

Zad. 1 Badano zmianę poziomu płac pracowników pewnego przedsiębiorstwa w latach 2001-2002. Dla 12-osobowej próby pracowników zatrudnionych w tym przedsiębiorstwie w 2001 r. otrzymano średnią płacę 1240 zł i odchylenie standardowe 110 zł, a dla 10-osobowej próby innych pracowników zatrudnionych w tym przedsiębiorstwie w 2002 r. otrzymano średnią płacę 1480 zł i odchylenie standardowe 140 zł. Zakładamy, że płace w poszczególnych latach miały rozkłady normalne o równych wariancjach. Czy na podstawie tych danych można uznać, że średnie płace w 2002 r. wzrosły w porównaniu z 2001 r.? Przyjąć poziom istotności 0,05.

Rozw.

- Niech zmienna losowa X oznacza płacę pracownika w roku 2001, $X \sim N(\mu_1, \sigma_1)$, przy czym parametry rozkładu nieznane
- Niech zmienna losowa Y oznacza płacę pracownika w roku 2002, $Y \sim N(\mu_2, \sigma_2)$, przy czym parametry rozkładu nieznane
- $\sigma_1 = \sigma_2 = \sigma$
- Próbkę płac pracowników w roku 2001: liczebność $n_1 = 12$, średnia z próbki $\bar{x} = 1240$ zł., próbkowe odchylenie standardowe $s_1 = 110$ zł.
- Próbkę płac pracowników w roku 2002: liczebność $n_2 = 10$, średnia z próbki $\bar{y} = 1480$ zł., próbkowe odchylenie standardowe $s_2 = 140$ zł.
- Poziom istotności: $\alpha = 0,05$,
- Pytanie: czy na podstawie danych zadania można uznać, że średnie pacy wzrosły w roku 2002 w porównaniu z rokiem 2001?

1. Hipotezy: $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 < \mu_2$

2. Statystyka testowa:

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ma rozkład t-Studenta o 20 stopniach swobody, oznaczany jako t_{20} , o ile H_0 prawdziwa, gdzie

$$S_p = \sqrt{S_p^2}, \quad S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

3. Wartość statystyki testowej:

$$t = \frac{1240 - 1480}{\sqrt{\frac{11 \cdot 110^2 + 9 \cdot 140^2}{12 + 10 - 2}} \sqrt{\frac{1}{12} + \frac{1}{10}}} = -4,50583$$

4. , $H_1: \mu_1 < \mu_2 \Rightarrow$ zbiór krytyczny $C = (-\infty; -t_{1-\alpha,20}] = (-\infty; -t_{0,95;20}]$, $t_{0,95;20} = 1,7247$. Zatem

$$C = (-\infty; -1,7247]$$

5. Decyzja i jej uzasadnienie: $t \in C$, więc odrzucamy hipotezę zerową i przyjmujemy alternatywną na

poziomie istotności 0.05. Wyniki pomiarów skazują, że średnie płace wzrosły.

Zad. 2 Dwie formacje geologiczne porównano pod względem zawartości pewnego minerału. Uzyskano następujące dane:

formacja I	7,6	11,1	6,8	9,8	4,9	6,1	15,1
formacja II	4,7	6,4	4,1	3,7	3,9		

Zakładamy, że rozkłady zawartości tego minerału w obu formacjach są normalne z odchyleniami standardowymi równymi odpowiednio: 2 i 1. Czy można stwierdzić, że średnia zawartość tego minerału w pierwszej formacji jest istotnie większa od zawartości w drugiej formacji? Przyjąć poziom istotności 0,05.

Rozw.

- Niech zmienna losowa X oznacza zawartość minerału w formacji I, $X \sim N(\mu_1, \sigma_1)$, przy czym $\sigma_1 = 2$, średnia nieznana
- Niech zmienna losowa Y oznacza zawartość minerału w formacji II, $Y \sim N(\mu_2, \sigma_2)$, przy czym $\sigma_2 = 1$, średnia nieznana

- Próbka dla formacji I: 7,6 11,1 6,8 9,8 4,9 6,1 15,1

Stąd liczebność $n_1 = 7$, średnia z próbki $\bar{x} = \frac{61,4}{7} = 8,771$

- Próbka dla formacji II: 4,7 6,4 4,1 3,7 3,9

Stąd liczebność $n_2 = 5$, średnia z próbki $\bar{y} = \frac{22,8}{5} = 4,56$

- Poziom istotności: $\alpha = 0,05$
- Pytanie: Czy można stwierdzić, że średnia zawartość tego minerału w pierwszej formacji jest istotnie większa od zawartości w drugiej formacji?

