
SPSS教程：配对样本t检验详解

作者：writer 来源：医咖会

本文原地址：<https://www.iikx.com/news/statistics/18046.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

SPSS教程：配对样本t检验详解。

一、问题与数据

研究者想验证一种新型运动饮料配方是否有助于提高人们的跑步距离。传统饮料配方为纯碳水化合物，而新型饮料为碳水化合物-蛋白质混合物。

为了比较两种运动饮料对人们跑步距离的影响差异，研究者招募了20名受试者，每人进行2项试验，每项试验受试者均在跑步机上运动2小时。2项试验中，同一受试者跑步前分别喝含纯碳水化合物饮料和碳水化合物-蛋白质混合饮料。同时，均衡所有受试者进行2项试验的先后顺序，使一半人先喝纯碳水化合物饮料，另一半人先喝碳水化合物-蛋白质混合饮料，分别记录其跑步距离。

碳水化合物饮料组的跑步距离记为carb变量，碳水化合物-蛋白质饮料组的跑步距离记为carb_protein变量。研究者想知道，是否2组的跑步距离有差异，即2种运动饮料对人们跑步距离的影响不同。从变量层面上，也就是看是否carb变量和carb_protein变量的均数存在差异(部分数据如下图)。

二、对问题的分析.

研究者想探索是否2个相关(配对)组别间的因变量均数存在差异，可以使用配对样本t检验。使用配对样本t检验时，需要考虑4个假设：

假设1：因变量为连续变量;

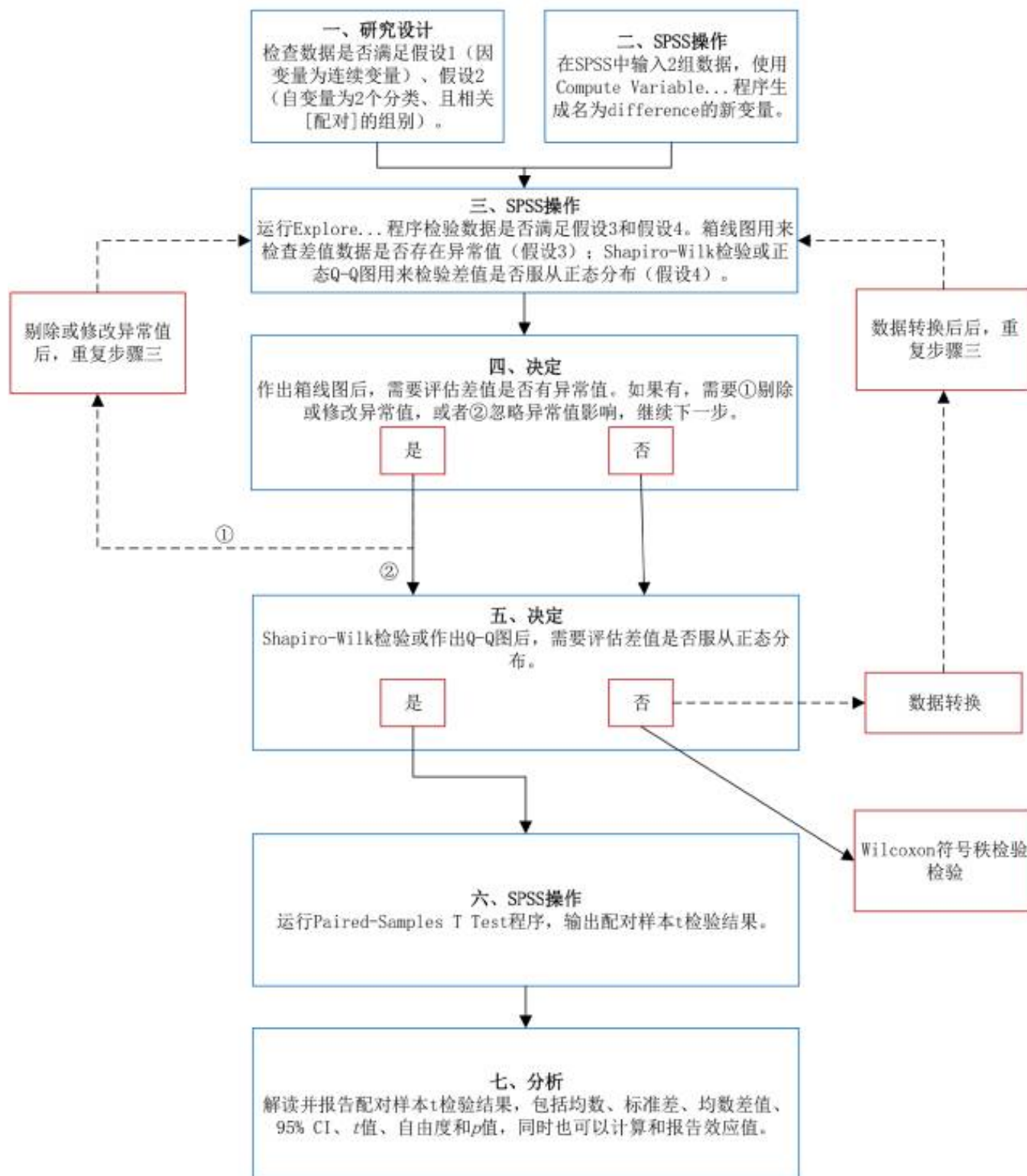
假设2：自变量包含2个分类、且相关(配对)非独立的组别;

假设3：2个相关(配对)组别间的因变量差值没有明显异常值;

假设4：2个相关(配对)组别间的因变量差值近似服从正态分布。

那么进行配对样本t检验时，如何考虑和处理这4个假设呢?

三、思维导图.



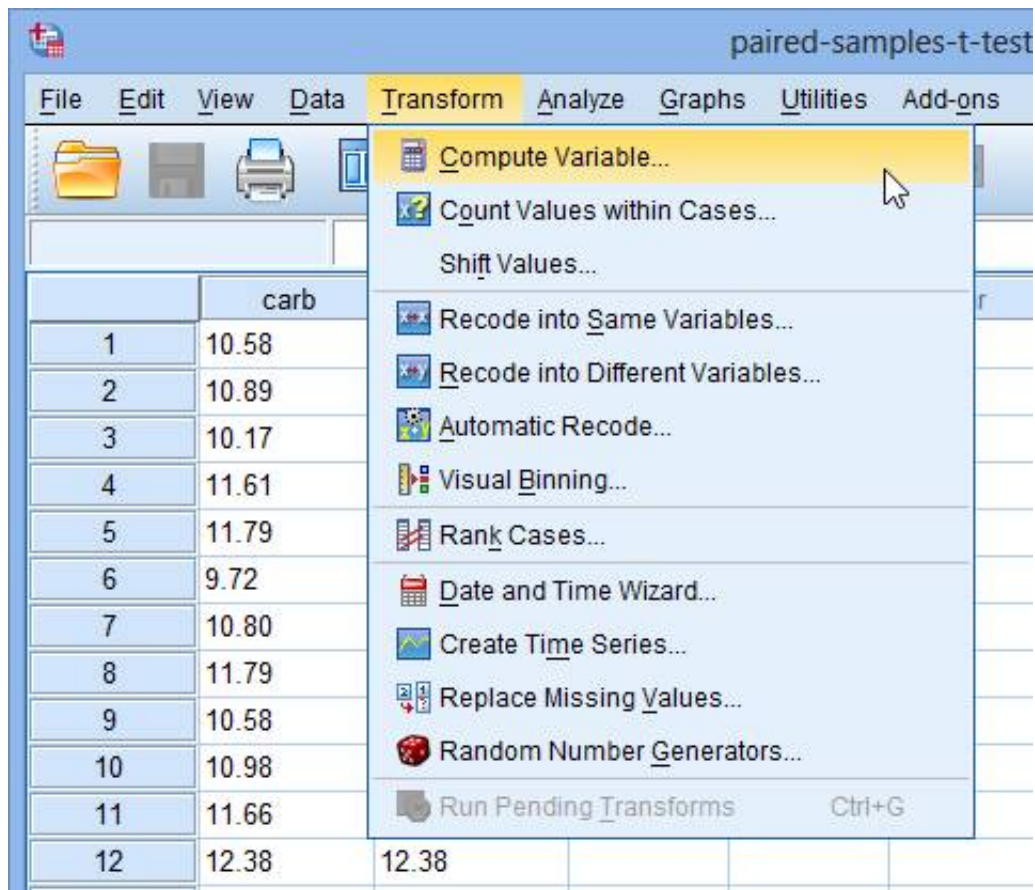
四、对假设的判断.

