

Statystyka i rachunek prawdopodobieństwa

Wykład 10: Estymacja przedziałowa

Mariusz Tarnopolski

Instytut Astronomii UMK

Statystyka ©2026



Estymacja przedziałowa

Estymacja przedziałowa polega na poszukiwaniu takiego przedziału (Q_1, Q_2) wewnątrz którego wartość parametru Q populacji leży z zadany­m prawdopodobieństwem. Niech (X_1, X_2, \dots, X_n) będzie próbą prostą. Wtedy

$$P [Q_1(X_1, \dots, X_n) < Q < Q_2(X_1, \dots, X_n)] = 1 - \alpha$$

gdzie:

- (Q_1, Q_2) nazywamy **dwustronnym przedziałem ufności**
- $1 - \alpha$ to **prawdopodobieństwo ufności**, zaś
- α to **poziom istotności** (najczęściej przyjmuje się $\alpha = 0.01$ lub ew. $\alpha = 0.05$)
- przedział ufności jest **losowy** (inny dla każdej próby)
- częstość (prawdopodobieństwo) błędnych oszacowań wynosi α
- zwiększanie prawdopodobieństwa ufności $1 - \alpha$ zwiększa przedział ufności (Q_1, Q_2)
- powiemy, że (Q_1, Q_2) z prawdopodobieństwem $1 - \alpha$ pokrywa nieznaną wartość parametru Q

Estymacja przedziałowa

Estymacja przedziałowa polega na poszukiwaniu takiego przedziału (Q_1, Q_2) wewnątrz którego wartość parametru Q populacji leży z zadany­m prawdopodobieństwem. Niech (X_1, X_2, \dots, X_n) będzie próbą prostą. Wtedy

$$P [Q_1(X_1, \dots, X_n) < Q < Q_2(X_1, \dots, X_n)] = 1 - \alpha$$

gdzie:

- (Q_1, Q_2) nazywamy **dwustronnym przedziałem ufności**
- $1 - \alpha$ to **prawdopodobieństwo ufności**, zaś
- α to **poziom istotności** (najczęściej przyjmuje się $\alpha = 0.01$ lub ew. $\alpha = 0.05$)
- przedział ufności jest **losowy** (inny dla każdej próby)
- częstość (prawdopodobieństwo) błędnych oszacowań wynosi α
- zwiększanie prawdopodobieństwa ufności $1 - \alpha$ zwiększa przedział ufności (Q_1, Q_2)
- powiemy, że (Q_1, Q_2) z prawdopodobieństwem $1 - \alpha$ pokrywa nieznaną wartość parametru Q

Estymacja przedziałowa polega na poszukiwaniu takiego przedziału (Q_1, Q_2) wewnątrz którego wartość parametru Q populacji leży z zadany­m prawdopodobieństwem. Niech (X_1, X_2, \dots, X_n) będzie próbą prostą. Wtedy

$$P [Q_1(X_1, \dots, X_n) < Q < Q_2(X_1, \dots, X_n)] = 1 - \alpha$$

gdzie:

- (Q_1, Q_2) nazywamy **dwustronnym przedziałem ufności**
- $1 - \alpha$ to **prawdopodobieństwo ufności**, zaś
- α to **poziom istotności** (najczęściej przyjmuje się $\alpha = 0.01$ lub ew. $\alpha = 0.05$)
- przedział ufności jest losowy (inny dla każdej próby)
- częstość (prawdopodobieństwo) błędnych oszacowań wynosi α
- zwiększanie prawdopodobieństwa ufności $1 - \alpha$ zwiększa przedział ufności (Q_1, Q_2)
- powiemy, że (Q_1, Q_2) z prawdopodobieństwem $1 - \alpha$ pokrywa nieznaną wartość parametru Q

Estymacja przedziałowa polega na poszukiwaniu takiego przedziału (Q_1, Q_2) wewnątrz którego wartość parametru Q populacji leży z zadany­m prawdopodobieństwem. Niech (X_1, X_2, \dots, X_n) będzie próbą prostą. Wtedy

$$P [Q_1(X_1, \dots, X_n) < Q < Q_2(X_1, \dots, X_n)] = 1 - \alpha$$

gdzie:

- (Q_1, Q_2) nazywamy **dwustronnym przedziałem ufności**
- $1 - \alpha$ to **prawdopodobieństwo ufności**, zaś
- α to **poziom istotności** (najczęściej przyjmuje się $\alpha = 0.01$ lub ew. $\alpha = 0.05$)
- przedział ufności jest **losowy** (inny dla każdej próby)
- częstość (prawdopodobieństwo) błędnych oszacowań wynosi α
- zwiększanie prawdopodobieństwa ufności $1 - \alpha$ zwiększa przedział ufności (Q_1, Q_2)
- powiemy, że (Q_1, Q_2) z prawdopodobieństwem $1 - \alpha$ pokrywa nieznaną wartość parametru Q

Estymacja przedziałowa polega na poszukiwaniu takiego przedziału (Q_1, Q_2) wewnątrz którego wartość parametru Q populacji leży z zadany­m prawdopodobieństwem. Niech (X_1, X_2, \dots, X_n) będzie próbą prostą. Wtedy

$$P [Q_1(X_1, \dots, X_n) < Q < Q_2(X_1, \dots, X_n)] = 1 - \alpha$$

gdzie:

- (Q_1, Q_2) nazywamy **dwustronnym przedziałem ufności**
- $1 - \alpha$ to **prawdopodobieństwo ufności**, zaś
- α to **poziom istotności** (najczęściej przyjmuje się $\alpha = 0.01$ lub ew. $\alpha = 0.05$)
- przedział ufności jest **losowy** (inny dla każdej próby)
- częstość (prawdopodobieństwo) błędnych oszacowań wynosi α
- zwiększanie prawdopodobieństwa ufności $1 - \alpha$ zwiększa przedział ufności (Q_1, Q_2)
- powiemy, że (Q_1, Q_2) z prawdopodobieństwem $1 - \alpha$ pokrywa nieznaną wartość parametru Q

Estymacja przedziałowa polega na poszukiwaniu takiego przedziału (Q_1, Q_2) wewnątrz którego wartość parametru Q populacji leży z zadany­m prawdopodobieństwem. Niech (X_1, X_2, \dots, X_n) będzie próbą prostą. Wtedy

$$P [Q_1(X_1, \dots, X_n) < Q < Q_2(X_1, \dots, X_n)] = 1 - \alpha$$

gdzie:

- (Q_1, Q_2) nazywamy **dwustronnym przedziałem ufności**
- $1 - \alpha$ to **prawdopodobieństwo ufności**, zaś
- α to **poziom istotności** (najczęściej przyjmuje się $\alpha = 0.01$ lub ew. $\alpha = 0.05$)
- przedział ufności jest **losowy** (inny dla każdej próby)
- częstość (prawdopodobieństwo) błędnych oszacowań wynosi α
- zwiększanie prawdopodobieństwa ufności $1 - \alpha$ zwiększa przedział ufności (Q_1, Q_2)
- powiemy, że (Q_1, Q_2) z prawdopodobieństwem $1 - \alpha$ pokrywa nieznaną wartość parametru Q

Estymacja przedziałowa polega na poszukiwaniu takiego przedziału (Q_1, Q_2) wewnątrz którego wartość parametru Q populacji leży z zadany­m prawdopodobieństwem. Niech (X_1, X_2, \dots, X_n) będzie próbą prostą. Wtedy

$$P [Q_1(X_1, \dots, X_n) < Q < Q_2(X_1, \dots, X_n)] = 1 - \alpha$$

gdzie:

- (Q_1, Q_2) nazywamy **dwustronnym przedziałem ufności**
- $1 - \alpha$ to **prawdopodobieństwo ufności**, zaś
- α to **poziom istotności** (najczęściej przyjmuje się $\alpha = 0.01$ lub ew. $\alpha = 0.05$)
- przedział ufności jest **losowy** (inny dla każdej próby)
- częstość (prawdopodobieństwo) błędnych oszacowań wynosi α
- zwiększanie prawdopodobieństwa ufności $1 - \alpha$ zwiększa przedział ufności (Q_1, Q_2)
- powiemy, że (Q_1, Q_2) z prawdopodobieństwem $1 - \alpha$ pokrywa nieznaną wartość parametru Q

Przedział ufności w 4 krokach:

- 1 określić badany parametr populacji
- 2 znaleźć rozkład prawd. jego estymatora w n -elementowej próbie prostej
- 3 wybrać poziom ufności
- 4 znaleźć granice przedziału ufności

Model 1. Przedział ufności dla nieznaney wartości przeciętnej μ populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ oraz σ jest **znane**.

