1.单个正态总体参数的置信区间

讨论正态分布中σ和μ的置信区间

1.1 σ已知时μ的置信区间

在这种情况下，μ的点估计为$\overline{x}$，分布为$N(\mu,\frac{\sigma^2}{n})$，显然枢纽量可选为:

$$G=\frac{\overline{x}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

c和d应满足$P(c\le G\le d)=\Phi(d)-\Phi(c)=1-\alpha$，由于标准正态分布为单峰对称的，为了让区间最小，即令$d-c$最小，不难发现$d=-c=\mu_{1-\alpha/2}$时成立，则变形后得到:

$$P_\mu(\overline{x}-\mu_{1-\alpha/2}\sigma/\sqrt{n}\le\mu\le\overline{x}+\mu_{1-\alpha/2}\sigma/\sqrt{n})$$

$$\text{即}\overline{x}\pm u_{1-\alpha/2}\sigma/\sqrt{n}$$

1.2σ未知时μ的置信区间

则只能用样本方差s替换σ，此时可用枢纽量:

$$t=\frac{\sqrt{n}(\overline{x}-\mu)}{s}\sim t(n-1)$$

t分布也是一个单峰对称的，则类似1.1中的推理，得到μ的$1-\alpha$置信区间为:

$$\overline{x}\pm t_{1-\alpha/2}(n-1)s/\sqrt{n}$$

此处$s^2=\frac{1}{n-1}\sum_{i=1}^n(x_i-\overline{x})^2$是$\sigma^2$的无偏估计

1.3 σ^2的置信区间

实际上σ^2未知而μ已知的情况是很罕见的，所以只讨论μ未知的情况。

则可选择枢纽量:

$$G=\frac{(n-1)s^2}{\sigma^2}\sim\chi^2(n-1)$$

由于χ^2分布是偏态分布，则寻找等尾置信区间：把α平分为两部分，在分布两侧各截面积为$\frac{\alpha}{2}$的部分，即采用χ^2的两个分位数$\chi^2_{\alpha/2}(n-1)$和$\chi^2_{1-\alpha/2}(n-1)$，则容易化简的出σ^2的$1-\alpha$的置信区间为:

$$[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)},\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}]$$

2.两个正态分布下的置信区间

即$N(\mu_1,\sigma_1^2)$，$N(\mu_2,\sigma_2^2)$，$\overline{x},\overline{y}$分别是他们的均值，$s^2=\frac{1}{m-1}\sum_{i=1}^m(x-\overline{x})^2,s_y^2=\frac{1}{n-1}\sum_{i=1}^n(y_i-\overline{y})^2$分别是他们的样本方差。

2.1 $\mu_1-\mu_2$的置信区间

有如下情况:

2.1.1 σ_1^2，σ_2^2已知

此时$\overline{x}-\overline{y}\sim N(\mu_1-\mu_2,\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n})$,显然可归一为标准正态分布，取枢纽量为:

$$u=\frac{\overline{x}-\overline{y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\sim N(0,1)$$

易得，$\mu_1-\mu_2$的$1-\alpha$置信区间为:

$$\overline{x}-\overline{y}\pm u_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}$$

2.1.2 $\sigma_1^2=\sigma_2^2=\sigma^2$未知

仍然有：$\overline{x}-\overline{y}\sim N(\mu_1-\mu_2,\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n})$,且:

$$\frac{(m-1)s_x^2+(n-1)s_y^2}{\sigma^2}\sim\chi^2(m+n-2)$$

且整体的样本方(两个正态分布的方差合在一起，某一个练习题有证明过):

$$s_w^2=\frac{(m-1)s_x^2+(n-1)s_y^2}{m+n-1}$$

那么可以取枢纽量:

$$G=\frac{\overline{x}-\overline{y}-(\mu_1-\mu_2)}{s_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(m+n-2)$$

