





12. STATISTIKTAGE BAMBERG | FÜRTH 2024

Zensus – jetzt und in Zukunft

Katrin Kräck, M.Sc. und Nina Storfinger, Dipl.Soz.Univ.

In diesem Jahr fanden bereits zum zwölften Mal die StatistikTage Bamberg|Fürth statt. Die Veranstaltung wird im Rahmen des Statistik Netzwerk Bayern vom Bayerischen Landesamt für Statistik gemeinsam mit dem Lehrstuhl für Statistik und Ökonometrie der Otto-Friedrich-Universität Bamberg organisiert. Sie dient der Vernetzung von amtlicher Statistik und Wissenschaft sowie weiterer Nutzergruppen amtlicher Daten. Im Juli 2012 fand die Premiere der Tagungsreihe statt, Thema damals so aktuell wie heute: „Methoden und Potenziale des Zensus“. Während vor zwölf Jahren die Methoden und Erkenntnisse der Zensusrunde 2011 erläutert und diskutiert wurden, stand in diesem Jahr selbstverständlich der aktuelle Zensus von 2022 im Fokus. So lautete das Motto der StatistikTage 2024 „Zensus – jetzt und in Zukunft“. Die mit hochrangigen Rednerinnen und Rednern besetzte Veranstaltung stieß schon im Vorhinein auf großes Interesse. Mehr als 130 Personen hatten ihre Teilnahme angekündigt.



„Der persönliche Kontakt und der Austausch nehmen bei den StatistikTagen schon immer eine große Rolle ein. Auch die diesjährige Veranstaltung bietet viele Möglichkeiten, mit den Expertinnen und Experten der amtlichen Statistik ins Gespräch zu kommen. Ich kann gerade die Nachwuchswissenschaftlerinnen und -wissenschaftler unter Ihnen nur ermutigen: Nutzen Sie diese Gelegenheit! Die StatistikTage sind heuer wieder ausnehmend hochkarätig besetzt – und auch die Netzwerke, die Sie mit Ihren Kolleginnen und Kollegen knüpfen, sind sehr bereichernd und für Ihre weitere Karriere außerordentlich wertvoll.“

*Prof. Dr. Kai Fischbach,
Präsident der Otto-Friedrich-Universität Bamberg*

Eröffnet wurde die Tagung am Nachmittag des 11. Juli zunächst durch den Präsidenten der Otto-Friedrich-Universität Bamberg, Herrn Prof. Dr. Fischbach, der in seiner Rede betonte, dass der Zensus eines der Themen der amtlichen Statistik sei, das die Gemüter bewegte. Dabei umspanne das Themenfeld viele Facetten und bewege gleichermaßen Fachwelt und Öffentlichkeit, wobei die StatistikTage wie in jedem Jahr den Rahmen bieten würden, sich über aktuelle Entwicklungen, praktische Anwendungsfälle und Perspektiven zu beraten. Dass die Veranstaltung nun schon zum zwölften Mal durchgeführt wird, zeige welchen Stellenwert sie für den Austausch und die Vernetzung von Forschenden wie auch Nutzerinnen und Nutzern der verschiedenen Institutionen weit

über die Region hinaus hat. Anschließend begrüßte auch der Präsident des Bayerischen Landesamtes für Statistik, Herr Dr. Gößl, das Publikum. Er hob die Aktualität des Veranstaltungsthemas heraus, da erst weniger als drei Wochen vor Beginn der StatistikTage die ersten Ergebnisse des Zensus 2022 veröffentlicht wurden. Außerdem stelle der Zensus 2022, das Großprojekt der amtlichen Statistik, eine wichtige Datengrundlage für Deutschland dar. Die Ergebnisse des Zensus 2022 geben Aufschluss darüber, wie viele Menschen in Deutschland leben und wie sie wohnen und arbeiten. Sie sind maßgebend für zahlreiche finanz- und gesellschaftspolitische Entscheidungen. Herr Dr. Gößl und Herr Prof. Fischbach sprachen außerdem ihren Dank an die Organisatoren der beiden Institutionen aus.



„Eine immer mobilere Bevölkerung an ihrem Hauptwohnsitz abzubilden, ist eine Schlüsselfrage für die Register und einen registerbasierten Zensus. Der Stichtag des Zensus im Mai 2022 fiel in die Schlussphase der Pandemie und in die ersten Monate des russischen Angriffskriegs in der Ukraine, also eine Phase nochmals erhöhter Mobilität. Daher ist es wichtig, nicht nur auf Deutschland beschränkt die Modelle oder die Strategie des Zensus, die Methodik und Durchführung genau zu betrachten und zu diskutieren – und zwar im durchaus kritischen Gespräch zwischen Wissenschaft und amtlicher Statistik.“

*Dr. Thomas Gößl,
Präsident des Landesamts für Statistik*

Nachdem im letzten Jahr, die StatistikTage zum Thema „Energie und Preise“ erstmals in den Räumlichkeiten des Bayerischen Landesamts für Statistik stattfanden, traf man sich in diesem Jahr wieder in Bamberg. Die AULA der Universität Bamberg, gelegen in einer ehemaligen Dominikanerkirche aus dem Jahr 1401, bot abermals ein imposantes Ambiente für die gut besuchte Tagung. Neben den Wandmalereien lockte auch der neu gestaltete Kreuzgang im Innenhof der AULA einige Teilnehmende in den Pausen zu einer kleinen Erkundungstour. Weitere Informationen zur Sanierung der AULA durch die Universität Bamberg können Sie hier nachlesen:

www.uni-bamberg.de/fileadmin/kommunikation/045-UNI-PUBLIKATIONEN/Wissen-zum-Mitnehmen/Aula-zusammen.pdf



Im Anschluss an die Begrüßungsworte der beiden Präsidenten sowie der beiden Organisatoren und Moderatoren der Veranstaltung, Herr Prof. Dr. Schmid, Inhaber des Lehrstuhls für Statistik und Ökonometrie der Otto-Friedrich-Universität Bamberg, und Herr Prof. Dr. Fürnrohr, Leiter der Abteilung Bevölkerung, Haushalte, Zensus, Finanzen, Schulen des Bayerischen Landesamt für Statistik und Leiter des Projekts Zensus 2022 in Bayern, startete der erste inhaltliche Block zum Thema „Zensus international und in Deutschland“ mit einer hochrangigen Keynote.

Herr Prof. Dr. Pfeffermann von der Hebrew University of Jerusalem und der University of Southampton leitete bis vor zwei Jahren neun Jahre lang als Generaldirektor das israelische Zentralbüro für Statistik und bereicherte mit seiner umfassenden Expertise die Veranstaltung. Er gab den Teilnehmenden in seiner Keynote einzigartige Einblicke in die Durchführung der Erhebung sowie die Nutzung von Zensusdaten in Israel.

