Testowanie hipotez i estymacja w sytuacji populacji skończonej rozmiaru

Kinga Kurowska
promotor dr inż Andrzej Giniewicz

Wydział Matematyki
Politechnika Wrocławska

Wrocław, 01.12.2016r.

Spis treści

Motywacja

Motywacja

- znamy wzory w przypadku nieskończonej populacji - rozkład Bernoulliego,

Motywacja

- znamy wzory w przypadku nieskończonej populacji - rozkład Bernoulliego,
- gdy badana populacja jest faktycznie duża to one się sprawdzają,

Motywacja

- znamy wzory w przypadku nieskończonej populacji - rozkład Bernoulliego,
- gdy badana populacja jest faktycznie duża to one się sprawdzają,
- jednak gdy badanie populacji wpływa na kolejne wyniki potrzebne jest założenie o skończonej populacji,

Motywacja

- znamy wzory w przypadku nieskończonej populacji - rozkład Bernoulliego,
- gdy badana populacja jest faktycznie duża to one się sprawdzają,
- jednak gdy badanie populacji wpływa na kolejne wyniki potrzebne jest założenie o skończonej populacji,
- zastowanie w medycynie.

Testowanie hipotez - teoria

Testowanie hipotez - teoria

- H_0 - hipoteza zerowa (jej prawdziwość testujemy),

Testowanie hipotez - teoria

- H_0 - hipoteza zerowa (jej prawdziwość testujemy),
- H_1 - hipoteza alternatywna,

Testowanie hipotez - teoria

- H_0 - hipoteza zerowa (jej prawdziwość testujemy),
- H_1 - hipoteza alternatywna,
- α - poziom istotności testu, czyli prawdopodobieństwo odrzucenia hipotezy zerowej, gdy jest prawdziwa,

Testowanie hipotez - teoria

- H_0 - hipoteza zerowa (jej prawdziwość testujemy),
- H_1 - hipoteza alternatywna,
- α - poziom istotności testu, czyli prawdopodobieństwo odrzucenia hipotezy zerowej, gdy jest prawdziwa,
- błąd I rodzaju - odrzucenie H_0, gdy jest prawdziwa,

Testowanie hipotez - teoria

- H_0 - hipoteza zerowa (jej prawdziwość testujemy),
- H_1 - hipoteza alternatywna,
- α - poziom istotności testu, czyli prawdopodobieństwo odrzucenia hipotezy zerowej, gdy jest prawdziwa,
- błąd I rodzaju - odrzucenie H_0, gdy jest prawdziwa,
- p wartość - najmniejszy poziom istotności testu, przy którym odrzucamy H_0,

Testowanie hipotez - teoria

- H_0 - hipoteza zerowa (jej prawdziwość testujemy),
- H_1 - hipoteza alternatywna,
- α - poziom istotności testu, czyli prawdopodobieństwo odrzucenia hipotezy zerowej, gdy jest prawdziwa,
- błąd I rodzaju - odrzucenie H_0, gdy jest prawdziwa,
- p wartość - najmniejszy poziom istotności testu, przy którym odrzucamy H_0,
- błąd II rodzaju - nie odrzucenie H_0, gdy jest fałszywa, dla zadanej alternatywnej wartości badanego parametru,

Testowanie hipotez - teoria

- H_0 - hipoteza zerowa (jej prawdziwość testujemy),
- H_1 - hipoteza alternatywna,
- α - poziom istotności testu, czyli prawdopodobieństwo odrzucenia hipotezy zerowej, gdy jest prawdziwa,
- błąd I rodzaju - odrzucenie H_0, gdy jest prawdziwa,
- p wartość - najmniejszy poziom istotności testu, przy którym odrzucamy H_0,
- błąd II rodzaju - nie odrzucenie H_0, gdy jest fałszywa, dla zadanej alternatywnej wartości badanego parametru,
- moc testu - prawdopodobieństwo odrzucenia H_0, gdy jest fałszywa, dla zadanej alternatywnej wartości badanego parametru;
  p-stwo błędu II rodzaju $=$ 1 - moc testu.

Populacja nieskończonego rozmiaru - losowanie ze zwracaniem

Kiedy populację nazywamy nieskończoną?

Populacja nieskończonego rozmiaru - losowanie ze zwracaniem

Kiedy populację nazywamy nieskończoną?

