

# 第3章 抽样分布总结

# 统计量

## ◆ 统计量 (statistic)

在抽样估计中，用来反映样本总体数量特征的指标称为样本指标，也称为样本统计量或估计量，是根据样本资料计算的、用以估计或推断相应总体指标的综合指标。

常见的样本统计量有：

变量总体	属性总体
样本平均数 $\bar{x}$	样本比例（样本成数） $p$
样本标准差 $s$ 或方差 $s^2$	样本比例标准差 $s_p$ 或方差 $s_p^2$

# 抽样分布的概念

**样本统计量**的概率分布称为抽样分布 (sampling distribution)

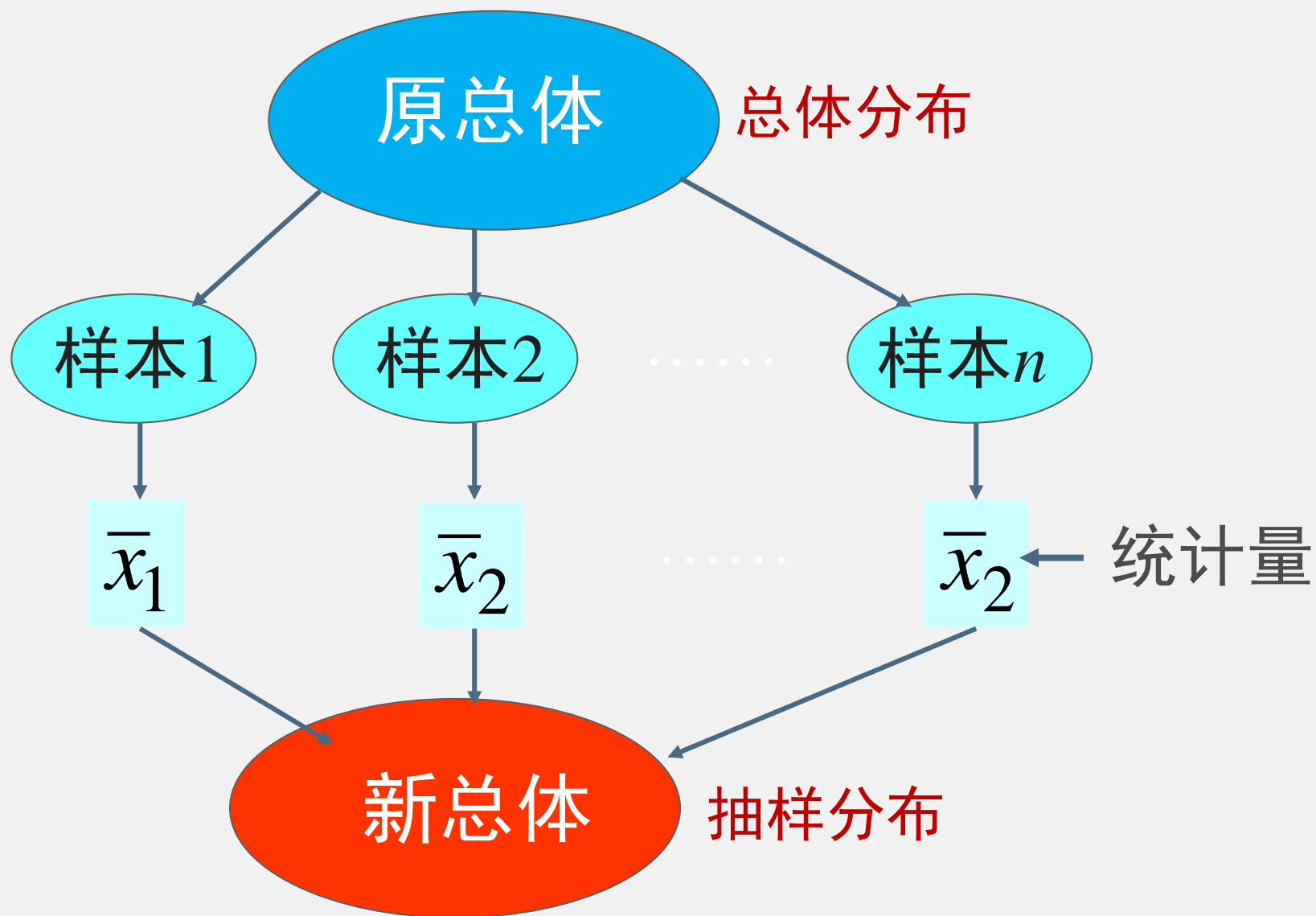
根据样本对总体做出估计和推断，并不是直接用样本本身，而是用样本的统计量来对总体做出估计和判断。但由于从总体中抽取的样本提供的信息仅是总体的一部分，因此它不能提供完全准确的信息，必然存在着一定的误差。即，对于**样本容量相同的多次随机抽样**，得到样本函数的观察值也是不同的，且其取值有一定的概率，即统计量也是一个随机变量，因而也有它的分布，称为抽样分布 (sampling distribution)。

