

Woche 06 Praktische Anwendungsbeispiele: Regression

Skript

Erarbeitet von
Katja Theune

Lernziele	1
Inhalt	2
Einstieg	2
Anwendungsbeispiel: Mietspiegel.....	2
Anwendungsbeispiel: Macht Geld glücklich?	3
Mehr Schein als Sein: Von Störchen und Babys	5
Abschluss	6
Weiterführendes Material	6
Quellen.....	6
Disclaimer.....	7

Lernziele

- Konkrete Anwendungsbeispiele der Regression in Praxis und Wissenschaft nennen
- Erläutern der Bedeutung der Einordnung von Ergebnissen

Inhalt

Einstieg

Wir haben uns doch alle schonmal bei der Wohnungssuche mit Mietpreisen auseinandergesetzt. In diesem Zusammenhang haben viele von euch bestimmt schon vom Mietspiegel gehört. Hier kommen häufig Regressionsverfahren zum Einsatz. Mieten sind also ein sehr lebensnahes Beispiel für die Anwendung von Regressionen aus der Praxis und unserem Alltag.

Oft beschäftigt uns auch unsere eigene Zufriedenheit und wie wir diese verbessern können. Die Zufriedenheitsforschung ist ein beliebtes Thema in der Wissenschaft und auch hier werden häufig Regressionsverfahren verwendet. Wir sehen also, dass man in Praxis und Wissenschaft gar nicht um das Verfahren der Regression herumkommt.

Anwendungsbeispiel: Mietspiegel

Die Mietpreisbremse ist ein großes Thema politischer Diskussionen geworden und einigen von euch bestimmt schon begegnet. So ist auch der sogenannte Mietspiegel in den Fokus gerückt. Er dient der Darstellung der ortsüblichen Vergleichsmiete. Sie gibt den üblichen Mietpreis für vergleichbare Wohnungen in einem bestimmten Gebiet an. Als Basis für einen qualifizierten Mietspiegel werden oft Regressionsverfahren verwendet. Hier betrachtet man vor allem die Zusammenhänge zwischen der Miethöhe und den einzelnen Wohnungsfeatures, die dann zusammen den Mietpreis bestimmen.

Bei einer Regression zur Mietpreisbestimmung ist unsere metrische Zielgröße dann z. B. die Netto-Kaltmiete. Zu berücksichtigende features sind nach Gesetz die Art, Größe, Ausstattung, Beschaffenheit und die Lage eines Gebäudes oder einer Wohnung.

Quelle [1]

