

서강대학교 정보통신대학원 통계기반 데이터 분석 5주

2019년 4월 정화민 교수

R 데이터를 엑셀파일로 변환

```
#엑셀파일로 저장하기: R데이터를 엑셀(excel) 데이터로 저장하는 기능을 제공하는 여러
개의 패키지가 있지만, 최근에 만들어진 writexl 패키지 사용.
install.packages("writexl")
library(writexl)
# 지난 시간 학습한 mtcars를 엑셀파일로 보내기
writexl::write xlsx(mtcars, path = "c:/RStudio/데이터분석과R프로그래밍/mtcars.xlsx")
#여러분들은 현재 워킹디렉토리를 확인
getwd()
# 워킹디렉토리를 지정
setwd("c:/RStudio/데이터분석과R프로그래밍")
# 엑셀 파일 읽어오는 패키지와 mtcars의 데이터 확인
Install.packages( " xlsx " )
Library(xlsx)
```

read.xlsx("mtcars.xlsx",1)

편차, 분산, 표준편차 공식 실습

#mtcars 엑셀 파일 변환 후 다음 공식 실습

평균

$$m = \frac{x_1 + x_2 + \dots + x_N}{N}$$

편차

$$V=rac{(x_1-m)^2+(x_2-m)^2+\cdots+(x_N-m)^2}{N}$$

표준편차

$$\sigma = \sqrt{rac{(x_1-m)^2+(x_2-m)^2+\cdots+(x_N-m)^2}{N}}$$

R 데이터 엑셀 변환, 엑셀 불러오기

```
3 install.packages("xlsx")
  4 library(xlsx)
  5 getwd()
  6 setwd("c:/RStudio/데이터분석과R프로그래밍")
   #만약, 필요한 패키지를 로딩중입니다: rJava Error : .onLoad가 loadNamespace()에서 'rJava'때문에 실패했습니
    다.라고 나오면 # java 재설치후 xlsx 패키지 다시 로딩해야 한다.
  8 #xlsx 패키지에서 제공되는 기본적인 엑셀 불러오기 함수는 read.xlsx() 이다. 이 함수는 read.csv() 함수와
    마찬가지로 header=T와 stringsAsFactors=T 옵션이 디폴트로 설정되며, read.xlsx("불러오려는 엑셀파일명",
    엑셀파일안에서 불러오려는 시트 순서).
  9 xlsx diet=read.xlsx("diet.xlsx",1)
 10 xlsx_diet
 12 #엑셀파일로 저장하기: R데이터를 엑셀(excel) 데이터로 저장하는 기능을 제공하는 여러 개의 패키지가 있지만,
    최근에 만들어진 writexl 패키지가 사용.
 13 install.packages("writexl")
 14 library(writexl)
 15 writexl::write xlsx(mtcars, path = "c:/RStudio/데이터분석과R프로그래밍/mtcars.xlsx")
 16 install.packages("xlsx")
 17 library(xlsx)
 18 read.xlsx("mtcars.xlsx",1)
 19
10:10 (Top Level) $
                                                                                            R Script $
Console C:/RStudio/데이터분석과R프로그레밍/ 🚕
trying URL 'https://cran.seoul.go.kr/bin/windows/contrib/3.5/xlsxjars 0.6.1.zip'
Content type 'application/zip' length 9485468 bytes (9.0 MB)
downloaded 9.0 MB
trying URL 'https://cran.seoul.go.kr/bin/windows/contrib/3.5/xlsx 0.6.1.zip'
Content type 'application/zip' length 458984 bytes (448 KB)
downloaded 448 KB
package 'xlsxjars' successfully unpacked and MD5 sums checked
package 'xlsx' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
      C:\Temp\RtmpOeRCHo\downloaded packages
> library(xlsx)
Warning message:
패키지 'xlsx'는 R 버전 3.5.3에서 작성되었습니다
> read.xlsx("mtcars.xlsx",1)
   mpg cvl disp hp drat
                        wt qsec vs am gear carb
1 21.0 6 160.0 110 3.90 2.620 16.46 0 1
2 21.0 6 160.0 110 3.90 2.875 17.02 0 1
3 22.8 4 108.0 93 3.85 2.320 18.61 1 1
4 21.4 6 258.0 110 3.08 3.215 19.44 1 0
5 18.7 8 360.0 175 3.15 3.440 17.02 0 0
6 18.1 6 225.0 105 2.76 3.460 20.22 1 0
                                             1
7 14.3 8 360.0 245 3.21 3.570 15.84 0 0
```

