5.5 Perspective

In this chapter we have presented the concept of a sampling distribution and have focused on the sampling distribution of $\overline{Y}$. Of course, there are many other important sampling distributions, such as the sampling distribution of the sample standard deviation and the sampling distribution of the sample median.

Let us take another look at the random sampling model in the light of Chapter 5. As we have seen, a *random* sample is not necessarily a *representative* sample.* But using sampling distributions, one can specify the degree of representativeness to be expected in a random sample. For instance, it is intuitively plausible that a larger sample is likely to be more representative than a smaller sample from the same population. In Section 5.1 and Section 5.2 we saw how a sampling distribution can make this vague intuition precise by specifying the probability that a specified degree of representativeness will be achieved by a random sample. Thus, sampling distributions provide what has been called "certainty about uncertainty."[14]

In Chapter 6 we will see for the first time how the theory of sampling distributions can be put to practical use in the analysis of data. We will find that, although the calculations of Chapter 5 seem to require the knowledge of unknowable quantities (e.g., μ and σ), when analyzing data one can nevertheless estimate the probable magnitude of sampling error using only information contained in the sample itself.

In addition to their application to data analysis, sampling distributions provide a basis for comparing the relative merits of different methods of analysis. For example, consider sampling from a normal population with mean μ. Of course, the sample mean $\overline{Y}$ is an estimator of μ. But since a normal distribution is symmetric, it is also the population median, so the sample *median* is also an estimator of μ. How, then, can we decide which estimator is better? This question can be answered in terms of sampling distributions, as follows: Statisticians have determined that, if the population is normal, the sample median is inferior to the sample mean in the sense that its sampling distribution, while centered at μ, has a standard deviation larger than $\frac{\sigma}{\sqrt{n}}$.

Consequently, the sample median is less efficient (as an estimator of μ) than the sample mean; for a given sample size n, the sample median provides less information about μ than does the sample mean. (If the population is not normal, however, the sample median can be much more efficient than the mean.)

---

*It is true, however, that sometimes the investigator can force the sample to be representative with respect to some variable (not the one under study) whose population distribution is known; for example, a stratified random sample as discussed in Section 1.3. The methods of analysis given in this book, however, are only appropriate for *simple* random samples and cannot be applied without suitable modification.

Supplementary Exercises 5.S.1–5.S.13

(*Note:* Exercises preceded by an asterisk refer to optional sections.)

5.S.1 In an agricultural experiment, a large field of wheat was divided into many plots (each plot being 7 × 100 ft) and the yield of grain was measured for each plot. These plot yields followed approximately a normal distribution with mean 88 lb and standard deviation 7 lb (as in Exercise 4.3.5). Let $\overline{Y}$ represent the mean yield of five plots chosen at random from the field. Find $\Pr\{\overline{Y} > 90\}$.

5.S.2 Consider taking a random sample of size 14 from the population of students at a certain college and measuring the diastolic blood pressure each of the 14 students. In the context of this setting, explain what is meant by the sampling distribution of the sample mean.

5.S.3 Refer to the setting of Exercise 5.S.2. Suppose that the population mean is 70 mm Hg and the population standard deviation is 10 mm Hg. If the sample size is 14, what is the standard deviation of the sampling distribution of the sample mean?

5.S.4 The heights of men in a certain population follow a normal distribution with mean 69.7 inches and standard deviation 2.8 inches.[15]

(a) If a man is chosen at random from the population, find the probability that he will be more than 72 inches tall.

(b) If two men are chosen at random from the population, find the probability that (i) both of them will be more than 72 inches tall; (ii) their mean height will be more than 72 inches.

5.S.5 Suppose a botanist grows many individually potted eggplants, all treated identically and arranged in groups of four pots on the greenhouse bench. After 30 days of growth, she measures the total leaf area Y of each plant. Assume that the population distribution of Y is approximately normal with mean $= 800\text{ cm}^2$ and SD $= 90\text{ cm}^2$.[16]

(a) What percentage of the plants in the population will have leaf area between 750 cm^2 and 850 cm^2?

