Mit Daten sicher durch den Zufall

–

Wie der Stochastikunterricht in der Sekundarstufe gemäß der nationalen Bildungsstandards aussehen könnte

Andreas Eichler, IMBF Markus Vogel, Halle 2013


Eigener Herd ...
Umfragen ergaben, den Deutschen schmeckt es zu Hause immer noch am besten.
Ich koche oder bereite Essen zu
Anteil der Befragten in Prozent
mehrmals täglich 20,3
mehrmals wöchentlich 34,5
einmal wöchentlich 5,0
mehrmals monatlich 6,4
einmal monatlich 2,9
seltener 12,3
nie 16,5
Quelle: TdW Intermedia 2000/2001 von 100 Prozent Gesamtbevölkerung 2,1 Prozent o. A.

Gliederung

1. Einstiegsbeispiel
2. Bildungsstandards, wieso-weshalb-warum?
3. Von der *Notwendigkeit* guter Daten
4. Von der *Brisanz*, die mit Daten und Zufall verbunden ist
5. Von der *Unsicherheit*, die dem Zufall innewohnt und trotz der dennoch gute Aussagen über Daten möglich sind
6. Von der *Schönheit*, Daten (mit dem Rechner) untersuchen zu können
7. Übergreifende Ideen

Gliederung

1. Einstiegsbeispiel
2. **Bildungsstandards, wieso-weshalb-warum?**
3. Von der *Notwendigkeit* guter Daten
4. Von der *Brisanz*, die mit Daten und Zufall verbunden ist
5. Von der *Unsicherheit*, die dem Zufall innewohnt und trotz der dennoch gute Aussagen über Daten möglich sind
6. Von der *Schönheit*, Daten (mit dem Rechner) untersuchen zu können
7. Übergreifende Ideen

2. Wieso – weshalb – warum

Sek. I

Leitidee Daten und Zufall (in beiden Sekundarstufen)

Schülerinnen und Schüler sollen erfahren, dass

Fragen an alltägliche empirische Phänomene gestellt und mit den *elementaren* Methoden der Sekundarstufen (und dem Rechner) beantwort werden können

und: Ideen statt Algorithmen zählen.

Sek. II

Ideen vor Algorithmen – Realistisches vor Konstruiertem

Produktion:
Bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla bla **im Durchschnitt 2%** bla bla bla bla bla bla bla bla bla bla bla bla. Bla bla bla bla bla bla. Wie groß ist die Wahrscheinlichkeit dafür, dass **von den 20** bla bla bla bla bla bla **genau 3** bla bla bla bla bla bla ?“

Text: irrelevant
Sachsituation: irrelevant
Daten: irrelevant
Modell der Binomialverteilung, 3 Zahlen: relevant
Rechner: (eher) irrelevant

1. Einstiegsbeispiel
2. Bildungsstandards, wieso-weshalb-warum**?**
3. **Von der *Notwendigkeit* guter Daten**
4. Von der *Brisanz*, die mit Daten und Zufall verbunden ist
5. Von der *Unsicherheit*, die dem Zufall innewohnt und trotz der dennoch gute Aussagen über Daten möglich sind
6. Von der *Schönheit*, Daten (mit dem Rechner) untersuchen zu können
7. Übergreifende Ideen

Von Daten zu „guten“ Daten

Sek. I

Erste einfache
- Befragungen
- Experimente
- Beobachtungen

Erste Beurteilungen;
Variation: Was wäre wenn?

Erste Annäherung an die Repräsentativität

Eigener **Herd ...**

Umfragen ergaben, den Deutschen schmeckt es zu Hause immer noch am besten:

Ich koche oder bereite Essen zu	Anteil der Befragten in Prozent
mehrmals täglich	20,3
mehrmals wöchentlich	34,5
einmal wochentlich	5,0
mehrmals monatlich	6,4
einmal monatlich	2,9
seltener	12,3
nie	16,5

Quelle: TdW Intermedia 2000/2001 von 100 Prozent Gesamtbevölkerung 2,1 Prozent o. A.

