Aufgabe 1

A

Die beiden Merkmale X und Y seien diskret und metrisch. Die simultane Untersuchung beider Merkmale ergibt folgende Kontingenztafel :

X \ Y	2	3	5	6
1	1	0	0	1
3	0	1	1	0
5	0	1	1	0
7	1	0	0	1

1) Berechnen Sie den Kontingenzkoeffizienten K !
2) Geben Sie den Wert des Korrelationskoeffizienten r nach Bravais–Pearson an !
3) Erläutern Sie kurz, warum sich die unter 1) und 2) erhaltenen Ergebnisse beträchtlich voneinander unterscheiden !
4) Erläutern Sie kurz, was man unter **partieller** Korrelation versteht !

Angenommen, der erhaltene Datensatz wäre das Ergebnis einer Erhebung zum Zwecke einer linearen Einfachregressionsanalyse, wobei X die Einflussgröße und Y die Zielgröße sei.

5) Bestimmen Sie die Werte für die Parameter $\hat{a}$ und $\hat{b}$ der Regressionsgeraden $\hat{y} = \hat{a} + \hat{b}x$!
6) Bestimmen Sie den Wert des Bestimmtheitsmaßes R^2 ! Interpretieren Sie den erhaltenen Wert !

B

An einer ausgewählten Probandengruppe wurde im Rahmen einer Korrelationsanalyse überprüft, ob sich ein Zusammenhang feststellen lässt zwischen den Merkmalen

X : „ Körpergröße "　　und　　Y : „ Manuelle Geschicklichkeit " [niedrig, mittel, hoch]

Angenommen, der Wert des berechneten Korrelationskoeffizienten ergibt : + 0,89 .

1) Welche Art von Korrelation wurde hier untersucht ? Interpretieren Sie diesen Wert !

An einer ausgewählten Probandengruppe wurde im Rahmen einer Korrelationsanalyse überprüft, ob sich ein Zusammenhang feststellen lässt zwischen den Merkmalen

X : „ Schuhgröße "　　und　　Y : „Manuelle Geschicklichkeit " [niedrig, mittel, hoch]

Zu diesem Zweck wurde jedoch zuvor die gesamte Probandengruppe hinsichtlich der „Schuhgröße" in drei Kategorien eingeteilt. Die jeweiligen Korrelationskoeffizienten ergaben folgende Werte :

Schuhgröße	kleiner als 35	zwischen 35 und 43	größer als 43
Korrelationskoeffizient zwischen X und Y	**+ 0,91**	**+ 0,67**	**+ 0,09**

2) Geben Sie kurz eine Begründung, wodurch die unterschiedlichen Werte der drei Korrelationskoeffizienten zustande kommen !

C

1) Erklären Sie **kurz** den Unterschied in der Datenerhebung zum Zwecke einer Korrelations– bzw. Regressionsanalyse !
2) Worüber gibt innerhalb der Durchführung einer linearen Regressionsanalyse
   a) der Residuenplot
   b) das Bestimmtheitsmaß R^2

   Aufschluss ?
3) Zeigen Sie mit Hilfe eines formalen Beweisganges, dass innerhalb einer linearen Einfachregressionsanalyse gilt :

$$\sum_{i=1}^{n} \hat{y}_i = \sum_{i=1}^{n} y_i \qquad \text{mit} \qquad \hat{y}_i = \hat{a} + \hat{b} \cdot x_i \ !$$

Aufgabe 2

I

An jedem Tag einer Arbeitswoche (Montag bis Freitag) muss Paul eine von einer Fußgängerampel geregelte Kreuzung überqueren, wenn er die U–Bahn Station verlässt.
Diese Ampel zeigt abwechselnd 12 Sekunden lang „ grün " und 48 Sekunden lang „ rot ". Paul trifft (von Arbeitstag zu Arbeitstag unabhängig voneinander) zu einem zufälligen Zeitpunkt an dieser Ampel ein.

