Grundlagen

Normalverteilung / Häufigkeitsverteilung / Histogramm

In einer Häufigkeitsverteilung wird die Häufigkeit gleicher Werte dargestellt. Angenommen es liegen in Spalte A die dargestellten Werte vor, bei denen es sich z.B. um Durchmesser einer gedrehten Welle handelt. Alle Werte gleicher Größe werden gezählt und die Häufigkeiten in die folgende Spalte geschrieben.

A	B
9,98	1
9,99	
9,99	2
10	
10	
10	3
10,01	
10,01	2
10,02	1

Zusammengefasst ergibt sich also folgende Tabelle:

A	B
9,98	1
9,99	2
10,00	3
10,01	2
10,02	1

Der Mittelwert $\bar{x}$ berechnet sich über $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

und die Standardabweichung s mit $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$

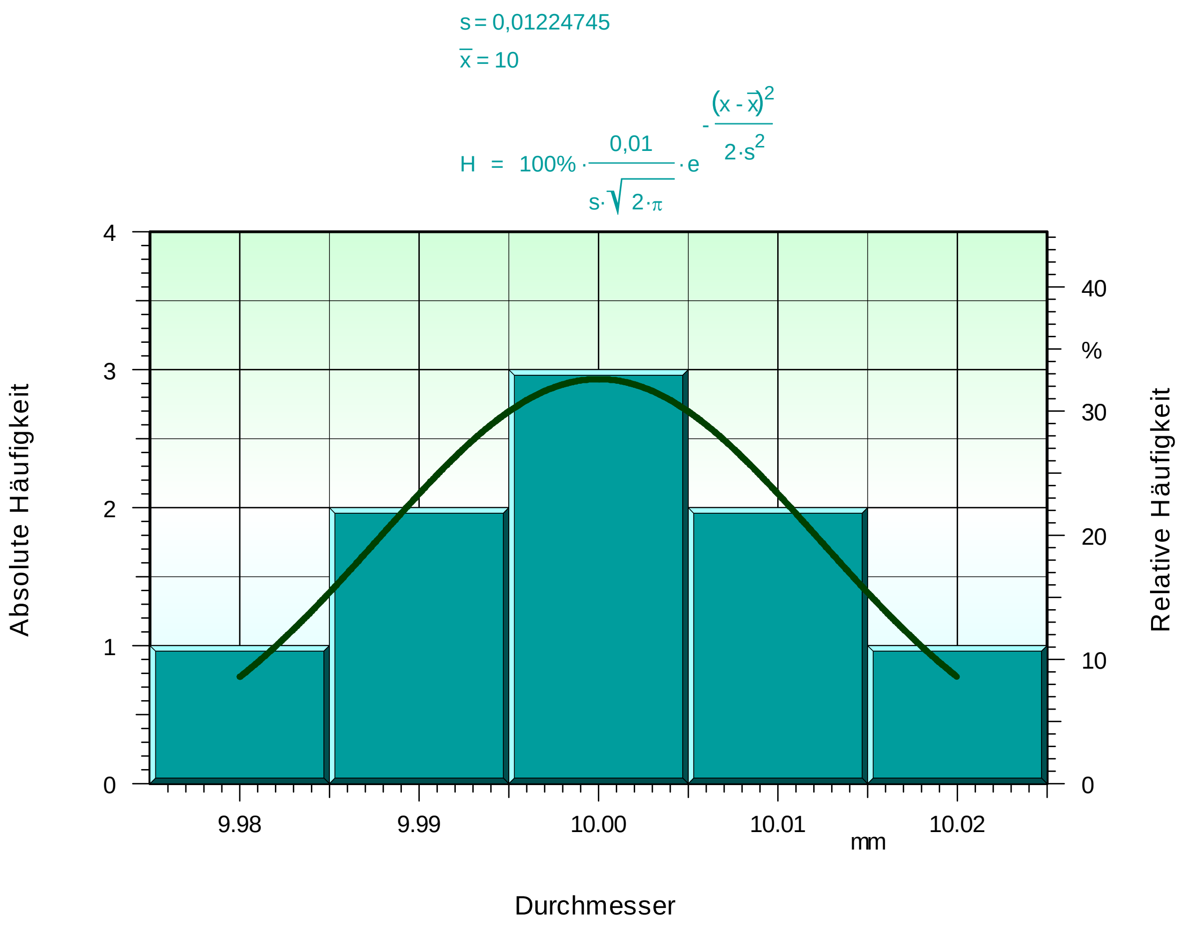
wobei n hier für die Anzahl der Daten steht. Mit diesen Angaben kann die sogenannte Gauß- oder Normalverteilung ermittelt werden, die als Kurve dargestellt wird (sogenannte Glockenkurve). Die Normalverteilung hat in der Praxis eine große Bedeutung. Sie ist der mathematisch idealisierte Grenzfall, der sich immer dann einstellt, wenn sich viele voneinander unabhängige Zufallseinflüsse addieren. Allgemein lautet die Dichtefunktion:

$$H = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-\bar{x})^2}{2s^2}} \quad \text{und für klassierte Daten} \quad H = K \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-\bar{x})^2}{2s^2}}$$

mit

H : Häufigkeit (normiert auf 1, in % mal 100)
s : Standardabweichung
$\bar{x}$: Mittelwert
K : Klassenbreite

Für die Approximation an klassierten Daten muss die Dichtefunktion um die Klassenbreite erweitert werden, damit die entsprechenden Einzelhäufigkeiten, bezogen auf die Einheiten, richtig berücksichtigt werden.



