

“

디지털 광고 업계에서 일하면서, 고객 데이터 수치를 깊이 있게 파악하고 싶었습니다.  
결국 퇴사를 선택했고 20년 8월부터 빅데이터 전문가 과정 교육을 받았습니다.  
R과 Python이 주 언어이며 데이터를 통해 인사이트를 얻는 일에 열정이 있습니다.  
현재는 자연어 처리, 그 중에서 텍스트 분류 감정분석 영역을 중점적으로 연구합니다.

”

## - CONTENTS -

1. Profile
2. Education
3. Certificate
4. Career
5. Project



- NAME 최나은 (NAEUN CHOI)
- PHONE 010-7111-2068
- E-MAIL naeun1218@gmail.com
- GITHUB <https://github.com/ne-choi>
- BLOG <https://ne-choi.github.io>
- KAGGLE <https://www.kaggle.com/nechoi>

 Learning R, SQL, and Python

 Interested in Data analysis & NLP

## ■ 학교 교육

### - 숙명여자대학교 학사 졸업

2014.03 - 2018.08

경영학부 전공 / 한국어문학부 복수전공

### - 강림직업전문학교 빅데이터 전문가반 졸업 예정

2020.08 - 2021.02 / 700시간 교육

평택시 4차 산업 빅데이터 전문가 양성 과정

R, Python, SQL 교육 및 프로젝트 진행, 공모전 3회 출전

언어	학습 교재명
R	R을 활용한 데이터 과학
	유쾌한 R 통계학
	ADsP 데이터 분석 준전문가
Python	Pandas 10분 완성
	파이썬 머신러닝 완벽 가이드
	텐서플로 2와 머신러닝으로 시작하는 자연어 처리
SQL	하루 10분 SQL
	구글 빅쿼리 완벽 가이드

## ■ 기타 교육

### - POSTEC 청년 AI · Big Data 아카데미 수료

2020.11 - 2021.01 / 103시간 교육

포항공과대학교 Smart Learn 수료증 발급

분류	강의명	학습 시간
빅데이터	빅데이터분석과 R프로그래밍 I	12시간
	빅데이터분석과 R프로그래밍 II	12시간
	데이터사이언스를 위한 통계학입문 I	12시간
	데이터사이언스를 위한 통계학입문 II	12시간
인공지능	머신러닝기법과 R프로그래밍 I	10시간
	머신러닝기법과 R프로그래밍 II	10시간
	Python 프로그래밍 I	12시간
	Python 프로그래밍 II	12시간
소프트웨어	Computational Thinking I	4시간
	Computational Thinking II	4시간
특강	AI 입문	2시간
	Computer Vision 입문	1시간

### - 수학의 신 선형대수학 온라인 강의 수료

2020.12 - 2021.01

### ■ 자격증

- ADsP 데이터분석 준전문가  
2020.12.22 / 한국데이터산업진흥원
- ERP 생산 정보관리사 2급  
2020.10.13 / 한국생산성본부
- ERP 물류 정보관리사 2급  
2020.10.13 / 한국생산성본부
- GTQ 그래픽기술자격 1급  
2019.04.12 / 한국생산성본부
- 컴퓨터활용능력 2급  
2017.12.08 / 대한상공회의소

### ■ 어학 능력

- TOEIC  
905점 / 2019.03.16 / ETS
- OPIc  
IH / 2020.03.28 / ACTFL

### ■ 수상 내역

- 소비트렌드 코리아2020 (KDX 유통소비데이터분석&시각화경진대회)  
우수상 / 2020.11.18 / KDX 한국데이터거래소, MBN

## ■ 정규직

### - (주)메조미디어 (CJ그룹 계열사)

기간: 2019.07 - 2020.07 (13개월)

부서: 디지털광고 1국2팀

직위: 정규직 사원

업무: 디지털 미디어 플래너

세부 내용: 광고 운영 및 성과 분석, 고객 타겟팅 데이터 관리, 미디어 믹스 작성, 광고주/대행사 관리, 비딩 참여, 지역 광고주 청약 담당

사용 툴: Google Analytics, Facebook 광고관리자, 네이버 NOSP 광고 시스템 등 캠페인 운영 및 캠페인 결과 분석 툴

교육: Google Analytics Academy, Facebook blueprint 수료, 다양한 광고 매체 설명회 참석

성과: 독립적 업무 수행, 월 10개 이상 캠페인 운영, 100개 이상 지역 광고주 청약 관리, 월 2회 이상 광고주/대행사 제안, 미팅 조율 및 참여

