Доп. занятие 2.

1. Доклад Никиты Чуйкина про информационные критерии.

2. Рассмотрим НКЛММР $Y_i = \beta_1 + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i$ с неслучайными регрессорами. Дополнительно известно, что на самом деле $\beta_2 = ... = \beta_k = 0$.

   (a) Найдите $\mathbb{E}(R^2)$.

   (b) Найдите $\mathbb{E}(R^2_{adj.})$.

3. (Универсиада по эконометрике, МГУ, 2016 год). Имеется временной ряд:

$$y_i = \theta \cdot i + \varepsilon_i + \varepsilon_0,\ i = 1,...,n,$$

   где $\mathbb{E}(\varepsilon_i) = 0, \mathbb{E}(\varepsilon_i^2) = \sigma^2$, $\mathbb{E}(\varepsilon_i \varepsilon_j) = 0$ при $i \neq j$, $i,j = 0,1,...,n$.

   (a) Вычислите дисперсию МНК–оценки параметра θ.

   (б) Будет ли эта оценка из пункта (а) состоятельной?

   (в) Будет ли она эффективной?

   (г) Предложите метод для получения эффективной оценки θ.

   (д) Пусть $n = 4$ и известно, что $y_1 = -1$, $y_2 = 4$, $y_3 = 6$, $y_4 = 8$. Вычислите эффективную оценку $\hat{\theta}$.

4. (Задача из семинара 9). Теоретическая регрессионная зависимость и выборочная корреляционная матрица центрированно-нормированных регрессоров X имеют вид:

$$y_i = \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i,$$

$$\widehat{Corr}(X) = \begin{pmatrix} 1 & 0.95 & 0 \\ 0.95 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

   (a Вычислите все главные компоненты. Сколько главных компонент надо выбрать, чтобы они объясняли не менее 70% общей дисперсии?

   (b) Вычислите матрицу факторной нагрузки. Проинтерпретируйте полученные результаты.

   - Процесс, порождающий данные, описывается уравнением

$$y_i = \beta x_i + \varepsilon_i,$$

$$\mathbb{E}(\varepsilon_i) = 0,\ \mathbb{E}(\varepsilon_i^2) = \sigma^2,\ \mathbb{E}(\varepsilon_i \varepsilon_j) = 0,\ i \neq j,\ i = 1,...,n.$$

Экспериментатор не имеет доступа к исходным данным, а может использовать лишь "групповые"данные. А именно, значения независимой переменной упорядочиваются по величине ($x_1 < x_2 < ... < x_n$), вычисляются средние значения в первой группе из n_1 наблюдений

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i, \bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i,$$

по второй группе — из n_2 наблюдений

$$\bar{x}_2 = \frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2} x_i, \bar{y}_2 = \frac{1}{n_2}\sum_{i=n_1+1}^{n_1+n_2} y_i,$$

и т.д. Всего есть J групп наблюдений, j-я группа имеет n_j наблюдений. Параметр β оценивается с помощью регрессии $\bar{y}_j$ на $\bar{x}_j$, $j = 1,...,J$. Вычислите среднее значение и дисперсию оценки. Оцените потерю эффективности в результате такой группировки данных.