- Gdy jest na tyle duża, aby uznać, że przebadanie jednego osobnika z populacji nie wpływa znacząco na kolejne wyniki - "zwracamy" go do populacji i losujemy ponownie.

Populacja nieskończonego rozmiaru - losowanie ze zwracaniem

Kiedy populację nazywamy nieskończoną?

- Gdy jest na tyle duża, aby uznać, że przebadanie jednego osobnika z populacji nie wpływa znacząco na kolejne wyniki - "zwracamy" go do populacji i losujemy ponownie.
- Przykładem może być: populacja kobiet na świecie, albo populacja rzutu kostką.

Model - schemat Bernoulliego

Gdy losujemy ze zwracaniem najprostszym modelem jest schemat Bernoulliego. Zmienna losowa $X \sim B(n, p)$, gdzie n to liczba prób, p prawdopodobieńtwo sukcesu oraz

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \; k = 0, 1, 2, \ldots$$

W języku statystyki n to wielkość próbki, a p to prawdopodobieństwo, że zachodzi dana cecha.

Populacja skończonego rozmiaru - losowanie bez zwracania

Kiedy populację nazywamy skończoną?

Populacja skończonego rozmiaru - losowanie bez zwracania

Kiedy populację nazywamy skończoną?

- Gdy jest na tyle mała, aby uznać, że przebadanie jednego osobnika z populacji wpływa znacząco na kolejne wyniki - losujemy bez zwracania go do populacji.

Populacja skończonego rozmiaru - losowanie bez zwracania

Kiedy populację nazywamy skończoną?

- Gdy jest na tyle mała, aby uznać, że przebadanie jednego osobnika z populacji wpływa znacząco na kolejne wyniki - losujemy bez zwracania go do populacji.
- Przykładem może być populacja osób chore na rzadką chorobę.

Model - rozkład hipergomentryczny

Gdy losujemy bez zwracania musimy uwzględnić to, że prawdopodobieństwo sukcesu zmienia się. Rozkładem odpowiadającym takiej sytuacji jest rozkład hipergeometryczny [1].

Zmienna losowa pochodzi z rokładu hipergeometrycznego, gdy $X \sim H(n, M, N)$, gdzie n to ilość próbki, M ilość osobników w populacji z daną cechą, a N ilość całej populacji oraz funkcja prawdopodobieństwa

$$h(k; n, M, N) = P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \ L \leqslant k \leqslant U, \tag{1}$$

gdzie $L = \max\{0, M - N + n\}$, $U = \min\{n, M\}$.

Sformuowanie statystyki

Niech X_1, X_2 niezależne zmienne losowe z rozkładu hipergeometrycznego $X_i \sim H(n_i, M_i, N_i)$, $P(X_i = k_i)$ oraz $p_i = M_i/N_i$, $i = 1, 2$.

Postawmy hipotezę H_0: $p_1 = p_2$ vs. H_1: $p_1 \neq p_2$.

Rozważmy statystykę

(2)
$$Z_{X_1,X_2} = \frac{X_1/n_1 - X_2/n_2}{\sqrt{V_{X_1,X_2}}},$$

gdzie V_{X_1,X_2} jest estymatorem wariancji zmiennej losowej $X_1/n_1 - X_2/n_2$ pod warunkiem H_0 dana wzorem

(3)
$$V_{X_1,X_2} = \left(\frac{N_1 - n_1}{n_1(N_1-1)} + \frac{N_2 - n_2}{n_2(N_2-1)}\right)\left(\frac{X_1+X_2}{n_1+n_2}\right)\left(1 - \frac{X_1+X_2}{n_1+n_2}\right).$$ [1]

Z Test

Z Centralnego Twierdzenia Granicznego statystyka Z_{X_1,X_2} pod warunkiem H_0 jest w przybliżeniu z rozkładu standardowego normalnego $N(0,1)$. Z test odrzuca H_0, gdy p wartość

$$P(|Z_{X_1,X_2}| \geqslant |Z_{k_1,k_2}| \; |H_0) = 2(1 - \Phi(|Z_{k_1,k_2}|)) \text{ [1]}$$

jest mniejsza niż poziom istotności α.

