

Technische Universität Dortmund

Wintersemester 2025/26

Wissenschaftliches Arbeiten: Bericht über Titanic-Datensatz

Einflussfaktoren auf die Überlebenschancen beim Untergang der Titanic

Verfasser:

Lukas König

Elaha Bahir

John Wilhelm

Annika Homm

06.02.2026

1 Einleitung

Der Untergang der Titanic ist ein weltbekanntes Ereignis in der Geschichte der Menschheit. Es ist Gegenstand zahlreicher historischer und statistischer Forschung. In diesem Bericht soll es darum gehen, durch eine explorative Datenanalyse signifikante Merkmale (wie Geschlecht, Alter, Klasse) der Passagiere auf der Titanic zu identifizieren und Zusammenhänge zum Überleben zu finden.

2 Daten und Methoden

- **Daten:** Zur Datenanalyse wurde der Datensatz titanic.csv und die Pakete "tidyverse" und "ggplot2" verwendet. Vor der Datenbereinigung wies die Variable *Age* 177 NA-Werte (ca. 19,86%) auf, *Embarked* besaß zwei leere Zellen (ca. 0,22%) und *Cabin* hatte 687 leere Zellen, was einer Quote von ca. 77,10% entspricht.
- **Methodik:** Leere Zeichenketten wurden als NA kodiert und nicht für die Analyse relevante Variablen (wie *PassengerID*, *Name*, *Ticket*, *Cabin*) wurden nach Merkmalsextraktion entfernt.
 - **Age (Alter in Jahren beim Untergang):** Fehlende Werte wurden durch das arithmetische Mittel der jeweiligen Titelgruppe ersetzt, um die demografische Struktur beizubehalten und keinen Bias zu kreieren.
 - **Embarked (Zustiegshafen):** Zwei fehlende Einträge wurden durch den Modus "SSouthhampton" (S) ersetzt.
 - **Cabin (Kabinennummer):** Aus der Kabinennummer wurden das **Deck** (*Deck*) und die **Schiffsseite** (*Side*) mittels Modulo-Operation extrahiert.
 - *Hinweis: Aufgrund der hohen NA-Quote der Variablen Deck und Side und da sie durch Passagiere der dritten Klasse überrepräsentiert (ca. 69,72%) ist, ist eine Analyse dieser Variablen verzerrt (Selection Bias). Im weiteren Verlauf des Berichts wird daher der Fokus auf die robusteren Variablen wie Geschlecht, Klasse und Alter gelegt.*
- **Datentransformation:** Die Variablen *Sex*, *Survived*, *Embarked*, *Side* und *Deck* wurden als Faktoren kodiert. Die Passagierklasse (*Pclass*) wurde als **ordered factor** definiert um die Rangfolge abzubilden.

3 Ergebnisse

3.1 Zusammenhang zwischen kategorialen Merkmalen und Überlebensstatus

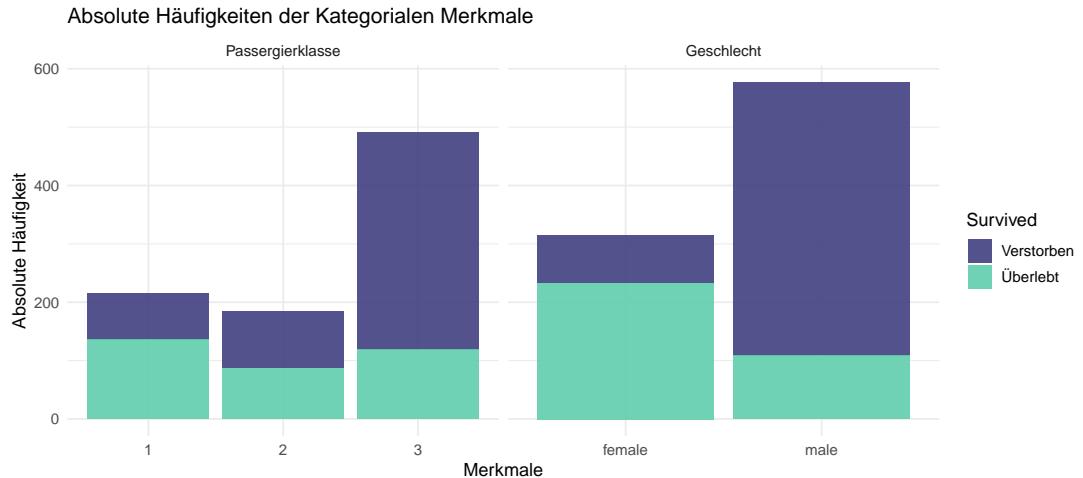


Abbildung 1: Abbildung 1: Absolute Häufigkeiten kategorialer Daten gruppiert nach Überlebensstatus

