Prawdopodobieństwo

2.1. Rzucamy dwukrotnie kostką do gry. Obliczyć prawdopodobieństwo, że suma oczek będzie większa od 9, jeżeli za pierwszym razem wypadło 6 oczek?
Odp. 1 2 .

2.2. W skrzyni znajduje się 12 elementów, z czego 6 jest dobrych a 6 wadliwych. W sposób losowy, bez zwracania wybieramy dwa elementy. Oblicz prawdopodobieństwo wylosowania za drugim elementu wadliwego, pod warunkiem, że za w pierwszym razem wybrano element dobry.
Odp. 6/11 .

2.3. Na dworcu kolejowym znajdują się dwoje schodów ruchomych. Pierwsze są sprawne z prawdopodobieństwem 1/2 , natomiast drugie 1/3 . Prawdopodobieństwo, że działają pierwsze schody, gdy zepsute są drugie wynosi 1/2 .
a) Jakie jest prawdopodobieństwo, że działają drugie schody, pod warunkiem, że nie działają pierwsze?
b) Jakie jest prawdopodobieństwo, że działają przynajmniej jedne schody?
Odp. a) 1/3 , b) 2/3 .

2.4. Rzucamy dwiema symetrycznymi kostkami do gry. Oblicz prawdopodobieństwo otrzymania sumy oczek równej 3, jeżeli na pierwszej kostce wypadła 1.
Odp. 1/6 .

2.5. Rozważmy rodziny z dwojgiem dzieci. Niech d oznacza dziewczynkę, c - chłopca. Zdarzeniami elementarnymi będą pary: (d,d), (d,c), (c,d), (c,c), gdzie pierwsza litera w parze oznacza płeć starszego dziecka, druga zaś młodszego. Zakładając, że wszystkie zdarzenia są jednakowo prawdopodobne obliczyć prawdopodobieństwo, że w losowo wybranej rodzinie z dwojgiem dzieci, jest dwóch chłopców, pod warunkiem, że w tej rodzinie jest co najmniej jeden chłopiec.
Odp. 1/3 .

2.6. Student dojeżdża na uczelnię rowerem średnio co drugi dzień, autobusem co trzeci dzień, a tramwajem co szósty. Jadąc rowerem, spóźnia się z prawdopodobieństwem raz na sześćdziesiąt razy jadąc autobusem - raz na dwadzieścia razy, a tramwajem raz na dziesięć razy. Oblicz prawdopodobieństwo, że student spóźni się na uczelnię.
Odp. 1/24 .

2.7. Potrzeby świerkowych sadzonek dla nadleśnictwa pokrywa produkcja dwóch szkółek leśnych. Pierwsza szkółka pokrywa 75% zapotrzebowania, przy czym na 100 sadzonek z tej szkółki 80 jest pierwszej jakości. Druga szkółka pokrywa 25% zapotrzebowania, przy czym na 100 sadzonek z tej szkółki 60 jest pierwszej jakości. Obliczyć prawdopodobieństwo, że losowo wybrana sadzonka jest pierwszej jakości.
Odp. 0, 75.

Zmienna losowa (dyskretna)

Przykład 3.3. Zmienna losowa X dyskretna ma funkcję prawdopodobieństwa określoną tabelą

x_i	-1	0	2
p_i	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$

Wyznaczyć dystrybuantę zmiennej losowej X.

Rozwiązanie. Dystrybuanta zmiennej losowej X jest funkcją F(x) = P(X ≤ x). Wtedy:

$F(-2) = P(X \le -2) = 0$
$F(-1) = P(X \le -1) = \frac{1}{2}$

$F(-0,5) = P(X \le -0,5) = \frac{1}{2}$
$F(0) = P(X \le 0) = P(-1) + P(0) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$
$F(1) = P(X \le 1) = P(-1) + P(0) = \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$
$F(3) = P(X \le 3) = P(-1) + P(0) + P(2) = \frac{1}{2} + \frac{1}{6} + \frac{1}{3} = 1$

Zatem:

$$F(x) = \begin{cases} 0 & dla\ x < -1 \\ \frac{1}{2} & dla -1 \le x < 0 \\ \frac{2}{3} & dla\ 0 \le x < 2 \\ 1 & dla\ x > 2 \end{cases}$$

Rozkład prawdopodobieństwa zmiennej losowej (dyskretnej)

Przykład 3.4. Zmienna losowa ma rozkład określony tabelą:

x_i	0	1	2	3
p_i	0,25	0,2	0,15	0,4

Wyznaczyć:

a) dystrybuantę rozkładu,
b) P(X < 3),
c) E(X),
d) $D^2(X)$,
e) σ.

