Statystyka z próby (jako zmienna losowa)

Statystyka z próby (jako zmienna losowa)

Definicja

Statystyką z próby nazywamy zmienną losową U, będąca funkcją obserwowanych w próbie zmiennych losowych $X_1, X_2, \ldots, X_n$:

$$U = f(X_1, X_2, \ldots, X_n).$$

Statystyka z próby (jako zmienna losowa)

Definicja

Statystyką z próby nazywamy zmienną losową U, będąca funkcją obserwowanych w próbie zmiennych losowych $X_1, X_2, \ldots, X_n$:

$$U = f(X_1, X_2, \ldots, X_n).$$

Rozkład U zależy od liczebności próby oraz od rozkładu badanej populacji.

Statystyka z próby (jako zmienna losowa)

Przykłady

$$\bar{X} = \frac{X_1 + X_2 \ldots + X_n}{n} \qquad \text{(średnia arytmetyczna z próby)}$$

$$S^2 = \frac{(X_1^2 + X_2^2 + \ldots + X_n^2) - n\bar{X}^2}{n-1} \qquad \text{(wariancja z próby)}$$

Przykład – statystyka będąca średnią arytmetyczną z próby

- X – zmienna losowa, opisująca badaną **cechę** w populacji generalnej, ma rozkład $\mathcal{N}(m, \sigma^2)$ (założenie)
- $(X_1, X_2, \ldots, X_n)$ – pobrana **próba losowa** jest **prosta** (założenie)

Przykład – statystyka będąca średnią arytmetyczną z próby

- X – zmienna losowa, opisująca badaną **cechę** w populacji generalnej, ma rozkład $\mathcal{N}(m, \sigma^2)$ (założenie)
- $(X_1, X_2, \ldots, X_n)$ – pobrana **próba losowa** jest **prosta** (założenie)
- $\bar{X}$ – **średnia arytmetyczna z próby (czyli statystyka z próby)** też ma rozkład normalny

Przykład – statystyka będąca średnią arytmetyczną z próby

- X – zmienna losowa, opisująca badaną **cechę** w populacji generalnej, ma rozkład $\mathcal{N}(m, \sigma^2)$ (założenie)
- $(X_1, X_2, \ldots, X_n)$ – pobrana **próba losowa** jest **prosta** (założenie)
- $\bar{X}$ – **średnia arytmetyczna z próby** **(czyli statystyka z próby)** też ma rozkład normalny

Badamy parametry rozkładu statystyki $\bar{X}$:

$$\mathbb{E}\left(\bar{X}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}nm = m,$$

$$D^2\left(\bar{X}\right) = D^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D^2(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n},$$

Przykład – statystyka będąca średnią arytmetyczną z próby

- X – zmienna losowa, opisująca badaną **cechę** w populacji generalnej, ma rozkład $\mathcal{N}(m, \sigma^2)$ (założenie)
- $(X_1, X_2, \ldots, X_n)$ – pobrana **próba losowa** jest **prosta** (założenie)
- $\bar{X}$ – **średnia arytmetyczna z próby (czyli statystyka z próby)** też ma rozkład normalny

Badamy parametry rozkładu statystyki $\bar{X}$:

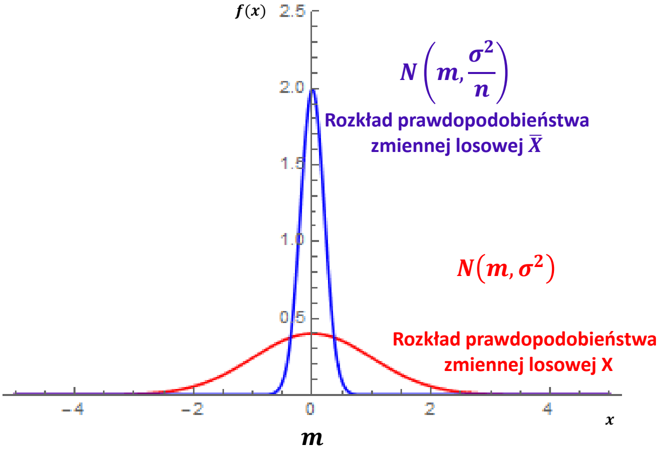
$$\mathbb{E}\left(\bar{X}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}nm = m,$$

$$D^2\left(\bar{X}\right) = D^2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D^2(X_i) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n},$$

Zatem **statystyka $\bar{X}$ ma rozkład $\mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$**.

