第6章　抽样分布理论

- 样本与统计量
- 抽样分布
- 样本均值和样本方差的分布

引　　言

随机变量及其所伴随的概率分布全面描述了随机现象的统计性规律。

概率论的许多问题中，随机变量的概率分布通常是已知的，或者假设是已知的，而一切计算与推理都是在这已知是基础上得出来的。

但实际中，情况往往并非如此，一个随机现象所服从的分布可能是完全不知道的，或者知道其分布概型，但是其中的某些参数是未知的。

例如：

某公路上行驶车辆的速度服从什么分布是未知的；

电视机的使用寿命服从什么分布是未知的；

产品是否合格服从两点分布，但参数——合格率p是未知的；

数理统计的任务则是以概率论为基础，根据试验所得到的数据，对研究对象的客观统计规律性做出合理的推断。

学习的基本内容

从第6章开始，我们学习数理统计的基础知识。数理统计的任务是以概率论为基础，根据试验所得到的数据，对研究对象的客观统计规律性作出合理的推断.数理统计所包含的内容十分丰富，本书介绍其中的参数估计、假设检验、方差分析、回归分析等内容.第6章主要介绍数理统计的一些基本术语、基本概念、重要的统计量及其分布，它们是后面各章的基础。

第1节

样本与统计量

5

样本与统计量

总体与样本

在数理统计中，把研究对象的全体称为总体（**population)**或母体，而把组成总体的每个单元称为个体。

抽样

要了解总体的分布规律，在统计分析工作中，往往是从总体中抽取一部分个体进行观测，这个过程称为抽样。

6

样本与统计量

子样

在抽取过程中，每抽取一个个体，就是对总体X进行一次随机试验，每次抽取的n个个体$X_1, X_2, \cdots, X_n$，称为总体X的一个容量为n的样本（sample）或子样；其中样本中所包含的个体数量称为样本容量。

子样 $X_1, X_2, \cdots, X_n$ 是n个随机变量，抽取之后的观测数据 $x_1, x_2, \cdots, x_n$ 称为样本值或子样观察值。

7

随机抽样方法的基本要求

代表性——即子样($X_1, X_2, \cdots, X_n$)的每个分量 X_i与总体X 具有相同的概率分布。

独立性——即 $X_1, X_2, \cdots, X_n$ 是相互独立的随机变量。

满足上述两点要求的子样称为简单随机子样.获得简单随机子样的抽样方法叫简单随机抽样.

从简单随机子样的含义可知，样本 $X_1, X_2, \cdots, X_n$ 是来自总体 X、与总体 X具有相同分布的随机变量.

8

简单随机抽样

例如：要通过随机抽样了解一批产品的次品率，如果每次抽取一件产品观测后放回原来的总量中，则这是一个简单随机抽样。

理论上要求采用有放回的抽取，而在实践中也仅当总体容量**N**较大时，才可以无放回的或同时的抽取。

没有特殊说明，今后所提到的样本都是简单随机子样。

9

统计量

定义　设（$X_1, X_2, \cdots, X_n$）为总体**X**的一个样本，$f(X_1, X_2, \cdots, X_n)$为不含任何未知参数的连续函数，则称 $f(X_1, X_2, \cdots, X_n)$为样本（$X_1, X_2, \cdots, X_n$）的一个统计量.

