Statystyka w analizie i planowaniu eksperymentu

Wykład 4

Podstawy testowania, test t-Studenta

Przemysław Biecek

Dla 1 roku studentów Biotechnologii

Przykład wprowadzający

Pytanie

Czy średnia liczba punktów ze statystyki dla losowo wybranego studenta jest równa 6?

Liczba punktów ze statystyki dla 10 losowo wybranych studentów to odpowiednio:

`5, 10, 6, 7, 6, 8, 8, 9, 7, 5.`

Średnia z tych liczb to 7.1, jak odpowiedzieć na pytanie?

Model eksperymentu (mechanizmu) losowego

- Obserwujemy niezależne realizacje pewnej zmiennej losowej o nieznanym nam rozkładzie.
- **Zakładamy**, że nieznany rozkład należy do pewnej rodziny rozkładów indeksowanych parametrem θ_0 należącym do zbioru Θ.
- Formułujemy pytanie (hipotezę) w terminach podzbioru Θ_0 zbioru Θ.

Hipotezy formułujemy parami. Pytanie dotyczące podzbioru Θ_0 nazywamy hipotezą zerową. Jednocześnie formułujemy hipotezę alternatywną, przeciwstawną (będącą dopełnieniem) do hipotezy zerowej.

Hipoteza zerowa

Hipotezą zerową nazywamy przypuszczenie, że $\theta \in \Theta_0$.

Hipoteza alternatywna

Hipotezą alternatywną nazywamy przypuszczenie, że $\theta \notin \Theta_0$.

Przykład kontynuacja

Interesuje nas liczba punktów ze statystyki zdobytych przez studentów. Zakładamy (np. na bazie obserwacji z poprzednich lat), że liczba punktów zdobytych przez studenta opisać można zmienną losową o rozkładzie normalnym z wariancją 4.
W tej bajce:

- Eksperyment losowy: napisanie kolokwiów przez studenta, podliczenie punktów,
- Za rodzinę rozkładów przyjmujemy rodzinę rozkładów normalnych $\mathcal{N}(\mu, 2^2)$, gdzie $\mu \in \mathcal{R}^+$,
- Za zbiór Θ_0 przyjmujemy jednoelementowy zbiór $\{6\}$.

Hipotezę zerową i alternatywną opisujemy

$$H_0: \quad \mu = 6,$$

$$H_A: \quad \mu \neq 6.$$

Test statystyczny

Testem statystycznym nazywamy regułę, która określa, dla jakich wyników eksperymentu $x \in \mathcal{X}$ należy podjąć decyzję o przyjęciu H_0, a dla jakich o odrzuceniu H_0 i przyjęciu H_A.

Reguła ta najczęściej przedstawiana jest w terminach statystyki testowej, czyli pewnej funkcji próby.

Test statystyczny dzieli przestrzeń $\mathcal{X}$ na dwa rozłączne podzbiory: obszar odrzucenia hipotezy zerowej (obszar krytyczny) $B \subset \mathcal{X}$ oraz jego dopełnienie, czyli obszar przyjęcia hipotezy zerowej B^C. Obszary te możemy opisać w terminach statystyki testowej $T : \mathcal{X} \to \mathcal{R}$

$$B = \{x \in \mathcal{X} : T(x) \geq c\},$$

$$B^C = \{x \in \mathcal{X} : T(x) < c\}.$$

Wartość c nazywamy wartością krytyczną.

Przykład testu

W naszym przykładzie dotyczącym studentów i ich ocen za statystykę testową przyjmijmy średnią z próby.

$$T(X) = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

Za test statystyczny możemy przyjąć następującą regułę:

- przyjmij hipotezę zerową gdy $5 \leq T(X) \leq 7$,
- odrzuć hipotezę zerową gdy $T(X) < 5$ lub $T(X) > 7$.

Wynik testowania

Bardzo Ważna Tabelka

Stan faktyczny	Decyzja	
	przyjąć H_0 $\psi(x) = 0$	odrzucić H_0 $\psi(x) = 1$
H_0 prawdziwa	decyzja poprawna	błąd I rodzaju
H_0 fałszywa	błąd II rodzaju	decyzja poprawna

Błędy w testowaniu

- Błąd I rodzaju popełniamy popełniamy jeżeli odrzucamy hipotezę zerową, gdy jest ona prawdziwa.
- Błąd II rodzaju popełniamy jeżeli nie odrzucamy hipotezy zerowej gdy jest ona fałszywa.

Prawdopodobieństwo popełniania błędu I rodzaju wynosi

$$Pr(\text{popełnimy błąd I rodzaju}) = P_{\theta_0}(B).$$

Przykład testu

Wyznaczmy prawdopodobieństwo popełnienia błędu I rodzaju.
Dla prawdziwej hipotezy zerowej

$$X_i \sim \mathcal{N}(6, 2^2).$$

Rozkład sumy dla wyników 10 studentów to

$$\sum_{i=1}^{10} X_i \sim \mathcal{N}(60, 10 * 2^2).$$

Rozkład średniej dla wyników 10 studentów to

$$T(X) = \frac{1}{10} \sum_{i=1}^{10} X_i \sim \mathcal{N}(6, \frac{2^2}{10}).$$

Przyjęty obszar krytyczny to $B = [-\infty, 5) \cup (7, \infty)$.
A więc prawdopodobieństwo popełnienia błędu pierwszego rodzaju wynosi

$$\begin{aligned} Pr(T(X) \in B) &= \Phi_{\mu=6,\sigma^2=0.4}(5) + 1 - \Phi_{\mu=6,\sigma^2=0.4}(7) \\ &= 0.05692315 + 0.05692315 \\ &\approx 0.114 \end{aligned}$$

Nasz przykład

Jakie przyjąć c, aby prawdopodobieństwo popełnienia błędu I rodzaju wynosiło 5%?

$$Pr(6 - c \leq T(X) \leq 6 + c) = 0.05$$

Należy wybrać odpowiednie kwantyle rozkładu normalnego.

$$Pr(4.760 \geq T(X)) = 0.025$$

więc $c = 1.24$.

