


| | | | |
|-----------------------------|--------------|-----------------|--------|
| 통계패키지활용 자료분석(2024년 겨울학기) | | | |
| 담 당 교 수 : 김 태 수 | | | |
| 강좌 번호 | 100961-31001 | 본인의 과제 자체 평가 | 9/10 점 |
| 기말 고사 : 대학생 성적 분석 및 시각화 | | | |

| | |
|---|-----|
| 이름 | 이지원 |
|  | |

| | |
|-------|---------------|
| 제 출 일 | 2024년 01월 14일 |
| 학 과 | 산업공학과 ITM전공 |
| 학 번 | 22102009 |

목차

I. 개요

- 1) 자료 불러들이기
- 2) 자료 확인 및 변수 분석
- 3) 분석 목표 및 분석 내용

II. R을 이용한 자료 분석 시각화

- 1) 학과에 따른 미적 학점 분포 시각화
- 2) 입학시기에 따른 미적 학점 분포 시각화
- 3) 출신고 지역에 따른 미적 학점 분포 시각화
- 4) 수학실력시험과 미적 학점 사이 상관관계 시각화
- 5) 미적1학점 및 미적2학점 사이의 상관관계 시각화

III. 결론

- 1) 결론
 - (1) 학과에 따른 미적 학점 분포
 - (2) 입학시기에 따른 미적 학점 분포
 - (3) 출신고 지역에 따른 미적 학점 분포
 - (4) 수학실력시험과 미적 학점 사이 상관관계
 - (5) 미적1학점 및 미적2학점 사이의 상관관계

IV. 부록: R script 첨부

I. 개요

1) 자료 불러들이기

기말고사.xlsx가 있는 파일 경로를 복사해 자료를 읽어들인다.

```
> library(readxl)
> Data <- read_excel("C:\\Users\\easy1\\Downloads\\기말고사.xlsx", "자료")
```

2) 자료 확인 및 변수 분석

```
> str(Data)
tibble [882 × 11] (S3: tbl_df/tbl/data.frame)
 $ 학번      : num [1:882] 1001 1002 1003 1004 1005 ...
 $ 단과대학  : chr [1:882] "대학A" "대학A" "대학A" "대학A" ...
 $ 학과      : chr [1:882] "D학과" "D학과" "D학과" "D학과" ...
 $ 재수여부   : chr [1:882] "현역" "재수" "현역" "현역" ...
 $ 입학시기   : chr [1:882] "정시" "정시" "수시" "수시" ...
 $ 입학전형방법: chr [1:882] "성적위주" "성적위주" "입학사정" "입학사정" ...
 $ 출신고지역 : chr [1:882] "E지역" "C지역" "A지역" "E지역" ...
 $ 영어실력시험: num [1:882] 78 91 80 75 67 85 58 72 74 78 ...
 $ 수학실력시험: num [1:882] 65 84 52 79 72 69 60 65 60 67 ...
 $ 미적1학점  : chr [1:882] "A+" "A+" "B0" "B+" ...
 $ 미적2학점  : chr [1:882] "A0" "A+" "B0" "A+" ...
```

- (1) Data\$학번: 학생의 학번(고유식별자)을 나타내는 변수이다.
- (2) Data\$단과대학: 학생이 속한 단과대를 나타내는 변수이다.
- (3) Data\$학과: 학생이 속한 학과를 나타내는 변수이다.
- (4) Data\$재수여부: 학생이 현역인지 재수생인지 여부를 나타내는 변수이다.
- (5) Data\$입학시기: 학생이 정시로 입학했는지 수시로 입학했는지 여부를 나타내는 변수이다.
- (6) Data\$입학전형방법: 학생이 입학한 전형방법(성적위주, 입학사정 등)을 나타내는 변수이다.
- (7) Data\$출신고지역: 학생의 출신 고등학교 지역을 나타내는 변수이다.
- (8) Data\$영어실력시험: 학생의 영어실력시험 점수를 나타내는 변수이다.
- (9) Data\$수학실력시험: 학생의 수학실력시험 점수를 나타내는 변수이다.
- (10) Data\$미적1학점: 학생의 미적1 과목 성적을 나타내는 변수이다.
- (11) Data\$미적2학점: 학생의 미적2 과목 성적을 나타내는 변수이다.

3) 분석 목표 및 분석 내용

(1) 분석 목표

수업시간에 배운 R과 R의 ggplot2, plotly, map등의 패키지를 활용하여 학생들의 입학정보와 시험 성적에 대한 데이터를 분석하고 시각화하여 결론을 도출할 수 있다.

