6 Testy t-Studenta

Test t-Studenta dla jednej próby

Rozważamy model jednej próby prostej z populacji o rozkładzie normalnym.

Uwaga: Założenie normalności rozkładów błędów możemy (ewentualnie) zastąpić założeniem mówiącym o dysponowaniu dużą próbą, tzn. $n > 100$.

Hipoteza zerowa: wartość oczekiwana (średnia) badanej cechy nie różni się istotnie od zadanej wartości.

$$H_0: \ \mu = \mu_0$$

Hipotezy alternatywne:

$$H_1: \ \mu \neq \mu_0$$
$$H_1: \ \mu > \mu_0$$
$$H_1: \ \mu < \mu_0$$

Statystyka testowa:

$$t = \frac{\bar{X} - \mu_0}{S}\sqrt{n}.$$

Rozkład statystyki testowej:

$$t|_{H_0} \sim t(n-1)$$

Testy dla dwóch prób

Posiadamy obserwacje jednej zmiennej (cechy) na jednostkach eksperymentalnych pochodzących z dwóch populacji (grup) lub posiadamy dwukrotne obserwacje tej samej zmiennej na tych samych jednostkach eksperymentalnych jednej populacji.

Rodzaje prób:

- Próby niezależne - obserwacje w poszczególnych populacjach (grupach) dokonywane są na różnych jednostkach eksperymentalnych.
- Próby zależne - obserwacje dokonywane są dwukrotnie na tych samych jednostkach eksperymentalnych.

Model: dwie próby proste niezależne z populacji o rozkładach normalnych

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i,\ i = 1, 2$$

gdzie

X_{ij} - j-ta obserwacja badanej cechy X w i-tej populacji (grupie),

μ_i - wartość oczekiwana (średnia, “prawdziwa” wartość) badanej cechy X w i-tej populacji (grupie),

ε_{ij} - błędy (reszty).

Założenia

O błędach zakładamy, że: - mają rozkłady normalne (dokładnie: są zmiennymi losowymi o rozkładach normalnych), - są niezależne (dokładnie: są niezależnymi zmiennymi losowymi), - mają wartość oczekiwaną równą zero (nie ma błędu systematycznego), tzn.

$$E(\varepsilon_{ij}) = 0, \quad j = 1, \ldots, n_i,\ i = 1, 2,$$

- w każdej z dwóch niezależnych prób mają jednakową, stałą i niezerową wariancję, tzn.

$$Var(\varepsilon_{ij}) = \sigma_i^2, \quad j = 1, \ldots, n_i,\ i = 1, 2.$$

Uwaga: Model ma cztery parametry: μ_1, μ_2, σ_1^2 i σ_2^2.

Test t–Studenta dla dwóch prób niezależnych

Uwaga: Założenie normalności rozkładów błędów możemy (ewentualnie) zastąpić założeniem mówiącym o dysponowaniu dużymi próbami, tzn. $n_1, n_2 > 100$.

Hipoteza zerowa: wartości oczekiwane (średnie) badanej cechy w dwóch populacjach (grupach) nie różnią się istotnie.

$$H_0: \ \mu_1 = \mu_2$$

Hipotezy alternatywne:

$$H_1: \ \mu_1 \neq \mu_2$$

$$H_1: \ \mu_1 > \mu_2$$

$$H_1: \ \mu_1 < \mu_2$$

Model z jednorodnymi wariancjami

Zakładamy dodatkowo, że $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Oznacza to, że w modelu mamy jedynie trzy parametry: μ_1, μ_2 i σ^2.

FAKT

Estymatorami nieobciążonymi parametrów modelu są statystyki:

$$\hat{\mu}_1 = \overline{X}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{1j}, \quad \hat{\mu}_2 = \overline{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j},$$

$$\hat{\sigma}^2 = S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2},$$

gdzie

$$S_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2, \quad i = 1, 2.$$

Test t–Studenta dla dwóch prób niezależnych o jednorodnych wariancjach

Statystyka testowa:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S}\sqrt{n},\ n = \frac{n_1 n_2}{n_1+n_2}.$$

Rozkład statystyki testowej:

$$t|_{H_0} \sim t(n_1+n_2-2)$$

Model z niejednorodnymi wariancjami

Zakładamy, że $\sigma_1^2 \neq \sigma_2^2$.

