통계조사론입문 HW3

2019150445 통계학과 신백록

텍스트이(가) 표시된 사진

자동 생성된 설명텍스트이(가) 표시된 사진

자동 생성된 설명테이블이(가) 표시된 사진

자동 생성된 설명

**2.**

2.1.

filename baseball "C:\baseball.csv";

**data** baseball;

infile baseball delimiter="," ;

input team $ leanguID $ player $ salary POS $ G GS InnOuts

PO A E DP PB GB AB R H SecB ThiB HR RBI SB CS BB

SO IBB HBP SH SF GIDP;

**data** baseball\_log; set baseball;

salary\_log=log(salary);

**RUN**;

**PROC** **FREQ** DATA=baseball\_log;

TABLES team ;

**RUN**;

**PROC** **SORT** DATA=baseball\_log ;

BY team;

**RUN**;

**data** a;

INPUT team $ \_TOTAL\_;

DATALINES;

ANA 26

ARI 28

ATL 28

BAL 25

BOS 27

CHA 26

CHN 29

CIN 27

CLE 28

COL 27

DET 26

FLO 26

HOU 25

KCA 27

LAN 24

MIL 25

MIN 25

MON 28

NYA 29

NYN 26

OAK 27

PHI 25

PIT 27

SDN 26

SEA 27

SFN 28

SLN 26

TBA 26

TEX 27

TOR 26

;

**RUN**;

**PROC** **SURVEYSELECT** SEED=**20201029** METHOD=SRS DATA=baseball\_log n=**150** OUT=baseball\_out;

STRATA team/ALLOC=prop;

**RUN**;

**PROC** **PRINT** DATA=baseball\_out;

**run**;

테이블이(가) 표시된 사진

자동 생성된 설명

테이블이(가) 표시된 사진

자동 생성된 설명

2.b.

**PROC** **SURVEYMEANS** MEAN CLM DATA=baseball\_out TOTAL=a;

VAR salary\_log;

WEIGHT samplingweight;

STRATA team;

테이블이(가) 표시된 사진

자동 생성된 설명**RUN**;

2.c.

**PROC** **SURVEYMEANS** MEAN CLM DATA=baseball\_out TOTAL=a;

VAR POS;

WEIGHT samplingweight;

STRATA team;

**RUN**;

테이블이(가) 표시된 사진

자동 생성된 설명

2.d.

**PROC** **SURVEYMEANS** MEAN CLM DATA=baseball\_out TOTAL=a;

VAR salary\_log;

WEIGHT samplingweight;

STRATA team;

BY team;

**RUN**;

각각의 sample variance=

ANA 0.595311^2

ARI 0.550919^2

ATL 0.698669^2

BAL 0.404605^2

BOS 0.197592^2

CHA 0.440330^2

CHN 0.626007^2

CIN 0.546414^2

CLE 0.162379^2

COL 0.581618^2

DET 0.386785^2

FLO 0.559405^2

HOU 0.571294^2

KCA 0.248324^2

LAN 0.431594^2

MIL 0.418652^2

MIN 0.604636^2

MON 0.350823^2

NYA 0.742482^2

NYN 0.304865^2

OAK 0.660898^2

PHI 0.397085^2

PIT 0.496472^2

SDN 0.479151^2

SEA 0.511487^2

SFN 0.383678^2

SLN 0.290694^2

TBA 0.350261^2

TEX 0.538986^2

TOR 0.396641^2

Sample variance가 각각의 층별로 차이가 꽤 나지만 cost의 정보가 주어져있지 않고 각 층의 N도 비슷하기 때문에 굳이 optimal alloc으로 표본을 추출하기보다 NEYMAN alloc을 사용하는 게 나아보인다.

2.e

**data** B;

INPUT team $ s;

CARDS;

ANA 0.595311

ARI 0.550919

ATL 0.698669

BAL 0.404605

BOS 0.197592

CHA 0.440330

CHN 0.626007

CIN 0.546414

CLE 0.162379

COL 0.581618

DET 0.386785

FLO 0.559405

HOU 0.571294

KCA 0.248324

LAN 0.431594

MIL 0.418652

MIN 0.604636

MON 0.350823

NYA 0.742482

NYN 0.304865

OAK 0.660898

PHI 0.397085

PIT 0.496472

SDN 0.479151

SEA 0.511487

SFN 0.383678

SLN 0.290694

TBA 0.350261

TEX 0.538986

TOR 0.396641

;

**RUN**;

**data** variance;

SET b;

\_VAR\_=s\*\***2**;

**RUN**;

**PROC** **SURVEYSELECT** SEED=**20201029** METHOD=SRS DATA=baseball\_log SAMPSIZE=**150** OUT=baseball\_NEYMAN;

STRATA team/ALLOC=NEYMAN VAR=variance;

**RUN**;

**PROC** **PRINT** DATA=baseball\_NEYMAN;

**RUN**;

테이블이(가) 표시된 사진

자동 생성된 설명테이블이(가) 표시된 사진

자동 생성된 설명