

订阅DeepL Pro以编辑此演示文稿。  
访问[www.DeepL.com/pro](https://www.deepl.com/pro?cta=edit-document)，了解更多信息。

**STAT200 - 初级统计学**

**教员。拉里-穆索里诺**

##### 第二章摘要

要创建一个频率分布表（用于整数数据）。

1. 注意：班级的数量与表格中的行数相同。
2. 用公式计算类的宽度。 
3. 注意：对于整数数据，四舍五入类宽到下一个较高的整数。
4. 创建一个行数等于班级数的表格。 为班级下限、班级上限、频率、相对频率等创建列。
5. 为1st 行设置类下限，等于数据集的最小值。
6. 然后计算 2nd 行的等级下限，即 1st 行的等级下限**加上**等级宽度。 加上班级宽度来计算其他行的班级下限。
7. 每一行的上层阶级限制比后面一行的下层阶级限制少一个。 你也可以把班级宽度加到每个班级上限上，以确定后续行的班级上限。
8. 每一行的频率：计算有多少数据值落在该行的等级下限和等级上限之间。
9. 每行的相对频率：取该类的频率，除以数据值的总数。
10. 每行的中点：取班级下限加班级上限，除以2
11. 每行的班级边界。 班级下限是班级下限-0.5
12. 每行的班级界线：班级上界是班级上限+0.5
13. 每行的累积频率：取该行的频率，加上之前所有行的频率。

重要的是对感兴趣的数据集进行图表/绘图，图表的例子包括。

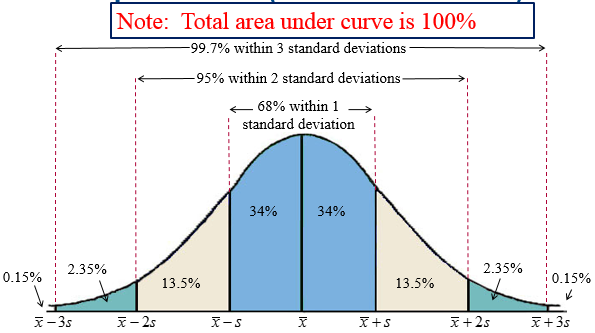
* 频率直方图。
* 注：为频率分布表。
* 点状图。
* 茎和叶图。
* 波谱图
* (x, y)数据的散点图

制图的好处。

* 识别异常值。
* 识别分布形状，如对称（钟形）、右偏、左偏。

|  |  |
| --- | --- |
| **样本平均数（样本平均数）** |  |
| **样本中位数** | * 将数据从最小值到最大值排序，那么中位数就是这个排序集合中的中间数据条目。 * 如果有偶数的数据值，取中间的两个数据值，相加后除以2来确定中位数。 |
| **范围** | (最大的数据值)-(最小的数据值) |
| **样本差异** |  |
| **样本标准差（注意这是样本方差的平方根）。** |  |
| **来自分组数据的样本平均值近似，即频率直方图** | 其中 是每个类别的中点， 是每个类别的频率。  请注意 |
| **来自分组数据的样本标准偏差近似值，即频率柱状图** | 其中 是每个类别的中点， 是每个类别的频率。 请注意 |
| **加权平均数，其中每个数据值都有一个相应的权重** | 其中***x***是数据值（例如分数）。  和***w***是权重（十进制格式）。 |
| **使用IQR识别异常值** | 识别数据集中的异常值。   1. 找到数据集的四分位数。 Q1 , Q2 , Q3 2. 找到IQR，其中IQR = Q3 - Q1 3. 计算出异常值的下限。 下限 = Q1 - 1.5\*IQR 4. 计算出离群值的上限。 上限 = Q3 + 1.5\*IQR 5. 任何低于下限或高于上限的数据值都是离群值。 |

**经验法则--用于钟形分布**



经验法则--对于钟形（对称）分布。

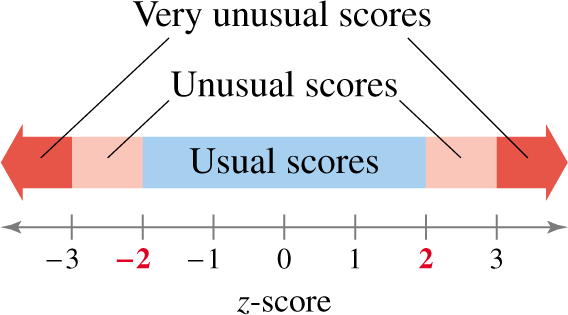
* ~68%的分布是在平均值的一个标准差之内。
* ~95%的分布是在平均值的两个标准差之内。
* ~99.7%的分布是在平均值的三个标准差之内。

***z***分数：衡量一个数据值离平均值有多远，相对于标准差的数量。

例如，如果一个数据值高于平均值两个标准差，那么相应的z分数将是z=2.0。

****

|  |  |
| --- | --- |
| **不寻常的数据值** | 与平均值相差2个标准差以上的数据值。 |
| **非常不寻常的数据值** | 与平均值相差3个标准差以上的数据值。 |



我们可以使用z-score来确定一个给定的数据值x与平均值的标准差的数量。

|  |  |  |
| --- | --- | --- |
| **解释z-scores** | 如果z-score为负数。 | X值小于平均值 |
| 如果z-score为零。 | X值等于平均值 |
| 如果z-score为正数。 | X值大于平均值 |
| 如果z-score大于2或小于-2 | x值是**不寻常的**，因为只有~5%的数据值是在± 2个标准差之外。 |
| 如果Z-分数大于3或小于-3 | x值**是非常不寻常的**，因为只有~0.3%的数据值是在± 3个标准差之外。 |

**第二章中使用的StatCrunch命令。**

|  |  |
| --- | --- |
| STAT🡪 数据🡪 排序 | 按升序或降序对一列数据进行排序。 |
| STAT🡪 摘要统计🡪 栏目 | 生成一列数据的汇总统计，如平均值、标准差、中位数、四分位数等。  注：在StatCrunch中，标准差是指*样本*标准差，而  未调整的标准差指的是*人口*标准差 |
| STAT🡪 计算器🡪 自定义 | 计算加权平均数。  还可以计算分组数据的平均值和标准偏差 |
| GRAPH🡪 直方图 | 为一列数据生成柱状图 |
| 图表🡪 茎和叶 | 为数据列生成茎叶图 |
| GRAPH🡪 点阵图 | 生成一列数据的点阵图 |
| GRAPH🡪 饼状图🡪 附摘要 | 生成饼状图 |
| GRAPH🡪 散点图 | 为(x, y)数据的表格生成散点图 |
| GRAPH🡪 箱形图 | 生成一列数据的箱形图 |