

Daten, Daten, Daten – was sie uns verraten

Die Leitidee „Daten und Zufall“ in der AHS Unter- und Oberstufe

STEFAN GÖTZ, EVELYN SÜSS-STEPANCIK, WIEN

Der Bordcomputer eines Autos liefert Daten u. a. über den mittleren Treibstoffverbrauch (ohne dass man selbst rechnen muss!). Diese werden im Allgemeinen aber nicht gesammelt und systematisch ausgewertet. – Was könnte aus einer solchen Analyse (z. B. mithilfe von Excel) erkannt werden? Es eröffnet sich eine Fülle von Möglichkeiten, „Datendetektiv“ zu spielen und praxisnahe Vermutungen (Erfahrungen) empirisch zu untermauern. Im Beitrag soll auf diese Weise ein moderner Zugang zur Stochastik im Mathematikunterricht präsentiert werden, der auf (realen) Daten basiert und der immer wieder bis zur zwölften Schulstufe im Sinne des Spiralprinzips aufgenommen werden kann. So wird der Forderung nach Anwendungsorientierung im Mathematikunterricht Rechnung getragen, ebenso wird an Erfahrungen der Schülerinnen und Schüler aus ihrer Umwelt angeknüpft bzw. werden Bezüge zur ihrer Lebenswelt hergestellt.

1. Einleitung

Der datenorientierte Zugang im Stochastikunterricht erfordert Überlegungen in dreierlei Hinsicht: statistische, technische und didaktische. Wir gehen dabei von der Hypothese aus, dass das Arbeiten mit *realen Daten* den Stochastikunterricht bereichert, weil auf diese Weise an die Alltagserfahrungen der Schülerinnen und Schüler angeknüpft werden kann: vgl. die erste Winter'sche Grunderfahrung (Winter 1995). Der Lehrplan Mathematik für die AHS-Unterstufe spricht ebenfalls von Grunderfahrungen:

„Der Mathematikunterricht soll folgende miteinander vielfältig verknüpfte Grunderfahrungen ermöglichen:

- Erscheinungen der Welt um uns in fachbezogener Art wahrzunehmen und zu verstehen [...].“ (Lehrplan 2000, S. 1)

Eine ähnliche Forderung findet sich im Lehrplan der AHS-Oberstufe („Erkenntnistheoretischer Aspekt“ in Lehrplan 2004, S. 1).

Der statistische Zugang verlangt grob gesprochen drei Schritte: die Erhebung der Daten, ihre Analyse und die kontextbezogene Interpretation der erhaltenen Ergebnisse. Auf den ersten Punkt werden wir bei der datenorientierten Vorgangweise besonderes Augenmerk legen. Es gibt verschiedene Möglichkeiten reale Daten für den Unterricht zu gewinnen.

Für die technische Komponente wollen wir ebenfalls Hinweise geben. Die Daten müssen verarbeitet, ausgewertet und dargestellt werden. Dafür gibt es kein „bestes“ Softwarepaket (für die Schule), es wird sich zeigen, dass wir je nach Aufgabenstellung bzw. Datenlage unterschiedliche Programme verwenden werden.

Bei der zugehörigen Vortragsvorbereitung sind wir immer wieder auf (mögliche) *Grundkompetenzen* zu verschiedenen Handlungsbereichen gestoßen, die hier verlangt werden, wie das die Anwendungsorientierung mit sich bringt.

Wir führen im Folgenden das Konzept des *Datendetektivs* bzw. der *Datendetektivin* ein, welches den Daten auf den Zahn fühlt und dabei Besonderheiten aufspürt.

2. Woher kommen die Daten?

Natürlich können die Daten selbst erhoben werden, dabei verdienen Erhebungen, die Schülerinnen und Schüler ohne Unterstützung durchführen können, besondere Beachtung. Beispielsweise können Packungen von Schokolinsen bezüglich ihrer Farbverteilung von Schülerinnen und Schülern ausgewertet werden (vgl. Abschnitt 7). Aber auch Körpergrößen oder origineller Schuhgrößen von Schülerinnen und Schülern stellen leicht zugängliche Datensätze dar (vgl. Götz & Süss-Stepancik 2012).

Eine weitere attraktive Möglichkeit besteht darin, Originaldaten aus dem Internet zu requirieren. Wir werden im Folgenden auf Messwerte einer Messstation auf Hawaii, die den CO₂-Gehalt in der Atmosphäre angeben, zurückgreifen (Abschnitt 5). Hierbei ist wie immer bei Verwendung von Informationen aus dem Internet auf die Seriosität der Quelle zu achten. In technischer Hinsicht müssen die Daten oftmals erst für die Verarbeitung mit in der Schule verwendeten Programmen aufbereitet werden. Möchte man Daten über einen größeren Zeitraum entweder miteinander vergleichen oder in ihrer zeitlichen Entwicklung darstellen, ist diese Möglichkeit der Datengewinnung im Allgemeinen wohl die einzig machbare.

Es soll an dieser Stelle nicht unerwähnt bleiben, dass es im Internet auch spezielle Lernangebote im Bereich Statistik gibt, die mit schon aufbereiteten Daten arbeiten. In wie weit Letztere ihre Authentizität durch diesen Vorgang bewahren, sei dahingestellt.

3. Benzin – die Erste

Die Entwicklung des Benzinpreises ist zurzeit in aller Munde:

„Unterschreiben gegen den Spritpreis-Wahnsinn



Abb. 1: www.heute.at vom 02. April 2012

Ein Blick auf die aktuellen Spritpreise lässt keine wahre Osterfreude aufkommen - Superbenzin und Diesel befinden sich auf Rekordhoch. Unterschreiben Sie gegen die Abzocke an den Zapfsäulen und gewinnen Sie einen vollen Autotank.“

Eine *erste – offene – Aufgabe* dazu könnte so lauten: „Recherchiere wie sich der Benzinpreis in den letzten 10 Jahren entwickelt hat und stelle deine Ergebnisse (grafisch) dar! Was fällt dir auf?“ Hier wird die Grundkompetenz WS 1.2¹ angesprochen.

Eine erste Recherche zur „Benzinpreisentwicklung Österreich“ mit **Google** liefert circa 192000 Ergebnisse. Eine a priori Durchsicht der Lehrenden könnte diesen Tipp zutage bringen: Unter dem Stichwort „*Treibstoffpreismonitor*“ finden wir beim Bundesministerium für Wirtschaft, Familie und Jugend (<http://www.bmwfj.gv.at>) brauchbare Daten. Abbildung 2a zeigt einen Ausschnitt davon. Mit Copy & Paste gelingt ein Transfer der Daten in Excel (Abbildung 2b) problemlos. Jetzt können die Daten analysiert werden. Wir stellen sie dazu als *Zeitreihe* von Oktober 2002 bis Oktober 2011 dar: Abbildung 3a.

¹ Die Bezeichnung der Grundkompetenzen entnehmen wir „Die standardisierte schriftliche Reifeprüfung in Mathematik“ (2012).

Nun kommt die Datendetektivin ins Spiel: Auffällig ist der Preissturz Ende 2008 / Anfang 2009 (Lupe). Was ist damals passiert (Grundkompetenz WS 1.1)? Eine weitere Internetrecherche beschert uns den in Abbildung 3b sichtbaren Ausschnitt aus WELT ONLINE (Lupe).

bmwfi
Bundesministerium für
Wirtschaft, Familie und Jugend

Stichtag	Diesel	Eurosuper	Normal	Super Plus
26.03.2012	€ 1.446	€ 1.491	€ 1.488	€ 1.633
19.03.2012	€ 1.442	€ 1.484	€ 1.482	€ 1.619
12.03.2012	€ 1.438	€ 1.472	€ 1.468	€ 1.599
05.03.2012	€ 1.424	€ 1.451	€ 1.449	€ 1.561
27.02.2012	€ 1.421	€ 1.436	€ 1.434	€ 1.571
20.02.2012	€ 1.422	€ 1.431	€ 1.430	€ 1.565

Abb. 2a: Die Preise verstehen sich in € je Liter.

