ĆWICZENIE 5
TESTY STATYSTYCZNE

Cel

Przedstawienie wybranych testów statystycznych, zasad wyboru właściwego testu, przeprowadzenia go oraz interpretacji wyników.

Wprowadzenie teoretyczne

Testem statystycznym nazywamy metodę umożliwiającą weryfikację hipotezy zerowej. W zależności od jej postaci i od rodzaju przeprowadzanego eksperymentu, należy wybrać odpowiedni test. Poniżej przedstawione zostały wybrane testy statystyczne oraz warunki, w jakich należy ich użyć.

Test t jest skrótem od nazwy **test t Studenta**. Testy t stosuje się do sprawdzania hipotez dotyczących wartości oczekiwanej rozkładów normalnych. W zależności od rodzaju danych wyróżnia się: test t dla jednej próby, test t dla prób niepowiązanych (niezależnych) oraz test t dla prób powiązanych (zależnych).

Test t dla jednej próby stosuje się dla próby losowej z populacji generalnej, której cecha ma rozkład normalny o nieznanej wariancji. Hipoteza zerowa jest postaci H_0: $\mu = \mu_0$, hipoteza alternatywna H_1: $\mu \neq \mu_0$, jest to więc test dwustronny. Statystyka testowa jest postaci

$$T = \frac{\sqrt{n}\left(\overline{X} - \mu_0\right)}{S},$$

gdzie $\overline{X}$ jest średnią arytmetyczną badanej cechy, n jest liczbą obserwacji, S jest odchyleniem standardowym cechy z próby. Obszar krytyczny (obszar odrzucenia) testu jest postaci

$$K = \left\{ t : |t| > t_{n-1,1-\frac{\alpha}{2}} \right\},$$

gdzie t jest wartością statystyki T, $t_{n-1,1-\alpha/2}$ oznacza kwanty rozkładu t rzędu $1-\alpha/2$ o $n-1$ stopniach swobody, α jest poziomem istotności.

Test t dla prób niepowiązanych stosuje się dla dwóch niezależnych prób losowych z dwóch populacji generalnych o rozkładach normalnych o tej samej, nieznanej wariancji. Hipoteza zerowa jest postaci H_0: $\mu_1 = \mu_2$, hipoteza alternatywna H_1: $\mu_1 \neq \mu_2$, gdzie μ_1 jest wartością oczekiwaną cechy w pierwszej populacji, μ_2 jest wartością oczekiwaną cechy w drugiej populacji. Statystyka testowa jest postaci

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}},$$

gdzie $\overline{X}_1$, $\overline{X}_2$ są średnimi arytmetycznymi badanej cechy w próbie pierwszej i drugiej, S_1^2, S_2^2 są wariancjami cechy w próbie pierwszej i drugiej, n_1, n_2 są liczebnościami prób. Obszar krytyczny testu jest postaci

$$K = \left\{ t : |t| > t_{n_1+n_2-2,1-\frac{\alpha}{2}} \right\},$$

gdzie t jest wartością statystyki T, $t_{n_1+n_2-2,1-\alpha/2}$ oznacza kwanty rozkładu t rzędu $1-\alpha/2$ o n_1+n_2-2 stopniach swobody, α jest poziomem istotności.

Test t dla prób powiązanych stosuje się dla zestawów par wiązanych *X* i *Y*, gdzie $D = Y - X$ oraz *D* ma rozkład normalny o średniej μ_D i nieznanej wariancji. Hipoteza zerowa jest postaci H_0: $\mu_X = \mu_Y$, hipoteza alternatywna H_1: $\mu_X \neq \mu_Y$. Statystyka testowa jest postaci

$$T = \frac{\sqrt{n}\overline{D}}{S_D},$$

gdzie $\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$, $S_D = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \overline{D}\right)^2}$, *n* jest liczbą obserwacji. Obszar krytyczny testu jest postaci

$$K = \left\{ t : |t| > t_{n-1,1-\frac{\alpha}{2}} \right\},$$

gdzie *t* jest wartością statystyki T, $t_{n-1,1-\alpha/2}$ oznacza kwanty rozkładu t rzędu $1-\alpha/2$ o $n-1$ stopniach swobody, α jest poziomem istotności.

