DOWNLOAD PDF METHODOLOGY AND DEMOGRAPHIC DATA

Chapter 1 : Statistical Language - Census and Sample

The American Community Survey (ACS) is a national survey that uses continuous measurement methods. In this survey, a series of monthly samples produce annual estimates for the same small areas (census tracts and block groups) formerly surveyed via the decennial census long-form sample.

Sampling[edit] A census is often construed as the opposite of a sample as its intent is to count everyone in a population rather than a fraction. However, population censuses rely on a sampling frame to count the population. This is the only way to be sure that everyone has been included as otherwise those not responding would not be followed up on and individuals could be missed. The fundamental premise of a census is that the population is not known and a new estimate is to be made by the analysis of primary data. The use of a sampling frame is counterintuitive as it suggests that the population size is already known. However, a census is also used to collect attribute data on the individuals in the nation. This process of sampling marks the difference between historical census, which was a house to house process or the product of an imperial decree, and the modern statistical project. The sampling frame used by census is almost always an address register. Thus it is not known if there is anyone resident or how many people there are in each household. Depending on the mode of enumeration, a form is sent to the householder, an enumerator calls, or administrative records for the dwelling are accessed. As a preliminary to the dispatch of forms, census workers will check any address problems on the ground. While it may seem straightforward to use the postal service file for this purpose, this can be out of date and some dwellings may contain a number of independent households. As these are not easily enumerated by a single householder, they are often treated differently and visited by special teams of census workers to ensure they are classified appropriately. Residence definitions[edit] Individuals are normally counted within households and information is typically collected about the household structure and the housing. For this reason international documents refer to censuses of population and housing. Normally the census response is made by a household, indicating details of individuals resident there. An important aspect of census enumerations is determining which individuals can be counted from which cannot be counted. Broadly, three definitions can be used: This is important to consider individuals who have multiple or temporary addresses. Every person should be identified uniquely as resident in one place but where they happen to be on Census Day , their de facto residence , may not be the best place to count them. Where an individual uses services may be more useful and this is at their usual, or de jure, residence. An individual may be represented at a permanent address, perhaps a family home for students or long term migrants. It is necessary to have a precise definition of residence to decide whether visitors to a country should be included in the population count. This is becoming more important as students travel abroad for education for a period of several years. Other groups causing problems of enumeration are new born babies, refugees, people away on holiday, people moving home around census day, and people without a fixed address. People having second homes because of working in another part of the country or retaining a holiday cottage are difficult to fix at a particular address sometimes causing double counting or houses being mistakenly identified as vacant. Another problem is where people use a different address at different times e. Census enumeration has always been based on finding people where they live as there is no systematic alternative - any list you could use to find people is derived from census activities in the first place. Recent UN guidelines provide recommendation on enumerating such complex households. Modern approaches take into account the problems of overcount and undercount , and the coherence of census enumerations with other official sources of data. An important aspect of the census process is to evaluate the quality of the data. In census circles this method is called dual system enumeration DSE. A sample of households are visited by interviewers who record the details of the household as at census day. These data are then matched to census records and the number of people missed can be estimated by considering the number missed in the census or survey but counted in the other. This way counts can be adjusted for non-response varying between different demographic groups. An explanation using a fishing analogy can be found in "Trout, Catfish and Roach Enumerator conducting a survey using a mobile phone-based questionnaire in rural Zimbabwe. Triple system enumeration has been proposed as an

improvement as it would allow evaluation of the statistical dependence of pairs of sources. However, as the matching process is the most difficult aspect of census estimation this has never been implemented for a national enumeration. It would also be difficult to identify three different sources that were sufficiently different to make the triple system effort worthwhile. The DSE approach has another weakness in that it assumes there is no person counted twice over count. In de facto residence definitions this would not be a problem but in de jure definitions individuals risk being recorded on more than one form leading to double counting. A particular problem here are students who often have a term time and family address. Everyone receives the short form questions. Thereby more data are collected but not imposing a burden on the whole population. This also reduces the burden on the statistical office. Recently there has been controversy in Canada about the cessation of the long form with the head, Munir Sheikh resigning. This allows a simulated census to be conducted by linking several different administrative databases at an agreed time. Data can be matched and an overall enumeration established accounting for where the different sources are discrepant. A validation survey is still conducted in a similar way to the post enumeration survey employed in a traditional census. Other countries which have a population register use this as a basis for all the census statistics needed by users. This is most common among Nordic countries but requires a large number of different registers to be combined including population, housing, employment and education. These registers are then combined and brought up to the standard of a statistical register by comparing the data in different sources and ensuring the quality is sufficient for official statistics to be produced. In Brazil, handheld devices were used by enumerators to locate residences on the ground. In many countries, census returns could be made via the Internet as well as in paper form. DSE is facilitated by computer matching techniques which can be automated, such as propensity score matching. In the UK, all census formats are scanned and stored electronically before being destroyed, replacing the need for physical archives. The record linking to perform an administrative census would not be possible without large databases being stored on computer systems. New technology is not without problems in its introduction. The US census had intended to use the handheld computers but cost escalated and this was abandoned, with the contract being sold to Brazil. Online response is a good idea but one of the functions of census is to make sure everyone is counted accurately. A system which allowed people to enter their address without verification would be open to abuse. Therefore, households have to be verified on the ground, typically by an enumerator visit or post out. Paper forms are still necessary for those without access to Internet connections. It is also possible that the hidden nature of an administrative census means that users are not engaged with the importance of contributing their data to official statistics. Alternatively, population estimations may be carried out remotely with GIS and remote sensing technologies. Unfortunately, many countries have outdated or inaccurate data about their populations and therefore, without accurate data are unable to address the needs of their population. Planners need this information for all kinds of development work, including: An accurate census can empower local communities by providing them with the necessary information to participate in local decision-making and ensuring they are represented. Uses of census data[edit] In the nineteenth century, the first censuses collected paper enumerations that had to be collated by hand so the statistical uses were very basic. The government owned the data and were able to publish statistics themselves on the state of the nation. Population estimates could be compared to those of other countries. By the beginning of the twentieth century, censuses were recording households and some indications of their employment. In some countries, census archives are released for public examination after many decades, allowing genealogists to track the ancestry of interested people. Archives provide a substantial historical record which may challenge established notions of tradition. It is also possible to understand the societal history through job titles and arrangements for the destitute and sick. There are a lot of politics that surround the census in many countries. In Canada in for example, the government under the leadership of Stephen Harper abolished the mandatory long-form census. The decision to cut the long-form census was a response to protests from some Canadians who resented the personal questions. Census data and research[edit] As governments assumed responsibility for schooling and welfare, large government research departments made extensive use of census data. Actuarial estimates could be made to project populations and plan for provision in local government and regions. It was also possible for central government to allocate funding on

