

32 Grafiken erstellen per Diagrammerstellung

Fortsetzung der Seiten in der 9. Auflage

32.19 Sequenz-, Autokorrelations- und Kreuzkorrelationsdiagramme

32.19.1 Sequenzdiagramme

Zur grafischen Darstellung von Zeitreihenwerten in der Zeit öffnet nach Öffnen die Datei MAKRO.SAV die Befehlsfolge

▷ „Analysieren, „Vorhersage“, „Sequenzdiagramme...“

die in Abb. 32.55 links dargestellte Dialogbox.

Beispiel. Es sollen die Zeitreihen BSP (Bruttosozialprodukt) und M1 (Geldmenge M1) aus der Datei MAKRO.SAV dargestellt werden. Die Variablen werden aus der Quellvariablenliste in das Eingabefeld „Variablen“ übertragen. In das Eingabefeld „Zeitachsenbeschriftung“ wird die Variable JAHR übertragen.

Die Variablen können auch in transformierter Weise dargestellt werden. Dafür stehen im Feld „Transformieren“ folgende Transformationen zur Auswahl:

- Natürlicher Logarithmus.* Logarithmus zur Basis e ($e \cong 2,7183$). Diese Option wird im Beispiel zur Darstellung der Variablen BSP und M1 gewählt. Bei der Wahl einer logarithmischen Skala kann der Verlauf der beiden Variablen im Diagramm besser verglichen werden.
- Differenz.* Man kann wählen, welche Differenz dargestellt werden soll. Voreingestellt ist die Differenz erster Ordnung, d.h. die Differenz zum vorhergehenden Wert. Die Differenz zweiter Ordnung erhält man durch Eintragen einer 2 in das Eingabefeld usw. Zweite Ordnung bedeutet z.B., dass die Differenz der Differenzen der ersten Ordnung gebildet wird (\Rightarrow Kap. 5.9).
- Saisonale Differenz.* Diese Option steht nur dann zur Verfügung, wenn zuvor mit Hilfe der Befehlsfolge „Daten“, „Datum definieren“ die Datenreihe als Zeitreihe definiert wurde (\Rightarrow Kap. 5.9). Es kann analog zur „Differenz“ auch die Differenz erster, zweiter usw. Ordnung abgebildet werden. Voreingestellt ist die Differenz erster Ordnung. Mit Periodizität wird die Häufigkeit von Zeitreihenwerten pro Periode angegeben. Bei Quartalsdaten z.B. ist die Periodizität gleich vier.

Wahlmöglichkeiten.

- ① *Ein Diagramm je Variable.* Hat man mehrere Variablen in das Eingabefeld „Variablen“ eingetragen, so wird für jede Variable eine Grafik erstellt.
- ② *Zeitlinien.* Mit diesem Untermenü kann man Bezugslinien auf die Zeitachse des Diagramms projizieren. Nach Klicken auf „Zeitlinien...“ öffnet sich die links in Abb. 32.56 dargestellte Dialogbox. In der Darstellung ist gegenüber der Voreinstellung eine Änderung vorgenommen. Es ist „Linie bei jedem Wechsel

von:“ gewählt (Voreinstellung „Keine Bezugslinien“) und als „Bezugsvariable“ ist KONJTIEF aus der Quellvariablenliste übertragen worden. Diese Variable bildet die konjunkturellen Tiefpunkte in den Jahren 1967, 1975 und 1982 ab. Für die Jahre 1960 bis 1966 hat die Variable den Wert 1, für 1967 bis 1974 den Wert 2, für 1975 bis 1981 den Wert 3 und für 1982 bis 1990 den Wert 4, d.h. der Wert ändert sich immer in dem Jahr eines Konjunkturtiefs. Die Bezugslinie wird bei jedem Wertewechsel der Variablen KONJTIEF in das Diagramm eingefügt, so dass die Jahre mit Konjunkturtiefen besonders gekennzeichnet sind. In Abb. 32.56 rechts ist das im Diagramm-Editor überarbeitete Diagramm mit Zeitlinien, in Abb. 32.55 rechts ohne Zeitlinien zu sehen.

Es ist auch möglich, eine Bezugslinie für eine bestimmte Beobachtung bzw. Zeitperiode in das Diagramm einzufügen. Dann wird in der Unterdialogbox „Sequenzdiagramme: Bezugslinien für Zeitachse...“ der Optionsschalter „Linie bei Zeitpunkt“ angeklickt und in das Eingabefeld von „Beobachtung“ die Zahl des Falles eingegeben, bei dem die Bezugslinie eingefügt werden soll. Wollte man im Beispiel eine Bezugslinie bei 1980 einfügen, so wäre 21 einzugeben, weil 1980 der 21te Fall ist. Ist per „Datum definieren“ im Datensatz eine Zeitvariable definiert und benutzt man diese für die Zeitreihe (\Rightarrow Kap. 5.9), so nennt das Eingabefeld die definierte Zeiteinheit. In diesem Fall ist der gewünschte Zeitpunkt so einzugeben, wie er auf der Zeitvariablen als Wert eingetragen ist.

③ *Format*. Klicken auf „Format...“ öffnet eine Dialogbox, in der folgende Spezifizierungen möglich sind:

- Zeit auf horizontaler Achse*. Mit Ausschalten dieser Option wird die senkrechte Achse der Grafik als Zeitachse verwendet.
- Diagramme einzelner Variablen*.
 - „Liniendiagramm“ (voreingestellt)
 - Mit „Flächendiagramm“ kann zu diesem Diagrammtyp gewechselt werden.¹
 - „Bezugslinie für Mittelwert der Zeitreihe“ fügt eine Linie in Höhe des Mittelwerts der Reihe ein.²
- Diagramm mit mehreren Variablen*. Option „Fälle zwischen Variablen verbinden“. Es werden für jede Zeitperiode (Beobachtung) Verbindungslinien zwischen den Variablen gezogen.

¹ Nur für Sequenzdiagramme mit einer Variable.

² Nur für Sequenzdiagramme mit einer Variable.

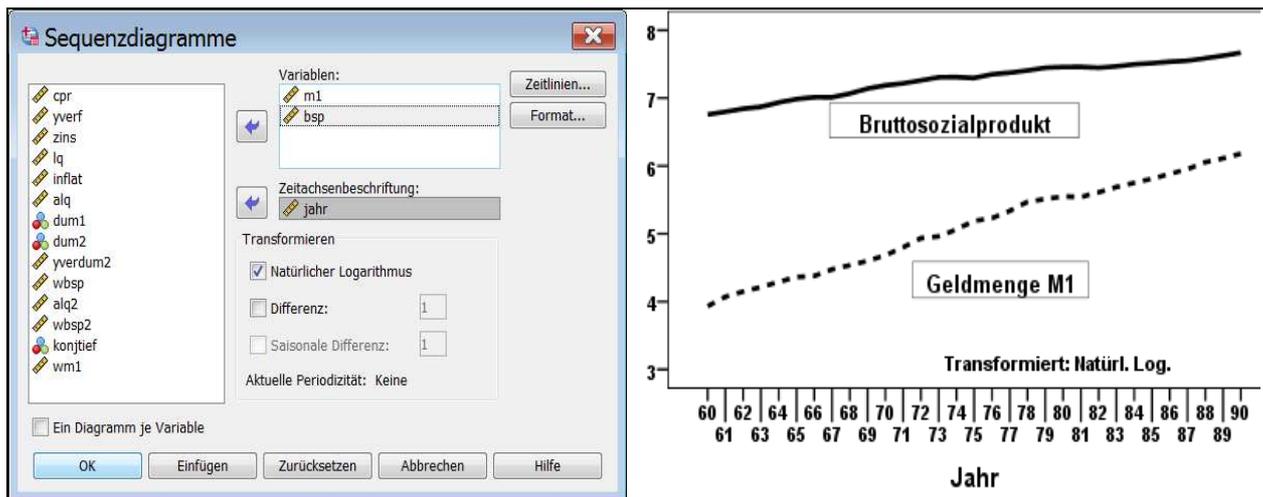


Abb. 32.55. Zeitreihendarstellung in logarithmischer Skala

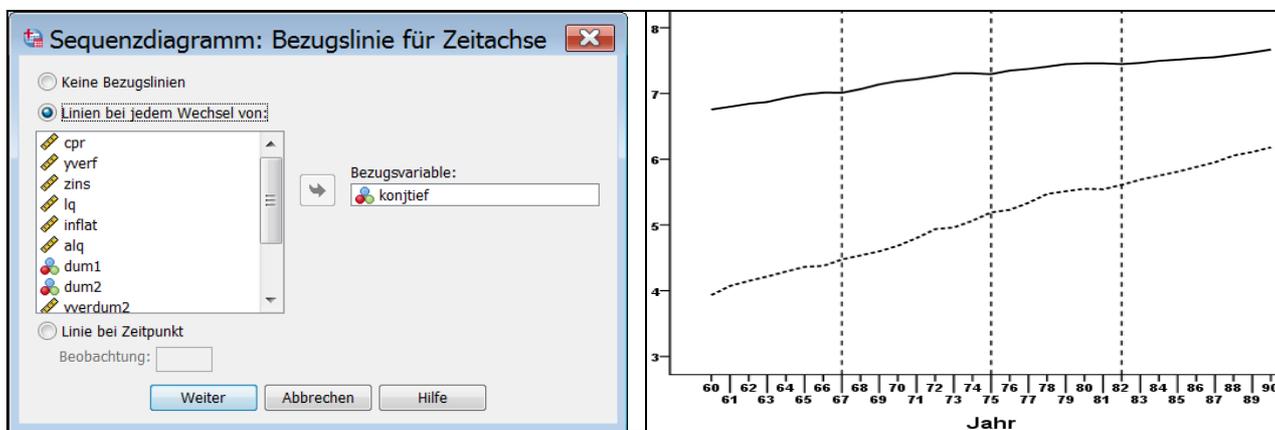


Abb. 32.56. Einfügen von Bezugslinien

32.19.2 Autokorrelationsdiagramme

In modernen Modellen der Zeitreihenanalyse wird die Entstehung einer Zeitreihe als ein stochastischer Prozess interpretiert. Die Struktur eines solchen Prozesses kann in verschiedenen Modellen [z.B. moving average oder kurz MA(k), autoregressiver Prozess oder kurz AR(k) der Ordnung k bzw. Mischformen] erfasst und spezifiziert werden. Für die Spezifizierung sowie Beurteilung eines derartigen Modells hat das *Autokorrelations-* sowie das *partielle Autokorrelationsdiagramm* eine wichtige Funktion, da das Muster dieser Diagramme Hinweise für die Modellierung gibt und zur Überprüfung eines gewählten Modells dient.

In einem Autokorrelationsdiagramm (Autokorrelationsfunktion) werden Autokorrelationskoeffizienten dargestellt. Autokorrelationskoeffizienten messen die Korrelation zwischen den Zeitreihenwerten in Periode t und den um 1, 2, 3, ...k Perioden zurückliegenden Zeitreihenwerten (lags). Der Schätzwert für den *Autokorrelationskoeffizienten* r_k (der Korrelation einer Zeitreihe mit sich selbst bei um k Perioden zurückliegenden Zeitreihenwerten) wird in Gleichung 32.3 angeführt.

