



Abschlussarbeit

Zain Khurram Chaudhary
Obada Ghazlan

20. August 2021

Inhaltsverzeichnis

1	Allgemeine Informationen	2
1.1	Ziel der Untersuchung	2
1.2	Beschreibung der Daten	2
1.3	Methode der Auswertung	3
1.4	Darstellung und Auswertung	3
2	Abschlussarbeit	4
2.1	Aufgabe 1	4
2.2	Aufgabe 2	5
2.3	Aufgabe 3	6
2.4	Aufgabe 4	6
2.5	Aufgabe 5	7
2.6	Aufgabe 6	8
2.7	Fazit	9

Kapitel 1

Allgemeine Informationen

1.1 Ziel der Untersuchung

Ziel der Untersuchung ist die Auswertung von zwei eingegebenen Datensätze (zwei Standorten in Deutschland) durch das Antwortens der folgenden Punkten

- Sind die eingegebene Datensätze bezüglich der Niederschlagssumme abhängig oder unabhängig?
- Sind die eingegebene Datensätze nach der Prüfung von Abhängigkeit bezüglich der Niederschlagssumme normalverteilt?
- Zeigen die eingegebene Datensätze bezüglich der Niederschlagssumme keinen signifikanten Unterschied?

1.2 Beschreibung der Daten

Jede Datei enthält Tageswerte verschiedener Parameter für ein ganzes Jahr (in Form einer Tabelle). Die ersten drei Spalten enthalten Angaben zum Datum. Die nachfolgenden Spalten sind die Werte für die Niederschlagsmenge (RSK) in mm, die Sonnenscheindauer (SDK) in Stunden, die mittlere Lufttemperatur (TMK), die maximale Lufttemperatur (TXK), die minimale Lufttemperatur (TNK) jeweils in °C.

1.3 Methode der Auswertung

Um die Daten bezüglich der Abhängigkeit zu prüfen, wird ein Rangkorrelationstest nach Kendall verwendet. Die Datenprüfung auf Normalität erfolgt mithilfe des Shapiro-Wilk Tests. Anschließend wird entschieden, welcher der folgenden vier Tests durchgeführt wird.

- Zwei Stichproben t-Test
- Wilcoxon Rangsummentest
- t-Test für gepaarte Stichproben
- Wilcoxon Vorzeichen-Rangtest

1.4 Darstellung und Auswertung

Die Auswertung der Ergebnisse wird durch graphische Darstellungen (Boxplots und Barplots) und Konsolenausgaben realisiert.

Kapitel 2

Abschlussarbeit

2.1 Aufgabe 1

Monat	Arkona	Norderney
JAN	62.1	73.7
FEB	51.3	98.4
MAR	26.9	34.2
APR	4.3	13.3
MAI	14.8	20.2
JUN	45.3	72.2
JUL	61.0	71.9
AUG	134.5	93.2
SEP	41.5	46.3
OKT	40.3	102.0
NOV	23.2	38.9
DEZ	38.1	89.5

Tabelle 2.1: die akkumulierte Niederschlagssumme jedes Monats.

Das Ziel der erste Aufgabe ist die akkumulierte Niederschlagssumme für jeden der zwölf Monate in einer Matrix zu ermitteln, indem wir nach Einlesen der Dateien (als Tabellen) die "Aggregate" Funktion benutzen, die die Daten so anordnet, dass für jeden Monat die Summe aller Niederschlagssumme der

einzelnen Daten ausgibt. Eine tabellarische Darstellung sieht man in der [Tabelle](#) oben.

2.2 Aufgabe 2

Das Ziel der zweiten Aufgabe ist die Niederschlagssumme eines jeden Monats graphisch in einem Balkendiagramm so zu erstellen, dass die Balken von dem gleichen Monats ausschließlich nebeneinander stehen, allerdings mit unterschiedlichen Farben.

Unter Nutzung der "barplot" Funktion geben wir die Niederschlagssumme als eine Matrix an und setzen wir die "besideäuf TRUE, damit die Balken nebeneinander stehen. Am Schluss müssen wir die Balken durch verschiedene Farben darstellen.

Die mögliche Darstellung sieht man in der [Abbildung](#) unten.

