

## Семинар 6

### Гетероскедастичность и нетипичные наблюдения

**Задание 1.** После оценивания линейной регрессионной модели  $y$  на предикторы  $x$ ,  $z$ ,  $d$  и  $w$  по 550 наблюдениям в критерии Уайта, используемом для проверки гетероскедастичности, было получено значение  $R^2$  из вспомогательной регрессии равное 0.4. Сформулируйте нулевую и альтернативную гипотезы, запишите спецификацию дополнительной (вспомогательной) модели, рассчитайте статистику критерия и сделайте вывод на основании  $p$ -value.

**Задание 2.** По данным 120 стран исследователь оценил зависимость  $Y$  от двух объясняющих переменных:  $X$  и  $Z$ . Исследователь предполагает, что вариация ошибок зависит пропорционально от  $X$ . Чтобы проверить это предположение, он упорядочил наблюдения по возрастанию  $X$ , исключил из середины выборки 30% наблюдений и оценил регрессии на первом и втором сегментах выборки с наименьшим и наибольшим  $X$ . Величины  $RSS$  для этих регрессий равны 2500 и 2125 соответственно. Рассчитайте статистику критерия Голдфелда-Квандта и соответствующее значение  $p$ -value. Какой вывод сделает исследователь на основе представленных результатов?

**Задание 3.** По указанным ниже значениям предиктора и отклика ( $x$  и  $y$  соответственно) в регрессионной модели  $y$  на  $x$  найдите матрицу проекции (hat matrix). Какие наблюдения по  $x$  обладают наибольшим потенциалом влияния?

$x$	2	1	0	-1
$y$	1	0	0	-3