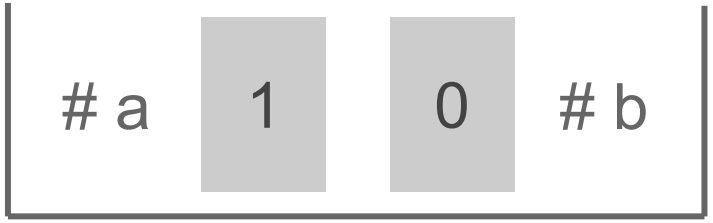
Sampling Distributions: Sample Average

Gaston Sanchez

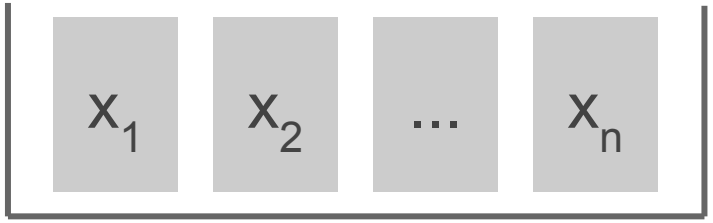
Two main types of boxes

Chap 21 (estimation)



Interest in:

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



Interest in:

- Sum of draws
- Average of draws

Chap 23 (estimation)

Formulas for Sum of draws (reminder)

$$\text{EV sum} = (\#\text{ of draws}) \times (\text{avg of box})$$

$$\text{SE sum} = \sqrt{\#\text{ draws}}\ (\text{SD box})$$

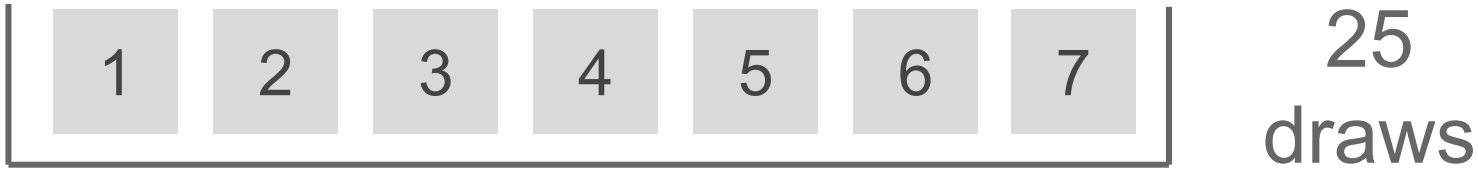
Formulas for Average

$$\text{EV } \mathbf{avg} = \text{avg of box}$$

$$\text{SE } \mathbf{avg} = \text{(SD box)} / \sqrt{\text{\# draws}}$$

First Example

25 draws at random with replacement from box:



Avg of box = (1 + 2 + … + 6 + 7) / 7 = 4

SD of box = 2

SE sum = $\sqrt{25}\,(2) = 10$

SE avg = 10 / 25 = 0.4

Example 1

SRS from box [1, 2, 3, 4, 5, 6, 7]

Avg in 25 draws = 4

Sample SD = 2

The average of the draws will be around **4**, give or take **0.4**

Demo: numbers from box

Sample Average

Sample Average

A random sample is taken from a box of unknown composition.

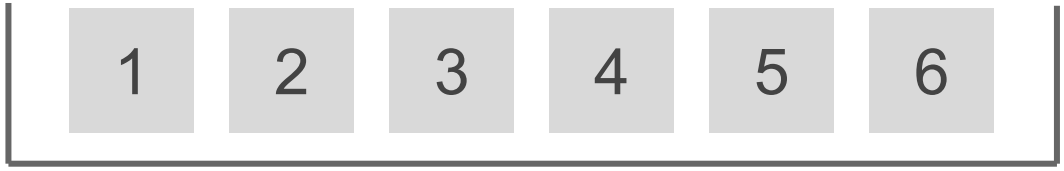
We want to estimate the average of the box.

We use the average of the draws as the statistic.

We use the SE for the sample average (i.e. bootstrap method) to compute a confidence interval for the estimated sample average.

Second Example

Roll a fair die 100 times. Chance of avg > 3.67?



