# 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2

## 明确本次课知识点，明确重点难点

### 【知识点目标】

* T检验
* 单样本T检验
* 独立样本T检验
* 配对样本T检验

### 【重点】

* T检验
* 单样本T检验
* 独立样本T检验
* 配对样本T检验

### 【难点】

* T检验
* 单样本T检验
* 独立样本T检验
* 配对样本T检验

## 复习巩固作业讲解

无

## 本次课程任务讲解

### 【知识点1】T检验

1. t检验基本思想和类型

T检验：用来观察分类变量和连续变量之间的关系，主要用来了解两个（两组）连续变量之间的差异性

两组连续变量是否属于同一总体

前置：假设检验--小概率反证法

假设检验的基本思想是小概率反证法思想。

小概率思想是指小概率事件（P<0.01或P<0.05）在一次试验中基本上不会发生。

反证法思想是先提出假设(检验假设H0)，再用适当的统计方法确定假设成立的可能性大小，如可能性小，则认为假设不成立，若可能性大，则还不能认为假设成立。

P值：用来检验小概率事件是否发生，如果p值<指定的边界，则意味着小概率事件发生，在t检验中，就意味着样本不属于同一总体

基本思想：先建立一个关于样本所属总体的假设，考察在假设条件下随机样本的特征信息是否属于小概率事件，若为小概率事件，则怀疑假设成立有悖于样本提供的特征信息，因此拒绝原假设。

H0：原假设

H1：备择假设

P<=边界值，认为原假设不成立，接受备择假设

1. 边界值，接受原假设

常见的P值一般为0.1,0.05,0.01

在t检验中，p值的意义是“大同小异”

P值<=边界值意味着“差异性显著”

T检验中，主要针对样本/总体的均值进行检验/比较，观察均值是否来自同一总体

### 【知识点2】单样本T检验

比较一个已知总体和一个从未知总体中抽取的样本之间的关系（差异性）

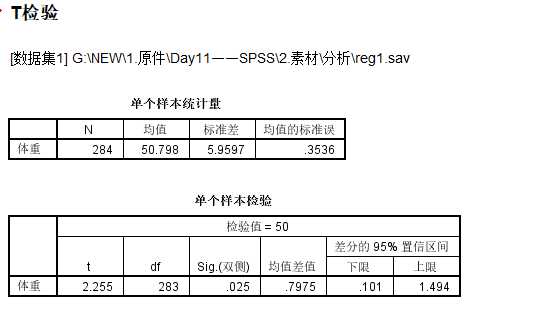
前置条件：

* 1. 要有一个已知的总体均数，总体均数：50
  2. 可以获得样本的均数和标准差（数据是连续性变量）
  3. 样本来自正态或近似正态的总体

操作步骤：分析--比较均值--单样本t检验



结果解读：



观察sig值，sig< 0.05，说明差异性显著

T=(均值-检验值)/（标准差/根号N）

总结：单一样本T检验：是用来比较一组数据的平均值和一个数值有无差异。例如：你选取了5个人，测定了他们的身高，要看这五个人的身高平均值是否高于、低于还是等于170cm,就需要用到这个检验方法

### 【知识点3】独立样本T检验

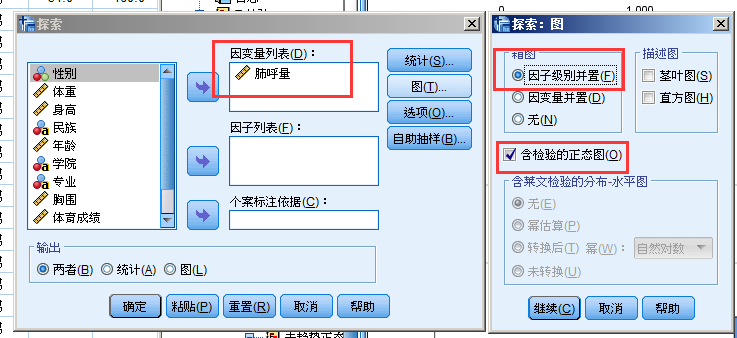
从两个不同的总体中分别取得一个样本，然后比较两个总体之间是否存在差异

前置条件：

* 1. 两组变量都要是连续变量
  2. 独立性
  3. 观测变量中不能有显著的异常值
  4. 观测变量要符合或近似符合正态分布
  5. 两组观测变量方差要对齐

