第六章 样本及抽样分布

- 随机样本
- 直方图和箱线图
- 抽样分布

§6.1 随机样本

§6.1.1 总体与个体

- 总体：试验全部可能的观察值；
- 个体：每一个可能的观察值（总体中的成员）；
- 总体的容量：总体中包含的个体数；
- 有限总体：容量有限的总体；
- 无限总体：容量无限的总体，通常将容量非常大的总体也按无限总体处理。

例1 现要研究某一个公司员工工资水平及其影响工资水平的因素. 这个公司的每个员工就是"个体", 而所有的员工构成一个"总体". 由于公司的员工总数是有限的, 因此, 是一个有限总体.

§6.1.2 样本

- 总体中的每一个个体是随机试验的一个观察值, 对于不同的个体来说有不同的取值, 这些取值可以构成一个分布, 因此可以看成一个随机变量.对总体的研究就是对一个随机变量**X**的研究， 统称为总体**X**.
- 数理统计主要任务是从总体中抽取一部分个体, 根据这部分个体的数据对总体分布给出推断. 被抽取的部分个体叫做总体的一个样本.

- 随机样本：从总体中随机地取n个个体，称为一个随机样本。

- 简单随机样本：满足以下两个条件的随机样本$(X_1, X_2, \cdots, X_n)$称为容量是n的简单随机样本。

  1. 每个X_i与X同分布；
  2. $X_1, X_2, \cdots, X_n$是相互独立的随机变量。

- 如何取得的样本才称是简单随机样本?

对于有限总体, 采用放回抽样就能得到简单随机样本.

但当总体容量很大的时候,放回抽样有时候很不方便, 因此在实际中当总体容量比较大时,通常将不放回抽样所得到的样本近似当作简单随机样本来处理.

对于无限总体, 一般采取不放回抽样.

§6.2 直方图和箱线图

为了研究总体分布的性质，人们通过随机试验得到许多观测值。一般来说，这些观测数据是杂乱无章的，为对它们进行统计分析，需将数据加以整理，如将数据排序、分类等。

本节将介绍通过图表整理数据的方法——直方图和箱线图。

§6.2.1 直方图

例1 下面给出了84个伊特拉斯坎（Etruscan）人男子的头颅的最大宽度（mm），现在来画这些数据的“频率直方图”.

141	148	132	138	154	142	150	146	155	158	150	140
147	148	144	150	149	145	149	158	143	141	144	144
126	140	144	142	141	140	145	135	147	146	141	136
140	146	142	137	148	154	137	139	143	140	131	143
141	149	148	135	148	152	143	144	141	143	147	146
150	132	142	142	143	153	149	146	149	138	142	149
142	137	134	144	146	147	140	142	140	137	152	145

步骤：

1. 找出最小值126，最大值158，现取区间[124.5,159.5]；

2. 将区间[124.5，159.5]等分为7个小区间，

小区间的长度记成 Δ, $\Delta = (159.5 - 124.5)/7 = 5$,

Δ称为组距；

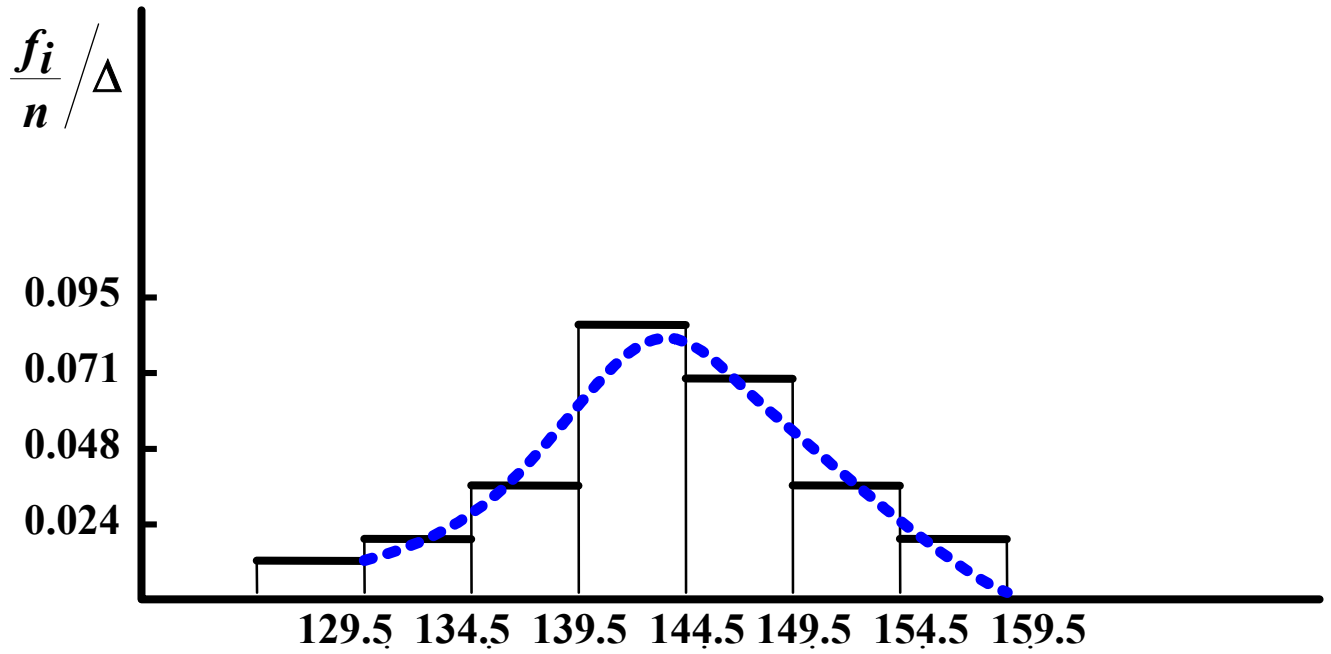
3. 小区间的端点称为组限, 数出落在每个小区间的数据的频数 f_i, 算出频率 f_i / n.

列表如下：

组 限	频 数	频 率	累计频率
124.5~129.5	1	0.0119	0.0119
129.5~134.5	4	0.0476	0.0595
134.5~139.5	10	0.1191	0.1786
139.5~144.5	33	0.3929	0.5715
144.5~149.5	24	0.2857	0.8572
149.5~154.5	9	0.1071	0.9643
154.5~159.5	3	0.0357	1.0000

现在自左向右依次在各 个小区间上作以 $\frac{f_i}{n}\Big/\Delta$ 为高的小矩形 ，这样的图形叫频率直方图.

