

Psychophysical Scale 적용시 오류에 관한 사례조사 (A case study on misuse of psychophysical scales)

곽지영*, 박성준*, 한성호*

ABSTRACT

Psychophysical data, in general, belong to one of the four scale categories : Nominal, Ordinal, Interval, and Ratio Scale. This paper introduces properties of the four scale categories and describes some psychophysical scales that attempt to measure subjective feeling or opinion of human. In addition, guidelines of analyzing and interpreting measured data are suggested. Some examples of analyzing and interpreting psychophysical data inappropriately are presented especially with category scales which have been used most widely in measuring subjective information.

1. 서론

제품이나 기계 시스템을 설계할 때, 사용자의 요구사항(Design Requirement)을 설계 초기단계에서부터 파악하는 것은 설계과정의 효율증진 측면과 사용 편의성(Usability) 제고의 측면에서 매우 중요한 개념으로 인식되고 있다. 그러나 사용자의 요구사항은 대부분이 정성적(Qualitative)인 내용을 담고 있어 설계사항(Design Specification)으로 사용하려면 사용자의 요구사항을 측정하고 정성적 측정치를 정량화(Quantification)하는 작업이 필수적으로 요구된다. 일반적으로 사용하는 측정방법으로는 크게 나누어, 직접 관찰법(Direct Observation)을 통하여 작업 수행도(Performance)나 행동(Behavior)의 발생빈도(Frequency), 지속시간(Duration) 등을 측정하는 방법과, 각종 Scaling 기법을 통해 측정대상의 성격을 정량화하는 방법이 있다(Meister, 1985).

Psychophysical Scaling 기법은 원래 소리의 크기(Loudness), 빛의 밝기(Brightness), 통증(Pain), 촉감(Touch), 추위(Cold), 미각(Taste), 등 물리적 자극에 대한 인간의 지각강도(Sensory Magnitude)를 수치화하고, 자극의 세기와 지각강도 간의 관계를 파악하기 위한 목적으로 사용되었으나(Gescheider, 1985), 이후의 많은 연구를 통해 이론적인 보장을 거듭하면서, 여러가지 사회적 현상이나 행동특성 등 심리적인 요소의 정량적 측정에 이용되고 있다. 예를 들어, 범죄의 심각성에 따른 형벌적용의 타당성에 대한 시민의 의견(Gescheider, Catlin, and Fontana, 1982), 그림의 미적 가치

* 포항공과대학교 산업공학과

(Aesthetic Value)(Ekman and Künnapas, 1962), 어떤 직업의 사회적 지위와 직업에 대한 선호도(Künnapas and Wikström, 1963), 공공재(Public Goods)의 유용성(Kemp, 1987), 각종 화장품에 대한 소비자의 의견(Moskowitz, 1984), 열차내 행동 및 열차의 진행 방향에 따른 승객이 느끼는 편안함의 정도(한성호 외 4인, 1993) 등은 인간이 느끼는 심리적 선호도나 의견을 정량적으로 파악하여 설계에 이용할 수 있음을 보여 주고 있다.

이와 같이 사용자의 요구사항을 설계에 반영하기 위한 측정방법은 여러가지가 있을 수 있으나, 측정결과는 일반적으로 명목척도(Nominal Scale), 서열척도(Ordinal Scale), 간격척도(Interval Scale), 비율척도(Ratio Scale) 중의 한 부류에 속하게 된다. 본 논문에서는 측정 데이터의 분류형태에 적합한 통계처리 방법 및 자료해석방법의 차이를 소개하며, 또한 부적절하게 사용된 측정 자료의 분석 실례를 통하여, 측정 자료의 올바른 사용방안을 제시하고자 한다.

2. 자료의 분류체계

Stevens(1946)는 기본적으로 각종 측정기법의 적용 결과 얻어지는 척도(Scale)를 명목(Nominal), 서열(Ordinal), 간격(Interval), 비율(Ratio)의 네가지로 분류하였다. 명목척도(Nominal Scale)에서 사용되는 문자, 숫자 등의 기호(Symbol)는 자료의 분류(Classification)나 식별(Identification)을 위한 것이므로, 어떤 대상을 표현하기 위해 숫자가 사용되더라도 이는 정량적인 측정의 의미는 제공하지 못한다. 명목척도의 예로는 노선버스의 번호나 운동선수에게 배정된 고유번호 등을 들 수가 있는데 이들은 모두 대상들 간의 구별을 위해 사용되는 번호이다. 측정된 데이터(Data)에 대하여 각종 변환(Transformation)을 실시할 때 변환된 데이터가 이전의 데이터가 갖는 성질을 그대로 유지하게 되는 것을 불변성(Invariance)이라고 정의하며, 명목척도의 데이터에 대해서는 동일성(Identity)을 잃지 않는 범위 내에서의 모든 변환이 불변성을 갖는다. 명목척도는 네가지 척도 중 가장 낮은 단계이므로 사용가능한 통계량(Statistics)이나 통계적 분석기법(Statistical Analysis Technique)의 적용에 있어서 가장 많은 제약을 받는다. 즉, 각 데이터는 단순한 분류 및 식별의 의미만을 제공하므로, 경우의 수(Number of Cases), 최빈값(Mode), Contingency Correlation 등이 통계량으로서 적절하다.