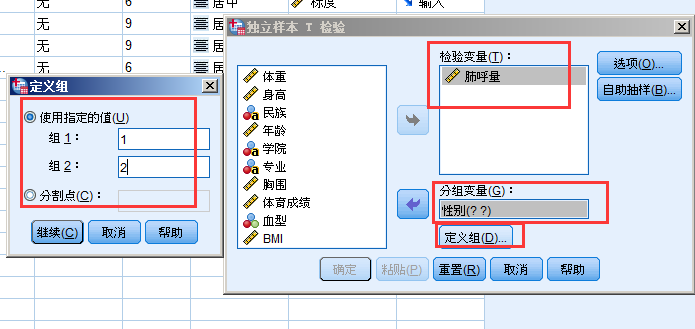
操作步骤：数据源：ttest1.sav

* 1. 通过探索过程观察异常值和检查正态性

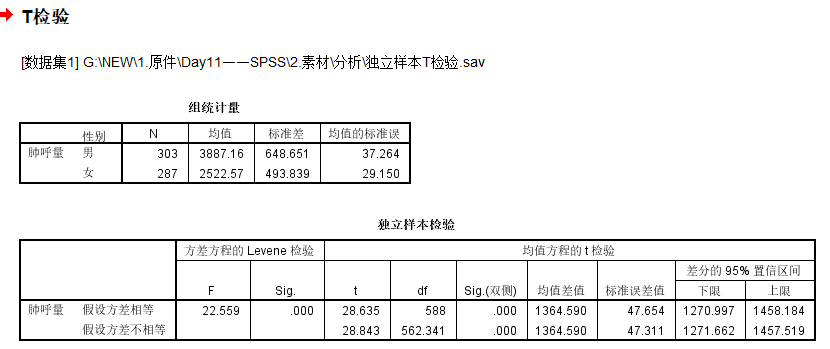


也可以根据性别来分组查看肺呼量

(2)分析--比较均值--独立样本t检验，注意组的设定



结果解读：

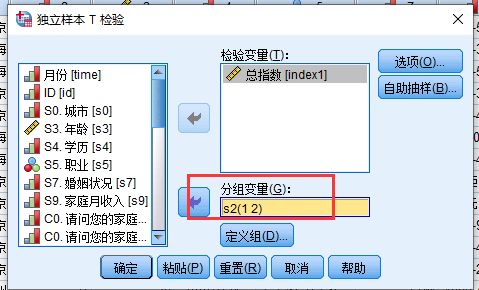
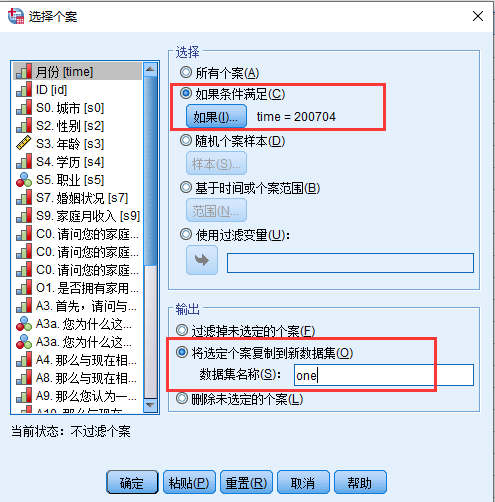


注意方差齐与方差不齐时如何观察

Sig>0.05,意味着方差相同，当前sig=0.000<0.05,意味着方差不等

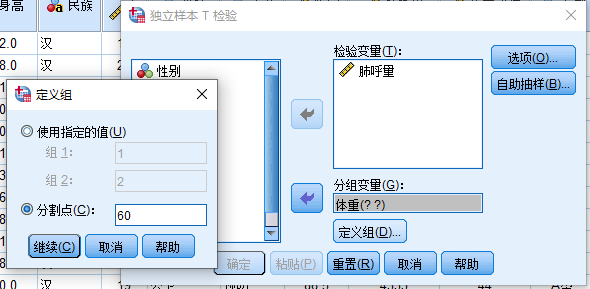
意味着男生和女士的肺活量存在差异，男生的肺活量比女士的肺活量更大

对CCSS数据针对200704不同性别的信心值指数的差异



对ttest数据针对不同体重的肺活量的差异





总结：独立样本T检验：用来看两组数据的平均值有无差异。比如：选取了5男5女，想看男女之间身高有无差异，这样，男的一组，女的一组，这两个组之间的身高平均值的大小比较可用这个方法