(2) 분석 내용

- 미적 1, 2학점과 다른 요소의 상관관계를 분석하고 어떤 요소가 제일 상관관계가 높은지 결론을 도출한다.
 - 학생들의 입학정보와 미적1, 2 학점의 상관관계를 분석한다.
 - 학생들의 수학실력시험 성적과 미적1, 2 학점의 상관관계를 분석한다.
- 미적1 학점과 미적 2학점의 상관관계를 분석한다.

II. R을 이용한 자료 분석 시각화

미적1, 2학점을 수치형으로 변환

문자열 형태였던 미적1학점과 미적2학점을 A+은 4.5, A0는 4.0, B+는 3.5, B0는 3.0, C+는 2.5, C는 2.0의 수치형 자료로 변환하였다.

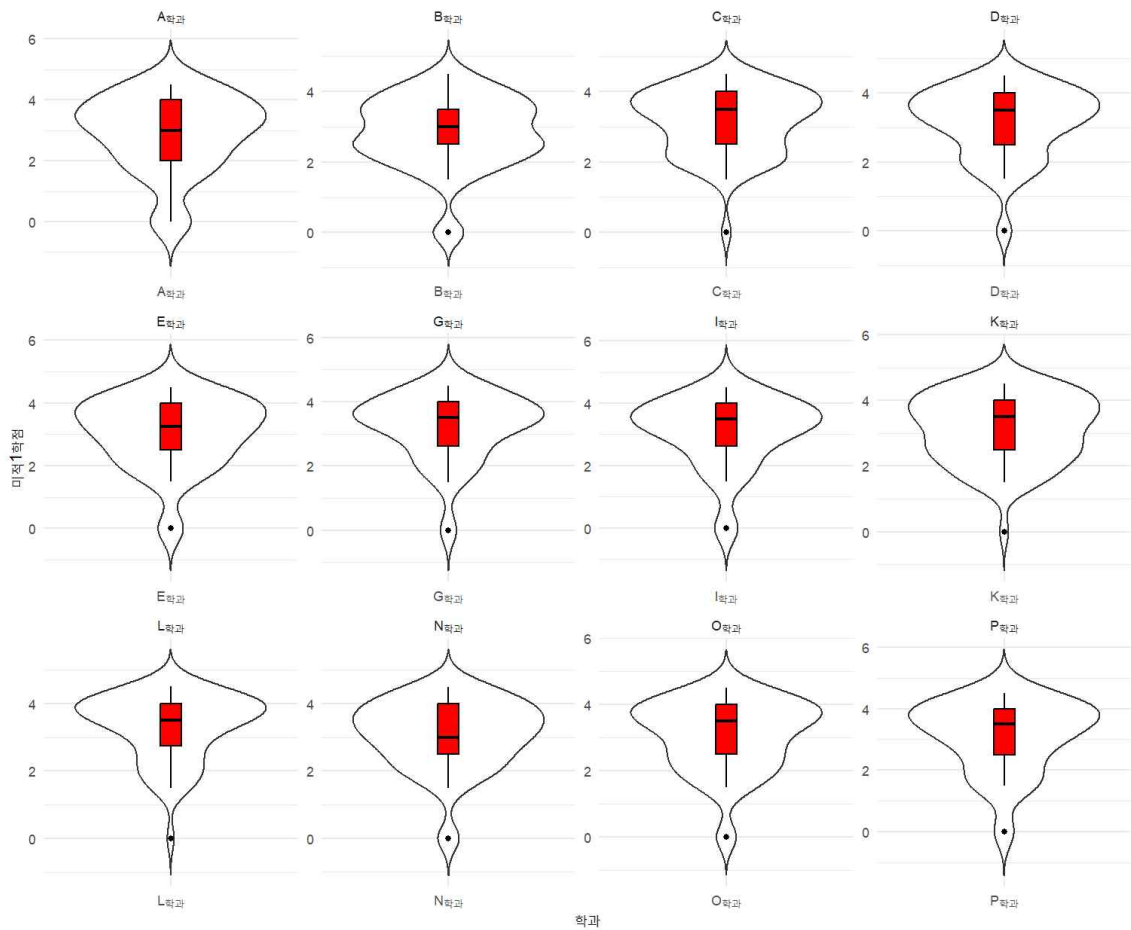
```
> convert_grades <- function(grade) {  
+   grade_mapping <- c("A+" = 4.5, "A0" = 4.0, "B+" = 3.5, "B0" = 3.0, "C+" =  
2.5, "C0" = 2.0, "D+" = 1.5, "F" = 0.0)  
+   return(grade_mapping[grade])  
+ }  
> Data$미적1학점_numeric <- sapply(Data$미적1학점, convert_grades)  
> Data$미적2학점_numeric <- sapply(Data$미적2학점, convert_grades)
```

