Window of Statistics

통계의 창

2017
WINTER
Vol.20



이슈

비정형 데이터를 주목하라

비정형 데이터의 가치를 캐다
비정형 데이터 의미
텍스트마이닝 : 의미 속의 의미 찾아내기

통계광장

한국의 4차 산업혁명, 어디까지 왔는가
보건의료계에 불어오는 빅데이터혁명
사회적 양극화 해소의 실마리가 되는 고용통계

PEOPLE

사회적 건강이 무엇인가
염유식 / 연세대 사회학과 교수

통계탐방

인구로 보는 대한민국 서비스

목차

통계의 창 2017 Winter Vol.20



ISSUE

002 비정형 데이터의 가치를 캐다
– 김장현 / 성균관대 인터랙션사이언스학과 교수

008 비정형데이터가 말하는 시그널에 주목하라
– 최재원 / 다음소프트 이사

012 텍스트는 어떻게 분석되어 지는가
– 김도훈 / 아르스 프락시아 대표



통계광장

016 한국의 4차 산업혁명, 어디까지 왔는가
– 김광석 / 삼정 KPMG 경제연구원 수석연구원

020 보건·의료계에서 불어오는 빅데이터 혁명
– 이강봉 / 통계의창 객원기자

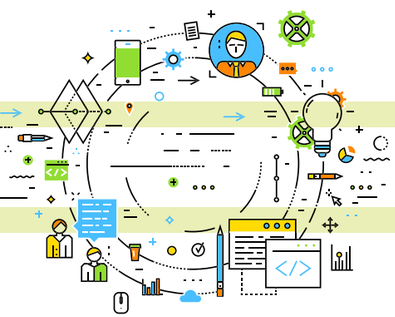


026 사회적 양극화 해소의 실마리가 되는 고용통계
– 빈현준 / 통계청 고용통계과장

030 데이터 스토리텔링… 데이터에 스토리를 입혀라
– 강양석 / 셀바스그룹 COO

036 AI 서비스가 시작되는 AI 플랫폼에 주목하라
– 유태준 / 마인즈랩 대표

042 경제를 이해하려면, 통계에 사람을 먼저 포함하라
– 김덕식 / 통계의창 객원기자



교육

048 R에 도전하자… 따라가다보면, 나도 R유저②
– 심송용 / 한림대 금융정보통계학과 교수

056 인포그래픽 나도 할 수 있다… 효과적으로 정보를 전달하는 방법
– 송정수 / 인포그래픽웍스 대표

060 국내 유일 통계교육기관… 통계교육원
– 이춘우 / 통계교육원 사무관

PEOPLE

064 건강한 삶, 행복한 삶은 당신 가까이에 있다
– 염유식 / 연세대 사회학과 교수



통계탐방

070 KOSIS '인구로 보는 대한민국 서비스'
– 이주원 / 통계청 통계서비스기획과 사무관



BOOK REVIEW

076 책을 베고 잠들면, 꿈은 당신이 안 가본 길로 안내한다
– 김영진 / 한국통계진흥원 부장

창가의 여유

080 미리가보는 평창 올림픽
– 김준래 / 통계의창 객원기자



084 흰 눈 위의 자유, 스노보드를 만나다
– 최용준 / 인간개발연구원 교육팀장

092 추운 겨울 아이들과 도란도란… 요리가 만들어지는 그림책
– 김현정 / 타샤의 책방

096 통계로 바라보는 세상이야기…
– 신동헌 / 통계의창 객원기자

발행일 | 2017년 12월 20일
발행인 | 임병권
발행처 | 통계교육원
기 획 | 심원보, 이승구

주 소 | 대전광역시 서구 한발대로 713(월평동) 통계센터 통계교육원
전 화 | 042-366-6151
팩 스 | 042-366-6498
이메일 | ljh0217@korea.kr

디자인 | 동아에스앤씨(02-3148-0745)
인 쇄 | 일진인쇄 (02-499-4500)
ISSN 2005-1379-00

©2017. 통계교육원

※ 「통계의창」에 실린 내용은 필자 개인의 의견이므로 필자의 소속기관이나 본지의 공식적인 견해를 대변하는 것은 아닙니다.

김장현 ◆ 성균관대 인터랙션사이언스학과 교수

1 비정형 데이터의 가치를 캐다



비정형 데이터란 무엇인가

비정형 데이터라는 말에서 '비정형'은 'unstructured'라는 영어 표현의 번역이다. 사실 여전히 지배적으로 사용되는 데이터의 형식은 소위 구조화된 데이터 또는 정형화된 데이터라고 불리우는, 변수와 응답자(케이스)가 각각 열과 행을 이루는 스프레드시트 형태이다. 이러한 형태 안에는 주로 수치로 표기된 데이터가 존재하고, 그러한 데이터가 정보의 인출, 탐색, 분석, 의사결정에 이르는 제반 과정을 거치며 회귀분석, 분산분석, 컨조인트 분석 등 다양한 통계적 기법에 의해 다루어지기 마련이다.

그런데 텍스트로 이뤄진 데이터의 경우 사전에 정의된 방식 또는 구조로 배열되지 않는 경우가 많다. 텍스트 안에서는 단어와 숫자, 기호 들이 전혀 예측할 수 없어 보이는 방식으로 배열되며 이런 무작위성 때문에 비구조화 또는 비정형 데이터라고 불린다. 물론 언어에 문법과 어법이 있지만, 모든 이가 동일한 상황을 동일하게 묘사하는 것도 아니고, 의견을 표시하는 방식도 제각각이기 때문에 말이든 글이든 비정형 데이터라고 불리우게 된 것이다.

하지만, 비정형 데이터라고 해서 반드시 예측 불가한 방식으로 구성되는 것은 아니다. 한 단어가 등장했을 때 다음 단어가 무엇이 될지는 확률적으로 계산이 가능하며, 이러한 계산을 통해 자동으로 다음 단어를 제안해주는 방식은 검색엔진 등에서 이미 널리 활용되고 있다. 이런 확률론적 관점에서는 사실 비정형처럼 보이는 데이터에도 숨은 정형(구조)이 있기 때문에 엄격한 의미에서 비정형 데이터는 거의 없다고 주장하기도 한다.

음악, 소리, 사진이나 동영상 역시 종종 비정형 데이터로 간주되는데, 사진에 포함된 색상이나 채도의 분포로 그 사진을 찍은 사람의 성격을 알아내고자 하는 시도나 그림에 등장하는 색상과 명도 등을 분석하여 그 그림이 어느 시대의 화풍을 반영하는지 자동으로 알아내려는 연구가 등장하는걸 보면 이런 데이터들 역시 그 속에 일정한 패턴과 구조가 내재되어 있음을 알 수 있다.

텍스트에 관한 연구는 어떻게 발전해가고 있는가

그렇다면 비정형 데이터들은 그동안 어떻게 연구되어 왔을까? 먼저 가장 오랜 역사를 가진 텍스트에 관한 연구전통부터 살펴보자. 학계에서는 전통적으로 '내용분석'이나 '담론분석'이 사용되었다. 전자의 경우 보통 세 명이상의 코더(코딩하는 사람)가 주어진 텍스트를 읽고 그 텍스트의 주된 특질이 사실 또는 사건의 건조한 전달인지 아니면 주장이나 관점인지 분석하는 경우가 일반적이다. 그리고 이러한 분석의 결과에 따라 저자가 어떤 특이한 틀을 갖고 글에서 다루는 대상을 바라보는지 도출해 내기도 하며, 이런 종류의 내용분석을 '프레이밍 분석' 또는 '프레임 분석'이라 지칭하기도 한다.

그런데 프레이밍 분석이 가질 수도 있는 단점은 연구에 참여한 복수의 코더들이 사실은 해당분야 교수 또는 박사후연구원과 그 연구실의 구성원, 또는 그러한 사람들에게 강의를 들은 학생인 경우가 적지 않다는 점이다. 특정 교수 밑에서 학문적 관점과 세계관을 습득한 연구원들이 거의 동질화된 관점에서 텍스트를 코딩하니, 아무리 코더간 일치도를 계산해봤자 높게

비정형데이터_ 다이퀘스트



나올 수 밖에 없다. 마치 강의를 열심히 들은 학생일 수록 교수의 가치관을 빨리 파악하기 마련이고, 그 학생이 좋은 점수를 받기 위해 자기 자신보다는 교수의 관점에 기반한 보고서를 내는 경우와 마찬가지이다.

주로 '양적 분석'으로 분류되는 내용분석과 달리 단어와 문장들이 주제의식을 구성하는 글이라는 그릇의 구성물이라는 전제에서 글에 숨어있는 이데올로기와 가치관을 해석해내고자 하는 '질적 분석'이 바로 담론분석이다. 담론분석의 관심사는 수치화된 텍스트의 특질이라기보다는 오히려 텍스트안에 감춰진 권력, 서열, 지배욕, 그리고 투쟁이다. 때문에 주로 비판이론가들이나 인류학자들이 많이 사용해왔다. 이러한 질적 분석은 단순히 단어의 등장 빈도나 등장 확률에 관심이 있기 보다는 텍스트에서 드러나는 주체와 객체, 그리고 권력자와 피권력자의 의도와 통제를 읽는다. 어쩌면 인간 사회에 대한 이해가 부족한 컴퓨터가 아직 따라잡기 어려운 영역이 바로 질적 분석이 아닐까 생각한다.

그런데 내용 분석에서도 컴퓨터를 활용해 차츰 깊이를 추구하는 경향이 등장하고 있다. 바로 '감성분석(sentiment analysis)'이라고 불리우는, 텍스트 전반에 드러나는 감정의 색조를 분석하는 경우다. 영문 데이터의 경우, 각 감정별로 해당하는 단어들의 리스트가 사전(dictionary) 형태로 공개되어 있어 텍스트 전반 또는 일부에서 드러나는 감정을 잡아내는게 점점 더 용이해지고 있다. 한국어의 경우 조사와 어미가 발달한 교착어인 데다가, 한자어에서 유래한 말과 고유한 우리 말이 혼재되어 있어서 이러한 감성사전의 단계에 이르는 리스트는 찾아보기가 쉽지 않다.

텍스트 데이터를 분석하는 방법들

비정형 데이터에 대한 가장 쉬운 접근은 텍스트를 하나의 단어주머니(bag of words)로 간주해서 단어순서나 언어적 구조를 무시한 채, 텍스트 안에 한가지 원소(단어)가 여러 번 등장할 수 있는 일종의 수학적 집합으로 보는 접근이다. 문서에 특정 단어가 1, 없으면 0으로 표기하는 방식이 많이 사용된다.

한 발짝 더 나아간 접근은 단어빈도(Term Frequency)에 주목하는 방식이다. 일반적으로는 텍스트안에 단어가 여러 번 등장할 수록 중요하다고 간주되는데, 그 이유는 아무리 긴 텍스트라도 소수의 자주 등장하는 단어들이 결국 전체 등장횟수의 상당부분을 점유한다는 사실이 알려져있기 때문이다. 다만, 영어의 'a' 나 'the' 처럼 횟수로는 가장 많은 부분을 차지하지만 의미가 별로 없는 경우는 분석에서 배제한다. 그런데 개별 단어보다 문장이 중요한 경우에는 단어들을 1개씩부터 n개씩까지 차츰 묶어보면서 단어쌍들간의 관계에 주목하는 방식도 있다. 이것을 N-그램 순열이라고 한다. 아무래도 개별 단어로 쪼개서 분석하는 것보다 그 연결패턴에 주목하기 때문에 더 깊은 관찰이 가능하다.

지난 1970년대부터 소개되어 최근에도 널리 사용되는 시멘틱 네트워크(의미망) 분석은 이러한 단어쌍들간의 거리와 동시출현 빈도를 바탕으로 텍스트를 단어기반의 네트워크로 표현하는 분석이다. 신문기사나 인터넷 댓글을 분석했을 때 특정 단어를 중심으로 지배적인 세력

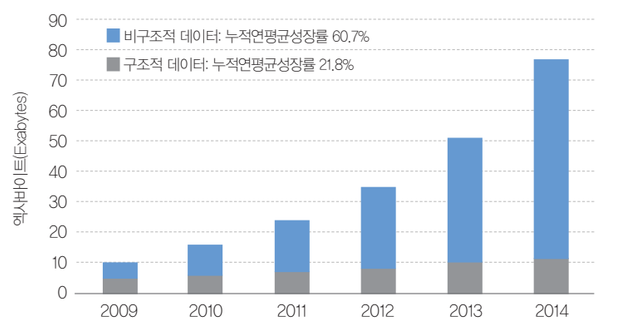
이 의미망 상에 등장하면 우리는 그것을 그 단어가 들어간 텍스트의 핵심주제로 받아들인다.

주어진 텍스트 데이터에 몇 개의 주제가 잠복해 있으며, 각 주제는 몇 개의 단어만 찾아내면 감지해낼 수 있다는 전제에서 출발하는 LDA(잠재디리크리할당)라는 주제모델(Topic Modeling)기법도 있다. 단어들의 연계성에 그치지 않고 그 연계성에 숨어있는 주제들까지 뽑아내는 단계에 이른 것이다. 이러한 주제모델은 텍스트의 시간적 변화까지 추적할 수 있는 동적 주제모델(Dynamic Topic Modeling)로까지 발전해 있다. 또한 최근에는 다양한 인공지능 알고리즘을 통해 텍스트를 분석하는 기법도 많이 사용된다.

비정형 데이터의 가치는 무엇인가

그렇다면 비정형데이터에 주목하는 이유는 무엇인지로 다시 돌아가 보자. 지금까지 수치화, 정형화된 데이터의 분석기법은 많이 발전해 왔으나 말로 이뤄진 데이터는 사람이 직접 읽고 판단하는 방식 외에는 뾰족한 수가 없었다. 매일같이 늘어나는 데이터의 상당 부분은 텍스트 또는 대화이고, 첨단 기술을 이용한 미디어들은 그러한 텍스트를 기반으로 서비스를 해야하는 상황에 이르렀다. 최근 유행하는 AI 스피커(인공지능 스피커) 역시 이용자와 스피커에 연결된 서버간 대화를 통해 이용자의 의도를 찾아내고 거기에 맞춰서 서비스를 제공해야 하는 것이다. 또한 항공권을 예약하는 사이트들이 이용자와 자동으로 채팅을 통해 발권까지 해주는 기능을 채용하고 있으며 이때 사용자와 사이트간에 오가는 데이터역시 비정규 데이터들이다. 또한 노인이나 환자들을 대상으로 정서적 안정감과 책읽어주기 등 다양한 서비스를 제공하는 휴머노이드 로봇들 역시, 인간과의 대화를 통해 소비자 만족을 달성한다. 여기에 또 비정규 데이터가 등장한다.

비정형데이터의 급증과 비중 증대



출처: EMC(2011)



인공지능과 비정형 데이터의 만남이 우리의 삶을 바꿀 것이다

인간과 기술사이의 인터페이스는 결국 0(OFF)과 1(ON)로 표현되는 디지털 숫자면 충분했다. 하지만 이제는 인간이 마치 다른 인간과 소통하듯 스마트미디어를 대하고 있으며 이런 과정에서 쏟아지는 엄청난 비정규데이터들이야 말로 '21세기의 석유'라고 불리우는 빅데이터의 주축을 이루게 되었다. 아직은 초보단계인 비정규데이터 분석의 발달은 곧 인공지능에 연결되며, 인공지능은 인간의 삶을 바꿔놓을 또 하나의 산업혁명이 될 공산이 크다. 인공지능이 비정규데이터에 익숙해져 가면 갈수록, 우리 스스로도 모르는 인간의 욕구와 비밀이 차츰 고개를 내밀게 될 것이다. 그러한 변화에 우리가, 그리고 우리의 아이들이 어떻게 적응해 나아갈지 기대도 되고, 고민도 되는 시점이다.

최재원 ◆ 다음소프트 이사

2 비정형데이터가 말하는 시그널에 주목하라



빅데이터가 이제 막 관심을 끌던 시기 정부가 주관한 빅데이터 경진대회 심사위원을 맡은 적이 있다. 초창기 심사하면서 관심 있게 보았던 부분은 참가한 학생들의 전공이었다. 전산학, 통계학, 문헌정보학, 신문방송학 등 전공들이 다양했다. 심사 전에는 비정형 데이터 분석을 위한 텍스트마이닝 기술을 사용할 줄 아는 전산학이나 추출된 데이터를 기반으로 상관성을 찾아내는 통계학 전공자들이 유리할 것이라는 생각을 갖고 심사했다. 하지만 나의 예상을 깨고 대상은 문헌정보학과 학생들이 차지했다. 초기 빅데이터는 기존에 데이터로 인식되지 못했던 것들이 데이터화 되면서 빅데이터라는 개념이 나오듯, 굳이 비유하자면 사람들 사이에서 말로만 전해지던 것들이 글로 표현되어 그 책이 도서관에 보관되면서 빅데이터라는 거대한 멀티미디어 도서관이 만들어진 것이다. 초창기 문헌정보학과가 빅데이터와 연관성을 갖는 이유에 대한 답은 분류에 있었다. 그 후에 치러진 경진대회에서 한 동안 수상자의 전공은 통계학과 출신으로 바뀌게 되었다. 정형데이터를 분석에 활용하면서 다양한 통계 기법을 통한 예측이 주요하게 작용됐지 때문이다. 하지만 최근 수상자는 통계학과 출신에서 사회학과나 심리학과 같은 인문학 출신자들이 수상하는 사례가 점점 많아지고 있다.

데이터는 시그널이다

최근 데이터 분석의 가치가 높아지면서, 빅 데이터를 포함한 데이터 과학(Data Science)에 대한 관심이 크다. 데이터 과학은 전문가를 위한 영역이라고 생각하지만, 원래는 가설을 세우고 검정하여 의미 있는 현상을 뽑아내는 과정 전체를 아우르는 말이다. 물론 이전부터 데이터가 쓰이지 않았던 적은 없다. 그러나 지금 부르는 데이터 과학이 단순히 엑셀 파일에 있는 숫자들을 그래프로 표현하는 걸 의미하진 않는다. 데이터를 단순히 분류하거나 분석하는 것 말고 데이터 속에 담긴 패턴이나 미래 예측에 도움이 되는 신호를 찾는 것 등을 아우르는 것이 데이터 과학인 것이다. 전문가 뿐만 아니라 비전문가도 데이터 과학으로 의미 있는 결과와 예측을 하는 것이 불가능한 일은 아니라는 것이다.

이 가운데 스몰 데이터는 개인의 취향이나 필요, 건강 상태, 생활 양식 등 사소한 행동에서 나오는 정보들을 말한다. 스몰데이터는 비정형 데이터가 많다. 개인에 대한 관찰을 통해 정확한 추리를 해내는 명탐정 셜록 홈즈는 뛰어난 스몰데이터 분석가라고 할 수 있다.

야구에서 왜 희생번트를 고집할까



희생번트_뉴스줌

데이터를 많이 활용하는 스포츠 종목인 야구의 경우, 데이터만 놓고 보면 희생번트 작전은 득점 보다 점수를 희생하는 작전이다. 이닝당 점수를 비교했을 때 무사 1루에서 희생번트를 했을 경우에는 0.75점, 하지 않았을 경우에는 0.93점이 나왔다는 것이다. 그러자 이에 많은 감독들이 반발했다. 박빙 승부에서 1점은 낸다는 의미를 통계학자는 알 수 없다는 것이다.

그런데 박빙 승부에서 1점을 낼 확률은 희생 번트를 했을 때 0.417, 하지 않았을 때는 0.4로 매우 근소했다는 것이다. 솔직히 통계적으로는 득점 확률이 낮지만 심리적으로 감독에게 부여된 작전기회에 희생번트 작전을 쓰지 않아 병살이 됐을 때 받을 비난을 의식하지 않을 수 없었을 것이다. 희생번트는 전적으로 데이터로 설명할 수 없는 작전 중 하나라고 볼 수 있다. 양 팀이 똑같이 27개의 아웃카운트를 갖고 시작하는 야구에서 1아웃의 가치와 1베이스 진루의 가치를 상황에 따라 달라질 수 있다. 통계적인 의미로 판단하기 보다는 주어진 데이터가 처해진 상황에서 어떻게 활용하고 해석하느냐에 따라 경기 결과를 좌우할 수 있다.

홍대와 가로수 길의 차이를 데이터는 알고 있다

빅 데이터는 이미 다양한 곳에서 그 위력을 과시하고 있다. 빠르게 변하는 산업 환경에 적응하기 위해서는 데이터 분석에 근거한 의사 결정이 필수적이라는 말이다. 빅데이터는 주로 인과성(causality)보다 상관성(correlation)을 보여준다. 때문에 그 동안은 어떻게 데이터를 저장하고, 어떻게 접근할 것인가에 초점을 맞추면서 단순히 언급량에 의한 데이터의 상관성을 중심으로 분석하는 경우가 대부분이었다. 하지만 최근에는 인과성에 의한 왜(why)를 찾고자 하는 목적으로 데이터를 이해하는 경우가 많아지고 있다. 예를 들어 보면 홍대와 가로수 길에 있는 똑같은 여성 의류 전문 매장에서 할인행사를 했다. 똑같은 제품에, 똑같은 할인폭이었지만 결과는 전혀 달랐다. 홍대 매장은 매출이 급등한 반면, 가로수 길 매장은 평소와 큰 차이가 없었다. 왜 그럴까? 답을 찾기 위해 두 매장의 신용카드 구매 기록 데이터를 분석해 보면 흥미로운 결과가 나와 있다. 홍대 매장에서는 대부분 여성의 카드가 사용되었지만, 가로수 길 매장에서는 남성의 카드로 결제하는 경우가 많았던 것이다. 즉 홍대에서는 자신의 돈으로 옷을 사 입는 여성고객이 대부분이라 할인율이 구매에 큰 영향을 줄 수 있었지만, 가로수길 매장에서는 남성의 카드, 즉 남자친구가 사주는 경우가 많아 할인율이 구매에 큰 영향을 주지 않을 수도 있다는 해석을 할 수 있다. SNS에 올라오는 비정형데이터를 분석해 보면 가로수길은 오빠가 사준 신발 이나 남자 친구가 사준 옷 등 선물에 대한 글과 사진이 많이 올라오고 있는 지역이었다. 남자들은 데이트할 때 피해야 될 지역 중에 하나가 바로 가로수길이다.

또한 실제로 실리콘밸리 기업들은 빅데이터를 신용평가에 활용하는 다양한 알고리즘을 실험하고 있다. 미국 하버드 대학의 아심 크와자 교수는 '맞춤법을 틀리지 않는 대출자는 틀리는 대출자에 비해 평균 15% 정도 덜 연체한다'는 연구결과를 바탕으로 신용평가 모델을 개발하기도 했다. 또 미국 핀테크 스타트업 렌도는 대출 희망자의 동의를 받은 뒤 SNS 계정에서 친구 목록, 결혼·연애 상태, 생년월일, 경력·학력, 출신지, 관심사 등 정보를 바탕으로 0~1000점까지 고객 신용 점수를 도출한다. 온라인 평판과 교우 관계가 대출 신용평가 점수에 반영이 되는 것이다. 최근에는 웨어러블 기술 발달로 운동습관, 심장박동수 등 건강상태에 대한 정보를 수집, 활용하는 기업들이 생겨나면서 이들의 정보를 대출 실행 정보로 활용하는 방법도 논의되고 있다.


가로수길_aljuarez

시그널을 파악하는 촉각은 분석 기술보다 해석이다

이처럼 빅데이터는 데이터를 추출하고 그 데이터에서 패턴을 찾아내는 분석의 과정과 그 패턴에서 의미를 찾아서 가치를 덧붙이는 해석의 과정으로 이뤄져 있다. 여기에서 분석 과정은 기술이나 통계 기법으로 해결되지만 해석 과정은 기술로 해결되지 않는다. 해석 과정에서 인간이나 사회에 대한 이해가 필요한 경우가 많기 때문이다. 어떻게 보면 빅데이터는 분석 기술 보다 해석 과정이 더 중요하다고도 볼 수 있다. 빅데이터에 대해 많은 분들이 오해하는 부분이 있다. 빅데이터(Big Data)는 이름처럼 큰 데이터만을 의미하는 것은 아니다. 물론 여러 가지 특징을 갖춘 데이터가 되려면 결과적으로 사이즈가 커질 수밖에 없겠지만 데이터 사이즈가 작더라도 소비자의 작은 행동 하나 하나 까지 파악해 그 무엇인가를 찾아 낼 수 있다면 그것도 빅데이터라 볼 수 있습니다. 빅데이터의 BIG은 Volume 보다 Value에 더 가까울 수 있다는 것이다. "건물이 무너졌다"는 트윗 백만개 보다 "건물이 무너질 것 같다"는 트윗 하나가 재난 분야에 있어서는 훨씬 더 의미가 있는 스마트 데이터가 될 수 있는 것이다.



빅데이터의 토대는 바로 디지털이다. 기계에서 생성되는 각종 로그 데이터나 시그널 데이터 등과 같은 머신데이터와 소셜 미디어나 온라인 포털 등 인간이 디지털에 남긴 흔적들이 모인 휴먼데이터가 빅데이터의 재료가 된다. 때문에 잘 해석된 빅데이터는 사회의 흐름이나 대중의 심리, 소비 트렌드 등을 파악할 수 있는 하나의 신호가 될 수 있다. 빅데이터로 찾아낸 '시그널'들은 우리가 새로운 비즈니스의 기회를 포착하는데 아주 중요한 길잡이가 될 수도 있다. 패턴 분석과 미래 전망은 물론 빠른 의사결정과 실시간 대응 지원, 트렌드 변화 분석을 통한 제품 경쟁력 확보, 타 분야와의 결합을 통한 새로운 가치창출이 가능해진다.

이제 빅데이터는 분석적 영역을 넘어 사회전반에 주어진 과제로 다가온 느낌이다. 누가 얼마나 많은 데이터를 분석하느냐보다 누가 먼저 데이터의 시그널을 잡느냐에 따라 모든 운명은 결정될 것이다.

김도훈 ◆ 아르스 프락시아 대표이사

3 텍스트는 어떻게 분석되어 지는가



텍스트 마이닝이란

비정형 데이터 분석이 각광을 받고 있다. 그도 그럴 것이, 빅데이터가 넘쳐나는 요즘 세상에서 세상에 존재하는 데이터의 80% 이상이 비정형 데이터로 추산된다(Chakraborty and Pagolu, 2014). 그 중에서도 텍스트는 가장 기본적이고 광범위한 비중을 차지하는 비정형 데이터이다. 이런 텍스트를 언어학, 수학, 통계학, 컴퓨터공학 등의 학문적 지식을 이용하여 특정 목적에 맞게 유의미한 정보를 추출하는 분석 및 처리 과정을 텍스트 마이닝(text mining)이라고 한다. 비록 요즘 텍스트 마이닝이 빅데이터의 신기원을 여는 기술인 양 각광을 받고 있지만, 그에 활용되는 학문적 지식들은 사실 1970년대부터 활성화된 것으로, 대체로 반세기 정도의 역사를 갖고 있다고 볼 수 있다. 반대로 얘기하면, 지난 50여년간 축적되어 온 지식들이 최근의 빅데이터 시대에 어느 정도의 현실 정합성과 효과성을 가지는 지는 새로운 경험적 관점과 평가를 요한다.

필자는 비정형 데이터 분석에 전문성을 갖고 있는 리서치/컨설팅 회사를 운영하고 있다. 유럽에서 공부할 당시 사회심리학, 네트워크 과학, 사회과학 방법론으로서의 텍스트 분석이 주 전공이었고, 한국에 돌아와서는 수학, 통계, 컴퓨터공학 등의 전문가들과 함께 데이터 분석에 기반한 컨설팅을 10여년 가까이 업으로 하고 있다. 그 과정에서 느꼈던 텍스트 마이닝 방법론의 실제적 효용과 한계, 앞으로의 도전 과제에 대해 짤막하게나마 언급해 보고자 한다.

텍스트를 분석하는 다양한 방법들

텍스트는 다루기 어려운 데이터다. 허스트(Hearst, 1999)는 텍스트가 모호하고 추상적인 단어로 가득차 있고, 같은 단어도 문맥에 따라 개념이 변하며, 유사한 개념(동의어, 유의어)도 표현하는 다양한 방법이 존재하고, 통계적 관점에서 단어 하나하나가 각각의 차원을 이루기 때문에 쉽게 처리하기 힘든 고차원성을 가진다고 지적한다. 분석한 주요 특징을 시각화하기 어려운 것은 덤이다. 이러한 어려운 특징을 가진 텍스트의 핵심 의미를 캐내는 방법을 정형화하기 위해 수십년간 다학제적 연구가 이루어져 왔다고 이해할 수 있다. 그 결과, 몇가지 방법들이 널리 통용되게 되었다. 다음 페이지는 그런 자연어 처리의 주요 개념과 방법들의 예시이다.



자연어 처리의 주요 개념과 방법들의 예시

형태소(단어) 분석 의미의 최소단위인 형태소(morpheme)이나 단어에 대한 분석은 텍스트를 분석하는 가장 보편적인 방법으로, 문법적 규칙 혹은 확률에 의한 품사 태깅(part of speech tagging), 개체명 인식(named entity recognition), 철자 교정, 단어 식별(tokenization) 기법 등을 이용한다.

문자열 분석 영어의 음운이나 한국어의 글자의 갯수(n-gram)를 지정하여 전체 텍스트 코퍼스를 분석함으로써 해당 문자열이 나왔을 때 그 다음에 어떤 글자가 나올지를 확률분포를 통해 예측한다.

핵심어구 추출 텍스트를 어휘적으로 상호 관련 있는 단어들로 나누는 청킹(chunking)을 통해 핵심어구를 추출하거나, 개체명 인식, 관계 추출(relation extraction) 등에 활용한다.

벡터 공간 모델 요즘 머신러닝 등에서도 종종 언급되는 벡터 공간 모델은 솔튼 등(Salton et al, 1975)에 의해 1970년대에 주창되었고, 지금도 활발히 활용되는 방법이다. 벡터 공간 모델에서 문서(document)와 단어(term)를 각각 벡터(vector)와 차원(dimension)에 대응시켜 통계적 가중치를 구한다. TF-IDF, 카이제곱 검정, 코사인 유사도 등을 통해 단어의 빈도수 분포에서 중요하지 않은 단어를 걸러내고, 문서의 유사도를 구한다.

동시출현 단어 분석 단어들 사이의 의미상의 관계성을 파악하기 위해 일정한 문맥 내에서 두 단어가 동시출현하는 빈도를 구한 후, 다양한 통계적 방법을 활용하여 유의미한 단어쌍을 추출해 낼 수 있다.

토픽 모델링 구조화되지 않은 대량의 텍스트로부터 숨겨져 있는 주제 구조를 발견하고 카테고리화 하기 위한 통계적 추론 알고리즘으로, LDA (Latent Dirichlet Allocation) 모델이 주로 활용된다. 각각의 창발적인(emergent) 주제를 각 행에 배열되는 단어들의 확률분포를 통해 표현한다.

텍스트의 감성 분석 감성 사전을 기반으로 분석 대상이 되는 전체 텍스트의 감성 비율을 정량화하는 기법이다. 통상적으로 긍정, 중립, 부정 등의 '평가어' 분석이 감성분석으로 이해되나 이는 실제론 기쁨, 우울, 화남 등 심리학적인 감성 카테고리에 기반한 정량화의 한 특수한 사례이다.

의미연결망 분석 키워드 동시출현 분석 기법에 기반하거나 그 한계를 넘어[1] 단어간의 의미 혹은 맥락상의 연결관계를 정의하고, 해당 연결 관계를 시각화하거나 중요한 컨셉을 네트워크 속의 위상(centrality)에 따라 추출하는 방법이다.

머신러닝 최근 '인공지능'이라는 마케팅 용어를 통해 텍스트 분석에서 활용되는 머신러닝 기법을 설명하는 경우들이 있으나, 엄밀히 얘기하면 머신러닝의 여러 기법들 역시 대부분 학계에서 이미 수십년간 알려진 것들이다. SVM(Support Vector Machine), 의사결정트리(Decision Tree, 랜덤 포레스트(Random Forest) 등이 대표적이고, 최근에는 딥 러닝(Deep Learning) 기법이 활용되고 있다.



1) 필자가 고유한 전문성을 가지고 학계와 현실세계에서 기여해 온 영역 중 하나는 의미연결망 분석(semantic network analysis)의 방법론적 모델을 개선하고 정교화 하는 작업이다. 기존의 방법론이 단순히 동시출현 키워드들을 통계적으로 추출한 후 연결시킨 결과를 데이터로 시각화하거나 키워드의 네트워크 중심성을 이용하여 의미를 해석하는 데 반해, 키워드의 코딩, 추출, 해석 방법론을 혁신하고 현실세계의 문제를 해결하는 사례들을 축적해 왔다. 자세한 내용은 Kim, L. and Kim, N. (2015); Kim, L (2013) 참고.

텍스트 마이닝 기술 이렇게 활용된다

실시간으로 데이터를 다루고, 고객에게 실제 도움이 되는 높은 품질의 분석 결과를 전달해야 하는 업계에서는 해당 기술들을 어떻게 활용하고 있을까? 사실, 행위자(업체, 연구자)에 따라 그 수준과 퍼포먼스가 천차만별이기 때문에 '업계'의 경험적 관점을 일반화하기는 어렵다. 그럼에도 불구하고, 나름 이 분야의 최전선에서 기존의 방법론을 테스트하고 새로운 방법론을 개발해 온 입장에서 몇가지 생각들을 정리해 본다.

첫째, 자연어 처리(Natural Language Processing)는 데이터 전처리-프로세싱-분석-해석으로 이어지는 텍스트 마이닝에서 제일 앞단을 차지하는 과정이다. 사회에서 사람들을 만나면, 자연어 처리 자체가 복잡한 데이터의 분석을 위한 프로세싱을 하는 전가의 보도인양, 잘못 이해하는 경우가 많다.

둘째, 앞서 설명한 벡터 공간 모델, 동시출현 단어, 토픽 모델링은 복잡한 텍스트의 모래에서 대강의 특징만을 추출한, 매우 거친 체와도 같다. 그 자체만으론 현실세계에서의 상식과 경험을 능가하는 통찰을 찾아내기 어렵다.

셋째, 감성분석, 의미연결망 분석, 머신러닝은 서로 융합되면서 보다 정교하게 발전할 여지가 많다. 단, 빅데이터의 크기가 방법론적으로 산적한 과제들을 자동적으로 해결해 주진 않는다. 새로운 이론적 가정과 방법론적 숙고가 필요하다.

