Lösung zu Aufgabe 2: Histogramm

a) Histogramm der Verteilung:

Zunächst werden die gegebenen Messwerte in aufsteigender Reihenfolge sortiert:

i	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
x	4,574	4,589	4,593	4,599	4,601	4,607	4,608	4,609	4,611	4,612	4,613	4,614	4,616	4,619	4,620
i	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*
x	4,620	4,621	4,622	4,625	4,625	4,627	4,628	4,630	4,631	4,634	4,639	4,640	4,641	4,642	4,651

Damit ein Histogramm aussagekräftig wird, muss die Klassenbreite geeignet gewählt werden. Als Richtwert sollte man den Wertebereich in mindestens $\sqrt{n}$ *gleich große Teile aufteilen.* Damit ergibt sich für die Klassenanzahl k die Forderung:

$$k \geq \sqrt{n}$$

Mit n = 30 lautet diese Forderung im vorliegenden Fall also:

$$k \geq \sqrt{30} \approx 5,477$$

Da die Klassenanzahl k ganzzahlig sein muss, wird also gefordert:

$$k \geq 6$$

Der von den Klassen des Histogramms umfasste Wertebereich muss alle Einzelwerte der Messreihe einschließen. Eine Abschätzung der Klassenbreite Δx kann daher mit Hilfe der minimalen Klassenanzahl k sowie der Spannweite der Verteilung als Differenz von kleinstem Wert $x_{i_{min}}$ und größtem Wert $x_{i_{max}}$ gemäß folgender Gleichung vorgenommen werden:

$$\Delta x \leq \frac{x_{i_{max}} - x_{i_{min}}}{k}$$

Aus den vorliegenden Messwerten ergibt sich:

$$x_{i_{min}} = 4,574$$

$$x_{i_{max}} = 4,651$$

$$\Delta x \leq \frac{4,651 - 4,574}{6} = 0,0128\overline{3}$$

Um die Erstellung des Histogramms zu erleichtern, ist es sinnvoll, die so berechnete Klassenbreite auf einen sinnvollen Wert zu runden. Im vorliegenden Fall wählen wir:

$$\Delta x = 0,01$$

Als Anfangswert x_0 des Histogramms, d.h. als Untergrenze der ersten Klasse, wird ein Wert gewählt, für den gilt:

$$x_0 < x_{i_{min}}$$

Im vorliegenden Fall bietet es sich an, den Wert von $x_{i_{min}} = 4{,}574$ auf den Wert $x_0 = 4{,}57$ abzurunden.

Nun werden alle Messwerte x_i jeweils jener Klasse m zugeordnet, innerhalb deren Klassengrenzen sie ihrem Wert nach liegen. Ein Messwert x_i wird also der Klasse m zugeordnet, wenn gilt:

$$x_m < x_i \le x_{m+1}$$

Die Klassengrenzen ergeben sich dabei ausgehend vom Anfangswert x_0 zu ganzzahligen Vielfachen der Klassenbreite Δx:

$$x_m = x_0 + m \cdot \Delta x \quad \text{mit} \quad m \in \{0,1,2,...,k\}$$

Es ergeben sich somit die in der 1. Spalte der nachfolgenden Tabelle aufgeführten Klassengrenzen sowie die in der 2. Spalte aufgetragenen absoluten Häufigkeiten Δn_m.

