

五、问题一的模型建立与求解

5.1 数据预处理

5.1.1 剔除异常数据

由题目以及查阅世界卫生组织所发布的一份有关儿童睡眠的指南表示，0~3个月大的婴儿每天应保持14~17小时的睡眠，4~11个月大的婴儿每天应保持12~16个小时的优质睡眠。在附件中发现编号为180的婴儿睡眠时间为99时99分，不仅远超于正常婴儿的睡眠时间，甚至于整晚的睡眠时间超过了24小时，而且时间不规范，属于异常数据，故剔除编号为180的数据。

5.1.2 对定类变量编码

在本题中，有诸如母亲年龄，教育程度等定量变量，也有婴儿行为特征的定类变量，为了方便数据处理，需要对婴儿行为特征这一变量进行编码，令安静型为0，中等型为1，矛盾型为2。

5.2 Spearman 相关性分析

5.2.1 Spearman 相关性系数

Spearman 相关系数适用于定量变量或定序变量两两之间的相关分析，利用两变量的秩次大小作线性相关分析，对原始变量的分布不作要求，当变量中至少存在一个有序变量时，可使用 Spearman 系数来描述两变量的相关性，它利用单调函数评价两个统计变量的相关性。

对于样本容量为 n 的样本，将 n 个原始数据 X_i, Y_i 转换成等级数据 $R(X_i), R(Y_i)$ ，则相关系数 r_s 为：

$$r_s = \rho_{R_X, R_Y} = \frac{cov(R(X), R(Y))}{\sigma_{R(X)} \sigma_{R(Y)}} \tag{1-1}$$

通常使用更简单的计算步骤计算等级相关系数：

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{1-2}$$

其中， ρ 是 Pearson 积矩相关系数， $cov(R(X), R(Y))$ 为等级变量的协方差， $\sigma_{R_X}, \sigma_{R_Y}$ 为等级变量的标准差， $d_i = R(X_i) - R(Y_i)$ 为每组观测中的两个变量等级差值。

由于定量数据不完全满足正态分布，因此本题中，应用 Spearman 相关系数和显著性 P 值来分别分析母亲的身体指标和心理指标是否对婴儿行为特征和睡眠质量有显著影响，并分析影响程度的大小和影响的方向性。

表 1.1 母亲身体指标与婴儿行为特征 Spearman 参数结果表

	母亲年龄	婚姻状况	教育程度	妊娠时间（周数）	分娩方式	CBTS	EPDS	HADS	婴儿行为特征	婴儿性别	婴儿年龄(月)	整晚睡眠时间(小时)	睡眠次数	入睡方式
母亲年龄	1(0.000***)	0.054(0.286)	0.171(0.001***)	-0.119(0.019**)	0.037(0.466)	-0.043(0.398)	-0.093(0.066*)	-0.072(0.158)	-0.096(0.060*)	-0.062(0.225)	-0.019(0.709)	0.014(0.779)	0.037(0.466)	-0.08(0.114)
婚姻状况	0.054(0.286)	1(0.000***)	0.096(0.059*)	-0.051(0.319)	0.006(0.903)	-0.023(0.653)	-0.053(0.299)	-0.068(0.178)	-0.019(0.706)	-0.076(0.134)	-0.029(0.567)	-0.028(0.587)	-0.034(0.501)	-0.103(0.041**)
教育程度	0.171(0.001***)	0.096(0.059*)	1(0.000***)	0.029(0.572)	-0.016(0.756)	-0.059(0.244)	-0.101(0.047**)	-0.097(0.056*)	0.056(0.272)	-0.005(0.929)	0.037(0.472)	-0.056(0.271)	0.085(0.095*)	-0.05(0.322)
妊娠时间（周数）	-0.119(0.019**)	-0.051(0.319)	0.029(0.572)	1(0.000***)	-0.176(0.000***)	-0.086(0.092*)	-0.055(0.277)	-0.113(0.026**)	0.002(0.968)	0.046(0.362)	0.000(0.994)	0.074(0.146)	0.091(0.073*)	0.051(0.318)
分娩方式	0.037(0.466)	0.006(0.903)	-0.016(0.756)	-0.176(0.000***)	1(0.000***)	-0.02(0.687)	-0.018(0.717)	-0.064(0.205)	0.004(0.944)	-0.11(0.030**)	-0.05(0.329)	0.016(0.749)	-0.062(0.223)	0.064(0.206)
CBTS	-0.043(0.398)	-0.023(0.653)	-0.059(0.244)	-0.086(0.092*)	-0.02(0.687)	1(0.000***)	0.781(0.000***)	0.711(0.000***)	0.126(0.013**)	0.071(0.160)	-0.073(0.151)	-0.129(0.011**)	0.076(0.133)	0.046(0.369)
EPDS	-0.093(0.066*)	-0.053(0.299)	-0.101(0.047**)	-0.055(0.277)	-0.018(0.717)	0.781(0.000***)	1(0.000***)	0.784(0.000***)	0.145(0.004***)	0.04(0.435)	-0.045(0.373)	-0.167(0.001***)	0.115(0.024**)	0.006(0.907)
HADS	-0.072(0.158)	-0.068(0.178)	-0.097(0.056*)	-0.113(0.026**)	-0.064(0.205)	0.711(0.000***)	0.784(0.000***)	1(0.000***)	0.129(0.011**)	0.03(0.552)	0.006(0.906)	-0.117(0.021**)	0.065(0.200)	0.06(0.240)
婴儿行为特征	-0.096(0.060*)	-0.019(0.706)	0.056(0.272)	0.002(0.968)	0.004(0.944)	0.126(0.013**)	0.145(0.004***)	0.129(0.011**)	1(0.000***)	-0.019(0.705)	0.343(0.000***)	-0.115(0.023**)	0.243(0.000***)	-0.001(0.978)
婴儿性别	-0.062(0.225)	-0.076(0.134)	-0.005(0.929)	0.046(0.362)	-0.11(0.030**)	0.071(0.160)	0.04(0.435)	0.03(0.552)	-0.019(0.705)	1(0.000***)	0.001(0.978)	-0.056(0.273)	0.077(0.127)	-0.024(0.641)
婴儿年龄(月)	-0.019(0.709)	-0.029(0.567)	0.037(0.472)	0.000(0.994)	-0.05(0.329)	-0.073(0.151)	-0.045(0.373)	0.006(0.906)	0.343(0.000***)	0.001(0.978)	1(0.000***)	0.187(0.000***)	-0.005(0.925)	0.18(0.000***)
整晚睡眠时间(小时)	0.014(0.779)	-0.028(0.587)	-0.056(0.271)	0.074(0.146)	0.016(0.749)	-0.129(0.011**)	-0.167(0.001***)	-0.117(0.021**)	-0.115(0.023**)	-0.056(0.273)	0.187(0.000***)	1(0.000***)	-0.309(0.000***)	0.234(0.000***)
睡眠次数	0.037(0.466)	-0.034(0.501)	0.085(0.095*)	0.091(0.073*)	-0.062(0.223)	0.076(0.133)	0.115(0.024**)	0.065(0.200)	0.243(0.000***)	0.077(0.127)	-0.005(0.925)	-0.309(0.000***)	1(0.000***)	-0.26(0.000***)
入睡方式	-0.08(0.114)	-0.103(0.041**)	-0.05(0.322)	0.051(0.318)	0.064(0.206)	0.046(0.369)	0.006(0.907)	0.06(0.240)	-0.001(0.978)	-0.024(0.641)	0.18(0.000***)	0.234(0.000***)	-0.26(0.000***)	1(0.000***)

