Eberswalde University
for Sustainable
Development

Informationstechnologien im Wald I

Bachelor SÖW 2. Semester

Dr. Evelyn Wallor

Organisatorisches

Dozentin: Dr. Evelyn Wallor
(Haus 11, Raum 315)
evelyn.wallor@hnee.de
(03334) 657-357

Literatur:

RUDOLF, M.; KUHLISCH, W. (2008). Biostatistik. Eine Einführung für Naturwissenschaftler. Pearson Studium. München.

HEDDERICH, J.; SACHS, L. (2012). Angewandte Statistik. Methodensammlung mit R, 14. Auflage. Springer- Verlag. Berlin.

OESTREICH, M.; ROMBERG, O. (2009). Keine Panik vor Statistik! Erfolg und Spaß im Horrorfach nicht-technischer Studiengänge. Vieweg und Teubner. Wiesbaden.

Organisatorisches

Moodle Lernraum: https://lms.hnee.de/course/view.php?id=4122

Einführung

Was ist Biometrie?

- altgriechisch *bíos* „Leben“ und *métron* „Maß“
- Die Biometrie ist eine Wissenschaft, die sich mit der Informationsgewinnung aus Daten von Lebewesen beschäftigt.
- Besser wären die Begriffe Biostatistik oder biometrische Statistik, da so eine deutliche Abgrenzung zu Verfahren der automatischen Personenerkennung (biometrischer Reisepass) möglich ist.
- Anwendungsgebiete: Biologie, Medizin, Genetik, Versicherungsmathematik, Land- und Forstwirtschaft, …

Motivation

Warum ist es notwendig statistische Sachverhalte zu verstehen?

- Prüfung bestehen
- Bachelorarbeit und Projektberichte
- Jeder ist in seinem Leben von Statistik umgeben
- Berufliche Notwendigkeit
  (Fachzeitschrift, Berichte)
- Ganz einfach: Spaß!

Einführung

Bestandteile statistischer Untersuchungen:

- Versuchsplanung
  - Untersuchungsziel definieren und Methoden festlegen
- Versuchsdurchführung
  - Datenerhebung
- Versuchsauswertung
  - Datenaufbereitung (ungültige Daten aussortieren)
  - Datenanalyse
  - Interpretation der Ergebnisse/ Schlussfolgerung

Teilbereiche

Deskriptive Statistik

Induktive Statistik

Teilbereiche der Statistik

Deskriptive Statistik

→ Deskriptiv = beschreibend, daher auch beschreibende Statistik

- Beschreibung der erhobenen Daten
  - Datenvalidierung
  - Zusammenfassung von Daten
  - Grafische Darstellung
  - Zusammenhang zwischen Merkmalen (Variablen)

Induktive Statistik

→ Induktiv = schließend, daher auch schließende Statistik, Inferenzstatistik

- Wie kann man von den erhobenen Daten auf die Grundgesamtheit schließen?
  - Statistische Modelle und Hypothesen
  - Wahrscheinlichkeitstheorie
  - Statistische Tests

Grundbegriffe

Grundgesamtheit vs. Stichprobe

Def.

Unter der **Grundgesamtheit** oder auch **Population** versteht man die Menge aller potentiellen statistischen Untersuchungseinheiten für eine bestimmte Fragestellung. Diese Einheiten weisen ähnliche sachliche, räumliche und zeitliche Identifikationsmerkmale auf.

- Endlich → alle Bäume im forstbotanischen Garten
- Unendlich → alle Bäume auf der Erde
- Hypothetisch → alle Bäume die jemals existiert haben

Problem: In den meisten Fällen ist es nicht möglich die Grundgesamtheit komplett zu erfassen bzw. zu vermessen.

↓

Datenmangel, Aufwand, Kosten, … → Stichprobe

Grundbegriffe

Grundgesamtheit vs. Stichprobe

Def.

Eine **Stichprobe** ist eine Teilmenge der Grundgesamtheit. Im Sinne des Untersuchungsziels sollte sie **repräsentativ** sein.

