

Modul 162

Daten analysieren und modellieren

Block 1: Daten



Inhalt

1	Daten	3
1.1	Was sind Daten?	3
1.2	Information und Nachricht	4
1.3	Redundanz	5
1.4	Dateiformate	6
1.4.1	Textbasierte Dateien	8
1.4.2	Spezielle Textformate zum Datenaustausch	9
1.4.3	Binäre Dateien	9
1.5	Datentypen	10
2	Datenstrukturen	13
2.1	Datenanalyse	13
2.2	Datenstrukturen	13
2.2.1	Elementare Datenstrukturen	13
2.2.2	Zusammengesetzte Datenstrukturen	14
2.2.3	Hierarchische Datenstruktur	14
2.2.4	Verarbeitungsbezogene Strukturierungsmerkmale	15
2.3	Programmstrukturen	15
2.4	Übung zu Datenstrukturen	16
3	Daten aufbereiten und auswerten (Statistik)	17
3.1	Mittelwert	17
3.2	Mittelwert mit Klassen	18
3.3	Median	19
3.4	Standardabweichung	20
3.5	Zusätzliche Aufgaben	21

1 Daten

1.1 Was sind Daten?

Im Alltag sind wir dauernd mit «Daten» konfrontiert. Die folgende Liste zeigt ein paar solcher Daten. Notieren Sie zu den einzelnen Daten Ihren möglichen Sinn!

13.15	- Uhrzeit - eine physikalische Grösse, z.B.: 13.15m, 13.15 kg usw. - Zahl mit Kommastellen
05.06.2010	- Datum - eine Produkt- / Artikelnummer - Bankkontonummer
0793451288	- Telefonnummer - Passwort - Zahl
Max Kummer	- Vorname und Nachname - Firmenname, Marken - Zeichenkette
	- Verbotsschild "Rauchen verboten" - Kunst, Art, Grafik, Symbol - Foto
夢	- chinesisches Schriftzeichen - Kalligraphie - Grafik, Kunst, Tattoo

Sobald die oben aufgeführten **Daten für den Betrachter einen Sinn ergeben**, z.B. ist «Max Kummer» mit grösster Wahrscheinlichkeit ein Vor- und Nachname, sprechen wir von einer **Information**.

1.2 Information und Nachricht

Der Begriff «Information» wird folgendermassen definiert:

Informationen sind im Sinne der Umgangssprache die Kenntnisse und das Wissen über Sachverhalte, Vorgänge, Zustände, Ereignisse usw. Sie können durch gesprochene und geschriebene Wörter, durch Tabellen (Datenbanken) und Diagramme oder durch Bilder und Grafiken dargestellt, gespeichert und verbreitet werden.

In der Informatik und in der Kommunikationstechnik werden die Informationen durch Zeichen dargestellt. Beispiele für solche Zeichen sind die Buchstaben des Alphabets, Ziffern, Interpunktionszeichen, Steuerzeichen usw. (siehe ANSI-, ASCII-Tabelle)

In der Kommunikationstechnik werden Zeichenfolgen (eine definierte Anzahl von einzelnen Zeichen) zur Übertragung einer Information verwendet. Dies bezeichnet man als **Nachricht**.

Nachricht: Wenn Daten übertragen werden
Information: Wenn Daten etwas beinhalten

1.3 Redundanz

In der Informatik werden Informationen als redundant bezeichnet, wenn sie mehrfach vorkommen.

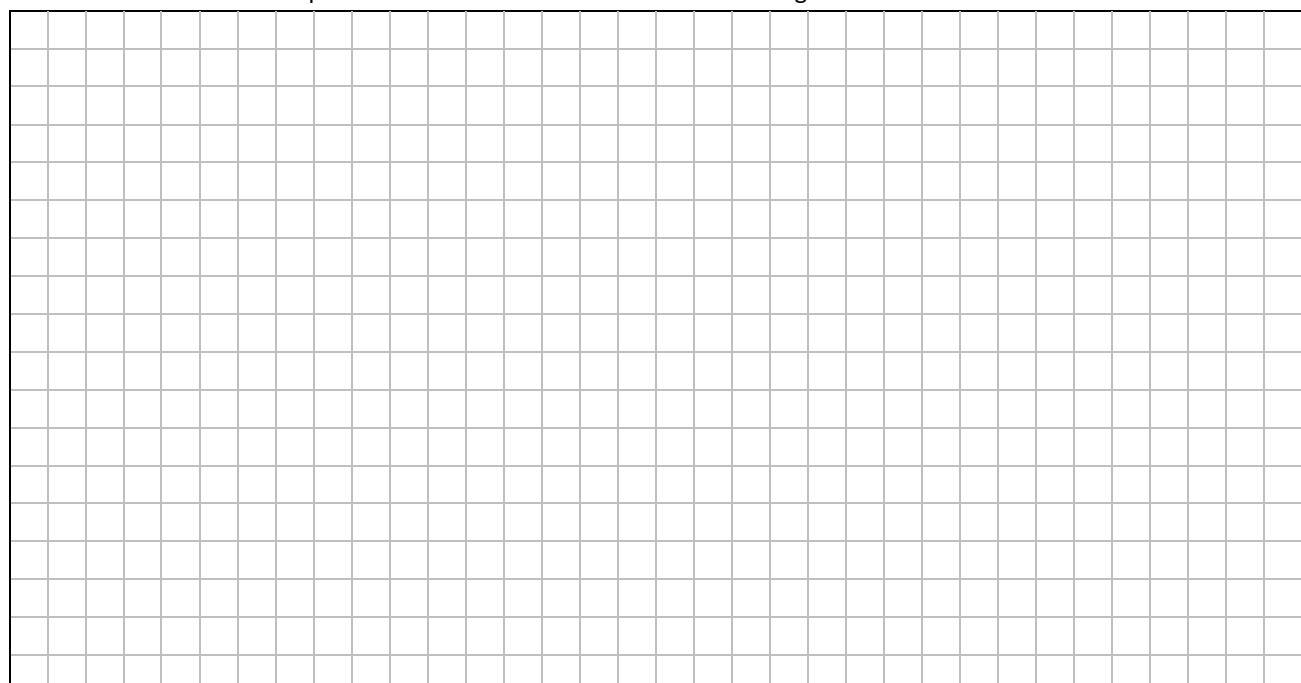
Im Alltag erlebt man dauernd redundante Informationen. So erfahren Sie zum Beispiel an einem Sonntag das Resultat des laufenden FCL-Matches im Radio, im Internet, von Kollegen per Handy oder sehen den Match live. Sie erhalten die gleiche Information mehrmals, ohne dass Sie etwas Neues erfahren. Diese Information ist redundant.

