

Chapter 8. Causal Research Design: Experimentation

Overview

Concept of Causality

; A causal inference relates to whether a change in one marketing variable produces a change in another variable.

- ; 인과적 추론(causal inference)은 하나의 마케팅 변수가 또다른 변수의 변화를 낳는지의 여부에 관계된다.
- ; 예를 들어 10%의 가격 인상이 매출에 영향을 미치는가를 조사하는 상황 등에서 쓰일 수 있다.

; Causality (인과 관계)

- ; When the occurrence of X increases the probability of the occurrence of Y.
- ; 사건 X의 발생으로 인해 사건 Y의 발생 확률이 증가하였을 때
- ; 다만 사람들이 그냥 쓰는 말과 과학자들이 쓰는 표현에는 차이가 있음에 유의하여야 한다.

Conditions for Causality (인과 관계 조건)

; 2개 변수간의 인과 관계 추론을 정당화 하기 위해서는 적어도 3가지의 아래 조건이 만족되어야 한다.

(1) Concomitant Variation (부수 변이)

; A condition for inferring causality that requires that the extent to which a cause, X, and an effect, Y, occur together or vary together as predicted by the hypothesis under consideration.

; 부수 변이는 이는 가정된 원인(presumed cause) X와 가정된 결과(presumed effect) Y가 가설에 의해 예측된대로 나타나고 변화될 때 생기게 된다.

; 예를 들어 매장 매출을 조사하는 시장 연구자가 조사 결과 매장 성과(performance)가 매출에 대한 원인(cause)라고 결론내렸다. 그것이 가능한 대답인가? 아니다. 매출에는 매장의 위치라든지 경쟁의 여부라든지 여러 다른 요인들(factors)이 존재하기 때문이다. 다만 성과와 매출량 사이에 연관성(association)이 존재한다는 말만을 할 수 있는 것이다.

; Example: Greater sales volume is due to preparation at a private school. Does school (X) cause sales (Y)?

(2) Time Order of Occurrence of Variables (변수 발생의 시간 순서)

; For one variable to hypothetically cause another, it must precede or occur simultaneously with the effect.

; 변수의 시간 발생은 직관적으로 볼 때 명백한 요구사항으로 보이지만, 사실 마케팅 상황에서 시간 순서를 결정하는 것이 항상 쉬운 일만은 아니다. 예를 들어 백화점에서 쇼핑을 자주 하는 사람이 Charge card(특정 점포에서만 사용하는 신용카드)를 가지고 있는 것인지, 아니면 특정 백화점의 Charge Card를 가지고 있는 사람이 쇼핑을 자주 하는 것인지 알아내기는 쉽지 않다. 이 경우 변수들의 시간 순서(Charge Card 소유 여부와 자주 쇼핑하는 것)는 명확하지 않다.

(3) Absence of Other Possible Causal Factors (다른 가능한 인과 요인들의 제거)

; The presence of additional or extraneous variables that impact the effect variable must be controlled in order to draw causal inferences.

Role of Evidence (증거의 역할)

; 위의 3가지 조건이 모두 만족된다고 하여 인과 관계가 존재한다는 의미는 아니다. 하지만 일관성있는 증거들이 축적되면 될수록 인과 간계가 존재한다는 확신은 강해진다. 통제된 실험은 모든 3가지 조건에 대한 강력한 증거를 제시할 수 있다.

; Even if all three conditions are present, causality cannot be proven. However, if all the evidence is strong and consistent, it may be reasonable to conclude there is a causal relationship. Controlled experiments can provide strong evidence on the three conditions. - Malhotra

What is Experimentation? (실험이란?)

; It is the primary method for establishing cause-and-effect relationships in marketing.

; Experiment (실험)

; The process of manipulating one or more independent variables and measuring their effect on one or more dependent variables, while controlling for the extraneous variables.

Independent Variables (독립 변수)

; Variables that are manipulated by the researcher and whose effects are measured and compared.

; 이러한 변수들은(treatment로도 알려져 있다) 가격 레벨, 패키지 디자인, 광고 테마 등을 포함할 수도 있다.

Test Units (실험 단위)

; Individuals, organizations, or other entities whose response to independent variables or treatments is being studied.

; 실험 단위는 소비자, 매장 및 지리적 지역을 포함할 수 있다.

Dependent Variable (종속 변수)

; Variables that measure the effect of the independent variables on the test units.

; 이 변수에는 매출, 이익 및 M/S가 포함될 수 있다.

Extraneous Variables (외부 변수)

; Variables, other than the independent variables, that influence the response of the test units.

; 외부 변수에는 매장 크기, 매장 위치 및 경쟁사의 노력 등이 포함된다.

Random Assignment to Experimental and Control Groups

; Random Assignment (무작위 할당)

; Involves randomly assigning test units to the experimental and control groups and is one of the most common techniques used to control for the effect of extraneous variables on the dependent variable.

; 나이, 수입, 혹은 브랜드 선호도와 같은 외부 변수의 영향을 최소화하는 것에 목적이 있다.

; Experimental Group (실험 집단)

; The group exposed to the manipulated independent variable.

; 실험 집단의 결과는 다른 정도의 조작을 가한 실험집단이나 통제 그룹과 비교할 수 있다.

; Control Group (통제 집단)

; This group is not exposed to the independent variable manipulation. It provides a point of comparison when examining the effects of these manipulations on the dependent variable.

Experimental Design (실험 설계)

; The set of experimental procedures specifying (1) the test units and sampling procedures, (2) independent variables, (3) dependent variables, and (4) how to control the extraneous variables.

Definition of Symbols

; X = the exposure of a group to an independent variable, treatment, or event, the effect of which are to be determined.

; O = the process of observation or measurement of the dependent variable on the test units or group of units.

; R = the random assignment of test units or groups to separate treatments.

; ① 왼쪽에서 오른쪽으로의 이동은 시간의 흐름을 나타낸다.

- ② 수평으로 배치된 기호들은 특정한 Treatment 집단임을 의미한다.
- ③ 수직으로 배치된 기호들은 동시에 발생한 사건이나 행동을 의미한다.
- 예제
 - 즉 X O₁ O₂는 처음에 한 번 treatment에 노출되었고, 그 이후 2번 측정(measure)되었다는 것을 의미한다.
 - R X₁ O₁
 - R X₂ O₂
 - 의 경우에는 2개의 다른 treatment에 각각 집단이 노출되었고 동시에 측정되었다는 것을 의미한다.

Validity in Experimentation (실험의 타당성)

실험을 수행시에 조사자는 2가지 목표를 가진다. 첫째는 독립 변수가 종속 변수에 미치는 영향에 대한 명확한 결론을 내는 것이다. (이는 내적 타당성이라 불린다) 둘째는 특수한 실험적 환경을 좀 더 많은 인구로 가져가서 타당하게 일반화를 시키는 것이다. 이는 외적 타당성이 성취되어야만 만족될 수 있다.

(1) Internal Validity (내적 타당성)

- A measure of accuracy of an experiment. It measures if the manipulation of the independent variables, or treatments, actually caused the effects on the dependent variables.
- Internal validity is threatened with the influences of extraneous variables are mixed with the independent variables. Without proper control of the extraneous variables, the researcher is unable to isolate the effect of the independent variable and thus cannot establish internal validity.

(2) External Validity (외적 타당성)

- A determination of whether the cause/effect relationships found in the experiment can be generalized.
- 만약 일반화 될 수 있다면 어떠한 인구 구성, 환경, 시간, 독립 변수 및 종속 변수들이 있어야 예측된 결과를 낼 수 있겠는가?
- 일반화시킬 수 없는 비현실적인 환경에서 이루어진 실험일 경우 외적 타당성에 대한 위험이 발생하게 된다. 실험실에서 이루어진 실험은 현장에서 이루어진 실험보다 외적 타당성이 부족하기 쉽다.

Extraneous Variables (외적 변수들)

① History

- Specific events that are external to the experiment but occur at the same time as the experiment.
- 예를 들어 실험을 수행하던 기간 중의 경제적 환경이 점차로 침체되어(decline) post-treatment 관측에 영향을 미칠 수 있다. 측정 사이의 시간 간격이 길수록 history가 실험에 영향을 미칠 가능성도 증가한다.
- Examples: Labor strikes, Stock-outs, Bad weather

② Maturation (성숙)

- An extraneous variable attributable to changes in the test units themselves that occur with the passage of time.
- 사람과 관련된 실험에서 maturation(성숙)은 사람들이 나이를 먹거나 보다 경험을 쌓거나 피로해지거나 지루해지거나 무관심해 지면서 발생한다. 몇 개월간에 걸친 추적과 시장 조사는 응답자들이 시간이 지나감에 따라 어떻게 변하는지 알기 어렵기 때문에 maturation 해지기 쉽다.
- Examples: growing older, hungrier, more tired

③ Testing Effects (시험 효과)

- The influence on respondent behavior of taking a measure on the dependent variable before and after the presentation of the treatment.
- 예를 들어 응답자는 전-처리(Pre-treatment) 및 후-처리(Post-treatment) 태도 사이에 일관성을 유지하려고 노력할 수 있다. 이러한 시험 효과로 인해 post-treatment 태도는 pre-treatment 태도에 비해 treatment 자체에 의해 보다 많이 영향 받는다.
- Example: ad awareness and prior measures

④ Instrumentation (도구)

- An extraneous variable involving changes in the measuring instrument, in the observers, or scores themselves.
- 예를 들어 매장 내 진열 (treatment)을 노출시키기 전과 후의 달러 매출을 측정하는 실험을 생각해 보자. 만약 측정들 간에 비일치적인 가격 변화가 있었다면 이는 instrumentation의 변화를 가져오게 되는데 이는 다른 가격 단위로 매겨지게 되기 때문이다.
- Examples: changing a scale from 5 to 7 points; is 2.5/5 equal to 3.5/7? Or, how can large scale experiments be conducted by one person?

⑤ Statistical Regression (통계적 회귀)

- An extraneous variable that occurs when test units with extreme scores move closer to the average score during the course of the experiment.
- 광고 실험에서, 어떤 응답자들은 매우 호의적이거나 매우 비호의적인 태도를 가졌다고 해 보자. 그런데 후-처리(Post-treatment) 실험에서 이들은 평균을 향하려는 경향을 가질 수 있다. 사람들의 태도는 연속적으로 변한다. 극단적인 태도를 가진 사람은 보다 변할 가능성이 많다. 이는 실험 결과를 혼동시키게 되는데, 관측된 효과가 처리(treatment) 자체보다는 통계적 회귀에 귀인하기 때문이다.
- Comment: what goes up must come down

⑥ Selection Bias (선택 편차)

- An extraneous variable attributable to the improper assignment of test units to treatment conditions.
- This bias occurs when selection or assignment of test units results in treatment groups that differ on the dependent variable before the exposure to the treatment condition.
- 만약 test units이 자기 자신의 그룹을 재귀 선택(self-select)하거나, 연구자의 판단에 근거하여 집단을 편성(assignment)하였을 경우 선택 편차가 나타날 수 있다. 예를 들어 각기 다른 2개의 매장에 각기 다른 매장 진열(merchandising display)(중고품과 신품)에 대해 구매 실험을 한다고 하여 보자. 이 2개의 집단은 시작점부터 동일하지 않을 수 있다. 이는 매장 크기와 같은 핵심 특성에 따라 달라질 수 있기 때문인데, 매장 크기는 매장 진열(merchandising display)에 관계없이 매출에 영향을 끼치기 때문이다.
- Comment: common among “natural” field experiments

⑦ Mortality (실패율)

- An extraneous variable attributable to the loss of test units while the experiment is in progress.
- 이는 많은 이유로 인하여 나타나는데, 예를 들어 실험 단위(test unit)가 실험을 계속하는 것을 거부하는 것과 같이 나타날 수 있다. 실패율은 결과를 혼동시키는데 이는 lost unit(즉 거부한 사람들)들이 남아 있는 사람들과 동일한 방법으로 처리(treatment)에 대해 응답했는지를 알아내기가 어렵기 때문이다.
- Examples: boredom, move away, death

Controlling Extraneous Variables (외부 변수 통제)

외부 변수들은 그것이 통제되지 않을 경우 실험의 내적/외적 타당성에 대해 심각한 위험이 된다는 점에서 confounding variable이라고 불리기도 한다.

(1) Randomization (무작위화)

- One method of controlling extraneous variables that involves randomly assigning test units to experimental groups by using random numbers. Treatment conditions are also randomly assigned to experimental groups.
- 예를 들어 응답자는 3개의 실험 집단 중 하나에 무작위로 할당(assign)될 수 있다. 이 때 3가지 버전 중의 하나의 시험 광고를 무작위로 선택해서 각 그룹에 제시한다. 무작위 할당의 결과로 외부 변수들은 각 처리 조건에 동일하게 나타난다. 또한 무작위화가 외부 변수의 측정에 영향을 주었는지를 확인하고 다른 실험 집단들과 비교하는 것이 가능하다. 따라서 무작위화는 외부 변수를 통제하는데 가장 많이 선호되는 방법이다.

(2) Matching (매칭)

- One method of controlling extraneous variables that involves matching test units on a set of key

background variables before assigning them to the treatment conditions.

; 매장 진열 실험에서, 매장들은 연간 매출, 크기, 장소에 따라 매치될 수 있다. 그리고 각 매치된 짝에서 하나의 매장이 각 실험 그룹에 할당될 수 있다.

; 단점

; ① Test units can be matched on only a few characteristics, so the test units may be similar on the variables selected but unequal on others.

; 실험 단위는 몇몇 특성에 대해서만 매치될 수 있기 때문에, 선택된 변수에 대해서는 유사하나 그렇지 않은 것에 대해서는 동등하지 않을 수 있다.

; ② If the matched characteristics are irrelevant to the dependent variable, then the matching effort has been futile(무의한).

; 만약 매치된 특성들이 종속 변수에 무관계한 것이라면, 매칭 효과는 없어진다.

; (3) Statistical Control (통계적 통제)

; One method of controlling extraneous variables by measuring the extraneous variables and adjusting for their effects through statistical methods.

; (4) Design Control (설계 통제)

; One method of controlling extraneous variables that involves using specific experimental designs.

A Classification of Experimental Designs (실험 설계의 분류)

; (1) Preexperimental Designs (예비 실험 설계)

; Designs that do not control for extraneous factors by randomization.

; 외부 인자 통제를 위해 무작위화를 사용하지 않는다. 따라서 이들은 많은 내적/외적 타당성의 위험을 받는다. 하지만 이러한 한계를 적절하게 note 함에 따라 탐색적 방법으로 사용될 때 그 가치를 더할 수 있다.

; (2) True Experimental Designs (실제 실험 설계)

; Experimental designs distinguished by the fact that the researcher can randomly assign test units to experimental groups and also randomly assign treatments to experimental groups.

; 여기에서 연구자들은 과제와 실험 집단을 임의로 할당한다. 따라서 이러한 설계는 외부 변수들에 대한 높은 통제도를 가지게 된다.

; (3) Quasi-experimental Designs (유사 실험 설계)

; Designs that apply part of the procedures of true experimentation, while lacking full experimental control.

; 이는 연구자가 독립 변수나 처리(treatment)를 완전하게 조작할 수 없으나 실제 실험 장치(apparatus)의 일부를 적용할 수 있을 때 사용된다. 이 설계는 보통 자연적인 환경에서 실험을 통제해야 할 필요가 있을 때 사용된다.

; (4) Statistical Design (통계적 설계)

; Design that allows for the statistical control and analysis of external variables.

; 통계적 설계는 그들의 특성과 사용에 따라서 분류된다.

Preexperimental Designs (예비 실험 설계)

① One-Shot Case Study (단발 사례 연구)

; A preexperimental design in which a single group of test units is exposed to a treatment X, and then a single measurement on the dependent variable is taken.

; X O₁

; 이러한 형태의 설계는, 주제가 자기-선택적(self-selected)이거나 연구자에 의해서 임의로 선택되었을 경우 non-random sampling process를 사용하도록 구성된다. 무작위화 없이 관측된 종속 변수는 몇몇 외부 변수들에게 영향을 받게 된다.

; 또한 이 설계 방식은 통제 집단(control group)이 없다. 통제 집단이 없기에 결과를 비교할 수 없다. 무작위화와 통

제 그룹의 부재로 인해, 이 설계 방식은 내적 타당성을 갖추기 힘들다. 여기에는 역사(history), 성숙(maturation), 선택(selection), 및 실패(mortality)에 대한 통제가 없다. 이러한 이유로 단발 사례 연구는 결론적(conclusive) 연구보다 탐색적(exploratory) 연구에 보다 적절하다.

② One-Group Pretest/Posttest Design (단일 집단 전실험/후실험 설계)

; A preexperimental design in which a group of test units is measured twice.

; O₁ X O₂

; Treatment effect : O₂ - O₁

; 여기에도 비교를 위한 통제 집단은 없다. 처리 효과(treatment effect)는 O₂ - O₁으로 계산된다. 비록 이 설계 방식이 사례 연구로서 보다 많이 고려되지만, 무작위화와 통제 집단이 없음으로 인한 외부 변수들로 인해 결론의 타당성에는 의심의 여지가 있다. 역사(history), 성숙(maturation), 실험(testing), 도구(instrumentation), 선택(selection), 실패(mortality), 및 회귀(regression)이 나타날 수 있다.

③ Static Group Design (정적 집단 설계)

; A preexperimental design in which there are two groups; the experimental group (EG), which is exposed to the treatment, and the control group (CG). Measurements on both groups are made only after the treatment, and test units are not assigned at random.

; EG: X O₁

; CG: O₂

; Treatment effect : O₁ - O₂

; 과제는 임의로 할당되지 않으며, 양 집단 모두 처리 후 측정(posttest)을 받게 된다. 처리 효과(treatment effect)는 통제 집단과 실험 집단의 차이인 O₁ - O₂로 계산된다. 무작위화의 부족은 몇몇 외부 효과에 실험이 노출되는 결과를 가져온다. 2개의 집단은 처리 이전에 다를 수 있기 때문에 선택 편차(selection bias)를 가져올 수 있다. 실패(Mortality) 효과도 나타날 수 있는데, 통제 집단에 비해 실험 집단에서 더 많은 실험 단위가 그만둘 수 있기(withdraw) 때문이다. ; 실제로 통제 집단은 아무런 처리를 받지 않은 집단이라기 보다는, 현 단계의 마케팅 활동을 받은 집단들로 정의된다. 많은 경우에서 마케팅 임력(가격과 같은)을 0으로 낮추는 것은 불가능하다.

True Experimental Design (실제 실험 설계)

① Pretest/Posttest Control Group Design (전실험/후실험 통제 집단 설계)

; A true experimental design in which the experimental group is exposed to the treatment but the control group is not. Pretest and posttest measures are taken on both groups. Test units are randomly assigned.

; EG: R O₁ X O₂

; CG: R O₃ X O₄

; Treatment Effect(TE) : (O₂ - O₁) - (O₄ - O₃)

; 과제는 각 실험/통제 집단에 임의로 할당된다. 각 집단에 대해 전처리 측정(pretreatment measure)이 행해진다. 또한 각 집단은 실험 집단에 대한 처리를 가하기 이전에 측정된다.

; 통제 집단의 사용과 무작위화는 대개의 외부 변수들을 통제 가능하게 한다. 외부 효과는 통제 및 실험 집단 양측에 균등하게 나타날 것으로 가정된다. 통제 및 실험 집단의 차이는 오직 처리(treatment)만을 반영하는 것이라고 생각할 수 있다.

; 전실험 측정을 사용함으로써 후실험 측정 때 interactive testing effect(상호 실험 효과)가 나타날 수 있다.

; Interactive Testing Effects (상호 실험 효과)

; Effect in which a prior measurement affects the test unit's response to the independent variable.