the basis of census data. Even into the mid twentieth century, census data was only directly accessible to large government departments. However, computers meant that tabulations could be used directly by university researchers , large businesses and local government offices. They could use the detail of the data to answer new questions and add to local and specialist knowledge. Now, census data are published in a wide variety of formats to be accessible to business, all levels of governance, media, students and teachers, charities and any citizen who is interested; researchers in particular have an interest in the role of Census Field Officers CFO and their assistants. Census data offer a unique insight into small areas and small demographic groups which sample data would be unable to capture with precision. For instance, when reporting data from a large city, it might be appropriate to give the average income for black males aged between 50 and Typically, census data are processed to obscure such individual information. Some agencies do this by intentionally introducing small statistical errors to prevent the identification of individuals in marginal populations; [25] others swap variables for similar respondents. Whatever measures have been taken to reduce the privacy risk in census data, new technology in the form of better electronic analysis of data poses increasing challenges to the protection of sensitive individual information. This is known as statistical disclosure control. Another possibility is to present survey results by means of statistical models in the form of a multivariate distribution mixture. As the final product does not contain any protected microdata, the model based interactive software can be distributed without any confidentiality concerns. Another method is simply to release no data at all, except very large scale data directly to the central government. Different release strategies between government have led to an international project IPUMS to co-ordinate access to microdata and corresponding metadata. Such projects also promote standardising metadata by projects such as SDMX so that best use can be made of the minimal data available. Egypt[edit] Censuses in Egypt first appear in the late Middle Kingdom and develops in the New Kingdom [27] Pharaoh Amasis , according to Herodotus , require every Egyptian to declare annually to the nomarch , "whence he gained his living".

Chapter 2 : Data Preparation and Analysis - Research Methodology Course

Methodology Research These documents describe the input data, methodology, and processes for the creation of population and housing unit estimates for the listed geographies.

EASI is a New York-based independent developer and marketer of desktop and internet demographic data and software solutions that provide demographic reports with unique search and analysis tools. Included with all software is a simple to learn mapping tool that performs street lookups, point maps, ring studies, create quintile analysis, and much more. EASI has been in business since with over 1, clients who use our databases, software, and on-line services. We use proprietary procedures, models, and algorithms to create that benchmark and then update and forecast the data elements in a manner that allows for accountability and accuracy; we call it Publication Quality. Publication Quality data must make sense. EASI works hard to be good "data accountants" and ensure that all tabulations logically sum to each other where expected, regardless of their distinct sources, or the rounding issues inherent in generating Block Group level data. These are our concerns, not our clients concerns. Published numbers, and numbers publishers, must be accountable. Clients often have questions, numbers are often surprising, and at EASI we believe our job does not end when we publish our data. We welcome the opportunities to answer client questions, and because we control every step of our updates in-house, we are able to do so thoroughly and quickly. In this discussion we will focus first upon the challenges created by the shortened Census Form, followed by a clear outline of how EASI will be able to provide you with ongoing quality estimates in spite of these issues. The Census, with its limited 10 question range, is a marked departure from the detailed Decennial data collection efforts of past years. This lack of detailed sample information will require EASI to maintain the high quality estimates of past decades, with less direct Census input than ever before. This change has caused statisticians and demographers to re-think many of the procedures that were previously used in producing updated demographic estimates. Fundamental data elements such as Population, Households, Race and Age will still be available from the Census. However many key indicators such as Income, Home Value, and Employment will not be released as part of the Decennial Census. Data providers will be now responsible for creating their own benchmarks for non-released demographics. Benchmarks will need to be developed from multiple, often diverse sources, creating more challenges for demographers than ever before. These are the data we will need to create the high quality, consistent Block Group demographics our clients have come to expect from us. The Census goal by law is true population counts for new Congress, redistricting, etc. In past decades, the SF3 Census File was a statistically significant sample covering the entire country, created once every 10 years; whereas ACS is a large multi year survey conducted annually. ACS represents a span of several years up to 5 years , rather than the "snapshot in time" that the Census represents i. What this really means is: The next most significant issues are the inherent inconsistencies found between multiple input sources as well as the specific definitional differences between Census and ACS data. These differences require far more expertise and attention to detail than ever before. Sources must be evaluated for accuracy and usefulness, inconsistencies must be resolved, and methods must be developed to integrate the best possible resources. For example, while the CPS Current Population Survey is a smaller sample than the ACS, it is optimized for accurate Population counts, and will be our source for updated Population and other variables. Several specific areas of concern include: ACS uses a "2 month rule" of length of stay whereas the Census question is "what is your usual residence". The Census includes no seasonal residences in their housing counts but ACS does. ACS income is inflation adjusted, Census income is not. ACS survey data 5 year version can be used for a Summary type file substitution on an annual basis compared with the previous decennial releases of SF3. Each year ACS adds a new year and drops an old year, keeping the survey fresh. As previously stated the most significant improvement in data collection is that the ACS is collected annually. This will keep the details of our annual estimates more up to date than ever before. The greatest advantages of utilizing ACS are: Each additional year will allow EASI to fine tune our benchmark data and ensure that the starting point for updates will be equal to or surpass what we used in previous censuses Update Accuracy: Substantial increase in the availability of more localized data for use in

