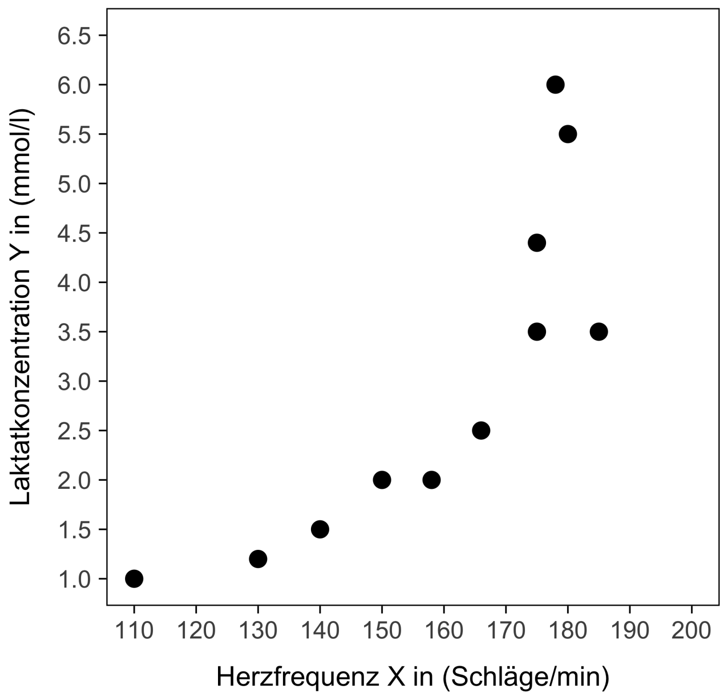
Aufgabe 1

Teil A

Moderne Fitnessuhren zeigen nach dem Training unter anderem die Herzfrequenz, die Geschwindigkeit und die erreichte Trainingszone an. Da die Trainingszonen an der Laktatkonzentration des Körpers orientiert sind, die nur im Labor direkt messbar ist, wird der Zusammenhang zwischen Laktat und Herzfrequenz genutzt, um die erreichte Trainingszone zu bestimmen. Bei ausgewählten Leistungssportlern werden hierfür die *Herzfrequenz X (Schläge/min)* und die *Laktatkonzentration Y (mmol/l)* unter verschiedenen Belastungsstufen auf dem Fahrradergometer im Labor gemessen. Diese Daten dienen als Grundlage, um die Trainingszonen zu schätzen. Die Messungen einer Leistungssportlerin wurden beispielhaft aufgezeichnet und in einem Streudiagramm visualisiert.

Messung	1	2	3	4	5	6	7	8	9	10	11
X	110	130	158	185	166	180	175	175	140	178	150
Y	1,0	1,2	2,0	3,5	2,5	5,5	3,5	4,4	1,5	6,0	2,0



Das arithmetische Mittel und die empirischen Varianz von X bzw. Y wurden bereits ermittelt:

$$\overline{x} = 158,8182\ (Schläge/min) \quad \text{und} \quad \overline{y} = 3,0091\ (mmol/l)$$
$$s_x^2 = 520,3306\ (Schläge/min)^2 \quad \text{und} \quad s_y^2 = 2,6772\ (mmol/l)^2$$

a) Wie sind die statistischen Merkmale X und Y skaliert?

X ist kardinal-diskret skaliert; Y ist kardinal-stetig skaliert.

b) Berechnen Sie die *empirische Kovarianz* s_{xy} zwischen X und Y. Verbalisieren Sie das Ergebnis.

Es gilt $s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}) \cdot (y_i - \overline{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy}$, also

$$\begin{aligned} s_{xy} = {} & \frac{1}{11} \cdot (110 \cdot 1,0 + 130 \cdot 1,2 + 158 \cdot 2,0 + 185 \cdot 3,5 + 166 \cdot 2,5 + 180 \cdot 5,5 \\ & + 175 \cdot 3,5 + 175 \cdot 4,4 + 140 \cdot 1,5 + 178 \cdot 6,0 + 150 \cdot 2,0) - 158,8182 \cdot 3,0091 \\ = {} & \frac{1}{11} \cdot 5595 - 158,8182 \cdot 3,0091 \\ \approx {} & 30,7365 \; (Schläge/min) \cdot (mmol/l) \end{aligned}$$

Auf Basis der Messungen, besteht zwischen der Herzfrequenz X und der Laktatkonzentration Y gemäß der empirischen Kovarianz ein *gleichsinniger/positiver linearer Zusammenhang*, d.h. höhere Werte in X (höhere Herzfrequenz) gehen mit höheren Werten in Y (höhere Laktatkonzentration) einher.

