

Online-Tools zum Lernen und Lehren von Statistik

Autor: Nils MEVENKAMP

Institutionen: MCI | Die Unternehmerische Hochschule®

Track: Technologie

Format: Interaktives Format

1. Hintergrund

Trotz oder vielleicht auch gerade wegen der Akademisierung der beruflichen Bildung (Annen, Maier 2023) und der zunehmenden Bedeutung von Kompetenz im Umgang mit Daten und statistischen Informationen auch jenseits von hochschulischen Kontexten (Ickstadt, Müller, Weinert 2022) kommt dem Fach Statistik in der Hochschullehre eine zentrale Bedeutung zu. Ähnlich wie in den Fächern Mathematik oder Informatik wird auch in der Statistikausbildung ein hoher Grad an Abstraktion gefordert (Mirolo et al. 2022) und die Fähigkeit, sich frei und sicher zwischen verschiedenen Abstraktionsebenen bewegen zu können (Rich, Yadav, Zhu 2019). Ein Grundkurs in deskriptiver Statistik wird z.B. die Beschreibung univariater Verteilungen thematisieren. Mit zunehmendem Abstraktionsgrad unserer Alltagsbeobachtungen sprechen wir dabei zunächst über Merkmale, Merkmalsträger und Variablen, dann über Skalenniveaus und Verteilungen, die wiederum durch Häufigkeiten bzw. Zentralitäts-, Streuungs- und Schiefeparameter beschrieben werden. Es folgt eine Diskussion von Grundgesamtheit und Stichprobe, Strukturvariablen und Repräsentativität, um abschließend den Begriff der Variation in seiner Beziehung zum Gesetz der großen Zahl zu diskutieren. In kürzester Zeit hat man so ein Abstraktionsniveau erreicht, das für viele Studierende früher wie heute schwer nachzuvollziehen ist. Zur Inferenzstatistik ist der Weg aber noch immer weit.

Nach meiner langjährigen Erfahrung als Hochschullehrer in der Methodenausbildung besteht eine der größten Schwierigkeiten für Studierende darin, die unterschiedlichen Abstraktionsebenen aufeinander zu beziehen und zwischen ihnen zu wechseln. Studierende besitzen zwar das zum Bestehen einfacher standardisierter Lernzielkontrollen notwendige Mikrowissen (z.B. in Form von Begriffsdefinitionen), können aber oft selbst triviale Entscheidungen nicht daran anknüpfen. Vielen Studierenden fällt beispielsweise die Wahl zwischen einem Balkendiagramm und einem Histogramm schwer, weil sie die Bedeutung des Skalenniveaus der zu visualisierenden Variable für ihre Entscheidung nicht sehen. Später zeigen sie dann z.B. eine Präferenz für Zählen statt Messen und favorisieren Kreuztabellen für bivariate Analysen, auch wenn z.B. der Vergleich von Durchschnittswerten besser geeignet wäre.

Ziel jeder Grundausbildung in Statistik muss sein, Studierende zu befähigen, gewissermaßen in Begriffen wie „Verteilung“ und „Variation“ zu denken. In den gängigen Lehrbüchern zur Statistik wird Methodenwissen ganz überwiegend nach Abstraktionsebenen in einzelne Kapitel zusammengefasst. Im Sinne einer konstruktivistischen, aktiven Natur des Lernens kann das so vermittelte Mikrowissen aber erst effektiv genutzt werden, wenn zugleich Erfahrungen in seiner Anwendung gemacht werden – und darin unterscheiden sich Studierende kaum von jüngeren Lernenden (Lajoie et al. 1995). Bei der Entwicklung geeigneter Lern- und Lehrtools muss es darum um dreierlei gehen:

- Abstraktionen müssen erfahrbar sein, in dem Sinne, dass ein selbstständiger und interaktiver Umgang mit unterschiedlichen Grundgesamtheiten, Stichproben, Variablen und Verteilungen ermöglicht wird. Parallelen zur Gamification (Boyle et al., 2014) sind hier erwünscht.
- Abstraktionsebenen und Deskriptionen müssen dabei integriert, d.h. immer alle zugleich sichtbar sein, um so die Zusammenhänge zwischen den einzelnen Ebenen erfahren zu können.