- Estymatorem μ jest statystyka $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- \bar{X} ma rozkład $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
- Dla ułatwienia weźmy statystykę $U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ (otrzymaną w wyniku standaryzacji) o rozkładzie $\mathcal{N}(0, 1)$
- Interesuje nas

$$P(u_1 < U < u_2) = 1 - \alpha$$

czyli

$$P\left(u_1 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < u_2\right) = 1 - \alpha$$

Model 1. Przedział ufności dla nieznaney wartości przeciętnej μ populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ oraz σ jest **znane**.

- Estymatorem μ jest statystyka $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- \bar{X} ma rozkład $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
- Dla ułatwienia weźmy statystykę $U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ (otrzymaną w wyniku standaryzacji) o rozkładzie $\mathcal{N}(0, 1)$
- Interesuje nas

$$P(u_1 < U < u_2) = 1 - \alpha$$

czyli

$$P\left(u_1 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < u_2\right) = 1 - \alpha$$

Model 1. Przedział ufności dla nieznaney wartości przeciętnej μ populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ oraz σ jest **znane**.

- Estymatorem μ jest statystyka $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- \bar{X} ma rozkład $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
- Dla ułatwienia weźmy statystykę $U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ (otrzymaną w wyniku standaryzacji) o rozkładzie $\mathcal{N}(0, 1)$
- Interesuje nas

$$P(u_1 < U < u_2) = 1 - \alpha$$

czyli

$$P\left(u_1 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < u_2\right) = 1 - \alpha$$

Model 1. Przedział ufności dla nieznannej wartości przeciętnej μ populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ oraz σ jest **znane**.

- Estymatorem μ jest statystyka $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- \bar{X} ma rozkład $\mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$
- Dla ułatwienia weźmy statystykę $U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ (otrzymaną w wyniku standaryzacji) o rozkładzie $\mathcal{N}(0, 1)$
- Interesuje nas

$$P(u_1 < U < u_2) = 1 - \alpha$$

czyli

$$P\left(u_1 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < u_2\right) = 1 - \alpha$$

- Załóżmy symetryczność przedziałów:

$$P(-\infty < U < u_1) = P(u_2 < U < +\infty) = \frac{\alpha}{2}$$

zatem $u_1 = u \left(\frac{\alpha}{2} \right)$, $u_2 = u \left(1 - \frac{\alpha}{2} \right)$, $u_1 = -u$.

- Zatem

$$P \left(-u < \frac{\bar{X} - \mu}{\sigma} \sqrt{n} < u \right) = 1 - \alpha$$

lub też

$$P \left(-u \frac{\sigma}{\sqrt{n}} + \bar{X} < \mu < u \frac{\sigma}{\sqrt{n}} + \bar{X} \right) = 1 - \alpha$$

- Weźmy $1 - \alpha = 0.95$ odpowiadające 95% przedziałowi ufności. Wartości u dla danego poziomu ufności $1 - \alpha$ (\equiv poziomowi istotności α) odczytujemy z tablic dystrybuanty standardowego rozkładu normalnego [tutaj $u = u \left(1 - \frac{\alpha}{2} \right) = 1.96$].
- Przedział ufności to

$$\left(-u \frac{\sigma}{\sqrt{n}} + \bar{X}, u \frac{\sigma}{\sqrt{n}} + \bar{X} \right)$$

- Załóżmy symetryczność przedziałów:

$$P(-\infty < U < u_1) = P(u_2 < U < +\infty) = \frac{\alpha}{2}$$

zatem $u_1 = u \left(\frac{\alpha}{2} \right)$, $u_2 = u \left(1 - \frac{\alpha}{2} \right)$, $u_1 = -u$.

- Zatem

$$P \left(-u < \frac{\bar{X} - \mu}{\sigma} \sqrt{n} < u \right) = 1 - \alpha$$

lub też

$$P \left(-u \frac{\sigma}{\sqrt{n}} + \bar{X} < \mu < u \frac{\sigma}{\sqrt{n}} + \bar{X} \right) = 1 - \alpha$$

- Weźmy $1 - \alpha = 0.95$ odpowiadające 95% przedziałowi ufności. Wartości u dla danego poziomu ufności $1 - \alpha$ (\equiv poziomowi istotności α) odczytujemy z tablic dystrybuanty standardowego rozkładu normalnego [tutaj $u = u \left(1 - \frac{\alpha}{2} \right) = 1.96$].
- Przedział ufności to

$$\left(-u \frac{\sigma}{\sqrt{n}} + \bar{X}, u \frac{\sigma}{\sqrt{n}} + \bar{X} \right)$$

- Załóżmy symetryczność przedziałów:

$$P(-\infty < U < u_1) = P(u_2 < U < +\infty) = \frac{\alpha}{2}$$

zatem $u_1 = u \left(\frac{\alpha}{2} \right)$, $u_2 = u \left(1 - \frac{\alpha}{2} \right)$, $u_1 = -u$.

- Zatem

$$P \left(-u < \frac{\bar{X} - \mu}{\sigma} \sqrt{n} < u \right) = 1 - \alpha$$

lub też

$$P \left(-u \frac{\sigma}{\sqrt{n}} + \bar{X} < \mu < u \frac{\sigma}{\sqrt{n}} + \bar{X} \right) = 1 - \alpha$$

- Weźmy $1 - \alpha = 0.95$ odpowiadające 95% przedziałowi ufności. Wartości u dla danego poziomu ufności $1 - \alpha$ (\equiv poziomowi istotności α) odczytujemy z tablic dystrybuanty standardowego rozkładu normalnego [tutaj $u = u \left(1 - \frac{\alpha}{2} \right) = 1.96$].
- Przedział ufności to

$$\left(-u \frac{\sigma}{\sqrt{n}} + \bar{X}, u \frac{\sigma}{\sqrt{n}} + \bar{X} \right)$$

- Załóżmy symetryczność przedziałów:

$$P(-\infty < U < u_1) = P(u_2 < U < +\infty) = \frac{\alpha}{2}$$

zatem $u_1 = u \left(\frac{\alpha}{2} \right)$, $u_2 = u \left(1 - \frac{\alpha}{2} \right)$, $u_1 = -u$.

- Zatem

$$P \left(-u < \frac{\bar{X} - \mu}{\sigma} \sqrt{n} < u \right) = 1 - \alpha$$

lub też

$$P \left(-u \frac{\sigma}{\sqrt{n}} + \bar{X} < \mu < u \frac{\sigma}{\sqrt{n}} + \bar{X} \right) = 1 - \alpha$$

- Weźmy $1 - \alpha = 0.95$ odpowiadające 95% przedziałowi ufności. Wartości u dla danego poziomu ufności $1 - \alpha$ (\equiv poziomowi istotności α) odczytujemy z tablic dystrybuanty standardowego rozkładu normalnego [tutaj $u = u \left(1 - \frac{\alpha}{2} \right) = 1.96$].
- Przedział ufności to

$$\left(-u \frac{\sigma}{\sqrt{n}} + \bar{X}, u \frac{\sigma}{\sqrt{n}} + \bar{X} \right)$$

Model 1. wydaje się być zbyt idealny — skąd niby znamy σ populacji ale nie znamy μ ?

- 1 Pomiar wielkości fizycznej przyrządem, którego dokładność jest znana z kalibracji — znany jest zatem rozrzut, estymujemy konkretne pomiary
- 2 Ogromna ilość wcześniejszych danych (agregacja, dane historyczne, z innych eksperymentów itp.) — σ traktuje się jako znane
- 3 Modele teoretyczne, symulacje (Monte Carlo)

W praktyce jednak istotnie częściej stosuje się inne modele 🍷

Model 2. Przedział ufności dla nieznannej wartości przeciętnej μ populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ oraz σ jest **nieznane**.

- σ estymujemy z próbki poprzez $S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ lub

$$S_2^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \text{ Wtedy statystyka}$$

$$T = \frac{\bar{X} - \mu}{S_1/\sqrt{n-1}} = \frac{\bar{X} - \mu}{S_2/\sqrt{n}}$$

ma rozkład t -Studenta o $n - 1$ stopniach swobody

- PDF rozkładu t -Studenta o n stopniach swobody ma postać

$$f(t; n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

gdzie Γ jest funkcją gamma (ciągłym uogólnieniem silni).

