

2과목	보건정보데이터분석 (36~60)
출제위원 : (방송대) 이태림	
출제범위 : 교재 전범위 (강의포함)	

36. ProMED-mail, Eurosurveillance, Wildlife Disease Information Node 등에서 세계 각국의 질병 발병 리포트들을 자동으로 수집하고 걸러내고 군집화하고 분류하여 구글지도의 API를 통해 시각화하여 가장 포괄적인 감염질환 정보 네트워크를 실시간으로 보여주는 시스템은? (3점)

- ① 의료정보학(Medical Informatics)
- ② 자가구조화지도(Self Organizing Map)
- ③ 건강지도(Health Map)
- ④ 지리정보시스템(GIS)

37. 다음 중 질병발생의 상대적 위험도를 나타내는 인덱스는? (2점)

- ① 오즈비
- ② 첨도
- ③ 조건부확률
- ④ 변동계수

38. 현재 의료행위가 의료기관을 중심으로 이루어지는 환경에서 모든 진료기록정보를 정보주체인 환자에게 넘겨주는 것은? (2점)

- ① 분산형 EHR
- ② EHR(Electronic Health Record)
- ③ EMR(Electronic Medical Record)
- ④ PHR(Personal Health Record)

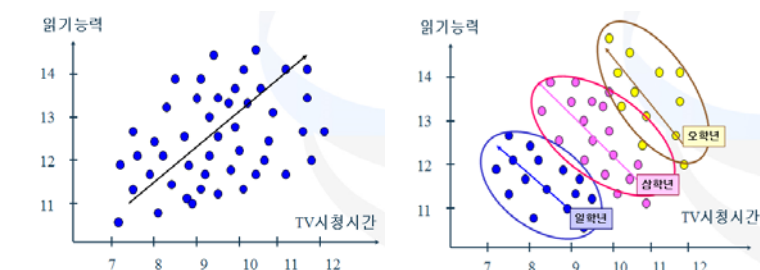
39. 데이터크리닝 과정에서 이상치를 발견하는 방법이 아닌 것은? (3점)

- ① 히스토그램 작성
- ② 산점도 작성
- ③ 결정계수
- ④ 점도표 작성

40. 흡연이 폐경을 앞당기는지 조사하여 로지스틱 모형을 구하기 위하여 흡연여부와 연령 폐경유무를 조사하였다. 이 연구에서 종속변수는 무엇인가? (3점)

- ① 각 연령별 흡연자 비율
- ② 폐경유무
- ③ 연령
- ④ 흡연유무

41. 연구자는 초등학생 등의 TV시청시간에 따라 문장 독해력을 알고자 산점도를 작성했더니 양의 상관관계가 나와 게임을 많이 할수록 독해력이 높다고 나왔다. 그런데 독해력에 영향을 끼칠 수 있는 요인으로 학년에 따라 나누어 산점도를 작성하니 학년별 산점도는 음의 상관관계가 작성되었다. 이러한 현상을 일컫는 말은 무엇인가? (3점)



- ① 중심극한정리
- ② 대수정리
- ③ 다중공선성
- ④ 심프슨 파라독스

42. 다음 R 함수 중 반복측정(repeated measure)을 할 때 사용하는 것은? (4점)

- ① survreg(mixed model) 함수
- ② library(multcomp) 함수
- ③ mean(cftest) 함수
- ④ t.test(levels) 함수

43. 산모의 흡연이 저체중아의 출산을 유발한다는 것을 증명하기 위해 다음과 같은 자료를 얻었다. 이 조사에서 관심있는 통계적 문제는? (3점)

흡연유무	저체중 1.5~3.0	정상체중 3.0~4.0
흡연	180	112
비흡연	148	420
	328	532

- ① 동질성 χ^2 검정
- ② 로지스틱 회귀
- ③ Fisher의 정확검정 χ^2
- ④ McNemar검정

44. 다음 중 공분산분석에서 잔차분석을 위해서 필요한 변수가 아닌 것은? (4점)

