Wykład 5: Testy statystyczne

Definicja. *Hipotezą statystyczną* nazywamy przypuszczenie dotyczące nieznanego rozkładu badanej cechy populacji, o prawdziwości lub fałszywości którego wnioskuje się na podstawie pobranej próby.

Przykład 5.1.

1. *Wysuwamy hipotezę, że badana cecha ma rozkład normalny.*
2. *Wiemy, że badana cecha ma rozkład normalny o nieznanej wartości średniej μ i znanym odchyleniu standardowym $\sigma = 1$. Wysuwamy hipotezę, że $\mu = 5$.*
3. *Dane są dwa zbiory obserwacji, np. wysokości plonów uzyskane podczas naważenia nawozem A i nawozem B.*
   (a) *Wysuwamy hipotezę, że oba zbiory można traktować jako pochodzące z populacji o tym samym rozkładzie.*
   (b) *Z wcześniejszych badań wiemy, że zbiory te można traktować jako pochodzące z populacji o rozkładach normalnych odpowiednio $\mathcal{N}(\mu_1, \sigma_1^2)$ i $\mathcal{N}(\mu_2, \sigma_2^2)$, gdzie parametry μ_1, μ_2, σ_1 i σ_2 są nieznane, ale takie, że $\sigma_1 = \sigma_2$. Wysuwamy przypuszczenie, że średnia wartość plonów przy nawożeniu nawozem A jest większa niż średnia wartość plonów przy nawożeniu nawozem B (tzn. że $\mu_1 > \mu_2$).*

Definicja. Hipotezy, które dotyczą wyłącznie wartości parametru lub parametrów rozkładu badanej cechy nazywamy *hipotezami parametrycznymi*. Hipotezy, które nie są hipotezami parametrycznymi nazywamy *hipotezami nieparametrycznymi*.

W przykładzie 5.1 hipotezy 2 i 3(b) są parametryczne, natomiast hipotezy 1 i 3(a) są nieparametryczne.

Definicja. *Hipotezą prostą* nazywamy hipotezę, która jednoznacznie określa rozkład badanej cechy. *Hipotezą złożoną* nazywamy hipotezę, która określa całą grupę rozkładów.

Hipoteza 2 z przykładu 5.1 jest hipotezą prostą. Pozostałe hipotezy w tym przykładzie są złożone.

W praktyce rozważamy dwie hipotezy: *hipotezę zerową* (będziemy ją oznaczać H_0) i *hipotezę alternatywną* (tą będziemy oznaczać H_1). Jeśli odrzucamy hipotezę zerową, to przyjmujemy hipotezę alternatywną i na odwrót.

Przykład 5.2.
Wiemy, że wysokość plonów przy nawożeniu starą metodą ma rozkład normalny o wartości średniej 5 i wariancji 1: $\mathcal{N}(5, 1)$ i że wysokość plonów przy nawożeniu nową metodą ma rozkład normalny o tej samej wariancji $\sigma^2 = 1$ ale o nieznanej wartości średniej μ: $\mathcal{N}(\mu, 1)$. Chcemy sprawdzić czy nowa metoda zwiększyła średnią wysokość plonów.

W tym celu będziemy testować $H_0 : \mu = 5$ przeciwko $H_1 : \mu > 5$.

Przeprowadzamy eksperyment losowy. Wynik takiego eksperymentu to próba czyli wektor losowy $\mathbb{X} = (X_1, X_2, \ldots, X_n)$ o wartościach w $\mathbb{R}^n$.

Definicja. *Statystyką testową* nazywamy funkcję próby $\delta(X_1, X_2, \ldots, X_n)$, która służy do weryfikacji H_0 przeciwko H_1. Zbiór wszystkich możliwych wartości funkcji δ dzielimy na dwa rozłączne zbiory W i W' takie, że:

- jeśli $\delta(x_1, x_2, \ldots, x_n) \in W$, to H_0 odrzucamy,
- jeśli $\delta(x_1, x_2, \ldots, x_n) \in W'$, to H_0 przyjmujemy.

W nazywamy *zbiorem krytycznym testu* (zbiorem odrzuceń H_0).

