

Kurzbericht: Analyse der Titanic-Daten

Meico Bastian Heil, Saskia Lapkowsky, Heetae Kim, Leticia Kwanga, and Sergei Nazarov
Technische Universität Dortmund
(Dated: 08.02.2026)

In dieser Arbeit wurde der Titanic-Datensatz im Rahmen einer Gruppenarbeit mithilfe von GitHub und R analysiert. Die Daten wurden zunächst bereinigt und sinnvoll transformiert, um fehlende Werte zu imputieren und neue Variablen zu erstellen. Anschließend wurden deskriptive und bivariate Analysen durchgeführt, um Faktoren zu identifizieren, die das Überleben der Passagiere beeinflussen, insbesondere Geschlecht, Passagierklasse und Ticketpreis.

I. EINLEITUNG

Ziel dieser Arbeit ist es, praktische Erfahrungen im gemeinsamen Arbeiten mit GitHub zu sammeln. Dazu werden in Gruppen einfache analytische Aufgaben an den Titanic-Daten bearbeitet, wobei LaTeX und R zur Dokumentation und Auswertung genutzt werden.

II. DATEN

Die Daten stammen aus der Titanic-Datenbank und wurden für die Analyse bereinigt und vereinheitlicht. Die wichtigsten Variablen nach der Aufbereitung sind in Tabelle I zusammengefasst.

Spalte	Bedeutung / Inhalt	Format
Survived	Überlebt? Ja (1) / Nein (0)	factor
Pclass	Passagierklasse (1 = 1., 2 = 2., 3 = 3.)	ordered factor
Sex	Geschlecht des Passagiers	factor
Age	Alter in Jahren	numeric
SibSp	Anzahl Geschwister / Ehepartner an Bord	numeric
Parch	Anzahl Eltern / Kinder an Bord	numeric
Fare	Ticketpreis	numeric
Embarked	Einschiffungshafen (C / Q / S)	factor
Anrede	Anrede / Titel aus dem Namen	factor

Tabelle I. Struktur der Titanic-Daten

Nach der Bereinigung wurden fehlende Werte behandelt und die Variablen in geeignete Formate konvertiert, sodass sie für die Analyse in R genutzt werden können.

III. METHODEN

Die Analyse der Titanic-Daten erfolgt im Rahmen einer explorativen deskriptiven Auswertung:

1. Deskriptive Beschreibung zentraler metrischer Variablen (Alter, Ticketpreis sowie Anzahl mitreisender Angehöriger).
2. Deskriptive Analyse kategorialer Merkmale (Überlebensstatus, Geschlecht, Passagierklasse, Einschiffungshafen).

3. Untersuchung des Zusammenhangs zwischen Überlebensstatus und ausgewählten kategorialen Variablen (Geschlecht, Passagierklasse).
4. Vergleich der Verteilungen von Alter und Ticketpreis zwischen Überlebenden und Nicht-Überlebenden.
5. Visuelle Analyse des Zusammenspiels von Passagierklasse, Einschiffungshafen, Überlebensstatus und Geschlecht.

IV. ERGEBNISSE

1. Deskriptive Statistik der metrischen Variablen

Die deskriptive Analyse der metrischen Variablen ergibt:

Variable	Mittelwert	Median	SD	Min	Max
Age	29.37	30.00	13.25	0.42	80.00
Ticketpreis	32.20	14.45	49.69	0.00	512.33

Das durchschnittliche Alter der Passagiere beträgt etwa 29 Jahre.

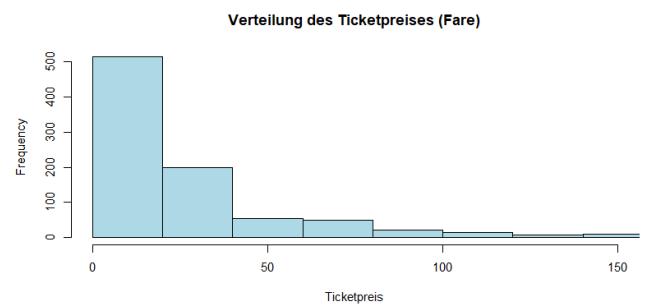


Abbildung 1. Histogramm der Ticketpreise (Fare)

Der Ticketpreis weist eine rechtsschiefe Verteilung auf. Der Mittelwert liegt bei 32, der Median nur bei 14.