	A	B	C	D	E
1	Stichtag	Diesel	Eurosuper	Normal	Super Plus
2	26.03.2012	1,446	1,491	1,488	1,633
3	19.03.2012	1,442	1,484	1,482	1,619
4	12.03.2012	1,438	1,472	1,468	1,599
5	05.03.2012	1,424	1,451	1,449	1,561
6	27.02.2012	1,421	1,436	1,434	1,571
7	20.02.2012	1,422	1,431	1,430	1,565

Abb. 2b: Die Preise verstehen sich in € je Liter.

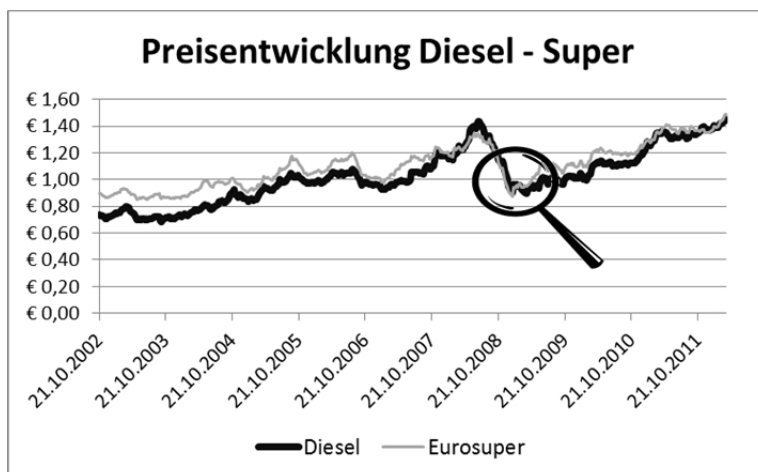


Abb. 3a

Benzinpreis fällt auf tiefsten Stand seit 31 Monaten

In der Folge der Finanzkrise ist der Ölpreis dramatisch gesunken. Das schlägt nun auch voll auf die Spritpreise durch. Autofahrer können sich über den niedrigsten Benzinpreis seit Anfang 2006 freuen. Der Preisabstand zum Diesel ist mit 2 Cent kleiner als gewöhnlich. Verantwortlich dafür könnte die hohe Nachfrage nach Heizöl sein.

WELT ONLINE

Abb. 3b: vom 20.10.2008

Abbildung 3b entnehmen wir, „... niedrigsten Benzinpreis seit Anfang 2006“. Wenn wir die Benzinpreise aus den beiden in Rede stehenden Zeiträumen vergleichen wollen, bieten sich *Boxplots* (4. bzw. 6. Klasse, Grundkompetenz WS 1.2) als adäquate Darstellungsmittel an: Abbildung 4.

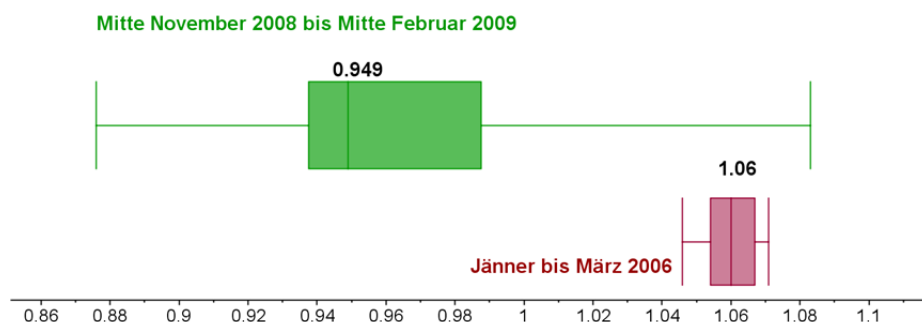


Abb. 4: mit GeoGebra erstellt

Im Sinne des Spiralprinzips kann dieser Vergleich in der 7. Klasse noch einmal aufgegriffen und mithilfe des *Vorzeichentests* quantifiziert werden. Dazu zeigt uns Tabelle 1 die relevanten Werte. Dabei fassen wir diese wöchentlichen Durchschnittswerte als Stichprobe im jeweiligen Zeitraum auf. Als Grundgesamtheit definieren wir die Benzinpreise in ganz Österreich.

Tab. 1: Preisentwicklung von 1 Liter Superbenzin			
Datum	Preis in €/Liter Super	Datum	Preis in €/Liter Super
27.03.2006	1,07	09.02.2009	0,961
20.03.2006	1,071	02.02.2009	0,941
13.03.2006	1,061	26.01.2009	0,956
06.03.2006	1,059	19.01.2009	0,936
27.02.2006	1,056	12.01.2009	0,939
20.02.2006	1,051	05.01.2009	0,876
13.02.2006	1,065	15.12.2008	0,909
06.02.2006	1,069	08.12.2008	0,942
30.01.2006	1,065	01.12.2008	0,976
23.01.2006	1,057	24.11.2008	0,999
16.01.2006	1,052	17.11.2008	1,047
09.01.2006	1,046	10.11.2008	1,083

Für 2008/2009 fällt auf, dass nur zwei von zwölf Werten über 1 € pro Liter liegen. Wir berechnen $\sum_{k=10}^{12} \binom{12}{k} \cdot 0,5^{12} < 0,02$ und schließen daraus, dass der Literpreis damals signifikant unter einem Euro gelegen ist (im Gegensatz zum ersten Quartal 2006). Die Grundgesamtheit in diesen beiden Zeiträumen scheint also nicht dieselbe zu sein.

4. Der fachdidaktische und bildungstheoretische Rahmen

2003 wurden in Deutschland die Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss von der KMK (2003) beschlossen. In diesem Dokument wird unter anderem die *Leitidee* „Daten und Zufall“ definiert. Dort finden wir die für unsere Intention des datenorientierten Zugangs relevanten Schüler/innentätigkeiten: „Die Schülerinnen und Schüler

- [...]
 - planen statistische Erhebungen,
 - [...]
 - beschreiben Zufallserscheinungen in alltäglichen Situationen,
 - bestimmen Wahrscheinlichkeiten bei Zufallsexperimenten.“ (S. 12)

In Österreich werden diese Tätigkeiten bei den *Bildungsstandards Mathematik für die achte Schulstufe M8* (2007) nicht direkt angesprochen, dennoch lassen sich Zusammenhänge finden:

- *Darstellen* meint die Übertragung gegebener mathematischer Sachverhalte in eine (andere) mathematische Repräsentation bzw. Repräsentationsform.
- *Modellbilden* erfordert über das Darstellen hinaus, in einem gegebenen Sachverhalt die relevanten mathematischen Beziehungen zu erkennen (um diese dann in mathematischer Form darzustellen), allenfalls Annahmen zu treffen, Vereinfachungen bzw. Idealisierungen vorzunehmen u. Ä.

Diese beiden Handlungsbereiche sind auch für unsere Zwecke höchst relevant, wie wir schon gesehen haben und noch werden. Beispielsweise kann aus der ursprünglichen Tabelle von Daten in Abschnitt 3 sowohl eine Zeitreihe als auch ein Boxplot generiert werden.

Im Sinne des *Spiralprinzips* (Abbildung 5) wollen wir auch zeigen bzw. haben wir schon gezeigt, dass der Inhaltsbereich „*Statistische Darstellungen und Kenngrößen*“ aus M8 in verschiedenen Jahrgangsstufen aufgegriffen werden kann. Selbiges gilt natürlich auch für die eben erwähnten Handlungsbereiche. Es ergibt sich so ein roter Faden von nicht unbedingt unmittelbar einsichtigen Zusammenhängen, die bei entsprechender Betonung zu einem langfristigen Lernprozesses motivieren können.

