LU02b – Standardabweichung und die 68–95–99.7-Regel

Ziel: Du kannst die Standardabweichung als Streuungsmass erklären, die 68–95–99.7-Regel anwenden und Werte mithilfe des **z-Scores standardisieren**.

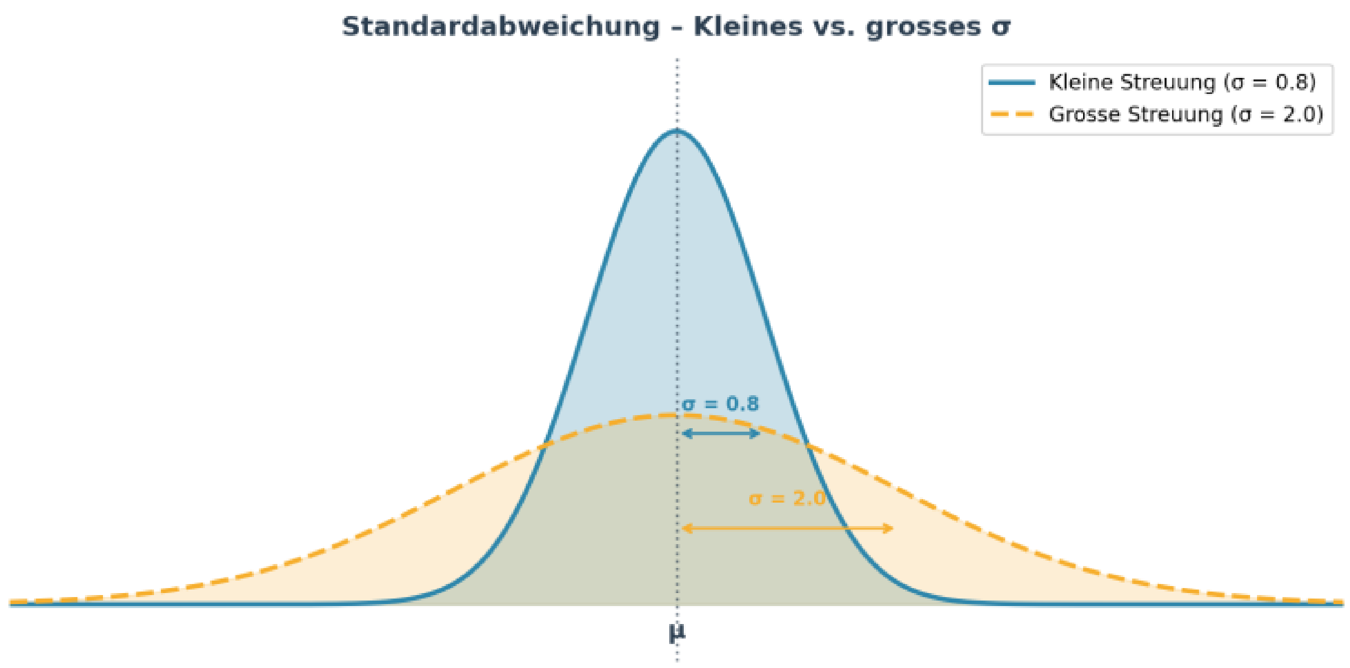
Kurztheorie (Merksätze)

- **σ (Sigma)** = Standardabweichung → misst, wie stark Werte um den Mittelwert streuen.
- **68 %** aller Werte liegen innerhalb von **± 1σ** vom Mittelwert.
- **95 %** aller Werte liegen innerhalb von **± 2σ** vom Mittelwert.
- **99.7 %** aller Werte liegen innerhalb von **± 3σ** vom Mittelwert.

---

1) Was ist die Standardabweichung?

Die **Standardabweichung σ** (Sigma) misst, wie stark die Werte eines Datensatzes um den **Mittelwert μ** herum streuen.



Sigma	Bedeutung	Kurvenform
Kleines σ	Werte liegen eng beieinander	Schmale, hohe Glocke
Grosses σ	Werte sind weit gestreut	Breite, flache Glocke

> **Merksatz:** Ein kleines σ bedeutet **Präzision** (z.B. eine gut kalibrierte Maschine). Ein grosses σ bedeutet **grosse Streuung** (z.B. unterschiedliche Testergebnisse in einer heterogenen Klasse).