E Test

E test odrzuca H_0, gdy p wartość

$$P(|Z_{X_1,X_2}| \geqslant |Z_{k_1,k_2}| \; |H_0) =$$
$$= \sum_{x_1=L_{x_1}}^{U_{x_1}} \sum_{x_2=L_{x_2}}^{U_{x_2}} h(x_1; n_1, \hat{M}_1, N_1) h(x_2; n_2, \hat{M}_2, N_2) \mathbb{1}(|Z_{x_1,x_2}| \geqslant |Z_{k_1,k_2}|), \; [1]$$

jest mniejsza niż poziom istotności α.

Przy czym $\hat{M}_i = [N_i \hat{p}]$, $\hat{p} = (k_1 + k_2)/(n_1 + n_2)$,
$L_{x_i} = \max\{0, \hat{M}_i - N_i + n_i\}$, $U_{x_i} = \min\{n_i, \hat{M}_i\}$, $i = 1, 2$.

Porównanie testów

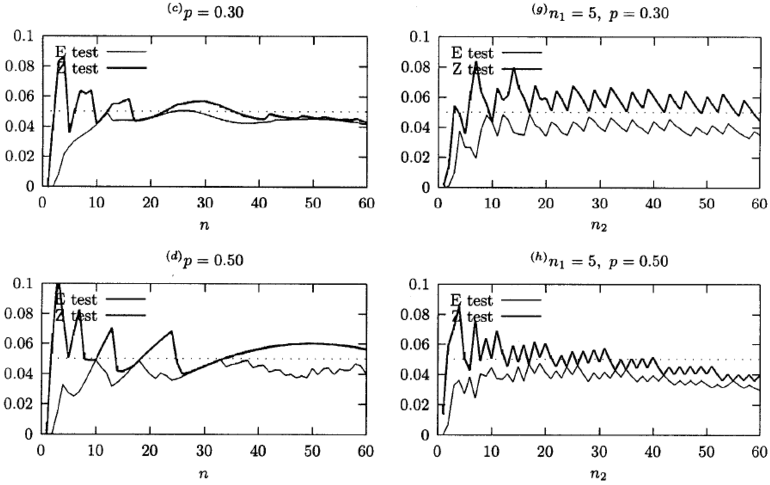

Figure 1. Exact Size of the Tests as a Function of the Sample Sizes at the Nominal Level α = 0.05; N_1 = N_2 = 100.

Porównanie testów

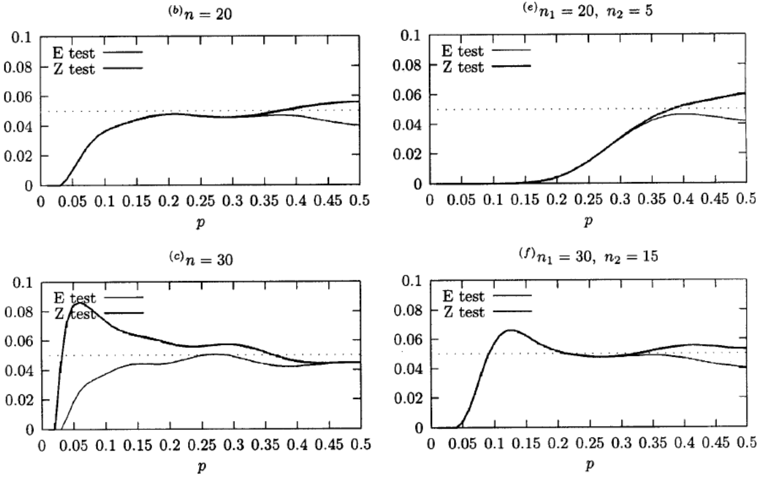

Figure 2. Exact Size of the Tests as a Function of $p = M_1/N_1 = M_2/N_2$ *at the Nominal Level* $\alpha = 0.05$; $N_1 = N_2 = 100$.

Co jeszcze?

Bibliografia

K. Krishnamoorthy, J. Thomson „Hypothesis Testing about Proportions in Two Finite Populations” The American Statistician, Vol. 56, No. 3 (2002): 215-222

J. P. Buonaccorsi, „A Note on Confidence Intervals for Proportions in Finite Populations”. The American Statistician, Vol. 41, No. 3 (1987): 215-218

F. Y. Edgeworth „On the Value of a Mean as Calculated from a Sample” Journal of the Royal Statistical Society, Vol. 81, No. 4 (Jul., 1918), pp. 624-632