A

Wie groß ist die Wahrscheinlichkeit dafür, dass in einer beliebigen Arbeitswoche

1) Paul bei „ grün " an dieser Ampel eintrifft ?
2) Paul kein einziges Mal an dieser Ampel warten muss, weil diese jedes Mal „ grün " anzeigt ?
3) Paul nur einmal bei „ grün " an dieser Ampel eintrifft ?
4) Paul nur am Donnerstag bei „ grün " an dieser Ampel eintrifft ?
5) Paul am Donnerstag zum ersten Mal in dieser Woche bei „ grün " an dieser Ampel eintrifft ?

B

Angenommen, Paul würde heute bei „ rot " an dieser Ampel eintreffen.
Sei X folgende Zufallsvariable :

X : „ Wartezeit bis die Ampel „ grün " anzeigt "

1) Wie ist X verteilt ? [Verteilungstyp und –parameter !]
2) Wie groß ist die Wahrscheinlichkeit, dass Paul mindestens 30 Sekunden auf „ grün " warten muss ?
3) Wie groß ist die Wahrscheinlichkeit, dass Paul noch höchstens 24 Sekunden auf „ grün " warten muss, falls er bereits eine halbe Minute bei „ rot " an der Ampel gewartet hat ?
4) Wie lange muss Paul im Durchschnitt an dieser Ampel warten, bis endlich „ grün " kommt ?

C

Diese Ampelanlage muss aus Sicherheitsgründen gewartet werden. Sie funktioniert durchschnittlich 1000 Stunden ohne Störung (d.h. bis zur ersten Störung). Störungen treten unabhängig voneinander auf, wobei es praktisch unmöglich ist, dass mehrere Störungen gleichzeitig auftreten.

1) Wie groß ist die Wahrscheinlichkeit, dass bei dieser Ampelanlage nach der letzten Wartung
   a) innerhalb der nächsten 400 Stunden 2 Störungen auftreten ?
   b) innerhalb der nächsten 900 Stunden mehr als 3 Störungen auftreten ?
2) Wie groß ist die Wahrscheinlichkeit, dass diese Ampelanlage nach der letzten Wartung mehr als 800 Stunden störungsfrei funktioniert ?

Angenommen, die Ampelanlage hat nach der letzten Wartung bereits 700 Stunden störungsfrei funktioniert.

3) Wie groß ist die Wahrscheinlichkeit, dass es noch höchstens weitere 700 Stunden dauert, bis die nächste (erste) Störung auftritt ?

II

In einer Klausur gelten für einen beliebigen Studenten folgende Wahrscheinlichkeiten für die in der Klausur erreichte Punktzahl :

Erreichte Punktzahl	4	3	2	1
Wahrscheinlichkeit	0,12	0,06	0,24	0,18

Notenschema

4 Punkte ≡ Note 1
3 Punkte ≡ Note 2
2 Punkte ≡ Note 3
1 Punkt ≡ Note 4
0 Punkte ≡ nicht bestanden

1) Bestimmen Sie tabellarisch die Wahrscheinlichkeitsverteilung der Zufallsvariablen

   Y : „ Erreichte Punktzahl eines Studenten, wenn dieser die Klausur besteht " !
2) Berechnen Sie den Erwartungswert und die Varianz von Y !

Aufgabe 3

A

In einem Physiklabor ist ein genormter Versuch installiert. Der Gesamtfehler (in %), der pro Versuch gemacht wird, setzt sich additiv zusammen aus dem Instrumentenfehler (I), dem Ablesefehler (A) und dem Fehler durch äußere Einflüsse (ÄE).
Man kann davon ausgehen, dass I , A und ÄE unabhängige Zufallsvariablen sind mit :

$$E(I) = 10 \quad \text{und} \quad Var(I) = 5$$
$$E(A) = 5 \quad \text{und} \quad Var(A) = 1$$
$$E(ÄE) = 5 \quad \text{und} \quad Var(ÄE) = 10$$

1) Berechnen Sie den Erwartungswert und die Varianz des Gesamtfehlers (G) pro Versuch !

