Доверителни интервали и обем на извадка

Лекция 6

Confidence Intervals (CI) and sample size

Параметри на генералната съвкупност

Параметри – неизвестни характеристики на генералната съвкупност

Параметър	Наименование
μ	Средна
p	Относителен дял

Оценки – известни характеристики, изчислени от извадката

2

Доверителен интервал (ДИ) за средна на генералната съвкупност

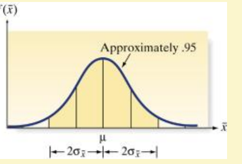
$\bar{X}$ е оценка на μ, и ние можем да използуваме Централната гранична теорема (ЦГТ) за да оценим колко точна е тази оценка

Според ЦГТ, 95% от всички $\bar{X}$ от извадките с обем n са в интервала $\bar{X} \pm 1.96\sigma_{\bar{X}}$

3

Доверителен интервал на средна на генерална съвкупност

$$\bar{x} \pm 2\sigma_{\bar{x}} = \bar{x} \pm \frac{2\sigma}{\sqrt{n}} \approx .95$$



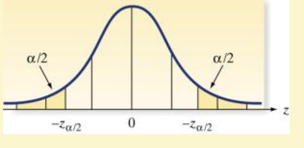
Ние обикновено не знаем σ , но при големи извадки s е добра оценка на σ

4

Равнище на значимост

Равнището на значимост е равно на $1-\alpha$

Където α е площта от нормалното разпределение, която остава в опашките



5

ДИ на средна

$$\bar{X} \pm Z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

Където: $\bar{X}$ е средната от извадката

s – стандартното оклонение от извадката

n – обема на извадката

Z – таблична стойност и зависи от доверителната вероятност (1-α)

6

Стойности на Z

Равнище на значимост	Двустранна	Стойност на Z
100(1-α)	α	**$Z_{\alpha/2}$**
90%	0.10	1.645
95%	0.05	1.96
99%	0.01	2.575

7

Интерпретация на CI:

Пример за интерпретация при 95% ДИ за средна

1. Ако правим достатъчно голям брой извадки с достатъчно голям обем и при всяка оценяваме средната, то 95% от всички доверителни интервали ще съдържат истинската средна (параметъра на генералната съвкупност) и само 5% няма да я съдържат.
2. При една единствена извадка няма начин да се разбере дали интервала съдържа истинската средна или не.

8

Пример

Да се построи 95% ДИ на средната

n = 900 и средна от извадката =100, стандартно отклонение =91.

$$\bar{X} \pm Z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) = 100 \pm 1.96\left(\frac{91}{\sqrt{900}}\right) = 100 \pm 6$$

95% ДИ: $(94 - 106)$

Интерпретация

9

Пример - продължение

- При същите условия да се построи 99% ДИ

$$\bar{X} \pm Z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) = 100 \pm 2.575\left(\frac{91}{\sqrt{900}}\right) = 100 \pm 7.8$$

99% ДИ: $(92.2 - 107.8)$

- **Интерпретация**

10

ДИ за средна при малки извадки

2 проблема
- ЦГТ не е валидна
- s не е добра оценка σ

11

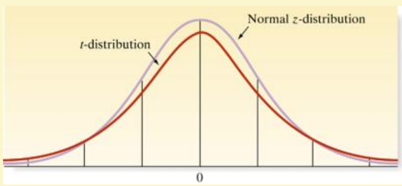
ДИ на средна при малки извадки

АКО разпределението в генералната съвкупност е приблизително нормално, то разпределението на оценките $\bar{X}$ може да се предположи че е приблизително t-разпределение на Студент.

12

t-разпределение на Стюдент

То прилича на нормалното и се определя от степените на свобода (n-1) равнището на значимост α.



13

ДИ за относителен дял
ДИ на Уолд (Wald)

Оценката се означава с $\hat{p}$

Средната от разпределението от оценките е неизместена оценка на параметъра.

Стандартното отклонение на разпределението на оценките е $\sigma_{\hat{p}} = \frac{\sqrt{pq}}{\sqrt{n}}$ където q=1-p

За големи по обем извадки разпределението на оценките е приблизително нормално.

