



Politechnika Wrocławska

Wydział Matematyki

Kierunek: Matematyka Stosowana

Specjalność: *nie dotyczy*

Praca dyplomowa — inżynierska

**TESTOWANIE HIPOTEZ I ESTYMACJA
W SYTUACJI POPULACJI SKOŃCZONEGO
ROZMIARU**

Kinga Kurowska

Słowa kluczowe:
testowanie hipotez
przedziały ufności
zastosowanie w medycynie

Krótkie streszczenie:

W pracy został rozważony problem testowania hipotez i estymacji w sytuacji populacji skończonego rozmiaru. Przedstawiono i porównano dwa schematy pobierania obserwacji, dla nieskończonej oraz skończonej populacji. Zostały opisane trzy testy statystyczne oraz przeprowadzone badania symulacyjne w celu porównania testów. Ostatecznie został wybrany jeden test, który jest najlepszy dla małych populacji.

| | | | |
|---------------|---|--------------|---------------|
| Opiekun pracy | dr inż. Andrzej Giniewicz | | |
| dyplomowej | <i>Stopień naukowy, imię i nazwisko</i> | <i>Ocena</i> | <i>Podpis</i> |

Do celów archiwalnych pracę dyplomową zakwalifikowano do: *

a) kategorii A (akta wieczyste),

b) kategorii BE 50 (po 50 latach podlegające ekspertyzie).

* *niepotrzebne skreślić*

pieczęć wydziałowa

Wrocław, rok 2017



Politechnika Wrocławska

Faculty of Pure and Applied Mathematics

Field of study: Applied Mathematics

Specialty: *not applicable*

Engineering Thesis

**HYPOTHESIS TESTING AND ESTIMATION
IN THE CASE OF FINITE POPULATION SIZE**

Kinga Kurowska

keywords:

hypothesis testing
confidence intervals
medical applications

Short summary:

The problem of hypothesis testing and estimation in the case of finite population size was considered in the thesis. Two schemes of sampling, for infinite and finite population, were presented and compared. Three statistical tests were described. Simulations studies were carried out in order to compare tests. Finally, one test was chosen, which is the best for small population size.

| | | | |
|------------|--|--------------|------------------|
| Supervisor | dr inż. Andrzej Giniewicz | | |
| | <i>Title, degree, name and surname</i> | <i>Grade</i> | <i>Signature</i> |

*For the purposes of archival thesis qualified to: **

- a) Category A (perpetual files)*
- b) Category BE 50 (subject to expertise after 50 years)*

** Delete as appropriate*

stamp of the faculty

Wrocław, 2017

Spis treści

| | |
|---|----|
| Wstęp | 5 |
| Rozdział 1. Schemat pobierania obserwacji | 7 |
| 1.1. Nieskończona populacja | 7 |
| 1.2. Skończona populacja | 7 |
| 1.3. Porównanie rozkładów | 8 |
| Rozdział 2. Przedstawienie testów | 13 |
| 2.1. Sformułowanie problemu | 13 |
| 2.2. Testy ze skończoną poprawką | 13 |
| 2.2.1. Test Z | 14 |
| 2.2.2. Test E | 15 |
| 2.2.3. Moc testu | 15 |
| 2.3. Test bez skończonej poprawki | 16 |
| 2.3.1. Test Zb | 16 |
| Rozdział 3. Analiza testów | 19 |
| 3.1. Porównanie testów ze skończoną poprawką | 19 |
| 3.1.1. Obszary krytyczne | 19 |
| 3.1.2. Błąd I rodzaju i moc testu | 20 |
| 3.2. Porównanie testu bez skończonej poprawki Zb z testem E | 22 |
| Podsumowanie | 29 |
| Spis rysunków | 31 |
| Bibliografia | 33 |

Wstęp

Początki teorii rachunku prawdopodobieństwa i statystyki sięgają XVI w. Zajmowano się wtedy analizą rzutu kostką oraz prawdopodobieństwem błędów pomiarowych. Już w XVII wieku Blaise Pascal sformułował i dowiódł własności trójkąta arytmetycznego oraz użył pojęcia kombinacji [3]. Na początku XVIII wieku opublikowane zostały prace Jacoba Bernoullego, w których zawarł wiele swoich tez na temat prawdopodobieństwa. Przez te kilka wieków teoria rachunku prawdopodobieństwa i statystyki znacząco się wzbogaciła i rozwinęła. Rozpoczęto rozważania na temat estymacji i testowania hipotez, które są w naszych czasach zasadniczą domeną statystyki.

W przypadku dyskretnym najczęściej testowane są proporcje populacji [5]. Chcemy się przekonać czy dana próbka ma jakąś konkretną proporcję albo dwie próbki mają tę samą proporcję elementów z badaną cechą. Znamy jest powszechnie teoria dotycząca testowania hipotez, gdy populacja jest nieskończona, a raczej na tyle duża, że możemy ją w przybliżeniu uznać za nieskończoną. Wtedy schemat próbkowania jest opisany jako losowanie ze zwracaniem. Jednak przypadek nieskończonej populacji nie wyczerpuje tematu testowania proporcji. Gdy populacja jest bardzo mała albo, gdy próbka jest niewiele mniejsza od całej populacji, schemat próbkowania opiera się o losowanie bez zwracania.

Warto zająć się teorią testowania hipotez dla skończonej populacji, ponieważ w określonych przypadkach testy ze skończoną poprawką dają dużo dokładniejszą informację o badanym przypadku niż testy zakładające nieskończoną populację. Ponadto zastosowanie tego typu testów ma duże znaczenie w medycynie, gdzie często rozważane populacje mają na tyle wyspecjalizowane cechy, że są uważane za małe.

W pierwszym rozdziale znajduje się opis schematu pobierania danych, w przypadku nieskończonej i skończonej populacji, oraz porównanie obu sposobów w oparciu o własności rozkładów próbek i estymację przedziałową proporcji. W rozdziale 2 przedstawiono testy bez skończonej poprawki oraz z jej uwzględnieniem. Rozdział 3 zawiera porównanie testów na podstawie prawdopodobieństwa błędu I rodzaju oraz mocy testu. Na końcu pracy znajduje się podsumowanie uzyskanych wyników.

Rozdział 1

Schemat pobierania obserwacji

Gdy rozważamy skończoną populację, należy dokładnie przyjrzeć się schematowi pobierania obserwacji. Jakimi rozkładami są opisane próbki oraz jakie są własności tych rozkładów. Wyznamy także przedziały ufności dla proporcji p w obu rozkładach. Przy czym, proporcja to stosunek wartości obserwacji do wielkości próbki, gdzie wartość obserwacji to liczba osobników z badaną cechą w próbce.

1.1. Nieskończona populacja

Gdy populacja jest bardzo duża, możemy traktować ją jako nieskończoną. Wobec tego losowanie kolejnych elementów próbki jest niezależne, czyli jest to losowanie ze zwracaniem. Zakładamy, że pobierane obserwacje pochodzą z rozkładu Bernoulliego $\mathcal{B}(n, p)$, gdzie n to rozmiar próby z nieskończonej populacji, a p to proporcja zdarzeń sprzyjających w populacji. Funkcja prawdopodobieństwa zmiennej losowej X z rozkładu dwumianowego jest równa

$$b(k; n, p) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad 0 \leq k \leq n. \quad (1.1)$$

1.2. Skończona populacja

Kiedy rozważamy populację skończoną, kolejne elementy próbki są losowane bez zwracania. Oznacza to, że prawdopodobieństwo sukcesu zmienia się w trakcie pobierania elementów obserwacji w zależności od wyboru poprzednich. W takiej sytuacji próbka pochodzi z rozkładu hipergeometrycznego $\mathcal{H}(n, M, N)$. Przy czym n jest rozmiarem próbki, M ilością osobników w populacji z daną cechą, a N rozmiarem populacji. Zmienna losowa X z rozkładu hipergeometrycznego ma funkcję prawdopodobieństwa określoną wzorem

$$h(k; n, M, N) = P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad L \leq k \leq U, \quad (1.2)$$

gdzie

$$L = \max\{0, M - N + n\}, \quad U = \min\{n, M\}. \quad (1.3)$$

Zauważmy, że wzór (1.2) ma prostą interpretację. Klasyczna definicja prawdopodobieństwa określa szansę zajścia zdarzenia A jako iloraz liczby

zdarzeń elementarnych w A przez liczbę zdarzeń elementarnych w Ω , czyli $P(A) = |A|/|\Omega|$. W rozważanym przypadku zdarzeniu A odpowiada sytuacja, w której próbka będzie zawierać k osobników z daną cechą. Zatem ilość zdarzeń elementarnych w A to iloczyn dwóch kombinacji bez powtórzeń. Mamy: $\binom{M}{k}$ możliwych wyborów k osobników z M posiadających daną cechę oraz $\binom{N-M}{n-k}$ możliwych wyborów $n-k$ osobników z $N-M$ nieposiadających tej cechy. Wobec tego

$$|A| = \binom{M}{k} \binom{N-M}{n-k}. \quad (1.4)$$

Natomiast zbiór wszystkich zdarzeń elementarnych w Ω to wybór losowej próbki n osobników z N -elementowej populacji, wobec tego

$$|\Omega| = \binom{N}{n}. \quad (1.5)$$

Ograniczenia nałożone na k są również naturalne. Dolne ograniczenie L jest równe maksimum z 0 i $M-N+n$. Będzie ono niezerowe, gdy $M-N+n > 0$. Przekształcając nierówność, otrzymujemy $n > N-M$, czyli rozmiar próbki jest większy od liczby osobników bez badanej cechy w populacji. W rezultacie mamy pewność, że w próbce będzie przynajmniej tyle osobników z daną cechą, ile wynosi różnica $n - (N-M)$. Ograniczenie górne jest równe minimum z n i M , co wynika z faktu, że nie może być więcej osób w próbce z daną cechą niż w całej populacji. Analiza wzoru (1.2) pokazuje, że rozkład hipergeometryczny jest ściśle związany z rozmiarem populacji.