공공데이터 분석 학회지 논문 투고 예



[통계법 제33조(비밀의 보호)]

① 통계의 작성과정에서 알려진 사항으로서 개인이나 법인 또는 단체 등의 비밀에 속하는 사항은 보호되어야 한다. ② 통계의 작성을 위하여 수집된 개인이나 법인 또는 단체 등의 비밀에 속하는 자료는 통계작성 외의 목적으로 사용 되어서는 아니 된다.

[통계법 제32조(성실응답의무)/제41조(과태료)]

정보화통계조사는 국가지정통계로서, 응답자는 조사사항에 성실히 응답할 의무가 있으며, 응답을 거부 방해 기피하거나 거짓으로 응답할 경우, 100만원 이하의 과태료가 부과될 수 있습니다.

2017년	정보화통계	l조사
-------	-------	-----

ID		
----	--	--

0	Ŀ	충	십	ш	177	۲

미래창조과학부와 한국정보화진흥원에서는 국가정보화기본법 제49조 및 통계법 제17조, 제18조에 의거하여 우리나라 사업체들의 정보화 수준 및 현황을 종합적으로 파악하기 위해 매년 전국의 모든 사업체를 모집단으로 하는 표본조사인 정보화통계조사를 실시하고 있습니다.

바쁘실 줄 아오나 정부의 효과적인 정보화 정책 수립에 도움이 될 수 있도록 적극적인 협조를 부탁드리겠습니 다. 아울러 작성해 주신 자료는 반드시 조사와 연구에 관련된 목적에만 사용될 것이며, 비밀은 철저히 보장될 것임을 약속 드립니다.

잠시만 시간을 내어 협조해 주시면 대단히 감사하겠습니다.

설문조사에 응해 주셔서 감사드리며, 귀 사업장의 평안과 번창하심을 기원합니다.

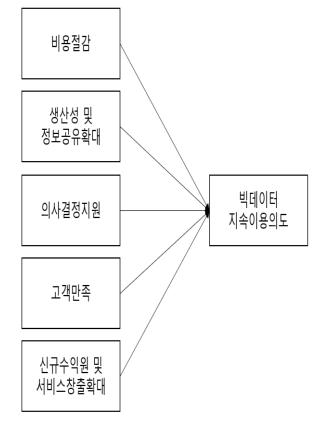
미래창조과학부 한국정보화진흥원

조사관련 문의처 [여왕타고리아 환유미 차장 (02-3415-5036), 조인회 대리 (02-3415-5210)

지역	①서울 ②부산 ③대구 ④인천 ⑤광주 ⑥대전 ⑦울산 ⑥세종 ⑤경기 웰강원 ⑪충북 ⑫충남 ⑬전북 ⑭전남 ⑬정북 ⑥경남 ⑰제주
사업체명 (면접원 기입)	
표본 번호 (면접원 기입)	-
집단 구분 <i>(면접원 기입)</i>	산업분류(5자리) 규모번호(1~6)
	업종구분(01~13) 제조업중분류(01~24) (** 업종구분에서 02.제조업인 경우만
사업 형태 (면접원 기입)	① 단독사업체 ② 본사/본점/본부 ③ 공장/지사(점)/영업소
조직 형태 (면접원 기입)	① 개인사업체 ② 회사법인 ③ 회사 이외 법인 ④ 비법인단체 ⑤ 국가/지방자치단체

