

# 图解SPSS数据分析(一)：基本界面操作和常用统计分析

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/statistics/9796.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！**

## 图解SPSS数据分析(一)：基本界面操作和常用统计分析

。对科研工作者来说，科研数据的统计分析是常常碰到的问题，从课题设计、SCI论文写作发表、研究生学位论文答辩到国家自然科学基金课题申请等任何一步都离不开数据分析。目前相关的统计分析软件很多，如SAS，SPSS，STATA，Minitab，Eviews，R语言等，相比这些软件，SPSS的操作界面简单，菜单式的操作方法，成为初学者不二的选择。因此SPSS是科研工作者必备的科研利器。

SPSS(Statistical Product and Service Solutions)是一个组合式软件包，它集数据录入、整理、分析功能于一身。2010年被IBM并购SPSS统一更名为IBM SPSS，每年8月更新版本，最新版本是2018年8月更新的SPSS 26.0。本文以SPSS 20.0为例进行讲解。

### 统计分析方法的选择：

分析方法的选择，抓住本质才是最重要的

。笔者认为，

方法选择的思考路径：研

究目的—可分析的变量—数据类型—分析方法

。研究目的很好理解但也很重要，一个数据可能发出几篇文章，那自然研究目的就有所不同了，相对应的统计方法自然也就不同了，再比如，欲比较均数之间的差异，你选择t检验或方差分析，若比较不同组之间率或构成比的差异，你可以选择卡方检验。

可分析的变量就是说，你在分析之前呢，先看看自己的数据，比如某个变量的缺失值达到50%，那这个变量就可以直接pass过去了，再比如编号ID之类的虽然需要，但也不是我们关心的变量。下面重点就来了，数据类型是初学者最容易忽视和搞不清楚的问题，但是这对于统计方法的选择至关重要，因为每个统计方法基本上都有相应的适用条件，说到这里不得不提一下变量类型。

### 基本变量类型如下：

计量资料(可以测量的、有单位的，g.身高、体重、年龄等)

计数资料：二分类、多分类(可以数个数的，g.性别：可以数出男多少人、女多少人等)

等级资料(程度呈递进的，g.一些实验室检验结果：可分为-、+、++、+++)

小编亲自总结了常用统计方法对变量的要求，如下表：

自变量 (X)	因变量 (Y)	统计方法
二分类	计量资料	T 检验
多分类	计量资料	方差分析
计数资料或等级资料	计数资料	卡方检验
计量资料	计量资料	线性相关
计量、计数、等级资料	计量资料	多元线性回归
计量、计数、等级资料	计数资料	Logistic 回归
计量、计数、等级资料	时间变量和终点事件	Cox 回归

