To Do:

Read Chapter 3.
Do end-of-chapter problems 1-7.

See instructor and TA office hours posted on Learn.

Piazza will be monitored by TAs and instructors during the study break.

The Steps of PPDAC

Problem: a clear statement of the study's objectives.

Plan: the procedures used to carry out the study including how the data will be collected.

Data: the physical collection of the data, as described in the Plan.

Analysis: the analysis of the data collected in light of the Problem and the Plan.

Conclusion: The conclusions that are drawn about the Problem and their limitations.

The Problem - Target Population or Target Process

Definition:

The target population or target process is the collection of units to which the experimenters conducting the empirical study wish the conclusions to apply.

Example: Vitamin D and Influenza A

In the vitamin D/influenza A study the units were children.

The target population is all children aged 6-15 in Japan at the time the study was conducted.

Example: Vitamin D and Influenza A

In the vitamin D/influenza A study, for each child (unit) in the target population, the variates of primary interest are whether the child received vitamin D or the placebo and whether or not the child contracted influenza A.

Other variates of interest defined for each unit were age, sex, and height.

The Problem: Attributes

Definition:

An attribute is a function of the variates over a population.

In the Problem step the questions of interest are specified in terms of the attributes of the target population.

Example: Vitamin D and Influenza A

The attributes of the target population of most interest in the vitamin D/influenza A example are:

1) the proportion of children in the target population who would contract influenza A if they took vitamin D

2) the proportion of children in the target population who would contract influenza A if they took the placebo.

The Problem - Types

The types of problems that an empirical study are designed to study can usually be categorized as one of three types:

Descriptive: to determine a particular attribute of the population (e.g. the national unemployment rate).

Causative: to determine the existence or nonexistence of a causal relationship between two variates (e.g. does a new hockey helmet reduce the risk of concussion).

Predictive: to predict the response of a variate for a given unit (e.g. predict e-cigarette weekly sales if the sales tax on them is doubled).

Example: Vitamin D and Influenza A

The problem of interest was whether taking vitamin D would reduce the proportion of school aged children in Japan who contract influenza A.

This is a causative problem.

Steps of PPDAC - The Plan

The purpose of the Plan step is to decide what units are available for study, what units will be examined, and what variates will be collected and how.

The Plan - Study Population or Study Process

Definition:

The study population or study process is the collection of units available to be included in the study.

Example: Vitamin D and Influenza A

The study population is all children aged 6-15 in Japan who attended the outpatient clinics during the time the study was conducted. (The accrual period was November 1 2008 to December 15 2009.)

The Plan - Study Error

Definition:

If the attributes in the study population differ from the attributes in the target population then the difference is called study error.

The Plan - Study Error

Since we do not know the values of the target population attributes or the study population attributes, the study error cannot be quantified.

In most cases we rely on expertise from other sources to determine whether conclusions derived from the study population (e.g. mice) may apply to the target population (e.g. humans).

Example of Study Error

Suppose the target population is all adults living in Canada and the study population is a list of people defined by their telephone number.

The study population excludes those people without telephones or with unlisted numbers.

If people without telephones are systematically different from people with telephones with respect to the variate of interest then their would be study error.

Example: Vitamin D and Influenza A

What is a possible source of study error in the Vitamin D/Influenza A study?

Example: Vitamin D and Influenza A

The target population was schoolchildren in Japan and the study population was schoolchildren who visited certain hospitals and certain doctors.

It could be that children who did not visit these hospitals and doctors are healthier and have a natural immunity to influenza A so taking Vitamin D would make no difference.

The Plan - Study Population or Study Process

Definition:

The sampling protocol is the procedure used to select a sample of units from the study population.

The number of units sampled is called the sample size.

The Plan - Sampling Protocol

Obtaining a “random sample” in practice is often very difficult and expensive and therefore other less rigorous sampling methods are used. (e.g. Canadian National Housing Survey).

The Plan - Sampling Protocol

In many surveys, the study population is a list of people defined by their telephone number and the sample is selected by calling a subset of the telephone numbers.

Sample size is usually a compromise between cost, availability and desired precision as calculated using a model.

Example: Vitamin D and Influenza A

What was the sampling protocol in the Vitamin D/Influenza study?

What was the sample size?

Example: Vitamin D and Influenza A

Schoolchildren aged 6-15 were asked to participate in the study by doctors at 12 hospitals and 8 doctors during the accrual period of November 1 2008 to December 15 2009.

Parents and children had to give written consent. The children had to be able to swallow pills and not have certain allergies, etc.

Example: Vitamin D and Influenza A

Sample size is a bit tricky.

430 children were recruited but only 334 were followed to the end of the study.

