参数估计

本章主要内容

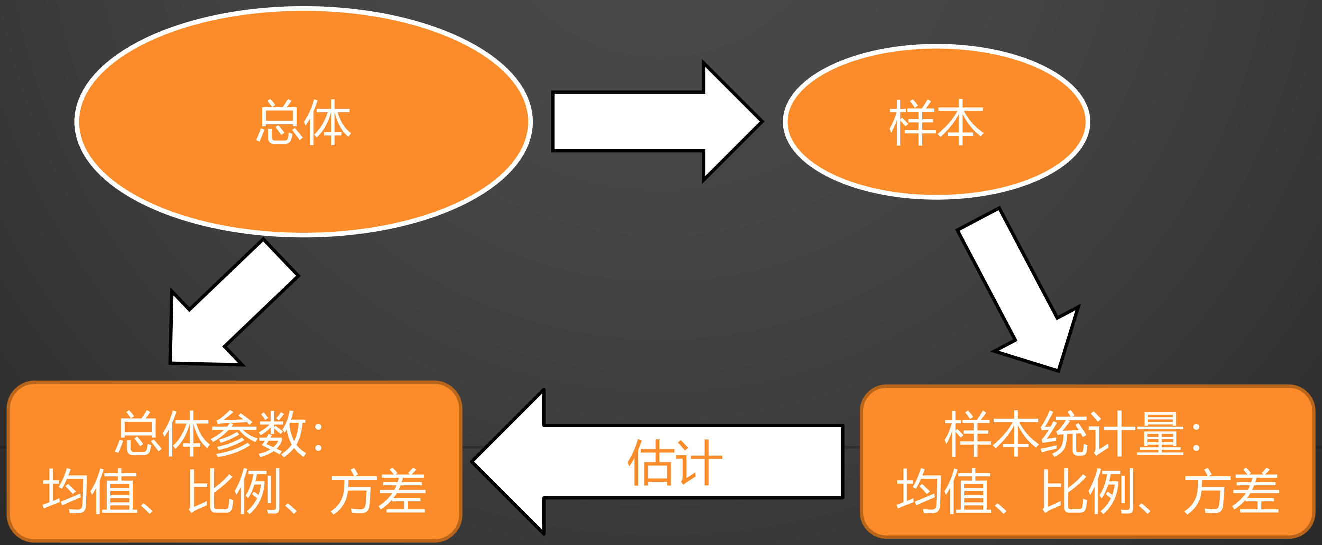
1. 参数估计的基本原理
2. 一个总体参数的区间估计
3. 两个总体参数的区间估计
4. 点估计：极大似然估计法

本章主要内容

1. 参数估计的基本原理
2. 一个总体参数的区间估计
3. 两个总体参数的区间估计
4. 点估计：极大似然估计法

参数估计基本原理

- 参数估计是推断统计的重要内容之一。它是在抽样及抽样分布的基础上，根据样本统计量来推断所关心的总体参数。
- 例如，想知道北京市民的平均收入，最准确的方式就是去询问每个北京市民的收入，然后得到平均值，但实际中往往不可行，因为北京市民有2000多万人，数量太大。所以一般是通过抽样调查以得到样本，统计出样本中的北京市民的平均收入，用样本的平均收入去估计总体的平均收入，这里总体的平均收入称为参数，所以这种方法叫作参数估计。



回顾：几个概念

关于参数估计的几个基本概念：

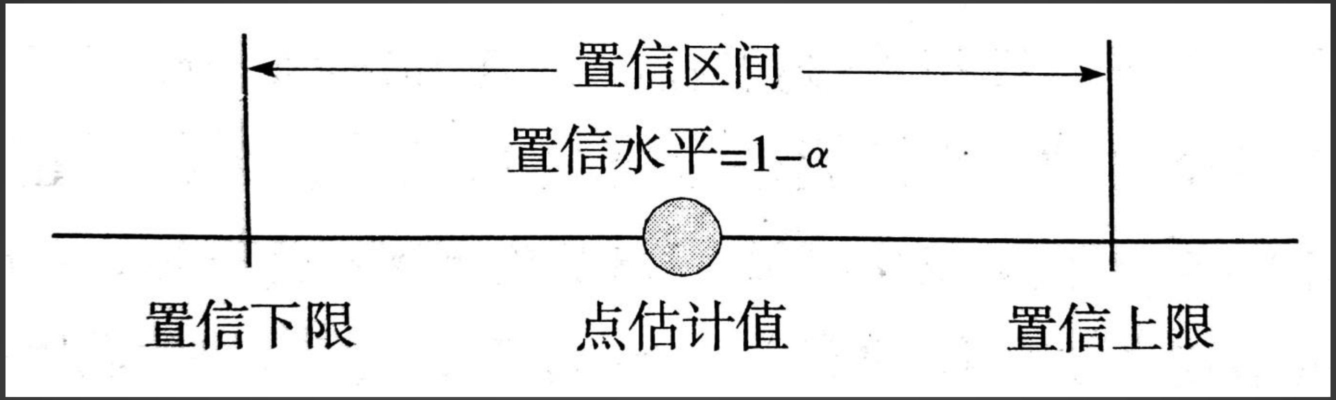
- 总体：全体北京市民称为总体
- 样本：抽取的部分北京市民称为样本
- 参数：全体北京市民的平均收入为参数
- 统计量（或估计量）：抽取的北京市民样本的平均收入为统计量（也可以叫做估计量）
- 估计值：假设计算出抽取的样本的平均收入为3000元，则3000就是估计值。

