3. Anpassung der Niederschlagsdaten

Korrelations- und Regressionsanalysen setzen n-dimensionale Normalverteilungen voraus, deren Parameter aus der Stichprobe geschätzt werden (*Balzer*, 1979). Während die Verteilungen der Klimamitteltemperaturen von Berlin-Dahlem dieser Bedingung für jeden einzelnen Kalendermonat gerecht werden (*Dettmann*, 2000), war dieses für die Niederschlagsmengen nicht zu erwarten. Allein das Vorhandensein einer unteren Schranke läßt auf eine andere Verteilung schließen. Wie *Wilks* (2000b) zeigte, lassen sich monatliche bzw. jährliche Niederschlagsmengen vielmehr durch eine Gammaverteilung approximieren. Um dennoch Niederschlagsprognosen auf der Basis von Regressionen erstellen zu können, wurde versucht, die vorliegenden Zeitreihen der Normalverteilung anzupassen. Zu diesem Zweck wurden aus den täglichen Niederschlagsmengen die Zahl der Tage mit meßbarem Niederschlag in jedem Monat konstruiert (vergl. Abschnitt 2.2). Diese Vorgehensweise erscheint zunächst paradox, da die resultierenden Verteilungen sogar zweiseitig begrenzt sind (weniger als 0 bzw. mehr als 28, 29, 30 oder 31 Tage mit Niederschlag sind nicht möglich). Im Vergleich zu den Monatssummen weisen diese Verteilungen jedoch eine deutlich geringere Schiefe auf; zusätzlich sind die extremen Klassen schwächer besetzt. Für beide Verteilungstypen wurde daher die Anpassung an die Normalverteilung mittels **Kolmogoroff-Smirnoff-Test** geprüft (vergl. *Sachs*, 1997). Dieser ist verteilungsunabhängig und setzt lediglich die Stetigkeit der Grundgesamtheit voraus. Er entspricht dem χ^2-Anpassungstest, ist zur Aufdeckung von Abweichungen in der Verteilungsform jedoch besser geeignet. Die zu prüfende Nullhypothese H_0 lautet: Die Stichprobe entstammt der theoretisch berechneten Verteilung $F_0(x)$. Überschreitet die Testgröße

$$D = \frac{\max\left|F_B - F_E\right|}{n} \tag{3.1}$$

eine bestimmte kritische Schranke, so ist sie zu verwerfen. Hierbei sind F_E bzw. F_B die Summenhäufigkeiten der erwarteten bzw. beobachteten absoluten Häufigkeiten, n gibt den Stichprobenumfang an.

Wird der **Kolmogoroff-Smirnoff-Test**, wie im vorliegenden Fall, speziell zur Anpassung an eine Normalverteilung durchgeführt, deren Mittelwert und Varianz aus der Stichprobe geschätzt werden müssen, so sind die resultierenden Ergebnisse sehr konservativ. Daher wurde hier die als **Lilliefors-Modifikation** bezeichnete Variante dieses Tests verwendet. Sie ist schärfer als die ursprüngliche Version, d.h. sie lehnt die Nullhypothese wesentlich leichter ab. Für $n > 30$ lassen sich danach die kritischen Schranken $S(\alpha)$ mit

$$d_n = \left(\sqrt{n} - 0{,}01 + 0{,}83/\sqrt{n}\right) \tag{3.2}$$

wie folgt approximieren: $S(0{,}20)=0{,}741/d_n$; $S(0{,}15)=0{,}775/d_n$; $S(0{,}10)=0{,}819/d_n$; $S(0{,}05)=0{,}895/d_n$; $S(0{,}01)=1{,}035/d_n$. Hierbei gibt α die Irrtumswahrscheinlichkeit des Tests an.

Zur Bestimmung der Anzahl der benötigten Klassen k wurde die empirische Formel von **Pankofsky** und **Brier** herangezogen. Sie lautet:

$$k = 5 \log n \, . \tag{3.3}$$

Die zur Bildung der Summenhäufigkeiten F_E benötigten erwarteten absoluten Häufigkeiten E in den einzelnen Klassen lassen sich über die Wahrscheinlichkeitsdichte der Standardnormalverteilung bestimmen:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \, . \tag{3.4}$$

Hierin ist z die bekannte Standardnormalvariable. Die absoluten Häufigkeiten E ergeben sich dann durch Multiplikation von f(z) mit einer Konstanten K, die zur Anpassung der Gesamtzahl der Erwartungshäufigkeiten dient. Sie wird durch die Gleichung

$$K = \frac{n\, b}{s} \tag{3.5}$$

berechnet. Dabei ist s die Standardabweichung der Stichprobe, b die Klassenbreite.

Der Erwartung entsprechend zeigten sich die Verteilungen der monatlichen Niederschlagsmengen allgemein schlecht an die Normalverteilung angepaßt. Für die Anzahl der Tage mit meßbarem Niederschlag gilt dies jedoch nicht (siehe Tabelle 3.1).

