Summary CHAPTER 2

Research in human–computer interaction examines human behavior in relation to computers or computer-related devices. There are three major types of research methods for studying human behavior: descriptive, relational, and experimental. The major strength of experimental research, compared to the other two types, is that it allows the identification of causal relationships between entities or events.

After a hypothesis is constructed, the design of an experiment consists of three components: treatments, units, and the assignment method. In an experiment, the process of sample selection needs to be randomized or counter-balanced, as does the assignment of treatments, or experiment conditions. Many methods can be used to randomly select samples or assign experiment conditions, including, but not limited to, the random digit table and software-generated randomization schemes.

Successful experimental research depends on well-defined research hypotheses that specify the dependent variables to be observed and the independent variables to be controlled. Usually a pair of null and alternative hypotheses are proposed and the goal of the experiment is to test whether the null hypothesis can be rejected or the alternative hypothesis can be accepted. Good research hypotheses should have a reasonable scope that can be tested within an experiment; clearly defined independent variables that can be strictly controlled; and clearly defined dependent variables that can be accurately measured.

Significance testing allows us to judge whether the observed group means are truly different. All significance tests are subject to two types of error. Type I errors refer to the situation when the null hypothesis is mistakenly rejected when it is actually true. Type II errors refer to the situation of not rejecting the null hypothesis when it is actually false. It is
generally believed that Type I errors are worse than Type II errors, therefore the alpha threshold that determines the probability of making Type I errors should be kept low. The widely accepted alpha threshold is 0.05.

With its notable strengths, experimental research also has notable limitations when applied in the field of HCI: difficulty in identifying a testable hypothesis, difficulty in controlling potential confounding factors, and changes in observed behavior as compared to behavior in a more realistic setting. Therefore, experimental research methods should only be adopted when appropriate.

Summary CHAPTER 3

Experiment design starts with a clearly defined, testable research hypothesis. During the design process, we need to answer the following questions:

- How many dependent variables are investigated in the experiment and how are they measured?
- How many independent variables are investigated in the experiment and how are they controlled?
- How many conditions are involved in the experiment?
- Which of the three designs will be adopted: between-group, within-group, or split-plot?
- What potential bias may occur and how can we avoid or control those biases?

When an experiment studies only one independent variable, we need to choose between the between-group design andthe within-group design. When there is more than one independent variable, we need to select among the between-group design, the within-group design, and the split-plot design.

The between-group design is cleaner, avoids the learning effect, and is less likely to be affected by fatigue and frustration. But this design is weaker due to the high noise level of individual differences. In addition, larger numbers of participants are usually required for a between-group design. The within-group design, on the other hand, effectively isolates individual differences and, therefore, is a much stronger test than the between-group design. Another bonus is that fewer participants are required. But within-group designs are more vulnerable to learning effects and fatigue. The appropriate design method needs to be selected based on the nature of the application, the participant, and the tasks examined in the experiment.

All experiments strive for clean, accurate, and unbiased results. In reality, experiment results are highly susceptible to bias. Biases can be attributed to five major sources: the measurement instruments, the experiment procedure, the participants, the experimenters, and the physical and social environment. We should try to avoid or control biases through accurate and appropriate measurement devices and scales; clearly defined and detailed experimental procedures; carefully recruited participants; well-trained, professional, and unbiased experimenters; and well-controlled environments.

Summary CHAPTER 4

Statistical analysis is a powerful tool that helps us find interesting patterns and differences in the data as well as identify relationships between variables. Before running significance tests, the data needs to be cleaned up, coded, and appropriately organized to meet the needs of the specific statistical software package. The nature of the data collected and the design of the study determine the appropriate significance test that should be used. If the data are normally distributed and intervally scaled, parametric tests are appropriate. When the normal distribution and interval scale requirements are not met, non-parametric tests should be considered.

A number of statistical methods are available for comparing the means of multiple groups. A simple t test allows us to compare the means of two groups, with the independent-samples t test for the between-group design and the paired- samples t test for the within-group design. A one-way ANOVA test allows us to compare the means of three or more groups when a between-group design is adopted and there is only one independent variable involved. When two or more variables are involved in a between-group design, the factorial ANOVA test would be appropriate. If a study adopts a within-group design and involves more than two conditions, the repeated measures ANOVA test would be appropriate. For studies that involve both a between-group factor and a within-group factor, the split-plot ANOVA test should be considered.

