Research Methodology: Descriptive Research

Yesoda Bhargava

December 22, 2020

Introduction

- The term **descriptive research** encompasses a variety of methodologies that are best suited to examining and trying to make sense of a situation or event *as it currently exists* in the world.
- It includes basically two types of research designs:
  1. Designs aimed at characterizing the general nature of an observed phenomenon. Example: Human Ageing Observation.
  2. Designs aimed at revealing associations among two or more phenomenon. Example: How human ageing differs in different populations and what could be the possible explanations.
- Caution: Descriptive Research does *not* involve changing or modifying a situation under investigation, nor it is intended to determine cause-and-effect relationships.
- In this lecture, we focus on quantitative approaches in descriptive research. The concepts also apply to qualitative approaches.

Research Designs to be discussed

- Observation studies
- Correlational research
- Developmental designs
- Experience-sampling methods
- Survey research

Observation Studies

- Unlike in an experiment or clinical trial, the existing situation is observed without changing anything, to understand what is happening.
- Example: Cigarette smoking causes cancer.
- Imagine a study in which children are randomly allocated to a "20 cigarettes a day for 50 years" group and "never smoke a cigarette in your life" group.
- Unethical and we cannot persuade subjects to stick to the treatment. Hence, we must observe the disease process as best as we can.
- Recommended reading: British Doctors Study (Di Cicco et al., 2016).

Mortality in relation to smoking: the British Doctors Study

- The British doctors study: the first strong statistical proof of association between smoking and many diseases.
- In 1950s, smoking was nowhere close to being associated with lung cancer.
- In the 1950s, two scientists Richard Doll and Bradford Hill, found out he risk of lung cancer was related to the number of cigarettes smoked per day (the risk was 25 times higher in those who smoked $>$25 cigarettes a day than in nonsmokers).
- But there was a problem. What?
- How was the association to be proved beyond any reasonable doubt? Controlled trial would be unethical.
- In 1951, Doll and Hill sent questionnaire to all the doctors resident in the UK.
- These doctors were followed up in time. Study continued for 50 years.
- New questionnaires were sent to the study subjects (in 1957, 1966, 1971, 1978, 1991 and 2001) in order to gather information on changes in smoking habits and medical history.
- Finding: The risk of death from lung cancer was related to the amount of tobacco smoked (the annual death rate was 0.07 per 1000 in nonsmokers and 3.15 per 1000 in men smoking 35 or more cigarettes per day).

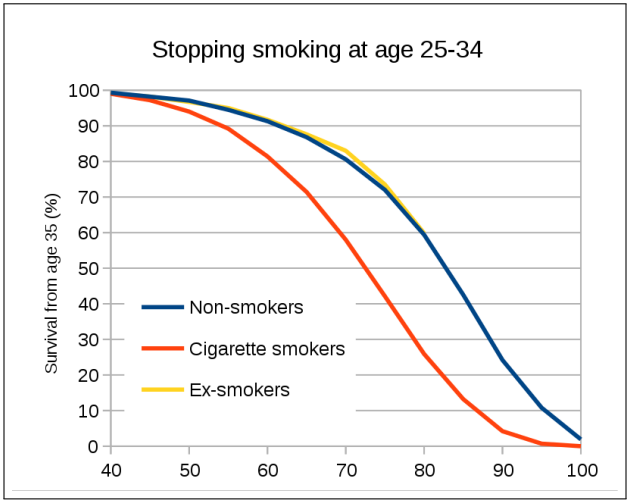

Figure 1: Survival from age 35 of non-smokers, cigarette smokers and ex-smokers who stopped smoking between 25 and 34 years old.

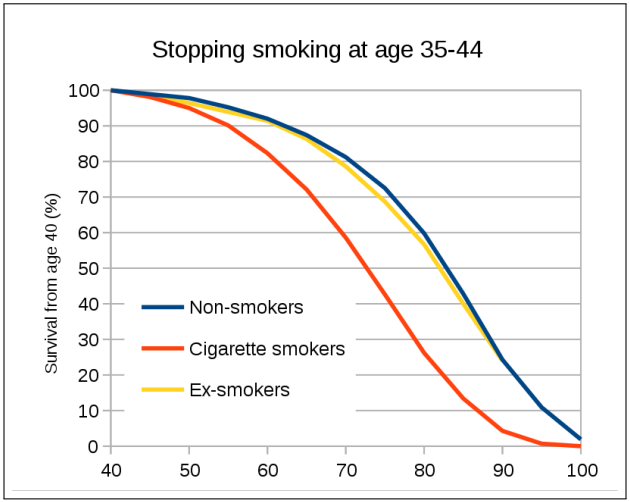

Figure 2: Survival from age 40 of non-smokers, cigarette smokers and ex-smokers who stopped smoking between 35 and 44 years oldxt

