

Chương 6: CHỌN MẪU NGHIÊN CỨU VÀ THIẾT KẾ BẢN CÂU HỎI

6.1. Chọn mẫu nghiên cứu

6.1.1. Lý do chọn mẫu

6.1.1.1. Một số định nghĩa

- Phần tử:

Một phần tử là một đơn vị trong đó thông tin về nó được thu thập và làm cơ sở cho việc phân tích. Thông thường trong lấy mẫu nghiên cứu marketing, những phần tử là con người, tuy vậy cũng có những loại phần tử khác như là: gia đình, cửa hàng hoặc doanh nghiệp.

- Tổng thể

Một tổng thể là sự tập hợp các phần tử. có 2 loại tổng thể:

- Tổng thể chủ đích (target population): Là một tổng thể được yêu cầu bởi đặc trưng thông tin cần nghiên cứu.

- Tổng thể lấy mẫu (sampling population): Là một tổng thể thực tế được chọn trên yêu cầu thông tin cần nghiên cứu.

- Cấu trúc mẫu (sampling frames)

Cấu trúc mẫu là một danh sách các phần tử lấy mẫu. Ví dụ: yêu cầu đánh giá trình độ trung bình của sinh viên năm thứ 4 của một trường đại học. Tổng thể chủ đích là tất cả sinh viên đang học năm thứ 4 đã theo học từ năm đầu tiên. Tuy nhiên, một số sinh viên đã bỏ học vì chuyển sang trường khác, hoặc vì lý do nào đó. Số sinh viên còn lại là tổng thể lấy mẫu. Danh sách các sinh viên này là cấu trúc mẫu. Mỗi sinh viên trong danh sách là một phần tử lấy mẫu.

6.1.1.2. Những lí do của việc chọn mẫu

Trong nghiên cứu marketing nói riêng và trong các lĩnh vực nghiên cứu khác nói chung, việc lấy mẫu để điều tra thay vì phải điều tra toàn bộ được thực hiện bởi lý do sau:

- Những người ra quyết định thường bị giới hạn về mặt thời gian, do đó họ phải dựa vào bất kỳ thông tin nào có thể dùng được trong thời gian đó.

- Đối với qui mô tổng thể nghiên cứu lớn, chi phí cho một cuộc điều tra toàn bộ rất lớn, sẽ gặp hạn chế về kinh phí. Vì vậy việc điều tra trên một mẫu sẽ có ưu thế hơn nhưng vẫn bảo đảm thu thập đầy đủ thông tin thích hợp.

- Trong một số trường hợp, việc tiến hành điều tra toàn bộ tổng thể vẫn không thể nâng cao độ chính xác của thông tin trong khi lại tốn kém chi phí và mất nhiều thời gian.

- Trong những tình huống mà việc kiểm tra, đo lường có thể phá hủy phần tử thì việc lấy mẫu là điều hiển nhiên. Ví dụ: kiểm tra các phim chụp ảnh...

6.1.2. Quy trình chọn mẫu

Số lượng những nhiệm vụ và những quyết định trong lấy mẫu, cũng như để hiểu được chúng, chúng ta cần xem xét chúng trong tổng thể của quy trình chọn mẫu. Do đó, chúng ta cần xem xét những giai đoạn khác nhau của việc chọn mẫu.

Bước 1. Xác định tổng thể trong đó một mẫu được lấy ra

Trong quá trình thiết kế nghiên cứu, nhà nghiên cứu đã xác định nguồn dữ liệu (đối tượng cần thu thập dữ liệu). Như vậy, nhà nghiên cứu đã xác định được tổng thể cần nghiên cứu cho dự án nghiên cứu.

Ví dụ. Kiểm định lý thuyết về hành vi tiêu dùng đối với người tiêu dùng sản phẩm bia tuổi từ 18-45 tại thành phố Hồ Chí Minh thì chúng ta đã xác định được đám đông nghiên cứu gồm tất cả người tiêu dùng sản phẩm bia tuổi từ 18-45 tại thành phố Hồ Chí Minh.

Bước 2. Thiết lập “Khung” cho tổng thể đó

Sau khi xác định được đám đông nghiên cứu thì công việc tiếp theo là xác định khung mẫu. Trong ví dụ trên, là danh sách liệt kê các người tiêu dùng bia tại thành phố Hồ Chí Minh có độ tuổi từ 18-45 cùng với thông tin cá nhân cần thiết cho mẫu như tên gọi, địa chỉ, độ tuổi, ..Như vậy trong quá trình chọn mẫu, người tiêu dùng nào thuộc vào mẫu thì nhà nghiên cứu có thể xác định và tiếp cận được họ để thu thập dữ liệu.

Bước 3. Chọn phương pháp lấy đơn vị mẫu

Có nhiều phương pháp chọn mẫu, chúng được chia thành hai nhóm chính bao gồm (1) Các phương pháp chọn mẫu theo xác suất thường gọi là chọn mẫu ngẫu nhiên, (2) Các phương pháp chọn mẫu không theo xác suất thường gọi là phi xác suất hay không

ngẫu nhiên. Trong kiểm định lý thuyết khoa học, để quyết định chọn mẫu theo phương pháp nào nhà nghiên cứu phải xem xét nhiều yếu tố như mục tiêu nghiên cứu, tính tổng quát hóa của kết quả nghiên cứu, thời gian và chi phí.

Bước 4. Xác định kích thước mẫu cần thiết

Kích thước mẫu là công việc không dễ dàng trong nghiên cứu khoa học. Kích thước cần cho nghiên cứu phụ thuộc các yếu tố như phương pháp xử lý, độ tin cậy cần thiết,...Chúng ta đã biết kích thước mẫu càng lớn càng tốt nhưng lại tốn chi phí và thời gian. Hiện nay, các nhà nghiên cứu xác định kích thước mẫu cần thiết thông qua các công thức kinh nghiệm cho từng phương pháp xử lý (chúng ta có thể tiếp cận các công thức kinh nghiệm qua sự phổ biến của nó trong các bài báo khoa học).

Bước 5. Viết các chỉ dẫn để nhận ra và chọn các phần tử thật của mẫu

Nếu chọn mẫu bằng phương pháp chọn mẫu theo xác suất thì sau khi chọn các phần tử cho mẫu, chúng ta phải tiến hành đánh dấu vị trí các phần tử trong mẫu để tổ chức và quản lý việc phỏng vấn. Phỏng vấn viên không được thay đổi phần tử mẫu đã được xác định. Trong trường hợp chọn mẫu bằng phương pháp chọn mẫu phi xác suất phỏng vấn viên được tự do thay thế phần tử tham gia vào mẫu, miễn sao cho phần tử đó thỏa mãn các tính chất cần có cho phương pháp chọn mẫu phi xác suất.

