Alternative Konzepte zur Laufzeitenermittlung:

Statistisch-theoretische Grundlagen

Dr. Jürgen Faik

- Reinfeld, 20.11.2007 -

1. Einleitung

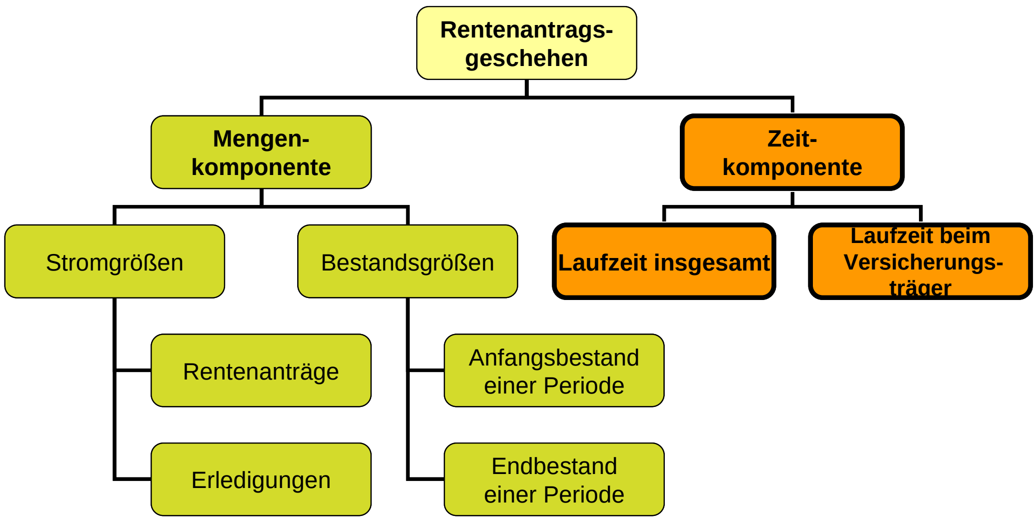
In der sozialpolitischen Diskussion, welche sich auf die gesetzliche Rentenversicherung (GRV) bezieht, stehen üblicherweise andere Aspekte als das Rentenantragsgeschehen im Vordergrund. So waren etwa im Rahmen der Rentenreformdiskussionen der vergangenen Jahre vorrangig Daten aus der Versicherten-, der Rentenzugangs- und der Rentenbestandsstatistik gefragt. Wenn von Verrentung gesprochen wird, muss aber stets bedacht werden, dass jede Rente *beantragt* werden muss. Dem Rentenantragsgeschehen kommt somit eine Lead-Funktion in Bezug auf das Verrentungsgeschehen zu. Diesem zeitlichen Vorlauf entsprechend, sind die Elemente der Rentenantragsstatistik durchaus wichtige sozialpolitische Indikatoren für künftige Rentenentwicklungen. Dieser Gedanke wird noch dadurch verstärkt, dass die Rentenantrags- und Erledigungsstatistik bei der Deutschen Rentenversicherung Bund - im Unterschied zu anderen, jahresbezogenen Rentenversicherungs-Statistiken wie z. B. der Versicherten-, der Rentenzugangs- oder der Rentenbestandsstatistik - monatsbezogen geführt wird und daher zeitlich vergleichsweise tief gegliederte Analysen zulässt. Hinzu kommt, dass die Rentenantrags- und Erledigungsstatistik sehr zeitnah aufbereitet wird: Die betreffenden Informationen eines spezifischen Monats stehen üblicherweise bereits Mitte des Folgemonats zur Verfügung. Aufgrund der spezifischen Vorzüge der Rentenantrags- und Erledigungsstatistik (große zeitliche Disaggregationstiefe und große Zeitnähe zum Erhebungszeitpunkt) eignet sie sich in besonderem Maße dazu, Entwicklungen des Verrentungsgeschehens frühzeitig zu erkennen und auf sie zu reagieren. Für den handelnden Sozialpolitiker ist die Kenntnis von Rentenantragskennziffern und ihren Bestimmungsgründen daher alles andere als irrelevant, wenngleich weitreichende statistisch fundierte Erklärungsansätze insofern unterbleiben müssen, als die Rentenantrags- und Erledigungsstatistik – als Folge ihres Designs als GRV-Arbeitsstatistik – nur eine unzureichende Anzahl an Erklärungsvariablen anbietet.

2. Begriffsbestimmungen

Vor einer analytischen Durchdringung des Rentenantragsgeschehens ist es sinnvoll, die verschiedenen Dimensionen des Rentenantragsgeschehens einer Systematisierung zu unterwerfen. Eine mögliche Systematisierung in diesem Zusammenhang ist die Trennung des Rentenantragsgeschehens in eine Mengen- und eine Zeitkomponente (siehe Abbildung 1).

Abbildung 1:

Übersicht über das Rentenantragsgeschehen in Deutschland:



Quelle: Faik 1999, S. 77

Im Rahmen der *Mengenkomponente* sind zunächst die einzelnen zu bearbeitenden Vorgangsarten voneinander zu differenzieren. Da die Mengenkomponente in meinem Vortrag keine größere Rolle spielt, wird auf diesen Ast von Abbildung 1 nicht näher eingegangen.

