WERYFIKACJA HIPOTEZ STATYSTYCZNYCH

Przeprowadzenie badań niewyczerpujących, (częściowych – prowadzonych na podstawie próby losowej), nie daje podstaw do formułowania stanowczych stwierdzeń dotyczących badanych zmiennych losowych. Daje jednak wystarczające podstawy do stawiania oraz weryfikowania różnego rodzaju przypuszczeń dotyczących zarówno klasy rozkładu, jak i jego parametrów. Tego rodzaju przypuszczenia nazywamy **hipotezami statystycznymi**.

WERYFIKACJA HIPOTEZ STATYSTYCZNYCH

Hipotezy, które dotyczą parametrów nazywać będziemy hipotezami **parametrycznymi**, zaś takie, które dotyczą klasy rozkładu zmiennej – a więc funkcyjnej postaci dystrybuanty nazywać będziemy hipotezami **nieparametrycznymi**. Procedury mające na celu sprawdzenie tak sformułowanych hipotez nazywać będziemy **testami statystycznymi** i odpowiednio do sprawdzanych hipotez wyróżniamy testy parametryczne oraz nieparametryczne.

WERYFIKACJA HIPOTEZ STATYSTYCZNYCH

Procedura przeprowadzenia testu sprowadza się, najprościej mówiąc, do postawienia hipotezy statystycznej, wyboru statystyki testowej (kryterium danego testu) oraz poziomu istotności testu.

Najczęściej poziom istotności testu jest oznaczany jako alfa i oznacza prawdopodobieństwo popełnienia błędu pierwszego rodzaju (który polega na odrzuceniu sądu właściwego, przy czym prawdopodobieństwo przyjęcia sądu błędnego oznaczane jest jako beta – czyli błąd drugiego rodzaju).

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Test niezależności chi-kwadrat stosowany jest w przypadku badania niezależności cech niemierzalnych (jakościowych) lub w przypadku badania niezależności cechy jakościowej z ilościową.

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Załóżmy, że przedmiotem badania jest populacja generalna. Z populacji tej pobrano n–elementową próbę (przy czym ważne jest, by $n>30$), której wyniki sklasyfikowano w postaci tablicy wg jednej cechy w r wierszach i wg drugiej cechy w k kolumnach. Wnętrze tablicy niezależności stanowią liczebności n_{ij} elementów próby, które spełniają jednocześnie kryteria zawarte w i–tym wierszu i k–tej kolumnie.

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Tablica niezależności jest podstawą weryfikacji nieparametrycznej hipotezy zerowej głoszącej, że w populacji nie ma zależności między cechami (zmiennymi) X i Y.

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Hipotezę tę można zapisać zgodnie z pojęciem niezależności zmiennych losowych w sposób następujący:

H0; P (X = xi, Y = yj) = P (X = xi)P(Y = yj), czyli, że cechy X i Y są niezależne oraz

H1; P (X = xi, Y = yj) ≠ P (X = xi)P(Y = yj), czyli, że cechy X i Y są zależne,

przy przyjętym poziomie istotności α.

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Do weryfikacji powyższych hipotez stosuje się statystykę χ^2, której wartość liczymy ze wzoru:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{k} \frac{\left(n_{ij} - \hat{n}_{ij}\right)^2}{\hat{n}_{ij}}$$

gdzie:

- n_{ij} – liczba elementów próby,
- $\hat{n}_{ij}$ – liczebności teoretyczne,
- k – liczba kolumn tablicy niezależności,
- r – liczba wierszy tablicy niezależności.

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Liczebności teoretyczne oblicza się wg wzoru:

$$\hat{n}_{ij} = \frac{\sum_{i=1}^{r} n_{ij} \cdot \sum_{j=1}^{k} n_{ij}}{n}$$

- gdzie:
- n_{ij} – liczba elementów próby,
- $\hat{n}_{ij}$ – liczebności teoretyczne,
- k – liczba kolumn tablicy niezależności,
- r – liczba wierszy tablicy niezależności.

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Z tablic rozkładu chi-kwadrat odczytujemy wartość statystyki χ^2 odczytaną przy poziomie istotności α i przy (r-1)(k-1) stopniach swobody,

obszar krytyczny (prawostronny) wyznacza nierówność:

$$\chi^2 \geq \chi^2_{\alpha;(r-1)(k-1)}$$

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Jeżeli $\chi^2 \geq \chi^2_{\alpha;\,(r-1)(k-1)}$ $\rightarrow$ H_o odrzucamy na rzecz hipotezy alternatywnej

Jeżeli $\chi^2 < \chi^2_{\alpha;\,(r-1)(k-1)}$ $\rightarrow$ Nie ma podstaw do odrzucenia H_o o niezależności cech

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Wartości chi-kwadrat zależą od:

- **natężenia (siły) związku** badanych cech – im większe różnice między liczebnością empiryczną a teoretyczną, tym większa wartość chi-kwadrat i tym samym większa zależność między cechami,
- od **wielkości próby**, przy czym chi-kwadrat liczymy tylko dla dużych prób,
- od **stopnia szczegółowości danych** (przy czym w każdym polu tabeli powinno być co najmniej 5, a więc czasami trzeba łączyć wiersze lub kolumny).

