Karlsruher Institut für Technologie

Institut für Volkswirtschaftslehre (ECON)

Lehrstuhl für Ökonometrie und Statistik

# Kapitel V - Graphische Darstellung von Häufigkeitsverteilungen

## Deskriptive Statistik

Prof. Dr. W.-D. Heller
Hartwig Senska
Carlo Siebenschuh

# Agenda

1. **Graphische Darstellungsmöglichkeiten**
2. Histogramm
3. Summenhäufigkeitsfunktion

# Graphische Darstellungsmöglichkeiten

**Ziele graphischer Darstellungen**

→ Größere Übersichtlichkeit

→ Größere Einprägsamkeit

→ Größere Attraktivität

**Aufgabe**

→ Zahlen (bei absoluten Häufigkeiten) bzw. Anteile (bei relativen Häufigkeiten) graphisch so repräsentieren, dass ihre Größe aus der zeichnerischen Darstellung verglichen werden kann.

# Graphische Darstellungsmöglichkeiten

## Ziele graphischer Darstellungen

→ Größere Übersichtlichkeit

→ Größere Einprägsamkeit

→ Größere Attraktivität

## Aufgabe

→ Zahlen (bei absoluten Häufigkeiten) bzw. Anteile (bei relativen Häufigkeiten) graphisch so repräsentieren, dass ihre Größe aus der zeichnerischen Darstellung verglichen werden kann.

# Graphische Darstellungsmöglichkeiten

**Ziele graphischer Darstellungen**

→ Größere Übersichtlichkeit

→ Größere Einprägsamkeit

→ Größere Attraktivität

**Aufgabe**

→ Zahlen (bei absoluten Häufigkeiten) bzw. Anteile (bei relativen Häufigkeiten) graphisch so repräsentieren, dass ihre Größe aus der zeichnerischen Darstellung verglichen werden kann.

# Graphische Darstellungsmöglichkeiten

**Ziele graphischer Darstellungen**

→ Größere Übersichtlichkeit

→ Größere Einprägsamkeit

→ Größere Attraktivität

**Aufgabe**

→ Zahlen (bei absoluten Häufigkeiten) bzw. Anteile (bei relativen Häufigkeiten) graphisch so repräsentieren, dass ihre Größe aus der zeichnerischen Darstellung verglichen werden kann.

# Graphische Darstellungsmöglichkeiten

Beispiel 5.1 - Vergleicht man die Zahlen 1 und 8, so kann man sie z.B. darstellen

(1) als 1 cm bzw. 8 cm lange Strecken (**eindimensional**)

(2) als Flächen mit den Flächeninhalten 1 $cm^2$ bzw. 8 $cm^2$ (**zweidimensional**)

(3) als Volumina mit den Volumeninhalten 1 $cm^3$ bzw. 8 $cm^3$ (**dreidimensional**)

(4) als Bildsymbole (**Piktogramm**)

# Graphische Darstellungsmöglichkeiten

Beispiel 5.1 - Vergleicht man die Zahlen 1 und 8, so kann man sie z.B. darstellen

(1) als 1 cm bzw. 8 cm lange Strecken (**eindimensional**)

(2) als Flächen mit den Flächeninhalten 1 $cm^2$ bzw. 8 $cm^2$ (**zweidimensional**)

(3) als Volumina mit den Volumeninhalten 1 $cm^3$ bzw. 8 $cm^3$ (**dreidimensional**)

(4) als Bildsymbole (**Piktogramm**)

# Graphische Darstellungsmöglichkeiten

Beispiel 5.1 - Vergleicht man die Zahlen 1 und 8, so kann man sie z.B. darstellen

(1) als 1 cm bzw. 8 cm lange Strecken (**eindimensional**)

(2) als Flächen mit den Flächeninhalten 1 $cm^2$ bzw. 8 $cm^2$ (**zweidimensional**)

(3) als Volumina mit den Volumeninhalten 1 $cm^3$ bzw. 8 $cm^3$ (**dreidimensional**)

(4) als Bildsymbole (**Piktogramm**)

# Graphische Darstellungsmöglichkeiten

Beispiel 5.1 - Vergleicht man die Zahlen 1 und 8, so kann man sie z.B. darstellen

(1) als 1 cm bzw. 8 cm lange Strecken (**eindimensional**)

(2) als Flächen mit den Flächeninhalten 1 $cm^2$ bzw. 8 $cm^2$ (**zweidimensional**)

(3) als Volumina mit den Volumeninhalten 1 $cm^3$ bzw. 8 $cm^3$ (**dreidimensional**)

(4) als Bildsymbole (**Piktogramm**)

# Graphische Darstellungsmöglichkeiten

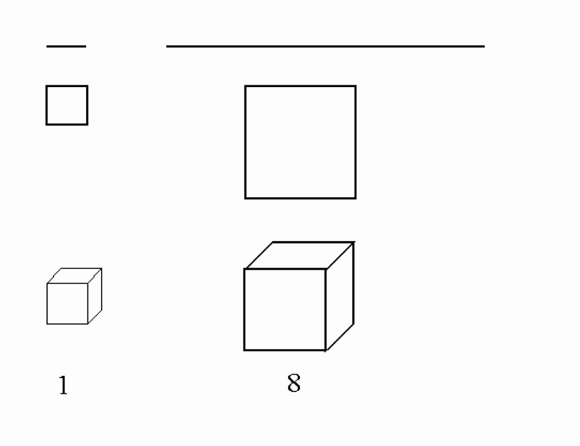


Abbildung 5.1 - Repräsentation von Zahlen durch Strecken, Flächen und Volumina

# Graphische Darstellungsmöglichkeiten

Abbildung 5.2 - Repräsentation von Zahlen durch Bildsymbole

# Graphische Darstellungsmöglichkeiten

Stab-, Linien- oder Säulenendiagramm (Balkendiagramm): Darstellung der Zahlen durch Stäbe bzw. Säulen über einer horizontalen Achse, auf der die Merkmalsausprägungen aufgetragen sind. Die Höhe der Stäbe bzw. Säulen spiegelt die absoluten bzw. relativen Häufigkeiten wider. ($\rightarrow$höhenproportional)

Flächendiagramm: Darstellung der Häufigkeiten durch Flächen entsprechender Größe ($\rightarrow$flächenproportional)

Volumenproportionale Darstellung: Darstellung der Häufigkeiten durch Volumina entsprechender Größe ($\rightarrow$volumenproportional)