<u>**Rozwiązanie**</u>

b) Mamy $P(X \le 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = 0,25 + 0,2 + 0,15 = 0,6$.

c) $E(X) = 0 \cdot 0,25 + 1 \cdot 0,2 + 2 \cdot 0,15 + 3 \cdot 0,4 = 1,70$

d) $[E(X)]^2 = 2,89$
$E(X^2) = 0^2 \cdot 0,25 + 1^2 \cdot 0,2 + 2^2 \cdot 0,15 + 3^2 \cdot 0,4 = 4,4$
$D^2(X) = E(X^2) - [E(X)]^2 = 4,4 - 2,89 = 1,51$.

e) $\sigma = \sqrt{1,51} \approx 1,23$

$$F(X) = \begin{cases} 0 & x \in (-\infty, 0) \\ 0,25 & x \in [0,1) \\ 0,45 & x \in [1,2) \\ 0,60 & x \in [2,3) \\ 1 & x \in [3,\infty) \end{cases}$$

Rozkłady zmiennej skokowej (rozkład Bernoulliego)

Przykład 3.5. W hali fabrycznej pracuje 5 maszyn. Każda z nich psuje się z prawdopodobieństwem p = 1/3 niezależnie od siebie. Wyznaczyć prawdopodobieństwa:
a) zepsuła się jedna maszyna, tj. P(X = 1),
b) żadna maszyna się nie popsuła, tj. P(X = 0),
c) zepsuły się trzy maszyny, tj. P(X = 3),
d) zepsuła się co najmniej jedna maszyna, tj. P(X ≥ 1),
e) zepsuła się co najwyżej jedna maszyna, tj. P(X ≤ 1),
f) zepsuło się więcej niż jedna maszyna, tj. P(X > 1).

<u>Rozwiązanie</u>:

a) P(X=1)=$\binom{5}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^4$=$\frac{80}{243} \approx 0{,}32922$
b) P(X=0)=$\binom{5}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^5$=$\frac{32}{243} \approx 0{,}13169$
c) P(X=3)=$\binom{5}{3}\left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^2$=$\frac{40}{243} \approx 0{,}16461$
d) P(X≥1)=1-P(X<1)=P(X=0)=0,86831
e) P(X≤1)=P(X=0)+P(X=1)=0,46091
f) P(X>1)=1-P(X≤1)=0,53909

Przykład 3.6. Środek owadobójczy zabija przeciętnie 90% owadów. Środek ten zastosowano na 10 owadach. Obliczyć prawdopodobieństwo, że co najwyżej dwa osobniki przeżyją.

<u>Rozwiązanie</u>:
Oznaczmy przez A zdarzenie, że owad przeżyje. Mamy więc p = 0, 1, q = 0, 9, n = 10.
P(X≤2)=P(X=0)+P(X=1)+P(X=2)=$\binom{10}{0}(0{,}1)^0(0{,}9)^{10}$+$\binom{10}{1}(0{,}1)^1(0{,}9)^9 + \binom{10}{2}(0{,}1)^2(0{,}9)^8$=0,9281

Rozkład Poissona

Przykład 3.7. Daltonizm stwierdza się o 1% mężczyzn. Obliczyć prawdopodobieństwo, że w próbie liczącej n = 100 mężczyzn
a) nie będzie ani jednego daltonisty,
b) będzie co najmniej trzech.

