

第 5 章 参数估计

- ❖ 5.1 参数估计的一般问题
- ❖ 5.2 一个总体参数的区间估计
- ❖ 5.3 两个总体参数的区间估计
- ❖ 5.4 样本容量的确定

参数估计与非参数估计

- ❖ 参数估计：运用样本对已知分布的未知参数进行估计。
- ❖ 参数检验：在总体分布形式已知（通常进行假定）的情况下，对总体分布的参数如均值、方差等进行推断的方法。
- ❖ 非参数估计：未知分布，运用样本对分布类型进行估计。
- ❖ 非参数检验：人们无法知道总体分布形态，或者无法对其作出简单假定，利用样本数据对总体分布形态等进行推断的方法。由于非参数检验方法在过程中不涉及有关总体分布的参数，因而得名为“非参数”检验。

5.1 参数估计的一般问题

- 一、估计量与估计值
- 二、点估计与区间估计
- 三、评价估计量的标准

估计量与估计值

(estimator & estimated value)

1 估计量：用于估计总体参数的随机变量

✧如样本均值，样本比率、样本方差等

✧例如：样本均值就是总体均值 μ 的一个估计量

参数用 θ 表示，估计量用 $\hat{\theta}$ 表示

估计值：估计参数时计算出来的统计量的具体值

✧如果样本均值 $\bar{x}=80$ ，则80就是 μ 的估计值

点估计

(point estimate)

1. 用样本的估计量直接作为总体参数的估计值

❧ 例如：用样本均值直接作为总体均值的估计

❧ 例如：用两个样本均值之差直接作为总体均值之差的估计

没有给出估计值接近总体参数程度的信息

点估计的方法有矩估计法、顺序统计量法、最大似然法、最小二乘法等

区间估计

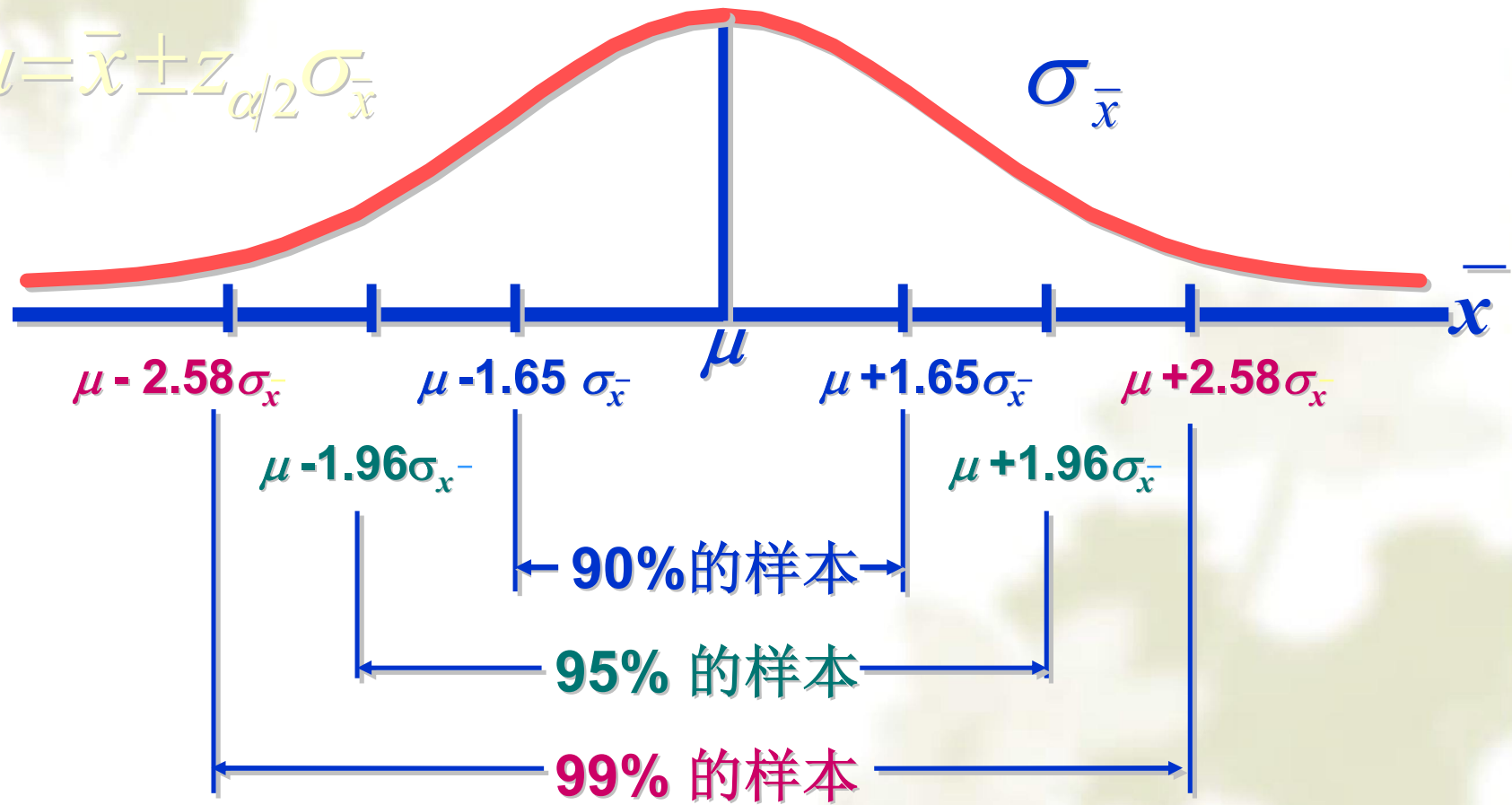
(interval estimate)

1 在点估计的基础上，给出总体参数估计的一个区间范围，该区间由样本统计量加减抽样误差而得到的。根据样本统计量的抽样分布能够对样本统计量与总体参数的接近程度给出一个概率度量。

