**Семинар «Модели бинарного выбора» 20 (23) апреля 2015**

**Самостоятельная работа**

Исходные данные приведены в файле **womenwk\_data.dta**

В файле собрана информация о 2000 женщинах, 657 из которых не работают.

Переменные:

wagefull – годовой доход женщины, тыс. долл.;

wage – годовой доход работающей женщины, тыс. долл.;

age – возраст, в годах;

married – состояние в браке (1 – если состоит, 0 – иначе);

children – наличие детей (порядковая переменная, принимает значения от 0 до 5);

education – число полных лет обучения, в годах.

Цель:

Выяснить, какие факторы (среди перечисленных) могут влиять на выбор женщины работать – не работать.

Перед началом работы не забудьте начать запись в файл: File→Log→Begin.

Сначала необходимо сгенерировать зависимую переменную **work**. Она принимает значение 1, если женщина работает, и 0 – если не работает.

Для создания этой переменной выполните следующие действия:

**gen work=0**

**replace work=1 if ( wagefull== wage)**

Рассчитайте описательные характеристики:

**sum age education married children work**

Оцените сначала параметры линейной вероятностной модели:

**reg work age education married children**

Каковы недостатки этой модели? Опишите.

Оцените теперь параметры пробит-модели:

**probit work age education married children**

Дайте содержательную интерпретацию построенной модели. Какие статистики указывают на ее качество?

Рассчитайте предельные эффекты объясняющих переменных:

**mfx compute**

Предельные эффекты можно также рассчитать, используя команду **dprobit**.

Интерпретируйте результаты расчетов.

Какова вероятность того, что женщина, имеющая средние по совокупности характеристики, будет работать?

Рассчитайте предельные эффекты объясняющих переменных, при условии, что у женщины нет детей:

**mfx compute, at ( children=0)**

Какова вероятность того, что женщина, не имеющая детей и обладающая по остальным параметрам средними характеристиками по выборке, будет работать?

Подтвердите свой ответ расчетами вручную.

Постройте сводную таблицу фактических и прогнозных исходов:

**estat classification**

Какова доля корректно предсказанных случаев?

Выведите значения информационных критериев Акаике и Шварца:

**estat ic**

Выполните все аналогичные расчеты для **logit**-модели.

Сделайте выводы.