

Milena DAMRAU, Bielefeld, Hernán VILLAMIZAR, Wien & Martin SKRODZKI, Berlin

## Eine Datenanalyse der Persistenz und Leistung von Schulkindern im Wettbewerb „Mathe im Advent“

Auf Grundlage der Daten von ca. 100.000 Schüler\*innen, die 2016 am Wettbewerb „Mathe im Advent“ teilgenommen haben, untersuchen wir den Einfluss verschiedener Faktoren auf Leistung und Persistenz der Schüler\*innen im Wettbewerb.

### Der Schülerwettbewerb „Mathe im Advent“

„Mathe im Advent“ (kurz MiA) ist ein digitaler Schülerwettbewerb mit 24 Aufgaben. Er richtet sich an Schüler\*innen der Klassen 4-9 aller Schularten. Die Teilnahme ist sowohl alleine, als auch gemeinsam als Klasse/Gruppe möglich. Für Einzelspieler\*innen ist die Teilnahme kostenlos, für Klassen kostenpflichtig (wenn sie einen Klassenpreis gewinnen wollen). Weitere Informationen zum Wettbewerb: [www.mathe-im-advent.de](http://www.mathe-im-advent.de).

### Auswertung statistischer Daten

Bei der Auswertung der statistischen Daten interessiert uns vor allem, welchen Einfluss die (bezahlte) Klassenteilnahme auf Persistenz und Leistung hat. Wir vergleichen hierbei mit zwei weiteren Faktoren (Geschlecht, Gewinnspieltteilnahme). Wir möchten wissen, ob Schüler\*innen durch die Klassenteilnahme mehr (richtige) Antworten abgeben als Einzelspieler\*innen und ob dies weiter durch die kostenpflichtige Teilnahme am Klassenspiel beeinflusst wird.

Abb. 1 zeigt, welcher Anteil der Klassenspieler\*innen bzw. Einzelspieler\*innen wie viele Lösungen abgegeben hat. Sehr viele Einzelspieler\*innen (etwa 24%) haben sich zwar bei MiA registriert, jedoch keine Lösung abgegeben. Bei den Klassenspieler\*innen sind dies deutlich weniger (knapp 5%). Auch ist der Anteil der Einzelspieler\*innen, die alle 24 Lösungen abgegeben haben, geringer als der der Klassenspieler\*innen (26% gegenüber 35%). Es

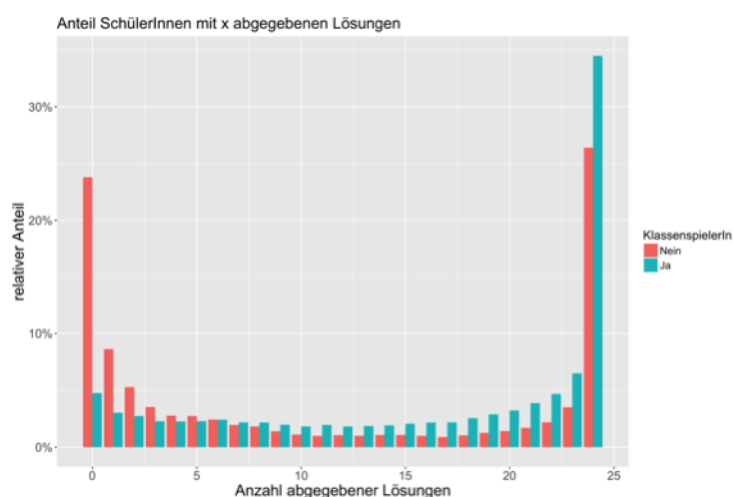


Abb. 1: Abgeg. Lösungen aller Schüler\*innen

wirkt so, als habe die Teilnahme als Klasse einen positiven Einfluss auf die Anzahl abgegebener Lösungen. Aber sind diese Gruppen überhaupt vergleichbar? Unter den Einzelspieler\*innen sind sowohl die zum Gewinnspiel angemeldeten, als auch diejenigen, die nur registriert waren, also nichts gewinnen konnten. Klassenspieler\*innen sind jedoch immer automatisch zum Gewinnspiel angemeldet.

Beschränken wir uns bei den Einzelspieler\*innen auf diejenigen, die zum Gewinnspiel angemeldet waren, sieht das Ergebnis anders aus (siehe Abb. 2). Der Anteil der Einzelspieler\*innen, die keine Lösung abgegeben haben ist zwar noch immer größer, als der der Klassenspieler\*innen, der Unterschied ist aber nicht mehr sehr groß (6% gegenüber 5%).

