概率论与数理统计

正态总体的抽样分布

主讲人：郑旭玲

信息科学与技术学院

抽样分布

- 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。
- 研究统计量的分布，是统计推断中十分重要且基础的工作。
- 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**。

抽样分布

设总体X的均值为μ, 方差为σ^2, $X_1, X_2, \cdots, X_n$是来自总体的一个样本，则

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

样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ ， $E(S^2) = \sigma^2$

- 本节给出正态总体一些重要抽样分布定理

01

单个正态总体的抽样分布

样本均值的分布

定理1

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

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

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) \sim \text{正态分布}$$

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

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

样本均值的分布

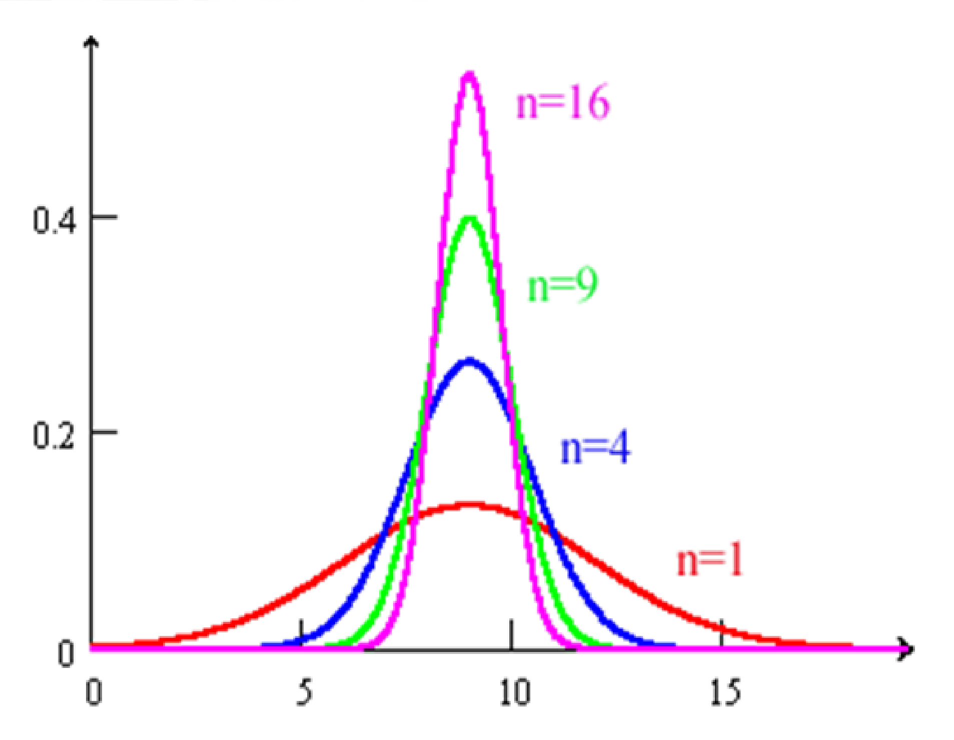
定理1

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

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

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

样本均值的分布



n **取不同值时样本均值 $\bar{X}$ 的分布**

样本均值的分布

例 设总体 $X \sim N(\mu, 4)$，有样本$X_1, X_2, \dots, X_n$ ，当样本容量 n 为多大时，使 $P(|\bar{X}-\mu| \le 0.1)=0.95$ 。

解： 由定理1， $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$P(|\bar{X}-\mu| \le 0.1) = P\left(\frac{-0.1}{2/\sqrt{n}} \le \frac{\bar{X}-\mu}{2/\sqrt{n}} \le \frac{0.1}{2/\sqrt{n}}\right)$$