<u>Rozwiązanie:</u>
Mamy n = 100, p = 0, 01, czyli λ = n·p = 1

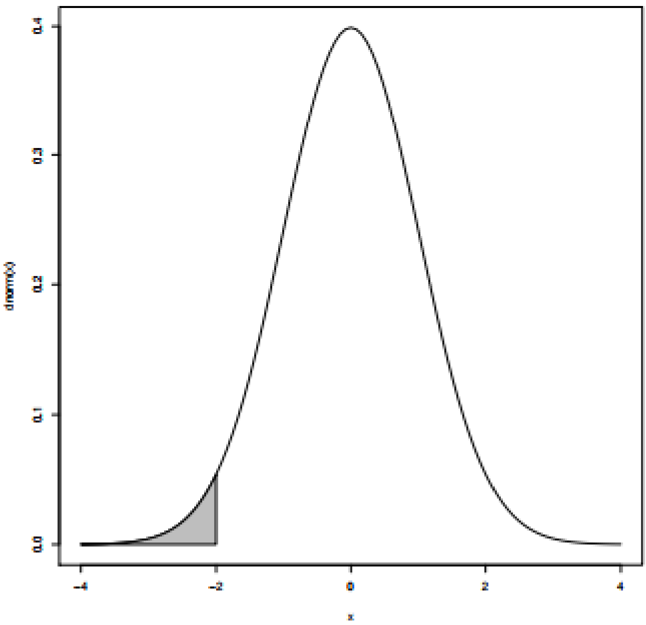
a) P(X=0)=$\frac{1}{0!}e^{-1}$=0,37
b) P(X≥3)=1-P(X<3) = 1 − [P(X = 0) + P(X = 1) + P(X = 2)]
P(X=0)=$\frac{1}{0!}e^{-1}$=0,36788
P(X=1)= $\frac{1}{1!}e^{-1}$=0,36788
P(X=2)= $\frac{1}{2!}e^{-1}$=0,18394
Stąd:
P(X≥3)= 1-[0,36788+0,36788+0,18394]=0,0803

Rozkłady zmiennej losowej ciągłej (rozkład normalny)

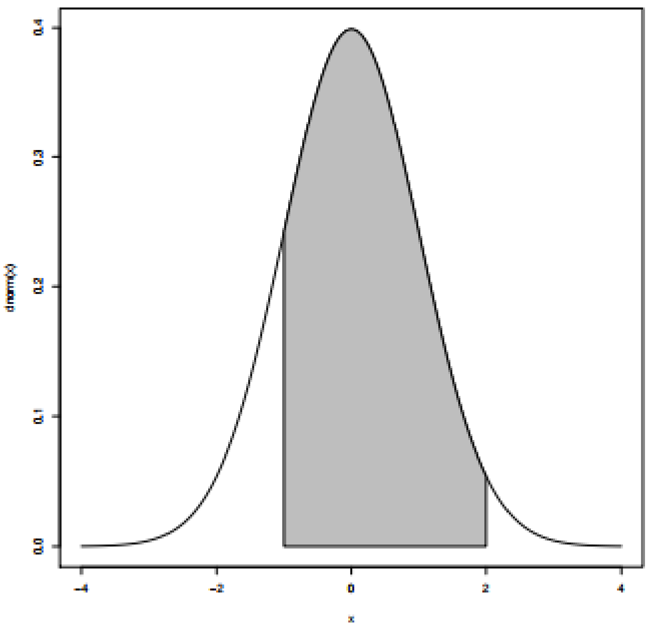
Przykład 3.8. Dla rozkładu N(0, 1) obliczyć
a) P(X < −2),
b) P(−1 ≤ X ≤ 2),
c) P(X > 6).

<u>Rozwiązanie</u>:
Mamy:

a) $P(X<2)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-2} e^{-\frac{t^2}{2}} = \Phi(-2) = 1-\Phi(2) = 1-0{,}97725 = 0{,}02275$



b) P(−1 ≤ X ≤ 3) = F(3) − F(1) = Φ(3) − (1 − Φ(−1)) = = Φ(3) + Φ(1) − 1 = 0, 9987 + 0, 8413 − 1 ≈ 0,84.



c) P(X > 6) = 1 − P(X ≤ 6) = 1 − F(6) = 1 − Φ(6) ≈ 1 − 1 ≈ 0

Przykład 3.10. Wydajność pracy jest mierzona liczbą detali wykonanych przez pracownika na danym stanowisku. Liczba detali dana jest zmienną losową X dla N(8, 2). Obliczyć P(X < 5).

