PRZEDZIAŁ UFNOŚCI

- informuje nas „na ile możemy zaufać danej zmiennej”

- pokazuje nam, że poszukiwana przez nas wartość mieści się w pewnym przedziale z założonym prawdopodobieństwem.

- jest mocno związany z teorią estymacji w statystyce (estymacja - szacowanie pewnych wartości w całej badanej zbiorowości na podstawie jedynie wycinka tej zbiorowości - próby.)

• Po co nam to?

Posłużymy się tu przykładem średniej do opisu. Chcemy obliczyć jaki jest poziom danej cechy w populacji. Co robimy?

PRZEPROWADZAMY BADANIA NA PRÓBIE „WYCINKU"

co dostajemy?

Wynik średni, średnią pewnej cechy

czy to wystarczy?

Nie przebadaliśmy całej populacji, chcemy obliczyć wartość na podstawie próby badawczej (części populacji), otrzymujemy średni poziom cechy, ale

NIE MOŻEMY NA JEGO PODSTAWIE WNIOSKOWAĆ W CAŁEJ POPULACJI TAKIEJ ŚREDNIEJ WARTOŚCI TEJ CECHY.

- Na ile jest ona zbliżona?

Nie wiadomo do końca, ALE MOŻNA WYZNACZYĆ

tzw. PRZEDZIAŁ UFNOŚCI

Na podstawie próby możemy wyznaczyć przedziały w których z założonym prawdopodobieństwem (np. 67%) mieści się prawdziwa wartość badanej miary.

np. Naukowiec chce sprawdzić poziom wiedzy wśród polskich studentów, przeprowadził badanie na pewnej próbie. Średni poziom wiedzy wyniósł 80. Za pomocą obliczeń statystycznych wykazał, że z prawd. 96% poziom wiedzy polskich studentów mieści się pomiędzy 76 – 84.

1. Przeprowadzamy badanie na próbie

2. Z pewnym prawdopodobieństwem podajemy w jakim PRZEDZIALE mieści się poszukiwana przez nas wartość.

Na przedział ufności wpływa:

- wybrane prawdopodobieństwo
- liczebność próby

Podsumowanie: PRZEDZIAŁ UFNOŚCI DOSTARCZA NAM ZAKRES (OD DO), W KTÓRYM Z ZAŁOŻONYM PRAWDOPOD. ZNAJDUJE SIĘ NASZA POSZUKIWANA WARTOŚĆ W POPULACJI.

PRZEDZIAŁ UFNOŚCI DLA ŚREDNIEJ

- Wariancja może być znana lub nieznana.

- Korzystamy wyliczając wartość dystrybuanty z rozkładu normalnego. (Centralne tw. graniczne)

! nie korzystamy jeżeli $n < 30$

Czego potrzebujemy? - jaki rozkład danej miary w populacji

- błąd standardowy

- założony przez nas poziom prawdopodobieństwa

$N(m, \sigma^2)$

$N(m, \frac{s^2}{n}) \rightarrow$ tw. dla średniej

Przedział ufności (x_1, x_2):

$$P(x_1 < m < x_2) = 1-\alpha$$

liczby, które możemy dodać i odjąć od naszej liczby

x_1, x_2 - f. wyznaczone na podstawie próby losowej

$1-\alpha$ - to prawdopodobieństwo, że rzeczywista wartość parametru m w populacji znajduje się w wyznaczonym przez nas przedziale ufności.

m - wartość śred. zmiennej losowej (cechy) X.

! zmienna X ma rozkład $N(m, \sigma^2)$, wtedy

$$\overline{X} \sim N\left(m, \frac{\sigma^2}{n}\right)$$

lub jest na tyle duża, że ma w przybliżeniu ten rozkład.

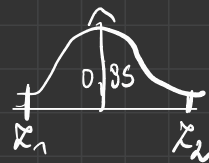
$\Rightarrow$ zmienna $\dfrac{\overline{X}-m}{\sigma}\sqrt{n}$ ~ma rozkład $N(0,1)$

Przykład: poziom ufności = 0,95

ZNAMY WARIANCJE i $N \geq 30$

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\bar{X}-m}{\sigma}\sqrt{n} < \underbrace{z_{1-\frac{\alpha}{2}}}_{\text{kwantyl}}\right) = 1-\alpha = 0{,}95$$

odchylenie standardowe



PRZEKSZTAŁCAMY:

$$P\left(-z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \bar{X}-m < z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha = 0{,}95$$

$$P\left(\bar{X} - z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}} < m < \bar{X} + z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}\right) = 0{,}95$$

otrzymujemy nierówność dla nieznanej średniej m

WAŻNE!