## 相关概念

- ◆ 总体分布：所有元素出现概率的分布.是简单意义上的随机变量对应的频次分布.
- ◆ 样本分布：样本分布有区别于总体分布,它是从总体中按一定的分组标志选出来的部分样本容量.选择的样本在随机变量上的对应的频次分布,样本分布实际上也在趋向总体分布.
- ◆ 抽样分布：是对**样本统计量**概率分布的一种描述方式. 抽样分布是一种概率分布,随机变量是样本统计量.

# 抽样分布的概念

就比如说调查一所中学的所有学生的身高,这就构成了总体,从中随机抽取300个人,这300个人就组成一个样本分布.之后再抽取**若干个**300人组成的样本,每个样本的平均数的分布就是抽样分布。





# 标准差已知时平均数的分布

---

## 1. 抽自正态分布总体

如果从**正态分布总体** $N(\mu, \sigma^2)$  进行独立随机抽样含量为 $n$ 的样本，则样本平均数 $\bar{Y}$ 服从均值为 $\mu$ , 方差为 $\sigma^2/n$ 的正态分布，记作：

$$\bar{Y} \text{ 服从 } N(\mu, \sigma^2/n)$$

将平均数 $\bar{Y}$ 标准化，则

$$u = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}}$$

$u$ 服从 $N(0, 1)$

其中 $\sigma/\sqrt{n}$ 正态总体中进行抽样为平均数的标准误差  
(standard error of mean)



## 2. 抽自非正态分布总体

### 中心极限定理(central limit theorem)

如果被抽总体不是正态分布总体，但具有平均数 $\mu$ 和方差 $\sigma^2$ ，随样本容量 $n$ 的不断增大，样本平均数 $\bar{Y}$ 的分布也越来越接近正态分布，且具有平均数 $\mu$ ，方差 $\sigma^2 / n$

不论总体为何种分布，只要是大样本，就可运用中心极限定理，认为样本平均数的分布是正态分布，在计算样本平均数出现的概率时，样本平均数可按下式进行标准化。

$$u = \frac{\bar{y} - \mu_x}{\sigma_x} = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}}$$



## 标准差未知时的平均数的分布

---

## 标准差未知时的样本平均数的分布----t 分布

若总体的方差是未知的，即标准差  $\sigma$  未知，可以用样本的标准差  $s$  代替总体的标准差  $\sigma$ ，

则变量  $u = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$  变为  $t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

此时  $s$  与  $\sigma$  的值相差比较大

$u$  符合  $N(0, 1)$  分布， $t$  则不服从标准正态分布，而是服从具有  $(n-1)$  自由度的  $t$  分布，其中  $\frac{s}{\sqrt{n}}$

