Stochastik für die Naturwissenschaften

Dr. C.J. Luchsinger

2. EINFÜHRUNG IN R UND BESCHREIBENDE STATISTIK (DESCRIPTIVE STATISTICS)

2.1. Einführung in R

kommt nach Ende Semester auch hier in Kurzform

2.2. Beschreibende Statistik

2.2.1 Merkmale und Skalen (Scales)

In diesem ersten Teil werden wir eine wichtige Klassifikation von Datentypen vornehmen. Es ist davon auszugehen, dass die meisten von Ihnen im Umgang mit Daten automatisch und gefühlsmässig richtig handeln würden. Ich nehme in dieser Vorlesung aus Zeitgründen und weil es intellektuell nicht sehr anspruchsvoll ist dieses Kapitel nur kurz zusammenfassend durch. Es ist aber wichtig: lesen Sie es also bitte im Verlauf der ersten Semesterwoche durch und lösen Sie aufmerksam die entsprechenden Aufgaben in den Übungen, weil Sie ab jetzt diese Begriffe alle kennen müssen!

In Umfragen/Untersuchungen kann man von verschiedenen Personen zum Beispiel unter anderem auch das **Geschlecht** registrieren. Es wird bei Menschen entweder männlich oder weiblich sein. In EDV-Systemen wird diese Information oft nicht ausgeschrieben als "männlich" oder "weiblich" abgelegt, sondern man codiert diese Information mit 0 und 1 (bzw. 1 und 0) und gibt in der Beschreibung der Statistik an, ob 0 bzw. 1 männlich oder weiblich bedeutet.

Sie haben dann eine Kolonne mit lauter Zahlen wie

$$\ldots, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1 \ldots \tag{1}$$

Eine weitere Zahlenfolge in der gleichen Erhebung kann zum Beispiel **Schweregrad einer Krankheit** sein. Eine Krankheit ist bei den Personen 0 = nicht vorhanden, 1 = in der leichtesten Form vorhanden, 2 = in einer mittelschweren Form vorhanden oder gar 3 = in schwerer Form vorhanden. Die entsprechenden Zahlenfolgen sehen dann also zum Beispiel auszugsweise folgendermassen aus:

$$\ldots, 0, 0, 3, 0, 2, 1, 3, 0, 0, 1 \ldots \tag{2}$$

Eine weitere Zahlenfolge in der gleichen Erhebung kann zum Beispiel die **Körpertemperatur der Person** sein. Diese wird hier in Celsius angegeben und sieht folgendermassen aus:

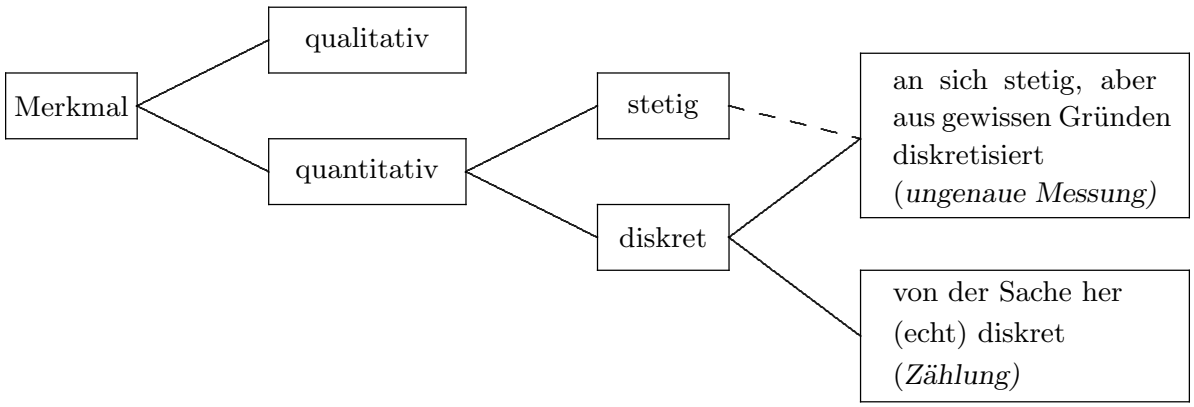
$$\ldots, 36, 36, 37, 39, 41, 36, 37 \ldots \tag{3}$$

Eine weitere Zahlenfolge in der gleichen Erhebung kann zum Beispiel das **Gewicht der Person** sein. Dieses wird hier in kg angegeben und sieht folgendermassen aus:

$$\ldots, 80, 66, 73, 104, 82, 51, 73 \ldots \tag{4}$$

Vergleichen Sie die 4 Datentypen miteinander. Überlegen Sie sich, ob zum Beispiel die Differenzen von 2 Zahlen eine konkrete Bedeutung haben oder gar das Verhältnis; (wo) ist es sinnvoll, das arithmetische Mittel zu bestimmen?

Eine erste Ordnung von Merkmalen sieht folgendermassen aus:



Des weiteren kann man sich überlegen, welche (Rechen-)Operationen wo sinnvoll sind. Dazu unterscheiden wir 4 Arten von Merkmalen:

Nominalskala	Keine Rangordnung; Rechenoperationen sinnlos (qualitatives Merkmal)
Ordinalskala	**Rangordnung festgelegt**; Rechenoperationen sinnlos (qualitatives Merkmal)
Intervallskala	Rangordnung festgelegt; **Abstände zwischen Zahlen haben Bedeutung**, Nullpunkt willkürlich, Bildung von Verhältnissen (und Prozenten) nicht sinnvoll (quantitatives Merkmal)
Verhältnisskala	Rangordnung festgelegt; Abstände zwischen Zahlen haben Bedeutung, **Nullpunkt absolut**, Bildung von Verhältnissen (und Prozenten) sinnvoll (quantitatives Merkmal)

Es folgen jetzt in der Vlsg ein paar Bemerkungen zum Begriff "abzählbar unendlich" (countably infinite).

https://schweizermonat.ch/sag-mir-wo-die-zahlen-sind-wo-sind-sie-geblieben/
https://schweizermonat.ch/stichprobenauswahl/

2.2.2 Darstellung von Versuchsergebnissen

DISKRET (discrete): Wir betrachten Prüfungsnoten (welche ev. sehr umfangreich sein können - vgl diese Vlsg mit über 700 Studis).