(b) Suppose each group of four plants can be regarded as a random sample from the population. What percentage of the groups will have a group mean leaf area between 750 cm^2 and 850 cm^2?

5.S.6 Refer to Exercise 5.S.5. In a real greenhouse, what factors might tend to invalidate the assumption that each group of plants can be regarded as a random sample from the same population?

***5.S.7** Consider taking a random sample of size 25 from a population in which 42% of the people have type A blood. What is the probability that the sample proportion with type A blood will be greater than 0.44? Use the normal approximation to the binomial with continuity correction.

5.S.8 The activity of a certain enzyme is measured by counting emissions from a radioactively labeled molecule. For a given tissue specimen, the counts in consecutive 10-second time periods may be regarded (approximately) as repeated independent observations from a normal distribution (as in Exercise 4.S.1). Suppose the mean 10-second count for a certain tissue specimen is 1,200 and the standard deviation is 35. For that specimen, let Y represent a 10-second count and let $\bar{Y}$ represent the mean of six 10-second counts. Both Y and $\bar{Y}$ are unbiased—they each have an average of 1,200—but that doesn't imply that they are equally good. Find $\Pr\{1{,}175 \le Y \le 1{,}225\}$ and $\Pr\{1{,}175 \le \bar{Y} \le 1{,}225\}$, and compare the two. Does the comparison indicate that counting for 1 minute and dividing by 6 would tend to give a more precise result than merely counting for a single 10-second time period? How?

5.S.9 In a certain lab population of mice, the weights at 20 days of age follow approximately a normal distribution with mean weight = 8.3 gm and standard deviation = 1.7 gm.[17] Suppose many litters of 10 mice each are to be weighed. If each litter can be regarded as a random sample from the population, what percentage of the litters will have a total weight of 90 gm or more? (*Hint:* How is the total weight of a litter related to the mean weight of its members?)

5.S.10 Refer to Exercise 5.S.9. In reality, what factors would tend to invalidate the assumption that each litter can be regarded as a random sample from the same population?

5.S.11 Consider taking a random sample of size 25 from a population of plants, measuring the weight of each plant, and adding the weights to get a sample total. In the context of this setting, explain what is meant by the sampling distribution of the sample total.

5.S.12 The skull breadths of a certain population of rodents follow a normal distribution with a standard deviation of 10 mm. Let $\bar{Y}$ be the mean skull breadth of a random sample of 64 individuals from this population, and let μ be the population mean skull breadth.

(a) Suppose $\mu = 50$ mm. Find $\Pr\{\bar{Y}$ is within ± 2 mm of $\mu\}$.

(b) Suppose $\mu = 100$ mm. Find $\Pr\{\bar{Y}$ is within ± 2 mm of $\mu\}$.

(c) Suppose μ is unknown. Can you find $\Pr\{\bar{Y}$ is within ± 2 mm of $\mu\}$? If so, do it. If not, explain why not.

5.S.13 Suppose that every day for 3 months Bill takes a random sample of 20 college students, records the number of calories they consume on that day, finds the average of the 20 observations, and adds the average to his histogram of the sampling distribution of the mean.

Suppose also that every day for 2 months Susan takes a random sample of 30 college students and records the number of calories they consume on that day (which is fairly symmetric), finds the average of the 30 observations, and adds the average to her histogram of the sampling distribution of the mean.

(a) Can we expect Bill's distribution and Susan's distribution to have the same shape? Why or why not? If not, how will the shapes differ?

(b) Can we expect Bill's distribution and Susan's distribution to have the same center? Why or why not? If not, how will the centers differ?

(c) Can we expect Bill's distribution and Susan's distribution to have the same spread? Why or why not? If not, how will the spreads differ?