Grafische Darstellung von Häufigkeitsverteilungen (1)

- Einführung -

Die grafische Darstellung von Häufigkeitsverteilungen dient...

- der Unterstützung des Lesens und Interpretierens von Daten.
- der Veranschaulichung mathematischer Begriffe und Zusammenhänge.

Häufigkeitsverteilung

Die Menge der Paare von Merkmalsausprägungen und Besetzungshäufigkeiten wird als Häufigkeitsverteilung bezeichnet.

Absolute und relative Häufigkeitsverteilungen werden dargestellt durch:

- Kreisdiagramm (oft benutzt, aber ungünstig)
- Säulen- oder Stabdiagramm
- Histogramm (mit Polygon)

Kumulative Häufigkeitsverteilungen werden dargestellt durch:

- Treppenkurve
- Summenpolygon

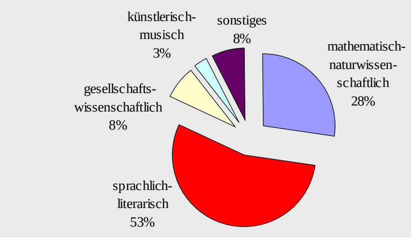
Grafische Darstellung von Häufigkeitsverteilungen (2)

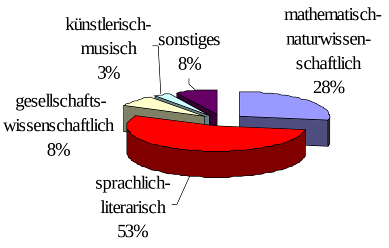
- Kreisdiagramm -

Kreisdiagramm

Das Kreisdiagramm wird häufig als eine geeignete Form der Darstellung für kategoriale Variablen angesehen. Aus wahrnehmungspsychologischen Gründen ist das Kreisdiagramm allerdings ungünstig. Noch problematischer sind dreidimensionale Darstellungen (s.u.). Im Kreisdiagramm wird jeder Ausprägung des Merkmals wird Kreissegment zugewiesen. Die Fläche jedes Segments spiegelt die relative bzw. prozentuale Häufigkeit seiner Ausprägung wider.

Beispieldaten: Erstsemesterbefragung WS99/00 Psychologie:
Merkmal: Leistungskursbereich des 1. LK (n = 105)





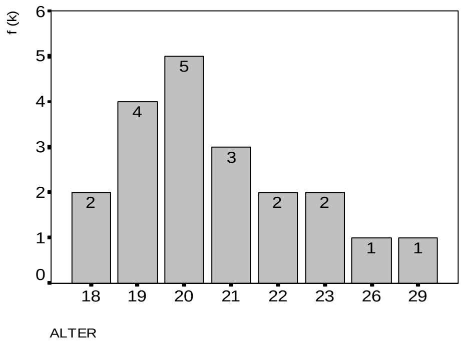
Grafische Darstellung von Häufigkeitsverteilungen (3)

- Säulendiagramm I -

Säulen- oder Stabdiagramm

In einem Koordinatensystem werden die Messwerte (oder Klassen) auf der x-Achse (Abszisse) und die Häufigkeiten der Messwerte (absolut oder relativ) auf der y-Achse (Ordinate) abgetragen. In einem Säulendiagramm lassen sich die Häufigkeiten der Klassen oder Kategorien besser vergleichen als in einem Kreisdiagramm.
Die Höhe des Rechteckes oder Stabes auf der Ordinatenachse entspricht der Häufigkeit des Auftretens der Merkmalsausprägung.

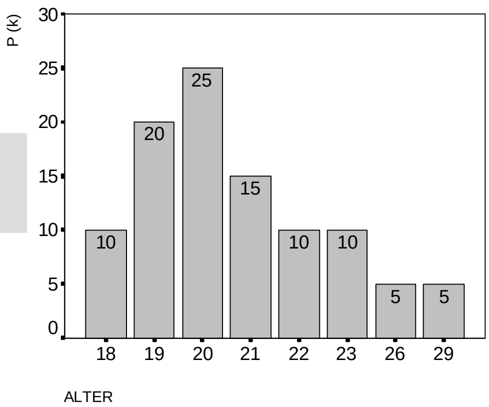
Merkmal: Alter in Jahren
- Absolute Häufigkeiten f_k
- Säulen berühren sich nicht!



