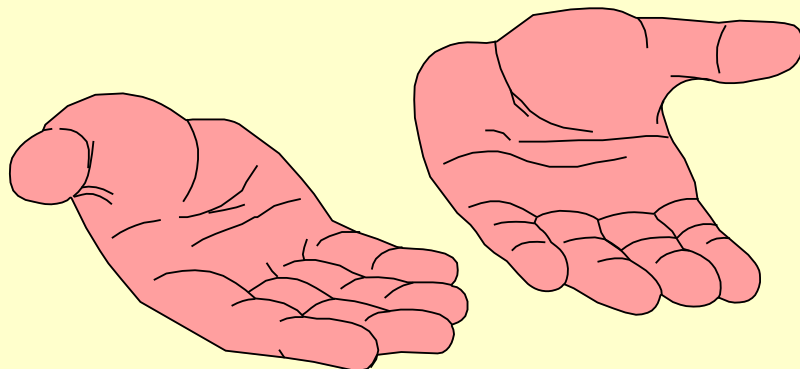


第六章 样本及抽样分布

- 引言
- 随机样本
- 抽样分布
- 正态总体的抽样分布定理



引言

数理统计学是数学的一个重要分支，它研究怎样有效地收集、整理和分析带有随机性的数据，以对所考察的问题作出推断或预测，并为采取一定的决策和行动提供依据和建议。

例 制衣厂为了合理的确定服装各种尺码的生产比例，需要调查人们身高的分布。现从男性成人人群中随机选取100人，得到他们的身高数据

(1) 试推断男性成人身高 X 的概率密度

(2)若已知 X 服从正态分布 $N(\mu, \sigma^2)$,试估计参数的 μ, σ^2 值

已知“总体”的分布类型,对分布中的未知参数所进行的统计推断属于“参数统计”。

概率论是在随机变量的分布已知条件下,研究随机变量的各种性质.例如求某个随机变量的分布函数、数字特征等等.

数理统计是在随机变量的分布完全未知,或不完全已知的情况下,通过对收集整理的数据进行分析,研究随机变量的分布并做出各种统计推断.

6.1 随机样本

一、总体与样本

总体：在统计学中,把研究对象的全体称为总体。

个体：总体中的每个研究对象称为个体。

在实际应用中人们关心的是研究对象(即总体)的某(些)项数量指标,而这个数量指标常常是事先无法预知的,所以它是一个随机变量。

从本质上讲,总体就是所研究的随机变量或随机变量的分布。

为了研究总体的分布以及其它统计特性, 必须按照一定的规则从总体中抽取一部分个体, 根据获得的个体数据来对总体的分布和统计特性做出推断.

2. **样本**来自总体的部分个体 X_1, \dots, X_n ;

如果满足:

(1) **同分布性**: $X_i, i=1, \dots, n$ 与总体同分布

(2) **独立性**: X_1, \dots, X_n 相互独立

则称为**容量为n 的简单随机样本**, 简称**样本**。

而称 X_1, \dots, X_n 的观察值 x_1, \dots, x_n 为样本观察值, 抽取样本的过程称为抽样。

来自总体 X 的随机样本 X_1, \dots, X_n 可记为

$$X_1, \dots, X_n \stackrel{i.i.d}{\sim} X \text{ 或 } F(x), f(x), P_k, \dots$$

显然，样本联合分布等于一维边缘分布的乘积

$$F^*(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i)$$

$$f^*(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i)$$

$$P\{X_1 = x_1, \dots, X_n = x_n\} = P\{X_1 = x_1\} \dots P\{X_n = x_n\}$$

后两者分别为连续型和离散型随机变量的似然函数。

例1 若 X_1, \dots, X_n 是总体 $X \sim B(1, p)$ 的样本,
求 (X_1, \dots, X_n) 的联合分布律.

解: 总体 X 的分布律为 关系

$$p(\mathbf{x}) = P\{X = \mathbf{x}\} = p^{\mathbf{x}}(1-p)^{1-\mathbf{x}}, \mathbf{x} = 0, 1.$$

所以 (X_1, \dots, X_n) 的联合分布律为

$$\begin{aligned} P\{X_1 = \mathbf{x}_1, \dots, X_n = \mathbf{x}_n\} &= \prod_{i=1}^n p(\mathbf{x}_i) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^n x_i} (1-p)^{n-\sum_{i=1}^n x_i} \quad x_i = 0, 1, \quad i = 1, \dots, n. \end{aligned}$$

例2 若 X_1, \dots, X_n 是总体 $X \sim N(\mu, \sigma^2)$ 的样本, 求
 (X_1, \dots, X_n) 的联合概率密度

