

Kolik je třeba mít pozorování, aby aritmetický průměr byl dobrým odhadem střední hodnoty?

a) z úvahy nerovnosti: $P(|\bar{X} - \mu| \leq \varepsilon) \geq 1 - \frac{\text{var } \bar{X}}{\varepsilon^2}$

$\text{var } \bar{X} = \frac{\sigma^2}{n}$, $\varepsilon > 0$ si zvolím

chci, aby ta pravděpodobnost byla alespoň $1 - \alpha = 0,95$

hodily bylo máti: $\sigma^2 = 1$, $\varepsilon = \frac{1}{2}$, pak:

$$P(|\bar{X} - \mu| \leq \frac{1}{2}) \geq 1 - \frac{\frac{1}{n}}{\frac{1}{4}} = 1 - \frac{4}{n} \geq 0,95$$

$$0,05 \geq \frac{4}{n}$$

$$n \geq \frac{4}{0,05} = 80$$

obecně: $1 - \frac{\text{var } \bar{X}}{\varepsilon^2} = 1 - \frac{\sigma^2}{n\varepsilon^2} \geq 1 - \alpha$

$$n \geq \frac{\sigma^2}{\alpha \varepsilon^2}$$

b) z intervalu spolehlivosti pro μ : $(\bar{X} - \mu_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{X} + \mu_{1-\alpha/2} \frac{\sigma}{\sqrt{n}})$

interval má délku $\bar{I} = 2 \mu_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$, z toho $n = \left(2 \mu_{1-\alpha/2} \frac{\sigma}{\bar{I}}\right)^2$

zvolím si, jaká délka intervalu je pro mě přijatelná, a z toho dopočítám n

máti. pro $\alpha = 0,05$; $\sigma^2 = 1$, $\bar{I} \leq 1$: $n \geq \left(2 \mu_{0,975}\right)^2 = (2 \cdot 1,96)^2 = 15,36 \approx 16$

Použití interval spolehlivosti je tedy výhodnější.

čo kedy ale nemám rozplyt?

Gladovím si dĺžku intervalu v násobkoch smýslodajnej odchýlky:

$$\bar{I} \leq k \cdot \sigma ; \text{ dostaneme } n \geq \left(\frac{2}{k} u_{1-\alpha/2} \right)^2$$

napr. pre $k=2$, tj. $\mu = \bar{X} \pm \sigma$, dostaneme $n \geq (u_{1-\alpha/2})^2$

pre $\alpha = 0,05$ $n \geq (1,96)^2 = 4$

pre $k=1$, tj. $\mu = \bar{X} \pm \frac{\sigma}{2}$, $n \geq (2 u_{1-\alpha/2})^2 = (2 \cdot 1,96)^2 = 16$

Existujú na to tabuľky, kde sa pre zvolenú k a $1-\alpha$ priamo
najde hodnota n .

