Programowanie narzędzi analitycznych Z02

Rafał Woźniak

Faculty of Economic Sciences, University of Warsaw

Warszawa, 14-10-2021

Schemat testowania hipotez statystycznych

1. policzyć statystykę testową t_{test}
2. przy prawdziwości H_0 statystyka testowa ma pewien znany rozkład
3. obszar krytyczny K zależny od hipotezy alternatywnej H_1
4. jeżeli $t_{test} \in K$, to dane są mało prawdopodobne
5. mało prawdopodobne dane mogą być podstawą do odrzucenia H_0

Testowanie hipotez - przykład

- $X_1, X_2, \ldots, X_n$ jest próbą losową z rozkładu normalnego $N(\mu, 1^2)$
- Hipoteza zerowa: $H_0 : \mu = \mu_0$
- Statystyka testowa: $U = \frac{\bar{X}-\mu_0}{\sigma}\sqrt{n}$ przy prawdziwej H_0 ma rozkład $N(0,1)$
- Alternatywa: $H_1 : \mu \neq \mu_0 \Rightarrow K_1 = \left(-\infty, -u(1-\frac{\alpha}{2})\right) \cup \left(u(1-\frac{\alpha}{2}), +\infty\right)$
- $\{X_1, \ldots, X_5\} = \{-0.6256417; -2.3999373; -0.2979613; -1.5966765; 1.3804601\}$
- $H_0 : \mu = 0,\ U_{test} = -1.583027$
- $K = (-\infty, -1.96)) \cup (1.96, +\infty)$
- $U_{test} \notin K$

Wartości krytyczne

- Odczytywanie z tablic statystycznych
- Wykorzystanie funkcji programu R
- Tablica z PNA_Polecenia_Z02.pdf

Rozkład	Kwantyl	Gęstość	...
Normalny	**q**norm(p, mean, sd)	**d**norm(x,mean, sd)	...
Beta	**q**beta(p, s1, s2)	**d**beta(x, s1, s2)	...
χ^2_n	**q**chisq(p, df)	**d**chisq(x, df)	...
Wykładniczy	**q**exp(p, rate)	**d**exp(x, rate)	...
t-Studenta	**q**t(p, df)	**d**t(x, df)	...
Jednostajny(0,1)	**q**unif(p, min, max)	**d**unif(x, min, max)	...
Gamma	**q**gamma(p, s, r)	**d**gamma(x, s, r)	...
F-Snedecora	**q**f(p, df1, df2)	**d**f(x, df1, df2)	...
Dwumianowy	**q**binom(p, s, p)	**d**binom(x, s, p)	...
Poissona	**q**pois(p, lambda)	**d**pois(x, lambda)	...

p-value (p-wartość)

- $H_0 : \theta = \theta_0$ oraz α jest poziomem istotności
- Obszar krytyczny $K = \{T(X) > c_\alpha\}$
- p-wartość jest równa $\mathbb{P}_{\theta_\mathbf{0}}(T(X) > t_{test})$
- Wnioskowanie:
  - Jeśli p-wartość$< \alpha$, to H_0 odrzucamy.
  - Jeśli p-wartość$> \alpha$, to nie ma podstaw do odrzucenia H_0.
  - Źródło: A. Boratyńska, *Wykłady ze statystyki matematycznej (II rok WNE), str. 84.*

p-value (p-wartość)

- $H_0 : \theta = \theta_0$ oraz α jest poziomem istotności
- Obszar krytyczny $K = \{T(X) > c_\alpha\}$
- p-wartość jest równa $\mathbb{P}_{\theta_0}(T(X) > t_{test})$
- Wnioskowanie:
  - Jeśli p-wartość$< \alpha$, to H_0 odrzucamy.
  - Jeśli p-wartość$> \alpha$, to nie ma podstaw do odrzucenia H_0.
  - Źródło: A. Boratyńska, *Wykłady ze statystyki matematycznej (II rok WNE), str. 84.*

p-value a hipoteza alternatywna

- $H_1 : \theta \neq \theta_0 \Rightarrow$ p-value $= \mathbb{P}_{\theta_0}(T(X) > |t_{test}|)$
- $H_1 : \theta < \theta_0 \Rightarrow$ p-value $= \mathbb{P}_{\theta_0}(T(X) < t_{test})$
- $H_1 : \theta > \theta_0 \Rightarrow$ p-value $= \mathbb{P}_{\theta_0}(T(X) > t_{test})$

Bardzo proszę żeby osoby, które nie miały programowania zapoznały się i spróbowały uruchomić kody z:

- Przemysław Biecek, Przewodnik po pakiecie R, 2014, (link), strony 37-45.
- Mikołaj Rybiński, Krótkie wprowadzenie do R dla programistów, z elementami statystyki opisowej, (link), strony 30-34.

Podstawowe testy są zaprogramowane w R. Część zagadnień trzeba zapisać samemu.

Zadanie 1* (Zadanie 6.12)
Niech $X_1, X_2, \ldots, X_n$, $n > 1$, będą niezależnymi zmiennymi losowymi o tym samym rozkładzie wykładniczym o gęstości

$$p_\lambda(x) = \lambda e^{\lambda x} \mathbb{1}_{[0,\infty)}(x),$$

$\lambda > 0$. Wiedząc, że zmienna losowa $T = 2\lambda \sum_{i=1}^{n} X_i \sim \chi^2_{2n}$ wyznaczyć a, b tak, aby

$$P\{T < a\} = P\{T > b\} = \frac{\alpha}{2}.$$

Wyznaczyć na tej podstawie przedział ufności dla parametru λ na poziomie ufności $1 - \alpha$.

Cel zajęć

Zadanie 2 (Modyfikacja zadania 7.5)
Niech $x_1, x_2, \ldots, x_n$ będą wynikami n-elementowej próby losowej pobranej z populacji, w której cecha X ma rozkład jednostajny na przedziale $(0, \theta)$. Do weryfikacji hipotezy $H : \theta = \theta_0$ przy alternatywie $K : \theta > \theta_0$ zaproponowano test: gdy $max(x_1, x_2, \ldots, x_n) = x_{n:n} \geq c$ hipotezę H odrzucamy na korzyść hipotezy K. Wykorzystując fakt, że statystyka $X_{n:n}$ ma rozkład o gęstości

$$f(z) = \begin{cases} \frac{n}{\theta^n} z^{n-1} & \text{gdy } z \in (0, \theta) \\ 0 & \text{w p.p.} \end{cases},$$

a) wyznacz stałą c tak, aby rozmiar testu był równy 0,1

b) dla realizacji 0.08436747 1.04285094 0.04980488 1.19777030 1.05726010 zweryfikuj hipotezę $H : \theta = \sqrt{2}$.