# Graphische Darstellungsmöglichkeiten

Stab-, Linien- oder Säulenendiagramm (Balkendiagramm): Darstellung der Zahlen durch Stäbe bzw. Säulen über einer horizontalen Achse, auf der die Merkmalsausprägungen aufgetragen sind. Die Höhe der Stäbe bzw. Säulen spiegelt die absoluten bzw. relativen Häufigkeiten wider. ($\rightarrow$höhenproportional)

Flächendiagramm: Darstellung der Häufigkeiten durch Flächen entsprechender Größe ($\rightarrow$flächenproportional)

Volumenproportionale Darstellung: Darstellung der Häufigkeiten durch Volumina entsprechender Größe ($\rightarrow$volumenproportional)

# Graphische Darstellungsmöglichkeiten

Stab-, Linien- oder Säulenendiagramm (Balkendiagramm): Darstellung der Zahlen durch Stäbe bzw. Säulen über einer horizontalen Achse, auf der die Merkmalsausprägungen aufgetragen sind. Die Höhe der Stäbe bzw. Säulen spiegelt die absoluten bzw. relativen Häufigkeiten wider. ($\rightarrow$höhenproportional)

Flächendiagramm: Darstellung der Häufigkeiten durch Flächen entsprechender Größe ($\rightarrow$flächenproportional)

Volumenproportionale Darstellung: Darstellung der Häufigkeiten durch Volumina entsprechender Größe ($\rightarrow$volumenproportional)

# Graphische Darstellungsmöglichkeiten

Beispiel 5.2 - Die Spieler eines Bundesligavereins (17 Spieler) wurden nach ihrem letzten Schul- bzw. Hochschulabschluss[1] befragt:

| Spieler | $s_1$ | $s_2$ | $s_3$ | $s_4$ | $s_5$ | $s_6$ | $s_7$ | $s_8$ | $s_9$ | $s_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Abschluss | A | H | H | H | U | R | B | H | B | B |

| Spieler | $s_{11}$ | $s_{12}$ | $s_{13}$ | $s_{14}$ | $s_{15}$ | $s_{16}$ | $s_{17}$ |
|---|---|---|---|---|---|---|---|
| Abschluss | R | R | FO | H | A | A | R |

Zugehörige Urliste ist also:

$$A, H, H, H, U, R, B, H, B, B, R, R, FO, H, A, A, R.$$

[1] H Hauptschule, B Berufsschule, R Realschule, A Abitur, FO Fachoberschule, FH Fachhochschule, U Universität

# Graphische Darstellungsmöglichkeiten

Beispiel 5.2 - Die Spieler eines Bundesligavereins (17 Spieler) wurden nach ihrem letzten Schul- bzw. Hochschulabschluss[1] befragt:

| Spieler | $s_1$ | $s_2$ | $s_3$ | $s_4$ | $s_5$ | $s_6$ | $s_7$ | $s_8$ | $s_9$ | $s_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Abschluss | A | H | H | H | U | R | B | H | B | B |

| Spieler | $s_{11}$ | $s_{12}$ | $s_{13}$ | $s_{14}$ | $s_{15}$ | $s_{16}$ | $s_{17}$ |
|---|---|---|---|---|---|---|---|
| Abschluss | R | R | FO | H | A | A | R |

Zugehörige Urliste ist also:

$$A, H, H, H, U, R, B, H, B, B, R, R, FO, H, A, A, R.$$

[1] H Hauptschule, B Berufsschule, R Realschule, A Abitur, FO Fachoberschule, FH Fachhochschule, U Universität

# Graphische Darstellungsmöglichkeiten

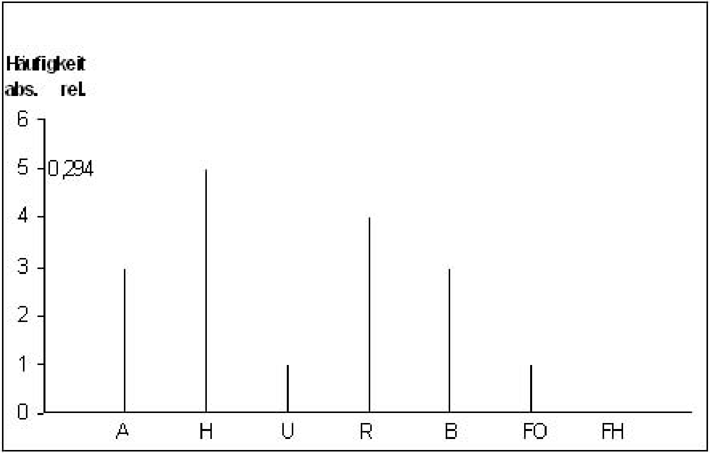


Abbildung 5.3 - Stabdiagramm zu Beispiel 5.2

# Graphische Darstellungsmöglichkeiten

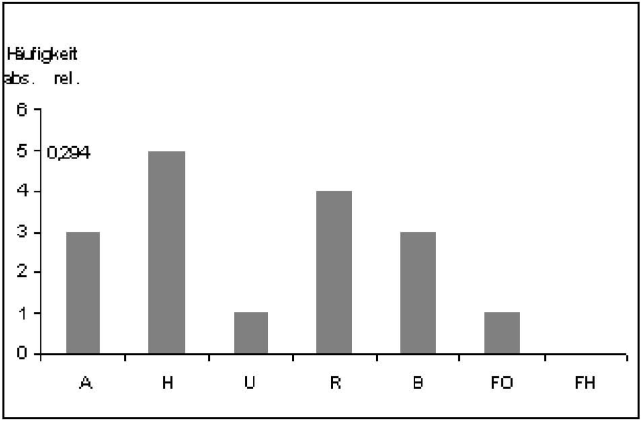


Abbildung 5.4 - Säulendiagramm zu Beispiel 5.2

# Graphische Darstellungsmöglichkeiten


Abbildung 5.5 - Säulendiagramm mit runden Säulen zu Beispiel 5.2

# Graphische Darstellungsmöglichkeiten

Anmerkung: Obwohl in Abbildung 5.4 bzw. 5.5 Flächen bzw. Volumina dargestellt sind, wird intuitiv nur die Höhe der Säule zum Vergleich herangezogen. Die Grundseite bzw. -fläche stimmt bei allen Säulen überein. Die weiteren Dimensionen dienen lediglich der besseren Erkennbarkeit gegenüber dem Stabdiagramm.

