

Manfred BOROVCNIK, Klagenfurt

Möglichkeiten von EXCEL in der Statistikausbildung

Das Poster soll einige statistische Konzepte illustrieren und aufzeigen, welche Möglichkeiten eine Tabellenkalkulation bietet, didaktische Ideen umzusetzen. Im Poster werden Ausschnitte aus Sequenzen zum Korrelationskoeffizienten, zum üblichen Zugang zur Beurteilenden Statistik (Alternativtest und Maximum likelihood) sowie zum Resampling-Ansatz gezeigt.

1. Dynamische Visualisierung

Denken ist von bildhaften oder strukturellen Vorstellungen begleitet. Fischbein (1987) spricht von Intuitionen. Primäre Intuitionen werden im besten Fall im Unterricht direkt aufgegriffen und – in Richtung offizieller Vorstellungen – verändert. Daraus entstehen begleitend sekundäre intuitive Vorstellungen, die ein flüssiges Handhaben der mathematischen Begriffe und ein tieferes Verständnis ermöglichen. Darüber hinaus haben sekundäre Intuitionen ein großes Gewicht im Hinblick auf die Akzeptanz der gelernten Begriffe bei den Lernenden. Wie dieses Wechselspiel zwischen Theorie und Intuitionen zum besseren Verständnis der Theorie beiträgt, kann man in Borovcnik (1992) nachlesen.

Intuitives Denken zu erfassen und zu beeinflussen ist schwierig. Hilfreich sind hier Darstellungen der Begriffe in Bildern und Materialien. Heute würde man wohl den Begriff „embodiment“ verwenden, Visualisierung ist ein Teil davon. An der Idee der Visualisierung reizt insbesondere die dynamische Variante. Ein Begriff wird also nicht statisch in einem Bild erfasst, es werden vielmehr wesentliche Eigenschaften eines Begriffes durch Hantieren mit der Graphik einsichtig. Der Autor sieht in animierten Graphiken ein wesentliches didaktisches Werkzeug, das angesprochene Wechselspiel zwischen Intuitionen und abstrakten Begriffen in Gang zu bringen und ein tieferes Begriffsverständnis anzubahnen.

2. Einige Beispiele

Die Dynamik kann man ausnutzen, wenn man empirisch eine Regressionsgerade an bestehende Punkte anpasst und spielerisch durch Verschieben einer vorläufigen Geraden die Abweichungsquadrate verkleinert, bis man das Optimum erreicht (siehe Fig. 1a).

Die Darstellung der Daten in Tabellenform lässt bestimmte Berechnungen besonders einfach miteinander vergleichen: Daten, Vorhersage für die Daten in Form von Mittelwerten bzw. in Form von linearen Prognosen (das sind die Werte aus der Regressionsgeraden). Durch Ausrechnen von ein-

fachsten Statistiken wie Mittelwerten und Varianzen dieser drei Datensätze parallel erschließt man eine elementare Interpretation des so genannten Bestimmtheitskoeffizienten R^2 , dem Quadrat des Korrelationskoeffizienten, welche grundlegend für die weitere Begriffsbildung ist. Dynamisiert man noch die begleitende Punktwolke durch Regler für einzelne Punkte, sieht man die Invarianz der empirisch gefundenen Beziehung ein (siehe Fig. 1b).

Weitere Beispiele liegen auf der Hand: Beim Testen von Hypothesen in der beurteilenden Statistik kann man die Wechselwirkungen zwischen Fehler 1. und 2. Art, der Lage von zwei Alternativhypothesen zueinander etc. dynamisch schön herausarbeiten. Nicht neu, aber in der Dynamik wesentlich überzeugender als die statische Lehrbuchgraphik von ehemals (siehe Fig. 1c).

Auch komplexere Methoden wie die Maximum Likelihood-Methode zur Schätzung von Parametern werden einsichtig, wenn man dynamisch alle Modelle, die zur Verfügung stehen, durchprobiert. Man sieht die Verteilungen graphisch durch ihre Stabdiagramme (oder Dichtefunktionen) und justiert den in Frage stehenden Parameter so lange, bis man die vorhandene Beobachtung in ihrer Wahrscheinlichkeit optimiert (siehe Fig. 1d).

Wenn man Vertrauensintervalle wiederholt simuliert, spiegelt sich die Überdeckungswahrscheinlichkeit als Prozentsatz von überdeckenden Intervallen auf lange Sicht wider.