Herr Prof. Dr. Pfeffermann berichtete über den letzten Zensus in Israel, der in den Monaten von April bis Dezember 2022 vom Zentralbüro für Statistik durchgeführt wurde. Der Methode des deutschen Zensus ähnelnd, wurde auch in Israel auf eine Kombination von Befragung und zeitgleichem Rückgriff auf administrative Daten gesetzt. Er erläuterte, welchen Problemen sie sich, insbesondere in Bezug auf die Nonresponse-Rate in bestimmten Bevölkerungsgruppen, gegenübersehen und auf welche statistischen Methoden zurückgegriffen wurde um dem Problem zu begegnen. Weitere Details finden sich auch auf den Folien von Herrn Prof. Dr. Pfeffermann ab Seite 33.

Anschließend folgte in diesem ersten inhaltlichen Block ein Vortrag zum Zensus 2022 in Deutschland. Herr Prof. Dr. Fürnrohr und Frau Dr. Woltering (beide Bayerisches Landesamt für Statistik) sprachen über Methode und Durchführung in Bayern. Nach einer Einführung zum Zensusmodell gaben sie einen Einblick in den Ablauf der Erhebungen in Bayern sowie einen Ausblick auf die noch folgenden Veröffentlichungen zu den Ergebnissen des Zensus 2022.

Prof. Dr. Pfeffermann



Danny Pfeffermann is Professor of Social Statistics at the University of Southampton, UK and Professor Emeritus of Statistics at the Department of Statistics and Data Science at the Hebrew University of Jerusalem, Israel. Until

two years ago, he served for 9 years as the National Statistician and Director General of Israel's Central Bureau of Statistics. His main research areas are analytic inference from complex sample analytic inference from complex sample surveys, seasonal adjustment and trend estimation, small area estimation, inference under informative sampling and non-response, mode effects and proxy surveys and more recently, estimation of population parameters from nonprobability samples. Professor Pfeffermann published more than 80 articles in leading statistical journals and co-edited the two-volume handbook on Sample Surveys. He was President of the Israel Statistical Society and of the International Association of Survey Statisticians (IASS) and is Fellow of the American Statistical Association (ASA), the International Statistical Institute (ISI) and the Institute of Mathematical Statistics (IMS). He received several prestigious international awards.



Vortragsfolien von Herrn Prof. Dr. Pfeffermann:

The 2022 Census in Israel: Strategy, Estimation and Evaluation

Danny Pfeffermann

Zensus – jetzt und in Zukunft- July 2024

Introduction

Israel had its previous census in **2008!!** The following, last census, started on **April 3, 2022** and ended on **31 December 2022**.

- ❖ It took about **7** years for planning, preparing and testing the new census.

Its main objectives were to provide rich information for small geographical areas (about 3,500 persons on average), in terms of number of residents by age and gender, transportation, work possibilities, education, wages not from work, etc.

The census information is used for allocation of Government funds via the socio-economic index, calculated based on it, for research, and it forms the sampling frame for many of our surveys.

2

Main goal- national population counts

- ❖ Population counts in small **statistical (geographical) areas**.

Israel is divided into **~3,500 statistical areas**, with an average number of almost **3,000** residents per area.

- ❖ Israel has a fairly accurate Central Population Register (**CPR**) of individuals, but there are on average about **15% inaccuracies** in the **CPR** regarding the correct living address.

The main reasons for the inaccuracies are due to people moving in or out of their reported area, but reporting late their change of address, while others who have an address of interest (tax benefits, favorite schools for children, parking facilities, etc.) do not report their change of address as long as the interest persists.

3

Plan of action

- 1- Draw a representative probability sample of **700K** residents from the population register (see below). Covers **7%** of the total population.
- 2- Each sampled unit is asked to provide information about all members of the household, so actual sample size is **~24%**.
- 3- Three possible modes of response: **Internet** → **Telephone** → **Personal (Face to face) interview**.
 - ❖ Small presents for people responding via the Internet.
- 4- Questionnaire designed electronically, such that information known from administrative files, (e.g., education), is added automatically, thus reducing **response burden**.
- 5- Enormous campaign in all communication systems before and during the census to encourage response to the questionnaire.

4

The Central Population Register (CPR) in Israel

- ❖ Established in **1948**
- ❖ Unique **ID** for each person.
- ❖ Extensive use by the public and private sectors.
- ❖ Includes: ID, name, address, date of birth, gender, family ties, marital status, religion, nationality, immigration date, country of birth, country of father's birth, border control information (departures and arrivals to Israel).
- ❖ New definitions of people living abroad for inclusion in the **CPR**.

5

Sampling design

We used an algorithm that constructs a **synthetic household frame** from the **CPR**. The algorithm integrates relevant data of individuals from available administrative data sources to distribute the **CPR** into **HHs**, whose members are believed to live in the same address.

The sample was drawn from the **synthetic frame (SF)** by use of stratified, two-stage cluster sampling, with many strata defined by **statistical area**, **size of administrative family (AF)**, persons classified in **SF** as living in the same address), **marital status** and **age of mother**. **Clusters** are the **AF**, **ultimate sampling units** are persons in them.

- ❖ One person sampled from each **AF**. (If the person sampled is a child, a parent responds).

6

Notation and estimation

N - **known** population size (**CPR**) on census day, (**3 April 2022**).

N_i - **unknown** number of residents in area **i** (**~3500** areas).

$P_i = N_i / N$ - true proportion of residents in area **i** .

z_{ij} = # of persons reported by sample unit **(i, j)** belonging to **AF**.

Estimation I - Direct estimators

$\hat{P}_i = \sum_{(i,j) \in S} I_{j \in i} w_{ij} z_{ij} / \sum_{(i,j) \in S} w_{ij}$; $I_{j \in i} = 1$ if person j resides in area i on census day, $w_{ij} = (1 / \pi_{ij})$, $\pi_{ij} = \Pr[(i, j) \in S]$.

Assuming **complete response**, the direct estimator of N_i is then,

$$\hat{N}_i = N \hat{P}_i.$$

7

Main worry: Nonresponse

We expected to encounter nonresponse in the census sample, mostly in the **Bedouin** and the **Ultra Orthodox Jewish** populations.

- ❖ A large portion of the Bedouins reside in scattered locations and/or they move sporadically from one location to another. Hard to locate their addresses and so only **face to face interviews**.
- ❖ Some of the Ultra-Orthodox Jews believe that the bible prohibits participation in a census. (**Not true**, several censuses reported in the bible).

8

Population division by ethnicity, Census 2022

Population Groups	Numbers (Thousands)
Total Population	9,601.7
<u>Jewish population:</u>	7,385.9
Ultra-Orthodox	1,127.0
Jewish population Exl. Ultra-Orthodox	6,258.9
<u>Arab population:</u>	2,003.2
Muslims	1713.1
Christians	141.4
Druze	148.7
Bedouin at the Negev desert	305.4
East Jerusalem	380.0
<u>Foreigners:</u>	212.6

9

Nonresponse (cont.)