Er ist analog zum Korrelationskoeffizienten nach Bravais-Pearson definiert (\Leftrightarrow Kap. 17.1):

$$r_k = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (32.3)$$

y_t = Zeitreihenwert in den Zeitperioden $t = 1, 2, \dots, n$

\bar{y} = arithmetisches Mittel

k = lag von k Perioden

Der *partielle Autokorrelationskoeffizient* misst analog zum partiellen Korrelationskoeffizienten die Stärke des Zusammenhangs zwischen Zeitreihenwerten in Periode t und um k Perioden zurückliegenden Werten, nachdem der Effekt der Korrelation der vorhergehenden lags statistisch eliminiert worden ist. Das Muster der partiellen Korrelationskoeffizienten bietet eine Hilfe zur Entscheidung über die Ordnung eines AR-Modells für die Zeitreihe.

Beispiel. Um ein Autokorrelationsdiagramm zu erzeugen, klickt man die Befehlsfolge

▷ „Analysieren“, „Vorhersage ▷“, „Autokorrelationen...“

zur Öffnung der in Abb. 32.57 links dargestellten Dialogbox. Für ein Anwendungsbeispiel wurde die Variable ZINS (Zinssatz) aus dem Datensatz MAKRO.SAV aus der Quellvariablenliste in das Feld „Variablen“ übertragen. In der Abb. 32.57 rechts wird das Korrelationsdiagramm dargestellt. Auf der X-Achse sind die Länge der lags und auf der Y-Achse die Autokorrelationskoeffizienten abgebildet. „ACF“ ist die Abkürzung für Auto-Correlation-Function. So wird der Ausweis der Autokorrelationskoeffizienten als Funktion der lag-Länge bezeichnet. In der Grafik wird auch der 95 %-Konfidenzbereich um den Wert Null ausgewiesen. Es zeigt sich, dass für lags in Höhe von 1 und 12 bis 14 die Koeffizienten außerhalb des Konfidenzbereiches liegen.

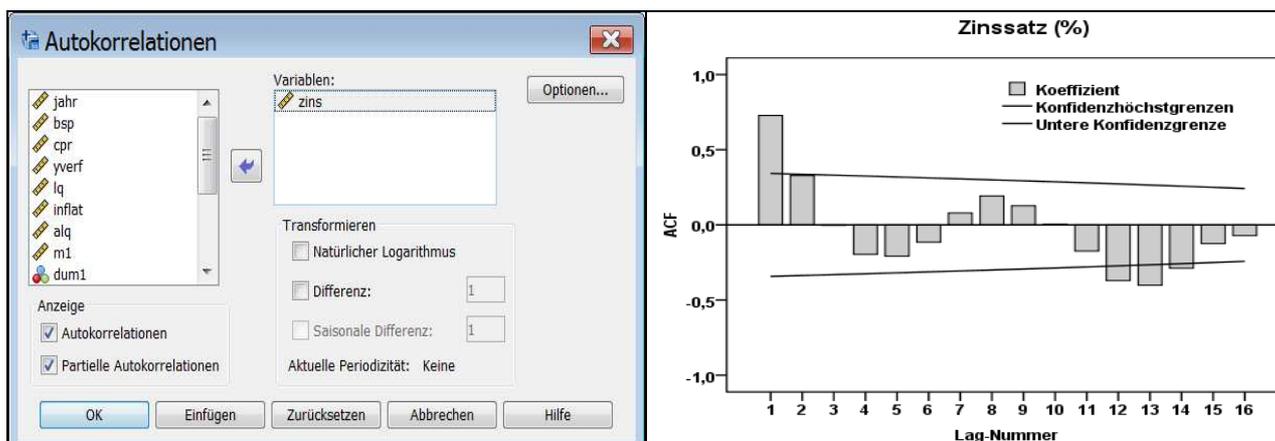


Abb. 32.57. Autokorrelationsdiagramm für die Zeitreihe Zinssatz

Folgende Spezifizierungen sind möglich:

- Transformieren.* Es können für die Variablen die gleichen Transformationen wie in Sequenzdiagrammen gewählt werden (\Rightarrow Kap. 32.19.1).
- Anzeige.*
 - *Autokorrelationen.* Es wird ein Autokorrelationsdiagramm erzeugt. In Abb. 32.57 rechts ist das Autokorrelationsdiagramm für die Variable ZINS dargestellt. Bei Erzeugung der dort dargestellten Grafik wurde außerdem per Schaltfläche „Optionen“ die Dialogbox „Autokorrelationen: Optionen“ geöffnet und dort 16 Zeitintervalle (lags) gewählt. Die Größe der Autokorrelationskoeffizienten entspricht der Höhe der Balken. Jeder Balken gehört zu einem bestimmten lag.
 - *Partielle Autokorrelationen.* Es wird ein partielles Autokorrelationsdiagramm erstellt.
- Optionen.* In einer Dialogbox können folgende Vorgaben festgelegt werden:
 - *Maximale Anzahl an Zeitintervallen.* Im Beispiel wurden 16 gewählt.
 - *Methode für Standardfehler.* Es stehen zwei Modelle zur Berechnung des Standardfehlers der Autokorrelationskoeffizienten für den Ausweis von Konfidenzbereichen zur Auswahl:

- *Unabhängigkeitsmodell.* In diesem Modell wird zur Berechnung eines Standardfehlers als H_0 -Hypothese angenommen, dass der Prozess zur Entstehung der untersuchten Zeitreihe y durch unabhängige Ziehungen aus gleichen Populationen entstanden ist. Man nennt einen derartigen Verlauf auch white noise. Der Standardfehler von r_k berechnet sich für dieses Modell (wenn keine Variablenwerte fehlen) nach der Gleichung

$$S_{rk} \cong \sqrt{\frac{1}{n} \left(\frac{n-k}{n+2} \right)} \quad (32.4)$$

- *Bartletts Approximation.* Diese Approximation ist dann angemessen, wenn für die Reihe als Modell ein moving-average-Prozess unterstellt werden kann. Nach Bartlett beträgt (wenn keine Variablenwerte fehlen und unter der Annahme eines MA-Prozesses der Ordnung $k-1$) der Standardfehler von r_k approximativ

$$S_{rk} \cong \sqrt{\frac{1}{n} \left(1 + 2 \sum_{j=1}^{k-1} r_j^2 \right)} \quad (32.5)$$

- *Autokorrelation für periodische Lags anzeigen.* Diese Möglichkeit zielt auf den Fall, dass die Daten z.B. saisonalen Schwankungen unterliegen und die Datenreihe mittels „Daten“, „Datum definieren“ als Zeitreihe mit z.B. Quartalen definiert worden ist (\Rightarrow Kap. 5.9).

Im Ausgabefenster von SPSS werden ergänzende Informationen zu den Autokorrelationsdiagrammen bereitgestellt. In Tabelle 32.9 ist ein Ausschnitt aus der Ausgabe für das Beispiel der Variablen ZINS wiedergeben. Für die lags werden mit „Autokorrelation“ die Autokorrelationskoeffizienten und mit „Standardfehler“ ihre Standardfehler aufgeführt. Mit „Box-Ljung-Statistik“ wird eine Test-

Prüfgröße bereitgestellt, die einen Hypothesentest zur Prüfung auf Vorliegen von Autokorrelation ermöglicht. Die Test-Statistik ist für den lag k definiert als:

$$Q_k = n(n+2) \sum_{j=1}^k \frac{r_j^2}{n-j} \quad (32.6)$$

n = Anzahl der Fälle

k = lag von k Perioden

r_j = Autokorrelationskoeffizient für einen lag von j Perioden

Für große n hat Q_k eine Chi-Quadrat-Verteilung mit $k-p-q$ Freiheitsgraden, wobei p und q die Ordnungen des autoregressiven bzw. moving-average-Prozesses sind.

Hohe Werte der Test-Prüfgröße sind ein Zeichen dafür, dass Autokorrelation vorliegt. Mit „Sig.“ wird in der SPSS-Ausgabe die Wahrscheinlichkeit (ein Signifikanzniveau) angeführt, mit der man sich bei Ablehnung der Hypothese H_0 (es besteht keine Autokorrelation) irren kann. In unserem Beispiel wird man für die Variable ZINS die H_0 -Hypothese ablehnen, da das angeführte Signifikanzniveau kleiner ist als eine gewöhnlich zur Prüfung verwendete Irrtumswahrscheinlichkeit von 5 % ($\alpha = 0,05$).

Im partiellen Autokorrelationsdiagramm für die Variable ZINS stellen die Balken die Größe des partiellen Korrelationskoeffizienten für lags der Länge von 1 bis 16 Perioden dar. Um den Wert Null wird der Zwei-Sigma-Konfidenzbereich markiert. Nur die Koeffizienten mit lags von 1 und 2 sind größer als der angezeigte Bereich. Diese Informationen deuten zusammen mit der Darstellung der Autokorrelationskoeffizienten darauf hin, dass die Zeitreihe eventuell mit einem autoregressiven Modell der Ordnung zwei prognostiziert werden kann. Zu dem partiellen Autokorrelationsdiagramm wird im Ausgabefenster auch eine Ausgabe analog der Tabelle 32.9 angeboten.

Tabelle 32.9. Ergebnisausgabe Autokorrelation

Autokorrelationen					
Zeitreihe:zins					
Lag	Autokorrelation	Standardfehler ^a	Box-Ljung-Statistik		
			Wert	df	Sig. ^b
1	,729	,171	18,098	1	,000
2	,328	,168	21,894	2	,000
3	-,004	,165	21,895	3	,000
4	-,196	,162	23,346	4	,000
5	-,208	,159	25,040	5	,000
6	-,115	,156	25,581	6	,000
7	,080	,153	25,852	7	,001
8	,193	,150	27,501	8	,001

a. Der angenommene zugrundeliegende Prozeß ist Unabhängigkeit (weißes Rauschen)

b. Beruht auf der asymptotischen Chi-Quadrat-Approximation.