캠페인 목적에 맞는  
미디어 믹스 작성  
(매체 선정, 고객 타겟팅)



캠페인 예약 세팅 및 집행



일일, 주간 리포트 작성  
타겟팅 효과 데이터 분석  
캠페인 라인 효율 확인



캠페인 최종 보고서 작성  
후 캠페인 타겟팅 제안

## ■ 아르바이트

### - 보임안경원

기간: 2018.02 - 2018.10 (9개월)

직위: 아르바이트

업무: 온라인 쇼핑몰 관리

세부 내용: 네이버 광고 관리, 고객 타겟팅, 네이버 애널리틱스 분석

사용 툴: 네이버 광고관리시스템, 블로그 인사이트 분석

성과: 오픈 6개월 만에 온라인 쇼핑몰 매출액 800만원 달성

## ■ 교육 업계

### - 고려 컴퓨터학원

기간: 2016.12 - 2017.10 (11개월)

직위: 아르바이트

업무: 수능반, 내신반 국어 강사



세부 내용: 고3 수능 국어 강의, 고1-2 내신 국어 강의, 글쓰기 교재 제작

### - 글쓰기 & 자기소개서 첨삭

기간: 2017년 ~

직위: 프리랜서

활동 사이트: 탈잉, 숨고, 교내 사이트

참고 자료:  

## ■ 대외활동

### - 경영학부 학생회

기간: 2014.03 - 2015.12 (21개월)

업무: MT 진행, 과방 지킴이, 전공 설명회 개최, 축제 주점 운영, 학부 간식 배부, 학생지도의 날 행사 진행

기여: 스태프 업무 지원, 행사 참석률 100%

### - 학생지원팀 소속 리더십그룹 블룸

기간: 2015.09 - 2017.06 (21개월)

(2016.09 - 2016.06 그룹원 / 2016.06 - 2017.06 기획부장)

업무: 입학식 도우미, 평생지도 교수제 홍보&관리, 리더십그룹 자체 행사 진행


기여: 각종 행사 기획안 작성, 결과 보고서 작성, 입학식 명찰과 팻말 제작

### - 국제협력팀소속 봉사동아리 GPM

기간: 2015.09 - 2016.06 (9개월)

업무: 외국인 유학생 대상 한국어와 전공 공부 교육, 학교 생활 멘토링

기여: 1:1 멘토, 행사 기획 아이디어 제시(유학생과 함께 하는 뷰티 클래스, 체육대회 등)

- 
- 리더십
  - 문제 해결 역량
  - 보고서 작성 능력
  - 커뮤니케이션 능력

- 다문화 경험
- 외국어 능력
- 행사 기획 능력

## ■ 대외활동

### - 경상대학(경영/경제) 회장단

기간: 2016.01 - 2016.12 (12개월)

업무: 입학식, OT, 시험 기간 간식 사업, 축제 기획, 학생지도의 날 체육대회 개최, 졸업식 등 행사 자율적으로 준비

기여: 단과대학 모든 행사 진행

### - 숙명여자대학교 비상대책위원회

기간: 2016.01 - 2016.12 (12개월)

업무: 총학생회장단 부재로 직무 대리

학교 축제 준비- 기획부 / 선거 준비- 중앙선거관리위원장


기여: 총학 부재 1년 동안 교내외 행사 진행, 선거 기관과 보궐선거 준비 및 진행

### - 숙명여자대학교 어학당 한국어 도우미

기간: 2016.06 - 2016.08 (3개월)