- Abstraktionen müssen inhaltlich sinnvoll sein und sich auf die Lebenswelt der Anwender:innen zurückführen lassen. Statistik ist schließlich kein Selbstzweck, sondern bietet einen Zugang zum Verständnis sozialer und/oder naturwissenschaftlicher Phänomene mit Hilfe von Massendaten. Die dazu notwendigen Kompetenzen werden heute oft unter dem Begriff der Data Literacy subsumiert (z.B. Stifterverband 2023).

Bei der vorgestellten Arbeit handelt es sich (noch) nicht um ein Forschungsprojekt, sondern vielmehr um einen Entwicklungs- und Erfahrungsbericht. Es geht dabei im Kern um die Frage, wie ein Lern- und Lehrtool für die Statistikausbildung gestaltet werden kann, das die freie und sichere Bewegung zwischen verschiedenen statistischen Abstraktionsebenen fördert und einen leichten Zugang zum kompetenten Umgang mit statistischen Abstraktionen bietet.

2. Umsetzung

Es wurde eine interaktive Online-Anwendung als R Shiny-App realisiert, die webbasiert und ohne weitere technische Anforderungen im Präsenz- wie im Online-Unterricht eingesetzt werden kann. Alle Teilnehmer:innen müssen dazu auf eine gemeinsame Webseite zugreifen. Für die Nutzung muss sich eine Person (i.d.R. der/die Lehrende) als Administrator:in anmelden, die dann für die Steuerung zuständig ist. Sie kann den anderen Teilnehmer:innen drei unterschiedliche Modi anbieten.

Im „Lecture Mode“ teilt die Administrator:in die dargestellte Grafik mit allen Teilnehmer:innen. Das Steuerungsmenü ist allein für die Administrator:in sichtbar und entsprechend bedienbar. Schritt für Schritt können dann von Seiten der Lehrenden ausgehend von Punkten (Merkmalsträgern) auf einem Zahlenstrahl (Variablenwerte) die verschiedenen Darstellungsoptionen der deskriptiven Statistik (Werteklassen, Klassenhäufigkeiten, Histogramme, Lage- und Streuungsparameter, Boxplots) entwickelt werden. Die Teilnehmer:innen sehen dabei alle dieselbe, von der Administrator:in definierte Darstellung auf ihren Monitoren, haben dabei aber zugleich die Möglichkeit, die Daten mit Hilfe von Kontextinformationen (Mouse Over) und Zoomfunktion zusätzlich auf individuelle Weise zu entdecken. Abb. 1 zeigt ein Szenario für eine Administrator:in (links) und vier Teilnehmer:innen (rechts), in dem am Beispiel der extrem rechtsschiefen Verteilung der Pro-Kopf-Wirtschaftsleistung aller Staaten dieser Erde die Aussagekraft häufigkeitsbasierter und verschiedener parametrischer Beschreibungen verglichen wurden. Als Datengrundlage dienen World Bank Open Data für 2018 (<https://data.worldbank.org/>).

Der „Game Mode“ stellt die interaktive Erweiterung des Lecture Modes dar. Hier können die Teilnehmer:innen per Mausklick miteinander interagieren. Eine naheliegende Anwendung zur Übung und Festigung des Verständnisses der erlernten statistischen Parameter besteht z.B. darin, die Teilnehmer:innen auf Basis gegebener Verteilungen in Form von Punktreihen oder Histogrammen verschiedene Lageparameter schätzen zu lassen. Dies bietet sich insbesondere für einfache Übungen zu Median und Quartilen auf Basis kleiner Datensätze an (aufgrund der Fallzahl von $n=9$ sind z.B. Daten der Statistik Austria zu den österreichischen Bundesländern sehr gut geeignet).

Abb. 2+3 zeigen eine Nutzungsvariante, die eher auf die inhaltliche Dimension der Statistik und Data Literacy zielt. Den Teilnehmer:innen wird dabei nur der Name einer Variablen und ihr Wertebereich genannt. Im Beispiel wird für die Lebenserwartung ab Geburt aller Länder dieser Erde ein Wertebereich von unter 50 bis über 80 Jahren angegeben, und die Teilnehmer:innen werden gebeten, den Median zu schätzen. Sie können nun mit der Maus die Stelle auf der X-Achse markieren, an der sie den Median vermuten (Abb. 2). Wenn alle ihre Wahl getroffen haben, wird die Eingabe gesperrt. Daraufhin können die Eingaben aller Teilnehmer:innen geteilt werden, so dass jede:r einzelne Teilnehmer:in die eigene Schätzung mit der Schätzung aller anderen vergleichen kann. Zusätzlich kann auch der korrekte Wert

angegeben werden (Abb. 3). Alle Teilnehmer:innen bekommen auf diese Weise sofort Feedback, und zwar sowohl (absolut) zur Genauigkeit ihrer eigenen Schätzentscheidung als auch (relativ) zu ihrer Leistung im Vergleich zu allen anderen Teilnehmer:innen. Letztere erfolgt selbstverständlich anonym.

