Accuracy of Percentages

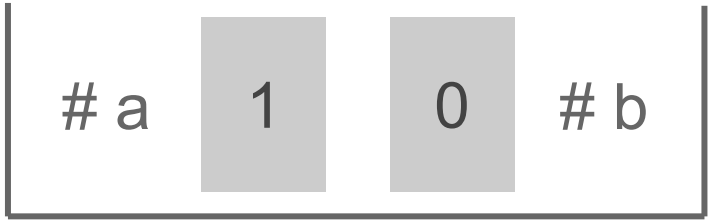
Gaston Sanchez

Confidence Intervals for percentages

(FPP chapter 21)

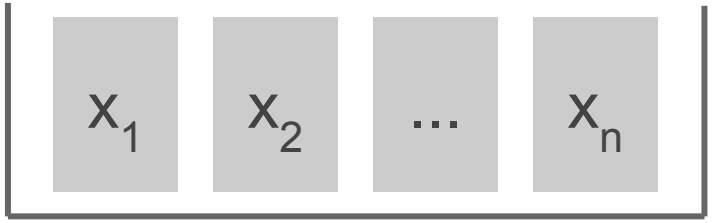
Two main types of boxes

FPP Chap 21 (estimation)



Interest in:

- Sum of draws (# 1's)
- Percentage of 1's



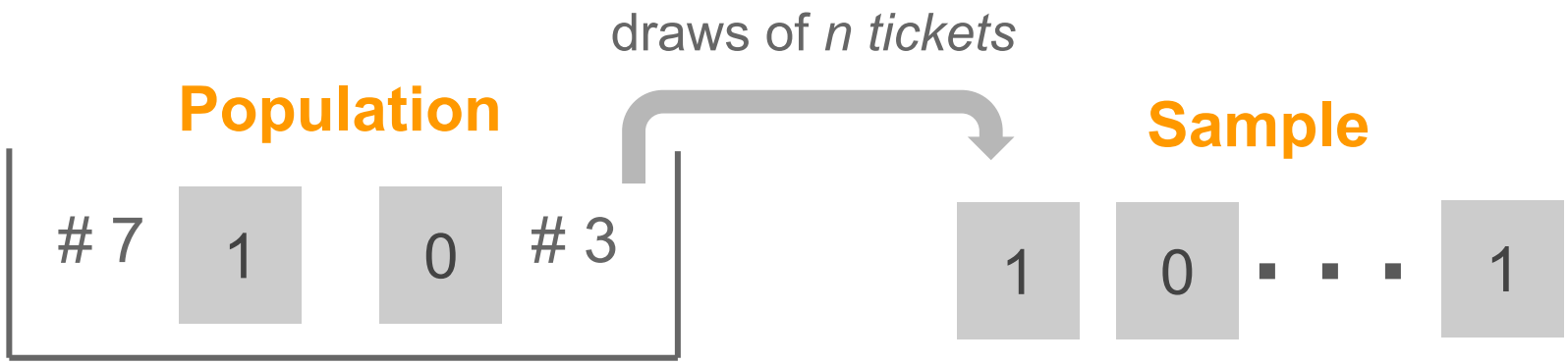
Interest in:

- Sum of draws
- Average of draws

FPP Chap 23 (estimation)

FPP Chapter 21: Deals with the **percentage** of draws from a box of tickets 1's and 0's

