

Investigation 3-6: Comparison of Sampling Methods (Minitab Exploration)

In this exploration, you will use a Minitab macro to examine some of the properties of the sampling methods described in Investigation 3-2.

The file `shoppingPop.mtw` contains prices for 139 foods. Treat this as your population.

(a) Calculate the population mean and describe the shape and variability of the population.

(b) Write a Minitab macro to take 1000 samples, each containing 30 products, from this population, calculate the mean for each sample, and accumulate the results in C4. Describe the shape, center, and spread of this sampling distribution.

(c) Column 2 contains information on the two strata (groups) – food items (labeled 1) and non-food items (labeled 0, where we have included spices and household goods). Use Minitab’s “unstack” command to separate these two groups:

```
MTB> unstack c2 c5 c6;      #places the results into columns 5 and 6
SUBC> subs c3.             #indicates that column 3 has the group identifiers
```

Note: These steps can be reversed using Minitab’s “stack” command:

```
MTB> stack c5 c6 c7;       #places the results into column 7
SUBC> subs c8.             #keeps track of which group they came from.
```

After running the unstack command, how many items are there in each group (e.g., `MTB> info c5 c6`)? What percentage of the population has been classified as a food item?

(d) Does there appear to be a difference in the price distributions between these two groups? Support your answer with appropriate numerical and graphical summaries. [*Hint: Boxplots are very helpful here.*]

(e) Now write a Minitab macro to take 1000 stratified samples of 30 products, with the stipulation that each sample must have 5 non-food products (from C5) and 25 food products (from C6). [Note that this food/non-food breakdown roughly matches the percentage breakdown in the population.] The macro should store the resulting samples into C7 and C8, restack the samples into C9, and then calculate the mean for the combined sample and store the results in C11. Describe this sampling distribution and compare it to the one generated in (b).

(f) Which sampling method(s), the simple random samples or the stratified samples, was/were unbiased? Which had less variability? Support your conclusion numerically and graphically and explain why these answers make sense.

(g) Create your own grouping variable (e.g., the first 69 items and the second 70 items) and explore a stratified sampling method using these strata. Does this stratification appear to improve the precision of the sample means? Explain.