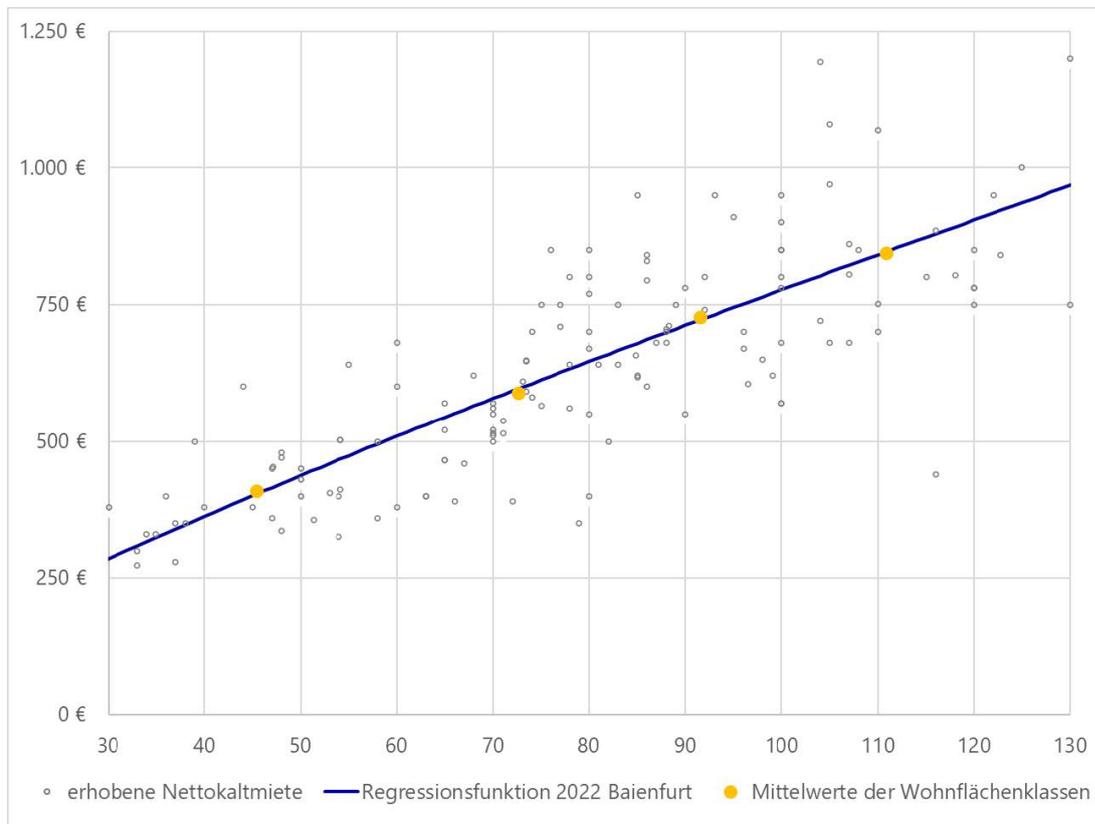
Einblendung verschiedene features und Zielgröße

Hier werden für die Mietspiegel häufig konkretere features wie der Wohnungstyp, die Wohnungsgröße, Ausstattungsmerkmale wie ein Balkon oder eine Badewanne, das Alter der Wohnung, vorgenommene Modernisierungen oder die Wohnlage verwendet. Ich denke, wir können uns alle vorstellen, dass diese features, und noch einige mehr, den Mietpreis beeinflussen.

Quelle [2]

Zur besseren Veranschaulichung hier mal das Ergebnis einer einfachen Regression mit der Netto-Kaltmiete als Zielvariable und nur der Wohnungsgröße als feature aus dem

Mietspiegelbericht für Ravensburg und Umgebung. Auf der horizontalen Achse ist die Wohnungsgröße in Quadratmetern und auf der vertikalen Achse die Netto-Kaltmiete abgetragen.



Quelle [3]

Wir sehen, dass tendenziell mit steigender Wohnungsgröße auch die Nettokaltmiete steigt und dass man mit einem Regressionsverfahren eine ansteigende Kurve bzw. Linie gut in die Punktwolke anpassen kann. Da wir bei einer multiplen Regression ja mehrere features gleichzeitig berücksichtigen, kann man so insbesondere dann auch für eine Wohnung mit bestimmten Merkmalen den voraussichtlichen Mietpreis prognostizieren.

Anwendungsbeispiel: Macht Geld glücklich?

Kommen wir jetzt zu einem anderen Anwendungsbeispiel. Macht Geld glücklich? Das ist eine Menschheitsfrage, mit der sich sowohl privat als auch in der Wissenschaft schon lange und ausgiebig beschäftigt wird.

Einigen von euch ist der sogenannte „World Happiness Report“ vielleicht ein Begriff. In ihm wird die Lebenszufriedenheit verschiedener Länder miteinander verglichen und in Zusammenhang zu verschiedenen Begebenheiten eines Landes gebracht.

Quelle [4]

Unsere Zielvariable ist hier also ein Wert für Zufriedenheit in einem Land. Er wird auf einer Skala von 0 bis 10 gemessen. 0 ist hier der schlechteste Wert und 10 der beste. Da ein Durchschnitt über alle beteiligten Personen eines Landes gebildet wird, haben wir hier eine metrische Zielvariable.

