1. Summary套路

- 为了XXX（总题干中的要求：制定一系列发展四个州洲际能源契约的目标），我们团队对数据进行了分析和预测。然后，我们XXX（吹水部分的题目：为州长提出了三项行动以实现能源契约的目标），具体来说：
- 第一部分：每一小问的做法+结果
- 第二部分：每一小问的做法+结果
- 第三部分：改变一种说法，复述题目

2. Data Processing套路

- First, we classify the data in the worksheet into three categories: redundant data, invalid data, and normal data. Then, as for redundant data, we remove them to make it more convenient for us to search the required data. As for invalid data, we delete them from the worksheet. As for normal data, we sort them and learn the meaning of their “指标名称”. The works we have done in data processing are as follows.
- redundant data：列出来
- invalid data：列出来

3. 第一部分

1）A小问（可视化分析，532~561行）：

- 能源生产总量和能源消耗总量：列表+分析（AZ和CA为供大于求的州，NM为供不应求的州，TX为供需平衡的州）
- 能源价格：列表+分析
- 能源结构：能源分成交通、商业、电力、工业和住宅五种用途，用它们所占总能源的比例画图+分析

2）B小问（回归预测模型）

- 画图，绘制能源消耗随时间变化的曲线图和能源结构（可再生资源、不可再生资源）随时间变化的柱形图（**感觉在凑字数**），再分析
- 以人口（人口角度）、GDP增长速率（经济角度）、电力现代化指数（工业角度）、是否近海（地理角度）的对数为自变量，可再生能源的比例为应变量，建立回归模型。求出参数，再根据实际情况分析模型的正确性

$$y_{it} = \beta_1 \ln x1_{it} + \beta_2 \ln x2_{it} + \beta_3 \ln x3_{it} + \beta_4 \ln x4_{it} + \boldsymbol{\varepsilon_{it}} \tag{1}$$

In the equation (1), $\boldsymbol{y_{it}}$ is the proportion of renewable energy consumption during the period of $\boldsymbol{t}$ in state $\boldsymbol{i}$, and $\boldsymbol{x1_{it}}$, $\boldsymbol{x2_{it}}$ and $\boldsymbol{x3_{it}}$ are the population during the period of $\boldsymbol{t}$ in state $\boldsymbol{i}$, the growth rate of GDP and the Grid Modernization Index respectively. We assign $\boldsymbol{\alpha_i = 1}$ when state $\boldsymbol{i}$ is close to the ocean. $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, $\boldsymbol{\beta_3}$ and $\boldsymbol{\beta_4}$ are the coefficients and $\boldsymbol{\varepsilon_{it}}$ is the residual term. Since the GDP data started in 1977, we all select the data after 1978 to get the result:

$$y_{it} = 0.023^{***} \ln x1_{it} + 0.002 \ln x2_{it} + 0.026^{***} \ln x3_{it} - 0.038\alpha_i - 0.073^{***}.$$

- 上述模型中地理因素不够显著，文字分析为什么地理因素对使用可再生能源有贡献（**针对模型中考虑不充分的地方，凑字数**）

3）C小问（TOPSIS和PROMETHEE评价类模型）

- 自定义多属性（查阅论文）：可再生能源潜力、能源结构、可再生能源比例和经济转化率
- 多属性决策：TOPSIS和PROMETHEE两种多属性决策方法合起来用

Let $\mathbf{S} = [\boldsymbol{CA}\ \ \boldsymbol{AZ}\ \ \boldsymbol{NM}\ \ \boldsymbol{TX}]$ be the set of the states, $\mathbf{A} = [\boldsymbol{A_1}\ \ \boldsymbol{A_2}\ \ \boldsymbol{A_3}\ \ \boldsymbol{A_4}]$ be the set of evaluation indexs, $\mathbf{X} = (\boldsymbol{x_{ij}})_{m\times n}$ be the decision matrix, where $\boldsymbol{x_{ij}}$ denotes the value of state $\boldsymbol{i}$ in index $\boldsymbol{j}$.

