2020년 겨울 계절학기 사회통계 기말고사

이름: 신백록

다음 제시되는 연구 설명을 살펴보고, 이후 제시되는 문제들에 응답하기 바랍니다. (총 4문제)

|  |
| --- |
| I. 연구 대상의 특성  - 2019 서울서베이 가구조사 자료 중 가구원 자료 (수업 실습 시 사용한 자료)  - **출생년도 기준 2000년 이전 출생자 (2000년 출생자 포함 할 것)**  II. 가설  - 문화 환경에 대한 만족도가 높을수록 일상생활에서 받는 스트레스가 낮을 것이다  - 여가 생활에 대한 만족도가 높을수록 일상생활에서 받는 스트레스가 낮을 것이다.  III. 주요 변수  1. 독립변수  - 문화 환경 만족도: 2019년 서울서베이 22번 문항 이용  (문22. 귀하는 서울시의 전반적인 문화환경 (문화시설, 문화프로그램, 비용 등)에 대해 어느 정도 만족하십니까?)  - 여가 생활 만족도: 2019년 서울서베이 26번 문항 이용  (문26. 귀하는 평소 여가생활에 어느 정도 만족하십니까?)  2. 종속변수  - 스트레스: 2019년 서울서베이 20번 문항 이용  (문20. 귀하는 지난 2주일 동안 일상생활에서 전반적으로 스트레스를 어느 정도 느꼈습니까?)  3. 통제변수  - 성별: 명목 척도로 구성 (기준 집단: 남성)  - 학력: 1 ~ 5의 값을 가지는 서열 척도로 구성 (중학교 졸업 이하: 1, 고등학교 졸업: 2, 전문대 졸업: 3, 대학교 졸업: 4, 대학원 졸업: 5)  - 월 평균 가구소득: 1 ~ 5의 값을 가지는 서열 척도로 구성 (200 만원 미만: 1,  200~400 만원: 2, 400~600 만원: 3, 600~800 만원: 4, 800 만원 이상: 5)  - 연령: 1 ~ 5의 값을 가지는 서열 척도로 구성 (20대 이하: 1, 30대: 2, 40대 3,   50대: 4, 60대 이상: 5) |
|  |

**문 1. 다음의 <표 1>을 완성하시오.**

<표 1> 연구대상의 인구사회적 특성

|  |  |  |  |
| --- | --- | --- | --- |
| 구분 | | 빈도(명) | 비율(%) |
| 성별 | 남성 | 20019 | 48.1% |
| 여성 | 21604 | 51.9% |
| 연령 | 20 대 이하 | 5187 | 12.5% |
| 30 대 | 7834 | 18.8% |
| 40 대 | 8283 | 19.9% |
| 50 대 | 8544 | 20.5% |
| 60 대 이상 | 11775 | 28.3% |
| 학력 | 중학교 졸업 이하 | 4373 | 10.5% |
| 고등학교 졸업 | 15692 | 37.7% |
| 전문대 졸업 | 7511 | 18% |
| 대학교 졸업 | 13494 | 32.4% |
| 대학원 졸업 | 553 | 1.33% |
| 월 평균 가구 소득 | 200 만원 미만 | 3442 | 8.27% |
| 200~400 만원 | 10948 | 26.3% |
| 400~600 만원 | 15084 | 36.2% |
| 600~800 만원 | 8732 | 21% |
| 800 만원 이상 | 3417 | 8.21% |

**문 2. 다음의 <표 2>를 완성하고, 내용을 요약하시오.**

<표 2> 소득집단 별 주요변수 기술통계

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 구분 | 문화 환경 만족도 | | 여가 생활 만족도 | | 스트레스 | |
| 평균 | 표준편차 | 평균 | 표준편차 | 평균 | 표준편차 |
| 200 만원 미만 | 3.19 | 0.672 | 2.95 | 0.793 | 3.23 | 1.04 |
| 200~400 만원 | 3.29 | 0.672 | 3.09 | 0.767 | 3.15 | 0.981 |
| 400~600 만원 | 3.29 | 0.683 | 3.13 | 0.767 | 3.18 | 0.980 |
| 600~800 만원 | 3.26 | 0.671 | 3.14 | 0.775 | 3.23 | 0.980 |
| 800 만원 이상 | 3.32 | 0.722 | 3.19 | 0.782 | 3.22 | 0.979 |