# Graphische Darstellungsmöglichkeiten

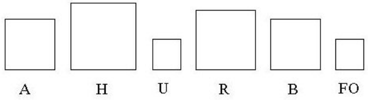


Abbildung 5.6 - Flächendiagramm zu Beispiel 5.2

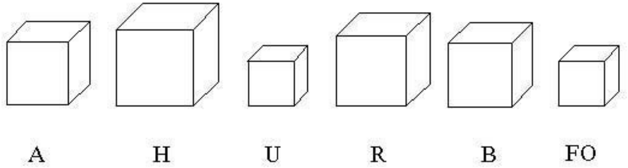


Abbildung 5.7 - Volumendiagramm zu Beispiel 5.2

# Graphische Darstellungsmöglichkeiten

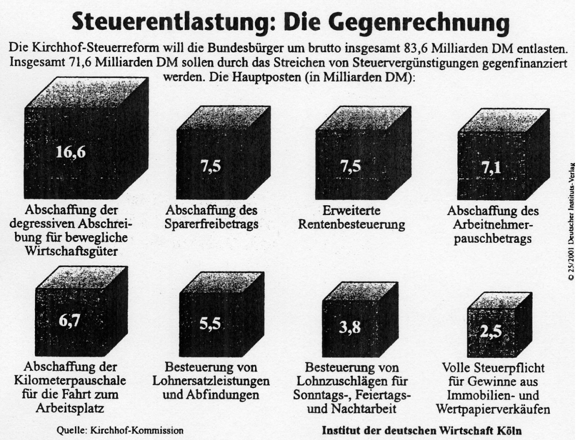


Abbildung 5.8 - Weitere Volumenproportionale Darstellung

# Graphische Darstellungsmöglichkeiten

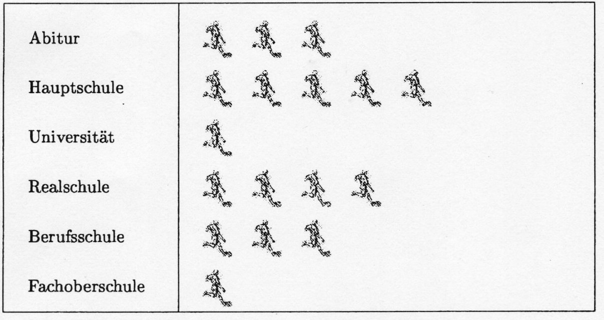


Abbildung 5.9 - Piktogramm zu Beispiel 5.2

# Graphische Darstellungsmöglichkeiten

Kreissektorendiagramm: Graphische Darstellung *relativer Häufigkeiten* durch sektorale Aufteilung einer Kreisfläche oder Kreisscheibe, wobei die Größe der Sektoren proportional zu den relativen Häufigkeiten ist.

Anmerkung: Dies ist nur eine Möglichkeit, eine geometrische Einheit (Strecke, Fläche, Volumen) entsprechend den Anteilen (relativen Häufigkeiten) aufzuteilen. Bei der Darstellung der relativen Häufigkeiten können ebenfalls Stab- und Säulendiagramme oder Flächendiagramme verwendet werden. Werden verschiedene statistische Massen hinsichtlich desselben Merkmals untersucht, so kann die unterschiedliche Größe der Massen bei der Größe der Kreisflächen berücksichtigt werden. Der Radius des Kreises ist dann proportional zur Wurzel aus der jeweiligen Anzahl statistischer Einheiten.

# Graphische Darstellungsmöglichkeiten

Kreissektorendiagramm: Graphische Darstellung *relativer Häufigkeiten* durch sektorale Aufteilung einer Kreisfläche oder Kreisscheibe, wobei die Größe der Sektoren proportional zu den relativen Häufigkeiten ist.

Anmerkung: Dies ist nur eine Möglichkeit, eine geometrische Einheit (Strecke, Fläche, Volumen) entsprechend den Anteilen (relativen Häufigkeiten) aufzuteilen.
Bei der Darstellung der relativen Häufigkeiten können ebenfalls Stab- und Säulendiagramme oder Flächendiagramme verwendet werden.
Werden verschiedene statistische Massen hinsichtlich desselben Merkmals untersucht, so kann die unterschiedliche Größe der Massen bei der Größe der Kreisflächen berücksichtigt werden. Der Radius des Kreises ist dann proportional zur Wurzel aus der jeweiligen Anzahl statistischer Einheiten.

# Graphische Darstellungsmöglichkeiten


Abbildung 5.10 - Beispiel für ein Kreissektorendiagramm

# Graphische Darstellungsmöglichkeiten

Bemerkung: Die graphische Darstellung sollte die Struktur der Menge der Merkmalsausprägungen berücksichtigen. Dies bedeutet, dass bei ordinalskalierten und kardinalskalierten Merkmalen die natürliche Reihenfolge bei der Darstellung berücksichtigt werden muss. Bei einem **Rangmerkmal** oder einem **diskreten quantitativen Merkmal** erfolgt dies z.B. in einem Stab- oder Säulendiagramm durch die natürliche Anordnung der Merkmalsausprägungen in horizontaler Richtung.

# Graphische Darstellungsmöglichkeiten

Beispiel 5.3 - Im Leistungskurs Mathematik eines Jahrgangs eines Gymnasiums wurden folgende Ergebnisse erzielt:

| Schüler | $s_1$ | $s_2$ | $s_3$ | $s_4$ | $s_5$ | $s_6$ | $s_7$ | $s_8$ | $s_9$ | $s_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Punktzahl | 8 | 14 | 9 | 13 | 8 | 12 | 9 | 11 | 12 | 9 |

| Schüler | $s_{11}$ | $s_{12}$ | $s_{13}$ | $s_{14}$ | $s_{15}$ | $s_{16}$ | $s_{17}$ | $s_{18}$ | $s_{19}$ | $s_{20}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Punktzahl | 12 | 14 | 10 | 12 | 9 | 7 | 11 | 12 | 13 | 9 |

# Graphische Darstellungsmöglichkeiten

Beispiel 5.3 - Im Leistungskurs Mathematik eines Jahrgangs eines Gymnasiums wurden folgende Ergebnisse erzielt:

| Schüler | $s_1$ | $s_2$ | $s_3$ | $s_4$ | $s_5$ | $s_6$ | $s_7$ | $s_8$ | $s_9$ | $s_{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Punktzahl | 8 | 14 | 9 | 13 | 8 | 12 | 9 | 11 | 12 | 9 |

| Schüler | $s_{11}$ | $s_{12}$ | $s_{13}$ | $s_{14}$ | $s_{15}$ | $s_{16}$ | $s_{17}$ | $s_{18}$ | $s_{19}$ | $s_{20}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Punktzahl | 12 | 14 | 10 | 12 | 9 | 7 | 11 | 12 | 13 | 9 |

