# 第三阶段 从数据分析到数据挖掘

# SPSS Modeler-[数据分析与挖掘1]

## 网校相关视频

第二章 SPSS Modeler模型篇

## 明确本次课知识点，明确重点难点

### 【知识点目标】

* SPSS Modeler 数据分析--常见差异性分析
* SPSS Modeler 数据分析--相关分析
* SPSS Modeler 数据分析—线性回归

## 本次课程任务讲解

### 【知识点1】SPSS Modeler 数据分析--常见差异性分析

#### 列联表与卡方检验

列联表（交叉表）用来观察分类变量对分类变量之间的关系，在SPSS Modeler中，使用“矩阵”节点可以实现列联表的功能，既能通过观察两个分类变量之间的实际频数进而计算卡方值判定两变量之间的关联关系，也能在两分类变量的交叉点显示与之相关的连续变量的统计结果。

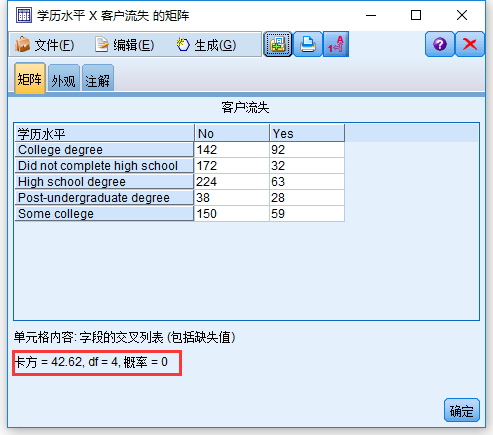
矩阵节点的使用

数据源：电信行业客户数据.xlsx

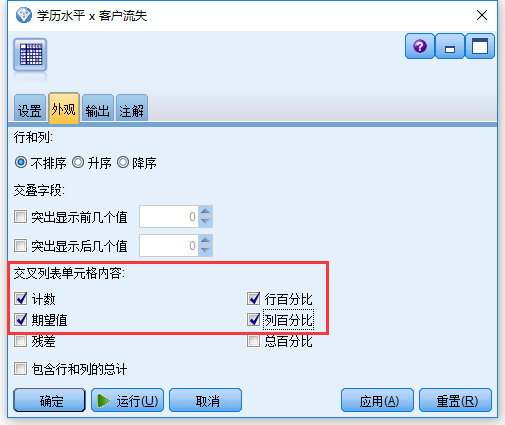
输出选项卡——矩阵

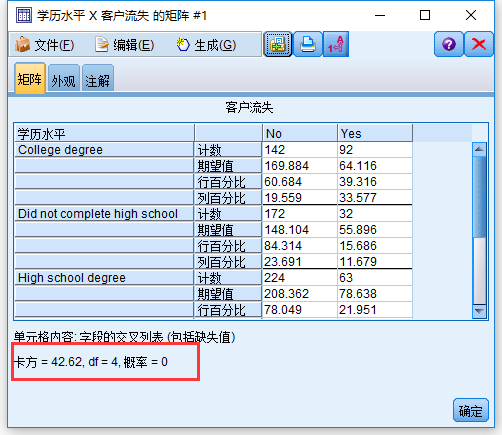


效果如下



可以设置交叉列联表单元格中显示的内容





