Metody probabilistyczne i statystyka - wykład jedenasty[1] Podstawy teorii estymacji: estymacja przedziałowa. Podstawy testowania hipotez.

dr Jarosław Kotowicz

Instytut Informatyki Uniwersytet w Białymstoku

wersja z roku ak. 2020/21

Spis treści

Wstęp. I

Rozważamy

1. ustalone populację i cechę w tej populacji, która ma rozkład o dystrybuancie F zależnej od pewnego parametru θ ze zbioru parametrów Θ,
2. pewne (znane) funkcje $\gamma\colon \Theta \to \mathbb{R}$,
3. model statystyczny $(\mathcal{X}, \mathcal{F}, \{P_\theta : \theta \in \Theta\})$.

Estymacja przedziałowa, to metoda służącą do oszacowania przedziału liczbowego, który prawdopodobnie zawiera wartość szacowanej (nieznanej) wielkości $\gamma(\theta)$ dla każdej wartości parametru θ.

Wstęp. II

Uwaga 1

1. *Przedział ten nazywa się przedziałem ufności.*[a]
2. *Przedziały ufności dla nieznanych wielkości (różnych funkcji γ) buduje się na podstawie rozkładu estymatorów tych wielkości.*
3. *Do wyznaczenia przedziałów ufności wykorzystuje się dokładny rozkład estymatora (rozkład ustalony na podstawie znajomości rozkładu zmiennej w populacji, z której losowana jest próba) lub rozkład graniczny (rozkład do którego dąży estymator w miarę wzrostu liczebności próby do nieskończoności).*

---

[a] Twórcą pojęcia „*przedział ufności*” był statystyk polskiego pochodzenia Jerzy Spława-Neyman.

Pojęcie przedziału ufności i jego interpretacja. I

Definicja 1

Niech dane będą liczba α *z przedziału* $]0,1[$, *próba losowa* $X_1,\dots,X_n \sim_{iid} F(\cdot;\theta)$ *oraz funkcja* γ *nieznanego parametru* θ. *Rozważmy dwie statystyki* $\underline{\gamma} = \underline{\gamma}(X_1,\dots,X_n)$ *i* $\overline{\gamma} = \overline{\gamma}(X_1,\dots,X_n)$. *Mówimy, że przedział* $]\underline{\gamma};\overline{\gamma}[$ *jest przedziałem ufności dla* $\gamma(\theta)$ *na poziomie ufności* $1-\alpha$, *jeżeli*

$$P_\theta(\{\underline{\gamma}(X_1,\dots,X_n) < \gamma(\theta) < \overline{\gamma}(X_1,\dots,X_n)\}) \geqslant 1-\alpha$$

dla każdego θ.

Wtedy liczbę $1-\alpha$ nazywamy współczynnikiem (poziomem) ufności.

Pojęcie przedziału ufności i jego interpretacja. II

Uwaga 2

W przypadku, gdy rodzina rozkładów $\{P_\theta : \theta \in \Theta\}$ *składa się z rozkładów ciągłych, to wówczas możemy żądać, aby był spełniony warunek*

$$P_\theta(\{\underline{\gamma}(X_1, \ldots, X_n) < \gamma(\theta) < \overline{\gamma}(X_1, \ldots, X_n)\}) = 1 - \alpha$$

dla każdego θ.

Interpretacja 1

Przy wielokrotnym pobieraniu prób losowych n elementowej i wyznaczeniu na ich podstawie przedziału $]\underline{\gamma}(X_1, \ldots, X_n), \overline{\gamma}(X_1, \ldots, X_n)[$ *średnio* $(1 - \alpha) \cdot 100\%$ *przedziałów pokryje nieznaną wartość* $\gamma(\theta)$ *tzn. wartość* $\gamma(\theta)$ *będzie należała do tego przedziałów, a* $\alpha \cdot 100\%$ *tej wartości nie pokryje.*

Pojęcie przedziału ufności i jego interpretacja. III

W zastosowaniach praktycznych oznacza to, iż mamy $(1-\alpha)\cdot 100\%$ ufności, że wyznaczony na podstawie próby przedział będzie *„pokrywał"* szacowaną wielkość $\gamma(\theta)$.

Uwaga 3

1. *Poziom ufności* $1-\alpha$ *jest bliski jedności.*
2. *Zwykle poziom ufności przyjmuje się na poziomie 0,9; 0,95; 0,98; 0,99.*

Podstawą konstrukcji przedziału ufności dla danej wielkości $\gamma(\theta)$ jest *„dobry"* estymator tego parametru - tj. nieobciążony lub asymptotycznie nieobciążony, zgodny, najefektywniejszy lub asymptotycznie najefektywniejszy.

I tak będziemy wykorzystywać do konstrukcji przedziału ufności w przypadku

1. wartości przeciętnej – średnią arytmetyczną z próby,

Pojęcie przedziału ufności i jego interpretacja. IV

2. wariancji σ^2 – odpowiednią wariancję z próby,
3. frakcji (prawdopodobieństwa p) – częstość wystąpienia danego zdarzenia.

Sposób konstrukcji przedziału ufności związany jest z rozkładem odpowiedniego estymatora. Rozkład ten jest zależny od założeń dotyczących rozkładu cechy w zbiorowości generalnej oraz od liczebności próby.

Uwaga 4

Dla jednej wielkości $\gamma(\theta)$ mogą występować różne postacie przedziałów ufności.