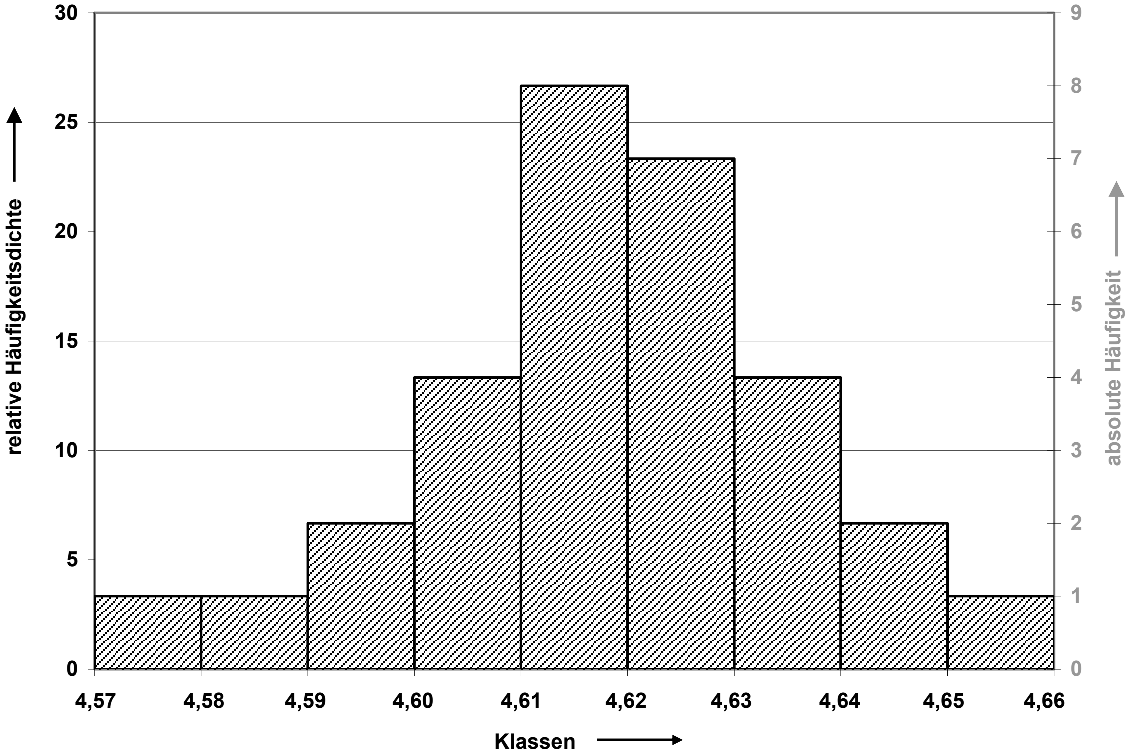
Klasse	Δn_m	$\frac{\Delta n_m}{n}$	$\frac{\Delta n_m}{n \cdot \Delta x}$	$\sum_{-\infty}^{m} \frac{\Delta n_m}{n}$
$4{,}57 < x_i \le 4{,}58$	1	$0{,}03\overline{3}$	$3{,}3\overline{3}$	$0{,}03\overline{3}$
$4{,}58 < x_i \le 4{,}59$	1	$0{,}03\overline{3}$	$3{,}3\overline{3}$	$0{,}06\overline{6}$
$4{,}59 < x_i \le 4{,}60$	2	$0{,}06\overline{6}$	$6{,}6\overline{6}$	$0{,}13\overline{3}$
$4{,}60 < x_i \le 4{,}61$	4	$0{,}13\overline{3}$	$13{,}3\overline{3}$	$0{,}26\overline{6}$
$4{,}61 < x_i \le 4{,}62$	8	$0{,}26\overline{6}$	$26{,}6\overline{6}$	$0{,}53\overline{3}$
$4{,}62 < x_i \le 4{,}63$	7	$0{,}23\overline{3}$	$23{,}3\overline{3}$	$0{,}76\overline{6}$
$4{,}63 < x_i \le 4{,}64$	4	$0{,}13\overline{3}$	$13{,}3\overline{3}$	$0{,}9$
$4{,}64 < x_i \le 4{,}65$	2	$0{,}06\overline{6}$	$6{,}6\overline{6}$	$0{,}96\overline{6}$
$4{,}65 < x_i \le 4{,}66$	1	$0{,}03\overline{3}$	$3{,}3\overline{3}$	1
$\sum$	30	1		

In dem zu erstellenden Histogramm soll die Fläche der einzelnen Balken als Maß für die relative Häufigkeit $\frac{\Delta n_m}{n}$ innerhalb der jeweiligen Klasse m dienen.

Daher werden die ermittelten absoluten Häufigkeiten Δn_m zunächst auf den Gesamtumfang $n = 30$ der Messreihe bezogen, wodurch sich die in der 3. Spalte der obigen Tabelle eingetragenen Zahlenwerte ergeben.

Die für die Erstellung des Histogramms benötigte Balkenhöhe ergibt sich, indem die relative Häufigkeit durch die Klassenbreite Δx dividiert wird. Das so erhaltene Maß $\frac{\Delta n_m}{n \cdot \Delta x}$ wird als relative Häufigkeitsdichte h_m bezeichnet. Die entsprechenden Zahlenwerte sind in der 4. Spalte der obigen Tabelle eingetragen.

Das resultierende Histogramm ist nachfolgend dargestellt:



Wie anhand der in obiger Tabelle aufgeführten Werte überprüft werden kann, ist die Fläche unter dem Histogramm, also die Summe der relativen Häufigkeiten, stets gleich 1.