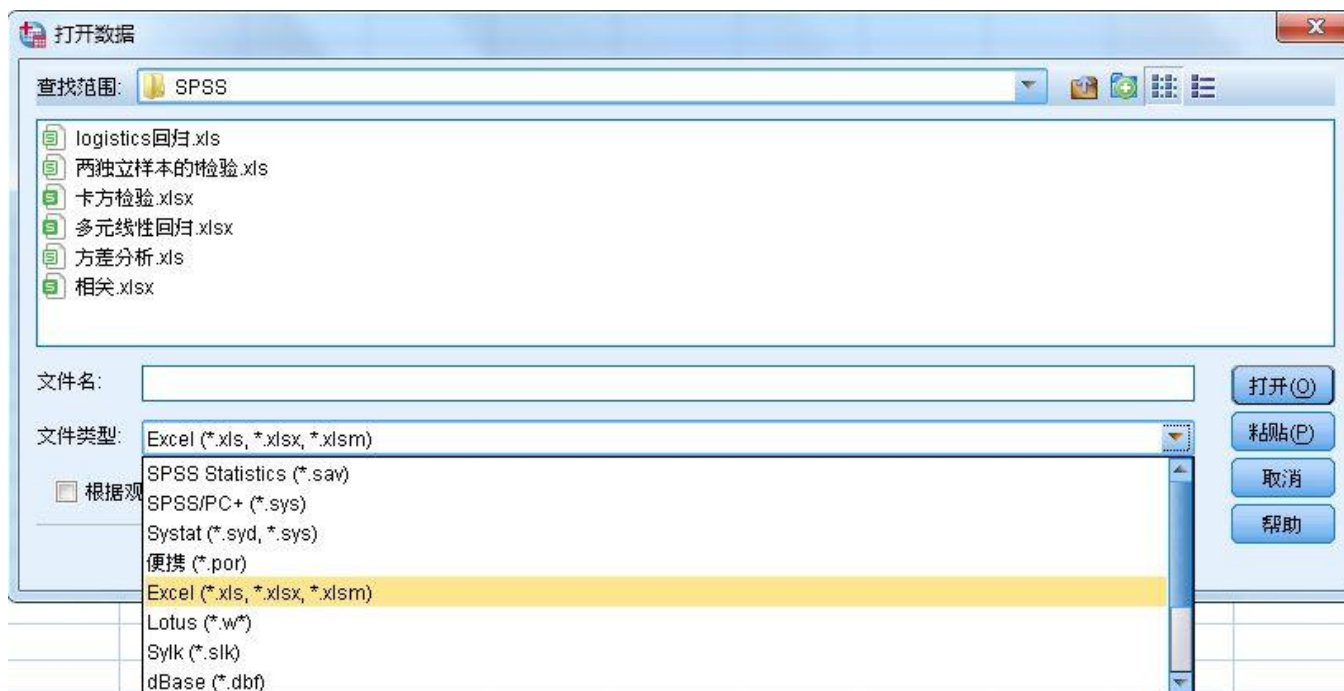
SPSS软件的数据管理：

数据管理是统计分析前必不可少的步骤，主要包括数据文件的建立、存取、核对和数据整理。现在就先让大家看看SPSS的界面。(常用的都圈起来了哟)

SPSS主窗口界面



我们常遇到数据大多为Excel文件，只需要依次点击：文件—打开—数据—选择Excel文件—打开，就OK了。(选择Excel文件时注意选择文件类型，如下图，当然如果你打开的是其他的文件类型，选择相应的文件类型就行了。)



下面我们简单介绍一下自己录入数据，举个例子吧，你想看看A、B药的治疗效果，已知服用A药患者64人愈合，21人未愈合；服用B药患者51人愈合，33人未愈合，你想看看两个药的疗效有无差异，你知道要用卡方检验，但却不知道如何将数据录入，下面咱就来看看吧。



图2

- ①点击：文件——新建，然后直接输入数据，结果如图1
- ②点击：变量视图，结果如图2
- ③在变量视图里，可以直接对变量的特征进行修改，主要修改前四项修改后如图3
- ④此时数据录入已完成，点击：数据视图，结果如图4

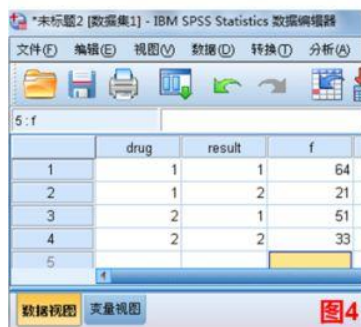
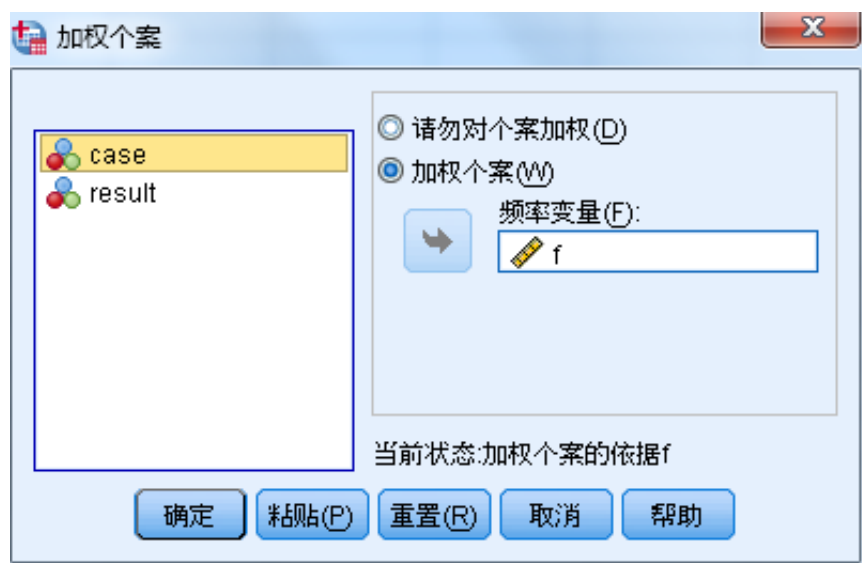


