第 六 章 数理统计的基本概念

第二节 抽样分布

1、四个常见的分布

2、抽样分布定理

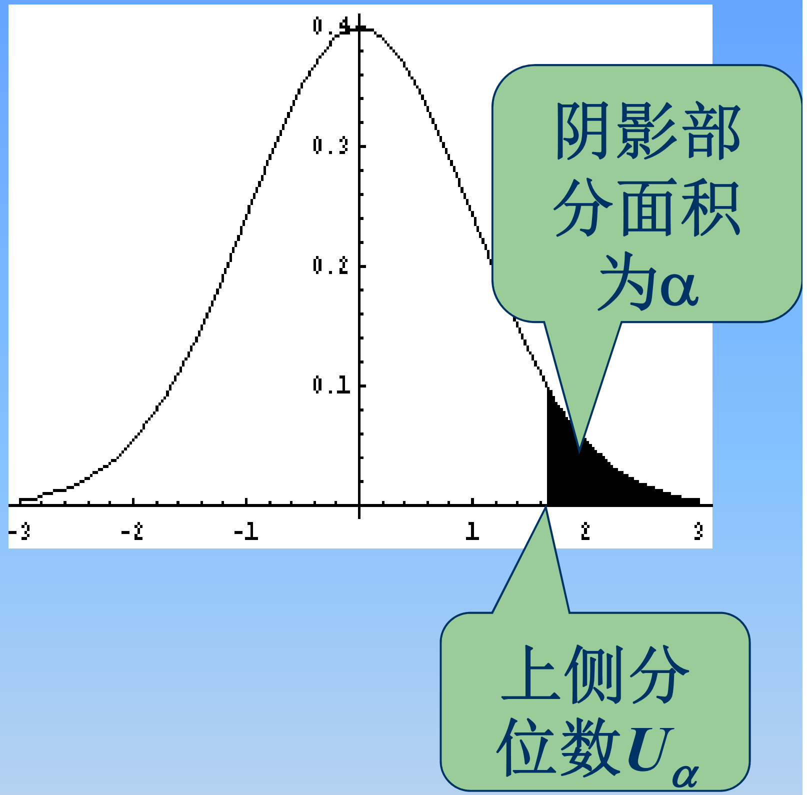
3、小结、思考

§2　常用的分布

一、四个常用分布

1. 标准正态分布　$X \sim N(0,1)$

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},x\in R$$



上侧分位数U_α(0< α<1）：

$$P\{X>u_\alpha\}=\int_{u_\alpha}^{+\infty}f(x)dx=\alpha$$

对于正态分布有：　$\Phi(u_\alpha)=1-\alpha$

2. χ^2 (卡方)分布

设随机变量X的概率密度函数为

$$f(x)=\begin{cases}\dfrac{1}{2\Gamma(\frac{n}{2})}(\dfrac{x}{2})^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x>0\\ 0, & x\le 0\end{cases},$$

其中, $\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx,\alpha>0$ 为Gama函数.

称随机变量X服从自由度为n 的χ^2分布,

记为: $\chi^2\sim\chi^2(n)$

Γ函数的主要性质: $\Gamma(1)=1,\Gamma(\frac{1}{2})=\sqrt{\pi},\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$

例：若$X\sim N(0,1)$， 则 $Y=X^2$ 的概率密度为：

$$f_Y(y)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}y^{-\frac{1}{2}}e^{-\frac{y}{2}}, & y>0;\\ 0, & y\le 0.\end{cases}$$

