

<b>Name:</b>	<b>Studiengang:</b> <input type="checkbox"/> B.A.   <input type="checkbox"/> MA.   <input type="checkbox"/> Mag.
<b>Vorname:</b>	<b>In FlexNow angemeldet:</b> <input type="checkbox"/> Ja   <input type="checkbox"/> Nein
<b>Matrikelnummer:</b>	
<b>Studienfächer:</b>	<b>Fachsemester Informationswissenschaft:</b>

**Allgemeine Hinweise:**

1. Überprüfen Sie bitte, ob Sie alle Seiten der Klausurangabe vollständig erhalten haben (Gesamtzahl: **8 Seiten**).
2. **Bearbeitungszeit: 90 Minuten**, maximal erreichbare **Punktezahl: 90**. Die jeweils erreichbare Punktezahl ist bei jeder Frage angegeben. Bitte teilen Sie Ihre Arbeitszeit entsprechend ein.
3. Denken Sie daran, die Daten oben einzutragen, **bevor** Sie mit der Bearbeitung beginnen. Bitte versehen Sie auch **jeden Papierbogen** mit Ihrem Namen und Ihrer Matrikelnummer.
4. Treffen Sie bitte die Auswahl Ihrer Antworten bei Multiple-Choice-Fragen **direkt** auf dieser Klausurangabe.
5. Verwenden Sie für die Beantwortung der Freitext-Fragen nur die offiziellen Klausurbögen für Ihre Lösungen und geben Sie an, auf welche Frage sich die Lösung jeweils bezieht. Bitte beginnen Sie für die Beantwortung einer Frage jeweils eine neue Seite.
6. Benutzen Sie keine Bleistifte, keine rot schreibenden Stifte und kein TippEx, o.ä.
7. Zugelassene Hilfsmittel: Foliensätze zur oben bezeichneten Veranstaltung sowie eigene handschriftliche oder maschinengeschriebene Notizen und Lehrbücher Ihrer Wahl. Sie dürfen diese Hilfsmittel nur auf Ihrem eigenen Arbeitsplatz ablegen.
8. Mobiltelefone sowie Computer am Arbeitsplatz - auch ausgeschaltet - sind nicht zugelassen.
9. Geben Sie keine mehrdeutigen (oder mehrere) Lösungen an. In solchen Fällen wird stets die Lösung mit der geringeren Punktzahl gewertet. Eine richtige und eine falsche Lösung ergeben also null Punkte.
10. Wenden Sie sich bei Unklarheiten in den Aufgabenstellungen immer an den Aufsichtsführenden. Hinweise und Hilfestellungen werden dann, falls erforderlich, offiziell für alle Teilnehmer durchgegeben.

Teil I) BENUTZERSTUDIEN (**MAX. 45 PUNKTE INSGESAMT**)

Hintergrund und Motivation:

Neue Studien zeigen, dass Benutzer in ihrer Freizeit im Internet oft aus Spaß oder zu ihrer Unterhaltung suchen, oder um sich die Zeit zu vertreiben. Solche Bedürfnisse stehen oft im Zusammenhang mit der Laune des Benutzers und seinem körperlichen Zustand.

Forscher A, B und C möchten eine Studie planen und durchführen, um zu lernen, welche Eigenschaften „Freizeitsuchanfragen“ haben (Länge, Frequenz, etc.), wie Benutzer mit Suchergebnissen interagieren, und wie lange Suchsitzungen dauern.

Aufgabe I.1)

Forscher A schlägt ein naturalistisches Verfahren vor. Er möchte eine Meta-Suchmaschine entwickeln, die passende Ergebnisse von YouTube, Twitter, und eine Sammlung von Nachrichten und Blogs präsentiert. Dieses System will er so breit wie möglich bekanntmachen, um eine möglichst große Teilnehmergruppe zu schaffen. Alle Systeminteraktionen sollen geloggt und analysiert werden, um zu lernen, wie sich die Benutzer verhalten.

