附件中数据给出了钢架四条腿部每侧应力的相关数据。首先使用因子分析，通过研究变量间的相关系数矩阵，将钢架每侧的应力变量反映成一个综合变量，这一过程又可称为降维。然后建立典型相关分析，从Line Temperature [°C]、Ambient Temperature OTLM [°C]、Humidity [%]、Angle [°] 、Altitude [m]和Vbat [V]中选取若干典型变量，其目的是分析电流与应力之间存在的关系。最后建立一元线性回归方程，更精确地定量描述应力与弧垂角度之间的关系。

利用因子分析进行应力数据的处理分析

因子分析是主成分分析的推广，它也是一种把多个变量化为少数几个综合变量的多元分析方法，其目的是用有限个不可观测的隐变量来解释原始变量之间的相关关系。

假设p维维随机向量

（清风，因子分析的原理一页）

因子分析模型的假设：（清风，假设一页）

KMO检验和巴特利特球形检验

**KMO检验**

KMO检验是 Kaiser, Meyer和 Olkin提出的，该检验是对原始变量之间的简单相关系数和偏相关系 数的相对大小进行检验，主要应用于多元统计的因子分析。KMO统计量是取值在0和1之间，当所有变量间的简单相关系数平方和远远大于偏相关系数平方和 时，KMO值越接近于1，意味着变量间的相关性越强，原有变量越适合作因子分析;当所有变量 间的简单相关系数平方和接近0时，KMO值越接近于0,意味着变量间的相关性越弱，原有变量越不适合作因子分析。

**巴特利特球形检验**

巴特利特球形检验是一种检验各个变量之间相关性程度的检验方法。巴特利特球形检验是以变量的相关系数矩阵为出发点的。它的原假设是相关系数矩阵是一个单位阵(不适合做因子分析，指标之间的相关性很差，不适合降维)，即相关系数矩阵对角线上的所有元素都是1，所有非对角线上的元素都为0。巴特利特球形检验的统计量是根据相关系数矩阵的行列式得到的。如果该值较大，且其对应的p值小于用户心中的显著性水平(一般为0.05)，那么应该拒绝原假设，认为相关系数不可能是单位阵，即原始变量之间存在相关性，适合于作因子分析。相反不适合作因子分析。

通过对附件中的数据建立预处理、特征工程、特征选择，变量从原来的8个变成了2个。而特征维度太高对主观法与客观法建立评价体系来说都很复杂，因此我们利用因子分析降维，将8个变量降成了2维。这里展示重要公共因子：

（一张表，因子分析的各个变量，详见华为杯论文）

处理一下下面的表

|  |  |  |
| --- | --- | --- |
| **KMO 和巴特利特检验（1侧）** | | |
| KMO 取样适切性量数。 | | .537 |
| 巴特利特球形度检验 | 近似卡方 | 106914.762 |
| 自由度 | 6 |
| 显著性 | <.001 |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| **KMO 和巴特利特检验（2侧）** | | |
| KMO 取样适切性量数。 | | .549 |
| 巴特利特球形度检验 | 近似卡方 | 178172.216 |
| 自由度 | 6 |
| 显著性 | <.001 |

原假设：数据不适合进行因子分析

备择假设：数据适合进行因子分析

对因子进行KMO 和巴特利特检验，由KMO 取样适切性量数和巴特利特检验的显著性水平可知，在99%的置信水平下，我们拒绝原假设，即认为原数据适合因子分析，通过了KMO 和巴特利特检验。

|  |  |  |
| --- | --- | --- |
| **公因子方差1侧** | | |
|  | 初始 | 提取 |
| S1 Ch1 Avg [MPa] | 1.000 | .590 |
| S2 Ch1 Avg [MPa] | 1.000 | .924 |
| S3 Ch1 Avg [MPa] | 1.000 | .799 |
| S4 Ch1 Avg [MPa] | 1.000 | .081 |
| 提取方法：主成分分析法。 | | |

|  |  |  |
| --- | --- | --- |
| **公因子方差2侧** | | |
|  | 初始 | 提取 |
| S1 Ch2 Avg [MPa] | 1.000 | .669 |
| S2 Ch2 Avg [MPa] | 1.000 | .674 |
| S3 Ch2 Avg [MPa] | 1.000 | .619 |
| S4 Ch2 Avg [MPa] | 1.000 | .797 |
| 提取方法：主成分分析法。 | | |

（写共因子数据结论）

由表和表可看出，通过主成分分析法，对于Ch1侧，提取的共因子可以反映S1脚59.0%的信息，S2脚92.4%的信息，S3脚79.9%的信息，S4脚8.1%的信息。对于Ch2侧，提取的共因子可以反映S1脚66.9%的信息，S2脚67.4%的信息，S3脚61.9%的信息，S4脚79.7%的信息。