Auch ein „Free Mode“ ist verfügbar. Hier wird allen Teilnehmer:innen das Menüsystem zur Verfügung gestellt (die Möglichkeit zur Wahl des „Play Modes“ ausgenommen). Damit können alle Teilnehmer:innen das System eigenständig nutzen, um individuelle explorative Datenanalyse anhand eigener Datensätze zu praktizieren.

3. Weiterentwicklung

Bei dem vorgestellten Tool handelt es sich um einen Prototypen, der derzeit noch nicht öffentlich nutzbar ist. Das Tool wird allerdings seit einigen Wochen erfolgreich in der Methodenlehre am MCI | Die Unternehmerische Hochschule® eingesetzt. Dabei zeigt sich, dass das interaktive Schätztool bei den Studierenden besonders beliebt ist. Nach vorläufiger eigener Einschätzung wirkt sich wiederholtes Spielen, z.B. zehn Minuten zu Beginn jeder Lehrveranstaltung, sehr positiv auf die Kompetenz von Studierenden aus, die Größenordnung statistischer Parameter inhaltlich abzuschätzen und Daten anhand ihrer Verteilungen einzuschätzen.

Als naheliegende Erweiterung ist derzeit die Implementierung von Methoden zum Umgang mit Nominal- und Ordinalskalen (Rankings, Balkendiagramme) in Arbeit. Des Weiteren ist geplant, die Schätzwerte der Studierenden im Zeitverlauf aufzuzeichnen. Eine Zeitreihe von Variationskoeffizienten der einzelnen Schätzzunden könnte beispielsweise ein geeignetes Verfahren darstellen, um den Teilnehmer:innen ein Feedback zu ihrem Kompetenzerwerb zu geben und das Tool zugleich auch einer Wirkungsevaluation zu unterziehen.

Das Tool bietet auch die Basis zur interaktiven Visualisierung von zentralem Grenzwertsatz, t-Verteilung und Konfidenzintervallen. Eine entsprechende Anwendung wurde ebenfalls bereits als Prototyp realisiert, die Integration von beschreibenden und inferenzstatistischen Prinzipien steht allerdings noch aus.

Literaturverzeichnis

- Annen, S. & Maier, T. (Hrsg.). (2023). Akademisierung, Hybridqualifikationen und Fachkräftebedarf. Ist die Konkurrenz zwischen akademisch und beruflich Qualifizierten Mythos oder Realität? (Berichte zur beruflichen Bildung, 1st). Leverkusen-Opladen: Budrich.
- Boyle, E. A., MacArthur, E. W., Connolly, T. M., Hainey, T., Manea, M., Kärki, A. et al. (2014). A narrative literature review of games, animations and simulations to teach research methods and statistics. *Computers & Education*, 74, 1–14. <https://doi.org/10.1016/j.compedu.2014.01.004>
- Ickstadt, K., Müller, H. & Weinert, H. (2022). Data Literacy. In C. Weihs (Hrsg.), *Statistische Datenanalyse im Journalismus: Fallstudien und wissenschaftliche Anforderungen zum Einsatz fortgeschrittener statistischer Methoden* (S. 29–42). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-64693-9_3
- Lajoie, S. P., Jacobs, V. R. & Lavigne, N. C. (1995). Empowering children in the use of statistics. *The Journal of Mathematical Behavior*, 14(4), 401–425. [https://doi.org/10.1016/0732-3123\(95\)90039-X](https://doi.org/10.1016/0732-3123(95)90039-X)
- Miroló, C., Izu, C., Lonati, V. & Scapin, E. (2022). Abstraction in Computer Science Education: An Overview. *Informatics in Education*, 20(4), 615–639. <https://doi.org/10.15388/infedu.2021.27>
- Rich, K. M., Yadav, A. & Zhu, M. (2019). Levels of Abstraction in Students' Mathematics Strategies: What Can Applying Computer Science Ideas about Abstraction Bring to Elementary Mathematics? *Journal of Computers in Mathematics and Science Teaching*, 38(3), 267–298.
- Stifterverband, 2023. Data Literacy Education. <https://www.stifterverband.org/data-literacy-education>.