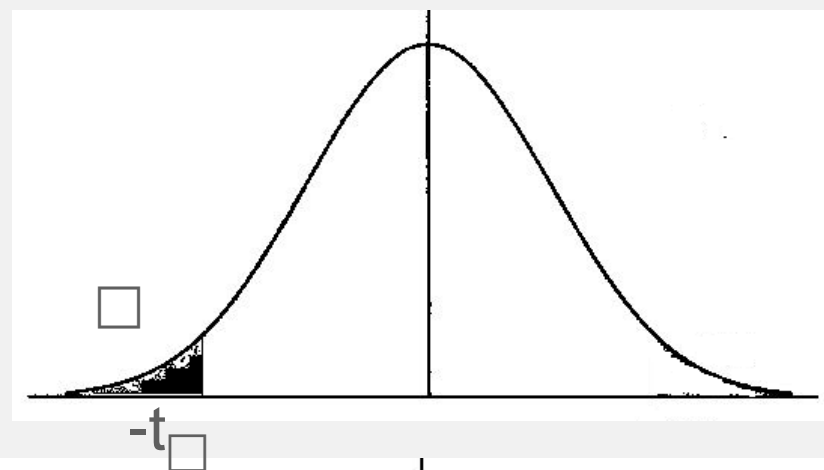
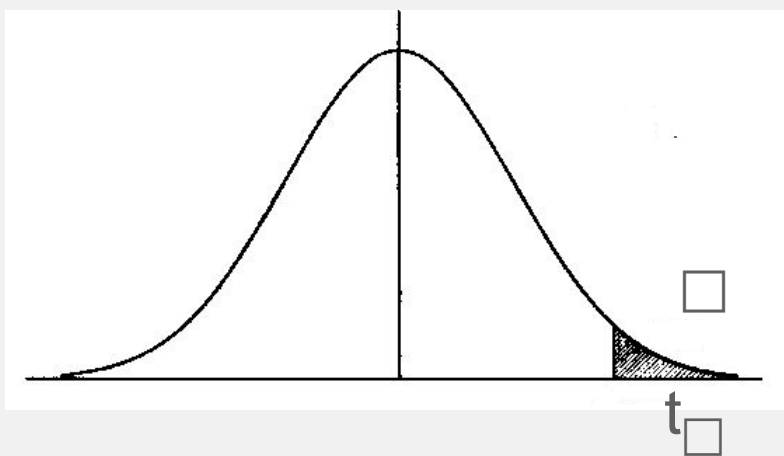
称为**样本标准误差**。

# t 分布的分位数：

单侧分位数 { 上侧分位数  
下侧分位数

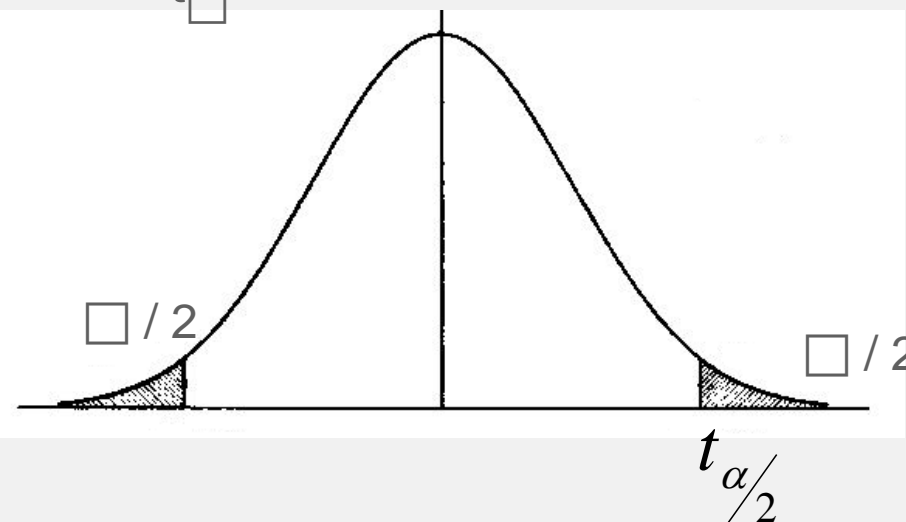
$$P(t \geq t_{\alpha}) = \alpha \quad t_{\alpha}$$

