

## Homework 2

류근관

2020년 가을학기 경제통계학

### Problem 1

한 프로야구 팀의 감독이 9회말 2아웃 만루의 동점상황에서 대타 선수를 기용하기 위해 벤치에 남아 있는 두 선수에 대한 자료를 살펴보았다.

	선수 A		선수 B	
	선발	대타	선발	대타
시즌 타석수	150	150	200	100
시즌 안타수	40	50	44	40
시즌 타율	0.300		0.280	

자료를 본 감독은 시즌타율이 좋은 선수 A를 대타로 기용하기로 결정하였다. 위와 같은 감독의 결정을 통계적 개념을 이용하여 평가하라.

### Problem 2

아래 표는 한 주당 가격이 각각 \$20, \$10인 A회사와 B회사 주식의 과거 수익률에 대한 통계치들이다. A회사 주식 100주와 B회사 주식 200주로 포트폴리오를 구성하는 경우 이러한 포트폴리오 수익률의 표준편차를 계산하시오.

	평균	표준편차	공분산
A회사주식	9%	20%	-200 (%) <sup>2</sup>
B회사주식	9%	20%	

### Problem 3

두 확률 변수 X와 Y간에 다음과 같은 선형관계가 성립한다고 가정하자.

$$Y = a + bX + u$$

위 식에서 a와 b는 상수이며 u는 평균이 0이고 분산이  $\sigma_u^2$ 인 확률변수이며 확률변수 X의 평균과 분산은 각각  $\mu_x$ ,  $\sigma_x^2$ 이다. 확률변수 X와 u는 확률적으로 독립임을 가정한다. 다음의 질문에 답하시오.

- (a) 확률변수  $X$ 가 주어진 경우 확률변수  $Y$ 의 조건부 기대값을 구하시오.
- (b) 확률변수  $Y$ 의 기대값을 구하시오.
- (c) 확률변수  $X$ 가 주어진 경우 확률변수  $Y$ 의 조건부분산을 구하시오.
- (d) 확률변수  $Y$ 의 분산을 구하시오.

#### Problem 4

다음과 같은 회귀분석 모형을 생각해보자.

$$y_i = a + bD_i + u_i, \quad i = 1, \dots, n$$

여기서 종속변수(피설명변수)  $y_i$ 는 여론조사에 참여한 유권자 중 다음 대선에서 보수당 후보에 투표하겠다고 답했으면 1의 값을 가지며 그렇지 않으면 0의 값을 가지는 더미변수이다. 독립변수(설명변수)  $D_i$ 는 유권자의 월평균 소득이 상위 50%에 속하면 1, 그렇지 않으면 0의 값을 취하는 더미변수이다. 오차항  $u_i$ 는 소득 이외에 유권자의 투표 성향에 영향을 미칠 수 있는 다른 요인들을 나타낸다. 여기서  $D_i$ 와  $u_i$ 는 확률적으로 독립이라고 가정하자.

아래 질문에 답하시오.

- (a) 위 회귀분석 모형에서 절편  $a$ 는 어떠한 의미를 가지는가? 말로 설명하시오.
- (b) 위 회귀분석 모형에서 기울기  $b$ 는 어떠한 의미를 가지는가? 말로 설명하시오.
- (c) 다음 대선에서 보수당 후보에 대한 유권자 지지율을  $p$ 라고 하자. 위의 회귀분석 모형을 토대로 전체 유권자 중 보수당 후보에 대한 지지율  $p$ 를  $a$ 와  $b$ 의 함수로 표현해보라. 단, 유권자 중 절반은 소득이 상위 50%에 속하고 나머지 절반은 소득이 하위 50%에 속한다고 가정하자.

#### Problem 5

월소비(원 단위)를 종속변수로 하고 월소득(원 단위)을 독립변수로 하는 회귀분석 모형을 추정된 결과 다음과 같은 결과를 얻었다고 하자.

$$\hat{y}_i = 300,000 + 0.3x_i, \quad i = 1, \dots, n$$

아래 질문에 답하시오.

(a) 위 회귀분석 모형에서 설명변수의 단위를 원 단위가 아닌 만 원 단위로 바꾸어 측정한 뒤 동일한 회귀분석을 시행할 경우 절편과 기울기에 대한 추정치가 어떻게 바뀌겠는가? 말로 설명하시오. (단, 종속변수의 단위는 변화시키지 않는다.)

(b) 위 회귀분석 모형에서 종속변수의 단위를 원 단위가 아닌 만 원 단위로 바꾸어서 동일한 회귀분석을 시행할 경우 절편과 기울기에 대한 추정치가 어떻게 바뀌겠는가? 말로 설명하시오. (단, 설명변수의 단위는 변화시키지 않는다.)

(c) 위 회귀분석 모형에서 설명변수와 종속변수의 단위를 모두 원 단위가 아닌 만 원 단위로 바꾸어서 동일한 회귀분석을 시행할 경우 절편과 기울기에 대한 추정치가 어떻게 바뀌겠는가? 말로 설명하시오.

### Problem 6

Ezstat.snu.ac.kr(→ statistical textbook → useful data set)에서 서울 강남구 아파트가격 자료를 다운로드 하여 다음 문제에 답하라.  
(데이터파일은 csv파일로 저장. **Python**을 이용하여 다음 문제를 답하시오.)

(a) 아파트 연령, 평 수 및 시세(중간값) 정보를 이용하여 아래 두 변수 간 결합분포표를 완성하라. (평당 가격은 시세(중간값)을 평 수로 나누어 구한다.)

연령 \ 평당 가격	400~850	850~1300	1300~1875
0~9			
10~18			
19~27			

(b) 아파트 연령을 x축, 평당 가격을 y축으로 하는, 회귀직선을 포함한 산포도를 그려보라. 그리고 두 변수 간 상관계수를 구하라. 산포도와 상관계수 값이 예상과 다르게 나온 경우 왜 그런 결과가 나올 수 있었는지 생각해보고 그 이유를 설명하라.

(Python에서 그린 그래프를 첨부하라.)

(c) 아파트 연령과 단지규모를 이용하여 아래의 두 변수 간 결합분포표를 완성하라.

아파트를 다음 단지규모별 기준에 따라 9개 범주로 분류하라.

범주9: 단지규모가 2900세대보다 큰 아파트

범주8: 단지규모가 2100세대보다 크고 2900세대보다 작거나 같은 아파트

범주7: 단지규모가 1400세대보다 크고 2100세대보다 작거나 같은 아파트

범주6: 단지규모가 1200세대보다 크고 1400세대보다 작거나 같은 아파트

범주5: 단지규모가 1100세대보다 크고 1200세대보다 작거나 같은 아파트

범주4: 단지규모가 900세대보다 크거나 같고 1100세대보다 작거나 같은

아파트

범주3: 단지규모가 800세대보다 크고 900세대보다 작은 아파트

범주2: 단지규모가 500세대보다 크고 800세대보다 작거나 같은 아파트

범주1: 단지규모가 500세대보다 작거나 같은 아파트

연령 \ 단지규모	0~9	10~18	19~27
범주1			
범주2			
범주3			
범주4			
범주5			
범주6			
범주7			
범주8			
범주9			

(d) 아파트 연령을 x축, 단지 규모를 y축으로 하는, 회귀직선을 포함한 산포도를 그

려보라. 그리고 두 변수 간 상관계수를 구하라. 산포도와 상관계수 값을 보고 그 값의 부호가 갖는 의미를 해석해보라.