Model 2. Przedział ufności dla nieznannej wartości przeciętnej μ populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ oraz σ jest **nieznane**.

- σ estymujemy z próbki poprzez $S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ lub

$$S_2^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2. \text{ Wtedy statystyka}$$

$$T = \frac{\bar{X} - \mu}{S_1/\sqrt{n-1}} = \frac{\bar{X} - \mu}{S_2/\sqrt{n}}$$

ma rozkład t -Studenta o $n - 1$ stopniach swobody

- PDF rozkładu t -Studenta o n stopniach swobody ma postać

$$f(t; n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

gdzie Γ jest funkcją gamma (ciągłym uogólnieniem silni).

Stopnie swobody to liczba niezależnych informacji w próbce, które pozostają po nałożeniu ograniczeń:

Stopnie swobody to liczba niezależnych informacji w próbce, które pozostają po nałożeniu ograniczeń:

- Mamy próbę prostą (X_1, \dots, X_n) — na starcie jest n niezależnych obserwacji.
- Jeśli nie znamy żadnych parametrów, wszystkie n obserwacji są swobodne.
- Jeśli estymujemy parametr z danych, to narzucamy jedno ograniczenie.
- Każdy estymowany parametr „zużywa” jeden stopień swobody.
- Np. jeśli estymujemy średnią \bar{X} to „tracimy” 1 stopień swobody \Rightarrow zostaje $n - 1$ stopni swobody

Stopnie swobody to liczba niezależnych informacji w próbce, które pozostają po nałożeniu ograniczeń:

- Mamy próbę prostą (X_1, \dots, X_n) — na starcie jest n niezależnych obserwacji.
- Jeśli nie znamy żadnych parametrów, wszystkie n obserwacji są swobodne.
- Jeśli estymujemy parametr z danych, to narzucamy jedno ograniczenie.
- Każdy estymowany parametr „zużywa” jeden stopień swobody.
- Np. jeśli estymujemy średnią \bar{X} to „tracimy” 1 stopień swobody \Rightarrow zostaje $n - 1$ stopni swobody

Stopnie swobody to liczba niezależnych informacji w próbce, które pozostają po nałożeniu ograniczeń:

- Mamy próbę prostą (X_1, \dots, X_n) — na starcie jest n niezależnych obserwacji.
- Jeśli nie znamy żadnych parametrów, wszystkie n obserwacji są swobodne.
- Jeśli estymujemy parametr z danych, to narzucamy jedno ograniczenie.
- Każdy estymowany parametr „zużywa” jeden stopień swobody.
- Np. jeśli estymujemy średnią \bar{X} to „tracimy” 1 stopień swobody \Rightarrow zostaje $n - 1$ stopni swobody

Stopnie swobody to liczba niezależnych informacji w próbce, które pozostają po nałożeniu ograniczeń:

- Mamy próbę prostą (X_1, \dots, X_n) — na starcie jest n niezależnych obserwacji.
- Jeśli nie znamy żadnych parametrów, wszystkie n obserwacji są swobodne.
- Jeśli estymujemy parametr z danych, to narzucamy jedno ograniczenie.
- Każdy estymowany parametr „zużywa” jeden stopień swobody.
- Np. jeśli estymujemy średnią \bar{X} to „tracimy” 1 stopień swobody \Rightarrow zostaje $n - 1$ stopni swobody

Stopnie swobody to liczba niezależnych informacji w próbce, które pozostają po nałożeniu ograniczeń:

- Mamy próbę prostą (X_1, \dots, X_n) — na starcie jest n niezależnych obserwacji.
- Jeśli nie znamy żadnych parametrów, wszystkie n obserwacji są swobodne.
- Jeśli estymujemy parametr z danych, to narzucamy jedno ograniczenie.
- Każdy estymowany parametr „zużywa” jeden stopień swobody.
- Np. jeśli estymujemy średnią \bar{X} to „tracimy” 1 stopień swobody \Rightarrow zostaje $n - 1$ stopni swobody

Stopnie swobody — związek z estymacją

- Rozważmy odchylenia od średniej:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \Rightarrow \sum_{i=1}^n (X_i - \bar{X}) = 0$$

tj. jeśli przesuwamy „układ odniesienia” do średniej, to odchylenia dodatnie i ujemne muszą się znosić.

- Oznacza to, że odchylenia nie są niezależne — jeśli znamy $n - 1$ odchyłeń $X_i - \bar{X}$, to n -te jest wyznaczone jednoznacznie.
- Warunek $\sum_{i=1}^n (X_i - \bar{X}) = 0$ ogranicza przestrzeń, w której „żyją” odchylenia (jako składowe wektora) do wymiaru $n - 1$; suma kwadratów odchyłeń jest normą tego wektora.
- Dlatego estymator wariancji:

$$S_2^2 = \frac{1}{n - 1} \sum_{i=1}^n (X_i - \bar{X})^2$$

ma $n - 1$ stopni swobody

Stopnie swobody — związek z estymacją

- Rozważmy odchylenia od średniej:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \Rightarrow \sum_{i=1}^n (X_i - \bar{X}) = 0$$

tj. jeśli przesuwamy „układ odniesienia” do średniej, to odchylenia dodatnie i ujemne muszą się znosić.

- Oznacza to, że odchylenia nie są niezależne — jeśli znamy $n - 1$ odchyleń $X_i - \bar{X}$, to n -te jest wyznaczone jednoznacznie.
- Warunek $\sum_{i=1}^n (X_i - \bar{X}) = 0$ ogranicza przestrzeń, w której „żyją” odchylenia (jako składowe wektora) do wymiaru $n - 1$; suma kwadratów odchyleń jest normą tego wektora.
- Dlatego estymator wariancji:

$$S_2^2 = \frac{1}{n - 1} \sum_{i=1}^n (X_i - \bar{X})^2$$

ma $n - 1$ stopni swobody

Stopnie swobody — związek z estymacją

- Rozważmy odchylenia od średniej:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \Rightarrow \sum_{i=1}^n (X_i - \bar{X}) = 0$$

tj. jeśli przesuwamy „układ odniesienia” do średniej, to odchylenia dodatnie i ujemne muszą się znosić.

- Oznacza to, że odchylenia nie są niezależne — jeśli znamy $n - 1$ odchyleń $X_i - \bar{X}$, to n -te jest wyznaczone jednoznacznie.
- Warunek $\sum_{i=1}^n (X_i - \bar{X}) = 0$ ogranicza przestrzeń, w której „żyją” odchylenia (jako składowe wektora) do wymiaru $n - 1$; suma kwadratów odchyleń jest normą tego wektora.
- Dlatego estymator wariancji:

$$S_2^2 = \frac{1}{n - 1} \sum_{i=1}^n (X_i - \bar{X})^2$$

ma $n - 1$ stopni swobody

Stopnie swobody — związek z estymacją

- Rozważmy odchylenia od średniej:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \Rightarrow \sum_{i=1}^n (X_i - \bar{X}) = 0$$

tj. jeśli przesuwamy „układ odniesienia” do średniej, to odchylenia dodatnie i ujemne muszą się znosić.

- Oznacza to, że odchylenia nie są niezależne — jeśli znamy $n - 1$ odchyłeń $X_i - \bar{X}$, to n -te jest wyznaczone jednoznacznie.
- Warunek $\sum_{i=1}^n (X_i - \bar{X}) = 0$ ogranicza przestrzeń, w której „żyją” odchylenia (jako składowe wektora) do wymiaru $n - 1$; suma kwadratów odchyłeń jest normą tego wektora.
- Dlatego estymator wariancji:

$$S_2^2 = \frac{1}{n - 1} \sum_{i=1}^n (X_i - \bar{X})^2$$

ma $n - 1$ stopni swobody

Estymujemy średnią \bar{X} oraz wariancję S_2^2 z próbki n -elementowej:

- Suma odchyłeń $\sum_i (X_i - \bar{X}) = 0$
- To ograniczenie „zużywa” 1 stopień swobody
- Pozostałe $n - 1$ niezależne odchylenia wchodzi w sumę kwadratów odchyłeń w wyrażeniu na S_2
- Dlatego statystyka $T = \frac{\bar{X} - \mu}{S_2 / \sqrt{n}}$ ma rozkład (t -Studenta) z $n - 1$ stopniami swobody