例如：设 (X_1, X_2, X_3) 是从正态总体 $N(\mu, \sigma^2)$ 中抽取的一个样本，其中 μ 为已知参数, σ为未知参数，

则　$X_1 + X_2 + 3\mu X_3$　　$X_1^2 + 3\mu X_2 X_3$　是统计量

$X_1 + \sigma X_2 + X_3^2$　　$X_1 X_2 X_3 + \sigma$　不是统计量

10

三个常用的统计量

设 $(X_1, X_2, \cdots, X_n)$ 是总体 X 的一个样本，

样本均值（**sample mean)**　反映总体**X**的取值平均

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

样本方差**(sample variance)** 反映总体**X**的离散程度

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

11

三个常用的统计量

设 $(X_1, X_2, \cdots, X_n)$ 是总体 X 的一个样本，

样本均方差或标准差

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

它们的观测值用相应的小写字母表示.反映总体**X**取值的平均，或反映总体**X**取值的离散程度。

12

三个常用的统计量

设 $(X_1, X_2, \cdots, X_n)$ 是总体 X 的一个样本，

极差（**Range)** 反映总体X的离散程度

$$\max_{1\le i\le n}\{X_i\} - \min_{1\le i\le n}\{X_i\}$$

变异系数（Coefficient of Variation） 比较不同量的离散程度

$$C_v = \frac{S}{|\overline{X}|}$$

13

经验分布函数

定义 设总体 X 的样本 $X_1, X_2, \cdots, X_n$ 的观测值为 $x_1, x_2, \cdots, x_n$，将其由小到大排列成 $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ 则称

$$F_n(x) = \begin{cases} 0 & x < x_{(1)} \\ \dfrac{k}{n} & x_{(k)} \le x < x_{(k+1)}, k = 1, 2, \cdots, n-1 \\ 1 & x_{(n)} \le x \end{cases}$$

为该样本的经验分布函数。

14

数据的简单处理

为了研究随机现象，首要的工作是收集原始数据.一般通过抽样调查或试验得到的数据往往是杂乱无章的，需要通过整理后才能显示出它们的分布状况。

数据的简单处理是以一种直观明了方式加工数据。

它包括两个方面——数据整理
计算样本特征数

15

数据的简单处理

数据整理：将数据分组 ⟶ 计算各组频数
作频率分布表 ⟶ 作频率直方图

计算样本特征数：

（1）反映集中趋势的特征数

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

中位数：数据按大小顺序排列后，位置居中的那个数或居中的两个数的平均数。

众数：样本中出现最多的那个数。

16

数据的简单处理

（2）反映分散程度的特征数：样本方差、样本标准差、极差、四分位差

极差——样本数据中最大值与最小值之差，$R = M - m$

四分位数——将样本数据依概率分为四等份的3个数据，依次称为第一、第二、第三四分位数。

第一四分位数Q_1：　$P\{X < Q_1\} = 0.25$

第二四分位数Q_2：　$P\{X < Q_2\} = 0.5$

第三四分位数Q_3：　$P\{X < Q_3\} = 0.75$

17

例1　为对某小麦杂交组合F_2代的株高X进行研究，抽取容量为100的样本，测试的原始数据记录如下(单位：厘米)，试根据以上数据，画出它的频率直方图，求随机变量X的分布状况。

87	88	111	91	73	70	92	98	105	94
99	91	98	110	98	97	90	83	92	88
86	94	102	99	89	104	94	94	92	96
87	94	92	86	102	88	75	90	90	80
84	91	82	94	99	102	91	96	94	94
85	88	80	83	81	69	95	80	97	92
96	109	91	80	80	94	102	80	86	91
90	83	84	91	87	95	76	90	91	77
103	89	88	85	95	92	104	92	95	83
86	81	86	91	89	83	96	86	75	92

18

第一．整理原始数据，加工为分组资料，作出频率分布表，画直方图，提取样本分布特征的信息.步骤如下：

1.找出数据中最小值m=69，最大值M=111，极差为 R=M－m=42

2.数据分组，根据样本容量n的大小，决定分组数k。

一般规律	30≤n≤40	5≤k≤6
	40≤n≤60	6≤k≤8
	60≤n≤100	8≤k≤10
	100≤n≤500	10≤k≤20

本例取k=9.

