

WISTAT - ein vielseitiger Datensatz für die sozialwissenschaftliche Methodenlehre

Lettau, Frank

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Lettau, F. (1994). WISTAT - ein vielseitiger Datensatz für die sozialwissenschaftliche Methodenlehre. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 34, 140-146. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-201313>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

WISTAT - Ein vielseitiger Datensatz für die sozialwissenschaftliche Methodenlehre

von Frank Lettau¹

Zusammenfassung

Die soziale Situation von Frauen findet immer stärkere Berücksichtigung in der bevölkerungswissenschaftlichen Forschung. Der vom Statistischen Büro der Vereinten Nationen zusammengestellte und vertriebene Datensatz WISTAT bietet hierzu eine interessante und breitangelegte Sammlung sozialer Indikatoren. Dieser Datensatz bietet zugleich vielfältige Möglichkeiten des Einsatzes im Rahmen von Lehrveranstaltungen der sozialwissenschaftlichen Methodenlehre.

Abstract

The social situation of women is of ever greater concern in population research. The WISTAT -dataset compiled and distributed by the United Nations Statistical Office contains a wide and interesting range of women-related social indicators. At the same time this dataset can be quite usefully employed within seminars for social science research methods and data analysis.

Das Kürzel WISTAT steht für 'Women's Indicators and Statistics'²; es handelt sich hierbei um eine vom Statistischen Büro der Vereinten Nationen³ zusammengestellte Sammlung von nach Nationalstaaten aggregierten Sozialen Indikatoren zur Situation der Frau. Zweck dieser Datensammlung ist die Dokumentation geschlechtsspezifischer sozialer Ungleichheit auf international vergleichender Basis (Untersuchungseinheiten sind N=178 Nationen). Der Datenumfang beträgt ca. 7.5 Mb und verteilt sich auf über 90 Spreadsheet-Dateien.⁴ Die thematische Ausrichtung dieses Datensatzes ist vor dem Hintergrund der von den

1 Dipl.-Soz. Dr. **Frank Lettau** ist Akad. Rat im Schwerpunkt 'Methoden der Politikwissenschaft' des Fachbereichs Politische Wissenschaft der FU Berlin, Innestr. 22, 14195 Berlin, Tel 030/838-4959.

2 Women's Indicators and Statistics Spreadsheet Database for Microcomputers (Version 2, May 1991).

3 United Nations, Dept. of Economic and Social Development, Director of the Statistical Division, New York-10017, U.S.A. Der Preis beträgt ca. 200 S.

4 Die Spreadsheets sind für das Tabellenkalkulationsprogramm LOTUS 1-2-3 geschrieben, sämtliche Dateien sind damit sog. '.wkl-files' (also mit der Namensendung '.wkl' versehen). Die Daten werden kontinuierlich überarbeitet und aktualisiert; eine 3. Version ist für den Herbst 1993 angekündigt.

Vereinten Nationen vorgenommenen Schwerpunktsetzung auf Verbesserung der Situation von Frauen zu sehen ⁵.

Neben der allgemeinen Aktualität der Thematik sind es zwei Aspekte, die diese Datensammlung interessant machen: Zum einen ist die Thematik in einen sowohl von der Anzahl als auch vom Inhalt her breiten Kranz von Variablen eingebettet, der sich über folgende neun Inhaltsbereiche erstreckt:

- 1) Population composition and distribution
- 2) Learning and educational services
- 3) Economic activity
- 4) Households, marital status and fertility
- 5) Housing conditions and human settlements
- 6) Health and health services
- 7) Public affairs and political participation
- 8) Crime and criminal justice
- 9) National product and expenditure,

wobei es (natürlich) ein durchgängiges Merkmal ist, daß sämtliche Informationen so weit wie möglich nach Geschlecht disaggregiert sind. Die große inhaltliche Breite des Datenangebots ermöglicht es, unterschiedlichste und sonst oft nur separat erhältliche Variablenbereiche zusammenzustellen und entsprechend komplexe multivariate Hypothesen zu überprüfen.