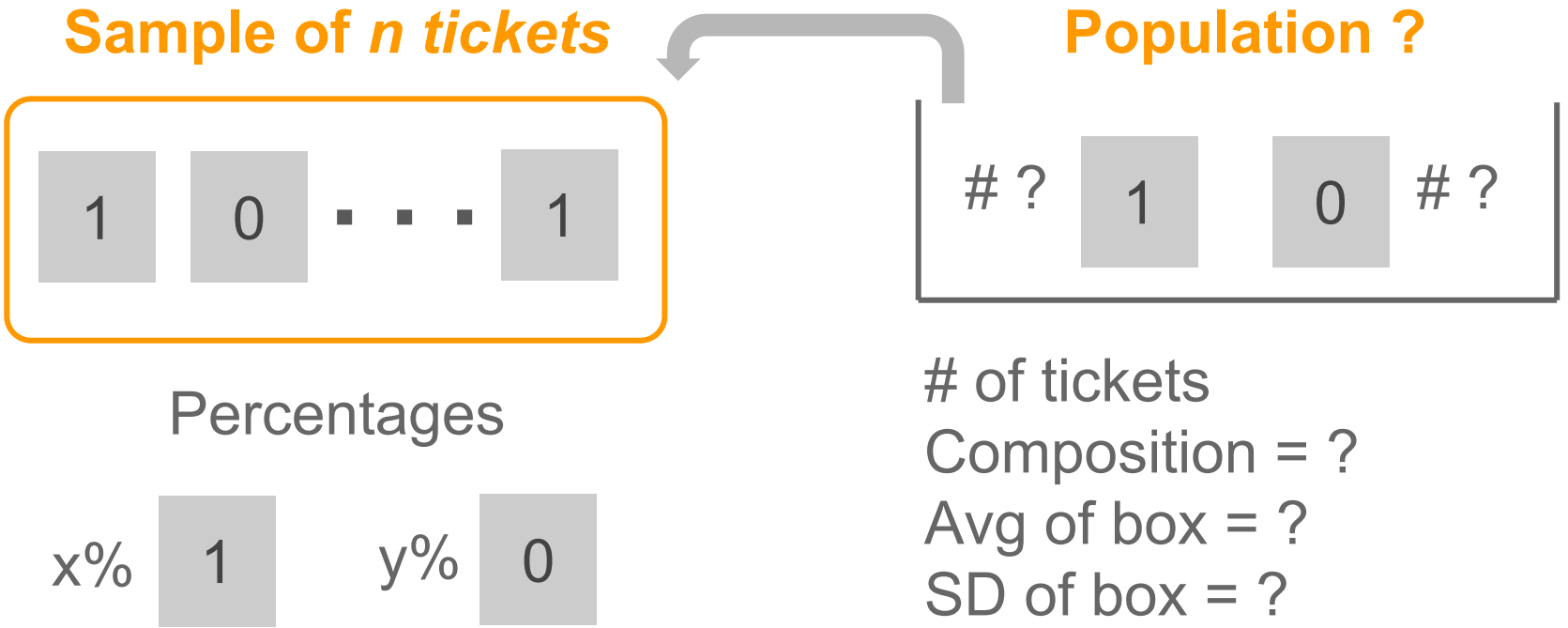
So far we've worked with box of known composition



10 tickets
Known composition
Avg of box
SD of box

Compute Statistic
Sum of 1's
Percentage 1's
SE of percentage

Now we turn in the other direction



What can we say about the percentages of the box

How accurate is an **estimated** percentage likely to be?