32.19.3 Kreuzkorrelationsdiagramm

In einem Kreuzkorrelationsdiagramm werden Korrelationskoeffizienten zur Messung der Stärke des Zusammenhangs zwischen zwei Zeitreihenvariablen dargestellt, wobei für die Korrelationen unterschiedliche lags (Zeitverzögerungen) zugrunde gelegt werden. Kreuzkorrelationsdiagramme bieten eine Entscheidungsgrundlage für die Frage, mit welchem lag eine Zeitreihe eine andere Zeitreihe am besten vorhersagen kann. Die Kreuzkorrelationskoeffizienten werden bei k lags wie folgt berechnet:

$$r_{xy}(k) = \frac{C_{xy}(k)}{S_x S_y} \quad (32.6)$$

Wobei:

$$\begin{aligned} C_{xy}(k) &= \text{Kovarianz bei } k \text{ lags} \\ &= \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}) \text{ für } k = 0, 1, 2, \dots \\ &= \frac{1}{n} \sum_{t=1}^{n+k} (y_t - \bar{y})(x_{t-k} - \bar{x}) \text{ für } k = -1, -2, \dots \end{aligned}$$

$$S_x = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^2} \quad \text{und} \quad S_y = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^2}$$

Der Standardfehler von $r_{xy}(k)$ beträgt unter der Annahme, dass die Zeitreihen nicht kreuzkorreliert sind und eine der Reihen approximativ white noise ist [Box and Jenkins (1976)]

$$S_{rk} \cong \sqrt{\frac{1}{n - |k|}} \text{ für } k = 0, \pm 1, \pm 2, \dots \quad (32.7)$$

Um ein Kreuzkorrelationsdiagramm zu erzeugen, klickt man die Befehlsfolge

▷ „Analysieren“, „Vorhersage ▷“, „Kreuzkorrelationen...“

zur Öffnung der in Abb. 32.58 links dargestellten Dialogbox. Für ein Anwendungsbeispiel aus dem Datensatz MAKRO.SAV sind die Variablen WM1 (Wachstumsrate der Geldmenge M1) und WBSP (Wachstumsrate des Bruttozonalprodukts) aus der Quellvariablenliste in das Feld „Variablen“ übertragen worden. In Abb. 32.58 rechts wird das Korrelationsdiagramm dargestellt. Auf der X-Achse sind die lags und auf der Y-Achse die Kreuzkorrelationskoeffizienten abgebildet. „CCF“ ist die Abkürzung für Cross-Correlation-Function. So wird der Ausweis der Kreuzkorrelationskoeffizienten als Funktion von lags bezeichnet. In der Grafik werden auch die 95 %-Konfidenzbereiche um den Wert Null ausgewiesen. Es zeigt sich, dass nur für den lag von eins der Koeffizient außerhalb des Konfidenzbereiches liegt.

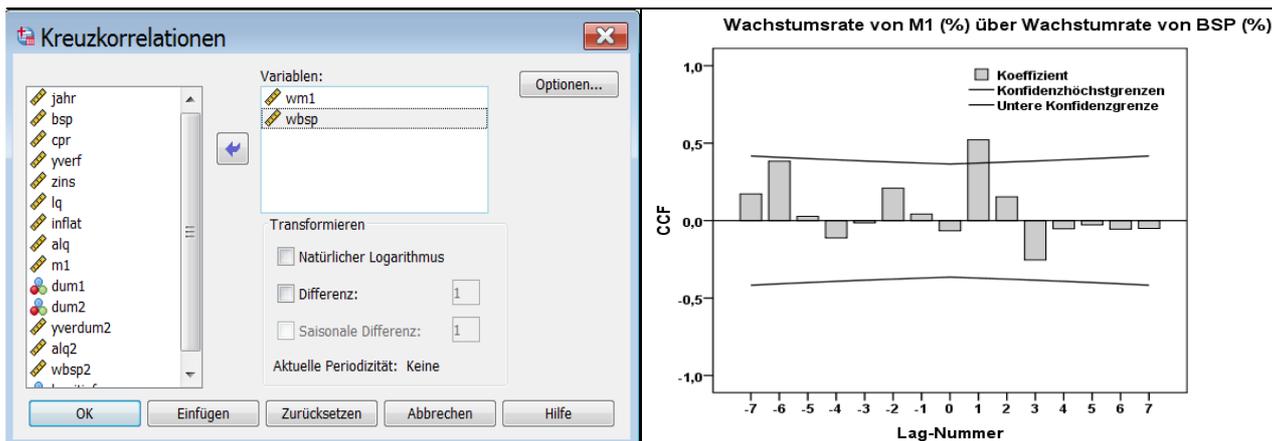


Abb. 32.58. Kreuzkorrelationsdiagramm für die Zeitreihen M1 und Wbsp

Für ein Kreuzkorrelationsdiagramm sind folgende Spezifizierungen möglich:

- Transformieren.* Es können für die Variablen die gleichen Transformationen wie in Sequenzdiagrammen gewählt werden (\Rightarrow Kap. 32.19.1).
- Optionen.* Klickt man auf „Optionen“, öffnet sich eine kleine Dialogbox. Es können folgende Vorgaben festgelegt werden:
 - *Maximale Anzahl von Lags.* Im Beispiel wurde 7 gewählt.
 - *Kreuzkorrelationen für periodische Lags anzeigen.* Diese Möglichkeit zielt auf den Fall, dass die Daten z.B. saisonale Schwankungen aufweisen und die Datenreihe mittels „Daten“, „Datum definieren“ als Zeitreihe mit z.B. Quartalen definiert worden ist (\Rightarrow Kap. 5.9).

Analog zu Autokorrelationsdiagrammen (\Rightarrow Kap. 32.19.2) werden im Ausgabe-fenster die Kreuzkorrelationskoeffizienten mit ihren Standardfehlern für die angeforderte Anzahl von lags in einer Tabelle aufgeführt und in einer einfachen Grafik dargestellt.

32.20 Regelkarten- und Pareto-Diagramme

32.20.1 Regelkartendiagramme.

Diese werden in der statistischen Qualitätskontrolle eingesetzt. Zur Überprüfung von laufenden Produktionsprozessen werden z.B. täglich produzierte Einheiten bzw. Stücke zufällig ausgewählt und auf ihre Qualität hin geprüft. Messwerte eines Qualitätsmerkmals (z.B. die Länge eines Werkstücks) können dann in Diagrammen abgebildet werden, um festzustellen, ob Abweichungen der Messwerte vom Normwert zufällig sind oder als fehlerhafte Veränderung im Produktionsprozess interpretiert werden müssen.

In der Regel werden die einer Qualitätsprüfung unterzogenen Einheiten bzw. Stücke in Bündeln (hier Untergruppen genannt) aus der laufenden Produktion ausgewählt und geprüft. Dabei werden zwei unterschiedliche Arten von Daten erfasst: entweder werden für die Werkstücke Daten mit metrischer Skala erhoben (z.B. die Länge des Werkstücks) oder es wird festgehalten, ob die Einheiten bzw.

Stücke fehlerhaft sind oder nicht. Davon abhängig sind unterschiedliche Diagrammtypen zu wählen. Weitere Kriterien für die Auswahl des geeigneten Diagrammtyps sind: ob ein Bündel (eine Untergruppe) eine große oder kleine Stückzahl enthält, ob die Gruppen eine konstante oder unterschiedliche Anzahl von Einheiten umfasst, ob verschiedene Fehler von fehlerhaften Stücken erfasst werden oder nicht. In Tabelle 32.10 wird eine Übersicht darüber gegeben.

Tabelle 32.10. Übersicht zu Arten von Regelkarten-Diagrammen

Art der Qualitätskontrolle	Einheiten je Untergruppe	Diagrammtyp
Metrische Daten	≥ 10	X-Quer und R
	< 10	X-Quer und s
	$= 1$	Einzelwerte und gleitende Spannweite
Fehlerhafte Stücke	konstant verschieden	p oder np p
Fehlerhafte Stücke mit je mehreren Fehlern	konstant verschieden	c oder u u

Um ein Regelkarten-Diagramm zu erstellen, öffnet man durch Klicken der Befehlsfolge

▷ „Analysieren“, „Qualitätskontrolle“, „Qualitätsregelkarten...“
die in Abb. 32.59 dargestellte Dialogbox.

Im Fall erhobener metrischer Daten sind die ersten beiden Diagramme wählbar („Variablendiagramme“), im Fall der Erhebung fehlerhafter Stücke die letzten beiden („Attributdiagramme“).

Bei beiden Datenarten werden die erhobenen Daten (Messwerte der geprüften Stücke bzw. Anzahl der fehlerhaften Stücke) in Spalten des Daten-Editors abgelegt.

Für beide Datenarten können die Daten im Daten-Editor unterschiedlich aufgebaut sein (Feld „Datenorganisation“):

- Fälle sind Einheiten.* Jeder Fall (d.h. eine Zeile im Daten-Editor) ist ein kontrolliertes Stück.
- Fälle sind Untergruppen.* Jeder Fall (d.h. eine Zeile im Daten-Editor) ist ein Fall einer Untergruppe oder eine Gruppe. Eine Untergruppe ist in der Regel eine in Bündeln entnommene Anzahl von zu kontrollierenden Stücken. Die Anzahl der Untergruppen muss nicht konstant sein.

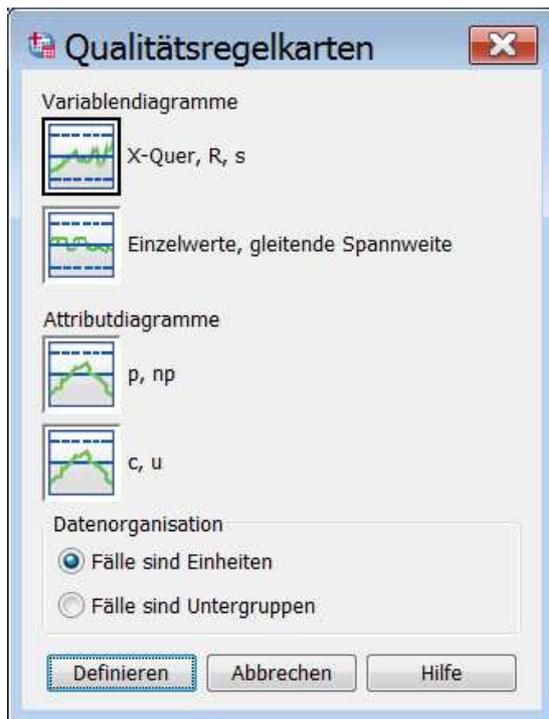


Abb. 32.59. Dialogbox zur Auswahl eines Regelkarten-Diagramms

Aus den Regelkarten-Diagrammen sind für die zwei Formen der Datenorganisation folgende vier Formen wählbar.

- *X-Quer, R, s*: Es handelt sich hierbei um verschiedene Diagramme:
 - *X-Quer-Diagramm*. Es wird für jede definierte Gruppe von Einheiten der Gruppen-Mittelwert einer metrischen Variablen (häufig mit dem Symbol \bar{x} bezeichnet) dargestellt.
 - *R-Diagramm*. Es wird für jede definierte Gruppe von Einheiten die Gruppenspannweite R (Range = die Differenz zwischen dem kleinsten und größten Wert in einer Gruppe) dargestellt.
 - *s-Diagramm*. Es wird für jede definierte Gruppe von Einheiten die Gruppenstandardabweichung s dargestellt.

Die Ergebnisse sind unabhängig davon, ob die Daten in der Form „Fälle sind Einheiten“ oder „Fälle sind Untergruppen“ organisiert sind, jedoch muss die Organisationsform korrekt angegeben werden, weil zur Berechnung der statistischen Kennzahlen einer Gruppe bei der ersten Organisationsform die Werte einer einzigen SPSS-Variablen in untereinander liegenden Zeilen verwendet werden, bei der zweiten Organisationsform dagegen die Werte mehrerer SPSS-Variablen, die in derselben Zeile stehen.
- *Einzelwerte, gleitende Spannweite*. Es handelt sich um zwei Diagramme zur Darstellung der Messwerte einer metrischen Variablen:
 - *Einzelwerte*. Dargestellt werden die Messwerte von einzelnen Stücken.
 - *Gleitende Spannweite*. Dargestellt wird jeweils die Differenz der Messwerte aufeinanderfolgender Stücke.

