

고려대학교 보건대학원 - 보건통계학 종합시험 (2019/9/27)

학과: _____ 학번: _____ 성명: _____

[문항 점수] 각 문항 4점씩 100점 만점 (정답에 유사하면 4점, 비교적 유사하면 2점, 무응답 0점)

[합격 기준] 70점 이상

[답안 작성법] 아래에 답할 것. 만일 여백이 모자라면 뒷면에 해당 문항 번호를 적고, 답을 기술하시오.

별지로 제공된 SPSS 분석결과는 1500명의 노인들의 건강검진 자료를 분석한 결과이다.

분석에 사용된 변수들의 설명은 다음과 같다:

Group	group (1=위험군 / 2=정상군)
Age	연령 (단위: year)
Sex	성별 (0=여자 / 1=남자)
BMI	체질량지수 (body mass index)
SBP	수축기혈압 (systolic BP, 단위: mmHg)
DBP	이완기혈압 (diastolic BP, 단위: mmHg)
preQOL	연구시작 시점 삶의 질 점수 (연속형)
postQOL	연구종료 시점 삶의 질 점수 (연속형)
Baseline_state	기저상태 (1=나쁨 / 2=중간 / 3= 좋음)
Recovery_01	건강 회복 여부 (1=회복되었음 / 0=회복되지 않음)
Recovery_12	건강 회복 여부 (1=회복되었음 / 2=회복되지 않음)

아래 각 물음에 답하시오. 모든 통계분석 결과는 유의수준 5% 하에서 판정하시오.

1. (분석1)은 각 변수들의 기초통계량 분석 결과이다. 이 결과를 사용해 아래 질문에 답하시오.

1. 위험군과 정상군 간 인원수의 오즈(odds)는 얼마인가? 즉, 정상군에 비해 위험군은 인원수가 몇 오즈 높은가? _____

2. 퍼센트(백분율)과 유효 퍼센트(유효 백분율)은 어떤 차이가 있는가?

3. 누적 퍼센트(누적 백분율)는 퍼센트(백분율)에 기초해서 작성되는가, 아니면 유효 퍼센트(유효 백분율)에 기초해서 작성되는가? _____

4. 변수 BMI의 "평균의 표준오차(SE)" 0.1422는 어떻게 계산된 값인가? 수식을 사용해 답할 것.

5. 위에 언급한 "평균의 표준오차(SE)"란 무엇인가? 즉, 표준오차의 개념에 대해 기술하시오.

6. 변수 Age의 경우, 평균 연령은 70.74세, 표준편차는 5.654세이다. 만일 변수 Age의 값들이 정규분포를 한다고 가정하면, 전체 1,500명 중 1,425명(95%)의 연령은 대략 몇 세와 몇 세 사이에 있을 것 같은가?

II. (분석2)는 Group과 Recovery 여부 간 관련성을 분석한 chi-square test 분석결과이다.

7. Group(위험군, 정상군) 간에 Recovery 비율 (즉, 회복율)을 비교하고자 한다.

(A) 위험군의 회복율은 얼마인가? _____

(B) 정상군의 회복율은 얼마인가? _____

(C) 이들 두 군 간 모집단 회복율이 서로 다른지의 여부를 평가하고자 한다. 이를 판단하기 위한 P-value (SPSS output에는 '유의확률'이라는 단어로 제시되고 있음)에 따르면 두 군의 회복율은 서로 다르다고 할 수 있는가? (이유도 적으시오.)

8. 이 자료는 SPSS output [카이제곱 검정] table에 제시된 분석결과들 중, "Pearson 카이제곱" 결과와 "Fisher의 정확검정" 결과 중 어떤 결과를 사용하는 것이 적절한가? 그 이유는?

9. SPSS output [위험도 추정값] table에 제시된 값 1.421의 의미는 무엇인가? 이 값에 대한 해석을 하시오.

10. 위 9번에서 제시된 odds ratio 값 1.421는 통계적으로 유의한 값이다. (즉, 모집단에서 두 군 간 회복 오즈는 차이가 있다고 할 수 있다.) 이렇게 판단할 수 있는 근거는 무엇인가?

III. (분석3)은 Group(위험군, 정상군) 간 평균 수축기혈압을 비교한 Student's t-test 결과이다.

11. 표본 자료에서 관찰된 두 군 간 수축기 혈압의 평균 차이는 얼마인가?

12. 해당 t-test에서 사용된 (모평균 비교에 관한) 귀무가설과 대립가설을 적으시오.

13. 두 군(위험군, 정상군) 간 수축기 혈압의 변동(분산)은 동일하다고 볼 수 있는가? 이에 관한 P-value(유의확률)를 제시하고, 분산의 동일성 여부에 관한 판정을 내리시오.

14. 두 군(위험군, 정상군) 간 수축기 혈압의 변동(분산)은 동일하다고 볼 수 있는가? 이에 관한 P-value(유의확률)를 제시하고, 평균의 동일성 여부에 관한 판정을 내리시오.

IV. (분석4)는 변수 Sex, Age, BMI 및 preQOL을 사용해서 postQOL(연구 종료시점 삶의 질 점수)를 예측하기 위한 다중회귀분석 결과이다.

15. 독립변수들 중 postQOL에 유의한 영향을 주는 변수들은? _____

16. 추정된 회귀식을 작성하시오.

17. [모형 요약] table의 "R 제곱" 0.835의 의미는 무엇인가?

18. 각 독립변수의 효과가 고려된 상태에서, 남자와 여자 간에 평균 postQOL 점수 차이는 어느 집단이 얼마나 높으며, 또한 그 차이는 통계적으로 유의한가?

19. 변수 preQOL의 회귀계수 추정값 0.962의 의미는 무엇인가?

20. 이 회귀모형에 사용된 독립변수들(Sex, AGE, BMI, preQOL) 중 어느 변수가 postQOL 값에 가장 큰 영향을 미치고 있는가? 그 이유는 무엇인가?

21. 변수 Age의 회귀계수 추정값 0.048에 관한 95% 신뢰구간은 (0.009, 0.087)로 얻어졌다. 이 신뢰구간의 의미는 무엇인가?

V. (분석5)는 Recovery_01 (1=회복되었음, 0=회복되지 않음)변수에 영향을 미치는 요인들을 파악하기 위해 다중로지스틱회귀분석(multiple logistic regression analysis)을 실시한 결과이다.

22. 이 결과는 "회복될 위험도(오즈)"에 관한 분석결과인가, 아니면 "회복되지 않을 위험도(오즈)"에 관한 분석결과인가? 그 이유는 무엇인가?

23. 범주형 변수 "Baseline_state", "Sex", "Group"들의 경우, 분석에 사용된 기준범주(reference group)는 각각 어떤 범주들인가?

(A) 변수 Baseline_state의 기준범주: _____

(B) 변수 Sex의 기준범주: _____

(C) 변수 Group의 기준범주: _____

24. 위험군과 정상군 간 회복(recovery) 오즈비(OR)는 얼마인가? 해당 값에 대한 해석은 어떻게 되는가?

25. Baseline_state (나쁨, 중간, 좋음) 범주들 간 회복(recovery) 오즈비(OR)의 크기를 비교하시오.