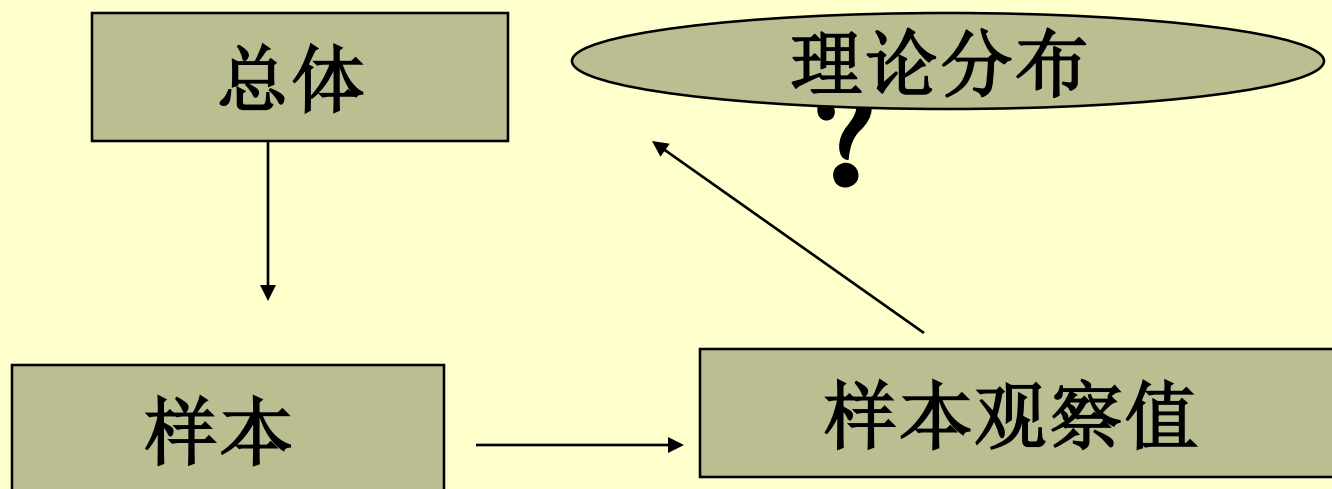
解: 总体 X 的概率密度为

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty.$$

所以 (X_1, \dots, X_n) 的联合概率密度为

$$\begin{aligned} f(\mathbf{x}_1, \dots, \mathbf{x}_n) &= \prod_{i=1}^n f(\mathbf{x}_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\ &= (2\pi)^{-\frac{n}{2}} (\sigma)^{-n} e^{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}} \quad -\infty < x_i < \infty, i = 1, \dots, n. \end{aligned}$$

3.总体、样本、样本观察值的关系



统计是从手中已有的资料——样本观察值，去推断总体的情况——总体分布。样本是联系两者的桥梁。总体分布决定了样本取值的概率规律，也就是样本取到样本观察值的规律，因而可以用样本观察值去推断总体

来自总体的样本含有总体的信息，是总体的代表. 我们作统计推断的依据就是样本. 但样本中包含的信息比较分散，一般不能直接用于统计推断.

为了把分散在样本中的信息集中起来，我们针对不同的问题构造各种适当的样本函数. 这种样本函数我们称为统计量。

二、统计量

定义：称样本 X_1, \dots, X_n 的函数 $g(X_1, \dots, X_n)$ 是总体 X 的一个统计量, 如果 $g(X_1, \dots, X_n)$ 不含未知参数.

设 (x_1, \dots, x_n) 是相应于样本 (X_1, \dots, X_n) 的样本观察值。

则称 $g(x_1, \dots, x_n)$ 是 $g(X_1, \dots, X_n)$ 的观察值。

注：统计量是随机变量。

几个常见的统计量

(1) 样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

样本均值是反映总体数学期望所在位置信息的一个统计量,是总体数学期望的无偏估计.

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X)$$

$$D(\bar{X}) = D\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n D(X_i) = \frac{D(X)}{n}$$

(2) 样本方差

样本标准差

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$S = \sqrt{S^2}$$

样本方差与样本标准差反映了数据取值分散与集中的程度,即反映了总体方差与标准差的信息.

样本方差是总体方差的无偏估计。 $E(S^2) = D(X)$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right]$$

证明:
$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2)$$

样本k阶矩

$$= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 \right)$$

$$= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - 2\bar{X}n\bar{X} + n\bar{X}^2 \right)$$

$$= \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right]$$

$$E\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right] = E\left(\frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2\right) = \frac{n-1}{n} D(X)$$

$$E(S^2) = D(X)$$

(3)样本**k**阶原点矩 $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$

样本**k**阶中心矩 $B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$

它们分别反映了总体**k**阶原点矩与**k**阶中心矩的信息.

样本**k**阶原点矩是总体**k**阶原点矩的无偏估计

样本**k**阶中心矩是总体**k**阶中心矩的无偏估计

(4) 最大次序统计量与最小次序统计量

次序统计量

统计量的观察值

$X_{(i)}$ 被称为样本的第*i*个次序统计量,它是样本

(X_1, \dots, X_n) 的满足如下条件的函数:

每当样本得到一组观察值 (x_1, x_2, \dots, x_n) 时,将它们从小到大排列为 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$,第*i*个值 $x_{(i)}$ 便是 $X_{(i)}$

的观察值, $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 称为该样本的次序统计量.

$X_{(1)}$ 称为该样本的最小次序统计量, $X_{(n)}$ 称为该样本的最大次序统计量, 也分别称两者为最小值与最大值.

它们的观察值分别为:

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

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

$$a_k = \frac{1}{n} \sum_{i=1}^n x_i^k, \quad k = 1, 2, \dots, s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k, \quad k = 1, 2, \dots$$

