《传播统计学》

独立样本t检验

教师：林志良
邮箱：linzhl@nfu.edu.cn
个人网站：www.zhilianglin.com

目录

- 独立样本t检验介绍
- 方差齐性
- 方差不齐
- 方差齐性检验：F检验
- 软件实操

独立样本t检验介绍

适用场景

- 比较两个总体均值是否有差异。涉及两个变量：其中一个变量是二分类型变量（自变量），另一个变量为数值型变量（因变量）
- 例如：
  - 比较男女生手游时间是否有**差异**。
  - 比较新闻专业学生和网新专业学生每天投入学习的时间是否有**差异**。
  - 比较党员同学和非党员学生学习成绩是否有**差异**。

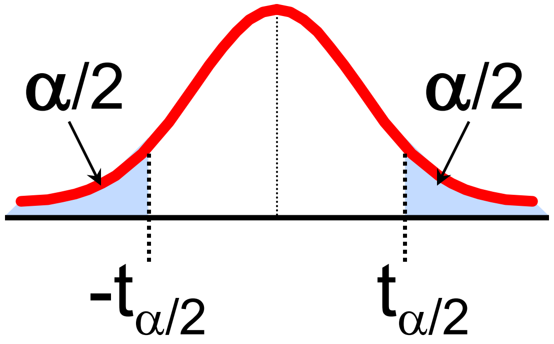
独立样本t检验介绍

假设检验

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

或

$H_0: \mu_1 - \mu_2 = 0$
$H_1: \mu_1 - \mu_2 \neq 0$



如果 $t < -t_{\alpha/2}$ 或 $t > t_{\alpha/2}$，拒绝H_0

独立样本t检验介绍

两种情形

检验两总体均值是否相等

σ_1 和 σ_2 未知,假设相等（**方差齐性**）

σ_1 和 σ_2 未知,假设不相等 （**方差不齐**）

怎么知道总体的方差是否相等?

做方差齐性检验，用F检验的统计方法

独立样本t检验介绍

单样本t检验公式： $t = \dfrac{\text{样本统计量} - \text{假设的总体参数}}{\text{估计标准误}}$

独立样本 t 检验公式：

$$t = \frac{\text{样本均值差异} - \text{总体均值差异}}{\text{估计标准误}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{SE_{(\bar{X}_1 - \bar{X}_2)}}$$

独立样本t检验的关键是如何得到均值差异的标准误。

方差齐性

方差齐性时

检验两总体均值是否相等

σ_1 和 σ_2 未知,假设相等（**方差齐性**）

σ_1 和 σ_2 未知,假设不相等 （**方差不齐**）

方差齐性

联合方差
(Pooled Variance)

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1{}^2 + (n_2 - 1)S_2{}^2}{(n_1 - 1) + (n_2 - 1)}$$

t值的自由度df=第一个样本的df+ 第二个样本的df

$= df_1 + df_2$

$= (n_1 - 1) + (n_2 - 1)$

$= n_1 + n_2 - 2$

例题

为了比较两种不同气味的阅读材料对学习效果的影响，心理学家随机选取20名被试，随机分成两组，准备了40组单词，一组阅读的材料是有香味的，另一组阅读的材料是没有香味的。以下数据为有香味组和无香味组能回忆起的单词数数据：

有香味组	19	20	24	30	31	32	30	27	22	25
无香味组	23	22	15	16	18	12	16	19	14	25

方差齐性

例题

基于以上数据，阅读有香味的阅读材料的记忆效果和阅读无香味的阅读材料的记忆效果有差异吗？

方差齐性

例题

H_0: $\mu_1 - \mu_2 = 0$ 或 ($\mu_1 = \mu_2$)

H_1: $\mu_1 - \mu_2 \neq 0$ 或 ($\mu_1 \neq \mu_2$)

$$\overline{X}_1 = 26, S_1 = 4.71$$

$$\overline{X}_2 = 18, S_2 = 4.21$$

方差齐性

例题

检验统计量为:

$$t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{(26-18)-0}{\sqrt{19.95\left(\frac{1}{10}+\frac{1}{10}\right)}}=\boxed{4}$$

$$S_p^2=\frac{(n_1-1)S_1{}^2+(n_2-1)S_2{}^2}{(n_1-1)+(n_2-1)}=\frac{(10-1)4.71^2+(10-1)4.21^2}{(10-1)+(10-1)}=19.95$$

