Dwie niezależne próby

Cecha Y w populacji 1 ma rozkład $N(\mu_1, \sigma_1)$.
Cecha Y w populacji 2 ma rozkład $N(\mu_2, \sigma_2)$.

Pytanie

Jaka jest różnica między średnimi w populacjach: $\mu_1 - \mu_2$?

Dwie niezależne próby

Badamy próbę o rozmiarze n_1 z populacji 1 otrzymując dane:

$$\bar{y}_1, \qquad s_1, \qquad SE_1 = \frac{s_1}{\sqrt{n_1}}$$

Badamy próbę o rozmiarze n_2 z populacji 2 otrzymując dane:

$$\bar{y}_2, \qquad s_2, \qquad SE_2 = \frac{s_2}{\sqrt{n_2}}$$

Chcemy wyznaczyć PU dla $\mu_1 - \mu_2$. Różnica średnich $\bar{y}_1 - \bar{y}_2$ jest estymatorem $\mu_1 - \mu_2$ i będzie środkiem PU. Potrzebujemy jeszcze wyznaczyć SE.

SE dla różnicy dwóch średnich

Istnieją dwa sposoby wyliczania SE dla różnicy dwóch średnich $\bar{y}_1 - \bar{y}_2$:

- *nieuśredniony* lub *nie łączony* (*unpooled*), który będziemy stosować w większości wypadków,
- *uśredniony* lub *łączony* (*pooled*), który będziemy stosować jedynie w sytuacji, gdy będzie to wyraźnie wymagane w zadaniu.

W sytuacji gdy $n_1 = n_2$ obie metody dają ten sam wynik.

SE dla różnicy dwóch średnich

Metoda zwykła (nie łączona)

Obliczamy SE dla każdej próby z osobna:

$$SE_1 = \frac{s_1}{\sqrt{n_1}} \qquad SE_2 = \frac{s_2}{\sqrt{n_2}}$$

a następnie nieuśrednione SE:

$$(N)SE = \sqrt{(SE_1)^2 + (SE_2)^2}$$

SE dla różnicy dwóch średnich

Metoda łączona

Obliczamy sumę kwadratów odchyleń dla obu prób:

$$SS_1 = \sum (y_{1,i} - \bar{y}_1)^2 \qquad SS_1 = \sum (y_{2,i} - \bar{y}_2)^2$$

uśrednioną wariancję:

$$s_c^2 = \frac{SS_1 + SS_2}{n_1 + n_2 - 2}$$

a następnie uśrednione (łączone) SE:

$$(U)SE = \sqrt{s_c^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = s_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Przykład 1

Dla próby rozmiaru $n_1 = 15$ z populacji 1 otrzymujemy $\bar{y}_1 = 75$ oraz $SS_1 = 600$.
Dla próby rozmiaru $n_2 = 10$ z populacji 2 otrzymujemy $\bar{y}_2 = 55$ oraz $SS_2 = 300$.
Obliczyć nieuśrednione oraz uśrednione SE dla $\bar{y}_1 - \bar{y}_2$.

Wyliczamy $s_1 = 6.55$ oraz $s_2 = 5.77$. Wówczas $SE_1 = \frac{s_1}{\sqrt{n_1}} = 1.69$ oraz $SE_2 = \frac{s_2}{\sqrt{n_2}} = 1.83$. Stąd:

$$(N)SE = \sqrt{(SE_1)^2 + (SE_2)^2} = 2.49$$

Dla metody łączonej wyliczamy $s_c = \frac{SS_1 + SS_2}{n_1 + n_2 - 2} = 6.26$. Wówczas:

$$(U)SE = s_c \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 2.55$$

PU dla różnicy dwóch średnich

PU dla pojedynczej średniej μ na poziomie ufności $(1-\alpha)$ wyliczaliśmy następująco:

$$\bar{y} \pm t_{\alpha/2} SE_{\bar{y}} = \text{estymator} \pm \text{kwantyl} \cdot SE$$

PU dla różnicy dwóch średnich $\mu_1 - \mu_2$ na poziomie ufności $(1-\alpha)$

$$(\bar{y}_1 - \bar{y}_2) \pm t(df)_{\alpha/2} SE_{\bar{y}_1 - \bar{y}_2}$$

gdzie liczba stopni swobody wynosi:

$$df = \frac{(SE_1^2 + SE_2^2)^2}{\frac{SE_1^4}{n_1 - 1} + \frac{SE_2^4}{n_2 - 1}}$$

$\min\{n_1 - 1, n_2 - 1\} \leq df \leq n_1 + n_2 - 2$ (w przybliżonych obliczeniach często stosujemy $n_1 + n_2 - 2$)

Przykład 1 c.d.

Dla próby rozmiaru $n_1 = 15$ z populacji 1 otrzymujemy $\bar{y}_1 = 75$ oraz $SS_1 = 600$.
Dla próby rozmiaru $n_2 = 10$ z populacji 2 otrzymujemy $\bar{y}_2 = 55$ oraz $SS_2 = 300$.
Skonstruuj 95% PU dla $\mu_1 - \mu_2$.