Die Formel für die Standardabweichung lautet:

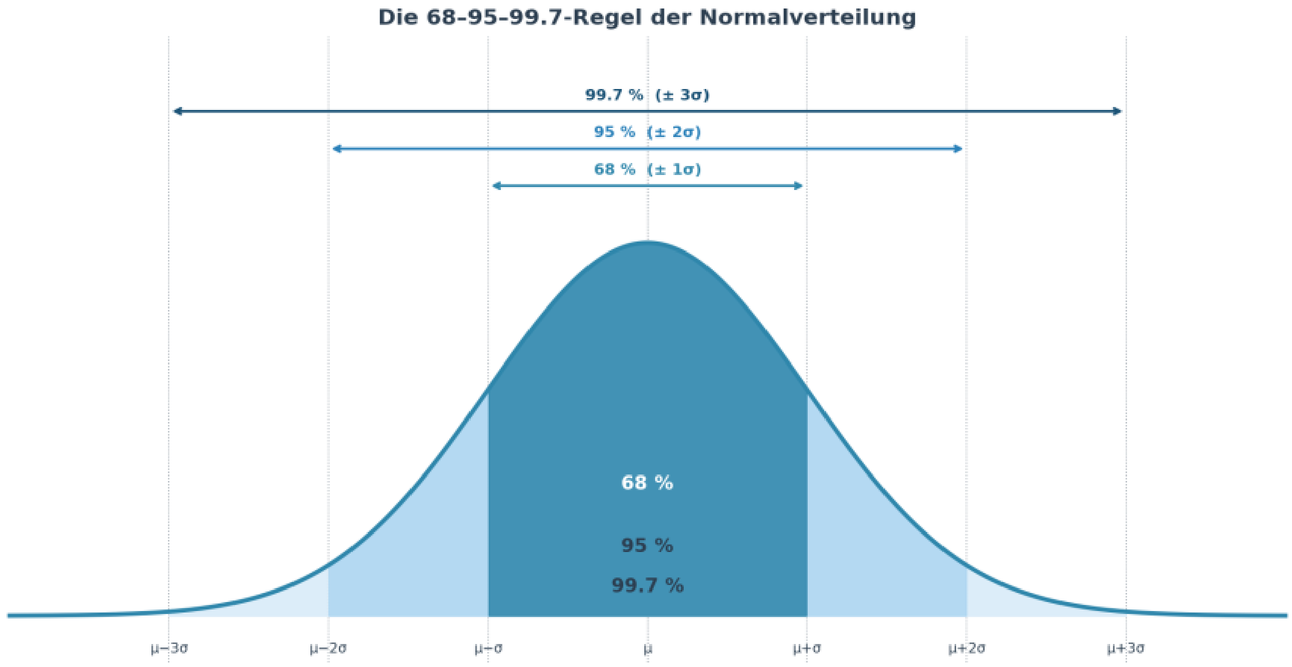
$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$

> **Hinweis:** In der Praxis verwenden wir oft einen Taschenrechner oder Software (Excel, Python) für die Berechnung.

---

2) Die 68–95–99.7-Regel

Bei einer normalverteilten Datenmenge gilt immer folgende **Faustregel**:



Bereich	Anteil der Werte	Bedeutung
$\mu \pm 1\sigma$	**68 %**	Wahrscheinlich – 68 von 100 Werten fallen hier rein
$\mu \pm 2\sigma$	**95 %**	Sehr wahrscheinlich – 95 von 100
$\mu \pm 3\sigma$	**99.7 %**	Fast sicher – 997 von 1000

Diese Regel wird auch **Empirische Regel** genannt und ist eines der wichtigsten Werkzeuge der Statistik.

2.1 Anwendungsbeispiel: Körpergrössen

In einer Schule sind **95 % der Schülerinnen und Schüler zwischen 1.1 m und 1.7 m gross**.