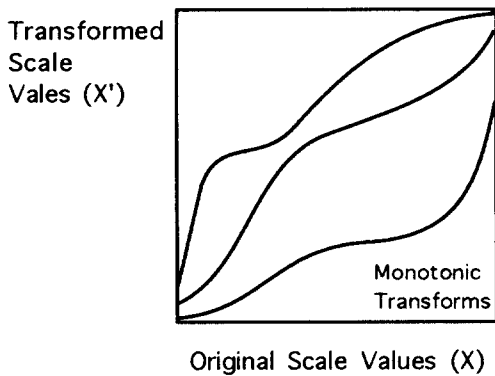
서열척도(Ordinal Scale)에서는 명목척도로 표현된 데이터가 갖는 분류와 식별의 의미와 함께 수치들간의 대소관계를 정의할 수가 있으므로, 그에 따라 측정대상에 대한 순위결정이 가능하다. 그러나 서열척도에서는 임의의 두 값의 차이나 두 측정대상 간의 거리에 관하여는 어떠한 의미도 갖지 않는다. 따라서 서열척도의 데이터에 대해서는 순서를 변화시키지 않는 범위 내에서 모든 종류의 변환, 즉, 그림 1의 (a)에 나타난 대로 모든 단조 증가함수(Monotonically Increasing Function)에 의한 변환이 가능하다. 서열척도의 예로는 각종 경기에서의 순위를 들 수가 있는데, 그 외에도 순서의 의미만을 갖는 데이터는 서열척도로 볼 수 있다. 적절한 통계량으로는 중앙값(Median), Percentile, Rank-order Correlation 등이 있으며, 통계적 검증을 위한 방법으로 앞에서 언급한 명목척도의 데이터와 마찬가지로, 비모수 검정(Nonparametric Test)을 이용하는 것이 적합하다.

서열척도에서는 측정대상을 표현하는 수치들 간의 순서적인 관계만이 의미가 있는데 반해, 간격척도(Interval Scale)는 순서 뿐만 아니라 그들간의 크기의 차이도 의미를 갖고, 척도의 연속성이 존재한다. 따라서 산술평균(Arithmetic Mean)과 표준 편차(Standard Deviation), Pearson Product-moment Coefficient 등을 통계량으로 사용할 수가 있고, 비모수검정 뿐만 아니라 회귀분석

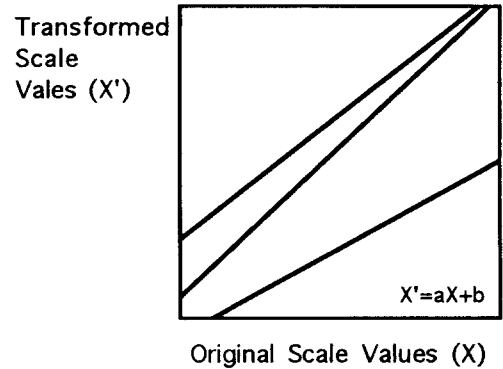
(Regression Analysis), 분산분석(Analysis of Variance, ANOVA) 등의 모수검정(Parametric Test)을 실시할 수 있다. 간격척도의 데이터에는 그림 1의 (b)와 같이 상수로 곱하거나 상수를 가감하는 직선의 함수에 의한 변환이 적용 가능하다. 대표적인 간격척도의 예로는 섭씨(Centigrade)온도와 화씨(Fahrenheit)온도를 들 수 있다.

비율척도(Ratio Scale)는 이전의 모든 척도의 성질을 갖고 있으며, 절대원점(Absolute Zero) 혹은 기준점(Origin)이 존재하여 서로 다른 두 값의 차이 뿐만 아니라 그들 간의 비(Ratio)도 의미를 갖는다. 비율척도의 예로서 무게(Weight)나 길이(Length), 절대온도 등을 들 수 있다. 비율척도 데이터에 대하여 불변성(Invariance)을 유지할 수 있는 변환은 상수로 곱하여 원점을 변화시키지 않는 1차 함수 뿐이며, 전술한 모든 통계량 이외에 기하평균(Geometric Mean), Coefficient of Variation 등이 사용가능하다. 그러나 비율척도의 데이터에 대해 하위 단계에서 이용가능한 변환을 실시하거나 통계량을 계산할 경우, 그 결과는 더이상 비율척도로서의 성질을 갖지 못한다. 예를 들어, 비율척도 데이터에 대하여 산술평균을 구할 경우, 그 결과는 비율척도의 성질을 잃고 간격척도의 데이터가 된다.

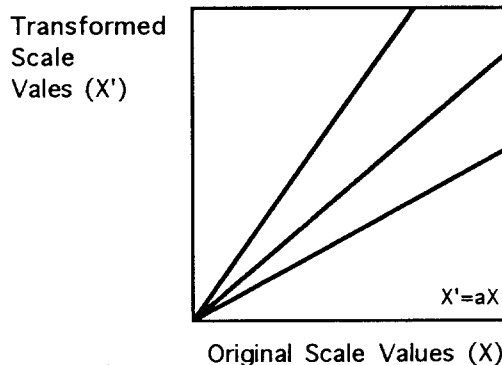
전술한 내용을 요약하면 표 1과 같다.



