

Laboratorio N.1
Introduccion a Los Metodos Estadisticos
Estimadores

Diana Carolina Arias Sinisterra Cod. 1528008
Kevin Steven Chica Garcia Cod. 1533173
Cesar Andres Saavedra Vanegas Cod. 1628466

Universidad Del Valle
Facultad De Ingenieria
Estadistica
Octubre
2017

Índice

1. Situacion 1	3
1.1.	4
1.2. Scripts R	4
1.3. Analisis Resultados	6
2. Situacion 1	6
3. Situacion 2	10
3.1. Punto A.	10
3.2. Punto B.	10
3.3. Punto C.	11
4. Situacion 3	12
4.1. Punto A.	12
4.2. Punto B.	12
5. Situacion 5	13
5.1. Punto A.	13
5.2. Punto B.	13
5.3. Punto C.	14

Índice de figuras

1. Titulo de la figura	3
2. Grafica del resultado de la estimacion del parametro real $\theta = 1200$	6
3. Grafica del sesgo de los distintos estimadores variando los tamaños de muestra . . .	7
4. Grafica de la varianza de los distintos estimadores variando los tamaños de muestra	8
5. Grafica del ECM de los distintos estimadores variando los tamaños de muestra . . .	9
6. Grafica de probabilidades de detectar el lote segun la media μ	14

Indice de tablas

1. Titulo de la tabla	3
---------------------------------	---

1. Situación 1

Despliegue de una tabla que se llama “Tabla1” junto a su número de referencia:

Tabla 1: Título de la tabla

Indicador	SO ₂	Temperatura
Promedio	0.006	24.093
Desviación estándar	0.007	2.950
Mediana	0.004	23.500
Minima	0.000	18.200
Máximo	0.127	32
Asimetría	5.415	0.451

Para referenciar una tabla, se utiliza el texto “ref” precedido de un *backslash*: así, se referenciará a la Tabla 1.

Despliegue de una imagen que se llama “Figura1” junto a su número de referencia:

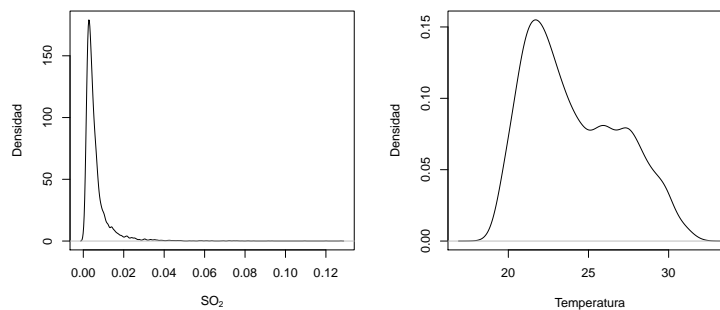


Figura 1: Título de la figura

Para referenciar una figura, se utiliza el texto “ref” precedido de un *backslash*: así, se referenciará a la Figura 6.

Expresión matemática en línea con el texto: $f(x) := ax^2 + bx + c$.

Representación de una ecuación, sin número de referencia:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Representación de una ecuación en una línea nueva, con número de referencia:

$$\bar{X} = \sum_{i=1}^n \frac{x_i}{n} \tag{1}$$

Para referenciar la ecuación (1). Se utiliza la etiqueta “eqref” precedido de un *backslash*

Para referenciar una cita bibliográfica se utiliza un archivo "Bibliografia.bib". Este contiene la información de las referencias utilizadas. Por ejemplo para citar dentro del texto: Según ? plantea que el modelo de regresión.....

1.1.

$$\left(\frac{365}{3}\right) \cdot \left(\frac{365}{4}\right)$$

Una lista:

- primer ítem de la lista.
- segundo ítem de la lista.

Número de una lista:

1. primer ítem de la lista.
2. segundo ítem de la lista.

Ejemplo para construir una matriz:

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

1.2. Scripts R

El ambiente *Verbatim* permite agregar código de R.

```
getwd()
k=70
cumpledif=1
p=0
q=0
for (j in 1:k){
  cumpledif=cumpledif*(1-((j-1)/365))
  q[j]=cumpledif
  p[j]=1-cumpledif
  cat(cumpledif,"\n")
}

windows()
pdf("Texmaker/Graficos/Punto2.pdf")
par(mfrow=c(1,3))
plot(p,col="red", xlab = "n", ylab = "Probabilidad", main = "Probabilidad de que dos o mas
tengan el mismo cumpleaños en funcion
de la cantidad de estudiantes n")
```

```

abline(h=0.5)
abline(v=23)

plot(q,col="blue", xlab = "n", ylab = "Probabilidad", main = "Probabilidad de que dos o mas
NO tengan el mismo cumpleai2os en funcion
      de la cantidad de estudiantes n")
abline(h=0.5)
abline(v=23)

plot(p,col="red", xlab = "", ylab = "")
par(new=TRUE)
plot(q,col="blue", xlab = "", ylab = "")

title(xlab = "n", ylab = "Probabilidad",main="Probabilidades de que dos o mas estudiantes
tengan el mismo cumpleai2os (rojo) y su
evento complementario (azul) en funcion de n")
dev.off()

#####simulacion punto 3a Exponencial

U = runif(1000, 0, 1) #Generar U
l = 4 #Parii2ametro de la exponencial Lambda = 4
X = -(1/l) * log(U)
windows()
pdf("Texmaker/Graficos/Punto3A.pdf")
par(mfrow=c(1,2))
plot(density(U),col="blue", xlab = "X", ylab = "Densidad", main = expression(Unif(0,1)))
plot(density(X),col="red", xlab = "X", ylab = "Densidad", main = expression(Exp(lambda=4)))
dev.off()

##### Punto 3b Poisson
x=0
for(j in 1:1000){
  lambda=7
  i = 0
  p = exp(-lambda)
  f = p
  u= runif(1000,0,1)
  while(u>=f){
    p=lambda*p/(i + 1)
    f = f + p
    i = i +1
    x[j] = i
  }
}

```

1.3. Analisis Resultados

2. Situacion 1

Los estimadores propuestos son:

$$\hat{\theta}_1 = 2\bar{X} - 1$$

$$\hat{\theta}_2 = X_{Max} + X_{Min} - 1$$

$$\hat{\theta}_3 = X_{Max}$$

$$\hat{\theta}_4 = Me(X)$$

Para evaluar los estimadores anteriores obtuvimos las siguientes cuatro graficas:

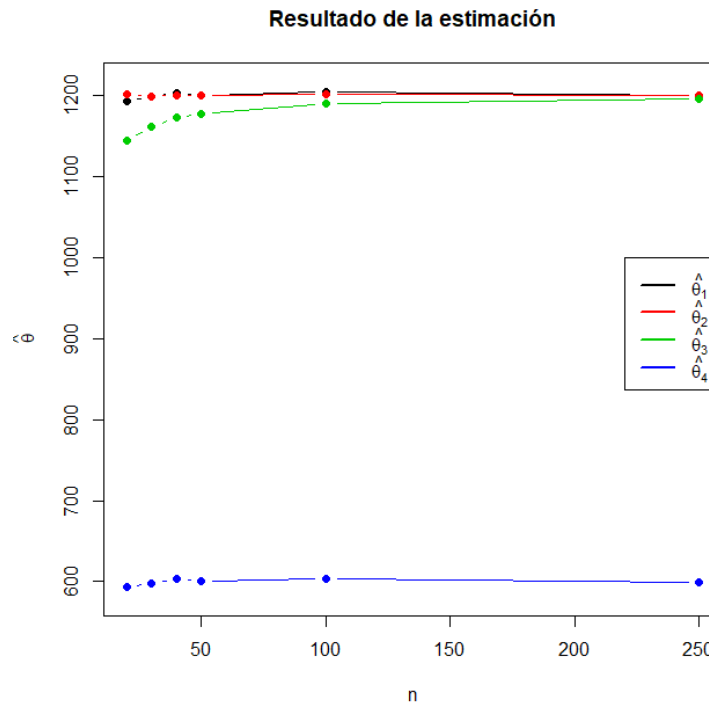


Figura 2: Grafica del resultado de la estimacion del parametro real $\theta = 1200$

Como podemos observar en la grafica, los estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ son los que mas se acercan al parametro real $\theta = 1200$ con cualquier tamaño de muestra, aunque el estimador $\hat{\theta}_3$ tambien se acerca pero despues de tamaño de muestra $n = 100$, lo cual nos dice que son mejores los dos anteriores. El estimador $\hat{\theta}_4$ esta muy lejos del parametro real, a simple vista podriamos decir que ese estimador serviria no para estimar θ en si, si no, para estimar la media de la distribucion en general que es

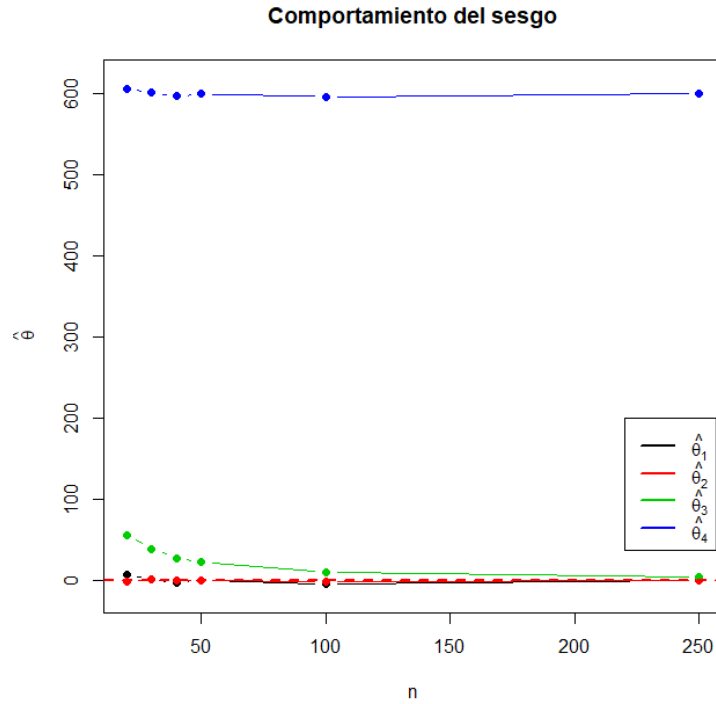


Figura 3: Grafica del sesgo de los distintos estimadores variando los tamaños de muestra

aproximadamente 600.

En esta grafica vemos que al igual que en la anterior, los estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ son los mejores, ya que tienen un sesgo casi igual a 0 para todos los tamaños de muestra, aunque vemos tambien que el sesgo del estimador $\hat{\theta}_3$ tambien tiende a 0 a medida que se aumenta los tamaños de muestra. El estimador $\hat{\theta}_4$ es el menos acertado ya que su sesgo es muy elevado. En general, basandonos en esta grafica, los mejores estimadores hasta ahora son $\hat{\theta}_1$ y $\hat{\theta}_2$

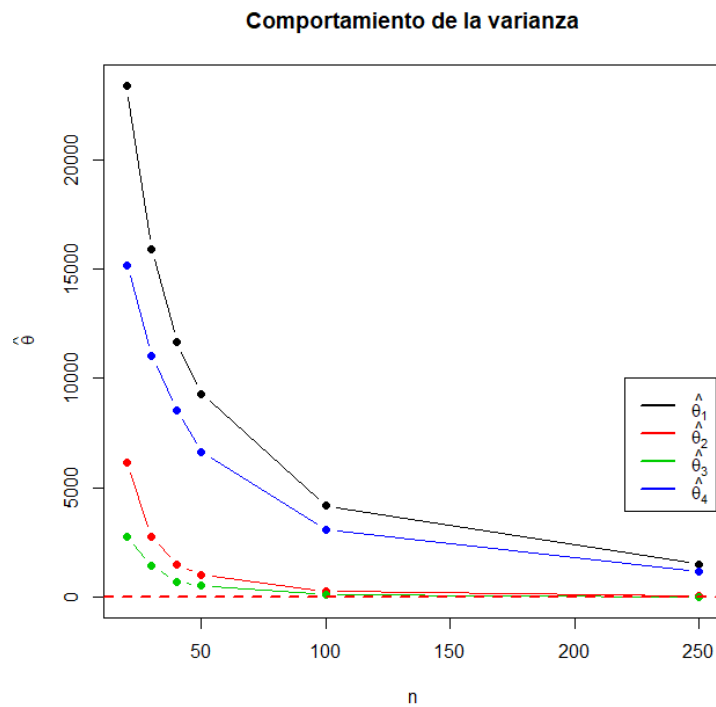


Figura 4: Grafica de la varianza de los distintos estimadores variando los tamaños de muestra

En esta grafica podemos observar que los estimadores $\hat{\theta}_2$ y $\hat{\theta}_3$ tienen menor varianza que los otros dos, pero notemos, que la diferencia de varianzas entre $\hat{\theta}_2$ y $\hat{\theta}_3$ es casi la mitad a favor de $\hat{\theta}_3$ en los tamaños de muestra mas pequeños, indicandonos esto, que el estimador de menor varianza es efectivamente $\hat{\theta}_3$.

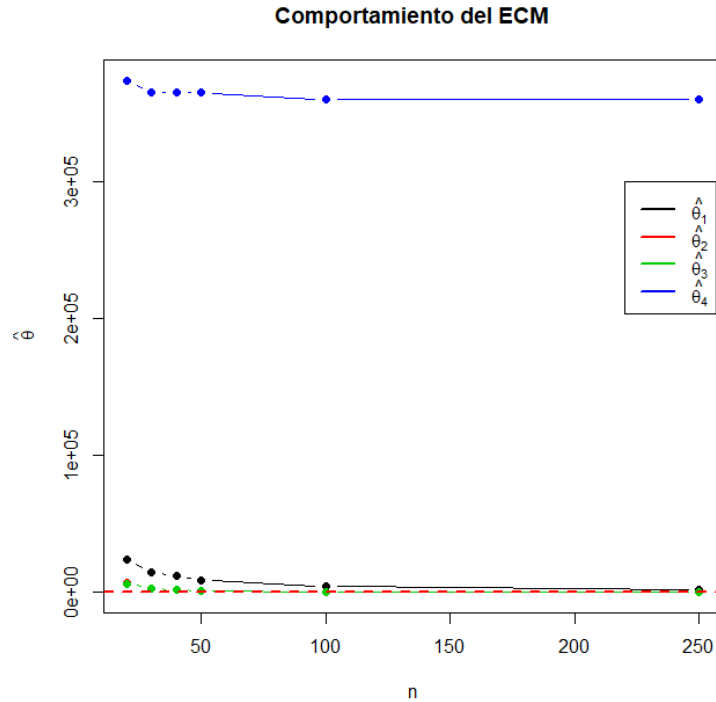


Figura 5: Grafica del ECM de los distintos estimadores variando los tamaños de muestra

En la grafica anterior, notamos que el menor Error cuadratico medio lo presentan los estimadores $\hat{\theta}_2$ y $\hat{\theta}_3$, con casi el mismo en todos los tamaños de muestra, por esa razon no logramos ver la linea y los puntos que representan el estimador $\hat{\theta}_2$, pero en el tamaño de muestra $n=5$ se alcanza a observar una pequeña porcion del punto rojo que representa el estimador $\hat{\theta}_2$, obstruido por el punto verde que representa el estimador $\hat{\theta}_3$, lo que nos asegura que no hay diferencias significativas entre estos estimadores con respecto al ECM. Entonces, para decidir cual de los dos estimadores es mejor, debemos ver las 4 graficas en conjunto, ya que el ECM no me deja claro esto, por ser tan parecido en ambos.

En conclusion, tomando en cuenta las 4 graficas, podriamos decir que para tamaños de muestra menor a 100, $\hat{\theta}_2$ es el mejor estimador para $\theta = 1200$ ya que su sesgo es menor que el de $\hat{\theta}_3$, su estimacion es mas cercana al valor real del parametro y ademas su varianza no es muy alta.

3. Situacion 2

3.1. Punto A.

Si $\hat{\theta}_1$ y $\hat{\theta}_2$ son dos estimadores insesgados tales que:

$$\hat{\theta}_3 = a\hat{\theta}_1 + (1 - a)\hat{\theta}_2$$

Aplico Esperanza a ambos lados

$$E[\hat{\theta}_3] = E[a\hat{\theta}_1 + (1 - a)\hat{\theta}_2]$$

$$E[\hat{\theta}_3] = E[a\hat{\theta}_1] + E[(1 - a)\hat{\theta}_2]$$

$$E[\hat{\theta}_3] = aE[\hat{\theta}_1] + (1 - a)E[\hat{\theta}_2] \text{ Con } \hat{\theta}_1 \text{ y } \hat{\theta}_2 \text{ estimadores insesgados}$$

$$E[\hat{\theta}_3] = a\theta + (1 - a)\theta$$

$$E[\hat{\theta}_3] = a\theta + \theta - a\theta$$

$$E[\hat{\theta}_3] = \theta$$

$\therefore \hat{\theta}_3$ Es un estimador insesgado.

3.2. Punto B.

El coeficiente de variacion de $\hat{\theta}_3$ es

$$CV[\hat{\theta}_3] = \frac{\sqrt{Var[\hat{\theta}_3]}}{E[\hat{\theta}_3]}$$

Hallamos la $Var[\hat{\theta}_3]$

$Var[\hat{\theta}_3] = Var[a\hat{\theta}_1 + (1 - a)\hat{\theta}_2] + 2Cov[a\hat{\theta}_1, (1 - a)\hat{\theta}_2]$, ya que no sabemos si $\hat{\theta}_1$ y $\hat{\theta}_2$ son independientes

Entonces, distribuyendo la Varianza y por las propiedades de Covarianza:

$$Var[\hat{\theta}_3] = Var[a\hat{\theta}_1] + Var[(1 - a)\hat{\theta}_2] + 2a(1 - a)Cov[\hat{\theta}_1, \hat{\theta}_2]$$

$$Var[\hat{\theta}_3] = a^2Var[\hat{\theta}_1] + (1 - a)^2Var[\hat{\theta}_2] + (2a - 2a^2)Cov[\hat{\theta}_1, \hat{\theta}_2]$$

Como $Var[\hat{\theta}_1] = \sigma_1^2$ y $Var[\hat{\theta}_2] = \sigma_2^2$ entonces

$$Var[\hat{\theta}_3] = a^2[\sigma_1^2] + (1 - a)^2[\sigma_2^2] + (2a - 2a^2)Cov[\hat{\theta}_1, \hat{\theta}_2]$$

Sabemos que $Cov[\hat{\theta}_1, \hat{\theta}_2] = E[\hat{\theta}_1\hat{\theta}_2] - E[\hat{\theta}_1] * E[\hat{\theta}_2]$, entonces:

Como $E[\hat{\theta}_1] = \theta$ y $E[\hat{\theta}_2] = \theta$ tenemos:

$$Cov[\hat{\theta}_1, \hat{\theta}_2] = E[\hat{\theta}_1\hat{\theta}_2] - \theta^2, \text{ por tanto:}$$

$$Var[\hat{\theta}_3] = a^2[\sigma_1^2] + (1 - 2a + a^2)[\sigma_2^2] + (2a - 2a^2)(E[\hat{\theta}_1\hat{\theta}_2] - \theta^2)$$

Como sabemos que $E[\hat{\theta}_3] = \theta$

∴ El coeficiente de variacion para $\hat{\theta}_3$ es:

$$CV[\hat{\theta}_3] = \frac{\sqrt{a^2\sigma_1^2 + (1-a)^2\sigma_2^2 + (2a-2a^2)(E[\hat{\theta}_1\hat{\theta}_2] - \theta^2)}}{\theta}$$

3.3. Punto C.

Como en este punto sabemos que $\hat{\theta}_1$ y $\hat{\theta}_2$ son independientes, entonces la $Cov[\hat{\theta}_1, \hat{\theta}_2] = 0$ y por tanto:

$$Var[\hat{\theta}_3] = Var[a\hat{\theta}_1 + (1-a)\hat{\theta}_2]$$

Distribuyendo la varianza y aplicando sus propiedades, nos queda:

$$Var[\hat{\theta}_3] = Var[a\hat{\theta}_1] + Var[(1-a)\hat{\theta}_2]$$

$$Var[\hat{\theta}_3] = a^2Var[\hat{\theta}_1] + (1-a)^2Var[\hat{\theta}_2]$$

$$Var[\hat{\theta}_3] = a^2\sigma_1^2 + (1-a)^2\sigma_2^2$$

Note que $Var[\hat{\theta}_3]$ se convierte en una funcion que depende de a , ya que, σ_1^2 y σ_2^2 son conocidas, entonces, debemos encontrar a , que haga minima dicha funcion, en otras palabras, todo se reduce a encontrar el minimo de la funcion. Para ello procedemos de la siguiente manera:

1. Encontramos la primera derivada de la funcion con respecto a la variable a :

$$f'(a) = 2a\sigma_1^2 - 2(1-a)\sigma_2^2$$

2. Igualamos el resultado de la primera derivada a 0 y despejamos la variable que nos interesa obtener, en este caso, despejamos a :

$$2a\sigma_1^2 - 2(1-a)\sigma_2^2 = 0$$

$$2a\sigma_1^2 - (2-2a)\sigma_2^2 = 0$$

$$2a\sigma_1^2 - 2\sigma_2^2 + 2a\sigma_2^2 = 0$$

$$2a\sigma_1^2 + 2a\sigma_2^2 = 2\sigma_2^2$$

$$a(2\sigma_1^2 + 2\sigma_2^2) = 2\sigma_2^2$$

$$a = \frac{2\sigma_2^2}{2\sigma_1^2 + 2\sigma_2^2}$$

$$\therefore a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

3. Ahora, debemos encontrar la segunda derivada y ver si es positiva o negativa para saber si encontramos un minimo o un maximo:

$$f''(a) = 2\sigma_1^2 + 2\sigma_2^2 > 0$$

Como nos dio que la segunda derivada parcial es siempre positiva, concluimos que el a hallado anteriormente es un minimo. Por lo tanto, para hacer que la $Var[\hat{\theta}_3]$ sea minima, debemos escoger a como $a = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$.

4. Situacion 3

4.1. Punto A.

Sea X una distribucion Poisson(λ) con, $n = 30$ y $E[x] = (\lambda)$

$$M'_1 = \frac{1}{30} \sum_{i=1}^{30} x_i$$

$$\mu'_1 = \lambda$$

$$M'_1 = \frac{1}{30} \sum_{i=1}^{30} x_i = \lambda = \mu'_1$$

$$\therefore \hat{\lambda} = \frac{1}{30} \sum_{i=1}^{30} x_i = \bar{X}$$

Ahora es necesario probar si es insesgado:

$$E[\hat{\lambda}] = E\left[\frac{1}{30} \sum_{i=1}^{30} x_i\right]$$

$$E[\hat{\lambda}] = \frac{1}{30} E\left[\sum_{i=1}^{30} \lambda_i\right]$$

$$E[\hat{\lambda}] = \frac{1}{30} 30\lambda$$

$$E[\hat{\lambda}] = \lambda$$

\therefore Un estimador insesgado para λ es $\hat{\lambda} = \frac{1}{30} \sum_{i=1}^{30} x_i = \bar{X}$

4.2. Punto B.

Sea $C = 3X + X^2$

$$E[C] = E[3X + X^2]$$

$$E[C] = E[3X] + E[X^2]$$

$$E[C] = 3E[X] + E[X^2]$$

Sabiendo que:

$$V[X] = E[X^2] - E^2[X]$$

Entonces:

$$E^2[X] = V[X] + E[X^2]$$

$$\therefore E^2[X] = \lambda + (\lambda)^2$$

Reemplazando $E^2[X]$ en:

$$E[C] = 3\lambda + E[X^2]$$

Obtenemos que: $E[C] = 3\lambda + \lambda + (\lambda)^2$

$$\therefore E[C] = 4\lambda + (\lambda)^2$$

5. Situacion 5

5.1. Punto A.

Tenemos la siguiente informacion del problema:

$\mu = 3800$ $\sigma = 400psi$ $n = 4$ y $\bar{X} = 3334$ el cual es el valor de la media sobre la cual se toma la decision.

Para responder a la pregunta sobre el riesgo que corre el proveedor bajo este criterio de decision, decidimos calcular la probabilidad de que $\bar{X} \leq 3334$, entonces:

Sabemos que $\bar{X} \sim N(\mu_x, \frac{\sigma}{\sqrt{n}})$, en este caso $\bar{X} \sim N(\mu_x = 3800, \frac{\sigma=400}{\sqrt{4}}) = (\mu_x = 3800, \sigma = 200)$

Entonces:

$P(\bar{X} \leq 3334) = P(Z \leq \frac{3334-3800}{200}) = P(Z \leq \frac{-466}{200}) = P(Z \leq -2.33) = 0.0099$ (Utilizando la tabla de la normal estandar)

Segun lo anterior, logramos ver que la probabilidad de que la media sea menor que 3344 es tan solo de 0.99%, entonces concluimos que el riesgo que corre el proveedor de que el cliente les devuelva un lote bajo esta regla de decision es muy bajo, de hecho, segun los resultados, de cada 100 lotes, solo le devolvera uno aproximadamente. El riesgo lo corre mas bien el comprador, ya que este, con la regla de decision establecida no va a lograr concluir casi nunca que el proceso esta fuera de control cuando en realidad lo esta. Por ultimo, viendolo en forma de errores tipo 1 y 2, concluimos que esta regla de decision cuenta con una probabilidad de cometer error tipo 2 muy alta, es decir, la mayoria de veces va a concluir que el proceso esta bajo control, cuando en realidad esta fuera de control.

5.2. Punto B.

Si se cambia la media a $\mu = 3700$, la probabilidad que existe de que un lote sea detectado bajo estas condiciones es la siguiente:

$P(\bar{X} \leq 3334) = P(Z \leq \frac{3334-3700}{200}) = P(Z \leq \frac{-366}{200}) = P(Z \leq -1.83) = 0.0336$ (Utilizando la tabla de la normal estandar)

Con lo anterior, podemos observar que la probabilidad de que un lote sea detectado bajo esas condiciones es de 3.36%, un poco mas alto que en el punto anterior, esto nos dice que entre mas cerca este μ del valor de la media muestral \bar{X} , mas alta va a ser esta probabilidad, y por consiguiente, las posibilidades de que el cliente detecte el lote van a ser mayores.

5.3. Punto C.

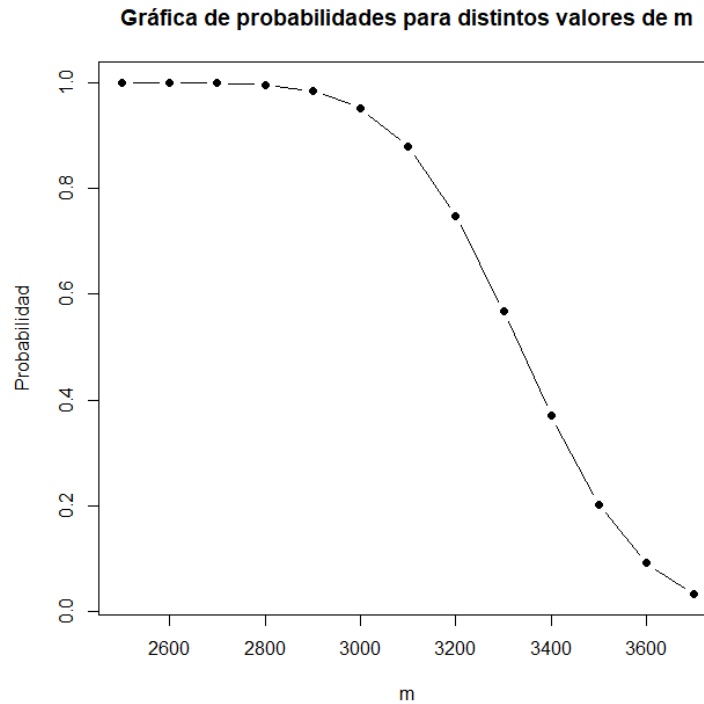


Figura 6: Grafica de probabilidades de detectar el lote segun la media μ

En esta grafica podemos ver que mientras el valor de la media μ aumenta, la probabilidad de detectar el lote va disminuyendo, esto pasa, porque la resistencia media con la cual se dice que el proceso de produccion esta bajo control es de 3800 psi, entonces a medida que la media muestral aumenta, nos vamos acercando mas a la media teorica (3800 en este caso), lo que hace menos probable que se devuelva un lote de varillas metalicas, ya que mientras mas cercana sea la media muestral a la poblacional, tenemos mas certeza para concluir que las varillas se estan haciendo con el proceso adecuado o bajo control.