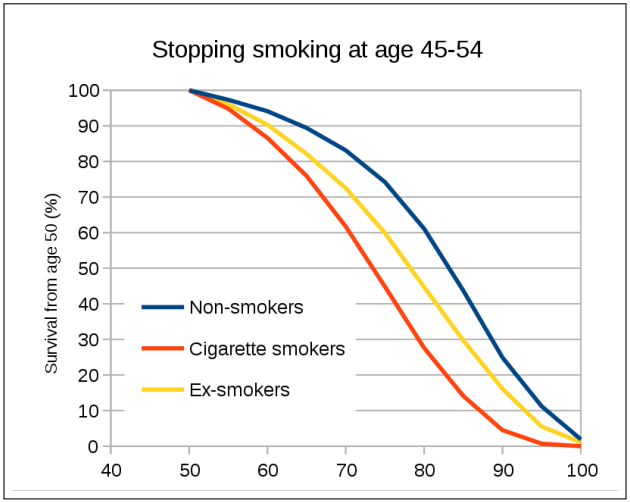

Figure 3: Survival from age 50 of non-smokers, cigarette smokers and ex-smokers who stopped smoking between 45 and 54 years old

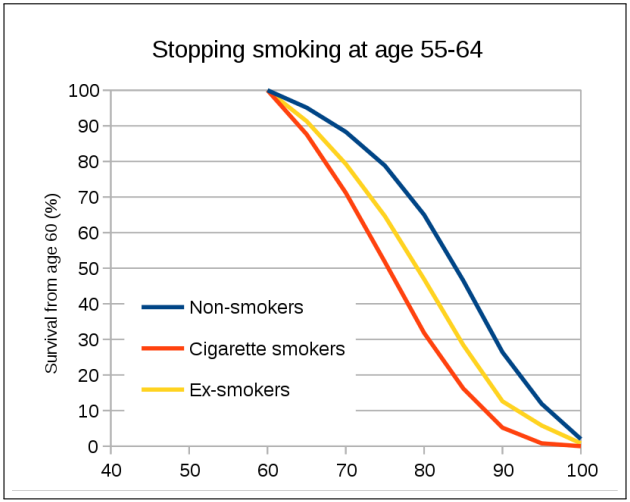

Figure 4: Survival from age 60 of non-smokers, cigarette smokers and ex-smokers who stopped smoking between 55 and 64 years old

Correlational Research

- A **correlational study** examines the extent to which differences in one variable are associated with differences in one or more variables.
- A correlation exists if, when one variable increases, another variable either increases or decreases.
- Knowing the value of one variable, then, enables us to predict the value of the other variable with some degree of accuracy.
- Quantitative data is gathered about two or more characteristics for a particular group of people.
- Eg. I want to check how the milage of a car depends on the weight of the car.
- Load mtcars dataset in R.
- Plot the data on x-y plane, often called as the scatter plot to allow visual inspection of the relationship between milage and weight. Fig. 5

