

# 총괄평가

## THEJOEUN IT ACADEMY

훈련과정명			회차	훈련생명	결재	팀장	원장
빅데이터전문가양성(파이썬,R)_B			1				
교과목	능력단위명	평가방법	평가일시		점수	평가자	
빅데이터 분석 결과 시각화	빅데이터 분석 결과 시각화	사례연구	2021. 3. 8.				

제출 파일 : 본인이름\_사례연구소스.R, 본인이름\_사례연구최종결과.html, 본인이름\_사례연구.Rmd 파일을 압축한 파일

1. '한국복지패널데이터'(SPSS, koweps\_hpc10\_2015\_beta5.sav)를 로드한 후 필요한 데이터 변수만을 select하여 변수명을 변경하시오. 단 필요한 필드로 성별은 gender, 태어난 연도는 birth, 혼인상태는 marriage, 종교는 religion, 월평균임금은 income, 직업코드는 code\_job, 지역코드는 code\_region로 필드명을 변경한다.

채점기준 ① 필요한 데이터를 다운받아 data.frame으로 load 하였다(4점).

채점기준 ② 필요한 필드만 select하였다(3점).

채점기준 ③ 필드의 변수명을 변경하였다(3점).

2. 1번 문제의 결과인 data.frame변수를 이용하여 성별에 따른 월급 차이가 있는지를 분석하시오.

채점기준 ① gender 필드 변수의 이상치가 있는지 확인하고 이상치 값 처리를 한다(1점).

채점기준 ② gender 필드 변수의 결측치를 확인한다(1점).

채점기준 ③ gender의 값이 1은 male로 2는 female로 변경하고 gender의 타입을 factor로 변경한다(1점).

채점기준 ④ 성별 비율을 도표로 나타내고 그래프로 시각화한다(2점).

채점기준 ⑤ income의 최소값, 1분위수, 중위수, 3분위수, 최대값, 결측치 등을 탐색하고, boxplot과 월급의 빈도그래프를 시각화한다(2점).

채점기준 ⑥ income이 0인 데이터는 이상치로 정하고, 이상치를 결측 처리한다(1점).

채점기준 ⑦ 결측치를 제외한 데이터를 이용하여 성별에 따른 월급차이가 있는지를 분석한다(2점).

3. 1번 문제의 결과인 data.frame변수를 이용하여 나이와 월급의 관계를 분석하여 몇 살 때 월급을 가장 많이 받는지 시각화하시오.

채점기준 ① birth, income 필드 변수의 이상치와 결측치를 확인한다(2점).

채점기준 ② birth변수를 이용하여 나이를 계산하고 이 값을 age 필드로 추가한다(2점).

채점기준 ③ x축을 나이, y축을 월급으로 지정하고 나이에 따른 월급의 변화가 표현되도록 막대그래프나 선 그래프로 시각화한다(3점).

채점기준 ④ 나이에 따른 월급의 차이가 있는지 관계를 분석한다(3점).

4. 1번 문제의 결과인 data.frame변수를 이용하여 연령대에 따른 월급의 차이가 있는지, 있으면 어떤 연령대가 월급이 가장 많은지 분석하시오. 단, 연령대는 30세 이하는 young, 30~60세는 middle, 61세 이상은 old로 분

류한다.

채점기준 ① 파생변수 agegrade를 필드로 추가한다(2점).

채점기준 ② agegrade 의 분포를 도표와 그래프로 시각화한다(3점).

채점기준 ③ 연령대 별 월급의 boxplot을 시각화한다(2점).

채점기준 ④ 실제로 연령대에 따른 월급 차이가 있는지 분석한다(3점).

5. 1번 문제의 결과인 data.frame변수를 이용하여 **성별에 따른 월급의 차이는 연령대 별로 다른지 분석**하시오.

채점기준 ① 성별, 연령대, 월급 데이터의 결측치를 확인한다(3점).