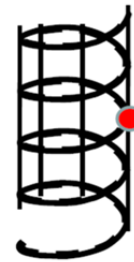


Abb. 5: Curriculumspirale aus Bruder 2012

5. CO₂ auf Hawaii (vgl. Eichler & Vogel 2009)

5.1 Eine erste (lineare) Annäherung

Auf der südlichsten und größten Insel von Hawaii befindet sich der Mauna Loa, der größte Vulkan auf Erden und immer noch aktiv². Auf ihm befindet sich ein Observatorium³, das unter anderem die Aufgabe hat, den CO₂-Gehalt der Atmosphäre zu messen. Der gewählte Standort ist dafür besonders geeignet:

„Die Lage der Messstation auf Hawaii ist für atmosphärische Messungen besonders geeignet, da die Luft in der Umgebung kaum durch lokale Einflüsse der Vegetation oder der Menschen beeinträchtigt wird.“ (http://de.wikipedia.org/wiki/Messstation_Mauna_Loa, Zugriff: 18.7.2012)

Dabei ergibt sich ein schon seit vielen Jahren typisches Bild des Entwicklungsverlaufes des CO₂-Gehalts in der Luft: Abbildung 6. Diese Kurve ist nach Charles David Keeling (1928 – 2005) benannt, ein Wissenschaftler der ebendort tätig gewesen ist. Deutlich (Achtung: Nulllinie nicht dargestellt!) erkennen wir einerseits den globalen Anstieg, andererseits eine periodische Schwankung des CO₂-Gehalts in der Atmosphäre.

Beide Phänomene wollen wir im Folgenden anhand der Originaldaten näher untersuchen. Diesmal ist es nicht so einfach, diese in ein weiterverwendbares Format (Excel) zu übertragen, sondern es musste ein Umweg über Word genommen werden, um den Stellenwert der einzelnen Daten zu erhalten: Abbildung 7a und 7b (für die Werte

des CO₂-Gehalts wurde die „interpolated“-Spalte verwendet).

Wir untersuchen die Entwicklung des CO₂-Gehalts in der Atmosphäre im Zeitabschnitt von 1995 bis 2004 (Abbildungen 8a bis c). Wenn wir linear interpolieren, ergibt sich eine Steigung von ungefähr 0,15 ppm pro Monat (für diesen Zeitraum, Abbildung 8a). Die periodischen Residuen (Abbildung 8b)

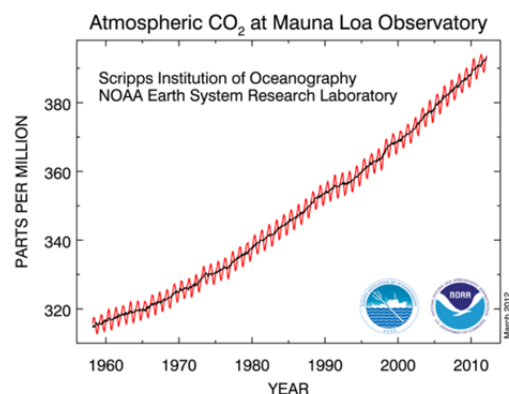


Abb. 6: *Keeling-Kurve*
<http://www.esrl.noaa.gov/gmd/obop/mlo/>
 (Zugriff: 18.7.2012)

² <http://www2.edc.org/sciencequest/2001/sq0110360/largest.html> (Zugriff: 18.7.2012)

³ <http://www.esrl.noaa.gov> (Zugriff: 18.7.2012)

verdienen eine eigene Betrachtung. Wir stellen mithilfe von GeoGebra eine Kreisfrequenz b von $0,52 \text{ Monat}^{-1}$ fest (Abbildung 8c, die Amplitude a und die Phasenverschiebung c können wir nicht ohne weiteres interpretieren).

Wenn wir eine Periodendauer von einem Jahr annehmen, ist die Kreisfrequenz $\omega = \frac{2\pi}{12} \approx 0,52$.

```
#
# NOTE: In general, the data presented for the last year are subject to change,
# depending on recalibration of the reference gas mixtures used, and other quality
# control procedures. Occasionally, earlier years may also be changed for the same
# reasons. Usually these changes are minor.
#
# CO2 expressed as a mole fraction in dry air, micromol/mol, abbreviated as ppm
#
# (-99.99 missing data; -1 no data for #daily means in month)
#
#          decimal      average      interpolated      trend      #days
#          date          (season corr)
#
1958  3  1958.208      315.71      315.71      314.61      -1
1958  4  1958.292      317.45      317.45      315.29      -1
1958  5  1958.375      317.50      317.50      314.71      -1
1958  6  1958.458      -99.99      317.10      314.85      -1
1958  7  1958.542      315.86      315.86      314.98      -1
1958  8  1958.625      314.93      314.93      315.94      -1
1958  9  1958.708      313.20      313.20      315.91      -1
1958 10  1958.792      -99.99      312.66      315.61      -1
1958 11  1958.875      313.33      313.33      315.31      -1
1958 12  1958.958      314.67      314.67      315.62      -1
1959  1  1959.042      315.62      315.62      315.70      -1
1959  2  1959.125      316.38      316.38      315.88      -1
1959  3  1959.208      316.71      316.71      315.61      -1
1959  4  1959.292      317.72      317.72      315.56      -1
1959  5  1959.375      318.29      318.29      315.50      -1
```

Abb. 7a: Die Originaldaten (Auszug)

Jahr	Monat	Monat	CO2 [ppm]
1958	März	3	315.71
1958	April	4	317.45
1958	Mai	5	317.5
1958	Juni	6	317.1
1958	Juli	7	315.86
1958	August	8	314.93
1958	September	9	313.2
1958	Oktober	10	312.66
1958	November	11	313.33
1958	Dezember	12	314.67
1959	Jänner	1	315.62
1959	Februar	2	316.38
1959	März	3	316.71

Abb. 7b: ... und was daraus wird.