- Rozkład t -Studenta jest rozkładem zmiennej losowej postaci $\frac{U}{\sqrt{Z}}\sqrt{n}$:
 - U jest zmienną losową o rozkładzie $\mathcal{N}(0, 1)$
 - Z jest zmienną losową o rozkładzie χ^2 o n stopniach swobody
 - U i Z są niezależne
 - dla $n \rightarrow \infty$, rozkład t -Studenta sprowadza się do standardowego rozkładu Gaussa, zaś
 - dla $n = 1$ otrzymujemy rozkład Cauchy'ego
 - kurtoza dla $n > 4$ wynosi $K = \frac{6}{n-4}$ — czyli rozkład t -Studenta ma cięższe ogony niż rozkład Gaussa (dla $2 < n \leq 4$ zachodzi $K = \infty$; dla $n \leq 2$ kurtoza nie istnieje)

- Rozkład t -Studenta jest rozkładem zmiennej losowej postaci $\frac{U}{\sqrt{Z}}\sqrt{n}$:
 - U jest zmienną losową o rozkładzie $\mathcal{N}(0, 1)$
 - Z jest zmienną losową o rozkładzie χ^2 o n stopniach swobody
 - U i Z są niezależne
 - dla $n \rightarrow \infty$, rozkład t -Studenta sprowadza się do standardowego rozkładu Gaussa, zaś
 - dla $n = 1$ otrzymujemy rozkład Cauchy'ego
 - kurtoza dla $n > 4$ wynosi $K = \frac{6}{n-4}$ — czyli rozkład t -Studenta ma cięższe ogony niż rozkład Gaussa (dla $2 < n \leq 4$ zachodzi $K = \infty$; dla $n \leq 2$ kurtoza nie istnieje)

- Rozkład t -Studenta jest rozkładem zmiennej losowej postaci $\frac{U}{\sqrt{Z}}\sqrt{n}$:
 - U jest zmienną losową o rozkładzie $\mathcal{N}(0, 1)$
 - Z jest zmienną losową o rozkładzie χ^2 o n stopniach swobody
 - U i Z są niezależne
 - dla $n \rightarrow \infty$, rozkład t -Studenta sprowadza się do standardowego rozkładu Gaussa, zaś
 - dla $n = 1$ otrzymujemy rozkład Cauchy'ego
 - kurtoza dla $n > 4$ wynosi $K = \frac{6}{n-4}$ — czyli rozkład t -Studenta ma cięższe ogony niż rozkład Gaussa (dla $2 < n \leq 4$ zachodzi $K = \infty$; dla $n \leq 2$ kurtoza nie istnieje)

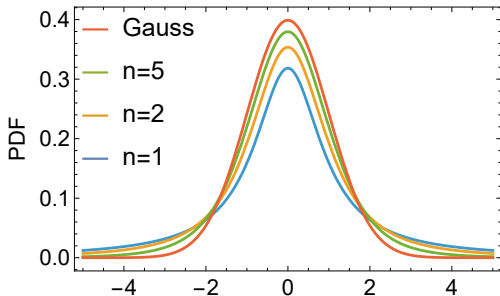
- Rozkład t -Studenta jest rozkładem zmiennej losowej postaci $\frac{U}{\sqrt{Z}}\sqrt{n}$:
 - U jest zmienną losową o rozkładzie $\mathcal{N}(0, 1)$
 - Z jest zmienną losową o rozkładzie χ^2 o n stopniach swobody
 - U i Z są niezależne
 - dla $n \rightarrow \infty$, rozkład t -Studenta sprowadza się do standardowego rozkładu Gaussa, zaś
 - dla $n = 1$ otrzymujemy rozkład Cauchy'ego
 - kurtoza dla $n > 4$ wynosi $K = \frac{6}{n-4}$ — czyli rozkład t -Studenta ma cięższe ogony niż rozkład Gaussa (dla $2 < n \leq 4$ zachodzi $K = \infty$; dla $n \leq 2$ kurtoza nie istnieje)

- Rozkład t -Studenta jest rozkładem zmiennej losowej postaci $\frac{U}{\sqrt{Z}}\sqrt{n}$:
 - U jest zmienną losową o rozkładzie $\mathcal{N}(0, 1)$
 - Z jest zmienną losową o rozkładzie χ^2 o n stopniach swobody
 - U i Z są niezależne
 - dla $n \rightarrow \infty$, rozkład t -Studenta sprowadza się do standardowego rozkładu Gaussa, zaś
 - dla $n = 1$ otrzymujemy rozkład Cauchy'ego
 - kurtoza dla $n > 4$ wynosi $K = \frac{6}{n-4}$ — czyli rozkład t -Studenta ma cięższe ogony niż rozkład Gaussa (dla $2 < n \leq 4$ zachodzi $K = \infty$; dla $n \leq 2$ kurtoza nie istnieje)

- Rozkład t -Studenta jest rozkładem zmiennej losowej postaci $\frac{U}{\sqrt{Z}}\sqrt{n}$:
 - U jest zmienną losową o rozkładzie $\mathcal{N}(0, 1)$
 - Z jest zmienną losową o rozkładzie χ^2 o n stopniach swobody
 - U i Z są niezależne
 - dla $n \rightarrow \infty$, rozkład t -Studenta sprowadza się do standardowego rozkładu Gaussa, zaś
 - dla $n = 1$ otrzymujemy rozkład Cauchy'ego
 - kurtoza dla $n > 4$ wynosi $K = \frac{6}{n-4}$ — czyli rozkład t -Studenta ma cięższe ogony niż rozkład Gaussa (dla $2 < n \leq 4$ zachodzi $K = \infty$; dla $n \leq 2$ kurtoza nie istnieje)

- Rozkład t -Studenta jest rozkładem zmiennej losowej postaci $\frac{U}{\sqrt{Z}}\sqrt{n}$:
 - U jest zmienną losową o rozkładzie $\mathcal{N}(0, 1)$
 - Z jest zmienną losową o rozkładzie χ^2 o n stopniach swobody
 - U i Z są niezależne
 - dla $n \rightarrow \infty$, rozkład t -Studenta sprowadza się do standardowego rozkładu Gaussa, zaś
 - dla $n = 1$ otrzymujemy rozkład Cauchy'ego
 - kurtoza dla $n > 4$ wynosi $K = \frac{6}{n-4}$ — czyli rozkład t -Studenta ma cięższe ogony niż rozkład Gaussa (dla $2 < n \leq 4$ zachodzi $K = \infty$; dla $n \leq 2$ kurtoza nie istnieje)

- Rozkład t -Studenta jest rozkładem zmiennej losowej postaci $\frac{U}{\sqrt{Z}}\sqrt{n}$:
 - U jest zmienną losową o rozkładzie $\mathcal{N}(0, 1)$
 - Z jest zmienną losową o rozkładzie χ^2 o n stopniach swobody
 - U i Z są niezależne
 - dla $n \rightarrow \infty$, rozkład t -Studenta sprowadza się do standardowego rozkładu Gaussa, zaś
 - dla $n = 1$ otrzymujemy rozkład Cauchy'ego
 - kurtoza dla $n > 4$ wynosi $K = \frac{6}{n-4}$ — czyli rozkład t -Studenta ma cięższe ogony niż rozkład Gaussa (dla $2 < n \leq 4$ zachodzi $K = \infty$; dla $n \leq 2$ kurtoza nie istnieje)



Przedział ufności konstruuje się w oparciu o

$$P\left(-t < \frac{\bar{X} - \mu}{S_2/\sqrt{n}} < t\right) = 1 - \alpha$$

czyli

$$P\left(-t \frac{S_2}{\sqrt{n}} + \bar{X} < \mu < t \frac{S_2}{\sqrt{n}} + \bar{X}\right) = 1 - \alpha$$

Przedział ufności konstruuje się w oparciu o

$$P\left(-t < \frac{\bar{X} - \mu}{S_2/\sqrt{n}} < t\right) = 1 - \alpha$$

czyli

$$P\left(-t \frac{S_2}{\sqrt{n}} + \bar{X} < \mu < t \frac{S_2}{\sqrt{n}} + \bar{X}\right) = 1 - \alpha$$

Przykład. Jaka jest średnia waga gumy do żucia? Po zważeniu 10 sztuk otrzymano $\bar{X} = 14$ g, $S^2 = 4$ g².

