STAT 241
Personal Notes

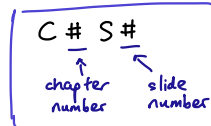* These notes are <u>strictly</u> my own interpretation of the course materials.

Marcus Chan

Taught by Michael Wallace

UW CS '25

Chapter 1: Introduction to Statistical Science

💡 Statistical science is the science of "empirical studies".

EMPIRICAL STUDY (C1S24)

💡₁ An "empirical study" is one where we learn by observation and/or experimentation.

💡₂ Note these involve uncertainty – repeated experiments generate different results.

💡₃ But we model these uncertainties using probability models.

UNIT (C1S25)

💡 A "unit" is an individual which we can take measurement(s).

POPULATION (C1S26)

💡 A "population" is a collection of units.

eg – all current UW undergrad students
– all donuts in Tim Hortons right now

*note: we need to be precise when defining populations, or any other terms!

eg if we said "all UW students" this is ambiguous, since it might include grads, alumni, etc

PROCESS (C1S27)

💡₁ A "process" is a system by which units are produced.

eg – hits on a particular website are units in a process
– claims made by insurance policy holders are units in a process

💡₂ Note that although populations & processes are collections of units:

① Populations are "static" (defined at one point in time), but

② Processes usually occur over time.

VARIATES (C1S32)

💡 "Variates" are characteristics of the units.

* we usually represent these by letters x, y & z.

CONTINUOUS VARIATES (C1S33)

💡 "Continuous variates" are those that can be measured (at least theoretically) to an infinite degree of accuracy.

eg height, weight, lifetime of a fuse, etc

DISCRETE VARIATES (C1S33)

💡₁ "Discrete variates" are those that can only take finitely or countably many values.

eg # of car accidents on a certain stretch of highway / yr, etc.

💡₂ Note that depending on how we measure a continuous variate, it may become discrete.

eg if we measure weight w/ a scale that only goes to 2dp, the resulting variate is discrete!

💡₃ Ultimately the distinction affects
① our assumptions of the data; and
② the probability models we use.

· for discrete variates, we usually use discrete prob models (eg Poisson)
· for cts variates, we usually use cts prob models (eg Gaussian)
· but there are exceptions. (C1S43)

CATEGORICAL VARIATES (C1S35)

💡 "Categorical variates" are those where the units fall into non-numeric categories, without any implied order.

eg hair color, university program

ORDINAL VARIATES (C1S35)

💡 "Ordinal variates" are those where an ordering is implied, but not necessarily from a numeric measure.

eg strongly disagree, ..., strongly agree;
small, medium, large;
etc

COMPLEX VARIATES (C1S37)

💡₁ "Complex variates" are those that are more unusual, and don't fall neatly into the other variate types.

eg open-ended responses to a survey question

💡₂ We usually need processing to convert these into one of the other types.

eg text processing to convert a tweet's content into "positive", "negative" or "neutral"

ATTRIBUTES [OF A POPULATION/PROCESS] (C1S48)

💡 "Attributes" of a population/process are functions of a variate which is defined for all units in said population/process.

eg (STAT 231 asmts) — mean # of completed asmts
— prop. of asmts subbed in last 24 hrs

(KW Humane Society) — prop. of dogs that arrive in good health
— mean # of owners of dogs in their care

TYPES OF EMPIRICAL STUDIES (C1S50)

SAMPLE SURVEY (C1S52)

💡 A "sample survey" is where information is obtained about a finite population by
① selecting a "representative" sample of units from the population; and
② determining the variates of interest for each unit in the sample.

eg — poll to predict who will win an election
— survey of potential consumers to compare products & state their preference (eg Coke vs Pepsi)

OBSERVATIONAL STUDY (C1S53)

💡₁ An "observational study" is where information about a population/process is collected without any change to the sampled units' variates.

eg a study of blood alcohol levels for students at a 8:30am Mon lecture

💡₂ Usually, the following are true:

Observational	Survey
① Pop^n of interest is infinite/conceptual	Pop^n is finite/real
② Data collected routinely over time	Data collected once
③ More passive (sit and see)	More "aggressive" (specific questions asked)

* but these are just guidelines — there are exceptions. (C1S55)

EXPERIMENTAL STUDY (C1S54)

💡 An "experimental study" is one where the experimenter intervenes and modifies some of the variates for the units in a study.

eg same example as above, but some students are warned beforehand, whereas some are not.