

# 假设检验基础

# 统计推断

## statistical inference

内容：



1. 参数估计  
(estimation of parameters)

包括：点估计与  
区间估计

2. 假设检验 (test  
of hypothesis)

由样本信息推断总体特征，除参数估计外，还会遇到这样的问题：

某一样本均数是否来自于已知均数总体？两个不同样本均数是否来自均数相同的总体？

要回答这类问题，更多的是用统计推断的另一方面 —— 假设检验(hypothesis test)。

本章涉及内容：

- 假设检验的概念及原理
- 假设检验的基本步骤
- $t$  检验方法
- 二项分布与poisson分布资料的 $Z$ 检验
- 假设检验的功效

# 第一节 假设检验的概念及原理

## 一、假设检验的概念：

一般科研程序： 假说---验证---对假说作出结论

统计上的假设检验：

**假设检验亦称为显著性检验，是判断样本指标与总体指标或样本指标与样本指标之间的差异有无显著性意义的一种统计方法。**

❖ 假设检验的原理：假设检验的基本思想是反证法和小概率的思想。

❖ 反证法思想：首先提出假设（由于未经检验是否成立，所以称为无效假设），用适当的统计方法确定假设成立的可能性大小，如果可能性小，则认为假设不成立，拒绝它；如果可能性大，还不能认为它不成立。

❖ 小概率思想：是指小概率事件在一次随机试验中认为基本上不会发生。

概率小于多少算小概率是相对的，在进行统计分析时要事先规定，即检验水准 $\alpha$ 。

# 假设检验的基本步骤

## 实例分析：

某医生测量了36名从事铅作业男性工人的血红蛋白含量，算得其均数为130.83g/L，标准差为25.74g/L。

## 问：

从事铅作业工人的血红蛋白是否不同于正常成年男性平均值140g/L？

# 造成两均数差别的原因有两种：

①两均数来自同一总体，其差别完全是抽样误差引起。

差别无显著性(差别无统计学意义)

②两均数并非来自同一总体，其差别不仅仅由抽样误差引起，而是本质不同。

差别有显著性(差别有统计学意义)

统计上通过假设检验 (hypothesis test) 或称显著性检验 (significance test) 来区分这两种原因。



# 假设检验的统计思想特点

---

- ① 统计检验的假设是关于总体特征的假设;
  - ② 用于检验的方法是以检验统计量的抽样分布为理论依据的;
  - ③ 作出的结论是概率性的, 不是绝对的肯定或否定。
-

# 步骤一：建立检验假设，确定检验水准

## ■ 1.两种假设：

无效假设（null hypothesis）：符号 $H_0$ ，一般设为某两个或多个总体参数相等，即认为他们之间的差别是由于抽样误差引起的。

备择假设（alternative hypothesis）：符号 $H_1$ 和 $H_0$ 的假设相互对立，即认为他们之间存在着本质的差异。 $H_1$ 的内容反映出检验的单双侧。

## ■ 2.确定双侧或单侧检验

## ■ 3.确定检验水准（size of a test）：亦称显著性水准（significance level），符号 $\alpha$ 。

## 单、双侧检验对应的假设（以符号表示）：

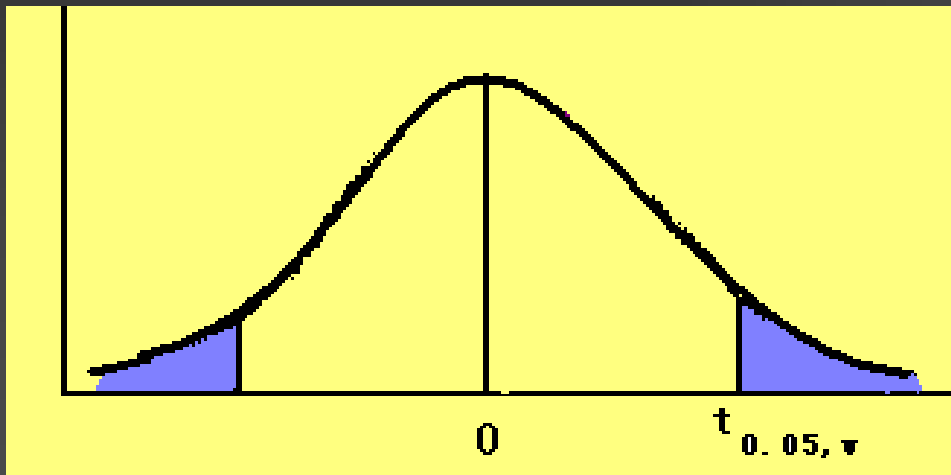
1. 样本均数（其总体均数为 $\mu$ ）与某已知总体均数 $\mu_0$ 作比较：

	目的	$H_0$	$H_1$
双侧检验	是否 $\mu \neq \mu_0$	$\mu = \mu_0$	$\mu \neq \mu_0$
单侧检验	是否 $\mu > \mu_0$ 或是否 $\mu < \mu_0$	$\mu = \mu_0$ $\mu = \mu_0$	$\mu > \mu_0$ $\mu < \mu_0$

2. 两样本均数（其总体均数分别为 $\mu_1$ 和 $\mu_2$ ）作比较：

	目的	$H_0$	$H_1$
双侧检验	是否 $\mu_1 \neq \mu_2$	$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$
单侧检验	是否 $\mu_1 > \mu_2$ 或是否 $\mu_1 < \mu_2$	$\mu_1 = \mu_2$ $\mu_1 = \mu_2$	$\mu_1 > \mu_2$ $\mu_1 < \mu_2$

