

---

例: 一个自动洗衣机生产商为一款产品提供了三种颜色: A, B 和C. 在卖出的100 该洗衣机中有40台颜色为A. 据此能否得出消费者喜欢颜色A的结论?

解:记 $p$ 为消费者喜欢颜色A的概率,  $X$ 表示100名消费者中喜欢A颜色的人数. 如果消费者对颜色没有偏好, 则 $p = 1/3$ , 否则 $p > 1/3$ . 因此可以使用二项检验来检验消费者是否偏好A颜色.

1. **Hypotheses:**  $H_0 : p = 1/3 \leftrightarrow H_1 : p > 1/3$

2. **Assumptions:** 在零假设下,  $X \sim B(100, 1/3)$

3. **Test statistic:** 检验统计量为 $X$ , 因 $n = 100$ , 故可以使用正态逼近方法近似 $p$ 值. 由于在 $H_0$ 下

$$z_{obs} = \frac{X_{obs} - np - 0.5}{\sqrt{np(1-p)}} = \frac{39.5 - 100/3}{\sqrt{100(\frac{1}{3})(\frac{2}{3})}} = 1.3081$$

大的 $z_{obs}$ 表明倾向于 $H_1$ .

4. **p value:**  $p = P(X \geq 40) \approx P(Z \geq z_{obs} | H_0) = 0.0954$ .

5. **Decision:** 由于 $p$ 值= 0.0954 > 0.05, 我们不能拒绝零假设 $H_0$ , 故接受零假设, 可认为消费者对颜色A没有偏好. 但由于 $p$ 值仅0.0954, 支持零假设的证据并不非常强烈.

in R

```
> n=100
> p0=1/3
> x=40
> binom.test(x,n,p0,alt="greater",0.95)

> z10=(x-0.5-n*p0)/sqrt(n*p0*(1-p0))
> sp=x/n
> z20=(sp-p0)/sqrt(p0*(1-p0)/n)
> p.value.exact=pbinom(x-1,n,p0,lower.tail=F) # exact p-value
> p.value.norm=pnorm(z10,lower.tail=F) #normal approx.
> p.value.ztest=pnorm(z20,lower.tail=F) #z.test
> c(sp,z10,z20)
[1] 0.400000 1.308148 1.414214
> c(p.value.exact,p.value.norm,p.value.ztest)
[1] 0.09662307 0.09541163 0.07864960
```

注:

1. 因为 $p$ 的95%置信区间包含了 $p_0 = 1/3$ , 故接受 $H_0$
2. binom.test报告的 $p$ 值是精确值, 其不同于正态逼近和大样本 $z$ 检验得到的 $p$ 值
3. 当 $n > 25$ 时, $z$ -test 检验统计量为

$$z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$$

---

## 习题

本次网上任务被抽中的同学有：滕志伟，周雨霖，武悦娇，陈柄任，周昞泽，赵娜，江浦平，张逸凡，杨紫旋，李凯，余德美，崔笛。请登录数据科学网对非参数统计课程的第二章的内容进行更新。提交方法点击相应栏目名称边的“编辑”，使用 $\text{\LaTeX}$ 输入数学符号和公式，可以参考其他课程内容“编辑”下的代码命令。

完成作业后请发送到：[nonpar.ustc@outlook.com](mailto:nonpar.ustc@outlook.com)，邮件标题格式：学号+姓名+第1次作业

1. 在一次安全月活动中，所检验的16辆车中有6辆是不安全的，试在显著性水平0.05下检验假设“这些车中有不多于10%的车是不安全的”。
2. 考虑二项分布 $T \sim \text{Bin}(n, p)$ 下的双边假设检验问题 $H_0 : p = 0.5 \leftrightarrow H_1 : p \neq 0.5$ ，求其似然比检验，若 $n = 10, T = 2$ ，给出检验的 $p$ 值。
3. 考虑二项分布 $T \sim \text{Bin}(n, p)$ 下的双边假设检验问题 $H_0 : p = p_0 \leftrightarrow H_1 : p \neq p_0$ ，求Score 检验和Wald 检验统计量并给出检验 $p$ 值。两者之间有什么关系？
4. 某学校初中三年级学生中随机调查20人的体重(单位: 公斤), 得到  
71, 67, 49, 59.5, 65.5, 51.5, 77, 61, 46.5, 68.5, 43, 59.5, 80.5,  
77, 79, 82.5, 40.5, 58.5, 64, 51.5  
试在显著性水平0.05下, 检验如下假设  
(1) 体重的中位数是51.5.  
(2) 95%分位数至少是75.  
(3) 30%分位数小于等于45.
5. 在某保险种类中, 2011年的索赔数额(单位: 万元)的随机样本为  
23.12, 23.64, 25.26, 27.42, 34.86, 37.98, 47.40, 52.37, 58.42, 63.71, 73.80, 75.06,  
93.60, 98.25  
已知2010年的索赔数额的中位数为48.32万元, 试在显著性水平0.05下检验2011年索赔数额的中位数是否比前一年有所变化。
6. 对两个相距50英里的机场进行为期一年的观测, 来研究他们因为天气原因而暂停航班服务之间是否存在显著差异。有286天两家机场均开放, 有62天因天气原因部分时间关闭, 有14天A机场关闭而B机场开放, 而有3天B机场关闭而A机场开放。因为天气原因, 两家机场开放的时间之间是否存在显著差异？
7. 光顾某食品店的22名顾客品尝甲和乙两种点心并选出喜欢的品种. 7名顾客喜欢甲品种, 12名顾客喜欢乙品种, 3名顾客没有特别偏好. 问在显著性水平0.05下, 这能否说明顾客有明显偏好差异？