Statystyka matematyczna - wykład ósmy[1]
Wnioskowanie przy innych schematach losowania. Podsumowanie wykładu.
kierunek: informatyka i ekonometria I°

dr Jarosław Kotowicz

Instytut Informatyki, Uniwersytet w Białymstoku

Spis treści

Metoda reprezentacyjna. I

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

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.

Definicja 3

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

Metoda reprezentacyjna. II

Definicja 4

Operatem losowania nazywamy ponumerowaną listę wszystkich jednostek losowania

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 $X_1, \ldots, X_n$ zależną z populacji o liczebności N. Nieobciążany estymator średniej to

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

gdzie

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

Losowanie ze skończonej populacji. II

Mamy

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

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

$$\begin{aligned}
\overline{X}_h &= \frac{1}{n_h}\sum_{i=1}^{n_h} X_{ih}\\
S_h^2 &= \frac{1}{n_h-1}\sum_{i=1}^{n_h}(X_{ih}-\overline{X}_h)^2.\\
\mathbb{D}^2(\overline{X}_h) &= \frac{S_h^2}{n_h}\frac{N_h-n_h}{N_h}.
\end{aligned}$$

Nieobciążony estymator średniej w populacji

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

Losowanie warstwowe. IV

Mamy dla niego

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

Nieobciążony estymator sumy wartości cechy w populacji τ jest

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

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
$\overline{M}$	–	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

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

Wariancja $\overline{X}_z$ wynosi

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

Losowanie zespołowe. III

gdzie

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

Estymatorem wariancji S_p^2 jest

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

i mamy wtedy

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

Losowanie zespołowe. IV

Uwaga 1

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

Estymatorem sumy wartości cechy w populacji τ jest

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

Mamy dla niego

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

Spis treści

Zagadnienia na egzamin. I

1. Podstawowe pojęcia statystyczne.
2. Próba losowa i statystyka z próby.
3. Rozkłady statystyk.
    1. Rozkład średniej i różnicy średnich.
    2. Rozkład wariancji i ilorazu wariancji.
    3. Rozkłady graniczne niektórych statystyk.
4. Teoria estymacji.
    1. Podstawy teorii estymacji – podstawowe pojęcia i ich rodzaje
    2. Rodzaje estymatorów i ich własności.
    3. Nierówność Rao - Cramera.
5. Estymacja punktowa.
    1. Metoda momentów konstrukcji estymatorów.
    2. Metoda największej wiarogodności konstrukcji estymatorów.
    3. Metoda najmniejszych kwadratów konstrukcji estymatorów.

Zagadnienia na egzamin. II

6. 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 wariacji σ^2 dla populacji normalnej o nieznanych wartości oczekiwanej i odchylenia standardowego.
   4. Przedział ufności dla parametru frakcji.
   5. Estymacja przedziałowa – problem minimalizacji próby.
7. Testowanie hipotez.
   1. Podstawowe pojęcia.
   2. Testy statystyczne (etapy konstrukcji, błędy, itd.).
8. Parametryczne testy istotności.
   1. Test istotności dla wartości średniej populacji generalnej.
   2. Test istotności dla wariancji.
9. Nieparametryczne testy istotności.

Zagadnienia na egzamin. III

1. Klasyfikacja testów.
2. Przykładowe test zgodności.
3. Test niezależności chi-kwadrat.

10. Korelacja cech.
    1. Miary zależności nieliniowej oparte na statystyce chi-kwadrat.
    2. Współczynnik korelacji liniowej Pearsona.
    3. Współczynnik korelacji rang Spearmana.
    4. Stosunki korelacyjne.
11. 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.

Zagadnienia na egzamin. IV

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.

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