마지막으로, 같은 맥락에서 머신러닝 역시 보다 나은 수퍼비전(supervision)을 위한 방법론적 천착과 데이터 테스트의 축적이 필요하다. 우리의 경험으로 봤을 때, 인공지능의 학습을 위한 초기의 데이터(seed data)를 정제하고, 인간의 통찰력과 가정을 적용하여 피드백을 받는 과정이 머신러닝의 품질과 실제 필드에서의 해석적 가치를 결정하기 때문이다.

참고문헌

- Chakraborty G and Pagolu M (2014) Analysis of Unstructured Data: Applications of Text Analytics and Sentiment Mining. *SAS Global Forum*.
- Hearst M (1999) Untangling Text Data Mining. *Association for Computational Linguistics*, University of Maryland, 20–16 June.
- Kim L and Kim N (2015) Connecting opinion, belief and value: semantic network analysis of a UK public survey on embryonic stem cell research. *JCOM*, 14 (01), A01.
- Kim L (2013) Denotation and connotation in public representation: semantic network analysis of Hwang supporters' internet dialogues. *Public Understanding of Science*, 22(3): 335 – 350.
- Salton G, Wong A and Yang C S (1975) A Vector Space Model for Automatic Indexing. *Communications of the ACM*, 18(11): 613–620.

김광석 ◆ 삼정KPMG 경제연구원 수석연구원

한국의 4차 산업혁명, 어디까지 왔는가

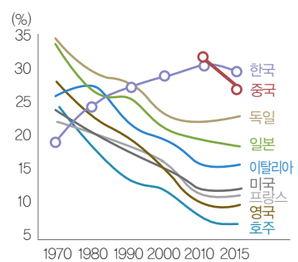




한국 경제에서 제조업의 영향력은 상당하다. 세계 주요국들과 비교해 보아도 제조업 경쟁력을 강화하는 것은 매우 중대한 과제일 수 있다. 지난 약 50년 동안 세계 각국은 금융업과 지식기반 서비스업 등을 중심으로 부가가치가 증대되면서 제조업이 총부가가치에서 차지하는 비중이 하락해 왔다. 미국은 1970년 약 23.6%에서 지속적으로 하락하여, 2015년 기준으로 12.0%에 달한다. 제조업 기반의 국가라고 할 수 있는 독일과 일본도 제조업이 총부가가치에서 차지하는 비중이 하락해 왔고, 주요국들도 마찬가지이다. 중국의 경우 제조업이 총부가가치에서 차지하는 비중이 2006년 32.5%로 최고점을 기록하였지만, 경제구조 개편과 서비스업 중심으로 재편되면서 제조업 비중은 상당 폭 하락하여 2015년 27.0%를 기록하고 있다. 한국도 서비스업을 중심으로 경제규모가 확대되면서 2010년 이후로 제조업 비중이 하락해 왔지만, 총부가가치의 29.5%가 제조업에서 비롯되고 있다.

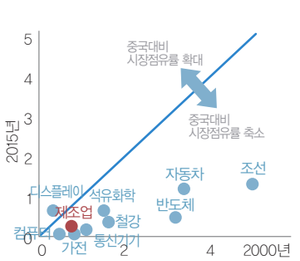
그럼에도 불구하고, 한국의 제조업 경쟁력은 약화되고 있다. 제조업 경쟁력을 판단하는 주요 대상국인 중국과 견주어 보았을 때, 제조업 전반에 걸쳐 한국의 세계 시장점유율이 중국 대비 축소되는 경향이 나타난다. 한국의 중국 대비 세계 시장점유율은 제조업 전체 평균이 2000년 0.73에서 2015년 0.23으로 떨어졌다. 디스플레이 업종을 제외한 모든 제조업 영역에서 유사한 특징을 보인다. 4차 산업혁명 대응을 통한 제조업 혁신이 요구되는 시점이다.

제조업이 총부가가치에서 차지하는 비중



자료 : UN Statistics

중국 대비 한국 제조업의 세계 시장점유율 변화



자료 : UN Comtrade

스마트 팩토리 도입과 산업 패러다임 변화



기업들이 기존 제조공정에 스마트 팩토리를 도입하면서 제조업의 혁신이 일어 나고 있다. 스마트 팩토리의 주요 기반 기술들로는 사이버물리시스템(CPS), 로보틱스, 3D 프린팅, 포그(Fog) 컴퓨팅, 사이버 보안 기술 등이 있다. 이러한 주요 기반 기술들이 제조 전 영역에 걸쳐 적용됨에 따라 제조업의 경쟁력이 놀라울 만큼 증대되는 제조 혁신이 나타나고 있다. 화학, 자동차, 철강, 항공, 식료품, 섬유 등 다양한 제조 산업에 걸쳐 스마트 팩토리를 도입하게 되면서, 생산성이 극대화 되고 기존에 소비자에게 제공하지 못하던 다양한 서비스 제공이 가능해 지고 있다.

이에 따라 제조업에 변화가 일고 있다. 실시간 주문형 맞춤형 생산이 가능해지고, 제조공정의 디지털화(Digitalization)가 가속화 되고 있다. 재고량을 극소화하고, 제품 불량률을 낮추며, 인건비가 절감되면서 생산성 혁신이 나타나고 있다. 생산라인뿐만 아니라 공급사슬 전 공정에 걸쳐 사물인터넷, 센서, 클라우드 기반의 초연결화가 가능해 지면서, 제조사와 부품 공급업자간의 유기적인 연결성이 강화 되고 있다. 3D 프린팅 활용을 위한 소재의 첨단화가 진행되고 있고, 내구성·내열성이 요구되는 로봇을 위한 첨단 소재나, 초정밀 공정을 위한 첨단 신소재 등에 관한 관심이 높아지고 있다. 마지막으로, 기존의 아날로그 생산 공정에서 필요한 기계, 부품 등의 자산 보안에서 빅데이터 중심의 사이버 보안으로 중점 보안분야가 이전되고 있다.

제조 혁신… 어떻게 가야하나

그럼 제조업의 혁신 방향은 어떻게 가야할까. 6가지 측면으로 제시해 본다. 첫째, 맞춤형 생산을 통해, 소비자 기호에 맞춤화된 제품과 서비스를 제공할 필요가 있다. 최근 주요 기업들은 D&A(Data and Analytics) 등에 기초해 실시간으로 소비자 기호에 맞춤화된 제품을 공급하고, 서비스를 제공하고 있다. 소비자들은 제품의 생산과정에까지 참여하는 적극적인 소비패턴을 보이고 있다. 기업들이 소비자의 다양한 기호변화에 대응하고, 소비자의 제품개발 및 유통과정 참여를 위해서는 실시간으로 유연하게 대응할 수 있는 스마트 팩토리 도입이 요구될 것이다.



둘째, 디지털 노동(Digital Labor) 패러다임 변화에 대응해야 한다. 고령화 및 생산가능인구 감소 현상을 겪고 있는 주요 선진국들뿐만 아니라, 인건비 상승 등에 부담을 느끼는 신흥국에서도 기업들은 제조 경쟁력을 갖추기 위해 디지털 노동을 적극 도입하고 있다. 자사의 생산 공정의 특수성을 감안하여, 적절한 스마트 팩토리 도입을 고려할 필요가 있다.

셋째, 생산성 혁신을 통해 경쟁력을 강화해야 한다. 기업들은 업종, 생산 품목, 가용 자원, 기존 생산 설비 등의 여건을 고려하여 스마트 팩토리의 기반 기술들을 적절히 제조과정에 도입하고 있다. 기업들은 생산성을 극대화 하고, 제조 경쟁력을 강화해 나가고 있다. 조립, 포장, 하역, 운반 등의 다양한 생산 공정에서 스마트 팩토리 도입을 통해 생산성을 증대시킬 가능성을 타진하고, 적절한 기반 기술들을 적용시켜야 한다.

넷째, 공급사슬 구조의 연결성을 강화할 필요가 있다. 복잡하게 얽혀 있는 공급사슬 구조하에서 자사만 독립적으로 스마트 팩토리를 구현하면 큰 의미가 없기도 하다. 공급사슬 전체가 유기적으로 연결되고 호환성을 강화해야만 실시간 부품 공급체계 및 물류관리 시스템이 확보될 수 있기 때문이다. 이를 위해서는 한국 정부의 '가치사슬 내 대기업 협력사 지원' 등을 활용할 필요도 있다.

다섯째, 첨단 소재를 확보해야 한다. 스마트 팩토리를 도입하는 과정에서 로보틱스가 내열성, 내구성을 보완하거나, 생산 공정 내 센서 및 IoT를 부착하는 영역 등에서 첨단 소재가 요구된다. 이는 3D 프린팅의 활용성을 증대시키기 위한 기초 기술에 해당하기도 한다. 첨단 소재 개발에 대한 R&D를 검토하고, 미래 유망 첨단 소재를 보유한 기업들을 M&A하는 등의 적극적인 행보가 필요한 시점이다.

여섯째, 사이버 보안을 강화해야 한다. IoT 기반의 데이터가 급증하고, 고객 기반의 데이터 및 다양한 비정형 데이터가 제조 공정에 활용되면서 사이버상에서 발생하는 위협 또한 증가하고 있다. 사이버 물리 시스템에 기반한 스마트 팩토리 제조 공정의 외부 해킹은 제품의 설계 정보 유출로 연결되고, 엄청난 양의 개인정보유출로까지 이어져 경제적·사회적 문제를 야기할 수 있다. 따라서 스마트 팩토리를 도입하는 기업들은 네트워크 보안, 클라우드 보안, 빅데이터 보안을 매우 중요한 과제로 인식할 필요가 있다.



스마트 팩토리의 기반 기술과 패러다임 변화

스마트 팩토리 기반 기술		제조업 패러다임 변화
사이버물리시스템(CPS)	제조 혁신	맞춤형 생산
로보틱스		디지털 노동(Digital Labor)
3D 프린팅		생산성 혁신
포그(Pog) 컴퓨팅		초연결화
사이버 보안 기술		소재의 첨단화
		자산 보안에서 사이버 보안으로

자료 : 삼정KPMG 경제연구원

이강봉 ◆ 통계의창 객원기자



보건·의료계에서 불어오는 빅데이터 혁명

인체 유전자정보, 핵심 통계 데이터로 떠올라

데이터을 통해 신종플루의 대비책을 진단하다

지난 2009년 봄 멕시코와 미국 등지에서 신종 바이러스 'H1N1/09'가 발생해 빠른 속도로 확산되기 시작했다. 사람·돼지·조류 인플루엔자 바이러스의 유전물질이 혼합돼 있는 새로운 형태의 바이러스였다.

처음에는 '돼지인플루엔자(돼지플루)' 또는 '돼지독감'이란 명칭을 사용했다. 그러나 돼지와 관련이 있다는 증거가 없다는 주장이 잇따라 제기되면서 세계보건기구(WHO) 실무자들은 '신종 인플루엔자A(H1N1)'로 명칭으로 호칭을 바꾸었다.

'신종플루'의 기세는 무서웠다. 전 세계로 확산되기 시작했다. 공포에 질린 WHO 실무자들은 최악의 사태를 생각해야 했다. 마지막 대비책에는 세계 주요 공항 폐쇄라는 극단적인 대책이 들어 있었다. 그러나 신종플루에 대한 대비책은 두 명의 이탈리아 과학자로 인해 큰 반발에 부딪혔다. 이들은 공항을 폐쇄한 후 신종플루 확산을 몇 주 정도 유보시킬 수는 있으나 근본적인 대책이 될 수 없다고 말했다. 오히려 경제적 혼란을 가져오고 세계를 더 큰 혼란에 몰아넣을 것이라고 주장했다.

이런 예측을 한 사람들은 의사들이 아니었다. 물리학자들이었다. 이들은 구글, 페이스북, 미국가안전국(NSA)과 같은 빅데이터 기업에서 일하는 연구원들처럼 수학을 효율적으로 활용해 신종플루 확산이 어떻게 이루어지고 있는지 그 상황을 정확히 예측할 수 있었다. 거기서 나온 데이터를 기반으로 WHO의 신종플루 대책이 현실과 맞지 않는 허황된 계획임을 정확히 짚어내고 있었다.



빅데이터로 비타민B의 비밀 밝혀내

이후 8년이 지났다. 그리고 지금 유전자 연구는 물론 에너지, 무역, 농업, 고고학 등 빅데이터를 적용하는 분야가 끝없이 확산되고 있다. 또한 이를 통해 그동안 수수께끼에 싸여 있던 미지의 세계들이 하나둘 그 실체를 드러내고 있다. 최근 들어 가장 큰 주목을 받고 있는 곳이 의료 분야다. 지난 8월 미국 오하이오주립대에서 놀라운 연구 결과를 내놓았다. 건강보조제로 자주 찾는 비타민 B6와 B12를 과다 복용할 경우 담배를 피우는 남성의 폐암 발병률이 4배 까지 높아진다는 것. 비타민 B6는 단백질 대사에 중요한 효소 구성 성분이고, B12는 핵산 합성과 조혈 작용에 관여하는 매우 중요한 영양소였다. 환자를 대상으로 축적한 빅데이터 분석을 통해 그 위험성이 밝혀진 것이다.

연구책임자 테오도르 브래스키(Theodore Brasky) 교수는 이 건강보조제를 10년 동안 다량으로 복용한 모든 남성의 폐암 발병 위험이 2배 정도 높았다고 밝혔다. 또 흡연자의 경우 위험이 3~4배 더 높아졌다고 경고했다.

이번 연구가 빅데이터 분석을 통해 이루어진데 대해 세계 의료계가 큰 놀라움을 표시했다. 지난 10년간 흡연자, 비타민 B 과다복용자, 폐암 환자 간의 상관관계를 추적해 이루어낸 데이터 사이언스의 개가였다.

흥미로운 사실은 이 연구에 'SEER(Surveillance, Epidemiology and End Results)'란 데이터베이스 시스템이 적용됐다는 사실이다. 'SEER'란 환자의 병기(illness, 病氣) 체험을 데이터화한 '요약병기(Summary Stage)'란 분류 방식을 말한다. 데이터베이스를 통해 암 환자 관련 데이터를 수집하기 시작한 것은 1973년이다. 그리고 최근 디지털, 빅데이터 분석 기술이 발전하면서 데이터베이스에 있는 기본 자료를 활용해 다양한 패턴의 연구를 수행하는 일이 가능해졌다.

실제로 미국국립암연구소(NCI)에서는 'SEER'를 통해 암이 발생한 장기로부터 암 증상이 어느 정도 진행됐는지를 데이터화하고 있었고, 오하이오주립대 연구진은 이 시스템을 활용해 비타민 B6와 B12의 비밀을 밝혀낼 수 있었다.

제약사들, 처방전 분석으로 신약 개발

주목할 점은 앞으로 빅데이터를 이용한 이런 유형의 연구가 확대될 것이라는 점이다. 의사 입장이 아니라 다수의 환자 관점에서 병기 체험과 관련된 데이터를 축적해 암과 관련된 또 다른 현상을 확인하는 일이 가능해졌다. 병원 업무 전반에 빅데이터 기술을 도입하는 병원도 늘고 있다. 캐나다 맥길대학 정신건강연구소는 최근 치매를 예측할 수 있는 새로운 알고리듬을 개발했다. 이를 통해 치매 증세가 나타나기 2년 전에 증상을 예고할 수 있다. 정확도는 84%. 이 시스템에는 인공지능(AI)이 동원되고 있다. 연구소 측은 "경도 인지장애(MCI) 환자 273명을 대상으로 촬영한 PET(양전자방출단층촬영) 영상 자료를 기반으로 알고리듬을 학습시켜 치매예측 장치를 만들어내는데 성공했다."고 밝혔다.

제약사들 역시 서둘러 빅데이터 기술을 도입하고 있다. 특히 많은 제약사들이 빅데이터 기술을 활용해 매우 강력한 의약품 판매 시스템을 개발 중인 것으로 알려지고 있다. 실제로 의약품 유통망은 수시로 상세한 처방정보가 축제되는 알고리즘이다.

의료법상 환자의 이름이 입력되지 않는 처방은 불가능하다. 이에 따라 제약사들은 현재 병원과 협의해 의사들의 처방전을 데이터화 하는 방안을 찾고 있는 중이다. 그럴 경우 제약사에서는 병원 · 약국 등에서 어떤 의약품이 처방되고 있는지 파악이 가능하다. 제약사들은 "의사들의 처방전을 데이터 분석함으로써 환자 치료에 큰 도움을 줄 수 있다."고 주장하고 있다. 또한 처방전 분석을 통해 각각의 의약품이 어떤 반응을 얻고 있는지 분석할 수 있으며, 평가결과를 기반으로 새로운 의약품을 개발할 수 있다고 보고 있다.

암 세포 데이터 축적해 원인 규명

빅데이터와 관련, 최근 의료계의 큰 관심사 중의 하나는 빅데이터를 통해 암의 비밀을 밝혀낼 수 있는지 그 여부다. 흥미로운 것은 이 작업에 MS, 구글 등 IT 공룡기업들과 함께 대형 병원들이 다수 참여하고 있다는 점이다.

실제로 마이크로소프트(MS)는 지난해 9월부터 암 세포 정보를 컴퓨터에 담기 시작했다. 최장 10년 동안 진행될 이 작업을 통해 암 세포의 비밀을 밝혀내고 그 정보를 활용해 암이 생기는 원인을 밝혀낸 후 암 치료법에 활용할 계획이다. 이처럼 방대한 양의 암 세포 데이터를 컴퓨터에 대량 입력하기는 MS가 처음이다. 연구에 참여하고 있는 크리스 비숍(Chris Bishop) 박사는 "생물학과 컴퓨터가 분필과 치즈처럼 다른 느낌을 주고 있지만 그 안에서 일어나는 일은 흡사한 양상을 보이고 있다"고 말했다.

그는 "지금과 같은 방식으로 컴퓨터 안에 방대한 양의 암세포 정보를 입력해나갈 경우 5~10년 후에는 암 치료를 위한 매우 중요한 정보들을 다양하게 획득할 수 있으며, 암 치료에 혁신을 가져올 수 있다"고 설명했다.

"5년 후 암 발병 원인 밝혀낼 수 있어"

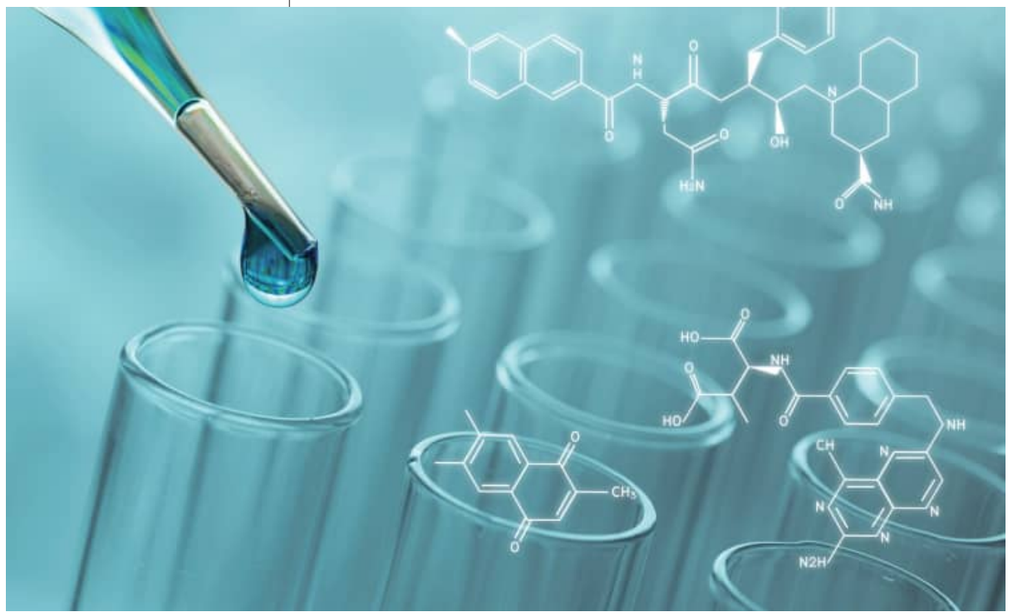
현재 이 프로젝트를 수행하고 있는 연구소는 케임브리지 대학 안에 설치돼 있다. 150명의 과학자와 소프트웨어 개발자들이 일을 하고 있는데 그중에는 저명한 생물학자, 프로그래머, 엔지니어 등이 다수 포함돼 있는 것으로 알려지고 있다. 연구팀장인 앤드류 필립스(Andrew Philips) 박사는 "암세포를 모방해 살아있는 컴퓨터(living computer)를 만들고 있다"고 말했다. "정보축적 작업을 5~10년 동안 수행할 경우 분자 차원에서 암을 정밀 진단할 수 있는 길이 열린다"고 말했다. 다음 단계는 이들 건강한 세포와 암 세포를 비교해 어느 부분에서 문제가 발생하고, 암 증상을 일으키는지 분석할 수 있도록 암 세포의 움직임을 모사하는 일이다. 이를 위해 연구팀이 개발한 소프트웨어 속에 암 세포 관련 정보를 입력하고 있는 중이다.

재스민 피셔(Jasmin Fisher) 책임연구원은 "이 작업이 마무리되면 그동안 몰랐던 암 증상의 원인을 알 수 있다"고 말했다. 그는 "일부 암의 경우 5년 정도면 분석이 완료되고, 10년이 지나면 거의 모든 암을 분석할 수 있을 것"으로 내다봤다.

MS에 근무하고 있는 많은 수의 컴퓨터과학자들은 지금처럼 컴퓨터가 암세포를 분석하고 있는데 대해 컴퓨터과학이 수행해야할 당연한 수순으로 인식하고 있다. 암과 같은 인류의 난치병들을 컴퓨터과학이 순차적으로 해결할 수 있다고 믿고 있다.

유전자 염기서열 분석해 치료법 개발

MS 연구소의 지넷 윙(Jeannette M. Wing) 부사장은 두 가지 전제 속에서 이번 프로젝트를 수행하고 있다고 말했다. 첫 번째는 암을 비롯해 어떤 질병이든지 컴퓨터를 통해 정보처리할 수 있다는 확신이다. 이에 따라 암세포를 비롯 다양한 질병 분석이 이루어질 전





망이다. 두 번째 전제는 '머신러닝' 같은 기계학습 프로그램을 통해 질병의 원인 분석을 인지하게 한 후 치료 방법을 추정해나갈 수 있다는 생각이다. 그럴 경우 그동안 의사들이 해오던 영역을 컴퓨터가 수행할 수 있게 된다.

윙 부사장은 "이런 목적을 위해 생명과학자들과 의료인, 컴퓨터과학자 등이 협의해 컴퓨터가 의료 활동을 할 수 있는 혁신적인 방안을 찾고 있다"고 말했다. 실제로 의료 기업과의 기술협력이 추진되고 있다. 또 윙 부사장은 "지금 개발하고 있는 모델을 신약을 개발하고 있는 제약회사에 적용하는 방안을 협의 중"이라고 말했다.

컴퓨터분석을 통해 불치병을 치료하려는 노력을 기울이고 있는 곳은 MS뿐만이 아니다. IBM의 인지컴퓨팅 왓슨(Watson)팀은 유전체 염기서열 자료 분석을 통해 새로운 치료방법을 개발하고 있는 중이다. 그러나 MS의 프로젝트는 불치병으로 알려진 암 세포를 대상으로 하고 있어 세계적인 주목을 받고 있는 중이다.

최근의 이런 변화는 국가 보건관리 시스템 전반에 큰 변화가 일어나고 있음을 말해주고 있다. 현재 부분적으로 이루어지고 있는 이 정보교환 시스템이 빅데이터화할 경우 국가는 물론 세계적으로 의료계와 환자 모두를 실시간으로 아우르는 일이 가능하다.

한국 정부, 의료빅데이터 플랫폼 구축

국내에서도 준비가 한창이다. 보건복지부는 최근 발표를 통해 보건의료 속성, 데이터 활용 의약연구의 중요성 등을 고려해 보건의료분야 정보를 목적으로 하는 특별법인 가칭 '보건의료 빅데이터 특별법' 제정을 추진한다고 밝혔다.

"기존 개인정보보호법의 경우 '건강정보'를 하나의 분류로 규정해 종류별 처리 방법이나 안전수칙, 허용범위 등이 세분화되지 않았다는 점이 존재했다."며, 보건의료 빅데이터의 활용에 제한을 막기 위한 조치라고 설명했다. 개인정보보호법에서의 '건강정보'는 키나 몸무게, 진료기록, 유전자 등이 구별되지 않은 상태에서 공통적으로 통칭되고 있



보건복지부 의료빅데이터 플랫폼 구축

는 상황이다. 복지부는 향후 법안 마련 시 중요내용 및 방향성을 공론화위원회 논의를 거쳐 특별법을 추진할 것이라고 밝혔다. 복지부는 또 올해 말이나 내년초 설치될 보건의료 빅데이터 자문위원회를 승계한 정책위원회를 만들고 기본계획 수립과 국가투자 및 지원 등 규정 마련에 들어갈 계획이다. 특별법 제정은 '보건의료 빅데이터 플랫폼 시범사업 추진계획'의 일환이다.

현재 정부는 이 추진계획을 위해 2018년 예산을 당초 계획 77억원보다 41억 원이 더 늘어난 115억 원을 투입할 예정이다. 정부의 이런 움직임은 국내 의료빅데이터 플랫폼 구축이 늦어지면서 국내 의료 경쟁력이 위축됐다는 우려에 따른 것이다.

사생활 침해 등 시민단체 반발 거세

지난해까지 국내에서는 '생명윤리안전에 관한 법률'에 따라 의료기관을 통하지 않고 유전자 검사 서비스를 받는 것이 금지돼 있었다. 의료계가 반발하면서 지난해 6월 체질량지수, 카페인대사, 혈압 등 12가지 항목에 대해 DTC(direct-to-consumer) 유전자 검사를 허용했다. 미국 등 선진국이 암 예방을 위해 유전자 검사를 활용하는 상황에서 비만, 탈모 검사 정도를 허용하고 있다는 비판이 이어졌다.

개인정보보호법 역시 유전자 검사의 걸림돌로 작용하고 있다. 유전자 정보를 분석하려면 방대한 양의 인간 유전체 정보가 필요하다. 그러나 개인정보보호법에 따라 개개인의 유전체 정보를 수집해 빅데이터화하는 것이 불가능하다. 일부 공공기관과 연구실에서 연구를 목적으로 유전체 데이터를 사용한 뒤 폐기하는 것이 일반적인 관례가 되고 있다. 그러는 사이 세계 유전자분석 시장은 급성장하고 있다. 이런 움직임에 대해 시민단체의 반발도 더욱 거세지고 있다. 11월27일 국회 보건복지위원회 일부 의원들과 시민단체들은 국회의원회관에서 '현 정부의 보건의료 빅데이터 추진 전략의 문제점'이란 주제로 토론회를 개최했다.이 자리에서 참여연대, 건강과대안, 인도주의실천의사협의회, 진보네트워크센터, 무상의료운동본부, 경제정의실천시민연합, 건강세상네트워크 등의 시민단체들은 "보건의료 빅데이터 사업 추진을 위해 신뢰성과 공익성을 확보해야 한다."고 주장했다.

인도주의실천의사협의회 정형준 정책국장은 특히 "정부가 어떠한 개인 동의나 법적 근거도 없이 산업 발전이라는 측면에서 개인정보를 마구잡이로 활용하겠다는 입장으로 플랫폼 예산 115억원을 배정했다"며 의문을 제기했다.

논란 속에 의료 빅데이터 산업 재편 중

논란이 이어지고 있는 가운데 의료정보의 빅데이터화는 유례가 없을 만큼 빠르게 진행되고 있는 중이다. 아툴 뷰트 UC샌프란시스코 컴퓨터헬스사이언스 연구소장은 최근 강연을 통해 "구글이 5년 후 제약회사가 된다면 믿으시겠습니까?"라고 질문하고 있다.

"앞으로 구글처럼 방대한 데이터와 플랫폼을 보유한 회사들이 제약·바이오산업을 장악할 것"이라는 것이 그의 견해다. 의료계 관계자들 역시 불과 수년 안에 첨단 데이터 기술이 병원, 제약사 등에 도입되고, 의료계 전반에 걸쳐 큰 변화를 줄 것으로 예견하고 있다.

빈현준 ◆ 통계청 고용통계과장

사회적 양극화 해소의 실마리가 되는 **고용통계**

새 정부의 가장 중요한 국정과제는 좋은 일자리를 많이 만들어 국민들에게 제공하는 것이다. 다시 말해, 기업과 공공부문에서 더 많은 양질의 일자리를 창출할 수 있도록 기반을 조성하는 것이라 할 수 있다. 문재인 대통령도 '17년 5월 10일 대통령에 당선되자마자 업무지시 1호로 대통령 직속 일자리위원회를 설치하도록 하였다. 업무지시 1호가 갖는 상징성에 의미를 더해 대통령 본인이 일자리위원회 위원장을 자처하면서 동시에 실시간으로 일자리 관련 통계를 직접 챙기기 위해 집무실에 일자리 상황판까지 설치하였다. 결국 일자리 관련 정책들이 성과를 거두고 있는지, 취약한 영역은 어디인지 판단하기 위해서 일자리 관련 통계를 지속적으로 모니터링 하겠다는 의미일 것이다.

하지만 통계청과 고용노동부 등에서 생산하는 여러 고용통계들이 서로 일치하는 경우도 있지만, 서로 다른 방향을 나타내는 경우도 많아 국민들이 혼란을 겪는 경우도 있다. 이는 결국 통계를 정확하게 해석하기 위해서는 각각의 통계가 가지는 의미나 기준을 정확하게 이해하는 것이 중요하다는 것을 의미한다.

고용통계는 노동시장에서의 수요와 공급에 대한 정보

우리나라는 현재 경제활동인구조사, 지역별고용조사, 사업체노동력조사, 고용행정통계, 일자리행정통계 등 다양한 통계들을 통해 고용상황을 설명하고 있다. 하지만, 이러한 통계들이 가지는 의미에 대해서는 일반 국민들도 생소할 것이다. 이를 이해하기 위해서는 우선 각 통계들이 보여주고자 하는 측면이 무엇인지 알아야할 필요가 있다.



물가는 상품시장에서 수요와 공급이 만나는 지점에서 형성되는 가격에 대한 정보를 담고 있고, 환율은 외환시장에서 거래되는 외국환의 수요와 공급이 결정되는 지점에서의 외국환의 가격에 대한 정보를 담고 있다. 고용통계도 마찬가지이다. 노동시장에서 수요와 공급이 만나는 지점에서의 가격과 거래량에 대한 정보를 담고 있는 것이다. 다만, 상품시장, 외환시장 등에서는 공급자가 기업 또는 정부이고, 수요자가 개인 또는 가구이지만, 노동시장에서는 개인 또는 가구가 공급자이고, 기업 또는 정부가 수요자 역할을 한다는 점이 차이이다. 이점을 제외하면, 다른 시장과 크게 다르지 않다.

노동시장에서 결정되는 가격이 실질임금이 되고, 그 거래량이 노동수급량이 된다는 점에서는 동일하다. 따라서, 앞서 언급된 여러 고용통계들도 궁극적으로는 노동의 거래가격인 실질임금과 거래량인 노동수급량을 보여주기 위해 작성되어지는 것이다. 물론 노동시장이 완전경쟁시장이고, 통계조사의 오차가 없다면, 어떠한 조사를 통해 나온 결과도 서로 동일할 것이다. 하지만, 이는 이론상으로만 존재할 뿐 현실은 완전경쟁시장도 아니며, 통계적 오차도 피할 수 없다. 따라서 우리나라 노동시장의 참 모습을 추정하기 위해서는 다양한 고용통계를 통해 노동시장의 다양한 측면을 살펴볼 필요가 있다.

노동통계의 본래의 목적을 이해하자

우리나라 가장 대표적인 고용통계는 경제활동인구조사가 있다. 매월 공표된다는 점에서 시의성이 있으며, 전국민, 전 산업을 아우른다는 점에서 가장 보편적으로 사용되고 있다. 경제활동인구조사의 표본은 전국 33,000가구 내 거주하는 개인으로 노동의 공급측면을 파악하기 위한 조사이다. 경제활동인구부가조사, 지역별고용조사, 이민자체류실태 및 고용조사 등도 노동의 공급측면을 파악하기 위한 조사이다. 이들 조사에서 노동을 공급한 사람 즉, 취업자를 어떻게 볼 것인가가 가장 핵심적인 사항이다. 취업자를 어떻게 볼 것인가에 따라 우리나라의 총 노동공급량이 달라지며, 그에 따른 실질임금도 변하게 된다. 이 점에 관한 기준은 현재 국제노동기구(ILO)가 규정하고 있는데, 기준기간(1주일) 동안 수입(임금 또는 이익)을 목적으로 1시간 이상 일한 사람을 취업자로 판단한다. 이 부분은 일반 국민들이 이해하는 상식과 다소 동떨어졌다는 비판을 받기도 한다. 다시 말해 1주일 동안 1시간만 일해도 취업자라는 비판에 직면하게 되는데, 명확한 것은 우리나라 뿐만 아니라 전 세계 대부분의 국가가 이 기준을 준수하고 있다는 점이다.

우선 이를 이해하기 위해서는 노동통계의 본래 목적을 이해할 필요가 있다. 노동통계는 좋은 일자리를 가진 개인을 파악하는 통계이기에 앞서, 한 나라 또는 지역에서 투입되거나 사용된 총 노동량을 측정하는 통계이다. 경제활동인구조사는 노동의 공급측면을 조사하는 통계이므로, 총 노동공급량을 측정하는 것이 본래의 목적이다. 따라서 특정 기준시간 이상 일한 경우만을 취업자로 간주한다면, 단시간 근로를 통해 총생산에 기여한 노동공급량이 제외되어 그 나라의 총 노동공급량이 과소하게 측정될 수밖에 없다. 그러한 점에서 1시간 기준(one hour criterion)은 노동통계 기준의 금과옥조로 여겨진다. 우리나라도 그러한 점에서 동일한 기준을 따른다. 다만, 예외로서 무급가족종사자가 있다.

