Zadania ze statystyki cz. 7.

Zad.1
Z populacji wyłoniono próbę wielkości 64 jednostek. Średnia arytmetyczna wartość cechy wyniosła 110, zaś odchylenie standardowe 16. Należy wyznaczyć przedział ufności parametru populacji μ, przy poziomie istotności $\alpha \leq 0{,}05$. Jaka będzie wielkość tego przedziału, gdy zmniejszymy poziom istotności do $\alpha \leq 0{,}01$.

Zad. 2
Z tej samej populacji wybrano próbę 26 jednostek. Średnia wyniosła 110, odchylenie standardowe wyniosło 16. Należy wyznaczyć przedział ufności parametru populacji μ, przy poziomie istotności $\alpha \leq 0{,}05$. Jaka będzie wielkość tego przedziału, gdy zmniejszymy poziom istotności do $\alpha \leq 0{,}01$.

Zad. 3
Przeprowadzono badania dwóch prób pochodzących z tej samej populacji, różniących się wielkością. Pierwsza próba liczyła sobie 100 a druga 400 jednostek. Uzyskano następujący rozkład proporcji badanej cechy w pierwszej próbie p=0,4, a w drugiej p=0,45. Jakie będą przedziały ufności dla uzyskanych frakcji, przy poziomie istotności $\alpha \leq 0{,}05$.

Zad. 4
Należy określić rodzaj testu (jedno czy dwustronny) oraz wartości krytyczne **z lub t** dla określonych hipotez i ich poziomów istotności:

a. $H_0: \mu_1 = \mu_2;\ \ H_1: \mu_1 \neq \mu_2;\ \ \alpha \leq 0{,}05\ \ \ N = 25$
b. $H_0: \mu_1 > \mu_2;\ \ H_1: \mu_1 \leq \mu_2;\ \ \alpha \leq 0{,}01\ \ \ N = 28$
c. $H_0: \mu_1 \leq \mu_2;\ \ H_1: \mu_1 > \mu_2;\ \ \alpha \leq 0{,}01$ N = 100
d. $H_0: \mu_1 - \mu_2 = 0;\ \ H_1: \mu_1 - \mu_2 > 0;\ \ \alpha \leq 0{,}001$ N = 10

Zad.5
Średnie zarobki osób zatrudnionych w pewnej fabryce wynoszą 3300 zł a odchylenie standardowe 1500 zł. Zarobki 250 pracowników pracujących w montowni tej fabryki wynoszą przeciętnie 3100 zł. Czy ich zarobki różnią się istotnie statystycznie od zarobków wszystkich pracowników (przy poziomie istotności $\alpha \leq 0{,}05$ i $\alpha \leq 0{,}01$)?

Zad. 6
Średnia ocen studentów I roku wszystkich kierunków studiów na Uniwersytecie Wrocławskim wyniosła 4,2. Przebadano próbę 20 studentów socjologii U.Wr. uzyskując średnią 4,27 oraz odchylenie standardowe 0,6 Czy różnica między średnimi jest istotna statystycznie, na poziomie istotności $\alpha = 0{,}01$ (czy jest efektem błędu losowego próby)?

Zad. 7
Przeprowadzono pewne badanie nad ocenami uczniów. Średnia ocena wszystkich uczniów w szkole wyniosła 4,1, zaś wylosowanej grupy 3,90. Odchylenie standardowe

średniej próby wyniosło 0,2. Widząc, że dla 100 osobowej próby różnica między średnimi jest istotna statystycznie przy poziomie istotności $\alpha \leq 0,05$, należy ustalić:

a. czy jest też istotna na poziomie istotności $\alpha \leq 0,01$.
b. Ile powinna wynosić minimalna różnica między średnimi, aby była istotna na poziomie $\alpha \leq 0,001$.
c. Jaka powinna być minimalna liczebność próby, aby różnica 0,1 była istotna statystycznie przy poziomie istotności $\alpha \leq 0,01$ (odchylenie standardowe 0,2).

Zad. 8
Badania dotyczące palenia wykazały, że nałóg ten dotyczy 72% studentów. Należy wykazać czy:

a. Problem palenia dotyczy większości studentów, wiedząc, że próba do badań wyniosła 150, przy poziomie istotności $\alpha \leq 0,05$.
b. Ilość palących wśród studentów jest większa niż w reszcie społeczeństwa, wiedząc, że palenie dotyczy 60% członków społeczeństwa, przyjmując poziom istotności $\alpha \leq 0,01$.

PODSUMOWANIE ZAJĘĆ

Estymacja i przedziały ufności

- Oszacowanie lub estymacja to dział wnioskowania statystycznego, będący zbiorem procedur statystycznych, pozwalających określić wielkość nieznanego parametru lub postać rozkładu zmiennej losowej na podstawie badania próby losowej.