一般采取等距分组（也可以不等距分组），组距等于比极差除以组数略大的测量单位的整数倍。

本例测量单位为1厘米

$$\frac{M-m}{k}=\frac{111-69}{9}\approx 4.7$$

组距向上取整为5

3．确定组限和组中点值。

一般根据算式：

$$\text{各组中点值} \pm \frac{1}{2} \times \text{组距} = \text{组的上限或下限}$$

注意：组的上限与下限应比数据多一位小数。

当取**a=67.5**，**b=112.5**（**a**略小于**m**，**b**略大于**M**，且**a**和**b**都比数据多一位小数），分组如下：

[67.5,72.5)	**[72.5,77.5)**	**[77.5,82.5)**
[82.5,87.5)	**[87.5,92.5)**	**[92.5,97.5)**
[97.5,102.5)	**[102.5,107.5)**	**[107.5,112.5)**

组中值分别为：**70 75 80 85 90 95 100 105 110**

21

4．将数据分组，计算出各组频数，作频数、频率分布表

组序	区间范围	频数f_j	频率 $W_j=f_j/n$	累计频率 F_j
1	[67.5,72.5)	2	0.02	0.02
2	[72.5，77.5）	5	0.05	0.07
3	[77.5，82.5）	10	0.10	0.17
4	[82.5，87.5）	18	0.18	0.35
5	[87.5，92.5）	30	0.3	0.65
6	[92.5，97.5）	18	0.18	0.83
7	[97.5，102.5）	10	0.1	0.93
8	[102.5，107.5）	4	0.04	0.97
9	[107.5，112.5）	3	0.03	1.00

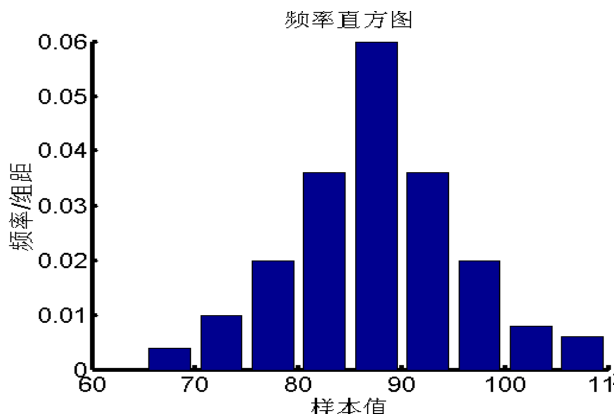
22

5.作出频率直方图

以样本值为横坐标，频率/组距为纵坐标；

以分组区间为底，以 $Y_j = \dfrac{W_j}{X_{j+1} - X_j} = \dfrac{W_j}{5}$ 为高

作频率直方图



23

从频率直方图可看到：靠近两个极端的数据出现比较少，而中间附近的数据比较多，即中间大两头小的分布趋势，——随机变量分布状况的最粗略的信息。

在频率直方图中， 每个矩形面积恰好等于样本值落在该矩形对应的分组区间内的频率，即

$$S_j = \frac{W_j}{X_{j+1} - X_j} \times \left(X_{j+1} - X_j\right) = W_j$$

频率直方图中的小矩形的面积近似地反映了样本数据落在某个区间内的可能性大小，故它可近似描述**X**的分布状况。

24

第二．计算样本特征数

1.反映集中趋势的特征数：样本均值、中位数、众数等

样本均值**MEAN**	中位数**MEDIAN**	众数
$\bar{X}=90.3$	91	$91, 94$

2.反映分散程度的特征数：样本方差、样本标准差、极差、四分位差等

样本方差	样本标准差	**Q1**	**Q3**	极差	四分位差
68.6970	**8.288**	**85.25**	**95**	**42**	**4.875**

上述差异特征统计量的值越小，表示离散程度越小.

第2节

抽样分布

统计量 $f(X_1, X_2, ..., X_n)$ 是样本 $X_1, X_2, ..., X_n$ 的不含任何未知参数的函数，它是一个随机变量

统计量的分布称为抽样分布。

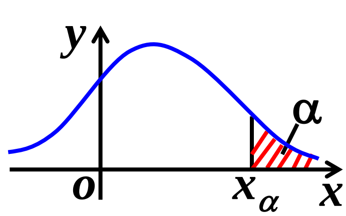
由于正态总体是最常见的总体，因此这里主要讨论正态总体下的抽样分布.