Beispiel: Spirelli kochen

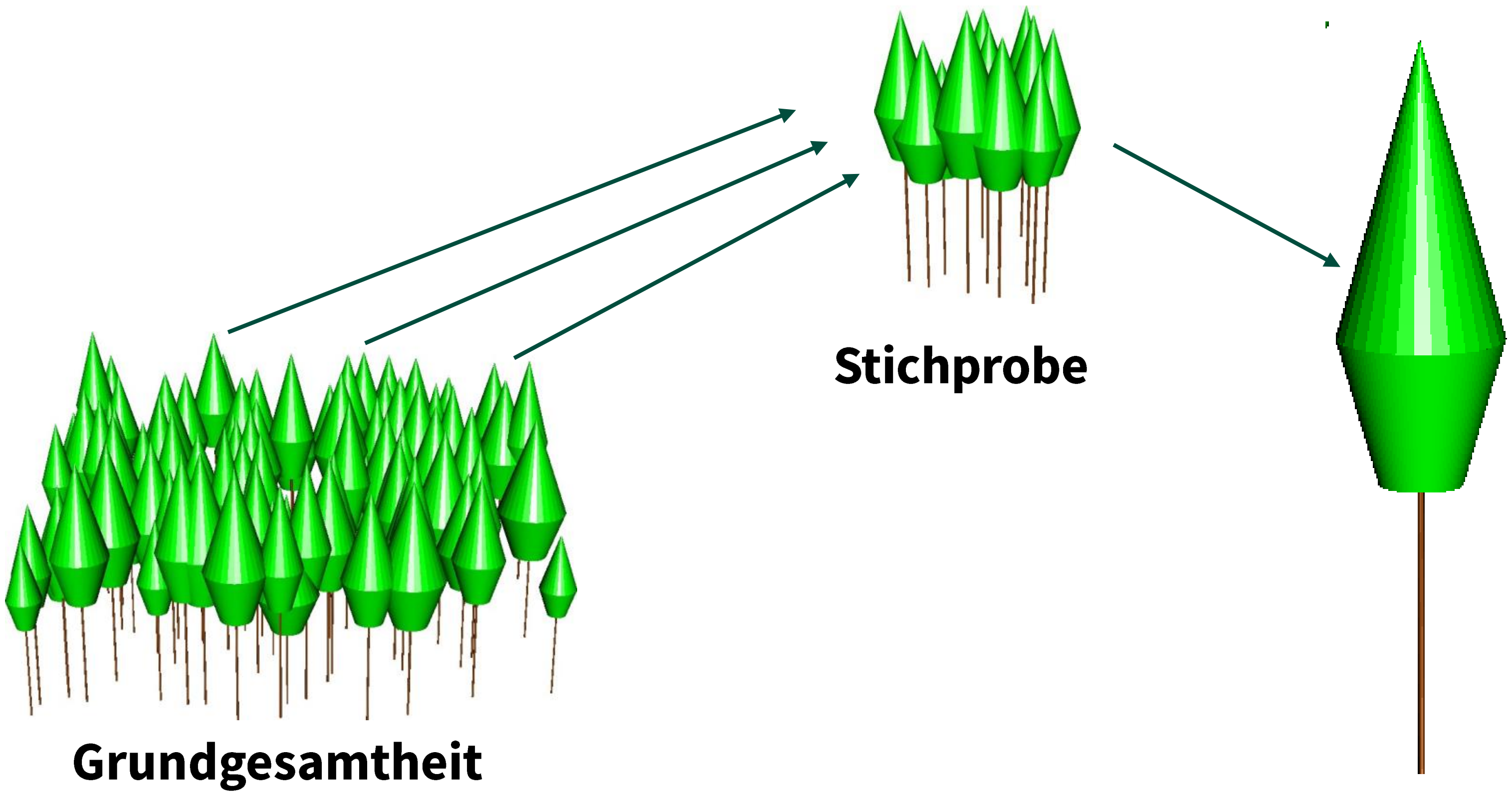
Wie stelle ich fest, ob sie 'al dente' sind?