Test U (test Manna-Whitneya) stosuje się do sprawdzenia hipotezy, że dwie niezależne próby pochodzą z tego samego rozkładu typu ciągłego (nie musi to być rozkład normalny). Hipoteza zerowa jest postaci H_0: $F_X = F_Y$, hipoteza alternatywna H_1: $F_X \geq F_Y$ lub H_1: $F_X \geq F_Y$, ale $F_X \neq F_Y$, gdzie F_X jest rozkładem cechy w populacji generalnej, z której pochodzi próba X, F_Y jest rozkładem cechy w populacji generalnej, z której pochodzi próba Y. Statystyka testowa jest postaci

$$U = \min(U_1, U_2),$$

$$U_1 = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} U_{ij} \text{, gdzie } U_{ij} = \begin{cases} 1 & gdy \quad X_i > Y_j \\ 0 & gdy \quad X_i \leq Y_j \end{cases},$$

$$U_2 = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} U_{ij} \text{, gdzie } U_{ij} = \begin{cases} 1 & gdy \quad X_i < Y_j \\ 0 & gdy \quad X_i \geq Y_j \end{cases}$$

Obszar krytyczny testu jest postaci

$$K = \left\{ u : u < U_{n_1,n_2,\frac{\alpha}{2}} \quad \text{lub} \quad u > n_1 n_2 - U_{n_1,n_2,\frac{\alpha}{2}} \right\},$$

gdzie $U_{n_1,n_2,\frac{\alpha}{2}}$ jest kwantylem rzędu $\alpha/2$ rozkładu statystyki U. Jeżeli $n_1 \geq 4$, $n_2 \geq 4$, $n_1+n_2 \geq 20$, to do sprawdzenia hipotezy zerowej stosujemy statystykę testową

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}},$$

która, przy założeniu, że hipoteza zerowa jest prawdziwa, przy $n_1 \to \infty$, $n_2 \to \infty$ ma rozkład N(0,1). Obszar krytyczny jest wówczas postaci

$$K = \left\{ z : |z| > z_{1-\frac{\alpha}{2}} \right\},$$

gdzie $z_{1-\frac{\alpha}{2}}$ jest kwantylem rzędu $1-\alpha/2$ rozkładu normalnego N(0,1).

Test Wilcoxona (test sumy rang Wilcoxona) stosuje się dla zestawów par wiązanych *X* i *Y*, gdzie *D = Y – X* oraz *D* ma rozkład nieznany. Jest to test równoważny testowi U, ale służy do analizy prób zależnych. Hipoteza zerowa jest postaci H_0: $F_X = F_Y$, hipoteza alternatywna H_1: $F_X \geq F_Y$ lub H_1: $F_X \geq F_Y$, ale $F_X \neq F_Y$, gdzie F_X jest rozkładem cechy w populacji generalnej, z której pochodzi próba X, F_Y jest rozkładem cechy w populacji generalnej, z której pochodzi próba Y. Statystyka testowa jest postaci

$$W = \min\left(\sum_{i=1}^{n_1} r_i, \sum_{i=n_1}^{N-n_1} r_i\right),$$

gdzie $\sum_i r_i$ jest sumą rang, zwaną **statystyką Wilcoxona**. Obszar krytyczny testu jest postaci

$$K = \left\{ w : w < W_{n_1,n_2,\frac{\alpha}{2}} \quad \text{lub} \quad w > n_1 n_2 - W_{n_1,n_2,\frac{\alpha}{2}} \right\},$$

gdzie $W_{n_1,n_2,\frac{\alpha}{2}}$ jest kwantylem rzędu *α/2* rozkładu statystyki W. Jeżeli liczba analizowanych par przekracza 15, wówczas do sprawdzenia hipotezy zerowej stosujemy statystykę testową

$$Z = \frac{W - \frac{N(N+1)}{4}}{\sqrt{\frac{N(N+1)(2N+1)}{24}}},$$

która, przy założeniu, że hipoteza zerowa jest prawdziwa, przy $n_1 \to \infty$, $n_2 \to \infty$ ma rozkład N(0,1). Obszar krytyczny jest wówczas postaci

$$K = \left\{ z : |z| > z_{1-\frac{\alpha}{2}} \right\},$$

gdzie $z_{1-\frac{\alpha}{2}}$ jest kwantylem rzędu *1-α/2* rozkładu normalnego N(0,1).