- *p, np*. Es handelt sich um zwei Diagramme zur Darstellung der Anzahl der fehlerhaften Stücke:
 - *p-Diagramm*. Dargestellt wird für jede Untergruppe die Häufigkeit von fehlerhaften Stücken in Form des Anteils p an allen Stücken der Untergruppe.
 - *np-Diagramm*. Dargestellt wird die absolute Anzahl der fehlerhaften Stücke in jeder Untergruppe ($p \cdot n = \text{Fehleranteil} \cdot \text{Stückzahl}$)
- *c, u*. Es handelt sich ebenfalls um zwei Diagramme zur Darstellung der Häufigkeit von fehlerhaften Stücken. Sie werden angewendet, wenn die Daten in anderer Form vorliegen, als den zwei bereits vorgestellten.

Im Folgenden werden einige dieser verschiedenen Diagrammformen anhand von Beispielen erläutert.

32.20.1.1 Diagrammtyp: X-Quer, R, s

Datenorganisation: Fälle sind Einheiten. *Beispiel:* In der Datei ZIGARETT.SAV, die ausschnittsweise in Abb. 32.60 im Dateneditorfenster dargestellt ist, sind mehrere metrische Variablen mit Messwerten von Zigaretten erfasst. Jeder Fall (eine Zeile im Daten-Editor) ist eine Zigarette aus einem Produktionsprozess A bzw. B. Stündlich wurden Zufallsproben von je 20 Zigaretten aus den laufenden Produktionsprozessen entnommen und geprüft. Insgesamt sind in dem Datensatz je Produktionsprozess Messwerte von 10 Proben à 20 Zigaretten enthalten. Die Variable PROBE mit den Werten 1 bis 10 dient zur Identifikation der Stichproben. Es soll die Variable DM_A, die den Durchmesser der auf der Anlage A produzierten Zigaretten erfasst, in Kontrolldiagrammen dargestellt werden.

	nr	probe	gew_a	zw_a	fv_a	dm_a	gew_b	zw_b	fv_b	dm_b	dm10_a	dm10_b
1	1	1	614	123	57,9	5,42	622	140	59,3	5,42	54,20	54,20
2	2	1	572	123	50,4	5,37	607	145	56,2	5,42	53,70	54,20
3	3	1	634	126	54,2	5,42	638	150	59,9	5,42	54,20	54,20
4	4	1	605	132	52,0	5,38	600	145	55,6	5,41	53,80	54,10
5	5	1	612	136	52,2	5,39	584	130	54,7	5,42	53,90	54,20
6	6	1	576	130	46,2	5,40	647	157	57,8	5,42	54,00	54,20
7	7	1	618	141	53,3	5,41	605	152	55,7	5,42	54,10	54,20

Abb. 32.60. Daten der Datei ZIGARETT.SAV im Daten-Editor (Ausschnitt)

Nach der Befehlsfolge „Analysieren“, „Qualitätskontrolle“, „Qualitätsregelkarten...“ wird die Auswahlkombination „X-Quer, R, s“ und „Fälle sind Einheiten“ angeklickt. Abb. 32.61 zeigt links die nach Klicken von „Definieren“ geöffnete Dialogbox mit einem Beispiel zur Grafikdefinition. Als Diagramm ist „X-Quer mit Bereich“ gewählt. Standardmäßig wird bei dieser Wahl auch das „R-Diagramm“ ausgegeben, da man i.d.R. beide Grafiken zusammen betrachtet. Die metrische Variable DM_A wurde in das Feld „Prozessmessung“ und die Variable PROBE in das Feld „Untergruppen definiert durch“ übertragen. Nach Klicken auf „OK“ werden die beiden Diagramme erzeugt. In der Abb. 32.61 rechts ist das erste Diagramm (das X-Quer-Diagramm) zu sehen. Für jede der zehn Stichproben à 20 Zigaretten wird der Mittelwert der Durchmesser dargestellt. Mit 5,3981 wird der Mittelwert aller Zigaretten als waagerechte Linie angezeigt. Um den Mittel-

wert werden als unterbrochene Linien der obere (UCL = upper control limit) und der untere (LCL = lower control limit) Kontrollwert in Form eines Drei-Sigma-Bereichs angezeigt. In der durch Mausklick auf „Optionen“ geöffnete Unterdialogbox „X-Quer, R, s: Optionen“ kann ein anderer Sigma-Bereich (beruhend auf der Normalverteilung) gewählt werden. In „Mindestgröße der Untergruppe“ kann eine Stichprobengröße vorgegeben werden. Ist die Stichprobengröße der Untergruppe kleiner, dann wird diese Untergruppe nicht in die Grafik und die Berechnungen einbezogen. Eine Option ermöglicht es zu wählen, ob fehlende Werte als Untergruppe angezeigt werden oder nicht. Letztere ist voreingestellt.

Es zeigt sich, dass die Mittelwerte einiger Proben aus dem abgesteckten Kontrollintervall herausfallen.

In dem zweiten Diagramm („R-Diagramm“) wird eine analoge Darstellung gegeben, mit dem Unterschied, dass für jede Stichprobe die Spannweite R (Range = Differenz zwischen größtem und kleinstem Wert in jeder Probe) auf der Y-Achse abgebildet wird.

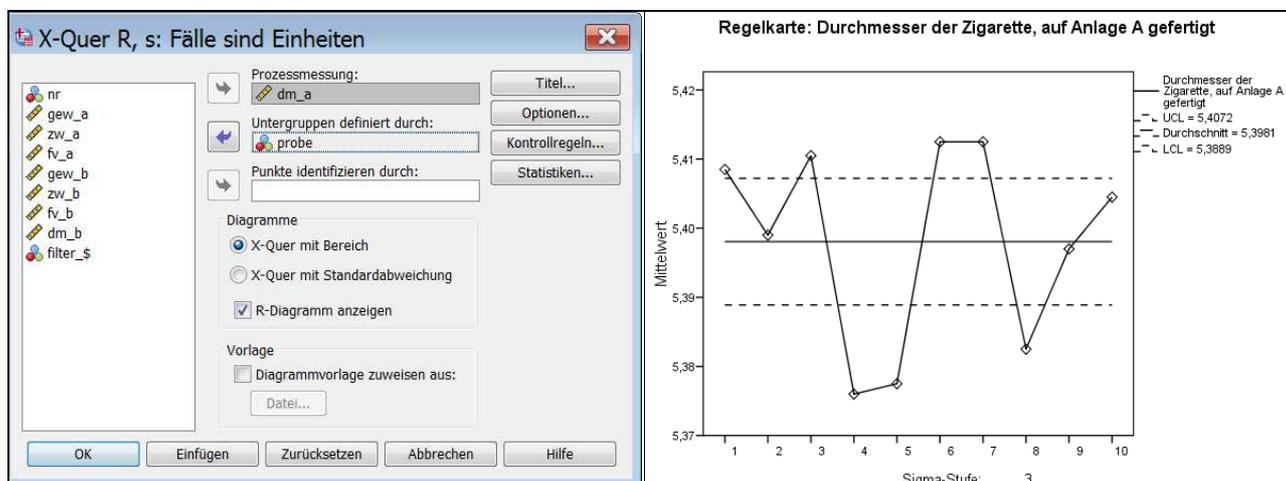


Abb. 32.61. Mittelwert des Durchmessers von Zigaretten in einem X-Quer-Diagramm

Würde „X-Quer mit Standardabweichung“ in der Dialogbox Abb. 32.61 links gewählt, dann würden ebenfalls zwei Diagramme erstellt: Das erste wäre das gleiche X-Quer-Diagramm, und das zweite (das „s-Diagramm“) bildete auf der Y-Achse die Standardabweichung s jeder Probe ab.

Datenorganisation: Fälle sind Untergruppen. *Beispiel.* Die Datei ZIGARET2.SAV enthält dieselben Fälle und Gruppen wie ZIGARETT.SAV (als Probevariable ist zur Vereinfachung nur GEW_A und GEW_B enthalten). Allerdings sind die Daten jetzt anders organisiert. Es gibt nur 10 Zeilen, für jede Gruppe eine. Die Daten der einzelnen Zigaretten sind nicht mehr in Zeilen untereinander geschrieben, sondern in Variablen nebeneinander. Da wir im Beispiel 20 Fälle haben, muss es für jede der zur Kontrolle verwendeten Variablen 20 Spalten geben, für jeden Fall eine. Im Beispieldatensatz gibt es für GEW_A also für jeden Fall eine eigene Spalte, bezeichnet als GEW_A_1, GEW_A_2 usw. (Abb. 32.62)

Nach der Befehlsfolge „Analysieren“, „Qualitätskontrolle“, „Qualitätsregelkarten...“ wird die Auswahlkombination „X-Quer, R, s“ und „Fälle sind Untergrup-

pen“ geklickt. Die Variablen GEW_A_1, GEW_A_2, GEW_A_3 usw. bis GEW_A_20 werden in das Eingabefeld „Stichproben“ übertragen, die Variable PROBE in das Eingabefeld „Untergruppenbeschriftung“. Es werden dieselben Diagramme erzeugt, wie oben beschrieben.

	probe	gew_a_1	gew_a_2	gew_a_3	gew_a_4	gew_a_5	gew_a_6	gew_a_7	gew_a_8
1	1	614	572	634	605	612	576	618	631
2	2	617	621	603	604	610	580	599	617
3	3	562	608	562	548	591	592	584	574
4	4	609	596	575	607	569	581	580	614
5	5	585	588	642	614	566	644	588	618
6	6	589	595	606	594	553	585	633	585
7	7	576	574	619	617	612	630	592	611
8	8	594	620	622	645	627	593	581	614
9	9	619	610	577	622	602	636	583	599
10	10	627	585	554	569	550	566	571	647

Abb. 32.62. Daten der Datei ZIGARET2.SAV im Daten-Editor (Ausschnitt)

32.20.1.2 Diagrammtyp: Einzelwerte, gleitende Spannweite

Datenorganisation: Fälle sind Einheiten. *Beispiel:* Zur Demonstration wird als darzustellende Prozessvariable wieder DM_A des Datensatzes ZIGARETT.SAV verwendet (\Rightarrow Abb. 32.60). Da auf der X-Achse des Diagramms die einzelnen Zigaretten dargestellt werden, haben wir unsere Darstellung mit der Befehlsfolge „Daten“, „Fälle auswählen“ auf die ersten neun Fälle beschränkt.³

Nach der Befehlsfolge „Analysieren“, „Qualitätskontrolle“, „Qualitätsregelkarten...“ wird die Auswahlkombination „Einzelwerte, gleitende Spannweite“ und „Fälle sind Einheiten“ angeklickt. Abb. 32.63 zeigt links die nach Klicken von „Definieren“ geöffnete Dialogbox zur Grafikdefinition. Als Diagramme sind „Einzelwerte und gleitende Spannweite“ gewählt. Die metrische Variable DM_A wurde in das Feld „Prozessmessung“ übertragen. Nach Klicken auf „OK“ werden zwei Diagramme erzeugt. Das erste Diagramm bildet für jeden Fall den Messwert von DM_A ab, zusätzlich einen Sigmabereich um den Mittelwert. Das zweite Diagramm stellt die gleitenden Spannweiten dar. Es ist in Abb. 32.63 rechts angezeigt. Da im Eingabefeld „Spanne“ der Dialogbox der voreingestellte Wert „2“ belassen worden ist, werden die Durchmesserunterschiede von im Datensatz aufeinander folgenden Zigaretten als Spannweite abgebildet. Die Spannweite kann durch Eintragen einer anderen Zahl erhöht werden. Wie auch in X-Quer-Diagrammen werden der Durchschnitt sowie ein Sigma-Bereich durch Kontrolllinien angezeigt. Die untere Kontrolllinie wird aber nur dargestellt, wenn sie im positiven Bereich liegt.