Angenommen, die Daten sind normalverteilt – gesucht: Mittelwert μ und Standardabweichung σ.

Schritt 1 – Mittelwert berechnen:

$$\mu = \frac{1.1 + 1.7}{2} = 1.4\,\mathrm{m}$$

Schritt 2 – Standardabweichung berechnen:

95 % entsprechen ± 2σ → der Gesamtbereich von 1.1 bis 1.7 m umfasst 4σ:

$$\sigma = \frac{1.7 - 1.1}{4} = \frac{0.6}{4} = 0.15\,\mathrm{m}$$

Kontrolle:

- 1σ-Bereich: 1.25 m – 1.55 m → ca. **68 %** der Schüler
- 2σ-Bereich: 1.10 m – 1.70 m → ca. **95 %** der Schüler ✓
- 3σ-Bereich: 0.95 m – 1.85 m → ca. **99.7 %** der Schüler

3) Prozentzonen der Standardnormalverteilung

Die Standardnormalverteilung (μ = 0, σ = 1) lässt sich in Zonen aufteilen:

Zone	Anteil in diesem Bereich	Kumulierter Anteil (ab −∞)
unter −3σ	0.1 %	0.1 %
−3σ bis −2σ	2.1 %	2.3 %
−2σ bis −1σ	13.6 %	15.9 %
−1σ bis 0	34.1 %	50.0 %
0 bis +1σ	34.1 %	84.1 %
+1σ bis +2σ	13.6 %	97.7 %
+2σ bis +3σ	2.1 %	99.7 %
über +3σ	0.1 %	99.9 %

Lesbeispiel

Dein Testergebnis liegt **0.5 Standardabweichungen über dem Durchschnitt** (z = +0.5). Wie viele Personen schnitten schlechter ab?

- Anteil unter dem Mittelwert (z < 0): **50 %**
- Anteil zwischen 0 und +0.5σ: **19.1 %**
- Total: **69.1 %** schnitten schlechter ab als du.

---

4) Praxisbeispiel: Zuckerbeutel-Produktion

Eine Abfüllmaschine hat: \(\mu = 1010\) g, \(\sigma = 20\) g.

Problem: Etwa 31 % der Beutel enthalten weniger als 1000 g!

Ziel: 1000 g soll bei **−2.5σ** liegen (dann nur noch 0.6 % zu leichte Beutel).

Option A – Mittelwert erhöhen:

$$\mu_{neu} = 1000 + 2.5 \times 20 = 1050 \text{ g}$$

Option B – Streuung verringern:

$$\sigma_{neu} = \frac{10}{2.5} = 4 \text{ g}$$

> **Fazit:** Mit der 68–95–99.7-Regel und dem z-Score können Produktionsprozesse **quantitativ gesteuert** werden – entweder durch Anpassen des Mittelwerts oder durch Verringern der Streuung.

Verständnisfragen

1. Ein Datensatz hat \(\mu = 50\) und \(\sigma = 10\). Welcher Anteil der Werte liegt zwischen **30 und 70**?
2. Berechne den z-Score für \(x = 75\), wenn \(\mu = 60\) und \(\sigma = 8\).
3. Was bedeutet ein z-Score von **−2.5** anschaulich?
4. Warum ist die Standardisierung nützlich, wenn man Werte aus **verschiedenen Verteilungen** vergleichen will?
5. Eine Maschine produziert Teile mit \(\mu = 100\) mm und \(\sigma = 2\) mm. Wie gross ist der Anteil der Teile, die **kürzer als 96 mm** sind?

---

5) Normalverteilung ohne Histogramm prüfen

Manchmal liegt kein Histogramm vor – man hat nur die **rohen Datenwerte**. Trotzdem lässt sich mit der empirischen Regel prüfen, ob die Daten näherungsweise normalverteilt (hügelförmig) sind.

> **Grundidee:** Wenn ein Datensatz wirklich normalverteilt ist, dann müssten ungefähr **68 %**, **95 %** und **99.7 %** der Werte in den entsprechenden σ-Bändern liegen. Wir berechnen das selbst – und vergleichen.