Im Alltag haben wir uns an diese Situation gewöhnt. In der Informatik gibt es Situationen in denen eine Redundanz der Daten erstrebenswert ist oder eben unter keinen Umständen gewünscht wird.

Überlegen Sie sich zwei Situationen in der Informatik bei denen eine Redundanz von Informationen gewünscht wird bzw. unerwünscht ist. Notieren Sie je eine Situation und erklären Sie kurz warum eine Redundanz erwünscht ist oder nicht!

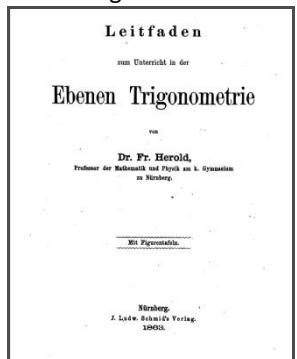
Redundanz erwünscht	Redundanz unerwünscht
<ul style="list-style-type: none">- Datensicherung, Backup- Server replizieren	<ul style="list-style-type: none">- Datenbank --> genau die gleiche Information mehrmals vorhanden ist

Redundanzen werden uns später in diesem Modul noch weiter beschäftigen



1.4 Dateiformate

Heutzutage sind fast alle Informationen elektronisch gespeichert. Sogar alte Bücher werden eingescannt und auf dem Internet zur Verfügung gestellt:



Wenn Sie in einem Internetbrowser «books.google.ch» aufrufen und das Stichwort «Trigonometrie» eingeben, finden Sie neben vielen anderen Büchern einen Leitfaden zur ebenen Trigonometrie von F. Herold aus dem Jahr 1863. (siehe nebenstehendes Bild)

Dies ist nur ein Beispiel wie heutzutage ganze Bücher systematisch eigescannt werden und nicht nur in irgendwelchen Bibliotheken verstaubten.

Bei der Speicherung der Daten unterscheidet man verschiedene Speicherformate. Die verschiedenen Dateiformate erkennt man auf einem PC-Betriebssystem an seiner Endung (Suffix), wie zum Beispiel dieses Dokument durch die Endung «.pdf» als ein Dokument im «Portable Document Format» erkennbar ist.

A large, empty grid consisting of many small squares, intended for handwritten notes or drawings.

Aufgabe: Erstellen Sie eine Liste mit mehr als 10 verschiedenen Dateiendungen, die für einen «normalen» PC-Benutzer wichtig sind. Notieren Sie zu jeder Endung das Programm, welches diese Endung verwendet, den Programmtyp und ein Verwendungsbeispiel, nachfolgendem Schema:

Endung	Programm / Typ	Verwendung
docx	Word / Textverarbeitung	Skript zum Modul 162
pptx	Powerpoint / Präsentationen	Präsentation über ein Projekt
xlsx	Excel / Verwaltung Daten (Tabellenkalkulationsprogramm)	Berechnungen, Notenlisten
pdf	Adobe Acrobat Reader / Textverarbeitung	Skript zum Modul 117 oder HA
jpg	Fotos / Bilder	Fotos in komprimierter Form
html	HTML / Visual Studio Code / Auszeichnungssprache	Für Erstellung und Bearbeitung von Webseiten
py	Python / Programmiersprache	Speicherung von der Datei vom Code
css	CSS / Visual Studio Code / Stylesheet-Sprache	Für Bearbeitung von Webseiten
java	Java / intelliJ / Programmiersprachen	Speicherung von der Datei vom Code
txt	Editor / Bearbeitungsprogramm von txt-Dateien	Um txt-Dateien zu bearbeiten, einfache und unformatierte Texte
mp4	Videodatei in MP4-Format / VLC Video	Videodatei als Erklärung

Sie kennen jetzt einige verschiedene Dateiformate. In der EDV unterscheidet man dabei grundsätzlich zwischen zwei verschiedenen Formaten:

- Dateien die mit einem einfachen Editor (Notepad) lesbar sind → **textbasierte Dateien**
- Dateien die nicht direkt lesbar sind → **binäre Dateien**

1.4.1 Textbasierte Dateien

Das klassische und älteste Dateiformat hat die Endung «txt». Die Informationen in dieser Datei sind als einzelne lesbare Zeichen gespeichert. Da ein Computer nur Zahlen speichern kann, verbirgt sich hinter jedem Zeichen «A», «B», «a», «b» usw. eine Zahl. Damit ein Editor nun weiß welches Zeichen er für eine gespeicherte Zahl darstellen soll, gibt es entsprechende Tabellen (→ Zeichentabellen bzw. Codepages). Diese Tabellen können pro Betriebssystem und verwendete Sprache unterschiedlich sein. Wir kennen hauptsächlich zwei solche Tabellen:

- ASCII-Tabelle (American Code for Information Interchange)
- ANSI-Tabelle (American National Standards Institute)

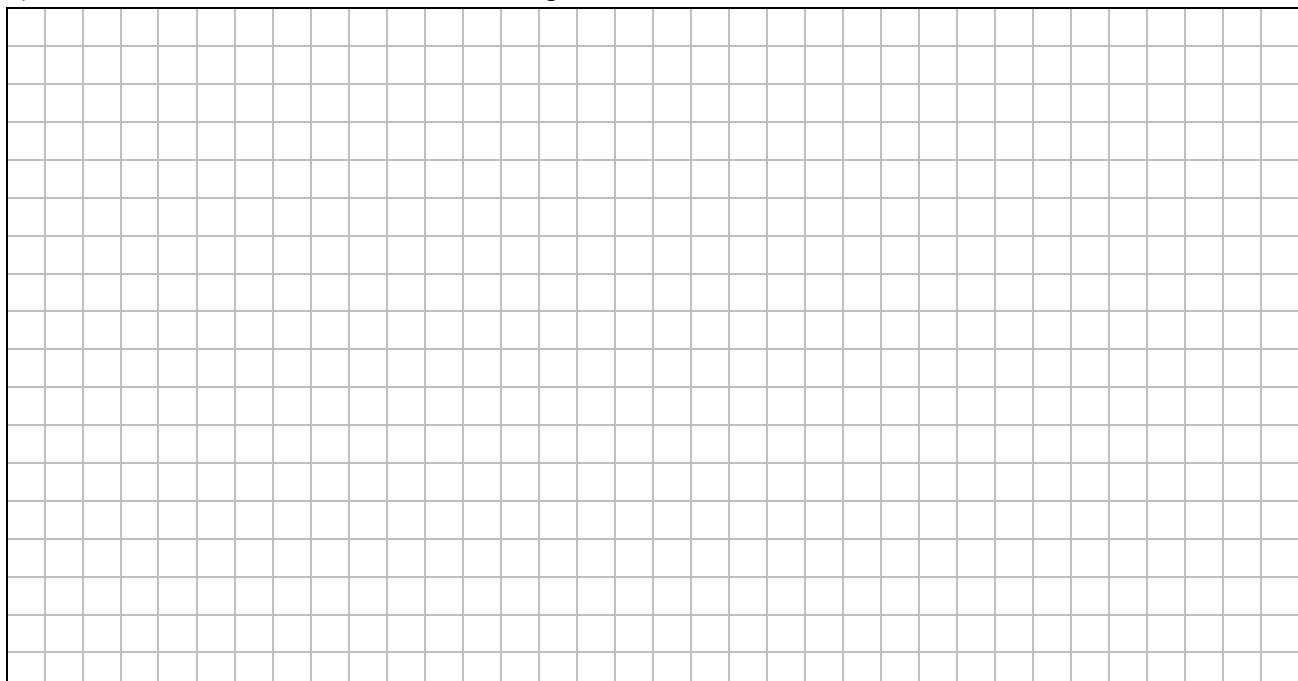
Die ASCII-Zeichentabelle wurde bei DOS-basierten PCs und bei Grosssystemen verwendet und dient vor allem zum Datenaustausch zwischen nicht kompatiblen Systemen. Die ANSI-Zeichentabelle wird beim Windowsbetriebssystem verwendet.

