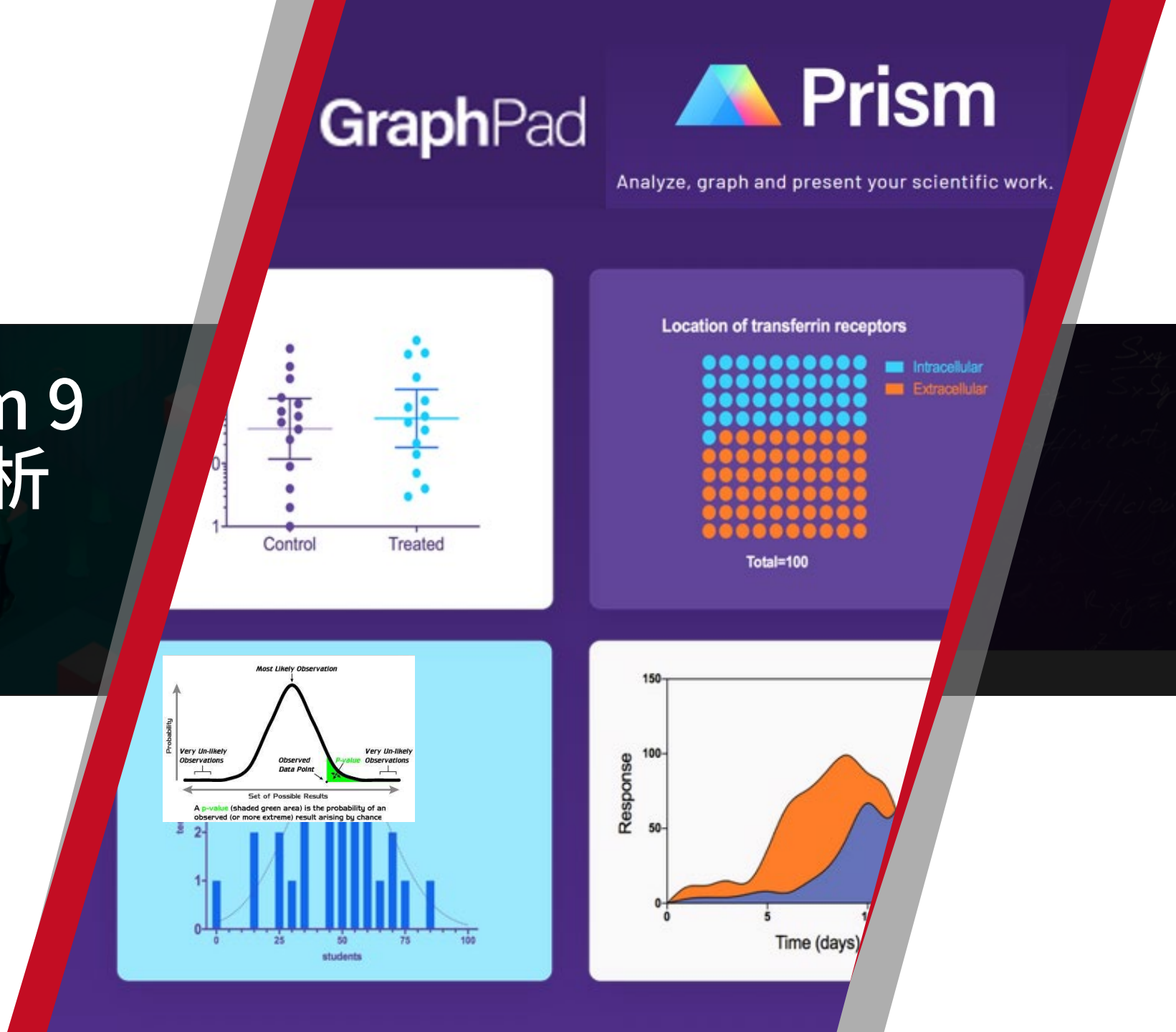
解螺旋

# GraphPad Prism 9 常用数据统计分析

## 科研技能 单元课 03 增刊

- 针对特定科研技能的专项突破





# 课程目录

## 数据统计分析概述

**在对于统计学方面而言，可谓是繁琐而冗长。所以在这里仅对一些常用的统计方法及软件应用进行简单教学，下面的概述更偏向于应用：**

➢ 参数检验与非参数检验:

1、参数检验是针对参数做的假设

非参数检验是针对总体分布情况做的假设，这是区分的一个重要特征

2、参数检验要利用到总体的信息（总体的分布、总体的一些参数特征，如方差），以总体分布和样本信息对总体参数做出推断

非参数检验不需要利用总体信息，以样本信息对总体分布做出推断

➢ 数据表示方法:

① 计量资料符合正态分布采用均数±标准差表示，例：本班男生身高为175±3cm

② 资料不符合正态分布采用中位数(四分位数间距)表示，例：该组患者胆红素水平为50（21，70）μM

③ 计数资料采用率或构成比表示，例：该组患者复发率为30%

➢ 假设检验:（假设检验的选择应该根据数据类型进行）

① 符合正态分布与方差齐性的两组间计量资料比较采用两独立样本t检验

② 不符合正态分布与方差齐性的两组计量资料比较采用非参数Mann-Whitney U检验

③ 配对设计计量资料比较，差值符合正态分布采用配对t检验；差值不符合正态分布采用非参数Willcoxon秩和检验（Sign检验）

④ 多组间计量资料比较符合条件采用单因素设计方差分析；组间两两比较采用LSD法（ SNK法/ Bonferroni法/Duncan法）；多组间计量资料比较不符合应用条件采用Kruskal-Wallis H法

⑤ 配伍组设计计量资料比较符合条件采用随机区组设计方差分析；不符合条件采用非参数Friedman检验

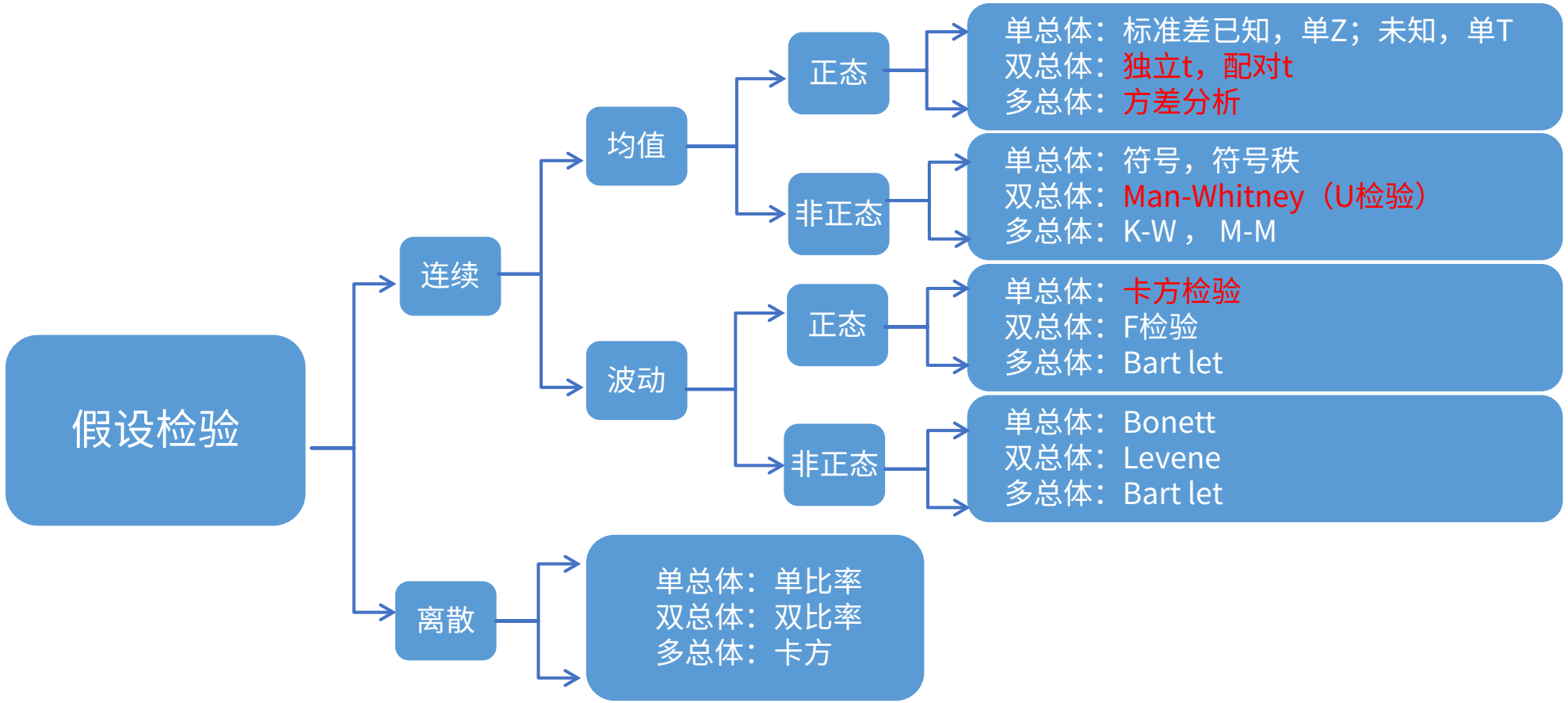
⑥ 等级资料组间比较采用非参数秩和检验

⑦ 成组四格表计数资料符合条件采用Pearson卡方检验；不符合条件采用Pearson连续校正卡方检验或Fisher 's确切概率法

⑧ 成组设计R*C表符合条件采用Pearson卡方检验；不符合条件采用MonteCarlo近似确切概率法

## 数据统计分析概述

- 也可以根据下表对适用的统计方法进行选择
- 本课程仅对红字内容进行介绍



## 数据统计分析概述

➢ 相关分析

① 双变量正态分布资料采用Pearson相关系数

② 双变量非正态分布或等级资料采用Spearman相关系数

③ 当面临两组多变量数据，并希望研究两组变量之间的关系时，就要用到典型相关分析（canonical correlation analysis）

➢ 回归分析:

① 线性回归：因变量为计量资料的影响因素分析采用多元线性回归

② Logistic回归：二分类因变量影响因素分析采用二元Logistic回归（Binary Logistic）

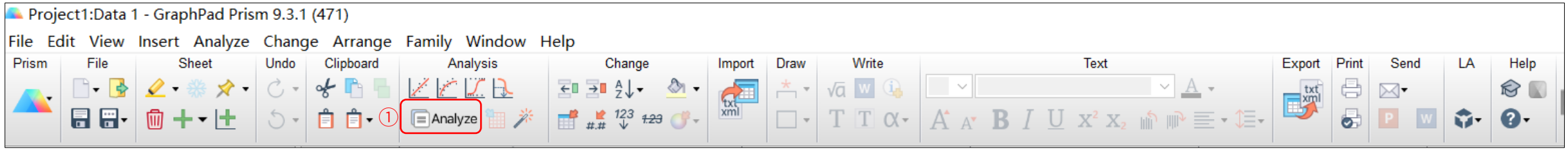
③ 有序因变量影响因素分析采用有序逻辑回归（Ordinal Logistic）

④ 无序多分类计数资料因变量影响因素分析采用多项Logistic回归（Multinominal Logistic）

➢ *P*值:

*P*＜0.05或*P*＜0.01具有统计学意义；如有可能，请将*P*值报告为一个精确到有几位数的数值，而非一个不等式。例如说“*P*值为0.0234”，而非“*P*＜0.05”

## 数据统计分析概述

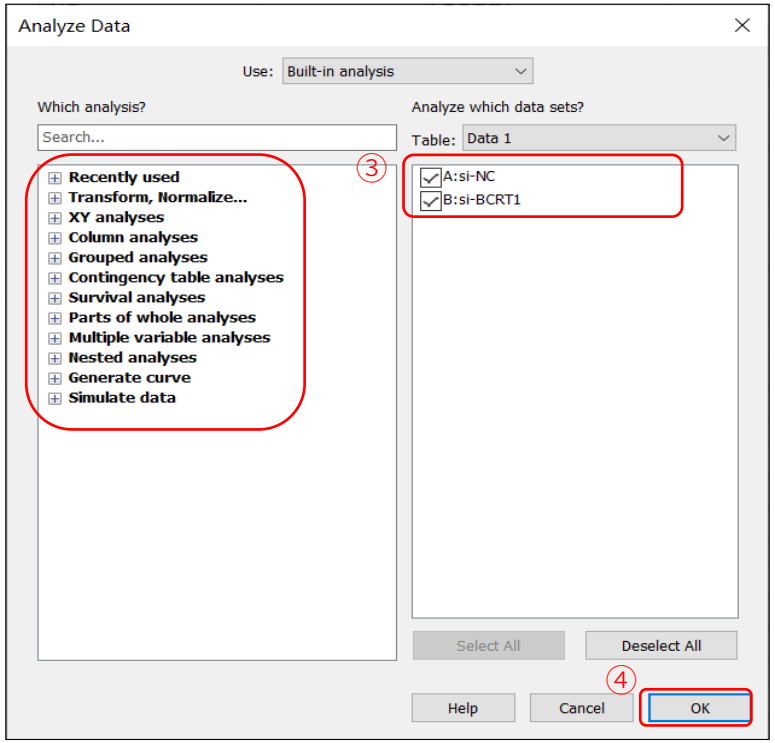



① GraphPad Prism 9里，在数据或者作图页面，点击顶部菜单栏的Analyze按钮，即可调出数据分析功能窗口

② 根据需要的分析功能对数据进行选择

③ 选择需要分析的数据

④ 选择OK



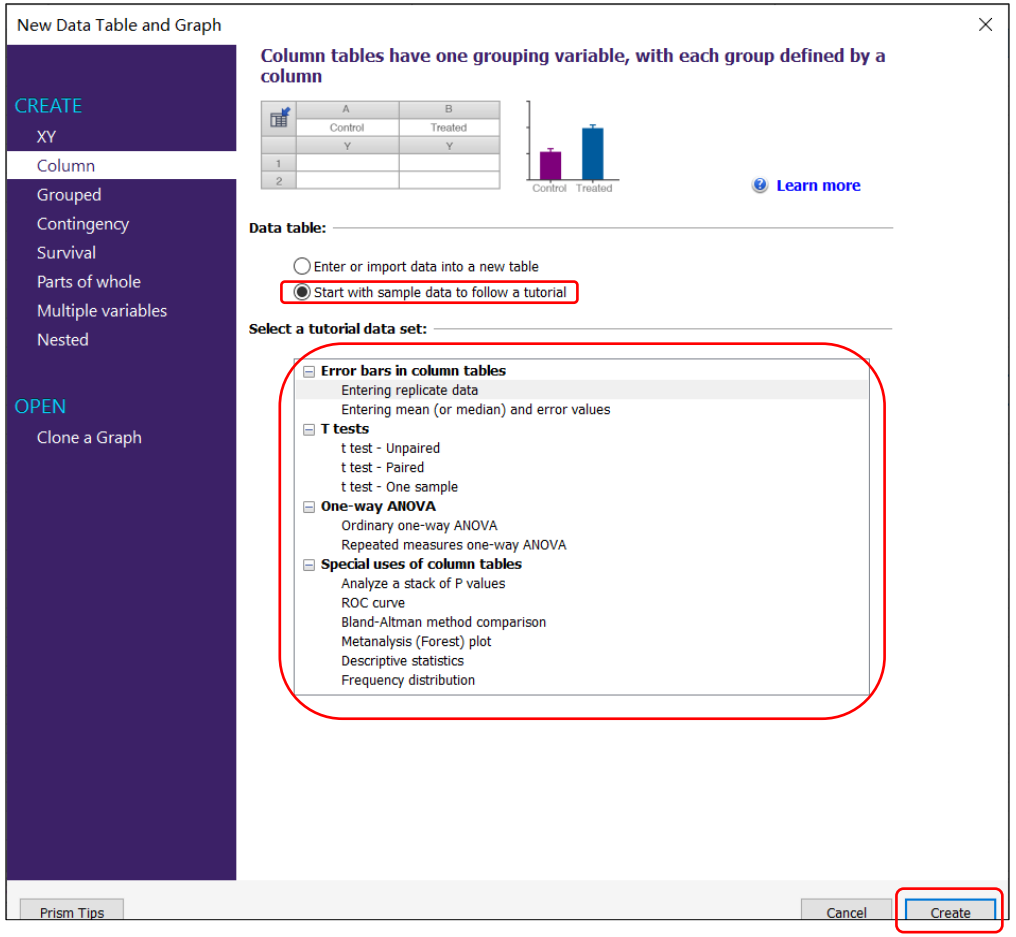

②

(1) Recently used 最近使用
(2) Transform， Normalized 转化，标准化
(3) XY analyses XY分析
(4) Column analyses Column分析
(5) Grouped analyses 成组分析
(6) Contingency table analyses 列联表分析
(7) Survival analyses 生存分析
(8) Parts of whole analyses 部分整体分析
(9) Multiple variable analyses 多变量分析
(10) Nested analyses 嵌套分析
(11) Generate curve 曲线生成
(12) Simulate data 模拟数据

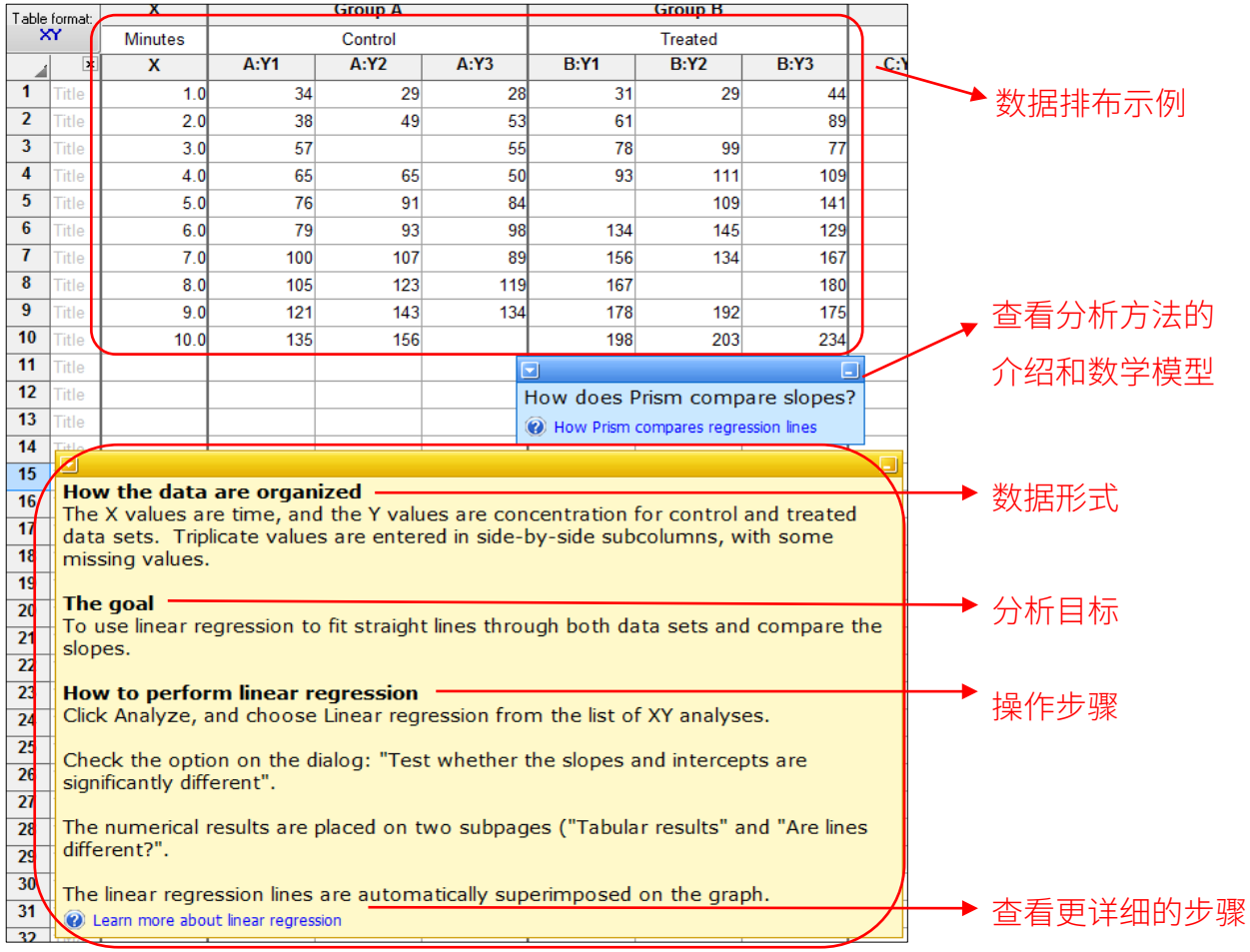
## 数据统计分析概述

- GraphPad Prism的数据分析功能强大，且针对不同的分析，还自带示例供参照学习：

① 在新建数据表格的窗口中，可选择某一分析的示例数据




② 点击“Create”后，新创建数据页面中即给出示例数据和说明




数据排布示例

查看分析方法的介绍和数学模型

数据形式

分析目标

操作步骤

查看更详细的步骤

# 课程目录

## 假设检验

- 假设检验（hypothesis test）主要是验证所选的模型和所解释的公式，在结构上、形式上、变化方向上是否能代表客观情况。一般采用最小平方法解模时必须进行统计假设检验，即应用统计推断的假设检验原理，通过特定的统计量进行各种显著性检验。假设检验的本质是利用样本对假设的真假进行判断

- 要进行统计假设的检验， 必须利用各种不同的判据， 即利用规则来选择。先假设总体某项假设成立，计算其会导致什么结果产生。若导致不合理现象产生，则拒绝原先的假设。若并不导致不合理的现象产生，则不能拒绝原先假设，从而接受原先假设。假定样本数据来自具有特定分布的总体，则使用参数检验。如果不能对数据集作出必要的假设，则使用非参数检验

- 假设检验的一般归类：

① 如果已知分布的均值和方差，则可以构造标准正态分布，即u检验

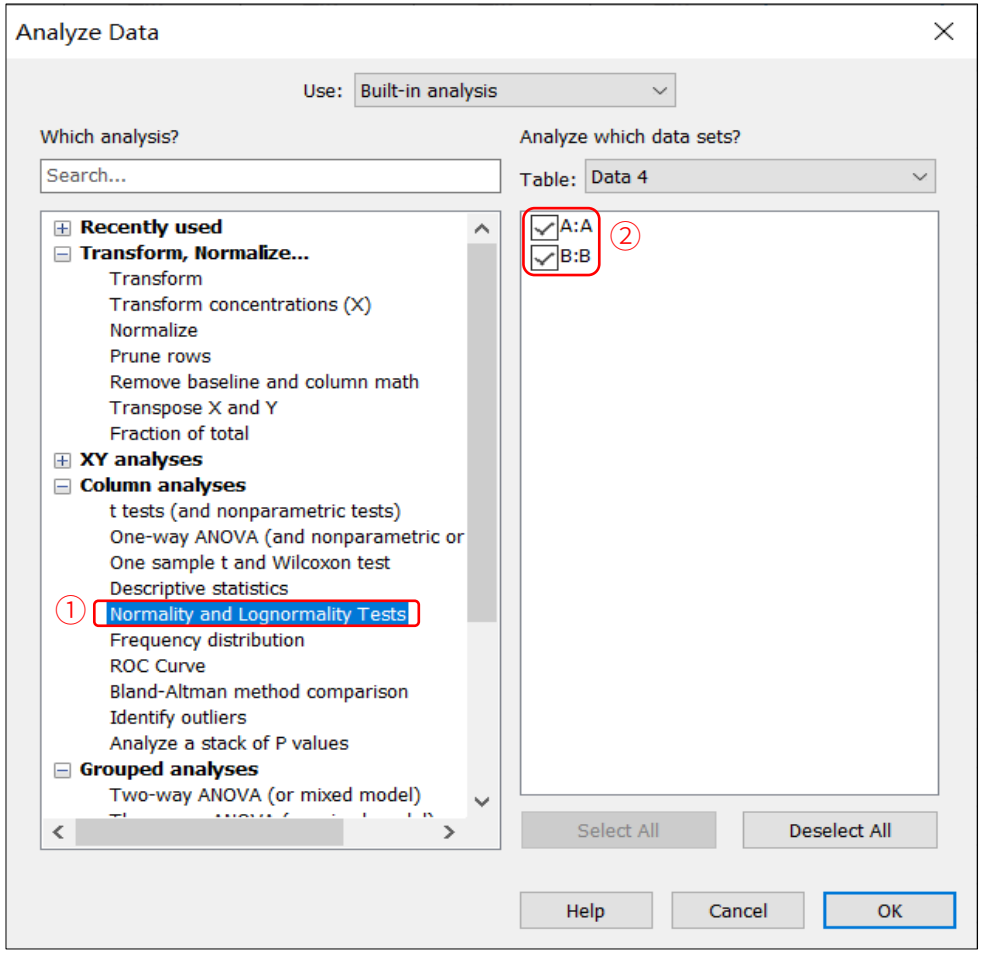
② 如果已知分布的均值，不知道方差，则可以用方差的无偏估计代替方差，即t检验法

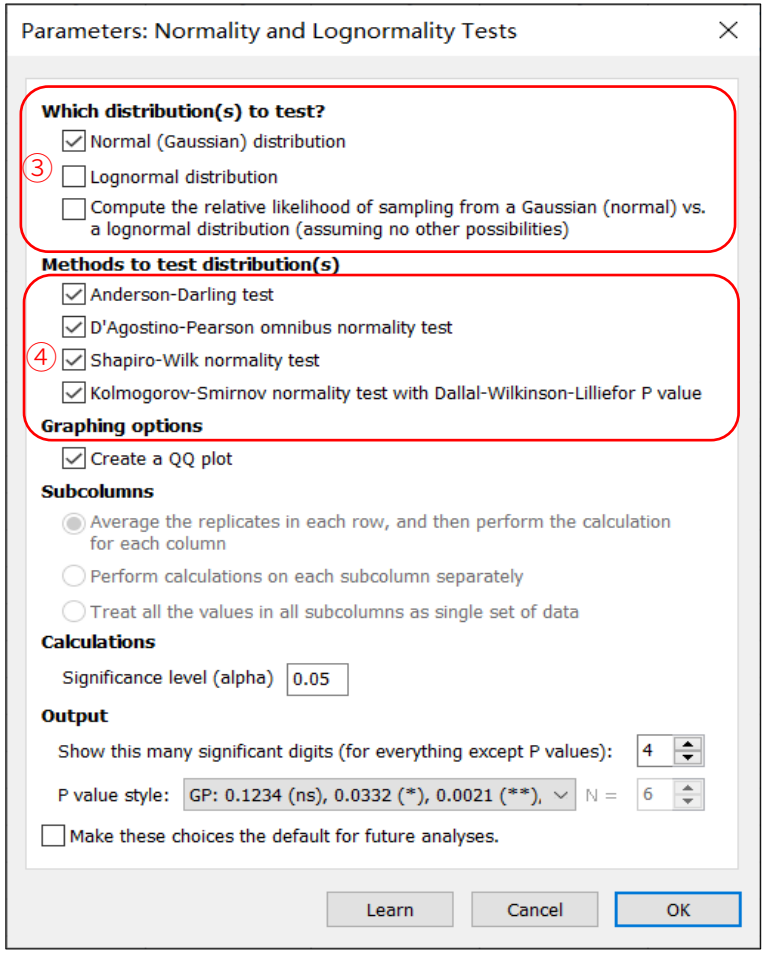
③ 如果均值、方差都不知，利用样本的方差，构造卡方分布，即卡方检验

④ 如果均值、方差都不知，存在两个分布的比较，构造F分布，查F分布表，即F检验

## 正态检测

1 在选择分析方法前，需要对数据的分布及方差进行检测；点击Analyze按钮，调出数据分析窗口




① 选择正态检验

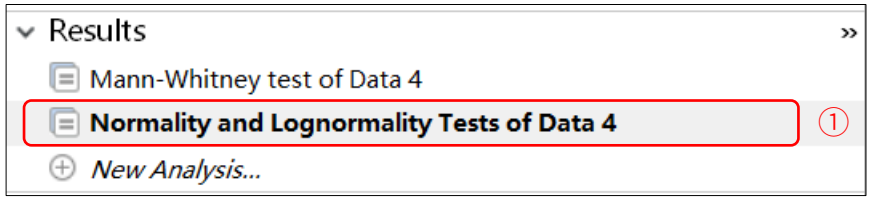
② 选择要检验的数据

③ 选择要检验什么分布

④ 选择检验的方法

针对同一种数据，四种方法的计算结果大同小异；虽然GraphPad Prism不推荐用Kolmogorov-Smirnov法，小样本多采用Shapiro-Wilk normality test

## 正态检测




① 在左边导航栏里找到结果列表：在上方Results中第二栏为检测结果，单击结果如右图

② 四种方法的检验结果

✓ 常用的是 Shapiro-Wilk 检验（SW）和Kolmogorov-Smirnov 检验（KS）

✓ 大多数情况下，三种方法的结论一致；不一致时，一般小样本（n<30，有争议，也有人认为n<200，还有人认为是n<2000）以SW法计算结果为准；大样本以KS法计算结果为准

③ 一般认为，当*p* value > 0.1（也有人认为*p*>0.05）时，数据服从正态分布

✓ 这里的0.1（或0.05）是置信度，理解：认为10%（5%）为小概率事件，即比这个几率小的事件，我们认为在实验中不会发生。设定*p* > 0.1认为数据符合正态分布，对正态分布的检验更保守

| | Normality and Lognormality Tests<br>Tabular results | A<br>A | B<br>B |
|---|---|---|---|
| 1 | **Test for normal distribution** | | |
| 2 | **Anderson-Darling test** | | |
| 3 | A2* | 0.5845 | 0.1982 |
| 4 | P value | 0.1225 | 0.8649 |
| 5 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 6 | P value summary | ns | ns |
| 7 | | | |
| 8 | **D'Agostino & Pearson test** | | |
| 9 | K2 | 6.224 | 0.6972 |
| 10 | P value | 0.0445 | 0.7057 |
| 11 | Passed normality test (alpha=0.05)? | No | Yes |
| 12 | P value summary | * | ns |
| 13 | | | |
| 14 | **Shapiro-Wilk test** | | |
| 15 | W | 0.9653 | 0.9747 |
| 16 | P value | 0.0857 | 0.8938 |
| 17 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 18 | P value summary | ns | ns |
| 19 | | | |
| 20 | **Kolmogorov-Smirnov test** | | |
| 21 | KS distance | 0.1025 | 0.1107 |
| 22 | P value | >0.1000 | >0.1000 |
| 23 | Passed normality test (alpha=0.05)? | Yes | Yes |
| 24 | P value summary | ns | ns |

## U检验

- 示例：文献： Partner R, Jones B, Tee J, Francis P. Playing through the pain: The prevalence of perceived shoulder dysfunction in uninjured rugby players using the Rugby Shoulder Score. Phys Ther Sport. 2022 Jan 16;54:53-57. doi: 10.1016/j.ptsp.2022.01.001. Epub ahead of print. PMID: 35065316. Table 2.