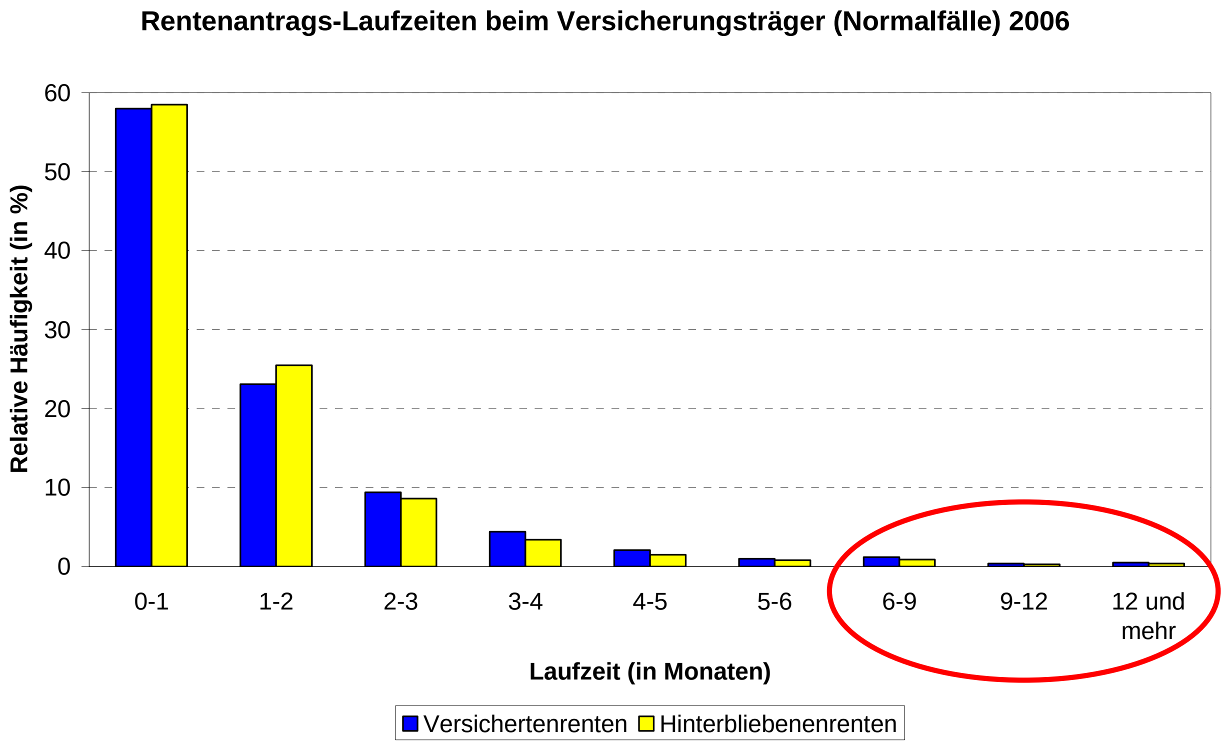
Die zweite Komponente des Rentenantragsgeschehens - die *Zeitkomponente* – steht im Fokus meines Vortrags. Sie umfasst die Rentenantragslaufzeiten, wobei üblicherweise zwischen der (durchschnittlichen) Rentenantragslaufzeit insgesamt und der (durchschnittlichen) Rentenantragslaufzeit beim Versicherungsträger unterschieden wird. Beide Größen unterscheiden sich um die sogenannte Vorlaufzeit voneinander – d. h. um die Zeit, die vergeht, bis der Rentenantrag nach dem Tag der Antragstellung beim verantwortlichen Rentenversicherungsträger eingegangen ist. Sinnvollerweise werden Rentenantragslaufzeiten nur für abgeschlossene, d. h. für erledigte oder für bewilligte Rentenanträge errechnet.

Zwischen der Mengen- und der Zeitkomponente des Rentenantragsgeschehens bestehen Rückkopplungseffekte. So geht z. B. ein hoher Rentenantragsbestand typischerweise mit recht langen (durchschnittlichen) Rentenantragslaufzeiten einher et vice versa. Hierauf werden wir später noch zu sprechen kommen.

3. Grafische Datenaufbereitung

Um einen ersten Einblick in die Thematik zu erhalten, können Häufigkeitstabellen wie die der Laufzeiten-Häufigkeiten grafisch umgesetzt werden. Hierbei eignet sich für klassifizierte Merkmale eine Darstellung in Balken- bzw. Stabform. Es ist für die Darstellung der einzelnen Merkmalsausprägungen im Prinzip unerheblich, welche Breite die Balken haben.

Abbildung 2:



Quelle: Interne Rentenantrags- und Erledigungsstatistik, Jahreswerte 2006, Tabelle 131

Auf der Ordinate sind in Abbildung 2 die relativen Häufigkeiten abgetragen. was sich gerade für den Vergleich zwischen Versicherten- und Hinterbliebenenrenten in diesem Beispiel anbietet, da die Fallzahlen zwischen beiden Gruppen unterschiedlich groß sind. Prinzipiell ist es aber auch möglich, bei einer Häufigkeitsdarstellung auf der Ordinate die absoluten Häufigkeiten zu verwenden.

Gerade wenn in einzelne Klassen nur wenige absolute Häufigkeiten fallen, werden diese Klassen vielfach zusammengefasst und weisen gegenüber den anderen Klassen eine größere Klassenbreite auf. In Abbildung 2 ist dies für Laufzeiten ab 6 Monaten der Fall. Es entsteht das Problem unterschiedlicher Klassenbreiten innerhalb einer Häufigkeitsverteilung. In diesem Fall ist die Häufigkeitsdarstellung streng genommen auf der Ordinate zu modifizieren.

