkcji zależnej od parametrów $f(y_i)\big(a_1(y_i), a_2(y_i), \dots, a_k(y_i)\big)$ przyjmując, że niepewność pomiarowa wynika tylko z niepewności zmiennej y

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

W przypadku gdy niepewności pomiarowe są nieznane, 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:

$$s^2 = \frac{1}{d}\sum_{i=1}^{N}\big(y_i - f(x_i, a_1, \dots, a_k)\big)^2 = \frac{1}{d}\sum_{i=1}^{N} \varepsilon_i^2. \tag{3}$$

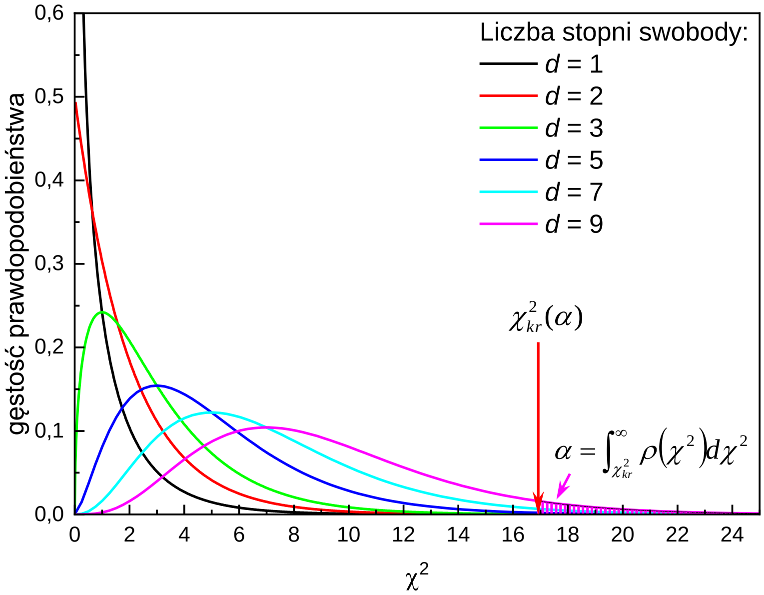
gdzie $d = N - k$ to liczba stopni swobody (liczba niezależnych wyników obserwacji (N) pomniejszonej o liczbę związków (k), które łączą te wyniki ze sobą). Jest to równoważne z zapostulowaniem, że $\chi^2 = N - k$, czyli wartości oczekiwanej rozkładu.

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 1), 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 1. Rozkład χ^2 w zależności od liczby stopni swobody (d)
z zaznaczonym dla d = 9, poziomem zgodności α = 0,05 i odpowiadającą mu wartością krytyczną χ^2_{kr}.

1. Jednym ze sposobów na walidację stosowanego modelu jest porównanie otrzymanej wartości χ^2 z wartością krytyczną (χ^2_{kr}) rozkładu dla zadanej liczby stopni swobody d i poziomu zgodności α (Tabela 1). Jeśli $\chi^2 > \chi^2_{kr}$ to odrzucamy model jako sprzeczny z danymi na zadanym poziomie ufności. Jeśli $\chi^2 < \chi^2_{kr}$ to przyjmujemy, że zaproponowany model jest niesprzeczny z naszymi danymi. Często stosowanym punktem odcięcia między wynikami istotnymi i nieistotnymi jest poziom zgodności 0,05. 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.
2. Jeśli dysponujemy oprogramowaniem, który umożliwia wyznaczanie dystrybuanty rozkładu (całki rozkładu prawdopodobieństwa), lepiej jest, w miejsce wartości krytycznej, podawać tzw. prawdopodobieństwo testowe (wartość p, p-wartość), prawdopodobieństwo że zmienna rozkładu χ^2 przyjmie wartość większą niż wartość χ^2 uzyskaną z danych. Jeśli wartość p 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 (jak również pozwala mu zorientować się w jakości i rygoryzmie Twych decyzji).
3. Ponieważ wartość χ^2 zależy od liczby stopni swobody, czyli liczby danych pomiarowych, często wygodniej jest do walidacji modelu użyć wartości zredukowanej χ^2, czyli wartości χ^2 przypadającej na stopień swobody:
$$\widehat{\chi^2} = \frac{1}{d}\chi^2, \tag{4}$$
gdzie, podobnie jak poprzednio, $d = N - k$ 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 przeszacowanie 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.

Warto podkreślić, że do zastosowania testu χ^2 Pearsona niezbędna jest wiedza na temat niepewności pomiarowych zmiennej zależnej. W przypadku braku takiej informacji metoda najmniejszych kwadratów zakłada niepewności równe średniej odchylenia standardowego reszt co prowadzi do wartości
$$\widehat{\chi^2} = \frac{1}{d}\sum_{i=1}^{N}\frac{\varepsilon_i^2}{s^2} = \frac{1}{d}\frac{1}{s^2}\sum_{i=1}^{N}\varepsilon_i^2 = \frac{1}{d}\frac{1}{\frac{1}{d}\sum_{i=1}^{N}\varepsilon_i^2}\sum_{i=1}^{N}\varepsilon_i^2 = 1$$
czyli wartości oczekiwanej, co uniemożliwia sprawdzenie modelu przy użyciu tego testu.

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.

Tabela 1. Wartości krytyczne rozkładu χ^2 dla zadanej liczby stopni swobody d i wybranych poziomów zgodności α.

Poziom zgodności α	Wartości krytyczne χ^2_{kr}, 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
0,001	10,83	13,82	16,27	18,47	20,51	22,46	24,32	26,12	27,88	29,59
0,003	9,00	11,83	14,16	16,25	18,21	20,06	21,85	23,57	25,26	26,90
0,005	7,88	10,60	12,84	14,86	16,75	18,55	20,28	21,95	23,59	25,19
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
0,100	2,71	4,61	6,25	7,78	9,24	10,64	12,02	13,36	14,68	15,99
	d = 11	d = 12	d = 13	d = 14	d = 15	d = 16	d = 17	d = 18	d = 19	d = 20
0,001	31,26	32,91	34,53	36,12	37,70	39,25	40,79	42,31	43,82	45,31
0,003	28,51	30,10	31,66	33,20	34,71	36,22	37,70	39,17	40,63	42,08
0,005	26,76	28,30	29,82	31,32	32,80	34,27	35,72	37,16	38,58	40,00
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
0,100	17,28	18,55	19,81	21,06	22,31	23,54	24,77	25,99	27,20	28,41
	d = 21	d = 22	d = 23	d = 24	d = 25	d = 26	d = 27	d = 28	d = 29	d = 30
0,001	46,80	48,27	49,73	51,18	52,62	54,05	55,48	56,89	58,30	59,70
0,003	43,52	44,94	46,36	47,76	49,16	50,55	51,93	53,31	54,68	56,04
0,005	41,40	42,80	44,18	45,56	46,93	48,29	49,65	50,99	52,34	53,67
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
0,100	29,62	30,81	32,01	33,20	34,38	35,56	36,74	37,92	39,09	40,26