An Example of Sampling Distributions: Ratio Estimation (Using SRS)

Reconsider the population (N=6): 2 6 8 10 10 12

The population mean is $μ=8$ and its standard deviation is $σ= \sqrt{64/6}\dot{=}3.27.$

For simple-random samples (SRS) of size *n* = 2, we again have the following sampling distribution:

 Distribution of $\overbar{X}$ for n=2

 $\overbar{X}$ Frq

 4 1 [1/15]

 5 1 [1/15]

 6 2 [2/15]

 7 2 [2/15]

 8 2 [2/15]

 9 3 [3/15]

 10 2 [2/15]

 11 2 [2/15]

The mean of the sampling distribution is of course 8 and the sampling variance is$ 64/15\dot{=}$ 4.27, and so the standard error is $\sqrt{64/15}\dot{=}$ 2.07.

Suppose we know all of the values of another variable associated with this variable. (For example, suppose this variable is height (in cm, after subtracting 160; 162-160=2, 166-6=6, etc.) and the second variable is body mass (in kg, after subtracting 60).) The associated or *auxiliary* variable has values as follows:

 A B C D E F

Primary Variable: 2 6 8 10 10 12

Auxiliary Variable: 4 8 16 8 16 20

An Example of Sampling Distributions III (Continuation) Page Two

Since we know all of the values of the auxiliary variable, let us rearrange the population in ascending order of this supplementary variable:

 A B D C E F

Primary Variable: 2 6 10 8 10 12

Auxiliary Variable: 4 8 8 16 16 20

We also know the average of the auxiliary variable, which is 12 here.

Now consider again the 15 possible SRSs of size *n* = 2, along with the corresponding values of the auxiliary variable:

PV: 2,6 / 2,10 / 2,8 / 2,10 / 2,12 / 6,10 / 6,8 / 6,10 / 6,12 / 10,8 / 10,10 / 10,12 / 8,10/ 8,12 / 10,12

AV: 4,8 / 4,8 / 4, 16 / 4,16 / 4,20 / 8,8 / 8,16 / 8,16 / 8,20 / 8,16 / 8,16 / 8,20 / 16,16 / 16,20 / 16,20

Using *x* for the auxiliary variable and *y* for the primary variable (the one whose mean we desire to estimate), the corresponding sample means are as follows:

$\overbar{y}$: 4 / 6 / 5 / 6 / 7 / 8 / 7 / 8 / 9 / 9 / 10 / 11 / 9 / 10 / 11

$\overbar{x}$: 6 / 6 / 10 / 10 / 12 / 8 / 12 / 12 / 14 / 12 / 12 / 14 / 16 / 18 / 18

If we were to know each and every value of both variables, then it would follow that

 $\frac{μ\_{y}}{μ\_{x}}=\frac{Nμ\_{y}}{Nμ\_{x}}=\frac{τ\_{y}}{τ\_{x}}=R$

where $τ\_{x}$ and $τ\_{y}$ are the respective population totals and *R* is the population ratio of such.

Thus, since

 $μ\_{y}=\frac{τ\_{y}}{τ\_{x}}μ\_{x}=Rμ\_{x} $

and we know $μ\_{x}$ [$μ\_{x}$ = 12 for our example], we could logically use

 $\hat{μ}\_{y}=\frac{\sum\_{i=1}^{n}y\_{i}}{\sum\_{i=1}^{n}x\_{i}}μ\_{x}=\frac{\overbar{y}}{\overbar{x}}μ\_{x}=rμ\_{x}$

as a *ratio estimator* of the population mean$ μ$.

An Example of Sampling Distributions III (Continuation) Page Three

For example, for the last of the 15 possible samples, we have as the ratio estimate

 $\hat{μ}\_{y}=\frac{11}{18}\left(12\right)=\frac{22}{3}\dot{=}7.33$

Similarly, we arrive at all 15 ratio estimates:

 8 12 6 7.2 7

 12 7 8 54/7 $\dot{=}$ 7.71 9

 10 66/7 $\dot{=}$ 9.43 6.75 20/3 $\dot{=}$ 6.67 22/3 $\dot{=}$ 7.33

Thus, the sampling distribution of $\hat{μ}\_{y}$ is as follows:

 Distribution of $\hat{μ}\_{y}$ for n=2

 $\hat{μ}\_{y}$ Frq

 6 1 [1/15]

 6.67 1 [1/15]

 6.75 1 [1/15]

 7 2 [2/15]

 7.2 1 [1/15]

 7.33 1 [1/15]

 7.71 1 [1/15]

 8 2 [2/15]

 9 1 [1/15]

 9.43 1 [1/15]

 10 1 [1/15]

 12 2 [2/15]

Not unexpectedly, the ratio estimator is not unbiased, as the average of the sampling distribution is seen to be 8.27 compared with the population average of 8.

Example of Sampling Distributions III (Continuation) Page Four

Nevertheless, calculating the sampling variance results in 3.23, and so the standard error is 1.80, compared with that of 2.07 for the corresponding SRS sampling distribution given earlier. (Also, it is of interest to note the range of the sampling distribution here is 12 – 6 = 6 whereas that for the SRS was 11-4 = 7.)

Although the ratio estimator is biased, such would be small in samples of practical size. (Even here, for n=2, the bias was only about 3 percent more than the true value of 8.) Thus, combined with the improvement in precision (1.80 < 2.07), this illustrates that having such supplementary information (the known auxiliary variable) could be useful in improving estimation.

Finally, it should be clear a ratio estimate is most appropriate when the relationship between y and x (and $\overbar{y}$ and $\overbar{x}) $has a positive linear correlation, with such a summarizing straight line passing through the origin. Relatedly, a consequence of the linear correlation being close to one would be small bias.

Historical Note

Although the first use of the term *ratio estimate* in writing was by W. Edwards Deming in 1950, the concept of ratio estimation was first used by Laplace in 1802 to estimate the population of France. (In her book *Sampling: Design and Analysis* (Second Edition), Sharon Lohr uses her English translation of some of what Laplace wrote in French on this as her chapter-opening epigraph for Chapter 4: Ratio and Regression Estimation.)