Błędy estymacji przedziałowej. I

Definicja 2

Połowa długości przedziału ufności parametru $\gamma(\theta)$ jest nazywana bezwzględnym błędem losowym lub maksymalnym błędem szacunku, jaki popełniamy, szacując przedziałowe parametr $\gamma(\theta)$ za pomocą estymatora T_n, błąd ten oznaczamy przez d_{T_n}.
Natomiast

$$B(T_n) = \frac{d_{T_n}}{T_n} \cdot 100\%$$

nosi nazwę względnego błędu losowego lub względnej precyzji.

Definicja 3

Dokładnością estymacji przedziałowej nazywamy długość przedziału ufności.

Błędy estymacji przedziałowej. II

Uwaga 5

1. *Oszacowanie wielkości $\gamma(\theta)$ za pomocą przedziału ufności jest tym lepsze, im mniejszy jest ten błąd.*
2. *Praktycznie przyjmuje się, że przy względnym błędzie losowym wynoszącym do 5% wnioskowanie statystyczne jest całkowicie „bezpieczne". Jeśli błąd ten jest powyżej 5%, lecz nie przekracza 10%, rezultaty oszacowania są wątpliwe, a przy błędzie powyżej 10% wnioskowanie jest całkowicie niepewne. W tym ostatnim przypadku należy bądź zwiększyć próbę, bądź przyjąć mniejszy poziom ufności. Jest to szczególnie ważne przy szacowaniu prawdopodobieństwa sukcesu p.*
3. *Chcemy, aby długość przedziału ufności była jak najmniejsza.*

O czym będzie na dalszej części wykładu

Będziemy budować i rozważać przedziały ufności w następujących sytuacjach

1. Przedział ufności dla wartości oczekiwanej w populacji
    1. normalnej ze znanym odchyleniem standardowym,
    2. normalnej z nieznanym odchyleniem standardowym,
    3. o nieznanym rozkładzie i znanym odchyleniu standardowym,
    4. o nieznanym rozkładzie i nieznanym odchyleniu standardowym.
2. Przedział ufności dla wariacji dla populacji normalnej o nieznanych wartości oczekiwanej i odchylenia standardowego.
3. Przedział ufności dla frakcji w rozkładzie Bernoulliego.

Spostrzeżenie

Estymatorem wartości oczekiwanej jest średnią z próby tzn.

$$\widehat{m} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}.$$

Wiemy, że średnia z próby jest estymatorem zgodnym, nieobciążonym, najefektywniejszym i dostatecznym parametru m.

Stanowi więc ona podstawę konstrukcji przedziału ufności dla tego parametru. Stąd wiadomo, przy podanych wyżej założeniach, niezależnie od liczebności próby, czyli zarówno dla prób małych, jak i dużych, rozkład średniej arytmetycznej z próby jest rozkładem normalnym z parametrami m i $\frac{\sigma}{\sqrt{n}}$.

Populacja normalna ze znanym odchyleniem standardowym. I

Rozkład cechy jest rozkładem $\mathcal{N}(m, \sigma)$. Przyjmijmy oznaczenie $\theta = (m, \sigma)$.

Rozważmy średnią z próby $\overline{X}$. Dokonując standaryzacji zmiennej losowej $\overline{X}$ tzn.

$$U = \frac{\overline{X} - m}{\sigma} \sqrt{n}$$

otrzymujemy

$$P_\theta(\{|U| < u_\alpha\}) = 1 - \alpha.$$

Stąd przekształcając wyrażenie $\left| \frac{\overline{X} - m}{\sigma} \sqrt{n} \right| < u_\alpha$ mamy

$$P_\theta\left(\left\{\overline{X} - \frac{u_\alpha \sigma}{\sqrt{n}} < m < \overline{X} + \frac{u_\alpha \sigma}{\sqrt{n}}\right\}\right) = 1 - \alpha, \tag{1}$$

gdzie liczba rzeczywista u_α jest taka, że $\Phi(u_\alpha) = 1 - \frac{\alpha}{2}$.

Populacja normalna ze znanym odchyleniem standardowym. II

Uwaga 6

Zauważmy, że u_α *jest kwantylem rzędu* $1 - \frac{\alpha}{2}$.

Populacja normalna z nieznanym odchyleniem standardowym. I

Rozkład cechy jest rozkładem $\mathcal{N}(m,\sigma)$. Przyjmijmy oznaczenie $\theta = (m,\sigma)$.

Jak już powiedzieliśmy estymatorem parametru m jest średnia z próby, której rozkład nie może być wyznaczony z powodu nieznajomości odchylenia standardowego σ.

Statystyką, którą tu stosujemy jest statystyka t-Studenta tzn.

$$T = \frac{\overline{X} - m}{S}\sqrt{n-1} \quad \text{lub} \quad T = \frac{\overline{X} - m}{\tilde{S}}\sqrt{n}.$$

Otrzymujemy

$$P_\theta\left(\left\{\overline{X} - \frac{t_{\alpha,n-1}S}{\sqrt{n-1}} < m < \overline{X} + \frac{t_{\alpha,n-1}S}{\sqrt{n-1}}\right\}\right) = 1-\alpha, \tag{2}$$

$$P_\theta\left(\left\{\overline{X} - \frac{t_{\alpha,n-1}\tilde{S}}{\sqrt{n}} < m < \overline{X} + \frac{t_{\alpha,n-1}\tilde{S}}{\sqrt{n}}\right\}\right) = 1-\alpha, \tag{3}$$

Populacja normalna z nieznanym odchyleniem standardowym. II

gdzie $t_{\alpha,n-1}$ jest takie, że $P_\theta(\{|T| < t_{\alpha,n-1}\}) = 1 - \alpha$.