Interessant ist diese Datensammlung zum zweiten auch deswegen, weil sie mehrfache Meßzeitpunkte berücksichtigt. Am ausführlichsten von allen Inhaltsbereichen sind die Bevölkerungsdaten ⁶ dokumentiert; die Meßzeitpunkte sind in diesem Fall:

1950 1960 1970 1975 1980 1985 1990 1995 2000 2025,

wobei die letzten drei Zeitpunkte keine Schätzungen, sondern Projektionen sind. Andere Variablen weisen hiervon verschiedene Meßzeitpunkte oder Meßzeitbereiche auf, wie z.B.:

1970 1980 latest available year,

oder

1965-70 70-75 75-80 80-85 85-90 90-95 95-00 2020-25

oder

year around 1970 year around 1980 latest available year,

5 Vgl. Report of the World Conference of the International Women's Year, Mexico City, 1975 (U.N. publication E.76.IV.1); United Nations Decade for Women 1975-1985, hierzu u.a. Report of the World Conference of the United Nations Decade for Women, Copenhagen, 1980 (U.N. publication E.80.IV.3).

6 Inhaltsbereich 1.1: Estimated and projected total, urban and rural population and population by age and sex.

wobei insbesondere das erstgenannte Format mit drei Meßzeitpunkten am häufigsten erscheint⁷. Erkennbar ist auf jeden Fall, daß je nach erhobener Variable unterschiedliche Meßzeitpunkte zu berücksichtigen sind; dies gilt auch für Variablen eines gemeinsamen Inhaltsbereichs (vgl. die o.g. Inhaltsbereiche 1-9). Die Tatsache der Existenz unterschiedlicher Meßzeitpunkte ist nicht erstaunlich, wenn man bedenkt, daß die einzelnen Variablen dieser Datensammlung aus ganz unterschiedlichen Länderkontexten stammen, von unterschiedlichen Unterorganisationen der Vereinten Nationen (z.B. WHO, ILO, UNESCO, UNFPA) bereitgestellt werden oder aus unterschiedlichen Datenbasen (z.B. National Accounts Database, Yearbook of Labour Statistics, Demographic Yearbook) entnommen werden⁸. Der große Vorteil besteht darin, überhaupt über zeitindizierte Variablen verfügen zu können, was gerade für die Überprüfung von Kausalhypothesen wichtig ist. Mit Hilfe zeitindizierter Variablen lassen sich auch dynamische Aspekte z.B. der Bevölkerungsentwicklung, der Urbanisierung, der Fertilität, des Bildungsgrades der Bevölkerung usw. leicht mit Hilfe einfachster Datenmodifikationen (Bildung von Differenzvariablen) einfangen.

Über die vorgenannten Nutzungsmöglichkeiten für die Methodenlehre und die Forschung hinaus gibt es auch didaktische Aspekte, die den WISTAT-Datensatz als gut geeignet für den Einsatz im Rahmen der sozialwissenschaftlichen Methodenlehre - hierbei insbesondere für datenorientierte Statistik-Lehrveranstaltungen des Hauptstudiums - erscheinen lassen: Die Aktualität und zweifellos gegebene Attraktivität der frauenbezogenen Thematik in Verbindung mit der Möglichkeit zu international vergleichender Betrachtung wirken sich besonders motivationsfördernd aus. Unbedingt erforderlich ist allerdings eine kurze allgemeine Einführung in die komplexe und vor allem etwas umfangreiche Struktur dieses Datensatzes sowie eine Einführung in die (rudimentäre) Benutzung von Lotus 1-2-3⁹, denn das Arbeiten mit WISTAT erfordert einen sicheren und wohlüberlegten Umgang mit Daten und Dateien bzw. bietet Gelegenheit, diesen zu erlernen. Da die einzelnen Variablen für z.B. einen SPSS-Lauf¹⁰ auf viele einzelne WISTAT-Dateien verteilt sind, müssen zunächst die Variablen und die sie enthaltenden Dateien bestimmt, dann mittels 'translate' vom Format .wkl nach SPSS portiert, als Systemdateien zwischengespeichert und später mit Hilfe von 'join match' zu einer Datei zusammengefügt werden.

Nach Vermittlung dieser eher arbeitstechnischen Aspekte kann relativ problemlos zur Vermittlung des eigentlichen Stoffes übergegangen werden, wobei sich im Idealfall in einer

7 Beim demnächst erscheinenden WISTAT-Update wird sich vermutlich in vielen Fällen die 3er-Stufung 70-80-latest um eine Stufe auf 70-80-90-latest erweitern.

