

## Контрольная работа №2

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.

Используя любые известные вам статистические методы, подробно и аргументированно ответьте на следующие вопросы.

1. Каждый участник оценивал собеседника по шести параметрам. Как эти оценки связаны с решением о желательности еще одной встречи? Какие факторы важны, а какие нет? Различается ли характер связи в зависимости от пола и расы участников?
2. Исследуйте, какие факторы (помимо оценки собеседника/-цы по шести параметрам) связаны с вероятностью решения о желательности еще одной встречи (например, пол участника, раса участника и собеседника/-цы, возраст участника и разница в возрасте участника и собеседника/-цы, наличие общих интересов, измеренных до встречи, специальность, порядок свиданий). Постройте наилучшую модель, проверьте на значимость эффекты взаимодействия, которые вам представляются интересными. Оцените силу эффектов (с помощью предсказанных вероятностей; для данного задания используйте модель с небольшим количеством предикторов). Подробно опишите результаты.
3. Оцените связь между полом, расой и специализацией участников. Поскольку в этом задании все характеристики участников являются неизменными, используйте базу данных, которая содержит одно наблюдение для каждого участника. Подробно опишите характер связи.