Testy statystyczne — teoria

przygotowanie: dr A. Goroncy, dr J. Karłowska-Pik

Niech $X_1, \ldots, X_n$ będzie próbą losową prostą z rozkładu P_θ, $\theta \in \Theta$ oraz niech $\alpha \in (0,1)$ będzie poziomem istotności (najczęściej 0,1, 0,05, czy 0,01).

Oznaczenia: Φ — dystrybuanta rozkładu $N(0,1)$,
$F_{t(n-1)}$ — dystrybuanta rozkładu t-Studenta z $n-1$ stopniami swobody,
n_i, n_{ij} — liczebności empiryczne (zaobserwowane),
n_i^0, n_{ij}^0 — liczebności teoretyczne,
$F_{\chi^2(k-1)}$ — dystrybuanta rozkładu χ^2 z $k-1$ stopniami swobody,

1. Test Studenta dla jednej średniej.
Hipoteza zerowa: Średnia wartość zmiennej jest równa określonej wartości a_0 ($a = a_0$).
Hipoteza alternatywna 1.: Średnia wartość zmiennej jest różna od określonej wartości a_0 ($a \neq a_0$).
Hipoteza alternatywna 2.: Średnia wartość zmiennej jest mniejsza od określonej wartości a_0 ($a < a_0$).
Hipoteza alternatywna 3.: Średnia wartość zmiennej jest większa od określonej wartości a_0 ($a > a_0$).

a) X ma rozkład normalny o znanej wariancji σ^2.

Statystyka testowa: $T_n = \sqrt{n}\dfrac{\bar{x} - a_0}{\sigma}$.

Obszar krytyczny 1.: $K = (-\infty, -\Phi^{-1}(1-\alpha/2)) \cup (\Phi^{-1}(1-\alpha/2), +\infty)$,

Obszar krytyczny 2.: $K = (-\infty, -\Phi^{-1}(1-\alpha))$,

Obszar krytyczny 3.: $K = (\Phi^{-1}(1-\alpha), +\infty)$.

b) X ma rozkład normalny o nieznanej wariancji σ^2.

Statystyka testowa: $T_n = \sqrt{n}\dfrac{\bar{x} - a_0}{s}$.

Obszar krytyczny 1.: $K = (-\infty, -F^{-1}_{t(n-1)}(1-\alpha/2)) \cup (F^{-1}_{t(n-1)}(1-\alpha/2), +\infty)$ dla $n \leqslant 30$,
$K = (-\infty, -\Phi^{-1}(1-\alpha/2)) \cup (\Phi^{-1}(1-\alpha/2), +\infty)$ dla $n > 30$,

Obszar krytyczny 2.: $K = (-\infty, -F^{-1}_{t(n-1)}(1-\alpha))$ dla $n \leqslant 30$,
$K = (-\infty, -\Phi^{-1}(1-\alpha))$ dla $n > 30$,

Obszar krytyczny 3.: $K = (F^{-1}_{t(n-1)}(1-\alpha), +\infty)$ dla $n \leqslant 30$,
$K = (\Phi^{-1}(1-\alpha), +\infty)$ dla $n > 30$.

c) X ma rozkład dowolny, istnieje D^2X, $n > 30$.

Statystyka testowa: $T_n = \sqrt{n}\dfrac{\bar{x} - a_0}{\sigma_0}$ lub $T_n = \sqrt{n}\dfrac{\bar{x} - a_0}{s}$, lub $T_n = \sqrt{n}\dfrac{\bar{x} - a_0}{\hat{s}}$,
gdzie σ_0 jest odchyleniem standardowym rozkładu przy założeniu prawdziwości hipotezy zerowej, o ile wariancja rozważanego rozkładu jest funkcją jego wartości oczekiwanej (np. w rozkładzie "0-1", dwumianowym, Poissona, geometrycznym itp.).

Obszar krytyczny 1.: $K = (-\infty, -\Phi^{-1}(1-\alpha/2)) \cup (\Phi^{-1}(1-\alpha/2), +\infty)$,

Obszar krytyczny 2.: $K = (-\infty, -\Phi^{-1}(1-\alpha))$,

Obszar krytyczny 3.: $K = (\Phi^{-1}(1-\alpha), +\infty)$.