Test Kruskala – Wallisa stosowany jest do porównywania wielu populacji na podstawie niezależnych prób pochodzących po jednej z każdej populacji, która ma rozkład ciągły. Służy do sprawdzenia hipotezy, że niezależnie pobrane próby pochodzą z tej samej populacji. Test ten jest uogólnieniem testu Manna – Whitneya. Jeżeli mamy *k* prób niezależnych, to hipoteza zerowa jest postaci H_0: $F_1 = F_2 = \ldots = F_k$, hipoteza alternatywna H_1: istnieje co najmniej jedna para *i*, *j* (*i ≠ j*, *i,j = 1,2,…,k)* taka, że $F_i \leq F_j$ lub $F_i \geq F_j$, ale $F_i \neq F_j$. Statystyka testowa jest postaci

$$T = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1),$$

gdzie $n = \sum_{i=1}^{k} n_i$, n_i jest liczebnością i-tej próby, R_i jest sumą rang dla i-tej próby. Przy prawdziwości hipotezy zerowej, rozkład statystyki T, przy $n_i \to \infty$ jest rozkładem χ^2 z *k-1* stopniami swobody. Obszar krytyczny jest postaci

$$K = \{ t : t > h_{k,n,1-\alpha} \},$$

gdzie *α* jest poziomem istotności, $h_{k,n,1-\alpha}$ jest kwantylem rzędu *1- α* rozkładu χ^2 o *k-1* stopniach swobody.

Testy χ^2 (testy chi – kwadrat) służą do sprawdzania hipotezy o jednakowym rozkładzie zmiennych losowych typu dyskretnego. Ze względu na ilość kryteriów klasyfikacji danych wyróżnia się test χ^2 dla klasyfikacji jednoczynnikowej oraz test χ^2 dla klasyfikacji dwuczynnikowej. Analizowane dane są najczęściej w postaci częstości wystąpień obserwacji należącej do danej kategorii.

Test χ^2 dla klasyfikacji jednoczynnikowej odnosi się do danych podzielonych ze względu na jeden czynnik. Hipoteza zerowa jest postaci H_0: częstości wystąpień badanej cechy są takie same we wszystkich grupach (podzielonych ze względu na jeden czynnik), hipoteza alternatywna H_1: przynajmniej w jednej parze grup częstotliwości wystąpienia badanej cechy różnią się między sobą. Statystyka testowa jest postaci

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - \overline{n})^2}{\overline{n}},$$

gdzie k jest liczbą grup, n_i jest liczbą elementów i-tej grupy, $\overline{n}$ jest średnią liczbą elementów wszystkich grup. Przy prawdziwości hipotezy zerowej, statystyka testowa ma rozkład χ^2 o *k-1* stopniach swobody. Obszar odrzucenia jest postaci

$$K = \{t : t > \chi^2_{k-1,1-\alpha}\},$$

gdzie α jest poziomem istotności, $\chi^2_{k-1,1-\alpha}$ jest kwantylem rzędu *1- α* rozkładu χ^2 o *k-1* stopniach swobody.

Test χ^2 dla klasyfikacji dwuczynnikowej odnosi się do danych podzielonych ze względu na dwa czynniki. Hipoteza zerowa jest postaci H_0: częstości wystąpień badanej cechy są takie same we wszystkich grupach (podzielonych ze względu na dwa czynniki), hipoteza alternatywna H_1: przynajmniej w jednej parze grup częstotliwości wystąpienia badanej cechy różnią się między sobą. Statystyka testowa jest postaci

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\left(n_{ij} - \frac{\sum_{i=1}^{k} n_{i.} \sum_{j=1}^{m} n_{.j}}{n} \right)^2}{\frac{\sum_{i=1}^{k} n_{i.} \sum_{j=1}^{m} n_{.j}}{n}},$$

gdzie k jest liczbą grup otrzymanych przez klasyfikację ze względu na pierwszy czynnik, m jest liczbą grup otrzymanych przez klasyfikację ze względu na drugi czynnik, n_{ij} jest liczbą elementów i-tej i j-tej grupy. Przy prawdziwości hipotezy zerowej, statystyka testowa ma rozkład χ^2 o *k-1,m-1* stopniach swobody. Obszar odrzucenia jest postaci

$$K = \{t : t > \chi^2_{k-1,m-1,1-\alpha}\},$$

gdzie α jest poziomem istotności, $\chi^2_{k-1,m-1,1-\alpha}$ jest kwantylem rzędu *1- α* rozkładu χ^2 o *k-1,m-1* stopniach swobody.