4	$4\frac{1}{2}$	5	3	$5\frac{1}{2}$	$4\frac{1}{2}$	$3\frac{1}{2}$	$5\frac{1}{2}$	3	4
$4\frac{1}{2}$	1	5	4	$5\frac{1}{2}$	$3\frac{1}{2}$	6	$4\frac{1}{2}$	5	$4\frac{1}{2}$
6	4	3	$4\frac{1}{2}$	5	3	6	4	$2\frac{1}{2}$	$4\frac{1}{2}$
$4\frac{1}{2}$	$3\frac{1}{2}$	$3\frac{1}{2}$	$5\frac{1}{2}$	4	$4\frac{1}{2}$	3	5	3	4

Anstelle obiger, ungeordneter *Urliste* oder *Rohdaten* (immer aufbewahren!) wollen wir die Daten übersichtlicher darstellen. Eine erste Zwischenstufe kann eine *Strichliste* sein. Anstelle von (n=) 40 (bzw über 700 Zahlen) haben wir jetzt lediglich eine Folge von 11 Zahlen.

1	|	1
$1\frac{1}{2}$		0
2		0
$2\frac{1}{2}$	|	1
3	~~||||~~ |	6
$3\frac{1}{2}$	||||	4
4	~~||||~~ ||	7
$4\frac{1}{2}$	~~||||~~ ||||	9
5	~~||||~~	5
$5\frac{1}{2}$	||||	4
6	|||	3

Es sind dies die *absoluten Häufigkeiten* (absolute frequencies) des betrachteten Merkmalswerts. Abstrakt können wir dies formalisieren in der folgenden Art: die möglichen Werte seien

$$\omega_1, \omega_2, \ldots, \omega_k.$$

In unserem Beispiel ist $\omega_1 = 1, \omega_2 = 1.5, \ldots, \omega_{11} = 6$ und somit $k = 11$. Die absolute Häufigkeit des Merkmalswerts ω_i bezeichnen wir mit H_i (hier also $H_1 = 1, \ldots, H_{11} = 3$). Es gilt dann immer

$$\sum_{i=1}^{k} H_i = n.$$

Aussagekräftiger als die *absoluten* Häufigkeiten H_i sind die *relativen* Häufigkeiten (relative frequencies) h_i; dazu teilen wir die H_i jeweils durch die Anzahl der Untersuchungsobjekte n:

$$h_i := \frac{H_i}{n} \tag{Theorie}$$

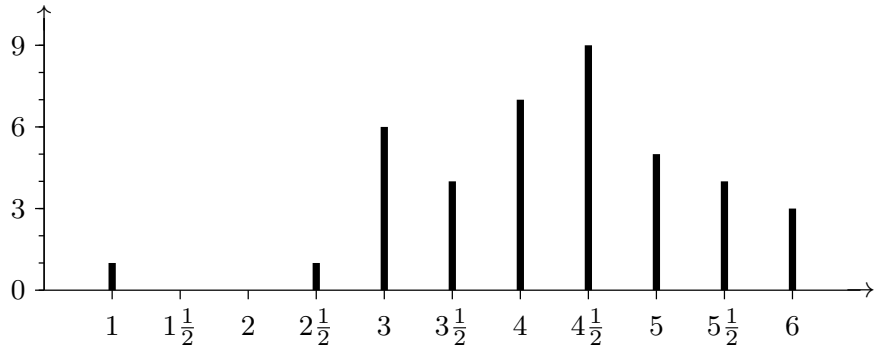
bzw

$$h_i := \frac{H_i}{n} * 100\%. \tag{Praxis}$$

Die Prüfungsnoten lassen sich damit kompakt zusammenfassen:

w_i	H_i	h_i	h_i (%)
1	1	0.025	2.5%
$1\frac{1}{2}$	0	0	0%
2	0	0	0%
$2\frac{1}{2}$	1	0.025	2.5%
3	6	0.150	15%
$3\frac{1}{2}$	4	0.100	10%
4	7	0.175	17.5%
$4\frac{1}{2}$	9	0.225	22.5%
5	5	0.125	12.5%
$5\frac{1}{2}$	4	0.100	10%
6	3	0.075	7.5%

Bilder sagen mehr als 1000 Worte und Zahlen; das *Stabdiagramm* (bar chart) gibt auf den ersten Blick wohl den schnellsten und besten Überblick:

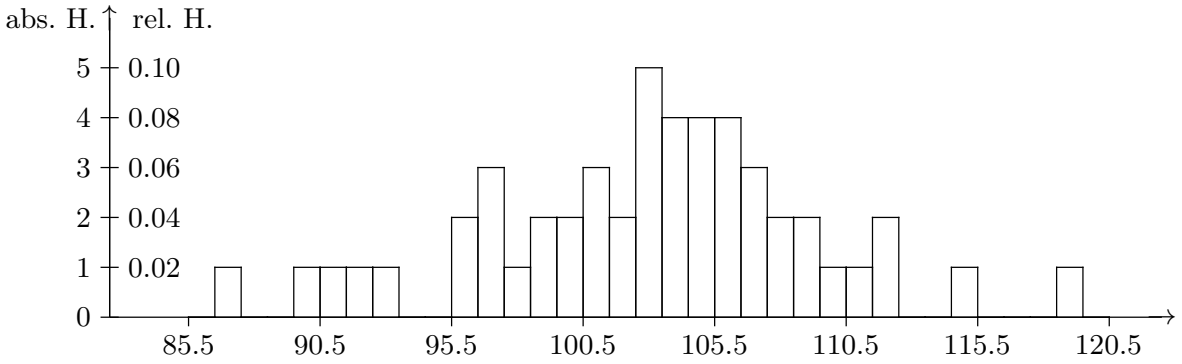


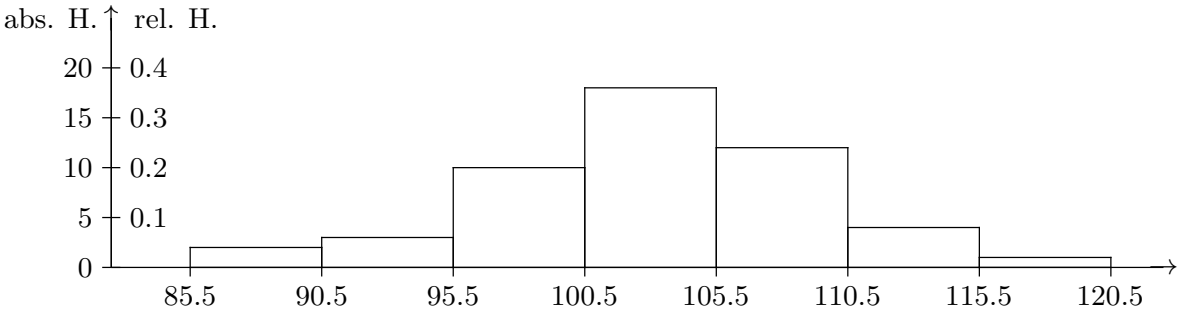
STETIG (continuous): In 2.1 haben wir kurz diskrete und stetige Daten unterschieden. Die Prüfungsnoten sind klar diskret. Wenn man 50 Küken wiegt, so ist das Gewicht (abgesehen von physikalisch / philosophischen Fragen) sicher stetig. Wir werden aber meist (unbewusst) Klassen bilden: wir fassen alle Küken mit Gewicht zwischen 103.5 und 104.5 zu 104 g zusammen. Wieder bilden wir die Urliste

100	87	101	107	102	105	91	104	103	102
99	96	104	93	105	107	103	106	96	111
104	92	107	101	109	90	106	97	103	112
101	103	108	105	105	110	97	109	112	103
108	98	104	106	97	119	99	115	100	106

und fassen die Strichliste und (absolute/relative) Häufigkeiten in einer Tabelle zusammen (wegen Seitenumbruch auf übernächster Seite).

Anstelle des Stabdiagramms (diskret - Stäbe berühren sich nicht) haben wir jetzt ein sogenanntes *Histogramm oder Blockdiagramm.* Die Balken liegen hier aneinander - dies soll illustrieren, dass eigentlich jeder Wert vorkommen kann. Selbsterklärende Begriffe sind dann auch *Klassenbreite (bin width) und Klassenmitte.*

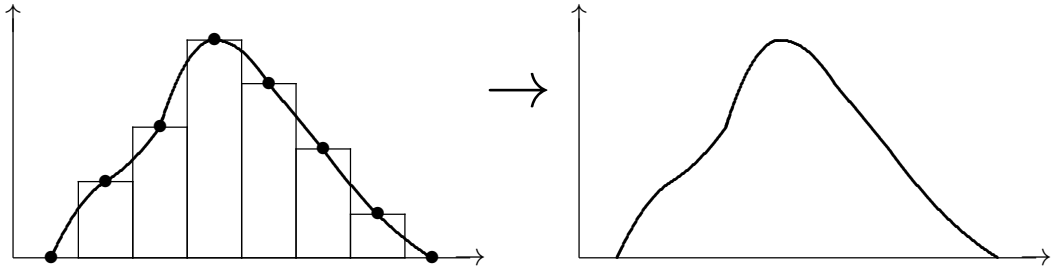




Man sollte nach Möglichkeit darauf achten, dass der Flächeninhalt in einem Histogramm total 1 ist (vor allem, wenn man thematisch auf eine Dichtefunktion (vgl Kapitel 4) hinarbeiten will). Beim unteren Histogramm stimmt das zwar nicht - dafür sind sie vergleichbar. Zur Frage der *Klassenbreite*: Frage ans Publikum: Wie gross sollte man diese wählen?

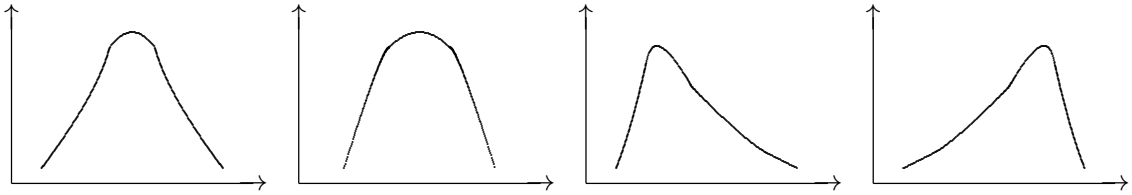
Messwert		absolute Häufigkeit	relative Häufigkeit (in %)
86		0	0
87	I	1	2
88		0	0
89		0	0
90	I	1	2
91	I	1	2
92	I	1	2
93	I	1	2
94		0	0
95		0	0
96	II	2	4
97	III	3	6
98	I	1	2
99	II	2	4
100	II	2	4
101	III	3	6
102	II	2	4
103	~~IIII~~	5	10
104	IIII	4	8
105	IIII	4	8
106	IIII	4	8
107	III	3	6
108	II	2	4
109	II	2	4
110	I	1	2
111	I	1	2
112	II	2	4
113		0	0
114		0	0
115	I	1	2
116		0	0
117		0	0
118		0	0
119	I	1	2
120		0	0
Total		50	100

Zur Idealisierung:



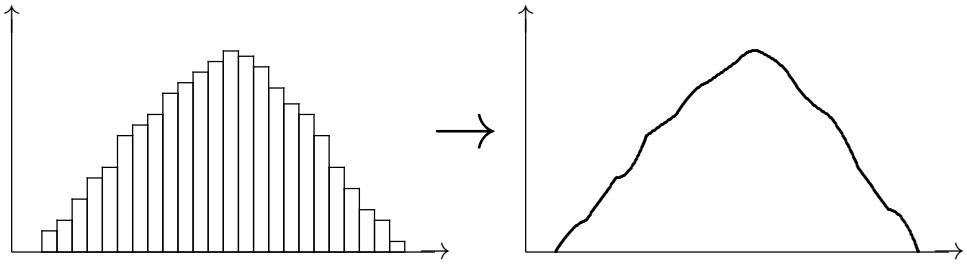
Dies tut man zum Beispiel dann, wenn man grob die Form der Häufigkeitsverteilung skizzieren will.