假设1：因变量为连续变量;假设2：自变量包含2个分类、且相关(配对)非独立的组别。和研究设计有关，需要根据实际情况进行判断。

假设3：2个相关(配对)组别间的因变量差值没有明显异常值。

对于配对样本t检验，异常值和正态性的假设检验都是基于2组间配对数值的差值进行的。因此，我们首先需要计算2组因变量的差值，并把它作为一个新变量储存，变量名为difference，具体操作如下：

1.在主菜单栏中点击Transform > Compute Variable...：

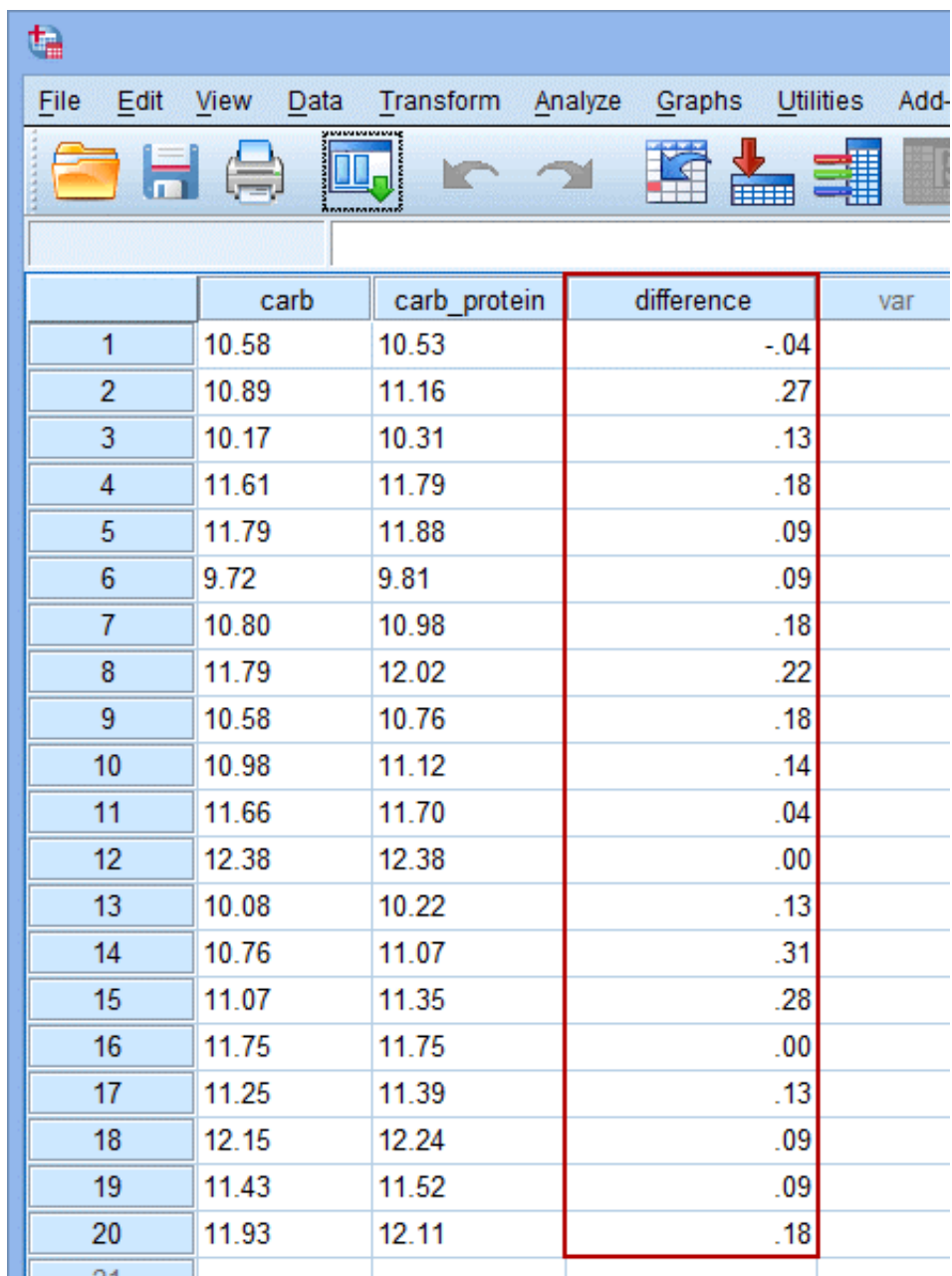


出现Compute Variable对话框：

2.在Target Variable:模块中输入difference, 即为新创建的变量名;在Numeric Expression:模块中输入carb_protein - carb, 即为2个配对组别间的因变量差值(也可以直接从左侧中部变量框中挑选变量进入Numeric Expression:模块, 并选择中间的运算符号和数字进行运算):

本例为用carb_protein变量值减去carb变量值，此顺序与研究设计和研究目的有关，通常用实验组的数值减去对照组的数值。本例关心的是新型运动饮料相比于传统运动饮料，是否可以提高跑步距离，因此传统碳水化合物饮料组应该作为对照组。如果2组差值为正数，则可以反映新型饮料有助于提高人们的跑步距离。

3.点击OK，返回Data View窗口，即可见到新变量difference：

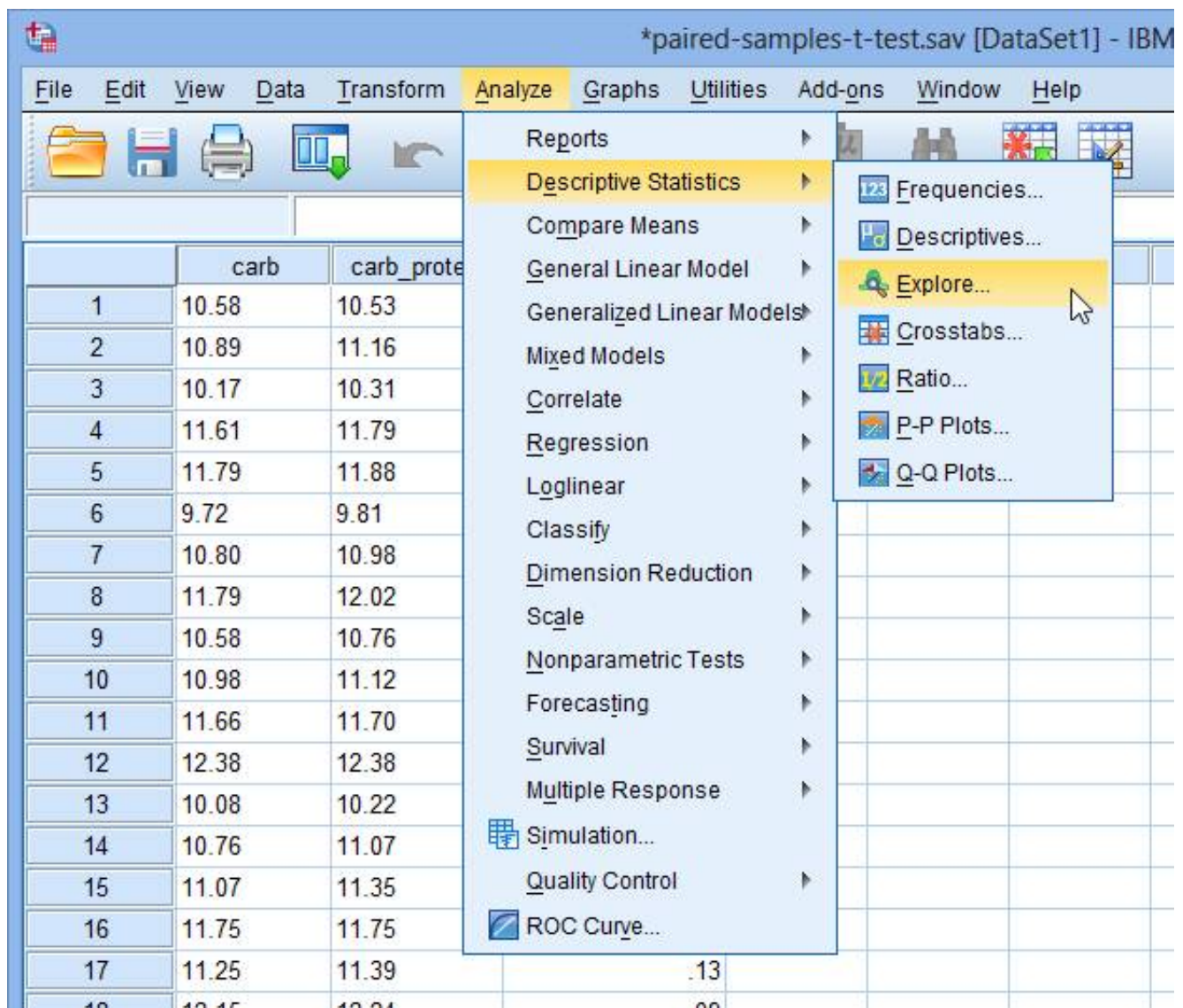


	carb	carb_protein	difference	var
1	10.58	10.53	-.04	
2	10.89	11.16	.27	
3	10.17	10.31	.13	
4	11.61	11.79	.18	
5	11.79	11.88	.09	
6	9.72	9.81	.09	
7	10.80	10.98	.18	
8	11.79	12.02	.22	
9	10.58	10.76	.18	
10	10.98	11.12	.14	
11	11.66	11.70	.04	
12	12.38	12.38	.00	
13	10.08	10.22	.13	
14	10.76	11.07	.31	
15	11.07	11.35	.28	
16	11.75	11.75	.00	
17	11.25	11.39	.13	
18	12.15	12.24	.09	
19	11.43	11.52	.09	
20	11.93	12.11	.18	
21				