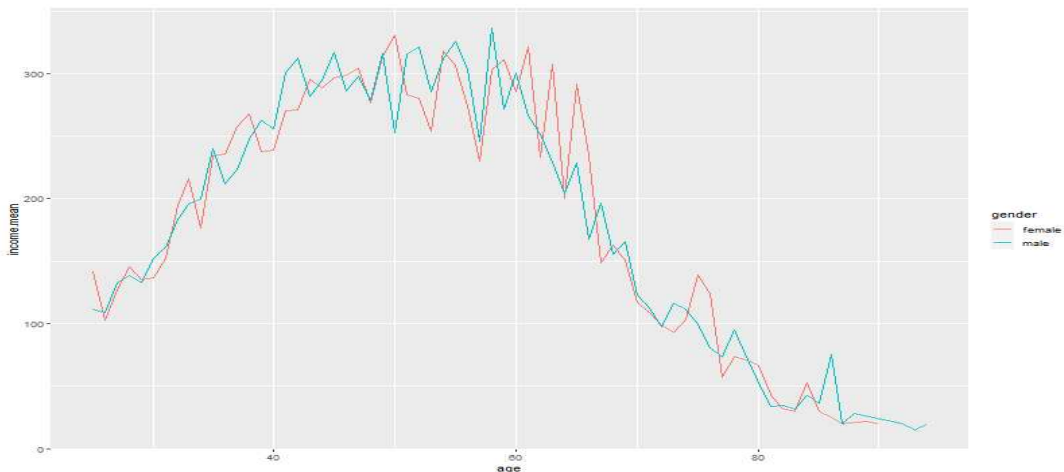
채점기준 ② 연령대별, 성별 월급의 평균과 표준편차, 빈도를 출력한다(3점).

채점기준 ③ 성별에 따른 월급의 차이가 연령대별로 다른지 시각화 한다(4점).

6. 1번 문제의 결과인 data.frame변수를 이용하여 **나이에 따른 월급 변화를 성별을 분리하여 시각화**하시오.

채점기준 ① 나이와 성별로 group\_by하여 월급평균, 월급표준편차, 월급중앙값, 최소값과 최대값, 빈도를 산출한다(5점).

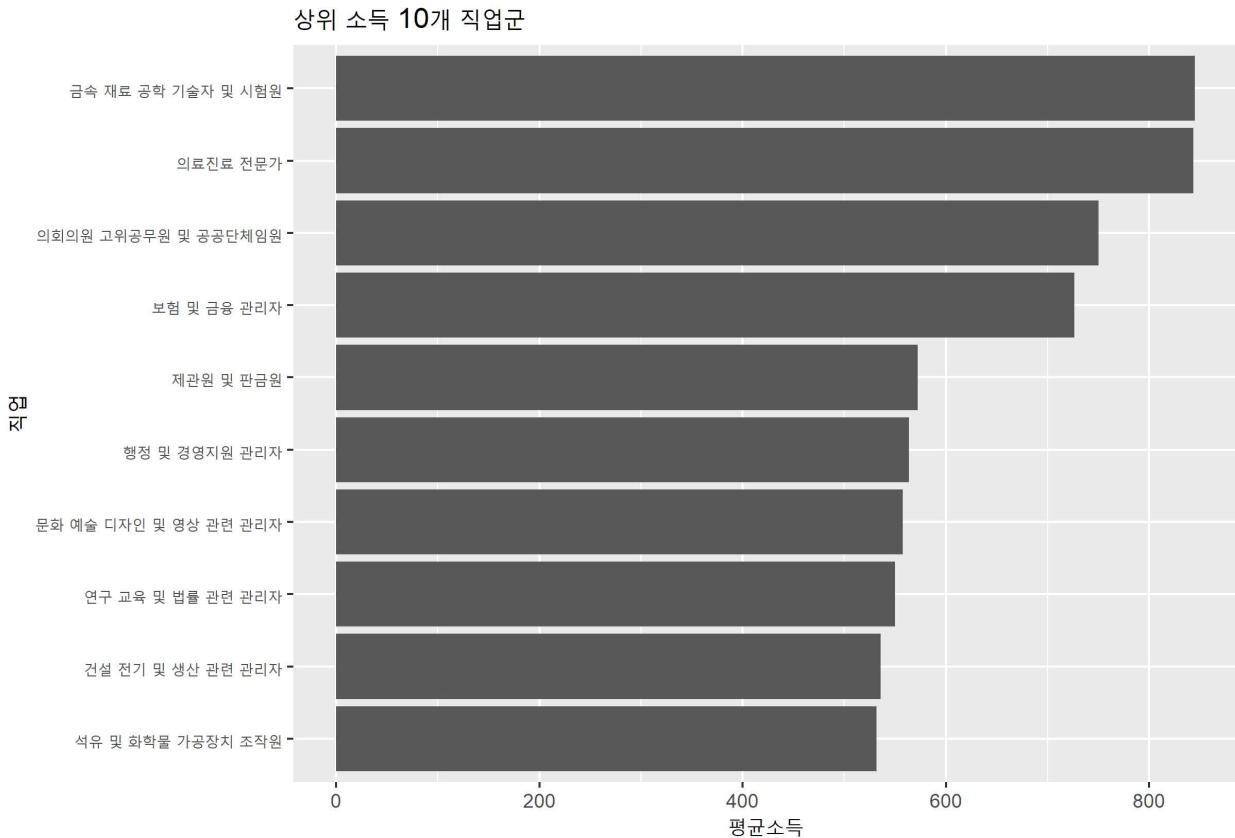
채점기준 ② 나이에 따른 월급평균의 추이를 아래와 같은 그래프를 시각화하고, 아래의 그래프를 파일로도 출력한다(5점).