点估计与区间估计

- 点估计：是指用样本统计量的某个取值直接作为总体参数的估计值。
- 例如，如果用抽取的部分北京市民样本的平均收入3000元直接作为全体北京市民的平均收入的估计值，这就是点估计，3000就是点估计值。
- 但在实际中，很少这样直接估计，因为估计总有误差，我们会说3000左右，这里的“左右”其实表述的是一个范围，也就是所谓的区间估计。

- 区间估计：是在点估计的基础上，给出总体参数估计的一个区间范围，该区间通常由样本统计量加减估计误差得到。
- 例如，在上方例子中，点估计值是3000，假设估计误差为500，则置信区间为（2500,3500）

区间估计



- 置信区间：由样本统计量所构造的总体参数的估计区间。
- 置信水平：区间估计的可信度，α是一个事先确定的概率值，通常取$\alpha = 0.05$或$\alpha = 0.1$。
- 置信上限：由点估计值加估计误差得到。
- 置信下限：由点估计值减估计误差得到。

本章主要内容

1. 参数估计的基本原理
2. 一个总体参数的区间估计
3. 两个总体参数的区间估计
4. 点估计：极大似然估计法

一个总体参数的区间估计

对于一个总体来说，关心的参数有：总体均值、总体比例和总体方差。

三类问题：

1. 总体均值的区间估计
2. 总体比例的区间估计
3. 总体方差的区间估计

一个总体参数的区间估计

对于一个总体来说，关心的参数有：总体均值、总体比例和总体方差。

三类问题：

1. 总体均值的区间估计
2. 总体比例的区间估计
3. 总体方差的区间估计

总体均值的区间估计

假设总体服从正态分布，且方差σ^2已知，则样本均值服从正态分布，且均值为μ，方差为$\frac{\sigma^2}{n}$，接着，将其转化为标准正态分布，即

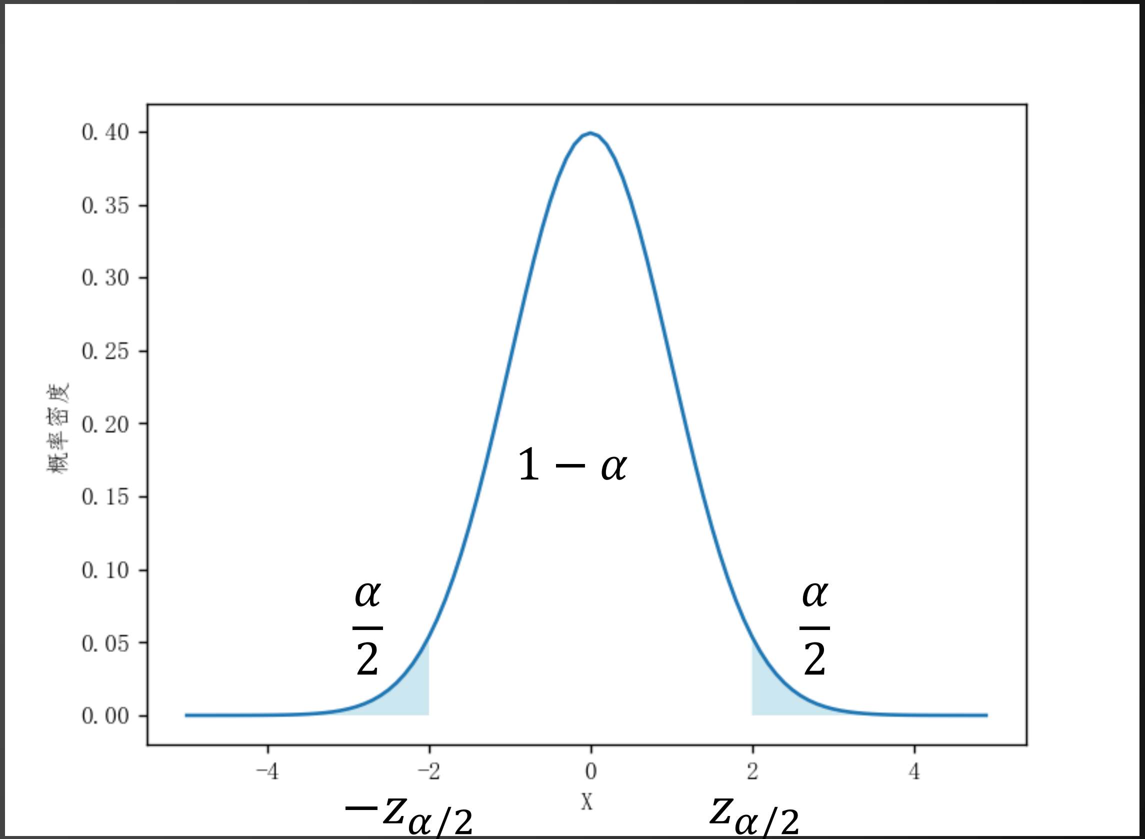
$$z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