Nonresponse occurred in all parts of the population, with differential response rates, possibly depending on the target census variable → not missing at random (**NMAR** → probability to respond depends on the target variable of inference and known covariates).

❖ Not accounting for **NMAR** nonresponse may result in **large bias**.

To deal with this problem, **I** proposed to apply **Sverchkov and Pfeffermann (SP 2018)** methodology, which permits estimating the **response probabilities** $\hat{Pr}(i, j)$ as a function of the target outcome variable and known covariates.

New response adjusted weights: $a_{ij} = [\pi_{ij} \times \hat{Pr}(i, j)]^{-1}$.

10

Estimation II- Combining the sample estimates with CPR

The direct estimator, $\hat{N}_{i,m} = N \times \sum_{(i,j) \in S} I_{j \in i} a_{ij} z_{ij} / \sum_{(i,j) \in S} a_{ij}$ with the response-adjusted weights a_{ij} does not use the area counts K_i in the **population register**.

❖ Although not fully accurate, the population register provides valuable information about the true counts, which should be taken into account. We considered therefore the use of **composite estimators** of the form,

$$\hat{N}_i^{Com} = A_i \hat{N}_{i,m} + (1 - A_i) K_i$$

❖ $\hat{N}_{i,m}$ has a **variance**, but (approximately) **no bias**. K_i has a **bias** (measurement error), but **no variance**.

11

Computation of A_i

$$\hat{N}_i^{Com} = A_i \hat{N}_{i,m} + (1 - A_i) K_i$$

$$MSE(\hat{N}_i^{com}) = E(\hat{N}_i^{com} - N_i)^2 = A_i^2 Var_D(\hat{N}_{i,m}) + (1 - A_i)^2 (K_i - N_i)^2.$$

The A_i , minimizing the **MSE** is, $A_{i,opt} = \frac{(K_i - N_i)^2}{(K_i - N_i)^2 + Var_D(\hat{N}_{i,m})}$.

In practice we don't know N_i , hence estimate $A_{i,opt}$ as,

$$\hat{A}_i = \frac{(K_i - \hat{N}_{i,m})^2}{(K_i - \hat{N}_{i,m})^2 + \hat{Var}_D(\hat{N}_{i,m})} \cdot \hat{Var}_D(\hat{N}_{i,m}) \text{ calculated by bootstrap.}$$

Estimation III- Calibration:

Calibrate the estimators \hat{N}_i^{Com} to **known population totals** from the **Register** such as age, gender, education categories etc. (See below.)

12

Example for nonresponse adjustment- data from 2008 Census

Target variable: $y_{ij} = 1$ if person j registered as living in area i is **married** and $y_{ij} = 0$ otherwise. **Noninformative** sampling in each area.

Let x_{ij} denote the age of person (i, j) and define $x_{1ij} = 1$, if $x_{ij} > 25$, $x_{1ij} = 0$ otherwise, $x_{2ij} = 1$ if $x_{ij} > 40$, $x_{2ij} = 0$ otherwise.

The following logistic models have been assumed to deal with **NMAR** nonresponse, for the outcomes **observed** for the responding units, and for the response probabilities. Data for $x_{ij} \leq 16$ have been removed.

$$Pr(y_{ij} = 1) = \text{logit}(\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + u_i); u_i \sim N(0, \sigma_u^2)$$

$$Pr(R_{ij} = 1) = \text{logit}(\gamma_0 + \gamma_1 x_{1ij} + \gamma_2 x_{2ij} + \gamma_3 y_{ij}).$$

13

Example (cont.)

Denote by R_i the subsample of responding units in Area i .

We considered the following estimators for the number of married people, M_i , among the K_i people registered as living in the area:

$$\hat{M}_{i,MCAR} = K_i \sum_{j \in R_i} y_{ij} / \sum_{j \in R_i} 1;$$

$$\hat{M}_{i,MAR} = \sum_{j \in R_i} y_{ij} + \sum_{k \notin R_i}^K \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1ik} + \hat{\beta}_2 x_{2ik} + \hat{u}_i)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1ik} + \hat{\beta}_2 x_{2ik} + \hat{u}_i)};$$

$$\hat{M}_{i,HB} = K_i \frac{\sum_{j \in R_i} y_{ij} [\hat{P}_r(R_{ij} = 1)]^{-1}}{\sum_{j \in R_i} [\hat{P}_r(R_{ij} = 1)]^{-1}} \rightarrow \text{The Hajek-Brewer estimator with the estimated response probabilities.}$$

$$\hat{M}_{i,SP} = \sum_{j \in R_i} y_{ij} + \sum_{k \notin R_i}^{\tilde{K}} \hat{R}a_{ik} \rightarrow \text{estimator developed by SP (2018).}$$

14

Results

In what follows I consider separately areas of size $K_i \leq 100$ and areas of size $100 < K_i \leq 200$, the smallest statistical areas.

Denote $E_i = (M_i - \hat{M}_i)$, where \hat{M}_i represents any of the above estimators.

The following table contains the **Bias** and **RMSE** over all areas **A** in each

group: **Bias** = $\sum_{i=1}^A E_i / A$; **RMSE** = $\left(\sum_{i=1}^A E_i^2 / A \right)^{0.5}$.

Estimator	Areas $100 < K_i \leq 200$		Areas $K_i \leq 100$	
	Bias	RMSE	Bias	RMSE
$\hat{M}_{i,MCAR}$	2.95	4.45	4.50	5.86
$\hat{M}_{i,MAR}$	2.25	3.26	3.32	4.52
$\hat{M}_{i,HB}$	0.88	3.28	0.68	4.02
$\hat{M}_{i,SP}$	0.79	2.18	0.67	3.05

15

What has been done in the 2022 Census

After a very complex editing process, the files for estimation consisted of **2,034,300** individuals and **527,400** households.

The **ICBS** used a very complex procedure for computing appropriate sampling weights for Individuals → Households → Statistical areas and Settlements (living places). Each Settlement comprises **1+** statistical areas.

A (loose) description of the main stages

In the **1st** stage, a **logistic model** has been fitted for estimating the response probabilities as a function of very large number of demographic and other variables (no area or settlement residence information).

16

What has been done in the 2022 Census (cont.)

In the **2nd** stage, sampling weights have been computed for each individual and household by accounting for the estimated response probabilities.

In the **3rd** stage, sampling weights have been computed for statistical areas and settlements within post-strata. These sampling weights have been calibrated using **731** calibration constraints by application of the **GREG** estimator. Define the final estimators as \hat{N}_i^{cal} .

In the **4th** stage, the design-based variances and hence **CV's** of the estimators have been estimated by use of linearization. In general, when **CV>30%**, the estimator is not published.

When **15%≤CV≤30%**, the estimator is published with a flag.

