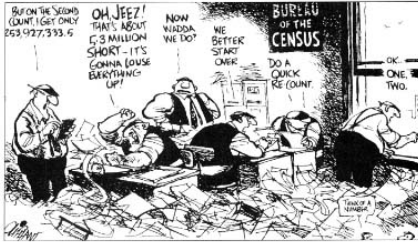


## Chance Error in Sampling



1

---

---

---

---

---

---

---

---



How different is the sample percentage from the population percentage?

The purpose of this chapter is to show how box models can be used to understand the error in simple random sampling.

2

---

---

---

---

---

---

---

---

## Sampling from a large population



---

---

---

---

---

---

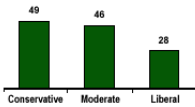
---

---

## Gallup Poll on Gun Ownership

Gun Ownership:  
By Political Ideology

% Who own a gun



**Liberals are least likely to say they have a gun in their home or on their property.**

*Do you have a gun in your home?*

*Do you have a gun anywhere else on your property such as in your garage, barn, shed or in your car or truck?*

± 3% Margin of Error  
October 14-17, 2002  
Sample Size= 1,002

4

---

---

---

---

---

---

---

---

---

---

## Review: The Sum of Draws from a Box (with replacement)

- The **expected value** of the sum is the number of draws times the box average
- The **standard error** of the sum is the square root of the number of draws times the box SD

5

---

---

---

---

---

---

---

---

---

---

## Special Case of a 0-1 Box

- Sum counts the number of times something happens
- Box average = fraction of tickets which equal 1
- Box SD =  $\sqrt{(\text{fraction of 0's}) \times (\text{fraction of 1's})}$

6

---

---

---

---

---

---

---

---

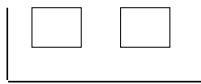
---

---

## Box Model for Simple Random Sampling

Population has 100 people, 20 of whom wear glasses. A sample of 10 is taken with replacement.

What is the box model for the number of people in the sample who wear glasses?



10 draws from this box

---

---

---

---

---

---

---

---

---

---

Of the 100 people, 35 are over the age of 30. What is the box model for counting the number of people in a sample of size 15 who are over the age of 30?



15 draws from this box

8

---

---

---

---

---

---

---

---

---

---

Of a population of 500,000 people, 20% are unemployed. A simple random sample of size 400 is taken.



How many people in the population are unemployed?

What's the box model?



400 draws from this box

What's the box average?

9

---

---

---

---

---

---

---

---

---

---

500,000 people, 20% are unemployed. A simple random sample of size 400 is taken.

How *many* people in the sample do you expect to be unemployed?

What *percent* of the sample do you expect to be unemployed?

10

---

---

---

---

---

---

---

---

With a simple random sample, the expected value of the sample percentage equals the population percentage.

11

---

---

---

---

---

---

---

---

Of a population of 500,000 people, 20% are unemployed. A simple random sample of size 400 is taken.



Expected number unemployed in sample?

What is the SD of the box?

What is the SE of the number of unemployed in the sample?

12

---

---

---

---

---

---

---

---

In the sample we expect \_\_\_\_ people to be unemployed, with an error of about \_\_\_\_.

80 people is what percent of the sample?

8 people is what percent of the sample?

In the sample we expect \_\_\_\_% of the people to be unemployed, with an error of about \_\_\_\_%.

13

---

---

---

---

---

---

---

---

The SE of the number was 8 and 8 is 2% of 400. So the SE of the percent in the sample is 2%.

This is what we did:

Percentage = 100 x number/sample size

SE of percentage =  $\frac{\text{SE of number}}{\text{sample size}} \times 100\%$

14

---

---

---

---

---

---

---

---

### A Formula

SE of percentage =  $\frac{\text{SE of number}}{n} \times 100\%$

=  $\frac{\sqrt{n} \text{ SD of box}}{n} \times 100\%$

=  $\frac{\text{SD of box}}{\sqrt{n}} \times 100\%$

Formula is exact for sampling with replacement and approximate without replacement.

15

---

---

---

---

---

---

---

---

Box SD = .4

What is the SE of the sample percentage for samples of size:

n=100?

n=400?

n=1600?

16

---

---

---

---

---

---

---

---

### Summary and Consequences

- The expected value of the sample percentage equals the population percentage
- The sample percentage becomes more accurate as the sample size increases. The SE of the sample percentage equals the SE of the sample sum divided by the sample size times 100

17

---

---

---

---

---

---

---

---

**?** Suppose that a town has a population of 100,000, and 40,000 of the residents own a gun. A random sample of 400 residents is taken. Of them we expect that \_\_\_\_% will own a gun, plus or minus \_\_\_\_% or so.

