Sampling Distribution of Sample Proportions

Example 1:

According to a Gallup survey, Gallup’s U.S. underemployment measure (March 8, 2012), which combines the percentage of workers who are unemployed and the percentage working part time but wanting full-time work in February, 2012, is 19.1%.

Suppose a sample of 100 workers were randomly selected on March 1, 2012 and were asked about whether or not they are underemployed.

Sample data are as follows (1 for ‘yes’; 0 for ‘no’):

```
1 0 0 0 0 1 0 0 0 0
0 0 0 1 0 0 0 0 0 0
1 0 0 0 1 0 0 0 1 0
0 1 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
1 0 0 0 0 1 0 0 0 0
0 0 0 1 0 0 0 0 0 0
1 0 0 0 1 0 0 0 1 0
0 1 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
```

Note: Sample proportion of ‘Yes’ responses is $16/100 = 0.16$

Describe the distribution of the population all sample proportions of size 100.
Calculate sample proportion of each sample and then form new population of sample proportions.
Solution:

The underlying population consists of millions of workers in the U.S. Data values in the underlying population would be millions of '1' and '0'. In other words, a worker is either underemployed or he/she is not.

For the underlying population, the proportion (or population mean) of underemployed workers is $p = 19.1\% = 0.191$.

Using data from the underlying population, we can form many, many samples of size 100 and then calculate the sample proportion of underemployed workers for each sample.

From theory, we know that:

1) The mean (or proportion) of all sample means (or proportions) with sample size of 100 is $\mu_x = p = 0.191$.

2) The standard error (or standard deviation of all sample proportions) =

$$\sigma_x = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.191(1-0.191)}{100}} = 0.0393.$$

In order to make the assumption that the distribution of the sample proportions is approximately normal, the following conditions have to be met:

Condition 1: Sample size (n) must be less than or equal to 5% of the population size (N), i.e., $n \leq 0.05N$.

(Condition 1 is needed to ensure that the sampled values are independent of each other.)

Condition 2: $n \cdot p \cdot (1-p) \geq 10$

The underlying population has many millions workers and the sample size (n) is only 100. We can see that the sample size (n) is definitely less than 5% of the population size (N).

And $n \cdot p \cdot (1-p) \geq 10$.

Since both Condition 1 and Condition 2 are met, we can assume that the population consisting of all sample proportions of size 100 is approximately normal and has a mean of 0.191 and standard deviation of 0.0393.