Musimy dobrze skonstruować statystykę testową i rozsądnie dobrać zbiór krytyczny, tak by podejmować decyzje zgodne z rzeczywistością. Jednak, ponieważ decyzje podejmujemy jedynie na podstawie próby, nie mamy całkowitej informacji o badanej populacji i w związku z tym zawsze jesteśmy narażeni na popełnienie błędu - podjęcie decyzji niezgodnej z rzeczywistością. Dokładniej, możemy popełnić jeden z dwóch błędów:

1. odrzucić H_0 w sytuacji, gdy jest ona prawdziwa (tzw. *błąd pierwszego rodzaju*);
2. przyjąć H_0 w sytuacji, gdy jest ona fałszywa (tzw. *błąd drugiego rodzaju*).

Chcielibyśmy by prawdopodobieństwa obu tych błędów były jak najmniejsze. Niestety, gdy przy ustalonej statystyce testowej, zmieniamy W tak by malał błąd pierwszego rodzaju, to błąd drugiego rodzaju rośnie i na odwrót. Postępujemy zatem tak:

- ustalamy z góry maksymalną wartość prawdopodobieństwa błędu pierwszego rodzaju, oznaczamy tą wartość α i nazywamy ją *poziomem istotności testu* (zwyczajowo przyjmuje się $\alpha = 0,01$ lub $\alpha = 0,05$, czasami $\alpha = 0,1$);
- statystykę testową i zbiór krytyczny W dobieramy tak by prawdopodobieństwo błędu pierwszego rodzaju nie przekraczało α i by prawdopodobieństwo błędu drugiego rodzaju było możliwie najmniejsze.

Zapisujemy to symbolicznie:

$$\underbrace{P(\text{odrzucimy } H_0|H_0)}_{\substack{\text{prawdop. odrzucenia } H_0 \text{ w sytuacji, gdy } H_0 \text{ jest prawdziwa} \\ \text{czyli prawdop. błędu pierwszego rodzaju}}} \quad \leqslant \quad \alpha \qquad \text{i}$$

$$\underbrace{P(\text{przyjmiemy } H_0|H_1)}_{\substack{\text{prawdop. przyjęcia } H_0 \text{ w sytuacji, gdy } H_1 \text{ jest prawdziwa} \\ \text{czyli prawdop. błędu drugiego rodzaju}}} \quad - \quad \text{możliwie najmniejsze}$$

Definicja. *Moc testu* parametrycznego (power of the test) to funkcja zmiennej θ (gdzie θ to badany parametr) dana wzorem

$$\begin{aligned} \beta(\theta) &= P(\text{odrzucimy } H_0|\theta) = \\ &= \text{ prawdop. odrzucenia } H_0 \text{ w sytuacji, gdy nieznany parametr przyjmuje wartość } \theta. \end{aligned}$$

W szczególności, jeśli $H_0 : \theta = \theta_0$, to

$$\begin{aligned} \beta(\theta_0) &= P(\text{odrzucimy } H_0|\theta_0) = P(\text{odrzucimy } H_0|H_0) = \\ &= \text{ prawdopodobieństwo błędu pierwszego rodzaju} \end{aligned}$$

i dla $\theta_1 \neq \theta_0$

$$\beta(\theta_1) = P(\text{odrzucimy } H_0|\theta_1) = P(\text{podejmiemy słuszną decyzję}).$$

Definicja. Najmniejszy poziom istotności, przy którym zaobserwowana wartość statystyki testowej prowadzi do odrzucenia H_0, nazywamy *p-wartością* (p-value) przeprowadzonego testu. Tzn.

$$\begin{aligned} p-\text{value} \leqslant \alpha \quad &\Rightarrow \quad \text{odrzucamy } H_0, \\ p-\text{value} > \alpha \quad &\Rightarrow \quad \text{przyjmujemy } H_0. \end{aligned}$$

UWAGA! Do wyników testowania statystycznego powinniśmy podchodzić z pewną rezerwą, zwłaszcza, gdy hipoteza alternatywna H_1 jest złożona. Jeśli statystyka testowa nie wpadnie do zbioru krytycznego, to stwierdzamy, że nie ma podstaw do odrzucenia H_0, co jescze nie oznacza, że H_0 należy uznać za prawdziwą.