1. **问题分析**

**1.1问题背景**

在博彩市场上，预测各支代表队奥运会夺牌数，一直以来都是一个经久不衰的课题。有研究指出，人口和财富水平均为夺牌数的关键因素。然而，其他少数几种指标也不容忽视，如主场优势、经济体制、往届表现等等。

**1.2问题分析**

本题要求使用已有数据，构建奥运奖牌预测模型。然后利用该模型预测2000年各国夺牌数，并与2000年各国实际夺牌数进行对比，检验该模型的优劣。

阅读相关文献后，总结得出人口、GDP、主场优势、经济体制、往届表现为决定各国夺牌数的重要因素。将这些因素作为自变量，夺牌数作为因变量，进行多元线性回归即可得到多元线性回归模型，进而以此预测2000年各国夺牌数。

1. **模型构建与求解**

**2.1模型构建**

（1）被预测变量：medaltot(奖牌数)

（2）解释变量：gdp 、pop(人口) 、host(是否为主办国)、planned(是否为计划经济) 、perform(之前的表现)

Medaltot=+gdp+pop+host+planned+perform

**2.2求解**

2.2.1计算往届表现

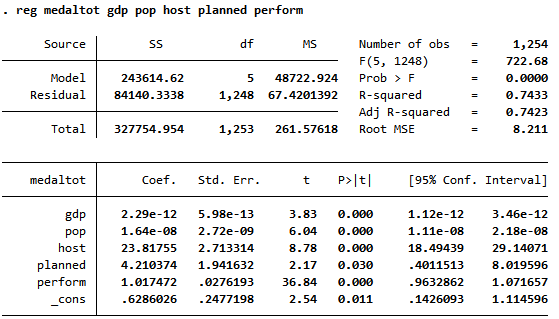
往届表现的衡量方式有很多，比如采用上一届夺牌数，亦或者往届夺牌数的平均值。还可以用往届夺牌数的加权平均值，距今时间更近则权重越大。本模型采用往届夺牌数的平均值来表示往届表示这一变量。数据包括1960-1996各国夺牌数与其他情况。如有一条某国1988年的数据，则该条数据的perform为1960-1984该国夺牌数的平均值。

具体数据采用Excel进行处理。先按年份升序排列所有数据，再使用以下函数求解得出每条数据的perform变量值。

SUMIFS($J$1:$J1,$C$1:$C1,C2)/COUNTIFS($C$1:$C1,C2)

2.2.2多元线性回归

使用stata进行多选线性回归，求解得出奖牌预测模型。回归结果如图一所示。



**图一 多选线性回归结果**

则可写出奖牌预测模型：

Medaltot=0.6286026+2.29e-12gdp+1.64e-08pop+23.81755host+4.210374planned+1.017472perform

1. **预测**

在网上搜集到了2000年奖牌榜前21个国家的各项数据，带入预测模型得出预测奖牌数，如表一。

为评价预测是否准确，采用均方误差，如以下公式。

计算得MSE=231.9717066。可见误差较大，模型还有待进一步改进。

**表一 2000年各国预测夺牌数与实际夺牌数**

|  |  |  |
| --- | --- | --- |
| 国家 | 预测奖牌数 | 实际奖牌数 |
| 美国 | 123.662 | 97 |
| 俄罗斯 | 92.66252 | 88 |
| 中国 | 45.06207 | 59 |
| 澳大利亚 | 21.63279 | 58 |
| 德国 | 81.21175 | 57 |
| 法国 | 23.16414 | 38 |
| 意大利 | 27.07585 | 34 |
| 荷兰 | 10.89905 | 25 |
| 古巴 | 15.47011 | 29 |
| 英国 | 25.63965 | 28 |
| 罗马尼亚 | 23.46731 | 26 |
| 韩国 | 4.246481 | 28 |
| 匈牙利 | 25.12041 | 17 |
| 波兰 | 21.2891 | 14 |
| 日本 | 35.24834 | 18 |
| 保加利亚 | 18.70078 | 13 |
| 希腊 | 3.139067 | 13 |
| 瑞典 | 1.37602 | 12 |
| 挪威 | 4.248528 | 10 |
| 埃塞俄比亚 | 3.364494 | 8 |
| 乌克兰 | 13.20977 | 23 |