ST2132 Tutorial 4 Solution 2210

1. $\mu = \mu_1$. $\sum_{i=1}^N (v_i - \mu)^2 = \sum_{i=1}^N v_i^2 - 2\mu \sum_{i=1}^N v_i + N\mu^2 = \sum_{i=1}^N v_i^2 - N\mu^2$, since $\sum_{i=1}^N v_i = N\mu$. Hence $\sigma^2 = \mu_2 - \mu_1^2$.

2. Let $h(m) = v_m^2$. Since M has a uniform distribution, $\mathrm{E}(X^2) = \mathrm{E}\{h(M)\} = \sum_{i=1}^N v_i^2 \frac{1}{N} = \mu_2$. From Survey Sampling I, $\mathrm{E}(X) = \mu$ which is μ_1 from (a). Hence $\mathrm{var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2 = \mu_2 - \mu_1^2$ which is equal to σ^2 by (a).

   [Why is $\mathrm{var}(X) = \sigma^2$ new? In the first lecture, σ^2 was defined as $\mathrm{var}(X)$. But here, σ^2 is defined as the variance of the population, not X. Similarly, $\mu = \mathrm{E}(X)$ by definition, but now μ is the population mean. Hence $\mathrm{E}(X) = \mu$ and $\mathrm{var}(X) = \sigma^2$ are theorems.]

3. (a) $\Pr(M_1 = 1, \ldots, M_N = N) =$

   $$\Pr(M_1 = 1) \times \Pr(M_2 = 2|M_1 = 1) \times \Pr(M_3 = 3|M_1 = 1, M_2 = 2)\times$$
   $$\cdots \Pr(M_N = N|M_1 = 1, \ldots, M_{N-1} = N-1)$$
   $$= \frac{1}{N} \times \frac{1}{N-1} \times \frac{1}{N-2} \cdots \times 1 = \frac{1}{N!}$$

   [Note that the third probability conditions on both $M_1 = 1$ and $M_2 = 2$, not just on $M_2 = 2$. Similarly for subsequent factors.]

   (b) $\Pr(M_1 = \pi(1), \ldots, M_N = \pi(N)) =$

   $$\Pr(M_1 = \pi(1)) \times \Pr(M_2 = \pi(2)|M_1 = \pi(1)) \times \Pr(M_3 = \pi(3)|M_1 = \pi(1), M_2 = \pi(2))\times$$
   $$\cdots \Pr(M_N = \pi(N)|M_1 = \pi(1), \ldots, M_{N-1} = \pi(N-1))$$

   The conditional probabilities are exactly like (a), so the answer is also $\frac{1}{N!}$.

   [If the notation seems abstract, think of a concrete rearrangement, like $M_1 = N, M_2 = N-1, \ldots, M_N = 1$.]

   (c) From (b), $(M_1, \ldots, M_N)$ has the same probability to be any permutation of $1, \ldots, N$.

4. (a) Each of the 10,000 undergraduates consumed a number of bubble teas during the previous week. The mean of the 10,000 numbers is μ.

   (b) The 400 numbers are realisations from RV's $X_1, \ldots, X_{400}$ which are a simple random sample from the population. μ is estimated as 5.4.

   (c) SE is $\mathrm{SD}(\bar{X}) = \sigma/\sqrt{400}$, where σ is the population SD. By the bootstrap, estimated SE is $5.3/20 \approx 0.3$.

   (d) The assumption in (b) is invalid. μ is likely to be smaller than 5.4, but it is hard to say by how much.

[We assume that the population size is much largaer than 400, so that an SRS is like drawing with replacement. Consequent to the assumption in (b), the data $x_1, \ldots, x_{400}$ are realisations of the RV's, and 5.4 is a realisation of the random sampmle mean $\bar{X}$. The SD of 5.3 is calculated with 400 in the denominator, though it does not matter so much if 399 were used instead.]

5. (a) (ii) $\mathrm{E}(\hat{p}) = p$, which is unknown and estimated as 0.69.

   (b) (ii) $\mathrm{SE} = \mathrm{SD}(\hat{p}) = \sqrt{p(1-p)/1500}$, which is also unknown, and estimated as

$$\sqrt{\frac{0.69 \times 0.31}{1500}} \approx 0.01$$

   [We also assume here that the population size is much larger than 1500.]