② Posttest-Only Control Group Design (후실험-유일 통제 집단 설계)

; A true experimental design in which the experimental group is exposed to the treatment but the control group is not and no pretest measure is taken. Test units are randomly assigned.

; EG: R X O₁

; CG: R O₂

; Treatment Effect(TE) : O₁ - O₂

; 후실험-유일 통제 집단 설계 방법은 어떠한 선행측정(premeasurement)을 포함하지 않는다.

; 이 설계 방법은 간명하기(simplicity) 시간(time), 비용(cost), 표본크기(sample-size) 측면에서 장점이 있다. 이러한 이유로 이는 마케팅 리서치의 실험 설계 중 가장 널리 쓰이는(popular) 것이다. 그러나 이 설계 방식에도 제한은있다. 비록 집단을 동질화하기 위해 무작위화를 사용했을 경우라도 이 집단의 동일성을 검증할 방법이 없다. 선택 편차(selection bias)와 실패율(mortality)의 문제가 나타날 수 있다. 선행측정이 없다는 것을 제외하고는 이는 전실험/후실험 통제 집단 설계와 매우 유사하다.

Quasi-Experimental Designs (유사 실험 설계)

; 유사 실험 설계는 조사자가 무작위화를 수행할 수 없거나 실험 처리 일정을 통제할 수 없는 상황에 적절한 방법이다. 그러나 실험자는 언제 그리고 누구에게 실험 측정을 하는지는 통제할 수 있어야 한다. 유사 실험 설계는 다른 실험 방법에 비해 빠르고 저비용의(less expensive) 방법이며, 특정 조사 환경에서는 자료를 수집하는 유일한 방법일 수도 있다. 하지만 전체적인 실험 통제가 부족하므로 연구자는 통제되지 않는 특정 변수 역시도 고려하여야 한다.

① Time Series Design (시간 연속적 설계)

; **A Quasi-experimental design that involves periodic measurements on the dependent variable for a group of test units. Then, the treatment is administered by the researcher or occurs naturally. After the treatment, periodic measurements are continued in order to determine the treatment effect.**

; O₁ O₂ O₃ O₄ X O₅ O₆ O₇ O₈ ...

; 이것은 연구자가 처리하고자 하는 대상에 대해 무작위화를 하는 등의 통제를 할 수 없기에 유사 실험이다. 또한 연구자는 처리 표시(presentation)의 시간 간격 통제를 할 수 없을 수도 있다. 여러번의 측정은 시험 효과(testing effect)를 가져올 수 있다. 역사(History)도 통제될 수 없다. 그럼에도 시간 연속적 설계는 부분적으로 자연적 환경에서 행동을 평가할 때 유용하다.

; Multiple Time Series Design (복수 시간 연속적 설계)

; **A time series design that includes another group of test units to serve as a control group**

; 이는 시간 연속적 설계와 유사하나 반복적으로 측정되는 통제 집단이 추가된다는 점이 다르다. 이 그룹은 실험 처리를 받지 않는다. 만약 통제 집단을 주의깊게 선정한다면, 이 설계 방식은 간단한 시간 연속적 실험을 전반적으로 개선할 수 있다.

Statistical Designs (통계적 설계)

; 통계적 설계는 외부 변수의 통계적 통제와 분석을 가능하게 하는 일련의 기본적인 실험으로 이루어져 있다. 다른 말로, 몇 개의 기본적인 실험들이 동시에 행해진다는 것이다. 따라서 통계적 설계는 사용되는 기본적인 설계에 영향을 미치는 무효(invalidity)의 동일한 원천에 의해 영향받게 된다.(무슨 소리야?)

; 통계적 설계의 이점:

- ① 하나 이상의 독립 변수들의 영향이 측정될 수 있다.
- ② 특정한 외부 변수가 통계적으로 통제될 수 있다.
- ③ 각 원인(subject)이 한 번 이상 측정될 때 경제적 설계(economical design)를 공식화할 수 있다.

Factorial Design (인자 설계)

; **A statistical experimental design used to measure the effects of two or more independent variables at various levels and to allow for interactions between variables.**

; 2개 이상 변수들의 동시적인 효과가 각각의 분리된 효과들의 합과 다를 때 상호성(interaction)이 나타났다고 한다.

; 예를 들어 한 사람이 커피를 좋아하고, 그 사람이 좋아하는 온도는 차가운 것이지만 그것이 그 사람이 차가운 커피를 좋아하지는 않는다는 것이다.

; 인자 설계는 표로서 나타낼 수 있다. 2-인자 설계에서 한 변수의 각 레벨은 row로 나타나고, 또다른 변수의 각 레벨은 column을 나타낸다. 인자 설계는 처리 변수의 가능한 모든 조합에 대한 각각의 셀을 포함한다.

; Analysis of variance(분산 분석)과 같은 통계적 방법이 처리 효과 및 상호성을 분석하기 위해 쓰인다. 인자 분석의 가장 큰 단점은 변수나 레벨의 수가 증가함에 따라 가능한 처리 조합의 수가 기하급수적(multiplicatively)으로 증가한다는 것이다. 그러나 조사자가 변수의 수와 레벨을 통제할 수 있을 때 그다지 큰 제약은 아니다.

; 조사 방법을 선택한다는 것은 통제에 대한 trade-off를 의미한다. 보통 높은 내적 타당성(internal validity)을 제공하는 설계는, 실험 결과의 일반성이나(generalizability) 외적 타당성(external validity)을 위협하는 높은 인공적인 환경에서 수행된다.

; 내적 및 외적 타당성의 최적화된 조합을 찾는 방법 가운데 하나는 다른 연구 시점에 다른 실험 설계를 사용하는 것이다. 예를 들어 높은 내적 타당성을 필요로하는 설계는 초기에 쓰일 수 있다. 이러한 방법으로 실제 처리 효과를 측정하는 보다 신뢰성 있는 방법을 얻을 수 있다. 연구의 나머지 단계에서는 결과를 일반화할 수 있는 보다 자연적인 환경을 사용할 수 있다.

Laboratory Versus Field Experiments

; Laboratory Environment (실험실 환경)

; **An artificial setting for experimentation in which the researcher constructs the desired conditions.**

; Central-location theater에서 광고 시험과 주방 실험하는 것은 실험실 실험의 일종이다. 왜냐하면 매우 인위적인 환경이기 때문에 외적 타당성의 필요가 제기된다. 현장 실험에 비해 연구실에서의 인위적인 환경은 응답자들이 독립적 변수보다는 상황 그 자체에 반응하는 반응 오류(reactive error)를 발생시킬 수 있다. 또한 실험실 환경은 demand artifact를 발생시킬 수 있다.

; 긍정적인 면으로 보자면, 실험실 실험은 현장 실험보다 복잡한 설계를 가능하게 한다. 이를 통해 보다 내적 타당성을 증대시킬 수 있다.

; Demand Artifacts (요구 인공물?)

; **Responses given because the respondents attempt to guess the purpose of the experiment and respond accordingly.**

; Field Environment (현장 환경)

; **An experimental location set in actual market condition.**

; 현장 환경(Field environment)은 행동(behavior), 태도(attitude), 지각(perception)의 측정을 수반한다. 조사자는 내적 타당성에 영향을 끼치는 외적 변수들에 대한 통제를 좀 더 줄여야 한다. 그러나 만약 내적 타당성이 유지된다면 실험실 환경에서 얻은 것보다 더 결과를 일반화 시킬 수 있다.

; 인터넷은 연구실 환경에서와 같은 통제된 실험이 가능한 메커니즘을 제공한다. 예를 들어 웹상의 광고 조사와 같은 경우, 각기 다른 사이트에 각기 다른 광고를 싣고 응답자들이 각기 다른 사이트에 들어가도록 한 다음에 그들의 인지도 변화를 조사할 수 있다.

; 인터넷 실험은 대안적인 웹 디자인의 효과성을 시험할 때도 쓰일 수 있다. 여러가지의 웹 디자인에 방문자들을 랜덤하게 노출시키고 그들의 마케팅 변수들을 고정해 놓은 다음에 구매도 변화를 측정하는 것이다. 그 경우 가장 효과적인 디자인을 회사의 웹 사이트로 결정하면 된다.

Factor	Laboratory	Field
Environment	Artificial	Realistic
Control	High	Low
Reactive Error	High	Low
Demand artifacts	High	Low
Internal validity	High	Low
External validity	Low	High
Time	Short	Long
Number of units	Small	Large
Ease of implementation	High	Low
Cost	Low	High

Selecting an Experimental Design

Limitations of Experimentation

(1) Time (시간)

； 많은 종류의 현장 실험은 적절한 시간이 있어야 정확해진다. 예를 들어 프로모션 캠페인이나 신제품 도입의 장기적 효과를 측정할 때 구매 행동은 다수의 구매 사이클을 통해 관측되어야 한다. 이러한 행태적 정보의 정확성은 시간이 지나면서 정확해진다. 이러한 부가적인 정확도는 제품 출시나 광고 캠페인 도입의 연기 비용에 견줄 수 있다.(trade-off)

(2) Cost (비용)

； 현장 환경에서의 신제품 조사는 매우 값비싸다. 이는 연구실 실험보다 훨씬 더 비싸게 든다. 신제품을 현장 시험하기 위해서 경영진은 직접적인 자료 수집이나 분석의 비용보다도 더 많은 것을 고려해야 한다. 제품은 제한된 규모에서 시작되어야 한다.(initiate) 제한된 기저 위에 매출 시점 관측 캠페인과 mass advertising이 개발되고 도입되어야 한다. 제한된 유통 채널도 열려야 한다. 테스트 마켓 실험은 대개 수백만 달러가 든다.

(3) Administration (통제)

； 외부 변수의 효과를 통제하는 것은 실험적 조사의 가장 핵심적인 부분이다. 연구실이 아닌 현장에서 조사를 수행할 때 원하는 정도까지의 통제에 다다른 것은 더 어려워진다. 현장 실험은 때때로 회사의 현재 진행중인 작업, 소매상/도매상의 협조 등등 때문에 더 어려워질 수 있다. 또한 경쟁자가 고의로 현장 실험 결과를 망쳐버릴 수도 있다.

Application: Test Marketing

； Test marketing (시험 마케팅)

； An application of a controlled experiment done in limited, but carefully selected, test markets. It involves replacing the planned national marketing program for a product in the test markets.

； Test markets (테스트 마켓)

； A carefully selected part of the marketplace that is particularly suitable for test marketing.

； 시험 마케팅의 2가지 목적 : ① 제품의 시장 수용도를 조사하기 ② 마케팅 믹스 변수들의 대안적인 단계들을 시험하기

； 표준 마켓 테스트를 설계하는 것은 어떠한 기준으로 테스트 마켓을 선택하고, 얼마나 많은 테스트 마켓을 사용하고, 얼마나 오래 시험할 것인가를 결정하는 것을 수반한다. 이러한 기준들은 모두 테스트 마켓을 선택할 때 고려된다. 시장의 대표성 (representativeness of market)은 가장 중요하다. 예를 들어 신제품 테스트에서 얻어진 정보들이 해당 제품의 궁극적인 시장 수용도를 예측하고 최종적인 제품 출시의 가/불가 결정에 사용될 수도 있다. 테스트 마켓은 테스트를 지원할만한 mass-market advertising이 커버 가능한 범위여야 한다. 이는 관측을 위해 돈을 낭비하는 것을 막는다. 추가적으로, 제한된 지역에서만 가능한 신제품 관측은 향후 출고에 대한 소비자의 부정적인 반응을 만들어 낼 수 있다.

； 대개 많은 테스트 마켓을 사용할수록 더 낫다. 자원이 제약되어 있더라도 적어도 2개의 테스트 마켓에서는 테스트되어야 한다. 하지만 외적 타당성이 중요한 문제라면, 적어도 4개의 테스트 마켓이 사용되어야 한다.

Summary Illustration Using the Opening Vignette

International Marketing Research

Technology and Marketing Research

Ethics in Marketing Research

； Debriefing (사후보고)

； After the experiment, the process of informing test subjects what the experiment was about and how the experimental manipulations were performed.

[Discussion]

1. An advocacy group wants to test the effectiveness of an abstinence education commercial. Two random samples, each of 250 respondents, are recruited in Chicago. One group is shown the abstinence education commercial. Then, attitudes toward abortion are measured for respondents in both groups.

a. Identify the independent and dependent variables in this experiment.

[[다시 한 번 풀어볼 것]]

independent : exposure to abstinence education commercial

dependent : person's attitudes toward abortion

b. What type of design was used?

[[다시 한 번 풀어볼 것]]

True experimental Design

； Definition of interesting commitment?

； reason of didn't pre-measure?

； randomization is good idea but what if we had a docteration? Measure change the belief.

； Reason for after-only. "MEASURE ATTITUDE OF CHANGE" is the really matters. pos-test only?

2. In the experiment just described, suppose the respondents had been selected by convenience rather than randomly. What type of design would result?

[[다시 한 번 풀어볼 것]]

Preexperimental, "static group"

3. State the type of experiment being conducted in the following situations.

a. A major distributor of office equipment is considering a new sales presentation program for its salespeople. The largest sales territory is selected, the new program is implemented, and the effect on sales is measured.

[[다시 한 번 풀어볼 것]]

Preexperimental

One-group Pretest?

b. Procter & Gamble wants to determine if a new package design for Tide detergent is more effective than the current design. Twelve supermarkets are randomly selected in Denver. In six of them, Tide is sold in the new packaging. In the other six, the detergent is sold in the old package. Sales for both groups of supermarkets are monitored for three months.

[[다시 한 번 풀어볼 것]]

Statiscal Design??

True experimental Design

[Lecture Note]

Binomial Test for “A Child’s Paper”

Ho: no difference, p = .5

Ha: positive difference, p > .5

a = .05 (95% confidence)

Reject Ho if Z > 1.645

$$Z = \frac{(x + .5 - np)}{\sqrt{(npq)}}$$

Where n = 288, p = .5, q =1-p = .5, x = 123

$$Z = \frac{((123 + .5) - (288 \times .5))}{\sqrt{(288 \times .5 \times .5)}}$$

$$Z = -19.5 / 8.5 = -2.3$$

Therefore, do not reject Ho

Example From Medical Research: Intranasal Lidocaine For Treatment Of Migraine

Objective: To evaluate effectiveness of treatment with lidocaine for migraine headaches
Design: Randomized, double-blind, placebo-controlled trial
RS O X O
RS O O
Patients: 53 received lidocaine; 28 placebo-controlled
Measure: Success determined by at least 50% reduction of headache within 15 minutes of treatment

	Successful	Not Successful
Treatment	29 (55%)	24 (45%)
Control	6 (21%)	22 (79%)

Results:

$$P(\chi^2 = 8.01, 1df) < .005$$

Chapter 9. Measurement and Scaling: Fundamentals and Comparative Scaling

Overview

Measurement and Scaling

- Measurement (측정)
 - The assignment for numbers or other symbols to characteristics of objects according to certain prespecified rules
 - 마케팅 조사에서, 숫자들은 둘 중 하나의 이유로 할당된다.
 - ① 생성된 자료의 통계적 분석을 허용하기 위해
 - ② 결과에 대한 정보 전달(communicate)을 돕기 위해
 - Measurement is the process of attaching numbers to objects or phenomena. That is, an unknown quality is compared with some standard of the same dimension.
- Scaling (척도)
 - The generation of a continuum upon which measured objects are located
- Data
 - Nonmetric or qualitative : Nominal scale, Ordinal scale
 - Metric or quantative : Interval scale, Ratio scale

Primary Scales and Levels of Measurement (기본 척도 및 측정 단계)

- (1) Nominal Scale (명목 척도)
 - A scale whose numbers serve only as labels or tags for identifying and classifying objects with a strict one-to-one correspondence between the numbers and the objects.
 - 명목 척도가 식별(identification)을 위한 수단으로 쓰이면, 측정된 대상과 할당된 각 숫자 사이에는 1대 1 관계만이 생성된다.
 - 명목 척도는 분류의 목적으로도 사용된다. 이 클래스들은 상호 배타적(mutually exclusive)이고 부분의 합이 전체를 구성(collectively exhaustive)하여야 한다.
 - Mutually exclusive : There is no overlap between classes; every object being measured falls into only one class.
 - 같은 class에 속한 것은 같은 번호를 가져야 하며, 어떠한 2개 클래스도 같은 번호를 가지지 않는다.
 - Collectively Exhaustive : All the objects fall into one of the classes.
 - 명목 척도에 할당된 번호들은 측정된 특성들의 상대적인 양을 반영하지는 않는다. 예를 들어 SSN이 높다고 우월한 측면이 있다는 것이 아니다. 이는 클래스에서도 마찬가지이다. 여성에게 1, 남성에게 2를 할당한다고 여성이 우월하다는 것은 아니다. 명목 척도의 숫자들은 오직 계수(count) 될 수만 있다. SSN을 평균내거나 응답자의 성별을 평균낸다고 의미있는 것이 아니다.

ex SSN, 미식 축구 선수의 번호, {male, female}, {European, American, Asian ancestry}

- Measurement at its weakest level.
- Numbers are assigned without numerical meaning.
- Some difference exists but no magnitude, direction, or quality implied.
- Statistics: chi-square analysis, Binomial test

Chi-square Test Example : Observed Product Usage by Gender

	Gender	
Product Usage	Male	Female
Light	50	10
Medium	80	70
Heavy	20	90

Ho: use of product is independent of gender

Ha: use and gender are related

let a = .05

Reject Ho if $P(\chi^2) < .05$

; **Binomial Test Example : Is Coke Preferred to Pepsi?**

Taste Coke and Pepsi:which do you prefer?

22/80 prefer Coke, and 58/80 prefer Pepsi

Ho: $P_{\text{coke}} = P_{\text{pepsi}} = .50$

Ha: the proportion is not .50

let a = .05

Reject Ho if $P(\text{binomial}) < .05$

Primary Scale	Basic Characteristics	Common Examples	Marketing Examples	Permissible Statistics
Nominal	Numbers identify and classify objects	SSN, numbering of football players	Brand numbers, store types, gender classification	Percentages, mode
Ordinal	Number indicate the relative positions of the objects but not the magnitude of differences between them	Quality rankings, rankings of teams in a tournament	Preference rankings, market position, social class	Percentile, median
Interval	Differences between objects can be compared; zero point is arbitrary	Temperature (Fahrenheit, Celsius)	Attitudes, opinions, index numbers	Range, mean, standard deviation
Ratio	Zero point is fixed; ratios of scale values can be computed	Length, weight	Age, income, costs, sales, market shares	Geometric mean, (All)

(2) Ordinal Scale (서열 척도)

; **A ranking scale in which numbers are assigned to objects to indicate the relative extent to which some characteristic is possessed. Thus, it is possible to determine whether an object has more or less of a characteristic than some other object.**

; 서열 척도는 상대적 태도(attitude), 의견(opinion), 지각(perception) 및 선호도(preference)를 측정하기 위해 사용된다. 이러한 형태의 측정 질문은 응답자에게 "...보다 큰", "...보다 작은"이라는 판단을 내리게 만든다. 응답자에게 다양한 스포츠의 선호도에 대해서 순위를 매기는 것이 이와 같은 경우이다.

; 서열 척도는 동등한 대상(equivalent object)이 같은 랭크를 받는다면 명목 척도와 동일하다. 대상 사이에서 서열화된 관계로 남아있는 한 어떠한 일련의 숫자로 할당될 수 있다. 이러한 점으로 인해 서열 척도는 대상의 기본적인 서열이 남아있는 한 어떻게든 변형될 수 있다. 서열 척도에서 숫자는 그들의 서열에 따라서만 달라지지 그들의 양에 따라서 달라지지는 않는다. 이러한 이유로 서열 척도에서 계수, 퍼센트, 평균 및 다른 요약 통계를 사용하는 것이 가능하다.