17

Census evaluation

Following, are several **cost free “bureau evaluations”**, which were applied. Due to budget limitations, the **ICBS** was not able to perform proper **post enumeration evaluations**.

1- Compare the **Register values K_i** with the calibrated sample estimates \hat{N}_i^{cal} . Although the K_i 's are inaccurate, it was expected that confidence intervals around \hat{N}_i^{cal} will cover most of them.

- ❖ Areas with large discrepancies between K_i and \hat{N}_i^{cal} required further consideration.
- ❖ This evaluation also allows testing the accuracy of the Register, towards the use of **administrative censuses** in the future.

18

Census evaluation (cont.)

2. Comparison to other administrative sources:

- ❖ Municipal TAX payments;
- ❖ Electricity payments;
- ❖ Student Registrations;
- ❖ Employment registrations (Tax authority).

It was expected that about **30%** of the population will have at least one additional address, other than the one in the **CPR**.

We predicted the true **CPR** residential address by **Machine Learning** methods. (Found effective in several pilot testing.)

19

Evaluation of sampling and estimation process

3. Compare the final census estimates for **large localities** to estimates obtained from other surveys, for which the estimates are sufficiently reliable. For example, to estimates obtained from the **LFS**, which includes home visits → the true addresses are **known**.

For a large locality L , one can estimate the size N_L from the **LFS** as follows. Estimate, $\hat{P}_{L|R}$ - proportion of individuals living in L out of those registered as living in L , and $\hat{P}_{R|L}$ - proportion of individuals registered in L out of those living in L . Hence, $N_L P_{R|L} = K_L P_{L|R}$ and $\hat{N}_L = K_L \hat{P}_{L|R} / \hat{P}_{R|L}$ is consistent for N_L .

20

Evaluation of data collection process

- 4.** Relevant indices for this process at the area level include:
- a.** Percent non-response in the area,
 - b.** Distribution of non-response across demographic classes,
 - c.** Percent of anchoring success,
 - d.** Zero cases

Further evaluation of estimation process

5. Apply a 'top down' evaluation procedure.

Basic idea: the sum of estimates of totals over all the areas nested in a large area L , should agree with the estimate of the total in L .

21

Evaluation of estimation process (cont.)

Denote the sampled units residing in **area i** by **j** , and the sampled units residing in the **locality L** by **k** . Ideally,

$$\sum_{i \in L} \sum_{j \in i} a_{ij} z_{ij} = \sum_{k \in L} a_k^{(L)} z_k. \quad (*)$$

A significant difference between the left- and right hand side can be attributed to many anchoring problems at the area level. In such cases, the weights a_{ij} can be modified to satisfy the constraint in (*). A modification often used, known as **pro-rata benchmarking** is,

$$a_{ij}^{(pr)} = a_{ij} \frac{\sum_{k \in L} a_k^{(L)} z_k}{\sum_{i \in L} \sum_{j \in i} a_{ij} z_{ij}}.$$

22

Some initial results from the 2022 census

1- The file contained **662,000** individuals, of which **528,000** reported on place of residence and **523,800** on Statistical Area.

2- Response rates:

National level- 81% (20% of households)

Jews and others*- 84%, Arabs- 54.4%,

Bedouins- 27.6%; 32% in recog. locations, 2% in scattered loc.

*** Others= Non-Arab Christians + unclassified by religion.**

3- Response rates by mode of response:

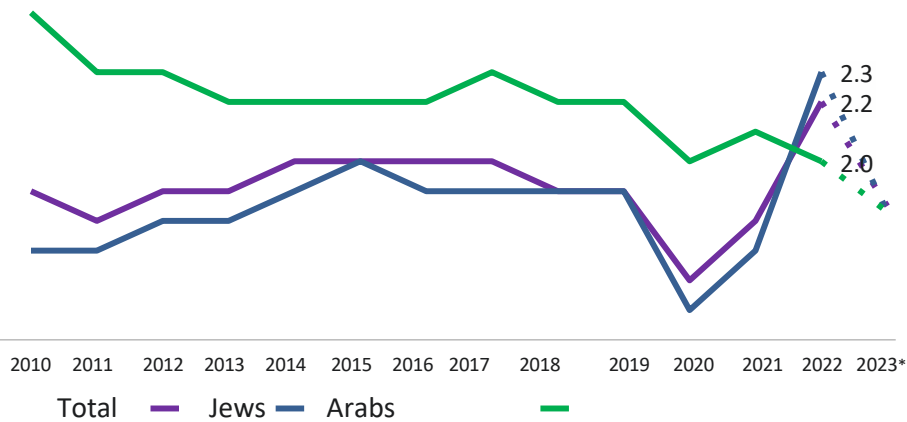
Internet- 38.5% (45.5% Jews and others, 7% Arabs)

Telephone- 31.5% (30% Jews and others, 35.2% Arabs)

Personal Interviews- 6.3% (5.5% Jews and others, 10% Arabs and Bedouins). (Bedouins only approached face to face.)