1.3. Porównanie rozkładów

Rozkład hipergeometryczny daje bardzo podobne wyniki do rozkładu Bernoulliego, gdy populacja jest duża, albo próbka stosunkowo mała. Natomiast przy małej populacji i dużej próbce różnica między tymi rozkładami jest znaczna.

Rozważmy to na medycznym przykładzie. Załóżmy, że jest grupa 20 osób, które są chore na jakąś bardzo rzadką chorobę oraz że 25% z nich ma szansę na wyzdrowienie. Chcemy dowiedzieć się, ile osób spośród przebadanych może wyzdrowieć. Weźmy 3 różne próbki o wielkościach n równych odpowiednio 10, 17 oraz 20. Możemy tę sytuację opisać za pomocą rozkładu Bernoulliego, wtedy badana zmienna losowa jest z rozkładu $\mathcal{B}(n, 0.25)$. Drugim sposobem jest rozkład hipergeometryczny $\mathcal{H}(n, 5, 20)$. Rysunki 1.1 – 1.3 na stronie 10 przedstawiają funkcję prawdopodobieństwa dla wymienionych przypadków.

Przeanalizujmy jakie wartości mogą przyjmować zmienne losowe z obu rozkładów. W rozkładzie dwumianowym zmienna w każdym z przypadków przyjmuje wartości od 0 do n . Na przykład na wykresie 1.1a argumenty funkcji to zbiór $\{0, 1, \dots, 9, 10\}$. Mimo że w populacji maksymalnie może wyzdrowieć 5 osób, dla rozkładu Bernoulliego prawdopodobieństwo $X = 6$ wynosi 0.09. Przedstawiona rozbieżność wynika z założenia, że pobieranie próbki to losowanie ze zwracaniem, czyli w próbce mogą znaleźć się dwie te same osoby. Natomiast w rozkładzie hipergeometrycznym zmienną losową

ograniczają L i U , zdefiniowane wzorem (1.3), które uwzględniają wielkość próbki. Porównując, na wykresie 1.1b zmienna losowa nie przyjmuje wartości większej niż 5. Dodatkowo na wykresie 1.2b najmniejszą wartością jest 2, ponieważ populacja zawiera 15 osób, które nie wyzdrowieją, zatem w 17-osobowej próbce jest pewne, że przynajmniej dwie będą zdrowe. Kolejne losowania są od siebie zależne, więc nie możemy drugi raz wylosować tej samej osoby.

Różnica między funkcjami prawdopodobieństwa obu rozkładów rośnie wraz ze wzrostem próbki. Skrajny przypadek przedstawia rysunek 1.3, na którym próbka równa się populacji. Wykres 1.3b idealnie obrazuje przypadek $n = N$. Pewne jest to, że w 20 osobach będzie dokładnie 5, które mogą wyzdrowieć. Natomiast na wykresie 1.3a prawdopodobieństwo, że $X = 5$ wynosi jedynie 0.2, ponieważ funkcja prawdopodobieństwa dla rozkładu dwumianowego nie uwzględnia wielkości populacji. Warto również zaznaczyć, że im większa próbka, tym $P(X = 5)$ dla rozkładu Bernoulliego jest coraz mniejsze, ponieważ mamy więcej elementów w obserwacji. Tymczasem dla rozkładu hipergeometrycznego jest odwrotnie, funkcja prawdopodobieństwa dla $k = 5$ rośnie, aż w końcu osiąga wartość 1. Gdy przebadamy więcej osobników, wzrasta nasza wiedza o próbce oraz jest bardziej prawdopodobne to, że znajdzie się w niej aż pięcioro zdrowych pacjentów.

Spójrzmy, jak wyglądają średnia i wariancja dla rozważanych rozkładów. Gdy $X \sim \mathcal{B}(n, p)$, to

$$E(X) = np, \quad Var(X) = np(1 - p). \quad (1.6)$$

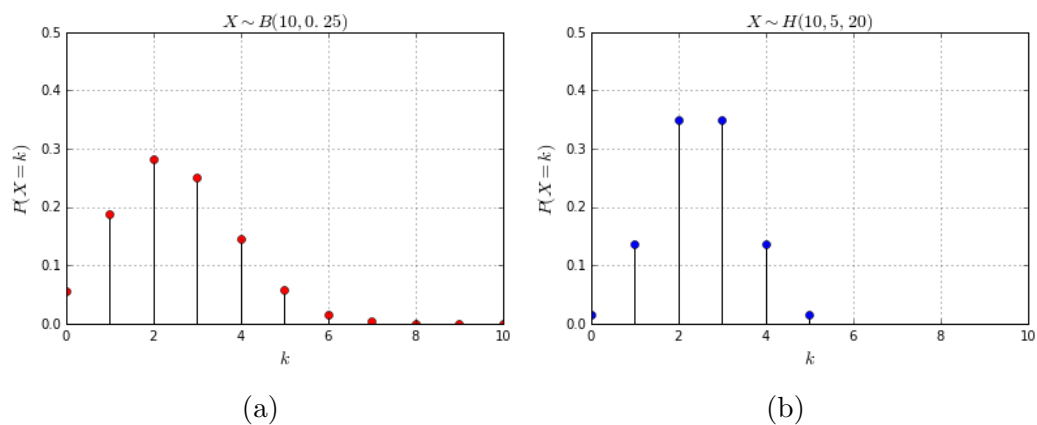
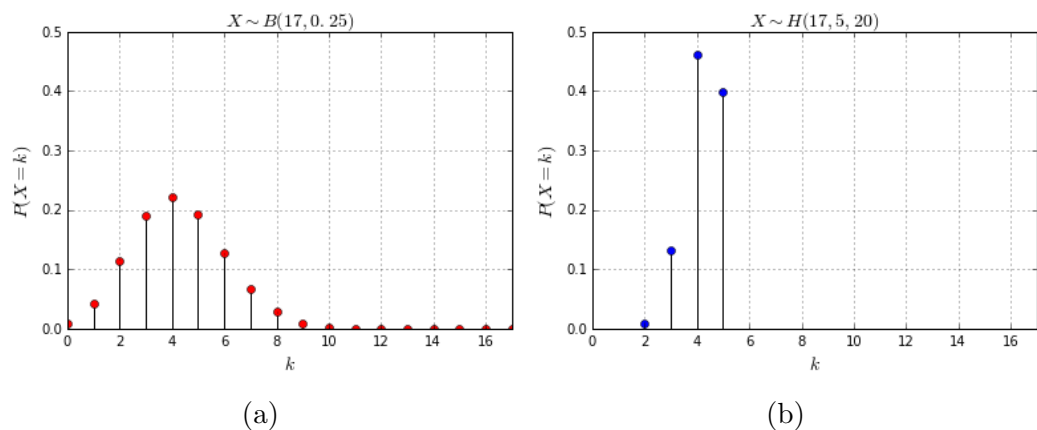
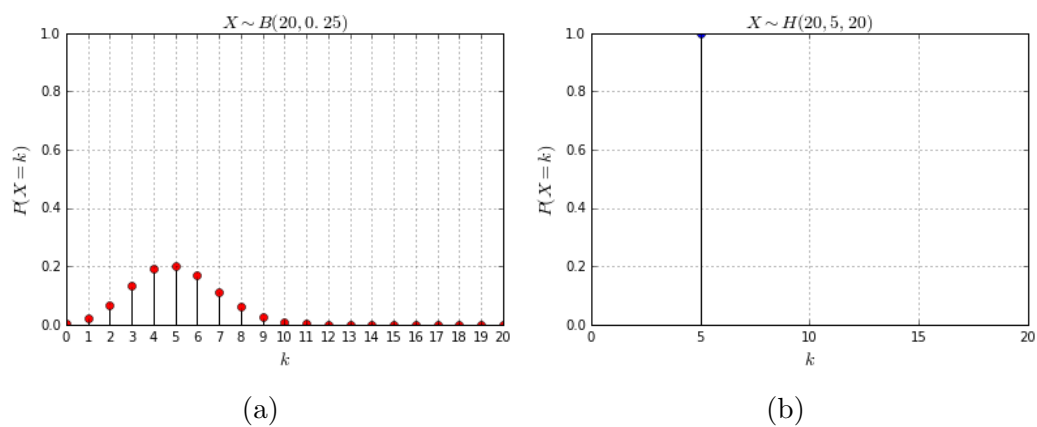
Podczas gdy $X \sim \mathcal{H}(n, M, N)$, to