✎ 比如，某班级平均分数在75~85之间，置信水平是95%

区间估计的图示

$$\mu = \bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}}$$



置信水平

将构造置信区间的步骤重复很多次，置信区间包含总体参数真值的次数所占的比率称为置信水平

表示为 $(1 - \alpha) \%$

❧ α 为是总体参数未在区间内的比率

常用的置信水平值有 99%，95%，90%

❧ 相应的 α 为 0.01，0.05，0.10

置信区间

(confidence interval)

由样本统计量所构造的总体参数的估计区间称为置信区间

统计学家在某种程度上确信这个区间会包含真正的总体参数，所以给它取名为置信区间

用一个具体的样本所构造的区间是一个特定的区间，我们无法知道这个样本所产生的区间是否包含总体参数的真值

- 我们只能是希望这个区间是大量包含总体参数真值的区间中的一个，但它也可能是少数几个不包含参数真值的区间中的一个

影响区间宽度的因素

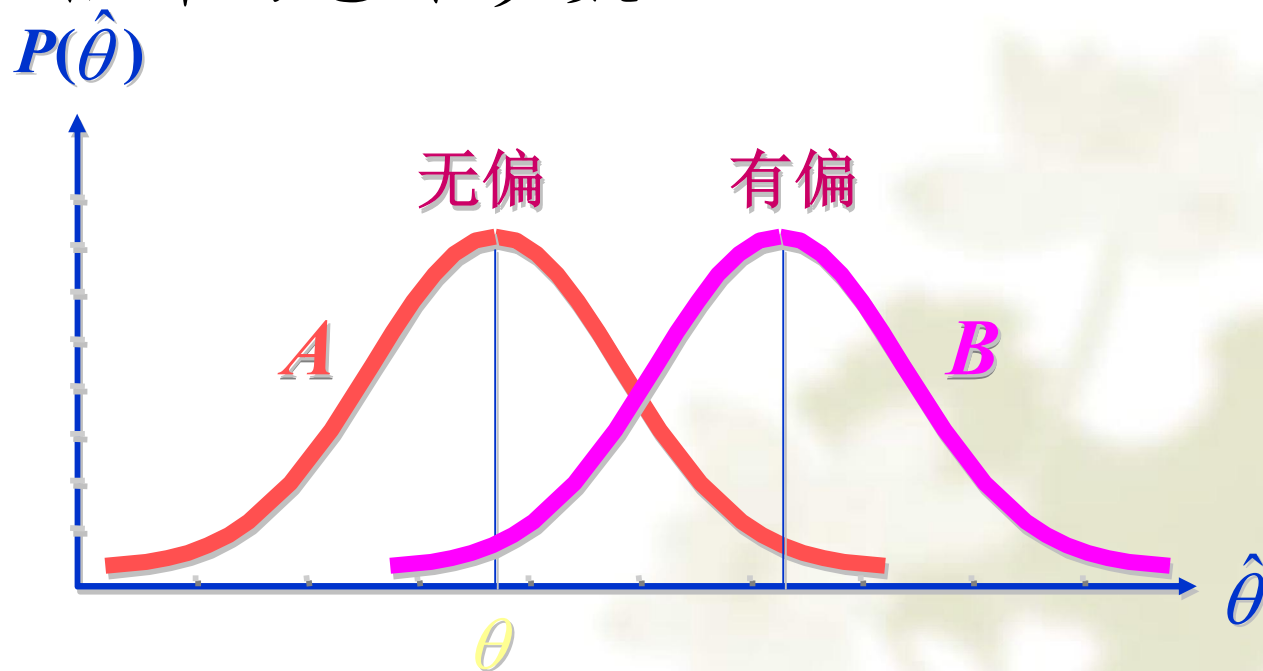
- ❖ 1. 总体数据的离散程度，用 σ 来测度
样本容量， $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- 3. 3. 置信水平 $(1 - \alpha)$ ，影响 z 的大小

评价估计量的标准

- ❖ 1、无偏性
- ❖ 2、有效性
- ❖ 3、一致性

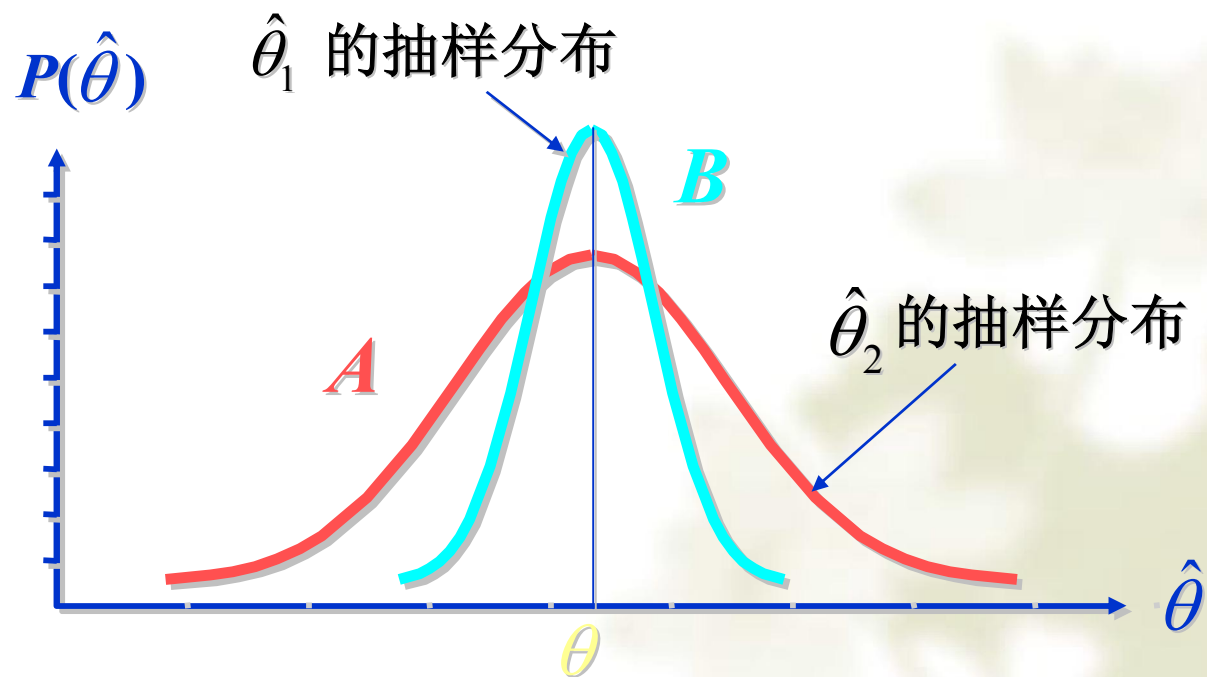
无偏性 (unbiasedness)