Aufgabe: Führen Sie die folgenden Schritte aus und beantworten Sie dann die untenstehenden Fragen:

- Starten sie den Betriebssystem-Editor, indem Sie unter dem Startmenü im Suchfeld «notepad» eintippen gefolgt von der Enter-Taste.
- Nun tippen Sie «Alt»+065 (Dies bedeutet: «Alt»-Taste drücken und gedrückt lassen und im numerischen Tastenfeld «0», «6» und «5» eingeben und die «Alt»-Taste loslassen)
→ Es sollte ein «A» erscheinen.
- Danach versuchen Sie «Alt»+0169. Wenn das «©»-Zeichen erscheint, haben Sie es richtig gemacht.

a) Erklären Sie warum man die Zeichen auch so eingeben kann?

b) Wie kann man das «©»-Zeichen sonst noch eingeben kann?



1.4.2 Spezielle Textformate zum Datenaustausch

Neben den unformatierten Textdateien haben sich auch andere Textbasierte Formate zum Datenaustausch zwischen verschiedenen Plattformen und Programmen etabliert:

- **rtf** (Rich Text Format) → für Textverarbeitungsprogrammen
- **xml** (Extensible Markup Language) → für Datenbanken
- **csv** (Comma-Separated Values) → für Tabellenkalkulationen

1.4.3 Binäre Dateien

Bei den meisten der **verwendeten Dateiformate** werden die **Informationen binär gespeichert**. Bei diesen Formaten ist es meistens so, dass immer ein binärer Header (Informationen am Anfang der Datei) oder Footer (Informationen am Ende der Datei) mit allgemeinen Informationen und danach die eigentlichen Daten gespeichert sind. Wie dieser Header aufgebaut ist, kann der jeweiligen Dokumentation entnommen werden.

Aufgabe: In einer mp3-Datei sind zusätzlich zu den Sounddaten auch noch weitere Informationen wie Interpret, Titel, usw. gespeichert. Finden Sie heraus wie und wo diese Informationen gespeichert sind. Suchen Sie dazu Informationen zum ID3v1-Tag:

Der Interpret ist oben beim Dokument gespeichert. Der Titel ist unten gespeichert. Man sieht kryptische Zeichen. Der ID3v-1 ist 128 Byte gross und steht am Ende einer MP3-Datei:

1.5 Datentypen

In der EDV werden die vorhandenen Daten in definierten Formaten gespeichert. Dabei unterscheidet man grob zwischen Text- und Zahlendaten. Je nach Anwendung (Tabellenkalkulation, Datenbank, Programmierung usw.) gibt es aber auch noch viele andere Datenformate oder Datentypen.

Die Programmiersprache C kennt dabei:

- integer ganze Zahlen
- float, double reelle Zahlen
- char einzelnes Zeichen

usw.

Aufgabe: Starten Sie die Tabellenkalkulation Excel und finden Sie heraus welche Datentypen bzw. Datenformate es bei Excel gibt.

a) Schreiben Sie die wichtigsten Datentypen auf!

b) Überlegen Sie sich wie die Daten in den einzelnen Feldern (Zellen) gespeichert werden!

a)

b)

Es gibt im Excel nur Text oder Zahl.

Wir haben festgestellt, dass Excel (Tabellenkalkulationsprogramm) pro Feld (Zelle) keinen fixen Datentyp kennt, sondern eigentlich nur eine Formatierung. Wie sieht es bei Datenbanken aus?

Aufgabe: Starten Sie Access und erstellen Sie eine leere Datenbank, z.B. mit dem Namen: «Test».

Stellen Sie die Ansicht auf „Entwurfsansicht“ um und übernehmen Sie den vorgeschlagenen Tabellennamen «Tabelle1»:

Erstellen Sie nun in der Entwurfsansicht zwei neue Felder:

- «Namen» → Felddatentyp: «Text»
- «Wert» → Felddatentyp: «Zahl»

Beachten Sie, dass es bei der Auswahl der Feldnamen

Einschränkungen gibt. So dürfen, wie bei der Programmierung auch, Eigennamen, wie «Text», «Name» usw. nur eingeschränkt verwendet werden.

Wechseln Sie danach die Ansicht wieder auf «Datenblattansicht» und geben Sie in den neuen Feldern Daten ein!

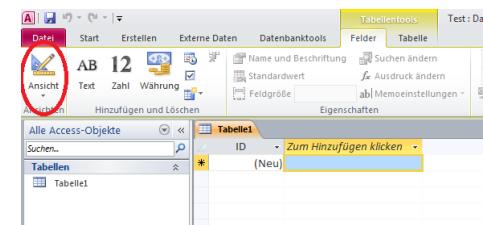


Tabelle1	
Feldname	Felddatentyp
ID	AutoWert
Namen	Text
Wert	Zahl

Versuchen Sie im «Wert»-Feld einen Text und im «Namen»-Feld eine Zahl einzugeben. Was stellen Sie fest?

In den Datenbanken kann man nur Werte speichern, welche auch ihrem Datentyp entsprechen --> es ist typensicher.