假设两个阴影部分的概率之和为α，有

$$P\{-z_{\alpha/2}<\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}<z_{\alpha/2}\}=1-\alpha$$

将上式变形，得

$$P\{\bar{x}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}<\mu<\bar{x}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\}=1-\alpha$$



关于置信区间的说明

根据上式，可以得出总体均值μ在$1-\alpha$置信水平下的置信区间为

$$(\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$$

其中， $\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$称为置信下限， $\bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$称为置信上限。

说明：

- α是一个事先确定的概率值，也称为风险值，通常取$\alpha=0.05$或$\alpha=0.1$，$1-\alpha$则是置信水平
- $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$是估计总体均值时的标准误差。

总体均值的区间估计

【例】一家食品生产企业以生产袋装食品为主，每天的产量大约为8000袋。按规定每袋的重量应为100g。为对产品重量进行检测，企业质检部门经常要进行抽样，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25袋，测得每袋重量如下表所示。

112.5	101	103	102	100.5
102.6	107.5	95	108.8	115.6
100	123.5	102	101.6	102.2
116.6	95.4	97.8	108.6	105
136.8	102.8	101.5	98.4	93.3

25袋食品的重量 单位：g

已知产品重量服从正态分布，且总体标准差为10g。试估计该天产品平均重量的置信区间，置信水平为95%

总体均值的区间估计

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

解：

已知$\sigma = 10, n = 25$，置信水平$1-\alpha = 95\%$，根据Excel公式或者查表，得$z_{\alpha/2} = 1.96$

根据样本数据计算：

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2634}{25} = 105.36$$

根据公式，得

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 105.36 \pm 1.96 \times \frac{10}{\sqrt{25}}$$

即，$105.36 \pm 3.92 = (101.44, 109.29)$

总体均值的区间估计

【例】一家保险公司收集到由36位投保人组成的随机样本，得到每位投保人的年龄数据如下表所示。

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

试建立投保人年龄90%的置信区间。

解:

由于总体方差未知，但是是大样本，可用样本方差来替代的总体方差。

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

根据样本数据计算样本均值和标准差如下:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 39.5$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} = 7.77$$

根据前面所讲的公式，有

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} = 39.5 \pm 1.645 \times \frac{7.77}{\sqrt{36}} = 39.5 \pm 2.13$$

所以，置信区间为$(37.4, 41.6)$

投保人平均年龄90%的置信区间为37.4~41.6岁。

方差未知、小样本

- 如果总体服从正态分布，则无论样本量如何，样本均值$\bar{x}$的抽样分布都服从正态分布。这时，只要总体方差σ^2已知，即时是在小样本的情况下，也可以按照前面的表达式建立总体均值的置信区间。
- 但是，如果总体方差σ^2未知，而且是在小样本情况下，则需要用样本方差s^2代替σ^2 ，这时，样本均值经过标准化以后的随机变量则服从自由度为$n-1$的t分布，即

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

- 因此，需要采用t分布来建立总体均值μ的置信区间。
- 根据t分布建立的总体均值μ在$1-\alpha$置信水平下的置信区间为：

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

例题

【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个，测得其使用寿命（单位：h）如下：

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

试建立该批灯泡平均使用寿命的95%的置信区间。

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

解:

根据样本数据，有

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{23840}{16} = 1490$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}} = \sqrt{\frac{9200}{16-1}} = 24.77$$

当$\alpha = 0.05$时，$t_{\alpha/2}(n-1) = 2.131$，所以得到平均使用寿命的置信区间为：

$$\overline{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = 1490 \pm 2.131 \times \frac{24.77}{\sqrt{16}} = 1490 \pm 13.2$$

置信区间为$(1476.8, 1503.2)$，所以这种灯泡平均使用寿命的95%的置信区间为$1476.8h \sim 1503.2h$。

总结：总体均值的区间估计

- 正态总体：总体σ已知，不论大样本还是小样本，都有

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- 非正态总体：只讨论大样本的情况，总体σ已知

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

如果总体σ未知，用样本标准差s代替总体标准差σ

- 正态总体：小样本，且总体σ未知

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

一个总体参数的区间估计

对于一个总体来说，关心的参数有：总体均值、总体比例和总体方差。

三类问题：

1. 总体均值的区间估计
2. 总体比例的区间估计
3. 总体方差的区间估计

总体比例的区间估计

这里只讨论大样本情况下总体比例的估计问题。由样本比例p的抽样分布可知，当样本量足够大时，比例p的抽样分布可用正态分布近似。

$$E(p) = \pi, \sigma_p^2 = \frac{\pi(1-\pi)}{n}$$

样本比例p经过标准化后的随机变量服从标准正态分布，即

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

与总体均值的区间估计类似，总体比例π在$1-\alpha$置信水平下的置信区间为：

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

【例】某城市想要估计下岗职工中女性所占的比例，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例的置信区间。

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

解：样本比例

$$p = \frac{65}{100} = 65\%$$

根据上面的式子，得

$$p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} = 0.65 \pm 1.96 \times \sqrt{\frac{0.65 \times (1-0.65)}{100}} = 0.65 \pm 0.0935$$

置信区间为：$0.65 \pm 0.0935 = (55.65\%, 74.35\%)$，该城市下岗职工中女性比例的95%的置信区间为 $55.65\% \sim 74.35\%$。

一个总体参数的区间估计

对于一个总体来说，关心的参数有：总体均值、总体比例和总体方差。

三类问题：

1. 总体均值的区间估计
2. 总体比例的区间估计
3. 总体方差的区间估计

总体方差的区间估计

这里只讨论正态总体方差的估计问题，根据前面所学的样本方差的抽样分布可知，样本方差服从自由度为$n-1$的χ^2分布，即

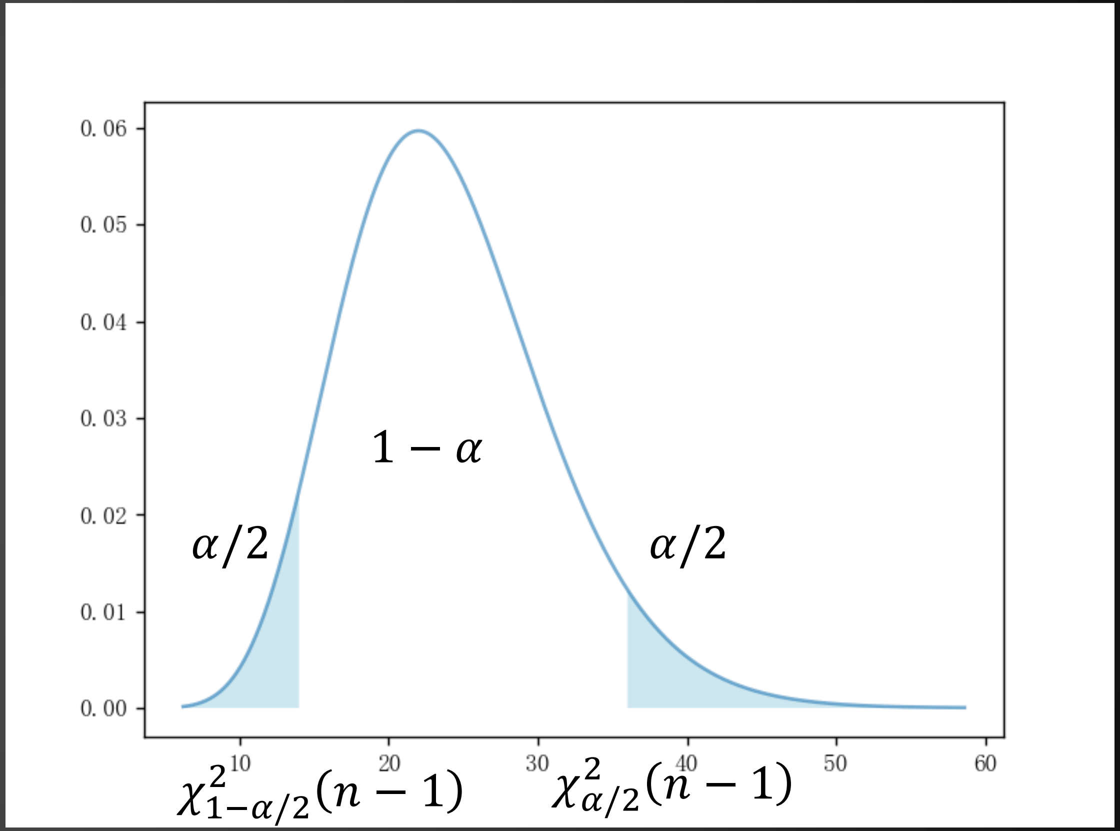
$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

