Семинар 3.
Модель множественной регрессии.
Матрицы: начало.

1. (еще немного о парной регрессии) Покажите, что для модели парной регрессии с константой $R^2 = \widehat{\rho}^2(y_i, \hat{y}_i)$.

Решение:
Напомним, что $R^2 = \frac{ESS}{TSS}$, а выборочная корреляция считается по формуле:

$$\hat{\rho}(y_i, \hat{y}_i) = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{(\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}.}$$

Вспомним, что для модели парной регрессии с константой $\bar{y} = \bar{\hat{y}}$, а также, что для любой модели регрессии $TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2, ESS = (\hat{y}_i - \bar{y})^2$. Тогда знаменатель выборочной корреляции в формуле выше равен $\sqrt{TSS}\sqrt{ESS}$. Осталось разобраться с числителем дроби. Заметим, что для любой модели регрессии справедливо разложение $y_i = \hat{y}_i + e_i$. Тогда

$$\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) = \sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y}) = \sum_{i=1}^{n}(\hat{y}_i - \bar{y} + e_i)(\hat{y}_i - \bar{\hat{y}}) =$$

$$= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} e_i(\hat{y}_i - \bar{y}).$$

Заметим, что если в модель включена константа, то вектор остатков e и вектор $y - \bar{y}\vec{i}$ (здесь $\bar{y}$ — число) ортогональны друг другу, следовательно, их скалярное произведение равно нулю, то есть $\sum_{i=1}^{n} e_i(\hat{y}_i - \bar{y}) = 0$. Это справедливо толкьо при наличии константы в модели регрессии, так как в этом случае единичный вектор $\vec{i}$ принадлежит пространству, образованному регрессорами модели.

Тогда для числителя получаем, что

$$\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = ESS.$$

В итоге

$$\hat{\rho}^2(y, \hat{y}) = \left(\frac{ESS}{\sqrt{TSS}\sqrt{ESS}}\right)^2 = \frac{ESS}{TSS} = R^2.$$

Данный результат справедлив и для модели множественной регрессии с константой. В модели регрессии без константы коэффициент детерминации не используется для оценка качества модели регрессии.

Немного теории.

Пусть r — случайный вектор размерности $[n\times 1]$, s — случайный вектор размерности $[k \times 1]$, A и b — неслучайные матрица и вектор соответственно, имеющие подходящие размерности.

Математическим ожиданием случайного вектора r называется вектор

$$\mathbb{E}(r) = \begin{pmatrix} \mathbb{E}(r_1) \\ \mathbb{E}(r_2) \\ \dots \\ \mathbb{E}(r_n) \end{pmatrix}.$$

Ковариационная матрица вектора r определяется следующим образом:

$$\mathbb{V}\text{ar}(r) = \begin{pmatrix} \mathbb{C}\text{ov}(r_1, r_1) & \mathbb{C}\text{ov}(r_1, r_2) & \dots & \mathbb{C}\text{ov}(r_1, r_n) \\ \mathbb{C}\text{ov}(r_2, r_1) & \mathbb{C}\text{ov}(r_2, r_2) & \dots & \mathbb{C}\text{ov}(r_2, r_n) \\ \dots & & \dots & \dots \\ \mathbb{C}\text{ov}(r_n, r_1) & \mathbb{C}\text{ov}(r_n, r_2) & \dots & \mathbb{C}\text{ov}(r_n, r_n) \end{pmatrix}.$$

Ковариационная матрица векторов r и s определяется следующим образом:

$$\mathbb{C}\text{ov}(r, s) = \begin{pmatrix} \mathbb{C}\text{ov}(r_1, s_1) & \mathbb{C}\text{ov}(r_1, s_2) & \dots & \mathbb{C}\text{ov}(r_1, s_k) \\ \mathbb{C}\text{ov}(r_2, s_1) & \mathbb{C}\text{ov}(r_2, s_2) & \dots & \mathbb{C}\text{ov}(r_2, s_k) \\ \dots & \dots & \dots & \dots \\ \mathbb{C}\text{ov}(r_n, s_1) & \mathbb{C}\text{ov}(r_n, s_2) & \dots & \mathbb{C}\text{ov}(r_n, s_k) \end{pmatrix}.$$

Свойства вектора математических ожиданий и ковариационной матрицы:

(a) $\mathbb{E}(Ar + b) = A\,\mathbb{E}(r) + b$

(b) $\mathbb{C}\text{ov}(r, s) = \mathbb{E}(rs^T) - \mathbb{E}(r)\,\mathbb{E}(s^T)$

