

Technische Universität Dortmund

Fakultät Statistik

Wintersemester 2022/23

Fallstudien I: Projekt 2

Deskriptive Analyse der Demografie einer klinischen Studie

Dozenten:

Prof. Dr. Guido Knapp

Yassine Talleb, M. Sc.

Verfasserin:

Julia Keiter

Gruppe 1:

Caroline Baer

Julia Keiter

Louisa Poggel

Daniel Sipek

17.11.2022

Inhaltsverzeichnis

1	Einleitung	1
2	Problemstellung	1
2.1	Datenmaterial	1
2.2	Ziele des Projekts	2
3	Statistische Auswertung	3
	Literatur	5
	Anhang	6

1 Einleitung

Besonders München gerät immer wieder in die Schlagzeilen, wenn von der (Un-)bezahlbarkeit von Mieten in deutschen Großstädten berichtet wird. Die Angebotsmieten in Euro pro Quadratmeter für Wohnungen in München sind seit 2012 bis zum 2. Quartal 2022 um mehr als 63 % gestiegen Statista Research Department (2022). Welche Kriterien bei der Mietpreisbildung berücksichtigt werden, wird als sogenannter Mietspiegel von vielen deutschen Städten abgebildet. Der Mietspiegel soll als sachliche Entscheidungshilfe dienen indem eine ortsübliche Vergleichsmiete, die nach BGB §558 insbesondere von

”den üblichen Entgelten, die in der Gemeinde oder einer vergleichbaren Gemeinde für Wohnraum vergleichbarer Art, Größe, Ausstattung, Beschaffenheit und Lage einschließlich der energetischen Ausstattung und Beschaffenheit in den letzten vier Jahren vereinbart oder, von Erhöhungen nach §560 abgesehen, geändert worden sind”,

als Orientierung gegeben wird. Inwiefern die Nettomieten von Variablen wie Art, Größe, Lage oder ähnlichem abhängig sind, soll in diesem Projekt statistisch untersucht werden. Nachdem in Kapitel 2 der Ausschnitts des Datensatzes zum Münchener Mietspiegels aus dem Jahr 2015 beschrieben hinsichtlich der Variablenverteilungen deskriptiv beschrieben wird, wird mithilfe von in Kapitel ?? vorgestellten statistischen Methoden die Variable Nettokaltmiete von 3065 zufällig ausgewählten Wohnungen aus 25 münchener Stadtbezirken in Abhängigkeit zu den übrigen Variablen gesetzt, die den Zustand der jeweiligen Wohnung beschreiben. Die Auswertung und Interpretation der Ergebnisse in Kapitel ?? ermöglicht schließlich die Zusammenfassung in Kapitel ?? und eine kurze Diskussion der Ergebnisse.

2 Problemstellung

2.1 Datenmaterial

Der zugrunde liegende Datensatz *mietspiegel2015.csv* wurde im Auftrag der Landeshauptstadt München von TNS Deutschland erhoben. Die Daten zu n=3065 Wohnungen stellen in 13 Variablen die Grundlage dar, eine Datenanalyse zur Erstellung des Mietspiegels für München für das Jahr 2015 durchzuführen. Die Daten hat die Stadt München nicht mehr online gestellt. Die Dokumentation zum Datensatz ist dort allerdings noch verfügbar (Sozialreferat der Landeshauptstadt München (2015)).

Die Daten wurden Mietspiegelinterview erhoben. Die Erhebung erfolgte zum einen Teil in Form eines persönlich-mündlichen Interviews der Mieter und zum anderen Teil in Form von schriftlichen Fragebögen, die an die Vermieter verschickt wurden. Die schriftlichen Fragebögen konnten analog oder digital ausgefüllt werden. Die Teilnehmenden wurden

entweder persönlich oder durch ein Telefoninterview befragt.

In der nominal skalierten Variable *Bezirk* sind die 25 Münchener Stadtbezirke aufgeführt, in denen sich die 3065 Wohnungen jeweils befinden. Die übrigen zwölf interessierenden Variablen sind mit den jeweiligen Skalenniveaus und Ausprägungen in Tabelle 1 aufgeführt.