Correlation analysis allows us to identify significant relationships between two variables. When three or more variables are involved and a quantitative model is needed to describe the relationships between the dependent variable and the independent variables, regression analysis can be considered. Different regression procedures should be used based on the specific goals of the study.

Non-parametric statistical tests should be used when the data do not meet the required assumptions of parametric tests. The Chi-square test is widely used to analyze frequency counts of categorical data. Although non-parametric tests have less strict requirements for the data, they are not assumption free and the data still need to be carefully examined before running any non-parametric tests.

Summary CHAPTER 5

Surveys are a very powerful tool for collecting data from many individuals. However, for survey data to be valid, there must be a number of different steps that take place. Survey questions must be well-worded and the survey design should make it easy for respondents to understand and use. Appropriate sampling methods, even if they are non-probabilistic, must be used to ensure a representative response. There must be a sufficient number of responses for the data to be considered valid. The key is to use surveys when they are appropriate to the research questions. However, other methods can also be useful in conjunction with surveys, such as focus groups, interviews, or time diaries.

Summary CHAPTER 8

Interviews and focus groups present substantial challenges for HCI researchers and practitioners. Writing questions, identifying appropriate respondents, conducting interviews, and analyzing data all require considerable skill and experience. For those of us who come to HCI from a technical background, the social science techniques and strategies that are involved may seem unfamiliar and somewhat daunting.

Despite these concerns, interviews and focus groups are invaluable tools for HCI researchers and practitioners, providing data into user and stakeholder needs and perceptions that would be difficult, if not impossible to get using other techniques. It's that simple – if you want to know what people want or what they think, you must ask them. For researchers, this might mean in-depth conversations aimed at building models to explain how systems are used and why. For designers and builders of interfaces, interviews can help build understanding of needs and reactions to interfaces. If you want to know why your last design failed, you can start by interviewing the users.

The choice of one-to-one interviews or focus groups involves trade-offs in time, expediency, depth, and difficulty. Focus groups let you hear from many people at once but with less depth from any given individual. You should consider the trade-off between this loss of depth and the potentially fuller understanding that may arise from a conversation between participants having multiple perspectives. Unfortunately, there are no guarantees: this intriguing dynamic conversation might not materialize. As the moderator of a focus group, you have a very important role to play: this is where the difficulty comes in. Skillful moderation can keep conversation focused and inclusive, increasing your chances of getting good data.

Interviews and focus groups might best be conducted as complements to other data collection approaches. Empirical studies, usability tests, ethnographic investigations, and case studies are among the methods that might be used alongside interviews. You can use multiple, complementary tactics to confirm findings or identify potential disconnects. Perhaps users prefer one interface design over another, even though it is slower. Why is that? Well-formed interview questions might help you understand the reasons.

If you feel intimidated by these challenges, start small. A simple, fully structured interview with closed questions will help you get started. As you become more comfortable with writing questions, talking to interviewees, and analyzing data, you might move on to interviews with less structure and greater challenges. Don't be ashamed to bring in some outside help. A colleague who is knowledgeable and experienced in interviewing can be an invaluable aide.

Summary CHAPTER 11

Text, multimedia, and other qualitative data are important sources of information for HCI researchers and practitioners. The procedure and techniques commonly used to analyze qualitative data are quite different from those applied to the analysis of quantitative data. Probably the most unique characteristic of content analysis is that it involves human coding. The absence of numeric data and direct measures makes qualitative data analysis more susceptible to biased interpretation or subjective manipulation. Therefore, it is critical to adopt well established procedures and techniques to ensure high-quality analysis that is both valid and reliable. Although there is disagreement regarding its implementation process and guidelines, grounded theory is widely used for qualitative data analysis. The major difference between grounded theory and other qualitative research strategies is its emphasis on theory development in continuous interplay between data collection and data analysis.

