Fachbereich Mathematik und Informatik
der Universität Marburg
Dr. Helga Lohöfer

Sommersemester 2012

Mathematik für Natur- und Sozialwissenschaftler

- Blatt 8 -

Abgabe: Mittwoch, den 13.06.2012, 12:10 Uhr vor der Vorlesung

Skript und Aufgabenblätter: Im Internet unter www.mathematik.uni-marburg.de/~lohoefer/
Lektürehinweis: Skript **„**Mathematische und statistische Methoden**"** Kap 7.3, 8.1, 9.3.4
Wichtige Begriffe: Dichtefunktion einer Variablen, Normalverteilung, Rechnen mit der Φ-Tabelle, Normalverteilung anstelle von Binomialverteilung bei zahlreichen Messungen.

1. Eine repräsentative Erhebung zur monatlichen Einkommenshöhe X in einer Kommune ergab bei 2800 befragten Haushaltungen je nach Feinheit der Einteilung in Merkmalsklassen E_i die folgenden absoluten Häufigkeiten H_i :

(I)				(II)			
E_i (in €)	H_i	h_i	p_i	E_i (in €)	H_i	h_i	p_i
0 – 600	28			0 – 300	0		
				300 – 600	28		
600 – 1200	252			600 – 900	84		
				900 – 1200	168		
1200 – 1800	644			1200 – 1500	280		
				1500 – 1800	364		
1800 – 2400	728			1800 – 2100	392		
				2100 – 2400	336		
2400 – 3000	280			2400 – 2700	196		
				2700 – 3000	84		
3000 – 3600	196			3000 – 3300	56		
				3300 – 3600	140		
3600 – 4200	476			3600 – 3900	224		
				3900 – 4200	252		
4200 – 4800	196			4200 – 4500	168		
				4500 – 4800	28		
> 4800	0			> 4800	0		

a) Tragen Sie in die obige Tabelle alle relativen Häufigkeitkeiten $h_i = h(E_i)$ ein $(0 \le h_i \le 1)$. (1,5)
b) Skizzieren Sie in einem gemeinsamen Koordinatensystem (Maßstab x-Achse: 1cm = 300€, y-Achse: 1cm = 0,10) die 2 Histogramme der relativen Häufigkeiten (gestrichelt = (I), durchgezogen = (II)). (2)
c) Tragen Sie in obige Tabelle auf 6 Nachkommastellen gerundet alle empirischen Dichten p_i ein. (1,5)
d) Skizzieren Sie in einem gemeinsamen Koordinatensystem (Maßstab: x-Achse: 1cm = 300€, y-Achse: 2,5cm = 0,000100) die 2 Graphen der empirischen Dichten (gestrichelt = (I), durchgezogen = (II)). (2)

2. In einem Staat rechnete das Wehrbeschaffungsamt lange Zeit mit folgenden statistischen Daten: Die Körpergröße X bei Männern galt als normalverteilt mit Erwartungswert μ = 180,5 cm, Standardabweichung σ = 8,0 cm.
   a) Wie viele Uniformen für Körpergröße 170cm $\le$ x $\le$ 175cm wurden benötigt, um eine Armee von 350 Tausend Mann einzukleiden? Tipp: Berechnen Sie zunächst $P(170 \le x \le 175)$. (2)
   b) Seither habe sich der Erwartungswert auf 183,2cm und die Standardabweichung auf 8,6cm vergrößert. Wie groß ist nun der Bedarf an Uniformen für Körpergröße 170cm $\le$ x $\le$ 175cm, wenn gleichzeitig die Armeestärke nur noch 185 Tausend Mann beträgt? (2)

bitte wenden

3. In einer gewissen Weltregion beträgt die Chance, älter als 50 Jahre zu werden, nur 70%. Wie wahrscheinlich ist es (in %, 1 Nachkommastelle) bei gleichbleibender Lebenserwartung, dass dort von 300 Kindern einer Schule mindestens 220 dieses Alter erreichen werden? (4)

Klausurtermine:
1. Klausur: Do, 19.07.12, 11:00 – 12:45, Hörsaalgebäude Chemie Lahnberge, Hans-Meerwein-Str., HS A + B
2. Klausur: Do, 11.10.12, 11:00 – 12:45, Hörsaalgebäude Chemie Lahnberge, Hans-Meerwein-Str., HS B

Stabdiagramm oder **Säulendiagramm:** Treten in einer Messreihe zur Variablen X nur wenige verschiedene Messwerte auf, diese aber vielfach, so wird für jeden dieser Messwerte die absolute Häufigkeit H oder die relative Häufigkeit h in einer Säule der Höhe H oder h visualisiert. Die Säulen stehen getrennt nebeneinander, ihre Breite ist irrelevant.

