

Rozważmy teraz dokładniej zagadnienie szacowania wartości oczekiwanej m zmiennej losowej X o rozkładzie normalnym $N(m, \sigma)$, w którym odchylenie standardowe σ jest znane.

Niech $X_1, X_2, \dots, X_n$ oznacza próbkę dla zmiennej losowej X oraz niech średnia z próbki, tj.

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n},$$

będzie estymatorem wartości oczekiwanej m . Zmienna $\bar{X}_n$ jest zmienną losową o rozkładzie normalnym $N\left(m, \frac{\sigma}{\sqrt{n}}\right)$. W twierdzeniu 5.2 udowodniliśmy bowiem, że $E(\bar{X}_n) = E(X) = m$, a ponadto mamy

$$\begin{aligned} D^2(\bar{X}_n) &= D^2\left(\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right) = \frac{1}{n^2} D^2(X_1 + X_2 + \dots + X_n) \\ &= \frac{1}{n^2} (D^2(X_1) + D^2(X_2) + \dots + D^2(X_n)) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}, \end{aligned}$$

czyli $D(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$.

Niech ε oznacza ustaloną liczbę dodatnią i obliczmy prawdopodobieństwo tego, że różnica pomiędzy naszym oszacowaniem $\bar{X}_n$ i szacowaną wartością m będzie większa od ε . Mamy

$$P\left(|\bar{X}_n - m| > \varepsilon\right) = P\left(\left|\frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}\right| > \frac{\varepsilon}{\sigma} \sqrt{n}\right).$$

Ponieważ zmienna losowa standaryzowana $\frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}$ ma rozkład normalny $N(0, 1)$, więc

$$\begin{aligned} P\left(\left|\frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}\right| > \frac{\varepsilon}{\sigma} \sqrt{n}\right) &= 1 - P\left(\left|\frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}\right| \leq \frac{\varepsilon}{\sigma} \sqrt{n}\right) \\ &= 1 - P\left(-\frac{\varepsilon}{\sigma} \sqrt{n} \leq \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\varepsilon}{\sigma} \sqrt{n}\right) = 1 - \left(\Phi\left(\frac{\varepsilon}{\sigma} \sqrt{n}\right) - \Phi\left(-\frac{\varepsilon}{\sigma} \sqrt{n}\right)\right) \\ &= 1 - \Phi\left(\frac{\varepsilon}{\sigma} \sqrt{n}\right) + \left(1 - \Phi\left(\frac{\varepsilon}{\sigma} \sqrt{n}\right)\right) = 2\left[1 - \Phi\left(\frac{\varepsilon}{\sigma} \sqrt{n}\right)\right], \end{aligned}$$

gdzie $\Phi(x)$ oznacza wartość dystrybuanty rozkładu normalnego $N(0, 1)$ w punkcie x . Z powyższego wzoru wynika, że prawdopodobieństwo popełnienia błędu ustalonej wielkości, którą oznaczyliśmy przez ε , jest malejącą funkcją liczebności próbki, tj. wielkości n i rosnącą funkcją odchylenia standardowego σ .

W naukach eksperymentalnych przyjęto błąd metody pomiarowej charakteryzować za pomocą odchylenia standardowego σ i z dwóch nieobciążonych metod pomiarowych za dokładniejszą uważać tę, dla której odchylenie standardowe σ jest mniejsze.

Przykład 5.3. Za pomocą pewnej metody wykonano pomiary przyspieszenia ziemskiego g . Otrzymano następujące wyniki (w m/s^2):

9,806633, 9,806682, 9,806673, 9,806665 i 9,806646.

Oszacowaniem stałej g jest średnia z tych pomiarów, która jest równa 9,806660. Zakładając, że błąd pomiaru jest zmienną losową o rozkładzie normalnym $N(0, 0,00002)$, tzn. że metoda jest nieobciążona, a jej dokładność charakteryzuje się odchyleniem standardowym $\sigma = 0,00002$, obliczymy prawdopodobieństwo, że błąd oszacowania otrzymany przez uśrednienie pięciu wyników pomiarów przekroczy $0,00001 \text{ m/s}^2$.