Man unterscheidet etwa (Merke: Fussball rollt nach...):



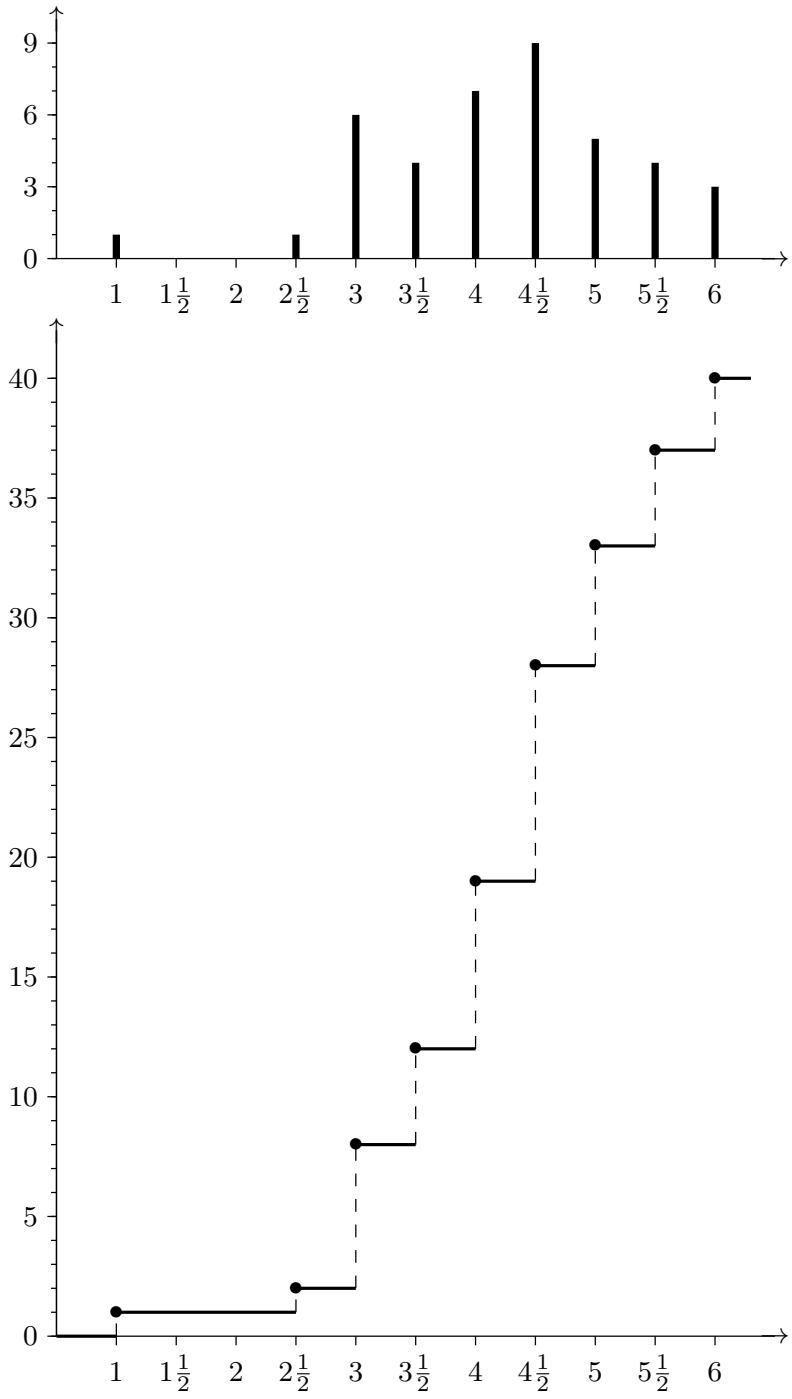
glockenförmig symmetrisch rechtsschief linksschief

Wir haben mit "$n \to \infty$":



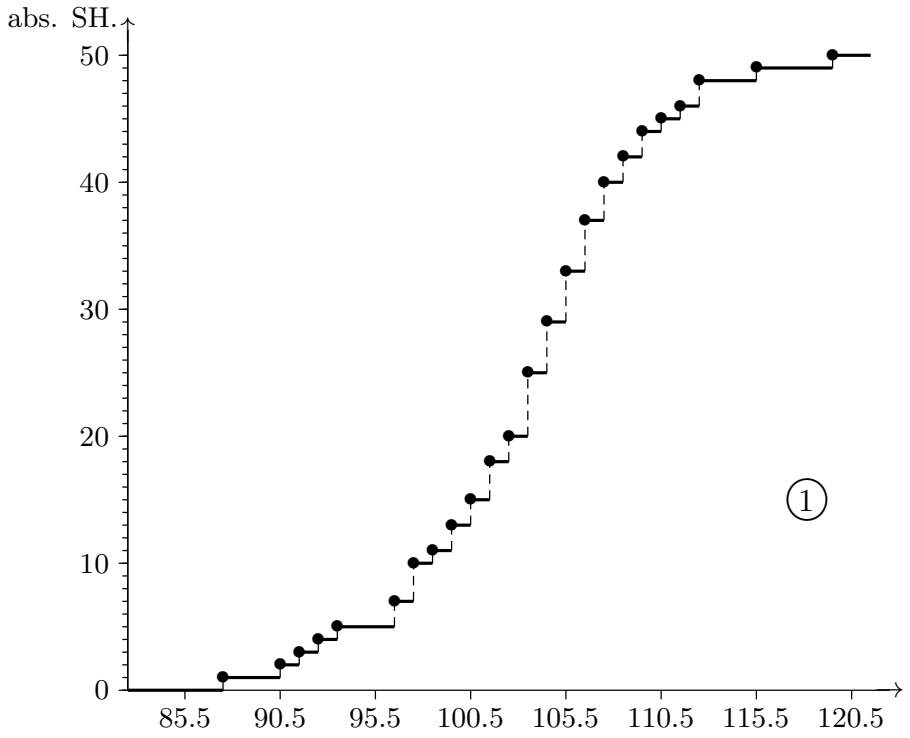
Eine weitere wichtige Darstellung ist die Summenhäufigkeitsverteilung (empirical cumulative distribution function). DISKRET: Wir betrachten erneut die Noten der Prüfung. Frage sei: **"Wie viele Personen haben eine ungenügende Note?"** Das ist die Frage nach der Anzahl (oder Anteil) Personen mit **Note** ≤ 3.5. Dies ist leider nicht direkt aus der Tabelle oder dem Stabdiagramm ablesbar. Man muss die Häufigkeit der Noten bis ≤ 3.5 addieren. Dies macht man vorteilhaft gleich für alle möglichen Fragen dieser Art und kommt dann zur *Summenhäufigkeit oder kumulativen Häufigkeit.* Wir gelangen so zu