1) 학과에 따른 미적 학점 분포 시각화

(1) 학과에 따른 미적1 학점 분포 시각화

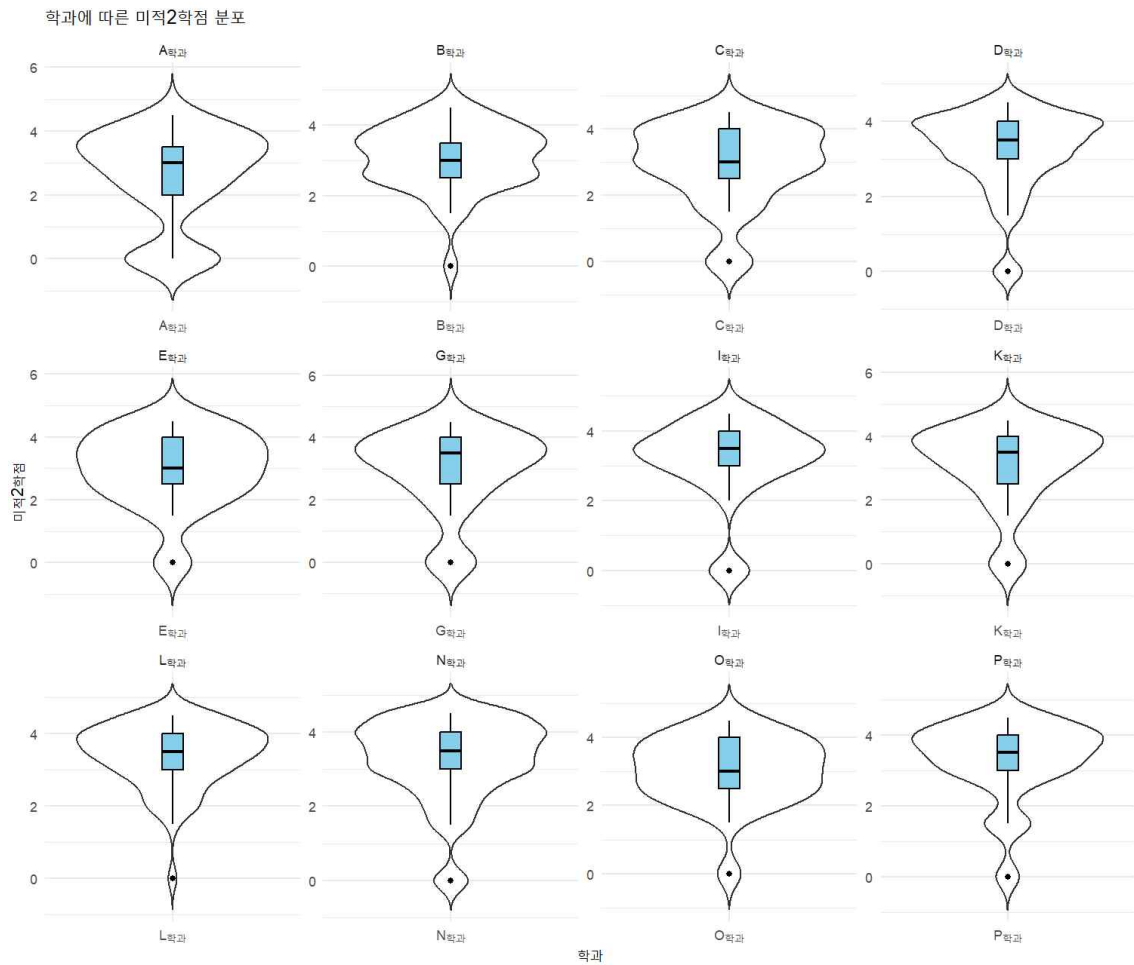
```
> ggplot(Data, aes(x=학과, y=미적1학점_numeric)) +  
+   geom_violin(trim=FALSE) +  
+   geom_boxplot(width=0.1, fill="red", color="black") +  
+   labs(title="학과에 따른 미적1학점 분포", x="학과", y="미적1학점") +  
+   theme_minimal() +  
+   facet_wrap(~학과, scales="free")
```