分别称为样本均值、样本方差、样本标准差、
样本 k 阶原点矩、样本 k 阶中心矩的**观察值**。

6.2 抽样分布

统计量的分布称为抽样分布。数理统计中常用到如下三个分布：

χ^2 分布、 t 分布和F分布。

一、 χ^2 分布

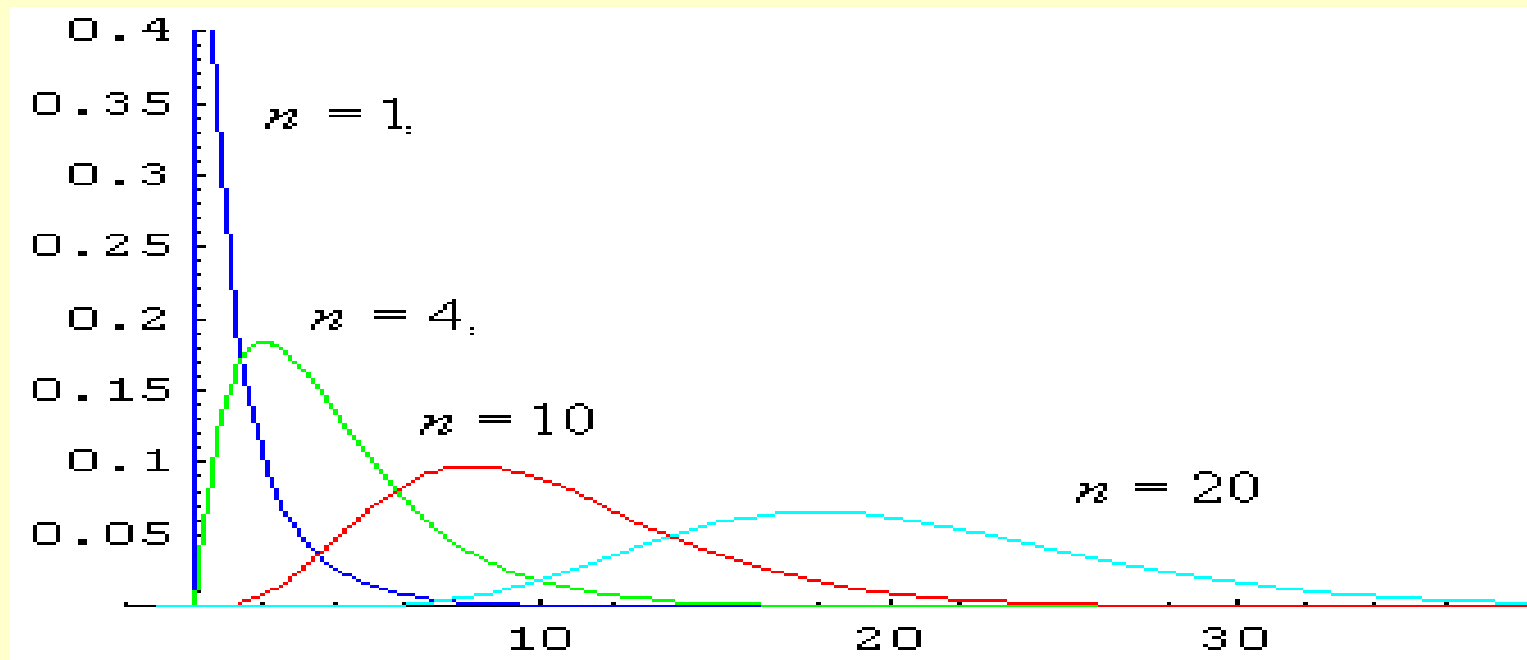
1. 构造 设 $X_1, \dots, X_n \stackrel{iid}{\sim} N(0,1)$, 则 $\chi^2 = \sum_{i=1}^n X_i^2 \sim \chi^2(n)$.

$\chi^2(n)$ 称为自由度为 n 的 χ^2 分布.

$$\chi^2(n) = \sum_{i=1}^n N(0,1)^2$$

2. χ^2 —分布的密度函数 $f(y)$ 曲线

$$f(y) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$



χ^2 分布的性质

3.性质:

a.分布可加性: 若 $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$,
 X, Y 独立, 则 $X + Y \sim \chi^2(n_1 + n_2)$

b.期望与方差: 若 $X \sim \chi^2(n)$, 则
 $E(X) = n$, $D(X) = 2n$

二、t分布

戈塞特

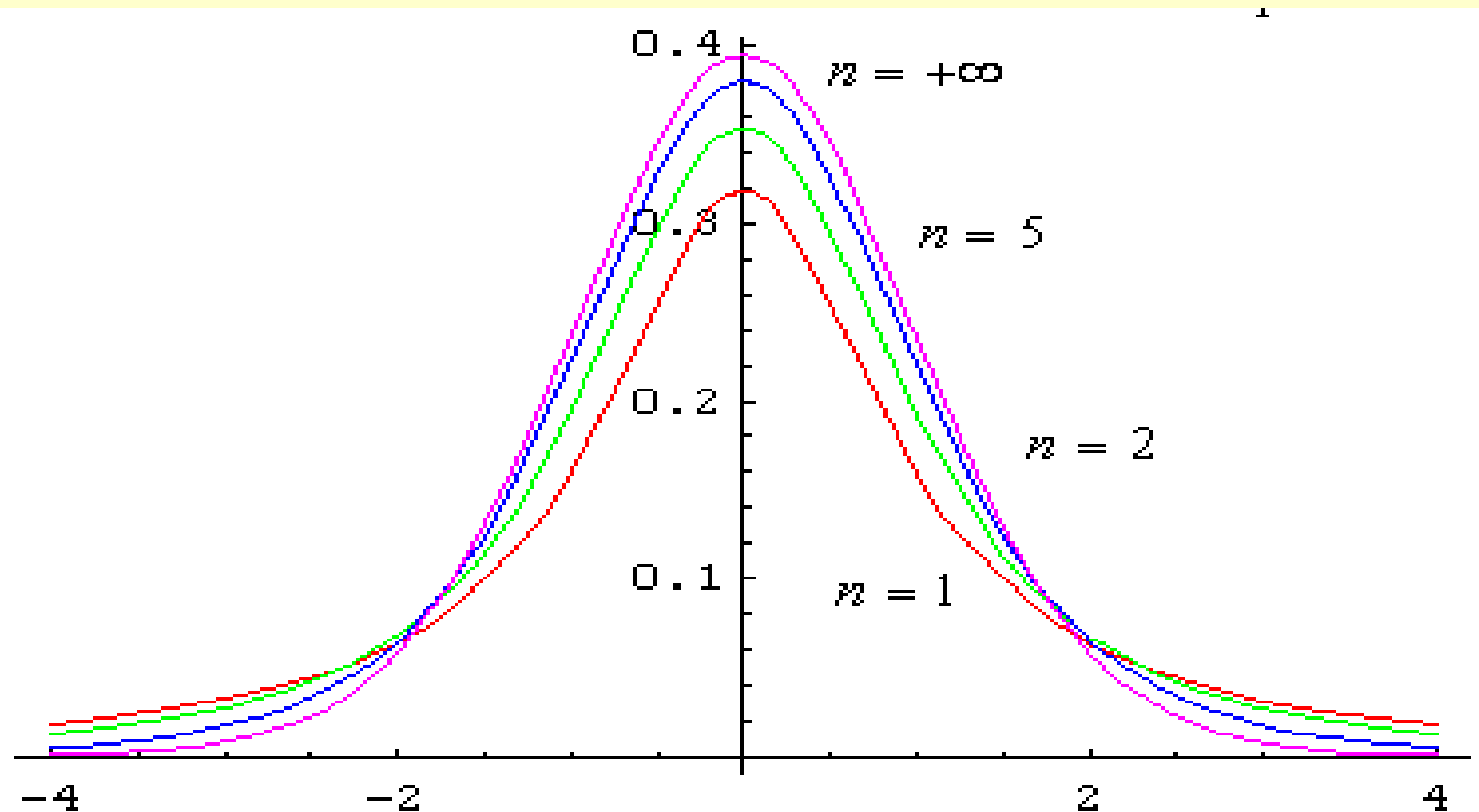
1.构造 若 $X \sim N(0, 1)$, $Y \sim \chi^2(n)$, X 与 Y 独立, 则