由于这些抽样分布的论证要用到较多的数学知识，故在本节中，我们主要给出有关结论，以供应用.

27

概率分布的分位数(分位点)

定义 对总体X和给定的α $(0<\alpha<1)$，若存在x_α，使$P\{X \geq x_\alpha\} = \alpha$，则称$x_\alpha$为$X$的分布的上侧$\alpha$分位数或上侧临界值. 如图.



$$P\{X \geq x_\alpha\} = \alpha$$

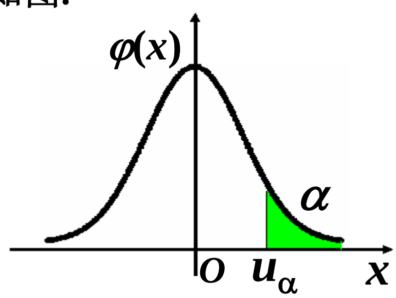
$$\int_{x_\alpha}^{+\infty} f(x)dx = \alpha$$

28

对标准正态分布变量U～$N(0, 1)$和给定α的，上侧α分位数是由：$P\{U\geq u_\alpha\} = \int_{u_\alpha}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \alpha$

即 $P\{U<u_\alpha\} = 1-\alpha$ $\Phi(u_\alpha) = 1-\alpha$ 确定的点u_α.

如图.



例如， $\alpha=0.05$，而

$P\{U\geq 1.645\} = 0.05$

所以， $u_{0.05} = 1.645$.

29

标准正态分布的分位数

在实际问题中， α常取0.05、0.01.

常用到下面几个临界值：

$u_{0.05} = 1.645$， $u_{0.01} = 2.326$

$u_{0.05/2} = 1.96$， $u_{0.01/2} = 2.575$

30

数理统计中常用的分布除正态分布外，还有三个非常有用的连续型分布，即

$\left.\begin{array}{l}\chi^2\text{分布}\\ t\text{分布}\\ F\text{分布}\end{array}\right\}$ 数理统计的三大分布(都是连续型).
它们都与正态分布有密切的联系.

在本章中特别要求掌握对正态分布、χ^2分布、t分布、F分布的一些结论的熟练运用. 它们是后面各章的基础.

31

一、χ^2分布

定义　设总体 $X \sim N(0,1)$，$(X_1, X_2, \ldots, X_n)$ 是 X 的一个样本，则称统计量 $Y = X_1^2 + X_2^2 + \cdots + X_n^2$ 服从自由度为n的 χ^2 分布，记作 $Y \sim \chi^2(n)$

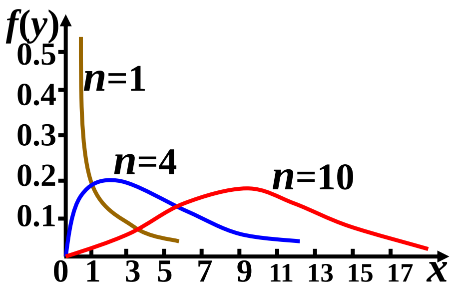
自由度是指独立随机变量的个数

$\chi^2(n)$ 分布的密度函数为

$$f(y)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}y^{n/2-1}e^{-y/2}, & y\ge 0\\ 0, & y<0\end{cases}\qquad \Gamma(n+1)=n!$$

32

$\chi^2(n)$ 分布密度函数的图形



其图形随自由度的不同而有所改变.

$$P\left\{\chi^2(n) \geq \chi_\alpha^2(n)\right\} = \alpha$$

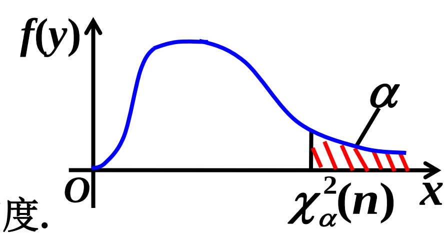
χ^2分布表(附表4(P_{325})).