图3

这样就可以做卡方了吗？No，常搞数据的同学可能已经意识到了，平时Excel导入的数据都是一行数据表示一个患者的记录，而我们录入的数据一行则表示一种药某种结局的人数，此时就需要对数据进行加权，这就是小编要强调的，操作如下：依次点击：数据—加权个案，出现如下对话框，选择加权个案，将f变量添加进去频率变量框中，如下图所示，点击：确定，就好了，就可以做卡方检验了。



常见统计分析方法实例及SPSS操作：

### 1、T检验(适用条件：随机样本，正态分布，方差齐性)：

T检验主要包括：单样本t检验、配对t检验、两独立样本的t检验。今天我们就以最为常见的两独立样本的t检验为例。简单地说就是比较两种不同的处理，效果的差异，比如下面的例1，比较两种降糖药的效果。

**例1** 为研究国产四类新药阿卡波糖胶囊的降血糖效果，某医院用**40名2型**糖尿病病人进行同期随机对照试验。试验者将这些病人随机等分到试验组(用阿卡波糖胶囊)和对照组(用拜唐苹胶囊)，分别测得试验开始前和8周后的空腹血糖，算得空腹血糖下降值见表1，能否认为该国产四类新药阿卡波糖胶囊与拜唐苹胶囊对空腹血糖的降糖效果不同？

**表1 试验值和对照组空腹血糖下降值 (mmol/L)**

试验组 $X_1$	-0.70	-5.60	2.00	2.80	0.70	3.50	4.00	5.80	7.10	-0.50
( $n_1=20$ )	2.50	-1.60	1.70	3.00	0.40	4.50	4.60	2.50	6.00	-1.40
对照组 $X_2$	3.70	6.50	5.00	5.20	0.80	0.20	0.60	3.40	6.60	-1.10
( $n_2=20$ )	6.00	3.80	2.00	1.60	2.00	2.20	1.20	3.10	1.70	-2.00

SPSS的操作如下：

(1)原始数据形式如下图，x表示血糖下降量，group(1：实验组，2：对照组)。



(2)依次点击：分析、均值比较、独立样本T检验，出现如下窗口，如下图。

(3)将x选入检验变量，将group选入分组变量，点击：定义组，组1定义为1、组2定义为2(主要取决于group变量里面的值)，如下图。

(4)点击：选项，如下图，(一般选默认即可，根据自己事情要求而定)，点击：继续。(5)点击：确定，结果就出来了。

The screenshot illustrates the SPSS interface for performing an independent sample t-test. On the left is the '数据编辑器' (Data Editor) window showing a dataset with two columns: 'x' and 'group'. The 'group' column has values 1 and 2. The main part of the image shows three dialog boxes in sequence, connected by yellow arrows:

- 独立样本 T 检验 (Independent-Sample T Test):** The '检验变量(T):' (Test Variable(s)) box contains 'x'. The '分组变量(G):' (Grouping Variable) box contains 'group'. The '定义组(D)...' (Define Groups...) button is highlighted with a red circle 2.
- 定义组 (Define Groups):** The '使用指定值(U)' (Use specified values) radio button is selected. '组 1(1):' (Group 1) is set to 1 and '组 2(2):' (Group 2) is set to 2. The '继续' (Continue) button is highlighted with a red circle 3.
- 独立样本 T 检验: 选项 (Independent-Sample T Test: Options):** The '置信区间百分比(C):' (Confidence interval percent) is set to 95%. The '按分析顺序排除个案(A)' (Exclude cases by analysis order) radio button is selected. The '继续' (Continue) button is highlighted with a red circle 4.