$$t = \frac{X}{\sqrt{Y/n}} \sim t(n) \qquad t(n) = \frac{N(0,1)}{\sqrt{\chi^2(n)/n}}$$

$t(n)$ 称为自由度为 n 的 t 分布。

$t(n)$ 的概率密度为

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < t < \infty$$



2. 基本性质:

(1) $f(t)$ 关于 $t=0$ (纵轴) 对称。

(2) $f(t)$ 的极限为 $N(0, 1)$ 的密度函数, 即

$$\lim_{n \rightarrow \infty} f(t) = \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, -\infty < x < \infty$$

三、F分布

1.构造 若 $U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$, U , V 独立, 则

$$F = \frac{U / n_1}{V / n_2} \sim F(n_1, n_2). \quad F(n_1, n_2) = \frac{\chi^2(n_1) / n_1}{\chi^2(n_2) / n_2}$$

称为第一自由度为 n_1 , 第二自由度为 n_2 的F分布, 其概率密度为

$$h(y) = \begin{cases} \frac{\Gamma(\frac{n_1+n_2}{2})(n_1/n_2)^{n_1/2} y^{\frac{n_1}{2}-1}}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})(1+\frac{n_1}{n_2}y)^{(n_1+n_2)/2}}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

四、连续型分布的分位点

1、设随机变量 $X \sim F(x)$, 给定常数 $\alpha: 0 < \alpha < 1$,

若存在 x_α , 满足 $P\{X > x_\alpha\} = \alpha$,

则称 x_α 为分布 $F(x)$ 的上(侧) α 分位点.

2、设随机变量 $X \sim N(0,1)$, 给定常数 $\alpha: 0 < \alpha < 1$,

若存在 Z_α , 满 $P\{X > Z_\alpha\} = \alpha$

则称 Z_α 为标准正态分布的上侧 α 分位点.

标准正态分布的分位点

α	0.001	0.005	0.01	0.025	0.05	0.10
z_α	3.090	2.576	2.327	1.96	1.645	1.282

$$z_\alpha = -z_{1-\alpha}$$

也可以利用标准正态分布的分布函数来求其分位点

$$\Phi(z_\alpha) = 1 - \alpha \quad \Phi(z_{\alpha/2}) = 1 - \alpha/2$$

标准正态分布函数表

N(0,1)分布函数 $\Phi(x)$

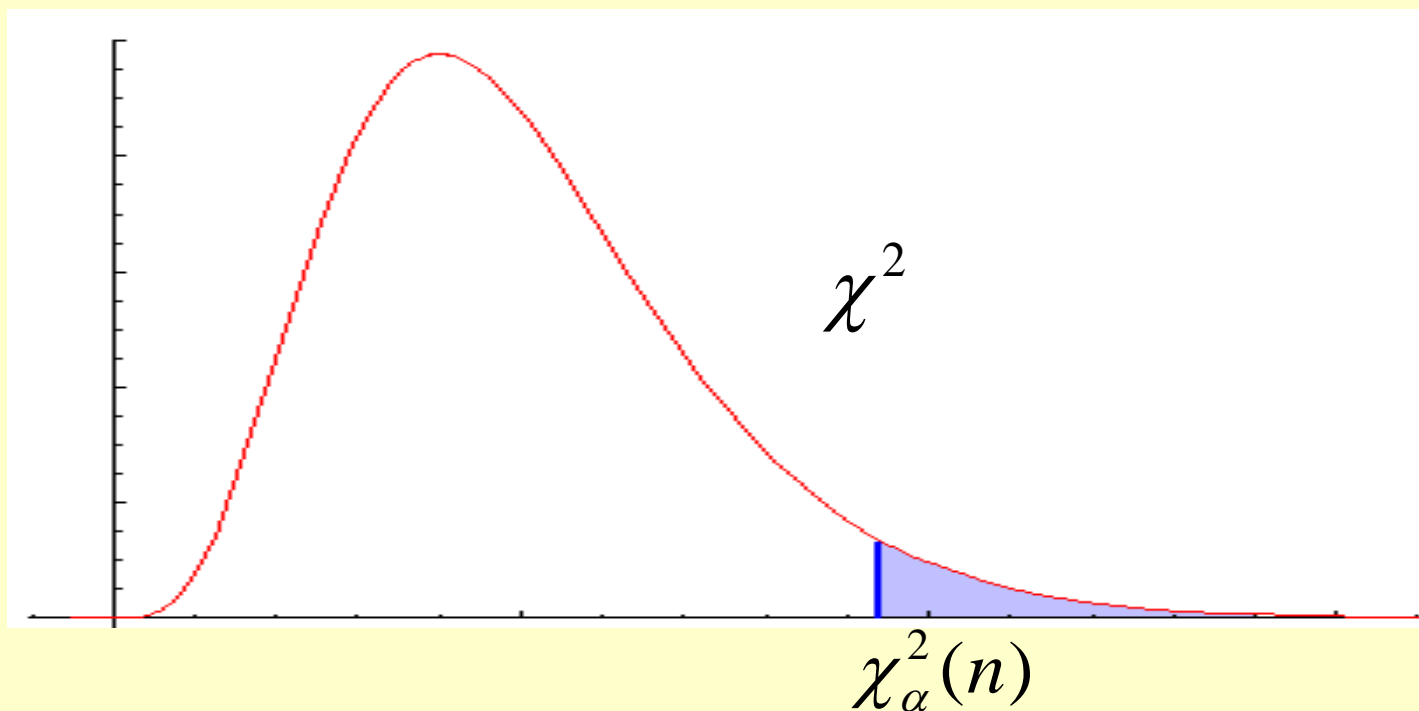
x	0	1	2	3	4	5	6	7	8	9	
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	0
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	0.1
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	0.2
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	0.3
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	0.4
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	0.5
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	0.6
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	0.7
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	0.8
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	0.9
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	1
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	1.1
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	1.2
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	1.3
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1.4
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	1.5
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	1.6
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	1.7
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	1.8
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	1.9
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	2
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	2.1
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	2.2
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	2.3
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	2.4
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	2.5
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	2.6
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	2.7
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	2.8
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	2.9
3	0.9987	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000	3
	3	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	