Einblendung verschiedene features und Zielgröße

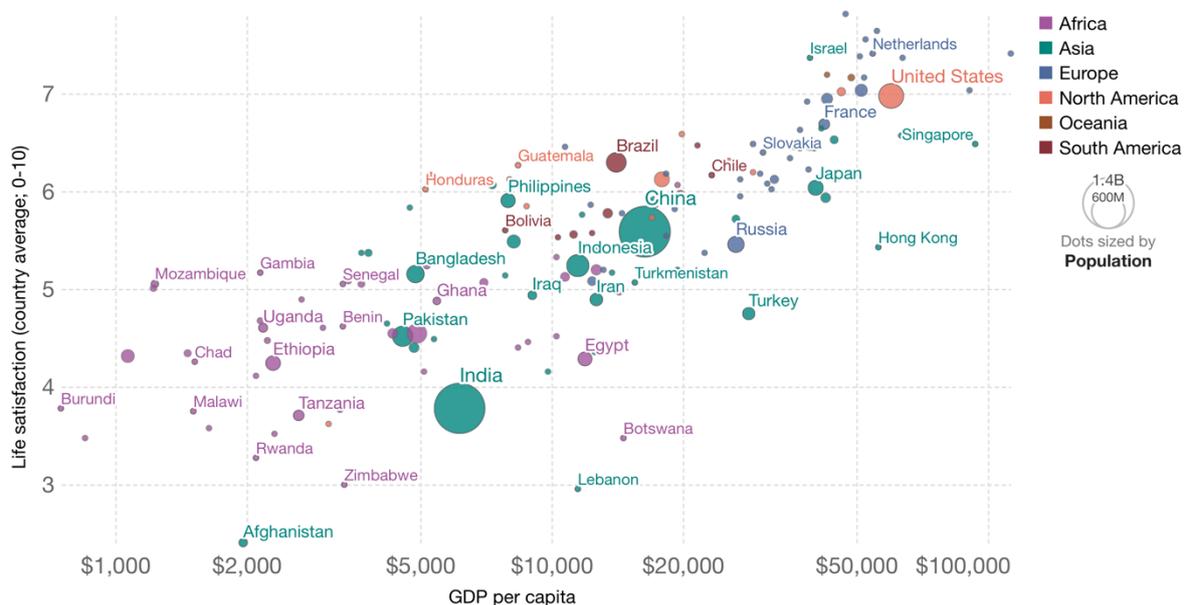
Als features, die die Lebenszufriedenheit bestimmen, werden z. B. die Lebenserwartung, der Grad an Freiheit in einem Land oder die Abwesenheit von Korruption betrachtet. Aber eben auch der Reichtum eines Landes bzw. hier das Bruttoinlandsprodukt pro Person. Im Happiness-Report wird mit einer multiplen linearen Regression die Zufriedenheit mit diesen features in Zusammenhang gebracht. Es zeigt sich z. B., dass mit steigendem Bruttoinlandsprodukt pro Person auch die Lebenszufriedenheit steigt.

Das lässt sich auch hier im Bild gut erkennen. Auf der horizontalen Achse ist das Bruttoinlandsprodukt pro Person und auf der vertikalen Achse die Lebenszufriedenheit abgetragen. Tendenziell haben Länder mit einem höheren Bruttoinlandsprodukt pro Person auch eine höhere Lebenszufriedenheit.