# Graphische Darstellungsmöglichkeiten

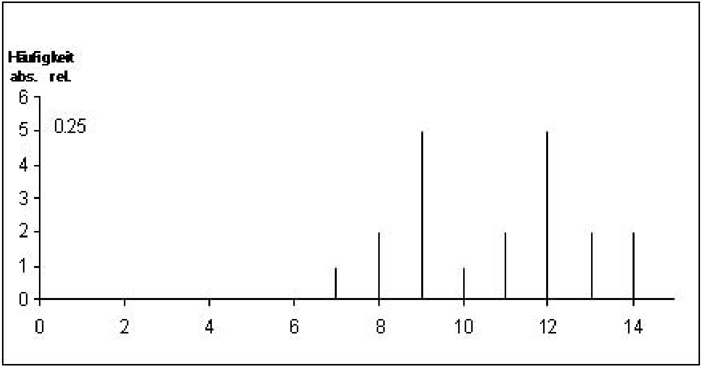


Abbildung 5.11 - Stabdiagramm zu Beispiel 5.3

# Graphische Darstellungsmöglichkeiten

Darstellung von Summenhäufigkeiten: Bei diskreten quantitativen Merkmalen genügt die Darstellung der Summenhäufigkeit als Punkt, zur besseren Übersichtlichkeit werden jedoch die Punkte durch waagerechte Linien nach rechts fortgesetzt.
($\rightarrow$Ähnlichkeit in Darstellung und Interpretation zur empirischen Verteilungsfunktion)

# Graphische Darstellungsmöglichkeiten

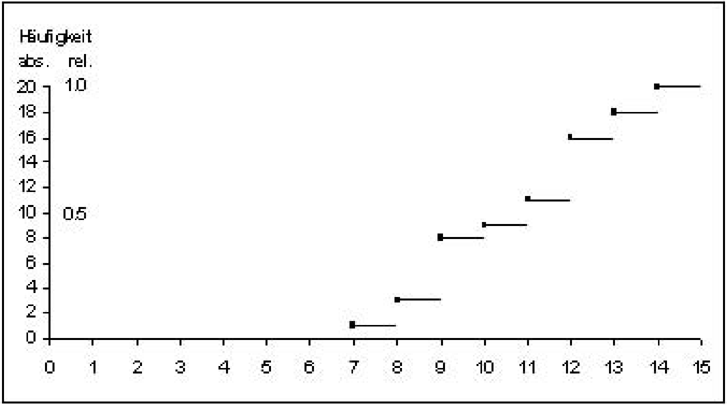


Abbildung 5.12 - Summenhäufigkeiten zu Beispiel 5.3

# Agenda

1. Graphische Darstellungsmöglichkeiten
2. **Histogramm**
3. Summenhäufigkeitsfunktion

# Histogramm

Bei der graphischen Darstellung von klassierten Merkmalen sind die Häufigkeiten für die jeweiligen Klassen aufzutragen. Jede Klasse entspricht einem Intervall auf der $x$-Achse, das durch die obere und untere Klassengrenze definiert ist. Jede obere Klassengrenze ist die untere Klassengrenze der nächsten Klasse, da jede Merkmalsausprägung in einer Klasse enthalten sein muß. Zur Darstellung der Häufigkeiten bildet man nun ein **Flächendiagramm** mit Rechtecken, wobei man als eine Seite des Rechtecks das Intervall der Klasse verwendet.

Achtung: Für offene Randklassen kann eine solche Darstellung nicht durchgeführt werden, sofern nicht eine festlegbare Klassenober- bzw. Klassenuntergrenze existiert. Ist der Anteil dieser Randklasse an der statistischen Masse gering, so kann man 0 als Höhe des Rechtecks wählen.

# Histogramm

Bei der graphischen Darstellung von klassierten Merkmalen sind die Häufigkeiten für die jeweiligen Klassen aufzutragen. Jede Klasse entspricht einem Intervall auf der $x$-Achse, das durch die obere und untere Klassengrenze definiert ist. Jede obere Klassengrenze ist die untere Klassengrenze der nächsten Klasse, da jede Merkmalsausprägung in einer Klasse enthalten sein muß. Zur Darstellung der Häufigkeiten bildet man nun ein **Flächendiagramm** mit Rechtecken, wobei man als eine Seite des Rechtecks das Intervall der Klasse verwendet.

Achtung: Für offene Randklassen kann eine solche Darstellung nicht durchgeführt werden, sofern nicht eine festlegbare Klassenober- bzw. Klassenuntergrenze existiert. Ist der Anteil dieser Randklasse an der statistischen Masse gering, so kann man 0 als Höhe des Rechtecks wählen.

# Histogramm

Sei $\Delta_I$ die Klassenbreite, $h(I)(p(I))$ die absolute (relative) Häufigkeit der Klasse $I$, so ist $\frac{h(I)}{\Delta_I}$ bzw. $\left(\frac{p(I)}{\Delta_I}\right)$ die Höhe des Rechtecks in einer geeigneten Längeneinheit.

Beispiel 5.4 - Bei den 40 Beschäftigten einer Behörde wurden folgende Altersangaben (in vollendeten Lebensjahren) ermittelt:

37, 58, 63, 17, 28, 46, 57, 26, 39, 47,
16, 62, 44, 39, 48, 27, 35, 59, 19, 26,
55, 36, 37, 48, 28, 46, 18, 62, 25, 37,
38, 45, 59, 61, 29, 36, 28, 42, 29, 37.

# Histogramm

Sei $\Delta_I$ die Klassenbreite, $h(I)(p(I))$ die absolute (relative) Häufigkeit der Klasse $I$, so ist $\frac{h(I)}{\Delta_I}$ bzw. $\left(\frac{p(I)}{\Delta_I}\right)$ die Höhe des Rechtecks in einer geeigneten Längeneinheit.

Beispiel 5.4 - Bei den 40 Beschäftigten einer Behörde wurden folgende Altersangaben (in vollendeten Lebensjahren) ermittelt:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 37, | 58, | 63, | 17, | 28, | 46, | 57, | 26, | 39, | 47, |
| 16, | 62, | 44, | 39, | 48, | 27, | 35, | 59, | 19, | 26, |
| 55, | 36, | 37, | 48, | 28, | 46, | 18, | 62, | 25, | 37, |
| 38, | 45, | 59, | 61, | 29, | 36, | 28, | 42, | 29, | 37. |

# Histogramm

Nach Klassierung ergibt sich die Häufigkeitstabelle:

| Alter | abs. Häufigkeit | rel. Häufigkeit |
|---|---|---|
| unter 20 | 4 | 0.1 |
| 20 bis unter 30 | 9 | 0.225 |
| 30 bis unter 40 | 10 | 0.25 |
| 40 bis unter 50 | 8 | 0.2 |
| 50 bis unter 60 | 5 | 0.125 |
| über 60 | 4 | 0.1 |

Anmerkung: Für die offenen Randklassen kann man als vertretbare untere bzw. obere Klassengrenze 15 bzw. 65 ansetzen, da in einer Behörde sicherlich keine Personen unter 15 und kaum Personen über 65 beschäftigt sein können.