示例解读

- 多个两组数据
- 数据表现形式为均值±方差或中位数（四分位差）
- 每种分组依据不同，结果源于不同个体（非配对），两组数据在个数上可以不相等

| | n | RSS Score |
|---|---|---|
| All Shoulders | 139 | 35 ± 20<br>23 (28) |
| No performance impact (Group 2) | 60 | 40 (22) |
| Performance impact (Group 3) | 17 | 61 (28) |
| | | p = 0.02 |
| Rugby League | 125 | 25 (26) |
| Rugby Union | 14 | 20 (25) |
| | | p = 0.41 |
| Professional | 58 | 40 (39) |
| Amateur | 81 | 20 (18) |
| | | p = 0.02 |
| Forwards | 73 | 29 (29) |
| Backs | 66 | 20 (21) |
| | | p = 0.036 |
| Previous Injury | 48 | 48 (41) |
| No Previous injury | 91 | 20 (13) |
| | | p < 0.001 |

| Table format: Grouped | | Group A | Group B |
|---|---|---|---|
| | | A | B |
| 1 | Title | 42 | 81 |
| 2 | Title | 78 | 106 |
| 3 | Title | 14 | 27 |
| 4 | Title | 43 | 90 |
| 5 | Title | 62 | 115 |
| 6 | Title | 29 | 95 |
| 7 | Title | 68 | 70 |
| 8 | Title | 77 | 54 |
| 9 | Title | 38 | 25 |
| 10 | Title | 15 | 43 |
| 11 | Title | 11 | 89 |
| 12 | Title | 54 | 3 |
| 13 | Title | 59 | 31 |
| 14 | Title | 34 | 52 |
| 15 | Title | 57 | 71 |
| 16 | Title | 29 | 49 |
| 17 | Title | 51 | 80 |
| 18 | Title | 80 | |
| 19 | Title | 70 | |
| 20 | Title | 57 | |
| 21 | Title | 37 | |

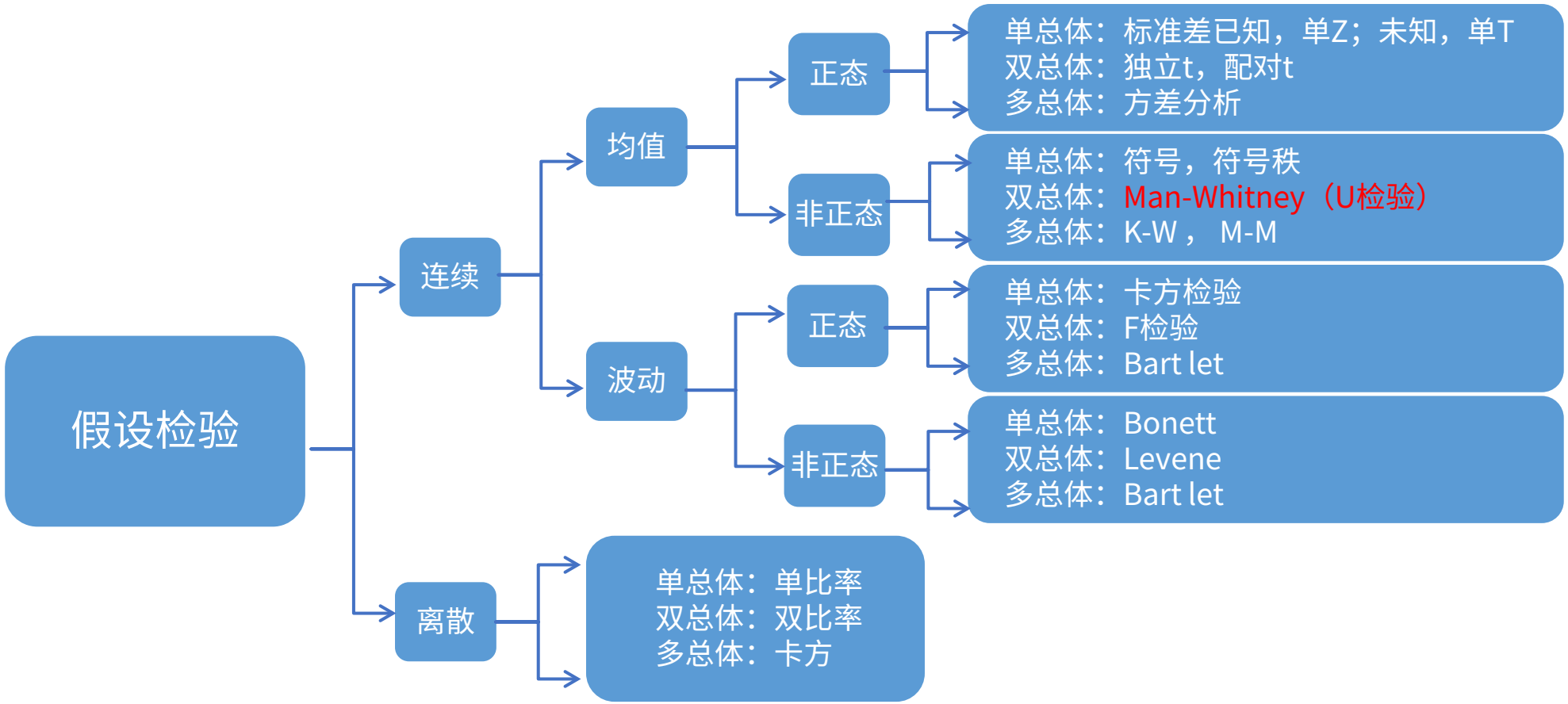
模拟数据：

No performance impact组的RSS Score 组60个数据；

performance impact组17个数据

## U检验

- 分析目标
  - ✓ 数据总体上看，是对shoulders的多个一维分组，如：有或无功能影响；成员来自橄榄球联会或橄榄球联盟；专业运动员或业余运动员；前锋或后卫；受过伤或没受过伤。所有的数据以均值±方差或中位数（四分位差）表示
  - ✓ 从实验结果上看，需要每个分组是否存在显著性差异，比较一维两组独立样本，要看数据是否符合正态分布（本组数据组明显不符合），从而使用U检验



## U检验

- 数据录入及作图请查看单元课03《GraphPad Prism 9 作图教程》，本教程为单元课03增刊，旨在讲述在作图后如何进行统计分析，作图操作不再重复

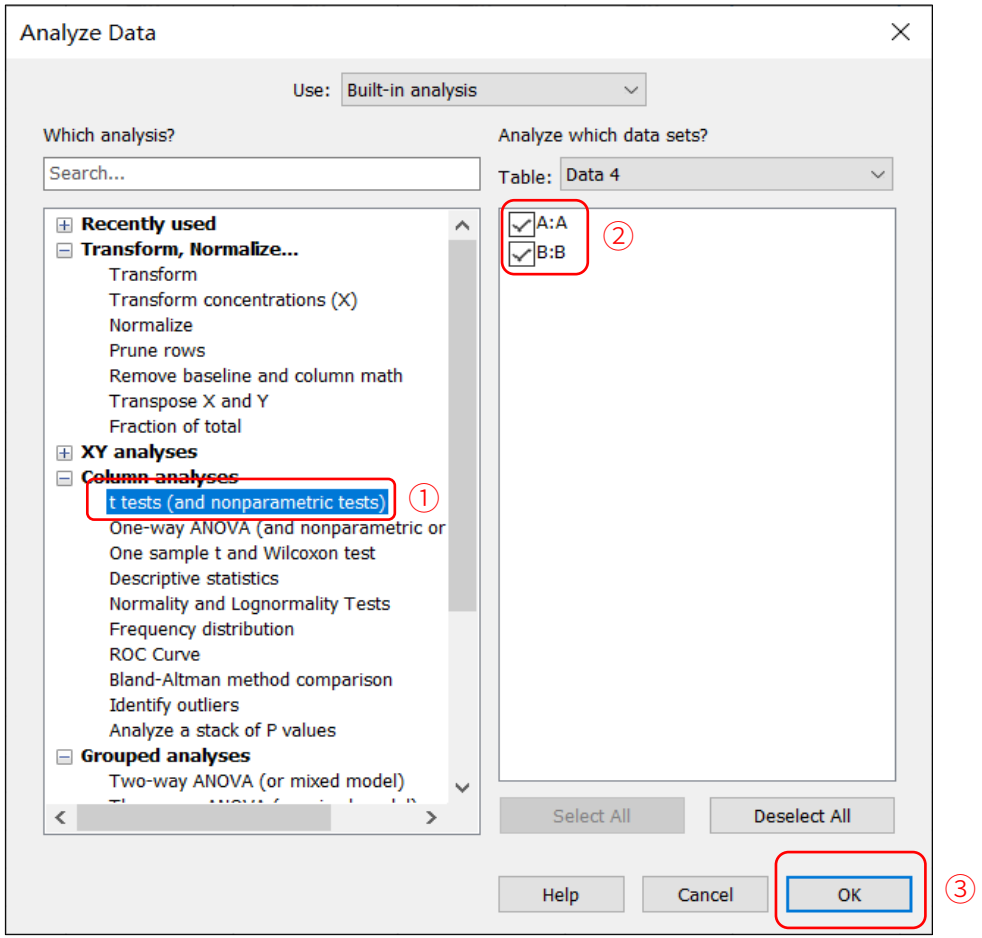
1 Built in analysis

点击Analyze按钮，调出数据分析窗口

① 点击t test（在选择t检验时，也是点击此选项，不要疑惑为什么做U检验时，点的是t检验）

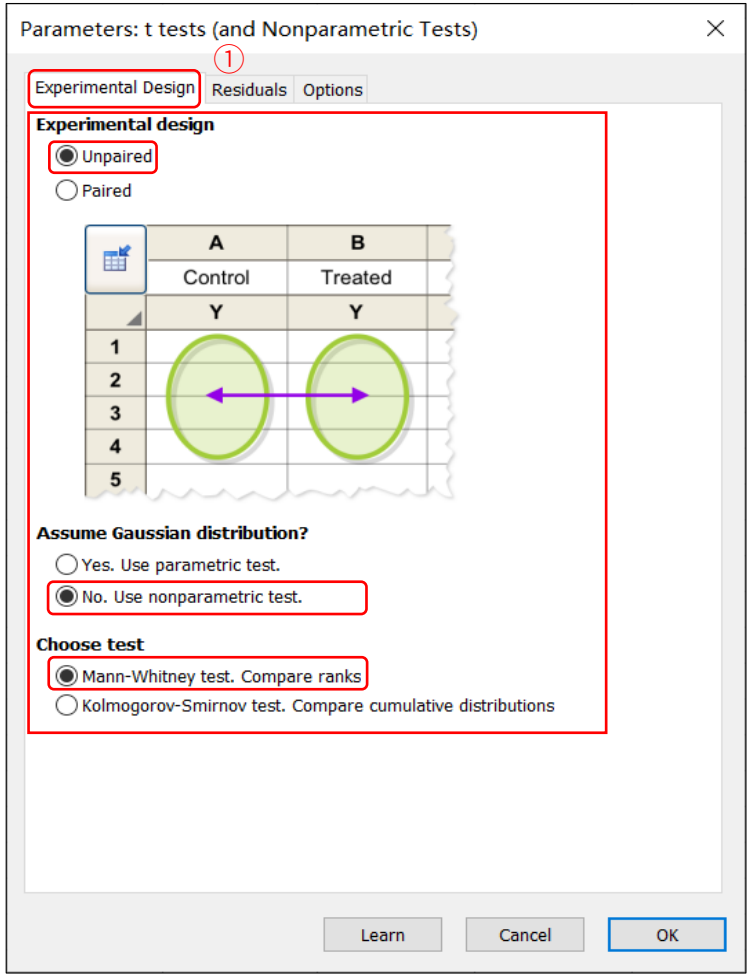
② 勾选AB两组数据

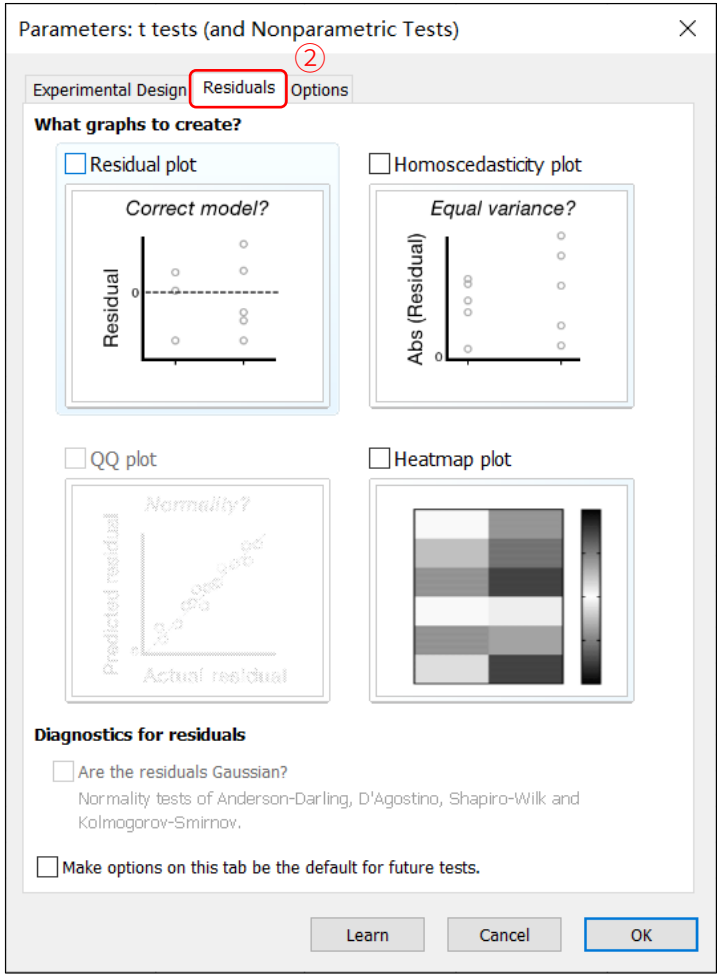
③ 点击OK



## U检验

### ❷ 参数设置1




参数设置包括实验设计，残差以及选项

①实验设计

U检验选择非配对，非参数检验，M-W法

②残差图选择（当前版本的Prism可以直接设置残差图）

勾选你需要的残差图种类

（1）实验设计

（2）非配对

（3）配对

（4）是否服从正态分布

（5）是，使用参数检验

（6）否，使用非参数检验

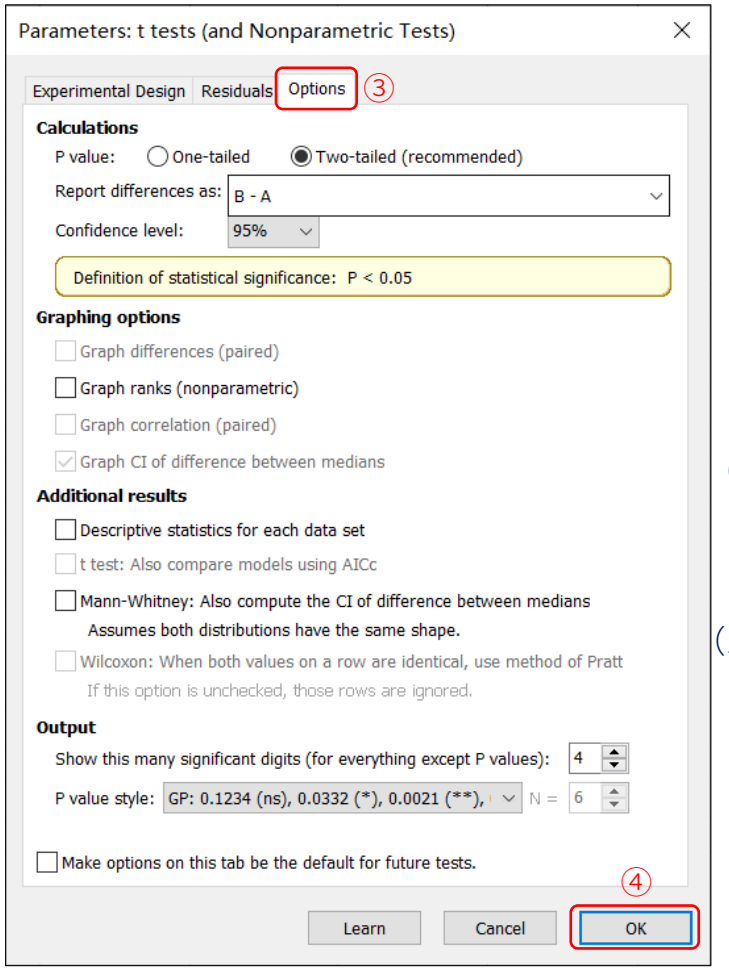
（7）选择检验方法

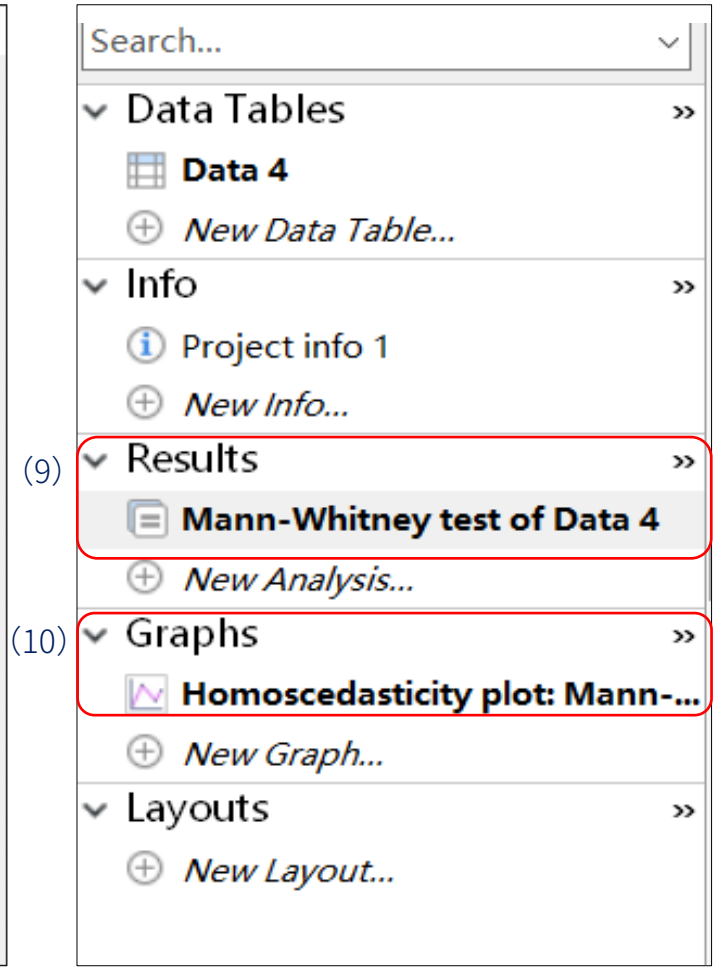
（8）M-W法，检验两个总体的分布是否有显著的差异（一般用这个）

（9）K-S法，检验一个分析变量是否符合某种分布或两个分析变量的分布是否相同

## U检验

### ② 参数设置2




参数设置包括实验设计，残差以及选项

①实验设计

U检验选择非配对，非参数检验，M-W法

②残差图选择（当前版本的Prism可以直接设置残差图）

勾选你需要的残差图种类

③选项

选择结果展示的方式

④点击OK后，可出现分析结果以及上一步中勾选创建的图案

（1）选择单侧检验或是双侧检验， 默认双侧，一般不建议修改

（2）选择比较的方式，谁与谁比

（3）置信水平，一般是95%，即定义$p$<5%时认为有显著差异

（4）绘图选项，可选择绘制统计分析的图，可以不选

（5）补充结果，可以不选

（6）设置统计结果的有效数字

（7）$p$值显示的方式

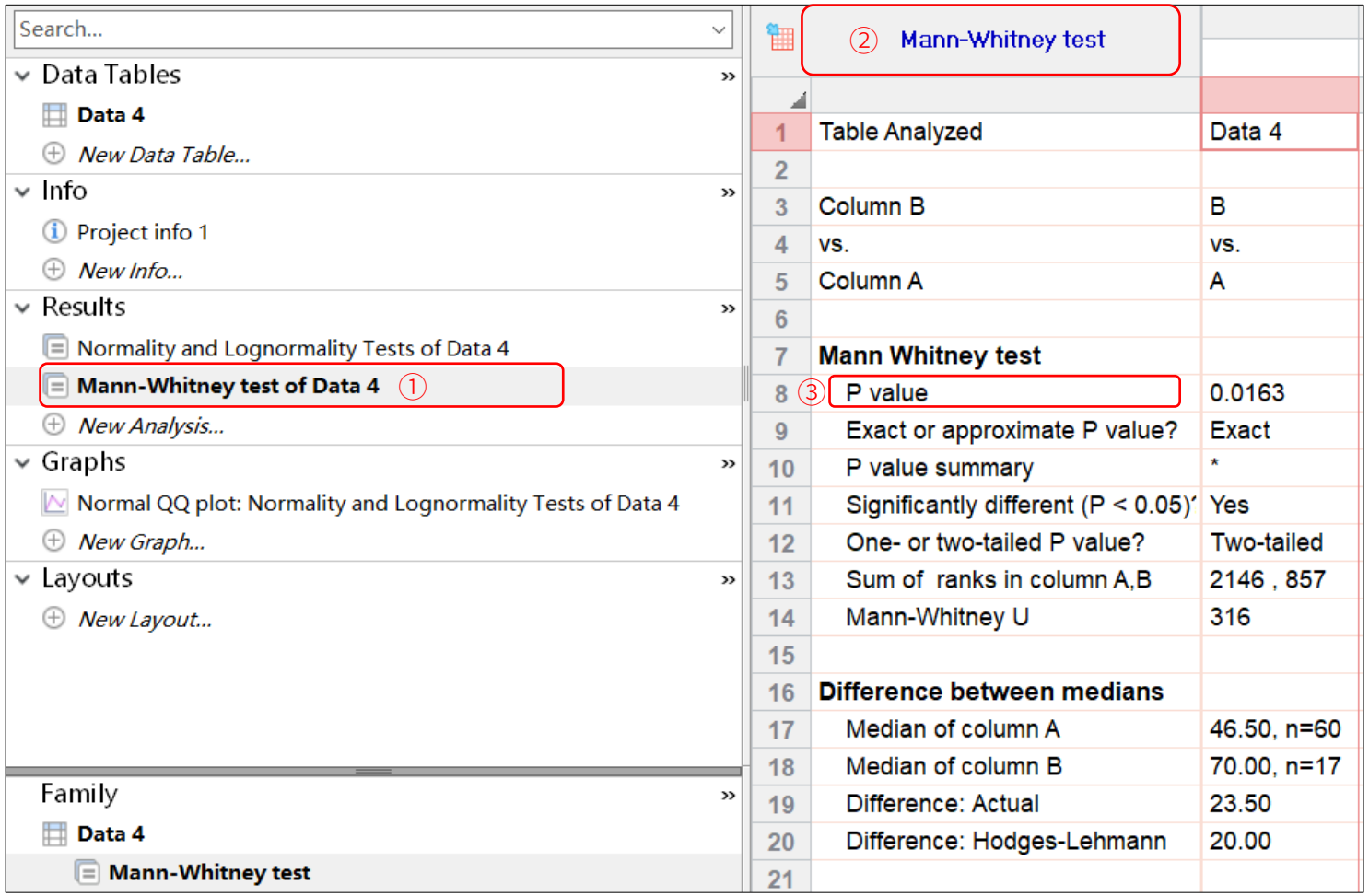
（8）勾选这个可将目前的选择设置为默认选项

（9）分析的结果

（10）残差图

## U检验

### ② 结果显示




① 点击M-W检测显示出右侧检验结果

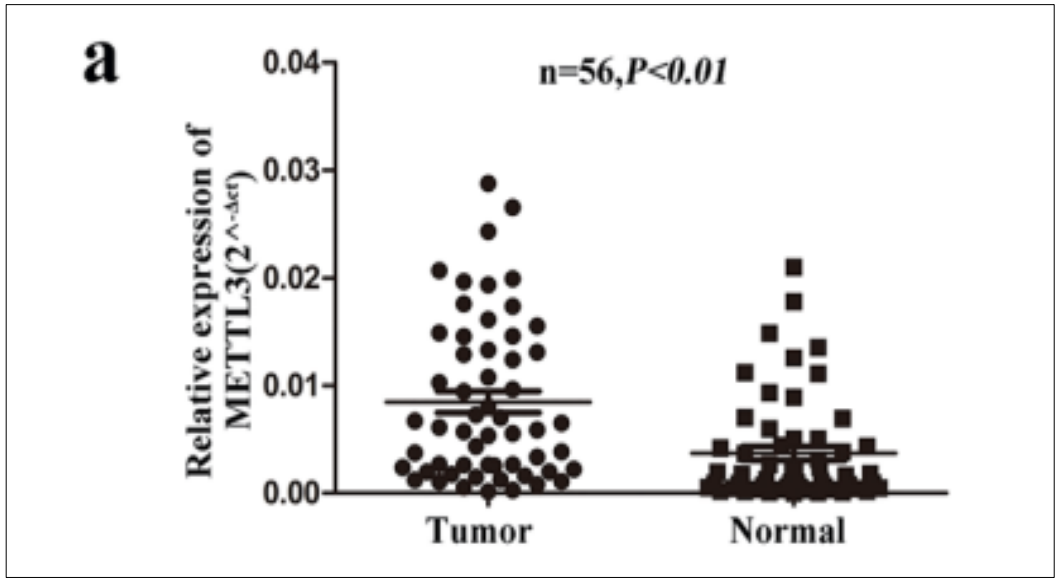
② 点击此按钮可再次调出检验设置窗口，修改检验设置

③ *p* 值=0.0163，两组数据具有显著性差异

## 独立t检验

- 示例：文献示例：Han J, Wang JZ, Yang X, Yu H, Zhou R, Lu HC, Yuan WB, Lu JC, Zhou ZJ, Lu Q, Wei JF, Yang H. METTL3 promote tumor proliferation of bladder cancer by accelerating pri-miR221/222 maturation in m6A-dependent manner. Mol Cancer. 2019 Jun 22;18(1):110. doi: 10.1186/s12943-019-1036-9. PMID: 31228940; PMCID: PMC6588935. Fig.1a

- 示例解读：
  - 该图是一个mettl3在正常组织与肿瘤组织中的mRNA含量的对比
  - 图中X轴仅以不同处理条件作为分类依据（Tumor，Normal），每个分类下，有几十次实验结果




- 模拟数据

| Table format: Column | Group A | Group B |
|---|---|---|
| | Tumor | Normal |
| 1 | Title | 0.01000 | 0.00960 |
| 2 | Title | 0.00100 | 0.00840 |
| 3 | Title | 0.00100 | 0.00390 |
| 4 | Title | 0.00100 | 0.00340 |
| 5 | Title | 0.00200 | 0.00640 |
| 6 | Title | 0.00250 | 0.00540 |
| 7 | Title | 0.00250 | 0.01460 |
| 8 | Title | 0.00250 | 0.00720 |
| 9 | Title | 0.02500 | 0.01220 |
| 10 | Title | 0.00550 | 0.01220 |
| 11 | Title | 0.02550 | 0.00920 |
| 12 | Title | 0.00500 | 0.00220 |
| 13 | Title | 0.02000 | 0.01410 |
| 14 | Title | 0.02000 | 0.02600 |