학과에 따른 미적1학점 분포



(2) 학과에 따른 미적2 학점 분포 시각화

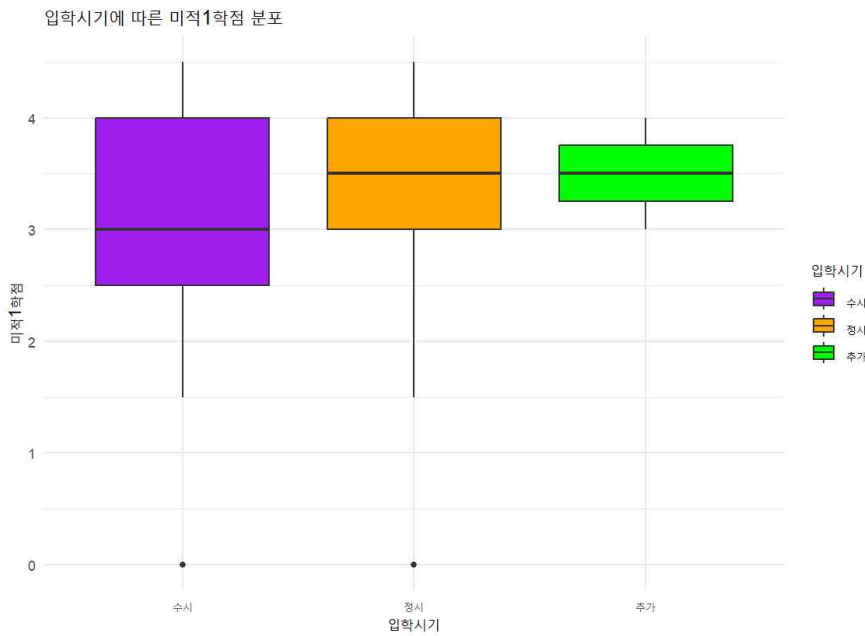
```
> ggplot(Data, aes(x=학과, y=미적2학점_numeric)) +
+   geom_violin(trim=FALSE) +
+   geom_boxplot(width=0.1, fill="skyblue", color="black") +
+   labs(title="학과에 따른 미적2학점 분포", x="학과", y="미적2학점") +
+   theme_minimal() +
+   facet_wrap(~학과, scales="free")
```



2) 입학시기에 따른 미적 학점 분포 시각화

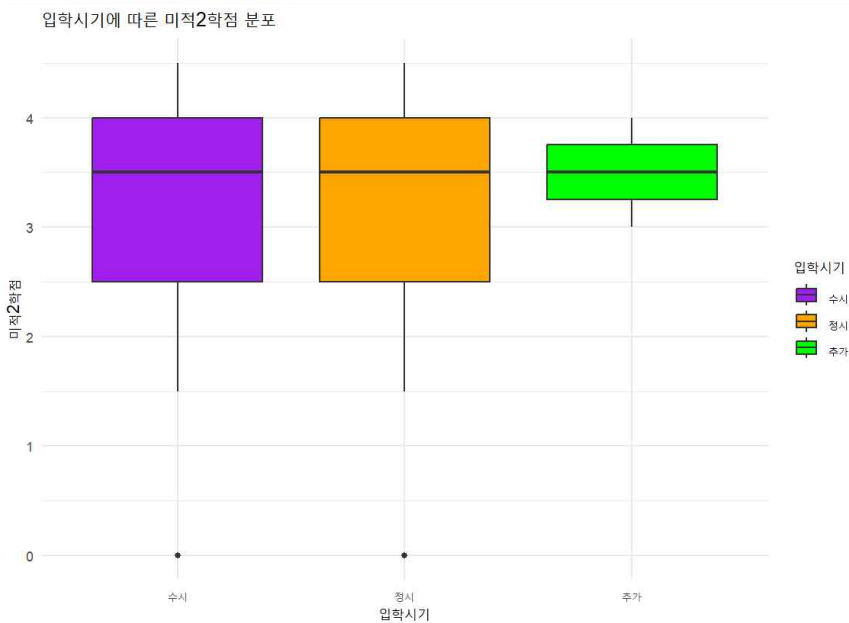
(1) 입학시기에 따른 미적1 학점 분포

```
> ggplot(Data, aes(x=입학시기, y=미적1학점_numeric, fill=입학시기)) +
+   geom_boxplot() +
+   scale_fill_manual(values=c("purple", "orange", "green")) +
+   labs(title="입학시기에 따른 미적1학점 분포", x="입학시기", y="미적1학점") +
+   theme_minimal()
```