(a) 서열척도(Ordinal Scales)



(b) 간격척도(Interval Scales)



(c) 비율척도(Ratio Scales)

그림 1. 각 척도의 불변성을 유지할 수 있는 변환의 예 (Gescheider, 1985)

표 1. 자료의 분류체계 요약

구분	적절한 통계량	통계적 검증법
명목척도(Nominal Scale)	경우의수(Number of Cases), 최빈값(Mode), Contingency Correlation 등	비모수검정 (Nonparametric Test) 예) χ^2 - Test
서열척도(Ordinal Scale)	중앙값(Median), Percentile, Rank-order Correlation 등	
간격척도(Interval Scale)	산술평균(Arithmetic Mean), 표준 편차(Standard Deviation), Pearson Product-moment Coefficient 등	모수검정 (Parametric Test) 예) 회귀분석(Regression), 분산분석(ANOVA)
비율척도(Ratio Scale)	기하평균(Geometric Mean), Coefficient of Variation 등	

3. Psychophysical Scales

Psychophysical Scale은 측정방법에 따라 크게 직접적(Direct)인 방법과 간접적(Indirect)인 방법으로 구분된다(Gescheider, 1985). 직접적인 방법은 관찰자(Observer)가 측정대상을 직접 평가하여 인지강도를 표현하는 방법이고, 간접적인 방법은 관찰자가 비교를 통하여 측정 대상을 간접적으로 평가함으로써 측정치를 얻어내는 방법이다. 간접적인 방법으로 측정치를 얻어내는 방법으로는 Successive DL Scale, Paired Comparison Scale, Ranking 및 Sorting에 의한 측정방법이 있고, Equisection Scale, Category Scale, Magnitude Scale 등은 직접적인 방법에 해당한다. Psychophysical Scale의 종류는 매우 다양하며, 그 중 현재까지도 많이 이용되고 있는 척도는 다음과 같다.

Successive DL Scales

Fechner의 법칙에 근거하여 개발된 척도로서, 실험을 통하여 Absolute Threshold(RL, Reiz Limen)와 Difference Threshold(DL, Differentz Limen)를 구한 후, RL로 부터 출발하여 Just Noticeable Difference(JND), 즉, 지각강도의 단위 증가분을 파악하여 이를 지각강도의 측정치로 이용한다. DL Scale은 자극의 전범위에 걸쳐 JND가 일정하다는 것과 DL의 크기가 자극의 세기에 항상 비

례한다는 Weber(1834)의 법칙이 옳다는 가정을 근거로한다. 그러나 일반적으로 RL 근처의 자극강도에 대해서는 Weber의 법칙이 성립하지 않음이 밝혀졌고, 이를 보강하기 위해 실험을 통하여 DL을 측정하여 이용할 경우 필요한 데이터를 얻기 위해 소요되는 실험횟수가 너무 많다는 단점이 있다.

Paired Comparison Scales

지각강도나 사회적 현상의 측정에 많이 이용되어 온 기법의 하나로, 자극을 비교의 대상이 되는 기준자극(Standard Stimulus)과 함께 제시하여, 피실험자가 어느 쪽의 자극이 더 크게 느껴지는지를 응답하게 한다. 서로 다른 두 자극 A와 B에 대한 전체 실험횟수 중 자극 A가 기준이 되는 자극 B보다 크다고 대답한 횟수의 비(Proportion of Frequency)를 계산하여 이를 이용하여 A와 B의 차이의 분포를 유도함으로써 측정치를 결정하게 되므로(Gescheider, 1985), 결과적으로 얻어지는 척도는 간격척도로 볼 수 있다. Paired Comparison Scale은 Thurstone(1927)의 Comparative Judgment의 법칙을 근거로 한다. Comparative Judgment의 법칙은 각 자극의 세기가 주관적 강도(Subjective Magnitude)의 분포로 표현가능하다는 점과 두 자극의 차이는 정규분포를 따른다는 점을 기본 가정으로 한다(Meister, 1985). Paired Comparison에 의해 n개의 자극을 평가하고자할 때에는 2^n 회를 반복해야 하므로 n이 클 경우, 실험평가가 불가능한 규모가 되는 단점이 있어, Torgerson(1958) 등에 의해 실험횟수를 감소시키는 방법에 대한 연구가 행해졌다.

Ranking

피실험자가 주어진 자극을 평가하여 지각강도의 크기에 따른 순서를 결정하는 방법으로, 일련의 자극에 대한 개개인의 Ranking 결과로부터 집단 전체의 Ranking 결과(Group Ranking)를 얻기 위해 사용한다(Meister, 1985). Ranking에 의한 평가결과로 얻어지는 척도는 서열척도이다.

Sorting

피실험자에게 모든 자극을 동시에 제시하여 Ranking과 Paired Comparison을 통해 Item들을 순서가 있는 일련의 항목으로 분류하는 방법이다(Meister, 1985). Sorting은 Ranking과 달리 모든 자극에 대해 순위를 결정하는 것이 아니라, 순서대로 제시된 일련의 항목에 맞추어 주어진 자극을 분류하는 것이다. Sorting에 의한 Scaling은 모든 자극을 동시에 고려해야 하므로 평가해야 할 자극의 수가 많을 경우 측정이 어려워진다. Sorting을 통해 얻어지는 척도는 Ranking이나 후에 언급할 Category Scaling의 경우와 마찬가지로 서열척도가 된다.