Model 5.

1. Hipotezy: $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 > \mu_2$

2. Statystyka testowa

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1),$$

jeśli H_0 jest prawdziwa

3. Wartość statystyki testowej

$$z = \frac{8,771 - 4,56}{\sqrt{\frac{4}{7} + \frac{1}{5}}} = 4,794$$

4. Zbiór krytyczny $C = [z_{1-\alpha}, \infty) = \{z: z \geq z_{1-\alpha}\}$, gdzie $1 - \alpha = 0,95$, stąd

$$C = [z_{0,95}, \infty) = [1,645, \infty)$$

5. **Decyzja i jej uzasadnienie:** $z \in C$ więc odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną na poziomie istotności 0,05: wyniki pomiarów wskazują na to, średnia zawartość tego minerału w pierwszej formacji jest większa od zawartości w drugiej na poziomie istotności 0,05.

Zad. 3 Podczas sprawdzianu z ortografii 8 losowo wybranych dzieci popełniło: 1, 3, 2, 7, 6, 5, 4, 8 błędów. Przez miesiąc, drogą licznych dyktand, ćwiczono ortografię, po czym powtórzono sprawdzian na tej samej grupie dzieci. Tym razem dzieci te popełniły, odpowiednio, następującą liczbę błędów: 0, 1, 3, 5, 5, 3, 2, 4. Zakładamy, że rozkład liczby popełnionych błędów jest normalny. Czy na podstawie przedstawionych danych można stwierdzić, że dyktanda wpływają na poprawę ortografii? Przyjąć poziom istotności 0,01.

Rozw.

- (X, Y) – zmienna losowa dwuwymiarowa charakteryzująca dziecko w badanej populacji, gdzie X – liczba błędów z ortografii na pierwszym sprawdzianie, Y – liczba błędów na drugim sprawdzianie po dodatkowych dyktandach
- $D = X - Y \sim N(\mu_D, \sigma_D)$, gdzie $\mu_D = E(X) - E(Y)$
- $n = 8$ – liczebność próbki
- Poziom istotności: $\alpha = 0,01$
- Pytanie: Czy można stwierdzić, że dyktanda wpływają na poprawę ortografii?

Dziecko	1	2	3	4	5	6	7	8
Sprawdzian 1	1	3	2	7	6	5	4	8
Sprawdzian 2	0	1	3	5	5	3	2	4
Różnica	1	2	-1	2	1	2	2	4

1. Hipotezy: $H_0: \mu_D = 0$, $H_1: \mu_D > 0$

2. Statystyka testowa

$$T = \frac{\bar{D}}{\frac{s_D}{\sqrt{n}}} \sim t_{n-1},$$

jeśli H_0 prawdziwa.

3. Wartość statystyki testowej

$$t = \frac{\bar{d}}{\frac{s_D}{\sqrt{8}}} = ?$$

Różnica	d_i	1	2	-1	2	1	2	2	4
---------	-------	---	---	----	---	---	---	---	---

$$\sum_{i=1}^8 d_i = 13, \quad \bar{d} = \frac{13}{8} = 1,625, \quad \sum_{i=1}^8 d_i^2 = 35$$

$$s_D^2 = \frac{1}{8-1} \sum_{i=1}^8 (d_i - \bar{d})^2 = \frac{1}{7} \left(\sum_{i=1}^8 d_i^2 - \frac{(\sum_{i=1}^8 d_i)^2}{8} \right) = \frac{1}{7} \left(35 - \frac{13^2}{8} \right) = 1,98214$$

$$s_D = \sqrt{1,98214} = 1,40788$$

$$t = \frac{\bar{d}}{\frac{s_D}{\sqrt{8}}} = \frac{1,625}{1,40788} \sqrt{8} = 3,26462$$

$$4. H_1: \mu_D > 0 \Rightarrow C = [t_{1-\alpha, n-1}, \infty) = [t_{0,99;7}, \infty) = [2,9980; \infty)$$

5. **Decyzja i jej uzasadnienie:** $t \in C$ więc odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną na poziomie istotności 0,01: wyniki sprawdzianów wskazują na to, że dyktanda wpływają na poprawę ortografii.

