

Testowanie hipotez statystycznych.

Krzysztof Topolski

Bioinformatyka

Wrocław, 3 kwietnia 2019

Temat.

Weryfikacja hipotez statystycznych dotyczących rozkładu.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Test zgodności χ^2 Pearsona.

Przypuśćmy, że mamy urnę w której znajduje się m kul. Każda kula ma jedną z k wyłączających się wzajemnie cech.

- m_1 ma cechę pierwszą,
- m_2 ma cechę drugą,
-,
- m_k ma cechę k -tą,

$$m_1 + m_2 + \dots + m_k = m.$$

Z urny losujemy ze zwracaniem n kul. Oznaczmy przez:

- N_1 liczbę kul, które mają cechę pierwszą,
- N_2 liczbę kul, które mają cechę drugą,
-,
- N_k liczbę kul, które mają cechę k -tą.

Wtedy wektor losowy $(N_1, \dots, N_k)$ ma rozkład wielomianowy

$$P\{N_1 = n_1, \dots, N_k = n_k\} = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k},$$

gdzie $p_i = \frac{m_i}{m}$, $i = 1, \dots, k$.

Statystyka

$$\chi_P^2 = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}$$

ma rozkład asymptotyczny χ_{k-1}^2 (chi kwadrat z $k - 1$ stopniami swobody).

Dowód dla przypadku $k = 2$

Z centralnego twierdzenia granicznego wiemy, że

$$Z = \frac{N_1 - np_1}{\sqrt{np_1(1 - p_1)}},$$

ma asymptotycznie standardowy rozkład normalny $N(0, 1)$.
Stąd

$$Z = \frac{(N_1 - np_1)^2}{np_1(1 - p_1)},$$

ma asymptotycznie rozkład chi kwadrat z jednym stopniem swobody χ_1^2 .

Dowód dla przypadku $k = 2$

Z centralnego twierdzenia granicznego wiemy, że

$$Z = \frac{N_1 - np_1}{\sqrt{np_1(1 - p_1)}},$$

ma asymptotycznie standardowy rozkład normalny $N(0, 1)$.
Stąd

$$Z = \frac{(N_1 - np_1)^2}{np_1(1 - p_1)},$$

ma asymptotycznie rozkład chi kwadrat z jednym stopniem swobody χ_1^2 .

Niech X będzie dyskretną zmienną losową, która przyjmuje k różnych wartości $x_1, x_2, \dots, x_k$. Niech

$$P\{X = x_i\} = p_i, \quad i = 1, \dots, k.$$

Niech $X_1, \dots, X_n$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie jak zmienna losowa X . Próba prosta z rozkładu dyskretnego o gęstości $\{p_i, i = 1, \dots, k\}$.

Niech X będzie dyskretną zmienną losową, która przyjmuje k różnych wartości $x_1, x_2, \dots, x_k$. Niech

$$P\{X = x_i\} = p_i, \quad i = 1, \dots, k.$$

Niech $X_1, \dots, X_n$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie jak zmienna losowa X . Próba prosta z rozkładu dyskretnego o gęstości $\{p_i, i = 1, \dots, k\}$.

Oznaczmy przez O_i liczbę tych zmiennych spośród $X_1, \dots, X_n$, które przyjęły wartość x_i a przez $E_i = np_i$ oczekiwaną liczbę wartości x_i . Wektor losowy $(O_1, O_2, \dots, O_k)$ ma rozkład wielomianowy z parametrami n i $(p_1, \dots, p_k)$ natomiast zmienna losowa

$$\chi_P^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i},$$

ma asymptotyczny rozkład chi kwadrat z $(k - 1)$ stopniami swobody χ_{k-1}^2 .

Test zgodności χ^2 dla hipotezy prostej

Rozważmy dyskretną zmienną losową X o nieznanym rozkładzie

$$P\{X = x_i\} = q_i, \quad i = 1, \dots, k,$$

q_i są nieznanne. Stawiamy hipotezę

$$H_0 : q_i = p_i \quad \text{dla każdego } i = 1, \dots, k.$$

Statystyka testowa to statystyka Pearsona χ_P^2

$$\chi_P^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i},$$

która w tym przypadku ma postać

$$\chi_P^2 = \sum_{i=1}^k \frac{(O_i - np_i)^2}{np_i}.$$

Test zgodności χ^2 dla hipotezy prostej

Rozważmy dyskretną zmienną losową X o nieznanym rozkładzie

$$P\{X = x_i\} = q_i, \quad i = 1, \dots, k,$$

q_i są nieznanne. Stawiamy hipotezę

$$H_0 : q_i = p_i \quad \text{dla każdego } i = 1, \dots, k.$$

Statystyka testowa to statystyka Pearsona χ_P^2

$$\chi_P^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i},$$

która w tym przypadku ma postać

$$\chi_P^2 = \sum_{i=1}^k \frac{(O_i - np_i)^2}{np_i}.$$

Przykład

Sprawdzamy, że cyfry z tabeli

i	0	1	2	3	4	5	6	7	8	9
O_i	33	32	29	31	37	34	34	26	32	32

są wygenerowane z rozkładu jednostajnego.