6.1.3. Phương pháp chọn mẫu

6.1.3.1. Phương pháp chọn mẫu phi xác suất

- Chọn mẫu thuận tiện

Theo cách chọn mẫu này, người nghiên cứu chọn ra các đơn vị lấy mẫu dựa vào “sự thuận tiện” hay “tính dễ tiếp cận”. Với phương pháp chọn mẫu thuận tiện, rất khó xác định tính đại diện của mẫu. Sự lựa chọn các đơn vị mẫu mang tính chủ quan của người nghiên cứu, vì thế độ chính xác và độ tin cậy không cao, ít được sử dụng rộng rãi.

- Chọn mẫu tích lũy nhanh

Theo phương pháp này, những đơn vị lấy mẫu (hay phần tử) ban đầu được lựa chọn bằng cách sử dụng các phương pháp xác suất, nhưng những đơn vị bổ sung tiếp đó được xác định từ thông tin được cung cấp bởi các đơn vị lấy mẫu ban đầu (quy nguyên). Dù phương pháp xác suất nào được sử dụng để lựa chọn những đơn vị lấy mẫu ban đầu,

thì toàn bộ mẫu vẫn được coi là mẫu phi xác suất vì những quy nguyên theo sau được chứa đựng trong mẫu ấy.

Kích thước mẫu và thời gian hao phí giảm đi là những thuận lợi chủ yếu của kỹ thuật lấy mẫu tích lũy nhanh. Tuy nhiên cách chọn mẫu “nhờ giới thiệu” này có thể có sai lệch vì những người được giới thiệu ra thường có một số đặc điểm tương đồng về nhân khẩu học hay tâm lý, sở thích. Do đó, phương pháp này chỉ được sử dụng khi các phần tử mà chúng ta muốn nghiên cứu rất khó tìm.

- Chọn mẫu phán đoán

Theo phương pháp chọn mẫu phán đoán, những đơn vị của mẫu được chọn dựa vào điều mà nhà chuyên môn suy nghĩ có thể thỏa mãn một tiêu chuẩn nào đó. Có hai hình thức lựa chọn phán đoán: lấy mẫu theo dư luận và phán đoán thống kê.

- Chọn mẫu kiểm tra tỷ lệ

Chọn mẫu kiểm tra tỷ lệ là phương pháp chọn mẫu mà trong đó người nghiên cứu cố gắng bảo đảm mẫu được lựa chọn có một tỷ lệ tương ứng với tỷ lệ tổng thể theo các tham số quan trọng nào đó (tuổi tác, giới tính, nghề nghiệp...). Các phần tử trong mẫu cũng được chọn theo chủ ý của người nghiên cứu chứ không phải dựa vào quy luật ngẫu nhiên. Chẳng hạn, nếu xác định kích thước mẫu cần điều tra là 100, và giới tính là một tham số quan trọng đối với nội dung điều tra (chẳng hạn việc sử dụng kẹo sôcôla); khi đó, nếu biết được tỷ lệ giới tính nữ- nam của tổng thể là 51:49 (tỷ lệ bách phân) thì mẫu được chọn sẽ có 51 nữ và 49 nam. Đây là một ví dụ đơn giản; trong thực tế, tùy thuộc nội dung điều tra, người ta xác định tỷ lệ theo nhiều tham số: tuổi tác - giới tính - thu nhập...

6.1.3.2. Phương pháp chọn mẫu xác suất

Phương pháp chọn mẫu xác suất thực hiện việc chọn các phần tử của mẫu dựa trên việc sử dụng các quy luật phân phối xác suất trong thống kê toán. Tuy nhiên, trong phần này do đối tượng nghiên cứu của môn học nên không trình bày tỉ mỉ như trong thống kê học, mà chủ yếu giới thiệu phương pháp để trên cơ sở đó có thể lựa chọn phương pháp chọn mẫu thích hợp cho từng mục tiêu nghiên cứu marketing.

- Chọn mẫu ngẫu nhiên đơn giản

Chọn mẫu ngẫu nhiên là một quá trình chọn lựa mẫu sao cho mỗi đơn vị lấy mẫu trong cấu trúc có một cơ hội hiện diện trong mẫu bằng nhau.

Chọn mẫu ngẫu nhiên có hai loại: chọn mẫu ngẫu nhiên có sự thay thế hoặc là không có sự thay thế. Trong lấy mẫu ngẫu nhiên có sự thay thế thì một phần tử đã được chọn luôn luôn được thay thế trước khi thực hiện sự lựa chọn kế tiếp. Cách này có khả năng lấy trên cùng một cá thể nhiều lần. Do vậy, trong nghiên cứu marketing, lấy mẫu ngẫu nhiên không thay thế được sử dụng chủ yếu.

Ưu điểm của phương pháp chọn mẫu ngẫu nhiên đơn giản là dễ hiểu, dễ thực hiện; trung bình mẫu là một sự tính toán khách quan của trung bình tổng thể nghiên cứu; phương pháp tính toán đơn giản, dễ dàng.

Nhược điểm của phương pháp chọn mẫu ngẫu nhiên:

- + Trong nhiều trường hợp, sự biến thiên của tổng thể nghiên cứu rất rời rạc và không theo quy tắc, thì lấy mẫu ngẫu nhiên không được dùng đến vì nó kém chính xác; mẫu có thể không mang tính đại diện hoặc bị lệch.

- + Để lựa chọn các phần tử, cần phải đánh dấu và lập danh sách toàn bộ tổng thể để sử dụng bảng số ngẫu nhiên, bốc thăm, quay số,... công việc này khó thực hiện được khi tổng thể là quá lớn.

- + Mẫu được chọn có thể bị phân tán, do vậy tốn kém chi phí và khó khăn trong đi lại khi thu thập dữ liệu.

Phương pháp chọn mẫu ngẫu nhiên đơn giản được áp dụng có kết quả khi tổng thể nghiên cứu không phân tán quá rộng về mặt địa lý; các phần tử trong tổng thể có khá nhiều sự đồng nhất về đặc điểm muốn nghiên cứu.

- Chọn mẫu ngẫu nhiên phân tầng

Khi tổng thể nghiên cứu được cấu tạo bởi nhiều tập hợp không đồng nhất liên quan đến những đặc điểm nghiên cứu, để thực hiện lấy mẫu cần phải phân tầng tổng thể nghiên cứu thành từng nhóm có những đặc điểm tương đồng. Lấy mẫu phân tầng là chọn một mẫu ngẫu nhiên đơn giản từ mỗi nhóm trong tổng thể nghiên cứu.

Tùy theo đặc điểm nghiên cứu, tổng thể có thể được phân tầng theo nhiều tiêu thức khác nhau; và có thể phân tầng một cấp (một tiêu thức) hoặc nhiều cấp (nhiều tiêu thức);

và khi chọn mẫu ngẫu nhiên phân tầng có thể theo tỷ lệ (tỷ lệ mẫu tương ứng với tỷ lệ tổng thể) hoặc không theo tỷ lệ.

