W1

* **变量类型**
* Categorical (Qualitative) 分类（定性）：

名义变量Nominal (no ranking)：性别、类型、是否

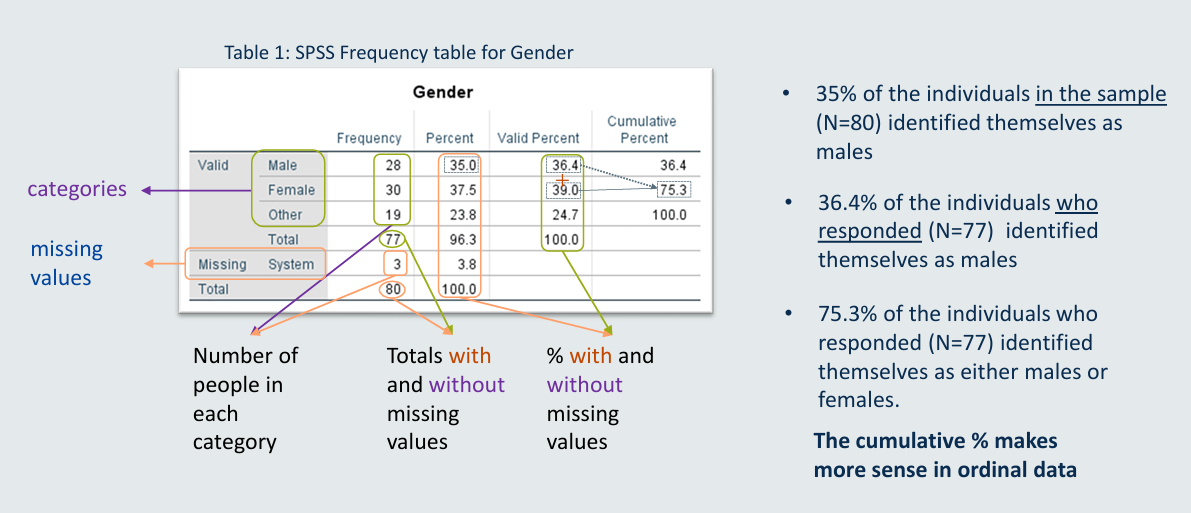
序数变量Ordinal (ranking)：表示类别的顺序，但缺乏数学意义；不等距

* Numerical (Quantitative) 数值变量（定量）

离散Discrete (set of values)：{1,2,3,4,5,6,7,8,9,10}（点）表计数

连续Continuous (all values in an interval)：【1,10】（线）表测量

* 注：序数和离散的区分：量表表程度（很反对、反对、中立、赞成、很赞成）、只有四五档的多为序数，多达10个程度、序号ID、如鞋码均为离散；身高、年龄、测量值为连续。
* 例题：communication: Quality of communication with the psychologist (1= very poor communication – 7= very good communication) 属于Categorical (Qualitative) 分类变量；而income（1：10000-20000，。。。，5:50000-60000,6:60000以上）属于离散数值变量
* 操作技巧：
* **Typo选出**：Data- Select Cases-输入条件-删除typo项-删除新生成筛选列
* **Box plot箱图**：Graphs-Legacy Dialogue-Box，奇异值显示的是序号，沿序号定位值内容
* **Describe题**：**数值型**：均值=xx（SD=xx），是否正态，是否关于均值对称，最小最大范围；**分类型**：每个类别从高到低各有几人（xx%），missing value几人（xx%）；**中位数众数**：中位数：一半人的xx都在中位数xx以上；众数：最多报告的是xx值；**箱图**：比较均值、范围、四分位、偏度、奇异值（outliers）。
* 不同变量的合适描述指标Descriptive indices和图表
* 分类变量Categorical：通过**frequencies and percentages，Bar Chart or Pie Chart**。
* 少于5点的序数同上，多于5点的同数值型，须考察是否正态。



In the sample：样本量总人数（多）

Who responded：去除missing value后的人数和比例（少）

* 连续数值型Continuous Numerical：查看是否正态，**Histogram, Box plot**。**Location: mean, median, mode; Dispersion: SD（Std. Deviation）, range, IQR（四分位数）**。
* 正态曲线Normal：众数、中位数、均值为同一个值（考察正态水平）；绕均值对称(symmetrical about the mean): 使用**mean and standard deviation**。
* 偏态Skew：偏左为正，偏右为负；使用**median and min-max**。

