

Statystyka matematyczna - wykład czternasty<sup>1</sup>  
Wnioskowanie przy innych schematach losowania.  
kierunek: matematyka I<sup>o</sup>  
specjalność: matematyka finansowa

dr Jarosław Kotowicz

Instytut Informatyki Uniwersytet w Białymstoku

# Spis treści

- 1 Próba losowa. Metod reprezentacyjna
- 2 Wnioskowanie przy innych schematach losowania
- 3 Zagadnienia na egzamin

# Losowania próby. I

Sekcja ta została opracowana na podstawie [2], [1], [3].

Metoda reprezentacyjna omawia sposoby pobierania próby oraz odpowiadające im metody wnioskowania co do rozkładu cechy jednowymiarowej.

Przypomnijmy:

## Definicja 1

*Jednostki losowania to indywidualne elementy populacji lub zespół elementów.*

## Definicja 2

*Losowy wybór próby to sposób wyboru, przy którym każda jednostka losowania ma stałe, znane prawdopodobieństwo wyboru do próby, lecz niekoniecznie jednakowe dla wszystkich elementów.*

## Losowania próby. II

### Definicja 3

*Losowanie nazywamy proces doboru próby losowej, a regułę tego wyboru nazywamy schematem losowania.*

### Definicja 4

*Operatem losowania nazywamy ponumerowaną listę wszystkich jednostek losowania*

### Definicja 5

*Losowanie jednostek do próby (ang. sampling) to technika pobierania jednostek do próby realizująca założenia losowego wyboru próby.*

# Techniki losowania próby. I

Techniki losowania jednostek to m.in.:

- 1 losowanie proste indywidualne
  - 1 niezależne,
  - 2 zależne,
- 2 losowanie zespołowe,
- 3 losowanie systematyczne,
- 4 losowanie warstwowe (wielostopniowe).

## Techniki losowania próby. II

### Definicja 6

*Losowanie proste indywidualne (ang. simple random sampling) polega na wyborze z populacji generalnej ostatecznych, indywidualnych jednostek badania w sposób bezpośredni, gdzie wszystkie jednostki losowania są jednocześnie jednostkami badania.*

### Definicja 7

*Losowanie proste indywidualne niezależne, inaczej losowanie ze zwracaniem (ang. simple random sampling with replacement), to losowanie proste indywidualne, w którym nadajemy każdej jednostce statystycznej jednakowe prawdopodobieństwo dostania się do próby. Jednostka raz wylosowana do badania bierze udział w dalszym losowaniu.*

## Techniki losowania próby. III

### Definicja 8

*Losowanie proste indywidualne zależne, inaczej losowanie bez zwracania (ang. simple random sampling without replacement), to losowanie proste indywidualne, w którym wynik kolejnego losowania zależy od wyniku losowania poprzedniego (przez uszczuplenie populacji o wylosowaną wcześniej jednostkę).*

## Techniki losowania próby. IV

### Definicja 9

*Losowanie zespołowe (ang. cluster sampling), to schemat losowania stosowany, gdy badana próba jest duża, a otrzymanie operatu losowania uwzględniającego wszystkie ostateczne jednostki badania jest kosztowne. Przed przystąpieniem do losowania należy precyzyjnie zdefiniować tzw. zespoły jednostek. Następnie losuje się zespoły, jakie mają zostać włączone do badania. Jeśli do ostatecznej próby zostaną włączone wszystkie jednostki z wylosowanych zespołów, losowanie nazywa się losowaniem jednostopniowym. Najczęściej jednak z zespołów, będącymi jednostkami losowania pierwszego stopnia, losuje się mniejsze zespoły, z nich jeszcze mniejsze itd., by na końcu tego łańcucha wylosowane zostały indywidualne jednostki badania. W zależności więc od ilości etapów, schemat ten nazywamy dwustopniowym, trójstopniowym, itd., czyli losowaniem wielostopniowym.*



