# Рекомендации по выполнению задания №6

1. Рассматриваются следующие модели данных

Si=β0+β1Mi+β2Pi+β3Ei+ɛi. (UR)

Si-0.5Pi=β0+β1(Mi+Ei)+ɛi. (R)

S – количество слов в ежедневной истории;
M – количество слов, предложенных мартышкой;
E – количество слов, предложенных слоненком;
P – количество слов, предложенных попугаем.

Вначале для первой из этих моделей по первым 23 наблюдениям получите оценки коэффициентов на Excel, а далее с помощью матриц получите значение предсказания и оценку дисперсии для этого предсказания и проверьте, используя t-статистику, качество этого предсказания (Р-значение). Проверьте себя тестом Чоу на предсказательную силу.

После, проверяем гипотезу H0: β1=β3 & β2=0.5 в первоначальной регрессии (без ограничений). Строится регрессия с ограничениями. Из обеих выбираем RSS’ы (=e’iei). Выполняем следующие тесты:
а) проверка с помощью F статистики (знакомая вам)
б) LR (отношение правдоподобия) – n(lnRSSr-lnRSSur)~χ2k (здесь и далее n=24, k=2)
в) W (Вальд) – n(RSSr-RSSur)/RSSur~χ2k
г) LM(множители Лагранжа) – на остатках от регрессии с ограничениями обсчитываем модель:
eir= ƴ0+ƴ1Mi+ƴ2Pi+ƴ3Ei+υi, и проверяем гипотезу H0:R2=0, с помощью nR2~χ2k.
также LM~ n(RSSr-RSSur)/RSSr, проверьте
Сделайте сначала на Excel (б) и (в) еще и через матрицы

Матричный метод в общем виде для модели Y=X+ и ограничения Q=q (р-строк, в нашем случае р=2):
оценка без ограничений b=(X’X)-1X’Y, дисперсия V(b)=^2(X’X)-1
дисперсия ограничения V(Qb)=(QV(b)Q’)-1
W=(Qb-q)’V(Qb)(Qb-q) ~ p
множители Лагранжа =V(Qb)(Qb-q), тест LM=’QV(b)Q’ ~ p  (сначала получите Q’, а затем «обкладывайте» дисперсию)

Далее (б) и (в) в Stata:

(б) Используете регрессию (М, Р, Е переменные) из меню, запоминаете результат , затем также для М+Е и далее LR тест (см. например Демидова стр.143)
reg S M P E
estimates store REG1
gen ME = M+E
Gen S1=S-0.5\*P
**reg S ME**
reg S1 ME
estimates store REG2
lrtest REG1 REG2, force с этим сработает

(в) начальная регрессия и тест > линейная гипотеза из postestimation для Вальда (см. например Демидова стр.130)
estimates restore REG1 или reg S M P E
test (M=E) (P=0.5)

Сопоставьте получившиеся результаты проверки тестов.

2. Две модели:
P(D=1|Si)=β0+β1Si+ɛi, (левая часть это просто столбец дамми)
P(D=1|Ei)=β0+β1Pi+ɛi

Сначала исследуется модель линейной вероятности. Не забудьте нарисовать график подбора (поставить соответствующую галочку в Excel).
Эти же модели для взвешенного случая ст.ош.=корень(Y^(1-Y^),
если Y^>0.99, то 0.99, если <0.01, то 0.01

Далее, используя Stata (только здесь) обсчитайте те же зависимости для логит, P(O=1|Xi)=ln(li/(1-li), li=β0+β1Xi+ɛi, X здесь и далее – соответственно S и E, (команда logit O X или в меню Statistics > Binary outcomes > Logistic regression)
и пробит P(O=1|Xi)=Ф(zi), zi=β0+β1Xi+ɛi, Ф(z)-интегральная функция вероятности нормального распределения, (команда probit O X или в меню Statistics > Binary outcomes > Probit regression) моделей (см. например Демидова стр.236-238).

Для средних значений вычислите предельные эффекты для обеих моделей.
margins, dydx(X) at((mean) X)

Постройте график значений P(O=1|X) на одном графике значения наблюдений, пробит и логит функции (наблюдения – точки, вычисленные значения – линии), отдельно для S и D.

Используя STATA проведите ROC анализ.
Сразу после команды каждой из команд probit или logit
Statistics > Binary outcomes > Postestimation > ROC curve after logistic/logit/probit/ivprobit
например:
probit O P
dydx(P) at((mean) P)
estat classification
lsens
lroc

Напишите аннотацию (резюме), общим объемом, не превышающим половину страницы, по результатам ваших исследований.

Кроме того напишите Пояснительную записку, в которой описываете ваши действия, выводы и пр. Объем любой, распечатки результатов можете утопить в тексте.

# Обсуждаемые теоретические вопросы.

1. Метод максимального правдоподобия. Принципиальная особенность. Свойства.
2. LR, W, LM тесты, постановка задачи, вычислительные формулы.
3. Линейная модель бинарного вывода. Проблемы.
4. Модели логит, пробит. Постановка задачи, предельные эффекты.
5. ROC кривые.