17

ДИ за относителен дял
(ДИ на Уолд)

1. Проверка дали извадката е голяма

   Ako np≥5 и nq ≥ 5

2. ДИ

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\text{като } \hat{q} = 1 - \hat{p}$$

18

Интерпретация на CI:

Пример за интерпретация при 95% ДИ за относителен дял

1. Ако правим достатъчно голям брой извадки с достатъчно голям обем и при всяка оценяваме относителния дял, то 95% от всички доверителни интервали ще съдържат истинския относителен дял (параметъра на генералната съвкупност) и само 5% няма да го съдържат.
2. При една единствена извадка няма начин да се разбере дали интервала съдържа истинския относителен дял или не.

19

Пример

- Да се построи 95% ДИ на относителен дял
- Известно е, че n=50 и относителният дял от извадката е 0.6

1. Проверка дали извадката е голяма

   np=50*0.6=30≥5 и nq =50*0.4=20≥ 5

   Отговор: извадката е голяма и формулата може да се приложи.

20

Пример-решение

- Да се построи 95% ДИ на относителен дял
- Известно е, че n=50 и относителният дял от извадката е 0.6

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.6 \pm 1.96\sqrt{\frac{0.6*0.4}{50}} = 0.6 \pm 0.14$$

$$95\%\text{ ДИ: } (0.46 - 0.74)$$

- **Интерпретация**

21

Пример - продължение

- При същите условия да се построи 99% ДИ

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.6 \pm 2.575\sqrt{\frac{0.6*0.4}{50}} = 0.6 \pm 0.18$$

99% ДИ: $(0.42 - 0.78)$

- **Интерпретация**

22

ДИ с корекция на Агрести-Кул*

Когато p е близко до 0 или 1, ДИ на Уолд е погрешен.

Корекция на Агрести-Кул (Agresti-Coull) :

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

*Не е включено в повечето български учебници

23

ДИ с корекция на Агрести-Кул

$\hat{p}$ е относителния дял от извадката

$x = n\hat{p}$ брой успехи; $n =$ обем на извадката

Изчислява се нова оценка на относителния дял:

$$\tilde{p} = \frac{x+2}{n+4};$$

ДИ (Агрести-Кул): $\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}};$

24

Пример

- Същото условие както ДИ на Уолт. Да се построи 95% ДИ на относителен дял с корекция на Агрести-Кул
- Известно е, че n=50 и относителният дял от извадката е 0.6

$$\hat{p} = \frac{x}{n} \Rightarrow x = n\hat{p} = 50*0.6 = 30;$$

$$\tilde{p} = \frac{x+2}{n+4} = \frac{30+2}{50+4} = 0.59$$

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} = 0.59 \pm 1.96\sqrt{\frac{0.59(1-0.59)}{50+4}} = 0.59 \pm 0.13;$$

95% ДИ (Агрести-Кул) $(0.46 - 0.72)$

25

Определяне обема на извадката

От формулата за ДИ $\bar{X} \pm Z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$ изразяваме n

$$n = \frac{(Z_{\alpha/2})^2 s^2}{\Delta^2};$$

където: Δ е максималната грешка

или също е половината на доверителния интервал

s - предполагаемото разсейване измерено с s

Z - табличната стойност определена от желаното ниво на равнището на значимост

26

Пример

- Колко единици трябва да наблюдаваме ако искаме да оценим средната на генералната съвкупност?
- Искаме максималната грешка да бъде равна на 2
- Известно е (или предполагаме) че стандартното отклонение е 6.
- Какъв трябва да бъде обема на извадката?

27

Обем на извадката

- Извадка

$$n = \frac{\left(Z_{\alpha/2}\right)^2 s^2}{\Delta^2} = \frac{\left(1.96\right)^2 6^2}{2^2} = 34.57 \approx 35$$

- Извадката трябва да бъде с обем най-малко 35 единици.

28

Пример - продължение

- При същите условия, но сега искаме максималната грешка да е 1.

$$n = \frac{\left(Z_{\alpha/2}\right)^2 s^2}{\Delta^2} = \frac{\left(1.96\right)^2 6^2}{1^2} = 138.29 \approx 139$$

- При същите условия, но сега искаме доверителната вероятност да е 99%.

$$n = \frac{\left(Z_{\alpha/2}\right)^2 s^2}{\Delta^2} = \frac{\left(2.575\right)^2 6^2}{2^2} = 59.67 \approx 60$$

29