Chapters 1 Class Notes – Introduction

Environmental Science Example – Open-top chamber experiments (photo from ARS website)

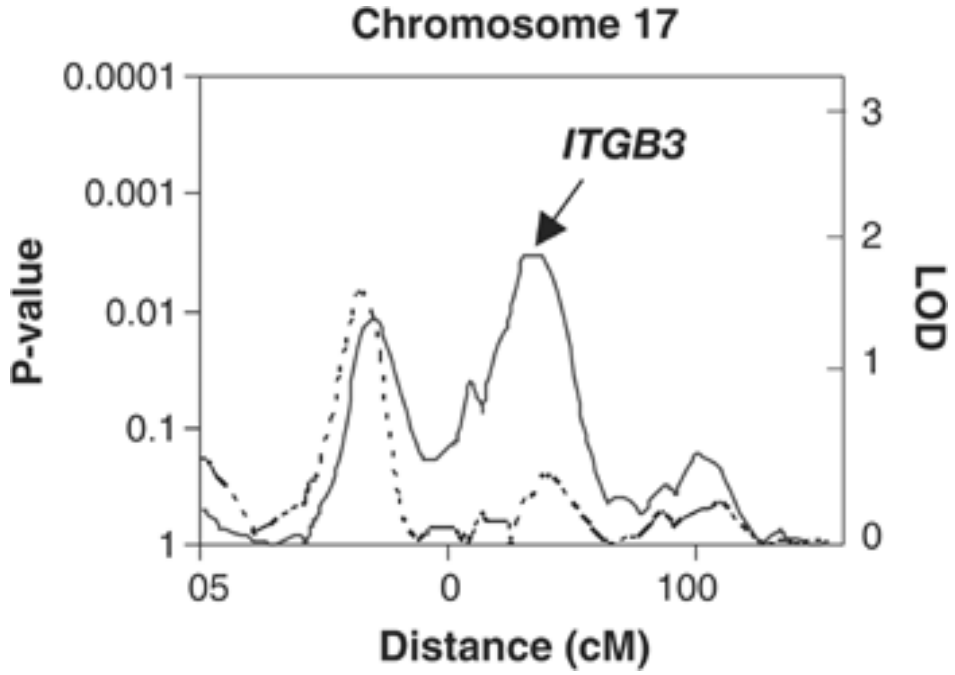
The variables here are

X = amount of ozone applied within the chamber (per day) over the course of the 10-week study;

Y = loblolly pine tree biomass (weight of organic pine tree matter) at end of study.

Both variables here are continuous (and therefore quantitative).

<u>Medical (Genetics) Illustration</u> – reference Bojesen et al (2003), "Integrin β_3 Leu33Pro Homozygosity and Risk of Cancer", *J. NCI*.



Women only 2 x 2 table:

		Outcome Status		
		With Cancer	Without Cancer	Total
Stratification	Non-carriers	501 (14.4%)	2983	3484
	Homozygotes	29 (21.5%)	106	135
	Total	530	3089	3619

The <u>variables</u> here are

X = stratification (either non-carrier or homozygote group)
Y = cancer status at end of study.

Both variables here are nominal (and therefore qualitative).

