G. Rohwer

Stichworte und Formeln für die Klausur zur Veranstaltung „Datengewinnung“ im Sommersemester 2008

1. Man sollte mit der Schreibweise $n! := 1 \cdot 2 \cdots n$ umgehen können (beachte: $0! := 1$), außerdem mit Summen- und Produktzeichen:

$$\sum_{i=1}^{n} x_i = x_1 + \cdots + x_n \quad \text{und} \quad \prod_{i=1}^{n} x_i = x_1 \cdots x_n$$

2. Man sollte folgende Begriffe aus der Mengenlehre kennen und durch Beipiele erläutern können: Menge; Element; Teilmenge; Vereinigungsmenge; Schnittmenge; Komplement einer Menge; leere Menge; Mengen, deren Elemente wiederum Mengen sind; Potenzmenge; das kartesische Produkt von zwei oder mehr Mengen; Partitionen.

3. Man sollte die Begriffe ‘Funktion’ (im mathematischen Sinn) und ‘Umkehrfunktion’ kennen, sie definieren und Beispiele angeben können. Man sollte mit Beispielen für Funktionen und Umkehrfunktionen rechnen können. Man sollte wissen, was injektive und surjektive Funktionen sind.

4. Man sollte erklären und durch Beispiele erläutern können, was logische und statistische Variablen sind und worin ihr Unterschied besteht. Man sollte insbesondere wissen und erklären können, dass statistische Variablen Funktionen sind. Man sollte Definitions- und Wertebereiche statistischer Variablen angeben können.

5. Man sollte Beispiele für ein-, zwei- und drei-dimensionale statistische Variablen angeben können.

6. Man sollte einfache Aufgaben zur Wahrscheinlichkeitsrechnung lösen können.

7. Man sollte wissen, was statistische Daten sind und wie man sie in Form einer Datenmatrix darstellen kann. In diesem Zusammenhang sollte man auch den Begriff eines realisierten Merkmalsraums definieren können.

8. Man sollte wissen, wie man statistische Daten durch Häufigkeitsverteilungen darstellen kann.

9. Man sollte den Begriff einer Relation (im mathematischen Sinn) kennen und durch Beispiele erläutern können.

10. Mögliche Eigenschaften von Relationen: reflexiv, symmetrisch und transitiv. Man sollte diese Eigenschaften definieren können. Man sollte Beispiele für Relationen angeben können, die diese Eigenschaften aufweisen bzw. nicht aufweisen. Man sollte den Begriff einer Äquivalenzrelation

kennen und durch Beispiele erläutern können.

11. Man sollte wissen, wie Relationen durch Adjazenzmatrizen oder durch Teilmengen kartesischer Produkte definiert bzw. repräsentiert werden können.

12. Man sollte den Begriff eines Graphen kennen, der aus Knoten und (gerichteten oder ungerichteten) Kanten besteht. Man sollte Beispiele für Graphen angeben können. Man sollte Graphen zeichnen können.

13. Man sollte bei ungerichteten Graphen ihre Komponenten und Cliquen bestimmen können.

14. Man sollte bei ungerichteten Graphen den Grad, bei gerichteten Graphen den Eingangs- und Ausgangsgrad ihrer Knoten berechnen können.

15. Man sollte sowohl bei ungerichteten als auch bei gerichteten Graphen ihre Dichte (Netzwerkdichte) berechnen können.

16. Man sollte wissen, wie sich Relationen und Graphen aller Art durch relationale Variablen repräsentieren lassen.

17. Man sollte den Begriff einer Abstandsfunktion kennen und durch Beispiele erläutern können. Außerdem sollte man wissen, welche Bedingungen für eine Semi-Metrik bzw. Metrik hinzukommen müssen, und auch dafür Beispiele angeben köonnen. Als Beispiele sollte man insbesondere kennen und berechnen können: die euklidische und die City-Block-Metrik (beide für Punkte in einer zweidimensionalen Ebene).

18. Man sollte wissen und durch Beispiele erläutern können, wie in der Methodenlehre die Begriffe ‘Index’ und ‘Indikator’ verwendet werden.

19. Man sollte die Unterscheidung zwischen additiven und nicht-additiven Indizes kennen und anhand von Beispielen erläutern können. Man sollte gewichtete und ungewichtete additive Indizes sowie die Unterscheidung zwischen verteilungsunabhängigen und verteilungsabhängigen Indizes kennen und anhand von Beispielen erläutern können.