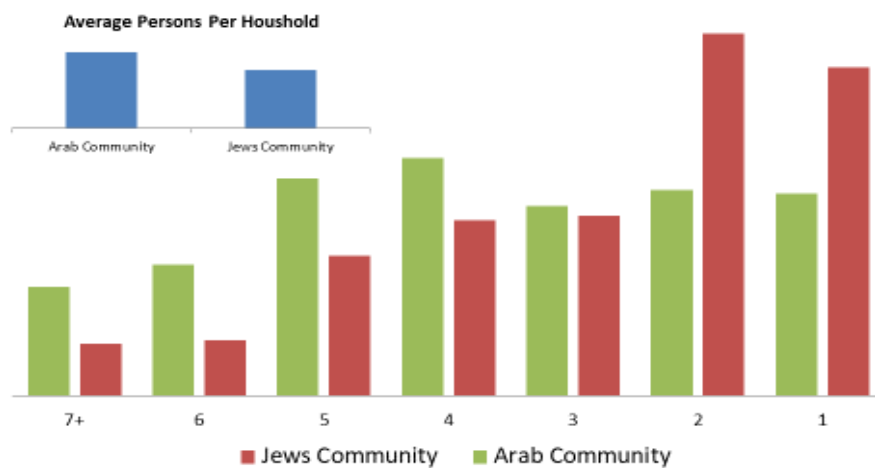
23

Growth rates of Jews and Arabs: 2010-2023



24

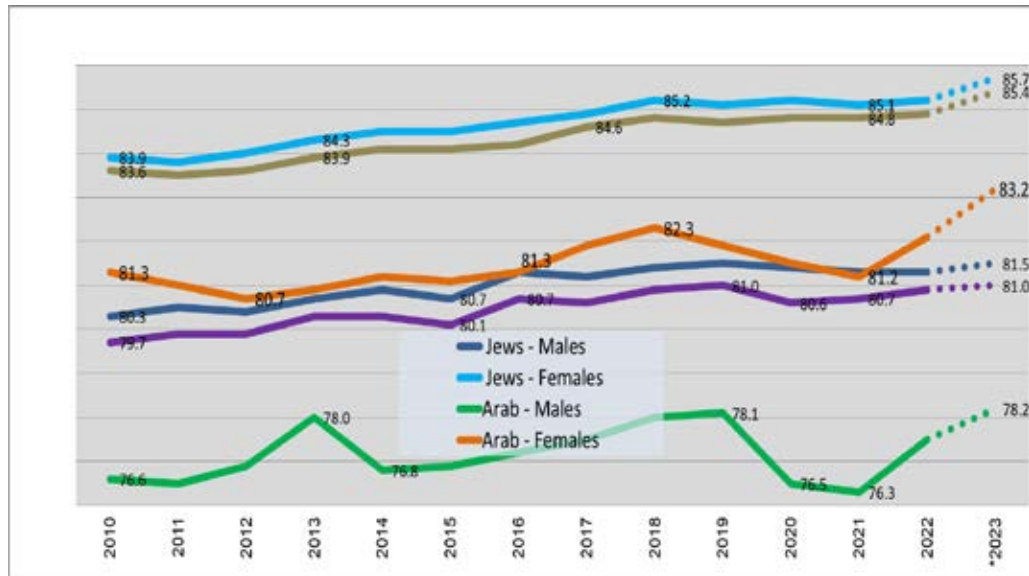
Household size by population groups, census 2022



Average size: Arabs- 3.75, Jews- 2.90.

25

Life expectation of Jews and Arabs- 2010-2023



26

Demographic indicators- International view

Indicators	ISRAEL (2023)	OECD	EU	Middle East & North Africa	WORLD
Growth Rate	1.85	0.34	0.04	1.46	0.79
Total Fertility Rate (TFR)	2.8	1.6	1.5	2.7	2.3
Percentage of age 0-14	27.5	17.2	14.8	29.8	25.3
Percentage of age 65+	12.6	18.0	21.3	5.4	9.8
Life Expectancy (E0)	83.2	78.7	80.3	72.8	71.3
Infant Mortality Rate	3.1	5.6	3.2	18.1	28.4
Percentage of born abroad	20.4	10.1	10.7	9.4	3.3

Source: World Bank Data Base: data of 2022 or the latest available.

27

Im zweiten inhaltlichen Block „Zensus 2022 – Modell und Durchführung“ erläuterte Herr Prof. Dr. Münnich (Universität Trier) in seinem Vortrag „Der deutsche Zensus im Wandel der Zeit“ insbesondere die Methodik zur Stichprobenziehung des Zensus 2022 und sich daraus ergebende Auswirkungen. Frau Dr. Schwitulla (Bayerisches Landesamt für Statistik) sprach über die Haushaltegenerierung im Zensus sowie erste empirische Untersuchungsergebnisse.

Der letzte Themenblock des ersten Veranstaltungstages stand unter dem Motto „Anwendung von Zensusdaten“. Zunächst zeigten Herr Prof. Dr. Kauermann und Herr Dr. Windmann (beide Ludwig-Maximilians-Universität München) einen Vergleich von Zensusdaten und Mietspiegeldaten in München. Anschließend präsentierte Frau Dittfurth (Bayerisches Landesamt für Umwelt) einen Werkstattbericht zu einem Projekt zur kleinräumigen Auswertung von Wärmedaten aus dem Zensus 2022 für den Energie-Atlas Bayern. Der Vortrag von Frau Dr. Würz (Otto-Friedrich-Universität Bamberg) bildete den Abschluss des ersten Tages. Sie stellte in ihrem Beitrag die Schätzung regionaler Einkommensindikatoren unter Transformationen und limitiertem Zugang zu Hilfsinformationen aus der Population unter Rückgriff auf Daten des Zensus 2011 vor.

Neben den Pausen zwischen den Vorträgen nutzten erfreulich viele Teilnehmende auch das gemeinsame Conference Dinner am Ende des ersten Veranstaltungstages, um sich weiter fachlich und auch privat auszutauschen und zu vernetzen. In schönem Ambiente auf dem Michelsberg kam man am Donnerstag noch einmal in großer Runde zusammen, um beim gemeinsamen Abendessen den Tag ausklingen zu lassen.

Dr. Ruth Brand



Dr. Ruth Brand (Jahrgang 1967) ist seit dem 1. Januar 2023 Präsidentin des Statistischen Bundesamtes und als solche mit den Aufgaben der Bundeswahlleiterin betraut. Nach Studium und Promotion der Wirtschaftswis-

senschaften in Hannover war sie als wissenschaftliche Mitarbeiterin an der Bundesanstalt für Arbeit in Nürnberg tätig. Im Jahr 2001 folgt der Wechsel in das Statistische Bundesamt, wo sie sich zunächst der faktischen Anonymisierung von statistischen Einzeldaten widmete. Danach war sie viele Jahre in der Landwirtschaftsstatistik tätig, bevor sie die Leitung der Abteilung „Gesundheit, Soziales, Bildung und Private Haushalte“ übernahm. Von April 2020 bis Dezember 2022 übernahm Dr. Brand die Leitung des Beschaffungsamtes des Bundesministeriums des Innern und für Heimat.

Der Freitag stand dann unter dem Motto „Zukunftsperspektiven des Zensus“. Zum Start in den zweiten Veranstaltungstag sprach Frau Dr. Brand, Präsidentin des Statistischen Bundesamtes, in ihrem Keynote-Vortrag über die Zukunft des Zensus in Deutschland.

Sie gab einen kurzen Rückblick auf den Zensus 2022 und die wertvollen Erfahrungen, die in der Vorbereitung, Durchführung und Auswertung gemacht wurden. Frau Dr. Brand erläuterte anschließend die sich wandelnden Anforderungen und Bedürfnisse an die amtlichen Daten am Beispiel der Bevölkerung in Deutschland, dem Gebäude- und Wohnungsbestand, dem Zusammenleben in Familien und Haushalten und zu Themen wie Arbeitsmarkt und Bildung. Ferner ging sie auf die Frage ein, wie diesen neuen Anforderungen Rechnung getragen werden kann. Ausschnitte ihres Vortrags können auf den Folien auf den Seiten 48 und 49 nachgelesen werden.