Zad. 4 Spośród studentów pewnego wydziału uczelni wylosowano niezależnie 7 studentów IV roku i otrzymano dla nich następujące średnie oceny uzyskane w sesji egzaminacyjnej na I i IV roku:

student	1	2	3	4	5	6	7
I rok	3,5	4,0	3,7	4,6	3,9	3,0	3,5
IV rok	4,2	3,9	3,8	4,5	4,2	3,4	3,8

Zakładamy, że rozkład ocen jest normalny. Czy te rezultaty potwierdzają hipotezę, że średnie wyniki po IV roku są lepsze niż po I roku? Przyjąć poziom istotności 0,05.

Rozw.

- (X, Y) – zmienna losowa dwuwymiarowa charakteryzująca losowo wybranego studenta IV roku w badanej populacji studentów, gdzie X oznacza średnią ocenę uzyskaną w sesji na I roku studiów, a Y jest średnią oceną uzyskaną w sesji na IV roku studiów
- $D = X - Y \sim N(\mu_D, \sigma_D)$, gdzie $\mu_D = E(X) - E(Y)$
- $n = 7$ – liczebność próbki
- Poziom istotności: $\alpha = 0,05$

Pytanie: Czy można stwierdzić, że średnie wyniki po IV roku są lepsze niż po I roku?

1. Hipotezy: $H_0: \mu_D = 0, \quad H_1: \mu_D < 0$
2. Statystyka testowa

$$T = \frac{\bar{D}}{\frac{s_D}{\sqrt{n}}} \sim t_{n-1},$$

jeśli H_0 prawdziwa.

3. Wartość statystyki testowej

$$t = \frac{\bar{d}}{\frac{s_D}{\sqrt{7}}} = ?$$

student	1	2	3	4	5	6	7
I rok	3,5	4,0	3,7	4,6	3,9	3,0	3,5
IV rok	4,2	3,9	3,8	4,5	4,2	3,4	3,8

Różnica	d_i	- 0,7	0,1	- 0,1	0,1	- 0,3	- 0,4	- 0,3
---------	-------	-------	-----	-------	-----	-------	-------	-------

$$\sum_{i=1}^7 d_i = -1,6, \quad \bar{d} = \frac{-1,6}{7} = -0,22857 \quad \sum_{i=1}^7 d_i^2 = 2,48$$

$$s_D^2 = \frac{1}{7-1} \sum_{i=1}^7 (d_i - \bar{d})^2 = \frac{1}{6} \left(\sum_{i=1}^7 d_i^2 - \frac{(\sum_{i=1}^7 d_i)^2}{7} \right) = \frac{1}{6} \left(2,48 - \frac{1,6^2}{7} \right) = 0,35238$$

$$s_D = \sqrt{0,35238} = 0,59362$$

$$t = \frac{\bar{d}}{\frac{s_D}{\sqrt{7}}} = \frac{-0,22857}{0,59362} \sqrt{7} = -1,01873$$

4. $H_1: \mu_D < 0 \Rightarrow C = (-\infty, -t_{1-\alpha, n-1}] = (-\infty, -t_{0,95;6}] = (-\infty, -1,9432]$

5. **Decyzja i jej uzasadnienie:** $t \notin C$ więc nie można odrzucić hipotezy zerowej i przyjąć hipotezy alternatywnej na poziomie istotności 0,05: nie ma podstaw do twierdzenia, że średnie wyniki po IV roku są lepsze niż po I roku.

Zad. 5 Badając wpływ nowego leku na poprawę stanu zdrowia chorych na cukrzycę, podano 300 losowo wybranym chorym ten nowy lek i u 240 z nich stwierdzono, po ustalonym okresie leczenia, powrót poziomu cukru w organizmie do normy. Natomiast w grupie 200 chorych leczonych lekami tradycyjnymi, cukier powrócił do normy u 124 pacjentów. Na poziomie istotności 0,01 zweryfikować hipotezę, że nowy lek jest skuteczniejszy od leków tradycyjnych.

Rozw.