Step1. Standardized Processing

●To eliminate different dimensions of different indexes, before data analysis, we first conduct Standardized Processing:

$$\boldsymbol{x'}_{ij} = \frac{\boldsymbol{x_{ij}} - \min \boldsymbol{x_{ij}}}{\max \boldsymbol{x_{ij}} - \min \boldsymbol{x_{ij}}}.$$

The non-dimensional decision matrix is denoted as $\mathbf{X'} = \mathbf{X} = \left(\boldsymbol{x'}_{ij}\right)_{m\times n}$.

Step2. Index weight definition based on entropy method

●Calculate the weight of $\boldsymbol{i}$ in the index $\boldsymbol{j}$: $\boldsymbol{p_{ij}} = \frac{1+x_{ij}}{\sum_{i=1}^{m}(1+x_{ij})}$.

●Calculate the entropy and the difference coefficient in the index $\boldsymbol{j}$:

$$\boldsymbol{e_j} = -(\ln \boldsymbol{m})^{-1} \sum_{i=1}^{m} \boldsymbol{p_{ij}} \ln \boldsymbol{p_{ij}},\ \ \boldsymbol{g_i} = \mathbf{1} - \boldsymbol{e_j}.$$

●Calculate the final weight of each index $\boldsymbol{w_j} = \frac{g_j}{\sum_{i=1}^{n} g_j}$ and get the weight set

$$\mathbf{W} = [\boldsymbol{w_1}\ \ \boldsymbol{w_2}\ \ \boldsymbol{w_3}\ \ \boldsymbol{w_4}].$$

Step3. Evaluation based on the TOPSIS method and the PROMETHEE method

●Calculate the weighted standardized decision matrix:

$$\mathbf{Z} = (\boldsymbol{z_{ij}})_{m\times n}, \boldsymbol{z_{ij}} = \boldsymbol{w_j} \times \boldsymbol{x_{ij}}.$$

●Determine the positive ideal solution $\mathbf{Z}^{+}$ and the negative ideal solution $\boldsymbol{Z}^{-}$:

$$\mathbf{Z}^{+} = (\mathbf{z}_1^{+}\ \ \mathbf{z}_2^{+}\ \ \mathbf{z}_3^{+}\ \ \mathbf{z}_4^{+}),$$

$$\mathbf{Z}^{-} = (\mathbf{z}_1^{-}\ \ \mathbf{z}_2^{-}\ \ \mathbf{z}_3^{-}\ \ \mathbf{z}_4^{-}),$$

where $\mathbf{Z}_j^{+} = \max\limits_{i}(\boldsymbol{x'}_{ij})$, $\mathbf{Z}_j^{+} = \min\limits_{i}(\boldsymbol{x'}_{ij})$.

●Calculate the Euclidean distance between each states' indexes and the positive ideal solution Z^+, the negative ideal solution Z^- respectively:

$$d_i^+ = \sqrt{\sum_{j=1}^{n}(z_{ij} - z_{ij}^+)^2}, d_i^- = \sqrt{\sum_{j=1}^{n}(z_{ij} - z_{ij}^-)^2}.$$

●Determine the preference function and preference exponential relationship. We define $d_j(a,b) = x'_{aj} - x'_{bj}$, preference function $P(d) = \begin{cases}1, d \geq 0 \\ 0, d \leq 0\end{cases}$, preference exponential relationship $\pi_{(d_{ab})} = \sum_{j=1}^{n} P_j(d_{ab})\omega_j$.

●Calculate the positive flow and the negative flow of each evaluation objective, and the formulas are as follows:

$$\Phi^+(d_i) = \frac{1}{m-1}\sum_{k=1}^{m}\pi(d_{ak}), \Phi^-(d_i) = \frac{1}{m-1}\sum_{k=1}^{m}\pi(d_{ak})$$

where $k = 1, 2 \ldots, m$, and $k \neq i$.

●For d_i^+, d_i^-,$\Phi^+(d_a)$, $\Phi^-(d_a)$, after normalization processing, we can obtain D_i^+,D_i^-, Φ_i^+, Φ_i^-.

