Scenario A1 - Exercise on data integration

In this exercise you will take on the role of researcher who has control over 1 particular source of data that by itself is not ideal to produce statistics on the topic of your interest. Every person in the classroom will receive an instruction paper, just like the one you are reading. However, every person will have control over a different data source, and will perhaps also have slightly different interests in the type of statistics that he/she wants to produce. The task in this exercise is to:

1. Find 2 other people who have the same scenario letter (a,b) but a different number (1,2,3) as you. Make groups of 3.
2. Find a place where you can discuss with eachother.
3. Figure out what data source each person controls. Please do not directly share the instructions below, but be sure to inform others about your data and desires.
4. Discuss ways to somehow combine or integrate these data sources to produce 'better' statistics. You can also choose to integrate 2 sources rather than 3. Be sure think about sampling, fieldwork, nonresponse and measurement.
5. You have 30 minutes for this exercise. Please make some notes about what you discussed, and the solution you find (if you find one!) for integrating data. Make sure you return in-class 30 minutes from the time you are reading this sentence.

Specific instructions for you:

You work at the National Statistical Institute of your country, and are in the lucky position that you have access to raw data from the population record/census data that contains the following details about the entire population of your country:

- Postal address
- Household size
- Whether person has any recognized children
- How many children person has
- Sex as established at birth
- Date of birth/age
- Household Income in the previous tax year
- Country of birth

The NSI you work for on top of this conducts a general social attitude survey in which only 2 questions are asked about health:

- Subjective health (1-5 scale going from very poor to excellent)
- Been in hospital for medical treatment in the past year

The NSI is interested in producing more health statistics on for example health behavior (exercise, physical activity), but unfortunately does not have funds to do more surveys. You would therefore be very happy if other people could conduct such a survey. You are willing to help with this where you can, e.g. by allowing use of the population register as long as the survey can be used to produce statistics for the entire population.

Scenario A2 - Exercise on data integration

In this exercise you will take on the role of researcher who has control over 1 particular source of data that by itself is not ideal to produce statistics on the topic of your interest. Every person in the classroom will receive an instruction paper, just like the one you are reading. However, every person will have control over a different data source, and will perhaps also have slightly different interests in the type of statistics that he/she wants to produce. The task in this exercise is to:

1. Find 2 other people who have the same scenario letter (a,b) but a different number (1,2,3) as you. Make groups of 3.
2. Find a place where you can discuss with eachother.
3. Figure out what data source each person controls. Please do not directly share the instructions below, but be sure to inform others about your data and desires.
4. Discuss ways to somehow combine or integrate these data sources to produce 'better' statistics. You can also choose to integrate 2 sources rather than 3. Be sure think about sampling, fieldwork, nonresponse and measurement.
5. You have 30 minutes for this exercise. Please make some notes about what you discussed, and the solution you find (if you find one!) for integrating data. Make sure you return in-class 30 minutes from the time you are reading this sentence.

Specific instructions for you:

You are a researcher working at a large hospital interested in studying the effects of physical activity on different kinds of health outcomes (e.g. disease, feelings of fitness) especially for older people. The hospital you work has access to lots of equipment to measure physical activity:

- Labs that can be used to measure physical condition (e.g. VO2-MAX)
- About 200 thigh-worn accelerometers that measure movement at a rate of 60 Hz. These devices can be used to measure physical activity in daily life with great precision
- You could potentially also use all other hospital equipment (e.g. CT scanners), but only when they are not used for treating other patients.

You have so far done quite a lot of research with patients that are recovering from various kinds of accidents/sickness for which they needed extensive hospital treatment (e.g. Covid patients). You monitored how these people recover, and monitored their physical condition.

You would actually like to extend this research to groups of 'healthy' patients outside the hospital. First because it would give you data on 'healthy' patients that you can use a benchmark against your hospital patients. But also because you think the accelerometers in particular work great. You can contribute expertise, the devices and time, but have no budget for doing these kind of studies outside of the hospital and are a bit frustrated by that.