TEST NIEZALEŻNOŚCI CHI–KWADRAT (TEST χ^2)

Przykład 1

Do badania wybrano 500 mieszkańców Rzeszowa, których poproszono o określenie, jakiego typu programy rozrywkowe oglądają w TV – kabarety czy relacje z festiwali. Wyniki odpowiedzi respondentów zostały przedstawione w tabeli niezależności. **Sprawdź, czy rodzaj oglądanych programów rozrywkowych i płeć respondenta są niezależne, przyjmując poziom istotności alfa= 0,05.**

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Płeć	Oglądane programy		RAZEM
	Kabarety	Festiwale	
Mężczyzna	30	80	110
Kobieta	170	220	390
RAZEM	200	300	500

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Rozpoczynamy od postawienia hipotezy zerowej i alternatywnej:

H0; $P(X = xi, Y = yj) = P(X = xi)P(Y = yj)$, czyli, że cechy X i Y są niezależne oraz

H1; $P(X = xi, Y = yj) \neq P(X = xi)P(Y = yj)$, czyli, że cechy X i Y są zależne,

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Następnie należy obliczyć liczebności teoretyczne:

$$\hat{n}_{11} = \frac{110 \cdot 200}{500} = 44$$

$$\hat{n}_{12} = \frac{110 \cdot 300}{500} = 66$$

$$\hat{n}_{21} = \frac{390 \cdot 200}{500} = 156$$

$$\hat{n}_{22} = \frac{390 \cdot 300}{500} = 234$$

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Do weryfikacji powyższych hipotez stosuje się statystykę χ^2, której wartość wynosi:

$$\chi^2 = \frac{(30-44)^2}{44} + \frac{(80-66)^2}{66} + \frac{(170-156)^2}{156} + \frac{(220-234)^2}{234} = 9{,}52$$

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Z tablic rozkładu chi-kwadrat odczytujemy wartość statystyki χ^2 dla poziomu istotności $\alpha=0,05$ i przy (2-1)(2-1) stopniach swobody, która wynosi 3,841. Ponieważ zachodzi nierówność:

$$\chi^2 > \chi^2_{0,05;1}$$

istnieją podstawy do odrzucenia hipotezy zerowej na rzecz hipotezy alternatywnej, czyli istnieje zależność między analizowanymi cechami.

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Przykład 2

Dla zbadania przypuszczenia, iż wyjazdy na letni wypoczynek zależą od wielkości miejscowości wylosowano w sposób niezależny 200 osób. Wyniki odpowiedzi respondentów zostały przedstawione w tabeli niezależności. Na poziomie istotności 0,05 zweryfikuj hipotezę, że wyjazdy na letni wypoczynek zależą od miejsca zamieszkania.

TEST NIEZALEŻNOŚCI CHI–KWADRAT (TEST χ^2)

Miejsce zamieszkania	Wyjazd na letni wypoczynek		Razem
	tak	nie	
do 30 tys. mieszkańców	14	36	**50**
30 tys. - 100 tys. mieszkańców	35	30	**65**
Powyżej 100 tys. mieszkańców	47	38	**85**
Razem	**96**	**104**	**200**

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Rozpoczynamy od postawienia hipotezy zerowej i alternatywnej:

H0; P (X = xi, Y = yj) = P (X = xi)P(Y = yj), czyli, że cechy X i Y są niezależne oraz

H1; P (X = xi, Y = yj) ≠ P (X = xi)P(Y = yj), czyli, że cechy X i Y są zależne,

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Następnie należy obliczyć liczebności teoretyczne:

$$\hat{n}_{11} = \frac{50 \cdot 96}{200} = 24$$

$$\hat{n}_{12} = \frac{50 \cdot 104}{200} = 26$$

$$\hat{n}_{21} = \frac{65 \cdot 96}{200} = 31,2$$

$$\hat{n}_{22} = \frac{65 \cdot 104}{200} = 33,8$$

$$\hat{n}_{31} = \frac{85 \cdot 96}{200} = 40,8$$

$$\hat{n}_{32} = \frac{85 \cdot 104}{200} = 44,2$$

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Do weryfikacji powyższych hipotez stosuje się statystykę χ^2, której wartość wynosi:

$$\chi^2 = \frac{(14-24)^2}{24} + \frac{(36-26)^2}{26} + \frac{(35-31{,}2)^2}{31{,}2} +$$

$$+ \frac{(30-33{,}8)^2}{33{,}8} + \frac{(47-40{,}8)^2}{40{,}8} + \frac{(38-44{,}2)^2}{44{,}2} = 10{,}72$$

TEST NIEZALEŻNOŚCI CHI-KWADRAT (TEST χ^2)

Z tablic rozkładu chi-kwadrat odczytujemy wartość statystyki χ^2 dla poziomu istotności $\alpha=0,05$ i przy (3-1)(2-1) stopniach swobody, która wynosi 5,991. Ponieważ zachodzi nierówność:

$$\chi^2 > \chi^2_{0,05;2}$$

istnieją podstawy do odrzucenia hipotezy zerowej na rzecz hipotezy alternatywnej, czyli istnieje zależność między analizowanymi cechami.