(2) 입학시기에 따른 미적2 학점 분포

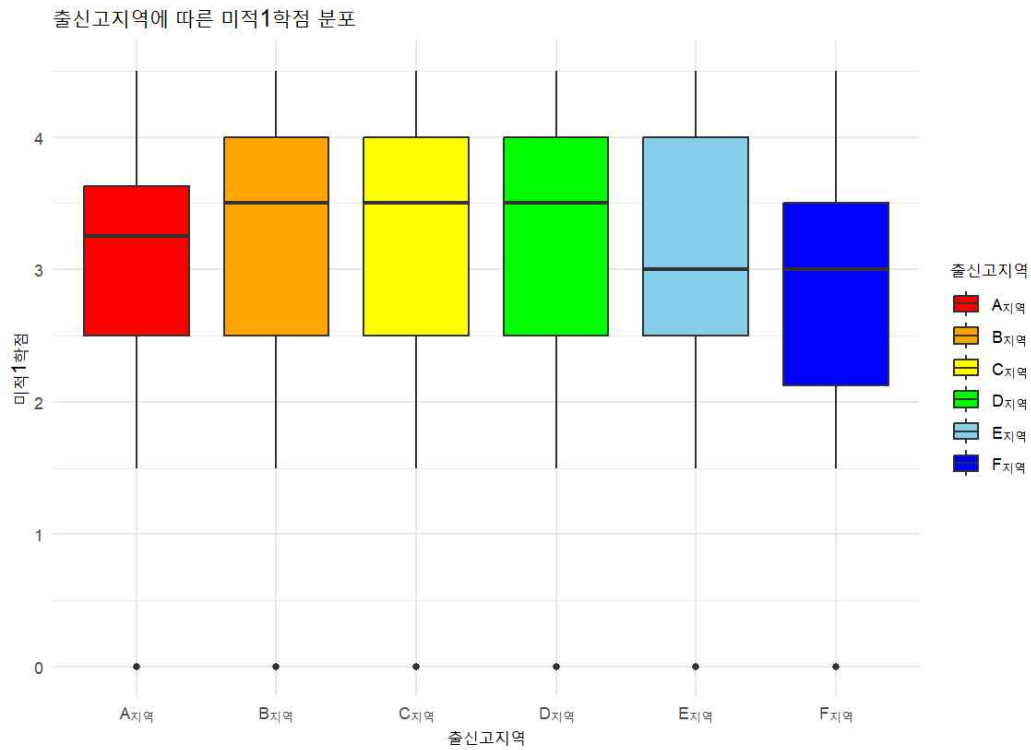
```
> ggplot(Data, aes(x=입학시기, y=미적2학점_numeric, fill=입학시기)) +
+   geom_boxplot() +
+   scale_fill_manual(values=c("purple", "orange", "green")) +
+   labs(title="입학시기에 따른 미적2학점 분포", x="입학시기", y="미적2학점") +
+   theme_minimal()
```



3) 출신고 지역에 따른 미적 학점 분포 시각화

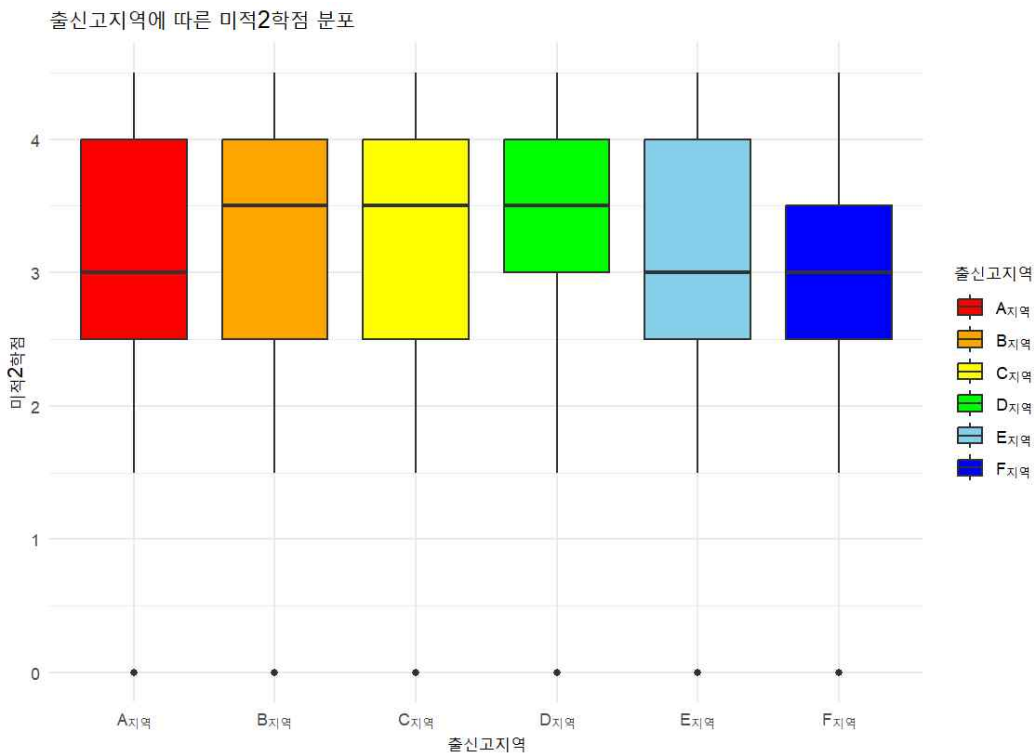
(1) 출신고 지역에 따른 미적1 학점 분포

```
> ggplot(Data, aes(x=출신고지역, y=미적1학점_numeric, fill=출신고지역)) +
+   geom_boxplot() +
+   scale_fill_manual(values=c("red", "orange", "yellow", "green", "skyblue", "blue")) +
+   labs(title="출신고지역에 따른 미적1학점 분포", x="출신고지역", y="미적1학점") +
+   theme_minimal()
```



