

통계학개론 REPORT



과목 : 통계학개론

교수님 : 심재창 교수님

제출기간 : ~2019.03.19

학과 : 응용신소재공학전공

학번 : 20141661

이름 : 신명섭

1. R 언어의 특징

- 통계계산과 그래픽을 위한 프로그래밍 언어이고 오픈소스 소프트웨어이다.
- 무료로 사용가능하고 무료라고 해서 다른 (EXCEL, Python, Spss 등) 통계 프로그램에 비해서 기능이 약하지 않다.
- 뛰어난 확장성을 가지고 있으며 초보자도 쉽게 접근해서 데이터를 다룰 수 있다.
- 플랫폼과의 호환이 가능하다. Microsoft Excel, Microsoft Access, MySQL, SQLite, Oracle 및 기타 프로그램에서 데이터를 가져올 수 있다.
- Windows, Mac OS X 및 Linux에서 실행 가능하다.
- 메모리 관리, 속도, 효율성이 조금 부족하고 보안 기능이 내장되어 있지 않다는 단점이 있다.
- 그래픽 관련 패키지를 설치하면 간단하게 다양한 그래프를 활용할 수 있으며 구글이나 네이버 지도를 불러오거나 이를 활용해 GIS 용도로 쓰는 것도 가능하다.

2. R언어로 동전을 1000개 던질 때 앞과 뒤가 나오는 경우의 수를 그래프로 나타내는 명령을 적으시오.

- `barplot(table(sample(1:2, 1000, TRUE)))`
- `barplot`은 막대 그래프로 데이터를 시각화시켜주는 도구입니다.
- `table`은 데이터를 한눈에 보기 쉽게 표처럼 만들어주는 도구입니다.
- `sample(범위(x), 개수(size), replace)`은 `x`를 `size`만큼 무작위로 추출하는 도구이다. `replace`는 `TRUE`(중복허용) or `FALSE`(중복을 허용하지 않는다. 모집단보다 큰 샘플을 가질 수 없다.) 중 선택을 해야 한다. *반드시 대문자로 해야 한다*

3. 도구로 푸는 통계, R (1~10과)의 유튜브 동영상을 학습하고, 소감을 적으시오.

1강 R 설치하기

- R을 설치하는 방법과 글씨의 크기를 조절하는 방법을 학습

2강 기본연산자와 기본함수

- 사칙연산을 하는 방법과 `n`제곱을 하는 방법과 기본함수, `help(모르는함수)`를 하면 함수에 대한 설명이 나오는 것을 학습하였다.

3강 변수에 숫자 데이터 입력

- 숫자를 입력하는 방법을 학습하였다. 변수의 이름을 정할 때 첫 글자는 문자 또는 ,로 시작해야하고 , 뒤에는 숫자가 올 수 없다.

4강 유용한 함수 1탄 (`min`, `max`, `range`, `length`, `sum`, `prod`)

- `min()` 최솟값을 구해주는 함수, `max()` 최댓값을 구해주는 함수, `range()` 범위를 알려주는 함수 최솟값과 최댓값을 표시해준다, `length()` 전체 요소의 개수를 구해

준다, sum() 전체 요소의 합을 구해준다, prod() 전체 요소의 곱을 구해준다.

5강 자료의 통계량 구하는 법(평균값, 중간값, 최빈값, 분산 + sort함수)

- sort함수는 오름차순, 내림차순정렬을 해주는 함수 디폴트로 오름차순이 되었다. 내림차순으로 하기 위해서는 sort(x,decreasing=FALSE)로 하면 된다. mean() 자료의 평균을 구해주는 함수. median() 중앙값을 구해주는 함수. 중앙값이란 크기순서대로 나열했을 때 순서가 중간인 값이다. mode() 최빈값을 구해주는 함수. 최빈값이란 자료에서 가장 자주 등장하는 값이다. var() 분산을 구해주는 함수.

6강 정규분포 함수(rnorm, dnorm, pnorm, qnorm)

- 난수함수(rnorm), 확률밀도함수(dnorm), 누적분포함수(pnorm), 분위수함수(qnorm)

7강 문자열(NA, NULL, Inf)

- 변수에 숫자 뿐만 아니라 문자도 사용가능하다. NA는 not available의 약자이다. 데이터 값이 존재하지 않을 때 사용. ex) 네 사람의 점수를 입력해야 하는데 한 사람의 점수를 모르는 경우 사용. a=c(88,98,90,NA), NULL은 데이터를 정의하지 않고 변수만 만들어 놓고 싶을 때 사용. Inf는 무한대를 의미

8강 진리값(TRUE, FALSE)

- 참과 거짓을 판단할 때 사용.

9강 비교연산자

- 부등호연산자 : >,<,>=,<= , 등호연산자 : ==(좌변과 우변이 같다),!=(좌변과 우변이 다르다)

10강 논리연산자 (and, or, not)

- and 연산자 : 두 피연산자가 모두 TRUE일 때만 TRUE 값을 반환하는 연산자이다. and 연산자는 &와 &&두 가지를 사용. 숫자의 경우 0만 FALSE로 인식 &는 원소단위의 개별연산 결과를 출력하고 &&는 첫 번째 원소간의 연산결과만을 출력.
- or 연산자 : 두 피연산자 중 하나만 TRUE여도 TRUE값을 반환하는 연산자. or연산자는 |와 ||두가지를 사용
- not 연산자 : !를 사용. not 연산자는 반대 논리값을 반환한다.

소감

통계학개론을 수강신청을 할 때 많은 고민이 있었는데 첫 수업을 듣고 나서 수강신청을 하길 잘 하였다는 생각이 들었습니다. 저는 컴퓨터공학과가 아닌 응용신소재공학 전공 학생이지만 교수님께서 타과생도 듣기 쉽게 잘 설명해주셔서 처음 접해보는 'R'이라는 통계프로그램에 쉽게 다가갈 수 있었고 흥미를 느꼈습니다. [도구로 푸는 통계, R] 유튜브를 학습하면서 느낀 점은 보통 인터넷 강의라고 하면 지루한데 짧은 시간 동안 핵심 부분만 알려주어서 재미있게 R에 대하여 배웠습니다. R 프로그램에 흥미가 생겨서 앞으로의 통계학개론 수업이 기대가 되고 기다려집니다. 열심히 해 보겠습니다.