³ Mit der Spezifizierung NR < 10 für „Falls Bedingung zutrifft“.

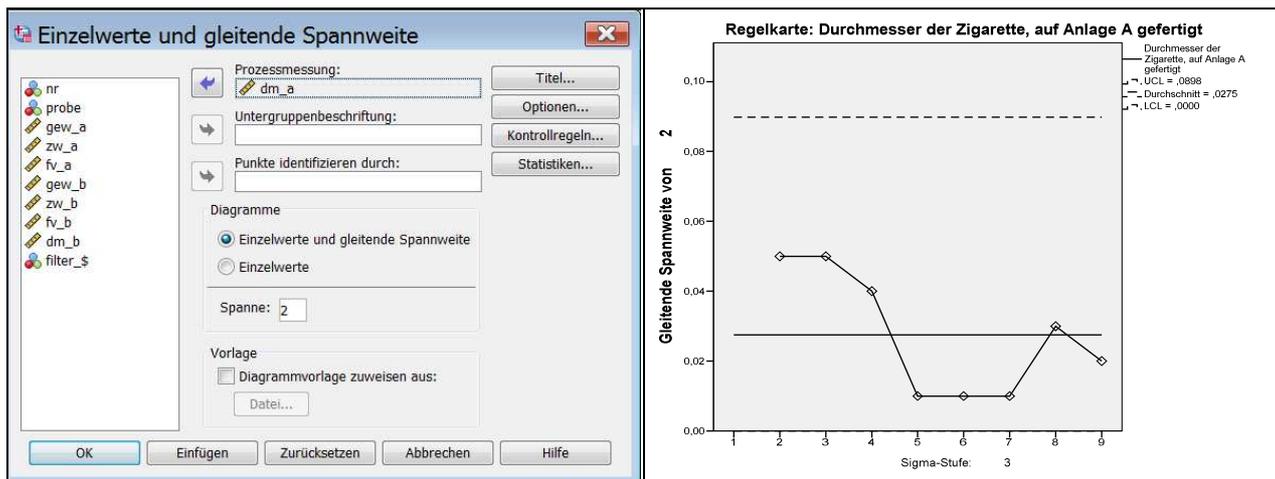


Abb. 32.63. Gleitende Spannweite von Zigarettendurchmessern

32.20.1.3 Diagrammtyp: p, np

Datenorganisation: Fälle sind Einheiten. Dieser Programmtyp wird gewählt, wenn die Qualitätsdaten in qualitativer Form vorliegen. Bei Wahl des p-Diagramms wird der Anteil und bei Wahl des np-Diagramms die absolute Anzahl von fehlerhaften Stücken in jeder Stichprobe grafisch dargestellt. Zur Anwendungsdemonstration wurde die Variable GEW_A aus dem Datensatz ZIGARETT (\Rightarrow Abb. 32.60), die das Gewicht der Zigaretten in mg misst, in eine Variable GEW1_A rekodiert: der Variablenwert „2“ bildet ein (für dieses Beispiel) normgemäßes, „1“ ein zu geringes und „3“ ein zu großes Gewicht ab (Die umkodierten Daten stehen in der Datei: ZIGARET1.SAV).

	dm10_b	dm1_a	fv1_a	zw1_a	gew1_a	fehl_a	dm1_b	fv1_b	zw1_b	gew1_b	fehl_b
1	54,20	2	3	2	2	4	2	2	2	2	0
2	54,20	2	2	2	2	0	2	2	2	2	0
3	54,20	2	2	2	2	0	2	2	2	2	0
4	54,10	2	2	2	2	0	2	2	2	2	0
5	54,20	2	2	2	2	0	2	2	2	2	0
6	54,20	2	1	2	2	3	3	2	3	3	9
7	54,20	2	2	2	2	0	2	2	2	2	0

Abb. 32.64. Daten der Datei ZIGARET1.SAV im Daten-Editor (Ausschnitt)

Nach der Befehlsfolge „Analysieren“, „Qualitätskontrolle“, „Qualitätsregelkarten...“ wird die Auswahlkombination „p, np“ und „Fälle sind Einheiten“ angeklickt. Abb. 32.65 zeigt links die nach Klicken von „Definieren...“ geöffnete Dialogbox mit einem Beispiel zur Grafikdefinition. Im Feld „Diagramme“ ist „p (Anteil der Abweichenden)“ gewählt worden. Die Variable GEW1_A wurde in das Feld „Merkmal:“ und die Variable PROBE in das Feld „Untergruppen definiert durch:“ übertragen. Um den Anteil der Zigaretten mit fehlerhaftem Gewicht an allen in einer Stichprobe enthaltenen Zigaretten anzuzeigen, wird im Auswahlfeld „Zu zählender Wert“ wurde „Fehlerfreie“ gewählt und der Wert „2“ in das Ein-

gabefeld „Wert“ eingetragen. Diese vielleicht nicht unmittelbar einsichtige Einstellung ergibt sich daraus, dass der Wert 2 für normgerechte Zigaretten steht und die Einstellung „Fehlerfrei“ bewirkt, dass der Anteil der nicht diesem Wert entsprechenden Fälle gezählt wird („Abweichende“ dagegen zählt die Fälle, die dem angegebenen Wert entsprechen). Nach Klicken auf „OK“ wird die Regelkarte erzeugt. In der Abb. 32.65

Rechts ist das Diagramm dargestellt. Für jede der 10 Proben à 20 Zigaretten wird die Quote der Zigaretten mit fehlerhaftem (zu geringes oder zu hohes) Gewicht dargestellt. Die waagrechte Linie „Zentrum = 0,09“ veranschaulicht die mittlere Quote der Zigaretten mit fehlerhaftem Gewicht. Um den Mittelwert werden als unterbrochene Linien der obere (UCL = upper control limit) und der untere (LCL = lower control limit) Kontrollwert in Form eines Drei-Sigma-Bereichs angezeigt. Da hier LCL im negativen Bereich liegen würde, wird diese Linie nicht abgebildet. Im Menü „Optionen“ kann ein anderer Sigma-Bereich gewählt werden.

Wird „np (Anzahl der abweichenden Einheiten)“ in dem in Abb. 32.65 dargestellten Dialogfenster gewählt, dann werden die fehlerhaften Stücke in absoluter Anzahl grafisch dargestellt. Diese Auswahl ist nur dann sinnvoll, wenn (wie in unserem Beispiel) die Anzahl der Stücke in jeder Stichprobe konstant ist.

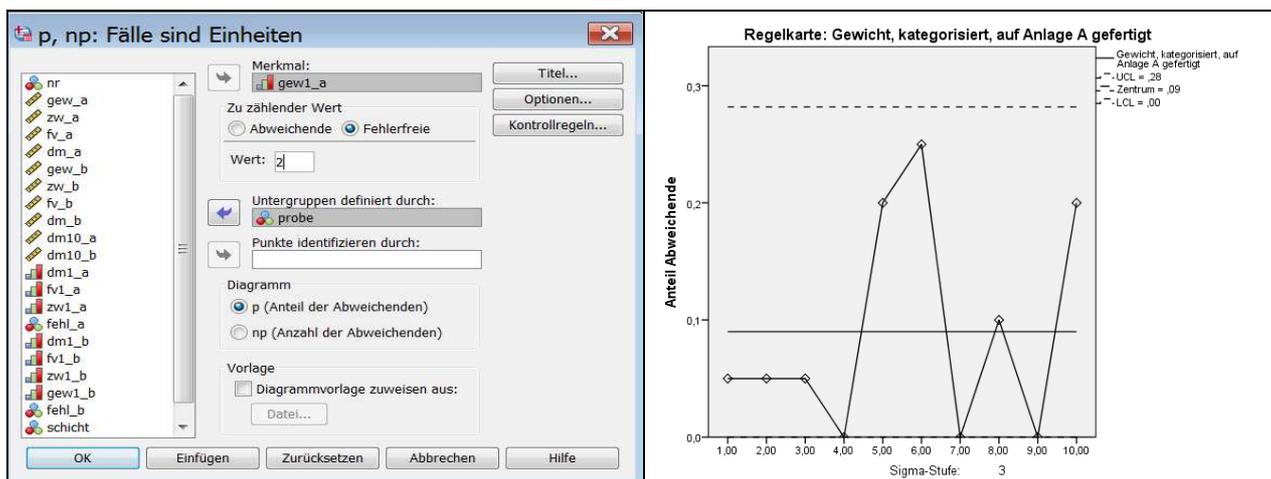


Abb. 32.65. Anteil nicht normgerechter Durchmesser von Zigaretten im p-Diagramm

Datenorganisation: Fälle sind Untergruppen. Dieser Diagrammtyp eignet sich für Qualitätsdaten, die in Form von Häufigkeiten vorliegen.⁴ *Beispiel:* Zur Demonstration werden die in Abb. 32.66 dargelegten Daten verwendet (Datei: ZWISCHF.SAV). Es handelt sich bei ZWISCHF um die Häufigkeit von unvorhergesehenen Zwischenfällen bei in sechs Monaten durchgeführten Operationen. Diese Daten entsprechen denen in Abb. 32.60, mit dem Unterschied, dass hier die wöchentlichen Operationen eines Monats zusammengefasst sind. Die Variable N gibt die Anzahl der Operationen pro Monat an.

⁴ Beachten Sie, dass „Fälle sind Gruppen“ hier etwas anderes bedeutet als bei „X-Quer, R, s“. Hier handelt es sich um aggregierte Daten für die Gruppen, nicht mehr um die Einzelfälle.

	monat	zwischenf	n
1	1	0	20
2	2	4	40
3	3	1	30
4	4	3	25
5	5	1	30
6	6	2	40

Abb. 32.66. Daten der Datei ZWISCHF.SAV im Daten-Editor (Ausschnitt)

Nach der Befehlsfolge „Analysieren“, „Qualitätskontrolle“, „Qualitätsregelkarten...“ wird die Auswahlkombination „p, np“ und „Fälle sind Untergruppen“ angeklickt. Abb. 32.67 zeigt links die nach Klicken von „Definieren“ geöffnete Dialogbox mit dem Beispiel zur Grafikdefinition. Im Feld „Diagramm“ ist „p (Anteil der Abweichenden)“ angeklickt. Die Variable ZWISCHF wurde in das Feld „Anzahl der Abweichenden:“ und die Variable MONAT in das Feld „Untergruppenbeschriftung“ übertragen. Im Feld „Umfang der Stichprobe“ wurde „Variable“ angeklickt und die Variable n in das Eingabefeld übertragen (sie enthält die Anzahl der Operationen je Monat). Diese Option ist zu wählen, wenn die Stichprobengröße je Untergruppe (hier innerhalb der Monate) variiert. Für den Fall gleicher Stichprobengröße je Untergruppe kann im Feld „Umfang der Stichprobe“ „Konstante“ markiert und anschließend die Stichprobengröße in das dafür vorgesehene Eingabefeld eingetippt werden.

