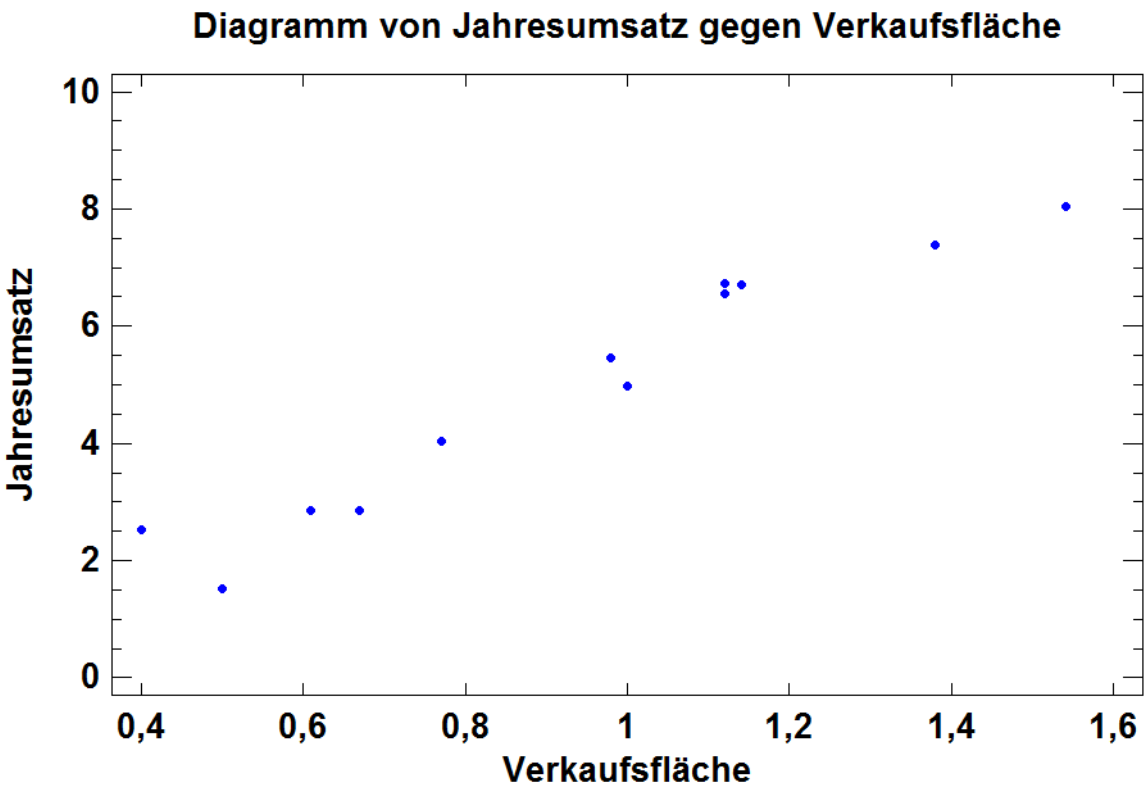
8. Lösung weitere Übungsaufgaben Statistik II WiSe 2019/2020

1. **Aufgabe:** Ein Unternehmen will den Zusammenhang zwischen den beiden Merkmalen Verkaufsfläche und Jahresumsatz untersuchen. In der folgenden Tabelle sind die Daten von 12 Filialen des Unternehmens zu finden:

Filiale i	Verkaufsfläche (in Tsd. m^2) x_i	Jahresumsatz (in Mio. €) y_i
1	0,40	2,53
2	1,00	4,97
3	1,12	6,55
4	1,14	6,71
5	1,12	6,72
6	1,54	8,05
7	0,77	4,03
8	0,98	5,47
9	1,38	7,38
10	0,67	2,86
11	0,50	1,52
12	0,61	2,85



Gehen Sie im Weiteren davon aus, dass es sich bei den Daten um Realisierungen normalverteilter Merkmale handelt.

a) Für die gewönliche Korrelation (Pearson) liefert Statgraphics folgendes Ergebnis.

Korrelationen

	Verkaufsfläche	Jahresumsatz
Verkaufsfläche		0,9700
		(12)
		0,0000
Jahresumsatz	0,9700	
	(12)	
	0,0000	

Korrelation
(Stichprobengröße)
p-Wert

Welche Hypothese wird getetest? Wie lautet die Testentscheidung bei $\alpha = 0{,}01$?

b) Testen Sie zum Niveau $\alpha = 0{,}05$, ob die Korrelation zwischen den beiden Merkmalen signifikant größer ist als 0,92.