$$= \Phi(0.05\sqrt{n}) - \Phi(-0.05\sqrt{n}) = 2\Phi(0.05\sqrt{n}) - 1$$

因$P(|\bar{X}-\mu| \le 0.1)=0.95$，即$2\Phi(0.05\sqrt{n})-1=0.95$

得$\Phi(0.05\sqrt{n})=(1+0.95)/2=0.975$

查标准正态分布表可知 $\Phi(1.96)=0.975$，即 $0.05\sqrt{n}=1.96$

于是得 $n=1536.6 \approx 1537$

样本方差的分布

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，

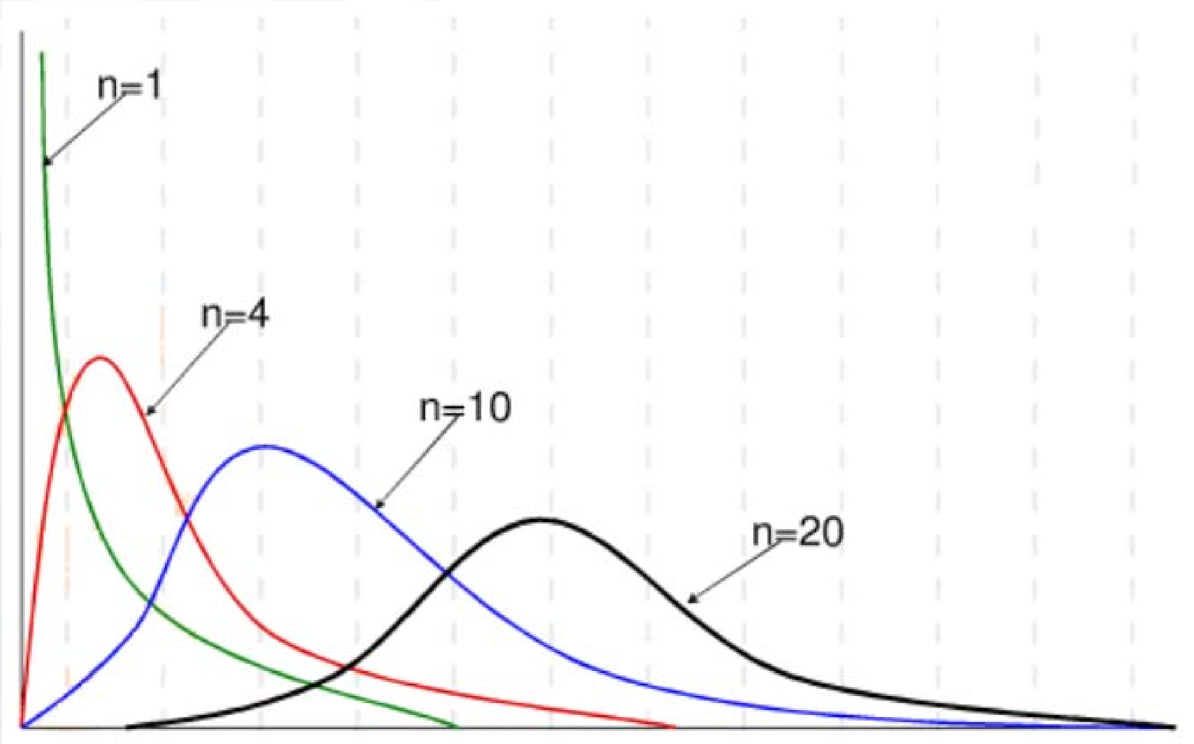
$\bar{X}$和S^2分别为样本均值和样本方差，则有

(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.

样本方差的分布



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

样本方差的分布

例 从正态总体$N(\mu, 0.5^2)$中抽取样本$X_1, \cdots, X_{10}$.

(1) 已知 $\mu = 0$, 求概率 $p\left\{\sum_{i=1}^{10} X_i^2 \geq 4\right\}$;

(2) μ 未知，求概率 $p\left\{\sum_{i=1}^{10} (X_i - \bar{X})^2 \geq 2.85\right\}$.

解： (1) 由 $\mu = 0$， 有 $X_i/0.5 \sim N(0,1)$

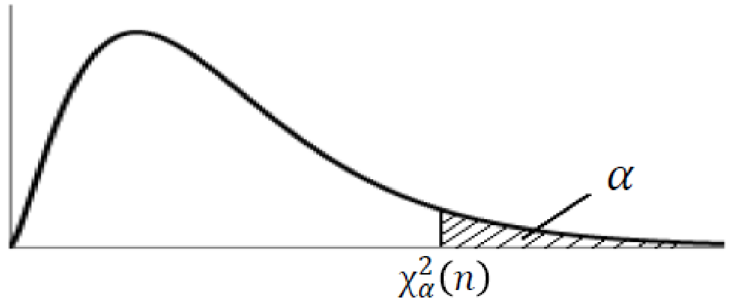
则 $Y \stackrel{\Delta}{=} \frac{1}{0.5^2}\sum_{i=1}^{10} X_i^2 \sim \chi^2(10)$

$$p\left\{\sum_{i=1}^{10} X_i^2 \geq 4\right\} = p\left\{\frac{1}{0.5^2}\sum_{i=1}^{10} X_i^2 \geq \frac{4}{0.5^2}\right\} = p\{Y \geq 16\}$$

样本方差的分布

附表 5 χ^2 分布表

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



n \ α	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.843	5.025	6.637	7.882
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.344	12.837
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.085	16.748
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.440	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.012	18.474	20.276
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.534	20.090	21.954
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.022	21.665	23.587
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

查表求$\chi^2_\alpha(10)=16$，可知$\alpha=0.1$　　由此可得 $p\left\{\sum_{i=1}^{10} X_i^2 \geq 4\right\} = 0.10$

样本方差的分布

例 从正态总体$N(\mu,0.5^2)$中抽取样本$X_1,\cdots,X_{10}$.