频率直方图


$\frac{f_i}{n}\Big/\Delta$
0.095
0.071
0.048
0.024
129.5
134.5
139.5
144.5
149.5
154.5
159.5

直方图

1. 用矩形的宽度和高度来表示频数分布的图形，实际上是用矩形的面积来表示各组的频数分布。
2. 在直角坐标中，用横轴表示数据分组，纵轴表示频数或频率，各组与相应的频数就形成了一个矩形，即直方图。
3. 直方图下的总面积等于1。

§6.2.2 箱线图

样本p分位数：

设有容量为n的样本观察值 $x_1, x_2, \cdots, x_n$，样本p分位数$(0 < p < 1)$记为x_p，它具有以下的性质：

⑴ 至少有 np 个观察值小于或等于 x_p；

⑵ 至少有 $n(1-p)$ 个观察值大于或等于 x_p.

样本 p 分位数可按以下法则求得. 将 $x_1, x_2, \cdots, x_n$ 按从小到大的顺序排列成 $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$.

1^{o} 若np不是整数，则只有一个数据满足定义中的两点要求，这一数据位于大于np的最小整数处，即为位于 $[np]+1$ 处的数.

2^{o} 若np是整数， 就取位于 $[np]$和$[np]+1$处的平均值.

$$x_p = \begin{cases} x_{([np]+1)}, & \text{当}np\text{不是整数，} \\ \dfrac{1}{2}[x_{(np)} + x_{(np+1)}], & \text{当}np\text{是整数.} \end{cases}$$

当$p=0.5$时，0.5分位数$x_{0.5}$也记为Q_2或M，称为样本中位数，即有

$$x_{0.5}=\begin{cases} x_{([\frac{n}{2}]+1)}, & \text{当n是奇数，} \\ \frac{1}{2}[x_{(\frac{n}{2})}+x_{(\frac{n}{2}+1)}], & \text{当n是偶数.} \end{cases}$$

0.25分位数$x_{0.25}$称为第一四分位数，又记为Q_1；

0.75分位数$x_{0.75}$称为第三四分位数，又记为Q_3.

例2 设有一组容量为18的样本如下（已经排过序）

122 126 133 140 145 145 149 150 157

162 166 175 177 177 183 188 199 212

求样本分位数：$x_{0.2}$，$x_{0.25}$，$x_{0.5}$.

解 (1)因为 $np = 18 \times 0.2 = 3.6,$

$x_{0.2}$位于第$[3.6]+1=4$处，即有 $x_{0.2} = x_{(4)} = 140.$

(2)因为 $np = 18 \times 0.25 = 4.5,$

$x_{0.25}$位于第$[4.5]+1=5$处，即有 $x_{0.25} = 145.$

(3)因为 $np = 18 \times 0.5 = 9$, $x_{0.5}$是这组数中间两个数的平均值，即有 $x_{0.5} = \frac{1}{2}(157+162) = 159.5.$

箱线图：

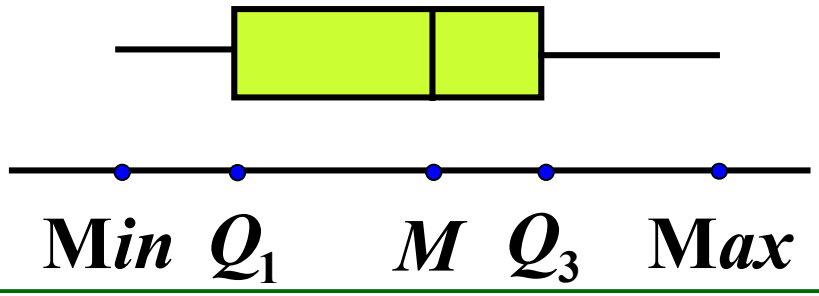
数据集的箱线图是由箱子和直线组成的图形，它是基于以下五个数的图形概括：最小值 **Min**，第一四分位数 Q_1，中位数M，第三四分位数 Q_3和最大值 **Max.**

箱线图的作法:

(1) 画一水平数轴，在轴上标上 Min，Q_1，M，Q_3，Max．在数轴上方画一个上、下侧平行于数轴的矩形箱子，箱子的左右两侧分别位于 Q_1，Q_3 的上方.

在M点的上方画一条垂直线段.线段位于箱子内部.

(2)自箱子左侧引一条水平线 Min; 在同一水平高度自箱子右侧引一条水平线直至最大值.



箱线图的特征：

1）中心位置：中位数所在的位置就是数据集的中心；

2）散布程度：全部数据落在[Min,Max]之内，在区间[Min,Q_1],[Q_1,M],[M, Q_3],[Q_3,Max]的数据个数约占1/4.区间较短时，表示落在该区间的点较集中，反之较为分散.

3）对称性：若中位数位于箱子的中间位置，则数据分布较为对称.若Min离M的距离较Max离M的距离大，则表示数据分布向左倾斜，反之表示数据向右倾斜，且能看出分布尾部的长短。

例3 以下是8个病人的血压（收缩压，**mmHg**）数据（已经过排序），试作出箱线图.

102　110　117　118　122　123　132　150

解 因为$np = 8 \times 0.25 = 2$, 故

$$Q_1 = \frac{1}{2}(110 + 117) = 113.5.$$

$\mathrm{Min} = 102,$

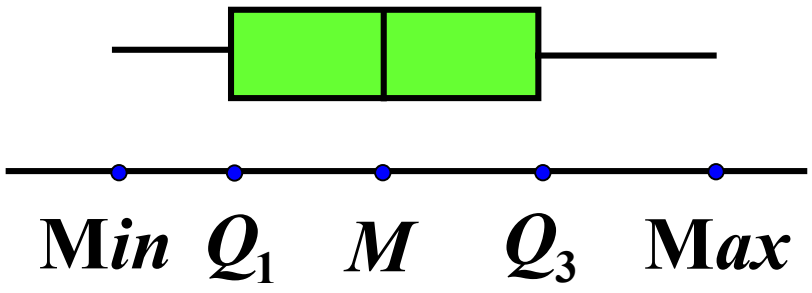
$\mathrm{Max} = 150,$

因为$np = 8 \times 0.5 = 4$, 故

$$x_{0.5} = Q_2 = \frac{1}{2}(118 + 122) = 120.$$

因为$np = 8 \times 0.75 = 6$, 故

$$x_{0.75} = Q_3 = \frac{1}{2}(123 + 132) = 127.5.$$



例4 下面分别给出了25个男子和25个女子的肺活量（以升计. 数据应经过排序）

女子组 **2.7 2.8 2.9 3.1 3.1 3.1 3.2 3.4 3.4**

3.4 3.4 3.4 3.5 3.5 3.5 3.6 3.7 3.7

3.7 3.8 3.8 4.0 4.1 4.2 4.2

男子组 **4.1 4.1 4.3 4.3 4.5 4.6 4.7 4.8 4.8**

5.1 5.3 5.3 5.3 5.4 5.4 5.5 5.6 5.7

5.8 5.8 6.0 6.1 6.3 6.7 6.7

试分别画出这两组数据的箱线图.