例题

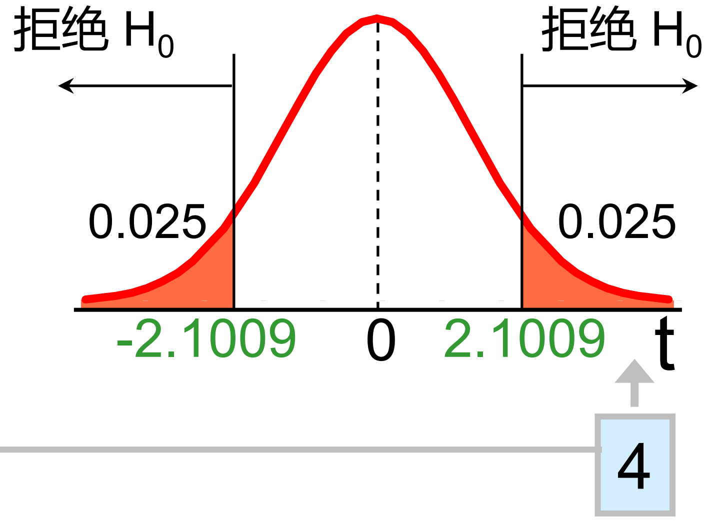
$H_0: \mu_1 - \mu_2 = 0$ 或 $(\mu_1 = \mu_2)$

$H_1: \mu_1 - \mu_2 \neq 0$ 或 $(\mu_1 \neq \mu_2)$

$\alpha = 0.05$

$df = 10 + 10 - 2 = 18$

临界值: $t = \pm 2.1009$



检验统计量: $t = \boxed{4}$

决策:

在 $\alpha = 0.05$的水平下拒绝H_0

结论:

两个均值有显著差异

方差不齐

方差不齐时

检验两总体均值是否相等

- σ1 和 σ2 未知,假设相等（**方差齐性**）
- σ_1 和 σ_2 未知,假设不相等 （**方差不齐**）

方差不齐

如果两总体的方差不相等，t统计量、自由度、临界值的计算公式都不同于方差相等的情况，以下为t'统计量及其自由度df的计算公式。（t'分布的临界值的计算公式略）

$$t' = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1^2} + \frac{S_2^2}{n_2^2}}}$$

$$df = \frac{\left(\frac{S_1^2}{n_1^2} + \frac{S_2^2}{n_2^2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1^2}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2^2}\right)^2}{n_2 - 1}}$$

方差不齐

例题：阅读有/无香味的阅读材料

$$t' = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1^2} + \frac{S_2^2}{n_2^2}}} = \frac{(26-18)-0}{\sqrt{\frac{4.71^2}{10^2} + \frac{4.21^2}{10^2}}} = 4$$

$$df = \frac{\left(\frac{S_1^2}{n_1^2} + \frac{S_2^2}{n_2^2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1^2}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2^2}\right)^2}{n_2 - 1}} = \frac{\left(\frac{4.71^2}{10^2} + \frac{4.21^2}{10^2}\right)^2}{\frac{\left(\frac{4.71^2}{10^2}\right)^2}{10 - 1} + \frac{\left(\frac{4.21^2}{10^2}\right)^2}{10 - 1}} = 17.78$$

方差齐性检验

F检验统计量

检验两总体方差

F 检验统计量

假设	**F**
$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$	S_1^2 / S_2^2

其中:

S_1^2 = 样本1的方差（样本方差更大）

n_1 = 样本1的样本量

S_2^2 =样本2的方差（样本方差更小）

n_2 =样本2的样本量

$n_1 - 1$ = 自由度1（分子）

$n_2 - 1$ = 自由度2（分母）

方差齐性检验

F分布

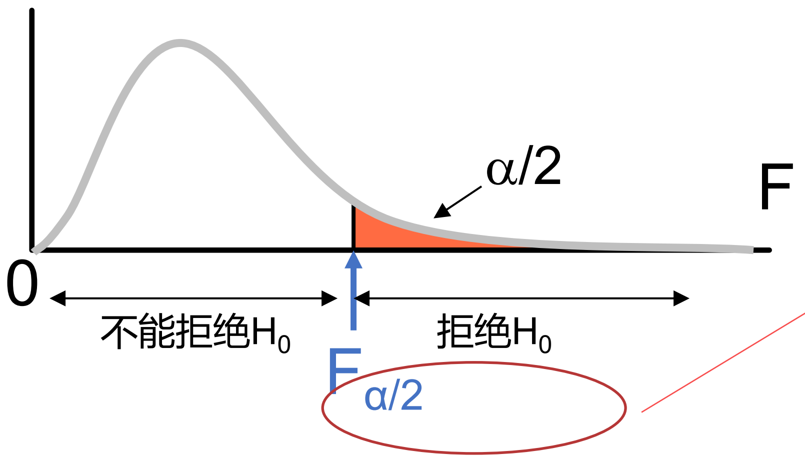
- F的临界值可通过查询F分布表得到
  - 临界值有两个，两者的关系是互为倒数
  - 我们查表一般查看大于1的那个临界值

