Zadanie 1

W pewnej firmie remontowej wartość materiału (w zł) zużytego do remontu mieszkania losowo wybranego klienta jest zmienną losowa X o rozkładzie normalnym $N(5000, 100)$. Dochód z remontu jest zmienną losową $Y = 0{,}5X + Z$, gdzie Z jest zmienna losową mającą wartość średnią 1000 oraz wariancję 20. X i Z są niezależnymi zmiennymi losowymi. Oblicz wartość średnią dochodu firmy z remontu losowo wybranego mieszkania.

Rozwiązanie

Wartość średnia dochodu firmy, to wartość oczekiwana zmiennej losowej Y.

$E(Y) = 0{,}5E(X) + E(Z) = 0{,}5*5000 + 1000 = 3500.$

Zadanie 2

Zanotowano czasy (w min.) wykonania pewnego projektu w konkursie programistycznym przez 16-tu losowo wybranych uczestników konkursu. Obliczono dla nich średni próbkowy czas wykonania projektu $\bar{x} = 105{,}5$ (min.) oraz próbkowe odchylenie standardowe $s = 20$ (min.). Wyznacz 95% przedział ufności dla wartości średniej czasu wykonania tego projektu, jeśli można założyć, że jest on zmienną losową o rozkładzie normalnym.

Rozwiązanie

Model 2

- *rozkład cechy w populacji normalny,*
- *odchylenie standardowe w populacji nieznane,*
- *mała próba*

Liczebność próby $n = 16$
Średnia próbkowa $\bar{x} = 105{,}5$
Odchylenie standardowe w próbce $s = 20$

Poziom ufności 1- $\alpha = 0{,}95$, stąd $\alpha = 0{,}05$, $\frac{\alpha}{2} = 0{,}025$, $1 - \frac{\alpha}{2} = 0{,}975$

Kwantyl $t_{1-\alpha/2;\, n-1} = t_{0{,}975;\, 15} = 2{,}1315$

Przedział ufności

$$[\ \bar{x} - t_{1-\alpha/2;\, n-1}\frac{s}{\sqrt{n}}\ ,\ \bar{x} + t_{1-\alpha/2;\, n-1}\frac{s}{\sqrt{n}}\ ]$$

$$[\ 105{,}5 - 2{,}1315\frac{20}{\sqrt{16}}\ ,\ \bar{x} + 2{,}1315\frac{20}{\sqrt{16}}\ ]$$

[94,84 , 116,16]

Mówiąc, że wyznaczony przedział [94,84 , 116,16] obejmuje średni czas wykonania projektu mamy 95% ufności, że sąd ten jest prawdziwy.

Zadanie 3

Wagi pięciu losowo wybranych noworodków wyniosły (w kg):

3,75 3,45, 3,50, 3,90, 3,25

Zakładając rozkład normalny wagi noworodka o odchyleniu standardowym 0,5 kg wyznacz 99% przedział ufności dla wartości średniej wagi noworodka.

Rozwiązanie

Model 1

- *rozkład cechy w populacji normalny,*
- *odchylenie standardowe w populacji znane,*

Liczebność próby $n = 5$
Średnia próbkowa $\bar{x} = 3{,}57$
Odchylenie standardowe w populacji $\sigma = 0{,}5$

Poziom ufności 1- $\alpha = 0{,}99$, stąd $\alpha = 0{,}01$, $\frac{\alpha}{2} = 0{,}005$, $1 - \frac{\alpha}{2} = 0{,}995$

Kwantyl $z_{1-\alpha/2} = z_{0{,}995} = 2{,}58$

Przedział ufności

$$[\ \bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\ ,\ \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\ ]$$

$$[\ 3{,}57 - 2{,}58\frac{0{,}5}{\sqrt{5}}\ ,\ 3{,}57 + 2{,}58\frac{0{,}5}{\sqrt{5}}\ ]$$

[2,994 , 4,146]

Mamy 99% ufności (zaufania, pewności), że wyznaczony przedział obejmuje wartość średnią wagi noworodka

Zadanie 4

Czy dla danych z zadania 2, na poziomie istotności 0,05 można stwierdzić, że wartość średnia czasu wykonania projektu przekracza 100 minut?