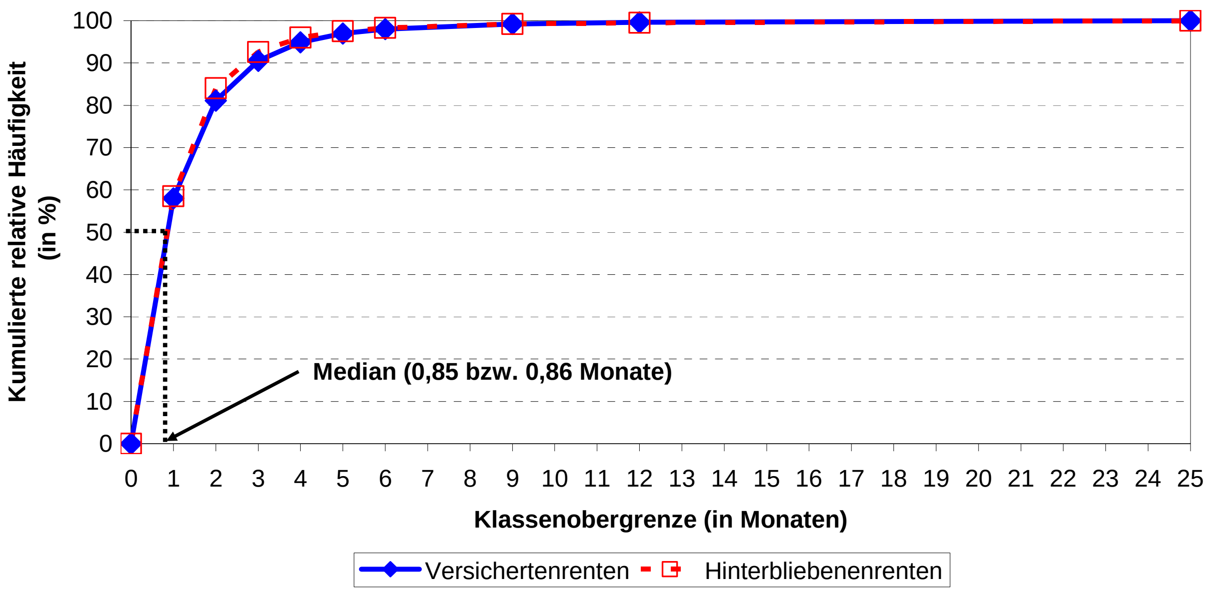
Verwendet man in Abbildung 2 z. B. naheliegenderweise eine normierte Klassenbreite von 1 Monat, müssten eigentlich die relativen Häufigkeiten der drei umkreisten Klassen durch die jeweilige Klassenbreite dividiert werden, was – nebenbei bemerkt – bei der nach oben offenen Klasse „12 und mehr Monate“ problematisch ist. Da die zugehörigen normierten relativen Häufigkeiten in Abbildung 2 für diese drei Klassen sehr gering und nicht mehr gut darstellbar sind, wurde hiervon Abstand genommen.

Auch ohne diese eigentlich gebotene Normierung wird die sehr stark linkssteile Häufigkeitsverteilung der Rentenantrags-Laufzeiten sowohl bei Versicherten- als auch bei Hinterbliebenenrenten sichtbar.

Bezieht man sich nicht auf die einfachen relativen bzw. absoluten Häufigkeiten, sondern auf die kumulierten relativen bzw. absoluten Häufigkeiten, erhält man ein Summenpolygon bzw. anders formuliert: eine Verteilungsfunktion. Von kumulierten absoluten bzw. kumulierten relativen Häufigkeiten spricht man im Falle sukzessive addierter absoluter bzw. relativer Häufigkeiten. In Abbildung 3 fragt man sich entsprechend, in wie viel Prozent aller Fälle die Rentenantragslaufzeit maximal 1 Monat, maximal 2 Monate, maximal 3 Monate usw. dauerte. Bei der grafischen Darstellung ist daher auf die jeweilige Klassenobergrenze Bezug zu nehmen. Beispielsweise ist an Hand von Abbildung 3 ersichtlich, dass in etwa 60 Prozent der Fälle sowohl bei den Versicherten- als auch bei den Hinterbliebenenrenten die Laufzeit maximal 1 Monat dauerte.

Abbildung 3:

Rentenantragslaufzeiten beim Versicherungsträger (Normalfälle) 2006, Verteilungsfunktionen



Quelle: Interne Rentenantrags- und Erledigungsstatistik, Jahreswerte 2006, Tabelle 131

Die 50-Prozent-Grenze des Summenpolygons bzw. der Verteilungsfunktion trennt die Verteilung in die unteren 50 % und die oberen 50 % der Verteilung. Diese Grenze wird auch als Median bezeichnet. In Abbildung 3 liegt sie bei den Versichertenrenten bei ca. 0,86 Monaten, und bei den Hinterbliebenenrenten bei ca. 0,85 Monaten, d. h. bei jeweils ca. 26 Tagen.

4. Mittelwerte

Es ist möglich, dass der Form nach gleiche Häufigkeitsverteilungen mit unterschiedlichen Niveaus der Merkmalsausprägungen einhergehen. Zur Differenzierung zweier Häufigkeitsverteilungen identischer Form, aber unterschiedlicher Lage ist die Angabe eines entsprechenden Lageparameters sinnvoll, welcher den Niveaueffekt in einer einzelnen Zahl beschreibt. Man spricht von Maßen der zentralen Verteilungstendenz oder von repräsentativen Wertgrößen bzw. einfacher: von Mittelwerten.

Die gebräuchlichsten Mittelwerte sind: Der Modus, der Median, das arithmetische und das geometrische Mittel. Auf sie wird nachfolgend jeweils kurz eingegangen.

Übersicht 1:

Mittelwertkonzepte:

Modus: $X_{Modus} = X_i, \quad f_i = \max$

Median, allgemein: $X_{Median} = X_{F(0,5)}$

Median, ungerades N: $X_{Median} = X_{\left[\frac{N+1}{2}\right]}$

Median, gerades N: $X_{Median} = \frac{1}{2} \cdot \left(X_{\left[\frac{N}{2}\right]} + X_{\left[\frac{N}{2}+1\right]} \right)$

Median, klassifizierte Daten: $X_{Median} = X_i^u + \frac{0,5 - F(X_i^u)}{F(X_i^o) - F(X_i^u)} \cdot (X_i^o - X_i^u)$

Arithmetisches Mittel, ungewichtet: $\overline{X} = \frac{1}{N} \cdot \sum_{i=1}^{N} X_i$

Arithmetisches Mittel, gewichtet: $\overline{X} = \frac{1}{N} \cdot \sum_{i=1}^{K} X_i \cdot h_i = \sum_{i=1}^{K} X_i \cdot f_i$

Geometrisches Mittel: $GM = \sqrt[N]{\prod_{i=1}^{N} X_i} \quad \Leftrightarrow \quad \log GM = \frac{1}{N} \sum_{i=1}^{N} \log X_i$

Quelle: Faik 2007, S. 267

4.1 Modus

Der Modus gibt den häufigsten Merkmalswert einer Häufigkeitsverteilung an. Liegen die Daten klassifiziert vor, wählt man zunächst die Klasse mit der größten relativen bzw. absoluten Häufigkeit aus und nimmt als Modus einen möglichst repräsentativen Wert der ausgewählten Klasse. In der Regel entscheidet man sich dabei für die Klassenmitte.

