Chapter 2: Collecting Data Using Surveys and Scientific Studies

2.1 The relative merits of the different types of sampling units depends on the availability of a sampling frame for individuals, the desired precision of the estimates from the sample to the population, and the budgetary and time constraints of the project.

2.2 She could conduct a stratified random sample in which the plants serve as the stratum. A simple random sample could then be selected within each plant. This would provide information concerning the differences between the plants along with the individual opinions of the employees.

2.3 The list of registered voters in the state could be used as the sampling frame for selecting the persons to be included in the sample.

2.4 a. No. The survey in which the interviewer showed the peanut butter should be the more accurate because it does not rely on the respondent's memory of which brand was purchased.

b. Both surveys may have survey nonresponse bias because an entire segment of the population (those not at home) cannot be contacted. Also, both surveys may have interviewer bias resulting from the way the question is posed (e.g., tone of voice). In the first survey, results may be biased by the respondent's ability to recall correctly which brand was purchased. The second survey may be biased by the respondent's unwillingness to show the interviewer the peanut butter jar (too intrusive), or by the respondent not recognizing that the peanut butter that had purchased was *low fat*.

2.5 a. Alumni (men only?) graduating from Yale in 1924.

b. No. Alumni whose addresses were on file 25 years later would not necessarily be representative of their class.

c. Alumni who responded to the mail survey would not necessarily be representative of those who were sent the questionaires. Income figures may not be reported accurately (intentionally), or may be rounded off to the nearest \$5,000, say, in a self-administered questionaire.

d. Rounding income responses would make the figure \$25,111 highly unlikely. The fact that higher income respondents would be more likely to respond (bragging), and the fact that incomes are likely to be exaggerated, would tend to make the estimate too high.

2.6 a. Water Temperature and Type of Hardener

b. Water temperature: $175°F$ and $200°F$
Type of Hardener: H_1, H_2, H_3

c. Manufacuring Plants

d. Plastic Pipe

e. Location on Plastic Pipe

f. 2 Pipes per Treatment

g. 6 Treatments: $(175°F, H_1)$, $(175°F, H_2)$, $(175°F, H_3)$
$(200°F, H_1)$, $(200°F, H_2)$, $(200°F, H_3)$

2.7 a.
- Factors: Location in Orchard, Location on Tree, Time of Year
- Factor Levels: Location in Orchard: 8 Sections
  Time of Year: Oct., Nov., Dec., Jan., Feb., March, April, May
  Location on Tree: Top, Middle, Bottom
- Blocks: None
- Experimental Units: Locaton on Tree during one of the 8 months
- Measurement Units: Oranges
- Replications: For each section, time of year, location on tree, there is one experimental unit. Hence 1 rep.
- Treatments: 192 combinations of 8-Sections, 8 Months, 3 Locations on Tree: (S_i, M_j, L_k), for $i = 1, \cdots, 8; j = 1, \ldots, 8; k = 1, \ldots, 3$

b.
- Factors: Type of Treatment
- Factor Levels: T_1, T_2
- Blocks: Hospitals
- Experimental Units: Wards
- Measurement Units: Patients
- Replications: 2 Wards per Treatment in each of the 8 Hospitals
- Treatments: T_1, T_2

c.
- Factors: Type of Treatment
- Factor Levels: T_1, T_2
- Blocks: Hospitals, Wards
- Experimental Units: Patients
- Measurement Units: Patients
- Replications: 2 Patients per Treatment in each of the Ward/Hospital combinations
- Treatments: T_1, T_2

d.
- Factors: Type of School
- Factor Levels: Public, Private-nonparochial, Parochial
- Blocks: Geographical region
- Experimental Units: Classroom
- Measurement Units: Students in Classroom
- Replications: 2 Classrooms per Each Type of School in each of the City/Region combinations
- Treatments: Public, Private-nonparochial, Parochial

2.8 a. Factors: Temperature, Type of Seafood

b. Factor Levels: Temperature$(0°C, 5°C, 10°C)$, Type of Seafood(Oysters, Mussels)

c. Block: none

d. Experimental Unit: Package of Seafood

e. Measurement Unit: Sample from package

f. Replications: 3 packages per temperature

h. Treatments: $(0°C, Oysters), (5°C, Oysters), (10°C, Oysters)$
$(0°C, Mussels), (5°C, Mussels), (10°C, Mussels)$

2.9 a. Randomized block design with blocking variable(5 Farms) and 48 treatments in a 3x4x4 factorial structure.

b. Completely randomized design with 10 treatments (Software Programs) and 3 reps of each treatment.

c. Latin square design with blocking variables (Position in Kiln, Day) each having 8 levels. The treatment structure is a 2x4 factorial structure (Type of Glaze, Thickness).

2.10 a. Design B. The experimental units are not homogeneous since one group of consumers gives uniformly low scores and another group gives uniformly high scores, no matter what recipe is used. Using Design A, it is possible to have a group of consumers which give mostly low scores randomly assigned to a particular recipe. This would bias this particular recipe. Using Design B, the experimental error would be reduced since each consumer would evaluate each recipe. That is, each consumer is a block and each of the treatments (recipes) are observed in each block. This results in having each recipe subject to consumers who give low scores and to consumers who give high scores.

b. This would not be a problem for either design. In design A, each of the remaining four recipes would still be observed by 20 consumers. In design B, each consumer would still evaluate each of the four remaining recipes.