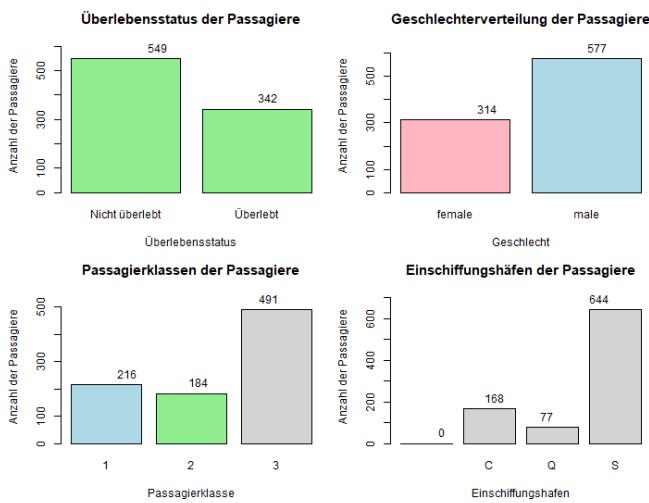


Abbildung 2. Übersicht der Passagierkategorien

2. Deskriptive Statistik der kategorialen Variablen

- **Überlebensstatus:** Die Mehrheit der Passagiere hat die Katastrophe nicht überlebt.

Kategorie	Anzahl	Anteil (%)
Nicht überlebt	549	61.6
Überlebt	342	38.4

- **Geschlecht:** Deutlich mehr männliche als weibliche Passagiere an Bord.

Geschlecht	Anzahl	Anteil (%)
weiblich	314	35.2
männlich	577	64.8

- **Passagierklasse:** Die Mehrheit reiste in der dritten Klasse.

Klasse	Anzahl	Anteil (%)
1. Klasse	216	24.2
2. Klasse	184	20.7
3. Klasse	491	55.1

- **Einschiffungshafen:** Die meisten Passagiere stiegen in Southampton ein.

Hafen	Anzahl	Anteil (%)
C	168	18.9
Q	77	8.6
S	644	72.3

Abb. 2 zeigt eine allgemeine Übersicht

3. Bivariate deskriptive Statistik für zwei kategoriale Variablen

Überlebensstatus nach Geschlecht: Unter den Überlebenden ist der Anteil der Frauen deutlich höher; unter den Nicht-Überlebenden überwiegen die Männer.

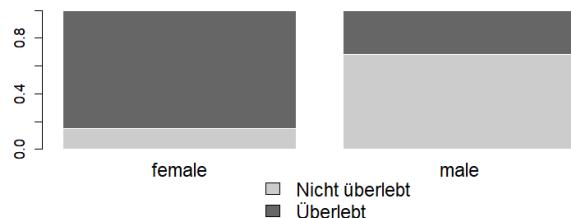


Abbildung 3. Überlebensstatus nach Geschlecht (Anteil)

Überlebensstatus nach Passagierklasse: In der 1. Klasse überleben deutlich mehr Passagiere, während die 3. Klasse den höchsten Anteil Nicht-Überlebender aufweist.

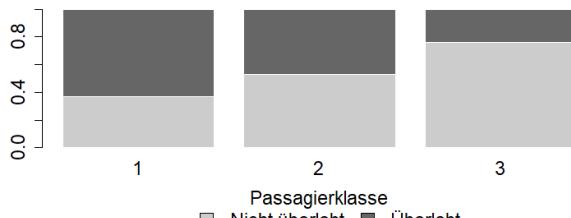


Abbildung 4. Überlebensstatus nach Passagierklasse (Anteil)

Passagierklasse nach Einschiffungshafen: Passagiere aus Cherbourg gehörten häufiger der 1. Klasse an, was auf einen höheren sozioökonomischen Status hindeutet (Abb. 5).