(1) 已知 $\mu=0$, 求概率 $p\left\{\sum_{i=1}^{10}X_i^2\geq 4\right\}$;

(2) μ 未知，求概率 $p\left\{\sum_{i=1}^{10}(X_i-\bar{X})^2\geq 2.85\right\}$.

解： (2) 由题设及定理2，$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

$$Z\overset{\Delta}{=}\frac{9S^2}{0.5^2}=\frac{1}{0.5^2}\sum_{i=1}^{10}(X_i-\bar{X})^2\sim\chi^2(9)$$

$$p\left\{\sum_{i=1}^{10}(X_i-\bar{X})^2\geq 2.85\right\}=p\left\{\frac{1}{0.5^2}\sum_{i=1}^{10}(X_i-\bar{X})^2\geq\frac{2.85}{0.5^2}\right\}=P\{Z\geq 11.4\}$$

样本方差的分布

利用Excel函数CHIDIST，求$\chi_{\alpha}^{2}(9)=11.4,$

可得$\alpha=0.25,$

由此，可得 $p\left\{\sum_{i=1}^{10}(X_i-\bar{X})^2 \geq 2.85\right\}=0.25$

样本均值的分布

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

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

样本均值的分布

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σ^2未知时，可用本定理计算样本均值$\bar{X}$.

样本均值的分布

例 设总体 X 服从正态分布 $N(12,\sigma^2)$, 抽取容量为 25 的样本，求样本均值 $\bar{X}$ 大于 12.5 的概率。

如果 (1) 已知 $\sigma=2$; (2) σ 未知，但已知样本方差 $S^2=5.57$。

解： （1）由定理1，可知 $\dfrac{\bar{X}-12}{2/\sqrt{25}} \sim N(0,1)$

$$\text{故 } P\{\bar{X}>12.5\}=P\left\{\frac{\bar{X}-12}{2/\sqrt{25}}>\frac{12.5-12}{2/\sqrt{25}}\right\}$$

$$=P\left\{\frac{\bar{X}-12}{0.4}>1.25\right\}$$

$$=1-\Phi(1.25)=1\text{-}0.8944=0.1056$$

样本均值的分布

例 设总体 X 服从正态分布 $N(12,\sigma^2)$，抽取容量为 25 的样本，求样本均值 $\bar{X}$ 大于 12.5 的概率。

如果 (1) 已知 $\sigma=2$；(2) σ 未知，但已知样本方差 $S^2=5.57$。

解： （2）由定理3，可知 $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}=\dfrac{\bar{X}-12}{S/\sqrt{25}}\sim t(24)$

$$\text{故 } P\{\bar{X}>12.5\}=P\left\{\frac{\bar{X}-12}{S/5}>\frac{12.5-12}{S/5}\right\}$$

$$=P\{T>1.059\}$$

查表求 $t_\alpha(24)=1.059$，可知 $\alpha=0.15$

故有 $P\{\bar{X}>12.5\}=0.15$

02

两个正态总体的抽样分布

两总体样本均值差、样本方差比的分布

定理4

设 $X \sim N(\mu_1,\sigma_1^2)$, $Y \sim N(\mu_2,\sigma_2^2)$, 且$X$与$Y$独立，$X_1,X_2,\ldots,X_{n_1}$是来自$X$的样本，$Y_1,Y_2,\ldots,Y_{n_2}$是取自$Y$的样本，$\bar{X}$ 和$\bar{Y}$分别是这两个样本的样本均值，S_1^2和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，$\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\boxed{\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2)$

S_W

两总体样本均值差、样本方差比的分布

例 设总体 $X\sim N(6,\sigma_1^2)$, $Y\sim N(5,\sigma_2^2)$, 有 $n_1=n_2=10$ 的独立样本，求二样本均值之差 $\bar{X}-\bar{Y}$小于 1.3 的概率。

如果 (1) 已知 $\sigma_1^2=1$, $\sigma_2^2=1$;

