Key lessons

Summary SS 2026 based on input by ten participants.

Topic 1 - Fundamentals

- Numerical data do not necessarily equate to quantitative research

  Qualitative research that provides numerical data (typically through counting) is still qualitative research, even if it looks like quantitative research. It's just transforming the qualitative responses into something to analyse and handle more easily.

- Research in social sciences: rather inductive

  Research in social sciences seems to differ quite from research in health care, where research tends to be quite linear (confirmatory, pre-structured; experimental design, study design pre-approved, ethics approval, treatment group and control group, outcome and description). In knowledge management, approaches are almost always inductive or even abductive.

- Abductive reasoning

  involves forming the most plausible explanation from incomplete observed evidence, working backward from an observation to a likely cause (e.g., assuming it rained because the grass is wet); however, in quantitative research, such explanations must still be systematically tested using rigorous methods.

- Validity is the cornerstone of meaningful research

  A study is only as good as the degree to which it measures what it claims to measure and draws conclusions that are truly warranted by the data.

  Research is a structured process that translates theoretical ideas into variables and hypotheses, tests them empirically, and evaluates the validity of the inferences drawn from the data.

  There are four types of validity which are conclusion, internal, construct, and external validity. Conclusion validity asks whether a relationship exists, internal validity asks whether it is causal, construct validity asks whether the constructs were properly operationalized, and external validity asks whether the findings can be generalized.

- Research is a systematic process

  Broad ideas must be transformed into testable questions through clear concepts, empirical observation and careful reasoning. Sound research depends not only on methods and data, but also on validity, ethical considerations and explicit assumptions about how the research can be conducted.

- Ecological fallacy (inferring from group to individual) and exception fallacy (inferring from exceptional cases to group)

  Always beware of these popular fallacies. Group differences tell very little, if anything, about individuals. In contrast, observations at the level of an individual should not be generalised to a group.

Topic 2 - Sampling:

- Importance of precise definition of total population (e.g., “clients served in the last 12 months” instead of just “clients”)

  Vague population definitions lead to biased samples, so researchers must precisely define who is included to ensure valid results.

- Importance of sampling strategy

  No statistical method, however sophisticated, can compensate for a flawed sampling strategy. The representativeness of your sample is what determines whether your findings generalize to the target population.

- External validity/generalisation

  Sampling is essential for external validity because it determines how well findings from a sample can be generalized to the target population. While probability sampling provides the strongest basis for generalization, nonprobability methods are often used in practice but require careful justification due to potential bias.

- Proximity Similarity Model

  An alternative but also complementary approach to the sampling model. A more pragmatic attempt at generalization.
  A random sample drawn from a specific total population does guarantee generalizability if the context differs.

- Non-probability sampling techniques

  Can still be a reasonable approach, e.g. when studying minority groups.

- Standard error

  A statistical concept expressing the variability in a sample statistic of interest (e.g. standard error of the mean). It is typically a combination of variation within the sample (its standard deviation) and sample size.

- Quality of sample

  The quality of sample determines the quality of the conclusions. A good sample allows generalization to a broader population, while a bad one undermines everything, no matter how well the rest of the study is done. Bias can enter at any point: a flawed sampling frame, poor selection procedures, or high nonresponse all damage conclusions. This is where external validity becomes central. It asks how far findings from one study can be generalized to other people, places, settings, and times. Importantly, external validity is not all-or-nothing. Generalization becomes weaker the more a new context differs from the original study. Even a technically sound sampling method does not guarantee strong external validity; the sample, setting, and timing all need to reflect the population of interest.

  Probability of sampling gives the strongest basis for generalization. Nonprobability sampling is often more practical but requires careful justification for why the sample is still meaningful. Neither is good or bad, what matters is whether the sampling decision honestly supports the conclusion or not.

Topic 3 – Measurement:

- Good measurement is essential for construct validity

  It ensures that theoretical constructs are properly translated into observable manifestations (importance of content validity).

- Measurement error, both random (e.g., individual mood) and systematic (e.g., loud traffic during data collection), can distort results

  When systematic, error can even shift the mean of a distribution, making it essential to minimize error through careful design, such as pilot testing, interviewer training, data checks, statistical adjustments, and the use of multiple measures (triangulation).

  True score theory highlights that all measurements contain error, which makes reliability important for capturing the consistency of a measure.

- The level of measurement (nominal, ordinal, interval, ratio) is not a mere technicality. *See also Survey Design*

  It directly constrains which statistical operations are permissible and meaningful. Any measurement instrument must be evaluated against reliability and validity to ensure it consistently and accurately captures the intended construct.

  The level of measurement is not always "intuitive," often not necessarily logically derivable, and often speculative. Variables to be measured may be quantitative in themselves, yet their measurement may only be performed at an ordinal level.