- 显著性水准，符号为  $\alpha$ ，是预先给定的概率值。是判定样本指标与总体指标或两样本指标间的差异有无统计学显著性意义的概率水准，在实际工作中， $\alpha$  常取0.05。 $\alpha$  可根据不同的研究目的给予不同的设置，如方差齐性检验，正态性检验  $\alpha$  常取0.1或0.2。



## 步骤二：选定检验方法，计算检验统计量

应根据研究设计类型和统计推断目的选用不同的检验方法。

研究设计类型	检验方法	检验统计量
配对资料	t检验	t值
成组设计两样本均数比较	t检验（小样本） u检验（大样本）	t值 u值
多个样本均数比较	F检验 (即方差分析)	F值

所有检验统计量都是在 $H_0$ 成立的前提条件下计算出来的。

## 步骤三：确定 $P$ 值，作出统计推断结论

- ◆  $P$  的含义是指从 $H_0$ 规定的总体中随机抽样，其检验统计量等于及大于现有样本获得检验统计量值的概率即 $H_0$ 成立的概率。当求得检验统计量的值后，一般可通过特制的统计用表直接查出 $P$  值。

根据获得的事后概率 $P$ ，与事先规定的概率— $\alpha$ 进行比较，看其是否为小概率事件而得出结论。

### 作出推断结论：

若 $P \leq \alpha$ ，结论为按 $\alpha$ 检验水准，拒绝 $H_0$ ，接受 $H_1$ ，有统计学意义(统计结论)。可认为.....不同， .....高于.....(专业结论)。

若 $P > \alpha$ ，结论为按 $\alpha$ 检验水准，不拒绝 $H_0$ ，无统计学意义(统计结论)。还不能认为.....不同(专业结论)。

- ◆  $P$  值大小反映了两均数的差异仅是由抽样误差造成的可能性大小。

## 第二节 $u$ 检验和 $t$ 检验

### 一、样本均数与总体均数的比较

1.  $u$  检验

2.  $t$  检验

### 二、配对设计的样本资料 —— $t$ 检验

### 三、完全随机设计的两样本均数的比较

1.  $u$  检验

2.  $t$  检验

# 一、样本均数与总体均数的比较

- 目的是推断样本所代表的未知总体均数与已知总体均数有无差别.

$u$ 检验:

$$u = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

→  $\sigma$ 已知, 或 $\sigma$ 未知但样本含量 $n$ 足够大.

$t$ 检验:

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$\nu = n - 1$$

→ 样本含量 $n$ 小, 且 $\sigma$ 未知



$\bar{x}$

例：某医生测量了36名从事铅作业男性工人的血红蛋白含量，算得其均数为130.83g/L，标准差为25.74g/L。问从事铅作业工人的血红蛋白是否不同于正常成年男性平均值140g/L？

## (1) 建立检验假设，确定检验水准

$H_0: \mu = \mu_0 = 140\text{g/L}$ ，从事铅作业男性工人平均血红蛋白含量与正常成年男性平均值相等。

$$H_1: \mu \neq \mu_0,$$

$$\alpha = 0.05。$$

## (2) 计算检验统计量

$$t = \frac{\bar{X} - \mu}{S_{\bar{X}}} = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} = \frac{130.83 - 140}{25.74 / \sqrt{36}} = -2.138$$

$$\nu = n - 1 = 36 - 1 = 35$$

### (3) 确定P值，作出推断结论

$$|t| = |-2.138| = 2.138$$
$$t_{0.05/2, 35} < 2.138 < t_{0.02/2, 35}$$

$$0.05 > P > 0.02$$

按 $\alpha=0.05$ 水准，拒绝 $H_0$ ，接受 $H_1$ ，有统计学意义。  
可认为从事铅作业男性工人的平均血红蛋白含量低于正常成年男性的。

P80 例5.1

P87 例5.4

P88 例5.5

例：根据大量调查，已知健康成年男子脉搏的均数为72次/分，某医生在一山区随机调查了25名健康成年男子，求得其脉搏均数为74.2次/分，标准差为6.0次/分。

问：能否据此认为该山区成年男子的脉搏均数高于一般成年男子的脉搏均数？

## 二、配对设计的样本资料—— $t$ 检验

- 配对设计(paired design)——一是将受试对象按一定条件配成对子，再随机分配每对中的两个受试对象到不同处理组。配对的因素是影响实验效应的主要因素。非处理因素。

# 配对设计资料常见有四种：

- ① 配对的两个受试对象分别接受两种处理后的数据
- ② 同一对象（如检品）用两种方法或仪器检验的结果
- ③ 同一受试对象处理前后结果的比较。
- ④ 同一受试对象两个部位数据的比较。

- 配对设计目的是推断两种处理或方法的结果有无差别.
- 差值均数与总体均数 0 的比较

$t$  检验:

$$t = \frac{\bar{d} - 0}{s_{\bar{d}}} = \frac{\bar{d}}{s_d / \sqrt{n}}$$

$$\nu = n - 1$$

$n$ 为对子数

应用条件是 $d$ 符合正态分布

例3-6 为比较两种方法对乳酸饮料中脂肪含量测定结果是否不同，某人随机抽取了10份乳酸饮料制品，分别用脂肪酸水解法和哥特里—罗紫法测定其结果如表3-3第(1)~(3)栏。问两法测定结果是否不同？



表 3-3 两种方法对乳酸饮料中脂肪含量的测定结果(%)

编号	哥特里—罗紫法	脂肪酸水解法	差值 $d$
(1)	(2)	(3)	(4)=(2)-(3)
1	0.840	0.580	0.260
2	0.591	0.509	0.082
3	0.674	0.500	0.174
4	0.632	0.316	0.316
5	0.687	0.337	0.350
6	0.978	0.517	0.461
7	0.750	0.454	0.296
8	0.730	0.512	0.218
9	1.200	0.997	0.203
10	0.870	0.506	0.364
			2.724