Rozwiązanie:
Mamy Y=$\frac{X-8}{2}$, czyli
P(X<5)=P$\left(\frac{X-8}{2} < \frac{5-8}{2}\right)$=P (Y < −1, 5) = F(−1, 5) = Φ(−1, 5) = 1- Φ(1, 5) = 1 − 0, 93319 = 0,06681

Przykład 3.11. Urządzenie złożone z dwóch bloków pracuje w ten sposób, że najpierw włączony jest pierwszy blok, a w chwili awarii tego bloku włącza się drugi blok. Czasy bezawaryjnej pracy bloków są zmiennymi losowymi o rozkładach odpowiednio N(60; 4) i N(80; 3). Obliczyć prawdopodobieństwo, że urządzenie będzie pracować co najmniej 150 godzin.

Rozwiązanie:
Z = X + Y , zatem Z ma rozkład N(60 + 80,$\sqrt{4^2+3^2}$) = N(140, 5).
P(Z > 150) = P$\left(\frac{Z-140}{5} \geq \frac{150-140}{5}\right)$= P (R ≥ 2) = 1 − P(R < 2) = 1 − Φ(2) = 1 − 0, 97725 ≈ 0, 02275.
Czyli około 2, 2%.

Zadania samodzielne:
3.1. Zmienna losowa skokowa X ma funkcję prawdopodobieństwa:

x_i	-1	1	4
p_i	0,5	0,4	0,1

Wyznaczyć dystrybuantę zmiennej losowej X i narysować jej wykres.

Odp. F(X)= 0 dla x ≤ -1; 0,5 dla -1 < x ≤ 1; 0,9 dla 1 < x ≤ 4; 1 dla x > 4

3.2. Zmienna losowa ciągła X ma gęstość:

$$f(x)=\begin{cases} 0 & x \leq -1 \\ \frac{1}{2} & -1 < x \leq 0 \\ x & 0 < x \leq 1 \\ 0 & x > 0 \end{cases}$$

Wyznaczyć dystrybuantę zmiennej X.

Odp.

$$f(x)=\begin{cases} 0 & x \leq -1 \\ \frac{1}{2}(x+1) & -1 < x \leq 0 \\ \frac{1}{2}(x^2+1) & 0 < x \leq 1 \\ 1 & x > 0 \end{cases}$$

3.3. Dystrybuanta zmiennej losowej ciągłej dana jest wzorem

$$F(x)=\begin{cases}0 & x\le 0\\ \frac{1}{16}x^2 & 0<x\le 4\\ 1 & x>4\end{cases}$$

Wyznaczyć funkcję gęstości zmiennej X oraz wyznaczyć wartość oczekiwaną i wariancję zmiennej X.

Odp.

$$f(x)=\begin{cases}0 & x\le 0\\ \frac{1}{8}x & 0<x\le 4, E(X)=\frac{8}{3}, \sigma^2(X)=\frac{8}{9}\\ 0 & x>4\end{cases}$$

3.4. Gęstością zmiennej losowej X jest funkcja

$$f(x)=\begin{cases}0 & x\le 1\\ \frac{1}{2} & 1<x\le 3\\ 0 & x>3\end{cases}$$

Wyznaczyć dystrybuantę zmiennej losowej X oraz wyznaczyć P(X > 2).

Odp.

$$F(x)=\begin{cases}0 & x\le 1\\ \frac{1}{2}x-\frac{1}{2} & 1<x\le 3, P(X>2)=\frac{1}{2}\\ 1 & x>3\end{cases}$$

3.5. Zmienna losowa X przyjmuje wartości 0, 1, 2, 3 z prawdopodobieństwami odpowiednio równymi 0, 1; 0, 1; 0, 2; 0, 6. Obliczyć wartość oczekiwaną i wariancję zmiennej X.

