



## ICS 500: Research Methods and Experiment Design in Computing

# Lecture

# Surveys

# Lecture Objectives

- ✓ Discuss the stages in survey-based research:
  - Setting the survey's objectives
  - Selecting the most appropriate survey design
  - Constructing the survey instrument
  - Assessing the reliability and validity of the survey instrument
  - Administering the instrument
  - Analyzing the collected data



# What is Survey?



- A survey is not just the instrument (the questionnaire or checklist) for gathering information.
- A survey is:
  - A comprehensive research method
  - for collecting information
  - from (a sample of) entities
  - to describe, compare or explain knowledge, factors, attitudes and behavior.

# When to do Survey?



- We use survey when we have to measure many variables (with multiple indicators) and test several hypothesis at one time.
- Survey can be used for research questions about factors, attitude, believes, opinions, characteristics, expectations, knowledge etc.


# Questionnaires vs. Surveys



- Are questionnaires and surveys the same?
  - Survey describes the process of conducting the research, which includes multiple steps
  - Questionnaire is one part of the survey process.
  
- What is a questionnaire?
  - A list of survey questions asked to respondents to extract specific information.
  - The data collection component of an overall survey

# Survey Stages



- Setting the survey's objectives
  - Selecting the most appropriate survey design
  - Constructing the survey instrument
  - Assessing the reliability and validity of the survey instrument
  - Administering the instrument
  - Analyzing the collected data.
- 

# 1. Set Objectives



- Each objective is simply a statement of the survey's expected outcomes or a question that the survey is intended to answer.
  - e.g., To identify the most useful features of a front-end development tool
  - To identify the most effective best practices for safeguarding organizations against insider threats.
- As the objectives are defined in more detail, you should be able to specify:
  - The hypotheses to be tested
  - What scope of survey project is appropriate to address the objectives
  - What resources are necessary to achieve the objectives

## 2. Design Survey



- Two common types of design
  - Cross-Sectional
  - Longitudinal
  
- The way survey will be administered
  - Self-administered questionnaires (postal or via Internet)
  - Telephone surveys
  - One-to-one interviews



## 2.1 Cross-Sectional Surveys



- Cross-sectional surveys are used to gather information on a population at a single point in time.
  - e.g., How parents feel about Internet filtering, as of March of 2024.
- A different cross-sectional survey questionnaire might try to determine the relationship between two factors, like religiousness of parents and views on Internet filtering.
  - Practicing parents are more interested in internet filtering etc

## 2.2 Longitudinal Surveys



- Longitudinal surveys gather data over a period of time.
- The researcher may then analyze changes in the population and attempt to describe and/or explain them.
- The three main types of longitudinal surveys are:
  - trend studies,
  - cohort studies, and
  - panel studies.

# Trend Studies

- Trend studies focus on a particular population, which is sampled and examined repeatedly.
- While samples are of the same population, they are typically not composed of the same people.
- Trend studies, since they may be conducted over a long period of time, do not have to be conducted by just one researcher or research project.
- A researcher may combine data from several studies of the same population in order to show a trend.
  - A yearly survey of KFUPM students asking about their attitudes towards Physical Education
- Population = KFUPM students
- Sample = different students chosen from population


# Cohort Studies



- Cohort studies also focus on a particular population, sampled and studied more than once.
  
- But cohort studies have a different focus.
  - For example, a sample of 2024 students of KFUPM (first year class cohort) could be questioned regarding their attitude towards PE
  - Three years later, the researcher could question another sample of 2024 students (same cohort – first year students), and study any changes in attitude ...
  
- A cohort study would sample the same cohort, every time but of course different people/ sample.

# Panel Studies



- Panel studies allow the researcher to find out why changes in the population are occurring, since they use the same sample of people every time.
  - That sample is called a panel.
    - Select a sample of KFUPM undergraduate students of 1st semester and ask them questions on their library usage.
    - Every year thereafter, the researcher would contact the same people, and ask them similar questions, and ask them the reasons for any changes in their habits.
  - Panel studies, while they can yield extremely specific and useful explanations, can be difficult to conduct.
  - They tend to be expensive, they take a lot of time, and they suffer from high attrition rates. Attrition is what occurs when people drop out of the study.
- 

# 3. Develop Instrument



- Survey instruments, which are usually questionnaires, are developed using the following steps:
  - Search the relevant literature
  - Construct an instrument
  - Evaluate the instrument
  - Document the instrument

## 3.1 Search the relevant literature



- Identify what other studies have been done on the topic.
- Determine how the previous studies' researchers collected their data and what questionnaires or other data collection mechanisms were used.
- Surveying the literature can help knowing:
  - if previous studies have developed relevant instruments or questions that we can adopt
  - if other researchers had problems with response rates, we will be aware of the need to adopt measures to address this problem.