- ❖ 无偏性：估计量抽样分布的数学期望等于被估计的总体参数



有效性 (efficiency)

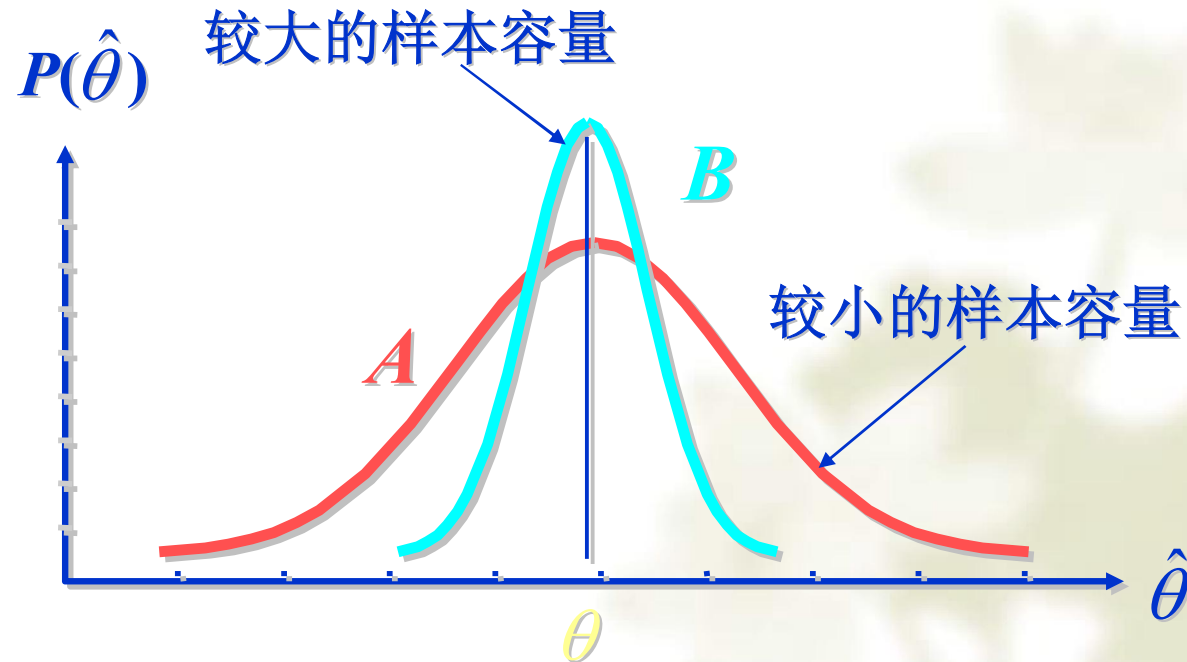
有效性: 对同一总体参数的两个无偏点估计量，有更小标准差的估计量更有效



一致性

(consistency)

- ❖ **一致性**：随着样本容量的增大，估计量的值越来越接近被估计的总体参数



5.2 一个总体参数的区间估计

- 一、总体均值的区间估计
- 二、总体比率的区间估计
- 三、总体方差的区间估计

一个总体参数的区间估计

总体参数	符号表示	样本统计量
均值	μ	\bar{x}
比率	π	p
方差	σ^2	s^2

总体均值的区间估计 (大样本)

❖ 1. 假定条件

☞ 总体服从正态分布, 且方差(σ^2) 未知

☞ 如果不是正态分布, 可由正态分布来近似 ($n \geq 30$)

使用正态分布统计量 z

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

总体均值 μ 在 $1-\alpha$ 置信水平下的置信区间为

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{或} \quad \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} \quad (\sigma \text{ 未知})$$

总体均值的区间估计

(例题分析)

【例】一家食品生产企业以生产袋装食品为主，为对产量质量进行监测，企业质检部门经常要进行抽检，以分析每袋重量是否符合要求。现从某天生产的一批食品中随机抽取了25袋，测得每袋重量如下表所示。已知产品重量的分布服从正态分布，且总体标准差为10g。试估计该批产品平均重量的置信区间，置信水平为95%

25袋食品的重量

112.5	101.0	103.0	102.0	100.5
102.6	107.5	95.0	108.8	115.6
100.0	123.5	102.0	101.6	102.2
116.6	95.4	97.8	108.6	105.0
136.8	102.8	101.5	98.4	93.3

总体均值的区间估计

(例题分析)

解：已知 $X \sim N(\mu, 10^2)$, $n=25$, $1-\alpha = 95\%$, $z_{\alpha/2}=1.96$ 。根据样本数据计算得： $\bar{x} = 105.36$

总体均值 μ 在 $1-\alpha$ 置信水平下的置信区间为

$$\begin{aligned}\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} &= 105.36 \pm 1.96 \times \frac{10}{\sqrt{25}} \\ &= 105.36 \pm 3.92 \\ &= (101.44, 109.28)\end{aligned}$$

该食品平均重量的置信区间为101.44g~109.28g

总体均值的区间估计

(例题分析)