Oznaczając przez $\bar{X}_5$ średnią z pomiarów, a przez g szacowaną wartość przyspieszenia ziemskiego, z podanego wzoru mamy

$$P(|\bar{X}_5 - g| > 0,00001) = 2 \left[1 - \Phi \left(\frac{0,00001}{0,00002} \sqrt{5} \right) \right] \approx 2(1 - \Phi(1,095)) \approx 0,2736.$$

Często stawiamy pytanie: ile należy wykonać pomiarów, aby oszacowanie interesującej nas wielkości za pomocą średniej z wyników tych pomiarów z prawdopodobieństwem co najwyżej równym p miało błąd nie większy niż ε . Należy zatem wyznaczyć wartość n tak, by

$$P(|\bar{X}_n - m| \leq \varepsilon) \geq p.$$

Na podstawie poprzedniego wzoru oznacza to, że wartość n powinna spełniać warunek

$$2 \left[1 - \Phi \left(\frac{\varepsilon}{\sigma} \sqrt{n} \right) \right] \leq 1 - p.$$

Przykład 5.4. Przy założeniach z poprzedniego przykładu obliczmy, ile należy wykonać pomiarów przyspieszenia ziemskiego metodą, w której odchylenie standardowe σ jest równe $0,00002 \text{ m/s}^2$, aby błąd wykonanego oszacowania z prawdopodobieństwem co najmniej 0,99 nie przekroczył $\varepsilon = 0,00001 \text{ m/s}^2$.

Na podstawie podanego wzoru mamy

$$2 \left[1 - \Phi \left(\frac{0,00001}{0,00002} \sqrt{n} \right) \right] \leq 0,01.$$

Stąd

$$\Phi \left(\frac{\sqrt{n}}{2} \right) \geq 0,995.$$

Korzystając z tablic rozkładu $N(0, 1)$ rozwiązujemy względem x równanie

$$\Phi(x) = 0,995$$

i otrzymujemy $x \approx 2,576$. Ponieważ funkcja Φ jest rosnąca, więc

$$\frac{\sqrt{n}}{2} \geq 2,576,$$

czyli $n \geq 27$.

Wiemy, że jeśli zmienna losowa X ma rozkład normalny $N(m, \sigma)$, to średnia $\bar{X}_n$ z próbki $X_1, X_2, \dots, X_n$ jest zmienną losową o rozkładzie normalnym $N\left(m, \frac{\sigma}{\sqrt{n}}\right)$, a zmienna losowa postaci $\frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}}$ ma rozkład $N(0, 1)$. Jeśli mamy ustaloną liczbę Q z przedziału $(0, 1)$, to można

wyznaczyć taką liczbę t_Q , by

$$P\left(\left|\frac{\bar{X}_n - m}{\sigma} \sqrt{n}\right| \leq t_Q\right) = Q.$$

Ponieważ

$$\begin{aligned} P\left(\left|\frac{\bar{X}_n - m}{\sigma} \sqrt{n}\right| \leq t_Q\right) &= P\left(-t_Q \leq \frac{\bar{X}_n - m}{\sigma} \sqrt{n} \leq t_Q\right) \\ &= \Phi(t_Q) - \Phi(-t_Q) = \Phi(t_Q) - (1 - \Phi(t_Q)) = 2\Phi(t_Q) - 1, \end{aligned}$$

więc liczbę t_Q otrzymuje się przez rozwiązanie równania

$$2\Phi(t_Q) - 1 = Q.$$

Wybierając dostatecznie dużą wartość Q możemy być prawie pewni, że zdarzenie

$$\frac{\bar{X}_n - m}{\sigma} \sqrt{n} \leq t_Q$$

nastąpi, czyli, że szacowana wartość oczekiwana m zmiennej losowej X będzie znajdowała się w przedziale

$$\left[\bar{X}_n - t_Q \frac{\sigma}{\sqrt{n}}, \bar{X}_n + t_Q \frac{\sigma}{\sqrt{n}}\right].$$

Przedział ten nazywa się *przedziałem ufności* dla parametru m , a liczbę Q – *poziomem ufności*. W praktyce przyjmuje się zwykle $Q = 0,95$ lub $Q = 0,99$.

Przykład 5.5. Wyznaczyć przedział ufności dla $Q = 0,95$ i danych z przykładu 5.3.

W przykładzie tym dla próbki o liczebności $n = 5$ otrzymaliśmy średnią $\bar{X}_5 = 0,80666$. Odchylenie standardowe metody pomiarowej było równe $\sigma = 0,00002$. Po podstawieniu do powyższego wzoru otrzymujemy przedział ufności $[9,806642; 9,806678]$ z poziomem ufności $Q = 0,95$.

Dotychczas zakładaliśmy, że odchylenie standardowe σ w odpowiednich rozkładach prawdopodobieństwa jest znane. W praktyce tak nie jest i powstają następujące pytania:

- jak należy szacować wartość oczekiwaną i jak obliczać dla niej przedział ufności, gdy z powodu braku znajomości parametru σ nie można określić przedziału ufności,
- jak oszacować parametr σ na podstawie próbkę?

Niech zmienna losowa X ma pewien rozkład z nieznaną wariancją σ^2 . Niech zmienne $X_1, X_2, \dots, X_n$ stanowią próbkę dla tej zmiennej, a $\bar{X}_n$ niech oznacza średnią z tej próbki. Wyrażenie

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

będziemy nazywać *wariancją z próbki*.

