

Histogramm

St. Lange¹, R. Bender²

¹ Abteilung für Medizinische Informatik, Biometrie und Epidemiologie der Ruhr-Universität Bochum

² Fakultät für Gesundheitswissenschaften, AG Epidemiologie und medizinische Statistik, Universität Bielefeld

Das Histogramm ist eine einfache Möglichkeit, einen Datensatz graphisch darzustellen. Hierbei werden die absoluten oder relativen Häufigkeiten, mit denen eine Merkmalsausprägung beobachtet wurde, als Rechtecke abgebildet. Liegen viele verschiedene Merkmalsausprägungen vor, ist es sinnvoll, mehrere Ausprägungen in Klassen zusammenzufassen. Wenngleich es für die Breite und Anzahl der Klassen grobe Empfehlungen gibt (zum Beispiel für die Anzahl der Klassen die Wurzel aus dem Stichprobenumfang: » $\sqrt{n}$ -Regel«), genügen im Allgemeinen, abgesehen von sehr großen Stichprobenumfängen, 8 bis 15 Klassen. Wichtiger als solche Empfehlungen sind auf jeden Fall inhaltliche Überlegungen, insbesondere auch im Hinblick auf die Wahl der Klassengrenzen. Bei der Darstellung von Altersverteilungen ist es beispielsweise üblich, Dekaden oder 5-Jahres-Intervalle zu verwenden, und die Klasse bei 5er oder 10er Jahreszahlen zu beginnen. Klassenbreiten von 7,74 Jahren, die sich bei strikter Anwendung der » $\sqrt{n}$ -Regel« bei einem Stichprobenumfang von 60 ergäben, sind eher nicht adäquat.

Tab.1 Altersverteilung von Personen, die in einem Londoner Stadtbezirk einen häuslichen Unfall erlitten haben.

Altersgruppe (Jahre)	Häufigkeit
[0–5)	206
[5–15)	154
[15–45)	247
[45–65)	111
[65–90]	95
Σ	813

Die eckigen Klammern bei den Intervallangaben bedeuten, dass der entsprechende Wert noch zum Intervall gehört, die runden Klammern bedeuten, dass der Wert nicht mehr zum Intervall gehört

In einem Histogramm haben die Klassen zumeist die gleiche Breite. Es kann aber auch sinnvoll sein, nicht oder nur schwach besetzte Klassen zusammenzufassen. In dieser Situation mit unterschiedlich breiten Klassen muss allerdings beachtet werden, dass bei einem Histogramm nicht die Höhe, sondern die Fläche der Rechtecke der beobachteten (absoluten oder relativen) Häufigkeit der jeweiligen Klasse entspricht. Um die richtigen Relationen herzustellen, muss bei unterschiedlichen Klassenbreiten dann die Höhe jeder Klasse durch die Breite dieser Klasse geteilt werden.

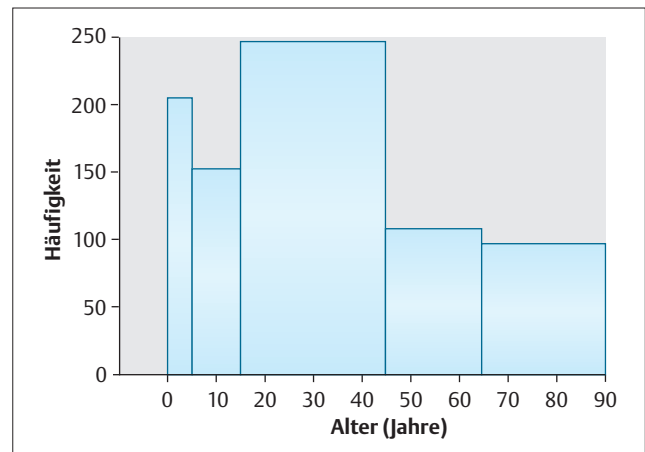


Abb.1 Fälschliches »Histogramm« bei ungleichen Klassenbreiten für die Daten aus Tabelle 1: Die Höhe der Rechtecke wird als Maß für die Häufigkeit verwendet.

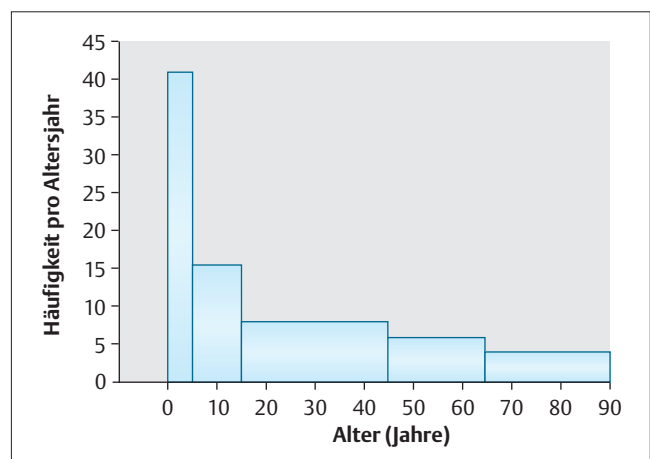


Abb.2 Richtiges Histogramm bei ungleichen Klassenbreiten für die Daten aus Tabelle 1: Die Fläche der Rechtecke wird als Maß für die Häufigkeit verwendet.

In **Tab.1** ist die Altersverteilung von Personen angegeben, die in einem Londoner Stadtbezirk einen häuslichen Unfall erlitten haben. Die meisten Opfer sind Erwachsene mit der größten Zahl von Unfallopfern im Altersbereich zwischen 15 und 44 Jahren. Die Breite der Altersklassen variiert zwischen 5 und 30 Jahren. Eine Darstellung, in der die Höhe der Rechtecke als Maß für die Häufigkeit verwendet wird, ergibt den Eindruck, dass Kleinkinder im Alter bis 5 Jahre bei den Opfern

nur wenig vertreten sind, was eigentlich nicht zu erwarten ist (**Abb. 1**). Die richtige Darstellung, bei der die Fläche der Häufigkeit entspricht (**Abb. 2**), und die Zahl von Unfallopfern pro Altersjahr (Höhe dividiert durch Breite) angegeben ist, korrigiert den fälschlichen Eindruck: Das größte Risiko haben Kleinkinder.

Tab. 2 zeigt die Übersetzungen.

Tab. 2 Übersetzungen (deutsch – englisch)

Histogramm	histogram
Absolute (relative) Häufigkeit	absolute (relative) frequency

kurzgefasst: In einem Histogramm werden Häufigkeiten als Rechtecke abgebildet. Zumeist haben hierbei die Rechtecke der verschiedenen Klassen die gleiche Breite. Sollte dies nicht der Fall sein, muss man die Höhe jeder Klasse durch die Breite teilen, um richtig interpretieren zu können.

Korrespondenz

Dr. Stefan Lange
Abteilung für Medizinische Informatik, Biometrie und
Epidemiologie
Ruhr-Universität
Universitätsstraße 150
44780 Bochum
E-Mail: stefan.f.lange@ruhr-uni-bochum.de