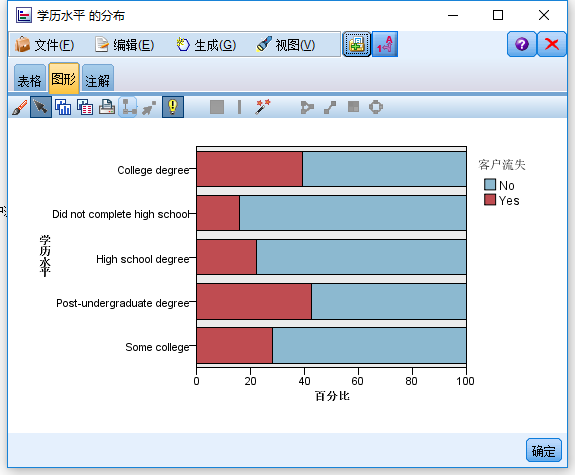
设置“函数”选项观察对连续变量的影响





也可以用分布图观察分类变量对分类变量的影响





#### T检验及方差分析

T检验和方差分析用来观察分类变量对连续变量的影响，在SPSS Modeler中，通过使用“均值”节点，可以方便地实现T检验和/或方差分析的效果。

独立样本T检验/方差分析

选择“在字段的组之间”，并设定分组的分类字段和测试的连续字段，连续字段可以有多个

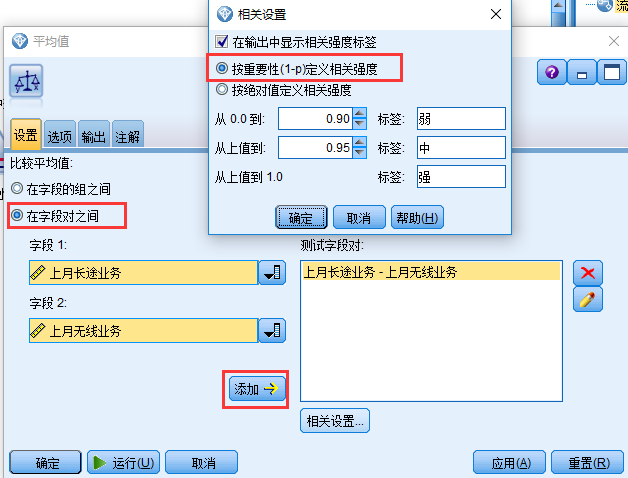




SPSS Modeler中以(1-P)值来描述显著性，显著性>0.95为通过显著性检验

配对样本的T检验

通过设置均值节点的“在字段对之间”选项，可以完成配对样本T检验





相关：0.064表示相关强度r值，r<0.3，表示不相关，强表示P值，显著性强

### 【知识点2】SPSS Modeler 数据分析--相关分析

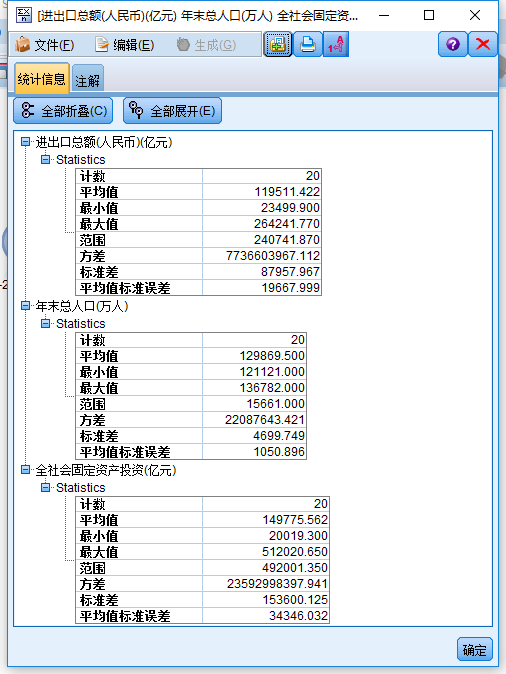
#### 相关分析

相关分析用来观察连续变量对连续变量的影响，在SPSS Modeler中，通过使用Statistics节点，可以完成对连续变量的统计量观察和连续变量之间的相关关系判定。