Twierdzenie 5.4. *Wariancja z próbki jest nieobciążonym i zgodnym estymatorem wariancji zmiennej losowej X , o ile ta ostatnia istnieje.*

Dowód. Rozważmy zmienną losową $U_i = (X_i - \bar{X}_n)^2$. Dla zmiennej losowej U_1 mamy

$$\begin{aligned} U_1 &= \left(X_1 - \frac{1}{n} \sum_{j=1}^n X_j \right)^2 = \left[(X_1 - m) - \frac{1}{n} \sum_{j=1}^n (X_j - m) \right]^2 \\ &= \left[\left(1 - \frac{1}{n} \right) (X_1 - m) - \frac{1}{n} \sum_{j=2}^n (X_j - m) \right]^2 \\ &= \left(1 - \frac{1}{n} \right)^2 (X_1 - m)^2 + \frac{1}{n^2} \sum_{j=2}^n (X_j - m)^2 - 2 \left(1 - \frac{1}{n} \right) \frac{1}{n} \sum_{j=2}^n (X_1 - m)(X_j - m) \\ &\quad + \frac{1}{n^2} \sum_{j=2}^n \sum_{\substack{k=2 \\ k \neq j}}^n (X_j - m)(X_k - m). \end{aligned}$$

Stąd

$$\begin{aligned} E(U_1) &= \left(1 - \frac{1}{n} \right)^2 E((X_1 - m)^2) + \frac{1}{n^2} \sum_{j=2}^n E((X_j - m)^2) \\ &\quad - 2 \left(1 - \frac{1}{n} \right) \frac{1}{n} \sum_{j=2}^n E((X_1 - m)(X_j - m)) + \frac{1}{n^2} \sum_{j=2}^n \sum_{\substack{k=2 \\ k \neq j}}^n E((X_j - m)(X_k - m)). \end{aligned}$$

Ponieważ zmienne $X_1, X_2, \dots, X_n$ są próbką, więc dla każdej wartości $j = 1, 2, \dots, n$ mamy

$$E((X_j - m)^2) = \sigma^2.$$

Z kolei z niezależności zmiennych losowych $X_1, X_2, \dots, X_n$ wynika, że

$$E((X_j - m)(X_k - m)) = 0 \quad \text{dla } j \neq k.$$

Otrzymujemy zatem

$$E(U_1) = \left(1 - \frac{1}{n} \right)^2 \sigma^2 + \frac{1}{n^2} \sum_{j=2}^n \sigma^2 = \frac{n-1}{n} \sigma^2.$$

Taki sam wzór otrzymamy dla zmiennych $U_2, U_3, \dots, U_n$, czyli

$$E(U_i) = \frac{n-1}{n} \sigma^2 \quad \text{dla } i = 1, 2, \dots, n.$$

Stąd

$$E(S_n^2) = \frac{1}{n-1} \sum_{i=1}^n E(U_i) = \frac{1}{n-1} \sum_{i=1}^n \frac{n-1}{n} \sigma^2 = \sigma^2,$$

czyli estymator S_n^2 jest nieobciążony. Dowód zgodności estymatora S_n^2 jest trudniejszy, bo zmienne losowe U_i nie są niezależne (pomijamy go). ■

Przykład 5.6. Dla danych z poprzednich przykładów, które dotyczyły przyciągania ziemskiego ze średnią z próbki równą 9,80666, oszacowaniem wariancji jest

$$\begin{aligned} S_5^2 = \frac{1}{4} & [(9,806633 - 9,806660)^2 + (9,806682 - 9,806660)^2 \\ & + (9,806673 - 9,806660)^2 + (9,806665 - 9,806660)^2 \\ & + (9,806646 - 9,806660)^2] \approx 4 \cdot 10^{-10}. \end{aligned}$$

W celu konstrukcji przedziału ufności dla wariancji σ^2 zmiennej losowej X o rozkładzie normalnym $N(m, \sigma)$ należy znać rozkład estymatora S_n^2 tej wariancji.

Twierdzenie 5.5. Niech zmienna losowa X ma rozkład normalny $N(m, \sigma)$ i niech $X_1, X_2, \dots, X_n$ oznacza n -elementową próbkę dla tej zmiennej, a S_n^2 – wariancję z próbki. Wówczas zmienna losowa

$$V_n = \frac{(n-1)S_n^2}{\sigma^2}$$

ma rozkład o dystrybuancie

$$F_n(v) = \begin{cases} 0 & \text{dla } v < 0, \\ c_{n-1} \int_0^v x^{\frac{n-2}{2}} \cdot \exp\left(-\frac{x}{2}\right) dx & \text{dla } v \geq 0, \end{cases}$$

gdzie c_{n-1} oznacza stałą.

