Konfidenzintervalle

statistik-online.ch

Ein **Konfidenzintervall** gibt einen Bereich an, in dem ein unbekannter Populationsparameter mit einer bestimmten Vertrauenswahrscheinlichkeit geschätzt wird.

1 Grundidee

Wir schätzen einen unbekannten Populationsmittelwert nicht mit einem einzigen Wert, sondern mit einem plausiblen Bereich.

Typische Fragestellung

Eine Stichprobe liefert den Mittelwert $\bar{x}$. Da Stichproben zufällig schwanken, ist $\bar{x}$ normalerweise nicht exakt gleich dem wahren Populationsmittelwert μ. Ein Konfidenzintervall beschreibt deshalb einen Bereich plausibler Werte für μ.

- Wie gross ist die durchschnittliche Körpergrösse in einer Population?
- In welchem Bereich liegt die durchschnittliche Bearbeitungszeit?
- Wie genau wurde der Mittelwert einer Stichprobe geschätzt?

Aufbau eines Konfidenzintervalls

Ein Konfidenzintervall besteht immer aus einem Punktschätzer und einer Fehlerspanne:

$$\text{Konfidenzintervall} = \text{Schätzer} \pm \text{Fehlerspanne}$$

Für den Mittelwert ist der Schätzer der Stichprobenmittelwert $\bar{x}$.

$$\text{Konfidenzintervall} = \bar{x} \pm \text{kritischer Wert} \cdot \text{Standardfehler}$$

2 z oder t?

Die Wahl zwischen z-Verteilung und t-Verteilung hängt davon ab, ob die Populationsstandardabweichung σ bekannt ist.

Situation	**Verteilung**	**Standardfehler**
σ bekannt	z-Verteilung	$\sigma/\sqrt{n}$
σ unbekannt, nur s bekannt	t-Verteilung	$s/\sqrt{n}$

Merksatz:

$$\sigma \text{ bekannt} \Rightarrow z \qquad \sigma \text{ unbekannt} \Rightarrow t$$

In der Praxis ist σ oft unbekannt. Dann wird die Stichprobenstandardabweichung s verwendet und das Konfidenzintervall mit der t-Verteilung berechnet. Bei grossen Stichproben werden z- und t-Werte sehr ähnlich, aber die t-Verteilung bleibt bei unbekanntem σ die saubere Standardwahl.

3 Konfidenzintervall mit z-Werten

Wenn die Populationsstandardabweichung σ bekannt ist, lautet das Konfidenzintervall für den Mittelwert:

$$\bar{x} \pm z_{krit} \cdot \frac{\sigma}{\sqrt{n}}$$

Symbol	Bedeutung
$\bar{x}$	Mittelwert der Stichprobe
σ	bekannte Populationsstandardabweichung
n	Stichprobengrösse
z_{krit}	kritischer z-Wert zum Konfidenzniveau

Konfidenzniveau	90%	95%	99%
z_{krit}	1.645	1.960	2.576

Beispiel mit bekanntem σ

Eine Stichprobe von $n = 64$ Personen hat den Mittelwert $\bar{x} = 172.4$ cm. Die Populationsstandardabweichung ist bekannt: $\sigma = 6.0$ cm. Gesucht ist ein 95%-Konfidenzintervall.

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{6.0}{\sqrt{64}} = 0.75$$

$$\text{Fehlerspanne} = z_{krit} \cdot SE = 1.960 \cdot 0.75 = 1.47$$

$$172.4 \pm 1.47 = [170.93,\ 173.87]$$

Interpretation: Der Populationsmittelwert wird mit 95% Konfidenz zwischen 170.93 cm und 173.87 cm geschätzt.

4 Konfidenzintervall mit t-Werten

Wenn die Populationsstandardabweichung σ unbekannt ist und nur die Stichprobenstandardabweichung s vorliegt, lautet das Konfidenzintervall:

$$\bar{x} \pm t_{krit} \cdot \frac{s}{\sqrt{n}}$$

Die Freiheitsgrade sind:

$$df = n - 1$$

df	90%	95%	99%
9	1.833	2.262	3.250
15	1.753	2.131	2.947
24	1.711	2.064	2.797
30	1.697	2.042	2.750

Beispiel mit unbekanntem σ

Eine Stichprobe von $n = 16$ Personen hat den Mittelwert $\bar{x} = 48.2$ und die Stichprobenstandardabweichung $s = 5.6$. Gesucht ist ein 95%-Konfidenzintervall.

$$df = 16 - 1 = 15 \qquad t_{krit} = 2.131$$

$$SE = \frac{s}{\sqrt{n}} = \frac{5.6}{\sqrt{16}} = 1.40$$

$$\text{Fehlerspanne} = t_{krit} \cdot SE = 2.131 \cdot 1.40 = 2.98$$

$$48.2 \pm 2.98 = [45.22,\ 51.18]$$

Interpretation: Der Populationsmittelwert wird mit 95% Konfidenz zwischen 45.22 und 51.18 geschätzt.

5 Interpretation

Ein 95%-Konfidenzintervall bedeutet nicht, dass der wahre Mittelwert mit 95% Wahrscheinlichkeit in genau diesem berechneten Intervall liegt. Der wahre Mittelwert ist fest, das Intervall ist zufallsabhängig.

Eine geeignete Interpretation lautet:

> Wenn man sehr viele Stichproben auf dieselbe Weise ziehen und jedes Mal ein 95%-Konfidenzintervall berechnen würde, dann würden langfristig ungefähr 95% dieser Intervalle den wahren Populationsmittelwert enthalten.

