**Auszug aus der Lösung einer Statistik-Altklausur**

**Aufgabe A**

A.1

Das Konfidenzintervall für den Pearson-Korrelationskoeffizienten berechnet sich mit der Formel

$$P\left(\frac{e^A-1}{e^A+1} \le \rho \le \frac{e^B-1}{e^B+1}\right) = 1-\alpha$$

mit $A = \ln\left(\frac{1+r}{1-r}\right) - \frac{2z}{\sqrt{n-3}}$ und $B = \ln\left(\frac{1+r}{1-r}\right) + \frac{2z}{\sqrt{n-3}}$ .

Die Werte r = 0,353, n = 102 und z = 2,58 für 1-α/2 = 0,995 eingesetzt, ergibt:

$$A = \ln\left(\frac{1+0{,}353}{1-0{,}353}\right) - \frac{2*2{,}58}{\sqrt{102-3}} = 0{,}7377 - 0{,}5186 = 0{,}2191$$

$$B = \ln\left(\frac{1+0{,}353}{1-0{,}353}\right) + \frac{2*2{,}58}{\sqrt{102-3}} = 0{,}7377 + 0{,}5186 = 1{,}2563$$

Das Konfidenzintervall lautet dann:

$$\left[\frac{e^{0{,}2191}-1}{e^{0{,}2191}+1};\frac{e^{1{,}2563}-1}{e^{1{,}2563}+1}\right] = \left[\frac{0{,}24450}{2{,}24450};\frac{2{,}5125}{4{,}5125}\right] = [0{,}1091;0{,}5568]$$

A.2

$H_0$: ρ = 0

$H_1$: ρ ≠ 0.

Zur inferenzstatistischen Prüfung dieses Hypothesenpaars können prinzipiell zwei Tests verwendet werden: zum einen der t-Test auf Nullkorrelation und zum anderen der Test auf Basis von Fishers Z-Statistik. Da das Ergebnis in Teilaufgabe A.3 mit dem unter A.1 berechneten Konfidenzintervall verglichen werden soll, wird trotz $H_0$: ρ = 0 und des dafür üblicherweise eingesetzten t-Tests hier der Test auf Basis von Fishers Z-Statistik verwendet, der bei beliebigem ρ verwendet werden kann, also auch bei ρ = 0.

Die für große Stichproben standardnormalverteilte Prüfgröße z dieses Tests berechnet sich über die Formel

$$z = \frac{\sqrt{n-3}}{2}\left(\ln\left(\frac{1+r}{1-r}\right) - \ln\left(\frac{1+\rho}{1-\rho}\right)\right) .$$

Die Werte n = 102, r = 0,353 und ρ = 0 eingesetzt, ergibt

$$z = \frac{\sqrt{102-3}}{2}\left(\ln\left(\frac{1+0{,}353}{1-0{,}353}\right) - \ln\left(\frac{1+0}{1-0}\right)\right) = 4{,}9749(0{,}7377-0) = 3{,}6702.$$

Die kritischen Werte für $\alpha = 0{,}01$ zweiseitig sind $z_{krit} = z_{1-\alpha/2} = z_{1-0{,}01/2} = z_{0{,}995} = 2{,}58$ und $-z_{krit} = -2{,}58$. Da $z = 3{,}6702 > 2{,}58$ ist der Test signifikant. $H_0$ kann auf dem Niveau $\alpha = 0{,}01$ verworfen werden.

A.3

Die Fragestellung von Teilaufgabe 3.2 hätte auch mit dem Konfidenzintervall aus A.1 beantwortet werden können, da der Signifikanztest zum gleichen α-Niveau genau dann signifikant wird, wenn das Konfidenzintervall die Null nicht enthält.

