# 第一阶段 业务分析师

# 建模数据分析师

## 网校相关视频

SPSS数据分析

## 明确本次课知识点，明确重点难点

### 【知识点目标】

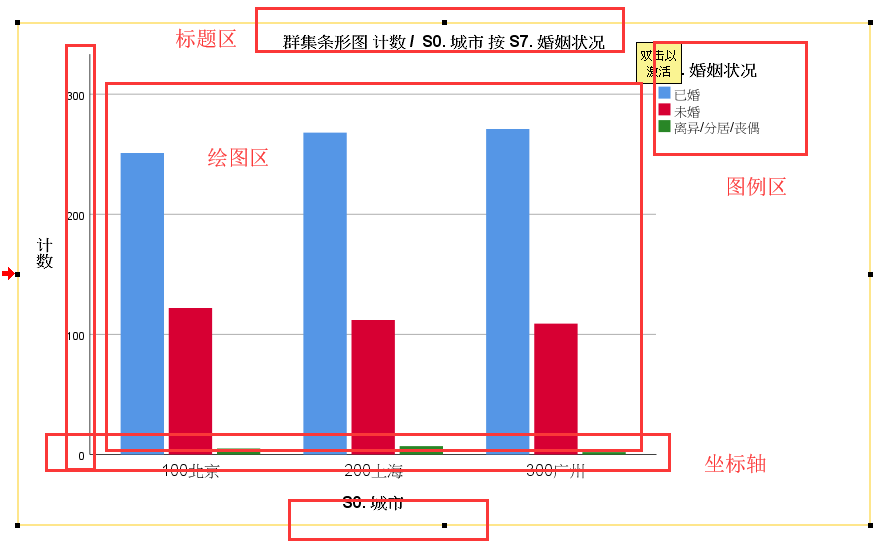
* SPSS数据可视化
* 直方图、分段直方图、直方图组的绘制
* 箱体图的绘制
* 饼图的绘制
* 条形图的绘制
* 线图的绘制
* 散点图的绘制
* 采用可视化方法

## 复习巩固作业讲解

无

## 本次课程任务讲解

### 【知识点1】SPSS数据可视化



标题区和注解区：位于图形的最上方和最下方；

坐标轴：用于表示响应变量的取值情况；

绘图区：被坐标轴包围，直接使用图形元素对数据进行呈现的区域；

图例区：位于整个图形的右侧。

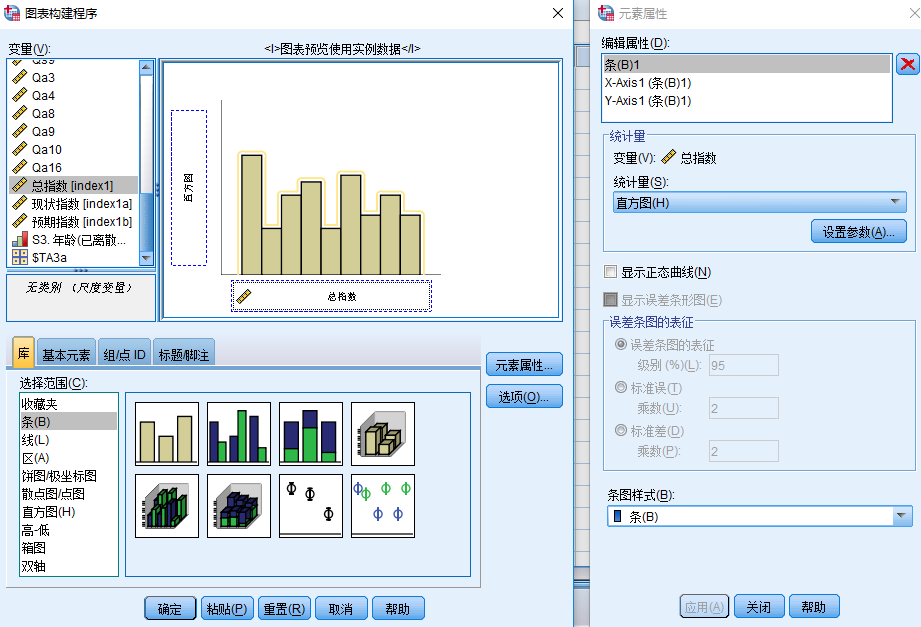
### 【知识点2】直方图、分段直方图、直方图组的绘制

直方图主要用来观察连续变量的分布情况

通过图表构建器创建直方图

1）直方图：观察数据分布“形状”