Abb. 1: Im „Lecture Mode“ für eine Administrator:in (links) und vier Teilnehmer:innen (rechts) werden die Darstellungen zentral gesteuert, individuelle Interaktivität ist dennoch möglich.

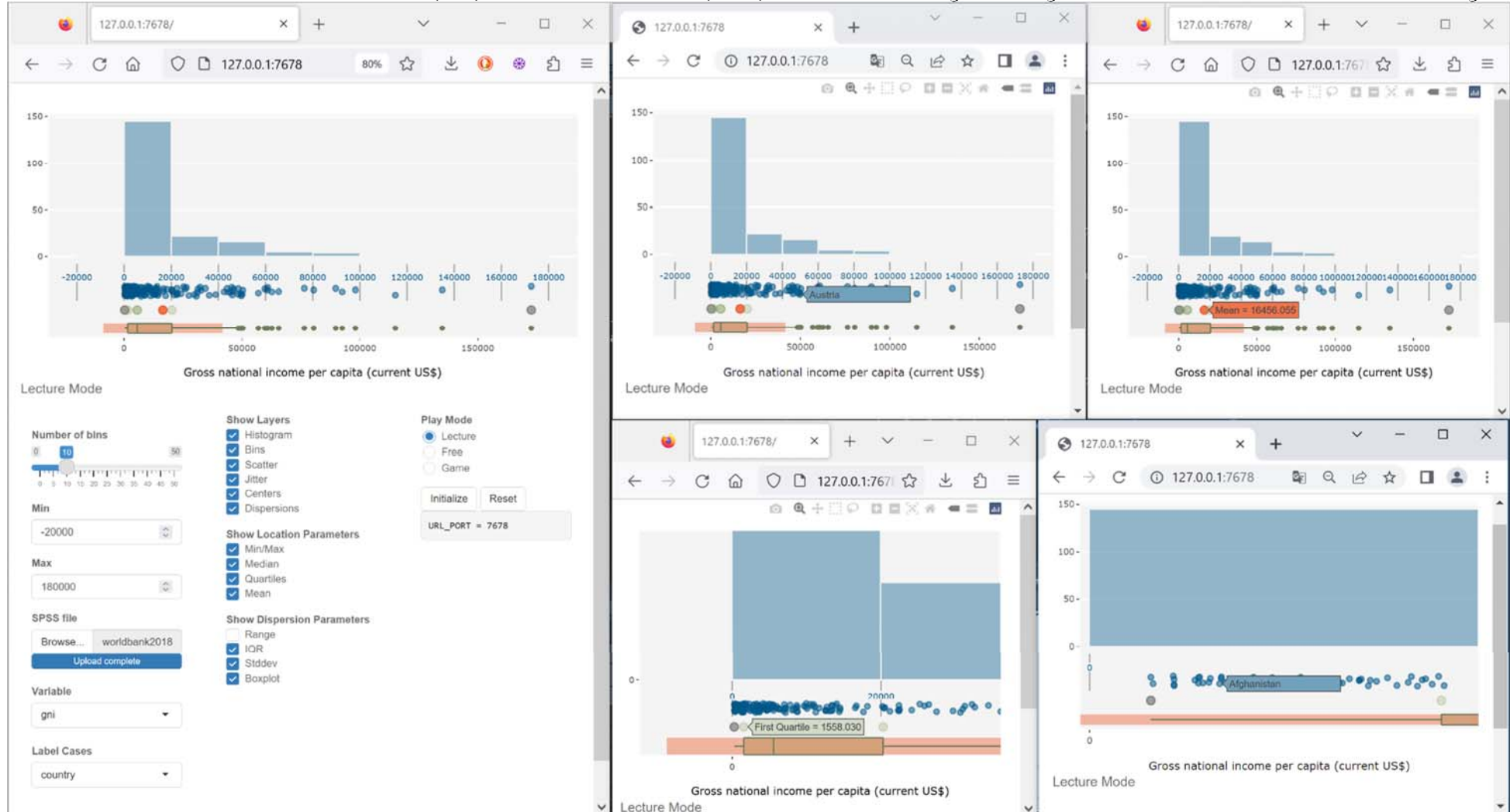


Abb. 2: Im „Game Mode“ können Teilnehmerinnen (rechts) unabhängig voneinander Parameter von Verteilungen schätzen. Auch die Administrator:in (links) kann sich beteiligen.