(c) $\mathbb{C}\text{ov}(Ar + b, s) = A\,\mathbb{C}\text{ov}(r, s)$

(d) $\mathbb{C}\text{ov}(r, As + b) = \mathbb{C}\text{ov}(r, s)A^T$

(e) $\mathbb{V}\text{ar}(r) = \mathbb{C}\text{ov}(r, r) = \mathbb{E}(rr^T) - \mathbb{E}(r)\,\mathbb{E}(r^T)$

(f) $\mathbb{V}\text{ar}(Ar + b) = A\,\mathbb{V}\text{ar}(r)A^T$

(g) $\mathbb{E}(r^TAr) = \text{trace}(A\,\mathbb{V}\text{ar}(r)) + \mathbb{E}(r^T)A\,\mathbb{E}(r)$

(h) Если вектора r и s имеют одинаковый размер, то $\mathbb{V}\mathrm{ar}(r+s) = \mathbb{V}\mathrm{ar}(r) + \mathbb{V}\mathrm{ar}(s) + \mathbb{C}\mathrm{ov}(r,s) + \mathbb{C}\mathrm{ov}(s,r)$

Условные ожидание и дисперсия определяются аналогично и обладают аналогичными свойствами.

Вспомним несколько свойств следа:

(a) $\mathrm{trace}(\alpha A+\beta B) = \alpha\,\mathrm{trace}(A)+\beta\,\mathrm{trace}(B)$, α,β — $const$,A,B — произвольные матрицы;

(b) $\mathrm{trace}(ABC) = \mathrm{trace}(CAB) = \mathrm{trace}(BCA)$, A,B,C — произвольные матрицы;

(c) $\mathrm{trace}(u) = u$, где u — скаляр.

2. (Математическое ожидание квадратичной формы) Пусть y — случайный вектор размерности $[n\times 1]$, а A — неслучайная матрица подходящей размерности. Докажите, что справедлива следующая формула для математического ожидания квадратичной формы:

$$\mathbb{E}(y^TAy) = \mathrm{trace}(A\,\mathbb{V}\mathrm{ar}(y)) + \mathbb{E}(y^T)A\,\mathbb{E}(y).$$

Решение:
Распишем математическое ожидание, используя свойства следа:

$$\mathbb{E}(y^TAy) = \mathbb{E}(\mathrm{trace}(y^TAy)) = \mathbb{E}(\mathrm{trace}(Ayy^T)) = \mathbb{E}(\mathrm{trace}(Ayy^T)) =$$

$$= \mathrm{trace}\,\mathbb{E}(Ayy^T) = \mathrm{trace}(A\,\mathbb{E}(yy^T)).$$

Вспомним, что

$$\mathbb{V}\mathrm{ar}(y) = \mathbb{E}(yy^T) - \mathbb{E}(y)\,\mathbb{E}(y^T),$$

откуда получаем:

$$\mathbb{E}(y^TAy) = \mathrm{trace}(A\,\mathbb{E}(yy^T)) = \mathrm{trace}\,A(\mathbb{V}\mathrm{ar}(y) + \mathbb{E}(y)\,\mathbb{E}(y^T)) =$$

$$= \mathrm{trace}(A\,\mathbb{V}\mathrm{ar}(y)) + \mathrm{trace}(A\,\mathbb{E}(y)\,\mathbb{E}(y^T)) = \mathrm{trace}(A\,\mathbb{V}\mathrm{ar}(y)) + \mathrm{trace}(\mathbb{E}(y^T)A\,\mathbb{E}(y)) =$$

$$= \mathrm{trace}(A\,\mathbb{V}\mathrm{ar}(y)) + \mathbb{E}(y^T)A\,\mathbb{E}(y).$$

3. Предположим, что все предпосылки теоремы Гаусса — Маркова выполнены. Используя матрицы $\mathrm{P} = X(X^TX)^{-1}X^T$ и $\pi = \vec{i}(\vec{i}^T\vec{i})^{-1}\vec{i}^T$, запишите матричные представления для TSS, ESS и RSS и вычислите их математические ожида-

ния.

