Оценяване на параметри

Точкови оценки

Параметърът е числова характеристика на популацията

Точковата оценка е едно число (стойност), което се използва, за да се оцени неизвестната стойност на параметъра.

Зависи от извадката и има характер на случайна величина

Параметър	Точкова оценка
Средна стойност μ	*Извадково средно* $\overline{x}$
Дисперсия σ^2	*Извадкова дисперсия* s^2
Вероятност p	*Извадкова пропорция* $\hat{p}$

Неизместеност на точкова оценка

Средната стойност на статистиката= параметъра

Пример 1. Извадковото средно $\overline{X} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$

Нека случайната извадка с обем n е от популация със средна стойност μ

$$E\overline{X} = \frac{EX_1 + EX_2 + \ldots EX_n}{n} = \frac{\mu + \mu + \ldots + \mu}{n} = \mu$$

$$EX_i = \mu$$

$$E\overline{X} = \mu$$

Извадковото средно е **неизместена оценка** на параметъра "*популационна средна стойност* μ"

Пример 2. Извадкова пропорция $\hat{p}$

Извадка с обем n е направена от алтернативна популация.

$$E\hat{p} = \frac{EX}{n} = p$$

Извадковата пропорция е **неизместена оценка** на параметъра "*вероятност p*"

Доверителни интервали

Интервал, построен на основата на извадката, който съдържа неизвестния параметър с вероятност близка до 1

Пример:

Време на полет : 11 часа± 15 мин ➔ 10 ч 45 мин ~ 11 ч 15 мин.

средна стойност(математическо очакване)= точното времетраене на полета - неизвестно

Ниво на доверие

Означение: $(1-\alpha)100\%$

примери:

90% дов. инт.,

95% дов. инт.

Доверителен интервал за μ (σ известно)

Случай 1. Нормална популация

Нека е направена извадка с обем *n* от *нормално разпределение* с неизвестна средна стойност μ, и известно стандартно отклонение σ

Как да построим доверителен интервал за μ с (1-α)100% ниво на доверие ?

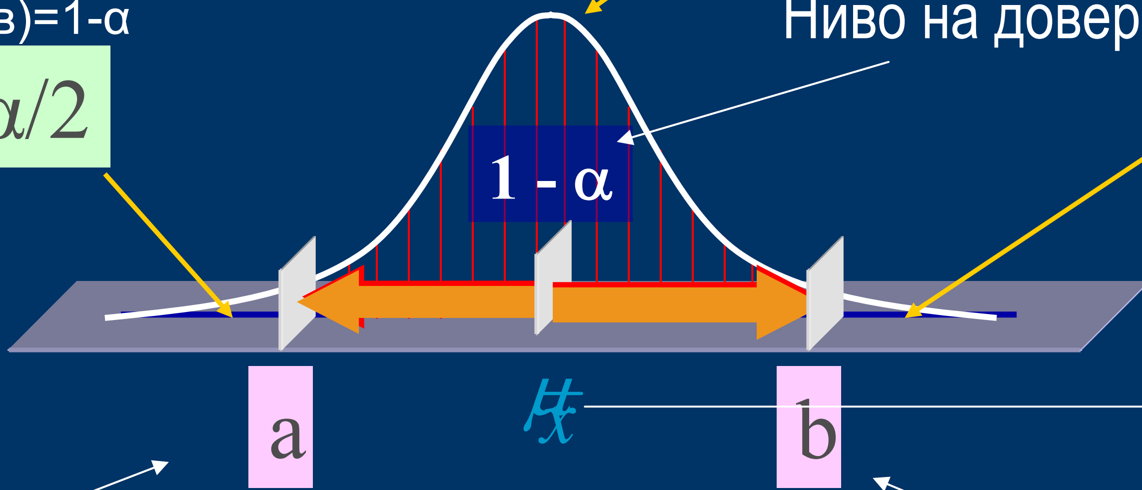
Точкова оценка на μ е извадковото средно