Przetestujemy hipotezę

$$H_0 : P\{X = i\} = \frac{1}{10}, \quad \text{dla każdego } i = 0, \dots, 9,$$

na poziomie istotności $\alpha = 0,01$.

Przykład

Sprawdzamy, że cyfry z tabeli

i	0	1	2	3	4	5	6	7	8	9
O_i	33	32	29	31	37	34	34	26	32	32

są wygenerowane z rozkładu jednostajnego.

Przetestujemy hipotezę

$$H_0 : P\{X = i\} = \frac{1}{10}, \quad \text{dla każdego } i = 0, \dots, 9,$$

na poziomie istotności $\alpha = 0,01$.

Przykład

Sprawdzamy, że cyfry z tabeli

i	0	1	2	3	4	5	6	7	8	9
O_i	33	32	29	31	37	34	34	26	32	32

są wygenerowane z rozkładu jednostajnego.

Przetestujemy hipotezę

$$H_0 : P\{X = i\} = \frac{1}{10}, \quad \text{dla każdego } i = 0, \dots, 9,$$

na poziomie istotności $\alpha = 0,01$.

W tabeli jest 320 cyfr a więc $n = 320$. Wszystkie częstości oczekiwane są takie same

$$E_i = \frac{n}{10} = \frac{320}{10} = 32.$$

Statystyka testowa ma wartość $\chi^2_P = 2,5$, natomiast kwantyl rzędu 0,99 dla rozkładu χ^2 z 9 stopniami swobody jest równy $\chi^2_{9;0,01} = 21,67$. Tak więc nie ma podstaw do odrzucenia hipotezy zerowej i przyjmujemy H_0 (p -wartość jest w tym przypadku równa $p = 0,981$).

Test zgodności χ^2 dla hipotezy prostej cd.

Test χ^2 możemy także stosować do sprawdzania hipotez o postaci rozkładu ciągłego. W tym celu należy wyznaczyć rozłączne klasy $A_1, \dots, A_k$ (podobnie jak przy konstrukcji histogramu) i obliczyć, przy założeniu hipotezy zerowej H_0 , prawdopodobieństwa

- $p_1 = P\{X \in A_1\},$
- $p_2 = P\{X \in A_2\},$
- $\dots,$
- $p_k = P\{X \in A_k\}.$

Niech n oznacza wielkość próby, a O_i liczbę obserwacji w i -tej klasie. Za pomocą testu χ^2 Pearsona możemy sprawdzić, czy zaobserwowane częstości $O_1, O_2, \dots, O_k$ podlegają rozkładowi z gęstością $\{p_i, i = 1, \dots, k\}$.

Test zgodności χ^2 dla hipotezy prostej cd.

Test χ^2 możemy także stosować do sprawdzania hipotez o postaci rozkładu ciągłego. W tym celu należy wyznaczyć rozłączne klasy $A_1, \dots, A_k$ (podobnie jak przy konstrukcji histogramu) i obliczyć, przy założeniu hipotezy zerowej H_0 , prawdopodobieństwa

- $p_1 = P\{X \in A_1\},$
- $p_2 = P\{X \in A_2\},$
- $\dots,$
- $p_k = P\{X \in A_k\}.$

Niech n oznacza wielkość próby, a O_i liczbę obserwacji w i -tej klasie. Za pomocą testu χ^2 Pearsona możemy sprawdzić, czy zaobserwowane częstości $O_1, O_2, \dots, O_k$ podlegają rozkładowi z gęstością $\{p_i, i = 1, \dots, k\}$.

Test zgodności χ^2 dla hipotezy prostej cd.

Test χ^2 możemy także stosować do sprawdzania hipotez o postaci rozkładu ciągłego. W tym celu należy wyznaczyć rozłączne klasy $A_1, \dots, A_k$ (podobnie jak przy konstrukcji histogramu) i obliczyć, przy założeniu hipotezy zerowej H_0 , prawdopodobieństwa

- $p_1 = P\{X \in A_1\},$
- $p_2 = P\{X \in A_2\},$
- $\dots,$
- $p_k = P\{X \in A_k\}.$

Niech n oznacza wielkość próby, a O_i liczbę obserwacji w i -tej klasie. Za pomocą testu χ^2 Pearsona możemy sprawdzić, czy zaobserwowane częstości $O_1, O_2, \dots, O_k$ podlegają rozkładowi z gęstością $\{p_i, i = 1, \dots, k\}$.