1. Gut umrühren → **2.** 3 Spirelli testen

Durch das Umrühren wird die kleine Stichprobe repräsentativ.

Grundbegriffe

Beispiel: Wald



Einzelobjekt mit verschiedenen **Merkmalen**:

- Baumart
- Baumhöhe h
- Durchmesser $d_{1,3}$
- Kronenansatz
- Kronenbreite
- Kronenschluss
- Vitalität
- soziale Stellung

Grundbegriffe

Merkmal (Variable)

Def.

Ein **statistisches Merkmal**, auch Variable genannt, ist eine bestimmte Eigenschaft einer statistischen Einheit (Einzelobjekt).

Bezeichnung:

Merkmal	Merkmalsausprägung
X	$x_1, x_2, \ldots, x_n$
(Großbuchstaben)	(Kleinbuchstaben + tiefgestellter Index)

Die Zahl n steht in diesem Zusammenhang für den Stichprobenumfang (Anzahl der Elemente einer Stichprobe).

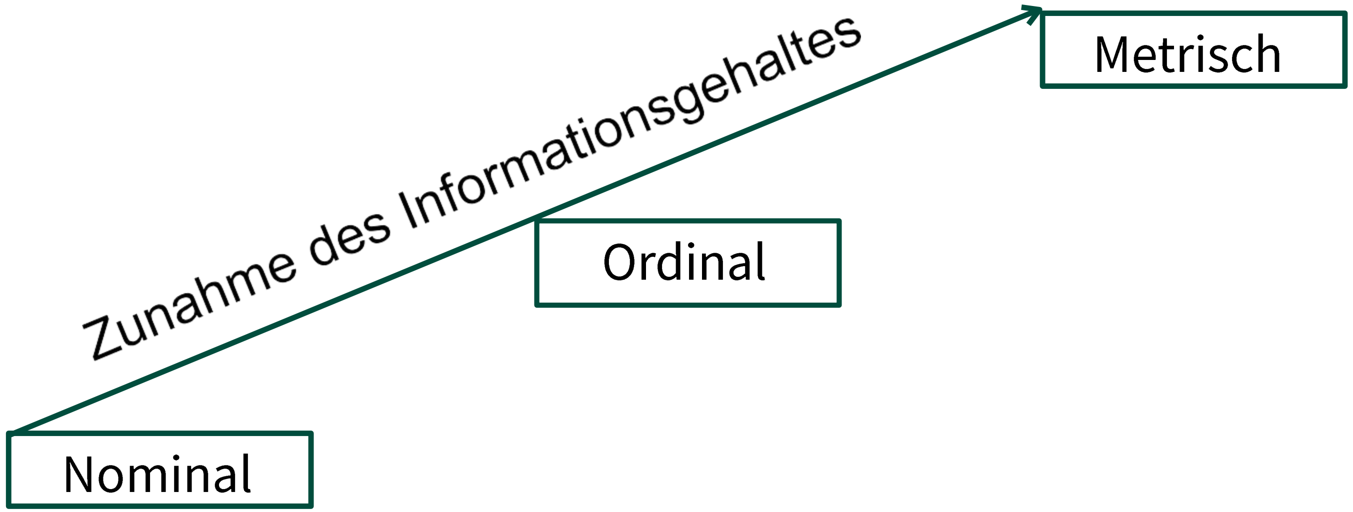
Def.

Der konkrete Wert, den eine Variable bei einer einzelnen statistischen Einheit annimmt, wird **Realisierung** genannt.

Skalenniveau

Skalierung oder Skalenniveau

Die Bestimmung des Skalenniveaus dient der ersten groben Einordnung der Daten und soll die Auswahl geeigneter statistischer Verfahren erleichtern.



Skalenniveau

Nominalskala

Def.

Ein statistisches Merkmal wird als **nominalskaliert** bezeichnet, wenn die Merkmalsausprägung lediglich eine Verschiedenartigkeit zum Ausdruck bringt.