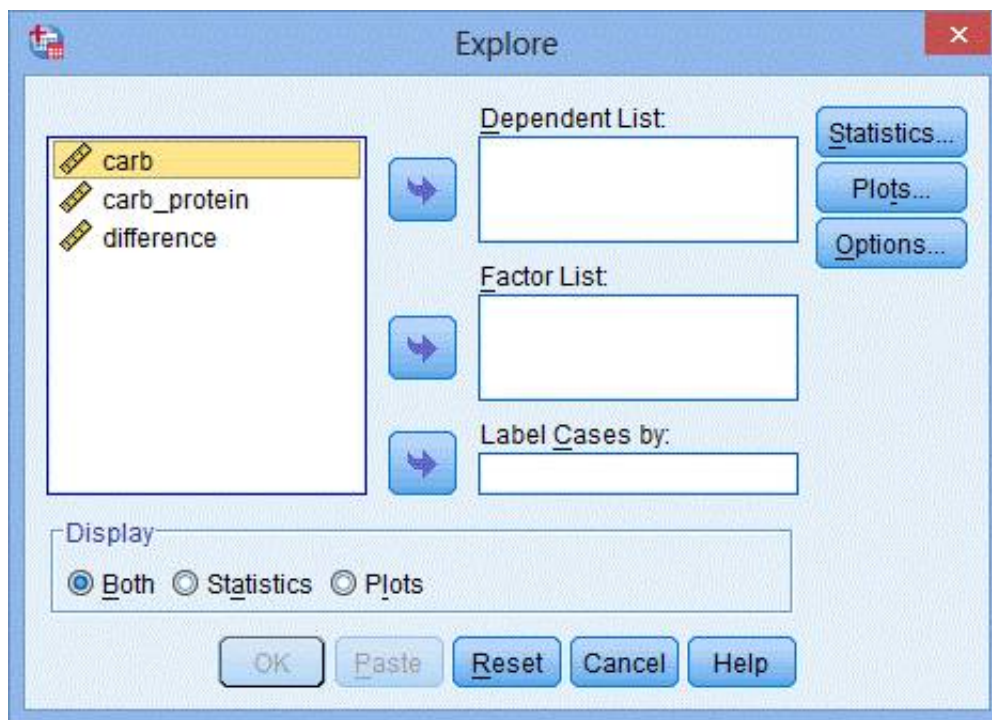
如果差值中的某些取值和其他值相比特别大或者特别小，则称之为异常值。异常值会影响差值组的均数和标准差，因此可能会对最终的统计结果产生很大的负面影响。对于小样本研究，异常值的影响尤其显著，必须检查差值组中是否存在明显异常值。

以下操作将说明如何在SPSS软件中利用Explore...程序检查异常值，以及检验数据是否服从正态分布：

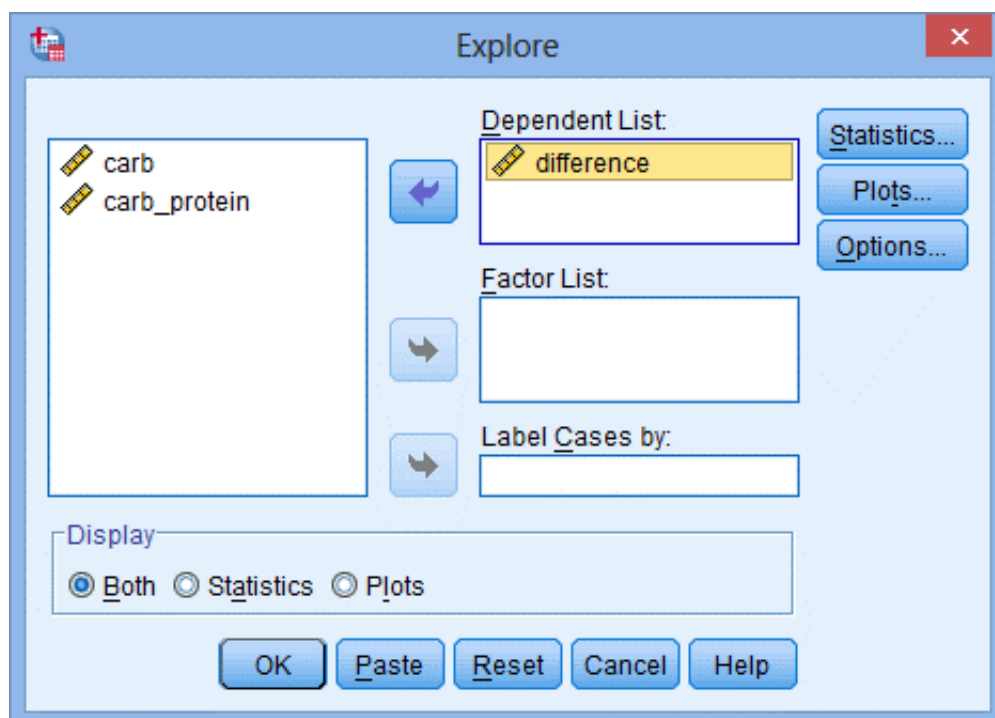
1.在主菜单栏中点击Analyze > Descriptive Statistics > Explore...：



出现Explore对话框：



2.把变量difference送入Dependent List模块中：

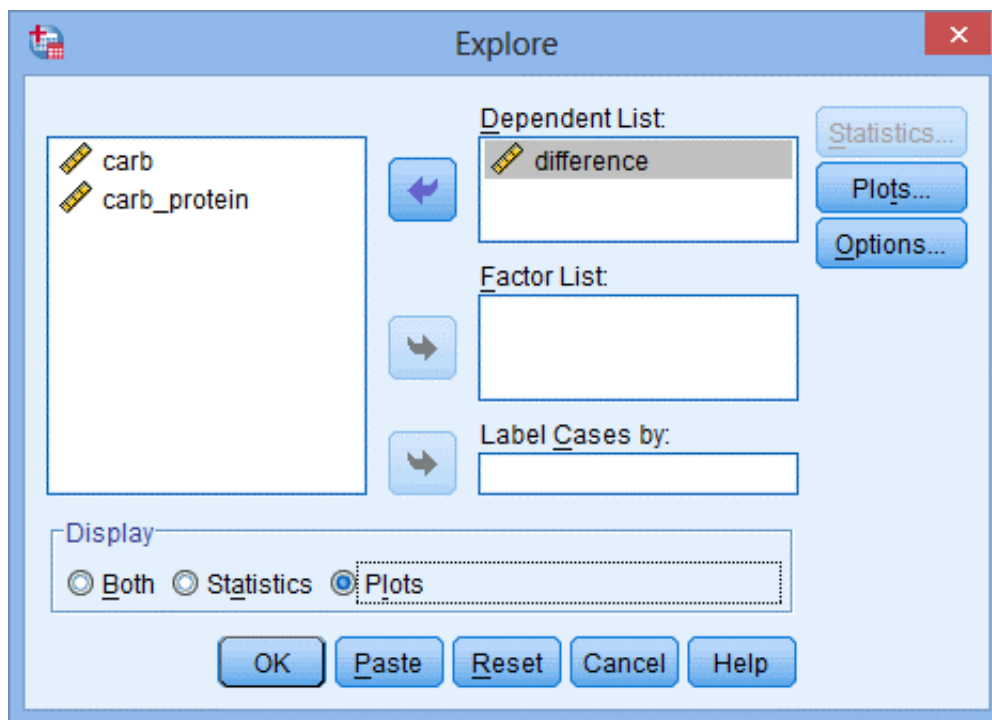


3.点击Plots..., 出现Explore: Plots对话框：

4.在Boxplots模块内保留系统默认选项Factor levels together，在Descriptive模块内取消选择Stem-and-leaf，在下方勾选Normality plots with tests：