## 3.2 Construct an Instrument



- The starting point in designing the survey instrument should always be the survey's purpose and objectives
- Simply converting a list of objectives into a set of questions rarely leads to a successful survey instrument.
- Carefully design
  - Questions – Types, Language, and Measuring the concepts
  - Answer to the questions




# Types of Survey Questions



- Open-ended or Verbal
  - The expected response is a word, phrase, or an extended comment
  - Responses can produce useful information, but analysis can present problems
  - Some form of content analysis may be required

# Designing Questions



- The language used is appropriate for the intended respondents and any possibly ambiguous terms are fully defined.
  - Use standard grammar, punctuation and spelling
  - Each question expresses one and only one concept; keep questions short but complete
  - Do not include vague questions or ambiguous qualifiers
  - Avoid asking question about events that occurred a long time in the past
  - Avoid asking sensitive questions that respondents may not be willing to answer
- 

# Types of Survey Questions



- Agreement scales
  - e.g., a response choice of the form: Strongly Disagree, Disagree, Neither Agree nor Disagree, Agree, Strongly Agree.
  
- Frequency scales
  - e.g., a response choice of the form: Never, Rarely, Sometimes, Occasionally, Most of the time.
  
- Evaluation scales
  - e.g., a response choice of the form: Terrible, Inferior, Passable, Good, Excellent.

# Types of Survey Questions (*cont'd*)

- Closed-ended, Likert Scale

Not very important

Very Important

1

2

3

4

5

- The respondents must indicate by selecting the number that best represents their attitude
- A quick and easy way to measure ranges of opinions


# Types of Survey Questions (*cont'd*)



- Closed-ended, Multiple Choice

Which learning activity did you like best?

- A. Small group
- B. Role play
- C. Mapping
- D. Computer modeling

- When you want respondents to pick the best answer or answers, consider this type
  - Be sure to include specific instructions
- 

# Types of Survey Questions (*cont'd*)

- Closed-ended, Ordinal

Please write a number between 1 and 5 next to each item below.

Put a 1 next to the item that is most important to you in selecting an **on-line university course**. Put a 5 next to the item that is LEAST important. Please use each number only once.

- \_\_\_\_ availability of instructor for assistance
- \_\_\_\_ tuition cost for the course
- \_\_\_\_ ability to work in groups with other students
- \_\_\_\_ quality and quantity of instructor feedback
- \_\_\_\_ number of students enrolled

- Great for rating things in relation to other things

- Again, be specific with instructions


# Types of Survey Questions (*cont'd*)



- Closed-ended, Categorical

How many hours do you study on a school night?

- 0-1
- 2-3
- 3+

- Be sure that categories do not overlap
  - Be sure to break down items into distinct and clear categories
  - Each respondent must “belong” in a specific category
- 

# Types of Survey Questions (*cont'd*)




- Closed-ended, Numerical
  
- How many years have you taught full time?
  - When the answer must be a real number, ask a numerical question
  - Use these, especially with large survey populations, to do an analysis of age range, years taught, etc.
  - Not as effective with smaller samples



# Which One Do I Use?




- Use Open-ended question when
    - you want to get the respondent's own words
  - Use Likert-Scale questions to
    - assess a person's feelings about something
  - Use Multiple-Choice questions when
    - there are a finite (limited) number of options
  - Use Ordinal questions to
    - rate things in relation to other things
  - Use Categorical questions when
    - the respondent must fall into one section or category
  - Use Numerical questions for
    - real numbers like age, number of months, etc. for large surveys
- 

# Contingency Questions

- Contingency questions can be any of the styles previously discussed
- Typically used when you want to filter out certain respondents
- Respondents are directed to other questions in the survey
- Too many jumps can confuse the reader
  
- Example:
  - Have you ever been bullied in school?
    - Yes
    - No
    - If “yes,” go to question #4