8 Ein viel wichtigerer Aspekt bei Entscheidungen über Einbeziehung oder Ausschluß bestimmter Meßzeitpunkte dürfte die Maximierung der Datenausbeute durch Minimierung der zu erwartenden 'missings' sein.

9 Das Problem hierbei ist, daß für die Zusammenstellung der interessierenden Variablen die jeweiligen WISTAT-Dateien zuvor gesichtet werden müssen. Die Einführung in 1-2-3 beschränkt sich auf diesen Aspekt. Für die Betrachtung von Dateien kann alternativ das Programm XTREE (Version XTGOLD) verwendet werden, welches das 1-2-3-Format automatisch konvertiert.

10 Die nachfolgenden Darstellungen beziehen sich auf die DOS-Version von SPSS.

Kombination aus Vorlesung (z.B. über multivariate Analyseverfahren) und Übung bzw. einem entsprechend geteilten Seminar erfahrungsgemäß leicht studentische Arbeitsgruppen bilden lassen. Für den zu erbringenden Leistungsnachweis konzentrieren sich die Studenten auf selbst ausgesuchte Fragestellungen bzw. bestimmte vorgegebene Aufgabenstellungen. Zu diesem Zweck können eigene 'maßgeschneiderte' Datensätze zusammengestellt und die in der Vorlesung erlernten Verfahren zur Anwendung gebracht werden.

Was den Variablentyp anbetrifft, so weisen die WISTAT-Variablen, wie dies bei Aggregatdaten häufig der Fall ist, metrisches Meßniveau auf. Die allermeisten Variablen sind entweder

- a) absolute Häufigkeiten (z.B. 'Nr. of female teachers at first level 1980'), oder
 - b) Durchschnittswerte (z.B. 'life expectancy at birth, males, 1965-70') oder
 - c) Anteilswerte (z.B. 'percentage of urban population with access to safe water, 1975').
- Zusätzlich gibt es, wie bereits angedeutet, vielfältige Möglichkeiten zur Bildung neuer Variablen; hierzu einige Illustrationen:

- *'percentage rural population 1990 - perc.r.p. 1970':*
ein möglicher Indikator für das Tempo von Urbanisierungstendenzen;
- *'age at 1st marriage (male) - age at 1st mar. (fem)':*
ein möglicher Indikator für geschlechtsspezifische soziale Ungleichheit (Divergenz im Mitspracherecht und in der Entscheidungskompetenz in der Ehe);
- *'Region' (Zusammenfassung der 178 Länder nach Wirtschaftsregionen oder geographischen Regionen):* Möglichkeiten zur Bildung verschiedener Gruppierungsvariablen, die entweder als erklärende oder auszupartialisierende Drittvariablen Verwendung finden können;
- *Indexbildungen:* z.B. Betrachtung und Auswertung der unterschiedlichen Möglichkeiten zur Bildung eines Index für die relative Benachteiligung der Frau unter Verwendung jeweils unterschiedlicher Kombinationen einzelner Variablen als multiple Indikatoren ¹¹.

Neben der offensichtlichen Aktualität einer frauenbezogenen Thematik gibt es ein weiteres und tieferliegendes Motiv für die inhaltlich breite Zusammenstellung dieses Datenpakets. Es besteht in der Erkenntnis der demographischen Forschung, daß soziale Aspekte wichtige sog. 'intermediäre Variablen' für die Vorhersage von Bevölkerungsprozessen ¹² sind. Dies gilt in besonderem Maße für die Lebensbedingungen von Frauen, also z.B. für Variablen wie primäre und sekundäre Schulbildung, Berufsausbildung oder (relatives) Heiratsalter.

¹¹ In diesem Zusammenhang lassen sich dann die verschiedenen Fragen in Zusammenhang mit 'multiple item measurement', mit der Definition eines Index, mit Fragen der Gewichtung und Standardisierung von Indikatoren und mit der Reliabilitäts- und Meßfehlerproblematik herausarbeiten.