Tabelle 1: erhobene Variablen mit Messniveau und Ausprägungen

Variable	Skalenniveau	Ausprägungen
Nettomiete (pro Monat in €)	metrisch, diskret	{174.75, ..., 6000.00}
Nettomiete (pro Monat in € und m^2)	metrisch, diskret	{2.47, ..., 22.13}
Wohnfläche (in m^2)	metrisch, diskret	{15, ..., 300}
Anzahl der Zimmer	metrisch, diskret	{1, ..., 8}
Baujahr	metrisch, diskret	{1918.0, ..., 2012.5}
Bezirkname	nominal	{Hadern, Laim,...}
Gute Lage	nominal, dichotom	{0, 1}
Beste Lage	nominal, dichotom	{0, 1}
Warmwasserversorgung	nominal, dichotom	{0, 1}
Heizung	nominal, dichotom	{0, 1}
Gefliestes Bad	nominal, dichotom	{0, 1}
Bad	nominal, dichotom	{0, 1}
Küche	nominal, dichotom	{0, 1}

Die Ausprägungen 0 und 1 in den nominal, dichotomen Variablen sind wie folgt zu interpretieren: Bei der Variable *Gute Lage* steht 0 für eine andere Lagekategorie und 1 für eine gute Lage. Bei der Variable *Beste Lage* steht 0 für eine andere Lagekategorie und 1 für eine beste Lage. Bei der Variable *Warmwasserversorgung vom Vermieter gestellt*, in der Tabelle und im Folgenden mit *Warmwasser* abgekürzt, steht 0 für ja und 1 für nein. Bei der Variable *Zentralheizung verfügbar*, in der Tabelle und im Folgenden mit *Heizung* abgekürzt, steht 0 für ja und 1 für nein. Bei der Variable *Gefliestes / Gekacheltes Bad*, in der Tabelle und im Folgenden mit *Fliese/Kachel* abgekürzt, steht 0 ja und 1 nein. Bei der Variable *Ausstattung des Bades*, in der Tabelle und im Folgenden mit *Bad* abgekürzt, steht 0 für normal und 1 für gehoben. Bei der Variable *Ausstattung der Küche*, in der Tabelle und im Folgenden mit *Küche* abgekürzt, steht 0 für normal und 1 für gehoben. Die Datenqualität ist aufgrund keinerlei fehlender Werte im gesamten Datensatz als sehr gut zu bewerten.

2.2 Ziele des Projekts

Ziel dieses Projekts ist, ein geeignetes multiples Regressionsmodell zur Schätzung der *Nettomiete* als Regressanten zu konstruieren. Da die Variable *Nettomiete in m^2* als Quotient der Variablen *Nettomiete* und *Wohnfläche* ein alternativer Regressant und keine unabhängige Variable ist, wird die *Nettomiete in m^2* in der Modellbildung nicht berücksichtigt. Welche Variablen einen signifikanten Einfluss zur Modellierung des Sachverhalts haben

(und damit Regressoren genannt werden), wird durch geeignete Test- und Variablenselektionsverfahren geprüft. Nachdem überprüft wird, welche Modell- und insbesondere Fehlerannahmen das finale Modell erfüllt, wird abschließend erläutert, inwiefern die Regressoren die *Nettomiete* im konstruierten Modell bilden.

3 Statistische Auswertung

Um einen Überblick über die Verteilungen der Variablen zu erhalten, sind in Tabelle 2 und 3 univariate Kenngrößen zu den Variablen aufgeführt.

Tabelle 2: univariate Kenngrößen für metrische Variablen

	Nettomiete (€)	Wohnfläche (m^2)	Zimmeranzahl	Baujahr
arithm. Mittel	763.06	71.98	2.70	1964.21
Median	700.00	70.00	3.00	1957.50
Minimum	174.75	15.00	1.00	1918.00
Maximum	6000.00	300.00	8.00	2012.50
Standardabw.	338.16	25.74	0.98	26.51
MAD	261.90	22.24	1.48	27.43

Da nah beieinanderliegende Lage- (arithmetisches Mittel vs. Median) bzw. Streuungsmaße (Standardabweichung vs. MAD) auf eine symmetrische Verteilung der Variable schließen lassen (Fahrmeir (2016), Seite 60), werden den metrisch skalierten Variablen vor allem diese Kenngrößen betrachtet.