## (1) 建立检验假设，确定检验水准

$H_0: \mu_d=0$ ，两种方法的测定结果相同

$H_1: \mu_d \neq 0$ ,

$\alpha=0.05$

## (2) 计算检验统计量

$$t = \frac{\bar{d} - \mu_0}{S_d / \sqrt{n}} = \frac{0.2724 - 0}{0.1087 / \sqrt{10}} = 7.925$$

$$v = 10 - 1 = 9$$

### (3) 确定 $P$ 值，作出推断结论

$$t=7.925 > t_{0.001/2, 9}=4.781$$

$$P < 0.001$$

按 $\alpha=0.05$ 水准，拒绝 $H_0$ ，接受 $H_1$ ，有统计学意义。两种方法对脂肪含量的测定结果不同，哥特里—罗紫法测定结果较高。

P102 表5-4

例6-2 某儿科采用静脉注射人血丙种球蛋白治疗小儿急性毛细支气管炎。用药前后患儿血清中免疫球蛋白IgG(mg/dl)含量如表所示。试问用药前后IgG有无变化？

**表6-1 用药前后患儿血清中免疫球蛋白IgG(mg/dl)含量**

序号	用药前	用药后	差值d
1	1206.44	1678.44	472.00
2	921.69	1293.36	371.67
3	1294.08	1711.66	417.58
4	945.36	1416.70	471.34
5	721.36	1204.55	483.19
6	692.32	1147.30	454.97
7	980.01	1379.59	399.58
8	691.01	1091.46	400.45
9	910.39	1360.34	449.95
10	568.56	1091.83	523.27
11	1105.52	1728.03	622.51
12	757.43	1398.86	641.44

假设检验步骤:

$$H_0: \mu_d = 0 \quad H_1: \mu_d \neq 0 \quad \alpha = 0.05$$

$$n = 12, \sum d = 5707.95, \bar{d} = \sum d / n = 5707.95 / 12 = 475.66$$

$$\sum d^2 = 2793182.166, S_d = \sqrt{\frac{2793182.166 - \frac{5707.95^2}{12}}{12 - 1}} = 84.2747$$

$$t = \frac{475.66 - 0}{84.2747 / \sqrt{12}} = 19.552 (mg / dl), \quad \nu = 12 - 1 = 11$$

查t 临界值表:  $t_{0.05/2,11}=2.201$

$t > t_{0.05/2,11}$ , 得  $P < 0.05$

按  $\alpha=0.05$  水准拒绝  $H_0$ , 接受  $H_1$ 。可认为用药后小儿IgG增高。

### 三、完全随机设计的两样本均数的比较

- 完全随机设计(**completely random design**) ——是将受试对象按随机化的方法分配到各个处理组中，观察实验效应。各个处理组的样本例数可以相等，也可以不等。
- 也称为成组设计或单因素设计。
- 目的是推断两样本各自代表的总体均数是否相等。根据样本含量 $n$ 的大小，分为u检验和t检验。



**u检验** ——两样本含量 $n_1$ 和 $n_2$ 均足够大，比如均大于 50 或 100

$$u = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

**t检验** ——两样本含量 $n_1$ 和 $n_2$ 均较小时

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_c^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

自由度  $v = n_1 + n_2 - 2$

例3-7 为研究国产四类新药阿卡波糖胶囊降血糖效果，某医院用40名II型糖尿病病人进行同期随机对照试验。试验者将这些病人随机等分到试验组(用阿卡波糖胶囊)和对照组(用拜唐苹胶囊)，分别测得试验开始前和8周后空腹血糖，算得空腹血糖下降值见表3-4，能否认为国产四类新药阿卡波糖胶囊与拜唐苹胶囊对空腹血糖的降糖效果不同？

表 3-4 试验组和对照组空腹血糖下降值(mmol/L)

试验组	-0.70	-5.60	2.00	2.80	0.70	3.50	4.00	5.80	7.10	-0.50
$X_1$										
$(n_1=20)$	2.50	-1.60	1.70	3.00	0.40	4.50	4.60	2.50	6.00	-1.40
对照组	3.70	6.50	5.00	5.20	0.80	0.20	0.60	3.40	6.60	-1.10
$X_2$										
$(n_2=20)$	6.00	3.80	2.00	1.60	2.00	2.20	1.20	3.10	1.70	-2.00

(1) 建立检验假设，确定检验水准

$$H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2$$

$$\alpha = 0.05$$

## (2) 计算检验统计量

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}, \quad \nu = n_1 + n_2 - 2$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2 + S_2^2}{n}}} = \frac{2.0650 - 2.6250}{\sqrt{\frac{3.0601^2 + 2.4205^2}{20}}} = -0.642$$

$$\nu = n_1 + n_2 - 2 = 2(n - 1) = 2(20 - 1) = 38$$

### (3) 确定 $P$ 值，作出推断结论

$$|t| = |-0.642| = 0.642 < t_{0.5/2, 38} = 0.681$$

$$P > 0.50$$

按 $\alpha=0.05$  水准，不拒绝 $H_0$ ，无统计学意义。  
还不能认为阿卡波糖胶囊与拜唐苹胶囊对空腹  
血糖的降糖效果不同。

P99 例5.10

P100 例5.11

例6-4 某口腔科测得长春市13—16岁居民男性20人的恒牙初期腭弓深度均值为17.15mm，标准差为1.59mm；女性34人的均值为16.92mm，标准差为1.42mm。根据这份数据可否认为该市13—16岁居民腭弓深度有性别差异？