4.2 Median

Wie bereits erwähnt, ergibt sich der Median an der 50-Prozent-Marke der kumulierten relativen Häufigkeiten. Liegen klassifizierte Daten vor, fällt der Median in die Klasse hinein, bei der eine kumulierte relative Häufigkeit von 0,5 erreicht wird. Man könnte dann den Klassenmittelwert als Median verwenden. Vielfach nutzt man indes in diesem Fall – auf Basis des Strahlensatzes – eine Näherungslösung für den Median.

In der Regel ist der Median resistent gegen eine nach unten offene unterste Klasse bzw. gegen eine nach oben offene oberste Klasse. Daher bietet er sich in unserem Falle einer nach oben offenen obersten Klasse in Abbildung 1 grundsätzlich als Indikator für die zentrale Verteilungslage an.

4.3 Arithmetisches Mittel

Das arithmetische Mittel gibt den Durchschnittswert einer Häufigkeitsverteilung durch Addition der Einzelwerte und anschließender Division durch die Anzahl aller Fälle an.

Das arithmetische Mittel ist nicht unbedingt ein konkreter Wert aus der Menge aller Merkmalsausprägungen. Anders formuliert, ist das arithmetische Mittel mitunter in Bezug auf die einzelnen Merkmalsausprägungen ein abstrakter Wert.

Bei klassifizierten Daten verwendet man als Merkmalsausprägungen die Klassenmitten. D. h.: Es wird unterstellt, dass die Merkmalsausprägungen innerhalb der Klassen keine gehäuften Abweichungen von der Klassenmitte nach oben bzw. unten aufweisen.

Das arithmetische Mittel reagiert sensitiv auf „Ausreißerwerte“ nach oben bzw. unten. Das ist beispielsweise dann problematisch, wenn man für die oberste nach oben offene Klasse einen konkreten Wert annehmen muss.

4.4 Geometrisches Mittel

Für Veränderungsraten ist das arithmetische Mittel kein geeignetes Maß. In einem solchen Fall verwendet man das geometrische Mittel.

Dieses ist nur für nicht-negative Werte definiert. Dies macht es erforderlich, bei seiner Berechnung nicht direkt auf die Veränderungsraten, sondern auf so genannte Wachstumsfaktoren abzustellen. Letztere ergeben sich daraus, dass man jeweils für zwei aufeinanderfolgende Merkmalsausprägungen den Quotienten bildet.

Das geometrische Mittel kann übrigens durch Logarithmierung in das arithmetische Mittel der logarithmierten Merkmalswerte überführt werden.

4.5 Empirische Mittelwertergebnisse

Die empirischen Ergebnisse ergeben für die Laufzeiten der Versicherten- wie Hinterbliebenenrenten jeweils einen Modus von 15 Tagen und einen Median von 26 Tagen. Das auf Grundlage der vorgegebenen Klassen – sowie unter der Annahme einer obersten Klassenmitte in Höhe von 20 Monaten – berechnete arithmetische Mittel beläuft sich bei den Versichertenrenten auf 43 Tage und bei den Hinterbliebenenrenten auf 40 Tage.

Übersicht 2a:

Mittelwerte, Rentenantragslaufzeiten der Versichertenrenten (Normalfälle) 2006

(berechnet aus gruppierten Daten):

- Modus: 0,5 Monate bzw. 15 Tage
- Median: 0,86 Monate bzw. 26 Tage
- Arithmetisches Mittel: 1,4 Monate bzw. 43 Tage
  (oberste Klassenmitte auf 20 Monate gesetzt;
  zum Vergleich arithmetisches Mittel aus Individualdaten:
  39 Tage)

Quelle: Eigene Berechnungen auf Basis der internen Rentenantrags- und Erledigungsstatistik, Jahreswerte 2006, Tabelle 131

Übersicht 2b:

Mittelwerte, Rentenantragslaufzeiten der Hinterbliebenenrenten (Normalfälle) 2006

(berechnet aus gruppierten Daten):

- Modus: 0,5 Monate bzw. 15 Tage
- Median: 0,85 Monate bzw. 26 Tage
- Arithmetisches Mittel: 1,3 Monate bzw. 40 Tage
  (oberste Klassenmitte auf 20 Monate gesetzt;
  zum Vergleich arithmetisches Mittel aus Individualdaten:
  37 Tage)

Quelle: Eigene Berechnungen auf Basis der internen Rentenantrags- und Erledigungsstatistik, Jahreswerte 2006, Tabelle 131

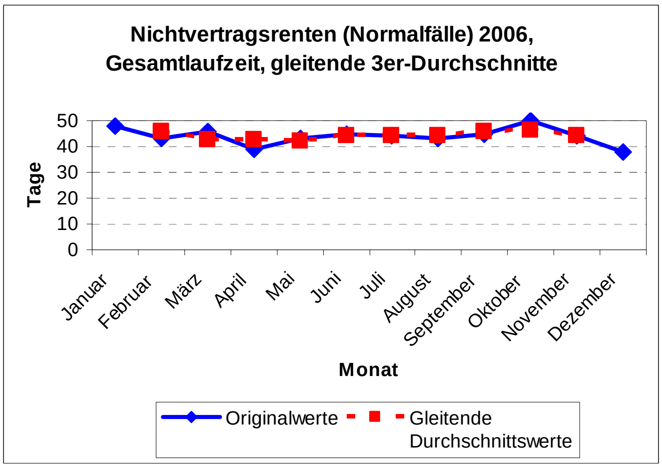
Gegenüber den exakten arithmetischen Mittelwerten aus den nicht-klassifizierten Daten sind diese Mittelwerte um 4 bzw. 3 Tage höher, was die Problematik der Verwendung klassifizierter Wertangaben andeutet.

