Week 7 Reading Guide Part 1 – Sampling

## Sampling

### 7.1 – Sampling Bowl

**Why was it important to mix the bowl before the balls were sampled?**

**Why is it that our 33 groups of friends did not all have the same numbers of balls that were red out of 50, and hence different proportions red?**

### 7.2 – Virtual Sampling

**Why couldn’t we study the effects of sampling variation when we used the virtual shovel only once? Why did we need to take more than one virtual sample (in our case 33 virtual samples)?**

**Why did we not take 1000 “tactile” samples of 50 balls by hand?**

**Looking at Figure 8, would you say that sampling 50 balls where 30% of them were red is likely or not? What about sampling 50 balls where 10% of them were red?**

**In Figure 12, shovels were used to take 1000 samples each, the proportion of the shovel’s balls that were red was calculated for each sample, and then the distribution of these 1000 proportions were visualized using a histogram. We did this for shovels with 25, 50, and 100 slots in them. As the size of the shovels increased, the histograms got narrower. In other words, as the size of the shovels increased from 25 to 50 to 100, what happened to the proportions?**

1. They varied less
2. They varied by the same amount
3. They varied more

**What summary statistic was used to quantify how much the 1000 proportions red varied?**

1. The interquartile range
2. The standard deviation
3. The range: the largest value minus the smallest.

### 7.3 – Sampling Framework

**In the case of the bowl activity, what is the population parameter? Do we know its value?**

**What would performing a census in the bowl activity correspond to? Why was a a census not performed?**

**What purpose do point estimates serve in general? What is the name of the point estimate specific to the bowl activity? What is its mathematical notation?**

**How did we ensure that our tactile samples using the shovel were random?**

**Why is it important that sampling be done at random?**

**What are we inferring about the bowl based on the samples using the shovel?**

**What purpose did the sampling distributions serve?**

**What does the standard error of the sample proportion** $\hat{p}$ quantify?

The table below is a version of Table 2, matching sample sizes $n$ to different *standard errors* of the sample proportion $\hat{p}$, but with the rows randomly re-ordered and the sample sizes removed. Fill in the table by matching the correct sample sizes to the correct standard errors.

| Sample Size | Standard Error of $\hat{p}$ |
| --- | --- |
| n = | 0.094 |
| n = | 0.045 |
| n = | 0.069 |

**What is the difference between an *accurate* and a *precise* estimate?**

**How do we ensure that an estimate is *accurate*?**

**How do we ensure that an estimate is *precise*?**

**In a real-life situation, we would not take 1000 different samples to infer about a population, but rather only one. Then, what was the purpose of our exercises where we took 1000 different samples?**

**Figure 15 with the targets shows four combinations of “accurate versus precise” estimates. Draw four *sampling distributions* of the sample proportion** $\hat{p}$ that correspond with each option.

**High precision, low accuracy**

**High precision, high accuracy**

**Low precision, low accuracy**

**Low precision, high accuracy**