统计作业3

1. **读书笔记**

假设检验的步骤：

构造假设（原假设、备择假设）-> 选定显著性水平 –> 计算检验统计量 ->1.p值法 2.临界值法 -> 是否拒绝原假设

原假设：

1. 当存在试验，验证新系统效果是否强于旧系统时，原系统结果为原假设。
2. 检验声明有效性时，原事实就是原假设。

备择假设：

1. 当存在试验，验证新系统效果是否强于旧系统时，新系统的研究假设为备择假设。
2. 检验声明有效性时，与原事实相反的假设就是备择假设。

两类错误：分为第一类错误和第二类错误。

1. 第一类错误：当H0事实上为真时，此时样本数据得出HO为假的结论，

做出拒绝H0的判断。

1. 第二类错误：当Ha事实为真时，即H0为假时，此时样本数据得出H0

为真的结论，做出接受H0的判断。

P值：

计算检验统计量的得出的概率，用来衡量样本数据对原假设的支持程度，p

值较小，样本数据对原假设的支持程度较低，导致拒绝H0；p值较大，样本数据对原假设的支持程度较高，通常不能拒绝H0。P值也可以称为观测显著性水平，根据与显著性水平的比较，判断p值是否足够小到可以做出拒绝H0的判断。

临界值：

拒绝原假设的最大的检验统计量。

不论是p值法还是临界值法都要区分是哪一种检验，下尾检验、上尾检验、双尾检验。

常用的假设检验及例子：

**单个总体均值的假设检验：**分为已知；未知。已知时，检验统计量服从正态分布；未知时，检验统计量服从t分布。

1. **题目分析**

图1 2011-2014年销售额前十的州排名

从图1中可以得知2011-2014年，订单销售额最高的州为NY，销售额总额达到167867元，销售额是第二名CA州的两倍，说明NY州对产品需求远大于其他州。统计出销售额排名前10州，便于了解企业订单主要流向。

做法：做行为state，列为totalprice求和项的数据透视表，对数据透视表按totalprice降序排序，同时筛选出totalprice前十名的state。根据表绘制条形图。

图2 totalprice分布直方图

图2纵向比较各年度及合计totalprice的频数分布。图2横向对各组左图频数最高的密集区域进行细分，分析其范围的频数分布。从图2可以看出，2013、2014年销售额主要分布在100元内，4年销售额也是主要分布在100元以内。2013、2014年及合计年度的销售额100元内，销售额大多为0到40。

图3 2011-2014年订单统计与销售额的条形图

从图3可以看出：2011年订单数最少，整个年度只有1单，2013年订单数最多，总额达到7931单。2011-2013年，订单数逐年增多，而2014年订单小于2013年。2013年销售额最高，从2011年-2013年，销售额逐年增加，2012-2013年变化幅度很大，而2014年比2013年销售额低。

随机抽出500个样本，使用样本均值进行假设检验。

原假设：均值 = 69.97



显著性水平= 0.05，因为p值（双尾检验）> 0.05，因此不能拒绝原假设（均值 = 69.97）