## 检验步骤:

$$H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2 \quad \alpha = 0.05$$

$$n_1 = 20, \bar{X}_1 = 17.15, S_1 = 1.59, n_2 = 34, \bar{X}_2 = 16.92, S_2 = 1.42$$

$$S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \frac{(20 - 1)(1.59)^2 + (34 - 1)(1.42)^2}{20 + 34 - 2} = 2.20$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_c^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{17.15 - 16.92}{\sqrt{2.20 \times \left( \frac{1}{20} + \frac{1}{34} \right)}} = 0.550$$

$$\nu = n_1 + n_2 - 2 = 20 + 34 - 2 = 52$$

查t 临界值表:  $t_{0.5/2,50}=0.679$

$t < t_{0.5/2,50}$ , 得  $P > 0.5$

按  $\alpha=0.05$  水准不拒绝  $H_0$ , 还不能认为该市13—16岁居民腭弓深度

$t_{0.4,50}=0.849$ ,  $t_{0.4,60}=0.848$   
采用内插法得:  $t_{0.4,52}$  的值

$$\frac{x - 0.849}{0.848 - 0.849} = \frac{52 - 50}{60 - 50}$$

$$t_{0.4,52} = 0.8478$$



# 两样本方差齐性检验

$$F = \frac{S_1^2 (\text{较大})}{S_2^2 (\text{较小})} \quad \begin{array}{l} \nu_1 = n_1 - 1 \\ \nu_2 = n_2 - 1 \end{array}$$

第一个样本方差既可能大于也可能小于第二个样本方差，是双侧检验。公式规定以较大方差作分子，F必然大于1，附表7只给出不对称F分布右侧界值，实对应双尾概率P。

例3-12 对例3-8，用F检验判断对照组和试验组病人8周时糖化血红蛋白HbA<sub>1c</sub>(%)下降值总体方差是否不等。

(1) 建立检验假设，确定检验水准。

$$H_0:$$

$$H_1: \sigma_1^2 = \sigma_2^2$$

$$\sigma_1^2 \neq \sigma_2^2$$

$$\alpha=0.10$$

## (2) 计算检验统计量

$$F = \frac{1.36^2}{0.70^2} = 3.775$$

$$\nu_1 = 20 - 1 = 19$$

$$\nu_2 = 20 - 1 = 19$$

## (3) 确定P值，作出推断结论

$$3.775 > 2.15 = F_{0.10}(20, 19)$$

$$P < 0.10$$

按 $\alpha=0.10$ 水准，拒绝 $H_0$ ，接受 $H_1$ ，有统计学意义。可认为对照组和试验组病人HbA<sub>1c</sub>(%)下降值总体方差不等。

例6-5 为探讨硫酸氧钒对糖尿病性白内障的防治作用，研究人员将已诱导糖尿病模型的20只大鼠随机分为两组。一组用硫酸氧钒治疗（DV组），另一组作对照观察（D组），12周后测大鼠血糖含量(mmol/L)。结果为DV组12只，样本均数为6.5mmol/L，标准差为1.34 mmol/L；D组8只，样本均数为13.7mmol/L，标准差为4.21 mmol/L。试问两组动物血糖含量的总体均数是否相同？

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad H_1 : \sigma_1^2 \neq \sigma_2^2, \quad \alpha = 0.05$$

$$F = \frac{S_1^2}{S_2^2} = \frac{4.21^2}{1.34^2} = 9.87, \quad \nu_1 = 8 - 1 = 7, \quad \nu_2 = 12 - 1 = 11$$

查F 临界值表3.2：  $F_{0.05,(7,11)}=3.76, F > F_{0.05,(7,11)}$ ，得  $P < 0.05$

按  $\alpha=0.05$  水准拒绝  $H_0$ ，接受  $H_1$ ，故可认为两个总体方差不相等。

方差齐性检验的另一个方面是比较两组观测数据的测量精度。（两个总体的变异程度是否相同）

例6-7 将同一瓶液样分成20份。将此20份样品随机分成两组，每组10份。用不同的方法分别检测液样中某物质的含量( $\text{mmol/L}$ )。结果两种方法测得样本均数相同，样本标准差分别为1.02与0.56。试问两法检测精度是否相同？

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad H_1 : \sigma_1^2 \neq \sigma_2^2, \quad \alpha = 0.05$$

$$F = \frac{S_1^2}{S_2^2} = \frac{1.02^2}{0.56^2} = 3.3176, \quad \nu_1 = 10 - 1 = 9, \quad \nu_2 = 10 - 1 = 9$$

查F 临界值表3.2:  $F_{0.05,(9,9)}=4.03, F < F_{0.05,(9,9)}$ , 得 $P>0.05$

按 $\alpha=0.05$ 水准不拒绝 $H_0$ , 故还不能认为两法检测结果精度不同。

## 总体方差不相等的 $t$ 检验

⑩数据变换后进行  $t$  检验

⑩秩转换的非参数检验

⑩近似  $t$  检验 —  $t'$  检验

(separate variance estimation  
t-test)



$t'$  检验

$$t' = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$\nu_1 = n_1 - 1$$

$$\nu_2 = n_2 - 1$$

Cochran & Cox法(1950)

对临界值校正,

Satterthwaite法(1946)

对自由度校正。

Aspin-Welch法(1947) p101

对自由度校正。

两样本所属总体方差不等且两总体均为正态分布

近似  $t$  检验-----  $t'$  检验

方差不齐时，两小样本均数的比较，可选择以下方法：

- 1) 采用适当的变量变换，使达到方差齐的要
- 2) 采用秩和检验。
- 3) 采用近似法检验  $t'$ （它包括对临界值和自由度校正两种）。

当所两小样本均数的比较时，其总体分布不呈正态分布时，可选择的方法有两种即变量变换和秩和检验。

#### 四、两样本所属总体方差不等且两总体均为正态分布

$$H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2 (\text{单侧} \mu_1 > \mu_2 \text{或} \mu_1 < \mu_2)$$