Ưu điểm của phương pháp chọn mẫu ngẫu nhiên phân tầng: sự phân nhóm có thể làm gia tăng mức độ chính xác của việc đánh giá các đặc điểm tổng thể nghiên cứu; thực hiện thuận tiện, phân tích số liệu khá toàn diện.

Nhược điểm của phương pháp này là cần phải lập danh sách các đơn vị lấy mẫu theo từng nhóm; tốn kém chi phí đi lại, đặc biệt khi tổng thể nghiên cứu trải rộng trên một vùng địa lý rộng lớn.

Với những ưu điểm và nhược điểm trên, phương pháp chọn mẫu ngẫu nhiên phân tầng thường được áp dụng khi tổng thể nghiên cứu có sự phân bố của đặc điểm nghiên cứu rất rời rạc, hay tập trung trên những điểm nhỏ bị phân tán của tổng thể.

- Chọn mẫu có hệ thống

Chọn mẫu có hệ thống với sự bắt đầu ngẫu nhiên là một phương pháp chọn mẫu được tiến hành bằng cách lấy từng đơn vị thứ k từ một tổng thể nghiên cứu có thứ tự. Đơn vị đầu tiên được chọn một cách ngẫu nhiên, k được gọi là khoảng cách lấy mẫu, số nghịch đảo $1/k$ là tỷ lệ lấy mẫu.

Ưu điểm của phương pháp chọn mẫu có hệ thống là mẫu được thiết lập dễ dàng, dễ thực hiện trên hiện trường (điều tra theo đường phố), mẫu được phân tán đều khắp tổng thể nghiên cứu và kết quả tính toán chính xác hơn so với lấy mẫu ngẫu nhiên đơn giản.

Nhược điểm của phương pháp chọn mẫu này là: có thể một mẫu được lấy chỉ bao gồm những đơn vị có cùng một dạng, và cần thiết phải có danh sách các đơn vị lấy mẫu theo thứ tự. Vì thế phương pháp chọn mẫu có hệ thống thường được áp dụng khi thứ tự của các đơn vị lấy mẫu là ngẫu nhiên, gần như có sự phân nhóm trong tổng thể nghiên cứu.

- Chọn mẫu theo cụm

Chọn mẫu theo cụm là phương pháp chọn mẫu được tiến hành bằng cách lấy những nhóm riêng biệt hoặc những cụm của những đơn vị nhỏ hơn. Những cụm của mẫu có thể được chọn bằng cách lấy mẫu ngẫu nhiên hay lấy mẫu có hệ thống với một sự khởi đầu ngẫu nhiên.

Tương tự với nhóm trong lấy mẫu ngẫu nhiên phân tầng, các cụm là các nhóm phụ riêng biệt với nhau cùng tạo nên tổng thể nghiên cứu. Tuy nhiên, không giống như nhóm, các cụm được tạo nên bởi những phần tử dị biệt, không đồng nhất, miễn sao mỗi nhóm sẽ là đặc trưng của tổng thể nghiên cứu. Ví dụ, nghiên cứu về sinh viên trong một trường đại học, thay vì chọn các phần tử là sinh viên theo kích thước mẫu, có thể chọn đơn vị lấy mẫu là lớp; do vậy không cần phải lập danh sách sinh viên, mà lập danh sách các lớp. Khi thực hiện điều tra, thì tất cả sinh viên trong một lớp được chọn đều được tiếp xúc.

Ưu điểm của phương pháp chọn mẫu theo cụm là không cần thiết phải xây dựng một danh sách tất cả các phần tử trong tổng thể nghiên cứu, mà cấu trúc đối với lấy mẫu theo cụm là một danh sách các cụm. Ngay cả khi danh sách các phần tử đã có sẵn, việc lấy mẫu theo cụm vẫn ít tốn kém hơn về chi phí.

Nhược điểm của phương pháp này là ở chỗ trong thực tế, lấy mẫu theo cụm không hiệu quả bằng lấy mẫu ngẫu nhiên hay phân tầng. Chẳng hạn, những hộ gần kề nhau thường có đặc điểm tương tự nhau hơn những hộ riêng biệt. Điều này sẽ ảnh hưởng đến tính đại diện của mẫu và được thể hiện qua sai sót chọn mẫu tăng.

Phương pháp chọn mẫu theo cụm được áp dụng khi danh sách đầy đủ các phần tử trong tổng thể nghiên cứu không có sẵn, hoặc khi chi phí điều tra thấp được xem là quan trọng hơn so với yêu cầu về sự chính xác.

- Chọn mẫu nhiều giai đoạn

Việc chọn mẫu được thực hiện qua hai hay nhiều giai đoạn. Trước hết, tổng thể nghiên cứu được phân ra thành những đơn vị của giai đoạn đầu tiên, từ đó tiến hành chọn mẫu và sau đó có thể tăng thêm nhiều giai đoạn qua việc phân chia tổng thể nghiên cứu thành nhiều cấp bậc của những đơn vị lấy mẫu tương ứng với mỗi giai đoạn lấy mẫu khác nhau.

Ví dụ: Giả định muốn chọn một mẫu 30 hộ từ một thành phố nào đó và thành phố được phân thành 10 khu phố, mỗi một khu phố có 10 hộ. Cách làm như sau:

- (1) Đánh số các khu phố từ 1 - 10.
- (2) Dùng bảng số ngẫu nhiên lấy 5 số ngẫu nhiên; tương ứng là các khu phố được chọn.

(3) Đánh số thứ tự các hộ, liên tục từ 1 đến 10 trong cấu trúc của các khu phố được chọn.

(4) Lấy 5 nhóm số ngẫu nhiên, với mỗi nhóm gồm 6 số trong từng khu phố được chọn.

(5) Chọn 6 hộ theo nhóm số ngẫu nhiên ban đầu; sau đó chọn 6 hộ ở nhóm số ngẫu nhiên tiếp theo ...đến khi đủ 30.

Ưu điểm của phương pháp lấy mẫu này là có tính hiệu quả và linh hoạt hơn lấy mẫu một giai đoạn. Ngoài trừ những đơn vị của giai đoạn thứ nhất, cấu trúc mẫu chỉ yêu cầu đối với những đơn vị đã chọn để lấy những đơn vị phụ.

Nhược điểm: Lý thuyết phức tạp khi áp dụng trên hiện trường; qui trình tính toán khó khăn cho những người không phải là chuyên viên thống kê.

Phương pháp lấy mẫu theo nhiều giai đoạn được áp dụng khi danh sách của các đơn vị lấy mẫu không có sẵn, tổng thể nghiên cứu trải rộng trên vùng rộng lớn.

