

为了对今后在 2025、2050 年各个州未发生政策改变的情况下对其能源概况进行预测，我们基于历史数据，对历史数据的各种能源分别进行线性回归，从而达到预测要求。

下面我们对符号进行定义，设 X 值为自变量，在本题中， X 为年份；设 Y 为因变量，即需要预测的能源种类的值， a 和 b 为相关系数。线性回归表达式为：

$$Y = aX + b$$

为求得 a, b ，我们需要求解

$$\min Q(a, b) = \sum_{i=1}^n (Y_i - (aX_i + b))^2$$

化简得到：

$$Q(a, b) = n\bar{Y}^2 - 2an\bar{X}\bar{Y} - 2bn\bar{Y} + a^2n\bar{X}^2 + 2abn\bar{X} + nb^2$$

为求得 $Q(a, b)$ 的最小值，分别对 Q 求 a 和 b 的偏导数，令其为 0。

$$\frac{\partial Q}{\partial a} = -2n\bar{X}\bar{Y} + 2an\bar{X}^2 + 2bn\bar{X} = 0 \quad \frac{\partial Q}{\partial b} = -2n\bar{Y} + 2an\bar{X} + 2nb = 0$$

从而得出：

$$a = \frac{\bar{X}\bar{Y} - \bar{X}\bar{Y}}{(\bar{X})^2 - \bar{X}^2} b = \bar{Y} - a\bar{X}$$

那么如何评价结果的好坏呢？我们可以通过 R^2 (coefficient of determination) 来判断回归方程的拟合程度。

The total sum of squares (proportional to the variance of the data):

$$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

The regression sum of squares, also called the explained sum of squares

$$SSR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

The sum of squares of residuals, also called the residual sum of squares:

$$SSE = \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 = \sum_{i=1}^n e_i^2$$

The most general definition of the coefficient of determination is

$$R^2 = 1 - \frac{SSE}{SST}$$

最后，我们得出如下结果：