models. This will generate greatly improved information about how very small areas such as Block Groups change demographically over time. In the past we were limited to variables which we knew we could reliably update during the 10 years between Census releases. Now that ACS is available annually at the local level, we are able to provide more detailed data than ever before. Specifically, more tables which can be differentiated by race and gender. The primary limitation of ACS: ACS survey results are released in single year files, in 3 year files, and in 5 year files. Because of the smaller ACS sample sizes there are reliability issues with some single year and 3 year data releases. Certain tables in each geography level may be excluded because of unacceptably high sampling errors. To overcome this, experienced vendors such as EASI will lean heavily on the more reliable 5 year ACS files and enhance the data with the 1 and 3 year files when analysis proves them to be reliable enough for Publication Quality estimates. This provides the optimal mix of accuracy and detail. The ability to create these tabulations from PUMS will allow EASI to create more significant benchmarked tables, as well as higher quality updates each year. EASI METHODOLGY Without divulging our exact proprietary methods, we want our clients to have an overview of how we will be handling the Census changes in order provide assurance that EASI will be doing the detailed and complex work necessary to resolve these Census changes in a manner that will be transparent to you and your applications of the data. During this procedure we capture data for over a thousand demographic variables. This wide data capture net will allow us to continue to publish all of the same data we have been publishing, and allow us to add additional tabulations to the EASI Data Library. Once we have gathered, cleaned, and organized all of our input data, EASI will follow this general outline to provide continuous quality estimates and forecasts, using procedures developed over the last 15 years, and modified for ACS input over the last 2 years: EASI will use as much 3 year data and 1 year data as is reliable enough for inclusion. Our goal as a vendor is to absorb changes as they arise and keep them transparent to our users. And as more ACS data become available our estimates will reflect the positive improvements that these data offer and, in many instances, to a degree not previously possible.

Chapter 3 : Methods | Pew Research Center

Demographic information provides data regarding research participants and is necessary for the determination of whether the individuals in a particular study are a representative sample of the target population for generalization purposes.

Survey Research Collecting survey data Survey researchers employ a variety of techniques in the collection of survey data. People can be contacted and surveyed using several different modes: The choice of mode can affect who can be interviewed in the survey, the availability of an effective way to sample people in the population, how people can be contacted and selected to be respondents, and who responds to the survey. In addition, factors related to the mode, such as the presence of an interviewer and whether information is communicated aurally or visually, can influence how people respond. Surveyors are increasingly conducting mixed-mode surveys where respondents are contacted and interviewed using a variety of modes. Survey response rates can vary for each mode and are affected by aspects of the survey design e. In addition to landline and cellphone surveys , Pew Research Center also conducts web surveys and mixed-mode surveys , where people can be surveyed by more than one mode. We discuss these types of surveys in the following sections and provide examples from polls that used each method. In addition, some of our surveys involve reinterviewing people we have previously surveyed to see if their attitudes or behaviors have changed. For example, in presidential election years we often interview voters, who were first surveyed earlier in the fall, again after the election in order to understand how their opinions may have changed from when they were interviewed previously. Cellphone surveys Telephone surveys have traditionally been conducted only by landline telephone. For certain subgroups, such as young adults, Hispanics and African Americans, the cell only rate is even higher. Research has shown that as the number of adults who are cell only has grown, the potential for bias in landline surveys that do not include cellphone interviews is growing. Cellphone surveys are conducted in conjunction with a landline survey to improve coverage. The data are then combined for analysis. In addition to the issues associated with sampling cellphones , there are also unique challenges that arise when interviewing people on their cellphones. One of the most important considerations when conducting cellphone surveys is that the costs are substantially higher than for a traditional landline survey. The cost of a completed cellphone interview is one-and-a-half to two times more than a completed landline interview. Although some of the fixed costs associated with landline surveys are not duplicated when a cellphone sample is added such as programming the questionnaire , other costs are higher data processing and weighting are more complex in dual-frame surveys. Cellphone surveys are more expensive because of the additional effort needed to screen for eligible respondents. A significant number of people reached on a cellphone are under the age of 18 and thus are not eligible for most of our surveys of adults. Cellphone surveys also cost more because federal regulations require cellphone numbers to be dialed manually whereas auto-dialers can be used to dial landline numbers before calls are transferred to interviewers. In addition, respondents including those to Pew Research surveys are often offered small cash reimbursements to help offset any costs they might incur for completing the survey on their cellphone. These payments, as well as the additional time necessary for interviewers to collect contact information in order to reimburse respondents, add to the cost of conducting cellphone surveys. Most cellphones also have caller identification or other screening devices that allow people to see the number that is calling before deciding to answer. People also differ considerably in how they use their cellphones e. Although people responding to landline surveys are generally at home, cellphone respondents can be virtually anywhere when receiving the call. Legal restrictions on the use of cellphones while driving, as well as concerns about safety, also have raised the issue of whether people should be responding to surveys on their cellphones while driving. In addition, people often talk on their cellphones in more open places where they may have less privacy; this may affect how they respond to survey questions, especially those that cover more sensitive topics. These concerns have led some surveyors including Pew Research Center to ask cellphone respondents whether they are in a safe place and whether they can speak freely before continuing with the interview. Lastly, the quality of connection may influence whether