Die Demografie der Passagiere am Bord der Titanic bestand zum Großteil aus Männern und der Passagierklasse 3. Mehr als die Hälfte, ca. 61,61% , haben den Untergang der Titanic nicht überlebt. In der ersten Passagierklasse, hat die Mehrheit überlebt, in der zweiten Passagierklasse war es fast ausgeglichen und in der dritten Klasse ist die große Mehrheit gestorben. Interessanterweise überlebten in jeder Klasse fast gleich viele, doch der Anteil der Menschen in der dritten Klasse ist weitaus größer als in den anderen zwei Klassen, die von der Größe her vergleichbar sind, wobei Klasse 1 noch etwas mehr vertreten ist. Ein weiterer großer Unterschied besteht zwischen den überlebten Frauen und Männern, denn Unter den Überlebenden gab es ca. doppelt so viele Frauen wie Männer. Der Anteil der Überlebenden bestand aus ca. 68,12% aus Frauen. Dem gegenüber gab es fast doppelt so viele Männer an Bord als Frauen. Unter den Männern überlebten nur ca. 31,87%.

3.2 Zusammenhang zwischen Ticketpreisen und Überleben

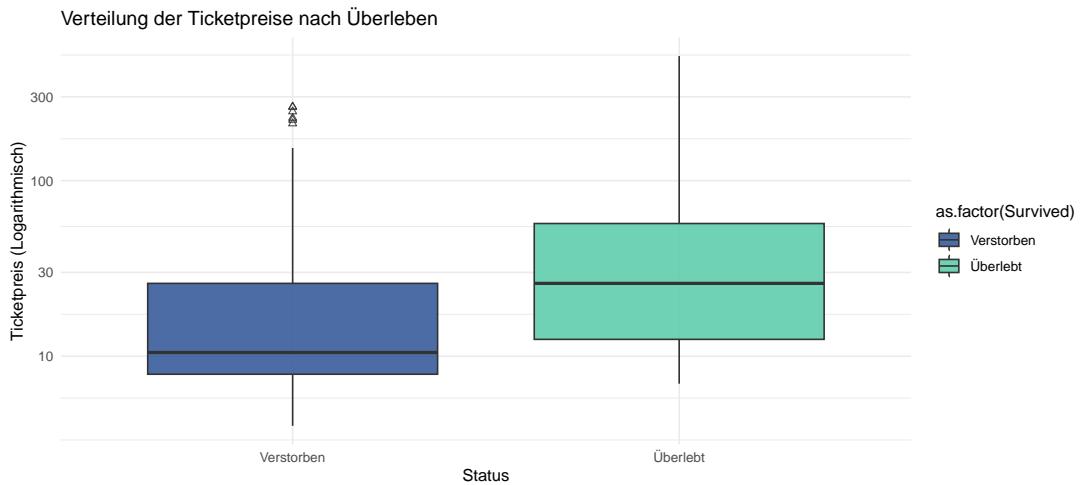


Abbildung 2: Abbildung 2: Boxplot der Ticketpreise gruppiert nach Überlebensstatus

Es gibt bei den Verstorbenen eine deutliche Tendenz nach unten, der Median orientiert sich an das untere Quartil und beide Quartile liegen unter den Quartilen der Überlebenden. Zudem gibt es bei den Verstorbenen auch viele Ausreißer nach oben, was darauf hindeutet, dass vereinzelt auch Passagiere mit teureren Tickets verstorben sind, allerdings verhältnismäßig wenige. Bei den Überlebten liegt der mediane Ticketpreis knapp unter 30 Währungseinheiten und somit fast 3 mal teurer als der mediane Ticketpreis der Verstorbenen (ca. 10 Währungseinheiten). Der Median liegt zudem mittig und es gibt keine Ausreißer.

