기술통계 연습문제 (2)

Wage 데이터 셋은 경제 및 노동 시장에 관련된 정보를 담고 있는 데이터셋이다.

이 데이터셋은 미국에서 수집된 임금에 대한 정보를 포함하고 있다.

https://data.hossam.kr/D02/wage.xlsx

year	age	maritl	race	education	region	jobclass	health	health_ins	logwage
2006	18.0	1. Never Married	1. White	1. < HS Grad	2. Middle Atlantic	1. Industrial	1. <=Good	2. No	4.3
2004	24.0	1. Never Married	1. White	4. College Grad	2. Middle Atlantic	2. Information	2. >=Very Good	2. No	4.3
2003	45.0	2. Married	1. White	3. Some College	2. Middle Atlantic	1. Industrial	1. <=Good	1. Yes	4.9
2003	43.0	2. Married	3. Asian	4. College Grad	2. Middle Atlantic	2. Information	2. >=Very Good	1. Yes	5.0
2005	50.0	4. Divorced	1. White	2. HS Grad	2. Middle Atlantic	2. Information	1. <=Good	1. Yes	4.3

각 변수는 다음과 같은 의미를 갖는다.

변수명	의미					
year	탄생년도					
age	나이					
maritl	결혼여부					
race	근로자의 인종					
education	교육수준					
region	지역					
jobclass	직군					
health	건상상태					
health_ins	건강보험 가입 여부					
logwage	임금(로그값)					
wage	임금					

- 이 데이터셋을 활용하여 다음 물음에 답하시오.
 - 1. 데이터를 로드하여 명목형 변수를 1, 2 등으로 레이블링 하시오. 값의 종류는 데이터프레임으로부터 조회하여 확인하시오.
 - 2. 레이블링 된 명목형 변수를 category 타입으로 변경하시오.
 - 3. 수치형 변수에 대한 요약 통계를 확인하고 설명하시오 (상자그림 제외)
 - 4. 명목형 변수에 대한 기술 통계를 수행하고 설명하시오.

- 5. 결혼 여부에 따른 임금 수준을 비교하고자 한다. 결혼 여부에 따라 서브플롯을 구성하여 임금 수준을 히스토그램으로 시각화 하고 설명하시오.
- 6. 교육 수준에 따른 임금에 대한 히스토그램을 시각화 하고 설명하시오. 교육수준별로 그래프를 나누어 서브플롯으로 제시해야 합니다.
- 7. 직군별 건강상태를 확인하고자 한다. 적절한 형태로 데이터를 재배치하고 설명하시오.
- 8. 교육 수준을 인종 비율에 따라 설명하고자 한다. 적절한 시각화 자료를 제시하고 설명하시오.