; **ex) Quality ranking, Rankings of teams in a tournament, education levels {LT high school, GT high school, college, ..}**

; **social class (upper > middle > lower)**

; **grades (a > b > c > d > e)**

; **Likert scale (strongly agree > agree > neutral > disagree > strongly disagree)**

; **Differentiates and orders; called the "ranking scale"**

; Characteristics:

; Items being measured are compared with each other, not some absolute standard; ; no absolute zero point

; Intervals between ranks are not equal

; Ordinal transitivity postulate applies: if a > b > c, then a > c

; Statistics: median, Spearman Rank Order Correlation

; **Spearman Rank Order Correlation Example**

Say social class and final grades in a course are measured for 5 students:

student	social class	grade
1	upper 3	a 4
2	lower 1	c 2
3	upper 3	d 1
4	middle 2	b 3
5	middle 2	e 0

Is there a correlation between social class and grades?

Ho: no correlation between social class and grades: $p = 0$

Ha: correlation: $p > 0$ or $p < 0$

let a = .05

Reject Ho if $P(p) < .05$

(3) Interval Scale (간격 척도)

; **A scale in which the numbers are used to rank objects such that numerically equal distances on the scale represent equal distances in the characteristic being measured.**

; 간격 척도는 서열 척도가 포함하는 모든 정보를 포함한다. 거기에 더해 이는 대상들 간의 차이 비교를 가능하게 한다. 1과 2의 차이는 2와 3의 차이와 같다. 시장 조사에서는 순위 척도로 얻은 태도에 대한 자료 (1=전혀 아니다 7=매우 그렇다)를 대개 간격 척도로 처리한다. 간격 척도에서는 각각이 동일한 외연에 대해 측정된 특성을 포함하고 있을 경우, 2개의 대상이 같은 숫자를 가질 수 있다.

; 간격 척도에서 0점(zero point)의 위치는 고정되어 있지 않다. 0점과 측정 단위는 임의적이다. 이는 온도의 측정에서와 같다. 화씨(Fahrenheit scale)와 섭씨(Celcius scale)는 각기 다른 0점과 다른 단위를 사용한다. 그러나 양쪽은 '온도'라는 같은 특성을 가지고 있다. 즉 $y=ax+b$ 의 식을 유지할 수만 있으면 척도의 특성도 유지된다. 여자 핸드백의 레이팅을 1, 3, 5로 매기든 22, 26, 28으로 매기든 같다는 의미이다.

; 0점이 고정되어 있지 않기 때문이 척도값의 비율을 취하는 것은 의미가 없다. 즉 Phoenix에서 90F 이고 Buffalo에서 60F 일 때 Phoenix가 Buffalo보다 1.5배 덥다고 말하는 것이 의미없는 것이기 때문이다.(Celcius scale로 변형되면 1

배가 되어 버린다.)

; 계수, 평균, 중간값 등에 더하여 대수적 평균, 표준 편차 등도 사용할 수 있다. 그러나 기하적 평균은 간격 척도에서는 의미없다.

; **ex) Temperature (Fahrenheit/Celcius)**
; IQ score
; Centigrade and Fahrenheit temperature
; “10 point scales of agreement” (0 through 9)

; Differentiates and orders to some standard units
; Represent differences not absolute quantities.
; Differences are isomorphic to arithmetic.
; No absolute zero.
; Statistics: mean, standard deviation, t-tests, ANOVA, Pearson correlation, regression and discriminant analysis

; Elaboration On Interval Scales

The following temperatures are equal....

Centigrade 0 10 30 100

Fahrenheit 32 50 86 212

Since there is no absolute zero, absolute comparisons can't be made. Thus even though 100 C is 10 times larger than 10 C it is not 10 times hotter. If that were the case, 212 F would be 10 times hotter than 50 F. But, there is an equal interval between scales:

$$C = (30-10) / (10-0) = 2$$

$$F = (86-50) / (50-32) = 2$$

; ANOVA Example

Say sales by shelf position are as follows:

eye level 50 units/week (average)

chest level 78 units/week (average)

knee level 12 units/week (average)

Is there an apparent relationship between sales and shelf level?

$$H_0: \mu_{eye} = \mu_{chest} = \mu_{knee}$$

Ha: at least two means differ

let $\alpha = .05$

Reject H_0 if $P(F) < .05$

(4) Ratio Scale (비율 척도)

; **This is the highest level of measurement. It allows the researcher to identify or classify objects, rank order the objects, and compare intervals or differences. It is also meaningful to compute ratios of scale values.**

; 비율 척도는 명목, 서열, 간격 척도의 모든 특성들을 포함한다. 추가적으로 절대 0점(absolute 0 point)이 명시된다. 이는 Scale의 Origin이 고정되었다는 것을 의미한다. 간격 척도와는 다르게, 척도값의 비율을 계산하는 것이 여기에서는 의미있다.

; 비율 척도는 비례(proportion)를 이용하여 변형될 수 있다. 변형식은 $y=bx$ 이고, 여기에서 b는 양의 상수이다. 간격 척도에서 등장하였던 a가 없어졌다는 점에 주목하라.

; 비율 척도에는 모든 통계적 방법이 적용될 수 있다. 거기에 더하여 기하적 평균 등도 활용 가능하다.

; **ex) Height, weight, age, income, sales, costs, m/s, number of customers, drive time, distance, Kelvin temperature (0 K = -273.16 C)**

; Differentiates, and orders in standard absolute units; true zero point represents the least possible amount.

; Statistics: any statistics that apply to interval scales apply here too

; What Components Impact On The Measurement Process?

1. true characteristic
2. stable characteristics of the respondent: intelligence, social class
3. short term characteristics of the respondent: fatigue, health, hunger
4. situational characteristics: physical surroundings
5. characteristics of the measurement process:
interviewer's personality,
type of survey: phone, mail, personal
6. characteristics of the measuring instrument: unclear instructions, confusing terms
7. mistaken responses caused by carelessness
8. characteristics of the analysis: coding, tabulating, data entry (8 = 3)

; Intentions v. Behavior

; Stated Intention = True Intention + Error

; examples

; 53% of intended automobile purchases actually took place as intended

; 30% of those who vow to exercise do so

; 25% of those who indicated intended museum patronage followed through

; A Troublesome Issue

; Using interval scale statistics for ordinal scale data

; There is some evidence that respondents tend to avoid extremes on any judgment scale, and sometimes they avoid a neutral or zero center position, thus forming a skewed or possibly a bimodal distribution.

; Other empirical research shows that polar adjectives frequently did not form interval scales.

; Certain statistics such as the t-test are affected rather dramatically when the assumption of linearity (equal intervals) is violated.

Selecting a Level of Measurement (측정 수준 선택하기)

① Nominal Scale

In the past year, who made a contribution to your household income?

1. Make head of the household.
2. Female head of the household
3. Both male and female head of the household
4. Other

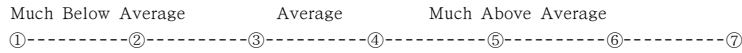
② Ordinal Scale

In the past year, what was your annual household income before taxes?

1. Less than \$20,001
2. \$20,001 to \$50,000
3. \$50,001 to \$100,000
4. \$100,001 to \$150,000
5. More than \$150,000

③ Interval Scale

In the past year, what was your annual household income before taxes?



④ Ratio Scale

In the past year, what was your annual household income before taxes?

; ① 에서 ④ 로 갈수록 얻을 수 있는 정보의 양이 증가하지만, 그만큼 응답자가 느끼는 인지적 부담(cognitive load)도 더 해지게 된다. 따라서 상충효과(trade-off)가 나타난다. 일반적인 톨은 조사자가 답을 얻을 수 있는 최소한의 단계만을 선택 하고 적절한 통계적 분석을 수행하는 것이다.

A Classification of Scaling Techniques

; Comparative Scales (상대적 척도)

; One of two types of scaling techniques in which there is direct comparison of stimulus objects with one another.

; ex) 응답자들은 Coke와 Pepsi 중 어느 것을 선호하는지 대답할 수 있다. 이 때 마케터들은 상대적 차이를 측정한다.

이는 서열 및 순위 측정만을 가지고 있어도 된다.

; 이는 때때로 nonmetric scaling 이라고 불리기도 한다.

; 상대적 척도의 가장 큰 장점은 대상들의 작은 차이가 발견될 수 있다는 점이다. 비교 절차는 응답자에게 두 대상 사이에서 선택하도록 강제한다. 또한 그들은 처음 판단이 나중까지 영향을 미치는 후광 효과(halo)나 나머지 효과(carryover)를 최소화하려는 경향이 있다.

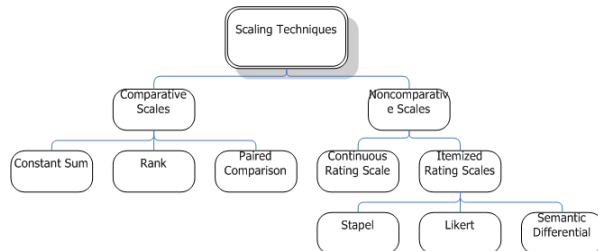
; 가장 큰 단점은 서열 자료 분석의 제약이다. 또한 대상들을 넘어서 일반화시킬 수 없다. 예를 들어서 Coke / Pepsi 를 비교하는 경우, 나중에 RC Cola로 바꾸었을 경우 조사자는 새로운 조사를 해야 한다. 이러한 단점들은 비상대적 척도를 이용함으로써 극복될 수 있다.

; Noncomparative Scales (비상대적 척도)

; One of two types of scaling techniques in which each stimulus object is scaled independently of the others.

; 이는 monadic 혹은 metric scale로도 불리며, 대상들은 각각에 대해 독립적으로 척도가 매겨진다. 결과 자료는 일반적으로 간격 척도가 된다. 예를 들어 응답자는 COke를 1-7까지 점수로 매길 수 있다. 이는 Pepsi와 RC Cola에 대해서도 유사하게 행해질 수 있다. 비상대적 척도는 연속적인 순위(continuous rating) 혹은 항목화된 순위 척도(itemized rating scale)를 포함한다.

; 비상대적 척도는 마케팅 조사에서 가장 널리 쓰이는 방법이다.



Comparative Scaling Techniques (상대적 척도 기법)

(1) Paired Comparison Scaling (짝 비교 척도)

; A comparative scaling technique in which a respondent is presented with two objects at a time and asked to select one object in the pair according to some criterion. The data obtained are ordinal in nature.

; 이는 2개의 짝을 동시에 주고 둘 중 하나를 고르도록 하는 기법이다.

; 짝 비교 척도는 고려중인 브랜드의 수가 5개 이하일 때 유용하다. 많은 수의 브랜드가 주어진다면, 비교의 횟수는 늘어나게 된다. 대안이 주어지는 순서에 따라서 bias가 나타날 수 있다. 또한 미심쩍은 점은, 짝 비교가 마켓 상황에서 잘 일어나지 않는다는 것이다. (대개 이 경우 여러 대안 중에서 선택하게 된다) 응답자는 하나의 대안을 다른 것보다 선호할 수 있다. 하지만 그것이 절대적인 의미에서 좋아한다는 것을 함축하지는 않는다.

(2) Rank Order Scaling (순위 서열 척도)

; A comparative scaling technique in which respondents are presented with several objects simultaneously and asked to order or rank them according to some criterion.

; 짝 비교 척도와 같이, 이 방법도 상대적인 특성을 가진다. 하지만 1위로 마감한 브랜드가 절대적으로 그렇지 않은 가능성도 있다. 즉, 그것은 "가장 적게 싫어하는" 것일 수도 있다는 의미이다. 순위 서열 척도는 서열화된 자료를 만들어 낸다.

; 이는 브랜드 사이의 선호도 및 브랜드간의 태도를 측정하는데 공통적으로 이용된다. 순위 서열 척도는 응답자가 대안들 사이에서 구별하도록 강제한다. 이러한 척도 절차는 쇼핑 환경에 보다 가깝다. 또한 이는 짝 비교 보다는 적은 시간이 걸린다. 또다른 장점은 쉽게 이해되고, 결과가 쉽게 전달 가능(communicate)하다는 것이다.

; 가장 큰 단점은 이것은 오직 서열 척도 자료만을 만들어낸다는 것이다.

(3) Constant Sum Scaling (고정 합 척도)

; A comparative scaling technique in which respondents are required to allocate a constant sum of units, such as points, dollars, chits, stickers, or chips, among a set of stimulus objects with respect to some criterion.

; 고정합 척도는 절대적인 0을 가진다. 따라서 고정합 척도 자료는 metric으로 취급된다. 다만 이것은 제한된 맥락 안에서 측정된 대상에 대해 적절할 수는 있어도, 이러한 결과가 일반화되거나 다른 대상에 대해서도 적용 가능한 것은 아니다. 엄밀히 말하자면, 고정합은 서열 척도로 고려되어야 한다.

; 고정 합 척도의 가장 큰 장점은 이것이 대안들 사이의 구별을 가능하게 하고 시간이 많이 걸리지 않는다는 점이다.

; 다만 가장 큰 단점은 응답자가 더 많이 혹은 적게 배분할 수 있다는 것이다.(108이나 94와 같이) 이런 사태가 발생하면 조사자는 100점이 되도록 조정하거나 응답자를 분석에서 제외해야 한다.

Relationship of Measurement and Scaling to the Marketing Research Process

; 측정과 척도는 Research Design 단계에서 행해진다. 조사자는 적절한 측정 단계(명목, 서열, 간격, 비율)를 선택해야 한다.

; 또한 조사자는 명목 및 서열 척도에서 측정될 정보의 적절한 척도를 선택해야 한다.

Summary Illustraion Using the Opening Vignette

International Marketing Research

Technology and Marketing Research

Ethics in Marketing Research

[Discussion]

A1. Identify the type of scale (nominal, ordinal, interval, or ratio) being used in each of the following. Please explain your reasoning.

- a. I like to solve crossword puzzles. (\Rightarrow interval)
- | Disagree | Agree |
|---------------------------|-------|
| ①-----②-----③-----④-----⑤ | |
- b. How old are you? _____ (\Rightarrow ratio)
- c. Please rank the following activities in terms of your preference by assigning rank 1 to 5, with 1 being most preferred and 5 being least preferred. (\Rightarrow ordinal)
- Reading magazines ____
 - watching television ____
 - Dating ____
 - Shopping ____
 - Eating out ____
- d. What is your Social Security number? _____ (\Rightarrow nominal)
- e. On an average weekday, how much time do you spend? (\Rightarrow ordinal)
- Less than 15 minutes ____
 - 15 to 30 minutes ____
 - 31 to 60 minutes ____

A6. Target and Wal-Mart are two of the major department stores. Develop a series of paired comparison scales comparing these two stores on store image characteristics. Identify the relevant store image characteristics by visiting the Web sites of these two stores (www.target.com, www.walmart.com)

- ① Cheap and Expensive
② Excitable and Calm

D1. "A brand could receive the highest median rank on a rank order scale of all the brands considered and still have poor sales." Discuss.

I think it can be possible if the specific brand is positioned in the consumer's mind as a high-quality and customers are thinking that the brand is too expensive to buy it. In this kind of situation, I'll argue that it is problem of consumer's perception.

A2. Show how intentions to purchase four brands of soft drinks (Coke, Pepsi, Dr. Pepper, and 7-Up) can be measured using ordinal, interval, and ratio scales.

(1) Ordinal scale

Please rank the following items in terms of your preference by assigning rank 1 to 4, with 1 being most preferred and 4 being least preferred.

- ① Coke
- ② Pepsi
- ③ Dr. Pepper
- ④ 7-Up

(2) Interval scale

I made up it with 4 sub questions and each of the subquestion has a statement like "I like Coke, or I like Pepsi, ...". And each question comes up with 5 points which is from 1 to 5. And 1 is Disagree, 5 is Agree. Respondent are required to mark all the 4 subquestion.

- ① I like Coke (① is Disagree, ⑤ is Agree)
①-----②-----③-----④-----⑤
- ② I like Pepsi
①-----②-----③-----④-----⑤
- ③ I like Dr. Pepper
①-----②-----③-----④-----⑤
- ④ I like 7-Up
①-----②-----③-----④-----⑤

(3) Ratio scale

Please score each of the soft drink brand in numbers but the sum of the scores should be 100.

- ① Coke _____
② Pepsi _____
③ Dr. Pepper _____
④ 7-Up _____

Chapter 10. Measurement and Scaling: Noncomparative Scaling Techniques

Overview

Noncomparative Scaling Techniques (비상대적 척도 기법)

; Noncomparative Scales

; One of two types of scaling techniques in which each stimulus object is scaled independently of the other objects in the stimulus set.

; 비상대적 척도는 때때로 monadic scale(단원 척도)이라고도 불리는데, 한 번에 하나의 대상만이 평가되기 때문이다. 비상대적 척도를 사용하는 응답자는 자기 자신의 순위 표준(own rating standard)를 사용한다. (Fatal attraction이라는 영화가 다른 영화가 아닌 그 자기 자신에 대해서만 평가되는 것과 같다)

Continuous Rating Scale (연속 순위 척도)

; Continuous Rating Scale

; A measurement scale in which respondents rate the objects by placing a mark at the appropriate position on a line that runs from one extreme of the criterion variable to the other.

; 이는 응답자가 2개의 점을 잇는 선분 사이에 표시를 하도록 한다. 따라서 이 때는 28.637과 같은 숫자도 나올 수 있다. 연속 순위 척도의 형태는 수직 혹은 수평 둘 다 가능하다. 연속 순위 척도는 때때로 graphic rating scale이라고 불리기도 한다.

; 범주 간의 거리는 일정하고 0점이 임의적이기 때문에, 이 형태의 척도는 간격 척도 자료를 생성한다.

; 연속 척도는 생성하기 쉽다. 하지만 컴퓨터 스크린 상에 보여지지 않는 한 이들을 점수매기는 것은 어렵고 신뢰할 수 없다. 이는 인터넷에서는 쉽게 구현된다.

Scale	Basic characteristics	Examples	Advantaes	Disadvantages
Continuous Rating Scale	Place a mark on a continuous line	Reaction to TV commercials	Easy to construct	Scoring can be cumbersome unless computerized
Likert Scale	Degree of agreement on a 1(strongly disagree) to 5(strongly agree) scale	Measurement of attitudes	Easy to construct, administer, and understand	More time-consuming
Semantic Differential	Seven-point scale with bipolar labels	Brand, product, and company images	Versatile	Difficult to construct appropriate bipolar adjectives
Stapel Scale	Unipolar 10-point scale, -5 to +5, without a neutral point (zero)	Measurement of attitudes and images	Easy to construct and administered over telephone	Confusing and difficult to apply

Itemized Rating Scales (항목 척도)

; A measurement scale that has numbers and/or brief descriptions associated with each category. The categories are ordered in terms of scale position.

; 이는 마케팅에서 가장 널리 쓰이는 척도이기도 하다.

(1) Likert Scale (리커트 척도)

; A measurement scale with five response categories ranging from "strongly disagree" to "strongly agree", which requires the respondents to indicate a degree of agreement or disagreement with each of a series of

statements related to the stimulus object.

; 자료들은 보통 간격 척도로 취급된다. 각 매장에 대한 각각의 점수를 결정하는 등의 일에 접근할 때, 높은(혹은 낮은) 점수가 호의적인 반응을 일관성있게 반영하도록 scoring procedure를 일관성있게 사용하는 것이 중요하다. 즉 부정형 질문에 대해서 응답자는 역으로 척도를 사용하여야 한다는 것을 의미한다.

