

Technische Universität Dortmund

Fakultät Statistik

Wintersemester 2022/23

Fallstudien I

Projekt 1:

Dozent: Prof. Dr. Guido Knapp

Verfasser:

Daniel Sipek

Gruppenmitglieder:

Louisa Poggel

Julia Keiter

ECaroline Baer

27.10.2022

Inhaltsverzeichnis

1	Einleitung	1
2	Problemstellung	1
3	Statistische Methoden	3
4	Statistische Auswertung	4
4.1	Ergebnisse	4
4.2	Interpretation	4
5	Zusammenfassung der Ergebnisse	4
5.1	Zusammenfassung	4
5.2	Diskussion	4
5.3	Ausblick	4
6	Literaturverzeichnis	4
7	Anhang	4

1 Einleitung

2 Problemstellung

Bei dem Datensatz handelt es sich um eine Beobachtungsstudie zur Wirksamkeit eines Medikaments für Personen mit chronischer kongestiver Herzinsuffizienz als zusätzliche Therapie zur Standardbehandlung. Die Studie wird doppelblind und placebo-kontrolliert durchgeführt. Die Daten wurden zum Zeitpunkt des Screenings erstellt. In dem Datensatz sind 15 Variablen vorhanden. Insgesamt haben 200 Patienten bei der Studie teilgenommen, welche multinational und multizentrisch durchgeführt wird. Deshalb werden die Variablen *Land* und *Zentrum* in den Datensatz aufgenommen. In diesem Fall hat die Variable *Land* bei allen 200 Patienten den Wert 10 und entspricht damit Deutschland. Um jeden Patient eindeutig zu identifizieren, gibt es die Variablen *Screeningnummer* und *Patientennummer*. Dabei gilt die *Patientennummer* als Identifikation für alle 200 Patienten und *Screeningnummer* als Identifikation für das jeweilige Zentrum. Die Patienten werden in zwei Medikationsgruppen geteilt, wo die eine Gruppe das Medikament zur Behandlung erhält und die andere Gruppe eine Placebo-Behandlung bekommen. Die Variable *Medikationsgruppe* gibt an, in welcher Gruppe sich die Patienten befinden. Zusätzlich wird angegeben, ob die Patienten zur Gruppe der *Safety-Analysis Population*, zur Gruppe der *Intention-To-Treat Population* und/oder zur Gruppe der *Per-Protocol-Analysis Population* gehören. In dem Datensatz werden ebenfalls demographische Angaben zum Patienten gemacht. Diese beinhalten das *Geschlecht*, die *Größe* in Zentimeter, das *Gewicht* in Kilogramm, und das *Alter* in Jahren. Außerdem wird der *Body-Mass-Index* mitaufgeführt mit der Einheit Kilogramm pro Quadratmeter. Der *Body-Mass-Index* stellt eine Abschätzung des Körperfettanteils dar. Es wird aus dem Quotienten aus Körpergewicht und Körpergröße zum Quadrat gebildet. Bei einem Wert von unter 18.5 gilt man als untergewichtig und ab einem Wert von 25 zählt man zu der Kategorie übergewichtig. Die Dauer der bestehenden Herzinsuffizienz (kurz *Dauer Herzinsuffizienz*) wird in Monaten im Datensatz aufgeführt und, ob ein Herzinfarkt sich ereignet hat, wird in der Variable *Herzinfarkt* aufgeführt.

Die Variablen *Zentrum*, *Screeningnummer* und *Patientennummer* sind numerisch, wobei es nur natürliche Zahlen beinhalten. Die Variable *Medikationsgruppe* ist dichotom mit den Ausprägungen 1 und 2, wobei 1 für die Gruppe steht, die die Behandlung mit dem Medikament erhalten und 2 die Gruppe ist, die die Placebo-Behandlung bekommen. Weitere dichotome Variablen sind *Safety-Analysis population*, *Intention-To-Treat*