배우자나 부모의 일을 도우면서 따로 임금을 받지 않는 가족이 무급가족종사자인데, 이 경우만 우리나라에서 예외적으로 1주일에 18시간 이상 일했을 경우 취업자로 간주하고 있다. 국제기준은 이 역시도 동일한 1시간 기준을 적용하도록 하고 있지만, 우리나라의 경우 농림어업이나 숙박 및 음식점업 등에 종사하는 무급가족종사자 비중이 높기 때문에 이러한 특수성을 감안한 기준을 사용하고 있다.

이러한 점은 미국도 유사한데, 미국도 현재 1주일에 15시간 이상 일한 무급가족종사자에 대해 취업자로 인정하고 있다. 이러한 예외를 제외하면, 경제활동인구조사는 노동의 공급측면에서 총 노동공급량을 측정하는데 유용하다. 다만, 산업 및 직업의 상세한 분류 자료가 제공되지 못해 고용상황을 심도 있게 분석하는 데는 한계가 있다.

노동수요측면을 조사하는 사업체노동력조사

다음으로 노동시장을 노동수요측면에서 조사하는 사업체노동력조사가 있다. 전국 25,000개 표본사업체를 대상으로 매월 조사하여, 현재 사업체에서 수요 되는 총 노동수요량을 측정하는 조사이다. 동 통계는 경제활동인구조사에서는 파악할 수 없는 빈 일자리, 노동이동 등을 노동 수요적 관점에서 파악할 수 있고, 세전임금도 장부에 의해 비교적 정확하게 조사할 수 있다는 강점이 있다. 다만 호출일일근로자, 무점포수리업자, 대리운전기사 등 고정 사업장 없이 사업을 영위하는 취업자를 포착하기 어렵다 점에서 포괄성에는 한계가 있다.

2016년부터 행정자료를 활용하여 일자리 행정통계 작성

그리고, 2016년부터 기존의 조사 자료가 아닌 행정자료를 활용해 작성되는 일자리 행정통계가 있다. 동 통계는 국민연금, 건강보험, 사회보험, 고용보험, 사업자등록 자료 등 30종의 행정자료를 연계하여 행정자료 DB를 구축한 후 통계표를 작성하는 과정을 거치게 된다. 앞서 언급된 두 종류의 통계와 달리 신규일자리, 대체일자리, 소멸일자리에 대한 정보를 상세히 제공해 주며, 무엇보다 현장조사에 소요될 인력과 예산을 절감해 다양한 고용정보를 생산할 수 있다는 점에서 강점이 있다.

또한 연계된 모든 사람과 사업체 정보를 통해 통계를 생산하므로, 기업규모, 산업중분류 등 상세한 자료를 제공해 준다. 다만, 행정자료가 완전히 제공된 이후에 작성된다는 점에서 시의성이 떨어진다는 약점이 있다. 또한 일자리에 대한 정보는 제공해 주지만, 일자리를 구하려는 실업자나 경제활동에 참여하고 있지 않은 비경제활동인구에 대한 정보를 제공해 주지 못한다는 점에서는 한계가 있다. 지금까지 우리나라 고용시장을 대표하는 통계에 대하여 개략적으로 살펴보았다. 다들 노동시장을 완벽하게 보여주지는 못하지만, 나름의 관점에서 노동시장을 충실히 측정하고 있는 통계들이다. 특히, 조사환경이 갈수록 악화되고, 개인의 사생활 보호의식이 강화되는 사회적 분위기를 감안할 때, 현장조사에 의한 고용통계 작성이 갈수록 어려워질 수 밖에 없다. 그러한 점에서 행정자료를 활용한 일자리 통계를 다양하게 작성하는 것에 대해 더욱 고민한 필요가 있다.



이제는 고용의 질이다

지금까지 우리 통계의 주 관심영역은 취업자 수, 일자리 개수 등 고용의 양과 관련된 지표를 얼마나 정확하게 생산해 내느냐 였다. 이는 외국도 크게 다르지 않다. 하지만, 1999년 ILO에서 남녀 모두에 대하여 자유, 평등, 안전, 인권을 갖춘 적절하고 생산적인 일을 할 수 있는 기회를 촉진하는 것을 목표로 좋은 일자리에 대한 논의를 시작한 이래 고용의 질에 대한 관심이 증가하고 있다. 이후 UNECE, OECD, ILO 등에서는 고용의 질을 측정할 수 있는 지표체계 및 지표안을 작성에 이를 권고하고 있는 실정이다.

이에 우리나라도 2010년에 고용의 질 지표를 연구하였고, 이에 대한 내부적 논의는 지속되었으나, 두드러진 성과를 거두진 못한 상황이었다. 하지만, 새 정부 출범 이후 국정과제에서 드러났듯이 좋은 일자리에 대한 논의가 시작되었고, 통계청에서도 고용의 양적 측면 뿐만 아니라 고용의 질적 측면을 측정하고 평가할 수 있는 고용의 질 지표체계에 대한 논의를 하게 되었다. 하지만, 고용의 질을 정확하게 측정한다는 것은 불가능에 가깝다. 다만 이를 국민들이 수용할 수 있도록 최대한 논리적 구조를 가진 지표체계를 작성해 사회적 합의를 도출하는 것이 중요하다. 특히, 일부 지표에 의해 전체 고용의 질 측정이 영향을 받거나, 고용의 질과는 무관하게 시간의 흐름에 따라 좋아지거나 나빠질 수밖에 없는 지표가 과도하게 선정되는 것을 피하는 것이 필요하다. 현재 통계청에서는 이러한 점들을 최대한 고려하여, 경제, 사회, 근로조건 및 환경 3개의 차원, 임금, 근로시간, 산업안전, 고용안정, 고용차별, 일과 생활의 균형, 사회안전망 등 7개 영역에서 지표들을 선정하였다. 아직은 초기 단계라 이를 우리나라 고용의 질을 측정할 수 있는 지표체계라 확정할 수는 없지만, 그 동안 파편적으로 논의되던 수준에서 보다 체계적으로 고용의 질을 측정하고자 한다는 점에서 진일보 한 것으로 평가할 수 있다.

고용통계의 장점과 한계점에 대한 충분한 이해 필요

우리나라가 IMF 외환위기를 겪으면서 실업에 대한 전국민적 인식이 달라졌고, 최근 들어 청년실업이 가속화되는 현상을 보면서, 소득주도 성장과 사회적 양극화 해소의 실마리로 일자리가 거론되고 있다. 그리고, 그 중심에 다양한 정책을 수립하고 평가하는 잣대로 여러 가지 고용통계가 사용될 것이다. 하지만, 이러한 고용통계들이 가지는 장점과 단점 그리고 질적인 측면을 담아내지는 못한다는 한계를 충분히 인식하고, 정책당국은 정책을 수립할 필요가 있어 보인다. 또한 통계를 생산하는 부처에서는 정책당국이 적절한 정책을 수립하고 집행할 수 있도록 정확한 통계를 시의성 있게 생산하는데 더욱 노력할 필요가 있다.



강양석 ◆ 셀바스그룹 COO



데이터 스토리텔링… 데이터에 스토리를 입혀라

데이터는 명확한 근거 없이 메시지를 주고받을 때 발생하는 '메시지의 모호함, 근거의 희박함, 논점의 두서없음을 해소하는 데 상당한 도움을 준다. 잘 짜여진 데이터를 근거로 활용하다보면, 그만큼 소통의 초점도 명확해진다. 이것이야말로 우리가 데이터를 기반으로 소통해야하는 마지막 단계라고 믿는다. 과연 그런지 사례를 통해 설명해보고자 한다.

데이터가 가진 수많은 장점을 더 극대화해주는 방법이 있다. 바로 스토리를 갖는 것이다. 스토리를 가진 정보는 그렇지 않은 정보보다 훨씬 침투력이 강하다. 그러므로 복수의 데이터를 자유자재로 활용하여 메시지 전달을 달성하는 데이터 스토리텔링 능력이야말로 데이터를 활용하여 의사소통하는 방법 가운데 가장 진화된 형태다. 데이터 하나하나가 가진 사실성이 한 장면 한 장면 극대화되면서 전체적으로 하나의 메시지로 수렴할 때의 설득력은 잘 구성된 영화 한 편을 보고 난 후 밀려오는 감동에 비견할 만하다.

간단한 예로 데이터 스토리텔링의 힘을 확인해보자. 미 국방부의 자료를 바탕으로 만들어진 다음 열한 개의 데이터는 '전쟁터보다 무서운 것은 마음의 병입니다'라는 메시지를 더욱 극적으로 전달하기위해 스토리를 형성하며 배열된 좋은 예시다.

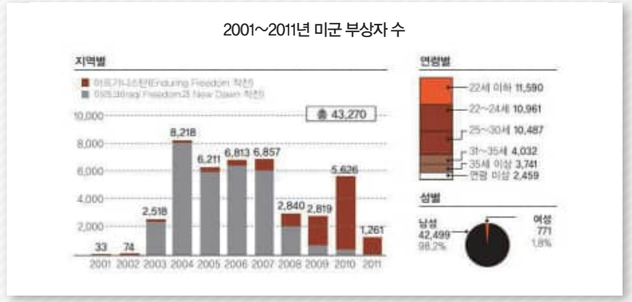
01

전 세계에 파병된
미군의 숫자는 2000년
초반부터 꾸준히
증가해왔습니다.



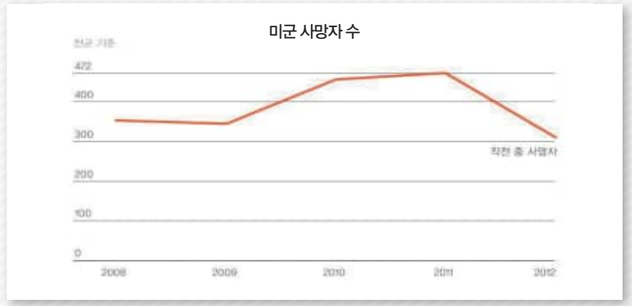
02

하지만 부상자 규모는
2000년대 중반
최고조에 이르다,
다행스럽게 최근
감소세에 있습니다.



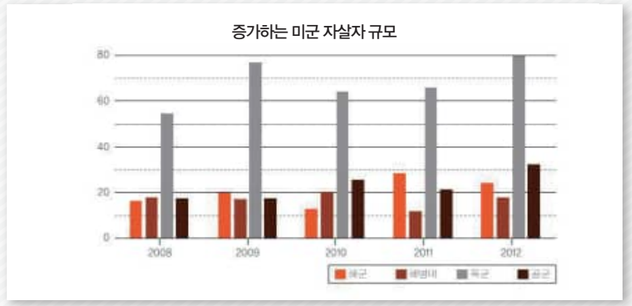
03

사망자 수 역시 최근
급감했습니다.



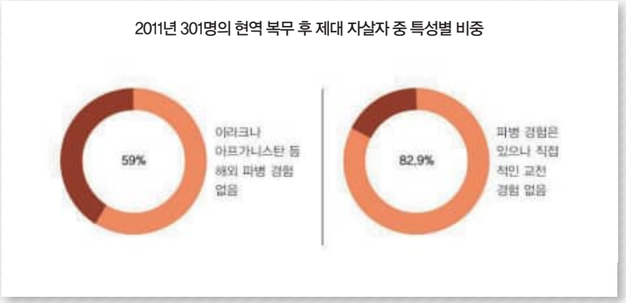
04

이상하게도
미군 전체에서 자살로
인한 사망자 수는
꾸준한 증가세를 보이고
있습니다.



05

이들 중 해외 파병이나
교전을 직접 경험한
사람의 비중은 그다지
높지 않습니다.



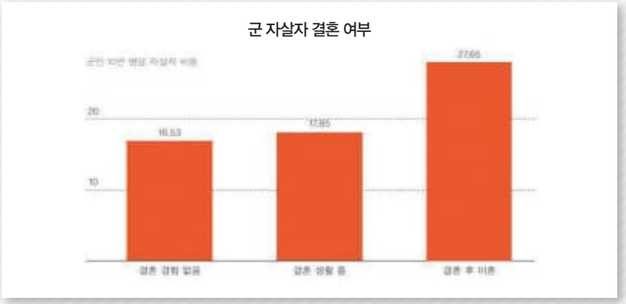
06

이들은 주로
혈기 왕성한 20대
청년으로…



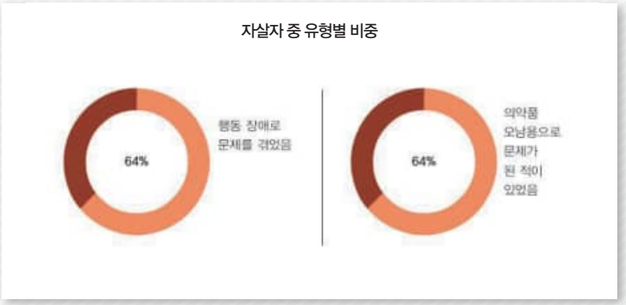
07

가정의 파괴와…



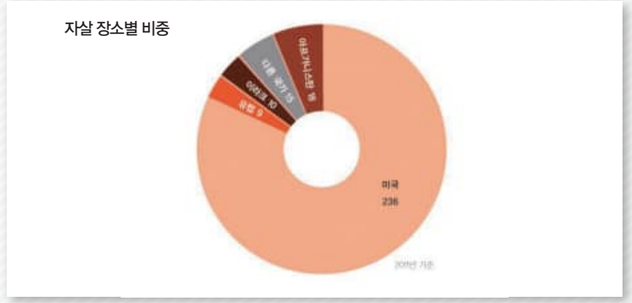
08

신체 장애 그리고
약물 오용 등으로
고통 받아왔던
이들이며…



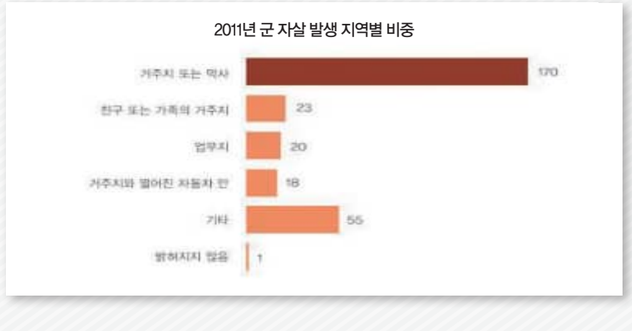
09

전쟁터가 아닌
미국땅에서…



10

그것도 자신의
막사 또는 거주지에서
극단적인 선택을
했다고 합니다.



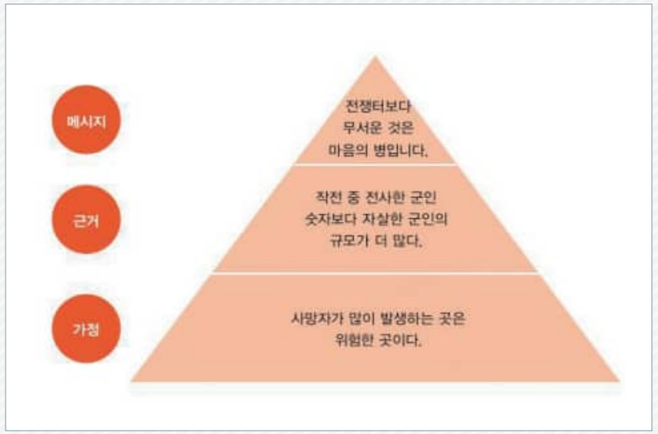
11

더욱 놀라운 것은
최근 그 극단적 선택을 한
군인의 규모가
작전 중 전사한 군인의
숫자를 앞질렀을
정도라는 것입니다.





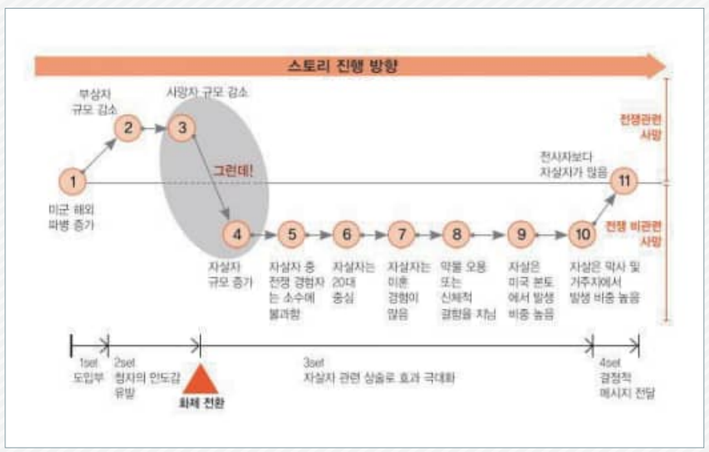
데이터로 줄거리를 만들어라



다수의 데이터가 연이어 활용되었다고 해서, 메시지 피라미드에 근본적인 변화를 가져오는 것은 아니다. 다만, 연관성이 있는 다수의 데이터를 동원해 사실성을 극대화하고 충분성을 확보하는 정도가 강화되었을 뿐이다. 즉, 메시지의 전달력이 높아지는 것이다.

이처럼 데이터들이 줄거리를 형성할 경우에는 메시지의 전달력이 극대화된다. 사실 미군 자살의 심각성은 열한 번째 데이터인 '미군 전사자와 자살자 규모의 역전' 하나로도 무리 없이 전달할 수 있다. 하지만 같은 메시지라도 좀더 드라마틱하게 전달하기 위해 나머지 데이터들이 나름의 목적을 가지고 포진된 것이다. 그럼 그 전달력이 어떻게 증폭될 수 있었는지 하나씩 뜯어 살펴보자.

이 데이터 스토리의 최종 메시지인 '전쟁보다 무서운 것은 마음의 병입니다'는 기본적으로 전쟁 관련 사망과 비(非)관련 사망을 대조시키는 형식을 띠고 있다. 이는 전쟁 비관련 사망 원인의 중요성을 극대화하기 위해 고안된 구조인데, 그도 그럴 것이 군인의 사망 원인 가운데 전사 이외 더 큰 원인이 있다는 사실은 우리의 상식과 적잖이 배치되기 때문에 듣는 이의 관심도를 순간적으로 높이게 된다.







데이터가 스토리를 만나면 강력한 힘이 생긴다



다시 말해서 극적으로 등장하는 열한 번째 데이터를 위해 사전에 여타 데이터들의 역할이 있었다는 것이다. 다른 데이터들은 역할별로 크게 네 덩어리로 나누어진다.

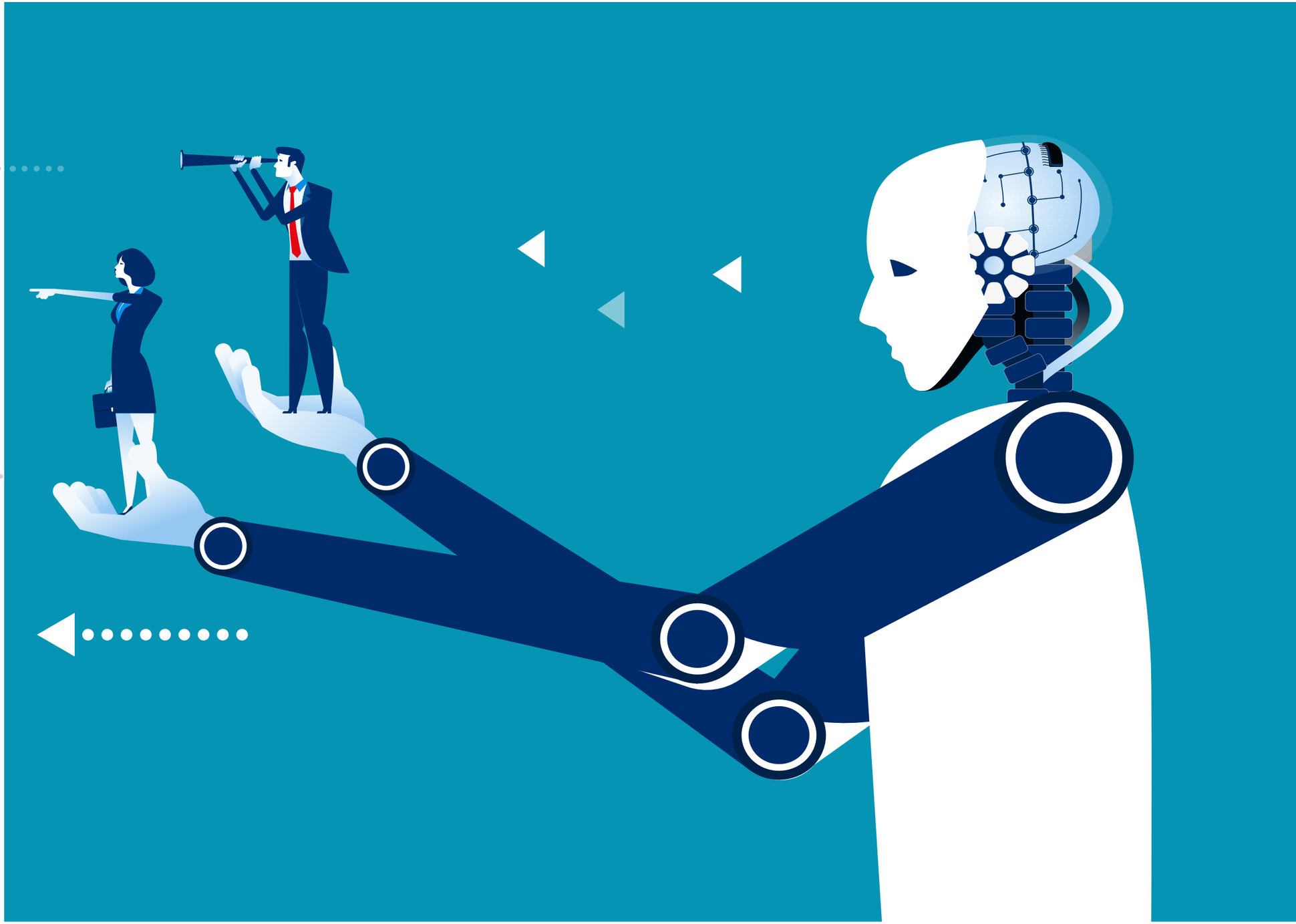
첫째 덩어리는 도입부에 해당하는 1번 데이터로, '미군, 파병, 증가'라는 키워드로 듣는 이의 관심을 환기한다.

두 번째는 전쟁 관련 사상자 규모를 설명하는 2, 3번 데이터로, 도입부의 미군 파병 증가 상황에도 불구하고 사상자 규모의 최근 감소세를 전달해 듣는 이의 안도감 형성에 기여한다. 하지만 곧바로 뒤에 등장하는 4번부터 10번 데이터는 그 안도감을 최대한 역이용한다. '마음의 병'이라는 더 무서운 사망 원인의 가능성을 표현하기 위해서 말이다. 즉, 2, 3번 데이터에 의해 형성된 안도감은 '마음의 병'이라는 새로운 사망 원인의 심각성을 극대화하기 위해 의도적으로 형성된 희생양인 셈이다. 마지막으로 11번 데이터는 그 두 원인을 직접 비교함으로써 설마 했던 듣는 이의 의구심을 강하게 확인시켜주며 강력한 메시지 전달이 완료되는 것이다.

만약 이 데이터들을 프레젠테이션 형식으로 전달한다면, 아마도 발표자는 3번과 4번 데이터 사이에서 듣는 이의 집중도를 최고조로 끌어올리고, 4번부터 10번까지의 데이터를 빠른 속도로 전개해 몰입도를 더욱 증가시킨 후, 극적으로 11번 데이터를 보여주며 메시지를 마무리했을 것이다. 이렇게 스토리를 잘 구성하면 듣는 이의 집중도 를 쥐락펴락할 수 있고, 집중도가 최고조로 유지됐다고 판단되는 그때, 메시지를 전달하면 된다. 이것이 바로 데이터가 스토리를 만났을 때 힘이 증폭되는 경우다.

유태준 ◆ 마인즈랩 대표

AI 서비스가 시작되는 AI 플랫폼에 주목하라



인공지능(AI)은 어느덧 시대의 화두가 되었다. 발 빠르게 AI를 도입한 금융권을 비롯해 산업 전반은 물론 공공 부문에서도 AI 기반의 서비스·솔루션 마련에 한창이다. AI 열풍 속에 일반인들의 인식 수준도 껑충 뛰었다. AI가 SF 영화에나 등장하는 로봇처럼 막연하게만 다루어졌던 이전과 달리, 이제는 대기업들이 연이어 출시하는 스피커 형태의 AI 기기, 각종 음성인식 및 챗봇 등으로 그 형태와 목적이 구체화되고 사용자의 일상에도 가까워지기 시작했다.

물론 지금까지 시중에 공개된 AI 서비스의 수준은 걸음마 단계에 불과하다. 하지만 클라우스 슈바프 다보스 포럼 회장이 지적했듯 우리가 살아왔고 일하던 방식을 AI가 근본적으로 뒤흔들 것이라는 데엔 의문의 여지가 없다. 인간의 개입이 최소화되거나 혹은 전혀 없더라도 빠르고 정확한 의사 결정과 실행이 가능해졌기 때문이다. 이 엄청난 규모와 복잡성을 지닌 영역의 핵심을 이해하기 위해서는 'AI 플랫폼'에 대한 이해가 필요하다.

왜 AI 플랫폼이 필요한가

인공지능 서비스를 가능하게 하는 여러 기술 요소를 한 곳에 통합한 뒤 이를 다양하게 조합하여 각 산업 영역에 적용할 수 있도록 구성된 것이 AI 플랫폼이다.

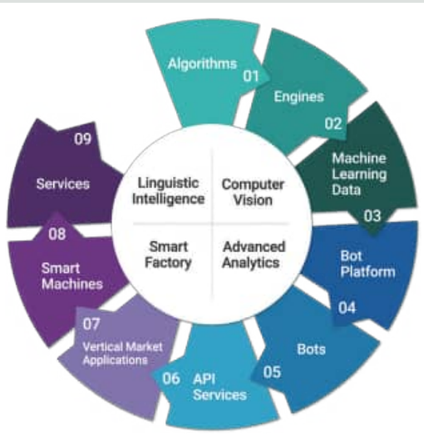
사람의 음성을 인식해 텍스트로 변환하거나(STT, Speech to Text) 텍스트를 다시 음성으로 변환하는 기술(TTS, Text to Speech), 언어를 이해하는 기술(NLU, Natural Language Understanding), 질의응답과 대화를 처리하는 기술 등이 대표적인 인공지능 기술 요소로, 이러한 기술 요소를 AI 엔진이라고 부른다. 이 AI 엔진들을 토대로 어떻게 모듈을 구성하느냐에 따라 AI 스피커, 챗봇, AI 가상 비서 등의 서비스를 구현할 수 있다.

그렇다면 왜 AI 플랫폼이 필요할까?

이 질문에 답하기 위해서는 AI 서비스 도입을 고민하는 대다수 기업과 기관들이 마주하는 문제점을 이해해야 한다. 대부분의 AI 서비스는 말 그대로 '예전에는 없었던 것'들이다. 챗봇이든, AI 스피커든, 가상 상담 서비스든 이 새로운 시스템을 구축하는 것 자체가 기술적으로 만만치 않다. 이처럼 자체적으로 AI 시스템을 만드는 것이 쉽지 않은 데다, 한 가지 AI 서비스만 필요한 것이 아니다 보니 여러 개의 시스템을 구축해야 하는 것도 문제다. 수많은 종류의 AI 서비스에 맞게 그때그때 시스템을 구축하는 것은 시간이나 비용 면에서 상당히 비효율적이기 때문이다.

또 AI는 계속해서 새로운 기술이 개발되는 영역인 만큼 최신 기술과 알고리즘을 어떻게 반영하느냐도 큰 숙제다. 엄청나게 빠른 속도로 최신 AI 알고리즘이 개발되고 있기 때문이다. 최신 기술 개발이 활발하게 이루어지는데도 이를 반영하지 못하면 결국 AI 서비스의 품질이 저하되는 결과로 이어진다. 하지만 개별 시스템마다 매번 최신 기술을 이해하고, 다시 이를 시스템에 적용하기 위해서는 엄청난 인적 · 물적 자원이 필요하다. 결국 AI 플랫폼은 가장 효율적으로, 가장 빠르게, 가장 최신의 AI 서비스를 구현할 수 있는 유일한 답인 셈이다.

[마인즈랩] AI Value Chain





이러한 이유로 AI 플랫폼은 전 세계 AI 기업들이 가장 집중하는 영역이기도 하다. '하나의 플랫폼에서 확장되는 수많은 서비스'라는 목적 아래, 개별적인 AI 애플리케이션보다는 AI 서비스를 총체적으로 제공하는 AI 플랫폼 개발에 힘쓴다.

IBM의 AI 플랫폼 왓슨(Watson)이 대표적으로, 어느 산업 분야에서든 그에 맞는 AI 서비스를 제공하는 것이 목적이다. 아마존의 알렉사(Alexa) 역시 가정용 AI 기기인 에코에 탑재되며 고객들에게는 쇼핑의 편의를 더하고 기업에게는 세일즈와 마케팅, 유통, 콘텐츠 네트워크를 위한 다각적 서비스를 제공하고 있다. 이처럼 전 세계 어느 산업에서나 다각적인 AI 서비스가 어떤 채널과 기기에서나 연속적으로 제공되기를 바란다. 그리고 이는 AI 플랫폼의 기술적 완성도에 달려있다.



아마존의 알렉사(Alexa) 역시 가정용 AI 기기

AI 플랫폼은 어떻게 활용되고 있을까

AI 플랫폼을 활용할 수 있는 영역은 무궁무진하지만, 기업·기관에 적용되는 사례는 크게 세 가지로 나누어 볼 수 있다.

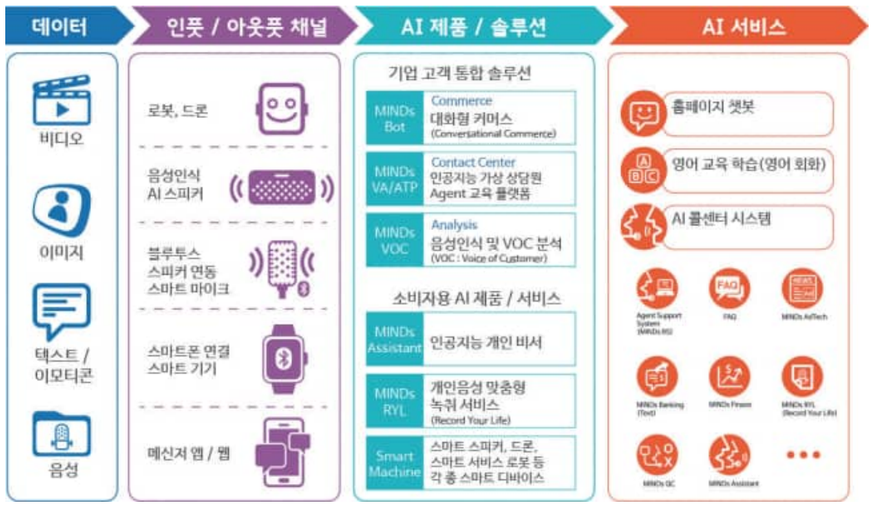
우선 단순 업무 지원(Smart Advisor)으로, 일종의 내부 직원용 서비스라고 이해하면 간편하다. 고객과 직원의 대화 중 주요 주제나 키워드를 인식해 AI가 최적의 답안을 추천해주면, 직원이 고객에게 직접 답변한다. 두 번째로 상담 업무 자동화(Virtual Agent)는 고객을 대상으로 24시간 내내 상담 서비스를 지원한다. 챗봇이 대표적으로, 실제 사람과 대화하듯 음성 대화나 채팅을 통해 상담 등 대화 기반의 서비스를 제공할 수 있게 된다. 마지막으로 업무 처리 자동화(AI Automation)는 사람이 수작업으로 진행하던 작업 일체를 AI를 활용해 말 그대로 자동화해버리는 것을 의미한다.

직접 고객을 응대해야 하는 대다수의 산업 분야에서는 상담 등 대고객 업무 자동화 과정에 관심을 둔다. 올해 금융권을 비롯해 숙박 예약, 배달 주문 등에서 챗봇이 쏟아져 나온 이유다.

최근에는 챗봇에 대한 공공기관의 관심도 높다. 대구광역시는 공공기관 최초의 AI 플랫폼 기반 민원 자동상담 시스템인 '뚜봇' 구축에 나섰다. 뚜봇은 여권, 차량 등록, 시정 안내 등의 분야에서 민원인의 답변에 24시간 답하는 챗봇이다. 꼭 채팅의 형태를 띤 챗봇이 아니고서도 언어를 이해하고 대화를 처리하는 AI 제반 기술을 활용하면 다양한 형태의 AI 기반 대고객 서비스가 가능해진다. 텍스트뱅킹은 계좌 이체·조회 등 간단한 금융 거래와 환율 조회·상품 추천 등의 금융 정보 제공이 앱 설치 없이 SMS 발송만으로 가능한 서비스다. 현재 '엄마 한테 10만원 보내', '오늘 엔화 환율 얼마야?'와 같은 말



[마인즈랩] 인공지능플랫폼 마음에이아이





을 즉시 이해하고 해당 명령을 실행하거나 답변할 수 있는 수준에 와 있다. 보험사나 카드사처럼 텔레마케팅이 활발한 분야나 콜센터에 접수된 민감한 민원을 즉시 처리해야 하는 제조사 등에서는 AI를 도입해 업무 처리를 자동화하고자하는 수요가 높다. 상담원이 표준 스크립트를 준수했는지, 고객의 요청 사항을 적절하게 처리했는지 등에 대한 검수가 필수적이기 때문이다.

AI가 없던 때에는 이 모든 과정을 수작업으로 일일이 녹취 파일을 2배속으로 청취하며 진행했다. 하지만 음성 인식과 텍스트 분석 기술 등 AI 기술을 기반으로 한 AI 플랫폼을 도입할 경우 해당 과정을 자동화하고 시각화할 수 있다. 이를 통해 업무의 효율과 검수 작업의 정확도가 획기적으로 개선된다.

AI 플랫폼을 넘어 AI 포털

AI 플랫폼은 이제 기술적으로 새로운 국면을 향해 나아가고 있다. 정보를 얻기 위한 사용자의 행동 양식에 근본적으로 변화를 줄 수 있도록, MRC(기계 독해, Machine Reading Comprehension)라는 최신 알고리즘이 개발되었기 때문이다. MRC란 인공지능이 텍스트를 독해하고 사람의 질문에 답할 수 있는 기술이다.

이전에는 정보를 얻기 위해 검색을 한 뒤 검색 포털이 제공하는 수많은 정보 중 연관성이 높아 보이는 것을 사용자가 직접 취사선택해야 했다면, MRC를 통해서는 그저 해당 정보를 AI에 곧장 물어보기만 하면 답을 얻을 수 있다.