# Histogramm

Nach Klassierung ergibt sich die Häufigkeitstabelle:

| Alter | abs. Häufigkeit | rel. Häufigkeit |
|---|---|---|
| unter 20 | 4 | 0.1 |
| 20 bis unter 30 | 9 | 0.225 |
| 30 bis unter 40 | 10 | 0.25 |
| 40 bis unter 50 | 8 | 0.2 |
| 50 bis unter 60 | 5 | 0.125 |
| über 60 | 4 | 0.1 |

Anmerkung: Für die offenen Randklassen kann man als vertretbare untere bzw. obere Klassengrenze 15 bzw. 65 ansetzen, da in einer Behörde sicherlich keine Personen unter 15 und kaum Personen über 65 beschäftigt sein können.

# Histogramm

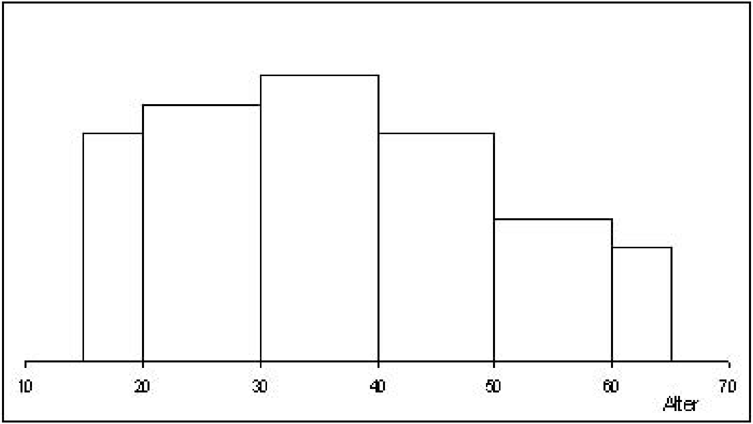


Abbildung 5.13 - Histogramm zu Beispiel 5.4

# Histogramm

Beispiel 5.5 - Für eine Verbrauchsstudie wurden die Nettojahreseinkommen von 100 Männern und 150 Frauen festgestellt:

| | Männer | | Frauen | |
|---|---|---|---|---|
| Einkommen in TEuro | abs. H. | rel. H. | abs. H. | rel. H. |
| unter 10 | 5 | 0.05 | 9 | 0.06 |
| 10 bis unter 20 | 15 | 0.15 | 30 | 0.20 |
| 20 bis unter 25 | 20 | 0.20 | 42 | 0.28 |
| 25 bis unter 30 | 25 | 0.25 | 48 | 0.32 |
| 30 bis unter 40 | 20 | 0.20 | 15 | 0.10 |
| 40 bis unter 60 | 10 | 0.10 | 6 | 0.04 |
| 60 bis unter 85 | 5 | 0.05 | 0 | 0 |
| $\sum$ | 100 | 1 | 150 | 1 |

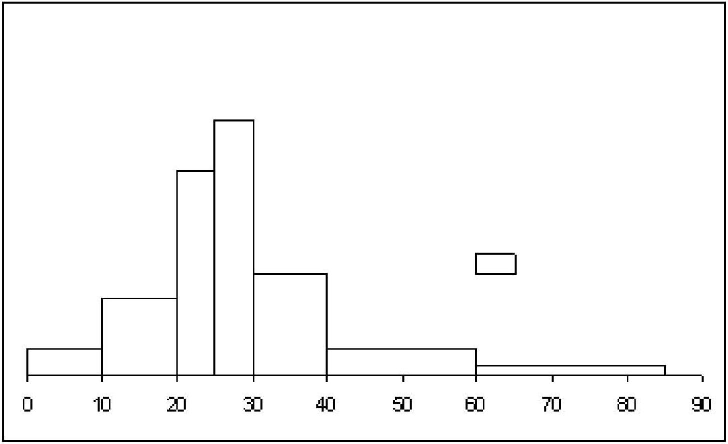


Abbildung 5.14 - Histogramm zur Einkommensverteilung der Männer

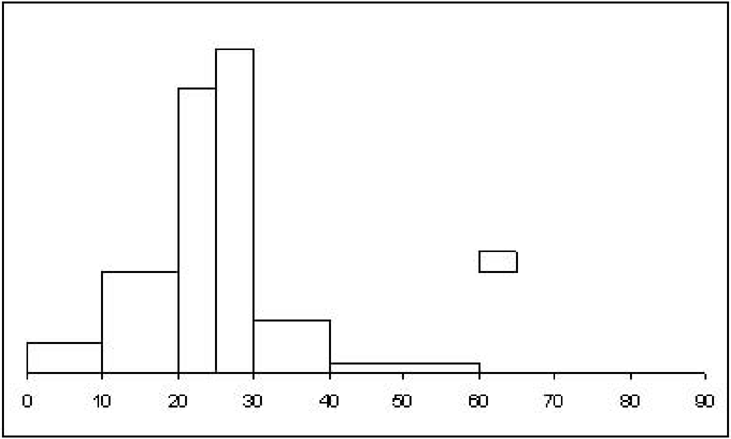


Abbildung 5.15 - Histogramm zur Einkommensverteilung der Frauen

# Histogramm

Anmerkung: Sollen zwei Häufigkeitsverteilungen an Hand der Histogramme verglichen werden, deren Grundgesamtheiten einen unterschiedlichen Umfang haben, so sind die relativen Häufigkeitsdichten zu verwenden. Auf diese Weise wird erreicht, dass eine Flächeneinheit demselben Anteil an der Grundgesamtheit entspricht. In diesem Fall können die Histogramme wie in Abbildung 5.16 übereinandergelegt werden.

**Häufigkeitsdichte**
Sie entspricht der Höhe des Rechtecks in einem Histogramm:

$$\frac{h(I)}{\Delta_I} \quad \text{bzw.} \quad \frac{p(I)}{\Delta_I}$$

Bei nicht übereinstimmender Klassenbreite ist die Höhe der Rechtecke nicht einfach proportional zur Häufigkeit, sondern zum Verhältnis aus Häufigkeit und Klassenbreite. Die Rechtecke sind demnach umso höher, je größer die Anzahl der Beobachtungen ist, die in einen Bereich vorgegebener Länge fallen. Damit ist die Höhe des Rechteckes davon abhängig, wie dicht die Beobachtungen beieinander liegen.