Scenario A3 - Exercise on data integration

In this exercise you will take on the role of researcher who has control over 1 particular source of data that by itself is not ideal to produce statistics on the topic of your interest. Every person in the classroom will receive an instruction paper, just like the one you are reading. However, every person will have control over a different data source, and will perhaps also have slightly different interests in the type of statistics that he/she wants to produce. The task in this exercise is to:

1. Find 2 other people who have the same scenario letter (a,b) but a different number (1,2,3) as you. Make groups of 3.
2. Find a place where you can discuss with eachother.
3. Figure out what data source each person controls. Please do not directly share the instructions below, but be sure to inform others about your data and desires.
4. Discuss ways to somehow combine or integrate these data sources to produce 'better' statistics. You can also choose to integrate 2 sources rather than 3. Be sure think about sampling, fieldwork, nonresponse and measurement.
5. You have 30 minutes for this exercise. Please make some notes about what you discussed, and the solution you find (if you find one!) for integrating data. Make sure you return in-class 30 minutes from the time you are reading this sentence.

Specific instructions for you:

You are a health researcher working at the national institute for public health. You regularly conduct surveys on health in the general population on topics like eating and drinking behavior, health problems people experience, and health attitudes. You are happy with your work and think the surveys you do are valuable. The government has recently given you institute a 20% increase in your budget (about 200k euros extra) to conduct these surveys. You could use this to for example increase the net sample size of the survey from 10.000 respondents annually to about 13.000, but would actually prefer to get more behavioral measures about the people in the survey. For example, it would be great if you could get information about eating behavior (diets), sports activities, smoking behavior, exercise, etc. that do not rely on self-report by the respondents. You know that the self-report questions you get know suffer from reliability issues, and underreports of bad behavior, as well as overreports of good behaviors.

Another problem you face is in sampling. You now conduct the survey by sending letters by post to households, but have trouble reaching particular kinds of groups: young people, males, lower educated groups, migrant groups, and the oldest old. It would be terrific if you could target these groups in particular by somehow getting information from e.g. hospitals or the government that you can use to target invitations more at these groups.

Scenario B1 - Exercise on data integration

In this exercise you will take on the role of researcher who has control over 1 particular source of data that by itself is not ideal to produce statistics on the topic of your interest. Every person in the classroom will receive an instruction paper, just like the one you are reading. However, every person will have control over a different data source, and will perhaps also have slightly different interests in the type of statistics that he/she wants to produce. The task in this exercise is to:

1. Find 2 other people who have the same scenario letter (a,b) but a different number (1,2,3) as you. Make groups of 3.
2. Find a place where you can discuss with eachother.
3. Figure out what data source each person controls. Please do not directly share the instructions below, but be sure to inform others about your data and desires.
4. Discuss ways to somehow combine or integrate these data sources to produce 'better' statistics. You can also choose to integrate 2 sources rather than 3. Be sure think about sampling, fieldwork, nonresponse and measurement.
5. You have 30 minutes for this exercise. Please make some notes about what you discussed, and the solution you find (if you find one!) for integrating data. Make sure you return in-class 30 minutes from the time you are reading this sentence.

Specific instructions for you:

You work at the ministry of transport in your country. Over the years you have been responsible for translating data from the travel diary survey into actionable insights for transport planning.
The travel diary asks respondents to keep a diary of all their travels for 14 days. This includes time and locations of the start and end of trips, as well as the mode of transport, purpose of the trip, and whether the trip was joint with other people.
You are happy doing the work you do, and do not like change. Some colleagues have urged you to look at new sources of data, such as data from Google, or GPS data that is collected with smartphones or travel loggers. You see that these data are perhaps useful, but it will mean your datasets will be structured completely differently.

What you would like to do is an extra study. You keep the travel diary survey, and next to that, a new study is designed where people are followed with their smartphone. Based on GPS data, a smartphone app will automatically generate a diary with stops and trips.
You would not like to spend a lot of money on it, at most 5.000 euros. This would suffice to perhaps use an already existing app in a relatively small sample of respondents.