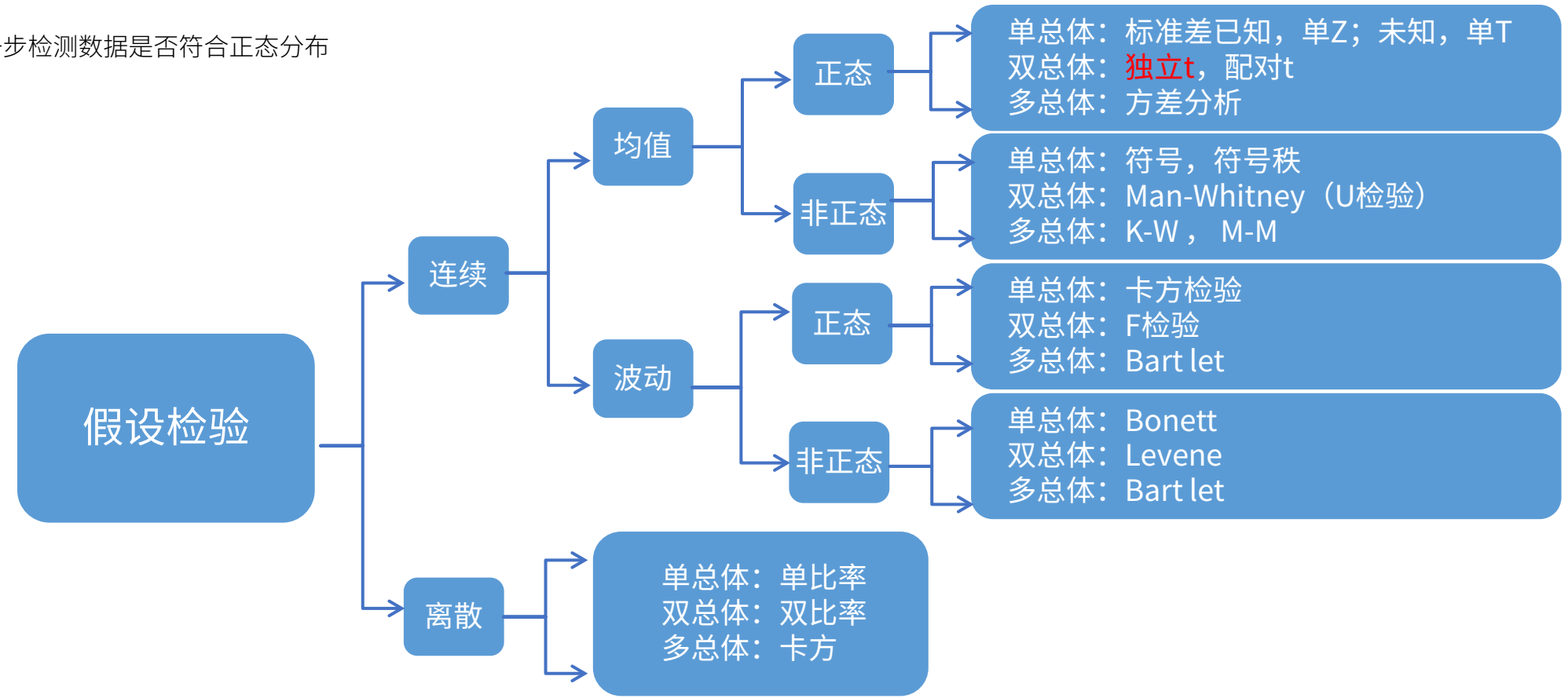
## 独立t检验

- 分析目标

✓ 从数据结构上看，是个一维分组，分为两组，分别为肿瘤组与正常人组

✓ 从实验设计上看，是独立设计；

✓ 下一步检测数据是否符合正态分布

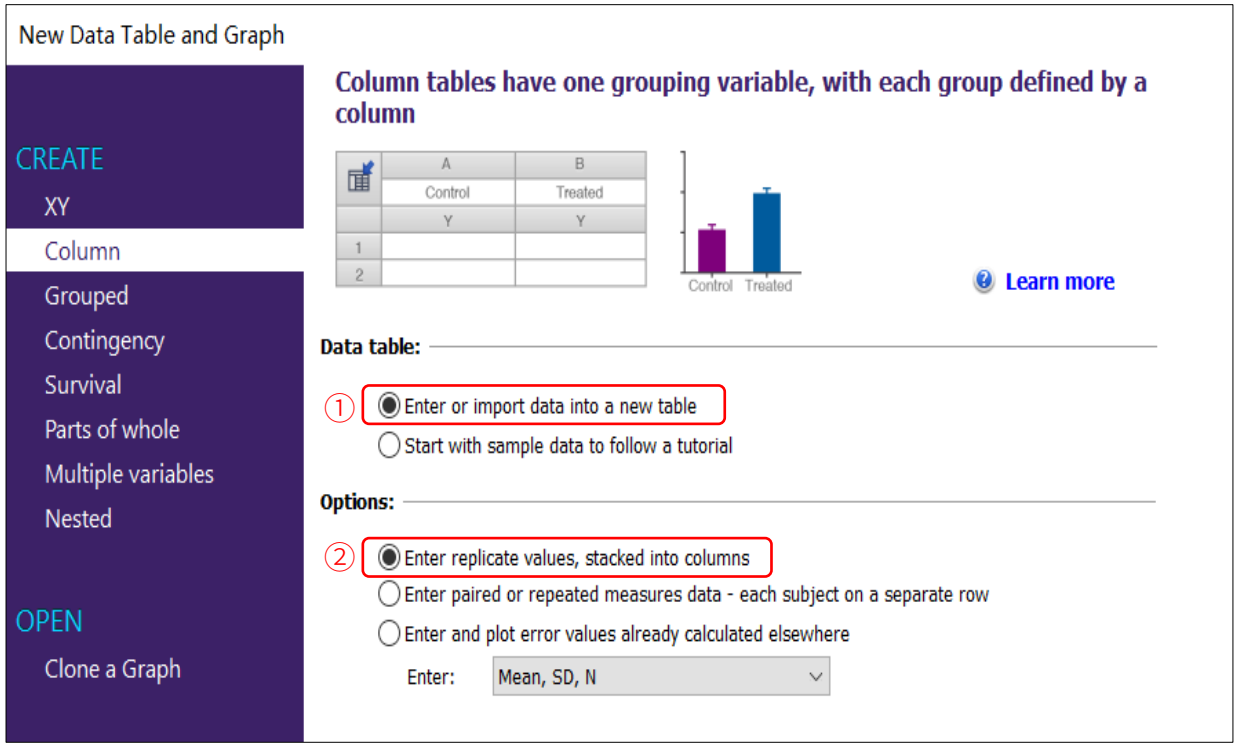


## 独立t检验

1 新建一个Column的数据表（重命名为“t test”），将数据按下图的格式输入

①选择新建表格；

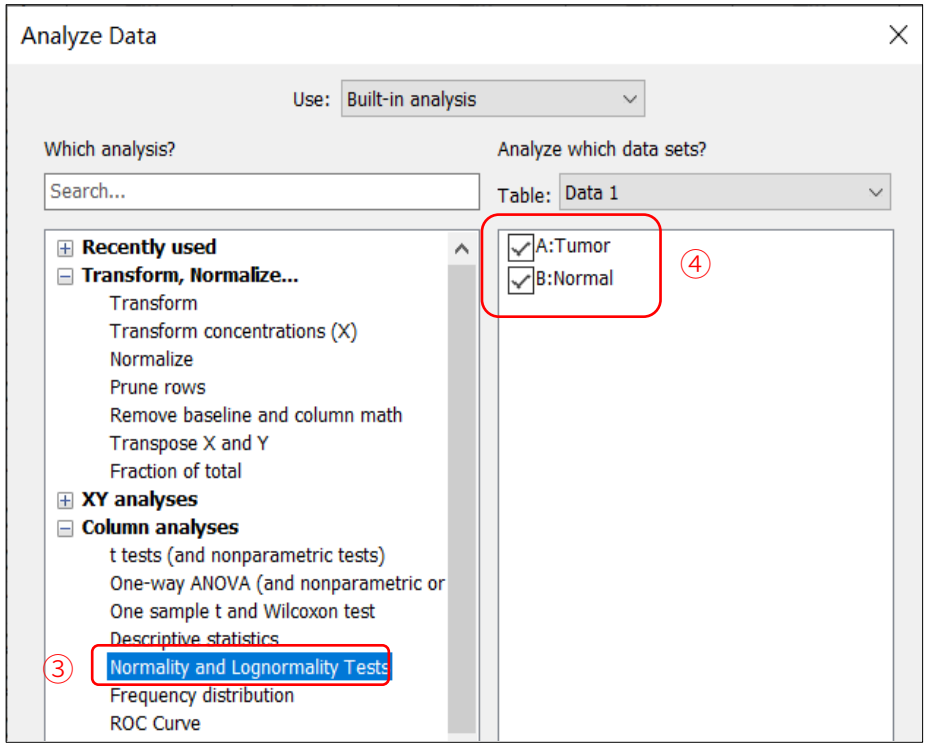
②输入每个值，excel可以拷贝；




2 按前述步骤检验正态分布

③点击分析后选择正态检测；

④勾选所有数据

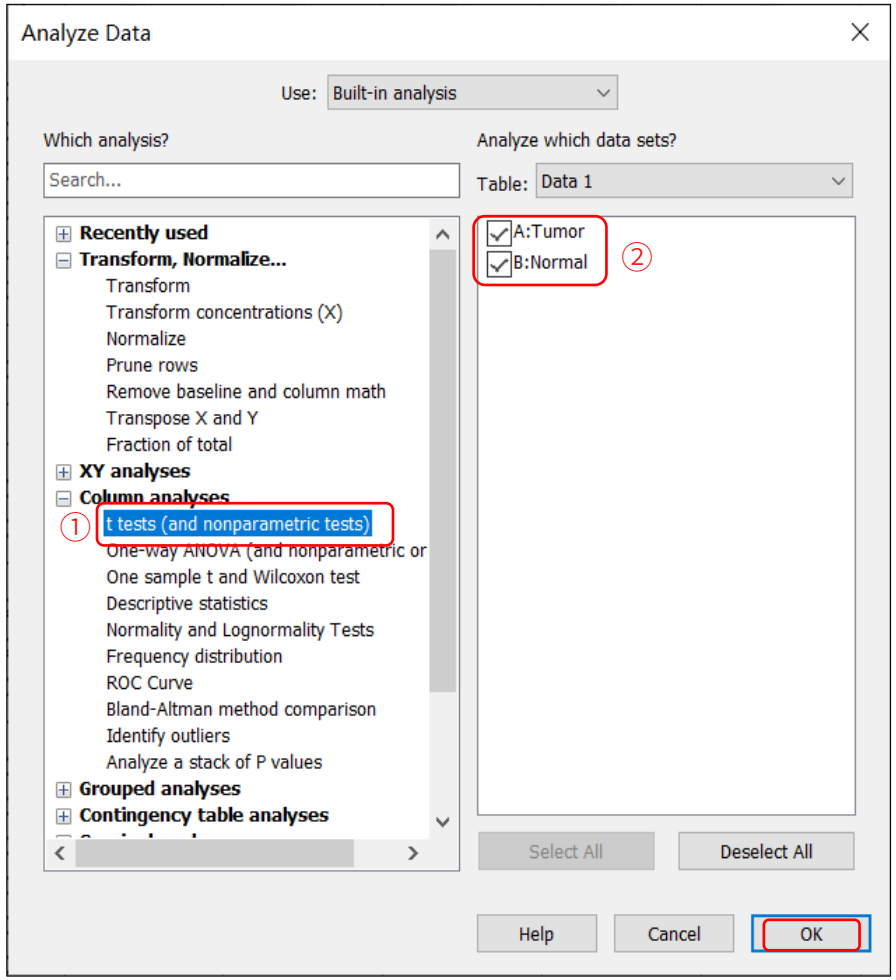


## 独立t检验

✓ 通过正态检验（P value均大于0.1）

| Normality and Lognormality Tests | A | B |
|---|---|---|
| | Tumor | Normal |
| **Test for normal distribution** | | |
| **Anderson-Darling test** | | |
| A2* | 0.5333 | 0.5251 |
| P value | 0.1650 | 0.1731 |
| Passed normality test (alpha=0.05)? | Yes | Yes |
| P value summary | ns | ns |
| | | |
| **D'Agostino & Pearson test** | | |
| K2 | 1.095 | 2.152 |
| P value | 0.5783 | 0.3409 |
| Passed normality test (alpha=0.05)? | Yes | Yes |
| P value summary | ns | ns |
| | | |
| **Shapiro-Wilk test** | | |
| W | 0.9651 | 0.9571 |
| P value | 0.1164 | 0.0513 |
| Passed normality test (alpha=0.05)? | Yes | Yes |
| P value summary | ns | ns |
| | | |
| **Kolmogorov-Smirnov test** | | |
| KS distance | 0.09936 | 0.09648 |
| P value | >0.1000 | >0.1000 |
| Passed normality test (alpha=0.05)? | Yes | Yes |
| P value summary | ns | ns |

3 回到column数据页面，点击顶部Analyze按钮，弹出的窗口里设置检验方法

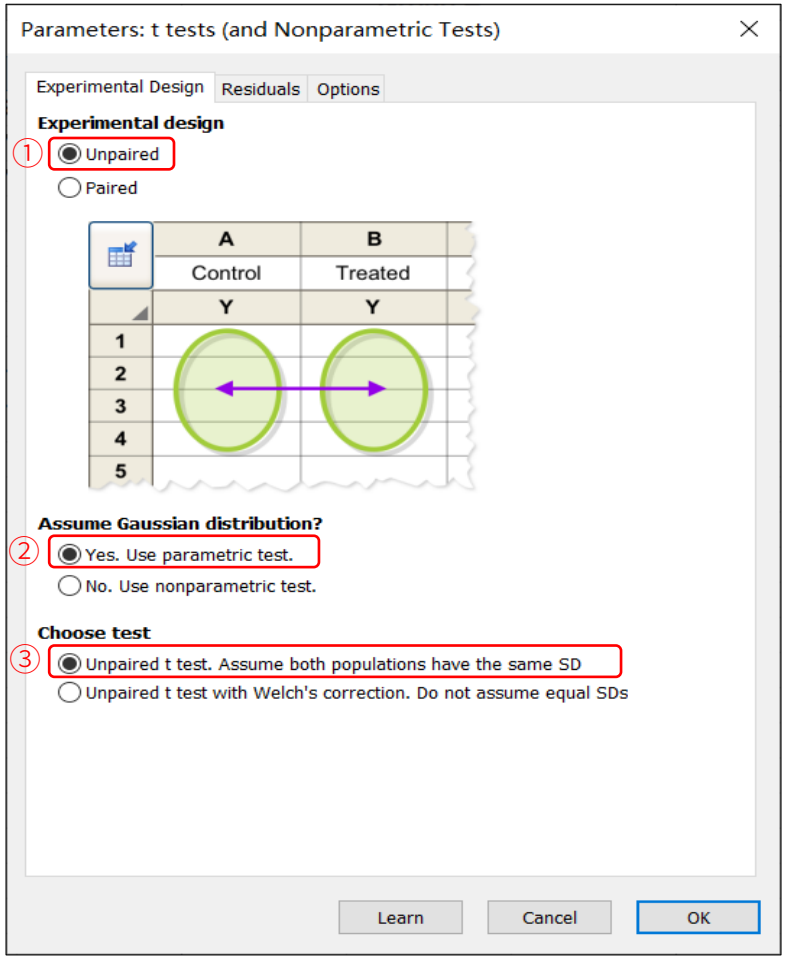


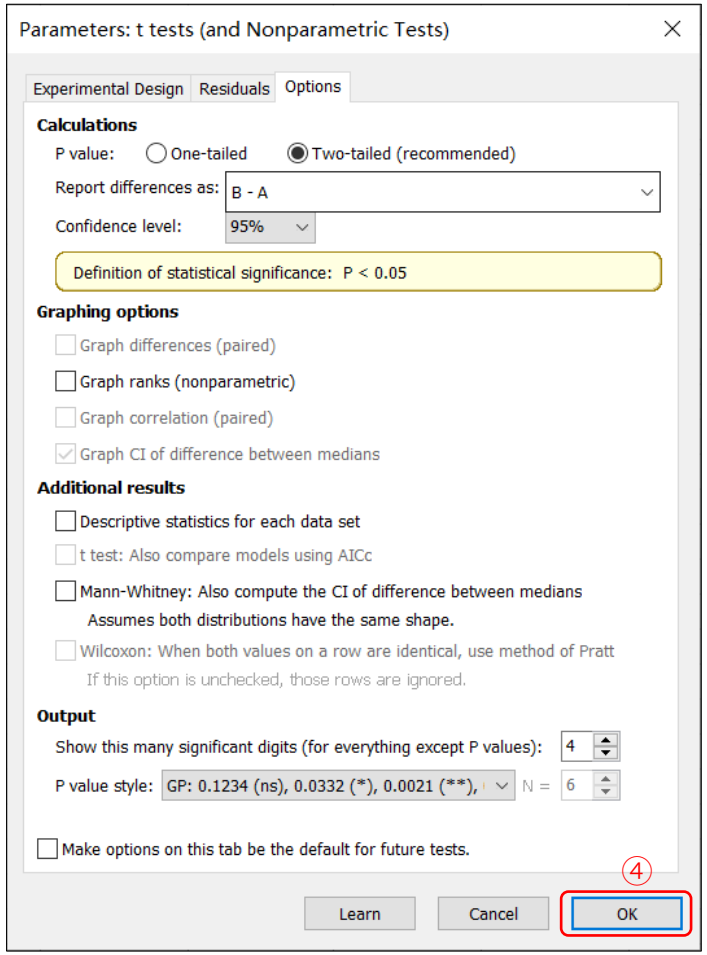

①选择T检验

②勾选所有数据

## 独立t检验

4 弹出的窗口里设置检验参数，然后点击“OK”




①选择非配对设计

②服从正态分布，勾选参数检验

③服从方差齐性，选择非配对t检验不需矫正

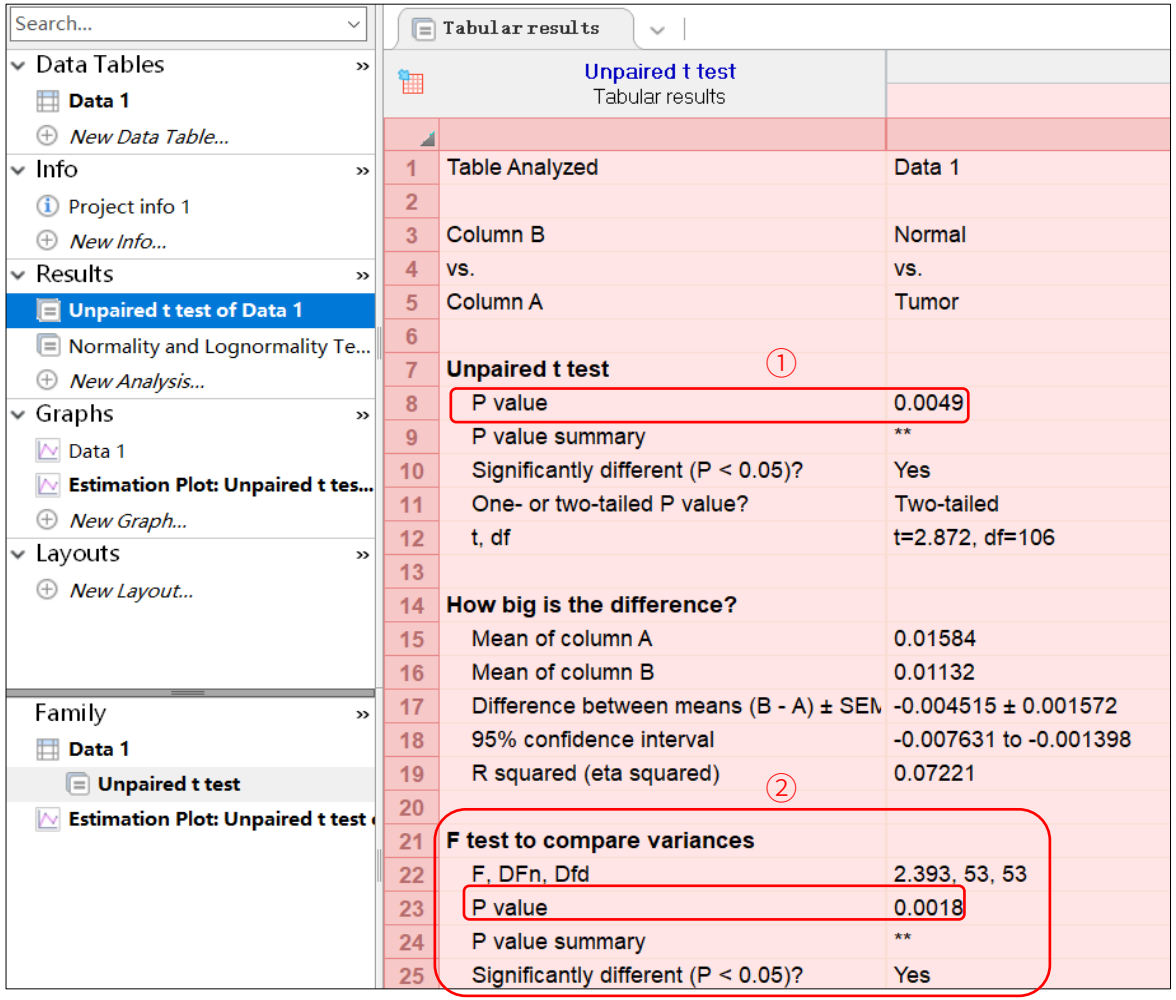
不知道数据是否方差齐性时，可先随意选择一个；分析结果里会显示F检验的结果，据此判断数据是否服从方差齐性，再来修改这个选项

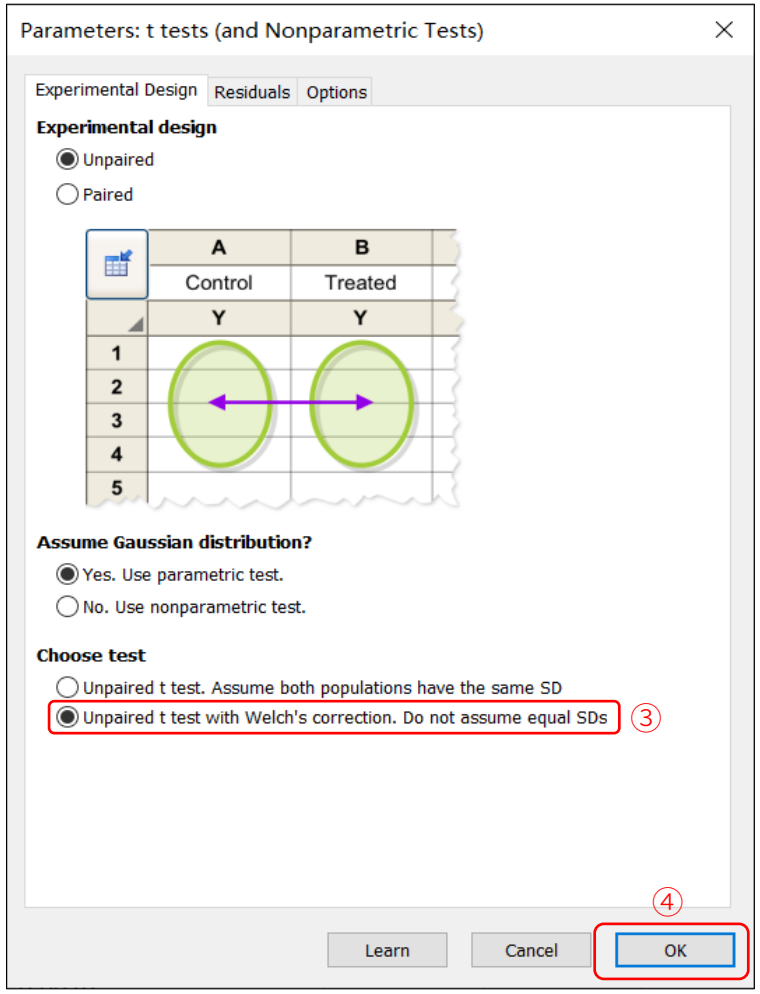
在大多情况下，选项中不需要再行修改

④点击OK

## 独立t检验

5 查看分析结果




①$p$ =0.0049，两组有显著性差异；

②F检验，一般要求$p$＞0.1（也有人认为$p$ ＞0.05），认为方差齐性；这里$p$ ＜0.1，可认为两组数据方差不齐，返回选择；

③选择Welch's correlation

④点击OK

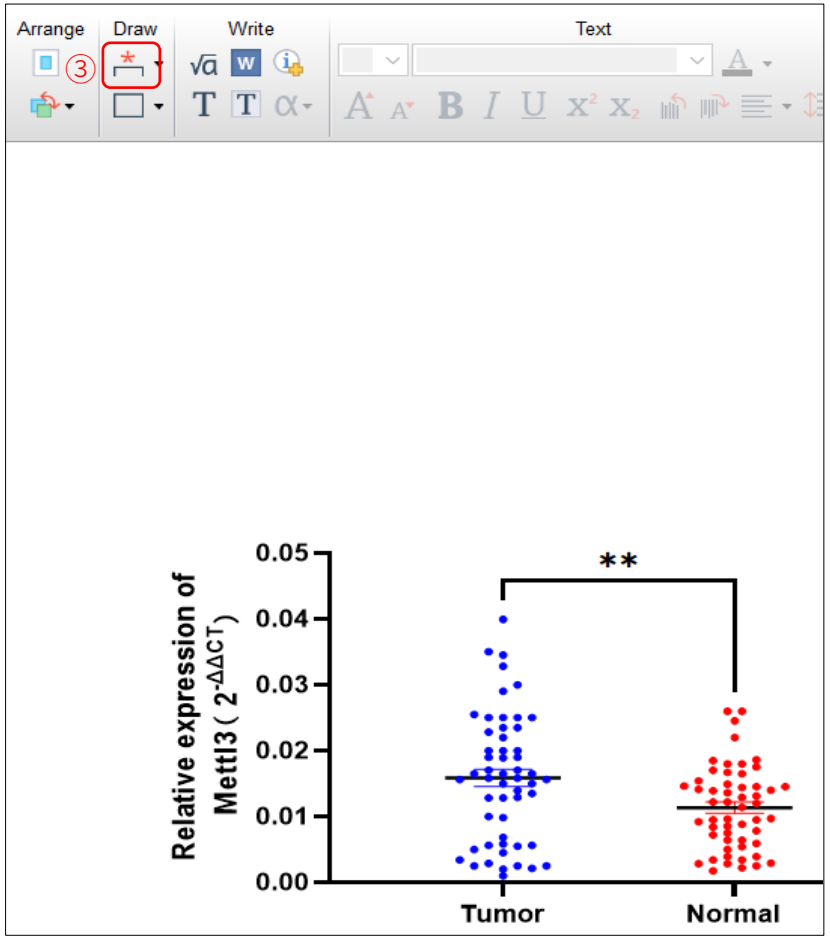
## 独立t检验

6 添加显著性差异

Welch's t test
Tabular results

• 矫正后的结果

| | | |
|---|---|---|
| 1 | Table Analyzed | Data 1 |
| 2 | | |
| 3 | Column B | Normal |
| 4 | vs. | vs. |
| 5 | Column A | Tumor |
| 6 | | |
| 7 | **Unpaired t test with Welch's correction** | |
| 8 | P value | 0.0051 ① |
| 9 | P value summary | ** |
| 10 | Significantly different (P < 0.05)? | Yes |
| 11 | One- or two-tailed P value? | Two-tailed |
| 12 | Welch-corrected t, df | t=2.872, df=90.71 ② |
| 13 | | |
| 14 | **How big is the difference?** | |
| 15 | Mean of column A | 0.01584 |
| 16 | Mean of column B | 0.01132 |
| 17 | Difference between means (B - A) ± SEM | -0.004515 ± 0.001572 |
| 18 | 95% confidence interval | -0.007637 to -0.001392 |
| 19 | R squared (eta squared) | 0.08337 |
| 20 | | |
| 21 | **F test to compare variances** | |
| 22 | F, DFn, Dfd | 2.393, 53, 53 |
| 23 | P value | 0.0018 |
| 24 | P value summary | ** |
| 25 | Significantly different (P < 0.05)? | Yes |
| 26 | | |




①$p$=0.0051，两组有显著性差异；

②矫正后；

③点击左侧工具栏Graph出现图像，点击此处，自动添加显著性差异符号（点击还可以对其进行格式修改。自动添加结果显示如图中所示）

## 配对t检验

- 对于配对设计的实验，就需要进行配对T检验
- 示例文献：Jafarimanesh H, Vakilian K, Mobasseri S. Thermo-therapy and cryotherapy to decrease the symptoms of restless leg syndrome during the pregnancy: A randomized clinical trial. Complement Ther Med. 2020 May;50:102409. doi: 10.1016/j.ctim.2020.102409. Epub 2020 Apr 19. PMID: 32444058. Table.2

- 示例解读：
  - ➢ 在37名孕妇，以同一人水疗前后的不宁腿指数作为两组，衡量治疗的效果

**Table 2**
Mean of RLS in the intervention groups.

| | Cold Mean ± SD | Warm Mean ± SD |
|---|---|---|
| RLS before | 18.65 ± 6.16 | 19.32 ± 5.87 |
| after | 11.02 ± 4.93 | 13.50 ± 4.74 |
| *P-valuc | 0.001 | 0.001 |
| *paired t-test | | |

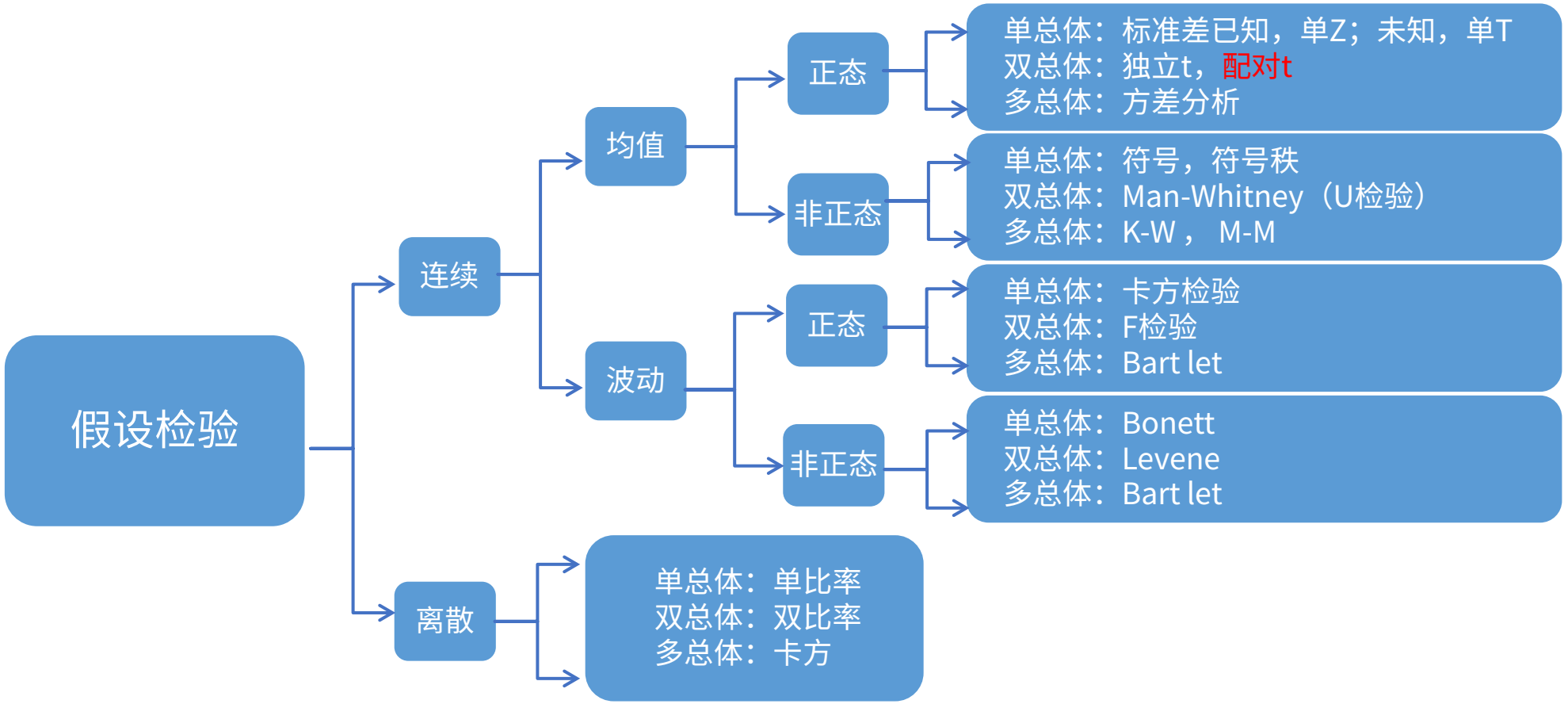
- 模拟数据

| | Group A Data Set-A | Group B Data Set-B |
|---|---|---|
| 1 | 31.7 | 14.5 |
| 2 | 16.9 | 15.2 |
| 3 | 18.8 | 15.6 |
| 4 | 20.0 | 12.9 |
| 5 | 27.1 | 15.7 |
| 6 | 17.1 | 9.4 |
| 7 | 21.9 | 7.1 |
| 8 | 31.3 | 2.4 |
| 9 | 20.0 | 3.8 |
| 10 | 30.8 | 16.3 |
| 11 | 16.9 | 16.6 |
| 12 | 19.8 | 15.8 |
| 13 | 12.3 | 7.5 |
| 14 | 21.4 | 11.2 |
| 15 | 17.9 | 12.7 |
| 16 | 4.6 | 11.7 |
| 17 | 18.7 | 14.0 |

## 配对t检验

- 分析目标

✓ 数据大体跟上一例差不多，不同的地方在于，本例中的两组数据，是配对关系，因此需要使用配对t检验。



## 配对t检验

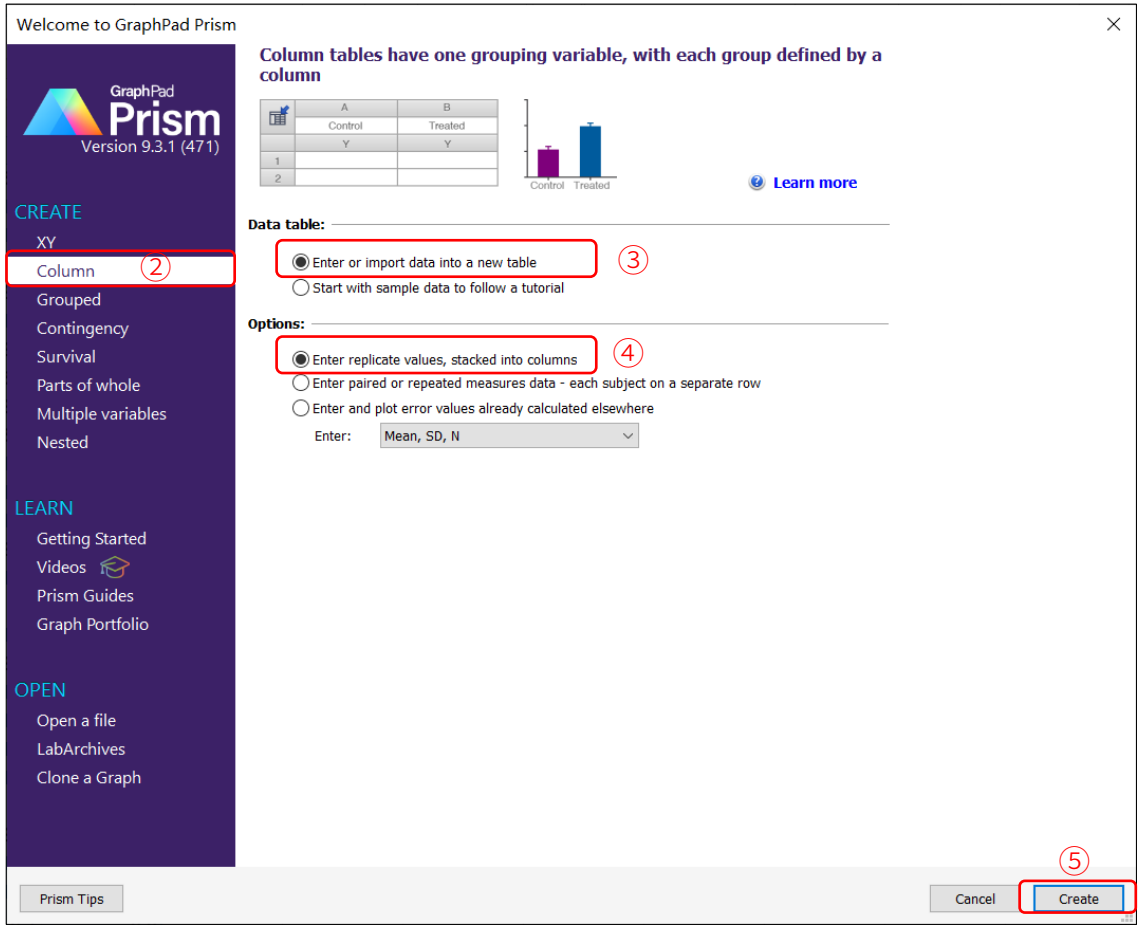
1 先在excel里计算两组数据的差值

| J | K | L |
|---|---|---|
| Before | After | After-before |
| 31.7 | 14.5 | -17.2 |
| 16.9 | 15.2 | -1.7 |
| 18.8 | 15.6 | -3.2 |
| 20.0 | 12.9 | -7.1 |
| 27.1 | 15.7 | -11.4 |
| 17.1 | 9.4 | -7.6 |
| 21.9 | 7.1 | -14.8 |
| 31.3 | 2.4 | -28.9 |
| 20.0 | 3.8 | -16.2 |
| 30.8 | 16.3 | -14.5 |
| 16.9 | 16.6 | -0.3 |
| 19.8 | 15.8 | -4.1 |
| 12.3 | 7.5 | -4.8 |
| 21.4 | 11.2 | -10.2 |
| 17.9 | 12.7 | -5.1 |
| 4.6 | 11.7 | 7.0 |
| 18.7 | 14.0 | -4.7 |
| 22.2 | 10.8 | -11.4 |
| 12.6 | 4.6 | -8.0 |
| 15.2 | 8.1 | -7.1 |