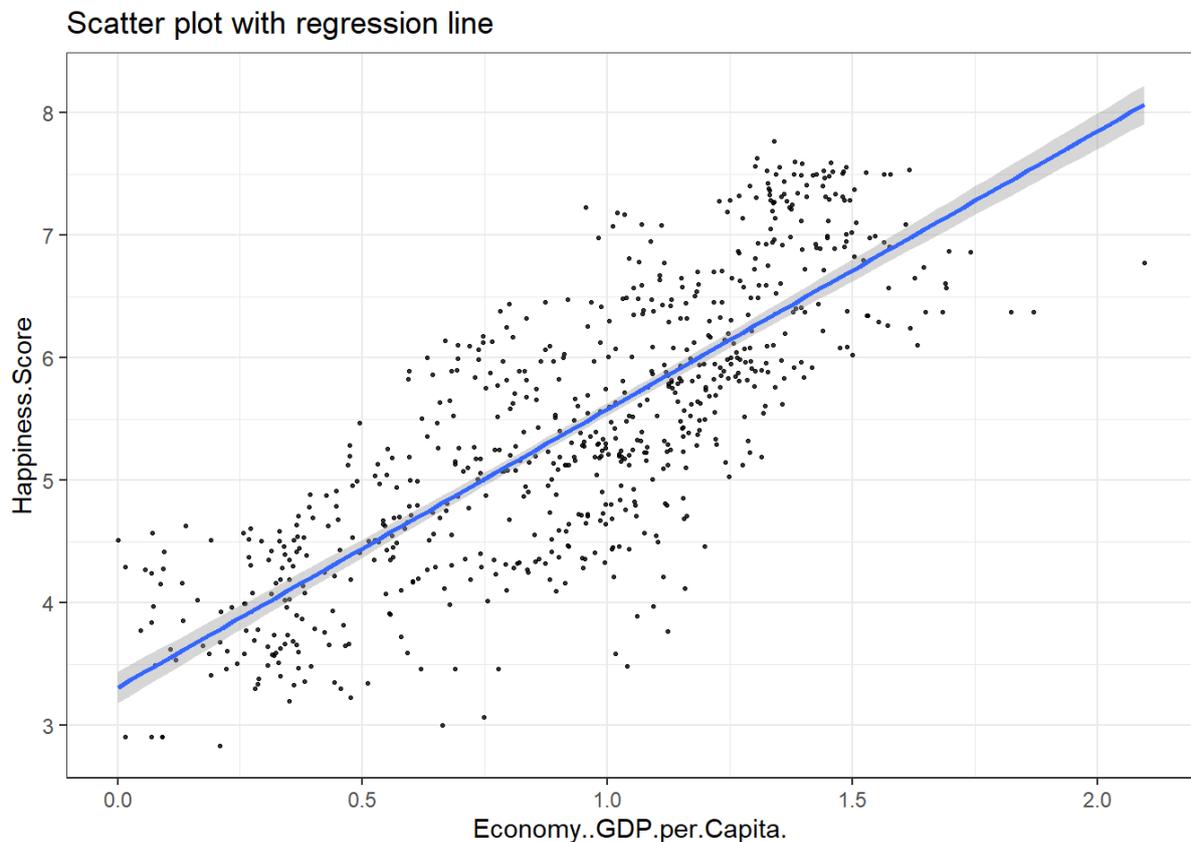
Self-reported life satisfaction vs GDP per capita, 2020

Our World in Data

The vertical axis shows the national average of the self-reported life satisfaction on a scale ranging from 0-10, where 10 is the highest possible life satisfaction. The horizontal axis shows GDP per capita adjusted for inflation and cross-country price differences.



Wir sehen uns hier jetzt auch nochmal das Ergebnis einer einfachen linearen Regression auf Basis dieser Daten an. Wir sehen, dass sich mittels dieses Regressionsverfahrens eine ansteigende gerade Linie gut in die Punktwolke einfügen lässt. Sie gibt den Zusammenhang von Bruttoinlandsprodukt und Zufriedenheit auch anschaulich wieder.