Grafische Darstellung von Häufigkeitsverteilungen (4)

- Säulendiagramm II -

Merkmal: Alter in Jahren
- Prozentuale Häufigkeiten P_k
- Säulen berühren sich nicht!



Grafische Darstellung von Häufigkeitsverteilungen (5)

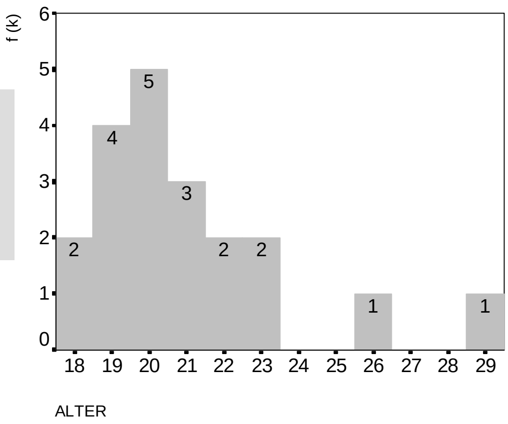
- Histogramm I -

Histogramm

Liegen Daten mit einer Ordnungsstruktur vor, kann das Histogramm angewendet werden.

Merkmal: Alter in Jahren

- Absolute Häufigkeiten f_k
- Säulen grenzen direkt aneinander
- Auch nicht besetzte Klassen werden berücksichtigt



Grafische Darstellung von Häufigkeitsverteilungen (6)

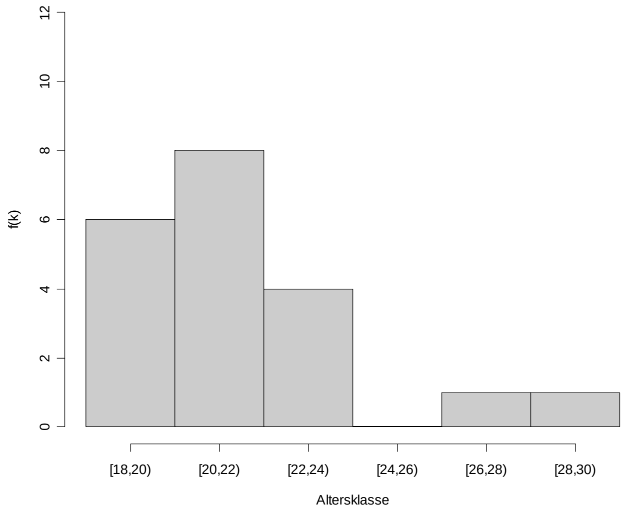
- Histogramm II -

Histogramm

Das Histogramm ist insbesondere für kategorisierte Daten die geeignete Darstellungsform.

Merkmal: Alter in Jahren kategorisiert

- Klassengrenzen eingetragen
- Absolute Häufigkeiten f_k



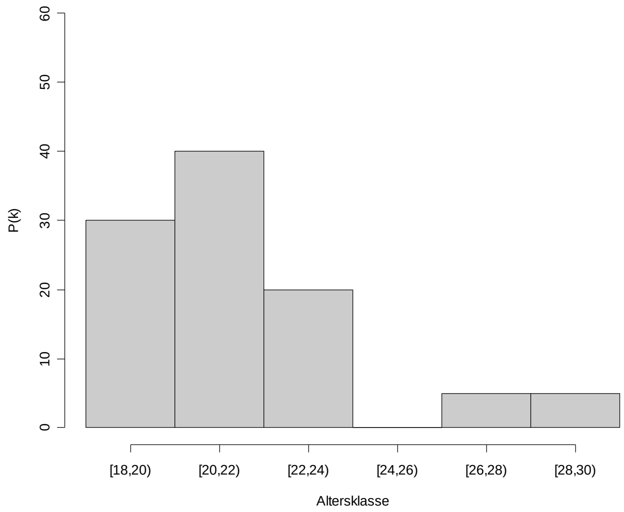
Grafische Darstellung von Häufigkeitsverteilungen (7)

- Histogramm III -

Histogramm

Der Anteil der Fläche unter dem Histogramm bis zu einem bestimmten Skalenpunkt an der Gesamtfläche entspricht der relativen (prozentualen) kumulativen Häufigkeit dieses Skalenpunktes.