# Techniki losowania próby. V

## Definicja 10

*Losowanie warstwowe (ang. stratified sampling) to dwustopniowa procedura, zgodnie z którą populację dzieli się na subpopulacje (podpopulacje), zwane warstwami, a następnie z poszczególnych subpopulacji niezależnie pobiera się w sposób losowy elementy do próby. W poszczególnych warstwach losowanie może być proste lub zależne.*

# Spis treści

- 1 Próba losowa. Metod reprezentacyjna
- 2 Wnioskowanie przy innych schematach losowania**
- 3 Zagadnienia na egzamin

# Metoda reprezentacyjna. I

## Omówimy

- 1 Losowanie ze skończonej populacji (ewentualnie niedużej).
- 2 Losowanie warstwowe (populacja składa się z części, których udział w populacji jest znany).
- 3 Losowanie zespołowe (jednostki populacji tworzą pewne naturalne grup, które wygodnie jest losować w celu dobrania próby).
- 4 Losowanie systematyczne (losowanie z dostępnego uszeregowania jednostek losowania do próby włącza się jednakowo odległe od siebie jednostki).

# Losowanie ze skończonej populacji. I

Losujemy próbę  $X_1, \dots, X_n$  zależną z populacji o liczebności  $N$ .  
Nieobciążony estymator średniej to

$$\bar{X}_s = \frac{1}{n} \sum_{i=1}^n X_i,$$

gdzie

$$m = \frac{1}{N} \sum_{i=1}^N x_i$$
$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - m)^2.$$

## Losowanie ze skończonej populacji. II

Mamy

$$\mathbb{D}^2(\bar{X}_s) = \frac{S^2}{n} \frac{N-n}{N}.$$

# Losowanie warstwowe. I

## Typy losowania warstwowego

- 1 proporcjonalne – liczba wylosowanych elementów z każdej warstwy jest proporcjonalna do liczebności elementów w warstwie.
- 2 nieproporcjonalne – ustalenie liczby losowanych elementów w każdej warstwie tak, aby zapewnić żądaną precyzję oszacowań.

Oznaczenia:

# Losowanie warstwowe. II

- $N$  – liczebność populacji
- $L$  – ilość warstw w populacji
- $N_h$  – liczebność  $h$ -tej warstwy w populacji
- $W_h$  – frakcja elementów  $h$ -tej warstwy w populacji
- $n$  – liczebność próby
- $n_h$  – liczebność  $h$ -tej warstwy w próbie
- $w_h$  – frakcja elementów  $h$ -tej warstwy w próbie
- $X_{ih}$  – wartość cechy  $X$   $i$ -tej wylosowanej jednostki w  $h$ -tej warstwie
- $m_h$  – średnia cechy  $X$  w  $h$ -tej warstwie w populacji
- $S_h^2$  – wariancja cechy  $X$  w  $h$ -tej warstwie w populacji.

## Losowanie warstwowe. III

$$\bar{X}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} X_{ih}$$

$$S_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (X_{ih} - \bar{X}_h)^2.$$

$$\mathbb{D}^2(\bar{X}_h) = \frac{S_h^2}{n_h} \frac{N_h - n_h}{N_h}.$$

Nieobciążony estymator średniej w populacji

$$\bar{X}_w = \sum_{h=1}^L W_h \bar{X}_h.$$



## Losowanie warstwowe. IV

Mamy dla niego

$$\mathbb{D}^2(\bar{X}_w) = \sum_{h=1}^L W_h^2 \frac{S_h^2}{n_h} \frac{N_h - n_h}{N_h}.$$

Nieobciążony estymator sumy wartości cechy w populacji  $\tau$  jest

$$\hat{\tau}_w = N\bar{X}_w = \sum_{h=1}^L N_h \bar{X}_h.$$

Mamy dla niego

$$\mathbb{D}^2(\hat{\tau}_w) = \sum_{h=1}^L N_h^2 \frac{S_h^2}{n_h} \frac{N_h - n_h}{N_h}.$$

# Losowanie zespołowe. I

Typy losowania zespołowego:

- ① jednostopniowe – losowanie pewnej liczby grup i tworzenie próby ze wszystkich jednostek wchodzących w skład wylosowanych grup,
- ② dwustopniowe lub wielostopniowe – losowanie pewnej liczby grup i wybieranie pojedynczych elementów lub mniejszych zespołów jednostek drogą dalszego losowania.