Uzupełnij poniższe punkty:

1. Badany parametr: μ - średni czas wykonania projektu
2. Hipoteza zerowa: H_0: $\mu = 100$,
3. Hipoteza alternatywna: H_1: $\mu > 100$
4. Statystyka testowa: $T = \frac{\bar{X} - \mu}{S}\sqrt{n}$ ma rozkład Studenta o 15 stopniach swobody, gdy H_0 jest prawdziwa
5. Wartość statystyki testowej: $t = \frac{105{,}5 - 100}{20}\sqrt{16} = 1{,}1$
6. Zbiór krytyczny: $C = \{t : t \geq t_{0{,}95;\, 15}\} = \{t : t \geq 1{,}753\}$
7. Wniosek:
   Ponieważ wartość statystyki testowej nie należy do obszaru krytycznego, nie ma podstaw do odrzucenia hipotezy zerowej. Nie możemy zatem twierdzić, że średni czas wykonania projektu przekracza 100 minut.

Zadanie 5

Na podstawie ankiety przeprowadzonej wśród studentów pewnej uczelni technicznej wyznaczono przybliżony 90% przedział ufności dla proporcji p studentów tej uczelni, którzy biegle władają językiem angielskim: [67,52 ; 72,48]. Jaka jest wartość estymatora $\hat{p}$ (oceny na podstawie próbki nieznanej proporcji p) proporcji studentów, którzy biegle władają językiem angielskim?

Rozwiązanie

Ponieważ $\hat{p}$ jest środkiem przedziału ufności, jego wartość wynosi . (67,52 + 72,48)/2 = 70.

Zadanie 6
W losowo wybranym półroczu liczba stażystów zatrudnionych w firmie jest zmienną losową X o funkcji prawdopodobieństwa określonej tabelą:

x	1	2	5
$p(x)$	0,7	0,2	0,1

a) Oblicz wartość oczekiwaną oraz wariancję liczby stażystów pracujących w firmie w losowo wybranym półroczu.
b) Oblicz wartość dystrybuanty $F(x)$ zmiennej losowej X w punkcie $x = 2{,}5$.

Rozwiązanie

$E(X) = 1*0{,}7 + 2*0{,}2 + 5*0{,}1 = 1{,}6$

$\mathrm{Var}(X) = (1-1{,}6)^2 * 0{,}7 + (2-1{,}6)^2 * 0{,}2 + (5-1{,}6)^2 * 0{,}1 = 1{,}44$

Dystrybuanta zmiennej losowej X

$$F(x) = \begin{cases} 0 & dla \quad x < 1 \\ 0{,}7 & dla \quad 1 \le x < 2 \\ 0{,}9 & dla \quad 2 \le x < 5 \\ 1 & dla \quad x \ge 5 \end{cases}$$

Dla $x = 2{,}5$ wartość dystrybuanty $F(2{,}5) = 0{,}9$.

Zadanie 8
Dla siedmioelementowej próbki jej wartości spełniają zależności:

$x_2 = x_3 = x_5 < x_1 = x_7 < x_6 = x_4$.

a) Wyznacz wartość mediany i górnego kwartyla.
b) Uprość wzór na średnią próbkową w przypadku danej próbki.

Rozwiązanie

Mediana (wartość środkowa) jest równa $M = Q_2 = x_1 = x_7$. Kwartyl górny $Q_3 = x_6$.