①
- =14.5-31.7
- =15.2-16.9

…

2 新建一个Column的数据表（重命名为“Paired t test”）




①在excel中利用公式计算After-before

②新建Column

③创建新table

④选择输入每个值

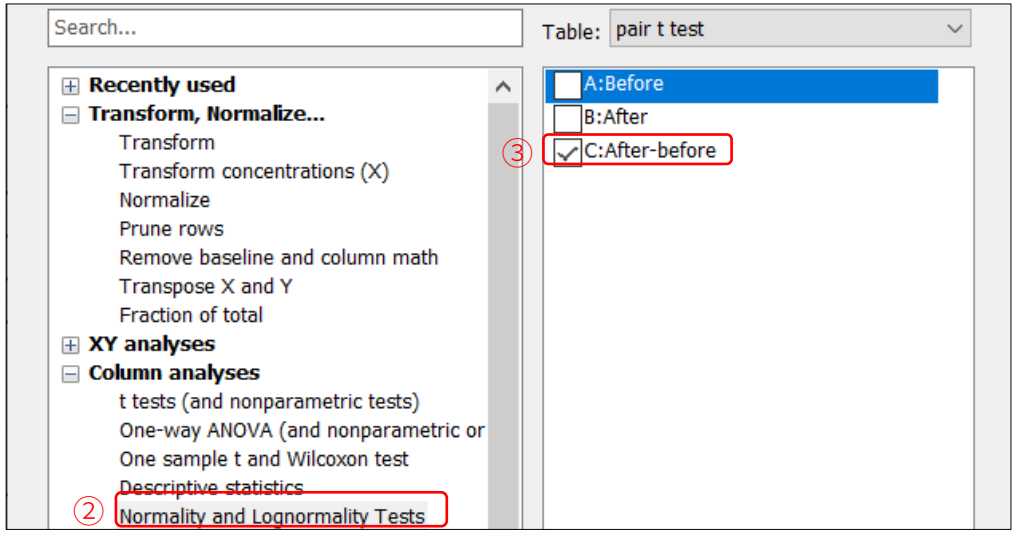
⑤点击OK

## 配对t检验

3 将数据按下图的格式输入

| ① | Group A | Group B | Group C |
|---|---|---|---|
| | Before | After | After-before |
| | Y | Y | Y |
| 1 | 31.7 | 14.5 | -17.2 |
| 2 | 16.9 | 15.2 | -1.7 |
| 3 | 18.8 | 15.6 | -3.2 |
| 4 | 20.0 | 12.9 | -7.1 |
| 5 | 27.1 | 15.7 | -11.4 |
| 6 | 17.1 | 9.4 | -7.6 |
| 7 | 21.9 | 7.1 | -14.8 |
| 8 | 31.3 | 2.4 | -28.9 |
| 9 | 20.0 | 3.8 | -16.2 |
| 10 | 30.8 | 16.3 | -14.5 |
| 11 | 16.9 | 16.6 | -0.3 |
| 12 | 19.8 | 15.8 | -4.1 |
| 13 | 12.3 | 7.5 | -4.8 |
| 14 | 21.4 | 11.2 | -10.2 |
| 15 | 17.9 | 12.7 | -5.1 |
| 16 | 4.6 | 11.7 | 7.0 |
| 17 | 18.7 | 14.0 | -4.7 |

4 按前述步骤，对两组的差值进行正态分布检验




①excel拷贝数据

②点击分析后，选择正态检测

③只对差值进行正态检测

④四种检测方法显示符合正态分布

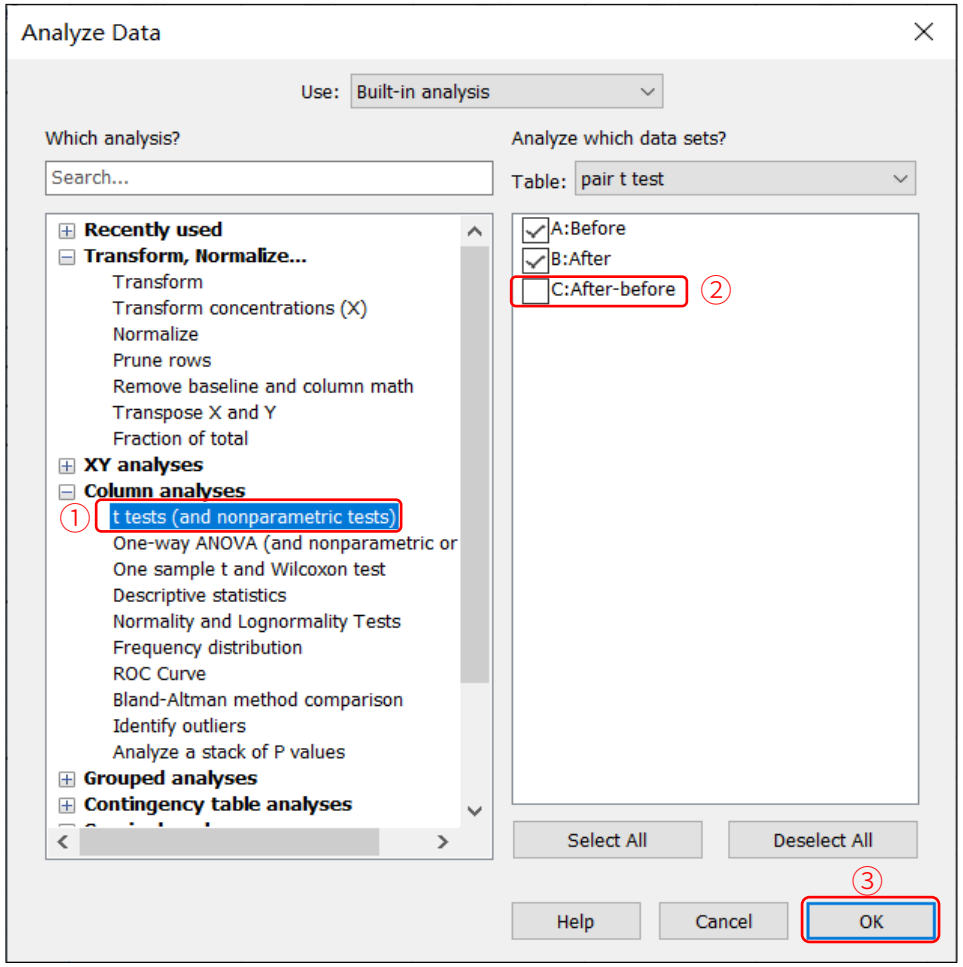
四种检测方法结果显示*P*值均大于0.1

| ④ | Test for normal distribution | |
|---|---|---|
| 2 | **Anderson-Darling test** | |
| 3 | A2* | 0.2536 |
| 4 | P value | 0.7142 |
| 5 | Passed normality test (alpha=0.05)? | Yes |
| 6 | P value summary | ns |
| 7 | | |
| 8 | **D'Agostino & Pearson test** | |
| 9 | K2 | 1.366 |
| 10 | P value | 0.5052 |
| 11 | Passed normality test (alpha=0.05)? | Yes |
| 12 | P value summary | ns |

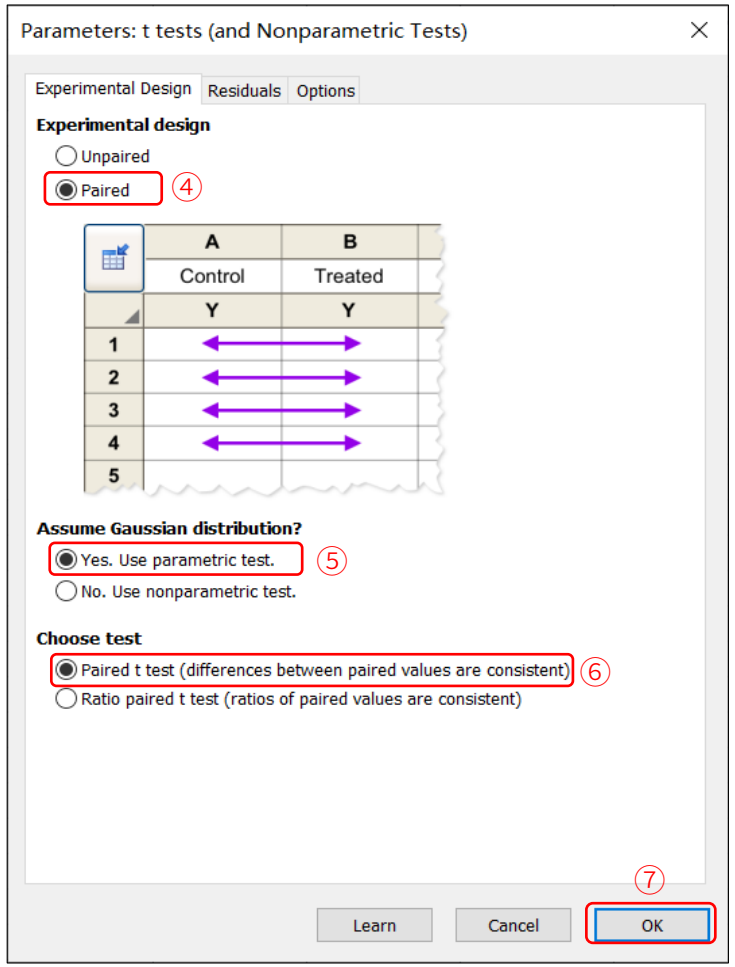
| 14 | **Shapiro-Wilk test** | |
|---|---|---|
| 15 | W | 0.9775 |
| 16 | P value | 0.6445 |
| 17 | Passed normality test (alpha=0.05)? | Yes |
| 18 | P value summary | ns |
| 19 | | |
| 20 | **Kolmogorov-Smirnov test** | |
| 21 | KS distance | 0.07687 |
| 22 | P value | >0.1000 |
| 23 | Passed normality test (alpha=0.05)? | Yes |
| 24 | P value summary | ns |

## 配对t检验

5 设置检验方法




6 设置检验参数




①选择t检验

②勾选差值组

③点击OK

④勾选配对设计

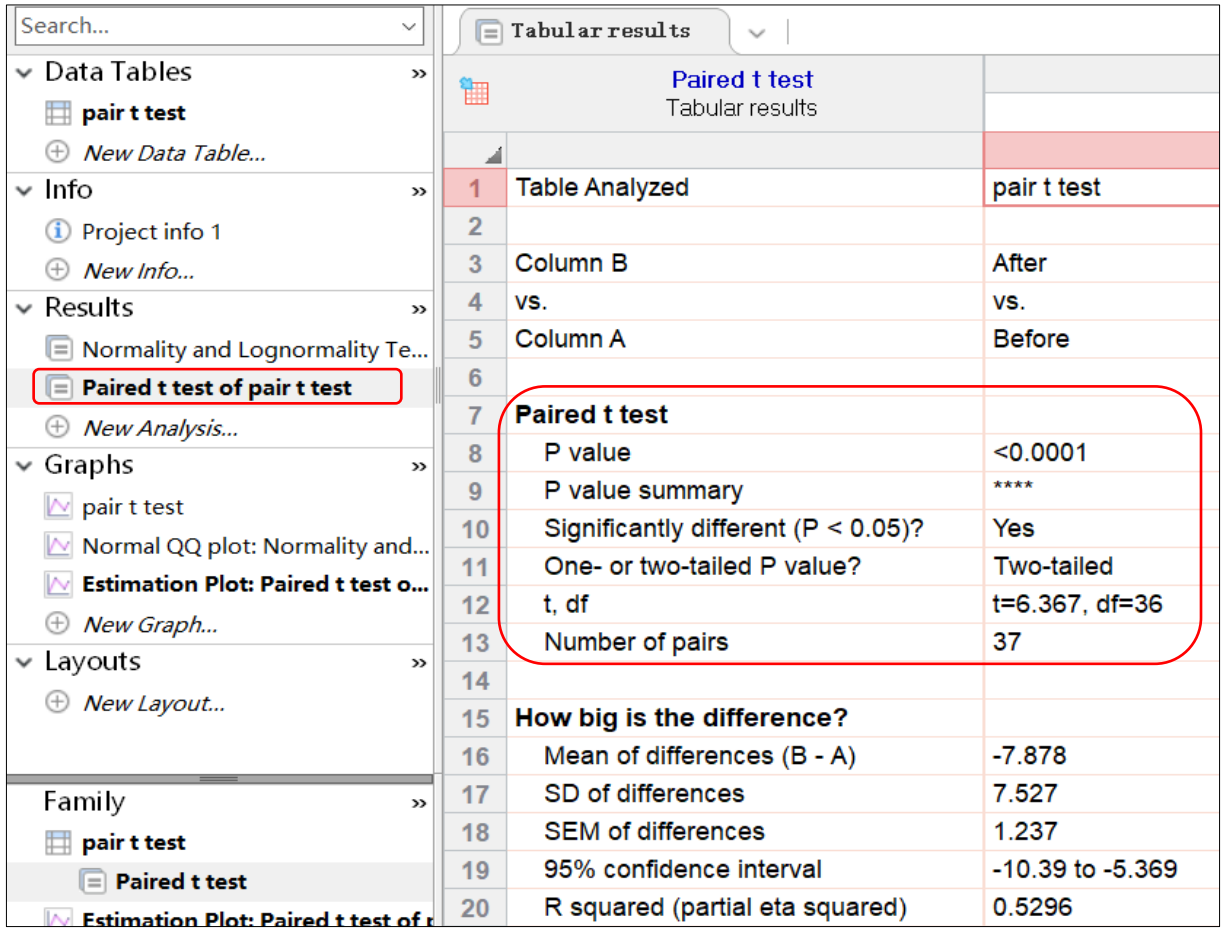
⑤服从正态分布

⑥选择配对t检验

⑦点击OK

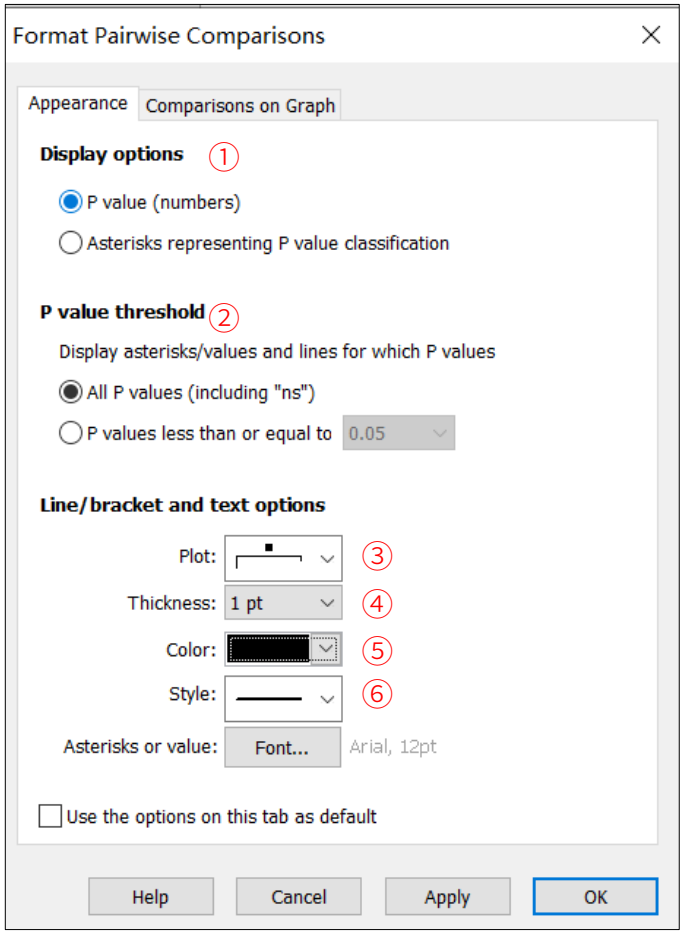
## 配对t检验

### 7 查看分析结果




| | | |
|---|---|---|
| 1 | Table Analyzed | pair t test |
| 2 | | |
| 3 | Column B | After |
| 4 | vs. | vs. |
| 5 | Column A | Before |
| 6 | | |
| 7 | **Paired t test** | |
| 8 | P value | <0.0001 |
| 9 | P value summary | **** |
| 10 | Significantly different (P < 0.05)? | Yes |
| 11 | One- or two-tailed P value? | Two-tailed |
| 12 | t, df | t=6.367, df=36 |
| 13 | Number of pairs | 37 |
| 14 | | |
| 15 | **How big is the difference?** | |
| 16 | Mean of differences (B - A) | -7.878 |
| 17 | SD of differences | 7.527 |
| 18 | SEM of differences | 1.237 |
| 19 | 95% confidence interval | -10.39 to -5.369 |
| 20 | R squared (partial eta squared) | 0.5296 |

### 8 显著性差异标注




显著性差异的做法就不再赘述

单击自动标注可对标注形式进行修改

①展示方式

②*P*值界限

③点的形式

④线厚度

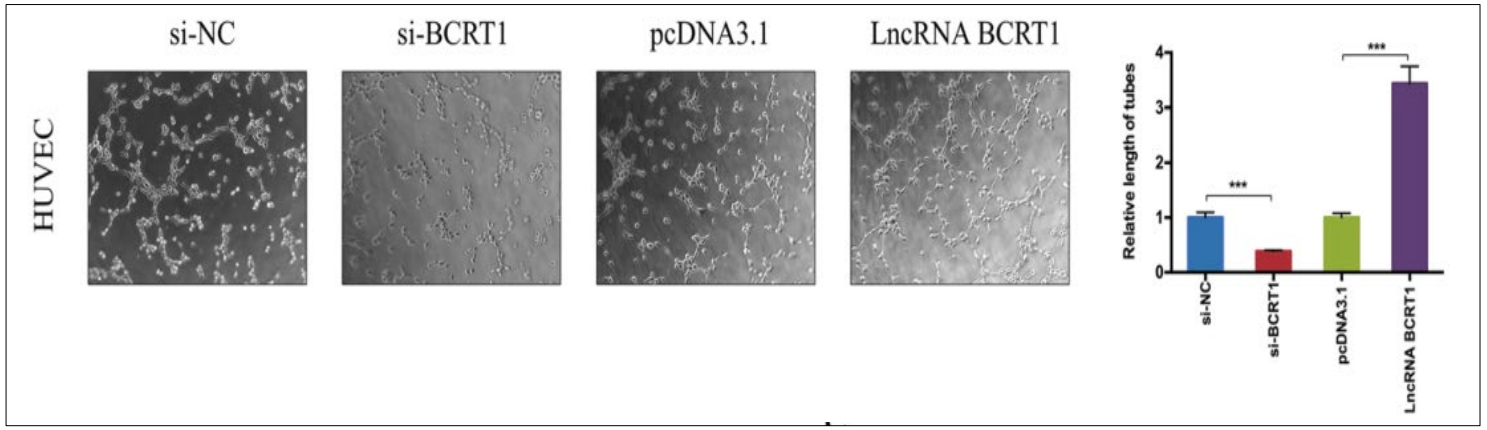
⑤颜色

⑥风格

## 方差分析

- 示例：文献Liang Y, Song X, Li Y, Chen B, Zhao W, Wang L, Zhang H, Liu Y, Han D, Zhang N, Ma T, Wang Y, Ye F, Luo D, Li X, Yang Q. LncRNA BCRT1 promotes breast cancer progression by targeting miR-1303/PTBP3 axis. Mol Cancer. 2020 May 8;19(1):85. doi: 10.1186/s12943-020-01206-5. PMID: 32384893; PMCID: PMC7206728. Fig.3C

- 示例解读：
  - 该图是一个小管形成实验，一般是在内皮细胞做实验，论证血管新生能力的大小
  - 图中X轴仅以不同处理条件作为分类依据（si-NC，si-BCRT1，pcDNA3.1，lncRNA BCRT1），每种分类下，有几次实验结果




- 模拟数据

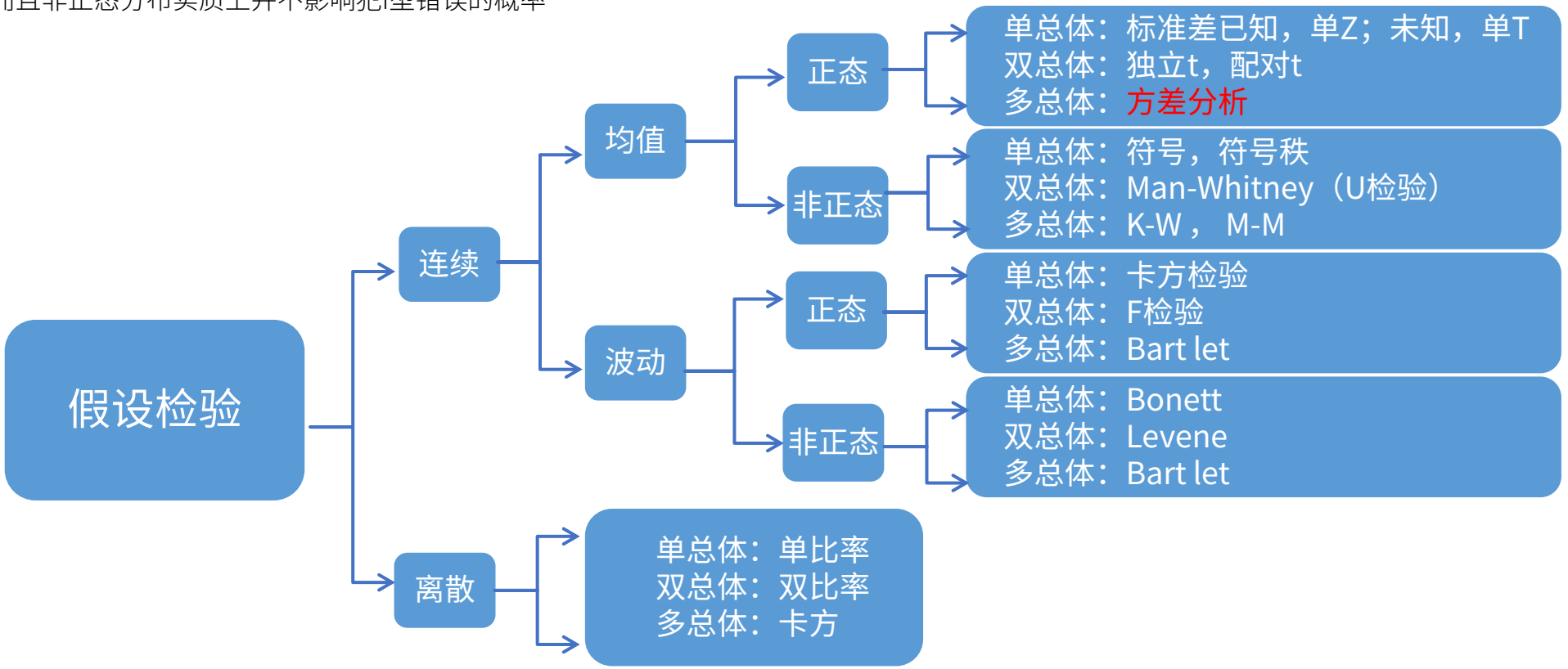
| si-NC | si-BCRT1 | pcDNA3.1 | lncRNA BCRT1 |
|---|---|---|---|
| 1.01 | 0.44 | 1.11 | 3.58 |
| 1.11 | 0.45 | 1.02 | 3.16 |
| 0.95 | 0.35 | 0.95 | 3.94 |

## 方差分析

- 分析目标

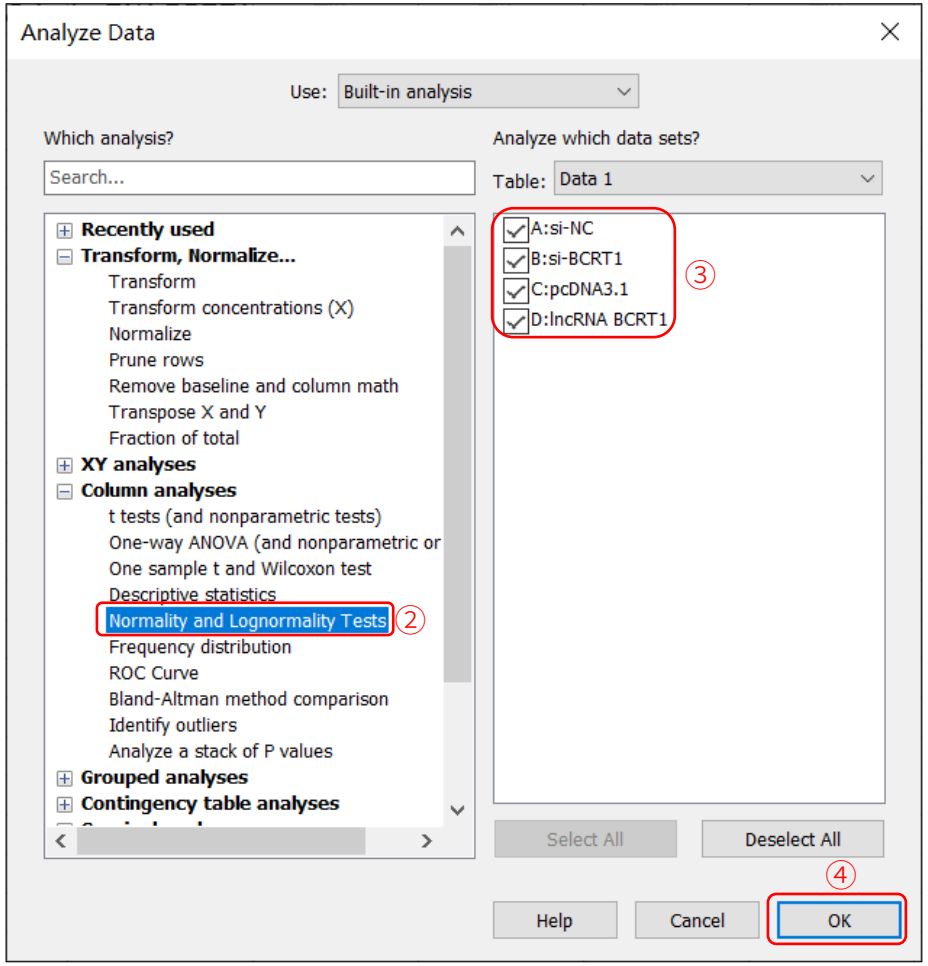
✓ 这是一个连续性的数据，以均值的形式表现，每组数据样本量少，经正态检测后是正态分布，符合方差分析的选择；

✓ GraphPad Prism中， 各组数据服从正态分布是使用One-way ANOVA（单因素方差分析）的前提，若数据不服从正态分布，有几个选择：①通过转换使得数据服从正态分布，如log、倒数等；②使用非参数检验；③ 直接进行分析：由于单因素方差分析对于偏离正态分布比较稳健，尤其是在各组样本量相等或近似相等的情况下，而且非正态分布实质上并不影响犯I型错误的概率

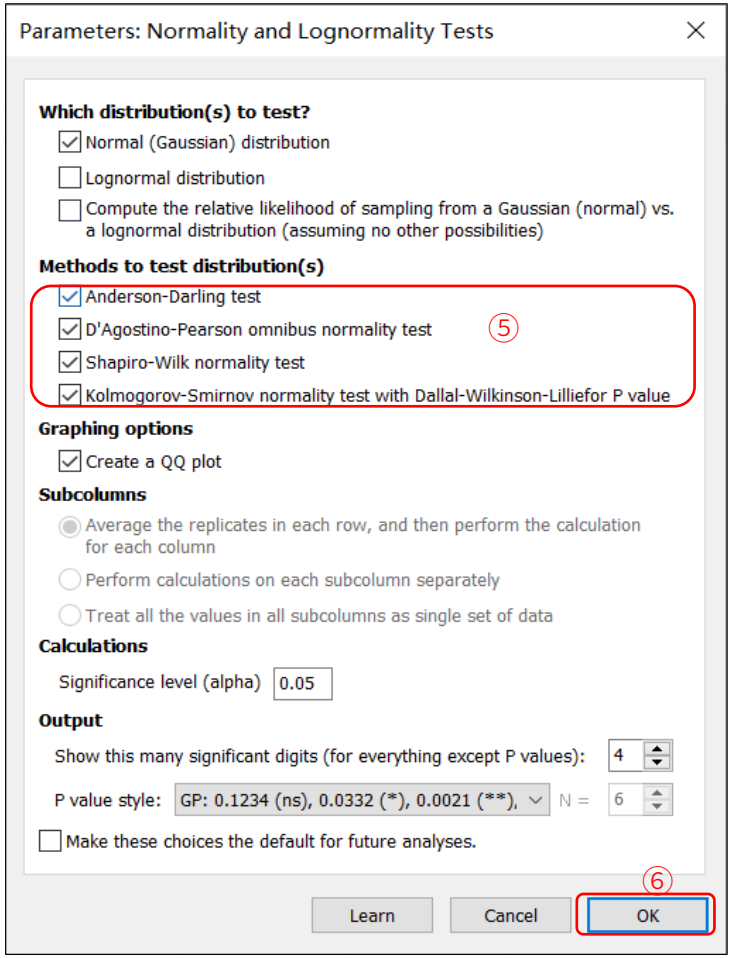


## 方差分析

1 点击Analyze按钮，调出数据分析窗口；分析方法选择Column analyses→Normality and Lognormality Tests勾选所有组别；点击“OK”




2 在弹出的窗口中，勾选正态分布检验的方法，点击“OK”




①excel拷贝数据

②点击分析后，选择正态检测

③勾选所有组别

④选择OK

⑤勾选检测方法

⑥选择OK

## 方差分析

3 查看结果

- 正态分布检验时，一般认为，

当$p$ value > 0.1 时，数据服从正态分布

- 由于样本量过小，只有S-W检测能够有效
- 四组数据均符合正态分布

| | Normality and Lognormality Tests<br>Tabular results | A<br>si-NC | B<br>si-BCRT1 | C<br>pcDNA3.1 | D<br>lncRNA BCRT1 |
|---|---|---|---|---|---|
| 1 | **Test for normal distribution** | | | | |
| 2 | **Anderson-Darling test** | | | | |
| 3 | A2* | N too small | N too small | N too small | N too small |
| 4 | P value | | | | |
| 5 | Passed normality test (alpha=0.05)? | | | | |
| 6 | P value summary | | | | |
| 7 | | | | | |
| 8 | **D'Agostino & Pearson test** | | | | |
| 9 | K2 | N too small | N too small | N too small | N too small |
| 10 | P value | | | | |
| 11 | Passed normality test (alpha=0.05)? | | | | |
| 12 | P value summary | | | | |
| 13 | | | | | |
| 14 | **Shapiro-Wilk test** | | | | |
| 15 | W | 0.9796 | 0.8242 | 0.9948 | 0.9980 |
| 16 | P value | 0.7262 | 0.1736 | 0.8624 | 0.9152 |
| 17 | Passed normality test (alpha=0.05)? | Yes | Yes | Yes | Yes |
| 18 | P value summary | ns | ns | ns | ns |
| 19 | | | | | |
| 20 | **Kolmogorov-Smirnov test** | | | | |
| 21 | KS distance | N too small | N too small | N too small | N too small |
| 22 | P value | | | | |
| 23 | Passed normality test (alpha=0.05)? | | | | |
| 24 | P value summary | | | | |
| 25 | | | | | |
| 26 | **Number of values** | 3 | 3 | 3 | 3 |