A.4

Die Aufgabenstellung macht den Eindruck, es handele sich um einen Vorher-Nachher-Vergleich. Dann wäre ein Signifikanztest für abhängige Stichproben angemessen. Da aber keine Informationen über die Differenzen zwischen erster und zweiter Stichprobe gegeben sind, wird die Aufgabe dahingehend interpretiert, dass es sich um einen Vergleich zweier unabhängiger Stichproben handelt. Es kann aufgrund der Stichprobengröße $n_1 = n_0 = n = 51$ ein approximativer z-Test für die Nullhypothese $\mu_1 \leq \mu_0$ zur Anwendung kommen. Die Prüfgröße $z_{emp}$ berechnet sich zu:

$$z_{emp}=\frac{\bar{x}_1-\bar{x}_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_0^2}{n_0}}}=\frac{\bar{x}_1-\bar{x}_0}{\sqrt{\frac{1}{n}\left(s_1^2+s_0^2\right)}}=\sqrt{n}\frac{\bar{x}_1-\bar{x}_0}{\sqrt{s_1^2+s_0^2}}=\sqrt{51}\frac{0{,}2635-0{,}0734}{\sqrt{0{,}83283^2+1{,}02828^2}}=1{,}0260.$$

Der kritische Wert bestimmt sich für $\alpha = 0{,}01$ einseitig zu $z_{krit} = z_{1-\alpha} = z_{1-0{,}01} = z_{0{,}99} = 2{,}33$.

Da $z_{emp} = 1{,}0260 < z_{krit} = 2{,}33$ ist, wird die Nullhypothese beibehalten.

A.5

Die Formel der partiellen Korrelation $r_{xy.z}$ lautet

$$r_{xy.z}=\frac{r_{xy}-r_{xz}*r_{yz}}{\sqrt{\left(\left(1-r_{xz}^2\right)\left(1-r_{yz}^2\right)\right)}}$$

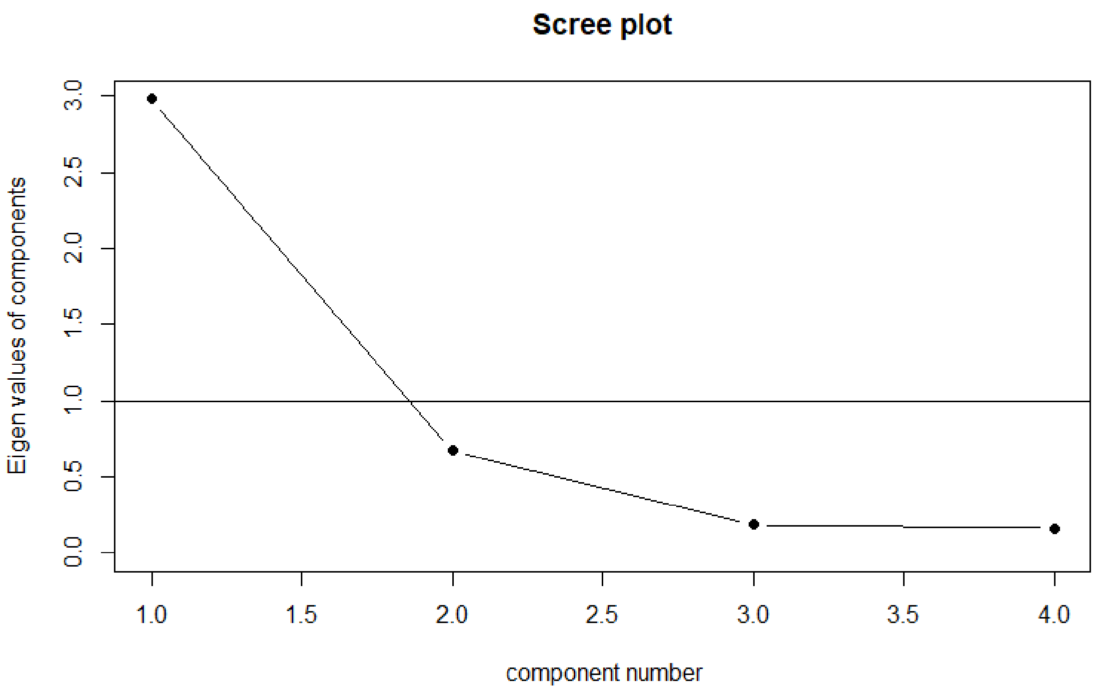
Mit $r_{xy} = 0{,}353$, $r_{xz} = -0{,}064$, $r_{yz} = 0{,}102$ ergibt sich

$$r_{xy.z}=\frac{0{,}353-\left(-0{,}064*0{,}102\right)}{\sqrt{\left(\left(1-\left(-0{,}064\right)^2\right)\left(1-\left(0{,}102\right)^2\right)\right)}}=\frac{-0{,}0065}{0{,}9980}=-0{,}0065$$

Die Korrelation zwischen X und Y unter Auspartialisierung des korrelativen Zusammenhangs dieser Variablen mit Z beträgt -0,0065, also fast gleich Null. Man kann also davon ausgehen, dass die Korrelation zwischen X und Y v.a. durch Z vermittelt wird.