# Histogramm

Anmerkung: Sollen zwei Häufigkeitsverteilungen an Hand der Histogramme verglichen werden, deren Grundgesamtheiten einen unterschiedlichen Umfang haben, so sind die relativen Häufigkeitsdichten zu verwenden. Auf diese Weise wird erreicht, dass eine Flächeneinheit demselben Anteil an der Grundgesamtheit entspricht. In diesem Fall können die Histogramme wie in Abbildung 5.16 übereinandergelegt werden.

**Häufigkeitsdichte**
Sie entspricht der Höhe des Rechtecks in einem Histogramm:

$$\frac{h(I)}{\Delta_I} \quad \text{bzw.} \quad \frac{p(I)}{\Delta_I}$$

Bei nicht übereinstimmender Klassenbreite ist die Höhe der Rechtecke nicht einfach proportional zur Häufigkeit, sondern zum Verhältnis aus Häufigkeit und Klassenbreite. Die Rechtecke sind demnach umso höher, je größer die Anzahl der Beobachtungen ist, die in einen Bereich vorgegebener Länge fallen. Damit ist die Höhe des Rechteckes davon abhängig, wie dicht die Beobachtungen beieinander liegen.

# Histogramm

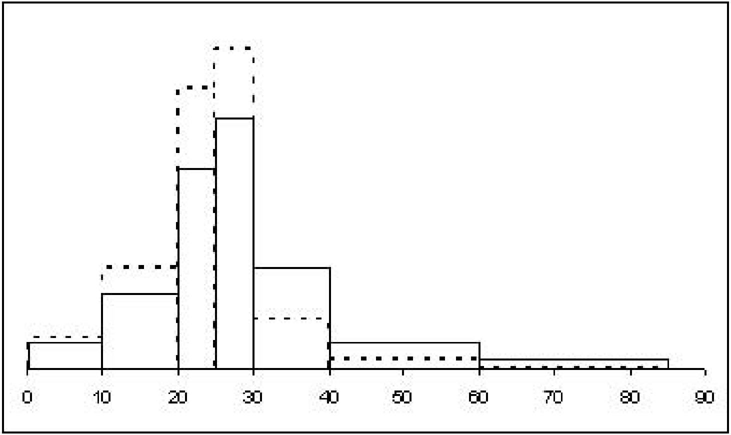


Abbildung 5.16 - Histogramm zur Einkommensverteilung der Frauen bzw. Männer

# Histogramm

Anmerkung: Stimmen die Klassenbreiten aller Klassen überein, so entspricht das Flächendiagramm auch einem Säulendiagramm. In diesem Fall sind die Höhen der Rechtecke aussagekräftig, d.h. die Ordinate kann mit einer Skala für die absoluten und/oder relativen Häufigkeiten versehen werden.

# Histogramm

Beispiel 5.6 - Gegeben sei folgende, gegenüber der ursprünglichen leicht modifizierte Einkommensverteilung der Männer:

| | Männer | |
|---|---|---|
| Einkommen in TEuro | abs. H. | rel. H. |
| unter 10 | 5 | 0.05 |
| 10 bis unter 20 | 15 | 0.15 |
| 20 bis unter 30 | 45 | 0.45 |
| 30 bis unter 40 | 20 | 0.20 |
| 40 bis unter 50 | 5 | 0.05 |
| 50 bis unter 60 | 5 | 0.05 |
| 60 bis unter 70 | 5 | 0.05 |
| $\sum$ | 100 | 1 |

# Histogramm

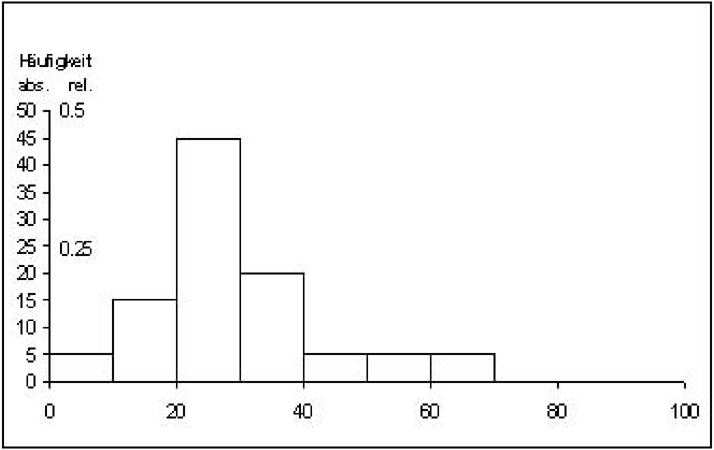


Abbildung 5.17 - Histogramm zur modifizierten Einkommensverteilung der Männer

# Histogramm

**Häufigkeitspolygon**

Polygonzug, der durch die Verbindung der Mittelpunkte der oberen Rechteckkanten entsteht.

# Histogramm

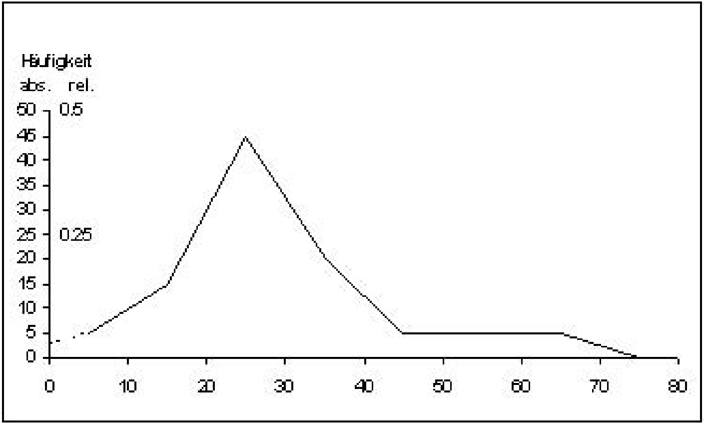


Abbildung 5.18 - Häufigkeitspolygon zur modifizierten Einkommensverteilung der Männer

# Histogramm

Beispiel 5.7 - Fragt man in Beispiel 5.5 etwa nach der Anzahl der männlichen Beschäftigten mit einem Einkommen zwischen 25000 und 45000 Euro, so ist aus der Tabelle zu ersehen, dass die Zahl zwischen 45 und 55 liegen muss und zu vermuten, dass sie zwischen 45 und 50 liegen dürfte. Die Anzahl der Flächeneinheiten beträgt 23.75, man erhält als Näherung 47.5 Personen.

