통조입 HW4

2019150445 통계학과 신백록

/\* 1.a),b),c) \*/

**DATA** cable;

INFILE 'C:\Stephens.dat';

INPUT district house houseval numold numyoung numtvs hiprice

hourstot hoursnew hoursprt hourskid hoursmov;

**RUN**;

**PROC** **PRINT** DATA=cable;

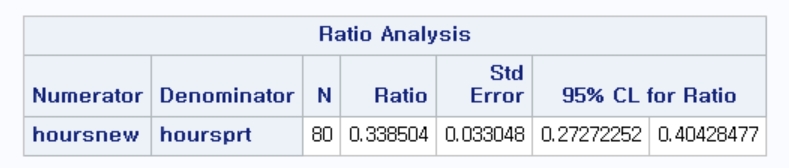
**RUN**;

**PROC** **SURVEYMEANS** DATA=cable;

VAR hoursnew hoursprt;

RATIO hoursnew/hoursprt;

TITLE 'Estimate of R for hoursnew over hoursprt';

**RUN**;

R\_hat=0.338504

se of R\_hat=0.033048

95%CI=(0.27272252,0.40428477)

1.d)

0.338504 hours spent on watching news/public affairs relative to hours spent on watching sports.

2.a)

/\*2.a)\*/

**PROC** **GPLOT** DATA=cable;

PLOT hiprice\*houseval;

SYMBOL V=circle;

TITLE 'Plot of hiprice versus houseval'

테이블이(가) 표시된 사진

자동 생성된 설명RUN;

2.b)

/\*2.b)\*/

**PROC** **CORR** DATA=cable;

VAR hiprice houseval;

TITLE 'correlation between hiprice and houseval';

테이블이(가) 표시된 사진

자동 생성된 설명**RUN**;

Cor=0.65491

2.c)

Intercept가 0에 가깝고 k번째 분산이 k가 커짐에 따라 증가하기에, 또한 유의미한 선형적인 양의상관관계가 있기에 Ratio estimator을 사용하는 것이 sensible하다.

2.d)

/\*2.d)\*/

**PROC** **SURVEYMEANS** DATA=cable;

VAR hiprice houseval;

RATIO hiprice/houseval;

TITLE 'Ratio estimation of hiprice over houseval';

**RUN**;

**DATA** ratio\_est;

INPUT r meanx;

ratio\_mean=meanx\*r;

CARDS;

0.000139 68045

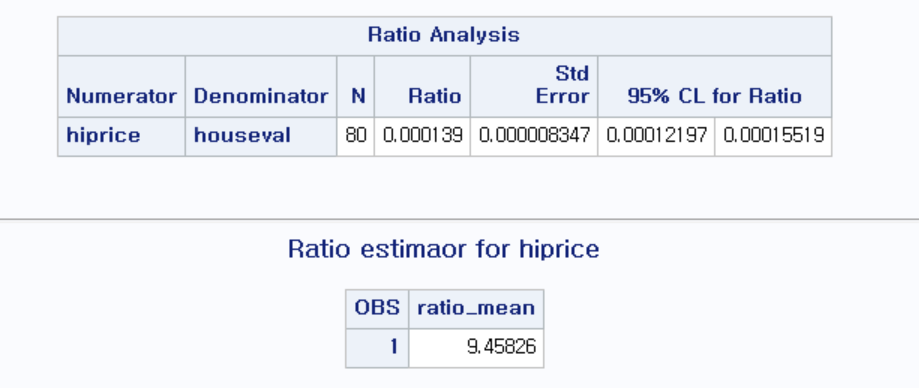
;

**RUN**;

**PROC** **PRINT** DATA=ratio\_est;

TITLE 'Ratio estimaor for hiprice';

VAR ratio\_mean;

**RUN**;

2.e)

/\*2.e)\*/

**PROC** **SURVEYMEANS** DATA=cable TOTAL=**31989**;

VAR hiprice houseval;

RATIO hiprice/houseval;

TITLE 'Ratio estimation of hiprice over houseval considering f';

**RUN**;

**DATA** ratio\_est\_se;

INPUT meanx se;

se\_ratio\_mean=meanx\*se;

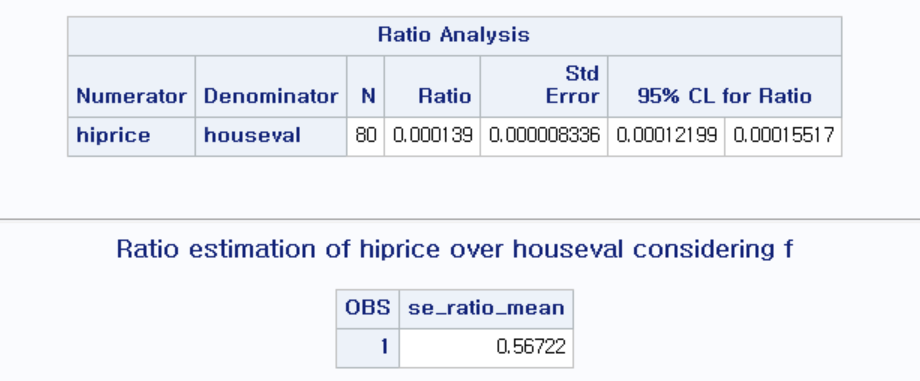
cards;

