概率统计第十三讲

点估计

本讲提要

- 数理统计引言
- 总体与样本
- 点估计
- 估计量的评价

引 言

- 数理统计是具有广泛应用的一个数学分支，
- 它以概率论为理论基础，
- 研究怎样有效地收集、整理和分析带有随机性的数据，
- 以对所考察的问题作出推断或预测，
- 并为采取一定的决策和行动提供依据和建议。

几个实际问题:

1. 估计产品寿命问题：根据用户调查获得某品牌洗衣机50台的使用寿命为5年，5.5年，3.5年，6.2年，……。根据这些数据希望得到如下推断:

A. 可否认为产品的平均寿命不低于4年？

B. 保质期设为多少年，才能保证有95%以上的产品过关？

商品日投放量问题：

1. 草莓的日投放量多少合理？
2. 如何安排银行各营业网点的现金投放量？
3. 快餐食品以什么样的速度生产最为合理等等。

例 制衣厂为了合理的确定服装各种尺码的生产比例，需要调查人们身高的分布。现从男性成人人群中随机选取100人，得到他们的身高数据为：…

1. 试推断男性成人身高X的概率分布；
2. 若已知X服从正态分布$N(\mu, \sigma^2)$，试估计参数的μ和σ^2值。

已知“总体”的分布类型，对分布中的未知参数所进行的统计推断属于“参数估计”。

总体与样本

1、总体

总体指研究对象的全体，通常指研究对象的某项数量指标。组成总体的元素称为个体。

从本质上讲，总体就是所研究的随机变量或随机变量的分布。

2、样本

来自总体的部分个体X_1，…，X_n如果满足：

1. 同分布性：X_i，i=1,…,n与总体同分布

2. 独立性：X_1，…，X_n相互独立；

则称为容量为n的简单随机样本，简称样本。

而称X_1，…，X_n的一次实现为样本观察值，记为x_1，…，x_n。

来自总体X的随机样本X_1，…，X_n可记为

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} X \text{或} F(x), p(x), \ldots$$

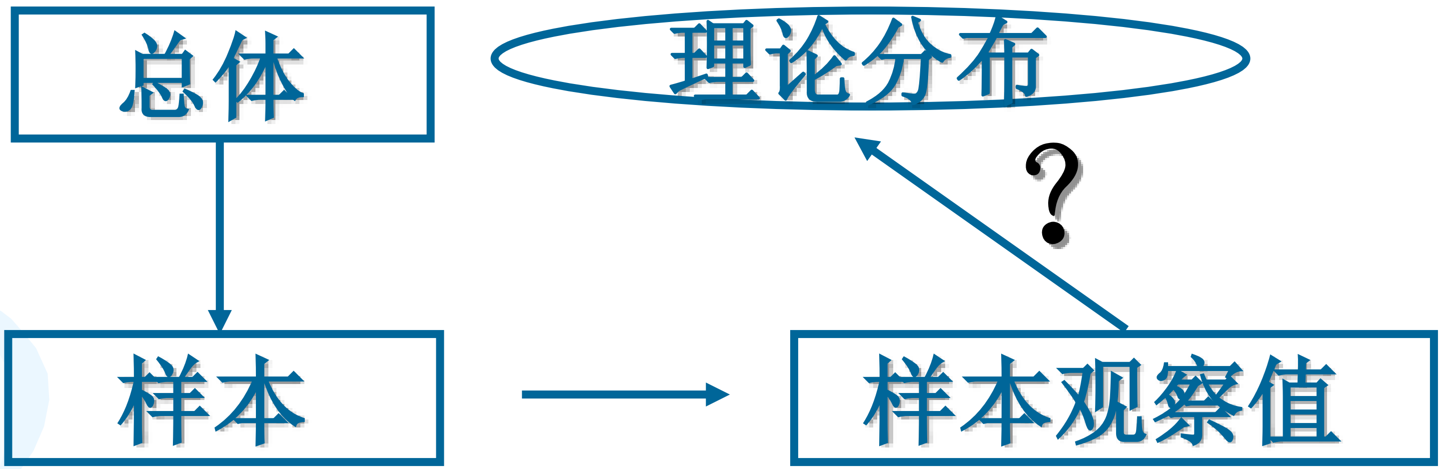
显然，样本联合分布函数为

$$F(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F(x_i),$$

样本联合密度函数为

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p(x_i).$$

3、总体、样本和样本观察值的关系



统计是从手中已有的资料——样本观察值，去推断总体的情况——总体分布。样本是联系两者的桥梁。总体分布决定了样本取值的概率规律，也就是样本取到样本观察值的规律，因而可以用样本观察值去推断总体。

统计量

定义：称样本x_1，…，x_n的函数$T(x_1,\ldots,x_n)$是总体的一个统计量，如果$T(x_1,\ldots,x_n)$不含未知参数。

几个常用的统计量：

1. 样本均值 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i,$

2. 样本方差 $s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

或 $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (无偏方差),

样本标准差 $s_n = \sqrt{s_n^2}$或$s = \sqrt{s^2}$ (无偏标准差).

