Inżynierskie zastosowanie statystyki – Ćwiczenia

Lista 6

1 Wstęp teoretyczny

Interesuje nas przetestowanie hipotezy $H_0 : \mu_1 - \mu_2 = \Delta_0$ dotyczącej różnicy pomiędzy wartościami średnimi (przeciętnymi) badanej cechy w dwóch populacjach. Δ_0 może być dowolną wartością, w szczególności może być równe 0. Dysponujemy dwiema próbami:

1. n_1-elementową próbą losową z populacji 1 $X_1 = (X_{11}; X_{12}; \ldots; X_{1n_1})$
2. n_2-elementową próbą losową z populacji 2 $X_2 = (X_{21}; X_{22}; \ldots; X_{2n_2})$.

Średnie arytmetyczne z obu prób wynoszą odpowiednio $\bar{X}_1$ i $\bar{X}_2$, a wariancje s_1^2 i s_2^2.

1.1 Model

1. Wariancje σ_1^2 i σ_2^2 znane. Badana cecha X populacji ma w dwóch populacjach rozkład normalny $\mathcal{N}(\mu_1, \sigma_1^2)$ i $\mathcal{N}(\mu_2, \sigma_2^2)$. Statystyka testowa:
$$U = \frac{\hat{X}_1 - \hat{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \tag{1}$$
przy założeniu prawdziwości hipotezy $H_o : \mu_1 - \mu_2 = \Delta_0$ ma rozkład $\mathcal{N}(0, 1)$.

2. Wariancje nieznane, ale równe. Badana cecha X populacji ma w dwóch populacjach rozkład normalny $\mathcal{N}(\mu_1, \sigma_1^2)$ i $\mathcal{N}(\mu_2, \sigma_2^2)$. Statystyka testowa:
$$T = \frac{\hat{X}_1 - \hat{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \tag{2}$$

gdzie

$$S_p = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \tag{3}$$

przy założeniu prawdziwości hipotezy $H_o : \mu_1 - \mu_2 = \Delta_0$ ma rozkład t-Studenta z $n_1 + n_2 - 2$ stopniami swobody.

3. Wariancje nieznane i różne. Badana cecha X populacji ma w dwóch populacjach rozkład normalny $\mathcal{N}(\mu_1, \sigma_1^2)$ i $\mathcal{N}(\mu_2, \sigma_2^2)$. Statystyka testowa:

$$T = \frac{\hat{X}_1 - \hat{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \tag{4}$$

przy założeniu prawdziwości hipotezy $H_o : \mu_1 - \mu_2 = \Delta_0$ ma rozkład t-Studenta z v stopniami swobody, gdzie

$$v = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \tag{5}$$

2 Lista zadań

1. Technolog jest zainteresowany skróceniem czasu schnięcia farby podkładowej. Dwie kompozycje farb są testowane. Preparat 1 ma standardowy skład chemiczny, a preparat 2 ma nowy składnik, który powinien skrócić czas suszenia. Z doświadczenia wiadomo, że odchylenie standardowe czasu suszenia wynosi 8 minut i to rozproszenie nie powinno ulec zmianie przez dodanie nowego składnika. Dziesięć elementów zostało pomalowanych preparatem 1, kolejne 10 - preparatem 2; wszystkie 20 elementów było malowanych w losowej kolejności. Średnie czasy suszenia dla obu prób wyniosły $\bar{x} = 121$ minut i $\bar{x}_2 = 112$ minut, odpowiednio. Jakie wnioski może wyciągnąć technolog odnośnie skuteczności nowego składnika na poziomie istotności $\alpha = 0.05$? Ile wynosi p-wartość w tym teście?

2. Dwa katalizatory zostały użyte w celu określenia ich wpływu na średnią wydajność procesu chemicznego. Katalizator 1 jest obecnie używany, ale katalizator 2 też jest akceptowalny. Ponieważ katalizator 2 jest tańszy, powinien zostać przyjęty,

o ile nie wpływa na wydajność procesu. Czy jest jakaś różnica między średnią wydajnością procesu na poziomie istotności $\alpha = 0.05$ i przy założeniu równości wariancji? $\bar{x}_1 = 92.255$, $\bar{x}_2 = 92.733$, $s_1 = 2.39$, $s_2 = 2.98$.