- F分布有两个自由度

方差齐性检验

F分布

H_0: $\sigma_1^2 = \sigma_2^2$
H_1: $\sigma_1^2 \neq \sigma_2^2$



注意：这里是α/2，因为图形左侧也有拒绝域和临界值的（只是不画出来）

如果$F > F_{\alpha/2}$，拒绝 H_0

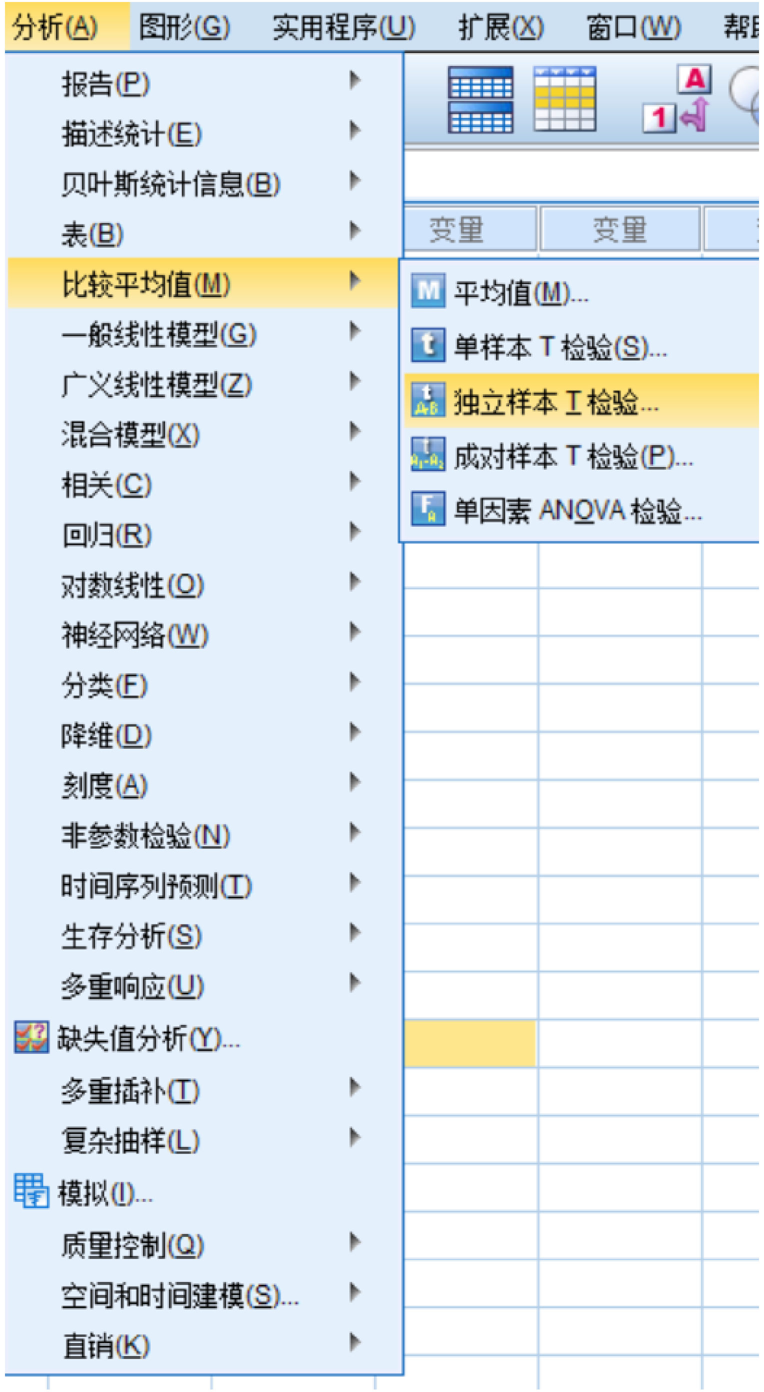
软件实操

注意：在SPSS上实操时，需要将数据整理成右图的形式（整洁的数据形式，tidy data）

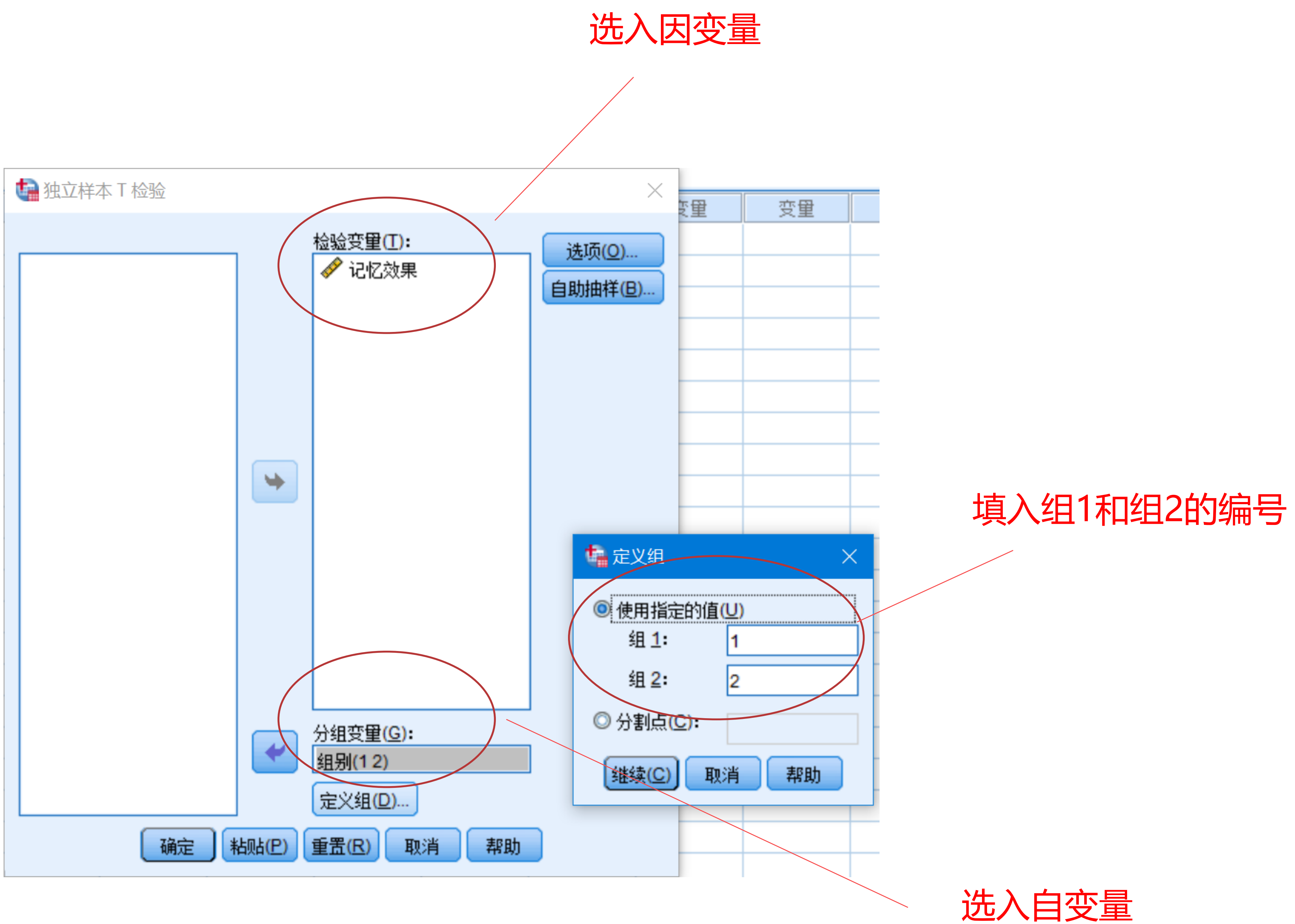
- 每一列是一个变量，每一行是一个被观测对象，每个单元格为具体的观测值

记忆效果	组别
19.00	1.00
20.00	1.00
24.00	1.00
30.00	1.00
31.00	1.00
32.00	1.00
30.00	1.00
27.00	1.00
22.00	1.00
25.00	1.00
23.00	2.00
22.00	2.00
15.00	2.00
16.00	2.00
18.00	2.00
12.00	2.00
16.00	2.00
19.00	2.00
14.00	2.00
25.00	2.00