Test zgodności χ^2 dla hipotezy prostej cd.

Test χ^2 możemy także stosować do sprawdzania hipotez o postaci rozkładu ciągłego. W tym celu należy wyznaczyć rozłączne klasy $A_1, \dots, A_k$ (podobnie jak przy konstrukcji histogramu) i obliczyć, przy założeniu hipotezy zerowej H_0 , prawdopodobieństwa

- $p_1 = P\{X \in A_1\},$
- $p_2 = P\{X \in A_2\},$
- $\dots,$
- $p_k = P\{X \in A_k\}.$

Niech n oznacza wielkość próby, a O_i liczbę obserwacji w i -tej klasie. Za pomocą testu χ^2 Pearsona możemy sprawdzić, czy zaobserwowane częstości $O_1, O_2, \dots, O_k$ podlegają rozkładowi z gęstością $\{p_i, i = 1, \dots, k\}$.

Test zgodności χ^2 dla hipotezy prostej cd.

Test χ^2 możemy także stosować do sprawdzania hipotez o postaci rozkładu ciągłego. W tym celu należy wyznaczyć rozłączne klasy $A_1, \dots, A_k$ (podobnie jak przy konstrukcji histogramu) i obliczyć, przy założeniu hipotezy zerowej H_0 , prawdopodobieństwa

- $p_1 = P\{X \in A_1\},$
- $p_2 = P\{X \in A_2\},$
- $\dots,$
- $p_k = P\{X \in A_k\}.$

Niech n oznacza wielkość próby, a O_i liczbę obserwacji w i -tej klasie. Za pomocą testu χ^2 Pearsona możemy sprawdzić, czy zaobserwowane częstości $O_1, O_2, \dots, O_k$ podlegają rozkładowi z gęstością $\{p_i, i = 1, \dots, k\}$.

Test zgodności χ^2 dla hipotezy prostej cd.

Test χ^2 możemy także stosować do sprawdzania hipotez o postaci rozkładu ciągłego. W tym celu należy wyznaczyć rozłączne klasy $A_1, \dots, A_k$ (podobnie jak przy konstrukcji histogramu) i obliczyć, przy założeniu hipotezy zerowej H_0 , prawdopodobieństwa

- $p_1 = P\{X \in A_1\}$,
- $p_2 = P\{X \in A_2\}$,
- $\dots$,
- $p_k = P\{X \in A_k\}$.

Niech n oznacza wielkość próby, a O_i liczbę obserwacji w i -tej klasie. Za pomocą testu χ^2 Pearsona możemy sprawdzić, czy zaobserwowane częstości $O_1, O_2, \dots, O_k$ podlegają rozkładowi z gęstością $\{p_i, i = 1, \dots, k\}$.

Metodę tę zilustrujemy następującym przykładem.

Przykład

Według producenta pojemność jego naczynie miareczkowego ma rozkład normalny z wartością średnią równą nominalnej pojemności oraz odchyleniu standardowym równym $0,01 \text{ mm}^3$. Aby to sprawdzić, przebadano pojemność 200 naczyń o nominalnej pojemności równej 30 mm^3 . Wyznaczono $k = 10$ rozłącznych klas $A_1, \dots, A_{10}$. Ich granice z zaobserwowanymi liczebnościami są podane w tabeli

Metodę tę zilustrujemy następującym przykładem.

Przykład

Według producenta pojemność jego naczynie miareczkowego ma rozkład normalny z wartością średnią równą nominalnej pojemności oraz odchyleniu standardowym równym $0,01 \text{ mm}^3$. Aby to sprawdzić, przebadano pojemność 200 naczyń o nominalnej pojemności równej 30 mm^3 . Wyznaczono $k = 10$ rozłącznych klas $A_1, \dots, A_{10}$. Ich granice z zaobserwowanymi liczebnościami są podane w tabeli

i	Klasa	O_i	E_i
1	$(-\infty, 29,980]$	2	4,55
2	$(29,980, 29,985]$	13	8,81
3	$(29,985, 29,990]$	20	18,37
4	$(29,990, 29,995]$	26	29,98
5	$(29,995, 30,000]$	40	38,29
6	$(30,000, 30,005]$	42	38,29
7	$(30,005, 30,010]$	26	29,98
8	$(30,010, 30,015]$	21	18,37
9	$(30,015, 30,020]$	10	8,81
10	$(30,020, +\infty)$	0	4,55

Niech X oznacza pomiar pojemności oraz niech $q_i = P\{X \in A_i\}$. W zapewnienia producenta oznaczają, że X ma rozkład normalny $N(30, 00, 0, 0001)$. My sprawdzamy hipotezę $H_0 : q_i = p_i$, w której

$$p_i = \Phi\left(\frac{a_i - 30}{0,01}\right) - \Phi\left(\frac{a_{i-1} - 30}{0,01}\right),$$

gdzie $A_i = (a_{i-1}, a_i]$ jest i -tą klasą.