Note	abs. H.	rel. H.	abs. SH.	rel. SH.
1	1	0.025	1	0.025
$1\frac{1}{2}$	0	0	1	0.025
2	0	0	1	0.025
$2\frac{1}{2}$	1	0.025	2	0.050
3	6	0.150	8	0.200
$3\frac{1}{2}$	4	0.100	12	0.300
4	7	0.175	19	0.475
$4\frac{1}{2}$	9	0.225	28	0.700
5	5	0.125	33	0.825
$5\frac{1}{2}$	4	0.100	37	0.925
6	3	0.075	40	1.000

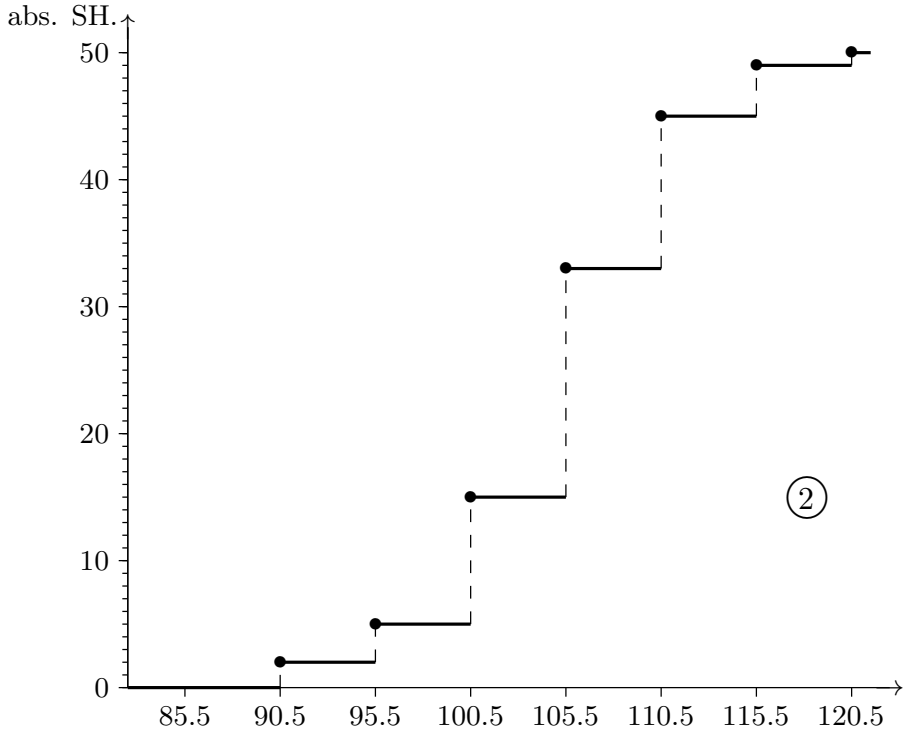


Wenn Sie sich jetzt daran stören, dass für diese 2 Graphiken eine ganze Seite verwendet werden musste: gut so! Denn es ist hiermit genau aufgezeigt, was passiert, wenn die Sprünge unten so hoch sein müssen wie die Balken oben. Machen Sie also beim Einstieg wenn möglich die Balken oben nicht so gross!

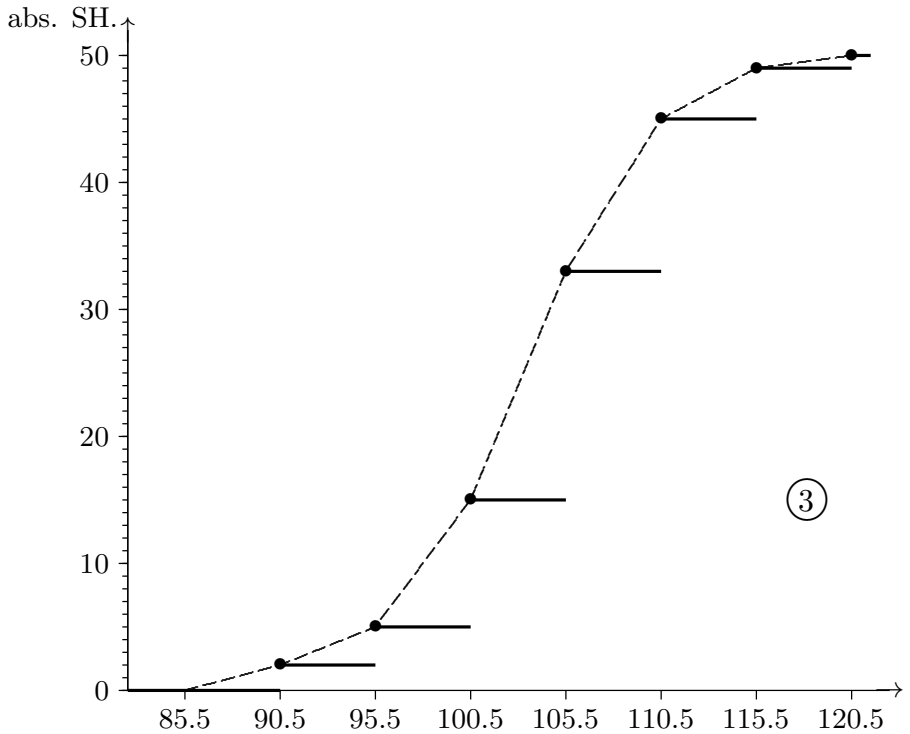
STETIG: Im Kükenbeispiel tritt hier ein kleines Problem auf: weil das Merkmal stetig ist und wir Klassen gebildet haben, kommen sämtliche Gewichte, welche eigentlich im Intervall $(86.5, 87.5]$ wären, auf einem Punkt 87 zu liegen. Wenn man dann die graphische Darstellung der Summenhäufigkeit bildet, kann man sich (zu Recht) darüber streiten, wo der Sprung stattfinden sollte (Klassenmitte oder Ende).