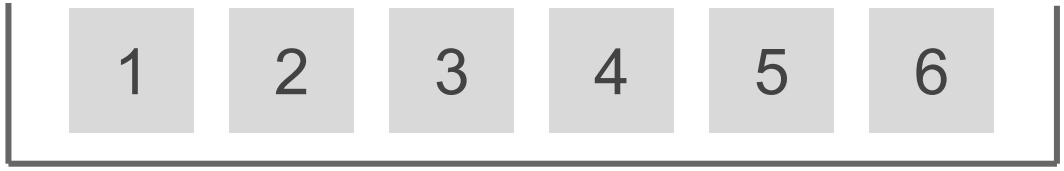
Avg of box = (1 + 2 + … + 6) / 6 = 3.5

SD of box = 1.7

SE sum = $\sqrt{100}\,(1.7) = 17$

SE avg = 17 / 100 = 0.17

Roll a fair die 100 times. Chance of avg > 3.67?

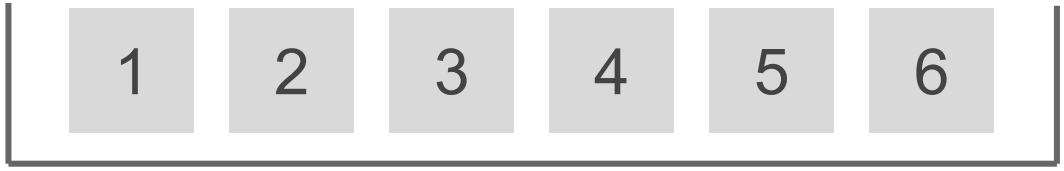


At least 25 draws

EV avg ± 2SE = 3.5 ± 2(0.17) within range

Thus, we can use normal approximation

Roll a fair die 100 times. Chance of avg > 3.67?