Uwaga 7

1. *Podobnie, jak w poprzednim wypadku $t_{\alpha,n-1}$ jest kwantylem rzędu $1 - \frac{\alpha}{2}$.*
2. *Dla $n > 120$ wartości $t_{\alpha,n-1}$ zastępuje się u_α.*
3. *Zwykle długość przedziału w przypadku znanego odchylenia standardowego jest mniejsza niż w przypadku, gdy odchylenie standardowe nie jest znane.*

Populacja o nieznanym rozkładzie, ale znanym odchyleniem standardowym

Rozkład graniczny estymatora wartości oczekiwanej ma rozkład normalnym.

W związku z tym przybliżamy średnią z próby rozkładem normalnym parametrami m i $\frac{\sigma}{\sqrt{n}}$. Dokonując standaryzacji tak, jak w pierwszym przypadku otrzymujemy

$$P_\theta\left(\left\{\overline{X} - \frac{u_\alpha \sigma}{\sqrt{n}} < m < \overline{X} + \frac{u_\alpha \sigma}{\sqrt{n}}\right\}\right) = 1 - \alpha, \tag{4}$$

dla dostatecznie dużych n tj. dla $n > 120$. Oznaczyliśmy $\theta = (m, \frac{\sigma}{\sqrt{n}})$.

Populacja o nieznanym rozkładzie z nieznanym odchyleniem standardowym

Dla $n > 120$ przyjmujemy $\sigma = S$ i sprowadzamy do ostatniego przypadku. Otrzymujemy

$$P_\theta\left(\left\{\overline{X} - \frac{u_\alpha S}{\sqrt{n}} < m < \overline{X} + \frac{u_\alpha S}{\sqrt{n}}\right\}\right) \approx 1 - \alpha. \tag{5}$$

Ponownie oznaczyliśmy $\theta = (m, \frac{\sigma}{\sqrt{n}})$.

Uwaga 8

Jeżeli próba jest mała, to nie możemy wyznaczyć przedziału ufności dla średniej.

Spostrzeżenia. I

Uwaga 9

1. *Przy zadanym poziomie ufności* $1 - \alpha$ *im większa jest liczebność, tym krótszy jest przedział ufności.*
2. *Przy ustalonej liczebności próby wraz ze wzrostem poziomu ufności rośnie rozpiętość (długość) przedziału ufności.*
3. *Im krótszy przedział, tym mniejszy błąd szacunku, co oznacza większy dokładność (precyzję) oszacowania.*

Spostrzeżenia. II

Przykład 1

Zakładając, że kwartalne wydatki na reklamę można uznać za cechę o rozkładzie $\mathcal{N}(m, \sigma)$, wylosowano do próby 100 zakładów usługowych i otrzymano następujący rozkład wydatków na reklamę:

Kwartalne wydatki	0-5	5-10	10-15	15-20
Liczba zakładów	10	20	40	30

- *Wyznacz na poziomie ufności* $1 - \alpha = 0{,}96$ *przedział ufności dla przeciętych kwartalnych wydatków na reklamę.*
- *Jaka będzie dokładność oszacowania, gdy poziom ufności będzie równy 0,9?*

Założenia

Uwaga 10

Przedział ufności dla wariancji można wyznaczyć tylko wówczas, gdy cecha X charakteryzująca zbiorowość generalną ma rozkład normalny $\mathcal{N}(m, \sigma)$.

Estymatorem wariancji jest wariancja z próby wyrażona jednym ze wzorów

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{6}$$

$$\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{7}$$

$$\dot{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2, \qquad \text{gdy znane jest } m. \tag{8}$$

Populacja normalna o znanej wartości oczekiwanej

Statystyką, którą tu stosujemy jest statystyka chi-kwadrat o n stopniach swobody tzn.

$$\chi^2 = \frac{n\dot{S}^2}{\sigma^2}.$$

Mamy wtedy

$$P_\theta(\{\chi^2 \geqslant \chi^2_{\frac{\alpha}{2},n}\}) = \frac{\alpha}{2} \quad \text{oraz} \quad P_\theta(\{\chi^2 \geqslant \chi^2_{1-\frac{\alpha}{2},n}\}) = 1 - \frac{\alpha}{2},$$

gdzie jak poprzednio $\theta = (m, \sigma)$. Otrzymujemy wtedy

$$P_\theta\left(\left\{\frac{n\dot{S}^2}{\chi^2_{\frac{\alpha}{2},n}} < \sigma^2 < \frac{n\dot{S}^2}{\chi^2_{1-\frac{\alpha}{2},n}}\right\}\right) = 1 - \alpha. \tag{9}$$

Populacja normalna o nieznanej wartości oczekiwanej. I

Statystyką, którą tu stosujemy jest statystyka chi-kwadrat o $n-1$ stopniach swobody tzn.

