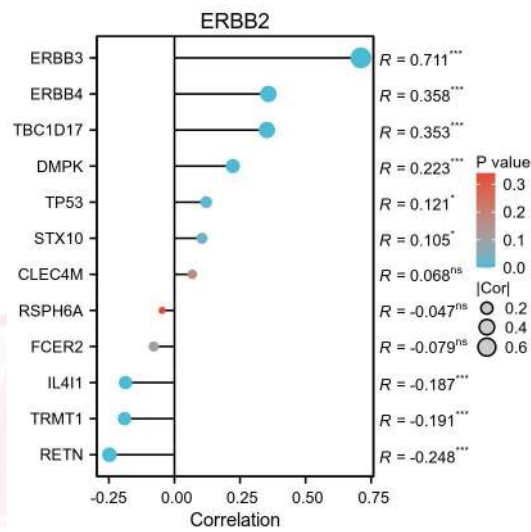


交互网络 - 相关性棒棒糖图[云]



网址: <https://www.xiantao.love>



更新时间: 2023.02.28

目录

基本概念	3
应用场景	3
分析过程	3
结果解读	6
数据格式	7
参数说明	8
数据处理	9
映射	11
点	13
线	14
标注	15
标题文本	16
图注	16
坐标轴	17
风格	18
图片	19
结果说明	20
主要结果	20
补充结果	21
方法学	22
如何引用	23
常见问题	24

基本概念

- 棒棒糖图：通过点图和线图组合成棒棒糖的形式
- 相关性棒棒糖图：分析 1 个主要变量和其他所有变量之间的相关性结果并用棒棒糖图形式展示结果

应用场景

基于云端数据 通过点、线的形式来展示主要变量与其他变量之间的相关性

分析过程

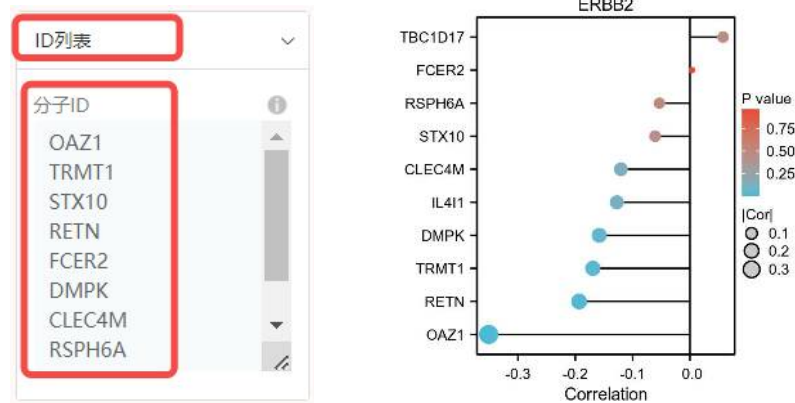
云端数据 → 相关性分析 → 可视化

- 云端数据：提供预清洗好的云端数据，不同平台的云端数据集的分子可能会有不同。 注意：选择了不同的平台，搜索出来的分子可能是不一样的
- 相关性分析
 - 将云端数据进行相关性分析
 - ◆ 在特殊参数[分子]中选择主要分子/变量/基因名

特殊参数

分子 ERBB2[ENSG00000141736.14]