TAB. 8 ROZSAHY VÝBĚRŮ PŘI PŘEDEPSANÉ DÉLCE INTERVALU SPOLEHLIVOSTI PRO μ

k	1 - α							k
	0,9	0,95	0,99	0,995	0,999	0,9995	0,9999	
0,03	12025	17074	29489	35020	48123	53848	67275	0,03
0,04	6764	9604	16588	19699	27069	30290	37842	0,04
0,05	4329	6147	10616	12608	17325	19386	24219	0,05
0,06	3007	4269	7373	8755	12031	13462	16819	0,06
0,07	2209	3136	5417	6433	8839	9891	12357	0,07
0,08	1691	2401	4147	4925	6768	7573	9461	0,08
0,09	1337	1898	3277	3892	5347	5984	7475	0,09
0,10	1083	1537	2654	3152	4332	4847	6055	0,10
0,11	895	1270	2194	2605	3580	4006	5004	0,11
0,12	752	1068	1844	2189	3008	3366	4205	0,12
0,13	641	910	1571	1865	2563	2868	3583	0,13
0,14	553	784	1355	1609	2210	2473	3090	0,14
0,15	481	683	1180	1401	1925	2154	2691	0,15
0,16	423	601	1037	1232	1692	1894	2366	0,16
0,17	375	532	919	1091	1499	1677	2096	0,17
0,18	335	475	820	973	1337	1496	1869	0,18
0,19	300	426	736	874	1200	1343	1678	0,19
0,20	271	385	664	788	1083	1212	1514	0,20
0,21	246	349	602	715	983	1099	1373	0,21
0,22	224	318	549	652	895	1002	1251	0,22
0,23	205	291	502	596	819	917	1145	0,23
0,24	188	267	461	548	752	842	1052	0,24
0,25	174	246	425	505	693	776	969	0,25
0,26	161	228	393	467	641	717	896	0,26
0,27	149	211	365	433	595	665	831	0,27
0,28	139	196	339	403	553	619	773	0,28
0,29	129	183	316	375	515	577	720	0,29
0,30	121	171	295	351	482	539	673	0,30
0,31	113	160	277	328	451	505	631	0,31
0,32	106	151	260	308	423	474	592	0,32
0,33	100	142	244	290	398	446	556	0,33
0,34	94	133	230	273	375	420	524	0,34
0,35	89	126	217	258	354	396	495	0,35
0,36	84	119	205	244	335	374	468	0,36
0,37	80	113	194	231	317	355	443	0,37
0,38	75	107	184	219	300	336	420	0,38
0,39	72	102	175	208	285	319	399	0,39
0,40	68	97	166	197	271	303	379	0,40
0,41	65	92	158	188	258	289	361	0,41
0,42	62	88	151	179	246	275	344	0,42
0,43	59	84	144	171	235	263	328	0,43
0,44	56	80	138	163	224	251	313	0,44
0,45	54	76	132	156	214	240	299	0,45
0,46	52	73	126	149	205	230	287	0,46
0,47	49	70	121	143	197	220	275	0,47
0,48	47	67	116	137	188	211	263	0,48
0,49	46	64	111	132	181	202	253	0,49
0,50	44	62	107	127	174	194	243	0,50
0,51	42	60	103	122	167	187	233	0,51
0,52	41	57	99	117	161	180	224	0,52

TAB. 8 ROZSAHY VÝBĚRŮ PŘI PŘEDEPSANÉ DÉLCE INTERVALU SPOLEHLIVOSTI PRO μ

TAB

k	1 - α							k
	0,9	0,95	0,99	0,995	0,999	0,9995	0,9999	
0,53	39	55	95	113	155	173	216	0,53
0,54	38	53	92	109	149	167	208	0,54
0,55	36	51	88	105	144	161	201	0,55
0,56	35	49	85	101	139	155	194	0,56
0,57	34	48	82	98	134	150	187	0,57
0,58	33	46	79	94	129	145	180	0,58
0,59	32	45	77	91	125	140	174	0,59
0,60	31	43	74	88	121	135	169	0,60
0,61	30	42	72	85	117	131	163	0,61
0,62	29	40	70	82	113	127	158	0,62
0,63	28	39	67	80	110	123	153	0,63
0,64	27	38	65	77	106	119	148	0,64
0,65	26	37	63	75	103	115	144	0,65
0,66	25	36	61	73	100	112	139	0,66
0,67	25	35	60	71	97	108	135	0,67
0,68	24	34	58	69	94	105	131	0,68
0,69	23	33	56	67	91	102	128	0,69
0,70	23	32	55	65	89	99	124	0,70
0,71	22	31	53	63	86	97	121	0,71
0,72	21	30	52	61	84	94	117	0,72
0,73	21	29	50	60	82	91	114	0,73
0,74	20	29	49	58	80	89	111	0,74
0,75	20	28	48	57	77	87	108	0,75
0,76	19	27	46	55	75	84	105	0,76
0,77	19	26	45	54	74	82	103	0,77
0,78	18	26	44	52	72	80	100	0,78
0,79	18	25	43	51	70	78	98	0,79
0,80	17	25	42	50	68	76	95	0,80
0,81	17	24	41	49	67	74	93	0,81
0,82	17	23	40	47	65	73	91	0,82
0,83	16	23	39	46	63	71	88	0,83
0,84	16	22	38	45	62	69	86	0,84
0,85	15	22	37	44	60	68	84	0,85
0,86	15	21	36	43	59	66	82	0,86
0,87	15	21	36	42	58	65	80	0,87
0,88	14	20	35	41	56	63	79	0,88
0,89	14	20	34	40	55	62	77	0,89
0,90	14	19	33	39	54	60	75	0,90
0,91	14	19	33	39	53	59	74	0,91
0,92	13	19	32	38	52	58	72	0,92
0,93	13	18	31	37	51	57	71	0,93
0,94	13	18	31	36	50	55	69	0,94
0,95	12	18	30	35	48	54	68	0,95
0,96	12	17	29	35	47	53	66	0,96
0,97	12	17	29	34	47	52	65	0,97
0,98	12	16	28	33	46	51	64	0,98
0,99	12	16	28	33	45	50	62	0,99
1,00	11	16	27	32	44	49	61	1,00
1,10	9	13	22	27	36	41	51	1,10
1,20	8	11	19	22	31	34	43	1,20