χ^2 分布分位点

设 $\chi^2 \sim \chi^2(n)$, 对于 α : $0 < \alpha < 1$, 若存在 $\chi_\alpha^2(n)$ 满足

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

则称 $\chi_\alpha^2(n)$ 为 $\chi^2(n)$ 分布的上 α 分位点。



χ^2 分布分位点

n	0.995	0.99	0.975	0.95	0.9	0.75	0.25	0.1	0.05	0.025	0.01	0.005	n
1	0.000	0.000	0.001	0.004	0.016	0.102	1.323	2.706	3.841	5.024	6.635	7.879	1
2	0.010	0.020	0.051	0.103	0.211	0.575	2.773	4.605	5.991	7.378	9.210	10.597	2
3	0.072	0.115	0.216	0.352	0.584	1.213	4.108	6.251	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	1.064	1.923	5.385	7.779	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	1.610	2.675	6.626	9.236	11.070	12.833	15.086	16.750	5
6	0.676	0.872	1.237	1.635	2.204	3.455	7.841	10.645	12.592	14.449	16.812	18.548	6
7	0.989	1.239	1.690	2.167	2.833	4.255	9.037	12.017	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	3.490	5.071	10.219	13.362	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	4.168	5.899	11.389	14.684	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	4.865	6.737	12.549	15.987	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	5.578	7.584	13.701	17.275	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	6.304	8.438	14.845	18.549	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	7.042	9.299	15.984	19.812	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	7.790	10.165	17.117	21.064	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	8.547	11.037	18.245	22.307	24.996	27.488	30.578	32.801	15
16	5.142	5.812	6.908	7.962	9.312	11.912	19.369	23.542	26.296	28.845	32.000	34.267	16
17	5.697	6.408	7.564	8.672	10.085	12.792	20.489	24.769	27.587	30.191	33.409	35.718	17
18	6.265	7.015	8.231	9.390	10.865	13.675	21.605	25.989	28.869	31.526	34.805	37.156	18
19	6.844	7.633	8.907	10.117	11.651	14.562	22.718	27.204	30.144	32.852	36.191	38.582	19
20	7.434	8.260	9.591	10.851	12.443	15.452	23.828	28.412	31.410	34.170	37.566	39.997	20
21	8.034	8.897	10.283	11.591	13.240	16.344	24.935	29.615	32.671	35.479	38.932	41.401	21
22	8.643	9.542	10.982	12.338	14.041	17.240	26.039	30.813	33.924	36.781	40.289	42.796	22
23	9.260	10.196	11.689	13.091	14.848	18.137	27.141	32.007	35.172	38.076	41.638	44.181	23
24	9.886	10.856	12.401	13.848	15.659	19.037	28.241	33.196	36.415	39.364	42.980	45.559	24
25	10.520	11.524	13.120	14.611	16.473	19.939	29.339	34.382	37.652	40.646	44.314	46.928	25
26	11.160	12.198	13.844	15.379	17.292	20.843	30.435	35.563	38.885	41.923	45.642	48.290	26
27	11.808	12.879	14.573	16.151	18.114	21.749	31.528	36.741	40.113	43.195	46.963	49.645	27
28	12.461	13.565	15.308	16.928	18.939	22.657	32.620	37.916	41.337	44.461	48.278	50.993	28
29	13.121	14.256	16.047	17.708	19.768	23.567	33.711	39.087	42.557	45.722	49.588	52.336	29
30	13.787	14.953	16.791	18.493	20.599	24.478	34.800	40.256	43.773	46.979	50.892	53.672	30
31	14.458	15.655	17.539	19.281	21.434	25.390	35.887	41.422	44.985	48.232	52.191	55.003	31
32	15.134	16.362	18.291	20.072	22.271	26.304	36.973	42.585	46.194	49.480	53.486	56.328	32
33	15.815	17.074	19.047	20.867	23.110	27.219	38.058	43.745	47.400	50.725	54.776	57.648	33
34	16.501	17.789	19.806	21.664	23.952	28.136	39.141	44.903	48.602	51.966	56.061	58.964	34
35	17.192	18.509	20.569	22.465	24.797	29.054	40.223	46.059	49.802	53.203	57.342	60.275	35
36	17.887	19.233	21.336	23.269	25.643	29.973	41.304	47.212	50.998	54.437	58.619	61.581	36
37	18.586	19.960	22.106	24.075	26.492	30.893	42.383	48.363	52.192	55.668	59.893	62.883	37
38	19.289	20.691	22.878	24.884	27.343	31.815	43.462	49.513	53.384	56.896	61.162	64.181	38
39	19.996	21.426	23.654	25.695	28.196	32.737	44.539	50.660	54.572	58.120	62.428	65.476	39
40	20.707	22.164	24.433	26.509	29.051	33.660	45.616	51.805	55.758	59.342	63.691	66.766	40
41	21.421	22.906	25.215	27.326	29.907	34.585	46.692	52.949	56.942	60.561	64.950	68.053	41
42	22.138	23.650	25.999	28.144	30.765	35.510	47.766	54.090	58.124	61.777	66.206	69.336	42
43	22.859	24.398	26.785	28.965	31.625	36.436	48.840	55.230	59.304	62.990	67.459	70.616	43
44	23.584	25.148	27.575	29.787	32.487	37.363	49.913	56.369	60.481	64.201	68.710	71.893	44
45	24.311	25.901	28.366	30.612	33.350	38.291	50.985	57.505	61.656	65.410	69.957	73.166	45