Quelle [6]

Wir haben also gesehen, dass reichere Länder bzw. Menschen glücklicher erscheinen. Aber ... könnte das vielleicht auch andersherum sein? Könnten glücklichere Menschen nicht auch reicher sein? Vielleicht, weil ihre Produktivität höher ist als bei unglücklichen Menschen? Hier kommen wir also zu einem Punkt, der uns bei Anwendungen von Regressionsverfahren und allgemein der Untersuchung von Zusammenhängen immer wieder beschäftigt. Ist das eine überhaupt Ursache und das andere Wirkung? Und falls ja, was wäre dann was?

Mehr Schein als Sein: Von Störchen und Babys

Im Anschluss an unser Beispiel von gerade möchte ich noch kurz an einem eher lustigen Beispiel erläutern, dass ein Zusammenhang in den Daten nicht unbedingt auch wissenschaftlich sinnvoll ist, denn gefundene Zusammenhänge zwischen features oder zwischen feature und Zielgröße sagen nicht unbedingt auch etwas über Ursache und Wirkung aus. Der gefundene Zusammenhang sagt uns nichts darüber, welche und ob überhaupt eine Sache die Ursache – man sagt auch kausal – für die andere ist. Beliebt ist hier das Beispiel zu der Legende, dass Störche die Babys bringen. Es konnte mit Daten über die Geburtenrate in verschiedenen Ländern und die Anzahl der dort brütenden Störche

mittels einer Regression gezeigt werden, dass mit steigender Anzahl an Störchenpaaren auch die Geburtenrate steigt. Das ist natürlich nicht wirklich zu ernst zu nehmen. Aber solche sogenannten Scheinzusammenhänge oder auch Scheinkausalitäten begegnen uns sehr häufig.

Quelle [7]

Bei diesen Scheinkausalitäten gibt es meist ein weiteres feature, was die beiden anderen betrachteten features bzw. feature und Zielgröße gleichermaßen beeinflusst. Bei unserem Beispiel geht man davon aus, dass dieses weitere feature die Region ist. So haben ländliche Regionen meist sowohl mehr Störche als auch eine höhere Geburtenrate. Das erklärt dann auch den Zusammenhang zwischen der Anzahl der Störche und der Geburtenrate.

Abschluss

Wir haben jetzt Anwendungsbeispiele von Regressionsverfahren in Praxis und Wissenschaft kennengelernt. Zudem wissen wir, dass es wichtig ist, unsere Ergebnisse, nicht nur von Regressionsverfahren, zu interpretieren und einzuordnen.

Weiterführendes Material

Videos/Kurse:

Von der Datenanalyse zur Datengeschichte – Datenanalyseergebnisse adressatengerecht kommunizieren, Datenanalyse II, Datenanalyse Theorie III: Korrelation und Kausalität.

<https://learn.ki-campus.org/courses/Datenanalyse-unibi2021/items/tnOSZxcf3lvv7DTq3GoPA>

Quellen

Quelle [1] § 558 Absatz 2 Satz 1 BGB

Quelle [2] Stadt Gelsenkirchen (2022). *Mietspiegel Gelsenkirchen 2022 – Methodenbericht*.

Quelle [3] ALP Institut für Wohnen und Stadtentwicklung (2022). *Dokumentation der Mietspiegel Ravensburg, Weingarten, Baienfurt, Baidnt und Berg 2022*.

- Quelle [4] Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (Hrsg.). (2022). *World Happiness Report 2022*. New York: Sustainable Development Solutions Network.
- Quelle [5] Our World in Data, Self-reported life satisfaction vs GDP per capita, 2020
<https://ourworldindata.org/grapher/gdp-vs-happiness>
- Quelle [6] RPubs, Analysis of World Happiness
https://rpubs.com/kyle_chen/WQD7004 *Analysis of World Happiness*
- Quelle [7] Matthews, R. (2000). Storks deliver babies ($p= 0.008$). *Teaching Statistics*, 22(2), 36-38.

Disclaimer

Transkript zu dem Video „Woche 06 Praktische Anwendungsbeispiele: Regression“, Katja Theune.

Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY](https://creativecommons.org/licenses/by/4.0/) 4.0 veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.