【例】一家保险公司收集到由36投保个人组成的随机样本，得到每个投保人的年龄(周岁)数据如下表。试建立投保人年龄90%的置信区间

36个投保人年龄的数据

23	35	39	27	36	44
36	42	46	43	31	33
42	53	45	54	47	24
34	28	39	36	44	40
39	49	38	34	48	50
34	39	45	48	45	32

总体均值的区间估计

(例题分析)

解：已知 $n=36$ ， $1-\alpha = 90\%$ ， $z_{\alpha/2}=1.645$ 。根据样本数据计算得： $\bar{x} = 39.5$ $s = 7.77$

总体均值 μ 在 $1-\alpha$ 置信水平下的置信区间为

$$\begin{aligned}\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} &= 39.5 \pm 1.645 \times \frac{7.77}{\sqrt{36}} \\ &= 39.5 \pm 2.13 \\ &= (37.37, 41.63)\end{aligned}$$

投保人平均年龄的置信区间为37.37岁~41.63岁

总体均值的区间估计 (小样本)

❖ 1. 假定条件

- ☞ 总体服从正态分布, 且方差(σ^2) 未知
- ☞ 小样本 ($n < 30$)

使用 t 分布统计量

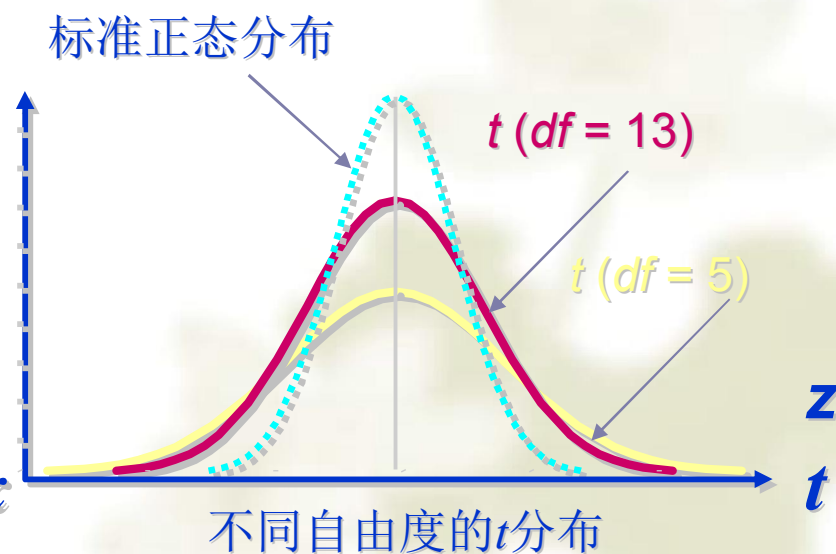
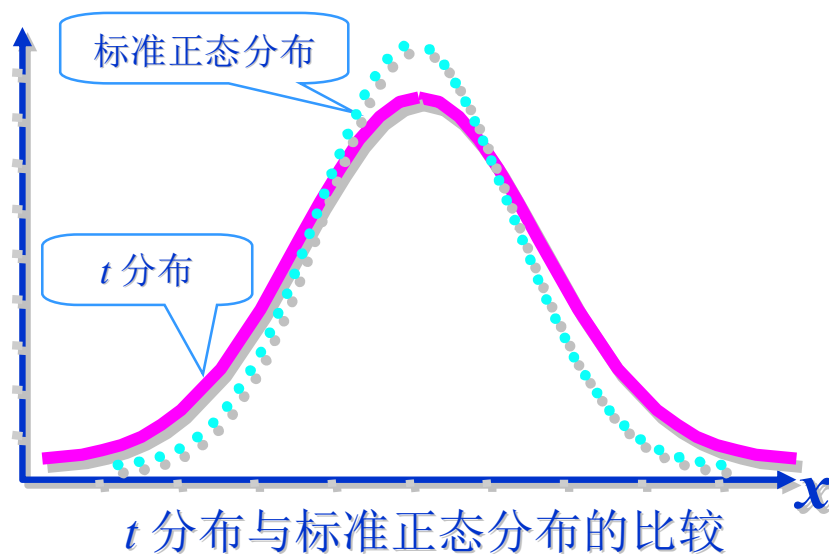
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

总体均值 μ 在 $1-\alpha$ 置信水平下的置信区间为

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

t 分布

- t 分布是类似正态分布的一种对称分布，它通常要比正态分布平坦和分散。一个特定的分布依赖于称之为自由度的参数。随着自由度的增大，分布也逐渐趋于正态分布



总体均值的区间估计

(例题分析)

【例】已知某种灯泡的寿命服从正态分布，现从一批灯泡中随机抽取16只，测得其使用寿命(小时)如下。建立该批灯泡平均使用寿命95%的置信区间

16灯泡使用寿命的数据

1510	1520	1480	1500
1450	1480	1510	1520
1480	1490	1530	1510
1460	1460	1470	1470

总体均值的区间估计

(例题分析)

解：已知 $X \sim N(\mu, \sigma^2)$, $n=16$, $1-\alpha = 95\%$, $t_{\alpha/2}=2.131$

根据样本数据计算得： $\bar{x} = 1490$, $s = 24.77$

总体均值 μ 在 $1-\alpha$ 置信水平下的置信区间为

$$\begin{aligned}\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} &= 1490 \pm 2.131 \times \frac{24.77}{\sqrt{16}} \\ &= 1490 \pm 13.2 \\ &= (1476.8, 1503.2)\end{aligned}$$

该种灯泡平均使用寿命的置信区间为1476.8小时~1503.2小时

总体比率的区间估计

❖ 1. 假定条件

- ∞ 总体服从二项分布
- ∞ 可以由正态分布来近似

使用正态分布统计量 z

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

总体比率 π 在 $1-\alpha$ 置信水平下的置信区间为

$$p \pm z_{\alpha/2} \sqrt{\frac{\pi(1-\pi)}{n}} \quad \text{或} \quad p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \quad (\pi \text{ 未知时})$$

