
◉Sheet ○Slides

Number

3

Done by:

Omar Fahed Salaymeh

corrected by:

Basel Abdeen

Doctor

Mahmoud Al-Hussami

بسم الله الرَّحْمَنِ الرَّحِيمِ

## Survey errors (slide 57):

1) The subject lies: a question may be too personal for the subject to give an honest answer, or the subject could try to give the socially acceptable response. There is a chance that you'll ask them and they won't tell you the real answer and instead give you false information, which can affect the analysis so we can't take the correct decision.

For example: "where did you eat last night?" the subject may have not eaten anything at all but he could make up a story. Other examples are: "Have you ever used an illegal drug?"; "Have you ever driven a car while intoxicated?"

2) The subject makes a mistake: the subject may not remember the answer, for example "What did you eat last month", "How much money do you have invested in the stock market?"

3) The interviewer makes a mistake: in recording or understanding subject's response.

4) The interviewer cheats: in most cases data collectors (or interviewers) get paid according to the amount of questionnaires they complete, so in order to speed things up they make up some answers and pretend the respondent submitted them. However, this is easily detected because you may have the same answers for every single questionnaire, or you may discover that the handwriting is the same in most of the questionnaires.

*There are two methods to administrate the questionnaire:

A) **Self-administered**: the data collector gives every single participant the questionnaire without any explanation because the questionnaire itself is annotated, so the participant reads the questionnaire and answers everything. Those types of surveys will most likely have low response rate, because people may not be inclined to complete the questionnaire.
B) **Interviewing the participants**: in order to get questionnaires filled out completely and achieve a high response rate, data collectors sit with participants and ask each question separately.

5) The interviewer influences the response: vocal intonation, age, sex, race, clothing of interviewer may influence response, and this happens in the second method as the participants may be affected by the interviewers. For example: An elderly woman dressed very conservatively asking young people about usage of illegal drugs may get different responses than a young interviewer wearing jeans with tattoos on her body and a nose ring.

(0:00-11:00)

## >>Types of Samples (slides 27-30):

The researcher has two ways to collect the data from the samples either by:

1) Non-probability Samples: - based on convenience or the researcher's judgment, where not all the participants in the accessing population have an equal chance of being selected. The problem here is that we do not know how representative our sample is, so there are limitations in generalizing the results to the entire population. Non-probability samples include:-

   A) *Convenience samples*: subjects are selected based their convenient accessibility and/or proximity to the researcher. For example: research conducted on the first 50 students to enter the classroom would be a convenience sample. However, these 50 students don't represent the whole class; therefore we have restrictions in generalizing the results to the entire population.

   B) *Quota sample*: interviewers are given quotas based on demographics. For instance, they may each be told to interview 100 subjects – 50 males and 50 females. Of the 50, say, 10 nonwhite and 40 white, thereby selecting participants from each segment based on a specified proportion.

   C) *Judgment sample:* based on the researcher's judgment as to what constitutes "representativeness" e.g., he/she might say these 20 stores are representative of the whole chain, so in this case the opinion of the researcher affects the sample.

   D) *Snowball sample* (not mentioned in slides): it is used in the researches in which sensitive, illegal, or deviant issues are involved e.g. drug abusers, in this type of sampling the researcher only knows a few participants who meet the specifications

of the research. However, those participants are likely to know others who share the characteristics that make them fit for the study.

We won't be using the Snowball sampling in our projects because they will probably be about student life in the university or a simple existing problem in the community (which is a secret in general).

*Note that we use "**Standardized Questionnaires"** in order to collect valid data.

(11:00-22:00)

2) Probability Samples: - every single subject or participant in the accessing or study population has *a known (or an equal) chance* to be selected, unless there is a set criteria to exclude certain individuals or groups (who do not fit the research). By this method we can generalize the results to the entire population.
Probability samples include:

A) *Simple random sample*: This is a sample collected in such a way that every element in the population has an equal chance of being selected.
So for instance; if we are doing a research on a class of a 100 students, we are not going to select the first 25 students to enter the classroom. Instead, we will:

1) Design a list that includes all the students in the classroom.

2) Give each student a number randomly by using a table of random numbers or a random number generator.

3) Now we start dealing with numbers, the starting number is chosen by lottery, the remaining 24 numbers are the ones that come after this number we chose by lottery. Or simply we may just pick all the numbers by lottery without using a table number as shown in the following picture.