$$P(t \leq -t_{\alpha}) = \alpha \quad -t_{\alpha}$$

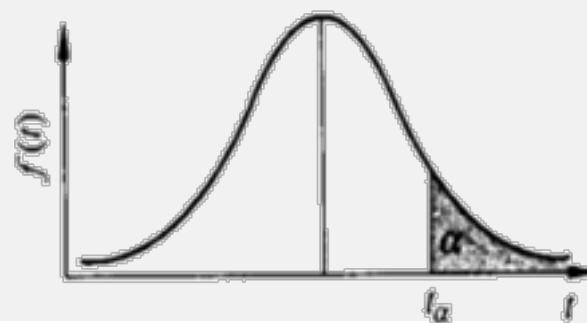


双侧分位数：  $P(|t| \geq t_{\alpha/2}) = \alpha$

$t_{\frac{\alpha}{2}}$  或者  $t_{\alpha} \text{ (双侧)}$

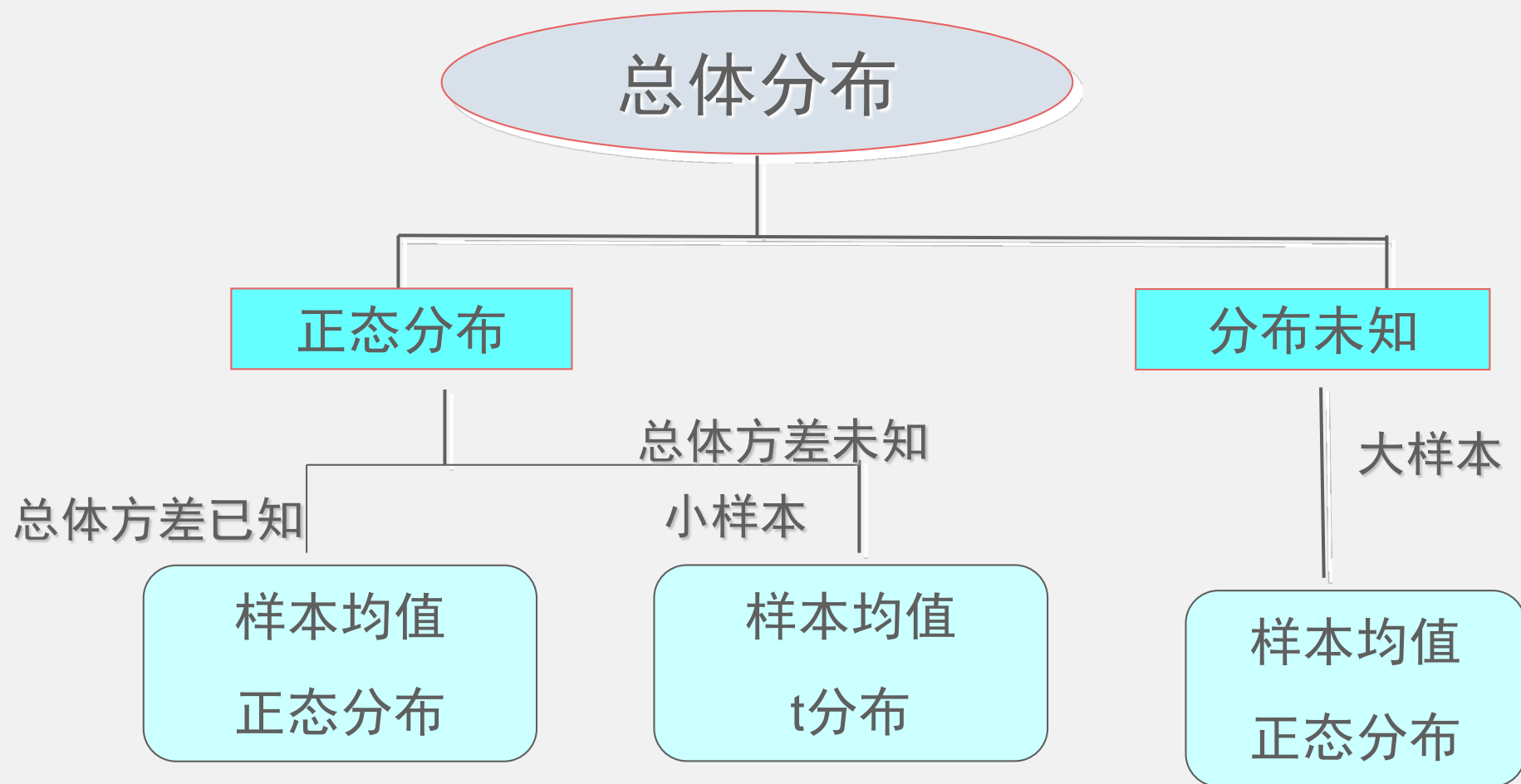


# t分布的临界值表



单侧	$\alpha=0.10$	0.05	0.025	0.01	0.005
双侧	$\alpha=0.20$	0.10	0.05	0.02	0.01
$V=1$	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012