结果如下：

## T检验

注：若 $P < 0.05$ ，则说明有差异，具体哪种药好，可以比较两组的均数。

[0001]

组统计量

	group	N	均值	标准差	均值的标准误
x	1	20	2.065	3.0601	.6843
	2	20	2.625	2.4205	.5412

在论文中，T检验的结果，主要提取均数±标准差（①），t值（③）、P值（④）。若②中 $p > 0.05$ ，则方差相等，取第一行的t值与p值；若②中 $p < 0.05$ ，则方差不相等，取第二行的t值与p值。本例的结果 $P = 0.525$ ，说明尚不能认为两种降糖药的效果有差别。

独立样本检验

		方差方程的 Levene 检验		均值方程的 t 检验						
		F	Sig.	t	df	Sig. (双侧)	均值差值	标准误差值	差分的 95% 置信区间	
									下限	上限
x	假设方差相等	.578	.452	-.642	38	.525	-.5600	.8724	-2.3262	1.2062
	假设方差不相等			-.642	36.086	.525	-.5600	.8724	-2.3293	1.2093

## 2、方差分析(适用条件：正态性、独立性、方差齐性)：

方差分析主要包括：完全随机设计的方差分析、随机区组设计的方差分析。今天我们就以最为常见的完全随机设计的方差分析为例。（方差分析与T检验最主要的区别就是，group大于2个组别）

**例2** 为研究钙离子对体重的影响作用，某研究者将36只肥胖模型大白鼠随机等分为三组，每组12只，分别给予常规剂量（0.5%）、中剂量组（1.0%）和高剂量组（1.5%）3种不同的高脂饲料，喂养9周，测其喂养前后体重的差值（表2）。问3种不同喂养方式下大白鼠体重改变是否不同？

表2 3种不同喂养方式下大白鼠体重喂养前后差值（g）

常规剂量（0.5%）	中剂量组（1.0%）	高剂量组（1.5%）
332.96	253.21	232.55
297.64	235.87	217.71
312.57	269.30	216.15
295.47	258.90	220.72
284.25	254.39	219.46
307.97	200.87	247.47
292.12	227.79	280.75
244.61	237.05	196.01
261.46	216.85	208.24
286.46	238.03	198.41
322.49	238.19	240.35
282.42	243.49	219.56

SPSS的操作如下：

(1)部分数据如下图 ,x表示体重差值,group(1常规剂量组 , 2中剂量组 , 3高剂量组)。

(2)依次点击：分析、均值比较、单因素ANOVA，将x选入因变量列表框，group选入因子框，如下图 。

(3)点击：两两比较，勾选LSD，显著性水平默认0.05，如下图 ，点击：继续。

(4)点击：选项，统计量勾选描述性，其他都默认，如下图 ，点击：继续。

(5)点击：确定，结果就出来了。

The image shows the SPSS interface for performing a one-way ANOVA. It includes a data view window with columns 'x' and 'group', and three sub-dialog boxes for configuring the analysis: '单因素方差分析' (One-Way ANOVA), '单因素 ANOVA: 两两比较' (One-Way ANOVA: Post Hoc), and '单因素 ANOVA: ...' (One-Way ANOVA: Options). Arrows indicate the sequence of clicks from the data view to the final options dialog.

	x	group
1	332.96	1
2	297.64	1
3	312.57	1
4	295.47	1
5	284.25	1
6	307.97	1
7	292.12	1
8	244.61	1
9	261.46	1
10	286.46	1
11	322.49	1
12	282.42	1
13	253.21	2
14	235.87	2
15	269.30	2
16	258.90	2
17	254.39	2
18	200.87	2
19	227.79	2
20	237.05	2
21	216.85	2
22	238.03	2
23	238.19	2
24	243.49	2
25	232.55	3
26	217.71	3
27	216.15	3
28	220.72	3
29	219.46	3

