Podstawowe statystyki

Średnia arytmetyczna:

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

Wariancja empiryczna:

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

Wariancja próbkowa:

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

Empiryczne odchylenie standardowe: $S = \sqrt{S^2}$

Próbkowe odchylenie standardowe: $S^* = \sqrt{S^{*2}}$

$$S^{*2} = S^2\,\frac{n}{n-1}$$

Estymacja przedziałowa

Przedziały ufności dla wartości oczekiwanej

Model I

Cecha ma rozkład $N(\mu,\sigma)$ i wariancja σ^2 jest znana.

$$P\left(\overline{x} - u\left(1-\frac{1}{2}\alpha\right)\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x} + u\left(1-\frac{1}{2}\alpha\right)\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$u\left(1-\frac{1}{2}\alpha\right)$ - kwantyl rozkładu normalnego $N(0,1)$ rzędu $1-\frac{1}{2}\alpha$

Model II

Cecha ma rozkład $N(\mu,\sigma)$, ale wariancja σ^2 jest nieznana.

$$P\left(\overline{x} - t\left(1-\frac{1}{2}\alpha, n-1\right)\frac{S}{\sqrt{n-1}} \le \mu \le \overline{x} + t\left(1-\frac{1}{2}\alpha, n-1\right)\frac{S}{\sqrt{n-1}}\right) = 1-\alpha$$

lub (równoważnie)

$$P\left(\overline{x} - t\left(1-\frac{1}{2}\alpha, n-1\right)\frac{S^*}{\sqrt{n}} \le \mu \le \overline{x} + t\left(1-\frac{1}{2}\alpha, n-1\right)\frac{S^*}{\sqrt{n}}\right) = 1-\alpha$$

$t\left(1-\frac{1}{2}\alpha, n-1\right)$ - kwantyl rzędu $1-\frac{1}{2}\alpha$ z rozkładu Studenta o $n-1$ stopniach swobody

Model III

Cecha ma rozkład nieznany lub inny niż normalny. Dla konstrukcji przedziału ufności w tym przypadku musimy mieć próbę dużego rozmiaru. Zwykle zakłada się że $n > 30$ (niektórzy podają $n \ge 100$).

$$P\left(\overline{x}-u\left(1-\frac{1}{2}\alpha\right)\frac{S^*}{\sqrt{n}}\le\mu\le\overline{x}+u\left(1-\frac{1}{2}\alpha\right)\frac{S^*}{\sqrt{n}}\right)=1-\alpha$$

lub równoważnie

$$P\left(\overline{x}-u\left(1-\frac{1}{2}\alpha\right)\frac{S}{\sqrt{n-1}}\le\mu\le\overline{x}+u\left(1-\frac{1}{2}\alpha\right)\frac{S}{\sqrt{n-1}}\right)=1-\alpha$$

$u\left(1-\frac{1}{2}\alpha\right)$ - kwantyl rozkładu normalnego $N(0,1)$ rzędu $1-\frac{1}{2}\alpha$

Przedziały ufności dla wariancji

Model I

Cecha ma rozkład $N(\mu,\sigma)$, μ - znane, σ - nieznane

$$P\left(\frac{nS^2}{\chi^2\left(1-\frac{1}{2}\alpha,n\right)}\le\sigma^2\le\frac{nS^2}{\chi^2\left(\frac{1}{2}\alpha,n\right)}\right)=1-\alpha$$

$\chi^2\left(1-\frac{1}{2}\alpha,n\right)$ - kwantyl rzędu $1-\frac{1}{2}\alpha$ z rozkładu χ^2 o n stopniach swobody, $\chi^2\left(\frac{1}{2}\alpha,n\right)$ - kwantyl rzędu $\frac{1}{2}\alpha$ z rozkładu χ^2 o n stopniach swobody,

Model II

Cecha ma rozkład $N(\mu,\sigma)$, μ,σ - nieznane

$$P\left(\frac{nS^2}{\chi^2\left(1-\frac{1}{2}\alpha,n-1\right)}\le\sigma^2\le\frac{nS^2}{\chi^2\left(\frac{1}{2}\alpha,n-1\right)}\right)=1-\alpha$$

lub (równoważnie)

$$P\left(\frac{(n-1)S^{*2}}{\chi^2\left(1-\frac{1}{2}\alpha,n-1\right)}\le\sigma^2\le\frac{(n-1)S^{*2}}{\chi^2\left(\frac{1}{2}\alpha,n-1\right)}\right)=1-\alpha$$

$\chi^2\left(1-\frac{1}{2}\alpha,n-1\right)$ - kwantyl rzędu $1-\frac{1}{2}\alpha$ z rozkładu χ^2 o $n-1$ stopniach swobody,

$\chi^2\left(\frac{1}{2}\alpha,n-1\right)$ - kwantyl rzędu $\frac{1}{2}\alpha$ z rozkładu χ^2 o $n-1$ stopniach swobody,

Przedziały ufności dla odchylenia standardowego – odchylenie standardowe to pierwiastek z wariancji, korzystając z tego faktu otrzymujemy:

Model II

Cecha ma rozkład $N(\mu,\sigma)$, μ,σ - nieznane

$$\left[\sqrt{\frac{nS^2}{\chi^2\left(1-\frac{1}{2}\alpha, n-1\right)}}, \sqrt{\frac{nS^2}{\chi^2\left(\frac{1}{2}\alpha, n-1\right)}}\right]$$

lub (równoważnie)

$$\left[\sqrt{\frac{(n-1)S^{*2}}{\chi^2\left(1-\frac{1}{2}\alpha, n-1\right)}}, \sqrt{\frac{(n-1)S^{*2}}{\chi^2\left(\frac{1}{2}\alpha, n-1\right)}}\right]$$

Przedziały ufności dla wskaźnika struktury

Badana cecha X ma w populacji generalnej rozkład zero-jedynkowy: $P(X=1)=p$, $P(X=0)=1-p$. p - wskaźnik struktury.

