Hinweise:

- Abgabeschluss für dieses Blatt ist **Mittwoch, 02.02.2022, 23:59:59 CET**. Nur Abgaben, die rechtzeitig über die Moodle-Seite zur Veranstaltung erfolgen, werden akzeptiert und bewertet.
- Gruppenabgaben mit maximal 4 Personen pro Gruppe sind erlaubt. Alle Beteiligten erhalten identische Bewertungen. Notieren Sie deutlich lesbar die vollständigen Namen und Matrikelnummern aller Beteiligten auf der ersten Seite Ihrer abgegebenen Lösung.
- Nur nachvollziehbare Lösungen können gewertet werden. Geben Sie für alle Berechnungen auch einen formalen Ansatz an um die volle Punktzahl zu erreichen.
- Runden Sie Ihre Ergebnisse auf 2 Nachkommastellen und kürzen Sie Brüche vollständig.
- Fassen Sie sich bei Textantworten kurz und formulieren Sie präzise mit korrekter Verwendung der in der Vorlesung eingeführten Fachbegriffe.
- Grafiken und Berechnungen können Sie gegebenenfalls auch mit R erzeugen bzw. durchführen. Übertragen Sie in diesem Fall den vollständigen, lauffähigen (!) und kommentierten Code für ihre Lösung sowie den relevanten R-Output in Ihre abgegebene Lösung.
- In der Angabe bezeichnet $\log(x) = \log_e(x)$ immer den natürlichen Logarithmus zur Basis e.

Aufgabe 1 24 Punkte

Zwei stetige Zufallsvariablen X und Y haben die gemeinsame Dichte

$$f_{X,Y}(x,y) = c\exp(-2(x+y))I(x \geq 0)I(y \geq 0).$$

(a) Bestimmen Sie den numerischen Wert der Konstante c.

(b) Bestimmen Sie die Randdichte von X.

(c) Bestimmen Sie die bedingte Dichte von $Y|X$.

(d) Bestimmen Sie $\rho(X,Y)$.

(e) Berechnen Sie $P(X < 0.5Y)$.

Aufgabe 2 16 Punkte

Seien X und Y Zufallsvariablen mit $E(X) = E(Y) = 0$ und $\text{Var}(X) = \text{Var}(Y) = 25$.
Sei $W = X + Y$ und $T = X - Y$.

(a) Bestimmen Sie $\text{Var}(W)$, $\text{Var}(T)$, $\text{Cov}(W,T)$ und $\rho(W,T)$ jeweils für den Fall, dass X und Y unabhängig sind.

(b) Bestimmen Sie $\text{Var}(W)$, $\text{Var}(T)$, $\text{Cov}(W,T)$ und $\rho(W,T)$ jeweils für den Fall, dass $\rho(X,Y) = -1/4$ gilt.

(c) Warum gilt in Szenario b) $\text{Var}(W) < \text{Var}(T)$? Geben Sie eine kurze inhaltliche Begründung, nicht nur eine rein formal-mathematische.

Aufgabe 3	29 Punkte

Wir schreiben das Jahr 2167. Wie alle rechtschaffenen Marskolonist:innen legt Hodlor Hodlorsdottir ihre kompletten Ersparnisse – kümmerliche 200 Muskcoins – in Kryptowährungen und NFTs an.
150 ihrer Muskcoins investiert sie in sogenanntes “Dogethereum”. Dieses sichere Investment garantiert eine zufällige Jahresrendite R_1, die gleichverteilt zwischen 6% und 8% ist.
Die verbleibenden 50 Muskcoins werden etwas spekulativer in einen NFT-ETF namens “Enefftetteff” investiert, hier kann Hodlor von einer prozentualen Jahresrendite R_2 ausgehen, die $\mathcal{N}(\mu = 8, \sigma^2 = 4)$-verteilt ist.
Als verantwortungsvoll diversifizierende Anlegerin hat Hodlor diese Investments ausgewählt, weil ihre Renditen stochastisch unabhängig sind.

(a) Stellen Sie den Wert V, den Hodlors Portfolio ein Jahr nach ihrer anfänglichen Investition hat, als Funktion der Renditen R_1 und R_2 dar.
Was ist der Erwartungswert und die Varianz von V?

(b) Wie können Sie die Wahrscheinlichkeit, dass Hodlor insgesamt eine prozentuale Jahresrendite R zwischen mindestens l% und höchstens h% erzielt, berechnen?
Hinweis: Gefragt ist hier ein möglichst weit entwickelter und vereinfachter Ausdruck für diese Wahrscheinlichkeit. Für ein funktionierendes R-Skript, mit dem Sie diese Wahrscheinlichkeit aussimulieren oder numerisch berechnen können, gibt es bis zu 3 Bonuspunkte.

