

I. Excel 부분합 계산

II. R 부분합 계산

1. 부분합 계산

- 부분합 계산이란 특정한 필드를 기준으로 데이터를 그룹별로 분류
- 각 그룹별로 합계뿐만 아니라 평균, 최댓값, 최솟값, 표본표준편차, 표준편차, 표본분산, 분산, 숫자개수 등 부분적인 계산을 자동으로 해 주는 기능을 말함

2. 실습

- 강기춘 홈페이지에서 subtotal.xlsx 파일을 다운로드
- 경영, 경제, 회계, 무역 등 4개학과
- 각 학과별 1-4학년 학생 20명

| | A | B | C | D | E | F | G |
|----|-----|----|------|------|------|----|----|
| 1 | 이름 | 출석 | 중간고사 | 기말고사 | 총점 | 학년 | 학과 |
| 2 | 강리라 | 9 | 44 | 18 | 42.8 | 1 | 경영 |
| 3 | 이동심 | 7 | 15 | 12 | 24.8 | 1 | 경영 |
| 4 | 강희영 | 9 | 75 | 76 | 78.4 | 1 | 경영 |
| 5 | 김민찬 | 7 | 62 | 76 | 69.2 | 1 | 경영 |
| 6 | 박자영 | 10 | 52 | 60 | 64.8 | 1 | 경영 |
| 7 | 이은영 | 6 | 16 | 20 | 26.4 | 2 | 경영 |
| 8 | 김은주 | 7 | 83 | 72 | 76 | 2 | 경영 |
| 9 | 김일원 | 9 | 16 | 78 | 55.6 | 2 | 경영 |
| 10 | 방인성 | 8 | 64 | 55 | 63.6 | 2 | 경영 |
| 11 | 문성철 | 5 | 42 | 67 | 53.6 | 2 | 경영 |
| 12 | 강지천 | 0 | 27 | 13 | 16 | 3 | 경영 |
| 13 | 박나한 | 9 | 73 | 65 | 73.2 | 3 | 경영 |
| 14 | 심은숙 | 5 | 40 | 70 | 54 | 3 | 경영 |
| 15 | 강민숙 | 7 | 30 | 64 | 51.6 | 3 | 경영 |
| 16 | 김영두 | 10 | 74 | 65 | 75.6 | 3 | 경영 |
| 17 | 강진혜 | 1 | 45 | 24 | 29.6 | 4 | 경영 |
| 18 | 강춘호 | 7 | 69 | 83 | 74.8 | 4 | 경영 |
| 19 | 이승원 | 10 | 35 | 58 | 57.2 | 4 | 경영 |
| 20 | 박주원 | 10 | 42 | 50 | 56.8 | 4 | 경영 |
| 21 | 김양욱 | 6 | 63 | 81 | 69.6 | 4 | 경영 |

- 예 1 : 학생 전체의 중간고사, 기말고사 및 총점의 평균을 계산
 - =average(c2:c81)로 중간고사 평균을 계산하고 나머지도 동일한 방법으로 계산

| | | | | | |
|----|--|---------|--------|------|--|
| 82 | | 50.3375 | 48.225 | 54.2 | |
|----|--|---------|--------|------|--|

- 예 2 : 학년별 중간고사, 기말고사 및 총점의 평균을 계산
 - 학년 순으로 그룹을 묶을 경우 학년 필드에 셀 포인터를 위치
 - [홈]-[편집] 그룹에서 [정렬 및 필터]를 클릭하여 [숫자 오름차순 정리]를 선택
 - 학년 순으로 정렬되면 [데이터]-[윤곽선] 그룹에서 [부분합]을 클릭
 - 부분합 대화상자에 아래 왼쪽 그림과 같이 입력
 - 학년별 중간고사, 기말고사 및 총점의 평균을 계산

? X

부분합

그룹화할 항목(A):

학년

사용할 함수(U):

평균

부분합 계산 항목(D):

출석

중간고사

기말고사

총점

학년

학과

새로운 값으로 대치(C)

그룹 사이에서 페이지 나누기(P)

데이터 아래에 요약 표시(S)

