

Teckenintervall

19 okt. 2026, 4 min läsning

UTKAST

#matematik #statistik #metod

Kurs: S0008M Förkunskaper: Statistisk inferens, Binomialfördelning

Teckenintervallet är ett konfidensintervall för medianen som bara bygger på observationernas tecken relativt ett värde. Metoden är fördelningsfri — den kräver inget normalfördelningsantagande — men ger till gengäld ett bredare intervall.

1. Förutsättningar

Observationerna $x_1, \dots, x_n$ ses som oberoende observationer av en **kontinuerlig** stokastisk variabel ξ med okänd median m . Mer än så behöver vi inte veta — inte vilken fördelning det är.

Medianen uppfyller $P(\xi < m) = P(\xi > m) = 0,5$. Varje observation hamnar alltså under m med sannolikheten 0,5, oberoende av de andra.

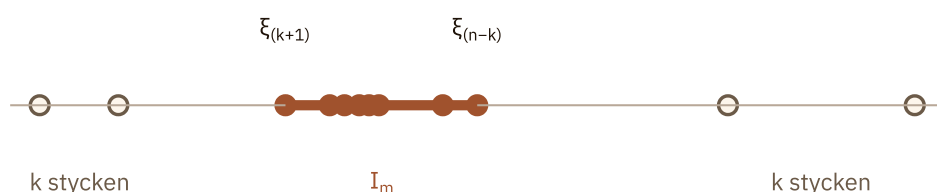
2. Ordnade observationer

Sortera observationerna i växande ordning och beteckna dem

$$\xi_{(1)} \leq \xi_{(2)} \leq \dots \leq \xi_{(n)}$$

så att $\xi_{(1)}$ är den minsta och $\xi_{(n)}$ den största. Teckenintervallet stryker de k minsta och de k största:

$$I_m = [\xi_{(k+1)}, \xi_{(n-k)}]$$



3. Konfidensgraden

Låt η = antalet observationer som är mindre än m . Varje observation är en **oberoende upprepning** med sannolikheten 0,5, så

$$\eta \in \text{Bin}(n, 0,5)$$

Intervallet missar m om m ligger till vänster om det – högst k observationer under m – eller till höger – högst k observationer över m . De två händelserna är disjunkta och lika sannolika av symmetriskäl:

$$P(m \in I_m) = 1 - 2P(\eta \leq k), \quad \eta \in \text{Bin}(n, 0,5)$$

Konfidensgraden kan inte väljas fritt

Eftersom k är ett heltal kan vi bara få vissa konfidensgrader. Vi väljer det största k som ger minst den konfidensgrad vi vill ha, till exempel 95 %.

Generell men inte effektiv

Teckenintervallet använder bara om varje observation ligger under eller över m , inte hur långt ifrån. Vet vi att observationerna är normalfördelade ger **normalfördelningsintervallen** kortare intervall med samma konfidensgrad.

Vanliga fel

- **Glömma faktorn 2.** Konfidensgraden är $1 - 2P(\eta \leq k)$; intervallet kan missa på båda sidor.
- **Räkna från fel ände.** Intervallet går från den $(k + 1)$:a minsta till den $(k + 1)$:a största observationen.

4. Lösta lektionsuppgifter

Lektion 11, övning 1 och 2 – Konfidensgrad för teckenintervall

Låt ξ vara en kontinuerlig stokastisk variabel med median m , och gör 10 oberoende mätningar.

- a) Bestäm $P(\xi < m)$.
- b) Vilken konfidensgrad har teckenintervallet $[\xi_{(1)}, \xi_{(10)}]$?
- c) Låt η = antalet mätvärden som är $< m$. Vilka värden kan η anta, och vilken fördelning har η ?
- d) Bestäm konfidensgraden för intervallet $[\xi_{(3)}, \xi_{(8)}]$.
- e) Generalisera till ett stickprov av storlek n och intervallet $[\xi_{(k+1)}, \xi_{(n-k)}]$.

 Lösning

a) Medianen delar massan på mitten: $P(\xi < m) = \boxed{0,5}$.

c) η kan anta värdena $0, 1, \dots, 10$. Varje mätning hamnar under m med sannolikheten $0,5$, oberoende av de andra, så $\eta \in \text{Bin}(10, 0,5)$.

b) $[\xi_{(1)}, \xi_{(10)}]$ är teckenintervallet med $k = 0$. Det missar bara om alla tio hamnar på samma sida om m :

$$P(m \in I) = 1 - 2P(\eta = 0) = 1 - 2 \cdot 0,5^{10} \approx \boxed{0,998}$$

d) $[\xi_{(3)}, \xi_{(8)}]$ stryker två observationer i varje ände, $k = 2$:

$$P(m \in I) = 1 - 2P(\eta \leq 2) = 1 - 2 \cdot \frac{\binom{10}{0} + \binom{10}{1} + \binom{10}{2}}{2^{10}} = 1 - 2 \cdot \frac{56}{1024} \approx \boxed{0,891}$$

e) Samma resonemang med $\eta \in \text{Bin}(n, 0,5)$ ger $P(m \in I) = 1 - 2P(\eta \leq k)$, avsnitt 3 ovan.

På räknaren. d) `1-2*binomCdf(10,0.5,0,2)`.

Facit: b) konfidensgraden 99,8 % (med bokens data i övning 8.7 blir intervallet $[140, 520]$); c) $\eta \in \text{Bin}(10, 0,5)$; d) 0,891.

≡ Lektion 12, övning 10 (Vännman 8.19 b) – Densitet utan normalfördelningsantagande

Densiteten (g/cm^3) hos ett material mäts 13 gånger. Vi antar ingenting om att mätvärdena är normalfördelade, bara att de är oberoende observationer av en kontinuerlig fördelning. Bilda ett konfidensintervall för fördelningens median m med konfidensgrad nära 98 %.

Lösning

Metod. Utan normalfördelningsantagande är teckenintervallet det som går att använda. Förutsättningen — oberoende observationer från en kontinuerlig fördelning — är uppfylld.

Konfidensgrad per k . Med $n = 13$ och $\eta \in \text{Bin}(13, 0,5)$:

k	0	1	2	3
$1 - 2P(\eta \leq k)$	0,9998	0,9966	0,9775	0,9077

$k = 2$ ger 97,8 %, det närmaste vi kommer 98 %. Intervallet blir $[\xi_{(3)}, \xi_{(11)}]$: stryk de två minsta och de två största mätvärdena.

Resultat. Med bokens mätvärden blir intervallet $[0,501; 0,511]$ med konfidensgrad 97,8 %.

Jämförelse. Med normalfördelningsantagande blev 98 %-intervallet $[0,502; 0,514]$, se **t-fördelningen**. Intervallen mäter olika saker — median respektive väntevärde — men för en normalfördelning sammanfaller de.

Uppgiften (Vännman 8.19 b, i 2:a uppl. 8.20 b) är återgiven ur det handskrivna lösningsförslaget, och intervallet och konfidensgraderna stämmer med det. Mätvärdena finns inte i valvet.

Se även

- Konfidensintervall
- Binomialfördelning
- Statistisk inferens

- Median och kvantiler

Resurser

- [Wikipedia: Sign test](#)
 - [Wikipedia: Order statistic](#)
-
-