Zulässige Relationen: gleich (==) / ungleich (!=) Beispiele:

- Baumart
  „Gemeine Kiefer“ = „Pinus sylvestris L.“ = „GKI“ = 111
  „Rot-Buche“ = „Fagus sylvatica L.“ = „RBU“ = 511
- Geschlecht
  „männlich“ = 1, „weibliche = 2
- Postleitzahl
  16225, 16227

Die Codierung durch Zahlen hat hier lediglich eine Bezeichnungsfunktion.

Skalenniveau

Ordinal- oder Rangskala

Def.

Ein statistisches Merkmal wird als **ordinalskaliert** bezeichnet, wenn die Merkmalsausprägung eine Verschiedenartigkeit und eine Rangfolge zum Ausdruck bringt. Die Abstände zwischen den Ausprägungen sind hingegen nicht interpretierbar.

Zulässige Relationen: gleich (==) / ungleich (!=) und Ordnung (<,>)

Beispiele:

- Baumklasse von Kraft (1884)
  „vorherrschend“ = 1, „herrschend“ = 2, „gering mitherrschend“ = 3, „beherrscht“ = 4, „ganz unterständig“ = 5
- Schadensstufe
  „ohne“, „gering“, „mittel“, „stark“

Skalenniveau

Metrische Skala

Def.

Eine metrische Skala liegt vor, wenn Ausprägungen durch Zahlen ausgedrückt werden, die eine Verschiedenartigkeit, eine Rangfolge und einen meß- bzw. quantifizierbaren Unterschied definieren.

Intervallskala		Verhältnisskala
	Zulässige Relationen	
gleich (==) / ungleich (!=) Ordnung (<,>)		gleich (==) / ungleich (!=) Ordnung (<,>)
	Rechenoperationen	
Addition (+), Subtraktion (-)		Addition (+), Subtraktion (-), Multiplikation (*), Division (/)
	Beispiele	
Temperatur [°C], Jahreszahlen, Intelligenzquotient [IQ]		Baumhöhe [m], Durchmesser [cm], Temperatur [K]

Skalenniveau

Übersicht Skalenniveau

Skalenniveau	==, !=	<,>	-,+	*,/
Nominal	✓	○	○	○
Ordinal	✓	✓	○	○
Intervall	✓	✓	✓	○
Verhältnis	✓	✓	✓	✓

Qualitative Merkmale

(mit Worten benannt)

Quantitative Merkmale

(durch Zahlen erfasst)

Diskret oder stetig?

Eine weitere Unterteilung der quantitativen Merkmale kann auf der Basis der möglichen Anzahl der Ausprägungen getroffen werden.

Def.

Ein **diskretes** Merkmal besitzt endlich oder abzählbar viele Ausprägungen.

Beispiele:

- Stammzahl N/ha
- Anzahl Baumarten

Def.

Ein Merkmal wird als **stetig** bezeichnet, wenn es theoretisch[1] überabzählbar (unendlich) viele Ausprägungen annehmen kann.

Beispiele:

- Durchmesser $d_{1,3}$
- Baumhöhe h
- Trockenmasse

[1]Durch Verbesserung der Messmethoden könnte man theoretisch eine sehr feine Auflösung und somit unendlich viele Ausprägungen erreichen.

Deskriptive Statistik

Lfd. Nr.	$d_{1,3}$ [cm]	Güte	Baumart
1	22,5	B	GKI
2	31,6	A	RBU
3	21,9	B	GKI
4	18,3	C	GKI
5	18,8	C	GKI
6	15,7	D	RBU
7	13,8	D	RBU
8	22,6	B	GKI
9	18,2	C	GFI
10	17,1	C	GFI
11	20,8	B	RBU
12	19,6	B	GKI

Urliste mit n=12

Lfd. Nr. → laufende Nummer
$d_{1,3}$ → Durchmesser in 1,3 m Höhe
Güte → Schaftqualität bis 6 m
GKI → Gemeine Kiefer
RBU → Rotbuche GFI →
Gemeine Fichte