Dowód pomijamy. ■

Definicja 5.4. Rozkład o dystrybuancie

$$G_k(v) = \begin{cases} 0 & \text{dla } v < 0, \\ c_k \int_0^v x^{\frac{k-1}{2}} \exp\left(-\frac{x}{2}\right) dx & \text{dla } v \geq 0, \end{cases}$$

tzn. rozkład o gęstości prawdopodobieństwa

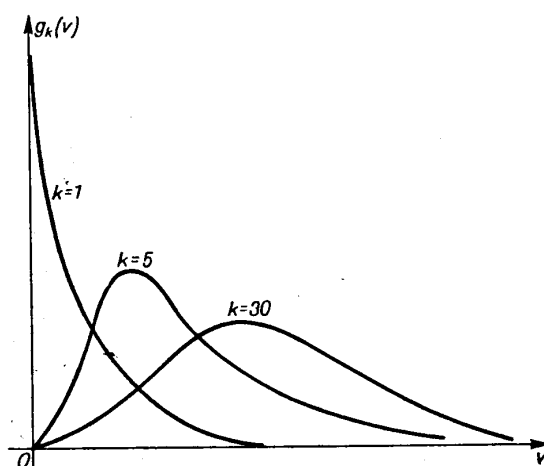
$$g_k(v) = \begin{cases} 0 & \text{dla } v < 0, \\ c_k v^{\frac{k-1}{2}} \exp\left(-\frac{v}{2}\right) & \text{dla } v \geq 0, \end{cases}$$

gdzie c_k oznacza stałą nazywa się *rozkładem chi-kwadrat* i jest oznaczany przez χ^2 . Rozkład ten zależy od parametru k , który nazywa się *liczbą stopni swobody*.

Gęstość rozkładu chi-kwadrat dla różnych wartości parametru k przedstawiono na poniższym rysunku. Przy rozkładzie chi-kwadrat korzysta się zazwyczaj z tablic. Na ogół tablice te podają wartości v , które są rozwiązaniami równania

$$G_k(v) = \alpha$$

dla różnych liczb α oraz k .



Na podstawie powyższej definicji twierdzenie 5.5 można wypowiedzieć następująco: przy przyjętych założeniach zmienna losowa $V_n = \frac{(n-1)S_n^2}{\sigma^2}$ ma rozkład chi-kwadrat z $n-1$ stopniami swobody.

W celu wyznaczenia przedziału ufności dla wariancji ustalmy poziom ufności Q . Ponieważ znamy rozkład zmiennej losowej V_n , więc potrafimy wyznaczyć dwie liczby v'_Q i v''_Q , by

$$P\left(v'_Q \leq \frac{(n-1)S_n^2}{\sigma^2} \leq v''_Q\right) = Q.$$

Ale nierówności

$$v'_Q \leq \frac{(n-1)S_n^2}{\sigma^2} \leq v''_Q$$

są równoważne nierównościom

$$\frac{1}{v''_Q} \leq \frac{\sigma^2}{(n-1)S_n^2} \leq \frac{1}{v'_Q},$$

czyli

$$\frac{(n-1)S_n^2}{v_Q''} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{v_Q'}.$$

Zatem przedział

$$\left[\frac{(n-1)S_n^2}{v_Q''}, \frac{(n-1)S_n^2}{v_Q'} \right]$$

jest przedziałem ufności dla wariancji σ^2 na poziomie ufności Q . Stąd przedziałem ufności dla odchylenia standardowego σ jest

$$\left[\frac{\sqrt{n-1} \cdot S_n}{\sqrt{v_Q''}}, \frac{\sqrt{n-1} \cdot S_n}{\sqrt{v_Q'}} \right].$$

Jeżeli w zmiennej losowej $\frac{\bar{X}_n - m}{\sigma} \sqrt{n}$ zastąpimy odchylenie standardowe σ oszacowaniem tego odchylenia, tj.

$$S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2},$$

to otrzymamy nową zmienną losową

$$T_{n-1} = \frac{\bar{X}_n - m}{S_n} \sqrt{n},$$

dla której mamy

Twierdzenie 5.6. Niech zmienna losowa X ma rozkład normalny $N(m, \sigma)$, $X_1, X_2, \dots, X_n$ oznacza n -elementową próbkę dla tej zmiennej, a $\bar{X}_n$ i S_n^2 – średnią i wariancję z tej próbki. Wówczas zmienna losowa T_{n-1} ma rozkład o dystrybucie

$$H_{n-1}(t) = c_{n-1} \int_{-\infty}^t \frac{dx}{\left(1 + \frac{x^2}{n-1}\right)^{\frac{n}{2}}},$$

gdzie c_{n-1} oznacza pewną stałą.