Na podstawie badań (obserwacji) próby szacuje się (estymuje się) wiele wartości w populacji, np.:

- wielkość poparcia dla partii X lub Y;
- odsetek osób kupujących dany produkt;
- liczbę osób dotkniętych ubóstwem.

Problem: dlaczego pojawiają się różne wyniki badań?

- W wyniku prowadzenia pomiarów na próbach zawsze istnieje pewna zmienność wyników, będąca efektem zmienności cechy w populacji. Zmienność wyników związana jest z błędem próby.

- Ze względu na poszukiwaną cechę rozkładu metody estymacji można podzielić na:

1. **estymację parametryczną** – szacuje się parametr populacji
2. **estymację nieparametryczną** – szacuje się proporcje rozkładu cechy w populacji

Ze względu na sposób szacowania poszukiwanej cechy rozkładu, można wyróżnić:

1. **estymację punktową** – szacowanie konkretnej wartości parametru lub proporcji rozkładu cechy w populacji.
2. **estymację przedziałową** – szacowanie przedziału ufności w jakim mieści się parametr lub proporcja rozkładu cechy w populacji.

Estymacja przedziałowa – przedziały ufności dla średnich z dużych prób

- Estymacja przedziałowa polega na wyznaczaniu pewnego przedziału liczbowego, wewnątrz którego z określonym prawdopodobieństwem znajduje się estymowany parametr.
- Przedział ten nazywamy **przedziałem ufności**, a prawdopodobieństwo tego, że znajdzie się w nim estymowany parametr wyznacza **współczynnik ufności**.
- Rozkład ze średnich z próby zbliża sie coraz bardziej do normalnego wraz ze wzrostem liczebności próby.
- Dlatego też dla dużych prób stosuje sie rozkład normalny w celu oszacowania wyniku.
- Do obliczenia błędu standardowego stosuje sie odchylenie rozkładu z próby średnich, czyli:

$$s_{\bar{x}} = \frac{s}{\sqrt{N}} \qquad\qquad s_{\bar{x}} = \frac{s}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}}$$

Przedział ufności dla średnich z dużych prób.

- Stosowany gdy N>30

$$P\left(\bar{x} - z_\alpha \frac{s}{\sqrt{N}} < \mu < \bar{x} + z_\alpha \frac{s}{\sqrt{N}}\right) = 1 - \alpha$$

z_α – *wartość krytyczna rozkładu normalnego dla poziomu* α

$\frac{s}{\sqrt{N}}$ – *wielkość błędu standardowego*

$1 - \alpha$ – *współczynnik ufności*

α – *poziom istotności*

Przykład

Dla wylosowanych w sposób losowy 100 studentów średnia z ocen z semestru zimowego wyniosła 4,1, odchylenie standardowe zaś 1,2. Jaki będzie przedział ufności tej średniej, przyjmując, że chcemy określić jego wielkość na poziomie istotności 0,05.

$$\alpha = 0{,}05 \quad 1 - \alpha = 0{,}95 \quad z_\alpha = 1{,}96 \quad N = 100$$

$$P\left(4{,}1 - 1{,}96\frac{1{,}2}{\sqrt{100}} < \mu < 4{,}1 + 1{,}96\frac{1{,}2}{\sqrt{100}}\right) = 1 - 0{,}05$$

$$P(4{,}1 - 0{,}2352 < \mu < 4{,}1 + 0{,}2352) = 0{,}95$$

$$P(3{,}8648 < \mu < 4{,}3352) = 0{,}95$$

- Gdy N <=30

$$P\left(\bar{x} - t_\alpha \frac{s}{\sqrt{N-1}} < \mu < \bar{x} + t_\alpha \frac{s}{\sqrt{N-1}}\right) = 1 - \alpha$$

$t_\alpha - wartość\ t\ dla\ okreslonego\ poziomu\ \alpha\ i\ df = N - 1$

Przykład

- Dla wylosowanych w sposób losowy 20 studentów średnia z ocen z semestru zimowego wyniosła 4,1, odchylenie standardowe zaś 1,2. Jaki będzie przedział ufności tej średniej, przyjmując, że chcemy określić jego wielkość na poziomie istotności 0,05.

$$t_{0,05} = 2{,}09\ przy\ df = 19 N = 20$$

$$P\left(4{,}1 - 2{,}09 \frac{1{,}2}{\sqrt{20-1}} < \mu < 4{,}1 + 2{,}09 \frac{1{,}2}{\sqrt{20-1}}\right) = 0{,}95$$

$$P(3{,}56 < \mu < 4{,}63) = 0{,}95$$

Przedziały ufności proporcji dla dużych prób

- Metoda estymacji θ przez p

$$P\left(p - z_\alpha \sqrt{\frac{pq}{N}} < \theta < p + z_\alpha \sqrt{\frac{pq}{N}}\right) = 1 - \alpha$$