Rozkład statystyki będącej średnią arytmetyczną z próby

$\mathbb{E}\left(\bar{X}\right) = \mathbb{E}(X) = m,$ $\quad D^2\left(\bar{X}\right)$ – maleje do zera wraz ze wzrostem liczebności próby



Standaryzacja statystyki $\bar{X}$ o rozkładzie normalnym

Standaryzacja statystyki $\bar{X}$ o rozkładzie normalnym

Standaryzacja statystyki $\bar{X}$ o rozkładzie normalnym

- Odejmujemy od $\bar{X}$ jej wartość średnią $\mathbb{E}(\bar{X}) = m$ i dzielimy otrzymaną zmienną przez odchylenie standardowe $D(\bar{X}) = \frac{\sigma}{\sqrt{n}}$:

$$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}}.$$

- Otrzymana zmienna losowa Z ma już rozkład $\mathcal{N}(0, 1)$.

Rozkład statystyki będącej średnią arytmetyczną z próby

Przykład

W wyniku badań ustalono, że rozkład płac pracowników w pewnym przedsiębiorstwie jest normalny, z wartością oczekiwaną 4,5 tys. zł i odchyleniem standardowym 0,5 tys. zł. Wybrano próbkę 16 pracowników. Jakie jest prawdopodobieństwo, że średnia płaca wylosowanych pracowników jest większa od 4 400 zł?

Rozkład statystyki będącej średnią arytmetyczną z próby

Przykład

W wyniku badań ustalono, że rozkład płac pracowników w pewnym przedsiębiorstwie jest normalny, z wartością oczekiwaną 4,5 tys. zł i odchyleniem standardowym 0,5 tys. zł. Wybrano próbkę 16 pracowników. Jakie jest prawdopodobieństwo, że średnia płaca wylosowanych pracowników jest większa od 4 400 zł?

$n = 16$

Rozkład statystyki będącej średnią arytmetyczną z próby

Przykład

W wyniku badań ustalono, że rozkład płac pracowników w pewnym przedsiębiorstwie jest normalny, z wartością oczekiwaną 4,5 tys. zł i odchyleniem standardowym 0,5 tys. zł. Wybrano próbkę 16 pracowników. Jakie jest prawdopodobieństwo, że średnia płaca wylosowanych pracowników jest większa od 4 400 zł?

$n = 16$
$\mathbb{E}(\bar{X}) = m = 4,5$ [tys. zł],

Rozkład statystyki będącej średnią arytmetyczną z próby

Przykład

W wyniku badań ustalono, że rozkład płac pracowników w pewnym przedsiębiorstwie jest normalny, z wartością oczekiwaną 4,5 tys. zł i odchyleniem standardowym 0,5 tys. zł. Wybrano próbkę 16 pracowników. Jakie jest prawdopodobieństwo, że średnia płaca wylosowanych pracowników jest większa od 4 400 zł?

$n = 16$

$\mathbb{E}(\bar{X}) = m = 4,5$ [tys. zł], $\quad D^2(\bar{X}) = \frac{\sigma^2}{n} = \frac{0{,}25}{16} = 0,015625, \quad D(\bar{X}) = \frac{0{,}5}{4} = 0,125.$

Rozkład statystyki będącej średnią arytmetyczną z próby

Przykład

W wyniku badań ustalono, że rozkład płac pracowników w pewnym przedsiębiorstwie jest normalny, z wartością oczekiwaną 4,5 tys. zł i odchyleniem standardowym 0,5 tys. zł. Wybrano próbkę 16 pracowników. Jakie jest prawdopodobieństwo, że średnia płaca wylosowanych pracowników jest większa od 4 400 zł?

$n = 16$

$\mathbb{E}(\bar{X}) = m = 4,5$ [tys. zł], $\quad D^2(\bar{X}) = \frac{\sigma^2}{n} = \frac{0{,}25}{16} = 0,015625, \quad D(\bar{X}) = \frac{0{,}5}{4} = 0,125.$

Zatem $\bar{X} \sim \mathcal{N}\,(4,5\ ,\ 0,015625)$ oraz

Rozkład statystyki będącej średnią arytmetyczną z próby

Przykład

W wyniku badań ustalono, że rozkład płac pracowników w pewnym przedsiębiorstwie jest normalny, z wartością oczekiwaną 4,5 tys. zł i odchyleniem standardowym 0,5 tys. zł. Wybrano próbkę 16 pracowników. Jakie jest prawdopodobieństwo, że średnia płaca wylosowanych pracowników jest większa od 4 400 zł?