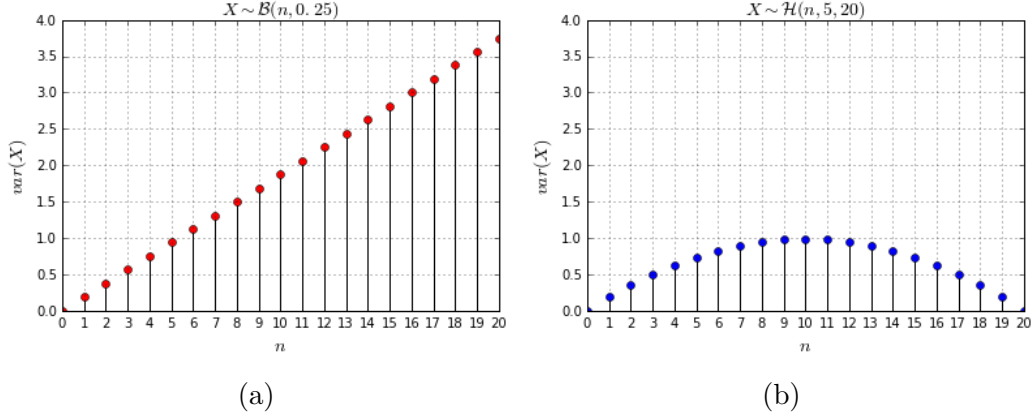
$$E(X) = n \frac{M}{N}, \quad Var(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N - n}{N - 1}. \quad (1.7)$$

Parametr p rozkładu dwumianowego odpowiada ilorazowi parametrów M/N w rozkładzie hipergeometrycznym. Wobec tego średnie obu rozkładów są sobie równe, ale wariancje różni dodatkowy składnik w rozkładzie hipergeometrycznym $(N - n)/(N - 1)$. Przeanalizujemy, jak ten czynnik wpływa na zróżnicowanie rozkładów. Rysunek 1.4 przedstawia wykresy wariancji dla rozważanego przykładu w zależności od wielkości obserwacji. W przypadku rozkładu dwumianowego wariancja stale rośnie wraz ze wzrostem próbki. Ostatecznie, gdy $n = 20$ jest ona największa. Jednakże, gdy przebadamy całą populację, oczekiwanym rezultatem jest wartość zerowa wariancji, ponieważ nie ma wtedy losowości. Taki wynik daje nam wykres 1.4b. Funkcja na początku rośnie, ale gdy wielkość próbki przekroczy połowę rozmiaru populacji, wariancja zaczyna maleć aż do zera. Odzwierciedla to fakt, że gdy coraz więcej wiemy o populacji, losowość uzyskanych wyników maleje.

Rozważmy także, jak wyglądają przedziały ufności dla zmiennych losowych z obu rozkładów. Załóżmy, że unormowana zmienna losowa pochodzi z rozkładu standardowego normalnego, czyli

$$\frac{X - E(X)}{\sqrt{Var(X)}} \sim \mathcal{N}(0, 1). \quad (1.8)$$

Rysunek 1.1: Funkcje prawdopodobieństwa $b(k; 10, 0.25)$ oraz $h(k; 10, 5, 20)$ Rysunek 1.2: Funkcje prawdopodobieństwa $b(k; 17, 0.25)$ oraz $h(k; 17, 5, 20)$ Rysunek 1.3: Funkcje prawdopodobieństwa $b(k; 20, 0.25)$ oraz $h(k; 20, 5, 20)$



Rysunek 1.4: Wariancja rozkładów Bernoulliego i hipergeometrycznego w zależności od rozmiaru próbki

Wobec tego dla $X \sim \mathcal{B}(n, p)$ przedział ufności parametru p jest równy

$$\left[\hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right], \quad (1.9)$$

gdzie $z_{1-\alpha/2}$ to kwantyl rozkładu normalnego rzędu $1 - \alpha/2$. Natomiast dla $X \sim \mathcal{H}(n, M, N)$ przedział ufności proporcji $p = M/N$ to

$$\left[\hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})(N-n)}{n(N-1)}}, \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})(N-n)}{n(N-1)}} \right]. \quad (1.10)$$

W obu przypadkach estymator $\hat{p} = k/n$, gdzie k to wartość obserwacji.

Kontynuując rozważany przykład, załóżmy teraz, że nie wiemy ile osób może wyzdrowieć w całej populacji, ale mamy wyniki badań. W grupie 10 osób wyzdrowiały 3. Chcemy się dowiedzieć ile osób może wyzdrowieć wśród wszystkich 20. Podsumowując, znamy parametry $N = 20$, $n = 10$ i wartość obserwacji $k = 3$, zatem możemy wyznaczyć przedziały ufności dla parametru p w obu rozkładach. Dla rozkładu dwumianowego

$$p \in [0.02, 0.58]. \quad (1.11)$$

Natomiast dla rozkładu hipergeometrycznego

$$p \in [0.09, 0.5]. \quad (1.12)$$

Prawdziwy parametr $p = 0.25$ faktycznie zawiera się w tych przedziałach. Zauważmy, że przedział dla rozkładu Bernoulliego jest szerszy, ze względu na brak skończonej poprawki w wariancji.

Rozdział 2

Przedstawienie testów

W niniejszym rozdziale znajduje się opis trzech testów. Dwa testy ze skończoną poprawką wykorzystują rozkład hipergeometryczny, a trzeci test, bez skończonej poprawki, opiera się o rozkład dwumianowy. Następnie omówiony jest sposób liczenia mocy dla wymienionych testów. Wymienione testy sprawdzają, czy dwie populacje mają te same proporcje badanej cechy. Zatem, aby wykonać test, potrzebne są próbki z obu populacji. Zakładamy również, że populacje są od siebie niezależne, tym samym próbki pochodzące z tych populacji także nie zależą od siebie.

2.1. Sformułowanie problemu

Założmy, że X_1 i X_2 są niezależnymi zmiennymi losowymi. Zaobserwowane wartości X_1 i X_2 oznaczmy odpowiednio k_1 i k_2 oraz proporcje w obserwacjach p_1 i p_2 . Będziemy testować

$$H_0: p_1 = p_2 \quad \text{przeciwko} \quad H_1: p_1 \neq p_2, \quad (2.1)$$

na podstawie wartości obserwacji i znanych parametrów populacji. Rozważmy unormowaną statystykę

$$Z_{X_1, X_2} = \frac{X_1/n_1 - X_2/n_2}{\sqrt{V_{X_1, X_2}}}, \quad (2.2)$$

gdzie V_{X_1, X_2} to estymator wariancji rozkładu zmiennej losowej $X_1/n_1 - X_2/n_2$, pod warunkiem prawdziwości H_0 , w połączonej próbie. Jego wzór zależy od rozkładu, z którego pochodzą zmienne losowe X_1 i X_2 . Wartość statystyki Z_{X_1, X_2} oznaczmy jako Z_{k_1, k_2} . Jest ona wyliczana według powyższych wzorów, poprzez zamienienie zmiennych losowych X_1 i X_2 odpowiednio ich wartościami k_1 i k_2 .

2.2. Testy ze skończoną poprawką

Jak już było wspomniane w podrozdziale 1.2, próbki w przypadku skończonej populacji pochodzą z rozkładu hipergeometrycznego. Zatem:

$$X_1 \sim \mathcal{H}(n_1, M_1, N_1), \quad X_2 \sim \mathcal{H}(n_2, M_2, N_2) \quad (2.3)$$

oraz proporcje są równe $p_1 = M_1/N_1$, $p_2 = M_2/N_2$. Znane parametry to rozmiary próbek n_1 i n_2 i wielkości populacji N_1 i N_2 .

W celu wyprowadzenia wariancji rozkładu $X_1/n_1 - X_2/n_2$ pod warunkiem $p_1 = p_2$ w połączonej próbie zapiszmy wariancję rozważanej zmiennej losowej w łącznej próbie, korzystając z własności wariancji oraz tego, że $Cov(X_1, X_2) = 0$ z niezależności X_1 i X_2

$$\begin{aligned} Var(X_1/n_1 - X_2/n_2) &= Var(X_1/n_1) + Var(X_2/n_2) = \\ &= Var(X_1)/n_1^2 + Var(X_2)/n_2^2. \end{aligned} \quad (2.4)$$

Wariancje X_1 i X_2 są równe:

$$Var(X_1) = n_1 p_1 (1 - p_1) (N_1 - n_1) / (N_1 - 1), \quad (2.5)$$

$$Var(X_2) = n_2 p_2 (1 - p_2) (N_2 - n_2) / (N_2 - 1). \quad (2.6)$$

Pamiętając, że zakładamy równość $p_1 = p_2$ zastąpmy oba parametry jednym równym p . Po podstawieniu otrzymujemy

$$\begin{aligned} V_{X_1, X_2} &= \frac{1}{n_1} p (1 - p) \frac{N_1 - n_1}{N_1 - 1} + \frac{1}{n_2} p (1 - p) \frac{N_2 - n_2}{N_2 - 1} = \\ &= p (1 - p) \left(\frac{N_1 - n_1}{n_1 (N_1 - 1)} + \frac{N_2 - n_2}{n_2 (N_2 - 1)} \right), \end{aligned} \quad (2.7)$$

przy czym p to proporcja liczby osobników z daną cechą do całości populacji w rozkładzie łącznym, zatem $p = (X_1 + X_2) / (n_1 + n_2)$. Ostatecznie otrzymujemy

$$V_{X_1, X_2} = \left(\frac{N_1 - n_1}{n_1 (N_1 - 1)} + \frac{N_2 - n_2}{n_2 (N_2 - 1)} \right) \left(\frac{X_1 + X_2}{n_1 + n_2} \right) \left(1 - \frac{X_1 + X_2}{n_1 + n_2} \right). \quad (2.8)$$

2.2.1. Test Z

Test Z jest oparty na aproksymacji rozkładem normalnym, zatem rozważana statystyka Z_{X_1, X_2} , pod warunkiem prawdziwości H_0 , jest w przybliżeniu z rozkładu standardowego normalnego $\mathcal{N}(0, 1)$. Wtedy p -wartość wyraża się wzorem

$$P(|Z_{X_1, X_2}| \geq |Z_{k_1, k_2}| | H_0) \approx 2(1 - \Phi(|Z_{k_1, k_2}|)), \quad (2.9)$$

gdzie Φ oznacza dystrybuantę rozkładu $N(0, 1)$. Test Z odrzuca hipotezę zerową, gdy p -wartość jest mniejsza od poziomu istotności α .