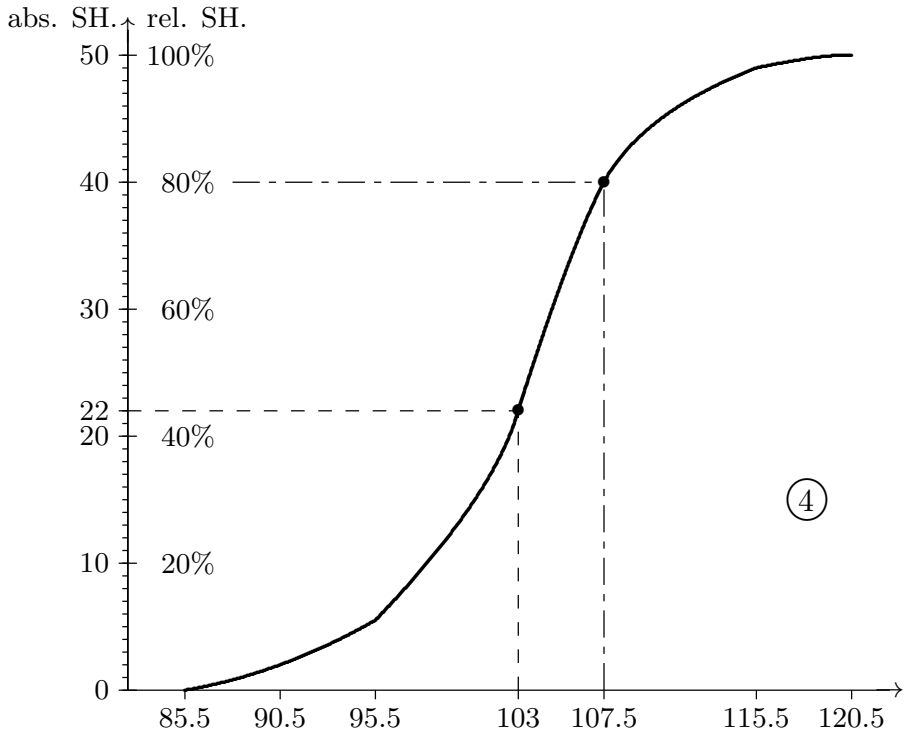
Weiter hat eine grosse Klassenbreite hier einen ganz klaren Nachteil; der Verlust an Information kommt klarer zum Vorschein als bei den Histogrammen:



Verbinden wir in der Figur 2 von (2.2.2.7) die Eckpunkte der "Treppe" durch Geradenstücke, so erhalten wir einen Streckenzug (gestrichelte Linie in Figur 3).



Wird die Klasseneinteilung verfeinert, d.h., unterwerfen wir Figur 1 demselben Prozess, so nähert sich der Streckenzug einer glatten Kurve. Verfeinern wir in Gedanken die Klasseneinteilung noch mehr, so erhalten wir als Idealisierung eine glatte Kurve, die etwa so aussieht:



Eine solche Kurve kann man dazu benützen, gewisse Daten auf einfache Weise abzulesen. Wir illustrieren das Vorgehen an Figur 4.

a) Wieviele Küken wiegen ≤ 103 Gramm?

   Wir gehen von 103 aus entlang der gestrichelten Geraden nach oben bis zur Kurve und dann nach links. Wir finden so die Antwort: 22.

b) Für welches Gewicht x gilt, dass 80% der Tiere höchstens x Gramm wiegen?

   Hier gehen wir von der Zahl 80% auf der Ordinate entlang der strichpunktierten Geraden bis zur Kurve und dann senkrecht nach unten. Als Antwort finden wir $x = 107.5$.

2.2.3 Statistische Masszahlen

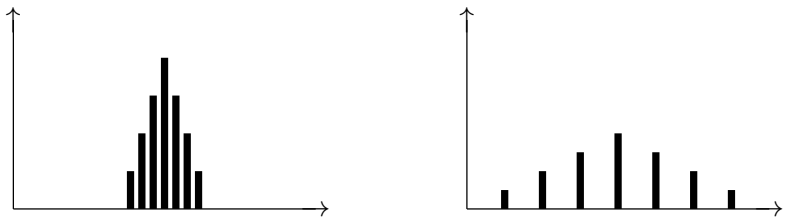
"Die Armen werden immer ärmer,
die Reichen immer reicher!"

Wer von Ihnen glaubt, dass diese Aussage richtig ist?

Wir haben auf der letzten Seite bei einer bekannten und trivialen Aussage gesehen, dass Daten sehr komplex sein können. Wünsche darf man haben: wir wollen komplexe Datensätze mit möglichst wenigen Zahlen möglichst umfassend beschreiben.
Für's erste dienen dazu **Lagemasse** (measures of location; Durchschnitt (average, mean), Median (median), Modus (mode))

Lagemass klein Lagemass gross

sowie **Streuungsmasse** (measures of scale; Variationsbreite, Interdezilbereich, Varianz (Variance), Standardabweichung (Standard deviation)).



kleine Streuung grosse Streuung

Betrachten wir erstmal das arithmetische Mittel (Durchschnitt) einer Folge von $n = 7$ Zahlen (zB Noten):

$$3.5, 4, 4, 6, 2.5, 6, 5.$$

Der Durchschnitt ist hier:
Allgemein definieren wir für eine Folge von n reellen Zahlen

$$x_1, x_2, \ldots, x_n$$

das arithmetische Mittel (oder auch Durchschnitt) als

$$\overline{x} := \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Im Buch 2.2.3 finden Sie noch den Fall der Berechnung eines Durchschnitts, wenn die Daten in Klassen eingeteilt sind.

In der einführenden Diskussion haben wir gesehen, dass bei Löhnen nicht einfach über den Durchschnitt gesprochen werden sollte. Mindestens der Median $\tilde{x}$ sollte auch angeschaut werden. Was ist der Median in obigem Datensatz von Noten

$$3.5, 4, 4, 6, 2.5, 6, 5?$$

Problem wenn n gerade Zahl?