Arithmetische Mittelwerte können bei Zeitreiheninformationen genutzt werden, um saisonale Schwankungen zu glätten. Man verwendet dann so genannte gleitende Durchschnittswerte. Diese sind dadurch charakterisiert, dass für eine bestimmte Anzahl aufeinanderfolgender Originalwerte der jeweilige Durchschnitt gebildet wird. In Abbildung 4 wurden die Monatswerte des Jahres 2006 für die durchschnittlichen Gesamtlaufzeiten der Nichtvertragsrenten (Normalfälle) auf der Basis dreier aufeinanderfolgender Werte, d. h. auf Basis von 3er-Durchschnitten, geglättet. Ersichtlicherweise geht bei einer solchen Vorgehensweise am Anfang und am Ende der Zeitreihe jeweils ein Beobachtungswert verloren.

Abbildung 4:

Gleitende (3er-)Durchschnitte, Gesamtlaufzeiten der Nichtvertragsrenten (Normalfälle) 2006 (in Tagen):

	Originalwerte	Gleitende Durchschnitts werte
Januar	48	
Februar	43	46
März	46	43
April	39	43
Mai	43	42
Juni	45	44
Juli	44	44
August	43	44
September	45	46
Oktober	50	46
November	44	44
Dezember	38	



Quelle: Eigene Berechnungen auf Basis der internen Rentenantrags- und Erledigungsstatistik, Jahreswerte 2006, Tabelle 233

Da die Original-Zeitreihe in Abbildung 4 nur eine geringe saisonale Variation aufweist, ist der Glättungseffekt vergleichsweise bescheiden.

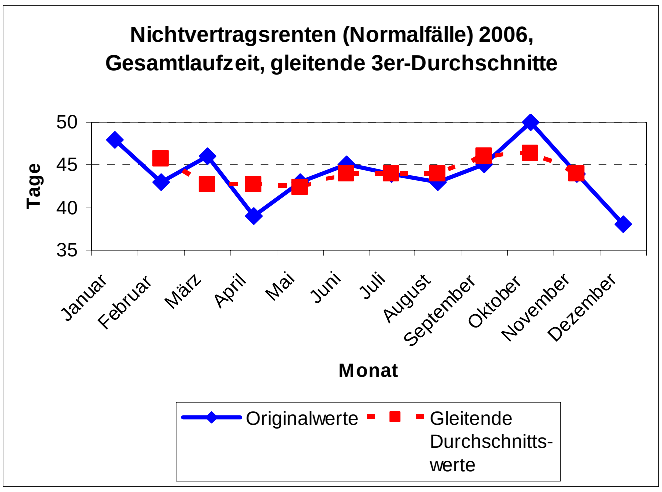
In nicht wenigen, statistisch allerdings eher unbedarften Darstellungen beginnen die Ordinatenwerte nicht im Ursprung, sondern bei einem höheren Wert, wodurch die einzelnen Werte stärker auseinander gezogen werden als bei einer ursprungsbezogenen Darstellung. Aus Illustrationszwecken habe ich in Abbildung 5 die Ordinate erst bei einem Wert von 35 Tagen starten lassen. Auf diese Weise sind zwar auf der einen Seite die Veränderungen zwischen den Monaten deutlicher zu sehen; es könnte aber auf der anderen Seite der unzulässige Eindruck entstehen, dass es sich um beträchtliche Wertunterschiede zwischen den einzelnen Beobachtungsmonaten handelt. Tatsächlich liegen zwischen maximaler und minimaler Laufzeit aber nur 11 Tage. Es ist daher seriöserweise von einer Ordinaten-Transformation wie der in Abbildung 5 abzuraten.

Abbildung 5:

ACHTUNG: Statistische Manipulation!

Gleitende (3er-)Durchschnitte, Gesamtlaufzeiten der Nichtvertragsrenten (Normalfälle) 2006 (in Tagen):

	Originalwerte	Gleitende Durchschnittswerte
Januar	48	
Februar	43	46
März	46	43
April	39	43
Mai	43	42
Juni	45	44
Juli	44	44
August	43	44
September	45	46
Oktober	50	46
November	44	44
Dezember	38	



Quelle: Eigene Berechnungen auf Basis der internen Rentenantrags- und Erledigungsstatistik, Jahreswerte 2006, Tabelle 233

Interessiert man sich für die durchschnittliche Gesamtlaufzeit der in den Abbildungen 4 und 5 angegebenen Originalwerte, liegt die Berechnung des arithmetischen Mittels im bekannten Sinne der Wertesumme aller

Monate dividiert durch die Anzahl der 12 Monate nahe. In diesem Falle ergibt sich ein (arithmetischer) Mittelwert in Höhe von 44 Tagen.

Interessiert man sich demgegenüber für die monatsbezogenen Veränderungen der Gesamtlaufzeit, ist, wie erwähnt, das geometrische Mittel als Indikator zu wählen. Es berechnet sich in unserem Fall als Produkt der angegebenen (insgesamt 11) Wachstumsfaktoren, aus dem die elfte Wurzel gezogen wird. Anschließend ist der Wert 1 zu subtrahieren, und man erhält dadurch in unserem Fall eine durchschnittliche monatsbezogene Veränderung von ca. -2,1 %. D. h.: Im Jahr 2006 haben sich die Gesamtlaufzeiten der Nichtvertragsrenten (Normalfälle) von Monat zu Monat *durchschnittlich* um etwa 2,1 % vermindert.