Jeżeli zakładamy, że $Z_{X_1, X_2} \sim \mathcal{N}(0, 1)$, możemy wyznaczyć przedział ufności statystyki Z_{X_1, X_2} . Wiemy, że

$$P\left(-z_{1-\alpha/2} \leq \frac{p_1 - p_2}{\sqrt{V_{X_1, X_2}}} \leq z_{1-\alpha/2}\right) = 1 - \alpha, \quad (2.10)$$

gdzie $-z_{1-\alpha/2}$ i $z_{1-\alpha/2}$ to kwantyle rozkładu normalnego. Wobec czego przedział ufności na poziomie $1 - \alpha$ dla różnicy proporcji $p_1 - p_2$ jest równy

$$\left[-z_{1-\alpha/2} \sqrt{V_{X_1, X_2}}, z_{1-\alpha/2} \sqrt{V_{X_1, X_2}} \right] \quad (2.11)$$

Przedział ufności w kontekście testowania to obszar akceptacji hipotezy zerowej. Jego dopełnieniem jest obszar krytyczny, czyli zbiór wartości różnicy proporcji $p_1 - p_2$, dla których test odrzuca H_0 .

2.2.2. Test E

Test E opiera się o rzeczywistą p -wartość, która, według artykułu K. Krishnamoorthy i J. Thomson z 2002 roku, jest równa [4]

$$\begin{aligned} P(|Z_{X_1, X_2}| \geq |Z_{k_1, k_2}| | H_0) &= E_{X_1, X_2}(\mathbb{1}(|Z_{X_1, X_2}| \geq |Z_{k_1, k_2}|) | H_0) = \\ &= \sum_{x_1=L_1}^{U_1} \sum_{x_2=L_2}^{U_2} h(x_1; n_1, N_1 p, N_1) h(x_2; n_2, N_2 p, N_2) \mathbb{1}(|Z_{X_1, X_2}| \geq |Z_{k_1, k_2}|), \end{aligned} \quad (2.12)$$

gdzie E_{X_1, X_2} to wartość oczekiwana łącznego rozkładu (X_1, X_2) , a p jest nieznaną wspólną proporcją pod warunkiem H_0 . Nie jest możliwe policzenie p -wartości wprost ze wzoru (2.12), ponieważ nie znamy parametru proporcji p . Krishnamoorthy i Thomson (2002) zaproponowali estymator p -wartości [4]

$$\begin{aligned} P(|Z_{X_1, X_2}| \geq |Z_{k_1, k_2}| | H_0) &\approx \\ &\approx \sum_{x_1=L_{x_1}}^{U_{x_1}} \sum_{x_2=L_{x_2}}^{U_{x_2}} h(x_1; n_1, \hat{M}_1, N_1) h(x_2; n_2, \hat{M}_2, N_2) \mathbb{1}(|Z_{x_1, x_2}| \geq |Z_{k_1, k_2}|), \end{aligned} \quad (2.13)$$

przy czym $\hat{p} = (k_1 + k_2)/(n_1 + n_2)$, $\hat{M}_i = [N_i \hat{p}]$, $L_{x_i} = \max\{0, \hat{M}_i - N_i + n_i\}$ oraz $U_{x_i} = \min\{n_i, \hat{M}_i\}$, $i = 1, 2$. Test odrzuca H_0 wtedy, gdy p -wartość wyliczona według wzoru (2.13) jest mniejsza od poziomu istotności α .

W przypadku testu E nie da się łatwo wyznaczyć przedziału ufności dla różnicy proporcji w oparciu o obszar krytyczny testu E. Jednakże w rozdziale 3 znajdują się numeryczne oszacowania przedziału ufności dla różnicy proporcji.

2.2.3. Moc testu

Moc testu to prawdopodobieństwo odrzucenia hipotezy zerowej, gdy jest ona nieprawdziwa. Wobec tego, spośród testów na zadanym poziomie istotności, interesują nas te o najwyższej mocy.

Moc obu testów można wyliczyć, korzystając z funkcji prawdopodobieństwa rozkładu hipergeometrycznego. Dla testu Z pod warunkiem hipotezy alternatywnej H_1 , zgodnie z rozważaniami Krishnamoorthy i Thomson (2002), moc jest równa [4]

$$\sum_{k_1=L_1}^{U_1} \sum_{k_2=L_2}^{U_2} h(k_1; n_1, M_1, N_1) h(k_2; n_2, M_2, N_2) \mathbb{1}(|Z_{k_1, k_2}| > z_{1-\alpha/2}), \quad (2.14)$$

gdzie $L_i = \max\{0, M_i - N_i + n_i\}$ i $U_i = \min\{n_i, M_i\}$, a $z_{1-\alpha/2}$ oznacza kwantyl rozkładu normalnego standardowego rzędu $1 - \alpha/2$.

Natomiast dla testu E, według Krishnamoorthy i Thomson (2002), moc zdefiniowana jest następująco [4]

$$\begin{aligned} & \sum_{k_1=L_1}^{U_1} \sum_{k_2=L_2}^{U_2} h(k_1; n_1, M_1, N_1) h(k_2; n_2, M_2, N_2) \times \\ & \times \mathbb{1} \left(\sum_{x_1=L_{x_1}}^{U_{x_1}} \sum_{x_2=L_{x_2}}^{U_{x_2}} h(x_1; n_1, \hat{M}_1, N_1) h(x_2; n_2, \hat{M}_2, N_2) \mathbb{1} (|Z_{x_1, x_2}| \geq |Z_{k_1, k_2}|) \leq \alpha \right), \end{aligned} \quad (2.15)$$

gdzie parametry są takie same jak we wzorach (2.13) i (2.14).

2.3. Test bez skończonej poprawki

Dla testu bez poprawki na skończony rozmiar populacji, zamiast rozkładu hipergeometrycznego stosujemy dwumianowy, więc X_1 i X_2 są niezależnymi zmiennymi losowymi o rozkładzie Bernoulliego:

$$X_1 \sim \mathcal{B}(n_1, p_1), \quad X_2 \sim \mathcal{B}(n_2, p_2). \quad (2.16)$$

Znane parametry to rozmiary próbek n_1 i n_2 .

Wariancję rozkładu $X_1/n_1 - X_2/n_2$ w łącznej próbie, pod warunkiem $p_1 = p_2$, możemy wyprowadzić analogicznie jak w podrozdziale 2.2, wychodząc od wariancji rozważanej zmiennej losowej

$$\text{Var}(X_1/n_1 - X_2/n_2) = \text{Var}(X_1)/n_1^2 + \text{Var}(X_2)/n_2^2. \quad (2.17)$$

Wariancje X_1 i X_2 są równe:

$$\text{Var}(X_1) = n_1 p_1 (1 - p_1), \quad (2.18)$$

$$\text{Var}(X_2) = n_2 p_2 (1 - p_2). \quad (2.19)$$

Zastępując p_1 i p_2 jednym parametrem równym p , otrzymujemy

$$V_{X_1, X_2} = p(1 - p)/n_1 + p(1 - p)/n_2 = p(1 - p)(1/n_1 + 1/n_2), \quad (2.20)$$

przy czym $p = (X_1 + X_2)/(n_1 + n_2)$.

2.3.1. Test Zb

Test Zb jest, podobnie jak omówiony wcześniej test E, oparty o estymator p -wartości, który, zgodnie z artykułem Storer i Kim z 1990 roku, jest równy [6]

$$\begin{aligned} & P(|Z_{X_1, X_2}| \geq |Z_{k_1, k_2}| \mid H_0) \approx \\ & \approx \sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} b(x_1; n_1, \hat{p}_1) b(x_2; n_2, \hat{p}_2) \mathbb{1} (|Z_{X_1, X_2}| \geq |Z_{k_1, k_2}|), \end{aligned} \quad (2.21)$$

gdzie $\hat{p} = (k_1 + k_2)/(n_1 + n_2)$. Test odrzuca hipotezę zerową, gdy p -wartość jest mniejsza od poziomu istotności α .

Przedziału ufności też nie da się prosto wyznaczyć tak jak dla testu E . Nie wiem czy jest sens o tym pisać. Na razie zostawiam tak, zobaczymy co z tymi wynikami w rozdziale 3

Moc testu wyraża się wzorem

$$\sum_{k_1=0}^n \sum_{k_2=0}^n b(k_1; n_1, p_1) b(k_2; n_2, p_2) \times \\ \times \mathbb{1} \left(\sum_{x_1=0}^{n_1} \sum_{x_2=0}^{n_2} b(x_1; n_1, \hat{p}_1) b(x_2; n_2, \hat{p}_2) \mathbb{1} (|Z_{x_1, x_2}| \geq |Z_{k_1, k_2}|) \leq \alpha \right). \quad (2.22)$$

Rozdział 3

Analiza testów

W celu porównania testów opisanych w rozdziale 2 napisano programy, które wyliczają prawdopodobieństwo błędu I rodzaju i moc testu oraz generują wykresy dla różnych parametrów.

