

# 금융통계 및 시계열 분석 제3강 요약

## 고급 자료 처리

### 자료형 변환

- 자료형 확인 : `is.xxx()` 명령 사용
- 자료형 변환 : `as.xxx()` 명령 사용
- 자료형 변환 방지 : `I()` 명령 사용

### apply 명령

- `apply(X, MARGIN, FUN)` : 행렬/어레이에서 루프 방향을 지정
- `lapply(X, FUN)` : 벡터/리스트에서 원소별로 루프
- `sapply(X, FUN)` : `lapply` 결과를 벡터로 변환
- `vapply(FUN, ...)` : `apply` 결과의 초기값을 미리 지정
- `mapply(FUN, ...)` : 복수 개의 벡터/리스트 루프
- `tapply(X, INDEX, FUN)` : 카테고리별로 루프
- `by(FUN, ...)` : 데이터프레임 입력
- `aggregate(FUN, ...)` : 데이터프레임 출력

### 자료 분할

- `split(data, index)` : 카테고리 인덱스 벡터에 따라 분할
- `cut(data)` : 수치값에 따라 분할

### Long/Wide 변환 : reshape2 패키지

- `melt(data, id.vars, variable.name, value.name)` : Wide -> Long
- `dcast(data, id.columns field.columns, value.var)` : Long -> Wide

### 데이터프레임 컬럼 분할

- `with(data, expression)` : express 사용시 data 이름 생략
- `attach(data)` : 워크스페이스에 컬럼 분할하여 로드
- `detach(data)` : 워크스페이스에 로드된 컬럼 삭제

### 데이터프레임 컬럼 변환/추가

- `transform(data, ...)` : 전체 변환 및 추가
- `ddply(data, category.column, transform, ...)` : 그룹별 변환 및 추가

### 데이터프레임 머지

- `merge(x, y, by.x, by.y)`
  - `x, y` : 데이터프레임
  - `by.x, by.y` : 각 데이터프레임에서 join할 컬럼 이름

## 기술통계

### 기술통계 방법론

- 요약통계 (summary statistics)
- 분할표 (table)
- 그래픽 (histogram, density plot)

### 요약통계

- `print()` : 전체 값 표시
- `head()` : 앞부분 값 표시
- `tail()` : 뒷부분 값 표시
- `summary()` : 평균/최대/최소 등
- `fivenum()` : 평균/최대/최소 등
- `quantile()` : 분위수
- `length(x)` : 샘플 갯수
- `max(x, na.rm=FALSE)` : 최대값
- `min(x, na.rm=FALSE)` : 최소값
- `which.min(x)` : 최대값 위치
- `which.max(x)` : 최소값 위치
- `mean(x, na.rm=FALSE)` : 산술 평균
- `median(x, na.rm=FALSE)` : 중간값
- `sum(x, na.rm=FALSE)` : 합계
- `sd(x, na.rm=FALSE)` : 표준편차
- `var(x, na.rm=FALSE)` : 분산
- `mad(x, na.rm=FALSE)` : 중간값절대변차

### 분할표

- `table(x, ...)`
- `xtabs(formula, data)`
- `ftable(x, ...)`
- `margin.table(x, margin)`
- `prop.table(x, ...)`
- `frequency(), deltat()` : 빈도 및 주기
- `window(ts, start, end)` : 일부 구간
- `lag(ts, k)` : k 단위만큼 지연
- `diff(ts, lag)` : 시계열 지연 차이
- `cbind(ts1, ts2)` : 시계열 합치기
- `ts.union(ts1, ts2)` : 시계열 합치기
- `ts.intersect(ts1, ts2)` : 데이터가 없는 부분 생략

### 그래프

- `stem(data)` : stem-leaf 플롯
- `hist(data)` : 히스토그램
- `density` : 커널 밀도