解 女子组 $\text{Min} = 2.7,\ \text{Max} = 4.2,\ M = 3.5,$

因 $np = 25 \times 0.25 = 6.25,\ Q_1 = 3.2.$

因 $np = 25 \times 0.75 = 18.75,\ Q_3 = 3.7.$

男子组 $\text{Min} = 4.1,\ \text{Max} = 6.7,\ M = 5.3,$

因 $np = 25 \times 0.25 = 6.25,\ Q_1 = 4.7.$

因 $np = 25 \times 0.75 = 18.75,\ Q_3 = 5.8.$

试作出箱线图?

疑似异常值：

在数据集中，某一个观察值不寻常地大于或小于该数据集中的其他数据,称为疑似异常值.

第一四分位数Q_1与第三四分数Q_3之间的距离：

$$Q_3 - Q_1 = IQR$$

称为四分位数间距.

若数据小于 $Q_1 - 1.5IQR$ 或大于$Q_3 + 1.5IQR$，则认为它是疑似异常值 .

修正箱线图的作法:

(1′) 同(1);

(2′) 计算$IQR = Q_3 - Q_1$, 若一个数据小于$Q_1 - 1.5IQR$或大于$Q_3 + 1.5IQR$, 则认为它是一个疑似异常值. 画出疑似异常值, 并以* 表示;

(3′) 自箱子左侧引一水平线段直至数据集中除去疑似异常值后的最小值, 又自箱子右侧引一水平线直至数据集中除去疑似异常值后的最大值.

例5 下面给出了某医院21个病人的住院时间（以天计），试画出修正箱线图（数据已经过排序）.

1 2 3 3 4 4 5 6 6 7 7 9 9

10 12 12 13 15 18 23 55

解 $\text{Min} = 1,\ \text{Max} = 55,\ M = 7,$

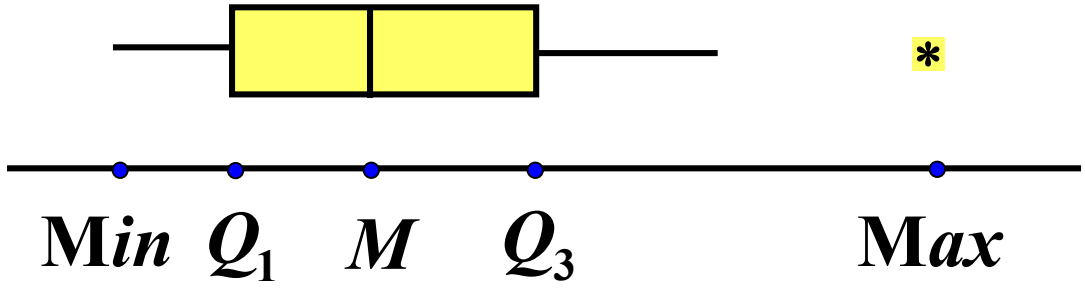
因 $21 \times 0.25 = 5.25,$ 得 $Q_1 = 4,$

又 $21 \times 0.75 = 15.75,$ 得 $Q_3 = 12,$

$IQR = Q_3 - Q_1 = 8,\quad Q_3 + 1.5IQR = 12 + 1.5 \times 8 = 24,$

$Q_1 - 1.5IQR = 4 - 12 = -8.$

观察值 $55 > 24,$ 故55 是疑似异常值，且仅此一个疑似异常值.



§ 6.3 抽样分布

§ 6.3.1 统计量

统计量：设$X_1,X_2,\ldots,X_n$是来自总体X的一个样本，$g(X_1,X_2,\ldots,X_n)$是$X_1,X_2,\ldots,X_n$的函数，若g中不含未知参数，则称$g(X_1,X_2,\ldots,X_n)$是一统计量.

因为$X_1,X_2,\ldots,X_n$都是随机变量，而统计量$g(X_1,X_2,\ldots,X_n)$是随机变量的函数，因此统计量也是随机变量.

常用统计量： 设（$X_1, X_2, \cdots, X_n$）为取自总体X的样本。常用的统计量如下：

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

2. 样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$，$S$为样本标准差

3. 样本矩　　k阶矩：$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \quad (k = 1, 2, \cdots)$

k阶中心矩：$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k \quad (k = 1, 2, \cdots)$

思考题：

设在总体$N\left(\mu,\sigma^2\right)$中抽取样本$\left(X_1,X_2,X_3\right)$,

其中μ已知，σ^2未知. 指出在

$(1)\ X_1+X_2+X_3$　$(2)\ X_2+2\mu$　$(3)\ \max\left(X_1,X_2,X_3\right)$

$$(4)\ \frac{1}{\sigma^2}\sum_{i=1}^{3}X_i^2 \quad (5)\ \left|X_3-X_1\right|$$

中哪些是统计量，哪些不是统计量，为什么？

答：只有(4)不是统计量。

§ 6.3.2 三大分布

- 统计量的分布称为抽样分布.
- 在数理统计中, 最重要的三个分布分别为:

χ^2 – 分布

t – 分布

F – 分布

χ^2分布

定义：设随机变量$X_1, X_2, \cdots X_n$相互独立，$X_i \sim N(0,1)\ (i=1,2,\cdots,n)$

则称 $\chi_n^2 = \sum_{i=1}^{n} X_i^2 \qquad (1)$

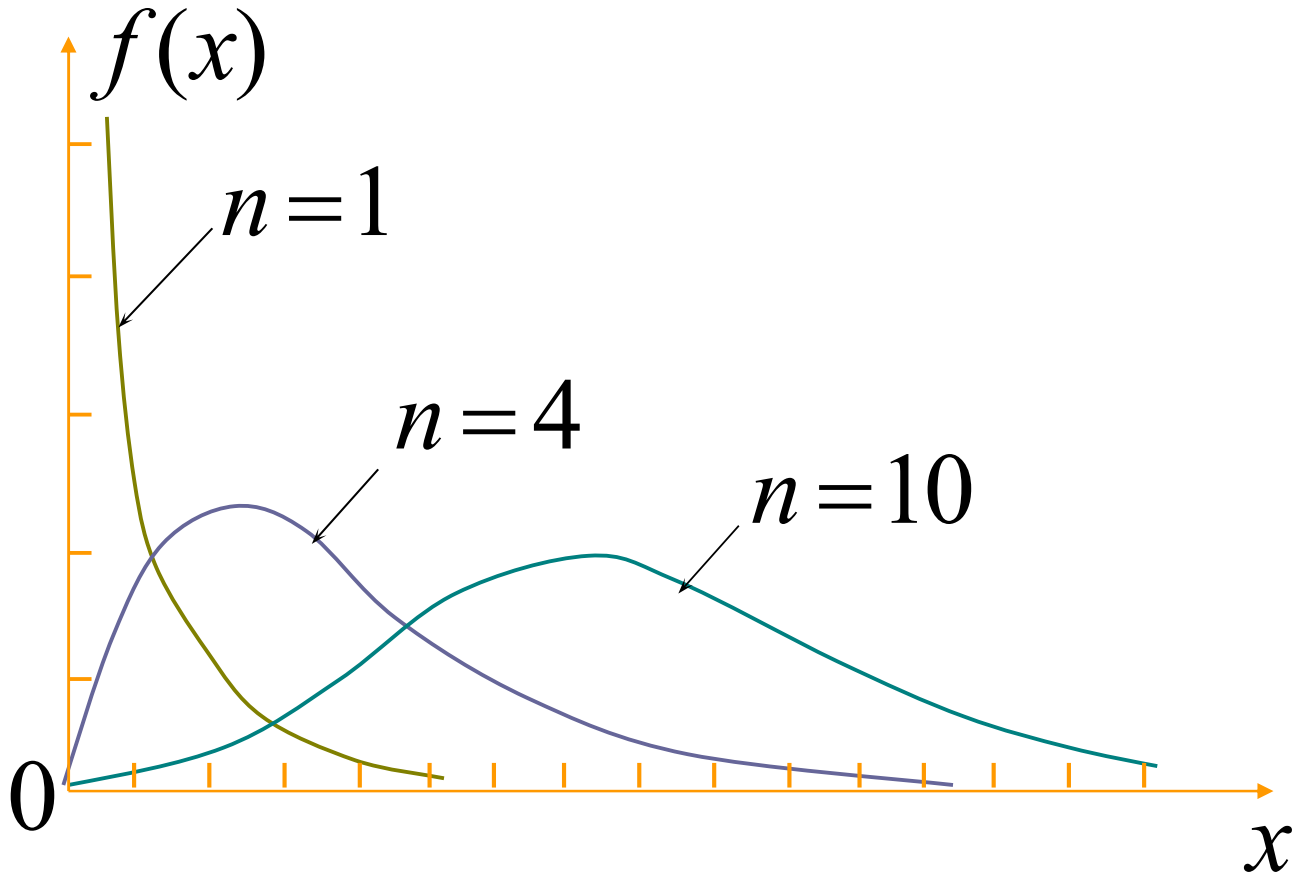
服从自由度为n的χ^2分布，记为$\chi^2 \sim \chi^2(n)$

自由度指(1)式右端包含的独立变量的个数.

$\chi^2(n)$分布的概率密度为：$f_n(y) = \begin{cases} \dfrac{1}{2\Gamma(n/2)}\left(\dfrac{y}{2}\right)^{\frac{n}{2}-1} e^{-\frac{y}{2}} & y>0 \\ 0 & y \le 0 \end{cases}$

其中$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$

χ^2分布的概率密度函数



推导

$\chi^2(n)$分布的概率密度为：$f_n(y)=\begin{cases}\dfrac{1}{2\Gamma(n/2)}\left(\dfrac{y}{2}\right)^{\frac{n}{2}-1}e^{-\frac{y}{2}} & y>0\\ 0 & y\le 0\end{cases}$

其中$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx$

提示：

1. 若随机变量X~N(0,1)， 则$Y=X^2$的概率密度为？

2. X_i^2~$\Gamma(\frac{1}{2},2)$,i=1,2,…,n

3. 由Γ分布可加性$\chi^2=\sum_{i=1}^{n}X_i^2$~$\Gamma(\frac{n}{2},2)$

χ^2分布的一些重要性质：

1. 设$\chi^2 \sim \chi^2(n)$,则有$E(\chi^2)=n, D(\chi^2)=2n$

2. 设$Y_1 \sim \chi^2(n_1), Y_2 \sim \chi^2(n_2)$,且$Y_1, Y_2$相互独立，则有$Y_1+Y_2 \sim \chi^2(n_1+n_2)$

性质2称为χ^2分布的可加性，可推广到有限个的情形：

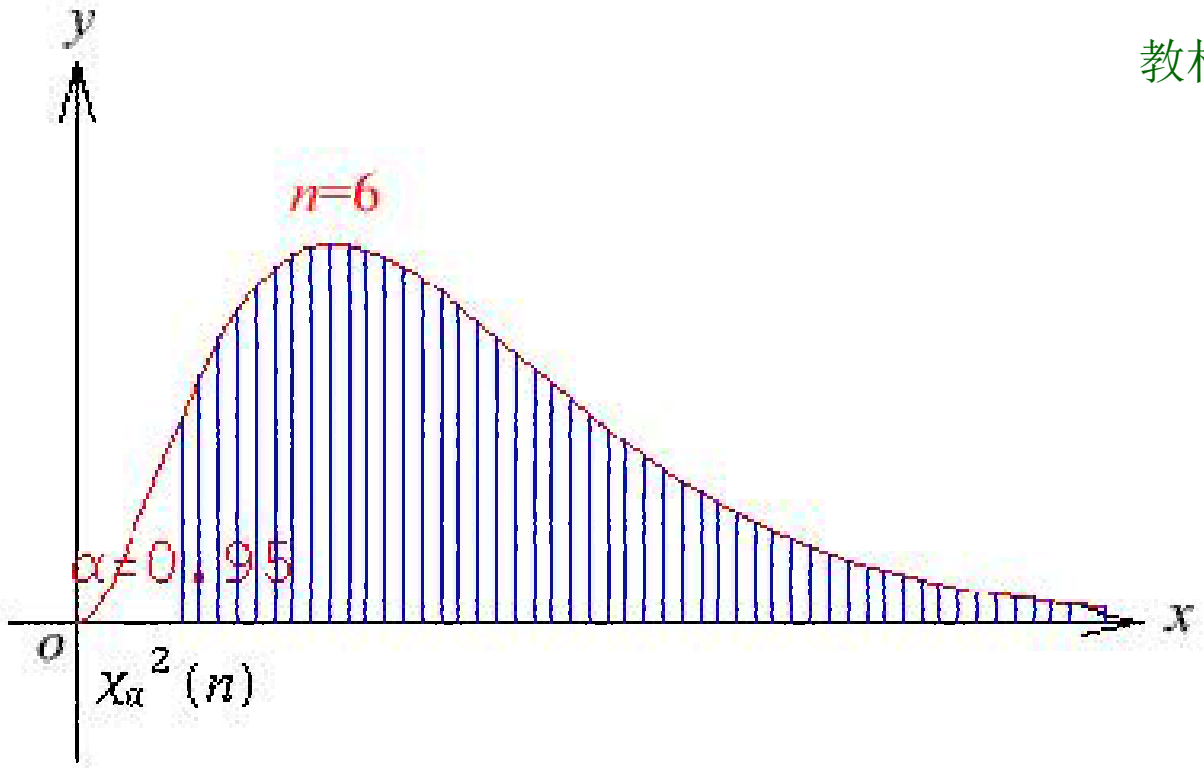
$Y_i \sim \chi^2(n_i), i=1,2,\cdots m$,并假设$Y_1, Y_2, \cdots Y_m$ 相互独立，