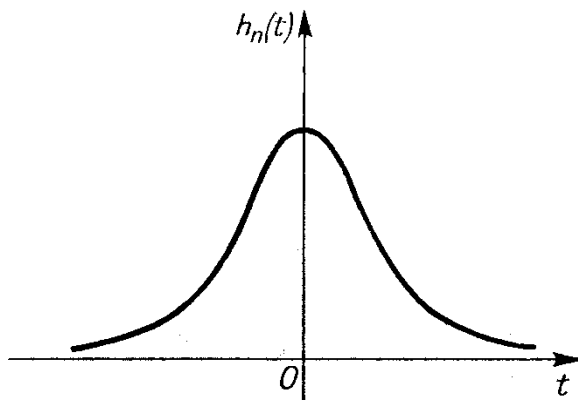
Dowód pomijamy. ■

Definicja 5.5. Rozkład z dystrybucją

$$H_n(t) = c_n \int_{-\infty}^t \frac{dx}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}$$

gdzie c_n oznacza stałą, nazywamy *rozkładem t -Studenta*. Liczba n jest parametrem tego rozkładu i nazywa się *liczbą stopni swobody*.

Gęstość rozkładu t -Studenta jest przedstawiona na poniższym rysunku.



Korzystając z definicji 5.5 twierdzenie 5.6 można sformułować następująco: przy przyjętych założeniach zmienna losowa T_{n-1} ma rozkład t -Studenta z $n - 1$ stopniami swobody.

Teoria przedziałów ufności dotycząca zmiennej losowej T_{n-1} jest identyczna z teorią przedziałów ufności związanych ze zmienną losową $\frac{\bar{X}_n - m}{\sigma} \sqrt{n}$ o rozkładzie normalnym $N(0, 1)$. Jedyna różnica polega na posługiwaniu się innymi tablicami. W przypadku rozkładu t -Studenta posługujemy się tablicami będącymi rozwiązaniami równania

$$H_n(t) = \alpha$$

dla wybranych wartości α oraz n .

Zadania

1. Błąd wskazania wysokościomierza w samolocie jest zmienną losową o rozkładzie normalnym z wartością oczekiwaną $m = 0$ i z odchyleniem standardowym $\sigma = 15$ m. Jakie jest prawdopodobieństwo, że błąd oceny wysokości będzie większy niż 30 m? Ile należy mieć takich wysokościomierzy w samolocie, żeby błąd oceny wysokości za pomocą średniej ich wskazań z prawdopodobieństwem 0,99 nie przekraczał 30 m?
2. Zbadano czas świecenia się czterech żarówek wylosowanych z partii żarówek wyprodukowanych w pewnym zakładzie. Otrzymano następujące wyniki (w godzinach): 1240, 1300, 1150 i 1210. Zakładając, że czas świecenia się żarówki jest zmienną losową o rozkładzie normalnym z odchyleniem standardowym równym 70 godzin, wyznaczyć jednostronny przedział ufności na poziomie ufności 0,90.
3. Z dużej partii słupków betonowych wybrano próbkę losową o liczności 50 sztuk. Średnia wytrzymałość na ściskanie osiowe obliczona w tej próbce wyniosła 248,31 kG/cm². Oszacować średnią wytrzymałość słupków. Zakładając, że cecha ma rozkład normalny oszacować wartość oczekiwaną tej zmiennej za pomocą przedziału ufności na poziomie 0,99.

4. Jaki będzie przedział ufności dla danych z poprzedniego zadania, gdy poziom ufności zmniejszy do 0,95?
5. W celu wyznaczenia błędu standardowego pewnego przyrządu pomiarowego (odchylenia standardowego tego przyrządu) dokonano pięciu pomiarów pewnej wielkości i otrzymano następujące wyniki: 8,15; 8,20; 8,04; 8,14 i 8,22. Oszacować wariancję wskazań przyrządu i obliczyć przedział ufności dla odchylenia standardowego tego przyrządu przy poziomie ufności 0,90.
6. Celem sprawdzenia dokładności wskazań pewnego przyrządu pomiarowego dokonano 10 pomiarów tej samej wielkości fizycznej i otrzymano następujące wyniki: 9,01; 9,00; 9,02; 8,99; 8,98; 9,00; 9,00; 9,01; 8,99; i 9,00. Oszacować wariancję wskazań przyrządu i podać przedział ufności dla jego błędu (odchylenia standardowego) na poziomie ufności równym 0,99 zakładając, że wskazania tego przyrządu mają rozkład normalny.
7. Za pomocą pięciu niezależnych pomiarów szacowano średnicę kątową planety Neptun i otrzymano następujące wyniki (w sekundach łuku): 2,31; 2,25; 2,28; 2,30 i 2,27. Zakładając, że wyniki pomiarów są błędem losowym o rozkładzie normalnym oszacować średnicę kątową Neptuna i podać przedział ufności dla tej średnicy na poziomie ufności 0,90. Jak zmieni się oszacowanie i przedział ufności, gdy obserwacje uzupełnimy nowym wynikiem pomiaru równym 2,30?