Equisection Scales

Equisection Scale에서는 우선 척도의 양 끝점이 되는 두 자극을 함께 제시하여 피실험자가 직접 자극의 세기를 조절하여 지각강도의 전 구간을 등간격으로 구분하는 지점을 표현하게 하여 측정치를 얻는다(Gescheider, 1985). 기본적으로 간격척도를 보장할 수 있으나, 적용범위가 제한되는 단점이 있으며, 피실험자에 의한 자극의 연속적 조절이 가능한 경우에만 적용가능하다.

Category Scales

Category Scaling은 강도를 표현하는 일련의 형용사나 어구들을 순서대로 제시하여 피실험자가 느낀 지각강도를 가장 유사하게 표현한 항목을 선택하도록 하는 방법을 쓴다. Sorting의 경우는 피실험자에게 모든 자극을 동시에 제시하는데 반해, Category Scaling에서는 한번 시행에 하나의 자극만을 제시하게 된다. Category Scale의 종류는 그림2와 같이 크게 두가지 유형으로 구분할 수 있다. 그림 2의 (a)와 같은 유형의 척도에서는 강도를 표현하는 일련의 형용사나 어구들을 순서대로 제시하여 피실험자가 느낀 지각강도를 가장 유사하게 표현하는 항목을 선택하도록 하는 방법을 쓴다. 이러한 유형의 척도에서는 항목들 간의 등간격성을 보장할 수 없으므로 간격척도로 보기 어렵다. 그림 2의 (b)와 같은 유형의 척도에서는 지각강도의 최고치와 최하치를 나타내는 형용사나 어구만을 제시하여 이들을 양 끝점으로 하는 구간을 세부항목으로 구분하여 관찰자의 지각강도와 유사한 점을 선택하게 하는 방법이다. 이러한 유형의 Category Scale은 간격척도로 보는 견해가 있으나, 실제로 척도의 연속성이 없으므로 역시 서열척도로 보는 것이 타당하다.

Category Scale은 전술한 다른 방법과 비교하여 측정이 용이하고, 평가대상의 갯수 만큼만 실험을 실시하면 측정치를 얻어낼 수가 있으므로, 자극의 평가와 분석이 모두 간단하여 여러 분야에서 가장 널리 적용되고 있으나, 다음과 같은 단점을 갖는다(Lodge, 1981). 첫째, 이미 언급한 바와 같이, 측정대상의 성질(Quality)을 표현하는 몇개의 제한된 항목 중에서 피실험자의 실제 지각강도와 가장 유사한 하나의 항목을 선택하게 되므로, 피실험자가 지각강도에서 약간의 차이를 느낀 경우에도 이를 표현하지 못하는 경우가 있어 정보의 손실 가능성이 있다. 둘째, 제한된 수의 항목을 제시함으로써 피실험자의 응답에 의도하지 않게 영향을 주게 되어 평가결과에 편향(Bias)을 초래할 수가 있다. 셋째, 각 항목을 표현하는 어구의 선정에 세심한 배려를 하지 않는 한 서열척도 수준의 측정만이 가능하다(Lodge, 1981; Osborne, 1976; Moskowitz, 1984). 따라서 전절에서 언급한 대로 간격척도 이상의 데이터를 근거로 하는 여러 강력한 통계기법들을 적용할 수가 없게 된다.

