

ИМЯ И ФАМИЛИЯ

ДАТА

ПЕРИОД

## Сопроводительные материалы для семей

### Статистические выводы

В этом модуле ваш учащийся будет использовать небольшую *выборку* данных для оценки информации о большой группе, называемой *генеральной совокупностью*, и применять моделирование для определения диапазона значений для оценки. Генеральная совокупность представляет собой весь набор объектов, представляющих интерес для решения задачи, а выборка — это небольшая группа в пределах этой генеральной совокупности.



Например, требуется определить *среднее значение* (или *среднее*) суммы, которую семьи в США ежемесячно тратят на продукты питания. Генеральная совокупность включает все семьи в США, но сбор информации со всех из них будет очень сложным и затратным, поэтому можно начать сбор данных с выборки из 50 семей.

Важный вопрос, на который необходимо дать ответ перед началом сбора информации из выборки, состоит в том, как получить эту выборку. Собранные данные могут существенно отличаться, если опрашивать семьи, совершающие

ИМЯ И ФАМИЛИЯ

ДАТА

ПЕРИОД

покупки в местном продуктовом магазине, или людей, идущих в дорогой ресторан. Аналогично сумма, затрачиваемая на продукты питания в Сан-Франциско, может значительно отличаться от суммы, затрачиваемой в сельских областях Айовы. Некоторые особенности трат могут таиться там, где мы их меньше всего ожидаем. Итак, как убедиться, что ваша выборка является репрезентативной относительно семей в США, не обращаясь к слишком большому количеству семей из групп, траты которых являются нехарактерными?

Решение состоит в применении *случайности*. Мы можем отобрать 50 семей случайным образом, например дав компьютеру выбрать их из базы данных случайным образом, не учитывая другие факторы. Это должно привести к снижению *ошибки случайной выборки*, которая может возникнуть, когда сбором информации занимаются люди, и с большой долей вероятности позволит добиться более точных соотношений между различными типами семей в США. Несмотря на то, что фактор случайности, вероятно, не исключит полностью ошибку случайной выборки, он существенно снизит ее по сравнению с выборкой без фактора случайности.

Специалисты выполнили подобные исследования и выяснили среднюю сумму, ежемесячно затрачиваемую на продукты питания. В отчете говорится, что средняя сумма, ежемесячно затрачиваемая на питание, составляет \$600, при этом *погрешность* составляет \$150. Погрешность используется, чтобы показать, что мы не ожидаем от каждой семьи в выборке затрат ровно \$600.

Для получения статистических результатов важно найти погрешность. Было бы безответственно рассматривать статистику, не давая погрешность для описания того, насколько значение может варьироваться. Во многих случаях при демонстрации графиков в новостях она указывается мелким шрифтом. Когда речь идет о графике рейтинга одобрения чиновника или опросов на следующих выборах, ищите что-то вроде  $\pm 3\%$ . Это означает, что процентные значения, показанные на графике, могут быть на 3 % больше или меньше показанных чисел.

**Ниже приводится задача, которую следует попробовать решить со своим учащимся:**

В городе ожидается голосование по поводу того, следует ли повышать налог на прибыль организаций на 2 % в целях повышения финансирования государственных общеобразовательных школ. В местных новостях демонстрируется изображение, показывающее, что 52 % электората поддерживают повышение налога, а в уголке указано: «погрешность  $\pm 3,5\%$ ». Репортер с уверенностью говорит о том, что налоги на организации будут повышены, потому что при любом результате голосования выше 50 % за налоги закон будет принят.

ИМЯ И ФАМИЛИЯ

ДАТА

ПЕРИОД

1. Репортер, получивший число 52 %, пришел к этому, объехав 4 из 20 разных районов города и спросив у жителей их мнение. Есть ли в его действиях какая-либо ошибка? Вы можете придумать более удачный способ сбора данных?
2. Что на этом изображении означает погрешность?
3. А вы уверены в том, что налоги будут повышены? Объясните свои рассуждения.

**Решения:**

1. Посещение всего 4 районов города может оставить неучтенными мнения многих избирателей из других районов, куда репортер не поехал. Более удачным способом сбора информации был бы случайный отбор нескольких домохозяйств в городе с целью узнать мнение их жильцов. Случайный отбор с большей долей вероятности позволит репортеру избежать ошибок случайной выборки, связанных с выбором посещенных районов.
2. Погрешность означает, что фактический процент поддерживающих повышение налогов может быть на 3,5% выше или на 3,5% ниже 52 %, указанных, исходя из выборки. Это означает, что фактическое процентное значение будет варьироваться между 48,5 % и 55,5 %.
3. Примерные ответы:
  - Я думаю, что вероятность повышения налогов все-таки велика. Однако фактическое процентное значение может быть всего лишь 48,5 %, исходя из погрешности, а может быть и 55,5 %. Большая часть возможных значений превышает 50 %, поэтому, я думаю, увеличение произойдет.
  - Я думаю, что пока неясно, произойдет ли повышение. Исходя из погрешности, фактическое процентное значение может составлять всего 48,5 %, в результате чего повышения не произойдет. Я также сомневаюсь в использованных репортером способах получения выборки, поэтому его отчет может быть неточным.



© CC BY 2019 Illustrative Mathematics®