# Questionnaire Format



- Leave a space for the respondents to comment on the questionnaire.
  - Use space between questions.
  - Use vertical format, spaces, boxes, arrows, etc. to maximize the clarity of questions.
  - Do not overwhelm the respondent with “clever” formatting techniques (particularly for Web Questionnaires).
  - Stick to a font size of 10–12.
  - Use a font that is easy to read.
  - Avoid lot of italics.
  - Use bolding, underlining or capitals judiciously and consistently for emphasis and instructions.
- 

# Inform participants



- What the purpose of the study is.
  - Who is sponsoring the study (and perhaps why).
  - Why it should be of relevance to them.
  - Why each individual's participation is important.
  - How and why each participant was chosen.
  - How confidentiality will be preserved.
  - A realistic estimate of the time required to complete the questionnaire.
- 

## 3.3 Evaluate the Instrument



- Evaluation is often called pre-testing, and it has several different goals to:
  - Check that the questions are understandable
  - Assess how long it takes to answer your questionnaire
  - Evaluate the validity and reliability of the instrument
  - Ensure that our data analysis techniques match our expected responses

## 3.3 Evaluate the Instrument (*cont'd*)

- The most common way to evaluate the validity of the questionnaire are focus groups or expert judgement.
  - The concept of validity is concerned with the extent to which your questionnaire measures what it aims to measure
- Focus groups (mediated discussion groups)
  - Assemble a group of people representing either those who will use the results of the survey or those who will be asked to complete the survey (or perhaps a mixture of the two groups).
  - The group members are asked to fill in the questionnaire and to identify any potential problems.

## 3.3 Evaluate the Instrument (*cont'd*)

- To evaluate the reliability of the questionnaire, the Cronbach's Alpha test is performed in the pilot study. The Cronbach's alpha test is a statistical measure used to assess a questionnaire's internal consistency or reliability. It measures how closely related a set of items are as a group,
- The Cronbach's alpha coefficient ranges from 0 to 1, with higher values indicating greater internal consistency. Generally, a Cronbach's alpha value of 0.7 or higher is considered acceptable for most research purposes
  - Example a pilot study with 20 participants and 15 questions

Case Processing Summary			
	N	%	
Valid	20	100	
Excluded	0	0	
Total	20	100	

Reliability Statistics	
Cronbach's Alpha	N of Items
0.871	15

# EXAMPLES

## 3.3 Evaluate the Instrument (*cont'd*)

Suppose we have the following survey items related to customer satisfaction with a smartphone:

- The smartphone meets my needs.
- The smartphone is of high quality.
- The price of the smartphone is reasonable.
- The customer service provided by the company is satisfactory.
- The smartphone packaging is attractive.

Respondents are asked to rate each item on a scale from 1 to 5, where 1 represents "strongly disagree" and 5 represents "strongly agree."

If respondents answer consistently to these questions then alpha will be on the high side and if respondents answer inconsistently then the alpha will be on the low side



## 3.4 Documentation



- Writing an initial descriptive document, called a questionnaire specification.
- It should include:
  - The objective(s) of the study.
  - A description the rationale for each question.
  - The rationale for any questions adopted or adapted from other sources, with appropriate citations.
  - A description of the evaluation process.
- Once the questionnaire is administered, the documentation should be updated to record information about:
  - Who the respondents were.
  - How it was administered.
  - How the follow-up procedure was conducted.
  - How completed questionnaires were processed.