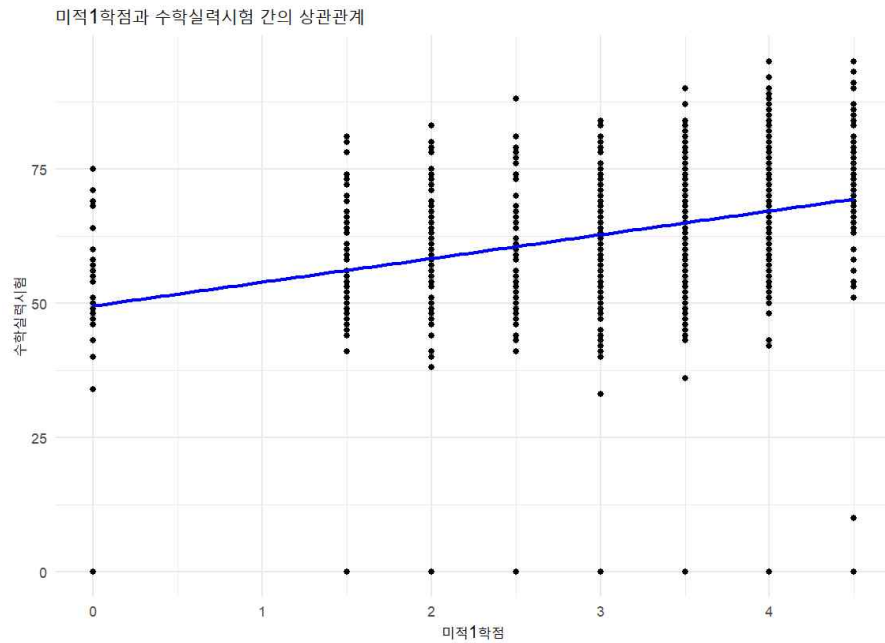
(2) 출신고 지역에 따른 미적2 학점 분포

```
> ggplot(Data, aes(x=출신고지역, y=미적2학점_numeric, fill=출신고지역)) +
+   geom_boxplot() +
+   scale_fill_manual(values=c("red", "orange", "yellow", "green", "skyblue", "blue")) +
+   labs(title="출신고지역에 따른 미적2학점 분포", x="출신고지역", y="미적2학점") +
+   theme_minimal()
```