5. 点击Continue, 返回Explore对话框;

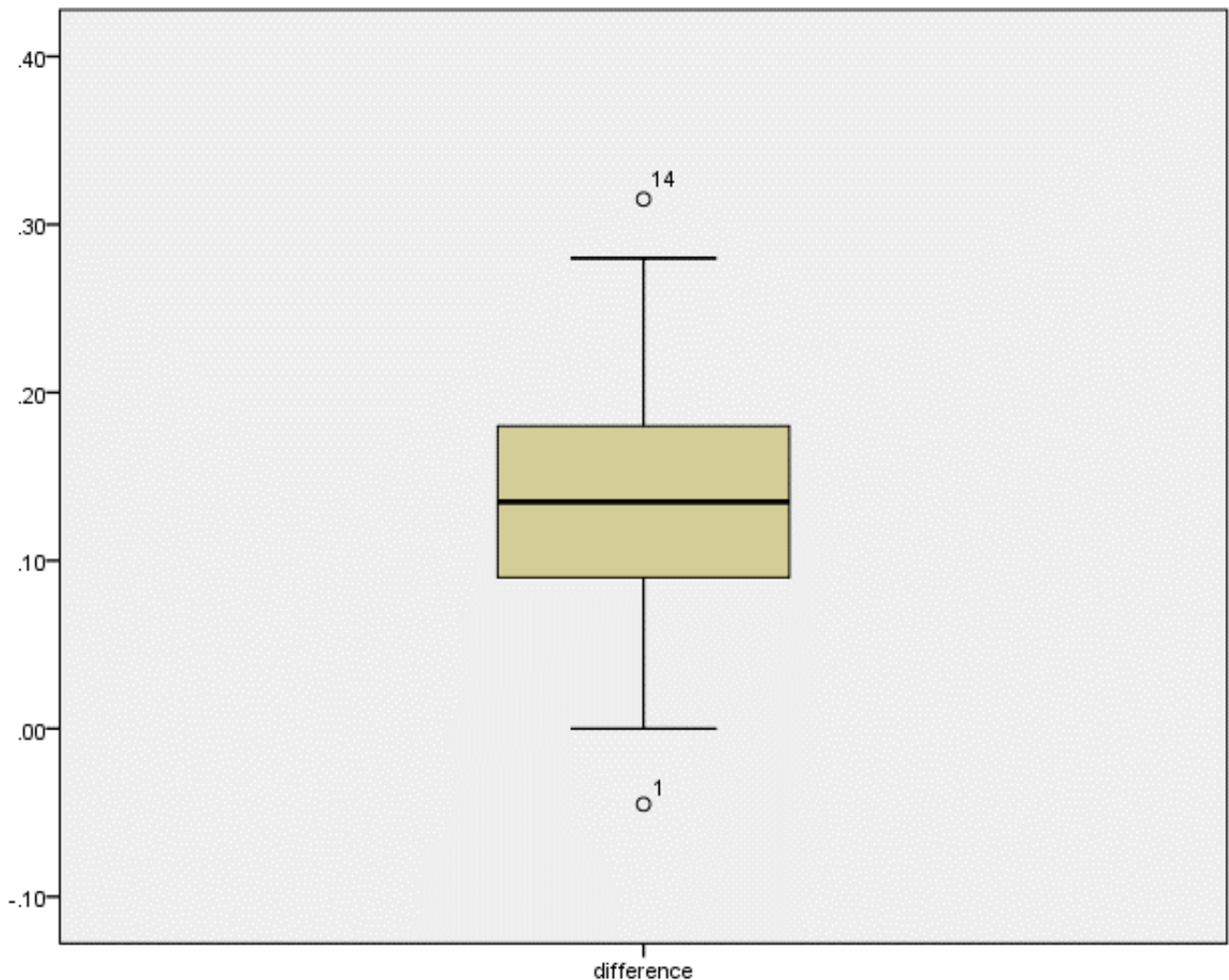
6. 在Display模块内点击Plots :



如果使用偏度和峰度(skewness and kurtosis)进行正态性判断，则保留Display模块内的默认选项Both或者选择Statistics。

7.点击OK，输出结果。

根据如下输出的箱线图，判断数据中是否存在异常值：



SPSS中将距离箱子边缘超过1.5倍箱身长度的数据点定义为异常值，以圆点(°)表示;距离箱子边缘超过3倍箱身长度的数据点定义为极端值(极端异常值)，以星号(*)表示。为容易识别，异常值均用其在Data View窗口的行数标出。

本例中，第1行(差值特别小)和第14行(差值特别大)的差值均为异常值，但是由于它们并非极端异常值，不会对2组均数差异产生过大影响，因此我们在接下来的分析中仍将其保留。

导致数据中存在异常值的原因通常有3种：

- 1.数据录入错误：首先应该考虑异常值是否由于数据录入错误所致。如果是，用正确值进行替换，并重新计算差值、重新进行所有检验;
- 2.测量误差：如果不是由于数据录入错误，接下来考虑是否因为测量误差导致(如仪器故障或超过量程)。通常情况下，大多数的测量误差是不可校正的;
- 3.真实的异常值：如果以上两种原因都不是，那最有可能源于真实的异常数据。这类异常值不好处理，但也没有理由将其当作无效值对待。目前它的处理方法比较有争议，尚没有一种特别推荐的方法。接下来，我们列举几种异常值的处理方法，供读者参考。

异常值的处理方法通常有2种：

1. 保留异常值：

- 1) 采用非参数Wilcoxon符号秩检验或符号检验;
- 2) 用非最极端的值(如第二大的值)来代替极端异常值;
- 3) 转换变量形式;
- 4) 将异常值纳入分析，并认为其对结果不会产生实质影响(比较有、无异常值的配对样本t检验结果)。

2. 剔除异常值：

我们也可以直接剔除异常值进行分析，但是需要提供所剔除异常点的信息，以便读者了解剔除的原因及其会对结果产生怎样的影响。

假设4：2个相关(配对)组别间的因变量差值近似服从正态分布。

正态性检验有很多方法，这里介绍最常用的2种方法：Shapiro-Wilk正态性检验和正态Q-Q图(其他还有偏度、峰度和直方图等)。

在假设3的判断中，我们在Explore: Plots对话框中勾选了Normality plots with tests，输出结果中会给出正态性检验的结果。

1. Shapiro-Wilk正态性检验.

如果样本量较小(<50)，或者研究者对正态Q-Q图以及其他图形方法的结果诠释不够有把握，推荐采用Shapiro-Wilk正态性检验。本例的Shapiro-Wilk检验结果如下：

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
difference	.120	20	.200 [*]	.971	20	.780