# Samples and Populations

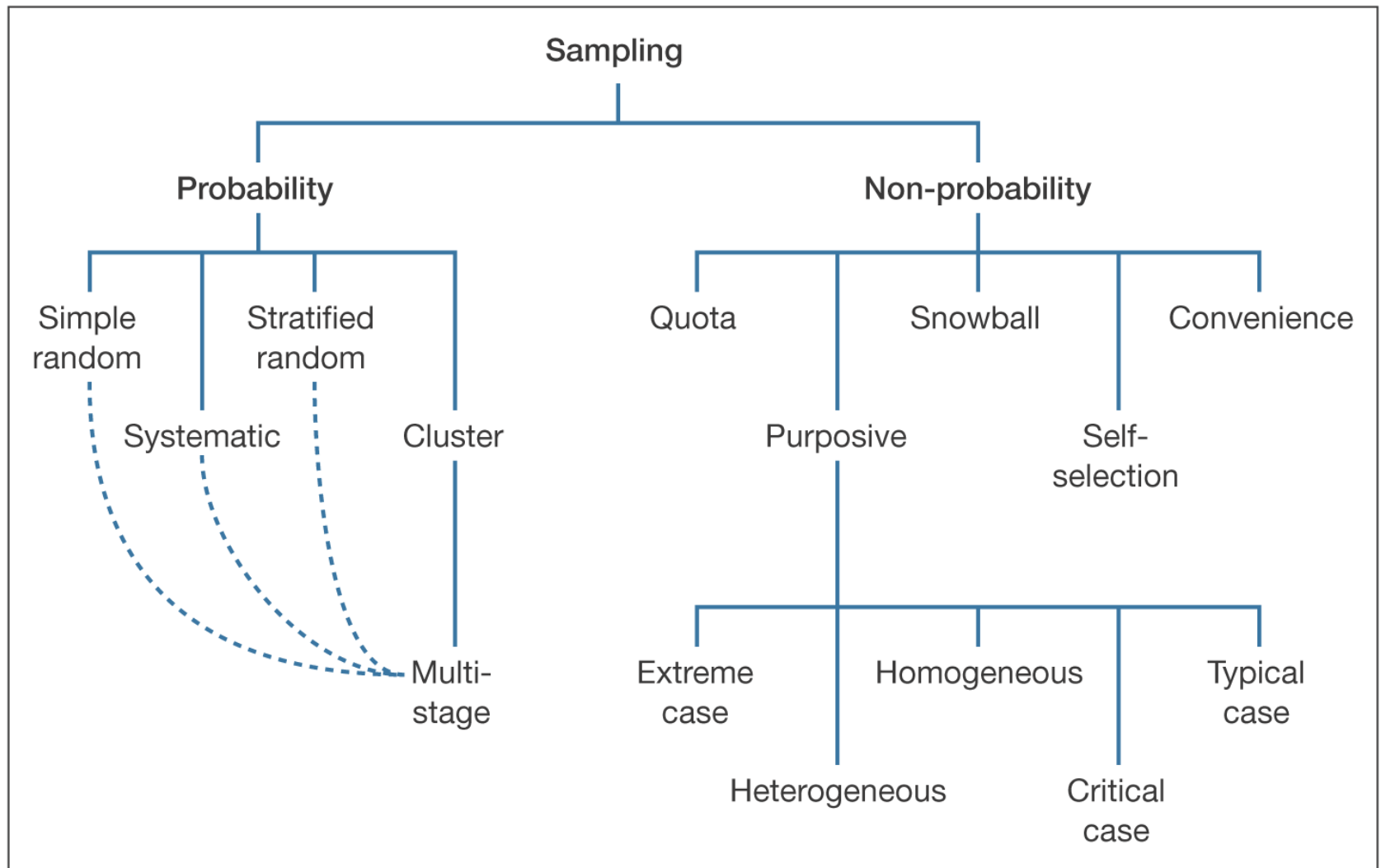


- To obtain a sample, define a target population.
  - The target population is the group or the individuals to whom the survey applies.
  - Seek those groups or individuals who are in a position to answer the questions and to whom the results of the survey apply.
  
- Ideally, a target population should be represented as a finite list of all its members called a sampling frame (In statistics, a sampling frame is the source from which a sample is drawn. It is a list of all those within a population who can be sampled)
  - For voting preferences, the electoral list is the sampling frame

# Samples and Populations (*cont'd*)

- A valid sample is a representative subset of the target population.
  - If we do not have a representative sample, we cannot claim that our results generalize to the target population
- Our initial assessment of the target population should arise from the survey objectives, not from a sense of who is available to answer our questions.
  - The more precisely the objectives are stated, the easier it will be to define the target population.
- We need to consider whether the analyses will lead to any meaningful conclusions, in particular:
  - Will the analysis results address the study objectives?
  - Can the target population answer our research questions?

# Sampling Methods



# Probabilistic Sampling Methods

- A **probability sampling** method is any method of sampling that utilizes some form of random selection.
  - In order to have a random selection method, you must set up some process or procedure that assures that the different units in your population have equal probabilities of being chosen.
- A **probability sample** is a sample in which every unit in the population has a chance (greater than zero) of being selected in the sample.
  - The aim of a probabilistic sample is to eliminate subjectivity and obtain a sample that is both unbiased and representative of the target population.
- Remember: we cannot make any statistical inferences from our data unless we have a probabilistic sample.