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \in N(\mu, \frac{\sigma^2}{n})$$

Избираме си ниво на доверие (1-α)100% и търсим доверителен интервал (а,в) така, че P(а<μ<в)=1-α

Плътност на извадковото средно

Ниво на доверие



Точкова оценка на μ е $\bar{x}$

Долна граница

Горна граница

$$2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$z_{\frac{\alpha}{2}} = \frac{a - \bar{x}}{\frac{\sigma}{\sqrt{n}}}$$

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

$$\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Пример

Нека X е количеството мляко, консумирано от българите дневно. Известно е, че това количество е нормално разпределено със стандартно отклонение 96 г. За да оценят средното количество, Асоцията на млекопроизводителите е направила случайна извадка от 576 българи и е получила средна стойност 133 грма дневно. Постройте 90% доверителен интервал.

Точкова оценка на неизвестната средна консумация е извадковото средно, $\bar{x} = 133$

1-α=0,9 α=0,1 α/2=0,05 $\Rightarrow$ $z_{0,05} = 1,645$

$$133 \pm 1,645 \frac{96}{\sqrt{576}} \Rightarrow 126,42 \leq \mu \leq 139,58$$

Интерпретация

На нивото на доверие

Пример

90% ниво на доверие

Ако построим 100 доверителни интервала, всеки на основата на различни извадки от една и съща популация, то можем да очакваме, че 90 от тях ще съдържат параметъра на популацията.

Доверителен интервал за μ (σ известно)

Случай 2. Популацията не е нормална, но обемът на извадката е достатъчно голям (по-голям от 30)

$$\left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

Пример: Направено е допитване до 100 деца в предучилищна възраст като са записани броя часове седмично прекарани пред телевизора и е намерено, че средното време е 27,197. Ако е известно, че стандартното отклонение на броя часове, прекарани пред телевизора за тази група от деца е 8.0 часа, то постройте 95% доверителен интервал на средното прекарано време седмично пред телевизора.

Решение

Параметърът, който ще оценяваме е μ –средното време, прекарано пред телевизора

Следователно, трябва да построим доверителен интервал за μ при условие, че не знаем нищо за разпределнието на популацията, но обемът на извадката n=100>30

$$\boxed{\bar{x} = 27{,}197}$$

Доколкото 1 - α =0,95, α = 0,05. то $\alpha/2$ = 0,025 и $Z_{0,025}$ = 1,96

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 27{,}191 \pm z_{0,025} \frac{8}{\sqrt{100}}$$

$$= 27{,}191 \pm 1{,}96 \frac{8}{\sqrt{100}} = 27{,}191 \pm 1{,}57 = [25{,}621;\ 28{,}761]$$

Връзка между нивото на доверие и дължината на доверителния интервал

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Нека променим само нивото на доверие=> променя се само z

Случай 1:Ниво на доверие 90%=(1-α)100%

$\alpha = 0{,}1$ $\alpha/2 = 0{,}05$

$$z_{0,05} = 1{,}645$$

$$\left[\bar{x} - 1{,}645\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1{,}645\frac{\sigma}{\sqrt{n}}\right]$$

Случай 2: Ниво на доверие 95% =(1-α)100%

$\alpha = 0{,}05$ $\alpha/2 = 0{,}025$

$$z_{0,025} = 1{,}96$$

$$\left[\bar{x} - 1{,}96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1{,}96\frac{\sigma}{\sqrt{n}}\right]$$

По-малко ниво на доверие = по-тесен доверителен интервал.

Ефект на обема на извадката върху ширината на доверителния интервал

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Нека променим само обемът на извадката => променя се само **n**

Случай 1: обем n=10

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{10}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{10}}\right]$$

Случай 2: обем n=100

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{100}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{100}}\right]$$

По-голям обем на извадката => по-тесен доверителен интервал.