* Những lưu ý đối với việc chọn mẫu

- Hiệu quả lấy mẫu

Một mẫu được coi là có hiệu quả thống kê hơn một mẫu khác (theo phương pháp chọn mẫu khác nhau) khi có cùng một kích thước mẫu nhưng sai số trung bình nhỏ hơn. Nói cách khác, một mẫu có hiệu quả thống kê hơn, khi ở mức chính xác đã cho (sai số tiêu chuẩn) thì có kích thước mẫu nhỏ hơn.

Một mẫu được coi là có hiệu quả kinh tế hơn mẫu khác, khi với một độ chính xác mong muốn, phí tổn điều tra là thấp hơn. Một cách lý tưởng, người nghiên cứu luôn muốn đạt được độ chính xác cao với một phí tổn thấp. Tuy nhiên, trên thực tế, sự chính xác có liên quan kích thước mẫu, nghĩa là sự chính xác càng lớn thì phí tổn càng cao.

- Kích thước mẫu trong chọn mẫu phi xác suất

Quyết định về kích thước mẫu trong chọn mẫu phi xác suất thường được xác định một cách chủ quan chứ không dựa theo công thức tính toán như chọn mẫu xác suất. Người nghiên cứu quyết định kích thước mẫu mà theo cảm tính của họ là đại diện cho tổng thể. Trong nhiều cuộc nghiên cứu, sự hạn chế về tài chính là yếu tố quan trọng nhất đối với việc xác định kích thước mẫu thích hợp.

- Khảo sát một tham số và khảo sát nhiều tham số

Khi xác định kích thước mẫu trong phương pháp xác suất, điều quan trọng là lựa chọn tham số nào để tính toán, bởi vì kích thước mẫu được xác định theo tham số này chưa chắc chắn là phù hợp với tham số khác.

Trên thực tế, khi sử dụng bản câu hỏi, người nghiên cứu thường chọn một vấn đề mà họ cho là quan trọng nhất (vấn đề then chốt của cuộc nghiên cứu), và dựa trên vấn đề đó, sẽ xác định kích thước mẫu tương xứng.

- Vấn đề về tỷ lệ trả lời

Hiếm có một cuộc điều tra nào mà số lượng bản câu hỏi gửi đi (theo kích thước mẫu được chọn) bằng với số lượng bản câu hỏi thu về, tức là tỷ lệ không trả lời luôn luôn tồn tại, và do vậy sẽ làm giảm kích thước mẫu, tương ứng là sai số gia tăng. Sự không trả lời có thể do là bản câu hỏi không đến được nơi nhận hoặc người nhận được bản câu hỏi từ chối trả lời. Chính vì thế, các nỗ lực nhằm gia tăng tỷ lệ trả lời luôn là vấn đề quan trọng để giảm bớt sai số do không trả lời. Sau đây là một số giải pháp nhằm cố gắng gia tăng tỷ lệ trả lời:

- Gửi thư trước để thông báo cho người phỏng vấn và đề nghị họ hợp tác.
- Khi gửi thư, cần lưu ý hình thức và chất lượng bao thư, tem.
- Sự động viên bằng tiền hoặc tặng phẩm sẽ làm gia tăng tỷ lệ trả lời.
- Đối với các vấn đề riêng tư cần đảm bảo việc dấu tên hay giữ bí mật.
- Nên có những cuộc tiếp xúc sau khi người trả lời đã nhận được bản câu hỏi.
- Huấn luyện, tuyển chọn nhân viên phỏng vấn kỹ lưỡng sẽ góp phần tăng hiệu quả tiếp xúc.

6.1.4. Quy mô mẫu

Về nguyên tắc, sau khi đã tuân thủ theo một cách chọn mẫu có tính khoa học, mẫu càng lớn thì kết quả thu được càng có độ tin cậy cao. Nhưng giá trị tối thiểu của mẫu là bao nhiêu? Giá trị này phụ thuộc vào các yếu tố sau (Schumacher & McMillan, 1993) (trích trong Lê Văn Hảo và cộng sự, 2016):

- Loại nghiên cứu: nếu nghiên cứu về sự tương quan giữa các mẫu con (là mẫu ứng với kết quả phân tầng cuối cùng) thì độ lớn tối thiểu của mỗi mẫu con là 15. Đối với

các nghiên cứu nặng về khảo sát (survey), kích thước tối thiểu của mỗi mẫu con là 100, của các mẫu phụ của mẫu con (nếu có) là từ 20-50.

- Số lượng biến khảo sát: nghiên cứu càng bao gồm nhiều biến khảo sát, kích thước của mẫu càng phải lớn. Kích thước tối thiểu của mỗi mẫu con cần gấp 4-5 lần số biến khảo sát (Hoàng Trọng, Chu Nguyễn Mộng Ngọc, 2008).

- Yêu cầu về tính chính xác: nghiên cứu đòi hỏi tính chính xác càng cao, kích thước của mẫu càng phải lớn.

- Tầm quan trọng của nghiên cứu: nghiên cứu càng có tầm quan trọng, kích thước của mẫu càng phải lớn.

- Năng lực tài chính: khả năng tài chính càng hạn hẹp, kích thước của mẫu càng lấy gần đến giá trị tối thiểu.

6.1.5. Ước lượng tham số tổng thể từ các trị số của mẫu

6.1.5.1. Các khái niệm

Ước lượng nghĩa là tính toán một cách gần đúng nhất giá trị của một đại lượng chưa biết dựa trên những thông tin đã có. Ta sẽ ước lượng cho những đại lượng đo lường về độ lớn như trung bình; đại lượng đo lường về độ dao động như phương sai, độ lệch chuẩn; đại lượng đo lường về khả năng như xác suất, tỷ lệ tổng thể.

Đại lượng cần đo lường là các tham số tổng thể (trung bình, phương sai, tỷ lệ), là chưa biết. Chính vì chưa biết nhưng lại cần biết để phân tích các quyết định nên cần ước lượng. Thông tin mà ta có đều lấy từ mẫu.

Ước lượng tham số là tính toán một cách gần đúng nhất giá trị của một tham số chưa biết trong tổng thể dựa trên thông tin từ một mẫu. Có nhiều tham số trong tổng thể, nhưng trong bài trước chỉ đề cập đến ba tham số chính, vì vậy tại đây ta cũng sẽ tập trung vào ba tham số này, vì vậy ta có ba bài toán:

- Ước lượng trung bình tổng thể: μ
- Ước lượng phương sai tổng thể: σ^2
- Ước lượng tỷ lệ tổng thể: p

Tham số độ lệch chuẩn tổng thể σ luôn phải tính toán thông qua phương sai tổng thể σ^2 do đó không cần tách thành bài toán riêng.