3.1. Porównanie testów ze skończoną poprawką

3.1.1. Obszary krytyczne

Najpierw porównano testy ze skończoną poprawką. Zaczniemy od wyznaczenia obszaru krytycznego dla testu Z i E na poziomie istotności $\alpha = 0.05$. Dla testu Z w podrozdziale 2.2.1 został wyznaczony wzór (2.11) na przedział ufności różnicy proporcji $p_1 - p_2$. Wiadomo, że obszar krytyczny to dopełnienie przedziału ufności. Wobec tego wyliczając kwantyl $z_{1-0.05/2}$ otrzymujemy, że zbiór wartości różnicy $p_1 - p_2$, dla których test Z odrzuca H_0 jest równy

$$(-\infty, -1.96\sqrt{V_{X_1, X_2}}) \cup (1.96\sqrt{V_{X_1, X_2}}, \infty). \quad (3.1)$$

Nie wiem jak konkretnie wyznaczyć te wartości, bo V_{X_1, X_2} jest różne w zależności od k_1 i k_2 .

Dla testu E nie jesteśmy w stanie w prosty sposób wyznaczyć obszaru krytycznego. Jednakże wiemy, że jest to zbiór wartości, dla których test E odrzuca hipotezę zerową oraz będzie on postaci

$$(-\infty, a) \cup (b, \infty). \quad (3.2)$$

Wobec czego możemy wyznaczyć wartości brzegowe a i b dla konkretnych parametrów rozkładów X_1 i X_2 , sprawdzając kiedy test E odrzuci H_0 dla każdych możliwych wartości k_1 i k_2 . Następnie, a można wyliczyć jako największą ujemną wartość różnicy $p_1 - p_2$, która została odrzucona, zaś b jako najmniejszą dodatnią wartość $p_1 - p_2$. W tabeli 3.1 zostały przedstawione wyniki dla różnych parametrów, gdy $p_1 = p_2$. Nie przedstawiono wyników w przypadku $p_1 \neq p_2$, ponieważ zmiana jednej z proporcji nie wpływa na zmianę wartości a i b .

Tu kilka wniosków... + w kolejnym podrozdziale odniesienie się do tych wniosków

Tabela 3.1: Wartości brzegowe obszaru krytycznego testu E dla różnicy proporcji $p_1 - p_2$, z różnymi parametrami rozkładów obserwacji

| n_1 | n_2 | M_1 | M_2 | N_1 | N_2 | $p_1 = p_2$ | a | b |
|-------|-------|-------|-------|-------|-------|-------------|------|------|
| 5 | 5 | 50 | 50 | 100 | 100 | 0.5 | -0.6 | 0.6 |
| 10 | 10 | 30 | 30 | 100 | 100 | 0.3 | -0.3 | 0.3 |
| 5 | 20 | 10 | 10 | 100 | 100 | 0.1 | -0.5 | 0.35 |
| 15 | 15 | 18 | 30 | 30 | 50 | 0.6 | -0.2 | 0.2 |
| 40 | 40 | 1800 | 500 | 3000 | 1000 | 0.5 | -0.1 | 0.1 |

3.1.2. Błąd I rodzaju i moc testu

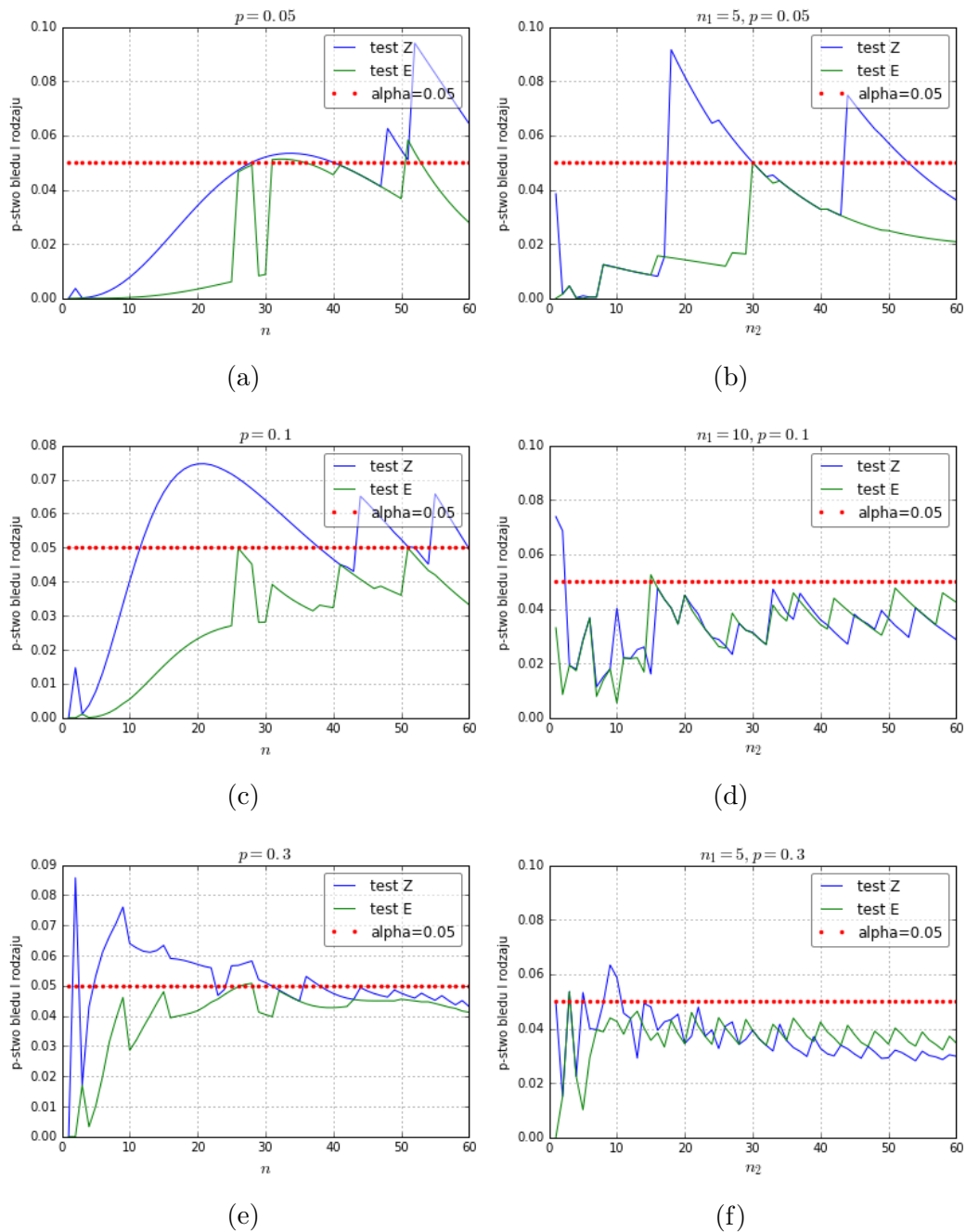
Rysunek 3.1 na stronie 21 przedstawia funkcję prawdopodobieństwa błędu I rodzaju w zależności od rozmiaru próbki dla różnych proporcji p . Po lewej stronie obserwacje z obu populacji są tej samej wielkości $n_1 = n_2 = n$, natomiast po prawej się różnią. Pierwszym spostrzeżeniem jest to, że funkcja dla testu E w większości przypadków jest mniejsza niż dla testu Z. Dodatkowo dla testu E prawdopodobieństwo błędu przekracza poziom istotności jedynie kilka razy, i to bardzo nieznacznie. Tymczasem dla testu Z funkcja częściej przyjmuje wartości powyżej α , dochodząc nawet do 0.95, jak na wykresie 3.1a. Warto wspomnieć także, że dla większych p (0.1 i 0.3) oraz wraz ze wzrostem wielkości próbki, funkcje są bliżej siebie i bardziej skupiają się przy poziomie istotności α .

Rysunek 3.2 na stronie 23 obrazuje również prawdopodobieństwo błędu I rodzaju, ale w zależności od proporcji p . Analizując wykresy funkcji, można dojść do analogicznych wniosków, jak dla rysunku 3.1. W każdym przypadku prawdopodobieństwo błędu I rodzaju testu Z jest większe od prawdopodobieństwa błędu testu E. Oprócz tego funkcja dla testu E przekracza poziom istotności jedynie na wykresie 3.2f, podczas gdy funkcja błędu testu Z tylko w jednym przypadku (3.2b) pozostaje całkowicie poniżej α .

Podsumowując, dla małych rozmiarów próbek test Z nie jest na poziomie istotności α , co wynika z zastosowania aproksymacji rozkładem normalnym do wyliczenia p -wartości. Przybliżenie rozkładem normalnym działa dobrze tylko dla dużych prób. Wobec tego test Z zbyt często odrzuca hipotezę zerową, gdy jest ona prawdziwa. Pamiętajmy, że analizujemy testy dla małych populacji, zatem to że test Z działa dla dużych prób nie jest wystarczające.

Rysunki 3.3 i 3.4 na stronach 24 i 25 zawierają wykresy mocy obu testów w zależności od rozmiaru próbki dla różnych proporcji oraz rozmiarów populacji. Druga populacja ma ustaloną proporcję p_2 , a pierwsza przyjmuje trzy wartości różne od p_2 . Dla obu testów zauważalny jest wzrost mocy wraz ze zwiększaniem się rozmiaru próbki. Ponadto moc testu wzrasta także dla proporcji bardziej oddalonych od siebie. W każdym przypadku, dla próby $n = 50$ i różnicy między proporcjami 0.3, moce mają wartość bliską 1.

Porównując, moc testu Z jest większa od mocy testu E, jednakże różnice nie są znaczne. Ponadto, z analizy błędu I rodzaju, wiemy, że test Z zbyt często odrzuca H_0 , co wpływa na wyższą moc testu. Natomiast test E utrzymuje



Rysunek 3.1: Prawdopodobieństwo błędu I rodzaju testów Z i E jako funkcja rozmiaru próbki n ; $\alpha = 0.05$; $N_1 = N_2 = 100$

poziom istotności α , wobec czego niewiele mniejsza moc testu nie oznacza, że test E jest gorszy.