3.样本*k*阶矩

原点矩 $\hat{a}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$,

中心矩 $\hat{b}_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$。

4.经验分布函数

用$S(x)$表示随机样本x_1，…，x_n中不大于x的样本个数。定义经验分布函数：

$$F_n(x) = S(x)/n。$$

Glivenko定理

$$\mathrm{P}\left(\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0 \right) = 1.$$

点估计

一、参数估计的概念

定义 设$x_1,\cdots,x_n$是总体的一个样本，其分布函数为$F(x;\theta)$，$\theta\in\Theta$；其中θ为未知参数，Θ为参数空间，若统计量$\theta'(x_1,\cdots,x_n)$可作为θ的一个估计,则称其为θ的一个估计量，记为$\hat{\theta}=\theta'(x_1,x_2,\ldots,x_n)$.

注: $F(x;\theta)$可用分布律或密度函数代替。

若$x_1,\cdots,x_n$是样本的一个观测值，则

$\hat{\theta}=\theta'(x_1,x_2,...,x_n)$称为$\theta$的估计值。

- 由于$\theta'(x_1,\cdots,x_n)$是实数域上的一个点，现用它来估计θ,故称这种估计为点估计
- 点估计的经典方法是
  1. 矩估计
  2. 最大似然估计
  3. Bayes估计

二、矩法估计（简称“矩法”）

1.用样本矩作为总体同阶矩的估计，即

若$\alpha_k = EX^k$，则$\hat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$；

若$\beta_k = E(X - EX)^k$，则$\hat{\beta}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$。

2.约定：若$\hat{\theta}$是未知参数θ的矩估计，则$g(\theta)$的矩估计为$g(\hat{\theta})$。

例1. 设x_1，…，x_n为取自总体$X\sim b(m,p)$的样本，其中m已知，$0 < p < 1$未知，求p的矩估计。

解: $EX = mp$, 故$p = EX/m = \mu/m$。

而$\hat{\mu} = \bar{x}$，故p的矩估计$\hat{p} = \bar{x} / m$。

Ex. x_1，…，x_n为取自参数为λ的指数分布总体X的样本，求λ的矩估计。

解：由$X \sim Exp(\lambda)$知$EX = 1/\lambda$，

所以$\lambda = 1/EX = 1/\mu$，得$\hat{\lambda} = 1 / \bar{x}$。

例2. 设$x_1, \ldots, x_n$为取自总体$X \sim N(\mu,\sigma^2)$的样本，求参数μ，σ^2的矩估计。

解：$EX = \mu$，$Var(X) = \sigma^2$,

所以，$\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$。

例3. 设$x_1, \ldots, x_n \overset{\text{i.i.d.}}{\sim} U(a, b)$，试求$\hat{a}$和$\hat{b}$。

解：$EX = (a+b)/2$，$Var(X) = (a-b)^2/12$。

$$\begin{cases} a = EX - \sqrt{3Var(X)}, \\ b = EX + \sqrt{3Var(X)}. \end{cases} \Rightarrow \begin{cases} \hat{a} = \bar{x} - \sqrt{3}s, \\ \hat{b} = \bar{x} + \sqrt{3}s. \end{cases}$$

例4. 设总体X的概率密度为 $p(x)=\frac{1}{2\sigma}e^{-|x|/\sigma}$,

$x_1, \ldots, x_n$为样本，求参数σ的矩估计。

解：$\mu \equiv EX = \int_{-\infty}^{\infty}\frac{x}{2\sigma}e^{-|x|/\sigma}dx = 0,$

$$\alpha_2 \equiv EX^2 = \int_{-\infty}^{\infty}\frac{x^2}{2\sigma}e^{-|x|/\sigma}dx = \frac{1}{\sigma}\int_0^{\infty}x^2e^{-x/\sigma}dx$$

$$= -\int_0^{\infty}x^2de^{-x/\sigma} = \int_0^{\infty}2xe^{-x/\sigma}dx$$

$$= -2\sigma\int_0^{\infty}xde^{-x/\sigma} = 2\sigma\int_0^{\infty}e^{-x/\sigma}dx = 2\sigma^2,$$

$$\Rightarrow \sigma = \sqrt{\alpha_2/2},\quad \hat{\sigma} = \sqrt{\hat{\alpha}_2/2} = \sqrt{\sum_{i=1}^{n}x_i^2/2n}。$$

三、最大似然估计

1、最大似然思想

1832年5月30日，法国著名数学家Galois与情敌决斗，离25步远用手枪射击，他的命中率为0.1, 情敌的命中率为0.9，第二天传闻有一人胃部中弹，24小时后去世。估计谁中弹身亡?