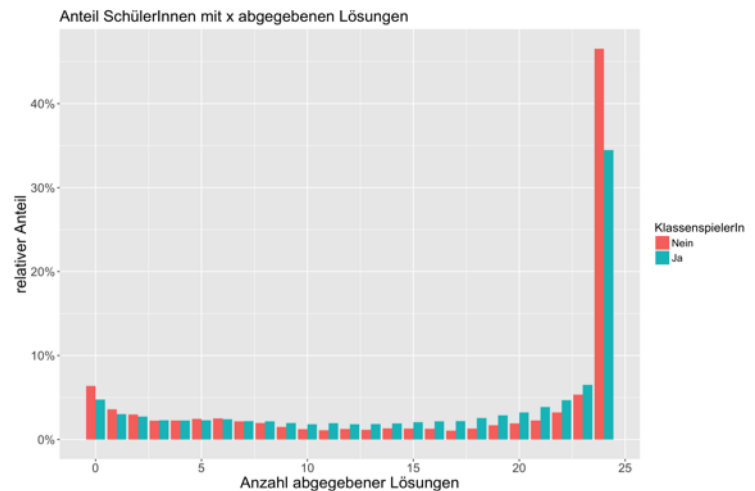


Abb. 2: Abgeg. Lösungen nur Gewinnspieler\*.

Der Anteil der Einzelspieler\*innen, die 24 Lösungen abgegeben haben, ist nun deutlich größer, als der der Klassenspieler\*innen. Ähnliche Ergebnisse erhalten wir bei der Anzahl richtiger Lösungen.

## Forschungsfragen und Überlegungen

Wir vermuteten dennoch, dass sowohl die Teilnahme als Klasse, als auch deren Zahlung einen positiven Einfluss auf Leistung und Persistenz hat.

Wie aber kann Leistung anhand der uns zur Verfügung stehenden Daten quantifiziert werden? Wir untersuchen die folgenden drei Größen: die Anzahl abgegebener und richtiger Lösungen, sowie das Verhältnis beider.

Wie wir gesehen haben, können wir nicht einfach die Leistung der Einzelspieler\*innen mit der Leistung der Klassenspieler\*innen vergleichen, da ein Selection Bias vorliegt. Wir brauchen Kontrollfaktoren. Einen solchen haben wir schon betrachtet: die Anmeldung zum Gewinnspiel. Des Weiteren haben wir die Größe „aktiv“ als Kontrollfaktor mit aufgenommen. „Aktive“ Spieler\*innen haben mindestens eine der 24 Lösungen abgegeben.

## Ordinary Least Squares – Ergebnisse

Um den möglichen Einfluss der Faktoren auf die Leistung zu untersuchen, haben wir die Methode der *Ordinary Least Squares* (OLS) genutzt (siehe z.B. Opfer, G., „Numerische Mathematik für Anfänger“).

Abb. 3 zeigt die Ergebnisse der OLS für die Anzahl abgegebener Lösungen. In Modell (4) wurden alle vier oben genannten Faktoren und die „Aktivität“ berücksichtigt. Der „aktiv“-Status korreliert am stärksten mit der Anzahl abgegebener Lösungen. Bei aktiven Spieler\*innen können ca. 11 abgegebene Lösungen mehr erwartet werden. Auch die Gewinnspielteilnahme korreliert stark mit der Anzahl abgegebener Lösungen. Die Klassenspieler\*innen zeigen ebenfalls eine positive Korrelation. Diese Effekte sind auch bei der Anzahl richtiger Lösungen und beim Verhältnis zu erkennen (ohne Abb.).

	Dependent variable:			
	Abgegebene Lösungen			
	(1)	(2)	(3)	(4)
weiblich	0.042 (0.061)	0.573*** (0.052)	0.399*** (0.047)	0.402*** (0.047)
KlassenspielerIn	5.538*** (0.064)	-0.388*** (0.064)	-0.565*** (0.057)	1.081*** (0.124)
Gewinnspielteilnahme		12.130*** (0.065)	8.181*** (0.063)	8.281*** (0.063)
aktiv			11.064*** (0.069)	11.022*** (0.069)
gezahlt:Nein				1.853*** (0.124)
Constant	10.855*** (0.049)	4.384*** (0.054)	-1.940*** (0.063)	-3.770*** (0.137)
Observations	100,997	100,997	100,997	100,997
R <sup>2</sup>	0.069	0.310	0.449	0.450
Adjusted R <sup>2</sup>	0.069	0.310	0.449	0.450

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Abb. 3: OLS Anzahl abgeg. Lösungen

Um den Einfluss der Zahlung einer Klasse auf die Leistung zu untersuchen, haben wir ebenfalls OLS Regressionen durchgeführt (ohne Abb.). Wir haben die Klassenspieler\*innen getrennt betrachtet, da Einzelspieler\*innen alle als „nicht bezahlt“ gelten. Die Zahlung einer Klasse korreliert leicht mit der Anzahl abgegebener Lösungen (etwas schwächer auch mit der Anzahl richtiger Lösungen, ohne Abb.). Die stärkste Korrelation ist auch hier der aktiv-Status. Zusammengefasst lässt sich sagen:

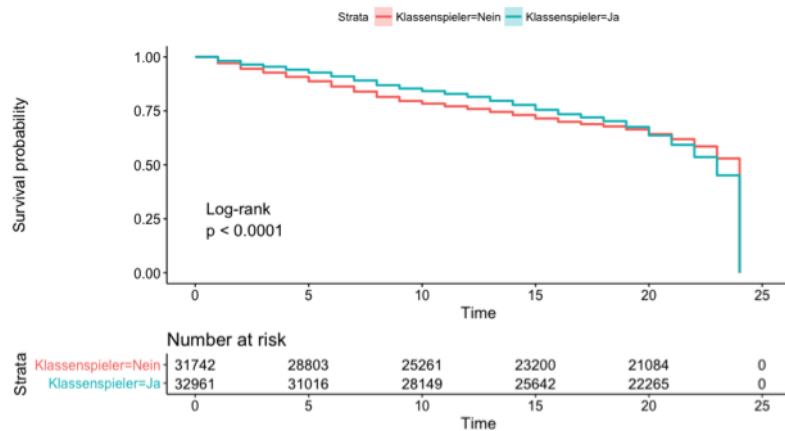
- Für „aktive“ Schüler\*innen, die am Gewinnspiel teilnehmen, ist zu erwarten, dass sie die meisten Aufgaben abgeben werden.
- Der Gruppeneffekt ist etwas stärker hinsichtlich der Anzahl abgegebener Lösungen als deren Richtigkeit.
- Zahlende Klassen sind „inklusive“. Sie enthalten aufgrund ihrer Größe (durchschnittlich ca. 16 Schüler\*innen gegenüber ca. 6) eine große Mischung an leistungsstärkeren und -schwächeren Schüler\*innen.

### Untersuchung der Persistenz

Vermutlich hängt die Wahrscheinlichkeit zu einem bestimmten Zeitpunkt auszusteigen vom bisherigen Abschneiden ab. Die Methode, die wir zur Untersuchung der Persistenz verwendet haben, ist der *Kaplan-Meier-Schätzer* (s. Kaplan & Meier, 1958). Dies ist eine Ereigniszeitanalyse (EZA) bzgl. der „Überlebenswahrscheinlichkeit“ (ÜW) der Individuen, in unserem Fall also die Wahrscheinlichkeit, dass ein\*e Spieler\*in weitere Antworten abgibt. Wir haben erst die zum Gewinnspiel angemeldeten Einzelspieler\*innen mit allen Klassenspieler\*innen verglichen. Dann haben

wir die Klassenspieler\*innen, die gezahlt haben, mit denen verglichen, die nicht gezahlt haben.

In Abb. 4 ist in rot der Graph der EZA für die Einzelspieler\*innen zu sehen, in blau der für die Klassenspieler\*innen. Ab der 2. Aufgabe haben Klassenspieler\*innen eine höhere ÜW, als Einzelspieler\*innen.



**Abb. 4:** Ereigniszeitanalyse Klassensp. vs. Einzelsp.

Dies deckt sich mit unseren Ergebnissen aus der OLS. Interessant wird es bei Aufgabe 20. Ab da haben Einzelspieler\*innen eine höhere ÜW. Wir vermuten, dass dies am Beginn der Weihnachtsferien liegt. Klassenspieler\*innen, die bis dahin Unterstützung in der Gruppe hatten, hatten diese dann nicht mehr. Die EZA bzgl. der Bezahlung (ohne Abb.) deckt sich ebenfalls mit dem Ergebnis der OLS: Die ÜW der Klassenspieler\*innen, die bezahlt haben, ist durchgehend größer, als die der Klassenspieler\*innen, die nicht bezahlt haben.

## Fazit und Ausblick

Das Spielen in einer Klasse scheint einen positiven Einfluss auf Leistung und Persistenz zu haben. Auch die Zahlung einer Klasse scheint zu einer größeren Beteiligung am Wettbewerb zu führen. Hier ist allerdings – aufgrund eines möglichen Selection Bias (z.B. könnte es sein, dass grundsätzlich „motiviertere“ bzw. größere Klassen auch eher bezahlen) – besondere Vorsicht geboten. Weitere Kontrollen, wie Untersuchungen von Klassen/Individuen über mehrere Jahre (Panel-Untersuchungen), wären notwendig. Um unsere Vermutung bzgl. der EZA und dem Beginn der Schulferien zu verifizieren, planen wir eine Aufschlüsselung nach Bundesländern (da der Beginn der Ferien unterschiedlich ist). Allgemein werden wir in näherer Zukunft Untersuchungen der MiA Daten bzgl. Bundesländern/Regionen und auch verschiedenen Schularten durchführen.

## Literatur

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association (JASA)*, 53 (282): 457–481.

Opfer, G. (2002). Numerische Mathematik für Anfänger. Eine Einführung für Mathematiker, Ingenieure und Informatiker. 4. Auflage. Vieweg, Braunschweig.