2. Test dla dwóch średnich i prób niezależnych

Hipoteza zerowa: Średnie wartości zmiennej są takie same w dwóch różnych populacjach ($a_1 = a_2$).
Hipoteza alternatywna 1.: Średnie wartości zmiennej są różne w badanych populacjach ($a_1 \neq a_2$).
Hipoteza alternatywna 2.: Średnia wartość zmiennej w pierwszej populacji jest mniejsza od średniej wartości zmiennej w drugiej populacji ($a_1 < a_2$).
Hipoteza alternatywna 3.: Średnia wartość zmiennej w pierwszej populacji jest większa od średniej wartości zmiennej w drugiej populacji ($a_1 > a_2$).

a) X ma w obu populacjach rozkład normalny o znanych wariancjach σ_1^2 i σ_2^2.

Statystyka testowa: $T_n = \dfrac{\bar{x_1} - \bar{x_2}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$.

Obszar krytyczny 1.: $K = (-\infty, -\Phi^{-1}(1-\alpha/2)) \cup (\Phi^{-1}(1-\alpha/2), +\infty)$,

Obszar krytyczny 2.: $K = (-\infty, -\Phi^{-1}(1-\alpha))$,

Obszar krytyczny 3.: $K = (\Phi^{-1}(1-\alpha), +\infty)$.

b) X ma w obu populacjach rozkład normalny o nieznanych, ale równych wariancjach σ_1^2 i σ_2^2.

Statystyka testowa: $T_n = \dfrac{\bar{x_1} - \bar{x_2}}{\sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \cdot \dfrac{n_1+n_2}{n_1 n_2}}}$.

Obszar krytyczny 1.: $K = (-\infty, -F^{-1}_{t(n_1+n_2-2)}(1-\alpha/2)) \cup (F^{-1}_{t(n_1+n_2-2)}(1-\alpha/2), +\infty)$,

Obszar krytyczny 2.: $K = (-\infty, -F^{-1}_{t(n_1+n_2-2)}(1-\alpha))$,

Obszar krytyczny 3.: $K = (F^{-1}_{t(n_1+n_2-2)}(1-\alpha), +\infty)$.

c) X ma w obu populacjach rozkład normalny o nieznanych wariancjach σ_1^2 i σ_2^2.

Statystyka testowa: $C_n = \dfrac{\bar{x_1} - \bar{x_2}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ (statystyka Cochrana i Coxa).

Obszar krytyczny 1.: $K = (-\infty, -c^{n_1,n_2}_{1-\alpha/2}) \cup (c^{n_1,n_2}_{1-\alpha/2}, +\infty)$,

Obszar krytyczny 2.: $K = (-\infty, -c^{n_1,n_2}_{1-\alpha})$,

Obszar krytyczny 3.: $K = (c^{n_1,n_2}_{1-\alpha}, +\infty)$,

gdzie

$$c^{n_1,n_2}_{1-\alpha} \approx \left(\frac{s_1^2}{n_1} F^{-1}_{t(n_1-1)}(1-\alpha) + \frac{s_2^2}{n_2} F^{-1}_{t(n_2-1)}(1-\alpha)\right) : \left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right).$$

d) X ma w obu populacjach rozkład o nieznanych wariancjach σ_1^2 i σ_2^2, próby mają liczebności większe bądź równe 100.

Statystyka testowa: $T_n = \dfrac{\bar{x_1} - \bar{x_2}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$.

Obszar krytyczny 1.: $K = (-\infty, -\Phi^{-1}(1-\alpha/2)) \cup (\Phi^{-1}(1-\alpha/2), +\infty)$,

Obszar krytyczny 2.: $K = (-\infty, -\Phi^{-1}(1-\alpha))$,

Obszar krytyczny 3.: $K = (\Phi^{-1}(1-\alpha), +\infty)$.