Auszug aus den Vortragsfolien von Dr. Ruth Brand:

12. StatistikTage Bamberg | Fürth | Zukunft des Zensus in Deutschland

destatis.de

Rückblick auf den Zensus 2022

- » Stichtagsverschiebung wegen Corona-Pandemie
- » Wanderungsbewegungen ausgelöst durch Ukraine-Krieg
- » Gebäude- und Wohnungszählung mit „only first“-Strategie
- » Stichprobenumfang der Haushaltebefragung
- » Aufwendige Datenaufbereitung

Lessons learned

 Statistisches Bundesamt (Destatis)

12. StatistikTage Bamberg | Fürth | Zukunft des Zensus in Deutschland

destatis.de

Zensus 2031



 Statistisches Bundesamt (Destatis)

Weggabelungen in die Zukunft



- » Rechtlicher Rahmen zeitnah erweitern wegen steigender Anforderungen (Registerzensusgesetz)
- » Vorteile der Verwaltungsdigitalisierung und Registermodernisierung nutzen, da wo sie sich bieten
- » Investitionen in Aufbau notwendig, Potenzial für Einsparungen

© nanoline icons by vuudu, CreativMarket / eigene Bearbeitung © Caviar-Premium Icons by Neway Lau, CreativMarket / eigene Bearbeitung

 Statistisches Bundesamt (Destatis)

Anschließend sprach Herr Prof. Dr. Schneider, Präsident der Deutschen Gesellschaft für Demographie (DGD), in seinem Vortrag über die Beziehung zwischen Wissenschaft und amtlicher Statistik. Er erläuterte dem Publikum, warum Daten aus den Sozialwissenschaften (wie beispielsweise SOEP- oder NEPS-Daten) den Großteil der Datengrundlage für wissenschaftliche Studien ausmachen – wenngleich die amtliche Statistik zu den größten Datenlieferanten in Deutschland zählt – und wo sich Ansatzpunkte finden ließen, die Beziehung von amtlicher Statistik und Wissenschaft zu verbessern.

Im letzten Vortrag der StatistikTage 2024 sprachen Frau Straub (Bayerisches Landesamt für Statistik) und Frau Setzer (Statistisches Bundesamt) über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder im Allgemeinen und auch speziell über Zugangsmöglichkeiten zu den Daten des Zensus 2022 und über sich daraus ergebende Forschungspotenziale.

Das letzte Highlight des Programms und gleichzeitig den Abschluss der Tagung bildete die Podiumsdiskussion zum Thema „Zukunft des Zensus in Deutschland“, moderiert von Herrn Prof. Dr. Zwick (Statistisches Bundesamt).

PODIUMSDISKUSSION ZUM THEMA „ZUKUNFT DES ZENSUS IN DEUTSCHLAND“

Prof. Dr. Markus Zwick, Statistisches Bundesamt

Das Bayerische Landesamt für Statistik und die Otto-Friedrich-Universität Bamberg organisierten im Rahmen des Statistik Netzwerks Bayern, die zwölften StatistikTage Bamberg|Fürth, diesmal zum Thema „Zensus – jetzt und in Zukunft“. Der erste Tag der StatistikTage präsentierte und diskutierte die Themen „Zensus international und in Deutschland“, Modell und Durchführung des Zensus 2022 sowie Möglichkeiten zur Anwendung der Zensusdaten. Der zweite Tag widmete sich insbesondere den Zukunftsperspektiven des Zensus. Abgeschlossen wurde die Veranstaltung mit der Podiumsdiskussion „Zukunft des Zensus in Deutschland“. An der Diskussion nahmen auf Seiten der amtlichen Statistik die Präsidentin des Statistisches Bundesamts, Dr. Ruth Brand sowie der Präsident des Bayerischen Landesamts für Statistik, Dr. Thomas Gößl teil. Die Wissenschaft wurde durch Prof. Dr. Ralf Münnich von der Universität Trier und Vorsitzender der Deutschen Statistischen Gesellschaft (DStatG) sowie Prof. Dr. Norbert Schneider von der Goethe-Universität Frankfurt und Vorsitzender der Deutschen Gesellschaft für Demographie (DGD) vertreten. Prof. Dr. Markus Zwick vom Statistisches Bundesamt moderierte die Diskussion.

In einer ersten Runde gaben die Diskutanten in einem Statement einen Blick in ihre Erwartungen bezüglich eines künftigen Zensus, der aus Sicht aller an der Diskussion Teilnehmenden, wesentlich stärker auf Registerdaten beruhen wird. Dr. Brand betonte dabei, dass die Auskunftgebenden stets zu den Merkmalen befragt werden müssen, an denen die Angaben über andere Quellen nicht oder nicht entsprechend der Qualitätsstandards der amtlichen Statistik vorhanden sind und somit neben der effizienten Umsetzung eines Zensus auch die Befragungsbelastung möglichst gering gehalten wird.

Dr. Gößl merkte in diesem Kontext an, dass die Statistischen Ämter der Länder letztendlich durch Verwaltungsakt die amtlichen Einwohnerzahlen feststellen. Daher sind bei einem Methodenwechsel immer auch dieses Ziel und die daraus folgenden Anforderungen an die Qualität auf tiefer regionaler Ebene zu erörtern und zu beantworten, damit am Ende die Ergebnisse auch gerichtsfest sind.

Die Zukunft des Zensus ist aus Sicht von Prof. Schneider unklar. Er stellte hierzu die Frage, wozu ein Zensus in 25 Jahren noch benötigt werde und worin dann ein Mehrwert bestünde und wie dieser ggf. vergrößert werden könne. Seine Vermutung war dahingehen, dass die Zusammenführung und Nutzung von Registerdaten zu einer Art Dauerbeobachtung führt, die keinen Zensus mehr erfordert.

Ergänzend hob Prof. Münnich hervor, dass aus seiner Sicht ohne jeden Zweifel in der Zukunft die Entwicklung in Richtung eines Registerzensus gehen werde. Allerdings sind hierbei noch rechtliche wie statistische Fragen zu klären. Es müsse auch darauf hingewiesen werden, dass die Register ohne permanente Auditierung kaum den notwendigen Qualitätskriterien entsprechen werden. Die Erwartung ist auch, dass nicht alle Informationsbedarfe über die Register zu decken sind und daher Befragungen auch weiterhin ergänzend eine Relevanz behalten, dies auch mit Hinblick auf die Qualitätssicherung.



Von links nach rechts: Prof. Dr. Markus Zwick, Dr. Ruth Brand, Dr. Thomas Gößl, Prof. Dr. Ralf Münnich und Prof. Dr. Norbert Schneider.

Die Diskussion entwickelte sich an der Frage weiter, wie der nächste Zensus 2031 konkret aussehen könnte und welche Informationsbedarfe damit zu bedienen sind. Aus Sicht des Statistischen Bundesamts ist es erforderlich, eine Antwort auf die Neuausrichtung des Systems der europäischen Bevölkerungs- und Wohnungsstatistik, sowie auf die steigenden Nutzerbedarfe auf nationaler Ebene und seitens der Wissenschaft zu finden. Die absehbaren Anforderungen können mit dem bisherigen System der Statistikproduktion im Bereich Bevölkerung nicht erfüllt werden. Sie setzen voraus, dass ein neues System zur Erfüllung des Datenbedarfs entwickelt wird, das über den Zensus hinausgeht, hier ist das Stichwort die Etablierung eines bevölkerungsstatistischen Gesamtsystems. Der Zensus 2031 wird ein erster Schritt in diese Richtung sein. Er wird noch stärker als bisher vorhandene Datenquellen in Verwaltung und Statistik nutzen, kombiniert mit weiterhin notwendigen, aber weniger umfangreiche Befragungen. In jedem Fall wird es nach jetziger Einschätzung 2031 wieder eine Gebäude- und Wohnungszählung geben.