The larger the value of D_i^- and Φ_i^+, the better the performance of the state. In contrast, the larger the value of D_i^+ and Φ_i^-, the worse the performance of the state. In this paper, we combine D_i^- and Φ_i^+ to evaluate each state's performance in the use of clean renewable energy: $S_i^+ = \varsigma\Phi_i^+ + \tau D_i^-$, where ς and τ depend on the preference of the decision maker and satisfy $\varsigma + \tau = 1, \varsigma, \tau \in [0,1]$.

In this paper, we define $\varsigma = \tau = \frac{1}{2}$.

- 分析标准的合理性

4）D小问（VAR向量自回归预测模型）

- 从总能耗、平均能源价格、石油价格、电价、发电量和用电量六个方面描述能源概况
- 建立向量自回归（VAR）模型，上述六个方面分别作为应变量，再考虑经济和人口因素，将前p年的应变量，前p年的平均能源价格，前p年的GDP增长率和前p年的人口数量作为自变量，建立回归模型，求出参数，分析其模型的可行性（准确率）

$TEC_t^{CA}, P_t^{CA}, V_{GDP(t)}^{CA}$ and POP_t^{CA} denote total energy consumption, average energy price, GDP growth rate and population in California in year t. They can be predicted by lags of $TEC_t^{CA}, P_t^{CA}, V_{GDP(t)}^{CA}$ and POP_t^{CA}. In order to obtain stationary data, we first standardize and difference the four indexes. Afterwards, we calculate the lag phase p according to the AIC criteria. The VAR model of energy usage prediction in California is as follows:

$$\mathrm{TEC}_t^{CA} = w_0 + \sum_{i=1}^{p} \alpha_i \mathrm{TEC}_{t-i}^{CA} + \sum_{i=1}^{p} \beta_i P_{t-i}^{CA} + \sum_{i=1}^{p} \gamma_i V_{GDP(t-i)}^{CA} + \sum_{i=1}^{p} \mu_i \mathrm{POP}_{t-i}^{CA} + \varepsilon_t,$$

where w_0 is constant, ε_t is random disturbance term with zero mean, constant variance and no simultaneous correlation. The prediction formulas of $P_t^{CA}, V_{GDP(t)}^{CA}$ and POP_t^{CA} can also be described in this way.

According to the AIC criteria, we calculate the lag phase $p = 1$ and we obtain the energy consumption prediction model:

$$\left(TEC_t^{CA},\ P_t^{CA},\ V_t^{CA},\ POP_t^{CA}\right) = \left(1,\ TEC_{t-1}^{CA},\ P_{t-1}^{CA},\ V_{t-1}^{CA},\ POP_{t-1}^{CA}\right) \cdot \Omega,$$

where $\Omega = \begin{pmatrix} 0.103 & 0.321 & 1.092 & 0.023 \\ 0.034 & 0.351 & -1.208 & 0.027 \\ -0.533 & -0.127 & -2.562 & 0.020 \\ 0.060 & -0.012 & -0.084 & -0.004 \\ 0.317 & -2.271 & -6.872 & 0.774 \end{pmatrix}$.

- 应用模型预测

4. 第二部分

1）A小问（带约束的优化问题，遗传算法）

- 建立带约束的优化问题，以电能进口花费+电能出口花费+石油花费最小为优化目标

$$\min \quad p_e^i E_{TC}^i + p_e^n E_{TC}^n + p_p P_{TC} + \frac{\eta}{E_{TC}^i}$$

$$s.t. \begin{cases} E_{TC}^i + E_{TC}^n + P_{TC} = TEC \\ \dfrac{E_{TC}^i + E_{TC}^n}{TEC} > \alpha \\ \dfrac{E_{TC}^n}{E_{TC}} > \beta \\ E_{TC}^i > 0 \end{cases}$$

- 遗传算法求解
- 分析结果

2）B小问

- 出口减税
- 可再生能源补贴
- 内部关系建设

5. 附：

1). topsis模型：

https://zhuanlan.zhihu.com/p/37738503

2). 遗传算法：

- https://www.zhihu.com/question/23293449

- https://blog.csdn.net/b2b160/article/details/4680853

可行性分析可以文字分析或者根据实际数据计算准确率/RMSE误差