In Abb. 32.67 rechts ist das Kontrolldiagramm dargestellt. Für jede der sechs Untergruppen (= Monate) wird der Anteil der Zwischenfälle bei Operationen dargestellt.

Würde „np (Anzahl der Abweichenden)“ in dem in Abb. 32.67 dargestellten Dialogfenster gewählt, dann würde die Anzahl der Zwischenfälle in absoluter Anzahl grafisch dargestellt. Diese Auswahl ist nur dann sinnvoll, wenn die Anzahl der Operationen (= Untergruppe) in jedem Monat konstant ist.

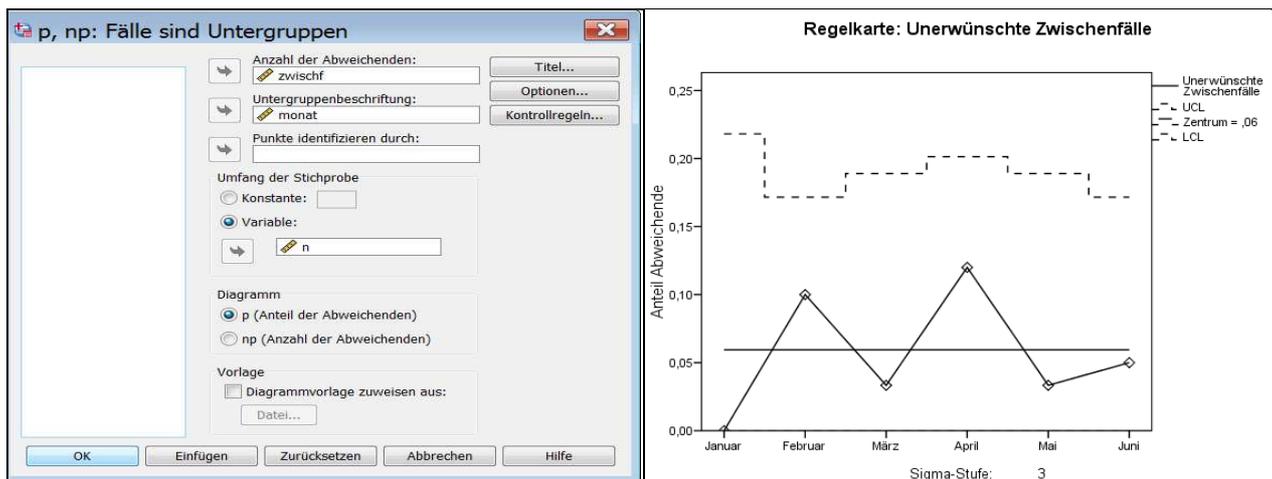


Abb. 32.67. Anteil der Zwischenfälle bei Operation im p-Diagramm

32.20.1.4 Diagrammtyp: c, u

Datenorganisation: Fälle sind Einheiten. Dieser Diagrammtyp ist bei einer anderen Datenlage geeignet. Die Daten liegen in Form von Häufigkeiten von Fehlern bzw. unerwünschter Ereignisse vor. *Beispiel:* In Abb. 32.68 wird ein Auszug aus der Datei ZWISCHF1.SAV gegeben. Bei den Fällen der SPSS-Datendatei handelt es sich um operative Eingriffe in einer Woche. Mit der Variablen ZWISCHF wird die Anzahl unerwünschter Zwischenfälle bei den Operationen erfasst. Eine zweite Variable MONAT (= Untergruppe) erfasst, in welchem Monat eine Operation stattgefunden hat. Bei Wahl des u-Diagramms wird für jeden Monat die Anzahl von Zwischenfällen je Woche und bei Wahl des c-Diagramms die absolute Anzahl von Zwischenfällen pro Monat (= in jeder Untergruppe) grafisch dargestellt.

Nach der Befehlsfolge „Analysieren“, „Qualitätskontrolle“, „Qualitätsregelkarten...“ wird die Auswahlkombination „c, u“ und „Fälle sind Einheiten“ angeklickt. Abb. 32.69 zeigt links die nach Klicken von „Definieren“ geöffnete Dialogbox mit dem Beispiel zur Grafikdefinition. Im Feld „Diagramm“ ist „c (Anzahl der Abweichungen)“ angeklickt. Die Variable ZWISCHF wurde in das Feld „Merkmal“ und die Variable MONAT in das Feld „Untergruppen definiert durch“ übertragen.

	monat	woche	zwischenf	n
1	1	1	0	5
2	1	2	0	4
3	1	3	0	6
4	1	4	0	5
5	2	1	1	10
6	2	2	0	8
7	2	3	2	12
8	2	4	1	10

Abb. 32.68. Daten der Datei ZWISCHF1.SAV im Daten-Editor (Ausschnitt)

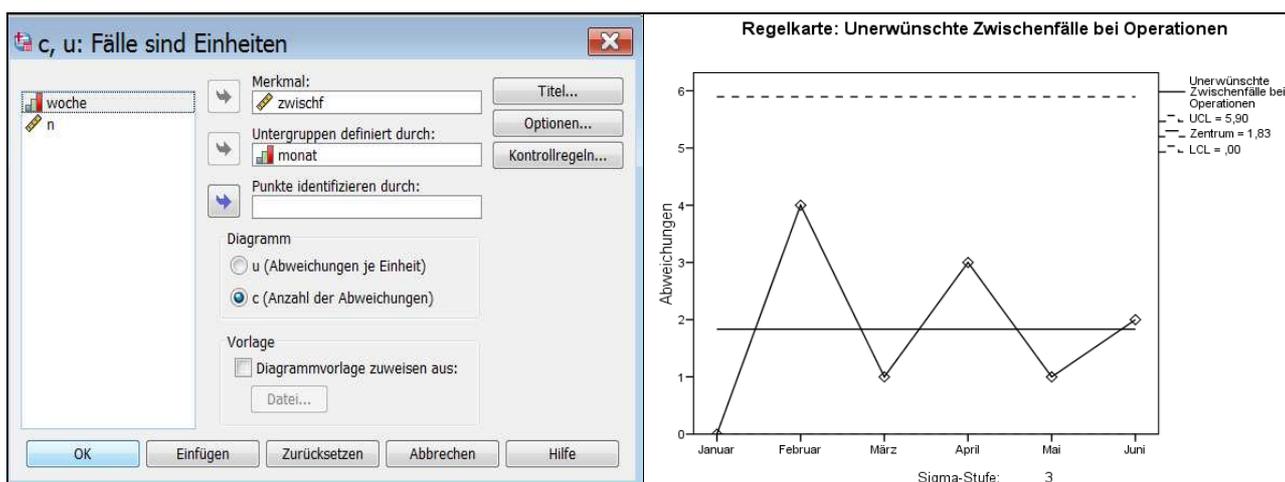


Abb. 32.69. Anzahl von Zwischenfällen im u-Kontrolldiagramm

In der Abb. 32.69 rechts ist das Diagramm dargestellt. Für jeden in der Datei enthaltenen Monat, wird die Anzahl von Zwischenfällen dargestellt. Die Angaben Zentrum, UCL und LCL sind oben erklärt. Würde „u (Abweichungen je Einheit)“ im Feld „Diagramm“ gewählt (\Rightarrow Abb. 32.69), dann würde für jeden Monat die Anzahl der Zwischenfälle je Woche (= Einheit) grafisch dargestellt.

32.20.1.5 Wahlmöglichkeiten

Für Regelkarten-Diagramme bestehen folgende Wahlmöglichkeiten:

Optionen.

- *Anzahl der Sigmas.* Es kann ein Sigma-Bereich (die Anzahl der Standardabweichungen oberhalb und unterhalb der Mittellinie) gewählt werden (vor eingestellt ist ein Drei-Sigma-Bereich).
- *Mindestgröße der Untergruppe.* Für X-Quer-, R- und s-Diagramme kann ein minimaler Umfang einer Untergruppe spezifiziert werden. Untergruppen mit kleinerem Umfang werden dann im Diagramm nicht dargestellt.
- *Fehlende Werte als Untergruppen anzeigen.* Wenn diese Option gewählt ist, werden fehlende Werte der Untergruppenvariable (einschließlich systemdefinierter fehlender Werte) im Diagramm als separate Gruppe angezeigt. Wenn diese Option nicht ausgewählt ist, werden Fälle mit systemdefinierten oder benutzerdefinierten fehlenden Werten für die Untergruppenvariable aus dem Diagramm ausgeschlossen.

Statistiken. Nur für die Diagrammformen „X-Quer, r, s“ und „Einzelwerte, gleitende Spannweite“ steht eine Unterdialogbox mit Statistiken zur Verfügung. U.a. können durch Eingabe von „Maximum“- und „Minimum“-Werten und/oder eines „Ziel“-wertes feste Kontrolllinien spezifiziert werden. Werden solche Werte eingegeben, dann stehen zusätzliche Statistiken zur Verfügung.

- *Fähigkeitssigma.* Das Fähigkeits-Sigma ist ein Maß der Streuung, das bei der Berechnung der Kennziffern für die Tauglichkeit angewendet wird. Im Feld „Tauglichkeitssigma“ kann man zwischen drei Formen der Berechnung wählen. Sigma wird entweder berechnet anhand von R-Quer (mittlere Spannweite), S-Quer (mittlere Standardabweichung) oder der Variation in den Untergruppen des Prozesses (nur bei „X-Quer, r, s“).
- *Indizes für Prozesstauglichkeit.* Dieses Auswahlfeld umfasst Auswahlkästchen für 12 Kennziffern, die allesamt die Qualität des Prozesses messen sollen (mehrere können gleichzeitig angefordert werden). Überwiegend handelt es sich um Differenzen, z.B. zwischen dem Zielwert und unteren Spezifikationsgrenze in Relation zum Fähigkeitssigma (in der Maßeinheit Fähigkeitssigma ausgedrückt). Der Prozess ist umso tauglicher, je höher dieser Wert liegt, denn dann liegt z.B. die untere Spezifikationsgrenze gemessen an der Streuung der Daten relativ weit vom Zielwert entfernt. (Nähere Informationen zu den einzelnen Kennziffern können Sie Drop-down-Fenstern entnehmen, die sich bei Klicken mit der rechten Maustaste auf die Option öffnen.)
- *Indizes für Prozessleistung.* In diesem Bereich können durch Markieren von Auswahlkästchen weitere Kennziffern angefordert werden. Sie sind analog zu den oben angegebenen, jedoch sind sie mit der geschätzten Standard-

abweichung des Prozesses normiert. *Beispiel:* $PP = \frac{\text{Maximum} - \text{Minimum}}{6 * \text{Standardabweichung}}$.