Deskriptive Statistik

Häufigkeitsverteilung des **nominal** skalierten Merkmals

Merkmals-ausprägung	Absolute Häufigkeit	Relative Häufigkeit	Prozentuale Häufigkeit
a_1	n_1	h_1	p_1
⋮	⋮	⋮	⋮
a_j	n_j	h_j	p_j
Summe	n	1,0	100 %

j = 1,…,k , wobei k die Anzahl **verschiedener** Merkmalsausprägungen ist.

$$h_j = \frac{n_j}{n} = \frac{absolute\ Häufigkeit\ n_j}{Stichprobenumfang\ n}, wobei\ gilt{:} \sum_{j=1}^{k} h_j = 1\ und\ 0 < h_j < 1$$

$$p_j = h_j \cdot 100\% = relative\ Häufigkeit \cdot 100\%$$

Deskriptive Statistik

Häufigkeitsverteilung am Beispiel der Baumart

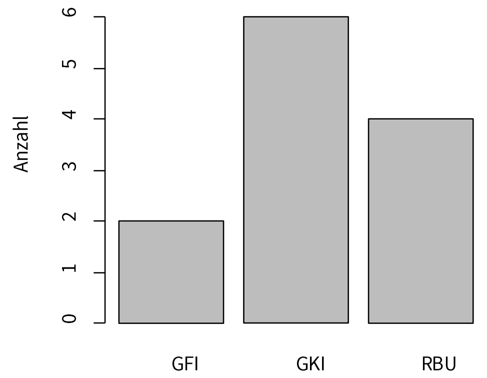
Merkmals-ausprägung	Absolute Häufigkeit	Relative Häufigkeit	Prozentuale Häufigkeit
GKI	6	6/12=0,500	50,0 %
GFI	2	2/12=0,167	16,7 %
RBU	4	4/12=0,333	33,3 %
Summe	12	1.0	100,00 %

Hinweis: Analog zu dieser Häufigkeitsverteilung können wir auch bei ordinalskalierten und metrisch-diskreten Merkmalen vorgehen.

Deskriptive Statistik

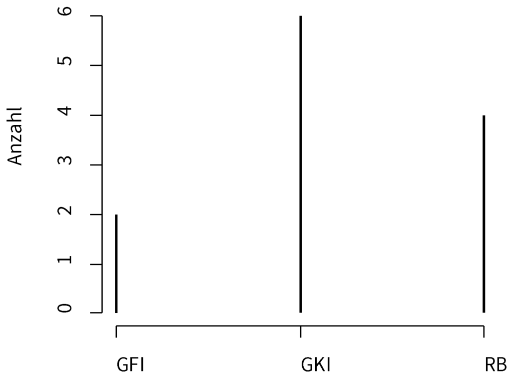
Grafische Darstellungsmöglichkeiten **nominaler** Variablen