当 $H_0$ 成立时，检验统计量(Satterthwaite近似法)

$$t' = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}, \quad \nu = \frac{\left( S_{\bar{X}_1}^2 + S_{\bar{X}_2}^2 \right)^2}{\frac{S_{\bar{X}_1}^4}{n_1 - 1} + \frac{S_{\bar{X}_2}^4}{n_2 - 1}}$$

例6-5 为探讨硫脲对糖尿病大鼠的防治作用,研究人员将已诱发的糖尿病大鼠随机分为两组。一组用硫脲治疗(实验组),另一组作对照观察(D组),测定大鼠血糖含量(mmol/L)。此资料总体方差不等。

结果为DV组12只,样本均数为6.5mmol/L,标准差为1.34 mmol/L; D组8只,样本均数为13.7mmol/L,标准差为4.21 mmol/L。试问两组动物血糖含量的总体均数是否相同?

$$\frac{S_1^2}{S_2^2} = \frac{4.21^2}{1.34^2} = 9.87$$

$$H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2 \quad \alpha = 0.05$$

$$n_1 = 8, \bar{X}_1 = 13.7, S_1 = 4.21, n_2 = 12, \bar{X}_2 = 6.5, S_2 = 1.34$$

$$t' = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{13.7 - 6.5}{\sqrt{\frac{4.21^2}{8} + \frac{1.34^2}{12}}} = 4.6817$$

$$\nu = \frac{\left(\frac{4.21^2}{8} + \frac{1.34^2}{12}\right)^2}{\frac{\left(\frac{4.21^2}{8}\right)^2}{8-1} + \frac{\left(\frac{1.34^2}{12}\right)^2}{12-1}} = 7.9542 \approx 8$$

查t 临界值表:  $t_{0.05/2,8}=2.306$

$t > t_{0.05/2,8}$ , 得 $P<0.05$

按 $\alpha=0.05$ 水准拒绝 $H_0$ , 接受 $H_1$ , 故可认为经硫酸  
氧钒治疗的大鼠与未治疗大鼠的血糖含量不同。

## 第四节 二项分布与poisson分布资料的z检验

### 一、二项分布资料的z检验

#### (一) 一组样本资料的z检验

如果二项分布的 $\pi$ 或 $1-\pi$ 不太小，则当 $n$ 足够大时，即阳性数与阴性数都大于等于5时，近似地有

$$X \sim N(n\pi, n\pi(1-\pi))$$

$$P \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

$$P = \frac{X}{n}$$

检验假设为：

$$H_0 : \pi = \pi_0 \quad H_1 : \pi \neq \pi_0$$

当 $H_0$ 成立时，检验统计量为：

$$Z = \frac{X - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} \sim N(0,1)$$

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \sim N(0,1)$$



当n不太大时，需作连续性校正：

$$Z = \frac{|X - n\pi_0| - 0.5}{\sqrt{n\pi_0(1 - \pi_0)}} \sim N(0,1)$$

$$Z = \frac{|p - \pi_0| - \frac{0.5}{n}}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}} \sim N(0,1)$$

例6-8 某医院称治疗声带白斑的有效率为80%，今统计前来求医的此类患者60例，其中45例治疗有效。试问该医院宣称的疗效是否客观？

$$H_0 : \pi = \pi_0 = 0.8 \quad H_1 : \pi < 0.8(\text{单侧}), \quad \alpha = 0.05$$

$$Z = \frac{|0.75 - 0.8| - \frac{0.5}{60}}{\sqrt{\frac{0.8 \times 0.2}{60}}} = 0.8069$$

按 $v=\infty$ 查t 临界值表：（单侧） $Z_{0.10, \infty}=1.2816$

$Z < Z_{0.10}$ ，得 $P>0.10$

按 $\alpha=0.05$ 水准不拒绝 $H_0$ ，故可认为该医院宣称的有效率尚属客观。

## (二) 两组独立样本资料的z检验

它的应用条件为当所比较的两组阳性数与阴性数都大于等于5时

检验假设为： $H_0: \pi_1 = \pi_2$      $H_1: \pi_1 \neq \pi_2$

当 $H_0$ 成立时，检验统计量为：

$$Z = \frac{|p_1 - p_2|}{\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \quad Z = \frac{|p_1 - p_2| - 0.5\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\sqrt{p_c(1-p_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$p_c = \frac{X_1 + X_2}{n_1 + n_2}$$

例6-9 用硝苯吡啶治疗高血压急症患者75例，有效者57例，用硝苯吡啶+卡托普利治疗同类患者69例，66例有效。试问两疗法的有效率是否相同？

$$p_1 = \frac{57}{75} = 0.76, p_2 = \frac{66}{69} = 0.95652, p_c = \frac{57 + 66}{75 + 69} = 0.85417$$

$$H_0 : \pi_1 = \pi_2 \quad H_1 : \pi_1 \neq \pi_2, \alpha = 0.05$$

$$\begin{aligned} Z &= \frac{|\mathbf{p}_1 - \mathbf{p}_2|}{\sqrt{\mathbf{p}_c(1 - \mathbf{p}_c)\left(\frac{1}{\mathbf{n}_1} + \frac{1}{\mathbf{n}_2}\right)}} \\ &= \frac{|0.76 - 0.95652|}{\sqrt{0.85417(1 - 0.85417)\left(\frac{1}{75} + \frac{1}{69}\right)}} = 3.33799 \end{aligned}$$