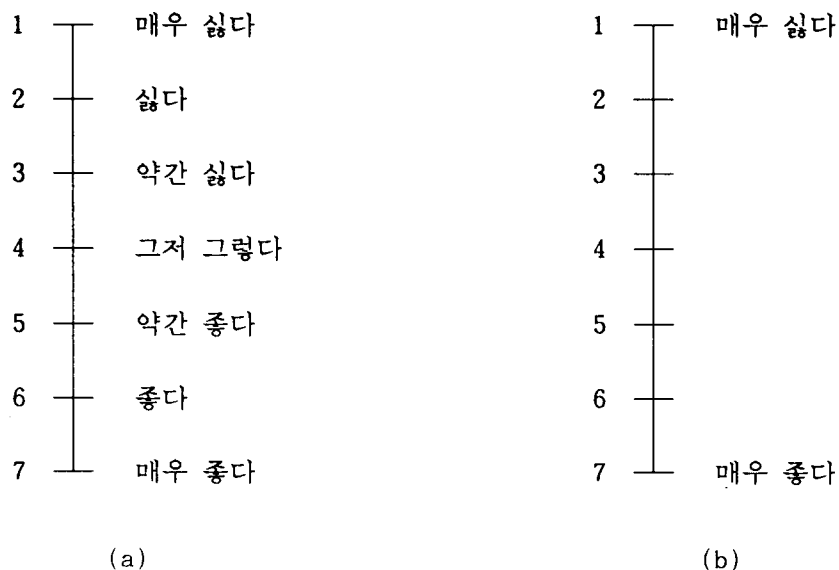


그림 2. Category Scale의 종류

Ratio Scales

비율판단(Ratio Judgment)에 의해 측정치를 결정하는 Ratio Scaling 기법 중 가장 많이 이용되는 방법으로는 Magnitude Estimation이 있다. 이 기법은 관찰자가 주어진 자극을 평가하여 기준자극의 지각강도에 대한 비율로 판단한 후, 기준자극에 배정된 값에 이 비율을 곱한 값을 응답하게 하여 이를 측정치로 이용하는 방법이다(Gescheider, 1985; 1988; Lodge, 1981). Magnitude Estimation은 실험을 수행하는 방법에 따라 피실험자에게 기준자극에 해당하는 값을 제시해 주는 Modulus Method와 이를 피실험자 스스로가 결정하게 하는 Free-Modulus Method로 구분되는데, Modulus Method의 경우 피실험자의 응답에 편중된 영향을 초래할 가능성이 커서(Gescheider, 1985) 최근의 여러 연구에서는 Free-Modulus Method가 주로 사용되고 있다.

Magnitude Estimation에 의해 얻어지는 데이터는 비율척도의 데이터이므로 모든 통계량이 의미를 가지며, 모든 통계적 검증기법의 적용이 가능하다. 또한 Magnitude Estimation은 측정에 이용할 수 있는 응답형식(Response Modality)이 다양하고, 어떠한 상황에도 적용이 가능하며, Cross-Modality Matching을 통해 척도의 타당성을 검토할 수 있는 장점으로 인하여, 감각기관과 관련된 지각강도의 측정 뿐 아니라 여러가지 사회현상이나 의견의 정량적 측정 및 측정대상간의 연관관계 파악에도 널리 이용되고 있는 가장 강력한 측정기법이라 할 수 있다.

그 외의 Ratio Scaling 기법으로는 Ratio Production(Fractionation), Ratio Estimation, Magnitude Production 등이 있으며, 이들 기법을 적용한 측정 결과는 Magnitude Estimation의 경우와 마찬가지로 비율척도 데이터의 성질을 갖는다.

4. Psychophysical Scale 적용 및 분석의 오류사례

앞에서 언급한 바와 같이, Category Scale의 두가지 유형 중 그림 2의 (a)와 같이 지각강도를 표현하는 일련의 어구를 순서대로 나열하여 피실험자가 한 항목을 선택하게 하는 방법을 사용할 경우, 각 항목들 간의 등간격성이 보장되지 않으므로 각 항목에 배정된 수치는 차이의 개념이 없는 서열척도의 값이 된다. 따라서 이 척도를 이용하여 측정한 데이터를 간격척도의 데이터처럼 취급하는 것은 바람직한 적용방법이 아니다. 예를 들어, 7점 척도의 경우 그림 2의 (a)에서 보듯이 중립위치(Neutral Position)인 '그저 그렇다'와 '약간 좋다'의 차이가 '좋다'와 '매우 좋다'의 차이와 반드시 같은 것은 아니므로, 이때 각 항목을 표현하는 기호로서 사용된 수치는 순서만이 의미가 있을 뿐이며, 수치의 차이에 대하여 언급하는 것은 타당하지 않다. 따라서 이러한 척도를 이용하여 측정한 데이터를 간격척도로 취급하는 것은 바람직하지 않다.

또한 그림 2의 (b)와 같이 최고치와 최하치를 나타내는 표현을 제시하여 이들을 양끝점으로 하는 구간을 다시 구분하여 구축한 척도를 측정에 이용할 경우, 척도의 연속성이 없으므로 역시 간격척도로 보기 어렵다. 즉, 간격척도의 성질을 갖기 위해서는 각 항목들 사이의 값, 예를 들어 2와 3 사이의 2.5, 2.8 등의 측정치도 의미를 가져야 하는데, Category Scaling에서는 이러한 측정치를 허용하지 않으므로, 엄밀한 의미에서 서열척도로 보는 것이 타당하다.

따라서 Category Scale로 측정한 데이터에 대해서는 각 항목의 발생빈도(Frequency Counts)와 그에 따른 백분율 분포(Percentage Distribution)를 분석에 이용하는 것이 최선의 방법이라 할 수가 있

다. 대표값으로는 최빈값, 중앙값 등을 이용하는 것이 바람직하며, 그 외 Contingency Correlation, Rank-order Correlation 등이 의미가 있는 통계량이 된다. 또한 통계적 검증을 위해서는 χ^2 - Test 등의 비모수 검정법(Nonparametric Test)을 이용하는 것이 바람직하다.

측정대상을 표현하거나 그들 간의 연관관계를 파악하고 모델링(Modelling)하는 데에는 모수검정에 해당하는 각종 통계기법들이 매우 유용하며, 이를 지원하는 패키지(Package)의 공급도 풍부하다. 이와 같은 이유로 인하여 Category Scaling을 측정기법으로 사용한 이전의 여러 연구들 중에는 측정결과를 간격척도의 데이터로 취급한 예가 빈번히 발견된다(Lodge, 1981). 이러한 오류를 그 원인이나 유형별로 분류하고 구체적 실례를 들어 설명하면 다음과 같다.

