四、四个抽样分布定理

- **数理统计中最重要的总体：$X \sim N(\mu,\sigma^2)$**
- **问题：如何由样本$X_1,X_2,\ldots,X_n$推断μ,σ^2?**
  - **统计推断通过抽样调查，从样本的统计值来估计总体的参数值，或检验关于总体特征的假设是否可接受。**
- **方法：构造统计量**

① 如何构造好的统计量$g(X_1,X_2,\ldots,X_n)$?

② $g(X_1,X_2,\ldots,X_n)$服从什么分布?

  - **统计量是样本的函数，也是随机变量，统计量所服从的分布称为抽样分布。**
  - **统计推断中最重要的结论：四个抽样分布定理**

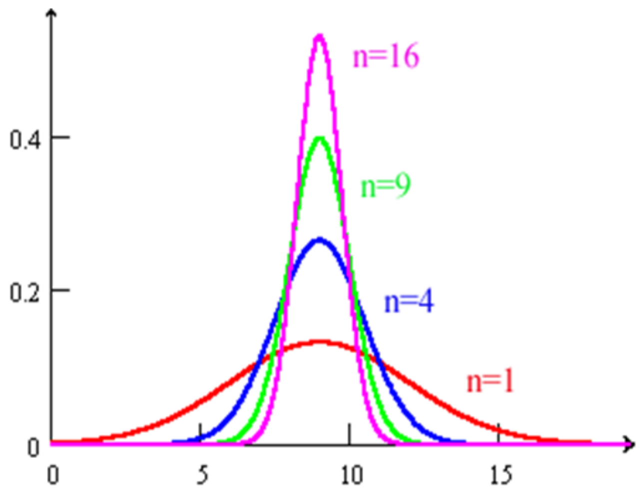
一、单个正态总体的抽样分布

定理1

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

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad 即 \quad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

【提示】在已知总体$\boldsymbol{\mu}$，$\boldsymbol{\sigma}^2$时，可用本定理计算样本均值$\bar{\boldsymbol{X}}$.