Na podstawie danych z tablicy, w których łączymy w jedną dwie dolne i dwie górne klasy otrzymujemy następującą wartość statystyki testowej $\chi_P^2 = 3,06$. Ponieważ przy 7 stopniach swobody p -wartość wynosi 0,879, nie ma podstaw do odrzucenia hipotezy zerowej H_0 , a w takim razie także hipotezy o rozkładzie normalnym $N(0, 1)$ pojemności naczyń miareczkowych.

Niech X oznacza pomiar pojemności oraz niech $q_i = P\{X \in A_i\}$. W zapewnienia producenta oznaczają, że X ma rozkład normalny $N(30, 00, 0, 0001)$. My sprawdzamy hipotezę $H_0 : q_i = p_i$, w której

$$p_i = \Phi\left(\frac{a_i - 30}{0,01}\right) - \Phi\left(\frac{a_{i-1} - 30}{0,01}\right),$$

gdzie $A_i = (a_{i-1}, a_i]$ jest i -tą klasą.

Na podstawie danych z tablicy, w których łączymy w jedną dwie dolne i dwie górne klasy otrzymujemy następującą wartość statystyki testowej $\chi^2_P = 3,06$. Ponieważ przy 7 stopniach swobody p -wartość wynosi 0,879, nie ma podstaw do odrzucenia hipotezy zerowej H_0 , a w takim razie także hipotezy o rozkładzie normalnym $N(0, 1)$ pojemności naczyń miareczkowych.

Uwaga

Przybliżenie rozkładem χ^2 można uznać za dopuszczalne, gdy $np_i \geq 5$ dla każdego $i = 1, \dots, k$.

Założmy, że interesuje nas jednorodność k rozkładów w sytuacji gdy obserwacje każdej z populacji mogą należeć do l kategorii. Oznaczmy przez $(p_1^{(i)}, \dots, p_l^{(i)})$ nieznany rozkład i -tej populacji, $i = 1, \dots, k$. Chcemy przetestować hipotezę

$$H_0 : p_i^{(1)} = p_i^{(2)} = \dots = p_i^{(k)} \quad \text{dla każdego } i = 1, 2, \dots, l.$$

Procedurę testowania opieramy na n_{ij} liczbach obserwacji z i -tej populacji należących do j -tej klasy, gdzie

$$\sum_{j=1}^I n_{ij} = n_{i..}$$

Przy zachodzeniu hipotezy zerowej wszystkie populacje mają ten sam rozkład $(p_1, p_2, \dots, p_I)$,

$$p_j^{(1)} = p_j^{(2)} = \dots = p_j^{(k)} = p_j, \quad \text{dla } j = 1, 2, \dots, I.$$

Estymatorem prawdopodobieństwa p_j jest ułamek

$$\frac{n_{.j}}{n},$$

gdzie

$$n_{.j} = \sum_{i=1}^k n_{ij} \quad \text{oraz} \quad n = \sum_{j=1}^I \sum_{i=1}^k n_{ij}.$$

Procedurę testowania opieramy na n_{ij} liczbach obserwacji z i -tej populacji należących do j -tej klasy, gdzie

$$\sum_{j=1}^l n_{ij} = n_{i..}$$

Przy zachodzeniu hipotezy zerowej wszystkie populacje mają ten sam rozkład $(p_1, p_2, \dots, p_l)$,

$$p_j^{(1)} = p_j^{(2)} = \dots = p_j^{(k)} = p_j, \quad \text{dla } j = 1, 2, \dots, l.$$

Estymatorem prawdopodobieństwa p_j jest ułamek

$$\frac{n_{.j}}{n},$$

gdzie

$$n_{.j} = \sum_{i=1}^k n_{ij} \quad \text{oraz} \quad n = \sum_{j=1}^l \sum_{i=1}^k n_{ij}.$$

Statystyką testową jest statystyka

$$\chi_P^2 = \sum_{j=1}^l \sum_{i=1}^k \frac{(n_{ij} - \frac{n_{.j}}{n})^2}{\frac{n_{.j}}{n}}.$$

Statystyka χ_P^2 ma asymptotyczny rozkład chi kwadrat z $(k-1)(l-1)$ stopniami swobody.