3.2. Porównanie testu bez skończonej poprawki Zb z testem E

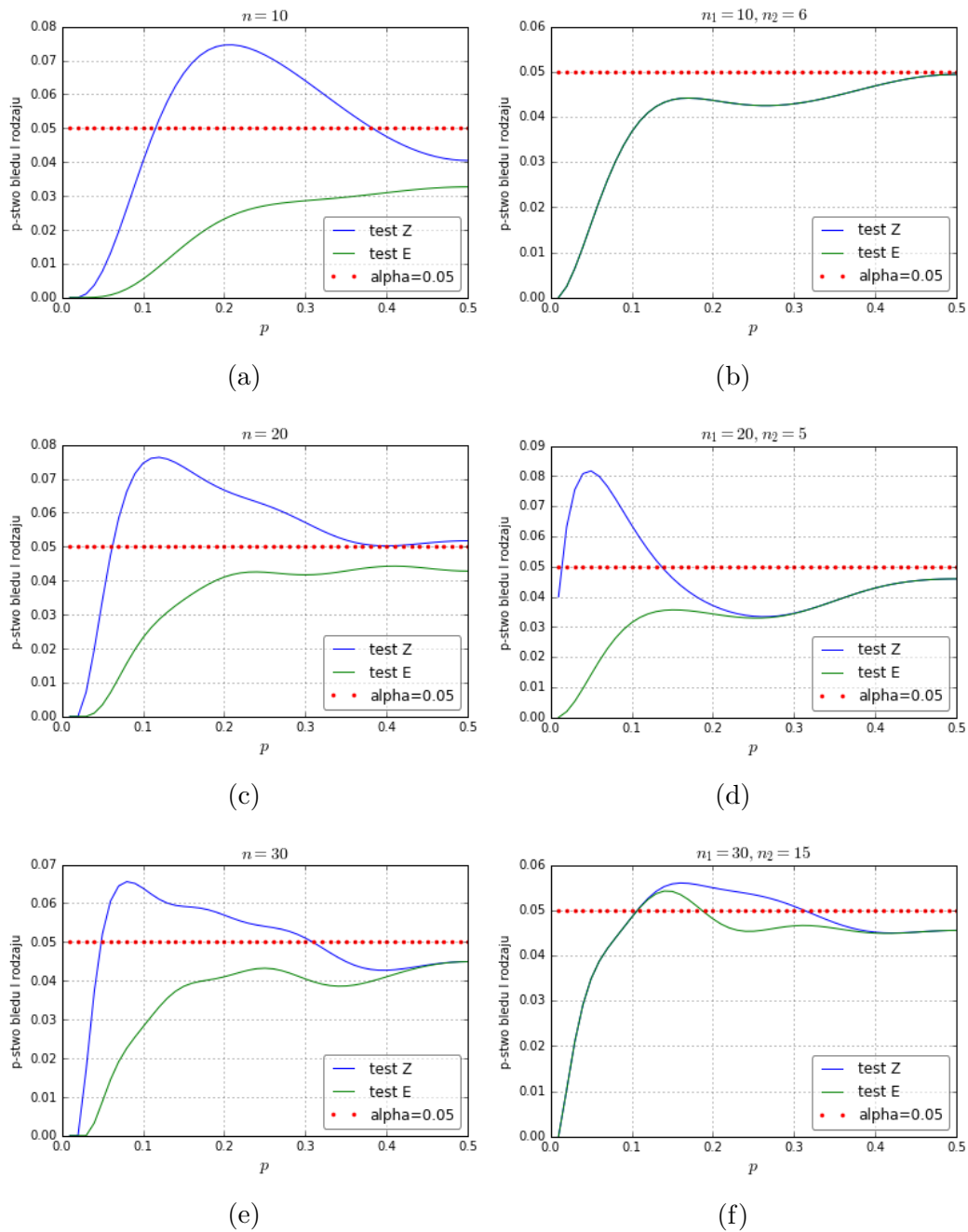
W niniejszym podrozdziale porównano test Zb z testem E. Nie włączono do analizy testu Z, z powodu wniosków opisanych w podrozdziale 3.1. Poniższe rozważania zostały przeprowadzone, aby potwierdzić zasadność stosowania testu ze skończoną poprawką dla małych populacji lub o stosunkowo dużej próbie.

Na rysunku 3.5 umieszczonym na stronie 26 znajdują się wykresy prawdopodobieństwa błędu I rodzaju analizowanych testów. Test E, jak stwierdzono w podrozdziale 3.1, nie przekracza poziomu istotności α . Jednakże dla testu Zb prawdopodobieństwo błędu I rodzaju znacząco przekracza poziom istotności. Funkcja błędu Zb przyjmuje wartości bliskie 0.2, gdzie największa znajduje się na wykresie 3.5d. Warto wspomnieć także, że funkcja jest największa w przypadku małych próbek. Na wykresach 3.5e i 3.5f widać, że mniej więcej do $n = 20$ test Zb jest powyżej poziomu α , następnie oscyluje, podobnie jak test E, w okolicach poziomu istotności.

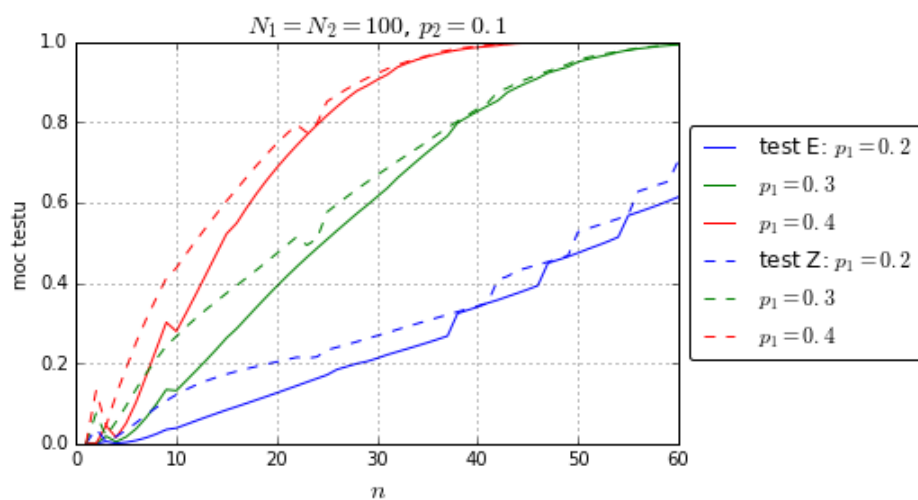
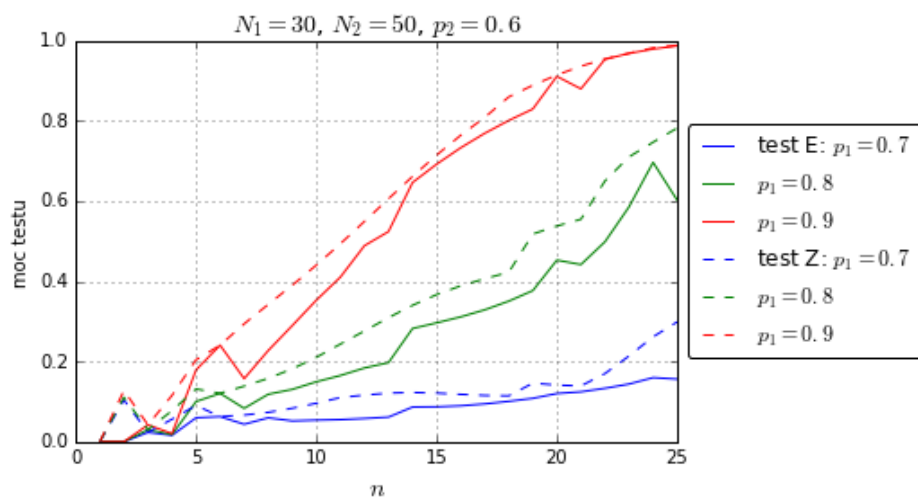
Podsumowując, test Zb nie jest adekwatny do rozważanego przypadku małej populacji gdy wartość n jest mała. Test Zb odrzuca hipotezę zerową dużo częściej, niż powinien test na poziomie α . Jest to spowodowane użyciem rozkładu dwumianowego w formułowaniu statystyki testowej. Faktycznie dla dużych próbek różnica jest nieznaczna, natomiast w sytuacji małej próbki rozbieżności są bardzo widoczne.