¹² Migration, Fertilität und Mortalität sind die Hauptparameter von Bevölkerungsprozessen und zugleich die 'großen' abhängigen Variablen der demographischen Forschung. Insbesondere im Hinblick auf die für Bevölkerungsprognosen wichtige Variable Fertilität gilt, daß sie als eine in einem Beziehungsgeflecht von verschiedenen sozialen Faktoren, Gesundheitssystem-Parametern und ökonomischen Rahmenbedingungen stehende abhängige Variable gesehen werden muß.

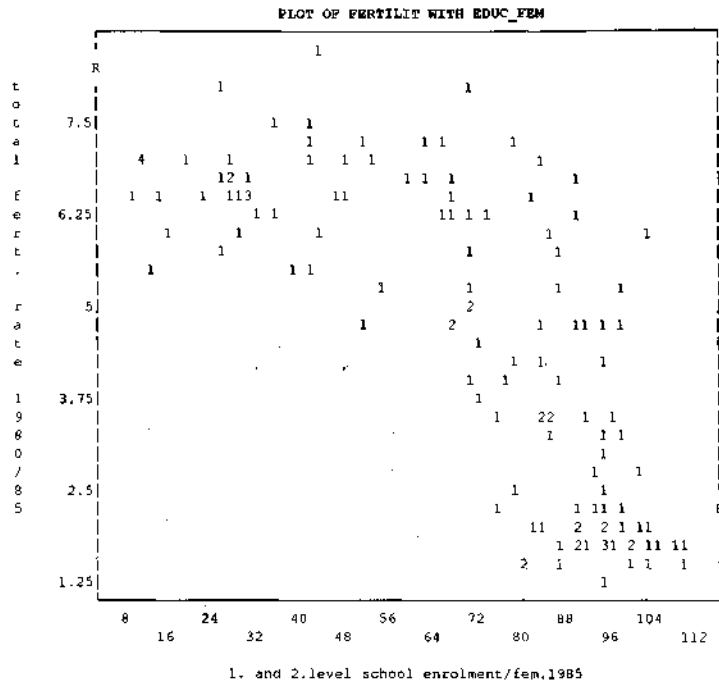
Als kleinstmögliche Demonstration für die Handhabung von WISTAT sei nachfolgend der Zusammenhang der Variablen Fruchtbarkeit (operationalisiert durch **'Total Fertility Rate 1980/85'** in Spalte **h** der WISTAT-Datei 43TFR.wkl) mit den beiden möglichen Prädiktoren **Vorhandensein von Schulbildung** (operationalisiert durch **'Female 1. and 2. Level School Enrolment'** in Spalte **z** der WISTAT-Datei 26ENRAT.wkl) und **Mitspracherecht in der Ehe** (operationalisiert durch die Differenz beim **'Age at First Marriage'** in den Spalten **m** und **o** der WISTAT-Datei 42SMAM.wkl) betrachtet. Die für die Zusammenstellung der Daten nötigen SPSS-Statements lauten:

```

set screen off.
* ! Einlesen einer Gewichtungsvariablen.
* ! Die zweite Spalte (b) jeder Wistat-Datei enthaelt
einen Laender-Code als Sortier- und Merge-Variable.
tra from '11pop85.wkl'/keep b g/range a30..ai207.
com bs_total=g.
var lab bs_total 'total pop.1985/both sexes'.
sort b(a).
sav/outfil 'file.pop'/keep b bs_total.
* ! Einlesen des Prädiktors Schulbildung.
tra from '26enrat.wkl'/keep b z/range a30..z207.
com educ_fem=z.
var lab educ_fem '1. and 2.level school enrolment/fem.1985'.
sort b(a).
sav/outfil 'file.ed'/keep b educ_fem.
* ! Einlesen des Prädiktors Differenz im Heiratsalter.
tra from '42smam.wkl'/keep b m o/range a30..o207.
com diff_age=(m-o).
var lab diff_age 'age-diff. at 1.marriage male/fem. 1980'.
sort b(a).
sav/outfil 'file.mar'/keep b diff_age.
* ! Einlesen der abh. Variablen Fruchtbarkeit.
tra from '43tfr.wkl'/keep b h/range a30..i207.
com fertilit=h.
var lab fertilit 'total fert. rate 1980/85'.
sort b(a).
sav/outfil 'file.fer'/keep b fertilit.
get fil 'country.sys'.
* ! Vorgefertigte 'Kopf'-Datei mit spez. Laenderinfor-
mationen.
join match/fil '*' /fil 'file.pop' /fil 'file.ed' /fil
'file.mar' /fil 'file.fer' /by b.
* ! Struktur von 'newfile': vgl. JOIN-statement.
sav/outfil 'newfile.sys'.