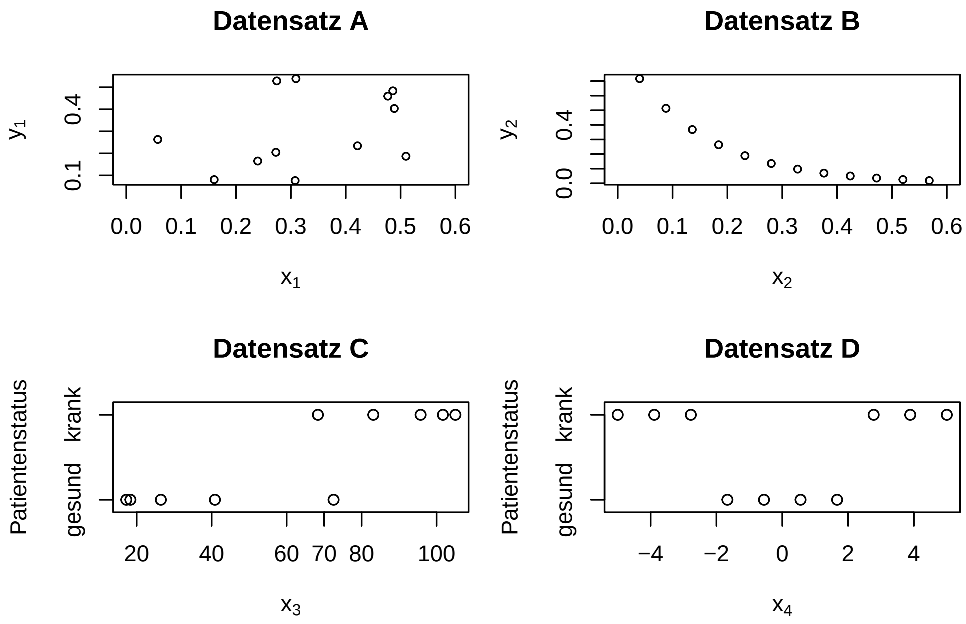
Oh nee – Marskommandantin Elon Grimes III hat mal wieder Schwachsinn über Dogethereum getwittert und die Märkte spielen verrückt. Die Korrelation zwischen R_1 und R_2 ist nun $\rho = -0.5$.

(c) Zeigen Sie, dass die Kovarianz zwischen R_1 und R_2 in diesem Szenario ca. -0.577 beträgt.

(d) Helfen Sie Hodlor, ihr Investmentkapital von 200 Muskcoins so zwischen Dogethereum und Enefftetteff aufzuteilen, dass die Varianz des Gesamtvermögens nach dem ersten Jahr (V) in diesem Szenario möglichst klein wird.
Für welche Aufteilung zwischen Dogethereum und Enefftetteff ist die Varianz von V minimal?
Wie groß ist das zu erwartende Gesamtvermögen nach einem Jahr für diese Aufteilung?

Aufgabe 4	16 Punkte

Betrachten Sie im Folgenden die unten dargestellten Datensätze.



(a) Ist der Korrelationskoeffizient nach Bravais-Pearson für die dargestellten Variablen aus den Datensätzen A und B jeweils größer, kleiner oder etwa gleich groß wie der Korrelationskoeffizient nach Spearman? Wie hoch ist der Korrelationskoeffizient nach Spearman in Datensatz B? Begründen Sie Ihre Antworten.

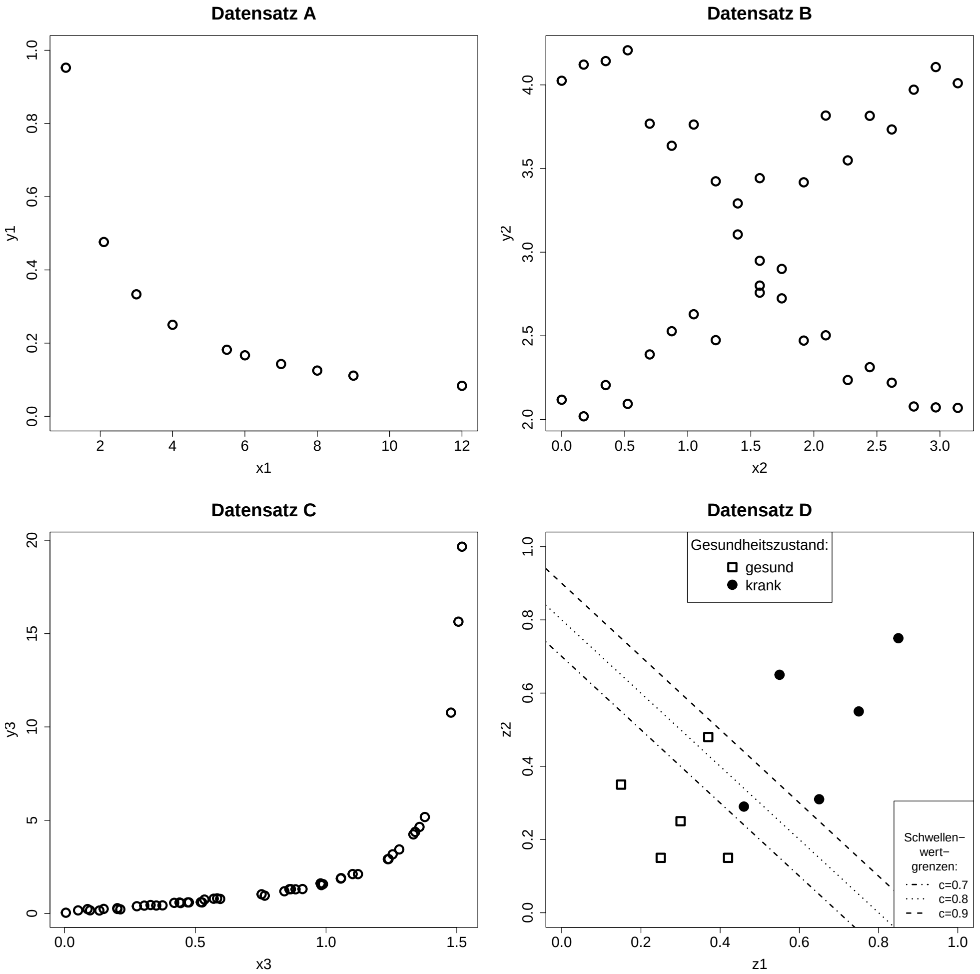
(b) Datensätze C und D zeigen auf der x-Achse jeweils einen diagnostischen Score und auf der y-Achse den beobachteten tatsächlichen Gesundheitszustand von 10 Personen ($y = 0 \Leftrightarrow$ "gesund"; $y = 1 \Leftrightarrow$ "krank"). Ist x_3 oder x_4 besser geeignet um den Gesundheitszustand der Personen vorherzusagen? Begründen Sie Ihre Antwort.