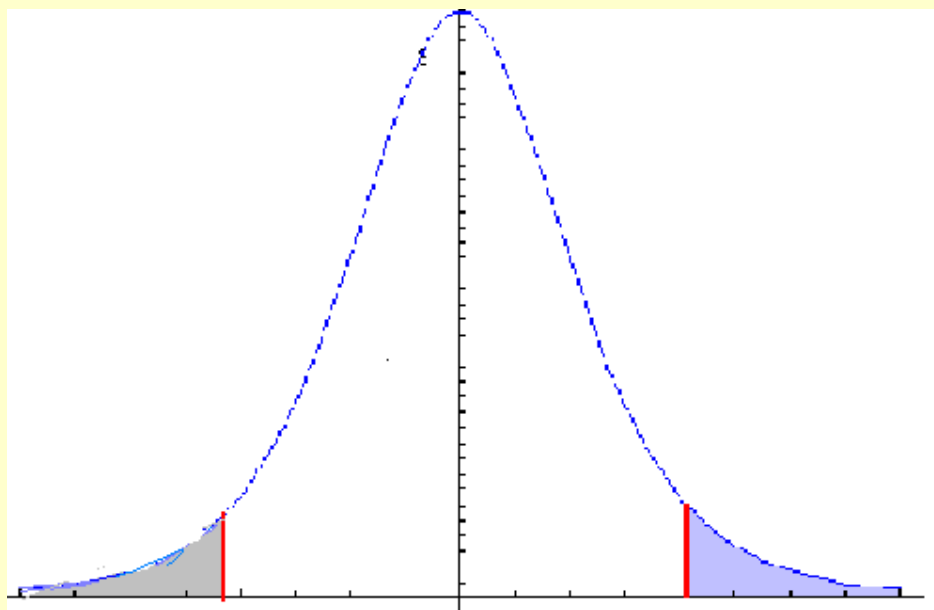
返回

t分布分位点

设 $t \sim t(n)$ ，对 $\alpha: 0 < \alpha < 1$ ，若存在 $t_\alpha(n)$ ，满足

$$P\{t > t_\alpha(n)\} = \alpha,$$

则称 $t_\alpha(n)$ 为 $t(n)$ 的上侧 α 分位点.



$$t_\alpha(n) = -t_{1-\alpha}(n)$$

$t_{1-\alpha}(n)$

$t_\alpha(n)$

t 分布分位点

n	0.25	0.1	0.05	0.025	0.01	0.005	n
1	1.0000	3.0777	6.3138	12.7062	31.8205	63.6567	1
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248	2
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409	3
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041	4
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321	5
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074	6
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995	7
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554	8
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498	9
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693	10
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058	11
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545	12
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123	13
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768	14
15	0.6912	1.3406	1.7531	2.1314	2.6025	2.9467	15
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208	16
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982	17
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784	18
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609	19
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453	20
21	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314	21
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188	22
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073	23
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969	24
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874	25
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787	26
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707	27
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633	28
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564	29
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500	30
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440	31
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385	32
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333	33
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284	34
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238	35
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195	36
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154	37
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116	38
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079	39
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045	40
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012	41
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981	42
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951	43
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923	44
45	0.6800	1.3006	1.6794	2.0141	2.4121	2.6896	45

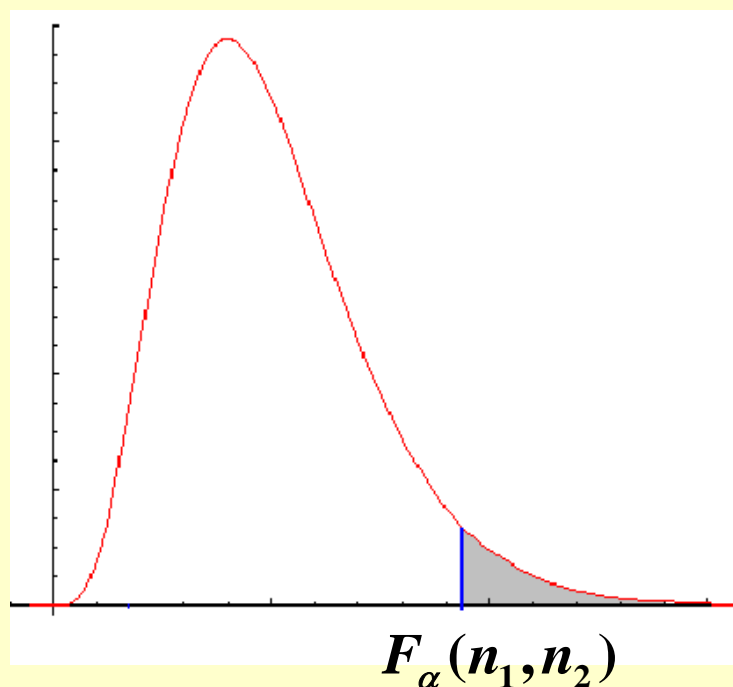
F分布分位点

设 $F \sim F(n_1, n_2)$ ，对于 $\alpha : 0 < \alpha < 1$ ，

若存在 $F_\alpha(n_1, n_2)$ ，满足

$$P\{F > F_\alpha(n_1, n_2)\} = \alpha,$$

则称 $F_\alpha(n_1, n_2)$ 为 $F(n_1, n_2)$ 的上侧 α 分位点



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

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

F分布分位点的性质的证明

F—分布的分位点

α		0.05				F分布分位点										n1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										</	
---	--	------	--	--	--	--------	--	--	--	--	--	--	--	--	--	----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----	--