则$\sum_{i=1}^{m} Y_i \sim \chi^2\left(\sum_{i=1}^{m} n_i\right)$

上分位点

对给定的概率$\alpha, 0<\alpha<1$,称满足条件$\int_{\chi_\alpha^2(n)}^{\infty} f_n(y)dy=\alpha$的点$\chi_\alpha^2(n)$为$\chi^2(n)$分布的上$\alpha$分位数,上$\alpha$分位数$\chi_\alpha^2(n)$的值可查$\chi^2$分布表

教材p386附表5

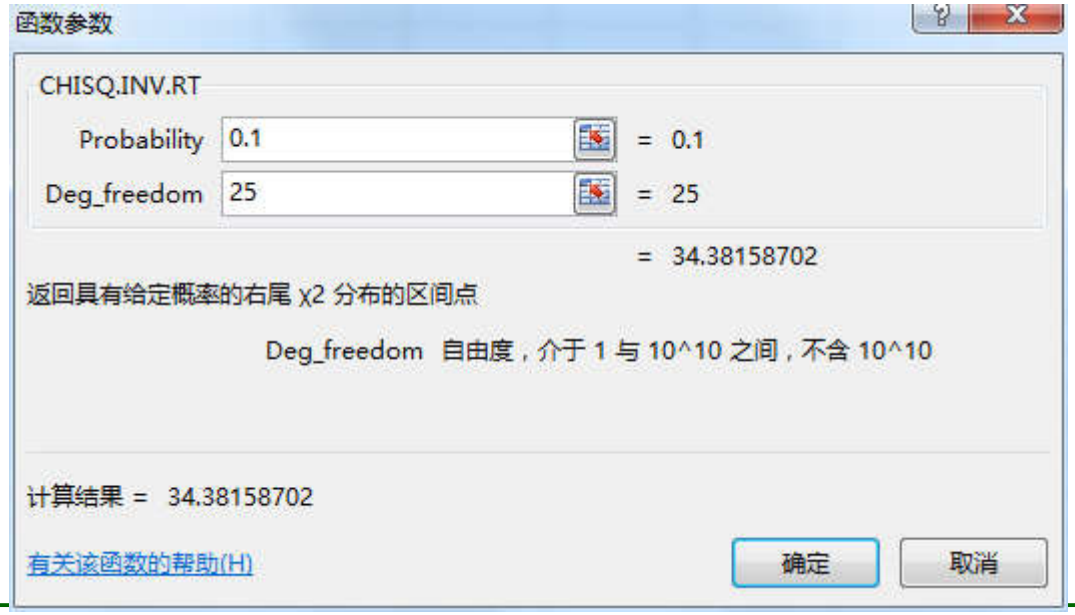


例1 通过Excel求$\chi^2_{0.1}(25)$

（1）具体如下: 在Excel表单的任一单元格输入“=”；

（2）在主菜单中点击“公式”，点击“插入函数”；

（3）在选择类别的下拉式菜单中选择“统计”，选择“CHISQ.INV.RT”点击“确定”在函数参数表单中输入Probability=0.1和 Deg_freedom=25



例2 设总体$X \sim N\left(\mu,\sigma^2\right),\mu,\sigma^2$ 已知。

$\left(X_1,X_2,\cdots,X_n\right)$是取自总体$X$的样本

求(1)统计量 $\chi^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ 的分布；

（2）设$n=5$, 若$a(X_1-X_2)^2+b(2X_3-X_4-X_5)^2 \sim \chi^2(k)$,

则a,b,k各为多少？

解：(1)作变换 $Y_i=\dfrac{X_i-\mu}{\sigma}\quad i=1,2,\cdots,n$

显然$Y_1,Y_2,\cdots,Y_n$相互独立，且$Y_i\sim N(0,1)\quad i=1,2,\cdots,n$

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

(2) $X_1-X_2\sim N(0,2\sigma^2),\dfrac{(X_1-X_2)^2}{2\sigma^2}\sim\chi^2(1)$

$a=\dfrac{1}{2\sigma^2},$

$b=\dfrac{1}{6\sigma^2},$

$k=2.$

$2X_3-X_4-X_5\sim N(0,6\sigma^2),\dfrac{(2X_3-X_4-X_5)^2}{6\sigma^2}\sim\chi^2(1)$

X_1-X_2与$2X_3-X_4-X_5$相互独立，

故$\dfrac{(X_1-X_2)^2}{2\sigma^2}+\dfrac{(2X_3-X_4-X_5)^2}{6\sigma^2}\sim\chi^2(2)$