Test F (test F Snedecora) stosuje się do testowania istotności zależności zmiennej zależnej *y* od zmiennej niezależnej *x*, opisanej wzorem $y = \beta_0 + \beta_1 x$. Hipoteza zerowa jest postaci H_0: zmienna *y* nie zależy od zmiennej *x*, hipoteza alternatywna H_1: zmienna *y* zależy od zmiennej *x*. Statystyka testowa jest postaci

$$F = \frac{\dfrac{\sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2}{k-1}}{\dfrac{\sum_{i=1}^{n} \left(\hat{y}_i - y\right)^2}{n-k}},$$

gdzie *n* jest liczbą obserwacji, $\hat{y}_i$ oznacza prognozę zmiennej zależnej otrzymaną z wzoru $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x$, *k* jest liczbą współczynników (β) w równaniu. Przy prawdziwości hipotezy zerowej, statystyka testowa ma rozkład $F_{k-1,n-k}$. Obszar odrzucenia jest postaci $K = \{F : F > F_{k-1,n-k,1-\alpha}\}$, gdzie $F_{k-1,n-k,1-\alpha}$ oznacza kwanty rzędu $1-\alpha$ rozkładu F o $k-1, n-k$ stopniach swobody.

Zadanie do wykonania

1. Odłowiono i zważono 8 srok w Białymstoku i 6 srok w Gliwicach. Podać nazwy wszystkich znanych testów dla ustalenia, czy sroki z tych dwóch miast różnią się ciężarem wiedząc, że ciężar srok ma rozkład normalny. Zapisać postawione hipotezy.
2. Pewien włościanin ze Słomnik kupił od Ukraińca z Żytomierza 400 jaj przepiórek japońskich, płacąc za nie po 45 groszy. Ukrainiec gwarantował średni wylęg jaj w 80%. Jakim testem można sprawdzić uczciwość Ukraińca? Ilość jaj, z których wylęgły się pisklęta, ma rozkład normalny. Zapisać postawione hipotezy.
3. Pewien badacz miał za zadanie ustalić, jakie jest stężenie metali ciężkich w tkankach roślin rosnących wzdłuż 80 km autostrady idącej z północy na południe. W tym celu postanowił pobrać równoliczne próby w 20 różnych miejscach. Jakiego testu należy użyć, aby porównać stężenie metali ciężkich w próbach zebranych na pierwszych 20 km i pozostałych 60 km autostrady? Jakiego testu należy użyć, aby porównać stężenie metali ciężkich w zebranych próbach? Zapisać postawione hipotezy.
4. Na 10 stanowiskach badawczych w okolicy budowy przyszłej zapory określono liczbę wszystkich gatunków chrząszczy, a następnie badania te powtórzono po napełnieniu zbiornika wodą. Jakim testem można sprawdzić, czy budowa zapory i inne zmiany, które w tym czasie nastąpiły, spowodowały istotne zmniejszenie lub zwiększenie liczby gatunków chrząszczy? Zapisać postawione hipotezy.
5. Aby ustalić zmniejszenie się ilości witaminy C w ziemniakach z upływem czasu, wybrano 8 ziemniaków zaraz po zbiorze i określono w nich ilość witaminy C w mg na 100 g ziemniaka. W miesiąc po zbiorze wybrano losowo 7 ziemniaków i zbadano je w taki sam sposób. Jakim testem można ustalić czy po miesiącu ilość witaminy C w ziemniakach zmniejsza się. Zapisać postawione hipotezy.
6. Podać przykłady eksperymentów, w którym zastosowanie znalazłby każdy z opisanych powyżej testów.

Źródła: Łomnicki A. „Wprowadzenie do statystyki dla przyrodników” PWN, Warszawa 2007,
Magiera R. „Modele i metody statystyki matematycznej” GiS, Wrocław 2002