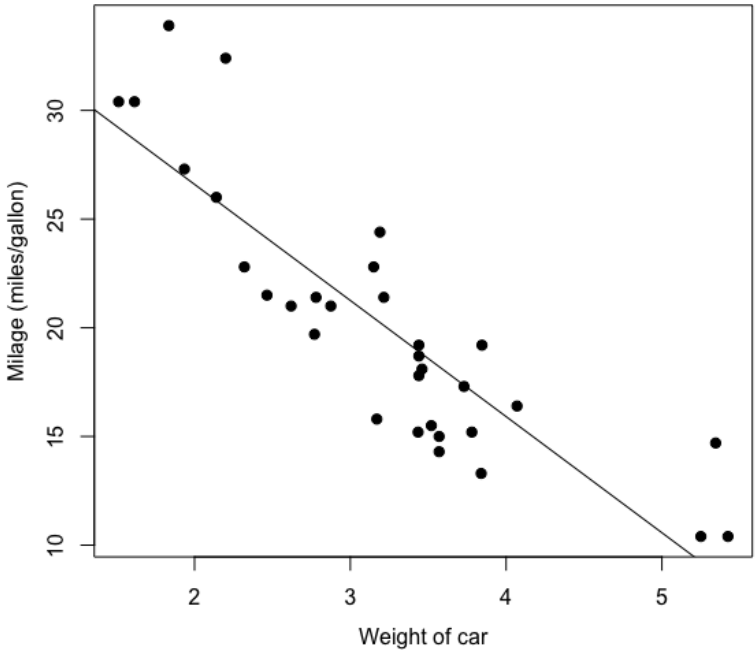

Figure 5: Example of a scatter plot: Correlation between Milage and Weight of Car.

Observations from Scatter plot

- Higher weight of car is associated with lower milage. There is a negative correlation; higher the weight of the car, lower the milage.
- Correlation value is -0.867.
- What if there was no correlation? How would a scatter plot look like?
- Scatter plot can only tell linear relationships, not non-linear.

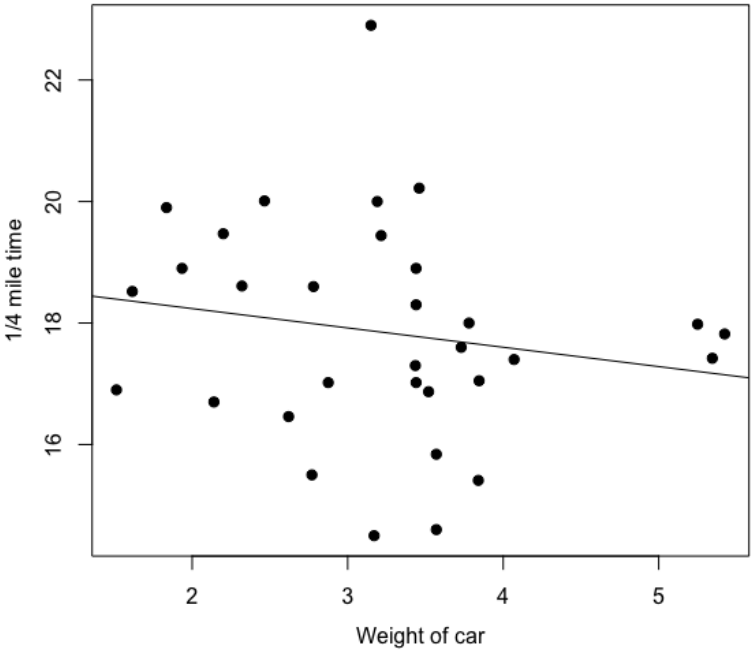

Figure 6: Example of a scatter plot: Correlation between 1/4 mile time and Weight of Car.

Interpretation of scatter plot 1/4 mile time and weight of car

- The data points are scattered without any visible pattern.
- Correlation value is -0.17.
- This scatter plot does not show any signs of relationship.
- What about this plot?

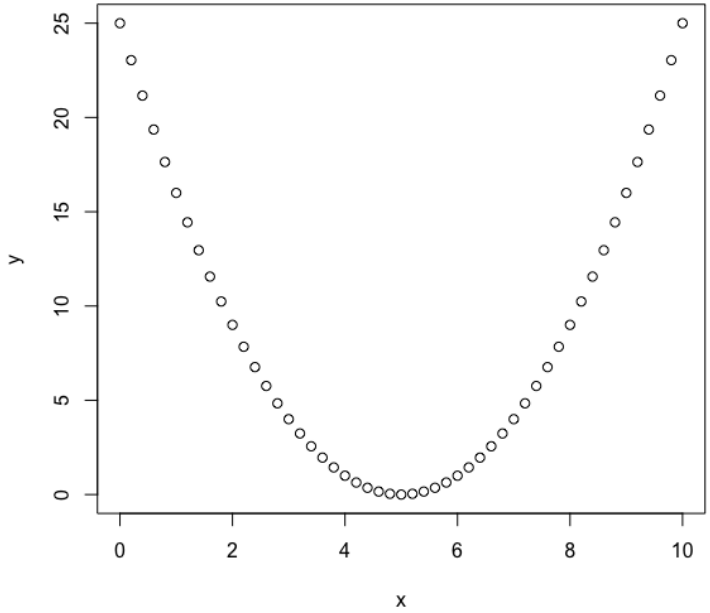

Figure 7: Example of a scatter plot. Correlation=0. Curvilinear relationship exists even though no linear relationship is seen through correlation value.