# 抽样分布与总体分布的关系





## 样本方差的分布

---



## 样本的分布

从  $N(\mu, \sigma^2)$  中以  $n$  为样本容量进行抽样，抽取的样本标准差  $s$  为连续型随机变量，当我们以

$\frac{(n-1)s^2}{\sigma^2}$  作为一个新的随机变量时，

称该随机变量为  $s^2$  的 **标准化** 的随机变量，且该随机变量仍为连续型的随机变量。

我们以  $\chi^2$  来命名新的随机变量，则有：

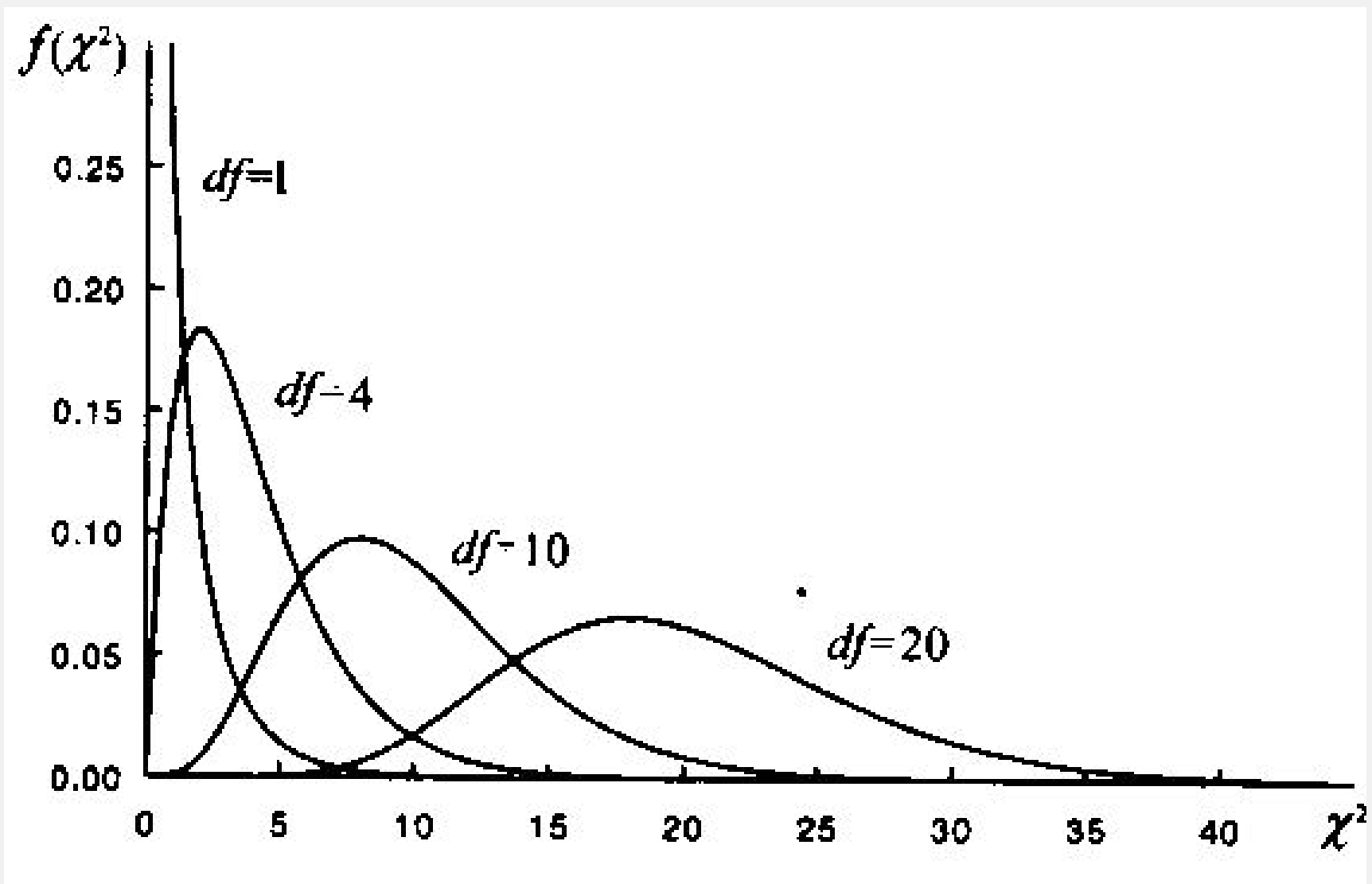
$$\chi_{df}^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{df \cdot s^2}{\sigma^2}$$

