(i) 白奴利分佈

假設 $X_1, X_2, \cdots, X_n$ 是一組從(母體)平均數 p，$0< p<1$，的白奴利隨機變數 X 取出的隨機樣本。則當樣本數 n 夠大($10\leq n\hat{p} \leq n-10$，$\hat{p}=\overline{X}$)時 $\hat{p} \approx N(p, p(1-p)/n)$，所以

$$P\left(\left|\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n}}\right|\leq z_{1-\alpha/2}\right)=$$

$P\left(\hat{p} - z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n} \leq p \leq \hat{p} + z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n}\right)$ $\approx 1 - \alpha$。因此實數區間

$[\hat{p} - z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n}, \hat{p} + z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n}]$是未知平均數 p 的一個近似信賴水準$100(1-\alpha)\%$ 的信賴區間。

若$100(1-\alpha)\% =95\%$，則 $z_{1-\alpha/2}=1.96$($\alpha=0.05$)，因此若要求 $z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})}/\sqrt{n} \leq 1.96/(2\sqrt{n})=3\%$，則 $n\geq 1068$(這就是你常常在媒體上看到抽樣人數約 1,000 人，正負誤差為百分之三的抽樣調查!)

例 假設年底台北市長選舉某候選人在 950 個表態公民中的支持度估計值為 0.35，則他真正支持度的近似 95%信賴區間為

$[0.35-1.96\frac{\sqrt{0.35(1-0.35)}}{\sqrt{950}}, 0.35+1.96\frac{\sqrt{0.35(1-0.35)}}{\sqrt{950}}]=[0.3197, 0.3803]$。

(ii) 其它分佈

假設 $X_1, X_2, \cdots, X_n$ 是一組從(母體)平均數 $-\infty< \mu <\infty$和(母體)變異數

$\sigma^2 > 0$ 的隨機變數 X取出的隨機樣本(μ 和 σ^2 均未知)。則當樣本數

n 夠大時 $\dfrac{\sqrt{n}(\overline{X}-\mu)}{\sigma}$ 近似 $N(0,1)$且 $(n\text{-}1)S^2/\sigma^2$近似 χ^2_{n-1}。

利用樣本標準變異 S來估計未知的σ，可得

$$P\left(\left|\frac{\overline{X}-\mu}{S/\sqrt{n}}\right| \leq z_{1-\alpha/2}\right)$$

$$= P(\overline{X} - z_{1-\alpha/2}S/\sqrt{n} \leq \mu \leq \overline{X} + z_{1-\alpha/2}S/\sqrt{n}) \approx 1 - \alpha。$$

所以信賴區間 $[\,\overline{x} - z_{1-\alpha/2}s/\sqrt{n},\ \overline{x} + z_{1-\alpha/2}s/\sqrt{n}]$是未知平均數 μ 的

一個近似信賴水準$100(1-\alpha)\%$ 的信賴區間。

同理，信賴區間 $\left[\dfrac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}, \dfrac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}\right]$ 是未知變異數 σ^2 的一個近

似信賴水準$100(1-\alpha)\%$ 的信賴區間。

例 假設 $X_1, X_2, \cdots, X_{30}$是一組從(母體)平均數 $\theta>0$ 的指數分佈隨機變

數 X取出的隨機樣本(θ未知)。若 $\overline{x} = 42$，s=12，則

[42-1.96(12)/ $\sqrt{30}$, 42+1.96(12)/ $\sqrt{30}$]=[37.706, 46.294]是θ的一個

近似 95%信賴區間。

假設檢定: 利用隨機樣本來對統計假設做拒絕或不拒絕的決策方

法。

“統計假設”是指對欲檢定的參數值的一種認定。分為虛無假設(null

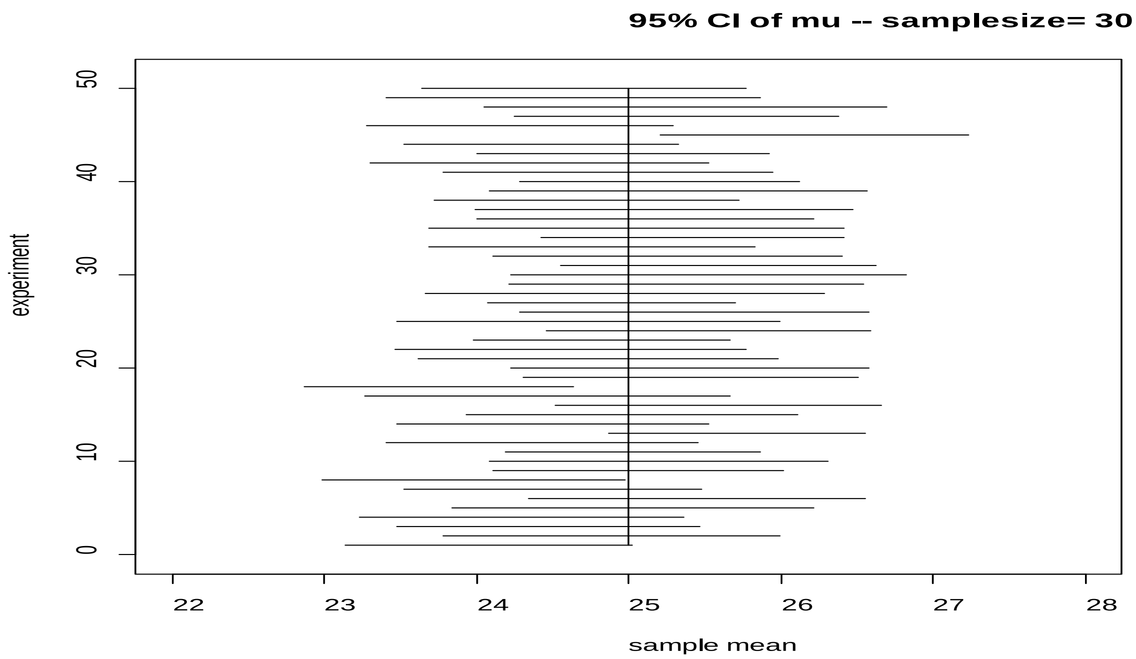
hypothesis)與對立假設(alternative hypothesis)，共有三種類型: 雙

