

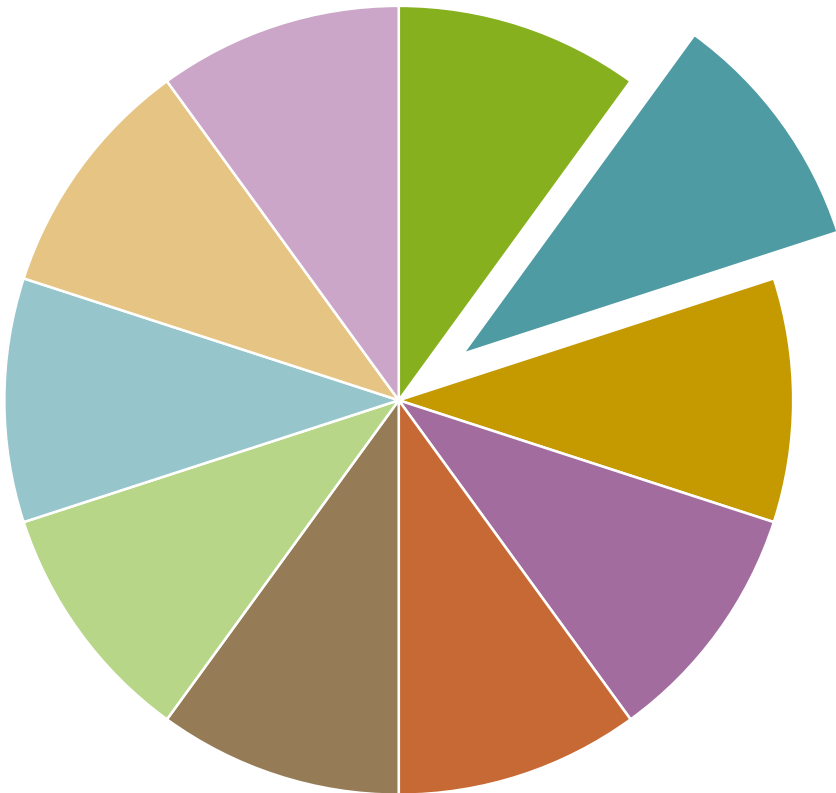
Barbara Więckowska

Próba badawcza

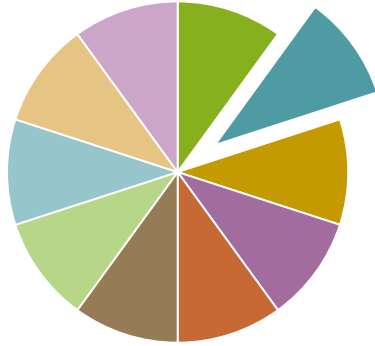
**dobór i wielkość**

czyli od czego zacząć przygodę ze statystyką

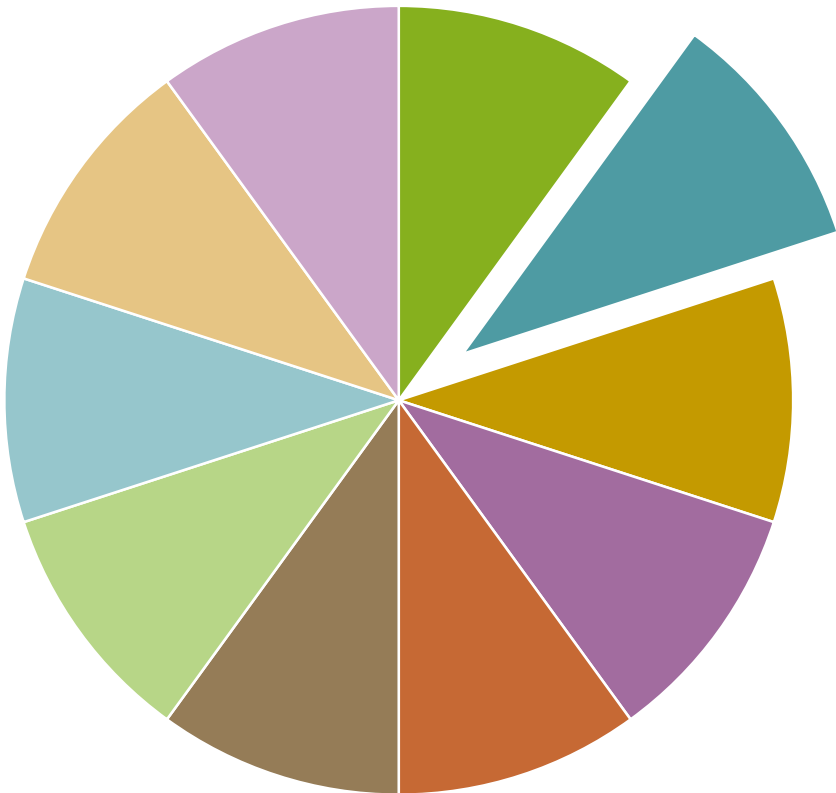
Populacja



Próba



Populacja



Próba



Wiedza o populacji



# Próba „reprezentatywna”

W pracy opisz dokładnie na czym polega w Twoim przypadku reprezentatywność

W szczególności:

- opisz sposób losowania np. losowanie proste, warstwowe, systematyczne
- podaj licznosc próby i jeśli szacowałeś tę wielkość podaj sposób

# Chcemy by odpowiednio liczna próba:

1

Odzwierciedlała wybrane cechy populacji z wskazaną dokładnością

2

Pozwoliła na zastosowanie wybranego testu statystycznego

3

Dała odpowiednią moc testu statystycznego do wykrycia zamierzonego efektu

Średnia  $\pm$  margines błędu

# Średnia $\pm$ margines błęd



# Średnia $\pm$ margines błęd



$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

**E**



# Średnia $\pm$ margines błędu



$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

populacyjne  
odchylenie standardowe

pierwiastek  
z liczności próby

wartość krytyczna  
dla 95%CI wynosi 1,96

**E**

# Średnia $\pm$ margines błęd



$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

E

# Średnia $\pm$ margines błędu



$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

# Średnia $\pm$ margines błędu



$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$



$$n = \left(z_{\alpha/2}\right)^2 \frac{\sigma^2}{E^2}$$

# Średnia $\pm$ margines błędu

Populacyjne odchylenie standardowe, opcje:

1. Wybrane na podstawie uprzednich badań
2. Wybrane na podstawie badania pilotażowego

Zgadywanie  
(max-min)/4

# Średnia $\pm$ margines błędu

## **Przykład**

Populacja:

Osoby chorujące na nadciśnienie w Polsce w latach 2005-2010, w wieku 20-40 lat.

Zainteresowani jesteśmy średnią masą ciała tych ludzi.

Ile osób wybrać by średnia masa obarczona była małym błędem?