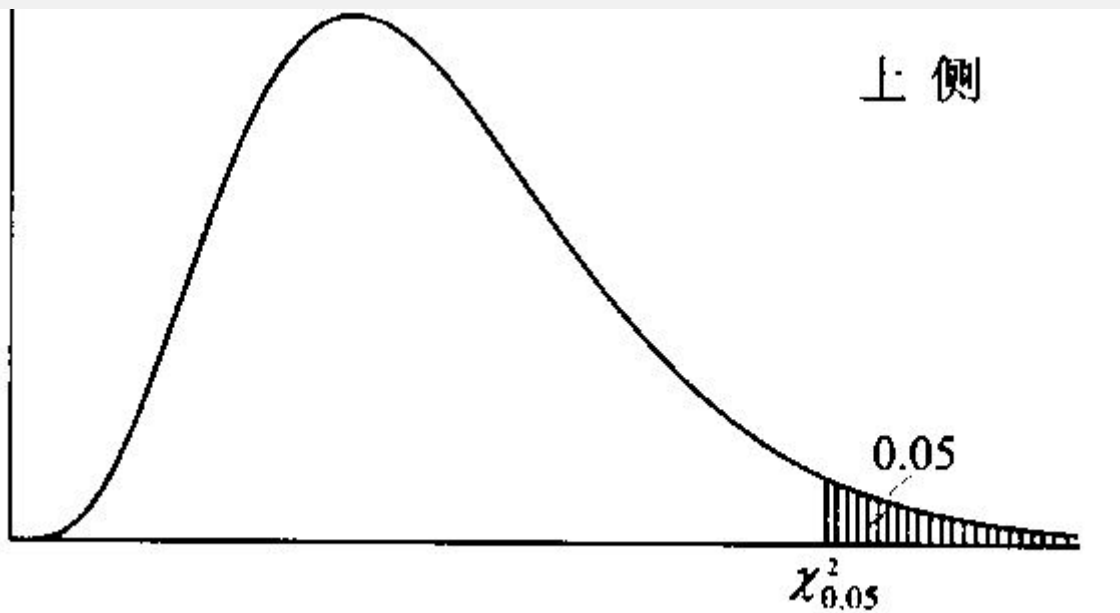
称上式为具有  $n-1$  自由度的卡方。

$\chi^2$  分布是概率分布曲线随自由度  $df$  而改变的一类分布，它的密度函数为：

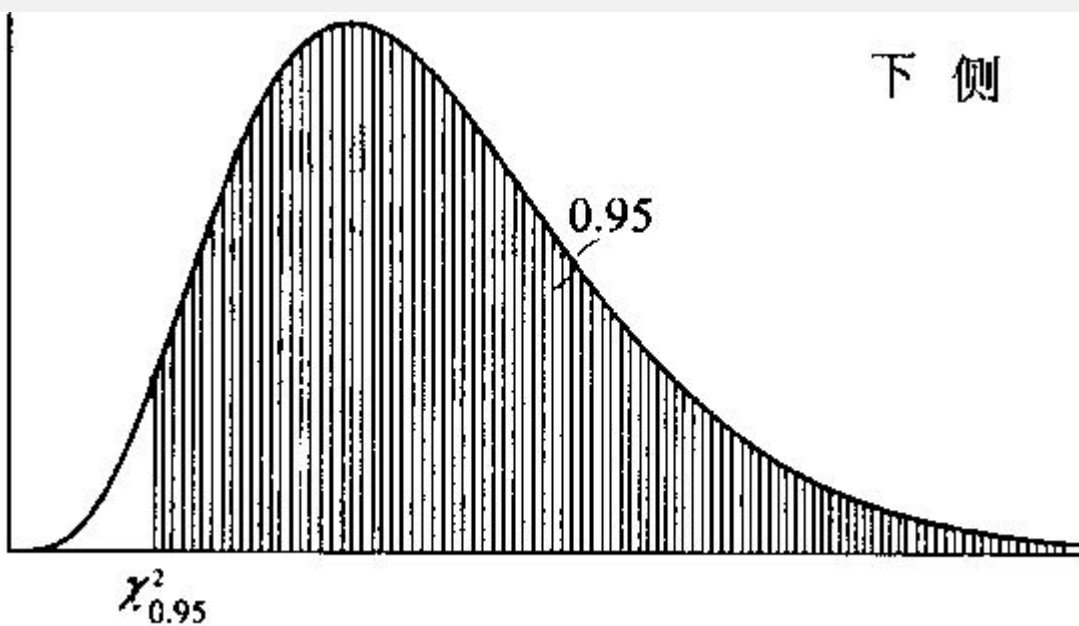
$$f_{df}(\chi^2) = \begin{cases} K(\chi^2)^{\frac{df}{2}-1} e^{-\frac{\chi^2}{2}}, & \chi^2 > 0 \\ 0 & \chi^2 < 0 \end{cases}$$