● 응답해 주실 때 꼭 지켜 주십시오

- 👔 질문지는 첫 페이지부터 <u>순서대로 차례차례</u> 응답해 주십시오. 질문 앞에 특별한 언급이 없는 한 모든 질문에 답해 주십시오. (겉표지를 포함해 1~25페이지로 구성되어 있습니다)
- ② 응답은 귀 사업장의 전산담당자/정보화담당자(대리급 이상)께서 해 주시기 바랍니다. 전산담당자가 없으실 경우, 총무담당자나 대표께서 직접 기입해 주셔도 됩니다.
- [3] 질문지에 응답하실 때 특별한 요구가 없으면 보기번호 중 한 개만 골라 주시기 바랍니다.
- [2] 질문 앞에 특별한 언급이 없는 한 모든 설문의 응답 기준 시점은 「2016년 12월 31일」입니다. 현재라는 표현의 질문에서도 모든 2016년 12월 31일 기준으로 응답해 주시기 바랍니다. 예) 귀 사업장은 현재 컴퓨터가 있습니까?⇒ 2016년 12월 31일 현재 컴퓨터 보유 여부를 말씀해 주시면 됩니다.
- *최근 1년간(2016년 1월 ~ 12월)을 기준으로 하는 문항은 문항에 별도로 표기가 되어 있습니다.



연구모형

분석예 (카이제곱검정)

<표 3> 지역별 빅 데이터 사용여부

TICE	빅 데이터	카이제곱/	
지역 	비해당(%)	해당(%)	р
서울	2853(21.7)	153(1.2)	
부산	910(6.9)	26(0.2)	
대구	553(4.2)	8(0.1)	
인천	602(4.6)	56(0.4)	
광주	345(2.6)	10(0.1)	
대전	288(2.2)	48(0.4)	
울산	285(2.2)	5(0.0)	
세종	47(0.4)	4(0.0)	
경기	2921(22.2)	65(0.5)	285.783
강원	319(2.4)	17(0.1)	0.000***
충북	393(3.0)	5(0.0)	
충남	485(3.7)	14(0.1)	
전북	431(3.3)	0(0.0)	
전남	531(4.0)	11(0.1)	
경북	699(5.3)	8(0.1)	
경남	877(6.7)	7(0.1)	
제주	152(1.2)	8(0.1)	
저 비	12,691	445(2.4)	
전체 	(96.6)	445(3.4)	

^{*:}p<0.05, **:p<0.01, **:p<0.001

<표 4> 업종별 빅 데이터 사용

	<u> </u>	카이제	
업종	이용하지	이용한다	곱
	않는다(%)	(%)	/p
농림수산업	551(4.2)	9(0.1)	
제조업	3558(27.1)	53(0.4)	
건설업	789(6.0)	8(0.1)	
도/소매업	909(6.9)	47(0.4)	
운수업	899(6.8)	11(0.1)	
숙박소매업	793(6.0)	22(0.2)	
출판,영상,사	547(4.2)	53(0.4)	
<u> </u>			
금융/보험업	620(4.7)	74(0.6)	311.853
부동산	587(4.5)	8(0.1)	0.000***
전문,	759(5.8)	43(0.3)	
_ 과학비스업			
사업시설관	969(7.4)	17(0.1)	
리		.=	
<u> 협회및단체</u>	524(4.0)	17(0.1)	
기타	1186(9.0)	83 (0.6)	
전체	12691(96.6)	445(3.4)	_

^{*:}p<0.05, **:p<0.01, ***:p<0.001

분석예 (교차분석, logistic Regression)

<표 6> 조사대상자의 조직형태

	빅 데이		
조직 형태	이용하지 않다(%)	이용한다(%)	계
 개인사 업체	4551(34.6)	48(0.4)	4599(35)
회사법 인	6066(46.2)	248(1.9)	6314 (48.1)
회사이 외법인	1177(9)	86(0.7)	1263(9.6)
<u> </u>	434(3.3)	8(0.1)	442(3.4)
국가지 방자치 단체	463(3.5)	55(0.4)	518(3.9)
전체	12691 (96.6)	445(3.4)	13136 (100)

<표 9> 로지스틱 회귀분석

항목	В	S.E	유의 확률	Exp(B)
비용절감	22.657	128.32 4	0.999	69.344
의사결정	22.005	218.55	0.000	
지원	22.605	1	0.999	65.653
생산성	-46.212	253.80	0.999	
정보공유	-40.212	9	0.999	0.000
고객만족	3.371	1.216	0.006**	29.117
신규수익	2 771	1.152	0.001**	42.402
서비스창출	3.771	1.152	0.001**	43.403
상수항	-2.620	0.035	0.000	0.073

^{*:}p<0.05, **:p<0.01, ***:p<0.001.

각 조별 프로젝트 데이터 확인 및 데이터 선정 이유 조별 설명 - 피드백