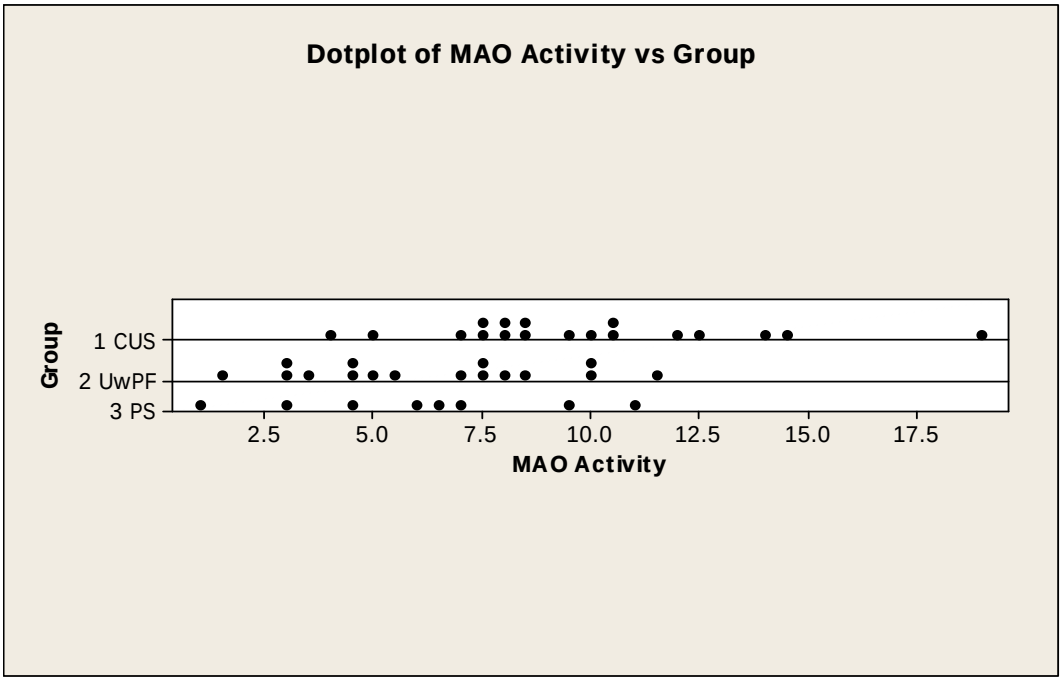
Note that the following table is somewhat different and that *the statistical analysis is very different*!

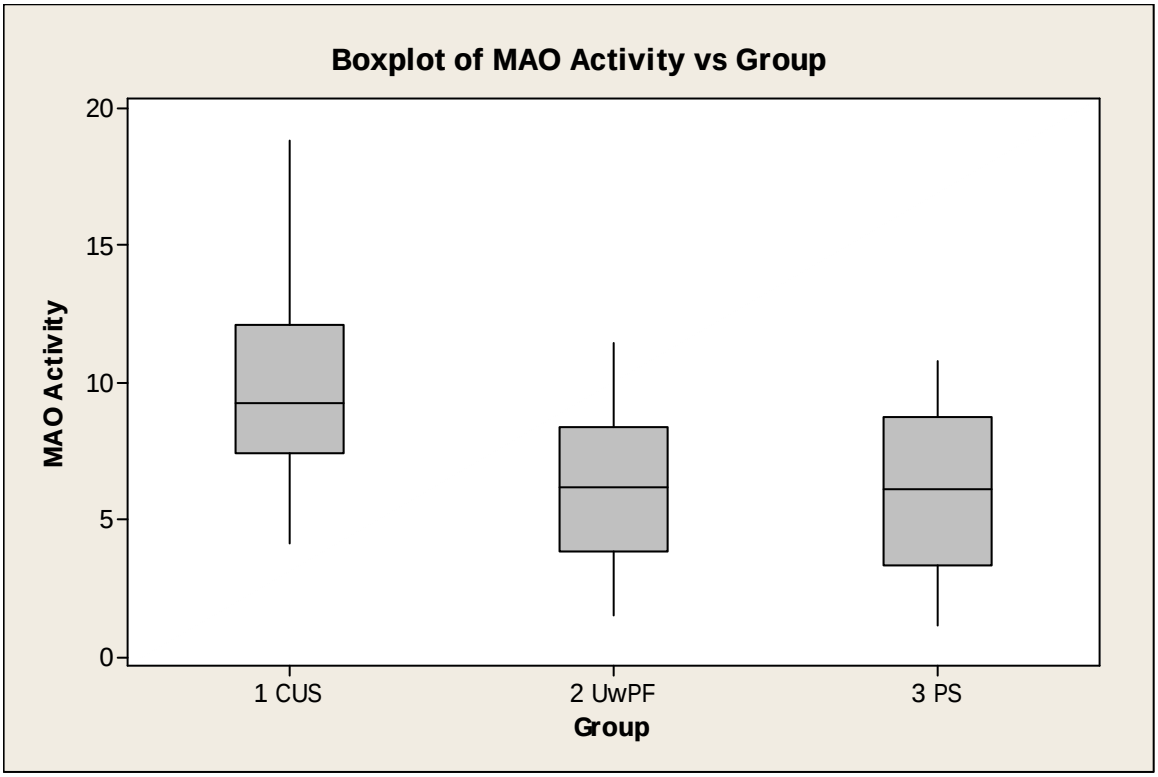
		Pain Relief				
		Complete	Substantial	Some	None	Total
Drug	A	5	10	7	3	25
	B	2	5	11	7	25
	Total	7	15	18	10	50

