

PPDAC: The statistical method

Corinne Riddell (instructor Alan Hubbard)

August 23, 2023

Problem

A clear statement of what we are trying to achieve.

Three main problem types

- **Descriptive:** learning about some particular attribute of a population
- **Causative/Etiologic:** do changes in an explanatory variable cause changes in a response variable?
- **Predictive:** how can we best predict the value of the response variable for an individual?

Plan

The procedures we use to carry out the study.

- **Census** or **sample** from the target population?
 - How was the sampling conducted?
- Is the study prospective or retrospective?
- Is the study observational or experimental?

Data

The data which is collected according to the **Plan**.

Analysis

The data is summarized and analysed to answer the questions posed by the **Problem**.

Conclusion

Conclusions are drawn about what has been learned about answering the **Problem**.

PPDAC Example: A vaping behavior study

PPDAC Example

Problem: Suppose we wish to study the vaping behavior of California residents aged 14-20 years. In particular, we are interested in the prevalence of current vaping by gender.

PPDAC Example

Plan: We need to first choose a time period, because we know that vaping behavior has changed immensely over time. It is infeasible to gather these data for all residents in California who are 14-20 years old. Instead we conduct a random sample of size n persons. We collect their age, gender, and vaping status.

Note that we need to decide how large n should be, and how to obtain the random sample. The latter question is very important if we want to ensure that our sample is representative of the population of interest. Time and money also constrain how the sample will be collected.

PPDAC Example

Data: Suppose that a random sample of 200 persons aged 14-20 was selected, yielding these data:

Gender	Number of vapers	Number of non-vapers	Total
Teen girls and women	32	66	98
Teen boys and men	27	75	102
Total	59	141	200

PPDAC Example

Analysis: The proportion of women in the sample who vape is $32/98 = 33\%$. The proportion of men in the sample who vape is $27/102 = 26\%$.

We would also like some idea as to how close this estimate is likely to be from the actual proportion in the population. If we selected a second random sample of the same size, we would likely estimate different proportions for men and women. We will learn how to estimate the precision of these estimates.

PPDAC Example

Conclusion: 33% of girls and women aged 14-20 and 26% of boys and men of the same age group are current vapers in California in 2018 (plus a measure of uncertainty).

Reference

The **PPDAC** method is described based on course notes from STAT 231 from the University of Waterloo (Ontario, Canada). Spring 2006 Course Packet.