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

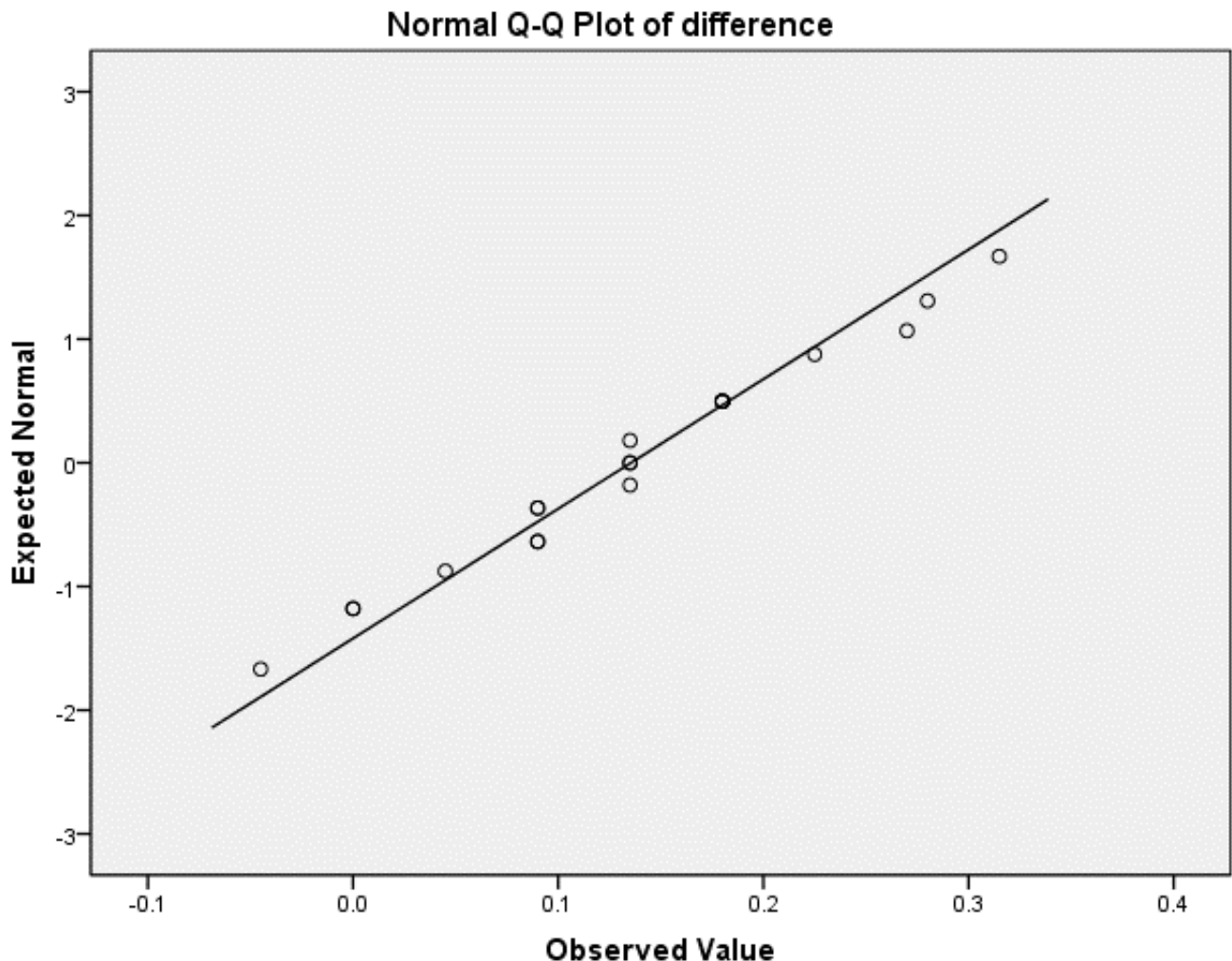
如果数据服从正态分布，显著性水平(Sig., 即P值)应该 ≥ 0.05 ;反之，P会 < 0.05 。Shapiro-Wilk检验的无效假设是数据服从正态分布，备择假设是数据不服从正态分布。因此，如果拒绝无效假设($P < 0.05$)，表示数据不服从正态分布;如果不能拒绝无效假设，则不能认为数据不服从正态分布。本例中 $P = 0.780$ ，因此不能认为2组差值不服从正态分布。

如果样本量大于50，推荐使用正态Q-Q图等图形方法进行正态判断。因为当样本量较大时，Shapiro-

Wilk检验会把稍稍偏离正态分布的数据也判断为有统计学差异，即认为数据不服从正态分布。

2. 正态Q-Q图.

对正态Q-Q图的直接观察可以更好地了解数据是否服从正态分布，但是不推荐对小样本数据采用Q-Q图进行正态性判断。本例差值的正态Q-Q图如下：



如果正态Q-Q图中的数值大致靠近图中的斜线分布，则可以认为服从正态分布;如果数值点并不是很好地沿着斜线分布，或者呈现不同的分布模式，则数据不服从正态分布。本例中差值的数据点大致沿着Q-Q图的斜线分布，可以认为2组的差值服从正态分布。

如果数据不服从正态分布，有如下4种方法进行处理：

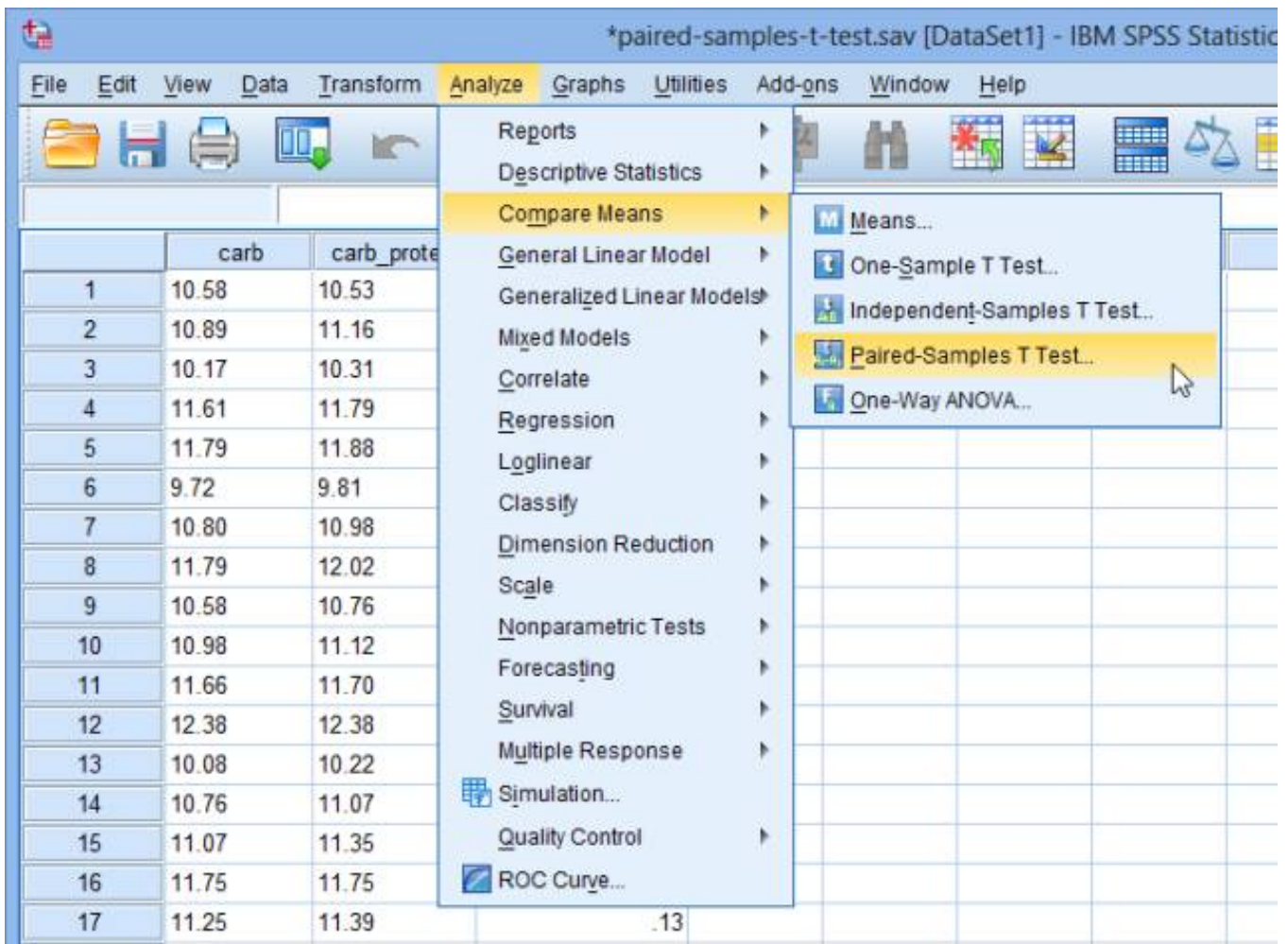
- 1.数据转换：对转换后呈正态分布的数据进行配对样本t检验，而且要对转换后的数据重新进行各种检验。对于一些常见的分布，有特定的转换形式，但是对于转换后数据的结果解释可能比较复杂;
- 2.使用非参数检验：可以使用Wilcoxon符号秩检验或符号检验等非参数检验方法;
- 3.直接进行分析：配对样本t检验对于稍偏离正态分布的数据比较稳健，而且非正态分布实质上并