- Reliability and Validity

  Reliability and construct validity together determine the quality of measurements with different roles. Reliability refers to the consistency or repeatability of a measure, ensuring that results are stable and not driven by random error, while constructing validity concerns whether the measure actually captures the theoretical concept it is intended to represent, linking observations to underlying theory. Critically, reliability is a necessary but not sufficient condition for validity measures can be consistent yet systematically wrong, meaning it is reliable but not valid. Therefore, the most important insight is that validity is the ultimate goal, as it determines the meaningfulness of research findings, while reliability serves as its foundation by minimizing measurement error.

  Reliability is necessary but not sufficient for validity. A scale can be highly consistent but still measure the wrong concept. We need to connect the abstract constructs to observable data through theory.

- Evaluation of validity is a complex task

  Measurement quality is also demonstrated in terms of investigating how well observed patterns match theoretical expectations.
  There are many pieces of evidence for validity (e.g. convergent, discriminant validity).

Topic 4 – Survey Research:

- There is no one best method fitting all purposes

  Since each option (mail, online, phone, or face-to-face) has its own pros and cons. For example, face-to-face interviews can give more detailed and higher quality answers, but they are expensive, while online surveys are cheaper and efficient, but often have lower response rates and less depth.
  Matching the survey method to the research context is essential for obtaining valid and actionable data.
  -> trade-off; multiple methods used in one study (potential impact of methods/comparability of results?)

- Importance survey design in survey-based research (questionnaires and interviews to collect data); Importance of item wording (question, statement)

  How a question is worded, ordered, and scaled can systematically bias responses, making methodological rigor in survey construction just as critical as subsequent data analysis.

  Clear question wording, suitable response formats, well-trained interviewers, and careful implementation are key to getting valid and unbiased responses.

  Designing an effective survey is a complex, theory-driven process where question quality directly determines data quality. A good survey requires carefully deciding what to ask, how to ask it, and how to structure it. This includes choosing appropriate question types (e.g., nominal, ordinal, Likert scales), avoiding common errors such as double-barrelled or biased questions, and ensuring wording is clear and neutral. Additionally, the sequencing and placement of questions must guide respondents logically and comfortably, while the choice of survey method (e.g., questionnaire vs. interview) must consider population, bias, and practical constraints. Overall, the central insight is that poorly designed questions lead to biased or meaningless data, whereas well-designed surveys enable valid and reliable measurement of constructs.

- Survey design – types of question, level of measurement

  Important to consider carefully which types of question you are going to use, also based on the level of measurement. You cannot change the way questions are asked later (importance of qualitatively oriented pre-testing!).

  Good survey research depends on careful question design, appropriate response formats and sensitivity to the respondent's situation.

- Online survey as the nowadays predominant type of survey

  Since Trochim and Donnelly published their book and launched their platform, online surveys have become the most prevalent type of survey. However, this comes at a price. Potential participants are inundated with survey requests. As a result, achieving satisfactory response rates can be very difficult. To increase response rates, it might be worth reconsidering the tried-and-tested method of using pen and paper.

Topic 5 – Scaling and Indexes:

- Purpose of scales and indexes

  Scales and indexes are used to measure abstract constructs by combining multiple indicators into a single score, since no single item can capture a complex concept on its own. While indexes bring together different components of a construct, scaling focuses on turning qualitative concepts into something that can be measured.

- Item purification required: Likert scaling: not all statements are suitable for scale construction. Items must be able to differentiate between respondents with different attitudes.

  Statements which everyone answers the same way (e.g., everyone agrees) are "undifferentiating" and should be removed, because they do not contribute to accurately measuring the attitude. *(Note:such items may still be useful for other subpopulations than the one under study)*

- Terminology matters: Likert-Scaling versus Likert Scale (named after Rensis Likert)

  Likert-scaling is a scaling technique, an approach to develop a scale/measurement instrument. The result is a Likert Scale. However, today, in most cases, the term Likert scale is used to denote the type of response scale (mostly agree to disagree, which is in line with Rensis Likert's ideas at least; but more and more used to imply any type of polytomous/multicategorical response scale irrespective of category characteristics).

- Selection of scaling technique/theory

  Thurstone, Likert, and Guttman scaling each offer a different set of assumptions about how latent constructs translate into observable responses — choosing the right approach shapes the validity and interpretability of the resulting scale.