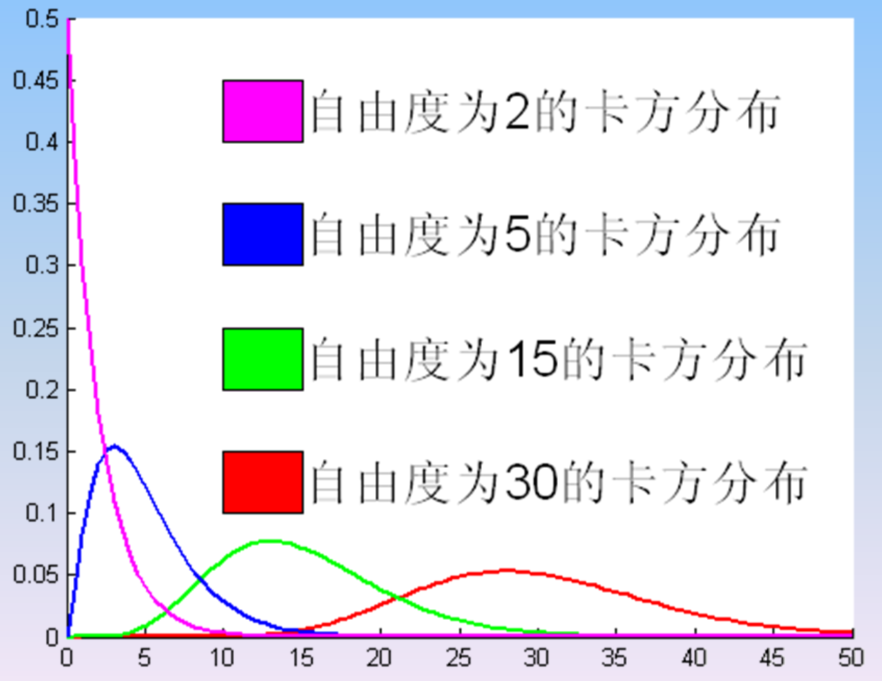
$$=\begin{cases}\dfrac{1}{2\Gamma(\frac{1}{2})}\left(\dfrac{y}{2}\right)^{\frac{1}{2}-1}e^{-\frac{y}{2}}, & y>0;\\ 0, & y\le 0.\end{cases}$$

即 $Y=X^2$ 服从 χ^2（1）分布。

定理6.2.1 设 X_1，X_2，…，X_n相互独立且都服从标准正态分布，则

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

即随机变量 χ^2 服从自由度为 n 的卡方分布。



例 统计量的分布(之一)

设 X_1，X_2，…，Xn 是来自正态总体 $X \sim N(\mu,\sigma^2)$ 的容量为 n 的样本，求下列统计量的概率分布：

1. $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 2. $Y = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2$

解：

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

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

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

证明: $$\chi^2 = \sum_{i=1}^{n} X_i^2$$

且 X_1，X_2，…，X_n 相互独立，$X_i \sim N(0,1)$

$$E(\chi^2) = \sum_{i=1}^{n} E(X_i^2) = \sum_{i=1}^{n} D(X_i) = n$$

$$D(\chi^2) = \sum_{i=1}^{n} D(X_i^2)$$

$$= \sum_{i=1}^{n} \{E(X_i^4) - [E(X_i^2)]^2\} = 2n.$$

性质2.(可加性) 设Y_1、Y_2相互独立且$Y_1 \sim \chi^2(n_1)$，

$Y_2 \sim \chi^2(n_2)$，则 $Y_1+Y_2 \sim \chi^2(n_1+n_2)$

性质3.(大样本分位数$\chi_\alpha^2(n)$) 当 n 足够大(如$n > 45$)时，有

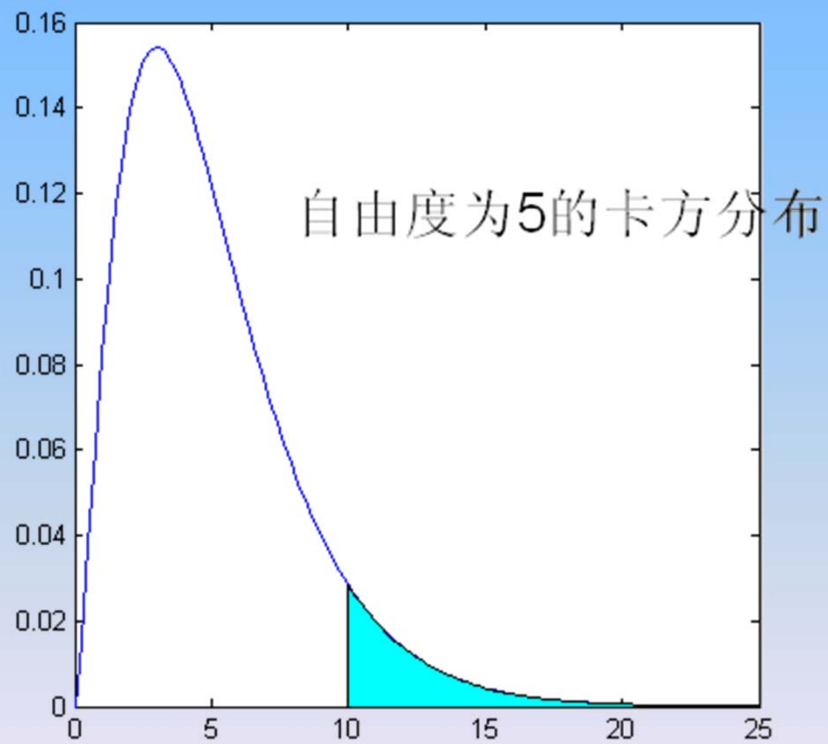
$$\chi_\alpha^2(n) \approx n + u_\alpha \sqrt{2n} \quad ,$$