Scenario B2 - Exercise on data integration

In this exercise you will take on the role of researcher who has control over 1 particular source of data that by itself is not ideal to produce statistics on the topic of your interest. Every person in the classroom will receive an instruction paper, just like the one you are reading. However, every person will have control over a different data source, and will perhaps also have slightly different interests in the type of statistics that he/she wants to produce. The task in this exercise is to:

1. Find 2 other people who have the same scenario letter (a,b) but a different number (1,2,3) as you. Make groups of 3.
2. Find a place where you can discuss with eachother.
3. Figure out what data source each person controls. Please do not directly share the instructions below, but be sure to inform others about your data and desires.
4. Discuss ways to somehow combine or integrate these data sources to produce 'better' statistics. You can also choose to integrate 2 sources rather than 3. Be sure think about sampling, fieldwork, nonresponse and measurement.
5. You have 30 minutes for this exercise. Please make some notes about what you discussed, and the solution you find (if you find one!) for integrating data. Make sure you return in-class 30 minutes from the time you are reading this sentence.

Specific instructions for you:

You work at a tech tart-up where you specialize in the development of smartphone applications. You recently developed an app where people can keep track of their travel behaviour using location data. They idea is that people keep the app on their phone, and let it run in the background. With time, the app will, produce all kinds of fun statistics, such as what a persons favorite shop is, how much they spend outdoors, how many kilometers they cycle in different months of the year, how much one travels compared to other people using the app, etc. The app will then feed these fun statistics back to respondents at random moments (about once a week).

You are very happy with the app, but your boss less so. So far, the app has made little money from in-app advertising, and your boss has encouraged you to find other ways to make money with the app. In order to recover development costs, you want to make at least 100.000 euros with the app over the next 5 years without investing much time into adapting the app, apart from regular maintenance. You are trying to see whether academic researchers or policy makers from the ministry are perhaps interested in buying your app. It would for them be a method to collect fine-grained location data over a long period of time.

Scenario B3 - Exercise on data integration

In this exercise you will take on the role of researcher who has control over 1 particular source of data that by itself is not ideal to produce statistics on the topic of your interest. Every person in the classroom will receive an instruction paper, just like the one you are reading. However, every person will have control over a different data source, and will perhaps also have slightly different interests in the type of statistics that he/she wants to produce. The task in this exercise is to:

1. Find 2 other people who have the same scenario letter (a,b) but a different number (1,2,3) as you. Make groups of 3.
2. Find a place where you can discuss with eachother.
3. Figure out what data source each person controls. Please do not directly share the instructions below, but be sure to inform others about your data and desires.
4. Discuss ways to somehow combine or integrate these data sources to produce 'better' statistics. You can also choose to integrate 2 sources rather than 3. Be sure think about sampling, fieldwork, nonresponse and measurement.
5. You have 30 minutes for this exercise. Please make some notes about what you discussed, and the solution you find (if you find one!) for integrating data. Make sure you return in-class 30 minutes from the time you are reading this sentence.

Specific instructions for you:

You work at a university, in the department of human geography. You are interested in why and how people make travel decisions. That is, whether they decide to travel, and themn how they travel (what route, with what mode of transport).
You were recently successful in getting a large grant to investigate how smartphone apps may provide a new method to generate travel data from individuals. You promised in your grant that doing data collection this way will eradicate the need for old fashioned diary surveys, that have so far been used. These travel diaries are very burdensome for respondents, and so it is hard to find anyone willing to do such studies. Moreover, it is widely known travel diary surveys suffer from recall errors, underreporting of trips, and misreporting of for example start and end times. Location data via smartphones will solve many of these problems. In your valorisation paragraph, you promised that you will convince the ministry to use this new method. You talked to them before, and noticed they were hesitant to change methods.
The budget for your grant is 500.000 euros. You want to hire a Ph.d student, and have time for yourself. You reserved about 25.000 euros to buy an app, or let someone develop an app, and about 25.000 euros for doing fieldwork with about 1000 respondents to test the new method.