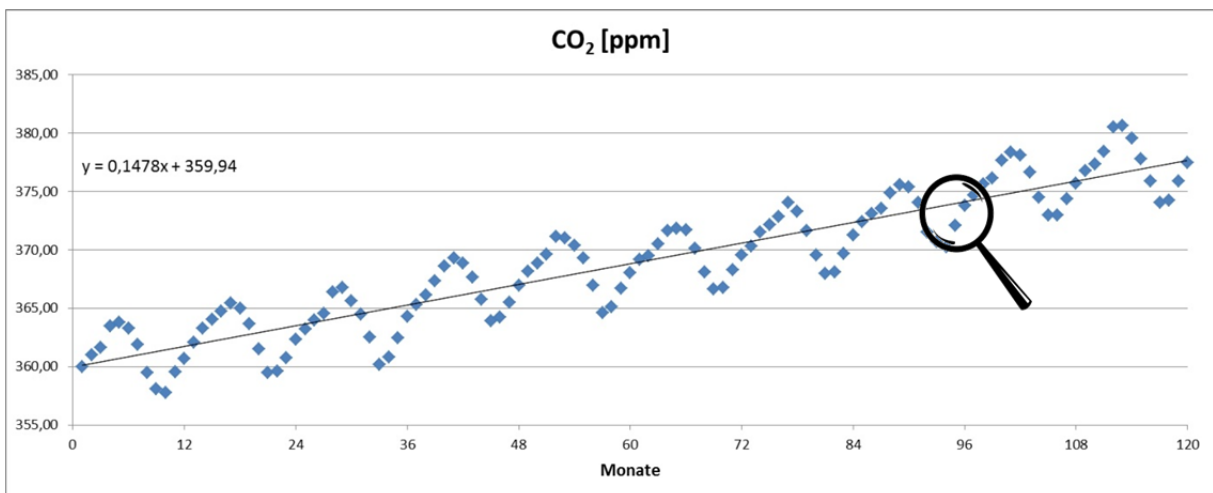


Abb. 8a: Lineare Interpolation des globalen Trends

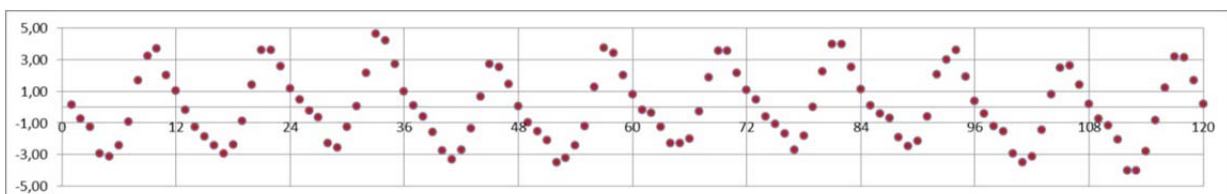
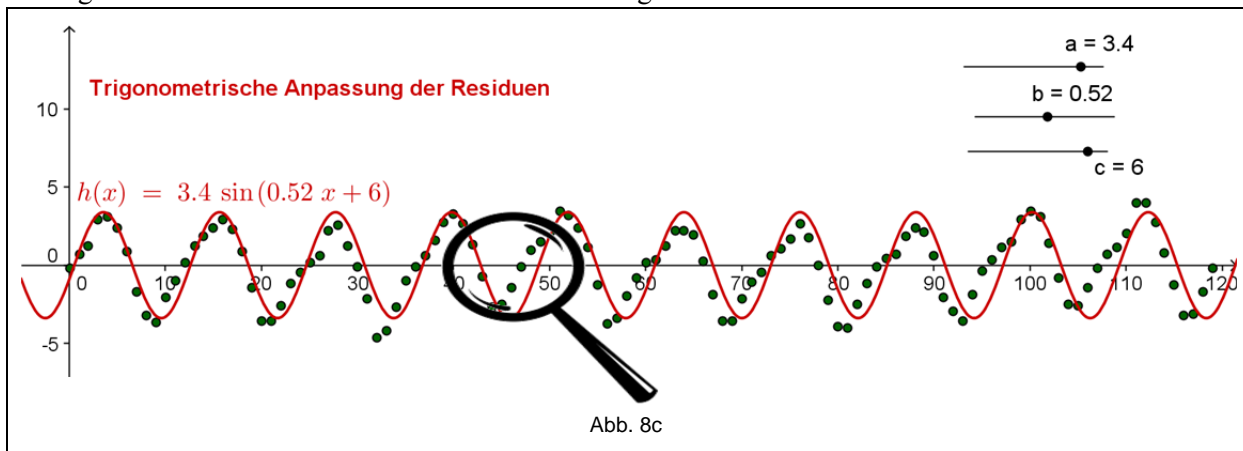


Abb. 8b: Daten = lineare Funktion + periodische Residuen

Woher kommen nun diese periodischen Schwankungen? – In der Zeitspanne vom Frühjahr bis zum Herbst (Sommer) *sinkt* der CO₂-Gehalt wegen der zunehmenden Belaubung, im Zeitraum Herbst-Frühjahr (Winter) dagegen *steigt* der CO₂-Gehalt aufgrund der Verrottung des organischen Materials. Da die Vegetation auf der Erde asymmetrisch verteilt ist, wirkt sich der Sommer auf der Südhalbkugel weniger stark aus als der Winter auf der Nordhalbkugel.



5.2 Der Schein kann trügen

Wenn wir uns den CO₂-Gehalt der Luft seit Aufzeichnungsbeginn ansehen, dann müssen wir erkennen, dass die lineare Modellierung nicht mehr passt (Abbildung 9).

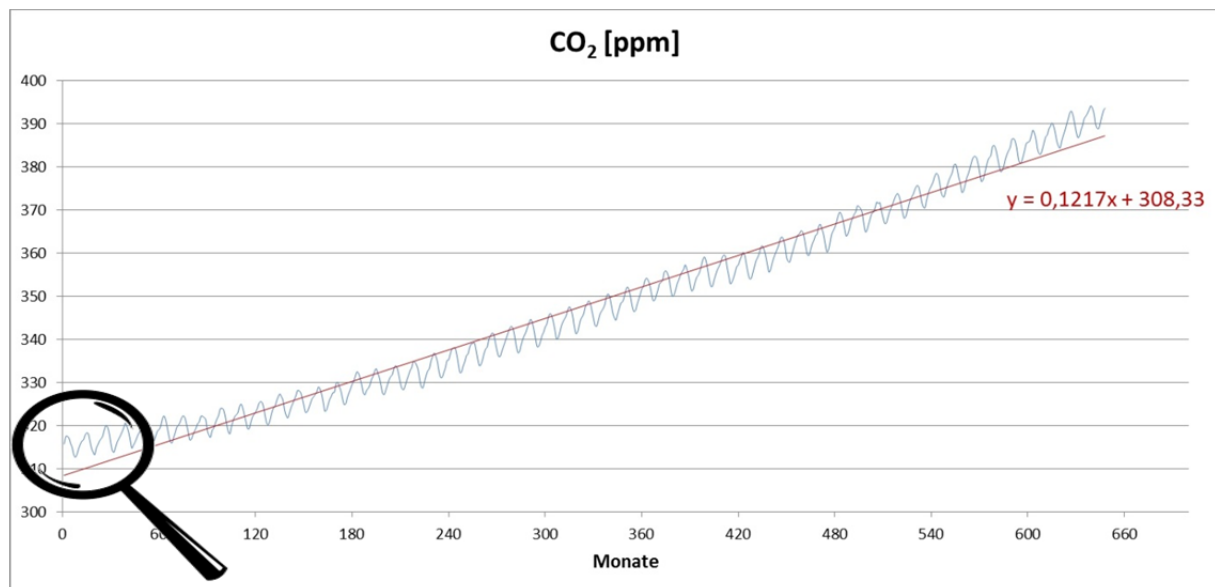


Abb. 9

Um dennoch einen langfristigen Trend modellieren zu können, greifen wir zur Methode der „Gleitenden Mittelwerte“ (vgl. Eichler & Vogel 2009, S. 115), um die Keeling-Kurve zu glätten. Dabei gehen wir von der einfachen Modellgleichung *Daten = Glatte Komponente + Saisonkomponente + Residuen* aus und berechnen $\bar{y}_t = \frac{0,5 \cdot y_{t-6} + y_{t-5} + \dots + y_{t+5} + 0,5 \cdot y_{t+6}}{12}$ für $t = 7$ bis 642. Wir setzen diesen Wert statt y_t für jedes Monat ein und erhalten Abbildung 10. Die Glättung ist also vollständig gelungen, was auch mittels Excel (Trendlinie – Gleitender Durchschnitt)

möglich wäre. Insgesamt haben wir nun eine Modellierung von innen aus den Daten heraus vorgenommen, während wir in 5.1 eine Modellierung von außen getätigt haben: Der Funktionstyp (linear) wurde a priori aufgesetzt.