χ^2分布的上α分位数

满足 $P\left\{\chi^2(n) \geq \chi_\alpha^2(n)\right\} = \int_{\chi_\alpha^2(n)}^{+\infty} f(y)dy = \alpha$

的数$\chi_\alpha^2(n)$为χ^2分布的上α分位数或上侧临界值，

其几何意义见图5-5所示.



其中$f(y)$是χ^2-分布的概率密度.

显然，在自由度n取定以后，$\chi_\alpha^2(n)$的值只与α有关.

例如，当n=21，α=0.05时，由附表4(P_{325})可查得，

$\chi_{0.05}^2(21)$ =32.67　即 $P\left\{\chi^2(21) \geq 32.671\right\} = 0.05.$

χ^2分布的数学期望与方差

设$\chi^2 \sim \chi^2(n)$，则$E(\chi^2)=n$，$Var(\chi^2)=2n$.

χ^2分布的可加性

设$\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$, 且$\chi_1^2$, χ_2^2相互独立,

则 $\chi_1^2+\chi_2^2 \sim \chi^2(n_1+n_2)$

35

二、t分布

定义5.4 设随机变量X～N(0,1)，Y～$\chi^2(n)$，且X与Y相互独立，则称统计量 $t=\dfrac{X}{\sqrt{\dfrac{Y}{n}}}$

服从自由度为n的t分布，记作t ～$t(n)$.

t分布的概率密度函数为

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

其图形如图6.2.3所示(P_{141})，其形状类似标准正态分布的概率密度的图形.

36

当n较大时，t分布近似于标准正态分布.

一般说来，当n>45时，t分布与标准正态分布$N(0, 1)$就非常接近.

t分布的数学期望与方差

设t～$t(n)$，则$E(t)=0$，$D(t)=\dfrac{n}{n-2}$. $(n>2)$

n=1时，t分布的数学期望不存在

n=1或2时，t分布的方差不存在

37

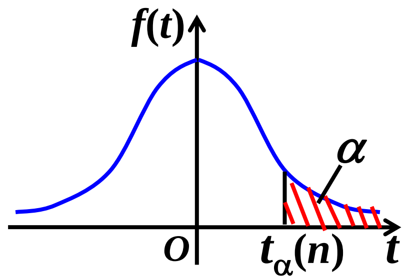
t分布的上α分位数

对于给定的$\alpha(0<\alpha<1)$，称满足条件

$$P\{t \ge t_\alpha(n)\} = \int_{t_\alpha(n)}^{+\infty} f(t)dt = \alpha$$

的数$t_\alpha(n)$为t分布的上α分位数或上侧临界值，

其几何意义如下图所示.



38

在附表3 (P_{324})中给出了t分布的临界值表.

例如，当n=15，α=0.05时，查t分布表得，

$t_{0.05}(15)=1.753$　　　$t_{0.05/2}(15)= 2.131$

其中$t_{0.05/2}(15)$由$P\{t(15)\geq t_{0.025}(15)\}=0.025$查得.

一般的t分布临界值表中，详列至n=30，当n>45就用标准正态分布$N(0, 1)$来近似.

即　$t_\alpha(n)\approx u_\alpha$,　　$n>45$.

39

三、F分布

定义5.5 设随机变量$X\sim\chi^2(n_1)$、$Y\sim\chi^2(n_2)$，且相互独立，则称随机变量$F=\dfrac{X/n_1}{Y/n_2}$服从第一自由度为n_1，第二自由度为n_2的F分布，

记作　$F\sim F(n_1, n_2)$.

概率密度函数 $$f(y)=\begin{cases} Ay^{\frac{n_1}{2}-1}\left(1+\dfrac{n_1}{n_2}y\right)^{-\frac{n_1+n_2}{2}}, & y\geq 0 \\ 0, & y<0 \end{cases}$$

其中$A=\dfrac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})}\left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}}$，其图形见图6.2.5.($P_{143}$)

40

性质：若$X \sim F(n_1, n_2)$，则$\dfrac{1}{X} \sim F(n_2, n_1)$.