Für die Darstellung müssen die Daten aufsteigen sortiert werden. Oft ergeben sich in der Praxis Messreihen, deren Daten sehr eng zusammenliegen. Genau gleich große Werte kommen nur selten oder gar nicht vor. Die Häufigkeitsverteilung würde jeden Wert nur einmal feststellen. In diesem Fall verwendet man eine Klassierung, d.h. man bildet Bereiche innerhalb derer sich die Daten befinden und erzeugt so eine Verbesserung der Häufigkeiten. Die Klassierung erfolgt nach der Formel:

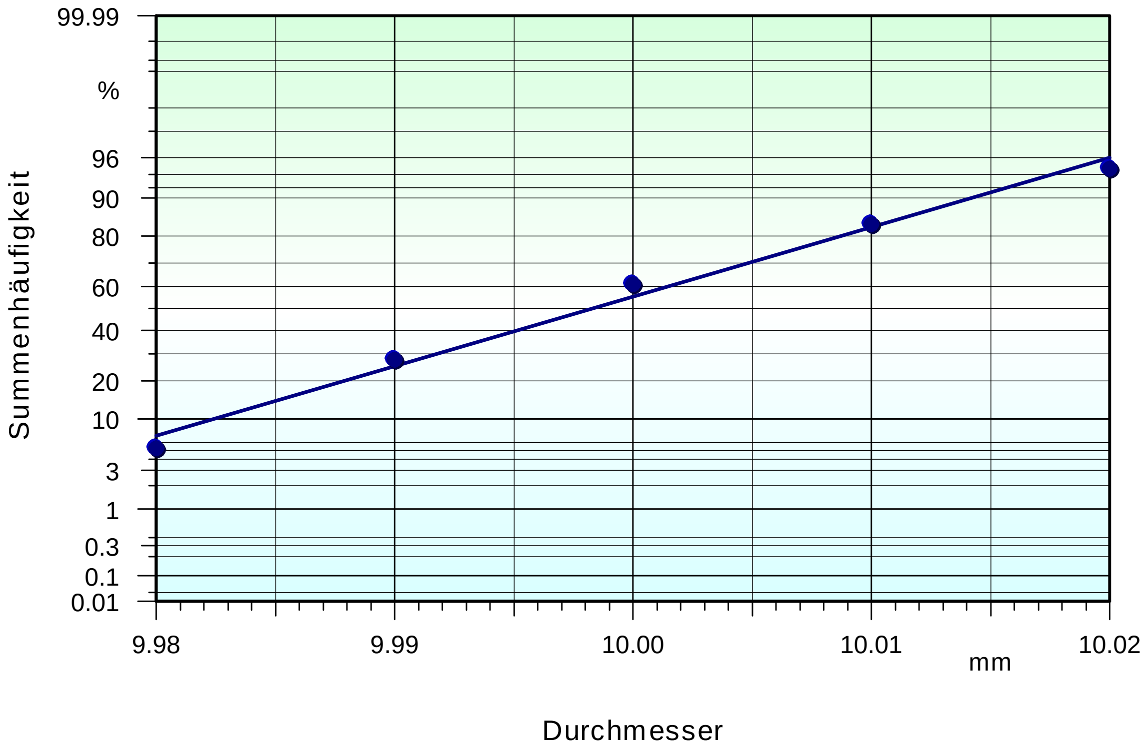
*Wert = Runden(Wert/Klassenbreite)*Klassenbreite*

Summenhäufigkeit / Wahrscheinlichkeitsnetz

Die Summenhäufigkeit oder das Wahrscheinlichkeitsnetz stellt die Summe der Häufigkeiten vom kleinsten Wert bis zum betrachteten Punkt x dar. Die Summenkurve ist das Integral der Dichtefunktion. Für die Normalverteilung gilt:

$$H = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-\bar{x})^2}{2s^2}}$$

Konkrete Datenwerte trägt man in ihren Häufigkeiten über den zugehörigen oberen Klassengrenzen als Summe auf (Erläuterung Klassen siehe Häufigkeitsverteilung). Die eingegebenen Werte für das Beispiel aus der Häufigkeitsverteilung sehen im Wahrscheinlichkeitsnetz so aus:



Diese Darstellung hat gegenüber der Häufigkeitsverteilung den Vorteil, dass für jedes Intervall leicht abgelesen werden kann, wie viel Prozent der Messwerte in ihm liegen (Abschätzung von Fehleranteilen außerhalb der Toleranz). Außerdem kann sehr leicht gezeigt werden, wie gut die Werte normalverteilt sind, nämlich dann, wenn sie möglichst nahe oder am besten auf der Summenkurve liegen.

Die Bestimmung der Häufigkeiten im Wahrscheinlichkeitsnetz erfolgt durch /23/:

$$H = \frac{i - 0{,}535206}{n - 0{,}070413} \cdot 100\%$$ mit i = Ordnungszahl der sortierten Werte

oder näherungsweise mit:

$$H = \frac{2i - 1}{2n} \cdot 100\%$$

Hinweis: Die Summenhäufigkeiten durch diese Gleichungen ergeben nicht exakt die aufaddierten Einzelhäufigkeiten, da hier auf Wahrscheinlichkeiten bezogen wird.
Normalerweise wird eine S-förmige Summenkurve zwischen den Punkten dargestellt. Dass diese hier eine Gerade ist, liegt daran, dass die Ordinate entsprechend logarithmisch verzerrt wurde.
Der Mittelwert (hier $\bar{x}$ = 10.0) liegt genau bei der Summenhäufigkeit von 50%. Der Bereich von $\bar{x} \pm s$ liegt bei 16% und 84% Häufigkeit.
In der Praxis wird oft die Summenhäufigkeit relativ zu den Streubereichen ±1s, ±2s und ±3s dargestellt. Dies bedeutet nichts anderes, als dass die X-Achse auf den Wert von s normiert und der Mittelwert auf 0 gesetzt wird.