; 높은 점수가 높은 호의도를 반영하는 것이라면, 부정형으로 진술된 문장들은 역코딩(reverse)할 필요가 있다. 이는 척도 최대값+1 에서 원래 값을 뺀으로서 이루어 질 수 있다. (예를 들어 척도점이 5개이고 답이 2인 경우 역코딩하면 (5+1)-2=4가 된다)

; 리커트 척도는 몇가지 장점이 있다. 우선 연구자가 만들고(construct) 관리(administer)하기가 쉽다. 또한 응답자들이 이해하기가 쉽다. 또한 우편, 전화, 개인 및 전자적 면접에 적합하다. 리코트 척도의 몇가지 변형은 마케팅에서 흔하게 사용되는데, 여기에는 척도점 수의 변형(7, 9등)이나 서술(descriptor)의 변형(중요도, 친밀도) 등이 있다.

; 가장 큰 단점은 리커트 스케일이 다른 항목 척도에 비해 완료하는데 시간이 더 많이 걸린다는 것이다. 응답자들은 짧은 구문이 아닌 전체 문장을 읽어야 한다.

(2) Semantic Differential Scale (의미 차별화 척도)

; A seven-point rating scale with end points associated with bipolar labels that have semantic meaning.

; 의미 차별화 척도를 사용할 때, 응답자는 보통 브랜드, 매장이나 다른 대상들을 양극 형용사(bipolar adjective) 사이에서 점수 매기게 된다.(예를 들면 "좋다", "싫다" 등)

; 응답자는 대상을 가장 잘 묘사하는 칸에 표시를 하게 된다. 각 질문에 대해 주의깊게 숙고하도록 하게 하기 위해, 방향을 반대로 질문을 배치하기도 한다. 이는 라벨을 읽지도 않고 한쪽에 칠해나가는, 이른바 지나치게 긍정적이거나 지나치게 부정적인 태도를 가진 일부 응답자의 경향을 통제하기 위해 사용된다.

; 의미차별화 척도에 사용된 개개의 항목은 -3 ~ +3 이나 1~7 척도로 점수매겨질 수 있다. 결과 자료는 간격 척도로 여겨지고, 프로파일 분석에 사용된다. 프로파일 분석에서 각 항목에 대한 평균이나 중간값이 계산되고, 플로팅(plot)되고, 통계적으로 분석된다. 결과를 플로팅함으로써 조사자는 측정된 대상 간의 전반적인 차이점이나 유사점을 발견할 수 있다.

; 의미 차별화는 그 가변성(versatility)으로 인해 시장 조사에서 매우 인기있게 쓰이는 방법이다. 이는 브랜드, 제품, 회사 이미지의 비교, 광고 및 홍보 전략 개발, 신제품 개발 연구 등에 사용된다. 가장 큰 단점은 척도를 만드는데 필요한 적절한 양극 형용사를 결정하는 것이다.

(3) Stapel Scale (스타펠? 척도)

; A scale for measuring attitudes that consists of a single adjective in the middle of an even-numbered range of values.

; 이는 대개 수직적으로 제시되고, -5부터 +5까지가 주어진다. 응답자는 중립을 선택하도록 허용되지 않는다. (즉 0점이 없다) 응답자는 숫자를 선택하여 각각의 단어가 얼마나 정확하게 혹은 부정확하게 대상을 묘사하는지를 나타낸다. 높은 숫자는 정확한 묘사, 낮은 숫자는 부정확한 묘사이다.

; 스타펠 척도로 얻어진 자료는 또한 간격 척도로 여겨진다. 스타펠 척도에서 하나의 형용사만을 사용하는 것이 의미 차별화 척도에 대해 가지는 장점으로, 사용된 형용사가 정말로 반대의 의미를 가졌는지 미리 테스트할 필요가 없기 때문이다. 이 척도의 간명함은 전화 인터뷰에서 나타난다. 하지만 일부 연구자들은 스타펠 척도가 혼란스럽고 어렵다고 생각한다.

; 이 척도는 과거보다도 폭넓은 적용을 가능하게 한다는 장점이 있다. 예를 들어 이는 전화 인터뷰에서 많이 쓰이는데, 이는 소비자 설문 조사에서 가장 인기 있는 방법이기도 하다.

Noncomparative Itemized Rating Scale Decisions (비상대적 항목 척도 결정)

; 비상대적 척도를 사용하는 장점 가운데 하나는 특정 연구 과제에 적응하는데 대한 유연성(flexibility)을 제공한다는 점이다.

Decision Factor	Guidelines
Number of categories	While there is no single optimal number, traditional guidelines suggest there should be between five and nine categories.
Balanced vs. unbalanced	In general, the scale should be balanced to obtain objective data.
Odd vs. even number of categories	If a neutral or indifferent scale response is possible for at least some of the respondents, an odd number of categories should be used.
Forced vs. nonforced	In situations where the respondents are expected to have no opinion, the accuracy of data may be improved by a nonforced scale.
Verbal description	An argument can be made for labeling all or many scale categories. The category descriptions should be located as close to the response categories as possible
Physical form	A number of options should be tried and the best one selected.

(1) Number of Scale Categories (척도 범주의 수)

； 조사자의 관점에서 척도 내의 범주의 수가 많을수록 브랜드, 대안 및 다른 대상들 간의 차이가 자세하게 된다.(finer)

； **Sensitivity (민감도)**

； **The sensitivity of a scale is closely related to reliability and is the ability to detect subtle differences in the attitude or the characteristic being measured.**

； 척도 범주 수를 증강면 민감도도 증가하게 된다. 하지만 많은 수의 범주는 응답자가 더 많은 정보 처리를 해야 함을 의미한다. 응답자가 질문에 대답할 때 처리할 수 있는 정보의 양에는 한계가 있다. 또다른 중요한 요인은 자료 수집 형태이다. 설문지를 볼 수 없는 전화 면접과 같은 경우 척도 범주의 수가 너무 크다면 매우 혼란스럽게 된다. 설문지의 공간 제한 역시도 또다른 제약조건이다.

； 최적화된 수(optimal number)가 없다면 조사자는 요구되는 정보의 양과 자료 수집 환경의 균형을 말아야 한다. 전통적인 지침에 따르면 5개보다 적거나 9개보다 많은 범주는 피하는 것이 좋다. 9개 이상의 척도 범주에서 응답자는 혼란스러워하거나 피로를 느끼기 때문에 자료의 질이 떨어지게 된다.

(2) Balanced Versus Unbalanced Scale (균형 대 비균형 척도)

； **Balanced Scale (균형 척도)**

； **A scale with an equal number of favorable and unfavorable categories.**

； 비균형 척도의 예는 "나쁘다"에 해당하는 것이 4개, "좋다"에 해당하는 것이 2개 있는 경우.

； 일반적으로 균형 척도가 보다 바람직하다. 만약 조사자가 응답자들이 부정적이거나 긍정적으로 풀려있다고 생각할 경우 비균형 척도를 사용하는 것도 바람직하다. 예를 들어서 고객 만족의 경우 보통 긍정적으로 풀리게 되는데(skew) 이는 많은 응답자들이 회사와 그 제품에 대해서 만족하기 때문이다. 따라서 이 경우 만족 범주의 수를 불만족 범주의 수 보다 많게 한다. 만약 비균형척도가 사용되면, 척도의 불균형도를 자료 분석시 고려하여야 한다.

(3) Odd or Even Numbers of Categories (짝수점 척도 및 홀수점 척도)

； 홀수점 척도가 쓰여진다면, 중간점은 보통 "중립"을 나타낸다. 중립 범주를 사용하는 것과 사용하지 않는 것은 응답에 지대한 영향을 끼친다. 리커트 스케일은 중립점을 가지는 홀수개의 범주를 가지는 균형 척도의 예이다.

； 만약 조사자가 응답자 인구가 중에서 실제로 중립적 의견을 고수하고 있는 사람들이 있다고 여겨질 때 홀수점 척도를 사용할 수 있다. 반면 조사자가 중립이나 무관심 응답이 없다고 믿는다면 짝수점 척도를 사용할 수 있다.

(4) Forced or Nonforced Choice (강제 혹은 비강제 선택)

； **Forced Rating Scale**

； **A rating scale that forces the respondents to express an opinion because a "no option" option is**

not provided.

； 만약 응답자의 상당수가 무의견일 때 강제 선택 척도를 사용한다면, 응답자는 척도의 중간점을 선택하는 경향이 있다. 다만 "무의견"을 중간점으로 찍는다는 것은 분산 측정을 훼손하는 결과가 될 수 있다. 이러한 경우 "무의견"이라는 범주를 하나 만드는 비강제 선택을 할 수 있다. 예를 들어 특정 브랜드에 대해 잘 모르는 경우 비강제 선택을 사용할 수 있다. 그 외의 경우에 "무의견" 범주는 피하는 것이 좋다.

(5) Nature and Degree of Verbal Description (척도점의 의미 기술)

； 척도점의 의미를 기술하는 방식에 따라 응답에 심각한 영향을 끼칠 수 있다. 척도 범주는 설명, 숫자, 그림 등이 될 수 있다. 다만 모든 범주에 대해 의미를 기술(verbal description)하는 것이 자료의 정확도나 신뢰성을 증대시키지는 않는다. 그러나 전부 혹은 상당수의 범주에 설명을 기술하는 것이 척도의 모호성을 줄인다는 주장도 있다. 범주 설명이 나 레이블은 가급적 응답 범주 가까이 위치해야 한다.

； 형용사의 강도도 응답에 영향을 끼친다. 예를 들어 "완전히 동의"(completely agree), "완전히 부정"(completely disagree) 등과 같은 형용사를 사용하면 극단을 피하려는 응답자의 경향 때문에 답변이 중간에 몰리는 경향이 있다. 반대로 좀 약한 형용사를 사용하면(generally agree, generally disagree)좀 더 균일하게(uniform and flat) 분산된다.

(6) Physical Form or Configuration (물리적 형태 혹은 조정)

； 척도가 제시되는 방법은 극적으로 달라질 수 있다. 척도는 수직 혹은 수평으로 제시될 수 있다. 박스나 점선, 숫자를 포함하거나 포함하지 않을 수도 있다. 숫자값이 쓰여졌을 경우 양수, 음수, 혹은 둘 다 일수 있다.

Multi-Item Scales (다항목 척도)

； **A scale consisting of multiple items, where an item is a single question or statement to be evaluated.**

； 리커트, 의미 차별화, 스타렐 척도들이 다항목 척도 사용의 예이다. 각각의 척도가 몇개의 항목을 포함하고 있다는 사실에 유의하라.

； **Construct (구성)**

； **A specific type of concept that exists at a higher level of abstraction than do everyday concepts.**

； ex) 브랜드 충성도, 제품 연관도, 태도, 만족 등등.

； 순서

- ① Develop the Construct
- ② Develop a theoretical definition
- ③ Develop an operational definition
- ④ Develop a Multi-Item scale
 - Genereate a pool of scale items
 - Reduce the pool of items based on judgment
 - Collect data
 - Purify the scale based on statistical analysis
- ⑤ Evaluate scale reliability and validity
- ⑥ Apply the scale and accumulate research findings

Scale Evaluation (척도 평가)

； 다항목 척도는 신뢰성(reliability)과 타당성(validity)을 평가받아야 한다.

； **Systematic Error (구조적 오류)**

； **Affects the measurement in a constant way and represents stable factors that affect the obseerved score in the same way each time the measurement is made.**

； **Random Error (임의 오류)**

； **Measurement error that arises from random changes that have a different effect each time the measurement is made.**

； * Total measurement error = Systematic error + Random error

(1) Reliability (신뢰성)

； The extent to which a scale produces consistent results if repeated measurements are made on the characteristic.

； 따라서 신뢰성은 측정이 얼마나 임의 오류로부터 자유로운가의 정도로 정의될 수 있다.

； 신뢰성은 구성 혹은 관심있는 변수를 반복적으로 측정하는 것으로 결정된다. 이러한 반복 절차를 통해 얻어진 점수의 높은 연관도가 증가할수록 척도는 보다 신뢰할 수 있게 된다.

； Reliability refers to the relative absence of variable errors of measurement; it is primarily a “mechanical” issue.

； Synonyms: dependability, stability, consistency, predictability, accuracy.

； It is possible for a measure not to be valid, but still be reliable. Ideally, a measure will be both reliable and valid.

① Test-Retest Reliability (시험-재시험 신뢰성)

； An approach for assessing reliability in which respondents are administered identical sets of scale items at two different times under as nearly equivalent conditions as possible.

； 재시험은 보통 원 시험 후 2~4주 이후에 이루어진다. 2회 측정 사이의 유사도 정도는 correlation coefficient (연관 계수)를 계산함으로써 이루어진다. c.c.가 높을수록 신뢰성은 증대된다.

； The same test is administered to the same individuals at two separate times, then the two scores are correlated.

； Major problems:

； Time lag (maturation of respondents, fatigue, hunger)

； Learning (the act of measurement may change the respondent, e.g., ad recall tests)

② Alternative-Forms Reliability (대안-서식 신뢰성)

； An approach for assessing reliability which requires two equivalent forms of the scale to be constructed and then measures the same respondents at two different times using the alternate forms.

； 2개의 동등한 서식을 사용하여 측정된 그들 사이의 correlation은 신뢰도 측정 수단을 제공한다.

； Also called “comparable forms,” alternate tests are administered to the same individuals at two separate times, then the two scores are correlated.

； Advantage: specific knowledge can't be used in the second test.

； Major problems:

； Time lag

； How does one develop a “comparable form?”

； 마케팅자들은 실제로 이거 거의 사용하지 않는다. 대신 Internal-consistency reliability 테스트를 한다.

③ Internal-Consistency Reliability (내적 일관 신뢰성)

； An approach used to assess the reliability of a summated scale and refers to the consistency with which each item represents the construct of interest.

； 이러한 형태의 척도에서 각 항목은 전체 척도로서 측정된 구성(construct)의 특정 측면들을 측정한다. 이 항목들은 특성을 지칭할 때 일관성을 가져야 한다.

； Split-half Reliability (반분 신뢰성)

； A form of internal-consistency reliability in which the items constituting the scale are divided into two halves, and the resulting half scores are correlated.

； 각 절반 사이에 높은 상관관계가 발견된다면 높은 내적 일관성을 가지고 있음을 가르킨다. 각 절반 사이의

상관관계는 집단이 어떻게 나뉘느냐에 따라 영향받는다.

； Coefficient Alpha (알파 계수)

； A measure of internal-consistency reliability that is the average of all possible split-half coefficients resulting from different splittings of the scale items.

； 이는 0~1까지이며 보통 0.6 이하의 값은 내적-일관성 신뢰성을 만족하지 못함을 지칭한다.

； Pragmatic form is “split half” reliability used by marketers.

； Think of one test with multiple items as two separate tests; correlate half of the items with the other half of the items.

； Generally, a “coefficient alpha” > .6 shows reliability

； Key advantages:

； No time lag

； Minimal sensitization of respondent, thus little chance for “learning”

； How To Improve Reliability? : To be interpretable, a test must be reliable.

； ① Increase sample size to reduce variance.

； ② Write test items unambiguously, as various interpretations increase error variance.

； ③ Add more items to form a scale. With few items, chance looms large.

； ④ Provide clear and standard instructions to reduce error measurement. Ambiguous instructions increase error variance.

(2) Validity (타당성)

； The extent to which differences in observed scale scores reflect true differences among objects on the characteristic being measured, rather than systematic or random errors.

； 완벽한 타당성을 갖춘 척도는 어떠한 측정 오류, 계측 오류 및 임의 오류도 포함하지 않게 된다.

； Validity is complex, controversial, and important. Basically, a valid measure is one that measures what it is supposed to measure.

； examples: age & calendar, length & ruler, weight & scale

； Less easy examples: attitudes? intentions? How can valid measures be made?

； It is possible for a measure to be valid, but not reliable.

① Content Validity (내용 타당성)

； A type of validity, sometimes called face validity, that consists of a subjective but systematic evaluation of the representativeness of the content of a scale for the measuring task at hand.

； 내용 타당성을 만족하게 되면 construct의 모든 차원(dimension)을 진술하게 된다. 이는 상식적인 척도 평가방법이다. 매장 이미지 구성을 측정하기로 설계된 척도는 이미지의 주요 차원(quality, variety, assortment of merchandise 등) 중 어느 하나라도 빼먹는다면 부적절한 것으로 여겨지게 된다. 내용 타당성 자체는 척도의 타당성을 측정하기에 적절하지 않다. (뒤이어 나오는 것들의 타당성도 함께 검증되어야 함)

； Also called “face validity,” it addresses common sense interpretation.

； By its very nature, it is judgmental.

； From a marketer's perspective, it is important to have content validity to be credible in the eyes of the client.

； But, even if a measure looks valid, does not mean that it is valid.

② Criterion Validity (기준 타당성)

； A type of validity that examines whether the measurement scale performs as expected in relation to other variables selected as meaningful criteria.

； 이 변수들은 criterion variables 라고 불리기도 한다. 여기에는 인구통계적(demographic), 심리분석적

(psychographic), 태도적(attitudinal), 그리고 행태적(behavioral) 특성들 혹은 다른 척도에서 얻은 점수(score)가 포함될 수 있다. 예를 들어 연구자는 다항목 척도를 사용해 시리얼 브랜드에 대한 태도를 측정할 수 있다. 태도 레벨에 근거하여 시리얼의 미래 구매가 예측될 수 있다. 이 때 스캐너 자료는 패널 구성원의 실제 시리얼 구매를 추적하기 위해 사용된다. 그리고 예측된/실제 구매는 태도 척도의 기준 타당성을 평가하기 위해 비교된다.

; Test or scale scores are compared with an attribute of interest. The higher the correlation between a measure and the attribute, the better the validity.
; Depending on the time sequence, there are three kinds of criterion validity:

; Predictive validity

; Example: Do SAT scores predict GPA?
; SAT is a predictor measured in time 1
; GPA is the criterion performance measured in time 2
; The problem is “restriction of range,” where no correlation might be found. (예를 들어 GPA > 2.0 & SAT > 1200 이어야만 입학/졸업 조건에 맞기 때문이다)

; Concurrent validity

; Can be used by marketers to streamline the measurement process.
; A long and a short version of a test are administered at the same time. If the results are “similar,” one might conclude that there is concurrent validity, and that the short version can then be used in future studies.
; However, in practice, it is seldom used. Consider the problem of developing a valid measure of income.
; 다음과 같은 상황도 있을 수 있다.
Measures of income made on four concurrent subsamples drawn from a master sample of 1,000 households....
Form 1: was income more than \$5,000, more than \$7,500, etc..
Form 2: zeroing or bracketing type of question: was income \$15,000 or more or would it be less than that, etc..
Form 3: was income less than \$5,000, less than \$7,500, etc..
Form 4: was income more than \$25,000, more than \$20,000, etc..
Results: US Census data median was \$15,235
Form 1 median was \$12,711 Form 2 median was \$14,961
Form 3 median was \$14,949 Form 4 median was \$17,184

; Postdictive validity

; Used in forensic medicine and case analysis to establish the facts of what happened -- it is the post mortem or autopsy.
; Outcome in time 1
; “Prediction” in time 2
; “We cannot escape history.”

③ Construct Validity (구성 타당성)

; A type of validity that addresses the question of what construct or characteristic the scale is measuring. An attempt is made to answer theoretical questions of why a scale works and what deductions can be made concerning the theory underlying the scale.

; The purpose of construct validity is to explain the theory of WHY a measure works. This is done through hypotheses testing and empirical studies.
; Examples:
What is a valid measure of social class?
How does family life cycle affect behavior?

