

기초통계

Park Beomjin¹

¹*University of Seoul*

1 다섯수치 요약(five number summary) + 평균

- 평균과 다섯수치 요약 값을 산출하기

```
proc means data = package noprint;
var pack;
output out = fivenums min = min p25 = q25 mean = mean median = median p75 = q75 n;
proc print data = fivenums;
run;
```

2 교차표

- 범주형 변수를 얻었을 때 교차표 만들기

```
data table1;
input age type count@@;
cards;
1 1 38 1 2 79
2 1 87 2 2 118
3 1 78 3 2 89
;
proc freq order = data;
weight count;
tables age * type /chisq expected nocol norow nopercnt;
run;

data table2;
input age type @@;
cards;
1 1 1 2
1 1 1 1
```

```

2 1 2 1
2 2 2 2
3 1 3 1
3 1 3 1
;
proc freq order = data;
tables age * type / chisq expected nocol norow nopercent;
run;

```

3 상관관계

- 두 개 혹은 그 이상의 변수들의 관계를 알아보기 위한 방법을 알아보자.
- 변수들간의 관계를 알아볼 수 있는 그래프로는 산점도가 있다.
- 또한 변수들간의 관계(선형관계)를 수치적으로 나타낸 값은 상관계수, 검정으로는 상관관계검정이 있다.
- 산점도 그리기

```

proc sgplot data = iris1;
scatter X = sepal_length y = sepal_width;
run;

```

```

proc sgplot data = iris1;
by species;
scatter X = sepal_length y = sepal_width;
run;

```

```

proc sgscatter data = iris1;
by species;
matrix sepal_length -- petal_width;
run;

```

- 상관계수 구하기

```

proc corr data = iris1;
var sepal_length sepal_width;

```

```
run;
```

```
proc corr data = iris1;  
var sepal_length -- petal_width;  
run;
```

- 상관관계 검정하기

```
proc corr data = iris1 FISHER(rho0 = 0);  
var sepal_length -- petal_width;  
run;
```

4 히스토그램

- 변수의 분포를 추정할 수 있고 시각적으로 나타내주는 히스토그램을 그려보자.

```
proc univariate data = iris noprint;  
var sepal_length sepal_width;  
histogram;  
run;
```