Średnia próbkowa $\bar{x} = \dfrac{2x_1 + 3x_2 + 2x_4}{7}$

Zadanie 9
Zbadano czasy wykonania (w sek.) pięciu losowo wybranych standardowych programów przy użyciu dwóch różnych systemów operacyjnych: A i B. Otrzymano wyniki:

Program	1	2	3	4	5
System A	5,0	4,5	7,0	7,0	8,5
System B	6,5	6,5	7,5	7,0	8,0

Można przyjąć, że różnica czasów wykonania losowo wybranych programów przy użyciu systemów A i B jest zmienną losową o rozkładzie normalnym. Czy można twierdzić, że wartość

średnia czasu wykonania losowo wybranego programu przy użyciu systemu A jest mniejsza niż przy użyciu systemu B? Przyjąć poziom istotności 0,05. Dokończ rozwiązanie.

1. Model: $D_i = X_i - Y_i$, $i = 1, 2,\ldots, 5$, są niezależnymi zmiennymi losowymi o rozkładzie $N(\mu, \sigma)$, gdzie $\mu = \mu_1 - \mu_2$, $\mu_1 = E(X_i)$ $\mu_2 = E(Y_i)$, $i = 1, 2, \ldots, 5$. Zmienna X_i oznacza czas wykonania programu i przy użyciu systemu A, zmienna Y_i oznacza czas wykonania programu i przy użyciu systemu B.
2. H_0: $\mu_D = 0$, H_1: $\mu_D < 0$,
3. Statystyka testowa: $T = \dfrac{\overline{D} - 0}{S}\sqrt{n}$ ma rozkład Studenta o 4 stopniach swobody ($n = 5$)
4. Obliczenia:
   Wartości zmiennych losowych $D_i = X_i - Y_i$

1	2	3	4	5
-1,5	-2	-0,5	0	0,5

   Średnia próbkowa $\bar{d} = \dfrac{-3{,}5}{5} = -0{,}7$
   Odchylenie standardowe z próbki $s = 1{,}04$
   Wartość statystyki testowej $t = \dfrac{-0{,}7}{1{,}04}\sqrt{5}$ = -1,51
   Zbiór krytyczny: [0,12 , 0,28]
5. Kwantyl $t_{0{,}95;\,4} = 2{,}1318$, stąd $C = \{t : t \le -2{,}1318\}$
6. Odpowiedź:
   Ponieważ wartość statystyki testowej nie należy do zbioru krytycznego, nie ma podstaw do odrzucenia hipotezy zerowej. Zatem nie ma podstaw by twierdzić, że wartość średnia czasu wykonania losowo wybranego programu przy użyciu systemu A jest mniejsza niż przy użyciu systemu B.

Zadanie 11
Zanotowano 9 czasów oczekiwania na połączenie z siecią teleinformatyczna (w sek.):

4,5 5,5 7,5 11,5 3,0 5,5 13,0 6,0 6,6.

Znajdź medianę i rozstęp międzykwartylowy dla zaobserwowanych czasów oczekiwania.

Rozwiązanie

Dane uporządkowane rosnąco

3,0 4,5 5,5 5,5 **6,0** 6,6 7,5 11,5 13,0
.

Mediana $x_{med} = 6{,}0$

Kwartyl dolny $Q_1 = (4{,}5 + 5{,}5)/2 = 5$

Kwartyl górny $Q_3 = (7{,}5 + 11{,}5)/2 = 9{,}5$

Rozstęp międzykwartylowy $IQR = Q_3 - Q_1 = 4{,}5$

Zadanie 13
Wśród stu losowo wybranych stacji paliw znalazło się 20 stacji, na których sprzedawane paliwo nie spełniało norm jakości. Wyznacz przybliżony 95% przedział ufności dla proporcji stacji paliw sprzedających paliwo nie spełniające norm jakości.