Thay vì phải viết với ba tham số μ , σ^2 , p riêng biệt, tạm thời dùng ký hiệu chung là θ (đọc là tê-ta). Khi viết tham số tổng quát θ thì ta hiểu có ba trường hợp chính là μ , σ^2 , p .

Khi ước lượng cho tham số θ dựa trên thông tin từ mẫu, có hai loại ước lượng là ước lượng điểm và ước lượng khoảng.

Ước lượng điểm: Trong thực tế ta thường dùng khái niệm ước lượng điểm như khi nói: “ước lượng cho lạm phát là 6,5%”; “ước lượng mức tăng trưởng kinh tế là 8%”, nghĩa là chỉ dùng một con số duy nhất để ước lượng.

Ước lượng điểm trong thống kê toán là tìm ra một giá trị, tính toán trên mẫu, do đó tùy thuộc mẫu mà kết quả sẽ có thể khác nhau.

Ước lượng khoảng: Bên cạnh ước lượng điểm “ước lượng lạm phát là 6,5%”, có cách nói thứ hai: “ước lượng mức lạm phát là trong khoảng 6 đến 7 phần trăm”. Đây là hình ảnh của một ước lượng khoảng. Khi ước lượng khoảng ta mong muốn khoảng đó chứa con số cần tìm là đúng với khả năng cao nhất. Từ đó có khái niệm ước lượng khoảng trong thống kê toán.

6.1.5.2. Ước lượng điểm

Ước lượng tham số bằng một giá trị tính toán trên mẫu gọi là ước lượng điểm cho tham số đó. Với mẫu ngẫu nhiên thì giá trị đó là một thống kê ngẫu nhiên, với mẫu cụ thể thì giá trị đó là một con số.

Ký hiệu ước lượng điểm của tham số θ là $\hat{\theta}$ (tê - ta mũ), đây là quy ước quốc tế của ước lượng điểm.

Với mẫu ngẫu nhiên $(x_1, x_2, \dots, x_n)$ thì thống kê có dạng: $\hat{\theta} = f(x_1, x_2, \dots, x_n)$ là một hàm số trên mẫu.

Với mẫu cụ thể $(x_1, x_2, \dots, x_n)$ thì thống kê có dạng $\theta_{qs} = f(x_1, x_2, \dots, x_n)$ là một con số. Chữ qs viết tắt của quan sát. Giá trị tính trên mẫu gọi là giá trị quan sát.

Tuy nhiên với một tham số của tổng thể, có nhiều cách để tìm được một ước lượng điểm. Trong số nhiều ước lượng điểm, cần tìm ra ước lượng điểm chính xác nhất. Để lựa chọn ước lượng tốt, cần có tiêu chuẩn đặt ra.

***Tiêu chuẩn lựa chọn ước lượng điểm**

Một ước lượng điểm là tốt khi nó không có sai lầm mang tính hệ thống và sai lầm ngẫu nhiên phải ở mức nhỏ nhất.

Ví dụ: Để ước lượng trung bình tổng thể μ , xét một mẫu ngẫu nhiên gồm hai người sẽ điều tra, hay mẫu kích thước là 2: $W = (X_1, X_2)$. Với mẫu này, có ba ý kiến như sau:

Ý kiến 1: Cho rằng người được điều tra đầu tiên là quan trọng nhất, do đó dùng công thức:

$$m_1 = \frac{2X_1 + X_2}{2}$$

Ý kiến 2: Cho rằng người được điều tra đầu tiên là quan trọng, nhưng khi chia cho tổng số thì phải coi người thứ nhất như hai người, do đó dùng công thức:

$$m_2 = \frac{2X_1 + X_2}{3}$$

Ý kiến 3: Cho rằng hai người được điều tra quan trọng như nhau, nên dùng công thức:

$$m_3 = \frac{X_1 + X_2}{2}$$

Trong ba công thức đó, công thức nào dùng là tốt nhất?

Để giải bài toán này, xét hai tiêu chuẩn của ước lượng điểm là tính không chệch và tính hiệu quả:

- Tính không chệch

Định nghĩa - Tính không chệch: Thống kê θ của mẫu gọi là ước lượng không chệch của tham số θ của tổng thể nếu kỳ vọng của nó bằng đúng giá trị tham số.

Vậy θ là ước lượng không chệch của θ thì: $E(\theta) = \theta$

Nếu $E(\theta) \neq \theta$ thì θ là ước lượng chệch của θ . Ước lượng chệch sẽ dẫn đến những sai lệch mang tính hệ thống, ước lượng cao quá hoặc thấp quá giá trị cần ước lượng. Nếu ước lượng chệch được dùng trong các ước lượng tham số khác nữa, thì kết quả sẽ càng sai lầm.

- Tính hiệu quả

Giả sử lệch θ_1, θ_2 là các ước lượng không chệch của θ , nếu $V(\theta_1) < V(\theta_2)$ thì ước lượng θ_1 được gọi là hiệu quả hơn ước lượng θ_2 .

Định nghĩa – Tính hiệu quả: Thống kê θ của mẫu gọi là ước lượng hiệu quả θ của tham số θ của tổng thể nếu θ là ước lượng không chệch và có phương sai nhỏ nhất. trong số các ước lượng không chệch của θ . Như vậy, ước lượng hiệu quả trước tiên phải là ước lượng không chệch. Ước lượng không chệch và hiệu quả được gọi là ước lượng tốt nhất.

Ví dụ 1. Tìm ước lượng không chệch, hiệu quả của trung bình tổng thể m trong ba ước lượng sau:

$$m_1 = \frac{2X_1 + X_2}{2} \quad m_2 = \frac{2X_1 + X_2}{3} \quad m_3 = \frac{X_1 + X_2}{2}$$

Giải: Cần nhớ lại kiến thức của bài trước: Biến ngẫu nhiên gốc X có $E(X) = m$ và $V(X) = \sigma^2$ thì với mẫu ngẫu nhiên thì mọi thành phần mẫu đều có: $E(X_i) = m$ và $V(X_i) = \sigma^2$. Ngoài ra còn cần các tính chất của kỳ vọng và phương sai đã học trong bài số 2. Do đó để xét tính không chệch, ta tính kỳ vọng của các ước lượng.

$$E(m_1) = E\left(\frac{2X_1 + X_2}{2}\right) = \frac{2E(X_1) + E(X_2)}{2} = \frac{2m + m}{2} = \frac{3}{2}m$$

$$E(m_2) = E\left(\frac{2X_1 + X_2}{3}\right) = \frac{2E(X_1) + E(X_2)}{3} = \frac{2m + m}{3} = m$$

$$E(m_3) = E\left(\frac{X_1 + X_2}{2}\right) = \frac{E(X_1) + E(X_2)}{2} = \frac{m + m}{2} = m$$

Vậy trong ba ước lượng thì ước lượng đầu tiên có kỳ vọng khác với m , nó là ước lượng chệch, không nên sử dụng.