68045 0.000008336

**RUN**;

**PROC** **PRINT** DATA=ratio\_est\_se;

VAR se\_ratio\_mean;

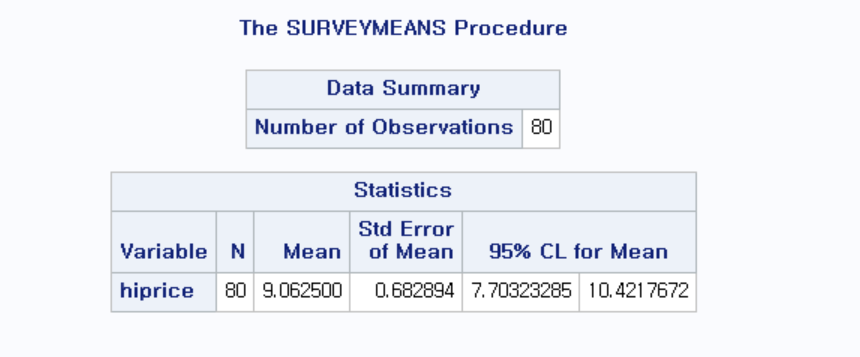
**RUN**;

2.f)

/\*2.f)\*/

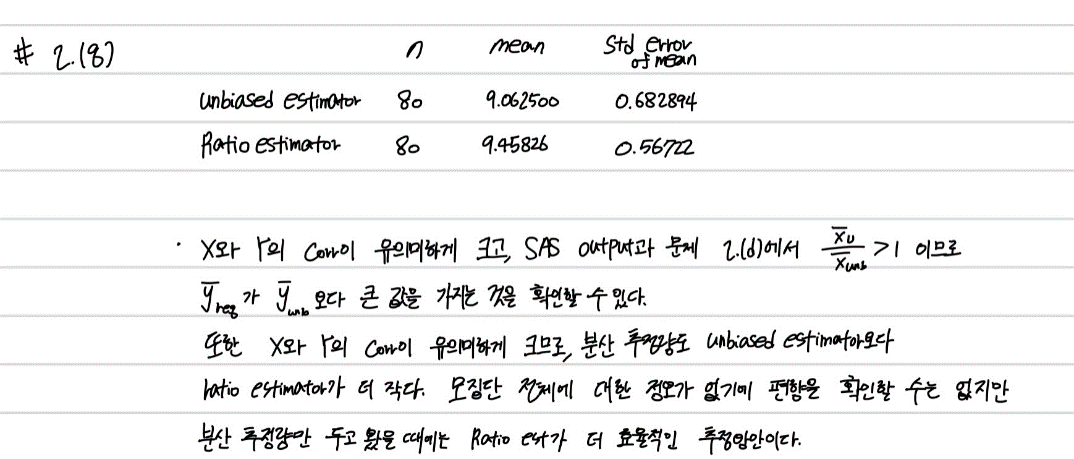
**PROC** **SURVEYMEANS** DATA=cable TOTAL=**31989**;

VAR hiprice;

**RUN**;

Sample mean=9.062500

Se of sample mean=0.682894

2.g)

3.a)

/\*3.a)\*/

**DATA** golf;

INFILE 'c:\golfsrs.dat';

INPUT rn state $ course $ holes type $ yearblt wkday18

wkday9 wkend18 wkend9 backtee rating par cart18

cart9 caddy $ pro $ ;

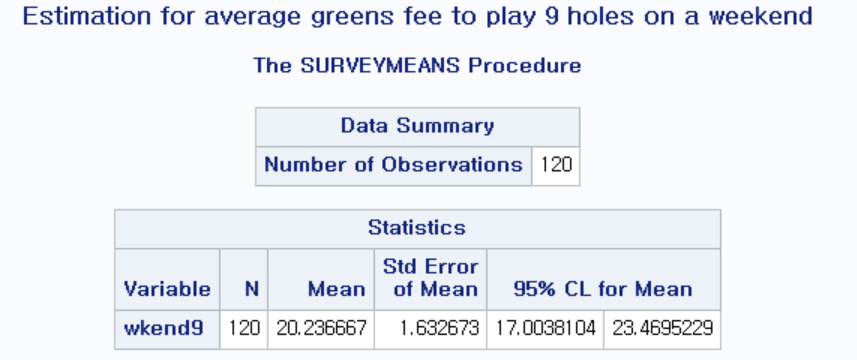
**Run**;

**PROC** **SURVEYMEANS** DATA=golf TOTAL=**16883**;

VAR wkend9;

TITLE 'Estimation for average greens fee to play 9 holes on a weekend';

**RUN**;



Sample mean=20.236667

3.b)

/\*3.b)\*/

**PROC** **GPLOT**;

PLOT wkend9\*backtee;

SYMBOL V=circle;