Von Daten zu „guten“ Daten

Sek. II

1. Ausgangsproblem:
Wie gelange ich zu einer repräsentativen Umfrage?

2. Vereinfachung mit Schulbezug:
Stellen Sie Kriterien auf, mit denen Sie x SchülerInnen Ihres Jahrgangs auswählen, die Ihren Jahrgang gut repräsentieren

3. Analogie/Überspitzung:
Stellen Sie willkürliche Kriterien auf, mit denen Sie 100 Studierenden aus einer Grundgesamtheit von 1082 Studierenden ziehen. Bestimmen Sie jeweils die relativen Häufigkeiten einer Merkmalsausprägung

Von Daten zu „guten“ Daten

Sek. II

(100 repräsentative Studierende zur Merkmalsausprägung „wohnen hochschulnah“

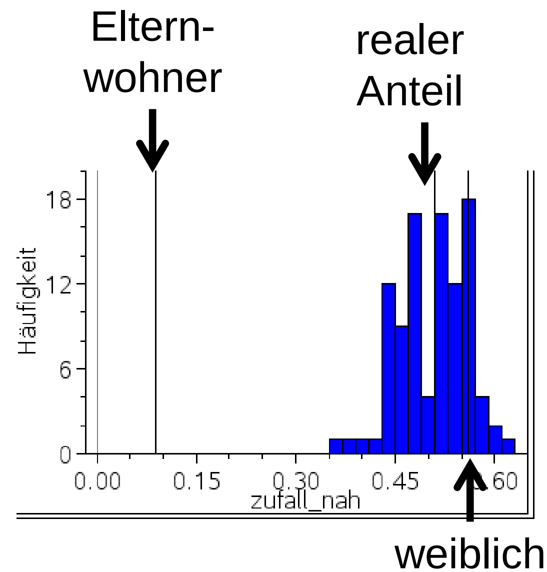
Eichler & Vogel, 2014
Biehler & Eichler, 2014

„wir nehmen die ersten 100 Studentinnen“
„wir nehmen die ersten 100 Bahnfahrer“
„wir nehmen die ersten 100, die bei den Eltern wohnen

Vergleich mit dem Populationsmittelwert

Eine einfache Zufallsstichprobe

Verhalten der Zufallsstichprobe in Abhängigkeit von der Stichprobe



1. Einstiegsbeispiel
2. Bildungsstandards, wieso-weshalb-warum?
3. Von der *Notwendigkeit* guter Daten
4. **Von der *Brisanz*, die mit Daten und Zufall verbunden ist**
5. Von der *Unsicherheit*, die dem Zufall innewohnt und trotz der dennoch gute Aussagen über Daten möglich sind
6. Von der *Schönheit*, Daten (mit dem Rechner) untersuchen zu können
7. Übergreifende Ideen

Sek. I Sek. II

Modell für die Würfel

Augenzahl	1	2	3	4	5	6
Wahrscheinlichkeit	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
Wahrscheinlichkeit	$0,05$	$0,1$	$0,35$	$0,35$	$0,1$	$0,05$

$$P(N) = P(Q) = 0{,}5$$

Es fällt die Augenzahl x

Aufgabe:

Die Spielleitung wählt verdeckt den normalen oder den quaderförmigen Würfel aus. Welcher ist es?

$$P(N|x) = \frac{P(x|N) \cdot P(N)}{P(x|N) \cdot P(N) + P(x|Q) \cdot P(Q)}$$

Verarbeiten der Information (x) und Anwendung der Formel von Bayes führt zu Neubewertung

Modell: Würfe des Würfels sind unabhängig!

Aus: SpiegelOnline vom 04.07.2012

HIV-Schnelltest für zu Hause: Wirkungsvolle Notlösung

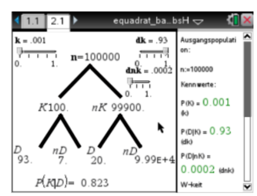
Von Irene Berres

Die Arzneimittelbehörde der USA hat einen HIV-Schnelltest für zu Hause zugelassen. Die vereinfachte Methode könnte dazu führen, dass sich mehr Menschen testen lassen. Die fehlende Beratung birgt jedoch auch Gefahren.