5.1 Vorgehen (Schritt für Schritt)

Schritt 1 – Mittelwert μ und Standardabweichung σ berechnen

Aus den Rohdaten: \(\mu = \bar{x}\) und \(\sigma\) wie gewohnt (Taschenrechner oder Excel).

Schritt 2 – Die drei Intervallgrenzen bestimmen

$$ [\mu - \sigma \;;\; \mu + \sigma] \qquad [\mu - 2\sigma \;;\; \mu + 2\sigma] \qquad [\mu - 3\sigma \;;\; \mu + 3\sigma] $$

Schritt 3 – Werte in jedem Band zählen

Wie viele Datenpunkte liegen innerhalb jedes Intervalls? → Anzahl durch Gesamtanzahl n dividieren → Prozentwert.

Schritt 4 – Mit der 68–95–99.7-Regel vergleichen

Band	Gemessener Anteil	Erwarteter Anteil	Beurteilung
$\mu \pm 1\sigma$	(selbst berechnet)	≈ **68 %**	Abweichung < 10 % → OK
$\mu \pm 2\sigma$	(selbst berechnet)	≈ **95 %**	Abweichung < 5 % → OK
$\mu \pm 3\sigma$	(selbst berechnet)	≈ **99.7 %**	Fast alle Werte → OK

Wenn alle drei Werte gut übereinstimmen: Daten sind **näherungsweise normalverteilt**.

> **Wichtig:** Bei kleinen Stichproben (n < 30) sind Abweichungen normal – die Regel funktioniert besser bei grossen Datensätzen.

5.2 Vollständiges Beispiel: Testergebnisse

Gegeben: Testergebnisse von 30 Schülerinnen und Schülern (Punkte von 100):

62, 65, 68, 68, 70, 71, 72, 72, 73, 74,

```
74, 75, 75, 76, 76, 77, 77, 78, 78, 79,
79, 80, 81, 82, 82, 83, 84, 85, 87, 91
```

Schritt 1 – Kennzahlen berechnen:

$$ \mu = 76.8 \qquad \sigma = 6.1 \qquad n = 30 $$

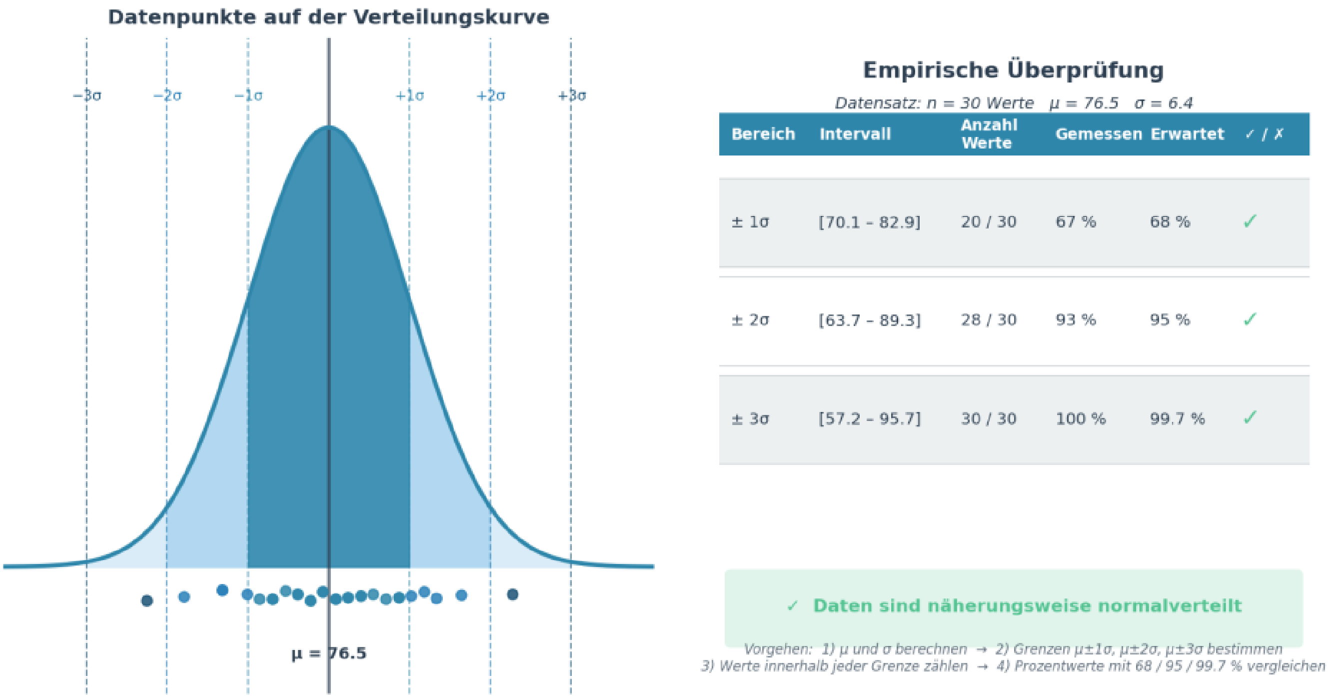
Schritt 2 – Intervallgrenzen:

$$ \mu \pm 1\sigma = [70.7 \;;\; 82.9] $$ $$ \mu \pm 2\sigma = [64.6 \;;\; 89.0] $$ $$ \mu \pm 3\sigma = [58.5 \;;\; 95.1] $$

Schritt 3 – Werte zählen:

Band	Intervall	Anzahl Werte	Anteil
$\mu \pm 1\sigma$	[70.7 – 82.9]	21 von 30	**70 %**
$\mu \pm 2\sigma$	[64.6 – 89.0]	28 von 30	**93 %**
$\mu \pm 3\sigma$	[58.5 – 95.1]	30 von 30	**100 %**

Schritt 4 – Vergleich und Urteil:



Band	Gemessen	Erwartet	Differenz	Urteil
μ ± 1σ	70 %	68 %	+2 %	✓ passt gut
μ ± 2σ	93 %	95 %	−2 %	✓ passt gut
μ ± 3σ	100 %	99.7 %	+0.3 %	✓ passt gut

Fazit: Alle drei Bänder stimmen gut mit den Erwartungswerten überein → Die Testergebnisse sind **näherungsweise normalverteilt**.

5.3 Gegenbeispiel: Stark schiefe Daten

Gegeben: Jahreseinkommen von 20 Personen (in TCHF):

42, 45, 47, 48, 50, 51, 52, 53, 55, 58,

```
60, 62, 65, 70, 80, 95, 110, 140, 200, 380
```

$$\mu = 88.2 \qquad \sigma = 82.5 \qquad n = 20$$

Band	Intervall	Anteil	Erwartet	Urteil
μ ± 1σ	[5.7 – 170.7]	18 / 20 = **90 %**	68 %	✗ zu viele
μ ± 2σ	[−76.8 – 253.2]	19 / 20 = **95 %**	95 %	~ passt zufällig
μ ± 3σ	[−159.3 – 335.7]	19 / 20 = **95 %**	99.7 %	✗ zu wenige

Fazit: Das 1σ-Band enthält 90 % statt 68 % → Die Daten sind **rechtsschief verteilt**, nicht normalverteilt. (Typisch für Einkommensdaten.)

5.4 Zusammenfassung: Schnellcheck

```
1. μ und σ berechnen
2. Grenzen bestimmen:  μ±1σ, μ±2σ, μ±3σ
3. Werte in jedem Band zählen → Prozent berechnen
4. Vergleichen:
     ≈ 68 %  im 1σ-Band?
     ≈ 95 %  im 2σ-Band?
     ≈ 99.7% im 3σ-Band?
5. Urteil: Alle drei passen → näherungsweise normalverteilt ✓
           Starke Abweichungen → eher nicht normalverteilt ✗
```

> **Einschränkung:** Diese Methode ist ein **pragmatischer Schnellcheck**, kein formaler statistischer Test. Für eine strenge Prüfung gibt es spezielle Tests (z.B. Shapiro-Wilk), die aber ausserhalb des Lernziels dieser Einheit liegen.