- 一般说，事件A发生的概率与参数$\theta\in\Theta$有关，θ取值不同，则$P(A)$也不同。
- 因而应记事件A发生的概率为$P(A|\theta)$
- 若A发生了，则认为此时的θ值应是在Θ中使$P(A|\theta)$达到最大的那一个，这就是**最大似然思想**。

I. 设总体X为离散型随机变量，它的分布律为

$$\mathrm{P}(X=a_k)=\mathrm{P}_\theta(a_k),\ \ k=1,2,\ldots$$

现有样本观察值$x_1,\ldots,x_n$,取值于$\{a_k,k=1,2,\ldots\}$.
则根据最大似然思想，如何用$x_1,\ldots,x_n$估计θ?

记$A=\{X_1=x_1,\ \ldots,\ X_n=x_n\}$,

则$\mathrm{P}(A\,|\,\theta)=\mathrm{P}_\theta(X_1=x_1,\ldots,X_n=x_n)=\prod_{i=1}^{n}\mathrm{P}_\theta(x_i).$

根据最大似然思想，θ值应是在Θ中使$\mathrm{P}(A\,|\,\theta)$达到最大的那一个，也就是使样本联合分布

$\prod_{i=1}^{n}\mathrm{P}_\theta(x_i)$达到最大。

II. 设总体X为连续型随机变量，概率密度$p(x;\theta)$，现有样本观察值x_1，x_2，…，x_n，则根据最大似然思想，如何用$x_1,\dots,x_n$估计θ?

记$\delta(x_1,\dots,x_n)$为包围点$(x_1,x_2,\dots,x_n)$的小球域，

$A=\{(X_1,X_2,\dots,X_n)\in\delta(x_1,x_2,\dots,x_n)\}$。

$$\text{则} P(A\mid\theta)=\int\cdots\int_{\delta(x_1,\dots,x_n)} p(x_1,\dots,x_n;\theta)\,\mathrm{d}x_1\cdots\mathrm{d}x_n$$

$$=\int\cdots\int_{\delta(x_1,\dots,x_n)} \prod_{i=1}^{n} p(x_i;\theta)\,\mathrm{d}x_1\cdots\mathrm{d}x_n。$$

根据最大似然思想，θ值应是在Θ中使$P(A|\theta)$即使样本联合密度达到最大的那一个。

2、似然函数与最大似然估计

设$x_1, x_2, ..., x_n \overset{\text{i.i.d.}}{\sim} p(x;\theta)$，$\theta \in \Theta$，则称

$$L(\theta) = L(x_1, x_2, ..., x_n;\theta) = \prod_{i=1}^{n} p(x_i;\theta)$$

为该总体的似然函数。

定义：若有$\hat{\theta} \in \Theta$，使得

$$L(\hat{\theta}) = \max_{\theta\in\Theta} L(\theta) \overset{\text{或}}{=} \sup_{\theta\in\Theta} L(\theta),$$

则称$\hat{\theta}$为θ的最（或极）大似然估计。

3、求最大似然估计的步骤*

设$x_1, \ldots, x_n \overset{\text{i.i.d.}}{\sim} p(x;\theta)$，$\theta \in \Theta$，求 $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$。

（1）做似然函数 $L(\theta) = \prod_{i=1}^{n} p(x_i;\theta)$，

（2）做对数似然函数 $\ln L(\theta) = \sum_{i=1}^{n} \ln p(x_i;\theta)$，

（3）列似然方程，令 $\dfrac{\mathrm{d}[\ln L(\theta)]}{\mathrm{d}\theta} = 0$，

若该方程有解，则从中得到最值为

$$\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)。$$

例5. 设x_1， …， x_n为取自参数为λ的Poisson分布总体X的样本，求λ的最大似然估计。

解：$$L(\lambda)=\prod_{i=1}^{n}P_{\lambda}(X=x_i)=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\frac{\lambda^{x_1+\cdots+x_n}}{e^{n\lambda}\prod_{i=1}^{n}x_i!},$$

$$\ln L(\lambda)=\sum_{i=1}^{n}x_i\ln\lambda-n\lambda-\sum_{i=1}^{n}\ln(x_i!),$$

$$\frac{\mathrm{d}[\ln L(\lambda)]}{\mathrm{d}\lambda}=\frac{1}{\lambda}\sum_{i=1}^{n}x_i-n=0\Rightarrow\hat{\lambda}=\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i.$$