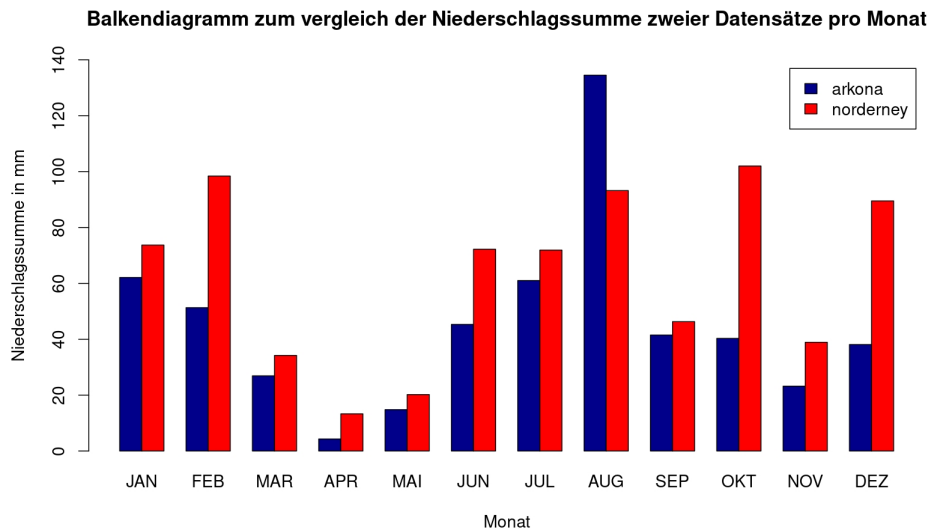


Abbildung 2.1: Balkendiagramm

2.3 Aufgabe 3

Das Ziel der dritten Aufgabe ist die zwölf Monatssummen der beiden Datensätze zusätzlich durch zwei Boxplots in einem gemeinsamen Diagramm darzustellen. Dieses läuft analog wie bei der zweiten Aufgabe, allerdings benutzt man die "boxplot" Funktion.

Die graphische Darstellung sieht man in das [Image](#) unten.

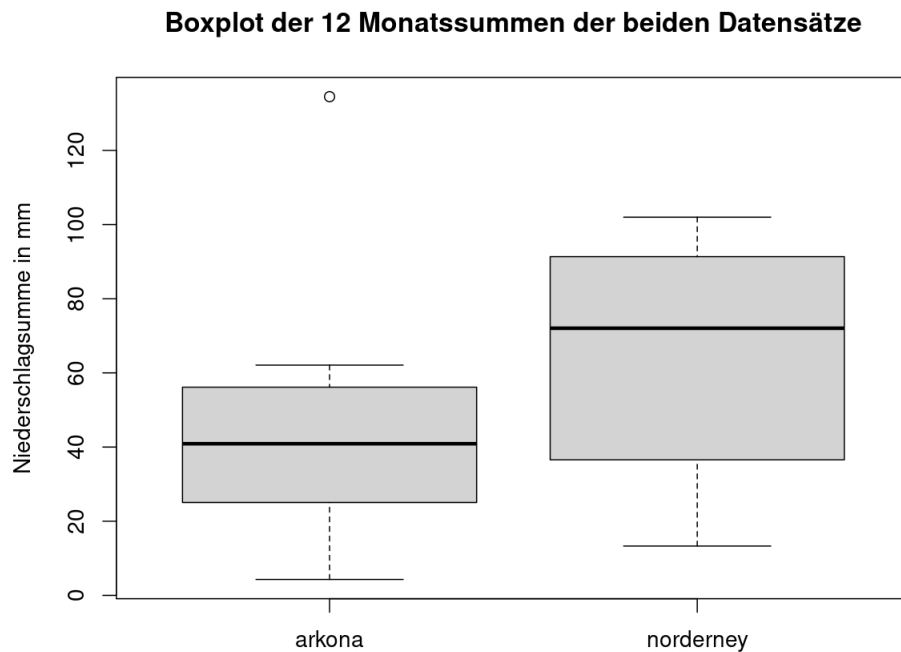


Abbildung 2.2: Boxplot

2.4 Aufgabe 4

Das Ziel der vierten Aufgabe ist zu prüfen, ob die 12 Monatssummen des Niederschläge der eingegebenen Datensätze mithilfe des Rangkorrelationstests nach Kendall, voneinander unabhängig sind.

Unsere Nullhypothese besagt, dass die beiden 12-monatigen Datensätze keine signifikante Korrelation aufweisen, d.h. dass die Daten unabhängig von einander sind.