$$\chi^2 = \frac{nS^2}{\sigma^2}.$$

Analogicznie, jak poprzednio mamy

$$P_\theta(\{\chi^2 \geqslant \chi^2_{\frac{\alpha}{2},n-1}\}) = \frac{\alpha}{2} \quad \text{oraz} \quad P_\theta(\{\chi^2 \geqslant \chi^2_{1-\frac{\alpha}{2},n-1}\}) = 1 - \frac{\alpha}{2}$$

oraz

$$P_\theta\left(\left\{\frac{nS^2}{\chi^2_{\frac{\alpha}{2},n-1}} < \sigma^2 < \frac{nS^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right\}\right) = 1 - \alpha. \tag{10}$$

Populacja normalna o nieznanej wartości oczekiwanej. II

Uwaga 11

Można również wykorzystać statystykę

$$\chi^2 = \frac{(n-1)\tilde{S}^2}{\sigma^2},$$

która ma również $n-1$ stopni swobody.

Przykład 2 ([1, Przykład 8.10])

Zakładamy, że czas pracy żarówek produkowanych ma rozkład $\mathcal{N}(750, \sigma)$. Na podstawie 16 elementowej próby losowej otrzymano $\dot{s}^2 = 2500$. Wyznaczymy 98-procentowy przedział ufności dla odchylenia standardowego czasu pracy żarówek.

ROZWIĄZANIE.

Populacja normalna o nieznanej wartości oczekiwanej. III

Najpierw wyznaczamy przedział ufności dla wariancji σ^2 na poziomie ufności $1 - \alpha = 0{,}98$ (m jest znane i $n < 30$). Ponieważ szukamy

$$\chi^2_{0,99;16} \quad \text{i} \quad \chi^2_{0,01;16},$$

więc z tablic rozkładu χ^2 mamy

$$\chi^2_{0,99;16} = 5{,}812 \quad \text{i} \quad \chi^2_{0,01;16} = 31{,}999.$$

Stad

$$35{,}36 < \sigma < 82{,}96.$$

Populacja normalna o nieznanej wartości oczekiwanej. IV

Uwaga 12

1. *Statystyka S ma rozkład asymptotycznie normalny o parametrach* $\mathcal{N}\left(m, \frac{\sigma}{\sqrt{2n}}\right)$.
2. *Z ostatniej obserwacji wynika, że przedział ufności dla odchylenia standardowego, gdy $n > 30$ wyznaczamy z jednego ze wzorów*

$$\frac{S}{1 + \frac{u_\alpha}{\sqrt{2n}}} < \sigma < \frac{S}{1 - \frac{u_\alpha}{\sqrt{2n}}}$$

lub

$$\frac{\dot{S}}{1 + \frac{u_\alpha}{\sqrt{2n}}} < \sigma < \frac{\dot{S}}{1 - \frac{u_\alpha}{\sqrt{2n}}}$$

w zależności od tego, czy znana, czy też nieznana jest wartość oczekiwana m. Natomiast u_α spełnia warunek $\Phi(u_\alpha) = 1 - \frac{\alpha}{2}$.

Przedział ufności frakcji (parametru p) w rozkładzie Bernoulliego. I

Uwaga 13

Stosujemy, gdy liczebność próbki wynosi co najmniej 100.

Stosujemy statystykę $\hat{p} = \frac{X}{n}$, gdzie X jest ilością wystąpień sukcesów w próbie n elementowej.

Rozkładem granicznym tej statystyki jest rozkład normalny $\mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$.

Dokonując jego standaryzacji, wykorzystując rozkład normalny oraz zastępując $\frac{p(1-p)}{n}$ przez $\frac{\hat{p}(1-\hat{p})}{n}$ otrzymujemy

$$P\left(\left\{\hat{p} - u_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + u_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right\}\right) \approx 1 - \alpha, \tag{11}$$

Przedział ufności frakcji (parametru p) w rozkładzie Bernoulliego. II

gdzie jak poprzednio u_α spełnia warunek $\Phi(u_\alpha) = 1 - \frac{\alpha}{2}$.

Spis treści

Motywacja i uwagi

Cel – kształtowanie warunków estymacji przedziałowej, aby otrzymać oszacowanie o żądanej dokładności przy jak najmniejszej próbie.

Rozpatrywać będziemy następujące problemy:

1. minimalizacja liczebności próby w estymacji przedziałowej wartości średniej w populacji o rozkładzie normalnym $\mathcal{N}(m, \sigma)$ cechy ze znanym odchyleniem standardowym,
2. minimalizacja liczebności próby w estymacji przedziałowej frakcji w populacji o rozkładzie dwumianowym cechy.

Minimalizacja wielkości próby w estymacji przedziałowej wartości oczekiwanej w populacji normalnej. I

Mamy wtedy przedział ufności $\left]\overline{X} - \frac{u_\alpha \sigma}{\sqrt{n}}, \overline{X} + \frac{u_\alpha \sigma}{\sqrt{n}}\right[$ o długości $\frac{2u_\alpha \sigma}{\sqrt{n}}$ (dokładność estymacji przedziałowej).

Dokładnością estymacji przedziałowej zależy od

1. wartości parametru σ określającego rozproszenie,
2. poziomu ufności $1 - \alpha$ określającego u_α,
3. liczebności próby.

Dokładność można modyfikować poprzez

- poziom współczynnika ufności,
- liczebność próby.