总体比率的区间估计

(例题分析)

【例】某城市想要估计下岗职工中女性所占的比率，随机地抽取了100名下岗职工，其中65人为女性职工。试以95%的置信水平估计该城市下岗职工中女性比率的置信区间

解：已知 $n=100$, $p=65%$, $1-\alpha = 95%$,
 $z_{\alpha/2}=1.96$

$$p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

$$= 65\% \pm 1.96 \times \sqrt{\frac{65\%(1-65\%)}{100}}$$

$$= 65\% \pm 9.35\%$$

$$= (55.65\%, 74.35\%)$$

该城市下岗职工中女性比率的置信区间为55.65%~74.35%

总体方差的区间估计

- ❖ 1. 估计一个总体的方差或标准差
- ❖ 2. 假设总体服从正态分布

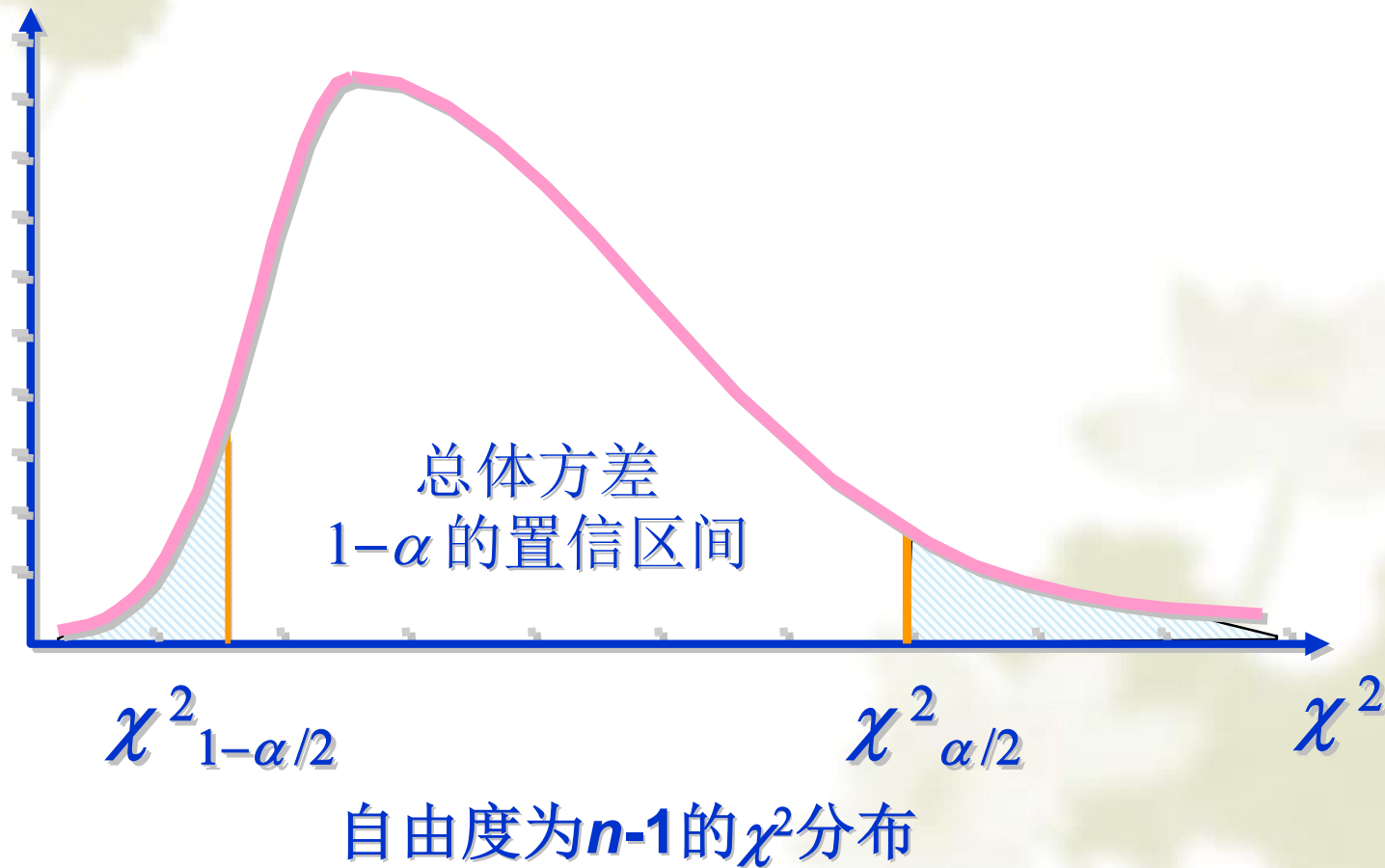
总体方差 σ^2 的点估计量为 s^2 , 且

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

总体方差在 $1-\alpha$ 置信水平下的置信区间为

$$\frac{(n-1)s^2}{\chi_{\alpha/2}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{1-\alpha/2}^2(n-1)}$$

总体方差的区间估计



总体方差的区间估计

(例题分析)

【例】一家食品生产企业以生产袋装食品为主，现从某天生产的一批食品中随机抽取了25袋，测得每袋重量如下表所示。已知产品重量的分布服从正态分布。以95%的置信水平建立该种食品重量方差的置信区间

25袋食品的重量				
112.5	101.0	103.0	102.0	100.5
102.6	107.5	95.0	108.8	115.6
100.0	123.5	102.0	101.6	102.2
116.6	95.4	97.8	108.6	105.0
136.8	102.8	101.5	98.4	93.3

总体方差的区间估计

(例题分析)

解: 已知 $n=25$, $1-\alpha=95%$, 根据样本数据计算得

$$s^2 = 93.21$$

$$\chi_{\alpha/2}^2(n-1) = \chi_{0.025}^2(24) = 39.364 \quad \chi_{1-\alpha/2}^2(n-1) = \chi_{0.975}^2(24) = 12.401$$