Das arithmetische Mittel (763 €) und der Median (700 €), sowie die Standardabweichung (338.16 €) und MAD (261.9 €) der Variable *Nettomiete* unterscheiden sich. Dies lässt auf eine asymmetrische Verteilung bzw. auf den Einfluss eines Ausreißers schließen. Diese Vermutung wird in Abbildung 1 bestätigt. Um die Verzerrung durch den Einfluss des Ausreißers zu umgehen, wird die Beobachtung mit *Nettomiete*= 60000 aus dem Datensatz entfernt, sodass sich eine neue Stichprobengröße von $n=3064$ ergibt. Die Lage- und Streuungsmaße für die metrischen Variablen *Wohnfläche*, *Zimmeranzahl* und *Baujahr* unterscheiden sich nur marginal, was auf eine ausgeglichene Verteilung dieser Variablen schließen lässt.

Tabelle 3: Deskriptive Kenngrößen für nominal, dichotome Variablen

	Gute Lage	Beste Lage	Warmwasser	Heizung	Fliese/Kachel	Bad	Küche
absolute Häufigkeit "0"	1980.00	2955.00	3039.00	2861.00	380.00	2704.00	2298.00
absolute Häufigkeit "1"	1085.00	110.00	26.00	204.00	2685.00	361.00	767.00
relative Häufigkeit "0"	0.65	0.96	0.99	0.93	0.12	0.88	0.75
relative Häufigkeit "1"	0.35	0.04	0.01	0.07	0.88	0.12	0.25
Modus	0	0	0	0	1	0	0

Die deskriptive Kenngrößen für die nominal, dichotomen Variablen zeigen, welche Zustände in den meisten der 3065 betrachteten Wohnungen vorliegen: 1085 Wohnungen befinden

sich in einer guten, 110 in einer besten Lage. Damit befinden sich 1870 Wohnungen, das sind 61 % der Grundgesamtheit, in einer anderen Lagekategorie. In 99 % der Wohnungen ist die Warmwasserversorgung vom Vermieter gestellt. 93 % der Wohnungen verfügen über eine Zentralheizung. Das Bad ist in 88 % der Wohnungen nicht gefliest und normal ausgestattet. In 75 % der Wohnungen ist die Küchenausstattung normal.

Literatur

Fahrmeir L., Heumann C. Künstler R. Pigeot I. Tutz G. (2016). *Statistik - Der Weg zur Datenanalyse*. Springer Berlin Heidelberg.

Sozialreferat der Landeshauptstadt München (2015). *Mietspiegel für München*© 2015. besucht am 17.11.2022. URL: <https://stadt.muenchen.de/infos/mietspiegel.html>.

Statista Research Department (2022). *3.6 Symmetrie- und Wölbungsmaße Entwicklung der Angebotsmieten für Wohnungen in München von 2012 bis zum 2. Quartal 2022*. besucht am 17.11.2022. URL: <https://de.statista.com/statistik/daten/studie/535280/umfrage/mietpreise-auf-dem-wohnungsmarkt-in-muenchen/>.

Anhang

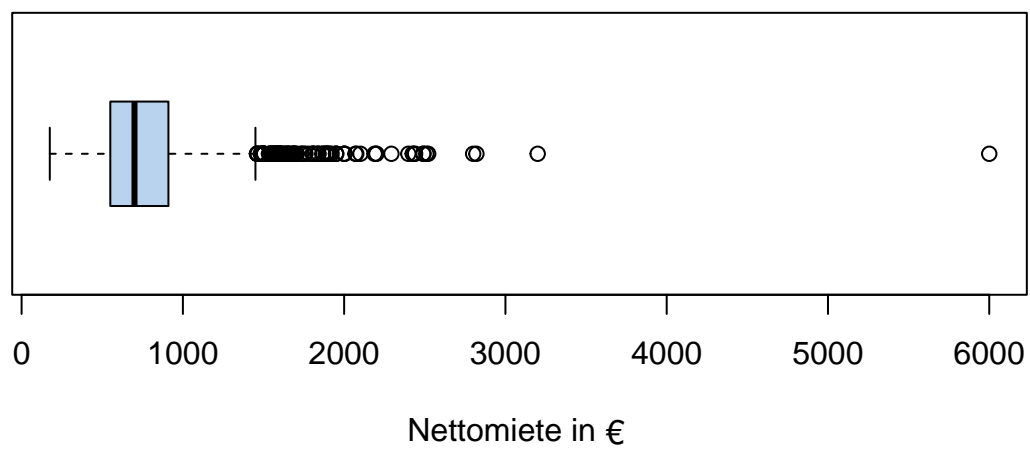


Abbildung 1: Boxplot für Variable *Nettomiete in €*