

TÊN

NGÀY

TIẾT HỌC

## Tài liệu Hỗ trợ Gia đình

### Suy luận thống kê

Trong bài học này, học sinh sẽ sử dụng một *mẫu* dữ liệu nhỏ để ước tính thông tin về một nhóm lớn hơn được gọi là *quần thể* và sử dụng mô phỏng để xác định phạm vi giá trị cho ước tính. Quần thể là toàn bộ tập hợp các đối tượng quan tâm cho một câu hỏi và mẫu là một nhóm nhỏ hơn trong quần thể đó.



Ví dụ: chúng ta có thể muốn tính ra số tiền *trung bình* (hoặc trung bình) mà các gia đình ở Hoa Kỳ chi tiêu cho thực phẩm hàng tháng. Quần thể bao gồm tất cả các gia đình ở Hoa Kỳ, nhưng việc thu thập thông tin từ mọi người sẽ rất khó khăn và tốn kém, vì vậy chúng ta có thể bắt đầu thu thập dữ liệu với mẫu gồm 50 gia đình.

Một câu hỏi quan trọng cần cân nhắc khi bắt đầu thu thập thông tin từ một mẫu là cách chọn mẫu. Dữ liệu bạn thu thập có thể rất khác nếu bạn hỏi những gia đình đang mua sắm tại một cửa hàng tạp hóa địa phương so với việc hỏi những người ở bên ngoài một nhà hàng sang trọng. Tương tự, số tiền chi cho thực phẩm ở San Francisco có thể rất khác so với số tiền chi tiêu ở vùng nông thôn Iowa. Thậm chí có thể có một số thói quen chi tiêu ẩn giấu theo những cách mà chúng ta chưa nghĩ tới. Vì vậy, làm thế nào

TÊN

NGÀY

TIẾT HỌC

để bạn chắc chắn rằng mẫu của mình đại diện cho các gia đình ở Hoa Kỳ mà không sử dụng quá nhiều gia đình từ các nhóm không điển hình trong chi tiêu của họ?

Giải pháp là sử dụng *tính ngẫu nhiên*. Chúng ta có thể chọn 50 hộ bằng cách sử dụng quy trình ngẫu nhiên chẳng hạn như dùng máy tính chọn ngẫu nhiên các hộ từ cơ sở dữ liệu mà không xem xét các yếu tố khác. Điều này sẽ làm giảm *yếu tố thiên vị* có thể xảy ra khi con người cố gắng lấy thông tin về con người và nó có thể sẽ bao gồm tỷ lệ chính xác hơn về các loại gia đình khác nhau ở Hoa Kỳ. Mặc dù tính ngẫu nhiên có thể không loại bỏ hoàn toàn sai lệch khỏi việc lựa chọn mẫu nhưng nó sẽ làm giảm đáng kể sai lệch hiện có khi so sánh với lựa chọn không ngẫu nhiên.

Các nhà nghiên cứu đã thực hiện những nghiên cứu như thế này và phát hiện ra rằng số tiền trung bình chi cho thực phẩm mỗi tháng là bao nhiêu. Một báo cáo cho biết số tiền trung bình chi cho thực phẩm mỗi tháng là 600\$ với *biên độ sai số* (MoE) là 150\$. Biên độ sai số được sử dụng để cho biết chúng ta không kỳ vọng mỗi gia đình trong mẫu sẽ chi chính xác 600\$.

Biên độ sai số rất quan trọng để tìm kiếm trong kết quả thống kê. Sẽ là vô trách nhiệm khi thảo luận về số liệu thống kê mà không đưa ra một giới hạn sai số để mô tả giá trị dự kiến sẽ thay đổi bao nhiêu. Nhiều biểu đồ có trong các bản tin sẽ báo cáo dưới dạng chữ in nhỏ trên đồ họa. Hãy tìm nội dung như  $\pm 3\%$  khi có hình ảnh về xếp hạng phê duyệt dành cho một quan chức hoặc các cuộc thăm dò trong cuộc bầu cử tiếp theo. Điều này có nghĩa là tỷ lệ phần trăm hiển thị trong đồ họa thực sự có thể thấp hơn hoặc cao hơn tới 3% so với con số hiển thị.

### Đây là một nhiệm vụ để thực hành với học sinh:

Một thị trấn sắp tổ chức cuộc bỏ phiếu về việc có nên tăng thuế thu nhập doanh nghiệp lên 2% để tăng kinh phí cho các trường công hay không. Tin tức địa phương công bố một hình ảnh cho thấy 52% số người bỏ phiếu ủng hộ việc tăng thuế và ở góc có hình ảnh “biên độ sai số  $\pm 3.5\%$ ”. Điều tra viên có vẻ tự tin rằng thuế doanh nghiệp sẽ tăng vì mọi giá trị lớn hơn 50% số phiếu ủng hộ thuế sẽ thì luật sẽ được thông qua.

1. Điều tra viên tìm ra con số 52% đã đạt được bằng cách lái xe đến 4 trong số 20 khu dân cư khác nhau xung quanh thị trấn và hỏi ý kiến của người dân. Có điều gì sai với cách thực hiện đó? Bạn có thể nghĩ ra cách nào tốt hơn để thu thập dữ liệu không?
2. Biên độ sai số có ý nghĩa gì trong hình ảnh này?
3. Bạn có nên tự tin rằng thuế sẽ được phê chuẩn tăng? Giải thích lý do.

### Lời giải:

1. Chỉ đi đến 4 khu phố trong thị trấn có thể sẽ bỏ sót ý kiến của nhiều cử tri ở các khu phố khác mà phóng viên chưa đi. Cách tốt hơn để thu thập thông tin có thể là chọn ngẫu nhiên một số hộ gia đình trong thị trấn để khảo sát ý kiến của họ.

TÊN

NGÀY

TIẾT HỌC

Việc lựa chọn ngẫu nhiên có nhiều khả năng tránh được bất kỳ thành kiến nào mà người báo cáo có về những khu vực lân cận sẽ đến thăm.

2. Biên độ sai số có nghĩa là tỷ lệ thực tế ủng hộ việc tăng thuế có thể cao hơn 3,5% hoặc thấp hơn 3,5% so với mức 52% được báo cáo dựa trên mẫu. Điều này có nghĩa là tỷ lệ phần trăm thực tế sẽ rơi vào khoảng từ 48,5% đến 55,5%.
3. Đáp an mẫu:
  - Tôi nghĩ vẫn có khả năng tăng thuế. Mặc dù tỷ lệ phần trăm thực tế có thể thấp tới 48,5% dựa trên tỷ lệ sai số nhưng cũng có thể cao tới 55,5%. Hầu hết các tỷ lệ phần trăm có thể xảy ra đều trên 50%, vì vậy tôi nghĩ tăng thuế sẽ xảy ra.
  - Tôi nghĩ vẫn chưa rõ liệu việc tăng thuế có xảy ra hay không. Dựa trên tỷ lệ sai số, tỷ lệ phần trăm thực tế có thể thấp tới 48,5%, điều này sẽ dẫn đến việc tăng thuế không xảy ra. Tôi cũng không chắc chắn về phương pháp thu thập mẫu trong báo cáo này của điều tra viên nên báo cáo có thể không chính xác lắm.



Bản quyền © CC BY 2019 của Illustrative Mathematics®