- ◆ 在参数[ID 列表]中输入云端数据中需要进行分析的分子/变量/基因名，输入多少个分子 ID 就对多少分子进行分析，如下：



如果输入分子 ID 少于 3 个或者输入分子 ID 与数据中匹配的个数少于 3 个，则不能进行分析：如下：



错误信息

ID列表少于3个，需要在ID列表中输入更多的分子ID才能进行分析和可视化

确定

错误信息

输入的ID中，正确识别出来的ID少于2个，无法绘制相关性和弦图

确定

- ◆ 相关性分析表：选择的特殊参数[分子]与主要参数[ID 列表]分子 ID 匹配到数据中的分子进行相关性分析

- 相关性系数、统计学检验 p 值等相关信息

相关性分析

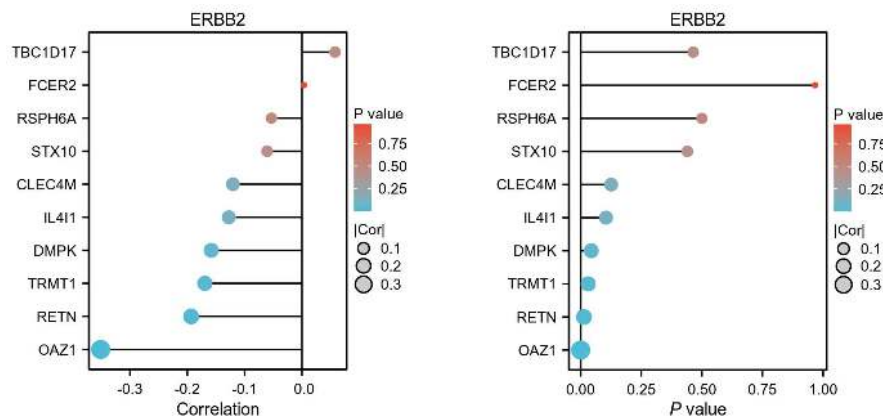
提供Pearson和Spearman统计方法的结果

主变量	次变量	自由度(df)	统计量-Pearson	相关系数-Pearson	p值-Pearson	统计量-Spearman	相关系数-Spearman	p
ERBB2	OAZ1	161	-3.7074	-0.280457	0.0003	9.75e+05	-0.350893	
ERBB2	TRMT1	161	-2.08243	-0.161952	0.0389	8.44e+05	-0.169294	
ERBB2	STX10	161	-0.968776	-0.0761287	0.3341	7.656e+05	-0.0607983	
ERBB2	RETN	161	-2.06727	-0.160803	0.0403	8.612e+05	-0.193218	
ERBB2	FCER2	161	-0.853311	-0.0670987	0.3948	7.194e+05	0.00326035	
ERBB2	DMPK	161	-3.29184	-0.25112	0.0012	8.359e+05	-0.158107	
ERBB2	CLEC4M	161	-1.19814	-0.0940084	0.2326	8.087e+05	-0.120386	
ERBB2	RSPH6A	161	-1.51854	-0.11883	0.1308	7.602e+05	-0.0531982	
ERBB2	TBC1D17	161	-0.381143	-0.0300247	0.7036	6.802e+05	0.0576477	
ERBB2	IL4I1	161	-1.42929	-0.111936	0.1549	8.139e+05	-0.127596	

- 将分析后得到的相关性系数与 p 值进行后续的相关性热图可视化



结果解读



- 横坐标表示其他变量（参数[ID 列表] 匹配到数据中的分子）（**图片为默认转置后**）
- 纵坐标表示主要变量（特殊参数[分子]）与其他变量（参数[ID 列表] 匹配到数据中的分子）之间的（相关系数（左侧）/p 值（右侧）），取决于映射的内容（**图片为默认转置后**）
- 每个点表示主要变量与其他变量之间所映射的内容（相关系数/p 值/无）
- 连接点的横线表示相关性系数（相关系数/p 值），横线越长，表示其对应的值越大），取决于映射的内容
 - **如左图**：横线越长，变量之间相关系数的绝对值越大，表示主要变量与其他变量之间的相关性越强
 - **如右图**：横线越长，表示主要变量与其他变量之间的统计学 p 值越大，即相关性越没那么显著
- 横坐标等于 0 位置的竖线（虚线）表示相关性系数（相关系数/p 值）分割线

数据格式

提供预清洗好的云端数据，不同平台的云端数据集的分子可能会有不同。注意：选择了不同的平台，搜索出来的分子可能是不一样的



参数说明

(说明：标注了颜色的为常用参数。)

数据参数



数据参数 重置参数

云端数据 ⓘ 食管癌 / TCGA / TCGA-ESCA / RNAseq / STAR / TPM @过滤:去除正常+去除无临床信息 @处理:log2(v...

- 云端数据：可以选择基于云端整理好的公共数据

特殊参数



特殊参数 重置参数

分子 ERBB2[ENSG00000141736.14]

- 可以选择进行分析过程中的主要变量/分子

主要参数

- 主要参数为分析以及可视化相关参数，如下[主要参数]

主要参数

ID 列表



- 分子 ID：输入与数据中匹配的分子，以进行后面分析以及可视化
 - 这部分分子可以来自「[单基因差异分析](#)」或者「[单基因相关性筛选](#)」两个模块筛选后再进行选择，建议是结合两者一起来看，如果想要热图结果好看一些，建议是从「单基因相关性筛选」模块中挑相关性高的分子进行可视化（因为相关趋势更加明显）
 - 分子 ID 不能少于 3 个或者分子 ID 与数据中匹配的个数少于 3 个
 - 一行一个 ID，可以是分子名，也可以是分子 ID，最多支持 20 个（最多支持 20 个分子进行相关性分析，也就是热图的行/列不能超过 20）