- 결과 요약

가장 먼저 눈에 띄는 것은 표준편차이다. 등분산성 검증을 해봐야 정확히 알겠지만 표준편차가 각 그룹 별 비슷하게 나타나는 것을 볼 수 있다.

평균을 보면, 200만원 미만의 저소득층은 문화 환경 만족도(3.19), 여가 생활 만족도(2.95)가 가장 낮게 나타났고, 스트레스는 가장 높게(3.23) 나타났다. 200만원 미만의 소득을 버는 집단은 돈 때문에 스트레스가 심화된 것으로 보이고, 다른 집단에 비해 문화, 여가에 시간과 돈을 투자하기 힘들기 때문에 이와 같은 차이가 벌어진 것으로 생각된다.

가장 소득을 많이 얻는 집단에서는 문화 환경 만족도(3.32), 여가 생활 만족도(3.19)가 다른 집단에 비해 가장 높았다. 위와는 반대로 문화와 여가에 사용할 돈과 시간적 여유가 충분해서 그런 것으로 생각된다.

다른 집단도 같이 보면, 문화 환경 만족도와 여가 생활 만족도는 소득과 정적(+) 관계를 이루는 것으로 보인다. 스트레스는 소득이 가장 적은 집단에서 가장 크게 나타나는 것은 맞지만, 소득이 많아진다고 줄어드는 것은 아닌 것처럼 보인다. 오히려 소득이 가장 적은 집단을 제외하고 보면, 소득이 많아지면 스트레스가 늘어나는 정적(+) 관계가 있다.

위 결과를 고려하면, 소득을 통제변수로 설정하는 것은 합리적이다.

**문 3. 다음의 <표 3>은 가설을 검증하기 위한 회귀분석 결과이다. 표를 완성하고 분석 결과를 요약하시오.**

<표 3> 문화 환경 및 여가 생활 만족도가 개인 스트레스에 미치는 영향

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 예측 변인 | | 종속 변인: 스트레스 정도 | | |
| Coef. | | S.E. |
| 통제 변인 | 성별 (기준변수: 남성) | -0.031 | \*\* | .0097 |
| 연령 | -0.024 | \*\*\* | .0039 |
| 학력 | 0.007 |  | .0053 |
| 월 평균 가구 소득 | 0.012 | \*\* | .0048 |
| 독립 변인 | 문화 환경 만족도 | -0.037 | \*\*\* | .0073 |
| 여가생활 만족도 | -0.14 | \*\*\* | .0064 |
| 상수 | | 3.78 | \*\*\* | .0379 |
| F-Value | | 107.9 \*\*\* | | |
| R-squared | | 0.01532 | | |
| N | | 41623 | | |
| \* p<.05, \*\* p<.01, \*\*\* p<.001 | | | | |

- 결과 요약

유의수준을 0.05로 할 때, 모델의 p-value가 0.05보다 작기에 본 모델은 통계적으로 유의하다. R^2값은 0.015, 표본은 41623개이다.

학력 변수는 p-value가 .05를 넘으므로 스트레스 변인에 통계적으로 유의한 영향을 끼치지 않는다. 따라서 학력변수는 해석하지 않는다.

다른 변수가 고정된 값을 가질 때, 여성이 스트레스를 평균적으로 0.031만큼 더 적게 받는 것을 알 수 있다. 또한 다른 변수가 고정되었을 때, 연령대가 한 집단 높아질수록 0.024만큼 스트레스를 덜 받는다. 사전분석에서 봤듯이, 다른 변수가 고정되었을 때 월 평균 가구 소득이 늘어나면(200단위) 스트레스도 0.012만큼 늘어나는 것을 알 수 있다.