6.3 正态总体的抽样分布定理

正态总体的抽样分布定理是置信区间与假设检验的理论基础

1. 若 $X_1, \dots, X_n \stackrel{i.i.d}{\sim} N(\mu, \sigma^2)$, 则

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

$$(2) t = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t(n-1)$$

$$(3) \chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \quad (4) \bar{X} \text{ 与 } S^2 \text{ 相互独立}$$

抽样分布定理的证明

2.若 $X_1, \dots, X_{n_1} \stackrel{i.i.d}{\sim} N(\mu_1, \sigma_1^2), Y_1, \dots, Y_{n_2} \stackrel{i.i.d}{\sim} N(\mu_2, \sigma_2^2)$,
且两样本独立. 则

$$(1) F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1);$$

(2) 进一步, 假定 $\sigma_1^2 = \sigma_2^2$, 就有,

$$t = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_w \sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2). \quad \text{其中}$$

$$S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \text{ 称为混合样本方差.}$$

练习题1

例1：设总体 $X \sim N(10, 3^2)$, X_1, \dots, X_n 是它的一个样本

$$Z = \sum_{i=1}^6 X_i \quad (1) \text{写出} Z \text{所服从的分布}; (2) \text{求} P(Z > 11).$$

$$Z \sim N(60, 54) \quad P(Z > 11) = 1 - \Phi\left(\frac{11 - 60}{\sqrt{54}}\right) = 1 - \Phi(-6.67) = 1$$

一般地，设随机变量 X_1, X_2, \dots, X_n 独立且 X_i 服从正态分布 $N(\mu_i, \sigma_i^2)$, $i=1, \dots, n$, 则

$$\sum_{i=1}^n a_i X_i + a_0 \sim N\left(\sum_{i=1}^n a_i \mu_i + a_0, \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

练习题2 设 X_1, \dots, X_n 是取自 $N(\mu, \sigma^2)$ 的样本,求样本方差 S^2 的期望与方差。

解：由抽样分布定理

知识点示意图

$$(3) \chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

再利用 χ^2 分布的性质

于是，得

$$\mathbf{E}\left(\frac{(n-1)S^2}{\sigma^2}\right) = (n-1)$$

$$\mathbf{E}(S^2) = \sigma^2$$

$$\mathbf{D}\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$$

$$\mathbf{D}(S^2) = \frac{2\sigma^4}{n-1}$$

练习题3 设 X_1, \dots, X_{10} 是取自 $N(0, 0.3^2)$ 的样本,

求 $P\{\sum_{i=1}^{10} X_i^2 > 1.44\}$

解: 由 χ^2 分布的定义, 得 $\sum_{i=1}^{10} \left(\frac{X_i}{0.3}\right)^2 \sim \chi^2(10)$

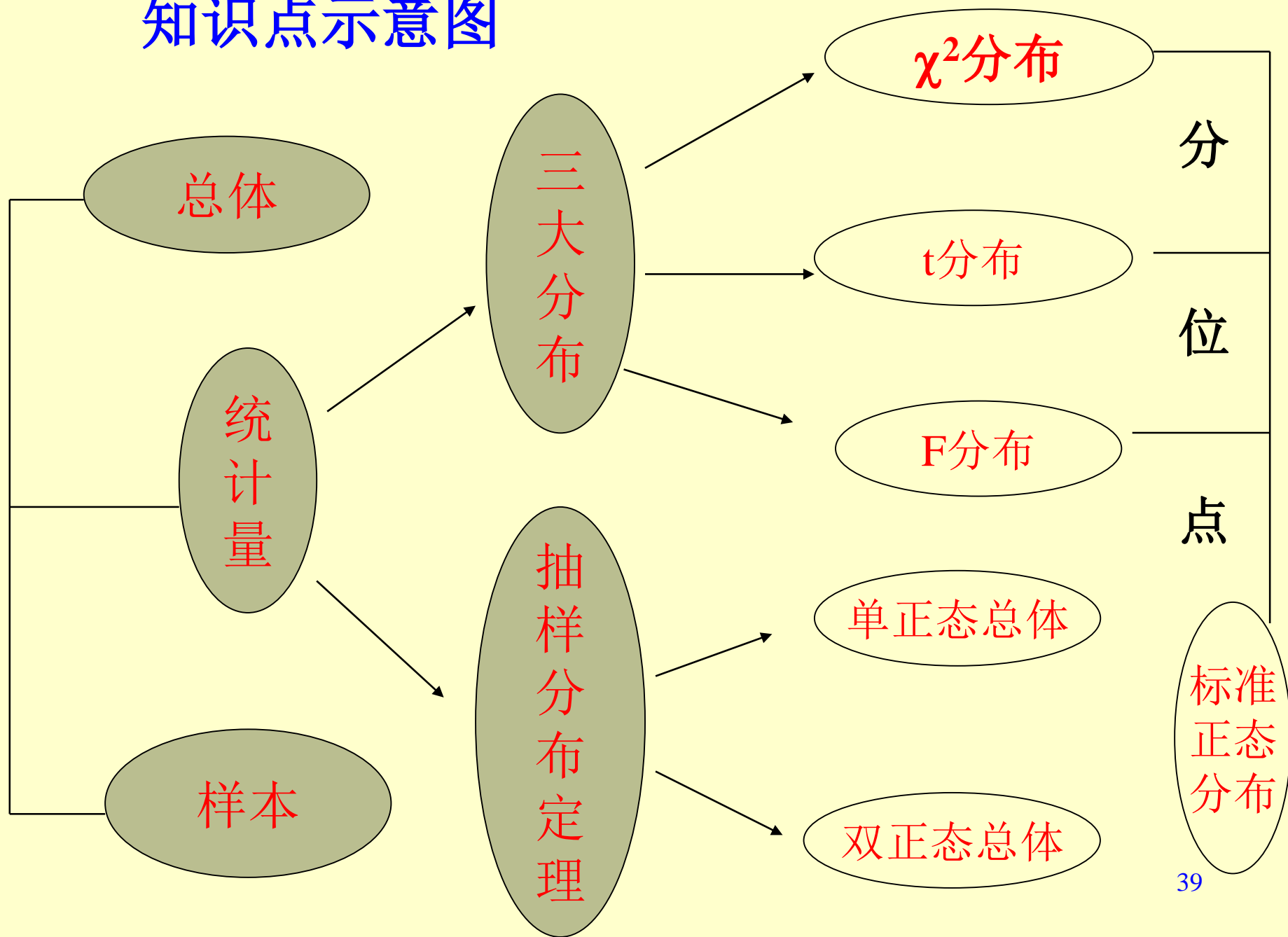
于是 $P\{\sum_{i=1}^{10} X_i^2 > 1.44\} = P\left\{\sum_{i=1}^{10} \left(\frac{X_i}{0.3}\right)^2 > 16\right\} = 0.1$