Oznaczenia:

$N$  – liczba zespołów populacji

$M_i$  – liczba elementów indywidualnych  $i$ -tego zespołu

$M$  – liczebność populacji

$\bar{M}$  – średnia liczba elementów w zespole

$m$  – średnia wartość cechy  $X$  w populacji

$m_i$  – średnia wartość cechy  $X$  w  $i$ -tym zespole populacji

$M_i$  – liczba elementów indywidualnych  $i$ -tego zespołu w populacji.

## Losowanie zespołowe. II

Założenie: losujemy bez zwracania  $n$  zespołów. Mamy wtedy

$X_{ik}$  – wartość cechy  $X$  w próbie dla  $k$ -tego elementu  
w  $i$ -tym wylosowanym zespole

$M_{(i)}$  – średnia liczba elementów  $i$ -tego zespołu w próbie.

Estymatorem średniej  $m$  w populacji jest

$$\bar{X}_z = \frac{\sum_{i=1}^n \sum_{k=1}^{M_i} X_{ik}}{\sum_{i=1}^n M_i} = \frac{\sum_{i=1}^n m_i M_i}{\sum_{i=1}^n M_i}.$$

Wariancja  $\bar{X}_z$  wynosi

$$\mathbb{D}^2(\bar{X}_z) = \frac{S_p^2}{n} \frac{N-n}{N},$$

## Losowanie zespołowe. III

gdzie

$$S_p^2 = \frac{1}{N-1} \sum_{j=1}^N (m_j - m)^2 \left( \frac{M_j}{\bar{M}} \right)^2 = \frac{1}{(N-1)\bar{M}^2} \sum_{j=1}^N (m_j M_j - m M_j)^2.$$

Estymatorem wariancji  $S_p^2$  jest

$$S_p^2 = \frac{1}{(n-1)\bar{M}_{(i)}^2} \sum_{i=1}^n (m_i M_i - \bar{X}_z M_j)^2$$

i mamy wtedy

$$\mathbb{D}^2(\bar{X}_z) = \frac{N-n}{Nn(n-1)\bar{M}_{(i)}^2} \sum_{i=1}^n (m_i M_i - \bar{X}_z M_j)^2.$$

## Losowanie zespołowe. IV

## Uwaga 1

Przy założeniu niejednakowej liczebności elementów w zespołach średnia  $\bar{X}_z$  jest estymatorem obciążonym.




Estymatorem sumy wartości cechy w populacji  $\tau$  jest

$$\hat{\tau} = M\bar{X}_z.$$

Mamy dla niego

$$\mathbb{D}^2(\hat{\tau}) = N^2\bar{M}^2\mathbb{D}^2(\bar{X}_z) = \frac{N(N-n)}{n(n-1)} \sum_{i=1}^n (m_i M_i - \bar{X}_z M_i)^2.$$

# Bibliografia

-  Czesław Bracha. *Teoretyczne podstawy metody reprezentacyjnej*. Warszawa: Wydaw. Naukowe PWN, 1996. ISBN: 8301121238.
-  *Założenia analizy regresji liniowej*. URL:  
<https://stat.gov.pl/metainformacje/sloownik-pojec/pojecia-stosowane-w-stat>  
(term. wiz. 27. 05. 2020).
-  Ryszard Zasępa. *Badania statystyczne metodą reprezentacyjną : zarys teorii i praktyki*. Warszawa: Państw. Wydawnictwo Naukowe, 1962.