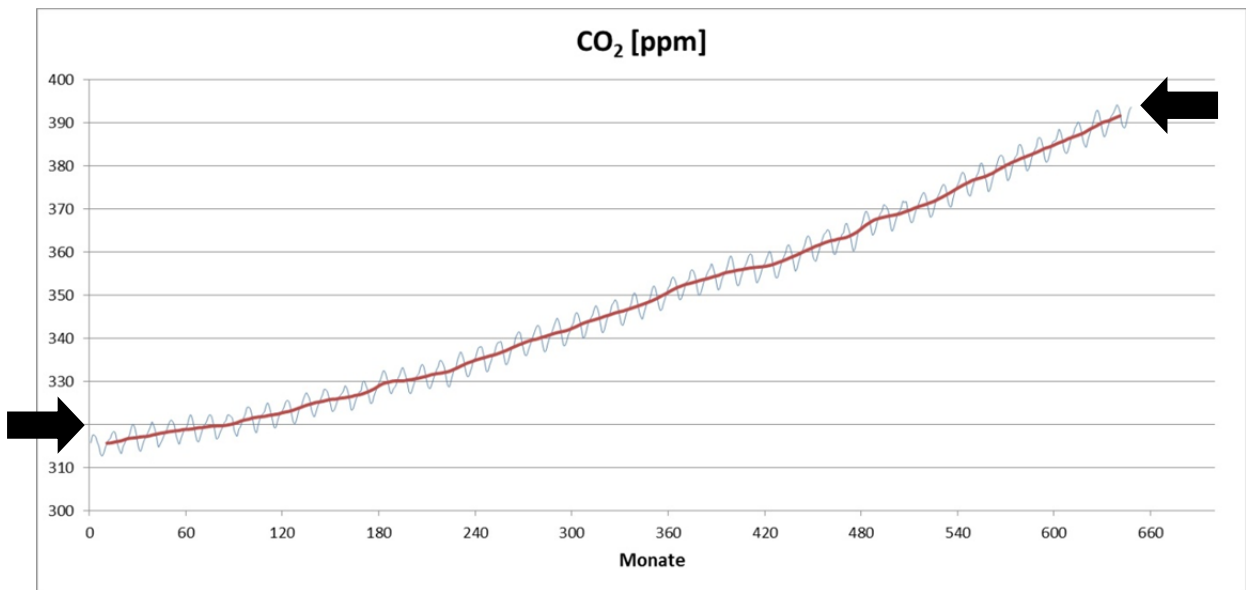


Abb. 10

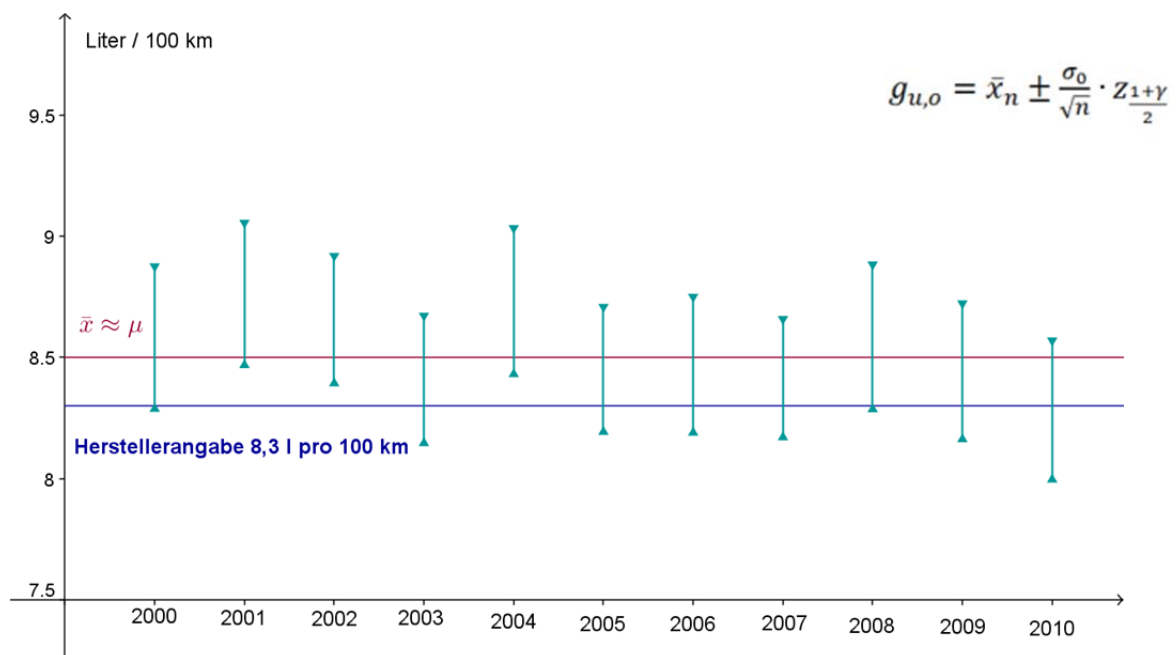


Abb. 11

6. Benzin – die Zweite

Seit vielen Jahren führt der Erstautor ein Fahrtenbuch für sein Auto der Marke Toyota Corolla G6R, Baujahr 1999, 4 Zylinder, 1,6 Liter Hubraum und 81 kW bzw. 110 PS. Dabei wird auch in akribischer Weise der durchschnittliche Benzinverbrauch pro 100 km festgehalten.

Eine Aufgabe dazu, siehe auch Grundkompetenz WS 4.1:

- (1) Berechne die Konfidenzintervalle des mittleren Benzinverbrauchs pro 100 km für die Jahre 2000 bis 2011! Nimm ein Konfidenzniveau von $\gamma = 0,95$ an!
Wie kann die Standardabweichung σ geschätzt werden?
Wovon hängt die Länge des Konfidenzintervalls ab?
- (2) Der Hersteller gibt einen Durchschnittsverbrauch von 8,3 Litern pro 100 km an. Geben die Daten dem Hersteller Recht? Teste einseitig!
Welche Gründe kann es geben, dass die Konfidenzintervalle die Herstellerangabe nicht immer überdecken?

Ein Ergebnis von (1) zeigt Abbildung 11. Dabei wurde σ aus der empirischen Standardabweichung von 267 Daten geschätzt: 0,67 Liter pro 100 km. Für $\gamma = 0,99$ ergeben sich entsprechend längere Konfidenzintervalle.

Für (2) formulieren wir folgende Hypothesen: $H_0: \mu \leq 8,3$ Liter pro 100 km und $H_1: \mu > 8,3$ Liter pro 100 km. Wir testen also rechtsseitig und erhalten $c = 8,37$ ($\alpha = 0,05$) und $c = 8,40$ ($\alpha = 0,01$) als kritische Werte (Abbildung 12). Die realen Daten liefern in ihrer Gesamtheit (1999 bis 2011) einen Mittelwert von $\bar{x} = 8,5$ Liter pro 100 km, daher kann die Nullhypothese (hoch) signifikant verworfen werden.

Eventuell könnte als Fortsetzung von (2) der Graph der Operationscharakteristik bzw. der Gütefunktion gezeichnet und interpretiert werden.