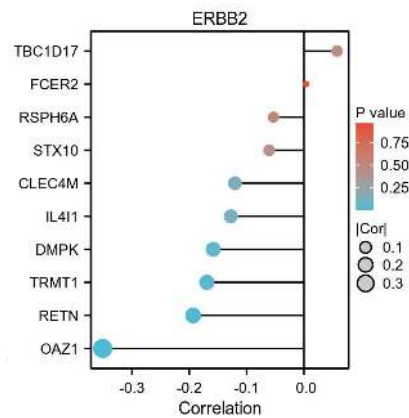
统计

统计

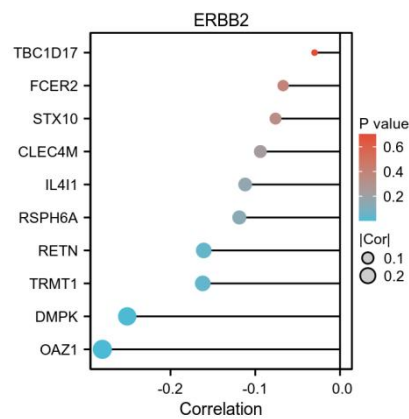
统计方法 Spearman

➤ 统计方法：可以选择主要变量与其他变量间进行相关性分析的方法

- spearman: Spearman(默认)为非参数检验方法，数据可以不需要满足正态性



- pearson: Pearson 为参数检验方法，数据需要满足双正态



映射

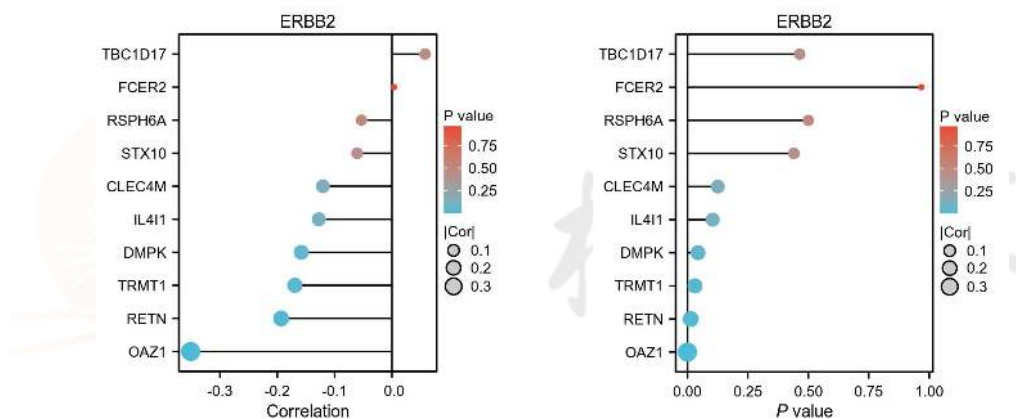
映射

y轴映射 相关系数

颜色映射 p值

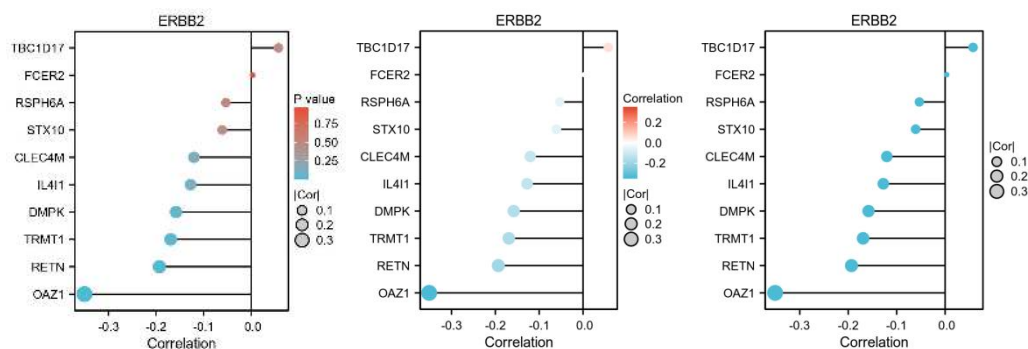
大小映射 相关系数绝对值

- y轴映射：可以选择y轴映射的数据，即是y轴代表的信息，默认为相关系数，还可以选择p值，如下图：(图片为默认转置后的结果)