5.4. Weryfikacja hipotez statystycznych

Niech będzie dana pewna zmienna losowa X , której rozkład albo pewne jego parametry nie są znane. Odnośnie tego rozkładu lub tych parametrów formułujemy pewne przypuszczenie. Każde takie przypuszczenie będziemy nazywać *hipotezą statystyczną* i oznaczać przez H .

W celu weryfikacji hipotezy H obserwujemy próbkę $X_1, X_2, \dots, X_n$ i definiujemy pewną funkcję δ na tej próbce. Niech Ω oznacza zbiór wszystkich wartości tej funkcji.

Ustalamy pewną małą liczbę α i konstruujemy taki podzbiór Ω_k zbioru Ω , aby prawdopodobieństwo zdarzenia polegającego na tym, że $\delta \in \Omega_k$, gdy hipoteza H jest prawdziwa, było nie większe od α . Jeżeli w wyniku próby zaobserwujemy taką wartość δ , która należy do zbioru Ω_k , to weryfikowaną hipotezę H odrzucamy jako nieprawdziwą, a w przeciwnym przypadku stwierdzamy, że wynik doświadczenia nie przeczy tej hipotezie.

Liczbę α nazywamy *poziomem istotności testu*, zbiór Ω_k – *zbiorem krytycznym testu*, a opisane postępowanie – *testem hipotezy H* .

Przykład 5.7. Wykonano nowy przyrząd do mierzenia wysokości drzew. Wiadomo, że błąd pomiaru tym przyrządem jest zmienną losową o rozkładzie normalnym z odchyleniem standardowym 0,1 m. W celu sprawdzenia, czy pomiary przyrządem nie są obarczone błędem systematycznym (tzn. czy błąd pomiaru ma wartość oczekiwaną równą 0) wykonano pięć pomiarów wzorcowego drzewa o znanej wysokości 12 m i otrzymano następujące wyniki: 12,32; 11,90; 11,93; 12,01 i 12,19. Średnia tych wyników jest równa $\bar{X}_5 = 12,07$, a więc błąd oszacowania wysokości wynosi 0,07 m.

Czy wynik ten upoważnia nas do stwierdzenia, że przyrząd wykazuje pewien błąd systematyczny, czy też zaobserwowane odchylenie od znanej wartości 12 m można usprawiedliwić losowymi wahaniami przyrządu?

Stawiamy hipotezę: przyrząd nie jest obciążony błędem systematycznym.

Naszymi danymi są:

- zmienna losowa X o rozkładzie $N(m, 0,1)$, gdzie wartość oczekiwana m jest nieznana, a hipoteza mówi, że $m = 12$, czyli $H: m = 12$,
- próbka o liczebności $n = 5$,
- funkcja δ , za którą przyjmujemy $\bar{X}_n$.

Wyznaczamy granicę, od której począwszy będziemy uważali obserwację $\bar{X}_n$ za sprzeczną z hipotezą, że rozpatrywana zmienna losowa ma rozkład $N(12, 0,1)$. Jako granicę weźmiemy taką liczbę λ , aby prawdopodobieństwo tego, że zaobserwujemy różnicę $|\bar{X}_n - m|$ większą niż λ było małe, np. $\alpha = 0,05$ (jest to przyjęty poziom istotności testu). Zbiorem krytycznym jest zbiór

$$\Omega_k = \{\delta: |\delta - 12| > \lambda\},$$

przy czym w naszym przypadku $\delta = \bar{X}_5$. Mamy

$$P(|\bar{X}_n - m| > \lambda) = P\left(\left|\frac{\bar{X}_n - m}{\sigma} \sqrt{n}\right| > \frac{\lambda}{\sigma} \sqrt{n}\right).$$

Zmienna losowa $U = \frac{\bar{X}_n - m}{\sigma} \sqrt{n}$ ma rozkład normalny $N(0, 1)$. Z tablic wyznaczamy liczbę u_α tak, by

$$P(|U| > u_\alpha) = \alpha = 0,05.$$

W tym celu wykonujemy kolejno następujące przekształcenia:

$$\begin{aligned} P(|U| > u_\alpha) &= 1 - P(|U| \leq u_\alpha) = 1 - P(-u_\alpha \leq U \leq u_\alpha) \\ &= 1 - (\Phi(u_\alpha) - \Phi(-u_\alpha)) = 2 - 2\Phi(u_\alpha). \end{aligned}$$