$t-$分布

设 $X \sim N(0,1)$，$Y \sim \chi^2(n)$，并且假设 X, Y 相互独立，

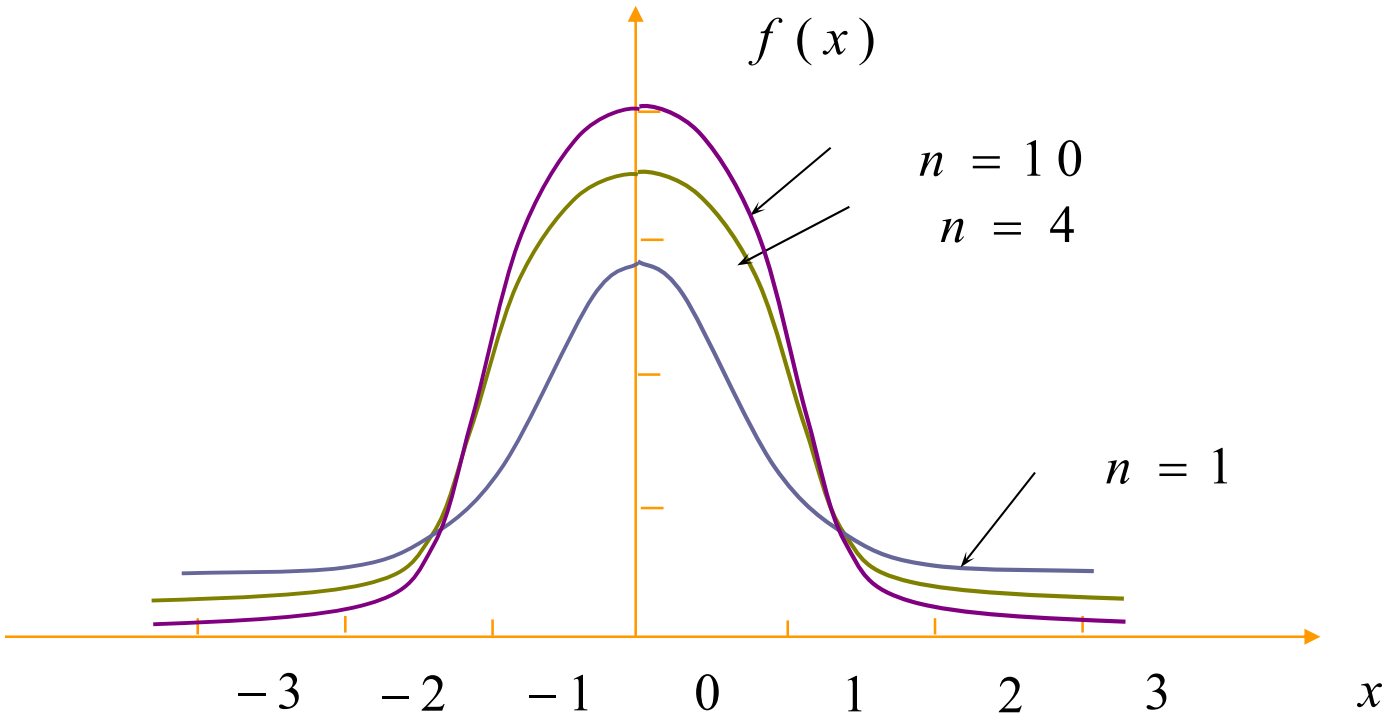
则称随机变量 $T = \dfrac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布。

记为 $T \sim t(n)$

$t(n)$ 分布的概率密度为：

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

t分布的概率密度函数

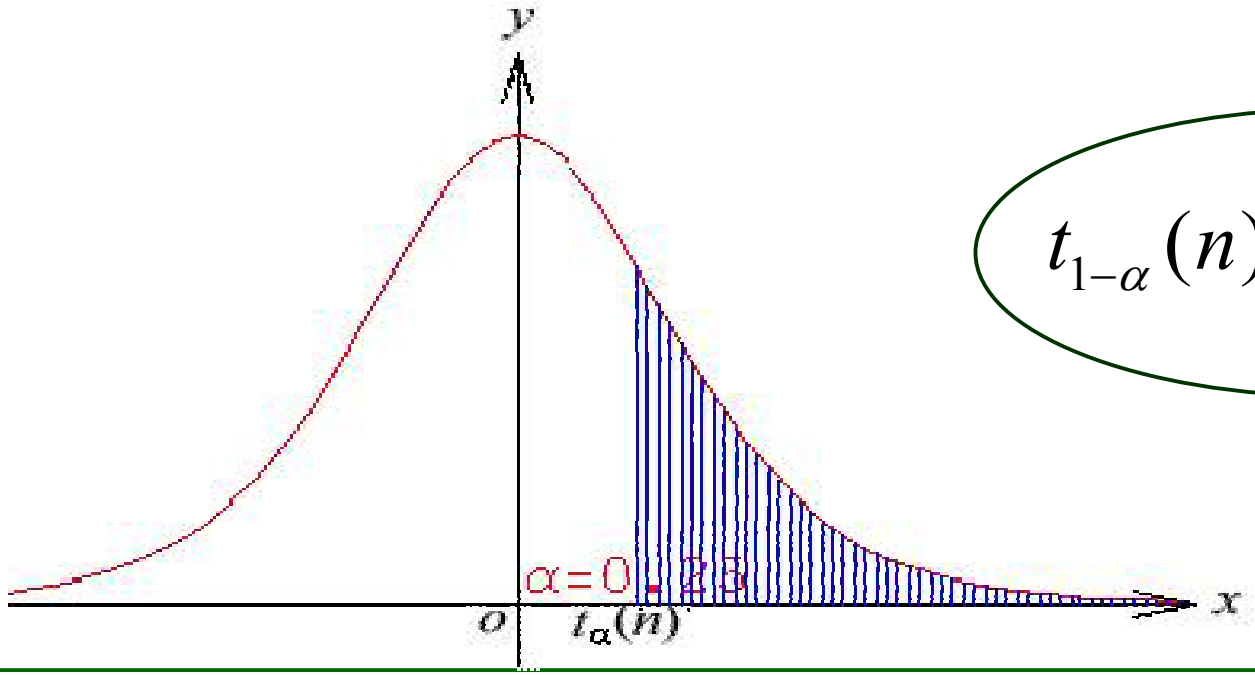


当n足够大时，t分布近似于N(0,1)分布

上分位点

对给定的$\alpha,\ 0<\alpha<1$,称满足条件$\int_{t_\alpha(n)}^{\infty} f(t,n)dt=\alpha$的点$t_\alpha(n)$为$t(n)$分布的上$\alpha$分位数。$t$分布的上$\alpha$分位数可查$t$分布表