The Plan - Sample Error

Definition:

If the attributes in the sample differ from the attributes in the study population then the difference is called sample error.

The Plan - Sample Error

The sample is only a subset of the units in the study population.

Different sampling protocols are likely to produce different sample errors.

Some protocols tend to have smaller errors than others.

The values of the study population attributes are unknown, so the sample error is unknown.

The model is used to quantify how large this error might be.

Example: Vitamin D and Influenza A

What is a possible source of sample error in the Vitamin D/Influenza A study?

Example: Vitamin D and Influenza A

The study population was schoolchildren who attended certain hospitals and certain doctors.

The sample consisted of children whose parents gave written consent, who could swallow pills, did not have allergies, etc.

It could be that children whose parents did not give written consent are systematically different than children of parents who do.

The Plan - What to Measure?

We measure the variates corresponding to any attributes of interest defined in the Problem step for the units in the sample.

We may also measure other variates that can aid the analysis.

In the vitamin D/influenza A study, the investigators also collected information on age, sex, and height.

The Plan - What to Measure?

When the value of a variate is determined for a given unit, errors are often introduced by the measurement system which determines the value.

The observed value and the "true" value are usually not identical.

Example: Vitamin D and Influenza A

How was the variate of main interest in the Vitamin D/Influenza study measured?

Example: Vitamin D and Influenza A

How was the variate of main interest in the Vitamin D/Influenza study measured?

A rapid influenza diagnostic test (RIDT) with a nose swab following the manufacture's protocol was used.

The Plan - Measurement Error

Definition:

If the measured value and the true value of a variate are not identical the difference is called measurement error.

Example: Vitamin D and Influenza A

What is a possible source of measurement error in the Vitamin D/influenza A study?

Example: Vitamin D and Influenza A

The RIDT might give false positives or false negatives.

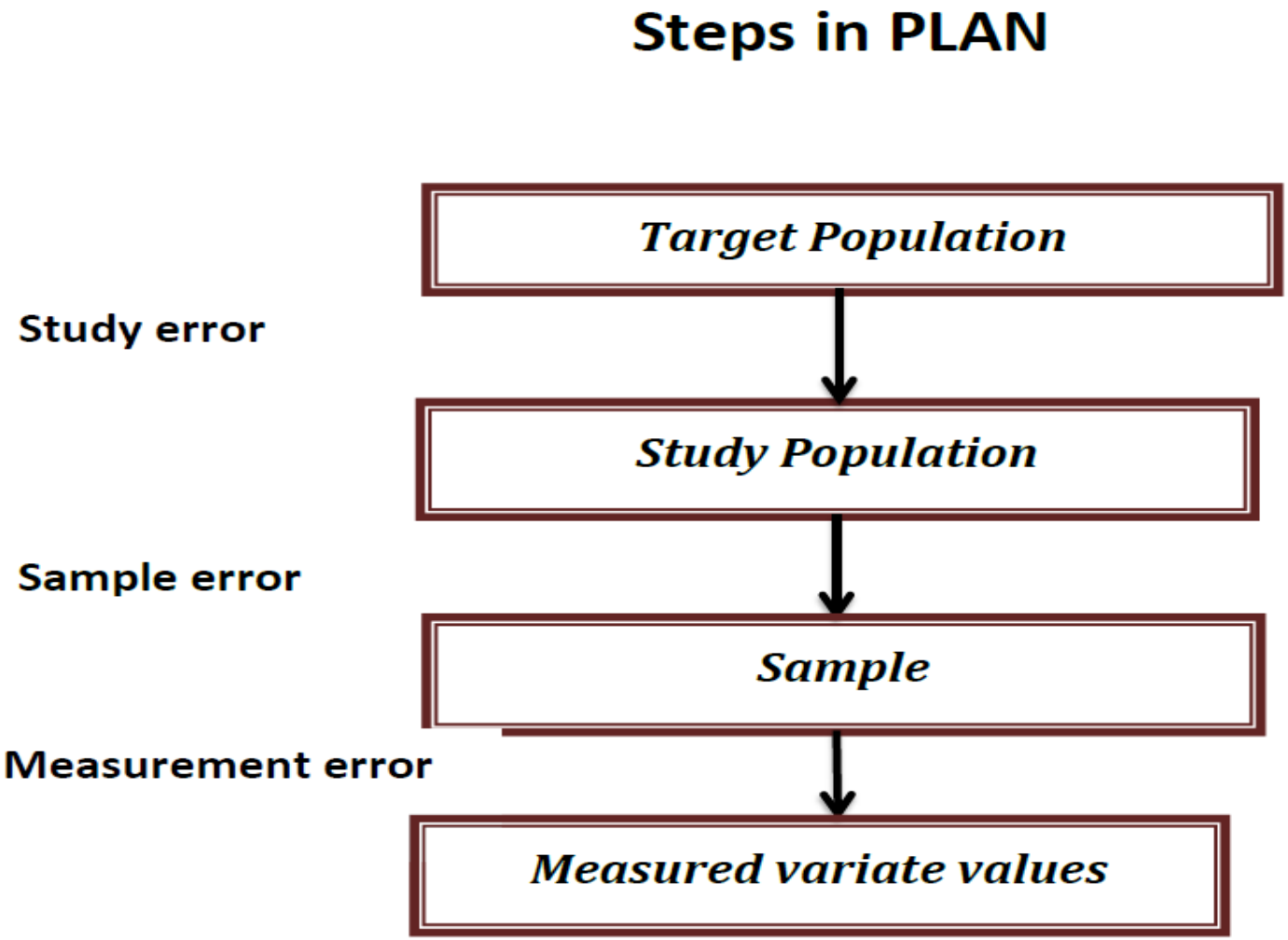
The RIDT might be administered incorrectly. (Is there a correct way to do a nose swab?)