按 $\nu=\infty$ 查t 临界值表： $Z_{0.001/2, \infty}=3.2905$

$Z > Z_{0.001/2}$ ，得 $P<0.001$

按 $\alpha=0.05$ 水准拒绝 $H_0$ ，接受 $H_1$ ，故可认为两种疗法有效率不同。

## 二、Poisson分布资料的z检验

当总体均数 $\lambda \geq 20$ 时，Poisson分布近似正态分布。

### (一) 一组样本资料的z检验

$$H_0 : \lambda = \lambda_0 \quad H_1 : \lambda \neq \lambda_0$$

当 $H_0$ 成立时，检验统计量为：

$$Z = \frac{X - \lambda_0}{\sqrt{\lambda_0}} \sim N(0,1)$$



## (一) 一组样本资料的z检验

例6-10 某地十年前计划到2000年把孕产妇死亡率将到25/10万以下。2000年监测资料显示，该地区平均而言，每10万例活产儿孕产妇死亡31人。问该地区降低孕产妇死亡的目标是否达到？

$$\lambda_0 = 25, X = 31$$

$$H_0 : \lambda = \lambda_0 = 25 \quad H_1 : \lambda > 25 \quad \alpha = 0.05$$

$$Z = \frac{X - \lambda_0}{\sqrt{\lambda_0}} = \frac{31 - 25}{\sqrt{25}} = 1.2$$

按 $v=\infty$ 查t 临界值表：(单侧)  $Z_{0.10, \infty} = 1.2816$

$$Z < Z_{0.10}, \text{ 得 } P > 0.10$$

按 $\alpha=0.05$ 水准不拒绝 $H_0$ ，故可认为该地区达到了预定目标。

例6-11 有研究表明，一般人群精神发良不全的发生率不3‰，今调查了有亲缘血统婚配关系的后代25000人，发现123人精神不全，问有亲缘血统婚配关系的后代其精神发良不全者的发病率是否人高于一般人群？

$$\lambda_0 = 25000 * 0.003 = 75, X = 123$$

$$H_0 : \lambda = \lambda_0 = 75 \quad H_1 : \lambda > 75 \quad \alpha = 0.05$$

$$Z = \frac{75 - \lambda_0}{\sqrt{\lambda_0}} = \frac{123 - 75}{\sqrt{75}} = 5.543$$

按 $\nu=\infty$ 查t 临界值表：(单侧)  $Z_{0.10, \infty} = 3.2905$

$$Z > Z_{0.0005}, \text{ 得 } P < 0.0005$$

按 $\alpha=0.05$ 水准拒绝 $H_0$ ，故可认为有亲缘血统婚配关系的后代其精神发育不全的发生率高于一般人群。

## (二) 两组独立样本资料的z检验

当两总体均数都大于20时，可应用正态近似原理。

$$H_0 : \lambda_1 = \lambda_2 \quad H_1 : \lambda_1 \neq \lambda_2$$

当 $H_0$ 成立时，检验统计量为：

●当两样本观测单位数相等时：

$$Z = \frac{X_1 - X_2}{\sqrt{X_1 + X_2}} \sim N(0,1)$$

●当两样本观测单位数不等时：

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

例6-12 甲、乙两检验师分别观察15名正常人末梢血嗜碱性白细胞数量。每张血片均观察200个视野。结果甲计数到嗜碱粒细胞26个，乙计数到29个。试问两位检验师检查结果是否一致？

$$H_0 : \lambda_1 = \lambda_2 \quad H_1 : \lambda_1 \neq \lambda_2, \quad \alpha = 0.05$$

$$Z = \frac{X_1 - X_2}{\sqrt{X_1 + X_2}} = \frac{26 - 29}{\sqrt{26 + 29}} = -0.40452$$

按 $v=\infty$ 查t 临界值表： $Z_{0.5/2, \infty}=0.6745$

$Z < Z_{0.5/2}$ ，得 $P>0.5$

按 $\alpha=0.05$ 水准不拒绝 $H_0$ ，故尚不能认为两检验师检查结果有差异。

例6-12 某车间改革生产工艺前，测得三次粉尘浓度，每升空气中 分别有38、29、36颗粉尘；改革工艺后，测取两次，分别为25、18颗粉尘。问工艺改革前后粉尘数有无差别？

$$\bar{X}_1 = \frac{38 + 29 + 36}{3} = 34.33, \quad n_1 = 3$$

$$\bar{X}_2 = \frac{25 + 18}{2} = 21.50, \quad n_2 = 2$$



$$H_0 : \lambda_1 = \lambda_2 \quad H_1 : \lambda_1 \neq \lambda_2, \quad \alpha = 0.05$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\bar{X}_1}{n_1} + \frac{\bar{X}_2}{n_2}}} = \frac{34.33 - 21.50}{\sqrt{\frac{34.66}{3} + \frac{21.50}{2}}} = 2.723$$

按 $v=\infty$ 查t 临界值表： $Z_{0.05/2, \infty}=1.96$

$Z > Z_{0.05/2}$ ，得 $P < 0.05$

按 $\alpha=0.05$ 水准拒绝 $H_0$ ，接受 $H_1$ ，故可认为工艺改革前后粉尘浓度不同，改革工艺后粉尘浓度较低。

## 应用泊松分布应注意的问题

1、泊松分布的 $X$ 虽然为样本计数，由于观察的单位（时间、面积、容积等）是固定的，可看成 $n$ 为1，故有均数的含义，按泊松分布的特性， $\mu = \sigma^2$ ，而标准误，在 $n$ 为1时，标准差与标准误相等。当观察单位不同时，标准误  $= \frac{\sigma}{\sqrt{n}}$ 。