업무: 어학당 멘토 활동, 한국어 공부&중간고사 준비 보조


기여: 멘티 2명 모두 서울 소재 대학교 진학

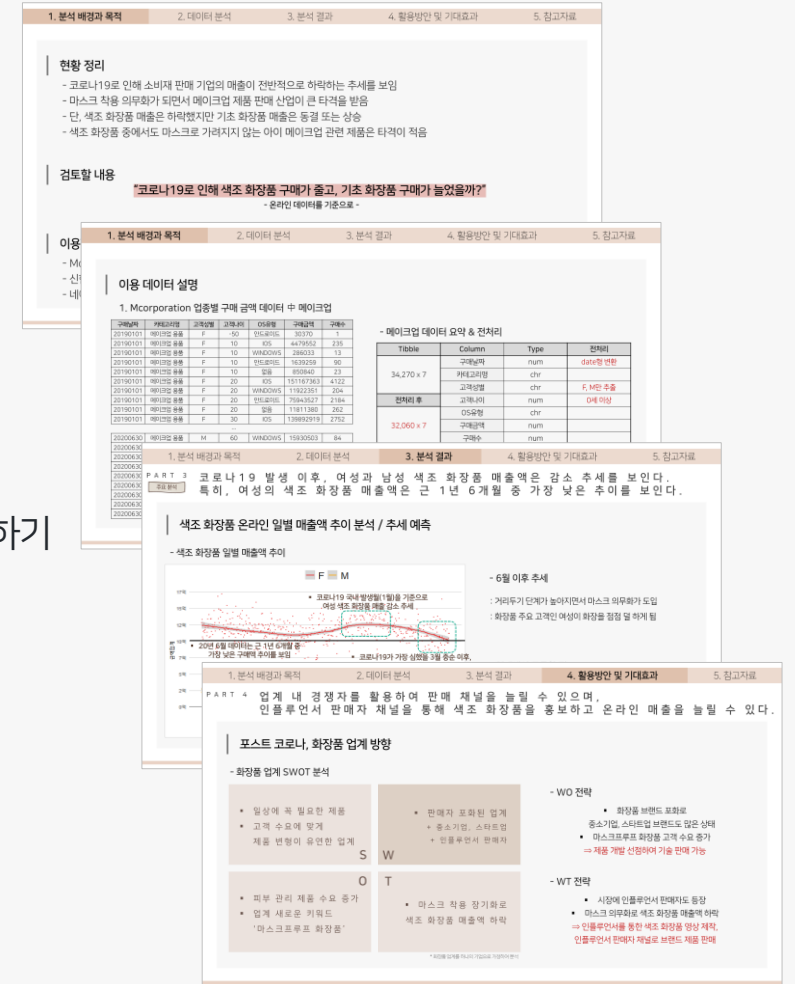
- 
- 리더십
  - 학생 자치 활동
  - 행사 기획 및 진행
  - 커뮤니케이션 능력
  - 타 단과대학과의 협업