Anmerkung: Bei dem Histogramm der Häufigkeitsverteilung besteht die Möglichkeit, die aufsummierte Häufigkeit mehrerer Klassen dadurch zu bestimmen, dass man die Fläche über den Klassen und unter der oberen Begrenzung durch das Histogramm bestimmt.

# Histogramm

Beispiel 5.7 - Fragt man in Beispiel 5.5 etwa nach der Anzahl der männlichen Beschäftigten mit einem Einkommen zwischen 25000 und 45000 Euro, so ist aus der Tabelle zu ersehen, dass die Zahl zwischen 45 und 55 liegen muss und zu vermuten, dass sie zwischen 45 und 50 liegen dürfte. Die Anzahl der Flächeneinheiten beträgt 23.75, man erhält als Näherung 47.5 Personen.

Anmerkung: Bei dem Histogramm der Häufigkeitsverteilung besteht die Möglichkeit, die aufsummierte Häufigkeit mehrerer Klassen dadurch zu bestimmen, dass man die Fläche über den Klassen und unter der oberen Begrenzung durch das Histogramm bestimmt.

# Histogramm

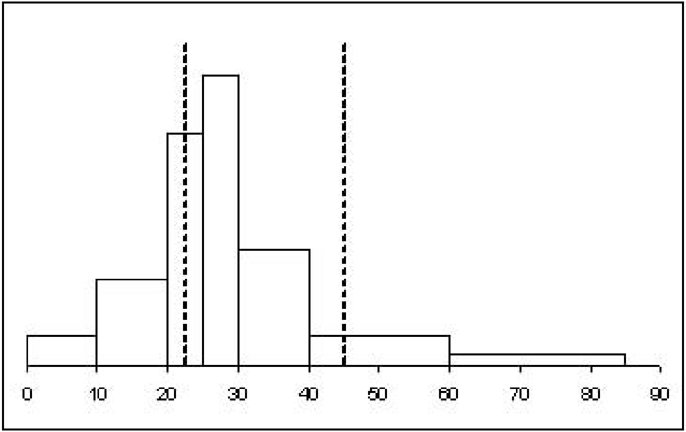


Abbildung 5.19 - Histogramm zur Einkommensverteilung der Männer zu Beispiel 5.7

# Summenhäufigkeitsfunktion

**(Absolute bzw. relative) Summenhäufigkeitsfunktion**:
Bei der graphischen Darstellung eines klassierten stetigen Merkmals werden die einzelnen Punkte bei den Klassengrenzen, da nur für diese eine Summenhäufigkeit definiert ist, näherungsweise durch Geradenstücke verbunden. Dadurch wird unterstellt, dass die absolute und relative Summenhäufigkeit zwischen den Klassengrenzen gleichmäßig zunimmt, d.h. man geht von der idealisierten Vorstellung aus, dass die Merkmalswerte gleichmäßig zwischen den Klassengrenzen verstreut liegen. Die so ermittelte Funktion nennt man die absolute bzw. relative Summenhäufigkeitsfunktion.

Anmerkung: Die relative Summenhäufigkeitsfunktion stimmt nicht mit der empirischen Verteilungsfunktion überein, sondern ist nur eine Näherung dieser.

# Summenhäufigkeitsfunktion

**(Absolute bzw. relative) Summenhäufigkeitsfunktion**:
Bei der graphischen Darstellung eines klassierten stetigen Merkmals werden die einzelnen Punkte bei den Klassengrenzen, da nur für diese eine Summenhäufigkeit definiert ist, näherungsweise durch Geradenstücke verbunden. Dadurch wird unterstellt, dass die absolute und relative Summenhäufigkeit zwischen den Klassengrenzen gleichmäßig zunimmt, d.h. man geht von der idealisierten Vorstellung aus, dass die Merkmalswerte gleichmäßig zwischen den Klassengrenzen verstreut liegen. Die so ermittelte Funktion nennt man die absolute bzw. relative Summenhäufigkeitsfunktion.

Anmerkung: Die relative Summenhäufigkeitsfunktion stimmt nicht mit der empirischen Verteilungsfunktion überein, sondern ist nur eine Näherung dieser.

# Summenhäufigkeitsfunktion

Beispiel 5.8 - Bei der Einkommensverteilung der Männer haben wir folgende Summenhäufigkeitstabelle:

| Klassengrenze | 10 | 20 | 25 | 30 | 40 | 60 | 85 |
|---|---|---|---|---|---|---|---|
| Summenhäufigkeit abs. | 5 | 20 | 40 | 65 | 85 | 95 | 100 |
| Summenhäufigkeit rel. | 0.05 | 0.20 | 0.4 | 0.65 | 0.85 | 0.95 | 1 |

# Summenhäufigkeitsfunktion

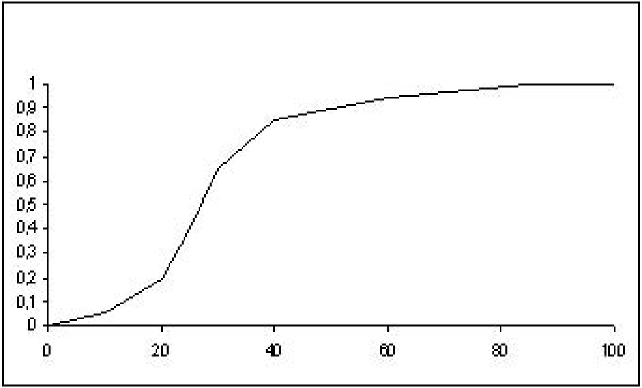


Abbildung 5.20 - Summenhäufigkeitsfunktion zur Einkommensverteilung der Männer

# Summenhäufigkeitsfunktion

**Formale Darstellung:**
Sei $p(I)$ die relative Häufigkeit der Klasse $I$ und $I$ habe die Klassengrenzen $\alpha_I$ und $\beta_I$. Dann nimmt die relative Summenhäufigkeit von $\alpha_I$ nach $\beta_I$ um den Betrag $p(I)$ zu. Sei $z$ ein Wert zwischen $\alpha_I$ und $\beta_I$. Verbindet man nun die Werte an den Stellen $\alpha_I$ und $\beta_I$ durch eine Gerade, so erhält $z$ den Wert $SF(z)$ mit:

$$SF(z) = F(\alpha_I) + \frac{z - \alpha_I}{\beta_I - \alpha_I} \cdot p(I)$$

# Summenhäufigkeitsfunktion

Beispiel 5.9 - Sei $I$ die Klasse der Einkommen von 20000 bis unter 25000 Euro, $z = 22000$ Euro. Für die weiblichen Beschäftigten erhält man mit $p(I) = 0.28$ und $\alpha_I = 20, z = 22$ und $\beta_I = 25$

$$SF(z) = F(\alpha_I) + \frac{z - \alpha_I}{\beta_I - \alpha_I} \cdot p(I) = 0.26 + \frac{22 - 20}{25 - 20} \cdot 0.28 = 0.372$$

also bis $z$ eine Zunahme der relativen Summenhäufigkeit um 0.112.

