Aufgabe C2 Landesabitur Hessen 2011 LK

In dieser Aufgabe geht es um das Benfordgesetz. Der amerikanische Physiker Benford hat entdeckt, dass bei großen Zahlenmengen die 1 als Anfangsziffer häufiger vorkommt als beispielsweise die 9 und aus seinen Beobachtungen folgende Gesetzmäßigkeit hergestellt:
Die Wahrscheinlichkeit p(k), dass in einer großen Zahlenmenge eine Zahl mit der Ziffer k beginnt berechnet sich zu $p(k) = \log_{10}\left(\frac{k+1}{k}\right)$

Aufgabe 1.1 (3 BE)

Ergänze die folgende Tabelle so, dass p(k) für $1 \le k \le 9$ eine Wahrscheinlichkeitsverteilung ist.

Anfangsziffer k	1	2	3	4	5	6	7	8	9
Wahrscheinlichkeit p(k)	0,30	0,18	0,12	0,10	0,08	0,07			

Für die gelben Felder gilt $p(7) = \log_{10}\left(\frac{7+1}{7}\right) = \log_{10}\left(\frac{8}{7}\right) \approx 0,058$

P(8) = 0,051 p(9) = 0,046

Die Summe aller Wahrscheinlichkeiten ergibt 1,01, die geringe Abweichung von der 1 lässt sich aus Rundungsfehlern erklären. Es gibt aber auch einen exakten Beweis über eine sogenannte Teleskopsumme:

$$\sum_{k=1}^{9} \log_{10}\left(\frac{k+1}{k}\right)$$
$$= \log_{10}\left(\frac{2}{1}\right) + \log_{10}\left(\frac{3}{2}\right) + \log_{10}\left(\frac{4}{3}\right) .... + \log_{10}\left(\frac{9}{8}\right) + \log_{10}\left(\frac{10}{9}\right)$$
$$= \log_{10}(2) - \log_{10}(1) + \log_{10}(3) - \log_{10}(2) + ... + \log_{10}(9) - \log_{10}(8) + \log_{10}(10) - \log_{10}(9)$$
$$= \log_{10}(10) - \log_{10}(1)$$
$$= 1$$

Aufgabe 1.2 (8 BE)

Die Binomialverteilung kann zu Grunde gelegt werden, der Erwartungswert ergibt sich dann zu $\mu = 100 \cdot 0{,}30 = 30$.

X sei die Anzahl der Zahlen mit Ziffer 1 am Anfang, dann gilt:

$$p(A) = p(X=30) = \binom{100}{30} \cdot 0{,}3^{30} \cdot 0{,}7^{70} \approx 0{,}0868$$

und mit $\sigma = \sqrt{100 \cdot 0{,}7 \cdot 0{,}3} \approx 4{,}6$

folgt für die Sigma-Umgebung $25{,}4 \leq X \leq 34{,}6$:

$$p(26 \leq X \leq 34) = F(100;0{,}3;34) - F(100;0{,}3;25) \approx 0{,}8371 - 0{,}1631 = 0{,}6740$$

Statt über die Tabelle der kumulierten Binomialverteilung kann das Ergebnis auch über den Taschenrechner erhalten werden. In diesem Fall sollte der Term $\sum_{x=26}^{34} \binom{100}{x} \cdot 0{,}3^{x} \cdot 0{,}7^{100-x}$

notiert werden.

Aufgabe 2.1 (10 BE)

Festlegen der Hypothesen bei n = 100

H_0 : p_0 = 0,3 die Nullhypothese beschreibt den Normalfall, das was bekannt ist bzw. das was man erwarten würde

H_1 : p_1 < 0,3 die Alternativhypothese beschreibt in diesem Fall die Vermutung, dass zu wenig Einsen an der ersten Stelle stehen.

H_0 wird abgelehnt, wenn sehr wenige Einsen auftreten, also ist ein linksseitiger Test angemessen mit einem vorgegebenen Signifikanzniveau von 5 %. Es gilt:

$p_{0,3}(X \le k) = F(100;0,3;k) \le 0,05$ wird für k = 22 wegen $p_{0,3}(X \le 22) = 0,0479$ zum ersten Mal unterschritten.

Das heißt, wenn weniger als 23 Einsen auftreten, lehnt man die Nullhypothese ab und vermutet eine Fälschung.

Mögliche Fehler bei dieser Entscheidung sind:

α - Fehler: Ho wird abgelehnt, obwohl die Hypothese zutreffend ist. Eine ehrlich angefertigte Bilanz wird zu Unrecht als gefälscht eingestuft.

β - Fehler: Die Nullhypothese Ho wird angenommen obwohl sie falsch ist. Eine gefälschte Bilanz wird als korrekt angesehen.

Aufgabe 2.2 (4 BE)

Eine Fälschung soll nicht erkannt werden. Es handelt sich also um einen β - Fehler.
β - Fehler können nur berechnet werden, wenn eine alternative Wahrscheinlichkeit vorliegt.
Dies ist hier der Fall, denn laut Angabe beginnt nur jede 5. Zahl mit der Anfangsziffer 1, für p_1 gilt also p_1 = 0,2. Daraus folgt:

β -Fehler : $p_{0,2}(X \ge 23) = 1 - p_{0,2}(X \le 22) = 1 - F(100;0,2;22) \approx 0,2611$

Mit einer Wahrscheinlichkeit von rund 26,1 % fällt diese Bilanzfälschung nicht auf.

Aufgabe 3 (5 BE)

In diesem Teil der Aufgabe wird nach Möglichkeiten gefragt, den als zu hoch empfundenen β - Fehler aus 2.2 zu reduzieren.

Die nächstliegende Lösung des Problems besteht darin, den α - Fehler zu vergrößern. Dadurch rückt die kritische Zahl k = 22 weiter nach „rechts“ und die Wahrscheinlichkeit des β - Fehlers, z.B. mehr als 25 Einser zu erhalten, nimmt bei konstantem p_1 = 0,2 natürlich ab, weil der Annahmebereich der Nullhypothese verkleinert wird.

Eine bessere Lösung wäre es natürlich, den Stichprobenumfang n zu erhöhen, z.B. auf n = 200 Zahlen. Dabei verkleinert man den β -Fehler, kann aber auch den α - Fehler klein halten. Wie schon in der Lösung zur diesjährigen Abitursaufgabe C1 ersichtlich, können Verteilungen, bei denen p_1 nur wenig von p_0 verschieden ist, bei größerem n schärfer getrennt werden. Die überlappenden

Bereiche beider Kurven werden bei größerem n immer kleiner, so dass die Verteilung für p_1 weniger in den Annahmebereich der Nullhypothese hineinragt.