

Implementing random sampling before creating the contingency table is an excellent way to ensure that our data is unbiased and representative. Here's how we can approach random sampling for our categories (Furniture, Office Supplies, Technology) and customer segments (Corporate, Consumer, Home Office).

### **Steps for Random Sampling**

#### **1. Define the Population:**

- Determine the entire population from which we will draw our sample. This could be the total dataset of sales transactions that includes various product categories and customer segments.

#### **2. Stratified Sampling:**

- Since we want to ensure representation from each product category and customer segment, we can stratify our data based on these two variables. This involves dividing the population into distinct groups (strata) and then randomly selecting samples from each group.

#### **3. Determine Sample Size:**

- Decide on the desired sample size for each category and segment. The sample size should be large enough to meet the assumption of expected frequencies (at least 5 per cell in the contingency table).
- A common approach is to calculate a proportionate sample based on the sizes of the categories.

#### **4. Random Selection:**

- Use a random sampling method to select individuals from each stratum. This could be done using programming libraries such as pandas in Python.

#### **5. Check Sample Size:**

- After sampling, check the number of observations in each category and segment to ensure that they meet the expected frequency requirement for the Chi-squared test.