Nasz przykład

Jakie przyjąć c, tak, żeby prawdopodobieństwo popełnienia błędu I rodzaju wynosiło 1%?

$$Pr(6 - c \leq T(X) \leq 6 + c) = 0.01$$

Należy wybrać odpowiednie kwantyle rozkładu normalnego.

$$Pr(4.370 \geq T(X)) = 0.005$$

więc $c = 1.63$.

Poziom istotności

Poziom istotności

Podczas testowania musimy określić poziom istotności, czyli maksymalne prawdopodobieństwo popełnienia błędu pierwszego rodzaju w sytuacji gdy testowana hipoteza zerowa okaże się być prawdziwa.

Podając wyniki testowania, należy wskazać, dla jakiego poziomu istotności, te wyniki zostały otrzymane.

Dwustronna czy jednostronna?

Dla ustalonej hipotezy zerowej, możemy w różny sposób określać hipotezy alternatywne.

- Dwustronna hipoteza alternatywna

$$H_A: \quad \mu \neq 6,$$

$$B = \{x \in \mathcal{X} : |T(x)| > c\},$$

- Jednostronna, lewostronna hipoteza alternatywna

$$H_A: \quad \mu < 6,$$

$$B = \{x \in \mathcal{X} : T(x) < c\},$$

- Jednostronna, prawostronna hipoteza alternatywna

$$H_A: \quad \mu > 6,$$

$$B = \{x \in \mathcal{X} : T(x) > c\},$$

P-wartość

P-wartość (ang. p–value) jest równa najmniejszemu poziomowi istotności, na którym dla wyniku x odrzuca się hipotezę H_0

$$p(x) = 1 - F_{\theta_0}(T(x)).$$

Dla zaobserwowanej średniej 7.1 p-wartość wynosi

- $p(x) = 0.082$ w przypadku alternatywy dwustronnej,
- $p(x) = 0.041$ w przypadku alternatywy jednostronnej, prawostronnej,
- $p(x) = 0.959$ w przypadku alternatywy jednostronnej, lewostronnej.

Ważne

- W procedurze testowania nigdy nie możemy udowodnić prawdziwości H0 – możemy ją jedynie odrzucić.
- Orzeczenie nieistotności wyniku nie oznacza akceptacji H0, lecz jedynie jej nieodrzucenie w danej sytuacji.
- Jeżeli nie mamy podstaw do zaprzeczenia H0, to nie może być ona odrzucona, ale nie oznacza to, że jest prawdziwa.

Gdy wariancja jest znana

Do testowania wartości średniej w podpopulacji, w sytuacji gdy wariancja jest znana wykorzystuje się test oparty na statystyce testowej

$$T(X) = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n}.$$

Przy prawdziwej hipotezie zerowej statystyka ta ma rozkład $\mathcal{N}(0,1)$. Obszary krytyczne wyznacza się ze wzorów

- dla dwustronnej hipotezy alternatywnej $W_\alpha = (-\infty, q_{\alpha/2}] \cup [q_{1-\alpha/2}, \infty)$
- dla lewostronnej hipotezy alternatywnej $W_\alpha = (-\infty, q_\alpha]$
- dla prawostronnej hipotezy alternatywnej $W_\alpha = [q_{1-\alpha}, \infty)$.

Gdy wariancja jest nie znana

Do testowania wartości średniej w podpopulacji, w sytuacji gdy wariancja jest nieznana wykorzystuje się test t-Studenta oparty na statystyce testowej

$$T(X) = \frac{\bar{X} - \mu_0}{S}\sqrt{n}.$$

Przy prawdziwej hipotezie zerowej statystyka ta ma rozkład t-Studenta o $n-1$ stopniach swobody. Obszary krytyczne wyznacza się ze wzorów

- dla dwustronnej hipotezy alternatywnej $W_\alpha = (-\infty, t_{\alpha/2}^{n-1}] \cup [t_{1-\alpha/2}^{n-1}, \infty)$
- dla lewostronnej hipotezy alternatywnej $W_\alpha = (-\infty, t_\alpha^{n-1}]$
- dla prawostronnej hipotezy alternatywnej $W_\alpha = [t_{1-\alpha}^{n-1}, \infty)$.

Co trzeba zapamiętać?

- Co to jest test statystyczny i statystyka testowa?
- Czym różni się hipoteza zerowa od hipotezy alternatywnej?
- Jakie rodzaje hipotez alternatywnych formułuje się najczęściej?
- Co to poziom istotności?
- Jakie błędy można popełnić podczas testowania?
- Co to obszar krytyczny?
- Jakie dwa testy poznaliśmy i czym one się różnią?