Merkmal: Alter in Jahren kategorisiert
- Klassengrenzen eingetragen
- Prozentuale Häufigkeit P_k



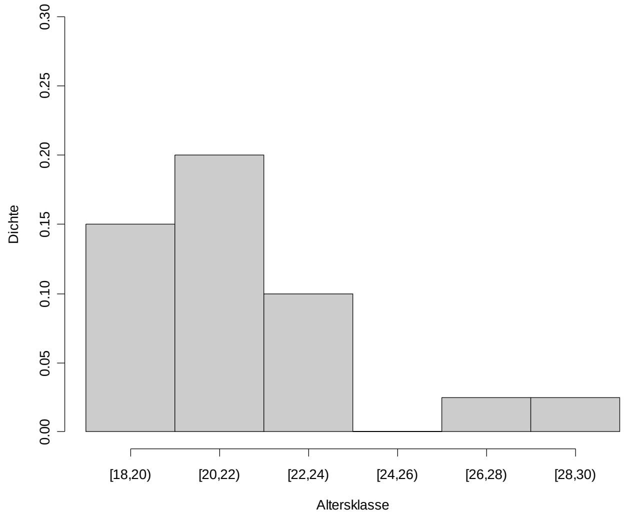
Grafische Darstellung von Häufigkeitsverteilungen (8)

- Histogramm IV -

Histogramm

Wird die relative Häufigkeit durch die Klassenbreite dividiert, erhält man die Dichte; dieses Maß ist auch für ungleich breite Klassen geeignet; die Gesamtfläche ist 1.

Merkmal: Alter in Jahren kategorisiert
- Klassengrenzen eingetragen
- Dichte = p(k)/Kl.Breite



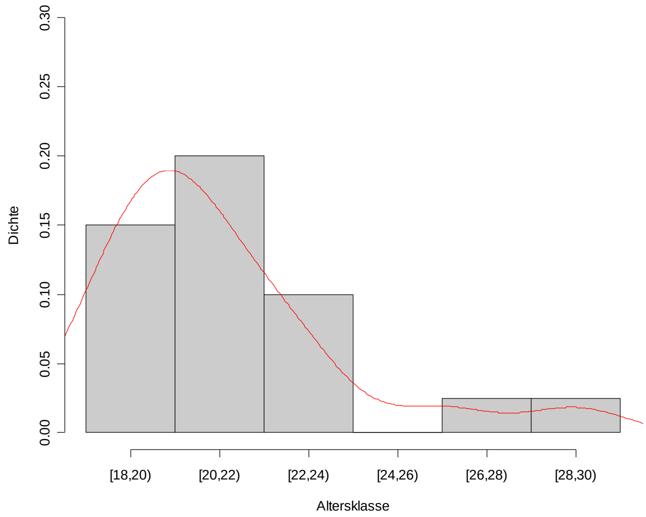
Grafische Darstellung von Häufigkeitsverteilungen (9)

- Histogramm V -

Histogramm und Kerndichtekurve

Eine moderne Alternative zum Histogramm ist die Kerndichtekurve. Zwar ist auch ihre Form von der Wahl einer „Bandbreite“ abhängig, die Verteilungsform ist aber wesentlich besser erkennbar.

