

Семинар 5.

Блочные матрицы и проверка гипотез.

Умножение блочных матриц. Если размеры блоков допускают операцию умножения, то:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \cdot \left[\begin{array}{c|c} E & F \\ \hline G & H \end{array} \right] = \left[\begin{array}{c|c} AE + BG & AF + BH \\ \hline CE + DG & CF + DH \end{array} \right].$$

Формула Фробениуса (блочное обращение).

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]^{-1} = \left[\begin{array}{c|c} A^{-1} + A^{-1}BH^{-1}CA^{-1} & -A^{-1}BH^{-1} \\ \hline -H^{-1}CA^{-1} & H^{-1} \end{array} \right],$$

где A — невырожденная квадратная матрица размерности $n \times n$, D — квадратная матрица размерности $k \times k$, $H = D - CA^{-1}B$.

Задание 1. Для блочной матрицы

$$\left[\begin{array}{c|c} A & B \\ \hline 0 & C \end{array} \right]$$

найдите обратную матрицу.

Предположения: $\det A_{n \times n} \neq 0$; $\det C_{k \times k} \neq 0$.

Задание 2. Рассмотрим модель множественной регрессии:

$$Y_i = \beta_1 + \beta_2 X_{i1} + \dots + \beta_k X_{ik-1} + \varepsilon_i.$$

Мы хотим проверить гипотезу следующего вида:

$$H_0 : \beta_{k-q+1} = \beta_{k-q+2} = \dots = \beta_k = 0.$$

Покажите, что для проверки гипотезы H_0 может быть использована статистика

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)} \sim F(q, n-k).$$

Задание 3. 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

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

Задание 4. В файле *dataflats.xlsx* хранятся данные о стоимости квартир в Москве (тыс.долл.).

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

$$price_sq_i = \beta_1 + \beta_2 livesp_i + \beta_3 dist_i + \varepsilon_i,$$

$$price_sq_i = \beta'_1 + \beta'_2 livesp_i + \beta'_3 dist_i + \beta'_4 metr dist_i + \varepsilon_i.$$

- (b) Для оцененных 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?
- (e) Постройте 90%-ые доверительные интервалы для коэффициентов обеих моделей.
- (f) Для каждой из моделей проверьте гипотезу о том: что расстояние до центра незначимо. 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? Вспомните выводы о соотношении оценок регрессии для исходных переменных и регрессии для стандартизированных показателей, полученные на предыдущем семинаре. Сравните теоретические выводы с полученными практическими результатами.