2、要利用正态近似法，要满足条件  $\mu > 20$ 。如果想满足此条件，可利用泊松分布的可加性原理，把若干个小单位合并成一个大单位。

3、两样本均数的比较时，要注意观察单位（时间、面积、容积、人口基数等）是否相等，若不相等，须化成相同的观察单位后再作比较。

### 第三节 假设检验的功效

#### 一、假设检验的两类错误

● **第I类错误**：如果实际情况与 $H_0$ 一致，仅仅由于抽样的原因，得到了较大的t值，使得统计量的观察值落到拒绝域，拒绝了实际上成立的 $H_0$ ，这类“弃真”的错误称为**第I类错误**。其概率大小用 $\alpha$ 表示， $\alpha$ 可以取单尾亦可以取双尾。

● **第II类错误**：如果实际情况与 $H_0$ 不一致，仅仅由于抽样的原因，得到了较小的t值，使得统计量的观察值落到接受域，不能拒绝原本错误的 $H_0$ ，这类“取伪”的错误称为**第II类错误**。其概率大小用 $\beta$ 表示， $\beta$ 只取单尾。

# 假设检验的两类错误

实际情况	检验结果	
	拒绝 $H_0$	不拒绝 $H_0$
$H_0$ 为真	第Ⅰ类错误( $\alpha$ ) 假阳性 (误诊)	结论正确( $1 - \alpha$ ) 置信度
$H_0$ 不真	结论正确( $1 - \beta$ ) 检验功效	第Ⅱ类错误( $\beta$ ) 假阴性 (漏诊)

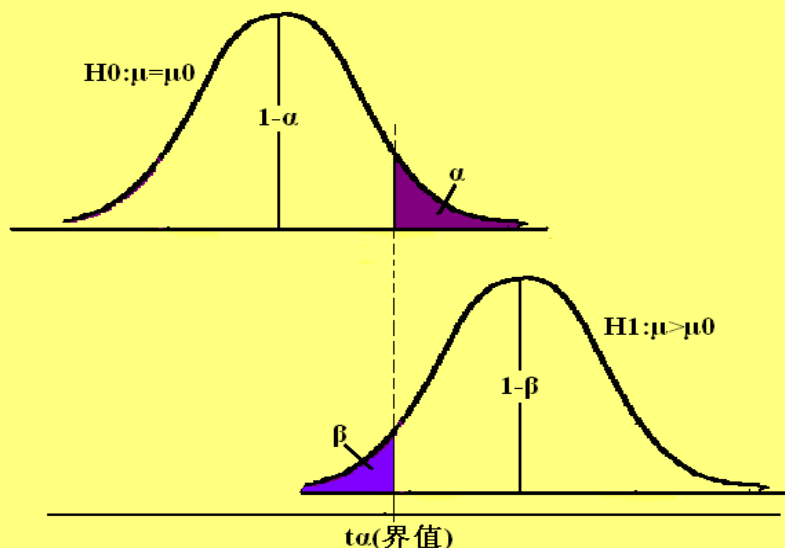


图6-3 假设检验的两类错误示意图

当样本含量 $n$ 一定时， $\alpha$ 越小， $\beta$ 越大；若想同时减少 $\alpha$ 和 $\beta$ ，只有增大样本含量。

$\alpha$ 是预先规定允许犯I型错误概率的最大值，可取单尾亦可取双尾。

II型错误的概率大小用 $\beta$ 表示， $\beta$ 只取单尾，因为 $H_0$ 不成立时检验统计量的精确分布往往难以确定，所以 $\beta$ 值的大小一般未知，须在知道两总体差值 $\delta$ （如 $\mu_1 - \mu_2$ 等）、 $\alpha$ 及 $n$ 时，才能算出。

$\beta$ 的意义是：如果 $H_0$ 并不成立，即所研究的总体与 $H_0$ 有实质差异（例如 $\mu_1 \neq \mu_2$ ），按照同样的方法在总体中重复抽样，那么在每100次检验结论中平均可以有 $100\beta$ 次接受 $H_0$ （犯第II类错误）。

$1-\beta$  称检验效能 (power of a test)，过去称把握度。为当两总体确有差异，按检验水准  $\alpha$  所能发现该差异的能力。 $1-\beta$  只取单尾。

$\alpha$  愈小， $\beta$  愈大； $\alpha$  愈大， $\beta$  愈小。若要同时减小  $\alpha$  和  $\beta$ ，唯一方法是增加样本含量  $n$ 。

若重点减少 $\alpha$  (如一般假设检验), 一般取 $\alpha=0.05$ ;

若重点减少 $\beta$  (如方差齐性检验, 正态性检验等), 一般取 $\alpha=0.10$ 或 $0.20$ 甚至更高。

拒绝 $H_0$ , 只可能犯I型错误, 不可能犯II型错误; 不拒绝 $H_0$ , 只可能犯II型错误, 不可能犯I型错误。



## 二、假设检验的功效

●  $1-\beta$ 称为假设检验的功效 当所研究的总体与 $H_0$ 确有差别时，按检验水准 $\alpha$ 能够发现它(拒绝 $H_0$ )的概率。

一般情况下对同一检验水准 $\alpha$ ，功效大的检验方法更可取。

在医学科研设计中，检验功效( $1-\beta$ )不宜低于0.75，否则检验结果很可能反映不出总体的真实差异，出现非真实的阴性结果。

当假设检验的结果根据 $P > 0.05$ 作出无统计学意义的“阴性”结论时，研究者则面临着犯第II类错误的可能性，应当考虑是否总体间的差异确实存在，有可能由于检验效能不足而未能把总体中确有的差异反映出来。近年来，许多国际会议对于假设检验中 $P > 0.05$ 作出的“阴性”结论，应附有第II类错误概率为据，因而有必要介绍其计算方法。

# 1. 一组样本资料t 检验的功效

$$Z_{\beta} = \frac{\delta\sqrt{n}}{\sigma} - Z_{\alpha}$$

例6-1 已知北方农村儿童前囟门闭合月龄为14.1月。某研究人员从东北某县抽取36名儿童，得囟门闭合月龄均值为14.3月，标准差为5.08月。问该县儿童前囟门闭合月龄的均数是否大于一般儿童？