c) Berechnen Sie den *Korrelationskoeffizienten nach Bravais-Pearson* r. Interpretieren Sie ihr Ergebnis statistisch exakt.

$$r := \frac{s_{xy}}{\sqrt{s_x^2 \cdot s_y^2}}, \quad \text{also} \quad r = \frac{30,7366}{\sqrt{520,3306 \cdot 2,6772}} \approx 0,8235$$

Auf Basis der Messungen, besteht zwischen der Herzfrequenz X und der Laktatkonzentration Y gemäß des Korrelationskoeffizienten nach Bravais-Pearson ein *starker gleichsinniger/positiver linearer Zusammenhang*.

d) Welches der beiden Maßzahlen s_{xy} und r eignet sich besser, um die Stärke eines möglichen linearen Zusammenhangs zwischen X und Y zu beurteilen? Begründen Sie Ihre Antwort.

Der Korrelationskoeffizient nach Bravais-Pearson r eignet sich besser, da er skalenunabhängig, dimensionslos und normiert ist.

e) Bestimmen Sie den *korrigierten Rangkorrelationskoeffizienten nach Kendall* τ^*. Interpretieren Sie Ihr Ergebnis statistisch exakt.

$$\tau^* := \frac{P-Q}{\sqrt{\frac{n(n-1)}{2}-\tau_x}\cdot\sqrt{\frac{n(n-1)}{2}-\tau_y}}, \quad \text{also} \quad \tau^* = \frac{48-4}{\sqrt{55-1}\cdot\sqrt{55-2}} \approx 0,8225$$

Auf Basis der Messungen, besteht zwischen der Herzfrequenz X und der Laktatkonzentration Y gemäß des korrigierten Rangkorrelationskoeffizienten nach Kendall ein *gleichsinniger/positiver monotoner* Zusammenhang. Die Stärke des Zusammenhangs liegt mit einem Wert von $0,8225$ im oberen Bereich des Werteintervalls $[-1,1]$.

Nebenrechnungen:
Es ist $n=11$, $P=48$, $Q=4$, $g_1=2$, $h_1=2$, $h_2=2$, folglich $\frac{11(11-1)}{2}=55$ und

$$\tau_x = \frac{1}{2}\sum_{i=1}^{k} g_i\cdot(g_i-1), \quad \text{also} \quad \tau_x = \frac{1}{2}\cdot(2\cdot(2-1)) = 1$$

$$\tau_y = \frac{1}{2}\sum_{j=1}^{l} h_j\cdot(h_j-1), \quad \text{also} \quad \tau_y = \frac{1}{2}\cdot\Big[2\cdot(2-1)+2\cdot(2-1)\Big] = 2.$$

Ermittlung von P, Q, g_1, h_1 und h_2 mittels abgebildetem Streudiagramm möglich. Alternativ via folgender Hilfstabelle:

Messung	1	2	9	11	3	5	7	8	10	6	4
X	110	130	140	150	158	166	175	175	178	180	185
Y	1,0	1,2	1,5	2,0	2,0	2,5	3,5	4,4	6,0	5,5	3,5
		+	+	+	+	+	+	+	+	+	+
			+	+	+	+	+	+	+	+	+
				+	+	+	+	+	+	+	+
					0	+	+	+	+	+	+
						+	+	+	+	+	+
							+	+	+	+	+
								0	+	+	0
									+	+	-
										-	-
											-

f) Warum müssen wir in e) die korrigierte Version des *Rangkorrelationskoeffizienten nach Kendall* berechnen?

Der *Rangkorrelationskoeffizient nach Kendall* muss in e) korrigiert werden, da Bindungen (z.B. $x_7 = x_8 = 175$ oder $y_4 = y_7 = 3,5$) in X und Y vorliegen.

g) Welches der beiden Zusammenhangsmaße r und τ eignet sich Ihrer Meinung nach besser zur Beschreibung des Zusammenhangs von X und Y in diesen beobachteten Daten? Begründen Sie Ihre Wahl unter Bezugnahme auf das Streudiagramm.

Da das Streudiagramm einen exponentiellen Zusammenhang zwischen X und Y zeigt, ist der Rangkorrelationskoeffizienten nach Kendall τ besser geeignet als der Korrelationskoeffizienten nach Bravais-Pearson r. τ erfasst *monotone* Zusammenhänge, unabhängig von deren Form (also auch nicht-lineare), während r nur *lineare* Zusammenhänge misst.

Teil B

Zusätzlich wird im Labor zu jeder in Teil A aufgeführte Messung *die auf dem Fahrradergometer erbrachte Leistung W* in Watt dokumentiert.