假设置信水平为$1-\alpha$，根据卡方分布的分位点，有

$$P\left\{\chi^2_{1-\frac{\alpha}{2}}(n-1) < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}(n-1)\right\} = 1-\alpha$$

变形后，得

$$P\left\{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right\} = 1-\alpha$$



总体方差的区间估计

所以，方差σ^2的一个置信水平为$1-\alpha$置信区间为

$$\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$$

如果，要得到标准差σ的置信区间，只需要两边开方

$$\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}}\right)$$

总体方差的区间估计：例题

【例】一家食品生产企业以生产袋装食品为主，每天的产量大约为8000袋。按规定每袋的重量应为100g。为对产品重量进行检测，企业质检部门经常要进行抽样，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25袋，测得每袋重量如下表所示。

112.5	101	103	102	100.5
102.6	107.5	95	108.8	115.6
100	123.5	102	101.6	102.2
116.6	95.4	97.8	108.6	105
136.8	102.8	101.5	98.4	93.3

已知产品重量服从正态分布，以95%的置信水平建立食品总体重量标准差的置信区间。

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$$

解：

根据样本数据计算出方差为：

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{2237.02}{25-1} = 93.21$$

接着，计算卡方分布的两个值：

$$\chi^2_{\alpha/2}(n-1) = \chi^2_{0.025}(25-1) = 39.3641$$
$$\chi^2_{1-\alpha/2}(n-1) = \chi^2_{0.975}(25-1) = 12.4011$$

所以，总体方差σ^2的置信区间为：

$$\frac{(25-1) \times 93.21}{39.3641} \leq \sigma^2 \leq \frac{(25-1) \times 93.21}{12.4011}$$

即$56.83 \leq \sigma^2 \leq 180.39$。所以，总体标准差的置信区间为$7.54 \leq \sigma \leq 13.43$，也就是说该企业生产的食品总体重量标准差的95%的置信区间为$7.54g \sim 13.43g$。

本章主要内容

1. 参数估计的基本原理
2. 一个总体参数的区间估计
3. 两个总体参数的区间估计
4. 点估计：极大似然估计法

两个总体参数的区间估计

对于两个总体来说，关心的参数有：两个总体的均值之差、两个总体的比例之差、两个总体的方差比

三类问题：

1. 两个总体的均值之差$\mu_1 - \mu_2$
2. 两个总体的比例之差$\pi_1 - \pi_2$
3. 两个总体的方差比σ_1^2 / σ_2^2

两个总体参数的区间估计

对于两个总体来说，关心的参数有：两个总体的均值之差、两个总体的比例之差、两个总体的方差比

三类问题：

1. 两个总体的均值之差$\mu_1 - \mu_2$
2. 两个总体的比例之差$\pi_1 - \pi_2$
3. 两个总体的方差比σ_1^2/σ_2^2

两个总体均值之差的区间估计

如果两个总体都为正态分布，或两个总体不服从正态分布但两个样本都为大样本($n_1 \geq 30$和$n_2 \geq 30$)，根据抽样分布的知识可知，两个总体均值之差$\overline{x}_1 - \overline{x}_2$的抽样分布服从期望为$\mu_1 - \mu_2$、方差为$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$的正态分布，而两个样本均值之差经标准化后则服从标准正态分布，即

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

当两个总体的方差σ_1^2和σ_2^2都已知时，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：

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

当两个总体的方差σ_1^2和σ_2^2未知时，可用两个样本方差s_1^2和s_2^2来代替，这时，置信区间为：

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

【例】某地区教育管理部门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，有关数据如下表。

中学1	中学2
$n_1 = 46$	$n_2 = 33$
$\bar{x}_1 = 86$	$\bar{x}_2 = 78$
$s_1 = 5.8$	$s_2 = 7.2$