例6-13 计算例6-1 检验的功效1- $\beta$ 。假定根据现有知识可以取 $\sigma=5$ 月， $\delta=0.5$ 月，单侧 $Z_{\alpha}=1.645$

由标准正态分布表查-1.045所对的上侧尾部面积，得到 $\beta=0.8519$ ， $1-\beta=0.1481$ 。说明该检验功效太即发现 $\delta=0.5$ 个月的差别的机会14.81%。

$$Z_{\beta} = \frac{\delta\sqrt{n}}{\sigma} - Z_{\alpha} = \frac{0.5 \times \sqrt{36}}{5} - 1.645 = -1.045$$

## 2. 两组独立样本资料t 检验的功效

由标准正态分布表查0.4057所对应的上侧尾部面积，得到 $\beta=0.3430$ ，于是 $1-\beta=0.6570$ 。说明该检验功效偏小，即发现 $\delta=1\text{mm}$ 的差别的机会仅有65.70%。

例6-14

可以取 $\sigma=1.5\text{mm}$

$\delta=1\text{mm}$ ,  $Z_\alpha=1.96$ ,  $n_1=20$ ,  $n_2=34$

$$Z_\beta = \frac{\delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} - Z_\alpha = \frac{1}{1.5 \times \sqrt{\frac{1}{20} + \frac{1}{34}}} - 1.96 = 0.4057$$

见有知识

例 用某药降低心肌梗死患者血压的随机双盲对照试验结果如下，假设检验结果 $t=1.54, P>0.1$ ,认为该药降低血压无效，问此结论是否可靠。

两组心肌梗死患者治疗后的收缩压

分组	n	均数	标准差
试验组	15	14.4	1.6
对照组	15	15.3	1.6

根据专业知识认为两均数相差若小于0.67kPa，应认为该差值无临床意义，故令 $\delta=0.67\text{kPa}$ ， $\alpha=0.05$ ，

$$Z_{0.05/2}=1.96$$

$$Z_{\beta} = \frac{\delta}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} - Z_{\alpha} = \frac{0.67}{1.6 \sqrt{\frac{1}{15} + \frac{1}{15}}} - 1.96 = -0.81$$

由标准正态分布表查-0.81所对应的上侧尾部面积，得到 $\beta=0.791$ ，于是 $1-\beta=0.2090$ 。说明该检验功效偏小，此结论并不可靠，还应增大样本含量进一步试验。

### 3.二项分布两组独立样本资料Z 检验的功效

$$Z_{\beta} = \frac{|\pi_1 - \pi_2| - Z_{\alpha} \sqrt{\pi(1-\pi) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}}$$

例6-15 试计算例6-9 中检验的功效。

假定根据现有知识可以取 $\delta=0.20$ ,  $\pi_1=0.76$ ,  $\pi_2=0.96$

$$\pi = (57+66)/(75+69)=0.854$$



$$Z_{\beta} = \frac{|\pi_1 - \pi_2|}{\sqrt{\pi}}$$

由标准正态分布表查1.547所对应的上侧尾部面积，得到 $\beta=0.0614$ ，于是 $1-\beta=0.9386$ 。说明该检验功效较高，即发现 $\delta=0.20$ 的差别的机会达到了93.86%。

$$= \frac{|\mathbf{0.76} - \mathbf{0.96}| - \mathbf{1.96} \sqrt{\mathbf{0.854(0.146)} \left( \frac{\mathbf{1}}{\mathbf{75}} + \frac{\mathbf{1}}{\mathbf{69}} \right)}}{\sqrt{\frac{\mathbf{0.76(0.24)}}{\mathbf{75}} + \frac{\mathbf{0.96(0.04)}}{\mathbf{69}}}} = \mathbf{1.547}$$

### 三、应用假设检验需要注意的问题

#### 1. 要有严密的研究设计

在抽样研究中，研究设计、搜集数据和统计分析是一个整体。每一种假设检验方法都是与相应的研究设计相联系的。组间应均衡，具有可比性，也就是除对比的主要因素(如临床试验用新药和对照药)外，其它可能影响结果的因素(如年龄、性别、病程、病情轻重等)在对比组间应相同或相近。

## 2. 应用检验方法必需符合其适用条件

一般 $t$  检验要求样本取自正态总体，而且各总体方差齐同。

配对设计计量资料：配对  $t$  检验。

完全随机设计两样本计量资料：

小样本(任一 $n_i \leq 60$ )且方差齐，两样本  $t$  检验；若方差不齐，近似  $t'$  检验。

大样本(所有 $n_i > 60$ )， $u$  检验。

### 3. 适当选择检验水准 $\alpha$

当样本量一定时， $\alpha$ 越小， $\beta$ 越大；反之亦然。若想同时减少 $\alpha$ 和 $\beta$ ，只有增大样本含量。

### 4. 结论不能绝对化

因统计结论具有概率性质，故“肯定”、“一定”、“必定”等词不要使用。在报告结论时，最好列出检验统计量的值，尽量写出具体的P值或P值的确切范围，而不简单写成 $P < 0.05$ ，以便读者与同类研究进行比较或进行循证医学时采用Meta分析。

## 4. 正确理解P值的意义

P值很小时“拒绝 $H_0$ ，接受 $H_1$ ”，但是不要把很小的P值误解为总体参数间差异很大。如果 $P < \alpha$ ，宜说差异“有统计学意义”，同时写出P的数值或相应的不等式。

## 5. 统计“显著性”与专业“显著性”

假设检验是为各专业服务的，统计结论必须和专业结论有机的相结合，才能得出恰如其分、符合客观实际的最终结论。

