

## Семинар 5.

1. Рассмотрим модель  $y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + \beta_4 x_{i3} + \varepsilon_i$ . При оценке модели по 24 наблюдениям оказалось, что  $RSS = 15$ ,  $\sum (y_i - \bar{y} - x_{i2} + \bar{x}_2)^2 = 20$ . На уровне значимости 1% протестируйте гипотезу

$$H_0 : \begin{cases} \beta_2 + \beta_3 + \beta_4 = 1 \\ \beta_2 = 0 \\ \beta_3 = 1 \\ \beta_4 = 0 \end{cases}.$$

2. Пусть  $y = X\beta + \varepsilon$  — регрессионная модель, где  $X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ ,  $y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$ ,

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}, E(\varepsilon) = 0, Var(\varepsilon) = \sigma^2 I.$$

На уровне значимости 5% проверьте гипотезу  $H_0 : \beta_1 + \beta_2 = 2$  против альтернативной  $H_a : \beta_1 + \beta_2 \neq 2$ :

- Приведите формулу для тестовой статистики.
  - Укажите распределение тестовой статистики при верной  $H_0$ .
  - Вычислите наблюдаемое значение тестовой статистики.
  - Укажите границы области, где основная гипотеза не отвергается.
  - Сделайте статистический вывод.
3. На основе опроса 25 человек была оценена следующая модель зависимости логарифма зарплаты ( $\ln W$ ) от уровня образования ( $Edu$ , в годах) и возраста ( $Age$ ).

$$\widehat{\ln W} = 1.7 + 0.5Edu + 0.06Age - 0.0004Age^2, \\ ESS = 90.3, RSS = 60.4.$$

Когда в модель были введены переменные  $Fedu$  и  $Medu$ , учитывающие уровень образования родителей, величина  $ESS$  увеличилась до 110.3.

- (a) Напишите спецификацию уравнения регрессии с учетом образования родителей.
  - (b) Сформулируйте и проверьте гипотезу о значимом влиянии уровня образования родителей на зарплату (уровень значимости 5%).
4. Рассмотрим следующую модель зависимости цены дома  $Price$  (в тысячах долларов), от его площади  $Hsize$  (в  $m^2$ ), площади участка  $Lsize$  (в  $m^2$ ), числа ванных комнат  $Bath$  и числа спален  $BDR$ :

$$\widehat{Price} = \hat{\beta}_1 + \hat{\beta}_2 Hsize + \hat{\beta}_3 Lsize + \hat{\beta}_4 Bath + \hat{\beta}_5 BDR, R^2 = 0.218, n = 23.$$

Напишите спецификацию регрессии с ограничениями для проверки статистической гипотезы:  $H_0 : \beta_4 = 20\beta_5$ . Дайте интерпретацию проверяемой гипотезе. Для регрессии с ограничениями был вычислен коэффициент  $R_R^2 = 0.136$ . Протестируйте нулевую гипотезу на уровне значимости 5%.

5. В файле `dataflats.xlsx` хранятся данные о стоимости квартир в Москве (тыс.долл.).

- (a) Оцените следующие модели регрессии для стоимости одного квадратного метра жилья:

$$\begin{aligned} price\_sq_i &= \beta_1 + \beta_2 livesp_i + \beta_3 dist_i + \varepsilon_i, \\ price\_sq_i &= \beta_1 + \beta_2 livesp_i + \beta_3 dist_i + \beta_4 metrdist_i + \varepsilon \end{aligned}$$

- (b) Для построенных моделей проверьте гипотезу о незначимости модели в целом.
- (c) Используя  $p$ -value коэффициентов, укажите для каждой из моделей, какие из переменных являются значимыми, а какие – незначимыми?
- (d) Проинтерпретируйте оценки коэффициентов при значимых переменных. Сходятся ли знаки данных оценок с интуицией и здравым смыслом?
- (e) Постройте 90%-ые доверительные интервалы для коэффициентов обеих моделей.
- (f) Для каждой из моделей проверьте гипотезу  $H_0 : \beta_3 = -0.1$ . Содержательно проинтерпретируйте результаты тестирования.
- (g) Для второй модели проверьте гипотезу  $H_0 : \beta_2 + \beta_4 = 0$ . Содержательно проинтерпретируйте результаты тестирования.
- (h) Переоцените модели регрессии на шакалированных данных. Какой из факторов оказывает наибольшее влияние на стоимость квартиры?
- (i) Какие из трех оцененных моделей могут быть сравнены по значению коэффициента детерминации  $R^2$ ? Выполните сравнение.

6. Домашнее задание. [Борзых Д.А., Вакуленко Е.С., Фурманов К.К. Эконометрика: РАБОТА С ДАННЫМИ НА КОМПЬЮТЕРЕ. ПРАКТИКУМ: Элементы теории. Практические задания. Ответы и решения].

Оценивается зависимость количества продаваемых чебуреков (в штуках) –  $qch$  от цены на чебуреки (в рублях) –  $pch$ , цены на шаурму (в рублях) –  $psh$  и цены на мороженое (в рублях) –  $pmor$  в виде линейной регрессии:

$$qch_i = \beta_1 + \beta_2 pch_i + \beta_3 psh_i + \beta_4 pmor_i + \varepsilon_i, i = 1, \dots, n.$$

В файле "*Regression\_9.xlsx*" приведены данные. Выполните следующие задания. Используйте 5%-ый уровень значимости.

- (a) Оцените данное уравнение регрессии и выпишите оцененное значение.
- (b) Является ли полученное уравнение регрессии значимым?
- (c) Используя  $p$  – *value* коэффициентов, укажите, какие из переменных являются значимыми, а какие – незначимыми?
- (d) Проинтерпретируйте оценки коэффициентов при значимых переменных. Согласуются ли знаки данных оценок с интуицией и здравым смыслом?
- (e) Протестируйте гипотезу  $H_0 : \beta_2 = -5$  против альтернативной гипотезы  $H_0 : \beta_2 - 5$ .
- (f) Протестируйте гипотезу  $H_0 : \beta_2 = -5$  против альтернативной гипотезы  $H_0 : \beta_2 < -5$ .
- (g) Протестируйте гипотезу  $H_0 : \beta_2 = -5$  против альтернативной гипотезы  $H_0 : \beta_2 > -5$ .
- (h) Проверьте гипотезу: "чем выше цена на шаурму, тем больше в среднем продается чебуреков". Подразумевается, что все прочие условия являются неизменными.