Доверителен интервал за μ (σ неизвестно)

Предположения

- **Стандартното отклонение на популацията е неизвестно**
- **Популацията е нормално разпределена**

Точкова оценка на σ е s

$$s^2 = \frac{\Sigma (X - \overline{X})^2}{n-1}$$

Знаем, че

$$\overline{X} \in N(\mu, \frac{\sigma^2}{n})$$

Но σ е неизвестно и не можем да го използваме.

Заместваме σ с точковата му оценка S.

Важно

Извадковата дисперсия **s²**, както и извадковото стандартно отклонение **s** са случайни величини.

Тогава

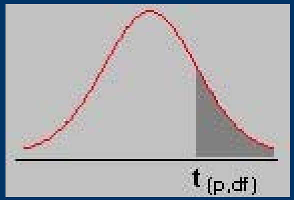
$$\frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} = \frac{\overline{X}-\mu}{s}\sqrt{n}$$

Случайна величина

е **частно** на две случайни величини

Има специално разпределение, което се нарича =>

t-разпределение



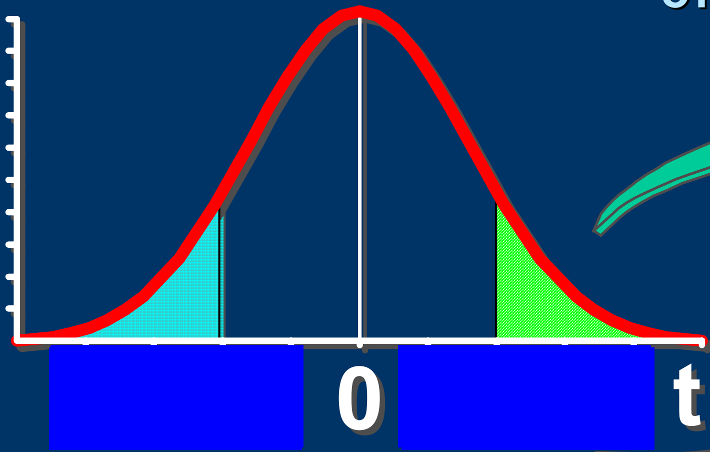
Има степени на свобода

Плътността е симетрична относно 0, т.е. 0=медиана

В t – таблица са дадени лицата под графиката на плътността

Нека ст.св. = 2

Намерете двете точки, които отсичат 20% от графиката на плътността



	Дясна опашка		
Ст. св.	.25	.10	**.05**
1	1.000	3.078	6.314
2	0.817	1.886	2.920
3	0.765	1.638	2.353

Означение

$$t_{n,\alpha}\ :\qquad P(T > t_{n,\alpha}) = \alpha$$

n - степени на свобода

α лице на дясна опашка

$t_{5;0,005} = ?$ $= 4{,}03214$

$t_{4;0,9} = ?$ $= -\ 1{,}533206$

df\p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.619 2
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688

Обратно към

Доверителен интервал за μ (σ неизвестно)

Предположения

- **Стандартното отклонение на популацията е неизвестно**
- **Популацията е нормално разпределена**

$$T = \frac{\overline{X} - \mu}{s}\sqrt{n} \in t(n-1)$$

t разпределение с **n-1** степени на свобода

Използваме таблицата за t -разпределение

Доверителния интервал :

Станд. Откл.

$$\overline{X} - t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{\frac{\alpha}{2}, n-1} \frac{S}{\sqrt{n}}$$

Тези стойности за получени от таблиците за t- разпределението

Пример

Организаторите на компютърна зала искат да знаят колко време седмично студентите прекарват в залата. За целта се избират случайно 16 студенти и се намира, че тяхното средно седмично време е 24 часа със стандартно отклонение 4 часа. Намерете 95% доверителен интервал на неизвестното средно време на всички студенти, ако се знае, че времето прекарано в компютърната зала е нормално разпределено.

Средната стойност на популацията е неизвестна. *Точкова оценка* на тази стойност е извадковото средно 24 часа.