Takeaways

- Always plot your data.
- Plotting plus statistical results in numerical forms provide a complete picture of the data.
- Correlation absence may not mean absence of relationship.
- It just means weak linear relationship, perhaps a non-linear relationship exists.
- This becomes clear with plotting the data points.
- Bottomline: Always plot your data.

Correlation is not causation

- Just because two variables are linearly associated does not mean one is the cause of another.
- Example: Weight of the car does not cause the milage to be less.
- The theoretical explanation for lower milage of cars with increase in weight could be due to some other mechanisms.
- High weight may not, per se, lead to lower milage.
- Another example: As ice cream sales increase, the rate of drowning deaths increases sharply. Therefore, ice cream consumption causes drowning.
- Explain how correlation does not imply causation in the above example.
- Cause-and-effect relationships are generally specified by long-term experimental studies in which subjects are followed over a period of time.

Developmental Designs

- In situations wherein a researcher wants to study how a particular characteristics changes with time, they use one of two developmental designs: Cross-sectional study or a longitudinal study.
- In **cross-sectional study**, people from different age groups are sampled and compared. Eg. A gerontologist might investigate how retired people in their 70s, 80s, and 90s tend to spend their leisure time.
- In a **longitudinal study**, a single group of people is followed over the course of several months or years, and data related to the characteristics under investigation are collected at various times.
- Eg. A psycholinguist might examine how children's spoken language changes between 6 months and 5 years of age.
- Another example: an educational researcher wants to explore relationship between academic achievement and socio-economic status. Recruits fifth graders and follows them for 10 years.
- Correlational studies are easier and more expedient to conduct than longitudinal studies.
- Loss to follow up in longitudinal studies.

Disadvantages of correlational studies

- Example: Imagine a study which sets to explore the critical thinking ability trend with age.
- Samples people aged 20 years and 70 years and measures their ability to critically analyse various scenarios.
- Suppose the critical thinking ability of 20 year olds is more than 70 year olds.
- What does this mean?
- Does the critical thinking ability declines with age?
- What are alternative explanations to refute the above claim?
- *Internal validity* issues in a cross-sectional study.
- *External validity* : ability to generalize the findings of the study. Cohort studies with longitudinal component are more relevant.