# Probabilistic Sampling Methods (*cont'd*)

- A **simple random sample** is one in which every member of the target population has the same probability of being included in the sample.
- A **stratified random sample** is obtained by dividing the target population into subgroups called strata. Each stratum is sampled separately.
  - In stratified sampling, the sampling is done on elements within each stratum.
  - Strata are used when we expect different sections of the target population to respond differently to our questions, or when we expect different sections of the target population to be of different sizes.
    - For example, we may stratify a target population on the basis of gender, because men and women often respond differently to questionnaires.

# Probabilistic Sampling Methods (*cont'd*)

- **Systematic sampling** involves selecting every  $n^{\text{th}}$  member of the sampling frame. If the list is random, then selecting every  $n^{\text{th}}$  member is another method of obtaining a simple random sample.
  
- **Cluster-based sampling** is the term given to surveying individuals that belong to defined groups.
  - For example, we may want to survey all members of a family group, or all patients at specific hospitals.
  - Randomization procedures are based on the cluster, not the individual.
  - We would expect members of each cluster to give more similar answers than we would expect from members of different clusters.

# Non-Probabilistic Sampling Methods

- Non-probability samples are created when respondents are chosen because they are easily accessible, or the researchers have some justification for believing that they are representative of the population.
- This type of sample runs the risk of being biased (that is, not being representative of the target population), so it is dangerous to draw any strong inferences from them.
- Certainly, it is not possible to draw any statistical inferences from such samples.

why we need ? }



# Why Non-Probabilistic Sampling?

- Three reasons for using non-probability samples:
  - The target population is hard to identify.
    - For example, if we want to survey software hackers, they may be difficult to find.
  - The target population is very specific and of limited availability.
    - For example, if we want to survey senior executives in companies employing more than 5000 software engineers, it may not be possible to rely on a random sample. We may be forced to survey only those executives who are willing to participate.
  - The sample is a pilot study, not the final survey, and a non-random group is readily available.
    - For example, participants in a training program might be surveyed to investigate whether a formal trial of the training program is worthwhile.

# Non-Probabilistic Sampling Methods (*cont'd*)

- **Convenience sampling** involves obtaining responses from those people who are available and willing to take part.
  - The main problem with this approach is that the people who are willing to participate may differ in important ways from those who are not willing.
  - For example, people who have complaints are more likely to provide feedback than those who are satisfied with a product or service
- **Snowball sampling** involves asking people who have participated in a survey to nominate other people they believe would be willing to take part. Sampling continues until the required number of responses is obtained.
  - This technique is often used when the population is difficult for the researchers to identify.
  - For example, we might expect software hackers to be known to one another, so if we found one to take part in our survey, we could ask him/her to identify other possible participants.

# Non-Probabilistic Sampling Methods (*cont'd*)



- **Quota sampling** is the non-probabilistic version of stratified random sampling.
  - The target population is split into appropriate strata (groups) based on known subgroups (e.g., gender, educational achievement, company size etc.).
  - Each stratum is sampled (using convenience or snowball techniques) so that number of respondents in each subgroup is proportional to the proportion in the population.
  - i.e., all groups have equal number of respondents

# Non-Probabilistic Sampling Methods (*cont'd*)



- **Self-selection sampling** is useful when we want to allow individuals or organizations, to choose to take part in research on their own accord.
  - Survey researchers may put a questionnaire online and subsequently anyone within a particular organization can take part.
  - The key component is that research subjects (or organizations) volunteer to take part in the research on their own accord. They are not approached by the researcher directly.

# Non-Probabilistic Sampling Methods (*cont'd*)

- In **Purposive sampling**, we sample with a purpose in mind.
- We usually would have one or more specific predefined groups we are seeking.
  - For example, have you ever run into people in a mall or on the street who are carrying a clipboard and who are stopping various people and asking if they could interview them? Most likely they are conducting a purposive sample (and most likely they are engaged in market research). They might be looking for females between 30-40 years old. They size up the people passing by and anyone who looks to be in that category they stop to ask if they will participate. One of the first things they're likely to do is verify that the respondent does in fact meet the criteria for being in the sample.
  - Purposive sampling can be very useful for situations where you need to reach a targeted sample quickly

Example | Carros do fogo


# Sample Size



- A major issue of concern when sampling is determining the appropriate sample size.
- There are two reasons why sample size is important.
  - An inadequate sample size may lead to results that are not significant statistically.
    - If the sample size is not big enough, we cannot come to a reasonable conclusion, and we cannot generalize to the target population.
  - Inadequate sampling of clusters or strata disables our ability to compare and contrast different subsets of the population.