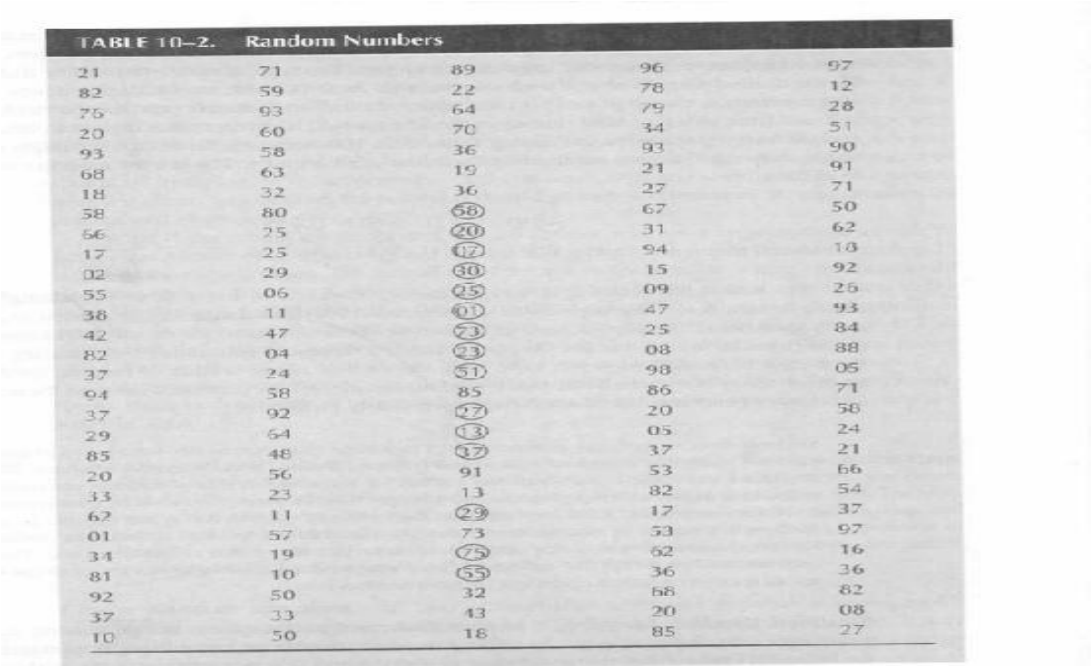

TABLE 10–2. Random Numbers

| | | | | |
|---|---|---|---|---|
| 21 | 71 | 89 | 96 | 97 |
| 82 | 59 | 22 | 78 | 12 |
| 76 | 93 | 64 | 79 | 28 |
| 20 | 60 | 70 | 34 | 51 |
| 93 | 58 | 36 | 93 | 90 |
| 68 | 63 | 19 | 21 | 91 |
| 18 | 32 | 36 | 27 | 71 |
| 58 | 80 | 66 | 67 | 50 |
| 66 | 25 | 20 | 31 | 62 |
| 17 | 25 | 12 | 94 | 10 |
| 02 | 29 | 30 | 15 | 92 |
| 55 | 06 | 25 | 09 | 26 |
| 38 | 11 | 01 | 47 | 93 |
| 42 | 47 | 73 | 25 | 84 |
| 82 | 04 | 23 | 08 | 88 |
| 37 | 24 | 51 | 98 | 05 |
| 94 | 58 | 85 | 86 | 71 |
| 37 | 92 | 27 | 20 | 56 |
| 29 | 64 | 13 | 05 | 24 |
| 85 | 48 | 37 | 37 | 21 |
| 20 | 56 | 91 | 53 | 66 |
| 33 | 23 | 13 | 82 | 54 |
| 62 | 11 | 29 | 17 | 37 |
| 01 | 57 | 73 | 53 | 97 |
| 34 | 19 | 75 | 62 | 16 |
| 81 | 10 | 55 | 36 | 36 |
| 92 | 50 | 32 | 68 | 82 |
| 37 | 33 | 43 | 20 | 08 |
| 10 | 50 | 18 | 85 | 27 |

****Response rate = total questionnaires answered/participants****

In quantitative studies, we enlarge the number of participants to get rid of the effect of the low response rate. Meanwhile, in qualitative studies we have to collect data until the data is saturated (until we see that answers are being repeated).

B) *Systematic random sample*: individuals are chosen at regular intervals from the sampling frame

- Ideally we randomly select a number to tell us the starting point.
- Example: every 5th household, or every 10th woman attending the conference.
- *Sampling fraction* = Sample size /study population.

*Interval size*= study population/Sample size.

**Note that participating in research studies is voluntary, so if someone refuses to participate we have to skip him/her**

C) *Stratified random samp*le: The population is sub-divided based on a characteristic and a simple random sample is conducted within each stratum.