또한 한국전자통신연구원에서 2차 개발을 진행 중인 AI인 엑소브레인(ExoBrain)과 같이 통계·법률·특허·금융 등 특정 전문 분야에서의 의사 결정과 판단을 지원하는 영역 역시 무궁무진한 발전 가능성이 잠재된 분야다.

이러한 기술적 발전을 뒷받침하기 위해서는 데이터의 확보가 무엇보다 중요하다. 어떤 상황에서든 수많은 질문에 답하고, 이를 바탕으로 정확한 의사결정을 즉각적으로 하기 위해서는 AI가 끊임없이 새로운 정보와 지식을 습득하고 이해하는 과정이 있어야 한다. 따라서 통계 데이터를 포함한 다양한 AI 학습 데이터의 축적이 필수적이다. 사람은 태어나서 어떤 말과 환경을 접하느냐에 따라 익히는 것이 달라진다. AI 역시 어떤 데이터를 접하고 학습하느냐에 따라 그 성능이 좌우된다. 다양한 분야의 학습 데이터가 절실한 이유다.

AI 플랫폼의 미래는 결국 '학습 데이터'에 달려있다. 특히 통계 자원과 같이 전문적인 영역의 데이터가 MRC를 포함한 인공지능 신기술들과 결합한다면 각종 전문 분야에서의 전문가 의사 결정을 지원하는 데서 엄청난 시너지 효과가 나올 것이다.

이렇게 AI 서비스가 더욱 폭넓게, 동시에 더욱 높은 성능으로 개선되려면 다양한 데이터에 대한 지원이 필요하다. AI 학습 데이터에 대한 국가적 관심이 더해진다면, 우리나라의 AI 산업은 결코 세계 시장에 뒤지지 않는 수준으로 발전할 것이라 확신한다. AI 플랫폼을 넘어 AI 포털로 진화하기 위해 다양한 AI 학습 데이터에 대한 국가적 관심이 환기되기를 바란다.

김덕식 ◆ 통계의창 객원기자

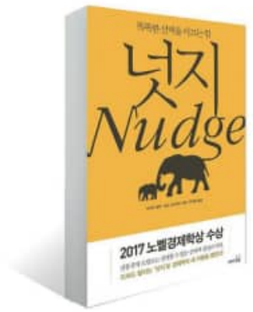
경제를 이해하려면, **통계에 사람을 먼저 포함하라**



"경제 행위자는 사람입니다. 경제 모델은 사람을 포함해야 한다는 인식이 가장 중요합니다." 2017년 10월 발표된 노벨경제학상의 영광을 차지한 리처드 H. 세일러(72) 미국 시카고대 교수는 수상 발표 직후 스웨덴 왕립과학원 노벨위원회와의 통화에서 이같이 밝혔다.

세일러 교수는 현실에 있는 심리적인 가정을 경제학적 의사결정 분석의 대상으로 통합하는 데 기여했다. 우리가 익히 들은 경제학은 합리적이고(rational), 계산적이며 이기적인 경제적 인간(homo economicus)이라는 가정에서 출발한 학문이다. 사람은 컴퓨터처럼 무제한의 계산을 실수 없이 해낸다고 믿고 있는 셈이다. 하지만 우리의 모습을 보면 이러한 가정에 의문이 생긴다. 경제현상은 사람의 심리를 떠나서 설명되기 어렵기 때문이다. 행동경제학은 기존 전통경제학에서 간과하고 있던 사람의 심리를 파고들었다.





**"경제 행위자는 사람입니다.
경제 모델은 사람을 포함해야 한다는 인식이 가장 중요합니다."**

– 리처드 H. 세일러(72) 미국 시카고대 교수

행동경제학이 주목받은 건 경제 주체의 비합리적·비이성적 의사 결정과 행동을 이해하는 도구로서 역할을 하면서다. 세일러 교수가 지난 2008년 쓴 베스트셀러 '넛지(Nudge)'에서 그 역할이 잘 설명돼 있다. '넛지'는 '팔꿈치로 가볍게 쿡쿡 찌르다'란 뜻이다. 규제나 강압보다 효율적인 힌트나 설계로 자율적 선택의 폭을 넓혀준다는 게 핵심이다. 가장 널리 알려진 것은 '소변기의 파리'다. 암스테르담 공항은 소변기에 파리 모양 스티커를 붙여놓는 아이디어만으로 소변기 아래 바닥에 떨어지는 소변량을 80%나 줄일 수 있었다고 이 책은 소개하고 있다.

세일러 교수는 노벨상 수상 직후 인터뷰에서 가장 성공적인 '넛지' 사례로 일부 국가에서 도입한 자동 연금 가입 제도를 꼽았다. 그는 "누구나 연금에 가입하도록 한 뒤 원치 않는 사람은 탈퇴할 수 있도록 정책을 설계했다. 그랬더니 연금이 필요하다는 것을 알면서도 당장 실행하지 않고 미루는 행동을 극복할 수 있다"고 설명했다.

그럼 왜 경제 현상을 분석할 때 사람을 먼저 생각해야하는지 하나씩 살펴보자.

휴리스틱 – 경제 활동은 심리에 의해서 움직인다

이 글을 읽고 있는 당신도 오늘 수많은 선택을 했을 것이다. 점심시간 메뉴를 결정할 때나 마트에서 과일을 고를 때를 생각해보자. 평소 먹던 것을 선택할 것인지 아니면 새로운 것에 도전할 것인가? 해당 품목이 다른 가게보다 저렴할까? 특별한 판촉행사가 있을까?

대부분의 사람들은 이러한 문제들에 많은 시간을 들이지 않는다. 상당수는 간단하게 늘 가던 곳에 가거나 보통 사던 과일을 선택한다. 이것이 바로 '휴리스틱(heuristic)'이다. 휴리스틱이란 시간이나 정보가 불충분해 합리적인 판단을 할 수 없거나 굳이 체계적이고 합리적인 판단을 할 필요가 없는 상황에서 신속하게 사용하는 어림짐작이다. 대부분 사람들의 경제적 활동은 휴리스틱에 기초한다. 그렇지 않다면 삶은 너무나 복잡해졌을 것이다. 사람들이 갖는 휴리스틱에는 대표성 휴리스틱이 있다. 사람들은 정보가 불완전한 상태에서 판단을 내리기 때문에 한쪽으로 치우친(bias) 잘못된 결정이 내려질 수 있는 위험을 갖고 있다. 대표성 휴리스틱이 이러한 특징을 지녔다. 대표성 휴리스틱은 어떤 사건이 전체를 대표한다고 보고 이를 통해 빈도와 확률을 판단하는 것을 의미한다. '하나를 보면 열을 안다'는 속담이 대표성 휴리스틱을 표현하는 좋은 예가 될 수 있다.

빅데이터 시대 더 각광받는 베이즈 통계

우리가 살다보면 불확실한 상황에서 선택을 해야만 하는 경우가 생긴다. 새로운 식당 선택처럼 비교적 간단한 결정 외에 이직이나 이사 등 삶 전체에 큰 영향을 미칠 선택까지 다양하다. 우리는 이러한 결정을 내릴 때 새로운 정보를 학습하려는 노력을 기울인다. 새로운 정보를 이용하면 불확실성을 조금이나마 낮출 수 있기 때문이다. 새로운 정보를 얻기 위해 TV, 라디오, 인터넷, 신문, 가족과 친구에게서 새로운 정보를 얻는다.

예를 들어 회사에 생긴 일로 맘이 상한 甲이 씁쓸한 마음을 달래고자 친구들과 술을 겸한 저녁식사를 계획했다. 그는 기분전환으로 오늘은 새로운 메뉴를 선택하려고 한다. 그는 친구들과 즐거운 시간을 보낼 수 있는 메뉴와 장소가 무엇인지 여러 채널을 통해 정보를 모을 것이다. 그는 새로운 정보를 어떻게 활용할까? 이 과정에서 우리는 통계의 힘을 빌린다. 해답의 실마리는 '베이즈 정리(Bayes' theorem)'다.

베이즈 정리는 영국의 수학자 겸 목사 토머스 베이즈(1701~1761)가 발견하고, 훗날 프랑스의 수학자이자 천문학자인 피에르 시몽 라플라스가 완성했다. 베이즈 정리가 다루는 확률은 주관적이다. 객관적인 수치가 아니라 인간의 심리에 의존한 것이기 때문이다. 베이즈 통계가 사상적인 면을 갖추고 있다고 평가받는 이유이기도 하다.

동전을 던져 앞면이 나올 확률이 50%지만, 찌그러진 동전을 던졌을 때는 확률이 달라진다. 이럴 경우 공장에서 찌그러진 동전이 생산될 확률을 구해 추가로 반영해야 한다. 이처럼 새로운 상황과 정보가 나오면 가장 가능성이 있는 것을 골라 적용해 확률을 수정해 나가는 것이 바로 베이즈 정리다. 이런 의미에서 베이즈 통계는 조건부 확률이라는 말로 표현되기도 한다.

조건부 확률에서는 새로운 정보를 알았을 때 확률의 개선이 일어나게 된다. 우리는 어떤 실험결과에서 나온 정보를 이용해 어떤 사건의 처음 확률을 개선시킬 수 있다. 여기서 처음 확률은 사전확률(prior probability)이라 하고, 개선된 확률을 사후확률(posterior probability) 이라고 한다. 甲은 줄어든 호주머니에서 기분 좋은 저녁식사 시간을 위해 새로운 정보를 계속 수집해 나가고, 새로운 정보에 근거해 저녁식사 당일 최종적인 의사결정을 내릴 것이다.



베이즈 이론이 이처럼 '과거 데이터를 기반으로 미래를 예측한다'는 특징 덕분에 데이터가 많을수록 확률도 정확해진다. 데이터가 바뀌면 예측도 저절로 수정된다. 컴퓨터의 논리구조와 유사하다. 초기에는 엉성했던 번역 프로그램들이 갈수록 정확해지는 이유가 여기에 있다. 이런 점에서 빅데이터 시대를 맞은 지금 베이즈 정리가 더욱 주목을 받는 원인이 될 수 있다.

베이즈 이론과 행동경제학의 발달과정에는 유사점이 있다. 기존 굳건한 학계의 비판을 극복했다는 점이다. 주관적인 믿음을 측정하기 때문에 합리적이지 않다는 이유에서 베이즈 정리는 오랫동안 금기시되기도 했다. 베이즈 정리를 언급하면 대학에서 자리를 얻을 수 없을 정도였다고 전해질 정도다. 하지만 최근 컴퓨터의 계산 능력이 급증하고, 더불어 사후확률의 계산을 몬테카를로 방법 혹은 변분추론과 같은 기법으로 해결할 수 있다는 것이 밝혀지면서, 베이즈 정리가 점차 확산됐다.



행동경제학 역시 주류 경제학과는 기본적인 가정이 달라 오랜 기간 설움을 겪었다. 세일러 교수의 올해 노벨상 수상은 곧 주류 경제학계로부터 이단 취급을 받아온 행동경제학의 쾌거로 평가받고 있다. 인간의 비합리성을 강조하는 행동경제학자들의 주장은 주류 경제학의 근간을 흔드는 것이기 때문에 쉽게 받아들여지지 않았기 때문이다. 주관적인 믿음을 측정하는 베이즈 정리와 경제학이 만나 사람의 심리적 특성을 활용하는 행동경제학이라는 꽃을 피우고 있다. 공통의 역사적 분모를 가진 두 분야가 만난 셈이다.

심리적 회계 – 운수 좋은 날 사람들의 선택은

세일러 교수가 개발한 이론으로 대표적인 것이 '심리적 회계(mental accounting)'다. 이를 통해 인간심리를 분석해 개인이 개별적으로 내리는 결정에 집중했고, 어떻게 재무적인 결론으로 이어지는지 설명했다. 비오는 날 택시 이용자가 늘기 때문에 택시가 더 잘 잡혀야 하지만 택시 운전사들은 하루 매상을 다 올렸다고 생각해 먼저 퇴근하면서 택시가 잡히지 않는다는 결론이다. 만약 택시 운전사들이 합리적으로 행동한다면 손님이 많은 날 운행을 더 많이 하고 맑은 날 쉬는 것이 옳지만 그렇게 행동하지 않을 수 있다는 설명이다.

사회적 선호현상 – 인간은 이기적인가 이타적인가

세일러 교수는 또 인간의 사회적 선호현상을 강조했다. 인간이 이기적인 본능에 따라 행동하는 것에 그치지 않고 함께 잘 살기 위해 이타적 행동을 할 때도 있다는 의미다. 세일러 교수가 심리학자 등과 함께 고안한 '독재자 게임'에서 관찰할 수 있다. 독재자 게임은 두 사람을 짝을 지어주고 돈을 준 뒤 한 사람은 분배자로, 한 사람은 수령자로 나눠 진행된다. 수령자는 돈을 나눈 결과에 어떤 저항도 할 수 없고 그대로 수용해야 한다. 주류경제학의 가정에 따라 인간이 이기적으로 행동한다면 분배자는 모든 돈을 독차지해야 한다. 하지만 이 실험을 수행한 결과 분배자들은 전체 돈의 25%가량을 수령자에게 줬다. 만약 분배자가 돈을 독점한다면 인간적 유대와 같은 사회적 경향이 무너지기 때문에 수령자에게 일부 떼어준다는 것이다.

세일러 교수의 연구는 지난 2008년 글로벌 금융위기 과정에서 부각했다. 인간의 합리성이 반드시 전체의 후생복리를 높이는 방향으로 귀결되지 않는다는 반성이 확산된 분위기 속에서 세일러 교수의 연구가 주목받은 것이다. 2008년 금융위기 당시 벌어졌던 어이없는 상황들이 바로 세일러 교수의 전공 분야였던 것이다. 금융위기 이후 그는 미국인들의 저축과 연금 플랜을 다듬어 빚더미에 앉은 미국을 구한 경제학자로 평가받기도 한다.

행동재무론 – 인간은 바보같은 행동을 하는가

행동경제학은 자본시장 연구에도 많은 영향을 미쳤다. 합리적 투자자에 기반을 둔 자본시장의 효율성이 1980년대 실증연구에서 다양한 시장 이상현상(market anomalies)들이 발견되면서 그 효용성에 의문이 들기 시작했다. 이를 해결하기 위해 현실적인 차원에서 투자자 심리의 비합리성을 고려한 행동재무론(behavioral finance)이 태동하기 시작했다. 투자자들을 비합리적으로 이끄는 주요 심리적 요인들로는 앞서 언급한 심리적 회계를 비롯해 과신(overconfidence), 현상유지(status quo) 본능, 정박효과(anchoring effect)와 군집행동(herding behavior) 등이 있다.

연구결과에 따르면 사람들은 자신의 능력을 지나치게 과신한다. 세일러 교수는 일반 투자자들이 가장 많이 하는 실수로 '자기 과신'을 꼽았다. 그는 "만약 스스로 잘 나가는 투자자라고 생각한다면, 자신의 실제 수익률이 얼마인지 계산해보라"면서 "대부분의 투자자는 거래수수료를 떼어내면 시장수익률에도 미치지 못한다"고 말했다.

합리적 인간이라면 좀 번거롭더라도 자기에게 유리한 쪽을 택한다. 하지만 현실에선 이미 익숙한 일을 바꾸려 하지 않는다. 이런 인간의 특징을 '현상유지편향'이라고 한다. 이는 익숙한 대상에 안주하려는 경향으로 주로 그 대상으로 포기함에 따른 아쉬움이나 두려움, 고통을 피하고자하는 본능에서 나온다. 투자자들은 다양하고 화려한 투자결과를 원하면서도 한 두 개의 주식종목에만 집착하게 되는 결과를 낳는다.

정박효과는 또 다른 노벨경제학상 수상자 대니얼 카너먼이 규정한 개념이다. 투자자가 숫자(기업매출액 등)에 꽂히면, 집착하기 시작한다는 것이다. 이어 새로운 정보를 접할 때 기존 수치를 조정하지만, 이미 애초 숫자에 집착한 뒤라 조정 폭이 충분히 크지 않다는 것이다. 과거에 알고 있던 지식을 고집하고, 새로운 정보를 쉽게 수용하지 않으려 한다. 특히 노년층으로 갈수록 정박효과는 강하게 나타난다.

군집행동은 한 사람의 행위가 다른 많은 사람에게 영향을 미치는 투자행위를 말한다. 무리에서 뒤처지는 걸 싫어해 다른 이들을 따라 하는 과정에서 나타난다. 투자자들이 공황 상태에 빠져 투매할 때는 물론 안전하게 다수의 행동에 편승할 때도 나타난다. 가격이 상승하면 우르르 사고, 하락해도 우르르 파는 현상이 벌어져 가격이 극단적인 모습으로 형성된다. 상승국면에선 언젠가 더 높은 가격을 부담하고 물량을 매수하고자 하는 세력이 나타날 거라는 기대로 매수에 나서고, 하락국면이면 가격이 바닥까지 떨어져 더는 하락할 공간이 없을 때까지 매도하는 원리다.







시장 이상현상 가운데 대표적인 사례가 '캘린더 효과'다. 캘린더 효과란 공휴일 다음날이나 주말을 쉬고 난 다음날의 시장 평균수익률이 다른 날의 수익률보다 높다는 것이다. 많은 전문가들이 과거 데이터를 가지고 휴일 다음날과 평일 사이에 수익률의 차이가 존재하는지를 통계적으로 분석해 보니 휴일을 보낸 다음날의 주가들이 유의미하게 높은 것으로 나타났다. 주식시장이 전통적인 경제이론처럼 효율적이라면 특정 요일과는 상관없이 유의미한 수익률 차이가 없어야 한다. 학자들은 캘린더 효과를 많은 투자자들이 합리적인 근거 없이 휴일이 지난 다음날 주식이 오를 것이라는 낙관적인 기대를 하는 성향이 크다는 데서 원인을 찾았다.

대표적인 캘린더 효과로는 1월 효과, 서머랠리(summer rally), 산타랠리(santa rally)와 추수감사절을 전후한 미국 증시의 오름세 등이다. 1월 효과는 연말·연초에 집중된 보너스들이 증시로 흘러들고, 새해를 맞아 주식 분석가들이 낙관적인 전망을 내놓으면서 1월의 주가 상승률이 다른 달에 비해 상대적으로 높게 나타나는 현상을 말한다. 서머랠리는 펀드매니저들이 긴 휴가를 떠나기 전에 가을 장세를 기대하고 미리 주식을 사 놓기 때문에 휴가를 앞둔 초여름에 주가의 단기 급등 현상이 나타나는 것을 말한다. 산타랠리는 연말에 소비 증가세가 나타나면서 연말장 종료 5일 전부터 이듬해 2일까지 증시가 강세를 보이는 현상이다. 또 미국에서는 전통적으로 추수감사절(11월 넷째 목요일)을 전후해 증시가 오른다. 이처럼 월별 주기에 따른 캘린더 효과 외에 '주말효과', '월요일 효과' 등으로 기간을 더욱 세분화해 증시 상황을 설명하기도 한다.

심송용 ◆ 한림대학교 금융정보통계학과 교수

나도 R유저 ②

R에 도전하자… 따라가다보면, **나도 R유저②**



R에 도전하자 ①에서 R의 설치부터 시작하여 몇 가지 기능을 따라 해보고 R을 끝내는 과정을 보았다. 이번엔 R에서 설정할 수 있는 색깔에 대해서 알아보기로 하자.

1 사용되는 몇 가지 함수의 소개

막대그래프

색깔의 출력을 확인하기 위해 막대그래프를 그려서 막대의 색을 바꾸는 방법으로 R의 색깔 설정을 확인하기로 하자. 막대그래프는 barplot 함수를 사용하여 그리며 일반적인 사용법은

```
barplot(height, space = NULL, names.arg = NULL, col = NULL, axes = TRUE,...)
```

로

- **height** 막대의 높이를 저장한 벡터 또는 행렬
- **space** 막대사이의 공간.
- **names.arg** 막대그래프의 막대 이름 설정
- **col** 막대의 색깔 설정
- **xlab, ylab** x축, 축의 이름
- **axes** TRUE/FALSE로 설정하며 빈도수에 해당하는 축을 그릴지 설정한다.



rep 함수와 반복

rep 함수는 R 개체를 반복하여 생성하게 하는 함수로

```
rep(x, times, ...)
```

로 사용하여 x를 times 만큼 반복하여 생성한다. 아래는 1을 다섯 번 반복 생성한다.

```
> rep(1, 5)
[1] 1 1 1 1 1
```

:(콜론)으로 1씩 증가하기

:(콜론)은 m:n과 같이 사용하며 m, m+1, m+2, …, n의 값을 얻는다. 일반적으로 m과 n은 정수를 사용하나 소숫점이 사용될 수도 있다.

나도 R유저 ②

```
> 0:10
[1] 0 1 2 3 4 5 6 7 8 9 10
```



등차수열과 seq 함수

등차수열을 만들 수 있는 함수로 seq 함수가 있으며 이 함수는

```
seq(from = 1, to = 1, by,  length.out, ...)
```

로 사용하여 from부터 to까지 by만큼 증가하는 수열을 생성한다. by 대신에 length를 설정하면 from부터 to까지의 length의 수 만큼의 등간격의 숫자를 생성한다.

```
> seq(1,10, by=2)
[1] 1 3 5 7 9
> seq(1,10, length=10)
[1] 1 2 3 4 5 6 7 8 9 10
```

length 함수와 원소의 개수

벡터를 매개변수로 length 함수를 호출하면 원소의 개수를 얻는다. 아래는 x에 1부터 10까지의 자연수를 저장하고 length 함수를 호출한 결과이며 원소의 개수가 10이므로 10을 얻는다.

```
> x <- 1:10
> length(x)
[1] 10
```

as.character 함수

숫자 등을 문자열 형식으로 바꾸고자 할 때 사용하는 함수이다. 이 함수의 결과는 문자열이므로 산술연산에는 사용하지 못한다.

```
> x <- 1:10
> as.character(x)
[1] "1"  "2"  "3"  "4"  "5"  "6"  "7"  "8"  "9"  "10"
> sum(as.character(x)) # 숫자가 아니므로 합이 계산 안됨 → 에러
```

2 색깔 설정의 방법

R에서 색깔 설정은 여러 가지 방법으로 할 수 있다. 다른 패키지를 추가하지 않고 R에 내장된 기본 함수만 사용하는 방법에 대해서 살펴보기로 하며, viridis, RColorBrewer 등의 패키지를 사용하면 좀더 확장된 방법으로 색을 정의할 수 있다.

나도 R유저 ②

색깔번호로

R에서는 0부터 8사이의 숫자로 색깔을 설정할 수 있으며 이 색깔은 0부터 8까지 순서대로 흰색(white), 검정, 빨강(red), 녹색(green), 파랑(blue), 시안(cyan), 마젠타(magenta), 노랑(yellow), 회색(grey 또는 gray)이다. 0번 흰색의 경우 엄밀하게는 흰색이 아니고 R 그래픽 설정에서 배경색(background color)이며 R 그래픽의 기본 설정이 배경색이 흰색이므로 대부분 흰색이 된다. 이 0부터 8번까지의 위의 아홉가지 색깔은 기본 파레트가 적용되었을 때의 색이며 R에서는 파레트 개념을 사용하여 색을 미리 정의할 수 있으며 파레트를 따로 정의하면 번호에 따른 색깔은 달라질 수 있다. 파레트는 이 원고의 뒤 부분에서 소개하기로 한다.

```
> barplot(rep(1,9), space=0, axes=F, names.arg=as.character(0:8), col=0:8)
```



색깔이름으로 설정하기

R언어의 색은 색깔의 이름으로 설정할 수도 있다. 색깔의 이름은 반드시 큰따옴표 또는 작은 따옴표를 사용하여 표현한다.

```
> barplot(rep(1,7), space=0, axes=F,
    col=c("red", "blue", "green", "cyan", "magenta", "yellow", "aquamarine"))
```





R에서 사용할 수 있는 색깔의 이름은 657개가 있으며 이 이름들은

```
> colors( )
```

명령으로 확인할 수 있다. 657개 중 앞의 10개와 끝의 10개의 이름은 다음과 같다.

```
> head(colors( ), 10)
[ 1 ] "white" "aliceblue" "antiquewhite" "antiquewhite1" "antiquewhite2"
[ 6 ] "antiquewhite3" "antiquewhite4" "aquamarine" "aquamarine1"  "aquamarine2"
> tail(colors( ), 10)
[ 1 ] "wheat2" "wheat3" "wheat4" "whitesmoke" "yellow"
[ 6 ] "yellow1" "yellow2" "yellow3" "yellow4" "yellowgreen"
```

빛의 삼원색과 rgb 함수

빛의 삼원색은 빨강(red), 녹색(green), 파랑(blue)이며 보통 RGB라고 부른다. R에서 이 세 가지 원색

나도 R유저 ②

의 농도를 설정하여 색을 만들 수 있으며 이를 위해 rgb 함수를 사용한다. 이 함수는

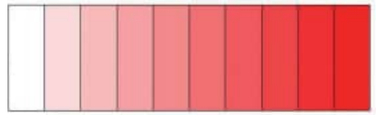
```
rgb(red, green, blue, alpha,...)
```

로 설정하며 red, green, blue에는 RGB의 농도를 설정한다. 예를 들어 rgb(1,0,0)은 red 만 최고 농도이고 다른 색은 없으므로 빨강, rgb(1,1,1)은 모든 색을 최고 농도로 사용하므로 흰색, rgb(0,0,0)은 모든 빛을 사용하지 않으므로 검정색이 된다. 다음 그림은 RGB의 기본색과 이들 중 두 색이 100%씩 사용되었을 때의 색깔(Y=yellowm M=magenta, C=cyan)을 그려본 것이다. 세 가지 모두 최대의 농도를 사용하면 흰색(W=white)이다.

alpha에는 투명도(opacity; 엄밀하게는 불투명도)를 설정하며 1은 불투명, 0은 투명이다.

다음은 기본 배경색인 흰색에 대해서 빨강의 농도만 최대로 나머지는 두 색은 0으로 설정하고 alpha값을 0부터 1까지 0.1씩 증가한 경우의 색을 확인하였다.

```
> mycol0 <- rgb(1, 0, 0, seq(0, 1, length=10))
> barplot(rep(1,10), space=0, axes=F, col=mycol0)
```



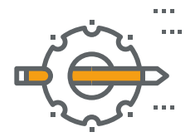


RGB 모델

투명도는 이전 색이 일부분 나타나는 경우이므로 배경색이 다르면 다른 색이 나타난다. 다음은 배경색은 검정, 글자색은 흰색으로 변경한 경우 투명도 설정에 따른 색의 표현이다. 이번에는 배경색이 검정이므로 가장 오른쪽의 색을 제외하고 나머지의 색은 위와 같이 mycol0로 색을 설정함에도 배경이 흰색일 때와 다른 색이 나타난다.

```
> win.graph( )
> par(bg="black", fg="white")
> barplot(rep(1,10), space=0, axes=F, col=mycol0)
```

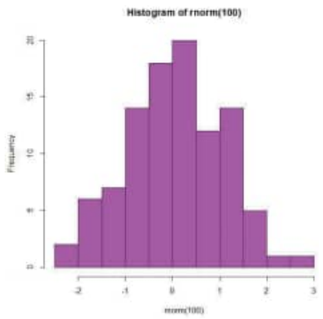




HTML 형식의 색깔 설정

Red, Green, Blue의 삼원색의 농도를 0-255사이의 16진수 (00-FF)의 값으로 원하는 색을 설정할 수도 있다. 이 색깔 설정은 여섯 개의 문자로 이루어지며 첫 두 문자는 빨강, 두 번째는 녹색, 세 번째는 파랑의 농도를 16진수로 표시한다. 예를 들어 빨간색의 경우 col="#FF0000"과 같이 설정한다. FF0000의 경우 빨간색은 FF(십진수 255), Green 및 Blue는 모두 0이므로 빨간색이다. 이와 같은 방법으로 설정할 경우 00FF00은 녹색(Green), 0000FF는 파랑(Blue), FFFFFF는 흰색, 000000는 검정색 등이다.

555555와 같이 이 삼원색의 모두 같은 값인 경우 회색(검정 및 흰색 포함)이 된다. 16진수이므로 반드시 #로 시작하고 큰 따옴표 또는 작은 따옴표를 사용하여야 한다. 예를 들어 아래는 히스토그램에서 FF00FF(=cyan) 색깔로 기둥을 그린 것이다. hist와 rnorm 함수는 R에 도전하자 ①에서 간단히 설명한 바 있다.




```
> hist(rnorm(100), col="#FF00FF")
```

색상, 명도, 채도와 hsv 함수

색깔을 설정하는 또다른 방법으로 색상(hue), 명도(value/brightness), 채도(saturation)를 설정하는 방법이 있다. 이 방법은 R에서 hsv 함수를 사용하여 설정한다. hsv 함수는

```
hsv(h = 1, s = 1, v = 1, alpha)
```

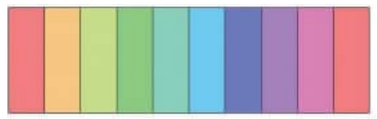
로 색상(h), 채도(s), 명도(v) 와 투명도를 0과 1사이의 값으로 설정한다. 색상을 0부터 1까지 등간격으로 설정한 결과는 다음과 같다.

```
> mycol1 <- hsv(seq(0,1, length=10), s=1, v=1, alpha=1)
> barplot(rep(1,10), space=0, axes=F, col=mycol1)
```



같은 색상에서 채도를 0.5로 낮추면 다음과 파스텔 계열의 색을 얻는다.

```
> mycol2 <- hsv(seq(0,1, length=10), s=0.5, v=1, alpha=1)
> barplot(rep(1,10), space=0, axes=F, col=mycol2)
```



밝기(value)를 0.5로 낮추면 어두운 색이 표현된다.

```
> mycol3 <- hsv(seq(0,1, length=10), s=1, v=0.5, alpha=1)
> barplot(rep(1,10), space=0, axes=F, col=mycol3)
```





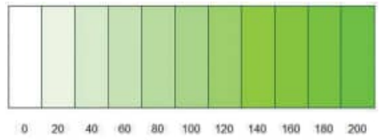
HCL 색깔 모형

HSV 색깔모형에서 색의 종류에 따라 채도가 다르게 느껴진다는 점을 개선하기 위한 색깔 모형으로 HSV와 다른 형식의 색상(Hue), 채도(Chroma) 명도(Luminance)를 설정하는 색깔 모형이다. 이 모형은 R언어에서 hcl 함수로 사용할 수 있으며

```
hcl(h = 0, c = 35, l = 85, alpha, ...)
```

h는 색상을 설정하며 0부터 360사이의 값을 설정한다. 양 끝점 포함하며 원의 둘레의 각도라고 생각하면 된다. 0은 빨강, 120은 green, 240은 blue이다. 360은 다시 빨강이다.
c는 양숫값으로 설정할 수 있는데 최대로 설정할 수 있는 값이 정해져 있지 않다. 이 값이 클수록 채도가 높아진다. l은 0부터 100사이의 값을 양 끝점 포함하여 설정하며 밝기를 설정한다.
alpha는 투명도의 설정으로 0은 투명, 1은 불투명으로 0과 1사이의 실수로 설정한다.
다음은 hcl 함수에서 색상이 green이고(h=120) 밝기가 최대(l=100)일 때 채도의 값을 0부터 200까지 바꾸어 색을 얻어 본 결과이다.

```
color.den <- seq(0, 200, length=11)
barplot(rep(1,length(color.den)), space=0, axes=F,
        names.arg=as.character(color.den), col=hcl(120, color.den, 100))
```





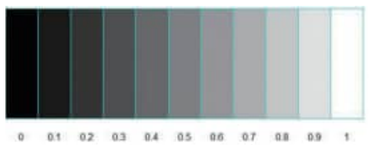
무채색과 grey/gray 함수

회색(흰색, 검정색 포함)을 grey 또는 gray 함수로 만들 수 있다. 이 함수는

```
gray(level, alpha = NULL) 또는 grey(level, alpha = NULL)
```

로 level은 0과 1사이의 숫자로 설정하며 0은 검정색, 1은 흰색이며 그 사이의 값은 회색이 된다. alpha는 투명도를 설정한다.

```
> par(fg="cyan")
> color.den <- seq(0,1, by=0.1)
> barplot(rep(1,length(color.den)), space=0, axes=F,
        names.arg=as.character(color.den), col=grey(color.den))
```



무채색(검정부터 흰색사이의 회색)은 rgb 함수의 red, green, blue의 값을 모두 같게 하여 얻을 수도 있다.

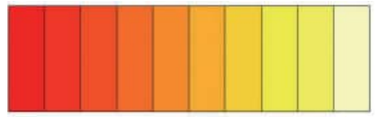
특정 규칙의 색깔함수

특정한 규칙을 가지는 색깔함수로 rainbow(무지개색), cm.colors(cyan에서 magenta 사이의 색 농도 변환), topo.colors(지도에서 바다와 육지의 깊이와 고도를 표시하는 색의 변화), heat.colors(열을 표시하는 색깔변화), terrain.colors(지도에서 육지의 고도를 표시하는 색의 변화) 함수가 있으며 이들은 자연수 n을 사용할 때 n개의 색을 해당 규칙에서 생성해준다. n외에 투명도 등의 매개변수를 추가로 설정할 수 있으므로 help 명령으로 확인해보기로 한다.