Übersicht 3:

Geometrisches Mittel, Gesamtlaufzeiten der Nichtvertragsrenten, Normalfälle (Wachstumsraten) 2006:

	Gesamtlaufzeit (in Tagen)	Wachstums-faktoren
Januar	48	
Februar	43	0,8958
März	46	1,0698
April	39	0,8478
Mai	43	1,1026
Juni	45	1,0465
Juli	44	0,9778
August	43	0,9773
September	45	1,0465
Oktober	50	1,1111
November	44	0,8800
Dezember	38	0,8636
	Geometrisches Mittel (in %):	-2,1014

Quelle: Eigene Berechnungen auf Basis der internen Rentenantrags- und Erledigungsstatistik, Jahreswerte 2006, Tabelle 233

5. Streuungsmaße

Ein gleicher Mittelwert kann mit einer unterschiedlich homogenen Häufigkeitsverteilung verbunden sein. Es stellt sich das Problem der Streuung. Analog zu den verschiedenen Mittelwertkonzepten stehen auch hier diverse (Streuungs-)Messziffern zur Verfügung.

Übersicht 4:

Streuungskonzepte:

Spannweite: $SPW = X_{\max} - X_{\min}$

Quartilsabstand: $QA = X_{0,75} - X_{0,25}$

Mittlere absolute Abweichung: $MAA = \frac{1}{N} \cdot \sum_{i=1}^{K} \left| X_i - \overline{X} \right| \cdot h_i$

Standardabweichung: $S = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{K} \left(X_i - \overline{X} \right)^2 \cdot h_i}$

Varianz: $S^2 = \frac{1}{N} \cdot \sum_{i=1}^{K} \left(X_i - \overline{X} \right)^2 \cdot h_i$

Variationskoeffizient: $V = \frac{S}{\overline{X}}$

Quelle: Faik 2007, S. 267-268

5.1 Spannweite

Die Spannweite (synonym: Range) z. B. gibt den Abstand zwischen der größten und der kleinsten Merkmalsausprägung an.

Bei klassifizierten Daten ist sinnvollerweise auf die Klassenmittelwerte der obersten und der untersten Klasse Bezug zu nehmen. Die Spannweite ist problematisch, da sie die Verteilung zwischen den beiden Extremwerten vernachlässigt. Bei nach unten offener unterster Klasse bzw. bei nach oben offener oberster Klasse kann dies mehr oder weniger große Ergebnisverzerrungen mit sich bringen. Da in Abbildung 1 eine nach oben offene oberste Klasse vorlag und die Werte dieser Abbildung auch den nachfolgend präsentierten Streuungsberechnungen zugrunde liegen, wurde wegen der skizzierten Problematik nachfolgend auf den Ausweis einer empirischen Spannweite verzichtet.

5.2 Quartilsabstand

Das Problem der Vernachlässigung von Merkmalsausprägungen stellt sich auch bei Verwendung des Maßes Quartilsabstand, wenngleich bei dieser Kenngröße Verzerrungen durch Extremwerte weniger gravierend als bei der Spannweite sind.

Beim Index des Quartilsabstandes werden die Merkmalsausprägungen auf Basis der kumulierten relativen Häufigkeiten in vier prozentual gleich große Bereiche (= in Quartile) einsortiert. Danach berechnet man den Quartilsabstand als Differenz des Wertes an der 75-Prozent-Marke (= Obergrenze des dritten Quartils) und des Wertes an der 25-Prozent-Marke (= Obergrenze des ersten Quartils).

Offensichtlich gehen nur das dritte und das unterste Quartil in die Berechnung des Quartilsabstands ein. Das zweite und das oberste Quartil mit den darin enthaltenen Merkmalsausprägungen werden hingegen vernachlässigt. Damit wird auch beim Quartilsabstand die Gesamtstreuung nicht vollständig berücksichtigt.

Bei klassifizierter Verteilung bietet sich zur Berechnung des Quartilsabstands – analog zur Ermittlung des Medians in einem solchen Fall – eine Näherungslösung an.

5.3 Mittlere absolute Abweichung

Das Problem der Vernachlässigung von Merkmalsausprägungen bei der Streuungsberechnung, welches sowohl der Spannweite als auch dem Quartilsabstand immanent ist, wird bei anderen Streuungsmaßen vermieden. Solche Streuungsmaße sind etwa jene, die auf Abweichungen von einem Mittelwert, z. B. vom arithmetischen Mittelwert, basieren. Hierbei dürfen aber nicht die einfachen Abweichungen vom arithmetischen Mittelwert verwendet werden, denn beim arithmetischen Mittel heben sich negative und positive Mittelwert-Abweichungen auf; es würde also nie eine Streuung ausgewiesen.

Um ein solch unsinniges Ergebnis zu vermeiden, verwendet man anstelle der einfachen beispielsweise die absoluten (d. h. betragsmäßigen) Abweichungen vom arithmetischen Mittelwert. Die einzelnen Abweichungen sind dabei also jeweils prinzipiell gleich bedeutsam.

5.4 Varianz und Standardabweichung

Eine Alternative zur Verwendung der absoluten Werteabweichungen vom arithmetischen Mittelwert ist die Quadrierung der entsprechenden Abweichungen. Hierdurch werden größere Abweichungen vom Mittelwert (absolut) stärker als kleinere berücksichtigt.

Das korrespondierende Maß heißt Varianz. Ihre Dimension ist die den Merkmalsausprägungen zugrunde liegende Einheit im Quadrat, also z. B. „Euro im Quadrat“ oder in unserem Fall: „Tage im Quadrat“.

