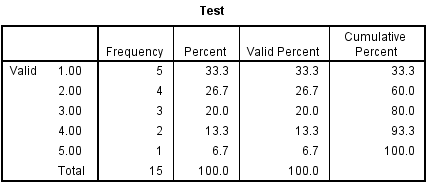
# 013：频度分布表(查看样本分布)

## 统计学解释

频数分布表：频数分布表，简称频数表，主要用来直接查看数据分布情况。对于离散型随机变量，直接计算每个值出现的频数并列表；对于连续型随机变量，则需要先将数据按等间隔分组，然后统计落在每一个组内的频数。

数组 [1,1,1,1,1,2,2,2,2,3,3,3,4,4,5] 在SPSS中的频数分布表如下：



其中：Frequency在这里指频数，Percent在这里指频率，Valid Percent指有效频率（即剔除了缺失值的频率，SPSS翻译为有效百分比），Cumulative Percent为累积频率（SPSS翻译为累积百分比）。

## 实现代码

现在，我们仅实现适用于离散型随机变量的频数分布表，其中包括频数、频率和累积频率。

实现思路：我们先计算各不同的值的出现频数，再通过频数计算频率和累积频率。

### 定义测试数据

data\_test=[1,1,1,1,1,2,2,2,2,3,3,3,4,4,5]

### 绘制频数分布表

from prettytable import PrettyTable #我们使用PrettyTable包来绘制表格

data\_set=set(data\_test);

size=len(data\_test)

# 计算频数

frequency={};

for i in data\_set:

frequency[i]=data\_test.count(i)

# 生成频数统计表

title=['Valid','Frequency','Percent','Cumulative Percent'] #定义表头

table=PrettyTable(title)

cumulative=0 #定义累积频率变量

for key,value in frequency.items():

valid\_percent=value/len(data\_test) #计算频率

cumulative+=valid\_percent #计算累积频率变量

array=[key,value,round(valid\_percent\*100,1),round(cumulative\*100,1)]

table.add\_row(array)

print(table)

结果

+-------+-----------+---------+--------------------+

| Valid | Frequency | Percent | Cumulative Percent |

+-------+-----------+---------+--------------------+

| 1 | 5 | 33.3 | 33.3 |

| 2 | 4 | 26.7 | 60.0 |

| 3 | 3 | 20.0 | 80.0 |

| 4 | 2 | 13.3 | 93.3 |

| 5 | 1 | 6.7 | 100.0 |

+-------+-----------+---------+--------------------+

其中：

table=PrettyTable(title) #定义PrettyTable表，并给出表头栏title

table.add\_row(array) #向PrettyTable表添加一行数据

round(x,1) #保留x的一位小数

## 实际应用

1.可以用作探索性研究，了解样本的分布类型和分布特征，以便选择合适的统计方法，进行进一步的计算和处理

2.可以用作数据清洗，发现特别大或特别小的异常值，再加以处理