Với hai ước lượng không chệch còn lại, xét tính hiệu quả qua phương sai. Nhớ rằng khi cho hằng số ra ngoài phương sai phải bình phương hằng số đó lên:

$$V(m_2) = V\left(\frac{X_1 + X_2}{3}\right) = \frac{2^2 V(X_1) + V(X_2)}{3^2} = \frac{4\sigma^2 + \sigma^2}{9} = \frac{5}{9}\sigma^2$$

$$V(m_3) = V\left(\frac{X_1 + X_2}{2}\right) = \frac{V(X_1) + V(X_2)}{2^2} = \frac{\sigma^2 + \sigma^2}{4} = \frac{1}{2}\sigma^2$$

Ta thấy $V(m_2) > V(m_3)$ nên ước lượng thứ ba hiệu quả hơn.

Vậy trong ba ước lượng thì ước lượng thứ nhất là chệch, ước lượng thứ hai tuy không chệch nhưng không hiệu quả như ước lượng thứ ba. Ước lượng thứ ba là không chệch và hiệu quả nhất.

Ví dụ 2. Biến ngẫu nhiên gốc X có trung bình là m , phương sai là σ^2 . Với mẫu ngẫu nhiên kích thước là 3: $W = (X_1, X_2, X_3)$, cho biết trong số các thống kê sau, hàm nào là ước lượng không chệch, hàm nào là ước lượng hiệu quả hơn:

$$G_1 = \frac{1}{2}X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3; \quad G_2 = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3; \quad G_3 = \frac{1}{3}(X_1 + X_2 + X_3)$$

Giải: Để xét tính không chệch, tính trung bình (kì vọng) của các thống kê. Do mẫu là ngẫu nhiên nên ta có: $E(X_1) = E(X_2) = E(X_3) = m$ và $V(X_1) = V(X_2) = V(X_3) = \sigma^2$. Ta có:

$$\begin{aligned} E(G_1) &= E\left(\frac{1}{2}X_1 + \frac{1}{2}X_2 + \frac{1}{3}X_3\right) = \frac{1}{2}EX_1 + \frac{1}{2}EX_2 + \frac{1}{3}EX_3 \\ &= \frac{1}{2}m + \frac{1}{2}m + \frac{1}{3}m = \frac{3}{2}m \neq m \end{aligned}$$

$$\text{Tương tự } E(G_2) = \frac{1}{2}m + \frac{1}{3}m + \frac{1}{6}m = m$$

$$E(G_3) = \frac{1}{3}m + \frac{1}{3}m + \frac{1}{3}m = m$$

Trong các thống kê thì G_1 là ước lượng chệch, G_2, G_3 là các ước lượng không chệch. Để xét tính hiệu quả, ta tính phương sai của các ước lượng không chệch. Ta có:

$$V(G_2) = \frac{1}{4}\sigma^2 + \frac{1}{9}\sigma^2 + \frac{1}{36}\sigma^2 = \frac{7}{18}\sigma^2$$

$$V(G_3) = \frac{1}{9}\sigma^2 + \frac{1}{9}\sigma^2 + \frac{1}{9}\sigma^2 = \frac{1}{3}\sigma^2$$

Do $V(G_2) > V(G_3)$ nên G_3 là ước lượng có phương sai nhỏ nhất, là ước lượng hiệu quả nhất trong số ba ước lượng. Qua hai ví dụ trên có thể rút ra nhận xét là:

Ước lượng cho trung bình tổng thể trên một tổ hợp của các thành phần mẫu thì.

- Tổ hợp đó là ước lượng không chệch khi tổng các hệ số bằng 1.
- Tổ hợp không chệch (tổng hệ số bằng 1) là hiệu quả nhất khi các hệ số bằng nhau.

Dựa vào các định lý và tính chất của biến ngẫu nhiên, người ta chứng minh được nếu biến ngẫu nhiên gốc phân phối chuẩn thì trung bình mẫu $\bar{X}$ là ước lượng không chệch, hiệu quả của trung bình tổng thể m (cũng chính là tham số μ); phương sai S^2 là ước lượng không chệch của phương sai tổng thể σ^2 ; tỷ lệ mẫu f là ước lượng không

chênh lệch của tần suất tổng thể p . Với một mẫu cụ thể, các ước lượng điểm sẽ được tính là một giá trị cụ thể.

6.1.5.3. Ước lượng khoảng

Trong phần trước, ước lượng điểm là một giá trị dùng để ước lượng cho tham số chưa biết. Trong nhiều trường hợp, ước lượng điểm là chưa đủ và có thể không đảm bảo độ chính xác cần thiết, khi đó cần tìm một khoảng giá trị để ước lượng cho tham số chưa biết. Với một khoảng giá trị dùng để ước lượng cho một tham số chưa biết, có thể không đảm bảo chính xác hoàn toàn, mà chỉ đúng với một xác suất đúng nhất định, và vẫn có khả năng sai.

- Khái niệm - Ước lượng khoảng: Ước lượng tham số bằng một khoảng tính toán trên mẫu, sao cho xác suất để khoảng đó chứa con số cần tìm là một giá trị đủ lớn, gọi là ước lượng khoảng cho tham số đó.

Ước lượng khoảng cho tham số θ là tìm một khoảng (θ_1, θ_2) sao cho: $P(\theta_1 < \theta < \theta_2)$ là con số đủ lớn. Nếu ký hiệu xác suất cho phép sai là α thì xác suất yêu cầu đúng là $1-\alpha$, ta có:

$$P(\theta_1 < \theta < \theta_2) = 1 - \alpha$$

Khi đó ta có các cách gọi như sau:

- + Khoảng (θ_1, θ_2) gọi là khoảng tin cậy của tham số θ
- + Giá trị $(1 - \alpha)$ gọi là độ tin cậy của ước lượng.
- + Đại lượng $1 = \theta_2 - \theta_1$ gọi là độ dài khoảng tin cậy

Ước lượng khoảng có độ dài khoảng tin cậy càng ngắn là càng tốt. Thông thường lấy độ tin cậy là 95% hay nói khác đi là xác suất đúng là 95%, cho phép sai là 5%.