F 分布的上α分位数

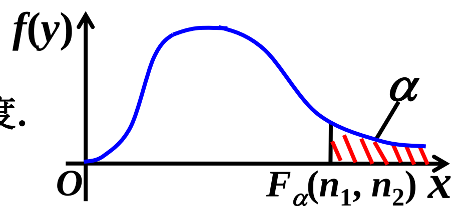
对于给定的$\alpha\,(0<\alpha<1)$，称满足条件

$$P\left\{F(n_1,n_2) \geq F_\alpha(n_1,n_2)\right\} = \int_{F_\alpha(n_1,n_2)}^{+\infty} f(y)dy = \alpha$$

的数$F_\alpha(n_1, n_2)$为F分布的上α分位数或上侧临界值，

其几何意义如图所示.

其中$f(y)$是F分布的概率密度.



F 分布的上α分位数

$F_\alpha(n_1, n_2)$的值可由F 分布表查得.

附表5($P_{326} \sim P_{330}$)分α=0.10、α=0.05、 α=0.025、 α=0.01 、α=0.005给出了F分布的上α分位数.

查表时应先找到相应的α值的表.

当时n_1=2, n_2=16时，有$F_{0.01}(2, 16)$= 6.226

在附表5中所列的α值都比较小，当α较大时，可用下面公式 $F_\alpha(n_1,n_2) = \dfrac{1}{F_{1-\alpha}(n_2,n_1)}$

例如，$F_{0.99}(16,2) = \dfrac{1}{F_{0.01}(2,16)} = \dfrac{1}{6.226} \approx 0.1606$

第3节

样本均值和样本方差的分布

一. 大样本情况下样本均值的分布

设总体的均值和方差分别为 μ,σ^2，$X_1,X_2,\dots,X_n$ 是来自总体的一个样本，当n充分大时

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\sim N\left(\mu,\frac{\sigma^2}{n}\right)$$

根据这个结论，不论总体是什么分布，只要样本容量足够大，就可以把样本均值看作正态分布！

因此关于样本均值的概率就可以用标准正态分布来计算。

例如 设某种零件的平均长度为**0.50cm**，标准差为**0.04cm**，从该种零件的总体中随机抽出**100**个样本，求这**100**个零件的平均长度小于**0.49cm**的概率.

解 由于$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$，所以 $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

因此所求概率为

$$P\{\overline{X} < 0.49\} = P\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < \frac{0.49-\mu}{\sigma/\sqrt{n}}\}$$

$$= P\{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < -2.5\} = \Phi(-2.5) = 1-\Phi(2.5) = 0.0062$$

45

二. 正态总体样本均值和样本方差的分布

结论1 设总体 $X \sim N\left(\mu, \sigma^2\right)$，$\left(X_1, X_2, \ldots, X_n\right)$ 是 X 的一个样本，则样本均值服从正态分布

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$U = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}} \sim N\left(0,1\right)$$

46

结论2 设$(X_1, X_2, \cdots, X_n)$取自正态总体$X \sim N(\mu, \sigma^2)$的样本，则

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$$

（P148第6题要用到此结论.）

证明 由已知，有

$X_i \sim N(\mu, \sigma^2)$且$X_1, X_2, \cdots, X_n$相互独立，

则 $\frac{X_i-\mu}{\sigma} \sim N(0,1)$且各$\frac{X_i-\mu}{\sigma}$相互独立，

由定义5.3得

$$\chi^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n).$$

47

结论3 设$(X_1, X_2, \cdots, X_n)$为来自正态总体$X \sim N(\mu, \sigma^2)$的样本，则

(1) 样本均值$\overline{X}$与样本方差S^2相互独立；

(2) $$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

从表面上看，$\sum_{i=1}^{n}(X_i-\overline{X})^2$是$n$个正态随机变量$X_i-\overline{X}$的平方和，但实际上它们不是独立的，它们之间有一种线性约束关系：

$$\sum_{i=1}^{n}(X_i-\overline{X}) = \sum_{i=1}^{n}X_i - n\overline{X} = 0$$

这表明，当这个n个正态随机变量中有$n-1$个取值给定时，剩下的一个的取值就跟着唯一确定了，故在这n项平方和中只有$n-1$项是独立的. 所以(5.8)式的自由度是$n-1$.

