

## Biostatistik

### Umgang mit kleinen Stichprobenumfängen bei der statistischen Analyse naturwissenschaftlicher Daten

1. Bei der Analyse von Getreideproben in einem Labor wurden folgende 24 Messwerte für den Kupfergehalt (gemessen in parts per million) ermittelt.

2.90	2.50	3.60	3.70	5.28	3.10	3.40	3.70
2.89	2.70	3.70	3.77	3.03	3.03	3.40	3.50
2.20	2.20	3.40	2.40	3.37	2.40	3.70	2.80

- (a) Beschreiben Sie die Daten anhand geeigneter Kennzahlen wie Mittelwert, Median und Stichprobenvarianz und stellen Sie die Daten graphisch dar.
  - (b) Geben Sie jeweils für den Mittelwert und die Varianz ein Konfidenzintervall unter der Annahme, die Daten stammen aus einer Normalverteilung  $\mathcal{N}(\mu, \sigma)$ , an.
  - (c) Laut gesetzlicher Vorschrift darf der mittlere Kupfergehalt eines Produkts den Grenzwert von 4.0 ppm nicht überschreiten. Formulieren Sie die Hypothesen für einen statistischen Test und führen Sie diesen durch (Signifikanzniveau: 5%). Dieser soll zeigen, dass der durchschnittliche Kupfergehalt die gesetzliche Obergrenze nicht überschreitet.
2. Nehmen wir an, dass in der Studie aus Kostengründen anstatt der 24 Proben nur 8 Proben analysiert wurden.

2.90	2.50	3.60	3.70	5.28	3.10	3.40	3.70
------	------	------	------	------	------	------	------

- (a) Wiederholen Sie Ihre Analysen und vergleichen Sie die Ergebnisse mit der vollständigen Stichprobe. Wie breit sind jetzt die Konfidenzintervalle? Wie verändert sich dadurch die Aussagekraft Ihrer Schätzung?
  - (b) Diskutieren Sie, ob aus Kostengründen auch die reduzierte Stichprobe von nur 8 Messwerten ausreichen würde.
  - (c) Welche Annahmen der klassischen Tests (t-Test, Varianzschätzer etc.) könnten bei kleinen n verletzt sein? Welche Alternativen gäbe es?
3. Angenommen, der erste Wert der jeweiligen Stichproben wurde versehentlich mit verschobener Dezimalstelle notiert. Aus 2.90 wurde 29.0.
    - (a) Wie wirkt sich dieser einzelne Fehler auf Ihre Analyse aus?
    - (b) Welche Strategien zur Absicherung gegen solche Fehler gibt es?

**Hinweis:** Die verwendeten Datensätze sowie Beispielimplementierungen zur Analyse sind in einem GitHub-Repository [GitHub-Link](#) hinterlegt. Die enthaltenen Python-Skripte und Jupyter-Notebooks sollen als Startpunkt dienen und können gern erweitert und für eigene Zwecke angepasst werden.