Messung	1	2	3	4	5	6	7	8	9	10	11
W	75	100	125	185	140	165	150	160	90	175	120

Der Rangkorrelationskoeffizienten nach Kendall τ zwischen X und W sowie zwischen Y und W wurde bereits ermittelt:

$$\tau_{xw} = 0,9175 \quad \text{und} \quad \tau_{yw} = 0,8335$$

h) Berechnen sie den *partiellen Rangkorrelationskoeffizient nach Kendall* $\tau_{xy\circ w}$ und interpretieren Sie Ihr Ergebnis statistisch exakt.

Es ist $\tau_{xy} = \tau^* = 0,8225$ und

$$\tau_{xy\circ w} := \frac{\tau_{xy} - \tau_{xw} \cdot \tau_{yw}}{\sqrt{(1-\tau_{xw}^2) \cdot (1-\tau_{yw}^2)}}, \quad \text{also}$$

$$\tau_{xy\circ w} = \frac{0,8225 - 0,9175 \cdot 0,8335}{\sqrt{(1-0,9175^2) \cdot (1-0,8335^2)}} \approx 0,2629$$

Wird die auf dem Fahrradergometer erbrachte Leistung W konstant gehalten, zeigt sich zwischen der Herzfrequenz X und Laktatkonzentration Y nach dem Rangkorrelationskoeffizienten nach Kendall weiterhin ein *gleichsinniger/positiver monotoner Zusammenhang*. Die Stärke des Zusammenhangs fällt jedoch geringer aus, wenn die Korrelation zwischen W und X bzw. W und Y berücksichtigt wird, denn $\tau_{xy\circ w} = 0,2629 < 0,8225 = \tau_{xy}$.

Teil C

Die Fitnessuhr schätzt nun anhand der während des Trainings gemessenen Herzfrequenz den Laktatwert und zeigt darauf basierend die erreichte Trainingszone an. Für einen Hobbyläufer wurden nach dem Joggen die *durchschnittliche Herzfrequenz* $\bar{X}$ *(Schläge/min)* sowie die *erreichte Trainingszone* Z auf einer Skala von 1 (sehr leichte sportliche Belastung) bis 5 (maximale sportliche Belastung) an fünf Tagen gemessen.

Tag	1	2	3	4	5
$\bar{X}$	170	140	135	165	120
Z	5	4	2	3	1

i) Berechnen Sie den *Korrelationskoeffizienten nach Spearman* r_{sp} zwischen $\bar{X}$ und Z. Interpretieren Sie ihr Ergebnis.

$$r_{sp} = 1 - \frac{6\sum_{i=1}^{n}(R_{\bar{x}_i} - R_{z_i})^2}{(n-1)\cdot n\cdot(n+1)}, \quad \text{also} \quad r_{sp} = 1 - \frac{6\cdot 2}{(5-1)\cdot 5\cdot(5+1)} = 0,9$$

Auf Basis der Messungen, besteht zwischen der durchschnittlichen Herzfrequenz $\bar{X}$ und der erreichte Trainingszone Z gemäß des Korrelationskoeffizienten nach Spearman ein *gleichsinniger/positiver monotoner* Zusammenhang. Die Stärke des Zusammenhangs liegt mit einem Wert von $0,9$ nahe am Maximalwert des Wertebereichs $[-1,1]$.

Mögliche Hilfstabelle:

Tag	5	3	2	4	1
$R_{\bar{x}}$	1	2	3	4	5
R_z	1	2	4	3	5
$R_{\bar{x}} - R_z$	0	0	-1	1	0
$(R_{\bar{x}} - R_z)^2$	0	0	1	1	0

j) Berechnen Sie den *Korrelationskoeffizienten nach Kendall* τ zwischen $\bar{X}$ und Z. Interpretieren Sie ihr Ergebnis.

$$\tau := \frac{P-Q}{\frac{n(n-1)}{2}}, \quad \text{also} \quad \tau = \frac{9-1}{\frac{5(5-1)}{2}} = 0,8$$

Auf Basis der Messungen, besteht zwischen der durchschnittlichen Herzfrequenz $\bar{X}$ und der erreichte Trainingszone Z gemäß des Rangkorrelationskoeffizienten nach Kendall ein *gleichsinniger/positiver monotoner* Zusammenhang. Die Stärke des Zusammenhangs liegt mit einem Wert von $0,8$ im oberen Bereich des Werteintervalls $[-1,1]$.

Mögliche Hilfstabelle:

Tag	5	3	2	4	1
$\bar{X}$	120	135	140	165	170
Z	1	2	4	3	5
		+	+	+	+
			+	+	+
				-	+
					+