Dr. Gößl unterstrich die Bedeutung des künftigen Systems der europäischen Bevölkerungs- und Wohnungsstatistik, über das in Brüssel unter Beteiligung von Bundesregierung und Bundesrat noch intensiv verhandelt wird. Klar sei, dass es 2031 nicht um einen Zensus auf Knopfdruck und ohne ergänzende Befragungen geht, sondern um eine Weiterentwicklung des registergestützten Zensus, wie er 2011 und 2022 durchgeführt wurde. Auch für die Bevölkerungsstatistik allgemein gelten die Anforderungen an die Qualität auf tiefer regionaler Ebene, die nur gemeinsam von den Statistischen Ämtern des Bundes und der Länder erreicht werden kann.

Seitens des Vorsitzenden der DStatG ist es kaum vorstellbar, bei der aktuellen Qualität der Register in Deutschland sowie eines derzeit noch fehlenden eindeutigen Identifikators ohne eingehende komplexe Forschung in 2031 einen Registerzensus durchzuführen. Auch aus Nachbarländern hört man, so Herr Prof. Dr. Münnich, dass auf kleinräumiger Ebene zu viele Fragen offenbleiben. Prof. Schneider ergänzte hierzu, dass sich die Welt rasch wandelt, Register aber erst einmal träge sind und man mit Registerdaten eher erfährt, was vorhanden ist und nicht unbedingt das was man wissen möchte.

Dies führte zu der Frage, ob Datenlücken, die sich bei einer Nutzung von Registern ergeben, durch die Nutzung nicht traditioneller Datenquellen, wie z. B. Fernerkundungsdaten, sowie moderne Methoden wie das maschinelle Lernen, zumindest teilweise geschlossen werden können. Frau Dr. Brand sieht den Einsatz von KI bei der Plausibilisierung und Aufbereitung von Daten schon jetzt als möglich an und betont, dass dies zum Teil schon konkret in der Planung ist, beispielsweise bei der Texterkennung und der darauf aufbauenden Klassifikation von Berufen. Weiter führte sie aus, dass zur Schließung von Datenlücken auch modellbasierte, datengetriebene statistische Methoden in Frage kommen. Es ist dabei aber zu berücksichtigen, dass die Funktionalität einer statistischen Methode von geeigneten Daten abhängt. Für den Registerzensus kommen zum Beispiel Modelle in Frage, die auf einer Kombination aus personen- und regionsspezifischen Informationen basieren.

Auf die klare Unterscheidung von bewährten Verfahren der KI wie dem maschinellen Lernen und Anwendungen der generativen KI bezogen sich die Anmerkungen von Dr. Gößl. Maschinelles Lernen wird in der amtlichen Statistik bereits erfolgreich eingesetzt, zuletzt bei der Haushaltegenerierung für den Zensus 2022. Dieses Teilprojekt hat das Bayerische Landesamt für Statistik fachlich und in der Konzeption und Entwicklung der Software für den Statistischen Verbund durchgeführt. Ganz anders seien die Anwendungen der generativen KI zu beurteilen, die bekanntermaßen halluzinieren und in der amtlichen Statistik nur mit großer Vorsicht geprüft werden können, soll das Ziel, verlässliche Information für alle zu liefern, nicht verfehlt werden.

Prof. Münnich merkte an, dass moderne ML- oder KI-Verfahren zu integrieren dort sinnvoll erscheint, wo neue Datenquellen kaum noch traditionell integriert werden können. Dabei muss aber stets auch berücksichtigt werden, wie man zu einem möglicherweise veränderten Qualitätsbegriff kommt – und wie sich dieser mit den European Statistics Code of Practice verbinden lässt. Betrachtet man aber vollständige Kohärenz auf allen Ebenen, inklusive der Problematik einer Haushaltegenerierung, dann macht es aus seiner Sicht Sinn, über Mikrosimulationen nachzudenken. Mittels statistischer Prädiktionsverfahren und unter Verwendung von mathematischen Optimierungsverfahren könnte so ein zensustypischer Datensatz erzeugt werden, der in sich kohärent ist und auch Schätzungen aus der Zensus-Stichprobe berücksichtigt. Neu zu integrieren wären Module zu Registerfehlern, die im finalen Zensus-Datensatz korrigiert sein müssen.

Seitens Prof. Schneider sollten künftig Menschen nicht mehr direkt befragt werden. Seine Erfahrungen in der Surveyforschung würden recht deutlich zeigen, dass Internet-gestützte, auch auf dem Smartphone funktionierende Eingabeplattformen effizient sind und Akzeptanz genießen. Hier könnte man die Chancen nutzen, Daten zu generieren, die in Registern nicht vorhanden sind. Ggf. könnten auch Video-gestützte Interviews als weitere Methode der Wahl durchgeführt werden.

Die folgende Diskussionsrunde fokussierte insbesondere auf die beiden Vorsitzenden der Fachgesellschaften mit der Frage, wie aus ihrer Sicht ihre Expertise in die Konzeptionsphase einer Registermodernisierung und eines Registerzensus bestmöglich eingebunden werden kann. Seitens der DStatG, so Prof. Münnich, ist sicher eine Methodenbegleitforschung essentiell und könnte selbstverständlich geliefert werden. Gerade die Bereiche Record Linkage, Matching, statistical Prediction, Microsimulation, Sampling, Calibration und Survey Statistik sind in der Deutschen Statistischen Gesellschaft sehr gut vertreten und können eingebracht werden. In diesem Zusammenhang wäre es natürlich sinnvoll, insbesondere wenn weitere Variablen in den Zensus integriert werden sollten, diese für eine Begleitforschung auch in Datenzentren (ggf. via Fernrechnen) zur Verfügung zu stellen.

Natürlich wünschen sich die wissenschaftlichen Fachgesellschaften, so Prof. Schneider, stärker in den Entscheidungsprozess eingebunden zu werden, welche Daten generiert werden sollen. Ebenso würden sich die Fachgesellschaften wünschen, stärker daran beteiligt zu werden, wie Sachverhalte definiert und kategorisiert werden, um die Vergleichbarkeit mit rein wissenschaftlichen Studien zu ermöglichen. Aus Sicht der Demografie und der Sozialwissenschaften wäre es daneben sinnvoll, die sozialstrukturelle Tiefe der Daten ebenfalls zu erhöhen. Hierbei können die Fachgesellschaften, wie die Deutschen Gesellschaft für Demographie, unterstützen.