Minimalizacja wielkości próby w estymacji przedziałowej wartości oczekiwanej w populacji normalnej. II

Uwaga 14

1. *Przypomnijmy, że połowę długości przedziału ufności nazywamy maksymalnym błędem szacunku.*
2. *W przypadku pierwszym skracanie przedziału może powodować zmniejszenie prawdopodobieństwa pokrycia parametru. W przypadku drugim zakładamy, że* $\frac{u_\alpha \sigma}{\sqrt{n}} \leqslant d$, *gdzie* d *jest zadaną z góry liczbą. Stąd* $n \geqslant \frac{u_\alpha^2 \sigma^2}{d^2}$.
3. *W przypadku, gdy odchylenie standardowe nie jest znane, to rozważamy nieobciążony estymator wariancji* $\tilde{S}^2$ *zamiast* σ^2 *oraz wyznaczamy wartość krytyczną* t_α *dla rozkładu* $t-$*Studenta o* $n-1$ *stopniach swobody zastępując nią* u_α, *gdzie* n *jest, jak zwykle, liczebnością próby.*

Minimalizacja wielkości próby w estymacji przedziałowej wartości oczekiwanej w populacji normalnej. III

Przykład 3 ([1, Przykład 8.13])

Załóżmy, że rozkład wagi uczniów pierwszych klas można ująć jako $N(m, \sigma)$. Na podstawie 10-elementowej próby otrzymano $\tilde{s}^2 = 16$. Ilu uczniów należy wylosować do próby, aby oszacować przeciętną wagę z maksymalnym błędem 0,5 kg na poziomie ufności $1 - \alpha = 0{,}98$?

ROZWIĄZANIE.

Korzystając z tablic rozkładu t-Studenta, dla 9 stopni swobody i $\alpha = 0{,}02$ otrzymujemy

$$t_{0{,}02;9} = 2{,}821.$$

Stąd

$$n = \frac{(2{,}821)^2 \cdot 16}{(0{,}5)^2} = 509{,}3,$$

a więc $n = 510$.

Minimalizacja próby w estymacji przedziałowej frakcji w populacji o rozkładzie dwumianowym

Mamy wtedy przedział ufności $\left]\hat{p} - u_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + u_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right[$. Analogicznie jak w punkcie pierwszym zakładamy, że $u_\alpha\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leqslant d$, gdzie d jest zadaną z góry liczbą.

Stąd $n \geqslant \frac{u_\alpha^2(1-p)p}{d^2}$.

Uwaga 15

Należy podkreślić, że jeżeli nie jest znany rząd wielkości parametru p, to ustalamy, że $p = \frac{1}{2}$.

Spis treści

Eksperyment opisany w [2]. I

Rink Hoekstra, Richard D. Morey, Jeffrey N. Rouder i Eric-Jan Wagenmakers w swojej pracy [2] rozważali następujący kwestionariusz, który wypełniali studenci, doktoranci i pracownicy naukowi uczelni w Niderlandach (Holandia).

Badacz przeprowadza eksperyment i stwierdza *„95% przedział ufności dla średniej wynosi od 0,1 do 0,4"*.

Proszę zaznaczyć każde z poniższych stwierdzeń jako *„prawda"* lub *„fałsz"*. Fałsz oznacza, że stwierdzenie nie wynika logicznie z cytowanego wyniku.

1. Prawdopodobieństwo, że średnia w populacji jest większa niż 0, wynosi co najmniej 95%.
2. Prawdopodobieństwo, że średnia w populacji wynosi 0, jest mniejsze niż 5%.
3. *„Hipoteza zerowa"*, że średnia w populacji równa się 0, jest prawdopodobnie nieprawidłowa.

Eksperyment opisany w [2]. II

4. Istnieje 95% prawdopodobieństwa, że średnia w populacji leży między 0,1 a 0,4.
5. Możemy być w 95% pewni, że średnia w populacji mieści się między 0,1 a 0,4.
6. Jeśli powtarzamy eksperyment w nieskończenie wiele razy, to w 95% przypadków średnia w populacji mieści się w przedziale od 0,1 do 0,4.
7. Jeśli powtarzamy eksperyment w nieskończenie wiele razy, to w 95% przypadków przedziały ufności skonstruowane na podstawie próby zawierają średnią w populacji.

Podsumowanie. I

95%-owy przedział ufności mówi o tym, że jeżeli losowalibyśmy wielokrotnie próbę losową (tzn. nieskończenie wiele razy) i na za każdym razem na podstawie próby wyznaczalibyśmy przedział ufności dla szacowanego parametru, to średnio 95 takich przedziałów na 100 będzie zawierało prawdziwą wartość szacowanego parametru.

Uwaga 16

Przypuśćmy, że otrzymaliśmy na podstawie próby losowej 95% przedział ufności $]A, B[$ *dla parametru* $\gamma(\theta)$*. Wtedy należy pamiętać i wiedzieć, że w statystyce matematycznej klasycznej (częstotliwościowej) zachodzą poniższe stwierdzenia.*

1. Prawdziwa wartość szacowanego parametru $\gamma(\theta)$ jest wartością stała i nieznaną.
2. Prawdziwa wartość szacowanego parametru $\gamma(\theta)$ w populacji albo jest w tym przedziale, albo jej tam nie ma.