- Statystyka testowa *t*-Studenta
- Wybieramy poziom ufności 95%, tj. $\alpha = 0.05$
- Z tablic statystycznych dla 9 stopni swobody odczytujemy kwantyl rzędu $1 - \frac{\alpha}{2} = 0.975$ jako $t = t(0.975) = 2.262$
- 95% symetryczny przedział ufności to

$$\left(-2.262 \frac{2 \text{ g}}{\sqrt{10}} + 14 \text{ g}, 2.262 \frac{2 \text{ g}}{\sqrt{10}} + 14 \text{ g}\right) = (12.57 \text{ g}, 15.43 \text{ g})$$

Przedział ufności konstruuje się w oparciu o

$$P\left(-t < \frac{\bar{X} - \mu}{S_2/\sqrt{n}} < t\right) = 1 - \alpha$$

czyli

$$P\left(-t \frac{S_2}{\sqrt{n}} + \bar{X} < \mu < t \frac{S_2}{\sqrt{n}} + \bar{X}\right) = 1 - \alpha$$

Przykład. Jaka jest średnia waga gumy do żucia? Po zważeniu 10 sztuk otrzymano $\bar{X} = 14$ g, $S^2 = 4$ g².

- Statystyka testowa t -Studenta
- Wybieramy poziom ufności 95%, tj. $\alpha = 0.05$
- Z tablic statystycznych dla 9 stopni swobody odczytujemy kwantyl rzędu $1 - \frac{\alpha}{2} = 0.975$ jako $t = t(0.975) = 2.262$
- 95% symetryczny przedział ufności to

$$\left(-2.262 \frac{2 \text{ g}}{\sqrt{10}} + 14 \text{ g}, 2.262 \frac{2 \text{ g}}{\sqrt{10}} + 14 \text{ g}\right) = (12.57 \text{ g}, 15.43 \text{ g})$$

Przedział ufności konstruuje się w oparciu o

$$P\left(-t < \frac{\bar{X} - \mu}{S_2/\sqrt{n}} < t\right) = 1 - \alpha$$

czyli

$$P\left(-t \frac{S_2}{\sqrt{n}} + \bar{X} < \mu < t \frac{S_2}{\sqrt{n}} + \bar{X}\right) = 1 - \alpha$$

Przykład. Jaka jest średnia waga gumy do żucia? Po zważeniu 10 sztuk otrzymano $\bar{X} = 14$ g, $S^2 = 4$ g².

- Statystyka testowa t -Studenta
- Wybieramy poziom ufności 95%, tj. $\alpha = 0.05$
- Z tablic statystycznych dla 9 stopni swobody odczytujemy kwantyl rzędu $1 - \frac{\alpha}{2} = 0.975$ jako $t = t(0.975) = 2.262$
- 95% symetryczny przedział ufności to

$$\left(-2.262 \frac{2 \text{ g}}{\sqrt{10}} + 14 \text{ g}, 2.262 \frac{2 \text{ g}}{\sqrt{10}} + 14 \text{ g}\right) = (12.57 \text{ g}, 15.43 \text{ g})$$

Przedział ufności konstruuje się w oparciu o

$$P\left(-t < \frac{\bar{X} - \mu}{S_2/\sqrt{n}} < t\right) = 1 - \alpha$$

czyli

$$P\left(-t \frac{S_2}{\sqrt{n}} + \bar{X} < \mu < t \frac{S_2}{\sqrt{n}} + \bar{X}\right) = 1 - \alpha$$

Przykład. Jaka jest średnia waga gumy do żucia? Po zważeniu 10 sztuk otrzymano $\bar{X} = 14$ g, $S^2 = 4$ g².

- Statystyka testowa t -Studenta
- Wybieramy poziom ufności 95%, tj. $\alpha = 0.05$
- Z tablic statystycznych dla 9 stopni swobody odczytujemy kwantyl rzędu $1 - \frac{\alpha}{2} = 0.975$ jako $t = t(0.975) = 2.262$
- 95% symetryczny przedział ufności to

$$\left(-2.262 \frac{2 \text{ g}}{\sqrt{10}} + 14 \text{ g}, 2.262 \frac{2 \text{ g}}{\sqrt{10}} + 14 \text{ g}\right) = (12.57 \text{ g}, 15.43 \text{ g})$$

Model 3. Przedział ufności dla nieznannej wartości przeciętnej μ populacji, w której nieznaną cechą ma dowolny rozkład o nieznannej wartości przeciętnej i wariancji, ale próbka jest stosunkowo duża (zwyczajowo $n \gtrsim 30$).

Model 3. Przedział ufności dla nieznannej wartości przeciętnej μ populacji, w której nieznaną cechą ma dowolny rozkład o nieznannej wartości przeciętnej i wariancji, ale próbka jest stosunkowo duża (zwyczajowo $n \gtrsim 30$).

Z CLT wynika, że statystyka $U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ ma asymptotyczny rozkład normalny $\mathcal{N}(0, 1)$. Z próbki szacujemy (punktowo) nieznanne odchylenie standardowe: $\sigma = S$.

Dalej postępujemy jak w modelu 1 otrzymując

$$P\left(-u < \frac{\bar{X} - \mu}{S} \sqrt{n} < u\right) = 1 - \alpha$$

$$P\left(-u \frac{S}{\sqrt{n}} + \bar{X} < \mu < u \frac{S}{\sqrt{n}} + \bar{X}\right) = 1 - \alpha$$

Interludium: rozkład χ^2

Rozważmy zmienną losową $Z = \sum_{i=1}^n X_i^2$, gdzie X_i są iid o rozkładzie $\mathcal{N}(0, 1)$. Wyprowadzimy jej PDF.

Interludium: rozkład χ^2

Rozważmy zmienną losową $Z = \sum_{i=1}^n X_i^2$, gdzie X_i są iid o rozkładzie $\mathcal{N}(0, 1)$. Wyprowadzimy jej PDF.

Krok 1. Niech $X \sim \mathcal{N}(0, 1)$, zatem ma PDF $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$; zdefiniujmy $Y = X^2$. Dla $y > 0$ transformacja $y = x^2$ prowadzi do $x = \pm\sqrt{y}$. Otrzymujemy

$$f_Y(y) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}}$$

Interludium: rozkład χ^2

Rozważmy zmienną losową $Z = \sum_{i=1}^n X_i^2$, gdzie X_i są iid o rozkładzie $\mathcal{N}(0, 1)$. Wyprowadzimy jej PDF.

Krok 1. Niech $X \sim \mathcal{N}(0, 1)$, zatem ma PDF $f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$; zdefiniujmy $Y = X^2$. Dla $y > 0$ transformacja $y = x^2$ prowadzi do $x = \pm\sqrt{y}$. Otrzymujemy

$$f_Y(y) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}}$$

Skoro $f_X(x)$ jest symetryczna, to

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \frac{\exp\left(-\frac{y}{2}\right)}{\sqrt{y}}$$

Jest to rozkład Gamma(α, θ) z $\alpha = \frac{1}{2}$, $\theta = 2$.

Krok 2. Rozpatrzmy rozkład sumy dwóch iid zmiennych losowych o rozkładach Gamma:

$$Y_1 \sim \text{Gamma}(\alpha, \theta), \quad Y_2 \sim \text{Gamma}(\beta, \theta), \quad \theta > 0$$

Ich PDFy to

$$f_{Y_1}(y_1) = \frac{1}{\Gamma(\alpha)\theta^\alpha} y_1^{\alpha-1} \exp\left(-\frac{y_1}{\theta}\right), \quad y_1 > 0$$

$$f_{Y_2}(y_2) = \frac{1}{\Gamma(\beta)\theta^\beta} y_2^{\beta-1} \exp\left(-\frac{y_2}{\theta}\right), \quad y_2 > 0$$

Interesuje nas zatem zmienna $S = Y_1 + Y_2$ o rozkładzie:

$$f_S(s) = \int_0^s f_{Y_1}(y_1) f_{Y_2}(s - y_1) dy_1$$

Interludium: rozkład χ^2

$$f_S(s) = \int_0^s f_{Y_1}(y_1) f_{Y_2}(s - y_1) dy_1$$

Podstawiając PDFy:

$$f_S(s) = \frac{\exp\left(-\frac{s}{\theta}\right)}{\Gamma(\alpha)\Gamma(\beta)\theta^{\alpha+\beta}} \int_0^s y_1^{\alpha-1} (s - y_1)^{\beta-1} dy_1$$

$$f_S(s) = \int_0^s f_{Y_1}(y_1) f_{Y_2}(s - y_1) dy_1$$

Podstawiając PDFy:

$$f_S(s) = \frac{\exp\left(-\frac{s}{\theta}\right)}{\Gamma(\alpha)\Gamma(\beta)\theta^{\alpha+\beta}} \int_0^s y_1^{\alpha-1} (s - y_1)^{\beta-1} dy_1$$