易解出$\mu_1-\mu_2$的$1-\alpha$置信区间为:

$$\overline{x}-\overline{y}\pm\sqrt{\frac{m+n}{mn}s_wt_{1-\alpha/2}(m+n-2)}$$

2.2 $\frac{\sigma_1^2}{\sigma_2^2}$的置信区间

由于:

$$\frac{(m-1)s_x^2}{\sigma_1^2}\sim\chi^2(m-1)$$
$$\frac{(n-1)s_y^2}{\sigma_2^2}\sim\chi^2(n-1)$$

则可构造枢纽量:

$$F=\frac{s_x^2/\sigma_1^2}{s_y^2/\sigma_2^2}\sim F(m-1,n-1)$$

F分布也是偏态分布，所以在选取时也是选择等尾部置信，即:

$$P(F_{\alpha/2}(m-1,n-1)\le\frac{s_x^2}{s_y^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}\le F_{1-\alpha/2}(m-1,n-1))=1-\alpha$$

易解出σ_1^2/σ_2^2的$1-\alpha$置信区间为:

$$[\frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{1-\alpha/2}(m-1,n-1)},\frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{\alpha/2}(m-1,n-1)}]$$

3.三大抽样分布

3.1 卡方分布

定义 3.1.1

设X_1,X_2，$\cdots,X_n$独立同分布于标准正态分布$N(0,1)$，则$\chi^2=X_1^2+\cdots+X_n^2$的分布称为自由度为$n$的$\chi^2$分布，记为$\chi^2\sim\chi^2(n)$

我们知道若$X\sim N(0,1)$,则$X^2\sim Ga(\frac{1}{2},\frac{1}{2})$，由伽马分布的可加性有$\chi^2\sim Ga(\frac{n}{2},\frac{1}{2})=\chi^2(n)$，由此可见$\chi^2(n)$是伽马分布的特例，其分布的密度函数为：

$$p(y)=\frac{(1/2)^{\frac{n}{2}}}{\Gamma(n/2)}y^{\frac{n}{2}-1}e^{-\frac{y}{2}}$$

该函数是一个偏态分布，期望$E(\chi^2)=n$,方差$Var(\chi^2)=2n$。

且有如下定理：

定理3.1.1

设$x_1,\cdots,x_n$是来自正态总体$N(\mu,\sigma^2)$的样本，其样本均值和样本方差分别为：

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i，s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

则有：

1. $\overline{x}$与s^2相互独立
2. $\overline{x}\sim N(\mu,\frac{\sigma^2}{n})$
3. $\frac{s^2}{\frac{\sigma^2}{n-1}}\sim\chi^2(n-1)$

证明在P285

3.2 F分布

定义3.2.1

设随机变量$X_1\sim\chi^2(m),X_2\sim\chi^2(n)$，二者相互独立，则称$F=\frac{X_1/m}{X_2/m}$的分布是自由度为$m$与$n$的$F$分布，记为$F\sim F(m,n)$，其中$m$称为分子自由度，$n$称为分母自由度

密度函数在p286

F分布的一个性质：

$$F_\alpha(n,m)=\frac{1}{F_{1-\alpha}(m,n)}$$

3.3 t分布

定义3.3.1

设随机变量X_1与X_2独立且$X_1 \sim N(0,1)$，$X_2 \sim \chi^2(n)$，则称$t = \frac{X_1}{\sqrt{X_2/n}}$的分布为自由度为$n$的$t$分布，记为$t \sim t(n)$

容易发现$t^2 = \frac{X_1}{X_2/n} \sim F(1,n)$，求导即可得t分布的密度函数，p289

两个变量总的方差：

$$s_w^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2}{m+n-2} = \frac{\sum_{i=1}^{m}(x_i - \overline{x}) + \sum_{i=1}^{n}(y_i - \overline{y})^2}{m+n-2}$$

$$\frac{(\overline{x} - \overline{y}) - (\mu_1 - \mu_2)}{s_w\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t(m+n-2)$$