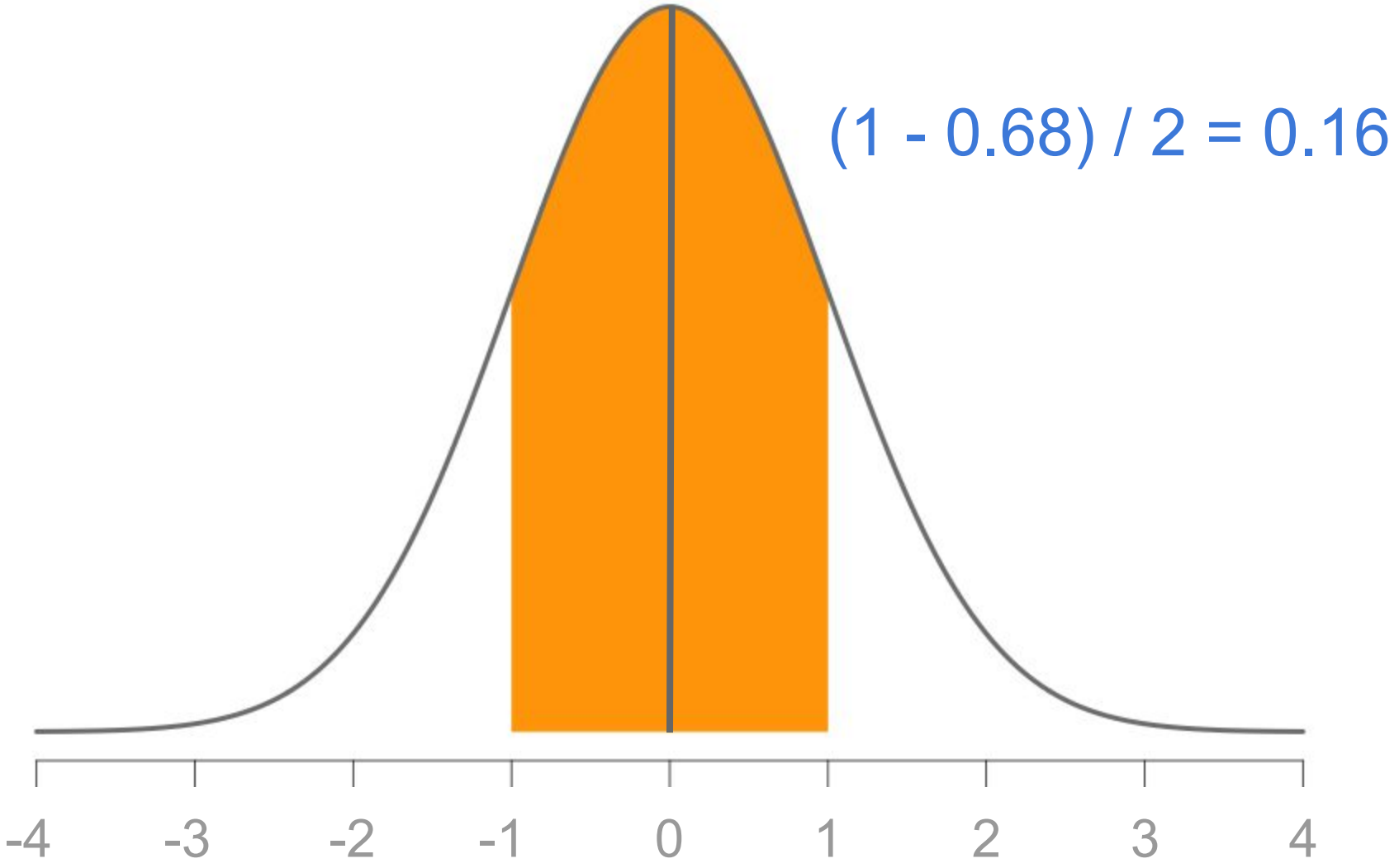


100 draws

$$SU_1 = \frac{3.5 - 3.5}{0.17} = 0$$

$$SU_2 = \frac{3.67 - 3.5}{0.17} = 1$$

68% observations within 1 SD of average



No need for continuity correction

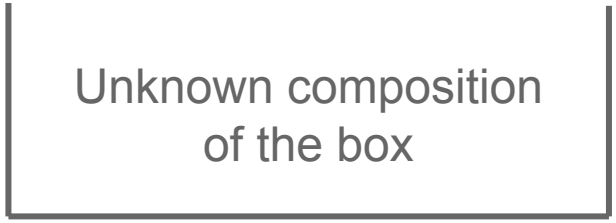
No continuity correction with averages since averages are continuous

Another Example

Turning to statistical estimation and inference

Average income of 25,000 families in a town?

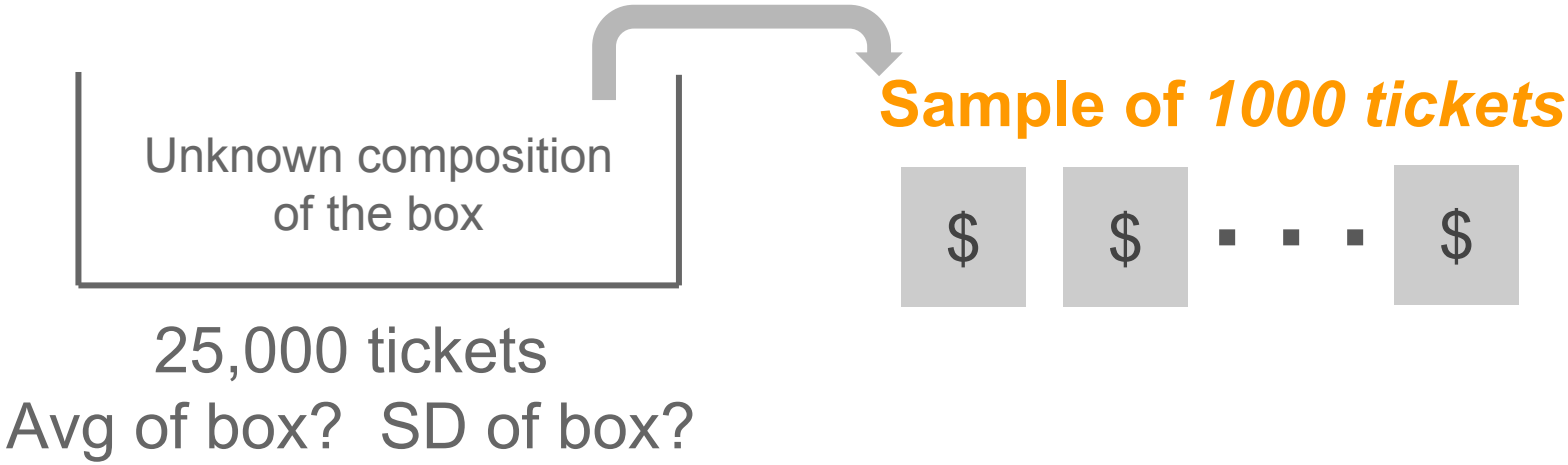
SRS of 1,000 families, with average income of $62,400, and SD of $53,000.



25,000 tickets
Avg of box? SD of box?