教材p385附表4



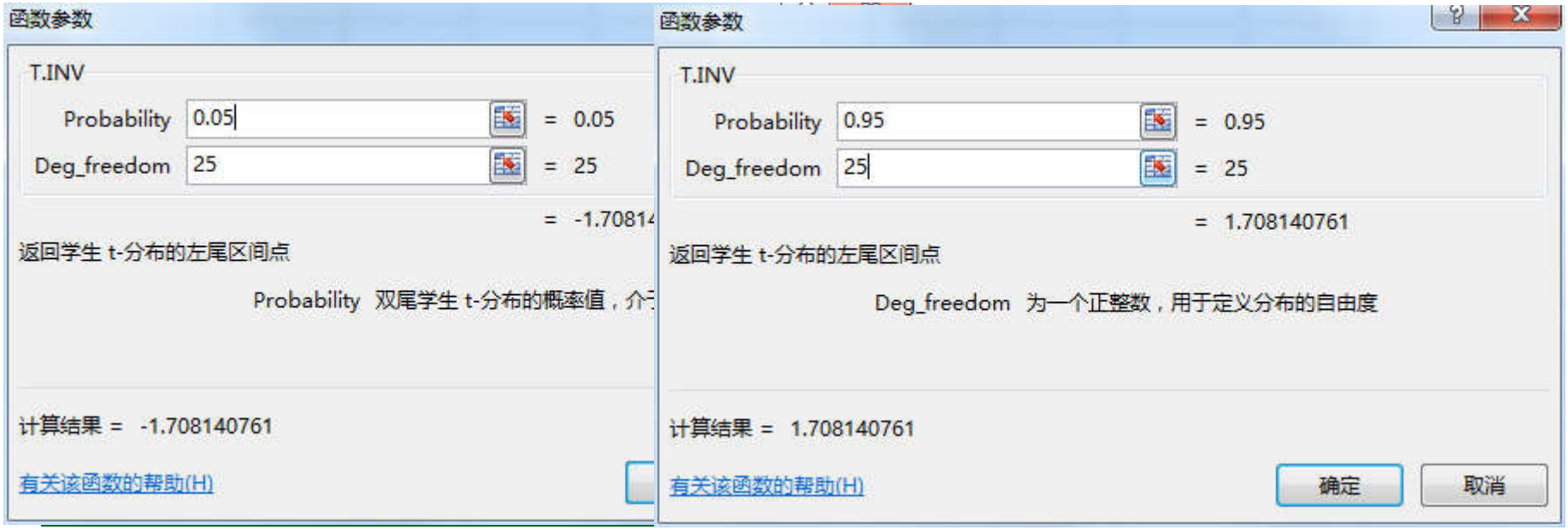
$$t_{1-\alpha}(n)=-t_\alpha(n)$$

例3 通过Excel求$t_{0.05}(25)$

（1）具体如下: 在Excel表单的任一单元格输入“=”；

（2）在主菜单中点击“公式”，点击“插入函数”；

（3）在选择类别的下拉式菜单中选择“统计”，选择“T.INV”点击“确定“在函数参数表单中输入Probability=0.95和Deg_freedom=25



F分布

定义：设$X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$，且$X, Y$独立，则

称随机变量$F = \dfrac{X/n_1}{Y/n_2}$服从自由度(n_1, n_2)的F分布，

记为$F \sim F(n_1, n_2)$

其中n_1称为第一自由度，n_2称为第二自由度

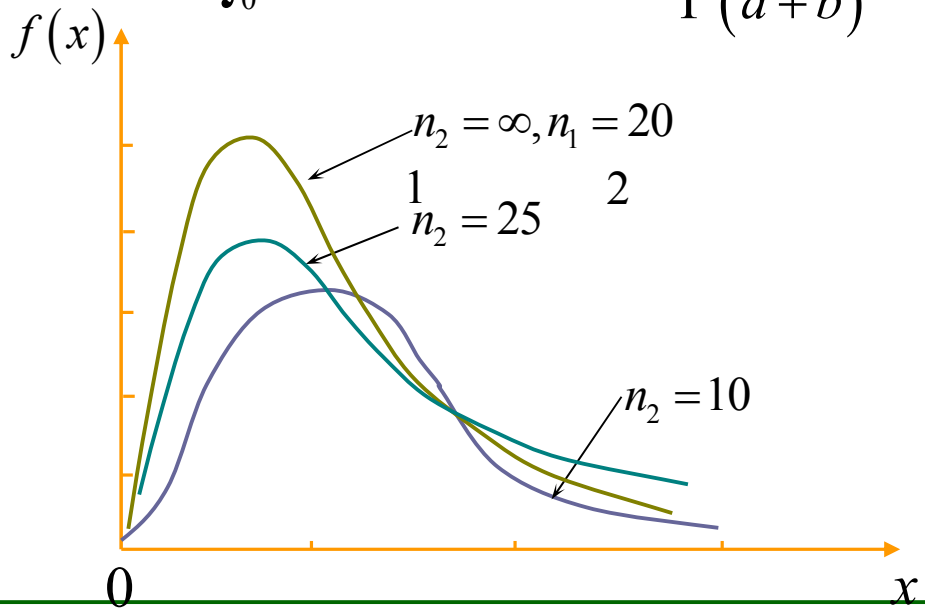
性质：$F \sim F(n_1, n_2)$,则$F^{-1} \sim F(n_2, n_1)$

F分布的概率密度函数

$F(n_1,n_2)$分布的概率密度为：

$$f(x;n_1,n_2)=\begin{cases}\dfrac{1}{B\left(n_1/2,\,n_2/2\right)}\,n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}x^{\frac{n_1}{2}-1}\left(n_2+n_1x\right)^{-\frac{n_1+n_2}{2}} & x>0\\ 0 & x\le 0\end{cases}$$

其中$B(a,b)=\int_0^1 x^{\alpha-1}(1-x)^{b-1}\,dx=\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$



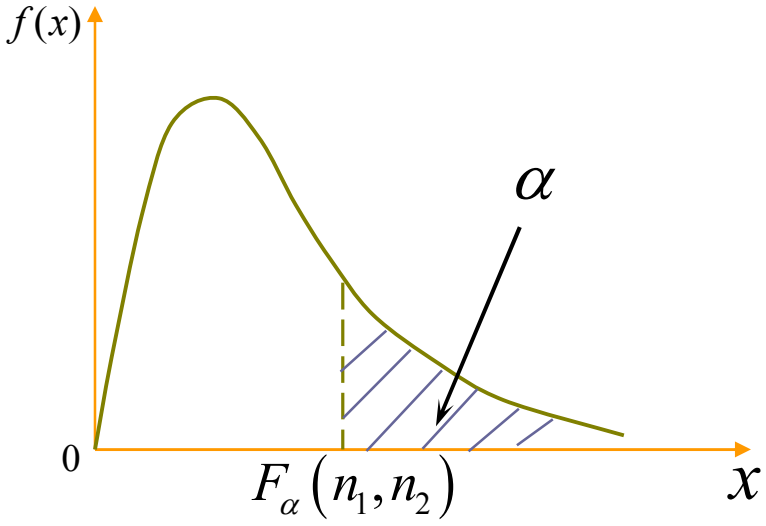
上分位点

对于给定的$\alpha,\ 0<\alpha<1$,称满足条件

$$\int_{F_\alpha(n_1,n_2)}^{\infty} f(x;n_1,n_2)dx=\alpha$$

的点$F_\alpha(n_1,n_2)$为$F(n_1,n_2)$分布的上α分位数。

$F_\alpha(n_1,n_2)$的值可查F分布表　教材p387附表6



$$F_{1-\alpha}(n_1,n_2)=[F_\alpha(n_2,n_1)]^{-1}$$

例4 通过Excel求$F_{0.1}(9,10)$

（1）具体如下: 在Excel表单的任一单元格输入“=”；

（2）在主菜单中点击“公式”，点击“插入函数”；

（3）在下拉式菜单中选择“统计” 选择 “F.INV.RT”点击“确定”在函数参数表单中输入 Probability=0.1和 Deg_freedom1=9, Deg_freedom2=10