- $P$值均大于0.1，服从正态分布

## 方差分析

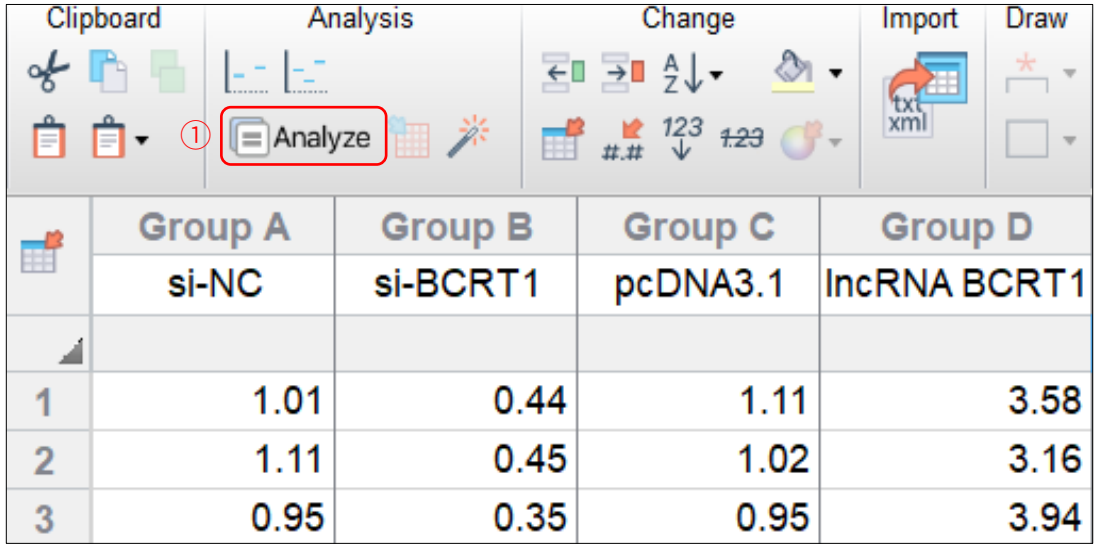
### 4 分析数据

①点击Analyze

②在Column analyses下找到One-way ANOVA

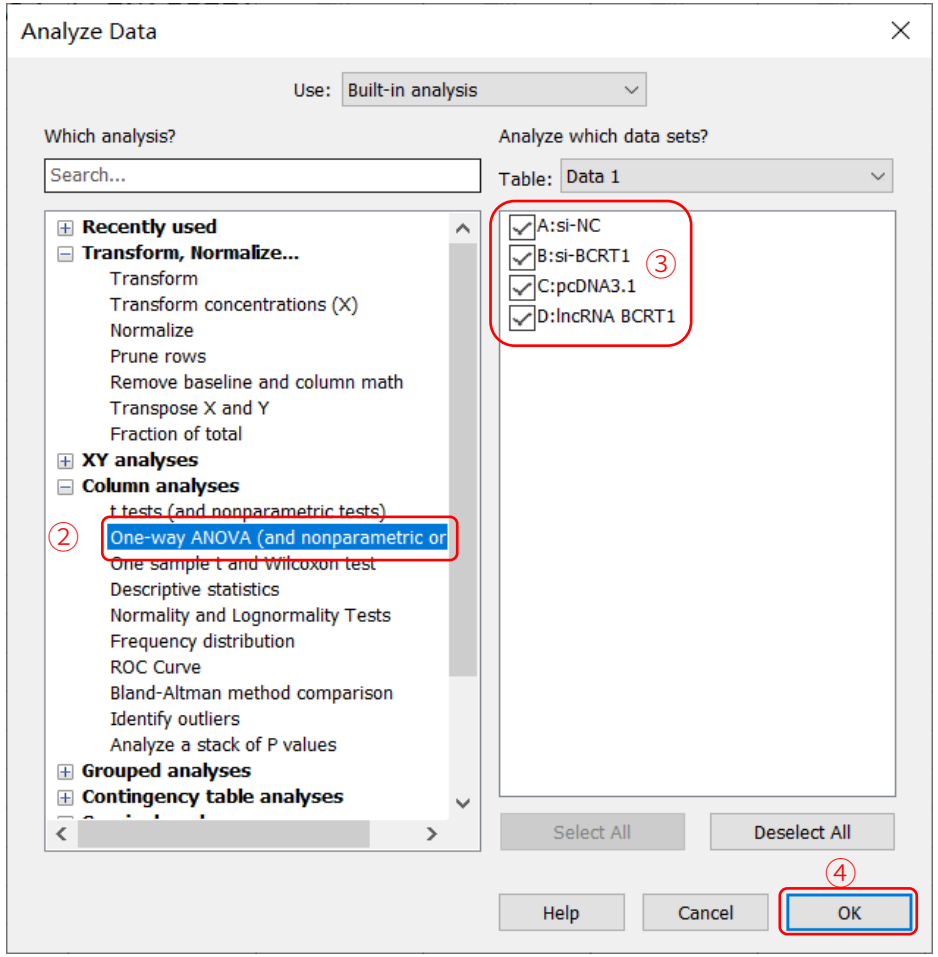
③默认已勾选了所有组别

④点击OK



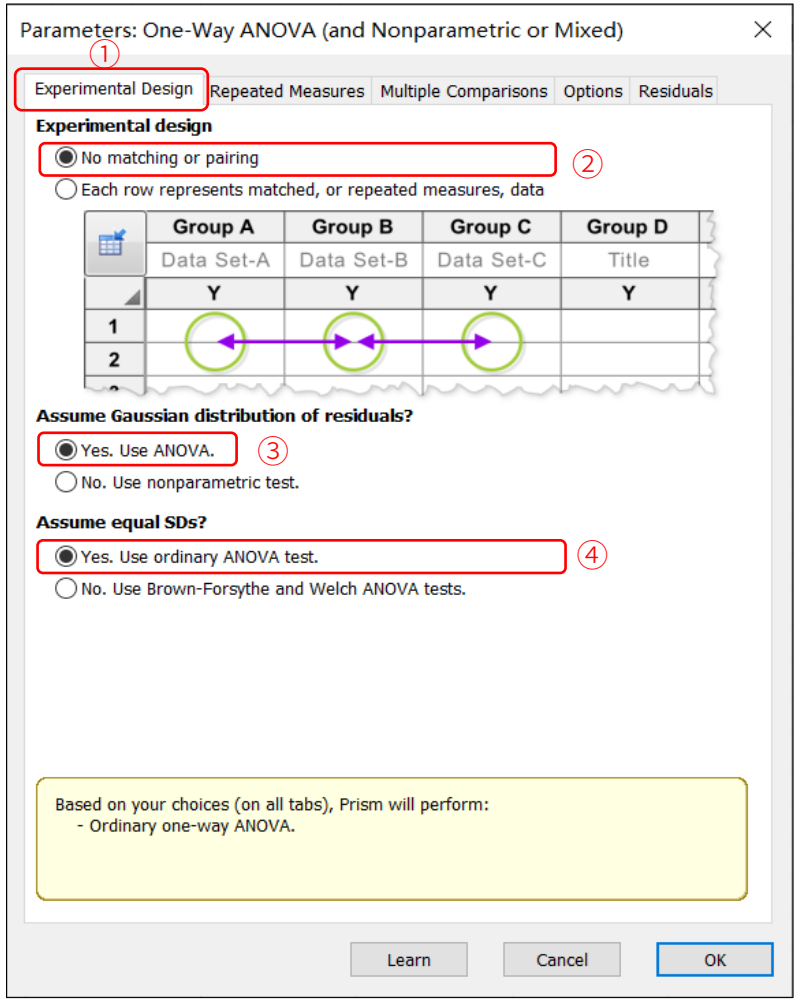

| | Group A | Group B | Group C | Group D |
|---|---|---|---|---|
| | si-NC | si-BCRT1 | pcDNA3.1 | lncRNA BCRT1 |
| 1 | 1.01 | 0.44 | 1.11 | 3.58 |
| 2 | 1.11 | 0.45 | 1.02 | 3.16 |
| 3 | 0.95 | 0.35 | 0.95 | 3.94 |

### 5 选择分析方法

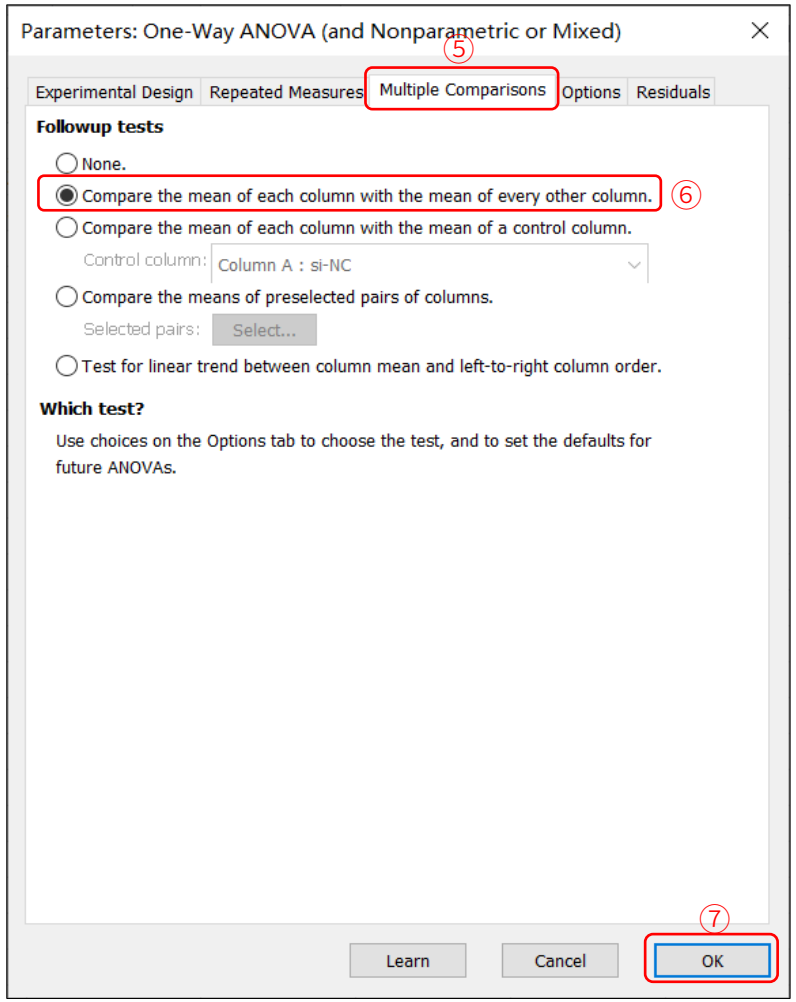


## 方差分析

### 6 Experimental Design




### 7 Multiple Comparisons




①点击 Experimental Design

②选择非重复或非配对测量

③服从正态分布，选择ANOVA

④方差齐，选择经典ANOVA

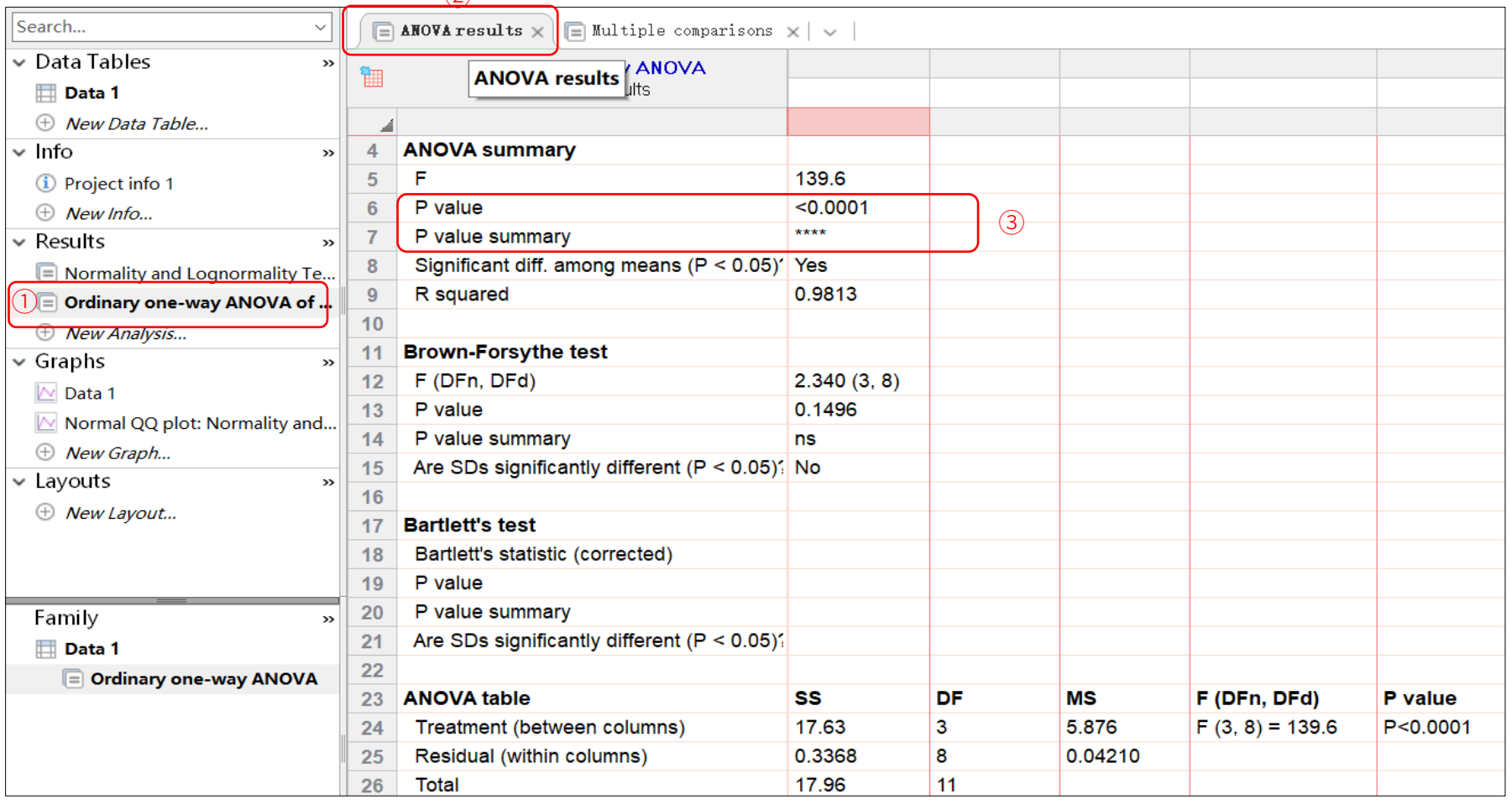
⑤选择Multiple Comparisons

⑥选择两两比较

⑦选择OK

## 方差分析

### 8 查看结果




| | | | | | | |
|---|---|---|---|---|---|---|
| 4 | **ANOVA summary** | | | | | |
| 5 | F | 139.6 | | | | |
| 6 | P value | <0.0001 | | | | |
| 7 | P value summary | **** | | | | |
| 8 | Significant diff. among means (P < 0.05)? | Yes | | | | |
| 9 | R squared | 0.9813 | | | | |
| 10 | | | | | | |
| 11 | **Brown-Forsythe test** | | | | | |
| 12 | F (DFn, DFd) | 2.340 (3, 8) | | | | |
| 13 | P value | 0.1496 | | | | |
| 14 | P value summary | ns | | | | |
| 15 | Are SDs significantly different (P < 0.05)? | No | | | | |
| 16 | | | | | | |
| 17 | **Bartlett's test** | | | | | |
| 18 | Bartlett's statistic (corrected) | | | | | |
| 19 | P value | | | | | |
| 20 | P value summary | | | | | |
| 21 | Are SDs significantly different (P < 0.05)? | | | | | |
| 22 | | | | | | |
| 23 | **ANOVA table** | **SS** | **DF** | **MS** | **F (DFn, DFd)** | **P value** |
| 24 | Treatment (between columns) | 17.63 | 3 | 5.876 | F (3, 8) = 139.6 | P<0.0001 |
| 25 | Residual (within columns) | 0.3368 | 8 | 0.04210 | | |
| 26 | Total | 17.96 | 11 | | | |

① 在左边导航栏里找到“Ordinary one-way …”

② 点击其目录下的“ANOVA”查看结果

③ 统计结果：*p* 值的意义和*p* 值。这里*p* 值<0.05表示：多组数据中，至少有一对（任意两组）或者多对具有显著性差异，至于具体是哪个组跟哪个组之间有差异，则需要查看多重比较的结果

## 方差分析

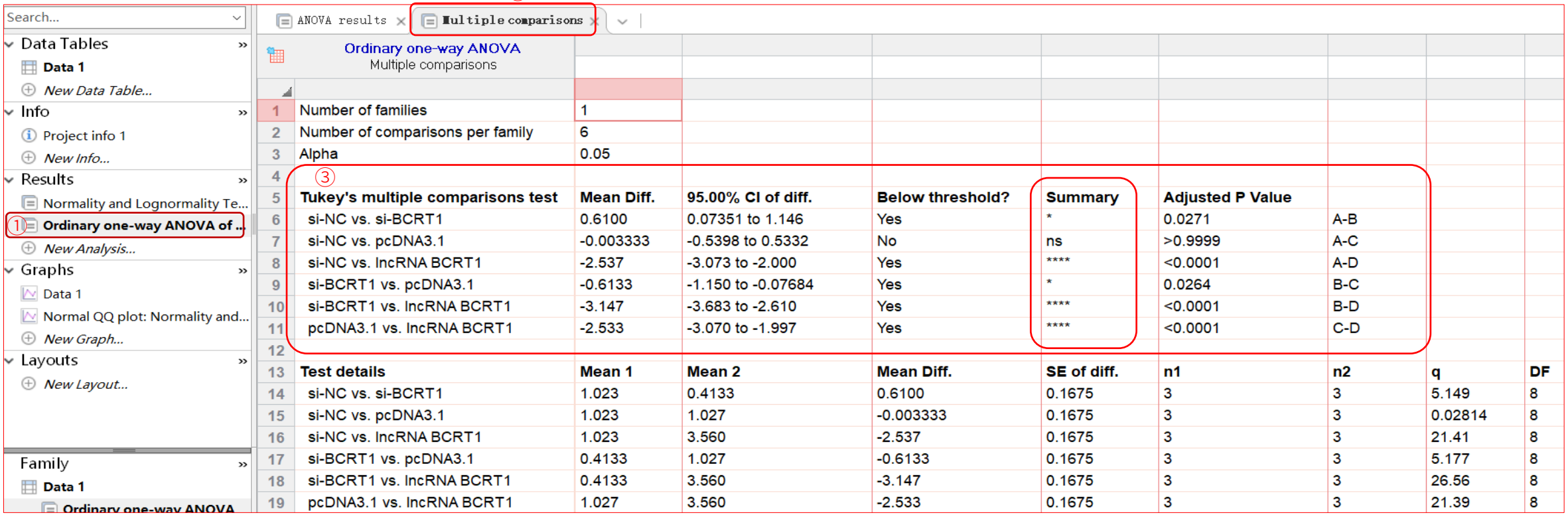
### 8 查看结果

① 在左边导航栏里找到“Ordinary one-way ANOVA”

② 点击其目录下的“Multiple comparisons”查看结果（“ANOVA“表格里是总体统计结果）

③ 统计结果：*p*值的意义和*p*值。第6-11行分别显示各组两两比较的结果

②

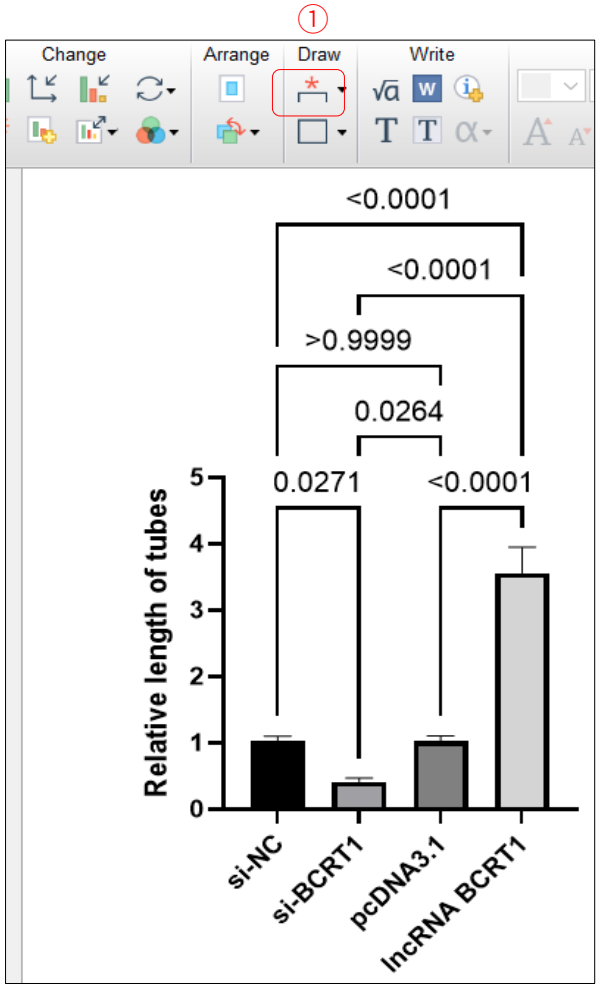


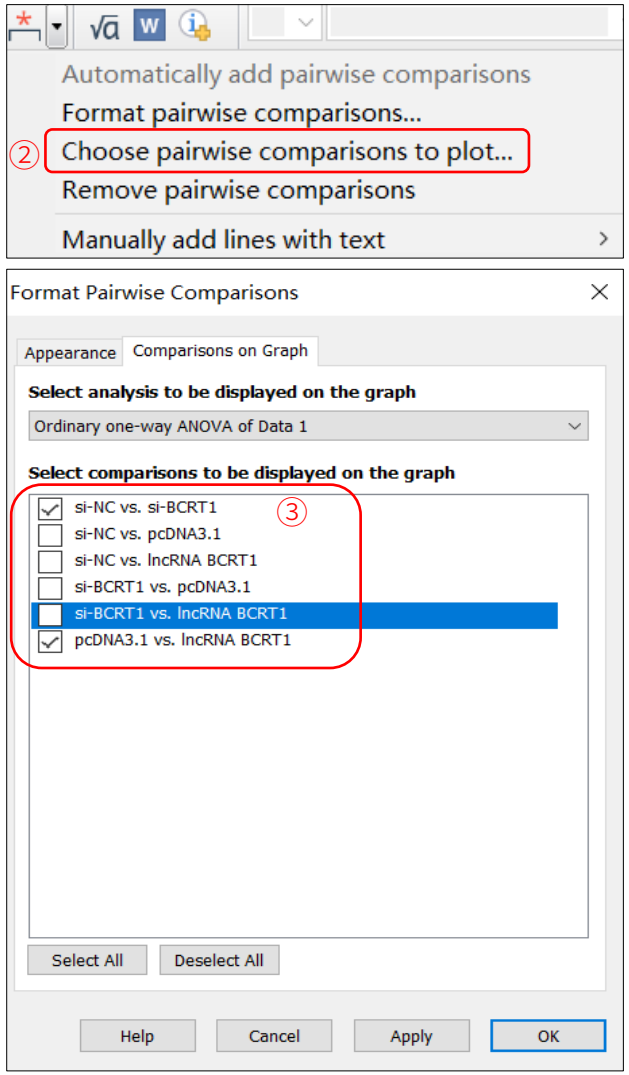

Ordinary one-way ANOVA
Multiple comparisons

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Number of families | 1 | | | | | | | |
| 2 | Number of comparisons per family | 6 | | | | | | | |
| 3 | Alpha | 0.05 | | | | | | | |
| 4 | ③ | | | | | | | | |
| 5 | **Tukey's multiple comparisons test** | **Mean Diff.** | **95.00% CI of diff.** | **Below threshold?** | **Summary** | **Adjusted P Value** | | | |
| 6 | si-NC vs. si-BCRT1 | 0.6100 | 0.07351 to 1.146 | Yes | * | 0.0271 | A-B | | |
| 7 | si-NC vs. pcDNA3.1 | -0.003333 | -0.5398 to 0.5332 | No | ns | >0.9999 | A-C | | |
| 8 | si-NC vs. lncRNA BCRT1 | -2.537 | -3.073 to -2.000 | Yes | **** | <0.0001 | A-D | | |
| 9 | si-BCRT1 vs. pcDNA3.1 | -0.6133 | -1.150 to -0.07684 | Yes | * | 0.0264 | B-C | | |
| 10 | si-BCRT1 vs. lncRNA BCRT1 | -3.147 | -3.683 to -2.610 | Yes | **** | <0.0001 | B-D | | |
| 11 | pcDNA3.1 vs. lncRNA BCRT1 | -2.533 | -3.070 to -1.997 | Yes | **** | <0.0001 | C-D | | |
| 12 | | | | | | | | | |
| 13 | **Test details** | **Mean 1** | **Mean 2** | **Mean Diff.** | **SE of diff.** | **n1** | **n2** | **q** | **DF** |
| 14 | si-NC vs. si-BCRT1 | 1.023 | 0.4133 | 0.6100 | 0.1675 | 3 | 3 | 5.149 | 8 |
| 15 | si-NC vs. pcDNA3.1 | 1.023 | 1.027 | -0.003333 | 0.1675 | 3 | 3 | 0.02814 | 8 |
| 16 | si-NC vs. lncRNA BCRT1 | 1.023 | 3.560 | -2.537 | 0.1675 | 3 | 3 | 21.41 | 8 |
| 17 | si-BCRT1 vs. pcDNA3.1 | 0.4133 | 1.027 | -0.6133 | 0.1675 | 3 | 3 | 5.177 | 8 |
| 18 | si-BCRT1 vs. lncRNA BCRT1 | 0.4133 | 3.560 | -3.147 | 0.1675 | 3 | 3 | 26.56 | 8 |
| 19 | pcDNA3.1 vs. lncRNA BCRT1 | 1.027 | 3.560 | -2.533 | 0.1675 | 3 | 3 | 21.39 | 8 |

## 方差分析

### 9 标记结果



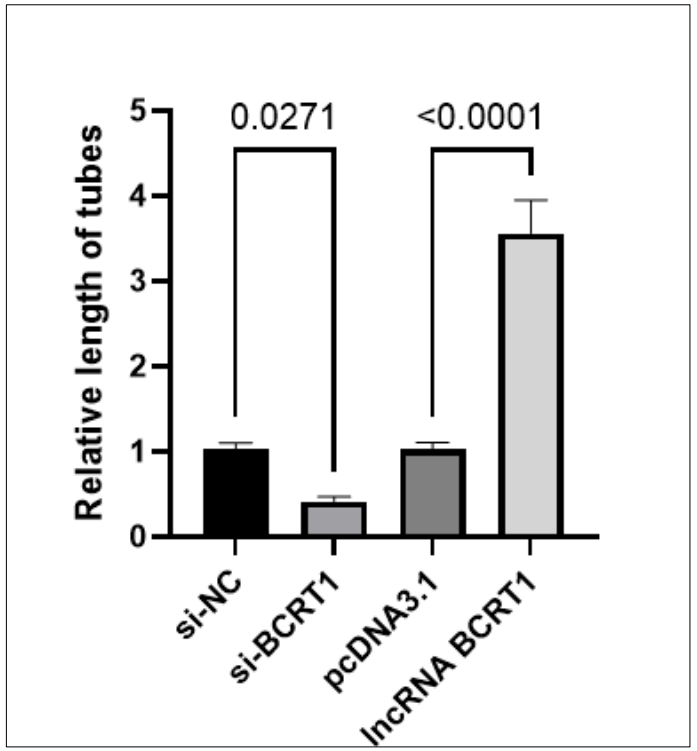




① 点击可自动出现所有的comparisons，如左侧图

② 点击三角形下拉选项，选择第三选项

③ 将不需要的选项勾去

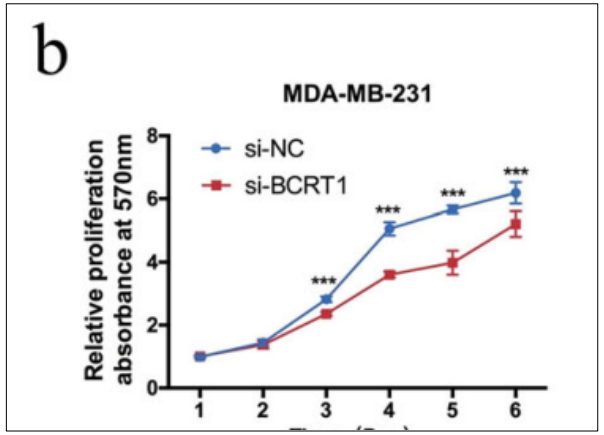
④ 结果如下图



## 方差分析

- 示例：文献Liang Y, Song X, Li Y, Chen B, Zhao W, Wang L, Zhang H, Liu Y, Han D, Zhang N, Ma T, Wang Y, Ye F, Luo D, Li X, Yang Q. LncRNA BCRT1 promotes breast cancer progression by targeting miR-1303/PTBP3 axis. Mol Cancer. 2020 May 8;19(1):85. doi: 10.1186/s12943-020-01206-5. PMID: 32384893; PMCID: PMC7206728. Fig.2b

- 示例解读：
  - 2个组别，分别是 si-NC, si-BCRT1；每个组别检测了0、1、2、3、4、5、6天的数据，每个数据有3个重复

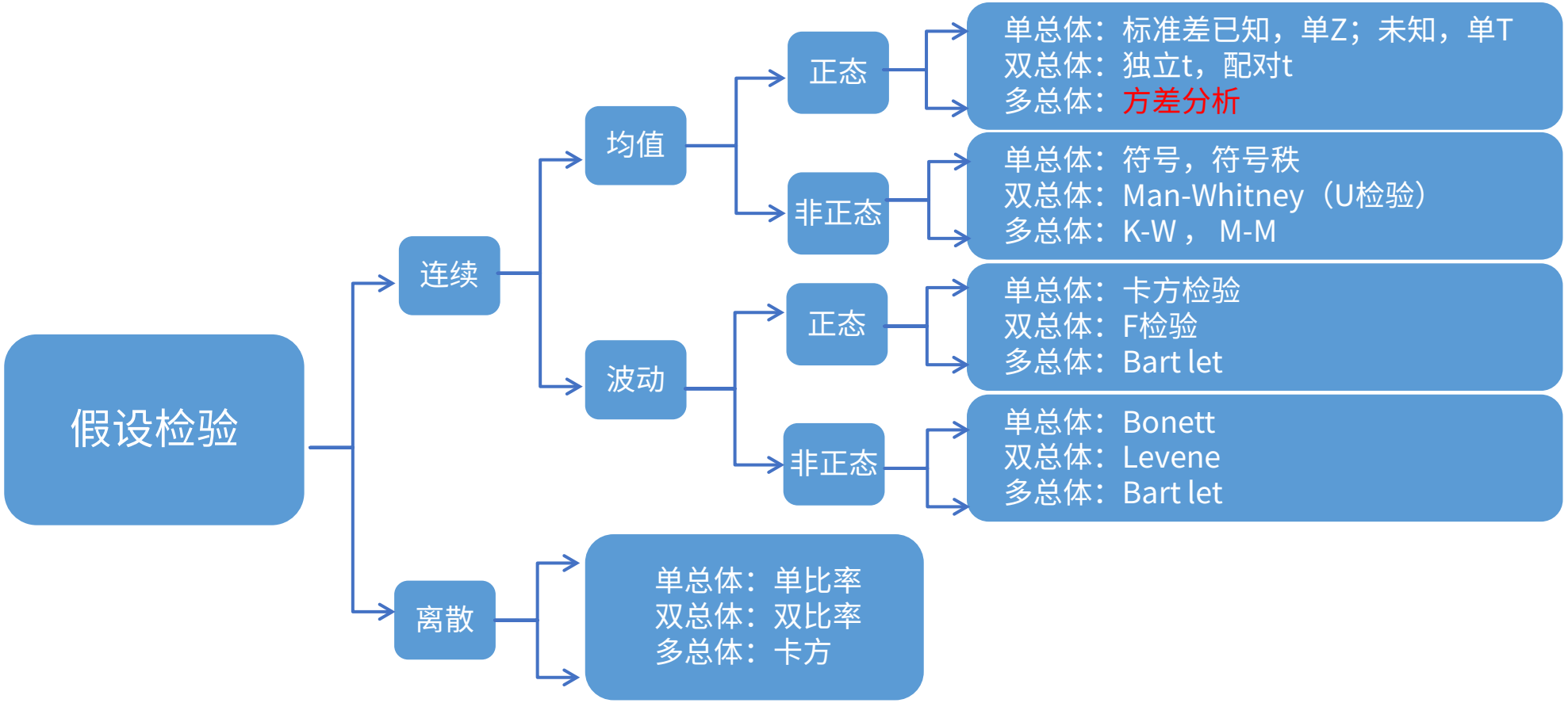



- 模拟数据

| Table format: XY | X | Group A | | | Group B | | |
|---|---|---|---|---|---|---|---|
| | day | si-NC | | | si-BCRT1 | | |
| | X | A:Y1 | A:Y2 | A:Y3 | B:Y1 | B:Y2 | B:Y3 |
| 1 | Title | 1.0 | 1.0000 | 1.0000 | 1.0000 | 1.0 | 1.0 | 1.0 |
| 2 | Title | 2.0 | 1.2000 | 1.6000 | 1.8000 | 2.0 | 2.3 | 2.4 |
| 3 | Title | 3.0 | 2.5000 | 2.6000 | 2.7000 | 4.3 | 4.4 | 4.6 |
| 4 | Title | 4.0 | 3.9000 | 3.7000 | 3.5000 | 5.1 | 5.2 | 5.4 |
| 5 | Title | 5.0 | 4.0000 | 4.1000 | 4.5000 | 5.4 | 5.7 | 5.9 |
| 6 | Title | 6.0 | 5.5000 | 5.6000 | 5.9000 | 6.2 | 6.3 | 6.5 |

## 方差分析

- 分析目标
  - ✓ 数据结构上，这是个二维分组，处理分组与时间点
  - ✓ 从实验结果上看，需要比较实验组和对照组相同时间点上的数据是否有差异，细胞间不需要比较，但需要比较的时间点有多个，双因素，用Two-way ANOVA