Basisrate (lt. Unaids)	Krank: P(K) = 0,001, P(K) = 0,999
Sensitivität (lt. Hersteller)	Wenn krank, dann Test+ (D): P(D|K) = 0,93; P($\bar{D}$|K) = 0,07
Spezifität (lt. Hersteller)	Wenn gesund, dann Test+ (D): P(D|$\bar{K}$) = 0,0002; P($\bar{D}$|$\bar{K}$) = 0,9998

Aufgabe: HIV

- Warum sollte man vor so einem Test noch warnen?



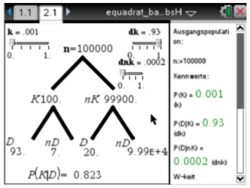
Aus: FAQs der Deutschen AIDS-Hilfe zum Schnelltest

„Bei den meisten HIV-Tests liegt die Spezifität in der Praxis bei etwa 99,7 %“

Basisrate (lt. Unaids)	Krank: P(K) = 0,001, P(K) = 0,999
Sensitivität (lt. Hersteller)	Wenn krank, dann Test+ (D): P(D|K) = 0,93; P($\overline{D}$|K) = 0,07
Spezifität (lt. Hersteller)	Wenn gesund, dann Test+ (D): P(D|$\overline{K}$) = **0,003**; P($\overline{D}$|$\overline{K}$) = **0,997**

Aufgabe: HIV

- Warum sollte man vor so einem Test noch warnen?
- Was passiert bei unsachgerechter Anwendung?



Aufgabe: HIV

- Grenzen der mathematischen Modellierung

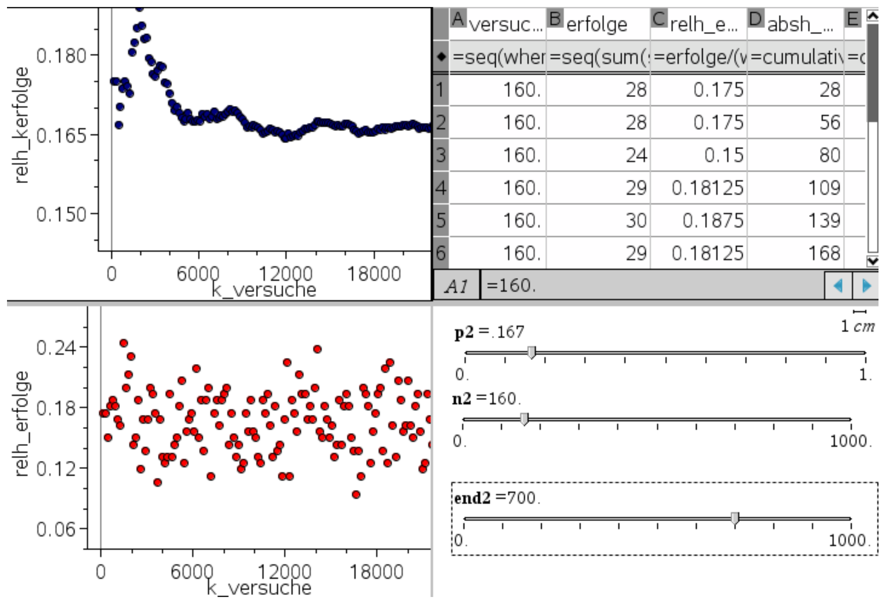
1. Einstiegsbeispiel
2. Bildungsstandards, wieso-weshalb-warum?
3. Von der *Notwendigkeit* guter Daten
4. Von der *Brisanz*, die mit Daten und Zufall verbunden ist
5. **Von der *Unsicherheit*, die dem Zufall innewohnt und trotz der dennoch gute Aussagen über Daten möglich sind**
6. Von der *Schönheit*, Daten (mit dem Rechner) untersuchen zu können
7. Übergreifende Ideen