Podstawienie $y_1 = su$, $dy_1 = sdu$:

$$\int_0^s y_1^{\alpha-1} (s - y_1)^{\beta-1} dy_1 = s^{\alpha+\beta-1} \int_0^1 u^{\alpha-1} (1 - u)^{\beta-1} du$$

Interludium: rozkład χ^2

$$f_S(s) = \int_0^s f_{Y_1}(y_1) f_{Y_2}(s - y_1) dy_1$$

Podstawiając PDFy:

$$f_S(s) = \frac{\exp\left(-\frac{s}{\theta}\right)}{\Gamma(\alpha)\Gamma(\beta)\theta^{\alpha+\beta}} \int_0^s y_1^{\alpha-1} (s - y_1)^{\beta-1} dy_1$$

Podstawienie $y_1 = su$, $dy_1 = sdu$:

$$\int_0^s y_1^{\alpha-1} (s - y_1)^{\beta-1} dy_1 = s^{\alpha+\beta-1} \int_0^1 u^{\alpha-1} (1 - u)^{\beta-1} du$$

Ostatnia całka jest definicją funkcji beta, która wyraża się przez funkcję Γ :

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Interludium: rozkład χ^2

Podstawiając z powrotem i upraszczając:

$$f_S(s) = \frac{1}{\Gamma(\alpha + \beta)\theta^{\alpha+\beta}} s^{\alpha+\beta-1} \exp\left(-\frac{s}{\theta}\right)$$

Interludium: rozkład χ^2

Podstawiając z powrotem i upraszczając:

$$f_S(s) = \frac{1}{\Gamma(\alpha + \beta)\theta^{\alpha+\beta}} s^{\alpha+\beta-1} \exp\left(-\frac{s}{\theta}\right)$$

Rozpoznajemy zatem, że $Y_1 + Y_2 \sim \text{Gamma}(\alpha + \beta, \theta)$

Interludium: rozkład χ^2

Podstawiając z powrotem i upraszczając:

$$f_S(s) = \frac{1}{\Gamma(\alpha + \beta)\theta^{\alpha+\beta}} s^{\alpha+\beta-1} \exp\left(-\frac{s}{\theta}\right)$$

Rozpoznajemy zatem, że $Y_1 + Y_2 \sim \text{Gamma}(\alpha + \beta, \theta)$

Krok 3. Przez indukcję: jeśli $Y_i = X_i^2 \sim \text{Gamma}(\alpha_i, \theta)$, to

$$Z = \sum_{i=1}^n Y_i = \sum_{i=1}^n X_i^2 \sim \text{Gamma}\left(\sum_{i=1}^n \alpha_i, \theta\right)$$

Interludium: rozkład χ^2

Podstawiając z powrotem i upraszczając:

$$f_S(s) = \frac{1}{\Gamma(\alpha + \beta)\theta^{\alpha+\beta}} s^{\alpha+\beta-1} \exp\left(-\frac{s}{\theta}\right)$$

Rozpoznajemy zatem, że $Y_1 + Y_2 \sim \text{Gamma}(\alpha + \beta, \theta)$

Krok 3. Przez indukcję: jeśli $Y_i = X_i^2 \sim \text{Gamma}(\alpha_i, \theta)$, to

$$Z = \sum_{i=1}^n Y_i = \sum_{i=1}^n X_i^2 \sim \text{Gamma}\left(\sum_{i=1}^n \alpha_i, \theta\right)$$

Pokazaliśmy w kroku 1., że $X_i^2 \sim \text{Gamma}\left(\frac{1}{2}, 2\right)$, zatem

$$Z = \sum_{i=1}^n X_i^2 \sim \text{Gamma}\left(\frac{n}{2}, 2\right) \equiv \chi_n^2$$

z PDFem: $f_Z(z) = \frac{1}{\Gamma\left(\frac{n}{2}\right) (\sqrt{2})^n} z^{\frac{n}{2}-1} \exp(-z/2)$

Model 4. Przedział ufności dla nieznannej wartości wariancji (ew. odchylenia standardowego) populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ o nieznanymi wartości przeciętnej i wariancji. Konstrukcja przedziału ufności opiera się na statystyce

$$\chi^2 = \frac{nS_1^2}{\sigma^2} = \frac{(n-1)S_2^2}{\sigma^2} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}$$

tj. na rozkładzie χ^2 z $n - 1$ stopniami swobody.

Model 4. Przedział ufności dla nieznannej wartości wariancji (ew. odchylenia standardowego) populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ o nieznanymi wartości przeciętnej i wariancji. Konstrukcja przedziału ufności opiera się na statystyce

$$\chi^2 = \frac{nS_1^2}{\sigma^2} = \frac{(n-1)S_2^2}{\sigma^2} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2}$$

tj. na rozkładzie χ^2 z $n - 1$ stopniami swobody. Postać PDF jest następująca:

$$f(\chi^2) = f(x) = \frac{1}{(\sqrt{2})^n \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp(-x/2)$$

Wartość oczekiwana i wariancja tego rozkładu wynoszą $\mathbb{E}(X) = n$, $V(X) = 2n$.

- Jeśli X_i są iid o rozkładzie $\mathcal{N}(0, 1)$, to zmienna $\chi^2 = \sum_{i=1}^n X_i^2$ ma rozkład χ^2 z n stopniami swobody
- Jeśli X_i są iid o rozkładzie $\mathcal{N}(\mu, \sigma)$ oraz $S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$, to zmienna $\frac{nS_0^2}{\sigma^2}$ ma rozkład χ^2 o n stopniach swobody
- Jeśli X_i ma rozkład $\mathcal{N}(\mu, \sigma)$ oraz $S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, to zmienna $\frac{nS_1^2}{\sigma^2}$ ma rozkład χ^2 o $n - 1$ stopniach swobody — jeden stopień swobody został „zużyty” na wyznaczenie wartości estymatora wartości przeciętnej

Przedział ufności konstruuje się w oparciu o

$$P \left(\chi_1^2 < \frac{nS_1^2}{\sigma^2} < \chi_2^2 \right) = 1 - \alpha$$

- Jeśli X_i są iid o rozkładzie $\mathcal{N}(0, 1)$, to zmienna $\chi^2 = \sum_{i=1}^n X_i^2$ ma rozkład χ^2 z n stopniami swobody
- Jeśli X_i są iid o rozkładzie $\mathcal{N}(\mu, \sigma)$ oraz $S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$, to zmienna $\frac{nS_0^2}{\sigma^2}$ ma rozkład χ^2 o n stopniach swobody
- Jeśli X_i ma rozkład $\mathcal{N}(\mu, \sigma)$ oraz $S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, to zmienna $\frac{nS_1^2}{\sigma^2}$ ma rozkład χ^2 o $n - 1$ stopniach swobody — jeden stopień swobody został „zużyty” na wyznaczenie wartości estymatora wartości przeciętnej

Przedział ufności konstruuje się w oparciu o

$$P \left(\chi_1^2 < \frac{nS_1^2}{\sigma^2} < \chi_2^2 \right) = 1 - \alpha$$

- Jeśli X_i są iid o rozkładzie $\mathcal{N}(0, 1)$, to zmienna $\chi^2 = \sum_{i=1}^n X_i^2$ ma rozkład χ^2 z n stopniami swobody
- Jeśli X_i są iid o rozkładzie $\mathcal{N}(\mu, \sigma)$ oraz $S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$, to zmienna $\frac{nS_0^2}{\sigma^2}$ ma rozkład χ^2 o n stopniach swobody
- Jeśli X_i ma rozkład $\mathcal{N}(\mu, \sigma)$ oraz $S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, to zmienna $\frac{nS_1^2}{\sigma^2}$ ma rozkład χ^2 o $n - 1$ stopniach swobody — jeden stopień swobody został „zużyty” na wyznaczenie wartości estymatora wartości przeciętnej

Przedział ufności konstruuje się w oparciu o

$$P \left(\chi_1^2 < \frac{nS_1^2}{\sigma^2} < \chi_2^2 \right) = 1 - \alpha$$

- Jeśli X_i są iid o rozkładzie $\mathcal{N}(0, 1)$, to zmienna $\chi^2 = \sum_{i=1}^n X_i^2$ ma rozkład χ^2 z n stopniami swobody
- Jeśli X_i są iid o rozkładzie $\mathcal{N}(\mu, \sigma)$ oraz $S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$, to zmienna $\frac{nS_0^2}{\sigma^2}$ ma rozkład χ^2 o n stopniach swobody
- Jeśli X_i ma rozkład $\mathcal{N}(\mu, \sigma)$ oraz $S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, to zmienna $\frac{nS_1^2}{\sigma^2}$ ma rozkład χ^2 o $n - 1$ stopniach swobody — jeden stopień swobody został „zużyty” na wyznaczenie wartości estymatora wartości przeciętnej

