

Hausaufgaben für LL-08, Statistik zum 2.4.2009

Aus dem Artikel „Der große Bewerbungsirrsinn“ (*Die Zeit* vom 19. März 2009 über Studieren in den USA): „Nun ist der Campus ihre Welt, die Seminare, die Hunderte Seiten Lektüre pro Tag [!], ...“.

Ein TFH-Student, der zur Zeit ein Auslandssemester in den Niederlanden macht, berichtet ebenfalls von extrem hohen Leseumfängen.

Tipp: Organisiert Eure Gruppenarbeit, um die Lesearbeit aufzuteilen und zu diskutieren.

Aufgabe 1

Lese Abschnitt 2.3 *Kumulierte Häufigkeiten und empirische Verteilungsfunktion*.

Zu Formel (2.12): Summiert werden alle die relativen Häufigkeiten, die zu Ausprägungen gehören, die kleiner oder gleich x sind. Das Symbol $:=$ bedeutet „ist per Definition gleich“, d.h. die linke Seite wird durch die rechte Seite definiert.

Zur Erklärung „empirisch“ auf Seite 23 oben: Eine ausführlichere Erklärung kann erst gegeben werden, wenn Kenntnisse der Wahrscheinlichkeitstheorie und der schließenden Statistik vorliegen. Grob: In der Praxis möchte man anhand einer Stichprobe schätzen, welches Wahrscheinlichkeitsmodell am besten die Variabilität der Daten beschreibt. Die empirische Verteilungsfunktion ist die Schätzung der Wahrscheinlichkeitsverteilung. Erstaunlicherweise findet man nur in wenigen Büchern dazu nähere Informationen (z.B. Lawson, Erjavec (2001) *Modern Statistics for Engineering and Quality Improvement*. Duxbury. Chapter 5 - Probability Plots).

Die Formel auf Seite 23 unten wird durch Aufgabe 2.6 b) verständlicher.

Bearbeite die Aufgaben 2.5 und 2.6.

Aufgabe 2

Lese Abschnitt 2.4 *Lageparameter*.

Man mache sich die Bedeutung der Formeln (2.16) und (2.17) klar. Wenn man sie richtig lesen kann, wird man erkennen, dass ihre Bedeutung dem „gesunden Menschenverstand“ entspricht.

Zu Beispiel 2.14: Es ist etwas merkwürdig, dass zwar die obere offene Klasse (sinnvoll) eingeschränkt wird, aber eine ebenso sinnvolle Einschränkung der unteren Klasse unterbleibt. Die Problematik wird zwar kommentiert, aber eine Einschränkung auf z.B. 16 Jahre erfolgt nicht.

Bearbeite Aufgaben 2.7, 2.9, 2.10 (Aufgabe 2.8 wurde in der Vorlesung bearbeitet)

Das Symbol Z für den Zentralwert ist eher unüblich, da Z später die sehr wichtige standard-normalverteilte Zufallsvariable bezeichnet. Auch ist die Bezeichnung Median viel gebräuchlicher als Zentralwert.

Eigenschaft b) auf Seite 32 mache man sich am Bild 2.5 klar (und man konstruiere ein weiteres Beispiel, in dem es eine Treppenstufe auf dem Niveau 0,5 gibt).

Bearbeite Aufgaben 2.11 und 2.12.

Die Begriffe sollten keine Verständnisschwierigkeiten bereiten. Sie sind trotzdem wichtig. Eine links-steile Verteilung wird auch oft als rechts-schief bezeichnet.

Bearbeite Aufgabe 2.13.

Aufgabe 3

Lese Abschnitt 2.5 *Streuungsparameter*.

Zur ersten Formel auf Seite 42: Sind alle Ausprägungen x_1, \dots, x_n gleich, so bedarf es keiner statistischer Methoden. Man darf also guten Gewissens davon ausgehen, dass in der Praxis $s^2 \neq 0$ ist. Da man in vielen Formeln durch diese Größe dividieren muss, ist das eine nützliche Annahme.

Zu Formel (2.39): Vgl. mit der Herleitung in der Vorlesung für σ^2 und der Anmerkung zur Fortpflanzung von Rundungsfehlern (die hier leider fehlt).

Berechne Beispiel 2.22 auch mit der ursprünglichen Formel (2.38) und der entsprechenden Tabelle (wie in der Vorlesung). Man sollte beide Berechnungsarten beherrschen. Je nach Zahlenmaterial ist die eine oder die andere Methode besser. Zum Lernen sollte man die Tabelle mit Papier und Taschenrechner erstellen; als Vorbereitung für die Praxis sollte man dieselbe Tabelle mit Excel erstellen.

Wichtig: Bemerkung b) auf Seite 44. Wir benötigen beide Formeln - die Varianz der Grundgesamtheit σ^2 und die der Stichprobe s^2 .

Bearbeite Aufgabe 2.15.

Aufgabe 4

Diese Aufgabe bezieht sich auf das E-Book *Einführung in die Statistik mit Excel und SPSS* von Duller (Springerlink).

Lese Abschnitt 6.4 *Die empirische Verteilungsfunktion*.

Zu Abbildung 6.20. Die senkrechten Verbindungslinien gehören streng genommen nicht zum Graph. Auch sollte deutlich gemacht werden, welcher Endpunkt der Treppenstufe zum Graph gehört und welcher nicht.

Aufgabe 5

(Gute Wiederholung des Stoffes. Die nahezu gleiche Themenwahl zeigt die grundlegende Bedeutung der Begriffe. Nachdem man den Sachs aufmerksam gelesen hat, sollte man dieses Kapitel schnell lesen können.)

Lese im E-Book von Duller Kapitel 7 *Maßzahlen*. Die Abschnitte zu SPSS können übersprungen werden. Es genügt, die Informationen zur Schiefe und Wölbung zu überfliegen.

WICHTIG: Dieses Buch verwendet die Bezeichnung s^2 für unser σ^2 . Bitte nicht verwirren lassen.