모두 제거(R) **확인** 취소

| | A | B | C | D | E | F | G |
|----|-----|----|---------|---------|-------|----|----|
| 1 | 이름 | 출석 | 중간고사 | 기말고사 | 총점 | 학년 | 학과 |
| 2 | 강리라 | 9 | 44 | 18 | 42.8 | 1 | 경영 |
| 3 | 이동심 | 7 | 15 | 12 | 24.8 | 1 | 경영 |
| 4 | 강희영 | 9 | 75 | 76 | 78.4 | 1 | 경영 |
| 5 | 김민찬 | 7 | 62 | 76 | 69.2 | 1 | 경영 |
| 6 | 박자영 | 10 | 52 | 60 | 64.8 | 1 | 경영 |
| 7 | 문봉기 | 10 | 51 | 91 | 76.8 | 1 | 경제 |
| 8 | 고이영 | 7 | 64 | 22 | 48.4 | 1 | 경제 |
| 9 | 김철수 | 9 | 41 | 31 | 46.8 | 1 | 경제 |
| 10 | 오시창 | 8 | 61 | 43 | 57.6 | 1 | 경제 |
| 11 | 강라세 | 8 | 52 | 50 | 56.8 | 1 | 무역 |
| 12 | 김건업 | 7 | 73 | 91 | 79.6 | 1 | 무역 |
| 13 | 백은희 | 7 | 54 | 47 | 54.4 | 1 | 무역 |
| 14 | 신진호 | 8 | 47 | 71 | 63.2 | 1 | 무역 |
| 15 | 김상대 | 10 | 60 | 80 | 76 | 1 | 회계 |
| 16 | 민양수 | 7 | 38 | 45 | 47.2 | 1 | 회계 |
| 17 | 고희선 | 5 | 37 | 34 | 38.4 | 1 | 회계 |
| 18 | 박신정 | 6 | 49 | 49 | 51.2 | 1 | 회계 |
| 19 | | | 51.4706 | 52.7059 | 57.44 | 1 | 평균 |
| 20 | 이은영 | 6 | 16 | 20 | 26.4 | 2 | 경영 |
| 21 | 김은주 | 7 | 83 | 72 | 76 | 2 | 경영 |
| 22 | 김왕원 | 9 | 16 | 78 | 55.6 | 2 | 경영 |
| 23 | 방인성 | 8 | 64 | 55 | 63.6 | 2 | 경영 |
| 24 | 문성철 | 5 | 42 | 67 | 53.6 | 2 | 경영 |
| 25 | 김행준 | 7 | 57 | 71 | 65.2 | 2 | 경제 |
| 26 | 고영준 | 8 | 71 | 85 | 78.4 | 2 | 경제 |
| 27 | 정열순 | 9 | 40 | 44 | 51.6 | 2 | 경제 |
| 28 | 박호민 | 7 | 47 | 40 | 48.8 | 2 | 경제 |
| 29 | 박규성 | 3 | 29 | 57 | 40.4 | 2 | 경제 |
| 30 | 강경은 | 9 | 25 | 29 | 39.6 | 2 | 무역 |
| 31 | 박신애 | 8 | 55 | 79 | 69.6 | 2 | 무역 |
| 32 | 이광림 | 9 | 72 | 68 | 74 | 2 | 무역 |
| 33 | 명한숙 | 6 | 86 | 46 | 64.8 | 2 | 무역 |
| 34 | 조재진 | 4 | 39 | 73 | 52.8 | 2 | 무역 |
| 35 | 원옥순 | 10 | 48 | 52 | 60 | 2 | 회계 |
| 36 | 김숙희 | 9 | 41 | 40 | 50.4 | 2 | 회계 |
| 37 | 김지안 | 8 | 36 | 34 | 44 | 2 | 회계 |
| 38 | 공병호 | 6 | 44 | 32 | 42.4 | 2 | 회계 |
| 39 | 김성욱 | 5 | 48 | 31 | 41.6 | 2 | 회계 |
| 40 | | | 47.95 | 53.65 | 54.94 | 2 | 평균 |
| 41 | 강지천 | 0 | 27 | 13 | 16 | 3 | 경영 |
| 42 | 박나한 | 9 | 73 | 65 | 73.2 | 3 | 경영 |
| 43 | 심은숙 | 5 | 40 | 70 | 54 | 3 | 경영 |