# Średnia $\pm$ margines błędu

$$E = 3kg$$

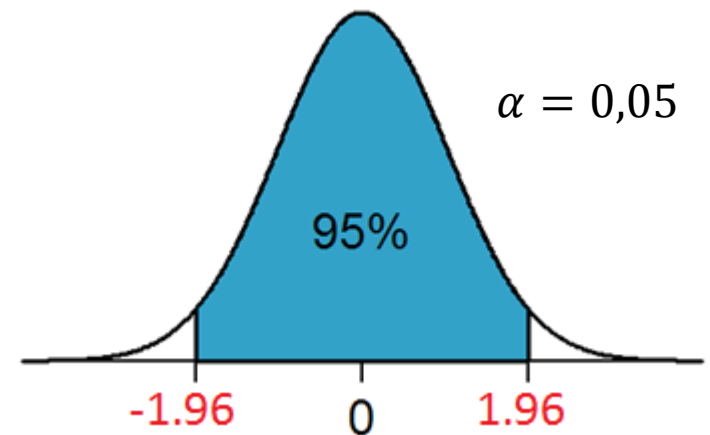
$$\sigma = 18kg$$

# Średnia $\pm$ margines błędów

$$E = 3kg$$

$$\sigma = 18kg$$

$$n = \left(z_{\alpha/2}\right)^2 \frac{\sigma^2}{E^2}$$





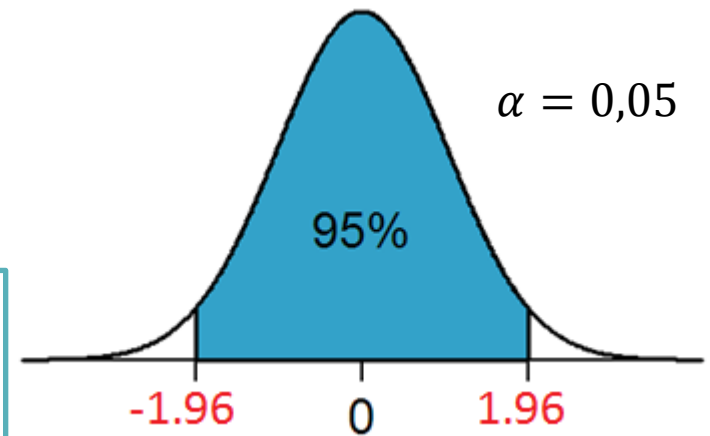
# Średnia $\pm$ margines błędów

$$E = 3kg$$

$$\sigma = 18kg$$

$$n = \left(z_{\alpha/2}\right)^2 \frac{\sigma^2}{E^2}$$

$$n = (1,96)^2 \frac{18^2}{3^2} = 138.3$$



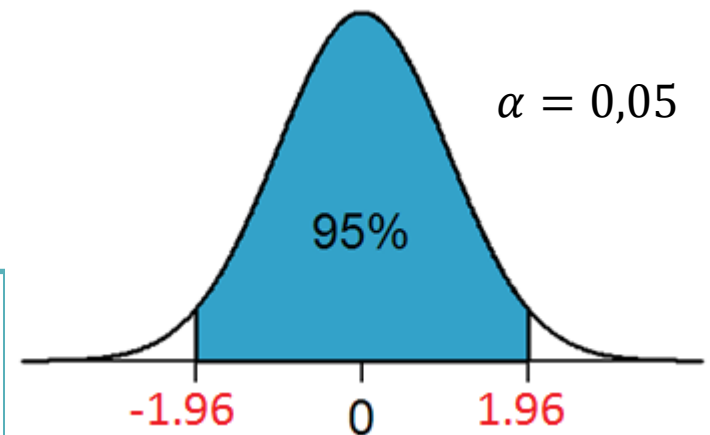
# Średnia $\pm$ margines błędu

$$E = 3kg$$

$$\sigma = 18kg$$

$$n = \left(z_{\alpha/2}\right)^2 \frac{\sigma^2}{E^2}$$

$$n = (1,96)^2 \frac{18^2}{3^2} = 138.3$$



By mieć 95% pewności, że w marginesie błędu ( $\pm 3kg$ ) zbudowanym wokół średniej z naszej próby znajduje się średnia populacyjna potrzebujemy zebrać przynajmniej 139 osób.