- Niech zmienna losowa X przyjmuje wartości 1, jeśli nowy lek poprawia stan chorego na cukrzycę, oraz 0 w przypadku przeciwnym. Niech $p_1 = P(X = 1)$, stąd

$$X \sim \text{Bin}(1, p_1), p_1 \in (0,1)$$

- Niech zmienna losowa Y przyjmuje wartości 1, jeśli tradycyjny lek poprawia stan chorego na cukrzycę, oraz 0 w przypadku przeciwnym. Niech $p_2 = P(Y = 1)$, stąd

$$Y \sim \text{Bin}(1, p_2), p_2 \in (0,1)$$

- Próbka o liczności $n_1 = 300$, $k_1 = 240$ pacjentów leczonych nowym lekiem uzyskało poprawę
- Próbka o liczebności $n_2 = 200$, $k_2 = 124$ pacjentów leczonych tradycyjnymi lekami uzyskało poprawę
- Poziom istotności: $\alpha = 0,01$
- Pytanie:** zweryfikować hipotezę, że nowy lek jest skuteczniejszy od leków tradycyjnych.

Założenia zadania sugerują model 11 – test o różnicy proporcji dwóch populacji, jeśli spełnione są warunki:

$$n_1 \hat{p}_1 = 300 \cdot \frac{240}{300} = 240 \geq 5, \quad n_1(1 - \hat{p}_1) = 300 \left(1 - \frac{240}{300}\right) = 60 \geq 5$$

$$n_2 \hat{p}_2 = 200 \cdot \frac{124}{200} = 124 \geq 5, \quad n_2(1 - \hat{p}_2) = 200 \left(1 - \frac{124}{200}\right) = 76 \geq 5$$

Możemy stosować model 11.

1. Hipotezy: $H_0: p_1 = p_2, \quad H_1: p_1 > p_2$

2. Statystyka testowa

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \text{bliski rozkładowi } N(0,1),$$

jeśli H_0 jest prawdziwa, oraz spełnione są warunki $n_1 \hat{p}_1 \geq 5, n_1(1 - \hat{p}_1) \geq 5, n_2 \hat{p}_2 \geq 5, n_2(1 - \hat{p}_2) \geq 5$ gdzie

$$\hat{p}_1 = \frac{X_1 + \dots + X_{n_1}}{n_1} := \frac{K_1}{n_1}, \quad \hat{p}_2 = \frac{Y_1 + \dots + Y_{n_2}}{n_2} := \frac{K_2}{n_2},$$

są proporcjami empirycznymi, $X_j = 1(0)$, jeśli j -ty pacjent leczonym nowym lekiem uzyskał poprawę (nie uzyskał poprawy), $Y_j = 1(0)$, jeśli j -ty pacjent leczonym tradycyjnymi lekami uzyskał poprawę (nie uzyskał poprawy),

$$\hat{p} = \frac{K_1 + K_2}{n_1 + n_2}$$

3. Wartość statystyki testowej:

$$z = \frac{\frac{240}{300} - \frac{124}{200}}{\sqrt{\frac{364}{500} \left(1 - \frac{364}{500}\right)} \sqrt{\frac{1}{300} + \frac{1}{200}}} = \frac{0,8 - 0,62}{\sqrt{0,728 \frac{5}{600}}} = \frac{0,18}{0,0406218} = 4,43112$$

4. Zbiór krytyczny $C = [z_{1-\alpha}, \infty) = [z_{0,99}, \infty) = [2,32635; \infty)$

5. **Decyzja i jej uzasadnienie:** $z \in C$ więc odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną na poziomie istotności 0,01: nowy lek jest skuteczniejszy od leków tradycyjnych, przy poziomie 0,01.

Zad. 6 Badania grupy krwi 200 losowo wybranych osób dały następujące wyniki:

- grupę A miało 37 osób,
- grupę B - 52 osoby,
- grupę AB - 66 osób,
- grupę 0 miało 45 osób.

Czy na podstawie tych wyników „można przyjąć” hipotezę o równomiernym rozkładzie wszystkich grup krwi? Przyjąć poziom istotności 0,05.

Rozw.

- X – grupa krwi losowo wybranej osoby z badanej populacji, cecha jakościowa mająca 4 możliwe kategorie: grupy krwi
- Liczebność próby (próbki) = 200
- Rozkład cechy X określony tabelą:

Grupa krwi	A	AB	B	0
Prawdop. grupy krwi	p_1	p_2	p_3	p_4

gdzie $p_i \in (0,1)$, $p_1 + p_2 + p_3 + p_4 = 1$

- Poziom istotności $\alpha = 0,05$

1. Hipotezy: $H_0: p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$,

H_1 : rozkład cechy X nie jest równomierny

2. Statystyka testowa

$$\chi^2 = \sum_{i=1}^4 \frac{(N_i - np_i)^2}{np_i} = \text{ma rozkład bliski rozkładowi } \chi_3^2$$

jeśli hipoteza zerowa jest prawdziwa i $N_i \geq 5$, $i = 1,2,3,4$. N_i – liczba osób z grupą krwi i -tą spośród

200-tu losowo wybranych osób z badanej populacji.