Podsumowanie. II

3. Przedział ufności nie mówi o tym, że na 95% prawdziwa wartość parametru $\gamma(\theta)$ jest gdzieś pomiędzy A a B.
4. Nie możemy odnosić się do parametru $\gamma(\theta)$ mówiąc o jakimkolwiek prawdopodobieństwie jego dotyczącym.
5. Nie możemy mówić, że na 95% wartość parametru w populacji zawiera się w jakimś przedziale.

Spis treści

Hipoteza statystyczna i zbiór hipotez dopuszczalnych. I

Testowanie hipotez statystycznych obejmuje zasady i metody sprawdzania określonych przypuszczeń, inaczej założeń, dotyczących parametrów lub postaci rozkładu cech statystycznych populacji generalnej na podstawie wyników z próby prostej.

Definicja 4

Hipotezą statystyczną nazywamy każdy sąd o zbiorowości generalnej, wydany bez przeprowadzenia badania całkowitego.

Uwaga 17

Tak więc hipoteza statystyczną jest dowolne przypuszczenie co do rozkładu cechy w populacji generalnej (jego postaci funkcyjnej lub wartości parametrów).

Hipoteza statystyczna i zbiór hipotez dopuszczalnych. II

Definicja 5

Zbiorem hipotez dopuszczalnych nazywamy zbiór wszystkich możliwych rozkładów, które mogą charakteryzować populację generalną.
Standardowo będziemy ten zbiór zapisywać $\{P_\theta : \theta \in \Theta\}$.

Uwaga 18

1. *Zauważmy, że elementy zbioru hipotez dopuszczalnych są indeksowane parametrem. Tak więc zbiór hipotez dopuszczalnych można utożsamić z przestrzenią parametrów.*
2. *Hipoteza statystyczna jest pewnym podzbiorem hipotez dopuszczalnych.*
3. *Hipotezę statyczną można więc zapisać następująco*

$$H : \theta \in \Theta_0 \wedge \Theta_0 \subseteq \Theta.$$

Klasyfikacja hipotez statystycznych. I

Podział hipotez ze względu na ilość rozkładów jakie może przyjmować hipoteza

1. prosta – hipoteza jednoznacznie wyznacza rozkład (podzbiór Θ_0, do którego ma należeć parametr, jest jednoelementowy);
2. złożona – hipoteza niejednoznacznie wyznacza rozkład (podzbiór Θ_0, do którego ma należeć parametr co najmniej dwuelementowy).

Podział hipotez ze względu czego dotyczą

1. parametryczna – dotyczy wartości parametrów rozkładu cechy w populacji generalnej;
2. nieparametryczna – (pozostałe hipotezy).

Klasyfikacja hipotez statystycznych. II

Przykład 4

Załóżmy, że cecha w populacji generalnej ma rozkład skokowy.
Wtedy zbiór hipotez dopuszczalnych $\{P_\theta : \theta \in \Theta\}$ *zawiera wszystkie możliwe rozkłady skokowe. Ponadto*

- *hipoteza, że cecha w populacji ma rozkład Poissona jest hipotezą nieparametryczna i złożona,*
- *hipoteza, że cecha w populacji ma rozkład Poissona z wartością oczekiwaną równą* 1 *jest hipotezą nieparametryczna i prosta,*
- *hipoteza, że cecha w populacji ma wartość oczekiwaną równą* 1 *jest hipotezą parametryczna i złożona.*

Klasyfikacja hipotez statystycznych. III

Przykład 5

Załóżmy, że cecha w populacji generalnej ma rozkład normalny.
Wtedy zbiór hipotez dopuszczalnych $\{P_\theta : \theta \in \Theta\} = \{\mathcal{N}(m, \sigma) : (m, \sigma) \in \mathbb{R} \times \mathbb{R}_+\}$.
Ponadto

- *hipoteza, że cecha w populacji ma rozkład z wartością oczekiwaną równą* 1 *jest hipotezą parametryczna i złożona,*
- *hipotez, że cecha w populacji ma rozkład z parametrem* $\theta = (1, 1)$ *jest hipotezą parametryczna i prostą.*

Hipoteza zerowa i alternatywna. I

Przy testowaniu hipotez formułuje się dwie hipotezy zerową i alternatywną.

Definicja 6

Hipotezą zerową, oznaczaną H_0, nazywamy hipotezę sprawdzaną (testowaną, weryfikowaną). Hipotezą alternatywną (konkurencyjną), oznaczaną H_1, nazywamy hipotezę, którą jesteśmy skłonni przyjąć, jeżeli odrzucamy hipotezę zerową.

Hipoteza zerowa i alternatywna. II

Uwaga 19

1. *Prawdziwość hipotezy zerowej jest oceniana na podstawie wyników próby losowej.*
2. *"Obie „konkurencyjne" hipotezy traktujemy nierównoprawnie. Zasadniczo, interpretacja jest taka: H_0 jest założeniem obowiązującym do czasu, gdy pojawi się dane doświadczalne sprzeczne (lub raczej „bardzo trudne do pogodzenia") z tą hipotezą. Z kolei, H_1 jest „ewentualnością, z którą powinniśmy się liczyć", jeżeli przyjdzie nam zrezygnować z hipotezy H_0."*[a]
3. *"Za klasyczną teorią testowania stoi ważna idea metodologiczna. Jest to zasada konserwatyzmu: nie należy rezygnować z ustalonej teorii (hipotezy zerowej), jeżeli nie ma po temu koniecznych lub przynajmniej bardzo wyraźnych powodów."*[b]

[a] Niemiro Wojciech. *Statystyka I*. 2014.
[b] Niemiro Wojciech. *Statystyka I*. 2014.