- 颜色映射：可以选择图中点映射的颜色信息，默认以p值作为映射，如图，颜色越偏向红色表示p值越大，颜色越偏向蓝色说明p值越小（如下左）

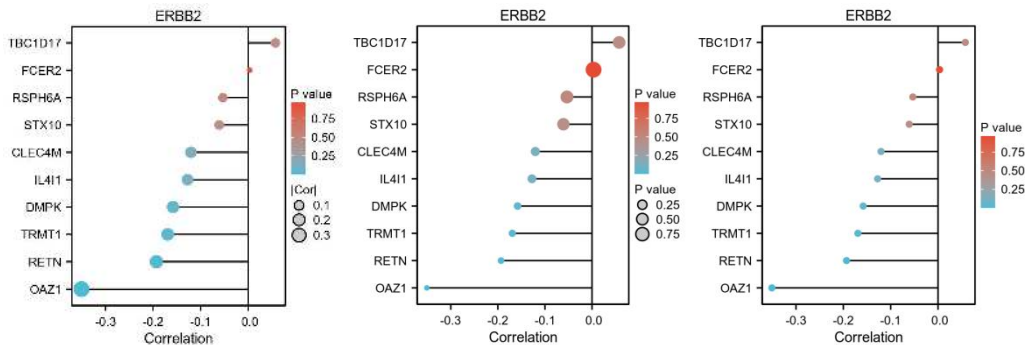
- 还可以选择以相关系数进行映射（如下中）
- 还可以选择不映射（所有都是一个颜色）（如下右）



- 大小映射：可以选择图中点映射的大小信息，默认以相关系数绝对值作为映射，如图，点越大，说明主要变量与其他变量之间相关系数绝对值越大，即他们之间相关性越强（正相关/负相关）（如下左）

■ 还可以选择以 p 值进行映射（如下中）

■ 还可以选择不映射（所有点的大小都是一样的）（如下右）



点

点

填充色

描边色

样式

圆形

大小比例

1

不透明度

1

- 填充颜色：可以修改图中各点的填充颜色
- 描边颜色：可以修改图中各点的描边颜色
- 样式：可以修改图中各点的样式（形状），默认为圆形，还可以选择正方形、菱形、三角形、倒三角形
- 大小比例：可以修改图中个点的大小比例，默认为 1
- 不透明度：可以修改图中各点不透明度，默认为 1，表示完全不透明,0 表示完全不透明

线



A configuration panel for lines, titled '线' (Line). It contains four settings: '颜色' (Color) with a black color swatch, '类型' (Type) with a dropdown menu showing '实线' (Solid line), '粗细' (Thickness) with a dropdown menu showing '0.75pt', and '不透明度' (Opacity) with a text input field showing '1'.

- 颜色：可以修改图中线的颜色
- 类型：可以选择连接点的线的类型，默认为实线，还可以选择虚线
- 粗细：可以选择修改线条的粗细
- 不透明度：可以修改线条的不透明度，默认为 1，表示完全不透明，0 表示完全透明

标注

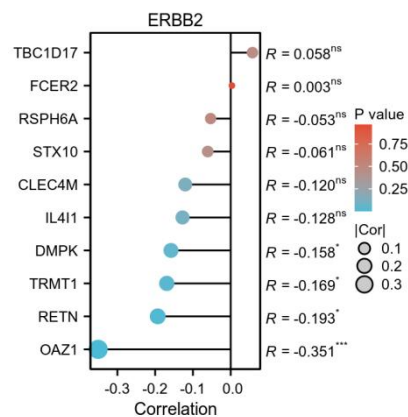
标注 ▼

内容 不标注 ▼

标注大小 6pt ▼

- 内容：可以选择是否对图形进行标注映射，默认为不映射，还可以选择映射：
如下（为相关系数-星号）

- 相关系数
- 相关系数-星号
- 星号、p 值科学计数法
- p 值数值(小于 0.05 自动<)
- p 值数值(小于 0.001 自动<)