3. Zusammenfassung

„Früher“ war eine Programmierumgebung notwendig, um dieses Medium und sein unerschöpfliches Potential zur Unterstützung von Lernprozessen zu nützen. Auch heute werden etwa Java oder andere graphische Werkzeuge dazu genutzt. Der Autor möchte aufzeigen, dass EXCEL alles bietet, solche Animationen herzustellen. Es ist auch viel leichter, diese Animationen in EXCEL (oder einer anderen Tabellenkalkulation) herzustellen als üblicherweise angenommen. Die graphischen Möglichkeiten sind besser als der generelle Ruf. Die Möglichkeit, auch die Erzeugung von Daten zu simulieren, lässt dynamisch davon abhängige Graphiken darstellen.

EXCEL bietet überdies einen direkten Zugang für Studierende,

- die animierten Graphiken nicht nur abzuspielen, sondern selbst herzustellen
- Berechnungen mit den Graphiken zu verknüpfen, welche auftauchende Fragen beantworten helfen,

- zu verstehen, was diese Graphiken wirklich leisten – sie bilden keine magische Black box.

Studierende (der Betriebswirtschaft oder der Mathematik) machen sich die Technologie rasch zu Eigen und schöpfen die Möglichkeiten sehr gut aus, wie die Erfahrungen des Autors zeigen. Darüber hinaus gehören heute Fertigkeiten in einer Tabellenkalkulation zu den Grundtechniken, welche schon eine sekundäre Ausbildung vermitteln soll. Sie sind in vielen beruflichen und privaten Situationen mit Vorteil einzubringen. Dagegen sind Fertigkeiten in speziellerer Software ohne Belang und können nur strukturell auf andere Situationen übertragen werden.

In der Ausbildung in Statistik wird der Einsatz von Rechenhilfsmitteln ohnehin erforderlich sein, um langatmige Berechnungen und lästiges Arbeiten mit Tabellen zu vermeiden. Da ist eine Tabellenkalkulation jedem noch so guten Taschenrechner überlegen. Die Verbreitung von EXCEL ist als weiterer Pluspunkt anzumerken. Für statistische Feinarbeit bietet das Add-on REXCEL einen fließenden Übergang zur statistischen Sprache R. Für unterrichtliche Zwecke (in der Sekundarstufe und in der universitären Grundausbildung) reicht eine Tabellenkalkulation vollauf. Nicht alle Blätter des Posters wurden hier beschrieben; das Projekt deckt viele weitere Aspekte der Ausbildung ab.

Literatur

- Bartz, S.: Excelblatt vereinfacht Stochastik. *Stochastik in der Schule* 27 (2007) 2, 25-29.
- Borovcnik, M.: *Stochastik im Wechselspiel von Intuitionen und Mathematik.* Mannheim: BI 1992.
- Borovcnik, M.: Das Sammelbildproblem – Rosinen und Semmeln und Verwandtes: Eine rekursive Lösung mit Irrfahrten. *Stochastik in der Schule* 27 (2007) 2, 19-24.
- Borovcnik, M., Neuwirth, E.: Ein rekursiver Zugang zur Binomialverteilung. <http://www.osg.or.at/> (2006)
- Christie, D.: Resampling mit Excel. *Stochastik in der Schule* 24 (2004) 3, 22-27.
- Fischbein, E.: *Intuition in Science and Mathematics. An Educational Approach.* Dordrecht: D. Reidel 1987.
- Lorenz, R. J.: *Grundbegriffe der Biometrie*, 4. Auflage. Stuttgart: Fischer 1996.
- Meyer, J.: *Excel-Files in der Lehrerfortbildung.*
Auf Anfrage beim Autor: J.M.Meyer@t-online.de
- Neuwirth, E.: <http://sunsite.univie.ac.at/mailman/listinfo/Improve-excel>
- Neuwirth, E., Arganbright, D.: *The Active Modeler: Mathematical Modeling with Microsoft Excel.* Brooks/Cole 2004.
- Reckelkamm, B.: Der Tanz der Residuen - Erarbeitung statistischer Grundbegriffe mit Hilfe von EXCEL. *Stochastik in der Schule* 24 (2004) 3, 14-21.
- Spreadsheets in Education: <http://www.sie.bond.edu.au/>

Bestimmen der Regression spielerisch

Optisches Anpassen der Parameter für die beste Regressionsgerade

Verbessern mit Hilfe der Interaktion

Modelle: $n_{alt} = \text{Mittel } y$
für Pro- gnosen $LS: y = a + b \cdot x$

optimale Werte zum Vergleich **Regler**

a 1,295 **aktuell** 1,709
b 0,822 **aktuell** 0,983

Anstatt die Parameter mathematisch zu optimieren, werden die Werte interaktiv angepasst. Man sieht die Auswirkungen unmittelbar.