(c) Berechnen Sie Sensitivitäten und Spezifitäten eines diagnostischen Tests auf der Basis von x_3 für Schwellenwerte $c \in \{80, 70, 60\}$ und zeichnen Sie die sich daraus ergebende ROC-Kurve.

(d) Zeichen Sie die ROC-Kurve eines diagnostischen Tests basierend auf $\tilde{x}_4 = (x_4)^2$ (keine Rechnung gefragt, nur Zeichnung). Geben Sie einen Schwellenwert c für $\tilde{x}_4$ an der eine möglichst genaue Diagnose ergibt. Welche Sensitivität und Spezifität erreicht der diagnostische Test für diesen Schwellenwert?

Aufgabe 5	29 Punkte

Für vier Datensätze mit jeweils zwei bzw. drei Merkmalen liegen die folgenden Streudiagramme vor:



a) Geben Sie bezüglich der Datensätze A, B und C jeweils Auskunft über den Korrelationskoeffizienten nach Bravais-Pearson, den Rang-Korrelationskoeffizienten nach Spearman sowie über Kendall's τ. Vervollständigen Sie dazu folgenden Tabelle indem Sie für jedes Zusammenhangsmass $r \in \{r_{XY}, r_{XY}^{SP}, \tau_{XY}\}$ für jeden der Datensätze A, B und C eine der folgenden Aussagen treffen:

- r ist exakt 0: schreiben Sie "$= 0$"
- r ist exakt 1: "$= 1$"
- r ist exakt -1: "$= -1$"
- r ist nahe 0: "≈ 0"
- r ist nahe 1: "≈ 1"
- r ist nahe -1: "≈ -1"
- r ist positiv: "> 0"
- r ist negativ: "< 0"

Falls mehrere Aussagen zutreffen geben Sie jeweils die *präziseste* der zutreffenden Aussagen an.

	r_{XY}	r_{XY}^{SP}	τ_{XY}
A			
B			
C			

b) Betrachten Sie in Datensatz A statt dem Merkmal Y_1 das Merkmal $Y_1' = 2 \cdot Y_1$ bzw. das Merkmal $Y_1'' = -Y_1$. (Wie) ändern sich die Zusammenhangsmasse zwischen Y_1' und X_1 bzw. Y_1'' und X_1

jeweils gegenüber den Zusammenhangsmassen zwischen Y_1 und X_1? Begründen Sie Ihre Antworten kurz.

c) Zwei Merkmale U und V weisen einen Zusammenhang der Form $V = \frac{1}{U}$ auf. Nennen Sie eine Transformation f von U mit $\tilde{U} = f(U)$ die bewirkt dass der Korrelationskoeffizienten nach Bravais–Pearson zwischen V und $\tilde{U}$ den Wert 1 annimmt.

In Datensatz D sind insgesamt 3 Merkmale von $n = 10$ Probanden abgetragen. $Z1$ und $Z2$ sind dabei zwei klinische Messwerte die in der Summe $Z1 + Z2$ als zusammengefasster Score S betrachtet werden sollen. Die Grenzen für die Schwellenwerte $c \in \{0.7, 0.8, 0.9\}$ bezogen auf den Score S sind in dem Streudiagramm ebenfalls abgetragen.

d) Berechnen Sie jeweils Sensitivität und Spezifität in Bezug auf den Gesundheitszustand für diese Schwellenwerte von S und zeichnen Sie die ROC–Kurve.

e) Wie hoch ist hierbei der AUC–Wert?

f) Warum ist es hier nicht zielführend, $Z1$ und $Z2$ zu $S = Z1 + Z2$ zusammenzufassen um einen möglichst zuverlässigen diagnostischen Test zu entwickeln? Was für ein Vorgehen schlagen Sie stattdessen vor?