Name: Ngu Yi Hui Course: SCOR
ID : 18ADB01438

UDPS2033 SAMPLE SURVEY AND SAMPLING TECHNIQUE

1. A town contains 500 housing areas, which could be classified as 250 middle income housing area, 150 low income housing area and the rest, affluent housing area. There is an average of 5 people in a house.

   We would like to conduct a survey to determine the prevalence of overweight/ obesity among adults in the town. Adults with Body Mass Index (BMI) above 25 kg/m^2 could be classifies as overweight, whereas BMI 30 kg/m^2 and above could be classified as being obese. BMI could be calculated as follows:

$$BMI = \frac{weight\ (kg)}{Height\ (m)^2}$$

   Previous studies showed that women, older adults aged 60 and above and those with high income are at the elevated risk of being overweight and obese.

   We could visit any house in any housing area since we have adequate research facilities and approval from the local authorities and people. We have made a sample size estimation and we need to interview 1000 adults to achieve the desired precision.

   Explain in detail how sampling should be conducted and list the measurements that should be taken. (10 marks)

   Stratified cluster sampling method should be performed in this case.

   Firstly, the population may be divided into 3 strata (middle income housing area, low income housing area, and affluent housing area). For the 3 strata, the housing areas will be randomly selected according to their proportions in population.

   For middle income housing area, we randomly select $\frac{250}{500}(20) = 10$ housing areas.
   For low income housing area, we randomly select $\frac{150}{500}(20) = 6$ housing areas.
   For middle income housing area, we randomly select $\frac{100}{500}(20) = 4$ housing areas.
   So, there will be total 20 housing areas selected in the 3 strata.

   Then, we randomly selected clusters of 10 houses (primary unit) from each housing area. All the people live in the house (secondary unit) are required to participate the survey. Since we have an average of 5 people in a house, we will finally interview around 1000 adults for the survey. 5 Marks

   Measurements that should be taken from the adults are gender, age, income, weight, and height. 2 Marks

2. A medication produces a certain chemical derivative in the blood two hours after consumption. A healthy human being usually has a minimal value of the derivative in the blood. The dosage injected (ml) and chemical derivative concentration (mg/dL) in the blood after two hours for seven patients is shown below.

Dosage (ml)	120	125	130	145	150	155	160
Derivative (mg/dL)	35	36	40	45	48	50	55

i) Explain with reasons the most suitable method to study the relationship between dosage given and the derivative concentration. (4 marks)

Regression Estimation is the most suitable method to study the relationship between dosage given and the derivative concentration.

Reasons:

1. Dosage (x) & derivative (y) are highly correlated with correlation coefficient = 0.989.
2. Dosage (x) contributes information for prediction of derivative (y).
3. A healthy human being usually has a minimal value of the derivative (y) in the blood, indicates that the relationship between derivative (y) and dosage (x) is not a linear through origin.

3 Marks

ii) If the mean dosage given to all adult patients treated is 148 ml, estimate the mean concentration of the derivative in the blood for the patient population. Give a bound to the estimate using α=0.05. (6 marks)

Regression Estimation:

$$b = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{694.286}{1471.43} = 0.4718$$

$$\hat{\mu}_{yL} = \bar{y} + b(\mu_x - \bar{x}) = 44.143 + 0.4718(148 - 140.714) = 47.581$$

Since N is not mentioned in this question, we assume it as a very large value. Thus, finite population correction (FPC) can be ignore.

$$V(\hat{\mu}_y) = \left(\frac{1}{n}\right)\left(\frac{1}{n-2}\right)\left[\sum_{i=1}^{n}(y_i-\bar{y})^2 - b^2\sum_{i=1}^{n}(x_i-\bar{x})^2\right]$$

$$= \left(\frac{1}{7}\right)\left(\frac{1}{5}\right)[334.8571 - (0.471845)^2(1471.4286)]$$

$$= 0.2075$$

$$B = 2\sqrt{V(\hat{\mu}_y)} = 2\sqrt{0.2075} = 0.911$$

∴ The mean concentration of the derivative in the blood for the patient population = 47.581 mg/dL. The bound to the estimate = 0.911.

6 Marks

3. Kathy received 400 products in boxes of 20 products on Monday. She wanted to determine the proportion of deformed products and decides to conduct a survey instead of inspecting all elements. The rate of deformities in five randomly selected boxes are as follows:

Box	1	2	3	4	5
Rate of deformity	0.1	0.15	0.3	0.1	0.25

i) Estimate the number of boxes she should inspect if she wants to make an estimate with error of estimation of 0.01 at 95% confidence level. (5 marks)

$M = 400$

$N = \frac{400}{20} = 20$

$\bar{M} = \frac{400}{20} = 20$

$B = 0.01$

$D = \frac{B^2\bar{M}^2}{4} = \frac{(0.01)^2(20)^2}{4} = 0.01$

$\sigma_p{}^2 = s_p{}^2 = \frac{\sum_{i=1}^{n}(a_i - \hat{p}m_i)^2}{n-1} = \frac{13.2}{4} = 3.3$

$n = \frac{N\sigma_p{}^2}{ND + \sigma_p{}^2} = \frac{(20)(3.3)}{(20)(0.01)+3.3} = 18.86 \simeq 19$

∴ The number of boxes Kathy should inspect = 19 boxes.

5 Marks

ii) She thinks that the sample size estimated is too large and will take so much of her time. Suggest ways to reduce the sample size. (2 marks)

Ways to reduce sample size:
1. Reduce the alpha level to 0.1
2. Increase the error of estimation

2 Marks

iii) In your opinion what are steps that should be taken while estimating sample size so that the estimation is not "manipulated" to be too little just to make ones work easy and consequently suffer lack of power. (3 marks)

We should design the research method carefully and obtain the appropriate sampling frame that covers the whole population so that the estimation is not "manipulated" to be too little just to make ones work easy. In addition, a higher significant level will help increasing the power. Therefore, we can set a higher significant level to avoid lack of power.

1 Mark

Question 1: 7 Marks
Question 2: 9 Marks
Question 3: 8 Marks
TOTAL: 24/30
CLO3:7
CLO4: 9
CLO5:8