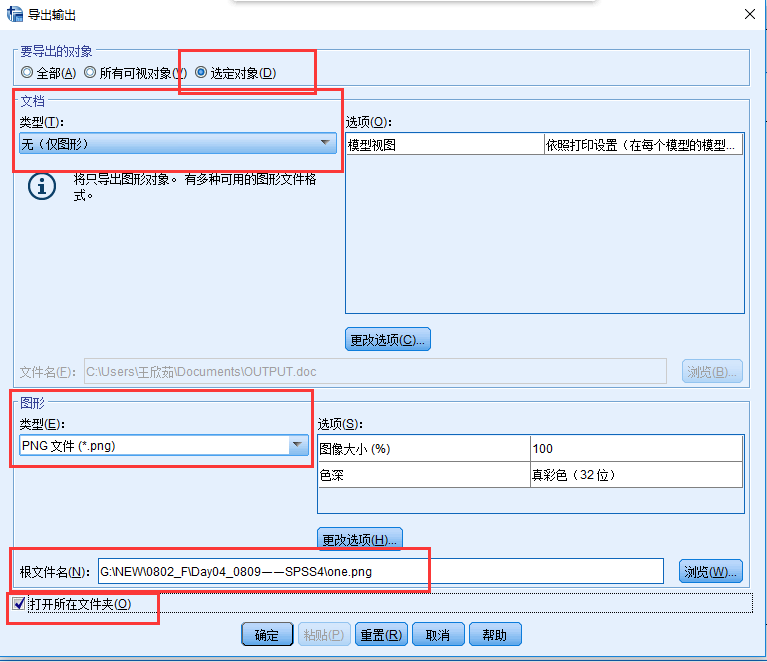
案例：*CCSS\_Sample.sav*,绘制消费者信心值的直方图



图表编辑器修改美化图形

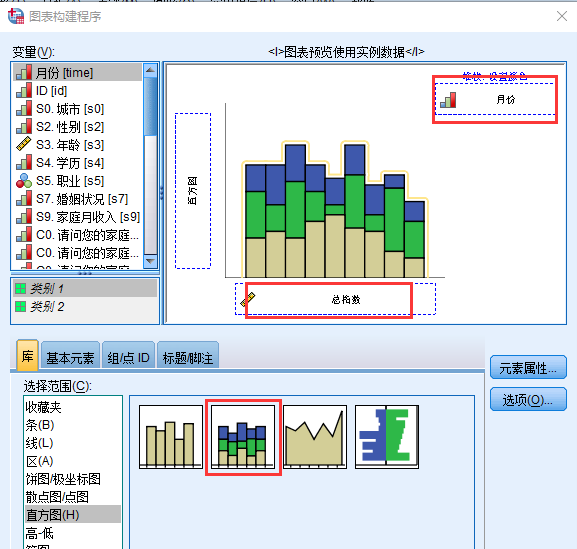
操作：双击激活图形，想要对某部分进行样式的设置，就点击该部分，就会有对应的菜单来进行操作