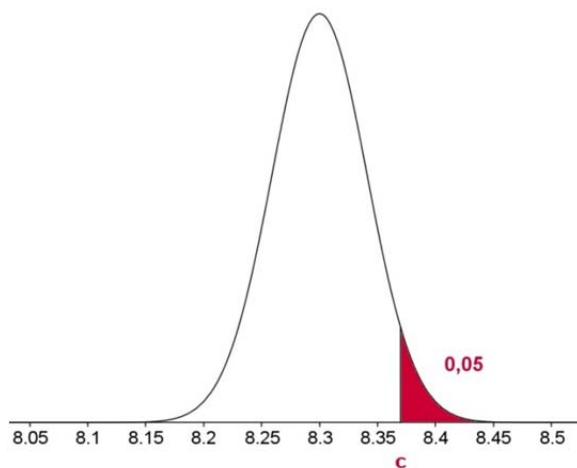


Abb. 12

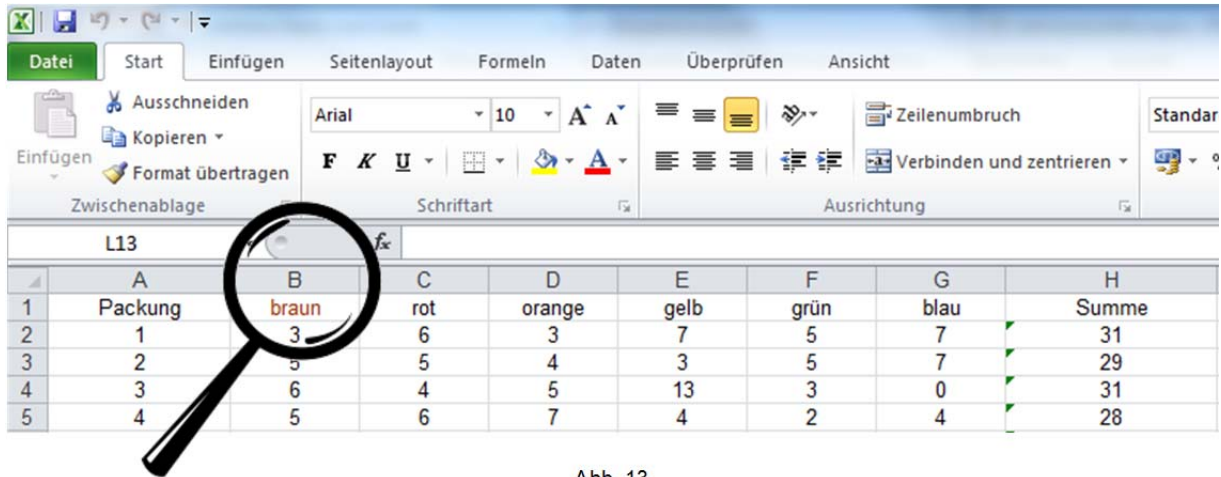
Abschließend wollen wir kurz im Sinne des Modellierens thematisieren, warum eigentlich die Normalverteilung beim Benzinverbrauch angenommen werden kann. Der Grund dafür ist, dass der Benzinverbrauch von vielen – jeweils nicht einseitig dominierenden – Faktoren abhängt:

- Fahrweise
- Jahreszeit
- Reifendruck
- Langstrecke – Stadtverkehr
- (Gegen-)Wind
- Fahrbahnbeschaffenheit
- Rollwiderstand der Reifen
- Steigung/Gefälle der Straße
- Sonstige Verbraucher (z. B. Klimaanlage, Tagfahrlicht) an/aus
- ...

Diese Einflüsse sind wohl auch dafür verantwortlich, dass sich die tatsächliche Grundgesamtheit von der Herstellerangabe unterscheidet und daher auch die Konfidenzintervalle für den mittleren Benzinverbrauch nicht immer den Wert 8,3 Liter pro 100 km überdecken. Der Test in Punkt (2) stützt diese Vermutung.

7. Aufbereitete Daten am Beispiel Schokolinsen

Eichler & Vogel (2009) stellen im Internet *Zusatzmaterialien* zur Verfügung. Wir sehen uns den Datensatz Schokolinsen als Excel-Datei (Abbildung 13) näher an. Insgesamt wurden 103 Packungen ausgezählt.



	A	B	C	D	E	F	G	H
1	Packung	braun	rot	orange	gelb	grün	blau	Summe
2	1	3	6	3	7	5	7	31
3	2	5	5	4	3	5	7	29
4	3	6	4	5	13	3	0	31
5	4	5	6	7	4	2	4	28

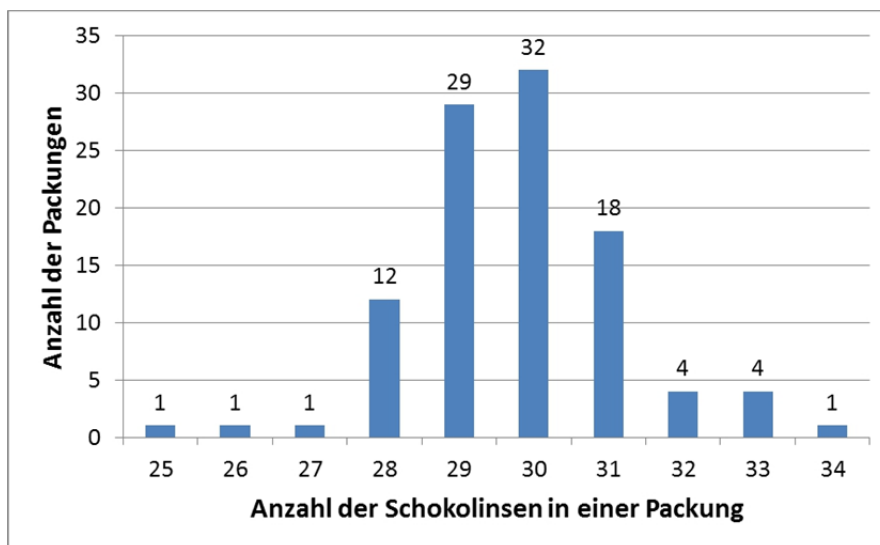
Abb. 13

7.1 Erzeugen von statistischen Darstellungen in der AHS-Unterstufe

Folgende *Aufgaben* könnten beispielsweise dazu gestellt werden (Grundkompetenzen WS 1.1 und WS 1.2):

- (1) Ist in jeder Packung gleich viel drin?
- (2) Suche zwei Packungen mit möglichst unterschiedlicher Farbverteilung und präsentiere sie jeweils mittels geeigneter Diagramme!
- (3) Stelle die Entwicklung der kumulativen Verteilung der relativen Häufigkeiten der Farbverteilungen aus den Einzelverteilungen grafisch dar!

Die Abbildungen 14 bis 16 zeigen mögliche Ergebnisse.



Mit Excel:

1. Häufigkeit:
gibt kumulative Werte
2. Differenzbildung
gibt gewünschte Daten

Abb. 14: Die ähnlichen Wertebereiche von x- und y-Achse erschweren das Lesen und Interpretieren dieses Diagramms.