- 다문화 경험
- 외국어 능력



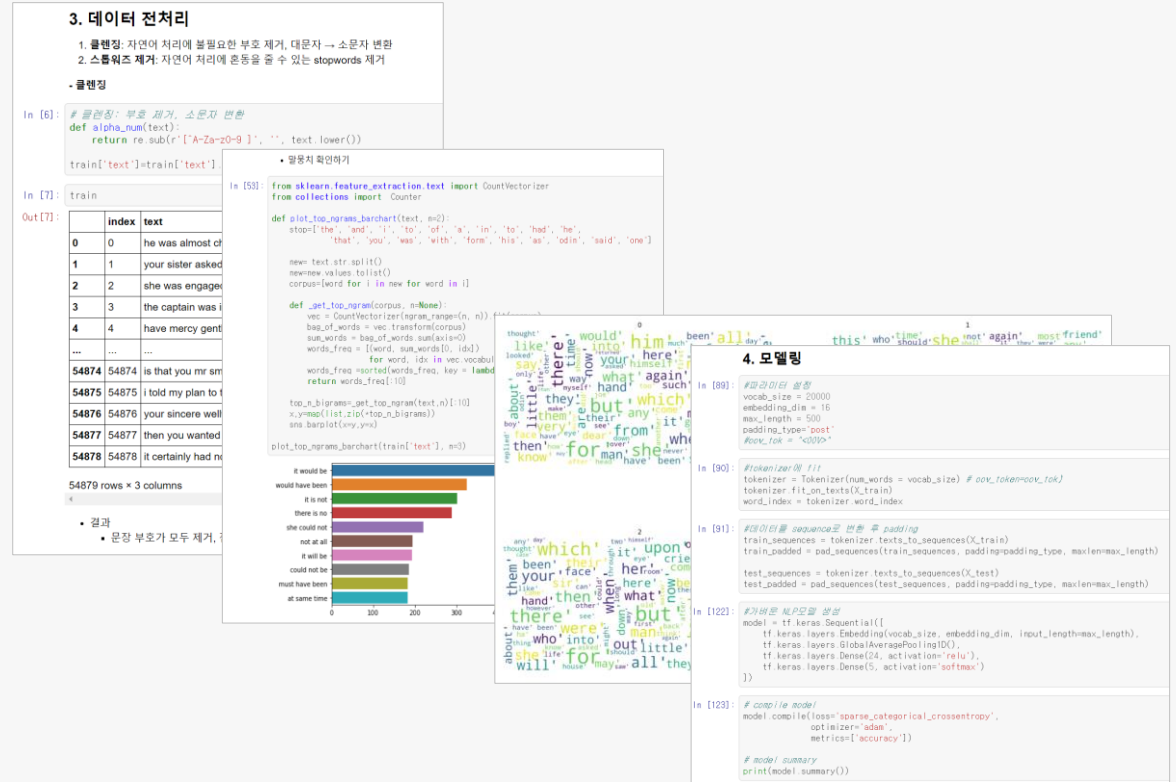
## ■ KDX 소비트렌드 코리아2020: EDA, 시각화, 분석

- 언어 및 작업툴: R, RStudio
- 인원: 3명
- 기간: 2020.10.12 - 2020.10.24
- 내용
  - 주어진 데이터와 외부 데이터를 사용하여 한국 소비 트렌드를 분석하고 인사이트 도출하기
  - 주제 선정 자유
- 선정 주제
  - "코로나19 발생 이후 색조 및 기초 화장품 수요 비교분석"
  - 코로나19로 인한 마스크 일상화/의무화 시대, 화장품 업계 매출 추이를 살펴보고 발전 방향 제언하기
- 기여: 데이터 전처리 및 시각화, ppt 제작
- 결과: 3등(우수상) → 매일경제 기사 보기
- 제출 파일: 