(2) σ_1^2, σ_2^2 未知，但二者相同，样本方差分别为

$S_1^2=0.9130$, $S_2^2=0.9816$。

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

由正态分布的可加性，有 $\bar{X}\text{-}\bar{Y}\sim N(\mu_1\text{-}\mu_2,\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2})$

两总体样本均值差、样本方差比的分布

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

代入μ_1=6，μ_2=5, $n_1=n_2=10$, $\sigma^2=1$ 时，有 $\dfrac{(\bar{X}-\bar{Y})-(6-5)}{\sqrt{\dfrac{1}{10}+\dfrac{1}{10}}}\sim N(0,1)$

故 $P\left\{\bar{X}-\bar{Y}<1.3\right\}=P\left\{\dfrac{(\bar{X}-\bar{Y})-(6-5)}{\sqrt{\dfrac{1}{10}+\dfrac{1}{10}}}<\dfrac{1.3-(6-5)}{\sqrt{\dfrac{1}{10}+\dfrac{1}{10}}}\right\}=P\{U<0.67\}$

$$=\Phi(0.67)=0.7486$$

两总体样本均值差、样本方差比的分布

例 设总体 $X\sim N(6,\sigma_1^2)$, $Y\sim N(5,\sigma_2^2)$, 有 $n_1=n_2=10$ 的独立样本，求二样本均值之差 $\bar{X}-\bar{Y}$小于 1.3 的概率。

如果 (1) 已知 $\sigma_1^2=1$, $\sigma_2^2=1$;

(2) σ_1^2, σ_2^2 未知，但二者相同，样本方差分别为 $S_1^2=0.9130$, $S_2^2=0.9816$。

解： (2) 由定理4(2)，可知 $\dfrac{\bar{X}-\bar{Y}-(6-5)}{S_W\sqrt{\dfrac{1}{10}+\dfrac{1}{10}}}\sim t(18)$

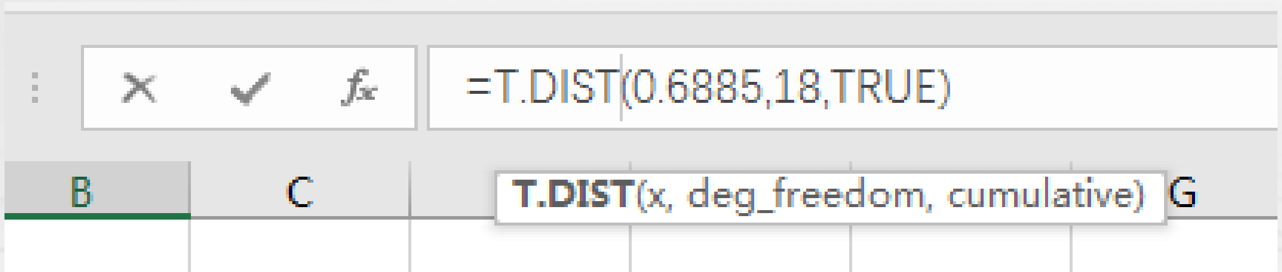
其中，$S_W^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}=\dfrac{9\times 0.9130+9\times 0.9816}{18}=0.9733^2$

两总体样本均值差、样本方差比的分布

故 $P\left\{\overline{X}-\overline{Y}<1.3\right\}=P\left\{\dfrac{(\overline{X}-\overline{Y})-(6-5)}{0.9733\sqrt{\dfrac{1}{10}+\dfrac{1}{10}}}<\dfrac{1.3-(6-5)}{0.9733\sqrt{\dfrac{1}{10}+\dfrac{1}{10}}}\right\}$

$=P\left\{t(18)<0.6884\right\}$

利用$Excel$函数$T.DIST$，可以求得$P\left\{t(18)<0.6884\right\}=0.75$



故有 $P\left\{\overline{X}-\overline{Y}<1.3\right\}=0.75$

两总体样本均值差、样本方差比的分布

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，求两样本方差之比 S_1^2 / S_2^2 大于1.272的概率。

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

可知 $\dfrac{S_1^2}{S_2^2} \times \dfrac{5}{3} \sim F(9,14)$

故 $P\left\{\dfrac{S_1^2}{S_2^2} > 1.272\right\} = P\left\{\dfrac{S_1^2}{S_2^2} \times \dfrac{5}{3} > 1.272 \times \dfrac{5}{3}\right\} = P\{F(9,14) > 2.12\}$

查F分布表，求$F_\alpha(9,14)=2.12$, 可得$\alpha=0.1$。即 $P\left\{\dfrac{S_1^2}{S_2^2} > 1.272\right\}=0.1$

小结

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。

谢谢大家