PRZEDZIAŁ UFNOŚCI TO: $\left(\bar{X} - \underbrace{z_{1-\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}}_{\text{margines błędu}}, \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$

Długość p. ufności: $2\Delta = 2 z_{\alpha}\frac{\sigma}{\sqrt{n}}$

NIE ZNAMY WARIANCJI LUB $n < 30$

- korzystamy z rozkładu t-studenta i obliczamy wariancję z próby (s)

PRZEDZIAŁ UFNOŚCI

$$\underbrace{P(L < t < P)}_{\text{wartości krytyczne}} = \underbrace{1-\alpha}_{\text{poziom ufności}}$$

wartości krytyczne

poziom ufności

% jak bardzo dokładny jest p. ufności

α – poziom istotności, czyli max. prawdopodob. że popełnimy błąd pierwszego rodzaju (odrzucenie prawdziwej hipotezy zerowej), stwierdzimy, że nasz przedział nie zawierał szukanego parametru choć go zawierał.

Przedział ufności dla średniej:

$$\text{Dla znanej } \sigma \text{ i dowolnego } n: (\overline{X} - u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})$$

$$\text{Dla nieznanego } \sigma \text{ i } n \leq 30: (\overline{X} - u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \overline{X} + u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}})$$

$$\text{Dla nieznanego } \sigma \text{ i } n > 30: (\overline{X} - t_{\alpha,n-1}\frac{s}{\sqrt{n}}, \overline{X} + t_{\alpha,n-1}\frac{s}{\sqrt{n}})$$

Minimalna liczebność próby:

$$n \geq \frac{u_\alpha^2 \sigma^2}{d^2}$$

Poziom istotności a poziom ufności:

Relacja między tymi pojęciami jest prosta.

poziom istotności : α

poziom ufności : $1-\alpha$

Typowe poziomy ufności:

Najczęściej spotykanymi poziomami ufności są 99%, 98%, 95% oraz 90%. **Im większy poziom ufności tym szerszy przedział** – skoro chcemy mieć większą pewność, że wartość parametru leży w naszym przedziale to musimy go zwiększyć.

Po co nam przedział ufności?

Przedział ufności informuje nas na ile możemy ufać naszym wyliczeniom, np. dotyczącymi średniej. Załóżmy, że chcemy wyliczyć średnią wagę studentów w Polsce. Jeżeli chcielibyśmy wyliczyć dokładną wartość musielibyśmy przebadać wszystkich studentów w Polsce. Oczywiście dałoby się to zrobić ale badania byłyby kosztowne i czasochłonne. Dlatego chcemy wybrać reprezentatywną grupę studentów i **estymować** (oszacować) wartość średniej. Jednak wybierając grupę, może się zdarzyć, że trafiliśmy lepiej lub gorzej. Z tego powodu zamiast podawać dokładną wartość możemy podać przedział i powiedzieć, że na 99% średnia jest w tym przedziale a to już daje nam obraz wagi całej populacji studentów

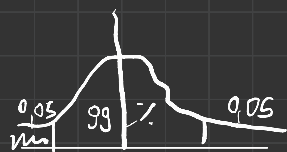
· Zapytano 16 studentów ile litrów kawy wypyli w tym tygodniu. Przedział ufności dla wartości oczekiwanej na poziomie 0,1 z próby wyniósł $(2, 8)$. Oblicz średnią i wariancję.

$n < 30$ i σ^2 nieznane – t-studenta

$$m \in \left(\bar{X} - t_{\alpha, n-1} \frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha, n-1} \frac{s}{\sqrt{n}} \right)$$

zarówno prawy i lewy kraniec jest tak samo odległy od $\bar{X}$.

$$\bar{X} = \frac{2+8}{2} = 5$$



$$t_{0,1,15} = 1,753$$

$$5 - 1,753 \cdot \frac{s}{4} = 2$$

$$s = \frac{3 \cdot 4}{1,753}$$

$$s^2 = \left(\frac{12}{1,753} \right)^2$$