其中

$$\chi_{0.1}^2(10) = 15.987$$

χ^2 分布分位点

知识点示意图



本节要求

- 知识点1、理解总体、样本与样本联合分布等概念；
- 2、理解统计量的概念，熟悉常见的统计量；
- 3、理解三大分布的定义；
- 4、理解三大分布与标准正态分布的分位点；
- 5、熟悉正态总体的抽样分布定理。

考点

- 1、会求样本联合分布律与联合概率密度；
- 2、判定一个随机变量是否服从三大分布；
- 3、会求分位点；
- 4、会利用正态分布的抽样分布定理解题。

F分布分位点的性质的证明

注:
$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_{\alpha}(n_2, n_1)}$$

证明: 设 $F \sim F(n_1, n_2)$, 则 $\frac{1}{F} \sim F(n_2, n_1)$

$$P\{F > F_{1-\alpha}(n_1, n_2)\} = 1 - \alpha$$

$$P\{\frac{1}{F} > F_{\alpha}(n_2, n_1)\} = \alpha$$

$$P\{\frac{1}{F} < \frac{1}{F_{1-\alpha}(n_1, n_2)}\} = 1 - \alpha$$

$$P\{\frac{1}{F} \geq \frac{1}{F_{1-\alpha}(n_1, n_2)}\} = \alpha$$

得证!

[返回](#)

抽样分布定理的证明

(1) 证明 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ **n** 个独立的正态随机变量的线性组合,故服从正态分布。

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu$$

$$D(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^n D(X_i) = \frac{\sigma^2}{n}$$

$$\therefore \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\Rightarrow \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

(2)证明:

$$\because \mathbf{Z} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1) \quad \chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1);$$

且 \mathbf{Z} 与 χ^2 独立,根据 t 分布的构造

$$\frac{\mathbf{Z}}{\sqrt{\frac{\chi^2}{n-1}}} \sim t(n-1)$$

$$\mathbf{t} = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t(n-1)$$

返回

戈塞特(William Sealey Gosset, 1876.6.13—1937.10.16)

英国统计学家。出生于英国肯特郡坎特伯雷市，求学于曼彻斯特学院和牛津大学，主要学习化学和数学。1899年，戈塞特进入都柏林的A.吉尼斯父子酿酒厂，在那里可得到一大堆有关酿造方法、原料（大麦等）特性和成品质量之间的关系的统计数据。

戈塞特是小样本统计理论的开创者。戈塞特在酿酒公司工作中发现，供酿酒的每批麦子质量相差很大，而同一批麦子中能抽样供试验的麦子又很少，每批样本在不同的温度下做实验，其结果相差很大。这样一来，实际上取得的麦子样本，不可能是大样本，只能是小样本。可是，从小样本来分析数据是否可靠？误差有多大？小样本理论就在这样的背景下应运而生。1905年，戈塞特利用酒厂里大量的小样本数据写了第一篇论文《误差法则在酿酒过程中的应用》。

在此基础上，**1907**年戈塞特决心把小样本和大样本之间的差别搞清楚.为此，他试图把一个总体中的所有小样本的平均数的分布刻画出来.做法是，在一个大容器里放了一批纸牌，把它们弄乱，随机地抽若干张，对这一样本做实验记录观察值，然后再把纸牌弄乱，抽出几张，对相应的样本再做实验观察，记录观察值.大量地记录这种随机抽样的小样本观察值，就可借以获得小样本观察值的分布函数.若观察值是平均数，戈塞特把它叫做**t**分布函数。**1908**年，戈塞特以“学生（**Student**）”为笔名在《生物计量学》杂志发表了论文《平均数的规律误差》。这篇论文开创了小样本统计理论的先河，为研究样本分布理论奠定了重要基础.被统计学家誉为统计推断理论发展史上的里程碑.戈塞特这项成果，不仅不再依靠近似计算，而且能用所谓小样本来进行推断.但是，应该指出：戈塞特推导**t**分布的方法是极不完整的，后来费希尔利用**n**维几何方法给出了完整的证明。

William Sealey Gosset在20世纪前三十余年是统计界的活跃人物，他的成就限于《均》文.同年他发表了在总体相关系数为0时，二元正态样本相关系数的精确分布，这是关于正态样本相关系数的第1个小样本结.

William Sealey Gosset的一些思想，对他日后与奈曼合作建立其假设检验理论有着启发性的影响.他说（引自《耐曼¾现代统计学家》）：“我认为现在统计学界中有非常多的成就都应归功于**William Sealey Gosset**.....”

戈塞特在其论著中，引入了均值、方差、方差分析、样本等概率、统计的一些基本概念和术语. 1907-1937年间，戈塞特发表了22篇统计学论文，这些论文于1942年以《“学生”论文集》为书名重新发行.

[返回](#)