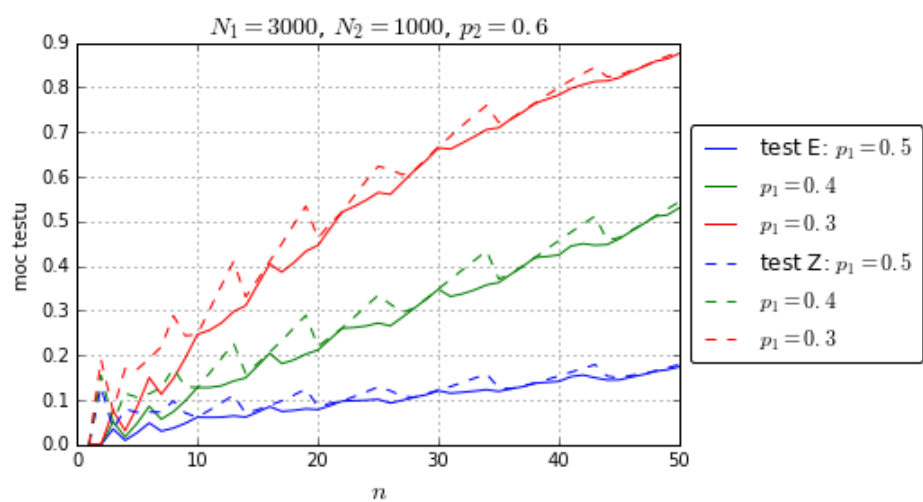
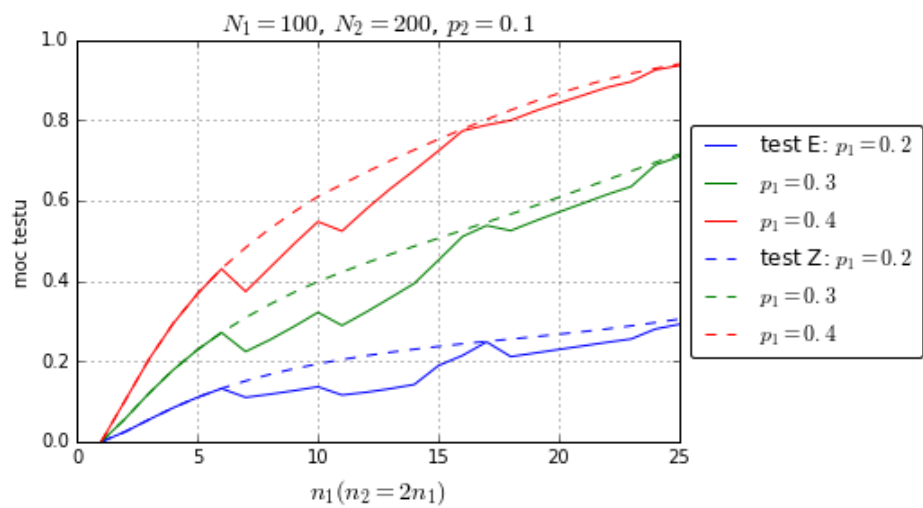
Rysunki 3.6 i 3.7 na stronach 27 i 28 przedstawiają wykresy mocy rozważanych testów w zależności od rozmiaru próbki dla różnych rozmiarów populacji oraz proporcji p . Moc testu Zb jest niewiele większa od mocy testu E. Jednakże podobnie jak w przypadku testu Z, wynika to ze zbyt częstego odrzucania H_0 . Biorąc pod uwagę ten fakt, wyższa wartość testu Zb nie świadczy o jego poprawności.

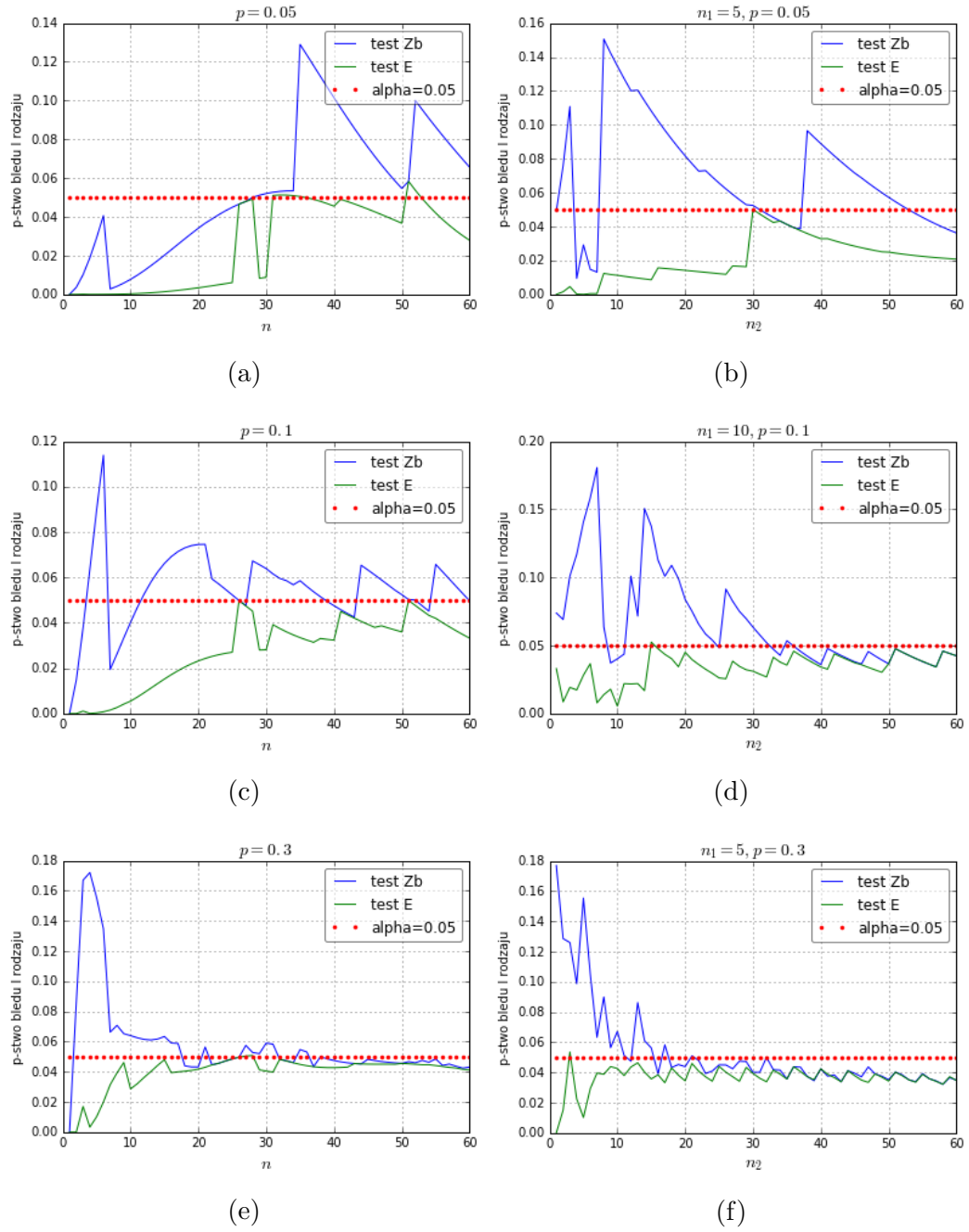
Ostatecznie pokazano, że test bez skończonej poprawki nie powinien być stosowany w przypadku małej populacji, ponieważ wtedy znacznie przekracza poziom istotności α . Najlepszym testem dla skończonej populacji okazał się test E, który dla różnych przypadków utrzymuje poziom istotności α .



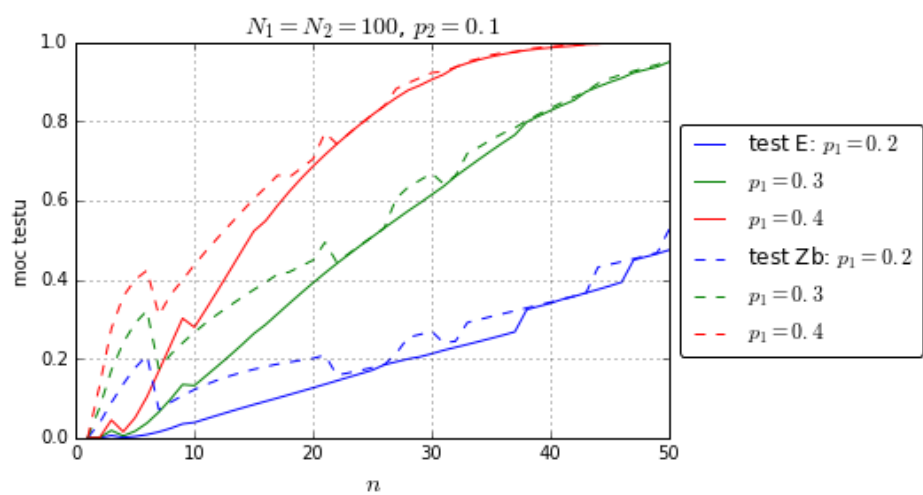
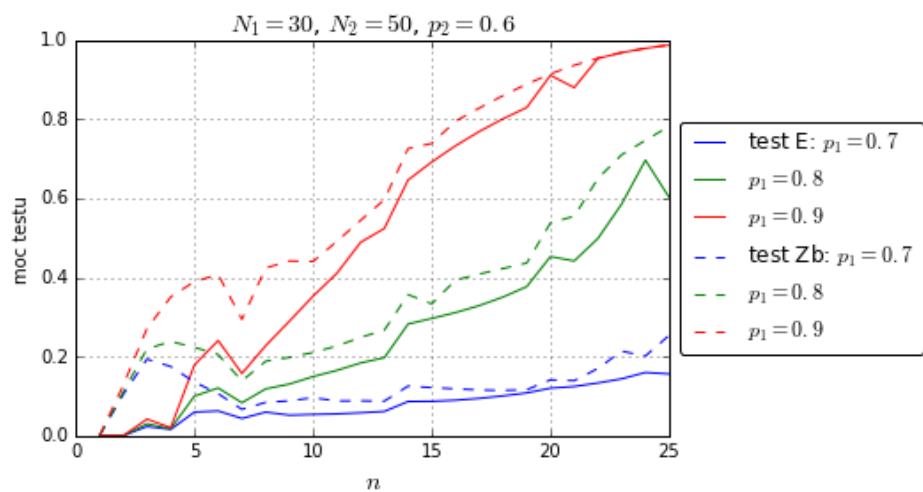
Rysunek 3.2: Prawdopodobieństwo błędu I rodzaju testów Z i E jako funkcja proporcji $p = M_1/N_1 = M_2/N_2$; $\alpha = 0.05$; $N_1 = N_2 = 100$

