总体均值假设检验

为什么要做假设检验?

我们在生活中经常会遇到对一个总体数据进行评估的问题，但我们又不能直接统计全部数据，这时就需要从总体中抽出一部分样本，用样本来估计总体情况。

假设检验定义

假设检验是先对总体参数提出一个假设值，然后利用样本信息判断这一假设是否成立

假设检验的假设

一种叫原假设，也叫零假设，用 H0 表示。原假设一般是统计者想要拒绝的假设。原假设的设置一般为：等于=、大于等于>=、小于等于<=。

另外一种叫备择假设，用 H1 表示。备则假设是统计者想要接受的假设。备择假设的设置一般为：不等于、大于>、小于<。

弃真错误、取伪错误

弃真错误也叫第I类错误或α错误：它是指 原假设实际上是真的，但通过样本估计总体后，拒绝了原假设。明显这是错误的，我们拒绝了真实的原假设，所以叫弃真错误，这个错误的概率我们记为α。这个值也是显著性水平，在假设检验之前我们会规定这个概率的大小。

取伪错误也叫第II类错误或β错误：它是指 原假设实际上假的，但通过样本估计总体后，接受了原假设。明显者是错误的，我们接受的原假设实际上是假的，所以叫取伪错误，这个错误的概率我们记为β。

显著性水平（Significance Level）

显著性水平是指当原假设实际上正确时，检验统计量落在拒绝域的概率，简单理解就是犯弃真错误的概率。这个值是我们做假设检验之前统计者根据业务情况定好的。

显著性水平α越小，犯第I类错误的概率自然越小，一般取值：0.01、0.05、0.1等

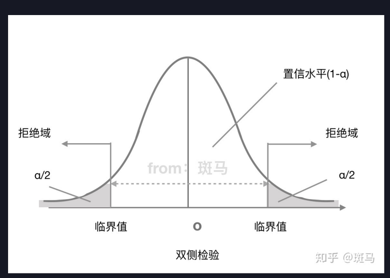
检验方式

检验方式分为两种：双侧检验和单侧检验。单侧检验又分为两种：左侧检验和右侧检验。

双侧检验：备择假设没有特定的方向性，形式为"≠"这种检验假设称为双侧检验

单侧检验：备择假设带有特定的方向性 形式为">""<"的假设检验，称为单侧检验 "<"称为左侧检验 ">"称为右侧检验

拒绝域



example

定义：拒绝域是由显著性水平围成的区域

拒绝域的功能主要用来判断假设检验是否拒绝原假设的。如果样本观测计算出来的检验统计量的具体数值落在拒绝域内，就拒绝原假设，否则不拒绝原假设。给定显著性水平α后，查表就可以得到具体临界值，将检验统计量与临界值进行比较，判断是否拒绝原假设。

检验统计量

样本数量大于30

- 总体标准差已知 → z = (X-μ)/[σ/√(n)]
- 总体标准差不知，用样本标准差估算总体 → z = (X - μ) / [s/√(n)].

样本数量小于30

- 总体接近正态分布并且总体标准差已知 → z = (X-μ)/[σ/√(n)]
- 总体接近正态分布并且总体标准差未知 → t= (X - μ) / [s/√(n)].

X-样本统计量
μ-总体均值
σ-总体标准差
s-样本标准差

假设检验步骤

1.提出原假设与备择假设
2.从所研究总体中出抽取一个随机样本
3.构造检验统计量
4.根据显著性水平确定拒绝域临界值
5.计算检验统计量与临界值进行比较

两种假设检验

假设检验根据业务数据分为两种：一个总体参数的假设检验和两个总体参数的假设检验

一个总体参数的假设检验：只有一个总体的假设检验

举个例子：学而思App原版本1转化率为 19%，学而思App版本2开发完成后，直接全量发布整体上线，过一段时间后统计转化率为27%，我们想判断版本2是否比版本1好，这时我们做的假设检验总体只有1个，全部用户。对于总体只有一个的称为一个总体参数的假设检验。

两个总体参数的假设检验：有两个总体的假设检验

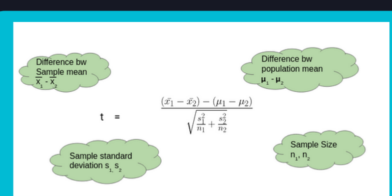
同样的例子：学而思App版本1和学而思App版本2同时上线，流量各50%，这时我们做的假设检验总体有2个，分别为命中版本1的全部用户与命中版本2的全部用户。