TAB. 8 ROZSAHY VÝBĚRŮ PŘI PŘEDEPSANÉ DÉLCE INTERVALU SPOLEHLIVOSTI PRO μ

k	1 - α							k
	0,9	0,95	0,99	0,995	0,999	0,9995	0,9999	
1,30	7	10	16	19	26	29	36	1,30
1,40	6	8	14	17	23	25	31	1,40
1,50	5	7	12	15	20	22	27	1,50
1,60	5	7	11	13	17	19	24	1,60
1,70	4	6	10	11	15	17	21	1,70
1,80	4	5	9	10	14	15	19	1,80
1,90	3	5	8	9	12	14	17	1,90
2,00	3	4	7	8	11	13	16	2,00
2,10	3	4	7	8	10	11	14	2,10
2,20	3	4	6	7	9	11	13	2,20
2,30	3	3	6	6	9	10	12	2,30
2,40	2	3	5	6	8	9	11	2,40
2,50	2	3	5	6	7	8	10	2,50
2,60	2	3	4	5	7	8	9	2,60
2,70	2	3	4	5	6	7	9	2,70
2,80	2	2	4	5	6	7	8	2,80
2,90	2	2	4	4	6	6	8	2,90
3,00	2	2	3	4	5	6	7	3,00
3,10	2	2	3	4	5	6	7	3,10
3,20	2	2	3	4	5	5	6	3,20
3,30	1	2	3	3	4	5	6	3,30
3,40	1	2	3	3	4	5	6	3,40
3,50	1	2	3	3	4	4	5	3,50
3,60	1	2	3	3	4	4	5	3,60
3,70	1	2	2	3	4	4	5	3,70
3,80	1	2	2	3	3	4	5	3,80
3,90	1	2	2	3	3	4	4	3,90
4,00	1	1	2	2	3	4	4	4,00
4,10	1	1	2	2	3	3	4	4,10
4,20	1	1	2	2	3	3	4	4,20
4,30	1	1	2	2	3	3	4	4,30
4,40	1	1	2	2	3	3	4	4,40
4,50	1	1	2	2	3	3	3	4,50
4,60	1	1	2	2	3	3	3	4,60
4,70	1	1	2	2	2	3	3	4,70
4,80	1	1	2	2	2	3	3	4,80
4,90	1	1	2	2	2	3	3	4,90
5,00	1	1	2	2	2	2	3	5,00

1 000 000 prvků (MN, M1, M3)