# Response Rates



- It is not enough to decide how many people to participate in survey.
  - We must also take steps to be sure that enough people return the survey to yield meaningful results.
  - Any reliable survey should measure and report its response rate, that is, the proportion of participants who responded compared to the number who were approached.
  - The validity of survey results is severely compromised if there is a significant level of non-response.
- 

# Response Rates (*cont'd*)



- If we have a large amount of non-response but we can understand why and can still be sure that our pool of respondents is representative of the larger population,
  - We can proceed with our analysis.
- If there is large non-response and we have no idea why people have not responded,
  - We have no way of being sure that our sample truly represents the target population.
  - It is even worse to have no idea what the response rate is.



# Analyzing Survey Data

## ■ Data Validation

- Before undertaking any detailed analysis, responses should be examined for consistency and completeness.
- It is important to have a *policy* for handling inconsistent and or incomplete questionnaires.
  - If we find that most respondents answered all questions, we may decide to reject incomplete questionnaires.
- We must investigate the characteristics of rejected questionnaires in the same way that we investigate non-response to ensure that we do not introduce any systematic bias.
- We may find that most respondents have omitted a few specific questions.
  - In this case, it is more appropriate to remove those questions from the analysis.

# Sample Size and response rate



Se manda al  
prof se salto  
muchas d'p'o.

Krejcie, R.V., & Morgan, D.W., (1970).  
Determining Sample Size for Research  
Activities. Educational and Psychological  
Measurement.

$N$	$S$	$N$	$S$	$N$	$S$
10	10	220	140	1200	291
15	14	230	144	1300	297
20	19	240	148	1400	302
25	24	250	152	1500	306
30	28	260	155	1600	310
35	32	270	159	1700	313
40	36	280	162	1800	317
45	40	290	165	1900	320
50	44	300	169	2000	322
55	48	320	175	2200	327
60	52	340	181	2400	331
65	56	360	186	2600	335
70	59	380	191	2800	338
75	63	400	196	3000	341
80	66	420	201	3500	346
85	70	440	205	4000	351
90	73	460	210	4500	354
95	76	480	214	5000	357
100	80	500	217	6000	361
110	86	550	226	7000	364
120	92	600	234	8000	367
130	97	650	242	9000	368
140	103	700	248	10000	370
150	108	750	254	15000	375
160	113	800	260	20000	377
170	118	850	265	30000	379
180	123	900	269	40000	380
190	127	950	274	50000	381
200	132	1000	278	75000	382
210	136	1100	285	100000	384

Note.— $N$  is population size.  $S$  is sample size.

Source: Krejcie & Morgan, 1970

- 
- [https://www.youtube.com/watch?v=8C\\_dK6XicNI](https://www.youtube.com/watch?v=8C_dK6XicNI)
  - <https://corporatefinanceinstitute.com/resources/knowledge/other/ordinal-data/>
  - <https://www.scribbr.com/statistics/ordinal-data/>
- 

# Summary

- Discuss the stages in survey-based research:
  - Setting the survey's objectives;
  - Selecting the most appropriate survey design
  - Constructing the survey instrument
  - Assessing the reliability and validity of the survey instrument
  - Administering the instrument
  - Analyzing the collected data.

the last class before the break  
Exam average 71%, max 90%

# References



- Paul Leedy and Jeanne Ormrod, 2015, “Practical Research: Planning and Design”, Pearson, ISBN 9781292095875, 1292095873
  - Loraine Blaxter, Christina Hughes and Malcolm Tight, “How to Research”, 4<sup>th</sup> Edition.
  - F. Shull et al. (eds.), Guide to Advanced Empirical Software Engineering, Springer 2008
  - O. Tamimi, “Principles of survey research,” SlideShare, 22-Mar-2014. [Online]. Available: <https://www.slideshare.net/omartamimi56/principles-of-survey-research>. [Accessed: 24-Oct-2020].
- 