Решение:

Прежде всего, необходимо помнить, что матрицы-проекторы P и $M = I - P$ — симметричные и идемпотентные. Матрица π — это частный случай матрицы-проектора P, когда $X = \vec{i}$. Получим матричное представление для TSS, ESS и RSS:

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2 = (y - \bar{y}\vec{i})^T (y - \bar{y}\vec{i}) =$$

$$= (y - \pi y)^T (y - \pi y) = ((I - \pi)y)^T ((I - \pi)y) = y^T (I - \pi)^T (I - \pi) y = y^T (I - \pi) y,$$

$$ESS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = (\hat{y} - \bar{y}\vec{i})^T (\hat{y} - \bar{y}\vec{i}) =$$

$$= (Py - \pi y)^T (Py - \pi y) = (y(P - \pi)y)^T ((P - \pi)y) = y^T (P - \pi)^T (P - \pi) y = y^T (P - \pi) y,$$

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = (y - \hat{y})^T (y - \hat{y}) =$$

$$= (y - Py)^T (y - Py) = ((I - P)y)^T (y(I - P)) = y^T (I - P)^T (I - P) y = y^T (I - P) y.$$

Таким образом, имеем:

$$TSS = y^T (I - \pi) y,$$

$$ESS = y^T (P - \pi) y,$$

$$RSS = y^T (I - P) y.$$

Заметим, что все три показателя представлены в виде квадратичной формы. В терминах предыдущей задачи матрица A равна $(I - \pi)$ для TSS, $(P - \pi)$ для ESS и $(I - P)$ для RSS. Здесь матрица I имеет размерность $[n \times n]$.

Предварительно заметим, что

$$trace(I) = n, \quad trace(P) = \mathrm{trace}(X(X^TX)^{-1}X^T) = trace((X^TX)(X^TX)^{-1}) = trace(I_k) = k.$$

Из определения матрицы π

$$\pi = \frac{1}{n} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

нетрудно убедиться, что $\mathrm{trace}(\pi) = 1$.

Используя формулу для математического ожидания квадратичной формы, полученную в предыдущем упражнении, можно легко посчитать интересующие нас математические ожидания.

Начнём с $\mathbb{E}(TSS)$:

$$\mathbb{E}(TSS) = \mathbb{E}(y^T(I-\pi)y) = \text{trace}((I-\pi)\,\mathbb{V}\text{ar}(y)) + \mathbb{E}(y^T)(I-\pi)\,\mathbb{E}(y) =$$

$$= \text{trace}(I-\pi)\sigma^2 I) + (X\beta)^T(I-\pi)(X\beta) = \sigma^2\,\text{trace}(I-\pi) + \beta^T X^T(I-\pi)X\beta =$$

$$= \sigma^2(\text{trace}(I) - \text{trace}(\pi)) + \beta^T X^T(I-\pi)X\beta = (n-1)\sigma^2 + \beta^T X^T(I-\pi)X\beta.$$

Проделаем аналогичные вычисления для $\mathbb{E}(ESS)$

$$\mathbb{E}(ESS) = \mathbb{E}(y^T(P-\pi)y) = \text{trace}((P-\pi)\,\mathbb{V}\text{ar}(y)) + \mathbb{E}(y^T)(P-\pi)\,\mathbb{E}(y) =$$

$$= \text{trace}((P-\pi)\sigma^2 I) + (X)^T(P-\pi)(X\beta) = \sigma^2\,\text{trace}(P-\pi) + \beta^T X^T(P-\pi)X\beta =$$

$$= \sigma^2(\text{trace}(P) - \text{trace}(\pi)) + \beta^T X^T(P-\pi)X\beta = (k-1)\sigma^2 + \beta^T X^T(P-\pi)X\beta$$

и для $\mathbb{E}(RSS)$

$$\mathbb{E}(RSS) = \mathbb{E}(y^T(I-P)y) = \text{trace}((I-M)\,\mathbb{V}\text{ar}(y)) + \mathbb{E}(y^T)(I-P)\,\mathbb{E}(y) =$$

$$= \text{trace}(I-P)\sigma^2 I) + (X\beta)^T(I-P)(X\beta) = \sigma^2\,\text{trace}(I-P) + \beta^T X^T(I-P)X\beta =$$

$$= \sigma^2(\text{trace}(I) - \text{trace}(P)) + \beta^T X^T(I-P)X\beta = (n-k)\sigma^2 + \beta^T X^T(I-P)X\beta = (n-k)\sigma^2,$$

так как (I-P)X $=$ MX $=$ 0.

Таким образом, получаем

$$\mathbb{E}(TSS) = (n-1)\sigma^2 + \beta^T X^T(I-\pi)X\beta,$$

$$\mathbb{E}(ESS) = (k-1)\sigma^2 + \beta^T X^T(H-\pi)X\beta,$$

$$\mathbb{E}(RSS) = (n-k)\sigma^2.$$

Заметим, что из $\mathbb{E}(RSS) = (n-k)\sigma^2$ можно получить несмещённую оценку для дисперсии случайной ошибки для случая множественной регрессии:

$$\hat{\sigma}^2 = \frac{RSS}{(n-k)}.$$

Важное примечание: k — общее число параметров модели, включая константу, если она предусмотрена в модели!