Sample of *1000 tickets*

Turning to statistical estimation and inference



SRS of 1,000 families with:

- Average of sample $62,400
- SD of sample $53,000

SE of Sample

SRS of 1,000 families with:

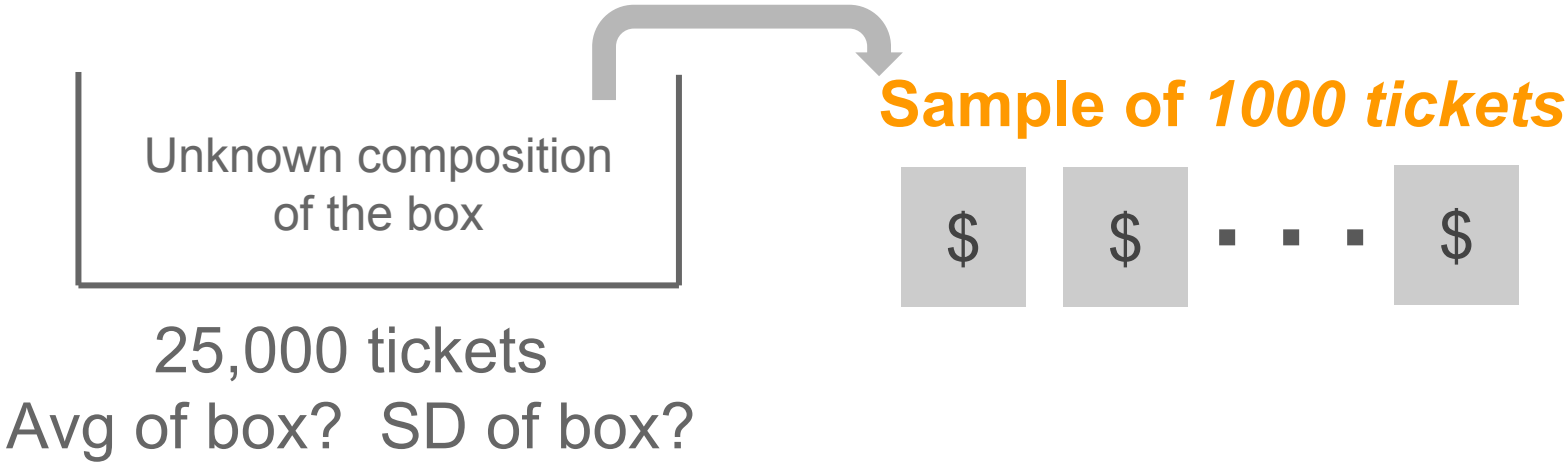
- Average of sample $62,400
- SD of sample $53,000

SD sum = 53,000

SE sum = $\sqrt{1000}$ (53,000) = 1,700,000

SE avg = 1,700,000 / 1,000 = 1,700

Turning to statistical estimation and inference



The average income of all 25,000 families can be estimated as $62,400 ± $1,700

Distribution of Sample Averages

From Central Limit Theorem, the sample avg follows a normal distribution

2 SEs either way from the sample average:

$62,400 ± 2($1,700) = $59,000 to $65,800

MSAT scores Example

Example

SRS of 400 Cal students

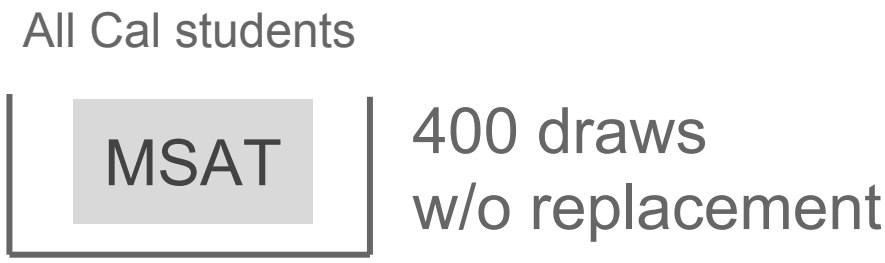
Avg MSAT in sample is 563

MSAT scores follow the normal curve

Sample SD = 90

The average MSAT score for all Cal students is around _statistic_ give or take _SE avg_

Example



SD box = ?

Bootstrap estimate of SD box

SE sum = $\sqrt{400}\,(90) = 1800$

SE avg = 1800 / 400 = 4.5

Example

SRS of 400 Cal students

Avg MSAT in sample is 563

Sample SD = 90

The average MSAT score for all Cal students is around 563 give or take 4.5 or so.

Example

95% CI for average MSAT for all Cal students is 563 ± 2(4.5)

CI: 554 to 572

99.7% CI for average MSAT for all Cal students is 563 ± 3(4.5)

CI: 549.5 to 576.5

True or False

About 68% of the students in the sample have MSAT scores between 473 and 653

(563 - 90, 563 + 90)

TRUE

We know twhat he sample Avg and SD are

True or False

About 68% of MSAT scores of all Cal students are between 558.5 and 567.5

(563 - 4.5, 563 + 4.5)

FALSE

We don’t know what the population Avg and SD are