Przedział ufności konstruuje się w oparciu o

$$P \left(\chi_1^2 < \frac{nS_1^2}{\sigma^2} < \chi_2^2 \right) = 1 - \alpha$$

- Jeśli X_i są iid o rozkładzie $\mathcal{N}(0, 1)$, to zmienna $\chi^2 = \sum_{i=1}^n X_i^2$ ma rozkład χ^2 z n stopniami swobody
- Jeśli X_i są iid o rozkładzie $\mathcal{N}(\mu, \sigma)$ oraz $S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$, to zmienna $\frac{nS_0^2}{\sigma^2}$ ma rozkład χ^2 o n stopniach swobody
- Jeśli X_i ma rozkład $\mathcal{N}(\mu, \sigma)$ oraz $S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$, to zmienna $\frac{nS_1^2}{\sigma^2}$ ma rozkład χ^2 o $n - 1$ stopniach swobody — jeden stopień swobody został „zużyty” na wyznaczenie wartości estymatora wartości przeciętnej

Przedział ufności konstruuje się w oparciu o

$$P \left(\chi_1^2 < \frac{nS_1^2}{\sigma^2} < \chi_2^2 \right) = 1 - \alpha$$

Założmy symetryczność (tj. równe prawdopodobieństwa w lewym i prawym ogonie): $P(\chi^2 < \chi_1^2) = P(\chi^2 > \chi_2^2)$, a więc $\chi_1^2 = \chi_1^2(\frac{\alpha}{2}, n - 1)$ jest kwantylem rzędu $\frac{\alpha}{2}$, zaś $\chi_2^2 = \chi_2^2(1 - \frac{\alpha}{2}, n - 1)$ jest kwantylem rzędu $1 - \frac{\alpha}{2}$ rozkładu χ^2 o $n - 1$ stopniach swobody. Kwantyle wyznaczamy na zadanym poziomie istotności α . Zatem:

Założmy symetryczność (tj. równe prawdopodobieństwa w lewym i prawym ogonie): $P(\chi^2 < \chi_1^2) = P(\chi^2 > \chi_2^2)$, a więc $\chi_1^2 = \chi_1^2(\frac{\alpha}{2}, n-1)$ jest kwantylem rzędu $\frac{\alpha}{2}$, zaś $\chi_2^2 = \chi_2^2(1 - \frac{\alpha}{2}, n-1)$ jest kwantylem rzędu $1 - \frac{\alpha}{2}$ rozkładu χ^2 o $n-1$ stopniach swobody. Kwantyle wyznaczamy na zadanym poziomie istotności α . Zatem:

$$P\left(\frac{nS_1^2}{\chi_2^2} < \sigma^2 < \frac{nS_1^2}{\chi_1^2}\right) = 1 - \alpha$$

czyli

$$P\left(S_1\sqrt{\frac{n}{\chi_2^2}} < \sigma < S_1\sqrt{\frac{n}{\chi_1^2}}\right) = 1 - \alpha$$

Model 5. Przedział ufności dla nieznannej wartości wariancji (ew. odchylenia standardowego) populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ o nieznanymi wartości przeciętnej i wariancji, ale próbka jest stosunkowo duża (zwyczajowo $n \gtrsim 50$).

Model 5. Przedział ufności dla nieznannej wartości wariancji (ew. odchylenia standardowego) populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ o nieznanymi wartości przeciętnej i wariancji, ale próbka jest stosunkowo duża (zwyczajowo $n \gtrsim 50$).

Tutaj można postępować jak w modelu 4 albo — dla ułatwienia obliczeń — zastosować przybliżenie rozkładu χ^2 rozkładem normalnym.

Model 5. Przedział ufności dla nieznannej wartości wariancji (ew. odchylenia standardowego) populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ o nieznanymi wartości przeciętnej i wariancji, ale próbka jest stosunkowo duża (zwyczajowo $n \gtrsim 50$).

Tutaj można postępować jak w modelu 4 albo — dla ułatwienia obliczeń — zastosować przybliżenie rozkładu χ^2 rozkładem normalnym.

Dokładniej, statystyka $\sqrt{2\chi^2} = \sqrt{2\frac{nS^2}{\sigma^2}} = \frac{S}{\sigma}\sqrt{2n}$ dla dostatecznie dużych n ma w przybliżeniu (uwzględniając tzw. poprawkę Edgewortha na skośność rozkładu χ^2 ; bez dowodu) rozkład $\mathcal{N}(\sqrt{2n-3}, 1)$. Zatem:

Model 5. Przedział ufności dla nieznannej wartości wariancji (ew. odchylenia standardowego) populacji, w której nieznaną cechą ma rozkład $\mathcal{N}(\mu, \sigma)$ o nieznanymi wartości przeciętnej i wariancji, ale próbka jest stosunkowo duża (zwyczajowo $n \gtrsim 50$).

Tutaj można postępować jak w modelu 4 albo — dla ułatwienia obliczeń — zastosować przybliżenie rozkładu χ^2 rozkładem normalnym.

Dokładniej, statystyka $\sqrt{2\chi^2} = \sqrt{2\frac{nS^2}{\sigma^2}} = \frac{S}{\sigma}\sqrt{2n}$ dla dostatecznie dużych n ma w przybliżeniu (uwzględniając tzw. poprawkę Edgewortha na skośność rozkładu χ^2 ; bez dowodu) rozkład $\mathcal{N}(\sqrt{2n-3}, 1)$. Zatem:

$$P\left(\frac{S\sqrt{n}}{\sqrt{2n-3} + u(1-\alpha/2)} < \sigma < \frac{S\sqrt{n}}{\sqrt{2n-3} - u(1-\alpha/2)}\right) = 1 - \alpha$$

gdzie $u(1 - \alpha/2)$ jest kwantylem rzędu $1 - \alpha/2$ rozkładu $\mathcal{N}(0, 1)$.

Przykład. Badanie opinii publicznej.

Wyniki takie (np. sondaże) są podawane w procentach, z dokładnością 3 pp. Te 3 pp. określają szerokość 95% przedziału ufności. Ile osób należy przeankietować by osiągnąć taką dokładność?

Przykład. Badanie opinii publicznej.

Wyniki takie (np. sondaże) są podawane w procentach, z dokładnością 3 pp. Te 3 pp. określają szerokość 95% przedziału ufności. Ile osób należy przeankietować by osiągnąć taką dokładność?

Rozw. Rozważmy rozkład dwupunktowy, tj. taki, w którym zmienna losowa przyjmuje jedną z dwóch wartości. Charakterystyką tego rozkładu jest wskaźnik struktury p (nazywany prawdopodobieństwem sukcesu), który jest częstością wystąpienia w populacji elementu o określonej właściwości.

Przykład. Badanie opinii publicznej.

Wyniki takie (np. sondaże) są podawane w procentach, z dokładnością 3 pp. Te 3 pp. określają szerokość 95% przedziału ufności. Ile osób należy przeankietować by osiągnąć taką dokładność?

Rozw. Rozważmy rozkład dwupunktowy, tj. taki, w którym zmienna losowa przyjmuje jedną z dwóch wartości. Charakterystyką tego rozkładu jest wskaźnik struktury p (nazywany prawdopodobieństwem sukcesu), który jest częstością wystąpienia w populacji elementu o określonej właściwości.

- ❶ **Estymacja punktowa.** Estymatorem jest stosunek liczby elementów wyróżnionych w próbie do liczności próbki:

$$\hat{p} = \frac{k}{n}$$

k ma rozkład dwumianowy o parametrach $\mathbb{E}(k) = np$, $V(k) = npq$, zatem rozkład \hat{p} ma momenty, odpowiednio, p oraz pq/n .

W myśl twierdzeń granicznych dla dużej próby rozkład zmiennej losowej $U = \frac{k/n - p}{\sqrt{p(1-p)/n}}$ dąży do rozkładu $\mathcal{N}(0, 1)$. Zatem \hat{p} ma

w przybliżeniu rozkład $\mathcal{N}\left(p, \sqrt{pq/n}\right)$.