软件实操


分析(A)
图形(G)
实用程序(U)
扩展(X)
窗口(W)
报告(P)
描述统计(E)
贝叶斯统计信息(B)
表(B)
比较平均值(M)
一般线性模型(G)
广义线性模型(Z)
混合模型(X)
相关(C)
回归(R)
对数线性(O)
神经网络(W)
分类(F)
降维(D)
刻度(A)
非参数检验(N)
时间序列预测(T)
生存分析(S)
多重响应(U)
缺失值分析(Y)...
多重插补(T)
复杂抽样(L)
模拟(I)...
质量控制(Q)
空间和时间建模(S)...
直销(K)
变量
变量
平均值(M)...
单样本 T 检验(S)...
独立样本 T 检验...
成对样本 T 检验(P)...
单因素 ANOVA 检验...

软件实操


选入因变量
独立样本 T 检验
检验变量(T):
记忆效果
选项(O)...
自助抽样(B)...
分组变量(G):
组别(1 2)
定义组(D)...
确定
粘贴(P)
重置(R)
取消
帮助
变量
定义组
使用指定的值(U)
组 1:
1
组 2:
2
分割点(C):
继续(C)
取消
帮助
填入组1和组2的编号
选入自变量

软件实操

组统计

	组别	个案数	平均值	标准 偏差	标准 误差平均值
记忆效果	1.00	10	26.0000	4.71405	1.49071
	2.00	10	18.0000	4.21637	1.33333

独立样本检验

		莱文方差等同性检验		平均值等同性 t 检验						
									差值 95% 置信区间	
		F	显著性	t	自由度	Sig.（双尾）	平均值差值	标准误差差值	下限	上限
记忆效果	假定等方差	.384	.543	4.000	18	.001	8.00000	2.00000	3.79816	12.20184
	不假定等方差			4.000	17.780	.001	8.00000	2.00000	3.79443	12.20557

方差齐性检验的p值，如果p<0.05，看第二行结果；如果p>0.05，则看第一行结果

注：SPSS计算出来的F值的方法不同于我们上面讲的方法

软件实操

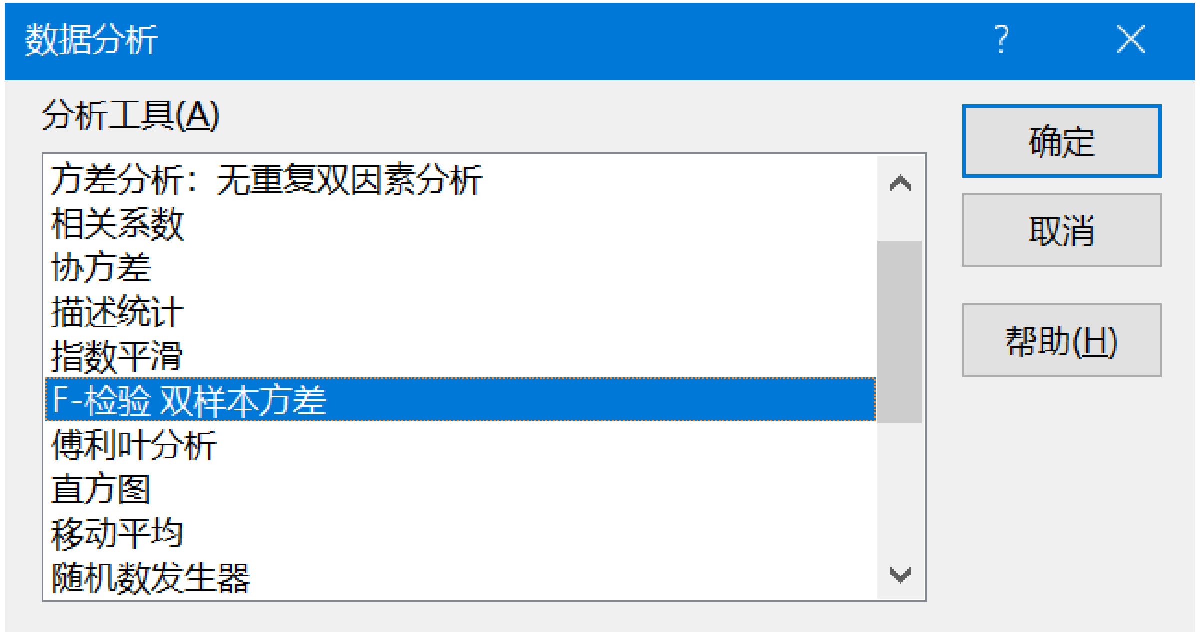
注：Excel的F检验用的是单侧检验，这在判断“方差齐性”问题上不是很常见。以下操作仅供参考，建议以SPSS为准。