3. Test dla dwóch średnich i prób zależnych
Hipoteza zerowa: Dwie zmienne zależne (o rozkładach normalnych) mają jednakowe średnie (inaczej: różnica $D = X - Y$ odpowiadających sobie wartości zmiennych ma średnią równą 0).
Hipoteza alternatywna 1.: Zmienne zależne mają różne średnie (inaczej: różnica $D = X - Y$ odpowiadających sobie wartości zmiennych ma średnią różną od 0).
Hipoteza alternatywna 2.: Pierwsza ze zmiennych ma średnią mniejszą niż druga (inaczej: różnica $D = X - Y$ odpowiadających sobie wartości zmiennych ma średnią ujemną).
Hipoteza alternatywna 3.: Pierwsza ze zmiennych ma średnią większą niż druga (inaczej: różnica $D = X - Y$ odpowiadających sobie wartości zmiennych ma średnią dodatnią).

Statystyka testowa: $T_n = \frac{\bar{d}}{s_d}\sqrt{n}.$

Obszar krytyczny 1.: $K = (-\infty, -F^{-1}_{t(n-1)}(1-\alpha/2)) \cup (F^{-1}_{t(n-1)}(1-\alpha/2), +\infty)$ dla $n \leqslant 30$,
$K = (-\infty, -\Phi^{-1}(1-\alpha/2)) \cup (\Phi^{-1}(1-\alpha/2), +\infty)$ dla $n > 30$,

Obszar krytyczny 2.: $K = (-\infty, -F^{-1}_{t(n-1)}(1-\alpha))$ dla $n \leqslant 30$,
$K = (-\infty, -\Phi^{-1}(1-\alpha))$ dla $n > 30$,

Obszar krytyczny 3.: $K = (F^{-1}_{t(n-1)}(1-\alpha), +\infty)$ dla $n \leqslant 30$,
$K = (\Phi^{-1}(1-\alpha), +\infty)$ dla $n > 30$.

4. Test chi-kwadrat zgodności
Założenia testu: Zmienna ma rozkład dyskretny, przyjmuje tylko wartości $l_1, \ldots, l_k$ z prawdopodobieństwami odpowiednio $p_1^0, \ldots, p_k^0$, które nie są znane.
Hipoteza zerowa: Zmienna ma rozkład dyskretny z określonymi prawdopodobieństwami $p_1^0, \ldots, p_k^0$.
Hipoteza alternatywna: Zmienna ma rozkład z innymi prawdopodobieństwami niż zadane.

Statystyka testowa: $\chi^2 = \sum_{i=1}^k \frac{(n_i - n_i^0)^2}{n_i^0} = \sum_{i=1}^k \frac{(n_i - np_i^0)^2}{np_i^0}.$

Obszar krytyczny: $K = (F^{-1}_{\chi^2(k-1)}(1-\alpha), +\infty)$.
Uwagi:

- Jeżeli rozkład teoretyczny zależy od d nieznanych parametrów, to parametry te wyznaczamy metodą największej wiarogodności, a liczbę stopni swobody zmniejszamy o d.
- Przybliżenie rozkładem chi-kwadrat uznajemy za dopuszczalne, gdy $np_i^0 \geqslant 5$, $i = 1, \ldots, k$, a za dobre, gdy $np_i^0 \geqslant 10$, $i = 1, \ldots, k$. Jeśli liczba kategorii jest duża (> 6), to zgadzamy się stosować przybliżenie rozkładem chi-kwadrat także wtedy, gdy dla jednej lub dwóch kategorii $1 \leqslant np_i^0 < 5$. Mało liczne kategorie można również łączyć z kategoriami sąsiednimi, redukując wówczas odpowiednio liczbę stopni swobody.
- W przypadku zmiennej o rozkładzie z ciągłą dystrybuantą dane grupujemy w k ($10k \leqslant n$) klas. Prawdopodobieństwa teoretyczne wyliczamy z dystrybuanty. Klasy staramy się dobrać tak, aby prawdopodobieństwa znalezienia się w klasie były równe $1/k$, a liczebności teoretyczne były co najmniej równe 5. Testujemy wówczas hipotezę zerową: Zmienna ma rozkład o podanej dystrybuancie.