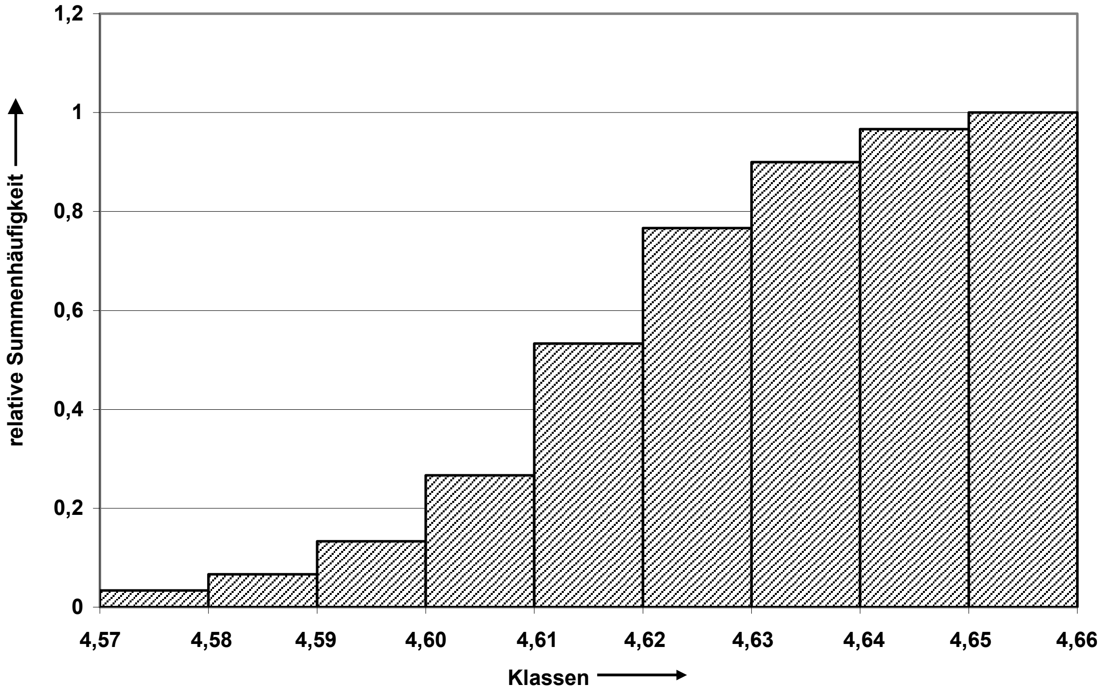
Prinzipiell ist es nicht erforderlich, dass alle Klassen des Histogramms die gleiche Breite Δx aufweisen. Ist dies jedoch wie in obigem Beispiel der Fall, ist die relative Häufigkeitsdichte der absoluten Häufigkeit proportional, da der Faktor $\frac{1}{n \cdot \Delta x}$ konstant ist. Abgesehen von der Beschriftung der y-Achse zeigt das Histogramm daher dasselbe Erscheinungsbild, wenn die absoluten Häufigkeiten innerhalb der einzelnen Klassen als Balkenhöhen aufgetragen werden (vgl. obige Abbildung).

b) relative Summenhäufigkeit der Verteilung:

Die relative Häufigkeit aller Messwerte x_i *mit* $x_i \leq x_{m+1}$ *wird als relative Summenhäufigkeit* S_m *bezeichnet:*

$$S_m = \sum_{-\infty}^{m} h_m \cdot \Delta x = \sum_{-\infty}^{m} \frac{\Delta n_m}{n}$$

Die Summation liefert für die vorliegende Verteilung die in der 5. Spalte der obigen Tabelle eingetragenen Zahlenwerte. Eine grafische Darstellung der relativen Summenhäufigkeit liefert nachfolgende Abbildung.



c) Schätzung von Erwartungswert und Standardabweichung der Grundgesamtheit

Bei der obigen Messreihe handelt es sich um eine Stichprobe vom Umfang n. Eine Stichprobe wird stets aus einer größeren Menge von Entitäten, der sogenannten Grundgesamtheit, entnommen. In Abhängigkeit vom Stichprobenumfang n wird die Stichprobe ein mehr oder weniger präzises Abbild der Verteilung der Grundgesamtheit liefern. Bei hinreichend groß gewähltem Stichprobenumfang n wird die Stichprobe repräsentativ für das statistische Verhalten der zugrundeliegenden Grundgesamtheit. Im Histogramm äußert sich dies beispielsweise dadurch, dass sich bei einer weiteren Erhöhung des Stichprobenumfangs n die Form des Histogramms nur noch geringfügig ändert.

Entsprechend kann eine Stichprobe dazu genutzt werden, die Lage- und Streuungsparameter der Verteilung der ihr zugrundeliegenden Grundgesamtheit abzuschätzen. Zu den Parametern Erwartungswert μ und Standardabweichung σ der Grundgesamtheit stellen der arithmetische Mittelwert $\overline{x}$ und die empirische Streuung S einer aus ihr entnommenen Stichprobe die besten Schätzwerte dar.

Der arithmetische Mittelwert $\overline{x}$ *von* n *Messwerten errechnet sich gemäß:*

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Die empirische Streuung S *von* n *Messwerten errechnet sich gemäß:*

$$S = +\sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

Im vorliegenden Fall ergibt sich:

$$\overline{x} = 4{,}6187$$

$$S \approx 0{,}01709$$

Da es sich bei den so berechneten Werten wie erwähnt nur um Schätzwerte für μ und σ handelt, stellt sich die Frage nach der Güte bzw. Unsicherheit dieser Schätzung. Ein Maß für die Unsicherheit dieser (besten) Schätzwerte liefert das sogenannte Konfidenzintervall, welches unter anderem vom Stichprobenumfang n abhängt. Die Bestimmung dieses Konfidenzintervalls wird im Rahmen nachfolgender Übungsaufgaben behandelt.