Rozwiązanie

Liczebność próby: $n = 100$
Liczba elementów $k = 20$

Estymator proporcji (proporcja z próbki): $\hat{p} = \frac{k}{n} = 0{,}2$

Warunki przybliżenia rozkładem normalnym:

$$\begin{cases} n\hat{p} = 20 > 5 \\ n(1-\hat{p}) = 80 > 5 \end{cases}$$

są spełnione

Poziom ufności 1- α = 0,95, stąd α = 0,05, $\frac{\alpha}{2} = 0{,}025$, $1 - \frac{\alpha}{2} = 0{,}975$

Kwantyl $z_{1-\alpha/2} = z_{0{,}975} = 1{,}96$

Przedział ufności

$$\left[\ \hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,\ \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ \right]$$

$$\left[\ 0{,}2 - 1{,}96\sqrt{\frac{0{,}2 * 0{,}8}{100}}\ ,\ 0{,}2 + 1{,}96\sqrt{\frac{0{,}2 * 0{,}8}{100}}\ \right]$$

[0,12 , 0,28]

Mamy 95% pewności, że liczba stacji, na których paliwo nie spełniało norm jakości jest liczbą z przedziału [0,12 , 0,28].

Zadanie 14
Poziom stresu wywołanego określonym bodźcem mierzony w teście psychotechnicznym dla kierowców jest zmienną losową o funkcji gęstości

$$f(x) = \begin{cases} \frac{1}{24}x & dla \quad x \in [4, 8] \\ 0 & dla \quad x \notin [4, 8] \end{cases}$$

Oblicz prawdopodobieństwo, że poziom stresu u losowo wybranego kierowcy przekroczy 6 (odpowiednich jednostek).

Rozwiązanie

Niech X będzie zmienną losową o podanej funkcji gęstości.

$$P(X > 6) = \int_6^\infty f(x)\,dx = \int_6^8 \frac{x}{24}\,dx = \frac{1}{48}x^2\Big|_6^8 = \frac{64-36}{48} = \frac{7}{12}$$

Zadanie 15
Zmienna losowa X ma rozkład normalny o wartości średniej 4 i odchyleniu standardowym 3. Niech $Y = 2X - 8$.

a) Znajdź $E(Y)$ oraz $Var(Y)$.
b) Wiedząc, że Y ma rozkład normalny znajdź $P(Y > 0)$.

Rozwiązanie

$E(Y) = 2E(X) - 8 = 2*4 - 8 = 0$

$Var(Y) = 4*Var(X) = 4*9 = 36.$

$Y \sim N(0, 6)$ $\quad P(Y > 0) = P(\frac{Y-0}{6} > 0) = P(Z > 0) = 1 - P(Z \le 0) = \Phi(0) = 0{,}5$

Zadanie 17
Dyrektor banku twierdzi, ze wartość średnia czasu obsługi klienta przy okienku kasowym wynosi 5 minut. Czasy obsługi różnych klientów są niezależnymi zmiennymi losowymi rozkładach normalnych z nieznaną wartością średnią oraz nieznanym odchyleniem standardowym. Na podstawie czasów obsługi 9 klientów obliczono średni czas obsługi $\bar{x} = 6{,}5$ minuty oraz odchylenie standardowe (próbkowe) s = 1,5 minuty. Czy na poziomie istotności 0,01 można zaprzeczyć twierdzeniu dyrektora?

Uzupełnij rozwiązanie:

1. H_0: $\mu = 5$, H_1: $\mu > 5$
2. $\alpha = 0{,}01$
3. Statystyka testowa $T = \frac{\bar{X} - \mu}{S}\sqrt{n}$ ma rozkład Studenta o 8 stopniach swobody, gdy H_0 jest prawdziwa
4. Wartość statystyki $t = \frac{6{,}5 - 5}{1{,}5}\sqrt{9} = 3$
5. Kwantyl $t_{0{,}99;\,8} = 2{,}8965$
6. Zbiór krytyczny $C = \{t : t \ge 2{,}8965\}$

Odpowiedź na pytanie i jej uzasadnienie:

Ponieważ wartość statystyki testowej należy do obszaru krytycznego odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej z prawdopodobieństwem popełnienia błędu pierwszego rodzaju nieprzekraczającym założonego poziomu istotności 0,01. Można zaprzeczyć twierdzeniu dyrektora. (Różnica między wartością hipotetyczną (5 min.) i wyznaczoną z próbki (6,5 min.) jest statystycznie istotna).