- 예 3 : 학과별 중간고사, 기말고사 및 총점의 평균을 파악
- 학과 순으로 그룹을 묶을 경우 학과 필드에 셀 포인터를 위치
- [홈]-[편집] 그룹에서 [정렬 및 필터]를 클릭하여 [텍스트 오름차순 정렬]를 선택
- 학과 순으로 정렬되면 [데이터]-[운곽선] 그룹에서 [부분합]을 클릭
- 부분합 대화상자에 아래 왼쪽 그림과 같이 입력
- 학과별 중간고사, 기말고사 및 총점의 평균을 계산

부분합 ? X

그룹화할 항목(A):

사용할 함수(U):

부분합 계산 항목(D):

- 이름
- 출석
- 중간고사
- 기말고사
- 총점
- 학년

새로운 값으로 대치(C)

그룹 사이에서 페이지 나누기(P)

데이터 아래에 요약 표시(S)

| | A | B | C | D | E | F | G | H |
|----|-----|----|-------|-------|-------|----|-----|----|
| 1 | 이름 | 출석 | 중간고사 | 기말고사 | 총점 | 학년 | 학과 | |
| 2 | 강리라 | 9 | 44 | 18 | 42.8 | 1 | 컴퓨터 | |
| 3 | 이동심 | 7 | 15 | 12 | 24.8 | 1 | 컴퓨터 | |
| 4 | 강희영 | 9 | 75 | 76 | 78.4 | 1 | 컴퓨터 | |
| 5 | 김민찬 | 7 | 62 | 76 | 69.2 | 1 | 컴퓨터 | |
| 6 | 박자영 | 10 | 52 | 60 | 64.8 | 1 | 컴퓨터 | |
| 7 | 이은영 | 6 | 16 | 20 | 26.4 | 2 | 컴퓨터 | |
| 8 | 김은주 | 7 | 83 | 72 | 76 | 2 | 컴퓨터 | |
| 9 | 김일원 | 9 | 16 | 78 | 55.6 | 2 | 컴퓨터 | |
| 10 | 방인성 | 8 | 64 | 55 | 63.6 | 2 | 컴퓨터 | |
| 11 | 문성철 | 5 | 42 | 67 | 53.6 | 2 | 컴퓨터 | |
| 12 | 강지천 | 0 | 27 | 13 | 16 | 3 | 컴퓨터 | |
| 13 | 박나한 | 9 | 73 | 65 | 73.2 | 3 | 컴퓨터 | |
| 14 | 심은숙 | 5 | 40 | 70 | 54 | 3 | 컴퓨터 | |
| 15 | 강민숙 | 7 | 30 | 64 | 51.6 | 3 | 컴퓨터 | |
| 16 | 김영두 | 10 | 74 | 65 | 75.6 | 3 | 컴퓨터 | |
| 17 | 강진혜 | 1 | 45 | 24 | 29.6 | 4 | 컴퓨터 | |
| 18 | 강훈호 | 7 | 69 | 83 | 74.8 | 4 | 컴퓨터 | |
| 19 | 이승원 | 10 | 35 | 58 | 57.2 | 4 | 컴퓨터 | |
| 20 | 박주원 | 10 | 42 | 50 | 56.8 | 4 | 컴퓨터 | |
| 21 | 김양욱 | 6 | 63 | 81 | 69.6 | 4 | 컴퓨터 | |
| 22 | | | 48.35 | 55.35 | 55.68 | | 컴퓨터 | 평균 |
| 23 | 문봉기 | 10 | 51 | 91 | 76.8 | 1 | 경제 | |
| 24 | 고이영 | 7 | 64 | 22 | 48.4 | 1 | 경제 | |
| 25 | 김철수 | 9 | 41 | 31 | 46.8 | 1 | 경제 | |
| 26 | 오시창 | 8 | 61 | 43 | 57.6 | 1 | 경제 | |
| 27 | 강행준 | 7 | 57 | 71 | 65.2 | 2 | 경제 | |
| 28 | 고영준 | 8 | 71 | 85 | 78.4 | 2 | 경제 | |
| 29 | 정열순 | 9 | 40 | 44 | 51.6 | 2 | 경제 | |
| 30 | 박호민 | 7 | 47 | 40 | 48.8 | 2 | 경제 | |
| 31 | 박규성 | 3 | 29 | 57 | 40.4 | 2 | 경제 | |
| 32 | 김주형 | 9 | 64 | 40 | 59.6 | 3 | 경제 | |
| 33 | 양은순 | 9 | 33 | 30 | 43.2 | 3 | 경제 | |
| 34 | 김미수 | 7 | 46 | 20 | 40.4 | 3 | 경제 | |
| 35 | 김영수 | 9 | 37 | 34 | 46.4 | 3 | 경제 | |
| 36 | 고유경 | 9 | 71 | 42 | 63.2 | 3 | 경제 | |
| 37 | 우진호 | 6 | 60 | 45 | 54 | 3 | 경제 | |
| 38 | 이금숙 | 10 | 66 | 50 | 66.4 | 4 | 경제 | |
| 39 | 이준용 | 10 | 72 | 55 | 70.8 | 4 | 경제 | |
| 40 | 장호철 | 8 | 80 | 64 | 73.6 | 4 | 경제 | |
| 41 | 강정희 | 10 | 63 | 39 | 60.8 | 4 | 경제 | |
| 42 | 고일동 | 1 | 53 | 27 | 34 | 4 | 경제 | |
| 43 | | | 55.3 | 46.5 | 56.32 | | 경제 | 평균 |
| 44 | 강라세 | 8 | 52 | 50 | 56.8 | 1 | 무역 | |
| 45 | 김건엽 | 7 | 73 | 91 | 79.6 | 1 | 무역 | |
| 46 | 백은희 | 7 | 54 | 47 | 54.4 | 1 | 무역 | |
| 47 | 신진호 | 8 | 47 | 71 | 63.2 | 1 | 무역 | |
| 48 | 강경은 | 9 | 25 | 29 | 39.6 | 2 | 무역 | |
| 49 | 박신애 | 8 | 55 | 79 | 69.6 | 2 | 무역 | |