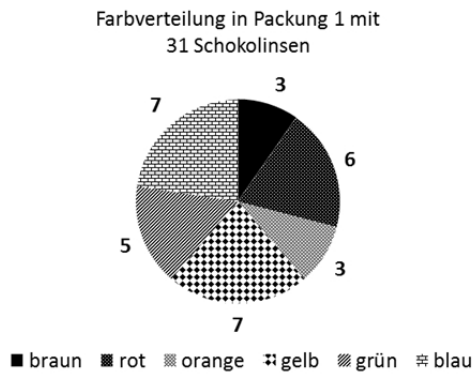


Abb. 15a: Kreisdiagramm für absolute Häufigkeiten

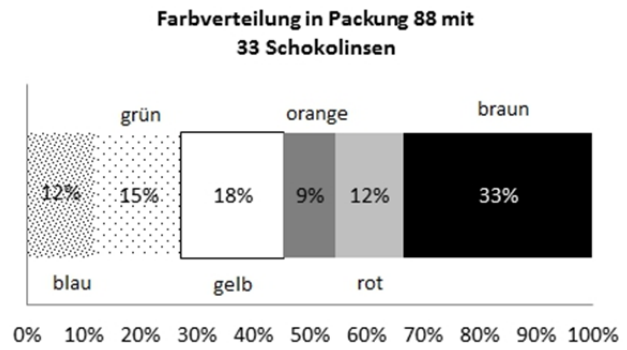


Abb. 15b: Prozentstreifen für relative Häufigkeiten

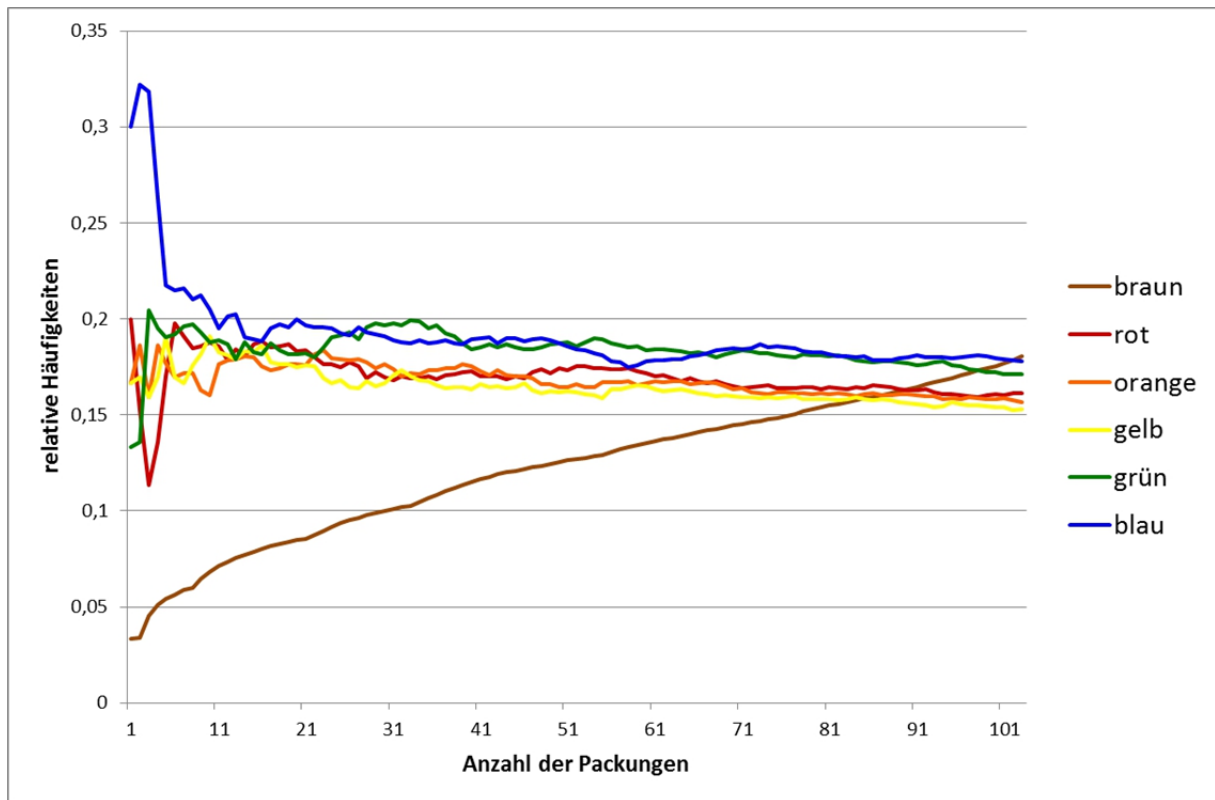


Abb. 16: Ein Beispiel (?) für das Empirische Gesetz der Großen Zahlen

7.2 Lernen aus Erfahrung: subjivistischer Wahrscheinlichkeitsbegriff (6. Klasse)

Tab. 2: Farbverteilung von zwei Schokolinsenpackungen

	blau	grün	gelb	orange	rot	braun
Packung 1	23%	16%	23%	10%	19%	10%
Packung 88	12%	15%	18%	9%	12%	33%

In Tabelle 2 sind die Farbverteilungen von zwei Schokolinsenpackungen (Nummer 1 und 88) festgehalten. Folgendes Szenario stellen wir uns nun vor:

Es wird aus *einer* der beiden Packungen mit Zurücklegen sukzessive je eine Schokolinse gezogen und ihre Farbe bekannt gegeben. Wie kann aufgrund der vorliegenden wachsenden Datenmenge immer sicherer eingeschätzt werden, aus welcher der beiden Packung gezogen wird?



1. A priori gehen wir von $P(88) = 0,5 = P(1)$ aus.
2. Angenommen, zuerst wird *Braun* gezogen:

$$P(88|Braun) = \frac{P(Braun|88) \cdot P(88)}{P(Braun|88) \cdot P(88) + P(Braun|1) \cdot P(1)} = \frac{0,33}{0,33+0,1} = 0,77 \text{ a posteriori.}$$

Damit ergibt sich eine neue a priori-Einschätzung: $P(88) = 0,77$ und $P(1) = 0,23$.

3. Angenommen, nun wird *Gelb* gezogen:

$$P(88|Gelb) = \frac{P(Gelb|88) \cdot P(88)}{P(Gelb|88) \cdot P(88) + P(Gelb|1) \cdot P(1)} = \frac{0,18 \cdot 0,77}{0,18 \cdot 0,77 + 0,23 \cdot 0,23} = 0,72 \text{ a posteriori.}$$

$P(88) = 0,72$ und $P(1) = 0,28$ ist dann die neue a priori-Einschätzung.

4. Angenommen, jetzt wird *Blau* gezogen: $P(88|Blau) = \dots = 0,57$, also $P(1) = 0,43$ usw.

Diese Aufgabe bietet mannigfache Variationsmöglichkeiten. Zum Beispiel könnte von einer anderen a priori-Verteilung ausgegangen werden und die daraus resultierenden a posteriori-Verteilungen könnten mit den ursprünglichen verglichen werden. Wenn sich die Daten nicht ändern, wird man sehen, dass sich diese mit wachsender Zahl der Ziehungen angleichen werden. Genauso können natürlich die (fiktiven) Daten verändert und der Einfluss auf die a posteriori-Einschätzung studiert werden.

Auf diese Weise kann auf sehr elementare Art der *subjektive Wahrscheinlichkeitsbegriff*, der ja im Lehrplan (Lehrplan 2004, S. 5) gefordert wird, thematisiert werden. Allerdings ist die *Regel von Bayes* nur für das Realgymnasium verpflichtend vorgesehen (S. 5).

Der Themenkreis „Schokolinsen“ eignet sich also für eine einfache *vertikale Vernetzung* über mehrere Jahrgangsstufen (vgl. Brinkmann & Siller 2012).

8. Weitere Vorschläge für Daten(gewinnung)

Im Seminar für LAK (Stochastik) des Erstautors im Wintersemester 2011/12 sind Sprungweiten von Papierfröschen (vgl. Eichler & Vogel 2009, S. 177 ff., Abbildung 17) und Spielzeugautos gemessen worden, dazu ist eine Sandkiste verwendet worden. So kann zur Frage der Unabhängigkeit von Merkmalen (Größe – Sprungleistung) ein Zufallsexperiment durchgeführt werden.



Abb. 17: kidsweb.de

Sprungweiten beim Schispringen können mittels Regression, Korrelation ausgewertet werden (1. Sprung – 2. Sprung) (vgl. Eichler & Vogel 2009, S. 89 ff.).

Zum subjektivistischen Wahrscheinlichkeitsbegriff können Quaderwürfel (Abbildung 18) im Vergleich zu Regelwürfeln herangezogen werden. Analog zu 7.2 wird hier nach dem Würfel gefragt (Regel oder Quader), mit dem die genannten Augenzahlen (Daten) geworfen werden (vgl. Eichler & Vogel 2009, S. 199 ff.).



Abb. 18: riemer-koeln.de

Die Feststellung der relativen Häufigkeiten des Auftretens der Augenzahlen beim Quaderwürfel kann als Schätzung der Wahrscheinlichkeitsverteilung herangezogen werden (frequentistischer Wahrscheinlichkeitsbegriff). Hier kann ja das Argument der Symmetrie anders als beim Regelwürfel nicht verwendet werden.

Das Wetter bietet eine quasi unerschöpfliche Quelle von Daten, die auf jedem Niveau statistisch ausgewertet werden können.

Das Werfen von vier Münzen („Bino-ley“, vgl. Eichler & Vogel 2009, S. 217 ff.) kann nach dem Auftreten von „Wappen“ klassifiziert werden: 0, 1, 2, 3 oder 4. Hält man die entsprechenden (absoluten) Häufigkeiten fest, so ergibt sich auf lange Sicht ein Muster, welches die Binomialverteilung vorbereitet.

9. Zusammenfassung

Abbildung 19 zeigt die wesentlichen Handlungen, die der datenorientierte Zugang fordert. Grob gesprochen und etwas vereinfacht entspricht der linke Strang dem, was in der Unterstufe gemacht werden kann, der rechte, der eine quantitative Auswertung zusätzlich mit sich bringt, ist der Oberstufe vorbehalten.