Um die gleiche Dimensionierung des Streuungsmaßes im Vergleich zu den Merkmalsausprägungen zu erhalten – in unserem Fall: „Tage“ –, wird aus der Varianz üblicherweise die Quadratwurzel gezogen, und man erhält die Standardabweichung.

5.5 Variationskoeffizient

Die bisherigen Streuungsmaße hatten allesamt eine bestimmte Einheit. Für Vergleiche zwischen Häufigkeitsverteilungen mit unterschiedlichem Mittelwert ist aber eine dimensionslose Größe vorzuziehen.

Verdoppeln sich beispielsweise sämtliche Merkmalsausprägungen, verdoppelt sich nicht nur das arithmetische Mittel, sondern auch die mittels Standardabweichung ausgewiesene Streuung. Solche problematischen Ergebnisse vermeidet der Variationskoeffizient; er ist üblicherweise als Quotient aus Standardabweichung und arithmetischem Mittel definiert.

5.6 Empirische Streuungsergebnisse

Gemäß Quartilsabstand ist die Streuung bei den Hinterbliebenenrenten um 3 Tage niedriger als bei den Versichertenrenten (37 versus 40 Tage). Auf Basis des umfassenderen Streuungsindikators Mittlere absolute Abweichung geht die betreffende Streuungsdifferenz mit 28 versus 31 Tagen in die gleiche qualitative Richtung.

Übersicht 5a:

Streuungsmaße, Rentenantragslaufzeiten der Versichertenrenten (Normalfälle) 2006
(berechnet aus gruppierten Daten):

- Quartilsabstand: 1,3 Monate bzw. 40 Tage
- Mittlere absolute Abweichung: 1,0 Monate bzw. 31 Tage
- Standardabweichung: 1,9 Monate bzw. 58 Tage
- Variationskoeffizient: 1,36 bzw. 136 %

Quelle: Eigene Berechnungen auf Basis der internen Rentenantrags- und Erledigungsstatistik, Jahreswerte 2006, Tabelle 131

Übersicht 5b:

Streuungsmaße, Rentenantragslaufzeiten der Hinterbliebenenrenten 2006
(berechnet aus gruppierten Daten):

- Quartilsabstand: 1,2 Monate bzw. 37 Tage
- Mittlere absolute Abweichung: 0,9 Monate bzw. 28 Tage
- Standardabweichung: 1,7 Monate bzw. 52 Tage
- Variationskoeffizient: 1,33 bzw. 133 %

Quelle: Eigene Berechnungen auf Basis der internen Rentenantrags- und Erledigungsstatistik, Jahreswerte 2006, Tabelle 131

Bei der Standardabweichung erhält man zwar auch das gleiche qualitative Ergebnis einer geringeren Streuung der Laufzeiten bei den Hinterbliebenenrenten im Vergleich zu den Versichertenrenten. Wegen der Gewichtung größerer Mittelwertabweichungen sind die angegebenen Werte für die Standardabweichung (58 Tage bei den Versichertenrenten, 52 Tage bei den Hinterbliebenenrenten) wesentlich höher, da für die oberste Klasse ein relativ hoher Mittelwert von 20 Monaten unterstellt wurde. Dies deutet einmal mehr die Problematik klassifizierter Daten an.

Der Variationskoeffizient schließlich zeigt, dass unter den getroffenen Annahmen die Standardabweichung sowohl bei Versicherten- als auch bei Hinterbliebenenrenten im Jahr 2006 jeweils etwa das 1,3- bis 1,4-fache des arithmetischen Mittelwerts betrug.

6. Rangkorrelationen

Bislang wurden nur univariate statistische Methoden behandelt. Die komplexe wirtschaftliche Realität ist aber durch die wechselseitige Beeinflussung einer Vielzahl von Variablen gekennzeichnet. Die Sozialwissenschaften haben dabei die Aufgabe, Zusammenhänge bzw. Erklärungsmuster für die einzelnen wirtschaftlichen Phänomene zu ergründen bzw. zu entwickeln.

Es ist daher bedeutsam, sich auch multivariaten statistischen Zusammenhängen bzw. Verfahren zu widmen. Aus didaktischen Gründen geschieht dies nachfolgend ausschließlich für den bivariaten Fall, d. h. für insgesamt lediglich zwei statistische Merkmale.

Für den Zusammenhang zwischen zwei ordinal-, d. h. rangskalierten Merkmalen lassen sich zwar keine quantitativen Aussagen über die Abstände zwischen den beiden Variablen X und Y treffen. Es können aber immerhin die jeweiligen Rangordnungen zueinander in Beziehung gesetzt und entsprechende Abweichungen festgehalten werden. Man spricht daher in diesem Fall auch von Rangkorrelation.

In Übersicht 6 sind für die einzelnen Merkmalsträger (in diesem Fall: für die einzelnen Versicherungsträger) die einzelnen X- und Y-Platzierungen zueinander in Beziehung gesetzt. Die Variable X steht für den normierten Durchschnittsbestand an Rentenanträgen im Falle von Nichtvertragsrenten (Normalfälle). Um die unterschiedliche Größe der Versicherungsträger zu berücksichtigen wurde der Durchschnittsbestand durch die Anzahl des Verwaltungspersonals beim jeweiligen Versicherungsträger dividiert. Die Variable Y steht in Übersicht 6 für die durchschnittlichen Rentenlaufzeiten bei den Nichtvertrags-Altersrenten (Normalfälle).