Kalendermonat	α bei Niederschlagsmengen	α bei Anzahl der Tage
Januar	0,01	–
Februar	0,05	0,05
März	0,01	0,15
April	0,01	0,15
Mai	0,01	–
Juni	0,10	–
Juli	0,01	–
August	0,05	0,10
September	0,10	–
Oktober	0,01	–
November	0,01	0,10
Dezember	0,05	–

Tabelle 3.1: Ergebnisse des **Lilliefors-Tests** zur Anpassung an eine Normalverteilung für die Verteilungen der monatlichen Niederschlagsmengen und der Anzahl der Tage mit meßbarem Niederschlag in Berlin-Dahlem im Zeitraum 1909–1998. Dargestellt sind die Irrtumswahrscheinlichkeiten α, mit denen sich die Nullhypothese der Normalverteilung ablehnen läßt.

Nur im Februar lehnt der Test die Nullhypothese auf dem 5%-Niveau ab. Für die Monatssummen hingegen wird dieses Niveau nur in den Monaten Juni und September nicht erreicht. In keinem Kalendermonat ist die Verteilung der Niederschlagsmengen der Normalverteilung besser angepaßt als jene der Zahl der Tage mit meßbarem Niederschlag.

Aufgrund dieser Resultate wurde für die Untersuchungen dieser Arbeit stets die Anzahl der Niederschlagstage verwendet. Um die Vergleichbarkeit der Ergebnisse zu gewährleisten, wurden diese auch dann benutzt, wenn die Analysemethode keine Bedingungen an die Verteilungsform stellte. In Anbetracht der Schärfe des **Lilliefors-Tests** erscheint die Anwendung von Regressionen auch auf die Februar-Verteilung aus statistischer Sicht tragbar. In Abbildung 3.1 ist diese im Vergleich zur Normalverteilung dargestellt. Es ist deutlich zu erkennen, daß lediglich die Klassen mit 5 bis 7 bzw. 15 bis 17 Niederschlagstagen eine nennenswerte Abweichung zur Normalverteilung aufweisen. Eine Unterbesetzung der Klassen mit leicht positiver Abweichung vom Mittelwert zugunsten trockener Monate ist festzustellen.

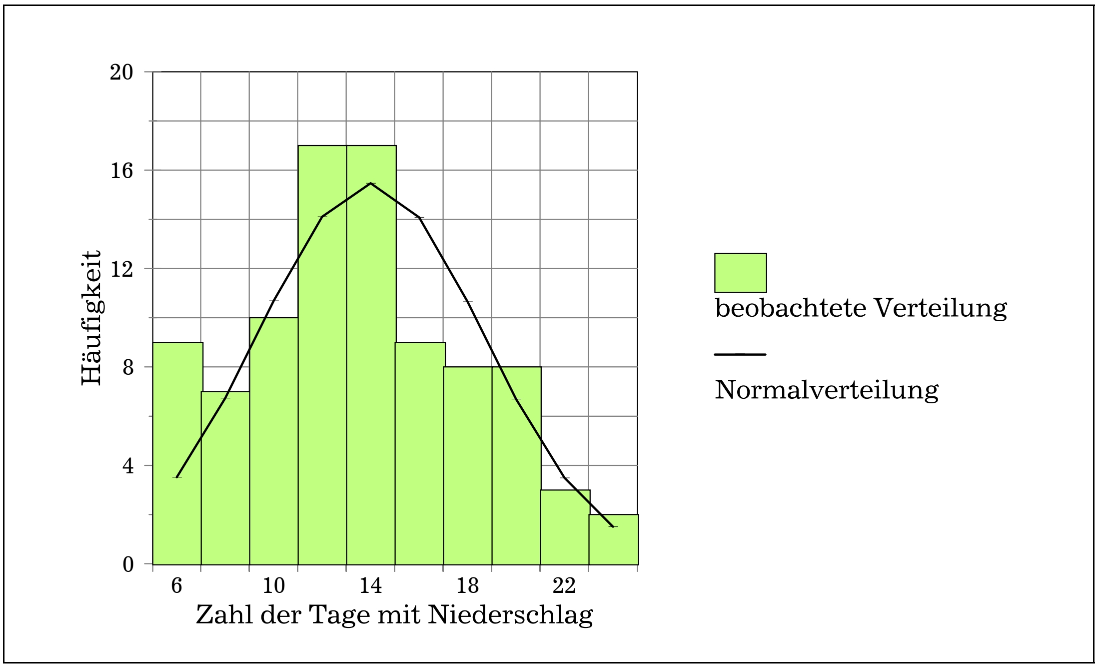

Abbildung 3.1: Histogramm der Häufigkeitsverteilung der jeweiligen Anzahl der Tage mit meßbarem Niederschlag in den Februarmonaten des Zeitraumes 1909–1998 im Vergleich zur aus den Parametern der Stichprobe berechneten Normalverteilung. Auf der x-Achse sind die Klassenmittel aufgetragen.