20. Man sollte anhand der Formel $Y^* = Y/H^\delta$ eine mögliche Definition von Äquivalenzeinkommen erläutern können. Und man sollte Aufgaben der folgenden Art lösen können: Angenommen, man hat sich bei einer Skala zur Berechnung von Haushaltsäquivalenzeinkommen für einen Wert $\delta = 0.7$ entschieden. Wie groß müsste das Haushaltseinkommen eines 4-Personen-Haushalts sein, damit dieser zu einem 2-Personen-Haushalt mit 2800 Euro pro Monat äquivalent ist?

21. Man sollte das Prinzip der dimensionalen Homogenität kennen und seine Bedeutung für die Konstruktion additiver Indizes erläutern können.

22. Man sollte wissen und durch Beispiele erläutern können, in welcher Weise mit der Konstruktion von Indizes (fast immer) eine Datenreduktion

verbunden ist.

23. Man sollte die von Louis Guttman aufgestellte Reproduzierbarkeitsforderung für Indizes kennen und erläutern können.

24. Man sollte den Begriff einer Rangordnung (für m Alternativen) kennen und durch Beispiele erläutern können. Man sollte insbesondere wissen, wie Rangordnungen durch numerische Vektoren repräsentiert werden können. In diesem Zusammenhang sollte man auch wissen, dass strikt äquivalente Vektoren die gleiche Rangordnung repräsentieren und was inverse Rangordnungen sind.

25. Man sollte wissen, dass die Anzahl möglicher Rangordnungen für m Alternativen, wenn keine Indifferenzen zugelassen sind, mit der Formel $m!$ (m Fakultät) berechnet werden kann.

26. Man sollte die Kemeny-Metrik für Rangordnungen kennen und anhand von Beispielen berechnen können. Hier noch einmal die Definition. Gegeben sind zwei Rangordnungen $\mathbf{r} = (r_1, \ldots, r_m)$ und $\mathbf{r}' = (r'_1, \ldots, r'_m)$. Für alle Indexpaare (i, j), wobei $1 \leq i < j \leq m$ ist, werden nun zunächst folgende Größen definiert:

$$\delta_{ij}(\mathbf{r}, \mathbf{r}') := \begin{cases} 0 & \text{wenn } (r_i, r_j) =_r (r'_i, r'_j) \text{ (d.h. } r_i = r_j \text{ und } r'_i = r'_j \\ & \text{oder } r_i < r_j \text{ und } r'_i < r'_j \text{ oder } r_i > r_j \text{ und } r'_i > r'_j) \\ 2 & \text{wenn } (r_i, r_j) \neq_r (r'_i, r'_j) \text{ (d.h. } r_i < r_j \text{ und } r'_i > r'_j \\ & \text{oder } r_i > r_j \text{ und } r'_i < r'_j) \\ 1 & \text{andernfalls} \end{cases}$$

Dann wird der Kemeny-Abstand von $\mathbf{r}$ und $\mathbf{r}'$ durch $d_r := \Sigma_{i<j}\, \delta_{ij}(\mathbf{r}, \mathbf{r}')$ definiert.

27. Man sollte feststellen können, ob eine Rangordnung zwischen zwei anderen Rangordnungen liegt; und umgekehrt: zu zwei vorgegebenen Rangordnungen – wenn möglich – eine dritte Rangordnung finden, die zwischen ihnen liegt.

28. Man sollte folgende Begriffe kennen und erläutern können: Grundgesamtheit, Stichprobe, Auswahlverfahren, zufälliges Auswahlverfahren, einfache Zufallsauswahl, systematische Zufallsauswahl, geschichtete Auswahlverfahren, mehrstufige Auswahlverfahren, Clusterstichprobe.

29. Man sollte das sozialstatistische Inferenzproblem kennen und erläutern können, außerdem – in diesem Zusammenhang – auch die Problematik der Idee einer „repräsentativen Stichprobe“ erläutern können.

30. Man sollte wissen, auf wieviel unterschiedliche Weisen n Elemente in einer Reihe angeordnet werden können.

31. Wieviel Stichproben des Umfangs n können aus einer Grundgesamtheit

mit N Elementen gebildet werden? Antwort:

$$\binom{N}{n} = \frac{N!}{n!\,(N-n)!}$$

Man sollte solche Binomialkoeffizienten für einfache Zahlenwerte berechnen können.

32. Man sollte ausgehend von der gegebenen Definition eines Auswahlgenerators $\mathcal{G}$ Inklusionswahrscheinlichkeiten $\pi(\omega) = \sum_{\mathcal{S} \ni \omega} \Pr[\mathcal{G}](\{\mathcal{S}\})$ berechnen können. Analog sollte man Inklusionswahrscheinlichkeiten zweiter Ordnung $\pi(\omega, \omega')$ berechnen können.