- ① Residual
- ② Standardized Residual
- ③ Observed Quantiles
- ④ Fitted Value

45. 반복처리 자료분석에서 처리 간의 차이 혹은 처리 간의 교호작용에 차이가 있는지를 검정하기 위해 사용하는 검정통계량은? (4점)

- ① t -통계량
- ② F -통계량
- ③ χ^2 -통계량
- ④ Z -통계량

46. 다음과 같이 서로 다른 측정단위로 잰 자료에 대해 각각의 왜도(skewness)를 비교했을 때 옳은 설명은? (4점)

<p><자료 1></p> <p>유방에서 제거한 종양의 직경(단위 cm)</p> <p>4.0 4.3 3.9 3.5 3.7 4.3 4.6 3.7</p> <p>$\bar{x} = 4$ $s = 0.47$ 왜도 = 0.9</p>	
<p><자료 2></p> <p>환자들의 콜레스테롤 농도(mg/100ml)</p> <p>160 182 270 267 165 178 280 168 272 218</p> <p>$\bar{x} = 216$ $s = 50.48$ 왜도 = 0.01</p> <p>$CV = \frac{s}{\bar{x}}$</p>	

- ① 자료1의 평균치는 작고 변이계수는 커서 분포가 평퍼짐하다.
- ② 자료2의 분포가 좌우대칭에 가깝다.
- ③ 자료2의 평균치가 크고 변이계수도 커서 분포가 자료1에 비해 뾰족하게 집중되어 있다.
- ④ 측정단위가 달라 비교할 수 없다.

47. 모수적 모형을 이용한 생존함수의 추정과 비교에 이용하는 분포가 아닌 것은? (3점)

- ① 지수분포: 위험률이 시간에 대해 일정한 경우 $h(t) = \lambda$
- ② 콰베르츠 분포: 위험함수를 시간에 대한 선형함수로 $\log h(t) = \mu + \alpha t$
- ③ 포아송분포: 사망이 희귀한 사건의 발생으로 $P(y_i | x_i \beta) = \frac{e^{-x_i \beta} (x_i \beta)^{y_i}}{y_i!}$
- ④ 와이블 분포: 위험함수가 $\log h(t) = \mu + \alpha \log t$ 인 분포

48. 생존시간에 영향을 미치는 요인(공변량) x_1, x_2, \dots, x_k 를 생각할 때 위험률이 시간에 대해 일정한 경우 생존함수 $S(t) = e^{-\lambda t}$, 확률밀도함수 $f(t) = \lambda e^{-\lambda t}$ 의 분포는? (3점)

- ① 로그-정규분포 ② 콰페르츠 분포
- ③ 와이블 분포 ④ 지수분포

49. 생존 시간을 몇 개의 구간으로 나누어 생존 자료를 요약하고 생존함수를 추정하기 위해서는 다음 중 어떤 것을 사용해야 하는가? (2점)

- ① COX의 회귀 모형
- ② 누적한계추정치
- ③ 생명표 방법
- ④ 지수분포를 이용한 모수적 방법

50. 생존함수의 모수적 모형의 적합도 검토를 위한 옳은 방법은? (2점)

- ① AIC 값을 비교하여 이 값이 더 작은 모형을 선택한다.
- ② 로그-우도를 출력한 다음 이들을 비교하여 절대값이 가장 큰 모형을 택할 수 있다.
- ③ 그래프를 이용하여 시간 t에 대한 $-\log S(t)$ 를 그렸을 때 그래프의 기울기가 큰 모형을 택한다.
- ④ 로그-정규분포 경우에 $\log t$ 에 대한 $\log((1-S(t))/S(t))$ 가 1을 통과하는 직선이 되는 경우 택한다.