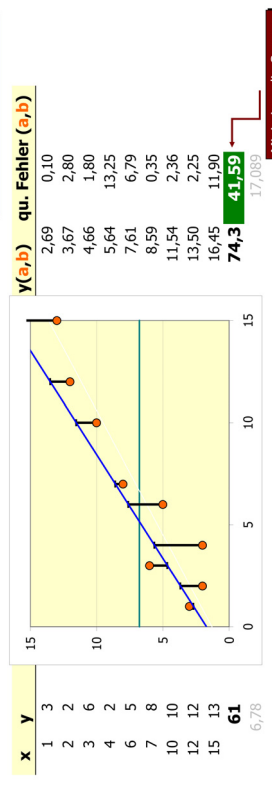


Fig. 1a

Dynamische Szenarien über das Verhalten von Käfern

Ein Käfer wird durch ein Labyrinth geschickt und kann bei einer am Ende platzierten Geruchsquelle landen – oder nicht. (Lorenz, Biometrie, 1988)

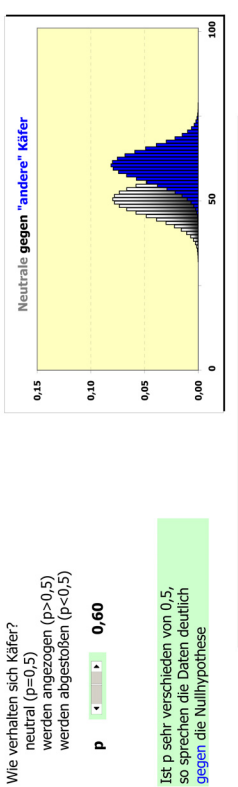
Szenarien: Wie verhalten sich Käfer - Wie bewährt sich die Entscheidungsregel Mit n=100 Käfern, die Werte für p werden dynamisch mit Reglern eingestellt.

Zahl der "Erfolge" – Verteilung in Abhängigkeit vom wahren p

Wie verhalten sich Käfer?
 neutral (p=0,5) werden angezogen (p>0,5) werden abgestoßen (p<0,5)

P 0,60

Ist p sehr verschieden von 0,5, so sprechen die Daten deutlich gegen die Nullhypothese



Die Folgen einer Vermutung, wie sich Käfer verhalten, auf die Anzahl der "Erfolge" können durch den Regler für p schnell erkannt werden.

Fig. 1c

Additivität für Quadratsummen (Varianzen) - unabhängig von speziellen Daten

Vorhersagen & Korrelation Vergleich von univ. & lineare Vorhersage Additivität für Quadratsummen

y-Daten werden aus x-Daten vorhergesagt: Vergleich von "einfachem" und linearem Modell Die Daten werden durch den Regler interaktiv verändert.

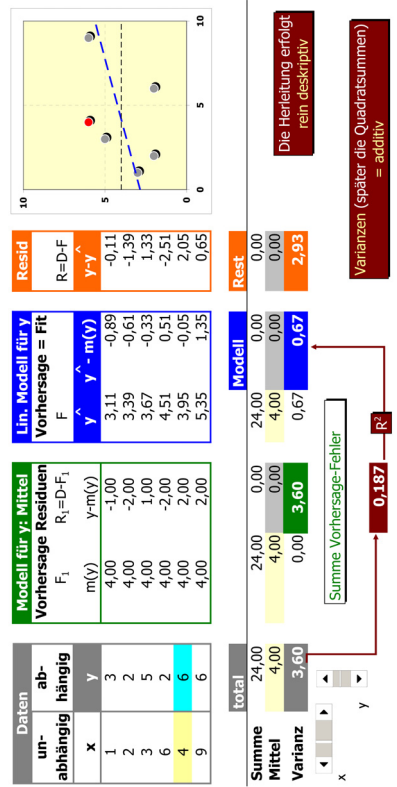


Fig. 1b

Maximum Likelihood-Schätzung von N in einer hypergeometrischen Verteilung

Graphische Bestimmung der Schätzung

Capture-Recapture-Problem

Es werden W = 10 Tiere markiert und der Population wieder untergemischt, nach Mischung (Zufall) werden n Tiere gefangen, darunter sind x markierte. Wie groß ist die Gesamtzahl aller Tiere N?

Population **N = ?** W markiert **N-W nicht markiert**
 Stichprobe **n = x** **12** **15** **5**

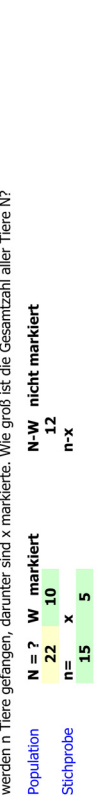


Fig. 1d

MLE = maximum likelihood = nehme jenen Wert für N, der die Beobachtung mit möglichst großer Wahrscheinlichkeit ausstärkt.