

Cheat sheet Práctica 4: Cálculo de Probabilidades

Juan Antonio Villatoro García – Grupo BioestadísticaR



**UNIVERSIDAD
DE GRANADA**

Todo el material para el conjunto de actividades de este curso ha sido elaborado y es propiedad intelectual del grupo BioestadísticaR formado por:

Antonio Martín Andrés

Juan de Dios Luna del Castillo,

Pedro Femia Marzo,

Miguel Ángel Montero Alonso,

Christian José Acal González,

Pedro María Carmona Sáez,

Juan Manuel Melchor Rodríguez,

José Luis Romero Béjar,

Manuela Expósito Ruíz,

Juan Antonio Villatoro García.

Todos los integrantes del grupo han participado en todas las actividades, en su elección, construcción, correcciones o en su edición final, no obstante, en cada una de ellas, aparecerán uno o más nombres correspondientes a las personas que han tenido la máxima responsabilidad de su elaboración junto al grupo de BioestadísticaR.

Todos los materiales están protegidos por la Licencia Creative Commons CC BY-NC-ND que permite "descargar las obras y compartirlas con otras personas, siempre que se reconozca su autoría, pero no se pueden cambiar de ninguna manera ni se pueden utilizar comercialmente".

Práctica 5: : Intervalos de confianza

Intervalos de confianza para la media μ de una variable normal cuando la varianza σ^2 es conocida

No hay una función de R que permita obtenerlo. Hay que calcularlo aplicando la fórmula:

$$\mu \in \bar{x} \pm z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

n: número de observaciones de la variable de interés
 α : nivel de significación.
 Z_{α} : valor correspondiente en la tabla de la $N(0,1)$



```
> pacientes=osteo[1:10,] #se selecciona a los 10 primeros pacientes
> sigma=sqrt(5.80) #desviación típica conocida
> media.muestral=mean(pacientes$imc) #estimación puntual de la media muestral
> n=length(pacientes$imc) #tamaño de la muestra
> z.alfa=1.96 #mirando la tabla de la N(0,1) y tomando alfa=0.05
> limite.inferior=media.muestral-z.alfa*sigma/sqrt(n)
> limite.superior=media.muestral+z.alfa*sigma/sqrt(n)
> intervalo=c(limite.inferior,limite.superior)
> intervalo
[1] 21.29450 24.27988
```

Intervalos de confianza para la media μ de una variable normal cuando la varianza σ^2 es desconocida

Función icm() de la librería BioestadísticaR

Caso general para una variable continua

icm(x, conf = 0.95, decs = 4)

x: variable de interés
conf: nivel de confianza
decs: precisión (número) de los decimales

```
> icm(osteo$imc, conf=0.99, decs=6)
```

Intervalo de confianza bilateral para la media de una VA normal

 Información muestral:

n: 94
 m: 23.92065
 s: 3.748202
 sem: 0.386598

Estimación:
 99 %-IC(μ): (22.904 , 24.9373)
 Precisión: 1.016648

Caso en el que la información muestral se presenta de forma resumida

icm(n, m, s, conf = 0,95, decs = 4)

n: tamaño de la muestra
m: media
s: desviación típica

```
> icm(n=250, m=187, s=25.4)
```

Intervalo de confianza bilateral para la media de una VA normal

 Información muestral:

n: 250
 m: 187
 s: 25.4
 sem: 1.6064

Estimación:
 95 %-IC(μ): (183.8361 , 190.1639)
 Precisión: 3.1639

Intervalos para variables discretas

Para variables discretas que se pueden aproximar a una normal, se puede hacer corrección por continuidad, añadiendo el parámetro **vac = FALSE**

imc(x, conf = 0.95, decs = 4, vac = FALSE)

```
> icm(c(1,4,3,2,5,6,5,4,3,2,6,9), vac=F)
```

Intervalo de confianza bilateral para la media de una VA normal

 Información muestral:

n: 12
 m: 4.1667
 s: 2.2088
 sem: 0.6376

Estimación:
 95 %-IC(μ): (2.7216 , 5.6117)
 Precisión: 1.4451

Práctica 5: : Intervalos de confianza

Determinar el tamaño de muestra para calcular la media en una variable normal

Varianza conocida: función `simple.size.mean()` de la librería `sampling book`

`sample.size.mean(e, s, level=0.95)`

- e:** precisión (d)
- s:** desviación típica
- level:** nivel de confianza

`> sample.size.mean(2, 25.4, level=0.95)`

sample.size.mean object: Sample size for mean estimate
Without finite population correction: N=Inf, precision
e=2 and standard deviation S=25.4

Sample size needed: 620

Varianza desconocida: función `nm()` de la librería `BioestadisticaR`

Opción 1 : `nm(n,m,s,d, conf)`

- n:** tamaño de la muestra
- m:** media en la muestra
- s:** desviación típica
- d:** precisión
- conf:** nivel de confianza

`> nm(n=250, m=187, s=25.4, d=2)`

Tamaño de muestra para la estimación de una media de una VA normal

Muestra piloto:

- n: 250
- m: 187
- s: 25.4
- sem: 1.6064
- precisión observada: 3.1639

Precisión deseada: 2
Tamaño muestral necesario: 626

Opción 2 : `nm(x, conf, d)`

x: vector con los valores de la muestra piloto

- conf:** nivel de confianza
- d:** precisión que se desea obtener

`> nm(osteosimc, d=0.5, conf = 0.99)`

Tamaño de muestra para la estimación de una media de una VA normal

Muestra piloto:

- n: 94
- m: 23.9206
- s: 3.7482
- sem: 0.3866
- precisión observada: 1.0166

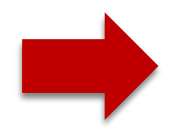
Precisión deseada: 0.5
Tamaño muestral necesario: 389

Intervalo de confianza para una proporción

Función `binom.confint()` de la librería `binom`

`binom.confint(x, n, conf.level=0.95, methods="all")`

- x:** número de éxitos
- n:** número de repeticiones del experimento
- conf.level:** nivel de confianza
- methods:** métodos a aplicar para obtener el intervalo (se puede dejar "all" y los obtiene todos)



```
> library(binom)
> binom.confint(56,458, conf.level=0.95, methods="all")
  method x  n  mean  lower  upper
1  agresti-coull 56 458 0.1222707 0.09520793 0.1556172
2  asymptotic 56 458 0.1222707 0.09226828 0.1522732
3  bayes 56 458 0.1230937 0.09365364 0.1534592
4  cloglog 56 458 0.1222707 0.09425095 0.1541354
5  exact 56 458 0.1222707 0.09369823 0.1558268
6  logit 56 458 0.1222707 0.09529258 0.1555733
7  probit 56 458 0.1222707 0.09480606 0.1548878
8  profile 56 458 0.1222707 0.09440081 0.1543536
9  lrt 56 458 0.1222707 0.09439903 0.1543533
10 prop.test 56 458 0.1222707 0.09440139 0.1566502
11 wilson 56 458 0.1222707 0.09537041 0.1554547
```

Práctica 5: : Intervalos de confianza



UNIVERSIDAD
DE GRANADA

Intervalo de confianza para una proporción (continuación)

Función icp() de la librería BioestadísticaR

Opción 1: `icp(x, n, conf=0.95)` → Indicando el número de éxitos

x: número de éxitos

n: número de repeticiones del experimento

conf: nivel de confianza

```
> icp(x=56, n = 458, conf = 0.90)
```

Intervalo de confianza para una proporción binomial

Información muestral:

n: 458

x: 56

Estimación puntual: p= 0.1223 , q=(1-p)= 0.8777

Método de Wilson:

95 %-IC(π): (0.0983 , 0.1509)

Semiapertura: 0.0263

Método de Wald:

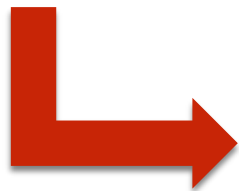
95 %-IC(π): (0.096 , 0.1485)

Precisión: 0.0263

Método de Wald ajustado:

95 %-IC(π): (0.1002 , 0.1509)

Precisión: 0.0254



Opción 2: `icp(factor,level)` → A partir de una variable factor

factor: vector de tipo factor

level: elemento del factor de referencia

conf: nivel de confianza

```
> icp(osteo$sexo, level="Hombre")
```

Intervalo de confianza para una proporción binomial

Información muestral:

n: 94

x: 45

Estimación puntual: p= 0.4787 , q=(1-p)= 0.5213

Método de Wilson:

95 %-IC(π): (0.3755 , 0.5837)

Semiapertura: 0.1041

Método de Wald:

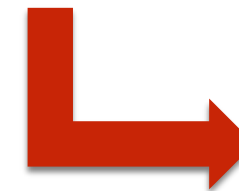
95 %-IC(π): (0.3724 , 0.585)

Precisión: 0.1063

Método de Wald ajustado:

95 %-IC(π): (0.3807 , 0.5785)

Precisión: 0.0989



Determinación del tamaño de muestra necesario para estimar una proporción

Función np() de la librería BioestadísticaR

Opción 1: `np(x, n, d)` → Número de éxitos de una muestra piloto

x: número de éxitos

n: número de repeticiones del experimento

d: precisión deseada

```
> np(x=15, n=90, d=0.03)
```

Tamaño de muestra para estimar una proporción binomial

Información muestral

n: 90

x: 15

Inferencia basada en el método de Wald ajustado:

95 %-IC(π): (0.103 , 0.2587)

precisión observada: 0.0778 (7.78 %)

Tamaño muestral requerido para d= 0.03 (3 %), conf.= 95 %

- Basado en la muestra piloto (p_o= 0.2587): n= 819

- Sin considerar la información previa: n= 1068

Opción 2: `np(factor, level, d)` → A partir de una variable factor

factor: vector de tipo factor

level: elemento del factor de referencia

d: precisión deseada

```
> np(osteo$sexo, level="Mujer", d=0.03)
```

Tamaño de muestra para estimar una proporción binomial

Información muestral

n: 94

x: 49

Inferencia basada en el método de Wald ajustado:

95 %-IC(π): (0.4215 , 0.6193)

precisión observada: 0.0989 (9.89 %)

Tamaño muestral requerido para d= 0.03 (3 %), conf.= 95 %

La muestra piloto es compatible con p= 0.5 : n= 1068