Wir setzen das Signifikanzniveau auf 10 % und prüfen mithilfe der zweiseitigen "cor.test"-Funktion für die Niderschlagssumme der einzelnen Monate der beiden Datensätze, ob der ausgegebene P-Wert kleiner als das Signifikanzniveau (10 %) ist. Wenn das der Fall ist, dann können wir die Nullhypothese ablehnen, und die Schlussfolgerung wäre dann, dass die Datensätze signifikante Korrelation aufweisen, d.h. die Daten sind doch abhängig voneinander. Wenn das nicht der Fall ist, dann wird die Nullhypothese nicht abgelehnt. Somit sind unsere Datensätze voneinander abhängig..

Für unsere Datensätze sieht die Ausgabe wie folgt aus:

"Nach Rangkorrelationstest nach Kendall ist der p-wert = 0.0138 und da der p-wert($2 \cdot p$) < 0.1 gilt, kann die Nullhypothese abgelehnt werden. Das heißt, die beiden 12-monatigen Datensätze weisen signifikante Korrelation auf"

also unsere Daten sind abhängig.

2.5 Aufgabe 5

Nachdem wir auf Abhängigkeit der beiden Datensätze geprüft haben, müssen wir sicherstellen, ob die Daten normalverteilt sind. Es gibt allerdings eine Entscheidung, die man treffen muss bevor der Test auf Normalverteilung durchgeführt wird.

- Sind die beiden Datensätze abhängig, so testet man die Paardifferenzen $X_i - Y_i$ auf Normalverteilung.
- Sind die beiden Datensätze unabhängig, so testet man die Datensätze separat auf Normalverteilung

Unsere Nullhypothese besagt, dass die Daten normalverteilt sind. Die zwei Entscheidungen folgen allerdings das gleiche Schema, indem man mithilfe der `shapiro.test()` den ausgegebene P-Wert mit dem Signifikanzniveau vergleicht. Ist das P-Wert kleiner als der 10 % Signifikanzniveau kann man die

Nullhypothese ablehnen und daraus schlussfolgern, dass die Daten nicht normalverteilt sind.

Für unsere Datensätze sieht die Ausgabe wie folgt aus.

"Nach Shapiro-Wilk Test ist der p -wert = 0.1676 und da der p -wert < 0.1 nicht gilt, kann die Nullhypothesen nicht abgelehnt werden. Das heißt, wir können annehmen dass die Daten normalverteilt sind."

2.6 Aufgabe 6

Nachdem wir auf Abhängigkeit und Normalität der beiden Datensätze geprüft haben, muss unser Programm in der Lage sein einer der 4 Tests, um unsere Daten weiter zu prüfen

- Zwei stichproben t-Test
- Wicoxon Rangsummentext
- t-Test für gepaarte Stichproben
- Wilcoxon Vorzeichen-Rangtest

Die Aufgabe sagt "Wählen Sie nun anhand des Ergebnisses des vorangegangenen Tests auf Normalverteilung einen entsprechenden Test aus, der entscheidet, ob sich die beiden Datensätze der Monatsniederschläge signifikant voneinander unterscheiden."

Die Nullhypothese besagt, dass die beiden Stationen keinen signifikanten Unterschied zeigen. Die Entscheidungen folgen allerdings dem selben Schema, indem man mithilfe des entsprechenden Tests den ausgegebene P-Wert mit dem Signifikanzniveau vergleicht. Ist der P-Wert kleiner als die 5 % des Signifikanzniveaus kann man die Nullhypothese ablehnen und daraus schlussfolgern, dass die Daten signifikanten einen Unterschied zeigen.

Für unsere Datensätze sieht die Ausgabe wie folgt aus.

"Nach t-Test für gepaarte stichproben ist das p -wert = 0.0466 und da p -wert < 0.05 True ist kann die Nullhypothesen abgelehnt werden. Das heißt, die Monatsniederschläge der beiden Stationen zeigen signifikanten Unterschied."

2.7 Fazit

Die Entscheidung, ob die angegebenen Daten signifikant voneinander unterscheiden ist ein sehr wichtiger Bestandteil der Statistik. Sie ermöglicht uns eine bessere Verständnis und Vorstellung der Daten zu haben. Indem man sich auf verschiedenen Aspekte bezieht, die erste Aspekt ist zu bestimmen, ob Abhängigkeit zwischen unseren Datensätze gibt. Die zweite Aspekt ist zu prüfen, ob die eingegebenen Datensätze normalverteilt sind, allerdings hängt dieser Aspekt von der einigen zuvor.

Nach der Prüfung von den vorherigen Aspekten kommt zum Schluss der Auswahl der geeigneten Test, der entscheidet, ob die angegebenen Datensätze signifikant voneinander unterscheiden.