an interview can be completed at that time, and interruptions may be more common on cellphones. Response rates are typically lower for cellphone surveys than for landline surveys. In terms of data quality, some researchers have suggested that respondents may be more distracted during a cellphone interview, but our research has not found substantive differences in the quality of responses between landline and cellphone interviews. Interviewer ratings of respondent cooperation and levels of distraction have been similar in the cell and landline samples, with cellphone respondents sometimes demonstrating even slightly greater cooperation and less distraction than landline respondents.

Chapter 4 : Demographics - SAGE Research Methods

Perhaps the most common and popular methods of direct collection of demographic data is the census. The census is commonly performed by a government agency and the methodology used is the individual or household enumeration.

This animation explains the concept of census and sample. If you are unable to access the video a Transcript. The animation requires Adobe Flash Player to run. There is no audio in this animation. How do we study a population? A population may be studied using one of two approaches: It is important to note that whether a census or a sample is used, both provide information that can be used to draw conclusions about the whole population. What is a census complete enumeration? A census is a study of every unit , everyone or everything, in a population. It is known as a complete enumeration, which means a complete count. What is a sample partial enumeration? A sample is a subset of units in a population, selected to represent all units in a population of interest. It is a partial enumeration because it is a count from part of the population. Information from the sampled units is used to estimate the characteristics for the entire population of interest. When to use a census or a sample? Once a population has been identified a decision needs to be made about whether taking a census or selecting a sample will be the more suitable option. There are advantages and disadvantages to using a census or sample to study a population: A sample must be robust in its design and large enough to provide a reliable representation of the whole population. Aspects to be considered when designing a sample include the level of accuracy required, cost, and the timing. Sampling can be random or non-random. In a random or probability sample each unit in the population has a chance of being selected, and this probability can be accurately determined. Probability or random sampling includes, but is not limited to, simple random sampling, systematic sampling, and stratified sampling. Random sampling makes it possible to produce population estimates from the data obtained from the units included in the sample. All members of the sample are chosen at random and have the same chance of being in the sample. A lottery draw is a good example of simple random sampling where the numbers are randomly generated from a defined range of numbers i. The first member of the sample is chosen at random then the other members of the sample are taken at intervals i. Relevant subgroups from within the population are identified and random samples are selected from within each strata. In a non-random or non-probability sample some units of the population have no chance of selection, the selection is non-random, or the probability of their selection can not be determined. In this method the sampling error cannot be estimated, making it difficult to infer population estimates from the sample. Non-random sampling includes convenience sampling, purposive sampling, quota sampling, and volunteer sampling Convenience sampling: Units are chosen based on their ease of access; Purposive sampling: The sample is chosen based on what the researcher thinks is appropriate for the study; Quota sampling: The researcher can select units as they choose, as long as they reach a defined quota; and Volunteer sampling: Collecting data about a population flowchart: Collecting Data about a Population Flowchart: Census and Sample Recommended: Read Data Sources next Further information:

Chapter 5 : Updated Demographicsâ€”Esri Demographics | ArcGIS

Two good examples are the recent emergence of "bio- demography," the incorporation of biological data and research questions into demographic studies; and "spatial demography," the incorporation of spatial data (geo-coded information and satellite-generated measures and images) into population studies.