51. 임상시험 계획위배가 발생한 경우 환자가 실제로 어떤 처리를 받았느냐에 관계없이 처음에 할당된 처리집단으로 간주하고 분석하는 방법은? (2점)

- ① 관측연구
- ② 집단축차검정
- ③ 임상시험계획위배
- ④ 처리의향분석

52. 평균 비교 시 환자수 계산을 위해 필요하지 않은 값은? (4점)

- ① 대조군과 실험군에서의 평균수준의 변화인 델타(δ)
- ② 두 평균치(μ_1, μ_2)
- ③ 유의수준(α)
- ④ 검정력($1-\beta$)

53. 임상시험을 성공적으로 수행하기 위해서 임상시험의 진행상황을 도중에 점검하는 것으로 임상시험이 임상시험계획서 대로 진행되고 있는지 시험약의 부작용 등을 검토하는 분석은? (3점)

- ① 중간분석
- ② 처리의향분석
- ③ 집단축차검정
- ④ 비열등성검정

54. 동등하다는 것을 대립가설로 놓고 두 효과가 다르다는 것을 귀무가설로 놓고 두 치료법의 효과를 검정하는 방법은? (3점)

- ① 집단축차검정 ② 중간분석
- ③ 동등성 검정 ④ 비열등성검정

55. OECD가 권고하는 개인정보 보호를 위한 8개 원칙에 해당되지 않는 것은? (2점)

- ① 수집제한과 정보정확성의 원칙
- ② 비공개와 책임의 원칙
- ③ 목적명확화와 이용제한의 원칙
- ④ 안전보호 및 개인참가의 원칙

56. 환자와 의사 양자 외에 다양한 전문가, 비전문가 참여자들이 네트워크 속에서 긴밀히 소통하는 공유된 의사결정 구조로 진행되는 것은? (2점)

- ① 자기구조화 의학 ② 네트워크 의학
- ③ Health map ④ 참여의학

57. 미래 의학의 방향을 나타내는 개념이 아닌 것은? (2점)

- ① 바이오 빅데이터에 기반한 예측의학(Predictive)
- ② 개인화된 바이오 빅데이터의 관리와 분석을 통한 맞춤의학(Personalized)
- ③ 개인정보 절대공개에 근거한 개인 맞춤형치료
- ④ 모바일 등 정보통신 네트워크를 통한 참여의학(Participatory)

58. 임상시험 연구의 4단계 중 독성이나 부작용 등을 관찰하는 단계는? (2점)

- ① 제1상(Phase I)
- ② 제2상(Phase II)
- ③ 제3상(Phase III)
- ④ 제4상(Phase IV)

59. 미래의 데이터의학의 구현을 위한 개인적인 빅 데이터에 포함되지 않는 것은? (2점)

- ① 유전체, 단백질, 시스템 생물학 데이터(바이오 데이터)
- ② OECD에서 발표하는 국제보건 의료 국가통계(국가보건통계)
- ③ 개인 건강정보, EMR, EHR, PHR(의료데이터)
- ④ 모바일 환경에서 생성되는 일상생활 및 생체정보(라이프 로그 데이터)

60. 다음과 같이 서로 다른 측정단위로 잰 자료에 대해 각각의 변동계수(CV)를 비교했을 때 옳은 설명은? (3점)

<자 료 1>

유방에서 제거한 종양의 직경(단위 cm)

4.0 4.3 3.9 3.5 3.7 4.3 4.6 3.7

$\bar{x} = 4$ $s = 0.47$

<자 료 2>

환자들의 콜레스테롤 농도(mg/100ml)

160 182 270 267 165 178 280 168 272 218

$\bar{x} = 216$ $s = 50.48$

$CV = \frac{s}{\bar{x}}$

- ① 자료1의 평균치는 작고 변이계수는 커서 분포가 평퍼짐하다.
- ② 자료2의 평균치가 크고 변이계수도 커서 분포가 자료1에 비해 뾰족하게 집중되어 있다.
- ③ 자료2의 콜레스테롤농도가 종양직경 보다 퍼짐정도가 크다.
- ④ 측정단위가 달라 비교할 수 없다.