

1. Ein Verbrauchermagazin berichtet die folgenden Korrelationen:

Der Pearson Korrelationskoeffizient für KFZ-Gewicht und Zuverlässigkeit ist -0.3

Für KFZ-Gewicht und Jährliche Servicekosten beträgt er 0.2

Welche Aussage trifft zu:

1. Schwerere Autos scheinen weniger zuverlässig zu sein.
2. Schwerere Autos verursachen tendenziell höhere Service-Kosten
3. KFZ-Gewicht und Zuverlässigkeit sind stärker Korreliert als KFZ-Gewicht und Service-Kosten

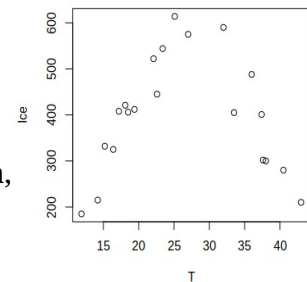
- | | |
|---------------------------------|---|
| <input type="checkbox"/> nur 1. | <input checked="" type="checkbox"/> 1. und 2. |
| <input type="checkbox"/> nur 2. | <input type="checkbox"/> 1. und 3. |
| <input type="checkbox"/> nur 3. | <input type="checkbox"/> 1., 2. und 3. |

2. Die Signifikanz von Pearson's r hängt vom Wert von r und vom Probenumfang ab. Warum ist sie unabhängig von der Varianz der Messwerte?

3. Neun Personen messen, wie lange sie die Luft anhalten können, wenn sie vorher normal geatmet bzw. hyperventiliert haben (Datensatz: **hypervent.csv**). Geben Sie die Korrelationskoeffizienten für die Zeiten nach Pearson und nach Spearman an. Vergleichen Sie die Werte und erklären Sie, warum der eine größer ist als der andere.

4. Im Software-Paket "R" ist der Datensatz "faithful" implementiert (**faith.csv**), der die Länge von Eruptionen des Old Faithful Geysiers im Yellowstone National Park gegen die Wartezeit zwischen den Eruptionen aufzeichnet. Können Sie für den Datensatz einen Korrelationskoeffizienten nach Pearson angeben? Begründen Sie unter Verwendung eines Tests! Können Sie einen Zusammenhang zwischen den beiden Parametern feststellen? Belegen Sie!

5. Für die Abhängigkeit des Speiseeis-Verkaufs von der Temperatur (**ice.csv**) wurde keine signifikante Korrelation gefunden. Die Umsätze als Funktion der Temp. zeigt Abb. 1. Dass Eisverkauf nicht von der Temperatur beeinflusst wird, ist wenig intuitiv. Wie könnte man prüfen, ob es einen Zusammenhang gibt?



6. Das Thunder Basin (Wyoming, USA) ist ein semi-arides Grasland. Die Größe der Antilopen-Population (/100; Datensatz: **antelope.csv**; "Population") wurde bezüglich der Abhängigkeit von Anzahl der Kitze im Frühling (/100; "Kitze") und des jährlichen Niederschlags (inch; "Niederschlag") untersucht.
- a) Prüfen Sie, ob jeweils signifikante Zusammenhänge bestehen. Wie gehen Sie vor? Geben Sie p-Werte an.
 - b) Prüfen Sie, ob Anzahl Kitze und Niederschlag einander stören
7. Die Datei "**blood.csv**" enthält Werte des systolischen Blutdrucks (sys), des Alters (age) und des Gewichts in kg (weight) von Patienten. Es ist davon auszugehen, dass der Blutdruck von beiden Parametern beeinflusst wird.
- a) Führen Sie eine multiple Regression durch und geben Sie an, ob die Korrelationen signifikant sind.
 - b) Bestimmen Sie anhand der Beta-Gewichte die Bedeutung der Parameter
 - c) Testen Sie die Parameter auf Redundanz