Odp. E(X) = 2, 3, σ^2(X) = 1,01.

3.6. Zmienna losowa X ma rozkład normalny N(1,5;2). Obliczyć prawdopodobieństwo:

a) P(X < −2, 5)
b) P(X > −0, 5)
c) P(0, 5 < X < 2).

Odp. a) 0,02275, b) 0,8413, c) 0,2902.

3.7. Masa gruszek odmiany klops ma rozkład normalny N(160, 30). Oblicz prawdopodobieństwo, że gruszka tego gatunku waży od 130 do 160 gramów.

Odp. 0,3413.

3.8. W populacji studentów uczęszczających na zajęcia ze statystyki dokonano pomiaru wzrostu mężczyzn. W wyniku badania stwierdzono, że zmienna losowa X wyrażająca wzrost studenta ma rozkład normalny N(178, 10). Oblicz prawdopodobieństwo, że
a) wzrost studenta jest mniejszy niż 188 cm,
b) wzrost studenta jest większy niż 172,
c) wzrost studenta jest większy niż 200 cm,
d) wzrost studenta należy do przedziału (166 cm, 186 cm).

<u>Odp</u>. a) 0,8413, b) 0,7257, c) 0,0139, d) 0,673.

Elementy statystyki opisowej

5.1. W fabryce w ciągu pięciu dni roboczych wyprodukowano pięć wyrobów o wadze: 12, 14, 16, 18, 20. Obliczyć średnią i odchylenie standardowe.

5.2. W pewnej szkole badano wzrost dziewcząt klas czwartych. Otrzymano wyniki: 140, 148 148, 148, 150, 150, 156, 156, 160, 160, 160, 160, 162, 163, 164, 166, 168, 169, 170, 175, 175, 180. Obliczyć: medianę, dominantę, kwartyle pierwszy i drugi oraz odchylenie ćwiartkowe.

5.3. Oceny studentów z przedmiotu statystyka przedstawia tabela

Ocena	3	3,5	4	4,5	5
Liczba studentów	25	30	10	15	20

Obliczyć: średnią, odchylenie standardowe, współczynnik zmienności, asymetrię.

5.4. Poniższa tabela przedstawia dane dotyczące wydajności pracy (w szt/h) pewnego wydziału zakładu produkcyjnego. Wyznaczyć medianę, dominantę, kwartyle pierwszy i trzeci oraz odchylenie ćwiartkowe.

Wydajność	40	50	60	30	25	20
Liczba pracowników	15	10	30	10	15	10

5.5. Dla poniższego szeregu rozdzielczego przedziałowego, przedstawiającego staż pracy pracowników pewnego przedsiębiorstwa, obliczyć: średnią, wariancję, odchylenie standardowe, współczynnik zmienności, asymetrię.

Przedział	1-5	5-10	10-15	15-20	20-25	25-30	30-35
Liczba elementów	10	20	25	30	35	10	20

5.6. Dla poniższego szeregu rozdzielczego przedziałowego obliczyć: medianę, dominantę, kwartyl pierwszy i drugi oraz odchylenie ćwiartkowe.

0-4	4-8	8-12	12-16	16-20	20-24	24-28
10	20	30	40	10	5	10

Estymacja

Przykład 6.1. Zakładając, że roczne wydatki na paliwo można uznać za cechę o rozkładzie N(μ, σ), pobrano próbę losową liczącą 100 małych zakładów. Uzyskano x = 12 oraz s = 4, 72 (w tys. zł). Wyznaczyć przedział ufności dla wartości oczekiwanej na poziomie ufności 1 − α = 0, 96.

Rozwiązanie. Ponieważ n > 100 oraz σ jest nieznane, więc korzystamy z przedziału postaci :

$$\overline{x} - t_\alpha \frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + t_\alpha \frac{\sigma}{\sqrt{n}},$$

Otrzymujemy

$$12 - t_\alpha \frac{4,72}{\sqrt{100}} < \mu < 12 + t_\alpha \frac{4,72}{\sqrt{100}},$$

, gdzie $\Phi(t_\alpha)$ = 1 – 0,04/2 = 0,98. Z tablic rozkładu normalnego otrzymujemy t_α = 2,05.
Zatem μ ∈ (11, 2; 12, 8).