Säulen- bzw. Balkendiagramm

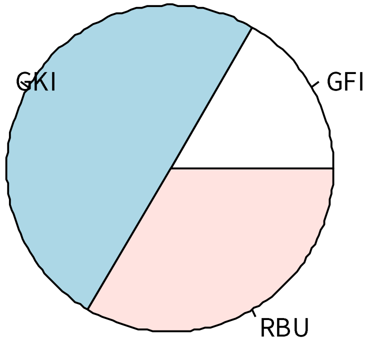


höhen-proportional

Stabdiagramm

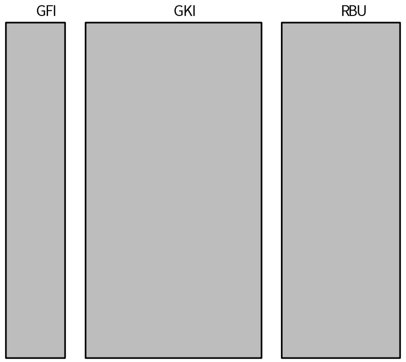


Kreisdiagramm



flächen-proportional

Rechteckdiagramm



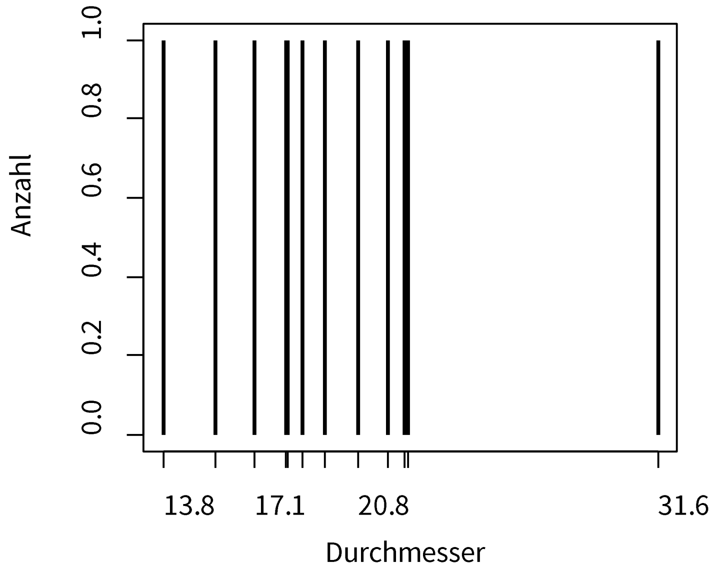
Deskriptive Statistik

Häufigkeitsverteilung **metrisch-stetiger** Variablen

Würde man bei metrisch-stetigen Variablen auf die selbe Art und Weise vorgehen wie bei nominalen Variablen, dann hätte eine Häufigkeitstabelle im schlimmsten Fall n Zeilen.

Ein Stabdiagramm würde in diesem Fall n Stäbe der Höhe 1 aufweisen.

Lösungsansatz: Klassenbildung bzw. Klassierung der Daten



Deskriptive Statistik

Häufigkeitsverteilung **metrisch-stetiger** Variablen

Grundsätze Klasseneinteilung:

- jeder Wert muss einer Klasse eindeutig zugeteilt werden können
- offene Klassen vermeiden
- sinnvolle Klassenbreiten nutzen und Extremfälle (für jeden Wert eine Klasse oder für alle Werte eine Klasse) vermeiden

Formel nach STURGES (1926): $k \approx 1 + 3{,}32 \cdot log(n)$

Beispiel Durchmesser $d_{1,3}$: $k \approx 1 + 3{,}32 \cdot log(12) \quad = 4.58 \approx 5$

Entsprechend dieser Berechnung müssten fünf Klassen gebildet werden.

Deskriptive Statistik

Häufigkeitsverteilung **metrisch-stetiger** Variablen

Doch wo fängt die erste Klasse an und wo hört die letzte Klasse auf?

Def.

Berechnung der **Spannweite R**:
Die Spannweite, auch Variationsbreite oder engl. range genannt, wird als Differenz zwischen dem Maximum und Minimum der Daten einer Stichprobe berechnet.

$$R = x_{max} - x_{min}$$

Beispiel Durchmesser $d_{1,3}$: $R = 31{,}6\ cm - 13{,}8\ cm = 17{,}8\ cm$

Deskriptive Statistik

Häufigkeitsverteilung **metrisch-stetiger** Variablen

Doch wo fängt die erste Klasse an und wo hört die letzte Klasse auf?

$$k \approx 1 + 3{,}32 \cdot log_{10}(12) = 4.58 \approx 5$$

$$R = 31{,}6\ cm - 13{,}8\ cm = 17{,}8\ cm$$

$$Klassenbreite = \frac{17{,}8\ cm}{5} \approx 3{,}6\ cm$$

Diese Klasseneinteilung wäre prinzipiell möglich, sie würde in der Praxis jedoch keine Anwendung finden.