What makes a good teacher?
; Three components for construct validation:
Suggest what constructs account for variation
Derive hypotheses about the theory
Test the hypotheses empirically
; The most difficult to establish, it requires convergent, discriminant, and nomological validity.
; 구성 타당성에는 수렴 타당성(Convergent validity), 판별 타당성(Discriminant validity), 및 법칙 타당성(Nomological Validity)이 포함된다.

; ㉔ Convergent Validity (수렴 타당성)

; A measure of construct validity that measures the extent to which the scale correlates positively with other measures of the same construct.

; ㉕ Discriminant Validity (판별 타당성)

; A type of construct validity that assesses the extent to which a measure does not correlate with other constructs from which it is supposed to differ.
; e.g., brand loyalty and variety seeking constructs differ

; ㉖ Nomological Validity (법칙 타당성)

; A type of validity that assesses the relationship between theoretical constructs. It seeks to confirm significant correlations between the constructs as predicted by theory.
; e.g., brands that are congruent with a shopper’s self-concept are more preferred is thought to be true

; 높은 레벨의 신뢰성은 구성 타당성의 증거로서 포함된다는 점에 유의한다.

Relationship Between Reliability and Validity

; 만약 측정이 완벽하게 타당하다면(valid), 이는 또한 완벽하게 신뢰할만 하다(reliable). 이러한 경우 계층 오류나 임의의 오류가 발생하지 않는다. 즉 perfect validity ⇒ perfect reliability 임을 의미한다. 만약 측정이 unreliable 하다면 이는 적어도 완벽하게 valid 하지는 않음을 의미한다. 임의 오류도 나타날 수 있다.
; 다만 측정히 완벽하게 신뢰할만 해도(reliable) 이것이 완벽히 타당하다(valid)는 것을 의미하지는 않는데, 이는 계층 오류가 발생할 수 있기 때문이다. 즉 신뢰성은 타당성을 위한 필요조건(necessary)이기는 하지만 충분조건(sufficient)은 아니다.

; Example From The Literature

; A three-item, six-point (disagree agree), Likert scale that assesses a person’s preference for shopping at malls. For example: “I enjoy going to big shopping malls.”
; Reliability: alpha of .759
; Validity: Factor analysis shows loadings of .7 or higher on the same dimension; content validity was deemed adequate.

; Reliability And Validity Concern For Key Market Assessments

; Typical information needed:
; demographics/socioeconomic characteristics
; psychographics and lifestyles
; personality
; motivation
; knowledge
; past behavior
; attitudes and opinions
; behavioral intentions

； What can we do to make these measures both reliable and valid?

Choosing a Scaling Technique

Summary Illustraion Using the Opening Vignette

International Marketing Research

Technology and Marketing Research

Ethics in Marketing Research

[Discussion]

1. Develop a Likert, semantic differential, and a Staple scale for measuring store loyalty.

2. Develop a multi-item scale to measure students' attitudes toward internationalization of the management curriculum. How would you assess the reliability and validity of this scale?

3. Construct a Likert scale for measuring Internet usage. Show how the rating-scale decisions were made.

Chapter 11. Questionnaire and Form Design

Overview

Importance of the Questionnnaire

； 설문지는 표준화(standardization)와 면접자간의 자료 비교성(comparability), 기록의 속도 및 정확성 증대, 자료 처리 촉진(facilitate) 등을 만족하여야 한다.

Questionnaire Definition

； Questionnaire

； A structured technique for data collection that consists of a series of questions, written or verbal, that a respondent answers.

； 3 specific objectives of questionnaire

； ① The overriding objective is to translate the researcher's information needs into a set of specific questions that respondents are willing and able to answer.

； ② A questionnaire should be written to minimize demands imposed on respondents.

； ③ A questionnaire should minimize response error.

Questionnaire Design Process

； Guidelines of developing questionnaire

； ① Specify the information needed

； ② Specify the type of interviewing method

； ③ Determine the content of individual questions

； ④ Design the questions to overcome the respondent's inability and unwillingness to answer

； ⑤ Decide on the question structure

； ⑥ Determine the question wording

； ⑦ arrange the questions in proper order

； ⑧ Choose the form and layout

； ⑨ Reproduce the questionnaire

； ⑩ Pretest the questionnaire

(1) Specify the Information Needed

； ⇒ Ensure that the information obtained fully addresses all the components of the problem.

； ⇒ Have a clear idea of the target population.

； 설문지는 대상 응답자를 고려하여 설계되어야 하며 그들의 교육적 수준과 경험도 고려되어야 한다. 질문에 사용된 언어와 맥락은 응답자에게 친숙한 것이어야 한다. 응답자의 특성을 고려하지 않은 설문지는 실패하게 된다. (이 경우 불명확 ("uncertain") 혹은 무의견 ("no opinion") 응답이 많다)

(2) Specify the Type of Interviewing Method

； ⇒ Review the type of interviewing method.

； 면접관-주도(interviewer-administered)의 질문지는 대화형(conversational style)으로 쓰여져야 한다. 우편 및 전자 설문지는 자가-관리적(self-administered)이다. 따라서 양측의 질문은 간명하고 주의깊게 쓰여져야 한다. CATI나 CAPI, Internet 등에서는 컴퓨터가 건너뛰기 패턴(skip pattern) 및 순서 편차를 제거하기 위한 질문 섞기(randomization of questoin) 등을 수행할 수 있다.

； 고급자 예제를 들어보자면, 설문 방식에 따라 순위 질문이 변하는지를 알 수 있다. 우편, 이메일, 인터넷 설문지에서는 응답자는 순위 서열을 기록한다. 반면 전화에서 7개 차 브랜드를 한꺼번에 평가하는 것은 쉽지 않기 때문에 각각의 브랜드에 대해서 평가하는 방식을 사용한다. 개인 면접의 경우에는 자동차 브랜드가 출력된 응답 카드를 주어 측정하기도 한다.

； 인터넷 설문은 CAPI 설문지의 많은 특성들을 공유한다. 설문지에서는 다양한 자극, 그래픽, 사진, 광고, 애니메이션, 소리,

비디오 등을 제공할 수 있다. 조사자는 자극을 가하는 시간의 양을 조절하고 각 응답자가 각 자극에 접근하는 것을 통제할 수 있다. 이런 인터넷 사용은 설문지의 범위와 복잡성을 크게 증대시킬 수 있다.
; 면접 방식의 형태는 개인적 질문의 질에도 영향을 준다.

(3) Determine the Content of Individual Questions

① Is the Question Necessary?

; 알면 좋기는 하겠지만 연구 과제에서 직접적으로 쓰이지 않을 질문들은 없애버리는 것이 좋다.
; 다만 여기에도 예외는 있다. 프로젝트의 목적이나 후원자를 감추기 위해 채워넣기 질문(filler question)을 추가할 수 있다. 예를 들자면 특정 브랜드에 대한 응답을 조사하기 위해 경쟁하는 모든 브랜드 이름을 집어넣어 누가 연구에 대한 후원사인지를 감추는 것이다. 면접의 초기에 응답자와의 관계를 쌓고 주의를 집중시키기 위해 간단한 대답하기 쉬운 중립적인 질문들을 던지는 것도 좋다. 때때로 신뢰성과 타당성을 평가하기 위한 목적으로 질문들을 반복적으로 사용하는 경우도 있다.

② Are Several Questions Needed Instead of One?

; 때때로 2개의 질문이 하나보다 낫다. 하지만 2개의 질문을 1개로 묶어서 질문하는 것은 해결책이 아니다.
; (Incorrect) Do you think Nike Town offers better variety and prices than other Nike sports?
; "yes"인 경우 명확하지만 "no"인 경우, 특히 하나는 맞고 하나는 틀린 경우가 문제가 된다.
; (Correct) Do you think Nike Town offers better variety than other Nike stores? / Do you think Nike Town offers better price than other Nike stores?

; ⇒ Do not use double-baralled question!!

; Double-barraled question (이중 질문)

; **A single question that attempts to cover two issues. Such questions can be confusing to respondents and can result in ambiguous responses.**

; 다수의 질문을 하나의 질문에 포함시키는 또다른 예는 "why" 질문이다. 이러한 질문은 자료를 얻어내는데 도움이 되지 못한다.

; (Incorrect) Why did you choose the Ford Explorer?

; Answer may be "It was the best buy", or "I liked it"

; (Correct) What was your main motivation in choosing the Ford Explorer? / How does the Ford Explorer compare to other SUV brands? / What features of the Ford Explorer appealed to you?

(4) Design the Question to Overcome the Respondent's Inability/Unwilling to Answer

① Is the Respondent Informed?

; 만약 연구 과제가 전문화된 경험이나 지식을 요구하는 것이라면, 걸러내기 질문(filter question)을 하도록 한다. 이를 통해 적절한 정보를 가지지 않은 사람을 걸러내도록 한다.

; Filter questions

; **Initial questions in a questionnaire that screen potential respondents to ensure they meet the requirements of the sample.**

; ⇒ If the respondents are not likely to be informed, filter questions that measure familiarity, product use, and past experience should be asked before questions about the topics themselves.

; 또한 질문에 "don't know" 선택지를 넣는 것도 도움이 된다. 이는 전반적인 응답율을 떨어뜨리지 않으면서도 알지 못한 상태에서의 응답(uninformed response)의 수를 줄여준다.

② Can the Respondent Remember?

; ⇒ Questions that do not provide the respondent with cues can underestimate the actual occurrence of an event.

; 기억할 수 없는 질문들을 던지지 않도록 한다. (예를 들자면 당신은 지난 4주동안 어떤 음료수를 얼마나 마셨는가?)
; 우리는 개인적으로 연관되어 있거나 unusual 한 사건들을 기억하는 경향이 있다. (결혼 기념일, 생일 등등) 또한 가까운 시간에 일어난 사건일수록 더 쉽게 회상할 수 있다. 마지막으로 들렸던 grocery market의 이름을 기억하는 것이 지난 3년간 들렸던 곳을 기억하는 것보다 쉽다.
; 회상을 돕는 질문도 있을 수 있고, 돕지 않는 질문도 있을 수 있다.

; 돕지 않는 질문(unaid recall) : "What brands of soft drinks do you remember being advertised last night on TV?"

; 돕는 질문(aid recall) : "Which of these brands were advertised last night on TV?" (소프트 드링크 브랜드의 목록을 같이 제시하면서)

⑤ Can the Respondent Articulate?

; 응답자는 때때로 분명히 응답하지 못할 수 있다. 그러나 이것이 응답자가 의견을 가지지 않는다는 것은 아니다. 예를 들어 매장의 이상적인 분위기를 묘사하는 일은 어렵다. 반면 여러가지 대안적인 매장 분위기 묘사를 제시해준다면, 그 중에서 가장 좋아하는 것을 고르는 것은 가능하다. 분명히 말할 것이 어려울 때, 응답자는 질문을 무시하고 설문지를 완성하는 것을 거부할 수도 있다. 그림, 도표, 지도 등의 시각 자료나 설명을 통해 응답자들이 제대로 응답할 수 있도록 도울 수 있다.

⑥ Effort Required of the Respondent

; ⇒ Minimize the effort required by the respondents

; (Incorrect) 가장 자주 쇼핑하는 매장부터 순서대로 목록을 적어주세요.

; (Correct) (매장 목록을 죽 보여주고 순위를 매기기)

⑦ Legitimate Purpose (적법한 방법)

; ⇒ Make the request for information seem legitimate

; 그들은 시리얼 마케팅을 하는데 나이, 수입과 직업에 왜 관심을 가지는지 궁금해 할 수 있다. 조사자는 이러한 의견을 예측하고 왜 자료가 필요한지를 설명해야 한다. 즉 "연령, 수입 및 직업별로 시리얼 브랜드의 선호와 소비도가 어떻게 달라지는지를 조사하기 위해 이러한 자료가 필요합니다.." 등의 말을 써 주면 좋다.

⑧ Sensitive Information

; ⇒ Is the information sensitive?

; 개인적이거나 높은 민감도를 가지는 정보들은 응답자에게서 얻어내기 어렵다. 이러한 것에는 수입, 가족 생활, 정치적/종교적 견해, 사고 및 범죄 연루 등이다. 응답자들은 이러한 질문에 당황하는데, 정확한 응답이 그들의 명성이나 자기-이미지를 위협한다고 생각하기 때문이다. 이러한 민감한 정보들은 설문지의 끝에 위치하는 것이 좋다. 즉 rapport와 과제의 적법성이 확보된 다음에 응답자들이 보다 정보를 주고 싶어하도록 만드는 것이다.

; (incorrect) What is you household's exact annual income?

; (correct) 이걸 4개(under \$25,000 / \$25,001-\$50,000 / \$50,001-\$75,000 / over \$75,000)로 나누어서 물어보는 것.

(5) Decide on the Question Structure

Unstructured Questions (비구조적 질문)

; **Open-ended questions that respondents answer in their own words.**

; 이는 자유-응답(free-response) 혹은 자유-대답(free-answer) 질문으로도 불린다.

; ex) What is your favorite pastime? / How would you describe the typical user of Land Rover SUVs?

; 이러한 개방형 질문들은 과제를 시작하는데 좋은 질문이다. 이러한 질문들은 응답자들이 이미 정의된 대안들로 응답을 제한시키는 bias 없이 그들의 태도와 의견을 표현할 수 있도록 한다. 따라서 이들은 근원적인 동기, 믿음, 태도 등을 식별하는데 유용할 수 있다.

; 비구조적 질문의 단점은 기록 오류(recording error), 자료 코딩 및 분석의 복잡성이다. 전화 응답의 경우 기록된 응답을 축약해서(verbatim) 기록하거나 요점만 기록하려 할 때 면접자 선입견(interviewer bias)가 나타날 수 있다. verbatim reporting 이 중요할 경우 테이프 레코더가 사용된다.

; 기록된 진술을 분류화하는 것은 또다른 bias의 근원이자 큰 결점이기도 하다. 사실 비구조적/열린 질문은 말이 많거나(talkative) 정확하게 말하는(articulate) 응답자에게 보다 많은 가중치를 주는 경향이 있다. 진술들을 요약하여 형식

화하여 조사하는 것은 시간도 많이 걸리도 비용도 만행 든다. 자기-통제적 질문지(편지, CAPI, 전자적)에서 비구조적 질문을 사용하는 것은 가치를 제한시키는데, 이는 응답자들이 말할 때보다 쓸때 보다 간략하게 쓰려는 경향이 있기 때문이다.
; 탐색적 조사나 여는 질문에서는 유용하지만 큰 규모의 설문조사에서는 기록의 복잡성과 정리의 어려움으로 단점이 커지게 된다.

; ⇒ Open-ended questions are useful in exploratory research and as opening questions

Structured Questions (구조적 질문)

; **Questions that prespecify the set of response alternatives and the response format. A structured question could be multiple-choice, dichotomous, or a scale.**

; ⇒ Use structured questions whenever possible

① Multiple-Choice Questions (선다형 질문)

- ; 선다형 질문의 경우 항목 척도를 사용할 때와 같은 문제점이 있다. 거기에 더해 ① 포함되어야 할 대안의 수 ② 순서(order) 혹은 위치(position) bias 의 문제가 추가된다.
- ; 선다형 질문의 선택들은 모든 가능한 대안들의 범위를 커버하여야 한다. 대안들은 상호 배타적(mutually exclusive)이며 부분의 합이 전체를 구성(collectively exhaustive)하여야 한다. 필요하다면 "기타" 항목도 포함시켜야 한다. 또한 하나만 선택해야 하는지, 복수개를 선택할 수 있는지도 명확히 해야 한다. 선택의 수가 많아질수록 응답하기 어려워진다.

; Order or position bias

- ; **A respondent's tendency to choose an alternative merely because it occupies a certain position on the page or in a list.**
- ; 보통 목록의 처음과 끝에 있는 것이 많이 선택된다. 숫자와 관련된 것이 나오면 중간을 선택하려는 경향이 있다. 이러한 order bias는 여러개의 설문지 형식을 준비하여 통제할 수 있다.
- ; 장점 : 선다형 질문은 응답자가 대답하기 쉽다. 또한 개방형 질문에 비해 분석, 요약(tabulate)이 쉽다. 면접관 선입견(interviewer bias)도 감소된다.
- ; 단점 : 효과적인 선다형 옵션을 개발하는 것이 어렵다. 적절한 응답 옵션을 알아내기 위해 탐색적 조사에서 개방형 질문을 해야 하는 경우도 있다. 응답자의 상당수가 "기타"를 선택하였을 경우, 대안 목록에 심각한 결점이 있다는 것을 의미한다. 옵션의 목록 자체도 bias를 내포할 수 있다.

; ⇒ In multiple-choice questions, the response alternatives should include the set of all possible choices and should be mutually exclusive.

② Dichotomous Questions (양분선택형 질문)

- ; **A structured question with only two response alternatives, such as yes and no.**
- ; 장점 : 가장 코딩하고 분석하기 쉽다.
- ; 단점 : 질문에 사용된 단어가 응답에 지대한 영향을 끼칠 수 있다. 즉 A than B 에서 A 와 B의 순서를 바꿈으로서 답이 달라질 수 있다는 것이다. 이러한 문제를 극복하기 위해서 설문지를 반으로 나누어 각기 단어 방향을 다르게 한 다음에 물어볼 수 있다. (split-ballot technique)

; ⇒ In a dichotomous question, if a substantial proportion of the respondents can be expected to be neutral, include a neutral alternative.

; ⇒ Consider the use of the split-ballot technique to reduce order bias in dichotomous and multiple-choice questions.

③ Scales

; 언급되었으니 넘어간다!

; ⇒ If the response alternatives are numerous, consider using more than one question.

Question Type	Advantages	Disadvantages	Comments
Unstructured	<ul style="list-style-type: none">Good as first questionsResponses are less biasedCan provide rich insights	<ul style="list-style-type: none">Potential for interviewer biasCoding is costly and time-consumingGreater weight to articulate respondentsUnsuitable for self-administered questionnaires	<ul style="list-style-type: none">Useful for exploratory research
Multiple-choice	<ul style="list-style-type: none">Interviewer bias is reducedEasy to code and analyzeImproved respondent cooperation	<ul style="list-style-type: none">Order or position biasDifficult to design response options	<ul style="list-style-type: none">Responses should be mutually exclusive and collectively exhaustiveUseful in large surveys
Dichotomous	<ul style="list-style-type: none">Same as multiple-choice	<ul style="list-style-type: none">Wording can bias the responses	<ul style="list-style-type: none">Use split-ballot technique
Scales	<ul style="list-style-type: none">Same as multiple-choice	<ul style="list-style-type: none">Difficult to design multi-item scales	<ul style="list-style-type: none">Scales should be evaluated for reliability, validity, and generalizability

(6) Determine the Question Wording

① Define the Issue

- ; ⇒ Define the issue in terms of who, what, when, where, why, and way (the six Ws)
- ; (incorrect) Which brand of toothpaste do you use?
- ; (correct) Which brand or brands of toothpaste have you personally used at home during the past month? In case of more than one brand, please list all the brands that apply.

② Use Simple Words

- ; ⇒ Use ordinary words. Words should match the vocabulary level of the respondents.
- ; 대부분의 미국인들은 고등학교 학력을 가지고 있음에 유의하도록 한다.
- ; (incorrect) Is the distribution of snack foods adequate?
- ; (correct) Are snack foods readily available when you want to buy them?

③ Use Unambiguous Words

- ; ⇒ Avoid ambiguous words: usually, normally, frequently, often, regularly, occasionally, sometimes
- ; 따라서 이런 단어들 대신 정확한 횟수를 나타내는 숫자로 찍어주면 좋다.

④ Avoid Leading or Biasing Questions

; ⇒ Avoid leading questions that clue the respondent to what the answer should be.