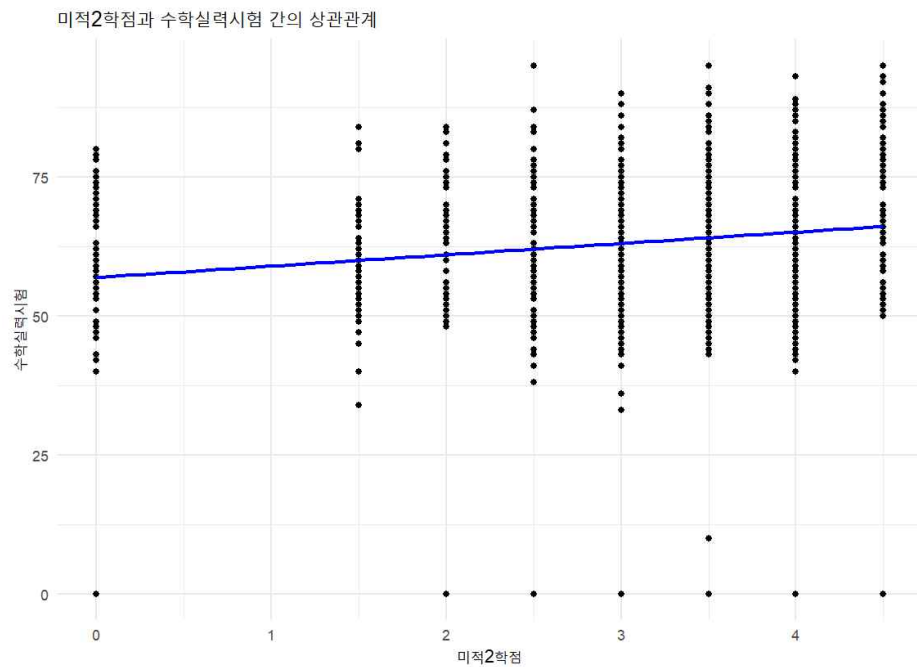


4) 수학실력시험과 미적 학점 사이 상관관계 시각화

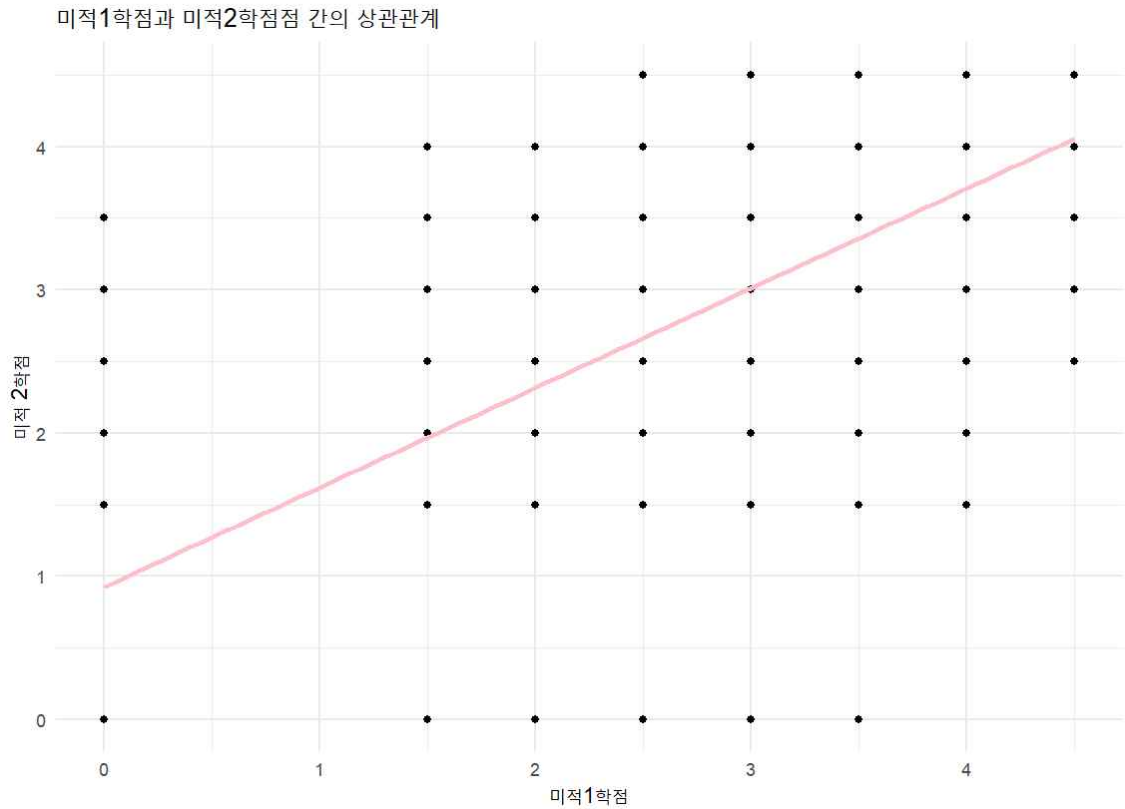
(1) 수학실력시험과 미적1 학점 사이 상관관계



(2) 수학실력시험과 미적2 학점 사이 상관관계



5) 미적1학점 및 미적2학점 사이의 상관관계 시각화



III. 결론 및 자기평가

1) 최종결론

(1) 학과에 따른 미적 학점 분포

- 미적1학점의 평균은 L, G, C, D, K, P, I, N, O, E, A, B순으로 높다.
- 미적2학점의 평균은 L, P, N, D, I, K, B, O, E, G, C, A순으로 높다.

(2) 입학시기에 따른 미적 학점 분포

- 미적1학점의 평균은 정시와 추가합격자가 수시합격자보다 높았다.
- 미적2학점의 평균은 수시, 정시, 추가합격자 모두 비슷한 양상으로 나타났다

(3) 출신고 지역에 따른 미적 학점 분포

- 미적1학점의 평균은 B, C, D, A, E, F순으로 크다.
- 미적2학점의 평균은 A, C, D, B, E, F순으로 크다.