不影响犯I型错误的概率。因此可以直接进行检验，但是结果中仍需报告对正态分布的偏离程度。

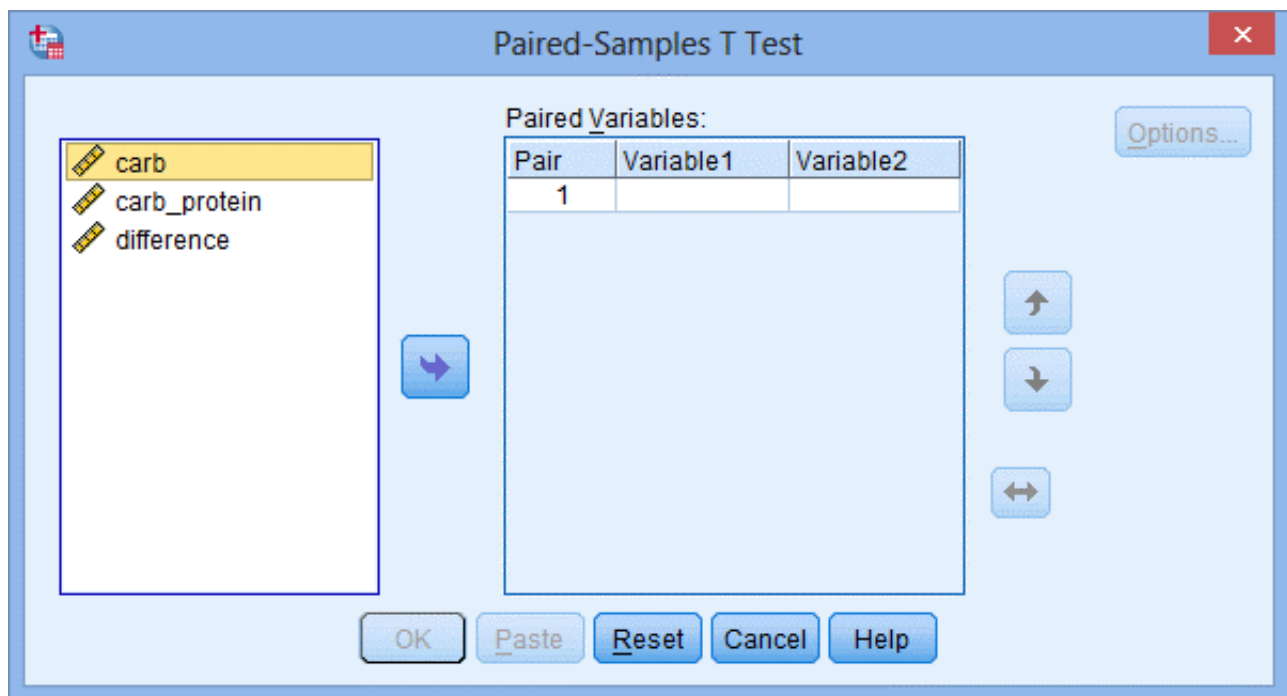
4.检验结果的比较：将转换后和未转换的原始数据分别进行配对样本t检验，并比较两者的结果；如果结论相同，则选择未转换的原始数据进行分析。

五、SPSS操作.

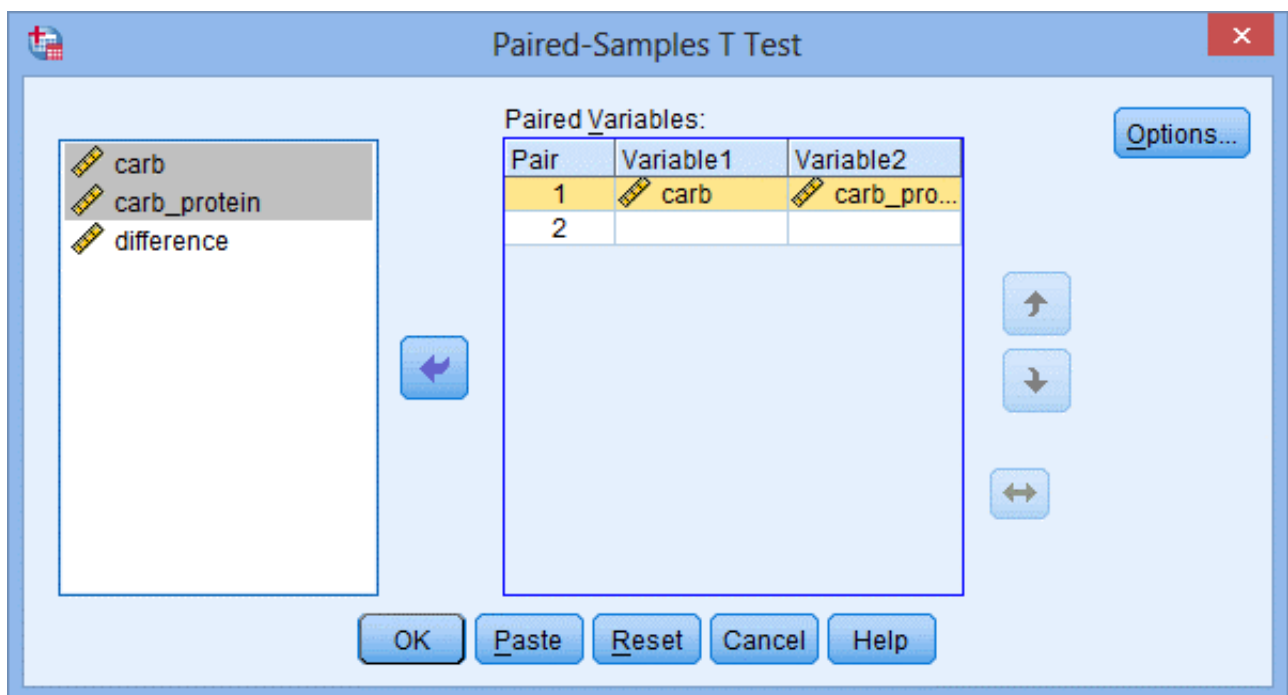
1.在主菜单栏中点击Analyze > Compare Means > Paired-Samples T Test...：



出现Paired-Samples T Test对话框：

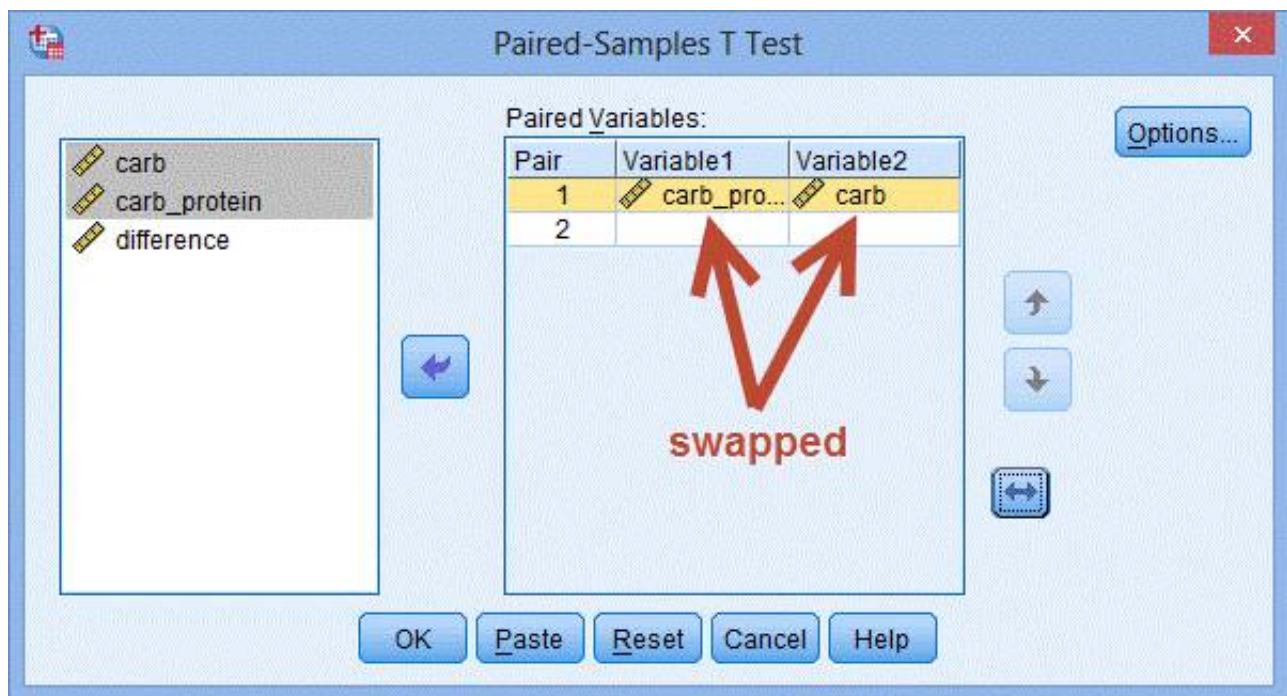


2.把变量carb和carb_protein送入Paired Variables:模块中(可以先后送入，也可以先选择一个变量后，按住shift键，再选择另一个变量同时送入)：