; Leading question

; **A question that gives the respondent a clue as to what the answer should be**

； (incorrect) Do you think that America should provide financial aid to poor foreign countries when it is not our responsibility to do so?
； (correct) Do you think that Armerica should provide financial aid to poor foreign countries?

； 또한 조사 스폰서의 신원을 제공하는 것도 같은 효과를 가져온다. 응답자들은 스폰서를 인식하게 되었을 때, 호의적으로 대답하는 경향이 있다.
； (incorrect) Is Colgate your favorite toothpaste?
； (correct) what is your favorite brand of toothpaste?

⑤ Balance Dual Statements

； ⇒ Use positive and negative statements.
； 리커트 척도 등에서 긍정형으로 물어보는지, 부정형으로 물어보는지에 대해서 차이가 나타난다고 알려져 있다. 따라서 2종류의 설문지를 준비하여 하나는 긍정형으로, 또다른 하나는 부정형으로 물어보면 된다.

(7) Arrange the Questions in Proper Order

① Opening Questions (시작 질문)

； ⇒ The opening questions shuld be interesting, simple, and nonthreatening.
； ⇒ Qualifying questions(검정 질문) should serve as the opening questions.
； 때때로 시작 질문에서 rapport를 형성하기 위해 간단한 질문을 하기도 한다.

② Type of Information (정보의 종류)

； Basic information (기본적 정보)

； Information that relates directly to the marketing research problem.

； Classification information (분류 정보)

； Socioeconomic and demographic characteristics used to classify respondents.

； Identification information (신원 정보)

； A type of information obtained in a questionnaire that includes name, address, and phone number.

； ⇒ Basic information should be obtained first, followed by classification, and, finally, identification information.

③ Difficult Questions (어려운 질문)

； ⇒ Difficult, sensitive, or complex questions should be placed late in the sequence.
； 분류(Classification) 섹션의 가장 마지막 질문은 보통 수입 (income) 정보이다. 같은 이유로 응답자의 전화번호는 신원 (Identification) 섹션의 가장 마지막 질문이 된다.

④ Effect on Subsequent Questions (후속 질문의 효과)

； Funnel approach (깔대기 접근방식)

； A strategy for ordering questions in a questionnaire in which the sequence starts with the general questions, which are followed by progressively specific questions, in order to prevent specific questions from biasing general questions.

； ⇒ General questions should precede specific questions.
； 다만 때로는 그 반대의 방식이 사용되기도 한다. 이는 응답자가 주제에 대한 명확히 형식화된 의견이 없고 주제에 대한 일반적인 질문에 어떻게 반응해야 하는지에 대한 common frame이 부족할 때 사용된다. 이 경우 구체적인 질문이 처음에 나오고 그 다음에 일반적인 질문들이 이어진다.

⑤ Logical Order (논리적 순서)

； Branching questions (분기형 질문)

； Questions used to guide an interviewer through a survey by directing the interviewer to different

spots on the questionnaire depending on the answers given.

； 분기형 질문을 사용할 경우 응답자에게 불필요한 질문들을 묻지 안하도록 한다. 이러한 건너뛰기 패턴은 CATI, CAPI, 인터넷 등에서 매우 복잡해진다.
； World Vision 예제 ： ① 기부부 우선순위와 동기 ② 월드비전 기관의 인지 여부 ③ 기관에 대한 인식 ④ 기부자와 의 의사소통 ⑤ 인구통계적 정보 의 순서대로 질문한다.
； 또한 주제가 바뀔 경우 짚막하게 이에 대한 설명을 달아주면 좋다.

； ⇒ Questions should be asked in a logical order.

Question Type	Nature	Function	Example
Qualifying/Screening Questions	Focus on respondent inclusion criteria	To determine if a respondent is eligible to participate in the survey	Who in your household does most of the shopping for groceries?
Introductory Questions/Warm-Ups	Broad, easy questions	To break the ice and put the respondent at ease	How often do you shop for groceries?
Main Questions: Easy	Related to the information needed but easy to answer	To focus on the survey topic and reassure the respondent that survey is easy	How important is each of the following factors in selecting a supermarket?
Main Questions: More difficult	Related to the information needed but may be difficult be to answer	To obtain the rest of the information needed	How would you rank order the following eight supermarkets in terms of your preference to shop?
Psychographics/Lifestyles	Not relevant in all survey	To obtain personality-related information	Please indicate your degree of dis/agreement with the following statement
Demographics	Personal information	To classify the respondents	What was your household's total annual income last year?
Identification Information	Name, address, telephone	To identify the respondent	Name:_____

(8) Choose the Form and Layout

； Precoding (사전 코딩)

； In questionnaire design, assigning a code to every conceivable response before data collection.

； ⇒ Divide a questionnaire into several parts.
； ⇒ Number the questions in each part.
； ⇒ Precode the questionnaire.
； ⇒ Serially number the questionnaires themselves.
； 사전 코딩을 통해 빠진 자료가 있는지 알아낼 수 있다. 다만 우편 설문지에서만은 예외이다. 응답자들의 익명성이 보장되어야 하기 때문에 질문에 identifier가 있다면 이 약속을 깨기 때문이다. 하지만 이러한 익명성이 결과에 약간 영향을 미친다고 주장하는 사람들도 있다.

(9) Reproduce the Questionnaire

； ⇒ The questionnaire should have a professional appearance.

- ; ⇒ Use a booklet format for long questionnaires.
- ; ⇒ Reproduce each question on a single page (or double-page spread)
- ; ⇒ Avoid the tendency to crowd questions to make the questionnaire look shorter. Place directions or instructions for individual questions as close as possible to the questions.

10) Pretest the Questionnaire

； Pretesting (사전 실험)

- ； The testing of the questionnaire on a small sample of respondents for the pupose of improving the questionnaire by identifying and eliminating potential problems before using it in the actual survey.
- ； ⇒ Always pretest.
- ； ⇒ Test all aspects of the questionnaire, including question content, wording, sequence, form and layout, question difficulty, and instructions.
- ； ⇒ Use respondents in the pretest that are similar to those who will be included in the actual survey.
- ； ⇒ Begin the pretest by using personal interviews.
- ； ⇒ Conduct the pretest by mail, telephone, or electronically if those methods are to be used in the actual survey.
- ； ⇒ Use a variety of interviewers for pretests.
- ； ⇒ The pretest sample size should be small, varying from 15 to 30 respondents for the initial testing.
- ； ⇒ After each significant revision of the questionnaire, conduct another pretest, using a different sample of respondents.
- ； ⇒ Code and analyze the responses obtained from the pretest.

Observational Forms

Summary Illustraion Using the Opening Vignette

International Marketing Research

Technology and Marketing Research

Ethics in Marketing Research

[Discussion]

1. Develop three double-barreled questions related to flying and passengers' airplane preferences. Also, develop corrected versions of each question.

- # Do you think American Airline offers better price and convenience than United Airline?
 - Do you think American Airline offers better price than United Airline?
 - Do you think American Airline offers better convenience than United Airline?

2. List at least 10 ambiguous words that should not be used in framing questions.

Usually, Normally, frequently, often, regularly, occasionally, sometimes, A lot of, A number of, commonly , repeatedly , habitually,

5. A new graduate hired by the marketing research department of a major telephone company is asked to prepare a questionnnaire to determine household preferences for telephone calling cards. The questionnaire is to be administereed in mall intercept interviews. Using the principles of questionnaire design, critically evaluate this questionnaire.

There should be potential for interviewer bias. The quesonnaire is concerning about the telephone card, so it is more likely to be telephone inverview.

1. Your name _____
2. Age _____
3. Marital status _____

⇒ It has to be nominal scale.

① Single
② Married
③ Separated
④ Divorced
⑤ Bereaved

4. Income _____

⇒ It needs too much effort for the respondent to give an answer. So it should be a ordinal scale.

What is your annual imcome?

① Below \$10,000
② \$10,000 - \$50,000
③ \$50,000 - \$100,000
④ \$100,000 - \$150,000
⑤ \$150,000 and above

⇒ 이런 것들은 뒤로 가는 것이 좋다. (개인 정보이기 때문)

5. Which, if any, of the following telephone calling cards do you have?

1. AT&T____
2. MCI____
3. Spring____
4. Others____

⇒ "When" and "Where" is not clear.

Which of the following telephone calling cards have you used at home during the past month? In case of more than one brand, please mark all the brands that apply.

6. How frequently do you use a telephone calling card?

Infrequently
Very frequently

①
②
③
④
⑤
⑥
⑦

⇒ Using interval scale is improper. It has to be changed to ordinal scale like this:

How much time do you use telephone calling card in a week?

- ① never use
- ② 1~5
- ③ 6~10
- ④ 10~15
- ⑤ 15 and above

7. How do you think of the telephone calling card offered by AT&T?

⇒ It is 'too' wide and open-ended question. It may not be in exploratory question, but is not proper in survey. My suggestion is followed.

[Semantic Differential Scale / Stapel Scale]

How do you think of the telephone calling card offered by AT&T?

Cheap	Expensive
Unreliable	Reliable
High quality	Low quality
Good Service	Poor Service
Easy	Difficult

8. Suppose your household were to select a telephone calling card. Please rate the importance of the following factors in selecting a card.

- a. Cost per call
- b. Ease of use
- c. Local and long-distance charges included in the same bill
- d. Rebates and discounts on calls
 - ⇒ Double-barrel question. It needs to be separated.
- e. Quality of telephone service
 - ⇒ It is too ambiguous. Meaning voice quality?
- f. Quality of customer service

9. How important is it for a telephone company to offer a calling card?

Not important						Very important
①	②	③	④	⑤	⑥	⑦

⇒ Very ambiguous question. Important for us or important for the telephone company?

10. Do you have children living at home?

⇒ Has to be changed in ordinal scale or ratio scale (how many children do you have at home?)
⇒ Also it has to be moved to the first or the end of questions.

obtained by visiting the Web sites of these companies (www.ibm.com, www.sony.com, www.dell.com, www.hp.com)

Bloomington's case on page 379

Develop an application of "Importance-Performance Analysis." (read article)

Review the attached "professional" questionnaires.

What are their strengths?

What are their weaknesses?

What kind of data analysis might be appropriate?

[Lecture Note]

7 Guidelines

; 1. Questions should be pretested

- ; Established by Zippo V. Rogers (1963), courts often cite this case when they review survey questionnaires.
- ; Through pretests, one can determine whether questions can be
 - ; handled properly by an interviewer
 - ; correctly understood by respondents
 - ; easily answered by respondents
- ; Eight revisions are not uncommon!

; 2. Questions that appear to predispose respondents must be avoided

- ; Example of a leading question from Universal City V. Nintendo (1984):
 - ; "To the best of your knowledge, was the Donkey Kong game made with the approval or under the authority of the people who produce the King Kong movies?"
- ; A better question would have been:
 - ; "As far as you know, who makes Donkey Kong?"

; 3. Question wording must be direct, clear, and unambiguous

- ; In Black & Decker v. N. A. Philips (1986), consumers said two portable vacuum cleaners shared several common features.
- ; The court declared that the survey lacked relevancy because its ambiguous questions did not distinguish between products and brand names.
- ; Example of an ambiguous term
 - ; "Big government" means...
 - ; welfare statism
 - ; corporatism
 - ; federal control
 - ; bureaucracy

; 4. Respondents must appear to be capable of understanding the topics that questions are raising

- ; It is inappropriate for a survey to layman respondents to use words such as:
 - ; compensatory damages
 - ; punished

6. IBM would like to conduct an Internet survey to determine the image of IBM PCs and the image of its major competitors (Sony, Dell, and Hewlett Packard). Develop such a questionnaire. Relevant information may be

; The court ruled in Firestone V. Crown Center (1985) that layman respondents, unlike lawyers, would not know the legal meaning of such terms.

; 5. Survey questions must relate directly to the legal research questions being litigated

- ; Wording may focus a respondents' attention improperly.
- ; Re: community obscenity standards...
 - ; Biased: "Is it right for adults to purchase magazines at adult book stores?"
 - ; Unbiased: "Are you personally opposed to the existence of adult book stores?"

; 6. Objective questions must include properly stated, complete sets of response scales

- ; In Upjohn V. American Home (1984), the court opined that it was unclear what an "agree" answer to a question actually meant because it was vague about time, place, and circumstance.
- ; Answers must...
 - ; be all inclusive
 - ; be mutually exclusive
 - ; reflect meaningful differences between categories

; 7. Respondents should be instructed explicitly not to guess

- ; Though a "don't guess" instruction cannot guarantee that respondents will comply, it will at least contribute to the face validity of the research.
- ; If a Structured Questionnaire is Indicated...
- ; Follow the Fred Morgan guidelines to generate evidence that can stand legal scrutiny
 - ; Select an appropriate data collection technique
 - ; Mail
 - ; Telephone
 - ; Demonstrate that the benefits of the information outweigh the costs, i.e., demonstrate that there is value in the exercise

Chapter 12. Sampling: Design and Procedures

Overview

Sample or Census

- ; Element**
 - ; Objects that posses the information the researcher seeks and about which the researcher will make inferences.
 - ; 대개 응답자이다.
- ; Population (모집단)**
 - ; The aggregate of all the elements, sharing some common set of characteristics, which compromise the universe for the purpose of the marketing research problem.
- ; 대개의 마케팅 리서치 과제의 목적은 population의 parameter나 특성에 대한 정보를 얻는 것이다.
- ; Census**
 - ; A complete enumeration of the elements of a population or study objects.
- ; Sample**
 - ; A subgroup of the elements of the population selected for participation in the study.

; 표본(Sample)을 고려하는 이유는 예산과 시간 제한 때문이다. 센서스(Census)는 비용 및 시간을 많이 소요하게 된다. 모집단(Population)이 큰 경우에는 어쩔 수 없이 표본만이 유일한 선택이 된다. 다만 B2B 와 같은 경우에는 모집단의 크기가 작으므로 센서스도 고려의 대상이 될 수 있다. 또한 모집단 사이에 큰 편차가 있을 경우 센서스를 하는 것이 좋다.

; 표본 오류의 비용은 비표본 오류(Interviewer bias등)와 비교 고찰해 볼 수 있다. 많은 B2B 연구에서 표본 오차는 센서스의 사용을 뒷받침한다. 그러나 다른 많은 연구에서 비표본 오차는 Total error에 큰 영향을 미치게 된다는 것이 알려져 있다. 센서스가 표본 오차를 제거해도 비표본오차가 Total error에 미치는 영향은 증가한다. (무슨소리야 -.-)그것이 미 통계청에서 센서스를 한 이후에도 정확도를 확인하기 위해 표본 설문조사를 실시하는 이유이다.

; 측정 절차가 제품의 파괴 혹은 소비를 야기하는 것일 경우에도 표본화가 적절하다. 이 경우 센서스를 한다는 것은 많은 양의 제품이 파괴되거나 소비된다는 것을 의미하며, 이는 비용을 크게 증가시키기 때문이다. 또한 심층 면접과 같이 개개의 사례에 집중해야 하는 경우 샘플링이 필요하다. 또한 연구 비밀을 유지해야 하는 경우와 같은 경우에도 샘플링이 센서스보다 우선시된다.

	Conditions Favoring the Use of	
Factors	Sample	Census
Budget	Small	Large
Time available	Short	Long
Population size	Large	Small
Variance in the characteristic	Small	Large
Cost of sampling error	Low	High
Cost of nonsampling errors	High	Low
Nature of measurement	Destructive	Nondestructive
Attention to individual cases	Yes	No

; Sampling v. Nonsampling Errors

- ; The size of a sampling error depends on
 - ; the size of the standard error of the mean or proportion, and
 - ; the level of confidence
- ; Nonsampling errors come from
 - ; nonresponse, and

; response bias of the researcher, interviewer, and the respondent

; Nonsampling Error: Frame Error

- ; In 1936 Literary Digest predicted Alfred Landon would defeat Franklin Roosevelt by 15 points in the race for the presidency.
- ; Sample size was 2.35 million people.
- ; But they were well educated Republicans, and were not representative of the voting universe!

; Fixing Frame Error Provides A Partial Solution

- ; Plus 1 random digit dialing
 - ; take an existing phone number and increase the last digit by one
 - ; unlisted numbers will now be included in the “frame”
- ; But... “nonresponse” will still be a problem!

; Example #1 Of Sampling Error

- ; The mean and standard deviation for the quality point averages of a random sample of 36 college seniors are calculated to be 2.6 and .3, respectively. Find the 95% confidence interval for the mean of the entire senior class.

$$n = 36, Z = 2.03, \sigma = .3, \bar{X} = 2.6$$

$$\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$2.6 - 2.03\left(\frac{.3}{6}\right) < \mu < 2.6 + 2.03\left(\frac{.3}{6}\right)$$

$$2.5 < \mu < 2.7$$

- ; In words, we are 95% confident that the true mean falls within this range.

; Example #2 Of Sampling Error

- ; In a random sample of 500 families owning TVs, it was found that 340 owned sets with 25”+ screens. Find a 95% confidence interval for the actual population proportion with screens this large.

$$n = 500, Z = 1.96, p = \frac{340}{500} = .68, q = 1 - p = .32$$

$$p - Z_{\alpha/2} \sqrt{\frac{p \cdot q}{n}} < P < p + Z_{\alpha/2} \sqrt{\frac{p \cdot q}{n}}$$

$$.68 - 1.96 \sqrt{\frac{(.68 \times .32)}{500}} < P < .68 + 1.96 \sqrt{\frac{(.68 \times .32)}{500}}$$

$$.64 < P < .72$$

- ; In words, we are 95% confident that the true proportion falls within this range.

; Major Question: Is the Sample Representative of the Population?

- ; Assume we appropriately selected, trained, and supervised the field workers. Is the sample collected representative of the population under study? That is, can we infer characteristics of the population based on the sample? This question must be addressed, irrespective of the data collection method utilized.

; Chi-Square Goodness Of Fit : Does the observed distribution fit the expected distribution?

Ho: The distributions are the same

Ha: The distributions differ

$$\alpha = .05, df = k - 1$$

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

reject Ho if $\chi_{calculated}^2 > \chi_{critical}^2$

[Q] Suppose we observed 144 winners at 8 post positions at a race track. Do some post positions produce more winners than others?

POST	1	2	3	4	5	6	7	8	Total
OBS	29	19	18	25	17	10	15	11	144

If post position does not matter, we would expect a uniform distribution, with 144/8=18 winners in each post position.

EXP	18	18	18	18	18	18	18	18	144
-----	----	----	----	----	----	----	----	----	-----

Ho: Winners follow a uniform distribution

Ha: Winners follow some other distribution

$$\alpha = .05, df = k - 1 = 8 - 1 = 7$$

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

reject Ho if $\chi_{calculated}^2 > \chi_{critical}^2 > 14.07$

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 16.3$$

Therefore, reject Ho, and conclude post position makes a difference in producing winners. Post positions 1 and 4 appear to be “best bets.”

; Why Take A Sample Instead Of A Census?

- ; **Cost:** If population is small in number (e.g., airline or auto industry), take a census
- ; **Time:** If results are needed quickly, take a sample
- ; **Accuracy:** Ceteris Paribus, a census is more accurate than a sample
- ; **Destructive Testing (bombs):** A sample is appropriate and essential

The Sampling Design Process (표본화 설계 절차)

(1) Define the Target Population (대상 모집단 정의)

; Target population (대상 모집단)

; **The collection of elements of objects that posses the information the researcher seeks and about which the researcher will make inferences.**

; 대상 모집단은 ① element ② sampling unit ③ extent ④ time frame에 대해서 정의되어야 한다.

; Element: Who will be sampled? Purchasing agents? Students? Voters?

; Sampling Unit: Where will the sampling be done? Corporate address? Supermarkets? Residential address?

; Extent: What is the geographic area of the sample? New York? Europe?