- Scales versus indexes

  Scales and indexes are both tools for representing constructs with a single numerical score. However, their quality depends strongly on how clearly the construct is defined, how explicitly items or components are combined and whether the underlying rules and dimensional assumptions are appropriate. A scale requires a formal scaling procedure, while an index depends on transparent and theoretically justified rules for combining components.
  Note: scales are used to measure a single dimension (a complex construct may require multiple scales to capture all its dimensions). An index summarises different variables, it does not need to be unidimensional and in many cases is not. While scales are compatible with the notion of an objective reality (critical realism, ontological claim), indexes are more compatible with constructivism. Indexes serve a purpose, they do not necessarily quantify a latent variable that "exists" per se.

- Evaluate versus descriptive assessments

  In personality perception but also perception of national characteristics, Peabody (1970) proposed an interesting approach using semantic differentials to disentangle evaluation and description. See:

Peabody, D. (1970). Evaluative and descriptive aspects in personality perception: A reappraisal. Journal of Personality and Social Psychology, 16(4), 639–646. https://doi.org/10.1037/h0030259

Topic 6 – Design:

- Causal inference

  The key idea is internal validity: whether observed effects can truly be attributed to the treatment rather than other factors. And this requires establishing temporal precedence, covariation, and ruling out alternative explanations. What really stands out is that many common designs (like single-group studies) are vulnerable to serious threats such as history, maturation, or regression to the mean, meaning that results can easily be misleading. Therefore, using strong designs, especially those with control groups and ideally random assignments, is crucial, because they reduce bias and help ensure that the observed reflects real causal effects rather than confounding influences.

- Regression to the mean can create the illusion of a treatment effect, even when no real effect exists

  Extreme groups tend to move closer to the average over time, which can be mistakenly interpreted as improvement. Always consider possibility of regression to the mean when selecting an extreme part of the sample.

- Importance of Internal validity

  The confidence that an observed effect is genuinely caused by the independent variable and not by confounding factors, is the central challenge of causal research design. Recognizing and systematically addressing threats to internal validity must happen at the design stage, not after data collection.

  Research design plays a central role in internal validity, as it determines whether observed effects can really be attributed to a treatment rather than to other explanations. In particular, single-group designs are highly vulnerable to such threats,

- Concept of endogeneity

  A major threat to conclusion validity. Whenever an observational design is used (no manipulation of independent variable by the researcher), endogeneity is a possibility. In a true experiment, the independent variable is exogenous (not explained by any other variable in the model). In case of endogeneity, the independent variable is related to the error term of a dependent variable implying an alternative flow of causality.

Topic 7 – Experimental design:

- Weaker versus stronger designs

  Weak designs, such as single-group studies, are highly vulnerable to threats like history, maturation, and regression to the mean, making results potentially misleading. Therefore, stronger designs, especially those using control groups and random assignment, are essential, as they reduce bias and allow for more credible identification of causal effects.

- Random assignment /True experiments (often referred to as RCT, randomized control trial) are considered the “gold standard” to test cause-and-effect relationships.

Random assignment as a key feature of true experiments, which creates comparable groups and allows to rule out alternative explanations.

The defining feature of a true experiment because it creates probabilistic equivalence between groups, making it the most powerful available tool for causal inference.

Factorial and block designs extend this logic to more complex questions while controlling for additional sources of variance.

- Random assignment versus random selection

  Important distinction. Random assignment is related to experimental design (how groups are formed) and of utmost importance in terms of internal validity. Random selection is related to sampling and therefore a matter of external validity (generalisation). Ideally, one implements both random assignment and random sampling. In many experiments though, random sampling is not a key issue.

- Different types of threat to internal validity

  Awareness of what can go wrong in an experiment (threat to internal/conclusion validity)/ what may compromise the results and the conclusions.

- Factorial designs/interactions

  Factorial designs enhance the signal, allow for considering more independent variables and/or more levels on each factor/independent variable.
  Typically analyses be means of ANOVA (analysis of variance), but regression analysis would also be possible. Oneway ANOVA can be used when there is only one factor (but at least three levels/groups; with merely two levels – i.e. two groups are compared – one may run a t-test for simplicity even though ANOVA would give you the same result). When there is more than one factor, a more general ANOVA procedure must be used. Then, interaction effects are also a possibility.
  A significant interaction effects means the effect of one factor cannot be inferred without considering levels of another factor.

- Enhance the signal or reduce the noise

  Two ways of increasing the chance to detect effects. Factorial designs aim to enhance the signal, whereas randomized block designs and covariance designs aim to reduce noise.

  Including co-variates may help identify an effect by explaining part of the otherwise unexplainable variation within the groups. It must be considered during the planning and design phase as data on the covariate(s) has to be collected.

Topic 8 – Quasi-experimental design:

- Quasi-experimental designs rely on design strategies other than random assignment, for example comparison groups or pre-post measures to approximate causal inference

  Quasi-experiments differ from true experiments in that they do not use random assignment, which makes causal inference more difficult. As a result, they are more vulnerable to selection threats to internal validity and use alternative design strategies, such as nonequivalent-groups or regression discontinuity, to deal with these issues as well as possible.