Model I

Liczność próby n: k elementów o wartości cechy równej 1, $n-k$ o wartości cechy równej 0.

$$\left[B_{\frac{\alpha}{2}}(k, n-k+1), B_{1-\frac{\alpha}{2}}(k+1, n-k)\right]$$

$B_q(n_1,n_2)$ - kwantyl rzędu q rozkładu Beta z parametrami n_1, n_2

Model II

Bardzo duża próba ($n \geq 100$)

$\hat{p}=\frac{k}{n}$ ma asymptotyczny rozkład normalny o parametrach $\mu=p$, $\sigma=\sqrt{\frac{p(1-p)}{n}}$, „przybliżony” przedział ufności dla wskaźnika struktury:

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

$u\left(1-\frac{1}{2}\alpha\right)$ - kwantyl rozkładu normalnego $N(0,1)$ rzędu $1-\frac{1}{2}\alpha$

Jeżeli istniej możliwość wyznaczenia kwantyli rozkładu Beta, to nie zaleca się korzystać z modelu przybliżonego.

Wyznaczanie minimalnej liczebności próby niezbędnej do uzyskania przedziału ufności o zadanej długości

Model I dla wartości oczekiwanej

Maksymalny błąd oszacowania (połowa długości przedziału ufności)

$$\Delta_{\max} = u\left(1 - \frac{1}{2}\alpha\right) \cdot \frac{\sigma}{\sqrt{n}}$$

Minimalna liczebność próby $n_{\min}$ zapewniająca, ze długość przedziału ufności nie przekroczy wartości d :

$$n_{\min} = \left[\left(\frac{2u\left(1-\frac{1}{2}\alpha\right)\cdot\sigma}{d}\right)^2\right] + 1$$

Symbol $[x]$ oznacza część całkowitą z liczby x (funkcja podłogi).

Model II dla wartości oczekiwanej

Maksymalny błąd oszacowania (połowa długości przedziału ufności)

$$\Delta_{\max} = t\left(1 - \frac{1}{2}\alpha, n-1\right)\frac{S}{\sqrt{n-1}}$$

Zaleca się pobrać z populacji próbę wstępną o liczebności n_0. Następnie oblicza się dla próby wstępnej średnią arytmetyczną $\overline{x}_0 = \frac{1}{n_0}\sum_{i=1}^{n_0} x_i$, odchylenie standardowe $S_0 = \sqrt{\frac{1}{n_0}\sum_{i=1}^{n_0}\left(x_i - \overline{x}_0\right)^2}$ oraz kwantyl $t\left(1-\frac{1}{2}\alpha, n_0 - 1\right)$. Wyznacza się wartość m ze wzoru

$$m = \left(\frac{2\cdot t\left(1-\frac{1}{2}\alpha, n_0-1\right)\cdot S_0}{d}\right)^2 + 1$$

Jeżeli

a) $m \le n_0$, to $n_{\min} = n_0$.

b) $m > n_0$ - należy dobrać losowo $n_1 = [m] + 1 - n_0$ elementów do próby, wówczas $n_{\min} = n_0 + n_1$. W tej sytuacji przedział ufności o długości nieprzekraczającej d ma postać

$$\left[\overline{x} - t\left(1-\frac{1}{2}\alpha, n_0-1\right)\cdot\frac{S_0}{\sqrt{n_0+n_1-1}}, \overline{x} + t\left(1-\frac{1}{2}\alpha, n_0-1\right)\cdot\frac{S_0}{\sqrt{n_0+n_1-1}}\right],$$

gdzie $\overline{x} = \frac{1}{n_0+n_1}\sum_{i=1}^{n_0+n_1} x_i$

Model III dla wartości oczekiwanej – postępujemy analogicznie jak w modelu I (wymagana jest duża liczebność próby) – w miejsce σ wstawiamy S^*.

Model II dla wskaźnika struktury

Mamy

$$n \geq \left(\frac{2u\left(1-\frac{1}{2}\alpha\right)\cdot\sqrt{\hat{p}\left(1-\hat{p}\right)}}{d} \right)^2$$

W przypadku braku przesłanek, jaką wartość może mieć $\hat{p}$, przyjmuje się najbardziej „rygorystyczny" wariant, czyli $\hat{p}=0.5$. Otrzymujemy

$$n_{\min} = \left[\left(\frac{u\left(1-\frac{1}{2}\alpha\right)}{d} \right)^2 \right] + 1$$

Wymagana liczebność próby nie mniejsza niż 100.

Szczegóły są dostępne w skrypcie: Jolanta Borowska, *Wnioskowanie statystyczne z programem Maple. Część I*, Częstochowa 2021