(4) 수학실력시험과 미적 학점 사이 상관관계

미적1학점과 미적2학점 모두 수학실력 시험과 유의미한 관계를 가지고 있다.

(5) 미적1학점 및 미적2학점 사이의 상관관계

미적1학점 및 미적2학점은 서로 유의미한 관계를 가지고 있다. 미적1학점이 좋을수록 미적2학점도 좋은 양상이다.

IV. 부록: R script 첨부

```
library(readxl)
Data <- read_excel("C:\\Users\\easy1\\Downloads\\기말고사.xlsx", "자료")
str(Data)
library(ggplot2)
#미적 1, 2학점을 수치형으로 변환
convert_grades <- function(grade) {
  grade_mapping <- c("A+" = 4.5, "A0" = 4.0, "B+" = 3.5, "B0" = 3.0,
"C+" = 2.5, "C0" = 2.0, "D+" = 1.5, "F" = 0.0)
  return(grade_mapping[grade])
}Data$미적1학점_numeric <- sapply(Data$미적1학점, convert_grades)
Data$미적2학점_numeric <- sapply(Data$미적2학점, convert_grades)
# 학과에 따른 미적1학점
ggplot(Data, aes(x=학과, y=미적1학점_numeric)) +
  geom_violin(trim=FALSE) +
  geom_boxplot(width=0.1, fill="red", color="black") +
  labs(title="학과에 따른 미적1학점 분포", x="학과", y="미적1학점") +
  theme_minimal() +
  facet_wrap(~학과, scales="free")
# 학과에 따른 미적2학점
ggplot(Data, aes(x=학과, y=미적2학점_numeric)) +
  geom_violin(trim=FALSE) +
  geom_boxplot(width=0.1, fill="skyblue", color="black") +
  labs(title="학과에 따른 미적2학점 분포", x="학과", y="미적2학점") +
```

```

    theme_minimal() +
    facet_wrap(~학과, scales="free")
#입학시기에 따른 미적1학점
ggplot(Data, aes(x=입학시기, y=미적1학점_numeric, fill=입학시기)) +
  geom_boxplot() +
  scale_fill_manual(values=c("purple", "orange", "green")) +
  labs(title="입학시기에 따른 미적1학점 분포", x="입학시기", y="미적1학점")
+
  theme_minimal()
#입학시기에 따른 미적2학점
ggplot(Data, aes(x=입학시기, y=미적2학점_numeric, fill=입학시기)) +
  geom_boxplot() +
  scale_fill_manual(values=c("purple", "orange", "green")) +
  labs(title="입학시기에 따른 미적2학점 분포", x="입학시기", y="미적2학점")
+
  theme_minimal()
#출신고지역에 따른 미적1학점
ggplot(Data, aes(x=출신고지역, y=미적1학점_numeric, fill=출신고지역)) +
  geom_boxplot() +
  scale_fill_manual(values=c("red", "orange", "yellow", "green", "skyblue",
"blue")) +
  labs(title="출신고지역에 따른 미적1학점 분포", x="출신고지역", y="미적1학
점") +
  theme_minimal()
#출신고지역에 따른 미적2학점
ggplot(Data, aes(x=출신고지역, y=미적2학점_numeric, fill=출신고지역)) +
  geom_boxplot() +
  scale_fill_manual(values=c("red", "orange", "yellow", "green", "skyblue",
"blue")) +
  labs(title="출신고지역에 따른 미적2학점 분포", x="출신고지역", y="미적2학
점") +
  theme_minimal()
#미적1학점과 수학실력시험 성적 상관관계
ggplot(Data, aes(x=미적1학점_numeric, y=수학실력시험)) +

```

```

geom_point() +
geom_smooth(method="lm", se=FALSE, color="blue") +
labs(title="미적1학점과 수학실력시험 간의 상관관계", x="미적1학점", y="수
학실력시험") +
theme_minimal()
#미적2학점과 수학실력 시험 성적 관계
ggplot(Data, aes(x=미적2학점_numeric, y=수학실력시험)) +
geom_point() +
geom_smooth(method="lm", se=FALSE, color="blue") +
labs(title="미적2학점과 수학실력시험 간의 상관관계", x="미적2학점", y="수
학실력시험") +
theme_minimal()
#미적1학점과 2학점 상관관계
ggplot(Data, aes(x=미적1학점_numeric, y=미적2학점_numeric)) +
geom_point() +
geom_smooth(method="lm", se=FALSE, color="pink") +
labs(title="미적1학점과 미적2학점점 간의 상관관계", x="미적1학점", y="미
적 2학점") +
theme_minimal()

```