통계량(Statistics) 사용의 오류

Category Scale의 데이터에 대한 통계량의 사용에 있어서 가장 많이 발견되는 오류는 피실험자가 응답한 항목에 대해 평균과 표준편차를 계산하여 이를 대표값으로 이용하는 경우이다. 최근의 예로서 Luria, Neri, Schlichting (1989)에 의하여 수행된 여러가지 색상의 VDT(Visual Display Terminal) 인광물질(Phosphor)에 있어서의 작업수행도 및 선호도에 관한 연구를 들 수 있다. 이 연구에서는 녹색, 황갈색, 황색, 적색, 청색 및 백색의 6 가지 색상의 인광물질로 만들어진 VDT Display에서 모양이 비슷한 문자들 속에서 Target Symbol(예. '1'과 'I')을 찾아내는 작업을 수행하게 하여 피실험자의 작업수행도(Performance)와 선호도(Preference)를 측정하였다. 각 인광물질에 대한 선호도의 평가에는 1(Poor)부터 10(Good)까지의 10점 Category Scale이 이용되었고, 평가결과를 파악하기 위해 평균(Mean)과 표준편차(Standard Deviation)를 계산하여 이용하였다. 그러나 전술한 대로 Category Scale을 이용한 평가결과 얻어지는 측정치는 서열척도의 데이터이므로 평균이나 표준편차는 대표값으로 사용하지 않는 것이 바람직하며, 최빈값(Mode)이나 중앙값(Median) 등이 통계량으로서 적절하다. 그 외에도 분석의 편의를 위해 Category Scale 데이터에 대해 평균과 표준편차를 구하여 통계량으로 이용한 사례는 흔히 발견된다.

통계기법 적용의 오류 : 분산분석의 적용

Category Scale로 측정한 데이터에 대하여 어떤 변수가 미치는 영향을 검증하기 위해 각종 통계기법을 적용함에 있어서 빈번히 발생하는 오류 중의 하나가 분산분석(Analysis of Variance, ANOVA)을 실시하는 경우이다. 이러한 예로서 Matthews(1987)의 연구를 들 수 있다. 이 연구에서는 CRT Display에서 Text와 배경색의 조합이 교정(Proof-reading)작업에 있어서 사용자의 수행도와 사용자가 느끼는 불편도(Discomfort)에 영향을 미치는지를 파악하기 위하여, 피실험자가 느끼는 불편도를 1(Low Discomfort)부터 7(High Discomfort)까지의 7점 Category Scale로 측정하였다. 8가지의 Text색상과 배경색의 조합에 관하여 피실험자가 평가한 불편도의 측정치가, 작업 수행 전과 후에 유의한 차이를 보이는지를 검증하기 위해 ANOVA를 실시하여 그 차이가 통계적으로 유의하다는 결론을 얻었다. 그러나, 전술한 대로, 이러한 유형의 Category Scale에 있어서 척도의 연속성이 없으므로 이를 간격척도로 보기는 어려우므로 모수검정법을 사용할 수 없다. 따라서 분산분석을 실시하여 각 인자의 효과를 검증하는 것은 잘못된 분석방법이며, 이러한 경우 Category Scale 대신 Magnitude Estimation을 이용한 평가를 통하여 비율척도의 데이터를 얻어 분산분석을 실시하는 것이 보다 적절한 접근방법이라 할 수

있다.

Category Scale 데이터에 대하여 분산분석을 실시한 또 다른 예로서 Rosenfeld, Tenenbaum, Ruskin, Halfon(1989)의 연구를 들 수 있다. 이 연구는 산업체 근로자로 구성된 피실험자군을 두 그룹으로 나누어 이들 두 그룹의 피실험자에게 본작업에 앞서 서로 다른 활동, 즉, 체계적인 운동 프로그램과 정신적인 사회활동의 두가지 작업을 실시하게 하여 피실험자가 느낀 업무의 효율성과 생산성의 차이를 파악하기 위해 실시되었다. 업무의 효율성과 생산성에 대한 주관적인 측정치를 얻기 위해 이 연구에서는 그림 3과 같은 6점 Category Scale이 사용되었고, 피실험자의 성별 및 연령과 수행한 활동의 종류에 따른 이들 측정치의 차이를 파악하기 위해 분산분석을 실시하였다. 그러나 이 연구에서 사용된 Category Scale의 경우, 앞에서 언급한 바와 같이 각 항목들 간의 등간격성이 보장되지 않으므로 측정치는 서열척도의 데이터로 보는 것이 타당하며, 앞의 예와 마찬가지로 각 변수의 영향을 파악하기 위해서는 Category Scaling 대신 Magnitude Estimation을 이용하여 측정하는 것이 보다 나은 접근방법이라 할 수 있다.