Aufgabe: Ergänzen Sie die folgende Tabelle mit den wichtigsten acht Felddatentypen von Access, die auch bei anderen Datenbanken (MySQL, SQL Server, usw.) vorkommen, indem Sie jeden Datentyp kurz beschreiben und ein Beispiel machen:

Felddatentyp	Beschreibung / Beispiel
Kurzer Text (bis Office 2013: Text)	String mit max. 255 Zeichen --> Vorname, Klassenbezeichnung usw.
Langer Text (bis Office 2013: Memo)	String mit mehr als 255 Zeichen --> Produktbeschreibung
Zahl	Zahlenwerte mit und ohne Kommastellen --> Gewicht, Länge usw.
Datum/Uhrzeit	Datum und Zeit --> 19.09.2022 13:55
AutoWert	automatische fortlaufende ganze Zahl --> Wird als eindeutige Identifikationsnummer verwendet --> Kundennummer
Ja/Nein	Boolean, Wahrheitswert --> Newsletter?, Nachessen mit Übernachtung?
Anlage	Bild, Sound (mp3) usw.
OLE-Objekt	Ein eingebettetes Objekt, wie z.B. eine Excel-Tabelle in einem Word-Dokument. E --> embedding (Einbettung) L --> linking (verlinken) O --> Object (Objekt)

2 Datenstrukturen

2.1 Datenanalyse

In der Datenanalyse geht es darum festzustellen, welche Daten für die Verarbeitung relevant sind. Dabei sind u. a. folgende Fragen zu klären:

- Welche Daten sollen ausgegeben werden?
→ Ausgabedaten, gewünschte Ergebnisse
- Welche Daten oder Dateien werden benötigt, um die Ausgabedaten erzeugen zu können?
→ Eingabedaten
- Welche Verarbeitungsregeln werden benötigt, um die Eingabedaten zu den gewünschten Ausgabedaten bzw. Ergebnissen zu verarbeiten?
→ Algorithmen, Rechenformeln

Im Zusammenhang mit der Datenanalyse sind die Datenstrukturen zu bestimmen.

2.2 Datenstrukturen

Sind die Daten für die Verarbeitung bekannt, so sind diese Daten zu strukturieren, wobei zu klären ist,

- welche Daten logisch zusammengehören (logische Struktur),
- in welcher Form und in welchem Format Daten zu behandeln sind (formale Struktur),
- wie viel Speicherplatz Daten intern oder auf externen Speichermedien beanspruchen (physische Struktur).

Die Strukturierung von Daten ist für alle Funktionen der Datenorganisation wie Speichern, Suchen, Sortieren und Ausgeben sowie für die Informationsverarbeitung von Bedeutung. Daten werden durch Namen bezeichnet, die Speicherplätze repräsentieren (z. B. ANR für die Artikelnummer, EP für Einzelpreis). Man unterscheidet Variablen und Konstanten. Variablen sind Daten, deren Werte sich im Verlauf der Verarbeitung verändern lassen (z. B. Einzelpreis). Der Name wird als Variablenname, der zugehörige Inhalt als Variablenwert bzw. Variableninhalt bezeichnet. Konstante sind Daten, deren Werte dauerhaft gleichbleiben und durch den Verarbeitungsprozess i. d. R. nicht verändert werden (z. B. Mehrwertsteuersatz). Man unterscheidet elementare und zusammengesetzte Datenstrukturen.

2.2.1 Elementare Datenstrukturen

Elementare Datenstrukturen bestehen aus **einfachen Datentypen**, d. h. eine Variable oder Konstante dieses Typs darf jeweils nur einen Wert annehmen.

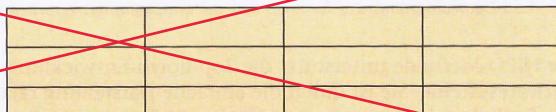
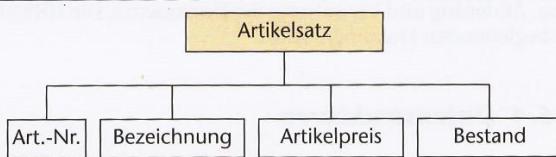
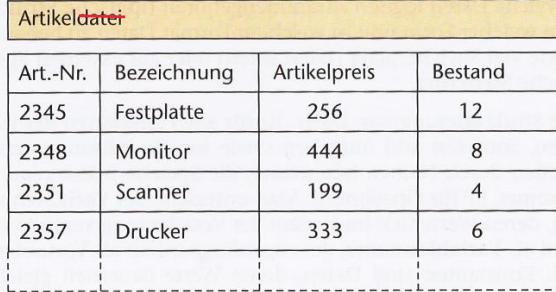
Dieser Wert kann eine Zahl, ein Zeichen oder ein Wahrheitswert sein. Datentypen sind durch ihre **Wertebereiche** charakterisiert (z. B. **Integer**: Eine Ganzzahl, die in 2 Bytes gespeichert wird, besitzt z. B. einen Wertebereich von **-32768 bis 32767**).

Einfacher Datentyp	Erläuterung	Möglicher Variablenwert
Integer	ganze Zahl	2
Real	gebrochene Zahl	2.45
String	Zeichenkette	Computer
Char	ein Zeichen	C
Boolean	Wahrheitswert	1 (für wahr/true) 0 (für falsch/false)

2.2.2 Zusammengesetzte Datenstrukturen

Zusammengesetzte Datenstrukturen bestehen aus mehreren elementaren Datentypen, die unter einem Namen angesprochen werden können (**strukturierte Datentypen**). Die nachfolgende Tabelle veranschaulicht zusammengesetzte Datenstrukturen.