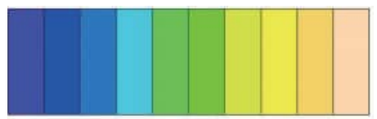
```
> barplot(rep(1,10), space=0, axes=F, col=rainbow(10))
```



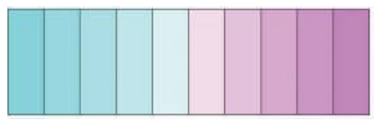
```
> barplot(rep(1,10), space=0, axes=F, col=heat.colors(10))
```



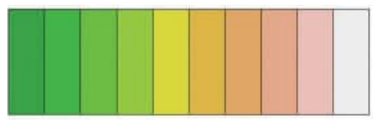
```
> barplot(rep(1,10), space=0, axes=F, col=topo.colors(10))
```



```
> barplot(rep(1,10), space=0, axes=F, col=cm.colors(10))
```



```
> barplot(rep(1,10), space=0, axes=F, col=terrain.colors(10))
```





palette 함수를 사용한 색의 설정

R은 사용자가 색을 정의할 수 있도록 palette 함수를 제공한다. 이 함수를 사용하여 색을 정하면 이 후부터 사용자가 정한 색을 번호로 정할 수 있다. 따로 정하지 않으며 0부터 8까지 설정할 수 있음은 앞에서 확인하였다. 파레트를 다르게 정의하더라도 0번 색은 배경색(기본값: 흰색)이다.
다음은 palette 명령없이, 즉 기본 파레트를 사용한 경우의 색이다. 기본 파레트인 경우 색의 종류는 배경색 포함 9개인데 기둥은 12개이므로 10번째 기둥(아래의 9번)부터는 다시 1번 색깔부터 반복됨을 확인할 수 있다.

```
> barplot(rep(1,12), space=0, axes=F, names.arg=as.character(0:11), col=0:11)
```





이번엔 cm.colors 함수에 의한 11개의 색을 파레트로 설정해보자. 파레트 설정은

```
> palette(cm.colors(11))
```

로 하고, 앞에서와 똑 같은 명령을 실행할 때 얻어지는 색은 cm.colors의 색깔임을 알 수 있다. 아래의 경우 palette함수의 cm.colors에 11개의 색이 얻어지므로 1번 색깔부터 다시 시작하지 않는다.

```
> barplot(rep(1,12), space=0, axes=F, names=as.character(0:11), col=0:11)
```



현재의 파레트의 색깔 번호별 색깔을 확인하려면 매개변수 없이 palette 함수를 호출한다. 위의 cm.colors를 사용한 palette 함수가 유효한 상태에서 palette 함수를 호출하면

```
> palette( )
[ 1 ] "#80FFFF" "#99FFFF" "#B3FFFF" "#CCFFFF" "#E6FFFF" "white"   "#FFE6FF"
[ 8 ] "#FFCCFF" "#FFB3FF" "#FF99FF" "#FF80FF"
```

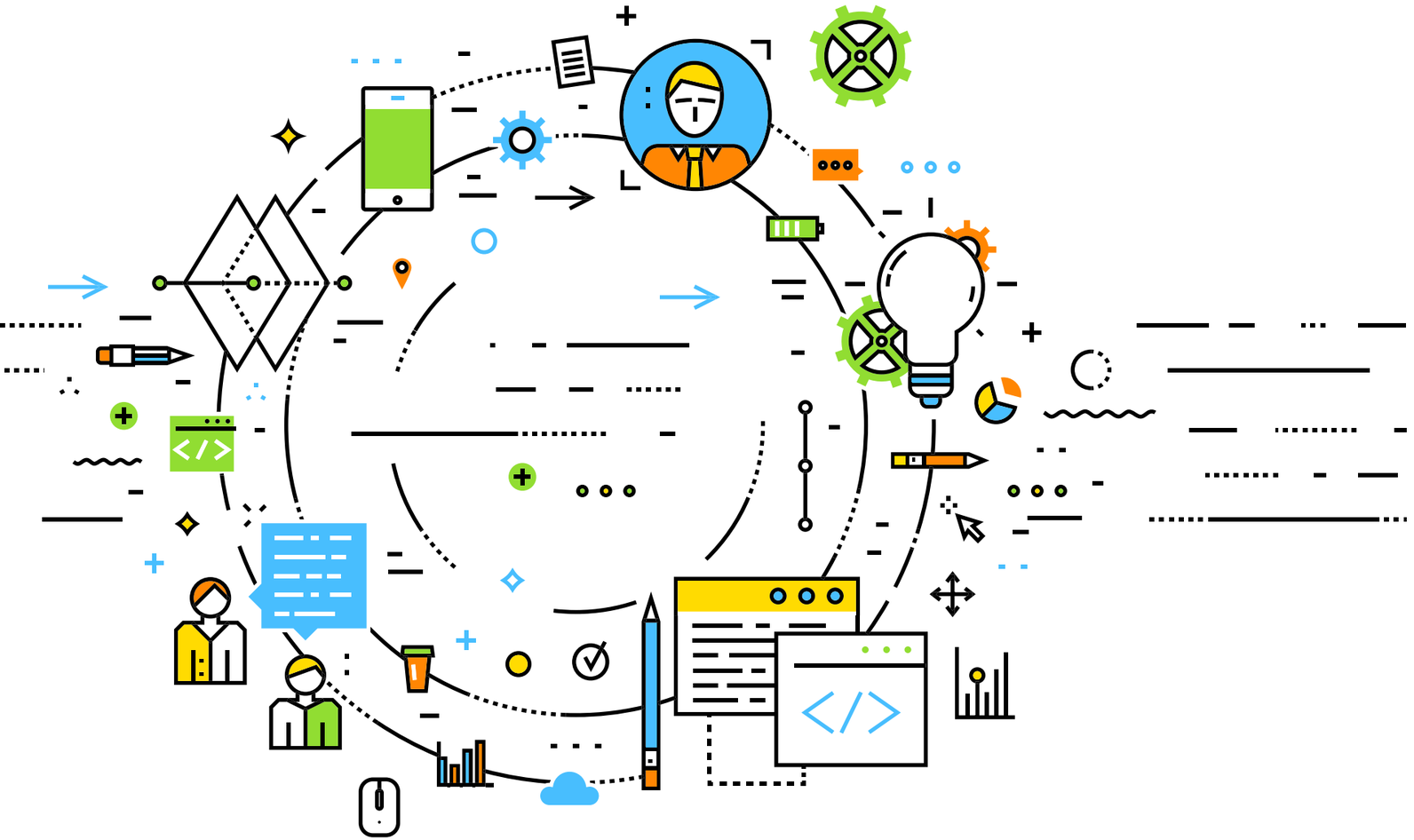
로 11개 색의 16진수 값을 얻을 수 있다.

palette 명령은 한 번 설정하면 다른 palette 명령이 나올 때까지 계속 유효하다. 파레트를 R의 기본 파레트로 바꾸려면 다음과 같이 매개변수에 "default"를 사용한다.

```
> palette("default")
```

송정수 ◆ 인포그래픽웍스 대표

효과적으로
정보를 전달하는 방법



Infographics

숫자 정보를 효과적으로 전달 하는 방법

다양한 통계데이터를 다루고 있는 우리들은 이를 다양한 방법으로 표현하여 전달하고 있다. 그러나 과연 '전달'의 측면에서 제대로 그 역할을 하고 있는 것일까? 시작에 앞서 하나의 이미지를 함께 보았으면 한다. 다음 그림을 보고 어떠한 생각이 드는가? 어떠한 데이터를 나타낸 것일까?
이 질문에 대한 대답은 뻔하다. '아무런 생각이 들 지 않는다'와 '무슨 데이터인지 모르겠다'일 것이다. 그렇다. 우리는 이 이미지를 가지고 '정보'라고 이야기 하지 않는다. 우리에게 어떠한 '정보'도 전달하고 있지 않은 숫자가 나열되어 있는 알 수 없는 그림일 뿐이다.

```
 4.35 3.17 3.06 1.37 0.19 0.11 0.03 0.05 0.20 1.22 2.86 3.09
 5.35 4.03 3.77 2.51 1.84 1.59 0.85 1.22 1.94 3.25 5.65 6.00
 1.53 1.36 2.69 3.64 3.32 3.78 3.66 4.22 3.82 2.41 2.92 2.47
 3.17 3.02 3.59 3.90 3.80 3.65 3.80 3.41 3.30 2.88 3.65 3.42
 2.01 2.08 2.39 2.85 6.21 9.33 5.70 7.58 7.63 5.64 2.66 1.83
```

그렇다면 '정보'는 무엇인가? 정보는 정보를 받는 사람이 무엇을 나타내고 있으며 나에게 전달하고자 하는 바가 무엇인지 알 수 있어야 한다. 보다 정확히 말하면 그 것을 활용할 수 있어야만 비로소 정보로서의 가치를 가진다. 그 속에서 있는 다양한 'Insight'를 통해 어떠한 현상이나 문제점 등을 직관적으로 발견할 수 있어야 '가치가 있는 정보'라고 이야기 할 수 있다. 알 수 없었던 숫자들은 사실 아래의 도표에서 추출해낸 숫자들이다. 아래 도표는 샌프란시스코를 포함한 미국의 주요 도시들의 월평균 강수량을 나타내고 있다.

AVERAGE RAINFALL (INCHES/MONTH)

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
San Fran	4.35	3.17	3.06	1.37	0.19	0.11	0.03	0.05	0.20	1.22	2.86	3.09
Seattle	5.35	4.03	3.77	2.51	1.84	1.59	0.85	1.22	1.94	3.25	5.65	6.00
Chicago	1.53	1.36	2.69	3.64	3.32	3.78	3.66	4.22	3.82	2.41	2.92	2.47
New York	3.17	3.02	3.59	3.90	3.80	3.65	3.80	3.41	3.30	2.88	3.65	3.42
Miami	2.01	2.08	2.39	2.85	6.21	9.33	5.70	7.58	7.63	5.64	2.66	1.83

▲ 미국의 주요 도시들의 월별 강수량을 나타낸 도표

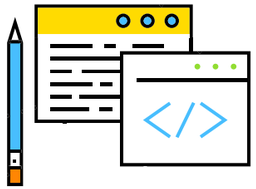
자, 이제 비로소 우리에게 어떠한 정보를 전달하고 있다. 예를 들어 뉴욕(New York)이라고 하는 도시는 1월(JAN)에 3.17인치(inch)라고 하는 강수량을 기록했다는 것을 알 수 있다.
데이터나 통계를 다루는 사람이라면 익숙한 도표의 모습이다. 우리는 엑셀을 비롯한 오피스 툴을 이용하여 유사한 형태의 도표를 많이 그려 낸다. 그렇다면 스스로에게 냉정하고 객관적인 시각에서 질문을 하나 던져 보자. 과연 해당 도표를 사람들이 볼 것인가? 해당 도표를 통해 어떠한 문제점이나 현상을 찾을 수 있는가? 앞서 이야기 했듯이 어떠한 현상이나 문제점을 직관적으로 발견할 수 있어야 비로소 가치 있는 정보라고 이야기 할 수 있다. 위 도표는 어디까지나 '전달자' 입장에서 만든 정보의 표현이다.
그렇다면 '수용자'의 입장에서 가치 있는 정보로 발전시켜 보자.

Average Rainfall (inches/month)

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
San Francisco	4.35	3.17	3.06	1.37	0.19	0.03	0.06	0.05	0.20	1.22	2.86	3.09
Seattle	5.35	4.03	3.77	2.51	1.84	1.59	0.85	1.22	1.94	3.25	5.65	6.00
Chicago	1.53	1.36	2.69	3.64	3.32	3.78	3.66	4.22	3.82	2.41	2.92	2.47
New York	3.17	3.02	3.59	3.90	3.80	3.65	3.80	3.41	3.30	2.88	3.65	4.42
Miami	2.01	2.08	2.39	2.85	6.21	9.33	5.70	7.58	7.63	5.64	2.66	1.83

▲ **1단계 변형** 이 단계에서는 데이터를 빠르게 읽을 수 있도록 디테일적인 부분들을 발전시켰다.

- 도시의 명칭을 우측으로 정렬
- 월(Month)에 대한 표기를 대문자에서 가독성과 시인성이 좋은 소문자로 변경
- 데이터 간의 행간 자간을 넓혀 줌으로써 인지를 도움
- 기준이 되는 월과 도시에 대한 명칭은 명도의 차이를 통해 톤을 낮춰줌으로써 데이터를 돋보이게 하는 효과



Average Rainfall (inches/month)

	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec
San Francisco	4.35	3.17	3.06	1.37	0.19	0.03	0.06	0.05	0.20	1.22	2.86	3.09
Seattle	5.35	4.03	3.77	2.51	1.84	1.59	0.85	1.22	1.94	3.25	5.65	6.00
Chicago	1.53	1.36	2.69	3.64	3.32	3.78	3.66	4.22	3.82	2.41	2.92	2.47
New York	3.17	3.02	3.59	3.90	3.80	3.65	3.80	3.41	3.30	2.88	3.65	4.42
Miami	2.01	2.08	2.39	2.85	6.21	9.33	5.70	7.58	7.63	5.64	2.66	1.83

▲ **2단계 변형** Insight를 줄 수 있는 가치 있는 정보로 만들어 보자. 방법은 어렵지 않다. 강수량에 따라 명도의 차이를 이용해 색을 넣었다. 강수량이 높은 곳은 짙은 색으로 낮은 곳에는 옅은 색으로 색을 넣었다. 이를 통해 정보의 수용자는 어느 도시가 강수량이 높은지, 특정 도시는 몇 월에 강수량이 높은지등을 쉽게 파악할 수 있다. 예를 들어 이 정보의 수용자는 특별한 노력을 하지 않아도 마이애미(Miami)는 여름철 비가 많이 오는 도시라는 것을 알 수 있다. 처음의 도표와 지금 발전시킨 도표는 수용자 입장에서는 아주 큰 차이가 있는 정보다.

처음 도표를 통해 정보를 'See'한다면 발전시킨 도표를 통해서는 'Understand' 즉, 이해 할 수 있다. 수 많은 통계 데이터를 효과적으로 전달하기 위해서는 철저하게 '수용자'의 입장에서 가치 있는 정보가 될 수 있도록 고민하고 발전시켜야 한다.



텍스트 정보를 효과적으로 전달 하는 방법

이번에는 텍스트 기반의 정보를 가지고 발전 시켜 보자. 먼저 다음 내용을 보자. 문장의 형태를 띄고 있는 텍스트 정보는 특히 수용자의 입장에서는 읽기 힘들며 귀찮은 정보의 형태이다.

우리 나라 직장인들의 한달 평균 독서 현황을 조사한 결과

1권 읽는다는 답변이 38%로 가장 많았고,
책을 많이 읽는 곳에 대한 답으로는 집이 43%로 가장 높았다.
한편 가장 책을 많이 읽는 사람은 7권이었다.

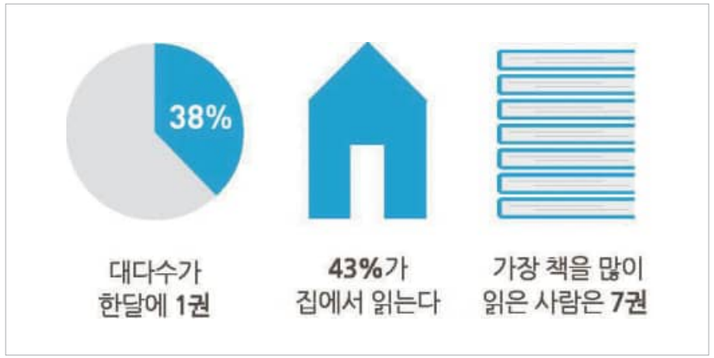
▶ **1단계** 위 내용에서 가장 핵심이 되는 내용은 강조 표시 되어 있는 바와 같이 1권–38%, 집–43%, 7권 이라고 하는 정보다. 먼저 핵심 정보를 추출하여 해당 정보를 중심으로 하여 재구성 해 본다.

우리 나라 직장인들의 한달 평균 독서 현황을 조사한 결과

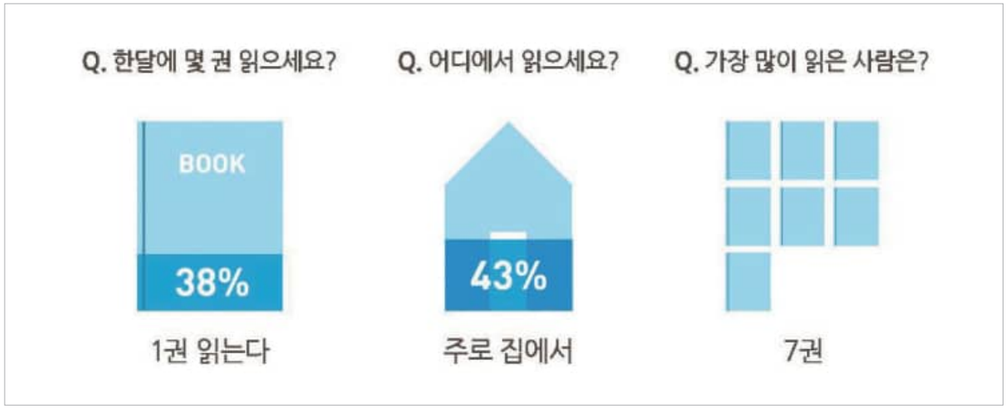
1권 읽는다는 답변이 **38%**로 가장 많았고,
책을 많이 읽는 곳에 대한 답으로는 집이 **43%**로 가장 높았다.
한편 가장 책을 많이 읽는 사람은 **7권**이었다.



◀ **2단계** 핵심 단어를 중심으로 정보를 재구성해보면 위와 같다.



◀ **3단계** 위 내용을 기반으로 시각적 요소를 더한다. 차트(Chart)나 픽토그램(Pictogram)의 시각적 언어를 포함하여 빠른 인지를 돕는다.





▲ **4단계** 3단계에서 아쉬운 부분은 정보의 Output은 있으나 Input을 파악하기에 부족함이 있다. 많은 사람들은 다수의 정보를 접할 때(위 사례에서는 총3가지 카테고리에 대한 정보를 전달하고 있다) 정보의 입력부분, 즉 무엇에 대한 사실을 전달하는지를 금방 잊어 버린다. 이를 효과적으로 돕기 위해서 좋은 방법은 질문과 답변(Q&A), 원인과 결과 또는 Before&After 등이 되겠다. 4단계 그림에서는 각 카테고리 별로 질문과 답변을 함께 표현함으로써 보다 직관적으로 정보를 전달 하고 있다.

정량적 데이터이든 정성적 데이터이든 정보 수용자의 입장에서 쉽고 빠르게 인지할 수 있으며 그 속에서 어떠한 Insight를 찾아낼 수 있는 가치 있는 정보로 만드는 고민이 동반될 때 비로소 '통계' 또한 그 가치를 지닐 수 있을 것이다.

이춘우 ◆ 통계교육원 사무관

국내 유일 통계교육기관… **통계교육원**

통계교육원은 국내 유일의 통계교육전문기관으로, 통계청 및 통계분야 직무에 종사하는 공무원, 통계작성기관 종사자를 대상으로 직무관련 통계전문교육을 수행하고 있다. 또한 학교와 사회단체 등 일반 국민들에 대한 통계교육을 담당한다. 통계교육원은 「세계 일류 통계전문 교육훈련 기관」라는 미션아래 글로벌 통계인재를 양성하고, 국가통계 발전을 선도하기 위한 통계전문인력을 육성하고, 대국민 통계소양 확산을 교육목표로 노력하고 있다.



통계교육원이 걸어온 길

통계교육원은 1991년 9월 13일, 통계전문 인력 양성을 위해 통계교육을 목적으로, 「통계연수원」으로 설립되었다. 1998년 12월 23일 국가교육훈련기관 조직개편에 따라 대전으로 이전하면서 '국가전문행정연수원 통계연수부'로 개편되었고, 2005년 1월 1일 다시 통계청 소속의「통계교육원」으로 개편되었다.

2007년 7월, 이러닝센터를 설치하여 통계교육 이러닝서비스를 시작하였으며, 2009년 8월20일, 현재의 통계센터로 이전하여 자리를 잡았다. 2013년 1월 4일부터는 이러닝 모바일서비스를 시작하면서 시공간 제약 없는 통계교육을 할 수 있게 되었다.

완벽한 교육환경 및 시설 완비

통계교육원은 일반강의실과 전문적인 전산교육을 위한 전산강의실, 영상강의를 위한 영상강의실, 참여형 토론수업을 위한 분임토의실, 연구실, 어학실습이 가능한 어학실습실 등의 강의시설과 6개의 동시통역실을 완비한 국제회의실, 300석 규모의 대강당, 국내·외 회의가 가능한 중회의실 등 회의 및 행사시설을 갖추고 있다. 또 효율적인 교육을 지원하기 위하여 최대 276명 수용이 가능한 숙박시설과 도서실, 정보사랑방 등 문화 및 휴게시설, 체력단련실, 탁구장, 배구장 및 농구장 등의 체육시설을 갖추고 있으며, 400석 규모의 식당 등 교육생을 위한 환경을 마련하고 있다.

교육과정… 집합교육, 이러닝(PC), 이러닝(모바일) 제공

통계교육원은 크게 집합교육과 이러닝 교육을 제공하고 있다. 2016년에는 148개 교육 과정이 개설되어 44,970명에게 921회에 걸쳐 교육을 실시하였다. 급변하는 시대 흐름에 발맞추어 빅데이터와 행정자료의 통계적 활용에 필요한 교육과정을 기존 2개에서 4개 과정으로 확대했으며, 한국데이터진흥원과 MOU를 체결해 양 기관의 교육과정에 교육생을 상호 파견하고 교육 노하우를 공유하고 있다.

또한 빅데이터 통계 작성에 필요한 통계패키지 프로그램으로 R 관련 교육과정을 기존 2개에서 4개 과정으로 확대해 전문 교육 수요 변화에 대응하고 있다.

❶ 집합 교육

통계교육원에서 실시되는 집합교육 과정은 초급과정 → 중급과정 → 고급과정으로 나누어 수준별로 접근이 가능하게 하였으며, 그 외 다양한 소양을 쌓을 수 있는 기타 과정을 운영하고 있다.

집합교육 신청절차

STEP1 교육대상자 추천요청	**교육생 선발 의뢰 : 통계교육원** 교육시작일 전전월 마지막 주에 교육과정, 교육대상, 기간, 교육비 등을 알림 (홈페이지 공지사항에 기재)
STEP2 교육대상자 선정	**교육대상자 추천 : 교육대상기관** 1) 기관 : 교육생 선발 기간 내에 공문으로 회신 2) 민간기관·개인 : 홈페이지 공지사항에 기재된 e-mail로 신청 가능
STEP3 교육대상자 확정통보	**교육대상자 확정통보 : 통계교육원** 교육시작 2주 전 공문으로 교육대상자 확정 통보 (홈페이지「집합교육생 선발확인」에서 확인 가능)

❷ 이러닝(PC) 교육

대전에 있는 통계교육원에 직접 찾아오지 않고도 집, 또는 사무실의 PC를 통해 다양한 통계교육을 받을 수 있다.

이러닝 교육은 다음과 같은 과정을 통해 청취할 수 있다.

1단계	PC를 통해 개별적으로 수강 신청 공무원의 경우 소속기관 교육담당자에게 상시학습시간 인정 가능 여부를 확인한 후 신청하시기 바랍니다.
2단계	수강 승인(교육생 확정)
3단계	PC를 통한 개별 학습
4단계	교육 수료(공무원의 경우 각 기관에 공문 발송)

담당자 연락처: 042-366-6223~4

❸ 이러닝(모바일) 교육

일상 생활에서 휴대하고 있는 모바일 환경에서 손쉽게 교육을 받을 수 있는 환경을 구축하였다. 통계교육원이러닝(모바일)에 접속하여 이용 가능하다.



❶ 통계교육원이러닝(모바일)에 접속후 로그인선택

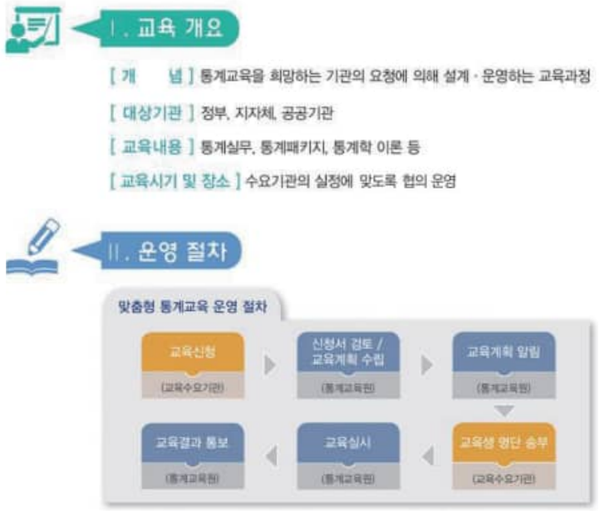


❷ 로그인 페이지에서 양식 입력

원하는대로, 환경에 맞는 교육 제공… 맞춤형 교육

맞춤형 교육

전문적인 통계교육을 희망하는 기관은 통계교육원에 교육신청을 하면 요구 사항에 맞는 맞춤형 교육이 가능하다. 홈페이지에서 전년도 실시한 교육 내용을 참고할 수 있다.



미래 인재 양성에 발벗고 나서다

4차 산업혁명 시대를 맞아 통계교육원은 미래 인재 양성을 위해 통계적 소양 증진에 힘쓰고 있다. 미래에는 빅데이터 환경으로 인하여 데이터에 대한 이해 및 통계적 능력은 필수로 갖춰야 하는 소양이 되었다. 이에 통계교육원은 「2015 개정 수학과 교육과정」에 맞춰 공동으로 실용 통계교육 협업 계획[1]을 발표(2017.6.7)하는 등 학교 통계교육을 위해 교육부와 협업을 진행하고 있다.

통계교육원에서는 교육용 통계패키지 「통그라미」를 개발해 교육 공학도구로 활용하도록 하였다. 「통그라미」는 2018년도부터 적용되는 교육과정의 중1 수학교과서에 공학도구로 소개되어 정규수업에 활용 예정에 있다. 검색창에서 '통그라미'로 검색(http://tong.kostat.go.kr)하면 누구나 쉽게 설문조사 및 통계분석을 수행할 수가 있다.

또한, 2017년 8월 28일에는 국회 의원회관에서 국회와 공동으로 학생들의 실용 통계교육 방향 논의를 위한 '4차 산업혁명시대, 초중등 실용 통계교육 방향 모색' 정책토론회를 개최하였으며, 2017년 11월 27일에는 The-K호텔에서 일선 교사 및 교육관계자 대상의 '실용 통계교육 활성화 추진 설명회'를 개최하였다.

통계교육원에서 그간 개발한 '고교 통계교육 교재' 및 '통계를 활용한 토론수업 방법' 등을 공유하며 통계교육의 필요성에 대한 공감대를 형성하였고, 통계교육 활성화를 도

1) (교육부) 교육과정과 교과서 개편, 교원 연수 실시, 교수학습 자료 개발
(통계청) 공학도구 「통그라미」 개발, 교육용 데이터 제공, 교원 연수 등



모하고 있다. 이외에도 '중등학교 통계교육 핵심교원 연수', 중학생과 고등학생 대상 통계아카데미, 초등학생을 위한 통계캠프, 사랑나눔통계교실을 운영하고 있으며, 통계연구학교, 교사연구회를 운영 및 지원하고 있다. 이외에도 전국학생통계활용대회, 대학생 재능기부단 등 다양한 프로그램을 운영 중이다.

세계인의 통계교육을 책임지다

2005년부터 통계청은 KOICA 위탁 통계연수과정을 운영하기 시작하였고, 2008년부터는 개발도상국 국가통계기관 종사자를 대상으로 통계실무기법 전수 및 통계기반 정책수립 교육을 지원하고 있다.

통계교육원은 3년간 8개 교육과정을 운영하여 캄보디아, 에콰도르, 이집트, 아프카니스탄 등 15개 국가의 119명이 통계실무과정을 이수했다. 또한 2013년에는 베트남의 통계 역량 강화 프로젝트 2년차 사업이 진행되었다. 이때 전문가 파견, 강사 육성, 교재 개발 및 교육기자재 보급 등의 통계 인력 양성 시스템을 구축해 자체 통계교육 운영 기반을 조성했다.

또한 2015년에는 스리랑카 통계청의 역량 강화 사업을 진행하며, 성과 향상을 위한 사전조사를 수행하고, 통계교육을 위한 강의실 설치 및 교육 기자재 보급 등 통계교육 인프라를 구축했다. 2016년부터 UN 아시아·태평양 통계연수소(UNSIAP)와 공조해 매년 2~3회 지역훈련과정을 운영하고 있으며, 2016년에는 캄보디아 등 17개국 공무원들을 대상으로 농산물 생산비 작성 과정을 운영하였다.

국민에게 한걸음 다가서는 통계교육원이 될 것을 약속

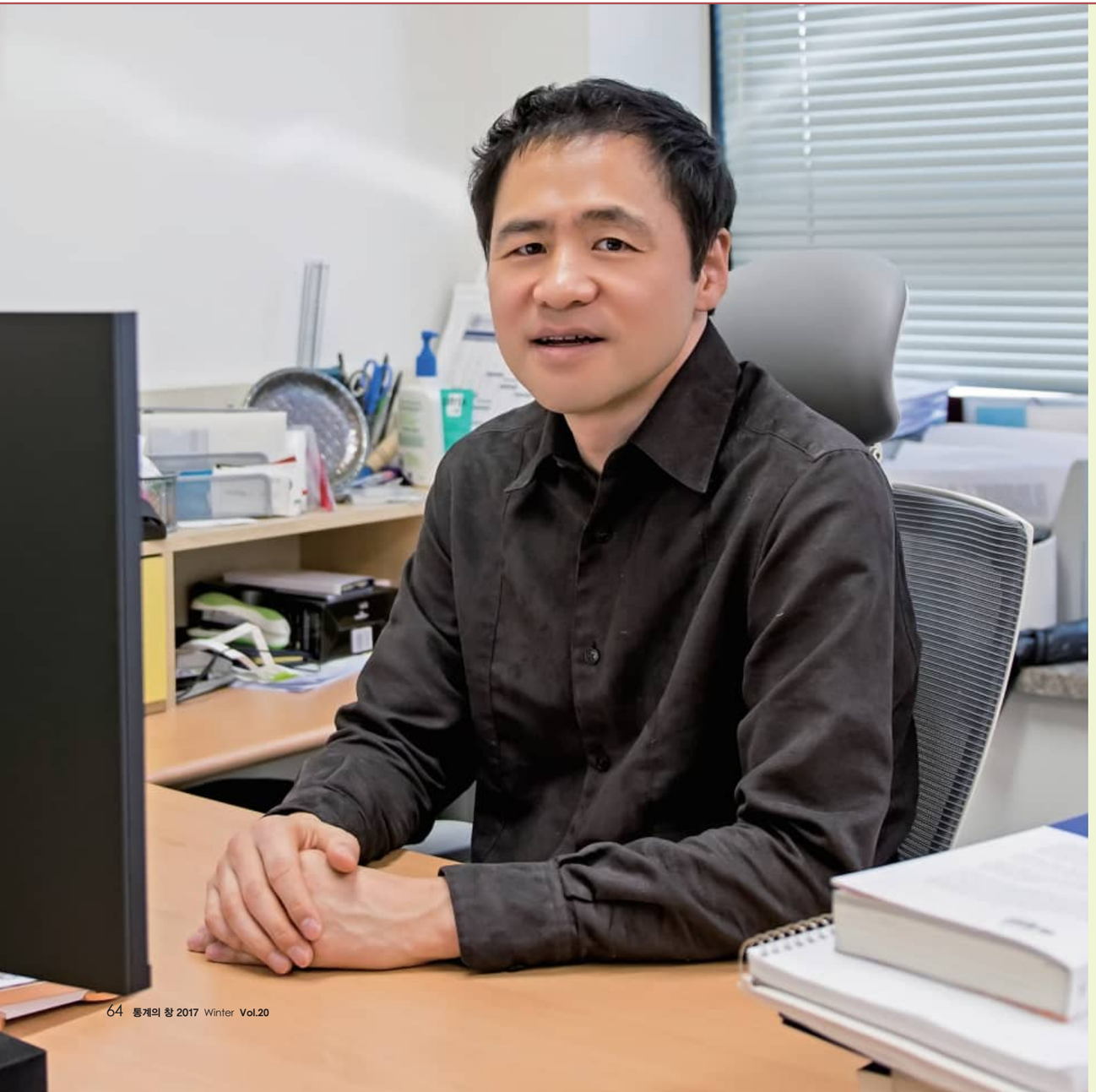
그간의 통계교육 활성화 기반 마련 노력을 바탕으로 2018년도 통계교육원은 '통계소양 확산', '전문역량 강화', '통계교육 기반 조성'을 위해 노력할 것이다. '통계소양 확산'을 위해서는 정부 및 공공기관 정책입안자를 대상으로 하는 통계교육과, 대국민 통계소양교육을 강화 할 것이며, 이를 위해 실용 통계교육을 위한 콘텐츠를 개발·보급하여 통계교육을 활성화 해 나갈 것이다. '전문역량 강화'를 위해 실시한 통계교육 프로그램 진단 결과를 반영하여 통계전문교육 과정을 재설계하고, 행정자료 및 빅데이터 분석 역량 제고 등을 위한 전문역량 교육을 강화할 예정이다.

또 통계교육원은 4차 산업혁명 시대 국가발전을 선도할 인재 양성 및 '통계교육 기반 조성'을 위해 지속적으로 노력을 해나갈 것을 약속한다.

건강한 삶, 행복한 삶은
당신 가까이에 있다

Interview

연세대학교 사회학과 교수
염유식



「친구 많을수록 골다공증 위험 높다」
이게 무슨 말인가. 신문에서 읽은 이 말은 단번에 나의 호기심을 사로잡았다. 관련 신문기사를 더 찾아보았다.
「말 상대가 많은 할아버지 고혈압 덜 걸린다」
「친구 많을수록 치매에 덜 걸려」

이 기사는 2011년부터 6년 째 사회연결망에 따른 건강관련 연구를 수행한 결과이다. 연세대학교 염유식 교수는 이러한 연구 결과를 네이처의 자매학술지 '사이언티픽 리포트'에 발표했다. 강화도 두개 면의 전체 고령자와 그 배우자를 대상으로 삼아 사회연결망을 조사하고 그들의 건강상태와 뇌의 기능성 자기공명 영상을 촬영하였다고 한다. 신문에 짧게 요약된 정보들로는 아쉬움이 있어 직접 찾아뵙고 자세한 얘기를 듣고 싶었다.
연세대학교 연구실에 들어서자 염유식 교수의 초상화 두 개가 먼저 눈에 들어왔다. 제자 분들이 그려준 것이라고 한다. 사진 기자와 나는 너무 젊게 보이신다고 한마디 하자 "사실 나이 많아요." 연말이라 연구 마감에 바쁘신 듯 조금은 초췌한 얼굴로 우리를 맞으시며 젊어 보인다는 말에 쑥스러운 듯 한마디 던지셨다. 연구실 한편에 키보드 건반 상자가 놓여 있어 음악도 하시느냐고 물으니 군대에 간 아들이 두고 간 것이라 했다. 아 그 정도 나이시구나.