Przykład 6.2. Poddano analizie wydatki na odzież w wiejskich rodzinach 5-osobowych. Z populacji tych rodzin wylosowano próbę 289-elementów. Na podstawie przeprowadzonych obserwacji ustalono przeciętną skalę wydatków na odzież x = 100 zł. Badania z lat ubiegłych wykazały, że rozkład wydatków na odzież jest rozkładem normalnym o stałej wariancji σ^2 = 576. Wyznaczyć przedział ufności średnich miesięcznych wydatków na odzież w wiejskich rodzinach 5-osobowych przyjmując poziom ufności 1 − α = 0,98.

Rozwiązanie. Korzystając z powyższego wzoru dla $\Phi(t_\alpha)$ = 0,99, t_α = 2,35. Zatem 96,682 < μ < 103,318. W rodzinach 5-osobowych miesięczne wydatki na odzież zawierają się w przedziale μ ∈ (96, 68zł; 103,32zł).

Testowanie hipotez statystycznych

Przykład 6.3. Załóżmy, że „długość życia" opon samochodowych ma rozkład normalny N(μ, σ). Producent twierdzi, że wartość przeciętna tej charakterystyki jest równa 50 tys. km. Na podstawie 100 losowo wybranych opon otrzymano x = 45 tys. km oraz s = 8 tys. km. Czy na poziomie istotności α = 0, 05 można uważać, że producent ma rację?

Rozwiązanie. Mamy H_0: μ = 50, H_1: μ ≠ 50.

Obliczamy teraz wartość statystyki testowej:

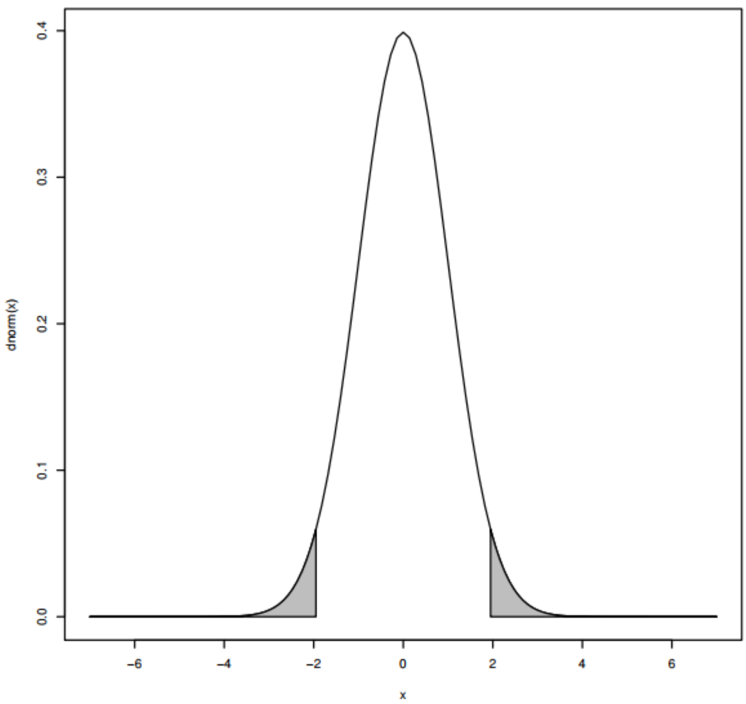
$$U = \frac{\overline{X} - \mu_0}{S}\sqrt{n} = \frac{45-50}{8}\sqrt{100} = -6,25$$

Budujemy dwustronny obszar odrzucenia dla α= 0,05.

