

Zadania na ocenę dostateczną.

Pytania.

1. Co to jest plan losowania?
2. Dla populacji $U = \{1, 2, 3\}$ wyznaczyć wszystkie próby nieuporządkowane.
3. Przypuśćmy, że mamy plan losowania o ustalonej liczebności próby $n = 6$. Ile wynosi suma prawdopodobieństw pierwszego rzędu?
4. Co to jest prawdopodobieństwo pierwszego rzędu? Podać definicję.
5. Co to jest błąd średniokwadratowy estymatora?

Zadania.

1. Z populacji 6000 budynków wylosowano (wg. losowania prostego bez zwracania) próbę rozmiaru 300. Obserwowaną cechą była kubatura budynku. Uzyskano wyniki: $\bar{y} = 1200$ oraz $s = 500$ [w m³]. Na poziomie ufności 0.95 oszacować średnią kubaturę budynku. Jaki powinien być minimalny rozmiar próby, aby błąd szacunku nie przekroczył wartości 60.?

2. Z populacji o liczebności $N = 1000$ wylosowano dziesięcioelementową próbę zgodnie ze schematem losowania prostego bez zwracania. Obserwowano wartości cechy Y oraz cechy dodatkowej X . Uzyskano wyniki:

x	48	44	51	53	56	59	39	45
y	51	34	44	45	46	49	32	33

Wiedząc, że $\bar{X} = 50$ wyznaczyć ocenę średniej populacyjnej oraz średniej błędu kwadratowego (dla zastosowanych estymatorów). Zastosować estymator ilorazowy oraz regresyjny. Czy są to estymatory nieobciążone średniej populacyjnej?

3. Populację o liczności 50000 podzielono na cztery warstwy względem kategorii wiekowej. Z każdej warstwy wylosowano próbę bez zwracania. Rozmiar próby był proporcjonalny do rozmiaru populacji (przy zaokrągleniu do najbliższej liczby całkowitej). Całowy rozmiar próby wynosił 200. Obserwowano wartość pewnej cechy. Na podstawie uzyskanych wyników oszacować na poziomie ufności 0.95 wartość średnią badanej cechy.

Kategoria wiekowa	N_h	\bar{y}_h	s_h
20-30	15000	3	1
30-40	12000	5	0.5
40-50	14000	8	0.9
50-60	9000	7	0.8

Dodatkowe zadania na ocenę dobrą.

4. W celu oszacowania wartości globalnej pewnej cechy, populacja licząca 50000 jednostek została podzielona na 100 jednostek losowania pierwszego stopnia. Według schematu losowania prostego bez zwracania wylosowano kolejno 5 jednostek losowania pierwszego stopnia. Z każdej takiej jednostki pobrano, według schematu losowania prostego bez zwracania, pewną liczbę jednostek drugiego stopnia. Otrzymano wyniki: $s_1^2 = 5799764$ oraz $s_2^2 = 1302$. Wyznaczyć optymalną lokalizację próby automatycznie wyważonej, przy założeniu, że koszt przypadający na jednostkę pierwszego stopnia wynosi 10 zł, a na jednostkę drugiego stopnia 1 zł. Przyjąć, że koszt całkowity badania nie może przekroczyć 300 zł.

5. W celu oszacowania wartości średniej pewnej cechy, populację liczącą 50 tysięcy jednostek podzielono na warstwy po 5, 10, 20 oraz 15 tysięcy elementów. Z każdej warstwy wybrano pewną liczbę jednostek. Na ich podstawie oszacowano wariancję dla każdej warstwy: dla pierwszej wyniosła ona 40, drugiej 50, trzeciej 48 oraz dla czwartej 35. Wyznaczyć optymalną alokację próby, przy założeniu, że dysponujemy kwotą 60000 zł, a koszt zebrania informacji od jednostki badania należącej do pierwszej, drugiej, trzeciej i czwartej warstwy wynosi odpowiednio 22, 10, 11 oraz 12 zł.

Dodatkowe zadania na ocenę bardzo dobrą.

6. Niech $\mathcal{U} = \{u_1, u_2, u_3\}$ oraz niech $\mathbf{Y} = [Y_1, Y_2, Y_3]$. Szacujemy $\bar{Y} = \frac{1}{3} \sum_{i=1}^3 Y_i$. Zgodnie z planem lpzz losujemy próbę dwuelementową. Estymator t określony jest w następujący sposób:

$$\begin{aligned}t(Y_1, Y_1) &= Y_1, \\t(Y_1, Y_2) &= \alpha_1 Y_1 + \beta_1 Y_2, \\t(Y_1, Y_3) &= \alpha_2 Y_1 + \gamma_1 Y_3, \\t(Y_2, Y_2) &= Y_2, \\t(Y_2, Y_3) &= \beta_2 Y_2 + \gamma_2 Y_3, \\t(Y_3, Y_3) &= Y_3,\end{aligned}$$

Dla jakich wartości parametrów α, β, γ estymator t jest nieobciążony. Wyznaczyć $D^2 t$.

7. Niech N elementowa populacja będzie podzielona na L warstw o liczebnościach N_h ($h = 1, \dots, L$). Zadanie polega na oszacowaniu wartości średniej \bar{Y} cechy Y w całej populacji. Zastosowano schemat losowania lwp:lpbz uzyskując z każdej warstwy próbę o liczebności odpowiednio n_h ($h = 1, \dots, L$). Niech $n = n_1 + \dots + n_h$. Rozważmy dwa estymatory średniej:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \text{ o wariancji } D^2(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{S^2}{n}$$

$$\bar{y}_w = \frac{1}{N} \sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} \text{ o wariancji } D^2(\bar{y}_w) = \sum_{h=1}^L \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$$

Przy jakich warunkach dotyczących wariancji obserwowanej cechy estymator \bar{y}_w jest lepszy od etymatora \bar{y} ?

8. Niech N elementowa populacja będzie podzielona na L warstw o liczebnościach N_h ($h = 1, \dots, L$). Zadanie polega na oszacowaniu wartości średniej \bar{Y} cechy Y w całej populacji. Zastosowano schemat losowania lwp:lpbz uzyskując z każdej warstwy próbę o liczebności odpowiednio n_h ($h = 1, \dots, L$). Niech $n = n_1 + \dots + n_h$. Rozważmy następujący estymator średniej:

$$\bar{y}_w = \frac{1}{N} \sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} \text{ o wariancji } D^2(\bar{y}_w) = \sum_{h=1}^L \left(\frac{N_h}{N}\right)^2 \left(1 - \frac{n_h}{N_h}\right) \frac{S_h^2}{n_h}$$

Zakładając, że koszt pobrania obserwacji w h -tej warstwie wynosi k_h oraz, że całkowity koszt badań nie może przekroczyć K wyznaczyć optymalną alokację próby.