‘코호트 연구’는 어떻게 진행되는 연구 인가요

넓은 의미로는 사회과학에서의 패널자료 연구와 동일하지만, 의료 쪽에서는 특히, 비슷한 나이대의 사람들을 패널의 형태로 전향추적조사하여 질병의 원인 등을 밝히는 연구를 지칭합니다.

지금 어떤 연구를 하고 계시는 지요

주로 노인들의 건강에 대해서 연구를 진행하고 있는데요. 사회학적 친구 관계에 대한 연구(KSHAP[1]), 도시와 농촌 비교 연구(KURE[2]), 대사 심혈관 질환에 관한 연구(CMERC[3]) 등입니다. 세 가지 연구 모두 가능한 자세하게 고령자들의 사회연결망을 설문의 형태로 수집

1) Korean Social Life, Health, and Aging Project 2) Korean Urban Rural Elderly
3) Cardiovascular and Metabolic Diseases Etiology Research Center

하고 있습니다. 고령자들이 중요한 이야기를 상의하는 (배우자를 포함한) 지인에 대한 인구통계학적 사항을 자세하게 7명까지 묻고, 이 지인들과의 관계정도도 측정합니다. 특히, KSHAP은 강화도 두 면의 모든 고령자들과 그 배우자들의 전수 조사를 하여 두 마을의 전체 사회연결망을 패널 형태로 측정하고 있습니다. 한 마을은 현재까지 4번 조사하였고 다른 마을은 2017년 1월에 첫 조사. 앞으로 4년간 두 마을 모두 한 차례 더 조사 예정입니다. 또한 채혈을 포함한 건강검진을 통하여 질병과 건강에서 사회적과정과 생의학적 과정이 분리될 수 없음을 밝히고자 합니다.

강화도를 연구 대상으로 삼은 이유가 있는 지요

사회 연결망을 조사하기 위해서는 인구 이동이 많은 도시 보다는 농촌 지역이 적합합니다. 또한 한 마을의 전체 연결망을 수집할 수 있어 전체적인 사회의 모습을 측정할 수 있습니다. 그리고 강화도가 서울에서 거리상 멀지 않아서 그곳으로 정했습니다.

연구 주제가 모두 건강입니다. 특별한 이유가 있으신지요

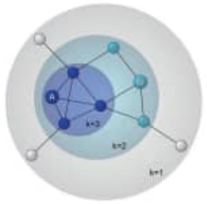
원래 전공은 사회연결망이었습니다. 대부분의 사회연결망 전공자들이 경제사회학에 관심을 가지는데, 저는 미국에서 공부할 때 첫 연구를 성병과 유방암에 관한 연구로 시작했습니다. 그런 인연으로 건강이 제 주된 관심사가 되었어요.

이런 코호트 연구의 어려움이 있다면 어떤 것인지요

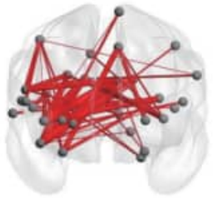
가장 어려운 점은 추적조사이다보니 시간이 지남에 따라 연락이 닿지 않는 고령자들이 늘어나게 된다는 점입니다. 최초의 인원을 가능하면 그대로 유지하여 높은 유지율을 유지하는 것이 가장 중요하고 힘듭니다.

신문에 실린 연구 결과를 보니 사회적 관계와 건강이 밀접한 관계가 있는 것으로 나타났는데요

한 마을의 친구 관계망에서 중심을 차지하는 고령자들은 뇌 건강이 좋다는 것으로 나타났어요. 또한 혼자 사는 할머니들도 양호하게 나타났습니다. 이는 관계에서 오는 스트레스가 적은 경우로 여겨집니다. 재미있는 것은 아들은 둔 노부부와 딸을 둔 노부부에서 차이가 나타났는데, 아들을 둔 경우 할머니의 우울증이 높게 나타났습니다. 할머니가 할아버지와 아들을 돌보아야하는 경우로 해석됩니다. 반면에 딸을 둔 경우에는 할머니는 차이가 없는데 할아버지의 우울증이 낮게 나타났습니다. 딸이 할아버지를 돌보아주는 관계로 해석됩니다. 즉 아들은 둔 경우 할머니는 할아버지와 아들 두 명을 보살펴야하므로 스트레스가 높고, 딸을 둔 경우 할아버지는 할머니와 딸로 부터 모두 보살핌을 받으니 스트레스가 낮은 것으로 여겨집니다.



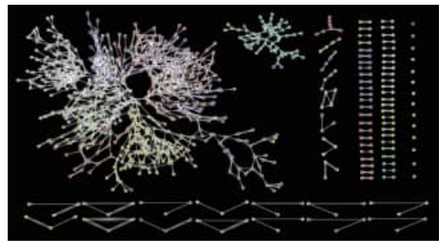
마을의 사회관계망에 대한 그림으로 중심부 파란 원 안에 들어간 사람들이 관계망 중심부에 있다



사회관계망 중심부에 있는 고령자일수록 뇌 연결성이 활발하게 나타났다

아버지 입장에서는 딸이 있어야한다는 것이 과학적으로 증명 된 거네요. 저는 아들 두 명인데 고민이 좀 되는 데요. 마을 관계망에는 특이할만한 사항이 있는지와 도시와 농촌의 차이는 있는지 궁금하네요

아래 그림에서 보면 위쪽 가운데에 있는 작은 집단(초록색)의 건강이 별로 좋지 않았습니다. 한 마을에서 집단적으로 격리된 지역이라 그렇지 않을까 생각하고 있습니다. 그리고 도시와 농촌에 대한 연구는 아직 진행 중입니다. 데이터를 모으고 있는 중이라 2년쯤 뒤에 결과가 나올 것 같습니다.



이 그림은 한 마을(면)의 전체 연결망을 표시한 것이다. 원은 개별 고령자를 나타내며, 원끼리 서로 선으로 연결이 되어 있는 것은 서로 친구나 지인관계임을 보여준다. 원의 색깔은 서로 다른 리(里)를 표시했다

현재 우리나라는 베이비부머 세대가 은퇴를 앞두고 있습니다. 은퇴해서 잘살려면 지금 아내한테 잘 해야 한다고 우스개처럼 얘기하는데요. 그런 연구 결과도 있는 지요

아직 우리나라는 연구 중이지만 미국에서 연구된 결과는 나와 있습니다. 몸의 만성 염증 정도를 재는 CRP(C-Reactive Protein)라는 것이 있는데요. 이것이 높을수록 면역력이 떨어져 있는 상태라고 이야기할 수 있습니다. 남자는 아내와 보내는 시간이 많을수록 몸의 염증 정도가 낮았고, 여자는 시간보다는 마음을 터놓고 얘기하거나 상대방에게 비난을 덜 받을수록 염증 정도가 낮았습니다. 즉 남자는 부부관계에서 양이 중요하고 여자는 관계의 질이 중요하다고 할 수 있겠습니다.

남자들이 행복하려면 아내한테 잘하고 딸을 두어야한다는 결론이네요. 전 두 부분에 다 못하고 있으니 걱정이 많이 되는데요. 요즘 1인 가구가 많아지는데요. 이런 현상은 어떻게 보시는 지요.

연구결과 1인 가구가 더 건강한 경우가 많았습니다. 다양한 사회활동을 통해 많은 친구들을 유지하면서 외부에서 오는 스트레스에서 벗어나는 것이 더 좋을 수 있다고 여겨집니다. 물론 빈곤 때문에 어쩔 수 없이 겪게 되는 1인 가구는 문제가 될 가능성이 높습니다. 1인 가구(living alone)와 친구 없이 혼자 살아가는 것(being alone)은 매우 다릅니다. 후자는 매우 위험하지만 전자는 반드시 그렇지는 않습니다.

아이들에게 경쟁해서 사회적으로 더 좋은 위치를 얻게 하는 것보다 더 중요한 것은 가까운 사람들과 어울리면서 살아가는 방법을 가르쳐 주는 것이 더 중요합니다. 제가 10년째하고 있는 우리나라 어린이 청소년 연구에 의하면, 엄마가 자녀들의 친구나 선생님, 주변인물을 많이 아는 자녀들일수록 더 행복했습니다.

교수님의 연구를 통해 사회적으로 어떤 메시지를 줄 수 있을까요

예를 들어 고령자 건강에서 가장 큰 문제 중의 하나는 치매입니다. 현재는 치매에 대해서 예방이 아니라 일단 질병에 걸린 후에 이루어지는 치료에 중점을 둔 정책이 펼쳐지고 있습니다. 이런 연구를 통해 치료에 앞서 예방 활동이 가능해지지 않을까요. 현재 치매에 대해 얘기하는 예방이 손가락 운동이나 고스톱을 장려하는 수준입니다. 인간관계는 우리가 생각하는 것보다 훨씬 더 고도의 인지 능력이 필요합니다. 이러한 복잡한 사회 활동을 유지하는 것 자체가 충분히 치매를 예방 할 수 있다고 믿습니다. 예를 들어 건강을 위해 금주하면서 혼자 헬스장에 가는 것 보다는 차라리 사람들과 어울려서 운동을 하고, 끝나면 적당한 음주도 하면서 사람들과 어울리는 것이 훨씬 더 건강한 삶을 가져올 것입니다. 또 다른 예를 들면 현대인은 몸무게에 대해서 너무 민감합니다. 의학적으로는 비만보다 아래 단계인 과체중은 경우에 따라, 건강에 아무런 영향을 안 미치거나 도리어 건강에 도움이 된다는 연구들도 있습니다. 또한 많은 분들이 몸무게를 걱정하며 보통 안 먹거나 특정 음식만을 먹어서 살을 빼는 경우가 많습니다. 실제 영양학적으로는 다양한 음식을 섭취하는 것은 건강을 위해 매우 중요합니다. 다양한 음식을 먹고 다채로운 육체적·사회적 활동을 하는 것을 권장하는 이유입니다.

교수님 말씀을 듣다보니 요즘 우리나라에 헬조선이라는 말이 많이 나오는데요. 행복에 대해서 어떻게 생각하시는지요

우리나라는 너무 목적 지향적인 면이 많습니다. 좋은 대학가야 되고, 강남에 집이 있어야하고, 좋은 직장이 있어야만 행복하다고 생각합니다. 그것을 얻기까지 지금은 참아야하죠. 하지만 그렇게 불행한 시절을 겪고 좋은 대학에 가면 행복할까요? 저는 그렇게 믿지 않습니다. 행복은 어떠한 조건을 만족하는 순간 소유하게 되는 무엇이 아니라, 그러한 조건을 만족하기 위해 가까운 사람들과 함께 고민하고 노력하는 과정이라고 믿습니다. 그런 점에서 우리나라 부모들은 아이들에게 불행하게 사는 법을 체계적으로 가르치고 있습니다.

제가 서울 시민 1,000명을 대상으로 수행한 연구결과에 따르면 월수입이 400만원까지는 행복도가 소득에 비례해서 증가합니다. 하지만 400만 원 이상의 수입은 더 이상 행복 정도를 늘려주지 않습니다. 400만 원 정도 수입이 되면 더 벌려고 발버둥 치기보다는 주변 사람들과 더 많은 시간을 보내세요. 아이들에게 경쟁해서 사회적으로 더 좋은 위치를 얻게 하는 것보다 더 중요한 것은 가까운 사람들과 어울리면서 살아가는 방법을 가르쳐 주는 것이 더 중요합니다. 제가 10년째하고 있는 우리나라 어린이 청소년 연구에 의하면, 엄마가 자녀들의 친구나 선생님, 주변인물을 많이 아는 자녀들일수록 더 행복했습니다. 이런 아이들은 자신의 고민이나 생각을 엄마와 많이 공유할 가능성이 높습니다.



염유식 교수가 걸어온 길 ...

- 현재 연세대학교 사회학과 정교수
- 2000. The University of Chicago. 사회학 박사
- 2000~2005. The University of Illinois at Chicago. 사회학과 조교수

우리나라 아버지들은 일에 쫓겨 사는 경우가 많은데요. 아버지를 위해서도 한 말씀 해주시죠

아버지가 같이 아이들과 게임해주는 집은 아이들이 게임 중독에 걸릴 확률이 크게 줄어듭니다. 아이들이 게임을 얼마나 오래하는지가 중요한 것이 아닙니다. 아버지들이 없는 시간이라도 쪼개어 자녀들과 게임이라도 함께 해주시는 게 중요합니다.

지금 우리나라가 고령사회에 접어들고 있어 사회적으로 이슈가 많이 되고 있는데요. 여기에 대해서 교수님의 생각은 어떠신지요

저도 고령사회에 대해서 어떻게 가야한다고 말하기가 힘든 면이 있습니다. 왜냐하면 인류가 지금처럼 오래 산적이 없어요. 처음으로 접하는 그런 새로운 유형의 시대라는 거죠. 따라서 저는 고령화 사회라고 하면, 생산력 감소, 고독사와 같이 자동적으로 연상되는 여러 사회적 문제를 그대로 인정하지는 않습니다. 새로운 유형의 사회는 새로운 시각으로 접근해야겠지요. 기존의 프레임으로 보면 문제투성이지만 이제 새로운 프레임으로 사회를 보기 시작할 때라 여겨집니다.

우리나라 사회 복지에 대해서 어떻게 생각하시는지요

젊을 때는 경제적 빈곤을 이겨낼 수 있지만 고령자들은 경제적 문제가 심각할 수 있습니다. 경제활동이 불가능한 경우가 많기 때문입니다. 다른 나라 사례에서 성공한 복지는 크게 두 가지를 꼽을 수 있습니다. 첫째 유럽형으로 높은 세금, 높은 복지입니다. 두 번째는 일본처럼 평생직장이 보장 되는 사회입니다. 이제 우리에게 적합한 복지 모델을 만들어야 합니다. 저는 세금을 더 걷어 복지를 강화하는 것이 맞다고 생각합니다. 더 걷어 들이는 세금이 투명하게 잘 쓰인다면 국민이 마다할 이유가 있을까요. 자신들도 나중에 혜택을 보는데 말이죠. 주변 사람이 행복해야 자신도 행복합니다. 주변 사람들이 불행한 사회는 아픈 사회입니다.

끝으로 앞으로의 연구 계획은 어떠신지요

앞에서 말씀드린 KSHAP 연구는 한국연구재단의 6년간의 지원 결과입니다. 앞으로 4년간은 지금과 같은 연구를 유지하고, 그 후에는 big data나 wearable device등의 측정을 본격적으로 도입하여, 건강과 질병의 연구에서 사회과학의 위치를 재정립하고자 합니다.

이주원 ◆ 통계청 통계서비스기획과 사무관

저출산·고령화로 인한 미래를 예측해보는

KOSIS '인구로 보는 대한민국'

통계청은 지난 9월 '인구로 보는 대한민국' 시각화콘텐츠를 국가통계포털(KOSIS, http://kosis.kr)을 통하여 서비스하였다. '인구로 보는 대한민국'은 저출산·고령화 등으로 인한 인구구조의 변화가 사회·경제적으로 다양한 영향을 미치고 있음을 알려주는 콘텐츠이다.



예견된 참사?… 미리 알고 준비한다면

인구의 모습과 변화는 미래사회를 전망할 수 있다. 현재 둔화되고 있는 인구증가속도, 저출산 및 빨라지는 고령화를 보면 인구의 변화가 사회·경제적으로 심각한 문제를 유발할 것으로 예측할 수 있다.

최근의 예를 들면 선생님은 우리나라 최고 선호 직업이었으나 교원임용 절벽으로 사회적으로 큰 혼란을 겪고 있다. 이는 통계로 보면 예견된 참사이다. 문제는 이러한 현상이 이제 시작이며, 학교 뿐 아니라 소아과, 학원 등 모든 국민들이 직·간접적으로 인구구조 변화로 인한 영향을 받게 될 것이다.

많은 이들이 이러한 상황을 전혀 인식하고 있지 않음에 따라 이러한 사실을 국민들이 잘 이해하고 공감하게 하는 서비스가 필요하다는 인식에서 이 서비스를 만들게 되었다.

2029년부터 사망＞출생… 생산가능인구 50년 뒤 반토막

'교원임용 절벽'은 예견된 참사였다

저출산으로 사라진 初中高, 현재 중학교 수보다 많은 3726곳

'고령사회의 덫' – 가족 붕괴시키는 '치매 간병 살인'

"청년실업자 4년 후에 130만명 넘을 수도"

"노인들 부양하러 태어났나요"– '목말 사회'

인구절벽 가속– 신생아 '30만 시대' 온다

생산인구 3763만→3168만→2062만… 日보다 빨리 줄어

국민과 함께 고민하고 토론하며 만든 「인구로 보는 대한민국」

이 서비스는 일반시민 및 전문가가 포함된 통계청 국민디자인단[1]이 기획에 참여하여 일반 국민의 아이디어와 목소리를 반영하였다. 또한 국민생각함을 통하여 국민들이 '인구'에 대해 어떠한 생각을 가지고 있는지 살펴보고 서비스 방향을 잡는데 참고하였다.

국민들의 생각을 바탕으로 크게 2가지 측면에서 서비스 방향을 검토하였다.

첫 번째, 미래를 예상할 수 있는 서비스 측면이다. 통계정보만으로는 일반국민이 인구구조 변화가 나와 내 가족에게 어떠한 영향을 끼치는지 까지는 알 수 없다. 서비스를 통하여 보다 체감할 수 있고 경각심과 미래설계를 할 수 있는 방법을 모색하였는데, 인구전문가의 기고문과 신문기사 등을 스크랩하여 서비스하자는 아이디어가 나왔다.

두 번째, 국민의 목소리를 담은 서비스 측면이다. 국민생각함의 많은 의견들을 보면서 느낀 것은 '국민들이 정말 하고 싶은 말이 많구나' 였다. 국민의 목소리를 담은 서비스를 위해 국민이 직접 정책 아이디어를 포스터로 만들어 소통토록 하자는 아이디어가 나왔다.

---

1) 국민디자인단이란 정부정책 공급자인 공무원과 수요자인 국민, 서비스디자이너 등이 함께 참여하여 정책을 수요자 중심으로 재해석하고 개발하는 활동을 하는 팀

크게 3개 부문을 통하여 미래의 모습을 담았다. 인구로 보는 대한민국은 숫자로 보는 인구, 인구로 보는 미래, 체험 마당으로 꾸며 졌다.



▲ 인트로화면

표 1 메뉴구성도

메뉴		내용
① 숫자로 보는 인구		인구와 가구, 출생과 사망, 혼인 등 6개 부문 30개 화면에서 26개 지표를 서비스
② 인구로 보는 미래	미래를 보는 시선	인구구조 변화로 인한 미래 모습을 기고문으로 서비스
	미디어가 본 미래	인구와 관련된 지면기사 스크랩 서비스
③ 체험마당	내가 꿈꾸는 세상	이용자가 바라는 인구정책 아이디어를 포스터로 작성
	인구피라미드	이용자가 인구변동 시나리오를 선택하여 미래 인구변화를 피라미드 움직임으로 체험



1 숫자로 보는 인구… 인구구조 변화를 한 눈에 이해할 수 있는 26개 지표

저출산, 고령화 등 인구구조 변화를 한 눈에 이해할 수 있도록 다양한 그래픽과 설명글을 제공하여 흥미와 이해 도모한다. 인구구조 변화를 볼 수 있는 지표를 선정·서비스하는데 부양비, 연령별 인구구조 등 30개 화면에서 26개 지표 서비스한다.



▲ 숫자로 보는 인구 화면



표 2 서비스 지표 및 화면

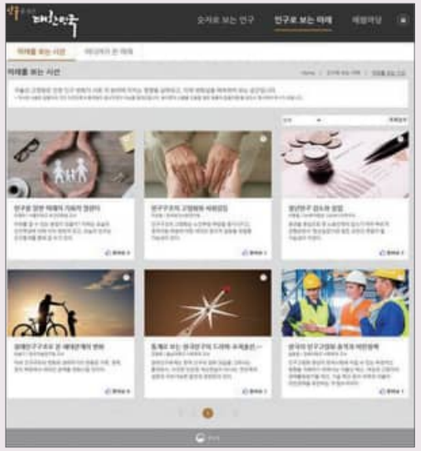
부문		화면
대분류	중분류	
인구와 가구	인구	총인구와 인구성장률, 시도별 인구, 시군구별 순위
	가구	총가구와 1인가구, 총가구와 고령자가구
출생과 사망	출생아수와 사망자수	출생아수와 합계출산율, 출생아수와 사망자수
	기대수명	기대수명, 시도별 기대수명
혼인	혼인	혼인건수와 조혼인율, 평균초혼연령
연령구조	연령별 인구구조	연령계층별 인구구성비
	생산가능인구(15–64세)	생산가능인구, 시도별 구성비, 시군구별 순위
	고령인구(65세 이상)	고령인구, 시도별 구성비, 시군구별 순위
	학령인구	학령인구, 시도별 구성비
	중위연령	중위연령, 시도별 중위연령, 시군구별 순위
부양비	부양비	부양비, 시도별 총부양비, 시군구별 순위
다문화	다문화 가구	다문화 가구구성
	다문화 가구원	다문화가구원, 국적별 결혼이민자, 국적별 귀화자

2 인구로 보는 미래… 전문가 기고문과 관련 신문 기사 발췌 서비스

인구의 변화가 미래 사회·경제적으로 미치는 영향을 알 수 있도록 인구 전문가에게 기고문을 의뢰하여 서비스한다. 또한 이미 진행되고 있는 저출산·고령화의 부정적인 현상, 해외 인구정책 사례, 인구구조 변화로 인한 미래 모습 등 관심을 유도할 수 있는 신문기사를 발췌·서비스한다.



▲ 신문기사 서비스 예시



▲ 인구로 보는 미래 中 미래를 보는 시선

3 체험마당… 국민의 목소리와 아이디어 공유

정부의 인구정책에 대해 국민의 목소리와 아이디어를 직접 포스터로 작성하여 공유할 수 있도록 서비스한다. 이용자는 어울리는 사진 또는 이미지를 배경으로 올리고 인구정책 아이디어를 입력하면 포스터가 만들어진다. 이를 SNS 등을 통하여 공유할 수 있으며, 게시된 포스터들을 보고 사회 곳곳에 다양한 인구정책이 필요함을 느낄 수 있다. 또한 인구변동 시나리오[2] 를 직접 설정하여 인구피라미드를 출력해보고 미래 인구를 예측할 수 있다.



▲ 체험마당 中 내가 꿈꾸는 세상



▲ 체험마당 中 인구피라미드

다양한 서비스 지속적으로 제공할 예정

통계청은 국민이 통계정보의 이해와 활용을 위한 일환으로 다양한 시각화콘텐츠를 개발 서비스하고 있다. 이번 '인구로 보는 대한민국' 서비스로 국민들이 저출산·고령화로 인한 현황을 이해하고 정책 및 미래 설계에 도움이 되기를 기대하며, 국민들로부터 나온 통계정보를 이용자들이 친숙하고 쉽게 많이 이용할 수 있도록 다양한 서비스를 지속적으로 할 예정이다.



2] 인구변동요인은 출생, 사망, 국제이동이며, 각 요인에 대한 가정(고위, 중위, 저위)에 따라 30개의 시나리오가 있음

김영진 ◆ 한국통계진흥원 부장

Review

책을 베고 잠들면, 꿈은 당신이 안 가본 길로 안내한다

사람에게도 저마나 다른 향기가 있듯이 책도 저마다의 향기를 가집니다. 몇 시간 동안이지만 서로 대화를 나누다 보면 그 향기의 잔상이 얼마동안 남아 있습니다. 잠시 차 한 잔을 마시며 그 향기에 대해서 얘기해보고자 합니다. 지금까지는 한권의 책을 골라 자세히 읽어나가는 느낌을 다루어 보았습니다. 이번호에서는 2017년에 나온 통계관련 책들 중에서 개인적으로 관심 있었던 책 몇 권에 대해서 다루어 보고자 합니다. 지금 이 글을 읽는 사람이면 통계에 관심이 있는 분이고 그런 분이면 서점에서 통계관련 책장 앞에 한번쯤 걸음을 멈추는 사람일테니요.

라이프 프로젝트

헬렌 피어슨 저
이영아 역
와이즈베리



아마 여러분들은 마시멜로 실험에 대해서 들어 보았을 것입니다. 몇 분 동안 마시멜로를 먹고 싶은 유혹을 이겨낸 아이가 나중 성인이 되었을 때 성공할 가능성이 높았다는 이야기죠. 그런데 이런 것을 어떻게 알 수 있었을까요. 지금 마시멜로 먹는 것을 참았다고 먼 훗날 벌어질 그 아이의 장래에 대해서 그렇게 주장하는 것이 너무 위험해 보입니다.

이런 주장을 가능하게 해주는 것이 바로 장기간동안 추적 조사하는 코호트연구 덕분입니다. 코호트 연구는 처음 조건이 주어진 집단(코호트)에 대하여 경과와 결과를 알기위해 계속 추적 조사하는 방법입니다.

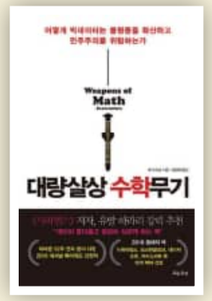
「무엇이 인생의 차이를 만드는가」 이런 부제목을 보고 이 책을 고른다면 실망 할지도 모릅니다. 인생의 차이는 무엇에서 오는 것일까에 대한 답을 찾기 위해 계속 책을 읽어 내려가지만 400쪽 가까이 되는 책의 대부분이 영국에서 진행되었던 코호트 연구의 진행 과정에 대해서 주로 서술되고 있기 때문이죠. 책 마케팅에 대해서 살짝 짜증도 나긴했지만 통계에 관련된 일을 하는 사람으로서는 이런 책이 나온 것이 반갑습니다. 우리가 지금 상식적으로 알고 있는 사실들이 어떤 노력을 통해 이루어진 결과인지 알 수 있게 된 것 만으로도 의미 있는 책입니다.

개인적으로 이 책을 읽으면서 놀랐던 것은 영국의 첫 코호트 연구가 1946년에 약 1만 7천여 명의 아기들을 대상으로 시작하여 12년 마다 5차례 시행되었다는 것입니다. 1946년 우리나라 상황을 생각하면 영국이 얼마나 앞서 이러한 문제에 접근 했는지 새삼 놀라울 뿐입니다. 이렇게 1946년부터 출생정보, 건강 상태, 지능, 학교 성적, 직업, 소득, 신체 발달에 대해서 추적 조사를 함으로써 수많은 연구 업적을 이룰 수 있었습니다. 이러한 결과를 통해 요즘 임신부들은 엽산, 요오드를 복용하며 아이들의 건강을 지킬 수 있었고, 또 출생조건이나 아이의 성향이 어떻게 미래의 삶에 영향을 주는지 예측 가능하게 해주었습니다.

코호트 연구는 우리의 삶에 대해서 필요한 연결고리나 좀 더 나은 삶을 위한 조건을 찾아내 주는 역할을 해줍니다. 사람들은 지금의 행동이나 생각이 미래에 어떤 결과를 가져다 줄지 알기 어렵고 그것을 느끼는 순간 되돌릴 수 없는 경우가 많기 때문입니다. 특히 건강이나 행복에 대한 부분은 이런 연구를 통해 많은 조언을 구할 수 있을 것 같습니다. 코호트 연구는 과정은 힘들지만 그 결과는 재미있습니다.

대량살상 수학무기

캐시 오닐 저
김정혜 역
흐름출판



자극적인 제목의 대량살상 수학무기라는 책은 제목은 수학무기이지만 목차를 보면 통계 모형에 관한 내용이라는 것을 알 수 있습니다.

통계가 수학이라는 범주에 속한다고 보면 이상할 것이 없지만 통계와 관련된 사람은 '대량살상 통계무기'라는 제목이었으면 좋았을 걸 하는 생각이 들긴 합니다.

이 책은 모형화 알고리즘에 따른 위험성을 경고하고 있습니다. 제목만 본다면 빅데이터의 위험성을 알리는 내용 같지만 읽어보면 오히려 앞으로 우리사회가 얼마나 데이터에 영향력에 놓이게 될지 짐작해볼 수 있습니다.

책 중간쯤까지 읽어 내려가면서 이런 모형에 대해서 이렇게 피해를 볼 수 있겠구나하는 생각하다가 문득 '어 이거 당연한 얘기 아니야' 하는 생각이 들었습니다. 인간의 인식은 불완전하고 모형도 이러한 불완전성을 그대로 반영한 결과이기 때문이죠.

예를 들어 아주 객관적이고 합리적인 방법으로 대학교 평가 모형을 만들었다고 해봅시다.

전국 대학 평가에서 서울대 순위가 5위였다고 한다면 여러분은 이 결과를 받아들일 수 있을 런지요. 아마 대부분의 사람들이 이게 뭐야. 이거 뭐가 잘못된 게 아니냐하고 의심할 수밖에 없을 것입니다. 또 아무리 좋은 개념의 접근도 데이터를 확보하지 못하면 그 변수를 사용하지 못합니다. 이러한 제한성 때문에 다양한 가치를 고려하기보다 확률적 가능성을 효율적으로 판단하는 데 그 목적을 둘 수밖에 없습니다.

이 책의 재미있는 점은 이런 모형의 위험성을 알리는 대신 또 이용하는 방법도 알려줍니다. 모형이 어떤 변수로 이루어졌는지 알면 역이용이 가능하다는 팁을 알려줍니다.

미래에는 많은 부분에서 기계에게 평가 받는 사회에 살고 있을 거라는 생각이 듭니다. 데이터가 쌓이게 되고 하드웨어가 발전해 나가면 당연히 사람들은 이것저것 시도를 할 것입니다.

AI가 학생들에게 성적이나 입학 상담을 코치해 주고, 회사 입사에서도 AI가 면접 대상자를 골라 주고 업무 평가도 AI한테 받는 날이 그리 멀게 느껴지지 않습니다.

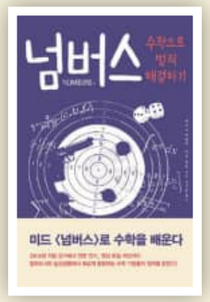
지금까지 우리는 빅데이터나, 사물 인터넷이 가져올 변화에 주목했습니다. 반면에 그런 변화가 몰고 올 그늘에 대해서는 크게 얘기하지 못했던 것 같습니다.

앞에서도 말씀드린 것처럼 이런 모형은 정보화가 발달함에 따라 생기는 새로운 문제는 아닙니다. 하지만 자신의 삶이 기계적 알고리즘으로 한 번에 판단되는 시대가 오면, 그 알고리즘의 사각 지대에 있는 사람은 새로운 신분제도처럼 넘을 수 없는 벽이 만들어지는 결과를 낳을 수 있을 것 같습니다. 과연 미래는 어떤 모습일까요.



넘버스

케이스 데블린, 게리 로든 저
정경훈 역
바다출판사



미드를 보면서 밤을 샌 적이 있으신지요. 저도 금요일 퇴근 시간 지하철에서 보기 시작한 미드에 주말까지 빠져들곤 했습니다. 그 중 통계에 관심 있는 사람이면 한번쯤 '어 이런 것도 있어'하는 미드가 있습니다. 바로「넘버스」입니다. 시즌6까지 나온 것으로 알고 있는데 저는 시즌5의 중간까지 보았습니다.

넘버스는 수학을 교육하기 위해서 만든 드라마라기보다는 대중적인 재미를 위해 만들었습니다. 그렇다 보니 수학을 깊이 있게 다루기보다는 스토리상 문제 해결의 열쇠로 들어가는 형식을 취하고 있습니다.

우리가 홈즈시리즈에 매료된 것처럼 범인을 추적해가는 과정은 흥미진진합니다. 그래서 수사물형태의 미드가 많이 있습니다. 뼈를 가지고 수사하는 '본즈', 심리 수사인 '멘탈리스트', 프로파일링 수사기법의 '크리미널 마인드', 과학 수사물 'CSI' 등이 있습니다.

CSI는 증거 기반의 추리입니다. 새로운 증거가 발견될 때마다 사건의 실체가 하나씩 벗겨지는 방식입니다. 전혀 예상하지 못한 곳에서 범인이 남기고간 단서를 잡아서 범인을 추적하는 재미가 있습니다. 크리미널 마인드는 범인의 프로파일을 추적해나가면서 범인 리스트에서 그 범위를 좁혀가는 방식입니다.

예를 들어 이런 사건은 50대 중반에 백인으로서 어렸을 때 가정 학대를 받았고 비슷한 사건으로 교도소에 간적이 있으며 전기관련 업종에 종사하고 있을 것이다라고 계속 범위를 좁혀가는 방식입니다.

넘버스는 패턴 추적에 가깝습니다. 범행이 지속됨에 따라 범인은 어떠한 발자취를 남깁니다. 그런 흔적에서 패턴을 읽고 추적하는 방식입니다.

이 책은 사건 해결에 실마리로 사용될 수학이나 통계이론에 초점을 맞추었습니다. 그래서 드라마적인 요소를 기대한 독자는 실망할 수도 있을 것 같네요. 하지만 수학이나 통계이론에 관심이 있다면 이것이 사건해결에서 어떤 힌트로 작용할 수 있는지 아는 것은 매력적으로 다가올 수 있습니다.

제가 드라마에서 인상 깊게 본 장면이 이 책에서도 첫 장으로 다루고 있습니다. 드라마 속에서 찰리는(수학천재) 앞으로 일어날 범죄 지점을 예측하려고 하지만 번번이 실패하고 맙니다. 절망감에 빠져 있던 찰리는 스프링클러에서 떨어지는 물을 보고 새로운 접근 방법을 생각해 냅니다. 다음 사건이 어디서 일어나는지 예측하기는 힘들지만 범인이 어디쯤 살고 있는지 추정은 가능하다는 것이죠.

수사 물을 좋아하는 사람이면 수사 물에도 공포물처럼 몇 가지 법칙이 있는 것을 알 수 있습니다. 범인은 범죄현장에 다시 나타난다거나, 범인이 잘 아는 장소에 시체를 유기한다거나, 또 자신의 위치가 발각되는 것이 두려워 자신의 거주지에서 일정거리를 두고 범행을 저지른다는 것들입니다. 이런 점에서 보면 스프링클러에서 떨어지는 물을 보고 수학적으로 범인의 거주지를 예측하는 것은 재미있는 접근법으로 다가옵니다.