图形的导出



2）分段直方图

案例：CCSS\_Sample.sav,绘制消费者信心值的直方图，希望观察各组段内各个月份的对比情况。

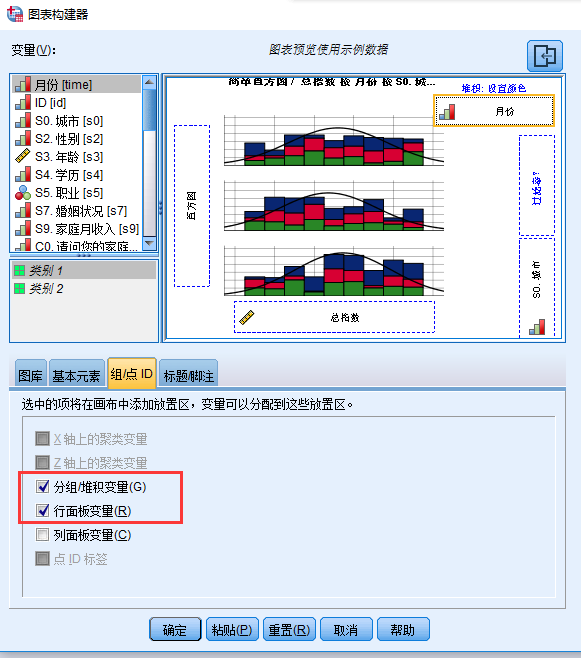


3）直方图组

案例：CCSS\_Sample.sav,绘制消费者信心值的直方图，希望观察北京、上海、广州三地受访者的信心分布有无差异。

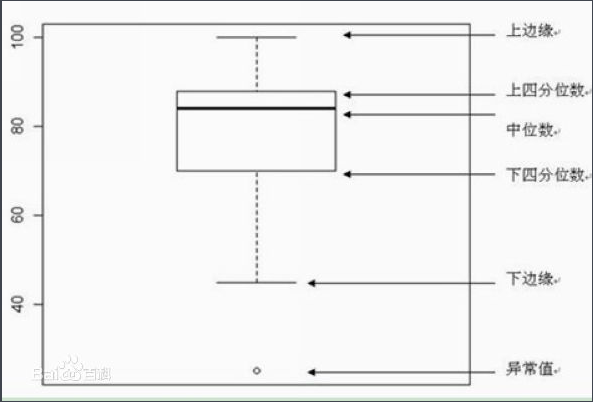


查看不同城市、不同月份的消费者信心指数的分布情况



### 【知识点3】箱体图的绘制

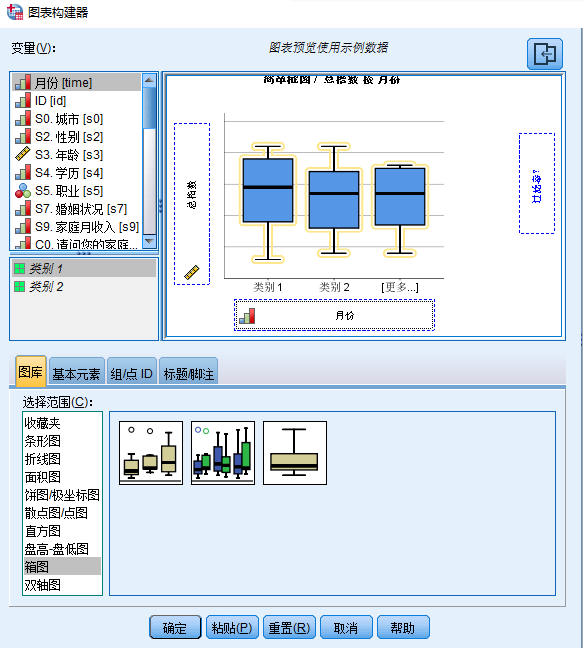
箱图除了用来观察连续变量的分布情况之外，也常用来观察连续变量中的异常值情况。通过为箱图设置分类变量，可以方便地观察连续变量在不同类别上的分布情况