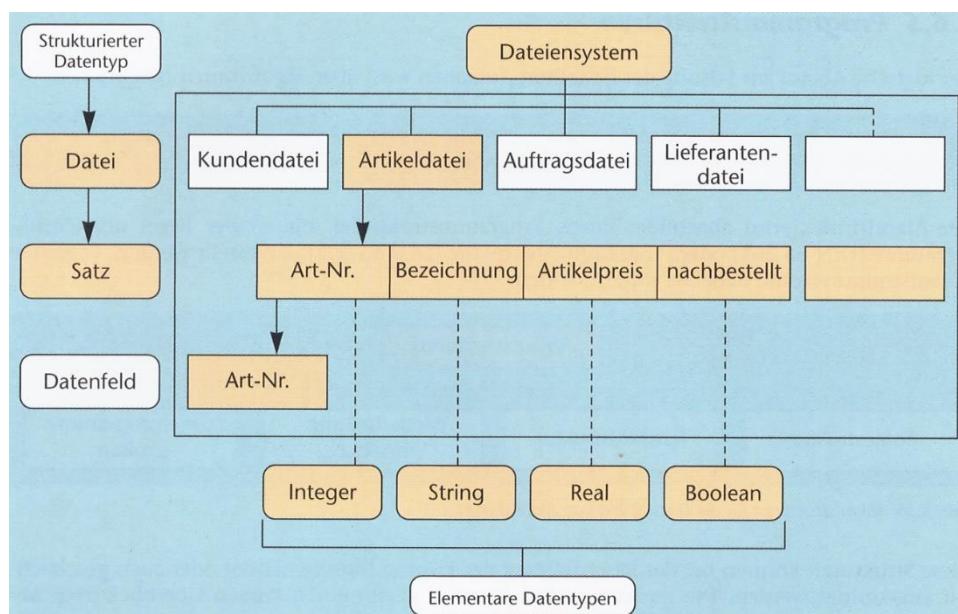
kommt am
Test

Strukturierter Datentyp	Beispiel
Feld (Array) Tabelle als Folge von Elementen, die alle den gleichen Datentyp (z. B. Integer) besitzen. Der Zugriff auf die Feldelemente erfolgt über Indizes (Tabellenverarbeitung).	Beispiel: Tabelle als Datenstruktur 
Satz (Record) Datensatz als Kombination von Datenfeldern unterschiedlichen Datentyps, die in einem logischen Zusammenhang stehen (z. B. Artikelsatz, Kundensatz)	
Datei (File) Tabelle als Zusammenfassung einer variablen Anzahl von gleichartig strukturierten Sätzen (z. B. Artikeldatei, Kundendatei). Der Zugriff auf die Datei erfolgt über den Dateinamen, auf die Sätze i. d. R. über einen Ordnungsbegriff, z. B. die Artikelnummer (Primärschlüssel).	Artikeldatei: 

Eine Zeile in
einer Tabelle

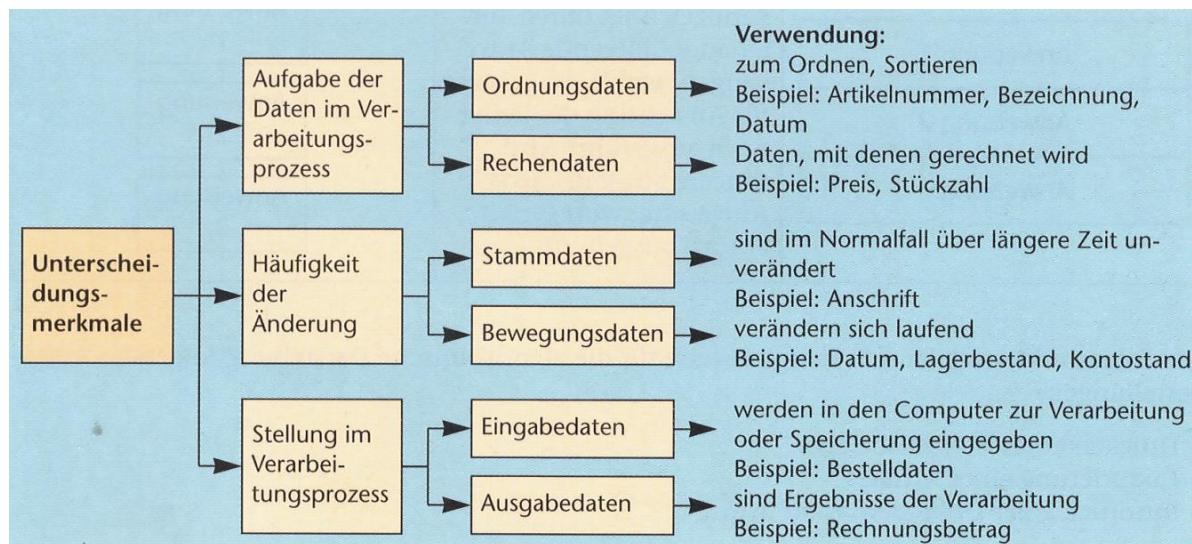
2.2.3 Hierarchische Datenstruktur

Die strukturierte Programmierung ist in der Vergangenheit überwiegend zur Entwicklung von komplexen Dateiverarbeitungssystemen eingesetzt worden. Beim Entwurf von Dateiverarbeitungsprogrammen kann die **hierarchische Datenstrukturierung** sinnvoll sein, um auf der Grundlage der betrieblichen Datenbasis die Schnittstellen im Funktions- und Datenentwurf beschreiben zu können. **Ausgangspunkt** für eine **hierarchische Datenstrukturierung** kann ein **Dateiensystem** sein, in dem mehrere Dateien über wechselseitige Zugriffsmechanismen und Datenelemente in Beziehung zueinanderstehen.



2.2.4 Verarbeitungsbezogene Strukturierungsmerkmale

Daten können zudem nach ihrer Aufgabe und Stellung im Verarbeitungsprozess sowie nach der Häufigkeit der Änderung unterteilt werden.



Aus der festgelegten Datenstruktur wird deutlich, welche Daten in den Funktionen zu verarbeiten sind. Im nächsten Entwurfsschritt sind die Lösungswege für die beschriebenen Funktionen darzustellen, um die Funktionen über geeignete Programmstrukturen abilden zu können. Die Programmstrukturen bilden den Kern der strukturierten Programmierung herkömmlicher Softwaresysteme.

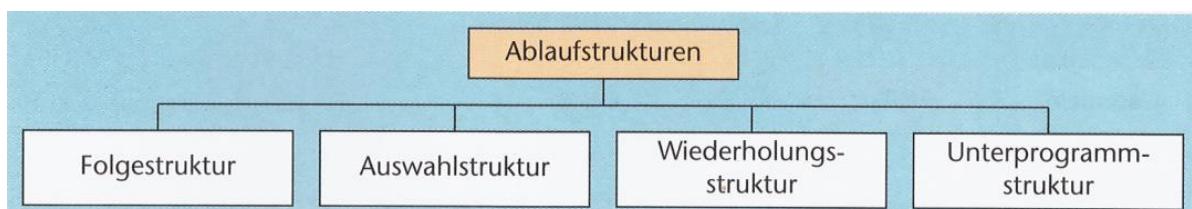
2.3 Programmstrukturen

Der logische Ablauf zur Lösung der gestellten Aufgaben wird über Algorithmen beschrieben.

Unter einem **Algorithmus** versteht man die eindeutige und vollständige Beschreibung der Lösungsschritte und deren zeitliche Verarbeitungsfolge.

