SPRING 1997

Statistics 200B
Lab 1: Welcome to Fort Evans

Fort Evans is a small city with a population of about 150,000 located in Silica County, California. The city is divided into six districts: Northeast, Southeast, North Central, South Central, Northwest, and Southwest. For this and some following labs you will conduct sample surveys to learn something about the demographics of the city. You might imagine that you are the city Census Director and that you don't have enough money in your budget to conduct a complete enumeration and description of the population, so you must use sampling to obtain information.

This week's lab is a warm up, or pilot study, in which you will familiarize yourself with the computer and take a quick peek at Fort Evans. Take a simple random sample of size 100 of the inhabitants, without replacement. For this sample:

1. How many people in your sample are in each of the six geographic regions?

2. Find the proportion of adults (people over 18) who are unemployed and looking for work.

3. Estimate the average income of a family in Fort Evans. Be careful here, noting that you are sampling individuals and not families.

You can look at other variables if you wish to do so. Compare your answers to the questions above to those of another class member. Do you have any sense of how much error there is in the average you computed in answer to (3) above as an estimate of the population average in Fort Evans? Do you have any idea how reliably you have estimated the proportion of unemployed adults looking for work? Write up the lab as a brief report explaining the problem, what you did, and what you concluded. Computer printout, including graphics, can be attached as appendices to your report. A report which consists of annotated computer printout is not acceptable. Remember that another human being has to read and understand your report.

The Fort Evans dataset you will be working with contains data on 155,197 persons for 11 variables:

Region District of residence, coded as follows:

1. Northeast
2. Southeast
3. North Central
4. South Central
5. Northwest
6. Southwest

Sex codes:

1. Male
2. Female

Age codes:

0-89 Age in years

90 90 years of age or older

Race codes:

1 White

2 Black

3 American Indian, Eskimo, or Aleut

4 Asian or Pacific Islander

5 Other

Ethnicity codes:

0 Not available

1 Mexican American

2 Chicano

3 Mexican (Mexicano)

4 Puerto Rican

5 Cuban

6 Central or South American

7 Other Spanish

8 All other

9 Don't know

Edlevel Educational level, coded as follows:

0 Children (age less than 15)

31 Less than 1st grade

32 1st, 2nd, 3rd, or 4th grade

33 5th or 6th grade

34 7th or 8th grade

35 9th grade

36 10th grade

37 11th grade

38 12th grade

39 High school graduate, high school diploma or equivalent

40 Some college but no degree

41 Associate degree in college – occupation/vocation program

42 Associates degree in college – academic program

43 Bachelors degree (e.g., BS, BA, AB)

44 Masters degree (e.g., MS, MA, MBA)

45 Professional school degree (e.g., MD, DDS, DVM, LLB, JD)

46 Doctoral degree (e.g., PhD, EdD)

Marital Marital status, coded as follows:

1. Married, civilian spouse present
2. Married, armed forces spouse present
3. Married, spouse absent (not sep.)

4. Widowed
5. Divorced
6. Separated
7. Never Married

Labstat Labor force status, coded as follows:

- **0** Children or Armed Forces
- **1** Working
- **2** With a job, not at work
- **3** Unemployed, looking for work
- **4** Unemployed, on layoff
- **5** Not in labor force
  - working without pay less than or equal to 15 hours per week
  - temporarily absent from without pay job
- **6** Not in labor force – unavailable
- **7** Other not in labor force

Fincome Family income in 1992. Codes: Dollar amount of income (positive or negative)

Numkids Number of own, unmarried children under the age of 18. Codes:

- **0–8** Number of children
- **9** 9 or more children

Fampers Number of persons in family. For primary families, this includes all persons in related subfamilies and related individuals living with the primary family. For an unrelated individual, fampers=1.

The Teaching Assistant will provide some information to help you get started with Splus. You can sample from the dataset using the command `sampler`. The Splus object that stores the returned sample is a matrix with number of rows equal to the sample size and eleven columns whose dimension names are region, sex, age, etc. Note that the function `sampler` is "homemade" and will not appear in the Splus Guide.