α= 0,05

α/2= 0,025

Решение

Какво е дадено? **Популацията е нормално разпределена**

Извадковото станд. откл. s=4

Използваме t-разпределение, степени на свобода=16-1=15

от t-таблицата

$$t_{15,0.025} = 2{,}131$$

$$\overline{X} \pm 2{,}131 \frac{s}{\sqrt{n}} = 24 \pm 2{,}131 \frac{4}{\sqrt{16}}$$

$$= 24 \pm 2{,}131$$

Доверителният интервал е (21,869; 26,131).

Връзка между нивото на доверие и дължината на доверителния интервал

Разглеждаме *t (15)*

Ниво на доверие =90%

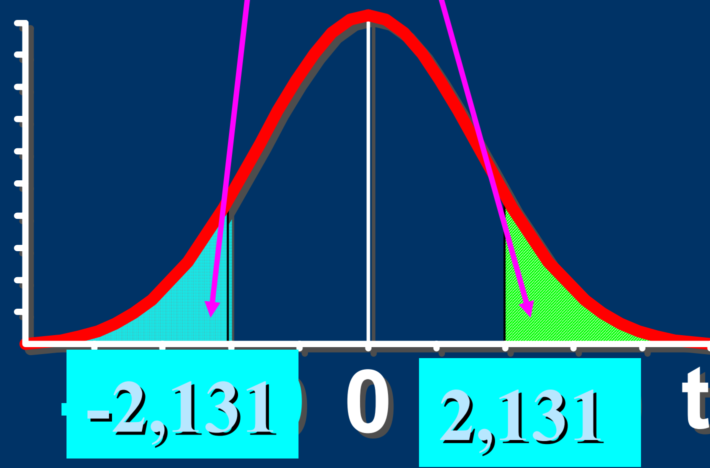
$\alpha = 0{,}1$

$\alpha /2 = 0{,}05$

Ниво на доверие =95%

$\alpha = 0{,}05$

$\alpha /2 = 0{,}025$



По=малко ниво на доверие (1-α) –
по-тесен доверителен интервал.

Ефект на обема на извадката върху ширината на доверителния интервал

$$\overline{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$$

Ако се променя само обемът, то в този случай се променя и стойността на t и стойността на s => не може да се направи извод за ширината

Сравни с доверителния интервал
с известна дисперсия!!!

Определяне обема на извадката

Колко голяма да е извадката, за да оценим средната стойност?

Оценяването става с доверителен интервал

Максималната грешка при оценяването = половината дължина на интервала

μ

Доверителен интервал

Случай 1. Знае се стандартното отклонение σ, $z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ грешката е E

$$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < E$$

$$n > \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2$$

Случай 2. Не се знае стандартното отклонение- предварително се правят някакви оценки.

Например, може да апроксимираме σ с s и да използваме нормалното разпределение (голямо n)

А ако се оценява процент , то

$$n > \left(\frac{z_{\frac{\alpha}{2}}\sqrt{p(1-p)}}{E}\right)^2 = \left(0{,}5\frac{z_{\frac{\alpha}{2}}}{E}\right)^2$$

Пример

Нека X е количеството мляко, консумирано от българите дневно. Известно е, че това количество е нормално разпределено със стандартно отклонение 16 г. Трябва да се направи извадка, с помощта на която да се построи 90% доверителен интервал, при които грешката да не е повече от 0,01 грама.

Колко е обемът на извадката????

$\sigma = 4$

$1\text{-}\alpha = 0{,}9$ $\alpha/2 = 0{,}05$ $\Rightarrow$ $z_{0,05} = 1{,}645$

$$n > \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2 = \left(\frac{1{,}645 * 4}{0{,}01}\right)^2 = 432\,964$$

Поне 432 965 българи трябва да се изберат по случаен начин

Ако увеличим грешката от 0,01 на 1 грам, то $n > \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2 = \left(\frac{1{,}645 * 4}{1}\right)^2 = 43{,}2964$

Само 44, обемът се намалява