Histogramm: Treten in einer langen Messreihe zur reellwertigen Variablen X sehr viele verschiedene Messwerte auf, so wird vergröbernd die x-Achse in einige Intervalle (= Merkmalsklassen) E_i (i=1,…,r) unterteilt, für jede Merkmalsklasse wird die absolute Häufigkeit H oder die relative Häufigkeit h der in ihr liegenden Messwerte bestimmt und über jedem Intervall eine Säule der Höhe H oder h errichtet. Die r Säulen stehen ohne Zwischenraum eng nebeneinander. Die Lage der Intervallgrenzen und die Breite der Intervalle entscheidet wesentlich über die jeweilige Säulenhöhe. Die Optik des Histogramms hängt also sehr stark von der willkürliche Einteilung der x-Achse ab und ist somit manipulierbar.

Dichtefunktion: Dies wird überwunden durch durch das Konzept der **empirischen Dichte** $p = \frac{h}{\text{Säulenbreite}}$ = relative Häufigkeit der Messwerte pro Einheit von X im Bereich des jeweiligen Intervalls. Ist x_i der Mittelpunkt des Intervalls E_i und p_i die zugehörige Dichte, dann verbindet man die Punkte $(x_i \mid p_i)$ (i=1,…,r) durch eine gebogene Linie zu einem Funktionsgraph. Verlängert man immer weiter die Messreihe ($n \to \infty$) und verkleinert dabei alle Merkmalsklassen ($r \to \infty$), so strebt die Folge dieser Graphen gegen den Graph einer Grenzfunktion:

y = p(x) = die theoretische Dichtefunktion von X.

Dies ist eine spezifische Funktion von X, und es gilt stets: Die Wahrscheinlichkeit, dass eine Einzelmessung von X einen Wert im Bereich $a \le x \le b$ liefert, ist gleich

$$P(a \le x \le b) = \int_a^b p(x)dx$$

Der Graph der Dichtefunktion kann eingipflig oder mehrgipflig sein, bei Eingipfligkeit schief oder symmetrisch.

Normalverteilung: Bei sehr vielen Variablen X, die im Wert von zahlreichen, aber jeweils nur schwachen Einflussgrößen abhängen, ist der Graph der Dichtefunktion eingipflig und symmetrisch („glockenförmig“), und wenn μ den Erwartungswert, σ die Standardabweichung von X bezeichnet, besitzt die Dichtefunktion von X die Berechnungsformel

$$p(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-0{,}5 \cdot \left(\frac{x-\mu}{\sigma}\right)^2}$$ X heißt in diesem Fall „normalverteilt“.

Diese Dichtefunktion besitzt als Stammfunktion $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ (siehe **Φ-Tabelle**), so dass folgt:

$$P(a \le x \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right), \text{ speziell } P(x \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - 0 \text{ und } P(a \le x) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Schätzformel für zahlreiche Messungen (Ersetzung der Binomialverteilung durch die Normalverteilung):

Es liege ein Fall von Binomialverteilung vor und die Anzahl n der Messungen sei derart groß, dass folgende Bedingung erfüllt ist (nachprüfen!):

$$n > \frac{9}{p \cdot (1-p)}$$

Dann gilt in guter Näherung die Schätzformel

$$P(a \le x \le b) = \sum_{k=a}^{b} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \approx \Phi\left(\frac{b+0{,}5-\lambda}{\sigma}\right) - \Phi\left(\frac{a-0{,}5-\lambda}{\sigma}\right)$$

Dabei ist $\lambda = n \cdot p$ und $\sigma = \sqrt{\lambda \cdot (1-p)}$. Diese Schätzung wird umso besser, je größer n ist.