Der Algorithmus wird abgebildet durch **Programmstrukturen**, die in der Regel über **Struktogramme** oder **Programmablaufpläne** dargestellt werden.

Folgende Ablaufstrukturen sind dabei zu unterscheiden:



2.4 Übung zu Datenstrukturen

Aufgabestellung: Wir analysieren die Daten, die unsere Klasse betreffen.

Ausgangslage: Klassenliste unserer Klasse.

Ziel: Konkrete Daten aus der Klassenliste und dem Unterricht zu diesen Strukturierungsmerkmalen zuordnen und die Verwendung mit Beispielen beschreiben. --> Auf Sephir bezogen

Strukturierungsmerkmal	Daten und deren Verwendung an Beispielen erklären
Ordnungsdaten	<ul style="list-style-type: none"> - Namen - Namen von Betrieb - Klassennamen - Geburtsdatum
Rechendaten	<ul style="list-style-type: none"> - mit oder ohne BM - Durchschnittsnote eines Moduls berechnen - Alter aus Geburtsdatum berechnen
Stammdaten	<ul style="list-style-type: none"> - Angaben wie E-Mail, Telefonnummer, Betrieb usw. - das Geburtsdatum ist nicht veränderbar - Die Fachrichtung (Applikationsentwicklung)
Bewegungsdaten	<ul style="list-style-type: none"> - Notenstand --> ändern sich öfters - Absenzen - Termine der ÜKs
Eingabedaten	<ul style="list-style-type: none"> - Bewertung des ÜKs - Arbeitsjournal - Profildaten (Passwort, E-Mail)
Ausgabedaten	<ul style="list-style-type: none"> - Inhalt einer E-Mail - Notenliste - Absenzenliste

3 Daten aufbereiten und auswerten (Statistik)

3.1 Mittelwert

Sie sehen eine Gruppe von zehn Lernenden zusammenstehen und fragen jeden einzelnen nach seinem Alter und notieren die Antworten auf. Ihre Liste sieht nach der Befragung wie folgt aus:

17, 16, 15, 19, 20, 17, 17, 16, 18, 17.

Aufgabe: Mittelwert berechnen.

- Berechnen Sie von Hand den Mittelwert der oben aufgeführten Liste!
- Kontrollieren Sie die Berechnung mit Excel!

The image shows a handwritten calculation and an Excel spreadsheet side-by-side. On the left, a handwritten equation $17+16+15+19+20+17+17+16+18+17 : 10 = \underline{\underline{17.2}}$ is framed in red. To its right is the formula $x = \frac{172}{10} = 17.2$. On the right, a screenshot of an Excel spreadsheet displays a list of ages from 1 to 10 in column A, and the sum 172 and mean 17.2 in row 11. The cells containing 172 and 17.2 are highlighted in yellow and blue respectively.

	A	B	C
1	17		
2	16		
3	15		
4	19		
5	20		
6	17		
7	17		
8	16		
9	18		
10	17		
11	172		<u><u>17.2</u></u>

Die Formel für die Berechnung eines Mittelwertes \bar{x} lautet, wenn n die Anzahl Werte und $x_1, x_2, x_3, x_4, \dots, x_n$ die Werte sind:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

3.2 Mittelwert mit Klassen

Häufig sind die einzelnen Werte nicht bekannt, sondern nur die Anzahl von Werten innerhalb einer Gruppe (Klasse). Zum Beispiel kennen wir von einer Prüfung nur die Anzahl der Lernenden welche die Note 6, 5.5, 5, 4.5 usw. erreicht haben. Um den Mittelwert von Klassenwerten zu berechnen verwenden wir die folgende Formel:

$$\bar{x} = \frac{k_1 \cdot n_1 + k_2 \cdot n_2 + k_3 \cdot n_3 + \dots + k_j \cdot n_j}{n} = \frac{1}{n} \cdot \sum_{i=1}^j k_i \cdot n_i$$

$$n = n_1 + n_2 + n_3 + \dots + n_j = \sum_{i=1}^j n_i$$

mit

j = Anzahl Klasse

k_i = Mitte der Klasse i

n_i = Anzahl Werte in der Klasse i

n = Anzahl Werte aller Klassen

\bar{x} = Mittelwert

Aufgabe: Bei einer Prüfung wurden die folgenden Resultate erzielt:

Note	3.5	4.0	4.5	5.0	5.5	6.0
Anzahl Lernende	3	2	5	6	2	1

a) Berechnen Sie von Hand den Klassenschnitt (Mittelwert) der oben aufgeführten Resultate!

$$n = 3+2+5+6+2+1 = 19$$

$$\bar{x} = \frac{3.5 \cdot 3 + 4 \cdot 2 + 4.5 \cdot 5 + 5 \cdot 6 + 5.5 \cdot 2 + 6}{19} = \frac{10.5 + 8 + 22.5 + 30 + 11 + 6}{19}$$

$$\bar{x} = \frac{88}{19} = 4.63$$

b) Bei der Berechnung mit Klassen findet man oft die Begriffe **absolute** und **relative Häufigkeit**. Beim obigen Beispiel sind die absoluten Häufigkeiten die Anzahl Lernende. Die relative Häufigkeit ist der prozentuale Anteil Lernende pro Note bezogen auf das Total Lernende.

→ Total hat es 19 Lernende, entspricht 100%. 3 Lernende haben eine 3.5, entspricht 15.8% usw.

$$3/19 * 100 = 15.8\%$$

Ergänzen Sie die Tabelle mit den fehlenden relativen Häufigkeiten:

$$2/19 * 100 = 10.5\%$$

Note	3.5	4.0	4.5	5.0	5.5	6.0
Absolute Häufigkeit (Anzahl Lernende)	3	2	5	6	2	1
Relative Häufigkeit	15.8%	10.5 %	26.3%	31.6%	10.5%	5.3%

3.3 Median

Der Median x_{Med} ist derjenige Wert, der in der Mitte steht, wenn alle Werte $x_1, x_2, x_3, \dots, x_n$ der Grösse nach geordnet sind.

Aufgabe: Bestimmen Sie aus der obigen Altersliste (17, 16, 15, 19, 20, 17, 17, 16, 18, 17 Jahre) den Median und kontrollieren Sie das Ergebnis mit Excel!

15, 16, 16, 17, 17, 17, 17, 18, 19, 20 Jahre

$$\text{Median} = (17 + 17) / 2 = 17 \text{ Jahre}$$

Wenn zum Berechnen des Medians die Anzahl Werte gerade ist, so muss der Median aus dem Durchschnitt der beiden mittleren Werte berechnet werden.

