

시험 안내

- 시험시간 19:00-22:00이며 22:00까지는 반드시 답안 업로드
  - 자료분석은 R이나 SAS 모두 사용 가능
  - 문제 번호와 함께 답안 작성
- 

폐질환인 ILD를 심각하게 진행 (progression)시키는 위험인자를 알아보기 위하여 자료를 수집하여 `ILD_final.csv`로 정리하였다. `ILD_final.csv`의 첫번째 행에 각 범주형 변수의 범주설명이 주어져있다. 변수들은 다음과 같다.

- `time` : 진행된 시간
- `status2`: 진행 여/부, 약간의 진행 (자료분석시 0과 2를 중도절단으로 처리)
- `age`, `sex`: 환자의 나이와 성별
- `FEV1FVC` : 폐기능검사에서 얻어지는 FEV1/FVC 비
- `honeycombing` : 폐 영상에 벌집모양 존재 여부

1. `ILD_final` 데이터를 R 이나 SAS로 변수명을 그대로 유지하여 읽어들이고 각 변수별로 요약하여 표로 정리하시오.
2. 진행까지의 시간에 대한 위험 함수와 누적 위험 함수를 그리고 그 모양을 설명하시오. 지수 모형이 적절한가?
3. 진행까지의 시간과 변수 `age`, `sex`, `smoking3`, `FEV1FVC` 및 `honeycombing`을 공변량으로 하는 모형을 Cox비례위험모형과 Weibull 비례위험 적합하시오.
  - 각 변수 하나씩 공변량으로 하는 5개의 일변량 비례위험모형 적합
  - 모든 변수들을 공변량으로 하는 다변량 비례위험모형 적합

Cox비례위험모형과 Weibull 비례위험 적합 결과를 표로 정리하시오. 공변량들이 사망까지의 시간과 통계적으로 유의하게 연관되어 있는가? 위험 증가와 관련된 변수 및 위험 감소와 관련된 변수는 무엇인가?

변수 이름	일변량 분석		다변량 분석	
	HR (95% CI)	p-value	HR (95% CI)	p-value
age				
sex				
smoking3				
Never	reference group			
Current				
Ex				
FEV1FVC				
honeycombing				

#### Cox 비례위험모형

변수 이름	일변량 분석		다변량 분석	
	HR (95% CI)	p-value	HR (95% CI)	p-value
age				
sex				
smoking3				
Never	reference group			
Current				
Ex				
FEV1FVC				
honeycombing				

#### Weibull 비례위험모형

4. **survminer** 패키지를 사용하여 성별 Kaplan-Meier 생존함수 추정량을 한 그래프에 성별로 다른 색을 사용하여 그리고 해석하고, log-rank 검정 결과 p-value도 그래프에 함께 나타내시오. smoking그룹별로도 동일한 분석을 수행하시오.
5. **survminer** 패키지의 적절한 함수를 사용하여 잔차분석을 통하여 위험함수를 설명하는 연속형 공변량 age과 FEV1FVC의 알맞는 함수 변환을 결정하시오. 이를 위해 사용된 잔차의 이름은 무엇인가?
6. 최적의 모형을 선택하고 그 결과를 표로 정리하고 해석하시오. 잔차분석을 통하여 비례성 가정, 변수변환, 이상치 등에 대하여 고려하고 이에 대하여 언급하시오