注1：若总体分布中含有多个未知参数，则可解方程组

$$\frac{\partial L}{\partial \theta_i}=0,\quad 或 \frac{\partial \ln L}{\partial \theta_i}=0$$

得出θ_i的最大似然估计$\hat{\theta}_i,\ i=1,2,\dots,m.$

例6. 设x_1，…，x_n为取自$N(\mu,\sigma^2)$总体的样本，求参数μ，σ^2的最大似然估计。

解:

$$L(\mu,\sigma^2)=\prod_{i=1}^{n} p(x_i;\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}},$$

$$\Rightarrow L(\mu,\sigma^2)=\frac{1}{(2\pi)^{n/2}\sigma^n}e^{-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}},$$

$$\ln L(\mu,\sigma^2)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}。$$

令 $$\begin{cases}\dfrac{\partial \ln L(\mu,\sigma^2)}{\partial\mu}=\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)=0,\\[2ex] \dfrac{\partial \ln L(\mu,\sigma^2)}{\partial\sigma^2}=-\dfrac{n}{2\sigma^2}+\displaystyle\sum_{i=1}^{n}\dfrac{\left(x_i-\mu\right)^2}{2\sigma^4}=0.\end{cases}$$

$$\Rightarrow \hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x},\quad \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=s_n^2.$$

注2：最大似然估计具有不变性：

- 若 $\hat{\theta}$ 是未知参数θ的最大似然估计，则 $g(\theta)$ 的最大似然估计为 $g(\hat{\theta})$.
- 于是正态总体 $N(\mu, \sigma^2)$ 的

1. 标准差σ的MLE是 s_n;
2. 分布函数 $F(x) = \Phi((x-\mu)/\sigma)$ 的MLE是 $\Phi((x-\bar{x})/s_n)$;
3. 总体α分位数 $x_\alpha = \mu+\sigma u_\alpha$ 的MLE是 $\bar{x}+s_n u_\alpha$，其中 u_α 为 $N(0,1)$ 分布的α分位数。

例7. 设x_1，…，x_n为取自参数为λ的指数分布总体的样本，$x>0$为一给定实数,求$p=F(x)$的最大似然估计。

解: $p=F(x)=\int_0^x \lambda e^{-\lambda t}\mathrm{d}t=1-e^{-\lambda x}$,

故若λ的最大似然估计为$\hat{\lambda}$, 则p的最大似然估计为$1-e^{-\hat{\lambda}x}$。先求λ的最大似然估计

$$L(\lambda)=\prod_{i=1}^{n}p(x_i;\lambda)=\prod_{i=1}^{n}\lambda e^{-\lambda x_i}=\lambda^n e^{-\lambda\sum_{i=1}^{n}x_i},$$

$$\ln L(\lambda) = \ln\left(\lambda^n e^{-\lambda\sum_{i=1}^{n} x_i}\right) = n\ln\lambda - \lambda\sum_{i=1}^{n} x_i,$$

令 $$\frac{\mathrm{d}[\ln L(\lambda)]}{\mathrm{d}\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0,$$

$$\Rightarrow \lambda = n\bigg/\sum_{i=1}^{n} x_i \quad \Rightarrow \hat{\lambda} = 1/\bar{x},$$

$$\therefore \hat{p} = 1 - e^{-x/\bar{x}}.$$

注3*：由似然方程解不出θ的似然估计时，可由定义通过分析直接推求。事实上 $\hat{\theta}$ 满足

$$L(\hat{\theta}) = \sup_{\theta\in\Theta} L(\theta).$$

例8. 设x_1， ⋯，x_n为取自$U(0, \theta)$总体的样本，$\theta > 0$未知，求参数θ的最大似然估计。

解：

$$L(\theta) = \prod_{i=1}^{n} p(x_i;\theta) = \begin{cases} \dfrac{1}{\theta^n}, & 0 < x_i < \theta, \\ 0, & \text{otherwise}. \end{cases}$$

$$\ln L(\theta) = \ln \theta^{-n} = -n \ln \theta,$$

令 $\dfrac{\mathrm{d}[\ln L(\theta)]}{\mathrm{d}\theta} = -\dfrac{n}{\theta} = 0,$ 无解！