功能：输出选项卡——Statistics节点

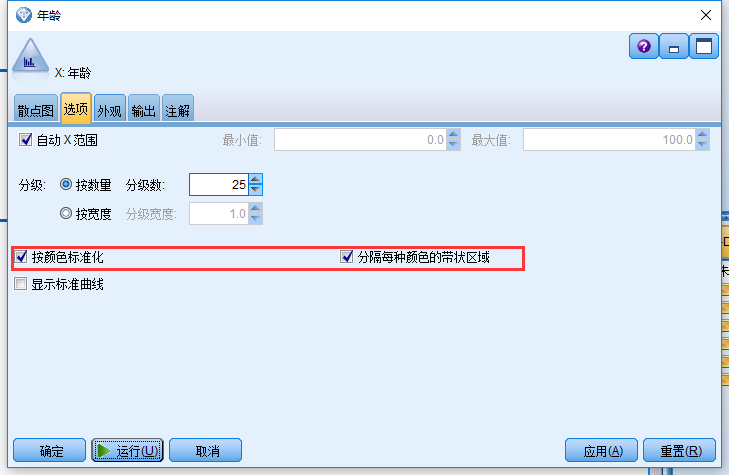
数据源：1995-2014GDP.xlsx

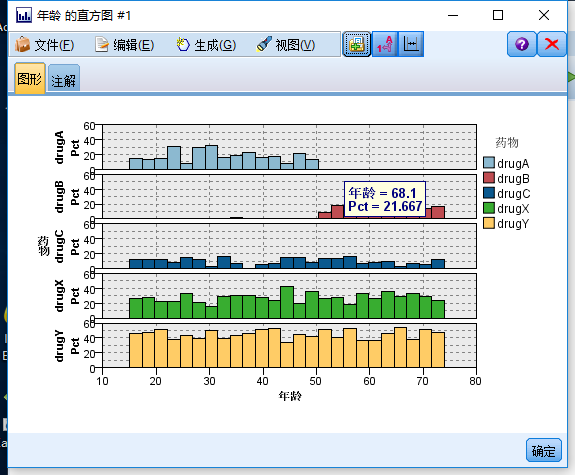




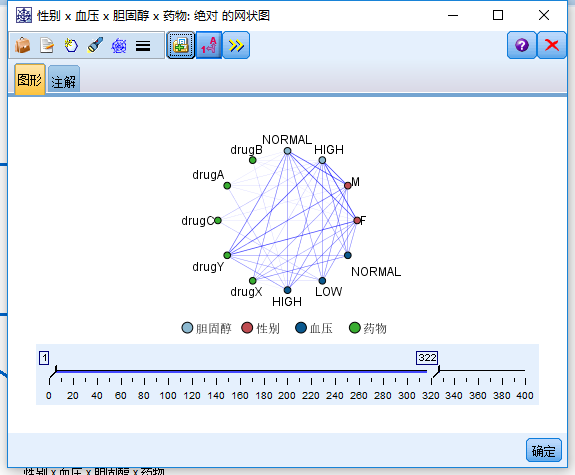
利用带颜色叠加或按分类变量分组的直方图，也可以观察连续变量与分类变量的之间的关系

数据源：历史数据，查看年龄和药物的关系





使用网络图，可以观察多个分类变量之间的关系



### 【知识点3】SPSS Modeler 数据分析—线性回归

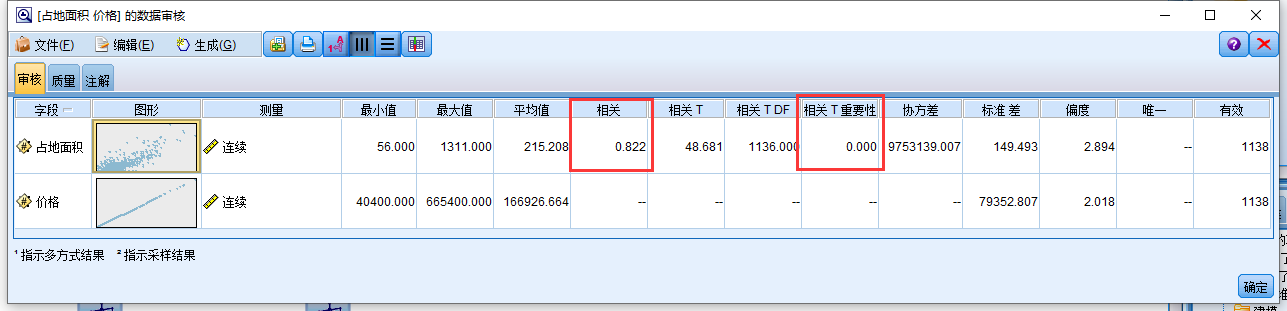
一元线性回归

形如Y=ax+b的方程用来描述因变量Y和自变量X之间的关系，通过模型中的“回归”节点完成。

设置目标



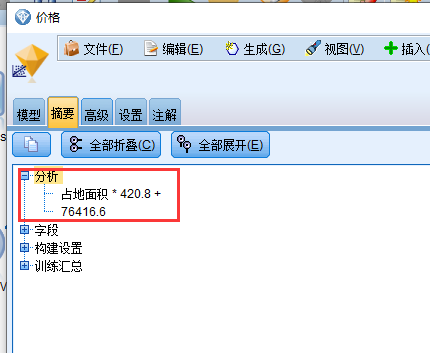
全面观察数据



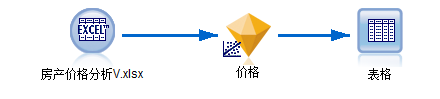
调用回归模型



模型结果

y=420.8x+76416.6

预测：



## 自主学习作业讲解

无

## 课程总结

* SPSS Modeler 数据分析--常见差异性分析
* 列联表与卡方检验
* 矩阵节点的使用
* T检验及方差分析
* 均值节点的使用
* SPSS Modeler 数据分析--相关分析
* 相关分析
* Statistics节点的使用
* SPSS Modeler 数据分析—线性回归

## 下次自主学习任务布置

#### 观看预习视频

第2章 模型篇

#### 课后作业

见作业文档