18

---

---

---

---

---

---

---

---

? A town has 40,000 households and 75,000 adults. 400 households are chosen at random and in each of the 400, the number of adults who own a gun is found. This total number, divided by 75,000 is used to estimate the percentage of adults in the town who own a gun. In fact, 55% of the adults own guns.

Answer the following question if you can, or if you cannot, explain why you cannot: In the sample the estimate of the percentage is likely to be off by \_\_\_% or so.

---

---

---

---

---

---

---

---

---

---

### Comparison of two situations:

Population size = 500,000, percent unemployed = 20%, Sample size = 400

SE of sample percentage:

Population size = 100,000, percent unemployed = 20%, sample size = 400

SE of sample percentage:

---

---

---

---

---

---

---

---

---

---

### Using the Normal Curve

Expected value of sample percentage = 20%

SE of sample percentage = 2%

What is the chance that sample percentage is more than 25%?

25% is \_\_\_\_\_ standard units greater than 20%

The chance is therefore about

---

---

---

---

---

---

---

---

---

---

What is the chance that the sample percentage is between .16 and .24?

.16 is \_\_\_\_\_ standard units

.24 is \_\_\_\_\_ standard units

The chance is therefore about \_\_\_\_\_ %

If the sampling were to be repeated many times, about \_\_\_\_\_ % of the sample percentages would be between .16 and .24.

22

---

---

---

---

---

---

---

---

## The Correction Factor

Results on SE of sample percentage pretended sampling was *with* replacement, but simple random sampling is done *without* replacement.

If the population size is large relative to the sample size, the with replacement approximation is good.

23

---

---

---

---

---

---

---

---

An exact formula for SE in sampling without replacement:

SE without = SE with x correction factor

$$\text{correction factor} = \sqrt{\frac{\text{population size} - \text{sample size}}{\text{population size} - 1}}$$

24

---

---

---

---

---

---

---

---



Population size = 100,000

sample size	correction factor
100	.9995
1000	.9950
5000	.9747
10,000	.9487

25

---

---

---

---

---

---

---

---

For estimating percentages the accuracy does not depend on the size of the population so long as the size of the sample is small relative to the population size.

26

---

---

---

---

---

---

---

---

**Liberals are least likely to say they have a gun in their home or on their property.**

*Do you have a gun in your home?*

*Do you have a gun anywhere else on your property such as in your garage, barn, shed or in your car or truck?*

± 3% Margin of Error

October 14-17, 2002

Sample Size= 1,002

Where does this 3% "margin of error come from?"

27

---

---

---

---

---

---

---

---

Suppose that ownership was equal to 50%. What would the SE of the sample percentage be with  $n=1005$ ?

---

---

---

---

---

---

---

---

## Where Have We Been?



- We can regard the percentage in a simple random sample as being obtained by drawing with replacement from a 0-1 box.
- The expected value of the sample percentage is the population percentage.
- The SE of the sample *percentage* can be found by dividing the SE of the sample *sum* by the sample size and multiplying by 100. (Convert numbers to percents.)

---

---

---

---

---

---

---

---

- If the sample size is large, the normal curve can be used to approximate probabilities.
- If the population size is large relative to the sample size, a simple random sample without replacement is very much like drawing with replacement.
- There is a correction factor for modifying the SE of the sample percentage if the sample size is not small relative to the population size.

---

---

---

---

---

---

---

---

? In a certain town there are 30,000 registered voters, of whom 12,000 are Democrats. A survey organization is about to take a simple random sample of 1000 registered voters.

There is about a 50% chance that the percentage of Democrats in the sample will be larger than \_\_\_\_.

There is about a 25% chance that the percentage of Democrats in the sample will be larger than \_\_\_\_.

---

---

---

---

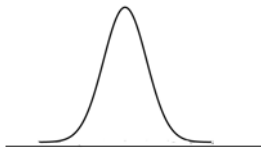
---

---

---

---

30,000 voters. 12,000 Democrat.  $n=1000$



EV of % Dem in sample =

SE of % Dem in sample =

---

---

---

---

---

---

---

---

EV of % Dem = . SE =

There is about a 50% chance that % Dem in sample will be greater than .

There is about a 25% chance that the percentage of Democrats in the sample will be larger than \_\_\_\_.

---

---

---

---

---

---

---

---

? **True or False:** City A has 100,000 voters and City B has 400,000 voters. Other things being equal, a sample of 0.1% of the voters from City A is about half as accurate as a sample of 0.1% of the voters from City B.

34

---

---

---

---

---

---

---

---

? Households in a city contain an average of 2.2 people, with an SD of 1.8. 15% of the households consist of just one person. A simple random sample of 500 households is taken.

There is about an 80% chance that between \_\_\_% and \_\_\_% of the sampled households consist of just one person.

35

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

36