- 2 **Estymacja przedziałowa.** Dla dostatecznie dużej próby przedział ufności wskaźnika struktury na poziomie ufności $1 - \alpha$ można przybliżyć jako

$$\left(\hat{p} - u(1 - \alpha/2) \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + u(1 - \alpha/2) \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right)$$

Dokładność na poziomie 3 pp. oznacza, że

$$P(\hat{p} - 0.03 \leq p \leq \hat{p} + 0.03) \geq 0.95$$

Tutaj nierówność „ ≥ 0.95 ” oznacza, że przedział ufności ma być „najgorszym scenariuszem” w sensie maksymalizacji wariancji. Dlatego w dalszych rozważaniach znajdziemy takie p by zrealizować ów „najgorszy scenariusz”.

Dla $p \in [0, 1]$ zachodzi $p(1 - p) \leq \frac{1}{4}$, zaś $p(1 - p)$ ma maksimum lokalne dla $p = 0.5$. Stąd $\sqrt{\frac{p(1-p)}{n}} \leq \sqrt{\frac{1}{4n}} \Rightarrow \sqrt{\frac{n}{p(1-p)}} \geq \sqrt{4n}$.

Dla $p \in [0, 1]$ zachodzi $p(1-p) \leq \frac{1}{4}$, zaś $p(1-p)$ ma maksimum lokalne dla $p = 0.5$. Stąd $\sqrt{\frac{p(1-p)}{n}} \leq \sqrt{\frac{1}{4n}} \Rightarrow \sqrt{\frac{n}{p(1-p)}} \geq \sqrt{4n}$.

Obliczmy:

$$\begin{aligned} P(\hat{p} - 0.03 \leq p \leq \hat{p} + 0.03) &= \\ P\left(-\frac{0.03}{\sqrt{p(1-p)/n}} \leq \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \leq \frac{0.03}{\sqrt{p(1-p)/n}}\right) &= \\ \int_{-0.03\sqrt{n/[p(1-p)]}}^{0.03\sqrt{n/[p(1-p)]}} \varphi(s) ds &\geq \\ \int_{-0.03\sqrt{4n}}^{0.03\sqrt{4n}} \varphi(s) ds &\geq 0.95 \end{aligned}$$

Dla $p \in [0, 1]$ zachodzi $p(1-p) \leq \frac{1}{4}$, zaś $p(1-p)$ ma maksimum lokalne dla $p = 0.5$. Stąd $\sqrt{\frac{p(1-p)}{n}} \leq \sqrt{\frac{1}{4n}} \Rightarrow \sqrt{\frac{n}{p(1-p)}} \geq \sqrt{4n}$.

Obliczmy:

$$\begin{aligned} P(\hat{p} - 0.03 \leq p \leq \hat{p} + 0.03) &= \\ P\left(-\frac{0.03}{\sqrt{p(1-p)/n}} \leq \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \leq \frac{0.03}{\sqrt{p(1-p)/n}}\right) &= \\ \int_{-0.03\sqrt{n/[p(1-p)]}}^{0.03\sqrt{n/[p(1-p)]}} \varphi(s) ds &\geq \\ \int_{-0.03\sqrt{4n}}^{0.03\sqrt{4n}} \varphi(s) ds &\geq 0.95 \end{aligned}$$

Aby ostatnia nierówność była spełniona, potrzeba by

$$0.03\sqrt{4n} \geq 1.96 \Rightarrow n \geq 1068$$

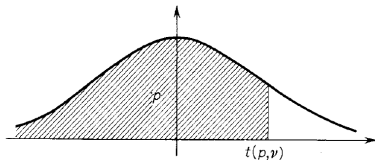
Badania sondażowe zatem bazują na próbie około 1100 osób.

Tablica 5. Wartości $\Phi(u)$ dystrybuanty rozkładu normalnego $N(0, 1)$

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	,5398	,5438	,5478	,5517	,5557	,5596	,5636	,5675	,5714	,5753
0,2	,5793	,5832	,5871	,5910	,5948	,5987	,6026	,6064	,6103	,6141
0,3	,6179	,6217	,6255	,6293	,6331	,6368	,6406	,6443	,6480	,6517
0,4	,6554	,6591	,6628	,6664	,6700	,6736	,6772	,6808	,6844	,6879
0,5	,6915	,6950	,6985	,7019	,7054	,7088	,7123	,7157	,7190	,7224
0,6	,7257	,7290	,7324	,7357	,7389	,7422	,7454	,7486	,7517	,7549
0,7	,7580	,7611	,7642	,7673	,7704	,7734	,7764	,7794	,7823	,7852
0,8	,7881	,7910	,7939	,7967	,7995	,8023	,8051	,8078	,8106	,8133
0,9	,8159	,8186	,8212	,8238	,8264	,8289	,8314	,8340	,8365	,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	,8643	,8665	,8686	,8708	,8729	,8749	,8770	,8790	,8810	,8830
1,2	,8849	,8869	,8888	,8907	,8925	,8944	,8962	,8980	,8997	,9015
1,3	,9032	,9049	,9066	,9082	,9099	,9115	,9131	,9147	,9162	,9177
1,4	,9192	,9207	,9222	,9236	,9251	,9265	,9279	,9292	,9306	,9319
1,5	,9332	,9345	,9357	,9370	,9382	,9394	,9406	,9418	,9429	,9441
1,6	,9452	,9463	,9474	,9484	,9495	,9505	,9515	,9525	,9535	,9545
1,7	,9554	,9564	,9573	,9582	,9591	,9599	,9608	,9616	,9625	,9633
1,8	,9641	,9649	,9656	,9664	,9671	,9678	,9686	,9693	,9699	,9706
1,9	,9713	,9719	,9726	,9732	,9738	,9744	,9750	,9756	,9761	,9767
2,0	0,9772	0,9779	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	,9821	,9826	,9830	,9834	,9838	,9842	,9846	,9850	,9854	,9857
2,2	,9861	,9864	,9868	,9871	,9875	,9878	,9881	,9884	,9887	,9890
2,3	,9893	,9896	,9898	,9901	,9904	,9906	,9909	,9911	,9913	,9916
2,4	,9918	,9920	,9922	,9925	,9927	,9929	,9931	,9932	,9934	,9936
2,5	,9938	,9940	,9941	,9943	,9945	,9946	,9948	,9949	,9951	,9952
2,6	,9953	,9955	,9956	,9957	,9959	,9960	,9961	,9962	,9963	,9964
2,7	,9965	,9966	,9967	,9968	,9969	,9970	,9971	,9972	,9973	,9974
2,8	,9974	,9975	,9976	,9977	,9977	,9978	,9979	,9979	,9980	,9981
2,9	,9981	,9982	,9982	,9983	,9984	,9984	,9985	,9985	,9986	,9986

Tablica 6. Kwantyle $u(p)$ rzędu p rozkładu normalnego $N(0, 1)$

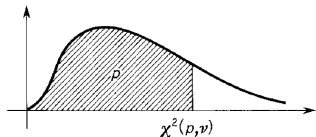
p	0,90	0,95	0,975	0,99	0,995
$u(p)$	1,28	1,64	1,96	2,33	2,58



T a b l i c a 7. Kwantyle $t(p, v)$ rzędu p rozkładu Studenta o v stopniach swobody

v	p				
	0,90	0,95	0,975	0,99	0,995
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	,638	,353	3,182	4,541	5,841
4	,533	,132	2,776	3,747	4,604
5	,476	,015	,571	,365	,032
6	1,440	1,943	2,447	3,143	3,707
7	,415	,895	,365	2,998	,499
8	,397	,859	,306	,897	,355
9	,383	,833	,262	,821	,250
10	,372	,812	,228	,764	,169
11	1,363	1,795	2,201	2,718	3,106
12	,356	,782	,179	,681	,054
13	,350	,771	,160	,650	,012
14	,345	,761	,145	,624	2,977
15	,341	,753	,131	,602	,947
16	1,337	1,746	2,120	2,583	2,921
17	,333	,740	,110	,567	,898
18	,330	,734	,101	,552	,878
19	,328	,729	,093	,539	,861
20	,325	,725	,086	,528	,845

Tablica 8. Kwantyle $\chi^2(p, \nu)$ rzędu p rozkładu χ^2 o ν stopniach swobody



ν	p							
	0,005	0,01	0,025	0,05	0,95	0,975	0,99	0,995
1	—	—	0,001	0,004	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	11,071	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	21,026	23,337	26,217	28,299
13	3,565	4,107	5,009	5,892	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	31,410	34,170	37,566	39,997