## 方差分析

1 选择分析方法

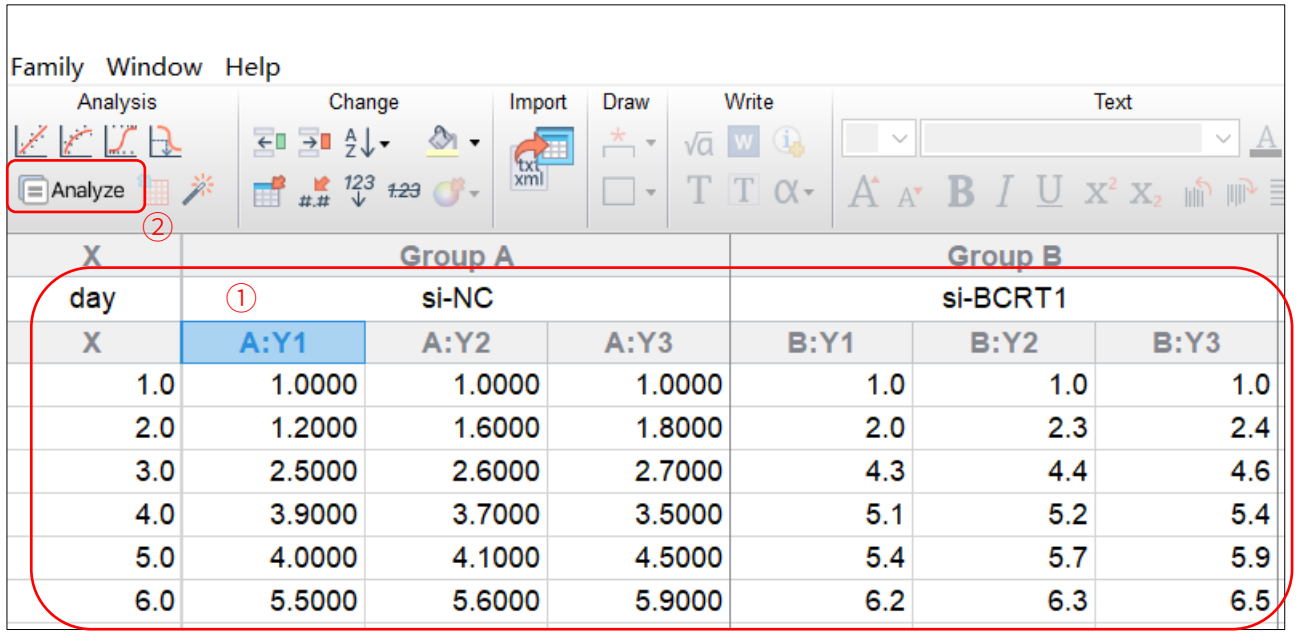
①excel拷贝数据

②点击分析后

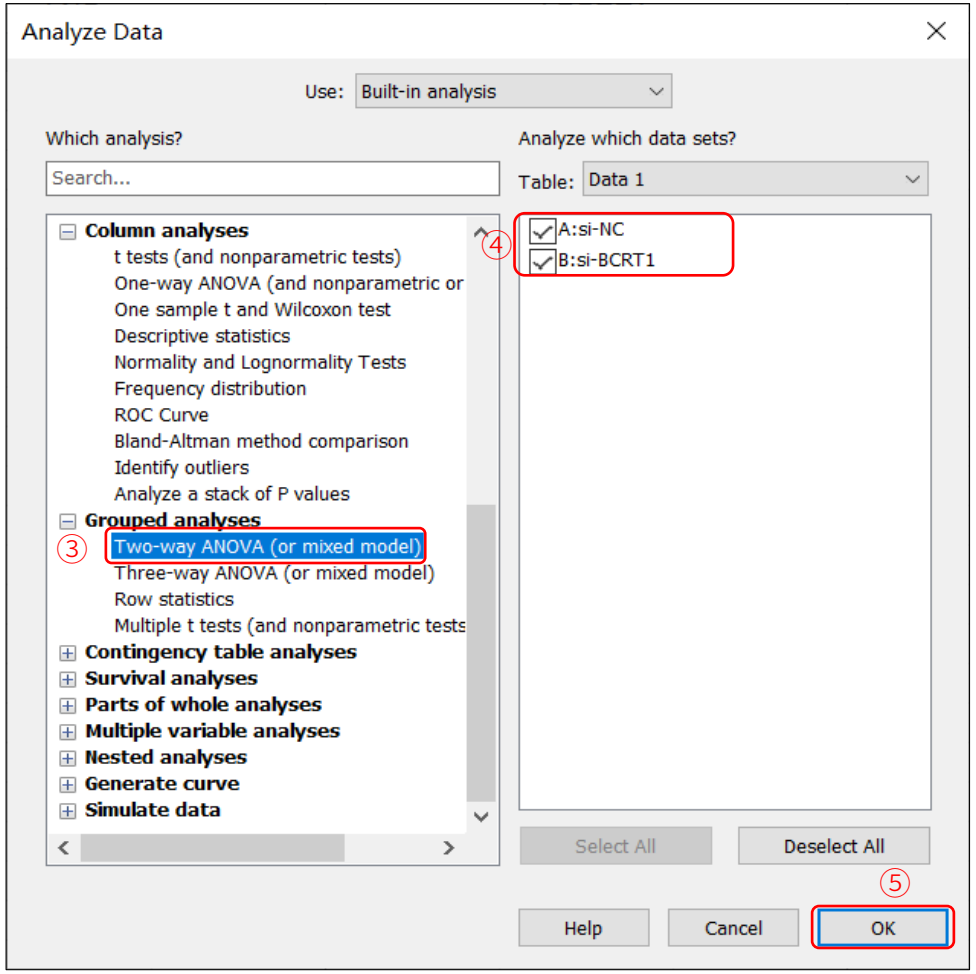
③勾选Grouped analyses下的Two-way ANOVA

④选择所有组别

⑤选择OK

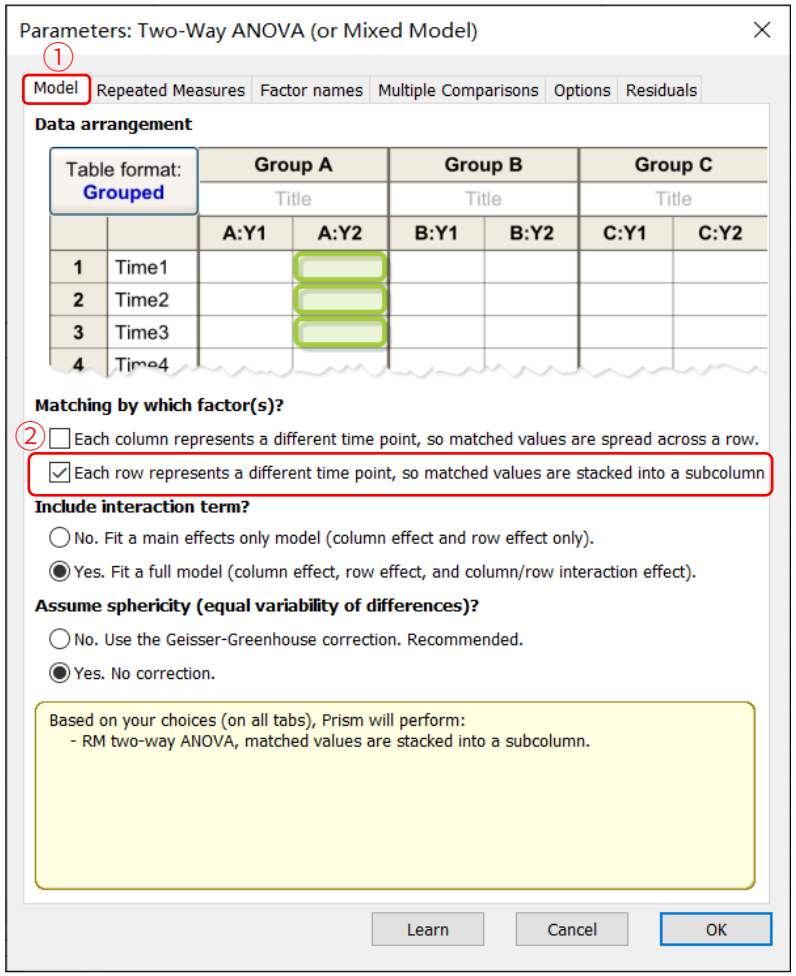


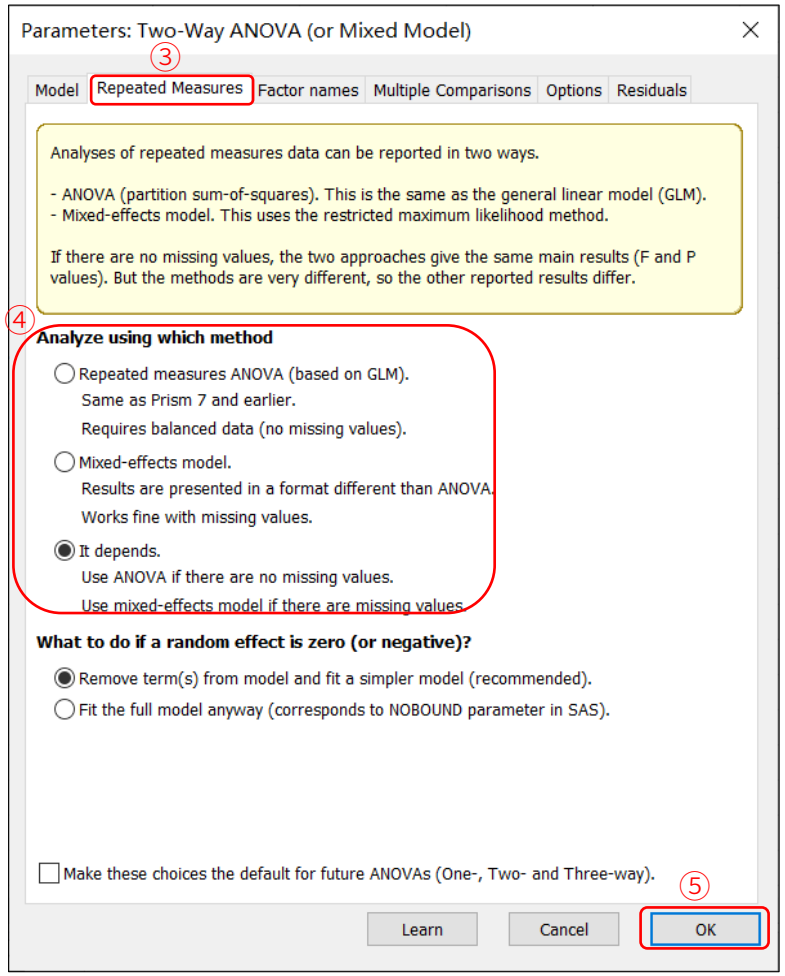

| X | Group A | | | Group B | | |
|---|---|---|---|---|---|---|
| day | si-NC | | | si-BCRT1 | | |
| X | A:Y1 | A:Y2 | A:Y3 | B:Y1 | B:Y2 | B:Y3 |
| 1.0 | 1.0000 | 1.0000 | 1.0000 | 1.0 | 1.0 | 1.0 |
| 2.0 | 1.2000 | 1.6000 | 1.8000 | 2.0 | 2.3 | 2.4 |
| 3.0 | 2.5000 | 2.6000 | 2.7000 | 4.3 | 4.4 | 4.6 |
| 4.0 | 3.9000 | 3.7000 | 3.5000 | 5.1 | 5.2 | 5.4 |
| 5.0 | 4.0000 | 4.1000 | 4.5000 | 5.4 | 5.7 | 5.9 |
| 6.0 | 5.5000 | 5.6000 | 5.9000 | 6.2 | 6.3 | 6.5 |



## 方差分析

### ② 选择实验参数




①选择Model

②选择行方向上是不同的时间点（重复测量堆叠在亚列内）

③选择Repeated Measures

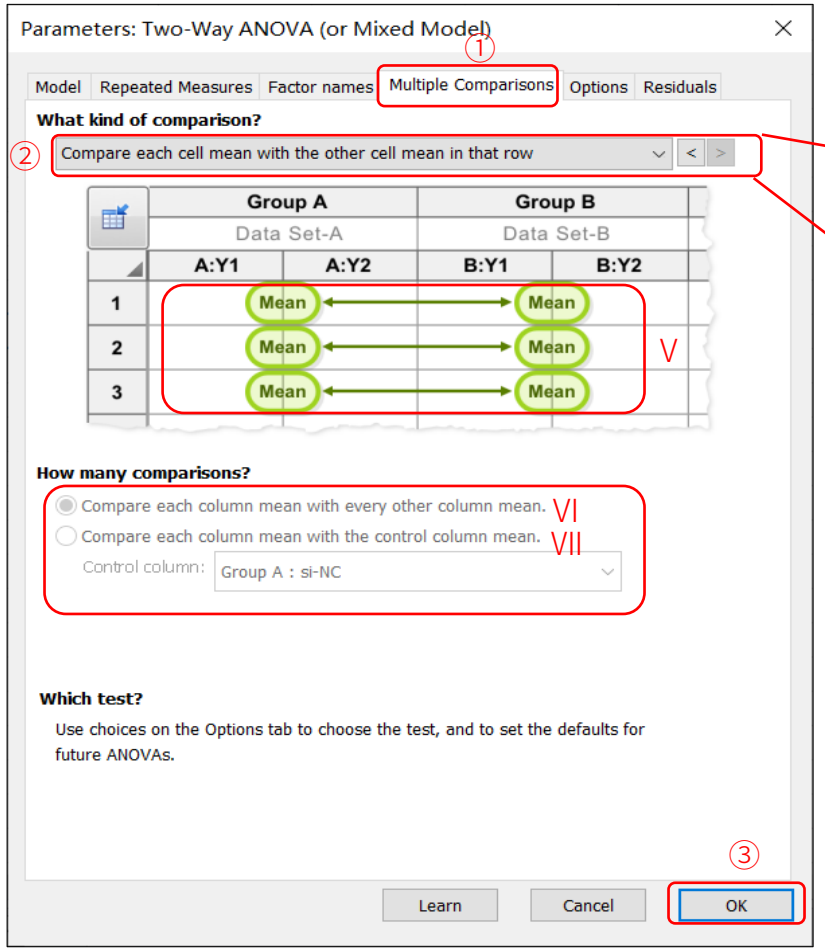
④选择分析方法

⑤选择OK

其它选项基于需要进行更改，一般情况下为默认选项

## 方差分析

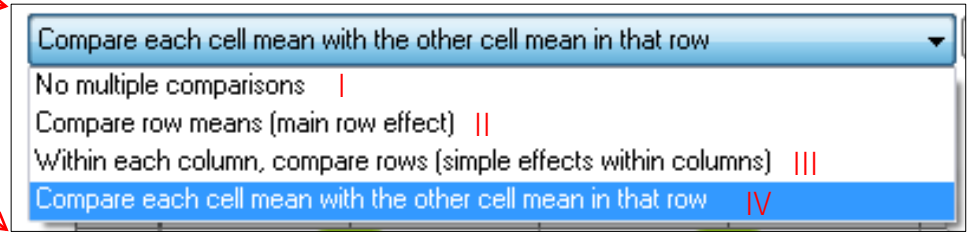
2 选择实验参数2




① 选择Multiple Comparisons

② 选择下拉菜单第四项

③ 选择OK




I. 无多重比较，即仅比较列间各组间总体上的差异

II. 在行间作多重比较，选择此项，可在下方选择是进行两两比较还是设定一个参照，其他行都跟这个参照比

III. 组内行间比较，即每组内进行行间多重比较

IV. 每行进行组别比较，即在每一行里的组别间进行比较

V. 比较方式的示意图

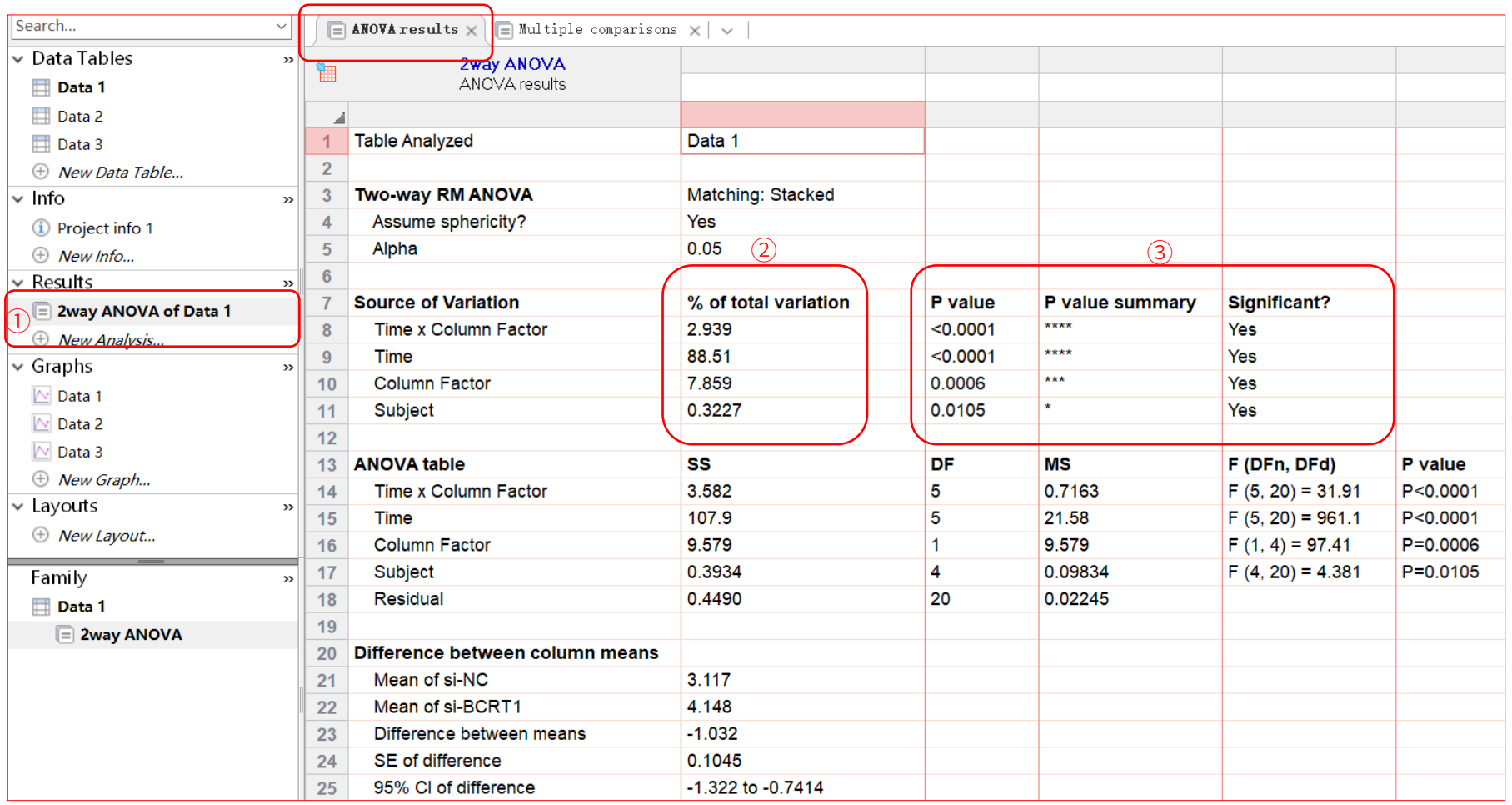
VI. 两两比较

VII. 设定一个参照，其他行都跟这个参照比

本例只有两个组别，因此不需要设置列间的多重比较

## 方差分析

### ③ 查看结果




| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | Table Analyzed | Data 1 | | | | |
| 2 | | | | | | |
| 3 | **Two-way RM ANOVA** | Matching: Stacked | | | | |
| 4 | Assume sphericity? | Yes | | | | |
| 5 | Alpha | 0.05 | | | | |
| 6 | | | | | | |
| 7 | **Source of Variation** | **% of total variation** | **P value** | **P value summary** | **Significant?** | |
| 8 | Time x Column Factor | 2.939 | <0.0001 | **** | Yes | |
| 9 | Time | 88.51 | <0.0001 | **** | Yes | |
| 10 | Column Factor | 7.859 | 0.0006 | *** | Yes | |
| 11 | Subject | 0.3227 | 0.0105 | * | Yes | |
| 12 | | | | | | |
| 13 | **ANOVA table** | **SS** | **DF** | **MS** | **F (DFn, DFd)** | **P value** |
| 14 | Time x Column Factor | 3.582 | 5 | 0.7163 | F (5, 20) = 31.91 | P<0.0001 |
| 15 | Time | 107.9 | 5 | 21.58 | F (5, 20) = 961.1 | P<0.0001 |
| 16 | Column Factor | 9.579 | 1 | 9.579 | F (1, 4) = 97.41 | P=0.0006 |
| 17 | Subject | 0.3934 | 4 | 0.09834 | F (4, 20) = 4.381 | P=0.0105 |
| 18 | Residual | 0.4490 | 20 | 0.02245 | | |
| 19 | | | | | | |
| 20 | **Difference between column means** | | | | | |
| 21 | Mean of si-NC | 3.117 | | | | |
| 22 | Mean of si-BCRT1 | 4.148 | | | | |
| 23 | Difference between means | -1.032 | | | | |
| 24 | SE of difference | 0.1045 | | | | |
| 25 | 95% CI of difference | -1.322 to -0.7414 | | | | |

① 在左边导航栏里找到“2way ANOVA of Data1”，点击其目录下的“ANOVA results”查看结果

② 各变量对总体方差变异的贡献

③ 差异性分析结果：*p*值的意义和*p*值。这里*p*值<0.05表示：该变量中至少有一对或者多对数据具有显著性差异，至于具体是哪两两数据之间有差异，则需要查看多重比较的结果

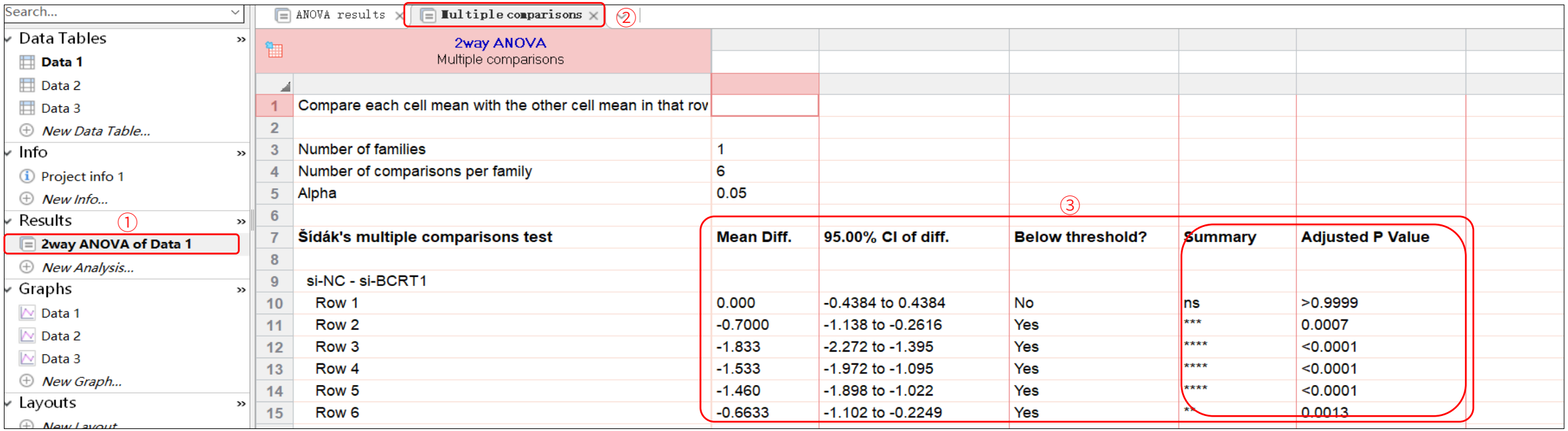
## 方差分析

### ❸ 查看结果

① 点击左边导航栏里 “2way ANOVA of Data1”

② 点击“Multiple comparisons”查看多重比较的结果

③ 多重比较的$p$值及其意义。“Row1”-“Row6”的$p$值分别对应于day1-day6，两组数据的显著性检验结果



## 卡方检验

- 示例：Steering Committee of the Physicians' Health Study Research Group. Preliminary report: Findings from the aspirin component of the ongoing Physicians' Health Study. N Engl J Med. 1988 Jan 28;318(4):262-4. doi: 10.1056/NEJM198801283180431. PMID: 3275899. Table1

- 示例解读：
  - 这是一项前瞻性研究。随机分配给受试者的两种治疗方法：安慰剂或阿司匹林。两组终点:心梗与未发生心梗。这些值是每个类别中的数目

**Table 1. Cardiovascular End Points in the Aspirin Arm of the Physicians' Health Study, According to Treatment Group.**

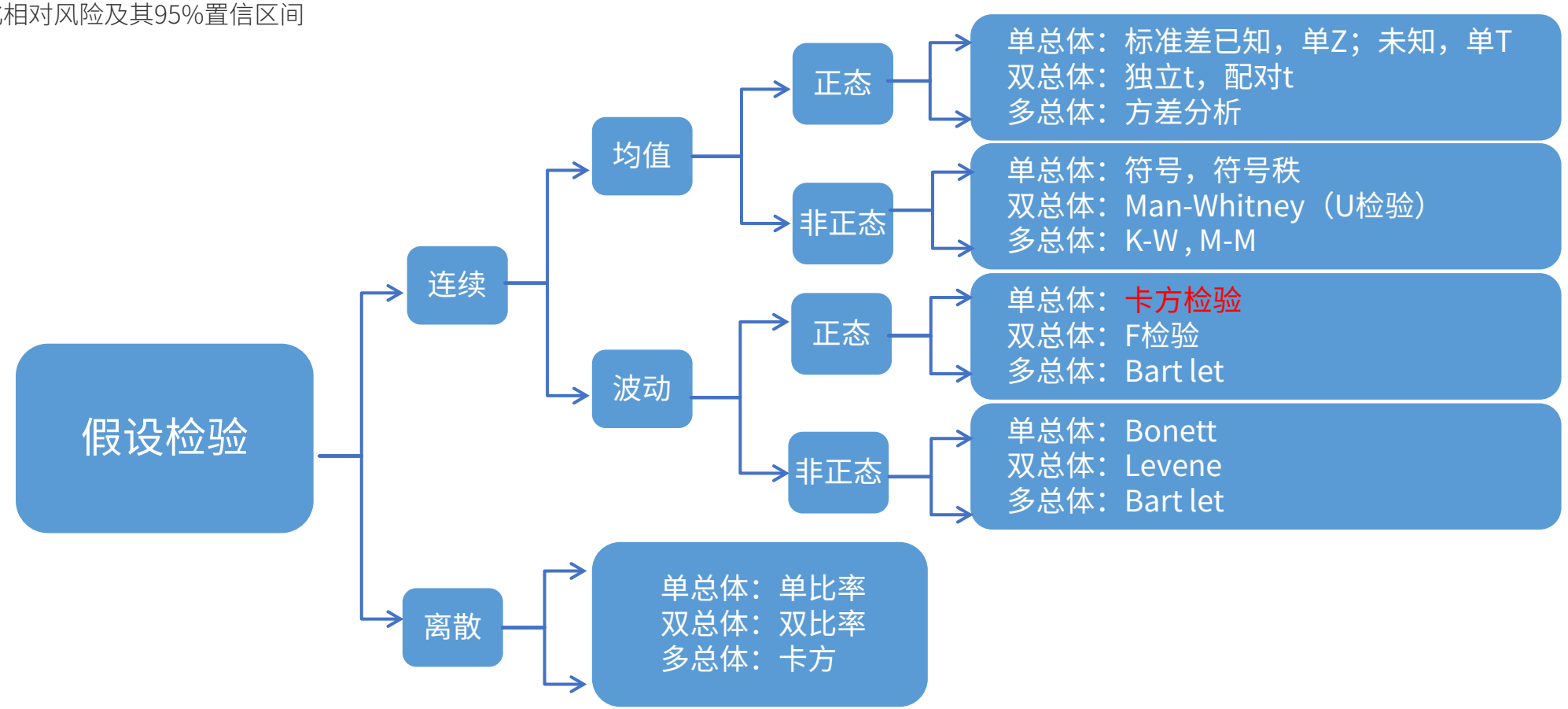
| End Point | Aspirin | Placebo | Relative Risk | 95% Confidence Interval | P Value |
|---|---|---|---|---|---|
| Myocardial infarction | | | | | |
| Fatal | 5 | 18 | 0.25 | 0.11–0.56 | 0.006 |
| Nonfatal | 99 | 171 | 0.56 | 0.44–0.71 | <0.00001 |
| Total | 104 | 189 | 0.53 | 0.42–0.67 | <0.00001 |
| Stroke | | | | | |
| Fatal | 6 | 2 | 3.00 | 0.75–11.98 | 0.16 |
| Nonfatal | 74 | 68 | 1.09 | 0.78–1.52 | 0.61 |
| Total | 80 | 70 | 1.15 | 0.84–1.58 | 0.41 |

- 模拟数据
  - 结果分为两组，分别是心梗与非心梗，治疗为使用安慰剂与使用阿司匹林

| Table format: Contingency | | Outcome A | Outcome B |
|---|---|---|---|
| | | Myocardial Infarction | No MI |
| 1 | Placebo | 189 | 10845 |
| 2 | Aspirin | 104 | 10933 |

## 卡方检验

- 分析目标
  - ✓ 评估心肌梗死发病率之间的差异是否大于预期的偶然性
  - ✓ 量化相对风险及其95%置信区间



## 卡方检验

1 选择分析方法

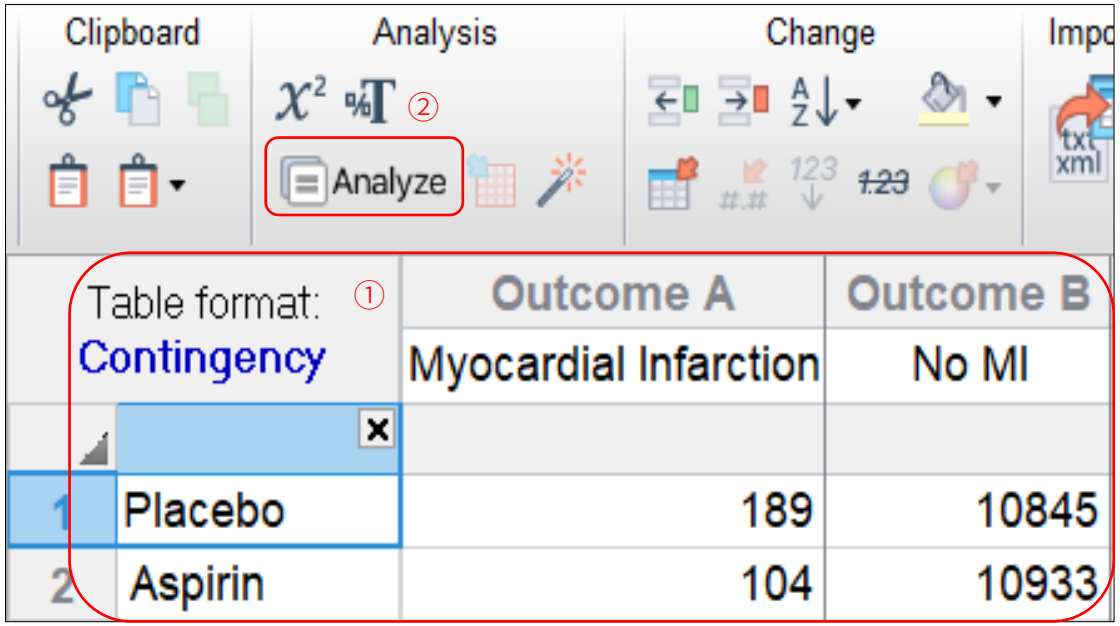
①excel拷贝数据

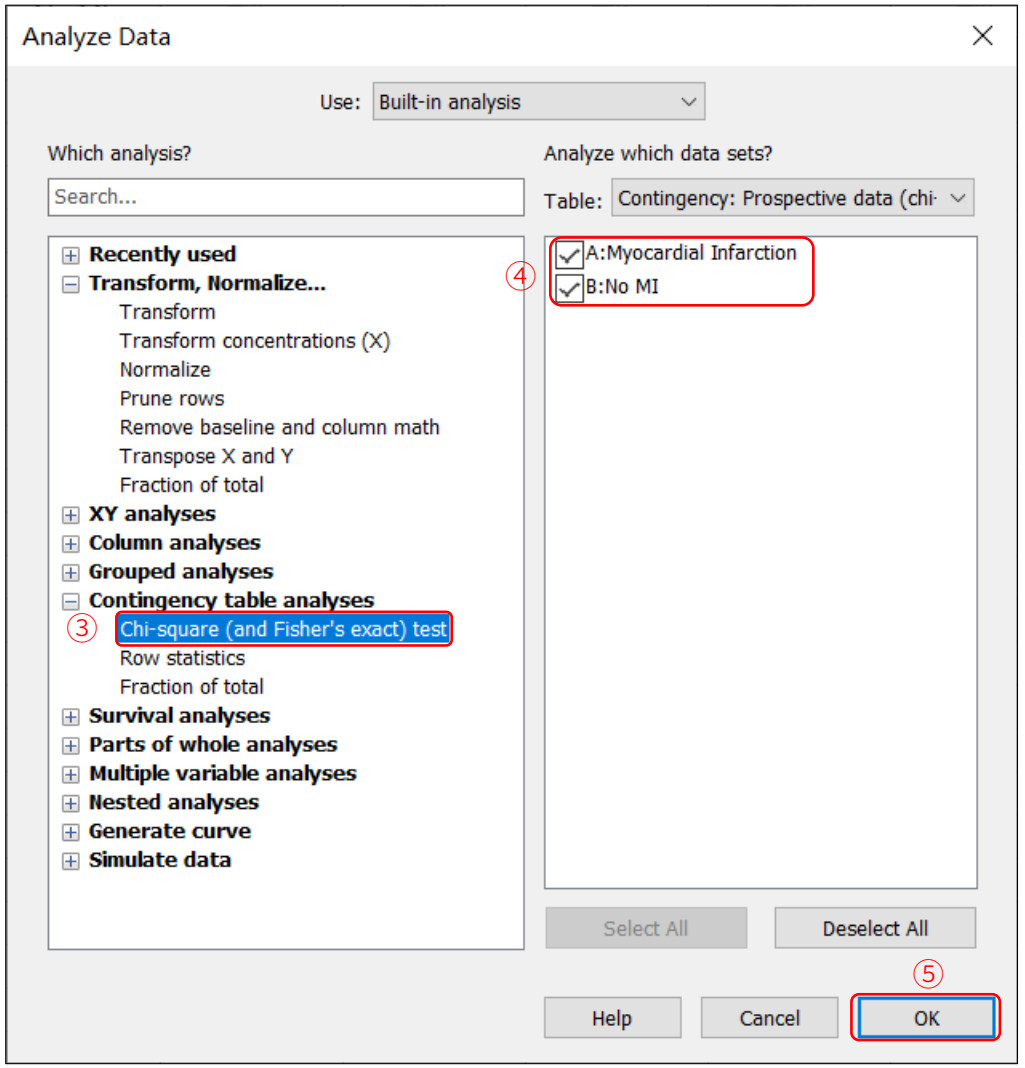
②点击分析后

③从列联表分析列表中选择“Chi-square (and Fisher's exact) test”

