Class Worksheet

April 18 and 23
Math 125 *Kovitz* 2025

Measuring Employment and Unemployment **(Chapter 22)**

Not a simple random sample.

Strata sampling.

Weighting the Sample.

Cluster samples.

Cluster samples are less informative than simple random samples of the same size. So the simple random sample formulas for the standard error do not apply.

The half-sample method may be used to estimate the standard error. Two halves of the sample are used to estimate the parameter. The difference of each estimate from their average is used to estimate the chance error.

The sample is rotated to avoid panel bias.

The data are carefully checked to avoid errors. This leads to high quality.

With samples of convenience, standard errors usually do not make sense because the method is not a probability method.

When a sample is taken by a probablity method, it is possible not only to estimate parameters, but also to figure the likely size of the chance errors in the estimates.

The formulas for the standard error have to take into account the details of the probability method used to draw the sample. The formulas which apply to simple random samples will usually underestimate the standard errors in cluster samples.

When bias operates more or less evenly across the sample, it cannot be detected just by looking at the data.

The Accuracy of Averages **(Chapter 23)**

When drawing at random from a box:

$$\text{EV for average of draws} = \text{average of box.}$$

$$\text{SE for average of draws} = \frac{\text{SE for sum}}{\text{number of draws}}.$$

If you wish, you may compute the SE for an average directly from the SD of the box by the formula

$$\text{SE for average} = (\text{SD of box})/\sqrt{\text{number of draws}}.$$

When drawing at random from a box, the probability histogram for the average of the draws will follow the normal curve, even if the contents of the box do not. The histogram must be put into standard units, and the number of draws must be reasonably large.

When drawing at random with replacement from a box of tickets, multiplying the number of draws by a factor (like 4) divides the SE for the average of the draws by the square root of that factor ($\sqrt{4} = 2$).

With a simple random sample, the SD of the sample can be used to estimate the SD of the box. The estimate is good when the sample is large.

The SE shows the likely size of the amount off. It is a give-or-take number.

For a given box model there are several SEs, each showing the likely size of a certain chance error. The corresponding formulas are:

$$\begin{aligned}
\text{SE for sum} &= \sqrt{\text{number of draws}} \times \text{SD of box} \\
\text{SE for average} &= \frac{\text{SE for sum}}{\text{number of draws}} = \frac{\text{SD of box}}{\sqrt{\text{number of draws}}} \\
\text{SE for count} &= \text{SE for sum, from a 0–1 box} \\
\text{SE for percent} &= \frac{\text{SE for count}}{\text{number of draws}} \times 100\% = \frac{\text{SD of zero-one box}}{\sqrt{\text{number of draws}}} \times 100\%
\end{aligned}$$

The SE for the sum is basic; the other formulas all come from that one. These formulas apply to draws made at random with replacement from a box.

Do not confuse the SD and the SE for the average.

- The SD says how far a number in the box is from average—for a typical number.
- The SE for the average says how far the sample average is from the population average—for a typical sample.

A reminder. The formulas for simple random samples should not be applied mechanically to other kinds of samples.

Problems to think about

(Chapter 22)

The Current Population Survey sample is split into two independent halves. From one half, the number of employed persons is estimated as 118.3 million; from the other half, it is estimated as 118.1 million. Combine these two estimates, and attach a standard error to the result.

(Chapter 23)

A hundred draws are made at random with replacement from a box.

(a) If the sum of the draws is 710, what is their average?

(b) If the average if the draws is 6.7, what is their sum?

Fill in the table below, for draws made at random with replacement from the box $\boxed{3\ 5\ 6\ 7\ 9}$.

Number of draws	*SE for sum of draws*	*SE for average of draws*	*Expected value for average of draws*
25			
100			
400			

A box contains 100,000 tickets. The numbers on these tickets average out to 50, and the SD is 30.

(a) A hundred tickets are drawn at random with replacement. The average of these draws will be around _____, give or take _____ or so.

(b) What if the draws are made without replacement?

(c) Find the chance (approximately) that the average of the draws will be in the range 20 to 80.

(d) Repeat, for the range 41 to 59.

Forty thousand draws will be made at random with replacement from a box of tickets. The average of the numbers in the box is 1500. The SE for the average of the draws is computed, and turns out to be 3. True or false:

(a) About 68% of the tickets in the box are in the range 1497 to 1503.

(b) There is about a 68% chance for the average of the forty thousand draws to be in the range 1497 to 1503.

Match each phrase in list A with one in list B.

List A	*List B*
population	draws
population average	average of the box
sample	box
sample average	number of draws
sample size	average of the draws

A university has 12,000 registered students. As part of a survey, 400 of these students are chosen at random. The average age of the sample students turns out to be 25.46 years, and the SD is 3.6 years.

(a) The average age of all 12,000 students is estimated as _____, give or take _____ or so.

(b) Find an approximate 95%-confidence interval for the average age of all 12,000 registered students.

A certain city has 7,000,000 residents. A survey team studying annual earned income choses 10,000 residents at random. The sample incomes average out to \$16,000, and the SD is \$6,000. A histogram is plottted for the sample incomes, and it does not follow the normal curve.

(a) If possible, find an approximate 68%-confidence interval for the average annual earned income of all the 7,000,000 residents of the city.

(b) True or false, and explain: about 68% of the 7,000,000 residents of the city had annual earned income between \$15,940 and \$16,060.

A state has 2,000,000 households. A consumer survey choses a simple random sample of 2,000 households and finds that the average number of boxes of laundry detergent purchased in the last year in the sample households was 7.26 and the SD was 2.83, and that 1,142 of the 2,000 households have a smoker.

(a) The average number of boxes of laundry detergent purchased per household in the last year in the entire state is estimated as _____; this estimate is likely to be off by _____ or so.

(b) If possible, find an approximate 95%-confidence interval for the average number of boxes of laundry detergent purchased per household in the last year in the entire state. If this isn't possible, explain why not.

(c) If possible find an approximate 99.7%-confidence interval for the percentage of households in the entire that have a smoker. If this isn't possible, explain why not.

Five hundred draws are made at random with replacement from a box. The sum of the draws is 1706. Can you estimate the average of the box? Can you attach a standard error to your estimate? Explain briefly.