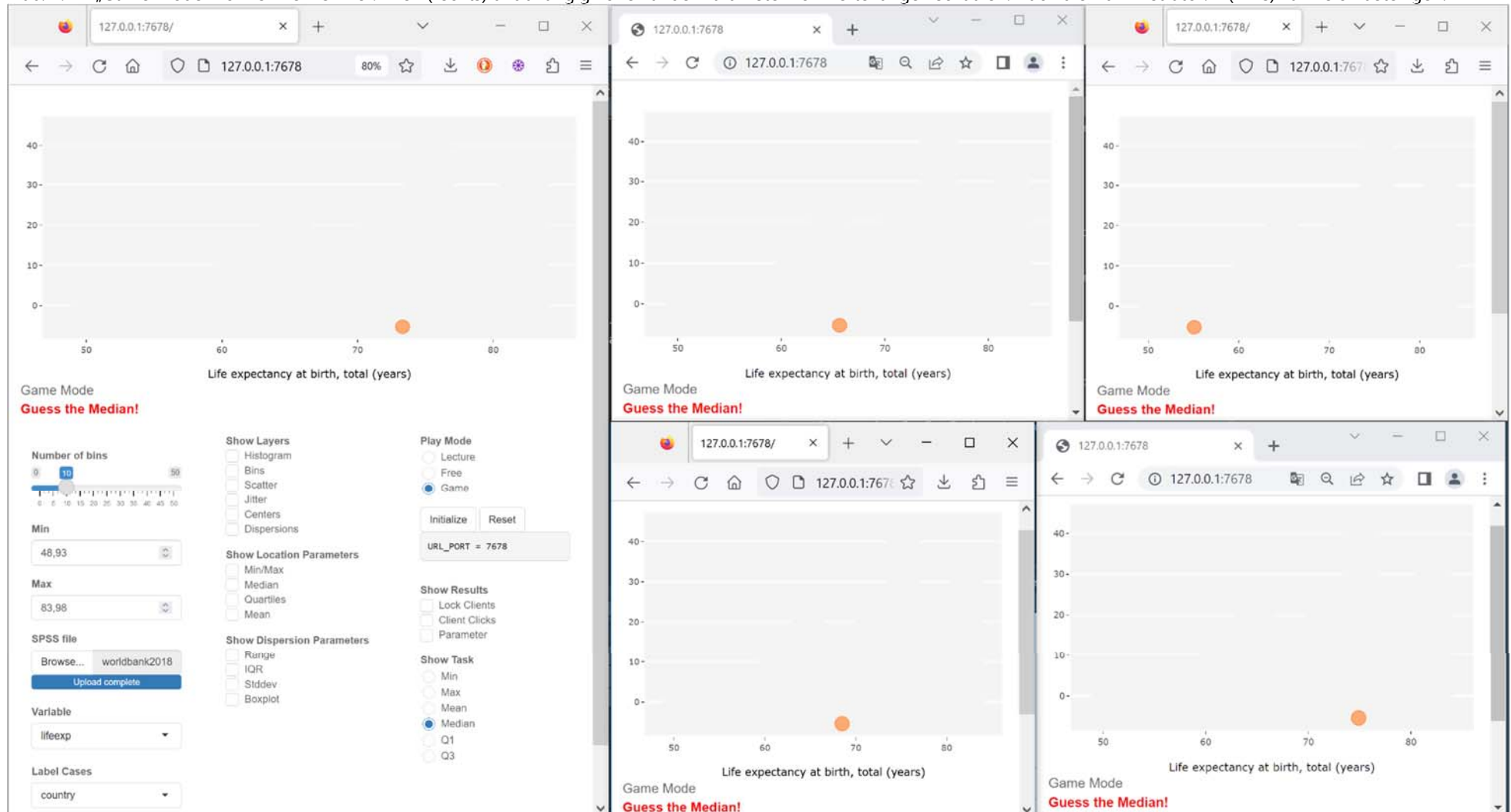


Abb. 3: Am Ende einer Spielrunde im „Game Mode“ kann die eigene Schätzung mit der richtigen Lösung und den Lösungen aller anderen Teilnehmer:innen verglichen werden.