Demographic data can be related to the Earth, the same as geographic data. Demographic Data usually represent geographical location, identification, or describe populations. This field of science and research can be applied to anything about the dynamic nature of the human population including how it changes over time and what factors are affecting the changes. This study also covers aspects of human population such as the size, structure, distribution, spatial and temporal changes in response to birth, death, aging or migration. Demographic data which are most commonly used include crude birth rate, general fertility rate, age-specific fertility rates, crude death rate, infant mortality rate, life expectancy, total fertility rate, gross reproduction rate and net reproduction ratio. Demographic data can be used in analyzing certain patterns and trends related to human religion, nationality, education and ethnicity. These data are also the basis for certain branches of studies like sociology and economics. Collection of demographic data can be broadly categorized into two methods: Direct demographic data collection is the process of collecting data straight from statistics registries which are responsible for tracking all birth and death records and also records pertaining to marital status and migration. Perhaps the most common and popular methods of direct collection of demographic data is the census. The census is commonly performed by a government agency and the methodology used is the individual or household enumeration. The interval between two census surveys may vary depending on the government conducting. In some countries, a census survey is conducted once a year or once every two years and still others do census once every 10 years. Once all the data collected are in place, information can already derived from individuals and households. The indirect method of demographic data collection may involve only certain people or informants in trying to get data for the entire population. For instance, one of the indirect demographic data methods is the sister method. In this method, a researchers only asks all the women on the number of their sisters who have died or have had children who have died at what age they died. From the collected data, the researchers will draw their analysis and conclusions based on indirect estimates on birth and death rates and then apply some mathematical formula so they can estimate trends representing the while population. Other indirect methods of demographic data collection may be to collect existing data from various organizations who have done a research survey and collate these data sources in order to determine trends and patterns. There are a lot of ways for demographic methods for modeling population processes. In fact, it is now a lot easier to get demographic data that can cover the whole planet while data users can drill down deep into the database to get more demographic data pertaining to very specific geographical area. With the popularity of the internet, looking for demographic data with corresponding analyses has become a lot easier and faster.

Chapter 6 : Gentrification Report Methodology

Esri 's Updated Demographics include current-year estimates and 5-year projections of U.S. demographic data. Esri develops the annual demographic datasets using a variety of sources, beginning with the latest base, then adding a mixture of administrative records and private sources to capture changes.

Data Preparation and Analysis Preparing Data After data collection, the researcher must prepare the data to be analyzed. Organizing the data correctly can save a lot of time and prevent mistakes. Most researchers choose to use a database or statistical analysis program e. Once the data has been entered, it is crucial that the researcher check the data for accuracy. This can be accomplished by spot-checking a random assortment of participant data groups, but this method is not as effective as re-entering the data a second time and searching for discrepancies. This method is particularly easy to do when using numerical data because the researcher can simply use the database program to sum the columns of the spreadsheet and then look for differences in the totals. One of the best methods of checking for accuracy is to use a specialized computer program that cross-checks double-entered data for discrepancies. Each descriptive statistic summarizes multiple discrete data points using a single number. They can tell the researcher the central tendency of the variable, meaning the average score of a participant on a given study measure. The researcher can also determine the distribution of scores on a given study measure, or the range in which scores appear. Additionally, descriptive statistics can be used to tell the researcher the frequency with which certain responses or scores arise on a given study measure. This amount of information is not enough information to conclude that vision correction affects economic productivity. Inferential statistics are necessary to draw conclusions of this kind. This means that for the most part, if a person is tall, they are likely to have a large shoe size, and conversely, if they are short, they are likely to have a smaller shoe size. Correlation can also be negative. For example, warmer temperatures outside may be negatively correlated with the number of hot chocolates sold at a local coffee shop. This is to say that as the temperature goes up, hot chocolate sales tend to go down. Although causality may seem to be implied in this situation, it is important to note that on a statistical level, correlation does not imply causation. A good researcher knows that there is no way to assess from correlation alone that a causal relationship exists between two variables. Determining causation is a difficult thing to do, and it is a common mistake to assert a cause-and-effect relationship when the study methodology does not support this assertion. Inferential Statistics Inferential statistics allow the researcher to begin making inferences about the hypothesis based on the data collected. This means that, while applying inferential statistics to data, the researcher is coming to conclusions about the population at large. Inferential statistics seek to generalize beyond the data in the study to find patterns that ostensibly exist in the target population. This course will not address the specific types of inferential statistics available to the researcher, but a succinct and very useful summary of them, complete with step-by-step examples and helpful descriptions, is available here. This difference must be due to the manipulation of the independent variable. No matter how well a researcher designs the study, there always exists a degree of error in the results. Statistical significance is aimed at determining the probability that the observed result of a study was due to the influence of something other than chance.

Chapter 7 : Survey methodology - Wikipedia

The series entitled "Research and Development - Methodology Reports from Statistics Sweden" presents results from research activities within Statistics Sweden. The focus of the series is on development of methods and.