Nennen Sie in Stichpunkten verschiedene Vor- und Nachteile dieses Verfahrens (**1,5 Punkte pro Grund, Max = 12**)

Vorteile	Nachteile

Aufgabe I.2)

Nehmen Sie zwei der Nachteile, die Sie oben genannt haben, und schlagen Sie vor, wie das Verfahren geändert oder verbessert werden könnte, um dieses Problem zu verhindern (**6 Punkte**).

Aufgabe I.3)

Forscherin B schlägt eine Laborstudie vor, in der sie das Nutzerverhalten mit zwei Systemen (Google und YouTube) untersuchen will. Die Teilnehmer sollen eine bestimmte Situation vorgegeben bekommen. Sie sollen sich in diese Situation hineindenken und sich passend verhalten. Eine Beispiel-Situation sehen Sie im Kasten unten. Interaktionen mit dem System sollen aufgezeichnet und später analysiert werden.

Beispiel:  
Stellen Sie sich vor, dass Sie daheim sind und auf einen Freund warten. Er wird Sie abholen und Sie werden zusammen ins Kino gehen. Bevor er kommt, haben Sie noch 15 Minuten Zeit. Daher gehen Sie online und suchen in Google nach etwas Unterhaltsamen.  
[Der Teilnehmer wird zu Google weitergeleitet]

Nennen Sie in Stichpunkten verschiedene Vor- und Nachteile dieser Methode (**1,5 Punkte pro Grund, Max = 12**)

Vorteile	Nachteile

Aufgabe I.4)

Wäre ein „Between groups“ oder ein „Within groups“ Design passender für die Studie der Forscherin B? Begründen Sie Ihre Antwort (**3 Punkte**).

Aufgabe I.5)

Forscher C schlägt Interviews vor, in denen Teilnehmer befragt werden sollen, wie Sie sich in solchen Situationen normalerweise verhalten. Die Fragen sollen sich auf die letzte Freizeitsuche der Teilnehmer und ihre Erfahrungen damit konzentrieren.

Nennen Sie in Stichpunkten die Vor- und Nachteile dieser Vorgehensweise (**1,5 Punkte pro Grund, Max = 6**)

Vorteile	Nachteile

## Aufgabe I.6)

Die Interviews sollen für Analysezwecke aufgenommen und transkribiert werden. Nennen Sie 3 Methoden, die geeignet sind, diese Art von Daten zu analysieren, und geben Sie Gründe an (einen pro Methode), warum die jeweilige Methode verwendet werden soll.

Die Antworten sollten formal so aussehen:

[Name der Methode; Warum würde der Forscher sie aussuchen (z.B. was könnte er mit der Methode bezwecken] **(6 Punkte)**

TEIL II STATISTISCHE DATENAUSWERTUNG (**MAX. 45 PUNKTE INSGESAMT**)Aufgabe II.1) ZUFALLSVARIABLE **5 Punkte**

Gegeben sei eine stetige Zufallsvariable  $X$ . Welche der nachstehenden Aussagen ist/sind richtig?

- A) Keine der Aussagen B) bis E) ist richtig.
- B) Die Fläche unterhalb der Dichtefunktion zu  $X$  summiert zu 1 auf.
- C) Sei  $f_X(x)$  die Dichtefunktion der stetigen Zufallsvariable. In diesem Fall liegen die Werte der Funktion  $f_X(x)$  nur im Intervall  $[0;1]$ .
- D) Sei  $f_X(x)$  die Wahrscheinlichkeitsfunktion dieser stetigen Zufallsvariable. In diesem Fall liegen die Werte der Funktion  $f_X(x)$  nur im Intervall  $[0;1]$ .
- E) Gegeben sei die Verteilungsfunktion  $F_X$  der stetigen Zufallsvariablen. In diesem Fall ist der Funktionswert  $F_X(x_0)$  die Wahrscheinlichkeit dafür, dass  $X$  den Wert  $x_0$  annimmt.

