📕 假设检验介绍

ch08 假设检验介绍

sec8.1 假设检验的逻辑

Definition

假设检验的定义:

通过样本数据来定性判断一个关于总体参数的假设是否成立

假设检验的逻辑流程

1. 提出一个关于总体某参数的假设.
2. **抽样之前的预设**: 如果样本来自总体, 那么样本的测量值和总体会存在差异, 但是差异不大.
3. 抽样实验过程,抽样方法选择, 抽样数据计算
4. 根据统计数据对假设进行判断. 根据第二步的预设, 统计值应该与总体对应值差别不大, 如果差距较大, 样本不嫩个很好的代表总体.

在整个逻辑流程中,计算只是很小的一部分内容. 有计算机协助, 统计结果几行代码就可以完成. 关键是要完全理解这个流程在干什么, 有什么需要注意的, 针对结果如何进行决策的问题.

总体参数会变化吗?

针对一个总体, 在某个时间点, 参数是不会变得. 但是加上时间轴, 参数就可能会发生改变.

比如从1950年到 1990年, 整个人口总体中作为参数的身高均数发生改变. 我们从1950,1990年随机抽取部分人测量升高, 两个样本的身高均值应该可以反应处总体的变化趋势.

人口方面的数据是可以通过普查的方法来获取. 但是有些总体在现实中是不存在的.

例如某制药公司开发了一种抗病毒感冒药. 公司想要评估药物是否对所有人都有效果. 那么这个总体就是未知的, 因为不可能对所有人都服用药物来检测效果. 服用药物的总体是否存在, 答案否定:不存在这样的总体. 只能假设一个服用药物的总体. 药效的总体参数怎么获取? 这就是假设检验要回答的问题.

随机选取多人试验药物效果,如果试验中样本容量比较大, 根据假设检验逻辑的第二条预设. 如果样本容量大, 样本的测量值会和总体差不多. 服用药物的样本会近似代表服用药物的总体. 如果药物对试验样本中的人有效果, 相应的, 对这些人代表的总体里所有的人都有效果.

总结为这样一个关系: **样本是否能够代表总体?如果样本能代表总体, 那么样本发生的变化也会出现在总体上.**

样本统计量的显著变化揭示出总体可能也发生同样的变化.

这里之所以下结论的语气很保守, 原因在于样本信息并不是总体的所有信息. 结论是基于已知的观察信息. 对于没有获取的信息在表述和决策时都要注意区分.

建立假设检验的实例

Example

example 1 在认知科学研究中,有学者发现对大脑某些区域轻微的电刺激会提高认知能力.

已知研究指出大脑的顶叶区域是与数学学习区域.

所以科学家设计实验项目:针对顶叶的电流刺激是否会提高数学符号的学习能力的研究.

实验方案和结果:

1. 随机选取两组志愿者参与实验研究
2. 随机分为两组, 一组用电流刺激大脑顶叶部位,一组用电流刺激其他脑区
3. 两组志愿者学习同样单位数学符号内容
4. 六个月以后检查两组志愿者对符号学习的能力, 发现刺激顶叶的这一组成绩优于刺激其他区域的组.

只针对一个学习项目的结果可能还难以得出令人信服的结论. 数学学习还包括其他方面的能力. 所以研究者想知道对顶叶的电刺激是否会提高数学标准化考试的成绩.

根据以前的统计资料, 学生总体在标准化考试里的参数为:平均分:$\mu = 80$ 标准差:$\sigma = 20$, 分数为正态分布.

研究者随机选取$n = 25$ 的学生进行试验, 被试者每天学习数学 30分钟,并且接受顶叶到电刺激, 历时 4 周,之后参加数学标准化考试.

$n = 25$ 个学生, 数量不算太多, 但是属于正态分布, 大致可以代表这个总体.会存在一些差异.

如果经过试验后的学生数学成绩平均分显著高于总体平均分时,可以得出结论, 电刺激对数学学习有帮助. 如果测试成绩平均分和总体均数差别不大, 那么可以认为电刺激对数学学习可能没有帮助

试验后的总体得分是未知的, 因为不可能对所有的学生都进行电刺激. 所以这个经过电刺激的总体是假设的.