其中u_α满足$\Phi(u_\alpha) = 1 - \alpha$

独立同分布中心极限定理

$\chi^2(n)$ 的上侧分位数($0< \alpha<1$):

$$P\{\chi^2 > \chi_\alpha^2(n)\} = \int_{\chi_\alpha^2(n)}^{+\infty} f_{\chi^2}(x)dx = \alpha$$



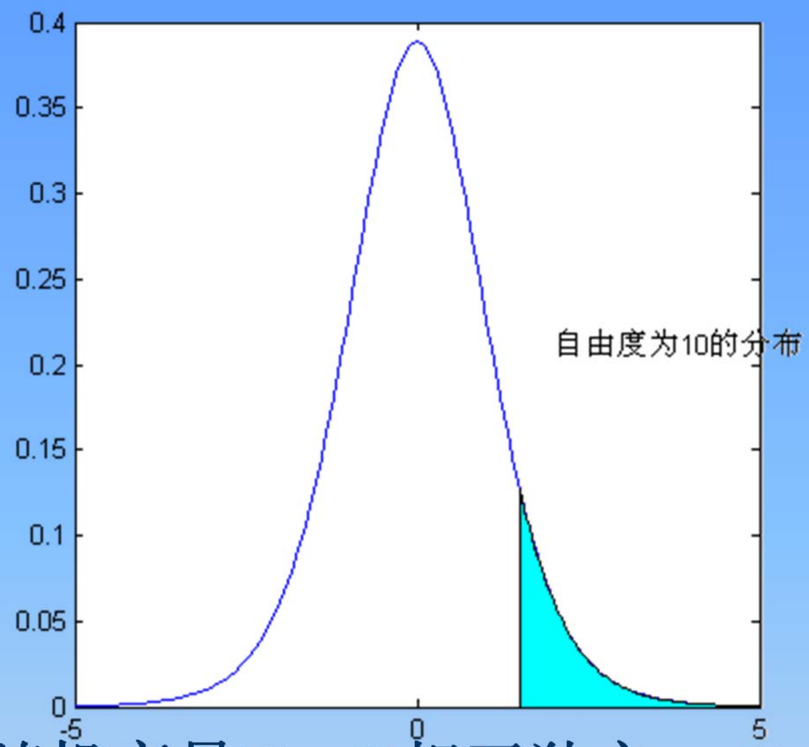
3. 自由度为 n 的 t 分布　　$T\sim t(n)$

(又称学生氏分布----第一个研究者以*Student*作笔名发表文章)

$$f(x)=\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}},\quad x\in R$$

- $t(n)$ 的上侧分位数 $t_\alpha(n)$ （$0<\alpha<1$）:

$$P\{T>t_\alpha(n)\}=\int_{t_\alpha(n)}^{+\infty}f_T(x)dx=\alpha$$



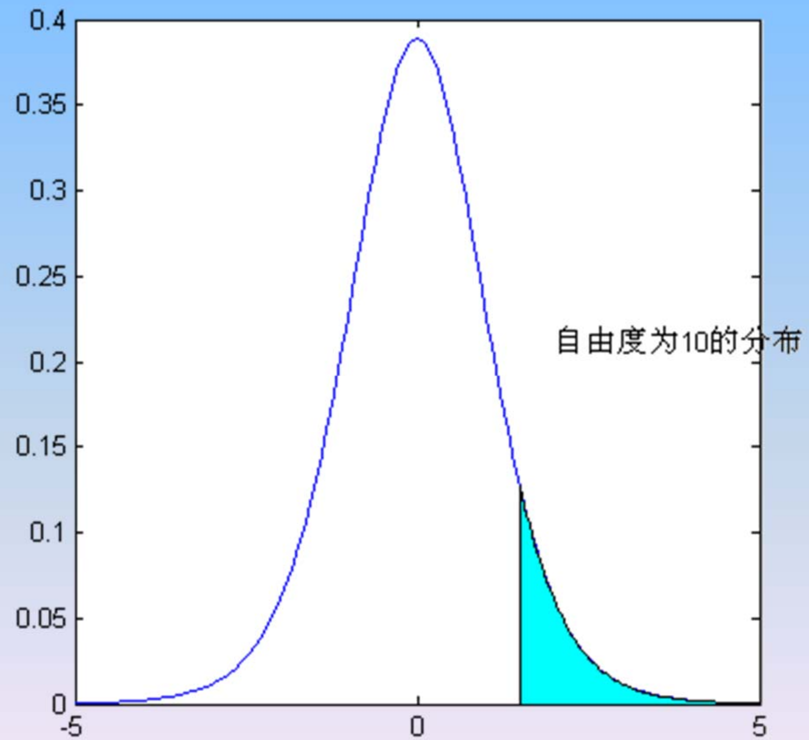
定理6.2.2 设 随机变量X，Y 相互独立， $X \sim N(0,1)$，$Y \sim \chi^2(n)$，则

$$T = \frac{X}{\sqrt{Y/n}} \sim t(n)$$

即随机变量 T 服从自由度为 n 的 t 分布。

T 分布的特点：

- 关于纵轴对称： $\boxed{t_{1-\alpha}(n) = -t_{\alpha}(n)}$
- *n* 较大时(*n*>45)， $t_{\alpha}(n) \approx u_{\alpha}$



例 查表计算： $t_{0.95}(20)=?$ $t_{0.95}(80)=?$

解： $t_{0.95}(20)=t_{1-0.05}(20)=-t_{0.05}(20)=-1.7247$

$$t_{0.95}(80)=-t_{0.05}(80)\approx -u_{0.05}=-1.645$$

4. F 分布 $F\sim F(n_1,n_2)$

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

称X服从第一自由度为n_1,第二自由度为n_2的F分布.

定理6.2.3 设 随机变量X，Y相互独立， $X\sim\chi^2(n_1)$ ，$Y\sim\chi^2(n_2)$，则

$$F=\frac{X/n_1}{Y/n_2}\sim F(n_1,n_2)$$

即随机变量 F服从第一自由度为n_1，第二自由度为n_2的F分布。

推论

$$\text{若}\quad F\sim F(n_1,n_2)\Rightarrow\frac{1}{F}\sim F(n_2,n_1)$$

$F(n_1, n_2)$的上侧分位数$F_\alpha(n_1, n_2)$ （$0<\alpha<1$）:

$$P\{F>F_\alpha(n_1,n_2)\}=\int_{F_\alpha(n_1,n_2)}^{+\infty}f_F(x)dx=\alpha$$

推论: 若 $F \sim F(n_1, n_2) \Rightarrow F_{1-\alpha}(n_1, n_2) = \dfrac{1}{F_\alpha(n_2, n_1)}$

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

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

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

例 统计量的分布(之二)

设 X_1，X_2，…，X_{n+m}是来自正态总体$X \sim N(0,\sigma^2)$的容量为 $n+m$的样本，求下列统计量的概率分布：

$$1. Y=\frac{1}{\sigma^2}\sum_{i=1}^{n+m}X_i^2 \qquad 2. Z=\sqrt{\frac{m}{n}}\sum_{i=1}^{n}X_i \Big/ \sqrt{\sum_{i=n+1}^{n+m}X_i^2} \qquad 3. \frac{1}{Z^2}$$

解：

1. $\frac{X_i}{\sigma} \sim N(0,1)$且所有$\frac{X_i}{\sigma}$相互独立$(i=1,2,\dots,n+m)$

故 $$\sum_{i=1}^{n+m}\left(\frac{X_i}{\sigma}\right)^2=\frac{1}{\sigma^2}\sum_{i=1}^{n+m}X_i^2=Y\sim\chi^2(n+m)$$

$$2.Z=\sqrt{\frac{m}{n}}\sum_{i=1}^{n}X_i\Big/\sqrt{\sum_{i=n+1}^{n+m}X_i^2}$$

$$2.\quad 因\sum_{i=1}^{n}X_i\sim N(0,n\sigma^2)\Rightarrow U=\sum_{i=1}^{n}X_i\Big/\sqrt{n\sigma^2}\sim N(0,1)$$