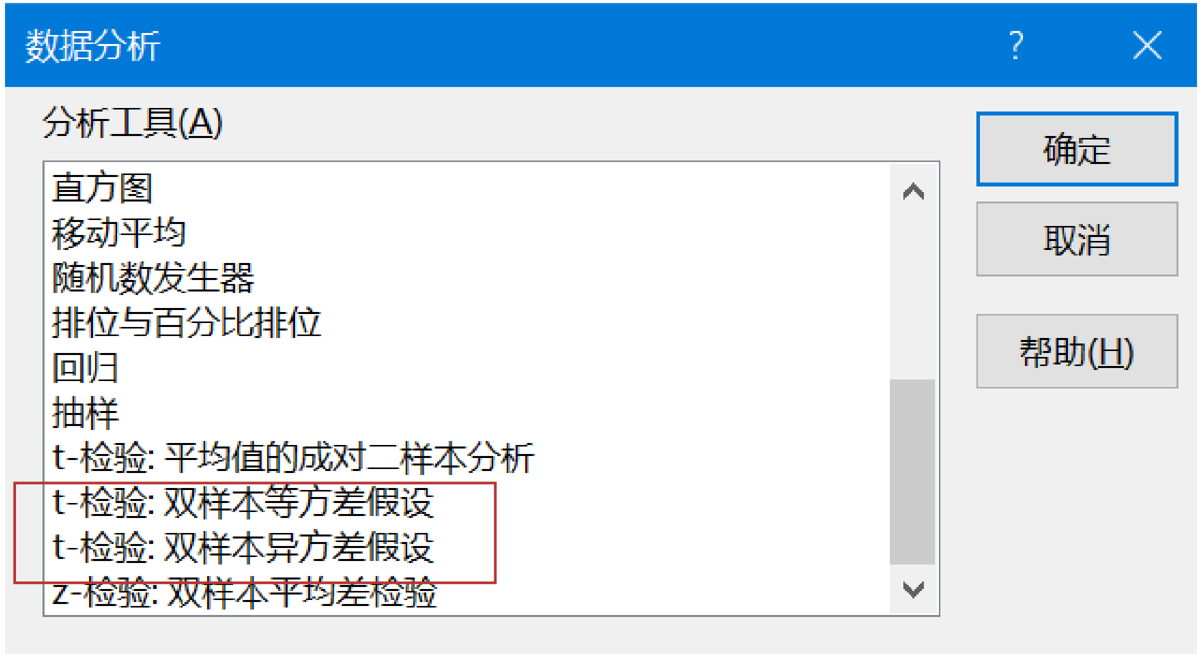
软件实操

【工具】>【数据分析】>【 F-检验 双样本方差分析】>选择数据范围

如果F检验的结果是p<0.05，则选择异方差；如果p>0.05，选择同方差

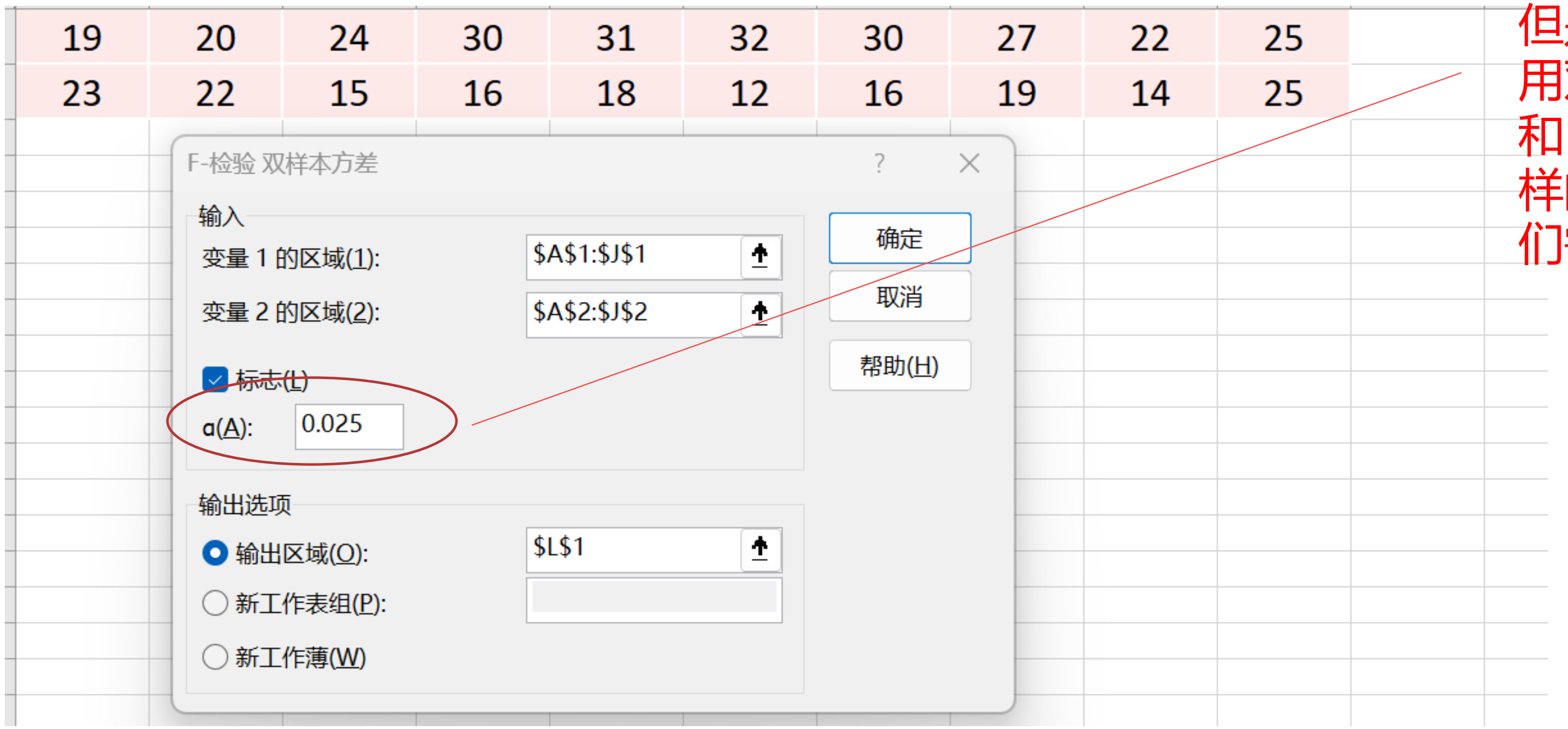
【工具】>【数据分析】>【 t检验：双样本等（异）方差假设】>选择数据范围





软件实操

F检验



Excel用的是单侧检验，但是我们一般更多的是用双侧检验。为了得到和 α=0.05 双边检验一样的结果，这里的α我们需要设置成0.025

软件实操

F检验

F检验

F-检验 双样本方差分析		
	变量 1	变量 2
平均	26	18
方差	22.22222	17.77778
观测值	10	10
df	9	9
F	1.25	
P(F<=f) 单尾	0.372501	
F 单尾临界	3.178893	

方差齐性检验，如果 $p<0.05$，使用【t检验：双样本异方差假设】；如果 $p>0.05$，则使用【t检验：双样本等方差假设】

软件实操

t检验：双样本等方差假设

t-检验: 双样本等方差假设

	变量 1	变量 2
平均	26	18
方差	22.22222	17.77778
观测值	10	10
合并方差	20	
假设平均差	0	
df	18	
t Stat	4	
P(T<=t) 单尾	0.00042	
t 单尾临界	1.734064	
P(T<=t) 双尾	0.00084	
t 双尾临界	2.100922	

我们一般看双尾的p值

谢谢！