48

结论3 设$(X_1, X_2, \cdots, X_n)$为来自正态总体 $X \sim N(\mu, \sigma^2)$的样本，则

(1) 样本均值$\overline{X}$与样本方差S^2相互独立；

(2) $$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

与结论2比较：

结论2 设$(X_1, X_2, \cdots, X_n)$为取自正态总体 $X \sim N(\mu, \sigma^2)$的样本，则 $\dfrac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$

49

结论4 设$(X_1, X_2, \cdots, X_n)$为来自正态总体 $X \sim N(\mu, \sigma^2)$的样本，则统计量

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

证 由于$\overline{X}$与S^2相互独立，且

$$U = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1), \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

由定义5.4得

$$\frac{\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \frac{\overline{X} - \mu}{S/\sqrt{n}} = T \sim t(n-1)$$

50

结论5 设$(X_1, X_2, \cdots, X_{n_1})$和$(Y_1, Y_2, \cdots, Y_{n_2})$分别是来自正态总体$N(\mu_1, \sigma^2)$和$N(\mu_2, \sigma^2)$的样本，且它们相互独立，则统计量

$$T=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_n\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

其中$S_n=\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$,

S_1^2、S_2^2分别为两总体的样本方差.

51

结论6 设n_1, S_1^2为正态总体$N(\mu_1,\sigma_1^2)$的样本容量和样本方差；n_2, S_2^2为正态总体$N(\mu_2,\sigma_2^2)$的样本容量和样本方差；且两个样本相互独立，则统计量

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

证明 由已知条件知

$$\frac{(n_1-1)S_1^2}{\sigma_1^2}\sim\chi^2(n_1-1),\qquad \frac{(n_2-1)S_2^2}{\sigma_2^2}\sim\chi^2(n_2-1)$$

且相互独立，由F分布的定义有

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

52

U分布

正态总体样本均值的分布

设总体 $X \sim N\left(\mu, \sigma^2\right)$，$(X_1, X_2, ..., X_n)$是 X 的一个样本，则样本均值服从正态分布

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$U = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

53

χ^2 分布

定义　设总体 $X \sim N(0,1)$，$(X_1, X_2, ..., X_n)$是 X 的一个样本，则称统计量 $\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$ 服从自由度为n的 χ^2 分布，记作 $\chi^2 \sim \chi^2(n)$

n个相互独立的标准正态分布之平方和服从自由度为n的 χ^2 分布

54

t分布

定义5.4　设随机变量$X\sim N(0,1)$，$Y\sim\chi^2(n)$，且X与Y相互独立，则称统计量$T=\dfrac{X}{\sqrt{\dfrac{Y}{n}}}$

服从自由度为n的t分布或学生氏分布，记作$T\sim t(n)$.

$$\frac{N(0,1)}{\sqrt{\dfrac{\chi^2(n)}{n}}}\sim t(n)$$

t分布的密度函数的图形相似于标准正态分布的密度函数.当n较大时，t分布近似于标准正态分布.

55

F分布

定义5.5 设随机变量$X\sim\chi^2(n_1)$、$Y\sim\chi^2(n_2)$，且与相互独立，则称随机变量 $F=\dfrac{X/n_1}{Y/n_2}$

服从第一自由度为n_1，第二自由度为n_2的F分布，

记作 $F\sim F(n_1, n_2)$.

$$\frac{\chi^2(n_1)/n_1}{\chi^2(n_2)/n_2}\sim F(n_1,n_2)$$

56

例1　设总体$X\sim N(0,1)$，X_1，X_2，…，X_n为简单随机样本，试问下列统计量各服从什么分布？

$$(1)\ \frac{X_1-X_2}{\sqrt{X_3^2+X_4^2}};\quad (2)\ \frac{\sqrt{n-1}X_1}{\sqrt{\sum_{i=2}^{n}X_i^2}};\quad (3)\ \frac{(\frac{n}{3}-1)\sum_{i=1}^{3}X_i^2}{\sum_{i=4}^{n}X_i^2}.$$