Angenommen von 100 Studenten führt jeder diesen Versuch genau einmal durch. Man kann davon ausgehen, dass dann der „ durchschnittliche Gesamtfehler ($\overline{G}$) pro Versuch “ normalverteilt ist.

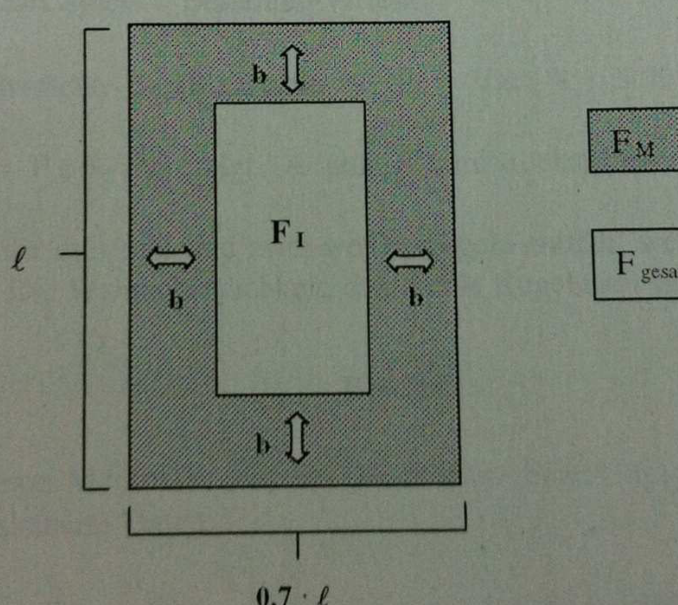
2) Wie groß höchstens (in %) ist der durchschnittliche Gesamtfehler pro Versuch mit einer Wahrscheinlichkeit von 37,07 % ?

B

Nach Murphys Gesetz geht schief, was schief gehen kann. Im Zusammenhang mit Landkarten bedeutet dies, dass die Orte, welche man auf der Karte sucht, fast immer auf den ungünstigen Stellen der Karte liegen.

Ungünstig heißt : entweder am Rand der Karte oder im Bereich der Hauptfaltung.

Zur Erklärung sei ein Atlas betrachtet, bei dem die Breite der Karten das **0,7**–fache der Höhe beträgt.
Als „ Murphy–Region “ F_M sei der gesamte äußere schraffierte Rand der Breite **b** bezeichnet :



$$\boxed{\text{Sei } b = \frac{\ell}{8}}$$

$$\boxed{F_{gesamt} = F_I + F_M}$$

Dabei kann von einer gleichmäßigen Verteilung der Orte über die gesamte Karte ausgegangen werden.
Ermitteln Sie unter Verwendung geometrischer Formeln zur Flächenberechnung

1) F_{gesamt}

2) F_I

3) F_M

Ermitteln Sie die Wahrscheinlichkeit dafür, dass ein zufällig ausgewählter Ort in der „ Murphy–Region “ liegt , d.h. berechnen Sie :

4) $P(F_M)$

Ökonomen & Wirtschaftsingenieure

Statistik I WS 2012 / 2013

Aufgabe MC

Block A

1) Der partielle Rangkorrelationskoeffizient τ nach Kendall ist eine dimensionslose Maßzahl. ✓

2) Führt man eine lineare Transformation ($x_i = a + b \cdot y_i$) eines kardinalen Datensatzes y_i durch mit $a = -\frac{\overline{y}}{s_y}$ und $b = \frac{1}{s_y}$, so gilt : $\overline{x} = 0$ und $s_x = 1$.