两个总体的假设检验：比较均值

Two-Sample t-Test：We do not know the
1.population variance
2.Our sample size is small, n < 30

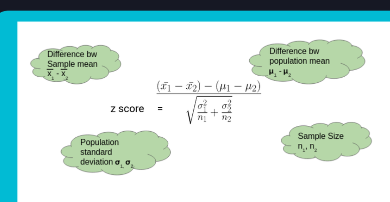
Two Sample Z Test：
1.We know the population variance, or
2.We do not know the population variance but our sample size is large n ≥ 30

Two sample t test



Summary

Two sample z test



Summary

总体成数的假设检验

一个总体成数的检验方法

在前面的篇章中我们知道，当 n 较大时，np>5，同时 n(1-p)>5，则$\hat{p}$的分布近似正态分布。因此，在检验一个总体成数时，可以构造 Z 统计量，即

$$Z=\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\sim N(0,1)$$

例子例1：某集团公司欲出台一项政策，有人估计支持率可达到80%以上。现随机抽取230人进行询问，其中有190人表示支持。试问在α>0.05的显著性水平下，对该支持率的估计是否可信?

解：本例是关于总体成数的检验问题，题中对p是否大于0.8感兴趣。因此，将p>0.8设为备择假设，即H0:p=0.8;H1:p>0.8。

已知n=230，是大样本，且np，n(1-p)皆大于5.所以可构造Z统计量

$$Z=\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\sim N(0,1)$$

式中，$\hat{p}=190/230\approx0.826$。将数据代入上式，即得

$$Z=\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}=\frac{0.826-0.8}{\sqrt{\frac{0.8(1-0.8)}{230}}}\approx0.986$$

两个个总体成数的检验方法

如果两个样本是独立的，同时 n_1p_1, $n_1(1-p_1)$大于 5 以及 n_2p_2, $n_2(1-p_2)$也大于 5，则$\hat{p}_1-\hat{p}_2$的抽样分布就近似于正态分布。可构造 Z 统计量，即

$$Z=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1}+\frac{p_2(1-p_2)}{n_2}}}\sim N(0,1)$$

在上式中，包含待检验参数p1-p2，所以Z统计量可用p1-p2=0的检验。

$$\bar{p}=\frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$$

则可以得到新的统计量 Z：

$$Z=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\bar{p}(1-\bar{p})(\frac{1}{n_1}+\frac{1}{n_2})}}\sim N(0,1)$$

但是如果H0:p1-p2=0，即p1=p2，则上式中的p1和p2就要按相等的比率计算。但此刻p1和p2都是未知的，所以要通过样本成数进行估计，即设`p=p1-p2，其中`p有下式计算得出:

例题

例2:随机调查了339名50岁以上的男性，其中205名吸烟者中有43人患慢性支气管炎；在134名不吸烟者中，有13人患慢性支气管炎。试在α>0.05时，检验吸烟者患此病的比率是否明显高于不吸烟者?

解：本例是关于两个总体成数差值的检验问题。设吸烟者为总体1，不吸烟者为总体2，根据题意可设

H0:p1-p2=0;H1:p1-p2>0

由题中知道，n1=205,n2=134是大样本。

$$\hat{p}_1=\frac{43}{205}\approx0.2098,\ \hat{p}_2=\frac{13}{134}\approx0.097$$

$$\bar{p}=\frac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}=\frac{43+13}{205+134}\approx0.165$$

因为 $n_1\bar{p}$, $n_1(1-\bar{p})$, $n_2\bar{p}$, $n_2(1-\bar{p})$都大于 5，所以可用 Z 统计量

$$Z=\frac{(\hat{p}_1-\hat{p}_2)-(p_1-p_2)}{\sqrt{\bar{p}(1-\bar{p})(\frac{1}{n_1}+\frac{1}{n_2})}}=\frac{0.2098-0.097}{\sqrt{0.165(1-0.165)(\frac{1}{205}-\frac{1}{134})}}\approx2.74$$

已知α=0.05，是右侧检验，查标准正态分布表，得Z0.05=1.64。因为Z=2.74>1.64=Z0.05，落在拒绝域，所以拒绝H0，接受H1，即吸烟者患慢性气管炎的比率明显的高于不吸烟者。