Wurde z.B. in der Datei ZIGARETT.SAV die untere Spezifikationsgrenze mit 5.37 festgelegt und die obere mit 5.44 und beträgt die Standardabweichung 0,01974, dann ergibt sich $PP = \frac{5,44 - 5,37}{6 * 0,01974} = 0,591$.

- *Ist-% außerhalb der Genzwerte.* Wird dieses Auswahlkästchen aktiviert, wird in der Ausgabe zusätzlich angegeben, wieviel Prozent der Messwerte außerhalb der durch Minimum und Maximum spezifizierten Grenzen liegen.
- Versorgung mit Titel und Fußnoten.
- Grafiklayout aus einer Vorlage übernehmen (⇒ Kap. 32.2).

32.20.2 Pareto-Diagramme

Pareto-Diagramme sind Balkendiagramme zur grafischen Darstellung von Häufigkeiten einer kategorialen Variablen, wobei in der Darstellung die Häufigkeiten der Kategorien der Größe nach geordnet werden: zuerst die Kategorie mit der größten Häufigkeit, dann die mit der zweitgrößten Häufigkeit usw. Wahlweise können durch eine Linie die kumulierten Häufigkeiten in dem Diagramm zusätzlich dargestellt werden. Ein Pareto-Diagramm wird sinnvoll immer dann verwendet, wenn eine Variable viele Kategorien hat und man daran interessiert ist, welche Kategorien die größten Häufigkeiten haben.

Um ein Pareto-Diagramm zu erstellen, öffnet man durch Klicken der Befehlsfolge

▷ „Analysieren“, „Qualitätskontrolle“, „Pareto-Diagramme...“

die in Abb. 32.70 dargestellte Dialogbox. Als Diagrammtypen sind ein einfaches und ein gestapeltes Diagramm wählbar, wobei beide Diagrammtyp drei unterschiedliche Arten von Daten darstellen können: „Häufigkeiten oder Summen für Kategorien einer Variablen“, „Summen verschiedener Variablen“ oder „Werte einzelner Fälle“.

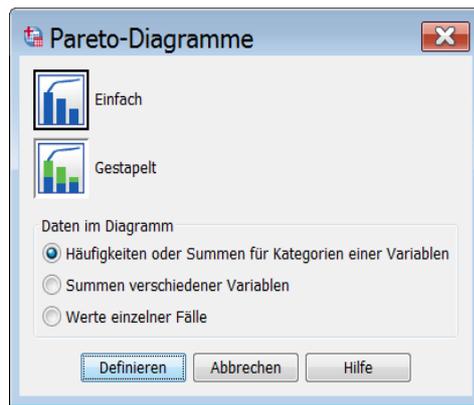


Abb. 32.70. Dialogbox zur Auswahl eines Pareto-Diagramms

Im Folgenden werden einige dieser Diagrammformen anhand von Daten zu Qualitätsmerkmalen von Zigaretten dargestellt. Es handelt sich dabei um festgestellte Mängel von geprüften Zigaretten. Die Mängel bzw. Fehler sind in Form von Abweichungen des Durchmessers, des Gewichts, des Zugwiderstandes und der Ventilation von Normgrenzwerten dieser Merkmale definiert. Die Normgrenzwerte wurden für diese Darstellungszwecke festgelegt.

32.20.2.1 Einfaches Pareto-Diagramm

Häufigkeiten oder Summen für Kategorien einer Variablen. *Beispiel:* Das folgende Übungsbeispiel bezieht sich auf die Variable FEHL_A der Datei ZIGARETT1.SAV.). Die Variable FEHL_A ist eine kategoriale Variable, die Mängelarten der auf der Anlage A produzierten Zigaretten erfasst, beispielsweise bedeutet der Variablenwert „1“ ein zu kleiner Durchmesser, „2“ ein zu großer Durchmesser, „3“ eine zu geringe Filterventilation, „5“ zu kleiner Zugwiderstand, „7“ zu kleines Gewicht usw.⁵ Die Fälle mit dem Variablenwert „0“ (ohne Fehler) werden mittels Menü „Daten, „Fälle auswählen“ ausgeschlossen⁶

Nach der Befehlsfolge „Analysieren“, „Qualitätskontrolle“, „Pareto-Diagramme...“ wird zunächst die Auswahlkombination „Einfach“ und „Häufigkeiten oder Summen für Kategorien einer Variablen“ angeklickt, darauf die Schaltfläche „Definieren“. Abb. 32.71 zeigt links die nach Klicken von „Definieren“ geöffnete Dialogbox mit einem Beispiel zur Grafikdefinition und rechts die daraus resultierende Grafik. Die Variable FEHL_A, in der die verschiedenen Mängel der Zigaretten kodiert sind, wurde aus der Quellvariablenliste in das Eingabefeld „Kategorienachse:“ übertragen. Im Auswahlfeld „Bedeutung der Balken“ ist „Häufigkeiten“ angeklickt. Damit wird eine Darstellung der Häufigkeiten der Fehlerarten in Form von Balken angefordert. Alternativ dazu kann auch „Summe der Variablen“ gewählt und eine metrische Variable in das vorgesehene Feld übertragen werden. Für jeden Variablenwert der kategorialen Variablen werden dann die Summen der

⁵ Diese Variable wurde durch Rekodierung aus verschiedenen metrischen Variablen der Datei ZIGARETT.SAV (⇒ ausschnittsweise in Abb. 32.60, Kap. 32.20.1) hergestellt.

⁶ Mit der Spezifikation FEHL_A > 0 für „Falls Bedingung zutrifft“.

metrischen Variablen dargestellt. Diese Option wäre z.B. dann angebracht, wenn in der Datendatei die Daten anders erfasst wären, wenn z.B. neben der kategorialen Variablen eine metrische Variable die Häufigkeiten von Fehlern erfassen würde. Die Option „Kumulierte Linie anzeigen“ ist aktiv geschaltet (Voreinstellung).

In der Abb. 32.71 rechts ist das Pareto-Diagramm dargestellt. Auf der X-Achse sind die Fehlerarten und auf der Y-Achse die Häufigkeiten (links als absolute Anzahl und rechts als prozentualer Wert) der Fehler abgebildet. Im Unterschied zu Balkendiagrammen wird die Reihenfolge der dargestellten Kategorien nach der Größe der Häufigkeit geordnet: der erste Balken stellt die Häufigkeit der Kategorie „mehrere Fehler“ dar, da diese die größte Häufigkeit hat. Auf den Balken werden die Häufigkeiten auch zahlenmäßig aufgeführt. Die Linie im Diagramm zeigt die kumulierten Häufigkeiten. Sie erscheint nur dann, wenn die Option „Kumulative Linie anzeigen“ gewählt worden ist.

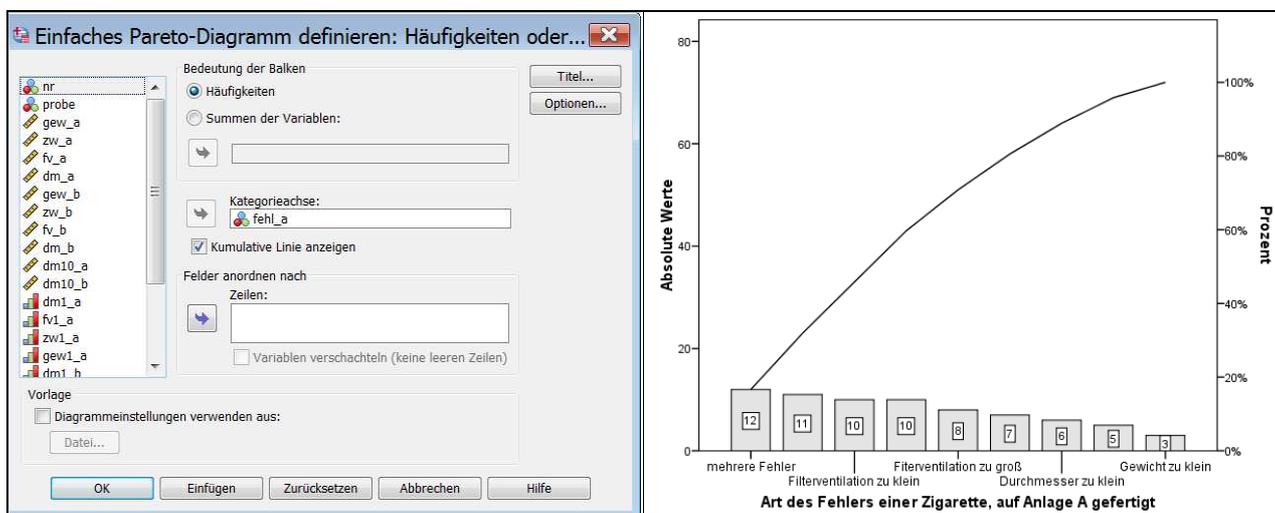


Abb. 32.71. Häufigkeit von Fehlern von Zigaretten im einfachen Pareto-Diagramm

Summen verschiedener Variablen. Beispiel. Für das folgende Demonstrationsbeispiel sind die Informationen zu Fehlern der Zigaretten in anderer Form aufbereitet. In der in Abb. 32.72 dargestellten Datei FEHLER.SAV stellt jeder Fall (eine Zeile im Daten-Editor) eine Probe von aus der laufenden Produktion entnommenen Zigaretten mit je zwanzig Zigaretten dar. Mit den Variablen N_FDM und N_FGW werden die Häufigkeiten (n) eines fehlerhaften Durchmessers (FDM) bzw. fehlerhaften Gewichts (FGW) erfasst.

	probe	n_fdm	n_fgw
1	1	1	3
2	2	1	0
3	3	1	5
4	4	0	4
5	5	4	2
6	6	5	3
7	7	0	3
8	8	2	1

Abb. 32.72. Daten der Datei FEHLER.SAV im Daten-Editor (Ausschnitt)

Nach der Befehlsfolge „Analysieren“, „Qualitätskontrolle“, „Pareto-Diagramme...“ wird die Auswahlkombination „Einfach“ und „Summen verschiedener Variablen“ angeklickt. Abb. 32.73 zeigt links die geöffnete Dialogbox mit einem Beispiel zur Grafikdefinition. Die Variablen N_FDM und N_FGW mit den Häufigkeiten eines fehlerhaften Durchmessers bzw. fehlerhaften Gewichts wurden aus der Quellvariablenliste in das Eingabefeld „Variablen:“ übertragen. Die Option „Kumulative Linie anzeigen“ ist nicht gewählt.

In der Abb. 32.73 rechts ist das Pareto-Diagramm dargestellt. Auf der X-Achse sind die beiden Variablen und auf der Y-Achse die summierten Häufigkeiten abgebildet. Im Unterschied zu Balkendiagrammen wird die Reihenfolge der dargestellten Balken nach deren Höhe geordnet.