n 取不同值时
样本均值 $\bar{X}$ 的分布

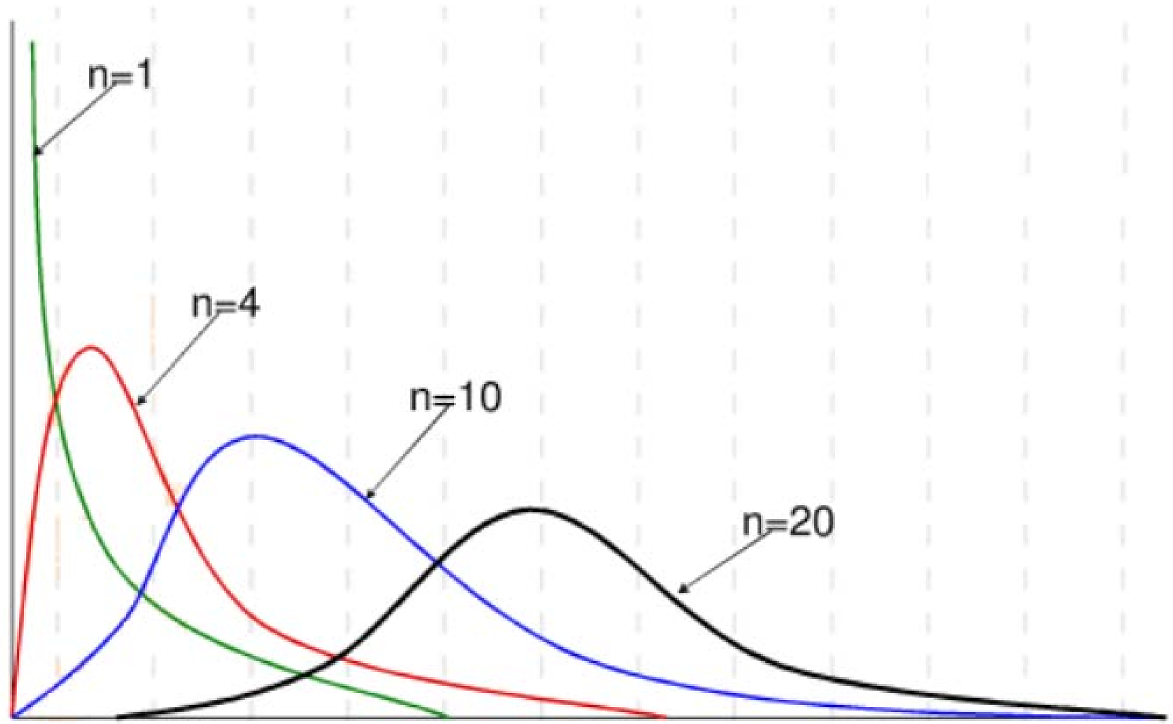
定理2

设 $X_1,X_2,\ldots,X_n$是来自正态总体 $N(\mu,\sigma^2)$的样本，$\bar{X}$和S^2分别为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) $\bar{X}$与S^2独立.

【提示】在已知总体σ^2时，可用本定理计算样本方差S^2.



n 取不同值时 $\dfrac{(n-1)S^2}{\sigma^2}$ 的分布

例　设 $X_1, X_2, \cdots, X_{10}$ 是总体 $X \sim N(\mu, 4)$ 的样本，求样本方差容量 S^2 大于2.622的概率。

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

解：由定理2，$\dfrac{(10-1)S^2}{4} \sim \chi^2(10-1)$

$$P(S^2 > 2.622) = P(\frac{9}{4}S^2 > \frac{9}{4} \times 2.622) = P(\frac{9}{4}S^2 > 5.8995)$$

由 χ^2 分布表，可知 $\chi^2_{0.75}(9) = 5.899$

则近似地有 S^2 大于2.622的概率为0.75。

x^2分布临界值表（卡方分布）

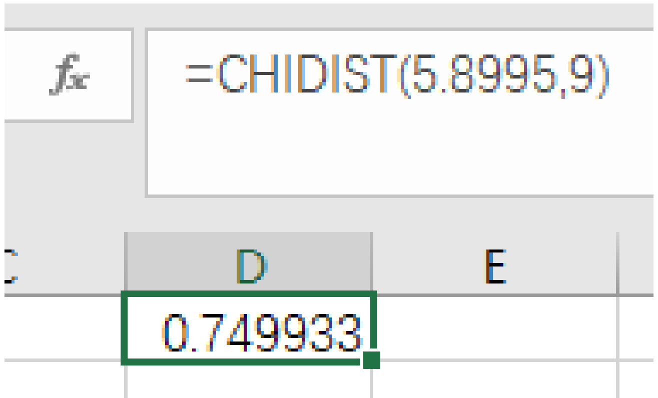
n'	P												
	0.995	0.99	0.975	0.95	0.9	0.75	0.5	0.25	0.1	0.05	0.025	0.01	0.005
1	…	…	…	…	0.02	0.1	0.45	1.32	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.02	0.1	0.21	0.58	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.07	0.11	0.22	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.34	12.84
4	0.21	0.3	0.48	0.71	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.2	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.4	5.07	7.34	10.22	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.7	3.33	4.17	5.9	8.34	11.39	14.68	16.92	19.02	21.67	23.59

$$P(S^2 > 2.622) = P(\frac{9}{4}S^2 > \frac{9}{4}\times 2.622) = P(\frac{9}{4}S^2 > 5.8995)$$

由 χ^2 分布表，可知 $\chi^2_{0.75}(9) = 5.899$

则近似地有 S^2 大于2.622的概率为0.75。

利用Excel函数CHIDIST，求$\chi_{\alpha}^{2}(9)=5.8995$,



可得$\alpha=0.75$,

由此，可得 S^2 大于2.622的概率为0.75

当总体方差 σ^2 未知时，可以用 S^2 来代替，同样也可以得到样本均值的分布。

由定理1、2，可知

$$\boxed{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}_{X} \sim N(0,1)\ ,\ \boxed{\frac{(n-1)S^2}{\sigma^2}}_{Y} \sim \chi^2(n-1)$$ **且相互独立**