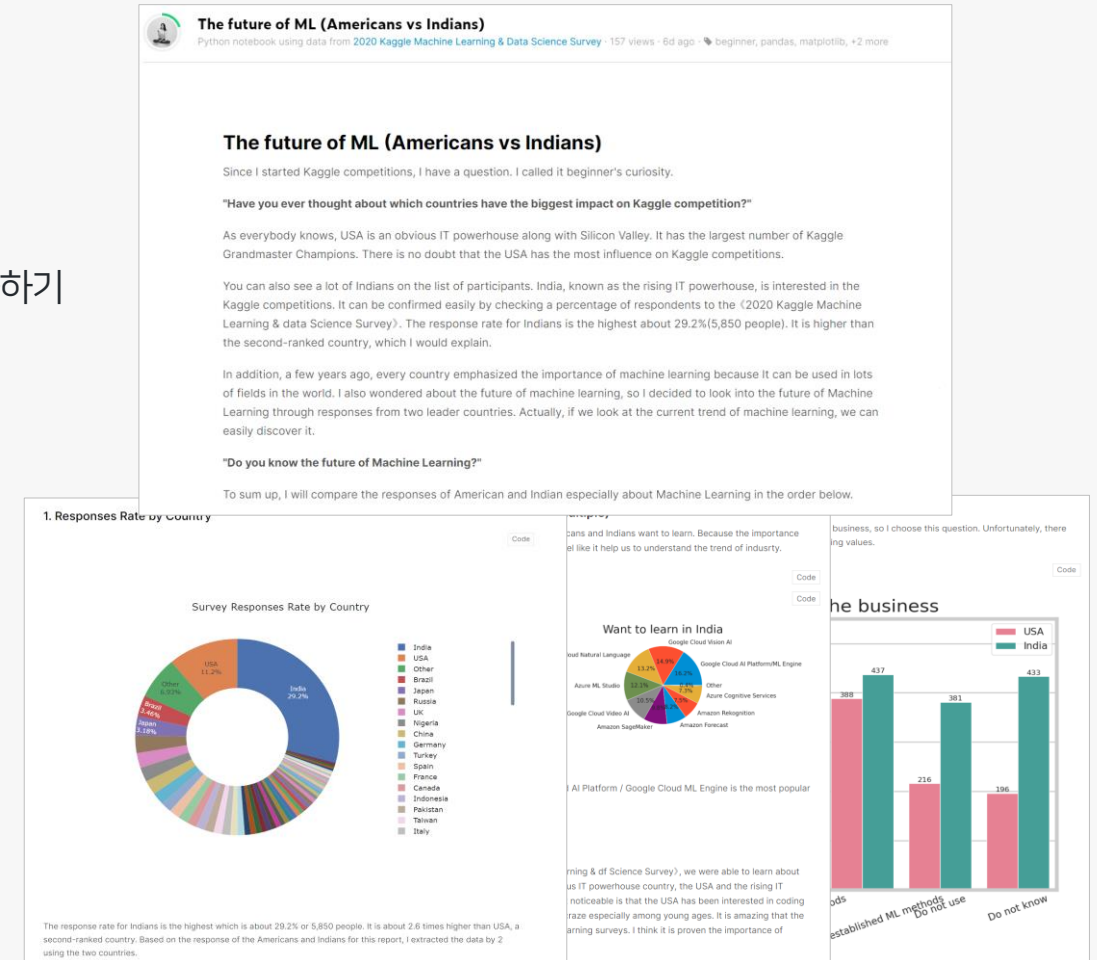
## ■ 소설 작가 분류 AI 경진대회: EDA & 모델링

- 언어 및 작업툴: python, jupyter notebook
- 인원: 2명
- 기간: 2020.12.03 - 2020.12.04
- 내용
  - 문체 분석 알고리즘 개발
  - 소설 속 문장문치 분석을 통한 저자 예측
    - 작가의 글을 분석하여 특징 도출
    - 취향 추천 시스템 활용 / 대필, 유사작 탐지
- 기여: 텍스트 분석 전처리, 모델링 코드 작성
- 결과: 158/287 (전체 참가 팀: 658팀)



## 2020 Kaggle Machine Learning & Data Science Survey: EDA, 시각화, 분석

- 언어 및 작업툴: python, jupyter notebook
- 인원: 1명
- 기간: 2020.12.30 - 2021.01.06
- 내용
  - 캐글에서 진행하는 연례 데이터 과학 설문 조사 데이터에서 인사이트 도출하기
  - 주제 선정 자유
- 선정 주제
  - "The future of ML (Americans vs Indians)"
- 기여: 데이터 전처리, 시각화 및 분석
- 캐글에서 보기: 



## ■ 교육 콘텐츠 제작

- NLP 강의안 제작
  - 텐서플로 2와 머신러닝으로 시작하는 자연어 처리 교안 정리
  - GTP3 등 자연어 처리 관련 최신 논문 정리

## - 자습 자료 작성

- ADsP 데이터 분석 준전문가 이론
- 포스텍 온라인 강의 강의안 정리
- etc.

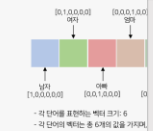
### 01. 단어 표현

#### 단어 표현(Word Representation)

: 텍스트를 자연어 처리를 위한 모델에 적용할 수 있도록, 언어적인 특성을 반영해 단어를 수치화하는 방법을 찾는 것  
: 주로 단어를 벡터로 표현해 수치화 → 단어 임베딩(word embedding) or 단어 벡터(word vector)로도 불림

#### 원-핫 인코딩(one-hot encoding) 방식

~ 각 단어를 0과 1로만 나타내는 벡터로 표현하는 방식

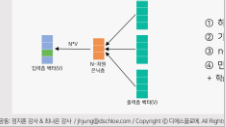


### 01. 단어 표현

#### CBOW 모델 학습 방법

- 주변 단어를 원-핫 벡터로 만들어 입력값으로 사용 (입력층 벡터)
- 가중치 행렬(weight matrix)을 각각의 원-핫 벡터에 곱하여 n-차원 벡터를 생성 (n-차원 은닉층)
- n-차원 벡터를 모두 다한 후, 개수로 나누어 평균 n-차원 벡터를 생성 (출력층 벡터)
- n-차원 벡터에 다시 가중치 행렬을 곱하여 원-핫 벡터와 같은 차원의 벡터로 만들
- 만들어진 벡터를 실제 예측하려는 단어의 원-핫 벡터와 비교하여 학습