Personal mall or street intercept survey Hybrids of the above. Research designs[edit] There are several different designs, or overall structures, that can be used in survey research. The three general types are cross-sectional, successive independent samples, and longitudinal studies. Successive independent samples studies[edit] A successive independent samples design draws multiple random samples from a population at one or more times. Such studies cannot, therefore, identify the causes of change over time necessarily. For successive independent samples designs to be effective, the samples must be drawn from the same population, and must be equally representative of it. If the samples are not comparable, the changes between samples may be due to demographic characteristics rather than time. In addition, the questions must be asked in the same way so that responses can be compared directly. Longitudinal studies[edit] Longitudinal studies take measure of the same random sample at multiple time points. Longitudinal studies are the easiest way to assess the effect of a naturally occurring event, such as divorce that cannot be tested experimentally. However, longitudinal studies are both expensive and difficult to do. This attrition of participants is not random, so samples can become less representative with successive assessments. To account for this, a researcher can compare the respondents who left the survey to those that did not, to see if they are statistically different populations. Respondents may also try to be self-consistent in spite of changes to survey answers. Questionnaires[edit] Questionnaires are the most commonly used tool in survey research. However, the results of a particular survey are worthless if the questionnaire is written inadequately. Nonresponse reduction[edit] The following ways have been recommended for reducing nonresponse [5] in telephone and face-to-face surveys: A short letter is sent in advance to inform the sampled respondents about the upcoming survey. The style of the letter should be personalized but not overdone. First, it announces that a phone call will be made, or an interviewer wants to make an appointment to do the survey face-to-face. Second, the research topic will be described. The interviewers are thoroughly trained in how to ask respondents questions, how to work with computers and making schedules for callbacks to respondents who were not reached. The interviewer should always start with a short introduction about him or herself. Also it can be useful to make clear that you are not selling anything: The questions asked must be clear, non-offensive and easy to respond to for the subjects under study. Brevity is also often cited as increasing response rate. A literature review found mixed evidence to support this claim for both written and verbal surveys, concluding that other factors may often be more important. Main interviewer traits that have been demonstrated to influence survey responses are race, [11] gender, [12] and relative body weight BMI. Hence, race of interviewer has been shown to affect responses to measures regarding racial attitudes, [14] interviewer sex responses to questions involving gender issues, [15] and interviewer BMI answers to eating and dieting-related questions. The explanation typically provided for interviewer effects is social desirability bias: Interviewer effects are one example survey response effects.

Chapter 8 : Descriptive Statistics

A census is the procedure of systematically acquiring and recording information about the members of a given calendrierdelascience.com term is used mostly in connection with national population and housing censuses; other common censuses include agriculture, business, and traffic censuses.

This appendix details the methods used in this study to project changes in the population size and geographic distribution of eight major religious groups from to It is organized in five sections. The first section explains how the baseline religious composition estimates were derived. The second section describes how key input data age and sex composition, fertility, mortality, migration and religious switching were gathered and standardized. The third part of this appendix introduces the projection methods and assumptions. The fourth section offers some important disclaimers about these projections. Estimating Religious Composition in Data Collection and Documentation Researchers acquired and analyzed religious composition information from about 2, data sources, including censuses, demographic surveys, general population surveys and other studies â€" the largest project of its kind to date. Censuses and nationally representative surveys can provide valid and reliable measures of religious landscapes when they are conducted following the best practices of social science research. Valid measurement in censuses and surveys also requires that respondents are free to provide information without fear of negative governmental or social consequences. However, variation in methods among censuses and surveys including sampling, question wording, response categories and period of data collection can lead to variation in results. Social, cultural or political factors also may affect how answers to census and survey questions are provided and recorded. The measure of religious identity in this study is sociological rather than theological. In order to have statistics that are comparable across countries, the study attempts to count individuals who self-identify with each religion. This includes people who hold beliefs that may be viewed as unorthodox or heretical by others who claim the same religion. It also includes people who do not regularly practice the behaviors prescribed by their religion, such as those who seldom pray or attend worship services. Pew Research Center staff standardized religion categories in all available censuses and surveys for each country. Censuses and surveys collect information on religious identity at different levels of specificity. For example, depending on the source, the most specific level of affiliation measured could be Christian, Protestant, Baptist or Southern Baptist. Researchers coded religious identities into standard categories that aggregate into the eight major global religious categories used in this report. Researchers sought a recent, reliable source â€" ideally, a census or large-scale demographic survey. Researchers favored sources in which religion was measured with a single question that permitted respondents to identify specific affiliations or no affiliation at all. In Vietnam, for example, the census and the Demographic and Health Survey did not adequately measure folk religion identities. Researchers instead relied on the Asian Barometer survey, which measured a wider range of religious identities, including identification with folk religions. Making Adjustments for Groups Not Adequately Measured As necessary, researchers made adjustments to the primary source s to account for omitted or underrepresented groups since small minority groups are sometimes not measured or not reported in surveys and censuses. Multiple survey sources, denomination counts and estimates produced by country experts for each nation were used to assess whether minority religious groups were omitted or undercounted in the selected primary source s. In cases where censuses and surveys lacked sufficient detail on minority groups, the study also drew on estimates provided by the World Religion Database, which takes into account other sources of information on religious affiliation, including statistical reports from religious groups themselves. Adjusting for Limitations in a Survey Questionnaire Usually, researchers assumed that members of underrepresented groups were included in the sample but were not adequately measured by the survey instrument. In a few cases, the study made adjustments based on evidence that political, legal or cultural dynamics in a country compromised the validity of self-reported religious identity. In India, for instance, there is evidence of a Christian undercount in the census; some Christians who belong to Scheduled Castes historically referred to as Untouchables or Dalits choose to identify as Hindu when completing official forms such as the census. Hinduism is the most common religion in India. Adjusting