试建立两所中学高考英语平均分数之差95%的置信区间。

解：

根据上面两个样本均值之差的公式，有

$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (86 - 78) \pm 1.96 \times \sqrt{\frac{5.8^2}{46} + \frac{7.2^2}{33}}$$

即$8 \pm 2.97 = (5.03, 10.97)$，所以两所中学高考英语平均分数之差95%的置信区间为5.03分~10.97分。

两个总体均值之差的估计：匹配样本

问题：需要比较两种组装产品的方法的时间差异。

- 一种方式是给每一种方法随机指派N个工人，然后比较这两组工人分别使用这两种方法组装产品的时间差异，虽然是随机指派，但是，这种比较方法存在问题，因为有可能将技术较好的工人指派给了方法1，而将技术较差的工人指派给了方法2，这样就会掩盖这两种方法组装产品所需时间的差异。
- 为了解决这一问题，可以使用匹配样本（matched sample）。可以先指派N个工人给方法1，记录使用方法1组装产品所需的时间，然后将这N个工人再指派给方法2，记录使用方法2组装产品所需的时间，这样就可以消除由于工人之间的差异性造成的两种方法组装时间上的差异。

两个总体均值之差的估计：匹配样本

在使用匹配样本进行估计时，可以先将匹配样本作差，得到一系列差值，如果这些差值服从正态分布，则将这些差值看成一个总体的样本均值的区间估计问题。

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

根据前面样本均值的区间估计，分为以下两种情况：

1. 如果是大样本，则两个总体均值之差$\mu_d = \mu_1 - \mu_2$在$1-\alpha$置信水平下的置信区间为：$\bar{d} \pm z_{\alpha/2}\frac{\sigma_d}{\sqrt{n}}$
2. 如果是小样本，且总体标准差未知，则两个总体均值之差$\mu_d = \mu_1 - \mu_2$在$1-\alpha$置信水平下的置信区间为：$\bar{d} \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}}$

- 说明：$\bar{d}$表示这些差值的均值，s_d表示这些差值的标准差。

两个总体均值之差的估计：匹配样本

【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。

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
5	91	74	17
6	49	51	-2
7	68	55	13
8	76	60	16
9	85	77	8
10	55	39	16

假定两套试卷分数之差服从正态分布，试建立两种试卷分数之差$\mu_d = \mu_1 - \mu_2$的95%的置信区间。

$$\bar{d} \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

解：根据样本数据，得

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n_d} = \frac{110}{10} = 11$$

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

查表或者通过Excel公式，得$t_{0.05/2}(9) = 2.2622$。根据公示得

$$\begin{aligned} \bar{d} \pm t_{\alpha/2}(n-1)\frac{s_d}{\sqrt{n}} &= 11 \pm 2.26 \times \frac{6.53}{\sqrt{10}} \\ &= 11 \pm 4.67 \end{aligned}$$

即（6.3,15.7），两套试卷平均分数之差的95%的置信区间为6.3分~15.7分。

两个总体参数的区间估计

对于两个总体来说，关心的参数有：两个总体的均值之差、两个总体的比例之差、两个总体的方差比

三类问题：

1. 两个总体的均值之差$\mu_1 - \mu_2$
2. 两个总体的比例之差$\pi_1 - \pi_2$
3. 两个总体的方差比σ_1^2/σ_2^2

两个总体比例之差的区间估计

由样本比例的抽样分布可知，从两个二项总体中抽出两个独立的样本，则两个样本比例之差的抽样分布服从正态分布。两个样本的比例之差经标准化后则服从标准正态分布。

$$Z=\frac{(p_1-p_2)-(\pi_1-\pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1}+\frac{\pi_2(1-\pi_2)}{n_2}}}\sim N(0,1)$$

所以，根据正态分布建立的两个总体比例之差$\pi_1-\pi_2$在$1-\alpha$置信水平下的置信区间为：

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

说明：当两个总体比例π_1和π_2未知时，用样本比例p_1和p_2来代替。

例题

【例】在某个电视节目的收视率调查中，从农村随机调查了400人，有32%的人收看该节目；从城市随机调查了500人，有45%的人收看了该节目。试以95%的置信水平估计城市与农村收视率之差的置信区间。

解：记城市收视率$p_1 = 45\%$，农村收视率$p_2 = 32\%$.