$\bar{y}_1 - \bar{y}_2 = 20$, $SE_1 = 1.69$, $SE_2 = 1.83$.

$$df = \frac{(SE_1^2 + SE_2^2)^2}{\frac{SE_1^4}{n_1 - 1} + \frac{SE_2^4}{n_2 - 1}} = 21.08$$

Stąd (dla nieuśrednionego SE) dostajemy przedział ufności:

$$20 \pm t(21)_{0.025} 2.49 = 20 \pm 2.08 \cdot 2.49 = 20 \pm 5.18$$

Dla uśrednionego SE dostajemy przedział ufności

$$20 \pm t(21)_{0.025} 2.55 = 20 \pm 2.08 \cdot 2.55 = 20 \pm 5.30$$

Przykład 2

Porównujemy wzrost roślin hodowanych w różnych warunkach oświetleniowych:

- mocne oświetlenie: $n_1 = 21$, $\bar{y}_1 = 24.6$, $SE_1 = 7.0$
- słabe oświetlenie: $n_2 = 22$, $\bar{y}_2 = 17.6$, $SE_2 = 5.0$

Skonstruuj 95% PU dla $\mu_1 - \mu_2$.

$$(N)SE = \sqrt{SE_1^2 + SE_2^2} = 8.6$$

$$df = \frac{(SE_1^2 + SE_2^2)^2}{\frac{SE_1^4}{n_1 - 1} + \frac{SE_2^4}{n_2 - 1}} = 36.6$$

Stąd (dla nieuśrednionego SE) dostajemy przedział ufności:

$$7.0 \pm t(37)_{0.025} 8.6 = 7.0 \pm 2.02 \cdot 8.6 = 7.0 \pm 17.4$$

Uwaga: PU to $(-10.4, 24.4)$, czyli zawiera zarówno liczby dodatnie, jak i ujemne.

Testowanie hipotez

Chcemy odpowiedzieć na pytanie naukowe dotyczące populacji w oparciu o próbę (informacja fragmentaryczna), np.

- Czy prawdopodobieństwo wyrzucenia orła wynosi $\frac{1}{2}$ (tzn. czy moneta jest uczciwa)?
- Czy prawdopodobieństwo sukcesu w próbie Bernoulliego wynosi p_0 (dla pewnej konkretnej wartości p_0)?
- Czy średnia w populacji wynosi 27?
- Czy średnie wartości cechy w obu populacjach są równe?
- Czy różnica między średnimi w obu populacjach wynosi μ_0?

Testowanie hipotez

Pytania dopuszczają tylko odpowiedzi TAK lub NIE, ale dotyczą całej populacji, do której nie mamy dostępu. Nasza decyzja (w oparciu o próbę) jest zagrożona błędem. Dlatego formułujemy ostrożne odpowiedzi:

- zamiast TAK, mówimy "w oparciu o badaną próbę nie możemy wykluczyć postawionej hipotezy"
- zamiast NIE, mówimy, "jest to mało prawdopodobne"

Później wprowadzimy ilościowy parametr usprawiedliwiania takich decyzji (p-wartość).

Analogia: czujnik dymu

Czujnik dymu ma ostrzegać przed pożarem. Czujnik reaguje na cząstki dymu w powietrzu.

- czujnik może być w dwóch stanach: CICHO lub GŁOŚNO
- dom może być dwóch stanach: NIE MA POŻARU lub JEST POŻAR

Na podstawie stanu czujnika dymu podejmujemy decyzję: CICHO – zostajemy, GŁOŚNO – uciekamy.

Analogia: czujnik dymu

Są dwie sytuacje prawidłowej reakcji i dwie sytuacje błędnej reakcji:

- Na ogół NIE MA POŻARU i czujnik jest CICHO, więc zostajemy (dobra decyzja).
- Czasami NIE MA POŻARU, ale czujnik jest GŁOŚNO, więc uciekamy (błędna decyzja – strata czasu) – błąd I-go rodzaju.
- Czasami JEST POŻAR, ale czujnik jest CICHO, więc zostajemy (błędna decyzja – niebezpieczeństwo) – błąd II-go rodzaju.
- Czasami JEST POŻAR i czujnik jest GŁOŚNO, więc uciekamy (dobra decyzja).

Hipoteza zerowa i hipoteza alternatywna

Stan podstawowy (NIE MA POŻARU) nazywamy *hipotezą zerową* i oznaczamy H_0.
Drugi możliwy stan (JEST POŻAR) nazywamy *hipotezą alternatywną* i oznaczamy H_A.
Decyzje zwykle opisujemy w odniesieniu do hipotezy zerowej H_0:

- decyzja „uciekamy" jest odrzuceniem H_0,
- decyzja „zostajemy" odpowiada nieodrzuceniu H_0.