```

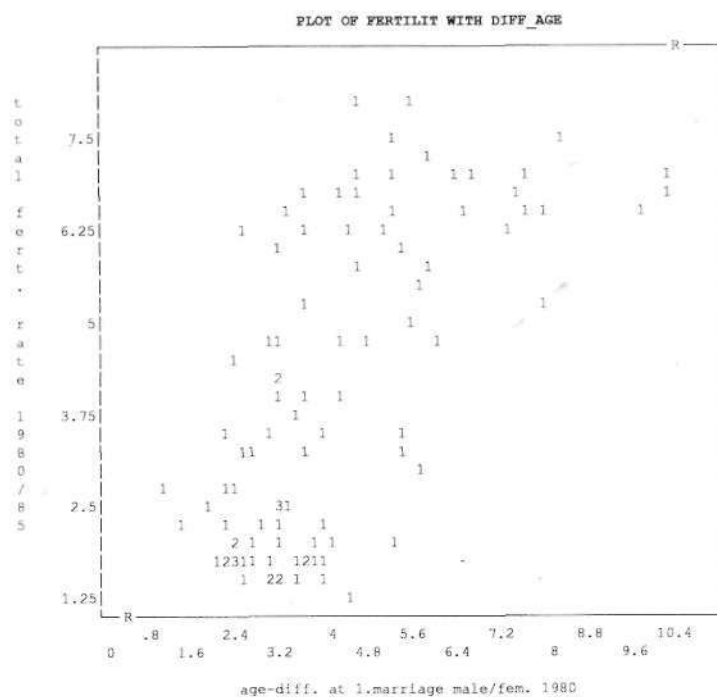
Ein einfacher Plot genügt, um den Zusammenhang zwischen Schulbildung (Schulbe-
suchsquote) und Fruchtbarkeitsrate zu veranschaulichen:



130 Fälle (ungewichtet)
 Korrelation: -0.7497; R^2 : 0.56
 Bei Gewichtung nach Ländergrößen (4.245.383.000 Fälle):
 Korrelation: -0.8017; R^2 : 0.64

Auch für den zweiten möglichen Prädiktor Mitspracherecht bzw. relative Unmündigkeit in der Ehe ergibt sich ein ähnlich starker Zusammenhang mit der Fruchtbarkeitsrate (siehe dazu die 2. Abbildung)

Auf die Einführung weiterer Variablen, die Inangriffnahme komplexerer Kausalmodelle und andere weitergehende Darstellungen muß im Rahmen dieses Kurzberichts verzichtet werden. Deutlich werden sollte auch nur, daß dieser Datensatz, der Teil des 'data dissemination program' der Vereinten Nationen ist, sowohl für die Forschung wie für die Lehre einiges bietet.



98 Fälle (ungewichtet)

Korrelation: 0.6603; R^2 : 0.44

Bei Gewichtung nach Ländergrößen (4.118.734.000 Fälle):

Korrelation: 0.7636; R^2 : 0.58

NB: Bei der ersten der beiden Graphiken mag überraschen, daß der Indikator 'percentage school enrolment' in einigen Fällen den Wert 100 übersteigt. Es manifestiert sich hier eine von dieser Art von Indikatoren bekannte Tendenz zur Überschätzung gegebener Verhältnisse. Dieses sog. 'Overreporting' kann dadurch zustande kommen, daß je nach Art des Bildungssystems und je nach Auswertung der nationalen 'school records' Mehrfachzählungen (z.B. Parallelfälle von Teilnahme an allgemeinbildender Schule *und* berufspraktischer Ausbildung *und* Konfessionsschule) möglich werden. Die auffälligsten Werte weisen in diesem Fall Finnland (107%), die Niederlande (109%) und Spanien (109%) auf, was natürlich nicht bedeutet, daß dieser Indikator bei Ländern mit geringeren 'enrolment ratios' nicht auch zur Überschätzung neigen kann.