

## Introduction to Statistical Inference

---

### 0.1 Set-up

Suppose a box contains 100 tickets, some of which are pink and some of which are blue (it's also possible to have all of one color). However, you don't know the proportion of each colors. Suppose that each pink ticket is worth \$1 and each blue ticket is worth \$5. So the box's value is something between \$100 and \$500. I will sell you the box for \$200. Because we know there are 100 tickets in total, a \$200 box would correspond to 75 pink tickets and 25 blue tickets. I won't tell you how many pink tickets are in the box. **Will you buy the box?**

I will let you take  $n = 8$  tickets from the box to help inform your decision. You can decide if you would like to draw the 8 tickets with or without replacement.

### 0.2 Discuss (part 1)

As a group, discussing the following: What are the benefits/drawbacks of drawing with versus without replacement? Which is preferable? Why? *Hint: maybe consider the extreme cases of being able to only draw 1 ticket ( $n = 1$ ), versus being able to draw  $n = 100$  tickets.*

### 0.3 Collect data

### 0.4 Discuss (part 2)

As a group, discuss the following:

1. What is your best estimate for the proportion of pink tickets in the box?
2. Why is this your best estimate?
3. How certain are you that this is exactly the true proportion?
4. How certain are you that this is close to the true proportion?
5. What margin of error would you be willing to accept? That is, how far away (in terms of decimal points) from the true proportion of pink tickets would you be happy with?
6. Suppose that there were actually 75 pink tickets. How likely (do you think) would it be to see a result exactly like the one we did?
7. How likely (do you think) would it be to see a result like the one you did, or more extreme (still supposing that there were actually 75 pink tickets)?
8. Can you come up with a way to quantify your uncertainty about the true value of  $p$  by using the 68-95-99.7 rule, along with the Central Limit Theorem (assume that  $n = 8$  samples is "large enough" for CLT to kick in)?
9. Will you buy the box?