**1. PPS Sampling**

**Q1:** A town has 6 shops with annual sales (in thousand Taka): 120, 80, 140, 200, 100, 60.

* Select a sample of 3 shops using **PPS sampling with replacement**.
* Compute the **Horvitz-Thompson estimator** for the total sales.

**2. Multi-stage Cluster Sampling**

**Q2:** Suppose you have 4 schools, each with 3 classes, and each class has 10 students.

* Select 2 schools (first stage), then 1 class per selected school (second stage), then 5 students per selected class (third stage).
* Estimate the **average score** of students if each student has a math score generated randomly from N(50, 10²).

**3. Double Sampling**

**Q3:** Assume we want to estimate the **average household income**.

* First, take a simple random sample of 100 households to collect **auxiliary information** (e.g., household size).
* Then, take a subsample of 40 households to collect **income data**.
* Use **regression estimation** to improve the income estimate.

**4. Sampling Errors**

**Q4:** Simulate a population of 1000 people with a binary variable Smoker (1 = yes, 0 = no, probability = 0.3).

* Draw 50 SRS samples of size 100.
* For each sample, compute the proportion of smokers.
* Plot the distribution of sample proportions and compare it with the true population proportion.
* Calculate **bias and mean squared error (MSE)**.

**5. Census vs Survey**

**Q5:** Generate a population of 500 households with an expenditure variable (rnorm(500, mean=15000, sd=3000)).

* Compute the true mean expenditure (census).
* Take a SRS sample of size 50 and compute the sample mean.
* Repeat 1000 times and plot the distribution of sample means.
* Comment on how well surveys approximate census results.

*(Hint: Use replicate() and mean() in R.)*

*(Hint: Use sampling::ppswr() in R.)*

*(Hint: Use sample() function stage by stage.)*

*(Hint: Use lm() for regression estimator in R.)*

*(Hint: Use replicate() and hist() in R.)*