下面提出实现假设检验的步骤

1. 第一步: 提出假设

   假设检验中提出两个完全相反的假设:虚无假设(零假设), 备择假设. 在一个假设检验中只能接受其中一种假设. 虚无假设(Null hypothesis, H_0): 假想总体的值和总体参数没有明显区别. 在本试验中虚伪假设为学生大脑电刺激的数学考试平均分和总体均分没有太大区别.

$$H_0 : \mu_{有刺激} = 80$$

备择假设(alternative hypothesis,H_1, H_a): 在本试验中, 备择假设描述为:大脑经过电刺激的学生数学考试成绩平均分和总体平均分明显不同.

$$H_1 : \mu_{有刺激} \neq 80$$

两种假设互斥.

2. 第二步:为下结论定下标准

根据零假设, 设定了假想的总体. 后续的的步骤就围绕这个假想的总体展开. 现在总这个假想总体出发, 进行抽样实验, 依照大数定律和中心极限定律, 均数的抽样分布为正态分布,当样本容量为$n = 25$时反复多次抽样实验获取的平均分数绝大部分机会出现在假想总体平均分附近.只有很少机会远离假想总体均值.

当一个样本均值远远偏离抽样均值分布中心的时候, 可以认为这个样本内的个体可能并不是来自于假想的总体, 而是从其他总体中抽取的. 现在需要为这个结论提供定量的依据

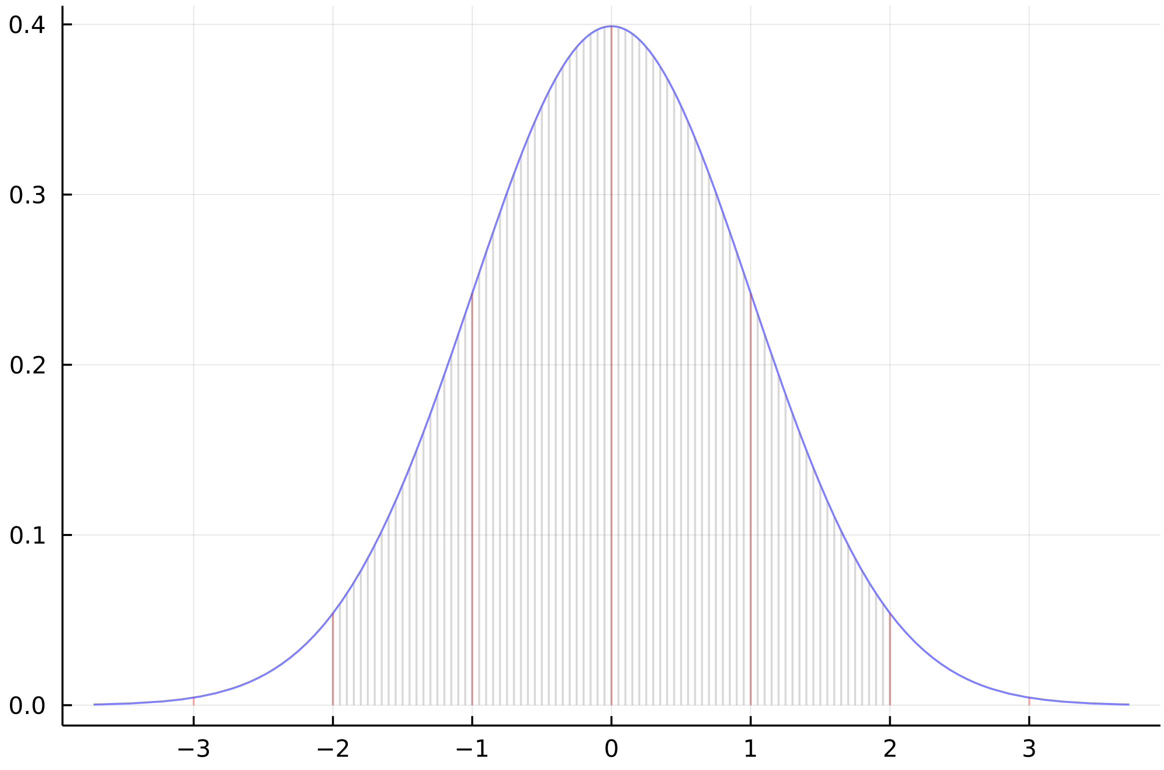
α 水平

常见的的α 水平有$0.05, 0.01, 0.001$ 以$\alpha = 0.05, 95\%$ 为例.见下图, 在标准正态分布中, 有5%的区域远离中心区域,大约在两个标准差以外. 样本均值落在这个区域的机会非常小, α 水平量化了判断零假设的依据. 如果样本均值离假想总体积均值超过两个标准差, 就认为样本个体来自于另外一个总体.