σ^2 置信度为95%的置信区间为

$$\frac{(25-1) \times 93.21}{39.364} \leq \sigma^2 \leq \frac{(25-1) \times 93.21}{12.401}$$

$$= 56.83 \leq \sigma^2 \leq 180.39$$

该企业生产的食品总体重量标准差的的置信区间为 $7.54\text{g} \sim 13.43\text{g}$

5.3 两个总体参数的区间估计

- ❖ 一、两个总体均值之差的区间估计
- ❖ 二、两个总体比率之差的区间估计
- ❖ 三、两个总体方差比的区间估计

两个总体参数的区间估计

总体参数	符号表示	样本统计量
均值之差	$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$
比率之差	$\pi_1 - \pi_2$	$p_1 - p_2$
方差比	σ_1^2 / σ_2^2	s_1^2 / s_2^2

两个总体均值之差的估计

(大样本)

❖ 1. 假定条件

- ☞ 两个总体都服从正态分布， σ_1^2 、 σ_2^2 已知
- ☞ 若不是正态分布，可以用正态分布来近似
($n_1 \geq 30$ 和 $n_2 \geq 30$)
- ☞ 两个样本是独立的随机样本

2. 使用正态分布统计量 z

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

两个总体均值之差的估计

(大样本)

- ❖ 1. σ_1^2 , σ_2^2 已知时, 两个总体均值之差 $\mu_1 - \mu_2$ 在 $1 - \alpha$ 置信水平下的置信区间为

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

2. σ_1^2 , σ_2^2 未知时, 两个总体均值之差 $\mu_1 - \mu_2$ 在 $1 - \alpha$ 置信水平下的置信区间为

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

两个总体均值之差的估计

(例题分析)

【例】某地区教育委员会想估计两所中学的学生高考时的英语平均分数之差，为此在两所中学独立抽取两个随机样本，有关数据如右表。建立两所中学高考英语平均分数之差95%的置信区间

两个样本的有关数据	
中学1	中学2
$n_1=46$	$n_2=33$
$\bar{x}_1=86$	$\bar{x}_2=78$
$S_1=5.8$	$S_2=7.2$

两个总体均值之差的估计

(例题分析)

解：两个总体均值之差在 $1-\alpha$ 置信水平下的置信区间

为

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$= (86 - 78) \pm 1.96 \times \sqrt{\frac{5.8^2}{46} + \frac{7.2^2}{33}}$$

$$= 8 \pm 2.97 = (5.03, 10.97)$$

两所中学高考英语平均分数之差的置信区间为
5.03分~10.97分

两个总体均值之差的估计

(小样本: $\sigma_1^2 = \sigma_2^2$)

❖ 1. 假定条件

- 两个总体都服从正态分布
- 两个总体方差未知但相等: $\sigma_1^2 = \sigma_2^2$
- 两个独立的小样本($n_1 < 30$ 和 $n_2 < 30$)

2. 总体方差的合并估计量

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

估计量 $\bar{x}_1 - \bar{x}_2$ 的抽样标准差

$$\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

两个总体均值之差的估计 (小样本: $\sigma_1^2 = \sigma_2^2$)

两个样本均值之差的标准化的

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

2. 两个总体均值之差 $\mu_1 - \mu_2$ 在 $1 - \alpha$ 置信水平下的置信区间为

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}(n_1 + n_2 - 2) \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

两个总体均值之差的估计

(例题分析)

【例】为估计两种方法组装产品所需时间的差异，分别对两种不同的组装方法各随机安排12名工人，每个工人组装一件产品所需的时间（分钟）下如表。假定两种方法组装产品的时间服从正态分布，且方差相等。试以95%的置信水平建立两种方法组装产品所需平均时间差值的置信区间

方法1		方法2	
28.3	36.0	27.6	31.7
30.1	37.2	22.2	26.0
29.0	38.5	31.0	32.0
37.6	34.4	33.8	31.2
32.1	28.0	20.0	33.4
28.8	30.0	30.2	26.5

两个总体均值之差的估计

(例题分析)

解：根据样本数据计算得

$$\bar{x}_1 = 32.5 \quad \bar{x}_2 = 28.8 \quad s_1^2 = 15.996 \quad s_2^2 = 19.358$$

合并估计量为：

$$s_p^2 = \frac{(12-1) \times 15.996 + (12-1) \times 19.358}{12+12-2} = 17.677$$

$$(32.5 - 28.8) \pm 2.0739 \times \sqrt{17.677 \times \left(\frac{1}{12} + \frac{1}{12} \right)} = 3.7 \pm 3.56$$

两种方法组装产品所需平均时间之差的置信区间为
0.14分钟~7.26分钟

两个总体均值之差的估计

(小样本: $\sigma_1^2 \neq \sigma_2^2$)

❖ 1. 假定条件

- 两个总体都服从正态分布
- 两个总体方差未知且不相等: $\sigma_1^2 \neq \sigma_2^2$
- 两个独立的小样本($n_1 < 30$ 和 $n_2 < 30$)

使用统计量

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(v)$$

两个总体均值之差的估计

(小样本: $\sigma_1^2 \neq \sigma_2^2$)

- ❖ → 两个总体均值之差 $\mu_1 - \mu_2$ 在 $1 - \alpha$ 置信水平下的置信区间为

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}(v) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

自由度 $v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$

两个总体均值之差的估计

(例题分析)

【例】沿用前例。假定第一种方法随机安排12名工人，第二种方法随机安排8名工人，即 $n_1=12$ ， $n_2=8$ ，所得的有关数据如表。假定两种方法组装产品的时间服从正态分布，且方差不相等。以95%的置信水平建立两种方法组装产品所需平均时间差值的置信区间

方法1		方法2	
28.3	36.0	27.6	31.7
30.1	37.2	22.2	26.5
29.0	38.5	31.0	
37.6	34.4	33.8	
32.1	28.0	20.0	
28.8	30.0	30.2	