$n = 16$
$\mathbb{E}(\bar{X}) = m = 4,5$ [tys. zł], $\quad D^2(\bar{X}) = \frac{\sigma^2}{n} = \frac{0,25}{16} = 0,015625, \quad D(\bar{X}) = \frac{0,5}{4} = 0,125.$
Zatem $\bar{X} \sim \mathcal{N}(4,5\ ,\ 0,015625)$ oraz

$$\mathbb{P}(\bar{X} > 4,4) = \mathbb{P}\left(\frac{\bar{X} - m}{\sigma/\sqrt{n}} > \frac{4,4 - 4,5}{0,125}\right) = \mathbb{P}\left(Z > \frac{-0,1}{0,125}\right) = \mathbb{P}(Z > -0,8) =$$

$$\overset{\text{z symetryczności } \mathcal{N}(0,1)}{=} \mathbb{P}(Z < 0,8) = 0,7881.$$

Studentyzacja statystyki $\bar{X}$

Studentyzacja statystyki $\bar{X}$

Studentyzacja statystyki $\bar{X}$

- Wykonywana jest, **gdy odchylenie standardowe σ w populacji nie jest znane**.

Studentyzacja statystyki $\bar{X}$

Studentyzacja statystyki $\bar{X}$

- Wykonywana jest, **gdy odchylenie standardowe σ w populacji nie jest znane**.

$$t = \frac{\bar{X} - m}{\frac{S}{\sqrt{n}}},$$

gdzie S jest odchyleniem standardowym z próby, tzn. $S = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)}$.

Studentyzacja statystyki $\bar{X}$

Studentyzacja statystyki $\bar{X}$

- Wykonywana jest, **gdy odchylenie standardowe σ w populacji nie jest znane**.

$$t = \frac{\bar{X} - m}{\frac{S}{\sqrt{n}}},$$

gdzie S jest odchyleniem standardowym z próby, tzn. $S = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)}$.

- Statystyka t jest niezależna od σ i ma rozkład Studenta o $(n-1)$ stopniach swobody.
- Liczba stopni swobody jest jedynym parametrem rozkładu Studenta.

Rozkład statystyki będącej średnią arytmetyczną z próby

- Krzywa gęstości rozkładu Studenta – przypomina krzywą rozkładu normalnego standardowego, choć jest bardziej <u>spłaszczona</u>.

Rozkład statystyki będącej średnią arytmetyczną z próby

- Krzywa gęstości rozkładu Studenta – przypomina krzywą rozkładu normalnego standardowego, choć jest bardziej spłaszczona.
- Przy wzroście liczby stopni swobody, **rozkład Studenta jest zbieżny do rozkładu normalnego standardowego** $\mathcal{N}(0,1)$.

Rozkład statystyki będącej średnią arytmetyczną z próby

- Krzywa gęstości rozkładu Studenta – przypomina krzywą rozkładu normalnego standardowego, choć jest bardziej spłaszczona.
- Przy wzroście liczby stopni swobody, **rozkład Studenta jest zbieżny do rozkładu normalnego standardowego** $\mathcal{N}(0,1)$.

- Dla $k > 30$ – dokładny rozkład t Studenta zastępowany jest rozkładem granicznym, czyli $\mathcal{N}(0,1)$.
- Dla $k \leqslant 30$ – korzystamy z tablic rozkładu t Studenta.

Estymator

Estymator

Definicja

Estymator parametru Q rozkładu zmiennej losowej X to taka statystyka

$$Z_n = f(X_1, X_2, \ldots, X_n),$$

której rozkład zależy od szacowanego parametru Q.

Estymator

Definicja

Estymator parametru Q rozkładu zmiennej losowej X to taka statystyka

$$Z_n = f(X_1, X_2, \ldots, X_n),$$

której rozkład zależy od szacowanego parametru Q.

Estymator służy do oceny wartości nieznanych parametrów populacji generalnej.

Estymator

Definicja

Estymator parametru Q rozkładu zmiennej losowej X to taka statystyka

$$Z_n = f(X_1, X_2, \ldots, X_n),$$

której rozkład zależy od szacowanego parametru Q.

Estymator służy do oceny wartości nieznanych parametrów populacji generalnej.

Przykłady

- $\bar{X}$ i S^2 – estymatory dla populacji generalnej o rozkładzie $\mathcal{N}(m, \sigma^2)$ z nieznanymi paramterami m i σ^2 (rozkłady $\bar{X}$ i S^2 zależą od nieznanych parametrów m i σ^2).