# Рекомендации по выполнению задания №4

**(часть1)**

Рассматривается следующая модель данных

 Pi=β0+β1Qi+β2Ii+β3Di+β4DQi+β5DIi+ɛi

P – средняя стоимость одного предмета в покупке;
Q – количество предметов в покупке;
I – доход покупателя;
D – фиктивная переменная для местоположения.

Первоначально предлагается с помощью тестов Chow на «устойчивость коэффициентов» и «предсказательную силу» проверить значимость использования фиктивной переменной т.е. различие уравнений для «центра» и «окраины».

Проверьте коэффициенты корреляции (найдете в «анализе данных»).

В дальнейшем предполагается упрощение модели: выпросить все незначимые, оставить все значимые коэффициенты. Для этого из «полной» модели «выбрасывается» самый незначимый коэффициент (максимальное P-value). Затем из полученной модели «выбрасывается» следующий самый незначимый и т.д. При этом следите (т.е. распечатайте в виде таблицы) за изменением R2adj и F-статистикой на все выброшенные относительно «полной модели» коэффициенты. Обратите внимание на изменение этих показателей, когда выбрасываются незначимые и значимые коэффициенты.

Аналогичные действия проведите и при обратном выборе фиктивной переменной. Окончательный выбор модели и значение фиктивной переменной при неком фиксированном уровне значимости (часто 5%) определяется наименьшим количеством коэффициентов.

# Замечания.

Указанные действия возможны при НЕЗАВИСИМОСТИ коэффициентов между собой. Последнее не справедливо для свободного члена, он всегда коррелирует с коэффициентами наклонов. Аналогично происходит и у коэффициентов с фиктивной переменной. Поэтому не спешите выкидывать β5 (перед чистой дамми), даже если он более незначимый в сравнении с другими коэффициентами с фиктивной переменной.

Используя STATA также последовательное исключение переменных для обоих выборов фиктивной переменой
потребуются всего две команды: импорт из Excel (меню file)и пошаговое исключение (меню Other> Stepwise estimation) каждый регрессор введите в свой Term, уровень значимости в «удаление» установите 0.03.

**(часть2)**

Рассматривается следующая модель данных

Qvi=β0+β1Pvi+β2Pni+β3Ii+ɛi

Qv – месячное потребление водки местного разлива;
Pv – цена водки;
Pn –цена наливки;
I – выплат по зарплате.

Первоначально проводится регрессия по всем переменным. Сопоставляется значимость F-статистики на адекватность регрессии со значимостями все коэффициентов наклона на предмет выявления возможной мультиколлиниарности. Определяетесь с уровнем значимости, на котором будете проводить дальнейшее исследование.

Проверяются все коэффициенты парной корреляции, отмечая сильные зависимости, и подсчитываются все VIF’ы, показывая какие коэффициенты «страдают» от МК.

Обсчитываются всевозможные упрощенные модели с одним, двумя регрессорами. Отбираете из них регрессии со значимыми (по вашему выбору) коэффициентами наклона.

Строится ортогональные квази-собственные векторы, проверяются коэффициенты корреляции для проверки правильной ортогонализации. А этих векторах строится регрессия и упрощается отбрасывая незначимые переменные.

Далее, на втором этапе, собираете расширенную выборку, используя данные ваших друзей (или врагов) и проверяете их на «одинаковость» с помощью теста Чоу. После чего выполняете все те же действия на расширенной выборке. Сравните значимости коэффициентов наклона и адекватности регрессии.

Используя STATA также получите значения VIF’ы. Введите данные (импорт из Excel), постройте регрессию (Statistics > Linear models and related > Linear regression) и сразу после получите VIF’ы (Statistics > Postestimation > Reports and statistics: Variance inflation factors).

**Дополнительно**

Напишите аннотацию (резюме), объемом не превышающим страницу, по результатам обоих ваших исследований.

**Обсуждаемые теоретические вопросы.**

1. Тесты Chow на «устойчивость коэффициентов» и «предсказательную силу».
2. Проверка гипотез на выброшенные коэффициенты и адекватность регрессии.
3. Модель регрессии в векторно-матричной форме. Оценки вектора коэффициентов и их дисперсионно-ковариационная матрица.
4. Нормированный R2.
5. Признаки мультиколлиниарности.
6. Выбор «удачных» моделей.
7. Построение ортогональных векторов.
8. Сопоставление результатов на увеличенной и начальной выборках.
9. VIF.