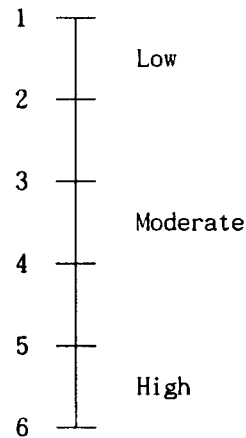


그림 3. Rosenfeld, et al.(1989)의 연구에서 사용된 6점 Category Scale

통계기법 적용의 오류 : 회귀분석의 적용

측정된 데이터를 이용하여 측정대상들 간의 연관관계를 예측하고 구체적인 수식을 제공하기 위해서는 회귀분석을 실시하게 된다. 그러나 회귀분석 또한 분산분석의 경우와 마찬가지로 간격척도 이상의 성질을 갖는 데이터에 대해서만 적용이 가능하므로, 서열척도인 Category Scaling 기법을 이용한 측정 데이터에 대해서는 적용할 수가 없는 것이 원칙이나 실제로 잘못 적용한 사례가 다수 존재한다. 그 사례로는 Beshir와 Ransey(1981)의 체감온도에 관한 연구를 들 수가 있다. 이 연구에서는 23.3, 32.3, 37.8, 43.4°C의 네가지 외부 온도조건에서 여러가지 작업을 수행하게 하여 각각의 경우 피실험자가 느낀 열감을 그림 4와 같은 9점 Category Scale을 이용하여 측정하였다. 평가 결과 얻어진 측정치에 대하여 회귀분석을 실시하여 실제 온도의 변화에 대한 주관적인 온도의 관계를 회귀식으로 표현하였다. 또한, 중립위치(Neutral Position)에 해당하는 온도를 피실험자가 가장 좋아하는 온도로 정

의하고, 이 회귀식을 이용하여 종속변수의 값을 4가 되게 하는 실제온도를 계산하였다. 그러나 이러한 모든 분석과정은 실제온도와 피실험자가 느낀 열감을 평가한 Category Scale의 데이터 간에 선형관계(Linear Relationship)가 성립한다는 가정을 바탕으로 한 것이나, 실제로 열감에 대한 인간의 지각 강도는 Power Function을 따르므로(Stevens and Marks, 1971; Stevens and Stevens, 1960) 잘못된 가정이라 할 수 있다. 따라서 이는 적합하지 못한 분석방법이라 할 수 있으며, 분석결과 얻어진 회귀식에도 큰 의미를 부여할 수 없다.

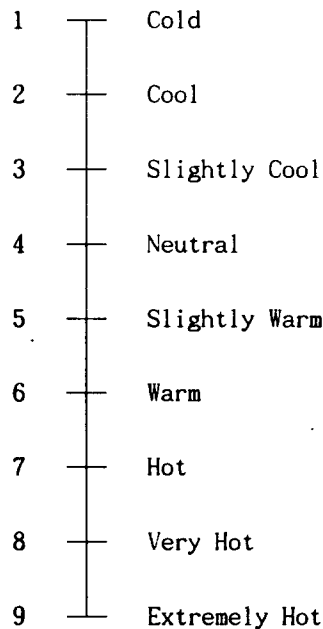


그림 4. 체감온도의 측정에 이용된 9점 Category Scale

5. 논의

본 논문에서는 Psychophysical Scale의 종류를 일반적인 자료의 분류체계와 연관지어 설명하였다. 그 중 측정이 용이하여 가장 많이 사용되어 온 Category Scale의 경우, 서열 척도 수준의 측정만이 가능하여 사용가능한 통계량이나 통계기법의 적용 등에서 많은 제약을 받게 됨에도 불구하고, 각 항목에 임의로 수치를 배정하고 이를 간격척도로 취급하여 분석을 실시한 사례를 조사하여 논하였다.

앞에서 언급한 오류들은 하위수준의 측정방법을 이용하여 얻어진 데이터에 대하여 상위수준의 측정치에 적용가능한 분석기법을 사용함으로써 발생한 것으로, 연구의 목적이 상위수준의 분석기법을 요구할 경우 측정 기법을 달리하여 간격척도 수준 이상의 데이터를 얻는 것이 바람직하다. 예를 들어 Magnitude Estimation을 사용하여 비율척도의 데이터를 얻었을 경우에는 거의 모든 통계량을 사용할 수가 있고, 모든 통계기법의 적용이 가능하게 된다. 그러나 모든 경우에 비율척도의 데이터가 필요한 것은 아니며, 실제로 어떤 측정기법을 사용하느냐는 측정의 목적에 따라 달라진다. 예를 들어, 제품 설계의 초기단계에서 시장조사 등을 통해 어떤 상품이 상대적으로 가장 선호되고 있는지를 파악하기

위한 측정에 있어서는 어떤 제품에 대한 절대적인 선호도를 파악하거나 그 관계를 수식화하려는 것은 아니므로 시간과 비용을 들여 비율척도의 데이터를 얻을 필요는 없다. 이런 경우에는 Category Scaling을 이용하여 발생빈도분포를 파악하여 분석에 이용하는 것이 보다 적절한 측정방법이라 할 수 있다.

반면, 본격적인 설계단계에서 구체적인 공학적 규격(Specification)을 결정함에 있어서 서로 다른 비용이 소요되는 여러 대안들에 대하여 구체적으로 몇배정도의 선호도차이가 있는지를 파악함으로써 비용 투자의 정당성을 분석(Cost Justification)하기 위하여, 혹은 측정대상들 간의 관계를 예측하고 모델링(Modelling)하기 위해서는 Magnitude Estimation을 이용하여 측정하는 것이 보다 적절한 접근방법이다.

따라서 사회적 현상이나 사용자의 의견 등의 심리적 요소의 측정에 있어서, 실험평가에 앞서 측정의 목적을 명백히 정의하여야 하고, 그에 따라 적절한 측정기법을 선택하여 실험평가를 실시한 후, 타당한 결론을 얻기 위해서는 허용되는 범위내의 분석을 실시하여야 하며, 결과의 해석에 있어서도 신중을 기하여야 한다. Psychophysical Scaling 기법의 적용에 있어서 요구되는 연구절차는 그림 5와 같이 요약될 수 있다.