Vorteil von $\tilde{x}$ gegenüber $\overline{x}$?

Im Buch finden Sie noch den Fall der Berechnung von $\tilde{x}$, wenn die Daten in Klassen eingeteilt sind.

Als letzter Ausdruck folgt der Modus: Im Zürcher Kantonsrat gilt in der Legislaturperiode 2023–2027 die folgende Sitzverteilung (Zahlen vom Wahlabend vor Parteiwechseln): AL 5, EDU 3, EVP 7, FDP 29, Grüne 19, Grünliberale 24, Die Mitte 11, SP 36, SVP 46. Der Modus liegt hier bei der SVP.

Arithmetisches Mittel, Median und Modus können alle gleich sein (Situation A) oder auch zum Beispiel alle verschieden (Situation B):

Ein Artikel vom Dozenten: https://schweizermonat.ch/durchschnitt-oder-median

Wir kommen zu den Streuungsmassen - zwei davon kann man sehr schnell abhaken: die Variationsbreite ist einfach der grösste Wert minus den kleinsten Wert (bei den Noten also $6 - 2.5 = 3.5$). Das ist einfach die Spannweite der Daten. Man sollte nie nur diese Zahl anschauen (Ausreisser! (outlier)). Wir können bei den der Grösse nach geordneten Daten einfach die kleinsten 10 % und die grössten 10 % der Daten ignorieren und die Spannweite der verbleibenden Daten berechnen - das ist der Interdezilbereich. Man ist dann schon robuster gegenüber Ausreissern.

Nach diesen (sehr einfachen) Massen kommen zwei zentral wichtige Masse: Varianz und Standardabweichung. Wir wollen ein sinnvolles Mass für die Streuung der Daten um das Lagemass (zum Beispiel um das arithmetische Mittel) definieren. Wenn wir Daten

$$x_1, x_2, \ldots, x_n$$

mit arithmetischem Mittel $\overline{x}$ haben, so kann man die Werte

$$(x_1 - \overline{x}), (x_2 - \overline{x}), \ldots, (x_n - \overline{x})$$

berechnen (Differenz zum arithmetischen Mittel). Berechnen Sie hiervon den Durchschnitt:

Nach diesem Fehlschlag fallen folgende Vorschläge ein, um ein sinnvolles Streuungsmass zu definieren:

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

und

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

Beide Masse werden (zu Recht) verwendet. Wir werden uns in dieser Vorlesung auf den zweiten Ausdruck mit dem Quadrat konzentrieren (es ist mathematisch übrigens einfacher).

Die Summe

$$SS_{xx} := \sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad (\textbf{S}\text{um of }\textbf{S}\text{quares x and x})$$

und Varianten werden in dieser Vorlesung noch oft vorkommen - untersuchen wir sie:

Wir definieren die empirische Varianz s^2 als

$$s^2 := \frac{1}{n-1} SS_{xx}.$$

Wir werden später in dieser Vlsg begründen, warum wir hier durch $(n-1)$ statt durch n teilen. Nach obigen Berechnungen haben wir also die folgenden drei gleichwertigen Formeln zur Auswahl:

$$\begin{aligned} s^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 \\ &= \frac{1}{n-1} \big(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 \big) \\ &= \frac{1}{n-1} \big(\sum_{i=1}^{n} x_i^2 - \frac{1}{n} (\sum_{i=1}^{n} x_i)^2 \big). \end{aligned}$$

Berechnen Sie bei den Noten $3.5, 4, 4, 6, 2.5, 6, 5$ die empirische Varianz (HA): 1.7024

Teilt R (oder Ihr TR) durch n oder $n-1$?

```
> ?var
> a<-c(2,3,4)
> mean(a)
[1] 3
> var(a)
[1] 1
```

Das letzte Streuungsmass, welches wir definieren, ist die empirische Standardabweichung (in R "sd(a)" für Englisch **s**tandard **d**eviation):

$$s := \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2} \geq 0.$$

Bemerkung zu Varianz und Standardabweichung: Die Varianz ist ein quadratisches Mass; wenn wir davon die Wurzel ziehen, erhalten wir ein Mass, welches mit den Daten direkt in Bezug gesetzt werden kann.

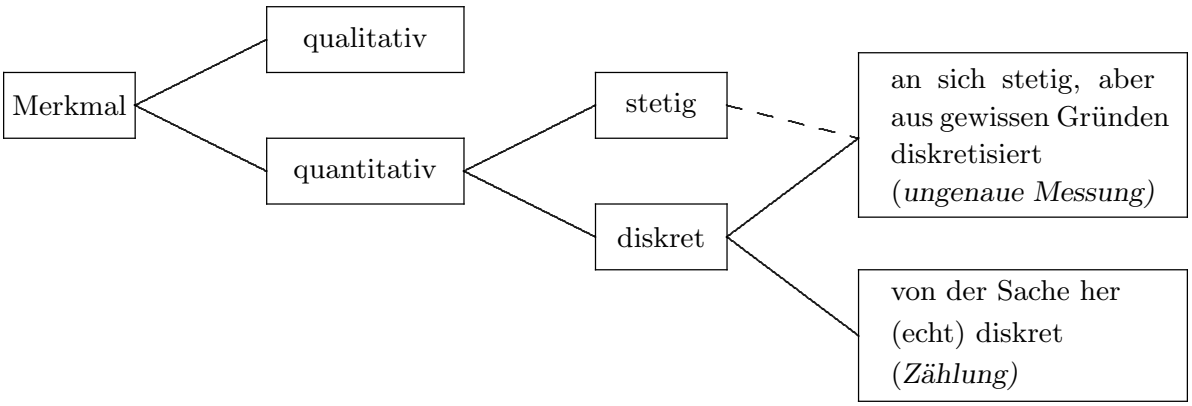
Aufgabe 2-13: Die Zahlen $x_1, x_2, \ldots, x_n$ seien gegeben. Gesucht ist die Zahl x, für welche der Ausdruck $\sum_{i=1}^{n}(x_i - x)^2$ minimal wird.

