Dr. Quapp: Statistik für Mathematiker mit SPSS

Hinweise zur 11. Übung – Stichproben –

1.] Für die Gewichte von Warenpackungen wird angenommen, dass sie $N(\mu, \sigma)$-verteilt sind mit unbekannten μ und σ. Es ergaben sich für eine Stichprobe vom Umfang 10 folgende Gewichte (in kg):
20,40, 20,25 20,00, 19,80, 20,05,
19,90, 20,50, 20,15, 20,20, 20,10.
Man bestimme ein Schätzintervall der Form $[a, \infty)$ für μ zum Niveau $0,99$.

Die Teststatistik in diesem Fall ist der T-Test

$$Y_n = \frac{\sqrt{n}\,(\bar{x}_n - \mu)}{\sqrt{s_n^2}} \sim t_{n-1} \quad \text{mit } s_n^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \ .$$

Es ist

$$1 - \alpha = P(Y_n \leq t_{n-1,\,1-\alpha}) = P(\bar{x}_n - t_{n-1,\,1-\alpha} \, \frac{s_n}{\sqrt{n}} \leq \mu)$$

Es ist somit das Intervall $[\bar{x}_n - t_{n-1,\,1-\alpha} \frac{s_n}{\sqrt{n}} \leq \mu, \ \infty)$ gesucht. Um in $-->$ *Explorative Datenanalyse* das einseitige Konfidenz-Intervall zu berechnen, kann man den Trick anwenden, ein doppelt so großes zweiseitiges Intervall von 2% zu betrachten, was ja 1% auf jeder Seite bedeutet. SPSS ergibt 98% CI für den Mittelwert $(19.9434; 20.3266)$. Eine Tabelle für die T-Verteilung ergibt mit $n = 10$ den kritischen Wert $t_{9;\,0.01} = 2.821$. Der Mittelwert der Daten ist 20.135, die Standardabweichung ist 0.2148, der Standardfehler also 0.0679. Die gesuchte linke Intervallgrenze ist also 20.135 - 2.821 $\sqrt{\frac{1}{10}}$ 0.2148 = 19.943 .

2.] Geben Sie folgendes Programm in SPSS ein .

```
MATRIX.
COMPUTE a={2,10,3,5;1,15,2,4;3,12,5,3;2,20,4,5}.
PRINT a.
COMPUTE c={1180;1001;1507;1574}.
Print c.
COMPUTE x=INV(a)*c.
PRINT x.
compute y=a*x.
Print y .
END MATRIX.
```

Was bewirkt das Programm? Kann man die Eigenwerte einer Matrix berechnen?

Das Programm bewirkt die Lösung eines linearen Gleichungssystems.
Mit einer symmetrischen Matrix b kann man auch Eigenwerte und Eigenvektoren berechnen:

```
MATRIX.
COMPUTE b={2,10,3,5;10,15,12,4;3,12,5,4;5,4,4,5}.
PRINT b.
CALL EIGEN(b,evec,ewerte) .
PRINT evec.
PRINT ewerte.
END MATRIX.
```

3.] Die Restlebenszeit X einer Person sei eine Zufallsgröße. Es sei bekannt, daß

$$P(X \geq t) = 1 - \left(\frac{t}{100}\right)^a, \quad 0 < t < 100.$$

a) Berechnen Sie die Dichte und den Erwartungswert von X.
b) Konstruieren Sie mit Hilfe der Momentenmethode eine Schätzung für a.
c) Unter Verwendung der Verteilungsfunktion von X kann man mit der Inversionsmethode Zufallszahlen erzeugen, die die Verteilung von X simulieren. Bestimmen Sie damit mit SPSS die Schätzung

$$\hat{a} := \frac{1}{100/\bar{X} - 1}$$

für a auf Grund einer Stichprobe vom Umfang $n = 2001$.

a) Es ist

$$P(X \geq t) = 1 - P(X < t) \quad \text{also ist} \quad F(x) = \left(\frac{x}{100}\right)^a,$$

also die Dichte

$$f(x) = a\left(\frac{x}{100}\right)^{a-1} \frac{1}{100}.$$

Der Erwartungswert ist dann

$$E\,X = \int_0^{100} (1 - \left(\frac{t}{100}\right)^a)dt = t - \frac{1}{a+1}\left(\frac{t}{100}\right)^{a+1} 100\Bigg|_0^{100} = 100\,\frac{a}{a+1}.$$

b) Verwendet man diese Resultat, kann man als Schätzung für a die Gleichung herumdrehen:

$$\hat{a} := \frac{1}{100/\bar{X} - 1}\ .$$

c) Wir konstruieren eine Verteilung für $F(x)$. Sei z.B. $a = 1/2$ gewählt. Zur Anwendung der Inversionsmethode bestimmen wir noch die Umkehrfunktion von $F(x) = \sqrt{x}\,/10$ zu

$$F^{-1}(u) = 100u^2$$

Berechnet man für die gefragten 2001 Zufallswerte vorerst $U_i \sim U(0,1)$, d.h. gemäß einer Gleichverteilung, dann sind die Zahlen

$$x_i = F^{-1}(u_i)$$

die gesuchten Zufallszahlen mit der Verteilung F. Im SPSS-Test ergab sich bei n=100: $E\,X$=32.54, und bei n=2004: $E\,X$=32.5524 (statt des theoretischen 100/3), somit ist die Schätzung $\hat{a} = 0.483$. Mittels eines Histogrammes kann man sich diese Werte veranschaulichen. Es sollte sich ein analoges Bild ergeben, wenn man $f(x) = 1/(20\sqrt{x})$ direkt über einer Achse von (0,100) zeichnet.