§6.3.3 正态总体的样本均值与样本方差的分布

定理6.3.1 设$X_1,X_2,\ldots,X_n$是来自正态总体$N(\mu,\sigma^2)$的样本，$\overline{X}$是样本均值，则有

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

定理6.3.2 设$X_1, X_2, \ldots, X_n$是来自正态总体$N(\mu, \sigma^2)$的样本，$\overline{X}$，S^2是样本均值和样本方差，则有

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

(2) $\overline{X}$ 与 S^2 相互独立.

定理6.3.3 设$X_1,X_2,\ldots,X_n$是来自正态总体$N(\mu,\sigma^2)$的样本，$\overline{X}$，S^2是样本均值和样本方差，则有

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

定理6.3.4 设$X_1,X_2,\ldots,X_n$与$Y_1,Y_2,\ldots,Y_n$是来自正态总体$N(\mu_1,\sigma_1^2)$和$N(\mu_2,\sigma_2^2)$的样本，且这两个样本相互独立.设$\overline{X_1},\overline{X_2},S_1^2,S_2^2$是样本均值和样本方差，则有

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

$$(2)\quad \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$$

（3）当$\sigma_1^2=\sigma_2^2=\sigma^2$时，

$$\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_1-\mu_2\right)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t\left(n_1+n_2-2\right).$$

其中$S_w^2=\frac{\left(n_1-1\right)S_1^2+\left(n_2-1\right)S_2^2}{n_1+n_2-2}$, $S_w=\sqrt{S_w^2}$

例5 设总体$X \sim N(\mu, \sigma^2), (X_1, \cdots, X_4)$

与$(Y_1, \cdots, Y_9)$是取自总体X的两个独立样本，

$\bar{X}, S_1^2$和$\bar{Y}, S_2^2$分别为样本均值和样本方差；

求(1) $D(S_1^2 - S_2^2)$;

(2) $a\dfrac{\bar{X}-\bar{Y}}{S_1} \sim t(k)$, 则$a, k$各为多少？

(3) $b\sum_{i=1}^{4}(X_i - \mu)^2 \Big/ S_2^2 \sim F(n_1, n_2)$, 则$b, n_1, n_2$各为多少？

解：（1）一般地，由 $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1),$

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

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

所以，$D(S_1^2)=\dfrac{2\sigma^4}{3},D(S_2^2)=\dfrac{\sigma^4}{4}$

因此，$D(S_1^2-S_2^2)=D(S_1^2)+D(S_2^2)=\dfrac{11\sigma^4}{12}.$

(2) $\because \bar{X} \sim N(\mu, \frac{\sigma^2}{4}), \bar{Y} \sim N(\mu, \frac{\sigma^2}{9})$, 且$\bar{X}$与$\bar{Y}$相互独立,

$$\therefore \bar{X} - \bar{Y} \sim N(0, \frac{13\sigma^2}{36}),$$

又$\frac{3S_1^2}{\sigma^2} \sim \chi^2(3)$, 且$\bar{X} - \bar{Y}$与$S_1^2$相互独立,

所以, $$\frac{6}{\sqrt{13}} \frac{\bar{X} - \bar{Y}}{\sigma} \Big/ \sqrt{\frac{3S_1^2}{3\sigma^2}} = \frac{6\sqrt{13}}{13} \frac{\bar{X} - \bar{Y}}{S_1} \sim t(3)$$

$$\Rightarrow a = \frac{6\sqrt{13}}{13}, k = 3.$$

$$(3)\ \frac{1}{\sigma^2}\sum_{i=1}^{4}(X_i-\mu)^2 \sim \chi^2(4),\quad \frac{8S_2^2}{\sigma^2} \sim \chi^2(8),$$

且$\sum_{i=1}^{4}(X_i-\mu)^2$与$S_2^2$独立，

$$\Rightarrow \frac{1}{4\sigma^2}\sum_{i=1}^{4}(X_i-\mu)^2 \Big/ \frac{8S_2^2}{8\sigma^2} = \frac{1}{4}\sum_{i=1}^{4}(X_i-\mu)^2 \Big/ S_2^2 \sim F(4,8),$$

$$\Rightarrow b=\frac{1}{4},(n_1,n_2)=(4,8).$$

练习1

下面列出了30个美国NBA球员的体重（以磅计，1磅=0.454kg）数据。这些数据是从美国NBA球队1990-1991赛季的花名册中抽样得到的。

225 232 232 245 235 245 270 225 240 240
217 195 225 185 200 220 200 210 271 240
220 230 215 252 225 220 206 185 227 236

（1）画出这些数据的频率直方图（提示：最大和最小观察值分别为271和185，区间[184.5,271.5]包含所有数据，将整个区间分为5等份，为计算方便，将区间调整为[179.5,279.5]）。

（2）作出这些数据的箱线图。

练习2

1.验证：(1)$P\{X \le F_{1-\alpha}\} = \alpha$；(2) $P\{F_{1-\alpha/2} < X \le F_{\alpha/2}\} = 1-\alpha$；其中$F_\alpha$为随机变量X的水平为$\alpha$的上侧分位数.

2.查表求标准正态分布的下列上侧分位数：

$$u_{0.4},\ u_{0.2},\ u_{0.1},\ u_{0.05}$$

3.设$(X_1,\ldots,X_n,X_{n+1})$是取自正态总体$N(\mu,\sigma^2)$的样本，求$Y = X_{n+1} - \frac{1}{n+1}\sum_{i=1}^{n+1} X_i$的分布.

4.设总体$X\sim N(0,2^2)$，$(X_1,\ldots,X_{14},X_{15})$是取自总体X的样本，求$Y = \frac{1}{2}\frac{x_1^2+\cdots+x_{10}^2}{x_{11}^2+\cdots+x_{15}^2}$的分布.

5.设总体X和总体Y相互独立且服从N(0,9)，$(X_1,\ldots,X_9)$和$(Y_1,\ldots,Y_9)$是分别取自X和Y的样本，求$U = \frac{X_1+\cdots+X_9}{\sqrt{Y_1^2+\cdots+Y_9^2}}$的分布.