### 【知识点4】配对样本T检验

处理分组的箱图：分类变量对连续变量的影响（分组的连续变量分布情况）

范例：针对CCSS\_Sample.sav, 用探索过程考察不同城市受访者年龄的分布情况。

针对配对前后的总体获取样本，比较两个样本之间的差异是否显著

前置条件

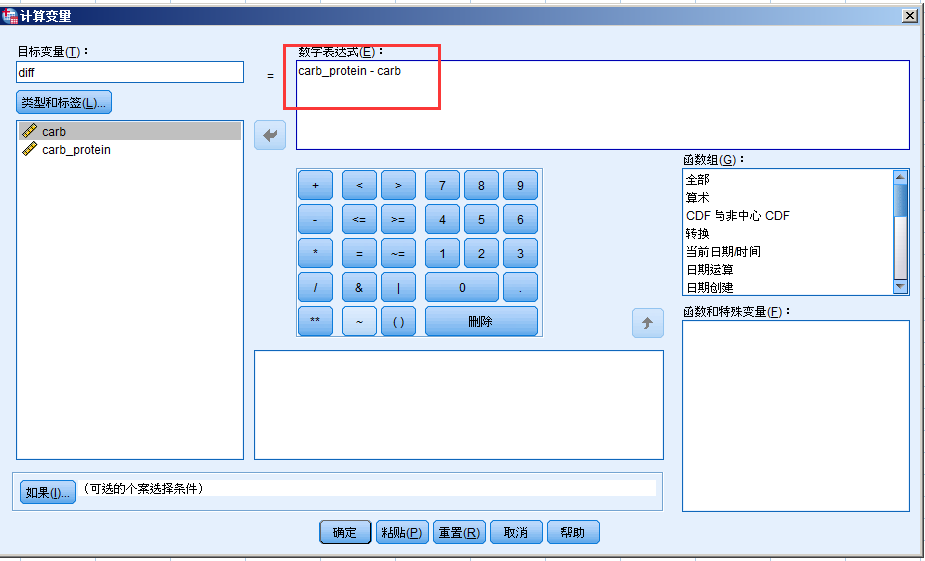
* 1. 连续变量
  2. 自变量要是二分类变量，并且要具有相关性
  3. 两组之间的因变量差值没有显著的异常值
  4. 两组之间的因变量差值要符合或近似符合正态分布

二分类变量即为那些结局只有两种可能性的变量，如有效与否，心梗，心血管不良事件，死亡等，一般将发生事件的人数除以样本量总数得到的事件发生率作为结局考察

数据源：配对样本T检验，carb普通饮料跑步的距离，carb\_protein运动饮料跑步的距离

操作步骤

* 1. 通过计算变量获得差值

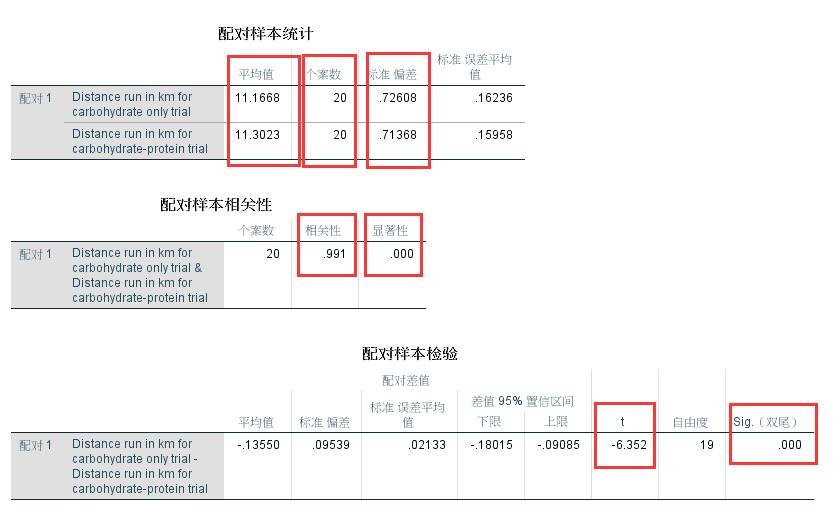


* 1. 检查差值的异常值情况和正态拟合（探索分析）
  2. 分析--比较均值—配对样本t检验



注意配对变量的设定

结果解读：



相关性=0.991，高度相关，显著性=0.000<0.05,说明这两个变量具有相关性

存在显著性差异。

平均值= -0.13550=普通-运动，说明喝运动饮料的确有助于改善跑步效果

配对样本T检验：用来看一组样本在处理前后的平均值有无差异。比如：选取了5个人，分布在饭前饭后测量了他们的体重，想检测吃饭对他们的体重有无影响，就需要用到该方法。

注意：配对样本T检验要求严格配对，也就是说，每一个人的饭前体重和饭后体重构成一对

总之：选取哪种T检验方法是由数据特点和结果要求来决定的

T检验会计算出一个统计量来，也就是T值

SPSS会根据T值来计算sig值，可以认为T值是一个中间过程产生的数据，只要看sig值就可以了。

如果sig<0.05，我们认为平均值相等的几率很小，差异显著，从而认为两组数据之间的平均值是不相等的

## 自主学习作业讲解

无

## 课程总结

* T检验
  + T检验的基本思想和分类
    - 假设检验
    - 显著性（P值）
    - 分类：单样本t检验、独立样本t检验，配对样本t检验
  + 单样本t检验
    - 前置条件
    - 操作步骤
    - 结果解读
  + 独立样本t检验
    - 前置条件
    - 操作步骤
    - 结果解读
  + 配对样本t检验
    - 前置条件
    - 操作步骤
    - 结果解读

## 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