Population, *Per-Protocol-Analysis Population* sowie *Geschlecht* und *Herzinfarkt*. Die drei Variablen *Safety-Analysis population*, *Intention-To-Treat Population* und *Per-Protocol-Analysis Population* sind mit 1 und 0 codiert. Dabei steht die 1 dafür, dass die Patienten zur jeweiligen Population gehören und die 0 dafür, dass sie nicht dazugehören. Bei dem *Geschlecht* bedeutet eine 1, dass der Patient männlich ist eine 2, dass der Patient weiblich ist. Bei dem *Herzinfarkt* deutet eine 1 darauf hin, dass dieser sich ereignet hat, wobei eine 2 bedeutet, dass kein Herzinfarkt eingetreten ist. Die Variablen *Größe*, *Gewicht*, *Alter*, *Body-Mass-Index* und *Dauer Herzinsuffizienz* sind kardinalskaliert. Dabei werden die *Größe* und *Gewicht* in dem Datensatz in ganzen Zahlen und *Alter*, *Body-Mass-Index* und *Dauer Herzinsuffizienz* in Dezimalzahlen angegeben. Das *Alter* und der *Body-Mass-Index* haben Fünf Nachkommastellen und die *Dauer Herzinsuffizienz* hat Sieben Nachkommastellen. Die Messgenauigkeit bei diesen Drei Variablen ist daher höher als bei den Variablen *Größe* und *Gewicht*, obwohl alle Fünf Variablen stetig sind. Insgesamt sind in dem Datensatz 3000 Daten, bei welchem 86 davon ein NA sind. Bei 36 Patienten befinden sich NA's in der Variable *Patientennummer* und *Medikationsgruppe*. 8 NA's sind in der Variable *Dauer Herzinsuffizienz* und die restlichen 6 befinden sich alle in dem Datensatz bei dem 179. Patienten. Bei diesem Patienten fehlen deshalb Werte zu den Variablen *Patientennummer*, *Medikationsgruppe*, *Geschlecht*, *Größe*, *Gewicht*, *Alter*, *Body-Mass-Index*, *Dauer Herzinsuffizienz* und *Herzinfarkt*.

Die interessierenden Variablen für dieses Projekt sind die demographischen Daten *Geschlecht*, *Größe*, *Gewicht* und *Alter*, sowie den *Body-Mass-Index*, die *Dauer Herzinsuffizienz* und den *Herzinfarkt*. Diese werden mithilfe Kenngrößen und grafischen Verfahren geeignet dargestellt. Dabei werden alle gescreenten Patienten betrachtet. Dies betrifft alle Patienten in dem Datensatz mit Ausnahme dem 179. Patient, zu welchem sich keine Daten finden lassen. Demnach reduziert sich der Stichprobenumfang auf 199 Teilnehmer. Als nächstes werden die randomisierten Personen betrachtet. Dies betrifft alle Patienten die zu einer der beiden Medikationsgruppen zugeordnet wurden. Daher entfallen die 35 weitere Teilnehmer, die keiner Gruppe angehören. Somit verkleinert sich der Stichprobenumfang auf 164 Teilnehmer. Es wird, mithilfe deskriptiver Maße und Grafiken, überprüft wie ähnlich die beiden Medikationsgruppen sind. Abschließend soll gewertet werden, wie gut die Randomisierung ereignet ist.

3 Statistische Methoden

$$\bar{x} = \sum_{i=1}^n x_i \quad i = \{1, \dots, n\}$$

Merkmal X ist mindestens ordinalskaliert n = Anzahl der Einzelbeobachtungen von X

$$\tilde{x}_{0,5} = \begin{cases} x_{n+1/2}, & \text{falls n ungerade} \\ 1/2(x_{n/2} + x_{(n/2)+1}), & \text{falls n gerade.} \end{cases}$$

Merkmal X sei mindestens ordinalskaliert n = Anzahl der Einzelbeobachtungen von X, $0 < p < 1$

$$\tilde{x}_p = \begin{cases} x_{[np+1]}, & \text{falls n ungerade} \\ 1/2(x_{[np]} + x_{[np+1]}), & \text{falls n gerade.} \end{cases}$$

Modalwert:

$$\bar{x}_{mod} = \max(h(x_i))$$

Spannweite:

$$x_{(n)} - x_{(1)}$$

Interquartilsabstand Differenz zwischen dem dritten und ersten Quartil:

$$Q_{0.25} = \tilde{x}_{0.75} - \tilde{x}_{0.25}$$

Varianz Merkmal X ist intervallskaliert:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$
$$s = +\sqrt{s^2}$$

Momentkoeffizient der Schiefe:

$$SK_M = \frac{1/n \sum_{i=1}^n (x_i - \bar{x})^3}{(1/n \sum_{i=1}^n (x_i - \bar{x})^2)^{3/2}}$$

Wölbung:

$$W_M = \frac{1/n \sum_{i=1}^n (x_i - \bar{x})^4}{(1/n \sum_{i=1}^n (x_i - \bar{x})^2)^2}$$

4 Statistische Auswertung

4.1 Ergebnisse

4.2 Interpretation

5 Zusammenfassung der Ergebnisse

5.1 Zusammenfassung

5.2 Diskussion

5.3 Ausblick

6 Literaturverzeichnis

7 Anhang