7. 1번 문제의 결과인 data.frame변수를 이용하여 직업별 월급의 차이가 나는지 분석하고, 만약 월급의 차이가 나면 **어떤 직업이 월급이 가장 많은지 상위 10개 직업군만 시각화**하시오.

채점기준 ① 직업별 월급 평균, 표준편차, 빈도를 평균월급 순으로 정렬하여 출력하여 직업별 월급의 추이를 분석한다(3점).

채점기준 ② 직업별 월급의 차이를 분석한 후, 상위 소득 10개 직업군을 도표로 출력하고, 아래와 같은 그래프로 시각화한다. 시각화한 그래프는 ggsave함수를 이용하여 top10.png라는 그림파일로 저장한다(4점).



채점기준 ③ 하위 소득 10개 직업군도 도표로 출력하고 시각화한다(4점).

8. 1번 문제의 결과인 data.frame변수를 이용하여 성별로 어떤 직업이 가장 많을지 분석하시오.

채점기준 ① 여성 최빈 직업 상위 10를 추출한다(5점).

채점기준 ② 남성 최빈 직업 상위 10를 추출한다(5점).

9. 1번 문제의 결과인 data.frame변수를 이용하여 **종교 유무에 따른 이혼률을 분석**하시오.

채점기준 ① 종교 데이터인 religion 필드의 이상치 및 결측치를 확인한다(1점).

채점기준 ② religion 필드가 1이면 "종교-유", 2이면 "종교-무"로 데이터를 변경한다(2점).

채점기준 ③ 종교 유무의 빈도를 시각화한다(1점).

채점기준 ④ 혼인 상태 데이터인 marriage 필드가 1이면 "기혼", 3이면 "이혼"으로, 그 외는 NA로 값을 같은 marriage\_group 파생변수를 추가한다(2점).

채점기준 ⑤ 종교 유무에 따른 이혼률을 분석한다(2점).

채점기준 ⑥ 분석한 결과를 도표와 그래프로 시각화한다(2점).

10. 1번 문제의 결과인 data.frame변수를 이용하여 지역별 연령대 비율을 분석하시오. **노년층이 많은 지역**은 어디인지 출력하시오.

채점기준 ① 결측치를 확인한다(2점).

채점기준 ② region 파생변수를 지역명으로 추가한다(2점).

1:서울 2:수도권(인천/경기) 3:부산/경남/울산 4:대구/경북 5:대전/충남 6:강원/충북 7:광주/전남/전북 /제주도

채점기준 ③ 지역별 연령대 비율을 분석한 도표 및 그래프를 시각화한다(3점).

채점기준 ④ 노년층이 많은 지역이 어디인지 시각화한다(3점).