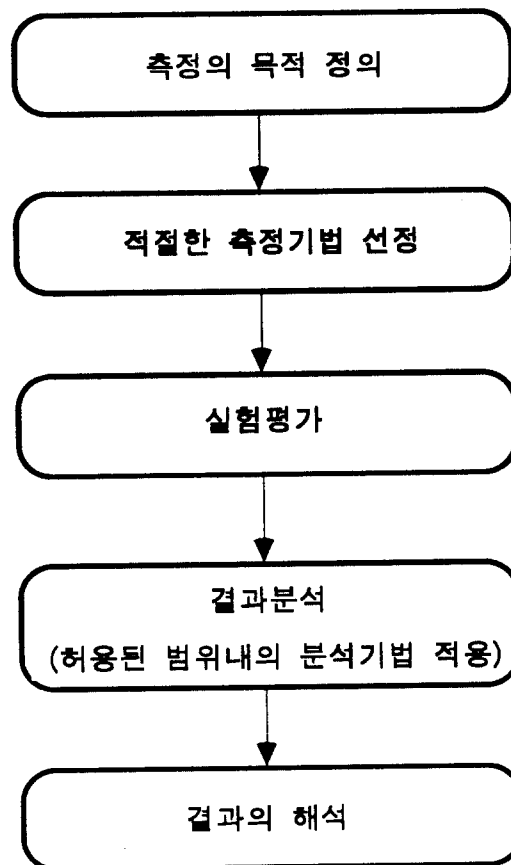


그림 5. Psychophysical Scaling 기법의 적용에 필요한 연구절차

6. 참고문헌

- 한성호, 정의승, 박성준, 곽지영, 최필성 (1993). "Psychophysics를 이용한 고속전철 객실 내장설계의 승객 선호도 평가", 대한인간공학회 춘계학술대회 논문집, 134-144.
- Beshir, M. Y., and Ramsey, J. D. (1981). "Comparison between male and female subjective estimates of thermal effects and sensations", *Applied Ergonomics*, Vol. 12, No. 1, 29-33.
- Ekman, G. and Künnapas, T. (1962). "Measurement of aesthetic value of 'direct' and 'indirect' methods", *Scandinavian Journal of Psychology*, Vol. 3, 33-39.
- Gescheider, G. A. (1985). *Psychophysics : Method, Theory, and Practice*, 2nd Ed., Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gescheider, G. A. (1988). "Psychophysical Scaling", *Annual Review of Psychology*, Vol. 39, 169-200.
- Gescheider, G. A., Catlin, E. C., and Fontana, A. M. (1982). "Psychophysical measurement of the judged seriousness of crimes and severity of punishments", *Bulletin of the Psychonomic Society*, Vol. 19, No. 5, 275-278.
- Kemp, S. (1991). "Magnitude estimation of the utility of public goods", *Journal of Applied Psychology*, Vol. 76, No. 4, 533-540.
- Künnapas, T. and Wikström, I. (1963). "Measurement of occupational preferences : A comparison of scaling methods", *Perceptual and Motor Skills*, Vol. 17, 611-624.
- Lodge, M. (1981). *Magnitude Scaling : Quantitative Measurement of Opinions*.
- Luria, S. M., Neri, D. F., and Schlichting, C. (1989). "Performance and preference with various VDT phosphors", *Applied Ergonomics*, Vol. 20, No. 1, 33-38.
- Matthews, M. L. (1987). "The influence of colour on CRT reading performance and subjective comfort under operational conditions", *Applied Ergonomics*, Vol. 18, No. 4, 323-328.
- Meister, D. (1985). "Subjective and objective methods", *Behavioral Analysis and Measurement Methods*.
- Moskowitz, H. R. (1984). *Cosmetic Product Testing : A Modern Psychophysical Approach*, New York, NY: Marcel Dekker, Inc.
- Osborne, D. J. (1976). "Examples of the use of rating scales in ergonomics research"
- Rosenfeld, O., Tenenbaum, G., Ruskin, H., and Halfon, S. (1989). "The effect of physical training on objective and subjective measures of productivity and efficiency in industry", *Ergonomics*, Vol. 32, No. 8, 1019-1028.
- Stevens, J. C. and Marks, L. E. (1971). "Spatial summation and the dynamics of warmth sensation", *Perception and Psychophysics*, Vol. 9, 291-298.
- Stevens, J. C. and Stevens, S. S. (1960). "Warmth and cold : Dynamics of sensory intensity", *Journal of Experimental Psychology*, Vol. 60, 183-192.
- Stevens, S. S. (1946). "On the theory of scales of measurement", *Science*, Vol. 103, 667-680.
- Torgerson, W. S. (1958). *Theory and Methods of Scaling*, New York, NY : Wiley.
- Thurstone, L. L. (1927). "A law of comparative judgment", *Psychological Review*, Vol. 34, 273-286.
- Weber, E. H. (1834). *De pulsu, Resorptione, Auditu et Tactu : Annotationes Anatomicae et Physiologicae*, Leipzig : Koehler.