5. 2. 2 Spearman 显著性检验

对于相关系数 r ，使用 Pearson 积矩中的费雪变换，对样本相关系数进行费雪变换后，可以用来检验关于总体相关系数 ρ 的假设，此时 ρ 的置信区间和假设检验可通过费雪变换获得：

$$F(r) = \frac{1}{2} \ln \frac{1+r}{1-r} = \operatorname{arctanh} r \tag{1-3}$$

如果 $F(r)$ 是 r 的费雪变换，则

$$z = \sqrt{\frac{n-3}{1.06}} F(r) \tag{1-4}$$

其中 r 在统计独立性 $\rho = 0$ 的零假设下近似服从标准正态分布，其显著性为：

$$t = r \sqrt{\frac{n-2}{1-r^2}} \tag{1-5}$$

其在零假设下近似副自由度为 $n - 2$ 的 t 分布
在表1.1中***、**、*分别代表 1%、5%、10%的显著性水平。

可使用相关系数热力图来更直观地表示相关系数地正负向以及相关性程度

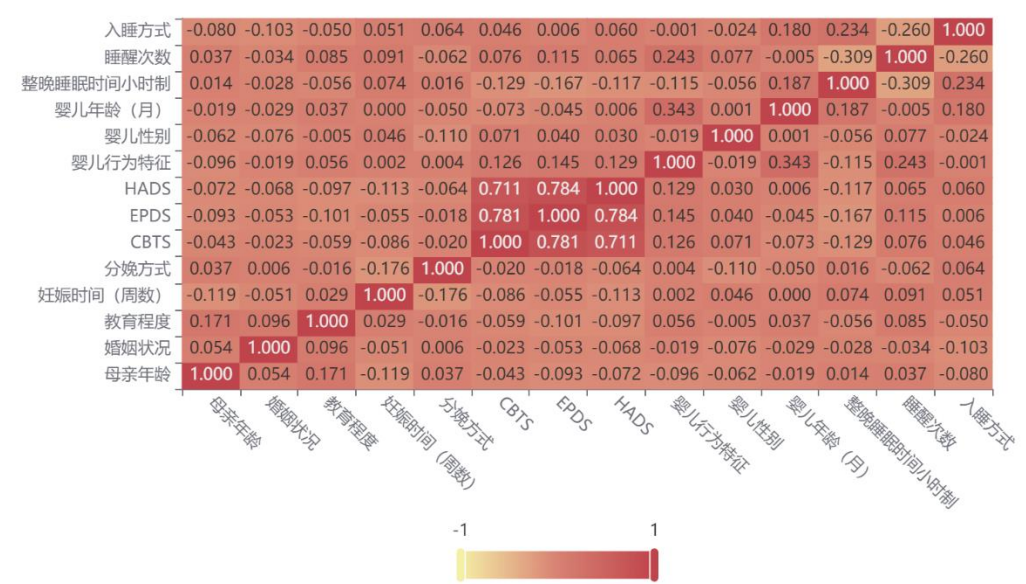


图 1.1 母亲身体指标与婴儿行为特征相关系数热力图

5. 3 问题一总结

母亲的身体指标和心理指标对婴儿的行为特征和睡眠质量有影响，由表 1.1 可知，在显著性水平为 5% 下，即拥有至少 95% 的把握认为，诸如婚姻状况与婴儿的入睡方式，母亲的心理焦虑程度与婴儿的行为特征以及婴儿的每晚的睡眠时有影响。