One of the variables here (Pain Relief) is ordinal.

<u>Text MAO and Schizophrenia Example on p.4</u>

Variables are Group (X) and MAO Activity (Y) – Groups are 1-CUS (Chronic undifferentiated schizophrenic), 2-UwPF (Undifferentiated with paranoid features), and 3-PS (paranoid schizophrenic)





Descriptive Statistics: MAO Activity

Variable	Group	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
MAO Activity	1 CUS	18	9.806	3.618	4.100	7.375	9.200	12.100	18.800
	2 UwPF	16	6.281	2.880	1.500	3.850	6.150	8.325	11.400
	3 PS	8	5.96	3.19	1.10	3.30	6.10	8.75	10.80

In §1.2 (p.7), we consider 'how' and 'when' to select individuals for a study, how to arrange experimental material in space and time, and how to allocate the experimental units (EU's) to treatment groups. An EU might be a person, a person on a specific day, the right or left hand of a person, and so on. In observational studies, the researcher merely observes –in experimental studies, the researcher intervenes with the 'treatment' or condition. Also, note the important discussion of anecdotal evidence on pp.7-8.

A Smoking and Birthweight example illustrates an observational study. Here:

- **Response variable: baby's birthweight**
- **Explanatory variable: whether or not mother smoked during pregnancy**
- **Extraneous variable(s): age, income, education, diet …**
- **Observational unit (OU): a mother-child pair**

As we encounter observational studies, we need to look for and consider possible sources of bias and then to imagine how these would impact the study and findings. We also need to ask ourselves whether the sample was truly random or simply a 'sample of convenience'.

In the Smoking study, it could well be the case that diet is *confounded* with whether or not a person smokes – so too might alcohol consumption (AC) be confounded with smoking status.

S → low birthweight

??? S ----→ low birthweight

AC

[Recent smoking studies – and all epidemiological studies in general – measure many variables (e.g., amount of coffee consumed, age at onset of menstruation, age, weight, height, blood type, religion, education, income…), and the correct statistical analysis controls for these additional variables. These studies do find a link between smoking and baby birthweight.]

But with all observational studies, remember

"Association is not causation."

Why then perform observation studies? First, may be the only way to study something like the effects of smoking (consider the alternative). Second, observational studies can be pooled or can give suggestions for subsequent experiments and research.

Spurious association: babies exposed to ultrasound in the womb were significantly lighter than babies not exposed to ultrasound. Does ultrasound cause reduced birthweight? [Of course not – the lurking variable in this study was whether or not a mother was experiencing problems during pregnancy.]

Case-Control Studies – are extremely popular in epidemiology and medicine. If we want to study the effects of diet on stomach cancer, we could identify cancer patients and corresponding control patients and then compare diets. We try to eliminate any lurking confounding variables by matching cancer and control patients. But be careful since these types of studies often involve samples of convenience – thereby rendering any conclusions possibly useless or questionable.

	Stomach Cancer	No Stomach Cancer
Good Diet		
Bad Diet		

In an experiment, the researcher intervenes and imposes treatment conditions on each EU. To remove bias (as much as possible), *randomization* is essential. For example, to study the effects of aspirin on relieving headache pain, we could conduct:

- **Experiment 1 – aspirin versus ibuprofen**
- **Experiment 2 – aspirin versus nothing**
- **Experiment 3 – aspirin versus placebo**

But each of these experiments has different goals.

