

Homework

Bùi Khánh Duy

2023-04-19

7.1

Trong ví dụ này, một mẫu hệ thống sẽ phù hợp hơn so với một mẫu ngẫu nhiên đơn giản.

Một mẫu hệ thống bao gồm việc chọn mỗi phần tử thứ k từ một danh sách hoặc chuỗi. Trong trường hợp này, các khoản vay thể chấp được đánh số theo thứ tự, có nghĩa là chúng đã được sắp xếp theo một chuỗi. Bằng cách chọn mỗi khoản vay thứ k , một mẫu đại diện có thể được thu thập.

Tuy nhiên, một mẫu ngẫu nhiên đơn giản bao gồm việc chọn một tập con ngẫu nhiên các phần tử từ toàn bộ quần thể. Mặc dù phương pháp này cũng có thể hiệu quả, nhưng nó có thể không hiệu quả bằng một mẫu hệ thống trong tình huống này. Điều này là do một mẫu ngẫu nhiên có thể không bao gồm một mẫu đại diện của các khoản vay được cấp trong vòng 20 năm. Hơn nữa, việc chọn một mẫu ngẫu nhiên có thể tốn thời gian và tài nguyên, đặc biệt là nếu số lượng khoản vay lớn.

Do đó, một mẫu hệ thống sẽ phù hợp hơn để ước tính tổng số tiền còn nợ trong tình huống này, vì nó hiệu quả hơn và đại diện cho toàn bộ các khoản vay.

7.2

Trong bài này, phương pháp lấy mẫu phân lớp sẽ là phương pháp phù hợp nhất để ước tính thu nhập trung bình của mỗi nhân viên.

Phương pháp lấy mẫu phân lớp bao gồm chia quần thể thành các nhóm con dựa trên các đặc điểm nhất định (trong trường hợp này là các đối tượng có thu nhập tương đương), sau đó lấy một mẫu ngẫu nhiên từ mỗi nhóm con. Bằng cách làm như vậy, mẫu sẽ đại diện cho quần thể và có thể dẫn đến ước tính chính xác hơn về thu nhập trung bình của mỗi nhân viên.

Phương pháp lấy mẫu hệ thống bao gồm việc chọn mỗi phần tử thứ k từ một danh sách hoặc chuỗi. Mặc dù phương pháp này có thể hiệu quả trong một số trường hợp, nhưng nó có thể không hiệu quả hoặc đại diện như phương pháp phân lớp trong tình huống này. Điều này là do các nhân viên được liệt kê theo thứ tự chữ cái trong từng nhóm thu nhập, vì vậy việc lấy mẫu hệ thống có thể không bao phủ toàn bộ phạm vi các mức thu nhập trong mỗi nhóm.

Phương pháp lấy mẫu ngẫu nhiên đơn giản bao gồm việc chọn một tập con ngẫu nhiên các phần tử từ toàn bộ quần thể. Mặc dù phương pháp này cũng có thể hiệu quả, nhưng nó có thể không hiệu quả như phương pháp phân lớp trong tình huống này. Điều này là do phương pháp lấy mẫu ngẫu nhiên đơn giản có thể không bao phủ được sự biến động của các mức thu nhập khác nhau trong từng nhóm thu nhập.

Các ưu điểm của phương pháp phân lớp là nó có thể dẫn đến một mẫu đại diện hơn và ước tính chính xác hơn về tham số của quần thể được quan tâm (trong trường hợp này là thu nhập trung bình của mỗi nhân viên). Tuy nhiên, những nhược điểm của phương pháp này là nó có thể phức tạp và tốn thời gian để thực hiện, và nó yêu cầu có kiến thức trước về các đặc điểm của quần thể (trong trường hợp này là các nhóm thu nhập). Các ưu điểm của phương pháp lấy mẫu hệ thống là nó có thể hiệu quả và dễ dàng để thực hiện. Tuy nhiên, những nhược điểm của phương pháp này là nó có thể không bao phủ toàn bộ sự biến động trong từng nhóm thu nhập và có thể không đại diện cho quần thể.

Các ưu điểm của phương pháp lấy mẫu ngẫu nhiên đơn giản là nó đơn giản và dễ dàng để thực hiện. Tuy nhiên, những nhược điểm của phương pháp này là nó có thể không đại diện cho quần thể và có thể không bao phủ toàn bộ sự biến động trong từng nhóm thu nhập.

Tóm lại, nên chọn phương pháp lấy mẫu phân lớp.

7.3

- a) Liệt kê tất cả cách chọn mẫu hệ thống 1-in-10 và tính toán chính xác variance cho từng mẫu. (Có 10 giá trị, có thể trùng nhau, với tỉ lệ 1/10 xuất hiện).

Các giá trị:

n = kích cỡ của từng mẫu

N = tổng kích cỡ

Các công thức:

$$\hat{p} = m/n$$

$$var(\hat{p}) = (1 - \frac{n}{N}) * \frac{\hat{p}*(1-\hat{p})}{n-1}$$

Các bộ mẫu:

Mẫu thứ 1: {1, 11, 21, 31}

$$\hat{p} = 0.75$$

$$var(\hat{p}) = 0.05625$$

Mẫu thứ 2: {2, 12, 22, 32}

$$\hat{p} = 1.0$$

$$var(\hat{p}) = 0.0$$

Mẫu thứ 3: {3, 13, 23, 33}

$$\hat{p} = 0.75$$

$$var(\hat{p}) = 0.05625$$

Mẫu thứ 4: {4, 14, 24, 34}

$$\hat{p} = 0.75$$

$$var(\hat{p}) = 0.05625$$

Mẫu thứ 5: {5, 15, 25, 35}

$$\hat{p} = 0.25$$

$$var(\hat{p}) = 0.05625$$

Mẫu thứ 6: {6, 16, 26, 36}

$$\hat{p} = 0.0$$

$$var(\hat{p}) = 0.0$$

Mẫu thứ 7: {7, 17, 27, 37}

$$\hat{p} = 0.0$$

$$var(\hat{p}) = 0.0$$

Mẫu thứ 8: {8, 18, 28, 38}

$$\hat{p} = 0.0$$

$$var(\hat{p}) = 0.0$$

Mẫu thứ 9: {9, 19, 29, 39}

$$\hat{p} = 0.25$$

$$var(\hat{p}) = 0.05625$$

Mẫu thứ 10: {10, 20, 30, 40}

$$\hat{p} = 0.25$$

$$var(\hat{p}) = 0.05625$$

b) Liệt kê tất cả cách chọn mẫu hệ thống 1-in-5 và tính toán chính xác variance cho từng mẫu.

Mẫu thứ 1: {1, 6, 11, 16, 21, 26, 31, 36}

$$\hat{p} = 0.375$$

$$var(\hat{p}) = 0.02678571428571429$$

Mẫu thứ 2: {2, 7, 12, 17, 22, 27, 32, 37}

$$\hat{p} = 0.5$$

$$var(\hat{p}) = 0.028571428571428574$$

Mẫu thứ 3: {3, 8, 13, 18, 23, 28, 33, 38}

$$\hat{p} = 0.375$$

$$var(\hat{p}) = 0.02678571428571429$$

Mẫu thứ 4: {4, 9, 14, 19, 24, 29, 34, 39}

$$\hat{p} = 0.5$$

$$var(\hat{p}) = 0.028571428571428574$$

Mẫu thứ 5: {5, 10, 15, 20, 25, 30, 35, 40}

$$\hat{p} = 0.25$$

$$var(\hat{p}) = 0.021428571428571432$$

7.4

$$n = 200$$

$$N = 2000$$

$$\hat{p} = \frac{\sum_{i=1}^{200} y_i}{n} = \frac{132}{200} = 0.66$$

$$var(\hat{p}) = (1 - \frac{n}{N}) * \frac{\hat{p} * (1 - \hat{p})}{n - 1} = (1 - \frac{1}{10}) * \frac{0.66 * 0.34}{199} \approx 0.0101$$

$$\Rightarrow \hat{p} - 2 * \sqrt{var(\hat{p})} < p < \hat{p} + 2 * \sqrt{var(\hat{p})}$$

$$\Rightarrow 0.5963 < p < 0.7237$$

7.5

$$B_0 = 0.01$$

Coi $p = \hat{p}$ như ở bài 7.4

$$p = 0.66q = 1 - p = 0.34n_0 = \frac{Npq}{(N-1) * \frac{B_0^2}{4} + pq} = 1635.718$$

=> Cần phải lấy ít nhất 1636 số nhân viên => cần chọn mẫu 1-1.22

7.6

```
data = c(12.0, 11.97, 12.01, 11.91, 11.98, 12.03, 11.87, 12.01, 11.98, 12.05, 11.87, 11.91, 11.75, 11.9)
```

```
N = 1800
```

```
n = 1800/50
```

```
hat_muy = sum(data)/n  
hat_muy
```

```
## [1] 11.94556
```

```
s_2 = var(data)  
s_2
```

```
## [1] 0.005808254
```

```
var_muy = (1-n/N)*(s_2/(n-1))  
var_muy
```

```
## [1] 0.0001626311
```

```
bound = 2*sqrt(var_muy)  
bound
```

```
## [1] 0.02550538
```

$$\Rightarrow \hat{\mu} - 2\sqrt{\text{var}(\hat{\mu})} < \mu < \hat{\mu} + 2\sqrt{\text{var}(\hat{\mu})}$$

$$\Rightarrow 11.92 < \mu < 11.97106$$

7.7

$$B_0 = 0.03$$

Vì σ chưa biết nên dùng s để thay thế.

$$n = \frac{N * s^2}{(N - 1) * \frac{B_0^2}{4} + s^2} = \frac{1800 * 0.005808254}{1799 * \frac{0.03^2}{4} + 0.005808254} = 25.46$$

Vậy cần lấy $n = 26$

7.8

$$\hat{\mu} = \frac{\sum y_i}{n} = \frac{90320}{45} = 2007.111111111$$

$$var(\hat{\mu}) = (1 - \frac{n}{N}) \frac{s^2}{n} = (1 - \frac{1}{32}) \frac{250^2}{45} = 1345.486$$

$$Bound = 2 * \sqrt{var(\hat{\mu})} = 2 * \sqrt{1345.486} = 73.3617338944$$

$$\Rightarrow 1,933.75941 < \mu < 2,080.47281$$

7.9

$$\hat{p} = \frac{\sum_{i=1}^{400} y_i}{400} = \frac{324}{400} = 0.81$$

$$var(\hat{p}) = (1 - \frac{n}{N}) \frac{\hat{p} * (1 - \hat{p})}{n - 1} = (1 - \frac{400}{2800}) * (0.81 * 0.19) / 399 = 0.0003306$$

$$Bound = 2 * \sqrt{var(\hat{p})} = 0.03637$$

7.10

Coi $p = \hat{p} = 0.81, q = 1 - p = 0.19$

$$n = \frac{Npq}{(N - 1) * \frac{B_0^2}{4} + pq} = \frac{3000 * 0.81 * 0.19}{2999 * \frac{0.015^2}{4} + 0.81 * 0.19} = 1431.2119$$

$\Rightarrow n = 1432$