Cohort-sequential study

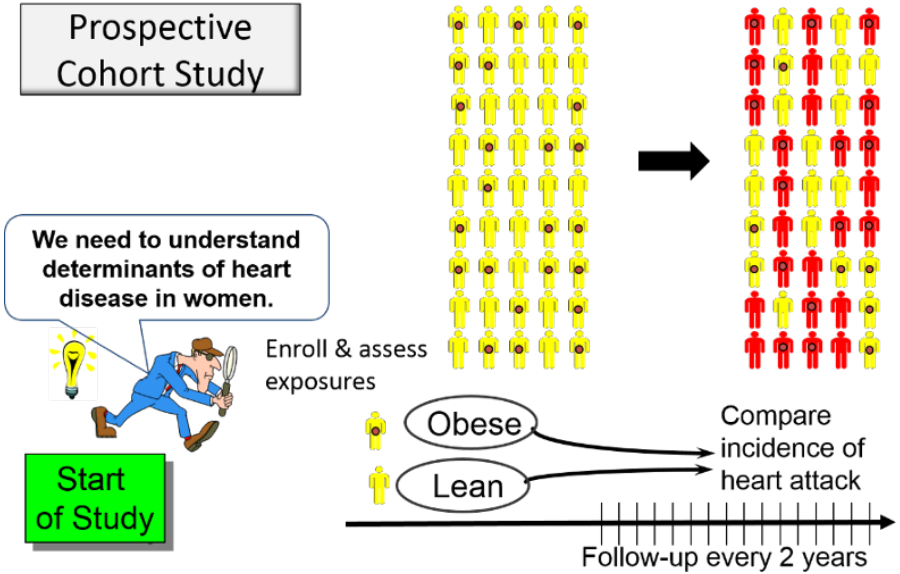

Figure 8: Diagram explaining cohort-sequential/cohort study

Experience Sampling Methods

- An approach in which a researcher collects frequent and ongoing data about people as they live their normal, everyday lives.
- Eg. Data collected by wearables, or mobile-based heart rate sensors
- Checking how the number of steps covered by people might be associated with their sleep patterns.
- Can be useful if researcher wants to collect longitudinal data to investigate any short-term *changes* in a characteristic as environmental o behavioural variables change.
- ESM can be combined with any methodology discussed above.

Example

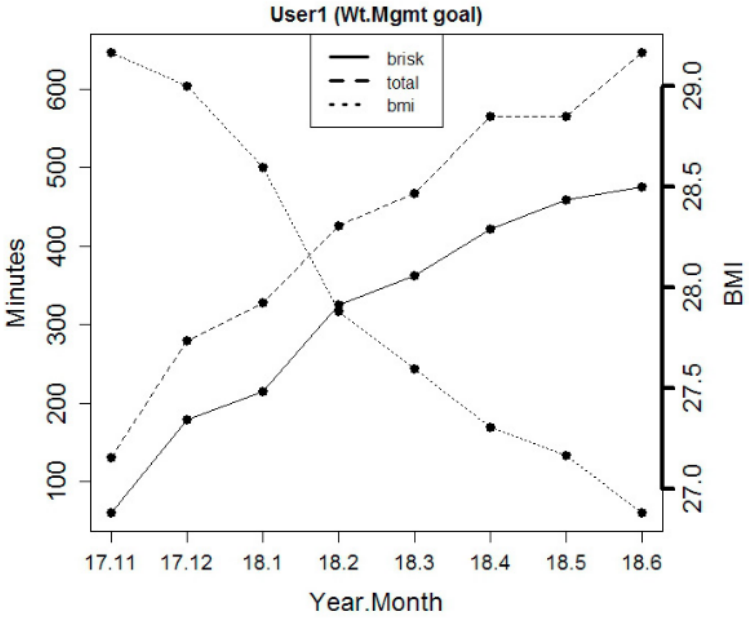

Figure 9: Scatter plot showing the trend in BMI and walking minutes for a user of smart health app over a duration. Source: (Bhargava and Nabi, 2020)

Survey Research

- A researcher obtains information about one or more groups of people-about their behaviours, opinions, attitudes or previous experiences.
- Involves collecting data about a sample of individuals, presumed to represent a much larger population.
- Survey research captures a fleeting moment in time. We must be cautious about generalizing the results for a longer time period.
- *There is nothing permanent but change. -Heraclitus.*
- Survey research typically employs: a face-to-face interview, a telephone interview, or a written questionnaire.

Face-to-face and Telephone interviews

- Standardized Quantitative Survey Research: everyone asked same set of questions.
- Structured Interview: the researcher asks certain questions and nothing more.
- Semi-structured interview: the researcher may follow the standard questions with one or more individually tailored questions to get clarification or probe a person's reasoning. This interview will have a qualitative element as well.
- Face-to-face interviews yield the highest response rates.
- Time and expense involved may be prohibitive.
- Telephone interviews are less time consuming and often less expensive. eg. BRFSS survey conducted by CDC.
- Sample will be biased to the extent that people w/o phones are part of the population under investigation.
- Video-conferencing interviews are also quickly becoming popular.

Questionnaires

- Take time to create them and formulate lucid and uncomplicated questions.
- Low response rate.
- Responses reflect their reading and writing skills and, perhaps, misinterpretation of one or more questions.
- Plan, construct and distribute carefully.
- As with every art, it becomes better with practice.

- Next lecture: Experimental Studies.
- Every lecture will be followed by ungraded multiple-choice questions quiz to assess student understanding of advanced concepts.

References

Bhargava, Y. and Nabi, J. (2020), 'The opportunities, challenges and obligations of fitness data analytics', *Procedia Computer Science* **167**, 1354 – 1362.

Di Cicco, M. E., Ragazzo, V. and Jacinto, T. (2016), 'Mortality in relation to smoking: the british doctors study.', *Breathe (Sheffield, England)* **12**(3), 275–275.