When analyzing text content, we need to develop a set of coding categories that accurately summarizes the data or describes the underlying relationships or patterns buried in the data. Depending on the specific context of the research question, a priori coding or emergent coding may be used to generate the coding categories. In order to produce high- quality coding, multiple coders are needed to code the data. During the coding process, the coders should constantly look for statements likely to carry valuable information, ask questions about the data, and make comparisons at various levels. Reliability control measures such as Cohen's Kappa should be calculated and evaluated throughout the coding process. Cohen's Kappa at or above 0.60 indicates satisfactory inter-coder reliability.

The basic guidelines for analyzing text content also apply to multimedia content. Due to the special nature of multimedia data, the analysis can be much more labor-intensive than for text data if a completely manual annotation procedure is adopted. In order to address that challenge, a number of techniques have been developed to assist the annotation of multimedia data. To date, the completely automated annotation techniques are highly error prone. Applications to facilitate manual coding have shown promising results and may serve as a useful tool for analyzing multimedia data.

Summary CHAPTER 14

Working with human subjects is one of the most challenging and informative aspects of HCI research. Finding appropriate participants; informing them of their rights; protecting their privacy; and answering their questions can be time- consuming and often tedious, but the results are more than worth the effort. Even when study participants criticize our designs or fail to confirm our cherished experimental hypotheses, they provide invaluable insight that provides a rigorous foundation for our work.

Whatever type of study you are running, it is never too early to plan for recruiting, informed consent documentation, and other aspects of human participation. Proper planning will keep your study from becoming one of the many that have been delayed by unforeseen circumstances including difficulty in finding participants, or delays in IRB approval.

Recruiting entails finding the right number of the right kinds of participants. For usability studies, ethnographic observations of users, interviews, focus groups, and other approaches aimed at gathering requirements or evaluating design proposals, this may mean understanding the audience of users and identifying a sample of participants that is broad enough to reflect the needs and behavior of potential users. Designers and professional developers conducting research of this sort might work with collaborators, marketing teams, professional organizations, or others with appropriate understanding and context to identify both the range of viewpoints that would be needed and possible sources of the appropriate individuals.

Empirical studies require consideration of both the diversity of potential participants and any confounding factors that might contribute to performance differences. Characteristics of desirable participants might both be informed by and influence experimental hypotheses. Students and researchers conducting these studies should be careful to plan their data analysis and recruiting together, to ensure that the participants will be selected to increase the power of the statistical analysis.

Appropriate respect for participants is a cornerstone of all research involving human subjects. Although designers and developers may not be required to secure the approval of institutional review boards, they should still endeavor to protect their subjects from any form of harm and to treat them with respect and dignity. These issues are particularly relevant for studies that involve deception. Even when not required by institutional policy to do so, designers and developers would be advised to use formal informed consent forms to help participants make informed decisions. Students and researchers should take the time – again, as early in the process as possible – to understand the regulations in force in their institution, and to make sure that their approvals are in order before starting any project.

Designers and developers may find online studies to be an attractive means of evaluating proposed interface designs. Students and researchers will undoubtedly continue to find the prospect of online research too enticing to resist. Before moving studies online, HCI professionals should be careful to validate that their proposed designs will provide the desired information. Pilot tests may be particularly useful in these cases.

Human subjects research in HCI can be an unpredictable and often unsettling process. Unforeseen problems, including misinterpreted tasks and goals, systems failures and missed appointments, are routine: it’s rare that a study (of any sort) goes off completely without a hitch. These matters can complicate data collection and interpretation: if a user chooses an interpretation of a written task that differs from your intent and then completes the task correctly, how do you interpret the result – is it correct or not? What should be done with results from a user who decides to withdraw from a study after completing only a portion of the tasks? As hard and fast rules for handling situations like these are few and far between, you may have to handle each issue on a case-by-case basis. The specific decisions that you make may be less important than how they are enforced: consistent application of policies and procedures will ensure your ability to make meaningful comparisons.

All participants in HCI research study should be well-treated and approached with an open mind. Participating in HCI studies should be fun and engaging whenever possible: by making our studies positive experiences, we encourage people both to participate and to provide useful feedback. As researchers, we should “expect the unexpected”: software will crash, devices won’t work, and (perhaps most distressingly) users will hate our beloved inventions. High-quality HCI research takes these setbacks in its stride, all the while striving to observe carefully while maintaining respect for the people who give a bit of their time to help our studies along. By watching and listening carefully, we can learn from what they do and how they do it. That, after all, is the point of conducting user studies.