**RUN**;

**PROC** **CORR**;

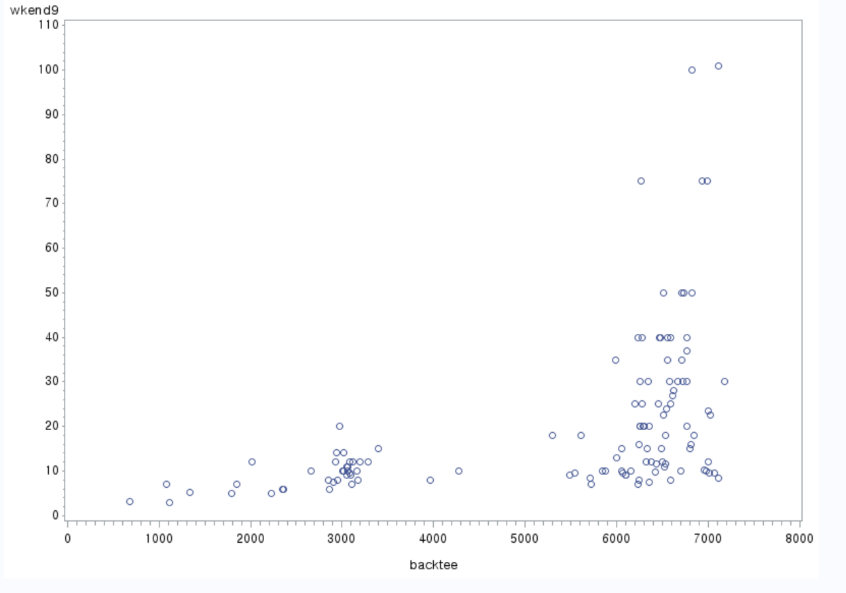
VAR wkend9 backtee;

**RUN**;

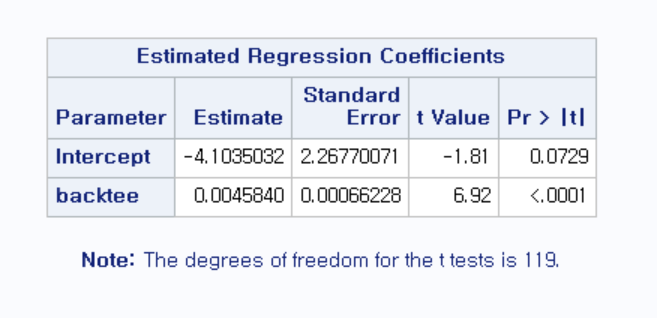
**PROC** **SURVEYREG** data=golf TOTAL=**16883**;

MODEL wkend9=backtee;

**RUN**;



cor=0.46217



The estimated regression equation is Wkend9=-4.1035032+0.0045840\*backtee

3.c)

보조정보에 대한 모평균, 즉 backtee의 모평균에 대한 정보가 없다. 따라서 wkend9의 mean regression estimate를 추정하기 어렵고, CI를 만들 수 없다.

3.d)

문제 3.b)에서의 plot과 0.46217의 correlation coefficient로 wkend9과 backtee는 양의 상관관계가 있는 것을 알 수 있다. 따라서 sample mean보다 regression estimator의 분산 추정량이 더 작을 것이고, 편향에 대한 정보가 없긴 하지만 n이 커질수록 편향이 줄어들어 regression estimator가 wkend9을 더 잘 추정할 것이다.

4.a)

/\*4.a)\*/

**data** cherry;

infile 'C:\cherry.txt' firstobs=**2**;

input diam height vol;

sampwt = **2967**/**31**;

obsnum = \_n\_;

**RUN**;

**proc** **surveymeans** data=cherry total=**2967**;

var diam vol;

ratio vol/diam;

**run**;

**data** ratio\_est\_cherry;

input r se totalx n;

ratio\_mean=totalx/n\*r;

ratio\_total=n\*ratio\_mean;

se\_ratio\_mean=totalx/n\*se;

se\_ratio\_total=n\*se\_ratio\_mean;

cards;

2.277331 0.130786 41385 2967

;

**proc** **print** data=ratio\_est\_cherry;

var ratio\_total se\_ratio\_total;

**run**;

**data** CI;

set ratio\_est\_cherry;

lower=ratio\_total-**1.96**\*se\_ratio\_total;

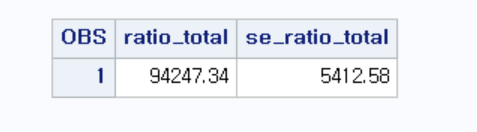
upper=ratio\_total+**1.96**\*se\_ratio\_total;

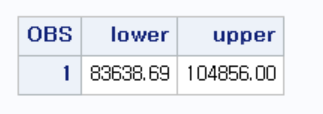
**run**;

**proc** **print** data=CI;

var lower upper;

**run**;





4.b)

/\*4.b\*/

**proc** **surveyreg** data=cherry total=**2967**;

model vol=diam;

estimate intercept **2967** diam **41385**;

**run**;

**data** CI\_reg;

input est se;

lower=est-**1.96**\*se;

upper=est+**1.96**\*se;

cards;

100039 2587.14

;

**run**;

**proc** **print** data=CI\_reg;

var lower est upper;

**run**;