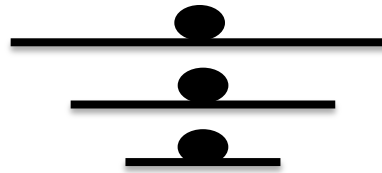
# Średnia $\pm$ margines błędu

Zmieniamy wartość  $E$

$E=3$  kg

$E=2$  kg

$E=1$  kg



$$n = \left( z_{\alpha/2} \right)^2 \frac{\sigma^2}{E^2}$$

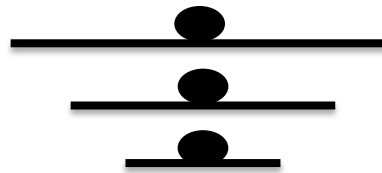
# Średnia $\pm$ margines błędu

Zmieniamy wartość  $E$

$E=3$  kg

$E=2$  kg

$E=1$  kg



$n=139$

$$n = \left( z_{\alpha/2} \right)^2 \frac{\sigma^2}{E^2}$$

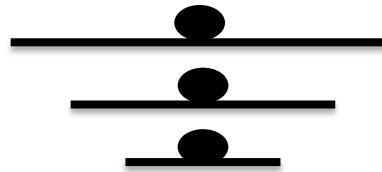
# Średnia $\pm$ margines błędów

Zmieniamy wartość  $E$

$E=3$  kg

$E=2$  kg

$E=1$  kg



$n=139$

$n=312$

$$n = \left( z_{\alpha/2} \right)^2 \frac{\sigma^2}{E^2}$$

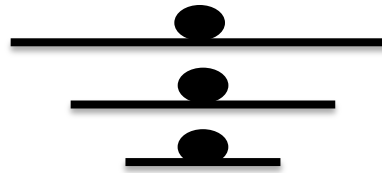
# Średnia $\pm$ margines błędu

Zmieniamy wartość  $E$

$E=3$  kg

$E=2$  kg

$E=1$  kg



$n=139$

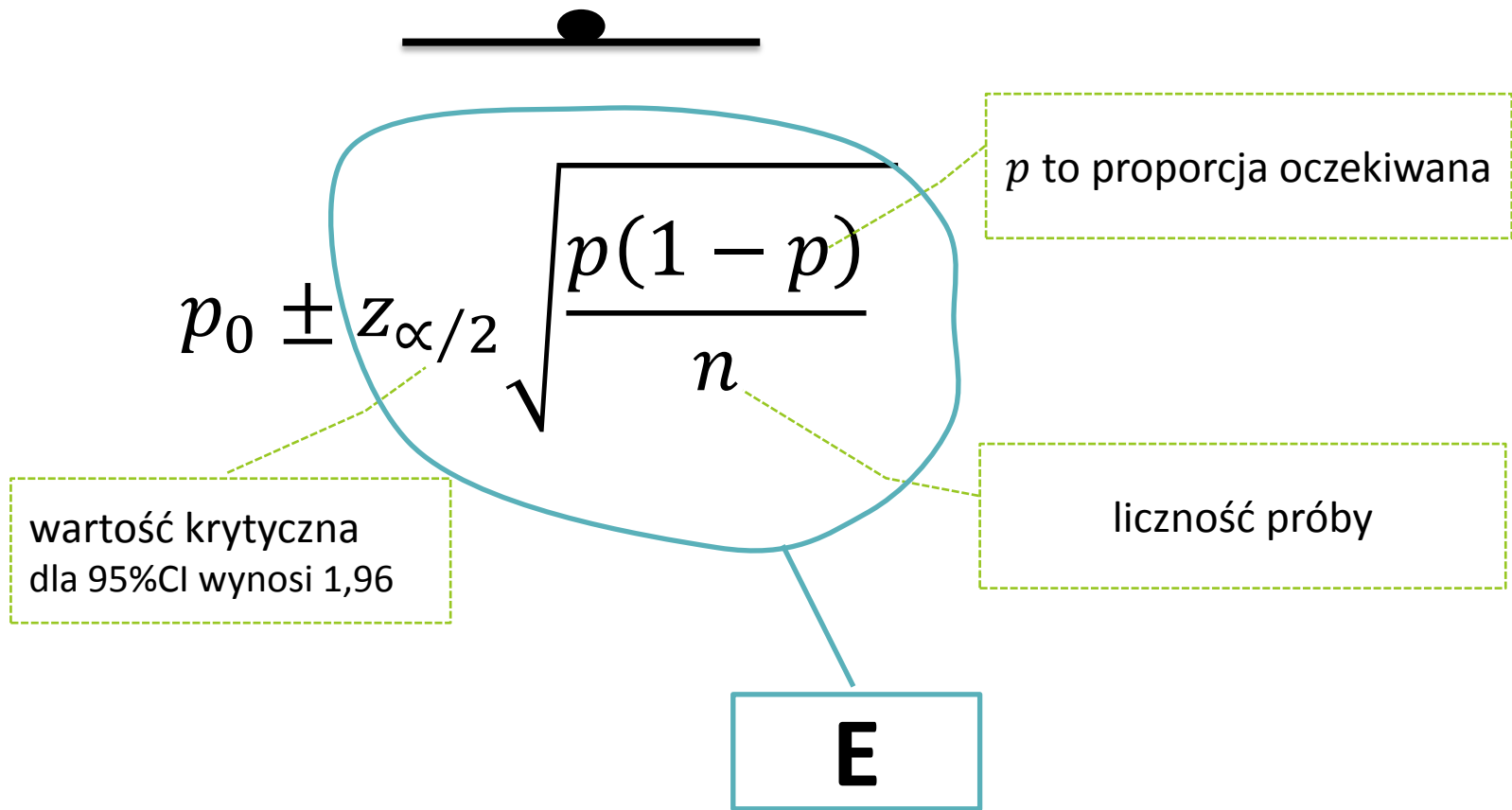
$n=312$

$n=1245$

$$n = \left( z_{\alpha/2} \right)^2 \frac{\sigma^2}{E^2}$$

Proporcja  $\pm$  margines błędu

# Proporcja $\pm$ margines błędu





# Proporcja $\pm$ margines błędu

$p_0 \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$

**E**

# Proporcja $\pm$ margines błędu



$$E = z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

# Proporcja $\pm$ margines błędu



$$E = z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$



Dla nieznanego  $p$

$$n = z_{\alpha/2}^2 \frac{1}{4E^2}$$

# Proporcja $\pm$ margines błędu

## **Przykład**

Populacja:

Uprawnieni do głosowania na prezydenta RP

Zainteresowani jesteście poparciem dla  
poszczególnych kandydatów

Ile osób wybrać by uzyskany odsetek obarczony był  
małym błędem?