개인적으로 이 책에서 아쉬운 점은 사건의 해결포인트가 되는 수학이나 통계이론으로 구성했으면 더 재미있지 않았을까하는 생각이 들었습니다. 게임이론, 카지노, 암호 등을 같이 다루다 보니 이 책만이 가질 수 있는 장점이 희석된 느낌도 들었습니다. 이런 개념들이 일반적인 통계관련 서적에서 많이 다루어지는 내용들이다보니 이 책만의 개성이 약해진 느낌이 들었습니다.

우연의 설계

뉴 사이언티스트 편
김성훈, 마이클 브룩스 역
반니



개인적으로 통계관련 서적 중 재미있게 읽는 주제는「우연」입니다. 우리는 본능적으로 어떤 일이 일어나면 원인을 분석하는 데 익숙하죠.

우리 주변에서 일어나는 다양한 일들을 우연의 관점에서도 해석이 가능합니다. 우연의 관점에서 바라보면 세상의 다양한 일들이 새롭게 해석되어지고 다른 방향으로 해결점이 생각나기도 합니다. 이런 얘기가 생각납니다. 임진왜란이 일어나지 않았다면 우리는 이순신 장군에 대해 알 수 있었을까. 반대로 생각하면 어느 시대나 이순신 장군 같은 인물은 존재하지만 그런 인물을 나오게 하는 환경이나 사건이 만들어지지 않는 한 부각될 수 없을 거라고요. 어쩌면 우리가 아는 인물이 로또를 몇 번 맞은 사람처럼 통계적으로 최극단치에 존재하는 인물이 아닐까하는 생각도 듭니다.

이 책은 주제별로 23인의 전문가가 쓴 칼럼을 묶어서 한 권의 책으로 엮었습니다. 그렇다보니 책 내용이 어려운 부분이 좀 있습니다. 하지만 2장「우연이 뇌에 미치는 영향」은 누구나 재미있게 읽을 수 있는 부분 같습니다.

「행운의 조건」은 행운을 자주 발견하는 사람들에 대한 특징에 대한 연구 결과를 다루고 있습니다. 행운을 자주 만나는 사람들은 인적 네트워크를 잘 형성하고 삶을 편안한 시각에서 바라보고 기회를 포착하는 능력이 뛰어나다고 합니다.

이런 예는「가위바위보」편에서 더 잘 이해 할 수 있습니다. 가위 바위 보 게임에서 이길 가능성을 1/3로보고 임한다면 승률이 별로 좋지 않다고 합니다. 왜냐하면 가위바위보 게임을 대할 때 사람들은 나름 어떤 생각을 가집니다. 예를 들어 '바위를 두 번 냈으니까 다음에는 바위가 아닐 거야' '저 친구는 초반에 가위를 내는 경향이 높아'와 같이 사람들은 어떤 패턴을 예측하려는 경향이 강하다는 것입니다. 이런 단순한 게임에서도 이기는 행운은 랜덤하게 주어지기 보다는 기술이 필요하다는 것을 보여줍니다.

이 책은 6개의 많은 얘기들을 들려주고 있습니다. 개인적으로 마지막 부분에 들려주는 부분이 인상 깊게 남습니다.「길을 잃어 보는 것도 좋아」

우리는 GPS기능 덕분에 이제 길을 잃는 경험을 하지 못할지도 모릅니다. 휴대폰만 켜면 지금의 위치를 알 수 있기 때문이죠. 우리는 낯선 곳을 여행한다고 흥분하지만 결국 누군가 소개한 장소에만 방문하고 누군가 추천한 물건을 구매합니다. 그리고 우리는 새로운 경험을 했다고 얘기합니다. 세상에 데이터가 많이 쌓일수록 우리는 정보가 인도해준 선택을 할 가능성이 높아 보입니다. 사람들은 불확실성을 좋아하지 않기 때문입니다. 하지만 불확실성은 새로운 즐거움을 안겨주기도 합니다.

저는 그런 경험을 한 적이 있습니다. 월드컵 축구 경기가 있던 날 새벽에 일어나 인터넷으로 경기를 관람했습니다. 그런데 인터넷 중계 장면이 지상파 TV중계보다 몇 초가 늦었습니다. 상대편으로 공을 몰고 가는데 그 순간 아파트에서 '아아아' 하는 탄식이 흘러나옵니다. 몇 초 뒤 상대편이 공을 가로채어 버립니다. '오' 하는 소리가 나고 몇 초가 지나면 공이 골대를 살짝 비켜서 날아갑니다. '와' 소리가 나고 몇 초가 지나면 득점을 합니다. 그때 그런 생각이 들었습니다. 미래를 아는 초능력이 있다면 그것은 축복이 아니라 불행일 거라고요. 우연은 세상을 재미있게 만들어줍니다.

통계를 아는 것은 새로운 즐거움과 행운을 부르는 방법일까요. 궁금하시면 이 책에 한 번 도전해 보시길 바랍니다.

김준래 ◆ 통계의창 객원기자

미리가보는 평창 동계올림픽

세계인의 축제인 '2018 동계올림픽 대회'가 두 달 앞으로 성큼 다가왔다. 오는 2월 9일부터 25일까지 강원도 평창을 넘어 전 세계를 뜨겁게 달굴 동계올림픽에 대해 전 세계의 관심이 모아지고 있다. 특히 평창 동계올림픽은 지난 1988년 서울에서 열렸던 하계올림픽 이후 30년 만에 우리나라에서 개최되는 올림픽이라는 점에서, 이번 행사에 적극 참여하고자하는 우리 통계인 들의 의지를 이번 기사에 담아 보았다.

평창 동계올림픽으로 그랜드슬램 달성하다

평창 동계올림픽 개막을 앞두고 가장 관심을 끄는 부분은 이번 대회에 과연 몇 개국이나 참가를 하느냐는 점이다. 평창 동계올림픽 조직위원회(이하 조직위)의 발표에 따르면 지난달 31일 접수된 예비 엔트리 등록 결과 92개국이 참가 의사를 밝힌 것으로 나타났다. 이는 지난 2014년에 열렸던 소치 동계올림픽의 참가국 88개국보다 4개 나라가 늘어난 것으로서, 현재 계획대로라면 최다 참가국 기록을 경신할 수 있을 것으로 보인다.

최다 국가 참가가 예상되는 만큼 대회 참여 인원도 어마어마할 것으로 보인다. 각 국의 참여 선수와 임원, 그리고 보도진을 다 합치면 90여 개국 2만 6000여명의 인원이 평창을 방문할 것으로 조직위 측은 전망하고 있다. 참가국 숫자와 더불어 눈길을 끄는 부분은 동계올림픽 유치를 통해 우리나라가 그랜드슬램을 달성한 국가가 되었다는 점이다. 그리고 기대 이상의 경제적 유발 효과도 누릴 것으로 예상되어 평창 동계올림픽의 전망을 밝게 하고 있다. 그랜드슬램 달성이란 동계올림픽을 비롯하여 하계올림픽과 월드컵, 그리고 세계육상선수권 등 4대 국제스포츠대회를 모두 개최한 경우를 말한다. 1988년 서울올림픽을 시작으로 2002년 한·일 월드컵과 2011년 세계육상선수권대회를 거쳐 2018년 평창 동계올림픽을 끝으로 그랜드 슬램을 달성하게 된 것이다. 이에 대해 조직위의 관계자는 "여태까지 그랜드슬램을 이룩한 국가는 프랑스, 독일, 이탈리아, 일본 등 4개국뿐이었다"라고 언급하며 "우리나라는 세계에서 5번째로 그랜드슬램을 달성한 국가가 되었다"라고 말했다. 경제적 유발 효과가 크다는 점도 고무적이다.

올림픽 같은 국제적 행사를 경제 효과로 환산하는 방법은 크게 두 가지인데, 첫 번째는 직접적 효과를 파악하는 것이다. 투자와 소비가 생산을 유발한다는 점을 고려할 때 직접적 효과는 21조 1천억 원에 달할 것으로 예측되고 있다. 두 번째 방법인 간접적 효과는 더 규모가 크다. 국가 이미지 제고 효과와 추가 관광 효과를 각각 11조 6천억 원과 32조 2천억 원 정도로 추정하고 있기 때문에 총 간접적 효과는 43조 8천억 원에 달할 것으로 조직위는 전망하고 있다.

단위 : 억원

구분	경제적 효과
직접적 효과(올림픽 관련 직접적 투자·소비효과)	21조1000
(투자효과)	(16조4000)
(소비지출효과)	(4조7000)
간접적 효과(올림픽 개최 이후 10년 동안의 효과)	43조8000
(세계적 관광지 부상에 따른 추가 관광 효과)	(32조2000)
(국가 이미지 제고 효과)	(11조6000)
총 경제적 효과	64조9000

평창 동계올림픽 개최에 따른 경제적 효과 〈현대경제연구원〉



빙상종목의 선전으로 동계스포츠 강국 중 하나로 등극

평창 동계올림픽 대회의 종목은 7개의 설상(雪上) 경기와 5개의 빙상(氷上) 경기, 그리고 봅슬레이처럼 썰매를 타는 종목으로 이루어져 있는 3개의 슬라이딩 경기로 이루어져 있다. 이 외에도 이번 평창올림픽에서는 ▲스노보드빅에어 ▲매스스타트 ▲컬링 믹스더블 ▲알파인 스키 혼성 단체전 등 총 6개 세부 종목이 추가되어 역대 대회에서 가장 많은 혼성 종목의 경기가 진행될 예정이다.

우리나라는 이번 동계올림픽에서 금메달 8개와 은메달 4개, 그리고 동메달 8개로 종합순위 4위를 달성한다는 목표를 세웠다. 하지만 역대 동계올림픽에서의 메달 순위를 살펴보면 최소 금메달 9개 이상을 따야 안정적인 4위를 확보할 것으로 예상되고 있다.

이를 위해 우리나라는 평창 올림픽에서 빙상 종목에서만 6~7개의 금메달을 목표로 하고 있다. 특히 강세를 보이고 있는 쇼트트랙에서 4개의 금메달을 노리고 있다. 동계올림픽하면 쇼트트랙, 대한민국하면 쇼트트랙이 떠오를 정도로 쇼트트랙은 이미 우리나라의 명실상부 효자종목으로 꼽히고 있다.

3년 전 러시아 소치에서 개최되었던 2014 소치동계올림픽에서는 8개의 메달 중 5개의 메달을 쇼트트랙에서 획득했고, 그 이전 대회인 2010년 벤쿠버 동계올림픽에서도 14개의 메달 중 8개의 메달을 쇼트트랙이 가져갔다. 또한 2006년 토리노 동계올림픽에서는 10개의 메달 중 9개의 메달을 쇼트트랙에서 획득하는 엄청난 성과를 거두기도 했다.



쇼트트랙 〈평창동계올림픽조직위원회〉

쇼트트랙과 관련한 역대 결과를 살펴보면 풍성한 통계 기록들이 많다는 것을 알 수 있다. 우리나라는 92년 알베르빌 동계올림픽부터 2010년 밴쿠버동계올림픽까지 총 19개의 금메달을 획득했는데, 이 중 우리나라에 최초로 금메달을 안겨준 선수는 김기훈 선수다. 그는 쇼트트랙이 첫 정식종목으로 채택된 92년 동계올림픽에서 1000m에 출전하여 최초로 금메달을 획득했고, 2006년 토리노 동계올림픽에서는 진선유 선수가 우리나라 선수로는 최초로 단일대회 금메달 3개를 거머쥐며 3관왕의 영예를 안기도 했다.


스피드스케이팅
〈평창동계올림픽조직위원회〉

진선유 선수가 한 대회 3관왕의 영예를 안았다면, 전이경 선수의 경우는 가장 많은 금메달을 보유하고 있는 것으로 유명하다. 그녀는 94년 릴레함메르 올림픽과 98년 나가노 올림픽에 연거푸 출전하여 총 금메달 4개와 동메달 1개를 목에 걸었다.

이처럼 우리나라가 동계올림픽의 강자로 떠오르는데 있어 쇼트트랙이 결정적 역할을 했다면, 얼음판 위에서 스케이트를 타고 속도를 겨루는 경기인 스피드 스케이팅은 앞으로 발전 가능성이 더 많은 종목이라 할 수 있다. 쇼트트랙에 비해 화려한 성적은 아니지만, 우리나라 선수들도 충분히 세계무대에서 통할 수 있다는 가능성을 보여준 종목이 바로 스피드 스케이팅이기 때문이다.

1992년 알베르빌 동계올림픽에서 김윤만 선수가 1000m에서 한국 최초로 은메달을 획득한 것을 시작으로 2006년 토리노 동계올림픽에서는 이강석선수가 500m에서 동메달을 추가하여 금메달에 대한 기대를 한껏 높여 주었다.

그런 기대감은 2010년에 와서 가능성이 아닌 현실이 되었다. 남자 500m 단거리에서 모태범 선수가 우리나라 스피드스케이팅 사상 최초를 금메달을 안겨주었고, 바로 다음날 이상화 선수가 여자 500m 단거리에서 금메달을 추가하며 대한민국에 큰 기쁨을 선사했다. 단거리에서만 빛을 발한 것이 아니다. 장거리에서 이승훈선수가 5000m에서 은메달, 10000m에서 금메달을 획득하여 아시아 선수로는 최초로 장거리에서 우승하는 기염을 토했다. 특히 이상화 선수의 경우는 다음 대회인 2014 소치동계올림픽에서 500m 단거리의 금메달을 차지하며 스피드스케이팅에서 2연패 달성이라는 쾌거를 이루기도 했다.

스피드 스케이팅과 관련하여 한 가지 흥미로운 사실은 이번 평창 동계올림픽부터 새로운 종목이 도입된다는 점인데, 바로 매스스타트라는 종목이다. 30명의 선수가 동시에 출발하여 남자는 35바퀴, 여자는 25바퀴를 도는 장거리 경기로서 레인구분이 엄격한 기존의 종목과 달리 자리싸움이 가능하기 때문에 쇼트트랙과 스피드스케이팅을 합친 것과 비슷하다는 평가를 받고 있다. 특히 이 종목은 앞에서 소개했던 2010년 밴쿠버올림픽의 10000m 금메달리스트인 이승훈선수가 절대강자로 군림하고 있어 이번 평창 동계올림픽에서도 좋은 성적을 낼 것으로 기대를 모으고 있다.

슬라이딩과 설상 종목에서도 선전할 것으로 기대

빙상 종목의 선전으로 과거 동계스포츠의 불모지라 여겼던 우리나라도 최근 들어서는 강국 중 하나로 여겨지고 있다. 다만 빙상 종목에 비해 드라이빙 종목이나 설상 종목에서는 단 한 개의 메달도 따지 못해 종목별 불균형이 심한 점은 아쉬운 대목이다.

그러나 이번 평창 동계올림픽에서는 그런 불균형이 많이 해소될 것으로 보이는데, 특히 드라이빙 경기에서 좋은 성과를 낼 것으로 보여 이목이 집중되고 있다. 이 경기에는 썰매 3총사로 불리는 봅슬레이와 루지, 그리고 스켈레톤이 포함되어 있는데, 관심이 모아지고 있는 종목은 단연 스켈레톤이다.스켈레톤은 우리 국민들에게는 상당히 생소한 종목이다. 선수층도 얇고 시설도 없었기 때문에 세계 스포츠 수준과는 현격한 차이를 보였던 것이 사실이다. 하지만 기대주인 윤성빈 선수가 등장하면서 현재 세계 1, 2위를 다투고 있는 상황이라 이제는 평창 동계올림픽에서 가장 확실한 금메달 후보로 꼽히고 있다.

올림픽에는 남, 여 종목을 통틀어 두 개의 메달이 걸려 있는데, 드라이빙 경기의 특성상 체중 규정이 까다롭다는 점을 주의해야 한다. 썰매와 선수 무게를 합쳐서 남자는 115kg, 여자는 92kg을 넘으면 안 되기 때문에 체중조절도 경기의 승패를 좌우할 주요 요인으로 작용할 전망이다. 조직위는 평창 동계올림픽에서 스켈레톤 종목을 시청할 관중들에게 관전 포인트를 정리하여 제공하고 있다. 기대종목이기는 하지만 아무래도 생소한 경기라 무엇을 주목해야 할지를 미리 알고 경기를 보라는 뜻이다. 조직위가 제시하고 있는 스켈레톤 종목의 포인트는 선수마다의 스타트 기록을 챙겨보라는 것이다. 이론적으로만 보면 스타트가 빠를 때 결승점에 도달할 때의 가속력이 가장 높아지기 때문에 출발선에서 출발한 선수들의 속도를 주시해 보면 기록을 대략 점칠 수 있다.

한편 빙상경기로 편성은 되어 있지만, 빙상 종목보다는 한편의 게임 같은 동계올림픽 종목도 있다. 바로 '얼음 위의 체스'라고 불리는 컬링 종목이다. 1996년 동계올림픽 이후 인기 종목으로 주목을 끌고 있는 컬링은 박진감 넘치는 다른 빙상 종목들과는 달리 체스를 두듯이 시종일관 머리를 써야 하기 때문에 또 다른 재미를 안겨주고 있다. 컬링은 각 팀에 소속된 4명의 선수들은 경기장에서 밑면이 오목하고 손잡이가 달린 둥근 돌(스톤)을 티(tee)라고 하는 목표 지점까지 밀어 보내는 경기다. 약 18.6kg의 스톤이 얼음 위를 미끄러져 가면, 빗자루를 든 두 명의 선수가 빗질을 연신 해가며 얼음 위의 얼음조각이나 눈가루 등을 제거해 가면서, 동시에 얼음을 녹인다. 얼음이 녹으면 스톤의 속도가 더 빨라지기 때문이다.


스켈레톤 〈평창동계올림픽조직위원회〉

경기의 승패는 가능한 한 원의 중심에 가깝게 돌을 밀어 보낸 팀이 승리하게 된다. 각 선수들은 상대편의 선수와 번갈아 돌을 2개씩 미끄러뜨려 보낼 수 있다. 상대 팀의 돌보다 티에 가깝게 돌을 보내면 그 때마다 1점씩 얻게 된다. 이에 대해 대한컬링경기연맹의 관계자는 컬링의 장점에 대해 "다양한 작전 구상을 통해 두뇌 회전을 좋게 하고, 상대팀과의 심리적 요인으로 인해 통찰력을 배양할 수 있다"라고 설명하며 "특히 책임감과 협동심이 균형을 이뤄야 경기에서 승리할 수 있기 때문에 팀워크가 어느 경기보다 중요한 종목"이라고 강조했다.

최용준 ◆ 인간개발연구원 교육팀장

흰 눈 위의 자유, **스노보드를 만나다**



내가 그녀를 만난 건 2월 중순 어느 날이었다. 그런데 이상하게도 그녀는 주말마다 다른 약속이 있다고 말했다. 서로에 대해 알아가기 시작하던 그 무렵, 직장마저 서로 가까웠던 그녀와 잠깐의 외근에도 커피 한 잔씩 하곤 했는데 주말만 되면 그녀는 사라졌다. 그녀는 스노보드를 탔다.

어린 시절 서태지와 아이들의 '필승'뮤직비디오에서 본 스노보드는 충격적이라 표현할 만큼 멋있었다. 언젠가는 스노보드를 타봐야겠다는 막연한 결심을 했다. 한 때는 비용이 많이 들 것이라는 핑계로 미뤄오던 스노보드를, 그녀와 함께하고 싶다는 이유로 진지하게 고민하게 되었다. 부끄러운 고백이지만 나는 지금도 롤러스케이트를 타지 못한다. 발밑이 고정되어 있지 않은 상황이 무섭다. 과연 내가 스노보드를 탈 수 있을까하는 걱정과 함께 겨울은 다가오고 있었다.

결국 참다못한 그녀의 후배가 한 시즌동안 무상으로 대여해주기로 하며 스노보드 준비에 박차를 가했다. 10년 넘게 스노보드를 탄 그녀는 한 명의 유격 조교를 붙여주었다. 그 유격 조교에게 처음으로 받은 조언은 바로 장비에 대한 부분이었다.

Tip 1 장비 렌탈? 구입?

스노보딩에 필요한 장비는 렌탈이 가능하지만 스노보드를 제대로 즐기고자 한다면 본인의 장비를 구입하는 것이 좋다. 렌탈의 경우 매번 다른 장비와 다른 세팅으로 접하게 되어 실력 향상에 방해가 된다.

장비 구입 시 상급자용 장비에 욕심은 부리지 말고 조작성이 수월한 입문자용 장비로 구입하는 것이 좋다. 스노보드 매장의 입문자용 기획전은 피하고 지인의 도움을 구하거나 "헝그리보더"나 "ATS 카페"와 같은 스노보드 커뮤니티에서 조언을 구하는 것이 좋다.



막상 스노보드를 타기로 마음을 먹으니 상당히 바빠졌다. 그녀에게 바로 멋진 모습을 보여주고 싶었기 때문이다. 뮤직비디어 '필승'에 나오는 장면처럼 멋진 점프를 보여주고자 스노보드 관련 커뮤니티들을 드나들며 정보를 수집했고, 김현식 프로와 박정환 프로의 강습영상을 대사마저 외울 정도로 수없이 반복해서 봤다. 어느새 나는 머리로서는 완벽한 프로의 기술을 구사하는 수준에까지 도달할 수 있었다.

Tip 2 정보수집요령 – 넘치는 정보 중 내게 필요한 것만 챙기는 법

스노보딩의 기초습득은 독학보다는 리조트에서 제공하는 기초강습을 추천한다. 장비의 올바른 착용법과 최초의 턴을 성공하기까지의 절차를 단계별로 배울 수 있다. 기초 강습을 받은 후 강습영상과 더불어 많은 정보를 접하는 것이 좋다.
유튜브나 비메오에는 수많은 강습영상이 있고 스노보딩에 관한 정보가 넘쳐난다. 단편적인 지식이나 짧은 팁 위주의 강좌보다는 체계적으로 구성된 내용을 통해 정보를 얻는 것이 좋다. 또한 최근 유행에 편중된 정보만 접하여 정보를 편식하기보다는 최근 5년 간의 자료를 두루 살펴보는 것이 좋다. 추천하는 검색 키워드는 "김현식", "박정환", "아이자와 모리오"이다.



내 인생의 첫 스키장에는 비가 내렸다. 슬로프는 상상 속의 흰 눈이 아닌 녹아버린 빙수와 같았다. 머리속에서는 나는 카빙턴과 알리를 하고 있었지만 바인딩을 체결 후 일어설 수조차 없었다.

그녀가 보는 앞에서 초급 슬로프의 삼분의 이를 엉덩이로 내려왔다. 슬로프의 반을 넘게 내려와서야 겨우 일어설 수 있었다. 한번 그렇게 내려오자 몸은 거의 녹초가 다되었다. 도저히 그녀의 얼굴을 보지 못했다. 불행인지 다행인지 내 옆에는 유격조교가 붙어서 하루가 저물도록 나를 끌고 다녔다. 그렇게 하루가 끝나갈 무렵 마침내 턴을 한번 성공하였고 나를 보던 그녀는 물개박수를 쳤다.

Tip 3 부츠의 선택

어떤 장비보다도 편안함이 우선되는 것이 부츠이다. 부츠는 묶는 방법에 따라 운동화처럼 끈으로 되어 있는 끈부츠, 다이얼을 돌리면 조여지는 보아부츠, 양쪽의 끈에 달린 손잡이를 당겨서 조이는 퀵레이스 부츠가 있다. 끈부츠는 한 번 묶으면 느슨해지지 않는 장점이 있으나 부츠를 착용하는데 많은 힘이 들어가니 보아부츠나 퀵레이스 부츠를 추천한다. 이 두가지 부츠는 느슨해질 때 한 번 더 조여주면 되는데 조이는 방법이 매우 수월하다.



▲ 끈부츠



▲ 보아부츠



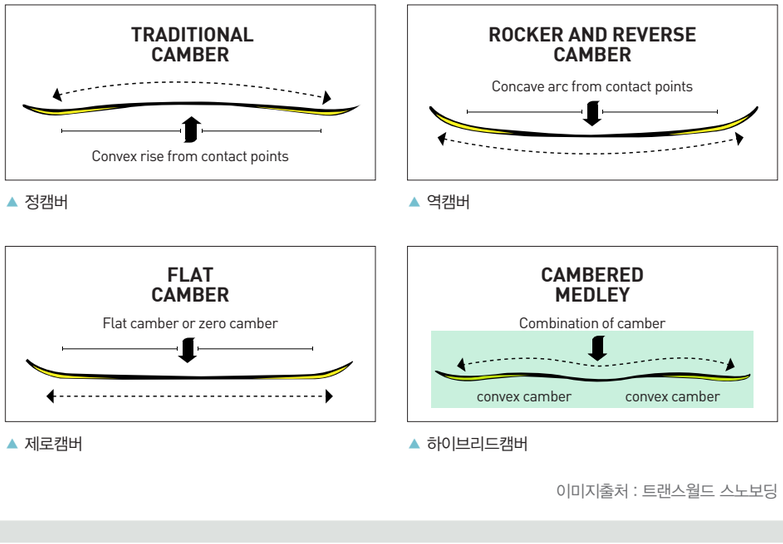
▲ 퀵레이스부츠

제품에 따라 단단한 정도가 매우 차이가 나는데 소프트한 부츠는 착용감이 좋고 발목사용에 용이하고 하드한 부츠는 강한 프레스를 주는데 용이한 장점이 있다. 입문 시에는 너무 단단한 부츠는 피하도록 한다.

턴을 한번 성공하자 '필승'뮤직비디오에 나오는 스노보더가 부럽지 않았다. 턴을 성공한 쾌감은 처음 자전거를 탔을 때, 처음 당구에서 쓰리쿠션을 성공했을 때와 비교할 수 없을 정도였다. 나는 그렇게 스노보드에 빠져 들었다.

최초 내 데크는 정캠버에 151cm인 데크였다. 데크는 중간 부분의 형태에 따라 구분할 수 있는데 데크의 중앙이 위를 향한 아치형의 정캠버와 아래를 향한 역캠버, 중앙 부분이 평평한 제로캠버(플랫 캠버)와 여러개의 아치가 합쳐진 하이브리드 캠버가 있다.

데크



▲ 정캠버 ▲ 역캠버

▲ 제로캠버 ▲ 하이브리드캠버

이미지출처 : 트랜스월드 스노보딩

정캠버인 데크는 데크 자체의 탄성을 활용하기 좋고 데크의 가장자리(엣지)를 활용하기 좋아 라이딩에 적합하다. 역캠버인 데크는 역엣지(보드의 진행방향과 엣지의 방향이 일치하지 않을 때 설면에 걸려 넘어지는 현상)의 위험을 줄일 수 있지만 엣지를 효율적으로 활용하기에는 제약이 있어 지빙이나 트릭에 적합하다. 데크의 중심부가 평평한 제로캠버는 정캠버와 역캠버의 중간적인 성격을 가지고 있다.

스탠스



▲ 덕 스탠스 ▲ 프론트 스탠스

이미지출처 : 라이더스파크

적절한 데크의 길이는 신장과 체중을 고려해서 결정하는데 입문 시에는 본인의 신장에서 15~20cm정도 짧은 데크를 결정하는 것이 좋다. 내 키는 175cm이다보니 다소 짧은 데크를 타게 되었는데 의외로 이 조건이 나에겐 큰 도움이 되었다. 짧은 데크는 조작성이 좋아 기초를 배우기에 용이했고 역캠버 데크로 배웠을 경우에 비해 더 많이 넘어지고 역엣지도 경험했지만 정캠버 데크운용을 통해 기초 확립에 도움이 되었기 때문이다.



스노보딩을 할 때 왼발을 앞으로 세팅을 하면 레귤러 스탠스, 오른발을 앞으로 세팅하면 구피스탠스라고 한다. 주로 오른발잡이는 레귤러스탠스, 왼발잡이는 구피스탠스로 스노보딩을 한다.

보통 스노보드를 시작할 때 바인딩은 앞발을 15°, 뒷발을 0°로 두고 시작하라고 한다. 그 이유가 궁금했지만 모두들 하나 같이 처음엔 그렇게 하는 거라고만 했다. 자세가 불편하여 각도를 바꾸려 하다 혼만 났다. 혼이 난 이유는 "초보는 15°, 0°로 타야하기 때문"이란다.

바인딩을 연결할 때 바인딩의 각도가 둘 다 진행방향으로 돌아가 있으면 프론트 스탠스(포워드 스탠스), 뒷발을 진행방향의 반대방향으로 세팅하면 덕스탠스라고 한다.

프론트스탠스는 상체가 진행방향 쪽을 보게 되어 시야확보에 유리하고 덕스탠스는 데크의 앞뒤를 바꿔 라이딩하는 스위치라이딩에 유리하며 보통 사람이 서 있을 때 양발이 좌우로 벌어지는 것과 같이 세팅하여 중심잡기에 유리하다.

요즘은 프론트스탠스가 유행하고 있지만 굳이 유행에 따르려고 할 필요는 없다. 그리고 입문이라고 해서 무조건 15°, 0°로 시작할 필요도 없다. 본인이 편안한 각도에서 시작하면 된다. 다만 어느 정도의 편안한 각도를 찾았다면 계속해서 바꾸려하지 말고 그 각도를 한 시즌 동안 유지하며 기초를 연습하는 것이 좋다.

나는 유격조교가 무서워 몰래 이리저리 각도를 바꿔가면서 이리저리 테스트를 해보았다. 첫 시즌 21°, –6°에서 편안함을 느꼈다.



Tip 4 데크와 바인딩의 선택

데크는 캠버 이외에도 플렉스FLEX라는 고유의 특성을 가진다. 플렉스는 데크가 휘어짐에 대해 단단한 정도를 의미하는데 소프트할수록 조작성이 좋고 단단할수록 안정감이 높아진다. 입문 시에는 다소 소프트하거나 중간정도의 단단한 데크를 추천한다.

바인딩은 최초 구입 시부터 높은 등급의 제품을 선택해도 무방하다. 특히 데크와 바인딩은 제품의 상태를 판별해줄 지인이 있다면 중고거래도 좋은 방법이다.

설면 위에서의 스노보딩은 매우 특별한 경험이 되었다. 처음 슬로프에서 미끄러져 내려갈 때 남들이 보기엔 걷는 속도보다도 느린 속도지만 내가 느끼는 속도감은 어마어마하다. 또한 몸의 축이 중력 방향이 아닌 슬로프 경사에 대해 수직이 되어야 하는데 이것이 맘처럼 쉽지가 않았다.

Tip 5 안전한 속도 적응요령

실력에 맞지 않게 속도를 내는 것은 위험하지만 스노보딩 자체는 어느 정도 속도가 나와야 제대로 즐길 수 있다. 속도감에 적응하기 위해서는 우선 멈추는 방법을 익혀야 한다. 완만한 경사에서 직활강을 하다 데크를 돌려 멈추는 연습을 한다. 또 평지를 이동할 때에는 한쪽 발만 체결하고 묶지 않은 발로 설면을 차서 이동하는 스케이팅을 하여 설면에서 미끄러지는데 익숙해져보자.
그리고 안전하게 넘어지는 연습도 필요하다. 넘어질 때에는 절대 손으로 설면을 짚지 않는다. 보호대를 믿고 엉덩이로 넘어져야 큰 부상을 피할 수 있다. 브레이킹 요령과 안전하게 넘어지는 요령을 익히면 보다 빠른 속도에서도 불안함을 극복할 수 있다.

스노보드를 경험한 첫 시즌 가장 힘들었던 것은 바로 조바심이었다. 그녀는 항상 초보 슬로프에서 나를 지켜보다 넘어지면 일으켜 세워 주었다. 고마웠지만 한편으론 맘에 부담이 되었다. 빨리 잘 타고 싶었다. 그래서 욕심을 부렸다. 한 번은 지친 상태에서 '한 번 더'를 반복하며 타던 중 다리에 힘이 빠져 중심을 잃고 돌진해오는 초보 보더를 피하지 못하고 충돌했었다. 정말이지 헬멧을 착용하지 않았더라면 큰일 날 뻔 한적이 있었다. 또 한 번은 무리해서 턴을 하던 중 스키어와 충돌했다. 슬로프에서 욕심은 절대 금물이며 항시 사고에 대비하여 패트롤을 호출할 수 있는 연락처를 숙지하는 것이 좋다. 또한 상의 주머니에는 어떤 물건도 넣지 않는 것이 좋다. 왼쪽 가슴 앞의 주머니에 차키를 넣어두고 앞으로 넘어져 갈비뼈가 골절되는 사고는 비일비재하다.



Tip 6 사고 발생 시 대처요령

1. 자주 가는 스키장의 패트롤 호출 번호를 미리 알아둔다.
2. 부상 시 주위 사람에게 도움을 청하고 패트롤을 호출한다.
3. 움직일 수 있다면 슬로프의 가장자리로 가서 장비를 풀어 슬로프 아래로 미끄러져 내려가지 않도록 하고 패트롤의 도움을 기다린다.
4. 충돌 사고 발생 시 목격자를 확보한다.
5. 부상여부를 우선적으로 파악하고 이후 장비의 손상여부도 확인한다.
6. 큰 사고 발생 시 스키장의 CCTV영상을 확보한다.

스노보드 동호회 카페에가 보면 많이 올라오는 글이 사고에 관한 것이다. 그 글들을 읽을 때마다 정말 멋보다도 안전이 우선이라는 것을 느끼고 느낀다.



Tip 7 슬로프 안전수칙

1. 보호대, 헬멧 등 안전 장비는 꼭 챙긴다.
2. 실력에 맞는 슬로프만 간다.
3. 슬로프 중간에 멈춰있지 않는다. 휴식이 필요할 경우 슬로프의 끝으로 이동한 후 휴식한다.
4. 항상 사방을 주시하고, 사람들이 몰린 곳으로 가지 않는다.
5. 위험 감지 시 큰 소리로 주위 사람에게 알린다.
6. 넘어질 때 손을 짚지 않는다.
7. 슬로프에서 데크를 풀어두지 않는다. 어쩔 수 없이 데크를 풀어야 할 상황이 되면 하이백을 세워 데크를 뒤집어두어 슬로프 아래로 미끄러져 내려가지 않도록 한다. 주인을 잃고 슬로프 아래로 미끌어져 내려간 데크는 큰 사고를 유발하는 흉기가 된다.
8. 음주 후 스노보드를 타지 않는다.
9. 슬로프 위에서는 절대 욕심을 부리지 않는다.

스노보드를 접한 지 몇 년이 지난 지금 리프트를 타고 올라가며 기초 연습 중인 사람들을 보면 참으로 보기 좋다는 생각이 든다. 입문 시 내가 초보임을, 그리고 기초연습 중임을 부끄러워했던 생각이 얼마나 의미 없는 것이었는지를 문득 깨닫게 된다. 슬로프 위에서 스노보드를 타는 자체가 즐거움이다. 아무리 추위를 싫어하는 사람일지라도 첫 턴을 성공하면 단언컨대 겨울만을 기다리게 될 것이다.