; Time: Over what period? The last three years? Last week? A specific week?

; Sampling unit (표본 단위)

; The basic unit containing the elements of the population to be sampled.

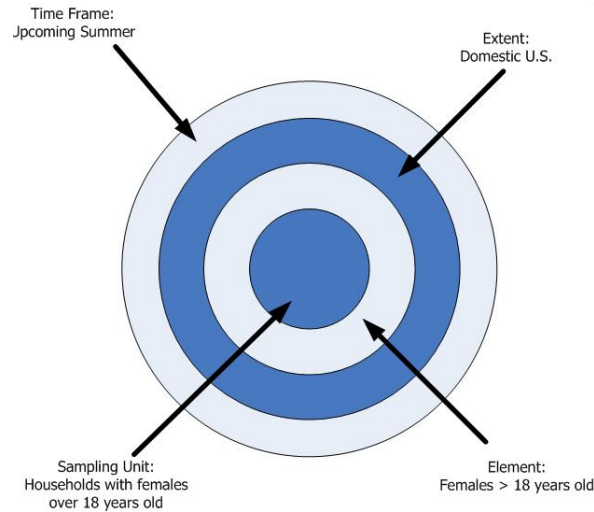
; Extent

; Refers to geographical boundaries.

; ex) Revlon 사례의 경우 미국 국내 시장만 해당됨.

; Time frame

; Time period of interest



(2) Determine the Sampling Frame (표본화 프레임 결정)

; Sampling frame

; A representation of the elements of the target population. It consists of a list or set of directions for identifying the target population.

; 이는 telephone book, telephone number generating program, association directory listing, mailing list, city directory, map 등등이 될 수 있다. Listing이 존재하지 않는다면 Compile해야 한다.

; 모집단 요소에서 목록을 compile 하는 절차는 어렵고 불완전하며, Sampling frame error를 가져올 수 있다.

; Sampling frame error

; An error occurred that elements may be omitted, or the list may contain more than the desired population.

; 예를 들어 전화번호부만 해도 3개의 sampling frame error를 내포하고 있다 : ① 목록에 없는 번호는 포함되어 있지 않다(당연하군...-a) ② 전화번호부가 출간된 이후 이사온 사람들의 전화번호도 포함되어 있지 않다. ③ 전화번호부가 출간된 이후 이사하여 비활성화된 전화번호를 포함하고 있다.

; 모집단과 표본 frame 간의 작은 차이는 무시될 수 있지만 대개의 경우 조사자는 sampling frame error를 인식하고 다루어야 할 필요가 있다.

; Sampling frame error를 다루는 3가지 방법

; ① The population can be redefined in terms of the sampling frame.

; ② The representativeness of the research frame can be verified during the data collection process.

; ③ The data can be statistically adjusted by weighting under- or overrepresented segments to achieve a more representative sample.

(3) Select a Sampling Technique (표본화 방법 선택)

; Nonprobability sampling (비확률? 표본화)

; Sampling technique that does not use chance selection procedures but instead, relies on the personal judgment of the researcher.

; 이 때 연구자는 표본을 자기가 원하는대로 임의로 선택하거나 표본에 포함시킬지를 의식적으로 결정할 수 있다. 대표적인 예로는 거리, 소매점 혹은 매장에서 사람들을 면접하는 것을 들 수 있다. 이 때 문제점은 객관적으로 표본 결과의 정확도(precision)를 측정하는 방법이 없다는 것이다.

; Precision (정확도)

; The level of uncertainty about the characteristic being measured. Greater precision implies smaller sampling error.

; Probability sampling (확률 표본화)

; A sampling procedure in which each element of the population has a fixed probabilistic chance of being selected for the sample.

; 모든 잠재적 표본이 동등하게 선택(selection)될 가능성을 가질 필요는 없고, 주어진 크기의 특정 표본을 선택하는 가능성을 명시하는 것이 가능하다. (?) 표본에 대해 신뢰 구간을 계산할 수 있고, 이는 모집단에 대한 샘플 결과가 대상 모집단에 대한 추론을 이끌어내기에 얼마나 통계적으로 의미있는지를 제시할 수 있다.

(4) Determine the Sample Size (표본 크기 결정)

; Sample size

; The number of elements to be included in a study.

; 표본 크기를 결정할 때에는 계량적(quantative) 및 비계량적(qualitative) 둘 다의 방식으로 숙고해 보아야 한다.

; 표본 크기를 결정할 때 중요한 질적 인자(qualitative factor)는 : ① the importance of the decision ② the nature of the research ③ the number of variables ④ the nature of analysis ⑤ sample sizes used in similar studies ⑥ resource constraints 등이 있다.

; 중요한 결정일수록 정보가 정확해야하며, 이는 보다 큰 표본을 택해야 함을 의미한다.

; 연구의 성질도 표본 크기에 영향을 미치는데, 표적 집단 면접과 같은 탐색적 연구에서는 적은 표본에 초점을 맞춘다. 하지만 서술적 설문조사와 같은 conclusive research에서는 표본 크기가 커야 한다.

; 변수의 수가 많을수록 표본의 수도 늘어나야 한다. 문제 정의 설문조사와 같은 경우는 보통 1000-2500 명의 표본이 필요하다.

; 진보적인 방법을 사용한 자료 조사와 같은 경우나 전체 모집단이 아닌 하위집단 분석과 같은 경우 큰 표본을 필요로 한다.

; 이전의 조사도 표본 크기를 예측하는데 도움을 준다.

; 마지막으로 표본 크기는 돈, 인적 자원 및 시간에 의해 제약받는다.

Type of Study	Minimum size	Typical Range
Problem-identification research (e.g. market potential)	500	1000-2500
Problem-solving research (e.g. pricing)	200	300-500
Product tests	200	300-500
Test-marketing studies	200	300-500
TV/radio/print advertising (per commercial or ad tested)	150	200-300
Test-market audits	10 stores	10-20 stores
Focus groups	2 groups	10-15 groups

(5) Execute the Sampling Process

；표본 설계의 다양한 세부사항들을 구현하는 것을 의미한다. 예를 들어 가구가 표본 단위일 경우, 가구에 대한 절차적 정의를 필요로 하게 된다. 예를 들어 비어있는 가구 단위에 대해서 다시 전화를 것 걸인가에 대해서도 명세화되어야 한다. 때때로 대상 모집단에 속하는 것을 확실하게 하기 위해 잠재적 응답자들을 제한하는 것도 필요하다. 이러한 경우 면접의 시작시에 자격 조건을 검사하는 질문(qualifying)을 던져야 한다. 예를 들어 heavy user를 대상으로 하는 면접일 경우 이를 식별하는 절차가 필요하다.

Nonprobability Sampling Techniques (비확률 표본 추출 기법)

(1) Convenience Sampling (편의 표본 추출)

；A nonprobability sampling technique that attempts to obtain a sample of convenient elements. The selection of sampling units is left primarily to the interviewer.
；때때로 응답자들은 적시 적정장소에 있다는 이유만으로 선정된다. 이러한 샘플링의 예는 ① use of students, church groups, and members of social organizations ② mall intercept interviews conducted without qualifying the respondents ③ department stores using charge account lists ④ tear-out questionnaire included in a magazine ⑤ people on the street interview ⑥ internet browser 등이 있다.
；장점 : 값싸고 빠르다는 점. 또한 표본 단위(sampling unit)들이 접근 가능하고, 측정 및 협력하기 쉽다는 점.
；단점 : 결과 표본이 정의 가능한 대상 모집단의 대표(representative)가 아니라는 것. 따라서 이 표본화 절차는 선택 편차(selection bias)의 문제점을 가지게 된다.
；편의 표본 추출은 모집단의 추론을 끌어내기 위한 서술 혹은 인과적 조사 방식에는 적합하지 않다. 하지만 개념을 생성하고, 통찰력을 얻으며 가설을 설정하는 탐색적 조사에서는 유용하다. 이들은 표적 집단, 선설현 설문지, 시현 연구 등에서 쓰일 수 있다.

(2) Judgmental Sampling (판단적 표본 추출)

；A form of convenience sampling in which the population elements are purposeively selected based on the judgment of the researcher.

；대표적인 예 : ① test markets selected to determine the potential of a new product ② purchase engineers selected in industrial marketing research because they are considered to be representative of the company ③ bellwether precincts selected in voting behavior research ④ expert witnesses used in court ⑤

department stores selected to test a new merchandising display system
；장점 : 저렴하고, 편리하며, 빠르다.
；단점 : 조사자의 숙련도(expertise)와 창의력(creativity)에 크게 의존하는 경향이 있다.
；일반적으로 모집단이 명시적으로 정의되지 않기 때문에 특정 모집단에 대한 일반화는 불가능하다. 이 표본화 기법은 폭넓은 모집단 일반화가 필요하지 않은 경우 가장 적절하다.

(3) Quota Sampling (할당 표본 추출)

；A nonprobability sampling technique that is a two-stage restricted judgmental sampling. The first stage consists of developing control categories or quotas of population elements. In the second stage, sample elements are selected based on convenience or judgment.

；1st stage : Using judgment to identify relevant categories such as age, sex, or race, the ressearcher estimates the distribution of these characteristics in the target population.
；2nd stage : Elements are selected using a convenience or judgment process.
；proportionate quota가 할당되어 모집단의 구성을 표본에 반영하도록 할 수 있다.
；이 표본화 기법에는 많은 잠재적 문제점이 있다. 의미있는(relevant) 특성들이 쿼터-설정 단계에서 간과되어 표본이 모집단을 적절히 통제된 특성 상에서 반영하지 못할 수 있다. 각 쿼터 내에서 선택되는 element들은 convenience or judgmental 이기 때문에 선택 편차(selection bias)가 발생할 수 있다. 불친절하거나 잘 차려입지 않은 사람, 혹은 바람직하지 않은 위치에 거주하는 사람들을 피할 수도 있다. 할당 표본 추출은 표본화 오류의 평가를 허용하지 않는 점에서도 제한이 있다.
；할당 표본 추출은 대표 표본을 상대적으로 낮은 비용에 얻을 수 있다. 할당 표본들은 상대적으로 추출하기 쉽다(convenient). 적절한 통제만 가해진다면 할당 표본 추출은 일반적인 probability sampling에 가까운 결과를 얻을 수 있다.

(4) Snowball Sampling (눈덩이 표집)

；A nonprobability sampling technique in which an initial group of respondents is selected, usually at random. Subsequence respondents are selected based on the referrals or information provided by the initial respondents. This process may be carried out in waves by obtaining referrals from referrals.

；이 방식은 probability sample에서 시작하지만 nonprobability sample에서 끝난다. 이는 소개된 응답자들이 자신을 소개한 사람들과 유사한 인구통계적 / 심리분석적 특성을 갖추는 경향이 있기 때문이다.
；이는 연구하는 특성이 상대적으로 드물거나 모집단에서 식별하기 힘들 경우 사용된다. 예를 들어 food stamp 와 같은 것은 비밀로 지켜진다. 흩어진 소수집단 등과 같은 경우도 referral 없이는 찾아내는 것이 힘들다. 산업 조사(industrial research)에서 스노우볼 표본 추출은 구매자/판매자 짝을 식별하기 위해 쓰인다.
；장점 : 모집단의 원하는 특성을 찾아낼 가능성을 증가시켜 준다. 상대적으로 적은 표본 분산도와 비용을 가진다.

Probability Sampling Techniques (확률 표본 추출 기법)

；Sampling efficiency

；Concept that reflects a trade-off between sampling cost and precision.
；대개 비용 증가는 정확성의 증대를 가져온다.

(1) Simple Random Sampling (단순 무작위 추출법)

；A probability sampling technique in which element in the population has a known and equal probability of selection. Every element is selected independently of every other element, and the sample is drawn by a random procedure from a sampling frame.
；모집단의 각 원소들은 알려지고 동등한 선택 확률을 가진다. 각 원소들은 다른 모든 원소들에 대해 독립적으로 선택된다. 예를 들자면 이름을 상자 안에 넣고 혼든 다음에 복권 추첨식으로 이름을 추출하는 것이 단순 무작위 추출법이다.
；전화 면접에서는 무작위-숫자 전화걸기(random-digit dialing) 기법으로 전화번호를 만들어낸다. 이 방식이 모든 가구의 전화들이 표본에 포함될 동등한 기회를 제공한다고 하여도 이런 식으로 생성된 모든 숫자가 전화번호가 되는 것은 아니다. 작동하지 않는 번호를 식별하고 제거하기 위한 몇가지 수정이 필요하다.
；장점 : 쉽게 이해되고 대상 모집단의 대표(representative)인 자료들을 생성해내는데 쓰일 수 있다. 대개의 통계적 추

론들은 단순 무작위 추출법이 사용되었음을 가정하고 접근한다.

； 단점 ： ① construct a sampling frame for SRS is difficult ② STS can be expensive and time-consuming because the sampling frame may be widely spread over a large geographical area ③ SRS often results in lower precision, producing samples with large standard error ④ samples generated by this technique may not be representative of the target population, particularly if the sample size is small. 평균적으로는 모집단의 대표로 뽑혀나올지 몰라도, 주어진 임의 표본들은 대상 모집단을 잘못 대표하고 있을지도 모른다. 이러한 이유로 SRS 기법은 마케팅에서는 잘 사용되지 않는다.

(2) Systematic Sampling (계통 추출법)

； A probability sampling technique in which the sample is chosen by selecting a random starting point and then picking every ith element in succession from the sampling frame.

； Sampling interval ： 원소들이 뽑혀 나오는 주기(frequency). 이는 모집단 크기 N을 샘플 크기 n으로 나누면서 만 들어진다.

； 만약 전화번호북이 sampling frame으로 쓰인다면 원소들은 알파벳 순으로 정렬된다. 어떤 경우에 이러한 순서는 연구자가 관심있는 속성에 대해 정렬될 수도 있다. 신용카드의 경우에는 잔고나 많은 순서대로 정렬될 수 있고, 산업의 경우에는 연 매출을 대상으로 정렬될 수도 있다. 모집단이 정렬된 방식에 따라 SRS와는 꽤 다른 표본 결과를 보여줄 수 있다. 예를 들어 연 매출 순으로 정렬된 기업의 경우, 크고 작은 회사들을 골고루 포함할 수 있다. 이에 비하면 SRS의 경우에는 대표성이 부족하다.

； 장점 ： 계통 추출법은 SRS보다도 더 비용이 적게 들고 값싼데, 임의 추출이 한 번만에 끝나기 때문이다. 또한 sampling frame에 대한 지식 없이도 적용될 수 있다. (예를 들어 매장에 들어오는 매 x 번째 사람만을 면접할 수도 있다) 이러한 이유로 계통 추출법은 소비자 우편, 전화, 매장 면접법에서 자주 쓰인다.

(3) Stratified Sampling (층화 추출법)

； A probability sampling technique that uses a two-step process to partition the population into subpopulations, or strata. Elements are selected from each stratum by a random procedure.

； ① 모집단은 strata라고 불리는 하위 집단으로 나뉜다. 모든 모집단 구성요소들은 하나의 stratum에 할당되고, 빠지는 예외는 없다. ② 각 stratum의 구성요소들은 임의로 선정된다. 이론적으로는 각stratum에서 원소를 선택하기 위해 SRS가 쓰일 수 있다. 하지만 실제로는 계통 추출법이나 다른 확률 추출법이 쓰인다.

； 이 때 비용을 증가시키지 않으면서도 정확도를 증가시킬 수 있다는 장점이 있다. 모집단은 stratification 변수들을 이용해 분할(partition) 된다. 이 strata는 4가지 기준으로 구성된다 : ① homogeneity(동질성) ② heterogeneity(이질성) ③ relatedness(동족성) ④ cost(비용)

； 다음과 같은 가이드라인이 준수되어야 한다:

- ； ① Elements within strata must be similar or homogeneous
- ； ② Elements between strata must differ or be heterogeneous
- ； ③ The stratification variables must be related to the characteristic of interest
- ； ④ The number of strata usually varies between two and six. Beyond six strata, any gain in precision in more than offset by the increased costs.

； 장점 ： 표본 추출 분산이 감소된다는 점, 표본 추출 비용이 감소된다는 점이다.(보다 측정하고 적용하기 쉽다)

； 변수들로는 인구 통계적 특성, 소비자의 종류, 기업의 크기, 산업의 종류 등이 쓰인다.

(4) Cluster Sampling (군집 추출법)

； A two-step probability sampling technique. First, the target population is divided into mutually exclusive and collectively exhaustive subpopulations called clusters. Then, a random sample of clusters is selected based on a probability sampling technique, such as simple random sampling. For each selected cluster, either all the elements are included in the sample or a sample of elements is drawn probabilistically.

； one-stage cluster sampling ： All the elements in each selected cluster are included in the sample.

； two-stage cluster sampling ： A sample of elements is drawn probabilistically from each selected cluster.

； 군집 추출법의 대표적인 예는 지역 표본 추출이다.

； Area sampling (지역 표본 추출)

； A common form of cluster sampling in which the clusters consist of geographic areas, such as

countries, housing tracts, blocks, and so forth.

； Single-stage area sampling ： Sampling all the elements within a particular area. (예를 들어 선택된 도시 블록 내의 모든 가구들을 조사하는 것)

； Two-stage area sampling ： Only a portion of the households within each block would be sampled.

； 장점 ： 편리성(feasibility) 및 저비용. 가장 비용 대 효과비(cost-effective)가 좋은 표본 추출 방식.

； 단점 ： 별개의 이질적인 군집들이 아니면 부정확한 표본이 생성된다. (예를 들어 한 블록 내의 가구들이 다르기 보다는 비슷할 수 있다) 이는 계산하거나 해석하기 쉽지 않다.

Cluster Sampling	Stratified Sampling
Only a sample of the subpopulations(clusters) is selected for sampling	All of the subpopulations (strata) are selected for sampling
Within a cluster, elements should be different (heterogeneous), whereas homogeneity or similarity is maintained between different clusters	Within a strata, elements should be homogeneous with clear differences (heterogeneity) between the strata
A sampling frame is needed only for the clusters selected for the sample	A complete sampling frame for the entire stratified subpopulations should be drawn
Increases sample efficiency by decreasing cost	Increases precision

Technique	Strengths	Weaknesses
Convenience sampling	Least expensive, least time-consuming, most convenient	Selection bias, sample not representative, not recommended for descriptive or causal research
Judgmental sampling	Low cost, convenient, not time-consuming	Does not allow generalization, subjective
Quota sampling	Sample can be controlled for certain characteristics	Selection bias, no assurance of representativeness
Snowball sampling	Can estimate rare characteristics	Time-consuming
Simple random sampling	Easily understood, results projectable	Difficult to construct sampling frame, expensive, lower precision, no assurance of representativeness
Systematic sampling	Can increase representativeness, easier to implement than STS, sampling frame not necessary	Can decrease representativeness
Stratified sampling	Includes all important subpopulations, precision	Difficult to select relevant stratification variables, not feasible to stratify on many variables, expensive
Cluster sampling	Easy to implement, cost-effective	Imprecise, difficult to compute and interpret results

Choosing Nonprobability Versus Probability Sampling

	Conditions favoring the use of	
Factors	Nonprobability Sampling	Probability Sampling
Nature of research	Exploratory	Conclusive
Relative magnitude of sampling and nonsampling errors	Nonsampling errors are larger	Sampling errors are larger
Variability in the population	Homogeneous (low)	Heterogeneous (high)
Statistical considerations	Unfavorable	Favorable
Operational considerations	Favorable	Unfavorable

Internet Sampling

Issues in Online Sampling

- ； 장점 ： 응답자는 편리하게 설문조사를 완료할 수 있다. 자료 수집이 빠르고 값싸다.
- ； 다만 인터넷 표본 추출과 관련된 가장 큰 이슈는 미국 내 많은 가구의 컴퓨터 소유와 인터넷 접속 부족으로 인한 대 표성의 문제이다.
- ； 또한 인터넷의 대량 사용자(heavy user)가 보다 포함될 확률이 높게 된다. 아무나 참여할 수 있도록 제한 없는 인터 넷 표본은 self-selection bias에 빠질 가능성이 높다
- ； 인터넷을 서핑하는 잠재적인 응답자를 표본 추출하는 것은 생성된 표본이 대상 모집단의 대표(representative)일 때 의미가 있다.
- ； sampling error를 피하기 위해 연구자는 응답자들이 선택되는 pool을 통제할 수 있어야 한다. 또한 응답자가 여러번 응답하지 않도록("stuff the ballot box") 조치를 취해야 한다. 이를 위해서 e-mail 설문조사를 하든지 아니면 e-mail 초대장을 보내어 Web에서 설문조사를 하게 하는 방법 등이 있을 수 있다.