독립 변인을 보면, 다른 변수가 고정되었을 때 문화 환경 만족도가 1단위 커지면, 스트레스는 0.037만큼 작아진다는 것을 알 수 있다. 또한 다른 변수가 고정되었을 때 여가생활 만족도가 1단위 커지면 스트레스는 0.14만큼 작아진다. 두 독립 변인 모두 종속변인에 정적(+) 영향을 미친 것이다.

결론적으로 독립 변인들의 p-value가 유의수준인 0.05보다 작고, 모델이 통계적으로 유의하기에 가설 1(문화 환경에 대한 만족도가 높을수록 일상생활에서 받는 스트레스가 낮을 것이다)과 가설 2(여가 생활에 대한 만족도가 높을수록 일상생활에서 받는 스트레스가 낮을 것이다)를 채택한다.

**문4. 위의 분석을 수행하는데 작성한 R 구문(스크립트)를 복사하여 붙여 넣으시오.**

#load data

seoul<-tibble(read\_xlsx('seoul.xlsx', sheet=1))

#select

seoul<- seoul %>%select(sq1\_3,sq1\_2,aq20,aq22,aq26,aq1,adq1,adq1a)

#rename

seoul<-seoul %>% rename(sex=2,birth=1,stress=3,culture=4,leisure=5,earn=6,edu1=adq1,edu2=adq1a)

#filter birth<=2000

seoul<-seoul %>% filter(birth<=2000)

#age var

seoul<-seoul %>% mutate(age=2019-birth) %>% mutate(age\_c=ifelse(age<30,1,

ifelse(age<40,2,

ifelse(age<50,3,

ifelse(age<60,4,5 )))))

#sex var

seoul<-seoul %>% mutate(sex\_c=ifelse(sex==1,'male','female')) %>%

mutate(sex\_c=factor(sex\_c, levels=c('male','female')))

seoul %>% glimpse()

#edu var

#졸업을 안한 경우 응답학력의 -1

#대학(6)은 -2

seoul<- seoul %>% mutate(temp=ifelse(edu2>1, ifelse(edu1==6,edu1-2,edu1-1),edu1))

#무학인 경우 결측이 되는 문제

seoul<- seoul %>%

mutate(temp=ifelse(edu1==1,1,temp))

seoul %>% .$temp %>% is.na() %>% sum() #결측값 제거

seoul<- seoul %>% mutate(edu\_c=ifelse(temp<=3,1,

ifelse(temp==4,2,

ifelse(temp==5,3,

ifelse(temp==6,4,5)))))

#earn var

seoul$earn %>% is.na() %>% sum() #결측값 없음

seoul<-seoul %>% mutate(earn\_c=ifelse(earn<5,1,

ifelse(earn<9,2,

ifelse(earn<13,3,

ifelse(earn<17,4,5)))))

#문1

seoul %>% group\_by(sex\_c) %>% summarize(n=n()) %>% mutate(prop=n/sum(n))

seoul %>% group\_by(age\_c) %>% summarize(n=n()) %>% mutate(prop=n/sum(n))

seoul %>% group\_by(edu\_c) %>% summarize(n=n()) %>% mutate(prop=n/sum(n))

seoul %>% group\_by(earn\_c) %>% summarize(n=n()) %>% mutate(prop=n/sum(n))

#문2

seoul %>% group\_by(earn\_c) %>% summarize(mean\_cul=mean(culture),sd\_cul=sd(culture),

mean\_lei=mean(leisure),sd\_lei=sd(leisure),

mean\_str=mean(stress), sd\_str=sd(stress))

#문3

lm(stress~sex\_c+age\_c+edu\_c+earn\_c+culture+leisure, seoul) %>% summary()