3) Im Pferdelotto gilt es, die vier schnellsten Pferde eines bestimmten Rennens mit ihrer Reihenfolge des Eintreffens ins Ziel vorherzusagen. Wenn insgesamt 10 Pferde an den Start gehen, gibt es somit 210 verschiedene mögliche Zieleinläufe.

4) Der Wert und das Vorzeichen des Korrelationskoeffizienten r nach Bravais–Pearson ist invariant gegenüber Lineartransformationen der Ausgangsdaten.

Block B

1) Ein Glücksspiel möge nur zwei Ausgänge haben : entweder man gewinnt 2 € oder man gewinnt nichts. Dann ist die Zufallsvariable X : „ Anzahl der gewonnenen Glücksspiele bei einmaliger Durchführung dieses Spiels “ Bernoulli–verteilt.

2) Für eine binomialverteilte Zufallsvariable X gilt : $Var(X) \leq E(X)$.

3) Aus $P(A/B) = P(B/A)$ folgt : A und B sind stochastisch unabhängig. f

4) Aus einer Urne, die drei rote und zwei weiße Kugeln enthält, werden gleichzeitig zwei Kugeln zufällig gezogen. Die Wahrscheinlichkeit, dass beide Kugeln rot sind, beträgt 20 % .

Block C

1) Wenn zwischen zwei Merkmalen nur die Gleichheits– bzw. Ungleichheitsrelation gilt, handelt es sich um nominal skalierte Daten. ✓

2) Aus n Objekten sollen r Elemente ($n > r$; $r \geq 1$) ausgewählt werden. Dann ist die Zahl der möglichen Auswahlen bei einer Variation mit Wiederholung größer als bei einer Variation ohne Wiederholung.

3) Der Wert des arithmetischen Mittels – berechnet aus bereits sortierten Daten – muss mit dem Wert des arithmetischen Mittels aus den noch nicht sortierten Daten (≅ Urliste) übereinstimmen. ✓

4) Ergibt sich bei metrischen Daten ein Rangkorrelationskoeffizient nach Kendall von $\tau = +1$, so muss der Korrelationskoeffizient nach Bravais – Pearson nicht den Wert $r = +1$ annehmen.

Block D

1) Für zwei kardinal skalierte Merkmale lässt sich stets der Rangkorrelationskoeffizient nach Kendall bestimmen.

2) Bei einer 5 x 6–Feldertafel für eine zweidimensionale Zufallsvariable Z = (X,Y) gibt es insgesamt 13 Wahrscheinlichkeitsverteilungen :
die jeweilige Wahrscheinlichkeitsverteilung (Randverteilung) von X bzw. Y ,
sechs bedingte Wahrscheinlichkeitsverteilungen (X / Y = y) und
fünf bedingte Wahrscheinlichkeitsverteilungen (Y / X = x) .

3) Bei einem ordinal skalierten Merkmal kann man nicht entscheiden, ob eine Merkmalsausprägung doppelt so groß ist wie eine andere Merkmalsausprägung. ✓

4) Eine statistische Größe bildet die Menge der Realisationsmöglichkeiten in die Menge der statistischen Einheiten ab.

Block E

1) Bei einer binomialverteilten Zufallsvariable X gilt stets : $P(4{,}9 \leq X \leq 5{,}1) \neq P(X = 5)$.

2) Eine Zufallsgröße ist eine Abbildung von der Ergebnismenge des zugrunde liegenden Zufallsexperiments in die reellen Zahlen. ✓

3) Innerhalb der Regressionsrechnung lässt sich mit Hilfe der Residuenanalyse untersuchen, ob der gewählte Ansatz für die Regressionsfunktion (z.B. nicht – linear) als erfüllt angesehen werden kann.

4) Um die Wahrscheinlichkeit des Zufallsereignisses (X < x) für beliebige x–Werte berechnen zu können, muss man von der Zufallsvariablen X die Wahrscheinlichkeitsverteilung kennen. f