Anmerkung: Aufgrund der durchgeführten Überlegungen ergibt sich zwischen Histogramm und Summenhäufigkeitsfunktion ein Zusammenhang. Die Fläche des Histogramms links von einem Punkt $z$ ergibt den Wert der Summenhäufigkeitsfunktion an der Stelle $z$, wenn die Gesamtfläche des Histogramms als Flächeneinheit verwendet wird.

# Summenhäufigkeitsfunktion

Beispiel 5.9 - Sei $I$ die Klasse der Einkommen von 20000 bis unter 25000 Euro, $z = 22000$ Euro. Für die weiblichen Beschäftigten erhält man mit $p(I) = 0.28$ und $\alpha_I = 20, z = 22$ und $\beta_I = 25$

$$SF(z) = F(\alpha_I) + \frac{z - \alpha_I}{\beta_I - \alpha_I} \cdot p(I) = 0.26 + \frac{22 - 20}{25 - 20} \cdot 0.28 = 0.372$$

also bis $z$ eine Zunahme der relativen Summenhäufigkeit um 0.112.

Anmerkung: Aufgrund der durchgeführten Überlegungen ergibt sich zwischen Histogramm und Summenhäufigkeitsfunktion ein Zusammenhang. Die Fläche des Histogramms links von einem Punkt $z$ ergibt den Wert der Summenhäufigkeitsfunktion an der Stelle $z$, wenn die Gesamtfläche des Histogramms als Flächeneinheit verwendet wird.

# Summenhäufigkeitsfunktion

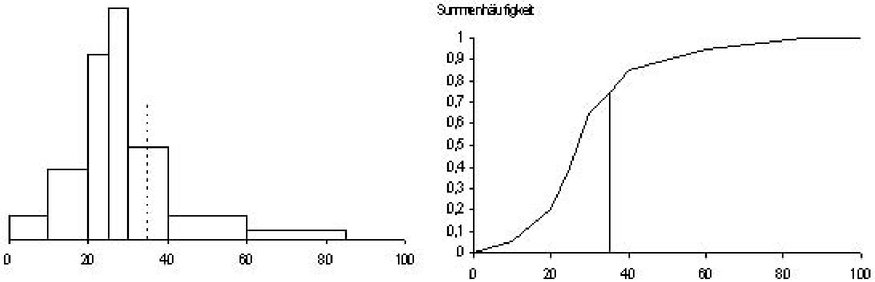


Abbildung 5.21 - Zusammenhang zwischen Histogramm und Summenhäufigkeitsfunktion

# Summenhäufigkeitsfunktion

Anmerkung: Die korrekten Werte erhält man demgegenüber aus der empirischen Verteilungsfunktion.

**Beispiel 4.4** - Die Messung der Körpergröße von 20 Personen ergab folgende Urliste:

$$1.49, 1.87, 1.91, 1.53, 1.68, 1.75, 1.66, 1.82, 1.76, 1.80,$$

$$1.92, 1.71, 1.77, 1.69, 1.57, 1.83, 1.84, 1.47, 1.79, 1.81.$$

Durch eine geeignete Klassierung erhält man folgende Tabelle:

| Klasse | 1<br>1.40 b.u. 1.50 | 2<br>1.50 b.u. 1.60 | 3<br>1.60 b.u. 1.70 | 4<br>1.70 b.u. 1.80 | 5<br>1.80 b.u. 1.90 | 6<br>1.90 b.u. 2.00 | $\sum$ |
|---|---|---|---|---|---|---|---|
| Absolute H. | 2 | 2 | 3 | 5 | 6 | 2 | 20 |
| Relative H. | 0.1 | 0.1 | 0.15 | 0.25 | 0.3 | 0.1 | 1 |

**Beispiel 4.4** - Die Messung der Körpergröße von 20 Personen ergab folgende Urliste:

$$1.49, 1.87, 1.91, 1.53, 1.68, 1.75, 1.66, 1.82, 1.76, 1.80,$$

$$1.92, 1.71, 1.77, 1.69, 1.57, 1.83, 1.84, 1.47, 1.79, 1.81.$$

Durch eine geeignete Klassierung erhält man folgende Tabelle:

| Klasse | 1<br>1.40 b.u. 1.50 | 2<br>1.50 b.u. 1.60 | 3<br>1.60 b.u. 1.70 | |
|---|---|---|---|---|
| Absolute H.<br>Relative H. | 2<br>0.1 | 2<br>0.1 | 3<br>0.15 | |
| Klasse | 4<br>1.70 b.u. 1.80 | 5<br>1.80 b.u. 1.90 | 6<br>1.90 b.u. 2.00 | $\sum$ |
| Absolute H.<br>Relative H. | 5<br>0.25 | 6<br>0.3 | 2<br>0.1 | 20<br>1 |

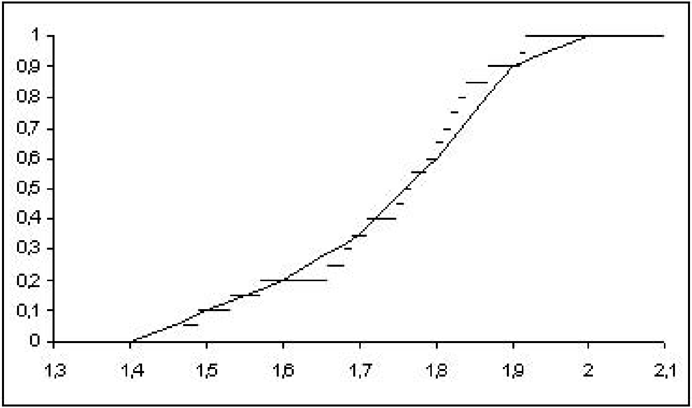


Abbildung 5.23 - Summenhäufigkeitsfunktion als Näherung der empirischen Verteilungsfunktion