- 标注大小：可以选择标注的大小，默认为 6pt

标题文本

标题 ▼

大标题

大标题内容

x轴标题

x轴标题内容

y轴标题

y轴标题内容

- 大标题：大标题文本
- x 轴标题：x 轴标题文本
- y 轴标题：y 轴标题文本

补充：在要换行的中间插入\n。如果需要上标，可以用两个英文输入法下的大括号括住，比如 $\{2\}$ ；如果需要下标，可以用两个英文输入法下的中括号括住，比如 $[2]$

图注

图注 ▼

是否展示

☒

图注位置

默认 ▼

- 是否展示：可以选择是否展示各指标映射的内容，即图注，默认展示
- 图注位置：可以选择图注的位置，默认表示默认展示在右侧，还可以选择上

坐标轴



坐标轴

x轴标注旋转 0

y轴范围+刻度 逗号隔开

- x 轴标注旋转：可选择并修改 x 轴对应刻度文本的旋转角度
- y 轴范围+刻度：可以控制 y 轴范围和刻度，可只提供 2 个值来控制范围。形如 0.1, 0.1, 0.2, 0.3 (最小值和最大值能不能可视化数据范围 20%，如果调整过大可能会无作用)



风格



- 边框：可以选择是否展示图片边框，默认展示
- 网格：可以选择是否展示网格，默认不展示
- 可以选择是否进行 x、y 轴颠倒，默认进行颠倒
- 文字大小：控制整体文字大小，默认为 7pt

图片



图片	▼
宽度 (cm)	6
高度 (cm)	7
字体	Arial ▼

- 宽度：图片横向长度，单位为 cm
- 高度：图片纵向长度，单位为 cm
- 字体：可以选择图片中文字的字体



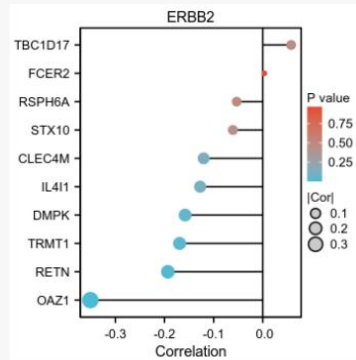
结果说明

主要结果

相关性棒棒糖图-云

相关性棒棒糖图: 分析ERBB2和其他所有变量之间的相关性结果并用棒棒糖图形式展示结果

统计方法: spearman



[相关性棒棒糖图.pdf](#)

[相关性棒棒糖图.tiff](#)

[相关性棒棒糖图.pptx](#)

[分析数据.xlsx](#)

具体相关性情况可查看<补充结果>



补充结果

相关性分析

提供Pearson和Spearman统计方法的结果

主变量	次变量	自由度(df)	统计量-Pearson	相关系数-Pearson	p值-Pearson	统计量-Spearman	相关系数-Spearman	p
ERBB2	OAZ1	161	-3.7074	-0.280457	0.0003	9.75e+05	-0.350893	
ERBB2	TRMT1	161	-2.08243	-0.161952	0.0389	8.44e+05	-0.169294	
ERBB2	STX10	161	-0.968776	-0.0761287	0.3341	7.656e+05	-0.0607983	
ERBB2	RETN	161	-2.06727	-0.160803	0.0403	8.612e+05	-0.193218	
ERBB2	FCER2	161	-0.853311	-0.0670987	0.3948	7.194e+05	0.00326035	
ERBB2	DMPK	161	-3.29184	-0.25112	0.0012	8.359e+05	-0.158107	
ERBB2	CLEC4M	161	-1.19814	-0.0940084	0.2326	8.087e+05	-0.120386	
ERBB2	RSPH6A	161	-1.51854	-0.11883	0.1308	7.602e+05	-0.0531982	
ERBB2	TBC1D17	161	-0.381143	-0.0300247	0.7036	6.802e+05	0.0576477	
ERBB2	IL4I1	161	-1.42929	-0.111936	0.1549	8.139e+05	-0.127596	

相关性.xlsx

相关系数为正，说明两个变量之间存在正相关关系；相关系数为负，说明两个变量之间存在负相关关系；

相关系数绝对值代表相关程度，0-0.3代表弱或者不相关；0.3-0.5代表弱相关；0.5-0.8代表中等程度相关；0.8-1代表强相关

相关是否有统计学意义还需要结合p值来查看

这里提供相关性分析表：可以查看主要变量与其他变量之间的相关系数与其对应的统计学 p 值

- 相关系数为正数，说明两个变量（主要变量与其他变量）之间可能存在正相关关系；相关系数为负数，说明两个变量可能存在负相关关系
 - 相关系数绝对值在 0.8-1.0 之间，说明两个变量之间强相关
 - 相关系数绝对值在 0.5-0.8 之间，说明两个变量之间中等程度相关
 - 相关系数绝对值在 0.3-0.5 之间，说明两个变量之间相关程度一般
 - 相关系数绝对值在 0.0-0.3 之间，说明两个变量之间弱相关或者不相关
- 相关是否有统计学意义还需要结合 p 值来查看