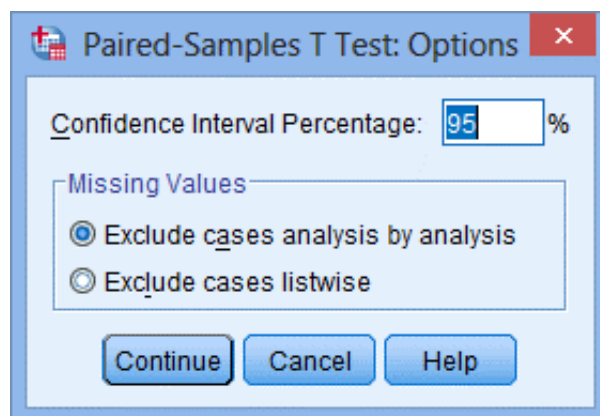


SPSS软件中配对样本t检验的差值是用Variable1减去Variable2，此步骤即用carb变量值减去carb_protein变量值。但是，我们之前提过，这里最好用carb_protein减去carb。

3.点击Paired Variables:模块中的黄色区域，激活右下部“双向箭头”按钮并点击，将会把carb变量和carb_protein变量的位置互换：



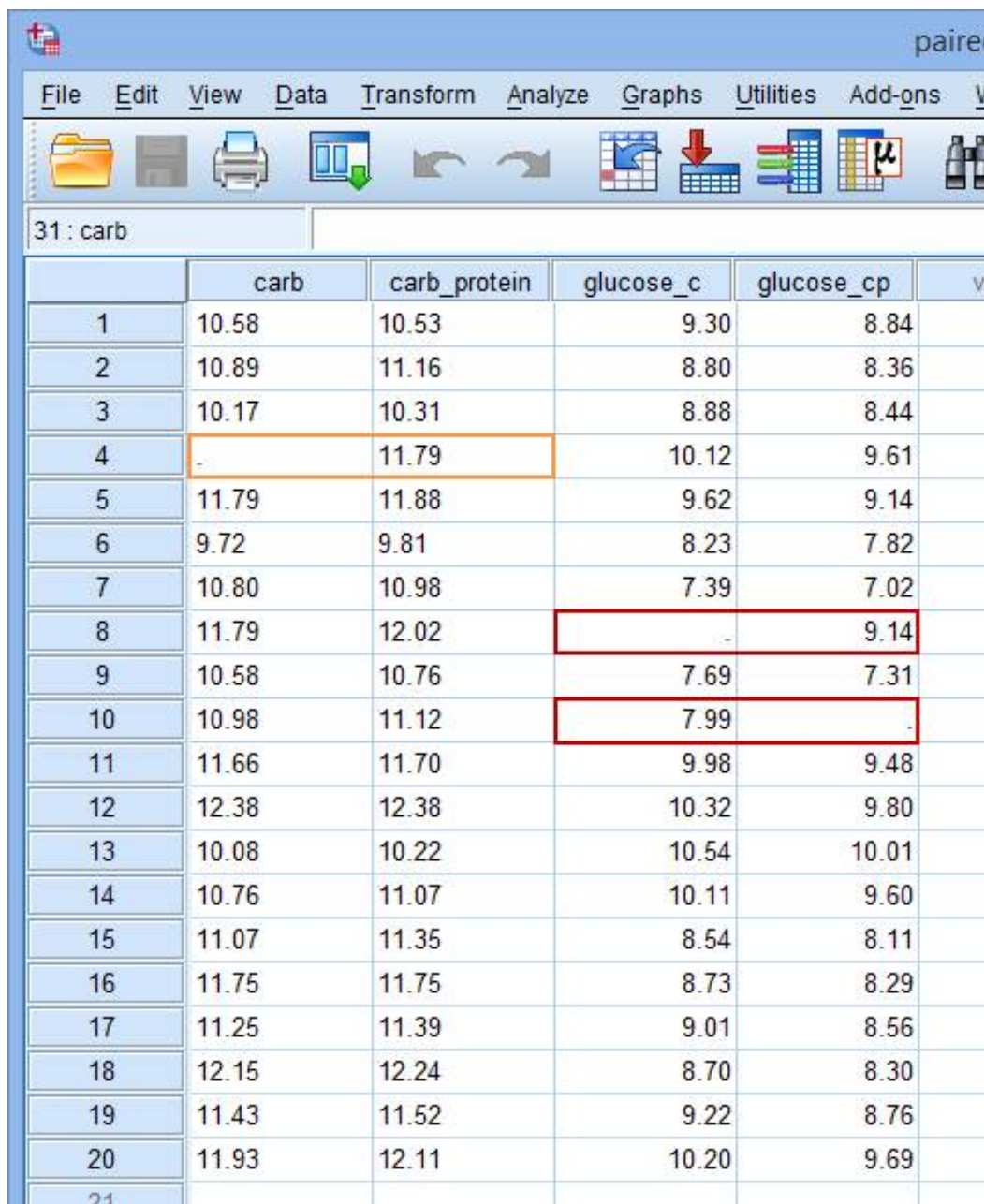
4. 点击Option...，出现Paired-Samples T Test: Options对话框：



保留系统默认的置信区间(95%)和缺失值选项(Exclude cases analysis by analysis)。

当我们只需要进行1次配对样本t检验时(如本例)，在Missing Values模块中选择Exclude cases analysis by analysis或者Exclude cases listwise时，结果是一致的。

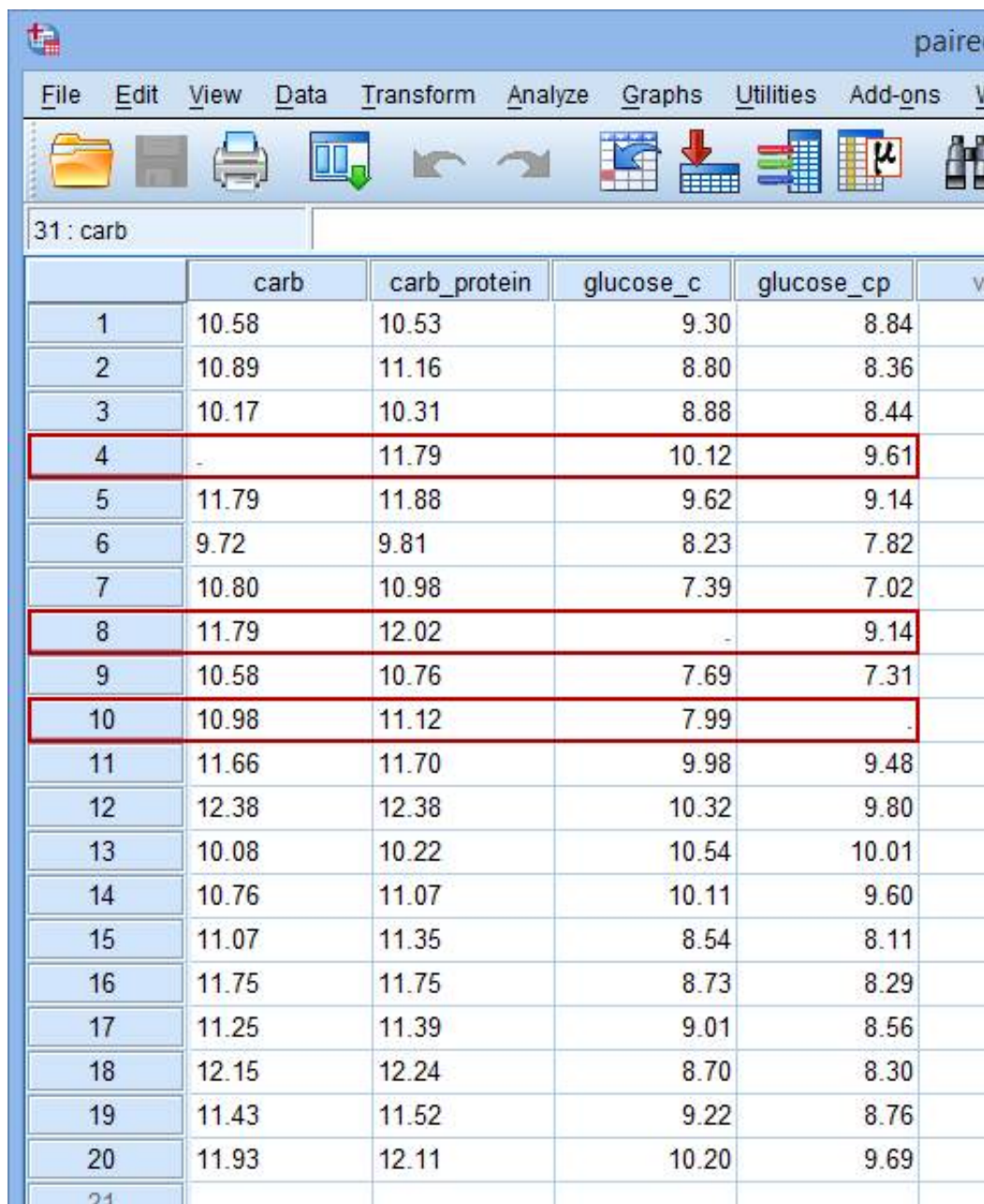
当我们的数据中要进行多次配对样本t检验时，Exclude cases analysis by analysis选项会分别剔除每次配对样本t检验的缺失值。比如下面的数据，SPSS软件将对19对的carb和carb_protein变量值进行配对样本t检验，排除第4行缺失的数据(橘色);而对glucose_c和glucose_cp变量的18对配对值进行配对样本t检验，排除第8行和第10行缺失的数据(红色)。