## $\chi^2$ 分布的分布曲线及分位数





$$P(\chi^2_{df} > \chi^2_{\alpha}) = \alpha$$



$$P(\chi^2_{df} > \chi^2_{1-\alpha}) = 1 - \alpha$$

# $\chi^2$ 分布

## 性质

$\chi^2$  分布随机变量的取值范围为  $(0, \infty)$

$\chi^2$  分布  $Y \sim \chi^2(n)$ , 则期望  $E(Y)=n$ , 均方  $\text{var}(Y) = 2n$

若  $Y_1 \sim \chi^2(n)$ ,  $Y_2 \sim \chi^2(m)$ , 且相互独立, 则

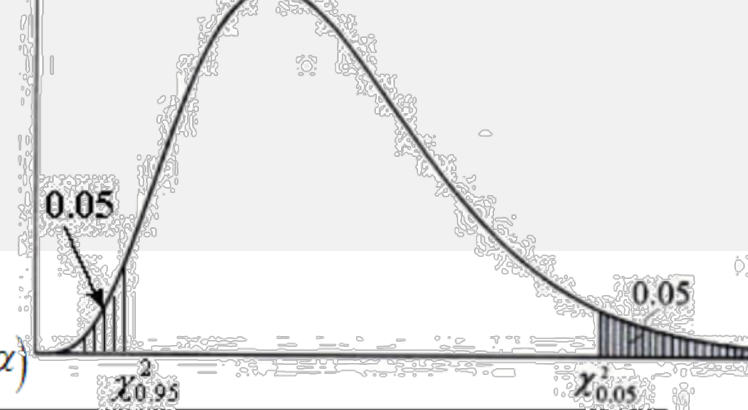
$$Y_1 \pm Y_2 \sim \chi^2(n \pm m)$$

$\chi^2$  分布为非对称分布, 其分布曲线的形状由自由度决定, 自由度越大, 分布越趋于对称

当  $n \rightarrow \infty$ ,  $\chi^2(n) \rightarrow N(n, 2n)$

# $\chi^2$ 分布上侧分位数表

$$P(X \geq \chi_{\alpha}^2) = \alpha$$



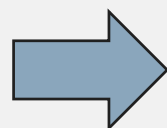
附表三  $\chi^2$  分布上侧分位数表

$$(P\{\chi^2(n) > \chi_{\alpha}^2(n)\} = \alpha)$$

$\alpha$ $n$	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
1	0.00004	0.00016	0.001	0.004	0.016	0.102	0.455	1.323	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	0.575	1.386	2.773	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	1.213	2.366	4.108	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	1.923	3.357	5.385	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	2.675	4.351	6.626	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	3.455	5.348	7.841	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	4.255	6.346	9.037	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	5.071	7.344	10.219	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	5.899	8.343	11.389	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	6.737	9.342	12.549	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	7.584	10.341	13.701	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	8.438	11.340	14.845	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	9.299	12.340	15.984	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	10.165	13.339	17.117	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	11.037	14.339	18.245	22.307	24.996	27.488	30.578	32.801