$\theta - proporcja\ cechy\ w\ populacji$
$p - proporcja\ cechy\ w\ próbie$
$q - odwrotność\ cechy\ w\ próbie, tak\ że\ q = 1 - p$

- Metoda „ostrożna"

$$pq = max \rightarrow p = q = 0{,}5$$

$$P\left(p - z_\alpha \sqrt{\frac{(0{,}5)(0{,}5)}{N}} < \theta < p + z_\alpha \sqrt{\frac{(0{,}5)(0{,}5)}{N}}\right) = 1 - \alpha$$

Przykład

Przed kampanią wyborczą przeprowadzono sondaż, którego celem było oszacowanie odsetka osób popierających kandydata partii XYZ na prezydenta. Wybrano 1000 osobową próbę. Wyniki sondażu wskazywały na poparcie 55% wyborców. Ile wynosić będzie przedział ufności, przy poziomie istotności $\alpha \leq 0{,}05$?

$$\alpha = 0{,}05 \quad 1 - \alpha = 0{,}95 \quad z_\alpha = 1{,}96 \quad N = 1000$$

$$P\left(p - z_\alpha\sqrt{\frac{pq}{N}} < \theta < p + z_\alpha\sqrt{\frac{pq}{N}}\right) = 1 - \alpha$$

$$p = 55\% = 0{,}55 \qquad q = 100\% - 55\% = 45\% = 0{,}45$$

$$P(0{,}55 - 1{,}95\sqrt{\frac{0{,}55 * 0{,}45}{1000}} < \theta < 0{,}55 + 1{,}95\sqrt{\frac{0{,}55 * 0{,}45}{1000}}) = 1 - 0{,}05$$

$$P(0{,}5192 < \theta < 0{,}5808) = 0{,}95$$

Testy statystyczne

Test *z* i *t* dla jednej próby

- Badają , czy różnica między średnimi z próby a populacji jest wystarczająco duża, by wyeliminować wahania wynikające z losowego doboru próby.
- Wymaga podjęcia decyzji o charakterze testu (jedno czy dwustronny).
- Wymaga przyjęcia określonego poziomu istotności α (prawdopodobieństwa popełnienia błędu I rodzaju), wyznaczającego **wartość krytyczną** testu (np. $\alpha \leq 0{,}05$, wartość krytyczna z = =+-1,96 przy teście dwustronnym lub 1,65 przy teście jednostronnym)

1. **Testy parametryczne - test z dla jednej próby przy znanym σ**

- Test z stosujemy gdy:
  1. próba jest duża (N>30)
  2. znane jest odchylenie standardowe populacji – **σ**
  3. populacja ma rozkład normalny

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{N}}}$$

$\bar{x} -$ *testowana wartość (średnia z próby)*
$\mu -$ *średnia populacji*
$\sigma -$ *odchylenie standardowe populacji*
$N -$ *wielkość próby*

2. **Testy parametryczne - test *z* dla jednej próby przy nieznanym σ**

- Test *z* stosujemy gdy:
    1. próba jest duża (N>30)
    2. nieznane jest odchylenie standardowe populacji – **σ**
    3. rozkład cechy w populacji jest normalny

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{N}}}$$

$s -$ *odchylenie standardowe próby*

3. **Testy parametryczne - test *t* dla jednej próby przy znanym σ**

- Test *t* stosujemy gdy:
  1. próba jest mała (N ≤ 30)
  2. gdy znane jest odchylenie standardowe populacji σ
  3. rozkład cechy jest nieznany

$$t = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{N}}} \quad dla\ df = N - 1$$

4. **Testy parametryczne - test *t* dla jednej próby przy nieznanym σ**

- Test *t* stosujemy gdy:
  1. próba jest mała (N ≤ 30)
  2. gdy nieznane jest odchylenie standardowe populacji σ
  3. rozkład cechy jest nieznany

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{N}}} \ \ dla\ df = N - 1$$

Test *z* dla frakcji w populacji

- **Test z dla frakcji stosujemy gdy:**
  1. **próba jest duża, co oznacza, że musi zostać spełniony warunek:**

     $$np > 5\ oraz\ nq > 5$$

  2. rozkład cechy może być nieznany w populacji

$$z = \frac{p - p_0}{\sqrt{\frac{p_0 q_0}{N}}}$$

$p - odsetek\ obserwacji\ w\ próbie\ spełniający\ oczekiwany\ warunek$
$p_0 - odsetek\ obserwacji\ spełniający\ warunek\ hipotezy\ zerowej\ H_0$
$q_0 - odsetek\ obserwacji\ nie\ spełniających\ warunku\ hipotezy\ zerowej\ H_0$
$n - liczebność\ frakcji\ n * p > 5\, n * q > 5$