Untersuchen Sie $\bar{x}$ und s^2 wenn $n = 1$:

Wichtig:

1. Lesen Sie jetzt das komplette Kapitel 2 im Buch selber durch.
2. Lösen Sie danach mindestens 5 Aufgaben hinten im Kapitel und vergleichen Sie mit den Lösungen am Schluss des Buches. Bei Bedarf lösen Sie mehr Aufgaben.
3. Gehen Sie in die Übungsstunde. Drucken Sie das Übungsblatt dazu *vorher* aus, lesen Sie *vorher* die Aufgaben durch und machen sich erste Gedanken dazu (zum Beispiel, wie man sie lösen könnte).
4. Dann lösen Sie das Übungsblatt: zuerst immer selber probieren, falls nicht geht: Tipp von Mitstudi benutzen, falls immer noch nicht geht: Lösung von Mitstudi anschauen, 1 Stunde warten, versuchen, aus dem Kopf heraus wieder zu lösen, falls immer noch nicht geht: Lösung von Mitstudi abschreiben (und verstehen - also sollte man insbesondere keine Fehler abschreiben!).
5. Lösen Sie die entsprechenden Prüfungsaufgaben im Archiv.

Lessons learnt:



Des weiteren kann man sich überlegen, welche (Rechen-)Operationen wo sinnvoll sind. Dazu unterscheiden wir 4 Arten von Merkmalen:

Nominalskala	Keine Rangordnung; Rechenoperationen sinnlos (qualitatives Merkmal)
Ordinalskala	**Rangordnung festgelegt**; Rechenoperationen sinnlos (qualitatives Merkmal)
Intervallskala	Rangordnung festgelegt; **Abstände zwischen Zahlen haben Bedeutung**, Nullpunkt willkürlich, Bildung von Verhältnissen (und Prozenten) nicht sinnvoll (quantitatives Merkmal)
Verhältnisskala	Rangordnung festgelegt; Abstände zwischen Zahlen haben Bedeutung, **Nullpunkt absolut**, Bildung von Verhältnissen (und Prozenten) sinnvoll (quantitatives Merkmal)

Stabdiagramm (diskret) vs Histogramm (stetig)

Summenhäufigkeitsverteilung

Lagemasse (Median, arith Mittel) vs Streuungsmasse (Varianz, Standardabweichung)

Klickerfragen zum Aufwärmen:

Frage 1: Welches der folgenden Merkmale ist kein Beispiel für eine nominalskalierte Variable?

(a) Haarfarbe einer Person
(b) Telefonnummer einer Person
(c) Gewicht einer Person

Frage 2: Welches der folgenden Merkmale ist qualitativ?

(a) Alter einer Person

(b) Beruf einer Person

(c) Anzahl der Geschwister einer Person

Frage 3: Welches der folgenden Merkmale ist diskret?

(a) Temperatur in Celsius

(b) Anzahl der Kinder pro Familie

(c) Körpergewicht einer Person

Frage 4: Gegeben ist eine Stichprobe von 50 Küken mit Gewichtsangaben in Gramm. Die Anzahl der Küken in bestimmten Gewichtsbereichen ist:

Gewicht (g): Anzahl Küken

90 - 99: 8

100 - 109: 22

110 - 119: 14

120 - 129: 6

Berechnen Sie die relative Häufigkeit der Küken im Bereich 100 - 109 g.

(a) 0.22

(b) 0.28

(c) 0.44

Frage 5: Die folgenden Prüfungsnoten wurden erfasst:

$3.5, 4.0, 4.0, 6.0, 2.5, 6.0, 5.5$

Berechne das arithmetische Mittel.

(a) 4.0

(b) 4.25

(c) 4.5

Frage 6: Die folgenden Prüfungsnoten wurden erfasst:

$2.5, 3.5, 4.0, 4.0, 5.0, 5.5, 6.0$

Was sind Median und Modus?

(a) Median = 4.0, Modus = 4.0

(b) Median = 4.5, Modus = 4.0

(c) Median = 4.0, Modus = 6.0

Frage 7: Welches der folgenden Zahlenpaare hat die grösste Standardabweichung?

(a) 3, 3, 4, 4, 5

(b) 1, 3, 5, 7, 9

(c) 5, 5, 5, 5, 5

Frage 8: Welche Lagemasse behandeln wir in dieser Vorlesung? Welche Streumasse behandeln wir in dieser Vorlesung?

Frage 9: Gegeben sind die folgenden fünf Werte:

$2, 4, 6, 8, 10$

Was ist das arithmetische Mittel dieser Zahlen?

(a) 5

(b) 6

(c) 7

Frage 10: Gegeben sind die folgenden Werte:

$2, 3, 3, 4, 4, 4, 5, 6, 6, 6, 6$

Was ist der Modus dieser Datenreihe?

(a) 4

(b) 5

(c) 6

Lösungen zu Klickerfragen:

Frage 1: (c) Gewicht einer Person; Erklärung: Nominalskalen haben keine natürliche Rangordnung - Haarfarbe und Telefonnummer sind nominal, da sie nur als Kategorien dienen - Gewicht ist metrisch (Verhältnisskala), da Differenzen und Verhältnisse bedeutungsvoll sind; Frage 2: (b) Beruf einer Person; Frage 3: (b) Anzahl der Kinder pro Familie; Frage 4: (c): absolute Häufigkeit durch Gesamtzahl der Beobachtungen , also $\frac{22}{50} = 0.44$; Frage 5: (c) $\bar{x} = \frac{3.5+4.0+4.0+6.0+2.5+6.0+5.5}{7} = \frac{31.5}{7} = 4.5$; Frage 6: (a) Median = 4.0, Modus = 4.0; Frage 7: (b) 1, 3, 5, 7, 9. Erklärung: Bei (a) liegen die Werte näher an ihrem Mittelwert als bei (b). (c) hat eine Standardabweichung von 0, weil die Werte gleich sind. ; Frage 8: 1. Median, arithmetisches Mittel, Modus; 2. Varianz, Standardabweichung (dazu noch Variationsbreite und Interdezilbereich); Frage 9: (b) $\bar{x} = \frac{\sum x_i}{n} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6$; Frage 10: (c) 6. Der Modus ist der Wert, der am häufigsten vorkommt.