Stąd wynika, że

$$2 - 2\Phi(u_\alpha) = 0,05,$$

a więc $\Phi(u_\alpha) = 0,975$ i z tablic odczytujemy, że $u_\alpha = u_{0,05} = 1,96$. Mamy zatem

$$P\left(\left|\frac{\bar{X}_n - m}{\sigma} \sqrt{n}\right| > 1,96\right) = 0,05$$

czyli $\frac{\lambda}{\sigma} \sqrt{n} = 1,96$, skąd wynika, że $\lambda = 1,96 \frac{\sigma}{\sqrt{n}}$ i ponieważ $\sigma = 0,1$ oraz $n = 5$,

więc mamy $\lambda \approx 0,088$. Zatem jeśli $|\bar{X}_5 - 12| > 0,088$, to należy uznać, że przyrząd wykazuje błąd systematyczny. Ponieważ dla $n = 5$ mamy $\bar{X}_5 = 12,07$, więc $|\bar{X}_5 - 12| = 0,07$, co nie pozwala na dyskwalifikację przyrządu jako obciążonego błędem systematycznym (zaobserwowany wynik nie jest elementem zbioru krytycznego). Brak zatem podstaw do zakwestionowania hipotezy.

Powstają dwa problemy:

- skoro wartość α jest prawdopodobieństwem popełnienia błędu, to dlaczego nie wybrano mniejszej liczby α ,
- jak należy konstruować zbiór krytyczny w ogólności?

Odpowiedź daje tzw. *funkcja mocy testu* (zwana też krótko *mocą testu*). Niech hipoteza dotyczy pewnego parametru ϑ (w przykładzie parametrem tym była wartość oczekiwana m). Oznaczmy przez $\delta \in \Omega_k$ zdarzenie polegające na tym, że odrzucimy weryfikowaną hipotezę, gdy interesujący nas parametr ma wartość ϑ (w podanym przykładzie: gdy wartość oczekiwana wskazań testowanego przyrządu pomiarowego jest równa m).

Definicja 5.6. Prawdopodobieństwo zdarzenia $\delta \in \Omega_k \mid \vartheta$ rozpatrywane jako funkcja parametru ϑ oznaczamy

$$M(\vartheta) = P(\delta \in \Omega_k \mid \vartheta)$$

i nazywamy *mocą testu*.

Liczbę α oraz zbiór krytyczny testu wybiera się tak, żeby moc testu spełniała pewne „rozsądne wymagania”. Co to znaczy wyjaśnimy na przykładzie.

Przykład 5.8. W poprzednim przykładzie przyjęliśmy za zbiór krytyczny zbiór

$$\Omega_k = \{\bar{X}_5 : |\bar{X}_5 - 12| > 0,088 \mid m\}.$$

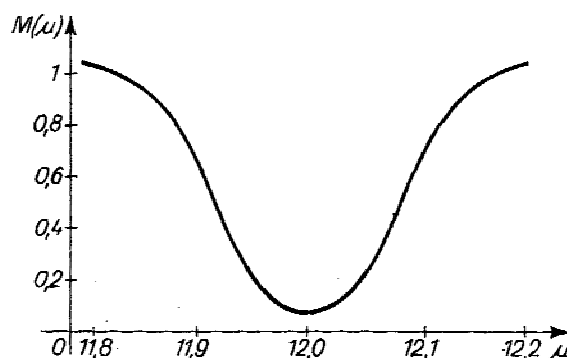
Zatem funkcją mocy tego testu jest

$$M(m) = P(|\bar{X}_5 - 12| > 0,088 \mid m),$$

czyli prawdopodobieństwo zdarzenia $|\bar{X}_5 - 12| > 0,088$, gdy rozkład zmiennej losowej jest normalny $N(m, 0,1)$. Wyznaczając to prawdopodobieństwo otrzymamy

$$M(m) = \Phi\left(\frac{11,912 - m}{0,1} \sqrt{5}\right) + 1 - \Phi\left(\frac{12,088 - m}{0,1} \sqrt{5}\right).$$

Wykres tej funkcji jest przedstawiony na poniższym rysunku.



Z wykresu widać, że jeśli $m = 12$, czyli gdy weryfikowana hipoteza jest prawdziwa, prawdopodobieństwo jej odrzucenia jest najmniejsze i rośnie wraz ze wzrostem obciążenia przyrządu (gdy $m > 12$ lub $m < 12$).

W praktyce najczęściej są stosowane testy wymienione poniżej.

1. Jednostronny test hipotezy o wartości oczekiwanej w rozkładzie normalnym przy znanej wariancji, czyli

$H: m = m_0$ przy hipotezach alternatywnych $K: m > m_0$ (lub $K: m < m_0$).

Danymi w tym teście są:

α – poziom istotności testu,

σ – odchylenie standardowe (otrzymane ze znanej wariancji),

n – liczebność próbki.

