

Анализ социологических данных – 1.

Задание к комиссии 27 сентября 2012 г.

База данных `dating.sav` содержит результаты эксперимента, проведенного в Колумбийском университете в Нью-Йорке в 2002-04 гг. (см. Fisman et al., Review of Economic Studies, 2008). Эксперимент заключался в серии “быстрых свиданий” (speed dating) между студентами и аспирантами университета. На каждую из 21 вечеринок приглашались от 10 до 40 участников, примерно поровну юноши и девушки. Девушки располагались за столиками, между которыми перемещались молодые люди. На общение между каждой парой отводились четыре минуты. После каждого “свидания” участники оценивали своего собеседника или собеседницу по шести параметрам (по 11-балльной шкале): внешняя привлекательность, искренность, интеллект, чувство юмора, амбициозность, наличие общих интересов. Также участники отмечали, хотели бы они встретиться с этим человеком еще раз. До вечеринки все участники заполняли анкету, в которой отвечали на вопросы о себе.

В базе данных каждая строчка соответствует одному короткому свиданию. Всего в базе 552 человека и 8378 свиданий. База содержит следующие переменные (все метки значений переменных имеются в базе данных):

- `id` – индивидуальный номер участника,
- `order` – порядковый номер собеседника на вечеринке (1 - говорил с ним/ней первым, 2 - вторым и т.д.),
- `male` – пол участника,
- `age` – возраст участника,
- `race` – раса участника,
- `age_o` – возраст собеседника или собеседницы,
- `race_o` – раса собеседника или собеседницы,
- `samerace` – фиктивная переменная, принимающая значение один, если участник и собеседник/собеседница относятся к одной расе,
- `int_corr` – коэффициент корреляции между интересами участника и собеседника/собеседницы, по данным анкет, заполненных до вечеринки. Чем выше корреляция, тем ближе интересы; отрицательная корреляция означает расхождение в интересах,
- `dec` – фиктивная переменная, принимающая значение один, если участник хочет встретиться с собеседником/собеседницей еще раз, и ноль в противном случае,

- `field` – специальность участника,
- `field_cd` – специальность участника, закодированная как числовая категориальная переменная,
- `attr` – оценка внешней привлекательности собеседника/-цы,
- `sinc` – оценка искренности,
- `intel` – оценка интеллекта,
- `fun` – оценка чувства юмора,
- `amb` – оценка амбициозности,
- `shar` – оценка близости интересов.

Строго говоря, наблюдения в этой базе данных нельзя рассматривать как независимые друг от друга, поскольку на каждого участника приходится несколько свиданий. Однако мы пренебрежем этим и будем рассматривать наблюдения как независимые.

1. Найдите средний и медианный возраст людей, принимавших участие в эксперименте. Найдите стандартное отклонение возраста.
2. С помощью линейной регрессии оцените зависимости оценки близости интересов с собеседником после встречи (`shar`) от предварительной оценки близости интересов до встречи (`int_corr`) и внешней привлекательности собеседника (`attr`). Проинтерпретируйте результаты статистически и содержательно.
3. С помощью логистической регрессии оцените зависимости вероятности того, что участники эксперимента хотят встретиться с собеседником еще раз (`dec`), от расы участника и оценок привлекательности, искренности, интеллекта, чувства юмора, амбициозности и близости интересов собеседника. Анализ проведите отдельно для мужчин и женщин. Приведите регрессионный вывод, проинтерпретируйте коэффициенты и их статистическую значимость. Дайте содержательную интерпретацию результатов. Как отличаются факторы, связанные с `dec` для мужчин и женщин?