解　(1) 因为$X_i\sim N(0,1)$，$i=1, 2, \cdots, n$. 所以

$$X_1-X_2\sim N(0, 2),\quad \frac{X_1-X_2}{\sqrt{2}}\sim N(0,1),$$

$$X_3^2+X_4^2\sim\chi^2(2),$$

故

$$\frac{X_1-X_2}{\sqrt{X_3^2+X_4^2}}=\frac{(X_1-X_2)/\sqrt{2}}{\sqrt{\frac{X_3^2+X_4^2}{2}}}\sim t(2).$$

57

例1　设总体$X\sim N(0,1)$，X_1，X_2，…，X_n为简单随机样本，试问下列统计量各服从什么分布？

$$(1)\ \frac{X_1-X_2}{\sqrt{X_3^2+X_4^2}};\quad (2)\ \frac{\sqrt{n-1}X_1}{\sqrt{\sum_{i=2}^{n}X_i^2}};\quad (3)\ \frac{(\frac{n}{3}-1)\sum_{i=1}^{3}X_i^2}{\sum_{i=4}^{n}X_i^2}.$$

解　(2) 因为$X_1\sim N(0,1)$，$\sum_{i=2}^{n}X_i^2\sim\chi^2(n-1)$

故

$$\frac{\sqrt{n-1}X_1}{\sqrt{\sum_{i=2}^{n}X_i^2}}=\frac{X_1}{\sqrt{\sum_{i=2}^{n}X_i^2\Big/(n-1)}}\sim t(n-1).$$

58

例1 设总体$X \sim N(0,1)$，X_1，X_2，…，X_n为简单随机样本，试问下列统计量各服从什么分布？

$$(1)\ \frac{X_1 - X_2}{\sqrt{X_3^2 + X_4^2}};\quad (2)\ \frac{\sqrt{n-1}X_1}{\sqrt{\sum_{i=2}^{n} X_i^2}};\quad (3)\ \frac{(\frac{n}{3}-1)\sum_{i=1}^{3} X_i^2}{\sum_{i=4}^{n} X_i^2}.$$

解 (3) 因为$\sum_{i=1}^{3} X_i^2 \sim \chi^2(3)$，$\sum_{i=4}^{n} X_i^2 \sim \chi^2(n-3)$，

所以

$$\frac{(\frac{n}{3}-1)\sum_{i=1}^{3} X_i^2}{\sum_{i=4}^{n} X_i^2} = \frac{\sum_{i=1}^{3} X_i^2 \Big/ 3}{\sum_{i=4}^{n} X_i^2 \Big/ (n-3)} \sim F(3,\ n-3).$$

59

例2 若$t \sim t(n)$，问t^2服从什么分布？

解 因为$t \sim t(n)$，可以认为

$$t = \frac{U}{\sqrt{V/n}},$$

其中$U \sim N(0,1)$，$V \sim \chi^2(n)$，

$$t^2 = \frac{U^2}{V/n},\qquad U^2 \sim \chi^2(1),$$

$$t^2 = \frac{U^2/1}{V/n} \sim F(1, n).$$

60

例3 设总体$X \sim N(\mu, 4^2)$， X_1，X_2，…，X_{10}是n=10简单随机样本， S^2为样本方差，计算$P\{S^2>30.078\}$.

解 因为n=10，$n-1$=9，$\sigma^2=4^2$，所以

$$\frac{9S^2}{4^2} \sim \chi^2(9).$$

因此 $P\{S^2 > 30.078\} = P\{\frac{9S^2}{4^2} > 16.918\} \underset{查表}{=} 0.05$

61

本章作业

P148 习题6

6

下节课交作业，请同学们做好作业，下次课不要忘了带过来

62