The need for a placebo group: The word *placebo* means "I shall please", and the placebo effect is 'curing' oneself simply by psychological means. Very similar are so-called 'sham treatment' often used by medical researchers (for example, injecting control animals with an inert substance such as saline solution). Also paramount is blinding (one's ignorance of which treatment has been applied), and when this applies to both MD and patient, it is called 'double blinding'.

Example 1.2.8 on p.12 demonstrates the need for a control group in experiments. The active substance here is Clofibrate, which is intended to lower cholesterol and thereby hopefully lower the chance of death from CHD. Here are the mortality rate data:

	Clofibrate		Placebo	
Adherence	n	5-yr mort. rate	n	5-yr mort. rate
≥ 80%	708	15.0%	1813	15.1%
< 80%	357	24.6%	882	28.2%

Without the control group, we'd just have the data above on the left, and conclude that Clofibrate lowers the mortality rate. But with the inclusion of the control group data, we see no difference between Clofibrate and Control – only that those that adhere to the protocol instructions differ from those that do not (i.e., there are two populations).

Historical controls can be used: results of AZT today with nothing yesterday, but these results are suspect since current-day patients usually show a better response due only to improved conditions.

Random Sample (of size n; denoted 'SRS') from a population (of size N): every member of the population has the *same chance of being chosen* and each member is *chosen independently*.
In practice, we can use the computer (e.g., in MTB, use CALC → Random Data) or use Table I on pp. 611-14 to help get a SRS.

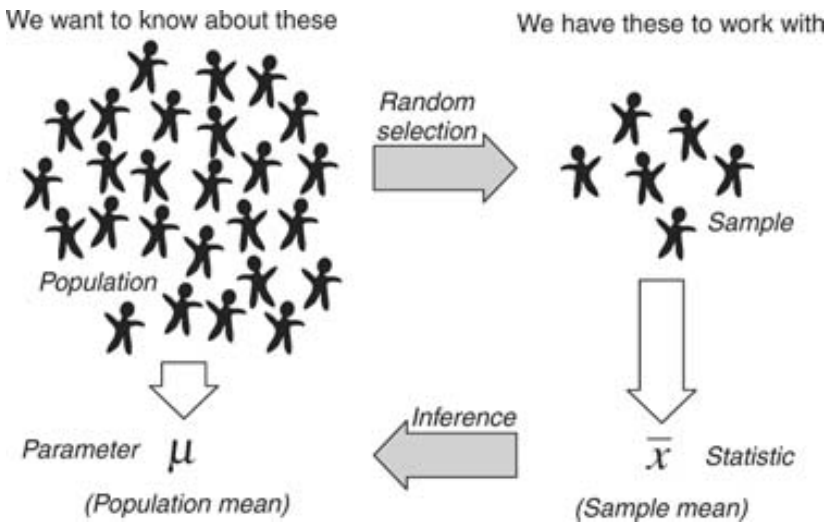
Why randomize?

→ To remove, or at least lessen, bias

Chance Error due to Sampling – you and I both choose random samples of size n from a given population and measure some quantitative variable. You calculate $\bar{y}_1$ and I calculate $\bar{y}_2$. Do you expect that $\bar{y}_1 = \bar{y}_2$? (More on this later in Chapter 5.)

As we hear about studies comparing two treatments or groups, we should (1) ask how the patients were selected, (2) note which variable(s) were measured, and (3) identify the two populations.

Carefully read §1.3 on Random Sampling (pp.15-23). To be discussed again in Chapter 2: the following diagram is helpful to understand an important application of Statistical Science called Statistical Inference. A representative *Sample* is taken from the *Population* (so as to remove any *Bias*), and the sample *Statistic* (such as sample mean) is used to estimate the population *Parameter*.



For example, a selected group of n = 25 Loyola undergraduates had an average age of $\bar{y}$ = 19.4 and $\hat{p}$ = 64.0% were Female. Making statements about *all* Loyola undergraduates falls under the heading of 'statistical inference'. It's important to consider if the sample is representative and consider whether the sampling methodology may result in any bias. Sometimes defining the population can be challenging. For example, if the sample was chosen "at random" by knocking on doors in the student dormitories, these issues may be very important.