

1. Consider the baseball dataset describing the population of baseball players in data file baseball.csv.
 - (a) Take a stratified random sample of 150 players from the file, using proportional allocation with the different teams as strata. Describe how you selected the sample.
 - (b) Find the mean of the variable $\text{logsal} = \ln(\text{salary})$, using your stratified sample, and give a 95% CI.
 - (c) Estimate the proportion of players in the data set who are pitchers, and give a 95% CI.
 - (d) Examine the sample variances of logsal in each stratum. Do you think optimal allocation (strata with same costs) would be worthwhile for this problem?
2. The data set agsrs.csv contains information on the number of farms in 1987 for the SRS of $n = 300$ counties from the population of the $N = 3078$ counties in the United States. In 1987, the United States had a total of 2,087,759 farms.
 - (a) Use ratio estimation to estimate the total number of acres devoted to farming in 1992, using the number of farms in 1987 as the auxiliary variable.
 - (b) Repeat (a), using regression estimation.
 - (c) Which method gives the most precision: ratio estimation with auxiliary variable acres87 , ratio estimation with auxiliary variable farms87 , or regression estimation with auxiliary variable farms87 ? Why?
 - (d) Estimate the total number of acres devoted to farming for each of two domains: (i) counties with fewer than 600 farms, and (ii) counties with 600 or more farms. Give standard errors for your estimates.

agsrs.dat Data from an SRS of size 300 from the U.S. 1992 Census of Agriculture. In columns 3-14, the value “-99” denotes missing data.

Column	Name	Value
1	county	county name
2	state	state abbreviation
3	acres92	number of acres devoted to farms, 1992
4	acres87	number of acres devoted to farms, 1987
5	acres82	number of acres devoted to farms, 1982
6	farms92	number of farms, 1992
7	farms87	number of farms, 1987
8	farms82	number of farms, 1982
9	largef92	number of farms with 1,000 acres or more, 1992
10	largef87	number of farms with 1,000 acres or more, 1987
11	largef82	number of farms with 1,000 acres or more, 1982
12	smallf92	number of farms with 9 acres or fewer, 1992
13	smallf87	number of farms with 9 acres or fewer, 1987
14	smallf82	number of farms with 9 acres or fewer, 1982
15	region	region of country (W = West, NC = North Central, S = South, N = Northeast)

baseball.dat Statistics on 797 baseball players, compiled by Jenifer Boshes from the rosters of all major league teams in November, 2004. Source: Forman, S. L. (2004). *Baseball-reference.com—Major league statistics and information*. Retrieved November 2004 from www.baseball-reference.com.

Column	Name	Value
1	team	team played for at beginning of the season
2	leagueID	AL or NL
3	player	a unique identifier for each baseball player
4	salary	player salary in 2004
5	POS	primary position coded as P, C, 1B, 2B, 3B, SS, RF, LF, or CF
6	G	games played
7	GS	games started
8	InnOuts	number of innings
9	PO	Put Outs
10	A	number of assists
11	E	Errors
12	DP	number of double plays
13	PB	number of passed balls (only applies to catchers)
14	GB	number of games that player appeared at bat
15	AB	number of at bats
16	R	number of runs scored
17	H	number of hits
18	SecB	number of doubles
19	ThiB	number of triples
20	HR	number of home runs
21	RBI	number of runs batted in
22	SB	number of stolen bases
23	CS	number of times caught stealing
24	BB	number of times walked
25	SO	number of strikeouts
26	IBB	number of times intentionally walked
27	HBP	number of times hit by pitch
28	SH	number of sacrifice hits
29	SF	number of sacrifice flies
30	GIDP	grounded into double play