两个总体均值之差的估计

解：根据样本数据计算得

$$\bar{x}_1 = 32.5 \quad s_1^2 = 15.996 \quad \bar{x}_2 = 27.875 \quad s_2^2 = 23.014$$

自由度为：

$$v = \frac{\left(\frac{15.996}{12} + \frac{23.014}{8} \right)^2}{\frac{(15.996/12)^2}{12-1} + \frac{(23.014/8)^2}{8-1}} = 13.188 \approx 13$$

$$(32.5 - 27.875) \pm 2.1604 \times \sqrt{\frac{15.996}{12} + \frac{23.014}{8}} = 4.625 \pm 4.433$$

两种方法组装产品所需平均时间之差的置信区间为
0.192分钟~9.058分钟

两个总体均值之差的估计 (匹配大样本)

假定条件

- 两个匹配的大样本($n_1 \geq 30$ 和 $n_2 \geq 30$)
 - 两个总体各观察值的配对差服从正态分布
- 两个总体均值之差 $\mu_d = \mu_1 - \mu_2$ 在 $1 - \alpha$ 置信水平下的置信区间为

$$\bar{d} \pm z_{\alpha/2} \frac{\sigma_d}{\sqrt{n}}$$

对应差值的均值

对应差值的标准差

两个总体均值之差的估计

(匹配小样本)

假定条件

- 两个匹配的大样本($n_1 < 30$ 和 $n_2 < 30$)
 - 两个总体各观察值的配对差服从正态分布
- 两个总体均值之差 $\mu_d = \mu_1 - \mu_2$ 在 $1 - \alpha$ 置信水平下的置信区间为

$$\bar{d} \pm t_{\alpha/2}(n-1) \frac{S_d}{\sqrt{n}}$$

两个总体均值之差的估计

(例题分析)

【例】由10名学生组成一个随机样本，让他们分别采用A和B两套试卷进行测试，结果如下表。试建立两种试卷分数之差 $\mu_d = \mu_1 - \mu_2$ 95%的置信区间

10名学生两套试卷的得分

学生编号	试卷A	试卷B	差值d
1	78	71	7
2	63	44	19
3	72	61	11
4	89	84	5
6	91	74	17
5	49	51	-2
7	68	55	13
	76	60	16
	85	77	8
10	55	39	16

STATISTICS

两个总体均值之差的估计

解: 根据样本数据计算得

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n_d} = \frac{110}{10} = 11 \quad s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n_d - 1}} = 6.53$$

$$\begin{aligned} \bar{d} \pm t_{\alpha/2}(n-1) \frac{s_d}{\sqrt{n}} &= 11 \pm 2.2622 \times \frac{6.53}{\sqrt{10}} \\ &= 11 \pm 4.67 \end{aligned}$$

两种试卷所产生的分数之差的置信区间为6.33分
~15.67分

两个总体比率之差的区间估计

- ❖ 1. 假定条件
 - 🌀 两个总体服从二项分布
 - 🌀 可以用正态分布来近似
 - 🌀 两个样本是独立的
- ❖ 2. 两个总体比率之差 $\pi_1 - \pi_2$ 在 $1 - \alpha$ 置信水平下的置信区间为

$$(p_1 - p_2) \pm z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

两个总体比率之差的估计

(例题分析)

【例】在某个电视节目的收视率调查中，农村随机调查了400人，有32%的人收看了该节目；城市随机调查了500人，有45%的人收看了该节目。试以90%的置信水平估计城市与农村收视率差别的置信区间

两个总体比率之差的估计

(例题分析)

解: 已知 $n_1=500$, $n_2=400$, $p_1=45\%$, $p_2=32\%$,

$$1-\alpha=95\%, \quad z_{\alpha/2}=1.96$$

$\pi_1-\pi_2$ 置信度为95%的置信区间为

$$\begin{aligned} & (45\%-32\%) \pm 1.96 \times \sqrt{\frac{45\% \times (1-45\%)}{500} + \frac{32\% \times (1-32\%)}{400}} \\ & = 13\% \pm 6.32\% = (6.68\%, 19.32\%) \end{aligned}$$