I-čas	P-čas	Čas	M1	Výměny	Porovnání	Q-čas	I-čas	P-čas	Čas	M3	Výměny	Porovnání
1.19	0.00	7.95	1.	4100362	23391548	6.84	1.17	0.00	8.01	1.	4094311	19850507
1.18	0.00	8.06	2.	4109733	23693092	6.94	1.18	0.00	8.12	2.	4126967	19089632
1.18	0.00	8.29	3.	4114931	22602791	6.77	1.19	0.00	7.96	3.	4098052	19640826
1.19	0.00	8.31	4.	4051794	24107207	7.03	1.18	0.00	8.21	4.	4093828	20067844
1.19	0.00	8.06	5.	4083053	23268242	6.88	1.19	0.00	8.07	5.	4064715	20258908
1.18	0.00	8.05	6.	4054783	24534597	7.10	1.18	0.00	8.28	6.	4108642	19373468
1.18	0.00	7.98	7.	4106845	22960107	6.84	1.17	0.00	8.01	7.	4090494	19629649
1.18	0.00	8.06	8.	4103625	22791147	6.78	1.17	0.00	7.95	8.	4082225	19802586
1.19	0.00	8.13	9.	4069077	23652237	6.87	1.18	0.00	8.05	9.	4100272	19489754
1.17	0.00	8.08	10.	4087163	23237522	6.77	1.18	0.00	7.95	10.	4115292	19326990
1.19	0.00	8.44	11.	4113420	23055301	6.82	1.18	0.00	8.00	11.	4105098	19565539
1.18	0.00	8.06	12.	4068719	24453598	7.07	1.19	0.00	8.26	12.	4111603	19424510
1.18	0.00	8.11	13.	4137795	22489281	6.78	1.18	0.00	7.96	13.	4108325	19427871
1.17	0.00	8.08	14.	4118836	22799494	6.79	1.18	0.00	7.97	14.	4107414	19406223
1.18	0.00	8.20	15.	4120516	22832905	6.79	1.18	0.00	7.97	15.	4082516	19754591
1.18	0.00	8.05	16.	4115326	22636947	6.75	1.19	0.00	7.94	16.	4113465	19339860
1.18	0.00	8.27	17.	4124258	22927205	6.78	1.18	0.00	7.96	17.	4110898	19387273
1.18	0.00	8.14	18.	4106038	22846174	6.80	1.18	0.00	7.98	18.	4115370	19328065
1.18	0.00	8.11	19.	4060726	24266218	7.00	1.18	0.00	8.18	19.	4103859	19486129
1.17	0.00	8.05	20.	4088484	23539988	6.93	1.17	0.00	8.10	20.	4082030	20203015
1.18	0.00	8.02	21.	4125048	22669153	6.83	1.19	0.00	8.02	21.	4099764	19585975
1.19	0.00	8.11	22.	4091661	23873178	6.93	1.18	0.00	8.11	22.	4111102	19445783
1.18	0.00	8.05	23.	4109483	22916798	6.84	1.19	0.00	8.03	23.	4096939	19678438
1.18	0.00	8.30	24.	4129074	22315426	6.63	1.17	0.00	7.80	24.	4114048	19278849
1.18	0.00	8.17	25.	4071611	23853352	7.03	1.18	0.00	8.21	25.	4103438	19616061
1.18	0.00	8.13	Ø	4098494	23268540	6.86	1.18	0.00	8.04	Ø	4101627	19578334

Jinou možností je selveninská odhad.

Je dvoustupňový:

1. Nejprve neseeme výběr $x_1, \dots, x_m$ předem stanoveného ^{počet} rozsahu m (tj. provedu m měření), a něj vypočtu $\bar{X}$ a S^2 .

2. Vypočtu n jako nejmenší celé číslo $\geq m$ splňující vztah

$$\frac{S^2}{n} \leq \frac{c^2}{b^2} \quad (\text{neboli } n \geq \frac{S^2 b^2}{c^2}),$$

kde b^2 je rozptyl rozdělení t_{m-1} (tj. $b^2 = \frac{m-1}{m-3}$ pro $m > 3$),

$$c = \frac{b l}{t_{m-1, 1-\alpha/2}}, \quad \text{kde } l \text{ je polovina pořádkové délky intervalu spolehlivosti pro } \mu.$$

Ukulečným datům m -m pozorování (experimentů).

Ka odhad μ se pak bere aritmetický průměr a celého výběru,

$$\text{tj. } \bar{X} = \frac{1}{m} \sum_{i=1}^m x_i.$$

Díky korekčnosti této metody je poměrně komplikovanější, protože se musí počítat s tím, že m je náhodná veličina!

Literatura: např. skript M. Kušera: Selveninská analýza

(SPN Praha, 1982) - je to trochu moc teoretické

nebo R.C. Rao: Lineární metody statistické indukce a jejich aplikace (Academia Praha, 1978)

na <http://www.karlin.mff.cuni.cz/~hlavka/vyuka/planex/>

kusera_seq.pdf

Příklad: quicksort, varianta M1

provedeme pouze první $m = 10$ měření

(8,01; 8,12; 7,96; 8,21; 8,07; 8,28; 8,01; 7,95; 8,05; 7,95)