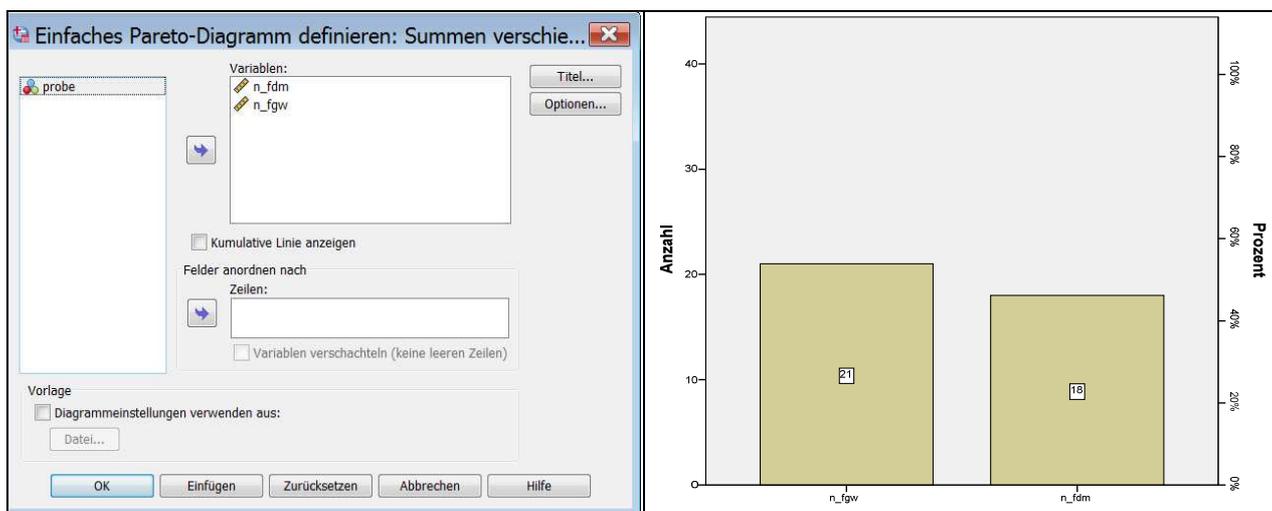


Abb. 32.73. Häufigkeiten fehlerhaften Gewichts bzw. Durchmessers im Pareto-Diagramm

Werte einzelner Fälle. In einem derartigen Diagramm werden für jeden Fall die Variablenwerte in der Reihenfolge ihrer Größe dargestellt: zuerst der Fall mit dem höchsten, dann mit zweithöchsten Wert usw.

Als „Achsenbeschriftung“ kann „Fallnummer“ oder eine Variable gewählt werden. Im zweiten Fall würde auf der X-Achse für jeden dargestellten Fall der Wertelabel dieser Variablen erscheinen.

32.20.2.2 Gestapeltes Pareto-Diagramm

Häufigkeiten oder Summen für Kategorien einer Variablen. *Beispiel:* Es wird wieder die kategoriale Variable FEHL_A mit Werten für verschiedene Fehlerarten der Datei ZIGARET1.SAV (erstellt aus den Messdaten der Datei ZIGARETT.SAV ⇒ Abb. 32.60) verwendet⁷. Stapelvariable ist die Variable SCHICHT, die mit den Werten „1“ und „2“ erfasst, ob die Zigaretten aus der Tages- oder Nachtschicht stammen. „Kumulative Linie anzeigen“ ist eingeschaltet.

Nach Auswahl der Fälle mit FEHL_A größer 0 („Daten“. „Fälle auswählen“, „Falls Bedingung zutrifft“), wird die Befehlsfolge „Analysieren“, „Qualitätskontrolle“, „Pareto-Diagramme...“ mit der Auswahlkombination „Gestapelt“ und „Häufigkeiten oder Summen für Kategorien einer Variablen“ ausgeführt. Abb. 32.74 zeigt links die nach Klicken von „Definieren“ geöffnete Dialogbox mit einem Beispiel zur Grafikdefinition. Die Variable FEHL_A wurde aus der Quellvariablenliste in das Eingabefeld „Kategorienachse“ übertragen. Im Auswahlfeld „Bedeutung der Balken“ ist wie im einfachen Pareto-Diagramm „Häufigkeiten“ angeklickt. Wie dort beschrieben ist, kann auch die andere Option gewählt werden. Die Variable SCHICHT wurde in das Eingabefeld „Stapel definieren durch:“ übertragen. Die Option „Kumulative Linie anzeigen“ ist eingeschaltet.

In der Abb. 32.74 rechts ist das aus dieser Einstellung resultierende Pareto-Diagramm dargestellt. Es gleicht einem einfachen Diagramm mit dem Unterschied, dass die Häufigkeit für jede Fehlerart nach den Werten der Stapelvariablen untergliedert wird. Nun kann man erkennen, wie sich die Häufigkeiten eines jeden Fehlers auf die Produktionszeiten Tages- und Nachtschicht aufteilen.

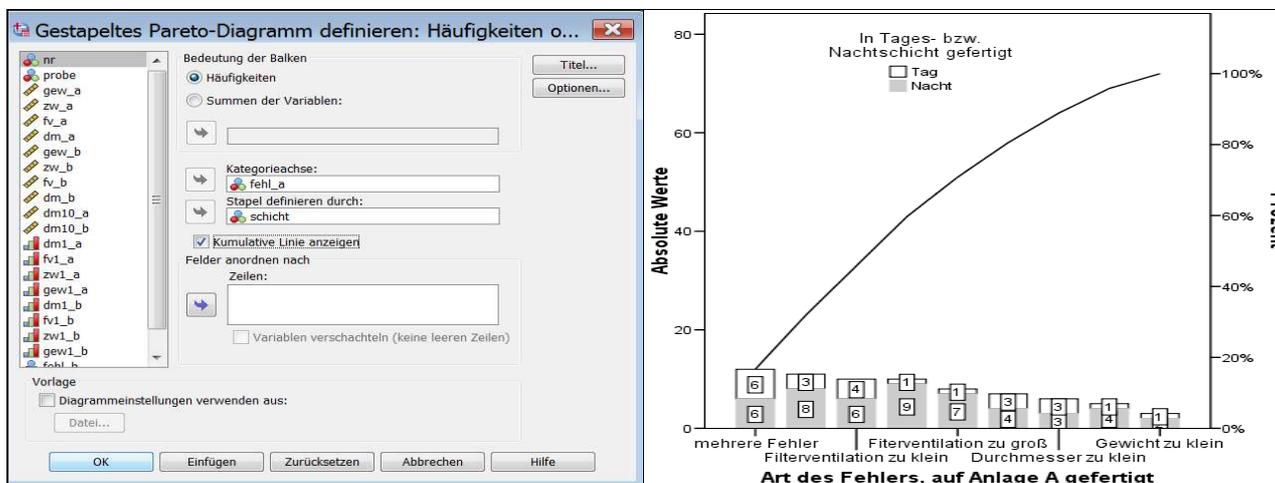


Abb. 32.74. Häufigkeit von Fehlern von Zigaretten untergliedert nach Tages- und Nachtschicht im gestapelten Pareto-Diagramm

Summen verschiedener Variablen. *Beispiel.* Das folgende Beispiel bezieht sich auf die in Abb. 32.72 dargestellte Datendatei (FEHLER.SAV), in der die Häufigkeiten von zwei Fehlerarten von Zigaretten erfasst sind. Jeder Fall der SPSS-Da-

⁷ Nur Fälle mit Werten bei FEHL_A größer 0.

tendatei ist eine Stichprobe von zwanzig Zigaretten, die aus der laufenden Produktion entnommenen wurden. Mit den Variablen N_FDM und N_FGW werden die Häufigkeiten (n) eines fehlerhaften Durchmessers (FDM) bzw. fehlerhaften Gewichts (FGW) in der jeweiligen Stichprobe erfasst.

Nach der Befehlsfolge „Analysieren“, „Qualitätskontrolle“, „Pareto-Diagramme...“ wird die Auswahlkombination „Gestapelt“ und „Summen verschiedener Variablen“ angeklickt. Abb. 32.75 zeigt links die geöffnete Dialogbox mit einem Beispiel zur Grafikdefinition und rechts die resultierende Grafik.

Werte einzelner Fälle. Bei der Auswahlkombination „Gestapelt“ und „Werte einzelner Fälle“ werden für jeden Fall die addierten Variablenwerte von zwei Variablen dargestellt. Die Reihenfolge der Darstellung der Balken orientiert sich wieder an der Höhe der Balken. Die Grafik entspricht der in Abb. 32.75 mit dem Unterschied, dass die Werte einzelner Fälle überlagert dargestellt werden.

32.20.2.3 Wahlmöglichkeiten

Für fast alle Pareto-Diagramme bestehen folgende Wahlmöglichkeiten

- Diagramme in Feldern (⇒ nur Zeilen, Kap. 32.4).
- Versorgung mit Titel und Fußnoten.
- Behandlung fehlender Werte (⇒ Kap. 32.2.1).
- Grafiklayout aus Vorlage übernehmen (⇒ Kap. 32.2.1).
- „Kumulative Linien anzeigen“. Ist diese Option aktiv, so wird im Diagramm eine Kurve der kumulierten Häufigkeiten bzw. Werte angezeigt (⇒ Abb. 32.74).

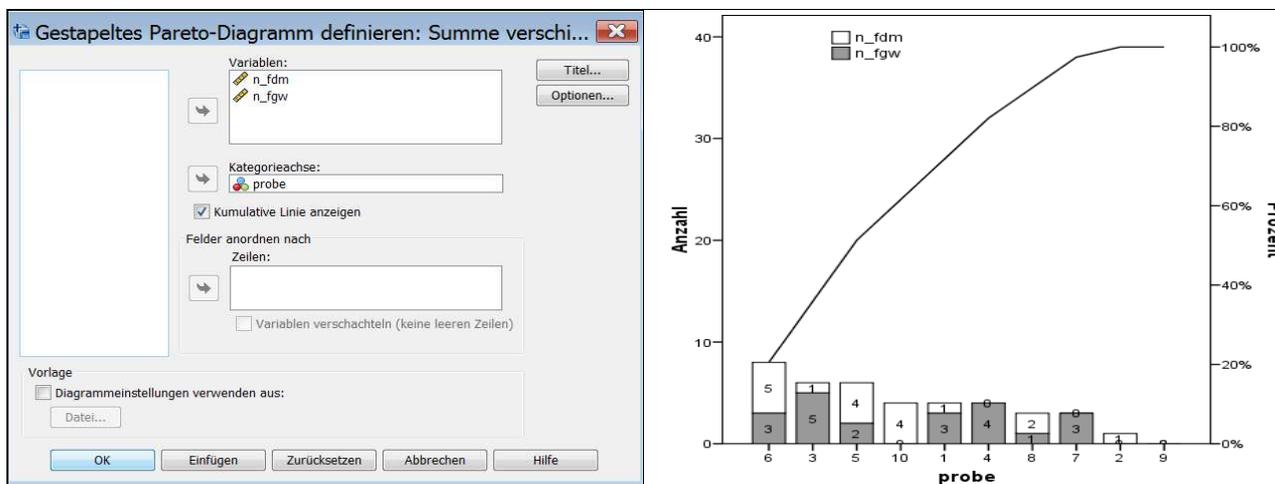


Abb. 32.75. Häufigkeiten fehlerhaften Gewichts bzw. Durchmessers im gestapelten Pareto-Diagramm