5. Test Kołmogorowa
Hipoteza zerowa: Zmienna ma rozkład o zadanej dystrybuancie F.
Hipoteza alternatywna: Zmienna ma rozkład o innej niż zadana dystrybuancie.
Wymagania testu: Ciągłość dystrybuanty.

a) $n \leqslant 100$

Statystyka testu: $D_n = \max\{D_n^+, D_n^-\}$,
gdzie $D_n^+ = \max_{1\leqslant i\leqslant n} \left|\frac{i}{n} - F(x_{(i)})\right|$, $D_n^- = \max_{1\leqslant i\leqslant n} \left|F(x_{(i)}) - \frac{i-1}{n}\right|$.
Obszar krytyczny: $(d_n(1-\alpha), 1]$ (odczytujemy z tablic Kołmogorowa -Smirnowa, jest to taka wartość, dla której $P(D_n \geqslant d_n(1-\alpha)) = \alpha$).

b) $n > 100$.
Statystyka testu: $\sqrt{n}D_n = \sqrt{n}\max\{D_n^+, D_n^-\}$ (czasem $(\sqrt{n} + 0,12 + 0,11/\sqrt{n})D_n$),
gdzie $D_n^+ = \max_{1\leqslant i\leqslant n} \left|\frac{i}{n} - F(x_{(i)})\right|$, $D_n^- = \max_{1\leqslant i\leqslant n} \left|F(x_{(i)}) - \frac{i-1}{n}\right|$.
Obszar krytyczny: $(\lambda_{1-\alpha}, +\infty)$, gdzie $\lambda_{1-\alpha}$ jest kwantylem rzędu $1-\alpha$ granicznego rozkładu Kołmogorowa.

Uwaga: W przypadku danych zgrupowanych w klasy bierzemy pod uwagę prawy koniec każdej z klas i zamiast podanych statystyk wyznaczamy wartość maksymalną statystyki $|F_n(x_i) - F(x_i)|$, gdzie F_n jest dystrybuantą empiryczną.

6. Test chi-kwadrat niezależności
Założenia testu: Cechy X, Y są jakościowe (nominalne lub o wartościach uporządkowanych).
Hipoteza zerowa: X, Y są zmiennymi niezależnymi.
Hipoteza alternatywna: X, Y są zależne.

Statystyka testowa: $\chi^2 = \sum_{j=1}^{k}\sum_{i=1}^{r} \frac{(n_{ij} - n_{ij}^0)^2}{n_{ij}^0}$, gdzie

$$\begin{aligned} r &\text{ — liczba kategorii zmiennej } X \text{ (liczba wierszy w tablicy kontyngencji)},\\ k &\text{ — liczba kategorii zmiennej } Y \text{ (liczba kolumn w tablicy kontyngencji)},\\ n_{ij} &\text{ — liczba wystąpień w próbie par obserwacji } (x_i, y_j),\\ n_{ij}^0 &= \frac{\sum_{j=1}^{k} n_{ij} \cdot \sum_{i=1}^{r} n_{ij}}{n},\\ n &= \sum_{i=1}^{r}\sum_{j=1}^{k} n_{ij}. \end{aligned}$$

Obszar krytyczny: $K = (F^{-1}_{\chi^2((r-1)(k-1))}(1-\alpha), +\infty)$.
Uwagi:

- Podobnie jak w teście chi-kwadrat zgodności, przybliżenie statystyki testowej rozkładem chi-kwadrat stosujemy, gdy liczebności teoretyczne prób w wierszach (kolumnach) są stosunkowo duże ($n_{ij}^0 \geqslant 5$).
- Gdy tablica kontyngencji ma rozmiar 2×2 i liczebności próby w wierszach (kolumnach) są zbyt małe, można oprzeć się na tzw. dokładnym teście Fishera (którego tu nie będziemy omawiać).
- W przypadku pary cech o uporządkowanych kategoriach test niezależności może okazać się zwodniczy. Może wówczas zajść potrzeba wprowadzenia odpowiedniej miary zależności między cechami (tego nie będziemy tu omawiać).