K toho $\bar{x} = 8,061$

$$s = \sqrt{\frac{1}{9} (\sum x_i^2 - 10 \cdot \bar{x}^2)} = \sqrt{\frac{1}{9} (649,9111 - 649,79721)} = 0,1125 \%$$

$$\text{rozptyl } t_{m-1} : \frac{m-1}{m-3} = \frac{9}{7} = 1,2857 = b^2$$

$$b = 1,1339$$

budu požadovat délku intervalu $2l = 0,095208$ na hladině 0,95% %.

$$t_{m-1, 1-\alpha/2} = t_{9, 0,975} = 2,2622$$

$$c = \frac{bl}{t} = 0,02386, \quad c^2 = 0,000569$$

$$\text{hledám } m \text{ tak, aby } \frac{s^2}{m} \leq \frac{c^2}{b^2}, \quad \text{či } m \geq s^2 \cdot \frac{b^2}{c^2} = 28,594$$

$$\text{Přesnější výpočet: } s^2 = \frac{1}{9} (\sum x_i^2 - 10 \bar{x}^2)$$

$$s^2 b^2 = s^2 \cdot \frac{9}{7} = \frac{1}{7} (\sum x_i^2 - 10 \bar{x}^2) = 0,01627$$

$$m \geq 28,594$$

Nějaké rozvahování když se více pýtlujeme, ale se
nřinamně.

Závěr: Gelvenčí analýza mi radí provést ještě dalších 19 měření.

(Ne důležitosti jsem tuto požadovanou délku dosáhl
o 25 měřeními.)

$$\alpha = 0,05$$

$$t_{q, 0,975} = 2,2622$$

www.mironet.cz

$$\begin{aligned} \% \text{ interval spolehlivosti } & \left(\bar{X} - t_{m-1, 1-\alpha/2} \frac{s}{\sqrt{m}} ; \bar{X} + t_{m-1, 1-\alpha/2} \frac{s}{\sqrt{m}} \right) \\ & \left(8,061 - 2,2622 \cdot \frac{0,1125}{\sqrt{10}} ; 8,061 + 2,2622 \cdot \frac{0,1125}{\sqrt{10}} \right) \end{aligned}$$

$$\left(8,061 - 0,0805 ; 8,061 + 0,0805 \right)$$

$$\left(7,9805 ; 8,1415 \right)$$

ma' dellu 0,161

%.%. tj. dellu, kterou ma'm x provedených 25 měření

Jedí ryjdu 12 lich ryjch 25 měření. Máme $\bar{x} = 8,044$, $s = 0,115$.

$$\text{rozptyl } t_{24} : \frac{24}{22} = 1,09 = b^2, \quad b = 1,04446$$

budu chtít interval délky $2l = 0,0545$ (což je polovina současné směrodatné odchylky)

$$\text{kvantil } t_{24; 0,975} = 2,0639$$

$$e = \frac{bl}{t} = 0,014549, \quad e^2 = 0,00021168$$

$$n \geq s^2 \frac{b^2}{e^2} = 68,156$$

Budu potřebovat celkem 69 měření, takže musím provést ještě 44.

Co kdybych chtěl interval délky $2l = 0,095208$ (tj. přesně to, co mám)?

$$e = 0,02409, \quad e^2 = 0,0005803$$

$$n \geq 24,8614$$

Takže to je úplně přesné.

Metoda relvenční analýzy by měla být přesnější, protože:

- 1) už máme nějakou představu o rozptylu
- 2) počítá se s kvantily t -rozdělení, nikoli normálního

Při normálním rozptylu je kolik délka intervalu spolehlivosti

$$\bar{I} = 2 t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}, \quad \text{neboli } n = \left(2 t_{n-1, 1-\alpha/2} \frac{s}{\bar{I}} \right)^2. \quad \text{Potom ale}$$

n předem neznám, nemůžeme bylo kvantily použít a aproximují je normálními, které jsou na n nerovnosti (ale pomaleji jsou o něco méně).

Jiste k předchozímu příkladu: Přes rybu k tělu ryby 25 měřené

a tedy celist interval délky $2L = 0,190416$ (tj. větší, než uř mám).

Pak. $c = 0,04818$, $c^2 = 0,00232$

$$n \gg 6,2185$$

Výsledkem tedy je, že na tuto porovnanou délku uř mám
měřené uř uř dost.