F(U)=1- α/2=0,975, skąd U=1,96. (dla rozkładu normalnego)

czyli (-∞;-1,96] ∪ [1,96;∞)

Oznacza to, że wartość statystyki testu wpada do zbioru krytycznego, czyli należy odrzucić hipotezę H_0 na rzecz hipotezy alternatywnej H_1. Innymi słowy producent nie ma racji twierdząc, że przeciętna długość życia opon wynosi 50 tys. km. Na poniższym obrazku na szaro został zaznaczony zbiór krytyczny.



Przykład 6.4. W pewnym rejonie morza dokonano 5 niezależnych pomiarów głębokości wody. Otrzymano średnią głębokość x = 770m oraz odchylenie standardowe s = 6,2. Na poziomie istotności α = 0,02 zweryfikować hipotezę, że średnia głębokość morza w tym rejonie wynosi μ = 775m, przyjmując że rozkład pomiarów głębokości w tym rejonie morza ma rozkład normalny.

Rozwiązanie. Testujemy hipotezę

H_0: μ = 775m

H1 : μ ≠ 775m

Statystyka testowa przyjmuje wartość

$$U = \frac{\bar{X} - \mu_0}{S}\sqrt{n} = \frac{770 - 775}{6,2}\sqrt{4} = -1,6$$

Wartość $t_{4;0,02}$ = 3,747 odczytujemy z tablic kwantyli rozkładu t-Studenta. Widzimy stąd, że wartość statystyki testowej nie należy do zbioru krytycznego, zetem nie ma podstaw do odrzucenia hipotezy zerowej.

6.1. Poddano analizie wydatki na opłaty za telefon TP S.A w 100 gospodarstwach domowych w pewnym mieście. Na podstawie przeprowadzonych obserwacji ustalono średnią miesięczną opłatę za telefon $\bar{x}$ = 95zł i odchylenie standardowe σ = 15zł. Zakładamy że wydatki mają rozkład normalny. Na poziomie ufności 1 − α = 0,98 wyznaczyć przedział ufności dla wartości przeciętnej miesięcznych opłat za telefon.

Odp. 91,55 < μ < 98,45.

6.2. W zakładzie „Alfa" zbadano staż pracowników fizycznych. Z populacji tych pracowników wylosowano próbę 169-elementową, z której obliczono $\bar{x}$ = 7,2 lat. Rozkład stażu pracowników fizycznych jest rozkładem normalnym z odchyleniem standardowym σ = 3,2 lat. Przyjmując współczynnik ufności zbudować przedział ufności dla nieznanego średniego stażu pracy w populacji pracowników fizycznych w tym zakładzie.

Odp. 6,634 < μ < 7,766.

6.3. Cecha X ma rozkład N(μ,σ), gdzie μ,σ są nieznane. Na podstawie próby 17 elementowej obliczono x = 60, s = 0,5. Zweryfikować hipotezę H_0: μ = 61, 5, wobec hipotezy alternatywnej H_1: μ ≠ 61,5 na poziomie istotności α = 0,05.

Odp. Odrzucamy H_0.

6.4. Z dużej partii słupów betonowych wybrano próbkę losową 64 słupów. Średnia wytrzymałość na ściskanie w tej próbie wynosiła $\bar{x}$ = 245 kG/cm . Odchylenie standardowe s = 5kG/cm . Zweryfikować hipotezę H_0: μ = 240kG/cm, wobec hipotezy alternatywnej H_1: μ ≠ 240kG/cm , na poziomie istotności α = 0, 01, przy założeniu, że wytrzymałość na ściskanie jest zmienną losową o rozkładzie normalnym.

Odp. Odrzucamy H_0.

6.5. W pewnym zakładzie wybrano losowo 10 pracowników. Otrzymano średni wiek $\bar{x}$ = 32 lata oraz odchylenie standardowe s = 4 lata. Zakładając, że wiek pracowników ma rozkład normalny zweryfikować hipotezę, na poziomie istotności α = 0,05, że średni wiek pracowników jest istotnie wyższy niż 30 lat. (Wskazówka: H_0 : μ = 30, H_1 : μ > 30).

Odp. Nie ma podstaw do odrzucenia H_0. tzn. nie możemy twierdzić, że średni wiek w przedsiębiorstwie jest istotnie większy od 30 lat.