	carb	carb_protein	glucose_c	glucose_cp	v
1	10.58	10.53	9.30	8.84	
2	10.89	11.16	8.80	8.36	
3	10.17	10.31	8.88	8.44	
4	-	11.79	10.12	9.61	
5	11.79	11.88	9.62	9.14	
6	9.72	9.81	8.23	7.82	
7	10.80	10.98	7.39	7.02	
8	11.79	12.02	-	9.14	
9	10.58	10.76	7.69	7.31	
10	10.98	11.12	7.99	-	
11	11.66	11.70	9.98	9.48	
12	12.38	12.38	10.32	9.80	
13	10.08	10.22	10.54	10.01	
14	10.76	11.07	10.11	9.60	
15	11.07	11.35	8.54	8.11	
16	11.75	11.75	8.73	8.29	
17	11.25	11.39	9.01	8.56	
18	12.15	12.24	8.70	8.30	
19	11.43	11.52	9.22	8.76	
20	11.93	12.11	10.20	9.69	
21					

Exclude cases analysis by analysis选项1次检验中的缺失值并不影响其他检验，这样会使每次分析配对数量最大化，但是也会导致每次配对样本t检验的样本量有差异。而Exclude cases listwise选项会使用所有分析、检验中无缺失值的样本，这样虽然会导致样本量的大幅下降，但也会保证所有分析的样本量一致。

比如上面的数据，在进行2次配对样本t检验时，SPSS软件就会剔除掉所有的缺失数据(第4、8、10行，红色)，最后仅对17个样本进行所有的检验：



	carb	carb_protein	glucose_c	glucose_cp
1	10.58	10.53	9.30	8.84
2	10.89	11.16	8.80	8.36
3	10.17	10.31	8.88	8.44
4	-	11.79	10.12	9.61
5	11.79	11.88	9.62	9.14
6	9.72	9.81	8.23	7.82
7	10.80	10.98	7.39	7.02
8	11.79	12.02	-	9.14
9	10.58	10.76	7.69	7.31
10	10.98	11.12	7.99	-
11	11.66	11.70	9.98	9.48
12	12.38	12.38	10.32	9.80
13	10.08	10.22	10.54	10.01
14	10.76	11.07	10.11	9.60
15	11.07	11.35	8.54	8.11
16	11.75	11.75	8.73	8.29
17	11.25	11.39	9.01	8.56
18	12.15	12.24	8.70	8.30
19	11.43	11.52	9.22	8.76
20	11.93	12.11	10.20	9.69
21				

5.点击Continue，返回Paired-Samples T Test对话框;

6.点击OK，输出结果。

六、结果解释.

1. 基本描述.

Paired Samples Statistics表格给出了数据的部分描述性统计结果：

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	carb_protein	11.3023	20	.71368	.15958
	carb	11.1668	20	.72608	.16236

各列变量名和含义对应如下：

列名	含义
Mean	平均数
N	例数
Std. Deviation	(样本) 标准差
Std. Error Mean	均数的标准误

本例中，受试者饮用碳水化合物-蛋白质混合饮料的平均跑步距离为11.3023 km，多于饮用纯碳水化合物饮料的平均跑步距离11.1668 km，而后的变异程度(标准差)(0.72608 km)要高于前者(0.71368 km)。我们在最终汇报描述性结果时应该报告平均数和标准差，而不是均数的标准误，同时要注意小数点位数的统一，比如都保留3位小数。

2. 配对样本t检验——差值结果.

Paired Samples

Test表格给出了2组均数的差值(carb_protein变量减去carb变量)，以及差值的变异程度指标：

本例中，2组配对试验的均数差值为0.13550 km，标准差为0.09539 km，标准误为0.02133 km，差值的95%置信区间(95% CI)为0.09085-0.18015。由于我们之前调整过carb和carb_protein变量的顺序，此时差值为正数表示carb_protein组的平均跑步距离大于carb组。

3. 配对样本t检验——检验结果.

配对样本t检验的检验结果见Paired Samples Test表格的最右侧3列：

从左到右分别为配对样本t检验的t值(t)、自由度(df)和p值(Sig. (2-tailed))。如果P<0.05，表示2个相关(配对)组别的均数差异具有统计学意义;反之，表示2个相关(配对)组别的均数差异无统计学意义。

本例中，P=.000，表示P<0.001，carb组和carb_protein组的均数差异具有统计学意义。还有另一种说法是总体人群中carb组和carb_protein组的跑步距离的差异不等于0。

均数差值的95%CI和p值之间有一定的对应关系。如果95%CI内不包含0，则P<0.05;如果95%CI内包含0，则P≥0.05。本例中，差值的95%CI为0.09085-0.18015，不包含0，因此2组差异具有统计学意义(P<0.05)。

4. 配对样本t检验——计算效应值.

现在一些杂志要求汇报统计学显著性水平的同时，还要求汇报效应值的大小。对于配对样本t检验，效应值(用d或Cohen ' d表示)等于均数差值(M)除以差值的标准差(SD)：

$$d = \frac{M}{SD}$$

M和SD在Paired Samples Test表格中已经给出，M=Mean，SD=Std. Deviation。本例的效应值大小d为：

$$d = \frac{0.13550}{0.09539} = 1.42$$

效应值是衡量研究结果实际意义的指标，Cohen ' d大小的强度分级标准如下：

效应值大小	强度
0.2	小
0.5	中
0.8	大

本例中效应值 $d=1.42$ ，强度大。但是，效应值的缺点是其实意义局限于特定研究对象，而且目前还没有完整规范的指南来阐述效应值强度的意义。关于效应值的计算方法有很多种，应结合我们的研究设计和研究类型进行适当选择。

七、撰写结论

总的来说，我们可以按照如下方式完整地报告结果：

数据以均数±标准差的形式表示。在跑步距离上，受试者饮用碳水化合物-蛋白质混合饮料(11.302 ± 0.717 km)比饮用纯碳水化合物饮料(11.167 ± 0.726 km)的跑步距离增加了0.136 km(95% CI：0.091-0.180 km)，差异具有统计学意义， $t(19)=6.352$ ， $P<0.001$ ， $d=1.42$ 。

如果再增加假设检验的内容，可以这样报告结果：

利用配对样本t检验来判断，受试者饮用碳水化合物-蛋白质混合饮料相比于饮用纯碳水化合物饮料的跑步距离差异是否有统计学意义。数据以均数±标准差的形式表示。

利用箱线图，发现了2个距离箱子边缘超过1.5倍箱身长度的异常值，但是由于这2个异常点的数值并非极端异常值，所以仍在后续分析中保留它们。经Shapiro-Wilk检验，2组差值的数据服从正态分布($P=0.780$)。

在跑步距离上，受试者饮用碳水化合物-蛋白质混合饮料(11.302 ± 0.717 km)比饮用纯碳水化合物饮料(11.167 ± 0.726 km)的跑步距离增加了0.136 km(95% CI：0.091-0.180 km)，差异具有统计学意义， $t(19) = 6.352$ ， $P<0.001$ ， $d=1.42$ 。

从无效假设和备择假设的角度出发，也可以这样报告结果：

饮用碳水化合物-蛋白质混合饮料和饮用纯碳水化合物饮料2组跑步距离的均数差值与0相比，差异具有统计学意义。因此，我们可以拒绝无效假设，接受备择假设，认为饮用碳水化合物-蛋白质混合饮料相比于饮用纯碳水化合物饮料有助于提高人们的跑步距离。

最后，我们可以用带有95% CI(error bar)的简单条形图来更加直观地呈现配对样本t检验的结果，感兴趣的读者可以自行绘制。

更多 统计方法 请访问 <https://www.iikx.com/news/statistics/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发