根据t分布的定义，可得

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

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

定理3

设$X_1, X_2, \ldots, X_n$是取自正态总体$N(\mu, \sigma^2)$的样本，

$\bar{X}$和S^2分别为样本均值和样本方差，

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

【提示】在总体$\boldsymbol{\sigma}^2$未知时，可用本定理计算样本均值$\bar{X}$.

例　设总体$X \sim N(3,\sigma^2)$，有n=10的样本，样本方差S^2=4，求样本均值 $\bar{X}$ 落在2.1253到3.8747之间的概率。

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

解：由定理3，$\dfrac{\bar{X}-3}{2/\sqrt{10}} \sim t(9)$

$$P(2.1253 \le \bar{X} \le 3.8747) = P\left(\frac{2.1253-3}{2/\sqrt{10}} \le \frac{\bar{X}-3}{2/\sqrt{10}} \le \frac{3.8747-3}{2/\sqrt{10}}\right)$$

$$= P\left(-1.3830 \le \frac{\bar{X}-3}{2/\sqrt{10}} \le 1.3830\right)$$

由t分布表，可得 $t_{0.1}(9)=1.3830$，由t分布的对称性及α分位点的定义，上述概率为

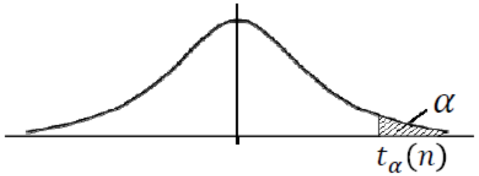
$$t_{1-\alpha}(n) = -t_{\alpha}(n) \qquad P(2.1253 \le \bar{X} \le 3.8747) = 1-2\times 0.1=0.8$$

例　设总体$X \sim N(3,\sigma^2)$，有n=10的样本，样本方差S^2=4，

。

附表 4　t 分布表

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



解:

n \ α	0.2	0.15	0.1	0.05	0.025	0.01	0.005
6	0.906	1.134	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.896	1.119	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.889	1.108	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.883	1.100	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.879	1.093	1.3722	1.8125	2.2281	2.7638	3.1693

$$= P\left(-1.3830 \le \frac{\bar{X}-3}{2/\sqrt{10}} \le 1.3830\right)$$

由t分布表，可得 $t_{0.1}(9) = 1.3830$，由t分布的对称性及α分位点的定义，上述概率为

$$t_{1-\alpha}(n) = -t_\alpha(n) \qquad P(2.1253 \le \bar{X} \le 3.8747) = 1-2\times 0.1 = 0.8$$

二、两个正态总体的抽样分布

定理4

设$\boldsymbol{X}\sim \boldsymbol{N}(\boldsymbol{\mu_1},\boldsymbol{\sigma_1^2})$，$\boldsymbol{Y}\sim \boldsymbol{N}(\boldsymbol{\mu_2},\boldsymbol{\sigma_2^2})$)，且$\boldsymbol{X}$与$\boldsymbol{Y}$独立，

$\boldsymbol{X_1},\boldsymbol{X_2},\dots,\boldsymbol{X_{n_1}}$是来自$\boldsymbol{X}$的样本，$\boldsymbol{Y_1},\boldsymbol{Y_2},\dots,\boldsymbol{Y_{n_2}}$是取自$\boldsymbol{Y}$的样本，

$\overline{\boldsymbol{X}}$和$\overline{\boldsymbol{Y}}$分别是这两个样本的样本均值，$\boldsymbol{S_1^2}$和$\boldsymbol{S_2^2}$

分别是这两个样本的样本方差，则有

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

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

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\underbrace{\boxed{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}}}_{S_W}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