Überlebensraten nach Anrede:

Anrede	Nicht überlebt	Überlebt
Mr.	436	81
Miss.	55	127
Mrs.	26	99
Master.	17	23
Andere	18	12

Die Anrede zeigt deutliche Unterschiede im Überleben: Miss und Mrs. sind mit höheren Überlebensraten



Abbildung 5. Passagierklasse nach Einschiffungshafen

verbunden, während die Mehrheit der Mr.-Passagiere nicht überlebte.

4. Bivariate deskriptive Statistik

Unterscheidet sich das Alter und der Ticketpreis der Passagiere in Abhängigkeit vom Überlebensstatus?

Alter: Überlebende sind im Median etwas jünger als Nicht-Überlebende.

Ticketpreis: Überlebende haben im Durchschnitt höhere Ticketpreise gezahlt

Dies deutet darauf hin, dass sowohl Alter als auch finanzieller Status einen Einfluss auf die Überlebenswahrscheinlichkeit hatten.

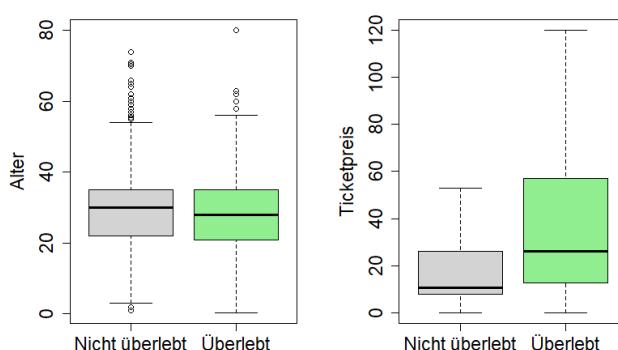


Abbildung 6. Alter und Ticketpreis nach Überlebensstatus

5. Visualisierung von mehreren kategorialen Variablen

Überlebensstatus nach Passagierklasse, Einschiffungshafen und Geschlecht: Frauen weisen in allen Klassen und Häfen höhere Überlebensanteile auf. Der Einfluss der Passagierklasse bleibt weiterhin bestehen.

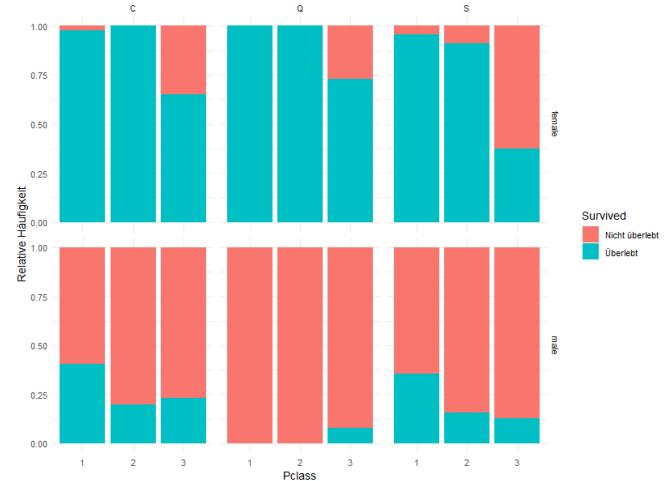


Abbildung 7. Überlebensstatus nach Passagierklasse, Einschiffungshafen und Geschlecht

6. Weitere Analyse

Unterscheidet sich das Alter der Passagiere zwischen den drei Klassen?

Passagiere der ersten Klasse sind im Median älter als die der dritten Klasse. Die Streuung ist ähnlich, aber in der dritten Klasse gibt es einige sehr junge Passagiere.

V. FAZIT

Alter, Geschlecht und Passagierklasse zeigen deutliche Zusammenhänge mit der Überlebenswahrscheinlichkeit. Frauen und Passagiere höherer Klassen überlebten häufiger, während das Alter einen moderaten Einfluss zeigt. Diese Faktoren sind damit die wichtigsten Prädiktoren für das Überleben auf der Titanic.