**案例：CCSS\_Sample.sav, 分月份考察消费者信心的分布**

**1.5倍四分位间距被定义为异常值，用o表示**

**3倍四分位间距被定义为极端值，用＊表示**

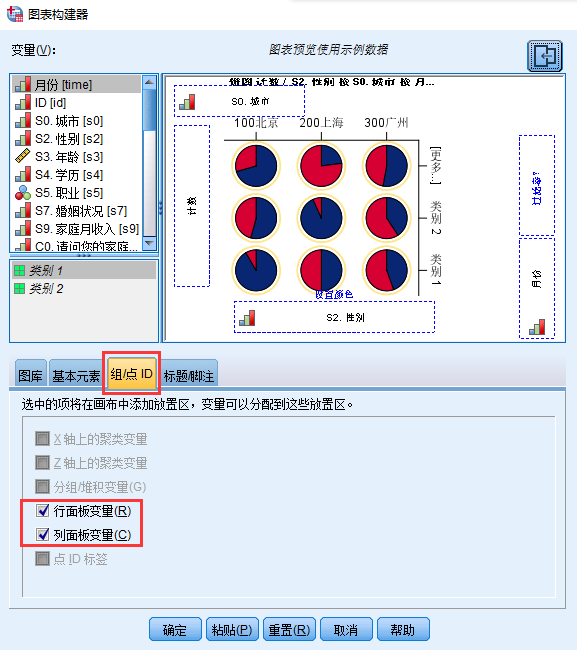


直方图和箱图除了可以观察连续变量自身的分布情况之外，还可以观察分类变量对连续变量的影响

### 【知识点4】饼图的绘制

目的：观察分类变量各类别占的比例

案例：CCSS\_Sample.sav, 分城市、分月份考察CCSS的性别比例是否存在一定的变化趋势



### 【知识点5】条形图的绘制

观察分类变量对连续变量的影响

利用排序和数据标签，可以很方便地观察分类变量中每个类别的具体差异情况。

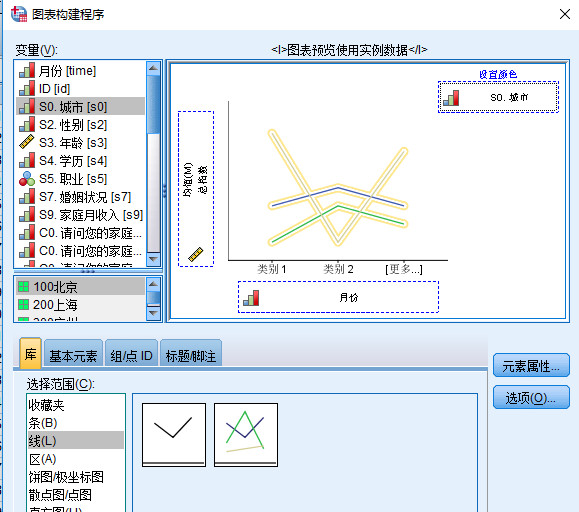
案例：CCSS\_Sample.sav,不同职业人群的消费者信心值的均值有无差异。

对条形图添加数据标签和按照降序排序

### 【知识点6】线图的绘制

折线图主要用来观察连续变量按照时间变化的趋势

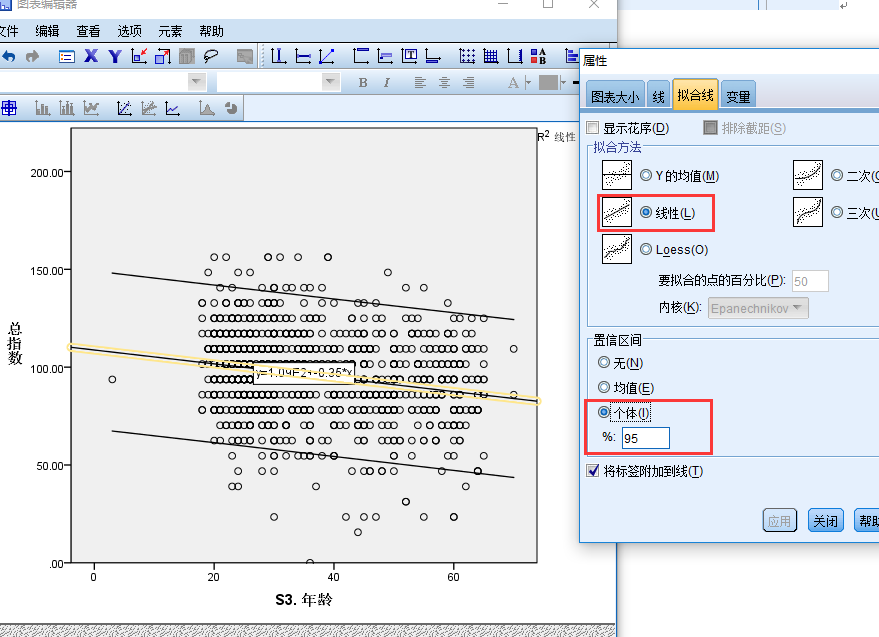
多重线图案例： CCSS\_Sample.sav,分城市考察不同月份总消费信心指数的均值变化有无内在趋势



### 【知识点7】散点图的绘制

散点图主要用来观察连续变量对连续变量的影响，可用于观察两个连续变量之间的相关性

案例文件：CCSS\_Sample.sav， 利用简单散点图考察s3年龄和总消费者信心指数间的数量关联趋势。



95%置信区间，意味着如果你用同样的步骤，去选样本，计算置信区间，那么100次这样的独立过程，有95%的概率你计算出来的区间会包含真实参数值，即大概会有95个置信区间会包含真值。

而对于某一次计算得到的某一个置信区间，其包含真值的概率，我们无法讨论。

**如果我们重复取样，每次取样后都用这个方法构造置信区间，有 95% 的置信区间会包含真值**

### 【知识点8】采用可视化方法

连续变量：直方图/箱图

分类变量：饼图

分类变量VS分类变量：分组的饼图

分类变量VS连续变量：分组的直方图/箱图，堆叠，柱状图/条形图，折线图

连续变量VS连续变量：散点图

## 自主学习作业讲解

无

## 课程总结

* SPSS数据可视化
* 直方图、分段直方图、直方图组的绘制
* 箱体图的绘制
* 饼图的绘制
* 条形图的绘制
* 线图的绘制
* 散点图的绘制
* 采用可视化方法

## 下次自主学习任务布置

#### 观看预习视频

数据建模分析师——SPSS数据分析

#### 课后作业

1. 参见直播平台
2. Qq群课后习题