- Phương pháp tìm ước lượng khoảng

Phương pháp tìm ước lượng khoảng tổng quát tính trên các mẫu ngẫu nhiên sẽ dựa vào các quy luật phân phối xác suất. Việc thực hiện chi tiết các biến đổi có thể xem trong Giáo trình kinh tế Lượng của Nguyễn Quang Đông năm 2012 của Đại học Kinh tế quốc dân. Với mẫu ngẫu nhiên, công thức ước lượng khoảng sẽ cho một khoảng ngẫu nhiên, và có thể viết dưới dạng xác suất. Với một mẫu cụ thể, thay số vào các đại lượng ngẫu nhiên, sẽ tính ra một khoảng cụ thể, và không gắn với xác suất. Để tránh nhầm lẫn, trong

cả hai trường hợp mẫu ngẫu nhiên hay cụ thể đều không viết với xác suất. Người học sẽ sử dụng các công thức đã được chứng minh, thay giá trị bằng số để tính ra kết quả cuối cùng (tham khảo Phương pháp nghiên cứu khoa học trong kinh doanh của Nguyễn Đình Thọ, 2011 nhà xuất bản Lao động Xã hội).

6.1.5.4. Ước lượng trung bình tổng thể

- Ước lượng điểm trung bình tổng thể

Ước lượng điểm không chệch cho trung bình tổng thể chính là trung bình mẫu. Ta có, trung bình tổng thể $E(X) = m$ nên $\bar{X}$ là ước không chệch của m . Phương sai của trung bình mẫu: $V(\bar{X}) = \frac{\sigma^2}{n}$

Chứng minh được khi tổng thể phân phối chuẩn thì $\bar{X}$ cũng là ước lượng hiệu quả nhất, hay là ước lượng tốt nhất.

- Ước lượng khoảng trung bình tổng thể phân phối Chuẩn

Ta chỉ xét bài toán ước lượng khoảng cho trung bình tổng thể khi tổng thể phân phối Chuẩn. Giả sử tổng thể có biến ngẫu nhiên gốc X phân phối chuẩn $X \sim N(\mu, \sigma^2)$, khi đó bình tổng thể sẽ được ký hiệu là μ , phương sai tổng thể cũng là phương sai biến ngẫu nhiên σ^2 .

Với mẫu W kích thước n , với độ tin cậy là $(1 - \alpha)$ cho trước.

Với W , tính được các thống kê đặc trưng mẫu $\bar{X}, S^2$.

Chúng ta xây dựng công thức ước lượng khoảng cho trung bình tổng thể như sau:

$$T = \frac{(\bar{x} - \mu) \sqrt{n}}{S} \sim T(n - 1)$$

6.1.5.5. Ước lượng phương sai tổng thể

- Ước lượng điểm

Phương sai tổng thể là $V(X) = \sigma^2$. Ước lượng không chệch cho phương sai tổng thể trong mẫu ngẫu nhiên chính là phương sai mẫu S^2 , vì ta đã có $E(S^2) = \sigma^2$. Tuy nhiên việc chứng minh tính hiệu quả của phương sai mẫu không dễ dàng, kể cả khi biến ngẫu nhiên gốc phân phối Chuẩn, đặc biệt khi không có thông tin gì về trung bình tổng thể.

- Ước lượng khoảng phương sai tổng thể phân phối Chuẩn

Khi X phân phối chuẩn, phương sai tổng thể σ^2 , cần ước lượng, với độ tin cậy $(1 - \alpha)$, dựa trên quy luật của thống kê phân phối Khi – bình phương bậc tự do $(n - 1)$. Từ đó có công thức ước lượng khoảng, hay khoảng tin cậy của phương sai tổng thể:

$$\frac{(n-1)S^2}{X_{\alpha/2}^2(N-1)} < \sigma^2 < \frac{(n-1)S^2}{X_{1-\alpha/2}^2(N-1)}$$

Trong đó hai giá trị ở dưới mẫu là giá trị tới hạn khi - bình phương bậc tự do $(n - 1)$ với mức là $(\alpha/2)$ và $(1-\alpha/2)$. Với mẫu cụ thể, thay S^2 bằng s^2 tính từ mẫu, nên khoảng tin cậy là:

$$\frac{(n-1)s^2}{X_{\alpha/2}^2(N-1)} < \sigma^2 < \frac{(n-1)s^2}{X_{1-\alpha/2}^2(N-1)}$$

6.1.5.6. Ước lượng tỷ lệ tổng thể

- Ước lượng điểm tỷ lệ tổng thể

Với một tổng thể kích thước N, có M phần tử chứa dấu hiệu A, thì $P = \frac{M}{N}$ là tỷ lệ tổng thể, hay tần suất tổng thể của dấu hiệu A. Nếu coi việc xuất hiện dấu hiệu A là một biến cố, thì p chính là xác suất của biến cố đó. Với một mẫu kích thước n, tỷ lệ mẫu hay tần suất mẫu là f, chứng minh được f là ước lượng điểm không chệch và hiệu quả nhất của p.

- Ước lượng khoảng tỷ lệ tổng thể

Cần ước lượng tần suất tổng thể p với độ tin cậy $(1 - \alpha)$ dựa trên một mẫu ngẫu nhiên kích thước n. Trong mẫu, tần suất mẫu của dấu hiệu A là $f = X_{A/n}$ với X_A là tần số ngẫu nhiên của dấu hiệu A trong mẫu.

Với mẫu ngẫu nhiên hoặc cụ thể, ký hiệu f dùng chung. Bài toán ước lượng khoảng cho tỷ lệ tổng thể thực hiện với mẫu có kích thước từ 100 trở lên: $n \geq 100$. Ước lượng khoảng, hoặc khoảng tin cậy có dạng đối xứng:

$$f - \frac{\sqrt{f(1-f)}}{\sqrt{n}} u_{\alpha/2} < P < f + \frac{\sqrt{f(1-f)}}{\sqrt{n}} u_{\alpha/2}$$

Ước lượng khoảng này có thể viết dưới dạng:

$$f - \varepsilon < P < f + \varepsilon$$

Trong đó ε gọi là sai số, và

$$\varepsilon = \frac{\overline{f(1-f)}}{\bar{n}} u_{\alpha/2}$$

Do đó nếu muốn sai số của ước lượng không vượt quá ε_0 thì kích thước mẫu tối thiểu là:

$$n' \geq \frac{f(1-f)}{\varepsilon_0^2} u_{\alpha/2}^2$$

6.2. Thiết kế bảng câu hỏi

6.2.1. Khái niệm

Bảng câu hỏi là một công cụ để thu thập dữ liệu định lượng, nó bao hàm một tập hợp các câu hỏi và câu trả lời theo một logic nhất định. Do vậy, xây dựng bảng câu hỏi là một tiến trình được chính thức hóa nhằm thu thập, ghi chép lại những thông tin xác đáng và được chỉ định rõ với sự chính xác và hoàn hảo tương đối.

Có hai dạng bảng câu hỏi, (1) bảng câu hỏi chi tiết dùng trong thu thập dữ liệu trong nghiên cứu định lượng và (2) dàn bài hướng dẫn thảo luận dùng trong nghiên cứu định tính. Bảng hỏi dùng trong nghiên cứu định lượng thường rất nhiều về mặt cấu trúc so với bảng hỏi trong nghiên cứu định tính.