证明:(1) 由定理2知,$\frac{(n_1-1)S_1^2}{\sigma_1{}^2} \sim \chi^2(n_1-1)$,$\frac{(n_2-1)S_2^2}{\sigma_2{}^2} \sim \chi^2(n_2-1)$

且两者独立，由F分布的定义，有：

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

(2) 由定理1, $\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$, $\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$,

且$\bar{X}$与$\bar{Y}$相互独立，所以$\bar{X}-\bar{Y} \sim N(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$,

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

当$\sigma_1^2=\sigma_2^2=\sigma^2$时，得 $U=\dfrac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim N\left(0,1\right)$

又由给定条件知：$\dfrac{\left(n_1-1\right)S_1^2}{\sigma^2}\sim\chi^2\left(n_1-1\right),\dfrac{\left(n_2-1\right)S_2^2}{\sigma^2}\sim\chi^2\left(n_2-1\right)$

且它们相互独立,故有χ^2分布的可加性知：

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

由附录2可知：U与V相互独立,

于是按t分布知：

$$\frac{U}{\sqrt{V/\left(n_1+n_2-2\right)}}=\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t\left(n_1+n_2-2\right)$$

例 两独立总体 $X \sim N(30,12), Y \sim N(20,18)$

从两总体中分别抽取容量为16, 25的样本

(1) 求两样本均值差在 (8, 12) 内的概率;

(2) 求两样本方差比不大于1.4的概率.

解:

(1) 由定理1可知

$$\boxed{\bar{X}-\bar{Y}} \sim N(30-20, \frac{12}{16}+\frac{18}{25}) = \boxed{N(10, 1.47)}$$

于是

$$P(8<\bar{X}-\bar{Y}<12) = \Phi(\frac{12-10}{\sqrt{1.47}}) - \Phi(\frac{8-10}{\sqrt{1.47}})$$

$$= 2\Phi(1.65) - 1 = 2\times 0.9505 - 1 = 0.9010$$

(2) 由定理4可 知

$$\frac{{S_1}^2/{S_2}^2}{12/18} \sim F(16-1,25-1)=F(15,24)$$

从而

$$\mathrm{P}(S_1^2/S_2^2 \leq 1.4)=\mathrm{P}(\frac{{S_1}^2/{S_2}^2}{12/18} \leq \frac{1.4}{12/18})$$

$$=\mathrm{P}(\frac{{S_1}^2/{S_2}^2}{12/18} \leq 2.1)$$

$$=0.95$$

例 总体 $X\sim N(\mu, 3)$，$Y\sim N(\mu, 5)$ 中分别抽取 $n_1=10,\ n_2=15$ 的独立样本，求二样本方差之比 S_1^2/S_2^2 大于 1.272 的概率。

解：由定理 4，$\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$

$$\frac{S_1^2}{S_2^2}\cdot\frac{5}{3}\sim F(9,\ 14)$$

$$P\left(\frac{S_1^2}{S_2^2}>1.272\right)=P\left(\frac{S_1^2}{S_2^2}\cdot\frac{5}{3}>1.272\times\frac{5}{3}\right)=P\left(\frac{S_1^2}{S_2^2}\cdot\frac{5}{3}>2.12\right)$$

由 F 分布表

$$F_{0.1}(9,\ 14)=2.12$$

即 $$P\left(\frac{S_1^2}{S_2^2}>1.272\right)=0.1.$$

单个正态总体的抽样分布

定理 1 ~ 3 给出了样本均值 $\bar{X}$ 和样本方差 S^2 的分布，可用于对正态总体的期望 μ 和方差 σ^2 进行统计推断（区间估计、假设检验）。

两个正态总体的抽样分布

定理 4 给出了样本均值差 $\bar{X}-\bar{Y}$ 和样本方差比 S_1^2/S_2^2 的分布，可用于对正态总体的期望差 μ_1-μ_2 和方差比 σ_1^2/σ_2^2 进行统计推断。