Statystyka matematyczna – parametryczne testy istotności (weryfikacja hipotez dotyczących parametrów rozkładu).

Zadanie

Maszyna mieszająca nawóz jest tak nastawiona, aby w każdych 100 kg nawozu było 10 kg azotanu. Zbadano dziesięć 100-kilogramowych worków. Procentowa zawartość azotanu była następująca: 9, 12, 11, 10, 11, 9, 11, 12, 9, 10. Czy na poziomie istotności $\alpha=0.05$ można uważać za słuszną hipotezę, że wartość przeciętna zawartości azotanu w worku jest równa 10 %, jeżeli hipotezą alternatywną jest hipoteza, że ta wartość przeciętna jest wyższa niż 10 % ?

Rozwiązanie

Zadanie to jest klasycznym przykładem tego, czym zajmuje się teoria testów. Naszym zadaniem jest, na podstawie badanej próbki, orzec o właściwości całej populacji; tutaj konkretnie jest to nieznana wartość oczekiwana.
Aby móc w ogóle rozwiązać to zadanie, należy ustalić tzw. model zagadnienia. W tym przypadku jest to tzw. model 2 weryfikacji hipotezy o nieznanej wartości oczekiwanej (rozkład cechy normalny, o nieznanych parametrach, liczność próby niewielka) – z tym, że by być uprawnionym do wykorzystania w tych warunkach modelu 2 należy założyć – gdyż nie jest to wprost powiedziane w zadaniu, że badana cecha – w tym przypadku zawartość azotanu – ma rozkład normalny o nieznanych parametrach.
Test statystyczny w każdym przypadku wymaga:

- postawienia hipotezy zerowej oraz przyjęcia hipotezy alternatywnej, przyjmowanej w wypadku odrzucenia hipotezy zerowej,
- wyliczenia, na podstawie próby, tzw. statystyki testowej
- wyznaczenia zbioru krytycznego, zależnego od założonego poziomu istotności, ustalonego modelu, oraz postaci przyjętej hipotezy alternatywnej
- dokonania właściwego testu statystycznego, polegającego na sprawdzeniu, czy wyliczona wartość statystyki testowej zawiera się w zbiorze krytycznym, czy też nie.

Postawienie hipotez

W naszym zadaniu postać hipotez mamy narzuconą z góry.
Hipotezą zerową (która zawsze zakładać musi równość nieznanej wartości cechy i pewnej wartości hipotetycznej) będzie hipoteza:

$$\mathbf{H_0} : \mu = \mu_0 = 10\ \%$$

Hipotezą alternatywną (która może zakładać, że nieznany parametr jest większy, mniejszy bądź różny od wartości hipotetycznej) zaś jest hipoteza:

$$\mathbf{H_k} : \mu > \mu_0$$

Wyznaczanie statystyki testowej.

Wobec naszego modelu obliczeń, wartość statystyki testowej wyraża się wzorem:

$$t_{obl} = \frac{\bar{x} - \mu_0}{S} \cdot \sqrt{n-1}$$

Zatem zacząć musimy od wyliczenia średniej arytmetycznej, oraz odchylenia standardowego z naszej próby:

Średnia arytmetyczna:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

$$\bar{x} = 10{,}4$$

Odchylenie standardowe:

$$s = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} x^2 - (\bar{x})^2}$$

$$s = 1{,}11$$

Wartość statystyki testowej:

$$t_{obl} = \frac{10{,}4 - 10}{1{,}11} \cdot \sqrt{9}$$

$$t_{obl} = 1{,}08$$

Zbiór krytyczny

W rozważanym modelu, wobec przyjętej hipotezy alternatywnej zbiór krytyczny wyraża się wzorem:

$$W = \langle t_{1-\alpha, n-1}, +\infty)$$

$t_{1-\alpha,\ n-1}$ – jest kwantylem rozkładu Studenta o n – 1 stopniach swobody , którego wartość odczytujemy z tablic. Parametr α stanowi - narzucony w treści zadania tzw. poziom istotności. Jest to prawdopodobieństwo popełnienia tzw. błędu I rodzaju, który polega na odrzuceniu hipotezy zerowej, pomimo, iż w rzeczywistości jest ona prawdziwa. Błędy w statystyce matematycznej biorą się stąd, że nie sposób bezbłędnie orzec co do właściwości populacji, na podstawie próbki, która zawsze jest tylko podzbiorem tejże populacji. Nie jest więc możliwe wyeliminowanie błędów. Aparat obliczeniowy umożliwia tylko przejęcie kontroli nad błędem pierwszego rodzaju.
Nasz kwantyl jest równy:

$$t_{0{,}95\,;\,9} = 1{,}83$$

Wobec czego zbiór krytyczny:

$$W = \langle 1{,}83; +\infty)$$

Ponieważ obliczona wartość statystyki testowej wynosi $t_{obl} = 1{,}08$, zatem prawdą jest, że:

$$t_{obl} \notin W$$

Wobec czego orzekamy, iż *brak jest podstaw do odrzucenia hipotezy zerowej*. Pozostajemy zatem w przekonaniu, iż wartość oczekiwana zawartości nawozu w workach wynosi istotnie 10 %. Uwaga, nie byłoby poprawne stwierdzenie, że przyjmujemy hipotezę zerową, gdyż w tej sytuacji jest możliwe, że popełnilibyśmy tzw. błąd II rodzaju, tj. przyjęliśmy hipotezę zerową, podczas, gdy w rzeczywistości jest ona fałszywa. Prawdopodobieństwo tego jest na ogół duże i, co gorsza, zastosowany aparat obliczeniowy nie zapewnia nam nad nim żadnej kontroli. Dlatego hipotezy zerowej nie przyjmujemy, a jedynie stwierdzamy, że brak jest podstaw do jej odrzucenia (co w praktyce wychodzi na to samo).

Przykład pochodzi z podręcznika *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, cz. II – statystyka matematyczna; W. Krysicki, J. Bartos, W. Dyczka, K. Królikowska, M. Wasilewski, Wydawnictwo Naukowe PWN, Warszawa 1995*

Koncepcja rozwiązania oraz objaśnienia: mgr inż. Sebastian Dziarmaga-Działyński