Merkmal: Alter in Jahren kategorisiert
- Klassengrenzen eingetragen
- Dichte = p(k)/Kl.Breite



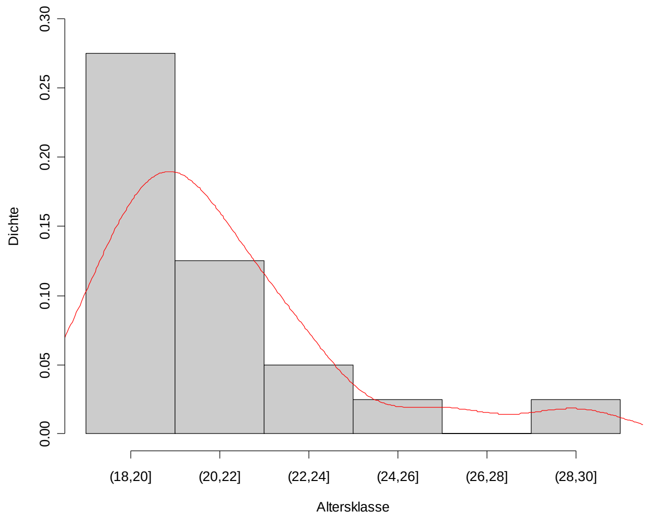
Grafische Darstellung von Häufigkeitsverteilungen (10)

- Histogramm VI -

Histogramm und Kerndichtekurve

Demgegenüber kann die Form des Histogramms sich schon bei einer Änderung der Klassengrenzen (rechter vs. linker Einschluss) drastisch ändern; die Dichtekurve hat dies Problem nicht:

Merkmal: Alter in Jahren kategorisiert
- Klassengrenzen eingetragen
- Dichte = p(k)/Kl.Breite



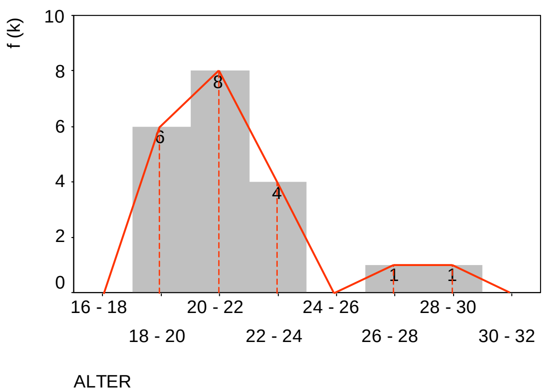
Grafische Darstellung von Häufigkeitsverteilungen (11)

- Histogramm VII -

Histogramm mit Polygon

Ist das Merkmal künstlich diskret können die Intervallmitten mit einem Linienzug verbunden werden. Hierbei wird angenommen, dass sich die Werte über die Intervallklassen gleich verteilen.

Merkmal:
Alter in Jahren kategorisiert
- Histogramm mit Polygon



Grafische Darstellung von Häufigkeitsverteilungen (12)

- Kumulative Häufigkeitsverteilungen I -

Summenhäufigkeit

Die Kumulation der absoluten Häufigkeiten gibt an, wie viele Merkmalsträger die Merkmalsausprägung x_i oder eine geringere besitzen.

Relative Summenhäufigkeit

Die Kumulation der relativen Häufigkeiten gibt den relativen Anteil der Merkmalsträger an, die die Merkmalsausprägung x_i oder eine geringere besitzen.

Empirische Verteilungsfunktion

Eine Funktion F(x) heißt empirische Verteilungsfunktion oder Summenhäufigkeitskurve des Merkmals X, wenn sie jeder reellen Zahl x den Anteil derjenigen Merkmalsträger zuordnet, deren Merkmalsausprägungen diese Zahl nicht überschreiten. Sie ist nur dann sinnvoll anwendbar, wenn eine Ordnungsrelation der Merkmalsausprägungen vorliegt.
F (x) ist monoton wachsend.

Ogive

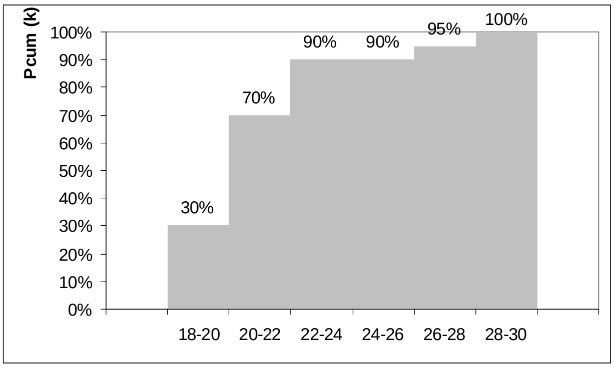
Graphische Darstellungen kumulierter Häufigkeiten nennt man Ogiven (sowohl das Summenpolygon als auch die Treppenkurve).