注意到 $L(\theta) = \begin{cases} \dfrac{1}{\theta^n}, & 0 < x_i < \theta, \\ 0, & \text{otherwise.} \end{cases}$

为使$L(\theta) \neq 0$，必须$0 < \max\{x_i\} < \theta$，故θ的值域为$(\max\{x_i\}, \infty)$，再由$L(\theta) = 1/\theta^n$关于θ单减，故θ越小，$L(\theta)$越大。于是$L(\max\{x_i\})=\sup L(\theta)$。

$$\therefore \hat{\theta} = \max\{x_i \mid i = 1, 2, \ldots, n\} = x_{(n)}.$$

点估计

矩估计法

最大似然估计法

基本步骤

基本步骤

估计量的评选标准

相合性

无偏性

有效性

一、相合性

定义：设$\hat{\theta}_n = \hat{\theta}(x_1, \ldots, x_n)$是$\theta$的估计量，

若$\hat{\theta}_n \xrightarrow{P} \theta$，则称$\hat{\theta}$是$\theta$的相合估计量。

例1.设$x_1, \ldots, x_n \overset{\text{i.i.d.}}{\sim} b(m, p)$, m已知，$0 < p < 1$,

讨论p的矩估计量的相合性。

解：由Chebyshev大数定律,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{P} \mathrm{E}X = mp.$$

$$\therefore \hat{p} = \bar{x}/m \xrightarrow{P} p.$$

故p的矩估计量是相合的。

二、无偏性

定义：设$\hat{\theta}=\hat{\theta}(x_1,\ldots,x_n)$为$\theta$的估计量，

若$E\hat{\theta}=\theta$，则称$\hat{\theta}$是θ的无偏估计量。

$$E\bar{x}=E\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n}\sum_{i=1}^{n}Ex_i=EX=\mu,$$

$$Es^2=E\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]=\frac{1}{n-1}E\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)$$

$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}Ex_i^2-nE\bar{x}^2\right)$$

$$=\frac{1}{n-1}\left[n(\mu^2+\sigma^2)-n(\mu^2+\sigma^2/n)\right]=\sigma^2.$$

例2. 设$x_1, x_2, \ldots, x_n \overset{\text{i.i.d.}}{\sim} U(0, \theta)$,

考察θ的矩估计和最大似然估计的无偏性。

解：θ的矩估计和最大似然估计分别为

$$\hat{\theta}_{\mathrm{M}} = 2\bar{x}, \quad \hat{\theta}_{\mathrm{MLE}} = x_{(n)}.$$

$$\mathrm{E}\hat{\theta}_{\mathrm{M}} = 2\mathrm{E}\bar{x} = 2 \times \mathrm{E}X = 2 \times \frac{\theta}{2} = \theta,$$

故θ的矩估计无偏，但最大似然估计是偏的，

注：若取 $\theta^* = \dfrac{n+1}{n}\hat{\theta}_{\mathrm{MLE}}$, 则θ*是θ的无偏估计。

三、有效性

定义：设$\hat{\theta}_1$和$\hat{\theta}_2$分别是参数θ的两个无偏估计，

若$\mathrm{Var}(\hat{\theta}_1) \le \mathrm{Var}(\hat{\theta}_2)$且存在$\theta \in \Theta$使得不等号

严格成立，则称$\hat{\theta}_1$比$\hat{\theta}_2$有效。

续例2. 已知$\hat{\theta}_M$与$\dfrac{n+1}{n}\hat{\theta}_{MLE}$都是$\theta$的无偏估计，

通过比较方差知后者比前者有效。

例3. 设$\bar{x}_1$和$\bar{x}_2$分别为取自总体X的容量为n_1和n_2的两个样本的样本均值，求证：对任意实数a，$b>0$，$a+b=1$，统计量$a\bar{x}_1+b\bar{x}_2$都是EX的无偏估计，并求a和b使所得统计量最有效。

解：$\mathrm{E}(a\bar{x}_1+b\bar{x}_2)=a\mathrm{E}\bar{x}_1+b\mathrm{E}\bar{x}_2=\mathrm{E}X,$

$$\mathrm{Var}(a\bar{x}_1+b\bar{x}_2)=a^2\mathrm{Var}(\bar{x}_1)+b^2\mathrm{Var}(\bar{x}_2)$$

$$=\left(\frac{a^2}{n_1}+\frac{b^2}{n_2}\right)\mathrm{Var}(X)\geq\frac{\mathrm{Var}(X)}{n_1+n_2},$$

$$\Rightarrow(a,b)=\left(\frac{n_1}{n_1+n_2},\frac{n_2}{n_1+n_2}\right).$$