Inżynierskie zastosowania statystyki - Ćwiczenia nr 9

1 Weryfikacja hipotezy odnośnie różnicy pomiędzy wartościami przeciętnymi badanej cechy dwóch populacji

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 $X_1 = (X_{11}, X_{12}, \ldots, X_{1n_1})$ oraz
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 Wariancje σ_1^2 i σ_2^2 znane

Badana cecha X populacji ma w dwóch populacjach rozkłady $N(\mu_1, \sigma_1^2)$ i $N(\mu_2, \sigma_2^2)$, odpowiednio, o znanych σ_1^2, σ_2^2 i nieznanych μ_1 i μ_2.

Statystyka testowa

$$U = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \tag{1}$$

przy założeniu prawdziwości hipotezy $H_0 : \mu_1 - \mu_2 = \Delta_0$ ma rozkład $N(0, 1)$.

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 $\bar{x_1} = 121$ 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. Załóżmy, że nie jest spełnione założenie odnośnie normalności populacji. Jak w takiej sytuacji przetestować hipotezę dotyczącą różnicy pomiędzy średnimi?

1.2 Wariancje nieznane, ale równe $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Badana cecha X populacji ma w dwóch populacjach rozkłady $N(\mu_1, \sigma_1^2)$ i $N(\mu_2, \sigma_2^2)$ o nieznanych, ale równych wariancjach σ_1^2, σ_2^2 i nieznanych μ_1 i μ_2.

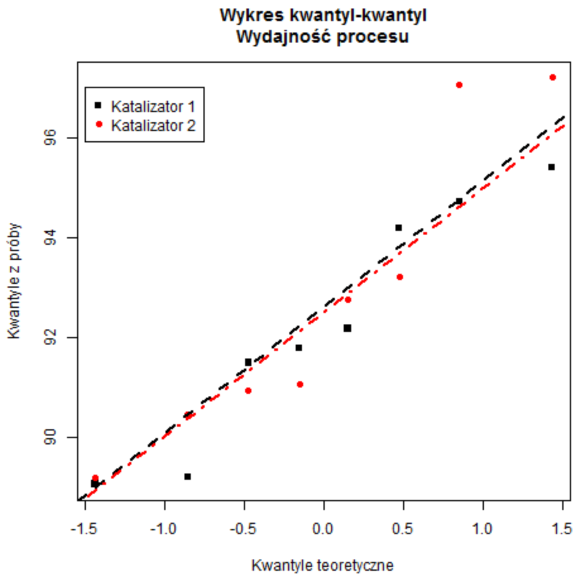
Statystyka testowa

$$T = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - (\mu_1 - \mu_2)}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \text{ gdzie } S_P^2 = \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2} \tag{2}$$

przy założeniu prawdziwości hipotezy $H_0 : \mu_1 - \mu_2 = \Delta_0$ ma rozkład t-Studenta z $n_1 + n_2 - 2$ stopniami swobody.

3. 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. Badanie przeprowadzone zostało w zakładzie pilotażowym a wyniki przedstawiono w poniższej tabeli. Czy jest jakaś różnica między średnią wydajnością procesu na poziomie istotności $\alpha = 0.05$ i przy założeniu równości wariancji?

Nr obserwacji	Katalizator 1	Katalizator 2
1	91.50	89.19
2	94.18	90.95
3	92.18	90.46
4	95.39	93.21
5	91.79	97.19
6	89.07	97.04
7	94.72	91.07
8	89.21	92.75
	$\bar{x}_1 = 92.255$ $s_1 = 2.39$	$\bar{x}_2 = 92.733$ $s_2 = 2.98$



1.3 Wariancje nieznane i różne $\sigma_1^2 \neq \sigma_2^2$

Badana cecha X ma w dwóch populacjach rozkłady $N(\mu_1, \sigma_1^2)$ i $N(\mu_2, \sigma_2^2)$ o nieznanych i niekoniecznie równych wariancjach σ_1^2, σ_2^2 i nieznanych μ_1 i μ_2.

Statystyka testowa

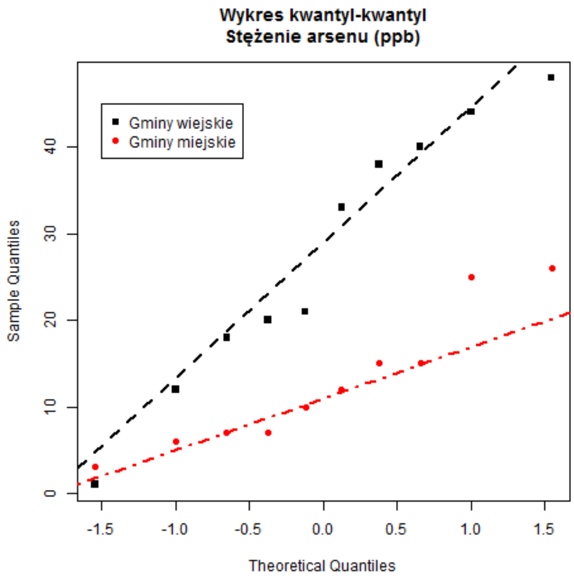
$$T = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{3}$$

przy założeniu prawdziwości hipotezy $H_0 : \mu_1 - \mu_2 = \Delta_0$ ma w przybliżeniu rozkład t-Studenta z v stopniami swobody, gdzie

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}} \tag{4}$$

4. 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. Dane oraz wykres kwantyl-kwantyl dla obu próbek stężenia arsenu przedstawiono poniżej.

Gminy miejskie		Gminy wiejskie	
$\bar{x}_1 = 12.5$		$\bar{x}_2 = 27.5$	
$s_1 = 7.63$		$s_2 = 15.3$	
Phoenix	3	Rimrock	48
Chandler	7	Goodyear	44
Gilbert	25	New River	40
Glendale	10	Apachie Junction	38
Mesa	15	Buckeye	33
Paradise Valley	6	Nogales	21
Peoria	12	Black Canyon City	20
Scottsdale	25	Sedona	12
Tempe	15	Payson	1
Sun City	7	Casa Grande	18



Z wykresu wynika, że o ile założenie normalności wydaje się dość rozsądne, to nachylenia obu prostych różnią się bardzo od siebie, co oznacza, że jest mało prawdopodobne, że wariancje populacji są takie same.

Chcemy określić, czy jest jakaś różnica w średnich stężeniach arsenu pomiędzy gminami miejskimi a gminami wiejskimi. Przeprowadź odpowiedni test.