1. summarise

- 학생 전체의 기초통계량을 계산
- 예 1 : 중간고사, 기말고사, 총점의 평균을 계산
- 예 2 : 중간고사, 기말고사, 총점의 표준편차를 계산

```
b3-ch2-3.R

library(openxlsx)
library(dplyr)

df<-read.xlsx("http://kanggc.iptime.org/book/data/subtotal-e.xlsx")
df

summary(df)

summarise(df,mean_mid = mean(mid), mean_final = mean(final), mean_total = mean(total))
summarise(df,sd_mid = sd(mid), sd_final = sd(final), sd_total = sd(total))
```

```
> summary(df)
  name          attend          mid          final          total          class          dept
Length:80      Min.   : 0.000      Min.   :15.00      Min.   : 4.00      Min.   : 9.6      Min.   :1.000      Length:80
Class :character 1st Qu.: 6.750      1st Qu.:39.00      1st Qu.:32.00      1st Qu.:45.7      1st Qu.:2.000      Class :character
Mode  :character Median : 8.000      Median :51.00      Median :46.00      Median :54.0      Median :3.000      Mode  :character
          Mean  : 7.388      Mean  :50.34      Mean  :48.23      Mean  :54.2      Mean  :2.575
          3rd Qu.: 9.000      3rd Qu.:63.00      3rd Qu.:65.50      3rd Qu.:64.8      3rd Qu.:3.250
          Max.  :10.000      Max.  :86.00      Max.  :91.00      Max.  :79.6      Max.  :4.000
```

```
> summarise(df,mean_mid = mean(mid), mean_final = mean(final), mean_total = mean(total))
  mean_mid mean_final mean_total
1  50.3375   48.225     54.2
```

```
> summarise(df,sd_mid = sd(mid), sd_final = sd(final), sd_total = sd(total))
  sd_mid sd_final sd_total
1 17.53907 21.21677 15.14808
```

2. summarise()(with group_by)