Hipotezę dyskwalifikuje się, gdy $\bar{X}_n > \lambda$, gdzie wartość λ jest ustalona z warunku

$$\Phi\left(\frac{\lambda - m_0}{\sigma} \sqrt{n}\right) = 1 - \alpha.$$

Moc testu określa wzór

$$M(m) = 1 - \Phi\left(\frac{\lambda - m}{\sigma} \sqrt{n}\right).$$

2. Dwustronny test hipotezy o wartości oczekiwanej w rozkładzie normalnym przy znanej wariancji, czyli

$H: m = m_0$ przy hipotezach alternatywnych $K: m \neq m_0$.

Danymi są te same wielkości, co poprzednio.

Hipotezę odrzucamy, gdy $|\bar{X}_n - m_0| > \lambda$, gdzie wartość λ wyznacza się z warunku

$$\Phi\left(-\frac{\lambda}{\sigma} \sqrt{n}\right) = \frac{\alpha}{2}.$$

Moc testu jest dana wzorem

$$M(m) = \Phi\left(\frac{m_0 - \lambda - m}{\sigma} \sqrt{n}\right) + 1 - \Phi\left(\frac{m_0 + \lambda - m}{\sigma} \sqrt{n}\right).$$

3. Jednostronny test hipotezy o wartości oczekiwanej w rozkładzie normalnym przy nieznanej wariancji, czyli

$H: m = m_0$ przy hipotezach alternatywnych $K: m > m_0$ (lub $K: m < m_0$).

Danymi w tym teście są:

α – poziom istotności testu,

n – liczebność próbki.

Hipotezę odrzucamy, gdy $\frac{\bar{X}_n - m_0}{S_n} \sqrt{n} > \lambda$, gdzie S_n^2 oznacza wariancję z próbki, a wartość λ wyznacza się z warunku

$$H_{n-1}(\lambda) = 1 - \alpha.$$

Funkcja $H_{n-1}(x)$ jest dystrybuantą rozkładu t -Studenta z $n - 1$ stopniami swobody. W celu wyznaczenia mocy testu trzeba rozważyć pewien nowy rozkład, bo gdy $m \neq m_0$, to zmienna losowa $\frac{\bar{X}_n - m_0}{S_n} \sqrt{n}$ nie ma rozkładu t -Studenta (pomijamy to zagadnienie).

4. Dwustronny test hipotezy o wartości oczekiwanej w rozkładzie normalnym przy nieznannej wariancji, czyli

$$H: m = m_0 \text{ przy hipotezach alternatywnych } K: m \neq m_0.$$

Danymi są te same wielkości, co poprzednio.

Hipotezę odrzucamy, gdy $\left| \frac{\bar{X}_n - m_0}{S_n} \sqrt{n} \right| > \lambda$, gdzie S_n^2 oznacza wariancję z próbki, a wartość λ wyznacza się z warunku

$$H_{n-1}(\lambda) = \frac{\alpha}{2}.$$

Oдноśnie mocy testu obowiązuje poprzednia uwaga.

5. Jednostronny test hipotezy o parametrze p zmiennej losowej X dwupunktowej (o rozkładzie zero-jedynkowym), tj. zmiennej dla której

$$P(X=1) = p, \quad P(X=0) = 1 - p.$$

Hipoteza ma postać

$$H: p = p_0 \text{ przy hipotezach alternatywnych } K: p > p_0 \text{ (lub } K: p < p_0).$$

Hipotezę weryfikuje się na podstawie próbki $X_1, X_2, \dots, X_n$. Suma S_n tych zmiennych ma rozkład Bernoulliego, tj.

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

przy czym dla dużej liczebności próbki (dużych wartości n) i małych wartości p stosuje się przybliżenie za pomocą rozkładu Poissona, tzn.

$$P(S_n = k) = \frac{(np)^k}{k!} \exp(-np).$$

W praktyce przybliżenie to stosuje się, gdy $p \leq 0,1$ oraz $np(1-p) \leq 9$. Gdy $np(1-p) > 9$, to zwykle rozkład dwumianowy przybliża się za pomocą rozkładu normalnego (możliwość taka wynika z centralnego twierdzenia granicznego – zob. następny rozdział):

$$P\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} < b\right) = \Phi(b) - \Phi(a).$$

Weryfikowaną hipotezę H odrzuca się, gdy $S_n > \lambda$, gdzie wartość λ jest najmniejszą liczbą ustaloną z zależności

$$P(S_n > \lambda, \text{ gdy } p = p_0) \leq \alpha.$$

Zauważmy, że równość w powyższej nierówności może nie być spełniona dla żadnej wartości λ , gdyż zmienna losowa S_n przyjmuje tylko nieujemne wartości całkowite.

Moc testu określa następujący wzór:

$$M(p) = P(S_n > \lambda \mid p),$$

przy czym za rozkład w ostatnich dwóch wzorach przyjmuje się (w zależności od okoliczności) jeden z wymienionych rozkładów, tzn. Bernoulliego, Poissona lub normalny.