Online Sampling Techniques

； ① Online intercept sampling

- ； 웹사이트에 방문하는 사람들을 가로채어 설문조사에 참여할 기회를 제공한다. 대개 Yahoo! 와 같이 방문자가 많은 곳에서 이루어진다.
- ； Nonrandom
 - ； Every visitor is intercepted. (웹사이트 트래픽이 적고 짧은 시간 내에 설문조사가 이루어지며, 인센티브 가 없을 경우)
 - ； 하지만 The result is convenience sample!
- ； Random
 - ； 임의로 선정하여 설문조사에 참여할지를 묻게 된다. 대개 SRS나 계통 추출법이 사용된다. 또한 임의화는 대표성을 향상시키고 같은 응답자로부터의 복수의 응답을 방지한다.

； ② Recruited online sampling

- ； 대개 보상(incentive)을 약속받는다. 멤버들은 상세한 심리분석적/인구통계적/인터넷 사용량/제품 소비 정보 등 을 가입시에 제공하게 된다.
- ； Panel
 - ； Recruited Panels
 - ； Opt-in Panels
 - ； 표본의 신뢰성(reliability)과 무결성(integrity)을 유지하기 위해, 다음 절차가 필요하다.
 - ； ① Password protection ： A respondent completes the survey only one time.
 - ； ② Reminder invitation ： To increase the number of respondents in the survey and to improve overall response rates.
 - ； ③ Summary of the survey findings ： To increase the number of respondents in the survey

- and to improve overall response rates.

； Nonpanel

- ； Requesting potential respondents to go online to answer a survey.
- ； 쪽지(flier)를 나누어주어서 설문조사에 참여하도록 하든지, e-mail 리스트를 임대하여 뿌리는 방법.
- ； 짧은 전화를 통해 스크리닝 하는 방법도 있다.

； ③ Other techniques

Summary Illustraion Using the Opening Vignette

International Marketing Research

Technology and Marketing Research

Ethics in Marketing Research

Chapter 13. Sampling: Final and Initial Sample Size Determination

Overview

- Normal distribution
 - A basis for classical inference that is bell-shaped and symmetrical in appearance. Its measures of central tendency are all identical.

Definitions and Symbols

- Parameter (모수 母數)
 - A summary description of a fixed characteristic or measure of the target population. A parameter denotes the true value that would be obtained if a census(a survey of the complete population), rather than a sample, was taken.
- Statistic (통계량)
 - A summary description of a characteristic or measure of the sample. The sample statistic is used as an estimate of the population parameter.
- Precision level (정확도)
 - When estimating a population parameter by using a sample statistic, the precision level is the desired size of the estimating interval. This is the maximum permissible difference between the sample statistic and the population parameter.
- Confidence interval (신뢰 구간)
 - The range into which the true population parameter will fall, assuming a given level of confidence.
- Confidence level (신뢰 수준)
 - The probability
- Random sampling error (임의 표본 오류)
 - The error due to the particular sample selected being an imperfect representation of the population of interest.

Variable	Population	Sample
Mean	μ	\bar{x}
Proportion	π	p
Variance	σ^2	s^2
Standard deviation	σ	s
Size	N	n
Standard error of the mean	$\sigma_{\bar{x}}$	$S_{\bar{x}}$
Standard error of the proportion	σ_p	S_p
Standardized variate (z)	$\frac{X - \mu}{\sigma}$	$\frac{X - \bar{x}}{s}$

The Sampling Distribution (표본 분포)

- The distribution of the values of a sample statistic computed for each possible sample that could be drawn from the target population under a specified sampling plan.
- Statistical inference (통계적 추론)
 - The process of generalizing the sample results to the population results

Central limit theorem (중심 극한 정리)

- As the sample size increases, the distribution of the sample mean of a randomly selected sample approaches the normal distribution.

Standard error (표준 오차)

- The standard deviation of the sampling distribution of the mean or proportion

z value

- The number of standard errors that a point is away from the mean

- 평균(mean)의 표본 분포는 정규 분포(normal distribution)이다. 사실 30개보다 적으면 이항분포에 가깝고 30개보다 많아야 정규 분포가 된다.

(2) 평균(mean) 표본 분포의 평균(mean)은 $\bar{X} = \frac{\left(\sum_{i=1}^n X_i\right)}{n}$ 이다.

- 표본 분포의 표준 편차는 평균(mean)이나 확률(proportion) 표준 오차(standard error)로 불린다.

평균의 표준 오차(편차) : $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ (즉 표준편차 / sqrt(개수))이다!

확률의 표준 오차(편차) : $\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$ (즉 sqrt(pq / n)이다)

- 표준 편차 s(혹은 σ)가 알려지지 않은 경우 다음의 식을 이용하면 된다.

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \quad (\text{즉 분산을 구한 다음, 그걸 } n-1 \text{로 나누어서 루트를 씌운다})$$

혹은

$$s = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}}{n-1}} \quad \text{이걸 이용한다.}$$

σ 가 s에 의해 예측되는 경우,

평균의 표준 오차 : est. $\sigma_{\bar{X}} = \frac{s}{\sqrt{n}}$ 가 된다.

* 교수님 note에서는 전체 range를 구한 다음에 6으로 나누는 방법도 사용한다.

- 확률의 표준 오차는 표본 확률 p를 모집단 확률의 예측자로 사용하여 다음과 같이 구할 수 있다.

확률의 표준 오차 : est. $s_p = \sqrt{\frac{p(1-p)}{n}}$ 가 된다.

- 두 점 사이의 곡선 아래의 표본 분포(정규 분포) 면적은 z value로 계산될 수 있다.

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

Statistical Approach to Determining Sample Size

- 표본 크기를 정할 때에는 몇가지 질적(qualitative) 인자가 고려되어야 한다. 여기에는 결정의 중요성, 연구의 성격, 변수

의 수, 분석의 성격, 유사한 사례에서 사용된 표본의 수, 발생률(incidence rate), 완성률(complete rate), 자원 제약 등이 있다.

; 통계적으로 결정되는 표본 크기는 총 혹은 최종 표본 크기이다. (제거되는 응답자를 제외한 최종 수를 말함)

; incidence rate과 complete rate을 고려하여 최초 표본 크기는 좀 더 커야 한다.

The Confidence Interval Approach (신뢰 구간 접근 방식)

- ; ① 신뢰구간 계산
- ; ② 평균의 표준 오차 계산
- ; ③ 신뢰구간 식에 대입

Steps	Means (평균)	Proportions (확률)
1. Specify the level of precision (정확도 수준을 결정)	$D = \pm \$5.00$	$D = p - \pi = \pm 0.05$
2. Specify the confidence level (신뢰수준을 결정)	CL = 95%	CL = 95%
3. Determine the z value associated with the CL (신뢰수준과 연관된 z-value를 결정)	z value = 1.96	z value = 1.96
4. Determine the standard deviation of the population (모집단의 표준 편차를 결정)	Estimate σ $\sigma = 55$	Estimate π $\pi = 0.64$
5. Determine the sample size using the formula for the standard error (표준 오차 공식을 이용하여 표본 크기 결정)	$n = \frac{\sigma^2 z^2}{D^2}$ $n = \frac{55^2 (1.96)^2}{5^2}$ <p>n = 465</p>	$n = \frac{\pi(1-\pi)z^2}{D^2}$ $n = \frac{0.64(1-0.64)(1.96)^2}{(0.05)^2}$ <p>n = 355</p>
6. If necessary, reestimate the confidence interval by employing s to estimate σ (필요하다면 s를 이용해 σ 를 예측해 신뢰구간을 다시 예측한다)	$= \bar{X} \pm z s_{\bar{X}}$	$p \pm z s_p$

Adjusting the Statistically Determined Sample Size

Incidence Rate (발생률)

- ; The rate of occurrence of persons eligible to participate in the study expressed as a percentage.
- ; 연구중인 특성을 갖춘 모집단의 인구 확률에 영향받는다. 낮은 발생률을 갖춘 특성을 연구하고 있다면, 높은 발생률을 갖춘 경우보다 많은 사람들을 만나야 할 것이다. 25~55살의 보석류 판매를 연구하고 있다면, 전체 인구 중에서 조건에 맞는 사람의 수가 75%라고 할 때 이 경우 발생률은 0.75가 된다. 이는 평균적으로 1명의 조건에 맞는 사람을 만나기 위해 1.33 (1.00 / 0.75) 명의 사람을 만나야 한다는 점을 의미한다. 조건을 갖춘 응답자에 대한 기준이 추가될 경우, 그 만큼 발생률은 낮아지고 만나야 할 사람의 수는 늘어난다. (예제는 쉬우니 skip)

Completion Rate (완성률)

- ; The percentage of qualified respondents who complete the interview. It enables researchers to take into account anticipated refusals by people who qualify.

; 면접 응답 성공률이 80%라고 예측하였을 경우 1개의 완성본을 얻기 위해 만나야 할 사람의 수는 1.25명 (1.00/0.8)이 된다.

Combined Adjustment

- ; Initial sample size = Final sample size / (Incidence rate X Completion rate)
- ; Response rate
- ; It is the number of completed interviews divided by the number of eligible respondents in the sample.
 - ; Response rate = Number of completed interviews / Number of eligible respondents in the sample

Online Sampling Adjustments

- ; Online 면접의 샘플 크기는 대개 통계적으로 결정된 것 보다 크다. 그 이유는 큰 표본을 수집하고 분석하는데 드는 marginal cost가 낮기 때문이다.
- ; 또한 큰 표본은 하위 집단이나 세그먼트 단계의 보다 상세한 분석을 가능하게 한다. 온라인 연구에 대한 가장 큰 과제는 모집단의 대표성을 반영하는 랜덤 표본을 생성하고, 응답률을 높이는 것이다. 이는 소비자 설문조사를 시행할 때 보다 문제가 된다.

Summary Illustration Using the Opening Vignette

International Marketing Research

Technology and Marketing Research

Ethics in Marketing Research

[Z-value table]

신뢰구간이 90% 일 경우(0.450) : Z = 1.65

신뢰구간이 95% 일 경우(0.475) : Z = 1.95

신뢰구간이 99% 일 경우(0.495) : Z = 2.58

[Lecture Note]

Sample Size Determination For Estimating Proportions

$$n = t^2 \frac{p(1-p)}{(p-P)^2}$$
 , where p = sample proportion, and P = population proportion

For “large” samples, t=1.96 for 95% confidence

A “typical” tolerance value for (p - P) = .025

If there is no prior knowledge regarding the sample population p, assume p=.5

$$n = 1.96^2 \frac{.5(1-.5)}{(.025)^2} = 1,537 \text{ observations}$$

If the population is finite, the following sample sizes are appropriate....

Example For Estimating Proportions

How many registered voters in the US must be sampled to estimate preference in a two-way presidential race

within 1% at the 95% confidence level?

$$(p-P) = .01, t = 1.96, p = .5$$

$$n = 1.96^2 \frac{0.5(1-0.5)}{0.012} = 9,604 \text{ observations}$$

Algebraic Relationships (확률의 샘플 크기 구하기)

Precision = (# of errors) * (size of standard error)

$$D = p - \pi = z\sigma_p \text{이다. 그런데 } \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \text{이므로,}$$

$$D = t \sqrt{\frac{p(1-p)}{n}} = \text{precision이 된다. (여기에서 t는 z와 같다고 생각하면 된다)}$$

$$D^2 = t^2 \frac{p(1-p)}{n}$$

$$nD^2 = t^2 p(1-p)$$

$$n = t^2 \frac{p(1-p)}{D^2} = \text{sample size for proportions}$$

Example: Determining A Sample Size For Estimating Proportions

To determine the effectiveness of the advertising campaign for a new VCR, management would like to know what percentage of the households are aware of the new brand. The advertising agency thinks that this figure is as high as 70%. The management would like a 95% confidence interval and a margin of error no greater than $\pm 2\%$.

$$n = t^2 \frac{p(1-p)}{D^2} = \text{sample size for proportions}$$

$$n = 1.96^2 \frac{.7(1-.7)}{.022} = 2,017 \text{ observations}$$

What If p Is Over Estimated?

In the previous example, suppose 2,017 observations were taken and it was found that only 25% were aware, not 70% as previously estimated.

Is the sample large enough for the $\pm 2\%$ tolerance specified?

$$D = t \sqrt{\frac{p(1-p)}{n}}$$

$$D = 1.96 \sqrt{\frac{.25(1-.25)}{2,017}} = 1.9\% \text{ (즉 이런 식으로 직접 D를 구해주어 정확도를 계산한다)}$$

Yes, when p is over estimated, the sample will be large enough.

What is the maximum sample needed?

$$n = 1.96^2 \frac{.5(1-.5)}{.022} = 2,401 \text{ observations}$$

Sample Size Determination For Estimating Means

$$n = t^2 \frac{s^2}{D^2}, \text{ where } s = \text{sample standard deviation}$$

For “large” samples, $t=1.96$ for 95% confidence

A “typical” tolerance value for $D = .025$

If there is no prior knowledge regarding the sample standard deviation, it can be estimated by dividing the range of values by 6.

Example For Estimating Means

Past experience indicates that weekly income is normally distributed with a standard deviation of \$500. How large of a sample is needed to estimate the income with 99% confidence and a \$60 tolerance?

$$t = 2.58 \text{ for } 99\% \text{ confidence, } s = \$500, D = \$60$$

$$n = t^2 \frac{s^2}{D^2} = 2.58^2 \frac{(500)^2}{60^2} = 463 \text{ observations}$$

Algebraic Relationships (Mean의 샘플 크기 구하기)

Precision = (# of errors) * (size of standard error)

$$D = \bar{x} - \mu = z\sigma_{\bar{x}}, \text{ 그런데 } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \text{이므로}$$

$$D = t \left(\frac{s}{\sqrt{n}} \right) = \text{precision}$$

$$D^2 = t^2 \frac{s^2}{n}$$

$$nD^2 = t^2 s^2$$

$$n = t^2 \frac{s^2}{D^2} = \text{sample size for means}$$

Example For Estimating Sample Sizes For Means

The management of a local restaurant wants to determine the average monthly amount spent by households in fancy restaurants. Some households do not spend anything at all, whereas other households spend as much as \$300 per month. Management wants to be 95% confident of the findings and does not want the error to exceed $\pm \$5$.

$$t = 1.96, s = \text{range}/6 = \$300/6 = \$50, D = \$5 \text{ (정확도)}$$

$$n = t^2 \frac{s^2}{D^2} = \text{sample size for means}$$

$$n = 1.96^2 \frac{(50)^2}{5^2} = 384 \text{ observations}$$

After the survey was conducted, the average expenditure was found to be \$90.30 and the standard deviation was \$45. Construct a 95% confidence interval.

$$\bar{x} = \$90.30, t = 1.96, s = \$45, n = 384$$

Precision = (# of errors) * (size of standard error)

$$D = t \frac{s}{\sqrt{n}} = \text{precision} = 1.96 \frac{45}{\sqrt{384}} = \$4.5$$

x ± precision, **\$90.30 ± \$4.50**

We are 95% confident that the true mean falls between: \$85.80 and \$94.80

Suppose A Mean And A Proportion Are To Be Estimated In The Same Survey

We wish to estimate with 95% confidence the proportion of a population that owns a TV within 3%, and the average number of hours viewed weekly within 1 hour.

$$n = t^2 \frac{p(1-p)}{D^2} = \text{sample size for proportions}$$

$$n = 1.96^2 \frac{.5(1-.5)}{.03^2} = 1,068 \text{ observations}$$

$$n = t^2 \frac{s^2}{D^2} = \text{sample size for means}$$

What is s? Suppose it is thought viewing ranges from 0 to 10 hours daily, or 0 to 70 hours weekly. The range of 70 divided by 6 suggests that s=11.7 hours.

$$n = 1.96^2 \frac{(11.7)^2}{1^2} = 526 \text{ observations.}$$

Thus, to estimate both the proportion and the mean as specified, make 1,068 observations.

Thus -1.0 < z < 1.0

3. The management of a local restaurant wants to determine the average monthly amount spent by households in fancy restaurants. Managements wants to be 95% confident of the finding and does not want the error to exceed plus or minus \$5.

a. After the survey was conducted, the average expenditure was found to be \$90.30 and the standard deviation was \$45. Construct a 95% percent confidence level.

$$\begin{aligned} &= \bar{X} \pm z s_{\bar{X}} \quad (\text{뭔가 자료가 부족해서 제대로 계산할 수 없음..}) \\ &= \$90.30 \pm 1.96 \frac{\$45}{\sqrt{n}} \end{aligned}$$

b. What sample size would have resulted in a standard deviation of \$45?

$$\begin{aligned} D &= z \frac{s}{\sqrt{n}} = 1.96 \frac{45}{\sqrt{n}} = 5 \\ n &= 311.1696, \text{ 즉 } n = 312 \end{aligned}$$

4. To determine the effectiveness of the advertising campaign for a new videocassette recorder, management would like to know what percentage of the households are aware of the new brand. The advertising agency thinks that this figure is as high as 70 percent. Management would like a 95% confidence interval and a margin of error no greater than plus or minus 2 percent.

a. What sample size should be used for this study?

$$\begin{aligned} n &= t^2 \frac{p(1-p)}{D^2} = 1.96^2 \frac{0.7(1-0.7)}{0.02^2} = 2016.84 \\ n &= 2017 \text{ observants.} \end{aligned}$$

b. Suppose that management wanted to be 99% confident but could tolerate an error of plus or minus 3%. How would the sample size change?

$$\begin{aligned} n &= t^2 \frac{p(1-p)}{D^2} = 2.58^2 \frac{0.7(1-0.7)}{0.03^2} = 1553.16 \\ n &= 1554 \text{ observants.} \end{aligned}$$

5. Assuming that n=100 and σ=5, compute the standard error of the mean.

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{100}} = 0.5$$

[Discussion]

1. Using Table 2 of the Appendix of Statistical Tables, calculate the probability that

a. z is less than 1.48

$$z < 1.48 = 0.4306 + 0.5 = 0.9308$$

b. z is greater than 1.90

$$z > 1.90 = 0.5 - 0.4713 = 0.0287$$

c. z is between 1.48 and 1.90

$$1.48 < z < 1.90 = 0.4713 - 0.4306 = 0.0407$$

d. z is between -1.48 and 1.90

$$-1.48 < z < 1.90 = 0.4713 + 0.4306 = 0.9019$$

2. What is the value of z if

a. 60% of all values of z are larger

$$z < 0.0255 = 0.1$$

$$\text{Thus } z > -0.0255$$

b. 10% of all values of z are larger

$$z < 1.28 = 0.4$$

$$\text{Thus } z > 1.28$$

c. 68.26% of all possible z values (symmetrically distributed around the mean) are to be contained in this interval.

$$z < 1.0 = 0.3413$$