ALQUSOUR ACADEMY
أكاديمية القصور

Q.A.J.U.S.T

# BIOSTATISTICS

## لطلبة الصيدلة والعلوم الطبية

ALQUSOUR ACADEMY

خاص للفصل الدراسي الأول 2016/2015

## Subject:

## Second Exam – Part One

Prod. Date: **3/12/2015** Pages: **9** Price: **30**

ساعـــات الـــدوام الرسمـــي

السبت - الخميس: 11:00 ظهراً - 12:00 ليلاً

الجمعـــــــة: 2:00 ظهراً - 12:00 ليلاً

**تحذير: محاضراتنا (الملخصات) متوفرة لدى أكاديمية القصور .**


اربد / بجانب اربيلا مول / سلام سنتر / الطابق الرابع

للإستفسار والتسجيل: 0785 70 60 08 / 0795 33 99 34

Like facebook.com/alqusouracademy.page

# Descriptive Statistics

## Collecting, Organizing and Presentation of Data

The first step is to collect data:

## Methods of data collection:

1) **Using available information**:
   - It is a secondary data analysis because it has been collected by somebody else not the investigator.
   - Its sources:
     - Health information system data.
     - Census data.
     - Unpublished reports.
     - Publications of archives, libraries or offices.
     - The study itself.

| Using Available Information | |
|---|---|
| Advantages | Disadvantages |
| * Inexpensive. | * Data are not always easy accessible .<br>Ex. Maybe the hospital manager only can look at records so we cannot have it. |
| * Permits examination of Past trends.<br>Ex. If we measure growth rate in different years it will differ , so to judge it we compare growth rate this year with past years.<br>(we can judge using new data collection method in the same way). | * Ethical issues regarding confidentiality may arise.<br>It is the patient right to keep his information. |
| | * Information may be incomplete and inaccurate.<br>(We did not collect it so not sure about its accuracy). |

Available information is not trusted because it is collected by different people and different methods.

**2) Observing:**

It involves systematic selection, watching and recording behavior and characteristics of living being or objects.

Observing could be:

a- Participant observation: ex, observing students behavior in class.

b- Non-participant observation: (from samples)
Ex. Analyze blood sample in lab, so it is working on samples.
- Measurements are called observations.
- Measurements need tools (ex. Scale or balance for weight).

**Important:** Observing is primary source of information.

| Advantages | Disadvantages |
|---|---|
| * More detailed, more accurate information. | * Ethical issues of privacy and confidentiality. (Patient has the right of privacy so me as observer will affect this). |
| * Collection of information not written in questionnaires. | * Observer Bias. Ex, if you believe that males are smarter than females this will affect your results unconsciously. |
| * Test in validity of responses to questionnaires. (in observation we see the reaction to questions so answers will be more accurate). . | * The presence of the observer can influence the situation. (people will work better in the presence of observer and this will affect results) |
| | * Extensive training of assistants is needed. This will need time and money. |

**3) Interviewing:**

- It is oral questions of respondents, either individually or as a group, face to face or over phone.
- degree of flexibility in interview depends on the level of researcher understanding of the problem.

| Interviewing | |
|---|---|
| Advantages | Disadvantages |
| * Suitable for illiterate. | * The presence of the interviewer can influence the responses. (interviewer bias) *Not honest for sensitive topics. |
| * Permits clarification by respondents. | |
| * Higher response rate than questionnaires. (usually answers all questions cannot ignore any question) | * time consuming. * less complete information compared to observation. |

**4) Written questionnaires:**

- It is written questions answered by respondents.
- Maybe distributed by mail or group.

**Administering Questionnaires**

| Advantages | Disadvantages |
|---|---|
| * Less expensive, less time. | * Can not be used with illiterate individuals. |
| * permits anonymity (I don't know who fill it) and more honest responses. | * No response rate could be high. (respondents may not answer all questions) |
| * No need for assistants. | * Questions may be misunderstood. |
| * Eliminates observer bias. | |

---

## Terminology used in sample surveys:

**Element:**

- Is the entity on which data are collected.

**Sampling units**: are groups of elements or elements themselves.

**Frame**: list of sampling units (all individuals) who is going to participate in the study.

Example: If I want to compare second year students with third year students, then my frame will include students in both years.

**Population (N):** the largest collection of entities for which there is an interest at a particular time.

- Contains all possible observations.