FAKT

Estymatorami nieobciążonymi parametrów modelu są statystyki:

$$\hat{\mu}_1 = \overline{X}_1 = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{1j}, \quad \hat{\mu}_2 = \overline{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j},$$

$$\hat{\sigma}_1^2 = \frac{1}{n_1-1}\sum_{j=1}^{n_1}(X_{1j} - \overline{X}_1)^2, \quad \hat{\sigma}_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2.$$

Test t–Studenta dla dwóch prób niezależnych o niejednorodnych wariancjach

Statystyka testowa:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}.$$

Rozkład statystyki testowej:

$$t|_{H_0} \sim t(m) \text{ (przybliżony)}, \; \frac{1}{m} = \frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1}, \; c = \frac{S_1^2}{n_1} / (\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}).$$

Uwaga. Test ten nosi również nazwę testu Welcha.

DEFINICJA

Niech $X \sim \chi^2(n)$ oraz $Y \sim \chi^2(m)$ będą niezależnymi zmiennymi losowymi.

Mówimy, że zmienna losowa

$$\frac{\frac{1}{n}X}{\frac{1}{m}Y}$$

ma **rozkład F-Snedecora** z n i m stopniami swobody (ozn. $F(n,m)$).

FAKT

$$f(x) = \frac{\Gamma \;\; \frac{n+m}{2}}{\Gamma \;\; \frac{n}{2} \;\; \Gamma \;\; \frac{m}{2}} \;\; \frac{m}{n}^{\frac{m}{2}} \; \frac{x^{\frac{n}{2}-1}}{x + \frac{m}{n}^{\frac{n+m}{2}}} I_{(0,\infty)}(x), \; x \in \quad .$$

Wybór modelu - test F dla dwóch wariancji

Hipoteza zerowa: wariancje badanej cechy w dwóch populacjach (grupach) nie różnią się istotnie.

$$H_0: \; \sigma_1^2 = \sigma_2^2$$

**Hipoteza alternatywna:

$$H_1: \; \sigma_1^2 \neq \sigma_2^2$$

Statystyka testowa:

$$F = \frac{S_1^2}{S_2^2}.$$

Rozkład statystyki testowej:

$$F|_{H_0} \sim F(n_1 - 1, n_2 - 1)$$

Model: dwie próby proste zależne z populacji o rozkładzie normalnym

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad j = 1, \ldots, n, \ i = 1, 2$$

gdzie

X_{ij} - obserwacja badanej cechy X na j-tej jednostce w i-tej próbie,

μ_i - wartość oczekiwana (średnia, “prawdziwa” wartość) badanej cechy X w i-tej próbie,

ε_{ij} - błędy (reszty).

Założenia

O błędach zakładamy, że: 1. mają rozkłady normalne (dokładnie: są zmiennymi losowymi o rozkładach normalnych), 2. są zależne (dokładnie: zależne są zmienne losowe ε_{1j} i ε_{2j} dla każdego j), 3. mają wartość oczekiwaną równą zero (nie ma błędu systematycznego), tzn.

$$E(\varepsilon_{ij}) = 0, \quad j = 1, \ldots, n_i, \ i = 1, 2,$$

4. w każdej z dwóch zależnych prób mają jednakową, stałą i niezerową wariancję, tzn.

$$Var(\varepsilon_{ij}) = \sigma_i^2, \quad j = 1, \ldots, n, \ i = 1, 2.$$

Mamy

$$X_{2j} - X_{1j} = (\mu_2 - \mu_1) + (\varepsilon_{2j} - \varepsilon_{1j}), \quad j = 1, \ldots, n.$$

Podstawiając

$$Z_j = X_{2j} - X_{1j}, \ \delta = \mu_2 - \mu_1, \ \varepsilon_j = \varepsilon_{2j} - \varepsilon_{1j},$$

sprowadzamy model dwóch prób zależnych do modelu jednej próby prostej

$$Z_j = \delta + \varepsilon_j, \quad j = 1, \ldots, n,$$

gdzie δ oznacza różnicę (zmianę) wartości oczekiwanych badanej cechy X w dwóch próbach, a założenia dotyczące błędów są identyczne jak w przypadku modelu jednej próby prostej z populacji o rozkładzie normalnym.

Funkcje związane z testemi t-Studenta:

t.test - test t-Studenta dla jednej próby oraz dla dwóch prób niezależnych i zależnych,

var.test - test F dla dwóch wariancji.