Grafische Darstellung von Häufigkeitsverteilungen (13)

- Kumulative Häufigkeitsverteilungen II -

Treppenkurve

Ist das Merkmal natürlich diskret (z.B. Kinderanzahl), so kann die kumulierte Häufigkeitsverteilung in Form einer Treppenkurve dargestellt werden. Zwischen den einzelnen Ausprägungen sind stufenartige Sprünge, da kein kontinuierliches Merkmal zugrunde liegt. In der Treppenkurve der kumulativen Verteilung entspricht die Sprunghöhe der Besetzungshäufigkeit.

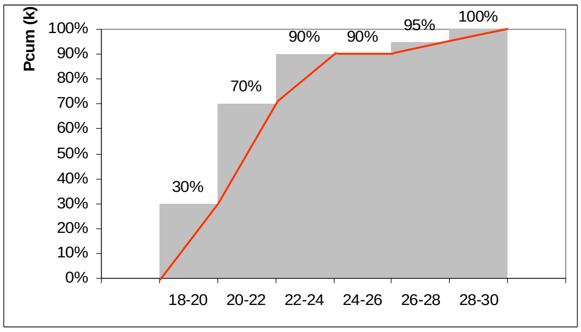


Grafische Darstellung von Häufigkeitsverteilungen (14)

- Kumulative Häufigkeitsverteilungen III -

Summenpolygon

Ist das Merkmal künstlich diskret (z.B. Alter), so kann die kumulierte Häufigkeitsverteilung in Form eines Summenpolygon dargestellt werden. Die kumulierten Häufigkeiten am Intervallanfang und -ende einer Klasse werden mit einer geraden Linie verbunden. Dem liegt die Annahme einer gleichmäßigen Verteilung der Messwerte über das Intervall zugrunde. Der Anstieg des Summenpolygons von Intervallanfang bis -ende einer Klasse ist proportional der Besetzungshäufigkeit dieser Klasse.



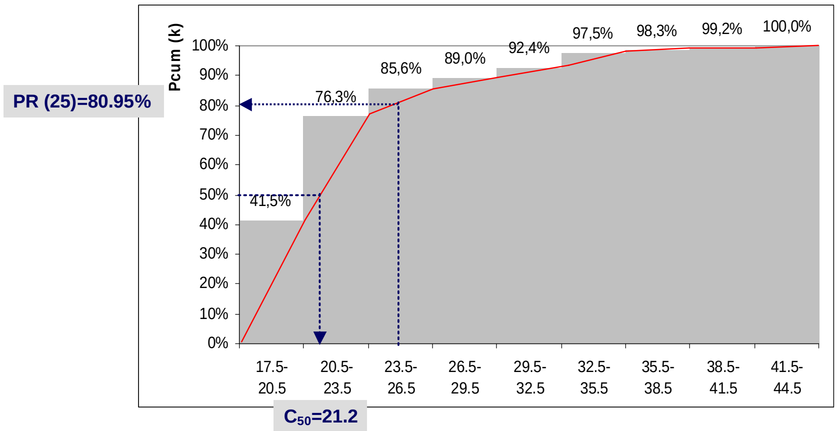
Grafische Darstellung von Häufigkeitsverteilungen (15)

- Kumulative Häufigkeitsverteilungen IV -

Summenpolygon

Grafische Bestimmung des PR (25) und C_{50}.

Beispiel: Altersverteilung der StudienanfängerInnen WS 99/00 (N=118)



Grafische Darstellung von Häufigkeitsverteilungen (16)

- Zusammenfassung -

Grafische Darstellung absoluter und relativer Häufigkeitsverteilungen:

- kategoriale Daten
- natürlich diskrete Daten
- künstlich diskrete Daten

Grafische Darstellung kumulativer Häufigkeitsverteilungen:

- natürlich diskrete Daten
- künstlich diskrete Daten