Example: Jordan.

**Sample (n):** is simply a part of population.

**Target Population:** Large group about which an investigator wishes to draw a conclusion (want to make inferences about).

**Sampled Population:** A subset of individuals that end up participating in your study? (those actually selected)

---

## Types of Surveys:

**(1) Surveys Involving Questionnaires:**

- Includes three common types:
  mail surveys, telephone surveys, and personal interview surveys.

- Survey cost are lower for mail and telephone surveys.

- With well-trained interviewers, higher response rates and longer questionnaires are possible with personal interviews. Personal interviews are more accurate.

**(2) Surveys Not Involving Questionnaires:**

a. Often, someone simply counts or measures the sampled items and records the results.
b. An example is sampling a company's inventory of parts to estimate the total inventory value or taking measurement of weight for the class.

---

To collect data we first have to select sample.

## Sampling methods:

- **The two categories of sampling methods are:**

(1) **Probabilistic sampling:** It involves ***random*** selection.

- Here, the probability of obtaining each possible sample can be computed.
- Probabilistic sampling includes the following types:

**1) Simple random sample:**

- We begin by developing a list (frame) of all elements in the population.
- Then a selection procedure, based on the use of random numbers (using random tables or computer).
- Is used to ensure that each element in the sampled population has the same probability of being selected.

**2) Stratified sampling :**

- Used for heterogeneous groups.
- The population is first divided into groups, called strata.
- Then for stratum, a simple random sample is selected.

- The data from the simple random samples are combined to develop an estimate of a population parameter.
- If the variability within each stratum is small, the sample will be more precise.

**3) Systematic random sampling:**

- Suitable to be used when we do not have a list of population. (example: emergency patients)
- Number the units in the population from 1 to N (estimate N).
- Decide on the n (sample size) that you want or need.
- K= N/n = the interval size.
- Randomly select an integer between 1 to k.
- Then take every $k^{th}$ unit.

Example:

If the population (N) is 1000 and I want a sample of 100, then the interval between persons selected is 1000/100 = 10 . then every 10 names ($10^{th}$) we will select one person to be included in the sample.

**4) Cluster sampling:**

- Suitable in surveys.
- The frame is list of areas or clusters.
- It tends to provide better results than stratified sampling when the elements within the clusters are **heterogeneous**.
- A primary application of cluster sampling involves area sampling.
- Example: select sample from Irbid: South. Then select sample of streets from south Irbid, then sample of houses, then sample of persons to interview.

**5) Multi-stage cluster sampling:**

- Starts from the country (national level).
- In this case, we have three or four stages in the sampling process and we can use both stratified and simple random sampling.

Example: Select geographical area in Jordan (Irbid for example), then select schools, then select classes, then selects students to be included in the final sample.

(2) **Non-probabilistic Sampling Methods: (not random)**

- The probability of obtaining each possible sample cannot be computed.
- Statistically valid statements cannot be made about the precision of the estimates.
- Sampling cost is lower compared with Probabilistic sampling and implementation is easier.

- Methods include:

a- Convenience Sampling:

- The units included in the sample are chosen because of accessibility. (ex. Making interview in the mall is easy for the investigator)
- In some cases, convenience sampling is the only practical approach. (ex. I only have access to KAUH then I collect data from it)

b- Judgment sampling:

- A knowledgeable person selects sampling units that he/she feels are most representative of the population.
- The quality of the result is dependent on the judgment of the person selecting the sample.
- Generally, no statistical statement should be made about the precision of the result.

☒ This means the researcher has some knowledge that his sample will be representative or not.

---

**Two types of errors can occur in conducting a survey:**

**1- Sampling error:**

- It is defined as the magnitude of the difference between the point estimate (statistic), developed from the sample, and the population parameter.

Example: The difference between sample and population mean is a sampling error.

- It occurs because not every element in the population is surveyed.

- It cannot occur in a census. (it is not sampling)
- It cannot be avoided, but it can be controlled.

**2- Non-sampling error:**

- It can occur in both a census and a sample survey.
- Examples include:

- Measurement error (ex. Not measuring blood pressure correctly).
- Errors due to non-response (we called respondent but he did not answer the phone).
- Errors due to lack of respondent knowledge (ex. Respondent read question but did not understand it).
- Selection error
- Processing error