Intervalový odhad - příklady

29. října 2013

1 Limonády

Na výstupu výrobní linky jsme odebrali dvacet náhodných limonád a zjistili jsme jejich objem. Určete 99% intervalový odhad pro jejich průměrný objem.
Data (v mililitrech): 223, 198, 212, 235, 202, 176, 193, 207, 201, 213, 189, 193, 215, 204, 199, 205, 224, 203, 187, 204.

Jedná se o odhad střední hodnoty, neznáme σ^2. Použijeme tedy vzorec:

$$\mu \in \overline{x} \pm \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}$$

$$t \sim St\,(n-1)\,.$$

Tento vzorec je tím přesnější, čím více dat máme. (Jeho platnost vychází z Centrální limitní věty. Při menším počtu dat by záleželo na konkrétním rozdělení průměrované veličiny.) Počet hodnot $n = 20$ považujeme za dostatečně velký, abychom jej mohli použít.

$\alpha = 0,01$, hledáme tedy kvantil $t_{0{,}995}\,(19)$. Ten najdeme v tabulkách:

Tabulka VI. **Kvantily rozdělení *t***

ν	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	1,638	2,353	3,182	4,541	5,841
4	1,553	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750

Spočteme střední hodnotu, rozptyl a směrodatnou odchylku výběrového souboru:

$$\begin{aligned} \overline{x} &= \frac{\sum x_i}{n} = 204,15, \\ s^2 &= \frac{\sum (x_i - \overline{x})^2}{n-1} = 190,13, \\ s &= \sqrt{s^2} = 13,79. \end{aligned}$$

Dosadíme do vzorce:

$$\mu \in \overline{x} \pm \frac{s}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}} = 204,15 \pm \frac{13,79}{\sqrt{20}} \cdot 2,861 = \langle 195, 213 \rangle .$$

99% interval spolehlivosti pro střední hodnotu objemu limonád je 195 až 213 mililitrů.

2 Volby

Přišel za mnou nejmenovaný politik, že by rád řekl svým spolustraníkům, že s 90% jistotou očekává lepší volební výsledek než x%. A jestli bych mu uměl ono x odhadnout. Vysvětlil jsem mu problémy kolem výběru kvalitního výběrového souboru, upozornil na to, že samotné publikované odhady razantně ovlivňují výsledek voleb a také, voliči že v odhadech kecají. Nakonec jsem provedl fundovaný odhad s těmito výsledky:

Počet respondentů: 458.
Hlasů pro onu stranu: 94.

Jedná se o odhad podílu a to o levostranný odhad. Použijeme proto vzorec:

$$\pi \in \left\langle p - \sqrt{\frac{p\,(1-p)}{n}} \cdot z_{1-\alpha},\ 1 \right\rangle,$$

$$z \sim N\,(0,1)\,.$$

Vzorec opět vychází z Centrální limitní věty. Dat je již dost, abysme jej s klidem mohli použít.

Výběrový podíl je $p = \frac{94}{458} = 0,205$.

$\alpha = 0,1$, hledáme tedy kvantil $z_{0,9}$:

Tabulka III.

Distribuční funkce normálního rozdělení

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dy$$

U	Φ(u)	u	Φ(u)	u	Φ(u)	u	Φ(u)
0,00	0,50000	0,40	0,65542	0,80	0,78814	1,20	0,88493
0,01	,50399	0,41	,65910	0,81	,79103	1,21	,88686
0,02	,50798	0,42	,66276	0,82	,79389	1,22	,88877
0,03	,51197	0,43	,66640	0,83	,79673	1,23	,89065
0,04	,51595	0,44	,67003	0,84	,79955	1,24	,89251
0,05	,51994	0,45	,67364	0,85	,80234	1,25	,89435
0,06	,52392	0,46	,67724	0,86	,80511	1,26	,89617
0,07	,52790	0,47	,68082	0,87	,80785	1,27	,89796
0,08	,53188	0,48	,68439	0,88	,81057	1,28	,89973
0,09	,53586	0,49	,68793	0,89	,81327	1,29	,90147
0,10	,53983	0,50	,69146	0,90	,81594	1,30	,90320
0,11	,54380	0,51	,69497	0,91	,81859	1,31	,90490
0,12	,54776	0,52	,69847	0,92	,82121	1,32	,90658
0,13	,55172	0,53	,70194	0,93	,82381	1,33	,90824
0,14	,55567	0,54	,70540	0,94	,82639	1,34	,90988
0,15	,55962	0,55	,70884	0,95	,82894	1,35	,91149
0,16	,56356	0,56	,71226	0,96	,83147	1,36	,91309
0,17	,56749	0,57	,71566	0,97	,83398	1,37	,91466
0,18	,57142	0,58	,71904	0,98	,83646	1,38	,91621

Dosadíme do vzorce:

$$\begin{aligned} \pi \quad \in \quad & \left\langle p - \sqrt{\frac{p\,(1-p)}{n}} \cdot z_{1-\alpha},\ 1 \right\rangle \\ & \left\langle 0,205 - \sqrt{\frac{0,205\,(1-0,205)}{458}} \cdot 1,28,\ 1 \right\rangle \\ & \langle 0,1809;\ 1 \rangle\,. \end{aligned}$$

S poznámkou, že mnohé události mohou poslat volební výsledek zcela jinam, jsem mu poslal výsledek výpočtu, totiž, že jeho strana na 90% získá přes 18% hlasů.