Sek. I

Vom empirischen Gesetz der großen Zahlen zum Konfidenzintervall (Hypothesentest): Fünf Stufen

Stufe I: empirisches Gesetz der großen Zahl

1. Die relativen Häufigkeiten stabilisieren sich
2. Die relativen Häufigkeiten schwanken offenbar in einem Intervall



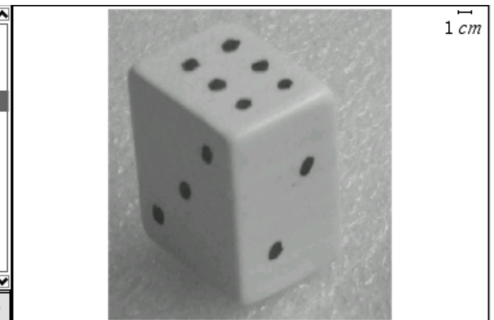
Vom empirischen Gesetz der großen Zahlen zum Konfidenzintervall (Hypothesentest): Fünf Stufen

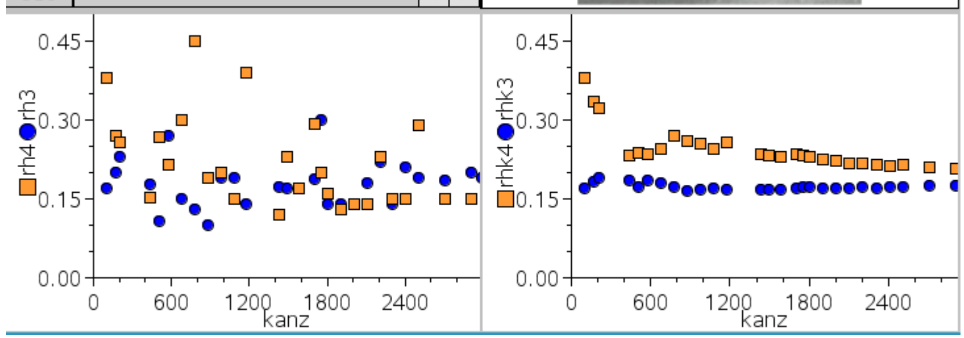
Sek. I

Stufe II: Vertrauen in das empirische Gesetz der großen Zahl für Punktschätzungen

	A anz	B anz3	C anz4	D rh3	E
◆				=seq(roun	=
23	100	14	15	0.14	
24	100	21	15	0.21	
25	100	19	29	0.19	
26	200	37	30	0.185	
27	200	40	30	0.2	
28	89	17		0.191	

C28



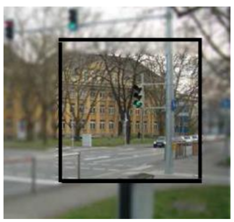


Für eine unbekannte Wahrscheinlichkeit ist die relative Häufigkeit nach einer langen Versuchsserie eine gute Schätzung

Vom empirischen Gesetz der großen Zahlen zum Konfidenzintervall (Hypothesentest): Fünf Stufen

Sek. II

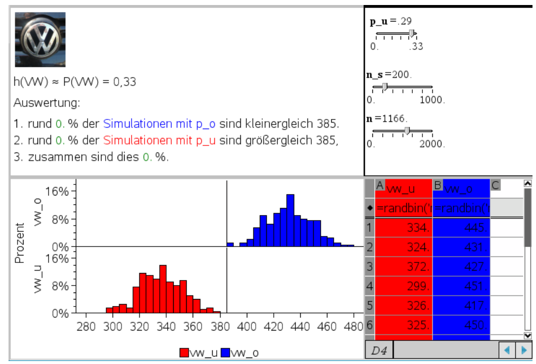
Stufe III: Wozu passt relative Häufigkeit eines Ereignisses?



In Braunschweig wurden bei 1166 PKW 385 der Marke VW gezählt, also rund 33%

Gibt es genau 33% VW in Braunschweig?

Könnten es auch 25% sein?