The Plan – Mearurement Error

The measurement systems may have to be studied separately (using their own PPDAC) to determine the magnitude of the measurement error.

See Examples in Case Study in Section 3.3 of Course Notes.

Steps in the Plan



The Plan

For an empirical study the Plan should indicate the study population, the sampling protocol, the variates which are to be measured, and the quality of the measurement systems that are intended for use.

Attention must be paid to the various types of error that may occur and how they might impact the conclusions.

The Plan

You will be asked to use PPDAC to critically examine the Conclusions from a study done by someone else.

You should examine each step in the Plan for strengths and weaknesses.

Example

The Toronto Police Department wants to know if black residents of Toronto are satisfied with police service in their neighbourhood. A questionnaire is prepared. A sample of 300 mailing addresses in a predominantly black neighbourhood is chosen and a uniformed police officer is sent to each address to interview an adult resident.

Example

The Toronto Police Department wants to know if black residents of Toronto are satisfied with police service in their neighbourhood. A questionnaire is prepared. A sample of 300 mailing addresses in a predominantly black neighbourhood is chosen and a uniformed police officer is sent to each address to interview an adult resident.

This is an example of what is called response bias.

Example

A Waterloo-based public opinion research firm was hired by the Ontario Ministry of Education to investigate whether the financial worries of Ontario university students varied by sex. Since the research firm was based in Waterloo it decided to study university students living in the Kitchener-Waterloo region in September 2016. An associate with the research firm randomly selected 250 university students attending a Laurier-Waterloo football game. The students were asked whether they agreed/disagreed with the statement "I have significant trouble paying my bills." Their sex was also recorded.

Data Step

The purpose of the Data step is to collect the data according to the Plan.

In order to do this the variates must be clearly defined and satisfactory methods of measuring them must be used.

Data Step

Mistakes can occur in recording or entering data into a data base.

For complex investigations, it is useful to put checks in place to avoid these mistakes, and detect those that are made.

Data Step

In many studies the units must be tracked and measured over a long period of time (e.g. consider a study examining the ability of aspirin to reduce strokes in which persons are followed for 5 years).

When data are recorded over time or in different locations, the time and place for each measurement should be recorded.

Data Step

Departures from the Plan may arise over time (e.g. persons may drop out of a long term medical study because of adverse reactions to a treatment). Such departures should be recorded.

These departures will also impact the Analysis and Conclusion.

Example: Vitamin D and Influenza A

Remember in the Vitamin D/Influenza study 430 children were recruited but only 334 were followed to the end of the study.

This is an example of missing or incomplete data.

The analysis of the data and the conclusions of the study should take this into account.

Data Step

When the amount of data is large, data base design and management is important.

Analysis Step

The Analysis step consists of the analyses of the data collected.

Analysis Step

The Analysis step should include numerical and graphical summaries of the data.

A key part of the analysis is the selection of an appropriate model that describes the data and how the data were collected.

Checking the fit of the model must be included.

Analysis Step

To answer the questions posed in the Problem step we usually formulate these questions in terms of the model parameters.

In the remainder of this course we will see many examples of formal analyses (estimation, tests of hypotheses).

Departures from the Plan that affect the Analysis must be noted.

Conclusion Step

In the Conclusion step, the questions posed in the Problem are answered to the extent permitted by the data. In other words the Conclusion step is directed by the Problem.

Potential study, sample or measurement errors, as described in the Plan step, should be discussed and quantified if possible.

Conclusion Step

Departures from the Plan that affect the Analysis must be addressed.

The limitations of the study must also be described.

TO DO!

Read on your own:

Section 3.3: Case Study