# Spis treści

- 1 Próba losowa. Metod reprezentacyjna
- 2 Wnioskowanie przy innych schematach losowania
- 3 Zagadnienia na egzamin

# Zagadnienia na egzamin. I

- 1 Podstawowe pojęcia statystyczne.
- 2 Elementy statystyki opisowej.
  - 1 Podstawowe pojęcia statystyki opisowej.
  - 2 Miary przeciętne.
  - 3 Miary zmienności.
  - 4 Miary asymetrii.
  - 5 Miary koncentracji.
- 3 Próba losowa i statystyka z próby.
- 4 Rozkłady statystyk.
  - 1 Rozkład średniej i różnicy średnich.
  - 2 Rozkład wariancji i ilorazu wariancji.
  - 3 Rozkłady graniczne niektórych statystyk.
- 5 Teoria estymacji.
  - 1 Podstawy teorii estymacji – podstawowe pojęcia i ich rodzaje



# Zagadnienia na egzamin. II

- 2 Rodzaje estymatorów i ich własności.
- 3 Nierówność Rao - Cramera.
- 6 Estymacja punktowa.
  - 1 Metoda momentów konstrukcji estymatorów.
  - 2 Metoda największej wiarygodności konstrukcji estymatorów.
  - 3 Metoda najmniejszych kwadratów konstrukcji estymatorów.
- 7 Estymacja przedziałowa.
  - 1 Podstawowe pojęcia związane z estymacją przedziałową.
  - 2 Przedział ufności dla średniej  $m$  w populacji normalnej i nieznanym rozkładzie
  - 3 Przedział ufności dla wariancji  $\sigma^2$  dla populacji normalnej o nieznanym wartości oczekiwanej i odchylenia standardowego.
  - 4 Przedział ufności dla parametru frakcji.
  - 5 Estymacja przedziałowa – problem minimalizacji próby.
- 8 Testowanie hipotez.

# Zagadnienia na egzamin. III

- 1 Podstawowe pojęcia.
- 2 Testy statystyczne (etapy konstrukcji, błędy, itd.).
- 9 Parametryczne testy istotności.
  - 1 Test istotności dla wartości średniej populacji generalnej.
  - 2 Test istotności dla wariancji.
- 10 Nieparametryczne testy istotności.
  - 1 Klasyfikacja testów.
  - 2 Przykładowe test zgodności.
  - 3 Test niezależności chi-kwadrat.
- 11 Korelacja cech.
  - 1 Miary zależności nieliniowej oparte na statystyce chi-kwadrat.
  - 2 Współczynnik korelacji liniowej Pearsona.
  - 3 Współczynnik korelacji rang Spearmana.
  - 4 Stosunki korelacyjne.

# Zagadnienia na egzamin. IV

## 12 Klasyczny model regresji liniowej.

- 1 Sformułowanie klasycznego modelu regresji liniowej i klasycznego modelu normalnej regresji liniowej.
- 2 Estymacja parametrów funkcji regresji.
- 3 Dokładność dopasowania prostej metodą najmniejszych kwadratów.
- 4 Twierdzenie Gaussa-Markowa.
- 5 Wnioskowanie o klasycznym modelu normalnej regresji liniowej.
- 6 Macierzowe ujęcie modelu regresji liniowej.
- 7 Klasyczny model regresji liniowej z wieloma zmiennymi niezależnymi.
- 8 Inne niż liniowe modele regresji.
- 9 Założenia modelu regresji liniowej i ich testowanie.

## 13 Jednoczynnikowa analiza wariancji.

- 1 Podstawowe pojęcia.
- 2 Założenia analizy wariancji.
- 3 Testy post-hoc.

# Zagadnienia na egzamin. V

- 14 Szeregi czasowe.
  - 1 Pojęcie szeregu czasowego.
  - 2 Wyrównywanie szeregów czasowych (metody).
  - 3 Dopasowywanie krzywych **MNK**.
  - 4 Wskaźnik wahań okresowych dla szeregu rozdzielczego.
  - 5 Eliminacja wahań okresowych i prognozowanie zjawiska dla przyszłych okresów.
- 15 Analiza dynamiki szeregów czasowych.
  - 1 Przyrosty i indeksy indywidualne.
  - 2 Syntetyczne wielkości charakteryzujące szeregi czasowe.
  - 3 Indywidualne i agregatowe indeksy cen, ilości i wielkości.
  - 4 Wielkości stosunkowe indywidualne i stosunkowe.