Konfidenzniveau 95%

Simulation: [0,30 ; 0,36]

Berechnung: [0,303;0,357]

Deutschland: 22 %

Niedersachsen: 35%

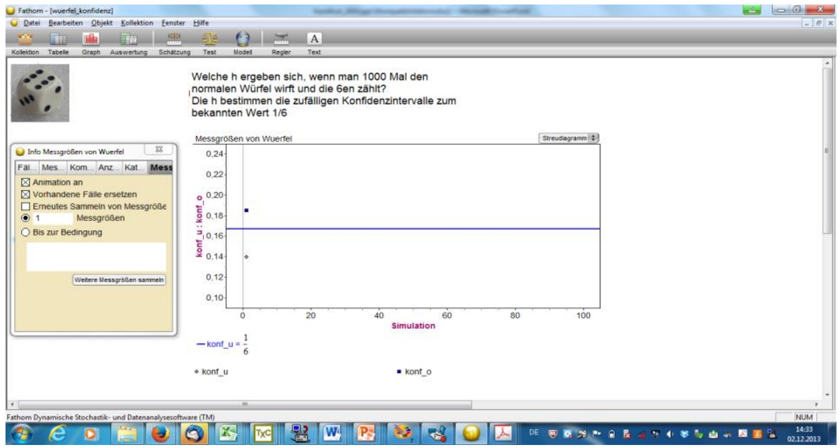
Sek. II

Vom empirischen Gesetz der großen Zahlen zum Konfidenzintervall (Hypothesentest): Fünf Stufen

Stufe IV: Was sagt ein Konfidenzintervall aus?

Konfidenzniveau 95%, [0,303;0,357]

Bei vielen (95%) solcher Stichproben würde man mit dem Konfidenzintervall den wahren Anteil erwischen.



Wie ist die Aussage, die SPD komme auf 28%?

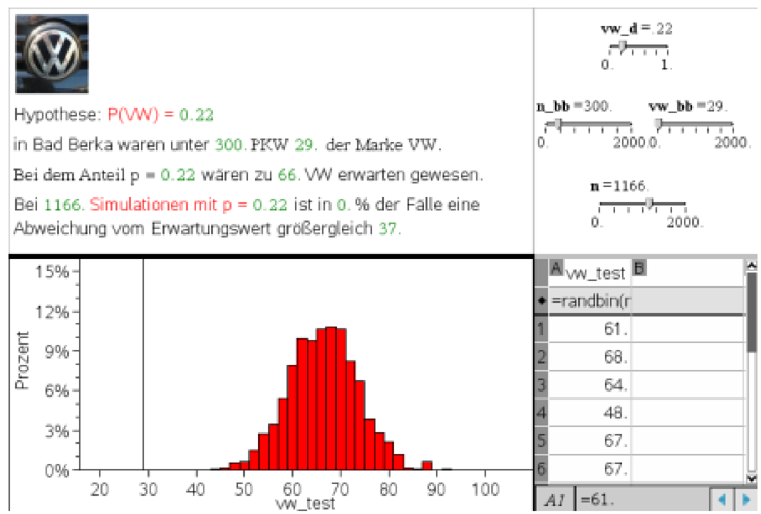
Sek. II

Vom empirischen Gesetz der großen Zahlen zum Konfidenzintervall (Hypothesentest): Fünf Stufen

Stufe V: Der kurze Weg zum Hypothesentest

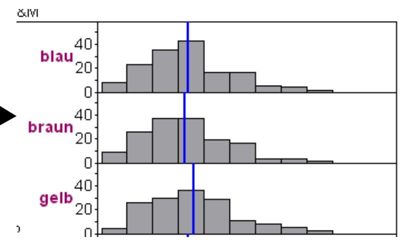
Passt der VW-Anteil in *Halle* eigentlich zum bundesweiten Markenanteil (0,22)?

Eine Schnell-Beobachtung hat unter 40 PKW 226 VW ergeben.



Vom empirischen Gesetz der großen Zahlen zum Konfidenzintervall (Hypothesentest): Fünf Stufen

Stufe V: Der kurze Weg zum Hypothesentest

 →  →  → 

Aufgabe: Untersucht die Inhalte dieser Tüten.

