

## Lösung zu Kapitel 9: Beispiel 1

Im April 2002 fand in Österreich ein Volksbegehren mit dem Ziel statt, soziale Rechte in der Verfassung festzuschreiben. Da dieses Volksbegehren durch die damals oppositionelle sozialdemokratische Partei (SP) unterstützt wurde, bezeichneten Kritiker des Volksbegehrens diese Unterstützung als vorweggenommenen Wahlkampf.

Das Datenfile `volksbegehren.csv` enthält Daten zu den Ergebnissen in Wien in folgenden Variablen:

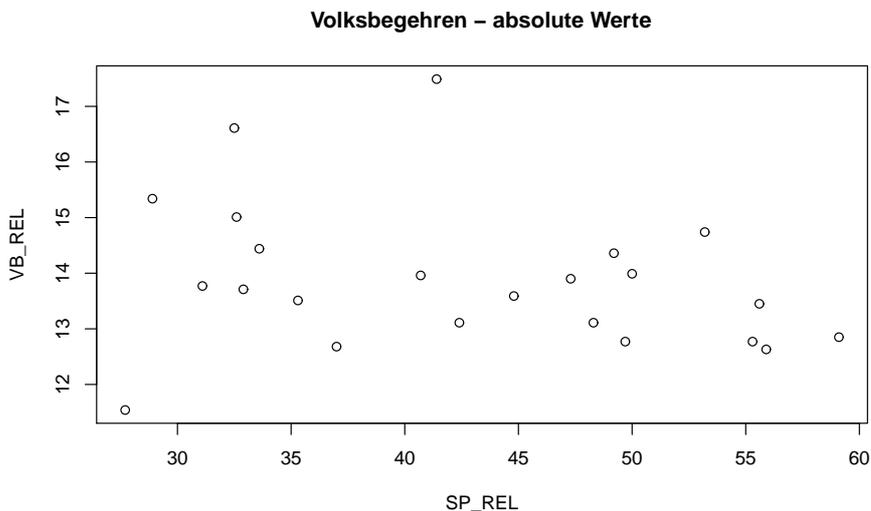
- `BEZIRK` Nummer des Bezirks in Wien
- `VB_ABS` Anzahl Unterschriften für das Volksbegehren
- `VB_REL` Anteil der Unterschriften an Wahlberechtigten
- `SP_ABS` Stimmen für die SP bei den Gemeinderatswahlen 2001
- `SP_REL` Stimmenanteil für die SP bei den Gemeinderatswahlen 2001

- Erstellen Sie ein Streudiagramm mit den absoluten Angaben (`SP_ABS` und `VB_ABS`) und ermitteln Sie den Korrelationskoeffizienten!
- Warum kann aus diesem Ergebnis noch nicht auf einen Zusammenhang zwischen Stärke der SP und Unterstützung für das Volksbegehren geschlossen werden?
- Führen Sie obige Untersuchung mit den relativen Angaben (`SP_REL` und `VB_REL`) durch!

Zur Verfügung stehen sowohl absolute wie auch relative Werte. Ein Blick auf die absoluten Zahlen verleitet zur Annahme, dass ein starker linearer Zusammenhang besteht (► Abbildung 1).

**R**

```
> vb <- read.csv2("volksbegehren.csv", header = T)
> attach(vb)
> plot(SP_ABS, VB_ABS, main = "Volksbegehren - absolute Werte")
```



**Abbildung 1:** Streudiagramm der absoluten Angaben von SP-Stimmen und Beteiligung am Volksbegehren.

Natürlich zeigt auch der Korrelationskoeffizient einen hohen und signifikant von 0 abweichenden Wert an.

R

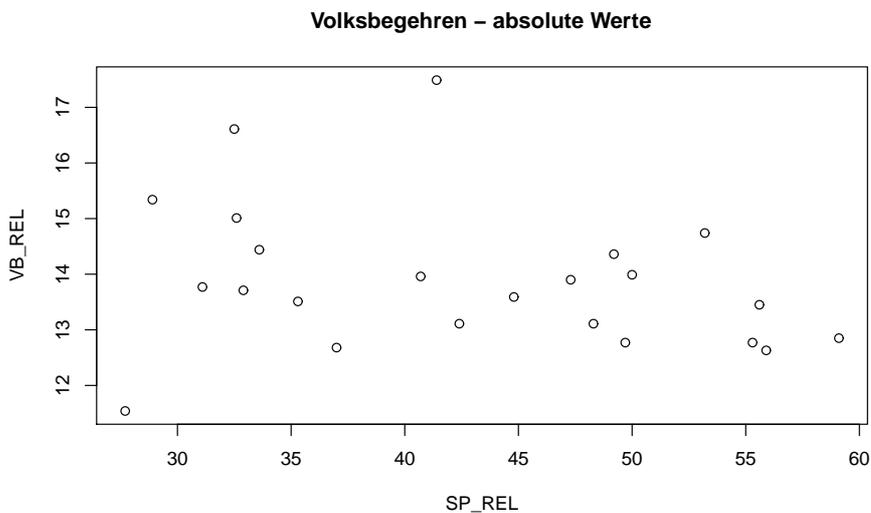
```
> cor.test(SP_ABS, VB_ABS)
```

Pearson's product-moment correlation

```
data: SP_ABS and VB_ABS
t = 17.8214, df = 21, p-value = 3.708e-14
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9259416 0.9867645
sample estimates:
      cor
0.9684938
```

Daraus abzuleiten, dass ein hoher Anteil an SP-Stimmen im Bezirk mit einer hohen Beteiligung am Volksbegehren einhergeht, ist etwas voreilig. Die Bezirke sind von der Einwohnerzahl recht unterschiedlich. In kleinen Bezirken werden vermutlich weniger Stimmen für die SP abgegeben als in großen; aber ähnlich wird es sich mit den Unterschriften für das Volksbegehren verhalten.

Diese Überlegung führt auch schon zur Lösung des Problems: Die Daten müssen unabhängig von der Bewohneranzahl, also in relativen Häufigkeiten, betrachtet werden. Im Streudiagramm (► Abbildung 2) ist nur ein schwacher - noch dazu ein negativer - Zusammenhang beobachtbar.



**Abbildung 2:** Streudiagramm der relativen Angaben zu SP-Stimmen und Beteiligung am Volksbegehren.

R

```
> plot(SP_REL, VB_REL, main = "Volksbegehren - relative Werte")
> cor.test(SP_REL, VB_REL)
```

Pearson's product-moment correlation

```
data: SP_REL and VB_REL
```

```
t = -1.2286, df = 21, p-value = 0.2328
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.6064279  0.1715571
sample estimates:
      cor
-0.2589576
```

Zum Schluss lösen wir die Verbindung zum Datensatz wieder auf:

**R**

```
> detach(vb)
```

In diesem Beispiel war die Ursache für die hohe Korrelation zwischen den Absolutwerten die – nicht berücksichtigte – Einwohnerzahl der Bezirke, die auf beide Variablen Einfluss hat. Berücksichtigt man diese Variable Einwohnerzahl (etwa durch Übergang auf relative Werte), ist die Korrelation nicht mehr von Bedeutung.

Solche Fälle von Scheinkorrelation (wichtige Variablen kommen im Modell nicht vor) treten nicht selten auf und führen vor allem bei kausaler Argumentation zu Fehlschlüssen.