Một bảng hỏi tốt sẽ giúp nhà nghiên cứu thu thập dữ liệu cần thiết với độ tin cậy cao. Một bảng hỏi phải thỏa mãn hai yêu cầu cơ bản sau đây:

- Phải có đầy đủ các câu hỏi mà nhà nghiên cứu muốn thu thập dữ liệu từ các câu trả lời.
- Phải kích thích được sự hợp tác của người trả lời.

6.2.2. Mục đích

Một bảng câu hỏi chỉ một tiến trình được chính thức hóa nhằm thu thập và ghi chép lại những thông tin xác đáng và được chỉ định rõ với sự chính xác và hoàn hảo tương đối. Nói cách khác, nó điều khiển quá trình đặt câu hỏi và giúp cho việc ghi chép được rõ ràng và chính xác. Do đó, mục đích của bảng câu hỏi gồm:

- Giúp người được phỏng vấn hiểu biết rõ ràng các câu hỏi.
- Khiến cho người được phỏng vấn muốn hợp tác và tin rằng những câu trả lời của họ sẽ được giữ kín.
- Khuyến khích những câu trả lời thông qua sự xem xét lại nội tâm kỹ hơn, lục lại trí nhớ hay liên hệ với những điều đã ghi chép.

- Hướng dẫn rõ ràng những điều mình muốn biết và cách trả lời.
- Xác định những nhu cầu cần thiết để phân loại và kiểm tra lại cuộc phỏng vấn.

6.2.3. Quy trình thiết kế bảng câu hỏi

Do sự chính xác và thích hợp của các dữ liệu thu thập được phụ thuộc chủ yếu vào bảng câu hỏi, vì vậy thiết kế một bảng câu hỏi tốt sẽ có tác dụng rất lớn đến sự thành công của một dự án nghiên cứu. Việc thiết kế một bảng câu hỏi được coi như là một nghệ thuật hơn là một khoa học, điều đó hàm ý rằng, bạn sẽ chẳng bao giờ sáng tạo ra được một bản câu hỏi phỏng vấn chỉ dựa vào việc đọc sách mà cần phải bắt tay vào soạn thảo và thực hành liên tục. Các nhà nghiên cứu marketing đã tổng kết được 8 bước cơ bản về trình tự thiết kế một bảng câu hỏi:

Bước 1: Xác định các dữ kiện riêng biệt cần tìm

Điểm đầu tiên khi bắt tay vào thiết kế bản câu hỏi là nhà nghiên cứu cần phải xem xét mục tiêu nghiên cứu để xác định chính xác cái gì cần phải được đo lường. Liệt kê những gì cần đo lường, có thể là danh sách những ý hỏi (items) và câu hỏi riêng biệt, những nhóm chữ hay từ chủ yếu. Nếu danh sách này quá dài, nhà nghiên cứu cần phải loại bỏ bớt những nội dung quá xa với mục tiêu nghiên cứu để tiết kiệm thời gian và chi phí. Sau đó, dự tính xem những biến số được đo lường sẽ được sử dụng như thế nào, nên dùng loại kỹ thuật phân tích nào để mang lại ý nghĩa cho dữ liệu.

Bước 2: Xác định phương pháp điều tra

Trong bước này, người nghiên cứu cần quyết định dùng phương pháp nào để tiếp xúc với người được phỏng vấn (phỏng vấn trực tiếp, gọi điện thoại, gửi thư hay Internet.). Các phương pháp tiếp xúc khác nhau sẽ yêu cầu nội dung, cũng như cấu trúc câu hỏi của bản câu hỏi khác nhau.

Bước 3: Đánh giá nội dung câu hỏi

Mục tiêu và nội dung của vấn đề nghiên cứu quyết định nội dung các câu hỏi trong bản câu hỏi. Thế nhưng, việc có được những thông tin thích đáng từ những câu trả lời hay không lại phụ thuộc rất lớn vào khả năng phác thảo bản câu hỏi của người nghiên cứu. Do vậy, khi xây dựng các câu hỏi, cần cân nhắc các tiêu chuẩn sau:

- Câu hỏi đặt ra có cần thiết hay không?

Mỗi một câu hỏi đưa ra người nghiên cứu cần phải tự hỏi là câu hỏi đó có đóng góp và làm rõ mục tiêu nghiên cứu hay không, nếu không thì nên loại bỏ những câu hỏi này. Trên thực tế, trong một bảng câu hỏi cũng có một số câu hỏi tuy không thực sự liên quan trực tiếp đến mục tiêu nghiên cứu nhưng nó có thể dẫn dắt, định hướng và giúp cho người phỏng vấn gợi nhớ lại thông tin và trả lời chính xác những thông tin đó.

- Người trả lời có hiểu được câu hỏi đó không?

Người trả lời không hiểu câu hỏi có thể do nhiều nguyên nhân, thông thường là do người nghiên cứu dùng các thuật ngữ không quen thuộc với người được hỏi; thiếu định nghĩa rõ ràng về các thuật ngữ sử dụng; hoặc đặt câu hỏi mơ hồ, bỏ qua những khác biệt về văn phong, thói quen giao tiếp giữa những người có sắc tộc hoặc văn hóa khác nhau. Trong bản câu hỏi, người nghiên cứu cũng nên dùng ngôn từ quen thuộc.

Câu hỏi nên được xây dựng đúng cú pháp, văn phạm, tránh dùng câu phức, tránh dùng tiếng lóng hay các thuật ngữ chuyên môn...

- Người trả lời có được những thông tin cần thiết để trả lời các câu hỏi đó không?

Người được hỏi không trả lời được hoặc trả lời không đúng do 2 nguyên nhân: thiếu kiến thức về vấn đề được hỏi hoặc không nhớ sự kiện do câu hỏi đòi hỏi sự hồi tưởng quá nhiều. Để khắc phục tình trạng này, nhà nghiên cứu có thể sử dụng một số biện pháp sau:

+ Hỏi nhiều câu hỏi để gợi lại trí nhớ.

+ Xác định khoảng thời gian rõ ràng, thời gian càng xa độ chính xác của câu trả lời càng giảm.

+ Hỏi các câu hỏi có tính chất liên tưởng, gợi sự liên quan giữa các sự kiện để người trả lời nhớ lại.

+ Đề nghị người trả lời nêu rõ sự kiện nào họ nhớ chính xác nhất và sự kiện nào còn mơ hồ.

- Người trả lời liệu có cung cấp các thông tin đó không?

Người trả lời không muốn trả lời hoặc trả lời sai một cách cố ý là do câu hỏi đi vào những vấn đề có tính chất riêng tư, những vấn đề bí mật không muốn tiết lộ, hoặc các câu