**Aufgabe B**

B.1




B.2

Für den Screeplot gibt es kein eindeutiges Extraktionskriterium. Häufig sucht man einen Knick im Eigenwertverlauf und verwendet die Hauptkomponenten bzw. Faktoren links vom Knick (entweder ohne oder mit der Hauptkomponente / Faktor am Knick). Anhand des oben abgebildeten Eigenwertsverlaufs kann man einen Knick des Eigenwertverlaufs am 2. Eigenwert identifizieren. Man kann daher eine oder zwei Hauptkomponenten extrahieren.

B.3

Die Kommunalität $h_{ij}^2$ eines Items i berechnet sich zu $h_i^2 = \sum_j \lambda_{ij}^2$ mit $\lambda_{ij}$ als der Ladung des i-ten Items auf dem j-ten Faktor. Für eine Einfaktorlösung vereinfacht sich die Formel zu $h_i^2 = \lambda_{i1}^2$ . Für jedes Item ist also seine Ladung auf der 1. Hauptkomponente zu quadrieren:

Item 1 = $0{,}918^2 = 0{,}8427$.

Item 2 = $0{,}922^2 = 0{,}8501$.

Item 3 = $0{,}927^2 = 0{,}8593$.

Item 4 = $0{,}655^2 = 0{,}3025$.

B.4

Das 4. Item paßt nicht gut zum Konstrukt, da sie im Vergleich zu den anderen Variablen deutlich niedriger auf der 1. Hauptkomponente und insgesamt stärker auf der 2. Hauptkomponente lädt. Dies sieht man auch deutlich am Komponenten- (Ladungs-) diagramm.

B.5

Das faktorenanalytische Modell lautet: $\mathbf{x} = \mathbf{\Delta}\,\boldsymbol{\xi} + \boldsymbol{\varepsilon}$.

Die Varianz-Kovarianz-Matrix Cov($\mathbf{x}$, $\mathbf{x}$) kann deshalb zerlegt werden:

$$\mathrm{Cov}(\mathbf{x}, \mathbf{x}) = \mathrm{Cov}(\mathbf{\Delta}\,\boldsymbol{\xi} + \boldsymbol{\varepsilon}, \mathbf{\Delta}\,\boldsymbol{\xi} + \boldsymbol{\varepsilon}) = \mathrm{Cov}(\mathbf{\Delta}\,\boldsymbol{\xi}, \mathbf{\Delta}\,\boldsymbol{\xi}) + \mathrm{Cov}(\mathbf{\Delta}\,\boldsymbol{\xi}, \boldsymbol{\varepsilon}) + \mathrm{Cov}(\boldsymbol{\varepsilon}, \mathbf{\Delta}\,\boldsymbol{\xi}) + \mathrm{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon})$$

$$= \mathrm{Cov}(\mathbf{\Delta}\,\boldsymbol{\xi}, \mathbf{\Delta}\,\boldsymbol{\xi}) + \mathrm{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}) = \mathbf{\Delta}\,\mathrm{Cov}(\boldsymbol{\xi}, \boldsymbol{\xi})\,\mathbf{\Delta}' + \mathrm{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}) = \mathbf{\Delta}\,\mathbf{\Phi}\,\mathbf{\Delta}' + \mathbf{V}.$$

mit $\mathbf{\Delta}$ als der Ladungsmatrix, $\boldsymbol{\xi}$ als dem Vektor der Faktorwerte, $\boldsymbol{\varepsilon}$ als dem Fehlervektor, $\mathbf{\Phi}$ als der Varianz-Kovarianz-Matrix der Faktoren und $\mathbf{V}$ als der Varianz-Kovarianz-Matrix der Fehler. $\mathrm{Cov}(\mathbf{x}, \mathbf{x}) = \mathbf{\Delta}\,\mathbf{\Phi}\,\mathbf{\Delta}' + \mathbf{V}$ bezeichnet man als Fundamentaltheorem der Faktorenanalyse.