④选择所有组别

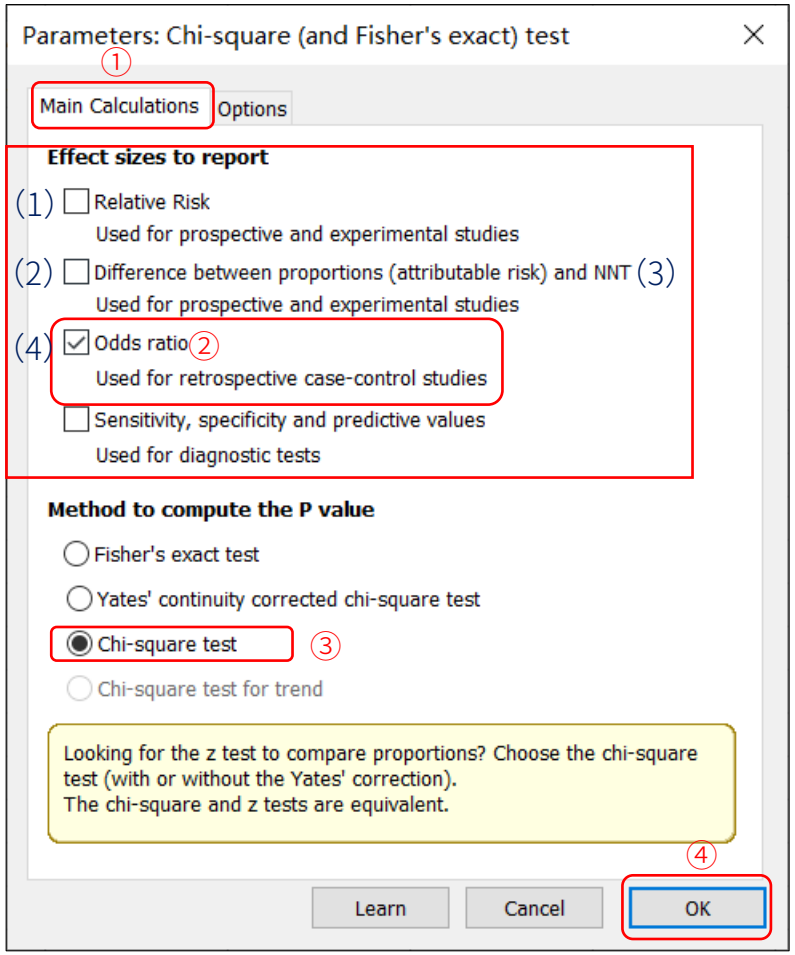
⑤选择OK



## 卡方检验

2 选择实验参数




①选择Main Calculation

②选择Odds ratio

③选择Chi-square test

④选择OK

（1）RR（Relative Risk，相对危险度）：暴露组研究对象的发病风险是非暴露组研究对象的多少倍

（2）AR（Attributable Risk，归因危险度）：暴露组发病率与非暴露组发病率之差，它反映发病归因于暴露因素的程度

（3）NNT（Number Needed to Treat，需治疗人数）：可把抽象的率转变为 1 个具体的频数，使临床试验结果转化为临床实践应用的指标，具有表达统计学意义及临床意义的双重作用，是一个衡量临床治疗效果、指导临床决策的有用工具

（4）OR（Odds Ratio，比值比）：在回顾性研究中可以评价暴露因素（阿司匹林）和疾病（心肌梗死）的关联强度

## 卡方检验

3 查看结果

| Contingency | A | B | C |
|---|---|---|---|
| 1 | **Table Analyzed** | Contingency: Prospective data (chi-square test) | | |
| 2 | | | | |
| 3 | **P value and statistical significance** | | | |
| 4 | Test | Chi-square | | |
| 5 | Chi-square, df | 25.01, 1 | | |
| 6 | z | 5.001 | | |
| 7 | P value | <0.0001 | | |
| 8 | P value summary | **** | | |
| 9 | One- or two-sided | Two-sided | | |
| 10 | Statistically significant (P < 0.05)? | Yes | | |
| 11 | | | | |
| 12 | **Effect size** | **Value** | **95% CI** | |
| 13 | Odds ratio | 1.832 | 1.439 to 2.338 | |
| 14 | Reciprocal of odds ratio | 0.5458 | 0.4277 to 0.6948 | |
| 15 | | | | |
| 16 | **Methods used to compute CIs** | | | |
| 17 | Odds ratio | Baptista-Pike | | |
| 18 | | | | |
| 19 | **Data analyzed** | **Myocardial Infarction** | **No MI** | **Total** |
| 20 | Placebo | 189 | 10845 | 11034 |
| 21 | Aspirin | 104 | 10933 | 11037 |
| 22 | Total | 293 | 21778 | 22071 |
| 23 | | | | |
| 24 | **Percentage of row total** | **Myocardial Infarction** | **No MI** | |
| 25 | Placebo | 1.71% | 98.29% | |
| 26 | Aspirin | 0.94% | 99.06% | |

# 课程目录

## 相关分析

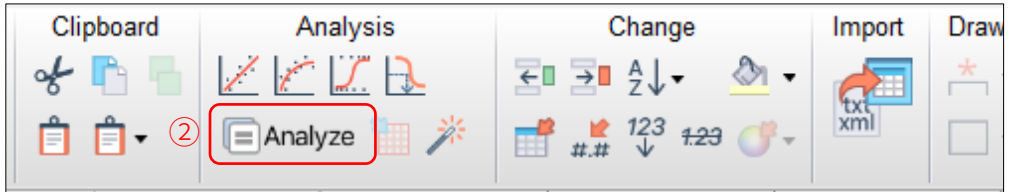
- 示例数据来源于Graphpad
- 每一行代表不同的日子。X柱表示臭氧水平。Y柱表示太阳辐射、风和温度
- 目标：了解这三个天气变量与臭氧水平之间的关系

| | X | Group A | Group B | Group C |
|---|---|---|---|---|
| ① | Ozone | Solar.R | Wind | Temp |
| | X | Y | Y | Y |
| 1 | 41 | 190 | 7.4 | 67 |
| 2 | 36 | 118 | 8.0 | 72 |
| 3 | 12 | 149 | 12.6 | 74 |
| 4 | 18 | 313 | 11.5 | 62 |
| 5 | | | 14.3 | 56 |
| 6 | 28 | | 14.9 | 66 |
| 7 | 23 | 299 | 8.6 | 65 |
| 8 | 19 | 99 | 13.8 | 59 |
| 9 | 8 | 19 | 20.1 | 61 |
| 10 | | 194 | 8.6 | 69 |
| 11 | 7 | | 6.9 | 74 |
| 12 | 16 | 256 | 9.7 | 69 |
| 13 | 11 | 290 | 9.2 | 66 |
| 14 | 14 | 274 | 10.9 | 68 |
| 15 | 18 | 65 | 13.2 | 58 |
| 16 | 14 | 334 | 11.5 | 64 |
| 17 | 34 | 307 | 12.0 | 66 |
| 18 | 6 | 78 | 18.4 | 57 |
| 19 | 30 | 322 | 11.5 | 68 |
| 20 | 11 | 44 | 9.7 | 62 |
| 21 | 1 | 8 | 9.7 | 59 |
| 22 | 11 | 320 | 16.6 | 73 |
| 23 | 4 | 25 | 9.7 | 61 |
| 24 | 32 | 92 | 12.0 | 61 |
| 25 | | 66 | 16.6 | 57 |
| 26 | | 266 | 14.9 | 58 |
| 27 | | | 8.0 | 57 |

①录入数据（此处可以非完整数据，也就是说允许数据的缺少）

②点击分析按钮

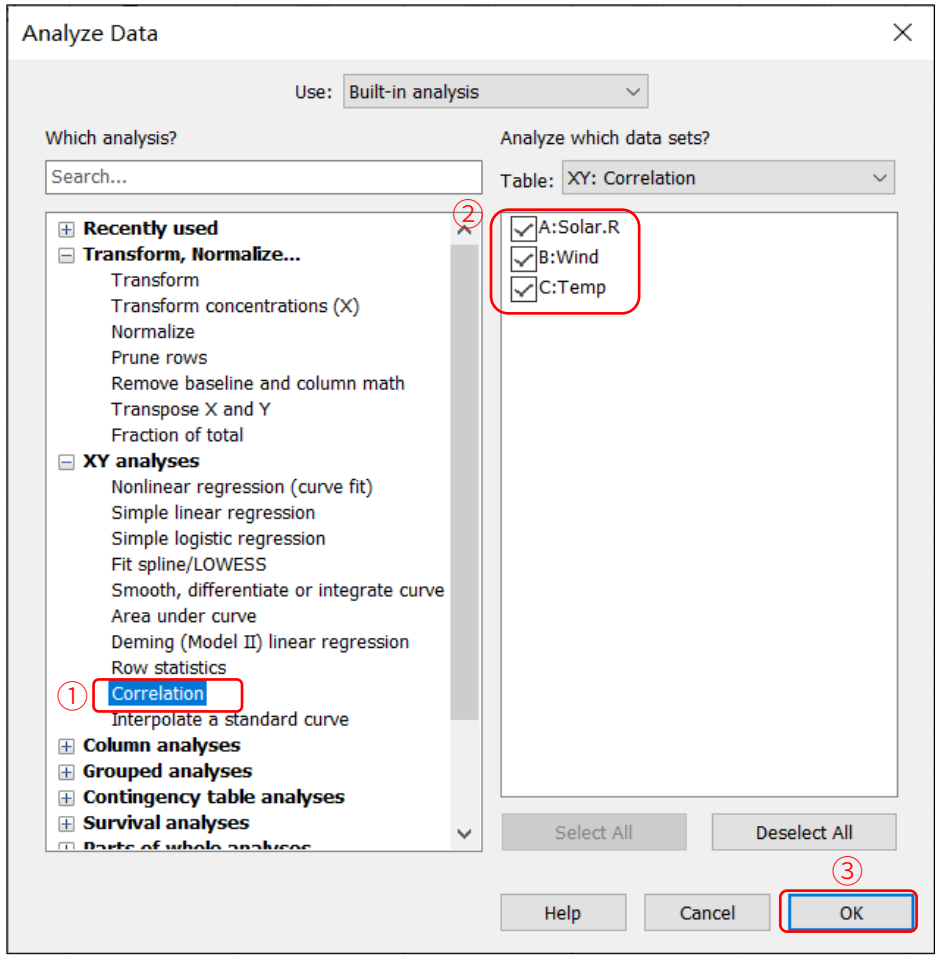
① 调出分析窗口



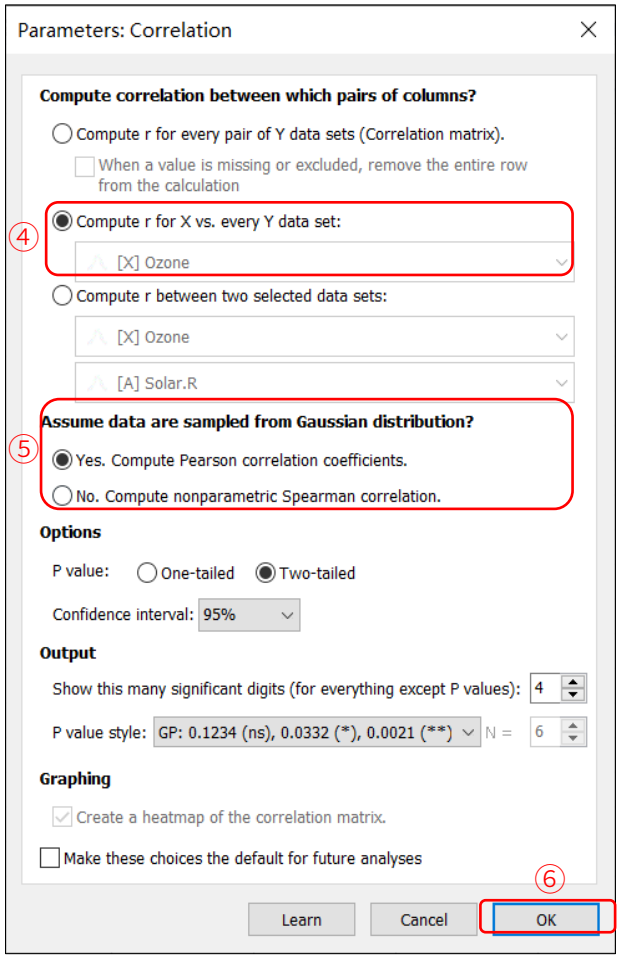

| | X | Group A | Group B | Group C |
|---|---|---|---|---|
| | Ozone | Solar.R | Wind | Temp |
| | X | Y | Y | Y |
| 1 | 41 | 190 | 7.4 | 67 |
| 2 | 36 | 118 | 8.0 | 72 |
| 3 | 12 | 149 | 12.6 | 74 |
| 4 | 18 | 313 | 11.5 | 62 |
| 5 | | | 14.3 | 56 |
| 6 | 28 | | 14.9 | 66 |
| 7 | 23 | 299 | 8.6 | 65 |
| 8 | 19 | 99 | 13.8 | 59 |
| 9 | 8 | 19 | 20.1 | 61 |
| 10 | | 194 | 8.6 | 69 |
| 11 | 7 | | 6.9 | 74 |
| 12 | 16 | 256 | 9.7 | 69 |

## 相关分析

### ② 选择分析方法




### ③ 分析参数选择




① 选择相关分析

② 勾选所有组别

③ 选择OK

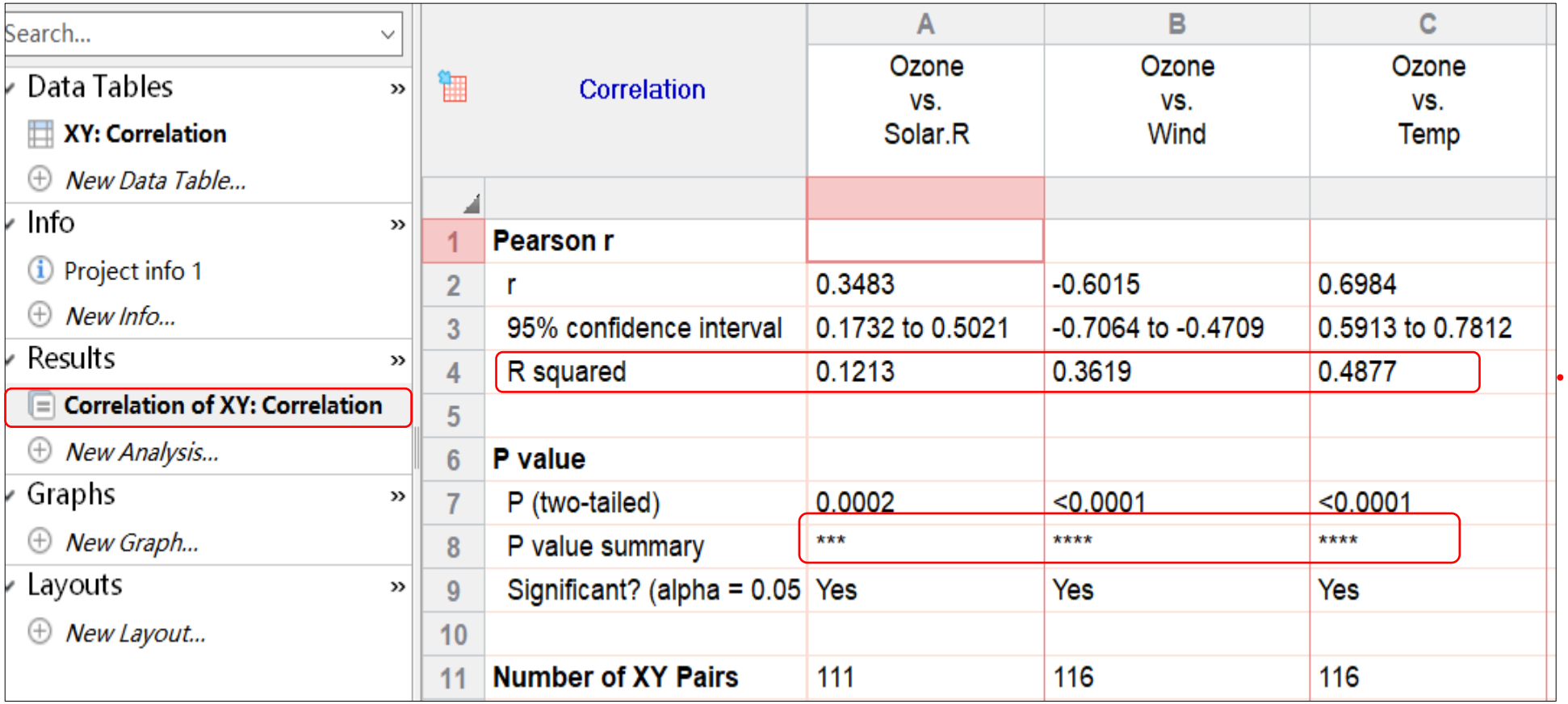
④ 选择臭氧作为X（三个天气变量Y与臭氧水平X之间的关系）

⑤ 根据是否符合高斯分布（即正态分布）选择皮尔森系数或者斯皮尔曼系数

⑥ 其它选项默认，选择OK

## 相关分析

4 查看分析结果




| | Correlation | A<br>Ozone vs. Solar.R | B<br>Ozone vs. Wind | C<br>Ozone vs. Temp |
|---|---|---|---|---|
| 1 | **Pearson r** | | | |
| 2 | r | 0.3483 | -0.6015 | 0.6984 |
| 3 | 95% confidence interval | 0.1732 to 0.5021 | -0.7064 to -0.4709 | 0.5913 to 0.7812 |
| 4 | R squared | 0.1213 | 0.3619 | 0.4877 |
| 5 | | | | |
| 6 | **P value** | | | |
| 7 | P (two-tailed) | 0.0002 | <0.0001 | <0.0001 |
| 8 | P value summary | *** | **** | **** |
| 9 | Significant? (alpha = 0.05 | Yes | Yes | Yes |
| 10 | | | | |
| 11 | **Number of XY Pairs** | 111 | 116 | 116 |

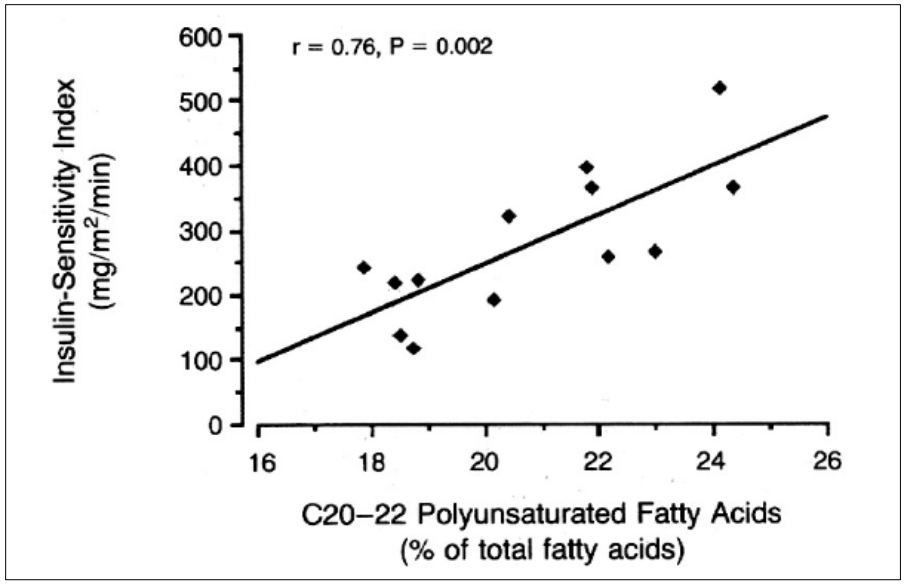
- R平方，即相关系数的平方，也就是决定系数，越接近1表示相关性越强

# 课程目录

## 线性回归分析

- 所有模型均将结果（Y）定义为一个或多项参数和一个独立变量（X）或几个独立变量的函数。目的是调整模型参数值，以找到最接近数据的直线或曲线
  例如，对于线性回归，目的是找到斜率和截距的最佳拟合值，以使直线接近数据。
  对标准化剂量-反应曲线进行非线性回归，目的是调整EC50的值（在最小和最大反应之间激发反应的浓度）和曲线斜率
- 线性回归示例：文献： Borkman M, Storlien LH, Pan DA, Jenkins AB, Chisholm DJ, Campbell LV. The relation between insulin sensitivity and the fatty-acid composition of skeletal-muscle phospholipids. N Engl J Med. 1993 Jan 28;328(4):238-44. doi: 10.1056/NEJM199301283280404. PMID: 8418404. Figure3a

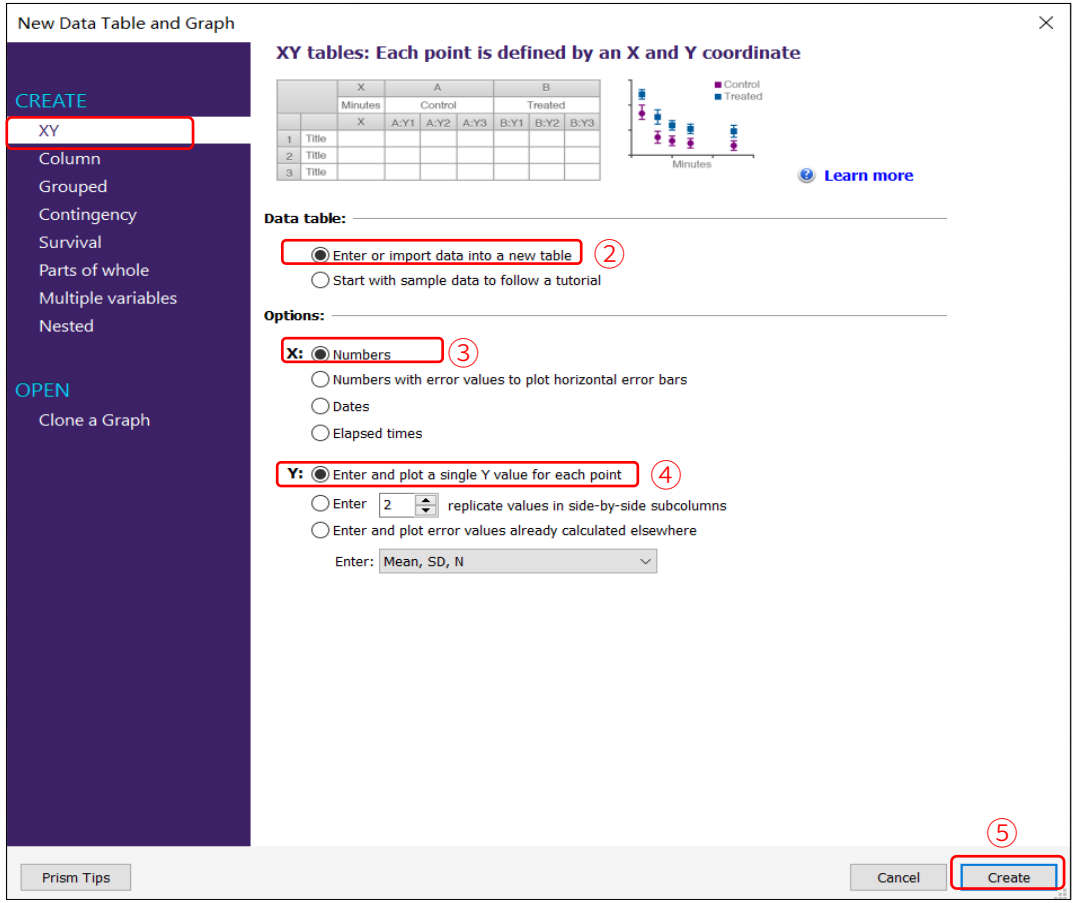
  - 示例解读：X值是肌肉中某一种脂肪酸的百分比（来自活组织检查），Y值是胰岛素敏感性
  - 目的：用线性回归来拟合数据中的直线
  - 在许多情况下，X是一个需要操作的变量，Y是一个需要测量的变量。在本例中这两个变量都被测量了，并且并不完全清楚哪个应该是X，哪个应该是Y



## 线性回归分析

### 1 建立XY数据表

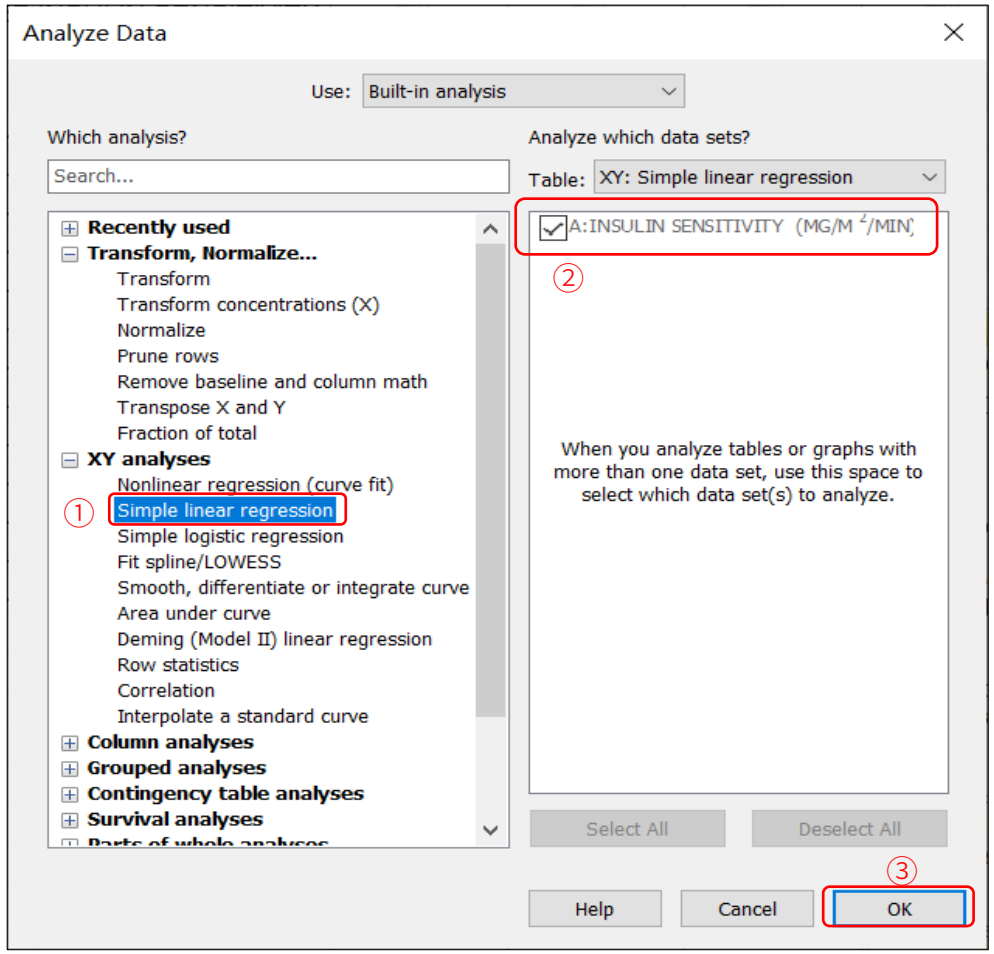
① 选择XY表；②选择新创建数据；③X选择数字；④Y中每个点输入独立的Y值；⑤点击Create后，右图内键入数据



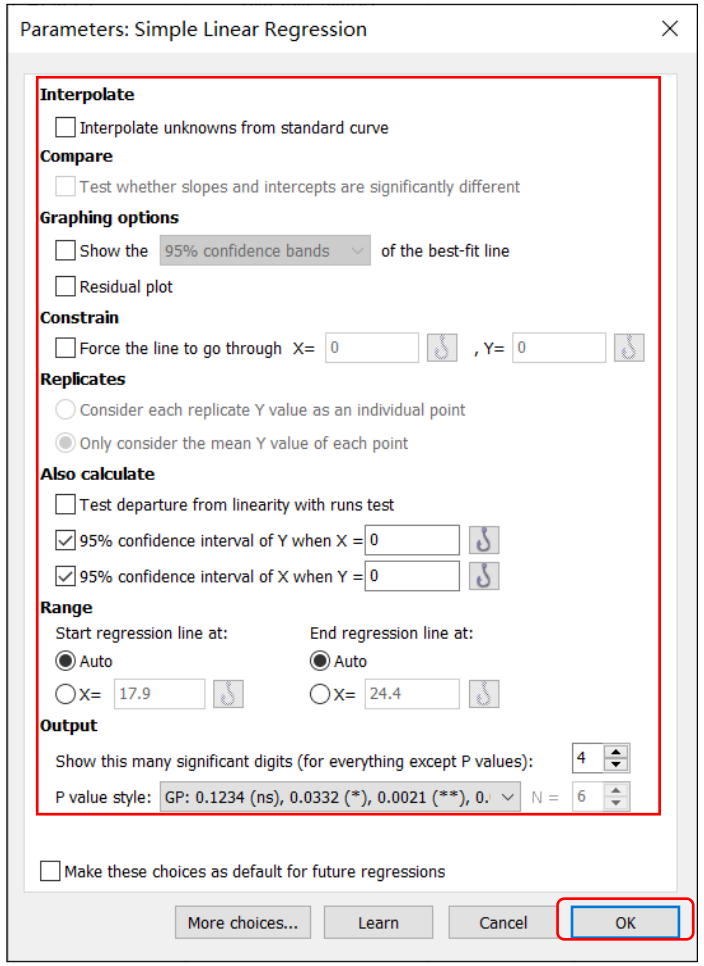

| Table format: XY | X | Group A |
|---|---|---|
| | %C20-22 POLYUNSATURATED FATTY ACIDS | INSULIN SENSITIVITY ($MG/M^2/MIN$) |
| | X | Y |
| 1 Title | 17.9 | 250 |
| 2 Title | 18.3 | 220 |
| 3 Title | 18.3 | 145 |
| 4 Title | 18.4 | 115 |
| 5 Title | 18.4 | 230 |
| 6 Title | 20.2 | 200 |
| 7 Title | 20.3 | 330 |
| 8 Title | 21.8 | 400 |
| 9 Title | 21.9 | 370 |
| 10 Title | 22.1 | 260 |
| 11 Title | 23.1 | 270 |
| 12 Title | 24.2 | 530 |
| 13 Title | 24.4 | 375 |