例 1 假設 $X_1, X_2, \cdots, X_{30}$ 是一組從(母體)平均數 μ =25 和(母體)變異數 σ^2 =9 的常態分佈隨機變數 X 取出的隨機樣本。本例將繪出 50 組平均數 μ 的 95%信賴區間。

```
> norm.generateRsample=function(mu,sigma,Nsample,Nsize)
+ {
+ normRandomSample=matrix(0,Nsample,Nsize)
+ for(i in 1:Nsize)
+ {
+ Rnumber=rnorm(Nsample,mu,sigma)
+ normRandomSample[,i]=as.matrix(Rnumber)
+ }
+ return(normRandomSample)
+ }
>
> Normsample=norm.generateRsample(25,3,50,30)
> dim(Normsample)
[1] 50 30
> round(mean(Normsample),4) #計算矩陣所有元素的總樣本平均值
[1] 24.9563
> round(var(as.vector(Normsample)),4) #計算矩陣所有元素的總樣本變異數，
[1] 9.0989
>              #需先將矩陣變成向量才能算，否則是計算樣本共變異數矩陣
```

```
> mu=25
> Nsample=50
> normCI=matrix(0,Nsample,4)
> colnames(normCI)=c("mean","SD","LL","UL")
> normCI[,1]=apply(Normsample,1,mean) #計算每一列的樣本平均值
> normCI[,2]=sqrt(apply(Normsample,1,var)) #計算每一列的樣本標準差
> normCI[,3]=normCI[,1]-qnorm(0.975)*normCI[,2]/sqrt(Nsize)
> #計算每一組樣本的LL
> normCI[,4]=normCI[,1]+qnorm(0.975)*normCI[,2]/sqrt(Nsize)
> #計算每一組樣本的UL
> xx1=seq(normCI[1,3],normCI[1,4],length=100) #先畫第1組樣本的信賴區間
> yy1=rep(1,100) #1重複100次的數列
> plot(xx1,yy1,type="l",xlim=c(22,28),ylim=c(0,Nsample+1),xlab="sample mean",
+ ylab="experiment")
> title(paste("95% CI of mu -- samplesize=",Nsize))
> for(i in 2:Nsample) #以下程式為畫第2到50組樣本的信賴區間
+ {
+ xx=seq(normCI[i,3],normCI[i,4],length=100)
+ yy=rep(i,100) #i重複100次的數列
+ lines(xx,yy)
+ }
```

```
> Xth=rep(mu,Nsample) #重複50次 mu 的數列
> Yth=1:Nsample #y=1,2,...,50
> lines(Xth,Yth) #加上mu=25的垂直線
```



例 2 假設 $X_1, X_2, \cdots, X_{30}$ 是一組從(母體)平均數 θ =5 的指數分佈隨機變數 X 取出的隨機樣本。本例將繪出 50 組變異數 θ^2 的近似 90%信賴區間。

```
> exp.generateRsample=function(lambda,Nsample,Nsize)
+ {
+ expRandomSample=matrix(0,Nsample,Nsize)
+ for(i in 1:Nsize)
+ {
+ Rnumber=rexp(Nsample,lambda)
+ expRandomSample[,i]=as.matrix(Rnumber)
+ }
+ return(expRandomSample)
+ }
>
> theta=5
> lambda=1/theta
> Expsample=exp.generateRsample(lambda,50,30)
```

```
> Nsample=50
> Nsize=30
> expCI=matrix(0,Nsample,3)
> colnames(expCI)=c("VAR","LL","UL")
> expCI[,1]=apply(Expsample,1,var) #計算每一列的樣本變異數
> expCI[,2]=(Nsize-1)*expCI[,1]/qchisq(0.95,(Nsize-1)) #計算每一組樣本的LL
> expCI[,3]=(Nsize-1)*expCI[,1]/qchisq(0.05,(Nsize-1)) #計算每一組樣本的UL
```

```
> xx1=seq(expCI[1,2],expCI[1,3],length=100) #先畫第1組樣本的信賴區間
> yy1=rep(1,100) #1重複100次的數列
> plot(xx1,yy1,type="l",xlim=c(0,130),ylim=c(0,Nsample+1),xlab="sample
+ variance",ylab="experiment")
> title(paste("90% CI of var -- samplesize=",Nsize))
> for(i in 2:Nsample) #以下程式為畫第2到50組樣本的信賴區間
+ {
+ xx=seq(expCI[i,2],expCI[i,3],length=100)
+ yy=rep(i,100) #i重複100次的數列
+ lines(xx,yy)
+ }
> Xth=rep(theta^2,Nsample) #重複50次 theta^2=100的數列
> Yth=1:Nsample #y=1,2,...,50
> lines(Xth,Yth) #加上theta^2=100的垂直線
```