确定好标准,就可以进行统计计算

2

```
@bind s Slider(range,default=2,show_value=true)
```




```
begin
 range2=-s:0.05:s   # 用户绘图细绿线, 间隔小
 val=[pdf(d,σ) for σ in range]
 cd=[pdf(d,i) for i in range2]
 plot(d,label=false,alpha=0.5,color=:blue,xticks=(-3:1:3) )
 plot!(repeat(range', 2), [zeros(1, n); val'], label = "", color = :red, alpha = 0.3,lw=1)
# plot!([s,s],[0, pdf(d,s)],label=false,color=:green, alpha=0.8,lw=2)
 plot!(repeat(range2', 2), [zeros(1, length(range2)); cd'], label =false, color = :gray, alpha = 0.3)

end
```

3. 第三步: 收集数据,计算样本统计量

   根据实验设计, 对每个学生个体进行脑区电刺激实验, 连续实验 4 周后, 收集学生的数学考试成绩,进行统计计算

   获取学生数学考试成绩平均值, 然后转换为 z分数用于结论判断

   z分数计算公式为:

$$z = \frac{\text{样本均值} - \text{假设总体均值}}{\text{均值标准误}}$$

4. 第四步: 下结论

根据实验统计, 接受电刺激的学生样本数学考试成绩均值为:

$$M = 89$$

总体参数为:

$$\mu = 80, n = 25, \sigma = 20$$

样本均值标准误为:

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = 4$$

样本均值$M = 89$ 对应的Z 分数为:

$$z = \frac{M - \mu}{\sigma_M} = \frac{9}{4} = 2.25$$

当α 设定为0.05时, Z 分数超过了 1.96 因为从假定总体($\mu = 80$)抽取到均值为89 的概率 极低. 可以认为样本来自于一个不同的总体. 拒绝零假设接受备择假设的. 电刺激导致样本均值高于假想总体的均值, 并且这种差异不能由抽样误差解释.

如果样本均值为$M = 84$, z分数为:

$$z = \frac{M - \mu}{\sigma_M} = \frac{4}{4} = 1.00$$

得到的 z分数为1.00, 那么 z分数位于假设总体均值的 1 个标准差之内, 抽样均值偏离假想总体均值的原因可能是因为抽样误差导致的. 所以接受零假设. 电刺激对数学学习没有影响

也就是如果样本均值落在假设总体均值两个标准差以内, 认为差异是由抽样误差引起的. 如果样本均值比假设总体均值高两个标准差, 差异可能不是由抽样误差引起的, 而是总体参数发生了变化.

就是一个假设检验的推理过程. 可能需要反反复复看多次才能比较好的理解这个流程.

sec8.2 假设检验中的不确定性和误差

由于假设检验使用的样本数据, 只能提供有限的关于总体参数的估算. 所以有可能会出现推断错误, 这是由概率决定的的不确定性. 假设检验中有有可能会犯两种错误.

第一类错误:

针对上上面的实验, 试验之前的学生总体是符合正态分布, 从中随机选择的$n = 25$ 个学生的成绩还有极大的概率出现在均值$\mu = 80$ 附近.但是这是随机随机抽取学生进行电刺激实验的时候,也可能以极小的概率抽取均值很高的一组学生. 这样以来即便是电刺激没有效果, 但是因为随机选取的学生数学成绩已经比平均值高. 我们会错误的把抽样的偶然性当做是试验效果.

> Definition
>
> 第一类错误:当实际情况零假设成立, 但是研究者拒绝零假设时就发生了一类错误.

例如制药公司研究一种聪明药, 在研究的时候刚好抽到了一些IQ 已经很高的学生, 因为实际成绩已经很高, 即便药物没有效果, 但是确使得样本均值高于正常IQ 均值. 因为选择样本的原因导致均值明显提高, 但是原因并不是因为药物的效果.

我们选择的α 水平也严格, 犯一类错误的机会越小, 因为α 越严格,从总体中抽出这样的样本越稀少.

第二类错误:

如果电刺激确实对数学有帮助, 但是我们在检测中根据数据却没能拒绝零假设.

对于一个药物或者社会学的统计实验, 推断其有效的风险非常大, 比如对于药物, 统计实验证明有效果的结果是会导致药物应用到所有适合人群中. 所以第一类错误的风向非常大, 因此结论需要进行多次抽样实验,并且加大样本容量. 第二类错误风险较小.

7

```julia
begin
    @variables μ,σ,d
    μ,σ=0,1
    d=Normal(μ,σ)
    range=-3σ:σ:3σ
    n=length(range)
    #data=rand(d,1000)
end
```