Decyzję podejmujemy w oparciu o zachowanie czujnika dymu, którego rolę będzie pełnić statystyka testowa, czyli pewna wielkość obliczona z próby.
Gdy czujnik jest GŁOŚNO, mówimy, że wynik testu jest „istotny" (tzn. istotny wynik powoduje odrzucenie H_0).
Gdy czujnik jest CICHO, to wynik testu jest „nieistotny" i nie odrzucamy H_0.

Hipoteza zerowa i hipoteza alternatywna

- Błąd I-go rodzaju: odrzucamy H_0, pomimo że jest prawdziwa (uciekamy, choć nie ma pożaru)
- Błąd II-go rodzaju: nie odrzucamy H_0, choć prawdziwa jest H_A (zostajemy, choć jest pożar)

Czujnik dymu ma pewną ustaloną czułość (reaguje na określoną ilość dymu). Jeśli czujnik:

- jest zbyt czuły, to często powoduje fałszywe alarmy (błędy I-go rodzaju),
- jest nie dość czuły, to nie będzie się włączał, kiedy powinien (błędy II-go rodzaju).

Zwiększając czułość zmniejszamy prawdopodobieństwo błedu II-go rodzaju, ale zwiększamy prawdopodobieństwo błędu I-go rodzaju.

Poziom istotności

Jak opisać czułość testu?

- poziom istotności α to prawdopodobieństwo błędu I-go rodzaju. Poziom istotności powinno się ustalić jeszcze przed przeprowadzeniem eksperymentu.
- β – prawdopodobieństwo błędu II-go rodzaju (zależy np. od wielkości pożaru)

Hipoteza zerowa i alternatywna

Hipoteza zerowa H_0 – np. $\mu = 0$, $\mu_1 = \mu_2$ (tzn. $\mu_1 - \mu_2 = 0$)
Hipoteza alternatywna H_A – np. $\mu \neq 0$, $\mu_1 \neq \mu_2$ (tzn. $\mu_1 - \mu_2 \neq 0$).
Aby kontrolować błąd I-go rodzaju, trzeba znać rozkład statystyki testowej przy H_0.
UWAGA: Odrzucenie H_0 oznacza, że wierzymy w H_A.
Nieodrzucenie H_0 oznacza, że nie mamy dość silnych dowodów przemawiających za H_A, ale nie oznacza udowodnienia prawdziwości H_0 (tego na ogół nie potrafimy zrobić przy pomocy statystyki).

Przykład

Przykład

Załóżmy, że mamy próbę z populacji o rozkładzie normalnym o (nieznanej) średniej μ. Chcemy przetestować hipotezę $H_0 : \mu = 5$ przeciw alternatywie $H_A : \mu \neq 5$.

Konstruujemy PU dla μ w oparciu o dane.

- jeśli PU nie zawiera 5, to odrzucimy H_0 na rzecz H_A,
- jeśli PU zawiera 5, to oznacza, że nie powinniśmy odrzucać H_0; ponieważ PU zawiera także wiele innych wartości, nie mamy dowodu, że H_0 jest prawdziwa

Przykład c.d.

PU na poziomie $(1-\alpha)$ jest dany wzorem

$$\bar{y} \pm t_{\alpha/2} SE$$

Aby sprawdzić, czy zawiera on wartość 5, należy sprawdzić czy:

$$t = \frac{\bar{y} - 5}{SE} \in (-t_{\alpha/2}, t_{\alpha/2})$$

Jeśli TAK, to statystyka t jest nieistotna i nie odrzucamy H_0.
Jeśli NIE, to statystyka t jest istotna i odrzucamy H_0.
Zbiór $(-\infty, -t_{\alpha/2}) \cup (t_{\alpha/2}, \infty)$ nazywamy *obszarem krytycznym* lub *obszarem odrzuceń* (jeśli statystyka testowa znajdzie się w obszarze krytycznym, to odrzucamy H_0). Zauważmy, że postać statystyki testowej zależy od H_0.
Statystyka testowa $\frac{\bar{y}-\mu}{SE}$ ma rozkład Studenta z $n-1$ stopniami swobody.

Poziom istotności

Poziom istotności α = prawdopodobieństwo błędu I-rodzaju (odrzucenie H_0, gdy jest prawdziwa, tzn. fałszywy dodatni wynik testu).
α wybieramy przed przystąpieniem do testowania (typowe wartości to 0.05, 0.01 lub 0.1). Wybór α powinien zależeć od konsekwencji błędów I-go i II-go rodzaju.
Uwaga:
Rozważaliśmy zbiór krytyczny $(-\infty, -t_{\alpha/2}) \cup (t_{\alpha/2}, \infty)$, bo H_A ($\mu \neq 5$) jest symetryczna. Gdybyśmy byli zainteresowani alternatywami jednostronnymi, np. $H_A: \ \mu < 5$, to obszar krytyczny miałby postać $(t_{\alpha/2}, \infty)$.