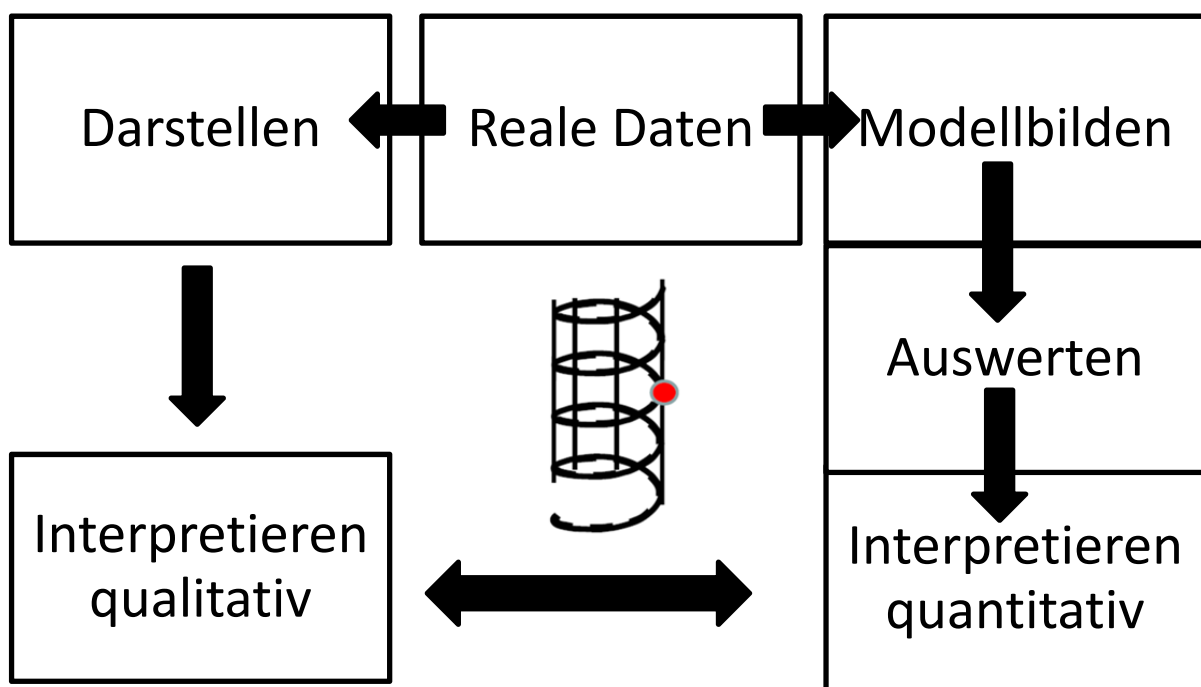


Abb. 19

Eine Klammer zwischen diesen beiden Strängen bildet die *Curriculumsspirale*, die ja gerade eine Wiederaufnahme einer bestimmten Thematik oder einer bestimmten Tätigkeit auf verschiedenen Ebenen (Jahrgangsstufen) fordert. Wenn wir davon ausgehen, dass am Ende jeder stochastischen Problemstellung die Interpretation der erhaltenen Ergebnisse stehen muss, ist an dieser Stelle des Modellbildungskreislaufes der gegenseitige Bezug von qualitativer und quantitativer Analyse besonders gegeben.

10. Resümee

In diesem Beitrag haben wir zwar kein neues Unterrichtskonzept entwickelt, aber dennoch Anlässe für den Einsatz des/der Datendetektivs/in anhand von realen Daten aufgezeigt. Es soll nicht verschwiegen werden, dass das Arbeiten mit realen Daten nicht immer die Hypothese stützt, die eventuell a priori aufgestellt worden ist. Zum Beispiel springen die Frösche nicht so, wie sie sollen, d. h. sie zeigen kein einheitliches Bild. Die Benzinverbrauchsdaten haben ergeben, dass die Angaben des Herstellers mit dem individuellen Fahrstil bzw. den Fahrgewohnheiten des Erstautors nicht kompatibel sind. Die am

Beginn in dem Zeitungsartikel propagierte Benzinpreissteigerung vor Ostern ist in der Betrachtung der globalen Entwicklung des Benzinpreises untergegangen.

Wie schon oben erwähnt, passiert der Datentransfer aus dem Internet in ein Softwareprogramm nicht immer ohne Probleme. Am Beispiel zum Mauna Loa haben wir weiters gesehen, dass das Analysieren der Residuen einen Wechsel des verwendeten Programms (von Excel zu GeoGebra) verlangt hat.

Allgemein kann festgehalten werden, dass Excel die Verarbeitung und Analyse *großer* Datenmengen gut bewältigt. Auf der anderen Seite sind Boxplot- und Stängel-Blatt-Diagramme nicht standardmäßig vorgesehen im Gegensatz zu GeoGebra. Wenn die Datenmenge nicht zu groß ist, bietet GeoGebra einen raschen Überblick über sie.

Weiters haben wir bemerkt, dass der datenorientierte Ansatz immer wieder Grundkompetenzen wie sie in *Die standardisierte schriftliche Reifeprüfung in Mathematik 2012* formuliert sind, einfordert. Die bildungstheoretische Orientierung des Inhaltsbereichs „Wahrscheinlichkeit und Statistik“ spricht von „Ergebnisse [...] im jeweiligen Kontext angemessen zu deuten und zu bewerten“ (S. 16).

Es ist evident, dass diesem Ansatz die Selbsttätigkeit der Schüler/innen inhärent ist.

Literatur

- Brinkmann, A., Siller, H-S. (2012): Vertikale Vernetzung über außermathematische Anwendungskontexte. In: Brandl, M.; Brinkmann, A.; Maaß, J.; Siller, H-S. (Hrsg.): *Mathe vernetzt. Anregungen und Materialien für einen vernetzten Mathematikunterricht*. Band 2. München: Aulis, 37–57.
- Bruder, R. (2012): *Konsequenzen aus den Kompetenzen?* Vortrag auf der Jahrestagung der GDM am 6.3.2012 in Weingarten. Online: <http://www.math-learning.com/files/120306wg.pdf> (Zugriff: 18.7.2012).
- Die standardisierte schriftliche Reifeprüfung in Mathematik* (Stand: April 2012). Online: <http://www.bifie.at/node/1442> (Zugriff: 18.7.2012).
- Eichler, A., Vogel, M. (2009): *Leitidee Daten und Zufall. Von konkreten Beispielen zur Didaktik der Stochastik*. Wiesbaden: Vieweg+Teubner. Online: http://www.leitideedatenundzufall.de/leitideedatenundzufall_index.html (Zugriff: 19.7.2012, Zusatzmaterialien).
- Götz, S., Süß-Stepancik, E. (2012): Welchen Mittelwert soll ich nehmen? Begriffsbildung im Stochastikunterricht. In: *mathematik lehren* 172, 20–24, 41.
- KMK (2003): *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss*. Online: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2003/2003_12_04-Bildungsstandards-Mathe-Mittleren-SA.pdf (Zugriff: 18.7.2012).
- Lehrplan (2000): *Lehrplan Mathematik für die AHS-Unterstufe*. Online: <http://www.bmukk.gv.at/medienpool/789/ahs14.pdf> (Zugriff: 18.7.2012).
- Lehrplan (2004): *Lehrplan Mathematik für die AHS-Oberstufe*. Online: http://www.bmukk.gv.at/medienpool/11859/lp_neu_ahs_07.pdf (Zugriff: 18.7.2012).
- M8 (2007): *Standards für die mathematischen Fähigkeiten österreichischer Schülerinnen und Schüler am Ende der 8. Schulstufe*. Online: http://www.uni-klu.ac.at/idm/downloads/Standardkonzept_Version_4-07.pdf (Zugriff: 18.7.2012).
- Winter, H. (1995): Mathematikunterricht und Allgemeinbildung. In: *Mitteilungen der Gesellschaft für Didaktik der Mathematik* Nr. 61, 37–46.

Verfasser/in

Stefan Götz
Universität Wien
Fakultät für Mathematik
Nordbergstraße 15
1090 Wien

stefan.goetz@univie.ac.at

Evelyn Süß-Stepancik
Universität Wien
Fakultät für Mathematik
Nordbergstraße 15
1090 Wien

evelyn.stepancik@univie.ac.at