3.3 Zusammenhang zwischen Familiengröße und Überleben

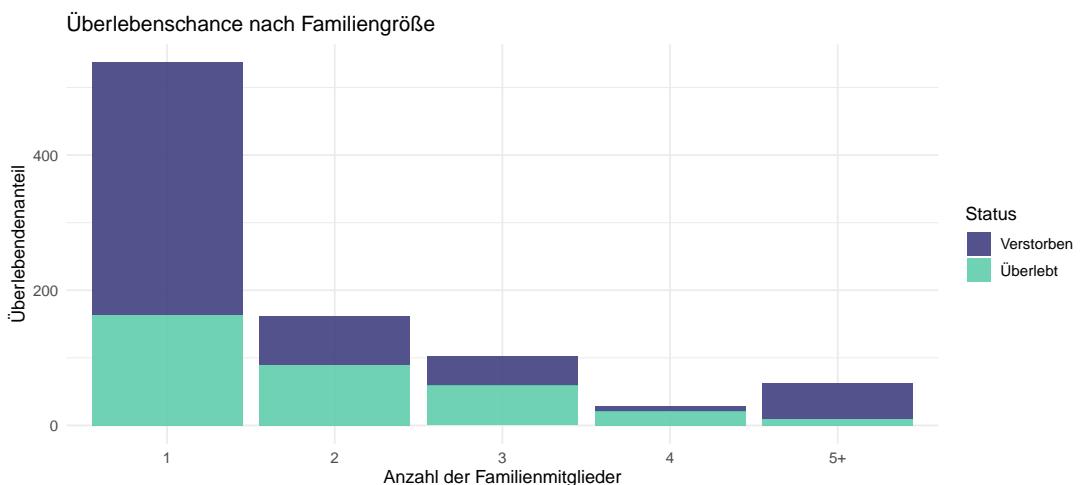


Abbildung 3: Abbildung 3: Balkendiagramm der Familiengröße gruppiert nach Überlebensstatus

bei der Mehrheit der Passagiere handelte es sich um Alleinreisende, wobei der Großteil von ihnen nicht überlebte. Auch Familien ab einer Größe von fünf und höher sind zu 85,10% verstorben, wobei diese eher selten vertreten waren.

3.4 Zusammenhang zwischen Alter und Überleben

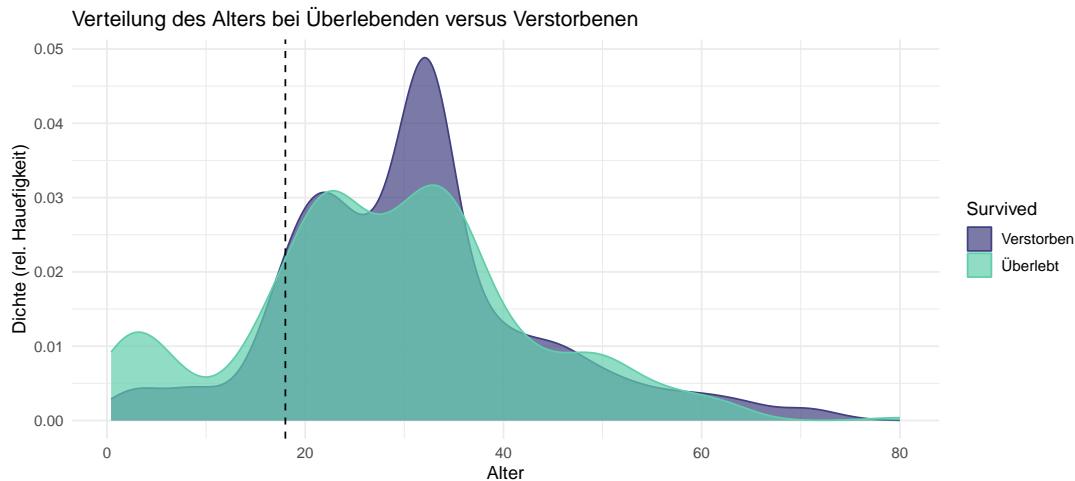


Abbildung 4: Abbildung 4: Dichteverteilungsfunktion des Alters gruppiert nach Überlebensstatus

Sowohl unter den Überlebenden, als auch den Verstorbenen liegt das Alter meist zwischen ca. 16-40 Jahren. Auffällig ist das lokale Maximum bei den Überlebenden zwischen etwas 0 und 10 Jahren, welcher bei den verstorbenen nicht vorhanden ist. Dies deutet darauf hin, dass Babys und kleine Kinder deutlich häufiger überlebt haben. Ein zweites lokales Maximum befindet sich bei den Verstorbenen von etwa 30 Jahren gibt, welcher methodisch bedingt (Datenimputation der Mittelwerte) und repräsentiert keine natürliche Häufung des Alters. Da die Datenimputation nach dem Mittelwert der jeweiligen Titel-kategorie erfolgt ist, ist die Interpretation, dass Kinder (z.B. mit dem Titel Master) häufiger überleben nicht gefährdet.

4 Fazit

Zusammenfassend lässt sich sagen, dass überwiegend Frauen überlebt haben, und das obwohl sie demografisch weniger an Bord vertreten waren. Zusätzlich haben auch Kinder deutlich häufiger überlebt, was auf das Prinzip von "Frauen und Kinder zuerst" hindeutet. Dies könnte auch eine Erklärung für die Ausreißer in Abbildung 2 sein, da es sich um reiche Männer handeln könnte, die trotz erster Klasse an Bord blieben. Große Familien haben häufig nicht überlebt, vermutlich weil sie nicht alle auf ein Rettungsboot gepasst haben und daher zurückbleiben mussten. Auch Alleinreisende sind zu einem großen Anteil

verstorben, eventuell weil ca. 60,33% von ihnen die dritte Passagierklasse besaß oder weil ca. 76,53% von ihnen männlich waren. Letztendlich lässt sich sagen, dass vermutlich sowohl die Moral (Frauen und Kinder zuerst), als auch das Geld die Passagiere gerettet hat.