Probabilidad y Estadística Matemática

Morfín Chávez Guillermo

1.

Considere $X_1, ..., X_n \sim Ber(p)$. Sean $L(X_1, ..., X_n) = \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ y $U(X_1, ..., X_n) = \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Muestre que (L, U) es un intervalo aproximado de confianza $1-\alpha$ para p. Donde $\hat{p}$ es la proporción de éxistos en una muestra aleatoria de tamaño n (siempre que n sea grande)

Supongamos que $X_1, .., X_n \overset{iid}{\sim} Ber(p)$, además nos basta saber que el estimador (o proporción) para la distribución *Bernoulli* está dado por $\hat{p} = \bar{X}$ y que el error estandar estimado esta dado por $\hat{\sigma}_{\hat{p}} = \sqrt{\bar{X}(1-\bar{X})/n}$.
Por la *Ley de Los Grandes Números* sucede que $\hat{\sigma}_{\hat{p}} \to \sigma_{\hat{p}}$ cuando $n \to \infty$ (para nuestra aproximación nos basta que n sea lo suficientemente grande) De modo que, definimos una variable aleatoria Z tal que:

$$\varphi = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} = \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1-p)}} = \frac{\hat{p} - p}{\sigma_{\hat{p}}}.$$

Y así, por el *Teorema Central del Límite* (bajo la misma condición sobre n) tenemos que

$$\varphi \overset{aprox.}{\sim} N(0,1)$$

De modo que:

$$\begin{aligned}
P(L < p < U) &= \left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \\
&= (-z_{\alpha/2}\hat{\sigma}_{\hat{p}} < p - \hat{p} < z_{\alpha/2}\hat{\sigma}_{\hat{p}}) \\
&= \left(-z_{\alpha/2} < \frac{\hat{p} - p}{\hat{\sigma}_{\hat{p}}} < z_{\alpha/2}\right) \\
&\approx \left(-z_{\alpha/2} < \frac{\hat{p} - p}{\sigma_{\hat{p}}} < z_{\alpha/2}\right) \\
&= P(\varphi < z_{\alpha/2}) - P(\varphi \leq -z_{\alpha/2}) \\
&= \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) \\
&= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.
\end{aligned}$$

■

2.

Escriba un programa que genere 80 números aleatorios con distribución Normal, $N(\mu = 7, \sigma = 2)$, y construya un intervalo de confianza al 95 % para μ. Repita el procedimiento anterior 1000 veces y registre los casos en que el intervalo cubre al valor 7. Verifique la cobertura del intervalo(al menos el 95 % de las veces el intervalo cubre el valor 7). Comente.

```
L = c()
U = c()
for(i in  1:1000)
{
  z <- qnorm(1-0.05/2)
  x <- rnorm(80,7,2)
  mean <- mean(x)
  S <- sqrt(sum((x-mean)^2)/80) #S<-sd(x)_para_aproximar_de_forma_sencilla
  L[i] <- mean-(z*S)/sqrt(80)
  U[i] <- mean+(z*S)/sqrt(80)
}
cbind(L, U) #para_visualizar_los_intervalos_
l<-L[L<=7]
u<-U[U>=7]
length(l)
length(u)
[1] 974 #resultado_para_L_que_satisfacen_el_intervalo
[2] 978 #resultado_para_U_que_satisfacen_el_intervalo
```

Debe resultar evidente el siguiente hecho: Dado que estamos simulando 1000 veces, para que se cumpla que al menos el 95 % de las veces el intervalo cubre $\mu = 7$ debe suceder que 950 de las ocasiones, como mínimo, el intervalo satisface $L \leq 7$ y a la vez $U \geq 7$, Lo que se verifica con la simulación. Nótese que al simular se obtiene que 26 de los mil casos no cumplen con $L \leq 7$ mientras que 22 de los casos no cumplen con $U \geq 7$. Además sabemos que para una misma iteración no sucede que $L_i > 7 > U_i$ (esto es claro ya que si $L_i = \bar{X} - (z * S_n)/\sqrt{80} > 7$ y $z, S_n > 0$, necesariamente $U_i > 7$ y viceversa, si $U_j < 7$ necesariamente $L_j < 7$). Lo anterior es útil para concluir que el total de casos para los que no se cumple el intervalo de confianza es $26 + 22 = 48$, lo que justifica la cobertura del intervalo.

3.

Bajo condiciones normales, el número promedio de llamadas personales que llegan a un conmutador en una compañía fue de 7.2 por hora. El gerente envió una carta a todos sus empleados pidiéndoles que el número de llamadas personales disminuyera.