城市与农村收视率差值的置信区间为
6.68%~19.32%

两个总体方差比的区间估计

❖ 1. 比较两个总体的方差比

用两个样本的方差比来判断

☞ 如果 S_1^2 / S_2^2 接近于 1, 说明两个总体方差很接近

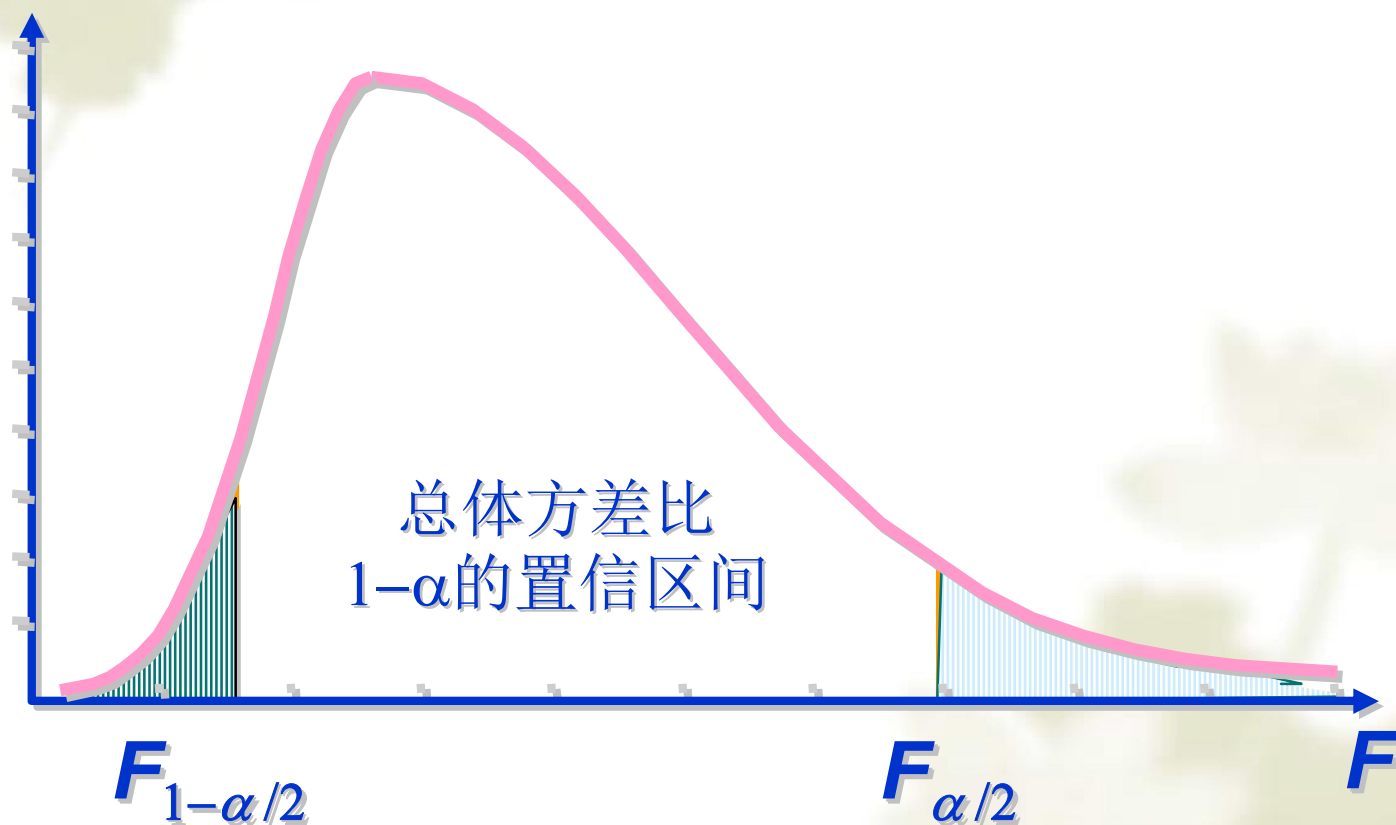
☞ 如果 S_1^2 / S_2^2 远离 1, 说明两个总体方差之间存在差异

总体方差比在 $1-\alpha$ 置信水平下的置信区间为

$$\frac{s_1^2 / s_2^2}{F_{\alpha/2}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2 / s_2^2}{F_{1-\alpha/2}}$$

$$F_{1-\alpha/2}(n_1, n_2) = \frac{1}{F_{\alpha/2}(n_2, n_1)}$$

两个总体方差比的区间估计



方差比置信区间示意图

两个总体方差比的区间估计

(例题分析)

【例】为了研究男女学生在生活费支出(元)上的差异，在某大学各随机抽取**25**名男学生和**25**名女学生，得到下面的结果：

$$\text{男学生: } \bar{x}_1 = 520 \quad s_1^2 = 260$$

$$\text{女学生: } \bar{x}_2 = 480 \quad s_2^2 = 280$$

试以**90%**置信水平估计男女学生生活费支出方差比的置信区间

两个总体方差比的区间估计 (例题分析)

解:根据自由度 $n_1=25-1=24$, $n_2=25-1=24$, 查得
 $F_{\alpha/2}(24)=1.98$, $F_{1-\alpha/2}(24)=1/1.98=0.505$
 σ_1^2 / σ_2^2 置信度为90%的置信区间为

$$\frac{260 / 280}{1.98} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{260 / 280}{0.505}$$

男女学生生活费支出方差比的置信区间为
0.47~1.84

5.4 样本容量的确定

- ❖ 一、估计总体均值时样本容量的确定
- ❖ 二、估计总体比率时样本容量的确定

估计总体均值时样本容量的确定

估计总体均值时样本容量 n 为

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2} \quad \text{其中: } E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

样本容量 n 与总体方差 σ^2 、允许误差 E 、可靠性系数 Z 或 t 之间的关系为

- 与总体方差成正比
- 与允许误差成反比
- 与可靠性系数成正比

估计总体均值时样本容量的确定

(例题分析)

【例】 拥有工商管理学士学位的大学毕业生年薪的标准差大约为2000元，假定想要估计年薪95%的置信区间，希望允许误差为400元，应抽取多大的样本容量？

估计总体均值时样本容量的确定

(例题分析)

解：已知 $\sigma=2000$ ， $E=400$ ， $1-\alpha=95\%$ ， $z_{\alpha/2}=1.96$
应抽取的样本容量为

$$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2} = \frac{(1.96)^2 \times 2000^2}{400^2}$$
$$= 96.04 \approx 97$$

即应抽取97人作为样本

估计总体比率时样本容量的确定

1 根据比率区间估计公式可得样本容量 n 为

$$n = \frac{(z_{\alpha/2})^2 \cdot \pi(1-\pi)}{E^2}$$

其中： $E = z_{\alpha/2} \sqrt{\frac{\pi(1-\pi)}{n}}$

2. E 的取值一般小于0.1
3. π 未知时，可取最大值0.5

估计总体比率时样本容量的确定

(例题分析)

【例】根据以往的生产统计，某种产品的合格率约为90%，现要求允许误差为5%，在求95%的置信区间时，应抽取多少个产品作为样本？

解：已知 $\pi=90%$ ， $\alpha=0.05$ ， $z_{\alpha/2}=1.96$ ， $E=5%$

应抽取的样本容量为

$$\begin{aligned}n &= \frac{(z_{\alpha/2})^2 \cdot \pi(1-\pi)}{E^2} \\ &= \frac{(1.96)^2 \times 0.9 \times (1-0.9)}{0.05^2} \\ &= 138.3 \approx 139\end{aligned}$$

应抽取139个产品作为样本