主要结果如下：



描述

X	N	均值	标准差	标准误	均值的 95% 置信区间		极小值	极大值
					下限	上限		
1	12	293.3683	24.62068	7.10738	277.7251	309.0116	244.61	332.96
2	12	239.4950	18.72159	5.40446	227.5999	251.3901	200.87	269.30
3	12	224.7817	23.24461	6.71014	210.0127	239.5506	196.01	280.75
总数	36	252.5483	36.93959	6.15660	240.0498	265.0469	196.01	332.96

单因素方差分析

X	平方和	df	均方	F	显著性
组间	31291.796	2	15645.898	31.355	.000
组内	16466.867	33	498.996		
总数	47758.663	35			

在论文中主要提取的结果为：均数、标准差、F值、P值。本例结果F=31.355, P<0.05, 则可以认为3种不同喂养方式下大白鼠体重改变不同。

## 在此之后检验

论文中一般不用写，主要用于比较不同组间的差异。本例中，1组体重该变量高于2组和3组，但是2组和3组的差异无统计学意义，不能认为2组和3组有差异。

因变量: X  
LSD

(I) group	(J) group	均值差 (I-J)	标准误	显著性	95% 置信区间	
					下限	上限
1	2	53.87333*	9.11954	.000	35.3195	72.4272
	3	68.58667*	9.11954	.000	50.0328	87.1405
2	1	-53.87333*	9.11954	.000	-72.4272	-35.3195
	3	14.71333	9.11954	.116	-3.8405	33.2672
3	1	-68.58667*	9.11954	.000	-87.1405	-50.0328
	2	-14.71333	9.11954	.116	-33.2672	3.8405

\*. 均值差的显著性水平为 0.05。

## 3、卡方检验(适用于两个样本率或构成比的检验)：

卡方检验一般可分为：四格表卡方检验、R×C的卡方检验、配对设计资料的卡方检验

废话不多说了，直接看例题吧。



**例3** 某院欲比较异梨醇口服液（试验组）和氢氯噻嗪+地塞米松（对照组）降低颅内压的疗效。将**200**例颅内压增高症患者随机分为两组，结果见表**3**。问两组降低颅内压的总体有效率有无差别？

**表3** 两组降低颅内压有效率的比较

组别	有效	无效	合计
试验组	99	5	104
对照组	75	21	96
合计	174	26	200

SPSS的操作如下：

- (1)数据如下图，group(1实验组，2对照组)，result(1有效，2无效)，f表示频数。
- (2)依次点击：分析、描述统计、交叉表，将group选入行框，result选入列框,如下图
- (3)点击：统计量，勾选卡方，如下图，点击：继续。
- (4)点击：单元格，勾选观察值、行，其他都默认，如下图，点击：继续。
- (5)其他都默认，点击：确定，结果就出来了。



主要结果如下：

## 交叉表

[0004]

案例处理摘要

	案例					
	有效的		缺失		合计	
	N	百分比	N	百分比	N	百分比
case*result	200	100.0%	0	0.0%	200	100.0%

case\*result 交叉制表

		result		合计
		1.00	2.00	
case 1.00	计数	99	5	104
	case 中的 %	95.2%	4.8%	100.0%
2.00	计数	75	21	96
	case 中的 %	78.1%	21.9%	100.0%
合计	计数	174	26	200
	case 中的 %	87.3%	13.0%	100.0%

在论文中主要提取的结果为：频数、百分比、Pearson卡方值、P值。本例结果 $\chi^2=12.857$ ， $P<0.05$ ，则可以认为两组降低颅内压的总体有效率有差别，实验组高于对照组。

卡方检验

	值	df	渐进 Sig. (双侧)	精确 Sig. (双侧)	精确 Sig. (单侧)
Pearson 卡方	12.857 <sup>a</sup>	1	.000		
连续校正 <sup>b</sup>	11.392	1	.001		
似然比	13.568	1	.000		
Fisher 的精确检验				.001	.000
线性组合	12.793	1	.000		
有效案例中的 N	200				

a. 0 单元格(0.0%) 的期望计数少于 5。最小期望计数为 12.48。

b. 仅对 2x2 表计算

## 参考资料：

《卫生统计学》第7版，方积乾主编，人民卫生出版社，2014年5月

《SAS统计软件应用》，贺佳主编，人民卫生出版社，2014年4月

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转发