#### Skip-gram 모델 학습 방법



### 02. 버트를 활용한 미세 조정 학습

#### 추가: 혼동 행렬 표와 F1 Score

##### - 혼동 행렬(Confusion Matrix)

관측값(Observed)	예측값(Predicted)	
	양성(Positive)	음성(Negative)
양성(Positive)	True Positive(TP)	False Negative(FN)
음성(Negative)	False Positive(FP)	True Negative(TN)

$$1. \text{Accuracy (정확도)} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$2. \text{Precision (정밀도)} = \frac{TP}{TP + FP}$$

$$3. \text{Recall (재현율)} = \frac{TP}{TP + FN}$$

$$4. \text{F1 Score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

\* Accuracy

: 전체 경우의 수 중에서 올바르게 예측한 비율

\* Precision

: 1이라고 예측한 것 중에 실제 1인 것의 비율

\* Recall (= sensitivity; 민감도)

: 실제 1인 것 중에서 내가 1이라고 맞춘 것의 비율

\* F1 Score

: 정확도와 재현율을 조화평균하여 하나의 수치로 나타낸 지표

### 2. 데이터 기술통계치 요약

- 데이터: 학업성취도(보통독립 고등학생 수취점수)

```
stud<-read.csv("week6_2/stud_math.csv")
head(stud)
dim(stud)
str(stud)
attach(stud)
```

#### summary

- summary(데이터명): 변수별로 요약통계량 제공
- 문자변수에는 빈도를 주고, 숫자변수에는 최솟값, 25%, 중위값, 75%, 최댓값 제공

```
# 1-1 Numerical analytics
summary(stud)
mean(G3) # 평균
sd(G3) # 표준편차 (분산의 제곱근)
var(G3) # 분산
```

#### 이산형 확률 변수

##### 1. 베르누이 확률분포(Bernoulli distribution)

- 결과가 2개만 나오는 경우
- 동전 던지기, 시험의 합격/불합격, 안타를 칠 확률

##### 2. 이항분포(Binomial distribution)

- 베르누이 시행을 n번 반복했을 때, k번 성공할 확률
- 5번 타석에 들어서서 3번 안타를 칠 확률 → n=5, k=3, 안타를 칠 확률 P(x)=타율
- 성공할 확률 P가 0이나 1에 가깝지 않고 n이 충분히 크면 정규분포에 가까워짐, 1/2에 가까우면 종 모양

##### 3. 기하분포(Geometric distribution)

- 성공확률이 p인 베르누이 시행에서 첫 번째 성공이 있기까지 x번 실패할 확률
- 5번 타석에 들어서서 3번째 타석에서 안타를 칠 확률

##### 4. 다항분포(Multinomial distribution)

- 세 가지 이상의 결과를 가지는 반복 시행에서 발생하는 확률 분포 (이항분포 확장한 것)

##### 5. 포아송분포(Poisson distribution)

- 시간과 공간 내에서 발생하는 사건의 발생빈도에 대한 확률분포
- 적에 오다가 5달 10개 나온다고 할 때, 한 페이지에 오다가 3개 나올 확률
- 최근 5달에서 10개의 출현을 했다고 할 때, 오늘 경기에서 출현할 자치 몇 개 나올

### 3. 상관분석

#### 상관분석(Correlation Analysis)

- 두 변수 간 관계의 정도를 알아보는 방법
- 상관계수(Correlation coefficient) 이용

#### 상관계수 특성

상관계수 범위	해석
$0.7 < r \leq 1$	강한 양(+)의 상관관계 있다
$0.3 < r \leq 0.7$	중간 양(+)의 상관관계 있다
$0 < r \leq 0.3$	약한 양(+)의 상관관계 있다
$r = 0$	상관관계 없음, 학년이 증가해도 성적은 변하지 않는다
$-0.3 \leq r < 0$	약한 음(-)의 상관관계 있다
$-0.7 \leq r < -0.3$	중간 음(-)의 상관관계 있다
$-1 \leq r < -0.7$	강한 음(-)의 상관관계 있다

#### 상관분석 유형

구분	특징	주의사항
계량	평균적으로 이상으로 측정된 두 변수의 상관관계 측정 방식	서열적이지 않은 두 변수 상관관계 측정 방식
특성	연속형 변수, 정규성 가설, 대칭분열의 사용	순서형 변수, 비정규성 가설, 순위 기준 상관관계 측정
상관계수	피어슨 r(직접상관계수)	순위상관계수(s, 로랑)

**고맙습니다**