当$\alpha = 0.05$时，$z_{\alpha/2} = 1.96$，所以置信区间为：

$$
\begin{aligned}
&(p_1 - p_2) \pm z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \\
&= (45\% - 32\%) \pm 1.96 \times \sqrt{\frac{45\% \times (1-45\%)}{500} + \frac{32\% \times (1-32\%)}{400}} \\
&= 13\% \pm 6.32\%
\end{aligned}
$$

即（6.68%，19.32%），城市与农村收视率之差的95%的置信区间为6.68%~19.32%。

两个总体参数的区间估计

对于两个总体来说，关心的参数有：两个总体的均值之差、两个总体的比例之差、两个总体的方差比

三类问题：

1. 两个总体的均值之差$\mu_1 - \mu_2$
2. 两个总体的比例之差$\pi_1 - \pi_2$
3. 两个总体的方差比σ_1^2/σ_2^2

两个总体方差比的区间估计

- 实际问题中，经常会遇到比较两个总体的方差问题，例如，比较用两种不同方法生产的产品的稳定性，比较不同测量工具的精度等。
- 根据前面所学的知识可知，两个总体的方差比服从F分布，即

$$\frac{s_1^2}{s_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}\sim F(n_1-1,n_2-1)$$

$$P\left\{F_{1-\alpha/2}(n_1-1,n_2-1)<\frac{s_1^2}{s_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}<F_{\alpha/2}(n_1-1,n_2-1)\right\}=1-\alpha$$

变形后，得

$$P\left\{\frac{s_1^2/s_2^2}{F_{\alpha/2}(n_1-1,n_2-1)}<\frac{\sigma_2^2}{\sigma_1^2}<\frac{s_1^2/s_2^2}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right\}=1-\alpha$$

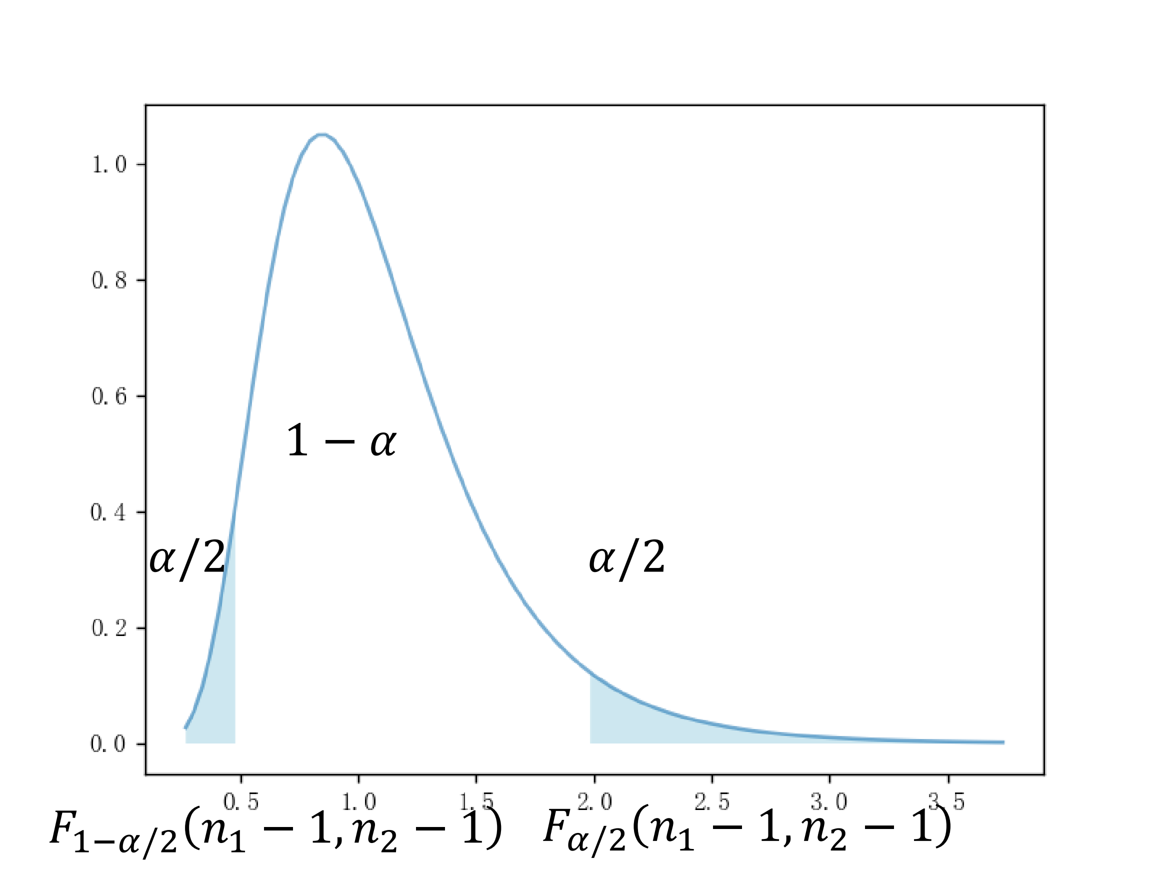
两个总体方差比的区间估计

所以，$\frac{\sigma_2^2}{\sigma_1^2}$的一个置信水平为$1-\alpha$的置信区间为

$$\left(\frac{s_1^2/s_2^2}{F_{\alpha/2}(n_1-1,n_2-1)},\frac{s_1^2/s_2^2}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right)$$