Recap

Statistic = Parameter + Chance Error + Bias

what we want to estimate

Likely size given by SE

Hard to measure

Estimate (statistic) likely to be equal to parameter, but off by an SE or so

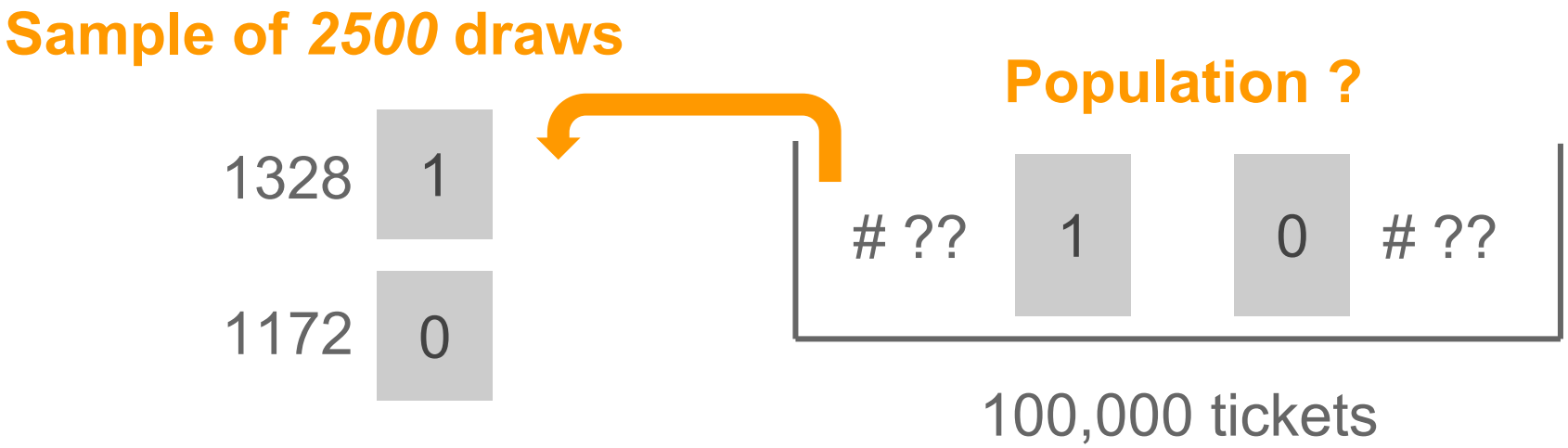
Recap

Statistic = Parameter + Chance Error + Bias

The amount of SE determines accuracy

Example
Page 377

Survey of 2,500 from 100,000 eligible voters



Survey of 2,500 from 100,000 eligible voters

Sample of *2500* draws

1328	1
1172	0

% 1's = 1328 / 2500 = 53%

SD sample = $\sqrt{0.53 \times 0.47}$ = 0.50

SD(box) by bootstrap method = 0.50

SE sum = $\sqrt{2500}$ x 0.50 = 25

SE percentage = 25 / 2500 = 1%

Proportion of voters is 53% give or take 1% = (52%, 54%)

Confidence Intervals

Confidence Intervals

A confidence Interval is used to give a **likely range for the parameter**

$$\text{Statistic} = \text{Parameter} + \text{Chance Error} + \text{Bias}$$

C.I. for the parameter

Confidence Intervals and Confidence Levels

C.I. at the **68%** level is:

statistic ± 1 SE

C.I. at the **95%** level is:

statistic ± 2 SE

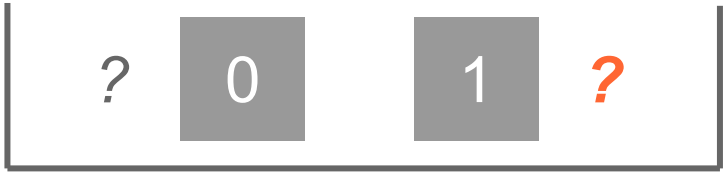
C.I. at the **99.7%** level is:

statistic ± 3 SE

Example: Field Poll (2010)

748 CA likely voters

43% support Barbara Boxer

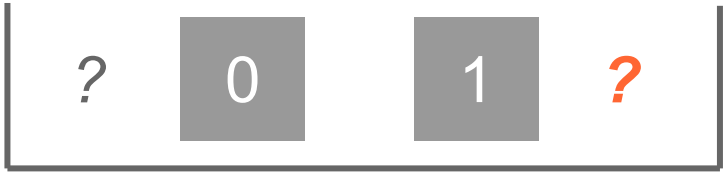


748 draws w/o replacement

Parameter: # voters for Boxer

Bootstrap method: estimate SD of box using SD of sample

Example: Field Poll (2010)



748 draws w/o replacement

parameter

SD of sample = $(1-0)\sqrt{(0.43)(0.57)} \approx 0.495$

SD of box by Bootstrap method

SE sum = $\sqrt{748}\,(0.495) = 13.54$ (with replacement)

SE % = 100% (SE sum / # draws) = 13.54 / 748 = 1.81%

Example: Field Poll (2010)

Estimated percentage = 43%

SE % = 1.81%

68% CI: 43% ± 1(1.81%) = 43% ± 1.81% = 41.2% to 44.8%

95% CI: 43% ± 2(1.81%) = 43% ± 3.6% = 39.3% to 46.6%

99.7% CI: 43% ± 3(1.81%) = 43% ± 5.43% = 37.5% to 48.4%

C.I. Demo

About Confidence Intervals

FPP chaps 17, 18, 20 are like throwing a dart at the target (i.e. **EV**)

FPP chaps 19, 21, 23 are like throwing a ring (i.e. **CI**) which may or may not cover an invisible stick (i.e. **parameter**)

C.I. Rules

About Confidence Intervals

It is WRONG to say: “There’s a 95% chance that the % of CA likely voters who support Boxer is between 39.4% and 46.6%”

It is CORRECT to say: “We are 95% **confident** that the % of CA likely voters who support Boxer is between 39.4% and 46.6%”

The parameter is NOT a random number

It feels like chance, but it’s not chance

Chance variability is in the sample process, NOT in the parameter.

To make a Confidence Interval

1) Check Normal curve makes sense:

- at least 25 draws
- statistic ± 2 SEs should be OK

2) No bias (or negligible)

SRS or draws with replacement means no selection bias

To make a Confidence Interval

3) CIs are for parameters

4) Chance is for sample values, sums of draws, random things

5) Average and SD are for data lists of numbers, tickets in a box

6) EV and SE are for sums of draws, sample values, sample averages, sample %'s