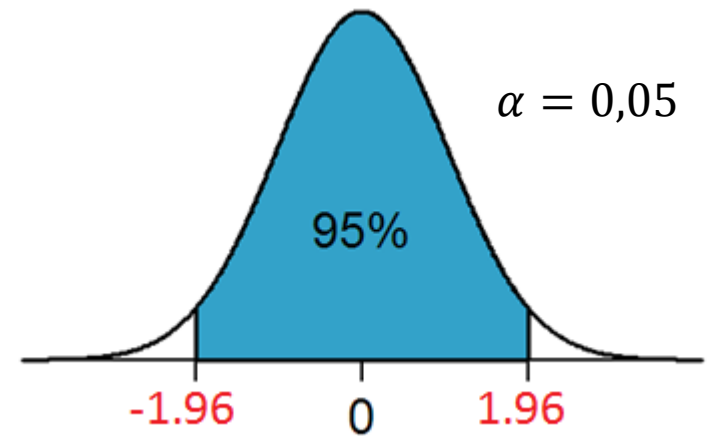
# Proporcja $\pm$ margines błędu

$$E = 2\% = 0,02$$

# Proporcja $\pm$ margines błędu

$$E = 2\% = 0,02$$

$$n = z_{\alpha/2} \frac{1}{4E^2}$$

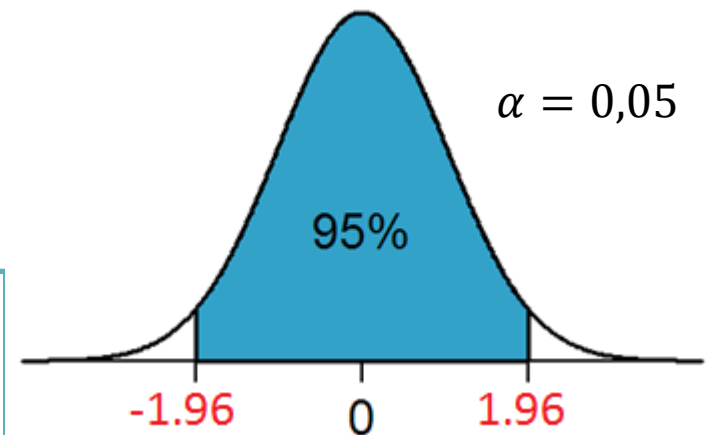


# Proporcja $\pm$ margines błędu

$$E = 2\% = 0,02$$

$$n = z_{\alpha/2} \frac{1}{4E^2}$$

$$n = (1,96)^2 \frac{1}{4 \times 0,02^2} = 2401$$

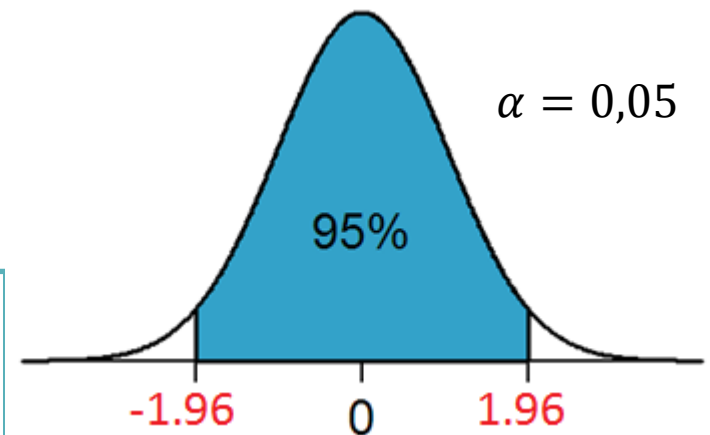


# Proporcja $\pm$ margines błędu

$$E = 2\% = 0,02$$

$$n = z_{\alpha/2} \frac{1}{4E^2}$$

$$n = (1,96)^2 \frac{1}{4 \times 0,02^2} = 2401$$



Przy próbie przynajmniej 2401 elementowej będziemy mieć 95% pewności, że błąd poparcia dla wybranego kandydata na prezydenta RP nie przekracza 2%



# Dziękuję

Barbara Więckowska  
Katedra i Zakład Informatyki i Statystyki, UM Poznań  
oraz PQStat Software ([www.pqstat.pl](http://www.pqstat.pl))  
[barbara.wieckowska@ump.edu.pl](mailto:barbara.wieckowska@ump.edu.pl)