Test statystyczny. I

Jak już wiemy, prawdziwość hipotezy zerowej weryfikujemy na podstawie wyników z próby. Dokonujemy tego za pomocą testu statystycznego, który w zależności od rodzaju hipotezy może być

1. testem parametrycznym (dla hipotez parametrycznych),
2. testem nieparametrycznym (w przypadku hipotez nieparametrycznych).

Definicja 7

Test statystyczny nazywamy regułę postępowania, która przyporządkowuje wynikom próby losowej decyzję przyjęcia lub odrzucenia hipotezy zerowej.

Test statystyczny. II

Uwaga 20

Formalna definicja testu statystycznego (dokładniej testu statystycznego niezrandomizowanego) jest następująca:
Testem hipotezy H_0 przeciw alternatywie H_1 nazywamy statystykę

$$\delta\colon \mathcal{X} \to \{0, 1\},$$

gdzie $\mathcal{X}$ jest przestrzenią realizacji (obserwacji). Natomiast wartość „1" interpretujemy jako decyzję o odrzuceniu hipotezy H_0, zaś „0" oznacza, że nie odrzucamy H_0.

Uwaga 21

Tak więc test statystyczny rozstrzyga, jakie wyniki próby pozwalają uznać, że nie ma podstaw do odrzucenia hipotezy zerowej, a jakie przemawiają za jej odrzuceniem i przyjęciem w to miejsce hipotezy alternatywnej.

Sprawdzian hipotezy i zbiór krytyczny. I

Definicja 8

Sprawdzianem hipotezy (statystyką testową) nazywamy statystykę T_n będącą funkcją próby losowej, o znanym rozkładzie, której wartość empiryczna t_n, policzona na podstawie próby losowej, pozwala na podjęcie decyzji, czy przyjąć, czy też odrzucić hipotezę zerową.

Uwaga 22

Jeżeli testem statystycznym jest δ hipotezy H_0 przeciwko H_1, to najczęściej mamy

$$\delta(X) = \mathbb{I}_{\{T(X)>c\}},$$

gdzie $X = (X_1, \ldots, X_n)$ jest próbą losową, $T(X)$ jest pewną statystyką testową, a c wartością krytyczną.

Sprawdzian hipotezy i zbiór krytyczny. II

Definicja 9

Zbiorem krytycznym Λ (obszarem odrzucenia) nazywamy zbiór tych wartości sprawdzianu hipotezy, które przemawiają za odrzuceniem hipotezy zerowej.

Uwaga 23

Zauważmy, że dla testu δ mamy

$$\Lambda := \{x \in \mathcal{X} : \delta(x) = 1\}.$$

1. W przypadku hipotez parametrycznych sprawdzianem hipotezy jest estymator lub określona funkcja estymatora tego parametru, którego dotyczy weryfikowana hipoteza. Postać tej zmiennej losowej zależy od tego, w jakich warunkach przebiegu testowanie, czyli jakie są założenia o rozkładzie cechy w zbiorowości generalnej oraz jak dużą próbą dysponujemy.

Sprawdzian hipotezy i zbiór krytyczny. III

2. Zbiór Λ może być w zależności od postaci hipotezy alternatywnej zbiorem jednostronnym (prawostronnym lub lewostronnym) albo zbiorem dwustronnym. Mówimy wtedy również, że test statystyczny jest jednostronny (prawostronny, lewostronny) lub dwustronny. Rozkład sprawdzianu hipotezy określa, z jakich tablic należy odczytać wartość krytyczną, wyznaczającą zbiór Λ, a zatem zbiór Λ zależy również od liczebności próby n, od tego, czy znamy parametry $(m, \sigma$ lub $p)$ w zbiorowości generalnej oraz od poziomu istotności α.

Etapy konstrukcji testu statystycznego. I

1. Formułowanie hipotezy podlegającej weryfikacji tzn. hipotezy zerowej
$$H_0 : \theta \in \Theta_0 \wedge \Theta_0 \subseteq \Theta.$$
2. Formułowanie hipotezy alternatywnej
$$H_1 : \theta \in \Theta_1 \wedge \Theta_1 \subseteq \Theta,$$
będącej zaprzeczeniem[2] hipotezy zerowej. Przyjmiemy ją za prawdziwą w przypadku odrzucenia hipotezy zerowej.
3. Określenie sprawdzianu hipotezy.
4. Określenie zbioru krytycznego hipotezy.

Etapy konstrukcji testu statystycznego. II

Uwaga 24

Niech W będzie przestrzenią próby[a]*, $t_n = T_n(x_1, \ldots, x_n)$ jej konkretna realizacją, Λ obszarem krytycznym. Jeżeli $t_n \in \Lambda$, to hipotezę zerową odrzucamy. Natomiast jeśli $t_n \in W \setminus \Lambda$, to nie ma podstaw do odrzucenia hipotezy zerowej.*

[a] Jest to zbiór wszystkich możliwych wyników próby.

[2] Nie musi być to całkowita negacja hipotezy zerowej.