Aufgabe: Die Verbrauchswerte von verschiedenen Autos in Litern pro 100 km lauten: 4.7, 5.6, 10.6, 5.9, 8.6, 4.2, 6.7 und 9.8 l/100 km.

- a) Berechnen Sie von Hand den Mittelwert der Verbrauchswerte!
 - b) Berechnen Sie von Median dieser Liste!
 - c) Kontrollieren Sie Ihre Berechnung mit Excel!

a) $n = 8$

$$x = \frac{4.7 + 5.6 + 10.6 + 5.9 + 8.6 + 4.2 + 6.7 + 9.8}{8} = \frac{56.1}{8} = 7.0125 \text{ l/100km}$$

b) $4.2, 4.7, 5.6, 5.9, 6.7, 8.6, 9.8, 10.6 = (5.9+6.7):2 = 6.3 \text{ l/100km}$

c)

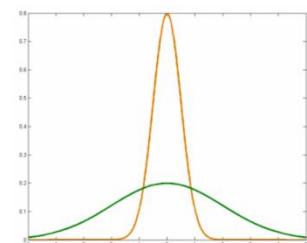
4.7	4.7
5.6	5.6
10.6	10.6
5.9	5.9
8.6	8.6
4.2	4.2
6.7	6.7
9.8	9.8
7.0125	6.3

3.4 Standardabweichung

Der englische Mathematiker Sir Francis Galton hat um 1860 den Begriff der Standardabweichung eingeführt: Die Standardabweichung ist ein Mass wieviel die untersuchten Werte $x_1, x_2, x_3, x_4, \dots, x_n$ um ihren Mittelwert streuen (abweichen). Der Wert der Standardabweichung wird umso grösser, je mehr die untersuchten Werte vom Mittelwert abweichen.

Berechnet wird die Standardabweichung nachfolgender Formel:

$$s = \sqrt{s^2} \text{ wobei } s^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$$



Dem Wert s^2 sagt man Varianz!

Aufgabe: Sie haben eine Liste mit den Körpergrössen von fünf verschiedenen Personen: 186 cm, 172 cm, 176 cm, 165 cm und 190 cm.

- Berechnen Sie die Standardabweichung s aus den obigen Werten!
- Kontrollieren Sie ihre Berechnung mit Excel (Nehmen Sie dazu die Excelformel STABWN)!

$\overline{x} = \frac{186\text{cm} + 172\text{cm} + 176\text{cm} + 165\text{cm} + 190\text{cm}}{5} = \frac{889\text{cm}}{5} = 177.8\text{cm}$ Varianz $s^2 = 1$ $5 * [(186\text{cm} - 177.8\text{cm})^2 + (172\text{cm} - 177.8\text{cm})^2 + (176\text{cm} - 177.8\text{cm})^2 + (165\text{cm} - 177.8\text{cm})^2 + (190\text{cm} - 177.8\text{cm})^2]$ (Zwischenresultat) $S^2 = 1$ $5 * 416.80 \text{ cm}^2 = 83.36 \text{ cm}^2$ $S = 9.13 \text{ cm}$ Worddokument!!!
--

3.5 Zusätzliche Aufgaben

Lösen Sie die folgenden Aufgaben auf einem separaten Blatt oder drucken Sie die mit Excel erstellte Berechnung aus!

$$(1754/8688)*100 = 20.2\%$$

Aufgabe 1 (von Hand)

$$(5578/8688)*100 = 64.2\%$$

Die Einwohner der Stadt Sursee hatten im Jahr 2008 die folgende Altersstruktur:

Altersgruppe in Jahren	0 - 19	20 - 64	65 - 79	80 - 100	
Anzahl Einwohner	1'754	20.2%	5'578	64.2%	1'008

11.6% 348 4% = 100%

Achtung: Um den Mittelwert von Klassen mit einem Größenbereich berechnen zu können, muss zuerst aus dem Bereich die Klassenmitte bestimmt werden.

- Berechnen Sie die Anzahl Einwohner **n = 8'688 Einwohner**
- Berechnen Sie den Altersdurchschnitt der Einwohner von Sursee **40.8 Jahre**
- Ergänzen Sie die obige Tabelle mit den relativen Häufigkeiten jeder Altersgruppe

$$((0+19)/2*1754+(20+64)/2*5578+(65+79)/2*1008+(80+100)/2*348)/8688$$

--> in Worddokument gelöst

Aufgabe 2 (Excel)

Eine Privatperson will sein Auto zum Verkauf im Internet ausschreiben. Dazu untersucht er zuerst die aktuell angebotenen Preise für seinen Fahrzeugtyp. Er erstellt sich dazu die folgende Liste mit den gefundenen Preisen in Franken:



7250, 7900, 6900, 8950, 9000, 8900, 8900, 7900
8800, 6800, 7900, 9950, 9200, 9900, 8300, 7900
9990, 9900, 8800, 9900, 9999, 7900, 7900, 8500
8900, 5990, 8600, 5500, 8000, 6900, 8900, 8900
8300, 5900, 5800, 8900, 9800, 9700, 7600, 5900
8550, 6900, 7900, 8500, 5800, 7900, 6450, 8900

Berechnen Sie aus diesen Preisen mit Excel:

- den Mittelwert
- den Median
- die Standardabweichung

Tipp: Damit Sie diese Werte nicht abtippen müssen, gehen Sie wie folgt vor:

- Erstellen Sie mit dem Editor eine Datei mit der Endung csv → Preise.csv
- Kopieren Sie die Werte von der Aufgabenstellung in diese Datei
- Öffnen Sie Excel und erstellen Sie eine leere Arbeitsmappe
- Wählen Sie unter dem Register „Daten“, „externe Daten abrufen“ und „aus Text“.
- Wählen Sie die zuvor erstellte csv-Datei aus.
- In der ersten Maske wählen Sie nun „getrennt“ aus und klicken auf „Weiter“
- In der zweiten Maske markieren Sie „Komma“ und klicken auf „Fertigstellen“
- Führen Sie die Berechnungen durch.

Aufgabe 3 (von Hand)

In einer mechanischen Werkstatt werden Spezialbohrer hergestellt. Nach jedem hundertsten Stück wird eines zu Kontrollmessungen entnommen. Dabei wurden die folgenden Masse ermittelt:



Durchmesser in mm	8.001	7.998	8.000	8.002	8.000	8.003	8.001	7.999	7.998	8.003
Länge in mm	120.1	120.0	120.4	119.8	120.1	119.8	120.0	119.9	120.2	119.7

a) Berechnen Sie den Mittelwert für die gemessenen Werte! **8.0005mm**

b) Berechnen Sie die Standardabweichung für die gemessenen Werte!