for Sampling Limitations In some situations, underrepresented groups are likely to be omitted from the sample itself. For example, recent migrants who may not be fluent in the language used in a survey often are missing in samples. Accounting for groups not included in the sample requires proportionately deflating survey data to account for underrepresented populations. For example, researchers made adjustments to survey-based estimates in Europe where they found evidence that some survey samples and population registers underrepresented Muslim migrants. In this study, researchers sought to ensure that primary sources were representative of the entire country. When this was not the case, it was usually due to concerns about the safety of interviewers and census takers or disputes about political boundaries. In such cases, researchers attempted to make appropriate adjustments or find an alternative data source that was nationally representative. For example, the Sri Lankan census was not conducted in a handful of northern and eastern districts because of perilous conditions due to armed conflict. After analyzing religion data from earlier censuses, researchers determined that the areas that were not covered by the census historically had a different makeup than the rest of the country. Researchers adjusted the census data for Sri Lanka based on census data covering regions omitted in the census. In a small number of countries where the census did not measure religious affiliation or where survey data on religious affiliation had sampling limitations, researchers used ethnicity data to estimate the religious affiliation of small groups. For example, ethnicity data from the Russian census was used together with Generations and Gender Survey data to estimate the proportion of Muslims in Russia. Making Adjustments for the Religious Affiliation of Infants Parents are sometimes hesitant to report a religious affiliation for their infant children even though they will claim a religion for the child when he or she is slightly older. Researchers observed evidence of this phenomenon in some Christian-majority countries where Christian parents were disproportionately describing their infants as religiously unaffiliated. This is evident when comparing census numbers over multiple years. While some of this change may be explained by mortality and migration, it is at least partly due to parents being more willing to describe their older children as Christians. In order to compensate for this measurement bias in Brazil and a few other countries where there was evidence of this phenomenon, researchers applied the religious composition of older children those years old to infants and young children those years old. This adjustment was made only where there was a substantial difference between the religious composition of the youngest age group and children ages Census agencies typically make adjustments for missing data before reporting results. Some census agencies, such as Statistics Canada, have historically imputed religion values for respondents who have not answered the census religion question. The likelihood that religion data will be missing increases when religion questions are labeled as optional, as is the case on censuses in countries such as Australia, the United Kingdom and the Czech Republic. This strategy allows the census agencies to demonstrate that answering their religion question was indeed optional. Therefore, after making any necessary adjustments for undercounted groups, religious shares were recalculated based on the population of all people who gave valid responses to the census or survey. The effect of this approach was to proportionately raise the shares of all religious groups, including the religiously unaffiliated. Following the procedures described above, researchers produced national-level estimates of the religious composition of each country for the year measured by the primary source. Projecting Earlier Data to Estimates based on data collected prior to have been projected forward to In those cases, researchers used additional data on differential fertility, age and sex composition as well as migration to project populations forward to , the base year for the projections in this report. The religious composition used for each country generally matches the estimates used in the report, except in cases where new sources, including recently released census data, allowed researchers to update estimates. Input Data for Population Projections The demographic projections in this report use data on age and sex composition, fertility, mortality, migration and religious switching. This section describes how these data were gathered and standardized for use in the projections. Age and Sex Structure Procedures Religious affiliation varies by age. In order to calculate the median ages of religious groups and carry out population projections, researchers assembled age structures for each of the eight religious groups in every country. Data on age structures were collected in 20 age categories measured in five-year increments with a top value of 95 and above for males and females e. Age structures were compiled in three steps. First, census or survey data

were used to capture the religious affiliation of each available age group. Second, survey data on religion by age were adjusted to account for small sample sizes. These steps are described in detail below. Estimating Religion by Age and Sex Researchers constructed initial age structures by analyzing survey data sets, census data sets and tables published by census agencies. While censuses usually enumerate religion for the entire population, including children, general population surveys do not usually include interviews with children. Since age structures require religious affiliation data for children, children were assigned religious affiliations when necessary based on the best methods available. For data sets that measured religious affiliation only for adult respondents, yet included the number and ages of children and other adults in the household, researchers were able to estimate the religious affiliation of remaining household members. In most cases, the religious affiliation of the respondent or head of the household was assigned to all additional members of the household who were not surveyed. For many countries, reliable age data were not available for all eight religious groups. Adjustments to Minimize Errors Due to Sample Size The reliability of survey estimates is partly dependent on the number of people surveyed the sample size. Since respondents who identify with each religious group are divided into 40 age and sex categories, the number of Buddhists, for example, in any one age-sex category may be small and produce less reliable estimates than a larger count would produce. This introduces significant variation in patterns of religious affiliation by age: Affiliation levels may bounce between highs and lows for consecutive age groups. To eliminate unlikely variation, researchers smoothed data using statistical procedures intended to reveal the general underlying pattern. However, in some cases, the age categories reported by census agencies are in year age groups or aggregated for all adults above a certain age, such as Researchers used statistical modeling techniques to distribute the composition of these aggregated age groups across the more specific five-year age categories used in this study. Matching Religion by Age and Sex to Overall Population by Age and Sex The overall religious affiliation resulting from the age structure procedures sometimes varies from the religious composition estimated for the country using the procedures described in the first section of this appendix. This difference exists for two reasons. First, the data sources used for the age-structure procedure may be different from the data sources used for the religious composition. Second, overall religious compositions were adjusted manually to account for undercounts and sampling issues. In order to match the overall religious composition figures to the data on religion by age and sex, the age structure was adjusted. The adjustment procedure used is often referred to as iterative proportional fitting IPF , or raking. Raking makes adjustments to the percentages of religious affiliation for each age group without significantly altering the underlying religious affiliation patterns by age group. When survey or census data on the differential religious composition of age-sex groups were not available, each age-sex group was assigned the same religious composition. Lack of differential religious composition data by age-sex group was most common in countries with very small populations. This is the case, for example, when census data with overall religious composition results are available but a detailed breakdown by age and sex is not released by the census bureau, in which case another source must be used to generate the age structure. Sources are also different when multiple waves of a survey series have to be combined in order to have a sample size large enough to construct reliable age structures. Age structures were further adjusted in countries where the age structure data source is much older than the source used for the religious composition of the country. In order to harmonize the data on overall religious affiliation with the age structure data, the latter is aged in five-year projections while holding the religious composition data constant. In a small number of countries, age structures were estimated based in part on ethnicity or citizenship data. For example, all six Gulf Cooperation Council GCC countries release information on the age distribution of citizens and non-citizens, but only Bahrain further breaks down this information by Muslims and non-Muslims. Estimating Fertility In many countries, there are substantial differences in the number of children born to women in different religious groups. Furthermore, religious groups often vary in the share of women in their population who are of childbearing age, and women in some groups may, on average, begin having children at younger or older ages than do women in other groups. Fertility data were gathered from censuses and surveys, and fertility rates were estimated via direct and indirect measures. Some censuses and surveys directly measure recent births or the number of children a woman has ever born by the time of the survey.