## 线性回归分析

### ② 选择分析方法




### ③ 设置统计参数（可默认）




① 选择简单线性回归

② 勾选所有组别

由于Y方向的组别只有1组，因此默认已选上这一组，且不能修改

若有多个组别，这里可以选择一到多个组别

③ 其它选项默认，选择OK

(1) 通过标准曲线填补缺失数据

(2) 检验斜率和截距是否显著不同

(3) 添加某个CI值的线

(4) 绘制残差图

(5) 可强迫拟合的直线经过某个点

(6) 当有重复实验时，这里可设置用平均值或者是各个独立的实验值进行分析

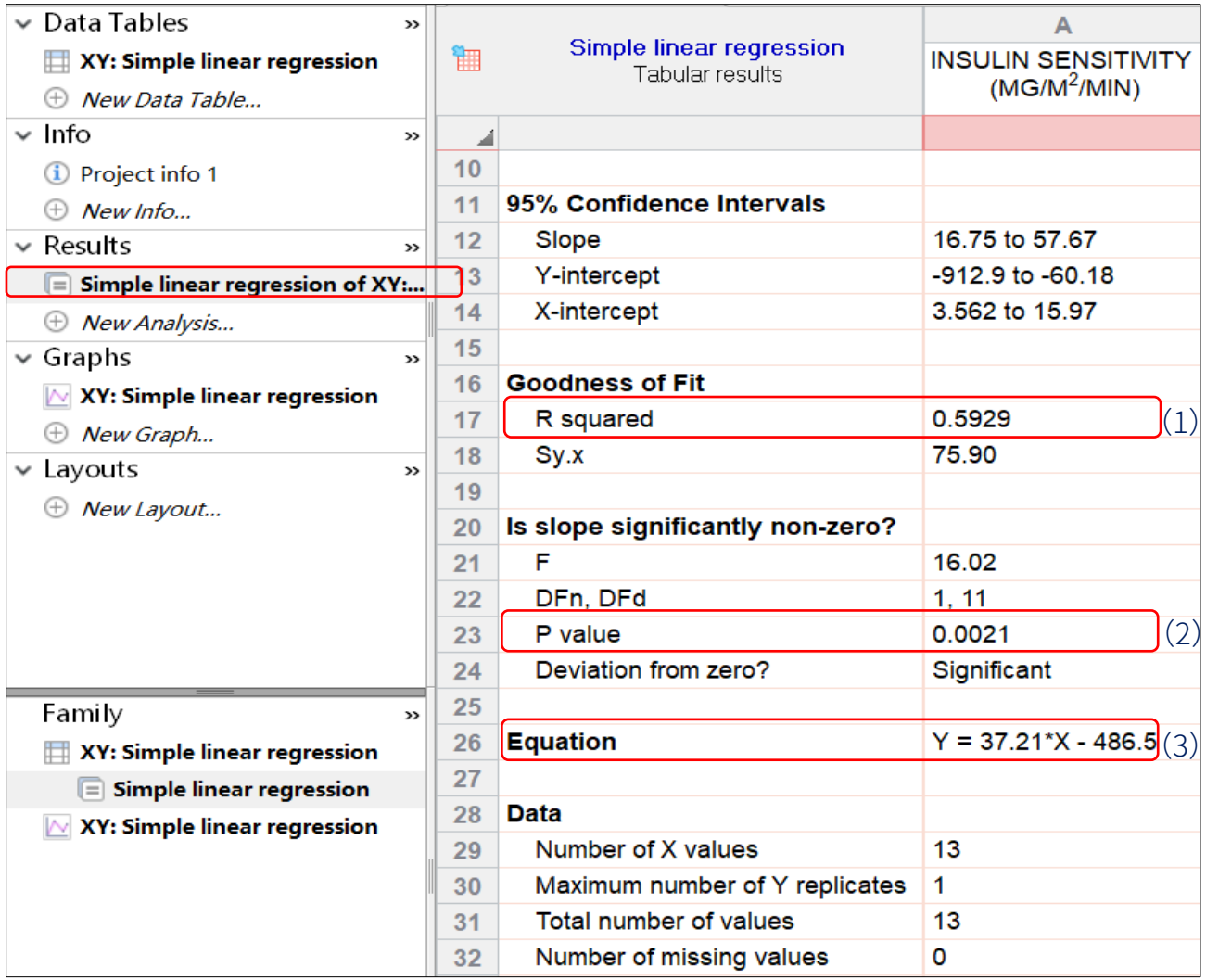
(7) 计算指定数据点的95%CI

(8) 可设置在某段数据范围内进行回归分析

(9) 显示回归分析中XY对应的值

## 线性回归分析

4 查看分析结果




Simple linear regression
Tabular results

| | | A |
|---|---|---|
| | | INSULIN SENSITIVITY (MG/M$^2$/MIN) |
| 10 | | |
| 11 | **95% Confidence Intervals** | |
| 12 | Slope | 16.75 to 57.67 |
| 13 | Y-intercept | -912.9 to -60.18 |
| 14 | X-intercept | 3.562 to 15.97 |
| 15 | | |
| 16 | **Goodness of Fit** | |
| 17 | R squared | 0.5929 (1) |
| 18 | Sy.x | 75.90 |
| 19 | | |
| 20 | **Is slope significantly non-zero?** | |
| 21 | F | 16.02 |
| 22 | DFn, DFd | 1, 11 |
| 23 | P value | 0.0021 (2) |
| 24 | Deviation from zero? | Significant |
| 25 | | |
| 26 | **Equation** | Y = 37.21*X - 486.5 (3) |
| 27 | | |
| 28 | **Data** | |
| 29 | Number of X values | 13 |
| 30 | Maximum number of Y replicates | 1 |
| 31 | Total number of values | 13 |
| 32 | Number of missing values | 0 |

（1）R squared，即相关系数的平方，也就是决定系数，越接近1表示相关性越强

（2）*p*值

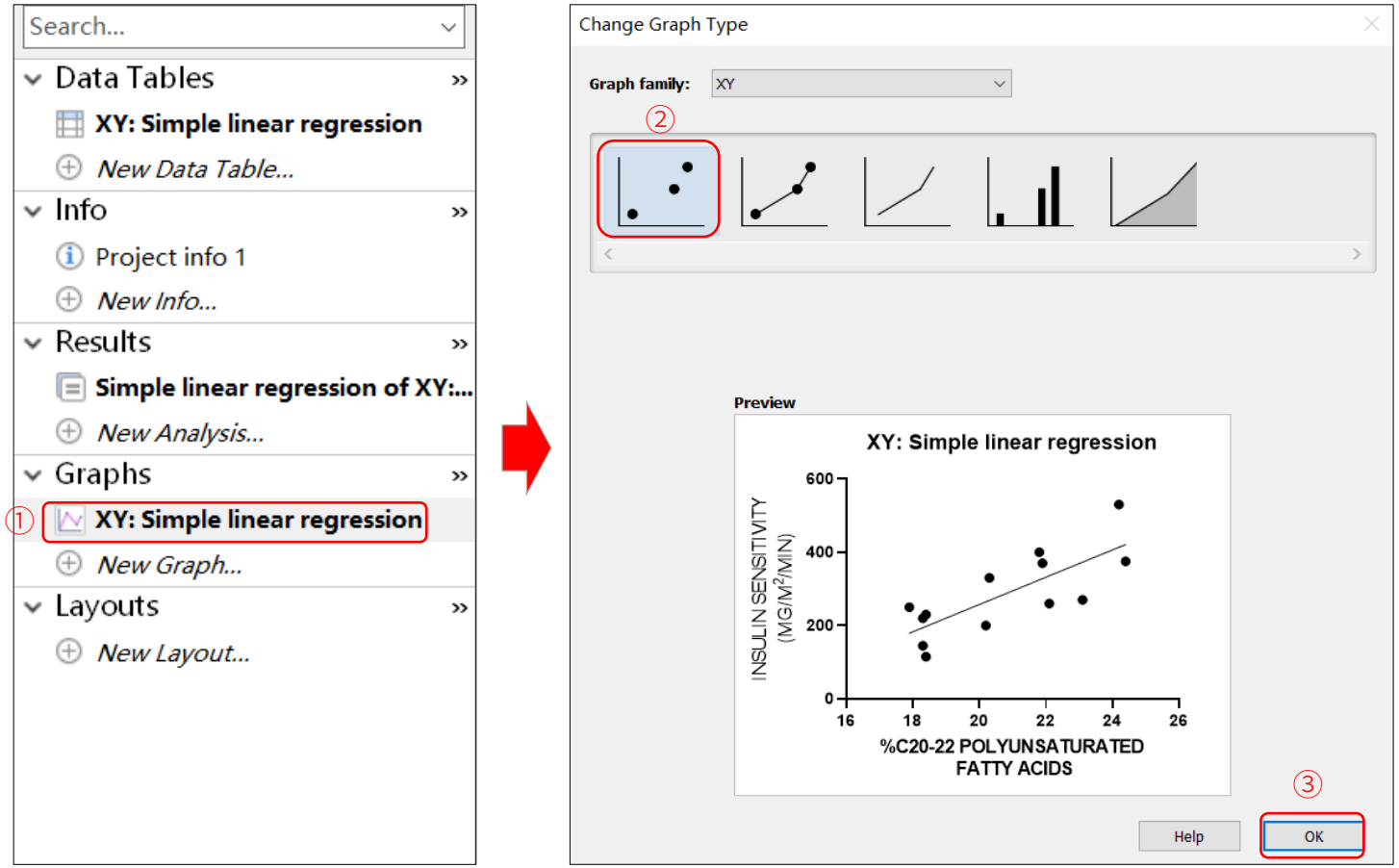
（3）回归方程：斜率＜0表示是负相关

斜率＞0表示正相关

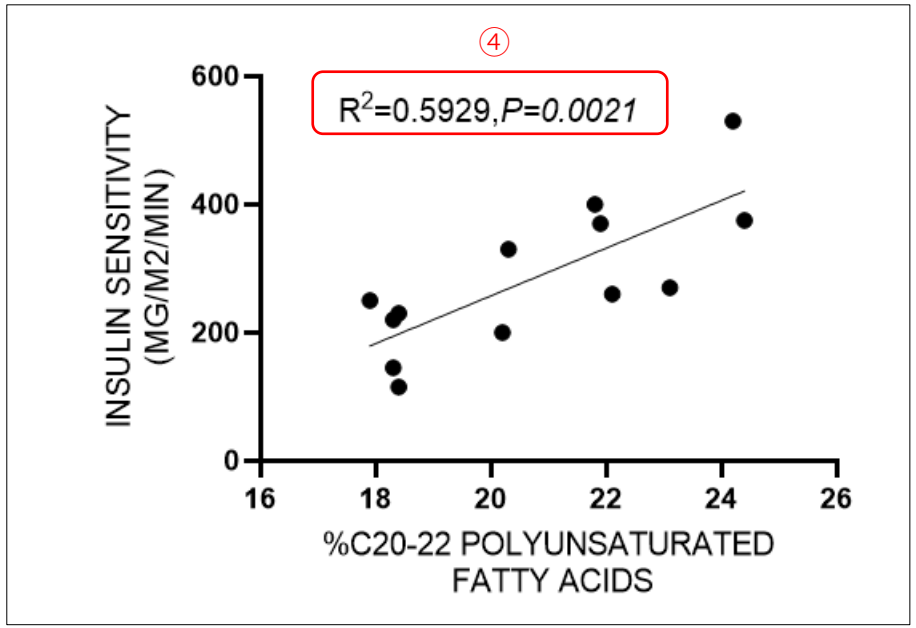
## 线性回归分析

### 5 点击图表视图

① 点击Graphs 下XY；②选择散点图；③点击OK； ④Y用文本框工具，将R平方和p值标记在图上




### 6 标记分析结果



## 非线性回归分析

- 所有模型均将结果（Y）定义为一个或多项参数和一个独立变量（X）或几个独立变量的函数。目的是调整模型参数值，以找到最接近数据的直线或曲线。

  例如，对于线性回归，目的是找到斜率和截距的最佳拟合值，以使直线接近数据。

  对标准化剂量-反应曲线进行非线性回归，目的是调整EC50的值（在最小和最大反应之间激发反应的浓度）和曲线斜率

- 非线性回归示例：

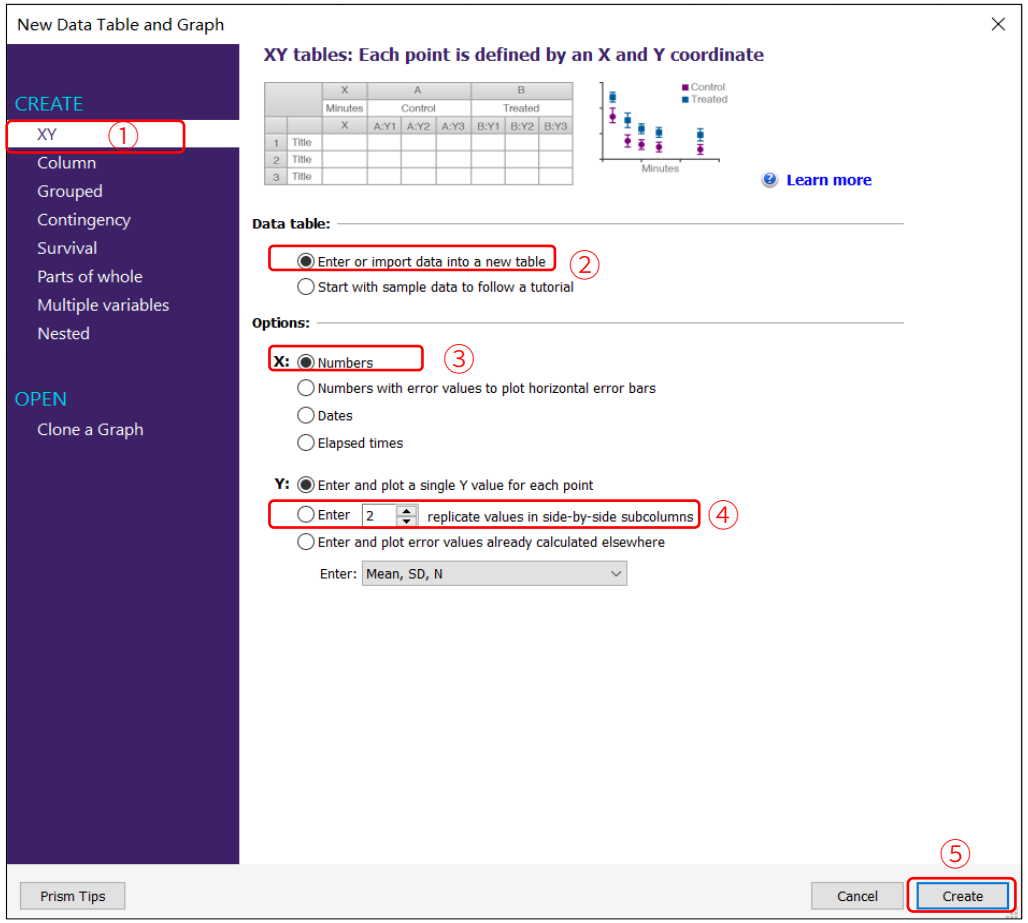
示例解读：

- X列记录时间；每个时间点相对应三次数据，用于数据质量。有些单元格是空白的，以表示缺失的数据
- 目的：拟合指数衰减曲线以确定衰减速率常数以及Y的起始值和平台值

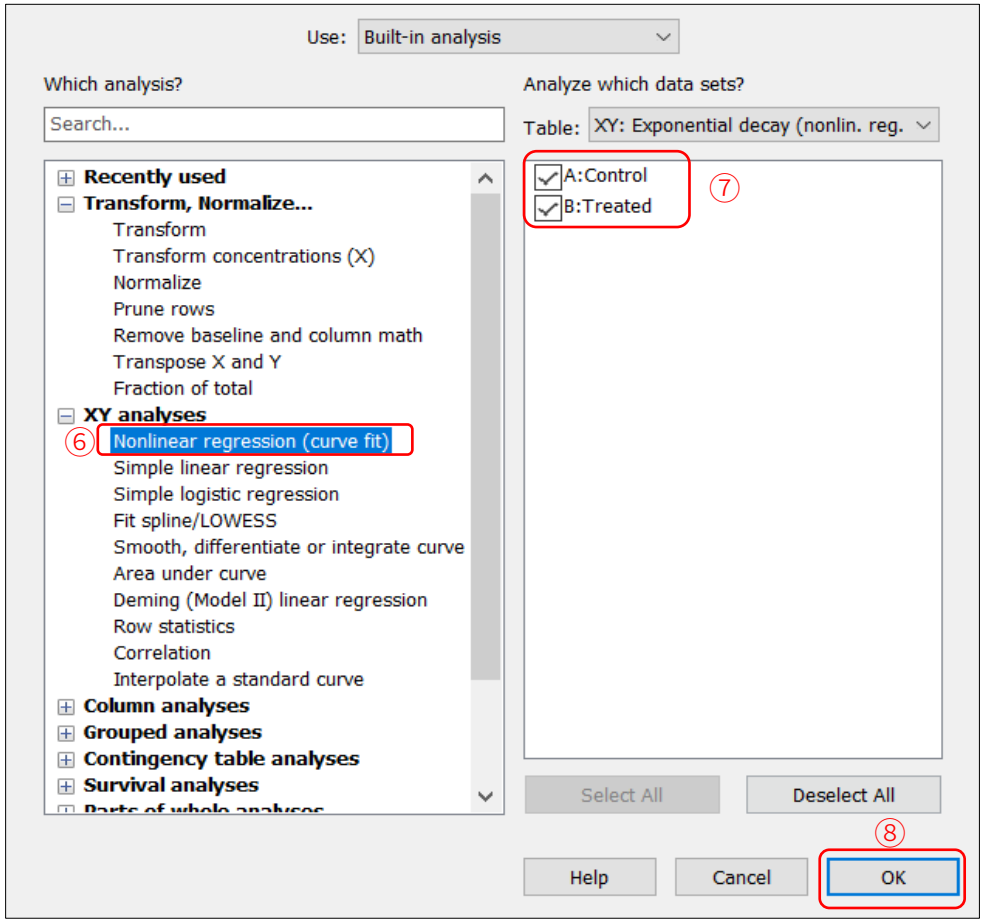
| Table format: XY | | X | Group A | | | Group B | | |
|---|---|---|---|---|---|---|---|---|
| | | Minutes | Control | | | Treated | | |
| | | X | A:Y1 | A:Y2 | A:Y3 | B:Y1 | B:Y2 | B:Y3 |
| 1 | Title | 1.0 | 8887 | 7366 | 9612 | 6532 | 7905 | 7907 |
| 2 | Title | 2.0 | 8329 | | 8850 | 5352 | 5841 | 6277 |
| 3 | Title | 3.0 | 7907 | 8810 | 8669 | 5177 | 4082 | 3157 |
| 4 | Title | 4.0 | 7413 | 8481 | 6489 | 3608 | | 4226 |
| 5 | Title | 5.0 | 7081 | 7178 | 5716 | 2559 | 3697 | 2816 |
| 6 | Title | 6.0 | 6249 | 6492 | | 1671 | 3053 | 2891 |
| 7 | Title | 8.0 | 5442 | 6172 | 6409 | 2264 | 1658 | 1879 |
| 8 | Title | 10.0 | 4020 | 3758 | 4138 | 1905 | 1302 | 1406 |
| 9 | Title | 14.0 | 4559 | 3146 | 2547 | 2994 | 1338 | 739 |
| 10 | Title | 20.0 | 3033 | 1587 | 2754 | 1444 | | 760 |
| 11 | Title | 25.0 | 2105 | 1707 | 2152 | 281 | 484 | 765 |
| 12 | Title | 30.0 | 1005 | 2156 | 1185 | 1103 | 1517 | 833 |
| 13 | Title | 50.0 | 820 | 1513 | 1591 | 1918 | 1128 | 1293 |

## 非线性回归分析

1 建立XY数据表，复制数据到表中



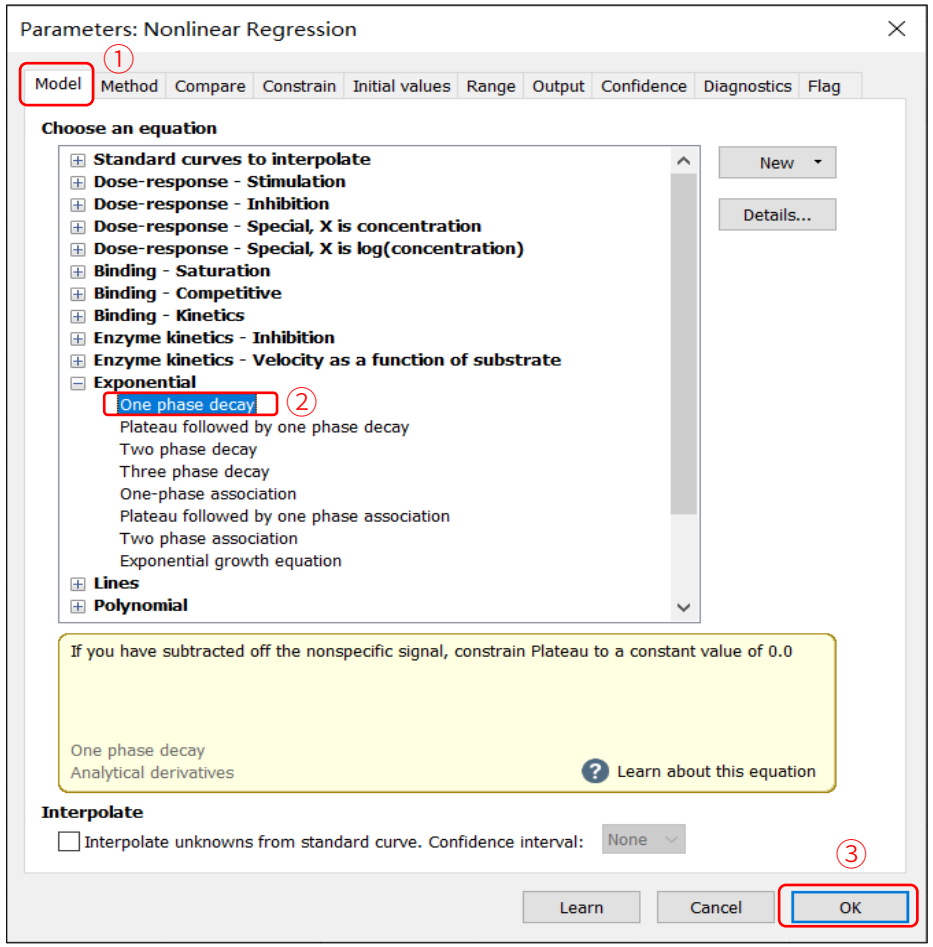

2 分析数据




① 选择XY表；
② 选择新创建数据；
③ X选择数字；
④ Y中输入三个的数据；
⑤ 点击Create
⑥ 选择非线性回归
⑦ 选择数据
⑧ OK

## 非线性回归分析

### ③ 选择模型




### ④ 结果显示

| | Nonlin fit Table of results | A Control | B Treated |
|---|---|---|---|
| 1 | **One phase decay** | | |
| 2 | **Best-fit values** | | |
| 3 | Y0 (1) | 9992 | 9593 |
| 4 | Plateau (2) | 987.0 | 1154 |
| 5 | K (3) | 0.08927 | 0.3042 |
| 6 | Half Life (4) | 7.765 | 2.278 |
| 7 | Tau (5) | 11.20 | 3.287 |
| 8 | Span (6) | 9005 | 8440 |
| 9 | **95% CI (profile likelihood)** | | |
| 10 | Y0 | 9301 to 10738 | 8419 to 10967 |
| 11 | Plateau | 201.0 to 1638 | 813.2 to 1480 |
| 12 | K | 0.06862 to 0.1134 | 0.2387 to 0.3862 |
| 13 | Half Life | 6.110 to 10.10 | 1.795 to 2.903 |
| 14 | Tau | 8.815 to 14.57 | 2.589 to 4.189 |
| 15 | **Goodness of Fit** | | |
| 16 | Degrees of Freedom | 34 | 34 |
| 17 | R squared | 0.9401 | 0.9196 |
| 18 | Sum of Squares | 16891590 | 12555343 |
| 19 | Sy.x | 704.8 | 607.7 |
| 20 | **Constraints** | | |
| 21 | K | K > 0 | K > 0 |
| 22 | | | |
| 23 | **Number of points** | | |
| 24 | # of X values | 39 | 39 |
| 25 | # Y values analyzed | 37 | 37 |

① 选择model

② 选择One phase decay

③ OK

（1） Y0是X（时间）为0时的Y值。与Y用相同的单位表示

（2） Plateau是无穷次时的Y值，与Y用相同的单位表示

（3） K是速率常数，用X轴时间单位的倒数表示

（4） 半衰期在X轴的时间单位内，它被计算为ln (2) /K

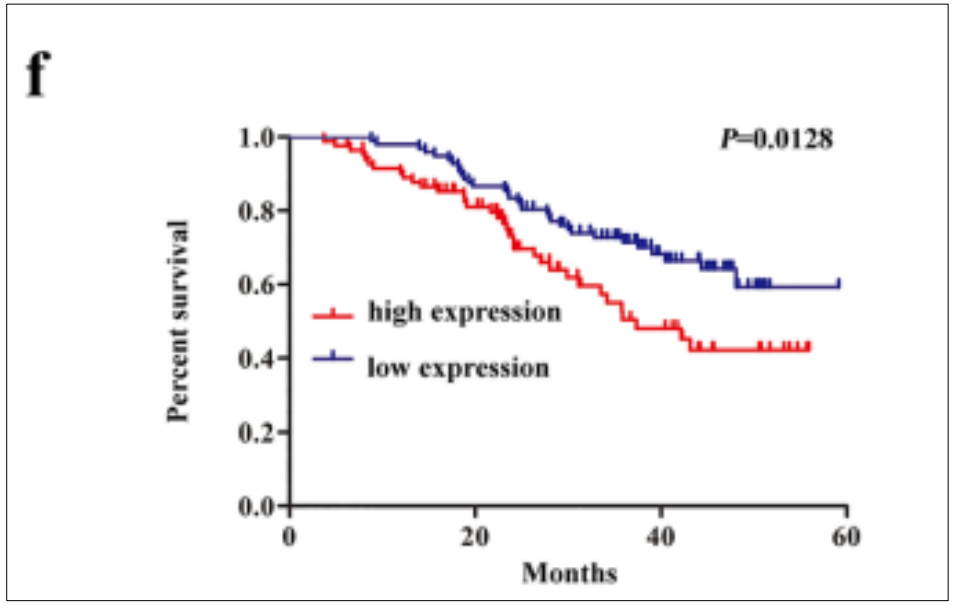
（5） Tau是时间常数，用与X轴相同的单位表示

（6） Span是Y0和Plateau之间的差值，用与Y值相同的单位表示

## 生存曲线差异分析

- 仅用于生存分析，由观察对象的生存时间和生存状态构成，1表示终点事件，0表示删失数据，注意分组数据的以“缺失数据”的形式进行表示
- 示例：文献Han J, Wang JZ, Yang X, Yu H, Zhou R, Lu HC, Yuan WB, Lu JC, Zhou ZJ, Lu Q, Wei JF, Yang H. METTL3 promote tumor proliferation of bladder cancer by accelerating pri-miR221/222 maturation in m6A-dependent manner. Mol Cancer. 2019 Jun 22;18(1):110. doi: 10.1186/s12943-019-1036-9. PMID: 31228940; PMCID: PMC6588935. Fig.1f

- 示例解读：
  - 图中实验是180例膀胱癌根据mettle3表达量的多少分为low 和high两组，对存活时间进行记录
  - 图中X轴作为生存时间，Y轴表示存活率




- 模拟数据

| Low expression | Survival time（months） | High expression | Survival time（months） |
|---|---|---|---|
| 1 | 17 | 1 | 10 |
| 1 | 38 | 1 | 3 |
| 1 | 26 | 1 | 10 |
| 1 | 34 | 1 | 6 |
| 1 | 36 | 1 | 27 |
| 1 | 25 | 1 | 7 |
| 1 | 20 | 1 | 18 |
| 1 | 14 | 1 | 26 |
| 1 | 27 | 1 | 18 |
| 1 | 20 | 1 | 14 |
| 1 | 15 | 1 | 16 |
| 1 | 29 | 1 | 22 |
| 1 | 18 | 1 | 15 |
| 1 | 20 | 1 | 25 |
| 1 | 42 | 1 | 26 |
| 1 | 26 | 1 | 32 |
| 1 | 14 | 1 | 40 |
| 1 | 36 | 1 | 34 |

## 生存曲线差异分析

| | Survival<br>Curve comparison | A | B |
|---|---|---|---|
| 1 | **Comparison of Survival Curves** | | |
| 2 | | | |
| 3 | **Log-rank (Mantel-Cox) test** | | |
| 4 | Chi square | 7.373 | |
| 5 | df | 1 | |
| 6 | P value | 0.0066 | |
| 7 | P value summary | ** | |
| 8 | Are the survival curves sig different? | Yes | |
| 9 | | | |
| 10 | **Gehan-Breslow-Wilcoxon test** | | |
| 11 | Chi square | 6.420 | |
| 12 | df | 1 | |
| 13 | P value | 0.0113 | |
| 14 | P value summary | * | |
| 15 | Are the survival curves sig different? | Yes | |
| 16 | | | |
| 17 | **Median survival** | | |
| 18 | Low expression | Undefined | |
| 19 | High expression | 43.00 | |
| 20 | | | |
| 21 | **Hazard Ratio (Mantel-Haenszel)** | **A/B** | **B/A** |
| 22 | Ratio (and its reciprocal) | 0.5589 | 1.789 |
| 23 | 95% CI of ratio | 0.3672 to 0.8506 | 1.176 to 2.723 |
| 24 | | | |
| 25 | **Hazard Ratio (logrank)** | **A/B** | **B/A** |
| 26 | Ratio (and its reciprocal) | 0.5619 | 1.780 |
| 27 | 95% CI of ratio | 0.3705 to 0.8523 | 1.173 to 2.699 |

GraphPad Prism在作生存曲线时，便自动对数据进行分析；对于生存曲线而言，在使用的时候更重要的是对数据的录入，相关内容的介绍在作图教程中有详细解答

- 给出了两种检验方法的结果：
- Log-rank检验（常用）：检验对不同生存时间点均取权重为“1”，即不考虑各观察时间点开始时存活的人数对统计模型的影响。相对重视远期效应
- Gehan-Breslow-Wilcoxon检验：在Log-rank检验的基础上增加了权重，按各个时间点存活的人数加权即开始存活人数相对多的时间点死亡情况的变化对整个模型的贡献较大，而开始存活人数相对少的时点死亡情况的变化对整个模型的贡献较小。对近期效应更加敏感。

  （这里不理解不要紧，只要记住结论即可）
- 两个方法的结论相同：可认为近期和远期都有差别（或无差别）
- 若Log-rank检验 $p$<0.05而Gehan-Breslow-Wilcoxon检验 $p$>0.05，可认为远期有差别而近期无差别，反之，近期有差别而远期无差别

## 小结：GraphPad Prism可简单便捷地实现科研数据处理

| 假设检验 | 相关分析 | 回归分析 |
| --- | --- | --- |
| U检验 | 皮尔森相关分析 | 线性回归 |
| T检验 | 斯皮尔曼方差分析 | 逻辑回归 |
| 卡方检验 | 典型相关分析 | 有序逻辑回归 |
| 方差分析 | | 多项逻辑回归 |


THE END

# GraphPad Prism 9 常用数据统计分析

## 科研技能 单元课 03 增刊

- 针对医生做基础科研过程中需要的特定科研技能，解螺旋制作了一系列专项学习突破的图文或视频教程，谓之“单元课”，此为第 3 期的增刊。

## 后续课程

## 敬请关注

意见反馈

搜索公众号：解螺旋 | 陪伴医生科研成长