1. Versuch:	**13,8**	**17,4**	**21**	**24,6**	**28,2**	**31,8**
Alternative:	**10**	**15**	**20**	**25**	**30**	**35**

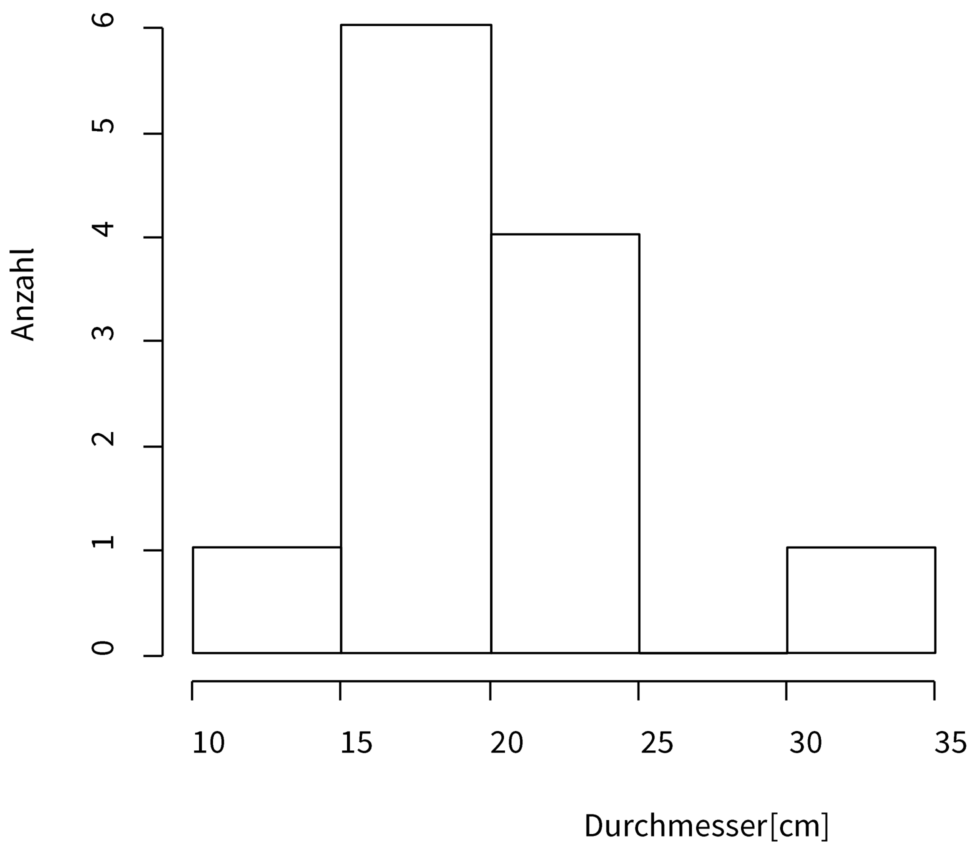
Deskriptive Statistik

Häufigkeitsverteilung am Beispiel des Durchmessers $d_{1,3}$

Klasse	Durchmesser Intervall	Absolute Häufigkeit	Relative Häufigkeit	Prozentuale Häufigkeit
1	$10 \leq x < 15$	1	1/12=0,083	$8,\overline{3}$ %
2	$15 \leq x < 20$	6	6/12=0,500	50,0 %
3	$20 \leq x < 25$	4	4/12=0,333	$33,\overline{3}$ %
4	$25 \leq x < 30$	0	0	0,0 %
5	$30 \leq x < 35$	1	1/12=0,083	$8,\overline{3}$ %
	Summe	12	1.0	100,00 %

Deskriptive Statistik

Häufigkeitsverteilung am Beispiel des Durchmessers $d_{1,3}$



Def.

Ein **Histogramm** ist eine flächenproportionale Darstellung für klassierte Daten. Während auf der x-Achse die Klassengrenzen aufgetragen sind, kann auf der y- Achse die absolute oder relative Häufigkeit verwendet werden.

• • •

Vielen Dank für die Aufmerksamkeit.

Larix decidua, Foto: Jens Jäpel
www.wikipedia.org (CC BY-SA 2.5)
no changes made

Fagus sylvatica, Foto: Jean-Pol Grandmont
www.wikipedia.org (CC BY 3.0)
no changes made