根据前面所学的F分布的知识，可知

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



【例】为研究男女学生在生活费支出（单位：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：

$$\left(\frac{s_1^2/s_2^2}{F_{\alpha/2}(n_1-1,n_2-1)},\frac{s_1^2/s_2^2}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right)$$

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

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

试以90%的置信水平估计男女学生生活费支出方差比的置信区间。

解：首先，将两个F值计算出来（查表或者利用Excel公式）：

$$F_{\alpha/2}(24,24) = F_{0.05}(24,24) = 1.98$$

$$F_{1-\alpha/2}(24,24) = F_{0.95}(24,24) = \frac{1}{1.98} = 0.505$$

根据相关公式，得

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

即$0.47 \le \frac{\sigma_1^2}{\sigma_2^2} \le 1.84$，男女学生生活费用支出方差比的90%的置信区间为0.47~1.84。

本章主要内容

1. 参数估计的基本原理
2. 一个总体参数的区间估计
3. 两个总体参数的区间估计
4. 点估计：极大似然估计法

点估计

点估计，两种常用的方法：矩估计法和极大似然估计法。

矩估计法：由前面的概率论知识可知，期望、方差都是特殊的矩，期望其实是一阶原点矩，方差其实是二阶中心距。

说白了，矩估计法其实就是用样本的均值（一阶原点矩）估计总体的均值

用样本的方差（二阶中心矩）去估计总体的方差。

例如，$X \sim N(\mu,\sigma^2)$，参数μ,σ^2均未知，则

$$\hat{\mu} = \bar{X}, \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$$

极大似然估计法

极大似然估计法，Maximum Likelihood Estimate, MLE，也叫最大似然估计法，通过构造样本的似然函数来求估计量。

极大似然估计法的求解步骤：

1. 写出似然函数：概率相乘；
2. 对似然函数取对数，将相乘变成相加；
3. 对似然函数中的参数求偏导数 ，并令其等于零，得到对数似然方程；
4. 求解对数似然方程，得到估计量的值。

极大似然估计法：例题

【例】有一个抽奖箱，里面有若干红球和白球，除颜色外，其他一模一样。我们每次从中拿出一个后记录下来再放回去，重复10次操作后发现，有7次抽到了红球，3次是白球，请估计红球所占的比例。

解：事件A：{重复10次操作后发现，有7次抽到了红球，3次是白球}，设红球所占的比例为p

p在取什么值的时候，$P(A)$最大?

1、写出似然函数： $P(A)=p^7(1-p)^3$

2、对似然函数取对数，得到对数似然函数

$$\ln P(A)=\ln(p^7(1-p)^3)=lnp^7+\ln(1-p)^3=7lnp+3ln(1-p)$$

3、对似然函数中的参数求导并令其等于零，得到对数似然方程

$$\frac{dlnP(A)}{dp}=\frac{7}{p}+\frac{3}{1-p}=0$$

4、求解对数似然方程解得，$p=\frac{7}{10}=0.7$

极大似然估计法：例题

【例】设$X \sim N(\mu,\sigma^2)$, μ,σ^2为未知参数，$x_1,x_2,\cdots,x_n$是来自X的一个样本值，求μ,σ^2的最大似然估计量。

解：X的概率密度为

$$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2\sigma^2}(x-\mu)^2]$$

1 似然函数为

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp[-\frac{1}{2\sigma^2}(x_i-\mu)^2]$$

$$=(2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2]$$

$$\boxed{e^m\cdot e^n=e^{m+n}}$$

2 对似然函数取对数，有

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

3 分别求L关于μ和σ^2的偏导数，并令其等于零，得到对数似然方程

$$\begin{cases} \frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - n\mu\right) = 0 \\ \frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \end{cases}$$

4 解上述对数似然方程

根据第一个方程，得到$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$，将其代入第二个方程，得$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$

所以，μ和σ^2的最大似然估计量分别为

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

可以看出，它们与矩估计量相同。