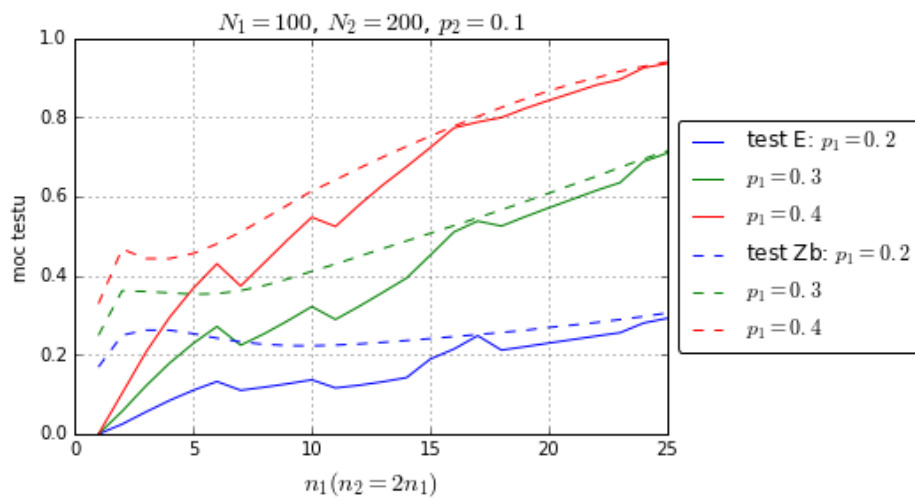
Rysunek 3.3: Moc testów Z i E jako funkcja rozmiaru próbki n , cz. 1

Rysunek 3.4: Moc testów Z i E jako funkcja rozmiaru próbki n , cz. 2

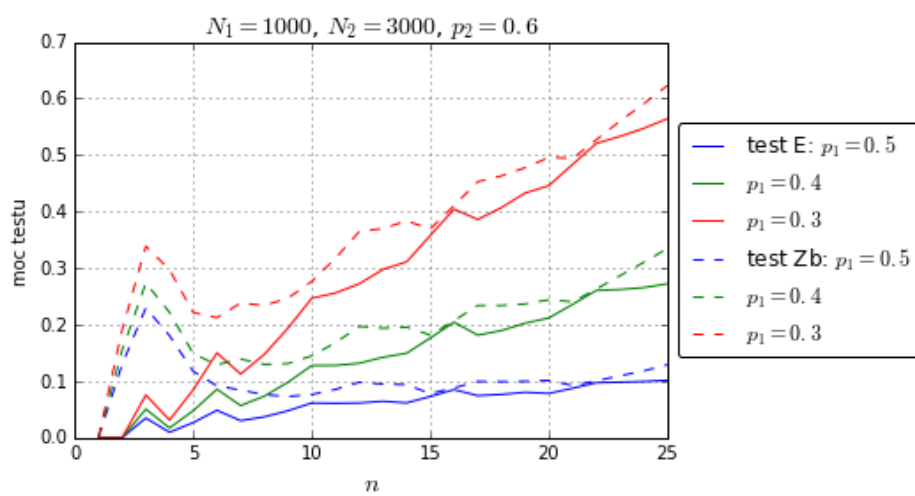


Rysunek 3.5: Prawdopodobieństwo błędu I rodzaju testów Zb i E jako funkcja rozmiaru próbki n ; $\alpha = 0.05$; $N_1 = N_2 = 100$

Rysunek 3.6: Moc testów Zb i E jako funkcja rozmiaru próbki n , cz. 1



(a)



(b)

Rysunek 3.7: Moc testów Zb i E jako funkcja rozmiaru próbki n , cz. 2

Podsumowanie

W pracy został rozważony problem testowania hipotez i estymacji w sytuacji populacji skończonego rozmiaru. W rozdziale 1 przedstawiono schematy pobierania obserwacji. W sytuacji nieskończonej populacji schemat opiera się o rozkład dwumianowy i próbka jest losowana ze zwracaniem. Natomiast, gdy populacja jest skończona, pobieranie obserwacji modeluje rozkład hipergeometryczny, przy czym elementy próbki są losowane bez zwracania. Po analizie obu rozkładów na konkretnym przykładzie sformułowano wnioski, że rozważane rozkłady dają podobne wyniki, jeśli populacja jest duża albo próbka stosunkowo mała. Jednakże, gdy obserwacja stanowi znaczną część populacji, albo sama populacja jest mała, to różnica między rozkładami staje się bardzo widoczna. Dzieje się tak, ponieważ rozkład hipergeometryczny uwzględnia wielkość populacji, a rozkład Bernoulliego nie uwzględnia takiej informacji.

Rozdział 3 zawiera analizę porównawczą testów opisanych w rozdziale 2, na podstawie wykresów prawdopodobieństwa błędu I rodzaju i mocy testu, dla różnych parametrów testowanych populacji. Porównanie dwóch testów ze skończoną poprawką wskazało na to, że test Z nie utrzymuje poziomu istotności α , ze względu na zastosowania aproksymacji rozkładem normalnym do statystyki testowej. Natomiast test E nie wykracza znacząco powyżej poziomu istotności oraz jego moc jest niewiele mniejsza od mocy testu Z. Tym samym, test E możemy uznać za dobry w przypadku małej populacji.

W celu pokazania zasadności stosowania testu ze skończoną poprawką do małych populacji, porównano także test E z testem bez skończonej poprawki. Na wykresach przedstawiających błąd I rodzaju możemy zauważyć, że test Zb znacząco przekracza poziom istotności α , szczególnie dla małych próbek. Oznacza to, że zbyt często odrzuca hipotezę zerową. Wobec czego stosowanie testu bez skończonej poprawki w sytuacji małej populacji wiąże się z dużymi błędami.

Problem skończonej populacji często występuje w medycynie. Lekarze badają przypadki, w których populacja jest mała, ze względu na konkretne cechy badanych osób. We Wrocławiu w Ośrodku Badawczo-Rozwojowym przy Wojewódzkim Szpitalu Specjalistycznym jeden z kardiologów przeprowadza badania na populacji dzieci z chorym sercem po operacji. Chce on sprawdzić jakie czynniki wpływają na wystąpienie powikłań pooperacyjnych u pacjentów. Niestety nie można na obecną chwilę przeprowadzić analizy danych, ponieważ badanie nie zostało jeszcze ukończone. Jednakże, aby sprawdzić, który czynnik jest najbardziej powiązany z powikłaniami, można zastosować opisany w pracy test E. Powinien dać on najlepsze wyniki spośród trzech testów przedstawionych w pracy.

Spis rysunków

| | | |
|-----|--|----|
| 1.1 | Funkcje prawdopodobieństwa $b(k; 10, 0.25)$ oraz $h(k; 10, 5, 20)$ | 10 |
| 1.2 | Funkcje prawdopodobieństwa $b(k; 17, 0.25)$ oraz $h(k; 17, 5, 20)$ | 10 |
| 1.3 | Funkcje prawdopodobieństwa $b(k; 20, 0.25)$ oraz $h(k; 20, 5, 20)$ | 10 |
| 1.4 | Wariancja rozkładów Bernoulliego i hipergeometrycznego w zależności od rozmiaru próbki | 11 |
| 3.1 | Prawdopodobieństwo błędu I rodzaju testów Z i E jako funkcja rozmiaru próbki n ; $\alpha = 0.05$; $N_1 = N_2 = 100$ | 21 |
| 3.2 | Prawdopodobieństwo błędu I rodzaju testów Z i E jako funkcja proporcji $p = M_1/N_1 = M_2/N_2$; $\alpha = 0.05$; $N_1 = N_2 = 100$ | 23 |
| 3.3 | Moc testów Z i E jako funkcja rozmiaru próbki n , cz. 1 | 24 |
| 3.4 | Moc testów Z i E jako funkcja rozmiaru próbki n , cz. 2 | 25 |
| 3.5 | Prawdopodobieństwo błędu I rodzaju testów Zb i E jako funkcja rozmiaru próbki n ; $\alpha = 0.05$; $N_1 = N_2 = 100$ | 26 |
| 3.6 | Moc testów Zb i E jako funkcja rozmiaru próbki n , cz. 1 | 27 |
| 3.7 | Moc testów Zb i E jako funkcja rozmiaru próbki n , cz. 2 | 28 |

Bibliografia

- [1] J. P. Buonaccorsi. A note on confidence intervals for proportions in finite populations. *The American Statistician*, 41(3):215–218, 1987.
- [2] F. Y. Edgeworth. On the value of a mean as calculated from a sample. *Journal of the Royal Statistical Society*, 81(4):624–632, Jul. 1918.
- [3] A. Hald. *A History of Probability and Statistics and Their Applications before 1750*. Wiley Interscience, 2003.
- [4] K. Krishnamoorthy, J. Thomson. Hypothesis testing about proportions in two finite populations. *The American Statistician*, 56(1):215–222, 2002.
- [5] E. L. Lehmann. *Testowanie hipotez statystycznych*. PWN, 1968.
- [6] B. E. Storer, C. Kim. Exact properties of some exact test statistics for comparing two binomial proportions. *Journal of the American Statistical Association*, 85:146–155, 1990.
- [7] W. Wang. Exact optimal confidence intervals for hypergeometric parameters. *Journal of the American Statistical Association*, 110(512):1491–1499, Dec. 2015.