Dr. G. Eichner
Math. Inst.
JLU Gießen

Übungen zu

R1: Grundlagen der Datenanalyse mit R

Blatt 8

SoSe 2024
13. 6. 2024
Abgabe: ≤3. 7. 2024, 14:00 Uhr

Beachten Sie die Bearbeitungszeit von *drei* Wochen für dieses Blatt!

1. Bestimmen Sie für die folgenden Ereignisse ihre Eintrittswahrscheinlichkeiten:

   a) Eine normalverteilte Zufallsvariable mit Erwartungswert 105 und Varianz 25 nimmt höchstens den Wert 112 an.

   b) Eine mit 7 Freiheitsgraden t-verteilte Zufallsvariable ist betraglich höchstens gleich 2.1.

   c) Eine auf $(-2, 2)$ uniformverteilte Zufallsvariable erreicht oder überschreitet den Wert -1.8 nicht.

   d) Eine F-verteilte Zufallsvariable mit 7 Zähler- und 5 Nennerfreiheitsgraden ist größer als 6.3.

2. Fragen nach Eintrittswahrscheinlichkeiten mit etwas Anwendungsbezug:

   a) Wie groß ist die Wahrscheinlichkeit, dass in einer Sequenz von 20 unabhängigen Würfen einer (unfairen) Münze, bei der die (konstante) Wahrscheinlichkeit für „Kopf" 2/3 ist, genau zehnmal „Kopf" geworfen wird?

   b) Ein Kunde kauft genau vier Sicherungen aus einer Lieferung von insgesamt 250 Stück. Die Ausschussquote der Lieferung betrage 6 %. Mit welcher Wahrscheinlichkeit erhält der Kunde keine einzige, genau eine, genau zwei, genau drei bzw. lauter defekte Sicherung(en)?

   c) Mit welcher Wahrscheinlichkeit kommt beim Roulette frühestens nach fünf Spielen die Farbe Rot?

   d) Die Analyse der Auslastung der Notaufnahme eines Krankenhauses ergab, dass die tägliche Anzahl an Patientinnen oder Patienten, die zwischen 18 und 19 Uhr eintreffen im (langfristigen) Mittel 6,9 beträgt. Wie groß ist die Wahrscheinlichkeit, dass an einem gegebenen Tag zwischen 18 und 19 Uhr mindestens zehn Patientinnen oder Patienten eintreffen?

   e) Ein „etabliertes" Diagnoseverfahren zur Abklärung einer speziellen medizinischen Fragestellung provoziere bekanntermaßen bei 10 % aller Patientinnen oder Patienten eine Unverträglichkeitsreaktion. Es werde ein neues, angeblich besseres Verfahren erprobt und 25 voneinander unabhängige Patientinnen oder Patienten diesem Verfahren unterzogen. Dabei trete nur bei einem von ihnen die erwähnte Reaktion auf.

      Mit welcher Wahrscheinlichkeit wäre ein mindestens so gutes Ergebnis bei dem etablierten (!) Verfahren zu beobachten gewesen?

   f) Die Lebensdauern in Jahren von Bauteilen eines gewissen Typs seien identisch exponentialverteilt zum Parameter $\lambda = 1/2$ (sodass ihre erwartete Lebensdauer $1/\lambda = 2$ Jahre beträgt). Drei dieser unabhängig arbeitenden (irreparablen) Bauteile seien in ein Sicherheitssystem eingebaut.

      Mit welcher Wahrscheinlichkeit ist die größte der drei Lebensdauern kleiner als 1, sprich fallen alle drei Bauteile noch vor dem Ende des ersten Jahres aus?

3. Quantile — theoretische und empirische:

   a) Wie lauten die oberen und unteren 5 %-, 2.5 %-, 1 %-, 0.5 %- und 0.1 %-Quantile der t-Verteilung mit 4 Freiheitsgraden? Wie lauten erstes und drittes Quartil sowie Median dieser Verteilung?

   b) Wie lauten die entsprechenden Quantile der Mietdaten aus Aufgabe 3 auf Blatt 6?

4. Die mathematische Signumsfunktion sgn ist für $x \in \mathbb{R}$ definiert durch

$$sgn(x) = \begin{cases} 1, \text{ falls } x > 0; \\ 0, \text{ falls } x = 0; \\ -1, \text{ falls } x < 0. \end{cases}$$

   Implementieren Sie in **R** eine Funktion namens `sgn` so, dass für einen `numeric`-Vektor `x` der Aufruf `sgn(x)` den Vektor der Signumswerte seiner Elemente zurückliefert!

   Versuchen Sie dies zum einen unter Verwendung einer `for`-Schleife, zum anderen mit `sapply` und zum dritten in vektorisierter Version (also ohne eine `for`- oder sonstige Schleife)!