120.0 mm

Aufgabe 4 (Excel)



Laut LUSTAT Statistik Luzern wurden in den Jahren 2000 bis und mit 2009 in der Stadt Luzern die folgenden Niederschlagsmengen in mm gemessen:

	Jan	Feb	Mrz	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez
2000	20	84	75	44	92	116	245	184	117	80	72	23
2001	88	30	172	170	92	241	141	188	168	77	76	33
2002	9	98	48	78	150	235	190	162	163	91	175	63
2003	48	31	63	72	127	147	124	61	50	132	70	33
2004	96	39	90	104	130	181	120	109	78	79	23	35
2005	42	42	47	144	110	91	163	293	79	70	33	29
2006	12	37	137	189	199	76	61	223	122	43	38	64
2007	43	56	94	18	183	188	222	218	79	41	46	76
2008	44	47	62	131	18	127	206	185	176	102	31	51
2009	31	53	69	25	69	194	174	121	39	75	113	105

Berechnen Sie

- a) Die totale Niederschlagsmenge pro Jahr!
- b) Die durchschnittliche Niederschlagsmenge pro Jahr inkl. Standardabweichung!
- c) Die durchschnittliche Niederschlagsmenge pro Monat über die Jahre 2000..2009!
- d) Die durchschnittliche Niederschlagsmenge über alles inkl. Standardabweichung!

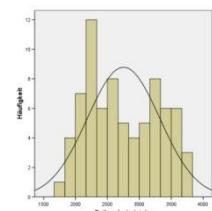
$$(562'948/4008351)*100 = 14\%$$

Aufgabe 5 (von Hand)

In der Schweiz gibt es im Jahr 2009 laut Zählung des Bundesamtes für Statistik 4'008'351 Wohnungen. Die folgende Tabelle zeigt die Wohnungen klassifiziert nach den Anzahl Zimmern. $4'008'351 * 0.264 = 1'058'205$

Anzahl Zimmer	1	2	3	4	5	6 und mehr	Total
Anzahl Wohnungen	241'690	562'948	1'058'205	1'086'263	618'664	440'165	4'008'351
Rel. Häufigkeit	6.0%	14%	26.4%	27.1%	15.5%	11%	100.0%

- a) Ergänzen Sie in der obigen Tabelle die fehlenden Werte!
- b) Berechnen Sie die mittlere Wohnungsgröße! Verwenden Sie zur Berechnung des Mittelwertes die Klasse „6 und mehr Zimmer“ als 6 Zimmer! $\frac{1'058'205}{440'165} = 3.65$ Zimmer
- c) Zeichnen Sie im Excel das dazugehörige Histogramm (Säulendiagramm) mit einer Trendlinie!



Aufgabe 6 (Excel)

Im letzten Jahr wurden in einem Museum die folgenden Besucherzahlen ermittelt:

Jan	Feb	Mrz	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Dez
1450	1201	1950	2134	2082	971	1892	2011	1509	1412	1934	2534

- a) Berechnen Sie die durchschnittliche Besucherzahl
- b) Berechnen Sie die minimale und maximale Besucherzahl
- c) Zeichnen Sie eine Säulengrafik mit allen Informationen inkl. Trendlinie

Aufgabe 7 (von Hand, Excel)

Bei einer Lehrabschlussprüfung haben die Kandidaten die folgenden Gesamtnoten erreicht.

Noten Bereich	> 1.0 und < 3.0	>= 3.0 und < 4.0	>= 4.0 und < 4.5	>= 4.5 und < 5.0	>= 5.0 und < 5.5	=>5.5 und <= 6
Anzahl Kandidaten	5	12	34	41	24	6

- a) Berechnen Sie den Durchschnittsnotenwert
 - b) ~~Zeichnen Sie eine Säulengrafik mit Trendlinie.~~
- $((1+3)/2*5+(3+4)/2*12+(4+4.5)/2*34+(4.5+5)/2*41+(5+5.5)/2*24+(5.5+6)/2*6)/122=4.52$

Aufgabe 8 (Excel)

Im März 2009 wurden am Flughafen in Houston, Texas, USA die folgenden Temperaturen gemessen:

Datum	Temperatur
01.03.2009	16° C
02.03.2009	19° C
03.03.2009	19° C
04.03.2009	24° C
05.03.2009	28° C
06.03.2009	28° C
07.03.2009	27° C
08.03.2009	28° C
09.03.2009	28° C
10.03.2009	29° C
11.03.2009	28° C
12.03.2009	10° C
13.03.2009	08° C
14.03.2009	09° C
15.03.2009	12° C
16.03.2009	24° C
17.03.2009	22° C
18.03.2009	26° C
19.03.2009	26° C
20.03.2009	27° C

- a) Berechnen Sie die durchschnittliche Temperatur
- b) Berechnen Sie die minimale und maximale Temperatur
- c) Berechnen Sie die Standardabweichung
- d) Zeichnen Sie eine Liniengrafik des Temperaturverlaufs für diesen Monat auf.

Aufgabe 9 (Excel)

Die folgende Tabelle zeigt die Anzahl Verkehrsunfälle im Kanton Luzern:

Jahr	nur Sachschaden	Todesfälle	Verletzte
1992	2'424	28	1'077
1993	2'419	27	1'031
1994	2'497	25	1'032
1995	2'449	31	1'048
1996	2'133	28	1'006
1997	2'261	23	1'119
1998	2'289	27	1'057
1999	2'407	25	1'140
2000	2'195	22	1'165
2001	2'346	18	1'294
2002	1'893	15	1'225
2003	1'751	20	1'279
2004	1'705	9	1'175
2005	1'676	12	1'198
2006	1'613	18	1'100
2007	1'598	15	1'090
2008	1'640	14	1'121
2009	1'669	22	1'133
2010	1'562	17	1'078
2011	1'392	12	1'095
2012	1'390	11	1'020

Kopieren Sie die Tabelle in ein Excel-Arbeitsblatt und führen Sie die folgenden Aufgaben aus:

- Fügen Sie am Ende der Tabelle eine weitere Spalte «Total» ein, mit einer Formel, welche die totale Anzahl Verkehrsunfälle pro Jahr berechnet
- Berechnen Sie für die vier Wert-Spalten jeweils die Durchschnittswerte, das Maximum und das Minimum. Fügen Sie dazu je eine neue Zeile ein.
- Erstellen Sie die folgenden Diagramme.