方法学

统计分析和可视化均在 R 4.2.1 版本中进行

涉及的 R 包：ggplot2 包（用于可视化）

处理过程：

- (1) 对数据中主变量和次要变量之间进行相关性分析
- (2) 分析结果用 ggplot2 包进行棒棒糖图可视化



如何引用

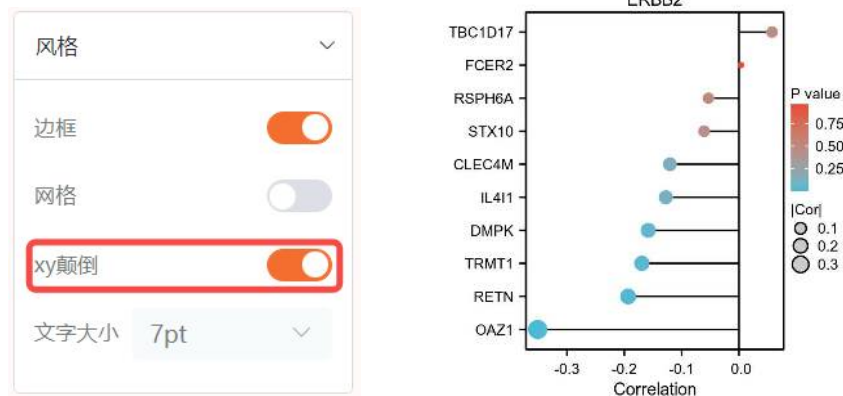
生信工具分析和可视化用的是 R 语言，可以直接写自己用 R 来进行分析和可视化即可，可以无需引用仙桃，如果想要引用仙桃，可以在致谢部分 (Acknowledge) 致谢仙桃学术 (www.xiantao love)。

方法学部分可以参考对应说明文本中的内容以及一些文献中的描述。

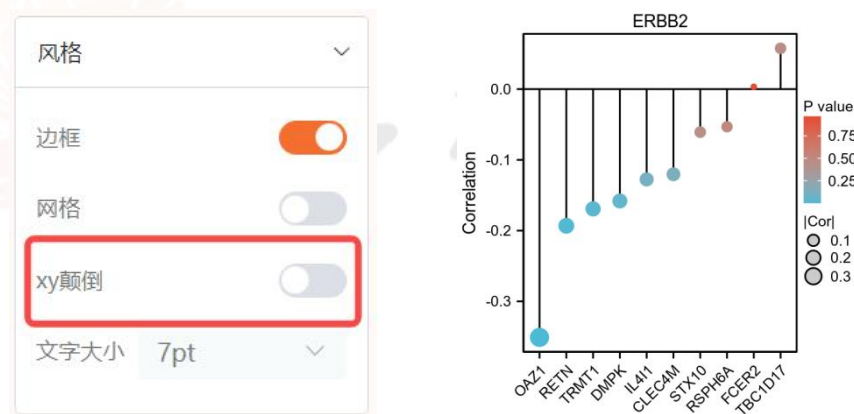


常见问题

1. 为什么数据说明与主要结果说明跟图的坐标对应不上?



答: 主要结果显示的是默认将图形进行转置之后得到的, 所以会跟数据说明与主要结果说明反过来了, 如下 (左侧为参数设置, 右侧为不转置结果)



2. 方法里面的 Spearman 和 Pearson 方法, 应该选择哪一个?

答: 两种方法均可以选择。Pearson 会要求数据是满足正态性, Spearman 因为是非参数的方法, 可以不需要满足。可以先选择非参数的 Spearman 相关进行尝试。

3. 相关系数多少为好?

答： 这个没有很统一的标准，可以参考以下：

➤ 相关系数强弱：

- 绝对值在 0.8 以上：强相关
- 绝对值在 0.5-0.8：中等程度相关
- 绝对值在 0.3-0.5：相关程度一般
- 绝对值在 0.3 以下：弱或者不相关