Błędy testowania hipotez. I

Błędy testowania hipotez dzielimy na

1. błąd **I**-go rodzaju, oznaczany α – odrzucenie na podstawie wyników z próby hipotezy zerowej, która jest prawdziwa

$$\Theta_0 \ni \theta \mapsto \alpha(\theta) := P_\theta(\{T_n \in \Lambda\}). \tag{12}$$

2. błąd **II**-go rodzaju, oznaczany β – przyjęcie na podstawie wyników z próby hipotezy zerowej, która jest fałszywa

$$\Theta_1 \ni \theta \mapsto \beta(\theta) := P_\theta(\{T_n \in W \setminus \Lambda\}) \quad \text{dla } \theta \in \Theta_1. \tag{13}$$

Błędy testowania hipotez. II

Uwaga 25

Dla testu $\delta = \mathbb{I}_{\{T(X)>c\}}$ *hipotezy* H_0 *przeciwko hipotezie* H_1*, gdzie* $\Theta = \Theta_0 \cup \Theta_1$ *rozpatruje się odwzorowanie*

$$\Theta \ni \theta \mapsto P_\theta(\{\delta(X) = 1\}) = 1 - \beta(\theta)$$

nazywane mocą testu i wtedy

- *błąd I-go rodzaju to* $\Theta_0 \ni \theta \mapsto \alpha(\theta) = 1 - \beta(\theta) = P_\theta(\{\delta(X) = 1\})$*,*
- *błąd II-go rodzaju to* $\Theta_1 \ni \theta \mapsto \beta(\theta) = P_\theta(\{\delta(X) = 0\})$*.*

Błędy testowania hipotez. III

		hipoteza zerowa H_0	
		prawdziwa	fałszywa
decyzja	przyjąć	brak błędu	błąd II-go rodzaju
dla H_0	odrzucić	błąd I-go rodzaju	brak błędu

Tabela: Opracowanie własne.

Uwaga w statystyce rozważa się macierz błędów inaczej tablice pomyłek

		klasa rzeczywista	
		pozytywna	negatywna
klasa	pozytywna	true positve (TP)	false positive (FP)
predykowana	negatywna	false negative (FN)	true negative (TN)

Błędy testowania hipotez. IV

Tabela: Opracowanie własne.

Uwaga 26

1. *Wartości α są bliskie zera i na ogół są równe 0,01; 0,02; 0,05; 0,1.*
2. *Dobry test statystyczny powinien mieć tę własność, że również β powinno być bliskie zera.*
3. *Wartości α i β są wzajemnie powiązane i zmniejszenie jednej z nich powoduje zwiększenie drugiej.*
4. *Testy konstruuje się tak, aby przy ustalonym α, minimalizować β.*

Błędy testowania hipotez. V

Przykład 6 ([1, Przykład 9.1])

Cecha X ma w zbiorowości generalnej rozkład $\mathcal{N}(m, 4)$. Na podstawie czteroelementowej próby należy zweryfikować hipotezę H_0, że wartość przeciętna jest równa 10, wobec hipotezy H_1, że wartość przeciętna wynosi 15. Przyjmujemy taką regułę postępowania, że hipotezę H_0 odrzucamy, gdy $\bar{X}$ obliczone na podstawie próby przyjmuje wartość większą od 13. Należy obliczyć prawdopodobieństwa popełnienia błędu I-go i II-go rodzaju oraz podamać ilustrację graficzną obliczonych prawdopodobieństw.

$$
\begin{aligned}
H_0 : &\qquad m = 10, \\
H_1 : &\qquad m = 15.
\end{aligned}
$$

Moc testu statystycznego. I

Definicja 10

Mocą testu, oznaczaną $M(\Lambda)$, nazywamy prawdopodobieństwo odrzucenia fałszywej hipotezy zerowej i przyjęcia w to miejsce prawdziwej hipotezy alternatywnej było największe, czyli prawdopodobieństwo podjęcia słusznej decyzji podczas weryfikowania hipotezy

$$M(\Lambda) = P_\theta(\{T_n \in \Lambda\}) \quad \theta \in \Theta_1.$$

Definicja 11

Testem najmocniejszym nazywamy taki, że jego moc $M(\Lambda)$ jest największa tzn. przy ustalonym α prawdopodobieństwo odrzucenia fałszywej hipotezy zerowej i przyjęcia w to miejsce prawdziwej hipotezy alternatywnej było największe.

Moc testu statystycznego. II

Uwaga 27

1. *Zauważmy, że* $M(\Lambda) = 1 - \beta(\theta)$.
2. *Oznacza to, że test jest tym mocniejszy, im mniejsze jest prawdopodobieństwo popełnienia błędu II-go rodzaju.*

Test zgodny

Definicja 12

Test statystyczny jest testem zgodnym, jeśli wraz ze wzrostem liczebności próby moc testu osiąga wartości coraz bliższe 1, tzn. gdy zachodzi relacja

$$\lim_{n\to\infty} M(\Lambda) = 1.$$

Bibliografia

[1] St. Ostasiewicz, Z. Rusnak i U. Siedlecka. *Statystyka. Elementy teorii i Zadania*. Wyd. 5 poprawione. Wrocław: Wyd. Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, 2003.

[2] Hoekstra Rink i in. "Robust misinterpretation of confidence intervals". W: *Psychonomic Bulletin & Review* 21.5 (2014), s. 1157–1164. ISSN: 1069-9384 (Print) 1531-5320 (Online). DOI: 10.3758/s13423-013-0572-3.

[3] Niemiro Wojciech. *Statystyka I*. 2014.