Übersicht 6:

Rangkorrelationen zwischen normiertem Durchschnittsbestand und den durchschnittlichen Rentenlaufzeiten bei den Nichtvertragsrenten (Normalfälle) 2006 - Altersrenten:

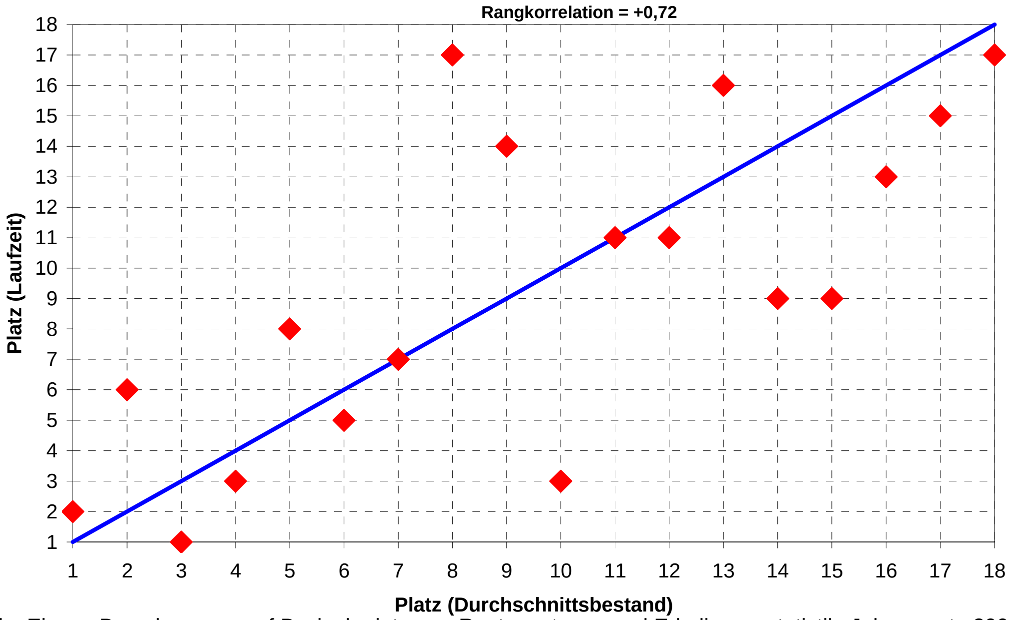
Versicherungsträger	Normierter Durchschnitts-bestand	Bestands-Platz	Laufzeit (Altersrenten)	Altersrenten-Platz
A	0,75	1	7	2
B	1,04	2	14	6
C	1,06	3	5	1
D	1,11	4	12	3
E	1,24	5	16	8
F	1,44	6	13	5
G	1,50	7	15	7
H	1,82	8	31	17
I	1,92	9	21	14
J	1,96	10	12	3
K	2,16	11	18	11
L	2,18	12	18	11
M	2,36	13	23	16
N	2,44	14	17	9
O	2,54	15	17	9
P	2,73	16	19	13
Q	2,95	17	22	15
R	4,00	18	31	17

Quelle: Eigene Berechnungen auf Basis der internen Rentenantrags- und Erledigungsstatistik, Jahreswerte 2006, Tabellen 243 und 900.S1

Wäre die Rangfolge gemäß der X-Variable unter den insgesamt 18 Versicherungsträgern exakt so wie nach der Y-Variable, läge eine vollständig positive Korrelation vor. Ersichtlicherweise ist im Falle von Übersicht 6 bzw. Abbildung 6 die Korrelation zwischen den Durchschnittsbeständen und den durchschnittlichen Laufzeiten bei den Versicherungsträgern nicht exakt Eins, aber doch – wie die einzelnen Beobachtungspunkte zeigen – nicht allzu weit weg von plus Eins. Man kann den konkreten Wert auch mittels des so genannten Rangkorrelationskoeffizienten berechnen, und es ergibt sich in unserem Fall ein Wert von immerhin +0,72.

Abbildung 6:

Rangkorrelationen zwischen normiertem Durchschnittsbestand bei den Nichtvertragsrenten (Normalfälle) vs. den durchschnittlichen Rentenlaufzeiten bei den Altersrenten 2006:



Quelle: Eigene Berechnungen auf Basis der internen Rentenantrags- und Erledigungsstatistik, Jahreswerte 2006, Tabellen 243 und 900.S1

Es ergibt sich das eingangs erwartete Ergebnis einer hoch-positiven Korrelation zwischen Durchschnittsbeständen und Laufzeiten.

7. Schlussbetrachtung

Halten wir abschließend fest: Bei klassifizierten Daten ergeben sich einige Fallstricke; insbesondere sind Ungenauigkeiten in den Ergebnissen nicht zu vermeiden. Bezüglich der in den Standardtabellen der Deutschen Rentenversicherung Bund ausgewiesenen Laufzeitenverteilungen kann m. E. mit einiger Berechtigung kritisiert werden, dass in die unterste Klasse „0-1 Monate“ ein zu großer Teil der Merkmalssumme hineinfällt. Dies versieht in diesem Spezialfall insbesondere die Berechnung von Modus und Median, aber auch die des Quartilsabstands mit einem hohen Ungenauigkeitsgrad. An dieser Stelle wäre eine verfeinerte Klassifizierung zielführend.

Über die Betrachtung univariater Statistiken hinaus erscheinen multivariate Statistiken in Form von Korrelationen bzw. weiterführend: in Form von Regressionen sinnvoll, um Zusammenhänge bzw. Erklärungen aufzuzeigen bzw. zu geben. Derartige Betrachtungen könnten in einer noch weitreichenderen Sicht zu Prognosezwecken genutzt werden. Gleichwohl ist stets sozusagen vor die Statistik der gesunde Menschenverstand zu setzen, um Scheinkorrelationen, d. h. Scheinzusammenhänge zu vermeiden und um wirklich gehaltvolle Aussagen zu erhalten.

Literatur:

Faik, Jürgen: Rentenanträge und ihre Erledigung in beiden Teilen Deutschlands – Eine Zeitverlaufsbetrachtung. In: Deutsche Rentenversicherung, Heft 1-2/1999, S. 75-94.

Faik, Jürgen: Elementare Wirtschaftsstatistik, Berlin 2007.