## 知识小结

总体  $X \sim N(\mu, \sigma^2)$



样本平均数

总体方差已知

$$u = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

总体方差未知  
(小样本)

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

样本方差

$$\chi_{df}^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{df \cdot s^2}{\sigma^2}$$



## 从两个正态总体中抽取的样本统计量的分布

---



## 1. 标准差 $\sigma_i$ 已知，两个平均数的和与差的分布

两个正态总体分别为： $N(\mu_1, \sigma_1^2)$  和  $N(\mu_2, \sigma_2^2)$ ，分别进行独立随机抽样含量为 $n_1$ 和 $n_2$ 的样本，则两个样本样本平均数的和与差的分布为：

$\bar{y}_1 \pm \bar{y}_2$ 服从均值为  $(\mu_1 \pm \mu_2)$ ，方差为  $(\sigma_1^2 / n_1 + \sigma_2^2 / n_2)$  的正态分布，记作：

$$\bar{y}_1 \pm \bar{y}_2 \text{ 服从 } N(\mu_1 \pm \mu_2, \sigma_1^2 / n_1 + \sigma_2^2 / n_2)$$

将 $\bar{y}_1 \pm \bar{y}_2$ 标准化, 则

$$u = \frac{(\bar{y}_1 \pm \bar{y}_2) - (\mu_1 \pm \mu_2)}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}}$$

$u$ 服从 $N(0, 1)$

## 2. 标准差 $\sigma_i$ 未知但相等，两个平均数的和与差的分布

两个正态总体分别为： $N(\mu_1, \sigma_1^2)$  和  $N(\mu_2, \sigma_2^2)$ ，分别进行独立随机抽样含量为 $n_1$ 和 $n_2$ 的样本。其中 $\sigma_1$ 与 $\sigma_2$ 未知，但是 $\sigma_1 = \sigma_2 = \sigma$ 则两个样本样本平均数的和与差的分布为：

$\bar{y}_1 \pm \bar{y}_2$ 服从 $df_1 + df_2$ 自由度的t分布。 $df_1 = n_1 - 1$ ； $df_2 = n_2 - 1$ 。

$$t_{df_1+df_2} = \frac{(\bar{y}_1 \pm \bar{y}_2) - (\mu_1 \pm \mu_2)}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

如果 $n_1=n_2$ ,

$$t_{2n-2} = \frac{(\bar{y}_1 \pm \bar{y}_2) - (\mu_1 \pm \mu_2)}{\sqrt{\frac{s_1^2 + s_2^2}{n}}}$$

### 3. 两个样本方差比的分布

两个正态总体分别为： $N(\mu_1, \sigma_1^2)$  和  $N(\mu_2, \sigma_2^2)$ ，分别进行独立随机抽样含量为  $n_1$  和  $n_2$  的样本。标准化的样本方差比的分布称为F分布

$$F_{df_1, df_2} = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$$