Lösung:
Untersucht werden zwei verbundene metrisch skalierte Merkmale (X, Y). Wobei

$$X \ldots \text{Verkaufsfläche}$$
$$Y \ldots \text{Jahresumsatz}$$

Die Stichprobengröße ist $n = 12$. Des weiteren soll der Merkmalsvektor $\begin{pmatrix} X \\ Y \end{pmatrix}$ einer Normalverteilung folgen.

a) Hier wird ein Test für den gewöhnlichen Korrelationskoeffizienten nach Pearson durchgeführt. Die zugehörigen Hypothesen lauten:

$$\underbrace{H_0 : \rho_{XY} = 0}_{\text{unkorreliert}} \qquad v.s. \qquad \underbrace{H_A : \rho_{XY} \neq 0}_{\text{korreliert}}$$

Der p-Wert lässt sich aus der Tabelle ablesen (dritte Zelle).
Es gilt $p = 0,0000 < 0,01 = \alpha \Longrightarrow H_0$ wird abgelehnt. D.h. die Korrelation zwischen den Merkmalen Verkaufsfläche (X) und Jahresumsatz (Y) ist signifikant von 0 verschieden.

Da der Merkmalsvektor $\begin{pmatrix} X \\ Y \end{pmatrix}$ als zweidimensional normalverteilt vorausgesetzt wird ist unkorreliert gleich unabhängig. D.h. es gibt eine signifikant Abhängigkeit zwischen den Merkmalen Verkaufsfläche (X) und Jahresumsatz (Y).

b) Es folgt der *Test auf die Größe von* ρ_{XY}:

1.) Das Hypothesenpaar lautet

$$H_0 : \rho_{XY} \leq 0,92 =: \rho_0 \qquad v.s. \qquad H_A : \rho_{XY} > 0,92 =: \rho_0$$

2.) Das Signifikanzniveau ist $\alpha = 0{,}05$.

3.) Die Teststatistik lautet

$$T = (Z - Z_0)\sqrt{n-3}$$

wobei

$$Z = \frac{1}{2}\ln\left(\frac{1+r_{xy}}{1-r_{xy}}\right)$$

und

$$Z_0 = \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right) + \frac{\rho_0}{2(n-1)}.$$

4.) Der kritische Bereich lautet

$$K = \{t \,|\, t \geq z_{1-\alpha}\}$$

Der Wert des Quantils lässt sich aus der Tafel ablesen:

$$z_{1-\alpha} = z_{0{,}95} = 1{,}6449.$$

5.) Den Korrelationskoeffizenten erhält man aus der Tabelle **Korrelation** aus (a): $r_{XY} = 0{,}97$. Damit erhält man

$$\begin{aligned} z &= \frac{1}{2}\ln\left(\frac{1+0{,}97}{1-0{,}97}\right) = 2{,}092296 \\ z_0 &= \frac{1}{2}\ln\left(\frac{1+0{,}92}{1-0{,}92}\right) + \frac{0{,}92}{2(12-1)} = 1{,}630845 \end{aligned}$$

Damit ergibt sich

$$t = (2{,}092296 - 1{,}630845)\sqrt{12-3} = 1{,}384$$

6.) Es gilt $t = 1{,}384 < 1{,}6449 \Longrightarrow t \notin K \Longrightarrow H_0$ wird angenommen.

D.h. die Korrelation zwischen X und Y ist nicht signifikant größer als 0,92.

---

2. Aufgabe: Eine Firma stellt in zwei Werken Schraubenspindelpumpen her und analysiert die Verkaufszahlen aus den letzten zehn Monaten.

i	1	2	3	4	5	6	7	8	9	10
x_i	404	391	400	332	380	322	405	35	383	431
y_i	261	250	257	236	251	221	246	22	280	305

x ist der Absatz aus dem ersten und y aus dem zweiten Werk. Der Firmenchef fragt sich, ob die Absatzzahlen der beiden Werke zusammenhängen.

a) Schätzen Sie aus der Stichprobe die Rangkorrelation von Kendall zwischen den Verkaufszahlen der beiden Werke.

j	1	2	3	4	5	6	7	8	9	10
x_j	35	322	332	380	383	391	400	404	405	431
y_j	22	221	236	251	280	250	257	261	246	305
q_j	0	0	0	2	4	1	1	1	0	0

$$\sum_{j=1}^{10} q_j = 2+4+1+1+1 = 9$$

$$r_{X,Y}^{(K)} = 1 - \frac{4 \cdot \sum\limits_{j=1}^{10} q_j}{n \cdot (n-1)} = 1 - \frac{4 \cdot 9}{10 \cdot 9} = \frac{3}{5} = \underline{0,6}$$

(Man kann auch hier Ränge wie im Aufgabenteil b) bestimmen, muss es aber nicht. Nach Bestimmung der Ränge ist die Bestimmung der q_j etwas einfacher.)