通过共因子方差可以看出，提取的公因子对各个变量的贡献度均较大，这表明了这两个公因子可以较好地反映各个变量的信息。

|  |  |
| --- | --- |
| **1侧成分矩阵a** | |
|  | 成分 |
| 1 |
| S1 Ch1 Avg [MPa] | .768 |
| S2 Ch1 Avg [MPa] | .961 |
| S3 Ch1 Avg [MPa] | -.894 |
| S4 Ch1 Avg [MPa] | -.285 |
| 提取方法：主成分分析法。 | |
| a. 提取了 1 个成分。 | |

|  |  |
| --- | --- |
| **2侧成分矩阵a** | |
|  | 成分 |
| 1 |
| S1 Ch2 Avg [MPa] | .818 |
| S2 Ch2 Avg [MPa] | -.821 |
| S3 Ch2 Avg [MPa] | .787 |
| S4 Ch2 Avg [MPa] | .893 |
| 提取方法：主成分分析法。 | |
| a. 提取了 1 个成分。 | |

因子载荷是变量与公共因子的相关系数，当某变量 在某公共因子中的载荷绝对值越大，表明该变量与 该公共因子更密切，即该公共因子更能代表该变量。由成分矩阵可得，合成的公共因子可以写成以下形式：

F1=0.768\* S1 Ch1 Avg [MPa]+0.961\* S2 Ch1 Avg [MPa]-0.894\* S3 Ch1 Avg [MPa]-0.285\* S4 Ch1 Avg [MPa]

F2=0.818\* S1 Ch2 Avg [MPa]-0.821\* S2 Ch2 Avg [MPa]+0.787\* S3 Ch2 Avg [MPa]+0.893\* S4 Ch2 Avg [MPa]

**利用典型相关分析分析电流与应力变化的关系**

为了研究电流与应力变化之间的关系，令电流为输入变量，应力变化为输出变量，采用典型相关分析法。

典型相关分析由Hotelling提出，其基本思想是利用主成分思想，分别找出输入变量与输出变量的线性组合，然后讨论线性组合之间的相关关系。。

首先在每组变量中找出变量的线性组合，使得两组的线性组合之间具有最大的相关系数;

然后选取和最初挑选的这对线性组合不相关的线性组合，使其配对，并选取相关系数最大的一对;

如此继续下去，直到两组变量之间的相关性被提取完毕为止。

被选出的线性组合配对称为典型变量，它们的相关系 数称为典型相关系数。典型相关系数度量了这两组变量之间联系的强度。

假设电流与应力数据服从联合正态分布。构造构造似然比统计量，对两组变量的相关性进行检验。

原假设H0:两组变量之间不存在相关关系

备择假设H1:两组变量之间存在相关关系。

利用SPSS软件进行典型相关分析，可以得到在99%的置信水平下，我们拒绝原假设，即认为两组变量之间存在相关性，可以进行典型相关分析。

|  |  |
| --- | --- |
| **典型相关性设置** | |
|  | 值 |
| 电流显著型指标 | LT°CDEV1 LT°CDEV2 AT°CDEV1 AT°CDEV2 ALm AG°DEV1 AG°DEV2 |
| 应力显著性指标 | HDEV1 HDEV2 VV |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **典型相关系数** | | | | | | | |
|  | 相关系数 | 特征值 | 威尔克统计 | F | 分子自由度 | 分母自由度 | 显著性 |
| 1 | .581 | .509 | .595 | 1159.890 | 21.000 | 123097.172 | .000 |
| 2 | .293 | .094 | .898 | 393.317 | 12.000 | 85740.000 | .000 |
| 3 | .131 | .018 | .983 | 150.353 | 5.000 | 42871.000 | .000 |
| H0 for Wilks 检验是指当前行和后续行中的相关性均为零  考虑到各个变量之间的量纲存在较大差异，为消除量纲对线性组合系数的影响，对线性组合系数进行标准化。标准化后典型相关变量对应的线性组合系数见下表。未标准化的典型相关变量对应的线性组合系数见附录。 | | | | | | | |
|  | | | | | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| **集合 1 标准化典型相关变量对应的线性组合系数** | | | |
| 变量 | 1 | 2 | 3 |
| LT°CDEV1 | .470 | 2.929 | 10.606 |
| LT°CDEV2 | -.160 | 1.845 | -10.650 |
| AT°CDEV1 | 8.011 | -5.494 | -1.090 |
| AT°CDEV2 | -7.885 | 1.109 | .882 |
| ALm | -.021 | .029 | -.087 |
| AG°DEV1 | .240 | -.865 | -.625 |
| AG°DEV2 | -.051 | -.773 | -.446 |

|  |  |  |  |
| --- | --- | --- | --- |
| **集合 2 标准化典型相关变量对应的线性组合系数** | | | |
| 变量 | 1 | 2 | 3 |
| HDEV1 | -1.360 | .060 | -1.642 |
| HDEV2 | 1.022 | -1.059 | 1.466 |
| VV | .556 | -.328 | -.836 |

