

Odhad priemernej spotreby paliva v automobile

Výrobca automobilu udáva, že priemerná spotreba paliva so 40% podielom diaľnice je 12,5 l/100km.

Testovací jazdec podrobil 14 automobilov meraniu spotreby v rôznych podmienkach.

Nameraná spotreba je v nasledovnej tabuľke:

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
l/100 km	12,8	13,5	14,2	13,6	14,1	14,5	13,6	13,9	14,3	15,1	13,7	13,4	13,9	14,2

1. Bodový odhad strednej hodnoty

Nevychýleným odhadom strednej hodnoty je aritmetický priemer:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{194,8}{14} = 13,91429 \text{ (l/100km)}$$

2. Bodový odhad rozptylu (disperzie)

Nevychýleným odhadom rozptylu je výberový rozptyl (disperzia)

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{4,017}{13} = 0,309$$

Výberová smerodajná odchýlka – odmocnina z rozptylu

$$S_x = \sqrt{S_x^2} = 0,55588$$

3. Intervalový odhad strednej hodnoty

Keďže poznáme iba odhad rozptylu základného súboru, použijeme vzťah

$$P(\bar{x} - \delta_x \leq \mu_x \leq \bar{x} + \delta_x) = 1 - \alpha$$

$$\text{kde} \quad \delta_x = \frac{S_x}{\sqrt{N}} t_{1-\frac{\alpha}{2}}$$

$t_{1-\frac{\alpha}{2}}$ je kvantil Studentovho rozdelenia s N-1 stupňami voľnosti.

Pre hladinu významnosti $\alpha=0,05$ $t_{1-\frac{\alpha}{2}} = 2,160369$

$$\delta_x = \frac{S_x}{\sqrt{N}} t_{1-\frac{\alpha}{2}} = \frac{\sqrt{0,309}}{\sqrt{14}} 2,16 = 0,320954$$

S pravdepodobnosťou 95% sa skutočná priemerná spotreba automobilu bude pohybovať v intervale

$$\langle \bar{x} - \delta_x; \bar{x} + \delta_x \rangle = \langle 13,59333 ; 14,23524 \rangle$$

Čo vplýva na veľkosť intervalu?

- **Hladina významnosti α (hodnota kvantilu)**

pre $\alpha=0,1$ $\langle \bar{x} - \delta_x; \bar{x} + \delta_x \rangle = \langle 13,65118 ; 14,17739 \rangle$

pre $\alpha=0,05$ $\langle \bar{x} - \delta_x; \bar{x} + \delta_x \rangle = \langle 13,59333 ; 14,23524 \rangle$

pre $\alpha=0,01$ $\langle \bar{x} - \delta_x; \bar{x} + \delta_x \rangle = \langle 13,46676 ; 14,36181 \rangle$

Nižšie α znamená širší interval – pri nižšej hodnote α sa musí do intervalu „vojst“ viac hodnôt. Hodnota α korešponduje s percentom hodnôt mimo intervalu.

- **Rozptyl nameraných údajov S_x^2**

Vyšší rozptyl znamená väčšiu šírku intervalu – čím je menší rozptyl v údajoch, tým presnejšie môžeme odhadovať strednú hodnotu.

- **Počet nameraných údajov N**

S rastúcim N sa interval zužuje – máme viac údajov a teda presnejšiu informáciu.

4. Testovanie hypotézy (t-test)

Testujeme hypotézu, že priemerná spotreba paliva so 40% podielom diaľnice je 12,5 l/100km, ako udáva výrobca.

Nulová hypotéza: $H_0: \mu = \mu_0$, kde $\mu_0 = 12,5$ (l/100km).

Nulovú hypotézu H_0 prijímame, ak bude platiť nerovnosť $|t| \leq t_{1-\frac{\alpha}{2}}$,

kde $t = \frac{\bar{x} - \mu_x}{S_x} \sqrt{N} \in \text{St}(N-1)$.

Z údajov výberového súboru vypočítame $t = \frac{13,91429 - 12,5}{0,5558} \sqrt{14} = 9,52103$.

Keďže $t_{1-\frac{\alpha}{2}} = 2,160369 \Rightarrow \boxed{|t| > t_{1-\frac{\alpha}{2}}}$

Na hladine významnosti 5% zamietame hypotézu o priemernej spotrebe 12,5 l/100km.

MATLAB:

funkcia `ttest` (Statistics and Machine Learning Toolbox)

`h=ttest(x)` Returns a test decision for the null hypothesis that the data in `x` comes from a normal distribution with mean equal to zero and unknown variance.



`h = 1` The result `h` is 1 if the test rejects the null hypothesis at the 5% significance level, and 0 otherwise.

Na hladine významnosti 5% zamietame hypotézu o nulovej strednej hodnote.

`h=ttest(x, 12.5)` *Returns a test decision for the null hypothesis that the data in x comes from a normal distribution with mean 12.5 and unknown variance.*



`h = 1`

Na hladine významnosti 5% zamietame hypotézu o priemernej spotrebe 12,5 l/100km.

`h=ttest(x, 13.8)`



`h = 0`

Na hladine významnosti 5% prijímame hypotézu o priemernej spotrebe 13,8 l/100km.