

[통계학 개론] 데이터의 그래프에 의한 요약

데이터의 종류와 자료정리 목적

자료의 구분

- 원자료(raw data): 표본에서 조사된 처음 자료
- 양적자료(quantitative data): 수의 크기로 조사된 자료(몸무게, 키, 나이)
- 질적자료(qualitative data): 특성에 의해 구분된 자료(성별, 종교). 수학적 연산을 할 수 없음

통계자료의 측정 척도

- 명목 척도(nominal scale): 남자 1, 여자 2
- 서열 척도(ordinal scale): 상, 중, 하
- 구간 척도(interval scale): 물가지수 (상대비교)
- 비율 척도(ratio scale): 몸무게, 세금액 (절대비교)

도수분포표의 작성과 해석

도수분포표 frequency table

자료의 각 값의 출현도수를 세거나 몇개의 구간으로 나누어 각 구간에 속하는 자료의 갯수를 세어 정리한 표

1. 자료의 개수를 세어 n 으로 나타낸다.
2. 최대값과 최소값을 찾아 범위를 구한다.
3. 급의 수를 정한다.
4. 급의 폭을 구한다.
5. 급의 경계값과 중심값을 구한다.
6. 도수분포용지를 준비하고 도수를 표시

히스토그램의 작성과 해석

- 자료의 중심위치 → 대표값
- 자료의 산포
- 봉우리의 갯수 → 봉우리가 2개 이상이면 분리 필요
- 이상치 점검 (outlier)

히스토그램의 유형

표준형태에 가까워 지도록 변형한다.

- 다중 히스토그램
- 양방향 히스토그램

질적데이터의 그래프

- 원그래프 (Pie Chart)

양적데이터의 그래프

줄기-잎그림 (stem-and-leaf plot)

1. 각 원자료를 줄기와 잎부분으로 나눈다. 일반적으로 잎은 한 자리수 마지막 자리이고, 줄기부분은 하나 이상의 자리수를 차지한다.
2. 줄기를 아래로 나열하여 쓴다.
3. 자료마다의 잎을 해당되는 줄기에 정렬한다.
4. 잎의 수치들을 크기순으로 정렬한다.

상자그림 Boxplot

점도표의 작성 Dot Diagram

1. 실선을 긋는다.
2. 직선상에 최소값과 최대값을 표시한다.
3. 최대값과 최소값 사이에 등간격으로 눈금을 표시한다.
4. 자료값을 실선 위에 해당 지점에 X 나 점(.)으로 표시한다.

분포특성 그래프

R Commander를 이용한 실습

R 프로그램

- 무료, 전 세계 많은 사용자가 다양한 함수와 패키지를 만들고 공유
- 오픈소스, 수많은 이용자가 자유롭게 분석기법 추가
- 다양한 분야의 통계분석 가능. 강력한 그래픽 기능
- 뛰어난 도움말 기능

R Commander

User Friendly 하게 만든 것

- R을 그래픽 환경(GUI) 활용할 수 있는 패키지

R Commander 설치

- R을 구동시킨 뒤 R Console에서 `install.packages("Rcmdr")`이라는 명령을 실행
- R Console의 윗줄 메뉴 중 Packages-Install Packages(s)를 클릭하여 원하는 패키지를 찾아 설치

R 설치

- [<https://cran.r-project.org/mirrors.html>]에서 다운 받을 미러를 선택한다.

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Files:

[R-3.3.2.pkg](#)

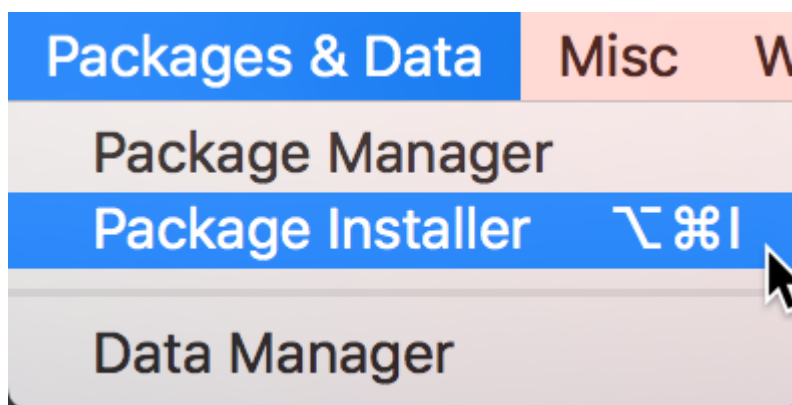
MD5-hash: 77e27f760b7559d18e30d69b6f2843ba
SHA1-

hash: 18446e3f65d31106cc2b62c41ac61cddb50c27ef
(ca. 71MB)

R 3.3.2 binary for Mac OS X 10.9 (Mavericks) and higher, signed package. Contains R 3.3.2 framework, R.app GUI 1.68 in 64-bit for Intel Macs, Tcl/Tk 8.6.0 X11 libraries and Texinfo 5.2. The latter two components are optional and can be omitted when choosing "custom install", it is only needed if you want to use the `tcltk` R package or build package documentation from sources.

Note: the use of X11 (including `tcltk`) requires [XQuartz](#) to be installed since it is no longer part of OS X. Always re-install XQuartz when upgrading your OS X to a new major version.

다운 받을 플랫폼을 선택한다. 파일을 다운 받고 설치를 시작한다.



질적 자료에 대한 그래프 표현

- pie : 원그래프
- barplot : 막대그래프

양적 자료에 대한 그래프 표현

- hist : 히스토그램
- stem : 줄기-잎그림
- dotchart : 점그래프
- boxplot : 상자그림

인포그래픽스, 미술적 소양,

통계학 : 필요한 데이터를 수집하고, 정리해 불확실한 사실에 대한 결론이나 일반적인 규칙성을 추구하

는 학문

필요한 자질, 배경, 스토리 텔링, 인문학 토양, 다양한 교양 필요

관련 문서

Plugin Backlinks: 아무 것도 없습니다.

From:

<http://jace.link/> - Various Ways

Permanent link:

<http://jace.link/open/%EB%8D%B0%EC%9D%B4%ED%84%B0%EC%9D%98-%EA%B7%B8%EB%9E%98%ED%94%84%EC%97%90-%EC%9D%98%ED%95%9C-%EC%9A%94%EC%95%BD>

Last update: 2020/06/02 09:25