3 Rozbitý stroj

Stroj lisuje trubky různého poloměru dle nastavení. Trubky stejného typu by však neměly mít rozptyl poloměrů větší než $13\,mm^2$. Pokud je rozptyl na $13\,mm^2$, je třeba stroj seřídit. My jsme zjistili na 28 kusech rozptyl $15\,mm^2$. Obsahuje 90% levostranný interval spolehlivosti ještě tolerovanou hodnotu $13\,mm^2$, nebo už je to signál k seřízení?

Jedná se o levostranný IS pro rozptyl, použijeme tedy vzorec:

$$\sigma^2 \quad \in \quad \left\langle \frac{(n-1)\,s^2}{\chi^2_{1-\alpha}}, \infty \right),$$

$$\chi^2 \sim Ch\acute{\imath}^2\,(n-1)\,.$$

Vzorec opět vychází z CLV, dat je dostatek pro jeho použití.

Z tabulky zjistíme kvantil $\chi^2_{0,9}\,(27)$:

Tabulka V pokračování

Kvantily rozdělení χ^2

v \ P	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
1	2,71	3,84	5,02	6,63	7,88	10,8	12,1
2	4,61	5,99	7,38	9,21	10,6	13,8	15,2
3	6,25	7,81	9,35	11,3	12,8	16,3	17,7
4	7,78	9,49	11,1	13,3	14,9	18,5	20,0
5	9,24	11,1	12,8	15,1	16,7	20,5	22,1
6	10,6	12,6	14,4	16,8	18,5	22,5	24,1
7	12,0	14,1	16,0	18,5	20,3	24,3	26,0
8	13,4	15,5	17,5	20,1	22,0	26,1	27,9
9	14,7	16,9	19,0	21,7	23,6	27,9	29,7
10	16,0	18,3	20,5	23,2	25,2	29,6	31,4
11	17,3	19,7	21,9	24,7	26,8	31,3	33,1
12	18,5	21,0	23,3	26,2	28,3	32,9	34,8
13	19,8	22,4	24,7	27,7	29,8	34,5	36,5
14	21,0	23,7	26,1	29,1	31,3	36,1	38,1
15	22,3	25,0	27,5	30,6	32,8	37,7	39,7
16	23,5	26,3	28,8	32,0	34,3	39,3	41,3
17	24,8	27,6	30,2	33,4	35,7	40,8	42,9
18	26,0	28,9	31,5	34,8	37,2	42,3	44,4
19	27,2	30,1	32,9	36,2	38,6	43,8	46,0
20	28,4	31,4	34,2	37,6	40,0	452	47,5
21	29,6	32,7	35,5	38,9	41,4	46,8	49,0
22	30,9	33,9	36,8	40,3	42,8	48,3	50,5
23	32,0	35,2	38,1	41,6	44,2	49,7	52,0
24	33,2	36,4	39,4	43,0	45,6	51,2	53,5
25	34,4	37,7	40,6	44,3	46,9	52,6	54,9
26	35,6	38,9	41,9	45,6	48,3	54,1	56,4
27	36,7	40,1	43,2	47,0	49,6	55,5	57,9
28	37,9	41,3	44,5	48,3	51,0	56,9	59,3
29	39,1	42,6	45,7	49,6	52,3	58,3	60,7
30	40,3	43,8	47,0	50,9	53,7	59,7	62,2

Dosadíme do vzorce:

$$\sigma^2 \quad \in \quad \left\langle \frac{(n-1)\,s^2}{\chi^2_{1-\alpha}}, \infty \right)$$

$$\left\langle \frac{(28-1)\cdot 15}{36,7}, \infty \right)$$
$$\langle 11, \infty)\,.$$

Interval (s rezervou) obsahuje tolerovanou hodnotu $13\,mm^2$, proto není na základě dat důvod přistoupit k seřízení stroje. Data však mohou být důvodem k obsáhlejšímu měření, abychom mohli určit rozptyl přesněji.

4 Vlasy

U stovky náhodně vybraných lidí jsem spočítal vlasy. Získal jsem tuto výběrovou střední hodnotu a výběrový rozptyl: $\mu = 111\,425$, $\sigma^2 = 1\,635\,857\,411$. Určete 95% interval spolehlivosti pro průměrný počet vlasů.

(103 498, 119 352)

5 Rozptyl

Odhadněte 95% pravostranný interval spolehlivosti pro rozptyl, pokud jsme na 15 datech naměřili výběrový rozptyl $s^2 = 147$.

(0, 313)

6 Nepoctivá kostka

Při 100 hodech kostkou jsme hodili průměr 4,2. Je možné, že kostka je přesto poctivá, tedy že její střední hodnota je 3,5? Vytvořte 99% levostranný interval spolehlivosti pro střední hodnotu. (Když kostka není poctivá, neznám ani rozptyl.) Výběrový rozptyl jsme naměřili 3,8.

Získáme interval $\langle 3,75, \infty)$, do kterého hodnota 3,5 již nepatří. Můžeme tedy tvrdit, že kostka není poctivá a to s jistotou (minimálně) 99%. Tedy pravděpodobnost, že kostka je poctivá, je maximálně 1%.