同时 $V=\sum_{i=n+1}^{n+m}\frac{X_i^2}{\sigma^2}\sim\chi^2(m)$, U与V相互独立

由t分布结构定理,

$$\frac{U}{\sqrt{V/m}}=\sqrt{\frac{m}{n}}\sum_{i=1}^{n}X_i\Big/\sqrt{\sum_{i=n+1}^{n+m}X_i^2}=Z\sim t(m)$$

$$2.\ Z=\sqrt{\frac{m}{n}}\sum_{i=1}^{n}X_i\Bigg/\sqrt{\sum_{i=n+1}^{n+m}X_i^2}\qquad 3.\ \frac{1}{Z^2}$$

3. 因$Z\sim t(m)$, 根据 t 分布结构定理，有

$$Z=\frac{U}{\sqrt{V/m}},$$

而且 $U\sim N(0,1)\Rightarrow U^2\sim\chi^2(1)$,

$V\sim\chi^2(m)$, U与V相互独立

故 $$\frac{V/m}{U^2/1}=\frac{1}{Z^2}\sim F(m,1)$$

#

二、抽样分布定理

定理6.2.4　设 $X_1, X_2, \ldots, X_n$ 是正态总体 $X \sim N(\mu, \sigma^2)$ 的样本，$\overline{X}, S^2$ 分别是样本均值和样本方差，则

(1) $\overline{X}$ 与 S^2 相互独立；　　(2) $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$；

(3) $\dfrac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$；　(4) $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

证明：(4) 由(2) $\Rightarrow U = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

由(3) $\Rightarrow V = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

由(1)可知：U和V是相互独立的；再由定理6.2.2可得

$$\frac{U}{\sqrt{V/(n-1)}}=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\Bigg/\sqrt{\frac{(n-1)S^2}{\sigma^2}\cdot\frac{1}{n-1}}$$

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

定理6.2.5： 设正态总体 X 与 Y 相互独立，

$X\sim N$（μ_1，σ_1^2），样本为：X_1， X_2， … X_{n1}，

样本均值和样本方差为$\overline{X},S_1^2$；

$Y\sim N$（μ_2，σ_2^2），样本为：Y_1， Y_2， … Y_{n2}，

样本均值和样本方差为 $\overline{Y},S_2^2$ 。

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

(2)当$\sigma_1^2=\sigma_2^2$时，

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

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

[分析] $\overline{X}-\overline{Y}$服从正态分布，$S_w^2$可化为$\chi^2$ 分布，二者组合而成的统计量应服从 t 分布。

证明：(2)　$\boxed{\text{因}\sigma_1^2=\sigma_2^2=\sigma^2}$

$$\text{故}\overline{X}-\overline{Y}\sim N(\mu_1-\mu_2,(\frac{1}{n_1}+\frac{1}{n_2})\sigma^2)$$

$$\text{令}U=\frac{\left(\overline{X}-\overline{Y}\right)-\left(\mu_1-\mu_2\right)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)$$

由χ^2分布的可加性，

$$V=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2}\sim\chi 2(n_1+n_2-2)$$

因 $\overline{X}, S_1^2$，$\overline{Y}, S_2^2$ 相互独立，故U 与 V也相互独立，从而

$$T = \frac{U}{\sqrt{V/(n_1+n_2-2)}}$$

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

#

小　结

定理6.2.1 设 X_1，X_2，…，X_n相互独立且都服从标准正态分布

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

定理6.2.2 设 随机变量X，Y相互独立,$X \sim N(0,1)$,$Y \sim \chi^2(n)$

$$T = \frac{X}{\sqrt{Y/n}} \sim t(n)$$

定理6.2.3 设 随机变量X，Y相互独立，$X \sim \chi^2(n_1)$，$Y \sim \chi^2(n_2)$

$$F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$$

二、抽样分布定理(正态总体的统计量分布)

定理6.2.4 设 $X_1, X_2, \ldots, X_n$ 是正态总体 $X \sim N(\mu, \sigma^2)$ 的样本，$\overline{X}, S^2$ 分别是样本均值和样本方差，则

(1) $\overline{X}$ 与 S^2 相互独立；　(2) $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$；

(3) $\dfrac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$；　(4) $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

定理6.2.5: 设正态总体 X 与 Y 相互独立，

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

$\boxed{(2)\text{当}\sigma_1^2 = \sigma_2^2\text{时},}$ $T = \dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2)$