3.1.

Sea X el número de llamadas registradas por periodo, ¿Que distribución de probabilidad se podría proponer para X? Durante periodos de 1 horas, los números de llamadas personales fueron:
1 4 5 5 6 8 5 8 10 11 8 5 8 13 11 4 11 7 14 7 4 10 6 5 3 8 6 5 6 7 4 8 10 7 13 10 8 8 5 5 8 3 7 7 8 12 5 5 7 5 6 10 9 9 8 13 5 10 7 7 4 5 8 7 8 6 12 9 8 6 6 5 6 3 7 6 6 8 9 9 6 7 11 11 7 8 7 7 10 6 5 6 6 4 7 4 11 6 5 7.

Es necesario notar que la muestra de datos presenta únicamente valores enteros, de modo que sería lógico tomar una distribución de probabilidad discreta. Si asumimos que el número de las llamadas que entran al conmutador son independientes podemos proponer una distribución binomial de parámetros p y m. Por tanto proponemos que $X \sim B(m, p)$.

3.2.

Para la distribución encontrada en 3.1., estime el o los valores de los parámetros.

Sabemos que los estimadores para la distribución binomial son:

$$\hat{m} = \frac{\bar{X}^2}{\bar{X} - S_{n-1}^2}; \qquad \hat{p} = \frac{\bar{X} - S_{n-1}^2}{\bar{X}}$$

De modod que:

```
x=c(1,4,5,5,6,8,5,8,10,11,8,5,8,13,11,4,11,7,14,7,4,10,6,5,3,8,6,5,6,7,4,8,
10,7,13,10,8,8,5,5,8,3,7,7,8,12,5,5,7,5,6,10,9,9,8,13,5,10,7,7,4,5,8,7,8,6,
12,9,8,6,6,5,6,3,7,6,6,8,9,9,6,7,11,11,7,8,7,7,10,6,5,6,6,4,7,4,11,6,5,7)
table(x) #para visualizar las frecuencias
y=mean(x)
S=sum((x-y)^2)/(100-1)
hat_m=(y^2)/(y-S)
[1]59.221405
hat_p=(y-S)/(y)
[2]0.12140880
```

3.3.

Use los valores estimados de los parámetros para estimar $P(X < 7.2)$ y $P(X > 7.2)$.

Para hacer el calculo podemos redondear el estimador $\hat{m}$ tal que $\hat{m} = 59$, así:

```
pbinom(7.2,59,hat_p)
[1] 0.5730051
pbinom(7.2,59,hat_p,lower.tail = FALSE)
[2] 0.4269949
```

3.4.

Use TCL para proponer la distribución de $\bar{X}$ y calcular$P(\bar{X} < 7.2)$.

Por el teorema central del límite sabemos que si tenemos una muestra aleatorias de n observaciones de una población con media μ y desviación estándar σ, entonces, cuando n es grande, la distribución muestral de $\bar{X}$ tendrá aproximadamente una distribución normal con una media igual a μ y una desviación estándar de $\sigma/\sqrt{n}$; por lo tanto:

$$\bar{X} \overset{aprox.}{\sim} N\left(\hat{m}\hat{p}, \sqrt{\frac{\hat{m}\hat{p}(1-\hat{p})}{n}}\right)$$

Y así

```
mu=hat_m*hat_p
sigma=sqrt((mu*(1-hat_p))/100)
pnorm(7.2,mu,sigma)
[1] 0.5158686
```

3.5.

Construya un intervalo de confianza al 99 % para el parámetro.

Suponemos que el parametro a estimar es μ; por lo tanto nuestro intervalo de confianza es:

$$\bar{X} \pm z_{\alpha/2}\frac{S_n}{n}$$

Sabemos que el nivel de confianza es $1 - \alpha = 0.99 \Rightarrow \alpha = 0.01$, así:

```
L=mean(x)-(qnorm(1-0.01/2)*(sd(x)/10))
U=mean(x)+(qnorm(1-0.01/2)*(sd(x)/10))
[1] 6.5426
[2] 7.8374
```

Por lo tanto el intervalo de confianza es: $(6.5426, 7.8374)$

3.6.

¿Qué podría concluir el gerente?

La conclusión más pertinente que podría hacerse, como resultado del análisis mostrado anteriormente, es que los empleados de la compañía no disminuyeron significativamente el numero de llamadas personales; más aun se mantuvieron prácticamente iguales.