Lecture ①

Variable:

a measurement that change from individual to another, or from time to time

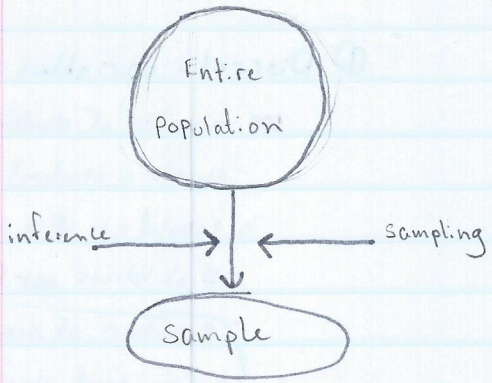
Example: Salary in Ontario (anual)

	Salary 2014	Salary 2015
John	$60 000 ↑ measurement	$65000
Phil	$53000 ↑ measurement	$42000
Ann	$120 000 ↑ measurement	$150 000

Variable

experimental Value

Sampling:



in sampling, we try to obtain some relevant information from the sample at hand to infer about the whole population.

why we need sampling (or statistics in general?)

Example 1: we can't measure body tempurature of every person on earth. Instead, we take a sample of say 100 healthy persons (from different places) and we measure the average, say 37°, then we conclude (using some math) that 95% of all healthy people have their tempurature between $37° \pm .3$ ← margine of error

Example 2: If we want to know time to death for a mouse exposed to a gamma radiation we'll have to wait until all mice die!!! (Not possible). Instead, we can observe, say 25 mice and wait number of hour to death for each mouse then we compute the average

If we find it for example, equal 32.3 hours

$$\frac{T_1 + \cdots T_{25}}{25} = 32.3\,h$$

Then we make a math free statement such as:

95% of mice die after 32.3hr with .5 margin of error

Example 3: To know the exact time numbers, we'll have to wait until election is over. Instead, we conduct some statistical surveys.

L, C, NDP

survey 1000 potential voters....

30% 40% 30%

From this we can build what we call (c) confidence intervals (I) for each party

L: 30% ± 2%	C: 40% ± 2.5%	NDP: 30% ± 1.8%
C.I 28%, 32%	C.I 37.5%, 42.5%	C.I 28.2%, 31.8%

(1) **univariate data:**

Salary = one single variable of interest of us

(2) **bivariate data:**

in addition to salary...

ex: where do you live?

(3) **multivariate data:**

		Salary	City	degree	M/F
a_1	John	$60000	ottawa	msc	m
a_2	Phil	$53000	Guelf	H.S	m
a_3	Ann	$120000	Toronto	phd	F

Descriptive Statistics

Statistics on the sample

Example: sample mean

If we compute the sample mean for John, phil and Ann, we get

$$= \frac{60\,000 + 53\,000 + 120\,000}{3}$$

$= \$77\,666$ ← Just the mean salary of 3 people picked at random

we also have sample variance, median, etc...

Inferential Statistics

are methods leading to some information about the true (unknow) mean, for example the entire population.

using inferential stat., we can make statements such as:

average salary in ontario in 2019 is between $55000 (lower) and $75 000 (upper)

Two types of variables:

① **Qualitative:** (categorical)

② **Quantitative:**

Variables that assume numbers (or numerical values)

ex · salary
· income
· temperature
· age
· time

Two types of quantitative data

① **Discrete variables**

ex · number of students in a class
· shirts a student has
· population of a town
· # children per family
· number of accidents per week on montreal street, ottawa
↳ be specific

② **Continuous:** over an interval