

Задачи к занятию 3

Задача 1. Имеется набор точек. Аппроксимируйте их линейной зависимостью $y = a + bx$ в MS Excel, используя линию тренда на точечной диаграмме, пакет анализа данных (если есть) и формулы коэффициентов регрессии:

$$\begin{cases} b = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \\ a = \bar{y} - b\bar{x} \end{cases}$$

Также рассчитайте RSS (сумму квадратов отклонений), TSS (общая сумма квадратов), R^2 (коэффициент детерминации) и F , пользуясь соотношениями

$$F_{\text{эмп}} = \frac{ESS/f_1}{TSS/f_2} = \frac{R^2}{1 - R^2} \frac{f_2}{f_1}; \quad f_2 = n - k; \quad f_1 = k - 1$$

$$R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}$$

и проверьте статистическую значимость регрессии.

Совет: используйте функции MS Excel КОВАР, ДИСП.Г, СРЗНАЧ, СУММКВРАЗН, СУММКВ.

x	10	8	13	9	11	14	6	4	12	7	5
y	8.04	6.95	7.58	8.81	8.33	9.96	7.24	4.26	10.84	4.82	5.68

Задача 2. Имеются экспериментальные данные по энтальпии смешения бинарного раствора (см. таблицу). Аппроксимируйте их моделью субрегулярного раствора $\Delta_{\text{mix}}H(x) = x(1-x)(A+Bx)$. Результат проиллюстрируйте графиком.

Совет: линеаризуйте данные, используя подстановку

$$A + Bx = \frac{\Delta_{\text{mix}}H}{x(1-x)}$$

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\Delta_{\text{mix}}H$	981	1576	2262	2885	2997	2840	2561	2020	1122

Задача 3. В GNU Octave сгенерировать выборку из 500 точек с координатами X_i , Y_i и Z_i . $X_i \in [0; 1]$, $Y_i \in [0; 1]$ и подчиняются равномерному распределению, $Z_i = 3 + 4X_i + 5Y_i + N(0; 1)$. Найти коэффициенты регрессии $\hat{\beta}$, их ковариационную матрицу $\text{cov}(\hat{\beta}, \hat{\beta})$ и доверительные интервалы. Рассчитать TSS , RSS , R^2 и F . Результат проиллюстрировать графиком.

Экспортировать данные в формат csv (функция `csvwrite`), загрузить их в MS Excel и провести регрессионный анализ с помощью пакета анализа данных.