Daten sammeln

darstellen

und Modell bilden: 18 Kugeln pro Packung 3 pro Farbe

Stimmt das Gleichverteilungs-modell auch bei Smarties?

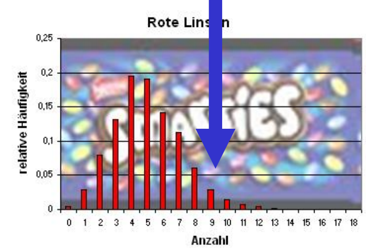
(40 Kugeln, 8 Farben)

→

Wahrscheinlichkeit von **genau drei** roten bei **40** Kugeln

Modell durcharbeiten

→



Hypothese testen

1. Einstiegsbeispiel
2. Bildungsstandards, wieso-weshalb-warum?
3. Von der *Notwendigkeit* guter Daten
4. Von der *Brisanz*, die mit Daten und Zufall verbunden ist
5. Von der *Unsicherheit*, die dem Zufall innewohnt und trotz der dennoch gute Aussagen über Daten möglich sind
6. **Von der *Schönheit*, Daten (mit dem Rechner) untersuchen zu können**
7. Übergreifende Ideen

Aufgabe:
Was für Eigenschaften haben Studierende/SchülerInnen?

Sek. I

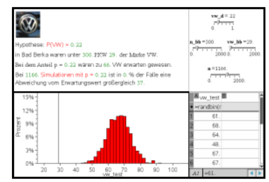
Sek. II

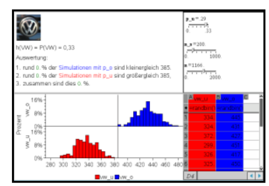
1. Einstiegsbeispiel
2. Bildungsstandards, wieso-weshalb-warum?
3. Von der *Notwendigkeit* guter Daten
4. Von der *Brisanz*, die mit Daten und Zufall verbunden ist
5. Von der *Unsicherheit*, die dem Zufall innewohnt und trotz der dennoch gute Aussagen über Daten möglich sind
6. Von der *Schönheit*, Daten (mit dem Rechner) untersuchen zu können
7. **Übergreifende Ideen**

Übergreifende Ideen

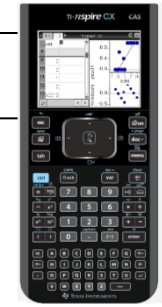
Die **Leitidee Daten und Zufall in der Sek. II** sollte zeigen, dass

- reale Fragestellungen beantwortet und nicht Algorithmen abgearbeitet werden
- mit Daten und Zufall Modelle der Realität gebildet werden können,
- auf Ideen fokussiert wird,
  - zu denen das Durcharbeiten eines Modells Mittel zum Zweck ist,
- Offenheit für Erweiterungen besteht,
  - aber auch Kürzungsmöglichkeiten in der Behandlungstiefe bestehen.



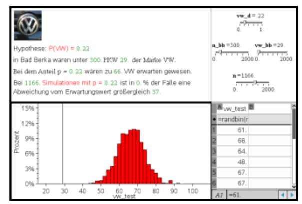


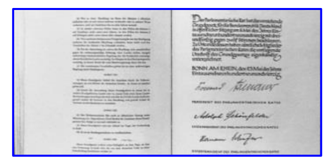
Rechner und Leitidee Daten und Zufall

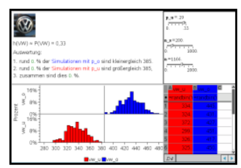


Rechner ersetzt nicht das Verständnis, aber hilft bei

- Bewältigen großer Datenmengen
- visuell gesteuerter Datenanalyse
- Simulation
- Unterscheidung Modell/Realität
- Elementarisierung konventioneller Methoden
- Erweiterung von Fragestellungen







Fathom Dynamic Data Software GeoGebra TinkerPlots DYNAMIC DATA EXPLORATION

Daten und Zufall

Danke für die Aufmerksamkeit