D) *Cluster Sample* (one stage) :it includes 2 steps, first we divide the study population into groups (clusters) then systematic or simple random sampling is used to randomly choose 1 cluster.

If conducting a study on ICU doctors in Amman, one would use exclusion and inclusion criteria to narrow down which hospitals may be selected for this study. For example: if

there are 100 hospitals in Amman, 20 of which have 300 beds or more, those 20 hospitals would meet the inclusion criteria to qualify for the study selection. After narrowing down the study selection based on systemic random and simple random sampling, 3 out of the 20 hospitals would be randomly chosen. Finally the questionnaires are administered to *ALL* ICU doctors.

(22:00-39:00)

## Variables (slide 31):

**Note:** the doctor didn't say anything about this slide so I just copied and pasted everything. (EDIT: the copied and pasted material was slightly edited)

Variables: Measurable characteristic of a person, object or phenomenon which can take on different values. They represent information that must be collected in order to meet the objectives of a study.

Examples: Weight, Distance, Monthly income, Number of children, Color, Outcome of disease, Types of food, Sex.

Variables allow clear definition of the core problem and influencing factors by introducing the concept of numeric value.

## Types of Variables (slides 32-33)

1) *Attribute Variable:* Pre-existing characteristics of the subject which the researcher simply observes and measures and can't manipulate. E.g. blood type.
2) *Active Variable:* Researcher creates or manipulates this. E.g. experimental drug (dosing of a drug)
3) *Continuous variable:* there are no gaps between the numbers (e.g., height, weight,) *(the values are real numbers).*
4) *Discrete variable:* there are gaps between the numbers (e.g., number of children) *(the values are natural numbers).*

To better understand the distinction between continuous and discrete variables, take this example:

A person's weight could include decimal points (continuous), but the number of children must be a *positive integer value* (discrete)… we can't say that someone has 3.5 brothers but we can say that someone weighs 75.5kg.

5) *Categorical variable:* has more than two values (e.g., marital status could be widowed, married, single, etc.)
6) *Dichotomous variable :* has only **two** values (e.g., gender, yes/no questions)

>>According to the **work/action** of the variables we divide them into:-

1) *Independent variable (IV):* the presumed cause (of a dependent variable)
2) *Dependent variable (DV):* the presumed effect (of an independent variable)
   Example: Smoking (IV) Lung cancer (DV)
3) *Confounding variable:* It has an effect on both the IV and the DV.
4) *Extraneous variable:* it has an effect only on the DV.

For example: in a study on a child's life where the child's life is the dependent variable, the parents are considered an independent variable as they influence the child's life. The family's income influences the parents' lives and the child's life (both IV and DV) so we consider it a *confounding variable*. While the school that the child attends only affects the child (DV) which makes it an *extraneous variable*.

**Levels of Measurement slides (34-39):**

***Note** the doctor didn't talk about these slides so I just copied and pasted everything. (EDIT: also slightly modified)

**1) Nominal:** the researcher just gives names to the values. Nominal measurement includes Categorical and dichotomous data

- Distinct categories such as gender, religion, marital status are symbols that have no quantitative value.
- Nominal measurement is the lowest level of measurement.
- Many characteristics can be measured on a nominal scale such as race, marital status, and blood type.
- We use appropriate statistics in descriptive statistics; e.g. frequency, percentage, and mode.

We cannot use an average. It would be meaningless here.

**2) Ordinal:**

- Involves using numbers to rank data points in ordered categories.

- The exact difference between the ranks cannot be specified since the ranks are not based on actual quantitative values.
- **Example: anxiety level: mild, moderate, severe.** Statistics used involve frequency distributions and percentages.
- Appropriate statistics: same as those for nominal data, plus the median; but not the mean.

**3) Interval:**

- They are real numbers and the difference between the ranks can be specified.
- Equal intervals, but no "true" zero.
- Involves assigning numbers that indicate both the order on an attribute and the distance between score values on the attribute.
- They are actual numbers on a scale of measurement.
- Example: two measured body temperatures could be 36.2 and 37.2 degrees Celsius. (difference: 1 degree Celsius)
- Appropriate statistics: same as for ordinal plus the mean,

**3) Ratio:** Is the highest level of data where data can be categorized and ranked. The difference between ranks can be specified, and a true or natural zero point can be identified.

- True zero point means that the value "0" indicates total absence of the quantity being measured.
- All scales, whether they measure weight in kilograms or pounds, start at 0. The 0 means something and is not arbitrary (SUBJECTIVE).
- Example: total amount of money.

**Levels of Measurement** (39:00-49:00)