Chapter 9 : UNSD â€" Methodology

This methodology white paper describes how Esri integrates data from the latest Consumer Expenditure Surveys (CEX)-the Diary Survey and the Interview Survey-from the Bureau of Labor Statistics with Tapestry Segmentation data to provide a comprehensive database about all consumer expenditures.

They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data. Descriptive statistics are typically distinguished from inferential statistics. With descriptive statistics you are simply describing what is or what the data shows. With inferential statistics, you are trying to reach conclusions that extend beyond the immediate data alone. For instance, we use inferential statistics to try to infer from the sample data what the population might think. Or, we use inferential statistics to make judgments of the probability that an observed difference between groups is a dependable one or one that might have happened by chance in this study. Descriptive Statistics are used to present quantitative descriptions in a manageable form. In a research study we may have lots of measures. Or we may measure a large number of people on any measure. Descriptive statistics help us to simplify large amounts of data in a sensible way. Each descriptive statistic reduces lots of data into a simpler summary. For instance, consider a simple number used to summarize how well a batter is performing in baseball, the batting average. This single number is simply the number of hits divided by the number of times at bat reported to three significant digits. A batter who is hitting. The single number describes a large number of discrete events. This single number describes the general performance of a student across a potentially wide range of course experiences. Every time you try to describe a large set of observations with a single indicator you run the risk of distorting the original data or losing important detail. Even given these limitations, descriptive statistics provide a powerful summary that may enable comparisons across people or other units. Univariate Analysis Univariate analysis involves the examination across cases of one variable at a time. There are three major characteristics of a single variable that we tend to look at: The distribution is a summary of the frequency of individual values or ranges of values for a variable. The simplest distribution would list every value of a variable and the number of persons who had each value. For instance, a typical way to describe the distribution of college students is by year in college, listing the number or percent of students at each of the four years. Or, we describe gender by listing the number or percent of males and females. In these cases, the variable has few enough values that we can list each one and summarize how many sample cases had the value. But what do we do for a variable like income or GPA? With these variables there can be a large number of possible values, with relatively few people having each one. In this case, we group the raw scores into categories according to ranges of values. For instance, we might look at GPA according to the letter grade ranges. Or, we might group income into four or five ranges of income values. One of the most common ways to describe a single variable is with a frequency distribution. Depending on the particular variable, all of the data values may be represented, or you may group the values into categories first e. Rather, the value are grouped into ranges and the frequencies determined. Frequency distributions can be depicted in two ways, as a table or as a graph. Table 1 shows an age frequency distribution with five categories of age ranges defined. The same frequency distribution can be depicted in a graph as shown in Figure 1. This type of graph is often referred to as a histogram or bar chart. Frequency distribution bar chart. Distributions may also be displayed using percentages. For example, you could use percentages to describe the: The central tendency of a distribution is an estimate of the "center" of a distribution of values. There are three major types of estimates of central tendency: Mean Median Mode The Mean or average is probably the most commonly used method of describing central tendency. To compute the mean all you do is add up all the values and divide by the number of values. For example, the mean or average quiz score is determined by summing all the scores and dividing by the number of students taking the exam. For example, consider the test score values: The Median is the score found at the exact middle of the set of values. One way to compute the median is to list all scores in numerical order, and then locate the score in the center of the sample. For example, if there are scores in the list, score would be the median. If we order the 8 scores shown above, we would get: Since both of these

scores are 20, the median is If the two middle scores had different values, you would have to interpolate to determine the median. The mode is the most frequently occurring value in the set of scores. To determine the mode, you might again order the scores as shown above, and then count each one. The most frequently occurring value is the mode. In our example, the value 15 occurs three times and is the model. In some distributions there is more than one modal value. For instance, in a bimodal distribution there are two values that occur most frequently. Notice that for the same set of 8 scores we got three different values -- If the distribution is truly normal i. Dispersion refers to the spread of the values around the central tendency. There are two common measures of dispersion, the range and the standard deviation. The range is simply the highest value minus the lowest value. The Standard Deviation is a more accurate and detailed estimate of dispersion because an outlier can greatly exaggerate the range as was true in this example where the single outlier value of 36 stands apart from the rest of the values. The Standard Deviation shows the relation that set of scores has to the mean of the sample. Again lets take the set of scores: We know from above that the mean is So, the differences from the mean are: