Apuntes Estadistica

DESKTOP-OCEEGEC

14 de mayo de 2025

Índice

1. Inferencia Estadistica

Estimador: $\hat{\theta}(X_1, \ldots, X_n)$ es una funcion de las v.a $X_1, \ldots, X_n$ $\overline{X} = \frac{X_1, \ldots, X_n}{n}$, y $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$

Estimacion: $\hat{\theta}$ es un valor que se obtiene cuando usamos la informacion de el estimador. En el caso de conocer σ^2 en una m.a de $N(\mu, \sigma^2)$, podemos usar los siguientes estimadores de μ

$$\hat{\mu_1} = \frac{X_1 + 2X_2 + \cdots + nX_n}{n^2} \qquad \hat{\mu_2} = \frac{X_1 + X_n}{2}$$
$$\hat{\mu_3} = \frac{X_1 + X_2 + \cdots + X_n}{n(n-1)} \qquad \hat{\mu_4} = \frac{X_1 + \cdots + X_n}{n}$$

Donde μ_1 subestima el valo de μ, μ_2 es insesgado de μ, μ_3 es un estimador sesgado de μ y μ_4 es insesgado de μ. Concluimos que μ_2 y μ_4 son estimadores insesgados de μ.

1.1. Insesgamiento

Un estimador es insesgado cuando su media coincide con el valor del parametro a estimar, osea, $E(\hat{\theta}(X_1, \ldots, X_n)) = \theta$, lo que quiere decir que $\sum_{i=1}^{n} \frac{\hat{\theta}_i}{n} = \theta$ es ser insesgado.

1.1.1. Ejemplo 1

Example 1 (Ex. 4): Let X and Y denote the strength of concrete beams and cylinders. The following data are obtained

X : 5.9, 7.2, 7.3, 6.3, 8.1, 6.8, 7.0, 7.6, 6.8, 6.5, 7.0, 6.3, 7.9, 9.0, 8.2, 8.7, 7.8, 9.7, 7.4, 7.7, 9.7, 7.8, 7.7, 11.6, 11.3, 11.8, 10.7.

Y : 6.1, 5.8, 7.8, 7.1, 7.2, 9.2, 6.6, 8.3, 7.0, 8.3, 7.8, 8.1, 7.4, 8.5, 8.9, 9.8, 9.7, 14.1, 12.6, 11.2.

Suppose $E(X) = \mu_1$, $V(X) = \sigma_1^2$; $E(Y) = \mu_2$, $V(Y) = \sigma_2^2$.

(a) Show that $\overline{X} - \overline{Y}$ is an unbiased estimator of $\mu_1 - \mu_2$. Calculate it for the given data.

1. $E(\overline{X} - \overline{Y}) = E(\overline{X}) - E(\overline{Y}) = \mu_1 - \mu_2$, entonces el estimador insesgado seria

$$\overline{x} - \overline{y} = 8{,}141 - 8{,}575 = 0{,}434$$

1.1.2. Ejemplo 2

Example 2 (Ex 8): In a random sample of 80 components of a certain type, 12 are found to be defective.
(a) Give a point estimate of the proportion of all not-defective units.
(b) A system is to be constructed by randomly selecting two of these components and connecting them in series. Estimate the proportion of all such systems that work properly.

1. $\hat{p} = \frac{80-12}{80} = 0{,}85$

1.1.3. Propiedades de Esperanza

1. $E(c) = c$
2. $E(cX) = cE(x)$
3. $E(X + c) = E(x) + c$
4. $E(aX + b) = aE(x) + b$
5. $E(X \pm Y) = E(X) \pm E(Y)$
6. Si X e Y son independientes, $E(X \cdot Y) = E(X) \cdot E(Y)$

1.2. Eficiencia Relativa

Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ estimadores insesgados de θ. Decimos que $\hat{\theta}_1$ es mas eficiente que $\hat{\theta}_2$ si $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$

1.3. Consistencia

Sea $\hat{\theta}$ un estimador insesgado de θ, es consistente con respecto a θ si

$$\lim_{n\to\infty} \mathrm{Var}(\hat{\theta}(X_1, \ldots, X_n)) = 0$$

, a medida que n se acerca al infinito, $P(\hat{\theta})$ es 1.

2. Estimacion Puntual

2.1. Resultados muestrales

1. $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$
2. $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim N(\mu, \frac{\sigma^2}{n})$
3. $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \sim \chi^2(n-1)$
4. $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$, se usa cuando σ^2 es desconocida

2.2. Estimadores

Estimadores que vamos a usar

- $\overline{x} = \sum_{i=1}^{n} x_i / n$
- $S^2 = \sum_{i=1}^{n}(x_i - \overline{x})/n - 1$
- $p = \frac{\text{numero de exitos}}{n}$ proporcion muestral

2.3. Distribucion Normal

Datos de una distribucion normal $N(\mu,\sigma)$ μ es la media, σ es la desviacion estandar Si tenemos $X_1,\dots,X_n$ m.a de $N(\mu,\sigma^2)$, entonces

1. $\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$
2. $\overline{X} \sim N(\mu, \sigma^2/n)$
3. $S^2 \sim \chi^2(n-1)$
4. Resultado 1: Si $\overline{x} = \sum_{i=1}^n x_i/n \sim N(\mu, \frac{\sigma^2}{n})$, lo usamos para hablar sobre μ donde σ^2 es conocido.
5. Resultado 2: Tenemos que $\frac{\overline{x}-\mu}{\frac{s}{\sqrt{n}}} = \sqrt{n}\frac{\overline{x}-\mu}{S} \sim \underbrace{t(n-1)}_{\text{grado de libertad}}$. Y lo usamos cuando σ^2 es desconocido. En R usamos

```
qt(area, g.l) # El percentil
pt(x, g.l) # el area acumulada
```

6. Resultado 3: Si $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, entonces lo usamos para hablar de σ^2, cuando este es el parametro de interes.

   En R usamos `qchisq(area,g.l)` y `pchisq(x,g.l)`

2.4. Distribucion Bernoulli

Con la distribucion de Bernoulli, tenemos resultados de variables $X \in \{0,1\}$, que forman una proporcion. Si tenemos $X_1,\dots,X_n$ m.a de Bernoulli(π), entonces: $E(X_i) = \pi$ y $Var(X_i) = \pi\cdot(1-\pi)$ p, proporcion muestral $= \frac{\sum_{i=1}^n X_i}{n}$

2.5. Teorema Central del Limite

Sean $X_1,\dots,X_n$ m.a F con $E(X_i) = \mu$ y $\mathrm{Var}(X_i) = \sigma^2$, entonces: $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N_{n\to\infty}(0,1)$, si es que n es lo suficientemente grande.

Asi, si $X_1,\dots,X_n$ m.a Bernoulli(π) con $E(X_i) = \pi$ y $\mathrm{Var}(X_i) = \pi(1-\pi)$, $p = \sum_{i=1}^n X_i/n$ tiene la siguiente distribucion: $\frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim aN(0,1)$, tambien tenemos que $\frac{p-\pi}{\sqrt{\frac{p(1-p)}{n}}} \sim aN(0,1)$

3. Intervalos de confianza

3.1. Trabajando con dos poblaciones

Sean $X_1,\dots,X_n$ m.a $N(\mu_1,\sigma_1^2)$ y $Y_1,\dots,Y_n$ m.a $N(\mu_2,\sigma_2^2)$ que son independientes

1. $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \sim N(0,1)$ Las poblaciones son conocidas
2. $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$, donde σ_1^2 y σ_2^2 son desconocidos pero iguales. Y $S^2p = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2}$
3. $\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{Sp\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(\eta)$, donde σ_1^2 y σ_2^2 son desconocidos y diferentes. donde $\gamma = \frac{(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2})^2}{\frac{(S_1^2/n_1)^2}{n_1-1}+\frac{(S_2^2/n_2)^2}{n_2-1}}$
4. $\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$

3.2. Para μ con σ o σ^2 conocido

```
# DATOS DE ENTRADA:
#
#   n TAMAÑO DE LA MUESTRA
#   prom PROMEDIO MUESTRAL
#   sigma DESVIACIÓN ESTÁNDAR POBLACIONAL
#   nc NIVEL DE CONFIANZA

# IC para mu (media poblacional) con sigma, sigma2 conocido
ICMU1=function(n,prom,sigma,nc=0.95)
  {
    z=qnorm(1-(1-nc)/2)
    EE=z*sigma/sqrt(n)
    LI=prom-EE
    LS=prom+EE
    cat("Un intervalo de confianza para MU","\n",
         "al nivel de confianza",nc*100,"%", "es:","\n",
    "Limite inferior:",LI,"Limite superior:",LS,"\n")
  }
```

3.3. Para μ con σ o σ^2 desconocido

```
# DATOS DE ENTRADA:
#
#   n TAMAÑO DE LA MUESTRA
#   prom PROMEDIO MUESTRAL
#   s DESVIACIÓN ESTÁNDAR MUESTRAL
#   nc NIVEL DE CONFIANZA

# IC para mu con sigma, sigma2 desconocido
ICMU2=function(n,prom,s,nc=0.95)
  {
    t=qt(1-(1-nc)/2,n-1)
    EE=t*s/sqrt(n)
    LI=prom-EE
    LS=prom+EE
    cat("Un intervalo de confianza para MU","\n",
         "al nivel de confianza",nc*100,"%", "es:","\n",
    "Limite inferior:",LI,"Limite superior:",LS,"\n")
  }
```

3.4. Para σ^2 (varianza poblacional)

```
# DATOS DE ENTRADA:
#
#   n TAMAÑO DE LA MUESTRA
#   s DESVIACIÓN ESTÁNDAR MUESTRAL
#   nc NIVEL DE CONFIANZA

# IC para sigma2 (varianza poblacional)
ICSIGMA2=function(n,s,nc=0.95)
  {
    v1=qchisq((1-nc)/2,n-1)
    v2=qchisq(1-(1-nc)/2,n-1)
    LI=(n-1)*s^2/v2
    LS=(n-1)*s^2/v1
    cat("Un intervalo de confianza para SIGMA^2","\n",
```

```
        "al nivel de confianza",nc*100,"%", "es:","\n",
   "Limite inferior:",LI,"Limite superior:",LS,"\n")
  }
```

3.5. Para π, proporcion poblacional

```
# DATOS DE ENTRADA:
#
#   n TAMAnO DE LA MUESTRA
#   p PROPORCIoN MUESTRAL
#   nc NIVEL DE CONFIANZA

# IC para pi (proporcion poblacional)
ICPI=function(n,p,nc=0.95)
  {
    z=qnorm(1-(1-nc)/2)
    EE=z*sqrt(p*(1-p)/n)
    LI=p-EE
    LS=p+EE
    cat("Un intervalo de confianza para PI","\n",
         "al nivel de confianza",nc*100,"%", "es:","\n",
    "Limite inferior:",LI,"Limite superior:",LS,"\n")
  }
```

3.6. Tamano de la muestra para una media poblacion μ

```
# SE NECESITA:
#     EE ERROR DE ESTIMACIoN
#     nc NIVEL DE CONFIANZA
#     sigma DESVIACIoN ESTaNDAR
#

# tamano de la muestra para media poblacional mu
NMU=function(EE,nc,sigma)
  {
    z=qnorm(1-(1-nc)/2)
    n=round((z/EE)^2*sigma^2,0)
    cat("El tamano de la muestra para estimar MU","\n",
         "al nivel de confianza",nc*100,"%","con un error
 de estimacion de",EE,", es:",n,"\n")
  }
```

3.7. Tamano de la muestra para una proporcion poblacional μ

```
# SE NECESITA:
#     EE ERROR DE ESTIMACIoN
#     nc NIVEL DE CONFIANZA
#     p PROPORCIoN CONOCIDA
#

# tamano de muestra para proporcion poblacional
NPI=function(EE,nc,p)
  {
    z=qnorm(1-(1-nc)/2)
    n1=round((z/EE)^2*p*(1-p),0)
    n2=round(1/4*(z/EE)^2,0)
    cat("El tamano de la muestra para estimar PI","\n",
         "al nivel de confianza",nc*100,"%","con un error
```

```
 de estimacion de",EE,", es:",n1,"\n")
    cat("El tamano de la muestra para estimar PI","\n",
         "al nivel de confianza",nc*100,"%","con un error
 de estimacion de",EE,",en el peor de los casos, es:",n2,"\n")
  }
```

3.8. Para dos poblaciones, diferencia de medias $\mu_1 - \mu_2$

3.8.1. Caso 1: Varianzas poblacionales conocidas

```
# DATOS:
# n1, n2 TAMAÑOS DE LAS MUESTRAS
# prom1, prom2 PROMEDIOS MUESTRALES
# sigma1, sigma2 DESVIACIONES POBLACIONALES
# nc NIVEL DE CONFIANZA (PROPORCIÓN)

# Diferencia de medias
## Caso1. IC para dos poblaciones, con varianzas poblacionales conocidas
difmed1=function(n1,n2,prom1,prom2,sigma1,sigma2,nc=0.95)
{
  z=qnorm(1-(1-nc)/2)
  EE=z*sqrt(sigma1^2/n1+sigma2^2/n2)
  LI= prom1-prom2-EE
  LS=prom1-prom2+EE
  cat("Un intervalo de confianza para MU1-MU2","\n",
         "al nivel de confianza",nc*100,"%", "es:","\n",
  "Limite inferior:",LI,"Limite superior:",LS,"\n")
}
```

3.8.2. Caso 2: Varianzas poblacionales desconocidas pero iguales

```
# DATOS:
# n1, n2 TAMAÑOS DE LAS MUESTRAS
# prom1, prom2 PROMEDIOS MUESTRALES
# s1, s2 DESVIACIONES ESTANDAR MUESTRALES
# nc NIVEL DE CONFIANZA (PROPORCIÓN)
#

# Diferencia de medias
## Caso2. IC para dos poblaciones, con varianzas poblacionales desconocidas pero iguales
difmed2=function(n1,n2,prom1,prom2,s1,s2,nc=0.95)
     {
        t=qt(1-(1-nc)/2,n1+n2-2)
  sp=sqrt(((n1-1)*s1^2+(n2-1)*s2^2)/(n1+n2-2))
  EE=t*sp*sqrt(1/n1+1/n2)
  LI= prom1-prom2-EE
  LS=prom1-prom2+EE
  cat("Un intervalo de confianza para MU1-MU2","\n",
         "al nivel de confianza",nc*100,"%", "es:","\n",
  "Limite inferior:",LI,"Limite superior:",LS,"\n")
     }
```

3.8.3. Caso 3: Varianzas poblacionales desconocidas y diferentes

```
# DATOS:
# n1, n2 TAMAÑOS DE LAS MUESTRAS
# prom1, prom2 PROMEDIOS MUESTRALES
# s1, s2 DESVIACIONES ESTANDAR MUESTRALES
```

```
# nc NIVEL DE CONFIANZA (PROPORCIÓN)
#

# Diferencia de medias
## Caso3. IC para dos poblaciones, con varianzas poblacionales desconocidas pero y diferentes
difmed3=function(n1,n2,prom1,prom2,s1,s2,nc=0.95)
     {
  eta=(s1^2/n1+s2^2/n2)^2/((s1^2/n1)^2/(n1-1)+
             (s2^2/n2)^2/(n2-1))
        t=qt(1-(1-nc)/2,eta)
  EE=t*sqrt(s1^2/n1+s2^2/n2)
  LI= prom1-prom2-EE
  LS=prom1-prom2+EE
  cat("Un intervalo de confianza para MU1-MU2","\n",
         "al nivel de confianza",nc*100,"%", "es:","\n",
  "Limite inferior:",LI,"Limite superior:",LS,"\n")
     }
```

3.8.4. Cuociente de varianzas poblacionales $\frac{\sigma_1^2}{\sigma_2^2}$

```
# DATOS:
# s1, s2 DESVIACION ESTANDAR MUESTRALES
# n1, n2 TAMAÑOS DE LAS MUESTRAS
# nc NIVEL DE CONFIANZA (PROPORCION)
#

# Diferencia de medias
## Cuociente de varianzas poblacionales
cuovar=function(n1,n2,s1,s2,nc=0.95)
     {
  f1=qf((1-nc)/2,n1-1,n2-1)
  f2=qf(1-(1-nc)/2,n1-1,n2-1)
  cuo=s1^2/s2^2
  LI=cuo/f2
  LS=cuo/f1
        cat("Un intervalo de confianza para SIGMA1^2/SIGMA2^2",
            "\n","al nivel de confianza",nc*100,"%", "es:","\n",
  "Limite inferior:",LI,"Limite superior:",LS,"\n")
     }
```

3.9. Para diferencia de proporciones $\pi_1 - \pi_2$

```
# DATOS:
#   n1, n2 TAMAÑOS DE LAS MUESTRAS
#   p1, p2 PROPORCIONES MUESTRALES
#   nc NIVEL DE CONFIANZA (PROPORCIÓN)

# IC para diferencia de proporciones pi1 - pi2
difprop=function(n1,n2,p1,p2,nc=0.95){
  z=qnorm(1-(1-nc)/2)
  v1=p1*(1-p1)/n1
  v2=p2*(1-p2)/n2
  EE = z*sqrt(v1+v2)
  LI=p1-p2-EE
  LS=p1-p2+EE
  cat("Un intervalo de confianza para PI1-PI2","\n",
         "al nivel de confianza",nc*100,"%", "es:","\n",
```

```
      "Limite inferior:",LI,"Limite superior:",LS,"\n")
}
```

Observacion:

- $LI < LS < 0 \rightarrow \pi_1 - \pi_2 < 0 \iff \pi_1 < \pi_2$
- $0 < LI < LS \rightarrow \pi_1 - \pi_2 > 0 \iff \pi_1 > \pi_2$
- $LI < 0 < LS \rightarrow \pi_1 - \pi_2 = 0 \iff \pi_1 = \pi_2$

4. Prueba de Hipotesis

Comenzamos con una hipotesis nula H_0, que contrapone a la hipotesis alternativa H_1.

4.1. Tipos de errores

- Tipo I

Este ocurre cuando H_0 es verdadera y la rechazamos a favor de H_1

- Tipo II

Este ocurre cuando no rechazamos la hipotesis nula cuando esta es falsa.

	$H_0 : T$	$H_0 : F$
No rechazar H_0	Correcta	Error Tipo II
Rechazar H_0	Error Tipo 1	Correcta

Se realizan afirmacion sobre el parametro H_0 y H_1

1. $H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$
2. $H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$
3. $H_0 : \theta \geq \theta_0 \quad \text{vs} \quad H_1 : \theta < \theta_0$

Entonces, podemos definir las probabilidades

- $P(\text{Error Tipo I}) = P(\text{Rechazar } H_0 | H_0 \text{ es verdad}) = \alpha$
- $P(\text{Error Tipo II}) = P(\text{No Rechazar } H_0 | H_0 \text{ es mentira}) = \beta(\theta_1)$

4.2. PH para media poblacional

```
## VARIABLES: MU0, Valor de MU en H0
##            H1=c("distinto","mayor","menor")
##            NS, Nivel de Significación (0.05)
##            n, tamaño de la muestra
##            prom, promedio muestral
##            Var, varianza muestral o poblacional
##            POB: TRUE si es Poblacional la varianza
##                 FALSE si es muestral la varianza

PHMU=function(MU0, H1, NS, n, prom, Var,POB=TRUE)
{
        cat("La Hipotesis alternativa es H1: Mu", H1, MU0,"\n")
        if(H1=="distinto")
        {
                if(POB)
                {
```

```
                print("Varianza Poblacional Conocida")
                E0=(prom-MU0)/sqrt(Var/n)
                cat("El valor de la estadística de prueba es:",E0,"\n")
                Z=qnorm(1-NS/2)
                if(abs(E0)>Z) print("SE RECHAZA H0")
                else print("NO SE RECHAZA H0")
                valor.p=2*(1-pnorm(abs(E0)))
                cat("El valor-p vale:",valor.p,"\n")
        }
        else {
                print("Varianza Poblacional Desconocida")
                E0=(prom-MU0)/sqrt(Var/n)
                cat("El valor de la estadística de prueba es:",E0,"\n")
                T=qt(1-NS/2,n-1)
                if(abs(E0)>T) print("SE RECHAZA H0")
                else print("NO SE RECHAZA H0")
                valor.p=2*(1-pt(abs(E0),n-1))
                cat("El valor-p vale:",valor.p,"\n")
                }
}
if(H1=="mayor")
{
        if(POB)
        {
                print("Varianza Poblacional Conocida")
                E0=(prom-MU0)/sqrt(Var/n)
                cat("El valor de la estadística de prueba es:",E0,"\n")
                Z=qnorm(1-NS)
                if(E0>Z) print("SE RECHAZA H0")
                else print("NO SE RECHAZA H0")
                valor.p=1-pnorm(E0)
                cat("El valor-p vale:",valor.p,"\n")
        }
        else {
                print("Varianza Poblacional Desconocida")
                E0=(prom-MU0)/sqrt(Var/n)
                cat("El valor de la estadística de prueba es:",E0,"\n")
                T=qt(1-NS,n-1)
                if(E0>T) print("SE RECHAZA H0")
                else print("NO SE RECHAZA H0")
                valor.p=1-pt(E0,n-1)
                cat("El valor-p vale:",valor.p,"\n")
                }
}
if(H1=="menor")
{
        if(POB)
        {
                print("Varianza Poblacional Conocida")
                E0=(prom-MU0)/sqrt(Var/n)
                cat("El valor de la estadística de prueba es:",E0,"\n")
                Z=qnorm(NS)
                if(E0<Z) print("SE RECHAZA H0")
                else print("NO SE RECHAZA H0")
                valor.p=pnorm(E0)
                cat("El valor-p vale:",valor.p,"\n")
        }
```

```
                else {
                        print("Varianza Poblacional Desconocida")
                        E0=(prom-MU0)/sqrt(Var/n)
                        cat("El valor de la estadística de prueba es:",E0,"\n")
                        T=qt(NS,n-1)
                        if(E0<T) print("SE RECHAZA H0")
                        else print("NO SE RECHAZA H0")
                        valor.p=pt(E0,n-1)
                        cat("El valor-p vale:",valor.p,"\n")
                        }
        }
}
```

Se quiere probar:

1. $H_0 : \mu = \mu_0 \quad \text{vs } H_1 : \mu \neq \mu_0$
2. $H_0 : \mu \leq \mu_0 \quad \text{vs } H_1 : \mu > \mu_0$
3. $H_0 : \mu \geq \mu_0 \quad \text{vs } H_1 : \mu < \mu_0$

Cuando σ^2 (poblacion) es :

1. conocido:
$$\text{Bajo } H_0 : E_0 = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$
2. desconocido:
$$\text{Bajo } H_0 : E_0 = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim t(n-1)$$

5. Ejercicios

5.1. Estimacion Puntual

5.1.1. 6.1

> **Ejercicio 6.1.** *Si el contenido en gramos de un determinado medicamento sigue una distribución normal $N(7.5, 0.3)$, calcular la probabilidad de que en una muestra de tamaño 5 se obtenga que la media es menor que 7.*

Usamos el segundo punto en Distribucion Normal
$X_1, \ldots, X_5$ m.a $N(7{,}5, 0{,}3)$, se quiere $P(\overline{X} < 7)$, donde $\overline{X} \sim N(7{,}5, \frac{0{,}3}{5})$

```
pnorm(7, 7.5, sqrt(0.3/5))
```

5.1.2. 6.2

> **Ejercicio 6.2.** *Una fábrica de pasteles elabora, en su producción habitual, un 3 % de pasteles defectuosos. Un cliente recibe un pedido de 500 pasteles de la fábrica. Calcular la probabilidad de que encuentre más del 5 % de pasteles defectuosos.*

Aplicamos un estimacion con Distribucion Bernoulli, donde nuestra proporcion muestral p sera $\frac{\sum_{i=1}^{500} X_i}{500}$ Queremos calcular, entonces aplicamos la proporcion muestral con $p = 0{,}05$ y $\pi = 0{,}03$ $P(p > 0{,}05) = P(Z > \frac{0{,}05 - 0{,}03}{\sqrt{\frac{0{,}03 \times 0{,}97}{500}}}) = P(Z > 2{,}621613)$

```
## 1-pnorm(2.621613) -- media 0 y ds 1 por defecto, busca la probabilidad que una variable aleatoria norm
1-pnorm(0.05, 0.03, sqrt(0.03*0.97/500)) # queremos las que estan por arriba
```

5.1.3. 6.3

Ejercicio 6.3. *Un ascensor limita el peso de sus 4 ocupantes a 300 kilogramos. Si el peso de un individuo sigue una distribución normal* $N(71,7)$*, calcular la probabilidad de que el peso de 4 individuos supere los 300 kilogramos.*

Usamos el primer punto de Distribucion Normal Queremos calcular $P(\sum_{i=1}^{4} X_i > 300)$ con $\sum_{i=1}^{4} X_i \sim N(4 \cdot 71, 4 \cdot 7)$

```
## pnorm(300, 284, sqrt(28))
1-pnorm(300, 284, sqrt(28))
```

Entonces $P(\sum_{i=1}^{4} X_i > 300) = 0{,}00124845445757082$

5.1.4. 6.4

Ejercicio 6.4. *Se prueba el rendimiento, en kilómetros por litro, de dos tipos de gasolina, encontrándose una desviación estándar de 1.23 kilómetros por litro para la primera gasolina y una desviación estándar de 1.37 kilómetros por litro para la segunda. Se prueba la primera gasolina en 35 vehículos, y la segunda en 42.*

Asumiremos que el rendimiento 1ª gasolina es $N(\mu_1, \sigma_1^2)$
rendimiento 2ª gasolina es $N(\mu_2, \sigma_2^2)$ } Independientes

a) ¿Cuál es la probabilidad de que la primera gasolina dé un rendimiento promedio mayor de 0.45 kilómetros por litro que la segunda gasolina? $P(\overline{X}_1 - \overline{X}_2 > 0.45)$

b) ¿Cuál es la probabilidad de que la diferencia en rendimientos promedio se encuentre entre 0.65 y 0.83 kilómetros por litro a favor de la primera gasolina? $P(0.65 < \overline{X}_1 - \overline{X}_2 < 0.83)$

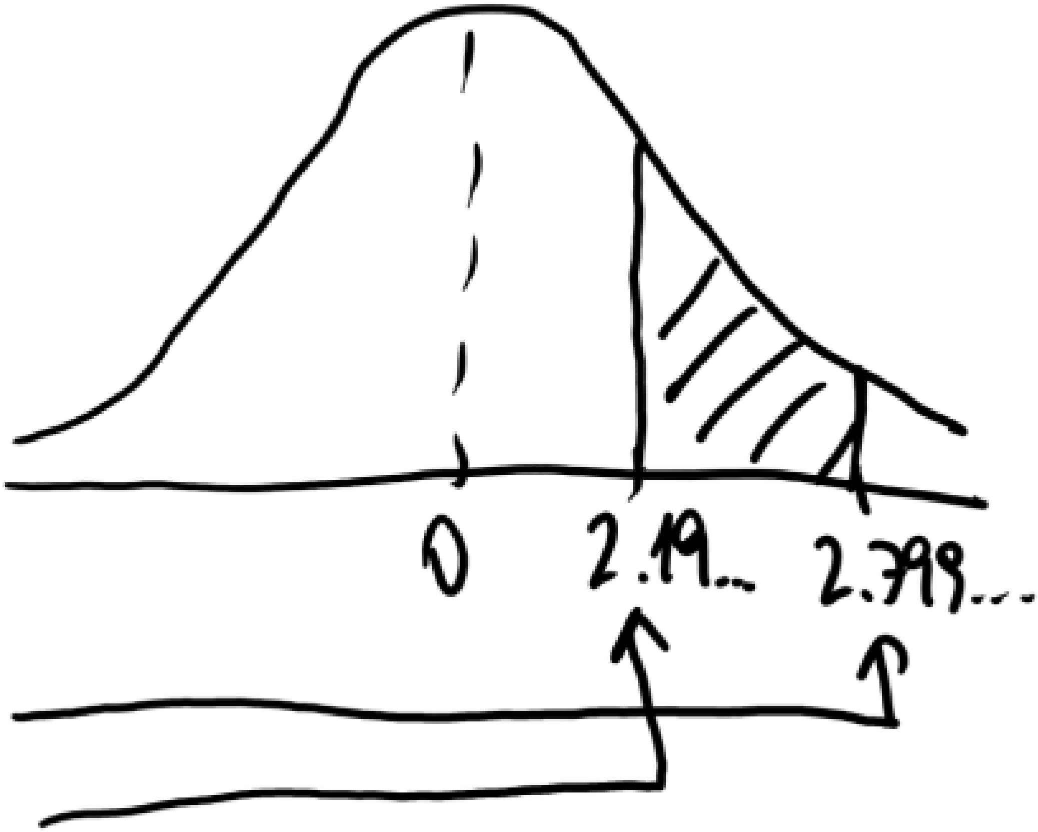
Usamos Caso 1: Varianzas poblacionales conocidas, asumiendo que ambas m.a son independientes y tambien tenemos que $\mu_1 = \mu_2 = \mu$

1. Datos: $\sigma_1 = 1{,}23$, $\sigma_2 = 1{,}37$, $\overline{X_1}$ = rendimiento prom gas 1, $\overline{X_2}$ = rendimiento prom gas 2. $n_1 = 35$, $n_2 = 42$ Queremos calcular $P(\overline{X_1} - \overline{X_2} > 0{,}45)$, aplicamos 1. de Trabajando con dos poblaciones $P(Z > \frac{0{,}45-0}{\sqrt{\frac{1{,}23^2}{35} + \frac{1{,}37^2}{42}}})$, como $\mu = \mu_1 = \mu_2 \rightarrow \mu_1 - \mu_2 = 0$

   ```
   1-pnorm(0.45/sqrt(( 1.23^2/35 ) + ( 1.37^2/42 )))
   ```

2. $P(0{,}65 < \overline{X_1} - \overline{X_2} < 0{,}83) = P(\overline{X_1} < 0{,}83) - P(\overline{X_2} < 0{,}65)$

   ```
   x1 = pnorm(0.83/sqrt(( 1.23^2/35 ) + ( 1.37^2/42 )))
   x2 = pnorm(0.65/sqrt(( 1.23^2/35 ) + ( 1.37^2/42 )))
   x1-x2
   ```



5.1.5. Se sabe que 2 % de las unidades fabricadas por A son defectuosas, y que el 25 % de las fabricadas por B son defectuosas.

Se necesitan 100 unidades de A y 150 de B, Cual es la probabilidad de que la proporcion de defectuosos muestrales de A supere a los de B?

Datos: $\pi_1 = 0{,}02$, $\pi_2 = 0{,}025$, $n_1 = 100$, $n_2 = 150$, p_1 : prop. defectuosos de A p_2 : prop. defectuosos de B Queremos calcular $P(p_1 - p_2 > 0)$

$$P(Z > \frac{0 - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}})$$

```
num=0-(0.02-0.025)
denom=sqrt((0.02*0.98/100)+(0.025*0.975/150))
1-pnorm(num/denom)
```

5.1.6. Si X ˜ N (40,10), calcular Pr (39 X 41) para n=10. ¿En qué intervalo se obtendrán el 95 % de los resultados?

5.1.7. guia .55

Estudios realizados por neurocientíficos del MIT revelan que la melatonina, segregada por la glándula pineal en el cerebro, funciona naturalmente como hormona inductora del sueño (Tampa Tribune, 1 de marzo de 1994). Voluntarios de sexo masculino recibieron distintas dosis de melanina o placebos y luego se colocaron en una habitación oscura a medio día, pidiéndoseles que cerraran los ojos y se durmieran. Lo que interesaba a los científicos del MIT era el tiempo Y (en minutos) que tardaba cada voluntario en quedarse dormido. Los investigadores determinaron que con el placebo (es decir, sin hormona), el tiempo medio para dormirse era de 15 minutos. Supón que con el tratamiento de placebo mu = 15 y sigma = 5. (a) Considera una muestra aleatoria de n = 20 hombres que reciben la hormona inductora del sueño. Sea y el tiempo medio en quedarse dormido en esta muestra. Si la hormona no es eficaz para inducir el sueño, describe la distribución de muestreo de y . (b) Calcula en el caso (a) p(y ≤ 6)

1. $n = 20, \mu = 15, \sigma = 5$. Por el teorema central del limite tenemos que

$$\overline{y} \sim N(\mu = 15, \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{20}} \approx 1{,}118)$$

2. $\mu = 15, \sigma = 5, n = 20, EE = \frac{\sigma}{\sqrt{n}}$

```
pnorm(6, 15, 5/sqrt(20))
```

5.2. Intervalos de Confianza

5.2.1. 6.5

Se ha comprobado que la concentración promedio de zinc que se saca del agua de un río a partir de una muestra de mediciones de zinc en 36 sitios diferentes es de 2.6 gramos por mililitro. Encontrar los intervalos de confianza del 95 % y 99 % para la concentración media de zinc en el río, suponiendo que la desviación típica de la población es 0.3.

Para μ con σ o σ^2 conocido

En primer caso con $\alpha = 0{,}05$

```
ICMU1(36, 2.6, 0.3)
```

En segundo caso con $\alpha = 0{,}01$

```
ICMU1(36, 2.6, 0.3, 0.99)
```

5.2.2. 6.6

Determinar un intervalo de confianza al nivel = 0.05 para la probabilidad de que un recién nacido sea niño, si en una muestra de tamaño 123 se han contabilizado 67 niños.

Tenemos $p = \frac{67}{123} = 0{,}54$ y $n = 123$.

```
ICPI(123, 0.54, 0.95)
```

5.2.3. 6.7

El encargado del departamento de producción de una fábrica recibe un lote de 2000 piezas necesarias para el montaje de un artículo. El fabricante de las piezas asegura que en este lote no hay más de 100 piezas defectuosas.

1. ¿Cuántas piezas hay que examinar para que, con un nivel de confianza del 95 %, el error que se cometa en la estimación de la proporción de piezas defectuosas no sea mayor que 0.05?

   Necesitamos calcular el tamano de la muestra, usamos `NPI()`, ademas tenemos que $p = \frac{100}{2000} = 0{,}05$, el error maximo $E = 0{,}05$ y $nc = 0.95$

   ```
   NPI(0.05, 0.95, 0.05)
   ```

2. Si se toma una muestra de 100 piezas elegidas al azar y se encuentran 4 defectuosas, determinar un intervalo de confianza para la proporción de defectuosas al nivel del 95 %.

   ```
   ICPI(100, 0.04, 0.95)
   ```

5.2.4. 6.8

El peso de los terneros de una granja se distribuye normalmente, con desviación típica de 10 kilogramos. Se toma al azar una muestra de 35 de ellos para transportarlos en un camión. Sabiendo que el peso medio resulta ser de 140 kilogramos, determinar un intervalo de confianza al 8 % de nivel de significación en el que oscilará el peso de los 35 terneros.

$X \sim N(\mu, 10)$. La suma de los pesos es $35X$, entonces $N(35 \cdot 140, 10\sqrt{35})$. Asi, tenemos que

$$IC = n\overline{X} \pm Z_{\frac{\alpha}{2}}\sigma\sqrt{n} = 35 \cdot 140 \pm 1{,}75 \cdot 10\sqrt{35} = 4900 \pm 103{,}53$$

5.2.5. 6.9

Dos compañías A y B fabrican el mismo tipo de cable. Un distribuidor desea conocer la diferencia promedio de la resistencia a la rotura de los mismos, para lo cual toma muestras de 100 cables de A y 50 cables de B. La muestra de los cables de la compañía A arroja una resistencia promedio a la rotura de 4500 kilogramos, mientras que los cables de la compañía B arrojan una resistencia promedio a la rotura de 4000 kilogramos. Se sabe, por experiencia, que la desviación típica de la resistencia a la rotura es de 300 kilogramos para la compañía A y de 200 kilogramos para la compañía B. Se pide estimar, con un nivel de confianza del 95 %, el intervalo de confianza de la diferencia de medias de la resistencia a la rotura entre los dos cables, si la resistencia a la rotura se distribuye normalmente para ambas compañías.

Comp. A: $n = 100, \overline{X}_A = 4500, \sigma_A = 300$ Comp. B: $n = 50, \overline{X}_B = 4000, \sigma_B = 200$ $nc = 0{,}95$

```
difmed1(n1=100, n2=50, prom1=4500, prom2=4000, sigma1=300, sigma2=200)
```

5.2.6. guia .1

El peso medio de los estudiantes secundarios sigue una distribución normal. Se toma una muestra aleatoria de 100 estudiantes y se obtiene una media de 65 kg con una desviación estándar de 9kg. Encuentre los límites para intervalos de confianza al 95 % y 99 % para:

1. La media poblacional si se sabe que la desviación estándar poblacional es de 10 kg $n = 100, \overline{x} = 65, s = 9, \sigma = 10$

   ```
   ICMU1(100, 65, 10)
   ```

   ```
   ICMU1(100, 65, 10, 0.99)
   ```

2. La media poblacional si no se conoce la desviación estándar poblacional. Además construya un intervalo de confianza al 95 % para la varianza poblacional

   ```
   ICMU2(100, 65, 9, 0.95)
   ```

   ```
   ICMU2(100, 65, 9, 0.99)
   ```

   ```
   ICSIGMA2(100, 9)
   ```

5.2.7. guia .4

En víspera de elecciones presidenciales se toma una muestra aleatoria de 1000 electores, de los cuales 628 dicen estar indecisos todavía. Se pide entonces un intervalo al 98 % de confianza de la proporción de personas que no saben aún por quien votar. $n = 1000, p = \frac{628}{1000} = 0{,}628$

```
ICPI(1000, 0.628, 0.98)
```

5.2.8. guia .6

Se debe estimar el grosor de las láminas de vidrio producidas en cierta fábrica. Se toma una muestra aleatoria de tamaño 100 y se encuentra un grosor promedio de 20mm. Suponiendo que se conoce la varianza poblacional y es igual a 1.44mm^2, se pide encontrar un intervalo de confianza del 95 % de confianza para el espesor promedio de las láminas de vidrio. $n = 100, \overline{x} = 20, \sigma^2 = 1{,}44$

```
ICMU1(100, 20, sqrt(1.44)) # sigma = \sqrt{sigma^2}
```

5.3. Prueba Hipotesis

5.3.1. fabricantes de herramientas

En nuestro caso: μ = tiempo medio de preparación por hora de operación

Un fabricante está considerando la adquisición de un nuevo equipo para la fabricación de herramientas y especifica que en promedio el equipo no debe requerir más de 10 min de tiempo de preparación por hora de operación. Un agente de compras visita una compañía en donde existe instalado otro equipo como el que se está considerando. En los registros que allí se tienen, se observa una muestra aleatoria de 40 horas de operación que incluye un total de 7 horas y 30 minutos de tiempo de preparación, con una desviación estándar de tiempo de preparación por hora de 3 minutos. Basándose en este estudio muestral ¿Podría decir que el equipo satisface las especificaciones de tiempo de preparación, con un nivel de significación de 1%?

$n = 40$ $\sum_{i=1}^{40} x_i = 450$ min $s = 3$ min $H_0: \mu \leq 10$ mins $(=)$ $H_1: \mu > 10$ min σ^2 es desconocido

```r
PHMU(MU0 = 10, H1 = "mayor", NS=0.01, n = 40, prom = 11.25, Var = 9, POB = FALSE)
```

$prom = \frac{450}{40}$

```
La Hipotesis alternativa es H1: Mu mayor 10
[1] "Varianza Poblacional Desconocida"
El valor de la estadística de prueba es: 2.635231
[1] "SE RECHAZA H0"
El valor-p vale: 0.006000591
```

Con un nivel de confianza del 99 %, rechazamos H_0 de que el tiempo promedio de preparacion es $\leq$ 10 minutos por hora. El equipo no cumple con la especificacion de tiempo de preparacion, ya que el tiempo prom observado 11.25 min/hora es mayor que el limite 10min/hora. Ademas, $p \approx 0{,}006 < \alpha = 0{,}01$