3. Wartość statystyki testowej $\chi^2_{obs} = ?$

nr	kategoria	Liczebność kategorii n_i	p_i	$np_i = 200p_i$	$(n_i - np_i)^2 / np_i$
1	A	37	0,25	50	$\frac{(37 - 50)^2}{50} = 3,38$
2	B	52	0,25	50	$\frac{(52 - 50)^2}{50} = 0,08$
3	AB	66	0,25	50	$\frac{(66 - 50)^2}{50} = 5,12$
4	O	45	0,25	50	$\frac{(45 - 50)^2}{50} = 0,5$
Suma =					$\chi^2_{obs} = 9,08$

4. Zbiór krytyczny $C = [\chi^2_{1-\alpha,3}, \infty) = [\chi^2_{0,95;3}, \infty) = [7,8147; \infty)$

5. **Decyzja i jej uzasadnienie:** $\chi^2_{obs} \in C$ więc odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną na poziomie istotności 0,05: można stwierdzić, że rozkład grup krwi nie jest równomierny (jednostajny).

Zad. 7 Wysłunięto przypuszczenie, że udziały w rynku trzech firm kosmetycznych: A, B i C wynoszą, odpowiednio, 5%, 24% i 11%. W celu weryfikacji tej hipotezy wylosowano 200 produktów kosmetycznych znajdujących się na rynku i otrzymano następujące wyniki:

Firma	A	B	C	pozostałe firmy
liczba produktów	20	50	14	116

Na poziomie istotności 0,01 zweryfikować wysuniętą hipotezę.

Rozw.

- X – producent losowo wybranego produktu kosmetycznego
- Liczebność próby (próbki) = 200
- Rozkład cechy X określony tabelą:

Firma	A	B	C	pozostałe
Prawdopodobieństwo wylosowania produktu danej firmy	p_1	p_2	p_3	p_4

gdzie $p_i \in (0,1)$, $p_1 + p_2 + p_3 + p_4 = 1$

- Poziom istotności $\alpha = 0,01$

- Hipotezy: $H_0: p_1 = 0,05; p_2 = 0,24; p_3 = 0,11; p_4 = 1 - (0,05 + 0,24 + 0,11) = 0,6$,
 $H_1: \text{rozkład cechy } X \text{ jest inny niż w } H_0$

- Statystyka testowa

$$\chi^2 = \sum_{i=1}^4 \frac{(N_i - np_i)^2}{np_i} = \text{ma rozkład bliski rozkładowi } \chi^2_3,$$

jeśli hipoteza zerowa jest prawdziwa i $N_i \geq 5$, $i = 1,2,3,4$. N_i – liczba produktów firmy i-tej w losowej próbie 200-tu produktów

- Wartość statystyki testowej:

$$\chi^2_{\text{obs}} = ?$$

nr	Firma	Liczebność kategorii n_i	p_i	$np_i = 200p_i$	$(n_i - np_i)^2 / np_i$
1	A	20	0,05	10	$\frac{(20 - 10)^2}{10} = 10$
2	B	50	0,24	48	$\frac{(50 - 48)^2}{48} = 0,0833$
3	C	14	0,11	22	$\frac{(14 - 22)^2}{22} = 2,9091$
4	Inne	116	0,60	120	$\frac{(116 - 120)^2}{120} = 0,1333$
Suma =					$\chi^2_{\text{obs}} = 13,1257$

- Zbiór krytyczny $C = [\chi^2_{1-\alpha,3}, \infty) = [\chi^2_{0,99,3}, \infty) = [11,3449; \infty)$
- Decyzja i jej uzasadnienie:** $\chi^2_{\text{obs}} \in C$ więc odrzucamy hipotezę zerową i przyjmujemy hipotezę alternatywną na poziomie istotności 0,05: można stwierdzić, że udziały w rynku kosmetyków badanych firm są inne niż podane w hipotezie zerowej.