|  |  |  |  |
| --- | --- | --- | --- |
| **集合 1 非标准化典型相关变量对应的线性组合系数** | | | |
| 变量 | 1 | ~~2~~ | ~~3~~ |
| LT°CDEV1 | .059 | ~~.365~~ | ~~1.323~~ |
| LT°CDEV2 | -.022 | ~~.248~~ | ~~-1.432~~ |
| AT°CDEV1 | 1.040 | ~~-.713~~ | ~~-.141~~ |
| AT°CDEV2 | -1.086 | ~~.153~~ | ~~.121~~ |
| ALm | -.001 | ~~.002~~ | ~~-.005~~ |
| AG°DEV1 | .358 | ~~-1.291~~ | ~~-.933~~ |
| AG°DEV2 | -.059 | ~~-.897~~ | ~~-.518~~ |

|  |  |  |  |
| --- | --- | --- | --- |
| **集合 2 非标准化典型相关变量对应的线性组合系数** | | | |
| 变量 | 1 | ~~2~~ | ~~3~~ |
| HDEV1 | -7.970 | ~~.354~~ | ~~-9.623~~ |
| HDEV2 | 6.266 | ~~-6.490~~ | ~~8.988~~ |
| VV | .834 | ~~-.492~~ | ~~-1.253~~ |

|  |  |  |  |
| --- | --- | --- | --- |
| **集合 1的典型相关变量 典型载荷** | | | |
| 变量 | 1 | 2 | 3 |
| LT°CDEV1 | .676 | .353 | -.386 |
| LT°CDEV2 | .643 | .359 | -.459 |
| AT°CDEV1 | .694 | .209 | -.419 |
| AT°CDEV2 | .632 | .227 | -.446 |
| ALm | .087 | .060 | -.150 |
| AG°DEV1 | .714 | .000 | -.223 |
| AG°DEV2 | -.726 | -.252 | .022 |

|  |  |  |  |
| --- | --- | --- | --- |
| **集合 2的典型相关变量 典型载荷** | | | |
| 变量 | 1 | 2 | 3 |
| HDEV1 | -.629 | -.769 | -.116 |
| HDEV2 | -.272 | -.943 | .190 |
| VV | .759 | -.145 | -.634 |

|  |  |  |  |
| --- | --- | --- | --- |
| **集合 1 交叉载荷** | | | |
| 变量 | 1 | 2 | 3 |
| LT°CDEV1 | .393 | .104 | -.051 |
| LT°CDEV2 | .373 | .105 | -.060 |
| AT°CDEV1 | .403 | .061 | -.055 |
| AT°CDEV2 | .367 | .067 | -.059 |
| ALm | .051 | .018 | -.020 |
| AG°DEV1 | .415 | .000 | -.029 |
| AG°DEV2 | -.422 | -.074 | .003 |

|  |  |  |  |
| --- | --- | --- | --- |
| **集合 2 交叉载荷** | | | |
| 变量 | 1 | 2 | 3 |
| HDEV1 | -.365 | -.225 | -.015 |
| HDEV2 | -.158 | -.276 | .025 |
| VV | .441 | -.043 | -.083 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **已解释的方差比例** | | | | |
| 典型变量 | 集合 1 \* 自身 | 集合 1 \* 集合 2 | 集合 2 \* 自身 | 集合 2 \* 集合 1 |
| 1 | .400 | .135 | .349 | .118 |
| 2 | .059 | .005 | .501 | .043 |
| 3 | .115 | .002 | .151 | .003 |

典型相关变量的典型载荷与已解释的方差比例见附录。

可得到标准化后的典型变量：

利用一元线性回归建立应力与弧垂角之间的线性关系

利用一元线性回归，结合题目已知条件，利用Stata软件进行回归分析，建立一元线性回归方程如下。

（显著性检验）

[1]侯宇,王伟,韦徵,等.输电线路动态增容技术研究及应用[J].电力系统自动化,2021,45(17):189-198.

[1]周赞东,苟晓侃,曾胜强.输电线路动态增容模块化计算技术研究[J].能源研究与管理,2017(02):55-57.DOI:10.16056/j.1005-7676.2017.02.013.

通过分析标准化典型相关变量对应的线性组合系数可知，对于Line Current [A]，LT [°C] DEV1和AT [°C] DEV1的线性组合系数的绝对值分别为2.929和5.494，表明其对Line Current [A]的影响较大。而Al[m]的线性组合系数仅为0.029，表明其对Line Current [A]的影响几乎可以忽略。对于ST[KPa]，影响较为显著的是H [%] DEV2与V [V]。

由上图可知，Line Current[A]在8月随昼夜变化，呈现出较强的周期性。在2020-8-12 12:00之前，Line Current[A]振幅较小。在2020-8-12 12:00之后，振幅出现了明显的增大趋势，最高峰值Line Current[A]达到了340[A]左右，这与夏季用电高峰有关。