

## Stata Quiz:

### [20대 총선 선거 데이터] (LI 2017S Midterm Q4 변형)

(Save your code as "quiz\_홍길동.do")

중앙선거관리위원회가 제공하는 자료를 사용하여 2016년 4월 실시된 20대 지역구 국회의원 개표 결과를 분석하는 문제입니다. 서울의 강남구와 서초구에 속하는 5개 지역구 국회의원 선거구 자료를 이용해서 다음 질문에 답하세요. CSV파일과 Excel 파일 중 본인이 편한 자료를 사용해서 분석을 진행하면 됩니다.

이 문제는 Stata 사용 능력을 보기 위한 것이기 때문에 엑셀에서 수작업으로 답을 맞추려고 하지 말고, 어느 단계까지는 관련된 숫자를 확인하는 과정이 담긴 코드를 작성하기 바랍니다.

1번 문제를 풀기 위해 “강남구갑” 자료를 Stata로 가져오고, 읍면동명이 합계, 거소·선상투표, 관외사전투표, 국외부재자투표, 잘못 투입·구분된 투표지인 행은 분석 대상에서 제외하세요.

[국회의원선거][서울특별시][강남구갑][강남구]		
읍면동명	투표구명	선거인수
합계		168,266
거소·선상투표		124
관외사전투표		6,347
국외부재자투표		1,233
신사동	소계	14,827
	관내사전투표	961
	신사동제1투	2,841
	신사동제2투	2,382
	신사동제3투	2,532
	신사동제4투	3,242
	신사동제5투	2,869
잘못 투입·구분된 투표지		0

1. “강남구갑” 자료를 사용하는 문제입니다. 투표수 변수를 볼 때 읍면동 내 소계 대비 관내사전투표 비율이 가장 높은 동은 어느 곳이며 그 비율은 얼마인가요?
2. 각 읍면동 하위 구분인 투표구명이 소계, 관내사전투표인 경우는 이제 분석 대상에서 제외하세요. 선거인수 대비 투표수를 투표율로 볼 때, 투표율이 가장 높은 두 곳의 투표구는 어느 곳이며, 그 비율은 얼마인가요?

이제 아래 3~5번 문제를 풀기 위해서 **강남의 5개 지역의 데이터를 하나의 데이터로 합치세요.** (Hint: 각 선거구별로 출마한 후보자의 수가 다릅니다. 하지만 자료를 잘 살펴보면 컬럼의 개수는 동일하고, Z에는 "계", AA에는 "무표 투표수", AB에는 "기권수" 정보가 담겨 있습니다.

1~2번 문제를 풀 때 했던 것처럼 읍면동명이 합계, 거소·선상투표, 관외사전투표, 국외부재자투표, 잘못 투입·구분된 투표지인 행은 분석 대상에서 제외하세요.

또한 각 읍면동 하위 구분인 투표구명이 소계, 관내사전투표인 경우는 분석 대상에서 제외하세요.

3. 이렇게 생성된 **5개 선거구의 총 투표소는 몇 곳이고 총 출마 인원은 몇 명인가요?**
4. 앞에서 삭제한 행은 제외하고 데이터에 남겨진 투표수만을 기준으로 답하세요.  
**"읍면동명"을 기준으로 투표수 대비 1위 후보자의 득표수의 비율이 가장 높은 행정동은 어디이며 그 비율은 얼마인가요?**
5. (난이도 높음) **"투표구명"을 기준으로 득표수가 가장 많은 1 위 후보자와 그 다음으로 많은 2 위 후보자의 득표수 차이를 나타내는 변수를 만들고, 그 차이가 가장 큰 3 곳의 투표구명을 적으세요.**