- 학년별, 학과별 등 그룹의 기초통계량을 계산
- 예 1 : 학년별 중간고사, 기말고사, 총점의 평균을 계산
- 예 2 : 학과별 중간고사, 기말고사, 총점의 평균을 계산

```
b3-ch2-3.R
이전 슬라이드와 동일
class <- df %>%
  group_by(class) %>%
  summarise(mean_mid = mean(mid),
            mean_final = mean(final),
            mean_total = mean(total))
class

dept <- df %>%
  group_by(dept) %>%
  summarise(mean_mid = mean(mid),
            mean_final = mean(final),
            mean_total = mean(total))
Dept
```

```
# A tibble: 4 x 4
  class mean_mid mean_final mean_total
  <dbl> <dbl> <dbl> <dbl>
1 1 51.5 52.7 57.4
2 2 48.0 53.6 54.9
3 3 47 38.8 48.5
4 4 55.6 49.8 57.3
```

```
# A tibble: 4 x 4
  dept mean_mid mean_final mean_total
  <chr> <dbl> <dbl> <dbl>
1 경영 48.4 55.4 55.7
2 경제 55.3 46.5 56.3
3 무역 46.0 51.5 54.2
4 회계 51.8 39.6 50.6
```

3. mutate()

- 새로운 변수(열)을 생성
- 실습(simulation 데이터)
- 예 1 : 중간고사와 기말고사를 합한 총점 변수 만들기
- 예 2 : 총점을 이용하여 평균 변수 만들기

```
dplyr-simul.R

library(dplyr)
set.seed(1234)
mydata<-data.frame(class = sample(3,10, replace=T),
                    id = sample(30,10, replace=T),
                    mid = as.integer(rnorm(10, mean=60, sd=10)),
                    final = as.integer(rnorm(10, mean=70, sd=8)))

mydata
mutate(mydata, total=mid+final, mean=total/2)
mydata<-mutate(mydata, total=mid+final, mean=total/2)

transmute(mydata, total=mid+final, mean=total/2)
mydata_trans<-transmute(mydata, total=mid+final, mean=total/2)
```

| | class | id | mid | final |
|----|-------|----|-----|-------|
| 1 | 2 | 26 | 68 | 70 |
| 2 | 2 | 6 | 73 | 69 |
| 3 | 1 | 15 | 43 | 75 |
| 4 | 3 | 14 | 53 | 66 |
| 5 | 1 | 20 | 60 | 70 |
| 6 | 1 | 14 | 67 | 70 |
| 7 | 2 | 30 | 53 | 75 |
| 8 | 2 | 24 | 68 | 78 |
| 9 | 3 | 30 | 63 | 56 |
| 10 | 2 | 4 | 63 | 52 |

| | class | id | mid | final | total | mean |
|----|-------|----|-----|-------|-------|------|
| 1 | 2 | 26 | 68 | 70 | 138 | 69.0 |
| 2 | 2 | 6 | 73 | 69 | 142 | 71.0 |
| 3 | 1 | 15 | 43 | 75 | 118 | 59.0 |
| 4 | 3 | 14 | 53 | 66 | 119 | 59.5 |
| 5 | 1 | 20 | 60 | 70 | 130 | 65.0 |
| 6 | 1 | 14 | 67 | 70 | 137 | 68.5 |
| 7 | 2 | 30 | 53 | 75 | 128 | 64.0 |
| 8 | 2 | 24 | 68 | 78 | 146 | 73.0 |
| 9 | 3 | 30 | 63 | 56 | 119 | 59.5 |
| 10 | 2 | 4 | 63 | 52 | 115 | 57.5 |

```
> mutate(mydata, total=mid+final, mean=total/2)
  class id mid final total mean
1     2 26 68  70   138 69.0
2     2  6 73  69   142 71.0
3     1 15 43  75   118 59.0
4     3 14 53  66   119 59.5
5     1 20 60  70   130 65.0
6     1 14 67  70   137 68.5
7     2 30 53  75   128 64.0
8     2 24 68  78   146 73.0
9     3 30 63  56   119 59.5
10    2  4 63  52   115 57.5
```

```
> transmute(mydata, total=mid+final, mean=total/2)
  total mean
1    138 69.0
2    142 71.0
3    118 59.0
4    119 59.5
5    130 65.0
6    137 68.5
7    128 64.0
8    146 73.0
9    119 59.5
10   115 57.5
```