# MATH1. Part II

## Probability and Statistics

# Chapter 13

## A Summary of Inference Methods

# 13.1 Introduction

When presented with a set of data, it is useful to ask a series of questions:

1. What question were the researchers attempting to answer when they collected these data?
2. What is the response variable in the study?
3. What predictor variables, if any, were involved?

# 13.1 Introduction

1. What **question** were the researchers attempting to answer when they collected these data?

- Data analysis is done for a **purpose**: to extract information and to aid decision making. When looking at data, it helps to bear in mind the purpose for which the data were collected.

  For example,

  – Were the researchers trying to compare groups, perhaps patients given a new drug and patients given a placebo?

  – Were they trying to see how two quantitative variables are related so that they can use one variable to make predictions of the other?

  – Were they checking whether a hypothesized model gives accurate predictions of the probabilities associated with a categorical variable?

- A good understanding of why the data were collected often clarifies the next question:

# 13.1 Introduction

2. What is the **response variable** in the study?

- **Type** of the data (Y) that being collected.

  For example,

  – If the researchers were concerned with the effect of a medication on blood pressure, then the likely response variable is Y = change in blood pressure of an individual (a continuous numeric variable).

  – If they were concerned with whether or not a medication cures an illness, then the response variable is categorical with two levels: yes if a person is cured, no if a person is not cured, or maybe even categorical with three or more ordered levels: fully cured, improved, no improvement.

# 13.1 Introduction

3. What **predictor variables**, if any, were involved?

- **Type** of the data (X) that being collected.

  For example,

  – If a new drug is being compared to a placebo, then the predictor variable is group membership: a patient is either in the group that gets the new drug or else the patient is in the placebo group.

  – If height is used to predict weight, then height is the predictor (and weight is the response variable).

  Sometimes there is no predictor variable.

  – For example, a researcher might be interested in the distribution of cholesterol levels in adults. In this case, the response variable is cholesterol.

# 13.1 Introduction

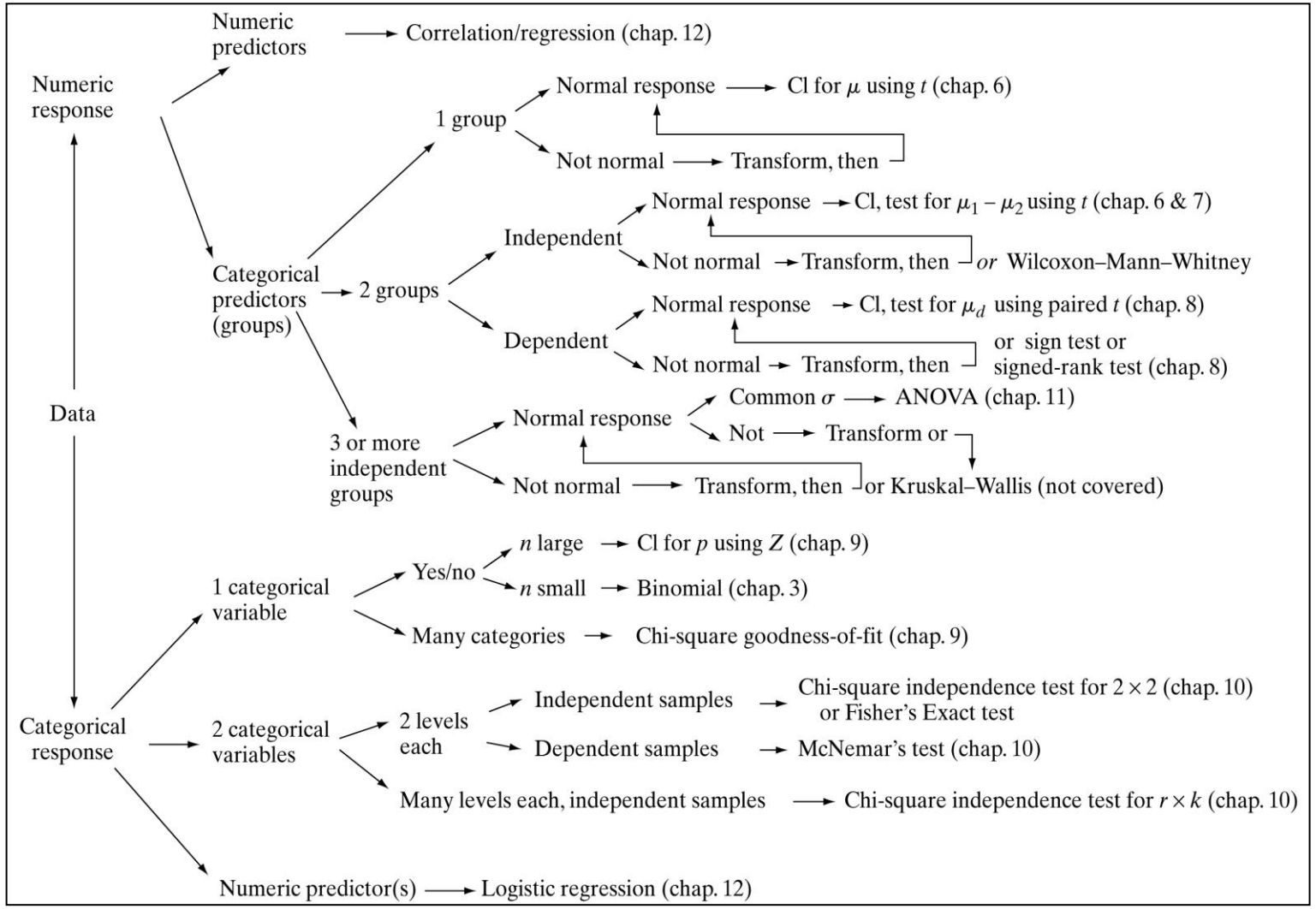


**Figure 13.1.1** A flowchart of inference methods

# Summary

## Chapter 13. A Summary of Inference Methods

- 13.1 Introduction