나는 이제 그녀와 속도를 맞추면서 슬로프를 내려온다. 때로는 그녀의 자세를 보며 잔소리도 한다. 그녀와 나는 평생 스노보드를 같이 타자고 약속을 했다. 내년 봄에 그녀와 나는 결혼을 한다. 봄에 결혼식을 하는 이유라면 겨울 중 주말에는 스노보드를 타야하기 때문이다. 처음에 그녀가 그랬고 이제는 둘이 함께.

Tip 8 슬로프에서의 에티켓

1. 리프트 하차 시 신속하게 하차장을 벗어나야 하며 넘어질 경우 당황하지 말고 신속하게 일어나 이동하도록 한다. 리프트는 안전요원이 멈춰준다.
2. 슬로프 시작점에서 일행과 횡대로 펼쳐 앉아 바인딩 체결하지 않는다.
3. 리프트 슬로프에선 절대 금연이다.
4. 슬로프의 중간에 멈추지 않는다. 슬로프 중앙에서 셀카를 찍지 않으며 사진을 찍어주기 위해 멈춰서있지 않는다.

김현정 ◆ 타샤의책방 대표

추운 겨울 아이들과 도란도란…
요리가 만들어지는 그림책

겨울은 날씨가 추워서 바깥 활동보다는 실내 활동이 많아지는 계절입니다. 그래서 아이와 마주앉아 도란도란 책을 읽으면서 입말 놀이나 간단한 요리 만들기 활동을 할 수 있는 음식 관련 그림책들을 소개하고자 합니다. 이 그림책들을 읽으면서 다양한 맛을 나타내는 단어도 찾아보고 카레라이스, 케이크, 빵, 피자, 김밥, 고구마 요리를 같이 해보면서 포근한 겨울을 보냈으면 좋겠습니다.



01 부모들은 아이들에게 야채 반찬을 먹이려고 애씁니다. 하지만 아이들은 반찬 속에 숨겨둔 야채 한 조각도 귀신같이 찾아 뱉어내곤 합니다. 이럴 때 부모는 어김없이 잔소리와 훈계를 늘어놓게 되지요. 야채 반찬을 상 위에 올리는 날, 〈채소가 최고야〉 그림책을 슬그머니 꺼내 아이에게 읽어주면 어떨까요.

〈채소가 최고야〉에는 우리가 매일 밥상에서 만나는 채소들이 주인공으로 나옵니다. 오늘은 채소들의 달리기 대회가 있는 날이에요. 감자, 브로콜리, 가지처럼 응원하는 채소들도 있습니다. 출발~! 알통이 올통볼통한 마늘이 맨 먼저 뛰어나오지만 파릇파릇 파슬리도 후다닥 달립니다. 순무가 단숨에 쑥쑨 선두로 달려나왔고요. 이런이런~! 달리기하던 단호박이 자전거를 타고 오는 오이와 부딪혀서 데굴데굴 강에 풍덩 빠져버립니다. 달리기 경기하던 토마토, 무, 당근, 셀러리들과 낚시하던 고구마아저씨까지 단호박을 구해내기 위해 다시 힘을 모읍니다. 영차영차! 결승선에서 박수가 터져 나옵니다. 과연 1등을 한 채소는 누구일까요?

이 책에는 '채소를 많이 먹어야 해. 그래야 건강해져.'라는 말은 한마디도 없습니다. 그런 뻔한 훈계는 아이들에게 지루할 뿐입니다. 아이들은 가지각색 색깔과 모양, 맛을 지닌 채소들이 앞서거니 뒤서거니 하며 달리기 실력을 겨루는 광경을 깔깔거리며 볼 겁니다. 그리고 채소들의 화합과 우정, 건강한 몸과 마음에 대해 깨닫게 됩니다. 책을 다 본 후에는 냉장고에서 채소들을 한아름 꺼내보세요. 채소들을 예쁘게 썰고, 카레가루를 푼 물에 채소들을 퐁당 넣어보고요. 그렇게 해서 채소들의 수영대회를 열어보는 거예요. 이름하여 '야채카레라이스' 요리를 만드는 거지요. 창밖에 내리는 하얀 눈도 감상하며 아이와 호호 불며 야채카레라이스를 먹으면 긴 겨울도 따스해질 거예요.

〈채소가 최고야〉
영양가 풍부한 야채 가득 카레라이스

글 이시즈 치히로
그림 야아무라 코지
천개의바람



02 작가는 자신의 아이들이랑 같이 고구마를 캘 때의 경험으로 이 책을 만들게 되었다고 합니다. 그림책을 넘기면 힘들게 고구마의 덩굴을 쭈욱 뽑아 올리는 장면이 시선을 사로잡습니다. 둥근 고구마, 길쭉한 고구마, 큰 고구마, 작은 고구마.... 한 뿌리에서 나온 고구마이지만 어쩜 이리 모양도 크기도 제각각인지요. 너무나 재미있는 건, 고구마들이 말끝마다 "~구마."라는 말을 붙이며 서로 대화를 한다는 겁니다.

'고구마는 둥글구마.' '고구마는 길쭉하구마.' '크구마.' '작구마.' 고구마의 생김새들이 이렇게 재미있구나 하고 책장을 넘겨보니, 더 재미난 이름의 고구마들이 나옵니다. '배불룩하구마' '굽었구마' '털났구마' '험상궂구마' 하나같이 참 다르게 생겼습니다. 이렇게 아이와 입말 놀이를 하면 너무나 재미있답니다.

이제 이 고구마들로 요리를 할 차례입니다. 고구마는 쪄 먹어도 맛있고, 구워 먹어도 맛있습니다. 기름에 잘 튀겨 먹으면 아삭아삭 씹는 맛이 끝내주지요. 그림책을 다 본 후 아이와 함께 다양한 고구마 요리를 해 먹으면서 형제자매들에 대한 이야기도 곁들여 해보세요. 고구마가 싫증이 난다면 감자를 가지고 해볼 수도 있답니다.

〈고구마구마〉
세상에서 가장 웃긴 고구마 말놀이

글 그림 사이다
반달



〈케이크 파티〉

특별한 날을 위한 케이크 만들기

글 그림 이소을
상상박스



03 이 책은 우리 집 세 아이들이 모두 좋아해서 너덜너덜해진 그림책 중 하나입니다. 아이들은 생일이나 기념일 때마다 케이크를 꼭 먹게 되지만, 케이크가 어떻게 만들어지는지는 잘 모릅니다. 〈케이크 파티〉에는 요리를 좋아하는 지니와 남동생 비니를 통해 케이크가 만들어지는 과정을 단계별로 보여줍니다. 주인공 지니가 어떤 요리를 할까 고민하자, 갑자기 오븐 안에서 빵이 '통~,통~,통~' 뛰어옵니다. 빵이 케이크가 되고 싶다며 지니에게 부탁을 하자 지니는 빵의 꿈을 돕기로 합니다. 그러자 냉장고 속 친구들도 케이크가 되고 싶다고 우르르르 튀어나옵니다. 지니는 귤껍질을 벗기고, 바나나를 가지런히 썰고, 딸기와 포도를 목욕시키고, 빵에 생크림을 입히고, 짤주머니로 생크림 쿠션을 만들고, 드디어 멋진 케이크가 만들어지면서 냉장고에서 나온 친구들의 꿈이 이루어집니다. 촛불도 불고, 노래도 부르고, 신나게 춤도 추고, 책을 읽으면서 케이크 친구들과 즐겁게 파티를 하고 싶어집니다.

준비물이 많이 필요한 케이크 만들기가 부담스럽다면

〈구리와 구라의 빵 만들기〉(나카기오 리에코 글/오무라 유리코 그림/한림출판사)에서처럼 밀가루만으로 만들 수 있는 커다란 빵을 만들어 보아도 좋습니다. 주인공으로 나오는 파란 모자와 옷을 입은 구리, 빨강 모자와 옷을 입은 구라는 그림선이 간단해서 아이들도 쉽게 따라 그릴 수 있는 들쥐 형제입니다.

이들은 어느 날 숲 속에서 커다란 알을 발견합니다. 이 알로 무엇을 할까 고민하다가 빵을 만들기로 결정합니다. 하지만 알이 너무 크고 무거워서 집으로 옮겨갈 수가 없습니다. 의논 끝에 집에서 빵 만드는 도구들을 숲에 가져옵니다. 빵 만드는 고소한 냄새가 숲에 가득 풍기고 하나둘 숲속 친구들이 모여들고 함께 나누어 먹습니다. 알껍데기로는 자동차를 만들어 요리도구들을 싣고 집으로 돌아옵니다. 따스한 색채와 아기자기한 동물들, 그리고 감동적인 결말 때문에 아이들이 아주 좋아하는 그림책입니다.

혹시 오븐이 없어서 빵 굽기가 어렵다면, 핫케이크 가루를 커다란 볼에 붓고 달걀과 우유를 넣어 아이와 휘휘 저으세요. 그리곤 커다란 프라이팬에 반죽을 부어 팬케이크를 만들어보세요. 온 가족이 각각 동물들이 되어 둘러앉아 폭신폭신한 팬케이크를 먹으며 이야기를 나눈다면 들쥐 형제가 모험을 한 숲 속에 있는 듯하겠지요?





04 요즘 아이들은 피자를 아주 좋아하지요. 그림책 속 주인공 아이도 피자를 아주 좋아하는 어린이입니다. 공부할 때도 피자를 먹고 싶어 하는 생각뿐입니다. 수학시간에 시간을 배우는 시계모양도 피자로 보입니다. 아이는 집에 오자마자 엄마에게 피자를 시켜달라고 조릅니다.

엄마는 마침 불량패스트푸드에 대한 뉴스기사를 본 뒤라서 직접 만들겠다고 아이 데리고 장을 보러 나섭니다. 피자라고는 식빵 피자 한 번 만들어 본 적 없는 엄마인데 말이지요. 엄마표 피자는 밀가루 반죽부터 시작되어 새콤새콤 소스도 만들고, 피자에 올라가는 토핑도 준비해서 오븐에 구워내었습니다. 하지만 처음 만든 엄마표 피자는 맛이 이상합니다. 곧 다가오는 아이의 생일날, 엄마는 케이크 대신 딸아이가 젤 좋아하는 피자 3단 케이크를 준비합니다. 과연 그 맛이 어땠을까요?

책 맨 뒤에는 '피자 레시피'가 자세히 나와 있어서 가정에서 아이와 함께 피자를 만들어 볼 수 있습니다. 너무 어렵지 않으니 꼭 해보세요. 급할 땐 종이로도 피자를 만들 수 있답니다. 배달되는 피자보다 엄마의 사랑과 정성으로 만드는 피자, 이번 겨울에 아이에게 꼭 만들어보세요.

〈앗 피자〉

서툴러서 더 귀여운, 요즘 엄마의 '피자' 도전기

글 그림 정호선
사계절



05 아이들이 뛰어노는 모습은 참 보기 좋습니다. 근심이나 걱정은 없고 맑고 환한 기운이 온몸에서 퍼져나오거든요. 그런데 그림책 속 피트는 우울한 표정을 짓습니다. 비가 와서 밖에 나가 놀지 못해 속상한 거지요. 아빠는 아들의 마음을 금세 눈치 챕니다. 아들의 심심함을 달래기 위해 '피자놀이'를 생각해 내지요. 아들 피트를 피자반죽처럼 반죽하기 시작하다가, 식탁 위에 올려놓고 진짜 피자를 만드는 것처럼 종이를 오려서 다양한 토핑을 뿌리는 놀이를 합니다.

아이들은 이 책을 보고난 후 피자놀이를 하자며 부모를 조르게 됩니다. 피자놀이로 한 판 충분히 즐겼다면 이번엔 〈돌돌말아 김밥〉(최지미 글그림/책읽는곰)처럼 '김밥놀이'를 해봅니다. 얇은 이불 한 장 위에 아이를 눕혀놓고, 김밥 속 다양한 친구들인 밥보, 단무지씨, 시금치양, 햄맨, 꼬마당근, 달걀지단을 불러모아 올려놓은 흉내를 낸 뒤 이불로 돌돌 말아서 이리저리 굴려주면 됩니다.

늦게 잠드는 아이를 일찍 재우고 싶다면 잠자리에서 꼭 해보세요. 아이들이 현실 속에서 피자나 김밥을 먹을 때마다 부모와 함께 몸놀이로 즐겼던 '피자놀이'와 '김밥놀이'의 순간들을 떠올리며 스스르 눈을 감게 될 테니까요.

〈아빠랑 함께 피자놀이를〉

아빠랑 함께 몸놀이하면서 스킨십을 느껴요

글 그림 윌리엄 스타이그
보림



신동헌 ◆ 통계의창 객원기자

통계로 바라보는 **세상이야기…**

01

멈추면 비로소 보여요, 갭이어



영국에서 처음 시작된 '갭이어(Gap Year)'는 학업을 잠시 중단하거나 병행하면서 봉사, 여행, 진로탐색, 교육, 인턴, 창업활동을 체험하며 흥미와 적성을 찾고, 앞으로의 진로를 설정하는 기간을 말합니다. '미국갭협회'의 연례보고서에 따르면, 2015년의 '갭이어' 학생은 전년보다 22% 증가하였는데, 갭이어 참여 학생 중 사회에서의 경험을 터득하고 개인의 성장을 느끼고자 갭이어를 선택한다는 응답자는 92%로 가장 많았고, 85%는 세계를 여행하며 다양한 문화를 접해보고 싶어서라고 답했습니다. 갭이어는 사실 '욜로'나 '휘게'와 같은 맥락이라고도 볼 수 있습니다. 굳이 세계여행이나 창업이 아니어도 일반화된 기준에서 벗어나 새로운 경험을 통해 진정한 나를 찾을 수 있는 인생의 쉼표 같은 시간을 가져보는 건 어떨까요?

02

노력은 결과와 과정을 빛내는 보석

한국교육과정평가원에서 매년 발표하는 '수능 연도별 응시현황'을 살펴보면 최근 5년간은 약 60만~65만 명이 지원·응시 하였으나, 2017년(2018 수능) 처음으로 59만 3,527명이 지원하며 60만 아래로 감소했습니다. 그렇다면 우리 학생들은 어떤 기대를 가지고 공부 할까요? 통계청에서 발표한 '사회조사 결과'를 살펴보면 대학 이상의 교육을 받으려는 학생의 주된 목적은 '좋은 직업을 갖기 위해서'가 51.1%(2016년)로 가장 높았고, 그 다음으로는 '자신의 능력과 소질 개발'을 위해 공부한다고 답한 학생들이 38.6%였습니다. 이밖에도 '주위의 기대 때문(3.2%)'이라는 대답과 '결혼, 친구관계 등 사회적 유리(4.0%)'가 목적이라 답변한 학생들도 있었는데, 어린 나이에 이러한 이유들로 공부한다는 게 안쓰럽기도 합니다.

03

우리는 틀리지도, 다르지도 않아요!

'한현민'이라는 모델을 아시나요? 얼마 전 타임지 선정 '2017년 가장 영향력 있는 10대 30인'에 당당하게 이름을 올린 만 16살의 한국인 모델입니다. 이 모델의 특별한 점은 나이지리아인 아버지와 한국인 어머니 사이에서 태어났다는 것인데요. 지난 11월 통계청에서 발표한 '2016년 다문화 인구동태 통계'를 살펴보면, 다문화 혼인은 2010년 35,098건에서 지난해 21,709건으로 지속해서 감소추세에 있습니다. 특히 '출신 국적별 혼인 비중'을 보면, 다문화 혼인을 한 외국 출신의 남편 국적은 중국(9.9%)이 가장 많았고, 미국(6.4%), 베트남(2.6%) 순으로 나타났으며, 외국 출신의 아내 국적은 베트남이 27.9%로 가장 많았고, 중국(26.9%), 필리핀(4.3%) 순이었습니다.

04

김장, 김치를 담그고 나누는 문화

김장은 한국의 문자 체계인 '한글'이나 '태극기'와 비교될 정도로 중요하다고 여겨집니다. 김치를 담그고 나누는 것, 특히 공동 작업인 김장은 한국인의 정체성을 재확인시켜주는 일입니다. 요즘도 직접 담그거나 친지에게 받는 사람들이 2015년 91.5%에서 2016년 89.0%로 많은 편이지만, 시판 김치를 구매하는 가구도 8.5%(2015)에서 11.0(2016)으로 늘고 있습니다. 또한, 절임과정의 번거로움을 해소하기 위해 신선배추보다 절임배추를 선호하는 추세가 꾸준히 증가하더니 지난해에는 절임배추(51%)가 신선배추(49%)를 추월했습니다. 올해는 배춧값이 작년 대비 낮아져 전체적인 김장비용은 4인 가구 22만5천 원(11월8일 기준)으로 추정됩니다. 여러분에게 김치 그리고 김장은 어떤 의미인가요?

05

존엄스런 죽음을 위한 준비, 웰다잉법

웰다잉(Well-dying)이라는 말 들어보셨나요? 웰다잉은 '살아온 날을 아름답게 정리하는, 평안한 삶의 마무리'를 일컫는 말로 삶의 마지막을 스스로 미리 준비하는 개념입니다. 보건복지부는 회복 가능성이 없는 환자가 스스로 연명 치료를 중단할 수 있는 '웰다잉법'의 본격 시행을 앞두고 10월 23일부터 내년 1월 15일까지 13개 기관에서 시범사업을 펼치고 있습니다. 서울대병원 암통합케어센터 윤영호 교수팀이 지난해 국내 암환자(1,001명)와 가족(1,006명), 일반인(1,241)명을 대상으로 죽음에 대한 인식을 조사한 결과 조사 대상자의 다수는(75.2%) '죽음과 함께 삶은 끝난다'고 답했다고 합니다. 하지만 90.9%는 '죽음은 고통이 아닌 삶의 완성으로 기억돼야 한다'고 답해 죽음에 대해 긍정적으로 인식하는 것으로 확인되었습니다.

06

쇼핑의 길을 안내하는 내비게이션 소비!



내비게이션 소비란 인공지능, 사물인터넷, 위치기반 서비스 등을 통해 소비자의 쇼핑과정을 파악하여 소비자의 요구에 맞춰 내비게이션처럼 정보를 제공하는 소비를 말합니다. 가장 대표적으로 인공지능을 활용한 챗봇은 수많은 대화 데이터를 기반으로 이용자들의 대화 패턴과 규칙을 분석하고 알고리즘이 형성되어 인공지능을 기반으로 사람과 대화를 나눌 수 있는 모바일 메신저를 말합니다. 그리고 아마존고의 마트시스템 역시 내비게이션 소비의 사례라고 할 수 있습니다. 계산대가 없는 아마존의 오프라인 매장에서는 아마존고 앱을 통해 QR코드를 찍고 입장하여 매장 선반에 있는 제품을 담으면 따로 계산원과 마주하지 않고도 바로 아마존 계정에 등록된 결제수단으로 구매 금액이 청구됩니다.

07

나눔의 미학, 사랑을 실천해요!

추위가 매서워질수록 사람들은 따뜻한 온기를 더 원합니다. 통계청에서 발표한 '2017년 사회조사 결과'를 살펴보면 지난 1년 동안 기부를 해 본 경험이 있는 13세 이상 인구는 26.7%로 2015년 29.9%에 이어 지속적인 감소 추세입니다. 기부의 인식을 바꾸기 위한 색다른 방식도 있는데요, 커피를 사 먹지 못하는 노숙자나 불우한 이웃을 위해 미리 여러 잔의 커피 값을 계산하는 '서스펜디드 커피(Suspended Coffee)'는 이탈리아에서 시작하여 현재는 세계 전역에서 활발히 이뤄지고 있고, 가격은 그대로지만 양이 절반인 제품을 구매하면서 제품 가격의 절반만큼 후원하는 브라질의 '행복을 위한 반(Half For Happiness)' 캠페인이나 우리들에게도 익숙한 '아이스 버킷 챌린지' 등 다양한 기부 방법들이 펼쳐지고 있다고 합니다.



08

연애합시다, 솔로는 외로워요!

'내가 솔로인 이유'에 대해 여성의 26%는 "사람들이 나의 진가를 모르기 때문"이라고 답했고, 남성의 23%는 "연애세포가 죽어서"라고 답했습니다. 많은 솔로들이 '연애는 못 하는 것이 아니라 안 하는 것이다.'라고 말하지만, 추운 겨울, 옆구리가 너무 시리지 않으신가요? 지난 2012년, 크리스마스이브를 맞이하며 외로운 솔로들이 짝을 찾기 위해 이벤트로 개최된 솔로대첩은 많은 이야깃거리를 남겼습니다. 사랑을 시작하기에 정답이 있다면 첫눈에 반하고 그 사람에게 잘 보이기 위해 노력하는 설렘이나 거절당했을 때의 슬픔과 같은 사랑의 과정에서 느끼는 감정들이 아무 의미도 없게 됩니다. 여러분 모두 사랑하는 사람과 따뜻하게 한 해를 마무리하며 보낼 수 있기를 통통 기자가 응원하겠습니다.

09

커피, 홈카페에서 즐겨볼까요?

하루에 커피를 몇 잔 드시나요? 농림축산식품부 통계에 따르면 20세 이상 성인이 1년에 377잔 즉, 하루에 한 잔 이상씩 커피를 마신다고 합니다. 최근에는 단순히 카페가 아닌 집에서 개성에 맞게 커피 머신 등을 이용하여 즐기는 사람이 증가하고 있습니다. 그렇다면 이렇게 과거와 달리 홈카페가 발달할 수 있었던 이유는 첫째, 고급커피에 대한 수요 증가, 둘째, '가성비', 셋째는 '넓은 선택의 폭'입니다. 특히 집에서 커피를 추출하는 방법이 드리퍼, 모카포트, 캡슐이나 에스프레소 머신 등 다양하기 때문입니다. 비록 커피를 사 마시는 것보다 조금 번거로울 수 있지만, 내 입맛에 맞는 커피를 직접 마실 수 있다는 점, 커피를 추출하면서 갖게 되는 즐거움 등을 생각할 때 나만의 홈카페는 어떨까요?



10

월화수목금금금, 아직도 과로사회?

'건설업·금융업 '과로사' 빈번… 960명 산재 신청, 버스·택시기사 등 특례업종 매달 3.6명씩 '과로사' 비극' 연일 과로사에 대한 뉴스가 나오고 얼마 전 국감에서도 근로기준법 개정안이 논의되었죠. 2016년 기준 우리나라 임금근로자의 연평균 2,052시간의 근무시간은 미국(1,789), 일본(1,724), 영국(1,694), 독일(1,289)과 비교했을 때 큰 차이를 보입니다. 이런 상황에서 기업이 직원의 건강을 체계적으로 관리하기에는 부족한 부분이 많은데요, 서울의대 스마트건강경영전략연구실의 '기업의 직원 건강관리 관련 일반 국민 인식조사결과(2017)'을 살펴보면 '과도한 업무 방지 및 충분한 휴식제공(27.1%)'을 첫째로 꼽았습니다. 피곤하지 않은 사회에서 일할 수 있는 환경이 조성되기를 바랍니다.

11

애드테크, 기술 들어갑니다!



"우리가 매체를 통해 하루에 수십 번 보는 것은 무엇일까요?" 정답은 바로 '광고'입니다. 요즘은 소비층이 IT 기기들을 많이 사용하고, 특히 PC 및 스마트폰을 이용하고 있기 때문에 아날로그를 통해 노출되는 광고는 줄어드는 추세에 맞춰 등장한 것이 광고와 IT, 그리고 빅데이터를 접목한 '애드테크(AD+Technology)'입니다. AD Tech는 보통 High Tech로 불리며 동시에 프로그래매틱 광고라고 합니다. 프로그래매틱 광고는 인터넷 이용자의 개인정보가 아닌 이용자가 사이트에 접속하면서 생긴 쿠키로, 이용자의 소비 행태를 예측해 이용자가 원할 것 같은 광고를 선택하여 보여주는 방식을 말합니다. 우리나라는 IT 강국이면서, 인터넷 보급이 높은 나라이기 때문에 관련 기술은 다양하게 발전하리라 생각합니다.

12

우리가 사는 지역의 직업구조는?

통계청에서 발표한 '2017 상반기 지역별고용조사'의 특별·광역시를 제외한 9개 도시군 지역의 직업별 구조를 보면, 시 지역에서는 서귀포시, 당진시, 나주시, 제주시, 영천시가 고용률 상위 5개로 나타났는데요. 취업자의 직업별 비중은 서귀포시(26.1%), 나주시(37.4%), 영천시(33.4%)의 '농림어업숙련종사자'가, 당진시(31.9%)는 '기능·기계조작·조립종사자', 제주시(26.4%)는 '서비스·판매종사자'가 가장 높게 나타났습니다. 5개 시 모두 '관리자·전문가 및 관련종사자' 및 '사무종사자' 비중이 시 지역 평균(각각 20.6%, 17.2%)보다 낮게 나타났고, '농림어업숙련종사자' 비중은 시 지역 평균(6.7%)보다 높게 나타났습니다. 직업별 고용률이 가장 높은 분야는 아무래도 농림어업숙련종사자로 보입니다.

13

제7회 대한민국 SNS 대상 '최우수상'



통계청이 지난 10월 19일, '대한민국 SNS 대상' 시상식에서 중앙부처 부문 최우수상을 받았습니다. SNS로 국민과 활발히 소통하고 있는 기업/공공기관을 찾아 널리 알리고 올바른 SNS 활용 문화 정착을 위해 노력하고 있어 관련 분야 최고 권위의 시상으로 인정받고 있습니다. 특히나 통계청은 작년에 이어 연속으로 수상을 했는데요. 올해는 공공부문 80개, 기업부문 15개 단체가 참가해 총 95개의 단체가 참여하여 총 11개의 기업과 32개의 공공기관이 선정되었습니다. 2011년을 시작으로 올해 7회째를 맞은 '대한민국 SNS 대상'은 헤럴드경제와 한국소셜콘텐츠협회가 주최하고, 과학기술정보통신부와 한국정보화진흥원, 한국블로그산업협회, 한국인터넷전문가협회가 후원하는 행사입니다.

14

엿보는 게 즐거워? 관찰예능 전성시대!

한국갤럽의 설문조사에 따르면, 2017년 9월 한국인이 좋아하는 TV 프로그램으로 MBC의 〈무한도전〉이 1위를 차지했고, 그다음으로 〈썰전〉, 〈삼시세끼〉, 〈나는자연인이다〉, 〈미운우리새끼〉, 〈아는형님〉, 〈효리네민박〉 등이 상위권을 석권하였습니다. 여러분들은 어떤 예능프로그램을 즐겨보시나요? 몇 년 전만 하더라도 리얼 버라이어티 프로그램이 대세였지만, 요즘에는 〈효리네민박〉, 〈삼시세끼〉, 〈미운우리새끼〉 같은 '관찰예능' 포맷의 예능프로그램이 인기입니다. 관찰예능은 제작진의 개입을 최소화하면서 관찰 카메라 형태로 구성된 예능 프로그램을 말합니다. 한국의 관찰예능은 서구의 관찰예능과 달리 간접체험을 할 수 있기 때문에 대리만족이라는 측면에서도 더욱 인기가 있어 보입니다.



15

'클릭' 한 번이면 현관문을 '똑똑'

지름신을 부르는 간편한 온라인 쇼핑, 여러분은 얼마나 자주 이용하세요? 작년대비 온라인 거래액은 14.9%, 모바일 거래액은 29.5% 증가했다고 합니다. 통계청 '2017년 8월 온라인쇼핑 동향' 조사결과를 살펴보면, 일주일에 평균 4.5회 온, 모바일 쇼핑몰 방문하는 것으로 드러났고, 온라인이나 모바일로 구매하는 가장 큰 이유로 '편리함'(53.1%)을 꼽았고, '가격/가치가 좋아서'(49.2), '결제가 편리해서'(28.1), '배송이 빨라서'(21.7), '쇼핑경험/할인혜택이 좋아서'(13.7), '상품구성이 좋아서'(9.6)이 그 뒤를 이었습니다. 이처럼 증가하는 온라인 쇼핑은 여행 및 예약 서비스부터 생활용품까지 다양한 상품을 갖추고 고객들의 지갑을 열고 있습니다. 지금 이 순간에도 클릭하고 계시는가요?



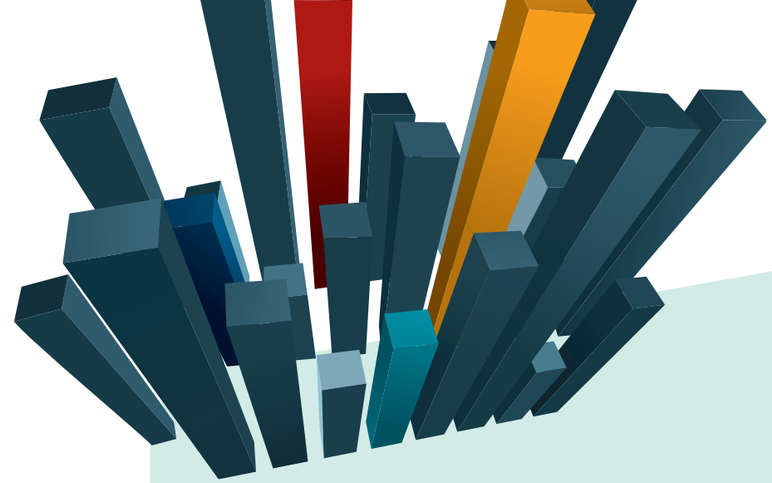
16

우리 사회를 비추는 거울 '사회조사'

'통계' 하면 어떤 생각이 떠오르나요? 많은 분이 '딱딱하다, 지루하다' 등을 얘기할 것 같은데요, '사회조사'를 살펴보면 통계가 다 그렇지 않다는 것을 알 수 있습니다. 사회조사는 1977년 「한국의 사회지표」의 주관적 지표를 보완하기 위한 목적으로 처음 시행했습니다. 초기에는 총 8개 영역(소득/소비, 고용/노사, 교육, 보건, 주택/환경, 사회, 문화/여가, 공안)으로 조사가 진행됐으며, 이후 사회지표 개편과 사회변화에 맞추어 주기적으로 개편됐습니다. 사회조사는 우리가 평소에 불편했던 것, 개선해야 한다고 느꼈던 것들을 수치를 통해 객관적으로 보여줌으로써 더 살기 좋은 세상을 만드는 것에 이바지하고 있습니다. 어렵다고 느껴졌던 통계가 한결 재밌고 또 한편으로는 고맙게 느껴지지 않나요?

통계청, 정부부처, 지방자치단체, 연구기관 등 모든 기관의 마이크로데이터를 한 곳으로...

보다 심도 있고 다양한 분석을 원한다면 지금 바로 MDIS를 클릭해보세요



■ **서비스 소개**

가. 서비스명 : 마이크로데이터통합서비스(MDIS, mdis.kostat.go.kr)

나. 제공통계수 : 총 116종(통계청 45종 및 통계작성기관 71종)

다. 제공형태 : 마이크로데이터(통계에 따라 사람, 사업체, 가구 기반 자료)

기준		주요 통계
통계청	인구 · 가구	경제활동인구조사, 가계동향조사, 국내인구이동통계, 사망원인통계, 가계금융복지조사, 지역별고용조사, 인구주택총조사, 인구동향조사, 생활시간조사, 사회조사 외 7종
	사업체 · 농어가	전국사업체조사, 광업 · 제조업조사, 농가경제조사, 기업활동조사, 농업총조사, 농산물생산비조사, 경제총조사, 어가경제조사, 도소매업조사, 운수업조사 외 12종
	행정통계 및 기타	귀농어귀촌인통계, 영리법인기업체행정통계, 신혼부부통계, 주택소유통계, 퇴직연금통계, 국부통계조사
통계작성기관		전국다문화가족실태조사, 가족실태조사, 자동차주행거리통계, 직종별사업체노동력조사, 보육실태조사, 지상기상통계, 국민여가활동조사, 외래관광객실태조사, 한부모가족실태조사, 청소년종합실태조사 외 61종

■ **서비스 내용**

가. 구분 : 자료의 민감성 정도에 따라 공공용, 승인용, 특수목적용으로 구분 운영

나. 수수료

- 무료 : 공공용 자료
- 승인용/특수목적용 : 선택제 수수료 부과

다. 서비스 방법

- 추출 · 다운로드 : MDIS 포털에서 직접 무료 다운로드
- 원격접근서비스 : 승인 후 이용자가 집 · 사무실 등에서 통계청 서버 접속 후 활용
- 이용센터 : 승인 후 지정된 장소를 방문 · 활용
- 주문형서비스 : 원하는 형태로 통계표 작성 및 제공

■ **문의**

- 연락처 : 재단법인 한국통계진흥원
- 전화 : (02) 512-0167 FAX : (02) 515-0240
- 주소 : (우) 06097 서울특별시 강남구 선릉로 612, 6층
- E-mail : MDIS@stat.or.kr

공공자료의 개방 및 공유 확대

MDIS
[mdis.kostat.go.kr]

통계청 45종 및 통계작성기관 71종의
통계자료로 제공확대



통계청에서 국가통계를 활용하세요!

통계청은 **통계개발·활용**에 필요한 모든 **정보와 도움**을 제공합니다.
다양한 국가통계정보 제공사이트를 활용하세요.

원하는 자료를 직접 분석 및 요청

MDIS
[mdis.kostat.go.kr]

온라인으로 「자료신청」 메뉴를 통하여
공공용 마이크로데이터를 무료로
직접 추출, 분석 가공 가능



국가통계 쉽게 찾기

KOSIS
[kosis.kr]

우리나라 주요 통계를 주제별, 기관별,
명칭별 등으로 알기 쉽게 분류해 제공



국가 발전 상황을 한눈에

국가지표체계
[www.index.go.kr]

국민의 관심이 크고 정책 수립에
활용 가능한 지표



지도 위 통계정보 살펴보기

SGIS
[sgis.kostat.go.kr]

인구, 가구, 주택, 사업체 통계 등 각종
통계를 지도(GIS) 위에서 한눈에 파악



국내 유일의 국가통계교육 전문기관

통계교육원
[sti.kostat.go.kr]

통계작성 및 활용 전문통계과정,
기관맞춤형과정, e-러닝 과정



통계청
Statistics Korea