Economics 103 – Statistics for Economists

Francis J. DiTraglia

University of Pennsylvania

Lecture # 14

Weighing a Random Sample

Bag Contains 100 Candies

Estimate total weight of candies by weighing a random sample of size 5 and multiplying the result by 20.

Your Chance to Win

The bag of candies and a digital scale will make their way around the room during the lecture. Each team (2 students) gets a chance to draw 5 candies and weigh them.

Team with closest estimate wins the bag of candy!

Weighing a Random Sample

Procedure

When the bag and scale reach your team, do the following:

1. Fold the top of the bag over and shake to randomize.
2. Randomly draw 5 candies without replacement.
3. Weigh your sample and record the result in grams.
4. Rodrigo will enter your result into his spreadsheet and multiply it by 20 to estimate the weight of the bag.
5. Replace your sample and shake again to re-randomize.
6. Pass bag and scale to next team.

Sampling Distributions and Estimation – Part I

Building a Bridge Between Probability and Statistics

Questions to Answer

1. How accurately do our sample statistics estimate the unknown population parameters?
2. How can we quantify the uncertainty in our estimates?

How We'll Proceed

1. Use sequence of iid RVs as a model for random sampling from a population.
2. Parameters of these RVs represent population parameters.
3. Use tools of probability theory to study the behavior of sample statistics.

Step 1: Random Variable as Model for Population

Treat Population as RV rather than list of objects

Old Way	New Way
Among 138 million voters, 69 million will vote for Hillary Clinton	Bernoulli($p = 1/2$) RV

Old Way	New Way
List of heights for 97 million US adult males with mean 69 in and std. dev. 6 in	$N(\mu = 69, \sigma^2 = 36)$ RV

In the second example, our model assumes that the distribution of height is symmetric and bell-shaped.

Recall: (Simple) Random Sample

Definition in Words

Select a sample of n objects from a population in such a way that:

1. Each member of the population has the same probability of being selected
2. The fact that one individual is selected does not affect the chance that any other individual is selected
3. Each sample of size n is equally likely to be selected

Definition in Math

$X_1, X_2, \ldots, X_n \sim$ iid $f(x)$ if continuous

$X_1, X_2, \ldots, X_n \sim$ iid $p(x)$ if discrete

Random Sample Means *Sample With Replacement*

- Without replacement $\Rightarrow$ dependence between samples
- But sample small relative to popn. $\Rightarrow$ dependence negligible.
- This means our candy experiment (in progress) isn't bogus.

Step 2: iid RVs Represent Random Sampling from Popn.

Who Will Vote for Hillary Clinton Example

Poll random sample of 1000 registered voters:

$$X_1, \ldots, X_{1000} \sim \text{ iid Bernoulli}(p = 1/2)$$

Heights of US Males Example

Measure the heights of random sample of 50 US males:

$$Y_1, \ldots, Y_{50} \sim \text{ iid } N(\mu = 69, \sigma^2 = 36)$$

Key Question

What do the properties of the population imply about the properties of the sample?

What does the population imply about the sample?

Suppose that exactly half of US voters plan to vote for Hillary Clinton. If you poll a random sample of 4 voters, what is the probability that *none of them* are Hillary supporters?

What does the population imply about the sample?

Suppose that exactly half of US voters plan to vote for Hillary Clinton. If you poll a random sample of 4 voters, what is the probability that *none of them* are Hillary supporters?

$$(1/2)^4 = 1/16 = 0.0625$$

What does the population imply about the sample?

Suppose that exactly half of US voters plan to vote for Hillary Clinton. If you poll a random sample of 4 voters, what is the probability that *exactly half* are Hillary supporters?

What does the population imply about the sample?

Suppose that exactly half of US voters plan to vote for Hillary Clinton. If you poll a random sample of 4 voters, what is the probability that *exactly half* are Hillary supporters?

$$\binom{4}{2} (1/2)^2 (1/2)^2 = 3/8 = 0.375$$

The rest of the probabilities...

Suppose that exactly half of US voters plan to vote for Hillary Clinton and we poll a random sample of 4 voters.

$$
\begin{aligned}
P(\text{Exactly 0 Hillary Voters in the Sample}) &= 0.0625 \\
P(\text{Exactly 1 Hillary Voters in the Sample}) &= 0.25 \\
P(\text{Exactly 2 Hillary Voters in the Sample}) &= 0.375 \\
P(\text{Exactly 3 Hillary Voters in the Sample}) &= 0.25 \\
P(\text{Exactly 4 Hillary Voters in the Sample}) &= 0.0625
\end{aligned}
$$

You should be able to work these out yourself. If not, review the lecture slides on the Binomial RV.