Aufgabe II.2) NORMALVERTEILUNGSANNAHME **10 Punkte**

Sie ziehen eine Stichprobe der Größe  $n=8$  der Kategorien des Online-Shops „amazon.de“ und bestimmen jeweils die Anzahl der Produkte in den jeweiligen Kategorien. Dabei ergeben sich folgende Werte:

Kategorie 1	10.686.494
Kategorie 2	19.413.280
Kategorie 3	10.586
Kategorie 4	981.979
Kategorie 5	4.232.219
Kategorie 6	1.866.832
Kategorie 7	443.176
Kategorie 8	142.796

Als statistischen Test möchten Sie einen t-Test durchführen und müssen deshalb zunächst prüfen, ob die Anzahl von Produkten je Kategorie einer Normalverteilung unterliegt.

Welchen Test zur Prüfung der Normalverteilungsannahme würden Sie weshalb anwenden?  
Warum ist die Durchführung eines solchen Tests bei diesen empirischen Daten eigentlich unnötig?

Aufgabe II.3) STATISTISCHE TESTS **5 Punkte**

Welche der folgenden Aussagen ist/sind richtig?

- A) Ein statistischer Test dient dem endgültigen Beweis der Nullhypothese.
- B) Ein statistischer Test dient zur Berechnung der Wahrscheinlichkeit der Nullhypothese.
- C) Falls eine Prüfgröße im Annahmebereich der Prüfgrößenverteilung liegt, muss die Nullhypothese beibehalten werden.
- D) Ein sog.  $\alpha$ -Fehler liegt dann vor, wenn die  $H_0$  zutrifft und daher die  $H_1$  verworfen werden muss.
- E) Keine der Aussagen A) bis D) ist korrekt.

Aufgabe II.4) SKALENARTEN **5 Punkte**

Welche der folgenden Aussage ist/sind falsch?

- A) Matrikelnummern an der Universität Regensburg sind ein Beispiel für ein metrisches Merkmal.
- B) Merkmale, die auf einer Verhältnisskala gemessen werden können, weisen für ihren Nullpunkt keine Ausprägung auf.
- C) Daten, welche auf einer Intervallskala erhoben werden können, können auch auf einer Verhältnisskala gemessen werden.
- D) Die Ausprägungen von Daten, welche auf einer Nominalskala gemessen werden, sind quantitativer Art.

Aufgabe II.5) Korrelationsrechnung **5 Punkte**

Da Ihr Kurs am 23. Dezember hätte stattfinden sollen, wurde im Rahmen einer Abstimmung durch Ihre Dozierenden der Grad erfragt (5-stufige Likert-Skala), ob Sie bereit sind, an diesem Tag im Kurs zu erscheinen. Zusätzlich wurde die Entfernung zu Ihrem Wohnort erfragt. Erläutern Sie, welches statistische Verfahren Sie weshalb anwenden würden, um einen Zusammenhang zwischen der Entfernung des Wohnortes und der Neigung zur Lehrveranstaltung am 23. Dezember zu untersuchen.

Aufgabe II.6) Hypothesen **10 Punkte**

Erläutern Sie, welche der Kriterien empirische Untersuchbarkeit, Generalisierbarkeit, Falsifizierbarkeit, Widerspruchsfreiheit, Freiheit von Tautologie, bei folgenden Aussagen erfüllt sind und welche nicht.

Sarkasmus ist in englischsprachigen Kundenrezensionen bei amazon.co.uk häufiger als in ihren deutschsprachigen Pendanten auf amazon.de

Aufgabe II.7) INTUITION UND STATISTIK **5 Punkte**

Im Rahmen eines informationswissenschaftlichen Experiments müssen die beiden Versuchsleiter für jede der 125 Teilnehmenden beurteilen, ob diese die Aufgabe, die sie erledigen sollten, erfolgreich abgeschlossen haben. Nachstehende Tabelle zeigt diese Daten, wobei das Pluszeichen die Einschätzung als erfolgreich abgeschlossen und das Minuszeichen die Einschätzung als nicht erfolgreich abgeschlossen wiedergibt.

		Rater B	Rater B	Gesamt
		+	-	
Rater A	+	118	5	123
Rater A	-	2	0	2
Gesamt		120	5	125

Es wird ein kappa-Wert von  $-0,0288$  errechnet. Nehmen Sie begründet Stellung zur Frage, weshalb dieser Wert kontraintuitiv ist.