When randomization is ethically or practically impossible, quasi-experimental designs such as the nonequivalent groups design or regression-discontinuity design offer a principled approximation, but they demand explicit acknowledgment and management of the alternative explanations they cannot rule out.

The switching-replication design is both meaningful and ethically justifiable, as it ensures that every participant (who does not drop out) can receive treatment.

- Power of regression-discontinuity design

  The regression-discontinuity design is a strong quasi-experimental design that even enables assignment based on a selected criterion before treatment instead of aiming for comparable groups.

- Causal inference without randomization

  Causal inference without randomization relies heavily on assumptions. The real challenge is not applying the method but defending its assumptions.

Topic 9 – Analysis Part 1: Data Preparation & Descriptive Statistics, Fundamentals of Hypothesis Testing paradigm

- Importance of careful data preparation and appropriate choice of descriptive statistics

  Matching central tendency and dispersion measures, as well as association measures to scale level (and other requirements), is a prerequisite for any trustworthy analysis, since errors or oversights here propagate through every subsequent step.

- There is always a trade-off between Type I and Type II errors. Consider matrix of decisions x scenarios of reality

  When trying to be very strict and reducing the chance of finding a false effect (Type I error), researchers increase the risk of overlooking a real effect (Type II error). Therefore, a compromise between these two types of mistakes is necessary.

  Data analysis is often challenged by noise, limited statistical power, and the risk of incorrect conclusions (Type I and Type II errors). To avoid misleading results, it is important to carefully examine the data using descriptive statistics and visualizations to better understand the underlying patterns.

- Correlation coefficients (e.g. Pearson's coefficient)

  Quantify the strength and direction of a linear relationship between two variables.
  Check requirements of a coefficient (scale level, type of relationship, normal distribution, etc.); beware of overinterpreting an association as a causal relationship (which requires appropriate research designs and not just the calculation of a correlation coefficient)

- Do not over-rely on a p-value

  p-values decide whether a difference (association, etc.) in the data is statistically significant.
  However, significance does not necessarily imply relevance (a small true effect may become significant in a big enough sample event though it has no practical relevance).
  Power (1 – type two error) should be considered (requires determination of an effect deemed relevant).

Additional evidence such as effect size (Cohen's d) helps assess the size of the effect relative to variation within the groups. Confidence intervals may also be informative. *Never "reverse" the probability that your data/difference is compatible or not compatible with chance (type one and type two error) to signal how likely your theory is true or not true. A single study tells very little about the theory being true. Replications are needed. Also note that a p-value only tells you how likely the data are given a hypothesis being true, which is perhaps very unlikely to begin with (the null-hypothesis). Why not calculate another p-value telling to whether the data are compatible with a specific difference (defined before analysing the data!)?*
Alternatives such as "second generation p-values" have also been proposed recently.

Topic 10 - Analysis Part 2: Inferential Statistics

- Statistical conclusions depend heavily on how the model is specified.

  Even with the same data, different assumptions (e.g., linear vs. nonlinear relationships) can lead to different results, which means that incorrect model choices can produce misleading conclusion, even when the data itself is correct.

  Statistical models are extensions of research design assumptions and statistical significance cannot compensate for poor design or model misspecification.

  The idea of "letting the data speak for themselves" is not really working for data rarely speak for themselves. Researchers make assumptions, include or exclude effects in a model, ask questions or do not ask a question, etc. We invoke rationalism as well as positivism (post-positivism, critical realism)

- Inferential statistics (t-tests, ANOVA, regression) allow conclusions to be extended from a sample to the broader population

  But their validity rests on meeting the underlying assumptions and on a research design that warrants causal or associative claims. Dummy variable coding is the practical bridge that lets categorical group membership enter regression models alongside continuous predictors.

  Conclusions depend on choosing an analysis that fits the research design. Different designs create different sources of variation, so the right method helps separate the treatment effect from noise or baseline differences, for example, when comparing two group means versus controlling for initial differences.

- Dummy variables/Dummy coding

  It can be used to differentiate between subgroups (in principle, nominal scale level) in regression analysis, e.g. a dummy variable can represent group assignment in a two-group design by having the value 1 for the treatment group and 0 for the control group. *Binary variables, in a certain sense, defy the determination of a scale level.*

- Statistical analysis and research design

  There is a close relationship between statistical analysis and research design. Different research designs require different analytical models of comparison and control.

- Use it or lose it

  One’s analytical skills (knowing which test to conduct, how to do it and interpret it, e.g. in SPSS) must be applied continuously. Otherwise, they could be lost.

  But don't worry. They'll return, and perhaps be even more resistant to being forgotten.