3. Stężenie arsenu w publicznych zasobach wody pitnej jest potencjalnym zagrożeniem dla zdrowia. W artykule w Arizona Republic (27 maja 2001 r.) przedstawiono dane dotyczące stężenia arsenu w wodzie pitnej w częściach na miliard (ppb) w 10 gminach miejskich w Phoenix i 10 gminach wiejskich Arizony. $\bar{x}_1 = 12.5$, $\bar{x}_2 = 27.5$, $s_1 = 7.63$, $s_2 = 15.3$. Zakładamy, że wariancję rozkładów są różne, czy jest różnica w średnich stężenia arsenu pomiędzy gminami miejskimi, a gminami wiejskimi.

4. Zakładając, że znamy wariancje dwóch populacji sprawdź prawdziwość hipotezy zerowej:

   (a) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 \neq 0$. Wariancje populacji: $\sigma_1 = 1$, $\sigma_2 = 1.2$. Dane z próby: $n_1 = 10$ $n_2 = 20$, $\bar{X}_1 = 2.4$, $\bar{X}_2 = 2.2$

   (b) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 \neq 0$. Wariancje populacji: $\sigma_1 = 2$, $\sigma_2 = 2$. Dane z próby: $n_1 = 100$ $n_2 = 10$, $\bar{X}_1 = 3.4$, $\bar{X}_2 = 3.1$

   (c) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 < 0$. Wariancje populacji: $\sigma_1 = 10$, $\sigma_2 = 1$. Dane z próby: $n_1 = 10$ $n_2 = 50$, $\bar{X}_1 = 10$, $\bar{X}_2 = 9.5$

   (d) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 > 0$. Wariancje populacji: $\sigma_1 = 1$, $\sigma_2 = 1.2$. Dane z próby: $n_1 = 30$ $n_2 = 30$, $\bar{X}_1 = 5$, $\bar{X}_2 = 5.1$

   (e) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 > 0$. Wariancje populacji: $\sigma_1 = 1$, $\sigma_2 = 2$. Dane z próby: $n_1 = 50$ $n_2 = 50$, $\bar{X}_1 = 4.1$, $\bar{X}_2 = 4$

5. Zakładając, że nie znamy wariancji dwóch populacji, ale wiemy, że są równe, sprawdź prawdziwość hipotezy zerowej:

   (a) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 \neq 0$. Dane z próby: $n_1 = 10$ $n_2 = 20$, $\bar{X}_1 = 2.5$, $\bar{X}_2 = 2$, $s_1 = 1.1$, $s_2 = 1.2$

(b) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 \neq 0$. Dane z próby: $n_1 = 30$ $n_2 = 30$, $\bar{X}_1 = 3.1$, $\bar{X}_2 = 3.2$, $s_1 = 2.1$, $s_2 = 2$

(c) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 < 0$. Dane z próby: $n_1 = 20$ $n_2 = 20$, $\bar{X}_1 = 14.4$, $\bar{X}_2 = 13.2$, $s_1 = 5.5$, $s_2 = 5.2$

(d) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 > 0$. Dane z próby: $n_1 = 30$ $n_2 = 20$, $\bar{X}_1 = 8.8$, $\bar{X}_2 = 9$, $s_1 = 2$, $s_2 = 2$

(e) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 > 0$. Dane z próby: $n_1 = 50$ $n_2 = 50$, $\bar{X}_1 = 4.4$, $\bar{X}_2 = 4.9$, $s_1 = 3$, $s_2 = 3.2$

6. Zakładając, że nie znamy wariancji dwóch populacji, ale wiemy, że nie są równe, sprawdź prawdziwość hipotezy zerowej:

   (a) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 > 0$. Dane z próby: $n_1 = 20$ $n_2 = 20$, $\bar{X}_1 = 2.4$, $\bar{X}_2 = 2.2$, $s_1 = 1$, $s_2 = 1.2$

   (b) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 \neq 0$. Dane z próby: $n_1 = 10$ $n_2 = 20$, $\bar{X}_1 = 3.4$, $\bar{X}_2 = 3.6$, $s_1 = 2$, $s_2 = 1.8$

   (c) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 < 0$. Dane z próby: $n_1 = 30$ $n_2 = 30$, $\bar{X}_1 = 5.4$, $\bar{X}_2 = 6$, $s_1 = 2.5$, $s_2 = 3$

   (d) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 \neq 0$. Dane z próby: $n_1 = 100$ $n_2 = 200$, $\bar{X}_1 = 10$, $\bar{X}_2 = 9.5$, $s_1 = 10$, $s_2 = 1$

   (e) $H_0 : \mu_1 - \mu_2 = 0$, $H_1 : \mu_1 - \mu_2 < 0$. Dane z próby: $n_1 = 20$ $n_2 = 60$, $\bar{X}_1 = 2.4$, $\bar{X}_2 = 3.2$, $s_1 = 6$, $s_2 = 2$