b) Schätzen Sie aus der Stichprobe die Rangkorrelation von Spearman zwischen den Verkaufszahlen der beiden Werke.
<u>Lösung:</u>

i	1	2	3	4	5	6	7	8	9	10
x_i	404	391	400	332	380	322	405	35	383	431
$Rang(x_i)$	8	6	7	3	4	2	9	1	5	10
y_i	261	250	257	236	251	221	246	22	280	305
$Rang(y_i)$	8	5	7	3	6	2	4	1	9	10
$d_i = Rang(x_i) - Rang(y_i)$	0	1	0	0	-2	0	5	0	-4	0
d_i^2	0	1	0	0	4	0	25	0	16	0

$$\sum_{i=1}^{10} (Rang(x_i) - Rang(y_i))^2 = \sum_{i=1}^{10} d_i^2 = 1+4+25+16 = 46$$

$$\begin{aligned} r_{X,Y}^{(S)} &= 1 - \frac{6 \sum\limits_{i=1}^{n} (Rang(x_i) - Rang(y_i))^2}{n(n^2-1)} \\ &= 1 - \frac{6 \cdot 46}{10 \cdot (100-1)} \\ &= 1 - \frac{46}{165} = \frac{119}{165} = \underline{0{,}7212} \end{aligned}$$

c) Auf welchen Zusammenhang deuten die geschätzten Rangkorrelationen hin?
<u>Lösung:</u> Beide Rangkorrelationen deuten auf einen monoton wachsenden Zsammenhang hin. Das ist sicher äußeren Einflüssen geschuldet. Sind diese Einflßse gut, dann haben beide Werke hohe Absatzzahlen.

3. Aufgabe: In der folgenden Tabelle finden Sie für die 7 einwohnerstärksten Städte in Deutschland die Einwohnerzahl (in Millionen) und die Anzahl der Flugpassagiere (in Millionen). Dabei wurde bei der Zahl der Flugpassagiere in Berlin die der 2 großen Flughäfen betrachtet.

Stadt	Einwohner (in Millionen)	Passagiere (in Millionen)
Berlin	3,422	27,98
Hamburg	1,746	14,76
München	1,408	39,70
Köln	1,034	9,45
Frankfurt am Main	0,701	59,57
Stuttgart	0,604	9,72
Düsseldorf	0,599	21,84

Bestimmen Sie die Rangkorrelation von Spearman zwischen der Anzahl der Einwohner und der Anzahl der abgefertigten Passagiere.

Lösung:
X - Einwohner
Y - Passagiere

Stadt	X-Einwohner	Rang(X)	Y-Passagiere	Rang(Y)
Berlin	3,422	7	27,98	5
Hamburg	1,746	6	14,76	3
München	1,408	5	39,70	6
Köln	1,034	4	9,45	1
Frankfurt am Main	0,701	3	59,57	7
Stuttgart	0,604	2	9,72	2
Düsseldorf	0,599	1	21,84	4

i	1	2	3	4	5	6	7
$Rang(x_i)$	7	6	5	4	3	2	1
$Rang(y_i)$	5	3	6	1	7	2	4
$d_i = Rang(x_i) - Rang(y_i)$	2	3	-1	3	-4	0	-3
d_i^2	4	9	1	9	16	0	9

$$\begin{aligned} r_{X,Y}^{(S)} &= 1 - \frac{6\sum_{i=1}^{n}(Rang(x_i) - Rang(y_i))^2}{n(n^2-1)} \\ &= 1 - \frac{6 \cdot 48}{7 \cdot (49-1)} \\ &= 1 - \frac{6}{7} = \frac{1}{7} = \underline{0{,}1429} \end{aligned}$$

---

4. Aufgabe: Für die 7 größten Flughäfen (nach Anzahl der abgefertigten Passagiere) wird der Zusammenhang zur Einwohnerzahl der Metropolregion der Stadt untersucht. Folgende Daten liegen vor.

Stadt	Passagiere in Millionen	Einwohner der Metropolregion in Millionen
Atlanta	95	6
Peking	82	21
London	70	15
Tokio	67	36
Chicago	66	9
Los Angeles	64	13
Paris	62	10

Bestimmen Sie die Rangkorrelation von Kendall zwischen der Anzahl der abgefertigten Passagiere und der Einwohnerzahl der Metropolregion der Stadt.

Lösung:

X - zufällige Anzahl der abgefertigten Passagiere
Y - zufällige Einwohnerzahl

j	1	2	3	4	5	6	7
Passagiere	62	64	66	67	70	82	95
Einwohnerzahl	10	13	9	36	15	21	6
q_j	2	2	1	3	1	1	0

$$\sum_{j=1}^{7} q_j = 2 + 2 + ... + 1 + 0 = 10$$

$$r_{X,Y}^{(K)} = 1 - \frac{4 \cdot \sum_{j=1}^{7} q_j}{n \cdot (n-1)} = 1 - \frac{4 \cdot 10}{7 \cdot 6} = \frac{1}{21} = \underline{0,0476}$$