Für die Kommunikation reicht oft:

> Der Populationsmittelwert wird mit 95% Konfidenz im angegebenen Bereich geschätzt.

6 Zusammenfassung

Situation	**Formel**	**Verteilung**
σ bekannt	$\bar{x} \pm z_{krit}\frac{\sigma}{\sqrt{n}}$	z-Verteilung
σ unbekannt	$\bar{x} \pm t_{krit}\frac{s}{\sqrt{n}}$	t-Verteilung

Schritt	**Frage**	**Entscheidung**
1	Ist σ bekannt?	Ja: z, Nein: t
2	Standardfehler	$\sigma/\sqrt{n}$ oder $s/\sqrt{n}$
3	Kritischer Wert	z-Tabelle oder t-Tabelle
4	Fehlerspanne	kritischer Wert $\cdot SE$
5	Intervall	$\bar{x}\pm$ Fehlerspanne

Aufgaben: Konfidenzintervalle

1 – Lieferzeit

Eine Stichprobe von $n = 36$ Lieferungen hat eine durchschnittliche Lieferzeit von $\bar{x} = 82.5$ Minuten. Die Populationsstandardabweichung ist bekannt: $\sigma = 9.0$ Minuten. Berechnen Sie ein 95%-Konfidenzintervall.

- Entscheiden Sie, ob z oder t verwendet wird.
- Berechnen Sie den Standardfehler.
- Berechnen Sie das Konfidenzintervall.

2 – Lernzeit

Eine Stichprobe von $n = 10$ Studierenden hat eine durchschnittliche Lernzeit von $\bar{x} = 24.8$ Stunden pro Woche. Die Populationsstandardabweichung ist unbekannt, die Stichprobenstandardabweichung beträgt $s = 3.0$. Berechnen Sie ein 95%-Konfidenzintervall.

- Entscheiden Sie, ob z oder t verwendet wird.
- Bestimmen Sie die Freiheitsgrade.
- Berechnen Sie das Konfidenzintervall.

3 – Testergebnis

Eine Stichprobe von $n = 25$ Personen erzielt im Durchschnitt $\bar{x} = 101.3$ Punkte. Die Populationsstandardabweichung ist unbekannt, die Stichprobenstandardabweichung beträgt $s = 12.5$. Berechnen Sie ein 99%-Konfidenzintervall.

- Entscheiden Sie, ob z oder t verwendet wird.
- Berechnen Sie den Standardfehler.
- Berechnen und interpretieren Sie das Intervall.

4 – Multiple Choice

Welche Aussagen sind korrekt? (Mehrfachantworten möglich)

- ❑ Wenn σ bekannt ist, wird für das Mittelwertintervall die z-Verteilung verwendet.
- ❑ Wenn σ unbekannt ist und nur s vorliegt, wird die t-Verteilung verwendet.
- ❑ Die Freiheitsgrade beim t-Konfidenzintervall sind $df = n - 1$.
- ❑ Ein 95%-Konfidenzintervall bedeutet, dass 95% aller Stichprobenmittelwerte im Intervall liegen.
- ❑ Eine grössere Stichprobe macht den Standardfehler kleiner.

Lösungen: Konfidenzintervalle

Lösung zu Aufgabe 1

Da σ bekannt ist, wird die z-Verteilung verwendet.

$$SE = \frac{9.0}{\sqrt{36}} = 1.50$$

Für ein 95%-Konfidenzintervall gilt $z_{krit} = 1.960$.

$$\text{Fehlerspanne} = 1.960 \cdot 1.50 = 2.94$$

$$82.5 \pm 2.94 = [79.56,\ 85.44]$$

Interpretation: Die durchschnittliche Lieferzeit wird mit 95% Konfidenz zwischen 79.56 und 85.44 Minuten geschätzt.

Lösung zu Aufgabe 2

Da σ unbekannt ist und nur s vorliegt, wird die t-Verteilung verwendet.

$$df = 10 - 1 = 9 \qquad t_{krit} = 2.262$$

$$SE = \frac{3.0}{\sqrt{10}} = 0.95$$

$$\text{Fehlerspanne} = 2.262 \cdot 0.95 = 2.15$$

$$24.8 \pm 2.15 = [22.65,\ 26.95]$$

Interpretation: Die durchschnittliche Lernzeit wird mit 95% Konfidenz zwischen 22.65 und 26.95 Stunden geschätzt.

Lösung zu Aufgabe 3

Da σ unbekannt ist, wird die t-Verteilung verwendet.

$$df = 25 - 1 = 24 \qquad t_{krit} = 2.797$$

$$SE = \frac{12.5}{\sqrt{25}} = 2.50$$

$$\text{Fehlerspanne} = 2.797 \cdot 2.50 = 6.99$$

$$101.3 \pm 6.99 = [94.31,\ 108.29]$$

Interpretation: Das durchschnittliche Testergebnis wird mit 99% Konfidenz zwischen 94.31 und 108.29 Punkten geschätzt.

Lösung zu Aufgabe 4

✓ Richtig – Bei bekannter Populationsstandardabweichung wird z verwendet.

✓ Richtig – Bei unbekannter Populationsstandardabweichung wird s verwendet und mit t gerechnet.

✓ Richtig – Beim t-Konfidenzintervall für den Mittelwert gilt $df = n - 1$.

✗ Falsch – Die Aussage beschreibt die Interpretation nicht korrekt.

✓ Richtig – Wenn n grösser wird, wird der Standardfehler kleiner.