Seitens der amtlichen Statistik betont Frau Dr. Brand, dass in die Konzeptionierung des Registerzensus bereits die fachliche Expertise unabhängiger Expertinnen und Experten einfließt, zum Beispiel in Form von wissenschaftlichen Gutachten für spezielle methodische Fragen. Auf diese Expertise werden die statistischen Ämter auch in Zukunft zurückgreifen. Die gesetzlich festgeschriebene Aufgabe der Methodenentwicklung wird allerdings immer beim Statistischen Bundesamt liegen.

Abschließend wagte die Diskussionsrunde den weiten Blick mit der Frage, wie der Zensus in 25 Jahren aussehen wird. In der Zukunftserwartung waren die Einschätzungen von Dr. Brand und Dr. Gößl sehr eng beieinander. Der Registerzensus soll Kernelement eines bevölkerungsstatistischen Gesamtsystems werden, mit dem die amtliche Statistik möglichst aufwandsarm und flexibel auf künftige Anforderungen reagieren kann. Beide erwarten, dass 2050 ein datensparsamer Multiple-Source-Ansatz verfolgt wird, der aus einer Kombination von Registern, Befragungen und anderen Verfahren der Datengewinnung besteht, so wie es zuletzt auch von der Kommission Zukunft Statistik bereits für mehrere Anwendungsfälle beschrieben wurde. Durch die Modernisierung der Registerlandschaft und die damit einhergehende Implementierung einer Identifikationsnummer in der Verwaltung bieten sich Möglichkeiten einer effektiven Umsetzung des Once-Only-Prinzips auch in der amtlichen Statistik. Frau Dr. Brand erwartet dabei auch, dass der Registerzensus Kernelement eines bevölkerungsstatistischen Gesamtsystems sein wird, mit dem die amtliche Statistik möglichst aufwandsarm und flexibel auf künftige Anforderungen reagieren kann. D. h. dass z. B. die Bevölkerungszahl nicht mehr alle zehn Jahre erhoben wird, sondern ggf. bei Bedarf auch unterjährig und dies mit weiteren interessierenden regional tief gegliederten Merkmalen.

Die Einschätzung seitens Prof. Schneider war, dass es letztlich primär nicht um die Frage geht, ob es in den kommenden Jahrzehnten Zensen geben wird und wie sie beschaffen sein werden. Das wird auf höchster politischer Ebene in Europa entschieden. Worum es aus seiner Sicht geht, ist die Frage, wie die amtliche Statistik ihren Output erhöhen und ihre Ressourcen möglichst effizient im Interesse der Nutzer einsetzen kann.



Als Fazit kann festgehalten werden, dass sich die erwarteten zukünftigen bevölkerungsstatistischen Informationsbedarfe nur mit qualitativ hochwertigen Registern und Verwaltungsdaten erfüllen lassen, die sich mit datenschutzkonformen direkten Identifikatoren untereinander und mit Surveys verknüpfen lassen. Zudem werden komplexe statistische Methoden und Verfahren notwendig sein, um aus der Integration der Daten Ergebnisse zu erstellen, die den Qualitätsanforderungen entsprechen. Zur Qualitätssicherung von Methoden und Konzepten bindet die amtliche Statistik die Expertise der empirischen Wissenschaft ein, die gesetzlich festgeschriebene Aufgabe der Methodenentwicklung nimmt dabei das Statistische Bundesamt im Benehmen mit den Statistischen Ämtern der Länder wahr. Zensustypische Ergebnisse liegen dann bei Bedarf auch unterjährig und in tiefer regionaler Differenzierung vor. Zu diesem Zielbild herrschte durchaus Einigkeit, aber auch darin, dass der Prozess dorthin wohl noch eine längere Wegstrecke ist, die in Teilen auch steinig sein wird. ■

Wir freuen uns, auch im nächsten Jahr die StatistikTage wieder stattfinden zu lassen, diesmal in den Räumlichkeiten des Bayerischen Landesamtes für Statistik in Fürth. Weitere Informationen zu Thema und Termin werden im Lauf des Jahres auf unserer Webseite bekannt gegeben: www.statistiknetzwerk.bayern.de.

“

Katrin Kräck, M.Sc. und Nina Storfinger, Dipl.Soz.Univ.

Statistik Netzwerk Bayern

Permanent wachsende Anforderungen an die Statistik sind ohne intensive Zusammenarbeit zwischen amtlicher und wissenschaftlicher Statistik nicht erfolgreich zu bewältigen. Das Bayerische Landesamt für Statistik erweitert seine Zusammenarbeit mit Einrichtungen der Wissenschaft und Forschung und fördert diese durch gemeinsame Veranstaltungen und Projekte. Zu diesem Zweck wurde im Februar 2013 das Statistik Netzwerk Bayern als Plattform für den gegenseitigen Austausch von Wissen und Erfahrung gegründet.

Mitglieder des Statistik Netzwerks Bayern sind:

- Bayerisches Landesamt für Statistik
- Bundesagentur für Arbeit
- Fachbereich Wirtschaftswissenschaften der Friedrich-Alexander-Universität Erlangen-Nürnberg
- Forschungszentrum des Bundesamts für Migration und Flüchtlinge
- ifo Institut – Leibniz-Institut für Wirtschaftsforschung an der Universität München e.V.
- Institut für Arbeitsmarkt- und Berufsforschung
- Institut für Geographie und Geologie der Julius-Maximilians-Universität Würzburg
- Institut für Statistik der Ludwig-Maximilians-Universität München
- Leibniz-Institut für Bildungsverläufe (LifBi)
- Max-Planck-Institut für Sozialrecht und Sozialpolitik – Munich Center for the Economics of Aging
- Ostbayerische Technische Hochschule Regensburg
- Otto-Friedrich-Universität Bamberg
- Technische Hochschule Deggendorf
- Universität Augsburg
- Wilhelm Löhe Hochschule

Weitere Netzwerkpartner können werden:

- Hochschulen und Fachhochschulen in Bayern sowie deren Einrichtungen und Lehrstühle/Professuren,
- Forschungseinrichtungen mit Sitz in Bayern, auch soweit sie unselbstständige Teile einer juristischen Person des öffentlichen oder des privaten Rechts sind.

Institutionen mit Sitz außerhalb Bayerns können je nach Forschungsinteresse und Projektlage als kooperierende Institutionen in die Arbeit des Statistik Netzwerks Bayern einbezogen werden. Im Gegensatz zu Netzwerkpartnern nehmen sie nicht an der Steuerung des Netzwerks teil. Sie können aber an den Angeboten und Veranstaltungen des Netzwerks teilhaben und werden regelmäßig über dessen Aktivitäten informiert.

Kontaktadresse bei Fragen oder Interesse an einer Beteiligung am Netzwerk:

Statistik Netzwerk Bayern, c/o Bayerisches Landesamt für Statistik, 90725 Fürth
www.statistiknetzwerk.bayern.de • statistiknetzwerk@statistik.bayern.de